

# Sesión 4

## Inferencia estadística

### 4.1. Planteamiento del problema

Se tiene un fenómeno aleatorio el cual queremos modelar matemáticamente para hacer inferencias sobre el comportamiento futuro del fenómeno estudiado.

Para dar solución a este problema se supondrá lo siguiente:

- El fenómeno aleatorio sigue una distribución de probabilidad conocida salvo ciertos parámetros  $f_X(x; \underline{\theta})$  con  $\underline{\theta} := (\theta_1, \dots, \theta_p)$  vector de parámetros desconocidos.
- Bajo el supuesto de que conocemos el modelo adecuado para nuestro problema, toda nuestra tarea se centrará en estimar al vector de parámetros  $\underline{\theta}$

Una vez determinado el vector de parámetros  $\underline{\theta}$  se tiene un modelo completamente definido y por tanto se puede hacer cualquier tipo de inferencia distribucional de los datos.

Con el fin de dar solución al problema de estimación definamos lo siguiente:

**Definición 4.1.1** (Espacio parametral). *Al conjunto de posibles valores que puede tomar los parámetros de un modelo se le conoce como espacio parametral y se le denota como  $\Theta$*

**Ejemplo 4.1.1.** *Supongamos que tenemos un fenómeno aleatorio que se puede modelar con una distribución  $N(1, \sigma^2)$  con  $\sigma^2$  desconocido. Es decir la función de densidad de nuestra población está dada por:*

$$f_X(x; \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-1)^2}$$

Entonces:

$$\Theta = \{\sigma^2 \in \mathbb{R} : \sigma^2 > 0\}$$

Observe que si conociéramos  $\sigma^2$  el modelo estaría completamente definido lo que nos permitiría por ejemplo obtener  $\mathbb{P}(X > 1)$  donde  $X$  es una observación futura del fenómeno que estamos modelando.

Nuestro objetivo es entonces estimar parámetros, esto lo podemos hacer de tres formas:

1. Estimación Puntual.
2. Estimación Intervalar.
3. Pruebas de Hipótesis.

A continuación se estudia a profundidad cada una de las técnicas de estimación.

## 4.2. Estimación Puntual

Dado que nuestro objetivo es estimar a los parámetros, necesitaremos herramientas que nos ayuden a obtener una buena estimación, en estadística nuestra principal fuente de información será una muestra de la población que pretendemos modelar. En la parte teórica supondremos que dicha muestra aún no la hemos observado y por tanto, serán variables aleatorias. Formalmente definiremos a una muestra aleatoria como:

**Definición 4.2.1** (Muestra Aleatoria). *Supongamos que tenemos una población que es modelada por una función de densidad dada por  $f_X(x; \underline{\theta})$  con  $\underline{\theta}$  un vector de parámetros desconocidos. Decimos que  $X_1, \dots, X_n$  es una muestra aleatoria (denotada por m.a.) de tamaño  $n$  si:*

- Cada  $X_i$  sigue la misma distribución dada por la función de densidad  $f_X(x; \underline{\theta})$
- $X_i$  es independiente de  $X_j$  con  $i \neq j$

En inferencia clásica  $\underline{\theta}$  se considera un vector de parámetros fijos pero desconocidos y por lo tanto no tiene asociada ningún tipo de distribución de probabilidad.

Nuestro objetivo es entonces encontrar métodos eficientes que ayuden a estimar de la mejor manera a los parámetros desconocidos. Para atacar este problema se define lo siguiente:

**Definición 4.2.2** (Estadística). *Sea  $X_1, \dots, X_n$  m.a. de cierto fenómeno aleatorio, una estadística es cualquier función de la muestra que no depende de parámetros desconocidos*

$$t_1 = t_1(X_1, \dots, X_n)$$

Ejemplo de estadísticas pueden ser:

$$\begin{aligned} t_1 &= \frac{\sum_{i=1}^n X_i}{n} \\ t_2 &= \max\{X_1, \dots, X_n\} \\ t_3 &= \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} \\ t_4 &= \left( \frac{\sum_{i=1}^n X_i}{n}, \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} \right) \end{aligned}$$

**Definición 4.2.3** (Estimador). *Un estimador, es una estadística, que esperamos cumpla con ciertas propiedades que ayuden a la estimación de un vector de parámetros desconocidos.*

La notación utilizada para identificar a un estimador de cierto vector de parámetros  $\underline{\theta} \in \mathbb{R}^p$  es:  $\hat{\underline{\theta}}$ , esto de cierta forma es la abreviación de:

$$\hat{\underline{\theta}} = \hat{\underline{\theta}}(X_1, \dots, X_n) = \left( \hat{\theta}_1(X_1, \dots, X_n), \dots, \hat{\theta}_p(X_1, \dots, X_n) \right)$$

El problema general implica estimar varios parámetros, sin embargo supondremos en la mayor parte de este curso que solo tenemos un parámetro desconocido en cuyo caso:

$$\hat{\theta} = \hat{\theta}(X_1, \dots, X_n) \text{ es una v.a. real}$$

Debe de quedar claro que un estimador, al ser función de una muestra aleatoria, es una variable aleatoria y por tanto debe de tener asociada una función de distribución de probabilidad, luego entonces surge de manera natural preguntarnos por algunas propiedades como su distribución o sus momentos.

Una de las primeras propiedades que deberíamos pedir al estimador es que al menos su contra dominio sea un subconjunto del espacio parametral, ya que la idea es que la distribución de dicho estimador esté centrada al rededor del parámetro que queremos estimar.

Para saber cuál estimador es mejor que otro, es necesario conocer las propiedades de cada uno y elegir aquél que más se adecue a nuestras necesidades. Con el objeto de conocer mejor estas propiedades es necesario saber cuando menos cuáles son sus momentos centrales y si es posible saber qué tipo de distribución está siguiendo.

Podemos motivar el estudio de la distribución de un estadístico de la siguiente manera. Sea  $X$  una variable aleatoria que nos mide alguna característica numérica de algún fenómeno aleatorio. En muchas ocasiones nos enfrentamos al problema de no saber completamente cuál es la función de distribución de esta variable aleatoria. Supongamos que sabemos que la variable aleatoria sigue una distribución  $F_X(x; \theta)$  pero con el parámetro  $\theta$  desconocido. Para obtener más información acerca de la distribución de  $X$  nosotros podemos repetir el experimento  $n$  veces de manera independiente y definir  $X_i$  la v.a. de la  $i$ -ésima repetición del experimento. Entonces  $X_1, \dots, X_n$  será una muestra aleatoria de la distribución en estudio con la cual se pretenderá estimar el valor de  $\theta$  a través de una función de ella. Ahora suponga que definimos un estadístico  $Y = u(X_1, \dots, X_n)$  para el cual podemos encontrar su función de densidad de probabilidad  $G_Y(u)$  y supongamos que esta función nos muestra que existe una probabilidad considerable de que  $Y$ , al realizar los experimentos, tenga un valor cercano al parámetro desconocido. Así, una vez que el experimento ha sido repetido en  $n$  ocasiones se obtiene una muestra observada  $x_1, \dots, x_n$  la cual al ser evaluada en la función de la muestra  $u(x_1, \dots, x_n)$  nos dará un número conocido al cual esperamos, bajo las condiciones arriba expuestas, sea cercano al valor desconocido del parámetro  $\theta$ .

Observe que de lo anterior para saber qué tan cercano será el valor estimado al parámetro desconocido  $\theta$ , es necesario conocer la forma de la distribución del estadístico que lo define, para así tener una medida de probabilidad de qué tan cercana será nuestra estimación.

**Ejemplo 4.2.1.** *Supongamos que tenemos  $X_1, \dots, X_n$  m.a. del modelo exponencial de parámetro  $\lambda$  desconocido, primero observemos que en este caso el espacio parametral está dado por  $\Theta = \mathbb{R}^+$ , queremos definir una función de la muestra que nos ayude a determinar cuál de todos los valores que están en el espacio parametral es el más apropiado para estimar a  $\lambda$ . Posibles estimadores para  $\lambda$  pueden ser:*

- $\hat{\lambda}(X_1, \dots, X_n) = \frac{n}{\sum_{i=1}^n X_i}$
- $\hat{\lambda}(X_1, \dots, X_n) = \min X_1, \dots, X_n$
- $\hat{\lambda}(X_1, \dots, X_n) = \frac{\sum_{i=1}^n X_i}{n}$

*Algo que no puede ser llamado estimador para  $\lambda$  sería :*

- $\hat{\lambda}(X_1, \dots, X_n) = \lambda \frac{n}{\sum_{i=1}^n X_i}$

*Este último caso no es estimador pues la expresión de la función involucra al parámetro desconocido  $\lambda$ .*

En ocasiones es difícil y ambicioso conocer completamente la distribución que sigue un estimador por lo que a veces solo nos conformamos conociendo algunos momentos como su esperanza y varianza.

Por ejemplo, supongamos que tenemos  $X_1, \dots, X_n$  m.a. de una cierta distribución  $F$  con media  $\mu$  y varianza  $\sigma^2$ , supongamos además que  $\mu$  es un parámetro de interés de la población que estamos estudiando, por lo que definimos el siguiente estimador:

$$\hat{\mu} = \hat{\mu}(X_1, \dots, X_n) = \frac{\sum_{i=1}^n X_i}{n} = \bar{X}$$

Ciertamente desconocemos la distribución de  $\hat{\mu}$  pues de hecho no se ha definido concretamente el modelo  $F$ , sin embargo es posible conocer los dos primeros momentos del estimador con la información que se tiene, pues:

$$\begin{aligned} \mathbb{E}(\hat{\mu}) &= \mathbb{E}(\bar{X}) = \frac{1}{n} \mathbb{E}\left(\sum_{i=1}^n X_i\right) = \mu \\ \text{Var}(\hat{\mu}) &= \text{Var}(\bar{X}) = \frac{1}{n^2} \text{Var}\left(\sum_{i=1}^n X_i\right) = \frac{\sigma^2}{n} \end{aligned}$$



Observe que en la última igualdad utilizamos fuertemente el hecho de la independencia de las variables aleatorias  $X_1, \dots, X_n$ . Debe de quedar claro entonces que no conocemos la distribución de  $\hat{\mu}$  sin embargo el hecho de conocer los dos primeros momentos nos brinda mucha información ya que ahora se sabe que en promedio  $\hat{\mu}$  tomará valores cercanos a  $\mu$  y que además la varianza de dicho estimador se va haciendo cada vez más pequeña conforme tenemos más muestra, esto quiere decir que  $\hat{\mu}$  es un estimador que conforme tiene más muestra ofrece mejores aproximaciones hacia el valor desconocido  $\mu$ . Estimadores que tienen esta propiedad son utilizados en estadística y se les denomina **Estimadores Insesgados**, los cuales se definen a continuación.

**Definición 4.2.4** (Estimador Insesgado). Decimos que  $\hat{\theta}$  es un estimador **insesgado** para  $\theta$  si

$$\mathbb{E}(\hat{\theta}) = \theta$$

Y cuando no pasa la igualdad decimos que el estimador es **sesgado**

Los estimadores insesgados entonces tiene la propiedad de que sus esperanza es el valor del parámetro desconocido, esto quiere decir que si repetimos muchas veces el experimento y se evalúa el estimador en cada una de las muestras que se van obteniendo, entonces en promedio el valor que se va a obtener es precisamente el valor desconocido  $\theta$ .

**Observación Importante:** Un estimador es una v.a. pues es función de la muestra aleatoria, sin embargo una vez observada la muestra y evaluada en la función que determina el estimador se obtiene un número, a dicho número se le conoce como estimación. Debe entonces quedar claro la gran diferencia que hay entre un estimador y la estimación.

**Ejercicio 4.2.1.** *Conteste lo siguiente:*

1. Suponga  $X_1, \dots, X_n$  m.a. de una distribución con media  $\mu$  y varianza  $\sigma^2$ . Pruebe que:

$$\hat{\sigma}_1^2 := \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}$$

Es un estimador sesgado para  $\sigma^2$  pero que:

$$\hat{\sigma}_2^2 := \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}$$

si es insesgado para  $\sigma^2$ , luego suponga que observamos la siguiente muestra:  $X_1 = 2, X_2 = 1, X_3 = 4, X_4 = 2, X_5 = 1$ , encuentre el valor estimado de  $\sigma^2$  de ambos estimadores.

2. Sea  $X_1, X_2, X_3, X_4, X_5$  muestra aleatoria de tamaño  $n = 5$  de una población con densidad dada por:

$$f_X(x; \theta) = \frac{\theta^x}{x!} e^{-\theta}; \quad \theta > 0; \quad x \in \{0, 1, 2, \dots, \}$$

Considere las siguientes funciones de la muestra:

$$\begin{aligned} \hat{\theta}_1 &= X_1 + X_2 + X_3 + X_4 + X_5 \\ \hat{\theta}_2 &= X_1 + \theta X_2 \\ \hat{\theta}_3 &= \frac{X_1 + X_2 + X_3 + X_4 + X_5}{5} \\ \hat{\theta}_4 &= \frac{2X_1 + 3X_2}{5} \end{aligned}$$

- Indique cuáles funciones de la muestra pueden ser considerados estimadores para  $\theta$  y cuáles no
- Indique cuáles estimadores son insesgados y cuáles sesgados.

Como ya se ha mencionado, encontrar la distribución exacta de estimadores a veces no es fácil de obtener, sin embargo en algunas ocasiones es posible encontrar algebraicamente la distribución exacta del estimador (ver apéndice). Por ejemplo, basado en ciertos teoremas se puede probar que cuando se tienen  $X_1, \dots, X_n$  m.a. del modelo  $N(\mu, \sigma^2)$  entonces:

$$\hat{\mu} = \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right); \quad \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} \sim \text{Gamma}\left(n-1, \frac{(n-1)\sigma^2}{n}\right)$$

El caso Normal es desde el punto de vista teórico muy importante pues permite encontrar distribuciones exactas de sus estimadores, sin embargo queda la pregunta abierta para el caso de distribuciones no Normales, afortunadamente contamos con distribuciones límites que nos ayudan a aproximar la distribución de la media poblacional a través de la Normal. Lo anterior está fuertemente sustentado en uno de los teoremas más importantes de la probabilidad que enunciamos a continuación.

**Teorema 4.2.1.** [Teorema del Límite Central] Sea  $X_1, \dots, X_n$  m.a. de una distribución  $F_x$  (Puede ser continua o discreta), tal que,  $\mathbb{E}(X) = \mu$  y  $\text{Var}(X) = \sigma^2$  ambos finitos. Entonces:

$$\bar{X} := \frac{\sum_{i=1}^n X_i}{n} \underset{\text{aprox}}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right)$$

O bien, otra forma de verlo es:

$$\sqrt{n} \left( \frac{\bar{X} - \mu}{\sigma} \right) \underset{\text{aprox}}{\sim} N(0, 1)$$

De hecho el teorema concluye que:

$$\sqrt{n} \left( \frac{\bar{X} - \mu}{\sigma} \right) \xrightarrow{n \rightarrow \infty} N(0, 1)$$

Algunas observaciones importantes del T.L.C

1. Es importante tener variables aleatorias independiente e idénticamente distribuidas y que el segundo momento exista (En series de tiempo no se aplica el T.L.C)
2. No importa de que distribución estemos muestreando, el teorema afirma que habrá una aproximación hacia la normalidad de  $\bar{X}$
3. Una de las preguntas más interesantes al momento de que vamos a aplicar el T.L.C es saber qué  $n$  debo tener para que afirmar que tengo ya una buena aproximación. La respuesta a esta pregunta no es fácil y se puede encontrar por medio de simulaciones sin embargo se muestra que el tamaño de muestra necesario depende de la forma de  $F_X$  sin embargo en la práctica esta última distribución es desconocida.

En la teoría de estimación el T.L.C. juega un papel interesante pues de hecho tiene la siguiente implicación.

**Teorema 4.2.2.** *Sea  $X_1, \dots, X_n$  m.a. de una población  $F_X$  tal que  $\mathbb{E}(X) = \mu$  y  $\text{Var}(X) = \sigma^2$ , suponga que se desea estimar a  $\mu$  por medio de la estadística  $\hat{\mu} = \bar{X}$  entonces, la distribución aproximada de este estimador está dada por:*

$$\hat{\mu} \overset{\text{aprox}}{\sim} N \left( \mu, \frac{\sigma^2}{n} \right)$$

El teorema anterior indica basicamente que si estamos interesados en estimar a la media de un modelo, entonces el estimador dado por  $\bar{X}$  es insesgado y además sigue una distribución aproximadamente Normal, en el caso de que se obtenga la muestra del modelo Normal, entonces la distribución del estimador es exacta.

### 4.2.1. Estimadores Insesgados de Mínima Varianza

En esta sección se explica cómo encontrar al mejor estimador para cierto parámetro desconocido, para responder esto primero es necesario un criterio que ayude a determinar cuál estimador es mejor que otro.

**Definición 4.2.5** (Error Cuadrático Medio). *Sea  $X_1, \dots, X_n$  una m.a. de un distribución  $F_X(\cdot; \theta)$ , con  $\theta$  un parámetro desconocido, supongamos que tenemos  $\hat{\theta}$  un estimador de  $\theta$ , definimos el Error*

cuadrático medio de  $\hat{\theta}$  como :

$$E.C.M.(\hat{\theta}) := \mathbb{E}\left(\left(\hat{\theta} - \theta\right)^2\right)$$

Observe que el E.C.M. mide la discrepancia cuadrática promedio que tiene nuestro estimador  $\hat{\theta}$  con respecto al verdadero valor del parámetro  $\theta$ . Es fácil ver además que :

$$E.C.M.(\hat{\theta}) = \text{Var}(\hat{\theta}) + \left(\mathbb{E}(\hat{\theta}) - \theta\right)^2$$

Donde  $\left(\mathbb{E}(\hat{\theta}) - \theta\right)$  se le conoce como el sesgo del estimador, observe además también que cuando el estimador es insesgado, entonces  $E.C.M.(\hat{\theta}) = \text{Var}(\hat{\theta})$ .

**Observación Importante:** El E.C.M. es una medida de qué tanto nos alejamos en promedio del verdadero valor del parámetro por lo que resulta lógico escoger siempre el estimador con menor  $E.C.M.$ , parece entonces que si queremos encontrar al mejor estimador para  $\theta$  necesitamos encontrar a  $\hat{\theta}$  tal que:

$$E.C.M.(\hat{\theta}) \leq E.C.M.(\hat{\theta}_1) \quad \forall \hat{\theta}_1 \text{ estimador de } \theta$$

Lo anterior no es tan fácil pues se puede probar que debido a que  $E.C.M.(\hat{\theta})$  depende de  $\theta$ , entonces se pueden encontrar dos estimadores  $\hat{\theta}_1$  y  $\hat{\theta}_2$  tal que:

$$E.C.M.(\hat{\theta}_1) < E.C.M.(\hat{\theta}_2) \quad \text{si } \theta \in \Theta_0$$

$$E.C.M.(\hat{\theta}_1) > E.C.M.(\hat{\theta}_2) \quad \text{si } \theta \in \Theta_1$$

Donde  $\Theta_0 \subset \Theta$  y  $\Theta_1 \subset \Theta$

**Ejemplo 4.2.2.** Sea  $X_1, \dots, X_n$  m.a. de una población  $F_X$  tal que  $\mathbb{E}(X) = \mu$  y  $\text{Var}(X) = \sigma^2$  finitos, supongamos que queremos estimar a  $\mu$  por medio de los siguientes estimadores:

$$\hat{\mu}_1 = \bar{X} \quad \hat{\mu}_2 = \mu_0$$

con  $\mu_0$  un numero conocido. Entonces:

$$E.C.M.(\hat{\mu}_1) = \frac{\sigma^2}{n} \quad E.C.M.(\hat{\mu}_2) = (\mu_0 - \mu)^2$$

Luego entonces si  $\mu = \mu_0$  se tiene que  $E.C.M.(\hat{\mu}_2) = 0$  y por tanto  $\hat{\mu}_2$  sería un mejor estimador que  $\hat{\mu}_1$  sin embargo si  $\mu$  está muy alejado de  $\mu_0$  tal que,  $(\mu_0 - \mu)^2 > \frac{\sigma^2}{n}$ , entonces  $\hat{\mu}_1$  es mejor estimador que  $\hat{\mu}_2$ .

Para solucionar este inconveniente del E.C.M. lo que se hace es restringir la búsqueda del mejor estimador entre todos los **insesgados**. Por lo tanto se desarrolla toda una teoría para encontrar los **estimadores insesgados de mínima varianza**. (Ver Apéndice)

**Ejercicio 4.2.2.** *Responda los siguientes ejercicios:*

1. Sea  $X_1, \dots, X_n$  muestra aleatoria de tamaño  $n$  de una población con distribución  $F_X(x)$  tal que  $\mathbb{E}(X_i) = \mu$  y  $\text{Var}(X_i) = \sigma^2$ . Pruebe que:

- Suponga  $\mu$  conocida, entonces  $S_1^2 := \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$  es estimador insesgado para  $\sigma^2$ .
- Suponga  $\mu$  desconocida, entonces  $S_2^2 := \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$  es un estimador sesgado para  $\sigma^2$ .
- Suponga  $\mu$  desconocida,  $S_3^2 := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$  es un estimador insesgado para  $\sigma^2$ .

2. Sea  $X_1, \dots, X_n$  muestra aleatoria de tamaño  $n$  de una población con distribución  $F_X(x)$  tal que  $\mathbb{E}(X_i) = \mu$  y  $\text{Var}(X_i) = \sigma^2$ . Sea  $\alpha_1 \dots \alpha_n \in \mathbb{R}$  tales que  $\sum_{i=1}^n \alpha_i = 1$ . Defina el siguiente estimador para  $\mu$ :

$$\hat{\mu} = \sum_{i=1}^n \alpha_i X_i$$

Pruebe que:

- a)  $\hat{\mu}$  es un estimador insesgado.
  - b) La varianza de  $\hat{\mu}$  se minimiza cuando  $\alpha_i = \frac{1}{n}$  para toda  $i \in \{1, \dots, n\}$ . (Este inciso muestra entonces que  $\bar{X}$  es el mejor estimador lineal). Cuando ocurre esto se dice que el estimador es BLUE (Best Linear Unbiased Estimator).
3. Sea  $X_1, X_2, X_3, X_4, X_5$  muestra aleatoria de tamaño  $n = 5$  de una población con densidad dada por:

$$f_X(x; \theta) = \frac{1}{\theta} e^{-\frac{x}{\theta}}; \quad \theta > 0; \quad x > 0$$

Considere las siguientes funciones de la muestra:

$$\begin{aligned} \hat{\theta}_1 &= X_1 + X_2 + X_3 + X_4 + X_5 \\ \hat{\theta}_2 &= X_1 + \theta X_2 \\ \hat{\theta}_3 &= \frac{X_1 + X_2 + X_3 + X_4 + X_5}{5} \\ \hat{\theta}_4 &= \frac{X_1 + 4X_2}{5} \end{aligned}$$

Conteste lo siguiente:

- Indique cuáles pueden ser considerados estimadores para  $\theta$  y cuáles no.

- Indique cuáles estimadores son insesgados y cuáles sesgados.
- Para los estimadores sesgados, calcule su sesgo.
- Para los estimadores insesgados, indique cuál tiene menor Error Cuadrático Medio y concluya cuál estimador es mejor.

4. Sea  $X_1, X_2, X_3$  muestra aleatoria de tamaño 3 de una población con distribución  $F_X(x)$  tal que  $\mathbb{E}(X_i) = \mu$  y  $\text{Var}(X_i) = 1$ . Calcular el Error Cuadrático Medio para los siguientes estimadores de  $\mu$ :

$$\hat{\mu}_1 = X_1 \quad ; \quad \hat{\mu}_2 = \frac{(3X_1 - 2X_2 + X_3)}{6}$$

Para cuáles valores de  $\mu$ , se tiene que  $\hat{\mu}_2$  es mejor estimador que  $\hat{\mu}_1$ ?

5. Sea  $X_1, \dots, X_n$  muestra aleatoria de un modelo  $N(\mu, \sigma^2)$  con  $\mu$  y  $\sigma^2$  desconocidos. Sabemos que  $S_3^2 := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$  es un estimador insesgado para  $\sigma^2$ . Demuestre que:

$$\text{Var}(S_3^2) = \frac{2\sigma^4}{n-1}$$

Esto muestra entonces que cuando  $\lim_{n \rightarrow \infty} \text{Var}(S_3^2) = 0$  lo cual es una propiedad interesante del estimador pues implica que conforme la muestra tiende a infinito, la precisión es cada vez mejor.

Basado en la teoría desarrollada se busca aquel estimador  $\hat{\theta}$  que además de ser insesgado cumpla con ser de mínima varianza, dichos estimadores se les denomina UMVUE's.

**Definición 4.2.6** (UMVUE). Al estimador insesgado de mínima varianza se le llama UMVUE por sus siglas en inglés *Uniformly Minimum-Variance Unbiased Estimator*

Resulta entonces que el objetivo se centrará en encontrar estimadores que sean **UMVUES** y teóricamente existen 2 caminos para encontrarlos:

1. **Cota Inferior de Cramer Rao.** Básicamente consiste en construir una cota inferior para la varianza de todos los estimadores insesgados de un modelo específico, luego, se busca un estimador  $\hat{\theta}$  tal que cumpla que su varianza es igual a dicha cota.
2. **Construcción de Estimadores a través de estadísticas suficientes.** Básicamente consiste en utilizar teoremas sobre estadísticas suficientes para construir vía esperanza condicional estimadores que cumplan con ser de mínima varianza.

A continuación se trabajan brevemente estos dos caminos.

### 4.2.2. Cota Inferior de Cramer-Rao

Se puede probar (*Ver Apéndice*) que bajo ciertas condiciones de regularidad es posible encontrar una cota inferior para la varianza de los estimadores insesgados de cierto parámetro y modelo.

**Teorema 4.2.3** (Cota Inferior de Cramer-Rao). *Sea  $X_1, \dots, X_n$  m.a. de un modelo con densidad  $f_X(x; \theta)$  y  $\theta$  un parámetro desconocido tal que  $\theta \in \Theta$ , definimos cota inferior de Cramer-Rao para los estimadores insesgados de  $\theta$  como:*

$$CICR(\theta) = \frac{1}{nI(\theta)} \leq \text{Var}(\hat{\theta})$$

Donde  $\hat{\theta}$  es un estimador insesgado arbitrario

En este caso  $I(\theta)$  se le conoce como Información de Fisher por Unidad Muestral y se define a continuación.

**Definición 4.2.7** (Información de Fisher por Unidad Muestral). *Dentro del contexto anterior definimos a **Información de Fisher por Unidad Muestral** denotada por  $I(\theta)$  como:*

$$I(\theta) = \text{Var}(U_i) = -\mathbb{E}\left(\frac{\partial^2}{\partial \theta^2} \text{Log}(f_{X_i}(X_i; \theta))\right)$$

Entonces si se encuentra un estimador insesgado  $\hat{\theta}^*$  tal que  $\text{Var}(\hat{\theta}^*) = CICR(\theta)$ , entonces  $\hat{\theta}^*$  es UMVUE, es decir, es el mejor estimador insesgado que podemos encontrar. (*Ver Apéndice*)

**Ejemplo 4.2.3.** *Sea  $X_1, \dots, X_n$  m.a. de un modelo Bernoulli de parámetro  $\theta \in (0, 1)$ . Entonces sabemos que  $f_{X_i}(x_i; \theta) = \theta^{x_i} (1 - \theta)^{1-x_i} \mathbf{1}_{\{0,1\}}(x_i)$ , por lo que:*

$$\text{Log}(f_{X_i}(X_i; \theta)) = x \text{Log}(\theta) + (1 - x) \text{Log}(1 - \theta)$$

Así:

$$\frac{\partial^2}{\partial \theta^2} \text{Log}(f_{X_i}(X_i; \theta)) = -\frac{x}{\theta^2} - \frac{1 - x}{(1 - \theta)^2}$$

Por lo que la información de Fisher queda como:

$$I(\theta) = -\mathbb{E}\left(-\frac{x}{\theta^2} - \frac{1 - x}{(1 - \theta)^2}\right) = \frac{1}{\theta(1 - \theta)}$$

Por lo tanto, para este modelo:

$$CICR(\theta) = \frac{1}{nI(\theta)} = \frac{\theta(1 - \theta)}{n}$$

Lo que concluye que cualquier estimador insesgado de  $\theta$ , digamos  $\hat{\theta}$  es tal que:

$$\text{Var}(\hat{\theta}) \geq \frac{\theta(1-\theta)}{n}$$

No es difícil probar con esto entonces que:

$$\hat{\theta}^* := \bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$

Es un UMVUE para  $\theta$  en el modelo Bernoulli( $\theta$ )

**Ejemplo 4.2.4.** Considere  $X_1, \dots, X_n$  m.a. de un modelo Exponencial de parámetro  $\theta \in (0, \infty)$ . Entonces sabemos que  $f_{X_i}(x_i; \theta) = \frac{1}{\theta} \exp^{-\frac{x_i}{\theta}} \mathbf{1}_{(0, \infty)}(x_i)$ .

- Suponiendo que se cumplen las condiciones de regularidad encuentre la CICR( $\theta$ ).

$$\text{Log}(f_{X_i}(X_i; \theta)) = -\text{Log}(\theta) - \frac{x}{\theta}$$

Ahora

$$\frac{\partial^2}{\partial \theta^2} \text{Log}(f_{X_i}(X_i; \theta)) = \frac{1}{\theta^2} - \frac{2x}{\theta^3}$$

Por lo que la información de Fisher queda como:

$$I(\theta) = -\mathbb{E}\left(\frac{1}{\theta^2} - \frac{2x}{\theta^3}\right) = \frac{1}{\theta^2}$$

Por lo tanto, para este modelo:

$$\text{CICR}(\theta) = \frac{1}{nI(\theta)} = \frac{\theta^2}{n}$$

- Considere  $\hat{\theta} = \bar{X}$  calcule  $\mathbb{E}(\hat{\theta})$  ¿ $\hat{\theta}$  es insesgado para  $\theta$ ?

$$\mathbb{E}(\hat{\theta}) = \frac{\sum_{i=1}^n \mathbb{E}(x_i)}{n} = \frac{\sum_{i=1}^n \theta}{n} = \frac{n\theta}{n} = \theta$$

Por lo tanto  $\hat{\theta}$  es insesgado para  $\theta$

- Calcule la  $\text{Var}(\hat{\theta})$  y el E.C.M.( $\hat{\theta}$ ) ¿ $\hat{\theta}$  es UMVUE para  $\theta$ ?

$$\text{Var}(\hat{\theta}) = \frac{\sum_{i=1}^n \text{Var}(x_i)}{n^2} = \frac{\sum_{i=1}^n \theta^2}{n^2} = \frac{n\theta^2}{n^2} = \frac{\theta^2}{n}$$

Como  $\hat{\theta}$  es insesgado



$$E.C.M.(\hat{\theta}) = \text{Var}(\hat{\theta}) = \frac{\theta^2}{n} = CICR(\theta)$$

Por lo tanto  $\hat{\theta}$  es UMVUE para  $\theta$

**Ejemplo 4.2.5.** Considere  $X_1, \dots, X_n$  m.a. de de una población  $N(\mu, \sigma^2)$   $\sigma^2$  conocida.

$$f_X(x; \mu) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}; \quad \sigma^2 > 0 \quad \mu \in \mathbb{R}$$

- Suponiendo que se cumplen las condiciones de regularidad encuentre la  $CICR(\mu)$ .

$$\text{Log}(f_{X_i}(X_i; \mu)) = -\frac{(x - \mu)^2}{2\sigma^2}$$

Ahora

$$\frac{\partial^2}{\partial \mu^2} \text{Log}(f_{X_i}(X_i; \mu)) = -\frac{1}{\sigma^2}$$

Por lo que la información de Fisher queda como:

$$I(\mu) = -\mathbb{E}\left(-\frac{1}{\sigma^2}\right) = \frac{1}{\sigma^2}$$

Por lo tanto, para este modelo:

$$CICR(\mu) = \frac{1}{nI(\mu)} = \frac{\sigma^2}{n}$$

- Considere  $\hat{\mu} = \bar{X}$  calcule  $\mathbb{E}(\hat{\mu})$  ¿  $\hat{\mu}$  es insesgado para  $\mu$ ?

$$\mathbb{E}(\hat{\mu}) = \frac{\sum_{i=1}^n \mathbb{E}(x_i)}{n} = \frac{\sum_{i=1}^n \mu}{n} = \frac{n\mu}{n} = \mu$$

Por lo tanto  $\hat{\mu}$  es insesgado para  $\mu$

- Calcule la  $\text{Var}(\hat{\mu})$  y el  $E.C.M.(\hat{\mu})$  ¿  $\hat{\mu}$  es UMVUE para  $\mu$ ?

$$\text{Var}(\hat{\mu}) = \frac{\sum_{i=1}^n \text{Var}(x_i)}{n^2} = \frac{\sum_{i=1}^n \sigma^2}{n^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

Como  $\hat{\mu}$  es insesgado

$$E.C.M.(\hat{\mu}) = \text{Var}(\hat{\mu}) = \frac{\sigma^2}{n} = CICR(\mu)$$

Por lo tanto  $\hat{\mu}$  es UMVUE para  $\mu$

**Ejercicio 4.2.3.** Resuelva lo siguiente:

1. Sea  $X_1, \dots, X_n$  m.a. de una población con densidad dada por:

$$f_X(x; \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}x^2}, \sigma^2 > 0$$

a) Calcule la Cota Inferior de Cramer Rao para esta familia y verifique que:

$$CICR(\sigma^2) = \frac{2\sigma^4}{n}$$

b) Defina un estimador de  $\sigma^2$  como  $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n X_i^2$  y demuestre que este estimador es insesgado.

c) ¿El estimador definido en el inciso anterior es UMVUE?

2. Sea  $X_1, \dots, X_n$  m.a. de una población con densidad Binomial ( $k, \theta$ )  $k$  conocida.

$$f_X(x) = \mathbb{P}(X = x) = \binom{k}{x} \theta^x (1 - \theta)^{k-x} \quad x \in \{0, 1, 2, \dots, k\} \quad \theta \in (0, 1)$$

a) Calcule la Cota Inferior de Cramer Rao para esta familia y verifique que:

$$CICR(\theta) = \frac{\theta(1 - \theta)}{nk}$$

b) Defina el estimador  $\hat{\theta} = \frac{1}{k} \bar{X}$ , pruebe que el estimador es insesgado.

c) Calcule la  $\text{Var}(\hat{\theta})$ . El estimador  $\hat{\theta}$  es UMVUE?

3. Sea  $X_1, \dots, X_n$  m.a. de una población con densidad Gamma ( $\alpha, \theta$ )  $\alpha$  conocida.

$$f_X(x; \theta) := \frac{1}{\theta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\theta}}; \quad x \in (0, \infty); \quad \alpha > 0; \quad \theta > 0$$

a) Calcule la Cota Inferior de Cramer Rao para esta familia y verifique que:

$$CICR(\theta) = \frac{\theta^2}{n\alpha}$$

b) Defina el estimador  $\hat{\theta} = \frac{1}{\alpha} \bar{X}$ , pruebe que el estimador es insesgado.

c) Calcule la  $\text{Var}(\hat{\theta})$ . El estimador  $\hat{\theta}$  es UMVUE?

4. Sea  $X_1, \dots, X_n$  m.a. de una población con densidad poisson de parámetro  $\lambda$ .

$$\mathbb{P}(X = x) = \frac{\lambda^x}{x!} e^{-\lambda} \quad x \in \{0, 1, 2, \dots\}$$

a) Calcule la Cota Inferior de Cramer Rao para esta familia y verifique que:

$$CICR(\lambda) = \frac{\lambda}{n}$$

b) Defina el estimador  $\hat{\lambda} = \frac{1}{n} \sum_{i=1}^n X_i$ , pruebe que el estimador es insesgado.

c) Calcule la  $Var(\hat{\lambda})$ . El estimador  $\hat{\lambda}$  es UMVUE?

5. Sea  $X_1, \dots, X_n$  m.a. de una población con densidad Normal de parámetros  $\mu, \sigma^2$  (ambos desconocidos).

a) Calcule la Cota Inferior de Cramer Rao para los estimadores de  $\sigma^2$ . Verifique que:

$$CICR(\sigma^2) = \frac{2\sigma^4}{n}$$

b) Defina un estimador de  $\sigma^2$  como  $\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$  y demuestre que este estimador es insesgado.

c) El estimador definido en el inciso anterior es UMVUE?

### 4.2.3. Construcción de Estimadores de Mínima Varianza

Una de las desventajas que tiene el método de Cota Inferior de Cramer Rao es que solamente nos indica si un estimador es UMVUE o no a través del cálculo de la cota y la varianza del estimador, en ningún momento indica cómo construir un UMVUE. Otro inconveniente es que se puede construir ejemplos donde estimadores UMVUE cuya varianza no alcanzan la Cota Inferior de Cramer Rao.

Existe otro camino para lograr encontrar estimadores insesgados de mínima varianza, sin embargo su detalle técnico implica conceptos fuera del alcance de estas notas como lo son Suficiencia, Suficiencia Minimal y Completes de una estadística.

Solo se mencionan los resultados más importantes de esta teoría.

Dentro de la teoría que se requiere para la construcción de UMVUE's se utilizan las denominadas **estadísticas completas**, las cuales juegan un papel muy importante pues son la estadísticas más importantes de los modelos paramétricos. Existe un teorema que nos garantiza que un estimador insesgado que sea función de las estadísticas completas será el de mínima varianza. Desafortunadamente encontrar estadísticas completas no es nada sencillo y se requieren del desarrollo de teoría bastante basta (*Ver apéndice*).

Afortunadamente existe una familia de distribuciones en la que es posible encontrar estadísticas completas de forma sencilla.

**Definición 4.2.8** (Familia exponencial Uniparamétrica). Sea  $f$  una función de densidad, decimos que  $f_X(x; \theta)$  pertenece a la familia exponencial uniparamétrica si:

$$f_X(x; \theta) = a(\theta) b(x) e^{c(\theta)d(x)} \quad x \in \mathbb{R}$$

Donde,  $a(\theta), c(\theta)$  son funciones reales de  $\theta$  y no depende de  $x$ , mientras que  $b(x), d(x)$  son funciones reales de  $x$  que no dependen de  $\theta$ .

**Ejemplo 4.2.6.** Verifiquemos que el modelo Bernoulli pertenece a la familia exponencial.

$$\begin{aligned} f_X(x; \theta) &= \theta^x (1 - \theta)^{1-x} \mathbb{1}_{\{0,1\}}(x) \\ &= e^{\ln(\theta^x (1-\theta)^{1-x}) \mathbb{1}_{\{0,1\}}} \\ &= e^{x \ln(\theta) + (1-x) \ln(1-\theta)} \mathbb{1}_{\{0,1\}} \\ &= e^{x \ln(\theta) + \ln(1-\theta) - x \ln(1-\theta)} \mathbb{1}_{\{0,1\}} \\ &= e^{x \ln\left(\frac{\theta}{1-\theta}\right)} (1 - \theta) \mathbb{1}_{\{0,1\}} \end{aligned}$$

Definiendo:

$$a(\theta) = (1 - \theta); \quad b(x) = \mathbb{1}_{\{0,1\}}(x); \quad c(\theta) = \ln\left(\frac{\theta}{1-\theta}\right); \quad d(x) = x$$

Obtenemos que:

$$f_X(x; \theta) = a(\theta) b(x) e^{c(\theta)d(x)}$$

Por lo tanto la densidad Bernoulli pertenece a la familia exponencial.

**Ejercicio 4.2.4.** Demuestre que:

1. El modelo Poisson( $\lambda$ ) pertenece a la familia exponencial. (Identifique adecuadamente quién es  $a(\lambda), c(\lambda), b(x), d(x)$ )

$$f_X(x) = \mathbb{P}(X = x) = \frac{\lambda^x}{x!} e^{-\lambda} \quad x \in \{0, 1, 2, \dots, \}$$

2. El modelo Binomial( $k, p$ ) pertenece a la familia exponencial, suponga para ello que el parámetro  $k$  es conocido. (Identifique adecuadamente quién es  $a(p), c(p), b(x), d(x)$ )

$$f_X(x) = \mathbb{P}(X = x) = \binom{k}{x} p^x (1-p)^{k-x} \quad x \in \{0, 1, 2, \dots, k\} \quad p \in (0, 1)$$

3. El modelo Exponencial ( $\lambda$ ) pertenece a la familia exponencial. (Identifique adecuadamente quién es  $a(\lambda), c(\lambda), b(x), d(x)$ )

$$f_X(x) = \lambda e^{-\lambda x} \quad x \in (0, \infty)$$

4. El modelo Geométrico( $p$ ) pertenece a la familia exponencial. (Identifique adecuadamente quién es  $a(p), c(p), b(x), d(x)$ )

$$f_X(x) = \mathbb{P}(X = x) = (1 - p)^{x-1} p \quad x \in \{1, 2, 3, \dots\}$$

5. El modelo Normal ( $\mu, \sigma^2$ ) pertenece a la familia exponencial. Suponga  $\sigma^2$  conocida. (Identifique adecuadamente quién es  $a(\mu), c(\mu), b(x), d(x)$ )

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \quad x \in \mathbb{R}$$

6. El modelo Gamma ( $\alpha, \beta$ ) pertenece a la familia exponencial. Suponga  $\alpha$  conocida. (Identifique adecuadamente quién es  $a(\beta), c(\beta), b(x), d(x)$ )

$$f_X(x) := \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}; \quad x \in (0, \infty)$$

La familia exponencial tiene la siguiente propiedad.

**Teorema 4.2.4** (Estadística Completa para la Familia Exponencial). *En la familia exponencial uniparamétrica, la estadística:*

$$T(X_1, \dots, X_n) = \sum_{i=1}^n d(X_i)$$

*es una estadística completa*

Las estadísticas completas son de gran utilidad por el siguiente teorema.

**Teorema 4.2.5** (Teorema de Lehmann Scheffé). *Sea  $X_1, \dots, X_n$  m.a. de  $f_X(x; \theta)$  y sea  $T_1$  una estadística completa, supongamos que existe una función  $g$  tal que el estimador  $\hat{\theta}^* = g(T_1)$  es insesgado, entonces  $\hat{\theta}^*$  es el UMVUE.*

A continuación se muestran ejemplos de cómo aplicar estos teoremas en los modelos más comunes:

**Ejemplo 4.2.7.** *Sabemos que cuando tenemos  $X_1, \dots, X_n$  m.a. Bernoulli( $\theta$ ) entonces  $T_1 = \sum_{i=1}^n X_i$  es una estadística completa, además sabemos que  $\frac{\sum_{i=1}^n X_i}{n}$  es insesgado por lo tanto se concluye mediante el Teorema de Lehmann Scheffé que  $\hat{\theta}^* = \frac{\sum_{i=1}^n X_i}{n}$  es UMVUE.*

**Ejemplo 4.2.8.** *Sea  $X_1, \dots, X_n$  m.a. del modelo Poisson( $\lambda$ ), se requiere el estimador UMVUE para  $\lambda$ . Primero se verifica que el modelo poisson pertenece a la familia exponencial.*

$$\begin{aligned} f_X(x; \lambda) &= \frac{\lambda^x}{x!} e^{-\lambda} \\ &= \frac{1}{x!} e^{-\lambda + x \ln(\lambda)} = \frac{1}{x!} e^{-\lambda} e^{\ln(\lambda)x} \end{aligned}$$

Por lo tanto definimos  $a(\lambda) = e^{-\lambda}$ ;  $b(x) = \frac{1}{x!}$ ;  $c(\lambda) = \ln(\lambda)$ ;  $d(x) = x$ , luego entonces, sabemos que  $\sum_{i=1}^n d(X_i) = \sum_{i=1}^n X_i$  es una estadística suficiente y completa para  $\lambda$ , por lo tanto si encontramos una función de  $\sum_{i=1}^n X_i$  que sea insesgada para  $\lambda$  habremos encontrado el UMVUE. Como  $\mathbb{E}(X_i) = \lambda$  se concluye entonces que  $\hat{\lambda} = \bar{X} = \frac{\sum_{i=1}^n X_i}{n}$  es un estimador insesgado función de una estadística completa y suficiente por lo tanto  $\hat{\lambda} = \bar{X}$  es UMVUE.

**Ejercicio 4.2.5.** Responda lo siguiente:

1. Encuentra una estadística completa para  $p$  del modelo Binomial( $k, p$ ) Suponga  $k$  conocida, construya a partir de dicha estadística el UMVUE para  $p$ .
2. Encuentra una estadística completa para  $\lambda$  del modelo Exponencial( $\lambda$ ) dado por:

$$f_X(x; \lambda) := \frac{1}{\lambda} e^{-\frac{1}{\lambda}x}; \quad x > 0$$

Construya a partir de la estadística completa el UMVUE para  $\lambda$ .

3. Encuentra una estadística completa para  $\mu$  del modelo Normal( $\mu, \sigma^2$ ). Suponga  $\sigma^2$  conocida, construya a partir de dicha estadística el UMVUE para  $\mu$ .

Dado que en la familia exponencial uniparamétrica siempre existe una estadística suficiente y completa, eso implica vía el proceso de Lehmanm Scheffé que siempre existirá un UMVUE para alguna transformación del parámetro desconocido. En efecto, como en la familia exponencial  $\sum_{i=1}^n d(X_i)$  es completa y suficiente entonces cualquier función de dicha estadística será UMVUE para cierta función de  $\theta$ . Por ejemplo, consideremos  $h(\theta) = \mathbb{E}(d(X_i))$  entonces  $\frac{\sum_{i=1}^n d(X_i)}{n}$  es UMVUE para  $h(\theta)$ , pues

$$\mathbb{E}\left(\frac{\sum_{i=1}^n d(X_i)}{n}\right) = \frac{n\mathbb{E}(d(X_1))}{n} = \mathbb{E}(d(X_1)) = h(\theta)$$

**Ejemplo 4.2.9.** Consideremos el modelo Exponencial:

$$f(x; \lambda) = \lambda e^{-\lambda x} \mathbb{1}_{(0, \infty)}^{(x)} \quad \lambda > 0$$

Es claro que pertenece a la familia exponencial pues  $a(\lambda) = \lambda$ ;  $b(x) = \mathbb{1}_{(0, \infty)}^{(x)}$ ;  $c(\lambda) = \lambda$ ;  $d(x) = x$ , luego entonces como,  $\mathbb{E}(X) = \frac{1}{\lambda}$  se sigue que la estadística:

$$\frac{\sum_{i=1}^n X_i}{n}$$

Es UMVUE para la cantidad  $\frac{1}{\lambda}$ .

**Ejercicio 4.2.6.** Considerando lo anterior responda lo siguiente:

1. Encuentra una estadística completa para  $p$  del modelo Geometrico( $p$ ), construya a partir de dicha estadística el UMVUE para  $\frac{1}{p}$ .

2. Encuentra una estadística suficiente y completa para  $\beta$  del modelo  $\text{Gamma}(\alpha, \beta)$ . Suponga  $\alpha$  conocida y construya a partir de dicha estadística el UMVUE para  $\frac{1}{\beta}$ .

#### 4.2.4. Construcción de Estimadores

A continuación se describen dos métodos para generar estimadores.

##### Método de Momentos

Este método se basa en el teorema de *Glivenko – Cantelli* el cual es conocido como el Teorema Fundamental de la Estadística.

**Teorema 4.2.6** (Teorema Glivenko-Cantelli (1933)). *Sea  $X_1, X_2, \dots$  una sucesión de variables aleatorias idénticamente distribuidas con función de distribución común dada por  $F(x)$ . Sea  $F_n(x)$  la función de distribución empírica formada por las primeras  $n$  variables aleatorias definida por:*

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{(-\infty, x]}(X_i)$$

*La ley fuerte de los grandes números afirma que la v.a.  $F_n(x)$  converge puntualmente y casi seguramente a  $F(x)$ , sin embargo, Glivenko-Cantelli demuestra que no solo hay convergencia puntual sino que hay una convergencia uniforme, es decir:*

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \rightarrow 0$$

El teorema indica entonces que conforme va creciendo la muestra se tiene que  $F_n$  se parece mucho a  $F$  para toda  $x \in \mathbb{R}$  entonces, como consecuencia de esta aproximación se puede mostrar que:

$$\sum_{i=1}^n \frac{x_i^r}{n} \approx \mathbb{E}(X^r)$$

A la cantidad  $\mathbb{E}(X^r)$  se le conoce como el  $r$ -ésimo momento poblacional mientras que  $\sum_{i=1}^n \frac{x_i^r}{n}$  se le conoce como el  $r$ -ésimo momento muestral, el teorema nos afirma entonces que los momentos poblacional y muestral son muy parecidos cuando la muestra es grande.

Ahora supongamos que  $X \sim F_X(\cdot; \underline{\theta})$  con  $\underline{\theta} = (\theta_1, \dots, \theta_p)$  un vector de parámetros. El método de momentos nos dice que las siguientes igualdades son aproximadas:

$$\begin{aligned}\mathbb{E}(X) &= a_1(\theta_1, \dots, \theta_p) \approx \sum_{i=1}^n \frac{x_i}{n} \\ \mathbb{E}(X^2) &= a_2(\theta_1, \dots, \theta_p) \approx \sum_{i=1}^n \frac{x_i^2}{n} \\ &\vdots \\ \mathbb{E}(X^p) &= a_p(\theta_1, \dots, \theta_p) \approx \sum_{i=1}^n \frac{x_i^p}{n}\end{aligned}$$

Lo cual genera un sistema de  $p$  ecuaciones con  $p$  incógnitas dadas por el vector de parámetros  $\underline{\theta} = (\theta_1, \dots, \theta_p)$ , el cual al resolverlo obtenemos:

$$\begin{aligned}\hat{\theta}_1 &= f_1(x_1, \dots, x_p) \\ &\vdots \\ \hat{\theta}_p &= f_p(x_1, \dots, x_p)\end{aligned}$$

Los cuales son denominados los estimadores de momentos. Considere los siguientes ejemplos:

**Ejemplo 4.2.10.** Sea  $X_1, \dots, X_n$  una muestra aleatoria del modelo Bernoulli ( $\theta$ ), entonces se sabe que el primer momento poblacional está dado por:  $\mathbb{E}(X) = \theta$  luego entonces la ecuación que queda para encontrar el estimador es:

$$\mathbb{E}(X) = \theta \approx \sum_{i=1}^n \frac{X_i}{n}$$

Luego entonces el estimador por momentos está dado por  $\hat{\theta} = \sum_{i=1}^n \frac{X_i}{n}$ .

**Ejemplo 4.2.11.** Sea  $X_1, \dots, X_n$  m.a. del modelo  $U(0, \theta)$ , entonces se sabe que  $\mathbb{E}(X) = \frac{\theta}{2}$  luego entonces igualando el primer momento poblacional y muestral se tiene que:

$$\mathbb{E}(X) = \frac{\theta}{2} \approx \sum_{i=1}^n \frac{X_i}{n}$$

Luego resolviendo la ecuación se obtiene que el estimador es:

$$\hat{\theta} = 2 \sum_{i=1}^n \frac{X_i}{n} = 2\bar{X}$$

Es fácil ver que este estimador es insesgado.

**Ejemplo 4.2.12.** Consideremos  $X_1, \dots, X_n$  m.a. del modelo  $N(\mu, \sigma^2)$ , en este caso se tienen 2



parámetros desconocidos por lo que se requiere generar y resolver un sistema de 2 ecuaciones con los dos primeros momentos:

$$\mathbb{E}(X) = \mu \approx \sum_{i=1}^n \frac{X_i}{n}$$

$$\mathbb{E}(X^2) = \mu + \sigma^2 \approx \sum_{i=1}^n \frac{X_i^2}{n}$$

Resolviendo el sistema:

$$\hat{\mu} = \bar{X} \quad \hat{\sigma}^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}$$

## Método de Máxima Verosimilitud

La verosimilitud es una función de  $\theta$  definida como sigue:

**Definición 4.2.9** (Verosimilitud-Muestra Observada). Sea  $X_1, \dots, X_n$  m.a. de un modelo con densidad  $f_X(x; \theta)$  y  $\theta$  un parámetro desconocido tal que  $\theta \in \Theta$ , definimos la función de verosimilitud  $\ell : \Theta \rightarrow \mathbb{R}$  como:

$$\ell(\theta; \underline{x}) := f_{\underline{X}}(x_1, \dots, x_n) = \prod_{i=1}^n f_{X_i}(x_i; \theta)$$

Método de máxima verosimilitud busca encontrar  $\hat{\theta}(x_1, \dots, x_p)$  tal que maximice la función de verosimilitud, la idea que hay detrás de este método es buscar aquel parámetro que maximice la plausibilidad o verosimilitud de haber observado la muestra  $(X_1 = x_1, \dots, X_n = x_n)$ , por ejemplo supongamos que en el modelo Bernoulli de parámetro  $\theta$  observamos la siguiente muestra con  $n = 5$

$$(X_1 = 0, X_2 = 0, X_3 = 0, X_4 = 0, X_5 = 0)$$

entonces sería inverosímil suponer que  $\theta \approx 1$ .

Maximizar a  $\ell(\theta; \underline{x})$  como función de  $\theta$  por lo general no es fácil, pues involucra el producto de densidades por lo que, en general, se decide no maximizar directamente a la verosimilitud sino mejor trabajar con el logaritmo de ésta debido a que el producto es transformado en suma:

$$L(\theta; \underline{x}) = \log \ell(\theta; \underline{x}) := \log f_{\underline{X}}(x_1, \dots, x_n) = \sum_{i=1}^n \log f_{X_i}(x_i; \theta)$$

Es sabido que una forma de encontrar el máximo de una función es derivando (**No es la única forma de hacerlo**) por lo que muchas veces para encontrar el estimador máximo verosímil se resuelve la siguiente ecuación:

$$\frac{\partial}{\partial \theta} L(\theta; \underline{x}) = 0$$

Sin embargo en el caso general, cuando tenemos  $k$  parámetros  $\underline{\theta} = (\theta_1, \dots, \theta_k)$  el problema de encontrar los estimadores máximo verosímiles se convierte en encontrar la solución al siguiente sistema de ecuaciones:

$$\begin{aligned} \frac{\partial}{\partial \theta_1} L(\underline{\theta}; \underline{x}) &= 0 \\ \frac{\partial}{\partial \theta_2} L(\underline{\theta}; \underline{x}) &= 0 \\ &\vdots \\ \frac{\partial}{\partial \theta_k} L(\underline{\theta}; \underline{x}) &= 0 \end{aligned}$$

Estos estimadores son muy utilizados en los modelos estadísticos más que nada por las siguientes propiedades:

- Son consistentes.
- Son asintóticamente eficientes.
- En general los estimadores no siempre son insesgados pero son asintóticamente insesgados.
- Los estimadores máximo verosímiles gozan la **propiedad de invarianza**, supongamos que tenemos  $\hat{\theta}_{MV}$  pero imaginemos que estamos interesados en un estimador para alguna transformación (no necesariamente biyectiva)  $g(\theta)$  entonces  $g(\hat{\theta})$  es el estimador máximo verosímil de  $g(\theta)$ . (Por ejemplo el estimador máximo verosímil para  $\mathbb{P}(X = 1) = e^{-\lambda}$  en el modelo poisson esta dado por  $\mathbb{P}(\widehat{X} = 1) = e^{-\hat{\lambda}}$ )

- Bajo ciertas condiciones de regularidad se puede probar que:

$$\hat{\theta}_{MV} \sim N\left(\theta, \frac{1}{nI(\theta)}\right)$$

donde  $I(\theta)$  es la información de Fisher por unidad muestral, es decir:

$$I(\theta) = -\mathbb{E}\left(\frac{\partial^2}{\partial \theta^2} \log(f_X(x; \theta))\right)$$

- Una generalización de lo anterior nos dice que bajos las mismas condiciones de regularidad se tiene que:

$$f(\hat{\theta}_{MV}) \sim N\left(f(\theta), \frac{(f'(\theta))^2}{nI(\theta)}\right)$$

A continuación se muestran tres ejemplos de cómo encontrar estimadores máximo verosímiles en tres familias paramétricas.

**Ejemplo 4.2.13.** Sea  $X_1, \dots, X_n$  m.a. aleatoria del modelo  $N(\mu; \sigma^2)$ , entonces la función de verosimilitud esta dada por:

$$\ell((\mu, \sigma^2); \underline{x}) := f_{\underline{X}}(x_1, \dots, x_n; (\mu; \sigma^2)) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2}$$

Luego entonces la Log-Verosimilud es:

$$L((\mu, \sigma^2); \underline{x}) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu)^2$$

Al derivar parcialmente e igualar a cero se obtiene el siguiente sistema de ecuaciones:

$$\begin{aligned} \sum_{i=1}^n (x_i - \mu) &= 0 \\ -\frac{n}{2} + \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 &= 0 \end{aligned}$$

Resolviendo dicho sistema de ecuaciones se obtienen los siguientes estimadores:

$$\begin{aligned} \hat{\mu}_{MV} &= \frac{1}{n} \sum_{i=1}^n X_i = \bar{X} \\ \hat{\sigma}_{MV}^2 &= \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \end{aligned}$$

Ahora consideraremos otro ejemplo que no requiere el uso de herramientas de cálculo para la maximización de la verosimilitud.

**Ejemplo 4.2.14.** Sea  $X_1, \dots, X_n$  m.a. aleatoria del modelo  $U(0, \theta)$ , entonces la función de verosimilitud esta dada por:

$$\ell(\theta; \underline{x}) := f_{\underline{X}}(x_1, \dots, x_n; \theta) = \prod_{i=1}^n \frac{1}{\theta} \mathbb{1}_{(0, \theta)}(x_i) = \frac{1}{\theta^n} \prod_{i=1}^n \mathbb{1}_{(0, \theta)}(x_i)$$

Sin embargo se sabe que:

$$\mathbb{1}_{(0, \theta)}(x_i) = \mathbb{1}_{(X_{(n)}, \infty)}(\theta)$$

con  $X_{(n)}$   $n$ -ésima estadística de orden. Lo que se concluye es que entonces la verosimilitud está dada por:

$$\ell(\theta; \underline{x}) = \frac{1}{\theta^n} \mathbb{1}_{(X_{(n)}, \infty)}(\theta)$$

Donde queda entonces  $\frac{1}{\theta^n}$  es una función decreciente (Ver Figura 4.2.14), por lo que el máximo

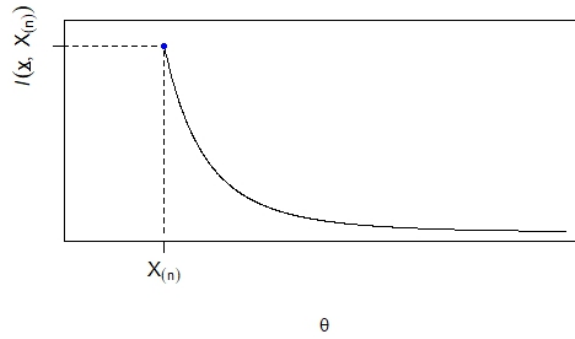


Figura 4.1: Función de verosimilitud de  $\ell(\underline{x}, \theta) = \frac{1}{\theta^n} \mathbb{1}_{(X_{(n)}, \infty)}(\theta)$

se obtiene cuando  $\theta = X_{(n)}$ , luego entonces:

$$\hat{\theta}_{MV} = X_{(n)} = \max \{X_1, \dots, X_n\}$$

Con lo anterior podemos observar que siempre puede maximizarse la función de verosimilitud, es decir, siempre existen los estimadores máximo verosímiles, sin embargo, el método que debemos utilizar dependerá de la distribución asociada. En el siguiente ejemplo veremos la aplicación de maximización por métodos numéricos. Cabe destacar que hace no muchos años estas técnicas no eran llamativas debido al gran número de cálculos aritméticos involucrados. Actualmente contamos con el poder computacional necesario, así, estas técnicas llaman la atención y resurgen para solucionar estos problemas. Por lo anterior, haremos uso de herramientas computacionales con el software estadístico **R-project** y utilizaremos el método numérico Newton-Raphson.

**Ejemplo 4.2.15.** Sea  $X_1, \dots, X_n$  m.a. con distribución Gamma  $\Gamma(\alpha, \beta)$  donde  $\alpha > 0, \beta > 0$  ambos desconocidos. Su función de densidad viene dada por:

$$f_X(x; \alpha, \beta) := \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}}; \quad x \in (0, \infty)$$

Nuestro objetivo es encontrar los EMV. Ahora, sabemos que:

$$\mathbb{E}(X) = \alpha\beta \quad \text{Var}(X) = \alpha\beta^2$$

Se puede probar que la Log-verosimilitud viene dada por:

$$L(\underline{x}; \alpha, \beta) = -n \text{Log}(\Gamma(\alpha)) - n\alpha \text{Log}(\beta) + (\alpha - 1) \sum_{i=1}^n \text{Log}(x_i) - \frac{1}{\beta} \sum_{i=1}^n x_i$$

Y las ecuaciones de verosimilitud son:

$$0 = \frac{\partial}{\partial \alpha} L(\underline{x}; \alpha, \beta) = -n \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} - n \text{Log}(\beta) + \sum_{i=1}^n x_i \quad (4.1)$$

$$0 = \frac{\partial}{\partial \beta} L(\underline{x}; \alpha, \beta) = \frac{1}{\beta} \left( -n\alpha + \frac{1}{\beta} \sum_{i=1}^n x_i \right) \quad (4.2)$$

Despejando a  $\beta$  de 4.2 tenemos que:

$$\beta = \frac{\bar{x}}{\alpha} \quad (4.3)$$

Si sustituimos 4.3 en 4.1 obtenemos:

$$0 = -n \text{Log}(\psi(\alpha)) - n \text{Log}(\bar{x}) + n \text{Log}(\alpha) + \sum_{i=1}^n \text{Log}(x_i) \quad (4.4)$$

Donde

$$\psi(\alpha) := \frac{d}{d\alpha} \text{Log}(\Gamma(\alpha)) = \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} = \int_0^{\infty} \left( \frac{e^{-t}}{t} - \frac{e^{-\alpha t}}{1 - e^{-t}} \right) dt$$

Se le llama la función digamma. Observemos pues que no es posible tener una expresión explícita para  $\hat{\alpha}_{MV}$  de  $\alpha$  que es raíz de 4.4, pero si podemos aproximarlos por métodos numéricos. Sea:

$$g(\alpha) = -n \text{Log}(\psi(\alpha)) - n \text{Log}(\bar{x}) + n \text{Log}(\alpha) + \sum_{i=1}^n \text{Log}(x_i) \quad (4.5)$$

De forma que la ecuación 4.4 es  $0 = g(\alpha)$ . El método Newton-Raphson aproxima la raíz mediante la siguiente sucesión:

$$\alpha_m = \alpha_{m-1} - \frac{g(\alpha_{m-1})}{g'(\alpha_{m-1})} = \alpha_{m-1} \left( 1 + \frac{\psi(\alpha_{m-1}) + \text{Log}(\bar{x}) + \text{Log}(\alpha_{m-1}) - \overline{\text{Log}(x_i)}}{1 - \alpha_{m-1} \psi'(\alpha_{m-1})} \right) \quad (4.6)$$

Donde:

$$\overline{\text{Log}(x_i)} = \sum_{i=1}^n \frac{\text{Log}(x_i)}{n}$$

$$g'(\alpha) = -n\psi'(\alpha) + \frac{n}{\alpha}$$

A la función  $\psi'()$  se le conoce como función trigamma. Tanto la digamma como la trigamma son funciones que ya se encuentran implementadas en R. Su puede tomar como valor inicial de la iteración el estimador por momentos para  $\alpha$ . Así  $\alpha_0 = \frac{\bar{x}^2}{s^2}$

```
#####
###Programa que encuentra los#####
```

```

###estimadores MV de una dis-#####
###tribucion Gamma ambos para-#####
### metros desconocidos #####
#####

emvalphaGamma = function(muestra){
# Calcula el e.m.v. del parametro alpha de una Gamma
m <- mean(muestra)
v <- var(muestra)
alpha0 <- 2/v # punto de inicio = estimador de los momentos
error <- 1
while(error>1E-4){
alpha1 <- GammaNewton(muestra,alpha0)
error <- abs(alpha1-alpha0)
alpha0 <- alpha1
}
alpha1
}

GammaNewton = function(muestra,alpha0){
# Sucesion que converge al e.m.v. del parametro alpha de la gamma
m <- mean(muestra)
mlog <- mean(log(muestra))
alpha1 <- alpha0 *(1+(digamma(alpha0)+log(m/alpha0)-mlog)/(1-alpha0*
trigamma(alpha0)))
}

```

Para probar la eficiencia de este método simularemos una m.a. de tamaño 50 de parámetros  $\alpha = 5$  y  $\beta = 2$

```

n = 50
alpha = 5
beta = 2
sample <-rgamma(n,shape=alpha,scale=beta)
m <- mean(sample)
emvalpha <- emvalphaGamma(X)
embeta <- m/emvalpha
writeLines(paste('alpha =',alpha,' e.m.v. (alpha) =',round(emvalpha,digits=4)))
writeLines(paste('beta =',beta,' e.m.v. (beta) =',round(embeta,digits=4)))

```

La salida producida es:

```

e.m.v. (alpha) = 5.1786
e.m.v. (beta) = 1.9241

```

Notemos que los resultados obtenidos son estimadores muy parecidos a los parámetros.

Es de vital importancia reconocer el hecho de que este método no es el único así como tampoco es el mejor. Recordemos que para tener una buena estimación necesitamos tener una buena muestra, por lo que debe recordarse que el muestreo debe ser realizado correctamente. Por último, existe paquetería implementada en R para estimación máximo verosímil por métodos numéricos. Investigue `LIBRARY(MAXLIK)`

**Ejemplo 4.2.16.** Sea  $X_1, \dots, X_n$  m.a. con distribución  $\exp(\lambda)$  con  $\lambda > 0$  y cuya función de densidad dada por:

$$f_X(x; \lambda) = \lambda e^{-\lambda x} \mathbb{1}_{(0, \infty)}(x)$$

Utilizando `LIBRARY(MAXLIK)` encuentre una aproximación al estimador MV para  $\lambda$ . Simularemos una m.a. de tamaño 10 de parámetro  $\lambda = 2$  y utilizaremos la función `MAXLIK()` utilizando el método Newton-Raphson y compararemos finalmente con la estimación proporcionada por el método de la segunda derivada.

```
n<-10
lambda <- 2
sample<-rexp(n,lambda)
#Generando funcion de verosimilitud
l.verosimilitud<-function(lambda){
  return(sum(log(dexp(sample,lambda))))
}
#Utilizando funcion maxLik()
library(maxLik)
estimation<-maxLik(l.verosimilitud,start=c(1),method="NR")
summary(estimation)
```

Se produce la siguiente salida:

```
-----
Maximum Likelihood estimation
Newton-Raphson maximisation, 5 iterations
Return code 1: gradient close to zero
Log-Likelihood: -3.488871
1 free parameters
Estimates:
      Estimate Std. error t value Pr(> t)
[1,]  1.9177    0.6063   3.163 0.00156 **
```

Si comparamos contra la estimación por el método de la segunda derivada tenemos que:

```
estimate<-1/mean(sample)
> estimate
[1] 1.917674
```

Tenemos que ambas estimaciones coinciden. Podemos ver en la Figura 4.2.16 que la estimación no dista mucho del verdadero valor del parámetro  $\lambda$ .

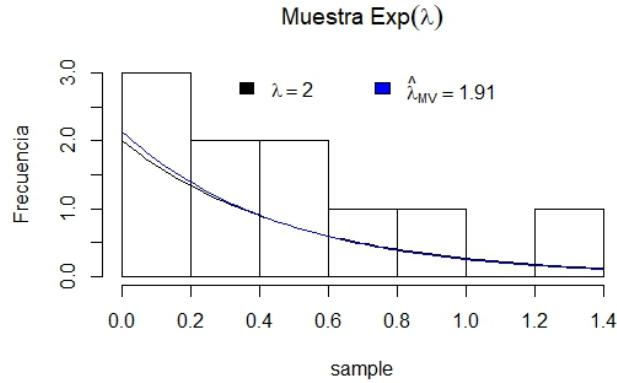


Figura 4.2: Estimación numérica  $Exp(\lambda)$

**Ejercicio 4.2.7.** 1. Sea  $X_1, \dots, X_n$  m.a. del modelo Geométrico( $p$ )

$$f_X(x) = \mathbb{P}(X = x) = (1 - p)^{x-1} p \quad x \in \{1, 2, 3, \dots\}$$

Encuentre el estimador de momentos y el de máxima verosimilitud y responda lo siguiente:

- ¿Los estimadores encontrados son insesgados?
- Simule una m.a. de tamaño 50 y realice un histograma incluyendo la función de densidad asociada al tipo de parámetros estimados y la densidad del parámetro verdadero.

2. Sea  $X_1, \dots, X_n$  m.a. del modelo Exponencial( $\lambda$ )

$$f_X(x) = \lambda e^{-\lambda x} \quad x \in (0, \infty)$$

Encuentre el estimador de momentos y el de máxima verosimilitud para  $\lambda$  y responda lo siguiente:

- ¿Los estimadores encontrados son insesgados?
- Simule una m.a. de tamaño 100 y obtenga el valor de la estimación del parámetro  $\hat{\lambda}_M$  además  $\hat{\lambda}_{MV}$  por método de la segunda derivada y método numérico.
- Realice un histograma incluyendo la función de densidad asociada al tipo de parámetros estimados y la densidad del parámetro verdadero.

3. Sea  $X_1, \dots, X_n$  m.a. de una población con función de densidad dada por:

$$f_X(x; \theta) = \frac{2}{\theta^2}(\theta - x), x \in [0, \theta]$$



- a) *Aplica el método de momentos para encontrar un estimador para  $\theta$ . (Denotemos a dicho estimador como  $\hat{\theta}_M$ ).*
- b) *Demuestre que  $\hat{\theta}_M$  es insesgado.*
- c) *Encuentre el Error Cuadrático Medio de  $\hat{\theta}_M$ .*
- d) *Simule una m.a. de tamaño 40 y realice un histograma incluyendo la función de densidad asociada al tipo de parámetro estimado por el inciso anterior y la densidad del parámetro verdadero.*

4. Sea  $X_1, \dots, X_n$  m.a. de una población con densidad dada por:

$$f_X(x; \theta) = \frac{1}{\theta} x^{\frac{1}{\theta}-1}, x \in [0, 1] \quad \theta > 0$$

- *Demuestre que la densidad pertenece a la familia exponencial.*
- *Demuestre que  $\sum_{i=1}^n \text{Log}(X_i)$  es una estadística completa para  $\theta$ .*
- *Aplica el método de momentos para encontrar un estimador para  $\theta$ . (Denotemos a dicho estimador como  $\hat{\theta}_M$ ).*
- *¿ $\hat{\theta}_M$  es insesgado?*
- *Aplica el método de máxima verosimilitud para encontrar un estimador para  $\theta$ . (Denotemos a dicho estimador como  $\hat{\theta}_{MV}$ ).*
- *Demuestre que  $\hat{\theta}_{MV}$  es insesgado.*
- *De los estimadores calcule el Error Cuadrático Medio.*
- *¿El estimador  $\hat{\theta}_{MV}$  es UMVUE? (Justifique su respuesta)*
- *Simule una m.a. de tamaño 30 y realice un histograma incluyendo la función de densidad asociada al tipo de parámetros estimados por los incisos anteriores y la del verdadero parámetro. Visualmente hablando ¿cuál estimador es mejor?*

En general sabemos que una de las ventajas de este tipo de estimadores es que tiene una **distribución asintótica normal** (Ver Apéndice).

$$\hat{\theta}_{MV} \overset{\text{approx}}{\sim} N\left(\theta, \frac{1}{nI(\theta)}\right)$$

Esto también se puede generalizar al caso de varios parámetros pero esta fuera del alcance de este curso. A continuación algunos ejemplos de la distribución asintótica de los estimadores máximos verosimiles.

**Ejemplo 4.2.17.** *Algunas distribuciones de la familia exponencial:*

- En el modelo Bernoulli ( $\theta$ ):

$$\hat{\theta}_{MV} = \bar{X}; \quad I(\theta) = \frac{1}{\theta(1-\theta)}; \Rightarrow \hat{\theta}_{MV} = \bar{X} \overset{approx}{\sim} N\left(\theta, \frac{\theta(1-\theta)}{n}\right)$$

- En el modelo Poisson ( $\lambda$ ):

$$\hat{\lambda}_{MV} = \bar{X}; \quad I(\lambda) = \frac{1}{\lambda}; \Rightarrow \hat{\lambda}_{MV} = \bar{X} \overset{approx}{\sim} N\left(\lambda, \frac{\lambda}{n}\right)$$

- En el modelo Exponencial ( $\lambda$ ):

$$\hat{\lambda}_{MV} = \frac{1}{\bar{X}}; \quad I(\lambda) = \frac{1}{\lambda^2}; \Rightarrow \hat{\lambda}_{MV} = \frac{1}{\bar{X}} \overset{approx}{\sim} N\left(\lambda, \frac{\lambda^2}{n}\right)$$

Veamos un ejemplo con R.

**Ejemplo 4.2.18.** Sea  $X_1, \dots, X_n$  m.a. del modelo  $Exp(\lambda)$  con  $\lambda > 0$ . Simularemos 10000 muestras de tamaño 1000 y  $\lambda = 2$ . Verificaremos pues la distribución asintótica de  $\hat{\lambda}_{MV}$ .

A cada una de estas muestras se hará la estimación MV para  $\lambda$  y posteriormente su histograma añadiendo la distribución asintótica normal.

```
lambda <- 2
n <- 1000
nSim <- 10000
est.mv <- rep(0, nSim)
for(i in 1:nSim){
  sample <- rexp(n, lambda)
  est.mv[i] <- 1/mean(sample)
}

library(lattice)
hist(est.mv, breaks=30, freq=FALSE, main=expression("Distribucion asintotica de " +
  (hat(lambda)[MV])), xlab="Clases", ylab="Densidad")
curve(dnorm(x, lambda, sqrt(lambda^2/n)), add=TRUE, col=2)
legend("topright", expression(N(2, frac(2^2, n))), fill="red", ncol=1, cex=1, bty="n")
```

De la Figura 4.2.18 es claro ver que la distribución asintótica es normal a medida que se incrementa el tamaño de muestra.

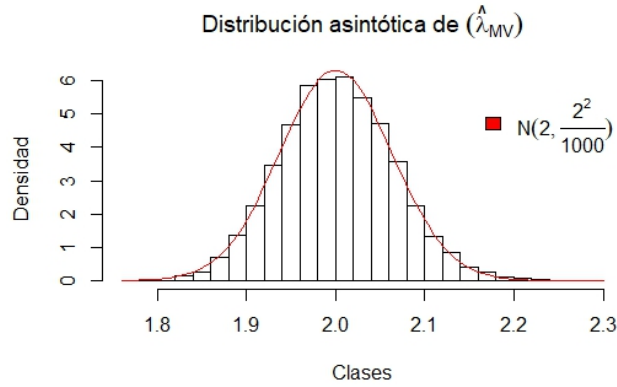


Figura 4.3: Distribución Asintótica de  $\hat{\lambda}_{MV}$  del modelo  $Exp(\lambda)$

**Observación:** Recuerde que para que funcione la aproximación se deben de cumplir las condiciones de regularidad, por ejemplo en el caso del modelo  $U(0, \theta)$  el soporte de la densidad depende del parámetro desconocido, por lo tanto no se cumple las condiciones y es **incorrecto** afirmar que

$$\hat{\theta}_{MV} = X_{(n)} \underset{\text{aprox}}{\sim} N\left(\theta, \frac{1}{nI(\theta)}\right)$$

A continuación un ejemplo de **invarianza**.

**Ejemplo 4.2.19.** Consideremos  $X_1, \dots, X_n$  m.a. del modelo *Poisson* ( $\lambda$ ), sabemos que  $\hat{\lambda}_{MV} = \bar{X}$ , sin embargo consideremos que estamos interesados en estimar  $\mathbb{P}(X = 0) = e^{-\lambda}$ . Observemos que  $e^{-\lambda}$  es una transformación biyectiva, entonces:

$$\mathbb{P}(\widehat{X = 0})_{M.V.} = e^{-\bar{X}}$$

Además por las propiedades asintóticas sabemos que:

$$\mathbb{P}(\widehat{X = 0})_{M.V.} = e^{-\bar{X}} \underset{\text{aprox}}{\sim} N\left(e^{-\lambda}, \frac{\lambda e^{-2\lambda}}{n}\right)$$

**Ejercicio 4.2.8.** Resuelva los siguientes ejercicios:

1. Considere una m.a. de tamaño  $n$  *Uniforme*  $(0, \theta)$  Encuentre los estimadores *MV* para la media y la varianza.
2. Considere una m.a. de tamaño  $n$  *Bin*  $(1, \theta)$  sabiendo que  $\bar{X}$  es un estimador completo para  $\theta$  encuentre el *UMVUE* para:
  - a)  $3\theta$

b)  $3\theta - 1$

c) Encuentre el estimador MV para  $\theta$  y  $\theta(1 - \theta)$

3. Considere una m.a. de tamaño  $n$ , estimar el parámetro  $\theta$  por el método de momentos. La función de densidad es la siguiente:

$$f(x; \theta) = \frac{2(\theta - x)}{\theta^2} \quad l_{0 < x < \theta} \quad (4.7)$$

Verifique si el estimador es insesgado.

4. Sea  $X_1, \dots, X_n$  una m.a. con densidad común

$$f(x; \theta) = e^{-(x-\theta)} \quad l_{\theta < x < \infty} \quad -\infty < \theta < \infty \quad (4.8)$$

- Encuentre el estimador de MV de  $\theta$
- Encuentre el estimador por momentos de  $\theta$

5. Sea  $X_1, \dots, X_n$  una m.a. con densidad común

$$f(x; \theta) = \theta x^{-2} \quad l_{\theta < x < \infty} \quad 0 < \theta < \infty \quad (4.9)$$

Encuentre el estimador MV para  $\theta$

6. Sea  $X_1, \dots, X_n$  una m.a. con densidad común  $N(\mu, \sigma^2)$ . Encuentre el UMVUE para

a)  $6\mu + 4\sigma^2$

b)  $\mu^2 - 5\sigma^2$

7. Sea  $\mu$  el verdadero valor del cociente intelectual de cierto estudiante. Para medir su C.I. realiza un test y se sabe que las puntuaciones obtenidas se distribuye normalmente, con media  $\mu$  y desviación estándar 5. El estudiante realiza el test y alcanza la puntuación 130. ¿Cuál es el estimador MV de  $\mu$ ?

## 4.3. Intervalos de Confianza

### 4.3.1. Introducción

El objetivo principal para construir un intervalo de confianza es encontrar dos estadísticas  $L(\underline{X})$  y  $U(\underline{X})$  tales que:

$$\mathbb{P}(L(\underline{X}) \leq \theta \leq U(\underline{X})) = 1 - \alpha$$

Donde  $\alpha$  toma valores pequeños (0.01, 0.05, 0.1).

Debe observarse que el límite inferior y superior al ser función de variables aleatorias entonces también son aleatorios, luego entonces debido a que  $\theta$  es considerado un parámetro fijo, la interpretación que debemos dar a la expresión anterior es **la probabilidad de que el intervalo cubra al parámetro desconocido  $\theta$**

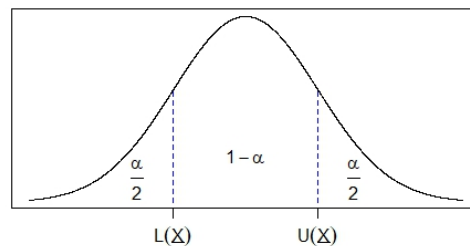


Figura 4.4: Intervalo de confianza

Por otro lado, una vez observada la muestra y dado que ya no tenemos variables aleatorias, entonces al intervalo generado por las estadísticas evaluadas en la muestra observada se le conoce como intervalo al  $(1 - \alpha) \times 100\%$  de confianza para  $\theta$ . En la *Figura 4.3.1* podemos observar una representación gráfica de un intervalo de confianza suponiendo una distribución simétrica.

Cabe destacar que los intervalos de confianza no son intervalos de probabilidad, éstos últimos están fuera del alcance de este curso. Para información más detallada consulte bibliografía relacionada con Estadística Bayesiana.

A continuación describimos el método tradicional para encontrar los estadísticos  $L(\underline{X})$  y  $U(\underline{X})$  que dan origen a un intervalo de confianza.

**Definición 4.3.1** (Cantidad Pivotal para  $\theta$ ). *Una cantidad pivotal es una función de la muestra  $X_1, \dots, X_n$  cuya regla de correspondencia debe de involucrar a  $\theta$  pero que su función de distribución **no** dependa de parámetros desconocidos.*

**Observación:** Sabiendo esto es claro que pueden existir muchos intervalos de confianza para un mismo parámetro, intervalos que cumplen con la propiedad deseada de cubrir al parámetro desconocido con una probabilidad  $1-\alpha$ , surge entonces la pregunta de saber con cuál intervalo quedarnos. La respuesta lógica es escoger aquel intervalo de longitud mínima. En general los intervalos óptimos estarán en función de estadísticas con propiedades importantes como suficiencia y completas. *Ver apéndice.*

### 4.3.2. Intervalos de confianza: Caso Normal

La distribución normal es considerada como la más fascinante de la teoría probabilística y definitivamente en estadística debido a sus características de forma y localización. Es por ello que nos centraremos en estudiar esta distribución con mayor detalle.

#### Intervalo para la media $\mu$ , $\sigma^2$ conocida

Supongamos que tenemos  $X_1, \dots, X_n$  m.a. del una población con distribución  $N(\mu, \sigma^2)$  y analicemos el caso en donde  $\sigma^2$  es conocida. Entonces definamos:

$$Q(\underline{X}) := \sqrt{n} \left( \frac{\bar{X} - \mu}{\sigma} \right) \sim N(0, 1)$$

$Q$  es entonces una cantidad pivotal para  $\mu$  de donde afirmamos también que:

$$\mathbb{P} \left( Z_{(\alpha/2)} \leq \sqrt{n} \left( \frac{\bar{X} - \mu}{\sigma} \right) \leq Z_{(1-\alpha/2)} \right) = 1 - \alpha$$

Donde  $Z_{(\alpha/2)}$  y  $Z_{(1-\alpha/2)}$  como siempre, representan los cuantiles  $(1-\alpha/2)$  de una distribución normal estándar (*Figura 4.3.2*). Ahora, por propiedades de simetría tenemos que  $Z_{(\alpha/2)} = -Z_{(1-\alpha/2)}$ .

Distribución normal estándar

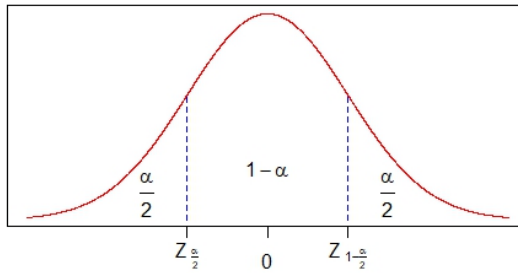


Figura 4.5: Distribución de la cantidad pivotal  $Q$  para el I.C. de  $\mu$ . Caso:  $N(\mu, \sigma^2)$  con  $\sigma^2$  conocida

Así, substituyendo y llevando a cabo un despeje del parámetro desconocido  $\mu$  obtenemos la siguiente expresión:

Intervalo de confianza para  $\mu$

$$\mathbb{P}\left(\bar{X} - \frac{\sigma}{\sqrt{n}}Z_{(1-\alpha/2)} \leq \mu \leq \bar{X} + \frac{\sigma}{\sqrt{n}}Z_{(1-\alpha/2)}\right) = 1 - \alpha \quad (4.10)$$

Cuya longitud es:

$$\begin{aligned} \text{Longitud} &= \bar{X} + \frac{\sigma}{\sqrt{n}}Z_{(1-\alpha/2)} - \left(\bar{X} - \frac{\sigma}{\sqrt{n}}Z_{(1-\alpha/2)}\right) \\ &= 2\frac{\sigma}{\sqrt{n}}Z_{(1-\alpha/2)} \end{aligned}$$

**Ejemplo 4.3.1.** Se ha obtenido una muestra de 25 alumnos del Posgrado para estimar la calificación media de los estudiantes. Se sabe que las calificaciones siguen una distribución normal con desviación estándar 2.01. Si la media estimada fue de 4.9, calcule el intervalo de confianza para la calificación media al 90% y 99%.

Generamos el siguiente código

```
#####
#Funcion que genera IC para mu N(mu,sigma^2), sigma conocida##
#####

norm.interval = function(n,media,var,alpha)
{
  z = qnorm(1 - alpha/2)
  EE = sqrt(var/n)
  c(media - z * EE, media + z * EE)
}
```

Produciendo la siguiente salida:

```
#Para alpha=10%
round(norm.interval(25,4.9,2.01^2,0.1),2)
[1] 4.24 5.56
> #Para alpha=1%
> round(norm.interval(25,4.9,2.01^2,0.01),2)
[1] 3.86 5.94
```

A continuación haremos un breve paréntesis sobre el tamaño de muestra necesario para hacer una estimación con cierto error haciendo uso de la teoría desarrollada en la construcción de intervalos de confianza y con el modelo propuesto.

### Tamaño de muestra

Vimos entonces que la mejor intervalo es el proporcionado en la ecuación (4.10) , imaginemos ahora que queremos construir un intervalo de longitud  $2d$  y con una confianza de  $1 - \alpha$  (Figura 4.3.2). A  $d$  se le conoce como la **precisión**.

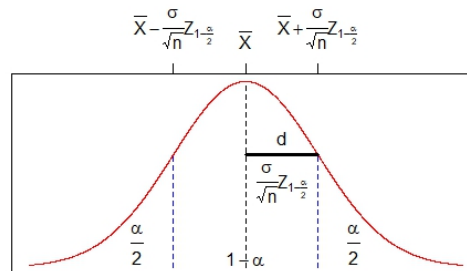


Figura 4.6: Precisión del I.C. de  $\mu$ . Caso:  $N(\mu, \sigma^2)$  con  $\sigma^2$  conocida

Surge de manera natural la pregunta de saber cuál deberá ser el tamaño de muestra necesario para cumplir este objetivo. La respuesta se obtiene despejando entonces a  $n$  de la siguiente igualdad:

$$2 \frac{\sigma}{\sqrt{n}} Z_{(1-\alpha/2)} = 2d \quad \Rightarrow \quad n = \frac{\sigma^2 Z_{(1-\alpha/2)}^2}{d^2}$$

Ahora notemos lo siguiente: si  $\frac{\sigma}{\sqrt{n}} Z_{(1-\alpha/2)} = d$  entonces por (4.10) obtenemos que:

$$\mathbb{P}(\bar{X} - d \leq \mu \leq \bar{X} + d) = 1 - \alpha \Rightarrow \mathbb{P}(|\bar{X} - \mu| \leq d) = 1 - \alpha$$

Luego entonces, esto nos dice que hemos encontrado el tamaño de muestra necesario para que el estimador  $\hat{\mu} = \bar{X}$  tenga un error menor a  $d$  unidades con una probabilidad de  $1 - \alpha$ .



**Ejemplo 4.3.2.** Supongamos que queremos inferir la estatura promedio de un mexicano mayor de 18 años. Para resolver este problema asumiremos la hipótesis distribucionales de que la estatura de los mexicanos adultos se pueden modelar por medio del modelo normal. Supongamos además para fines prácticos que conocemos la desviación estándar del modelo  $\sigma = 10$ . ¿Qué tamaño de muestra se requiere para poder estimar  $\mu$  si se desea tener un error menor a 1 cm con una confianza del 95 %?. La respuesta según la teoría desarrollada nos dice que la  $n$  requerida es:

$$n = \frac{10^2 * (1.96)^2}{1} = 384.16$$

Debido a que el tamaño de muestra solo toma valores enteros, entonces tomamos  $n = 385$ . Lo que nos afirma entonces que debemos tener una muestra de ese tamaño para que nuestro intervalo de confianza tenga una longitud de 2cm, lo que se traduce en un error menor a 1cm en nuestra estimación con una confianza del 95 %

### Intervalo para la media $\mu$ , $\sigma^2$ desconocida

Cuando  $\sigma^2$  es desconocida entonces la expresión:

$$\sqrt{n} \left( \frac{\bar{X} - \mu}{\sigma} \right) \sim N(0, 1)$$

Deja de ser cantidad pivotal para  $\mu$ , pues  $\sigma^2$  es desconocida y recordemos que la única cantidad desconocida debe de ser  $\mu$ , por lo tanto debemos de buscar alguna otra cantidad, para ello recordemos que:

$$\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2} \sim \chi_{(n-1)}^2$$

Además recordemos que si  $Z \sim N(0, 1)$  y  $Y \sim \chi_{(n-1)}^2$  (independientes) entonces

$$\frac{Z}{\sqrt{\frac{Y}{n-1}}} \sim t_{(n-1)} \text{ T student}$$

Por lo tanto concluimos que: (Obs: Recordar la independendencia entre  $\bar{X}$  y  $S^2$ )

$$\frac{\sqrt{n} \left( \frac{\bar{X} - \mu}{\sigma} \right)}{\sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2} / (n-1)}} = \sqrt{n} \left( \frac{\bar{X} - \mu}{\hat{\sigma}} \right) \sim t_{(n-1)}$$

Donde  $\hat{\sigma}^2 := \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{(n-1)}$ , la cual ya puede ser considerada cantidad pivotal (Figura 4.3.2) para  $\mu$ .

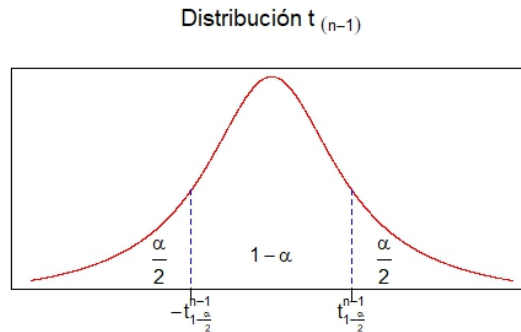


Figura 4.7: Distribución de la cantidad pivotal para el I.C. de  $\mu$ . Caso:  $N(\mu, \sigma^2)$  con  $\sigma^2$  desconocida

Entonces nuestro intervalo de confianza después de llevar a cabo el despeje estaría dado por:

$$\mathbb{P}\left(\bar{X} - \frac{\hat{\sigma}}{\sqrt{n}}t_{(n-1)}^{1-\alpha/2} \leq \mu \leq \bar{X} + \frac{\hat{\sigma}}{\sqrt{n}}t_{(n-1)}^{1-\alpha/2}\right) = 1 - \alpha$$

Donde  $t_{(n-1)}^{1-\alpha/2}$  es el cuantíl  $1 - \alpha/2$  de una distribución  $t$ -student con  $n - 1$  grados de libertad.

**Ejemplo 4.3.3.** *Se ha obtenido una muestra de 15 vendedores de seguros para estimar el valor medio de las ventas de los agentes de la compañía. Se presume que la distribución de las ventas sigue una distribución normal. Se ha estimado que la media y la varianza son 5 y 4 respectivamente. Calcule el intervalo de confianza al 95 % y 97 % para la venta media utilizando **R-project***  
 Tenemos el siguiente código:

```
#####
#Funcion que genera IC para mu de N(mu,sigma^2), sigma desconocida##
#####

interval = function(n,media,var,alpha)
{
  t = qt(1 - alpha/2,n-1)
  EE = sqrt(var/n)
  c(media - t * EE, media + t * EE)
}
```

Produciendo la siguiente salida:

```
> #Para alpha=5%
> round(interval(15,5,4,0.1),2)
[1] 4.09 5.91
> #Para alpha=3%
> round(interval(15,5,4,0.03),2)
[1] 3.75 6.25
```

## Intervalo para $\sigma^2$ , $\mu$ conocida

Suponiendo  $\mu$  conocida, consideremos la siguiente cantidad pivotal.

$$\sum_{i=1}^n \frac{(X_i - \mu)^2}{\sigma^2} \sim \chi_n^2$$

La anterior es claro que cumple las condiciones de ser llamada cantidad pivotal pues recordemos que  $\mu$  es conocida, luego entonces considerando  $a$  y  $b$  dos cantidades tales que:

$$\mathbb{P} \left( a \leq \sum_{i=1}^n \frac{(X_i - \mu)^2}{\sigma^2} \leq b \right) = 1 - \alpha \quad (4.11)$$

Al despejar a  $\sigma^2$  de esta expresión se obtiene el siguiente intervalo de confianza:

$$\mathbb{P} \left( \sum_{i=1}^n \frac{(X_i - \mu)^2}{b} \leq \sigma^2 \leq \sum_{i=1}^n \frac{(X_i - \mu)^2}{a} \right) = 1 - \alpha \quad (4.12)$$

Donde  $a$  y  $b$  se escogen de tal manera que se cumpla la ecuación (4.11) y que se obtenga la longitud mínima en la ecuación (4.12). Este problema desafortunadamente no arroja una solución algebraica para  $a$  y  $b$  por lo que su solución se obtiene por métodos numéricos. Algunos libros por facilidad optan por definir:

$$a = \chi_n^{2(\alpha/2)} \quad y \quad b = \chi_n^{2(1-\alpha/2)}$$

Quedando el intervalo de confianza como sigue:

$$\mathbb{P} \left( \sum_{i=1}^n \frac{(X_i - \mu)^2}{\chi_n^{2(1-\alpha/2)}} \leq \sigma^2 \leq \sum_{i=1}^n \frac{(X_i - \mu)^2}{\chi_n^{2(\alpha/2)}} \right) = 1 - \alpha$$

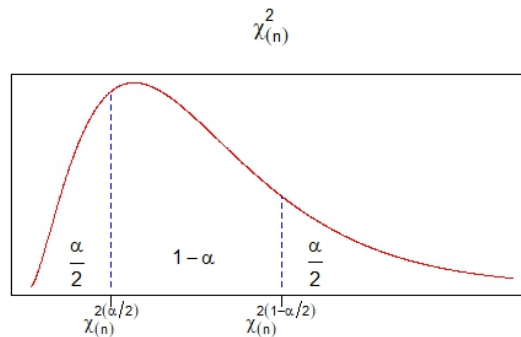


Figura 4.8: Distribución de la cantidad pivotal para el I.C. de  $\sigma^2$ . Caso:  $N(\mu, \sigma^2)$  con  $\mu$  conocida

Sin embargo debe de quedar claro que este intervalo (*Figura 4.3.2*) no necesariamente genera al óptimo (de longitud menor) sin embargo es el más práctico para fines didácticos.

**Ejemplo 4.3.4.** *Un fabricante de fibras sintéticas desea estimar la varianza de la tensión de ruptura de una fibra. Se diseña un experimento en el que se observan las tensiones medias de ruptura, en libras, de 16 hilos del proceso, seleccionados aleatoriamente. Las tensiones son 20.8, 20.6, 21.0, 19.9, 20.2, 19.6, 19.8, 20.9, 21.1, 20.4, 20.6, 19.7, 19.6, 20.3 y 20.7. Suponga que la tensión de ruptura de una fibra se encuentra modelada por una distribución normal con media de 20.3 Construya un intervalo de confianza estimado del 98 % para el valor real de la varianza de tensión de ruptura de la fibra utilizando **R**.*

```
#####
#Funcion que genera IC para sigma^2 N(mu,sigma^2), mu conocida##
#####

interval = function(datos,media,alpha)
{
  V = sum((datos-media)^2)
  c(V/qchisq(1-alpha/2,length(datos)),V/qchisq(alpha/2,length(datos)))
}
```

*Que produce la siguiente salida:*

```
> datos<-c(20.8,20.6, 21.0,19.9,20.2,19.6,19.8,20.9,21.1,20.4,20.6,19.7,19.6,20.3,20.7)
> #Para alpha=2%
> round(interval(datos,20.3,0.02),2)
[1] 0.13 0.74
```

### Intervalo para $\sigma^2$ , $\mu$ desconocida

Si ahora suponemos  $\mu$  desconocida simplemente procederemos a estimar a  $\mu$  por medio de  $\bar{X}$  y sustituyendolo en la cantidad pivotal anterior , de esta manera se obtiene la siguiente cantidad pivotal:

$$\sum_{i=1}^n \frac{(X_i - \bar{X})^2}{\sigma^2} \sim \chi_{n-1}^2$$

Y de forma análoga el intervalo de confianza queda como:

$$\mathbb{P} \left( \sum_{i=1}^n \frac{(X_i - \bar{X})^2}{b} \leq \sigma^2 \leq \sum_{i=1}^n \frac{(X_i - \bar{X})^2}{a} \right) = 1 - \alpha \quad (4.13)$$

Donde nuevamente  $a$  y  $b$  son cantidades que hacen que cumpla la confianza y se tenga la cobertura mínima, y como en el punto anterior, a veces  $a$  y  $b$  son calculados como los siguientes cuantiles para fines prácticos :

$$a = \chi_{n-1}^{2(\alpha/2)} \quad y \quad b = \chi_{n-1}^{2(1-\alpha/2)}$$

**Ejemplo 4.3.5.** Considere el ejemplo anterior y suponga que desconoce la media de ruptura de la fibra. Calcule el intervalo de confianza para la varianza al 98 % utilizando **R**. ¿Qué nota con respecto al intervalo anterior? Justifique su respuesta.

Tenemos el siguiente código:

```
#####
#Funcion que genera IC para sigma^2 N(mu,sigma^2), mu desconocida##
#####

interval = function(datos,alpha)
{
  V = sum((datos-mean(datos))^2)
  c(V/qchisq(1-alpha/2,length(datos)-1),V/qchisq(alpha/2,length(datos)-1))
}
```

Que produce la siguiente salida:

```
> datos<-c(20.8,20.6, 21.0,19.9,20.2,19.6,19.8,20.9,21.1,20.4,20.6,19.7,19.6,20.3,20.7)
> #Para alpha=2%
> round(interval(datos,0.02),2)
[1] 0.13 0.82
```

Observemos que el intervalo de confianza ha incrementado su longitud aunque tienen el mismo nivel de confianza, esto se debe al hecho de hacer tanto la estimación de la varianza como la de la media por lo que perdemos un grado de libertad en la distribución de la cantidad pivotal.

A continuación construiremos intervalos de confianza para la diferencia de medias en 2 poblaciones Normales, este un punto muy importante ya que a través de estos intervalos se concluyen diferencias de medias significativas que muchas veces son utilizadas para identificar la eficacia de tratamientos.

### Intervalo para la diferencia de Medias en 2 dos poblaciones. Caso 1: Varianzas conocidas

Supongamos que tenemos  $X_1, \dots, X_{n_1}$  y  $Y_1, \dots, Y_{n_2}$  dos muestras aleatorias independientes entre sí de dos poblaciones normales  $N(\mu_x, \sigma_x^2)$ ,  $N(\mu_y, \sigma_y^2)$  respectivamente. Además supongamos que tanto  $\sigma_x^2$  como  $\sigma_y^2$  son conocidas. Nuestro objetivo es lograr construir un intervalo de confianza para la diferencia de medias  $(\mu_x - \mu_y)$ . Para ello primero buscaremos una cantidad pivotal de la siguiente manera, sabemos que por ser v.a. normales que:

$$\bar{X} \sim N\left(\mu_x, \frac{\sigma_x^2}{n_1}\right) \quad y \quad \bar{Y} \sim N\left(\mu_y, \frac{\sigma_y^2}{n_2}\right)$$

Además por la independencia del muestreo entre las dos poblaciones se tiene que  $\bar{X}$  es independiente de  $\bar{Y}$  por lo que es fácil ver que:

$$\frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_x^2}{n_1} + \frac{\sigma_y^2}{n_2}}} \sim N(0, 1)$$

Lo anterior ya es una cantidad pivotal la diferencia de medias  $(\mu_x - \mu_y)$ , pues recordemos que estamos suponiendo que tanto  $\sigma_x^2$  como  $\sigma_y^2$  son conocidas no necesariamente iguales. Luego entonces de esta cantidad pivotal obtenemos el siguiente intervalo:

$$\mathbb{P} \left( (\bar{X} - \bar{Y}) - \sqrt{\frac{\sigma_x^2}{n_1} + \frac{\sigma_y^2}{n_2}} Z_{1-\alpha/2} \leq \mu_x - \mu_y \leq (\bar{X} - \bar{Y}) + \sqrt{\frac{\sigma_x^2}{n_1} + \frac{\sigma_y^2}{n_2}} Z_{1-\alpha/2} \right) = 1 - \alpha$$

Donde denotamos a  $Z_{1-\alpha/2}$  como el cuantíl  $1 - \alpha/2$  de la distribución Normal Estándar.

**Ejemplo 4.3.6.** Se observa la eficiencia de dos departamentos asignándole a cada uno de ellos diez tareas y midiendo su rendimiento en ellas. Los resultados están a continuación:

<b>Departamento 1</b>	0.6	1.2	0.9	1.9	2.0	0.6	0.9	2.0	0.8	1
<b>Departamento 2</b>	0.4	1.3	1.1	2.1	1.9	0.5	1.1	1.7	0.8	1.1

Suponiendo las puntuaciones como variables normales e independientes entre sí.

Considere  $\sigma_x^2 = \sigma_y^2 = 0.32$  calcule el intervalo de confianza del 90 % para la diferencia media de eficiencia.

Tenemos el siguiente código:

```
#####
#Funcion que genera IC para diferencia de medias, varianzas conocidas##
#####

interval = function(datosx,datosy,varx,vary,alpha)
{
  Est = mean(datosx)-mean(datosy)
  EE = sqrt((varx/length(datosx))+((vary/length(datosy))))*qnorm(1-alpha/2)
  c(Est-EE,Est+EE)
}
```

Que produce la siguiente salida:

```
> datosx<-c(.6,1.2,.9,1.9,2,.6,.9,2,.8,1)
> datosy<-c(.4,1.3,1.1,2.1,1.9,.5,1.1,1.7,.8,1.1)
> #Para alpha=10%
> round(interval(datosx,datosy,0.32,0.32,0.1),4)
[1] -0.4261 0.4061
```

## Intervalo para la diferencia de Medias en 2 dos poblaciones; Caso 2: Varianzas desconocidas pero iguales

Bajo el mismo contexto anterior pero asumiendo que  $\sigma_x^2 = \sigma_y^2 = \sigma^2$  se obtiene que:

$$\frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}} \sim N(0, 1)$$

$$\frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim N(0, 1)$$

Sin embargo al ser  $\sigma^2$  desconocida, la expresión anterior no puede ser considerada una cantidad pivotal para la diferencia de medias  $\mu_x - \mu_y$ , la forma en como atacaremos este problema es la siguiente, primero notemos que tenemos muestra de dos poblaciones normales, entonces:

$$\sum_{i=1}^{n_1} \frac{(X_i - \bar{X})}{\sigma^2} \sim \chi_{(n_1-1)}^2 \quad \sum_{i=1}^{n_2} \frac{(Y_i - \bar{Y})}{\sigma^2} \sim \chi_{(n_2-1)}^2$$

Entonces por independencia entre ambas muestras podemos concluir que:

$$\sum_{i=1}^{n_1} \frac{(X_i - \bar{X})}{\sigma^2} + \sum_{i=1}^{n_2} \frac{(Y_i - \bar{Y})}{\sigma^2} = \frac{\sum_{i=1}^{n_1} (X_i - \bar{X})^2 + \sum_{i=1}^{n_2} (Y_i - \bar{Y})^2}{\sigma^2} \sim \chi_{(n_1+n_2-2)}^2$$

Tenemos entonces una cantidad con distribución  $N(0, 1)$  y otra con distribución  $\chi^2$ , suponiendo independencia entre estas dos cantidades se concluye que:

$$\frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \bigg/ \sqrt{\frac{\sum_{i=1}^{n_1} (X_i - \bar{X})^2 + \sum_{i=1}^{n_2} (Y_i - \bar{Y})^2}{\sigma^2(n_1 + n_2 - 2)}} \sim t_{n_1+n_2-2}$$

Definiendo  $\hat{\sigma}_p^2$  como:

$$\hat{\sigma}_p^2 = \frac{\sum_{i=1}^{n_1} (X_i - \bar{X})^2 + \sum_{i=1}^{n_2} (Y_i - \bar{Y})^2}{n_1 + n_2 - 2} = \frac{(n_1 - 1)\hat{\sigma}_x^2 + (n_2 - 1)\hat{\sigma}_y^2}{n_1 + n_2 - 2}$$

Obtemos que:

$$\frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\hat{\sigma}_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2}$$

Lo anterior ya es una cantidad pivotal la cual obtiene el siguiente intervalo de confianza para la diferencia de medias:

$$\mathbb{P} \left( (\bar{X} - \bar{Y}) - \hat{\sigma}_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} t_{(n_1+n_2-2)}^{(1-\alpha/2)} \leq \mu_x - \mu_y \leq (\bar{X} - \bar{Y}) + \hat{\sigma}_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} t_{(n_1+n_2-2)}^{(1-\alpha/2)} \right) = 1 - \alpha$$

**Ejemplo 4.3.7.** *Continuando con el Ejemplo 4.3.6 Supongamos ahora que desconocemos las varianzas pero hay evidencia que nos dice que son iguales. Calcule el intervalo de confianza para la diferencia de medias al 90 %. Así, tenemos el siguiente código:*

```
#####
#Funcion que genera IC para dif de medias, varianzas desconocidas e iguales##
#####
interval = function(datosx,datosy,alpha)
{
  Est = mean(datosx)-mean(datosy)
  n1 =length(datosx)
  n2 =length(datosy)
  varp= ((n1-1)*var(datosx)+(n2-1)*var(datosy))/(n1+n2-2)
  EE = varp*sqrt((1/n1)+(1/n2))*qt(1-alpha/2,n1+n2-2)
  c(Est-EE,Est+EE)
}
```

```
> #Para alpha=10%
> round(interval(datosx,datosy,0.1),4)
[1] -0.4483 0.4283
```

### Intervalo para la diferencia de Medias en 2 dos poblaciones; Caso 3: Varianzas desconocidas y no iguales

Este quizás sea el caso mas común al que nos estaremos enfrentando, pero desafortunadamente no existe una cantidad pivotal que se pueda utilizar, por lo general para atacar este problema se utilizan los estimadores de  $\sigma^2$  de cada población y se obtiene la siguiente cantidad pivotal:

$$\frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{\frac{\hat{\sigma}_x^2}{n_1} + \frac{\hat{\sigma}_y^2}{n_2}}} \underset{\text{aprox}}{\sim} t_\nu$$

Donde:

$$\nu = \left( \frac{\hat{\sigma}_x^2}{n_1} + \frac{\hat{\sigma}_y^2}{n_2} \right)^2 \bigg/ \left( \frac{\left( \frac{\hat{\sigma}_x^2}{n_1} \right)^2}{n_1 - 1} + \frac{\left( \frac{\hat{\sigma}_y^2}{n_2} \right)^2}{n_2 - 1} \right)$$



Luego entonces un intervalo de confianza aproximado viene dado por:

$$\mathbb{P} \left( (\bar{X} - \bar{Y}) - t_{(\nu)}^{(1-\alpha/2)} \sqrt{\frac{\hat{\sigma}_x^2}{n_1} + \frac{\hat{\sigma}_y^2}{n_2}} \leq \mu_x - \mu_y \leq (\bar{X} - \bar{Y}) + t_{(\nu)}^{(1-\alpha/2)} \sqrt{\frac{\hat{\sigma}_x^2}{n_1} + \frac{\hat{\sigma}_y^2}{n_2}} \right) = 1 - \alpha$$

**Ejemplo 4.3.8.** *Continuando con el Ejemplo 4.3.6 supongamos que tenemos total desconocimiento de los parámetros pero estamos interesados en calcular el intervalo de confianza para la diferencia de medias con un nivel de significancia del 90 %. Utilice **R**. Tenemos el siguiente código:*

```
#####
#Funcion que genera IC para dif de medias, varianzas desconocidas NO iguales##
#####

interval = function(datosx,datosy,alpha)
{
  Est = mean(datosx)-mean(datosy)
  n1 = length(datosx)
  n2 = length(datosy)
  E1 = (var(datosx)/n1)+(var(datosy)/n2)
  E2 = (((var(datosx)/n1)^2)/(n1-1))+(((var(datosy)/n2)^2)/(n2-1))
  v = (E1^2)/E2
  c(Est-sqrt(E1)*qt(1-alpha/2,v),Est+sqrt(E1)*qt(1-alpha/2,v))
}
```

Que produce la siguiente salida:

```
> #Para alpha=10%
> round(interval(datosx,datosy,0.1),4)
[1] -0.4483 0.4283
```

Hacer intervalos de confianza para la diferencia se puede utilizar como sigue, imaginemos que tenemos 2 poblaciones y queremos ver si  $\mu_x > \mu_y$  a través una muestra de cada población, una forma de probarlo es construir un intervalo de confianza para la diferencia, luego entonces, si el número 0 se encuentra dentro del intervalo se dice que no hay diferencia significativa, pero si resulta que el 0 está a la izquierda del intervalo entonces  $0 < (\mu_x - \mu_y)$  lo que concluye que estadísticamente  $\mu_x > \mu_y$ .

**Ejemplo 4.3.9.** *Del Ejemplo 4.3.8 tenemos que el IC al 90 % está dado por  $(-0.4483, 0.4283)$  Observemos que este intervalo cubre al cero, por lo que tenemos que las medias de eficiencia de los dos departamentos son estadísticamente iguales.*

Supongamos ahora que estamos interesados en verificar la igualdad entre las varianzas de dos poblaciones, para atacar este problema se acostumbra construir intervalos para el cociente de

varianzas, luego entonces si el 1 se encuentra en dicho intervalo se dice que no hay diferencia significativa en cuanto a variabilidad. A continuación se exponen las cantidades pivotaes para generar los intervalos asociados a los distintos casos.

### Intervalo de confianza para el cociente de varianzas. Caso 1: Medias Conocidas

Si las medias son conocidas sabemos que :

$$\sum_{i=1}^{n_1} \frac{(X_i - \mu_x)^2}{\sigma_x^2} \sim \chi^2_{(n_1)} \quad y \quad \sum_{i=1}^{n_2} \frac{(Y_i - \mu_y)^2}{\sigma_y^2} \sim \chi^2_{(n_2)}$$

Por la independencia de los muestreos se tiene que al hacer el cociente de las  $\chi^2$ 's se tiene que:

$$\sum_{i=1}^{n_2} \frac{(Y_i - \mu_y)^2}{\sigma_y^2 n_2} \bigg/ \sum_{i=1}^{n_1} \frac{(X_i - \mu_x)^2}{\sigma_x^2 n_1} \sim F_{n_2, n_1}$$

Por lo tanto

$$\left( \frac{\sigma_x^2}{\sigma_y^2} \right) \left( \frac{\sum_{i=1}^{n_2} \frac{(Y_i - \mu_y)^2}{n_2}}{\sum_{i=1}^{n_1} \frac{(X_i - \mu_x)^2}{n_1}} \right) \sim F_{n_2, n_1}$$

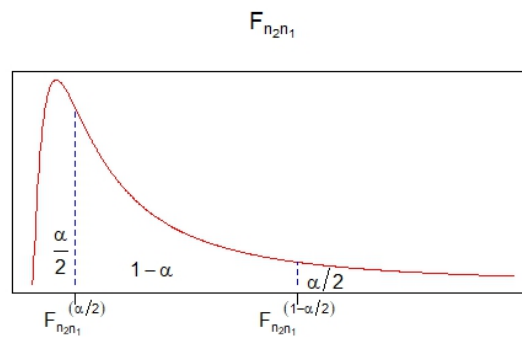


Figura 4.9: Distribución de la cantidad pivotal para el I.C. de cociente de varianzas. Caso: Medias Conocidas

La cual ya es una cantidad pivotal para el cociente de varianzas  $\left( \frac{\sigma_x^2}{\sigma_y^2} \right)$ , luego eligiendo  $a$  y  $b$  dos número reales tales que:

$$\mathbb{P} \left( a \leq \left( \frac{\sigma_x^2}{\sigma_y^2} \right) \left( \frac{\sum_{i=1}^{n_2} \frac{(Y_i - \mu_y)^2}{n_2}}{\sum_{i=1}^{n_1} \frac{(X_i - \mu_x)^2}{n_1}} \right) \leq b \right) = 1 - \alpha$$

Se obtiene el siguiente intervalo:

$$\mathbb{P} \left( \frac{\sum_{i=1}^{n_1} \frac{(X_i - \mu_x)^2}{n_1}}{\sum_{i=1}^{n_2} \frac{(Y_i - \mu_y)^2}{n_2}} a \leq \left( \frac{\sigma_x^2}{\sigma_y^2} \right) \leq \frac{\sum_{i=1}^{n_1} \frac{(X_i - \mu_x)^2}{n_1}}{\sum_{i=1}^{n_2} \frac{(Y_i - \mu_y)^2}{n_2}} b \right) = 1 - \alpha$$

Sin embargo la obtención de  $a$  y  $b$  se hace por medio de métodos numéricos para obtener el intervalo óptimo (de menor longitud), sin embargo en la literatura se acostumbra hacer:

$$a = F_{n_2, n_1}^{\alpha/2} \quad b = F_{n_2, n_1}^{1-\alpha/2}$$

Donde  $F_{n_2, n_1}^{\alpha}$  es el cuantil  $\alpha$  de la distribución  $F$  con  $n_2, n_1$  grados de libertad.

**Ejemplo 4.3.10.** *Continuando con el Ejemplo 4.3.6 supongamos que sabemos que la media de eficiencia del departamento 1 es 1.7 mientras que la del departamento 2 es 1.75 Con un nivel de significancia del 90 % calcule el intervalo de confianza para el cociente de medias ¿ Es razonable considerar que las varianzas son iguales? Justifique su respuesta. Utilice **R**. Tenemos el siguiente código:*

```
#####
#Funcion que genera IC para cociente de vaianzas. Caso: Medias Conocidas#####
#####

interval = function(datosx,medx,datosy,medy,alpha)
{

  n1 = length(datosx)
  n2 = length(datosy)
  Est = sum((datosx-medx)^2)/sum((datosy-medy)^2)*(n2/n1)
  c(Est*qt(alpha/2,n2,n1),Est*qt(1-alpha/2,n2,n1))
}
```

Que produce la siguiente salida:

```
#Para alpha=10%
round(interval(datosx,1.7,datosy,1.75,0.1),4)
[1] 0.3110 2.7588
```

Notemos que el I.C. al 90 % para el cociente de varianzas cubre al uno, por lo que estadísticamente se dice que las varianzas de ambos departamentos son iguales.

### 4.3.3. Intervalo de Confianza para poblaciones no Normales

Hasta ahora fuimos capaces de encontrar cantidades pivotaes para los parámetros de una distribución normal, sin embargo muchas veces nos enfrentaremos al problema de querer encontrar

intervalos de confianza en familias distintas a la normal, pero aún, a veces estaremos interesados en construir intervalos de confianza para funciones de cierto parámetro. Hemos visto que la clave para encontrar el intervalo de confianza se centra en poder localizar una cantidad pivotal adecuada, sin embargo en muchas ocasiones esto es una tarea difícil más que nada debido no hay una metodología para encontrarlos.

**Ejemplo 4.3.11.** *Supongamos que tenemos  $X_1, \dots, X_n$  una m.a. de una población  $U(0, \theta)$  se puede probar que:*

$$T = \frac{X_{(n)}}{\theta}$$

*Es una cantidad pivotal y por tanto:*

$$\mathbb{P} \left( \frac{X_{(n)}}{\xi_{1-\alpha/2}} \leq \theta \leq \frac{X_{(n)}}{\xi_{\alpha/2}} \right) = 1 - \alpha$$

*Es un intervalo de confianza al  $(1 - \alpha) 100\%$  para  $\theta$ , donde  $\xi_\alpha$  es el cuantíl  $\alpha$  de la distribución asociada a  $T = \frac{X_{(n)}}{\theta}$ . Con función de densidad dada por:*

$$f_T(t) = nt^{n-1} \mathbb{1}_{(0,1)} \quad n > 1$$

**Ejemplo 4.3.12.** *Supongamos que tenemos  $X_1, \dots, X_n$  m.a. de una población  $Exp(\lambda)$ , entonces se puede probar que:*

$$T_2 = \lambda \left( \sum_{i=1}^n X_i \right) \sim Gamma(n, 1)$$

*Es cantidad pivotal y por tanto:*

$$\mathbb{P} \left( \frac{\sum_{i=1}^n X_i}{\xi_{1-\alpha/2}} \leq \lambda \leq \frac{\sum_{i=1}^n X_i}{\xi_{\alpha/2}} \right) = 1 - \alpha$$

*Es un intervalo de confianza al  $(1 - \alpha) 100\%$  para  $\lambda$ , donde  $\xi_\alpha$  es el cuantíl  $\alpha$  de la distribución asociada a  $T_2 \sim Gamma(n, 1)$ .*

Existe un método general para encontrar cantidades pivotaes, sin embargo debido a su complejidad en ocasiones será difícil llevar a cabo el despeje del parámetro desconocido. El método se basa en el siguiente resultado:

**Teorema 4.3.1.** *Sea  $X_1, \dots, X_n$  m.a. de una población con función de distribución  $F_X(x; \theta)$ , (continua) con  $\theta$  un parámetro desconocido. Entonces:*

$$Q_1 = - \sum_{i=1}^n \log(F_X(X_i; \theta)) \sim Gamma(n, 1)$$

$$Q_2 = - \sum_{i=1}^n \log(1 - F_X(X_i; \theta)) \sim Gamma(n, 1)$$

Entonces  $Q_1$  y  $Q_2$  son ambas cantidad pivotaes para  $\theta$

Para ver la prueba formal consulte el *Apéndice*.

Cuando no podemos encontrar una cantidad pivotal adecuada, muchas veces debemos de recurrir a resultados asintóticos como el **Teorema del Límite Central** (4.2.1) o bien recordar la propiedad distribucional asintótica con la que cuentan los estimadores máximo verosímiles  $\hat{\theta}_{MV} \stackrel{approx}{\sim} N\left(\theta, \frac{1}{nI(\theta)}\right)$ .

Por ejemplo, gracias al *TLC* (4.2.1) podemos construir intervalos de confianza para la media  $\mu$  de cualquier distribución  $F_X(x)$  con segundo momento finito de la siguiente manera:

- Sabemos que bajo las condiciones del *TLC* (4.2.1) podemos afirmar que :

$$\bar{X} \stackrel{approx}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right) \Rightarrow \sqrt{n} \left(\frac{\bar{X} - \mu}{\sigma}\right) \stackrel{approx}{\sim} N(0, 1)$$

- Si suponemos  $\sigma^2$  conocida la última expresión puede utilizarse como cantidad pivotal y afirmar que:

$$\mathbb{P}\left(\bar{X} - \frac{\sigma}{\sqrt{n}}Z_{(1-\alpha/2)} \leq \mu \leq \bar{X} + \frac{\sigma}{\sqrt{n}}Z_{(1-\alpha/2)}\right) \stackrel{approx}{=} 1 - \alpha$$

es un intervalo aproximado al  $(1 - \alpha)100\%$  de confianza para  $\mu$

- Si suponemos que  $\sigma^2$  es desconocida, se hace una segunda aproximación y se estima  $\sigma^2$  por medio del estimador:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}$$

Quedando el intervalo de confianza como:

$$\mathbb{P}\left(\bar{X} - \frac{\hat{\sigma}}{\sqrt{n}}Z_{(1-\alpha/2)} \leq \mu \leq \bar{X} + \frac{\hat{\sigma}}{\sqrt{n}}Z_{(1-\alpha/2)}\right) \stackrel{approx}{=} 1 - \alpha$$

Sin embargo algunos autores aconsejan cambiar los cuantiles de la Normal por los de una *t-student* con  $(n - 1)$  grados de libertad, no obstante si tenemos un muestra grande ambos cuantiles son muy similares y es indiferente entre usar una *t* o una Normal para construir el intervalo. (Ver figura 4.3.3)

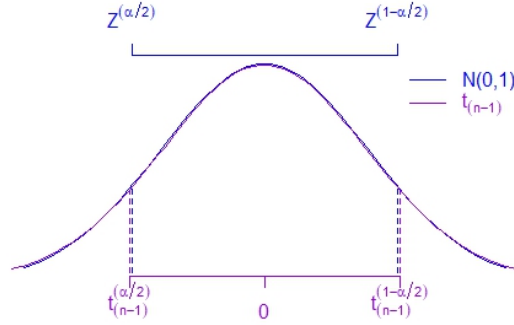


Figura 4.10: Comparación distribución normal estándar vs  $t_{(n-1)}$  para  $n > 30$

A continuación veremos el ejemplo de cómo construir un intervalo de confianza basados en estimadores máximo verosímiles, la idea del método recae en la siguiente aproximación:

$$\hat{\theta}_{MV} \overset{approx}{\sim} N\left(\theta, \frac{1}{nI(\theta)}\right) \Rightarrow \sqrt{nI(\theta)} (\hat{\theta}_{MV} - \theta) \overset{approx}{\sim} N(0, 1)$$

Debe de quedar claro que la última expresión ya es una cantidad pivotal (aproximada) para  $\theta$ , sin embargo despejar  $\theta$  de la ecuación anterior a veces puede llegar a ser difícil, en muchas ocasiones y para evitar el despeje de  $\theta$  en la cantidad pivotal se acostumbra llevar a cabo una segunda aproximación y afirmar que:

$$\hat{\theta}_{MV} \overset{approx}{\sim} N\left(\theta, \frac{1}{nI(\hat{\theta}_{MV})}\right) \Rightarrow \sqrt{nI(\hat{\theta}_{MV})} (\hat{\theta}_{MV} - \theta) \overset{approx}{\sim} N(0, 1)$$

De donde claramente el intervalo de confianza generado es de la forma:

$$\mathbb{P}\left(\hat{\theta}_{MV} - \frac{1}{\sqrt{nI(\hat{\theta}_{MV})}} Z_{1-\alpha/2} \leq \theta \leq \hat{\theta}_{MV} + \frac{1}{\sqrt{nI(\hat{\theta}_{MV})}} Z_{1-\alpha/2}\right) \overset{approx}{=} 1 - \alpha$$

**Ejemplo 4.3.13.** Sea  $X_1, \dots, X_n$  una m.a. del modelo Poisson ( $\lambda$ ). Construiremos un intervalo de confianza aproximado para  $\lambda = \mathbb{E}(X)$ , primero notemos que:

$$\hat{\lambda}_{MV} = \bar{X} \Rightarrow \bar{X} \overset{approx}{\sim} N\left(\lambda, \frac{1}{nI(\lambda)}\right)$$

Como en este modelo:

$$I(\lambda) = -\mathbb{E}\left(\frac{\partial^2}{\partial \lambda^2} f_X(X; \lambda)\right) = \frac{1}{\lambda}$$

Se concluye entonces que la cantidad pivotal es:

$$\frac{\hat{\lambda}_{MV} - \lambda}{\sqrt{\frac{\lambda}{n}}} \underset{\text{aprox}}{\sim} N(0, 1)$$

Despejar  $\lambda$  no es sencillo pero se puede llevar a cabo:

$$\begin{aligned} \mathbb{P}\left(\left|\frac{\hat{\lambda}_{MV} - \lambda}{\sqrt{\frac{\lambda}{n}}}\right| \leq Z_{1-\alpha}\right) &= 1 - \alpha \\ \mathbb{P}\left(\left(\frac{\hat{\lambda}_{MV} - \lambda}{\sqrt{\frac{\lambda}{n}}}\right)^2 \leq Z_{1-\alpha}^2\right) &= 1 - \alpha \\ \mathbb{P}\left((\hat{\lambda}_{MV} - \lambda)^2 \leq Z_{1-\alpha}^2 \frac{\lambda}{n}\right) &= 1 - \alpha \\ \mathbb{P}\left(\hat{\lambda}_{MV}^2 - 2\hat{\lambda}_{MV}\lambda + \lambda^2 \leq Z_{1-\alpha}^2 \frac{\lambda}{n}\right) &= 1 - \alpha \\ \mathbb{P}\left(n\hat{\lambda}_{MV}^2 - 2n\hat{\lambda}_{MV}\lambda + n\lambda^2 \leq Z_{1-\alpha}^2 \lambda\right) &= 1 - \alpha \\ \mathbb{P}\left(n\lambda^2 - (Z_{1-\alpha}^2 + 2n\hat{\lambda}_{MV})\lambda + n\hat{\lambda}_{MV}^2 \leq 0\right) &= 1 - \alpha \end{aligned}$$

La última expresión es una ecuación de segundo grado para  $\lambda$  la cual puede ser resuelta via la formula tradicional para encontrar raíces, haciendo lo siguiente:

$$\begin{aligned} a &= n \\ b &= -(Z_{1-\alpha}^2 + 2n\hat{\lambda}_{MV}) \\ c &= n\hat{\lambda}_{MV}^2 \end{aligned}$$

Obteniendo finalmente el siguiente intervalo de confianza:

$$\mathbb{P}\left(\frac{-b - \sqrt{b^2 - 4ac}}{2a} \leq \lambda \leq \frac{-b + \sqrt{b^2 - 4ac}}{2a}\right) = 1 - \alpha$$

Otra forma de obtener un intervalo de confianza y para no pasar por todo el proceso de despeje es

hacer otra aproximación y utilizar el hecho que:

$$\frac{\hat{\lambda}_{MV} - \lambda}{\sqrt{\frac{\hat{\lambda}_{MV}}{n}}} \underset{aprox}{\sim} N(0, 1)$$

De donde fácilmente el intervalo de confianza obtenido está dado por:

$$\mathbb{P} \left( \hat{\lambda}_{MV} - \sqrt{\frac{\hat{\lambda}_{MV}}{n}} Z_{1-\alpha/2} \leq \lambda \leq \hat{\lambda}_{MV} + \sqrt{\frac{\hat{\lambda}_{MV}}{n}} Z_{1-\alpha/2} \right) = 1 - \alpha$$

Debe de quedar claro que ambos son intervalos aproximados, sin embargo el segundo requirió una segunda aproximación, lo cual sugiere que su cobertura esperada no sera tan buena como la del primer caso sin embargo via simulaciones se puede probar que ambos intervalos son iguales con  $n$  mayores a 30.

#### 4.3.4. Intervalo de Confianza para transformaciones de los parámetros

Imaginemos que ahora no estamos interesados es construir un intervalo para  $\theta$  sino que estamos interesados en construir un intervalo para alguna transformación  $f(\theta)$ , para ello podemos recurrir a la siguiente aproximación:

$$f(\hat{\theta}_{MV}) \underset{aprox}{\sim} N \left( f(\theta), \frac{(f'(\theta))^2}{nI(\theta)} \right) \Rightarrow \frac{f(\hat{\theta}_{MV}) - f(\theta)}{\sqrt{\frac{(f'(\theta))^2}{nI(\theta)}}} \underset{aprox}{\sim} N(0, 1)$$

La última expresión ya es una cantidad pivotal para  $f(\theta)$  sin embargo, casi siempre esta ecuación será difícil de trabajar por lo que en general lo que se hace es hacer una segunda aproximación y utilizar la siguiente cantidad pivotal:

$$\frac{f(\hat{\theta}_{MV}) - f(\theta)}{\sqrt{\frac{(f'(\hat{\theta}_{MV}))^2}{nI(\hat{\theta}_{MV})}}} \underset{aprox}{\sim} N(0, 1)$$

De donde el intervalo de confianza aproximado está dado por:

$$\mathbb{P} \left( f(\hat{\theta}_{MV}) - \sqrt{\frac{(f'(\hat{\theta}_{MV}))^2}{nI(\hat{\theta}_{MV})}} Z_{1-\alpha/2} \leq f(\theta) \leq f(\hat{\theta}_{MV}) - \sqrt{\frac{(f'(\hat{\theta}_{MV}))^2}{nI(\hat{\theta}_{MV})}} Z_{1-\alpha/2} \right) = 1 - \alpha$$



**Ejemplo 4.3.14.** Sea  $X_1, \dots, X_n$  m.a. del modelo Gamma suponiendo  $\alpha$  conocido, constituir un intervalo de confianza para  $\mathbb{P}(X > x)$  dado por:

$$\mathbb{P}(X > x) = \int_x^\infty \frac{\beta^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\beta t} dt$$

Como  $\alpha$  es conocido debe de quedar claro que  $\mathbb{P}(X > x)$  queda en función de  $\beta$  y por tanto lo que nos piden es construir un intervalo de confianza para una transformación del parámetro desconocido  $\beta$ .

Primero obtenemos el estimador máximo verosímil y la información de fisher por unidad muestral correspondiente.

$$\hat{\beta}_{MV} = \frac{\alpha}{\bar{X}}; \quad I(\beta) = -\mathbb{E}\left(\frac{\partial^2}{\partial \beta^2} f_X(X; \beta)\right) = \frac{\alpha}{\beta^2}$$

Definamos  $f(\beta) := \mathbb{P}(X > x)$ , entonces:

$$f(\hat{\beta}_{MV}) \stackrel{approx}{\sim} N\left(f(\beta), \frac{(f'(\beta))^2 \beta^2}{n\alpha}\right)$$

De donde la cantidad pivotal nos queda como:

$$\frac{f(\hat{\beta}_{MV}) - f(\beta)}{\sqrt{\frac{(f'(\beta))^2 \beta^2}{n\alpha}}} \stackrel{approx}{\sim} N(0, 1)$$

Debido a la complejidad que tiene la función  $f$  haremos una segunda aproximación obteniendo la siguiente cantidad:

$$\frac{f(\hat{\beta}_{MV}) - f(\beta)}{\sqrt{\frac{(f'(\hat{\beta}_{MV}))^2 \hat{\beta}_{MV}^2}{n\alpha}}} \stackrel{approx}{\sim} N(0, 1)$$

De donde el intervalo de confianza esta dado por:

$$\mathbb{P}\left(f(\hat{\beta}_{MV}) - \sqrt{\frac{(f'(\hat{\beta}_{MV}))^2 \hat{\beta}_{MV}^2}{n\alpha}} Z_{1-\alpha/2} \leq f(\beta) \leq f(\hat{\beta}_{MV}) + \sqrt{\frac{(f'(\hat{\beta}_{MV}))^2 \hat{\beta}_{MV}^2}{n\alpha}} Z_{1-\alpha/2}\right) = 1-\alpha$$

Donde recordemos que:

$$f(\beta) := \mathbb{P}(X > x) = \int_x^\infty \frac{\beta^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\beta t} dt$$

$$f(\hat{\beta}_{MV}) := \mathbb{P}(\widehat{X} > x) = \int_x^\infty \frac{\hat{\beta}_{MV}^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\hat{\beta}_{MV}t} dt$$

$$f'(\beta) := \frac{\partial}{\partial \beta} \mathbb{P}(\widehat{X} > x) = \frac{\partial}{\partial \beta} \int_x^\infty \frac{\beta^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\beta t} dt$$

En la última igualdad podemos aplicar las condiciones de regularidad y llegar a que:

$$f'(\beta) = \int_x^\infty \frac{\partial}{\partial \beta} \frac{\beta^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\beta t} dt$$

### 4.3.5. Intervalos via Simulación

La simulación de variables aleatorias permite hacer aproximaciones muy precisas del comportamiento distribucional de la población, por ejemplo imaginemos que sabemos que  $X \sim F(\mu, \sigma^2)$  con  $F$  alguna función de distribución conocida. Supongamos que nos interesa saber el valor de  $\mu = \mathbb{E}(X)$ . Sabemos que para conocer dicho valor tenemos que llevar a cabo el cálculo de la esperanza lo cual en ocasiones nos lleva a problemas de integración difíciles. Supongamos que somos capaces de simular observaciones de la v.a.  $X$ , es decir, tenemos un algoritmo que no genera  $x_1, \dots, x_n$  observaciones de la v.a.  $X$ , entonces por la ley fuerte de los grandes números sabemos que:

$$\mathbb{E}(X) \approx \bar{x}$$

De hecho sabemos que si  $n$  tiende a infinito entonces  $\bar{x} \rightarrow \mu = \mathbb{E}(X)$ , por lo tanto una forma de aproximar el valor desconocido  $\mu$  es simular **muchas** observaciones de la v.a.  $X$  y luego asumir que  $\mathbb{E}(X) \approx \bar{x}$ .

Otra de las grandes ventajas de poder simular de una v.a. es que nos permite aproximar el comportamiento de transformaciones de variables aleatorias, imaginemos que nos interesa conocer el comportamiento distribucional de la v.a.  $Y = \sin(U)$  donde  $U \sim Uniforme(0, 2\pi)$ . Una forma de resolver el problema es realizar el cambio de variable correspondiente y encontrar  $f_Y(y)$ , sin embargo, via simulación podemos aproximar las cantidades de interés de  $Y$ , para ello supongamos que podemos simular observaciones de la v.a.  $U(0, 2\pi)$  de tal forma que tenemos  $u_1, \dots, u_n$  muestra observada de la v.a.  $U$ , para cada muestra observada procedemos a transformar  $y_1 = \sin(u_1), \dots, y_n = \sin(u_n)$ , entonces se prueba que la muestra  $y_1, \dots, y_n$  son observaciones (simulaciones) de la v.a.  $Y$ . Entonces si estamos interesados por ejemplo en encontrar a  $\mathbb{E}(Y) = \mathbb{E}(\sin(U))$ , entonces bastaría con simular una gran cantidad de observaciones la v.a.  $U$  luego transformarla para obtener  $y_1, \dots, y_n$  finalmente sabemos que una buena aproximación para  $\mathbb{E}(Y)$  es:

$$\mathbb{E}(Y) = \mathbb{E}(\sin(U)) \approx \bar{y}$$

En inferencia estadística la simulación juega un papel importante pues nos permite aproximar la distribución del estimador  $\hat{\theta}$ , como vimos, el hecho de conocer la distribución del estimador nos

permite construir cantidades pivotaes. Para llevar a cabo la aproximación por medio de simulaciones veremos dos casos, el Bootstrap Paramétrico y el no paramétrico.

## Bootstrap Paramétrico

El bootstrap es una técnica estadística popularizada por Bradley Efron en 1979. Aunque notablemente simple de implementar, el bootstrap no sería factible sin el poder computacional moderno. La idea clave es realizar cálculos sobre los propios datos para estimar la variación de las estadísticas que se calculan a partir de los mismos datos. Tales técnicas existían antes de 1979, pero Efron amplió su aplicabilidad y demostró cómo implementar la técnica Bootstrap con eficacia usando las computadoras. También acuñó el término "bootstrap". Nuestra aplicación principal del bootstrap será estimar la variación de las estimaciones puntuales, es decir, para estimar los intervalos de confianza. Anteriormente hemos trabajado con los distintos casos en los que la distribución es normal, ahora supongamos que nos encontramos en el caso en el que la distribución no es normal, aunque hemos probado que siempre que exista el primer momento de la v.a. aún podemos estimar puntualmente a  $\mu$  por  $\bar{x}$ . Pero, ¿cómo podemos encontrar un intervalo de confianza para  $\mu$ ? Nuestra respuesta será usar el bootstrap.

De hecho, esta técnica puede implementarse para intervalos de confianza de otras estadísticas diferentes de la media que incluso para el modelo normal pueden ser difíciles de manejar.

En estadística se considera que una **muestra** es un subconjunto de elementos tomados de un conjunto, también se entiende por **muestra aleatoria** a aquella muestra donde los elementos son elegidos aleatoriamente, a la elección de elementos de la muestra se le conoce comúnmente como **muestreo** básicamente podemos hablar en términos generales que podemos hacer el muestreo considerando el hecho que una vez seleccionado el elemento en muestra, podamos seleccionarlo de nuevo, por otro lado, también podemos pensar que una vez seleccionado el elemento en muestra, no sea seleccionado de nuevo posteriormente. A la muestra obtenida por el primer método se le conoce como **muestra aleatoria con reemplazo** mientras que por el segundo método se le conoce como **muestra aleatoria sin reemplazo**, o bien **muestra aleatoria simple**.

**Observación:** En la práctica si tomamos una muestra muy grande entonces no importa si tomamos muestras con o sin reemplazo. Por ejemplo: si seleccionamos al azar 100 de 110 millones de celulares en el mundo, entonces es tan poco probable que el mismo celular se escoja dos veces por lo que no hay diferencia real entre muestreo con o sin reemplazo.

Ahora, supongamos que tenemos una muestra aleatoria con función de distribución conocida pero cuyo parámetro es desconocido, una vez que estimamos puntualmente al parámetro tenemos "totalmente especificada" la distribución de los datos, con esta distribución podríamos generar o simular nuevas muestras aleatorias del mismo tamaño que la original, a este proceso se le conoce como **remuestreo** mientras que a la distribución utilizada se le conoce también como **distribución de remuestreo**.

**Observación:** El tamaño de muestra del remuestreo puede ser cualquiera, sin embargo, algunos autores recomiendan por lo menos 1000 remuestreos. A continuación presentaremos el algoritmo bootstrap para calcular intervalos de confianza.

### 4.3.6. Algoritmo Bootstrap

Sea  $X_1, X_2, \dots, X_n$  m.a. de algún modelo  $F_X(x; \theta)$  una muestra bootstrap es una remuestra  $X_1^*, X_2^* \dots X_n^*$  de tamaño  $n$  tomada de la distribución de remuestreo  $F_X(x; \hat{\theta})$ . Para una estadística  $\hat{\theta}$  calculada con los datos de la muestra original, podemos definir una estadística  $\hat{\theta}^*$  calculada de la misma fórmula de  $\hat{\theta}$  pero con los datos del remuestreo.

La técnica bootstrap paramétrico se usa como sigue:

1. Sea  $x_1, x_2, \dots, x_n$  muestra observada de de una distribución  $F_X(x; \theta)$ .
2.  $\hat{\theta}$  es una estadística calculada a partir de la muestra que estima a  $\theta$ .
3. Nuestras muestras bootstrap son tomadas con  $F_X(x; \hat{\theta})$ .
4. Para cada muestra bootstrap  $x_1^*, x_2^* \dots x_n^*$  calculamos  $\hat{\theta}^*$  y la diferencia bootstrap como  $\delta^* = \hat{\theta}^* - \hat{\theta}$ .
5. Hacer el histograma de  $\delta^*$ .
6. Calcular los cuantiles al  $\delta_{\frac{\alpha}{2}}^*$  y  $\delta_{1-\frac{\alpha}{2}}^*$ .
7. Entonces el principio bootstrap dice que la distribución de  $\delta^*$  aproxima la distribución de  $\delta = \hat{\theta} - \theta$ .
8. Use las diferencias bootstrap para calcular el intervalo de confianza de  $\theta$ . Como:

$$[\hat{\theta} - \delta_{1-\alpha/2}^*, \hat{\theta} - \delta_{\alpha/2}^*]$$

El interés real es que al poder estimar la variación de  $\delta$  por la de  $\delta^*$  podemos utilizar el error estándar de esta estimación para estimar el intervalo de confianza de  $\theta$ .

Por otro lado, es de vital importancia que el tamaño de la remuestra sea igual que la muestra original pues la variación de la estadística dependerá de  $n$ .

**Ejemplo 4.3.15.** *Suponga que tiene  $x_1, x_2, \dots, x_{300}$  datos que se distribuyen  $Exp(\lambda)$  con la parametrización usual. Estos datos arrojaron que  $\bar{x} = 2$ . Calcule el intervalo de confianza para  $\lambda$  al 95 % por el método bootstrap.*

*Generamos el siguiente código*

```
#####
#####Bootstrap parametrico Exp(lambda)#####
#####
alpha<- .05
n<-300
xbarra<-2
lambdaest<-1/xbarra
#Generando nboot muestras bootstrap de tamaño n
nboot<-1000
x<-rexp(n*nboot,lambdaest)
muestraboot<-matrix(x,nrow=n,ncol=nboot)
#Calculando lambda estimada asterisco para cada remuestra
lambdaaste<-1/colMeans(muestraboot)
#Calculando las diferencias delta asterisco
deltaaste<-lambdaaste-lambdaest
#Graficando el histograma de delta asterisco
hist(deltaaste,main=expression(delta^{ "*"}),xlab="Clases",ylab="",freq=FALSE,col="blue")
#Calculando los cuantiles alpha/2 y 1-alpha/2 de la distribución delta asterisco
d<-quantile(deltaaste,c(alpha/2,1-alpha/2))
#Calculando el intervalo de confianza al (1-alpha)% para lambda
ic<-lambdaest-c(d[2],d[1])
#Renombrando entradas
names(ic)<-c(names(ic)[2],names(ic)[1])
```

Que produce la siguiente salida:

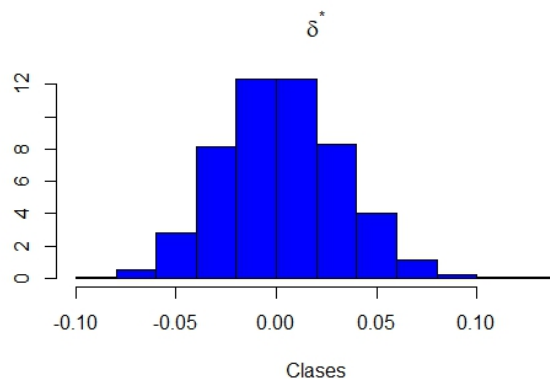


Figura 4.11: Histograma de  $\delta^*$

```
2.5%      97.5%
0.4379306 0.5524665
```

Observando la Figura 4.3.15 como la distribución de  $\delta^*$  está centrada alrededor del cero. Esta condición es necesaria dada la construcción que se ha realizado.

## Otras estadísticas

Para el caso en que sea de nuestro interés el estimar intervalos de confianza por el método bootstrap para otra estadística  $h(\theta)$  basta con encontrar un estimador  $\widehat{h(\theta)}$  y evaluar en la muestra original, después al hacer el remuestreo con la distribución de remuestreo  $F_X(x; \hat{\theta})$  calcularemos entonces  $\widehat{h(\theta)}^*$ . Calculamos las diferencias  $\delta^* = \widehat{h(\theta)}^* - \widehat{h(\theta)}$  y el intervalo queda como sigue:

$$[\widehat{h(\theta)} - \delta_{1-\alpha/2}^*, \widehat{h(\theta)} - \delta_{\alpha/2}^*]$$

**Ejemplo 4.3.16.** *Suponga estamos bajo el mismo contexto que el Ejemplo anterior, pero ahora se desea estimar la varianza del modelo y encontrar un intervalo de confianza al 95 %.*

*Solución:* Tenemos que para este modelo  $\hat{\lambda}_{MV} = \bar{x}$ , así, por principio de invarianza tenemos que el estimador MV para  $h(\lambda) = \frac{1}{\lambda^2}$  está dado por:

$$\widehat{h(\lambda)}_{MV} = h(\hat{\lambda}_{MV}) = \frac{1}{\bar{x}^2}$$

Así generamos el siguiente código:

```
#####  
#####Bootstrap parametrico lambda^-2#####  
#####  
alpha<-.05  
n<-300  
xbarra<-2  
lambdaest<-1/xbarra  
hlambdaest<-xbarra^-2  
#Generando nboot muestras bootstrap de tamaño n  
nboot<-1000  
x<-rexp(n*nboot,lambdaest)  
muestraboot<-matrix(x,nrow=n,ncol=nboot)  
#Calculando hlambda estimada asterisco para cada remuestra  
hlambdaaste<-(colMeans(muestraboot))^-2  
#Calculando las diferencias delta asterisco  
deltaaste<-hlambdaaste-hlambdaest  
#Graficando el histograma de delta asterisco  
hist(deltaaste,main=expression( delta^{"*"}),xlab="Clases",ylab="",freq=FALSE,col="yellow")  
#Calculando los cuantiles alpha/2 y 1-alpha/2 de la distribución delta asterisco  
d<-quantile(deltaaste,c(alpha/2,1-alpha/2))  
#Calculando el intervalo de confianza al (1-alpha)% para lambda  
ic<-hlambdaest-c(d[2],d[1])  
#Renombrando entradas  
names(ic)<-c(names(ic)[2],names(ic)[1])
```

Que produce la siguiente salida:

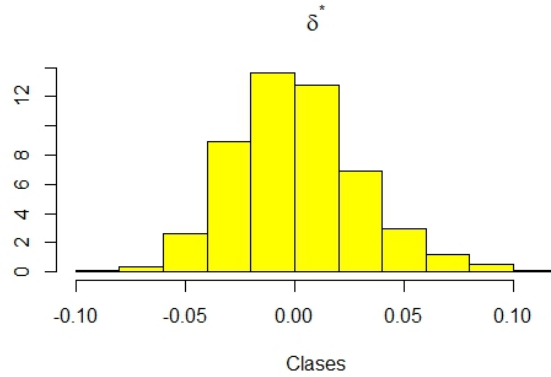


Figura 4.12: Histograma de  $\delta^*$  para estimar  $h(\lambda) = \frac{1}{\lambda^2}$

2.5%      97.5%  
 0.1825834   0.2988793

**Ejercicio 4.3.1.** *Resuelva lo siguiente:*

1. Sea  $X_1, \dots, X_n$  m.a. del modelo *Uniforme*(0,  $\theta$ ) con  $\theta > 0$ .

$$f_X(x) := \frac{1}{\theta} \mathbf{1}_{(0,\theta)}(x)$$

*El objetivo del ejercicio es encontrar un intervalo de confianza para  $\theta$ .*

- a) *Defina la variable aleatoria  $T$  como:*

$$T = \frac{X_{(n)}}{\theta} = \frac{\text{máx}\{X_1, \dots, X_n\}}{\theta}$$

*y muestre que la función de densidad de  $T$  está dada por:*

$$f_T(t) = nt^{n-1} \mathbf{1}_{(0,1)}(t)$$

*Hint: Recuerde que la densidad del máximo estadístico de orden en esta familia está dada por:*

$$f_{X_{(n)}}(x) = \frac{nx^{n-1}}{\theta^n} \mathbf{1}_{(0,\theta)}(x)$$

*Por lo tanto:*

$$f_T(t) = \frac{\partial}{\partial t} F_T(t) = \frac{\partial}{\partial t} \mathbb{P}(T \leq t) = \frac{\partial}{\partial t} \mathbb{P}\left(\frac{X_{(n)}}{\theta} \leq t\right) = \frac{\partial}{\partial t} \mathbb{P}(X_{(n)} \leq t\theta) = \frac{\partial}{\partial t} \int_0^{t\theta} f_{X_{(n)}}(x) dx$$

- b) Usando el ejercicio anterior sabemos que  $T$  es una cantidad pivotal, utilice esta cantidad pivotal para encontrar un intervalo al  $(1 - \alpha)\%$  de confianza para  $\theta$ .
- c) Calcule el IC al  $(1 - \alpha)\%$  para  $\theta$  utilizando la propiedad asintótica de los estimadores máximo verosímiles.
- d) Utilizando el **Teorema 4.3.1** construya una cantidad pivotal para  $\theta$  y obtenga un intervalo al  $(1 - \alpha)\%$  de confianza para  $\theta$ .
- e) Genere un programa en **R** que calcule el IC para  $\theta$  para cualquier  $\alpha$ . Considerando los incisos anteriores.
- f) Genere un programa en **R** que calcule el IC para  $\theta$  para cualquier  $\alpha$ . Utilizando el método bootstrap.
- g) Simule una m.a. con  $n = 50$  del modelo  $U(0, \theta)$  con  $\theta = 10$ . Suponga ahora que desconoce el valor de  $\theta$ . Calcule el IC al 90% para  $\theta$  utilizando los programas elaborados en incisos anteriores y comente los resultados.

2. Sea  $X_1, \dots, X_n$  m.a. del modelo  $Exp(\lambda)$  con  $\lambda > 0$ .

$$f_X(x) := \lambda e^{-\lambda x} \mathbf{1}_{(0, \infty)}(x)$$

a) Defina la variable aleatoria  $T$  como:

$$T = \lambda \sum_{i=1}^n X_i$$

Usando la función generadora de momentos prueba que  $T \sim \text{Gamma}(n, 1)$

- b) Usando el inciso anterior sabemos que  $T$  es una cantidad pivotal, utilice esta cantidad pivotal para encontrar un intervalo al  $(1 - \alpha)\%$  de confianza para  $\lambda$ .
- c) Calcule el IC al  $(1 - \alpha)\%$  para  $\lambda$  utilizando la propiedad asintótica de los estimadores máximo verosímiles.

- Encuentra un intervalo de confianza aproximado al  $(1 - \alpha)\%$  de confianza para  $\lambda$  usando el hecho de que  $\hat{\lambda}_{MV} \overset{*}{\sim} N\left(\lambda, \frac{1}{nI(\lambda)}\right)$ , es decir, utilice la cantidad pivotal dada por:

$$\frac{\hat{\lambda}_{MV} - \lambda}{\sqrt{\frac{1}{nI(\lambda)}}} \overset{*}{\sim} N(0, 1)$$

- Encuentra un intervalo de confianza aproximado al  $(1 - \alpha)\%$  de confianza para  $\lambda$  usando una segunda aproximación suponiendo ahora que  $\hat{\lambda}_{MV} \overset{**}{\sim} N\left(\lambda, \frac{1}{nI(\hat{\lambda}_{MV})}\right)$ , es



decir, utilice la cantidad pivotal dada por:

$$\frac{\hat{\lambda}_{MV} - \lambda}{\sqrt{\frac{1}{nI(\hat{\lambda}_{MV})}}} \stackrel{**}{\sim} N(0, 1)$$

- d) Utilizando el **Teorema 4.3.1** construya una cantidad pivotal para  $\lambda$  y obtenga un intervalo al  $(1 - \alpha)\%$  de confianza para  $\lambda$ .
- e) Genere un programa en **R** que calcule el IC para  $\lambda$  para cualquier  $\alpha$ . Considerando los incisos anteriores.
- f) Simule una m.a. con  $n = 100$  del modelo  $Exp(\lambda)$  con  $\lambda = 7$ . Suponga ahora que desconoce el valor de  $\lambda$ . Calcule el IC al 98 % para  $\lambda$  utilizando los programas elaborados en incisos anteriores y comente los resultados.
3. Considera  $X_1, \dots, X_{n_x}$  m.a. de  $N(\mu_x, \sigma_x^2)$  y  $Y_1, \dots, Y_{n_y}$  m.a. de  $N(\mu_y, \sigma_y^2)$ , suponga  $\mu_x$  y  $\mu_y$  desconocidas.

a) Muestre que:

$$\left( \frac{S_x^2}{S_y^2} F_{n_y-1, n_x-1}^{\alpha/2}, \frac{S_x^2}{S_y^2} F_{n_y-1, n_x-1}^{1-\alpha/2} \right)$$

Es un intervalo al  $(1 - \alpha)\%$  de confianza para el cociente  $\frac{\sigma_x^2}{\sigma_y^2}$ . Donde:

$$S_x^2 = \frac{1}{n_x - 1} \sum_{i=1}^{n_x} (x_i - \bar{x})^2; \quad S_y^2 = \frac{1}{n_y - 1} \sum_{i=1}^{n_y} (y_i - \bar{y})^2;$$

$F_{n_y-1, n_x-1}^{1-\alpha/2}$  y  $F_{n_y-1, n_x-1}^{\alpha/2}$  son los cuantiles de una distribución  $F$  con  $(n_y - 1, n_x - 1)$  grados de libertad.

- b) Genere un código en **R** para calcular el intervalo de confianza para el cociente de varianzas a partir del inciso anterior.
- c) Genere un código en **R** que calcule el IC al  $(1 - \alpha)\%$  para el cociente de varianzas por el método bootstrap.
- d) Del Ejemplo 4.3.6 suponga que se tiene total desconocimiento de los parámetros pero estamos interesados en saber si las varianzas son iguales. ¿ Se puede afirmar a un nivel de confianza del 95 % que las varianzas son iguales? y ¿ qué observaciones puede hacer con respecto de los dos métodos y justifique cuál elegiría?
4. Se quiere verificar que la estatura promedio de las mujeres es más pequeña que la estatura promedio de los hombres. Suponga que la distribución de la estatura de tanto mujeres como hombres es modelada por una distribución Normal con misma varianza conocida para ambas

poblaciones. Es decir

$$X \sim N(\mu_x, \sigma^2) \quad Y \sim N(\mu_y, \sigma^2)$$

Donde  $X$  modela la estatura de las mujeres y  $Y$  la estatura de los hombres.

Se llevo a cabo un muestreo aleatorio en las dos poblaciones de forma independiente, el tamaño de muestra en ambos muestreos fue de 100 y se obtuvieron los siguientes resultados:

$$\bar{x} = \frac{\sum_{i=1}^{100} x_i}{100} = 169\text{cm} \quad \bar{y} = \frac{\sum_{i=1}^{100} y_i}{100} = 175\text{cm}$$

Si la varianza de ambas poblaciones es  $\sigma^2=25$ . Construya un intervalo al 95 % de confianza para la diferencia de medias  $\mu_x - \mu_y$ . Concluya si en efecto tenemos evidencia como para decir que la estatura promedio de las mujeres es más pequeña que la estatura promedio de los hombres.

5. Se desea llevar a cabo un muestreo para estimar el peso promedio en kilogramos de los hombres mayores de 18 años en México, suponiendo que el peso de esta población sigue una distribución  $N(\mu, \sigma^2)$  con  $\sigma^2$  conocida e igual a 36 ( $\sigma^2 = 36$ ). Calcule el tamaño de muestra necesario para estimar la media  $\mu$  con un error de  $d = 0.5\text{kg}$  y una confianza de 95 %. ( $\alpha = 0.05$ ).
6. Suponga que una aseguradora modela el número de siniestros semanales usando un modelo Poisson de parámetro desconocido  $\lambda$ . La aseguradora pretende encontrar un intervalo al 95 % de confianza para la probabilidad de observar 0 siniestros en una semana:

$$\mathbb{P}(X = 0) = e^{-\lambda}$$

Supongamos que la aseguradora observa la siguiente muestra:

(3, 6, 6, 2, 4, 6, 1, 5, 6, 6, 6, 7, 5, 5, 9, 2, 8, 6, 5, 10, 7, 4, 2, 5, 4, 5, 1, 4, 2, 2)

- Encuentre el valor estimado de  $\mathbb{P}(X = 0) = e^{-\lambda}$  y construya el intervalo de confianza (aproximado) por propiedades asintóticas de los estimadores máximo verosímiles.
- Calcule el IC al 95 % por el método bootstrap.

7. Sea  $X_1, \dots, X_n$  m.a. del modelo:

$$f_X(x; \lambda) := \frac{1}{\lambda} e^{-\frac{x}{\lambda}} \mathbf{1}_{(0, \infty)}(x)$$

- a) Encuentra un intervalo de confianza aproximado al  $(1 - \alpha)$  % de confianza para  $\lambda$  usando

el hecho de que  $\hat{\lambda}_{MV} \overset{*}{\sim} N\left(\lambda, \frac{1}{nI(\lambda)}\right)$ , es decir, utilice la cantidad pivotal dada por:

$$\frac{\hat{\lambda}_{MV} - \lambda}{\sqrt{\frac{1}{nI(\lambda)}}} \overset{*}{\sim} N(0, 1)$$

- b) Encuentra un intervalo de confianza **aproximado** al  $(1-\alpha)\%$  de confianza para  $f(\lambda) = P(X > 1)$ .

*Hint: Recuerde que por propiedades asintóticas de los estimadores máximo verosímiles se tiene que:*

$$f(\lambda_{MV}) \overset{*}{\sim} N\left(f(\lambda), \frac{(f'(\lambda))^2}{nI(\lambda)}\right)$$

Con  $f$  una transformación diferenciable.

- c) Encuentre un intervalo de confianza al  $(1-\alpha)\%$  para  $\lambda$  utilizando el **Teorema 4.3.1**.
- d) Para todos los ejercicios anteriores genere un código en **R** que calcule el IC al  $(1-\alpha)\%$ .
- e) Genere un código en **R** que calcule un intervalo de confianza al  $(1-\alpha)\%$  para  $\lambda$  y  $f(\lambda)$  utilizando el método bootstrap.
- f) Simule una m.a. de tamaño  $n=46$  del modelo definido arriba con  $\lambda = 23.4$ . Ahora suponga que desconoce al parámetro. Tanto para  $\lambda$  como para  $h(\lambda)$  calcule los intervalos del confianza al 90%. ¿qué observaciones puede hacer?

## 4.4. Pruebas de hipótesis

### 4.4.1. Preliminares

En esta sección estudiaremos una tercer forma de llevar a cabo inferencia sobre parámetros desconocidos de una población. Comencemos por definir que es una hipótesis.

**Definición 4.4.1.** Una hipótesis es una aseveración o afirmación sobre el comportamiento distribucional de ciertas poblaciones de interés.

A forma de ejemplo supongamos que la v.a.  $X$  modela el tiempo de vida del componente 1 y  $Y$  modela el tiempo de vida de otro componente 2. Por definición sabemos que existe una función de distribución asociada a las v.a. aleatorias definidas, por lo que quizás una hipótesis inicial que podríamos tener es afirmar que  $X$  sigue una distribución Gamma con ciertos parámetros desconocidos. En términos de notación esto lo pondremos de la siguiente forma:

$$H_0 : X \sim \text{Gamma}(\alpha, \beta)$$

Ahora, supongamos que ya tenemos cierto conocimiento del fenómeno y que conocemos que el modelo *Gamma* ajusta bien pero se desconoce los parámetros, entonces algunas hipótesis pueden ser:

$$\begin{aligned} H_0 &: \alpha = 1 \\ H_0 &: \alpha > 1 \\ H_0 &: (\alpha, \beta) = (1, 2) \end{aligned}$$

A veces también estaremos interesados en probar hipótesis sobre dos poblaciones, por ejemplo imaginando que  $X \sim \text{Gamma}(\alpha_x, \beta_x)$  y  $Y \sim \text{Gamma}(\alpha_y, \beta_y)$ , una prueba de hipótesis muy interesante sería ver si el tiempo de vida de los dos componentes es el mismo en promedio, esto lo podríamos plantear de la siguiente forma:

$$H_0 : \mu_x = \mu_y \quad \Rightarrow \quad H_0 : \frac{\alpha_x}{\beta_x} = \frac{\alpha_y}{\beta_y}$$

La idea de una prueba de hipótesis es extraer una muestra del fenómeno y luego con esta información verificar la plausibilidad de la Hipótesis en cuyo caso a la **luz** de los datos tomar una decisión sobre si rechazo o no rechazo mi hipótesis.

En este curso nos enfocaremos a pruebas sobre los parámetros suponiendo entonces que el fenómeno estudiado ya tiene un modelo específico y solo se desconoce los parámetros.

Establecer una hipótesis sobre  $\theta$  supone dividir el espacio parameral  $\Theta$  en 2 subconjuntos  $\Theta_0$  y  $\Theta_1$  tal que:

$$\Theta = \Theta_0 \cup \Theta_1 \quad \Theta_0 \cap \Theta_1 = \emptyset$$

Donde:

$$\Theta_0 := \{\theta \in \Theta | H_0 \text{ es cierta}\} \quad \Theta_1 := \{\theta \in \Theta | H_0 \text{ no es cierta}\}$$

Así definido queda claro que entonces  $\Theta_0, \Theta_1$  forman una partición del espacio parametral  $\Theta$ . A la hipótesis que queremos contrastar la llamaremos hipótesis nula y la denotamos por  $H_0$ , por otro lado llamaremos la hipótesis alternativa  $H_1$  a la formada por los valores de  $\Theta$  tales que están en  $\Theta_1$ .

**Ejemplo 4.4.1.** *Suponga que  $X \sim \text{Bernoulli}(\theta)$  y desconocemos completamente a  $\Theta$ , en ese caso queda claro que  $\Theta := \{(0, 1)\}$ . Si planteamos la siguiente hipótesis  $H_0 : \theta \leq 0.3$ , entonces quedaría claro que  $\Theta_0 = \{(0, 0.3]\}$  y por completes entonces  $\Theta_1 = \{(0.3, 1)\}$  por lo que  $H_1 : \theta > 0.3$ . Por fines prácticos en la literatura decimos que cuando planteamos la hipótesis  $H_0 : \theta \leq 0.3$  entonces*

estamos contrastando las hipótesis  $H_0$  y  $H_1$  y lo escribimos así:

$$H_0 : \theta \leq 0.3 \quad vs \quad H_1 : \theta > 0.3$$

Imaginemos que tenemos cierta información adicional y sabemos que  $\theta$  solo puede tomar dos posibles valores  $\theta_0$  y  $\theta_1$ . En este caso  $\Theta := \{\theta_0, \theta_1\}$ . Bajo este contexto tenemos entonces que tomar una decisión sobre cuál parámetro es el correcto. Planteamos entonces la prueba:

$$H_0 : \theta = \theta_0 \quad vs \quad H_1 : \theta = \theta_1$$

Notemos que en este caso, la prueba anterior se reduce a poner a competir dos modelos:

$$H_0 : X \sim \text{Bernoulli}(\theta_0) \quad vs \quad H_1 : X \sim \text{Bernoulli}(\theta_1)$$

**Definición 4.4.2** (Hipótesis Simple). *Decimos que una hipótesis es simple si determina completamente a la función de distribución, en caso contrario decimos que la hipótesis es compuesta.*

Así por ejemplo en el caso anterior:

$$H_0 : X \sim \text{Bernoulli}(\theta_0) \quad vs \quad H_1 : X \sim \text{Bernoulli}(\theta_1)$$

Estamos probando dos hipótesis simples, sin embargo suponga que bajo el mismo contexto ahora se nos plantea que  $\Theta = (0, 1)$  con las siguientes hipótesis:

$$H_0 : \theta = \theta_0 \quad vs \quad H_1 : \theta \neq \theta_0$$

Luego entonces, en este último caso estamos contrastando una hipótesis simple contra una compuesta y finalmente si se plantean la hipótesis:

$$H_0 : \theta \leq \theta_0 \quad vs \quad H_1 : \theta > \theta_0$$

se contrasta una hipótesis compuesta contra compuesta.

Una vez planteado tanto el espacio parametral como la hipótesis nula, la siguiente pregunta natural que surge en la teoría es saber como decidir si rechazamos o no rechazamos nuestra hipótesis. Resulta lógico que para tomar nuestra decisión tenemos que tener una muestra de la población que se está estudiando.

**Definición 4.4.3** (Espacio Muestral). *Sea  $X_1, X_2, \dots, X_n$  muestra aleatoria de un fenómeno aleatorio con función de densidad  $f_X(x; \theta)$ . El espacio muestral es un subconjunto de  $\mathbb{R}^n$  que contiene*

a todos los posibles valores de la muestra de tamaño  $n$  y lo denotamos como  $\mathfrak{X}$ .

$$\mathfrak{X} = \{\underline{x} = (x_1, \dots, x_n) \in \mathbb{R}^n \mid x_i \in \{\text{Soporte de } X_i\}\}$$

### Caso: Simple vs Simple

Imaginemos que tenemos un fenómeno modelado por una densidad *Bernoulli* ( $\theta$ ), supongamos que  $\Theta = \{0.1, 0.8\}$  y nos plantemos la hipótesis  $H_0 : \theta = 0.1$ , entonces estaremos contrastando las hipótesis:

$$H_0 : \theta = 0.1 \quad vs \quad H_1 : \theta = 0.8$$

Supongamos ahora que observaremos una muestra aleatoria de tamaño  $n=3$ ,  $(X_1, X_2, X_3)$ . Listemos las posibles muestras que recibiremos del fenómeno.

$$\mathfrak{X} := \{(0, 0, 0), (0, 0, 1), (0, 1, 0), (1, 0, 0), (1, 1, 0), (0, 1, 1), (1, 0, 1), (1, 1, 1)\}$$

De estas muestras parece que el caso  $(X_1 = 0, X_2 = 0, X_3 = 0)$  es mas compatible con la hipótesis nula es decir observar esa muestra nos debería llevar a no rechazar  $H_0$ , mientras que el caso  $(X_1 = 1, X_2 = 1, X_3 = 1)$  no es muy compatible con  $H_0$  y por tanto observar esa muestra nos debe de llevar a rechazar nuestra hipótesis nula. Parece entonces que para crear una regla de decisión debemos llevar a cabo una partición de  $\mathfrak{X}$ . Supongamos entonces que participamos  $\mathfrak{X}$  en dos conjuntos denotados por:

$$\mathcal{C}^* = \{\underline{x} \in \mathfrak{X} \mid \text{Rechazo } H_0\} \quad y \quad \mathcal{C} = \{\underline{x} \in \mathfrak{X} \mid \text{No Rechazo } H_0\};$$

(Muchas veces a  $\mathcal{C}^*$  se le conoce como la región crítica)

En nuestro ejemplo supongamos que hacemos la siguiente partición:

$$\mathcal{C} = \{(0, 0, 0), (0, 0, 1), (0, 1, 0), (1, 0, 0)\} \quad y \quad \mathcal{C}^* = \{(1, 1, 0), (1, 0, 1), (1, 1, 1)\};$$

Es decir:

$$\mathcal{C}^* = \left\{ \underline{x} \mid \sum_{i=1}^3 x_i > 1 \right\} \quad y \quad \mathcal{C} = \left\{ \underline{x} \mid \sum_{i=1}^3 x_i \leq 1 \right\}$$

Bajo esta partición tenemos la siguiente regla de decisión:

$$\text{No Rechazo } H_0 \text{ si } \sum_{i=1}^n x_i \leq 1 \text{ y Rechazar } H_0 \text{ si } \sum_{i=1}^n x_i > 1$$

Pero, ¿nuestra elección de  $\mathcal{C}$  es correcta?. Necesitamos desarrollar teoría que nos ayude a determinar si nuestra partición es correcta o no.

Cuando optamos por Rechazar  $H_0$  existe una probabilidad de equivocarse en la decisión tomada o

bien también existe una probabilidad de equivocarse por no rechazar  $H_0$ . No parecería descabellado entonces buscar aquellas regiones  $\mathcal{C}$  y  $\mathcal{C}^*$  que minimicen la probabilidad de equivocarse.

	Realidad	
Decisión	$H_0$ cierta	$H_1$ cierta
No rechazo $H_0$	No hay error	Error tipo II
Rechazo $H_0$	Error tipo I	No hay error

**Definición 4.4.4** (Tipos de Error). *Decimos que en una prueba de hipótesis cometemos el **error tipo I** si rechazamos  $H_0$  cuando en realidad  $H_0$  era cierta y decimos que cometemos el **error tipo II** si no rechazamos  $H_0$  cuando en realidad era falsa.*

Supongamos que ya tenemos definido a  $\mathcal{C}$  y  $\mathcal{C}^*$  e imaginemos que aún no observamos la muestra, como la muestra no se ha observado entonces tiene sentido preguntarnos por las probabilidades del error tipo I y error tipo 2. Definamos  $\alpha$  y  $\beta$  como sigue:

$$\begin{aligned}\alpha &: = \mathbb{P}(\text{Error tipo I}) = \mathbb{P}(\text{Rechazar } H_0 | H_0 \text{ Cierta}) = \mathbb{P}(\underline{X} \in \mathcal{C}^* | H_0 \text{ Cierta}) \\ \beta &: = \mathbb{P}(\text{Error tipo II}) = \mathbb{P}(\text{No Rechazar } H_0 | H_1 \text{ Cierta}) = \mathbb{P}(\underline{X} \in \mathcal{C} | H_1 \text{ Cierta})\end{aligned}$$

Nuestro objetivo es entonces encontrar  $\mathcal{C}$  y  $\mathcal{C}^*$  que minimicen  $\alpha$  y  $\beta$  de manera simultánea.

**Definición 4.4.5** (Tamaño de la prueba). *Sea  $\mathcal{C}^*$  una región de rechazo en una prueba de hipótesis (simple vs simple), decimos que la prueba con  $\mathcal{C}^*$  como su región de rechazo, tiene tamaño  $\alpha$  si*

$$\mathbb{P}(\underline{X} \in \mathcal{C}^* | H_0 \text{ Cierta}) = \alpha$$

En el ejemplo que hemos venido revisando tenemos que:

$$H_0 : \theta = 0.1 \quad \text{vs} \quad H_1 : \theta = 0.8$$

$$\mathcal{C}^* = \left\{ \underline{x} \mid \sum_{i=1}^3 x_i > 1 \right\} \quad \text{y} \quad \mathcal{C} = \left\{ \underline{x} \mid \sum_{i=1}^3 x_i \leq 1 \right\}$$

entonces:

$$\begin{aligned}
 \alpha &= \mathbb{P}(\underline{X} \in \mathcal{C}^* | H_0 \text{ Cierta}) = \mathbb{P}\left(\sum_{i=1}^3 X_i > 1 \mid \theta = 0.1\right) \\
 &= \mathbb{P}\left(\sum_{i=1}^3 X_i > 1 \mid X_i \sim \text{Ber}(0.1)\right) = 0.028 \\
 \beta &= \mathbb{P}(\underline{X} \in \mathcal{C} | H_1 \text{ Cierta}) = \mathbb{P}\left(\sum_{i=1}^3 X_i \leq 1 \mid \theta = 0.8\right) \\
 &= \mathbb{P}\left(\sum_{i=1}^3 X_i \leq 1 \mid X_i \sim \text{Ber}(0.8)\right) = 0.104
 \end{aligned}$$

**Definición 4.4.6** (Potencia de la prueba). *En una prueba de hipótesis simple vs simple, a la cantidad  $1 - \beta$  se le conoce como la potencia de la prueba*

Por lo anterior la potencia de esta prueba (con  $\mathcal{C}$  y  $\mathcal{C}^*$  definidas) está dada por  $1 - \beta = 0.896$ . Ahora supongamos que modificamos la región crítica:

$$\mathcal{C}^* = \left\{ \underline{x} \mid \sum_{i=1}^3 x_i > 2 \right\} \quad y \quad \mathcal{C} = \left\{ \underline{x} \mid \sum_{i=1}^3 x_i \leq 2 \right\}$$

Bajo esta nueva configuración se puede verificar que ahora:

$$\begin{aligned}
 \alpha &= \mathbb{P}\left(\sum_{i=1}^3 X_i > 2 \mid \theta = 0.1\right) = 0.001 \\
 \beta &= \mathbb{P}\left(\sum_{i=1}^3 X_i \leq 2 \mid \theta = 0.8\right) = 0.488
 \end{aligned}$$

Observamos que con esta nueva región crítica hemos logrado reducir  $\alpha$  sin embargo  $\beta$  aumentó considerablemente ocasionado una disminución en la potencia de la prueba.

**En general se puede probar que minimizar  $\alpha$  aumenta la probabilidad del error tipo 2 y viceversa** luego entonces en la práctica se opta entonces por fijar un  $\alpha$  y luego buscar la región crítica  $\mathcal{C}^*$  que minimice  $\beta$  (maximice la potencia) (Obs: **En el contraste siempre se controla el error tipo I fijando un  $\alpha$  previo a observar la muestra** )

**Ejemplo 4.4.2.** *Sea  $X_1$  una m.a. del modelo  $N(\theta, \sigma^2 = 1)$  y suponga que  $\Theta = \{0, 3\}$ . Supongamos que se nos plantea la siguiente prueba simple vs simple:*

$$H_0 : \theta = 0 \quad vs \quad H_1 : \theta = 3$$



Definamos la siguiente región crítica

$$\mathcal{C}^* = \{\underline{x} | x_1 > 2\} \quad y \quad \mathcal{C} = \{\underline{x} | x_1 \leq 2\}$$

Entonces:

$$\begin{aligned}\alpha &= \mathbb{P}(X_1 > 2 | \theta = 0) = \mathbb{P}(X_1 > 2 | X_1 \sim N(0, 1)) = 0.0227 \\ \beta &= \mathbb{P}(X_1 \leq 2 | \theta = 3) = \mathbb{P}(X_1 \leq 2 | X_1 \sim N(3, 1)) = 0.159 \\ 1 - \beta &= 0.841\end{aligned}$$

Ahora definamos otra región crítica:

$$\mathcal{C}^* = \{\underline{x} | 1.5 < x_1 < 1.705\} \quad y \quad \mathcal{C} = \{\underline{x} | x_1 \leq 1.5 \text{ o } x_1 \geq 1.705\}$$

Entonces:

$$\begin{aligned}\alpha &= \mathbb{P}(1.5 < X_1 < 1.705 | \theta = 0) = \mathbb{P}(1.5 < X_1 < 1.705 | X_1 \sim N(0, 1)) = 0.0227 \\ \beta &= \mathbb{P}(X_1 \notin (1.5, 1.705) | \theta = 3) = \mathbb{P}(X_1 \notin (1.5, 1.705) | X_1 \sim N(3, 1)) = 0.969 \\ 1 - \beta &= 0.031\end{aligned}$$

Notemos que ambas pruebas generan una probabilidad del error tipo 1 similar  $\alpha = 0.0227$ , sin embargo la segunda región tiene una potencia mucho menor que la primera, luego entonces entre las dos regiones aquí expuestas preferimos la primera por tener mayor potencia. El objetivo en nuestra teoría es entonces dada  $\alpha$  encontrar la región con mayor potencia.

### Teorema de Neyman Pearson

El Teorema de Neyman Pearson nos dice cómo encontrar la región de rechazo más potente dada un  $\alpha$  fijo y conocido en el caso de una prueba de hipótesis simple vs simple.

**Teorema 4.4.1** (Teorema de Neyman Pearson). Sea  $\underline{X} = (X_1, \dots, X_n)$  m.a. de  $f_X(x; \theta)$  y  $\alpha \in (0, 1)$ . Supongamos que  $\Theta = \{\theta_0, \theta_1\}$  y planteamos la prueba de hipótesis:

$$H_0 : \theta = \theta_0 \quad vs \quad H_1 : \theta = \theta_1$$

Entonces la región de rechazo más potente  $\mathcal{C}^*$  tal que  $\mathbb{P}(\underline{X} \in \mathcal{C}^* | H_0 \text{ Cierta}) = \alpha$  está dada por:

$$\mathcal{C}^* = \left\{ \underline{x} \in \mathfrak{X} \left| \frac{f_X(x_1, \dots, x_n; \theta_0)}{f_X(x_1, \dots, x_n; \theta_1)} \leq k_\alpha \right. \right\}$$

Donde  $k_\alpha$  es número real tal que  $\mathbb{P}(\underline{X} \in \mathcal{C}^* | H_0 \text{ Cierta}) = \alpha$

En *Apéndice* encontrará la prueba detallada. Veamos un ejemplo de cómo utilizar este Teorema:

**Ejemplo 4.4.3.** Sea  $X_1, \dots, X_n$  m.a. de  $\text{Exp}(\lambda)$ , supongamos que  $\Theta = \{1, 5\}$ , y se plantea la hipótesis:

$$H_0 : \lambda = 1 \quad \text{vs} \quad H_1 : \lambda = 5$$

Se nos pide encontrar la mejor región de rechazo tal que  $\alpha = 0.05$ . El teorema de Neyman- Pearson nos indica que:

$$\mathcal{C}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{f_X(x_1, \dots, x_n; \lambda = 1)}{f_X(x_1, \dots, x_n; \lambda = 5)} \leq k_\alpha \right\}$$

Como:

$$f_X(x_1, \dots, x_n; \lambda = 1) = \prod_{i=1}^n e^{-x_i} = e^{-\sum_{i=1}^n x_i}$$

$$f_X(x_1, \dots, x_n; \lambda = 5) = \prod_{i=1}^n 5e^{-5x_i} = 5^n e^{-5\sum_{i=1}^n x_i}$$

Al hacer el cociente obtenemos que:

$$\mathcal{C}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{e^{-\sum_{i=1}^n x_i + 5\sum_{i=1}^n x_i}}{5^n} \leq k_\alpha \right\}$$

Esta región es equivalente a:

$$\mathcal{C}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \sum_{i=1}^n x_i \leq \frac{1}{4} \log(k_\alpha 5^n) \right\}$$

Haciendo  $k_\alpha^* = \frac{1}{4} \log(k_\alpha 5^n)$  obtenemos que la mejor región de rechazo de tamaño  $\alpha$  está dada por:

$$\mathcal{C}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \sum_{i=1}^n x_i \leq k_\alpha^* \right\}$$

Supongamos que fijamos un  $\alpha = 0.05$ , entonces queremos que:

$$\mathbb{P}(\underline{X} \in \mathcal{C}^* \mid H_0 \text{ cierta}) = 0.05 \Rightarrow \mathbb{P}\left(\sum_{i=1}^n X_i \leq k_\alpha^* \mid H_0 \text{ cierta}\right) = 0.05$$

Lo que necesitamos entonces es encontrar  $k_\alpha^*$  tal que se cumpla con lo siguiente:

$$\mathbb{P}\left(\sum_{i=1}^n X_i \leq k_\alpha^* \mid \lambda = 1\right) = 0.05$$

Como **bajo**  $H_0$   $\lambda = 1$  se sigue que  $\sum_{i=1}^n X_i \sim \text{Gamma}(n, \lambda = 1)$ , entonces el problema de

encontrar  $k_\alpha^*$  se reduce a encontrar un el cuantíl de la distribución Gamma de esos parámetros. Sea  $\gamma_\alpha$  el cuantíl  $\alpha$  de la de la densidad Gamma  $(n, 1)$ , entonces sabemos que:

$$\mathbb{P}\left(\sum_{i=1}^n X_i \leq \gamma_\alpha \mid \lambda = 1\right) = 0.05$$

Por lo tanto queda claro que  $k_\alpha^* = \gamma_\alpha$  por lo que finalmente la regla de decisión asociada a esta prueba es rechazar  $H_0$  si

$$\mathcal{C}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \sum_{i=1}^n x_i \leq \gamma_\alpha \right\}$$

Luego entonces el Teorema de Neyman-Pearson asegura que esta es la mejor región de rechazo para la hipótesis planteada de tamaño  $\alpha$ , es decir, es la región que tiene una probabilidad de error tipo I igual a  $\alpha$  con mayor potencia. En este caso para calcular la potencia de la prueba una vez encontrado  $\gamma_\alpha$  se calcula como:

$$1 - \beta = 1 - \mathbb{P}(\underline{X} \in \mathcal{C}^{*c} \mid H_1 \text{ cierta}) = \mathbb{P}(\underline{X} \in \mathcal{C}^* \mid H_1 \text{ cierta}) \quad (4.14)$$

$$= \mathbb{P}\left(\sum_{i=1}^n X_i \leq \gamma_\alpha \mid \lambda = 5\right) \quad (4.15)$$

Debe de observarse que la región de rechazo quedó en función de la estadística  $\sum_{i=1}^n X_i$  a la cual se le conoce como **estadístico de prueba** y se espera tenga una distribución conocida bajo  $H_0$  para poder determinar la región de rechazo de forma exacta.

En una prueba de hipótesis vista desde el punto de Neyman-Pearson, como decisor se fija un  $\alpha$  antes de observar la muestra y con ello se calcula la región con la que se rechazará o no la hipótesis, **Fisher** por otro lado atacó el problema de otra manera, para él no es necesario determinar un  $\alpha$  antes de observar la muestra, para **Fisher** lo más importante es observar primero la muestra y luego evaluar que tan **compatible** es esta muestra con la hipótesis. La forma en que se determina la compatibilidad de la muestra es evaluar que tan probable es observar esa muestra bajo la hipótesis nula basado en el estadístico obtenido por el Teorema de Neyman-Pearson.

**Definición 4.4.7 (P-Value).** . En estadística clásica la forma en cómo medimos la compatibilidad de una muestra con la hipótesis nula es por medio del calculo de la probabilidad de observar la muestra bajo  $H_0$  basados en algún estadístico de prueba, a dicho número lo denominamos **p-value**. Así pues, *p-values* chicos nos hablan de poca compatibilidad de la muestra con la hipótesis nula y por tanto se podría optar por rechazarla, por otro lado, en caso de observar *p-values* altos, esto nos habla de que la muestra goza de alta compatibilidad con nuestra hipótesis nula y por tanto no habría evidencia para rechazarla.

En nuestro ejemplo anterior trabajamos con  $X_1, \dots, X_n$  m.a. de  $Exp(\lambda)$ , y nos planteamos la

hipótesis :

$$H_0 : \lambda = 1 \quad vs \quad H_1 : \lambda = 5$$

El teorema de Neyman-Pearson nos llevó a concluir que la mejor región de rechazo de tamaño  $\alpha$  es:

$$\mathcal{C}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \sum_{i=1}^n x_i \leq \gamma_\alpha \right\}$$

El estadístico de prueba asociado al contraste es  $\sum_{i=1}^n X_i$  el cual bajo  $H_0$  tiene una distribución completamente conocida  $Gamma(n, \lambda = 1)$  Neyman-Person afirma que debemos rechazar  $H_0$  si se observan valores pequeños de este estadístico, Por otro lado, la forma en como lo ve Fisher es primero observar la muestra y luego calcular la probabilidad bajo  $H_0$  de que el estadístico tome un valor igual o más crítico que el de la estimación del estadístico con la muestra observada. Por ejemplo supongamos que observamos la muestra y tenemos que:

$$\sum_{i=1}^n x_i = 5$$

Entonces el **p-value** se podrá calcular como:

$$p - value = \mathbb{P} \left( \sum_{i=1}^n X_i \leq \sum_{i=1}^n x_i = 5 \mid \lambda = 1 \right)$$

Luego entonces debe de quedar claro que un  $p - value$  menor a  $\alpha$  nos hacen rechazar, mientras que valores superiores a  $\alpha$  nos hacen no rechazar la hipótesis nula. Esto nos lleva a la conocida regla del  $p-value$ .

*Si  $p - value \leq \alpha$  entonces **Se rechaza  $H_0$***   
*Si  $p - value > \alpha$  entonces **No se rechaza  $H_0$***

**Ejemplo 4.4.4.** Sea  $X_1, \dots, X_{10}$  m.a del modelo Poisson de parámetro  $\lambda$ . Se plantea la hipótesis:

$$H_0 : \lambda = 1 \quad vs \quad H_1 : \lambda = 3$$

- Suponga que se fija  $\alpha = 0.048$ , encontrar la mejor región de rechazo.
- Para la región de rechazo encontrada calcular su potencia.
- Suponga que observa una muestra tal que  $\sum_{i=1}^{10} x_i = 8$ , calcule el  $p-value$ .
- Concluya con la información anterior si rechaza o no  $H_0$ .

*Solución.*

Con la información proporcionada en el primer inciso sabemos que debemos utilizar el enfoque de Neyman-Pearson, así, haciendo el cociente de verosimilitudes obtenemos que:

$$\mathcal{C}^* = \left\{ \underline{x} \in \mathfrak{X} \left| \frac{f_X(x_1, \dots, x_{10}; \lambda = 1)}{f_X(x_1, \dots, x_{10}; \lambda = 3)} \leq k_\alpha \right. \right\}$$

Como:

$$f_X(x_1, \dots, x_{10}; \lambda = 1) = \prod_{i=1}^{10} \frac{1^{x_i}}{x_i!} e^{-1} = e^{-10} \prod_{i=1}^{10} \frac{1}{x_i!}$$

$$f_X(x_1, \dots, x_{10}; \lambda = 3) = \prod_{i=1}^{10} \frac{3^{x_i}}{x_i!} e^{-3} = e^{-30} \prod_{i=1}^{10} \frac{3^{x_i}}{x_i!}$$

Al hacer el cociente obtenemos que:

$$\mathcal{C}^* = \left\{ \underline{x} \in \mathfrak{X} \left| \frac{e^{-10} \prod_{i=1}^{10} \frac{1}{x_i!}}{e^{-30} \prod_{i=1}^{10} \frac{3^{x_i}}{x_i!}} \leq k_\alpha \right. \right\}$$

Esta región es equivalente a:

$$\mathcal{C}^* = \left\{ \underline{x} \in \mathfrak{X} \left| 3^{-\sum_{i=1}^{10} x_i} \leq k_\alpha e^{20} \right. \right\}$$

$$\mathcal{C}^* = \left\{ \underline{x} \in \mathfrak{X} \left| -\sum_{i=1}^{10} x_i \log(3) \leq \log(k_\alpha e^{20}) \right. \right\}$$

$$\mathcal{C}^* = \left\{ \underline{x} \in \mathfrak{X} \left| \sum_{i=1}^{10} x_i \geq -\frac{\log(k_\alpha e^{20})}{\log(3)} \right. \right\}$$

Finalmente, haciendo  $k_\alpha^* = -\frac{\log(k_\alpha e^{20})}{\log(3)}$  llegamos que la región óptima de rechazo es:

$$\mathcal{C}^* = \left\{ \underline{x} \in \mathfrak{X} \left| \sum_{i=1}^{10} x_i \geq k_\alpha^* \right. \right\}$$

Como  $\alpha = 0.048$ , entonces buscamos  $k_\alpha^*$  tal que se cumpla que:

$$\mathbb{P}\left(\sum_{i=1}^{10} X_i \geq k_\alpha^* \mid \lambda = 1\right) = \alpha = 0.048$$

Como bajo  $H_0$ ,  $\lambda = 1$  se sigue que  $\sum_{i=1}^{10} X_i \sim \text{Poisson}(10)$ . Buscamos entonces un punto en el cual la distribución Poisson(10) deje en su cola derecha 0.048 de probabilidad. Usando el programa

**R** sabemos que si  $Y \sim \text{Poisson}(10)$  entonces:

```
> qpois(1-.048,10)
```

```
[1] 16
```

$$\mathbb{P}(Y \geq 16) = 0.048$$

Por lo que en nuestro caso  $k_\alpha^* = 16$  y por tanto la región más potente de tamaño  $\alpha = 0.048$  está dada por:

$$\mathcal{C}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \sum_{i=1}^{10} x_i \geq 16 \right\}$$

Ahora calculemos la potencia de esta región.

$$\begin{aligned} 1 - \beta &= 1 - \mathbb{P}(\underline{X} \in \mathcal{C}^{*c} \mid \lambda = 3) = \mathbb{P}(\underline{X} \in \mathcal{C}^* \mid \lambda = 3) \\ &= \mathbb{P}\left(\sum_{i=1}^n X_i \geq 16 \mid \lambda = 3\right) = 0.9980525 \end{aligned}$$

Utilizando **R** con el siguiente código:

```
> 1-ppois(15,30)
```

```
[1] 0.9980525
```

Suponiendo que observamos  $\sum_{i=1}^{10} x_i = 8$  entonces calculamos el  $p$ -value

$$p - \text{value} = \mathbb{P}\left(\sum_{i=1}^{10} X_i \geq \sum_{i=1}^{10} x_i = 8 \mid \lambda = 1\right) = 0.7797794$$

Utilizando **R** con el siguiente código:

```
> 1-ppois(7,10)
```

```
[1] 0.7797794
```

Esto nos indica que hay mucha compatibilidad de la muestra observada con la hipótesis nula además como  $\alpha = 0.048 < p - \text{value}$  se concluye que no se debe de rechazar  $H_0$ .

#### 4.4.2. Generalización $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta_1$

El objetivo es ahora poder generar regiones óptimas de rechazo para los distintos casos a los que nos podemos enfrentar al hacer inferencia sobre los parámetros de un modelo. Por ejemplo:

$$H_0 : \theta = \theta_0 \quad \text{vs} \quad H_1 : \theta > \theta_0$$

$$H_0 : \theta = \theta_0 \quad vs \quad H_1 : \theta < \theta_0$$

$$H_0 : \theta = \theta_0 \quad vs \quad H_1 : \theta \neq \theta_0$$

$$H_0 : \theta \leq \theta_0 \quad vs \quad H_1 : \theta > \theta_0$$

$$H_0 : \theta \geq \theta_0 \quad vs \quad H_1 : \theta < \theta_0$$

En general todo lo anterior pueden resumirse al caso:

$$H_0 : \theta \in \Theta_0 \quad vs \quad H_1 : \theta \in \Theta_1$$

Donde  $\Theta = \Theta_0 \cup \Theta_1$  (Una partición del espacio parametral de nuestro problema). En el caso simple vs simple sabemos que  $\Theta_0 = \{\theta_0\}$  y  $\Theta_1 = \{\theta_1\}$  y además basados en el Teorema de Neyman Pearson se conoce que la mejor región de Rechazo para la prueba es de la forma:

$$\mathcal{C}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{f_X(x_1, \dots, x_n; \theta_0)}{f_X(x_1, \dots, x_n; \theta_1)} \leq k_\alpha \right\}$$

Esta región es equivalente a:

$$\mathcal{C}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{f_X(x_1, \dots, x_n; \theta_0)}{\max_{\theta \in \Theta} \{f_X(x_1, \dots, x_n; \theta)\}} \leq k_\alpha^* \right\}$$

La forma en como vamos a generalizar esta idea de Neyman-Person es por medio de la generación de regiones via un cociente de verosimilitudes conocido en inglés como L.R.T (Likelihood Ratio Test), el cual hablaremos a continuación.

### Region L.R.T Likelihood Ratio Test

Comencemos definiendo la región de rechazo via L.R.T.

**Definición 4.4.8** (Región L.R.T (Cociente de verosimilitudes generalizado)). *Suponga que tenemos  $X_1, \dots, X_n$  m.a de  $f_X(x; \theta)$ . Imaginemos que planteamos la prueba de hipótesis*

$$H_0 : \theta \in \Theta_0 \quad vs \quad H_1 : \theta \in \Theta_1$$

Entonces la región de rechazo formada por:

$$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{\sup_{\theta \in \Theta_0} f_X(x_1, \dots, x_n; \theta)}{\sup_{\theta \in \Theta} \{f_X(x_1, \dots, x_n; \theta)\}} \leq k_\alpha^* \right\}$$

Se le denomina la Región L.R.T

Para simplificar la expresión que aparece en la definición muchas veces se renombra a dicho

cociente con la letra  $\lambda$  de tal manera que:

$$\lambda(x_1, \dots, x_n) := \frac{\sup_{\theta \in \Theta_0} f_X(x_1, \dots, x_n; \theta)}{\sup_{\theta \in \Theta} \{f_X(x_1, \dots, x_n; \theta)\}}$$

**Obs:**  $0 \leq \lambda(x_1, \dots, x_n) \leq 1$

Notemos que en el denominador del cociente se esta maximizando a la verosimilitud sobre todo el espacio parametral, si suponemos que tenemos un completo desconocimiento del parametro  $\theta$  entonces dicho supremo sabemos que se obtiene evaluando la verosimilitud en el estimador máximo verosímil, es decir:

$$\sup_{\theta \in \Theta} \{f_X(x_1, \dots, x_n; \theta)\} = f_X(x_1, \dots, x_n; \hat{\theta}_{M.V.})$$

Entonces:

$$\mathcal{C}_{LRT}^* = \left\{ x \in \mathcal{X} \left| \frac{\sup_{\theta \in \Theta_0} f_X(x_1, \dots, x_n; \theta)}{f_X(x_1, \dots, x_n; \hat{\theta}_{M.V.})} \leq k_\alpha^* \right. \right\}$$

La idea de la región LRT en este caso es simple, debemos rechazar  $H_0$  si la estimación máximo verosímil se aleja mucho de la región  $\Theta_0$ , es decir entre la estimación máximo verosímil se aleje mas de la región  $\Theta_0$  mas evidencia hay para rechazar la hipótesis nula  $H_0 : \theta \in \Theta_0$ . Ahora bien, ¿Como obtenemos  $k_\alpha^*$ ?, la idea nuevamente será controlar el error tipo 1 y buscar una constante para  $k_\alpha$  tal que se tenga una probabilidad del error tipo 1 de magnitud  $\alpha$ . El problema al que nos enfrentamos ahora es que bajo  $H_0$  no necesariamente se tiene completamente definida a la función de distribución, es por ello que debemos re-definir lo que entenderemos por **tamaño de la prueba**.

**Definición 4.4.9** (Tamaño de la prueba). Sea  $\underline{X} = (X_1, \dots, X_n)$  m.a. de  $f_X(x; \theta)$  y  $\mathcal{C}^*$  una región de rechazo para la prueba

$$H_0 : \theta \in \Theta_0 \quad vs \quad H_1 : \theta \in \Theta_1$$

Definimos el tamaño de la prueba como:

$$\sup_{\theta \in \Theta_0} \mathbb{P}(\underline{X} \in \mathcal{C}^* | \theta) = \alpha$$

Observe que la la definición anterior hace que la maxima probabilidad de cometer el error tipo 1 sea precisamente  $\alpha$ .

Otra forma en que comúnmente se define el tamaño de la prueba es mediante la denominada **función potencia** de la prueba, la cual definimos a continuación.

**Definición 4.4.10** (Función Potencia). Sea  $\underline{X} = (X_1, \dots, X_n)$  m.a. de  $f_X(x; \theta)$  y  $\mathcal{C}^*$  una región de rechazo para la prueba

$$H_0 : \theta \in \Theta_0 \quad vs \quad H_1 : \theta \in \Theta_1$$



Definimos la función potencia como:

$$\pi_{\mathcal{C}^*}(\theta) = \mathbb{P}(\underline{X} \in \mathcal{C}^* | \theta)$$

Entonces con estas dos definiciones, es claro que el tamaño de una prueba se calcula como:

$$\alpha = \sup_{\theta \in \Theta_0} \pi_{\mathcal{C}^*}(\theta)$$

Una observación importante es que dada  $\mathcal{C}^*$  buscamos una función potencia  $\pi_{\mathcal{C}^*}(\theta)$  que se comporte de forma que:

$$\begin{aligned} \text{si } \theta \in \Theta_0 &\Rightarrow \pi_{\mathcal{C}^*}(\theta) \approx 0 \\ \text{si } \theta \in \Theta_1 &\Rightarrow \pi_{\mathcal{C}^*}(\theta) \approx 1 \end{aligned}$$

Es decir, queremos una región de rechazo que haga que la probabilidad de rechazar  $H_0$  cuando  $\theta \in \Theta_0$  sea pequeña (para no rechazar equivocadamente) y que la probabilidad de rechazar  $H_0$  cuando  $\theta \in \Theta_1$  sea alta (para rechazar acertadamente). Parece entonces que entre dos regiones de rechazo  $\mathcal{C}_1^*$ ,  $\mathcal{C}_2^*$  de tamaño  $\alpha$  escogeremos aquella que genere una mejor función potencia, es decir parecida a la función potencia ideal.

**Ejemplo 4.4.5.** Supongamos que tenemos  $X_1$  m.a. de tamaño 1 del modelo  $N(\mu, \sigma_0^2 = 1)$  y planteamos la prueba de hipótesis:

$$H_0 : \mu \leq 0 \quad \text{vs} \quad H_1 : \mu > 0$$

Definamos la siguientes regiones de rechazo:

$$\begin{aligned} \mathcal{C}_1^* &= \{x \in \mathfrak{X} | x > 2\} \\ \mathcal{C}_2^* &= \{x \in \mathfrak{X} | 1.5 < x < 1.705\} \end{aligned}$$

Entonces si graficamos ambas funciones potencia observaríamos que la mejor región de rechazo está dada por  $\mathcal{C}_1^*$  (Ver figura 4.4.5). Además se puede verificar que ambas regiones tienen el mismo tamaño  $\alpha = 0.0227$

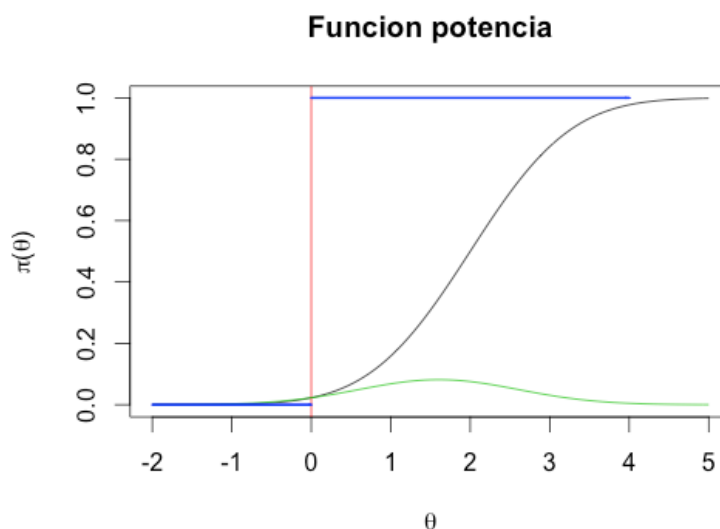


Figura 4.13: Función potencia

En resumen, sabemos que nuestra mejor **apuesta** para encontrar una región óptima de rechazo es llevar a cabo el cociente L.R.T. es decir, definir nuestra la región critica como:

$$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{\sup_{\theta \in \Theta_0} f_X(x_1, \dots, x_n; \theta)}{\sup_{\theta \in \Theta} \{f_X(x_1, \dots, x_n; \theta)\}} \leq k_\alpha^* \right\}$$

Donde  $k_\alpha$  es una constante que asegure que el tamaño de la prueba sea  $\alpha$ , es decir, el objetivo es encontrar  $k_\alpha$  que haga que:

$$\sup_{\theta \in \Theta_0} \mathbb{P}(\underline{X} \in \mathcal{C}^* | \theta) = \sup_{\theta \in \Theta_0} \mathbb{P}\left(\frac{\sup_{\theta \in \Theta_0} f_X(X_1, \dots, X_n; \theta)}{\sup_{\theta \in \Theta} \{f_X(x_1, \dots, x_n; \theta)\}} \leq k_\alpha | \theta\right) = \alpha$$

$$\sup_{\theta \in \Theta_0} \mathbb{P}(\lambda(X_1, \dots, X_n) \leq k_\alpha | \theta) = \alpha$$

Como era de esperarse, los estadísticos de prueba que arroja este tipo de cocientes tienen el inconveniente de que son muy complejos y por tanto llegar a conocer su distribución es muy difícil. Para fines de este curso, veremos algunos ejemplos para distribución normal ya que son las más comunmente usadas y de fácil comprensión, además, posteriormente se enuncian las regiones L.R.T. de esta familia.

**Ejemplo 4.4.6** (L.R.T para  $\mu$ ,  $\sigma_0^2$  conocida). Sea  $X_1, \dots, X_n$  m.a. de una población  $N(\mu, \sigma_0^2)$  con  $\sigma_0^2$  conocida. Encuentre la región L.R.T para la prueba:

$$H_0 : \mu = \mu_0 \quad vs \quad H_1 : \mu \neq \mu_0$$

En este caso  $\Theta = \mathbb{R}$  y por tanto  $\Theta_0 = \{\mu_0\}$  y  $\Theta_1 = \mathbb{R} \setminus \{\mu_0\}$ .

Sabemos que:

$$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \lambda(\underline{x}) = \frac{\sup_{\mu \in \Theta_0} f_X(x_1, \dots, x_n; \mu)}{f_X(x_1, \dots, x_n; \hat{\mu}_{M.V.})} \leq k \right\}$$

En este caso:

$$\sup_{\mu \in \Theta_0} f_X(x_1, \dots, x_n; \mu) = f_X(x_1, \dots, x_n; \mu_0) = (2\pi\sigma_0^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma_0^2} \left(\sum_{i=1}^n (x_i - \mu_0)^2\right)\right)$$

$$f_X(x_1, \dots, x_n; \hat{\mu}_{M.V.}) = (2\pi\sigma_0^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma_0^2} \left(\sum_{i=1}^n (x_i - \bar{x})^2\right)\right)$$

Por lo tanto la región de rechazo queda como:

$$\lambda(\underline{x}) = \frac{\exp\left(-\frac{1}{2\sigma_0^2} \left(\sum_{i=1}^n (x_i - \mu_0)^2\right)\right)}{\exp\left(-\frac{1}{2\sigma_0^2} \left(\sum_{i=1}^n (x_i - \bar{x})^2\right)\right)} \leq k$$

Despejando y tras unos pasos algebraicos

$$\frac{\exp\left(-\frac{1}{2\sigma_0^2} \left(\sum_{i=1}^n (x_i - \mu_0)^2\right)\right)}{\exp\left(-\frac{1}{2\sigma_0^2} \left(\sum_{i=1}^n (x_i - \bar{x})^2\right)\right)} \leq k \Leftrightarrow \exp\left(\frac{1}{2\sigma_0^2} \left(\sum_{i=1}^n (x_i - \bar{x})^2 - \sum_{i=1}^n (x_i - \mu_0)^2\right)\right) \leq k$$

$$\left(\sum_{i=1}^n (x_i - \bar{x})^2 - \sum_{i=1}^n (x_i - \mu_0)^2\right) \leq 2\sigma_0^2 \log(k)$$

$$\sum_{i=1}^n x_i^2 - n\bar{x}^2 - \sum_{i=1}^n x_i^2 + 2\mu_0 \sum_{i=1}^n x_i - n\mu_0^2 \leq 2\sigma_0^2 \log(k)$$

$$-n\bar{x}^2 + 2\mu_0 \sum_{i=1}^n x_i \leq 2\sigma_0^2 \log(k) + n\mu_0^2 \Leftrightarrow -n\bar{x}^2 + 2n\mu_0\bar{x} \leq 2\sigma_0^2 \log(k) + n\mu_0^2$$

$$\bar{x}^2 - 2\mu_0\bar{x} \geq -\frac{1}{n} (2\sigma_0^2 \log(k) + n\mu_0^2) \Leftrightarrow \bar{x}^2 + 2\mu_0\bar{x} + \mu_0^2 \geq -\frac{1}{n} (2\sigma_0^2 \log(k) + n\mu_0^2) + \mu_0^2$$

$$(\bar{x} - \mu_0)^2 \geq -\frac{1}{n} (2\sigma_0^2 \log(k)) \Leftrightarrow |\bar{x} - \mu_0| \geq k^*$$

Finalmente la región L.T.R. queda de la siguiente forma:

$$\mathcal{C}_{LRT}^* = \{\underline{x} \in \mathfrak{X} \mid |\bar{x} - \mu_0| \geq k^*\}$$

Observe como la región de rechazo llega a algo intuitivamente obvio, debemos rechazar  $H_0$  cuando la estimación máximo verosímil se aleje demasiado del valor  $\mu_0$  (El valor bajo  $H_0$ ). El problema es ahora encontrar  $k^*$  tal que el error tipo 1 sea igual a  $\alpha$  sin embargo para encontrar dicho valor necesitamos conocer exactamente la distribución de  $|\bar{X} - \mu_0|$  bajo  $H_0$ . Afortunadamente para el

caso Normal sabemos que si  $H_0$  es cierta se tiene que :

$$\bar{X} \sim N\left(\mu_0, \frac{\sigma_0^2}{n}\right) \Rightarrow \frac{\bar{X} - \mu_0}{\sqrt{\frac{\sigma_0^2}{n}}} \sim N(0, 1)$$

Entonces podemos sacar provecho de esta característica distribucional redefiniendo nuestra región de rechazo como:

$$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \left| \left| \sqrt{n} \left( \frac{\bar{x} - \mu_0}{\sqrt{\sigma_0^2}} \right) \right| \geq \sqrt{n} \frac{k^*}{\sqrt{\sigma_0^2}} = k^{**} \right. \right\}$$

Bajo esta nueva definición tenemos que encontrar el valor de  $k^{**}$  se simplifica pues:

$$\begin{aligned} \alpha &= \mathbb{P}(\underline{X} \in \mathcal{C}_{LRT}^* | H_0 \text{ cierta}) = \mathbb{P}\left(\left| \sqrt{n} \left( \frac{\bar{X} - \mu_0}{\sqrt{\sigma_0^2}} \right) \right| \geq k^{**} \mid \mu = \mu_0\right) \\ &= 1 - \mathbb{P}\left(-k^{**} \leq \sqrt{n} \left( \frac{\bar{X} - \mu_0}{\sqrt{\sigma_0^2}} \right) \leq k^{**} \mid \mu = \mu_0\right) \end{aligned}$$

Entonces:

$$\mathbb{P}\left(-k^{**} \leq \sqrt{n} \left( \frac{\bar{X} - \mu_0}{\sqrt{\sigma_0^2}} \right) \leq k^{**} \mid \mu = \mu_0\right) = 1 - \alpha$$

De donde concluimos que  $k^{**} = Z_{(1-\alpha/2)}$ , donde  $Z_{(1-\alpha/2)}$  es el cuantil  $1 - \alpha/2$  de la distribución  $N(0, 1)$ . Por lo tanto concluimos que la región de rechazo L.R.T. de tamaño  $\alpha$  está dada por:

$$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \left| \left| \sqrt{n} \left( \frac{\bar{x} - \mu_0}{\sqrt{\sigma_0^2}} \right) \right| \geq Z_{(1-\alpha/2)} \right. \right\}$$

Notemos entonces que nuestro estadístico de prueba en este caso es:

$$T = \left| \sqrt{n} \left( \frac{\bar{X} - \mu_0}{\sqrt{\sigma_0^2}} \right) \right|$$

Ahora imaginemos que  $\sigma_0^2 = 1$  y que estamos probando

$$H_0 : \mu = 1 \quad \text{vs} \quad H_1 : \mu \neq 1$$

y que recibimos una muestra  $x_1, \dots, x_{10}$  tal que  $\sum_{i=1}^{10} x_i = 20$ , entonces, el valor del estadístico evaluado en la muestra es igual a:

$$t = \left| \sqrt{10} (2 - 1) \right| = 3.1622$$

Suponiendo  $\alpha = 0.05$ , sabemos que la región de rechazo queda de la forma:

$$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \left| \left| \sqrt{n} \left( \frac{\bar{x} - 1}{\sqrt{1}} \right) \right| \geq 1.96 \right. \right\}$$

como  $3.1622 > 1.96$  entonces concluimos que podemos rechazar  $H_0$ , tenemos suficiente evidencia estadística como para rechazar  $H_0$  y la probabilidad de equivocarse esta controlada y es menor a 0.05. Ahora, desde el punto de vista de Fisher, el  $p$ -value se encontraría de la siguiente manera:

$$\mathbb{P} \left( \left| \sqrt{n} \left( \frac{\bar{X} - 1}{\sqrt{1}} \right) \right| \geq 3.1622 \mid \mu = 1 \right) = 0.0007829099$$

De esta manera dado que  $\alpha = 0.05 > 0.0007829099 = p$ -value, se rechaza  $H_0$  pues desde el punto de vista de Fisher, la muestra recibida es incompatible con la hipótesis nula pues la probabilidad de observar una muestra así bajo el supuesto de que  $\mu = 1$  es muy baja ( $p$ -value = 0.0007829099)

**Ejemplo 4.4.7** (L.R.T para  $\sigma^2$ ,  $\mu$  conocida). Suponga que ahora en el caso normal nos interesa a hacer inferencia sobre la varianza  $\sigma^2$  y se nos plantea la hipótesis.

$$H_0 : \sigma^2 = \sigma_0^2 \quad vs \quad H_1 : \sigma^2 \neq \sigma_0^2$$

Supondremos además que  $\mu = \mu_0$  conocida. Se nos pide encontrar la región L.R.T. para esta prueba.

En este caso  $\Theta = \mathbb{R}^+$  y por tanto  $\Theta_0 = \{\sigma_0^2\}$  y  $\Theta_1 = \mathbb{R}^+ \setminus \{\sigma_0^2\}$ .

Sabemos que:

$$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \left| \lambda(\underline{x}) = \frac{\sup_{\mu \in \Theta_0} f_X(x_1, \dots, x_n; \sigma^2)}{f_X(x_1, \dots, x_n; \hat{\sigma}_{M.V.}^2)} \leq k \right. \right\}$$

En este caso:

$$\sup_{\sigma^2 \in \Theta_0} f_X(x_1, \dots, x_n; \sigma^2) = f_X(x_1, \dots, x_n; \sigma_0^2) = (2\pi\sigma_0^2)^{-\frac{n}{2}} \exp \left( -\frac{1}{2\sigma_0^2} \left( \sum_{i=1}^n (x_i - \mu_0)^2 \right) \right)$$

$$f_X(x_1, \dots, x_n; \hat{\sigma}_{M.V.}^2) = (2\pi\hat{\sigma}_{M.V.}^2)^{-\frac{n}{2}} \exp \left( -\frac{1}{2\hat{\sigma}_{M.V.}^2} \left( \sum_{i=1}^n (x_i - \mu_0)^2 \right) \right)$$

Por lo tanto la región de rechazo queda como:

$$\lambda(\underline{x}) = \frac{(2\pi\sigma_0^2)^{-\frac{n}{2}} \exp \left( -\frac{1}{2\sigma_0^2} \left( \sum_{i=1}^n (x_i - \mu_0)^2 \right) \right)}{(2\pi\hat{\sigma}_{M.V.}^2)^{-\frac{n}{2}} \exp \left( -\frac{1}{2\hat{\sigma}_{M.V.}^2} \left( \sum_{i=1}^n (x_i - \mu_0)^2 \right) \right)} \leq k$$

Recordemos ahora que en este contexto:

$$\hat{\sigma}_{M.V.}^2 := \frac{1}{n} \sum_{i=1}^n (x_i - \mu_0)^2$$

Por lo tanto:

$$-\frac{1}{2\hat{\sigma}_{M.V.}^2} \left( \sum_{i=1}^n (x_i - \mu_0)^2 \right) = -\frac{n}{2} \quad y \quad -\frac{1}{2\sigma_0^2} \left( \sum_{i=1}^n (x_i - \mu_0)^2 \right) = -\frac{n}{2} \frac{\hat{\sigma}_{M.V.}^2}{\sigma_0^2}$$

Entonces:

$$\lambda(\underline{x}) = \frac{(2\pi\sigma_0^2)^{-\frac{n}{2}} \exp\left(-\frac{n}{2} \frac{\hat{\sigma}_{M.V.}^2}{\sigma_0^2}\right)}{(2\pi\hat{\sigma}_{M.V.}^2)^{-\frac{n}{2}} \exp\left(-\frac{n}{2}\right)} \leq k$$

Lo cual es equivalente a:

$$\left(\frac{\hat{\sigma}_{M.V.}^2}{\sigma_0^2}\right)^{\frac{n}{2}} \exp\left(-\frac{\hat{\sigma}_{M.V.}^2}{\sigma_0^2}\right)^{\frac{n}{2}} \leq \exp\left(-\frac{n}{2}\right) k$$

Elevando a la  $-\frac{n}{2}$  en ambos lados:

$$\left(\frac{\hat{\sigma}_{M.V.}^2}{\sigma_0^2}\right) \exp\left(-\frac{\hat{\sigma}_{M.V.}^2}{\sigma_0^2}\right) \leq \left(\exp\left(-\frac{n}{2}\right) k\right)^{-\frac{n}{2}} = k^*$$

Por lo tanto la la región de rechazo queda como:

$$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \left(\frac{\hat{\sigma}_{M.V.}^2}{\sigma_0^2}\right) \exp\left(-\frac{\hat{\sigma}_{M.V.}^2}{\sigma_0^2}\right) \leq k^* \right\}$$

Ahora notemos que para que  $\left(\frac{\hat{\sigma}_{M.V.}^2}{\sigma_0^2}\right) \exp\left(-\frac{\hat{\sigma}_{M.V.}^2}{\sigma_0^2}\right)$  sea pequeño, requerimos que:

$$\frac{\hat{\sigma}_{M.V.}^2}{\sigma_0^2} < k_1 \quad \text{??????} \quad \frac{\hat{\sigma}_{M.V.}^2}{\sigma_0^2} > k_2$$

Lo anterior se concluye analizando a la función  $f(x) = xe^{-x}$  la cual toma valores cercanos a cero cuando  $x \rightarrow 0$  o  $x \rightarrow \infty$ . Concluimos que debemos rechazar  $H_0$  si :

$$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{\hat{\sigma}_{M.V.}^2}{\sigma_0^2} < k_1 \quad o \quad \frac{\hat{\sigma}_{M.V.}^2}{\sigma_0^2} > k_2 \right\}$$

Notemos además que bajo  $H_0$  tenemos la siguiente propiedad distribucional:

$$\frac{\sum_{i=1}^n (x_i - \mu)^2}{\sigma_0^2} \sim \chi_{(n)}^2 \Rightarrow \frac{n\hat{\sigma}_{M.V.}^2}{\sigma_0^2} \sim \chi_{(n)}^2$$

Por lo que nos conviene redefinir nuestra región de rechazo como:

$$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \left| \frac{n\hat{\sigma}_{M.V.}^2}{\sigma_0^2} < k_1 \quad \text{o} \quad \frac{n\hat{\sigma}_{M.V.}^2}{\sigma_0^2} > k_2 \right. \right\}$$

Luego entonces, sacando uso de la propiedad distribucional de  $\frac{n\hat{\sigma}_{M.V.}^2}{\sigma_0^2}$  la región de rechazo de tamaño  $\alpha$  que obtenemos es :

$$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \left| \frac{n\hat{\sigma}_{M.V.}^2}{\sigma_0^2} < \chi_{(n)}^{2(\alpha/2)} \quad \text{o} \quad \frac{n\hat{\sigma}_{M.V.}^2}{\sigma_0^2} > \chi_{(n)}^{2(1-\alpha/2)} \right. \right\}$$

Donde  $\chi_{(n)}^{2(1-\alpha/2)}$  y  $\chi_{(n)}^{2(\alpha/2)}$  son los cuantiles  $1 - \alpha/2$  y  $\alpha/2$  respectivamente de una distribución  $\chi_{(n)}^2$

### 4.4.3. Regiones Críticas para distribuciones Normales (L.R.T.)

A continuación escribimos las regiones críticas obtenidas por el procedimiento del cociente de verosimilitudes (L.R.T.)

#### Contrastes para $\mu$ , suponiendo $\sigma^2$ conocida

**Entrada:**  $X_1, \dots, X_n$  m.a. de un modelo  $N(\mu, \sigma^2)$

**Estadístico de Prueba:**

$$T = \frac{\sqrt{n}(\bar{X} - \mu_0)}{\sigma} \sim N(0, 1)$$

$H_0$	$H_1$	$\mathcal{C}_{LRT}^*$ de tamaño $\alpha$
$\mu = \mu_0$	$\mu > \mu_0$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \left  \frac{\sqrt{n}(\bar{x} - \mu_0)}{\sigma} > Z_{(1-\alpha)} \right. \right\}$
$\mu = \mu_0$	$\mu < \mu_0$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \left  \frac{\sqrt{n}(\bar{x} - \mu_0)}{\sigma} < Z_{(\alpha)} \right. \right\}$
$\mu = \mu_0$	$\mu \neq \mu_0$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \left  \frac{\sqrt{n}(\bar{x} - \mu_0)}{\sigma} > Z_{(1-\frac{\alpha}{2})} \right. \right\}$
$\mu \leq \mu_0$	$\mu > \mu_0$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \left  \frac{\sqrt{n}(\bar{x} - \mu_0)}{\sigma} > Z_{(1-\alpha)} \right. \right\}$
$\mu \geq \mu_0$	$\mu < \mu_0$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \left  \frac{\sqrt{n}(\bar{x} - \mu_0)}{\sigma} < Z_{(\alpha)} \right. \right\}$

**Ejemplo 4.4.8.** Una fábrica de llantas tiene dos turnos de operarios, turno de día y turno mixto. Se selecciona una muestra aleatoria de 100 llantas producidas obteniéndose los siguientes datos por cada turno.

Para el turno de día tenemos que la media es 25006.04 y para el turno mixto la media fue de 24941.12 la desviación estándar en ambos casos es 4000

Se sabe que la producción de llantas sigue una distribución normal pero no se conocen y se le ha pedido al equipo de análisis de calidad ayudar al gerente a sacar conclusiones de cada una de las siguientes preguntas con una significancia del 10 %.

1. ¿ Es la duración promedio de las llantas producidas en el turno de día igual a 25 000 millas?  
En este caso se plantea la siguiente prueba de hipótesis

$$H_0 : \mu = 25000 \quad vs \quad H_1 : \mu \neq 25000$$

Sabemos que:

$$\bar{X}_{\text{día}} = 25006.04 \quad y \quad \sigma_{\text{día}} = 4000$$

Por lo que el estadístico de prueba queda como:

$$|T| = \left| \frac{\sqrt{100} (25006.04 - 25000)}{4000} \right| = 0.0151$$

Tenemos que se rechaza  $H_0$  si  $|T| > Z_{1-\frac{0.1}{2}} = 1.64$  y como  $0.0151 < 1.64$  entonces no se rechaza  $H_0$ , por lo que se puede decir que la duración promedio de las llantas producidas en el turno de día es igual a 25 000 millas.

2. ¿ Es la duración promedio de las llantas producidas en el turno mixto a lo más de 25 000 millas?

En este caso se plantea la siguiente prueba de hipótesis

$$H_0 : \mu \leq 25000 \quad vs \quad H_1 : \mu > 25000$$

Sabemos que:

$$\bar{X}_{\text{mixto}} = 24941.12 \quad y \quad \sigma_{\text{mixto}} = 4000$$

Por lo que el estadístico de prueba queda como:

$$T = \frac{\sqrt{100} (24941.12 - 25000)}{4000} = -0.1472$$

Tenemos que se rechaza  $H_0$  si  $T > Z_{1-0.1} = 1.281552$  y como  $-0.1472 < 1.281552$  entonces no se rechaza  $H_0$ , por lo que se puede decir que no existe suficiente evidencia estadística para rechazar que la duración promedio de las llantas producidas en el turno mixto es a lo más 25 000 millas.

#### 4.4.4. Contrastes para $\mu$ , suponiendo $\sigma^2$ desconocida

**Entrada:**  $X_1, \dots, X_n$  m.a. de un modelo  $N(\mu, \sigma^2)$

**Estadístico de Prueba:**

$$T = \frac{\sqrt{n} (\bar{X} - \mu_0)}{\hat{\sigma}} \sim t_{(n-1)}; \quad \text{con} \quad \hat{\sigma} = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}}$$



$H_0$	$H_1$	$\mathcal{C}_{LRT}^*$ de tamaño $\alpha$
$\mu = \mu_0$	$\mu > \mu_0$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{\sqrt{n}(\bar{x} - \mu_0)}{\hat{\sigma}} > t_{(n-1)}^{(1-\alpha)} \right\}$
$\mu = \mu_0$	$\mu < \mu_0$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{\sqrt{n}(\bar{x} - \mu_0)}{\hat{\sigma}} < t_{(n-1)}^{(\alpha)} \right\}$
$\mu = \mu_0$	$\mu \neq \mu_0$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{\sqrt{n}(\bar{x} - \mu_0)}{\hat{\sigma}} > t_{(n-1)}^{(1-\frac{\alpha}{2})} \right\}$
$\mu \leq \mu_0$	$\mu > \mu_0$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{\sqrt{n}(\bar{x} - \mu_0)}{\hat{\sigma}} > t_{(n-1)}^{(1-\alpha)} \right\}$
$\mu \geq \mu_0$	$\mu < \mu_0$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{\sqrt{n}(\bar{x} - \mu_0)}{\hat{\sigma}} < t_{(n-1)}^{(\alpha)} \right\}$

**Ejemplo 4.4.9.** Conocemos que las alturas  $X$  de los individuos de una ciudad, se distribuyen de modo gaussiano. Deseamos contrastar con un nivel de significación de  $\alpha = 5\%$  si la altura media es diferente de 174 cm. Para ello nos basamos en un estudio en el que con una muestra de  $n=25$  personas se obtuvo: media 170 y desviación 10.206 ambas estimadas.

Tenemos que en este caso la prueba de hipótesis está dada por:

$$H_0 : \mu = 174 \quad vs \quad H_1 : \mu \neq 174$$

Sabemos que:

$$\bar{X} = 170 \quad y \quad \hat{\sigma} = 10.206$$

Por lo que el estadístico de prueba queda como:

$$|T| = \left| \frac{\sqrt{25} (174 - 170)}{10.206} \right| = 1.959632$$

Tenemos que se rechaza  $H_0$  si  $|T| > t_{(25-1)}^{1-\frac{0.05}{2}} = 2.063899$  y como  $1.959632 < 2.063899$  entonces no se rechaza  $H_0$ , por lo que se puede decir que no existe suficiente evidencia estadística para rechazar que la la altura media es de 174cm.

### Contrastes para $\sigma^2$ , suponiendo $\mu$ conocida

**Entrada:**  $X_1, \dots, X_n$  m.a. de un modelo  $N(\mu, \sigma^2)$

**Estadístico de Prueba:**

$$T = \frac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma_0^2} \sim \chi_{(n)}^2$$

$H_0$	$H_1$	$\mathcal{C}_{LRT}^*$ de tamaño $\alpha$
$\sigma^2 = \sigma_0^2$	$\sigma^2 > \sigma_0^2$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{\sum_{i=1}^n (x_i - \mu)^2}{\sigma_0^2} > \chi_{(n)}^{2(1-\alpha)} \right\}$
$\sigma^2 = \sigma_0^2$	$\sigma^2 < \sigma_0^2$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{\sum_{i=1}^n (x_i - \mu)^2}{\sigma_0^2} < \chi_{(n)}^{2(\alpha)} \right\}$
$\sigma^2 = \sigma_0^2$	$\sigma^2 \neq \sigma_0^2$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{\sum_{i=1}^n (x_i - \mu)^2}{\sigma_0^2} < \chi_{(n)}^{2(\frac{\alpha}{2})} \quad o \quad \frac{\sum_{i=1}^n (x_i - \mu)^2}{\sigma_0^2} > \chi_{(n)}^{2(1-\frac{\alpha}{2})} \right\}$
$\sigma^2 \leq \sigma_0^2$	$\sigma^2 > \sigma_0^2$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{\sum_{i=1}^n (x_i - \mu)^2}{\sigma_0^2} > \chi_{(n)}^{2(1-\alpha)} \right\}$
$\sigma^2 \geq \sigma_0^2$	$\sigma^2 < \sigma_0^2$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{\sum_{i=1}^n (x_i - \mu)^2}{\sigma_0^2} < \chi_{(n)}^{2(\alpha)} \right\}$

**Ejemplo 4.4.10.** Sospechamos que nuestro cromatógrafo está estropeado, para comprobarlo queremos determinar si los resultados que nos proporciona son lo suficientemente precisos. Para ello, realizamos una serie de 8 mediciones del contenido de una solución de referencia que, sabemos, contiene 90mg de un determinado compuesto. Los resultados que obtenemos son:

$$93.3, 86.8, 90.4, 90.1, 94.9, 91.6, 92.3, 96.5$$

Se sabe que los datos siguen una distribución normal, y se quiere saber si la varianza es de al menos 5.44 a una significancia del 5%.

En este caso proponemos la siguiente prueba de hipótesis

$$H_0 : \sigma^2 \geq 5.44 \quad vs \quad H_1 : \sigma^2 < 5.44$$

Además sabemos que  $\bar{X} = 90$  por lo que el estadístico de prueba queda como:

$$T = \frac{\sum_{i=1}^8 (x_i - 90)^2}{5.44} = \frac{(93.3 - 90)^2 + \dots + (96.5 - 90)^2}{5.44} = 17.5386$$

Tenemos que se rechaza  $H_0$  si  $T < \chi_{(8)}^{2(0.05)} = 2.732637$  y como  $2.732637 < 17.5386$  entonces no se rechaza  $H_0$ , por lo que se puede decir que no existe suficiente evidencia estadística para rechazar que la varianza es de al menos 5.44

### Contrastes para $\sigma^2$ , suponiendo $\mu$ desconocida

**Entrada:**  $X_1, \dots, X_n$  m.a. de un modelo  $N(\mu, \sigma^2)$

**Estadístico de Prueba:**

$$T = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma_0^2} \sim \chi_{(n-1)}^2$$

$H_0$	$H_1$	$\mathcal{C}_{LRT}^*$ de tamaño $\alpha$
$\sigma^2 = \sigma_0^2$	$\sigma^2 > \sigma_0^2$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma_0^2} > \chi_{(n-1)}^{2(1-\alpha)} \right\}$
$\sigma^2 = \sigma_0^2$	$\sigma^2 < \sigma_0^2$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma_0^2} < \chi_{(n-1)}^{2(\alpha)} \right\}$
$\sigma^2 = \sigma_0^2$	$\sigma^2 \neq \sigma_0^2$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma_0^2} < \chi_{(n-1)}^{2(\frac{\alpha}{2})} \text{ o } \frac{\sum_{i=1}^n (x_i - \mu)^2}{\sigma_0^2} > \chi_{(n-1)}^{2(1-\frac{\alpha}{2})} \right\}$
$\sigma^2 \leq \sigma_0^2$	$\sigma^2 > \sigma_0^2$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma_0^2} > \chi_{(n-1)}^{2(1-\alpha)} \right\}$
$\sigma^2 \geq \sigma_0^2$	$\sigma^2 < \sigma_0^2$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma_0^2} < \chi_{(n-1)}^{2(\alpha)} \right\}$

**Ejemplo 4.4.11.** Una empresa dedicada a proveer transporte foráneo de pasajeros para ejecutivos desea dar una imagen de confiabilidad haciendo que sus conductores sean puntuales en sus tiempos de llegada a la terminal. La empresa tiene por objetivo que haya poca variabilidad en dichos tiempos. En términos de varianza de los tiempos de llegada la empresa ha implementado cierta estrategia con lo cual se busca que la varianza sea de 4 minutos o menos para esto se ha tomado una muestra aleatoria de 24 llegadas teniendo que la varianza muestral es de 4.9 y que el tiempo de llegadas sigue una distribución normal. Con una significancia del 5 % determine si se ha cumplido el objetivo de la empresa.

Tenemos que dada la problemática descrita estamos en el caso donde buscamos hacer inferencia sobre la varianza y además ambos parámetros son desconocidos por lo que la prueba de hipótesis queda como:

$$H_0 : \sigma^2 \leq 4 \quad \text{vs} \quad H_1 : \sigma^2 > 4$$

Además sabemos que  $\hat{\sigma}^2 = 4.9$  por lo que el estadístico de prueba queda como:

$$T = \frac{\sum_{i=1}^{24} (x_i - \bar{x})^2}{4} = \frac{(24-1)\hat{\sigma}^2}{4} = \frac{23 * 4.9}{4} = 28.175$$

Tenemos que se rechaza  $H_0$  si  $T > \chi_{(24-1)}^{2(1-0.05)} = 35.17246$  y como  $28.175 < 35.17246$  entonces no se rechaza  $H_0$ , por lo que se puede decir que no existe suficiente evidencia estadística para rechazar que la varianza es de a lo más 4 por lo que la empresa ha cumplido con su objetivo.

Es importante hacer mención que para las siguientes pruebas de hipótesis se debe considerar como supuesto general que las muestras son independientes entre sí.

#### 4.4.5. Contrastes para la diferencia de medias $\mu_x - \mu_y$ suponiendo $\sigma_x^2$ , $\sigma_y^2$ conocidas

**Entrada:**  $X_1, \dots, X_n$  m.a. de  $N(\mu_x, \sigma_x^2)$  y  $Y_1, \dots, Y_n$  m.a. de  $N(\mu_y, \sigma_y^2)$ ;

**Estadístico de Prueba:**

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{\left(\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{n}\right)}} \sim N(0, 1)$$

$H_0$	$H_1$	$\mathcal{C}_{LRT}^*$ de tamaño $\alpha$
$(\mu_x - \mu_y) = d_0$	$(\mu_x - \mu_y) > d_0$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{(\bar{x} - \bar{y}) - d_0}{\sqrt{\frac{1}{n}(\sigma_x^2 + \sigma_y^2)}} > Z_{(1-\alpha)} \right\}$
$(\mu_x - \mu_y) = d_0$	$(\mu_x - \mu_y) < d_0$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{(\bar{x} - \bar{y}) - d_0}{\sqrt{\frac{1}{n}(\sigma_x^2 + \sigma_y^2)}} < Z_{(\alpha)} \right\}$
$(\mu_x - \mu_y) = d_0$	$(\mu_x - \mu_y) \neq d_0$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \left  \frac{(\bar{x} - \bar{y}) - d_0}{\sqrt{\frac{1}{n}(\sigma_x^2 + \sigma_y^2)}} \right  > Z_{(1-\frac{\alpha}{2})} \right\}$
$(\mu_x - \mu_y) \leq d_0$	$(\mu_x - \mu_y) > d_0$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{(\bar{x} - \bar{y}) - d_0}{\sqrt{\frac{1}{n}(\sigma_x^2 + \sigma_y^2)}} > Z_{(1-\alpha)} \right\}$
$(\mu_x - \mu_y) \geq d_0$	$(\mu_x - \mu_y) < d_0$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{(\bar{x} - \bar{y}) - d_0}{\sqrt{\frac{1}{n}(\sigma_x^2 + \sigma_y^2)}} < Z_{(\alpha)} \right\}$

**Ejemplo 4.4.12.** *Un microbiólogo desea averiguar si hay diferencia en el tiempo que tarda en producir yogur utilizando dos tipos de bacterias: lactobacillus acidophilus (A) y bulgaricus (B). Se prepararon siete remesas de yogur con cada tipo de lactobacilo. A continuación se muestra el tiempo (en horas) hasta que se produjo cada remesa:*

Cultivo A	6.8	6.3	7.4	6.1	8.2	7.3	6.9
Cultivo B	6.1	6.4	5.7	5.5	6.9	6.3	6.7

Suponiendo que la distribución de ambos conjuntos de observaciones se puede considerar normal y que  $\sigma_{\text{cultivoA}}^2 = 0.5$  y que  $\sigma_{\text{cultivoB}}^2 = 0.25$ , se desea saber si el tiempo que se tarda en producir yogur con el cultivo A es menor o igual que el del cultivo B. Considere  $\alpha = 5\%$

Si consideramos a la variable  $X$  como el tiempo que tarda en producir yogur con el cultivo A y a la variable  $Y$  como el tiempo que tarda en producir yogur con el cultivo B y además tenemos conocimiento del valor de las varianzas para cada población. Por otro lado, solo se nos pide saber si el tiempo que se tarda en producir yogur con el cultivo A es menor o igual que el del cultivo B por lo que  $d_0 = 0$  entonces tenemos que la prueba de hipótesis queda como:

$$H_0 : (\mu_x - \mu_y) \leq 0 \quad \text{vs} \quad H_1 : (\mu_x - \mu_y) > 0$$

Por otro lado, haciendo el calculo de:

$$\bar{X} = 7 \quad \bar{Y} = 6.228571 \quad \sigma_X^2 = 0.5 \quad \sigma_Y^2 = 0.25$$

Con estadístico de prueba dado por:

$$T = \frac{(7 - 6.228571) - 0}{\sqrt{\left(\frac{0.5}{7} + \frac{0.25}{7}\right)}} = 2.356753$$

Tenemos que se rechaza  $H_0$  si  $T > Z^{1-0.05} = 1.64$  y como  $1.64 < 2.356753$  entonces se rechaza  $H_0$ , por lo que se puede decir que el tiempo que se tarda en producir yogur con el cultivo A es mayor que el que se tarda produciendo con el cultivo B.

**Contrastes para la diferencia de medias  $\mu_x - \mu_y$  suponiendo  $\sigma_x^2, \sigma_y^2$  desconocidas pero iguales  $\sigma_x^2 = \sigma_y^2$**

**Entrada:**  $X_1, \dots, X_{n_1}$  m.a. de  $N(\mu_x, \sigma^2)$  y  $Y_1, \dots, Y_{n_2}$  m.a. de  $N(\mu_y, \sigma^2)$ ;

**Estadístico de Prueba:**

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{\hat{\sigma}_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim t_{(n_1+n_2-2)}$$

Donde  $\hat{\sigma}_p^2$  está dada por:

$$\hat{\sigma}_p^2 := \frac{\sum_{i=1}^{n_1} (X_i - \bar{X})^2 + \sum_{i=1}^{n_2} (Y_i - \bar{Y})^2}{n_1 + n_2 - 2} = \frac{(n_1 - 1) \hat{\sigma}_x^2 + (n_2 - 1) \hat{\sigma}_y^2}{n_1 + n_2 - 2}$$

$H_0$	$H_1$	$\mathcal{C}_{LRT}^*$ de tamaño $\alpha$
$(\mu_x - \mu_y) = d_0$	$(\mu_x - \mu_y) > d_0$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{(\bar{x} - \bar{y}) - d_0}{\sqrt{\hat{\sigma}_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} > t_{(n_1+n_2-2)}^{(1-\alpha)} \right\}$
$(\mu_x - \mu_y) = d_0$	$(\mu_x - \mu_y) < d_0$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{(\bar{x} - \bar{y}) - d_0}{\sqrt{\hat{\sigma}_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} < t_{(n_1+n_2-2)}^{(\alpha)} \right\}$
$(\mu_x - \mu_y) = d_0$	$(\mu_x - \mu_y) \neq d_0$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{(\bar{x} - \bar{y}) - d_0}{\sqrt{\hat{\sigma}_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} > t_{(n_1+n_2-2)}^{(1-\frac{\alpha}{2})} \right\}$
$(\mu_x - \mu_y) \leq d_0$	$(\mu_x - \mu_y) > d_0$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{(\bar{x} - \bar{y}) - d_0}{\sqrt{\hat{\sigma}_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} > t_{(n_1+n_2-2)}^{(1-\alpha)} \right\}$
$(\mu_x - \mu_y) \geq d_0$	$(\mu_x - \mu_y) < d_0$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{(\bar{x} - \bar{y}) - d_0}{\sqrt{\hat{\sigma}_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} < t_{(n_1+n_2-2)}^{(\alpha)} \right\}$

**Ejemplo 4.4.13.** Para encontrar si un nuevo suero detiene la leucemia, se seleccionan nueve ratones, todos con una etapa avanzada de la enfermedad. Cinco ratones reciben el tratamiento

y cuatro no. Los tiempos de sobrevivencia en años, a partir del momento en que comienza el experimento son los siguientes:

Con tratamiento	2.1	5.3	1.4	4.6	0.9
Sin tratamiento	1.9	0.5	2.8	3.1	

¿ Se puede decir en el nivel de significancia del 0.05 que el suero es efectivo? Suponga que las dos poblaciones se distribuyen normalmente con varianzas iguales.

Del enunciado resulta claro que las muestras son independientes y las varianzas desconocidas pero iguales. Sea  $X$  el tiempo de sobrevivencia de los ratones que recibieron tratamiento y sea  $Y$  el tiempo de sobrevivencia de los ratones que no recibieron tratamiento. Para el contraste hay que considerar que el suero es efectivo si la media de ratones que fueron sometidos al tratamiento es mayor con respecto del tiempo medio de supervivencia de los ratones que no tuvieron tratamiento, así tenemos que  $d_0 = 0$ . Por lo tanto la prueba de hipótesis queda como:

$$H_0 : (\mu_x - \mu_y) \geq 0 \quad \text{vs} \quad H_1 : (\mu_x - \mu_y) < 0$$

Por otro lado, haciendo el calculo de:

	Media muestral	Cuasivarianza muestral	Tamaño de muestra
X	2.86	3.883	5
Y	2.07	1.365	4

Donde  $\hat{\sigma}_p^2$  está dada por:

$$\hat{\sigma}_p^2 = \frac{(5-1) * 3.883 + (4-1) * 1.365}{5+4-2} = 2.802786$$

Con estadístico de prueba dado por:

$$T = \frac{(2.86 - 2.07) - 0}{\sqrt{2.802786 \left(\frac{1}{5} + \frac{1}{4}\right)}} = 0.6989859$$

Tenemos que se rechaza  $H_0$  si  $T < t_{(5+4-2)}^{0.05} = -1.894579$  y como  $-1.894579 < 0.6989859$  entonces no se rechaza  $H_0$ , por lo que se puede decir que el suero es efectivo en el tratamiento de la leucemia.

**Contrastes para el cociente de varianzas  $\frac{\sigma_x}{\sigma_y}$  suponiendo  $\mu_x, \mu_y$  conocidas**

**Entrada:**  $X_1, \dots, X_{n_1}$  m.a. de  $N(\mu_x, \sigma_x^2)$  y  $Y_1, \dots, Y_{n_2}$  m.a. de  $N(\mu_y, \sigma_y^2)$ ;

**Estadístico de Prueba:**

$$F = \frac{\frac{1}{n_2} \sum_{i=1}^{n_2} (Y_i - \mu_y)^2}{\frac{1}{n_1} \sum_{i=1}^{n_1} (X_i - \mu_x)^2} \sim F_{(n_2, n_1)}$$

$H_0$	$H_1$	$\mathcal{C}_{LRT}^*$ de tamaño $\alpha$
$\frac{\sigma_x^2}{\sigma_y^2} = 1$	$\frac{\sigma_x^2}{\sigma_y^2} > 1$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{\frac{1}{n_2} \sum_{i=1}^{n_2} (y_i - \mu_y)^2}{\frac{1}{n_1} \sum_{i=1}^{n_1} (x_i - \mu_x)^2} > F_{(n_2, n_1)}^{1-\alpha} \right\}$
$\frac{\sigma_x^2}{\sigma_y^2} = 1$	$\frac{\sigma_x^2}{\sigma_y^2} < 1$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{\frac{1}{n_2} \sum_{i=1}^{n_2} (y_i - \mu_y)^2}{\frac{1}{n_1} \sum_{i=1}^{n_1} (x_i - \mu_x)^2} < F_{(n_2, n_1)}^\alpha \right\}$
$\frac{\sigma_x^2}{\sigma_y^2} = 1$	$\frac{\sigma_x^2}{\sigma_y^2} \neq 1$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{\frac{1}{n_2} \sum_{i=1}^{n_2} (y_i - \mu_y)^2}{\frac{1}{n_1} \sum_{i=1}^{n_1} (x_i - \mu_x)^2} < F_{(n_2, n_1)}^{\frac{\alpha}{2}} \text{ o } \frac{\frac{1}{n_2} \sum_{i=1}^{n_2} (y_i - \mu_y)^2}{\frac{1}{n_1} \sum_{i=1}^{n_1} (x_i - \mu_x)^2} > F_{(n_2, n_1)}^{1-\frac{\alpha}{2}} \right\}$
$\frac{\sigma_x^2}{\sigma_y^2} \leq 1$	$\frac{\sigma_x^2}{\sigma_y^2} > 1$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{\frac{1}{n_2} \sum_{i=1}^{n_2} (y_i - \mu_y)^2}{\frac{1}{n_1} \sum_{i=1}^{n_1} (x_i - \mu_x)^2} > F_{(n_2, n_1)}^{1-\alpha} \right\}$
$\frac{\sigma_x^2}{\sigma_y^2} \geq 1$	$\frac{\sigma_x^2}{\sigma_y^2} < 1$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{\frac{1}{n_2} \sum_{i=1}^{n_2} (y_i - \mu_y)^2}{\frac{1}{n_1} \sum_{i=1}^{n_1} (x_i - \mu_x)^2} < F_{(n_2, n_1)}^\alpha \right\}$

**Ejemplo 4.4.14.** Dos proveedores fabrican un engrane de plástico utilizado en una impresora láser. Una característica importante de estos engranes es la resistencia al impacto la cual se mide en pies-libras. Una muestra aleatoria de engranes suministrados ambos proveedores arrojando los siguientes resultados:

Fabricante 1	294.3863	281.7431	278.9542	288.6915	292.5826	283.7507	298.0782	286.4465
Fabricante 2	245.1440	249.8505	242.8463	242.1844	257.2024	253.4012		

Se sabe que la resistencia al impacto en ambos casos sigue una distribución normal, también se sabe que la resistencia promedio de los engranes fabricados es de 290 y 250 para el fabricante 1 y el fabricante 2 respectivamente. Se desea saber con una confianza del 90 % si la variación de niveles de resistencia de los engranes son iguales.

Sea  $X$  la resistencia al impacto de los engranes producidos por el fabricante 1, así como  $Y$  la resistencia al impacto de los engranes producidos por el fabricante 2. En este caso tenemos que la prueba de hipótesis viene dada por:

$$H_0 : \frac{\sigma_x^2}{\sigma_y^2} = 1 \quad \text{vs} \quad H_1 : \frac{\sigma_x^2}{\sigma_y^2} \neq 1$$

Con estadístico de prueba dado por:

$$F = \frac{\frac{1}{6} \sum_{i=1}^6 (Y_i - 250)^2}{\frac{1}{8} \sum_{i=1}^8 (X_i - 290)^2} = \frac{33.21747}{41.84327} = 0.7938547$$

Tenemos que se rechaza  $H_0$  si  $F < F_{(6,8)}^{0.1} = 0.2792843$  o  $F > F_{(6,8)}^{1-0.1} = 4.146804$  y como  $0.2792843 < 0.7938547 < 4.146804$  entonces no se rechaza  $H_0$ , por lo que se puede decir que no hay evidencia suficiente para rechazar que la variación en los niveles de resistencia de los engranes es la misma para ambos fabricantes.

#### 4.4.6. Contrastes para el cociente de varianzas $\frac{\sigma_x}{\sigma_y}$ suponiendo $\mu_x, \mu_y$ desconocidas

**Entrada:**  $X_1, \dots, X_{n_1}$  m.a. de  $N(\mu_x, \sigma_x^2)$  y  $Y_1, \dots, Y_{n_2}$  m.a. de  $N(\mu_y, \sigma_y^2)$ ;

**Estadístico de Prueba:**

$$F = \frac{\frac{1}{n_2-1} \sum_{i=1}^{n_2} (Y_i - \bar{Y})^2}{\frac{1}{n_1-1} \sum_{i=1}^{n_1} (X_i - \bar{X})^2} \sim F_{(n_2-1, n_1-1)}$$

$H_0$	$H_1$	$\mathcal{C}_{LRT}^*$ de tamaño $\alpha$
$\frac{\sigma_x^2}{\sigma_y^2} = 1$	$\frac{\sigma_x^2}{\sigma_y^2} > 1$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{\frac{1}{n_2-1} \sum_{i=1}^{n_2} (y_i - \bar{y})^2}{\frac{1}{n_1-1} \sum_{i=1}^{n_1} (x_i - \bar{x})^2} > F_{(n_2-1, n_1-1)}^{1-\alpha} \right\}$
$\frac{\sigma_x^2}{\sigma_y^2} = 1$	$\frac{\sigma_x^2}{\sigma_y^2} < 1$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{\frac{1}{n_2-1} \sum_{i=1}^{n_2} (y_i - \bar{y})^2}{\frac{1}{n_1-1} \sum_{i=1}^{n_1} (x_i - \bar{x})^2} < F_{(n_2-1, n_1-1)}^\alpha \right\}$
$\frac{\sigma_x^2}{\sigma_y^2} = 1$	$\frac{\sigma_x^2}{\sigma_y^2} \neq 1$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{\frac{1}{n_2-1} \sum_{i=1}^{n_2} (y_i - \bar{y})^2}{\frac{1}{n_1-1} \sum_{i=1}^{n_1} (x_i - \bar{x})^2} < F_{(n_2-1, n_1-1)}^{\frac{\alpha}{2}} \quad \text{o} \quad \frac{\frac{1}{n_2-1} \sum_{i=1}^{n_2} (y_i - \bar{y})^2}{\frac{1}{n_1-1} \sum_{i=1}^{n_1} (x_i - \bar{x})^2} > F_{(n_2-1, n_1-1)}^{1-\frac{\alpha}{2}} \right\}$
$\frac{\sigma_x^2}{\sigma_y^2} \leq 1$	$\frac{\sigma_x^2}{\sigma_y^2} > 1$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{\frac{1}{n_2-1} \sum_{i=1}^{n_2} (y_i - \bar{y})^2}{\frac{1}{n_1-1} \sum_{i=1}^{n_1} (x_i - \bar{x})^2} > F_{(n_2-1, n_1-1)}^{1-\alpha} \right\}$
$\frac{\sigma_x^2}{\sigma_y^2} \geq 1$	$\frac{\sigma_x^2}{\sigma_y^2} < 1$	$\mathcal{C}_{LRT}^* = \left\{ \underline{x} \in \mathfrak{X} \mid \frac{\frac{1}{n_2-1} \sum_{i=1}^{n_2} (y_i - \bar{y})^2}{\frac{1}{n_1-1} \sum_{i=1}^{n_1} (x_i - \bar{x})^2} < F_{(n_2-1, n_1-1)}^\alpha \right\}$

**Ejemplo 4.4.15.** Continuando con el ejemplo 4.4.14 supongamos ahora que se tiene total desconocimiento del valor de los parámetros realice la misma prueba de hipótesis a un nivel del 90 %.

$$H_0 : \frac{\sigma_x^2}{\sigma_y^2} = 1 \quad \text{vs} \quad H_1 : \frac{\sigma_x^2}{\sigma_y^2} \neq 1$$

Con estadístico de prueba dado por:

$$F = \frac{\frac{1}{(6-1)} \sum_{i=1}^6 (Y_i - \bar{Y})^2}{\frac{1}{(8-1)} \sum_{i=1}^8 (X_i - \bar{X})^2} = \frac{36.93366}{43.60406} = 0.8470232$$

Tenemos que se rechaza  $H_0$  si  $F < F_{(6-1, 8-1)}^{\frac{0.1}{2}} = 0.2050915$  o  $F > F_{(6-1, 8-1)}^{1-\frac{0.1}{2}} = 3.97152$  y como  $0.2050915 < 0.8470232 < 3.97152$  entonces no se rechaza  $H_0$ , por lo que se puede decir que no hay evidencia suficiente para rechazar que la variación en los niveles de resistencia de los engranes es la misma para ambos fabricantes.

#### 4.4.7. Contrastes para la igualdad de medias de k poblaciones normales (varianzas conocidas)

Una de las pruebas más utilizadas en la práctica es la que nos permite ver si existe alguna diferencia en las medias de  $k$  poblaciones normales, por ejemplo, imaginemos que estamos realizando



un experimento con 3 tipos de nuevos anti hipertensivos, para ello se forman 3 grupos de personas con hipertensión, a cada uno de los grupos se le aplica un tipo de anti hipertensivo y después de cierto tiempo se cuantifica la eficacia del medicamento midiendo la presión de cada individuo en todos los grupos. Imaginemos que se desea verificar si existe un anti hipertensivo que sea mejor que los demás. Para ello se decide modelar  $X_{i,j}$  = la presión arterial del individuo  $j$  ( $j \in \{1, \dots, n_i\}$ ) en la población  $i$  ( $i \in \{1, 2, 3\}$ ). Si suponemos que

$$X_{i,j} \sim N(\mu_i, \sigma_i^2) \quad i \in \{1, 2, 3\} \quad j \in \{1, \dots, n_i\}$$

Entonces el problema se podría reducir a contrastar la siguientes hipótesis:

$$H_0 : \mu_1 = \mu_2 = \mu_3 \quad vs \quad H_1 : \exists i, j \text{ tal que } \mu_i \neq \mu_j$$

La región de rechazo para este tipo de pruebas puede ser encontrada utilizando la técnica L.T.R., en el cual si suponemos varianzas conocidas se obtiene lo siguiente:

**Contraste:**

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_k \quad vs \quad H_1 : \exists i, j \text{ tal que } \mu_i \neq \mu_j$$

**Entrada:**

$$\begin{array}{ll} X_{1,1}, \dots, X_{1,n_1} & \text{m.a. de } N(\mu_1, \sigma_1^2) \\ X_{2,1}, \dots, X_{2,n_2} & \text{m.a. de } N(\mu_2, \sigma_2^2) \\ & \vdots \\ X_{k,1}, \dots, X_{k,n_k} & \text{m.a. de } N(\mu_k, \sigma_k^2) \end{array}$$

**Estadístico de Prueba:**

$$T = \sum_{i=1}^k \frac{n_i}{\sigma_i^2} (\bar{X}_i - \bar{X}_{**})^2 \sim \chi_{(k-1)}^2$$

Donde:

$$\bar{X}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{i,j} \quad \bar{X}_{**} = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^{n_i} X_{i,j} \quad \sum_{i=1}^k n_i = n$$

**Región de Rechazo:**

$$\mathcal{C}_{LRT}^* = \left\{ \sum_{i=1}^k \frac{n_i}{\sigma_i^2} (\bar{x}_i - \bar{x}_{**})^2 > \chi_{(k-1)}^{2(1-\alpha)} \right\}$$

**Ejemplo 4.4.16.** *Se llea cabo un estudio para determinar si existe diferencia en la temperatura media en las cuatro estaciones diferentes del año en el río Balsas. El termómetro mide en grados celsius. Se tomaron muestras aleatorias independientes que arrojaron los siguientes resultados:*

Invierno	25.6	23.2	21.7	21.7	22.8	23.9	24.5	23.8	24.5	24.6	25.4	26.1	24.9
Primavera	25.4	22.7	26.6	22.5	22.4	24.3	25.8	23.8	25.0				
Verano	21.8	21.7	20.7	22.0	24.7	26.1	23.7	25.4	22.4	23.8	25.1	24.1	
Otoño	24.8	24.7	24.7	24.8	23.7	25.3	23.3						

Nombraremos como la variable 1 a la temperatura del río Balsas durante el invierno y de forma análoga para las tres restantes. Adicionalmente se sabe que la distribución asociada en todos los casos es normal y la desviación estándar para cada una de ellas es:

$$\sigma_1 = 1.5 \quad \sigma_2 = 1.2 \quad \sigma_3 = 1.4 \quad \sigma_4 = 1.3$$

Con un nivel del 95 % determine si la temperatura media del río Balsas es indiferente a la estación del año.

Se tiene que la prueba de hipótesis es:

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 \quad \text{vs} \quad H_1 : \exists i, j \text{ tal que } \mu_i \neq \mu_j$$

Tenemos que en este caso  $k = 4$  y que:

$$n_1 = 13 \quad n_2 = 9 \quad n_3 = 12 \quad n_4 = 7 \quad \Rightarrow \quad n = 41$$

con

$$\bar{X}_1 = 24.05385 \quad \bar{X}_2 = 24.27778 \quad \bar{X}_3 = 23.45833 \quad \bar{X}_4 = 24.47143$$

Por otro lado obtenemos  $\bar{X}_{**} = \frac{984}{41} = 24$

Por otro lado, calculando el estadístico de prueba

$i$	$n_i$	$\sigma_i^2$	$\frac{n_i}{\sigma_i^2}(\bar{X}_i - \bar{X}_{**})^2$
1	13	1.5 <sup>2</sup>	0.01675214
2	9	1.2 <sup>2</sup>	0.4822531
3	12	1.4 <sup>2</sup>	1.946039
4	7	1.3 <sup>2</sup>	1.709576
$T$			4.15462

Recordemos que se rechaza  $H_0$  si  $T > \chi_{(4-1)}^{2(1-0.05)} = 7.814728$  y como  $4.15462 < 7.814728$  por lo que no se rechaza  $H_0$  y podemos decir que la temperatura promedio del río Balsas es indiferente a la estación del año.

#### 4.4.8. Contrastes para la igualdad de medias de k poblaciones normales (varianzas desconocidas pero iguales) ANOVA

El problema de la prueba anterior es que por lo general las varianzas van a ser desconocidas en el problema por lo que es necesario crear una regla de decisión para el caso en que las varianzas son desconocidas.

Atacar el problema de varianzas desconocidas es complejo y el estadístico de prueba se vuelve intratable cuando suponemos varianzas desiguales en cada población, afortunadamente cuando se supone homocedasticidad (igualdad de varianzas) los cálculos quedan mas ordenados y pueden ser presentados en una tabla llamada ANOVA (análisis de varianza)

**Contraste:**

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_k \quad vs \quad H_1 : \exists i, j \text{ tal que } \mu_i \neq \mu_j$$

**Entrada:**

$$\begin{array}{ll} X_{1,1}, \dots, X_{1,n_1} & \text{m.a. de } N(\mu_1, \sigma^2) \\ X_{2,1}, \dots, X_{2,n_2} & \text{m.a. de } N(\mu_2, \sigma^2) \\ & \vdots \\ X_{k,1}, \dots, X_{k,n_k} & \text{m.a. de } N(\mu_k, \sigma^2) \end{array}$$

**Obs:** Observe que estamos suponiendo igualdad de varianzas en las poblaciones (homocedasticidad) además para poder realizar esta prueba se requiere que  $n_i > 2$  para toda  $i \in \{1, 2, \dots, k\}$ .

**Estadístico de Prueba:**

$$F = \frac{(n - k) \sum_{i=1}^k n_i (\bar{X}_i - \bar{X}_{**})^2}{(k - 1) \sum_{i=1}^k \sum_{j=1}^{n_i} (X_{i,j} - \bar{X}_i)^2} \sim F_{k-1, n-k}$$

Donde:

$$\bar{X}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{i,j} \quad \bar{X}_{**} = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^{n_i} X_{i,j} \quad \sum_{i=1}^k n_i = n$$

**Región de Rechazo:**

$$\mathcal{C}_{LRT}^* = \left\{ \frac{(n - k) \sum_{i=1}^k n_i (\bar{x}_i - \bar{x}_{**})^2}{(k - 1) \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{i,j} - \bar{x}_i)^2} > F_{k-1, n-k}^{1-\alpha} \right\}$$

Una forma en como los paquetes estadísticos presentan esta prueba es mediante la tabla ANOVA:

Variación	Suma de Cuadrados	G. de Lib	Cuadrados Medios	Estadístico F
Tratamientos	$\sum_{i=1}^k n_i (\bar{x}_i - \bar{x}_{**})^2$	$k - 1$	$\frac{\sum_{i=1}^k n_i (\bar{x}_i - \bar{x}_{**})^2}{k-1}$	$\frac{(n-k) \sum_{i=1}^k n_i (\bar{x}_i - \bar{x}_{**})^2}{(k-1) \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{i,j} - \bar{x}_i)^2}$
Error	$\sum_{i=1}^k \sum_{j=1}^{n_i} (x_{i,j} - \bar{x}_i)^2$	$n - k$	$\frac{\sum_{i=1}^k \sum_{j=1}^{n_i} (x_{i,j} - \bar{x}_i)^2}{n-k}$	
Total	$\sum_{i=1}^k \sum_{j=1}^{n_i} (x_{i,j} - \bar{x}_{**})^2$	$n - 1$	$\frac{\sum_{i=1}^k \sum_{j=1}^{n_i} (x_{i,j} - \bar{x}_{**})^2}{n-1}$	

**Ejemplo 4.4.17.** A continuación se presenta un código en R que implementa esta prueba.

```
#####
##### Prueba de hipotesis para igualdad de medias #####
##### k poblaciones normales varianzas iguales #####
##### y desconocidas #####
#####

#-- Funcion que recibe por parametro una matriz cuyo renglon i-esimo
#-- son las observaciones de la poblacion i-esima
prueba<-function(data){
size<-dim(data)
ni<-size[2]-apply(apply(data,1,is.na),2,sum)
if (sum(ni[ni<2])>0 ) {
  cat("Error: You need a sample bigger than one for any population i where k>2")
} else {
xbar<-apply(data,1,mean,na.rm=TRUE)
xbarasterix<-mean(data,na.rm=TRUE)
SCTR<-sum(ni*(xbar-xbarasterix)^2,na.rm=TRUE)
SCE<-sum((ni-1)*apply(data,1,var,na.rm=TRUE))
SCT<-sum((data-xbarasterix)^2,na.rm=TRUE)
Var<-c("Treatment","Error","Total")
SumSq<-c(SCTR,SCE,SCT)
Df<-c(size[1]-1,sum(ni)-size[1],sum(ni)-1)
SumMeanSq<-SumSq/Df
Fvalue<-c(SumMeanSq[1]/SumMeanSq[2],NA,NA)
res<-data.frame(Var,SumSq,Df,SumMeanSq,Fvalue)
return(res)
}
}
```

Para probar esta función simularemos 5 muestras de poblaciones normales cuyos tamaño de muestra son 100, 2, 26, 100, 88 de poblaciones normales independientes  $N(0, 1)$ . Supongamos para este ejercicio que tenemos desconocimiento total de los parámetros pero se desea hacer constrañte sobre si las medias son iguales con  $\alpha = 5\%$ . Por lo utilizando la función implementada arriba tenemos elaborado el siguiente código:

```
datos<-matrix(rnorm(500),nrow=5,ncol=100)
datos[2,3:100]<-rep(NA,98)
datos[3,27:100]<-rep(NA,74)
```

```

datos[5,89:100]<-rep(NA,12)
fin<-prueba(datos)
fin
alpha<-0.05
cuant<-qf(1-alpha,fin$Df[1],fin$Df[2])
if(fin$Fvalue[1]>cuant) {cat("Como",fin$Fvalue[1],"es mayor que",cuant,+
  "Se rechaza H0")}
} else {cat("Como",fin$Fvalue[1],"es menor que",cuant, "No se rechaza H0")}

```

Que produce la siguiente salida:

Como 1.72022 es menor que 2.400678 No se rechaza H0

**Ejercicio 4.4.1.** Resuelva lo siguiente:

1. Implemente una función en **R** por cada una de las pruebas de hipótesis para caso normal.  
Hint: consulte la documentación en **R** de las funciones **switch()** y **apply()**.

2. Sea  $X_1$  una observación del modelo:

$$f_X(x) = \theta x^{\theta-1} \mathbf{1}_{(0,1)}(x); \quad \theta > 0$$

Se desea contrastar las siguientes hipótesis:

$$H_0 : \theta = 2 \quad vs \quad H_1 : \theta = 1$$

- Encontrar la región de rechazo mas potente suponiendo que deseamos una probabilidad de error tipo 1  $\alpha = 0.05$ .
  - Para la región de rechazo encontrada en el inciso anterior, calcular la probabilidad del error tipo 2 y encuentre la potencia de la prueba
3. Sea  $X_1, \dots, X_5$  m.a. del modelo  $\text{Normal}(0, \sigma^2)$ . Se desea contrastar las siguientes hipótesis:

$$H_0 : \sigma^2 = 1 \quad vs \quad H_1 : \sigma^2 = 2$$

- Encontrar la región de rechazo mas potente suponiendo que deseamos una probabilidad de error tipo 1  $\alpha = 0.1$ .
- Para la región de rechazo encontrada en el inciso anterior, calcular la probabilidad del error tipo 2 y encuentre la potencia de la prueba
- Suponga que observa la siguiente muestra: (4.14, 0.78, -1.28, 1.30, 1.33), ¿rechazaría  $H_0$ ?, encuentre el p-value asociado a esta prueba.

4. Sea  $X_1$  una observación del modelo:

$$f_X(x) = (2\theta x + 1 - \theta) \mathbf{1}_{(0,1)}(x); \quad \theta \in [-1, 1]$$

Se desea contrastar las siguientes hipótesis:

$$H_0 : \theta = 0 \quad \text{vs} \quad H_1 : \theta = 1$$

- Encontrar la región de rechazo mas potente suponiendo que deseamos una probabilidad de error tipo 1  $\alpha = 0.05$ .
- Para la región de rechazo encontrada en el inciso anterior, calcular la probabilidad del error tipo 2 y encuentre la potencia de la prueba
- Suponga que observa la siguiente muestra: (0.5), ¿rechazaría  $H_0$ ?, encuentre el p-value asociado a esta prueba.

5. Sea  $X_1, \dots, X_7$  m.a. del modelo  $\text{Gamma}(2, \beta)$

$$f_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \mathbf{1}_{(0,\infty)}(x); \quad \beta > 0$$

Se desea contrastar las siguientes hipótesis:

$$H_0 : \beta = 1 \quad \text{vs} \quad H_1 : \beta = 4$$

- Encontrar la región de rechazo mas potente suponiendo que deseamos una probabilidad de error tipo 1  $\alpha = 0.05$ .
- Para la región de rechazo encontrada en el inciso anterior, calcular la probabilidad del error tipo 2 y encuentre la potencia de la prueba.
- Suponga que observa la siguiente muestra: (0.650, 1.048, 3.109, 0.526, 0.558, 1.228, 1.320), ¿rechazaría  $H_0$ ?, encuentre el p-value asociado a esta prueba.

6. INFOTEC utiliza bombillas especiales para poder iluminar los distintos salones de sus instalaciones. Actualmente utiliza bombillas fabricadas por la **compañía 1** la cual afirma que sus bombillas tienen una vida útil de 900 horas en promedio, la **compañía 2** afirma que ellos fabrican una nueva bombilla con una tecnología mas eficiente que hace que sus bombillas duren mas de 900 horas en promedio. INFOTEC desea tomar una decisión sobre si cambia de proveedor o no, para ello optó por hacer un experimento sencillo, compró 20 bombillas a cada compañía y las colocó en distintos salones para medir su durabilidad.

Los datos de este experimento se encuentran la siguiente tabla. Con esta información responda lo siguiente suponiendo normalidad e igualdad de varianzas en ambas poblaciones.

COMPAÑIA1	COMPAÑIA2
856.33	866.77
969.69	907.61
828.62	831.06
962.91	937.76
917.56	963.75
938.78	941.05
930.21	980.58
759.69	940.63
809.59	926.95
802.53	862.5
779.71	858.92
728.85	945.34
889.03	775.36
797.97	960.19
845.82	891.21
767.92	939.64
824.37	978.08
940.06	922.53
891.85	973.22
897.06	995.29

- ¿ Se tiene evidencia estadística como para afirmar que las bombillas de las **compañía 1** en efecto tienen una vida promedio de 900 horas?. Para responder esto lleve a cabo la prueba de hipótesis:

$$H_0 : \mu_1 = 900 \quad vs \quad H_1 : \mu_1 \neq 900$$

Donde  $\mu_1$  representa el tiempo de vida promedio de las bombillas de la **compañía 1**. Calcule el p-value de la prueba y asumiendo un  $\alpha = 0.05$  concluya.

- ¿Se tiene evidencia estadística como para afirmar que las bombillas de la **compañía 2** tienen una vida promedio mayor a 900 horas?. Para responder este lleve a cabo la prueba de hipótesis:

$$H_0 : \mu_2 \leq 900 \quad vs \quad H_1 : \mu_2 > 900$$

Donde  $\mu_2$  representa el tiempo de vida promedio de las bombillas de la **compañía 2** . Calcule el p-value de la prueba y asumiendo un  $\alpha = 0.05$  concluya.

- Finalmente, se desea tomar la decisión final para saber si se cambia o no de proveedor

de bombillas. Para responder esto lleve a cabo la prueba de hipótesis.

$$H_0 : \mu_1 = \mu_2 \quad vs \quad H_1 : \mu_1 < \mu_2$$

Calcule el p-value de la prueba y asumiendo un  $\alpha = 0.05$  concluya.

7. Una compañía de jugos compró una máquina que llena botellas. El manual de instructivo dice que la máquina está configurada para llenar las botellas en promedio con 1 litro de líquido. La compañía está interesada en verificar esta afirmación, para ello hizo que la máquina llenara  $n = 10$  botellas y luego midió la cantidad exacta (en litros) con la que fueron llenadas. La información se encuentra en la siguiente tabla.

BOTELLA	LITROS SURTIDOS
1	0.96
2	1.01
3	1.14
4	0.87
5	1.08
6	1.05
7	1.13
8	0.99
9	1.06
10	1.05

Para probar la aseveración y dado que la compañía esta preocupada de que la máquina este surtiendo mas líquido de lo esperado se planteó la siguiente prueba de hipótesis

$$H_0 : \mu \leq 1 \quad vs \quad H_1 : \mu > 1$$

Asumiendo normalidad y varianza conocida ( $\sigma = 0.1$ )La compañía propuso la siguiente región de rechazo:

$$C = \left\{ \sqrt{n} \left( \frac{\bar{X} - 1}{0.1} \right) > 1.8 \right\}$$

- Con los datos observados, calcule el p-value y concluya si rechaza o no la hipótesis nula. (utilice  $\alpha = 0.5$  como su nivel de significancia para decidir)
- Realice el contraste nuevamente pero ahora suponga  $\sigma^2$  desconocida (utilice  $\alpha = 0.5$  como su nivel de significancia para decidir)



## 4.5. Modelos Lineales

El objetivo en esta sección es comprender el alcance del análisis de regresión así como los supuestos que debe de cumplir el modelo establecido, en este sentido el lector debe de ser capaz de ajustar un modelo de regresión, evaluando y mejorando el desempeño del mismo.

Los modelos matemáticos buscan explicar y entender distintos fenómenos que ocurren en el mundo que nos rodean. En general estos fenómenos pueden ser clasificados en deterministas y probabilísticas, estos últimos son de especial interés para la estadística ya que son lo que se pretenden modelar.

- **Probabilistas:** Las mismas entradas no necesariamente producirán la mismas salidas pues contempla la interacción con variables que no podemos controlar (azar).

Por ejemplo: Sea  $X$  el número de águilas en 6 lanzamientos independientes de una moneda "honesta", se propone el modelo probabilístico:

$$X \sim \text{Bin}(6, 0.5); \quad \mathbb{E}(X) = 3; \quad \text{Var}(X) = 1.5$$

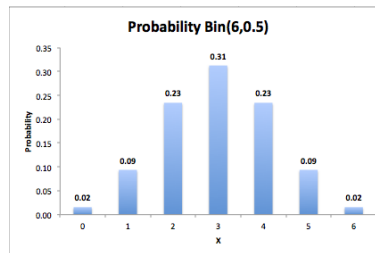


Figura 4.14: Binomial Density

En general un modelo busca encontrar una relación funcional entre ciertas variables independientes  $(z_1, z_2, \dots, z_q)$  y una variable respuesta  $Y$  la cual queremos explicar o modelar.

$$Y = f(z_1, z_2, \dots, z_q) \quad (\text{Determinista})$$

$$Y = f(z_1, z_2, \dots, z_q) + \varepsilon \quad (\text{Probabilístico})$$

Una vez planteado el problema surgen de manera natural las siguientes preguntas:

- ¿Qué función  $f$  tomar?
- ¿El modelo es adecuado?
- ¿Tenfo el mejor modelo para estos datos?

- ¿Qué supuestos debemos tomar en cuenta?
- ¿Las variables  $z_i$  si explican a  $Y$ ?
- ¿Mi modelo está respaldado con buenos datos?

Por ejemplo, se desea saber si la estatura máxima del padre influye en la estatura máxima del hijo, es decir, queremos modelar la estatura del hijo en función de la estatura del padre.

$$Y = f(z)$$

Donde  $z$  = estatura del padre y  $Y$  = estatura del hijo.

¿Existe un modelo determinista?

$$Y = f(z) + \varepsilon$$

Dada una variable respuesta  $Y$  y un conjunto de covariables  $\underline{z} = (z_1, z_2, \dots, z_q)$ , surge de manera

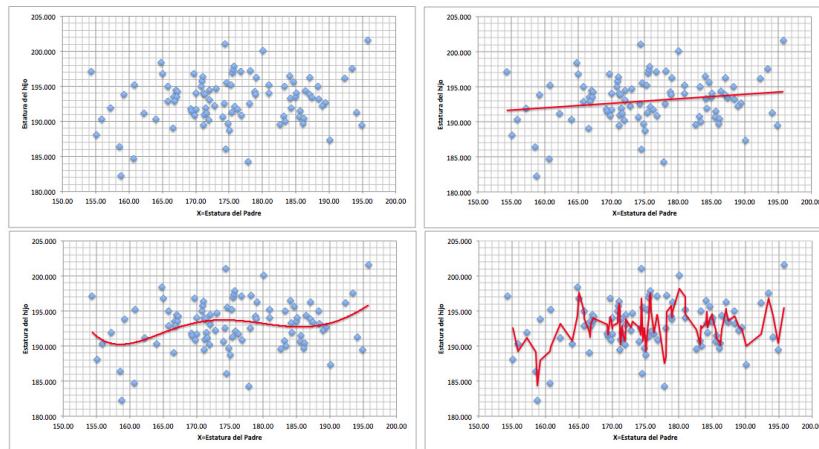


Figura 4.15: Estaturas

natural preguntarnos cuál deberá ser la relación funcional para modelar dicha relación.

Una forma de modelar podría ser:

$$\mathbb{E}(Y | \underline{z}) = \mu(\underline{z})$$

donde, en general,  $\mu(\cdot)$  es una función desconocida. En la práctica es común aproximar a  $\mu(\cdot)$  a través de una función más simple:

$$\mu(\underline{z}) = \psi(\underline{z}; \underline{\beta})$$

donde  $\underline{\beta} = (\beta_0, \beta_1, \dots, \beta_k)^T$  denota a un vector de parámetros desconocidos.

La forma mas simple para modelar la relación es suponer una función lineal de  $\underline{\beta}$  es decir:

$$\psi(\underline{z}; \underline{\beta}) = \beta_0 + \beta_1 s_1(\underline{z}) + \dots + \beta_k s_k(\underline{z})$$

Donde  $s_i : \mathbb{R}^q \rightarrow \mathbb{R}$  son funciones conocidas.

Finalmente esta función  $\psi(\underline{z}; \underline{\beta})$  es tratada como si fuera la verdadera función de regresión  $\mu(\cdot)$ , por lo que el problema se reduce en hacer inferencias sobre el valor del vector de parámetros  $\underline{\beta}$ .

$$\mathbb{E}(Y | \underline{z}) = \mu(\underline{z}) = \beta_0 + \beta_1 s_1(\underline{z}) + \dots + \beta_k s_k(\underline{z})$$

Muchas veces, en la vida real el modelo lineal no ajustará los datos por lo que tenemos que ver modelos lineales generalizados (siguiente sección) o bien proponer modelos no lineales. Modelos Lineales:

- $\psi(\underline{z}; \underline{\beta}) = \beta_0 + \beta_1 z_1$
- $\psi(\underline{z}; \underline{\beta}) = \beta_0 + \beta_1 z_1 + \beta_2 z_2 + \dots + \beta_p z_p$
- $\psi(\underline{z}; \underline{\beta}) = \beta_0 + \beta_1 z_1 + \beta_2 z_1^2 + \dots + \beta_p z_1^p$
- $\psi(\underline{z}; \underline{\beta}) = \beta_1 \sin(z_1) + \beta_2 \cos(z_2)$

Modelos no Lineales:

- $\psi(\underline{z}; \underline{\beta}) = \beta_0 + \sin(\beta_1 z_1)$
- $\psi(\underline{z}; \underline{\beta}) = \beta_0 + \cos(\beta_1 z_1)$
- $\psi(\underline{z}; \underline{\beta}) = \tan(\beta_1 z_1) + \beta_2 z_2$

También existen en la literatura modelos denominados **linealizables**. Diremos que un modelo es linealizable si existe una función  $g$  invertible tal que:

$$g(\psi(\underline{z}; \underline{\beta})) = \beta_0 + \beta_1 s_1(\underline{z}) + \dots + \beta_k s_k(\underline{z})$$

Por ejemplo:

- $\psi(\underline{z}; \underline{\beta}) = \beta_0 \exp(\beta_1 z_1) \Rightarrow \ln(\psi(\underline{z}; \underline{\beta})) = \ln(\beta_0) + \beta_1 z_1 = \beta_0^* + \beta_1 z_1$
- $\psi(\underline{z}; \underline{\beta}) = \log(\beta_0 + \beta_1 \cos(z_1)) \Rightarrow \exp(\psi(\underline{z}; \underline{\beta})) = \beta_0 + \beta_1 \cos(z_1)$

Nuestro modelo lineal supone entonces que:

$$\mathbb{E}(Y | \underline{z}) = \mu(\underline{z}) = \beta_0 + \beta_1 s_1(\underline{z}) + \dots + \beta_k s_k(\underline{z})$$

Esto lo podemos escribir en términos de variables aleatorias como:

$$Y | \underline{z} \sim F(\mu(\underline{z}), \sigma^2) \quad (\sigma^2 > 0, \text{desconocida})$$

Ahora imaginemos que recibiremos  $n$  observaciones de este modelo para distintos niveles de las covariables  $\underline{z}$ , entonces podemos escribir

$$Y_i = \beta_0 + \beta_1 s_1(z_i) + \dots + \beta_k s_k(z_i) + \varepsilon_i \quad \varepsilon_i \stackrel{iid}{\sim} F(0, \sigma^2) \quad (4.16)$$

Definamos:

$$x_{ij} = s_j(z_i) \quad i \in \{1, \dots, n\}; \quad j \in \{1, \dots, k\}$$

Entonces el modelo (4.16) lo podemos escribir como:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + \varepsilon_i \quad \varepsilon_i \stackrel{iid}{\sim} F(0, \sigma^2) \quad (4.17)$$

### Ejemplo:

Suponiendo que sólo tenemos una covariable  $z \in \mathbb{R}$ , y haciendo  $s_j(z) = z^j$  entonces (4.17) toma la forma:

$$Y_i = \beta_0 + \beta_1 z_i + \dots + \beta_k z_i^k + \varepsilon_i \quad \varepsilon_i \stackrel{iid}{\sim} F(0, \sigma^2) \quad (4.18)$$

$$Y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + \varepsilon_i \quad \varepsilon_i \stackrel{iid}{\sim} F(0, \sigma^2) \quad (4.19)$$

El modelo anterior pretende explicar el valor de  $Y$  a través de una función polinomial de la covariable  $z$ .

Obs: Al modelo (4.17) se le conoce como **Regresión Lineal Multiple** mientras que cuando  $k = 1$  lo denominamos el modelo **Lineal Simple** el cual se detalla a continuación.

#### 4.5.1. Regresión Lineal Simple

Se analiza de forma teórica y práctica el modelo lineal simple es decir, se asumirá que los datos que deseamos modelar fueron generados por la ecuación:

$$Y = \beta_0 + \beta_1 x_1 + \varepsilon$$

$$\mathbb{E}(Y) = \beta_0 + \beta_1 x_1 \quad \text{Var}(Y) = \sigma^2 \quad Y \sim F(\beta_0 + \beta_1 x_1, \sigma^2)$$

Se supondrá entonces que se observará una muestra de este modelo para distintos niveles de la covariable  $x_1$ , es decir se cuenta con observaciones  $(y_1, x_{11}), \dots, (y_n, x_{n1})$  de tal manera que:

$$y_i = \beta_0 + \beta_1 x_{i1} + \varepsilon_i$$

Adicionemos una hipótesis importante al modelo:

- El componente aleatorio de una observación  $\varepsilon_i$  *no* está correlacionado con el componente

aleatorio de otra observación, es decir:

$$\text{Cov}(\varepsilon_i, \varepsilon_j) = 0 \quad i \neq j$$

Tenemos los ingredientes necesarios para hacer inferencia desde el punto de vista estadístico:

- Una distribución paramétrica  $Y_i \sim F(\beta_0 + \beta_1 x_{i1}, \sigma^2)$
- Parámetros desconocidos  $(\beta_0, \beta_1, \sigma^2)$
- Una muestra de observaciones de la cual queremos obtener información para estimar los parámetros desconocidos

¿Cómo construir estimadores?

- Máxima Verosimilitud (Requiere saber la forma de la distribución F)
- Momentos (Requiere variables aleatorias independientes e idénticamente distribuidas sin embargo para fines del modelo lineal no se puede aplicar porque el modelo no asume que las variables sean idénticamente distribuidas)
- Mínimos Cuadrados (Busca encontrar los valores de los parámetros que reduzcan las diferencias cuadráticas entre el valor observado y esperado por el modelo)

Obs: Para fines de notación y únicamente para el modelo lineal simple se asumirá que  $x_{i1} = x_i$   
A continuación discutimos los métodos de estimación:

### Estimación por mínimos cuadrados

Para entender este método olvidemos un poco la teoría estadística y supongamos que tenemos el siguiente problema, se busca encontrar la ecuación de la recta  $Y = \beta_0 + \beta_1 x$  que más se ajuste a nuestros datos.

En términos matemáticos buscamos minimizar una función de dos variables  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ , con la siguiente forma:

$$f(\beta_0, \beta_1) = \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i))^2$$

Donde  $(y_1, x_1), (y_2, x_2), \dots, (y_n, x_n)$  son los datos observados. Observe entonces que estamos minimizando el cuadrado de la distancia entre el valor que arroja el modelo  $\hat{Y}$  y el valor observado  $Y$ , es decir  $(Y - \hat{Y})^2$ . A los estimadores encontrados bajo esta metodología se les conoce como **Estimadores de Mínimos Cuadrados**.

Utilizando la correspondiente teoría de cálculo de varias variables se puede probar que los parámetros que minimizan la suma de cuadrados es:

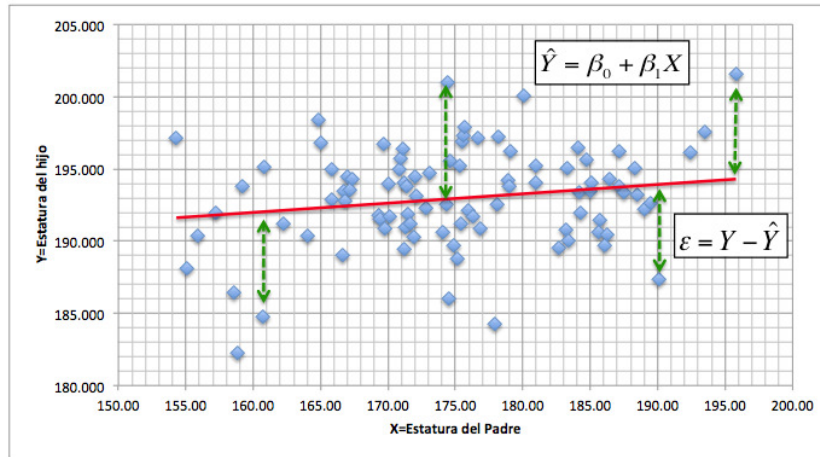


Figura 4.16: Estaturas

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Con  $S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2$  y  $S_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$

En este caso a  $\hat{\beta}_1$  y a  $\hat{\beta}_0$  se les denomina los estimadores de Mınimos Cuadrados.

**Ejercicio 4.5.1.** Verifique que los estimadores de Mınimos Cuadrados son combinaciones lineales de las observaciones  $y_i$ , es decir, pruebe que los estimadores pueden ser representados como

$$\hat{\beta}_1 = \sum_{i=1}^n c_i y_i \quad \hat{\beta}_0 = \sum_{i=1}^n d_i y_i$$

Donde  $c_i$  y  $d_i$  son constantes que no dependen de  $y_i$

En general se puede probar que los estimadores cumplen con lo siguiente:

- Los estimadores son combinaciones lineales de las v.a.  $y_i$ 's pues:

$$\hat{\beta}_1 = \sum_{i=1}^n \left( \frac{x_i - \bar{x}}{S_{xx}} \right) y_i \quad \hat{\beta}_0 = \sum_{i=1}^n \left( \frac{1}{n} - \left( \frac{x_i - \bar{x}}{S_{xx}} \right) \bar{x} \right) y_i$$

- Son insesgados es decir:

$$\mathbb{E}(\hat{\beta}_1) = \beta_1 \quad \mathbb{E}(\hat{\beta}_0) = \beta_0$$

- La varianza de los estimadores es:

$$\text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{S_{xx}} \quad \text{Var}(\hat{\beta}_0) = \left( \frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right) \sigma^2$$

- Se puede probar que los estimadores son los mejores estimadores lineales (BLUES). (Teorema de Gauss-Markov)

## Estimación por máxima verosimilitud

Para poder aplicar esta técnica se requiere agregar el siguiente supuesto:

$$\underline{\varepsilon} \sim N_n(\underline{0}, \sigma \mathbf{I})$$

. Las consecuencias de este importante supuesto son muchas, a continuación exponemos las más importantes:

- Como  $cov(\varepsilon_i, \varepsilon_j) = 0$  entonces  $\varepsilon_i$  es independiente de  $\varepsilon_j$
- Como  $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$  entonces  $y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$
- $y_1, \dots, y_n$  son independientes, pero no son idénticamente distribuidos.

Dado que tenemos la forma de la densidad de  $\varepsilon_i$ , entonces la función verosimilitud puede encontrarse como:

$$\begin{aligned} \mathcal{L}(\beta_0, \beta_1, \sigma^2; \underline{x}, \underline{y}) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2}\right) \\ &= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left(-\sum_{i=1}^n \frac{(y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2}\right) \end{aligned}$$

Bajo los principios de cálculo se prueba que los estimadores por máxima verosimilitud son:

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}; \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}; \quad \hat{\sigma}_{MV}^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2;$$

**Ejercicio 4.5.2.** Demuestre que bajo los supuestos del modelo lineal los estimadores son insesgados, es decir, pruebe que:

$$\mathbb{E}(\hat{\beta}_1) = \beta_1; \quad \mathbb{E}(\hat{\beta}_2) = \beta_2;$$

Por el ejercicio anterior se sabe que  $\hat{\beta}_1$  y  $\hat{\beta}_0$  cumplen la propiedad de insesgamiento, sin embargo en el caso del estimador para  $\sigma^2$  esto no ocurre pues:

$$\mathbb{E}(\hat{\sigma}_{MV}^2) = \frac{(n-2)}{n} \sigma^2$$

Para solucionar lo anterior comúnmente se suele usar al estimador insesgado  $\hat{\sigma}^2$  definido como:

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

## Análisis Inferencial de los parámetros

Hasta este momento hemos encontrado los estimadores puntuales del modelo lineal simple, a continuación se presenta la teoría respectiva para hacer inferencia desde el punto de pruebas de hipótesis e intervalos de confianza

### Pruebas de hipótesis en el modelo lineal

El objetivo es contrastar hipótesis de la forma:

$$H_0 : \beta_1 = b_1 \quad vs \quad H_1 : \beta_1 \neq b_1; \quad H_0 : \beta_1 \leq b_1 \quad vs \quad H_1 : \beta_1 > b_1$$

$$H_0 : \beta_1 \geq b_1 \quad vs \quad H_1 : \beta_1 < b_1$$

Sacando provecho de la normalidad y de que los estimadores máximo verosímiles son combinaciones lineales de las observaciones  $y_i$  se prueba el siguiente resultado:

$$t = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}} \sim t_{n-2}$$

Sacando provecho de esta distribución entonces se utilizan las siguientes reglas de decisión.

Si se plantea la hipótesis:

$$H_0 : \beta_1 = b_1 \quad vs \quad H_1 : \beta_1 \neq b_1$$

Entonces la regla es rechazar  $H_0$  cuando  $|t| \geq \tau_{n-2}^{1-\alpha/2}$ , donde  $t = \frac{\hat{\beta}_1 - b_1}{\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}}$  ;



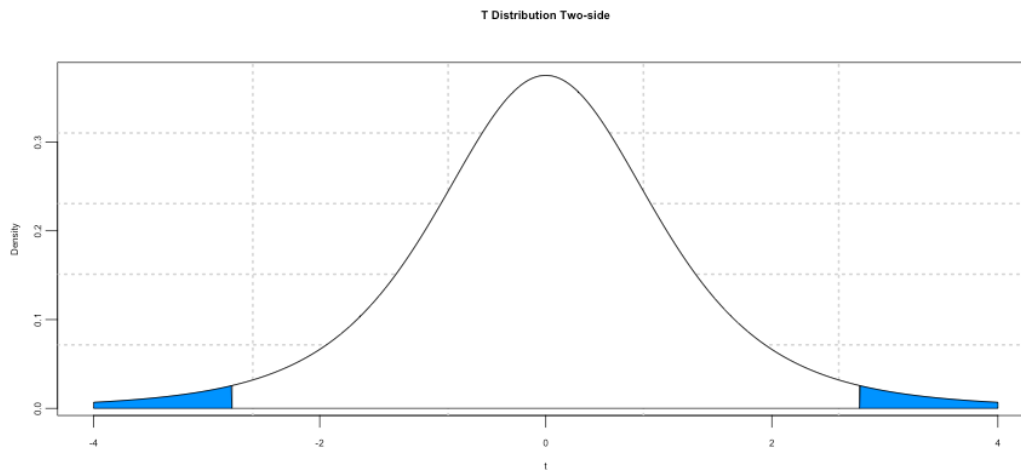


Figura 4.17: Región de Rechazo

**Importante:** Esta es la prueba de hipótesis mas importante dentro del análisis de regresión lineal simple pues cuando hacemos  $b_1 = 0$  nos ayuda a determinar si la pendiente de la recta ajustada es *estadísticamente* diferente de cero lo que se traduce en verificar si hay un efecto de la variable  $X_1$  en la variable  $Y$ .

Si se plantea la hipótesis:

$$H_0 : \beta_1 \leq b_1 \quad vs \quad H_1 : \beta_1 > b_1$$

Entonces la regla es rechazar  $H_0$  cuando  $t \geq \tau_{n-2}^{1-\alpha}$ , donde

$$t = \frac{\hat{\beta}_1 - b_1}{\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}} ;$$

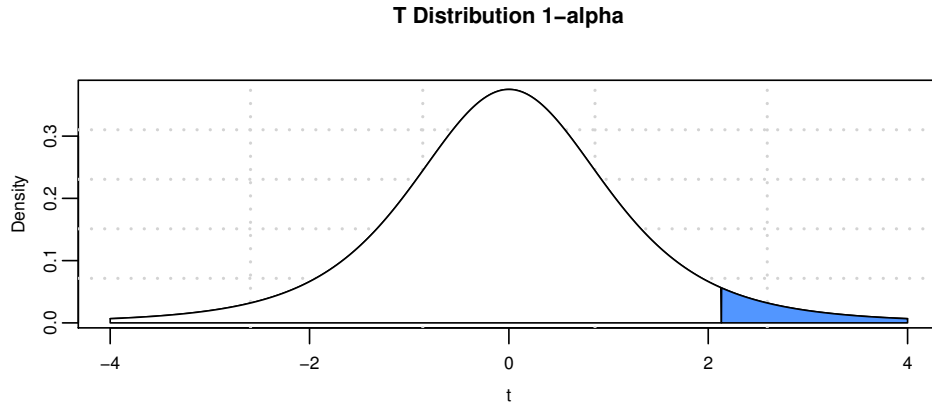


Figura 4.18: Región de Rechazo

Si se plantea la hipótesis:

$$H_0 : \beta_1 \geq b_1 \quad vs \quad H_1 : \beta_1 < b_1$$

Entonces la regla es rechazar  $H_0$  cuando  $t \leq \tau_{n-2}^\alpha$ , donde

$$t = \frac{\hat{\beta}_1 - b_1}{\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}} ;$$

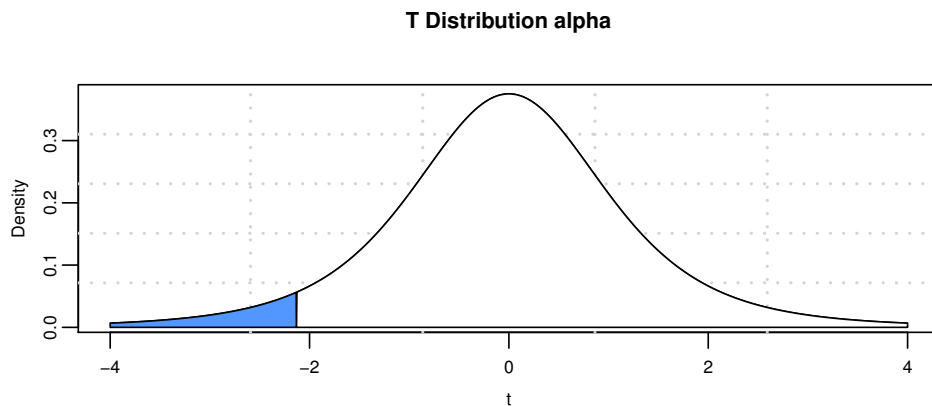


Figura 4.19: Región de Rechazo

Toda la teoría desarrollada hasta ahora fue para analizar el comportamiento de  $\beta_1$ , se puede demostrar haciendo un análisis similar que para  $\beta_0$  se tiene que:

Prueba	Región de Rechazo
$H_0 : \beta_0 \leq b_0$ vs $H_1 : \beta_0 > b_0$	$\frac{\hat{\beta}_0 - b_0}{\sqrt{\hat{\sigma}^2 \left( \frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right)}} > \tau_{n-2}^{1-\alpha}$
$H_0 : \beta_0 \geq b_0$ vs $H_1 : \beta_0 < b_0$	$\frac{\hat{\beta}_0 - b_0}{\sqrt{\hat{\sigma}^2 \left( \frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right)}} < \tau_{n-2}^\alpha$
$H_0 : \beta_0 = b_0$ vs $H_1 : \beta_0 \neq b_0$	$\frac{ \hat{\beta}_0 - b_0 }{\sqrt{\hat{\sigma}^2 \left( \frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right)}} > \tau_{n-2}^{1-\alpha/2}$

Dentro del modelo lineal, la varianza de los errores  $\sigma^2$  es un parámetro desconocido y por tanto también se puede llevar a cabo inferencia

Tomando en cuenta que:

$$\frac{\sum_{i=1}^n (y_i - \hat{y}_i)}{\sigma^2} \sim \chi_{n-2}^2$$

Se puede verificar que las regiones de rechazo correspondientes son:

Prueba	Región de Rechazo
$H_0 : \sigma^2 \leq s$ vs $H_1 : \sigma^2 > s$	$\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{s} > \chi_{n-2}^{2(1-\alpha)}$
$H_0 : \sigma^2 \geq s$ vs $H_1 : \sigma^2 < s$	$\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{s} < \chi_{n-2}^{2(\alpha)}$
$H_0 : \sigma^2 = s$ vs $H_1 : \sigma^2 \neq s$	$\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{s} > \chi_{n-2}^{2(1-\alpha/2)}$ o $\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{s} < \chi_{n-2}^{2(\alpha/2)}$

## Intervalos de Confianza

La estadística  $t = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}}$  es una cantidad pivotal y por tanto puede ser utilizada para construir intervalos de confianza correspondientes:

$$\mathbb{P} \left( -\tau_{n-2}^{1-\alpha/2} \leq \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}} \leq \tau_{n-2}^{1-\alpha/2} \right) = 1 - \alpha$$

$$\mathbb{P} \left( \hat{\beta}_1 - \sqrt{\frac{\hat{\sigma}^2}{S_{xx}}} \tau_{n-2}^{1-\alpha/2} \leq \beta_1 \leq \hat{\beta}_1 + \sqrt{\frac{\hat{\sigma}^2}{S_{xx}}} \tau_{n-2}^{1-\alpha/2} \right) = 1 - \alpha$$

Por lo tanto un intervalo de confianza al  $(1 - \alpha)\%$  es:

$$\left( \hat{\beta}_1 - \sqrt{\frac{\hat{\sigma}^2}{S_{xx}}} \tau_{n-2}^{1-\alpha/2}, \hat{\beta}_1 + \sqrt{\frac{\hat{\sigma}^2}{S_{xx}}} \tau_{n-2}^{1-\alpha/2} \right)$$

Donde  $\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}$  se le conoce como error estándar del estimador.

En cuanto a los otros 2 parámetros del modelo lineal simple se prueba que los intervalos de confianza para  $\beta_0$  y  $\sigma^2$  son

$$\left( \hat{\beta}_0 - \tau_{n-2}^{(1-\alpha/2)} \sqrt{\hat{\sigma}^2 \left( \frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right)}, \hat{\beta}_0 + \tau_{n-2}^{(1-\alpha/2)} \sqrt{\hat{\sigma}^2 \left( \frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right)} \right)$$

$$\left( \frac{1}{\chi_{n-2}^{2(1-\alpha/2)}} \sum_{i=1}^n (y_i - \hat{y}_i)^2, \frac{1}{\chi_{n-2}^{2(\alpha/2)}} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \right)^*$$

## Analisis Inferencial de la variable respuesta

Uno de los usos mas importantes del análisis de regresión es estimar la respuesta media para un valor particular de la variable explicativa  $X_1 = x_*$ , es decir, estimar  $\mathbb{E}(Y | X_1 = x_*)$ . Definamos a  $x_*$  como cualquier valor de la variable explicativa (**por lo general pedimos que  $x_*$  esté dentro del rango de los datos originales de  $X_1$** ), en nuestro modelo sabemos que:

$$\mathbb{E}(Y | X_1 = x_*) = \beta_0 + \beta_1 x_* \quad (4.20)$$

Como  $\beta_0$  y  $\beta_1$  son desconocidos se sigue que el valor para la respuesta media es desconocido  $\mathbb{E}(Y | X_1 = x_*)$  por lo que procedemos a estimarlo. La forma natural de estimar este valor es **enchufar** los estimadores que ya tenemos para  $\beta_0$ ,  $\beta_1$  en la expresión (4.20) obteniendo el siguiente estimador:

$$\mathbb{E}(Y | X_1 = x_*) = \hat{\beta}_0 + \hat{\beta}_1 x_*$$

Surge entonces la necesidad de verificar propiedades de nuestro estimador:

- ¿Es insegado?
- ¿Cuál es su varianza?
- ¿Qué distribución sigue?

Bajo el supuesto de normalidad de los errores, uno puede verificar que:

$$\mathbb{E}(Y | \widehat{X}_1 = x_*) \sim N \left( \beta_0 + \beta_1 x_*, \sigma^2 \left( \frac{1}{n} + \frac{(x_* - \bar{x})^2}{S_{xx}} \right) \right)$$

De donde se concluye que  $\mathbb{E}(Y | \widehat{X}_1 = x_*)$  es insesgado

Luego entonces recordando que  $\hat{\sigma}^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-2}$  y que  $\frac{(n-2)\hat{\sigma}^2}{\sigma^2} \sim \chi_{(n-2)}$  entonces podemos probar que:

$$\frac{\left( \mathbb{E}(Y | \widehat{X}_1 = x_*) - \mathbb{E}(Y | X_1 = x_*) \right)}{\sqrt{\hat{\sigma}^2 \left( \frac{1}{n} + \frac{(x_* - \bar{x})^2}{S_{xx}} \right)}} \sim t_{(n-2)}$$

La anterior es una cantidad pivotal y por tanto puede ser utilizada para crear intervalos de confianza.

El Intervalo al  $(1 - \alpha)\%$  de confianza para la respuesta media es:

$$\left( \mathbb{E}(\widehat{Y} | x_*) - \tau_{(n-2)}^{(1-\frac{\alpha}{2})} \sqrt{\hat{\sigma}^2 \left( \frac{1}{n} + \frac{(x_* - \bar{x})^2}{S_{xx}} \right)}, \mathbb{E}(\widehat{Y} | x_*) + \tau_{(n-2)}^{(1-\frac{\alpha}{2})} \sqrt{\hat{\sigma}^2 \left( \frac{1}{n} + \frac{(x_* - \bar{x})^2}{S_{xx}} \right)} \right)$$

Observaciones:

- La longitud del intervalo se minimiza si  $S_{xx}$  es grande. Obs, en este contexto  $S_{xx}$  mide la variabilidad de las variables independientes
- Otra forma de minimizar la longitud del intervalo es colocar  $x_* = \bar{x}$
- Conforme nos alejamos de  $\bar{x}$  la longitud del intervalo es mas grande, lo que se traduce a que entre mas nos alejemos de la zona de muestreo mayor sera la incertidumbre de nuestra estimación.

Ahora suponga que no nos interesa estimar a  $\mathbb{E}(Y | X_1 = x_*)$  sino que queremos estimar el valor directamente de la respuesta  $Y$  cuando  $X$  toma cierto valor  $x_0$ .

Queremos entonces una estimación para la variable respuesta  $y_0 = \beta_0 + \beta_1 x_0 + \varepsilon_0$ . Una estimación natural dadas las hipótesis que hemos venido manejando es

$$\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$$

Notemos que como  $y_0$  es una nueva observación entonces  $y_0$  es independiente de las observaciones anteriores  $(y_1, \dots, y_n)$  y como los estimadores  $\hat{\beta}_0, \hat{\beta}_1$  son combinaciones lineales de las observaciones  $(y_1, \dots, y_n)$ , se sigue entonces que  $y_0$  es independiente de  $\hat{y}_0$ , por lo tanto  $(y_0 - \hat{y}_0)$  sigue una

distribución normal. Con ayuda de lo anterior y después de unos pasos algebraicos se muestra que:

$$\frac{(y_0 - \hat{y}_0)}{\sqrt{\hat{\sigma}^2 \left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right)}} \sim t_{(n-2)}$$

Se deduce que el Intervalo de predicción al  $(1 - \alpha\%)$  para la respuesta cuando al variable independiente es una nueva observación:

$$\left( \hat{y}_0 - \tau_{(n-2)}^{(1-\frac{\alpha}{2})} \sqrt{\hat{\sigma}^2 \left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right)}, \hat{y}_0 + \tau_{(n-2)}^{(1-\frac{\alpha}{2})} \sqrt{\hat{\sigma}^2 \left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right)} \right)$$

Observaciones:

- Este intervalo es similar al encontrado anteriormente solo que agrega mas variabilidad debido a que se esta evaluando el modelo en una nueva observación que no se tenia en muestra.
- La longitud del intervalo se minimiza si  $S_{xx}$  que mide la variabilidad de las variables independientes es grande
- Otra forma de minimizar la longitud del intervalo es poner  $x_0 = \bar{x}$ .
- Conforme nos alejamos de  $\bar{x}$  la longitud del intervalo es mas grande, es decir se tiene mayor incertidumbre.
- Teóricamente con este intervalo podemos inferir el valor de la variable respuesta para cualquier valor de la variable independiente, sin embargo siempre debemos tomar en cuenta el contexto del fenómeno que estamos estudiando.
- En este caso se esta construyendo un intervalo para la variable aleatoria  $y_0$ , en el contexto de inferencia clásica al intervalo que cubre a una variable aleatoria se le conoce como **Intervalo de Predicción** y no de confianza.

## Validación del Modelo Lineal Simple

Una vez estimado el modelo de regresión lineal es importante llevar a cabo la verificación de las hipótesis que fueron impuestas dentro de la teoria, esto es, se tiene que verificar la linealidad de los datos, la normalidad de los errores, la homocedasticidad y su independencia. Es muy importante verificar todos estos supuestos previos a cualquier inferencia que se quiera hacer con el modelo calculado.

Para verificar la hipótesis de normalidad homocedasticidad y la independencia de los errores se utilizan los gráficos de los residuos, donde cada residuo, denotado por  $e_i$ , está definido como  $e_i = y_i - \hat{y}_i$ . Generalmente se gráfica  $(x_i, e_i)$ .

## Normalidad

Parte fundamental de la validación del modelo es verificar la normalidad de los residuales, para ello contamos con pruebas de bondad de ajuste, sin embargo una primera prueba se hace llevando a cabo un histograma de los residuales y verificar que siguen una distribución parecida a la normal.

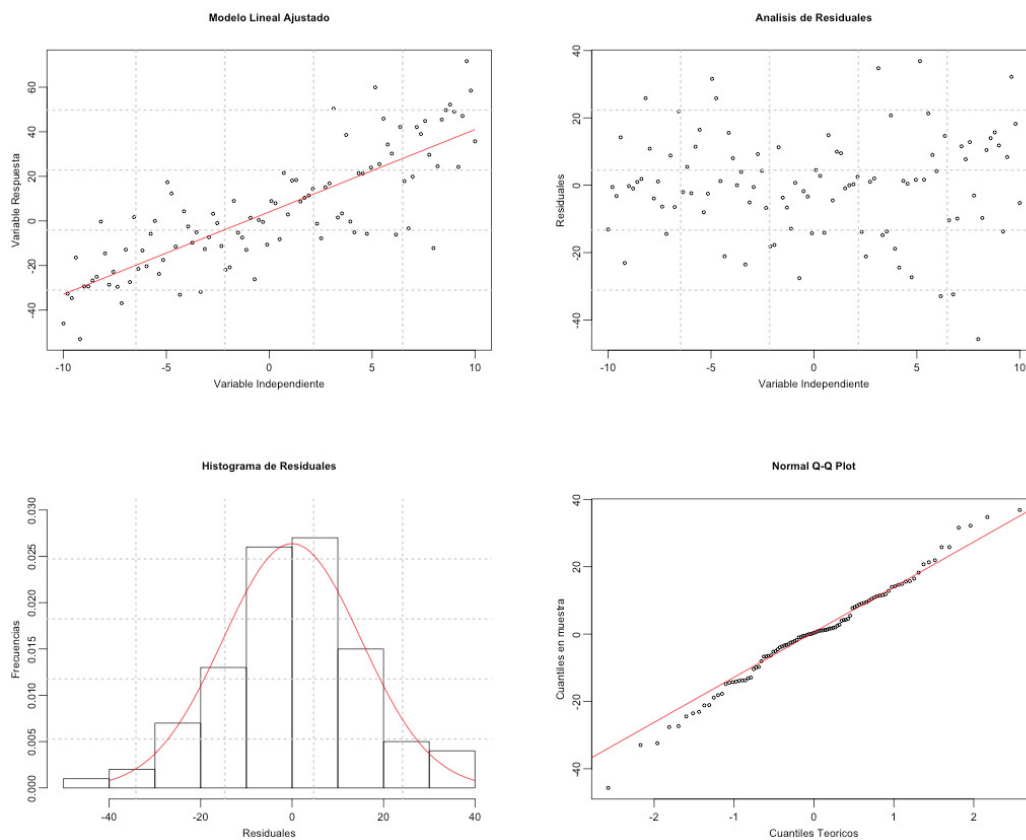


Figura 4.20: Prueba de Normalidad

Existen muchas pruebas para verificar la normalidad, las mas usadas son:

- Shapiro-Wilk Normality Test (`shapiro.test(x)`)
- Anderson-Darling test for normality (`ad.test(x)`)
- Kolmogorov-Smirnov Tests - Lilliefors (`lillie.test(x)`)

Ejemplo:

- Shapiro-Wilk normality test

data: res

W = 0.9887, p-value = 0.558

- Anderson-Darling normality test

data: res

A = 0.466, p-value = 0.2475

Nota importante: en el contexto del modelo lineal **no** podemos aplicar teóricamente las pruebas de bondad de ajuste pues se puede probar que los residuales no son independientes lo cual es un requisito indispensable para que las pruebas funcionen correctamente.

## Homocedasticidad (Igualdad de Varianzas)

Generalmente para verificar que tenemos varianza constante se llevan a cabo pruebas visuales verificando el comportamiento de los residuales conforme variamos la variable independiente

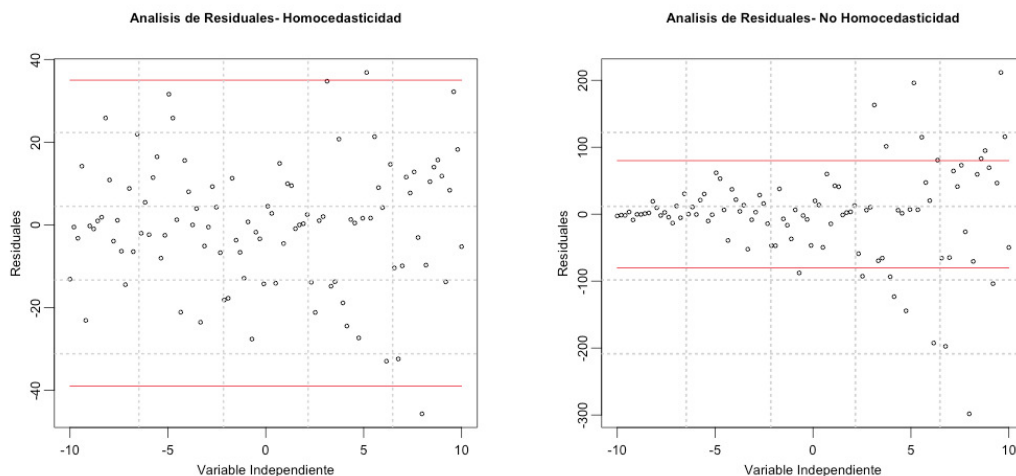


Figura 4.21: Homocedasticidad

Existen pruebas formales para llevar a cabo el contraste (Ej. Bartlett, Levene), pero requieren tener definido grupos en la variable independiente  $X$ , por lo general esta prueba es utilizada en diseño de experimentos pero podemos adaptarla al caso de regresión definiendo grupos arbitrarios, el problema que tiene esto es que deja que el analista determine el número de grupos así como el método para formarlos, lo cual puede no ser objetivo y llegar a cometer errores.

**Ejercicio 4.5.3.** Se deja como ejercicio que el alumno investigue la prueba Breusch-Pagan y encuentre el correspondiente código en R para ejecutarla



## Independencia

Generalmente para verificar que tenemos independencia de los errores se suele hacer una gráfica de los residuales contra el orden de obtención de los datos que casi siempre es contra la variable  $X_1$ . Luego en el gráfico generado se busca algún tipo de patrón que evidencie la dependencia de los errores conforme se van realizando las mediciones.

Sin embargo a veces no es fácil detectar un patrón por lo que se pueden aplicar pruebas estadísticas para detectar patrones en los datos, las pruebas más usuales son las que utilizan la función de autocorrelación (**ACF**) y la prueba no paramétrica de **Rachas** que va midiendo el cambio de los signos de los errores.

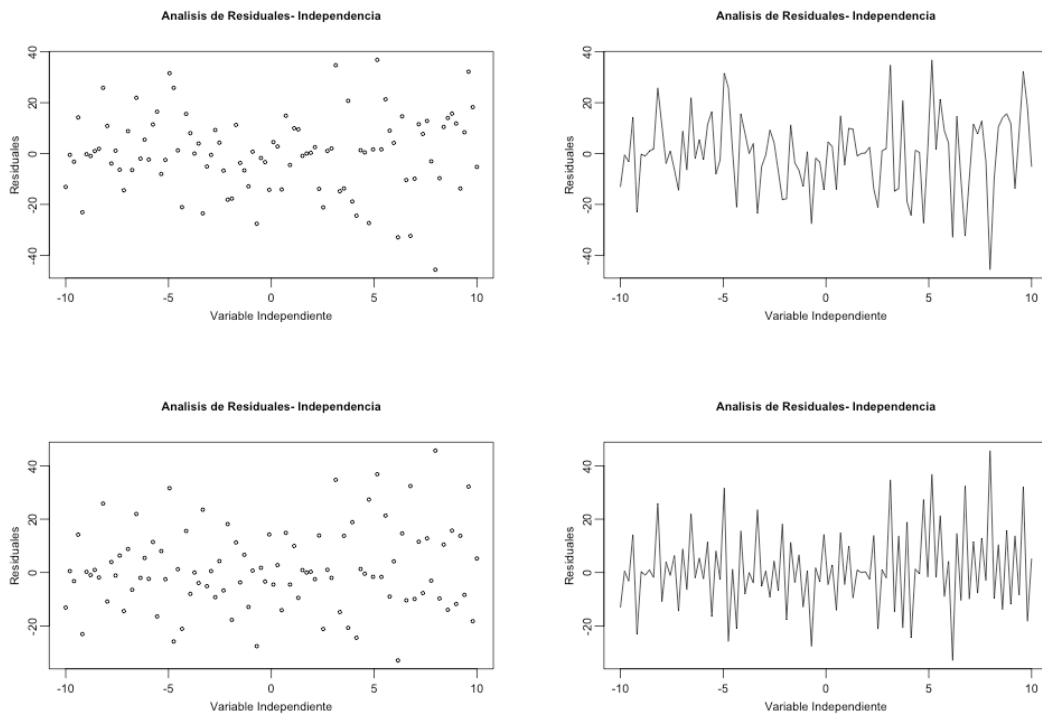


Figura 4.22: Rachas

- Función de Autocorrelación (ACF). En estadística, la autocorrelación de una serie de datos discreta de un proceso  $X_t$  no es más que el coeficiente de correlación de dicho proceso con una versión desplazada de la propia serie. La forma en como se calcula es la siguiente: suponiendo que tenemos a la serie de datos  $(e_1, e_2, \dots, e_n)$  entonces la función de autocorrelación con rezago  $k$  se obtiene como:

$$\rho_k = \frac{\sum_{i=1}^{n-k} (e_i - \bar{e})(e_{i-k} - \bar{e})}{\sum_{i=1}^n (e_i - \bar{e})^2}$$

Si  $\rho_k$  se aleja del valor 0 hay evidencia de que cada  $k$  observaciones hay un patrón de los residuales pues dichas observaciones están muy correlacionadas lo que implicaría dependencia de los residuales. La idea entonces es contrastar la hipótesis  $\rho_k = 0$  vs  $\rho_k \neq 0$ , afortunadamente

R tiene programada la prueba en la función `acf`

Otra forma que tenemos de verificar que los residuales son independientes es probar la aleatoriedad con la que van cambiando de signo los errores, si los residuales son independientes se esperaría que el cambio de signo del residual conforme se va obteniendo la muestra es aleatorio.

Se considera una Racha de tamaño  $k$  a la secuencia de  $k$  de valores consecutivos de un mismo signo siempre y cuando estos sean precedidos y seguidos por valores con signo opuesto a la de la Racha.

$\underbrace{+, +}, \underbrace{-, -}, \underbrace{+, +}, \underbrace{-, -}$

La idea de la prueba es contar el número de rachas en la muestra, luego un número reducido o grande de rachas es indicio de que las observaciones no se han obtenido de forma aleatoria .

Si la muestra es grande y la hipótesis de aleatoriedad es cierta la distribución muestral del número de rachas  $R$ , puede aproximarse mediante una distribución Normal de parámetros:

$$\mu_R = \frac{2n_1n_2}{n} \quad \sigma_R^2 = \frac{2n_1n_2(2n_1n_2 - n)}{n^2(n - 1)}$$

La prueba de Rachas se encuentra programada de R y se encuentra dentro de la librería `tseries` bajo el nombre `runs.test`, esta función recibe un vector con valores binarios indicando el signo del residual.

Ejemplo:

```
library(tseries)
runs.test(as.factor(res>0))

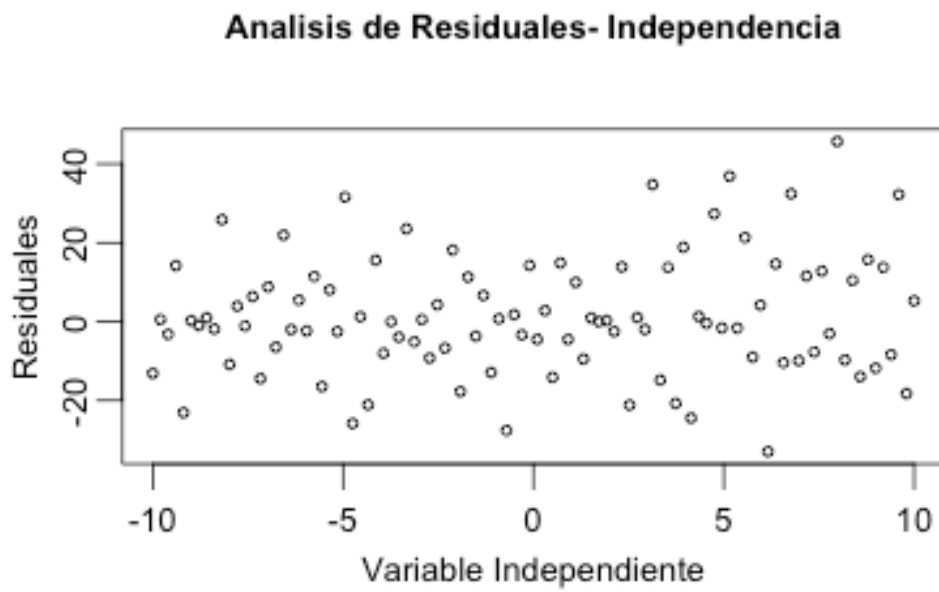
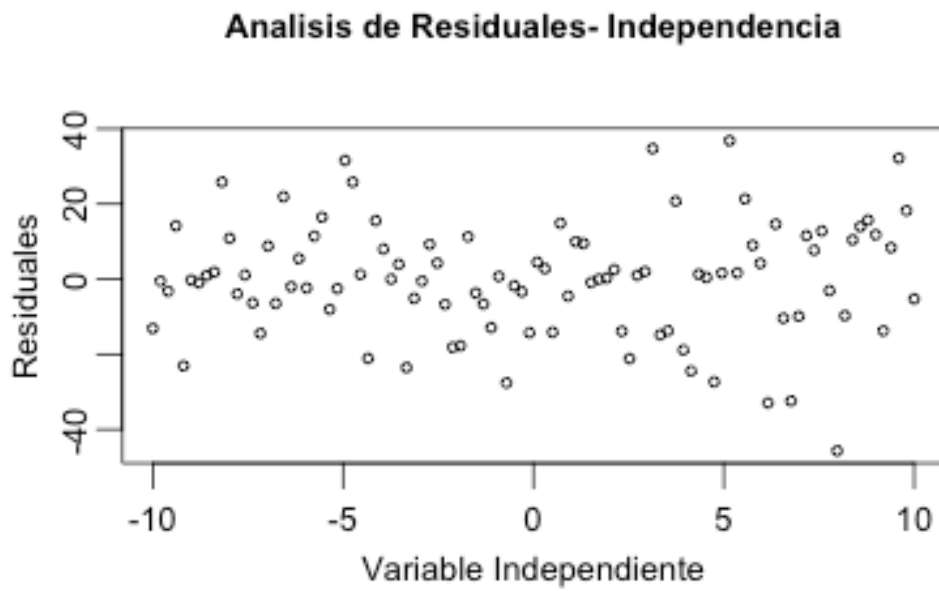
Runs Test

data:  as.factor(res > 0)
Standard Normal = 0.004, p-value = 0.9968
alternative hypothesis: two.sided

library(tseries)
runs.test(as.factor(res2>0))

Runs Test

data:  as.factor(res2 > 0)\
Standard Normal = 9.8499, p-value < 2.2e-16\
alternative hypothesis: two.sided\
```



H

Figura 4.23: Prueba de Racahas en Residuales

## Ejercicios

**Ejercicio 4.5.4.** Sea  $\hat{\beta}_1$  y  $\hat{\beta}_0$  los estimadores Máximo Verosimilias encontrados en el modelo de regresión lineal simple, demuestre que:

$$\mathbb{E}(\hat{\beta}_0) = \beta_0 \quad \text{Var}(\hat{\beta}_0) = \left( \frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right) \sigma^2$$

**Ejercicio 4.5.5.** Se tiene el modelo de regresión lineal simple  $y = \beta_0 + \beta_1 x + \varepsilon$ , con  $\mathbb{E}(\varepsilon_i) = 0$ ,  $\text{Var}(\varepsilon_i) = \sigma^2$  y tal que  $\text{cov}(\varepsilon_i, \varepsilon_j) = 0$  con  $i \neq j$ . Demostrar que:

$$\begin{aligned} \sum_{i=1}^n y_i &= \sum_{i=1}^n \hat{y}_i \\ \sum_{i=1}^n x_i e_i &= 0 = \sum_{i=1}^n \hat{y}_i e_i \\ \text{Cov}(\hat{\beta}_0, \hat{\beta}_1) &= \frac{-\bar{x}\sigma^2}{S_{xx}} \\ \text{Cov}(\bar{y}, \hat{\beta}_1) &= 0 \\ \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 &= \hat{\beta}_1^2 S_{xx} \\ \mathbb{E} \left( \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \right) &= \sigma^2 + \beta_1^2 S_{xx} \end{aligned}$$

Donde  $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$  y  $e_i$  es el residual  $i$ , es decir  $e_i = y_i - \hat{y}_i$

**Ejercicio 4.5.6.** Se tiene el modelo de regresión lineal simple:

$$y = \beta_0 + \beta_1 x + \varepsilon$$

Suponga que  $\beta_0$  es conocido, observe entonces que el único parámetro a estimar es  $\beta_1$

- Determinar el estimador de  $\beta_1$  por mínimos cuadrados para este modelo.
- ¿Cuál es la varianza del estimador de la pendiente que encontró en el punto anterior? Compare esta varianza con la varianza del estimador  $\hat{\beta}_1$  cuando el valor de  $\beta_0$  es desconocido. ¿Cuál tiene menor varianza?

**Ejercicio 4.5.7. Regresión por el origen** Suponga que se tiene el siguiente modelo lineal:

$$y = \beta_1 x + \varepsilon \quad \varepsilon \sim N(0, \sigma^2)$$

Es decir, el modelo asume que  $\beta_0 = 0$  por lo que la recta ajustada siempre pasa por el origen.

- Encuentre el estimador Máximo Verosímil para  $\beta_1$  y  $\sigma^2$
- Calcule la esperanza y varianza del estimador de  $\beta_1$
- ¿Cómo se distribuye el estimador para  $\beta_1$  y con qué parámetros?

**Ejercicio 4.5.8.** En el modelo de regresión lineal se define el coeficiente de determinación como  $R^2 = \frac{SCR}{SCT}$ , por otro lado se define el **Coefficiente de Correlación de Pearson** como :

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$$

Demuestre entonces que:

$$r^2 = R^2$$

**Ejercicio 4.5.9. Intervalo de Confianza para  $\beta_0$ .** Considere el estimador máximo verosímil para  $\beta_0$ .

- Demuestre que:  $\hat{\beta}_0 \sim N\left(\beta_0, \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right) \sigma^2\right)$
- Asumiendo el hecho que  $\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sigma^2} = \frac{(n-2)}{\sigma^2} \hat{\sigma}^2 \sim \chi_{n-2}^2$ , demuestre que:

$$\frac{\hat{\beta}_0 - \beta_0}{\sqrt{\hat{\sigma}^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)}} \sim \tau_{(n-2)}$$

- Observe que lo anterior es una cantidad pivotal, muestre entonces que un intervalo al  $100(1 - \alpha)\%$  de confianza para  $\beta_0$  es

$$\left( \hat{\beta}_0 - \tau_{n-2}^{(1-\alpha/2)} \sqrt{\hat{\sigma}^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)}, \hat{\beta}_0 + \tau_{n-2}^{(1-\alpha/2)} \sqrt{\hat{\sigma}^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)} \right)$$

**Ejercicio 4.5.10. Intervalo de Confianza para  $\sigma^2$ .** Dentro del modelo lineal simple, la varianza de los errores  $\sigma^2$  es un parámetro desconocido y por tanto también se puede llevar a cabo inferencia sobre él.

Tomando en cuenta que:

$$\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sigma^2} = \frac{(n-2)}{\sigma^2} \hat{\sigma}^2 \sim \chi_{n-2}^2$$

Demuestre entonces que un intervalo al  $100(1 - \alpha)\%$  de confianza para  $\sigma^2$  es:

$$\left( \frac{1}{\chi_{n-2}^{2(1-\alpha/2)}} \sum_{i=1}^n (y_i - \hat{y}_i)^2, \frac{1}{\chi_{n-2}^{2(\alpha/2)}} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \right)^*$$

**Ejercicio 4.5.11.** Una empresa dedicada a la venta de equipo deportivo desea modelar su ventas mensuales para pronosticar las ventas del siguiente mes. (La base de sus ventas se puede descargar desde este link)

- Ajuste un modelo lineal de la forma

$$y = \beta_0 + \beta_1 x + \varepsilon$$

y calcule la suma de los cuadrados de los errores.

- Se sabe que la venta de artículos deportivos tiene una periodicidad de 2 años (48 meses) por lo que se propone ajustar un modelo de la forma de la forma:

$$y = \beta_0 + \beta_1 \sin\left(\frac{2\pi}{48}x\right) + \beta_2 \cos\left(\frac{2\pi}{48}x\right) + \varepsilon$$

Ajuste el modelo por medio de mínimos cuadrados haciendo uso de la función `nlm()`. Una vez ajustado calcule la suma de los cuadrados de los errores

- De los dos modelos cuál elegiría y cuál sería su pronóstico de la venta del siguiente mes.

**Ejercicio 4.5.12.** La tabla `FootballLeague.csv` contiene los datos sobre el desempeño de los equipos de la liga nacional de fútbol de E.U.A. durante 1976. Se sospecha que el número de yardas ganadas por los oponentes (Variable  $x_8$ ) tiene un efecto sobre el número de juegos ganados (Variable  $y$ ).

- Ajuste un modelo lineal simple que relacione el número de juegos ganados con el número de yardas ganadas por los oponentes, es decir ajuste el modelo:

$$y_i = \beta_0 + \beta_1 x_8 + \varepsilon$$

- Construya el intervalo de confianza ( $\alpha=0.05$ ) para  $\beta_1$ . Verifique si el 0 está en el intervalo construido, dada la información obtenida ¿ Puede asegurar que el número de yardas ganadas tiene un efecto sobre el número de juegos ganados?
- Construya el intervalo de confianza ( $\alpha=0.05$ ) para  $\beta_0$ .
- Bajo el contexto de este problema, ¿Cuál sería la interpretación que se le daría a  $\beta_0$ ?
- Construya un intervalo de confianza ( $\alpha=0.05$ ) para  $\sigma^2$ . (Punto Extra. Encuentre el intervalo de confianza con menor longitud)
- Construya la tabla ANOVA correspondiente y lleve a cabo la prueba de hipótesis

$$H_0 : \beta_1 = 0 \quad vs \quad H_1 : \beta_1 \neq 0$$

Obtenga el  $p$ -value de la prueba. Tomando ( $\alpha = 0.05$ ) ¿Rechazaría  $H_0$ ?

- ¿Qué porcentaje de la variabilidad total de  $y$  es explicada por el modelo que ajustó?
- Encuentra un intervalo al 95 % de confianza para  $\mathbb{E}(Y \mid X = 2000)$ .
- Encuentra un intervalo al 99 % de confianza para el número de juegos ganados para un equipo que logró que sus oponentes ganasen 2072 yardas

Del modelo ajustado en el punto anterior lleve a cabo un análisis de los residuales, es decir verifique lo siguiente:

- Normalidad de los Residuales (Bondad de Ajuste)
- Homocedasticidad
- Independencia (ACF, Prueba de Rachas)

**Ejercicio 4.5.13.** La siguiente tabla ANOVA fue obtenida por un paquete estadístico:

```
anova(lm(y~x))\
Analysis of Variance Table\
```

*Response: y*

<i>Fnte. de Var.</i>	<i>Df.</i>	<i>Sum Sq</i>	<i>Mean Sq</i>	<i>F value</i>	<i>Pr(&gt;F)</i>
<i>x</i>	1	20.107	20.1069	4.1248	0.05649
<i>Residuals</i>	19	92.618	4.8746		
<i>Total</i>	1	4			

Se sabe además que  $S_{xx} = 770$ . Responda lo siguiente:

- ¿Con cuántas observaciones se hizo el ajuste?
- Tomando  $\alpha = 0.05$ . ¿Rechazaría la hipótesis  $H_0 : \beta_1 = 0$ ?
- De un estimador insesgado para  $\sigma^2$
- Construya un intervalo de confianza al 95 % para  $\sigma^2$
- ¿Cuanto vale  $S_{yy} := \sum_{i=1}^n (y_i - \bar{y})^2$ ?
- De un estimador para  $|\beta_1|$
- Calcule el error estándar del estimador de  $\beta_1$
- ¿Qué porcentaje de la variabilidad es explicada por el modelo?

**Ejercicio 4.5.14.** Considere el modelo de regresión lineal  $y = 50 + 10x + \varepsilon$ , donde  $\varepsilon \sim N(0, \sigma^2 = 16)$ . Genere 1 000 muestras de  $n = 19$  observaciones cada una tal que  $x \in \{1, 1.5, 2, 2.5, 3, \dots, 9.5, 10\}$ .

- Para cada muestra calcule los estimadores máximo verosímiles  $(\hat{\beta}_0, \hat{\beta}_1)$  de tal forma que se obtengan 1 000 estimaciones para  $\beta_0$  y 1 000 estimaciones para  $\beta_1$  y elabore un par de histogramas. ¿Los histogramas que obtiene son congruentes con la teoría que indica que ambos estimadores siguen una distribución normal?
- Para cada muestra calcule un intervalo al 95 % de confianza para  $\beta_1$ . ¿Cuántos de estos intervalos contiene al verdadero valor  $\beta_1 = 10$ ?
- Para cada muestra, encuentre el estimador de  $\mathbb{E}(y | x = 5)$ . Construya un histograma. ¿El histograma que obtiene es congruente con la teoría que indica que el estimador siguen una distribución normal?
- Para cada muestra construya un intervalo de confianza a

## 4.5.2. Análisis de Regresión Múltiple

Ahora nos enfrentamos al problema de tener  $k$  funciones de las covariables y una variable respuesta. Entonces se plantea ajustar un modelo lineal de la forma:

$$Y_i = \beta_0 + \beta_1 s_1(\underline{z}_i) + \beta_2 s_2(\underline{z}_i) + \dots + \beta_k s_k(\underline{z}_i) + \varepsilon_i$$

Donde definiendo:

$$x_{ij} = s_j(\underline{z}_i) \quad i \in \{1, \dots, n\}; \quad j \in \{1, \dots, k\}$$

se transforma en el modelo:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \varepsilon_i$$

Donde  $\varepsilon_i$  una variable aleatoria que representa el error en la relación.  $\beta_i$  son parámetros desconocidos en el modelo y tales que

$$\varepsilon_i \sim F(0, \sigma^2) \quad \mathbb{E}(\varepsilon_i) = 0 \quad \text{var}(\varepsilon_i) = \sigma^2;$$

Obs: Note que las  $k$  funciones de las covariables  $\underline{z}$  dan origen a  $k$  variables que denotamos por  $x$ , de ahí que digamos que se ajusta un modelo con  $k$  variables.

Supongamos entonces que observaremos  $n$  observaciones de las  $k$  funciones de las variables



explicativas con su respectiva variable respuesta, es decir:

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \dots + \beta_k x_{1k} + \varepsilon_1 \\ y_2 &= \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \dots + \beta_k x_{2k} + \varepsilon_2 \\ &\vdots \\ y_n &= \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \dots + \beta_k x_{nk} + \varepsilon_n \end{aligned}$$

En nuestro modelo general nuevamente impondremos la hipótesis de no correlación entre los errores, es decir  $Cov(\varepsilon_i, \varepsilon_j) = 0$  con  $i \neq j$ . Se plantea el problema entonces de encontrar los  $k+2$  parámetros asociados al modelo lineal:

$$(\beta_0, \beta_1, \dots, \beta_k, \sigma^2)$$

Para facilitar notación y dar una solución mas elegante al problema se establece la siguiente notación matricial. ( $p = k + 1$ )

$$\underline{Y} = \mathbf{X}\underline{\beta} + \underline{\varepsilon}$$

Donde:

$$\underline{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}_{n \times 1} \quad \mathbf{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix}_{n \times p} \quad \underline{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}_{n \times 1}$$

El vector de parámetros:

$$\underline{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}_{p \times 1}$$

A continuación se discuten los métodos de estimación para este modelo general.

### Estimación por mínimos cuadrados

El objetivo de esta técnica es minimizar la diferencia al cuadrado entre los valores reales y los ajustados por el modelo, para ello definamos el vector de valores ajustados como:

$$\underline{\hat{Y}} = \mathbf{X}\underline{\hat{\beta}}$$

Luego entonces el objetivo es minimizar respecto a  $\underline{\hat{\beta}}$  a la siguiente expresión:

$$\begin{aligned}
f(\underline{\hat{\beta}}) &= \sum_{i=1}^n (y_i - \hat{y}_i)^2 = (\underline{Y} - \underline{\hat{Y}})^t (\underline{Y} - \underline{\hat{Y}}) = (\underline{Y} - \mathbf{X}\underline{\hat{\beta}})^t (\underline{Y} - \mathbf{X}\underline{\hat{\beta}}) \\
&= \underline{Y}^t \underline{Y} - 2\underline{\hat{\beta}}^t \mathbf{X}^t \underline{Y} + \underline{\hat{\beta}}^t \mathbf{X}^t \mathbf{X} \underline{\hat{\beta}}
\end{aligned}$$

Utilizando técnicas de cálculo vectorial se prueba que el estimador por mínimos cuadrados se obtiene al solucionar la denominada ecuación normal:

$$\mathbf{X}^t \mathbf{X} \underline{\hat{\beta}} = \mathbf{X}^t \underline{Y} \quad (4.21)$$

$$\underline{\hat{\beta}} = (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \underline{Y} \quad (4.22)$$

Debe notarse que los estimadores  $\hat{\beta}_i$  son combinaciones lineales de las variables respuestas y que se tiene solución única a la ecuación normal cuando la matriz  $\mathbf{X}\mathbf{X}^t$  tiene inversa lo cual ocurre si y sólo si  $\mathbf{X}$  tiene rango completo. (Tarea: Investigar cuando una matriz tiene rango completo) lo que equivale a pedir que las variables  $(X_1, X_2, \dots, X_k)$  son linealmente independientes. Esto impone una nueva hipótesis al modelo lineal, la cual nos indica que una variable no puede ser combinación lineal de otras, en caso de que esto ocurra, se deberá analizar el modelo y reducir el número de variables.

Se prueba entonces que el estimador por mínimos cuadrados está dado por  $\underline{\hat{\beta}} = (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \underline{Y}$ , luego como  $\underline{\hat{Y}} = \mathbf{X} \underline{\hat{\beta}}$  entonces:

$$\underline{\hat{Y}} = \mathbf{X} (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \underline{Y}$$

Si se define:

$$\mathbf{H} = \mathbf{X} (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t$$

Entonces se tiene que  $\underline{\hat{Y}} = \mathbf{H}\underline{Y}$ . En la literatura a la matriz  $\mathbf{H}$  se le conoce como la matriz sombrero (hat) y tiene la propiedad de mapear al vector de valores observados en el vector de valores ajustados. Algunos resultados importantes del modelo requieren tener conocimiento de vectores aleatorios y álgebra matricial por lo que se invita al lector estudiar la sección de temas especiales de regresión.

- $\mathbf{H}$  es simétrica
- $\mathbf{H}$  es idempotente ( $\mathbf{H}\mathbf{H} = \mathbf{H}^2 = \mathbf{H}$ )
- El vector de residuales  $\underline{e}$  pueden expresarse como  $\underline{e} = (\mathbf{I} - \mathbf{H})\underline{Y}$
- $\mathbb{E}(\underline{Y}) = \mathbb{E}(\mathbf{X}\underline{\beta} + \underline{\varepsilon}) = \mathbf{X}\underline{\beta}$

- $\text{Var}(\underline{Y}) = \text{Var}(\mathbf{X}\underline{\beta} + \underline{\varepsilon}) = \text{Var}(\underline{\varepsilon}) = \mathbf{I}\sigma^2$

- (Insesgamiento de  $\hat{\underline{\beta}}$ ):

$$\begin{aligned}\mathbb{E}(\hat{\underline{\beta}}) &= \mathbb{E}\left((\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\underline{Y}\right) = (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbb{E}(Y) \\ &= (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{X}\underline{\beta} = \underline{\beta}\end{aligned}$$

- (Matriz de Varianzas- Covarianzas para  $\hat{\underline{\beta}}$ ):

$$\begin{aligned}\text{Var}(\hat{\underline{\beta}}) &= \text{Var}\left((\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\underline{Y}\right) \\ &= (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\text{Var}(\underline{Y})\left((\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\right)^t \\ &= (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{I}\sigma^2\mathbf{X}\left(\mathbf{X}^t\mathbf{X}\right)^{-1} \\ &= \sigma^2(\mathbf{X}^t\mathbf{X})^{-1}\end{aligned}$$

Observe que si  $\mathbf{C} = (\mathbf{X}^t\mathbf{X})^{-1}$  entonces  $\text{Cov}(\hat{\beta}_i, \hat{\beta}_j) = \sigma^2 c_{(i+1),(j+1)}$  donde  $c_{(i+1),(j+1)}$  es el elemento  $(i+1), (j+1)$  de la matriz  $\mathbf{C}$ , con  $i, j \in \{0, 1, \dots, k\}$

Teoricamente se puede mostrar ademas que el vector de estimadores  $\hat{\underline{\beta}}$  cumple con la condicion de ser mejor estimador lineal insesgado. (Teorema de Gauss-Markov)

## Estimación máxima verosimilitud

Hasta ahora las hipótesis que hemos utilizado en el modelo lineal son:

- La matriz  $\mathbf{X}$  es de rango completo  $\text{Rank}(\mathbf{X}) = p$
- $\varepsilon_i$  es variable aleatoria con distribución  $F$  para toda  $i$
- $\mathbb{E}(\varepsilon_i) = 0$
- $\text{Var}(\varepsilon_i) = \sigma^2$  (homocedasticidad)
- $\text{Cov}(\varepsilon_i, \varepsilon_j) = 0 \quad i \neq j$

Ahora impondremos una condición fundamental para el desarrollo de toda la teoría inferencial.

- $\underline{\varepsilon}$  sigue una distribución Normal Multivariante.

Consecuencias:

- $\varepsilon_i \sim N(0, \sigma^2)$
- $\varepsilon_i$  es independiente de  $\varepsilon_j$
- $\underline{\varepsilon} \sim N_n(\underline{0}, \sigma^2\mathbf{I}_n)$

- Como  $\underline{Y} = \mathbf{X}\underline{\beta} + \underline{\varepsilon}$  entonces  $\underline{Y} \sim N_n(\mathbf{X}\underline{\beta}, \sigma^2\mathbf{I}_n)$

Dado que ahora conocemos la forma de distribución de  $\underline{Y}$ , podemos encontrar la función de verosimilitud

$$\underline{Y} \sim N_n(\mathbf{X}\underline{\beta}, \sigma^2\mathbf{I}_n) \Rightarrow f(\underline{Y}; \underline{\beta}, \sigma^2) = \frac{1}{(2\pi)^{\frac{n}{2}} |\sigma^2\mathbf{I}_n|^{\frac{1}{2}}} e^{-\frac{1}{2}(\underline{Y}-\mathbf{X}\underline{\beta})^t (\sigma^2\mathbf{I}_n)^{-1} (\underline{Y}-\mathbf{X}\underline{\beta})}$$

Como  $|\sigma^2\mathbf{I}_n| = \sigma^{2n}$  y  $(\sigma^2\mathbf{I}_n)^{-1} = \frac{1}{\sigma^2}\mathbf{I}_n$  entonces la verosimilitud es:

$$\mathcal{L}(\underline{\beta}, \sigma^2; \underline{Y}) = (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}(\underline{Y}-\mathbf{X}\underline{\beta})^t (\underline{Y}-\mathbf{X}\underline{\beta})}$$

Maximizando con respecto a  $(\underline{\beta}, \sigma^2)$  se puede concluir que:

$$\hat{\underline{\beta}}_{MV} = (\mathbf{X}^t\mathbf{X})^{-1} \mathbf{X}^t\underline{Y}$$

La ventaja de la estimación por máxima verosimilitud es que nos permite encontrar un estimador para el parámetro desconocido  $\sigma^2$

$$\hat{\sigma}_{MV}^2 = \frac{(\underline{Y} - \mathbf{X}\hat{\underline{\beta}}_{MV})^t (\underline{Y} - \mathbf{X}\hat{\underline{\beta}}_{MV})}{n}$$

Definiendo  $\hat{\underline{Y}} := \mathbf{X}\hat{\underline{\beta}}_{MV}$  entonces:

$$\hat{\sigma}_{MV}^2 = \frac{(\underline{Y} - \hat{\underline{Y}})^t (\underline{Y} - \hat{\underline{Y}})}{n} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Dado que el estimador máximo verosímil para  $\underline{\beta}$  coincide con el estimador por mínimos cuadrados entonces:

$$\mathbb{E}(\hat{\underline{\beta}}_{MV}) = \underline{\beta} \quad \text{Var}(\hat{\underline{\beta}}_{MV}) = \sigma^2 (\mathbf{X}^t\mathbf{X})^{-1}$$

Luego como  $\underline{Y} \sim N_n(\mathbf{X}\underline{\beta}, \sigma^2\mathbf{I}_n)$  y como  $\hat{\underline{\beta}}_{MV} = (\mathbf{X}^t\mathbf{X})^{-1} \mathbf{X}^t\underline{Y}$  se sigue que:

$$\hat{\underline{\beta}}_{MV} \sim N_p(\underline{\beta}, \sigma^2 (\mathbf{X}^t\mathbf{X})^{-1})$$

Una propiedad que tiene este estimador es que es sesgado

$$\begin{aligned}\mathbb{E}(\hat{\sigma}_{MV}^2) &= \mathbb{E}\left(\frac{(\underline{Y} - \hat{\underline{Y}})^t (\underline{Y} - \hat{\underline{Y}})}{n}\right) = \frac{\sigma^2}{n} \mathbb{E}\left(\frac{(\underline{Y} - \hat{\underline{Y}})^t (\underline{Y} - \hat{\underline{Y}})}{\sigma^2}\right) \\ &= \sigma^2 \frac{n-p}{n}\end{aligned}$$

Sin embargo a partir de la expresion anterior facilmente se puede encontrar un estimador insesgado para  $\sigma^2$

$$\hat{\sigma}^2 = \frac{n}{n-p} \hat{\sigma}_{MV}^2 = \frac{(\underline{Y} - \hat{\underline{Y}})^t (\underline{Y} - \hat{\underline{Y}})}{n-p} = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-p}$$

## Análisis Inferencial de los parámetros

Hasta este momento hemos encontrado los estimadores puntuales del modelo lineal multiple, a continuación se presenta la teoria respectiva para hacer inferencia correspondiente.

## Pruebas de Hipótesis

En el modelo de regresión lineal multiple podemos llevar a cabo pruebas de hipótesis para los parámetros de forma conjunta o individual, en esta sección se resuelve el caso individual.

El objetivo es contrastar hipótesis de la forma:

$$H_0 : \beta_i = b_i \quad vs \quad H_1 : \beta_i \neq b_i; \quad H_0 : \beta_i \leq b_i \quad vs \quad H_1 : \beta_i > b_i$$

$$H_0 : \beta_i \geq b_i \quad vs \quad H_1 : \beta_i < b_i \quad i \in \{0, 1, \dots, k\}$$

Como  $\hat{\underline{\beta}}_{MV} \sim N_p(\underline{\beta}, \sigma^2 (\mathbf{X}^t \mathbf{X})^{-1})$ , entonces haciendo  $\mathbf{C} = (\mathbf{X}^t \mathbf{X})^{-1}$  y definiendo  $C_{ij}$  al elemento  $(i, j)$  de la matriz  $\mathbf{C} \in \mathbb{R}^{p \times p}$  se tiene que:

$$\hat{\beta}_i \sim N(\beta_i, \sigma^2 C_{(i+1)(i+1)}) \Rightarrow \frac{\hat{\beta}_i - \beta_i}{\sqrt{\sigma^2 C_{(i+1)(i+1)}}} \sim N(0, 1) \quad i \in \{0, \dots, k\}$$

De donde es posible mostrar que:

$$t = \frac{\hat{\beta}_i - \beta_i}{\sqrt{\hat{\sigma}^2 C_{(i+1)(i+1)}}} \sim \tau_{n-p} \quad i \in \{0, \dots, k\} \quad (4.23)$$

La expresión (4.23) es muy útil pues a través de ella se obtienen las regiones de rechazo para las pruebas individuales como se muestra a continuación.

$$H_0 : \beta_i = b_i \quad vs \quad H_1 : \beta_i \neq b_i$$

La regla es rechazar  $H_0$  cuando  $|t| \geq \tau_{n-p}^{1-\alpha/2}$ , donde  $t = \frac{\hat{\beta}_i - b_i}{\sqrt{\hat{\sigma}^2 C_{(i+1)(i+1)}}$

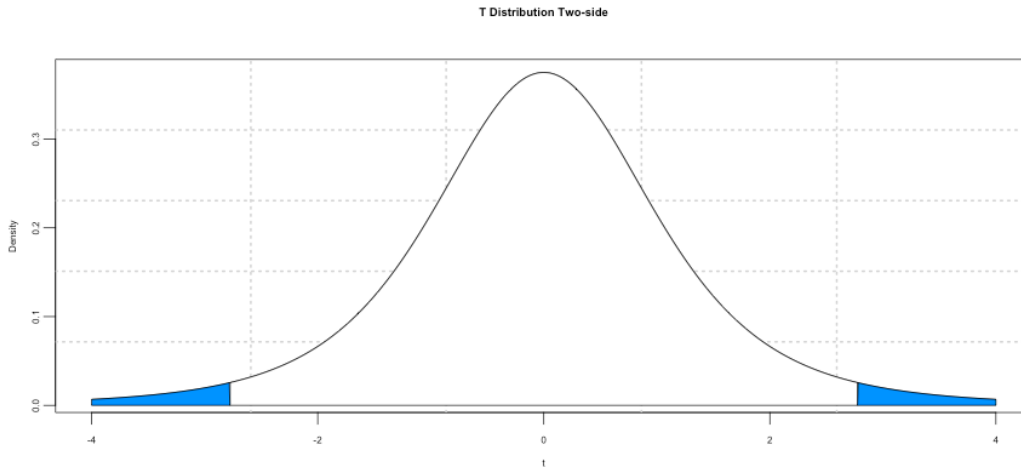


Figura 4.24: Región de Rechazo

**Importante:** Esta es la prueba de hipótesis mas importante dentro del análisis de regresión lineal pues cuando hacemos  $b_i = 0$  nos ayuda a determinar si la variable asociada a ese coeficiente es *estadísticamente* diferente de cero lo que se traduce en verificar si hay un efecto de la variable  $x_i$  en la variable  $Y$ .

Cuando se plantea la hipótesis:

$$H_0 : \beta_i \leq b_i \quad vs \quad H_1 : \beta_i > b_i$$

La regla es rechazar  $H_0$  cuando  $t \geq \tau_{n-p}^{1-\alpha}$ , donde

$$t = \frac{\hat{\beta}_i - b_i}{\sqrt{\hat{\sigma}^2 C_{(i+1)(i+1)}} \quad ;$$

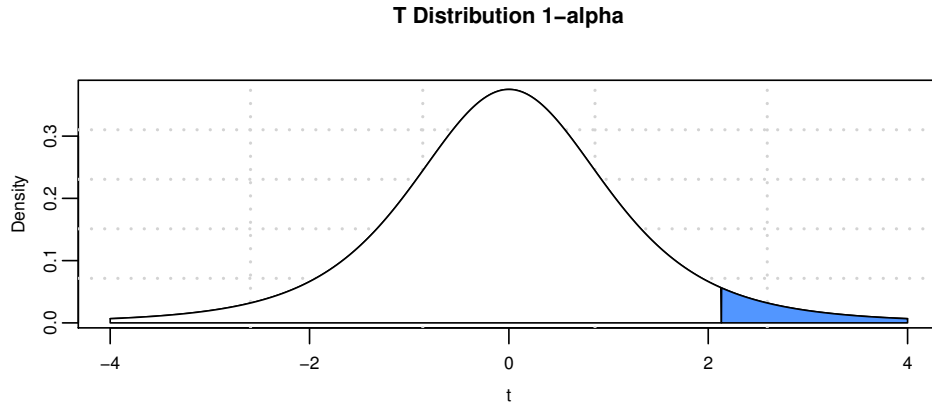


Figura 4.25: Región de Rechazo

Cuando se plantea la hipótesis:

$$H_0 : \beta_i \geq b_i \quad vs \quad H_1 : \beta_i < b_i$$

La regla es rechazar  $H_0$  cuando  $t \leq \tau_{n-p}^\alpha$ , donde

$$t = \frac{\hat{\beta}_i - b_i}{\sqrt{\hat{\sigma}^2 C_{(i+1)(i+1)}}} \quad ;$$

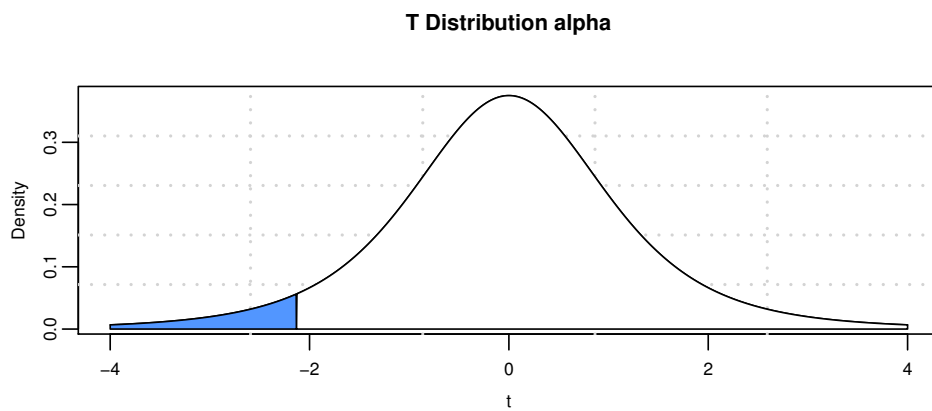


Figura 4.26: Región de Rechazo

En el modelo lineal múltiple teóricamente también podemos hacer inferencia para el parámetro

$\sigma^2$ , en cuyo caso nos apoyaremos fuertemente en el siguiente resultado:

$$\frac{(\underline{Y} - \hat{\underline{Y}})^t (\underline{Y} - \hat{\underline{Y}})}{\sigma^2} = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sigma^2} \sim \chi_{n-p}^2$$

Con ayuda del resultado anterior se puede determinar las siguientes regiones de rechazo:

Prueba	Región de Rechazo
$H_0 : \sigma^2 \leq s$ vs $H_1 : \sigma^2 > s$	$\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{s} > \chi_{n-p}^{2(1-\alpha)}$
$H_0 : \sigma^2 \geq s$ vs $H_1 : \sigma^2 < s$	$\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{s} < \chi_{n-p}^{2(\alpha)}$
$H_0 : \sigma^2 = s$ vs $H_1 : \sigma^2 \neq s$	$\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{s} > \chi_{n-p}^{2(1-\alpha/2)}$ o $\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{s} < \chi_{n-p}^{2(\alpha/2)}$

Un problema que tienen las pruebas a cada coeficiente es que da por hecho que los restantes coeficientes están en el modelo, surge entonces la necesidad de desarrollar una teoría para pruebas de hipótesis de un subconjunto de coeficientes. Consideremos el modelo lineal general:

$$\underline{Y} = \mathbf{X}\underline{\beta} + \underline{\varepsilon}$$

Queremos determinar si un subconjunto de tamaño  $r < p$  contribuye en la regresión, para ello particionemos al vector de coeficientes y la matriz  $\mathbf{X}$  como sigue:

$$\underline{\beta} = \begin{pmatrix} \underline{\beta}_1 \\ \underline{\beta}_2 \end{pmatrix}_{p \times 1} \quad \mathbf{X} = \begin{pmatrix} \mathbf{X}_1 & \mathbf{X}_2 \end{pmatrix}_{(n \times k)}$$

Donde  $\underline{\beta}_1$  es un vector con  $p - r$  entradas y  $\underline{\beta}_2$  es un vector con  $r$  entradas mientras que la partición de la matriz  $\mathbf{X}$  hace que  $\mathbf{X}_1$  sea una sub-matriz de  $n \times (p - r)$  mientras que  $\mathbf{X}_2$  es una sub-matriz de  $n \times r$ . Con esto el modelo lineal general puede expresarse como:

$$\underline{Y} = \mathbf{X}_1 \underline{\beta}_1 + \mathbf{X}_2 \underline{\beta}_2 + \underline{\varepsilon}$$

Nos interesa una prueba de hipótesis de la forma:

$$H_0 : \underline{\beta}_2 = \underline{0} \quad vs \quad H_1 : \underline{\beta}_2 \neq \underline{0}$$

En términos de modelos estamos haciendo una prueba como sigue:

$$H_0 : \underline{Y} = \mathbf{X}_1 \underline{\beta}_1 + \underline{\varepsilon} \quad vs \quad H_1 : \underline{Y} = \mathbf{X}_1 \underline{\beta}_1 + \mathbf{X}_2 \underline{\beta}_2 + \underline{\varepsilon}$$

Queremos entonces una estadística de prueba que nos ayude a determinar si los coeficientes dentro del vector  $\underline{\beta}_2$  son significativos. Para dar solución al problema nuevamente recurriremos a la des-



composición de la suma de cuadrados. Para el modelo completo  $\underline{Y} = \mathbf{X}\underline{\beta} + \underline{\varepsilon}$  sabemos que  $SCE$  es una medida del error del modelo completo, llamemos  $SCE_{H_1}$  a dicha suma. Ahora ajustemos el modelo reducido bajo  $H_0$  es decir elaboremos la tabla ANOVA del modelo  $\underline{Y} = \mathbf{X}_1\underline{\beta}_1 + \underline{\varepsilon}$ , con dicha tabla encontramos la  $SCE$  asociada y la denominamos  $SCE_{H_0}$  la cual es una medida del ajuste del modelo reducido (bajo  $H_0$ ). Definamos ahora  $SCE_{H_1|H_0}$  como

$$SCE_{H_1|H_0} = SCE_{H_0} - SCE_{H_1}$$

Intuitivamente  $SCE_{H_1|H_0}$  mide el error que logramos explicar por haber incluido los coeficientes de  $\underline{\beta}_2$  al modelo y por tanto entre mas grande sea este número hay evidencia de que  $H_0$  deberá ser rechazada. Se puede probar usando formas cuadráticas que:

$$\frac{SCE_{H_1|H_0}}{\sigma^2} = \frac{SCE_{H_0} - SCE_{H_1}}{\sigma^2} \sim \chi_r^2$$

Donde recordemos que  $r$  son el número de parámetros que tiene el vector  $\underline{\beta}_2$ , luego entonces siendo consistentes con lo que hemos estado haciendo, la estadística que surge para probar la hipótesis es:

$$F = \frac{\frac{SCE_{H_0} - SCE_{H_1}}{r\sigma^2}}{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sigma^2(n-p)}} = \frac{(SCE_{H_0} - SCE_{H_1})}{r\hat{\sigma}^2} \sim \mathcal{F}_{(r, n-p)}$$

Luego entonces el procedimiento para hacer la prueba es el siguiente:

Suponga que se plante la hipótesis:

$$H_0 : \underline{\beta}_2 = \underline{0} \quad vs \quad H_1 : \underline{\beta}_2 \neq \underline{0}$$

Entonces la forma en como se procede es:

- Ajustar el modelo completo y calcular  $SCE_{H_1}$  y  $SCME := \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-p}$
- Ajustar el modelo reducido bajo  $H_0$  donde encontramos  $SCE_{H_0}$
- Calcular el estadístico F y compararlo con el cuantíl  $\mathcal{F}_{r, n-p}^{1-\alpha}$  Donde:

$$F = \frac{\frac{SCE_{H_0} - SCE_{H_1}}{r\sigma^2}}{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sigma^2(n-p)}} = \frac{(SCE_{H_0} - SCE_{H_1})}{r\hat{\sigma}^2} \sim \mathcal{F}_{(r, n-p)}$$

A continuación plantearemos un caso particular muy importante para el modelo lineal. Supon-

gamos que particionamos al vector de parámetros  $\underline{\beta}$  como sigue:

$$\underline{\beta}_1 = (\beta_0) \quad \underline{\beta}_2 = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix}_{p \times 1}$$

La prueba de hipótesis asociada bajo esta partición se le conoce como **prueba de significancia de la regresión** y es calculada por casi todos los paquetes estadísticos con el fin de determinar si existe una relación entre las variables explicativas y la variable respuesta:

$$H_0 : \beta_1 = \beta_2 = \dots \beta_k = 0 \quad H_1 : \beta_j \neq 0 \text{ para algún } j \in \{1, \dots, k\}$$

En este caso particular el estadístico de prueba asociado toma la siguiente forma:

$$F = \frac{\frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sigma^2(p-1)}}{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sigma^2(n-p)}} = \frac{(n-p) \sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{(p-1) \sum_{i=1}^n (y_i - \hat{y}_i)^2} \sim \mathcal{F}_{(p-1, n-p)}$$

Luego entonces rechazamos  $H_0$  si el valor de nuestra estadística  $F$  es mas grande que el cuantil  $\mathcal{F}_{(p-1, n-p)}$

La prueba  $F$  que mostramos es sin duda muy importante en el análisis de regresión pues nos indica si realmente existe algún tipo de efecto entre las variables independientes y la variable respuesta, si no rechazamos esta prueba y nos inclinamos por decir que  $H_0$  es cierta entonces el modelo de regresión que estamos planteando no es válido. Debido a su importancia, esta prueba esta programada en prácticamente en todos los paquetes estadísticos y se encuentra resumida en la tabla ANOVA.

Var.	S. C.	G. Lib	S.C.M.	F
(SCR)	$\sum_{i=1}^n (\hat{y}_i - \bar{y})^2$	$p - 1$	$\frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{p-1}$	$\frac{(n-p) \sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{(p-1) \sum_{i=1}^n (y_i - \hat{y}_i)^2}$
(SCE)	$\sum_{i=1}^n (y_i - \hat{y}_i)^2$	$n - p$	$\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-p}$	
(SCT)	$\sum_{i=1}^n (y_i - \bar{y})^2$	$n - 1$	$\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1}$	

En resumen, la tabla ANOVA nos ayuda a contrastar la hipótesis mas importante de la regresión.

$$H_0 : \beta_1 = \beta_2 = \dots \beta_k = 0 \quad H_1 : \beta_j \neq 0 \text{ para algún } j \in \{1, \dots, k\}$$

Y se rechaza  $H_0$  al nivel de significancia  $\alpha$  si el valor del estadístico  $F$  toma valores mas grandes que el cuantil  $\mathcal{F}_{(p-1, n-p)}^{1-\alpha}$ .

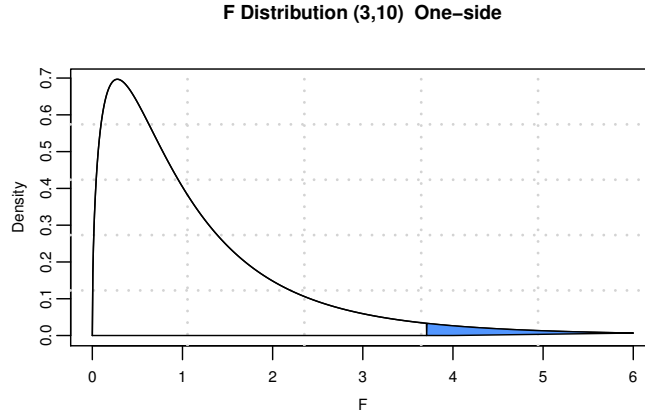


Figura 4.27: Región de Rechazo

Por otro lado, el coeficiente de determinación es un número entre 0 y 1 que comúnmente arrojan los paquetes estadísticos y lo denotan como  $R^2$  el cual sirve como una medida del ajuste del modelo. Este valor puede ser obtenido de la siguiente manera utilizando la Tabla ANOVA:

$$R^2 = \frac{SCR}{SCT} = 1 - \frac{SCE}{SCT}$$

El coeficiente de determinación tiene el problema de que siempre crece conforme vamos incluyendo mas variables lo cual crea un conflicto al momento de querer seleccionar modelos pues si utilizamos al coeficiente de determinación como un indice para determinar el mejor modelo caeríamos en la regla de ajustar siempre el modelo con mayor número de variables lo cual no puede ser siempre lo mejor pues sobreparametriza el modelo. Para resolver este problema los paquetes también arrojan un coeficiente de determinación ajustado que penaliza el número de variables que tiene el modelo. Se define como:

$$R_{Adj}^2 = 1 - \frac{SCE/(n-p)}{SCT/(n-1)} = 1 - \frac{SCME}{SCMT}$$

El caso mas general de pruebas de hipótesis es el siguiente:

$$H_0 : \mathbf{T}\underline{\beta} = \underline{0} \quad vs \quad H_1 : \mathbf{T}\underline{\beta} \neq \underline{0}$$

Donde  $\mathbf{T}$  es una matriz de  $(r \times p)$ . Este tipo de pruebas surgen cuando queremos verificar si alguna combinación lineal de los parámetros es igual cero, por ejemplo, suponga que tenemos que el modelo lineal:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon$$

Y queremos plantear la hipótesis

$$H_0 : \beta_1 + \beta_3 = 0 \quad \text{vs} \quad H_1 : \beta_1 + \beta_3 \neq 0$$

Entonces definimos  $\mathbf{T}$  como sigue:

$$\mathbf{T} = \begin{pmatrix} 0 & 1 & 0 & 1 & 0 \end{pmatrix} \Rightarrow T\beta = \beta_1 + \beta_3$$

La forma de atacar este problema es similar, es decir ajustar el modelo reducido y completo, extraer una medida de ajuste de cada modelo y comparar su diferencia con un cuantíl de la distribución  $\mathcal{F}$ . Queda fuera del alcance del curso para mas información revisar el libro *Introduction to Linear Regression Analysis* en la sección 3.3.4.

## Intervalos de confianza

La expresión (4.23) es una cantidad pivotal y por tanto puede ser utilizada para construir intervalos de confianza:  $i \in \{0, 1, \dots, k\}$

$$\mathbb{P} \left( -\tau_{n-p}^{1-\alpha/2} \leq \frac{\hat{\beta}_i - \beta_i}{\sqrt{\hat{\sigma}^2 C_{(i+1)(i+1)}}} \leq \tau_{n-p}^{1-\alpha/2} \right) = 1 - \alpha$$

$$\mathbb{P} \left( \hat{\beta}_i - \sqrt{\hat{\sigma}^2 C_{(i+1)(i+1)}} \tau_{n-p}^{1-\alpha/2} \leq \beta_i \leq \hat{\beta}_i + \sqrt{\hat{\sigma}^2 C_{(i+1)(i+1)}} \tau_{n-p}^{1-\alpha/2} \right) = 1 - \alpha$$

Por lo tanto un intervalo de confianza al  $(1 - \alpha) \%$  es:

$$\left( \hat{\beta}_i - \sqrt{\hat{\sigma}^2 C_{(i+1)(i+1)}} \tau_{n-p}^{1-\alpha/2}, \hat{\beta}_i + \sqrt{\hat{\sigma}^2 C_{(i+1)(i+1)}} \tau_{n-p}^{1-\alpha/2} \right)$$

Donde  $\sqrt{\hat{\sigma}^2 C_{(i+1)(i+1)}}$  se le conoce como error estándar del estimador.

Por otro lado el intervalo de confianza para  $\sigma^2$  es:

$$\left( \frac{1}{\chi_{n-p}^2 (1-\alpha/2)} \sum_{i=1}^n (y_i - \hat{y}_i)^2, \frac{1}{\chi_{n-p}^2 (\alpha/2)} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \right)$$

## Análisis inferencial de la variable respuesta

Consideremos una observación de las covariables  $(X_1, X_2, \dots, X_k)$  y construyamos al vector  $\underline{x}_0$  a partir de dicha observación:

$$\underline{x}_0 = \begin{pmatrix} 1 \\ x_{01} \\ x_{02} \\ \vdots \\ x_{0k} \end{pmatrix}$$

Supongamos que ahora queremos hacer inferencia de la variable respuesta  $y$  dado que se observa  $\underline{x}_0$ , el valor ajustado por nuestro modelo para este punto es el siguiente:

$$\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_{01} + \dots + \hat{\beta}_k x_{0k} = \underline{x}_0^t \hat{\underline{\beta}}$$

Queremos encontrar un intervalo para  $\mathbb{E}(y \mid \underline{x}_0)$ , queda claro que por ser  $\hat{\underline{\beta}}$  insesgado que  $\hat{y}_0$  también un estimador insesgado para  $\mathbb{E}(y \mid \underline{x}_0)$ , en efecto pues:

$$\mathbb{E}(\hat{y}_0) = \mathbb{E}(\underline{x}_0^t \hat{\underline{\beta}}) = \underline{x}_0^t \mathbb{E}(\hat{\underline{\beta}}) = \underline{x}_0^t \underline{\beta} = \mathbb{E}(y \mid \underline{x}_0)$$

Por otro lado su varianza es:

$$\text{Var}(\hat{y}_0) = \text{Var}(\underline{x}_0^t \hat{\underline{\beta}}) = \underline{x}_0^t \text{Var}(\hat{\underline{\beta}}) \underline{x}_0 = \sigma^2 \underline{x}_0^t (\mathbf{X}^t \mathbf{X})^{-1} \underline{x}_0$$

Surge la pregunta de saber como se distribuye  $\hat{y}_0$ , como  $\hat{\underline{\beta}} = (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \underline{Y}$  entonces:

$$\hat{y}_0 = \underline{x}_0^t \hat{\underline{\beta}} = \underline{x}_0^t (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \underline{Y} = \mathbf{C} \underline{Y}$$

Como  $\underline{Y}$  es un vector normal-multivariado se sigue que  $\hat{y}_0$  sigue una distribución normal, luego entonces concluimos que:

$$\hat{y}_0 \sim N\left(\mathbb{E}(y \mid \underline{x}_0), \sigma^2 \underline{x}_0^t (\mathbf{X}^t \mathbf{X})^{-1} \underline{x}_0\right)$$

Por lo tanto:

$$\frac{\hat{y}_0 - \mathbb{E}(y \mid \underline{x}_0)}{\sqrt{\sigma^2 \underline{x}_0^t (\mathbf{X}^t \mathbf{X})^{-1} \underline{x}_0}} \sim N(0, 1)$$

Asumiendo que  $(n-p) \frac{\hat{\sigma}^2}{\sigma^2} = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sigma^2} \sim \chi_{(n-p)}^2$  entonces:

$$\frac{\hat{y}_0 - \mathbb{E}(y \mid \underline{x}_0)}{\sqrt{\hat{\sigma}^2 \underline{x}_0^t (\mathbf{X}^t \mathbf{X})^{-1} \underline{x}_0}} \sim \tau_{(n-p)}$$

La cual ya es una cantidad pivotal de donde construiremos el intervalo de confianza para la res-

puesta media  $\mathbb{E}(y \mid \underline{x}_0)$ .

$$\mathbb{P} \left( -\tau_{(n-p)}^{1-\frac{\alpha}{2}} \leq \frac{\hat{y}_0 - \mathbb{E}(y \mid \underline{x}_0)}{\sqrt{\hat{\sigma}^2 \underline{x}_0^t (\mathbf{X}^t \mathbf{X})^{-1} \underline{x}_0}} \leq \tau_{(n-p)}^{1-\frac{\alpha}{2}} \right) = 1 - \alpha$$

Despejando obtenemos que el Intervalo al  $(1 - \alpha)$  % de confianza para la respuesta media es:

$$\left( \hat{y}_0 - \tau_{(n-p)}^{(1-\frac{\alpha}{2})} \sqrt{\hat{\sigma}^2 \underline{x}_0^t (\mathbf{X}^t \mathbf{X})^{-1} \underline{x}_0}, \hat{y}_0 + \tau_{(n-p)}^{(1-\frac{\alpha}{2})} \sqrt{\hat{\sigma}^2 \underline{x}_0^t (\mathbf{X}^t \mathbf{X})^{-1} \underline{x}_0} \right)$$

Ahora supongamos que queremos construir un intervalo de predicción para una nueva observación  $y_0$  (cuando las covariables toman el valor de  $\underline{x}_0$ ) Sabemos que una estimación natural para la variable respuesta  $y_0$  es

$$\hat{y}_0 = \underline{x}_0^t \hat{\underline{\beta}}$$

Notemos que como  $y_0$  es una nueva observación entonces  $y_0$  es independiente de las observaciones anteriores  $(y_1, \dots, y_n)$  y como los estimadores  $\hat{\beta}_0, \dots, \hat{\beta}_k$  son combinaciones lineales de las observaciones  $(y_1, \dots, y_n)$ , se sigue entonces que  $y_0$  es independiente de  $\hat{y}_0$ , por lo tanto  $(y_0 - \hat{y}_0)$  sigue una distribución normal. Luego como:

$$\begin{aligned} \mathbb{E}((y_0 - \hat{y}_0)) &= \mathbb{E}(\underline{x}_0^t \underline{\beta} + \varepsilon_0 - \underline{x}_0^t \hat{\underline{\beta}}) = 0 \\ \text{Var}((y_0 - \hat{y}_0)) &= \text{Var}(y_0) + \text{Var}(\hat{y}_0) = \sigma^2 + \text{Var}(\underline{x}_0^t \hat{\underline{\beta}}) \\ &= \sigma^2 + \sigma^2 \underline{x}_0^t (\mathbf{X}^t \mathbf{X})^{-1} \underline{x}_0 \\ &= \sigma^2 \left( 1 + \underline{x}_0^t (\mathbf{X}^t \mathbf{X})^{-1} \underline{x}_0 \right) \end{aligned}$$

Finalmente hemos probado que:

$$(y_0 - \hat{y}_0) \sim N \left( 0, \sigma^2 \left( 1 + \underline{x}_0^t (\mathbf{X}^t \mathbf{X})^{-1} \underline{x}_0 \right) \right)$$

Asumiendo que  $(n-p) \frac{\hat{\sigma}^2}{\sigma^2} = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sigma^2} \sim \chi_{(n-p)}^2$  entonces:

$$\frac{(y_0 - \hat{y}_0)}{\sqrt{\hat{\sigma}^2 \left( 1 + \underline{x}_0^t (\mathbf{X}^t \mathbf{X})^{-1} \underline{x}_0 \right)}} \sim t_{(n-p)}$$

De donde obtenemos una cantidad pivotal para construir el siguiente intervalo de confianza para nuevas observaciones:

$$\left( \hat{y}_0 - \tau_{(n-p)}^{(1-\frac{\alpha}{2})} \sqrt{\hat{\sigma}^2 + \hat{\sigma}^2 \underline{x}_0^t (\mathbf{X}^t \mathbf{X})^{-1} \underline{x}_0}, \hat{y}_0 + \tau_{(n-p)}^{(1-\frac{\alpha}{2})} \sqrt{\hat{\sigma}^2 + \hat{\sigma}^2 \underline{x}_0^t (\mathbf{X}^t \mathbf{X})^{-1} \underline{x}_0} \right)$$

Observaciones de los intervalos de confianza:

- Este intervalo es similar al encontrado anteriormente solo que agrega mas variabilidad debido a toma en cuenta la variabilidad de los estimadores así como de la variabilidad que tiene el modelo en su definición  $\varepsilon_0$
- Hay que tener mucho cuidado con la extrapolación como se muestra en la siguiente grafica

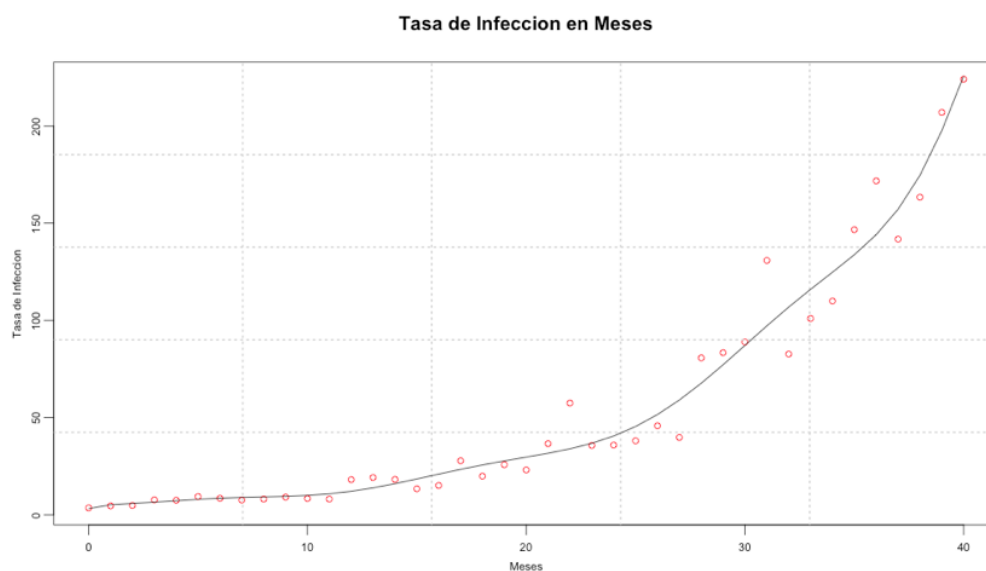


Figura 4.28: Problemas con extrapolar

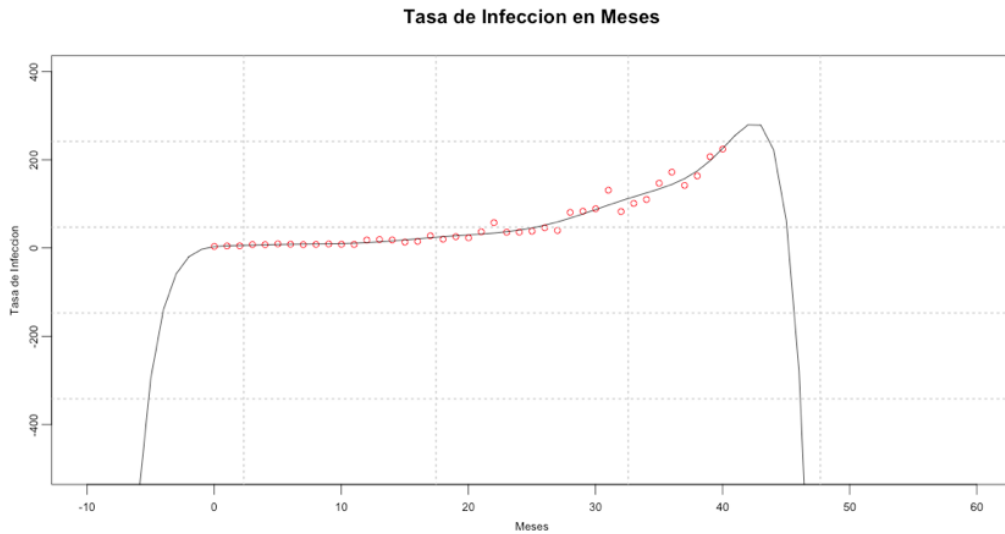


Figura 4.29: Problemas para extrapolar

Al hacer inferencia sobre nuevas respuestas en un punto dado  $\underline{x}_0^t$ , se debe de tener cuidado de la extrapolación pues es posible que los datos ajusten bien pero solo en la región en donde se tomó la muestra y no fuera de ella. Surge el problema de cómo determinar la región donde el modelo es válido. Ejemplo, supongamos que tenemos solo dos variables explicativas:

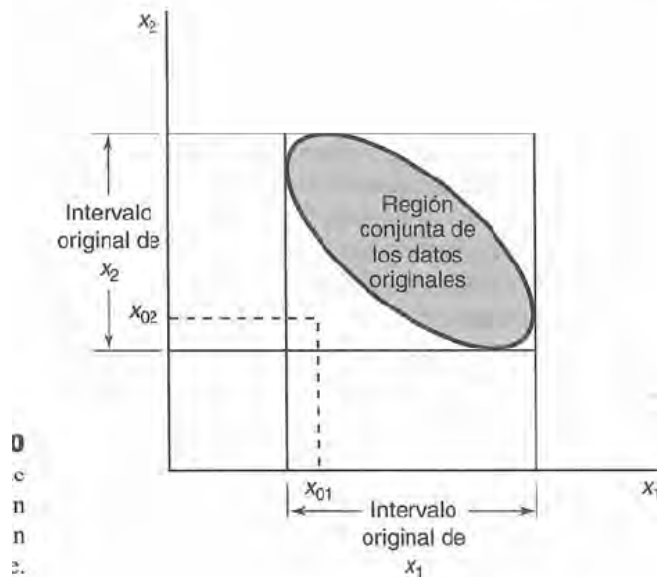


Figura 4.30: Región de extrapolarion

Necesitamos encontrar la región mínima en el espacio de las variables explicativas que contenga a todas las observaciones en muestra. Encontrar dicha región es difícil por lo que se estila encontrar



a la *elipse* de menor área que cubra a todos los puntos Cook [1979] y Weisberg [1985], consideran esta elipse

$$\underline{x}^t (\mathbf{X}^t \mathbf{X})^{-1} \underline{x} \leq h_{max}$$

Donde  $h_{max} := \text{máx} \{h_{11}, \dots, h_{nn}\}$  y  $h_{ii}$  son los elementos de la diagonal de la matriz Sombrero  $\mathbf{H}$ . Ellos demuestran que esta elipse encierra a todos los puntos de la muestra pero no garantizan que sea la mínima, sin embargo es una buena aproximación para solucionar el problema.

Luego entonces, suponga que queremos inferir a  $y$  en el punto  $\underline{x}_0$ , para verificar si nuestra predicción es válida, tenemos que calcular:

$$h_{00} = \underline{x}_0^t (\mathbf{X}^t \mathbf{X})^{-1} \underline{x}_0$$

Luego si:

- $h_{00} > h_{max}$  entonces,  $x_0$  está fuera de la región de muestra
- $h_{00} \leq h_{max}$  entonces,  $x_0$  está cercano a la región de muestra por lo que la estimación de  $y$  se considera buena

### 4.5.3. Validación del Modelo Lineal Multiple

En General, no se pueden detectar desviaciones respecto a las premisas básicas examinando los estadísticos estándar de resumen como por ejemplo los estadísticos  $t$ ,  $F$  o  $R^2$ , para validar correctamente el modelo se tiene que revisar los siguientes puntos:

- La relación entre la respuesta  $y$  y los regresores es lineal, (al menos en forma aproximada)
- Independencia (ACF, DurbinWatson, Rachas)
- Varianza Constante (Homocedasticidad, Bartlett, Levene)
- Normalidad (QQPlot, Histogramas )
- Multicolinealidad
- *Outliers* y observaciones influyentes

#### Linealidad

Recordemos que el modelo lineal que ajustamos es de la forma:

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + \varepsilon_i$$

Si  $k = 1$  (Modelo lineal simple) es relativamente fácil verificar linealidad, basta con realizar una gráfica de los puntos  $(y_i, x_{i1})$  para verificar si en realidad existe una recta asociada a la relación.

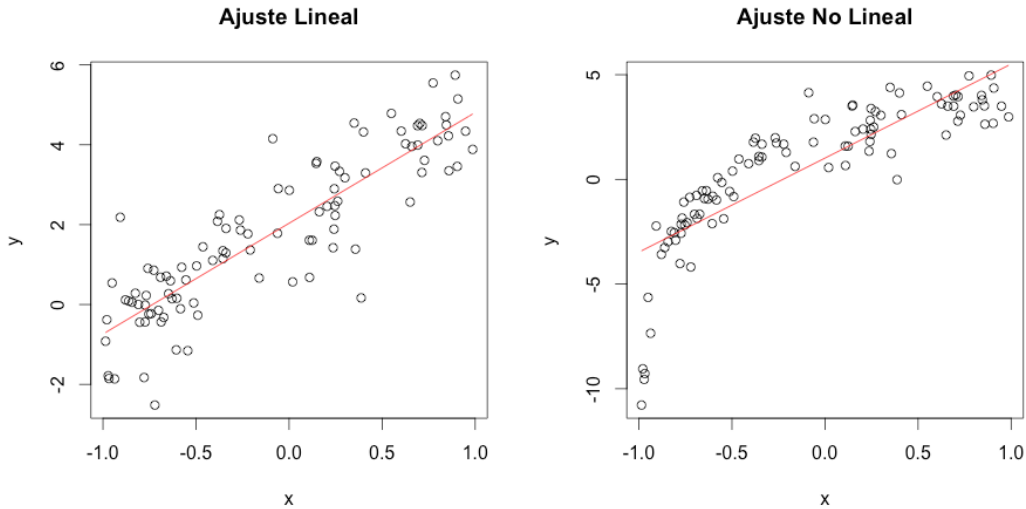


Figura 4.31: Relacion Lineal

Cuando tenemos 2 o mas variables explicativas la verificación de linealidad se complica debido a que se requiere graficar en dimensiones mayores. Una solución inicial al problema es graficar cada una de las variables explicativas contra la variable respuesta y verificar ahí la linealidad, sin embargo dichos graficos pueden llevarnos a conclusiones erróneas cuando los coeficientes tienen magnitudes distintas. En la siguiente figura se muestran las gráficas de las covariables vs la variable respuesta de un modelo **lineal**

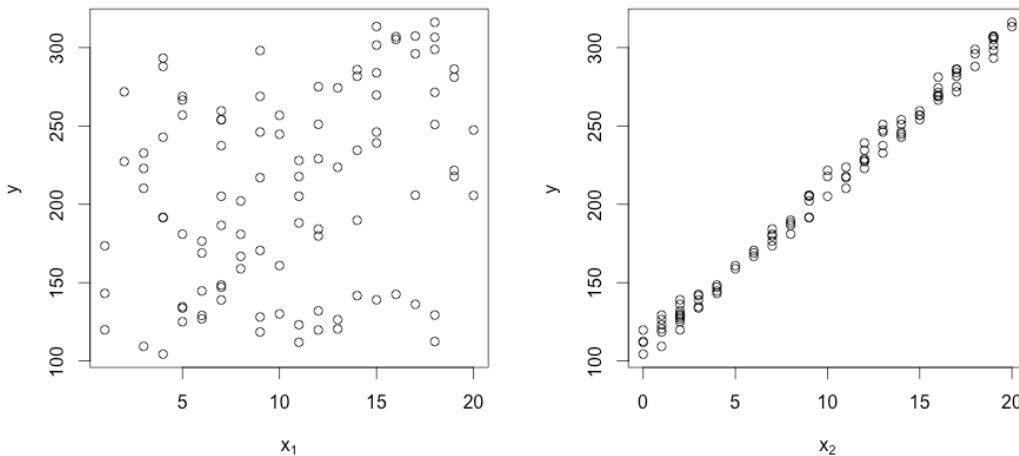


Figura 4.32: Relacion Lineal

Para solucionar esto recurriremos al análisis de residuales parciales. Supongamos que queremos ver si existe una relación lineal entre la variables  $X_q$  y  $Y$ , entonces lo que hacemos es ajustar el modelo **eliminando** a la variable  $X_q$  del modelo, luego encontramos los residuales de dicho modelo (denotamos por  $e_i^{(q)}$ ) y luego graficamos esos residuales contra los valores de la variables  $X_q$ , es

decir visualizamos los puntos  $(x_{iq}, e_i^{(q)})$ . Si la relación es lineal entre  $X_q$  y  $Y$  se espera que esta gráfica de residuales parciales presente una relación lineal, en caso contrario, es una indicación de que esa variable posiblemente necesite ser transformada para ayudar al ajuste.

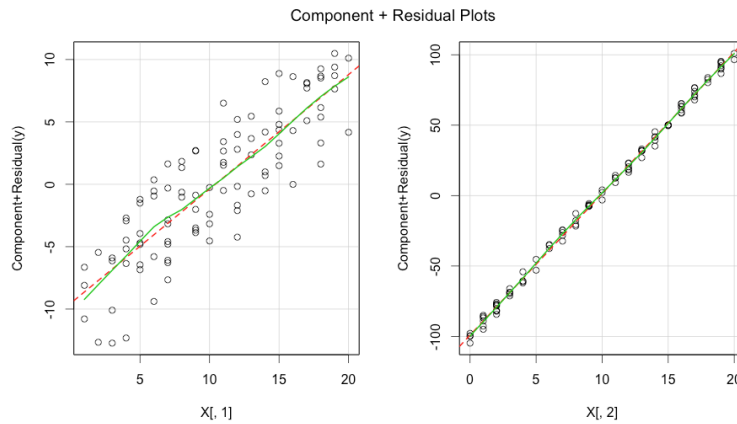


Figura 4.33: Residual Parcial

## Independencia

En el modelo de regresión lineal simple se plantearon las pruebas para identificar independencia utilizando el ACF (Función de autocorrelación) así como la prueba no paramétrica de Rachas. Otra prueba muy utilizada para la identificación de independencia (no correlación) es utilizar la prueba de *Durbin Watson*. Esta prueba pretende ver si los valores presentan algún tipo de dependencia en cuanto al orden de obtención. La prueba se plantea como sigue: Supongamos que los errores  $\varepsilon_i$  siguen un modelo AR(1) entonces:

$$\varepsilon_i = \rho\varepsilon_{i-1} + \delta_i \quad \delta_i \sim N(0, \sigma^2)$$

Luego entonces planteamos:

$$H_0 : \rho = 0 \quad vs \quad H_1 : \rho \neq 0$$

Estadístico de prueba:

$$d = \frac{\sum_{i=2}^n (e_i + e_{i-1})^2}{\sum_{i=1}^n e_i^2}$$

El comando utilizado en R es `durbinWatsonTest(model)` y esperamos ver un p-value alto para no rechazar. (Ojo: Estas pruebas suponen Normalidad)

## Homocedasticidad

Generalmente para verificar que tenemos varianza constante se llevan a cabo pruebas visuales verificando el comportamiento de los residuales conforme se van observando, o bien graficar contra

$\hat{y}_i$ .

Dado que ahora tenemos mas variables explicativas, también se suele graficar los residuales contra cada una de las variables explicativas y verificar le homocedasticidad ahi.

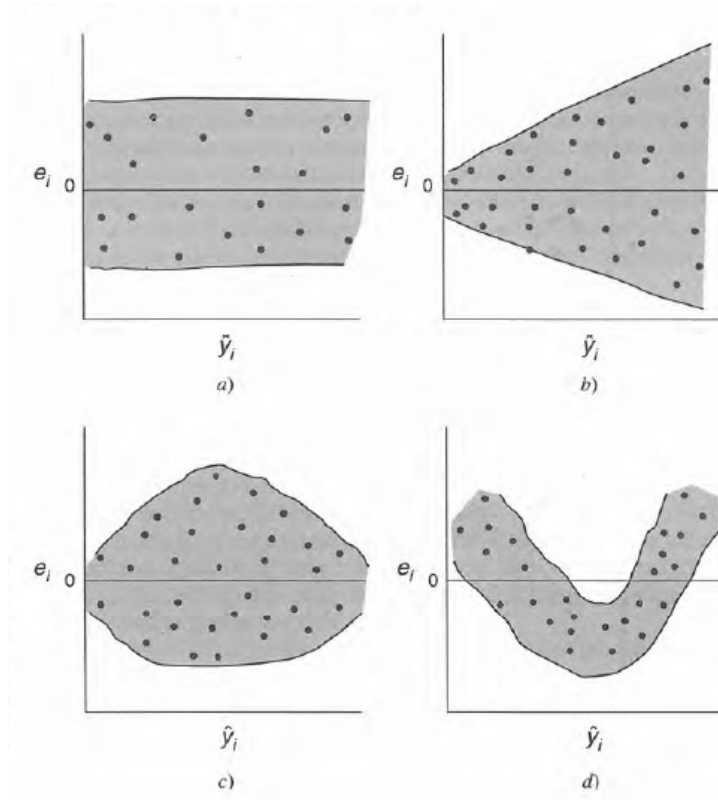


Figura 4.34: Homocedasticidad

Hay que recordar que existen pruebas formales para verificar la homocedasticidad (Ej. Bartlett, Levene) pero requieren tener definido grupos lo cual puede quitarle credibilidad a la prueba. Ver prueba `ncvTest`.

## Normalidad

- Realizar histogramas de los residuales esperando ver un comportamiento normal con media en 0
- Realizar la grafica QQ-Plot
- Pruebas de Bondad de Ajuste (Solo como medida del ajuste)

## Multicolinealidad

- Una forma muy simple para detectar multicolinealidad es inspeccionar la matriz de correlaciones entre las variables explicativas. (`cor(X)` , `pairs(X)`)

- Hacer regresiones lineales de  $X_q$  con el resto de las variables y obtener el coeficiente de determinación de dicha regresión  $R_q^2$ , un coeficiente de determinación cercana a 1 nos dice que hay un ajuste muy bueno por lo que se puede decir que:

$$\underline{X}_q \approx \alpha_0 \underline{X}_0 + \dots + \alpha_{q-1} \underline{X}_{q-1} + \alpha_{q+1} \underline{X}_{q+1} + \dots + \alpha_k \underline{X}_k$$

Una forma de obtener índices es calcular lo que se denomina el  $VIF_q$  (Factor de Inflación de la varianza) definido como:

$$VIF_q = \frac{1}{1 - R_q^2} \quad q \in 0, 1, \dots, k$$

Luego entonces si  $R_q^2 \approx 1$  eso se traduce en un  $VIF_q$  grande, en la práctica se suele tomar como punto de corte 5 ó 10, es decir obtener un  $VIF_q > 10$  nos habla de que la variable  $X_q$  puede ser expresada aproximadamente como combinación lineal de las restantes y por tanto podemos tener problemas numéricos en las estimaciones.

- Otra forma de detectar la multicolinealidad es con el índice de condición de una matriz  $(\mathbf{X}^t \mathbf{X})$  el cual se define como:

$$\kappa = \frac{\lambda_{max}}{\lambda_{min}} \quad \text{kappa}(t(X) \% * \%X)$$

Generalmente un índice de condición menor a 100 no tiene problemas de multicolinealidad, si está entre 100 y 1000 implica una moderada multicolinealidad y si excede a 1000 entonces hay problemas importantes y podemos tener problemas numéricos

## Outliers y Observaciones influyentes

Hay dos propósitos principales para tratar de identificar las observaciones que no pertenecen al modelo

- Proteger la integridad del modelo de los efectos de los puntos que no pertenecen a éste. (Observaciones que no pertenecen al modelo) Frecuentemente exhiben residuales grandes.
- Identificar las deficiencias del modelo . Posiblemente se necesite incluir nuevas variables o se requiere llevar a cabo transformaciones.

El primer paso para identificar observaciones atípicas es tener una idea de la región en donde estamos muestreando, la idea es encontrar una métrica que nos ayude a decidir si una observación de las variables independientes está muy alejada de la región de muestreo. En general una observación que se encuentra lejos de donde esta la masa de puntos ocasiona que el modelo cambie de manera significativa (Aunque no necesariamente).

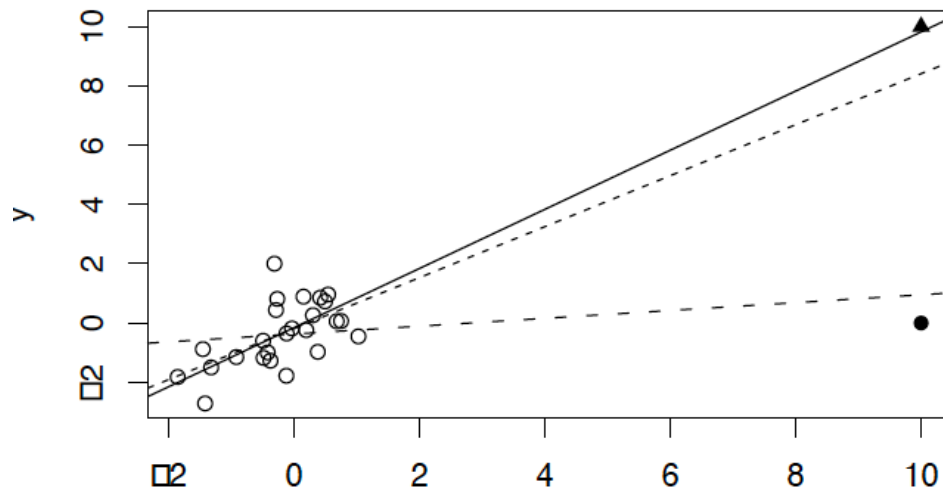


Figura 4.35: Observaciones Influyentes

Para detectar los puntos de muestreo que están alejados se utiliza la matriz sombrero,  $\mathbf{H} = \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t$ . A los elementos de la diagonal de  $\mathbf{H}$  ( $h_{ii}$ ) se les conoce como el *leverage* o la influencia de cada observación. De este modo un *leverage* grande nos habla de una observación alejada de la masa de los puntos de muestreo.

Como:

$$tr(\mathbf{H}) = \sum_{i=1}^n h_{ii} = p$$

Por lo tanto, en promedio los *leverages* toman el valor de  $\frac{p}{n}$ . Existe una regla muy utilizada que dice que aquellos  $h_{ii}$  que sean mayores a dos veces al promedio son puntos alejados y por tanto deben de ser analizados. ( $h_{ii} > 2\frac{p}{n}$ ) En la siguiente gráfica se observa como 4 observaciones se sale de la region definida, esas cuatro observaciones deben de ser analizadas y verificar si es correcto que estén en muestra. Si es correcta la observación lo siguiente es verificar si se trata de un outlier para el modelo o bien una observación influyente.

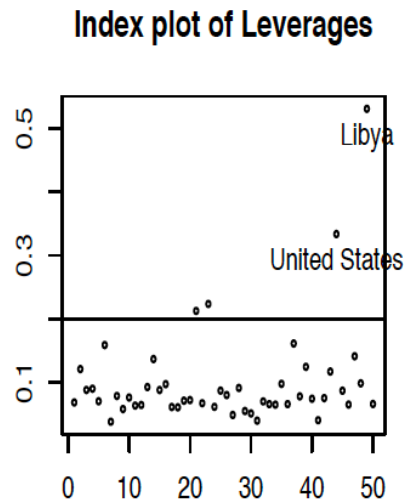


Figura 4.36: Observaciones Influyentes

Outlier: Es un punto que no ajusta en el modelo.

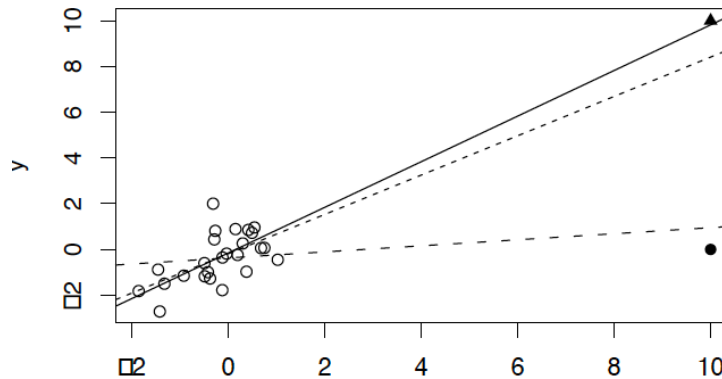


Figura 4.37: Observaciones Atípicas

Surge entonces la necesidad de saber cómo detectar outliers, un primer intento para encontrar este tipo de observaciones consiste en *studentizar* a los residuales:

$$r_i = \frac{e_i}{\hat{\sigma}\sqrt{1 - h_{ii}}}$$

La idea de esto es homologar la varianza, pues teóricamente los residuales no tiene varianza constante. Luego se espera que estos residuales  $r_i$  se encuentren en una banda entre  $-3$  y  $3$ , aquellos fuera de estas bandas serán candidatos a ser analizados como outliers del modelo. (Posibles causas: ¿Falta de Ajuste del modelo? ¿Errores de captura? )

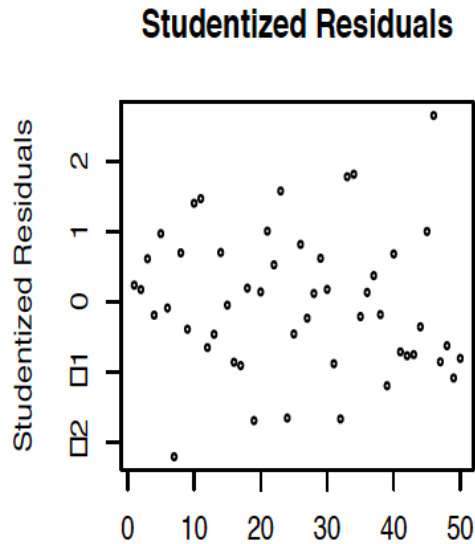


Figura 4.38: Observaciones atípicas

El método anterior tiene el problema de que posiblemente el modelo ajustado este siendo influenciado por el dato outlier y por tanto no sea detectado en los residuales, para solucionar esto se estila utilizar lo que se denomina el residual de *jackknife* (Residual de validación cruzada)

$$t_i = \frac{y_i - \hat{y}_{(i)}}{\hat{\sigma}_{(i)} \left( 1 + \underline{x}_i^T (\mathbf{X}_{(i)}^T \mathbf{X}_{(i)})^{-1} \underline{x}_i \right)^{1/2}}$$

Donde  $\hat{y}_{(i)} = \underline{x}_i^t \hat{\beta}_{(i)}$  es el valor ajustado de la  $i$ -ésima observación quitando esa misma observación del ajuste, por lo que  $\hat{\beta}_{(i)}$  y  $\hat{\sigma}_{(i)}$  son los estimadores de  $\underline{\beta}$  y  $\sigma$  quitando la  $i$ -ésima observación. De la misma forma  $\mathbf{X}_{(i)}$  es la matriz de diseño sin el  $i$ -ésimo renglón. Afortunadamente existe una forma fácil de calcular este residual por medio del residual studentizado.

$$t_i = r_i \left( \frac{n - p - 1}{n - p - r_i^2} \right)^{1/2}$$

La gran ventaja de esta última fórmula es que no requiere estar ajustando  $n$  regresiones. (Ver función: `rstudent`)

Finalmente, se puede probar que bajo el supuesto de que  $\underline{\varepsilon} \sim N_n(\underline{0}, \sigma^2 \mathbf{I}_n)$  se tiene que:

$$t_i \sim t_{n-p-1}$$

Luego entonces, para detectar un outlier se suele utilizar la corrección de Bonferroni y comparar contra el cuantil  $\alpha/n$  de una distribución  $t$ . Por ejemplo si  $\alpha = 0.05$  entonces se dice que la  $i$ -ésima



observación es outlier si

$$|t_i| > t_{n-p-1}^{(1-\alpha^*)} \quad \text{con } \alpha^* = \frac{0.05}{2n}$$

Notas:

- Dos o mas outliers juntos pueden ocultarse entre ellos.
- Un outlier en un modelo puede dejar de serlo en otro cuando hubieron transformaciones, se recomienda hacer el análisis de outliers cada vez que se hace alguna transformación al modelo
- Revisar funcion `outlierTest`.

Un punto influyente es aquel que siendo removido del modelo causa un cambio importante en todo el ajuste. Algunas medidas para detectar la influencia de una observación son:

- Cambio en el vector de los coeficientes ajustados:  $\underline{\hat{\beta}} - \underline{\hat{\beta}}_{(i)}$
- Cambio en el vector de observaciones  $\mathbf{X}^t (\underline{\hat{\beta}} - \underline{\hat{\beta}}_{(i)}) = \underline{\hat{Y}} - \underline{\hat{Y}}_{(i)}$

Surge el problema de definir una distancia que nos ayude a determinar la diferencia entre vectores. Distancia de Cook:

$$D_i = \frac{(\underline{\hat{\beta}} - \underline{\hat{\beta}}_{(i)})^T (\mathbf{X}^T \mathbf{X}) (\underline{\hat{\beta}} - \underline{\hat{\beta}}_{(i)})}{p \hat{\sigma}^2}$$
$$D_i = \frac{1}{p} r_i^2 \frac{h_i}{1 - h_i}$$

(R: `cooks.distance`; `lm.influence`) Un regla usada para la detección de observaciones influyentes es considerar  $D_i > 1$  (Cook, R. Dennis; and Weisberg, Sanford (1982)). Pero se ha visto que es muy conservadora, algunos entonces utilizan  $D_i > 4/n$  (Bollen, Kenneth A.; and Jackman, Robert W. (1990);) Belsey, Kuh y Welch introducen otras dos medias de influencia. La primera es una estadística que indica cuanto cambia el coeficiente de regresión  $\hat{\beta}_j$  cuando la  $i$ -ésima observación es removida.

$$DFBETA_{j,i} = \frac{\hat{\beta}_j - \hat{\beta}_{j(i)}}{\sqrt{\hat{\sigma}_{(i)}^2 C_{jj}}}$$

Donde  $C_{jj}$  es el  $j$ -ésimo elemento de la matriz  $(\mathbf{X}^t \mathbf{X})^{-1}$ . Luego entonces un valor (en magnitud) grande de  $DFBETA_{j,i}$  indica que la observación  $i$  tiene un influencia sobre el coeficiente  $j$  Los autores sugieren un punto de corte igual a  $2/\sqrt{n}$

$$|DFBETA_{j,i}| > 2/\sqrt{n}$$

También podemos medir la influencia de una observación sobre los valores ajustados  $\hat{y}_i$ . El diagnóstico propuesto es:

$$DFFIT_i = \frac{\hat{y}_i - \hat{y}_{(i)}}{\sqrt{\hat{\sigma}_{(i)}^2 h_{ii}}}$$

Donde  $\hat{y}_{(i)}$  es el valor ajustado para  $y_i$ , obtenido sin usar la  $i$ -ésima observación. El denominador no es más que una estandarización pues se puede probar que  $\text{Var}(\hat{y}_i) = \sigma^2 h_{ii}$ . (Recuerde que  $h_{ii}$  son los elementos de la diagonal de la matriz Sombrero).

La interpretación para el  $DFFIT$  es que mide la cantidad de desviaciones estándar que cambia el valor ajustado  $\hat{y}_i$  si se elimina la observación  $i$ . Obviamente un valor grande (magnitud) nos indica una influencia de la observación  $i$ . Los autores sugieren que merece investigarse toda observación tal que:

$$|DFFIT_i| > 2\sqrt{p/n}$$

Los diagnósticos que hemos visto permiten conocer el efecto de las observaciones sobre los coeficientes y las estimaciones y no proporcionan información sobre la *precisión* de la estimación. Recordemos que la teoría inferencial nos indica que un estimador es mejor que otro si este último tiene una menor varianza, luego entonces resultaría útil saber si quitando una observación la estimación mejora.

Surge entonces la necesidad de tener una medida escalar de la varianza de un vector. Lo que se utiliza como una medida escalar es utilizar el determinante de la matriz de varianzas y covarianzas. Definimos entonces la *varianza generalizada* como:

$$GV(\hat{\underline{\beta}}) = \left| \text{Var}(\hat{\underline{\beta}}) \right| = \left| \sigma^2 (\mathbf{X}^t \mathbf{X})^{-1} \right|$$

Finalmente para tener una medida de cuánto ganamos o perdemos por quitar una observación definimos lo siguiente:

$$COVRATIO_i = \frac{\left| (\mathbf{X}_{(i)}^t \mathbf{X}_{(i)})^{-1} \sigma_{(i)}^2 \right|}{\left| (\mathbf{X}^t \mathbf{X})^{-1} \sigma^2 \right|}$$

Queda claro entonces que un  $COVRATIO_i < 1$  indica que la  $i$ -ésima observación mejora la precisión de la estimación. No es fácil obtener valores de corte para este índice pero Belsley, Kuh y Welsh (1980) sugieren que si:

$$COVRATIO_i > 1 + 3p/n \quad \text{o} \quad COVRATIO_i < 1 - 3p/n$$

Entonces se debe considerar al punto  $i$  como influyente. (Estos valores solo se recomiendan para muestras grandes).

<http://www.statmethods.net/stats/rdiagnostics.html>

## Ejercicios

**Ejercicio 4.5.15.** *En el modelo lineal multiple:*

$$\underline{Y} = \mathbf{X}\underline{\beta} + \underline{\varepsilon}; \quad \underline{\varepsilon} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I}_{n \times n})$$

*Demuestre que los estimadores de  $\sigma^2$  definidos a continuación tienen distribución Gamma*

$$\hat{\sigma}^2 := \frac{(\underline{Y} - \hat{\underline{Y}})^T (\underline{Y} - \hat{\underline{Y}})}{n - p} \quad \hat{\sigma}_{MV}^2 := \frac{(\underline{Y} - \hat{\underline{Y}})^T (\underline{Y} - \hat{\underline{Y}})}{n}$$

**Ejercicio 4.5.16.** *Descomposición de la suma de cuadrados en la regresión lineal multiple:*

- *En el modelo lineal multiple:*

$$\underline{Y} = \mathbf{X}\underline{\beta} + \underline{\varepsilon}$$

*Demuestre que  $\mathbf{H}\mathbf{X} = \mathbf{X}$*

- *Sean  $\underline{x}_0, \dots, \underline{x}_k$  las  $p$  columnas de la matriz  $\mathbf{X}$ , es decir:*

$$\mathbf{X} = (\underline{x}_0 | \underline{x}_1 | \dots | \underline{x}_k)$$

*Demuestre entonces que:*

$$\mathbf{H}\underline{x}_i = \underline{x}_i \quad i \in \{0, 1, \dots, k\}$$

- *Verifique que en el modelo lineal multiple se cumple que:*

$$\sum_{i=1}^n e_i = 0$$

$$\sum_{i=1}^n e_i \hat{y}_i = 0$$

*Donde  $e_i = y_i - \hat{y}_i$  es el residual  $i$  y  $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \dots + \hat{\beta}_k x_{ik}$*

*Hint: Defina al vector columna de 1's como  $\mathbf{1} = (1, 1, \dots, 1)^T$ , luego entonces observe que:*

$$\sum_{i=1}^n e_i = \underline{e}^T \mathbf{1} \quad \text{Con } \underline{e} = (e_1, e_2, \dots, e_n)^T$$

$$\sum_{i=1}^n e_i \hat{y}_i = \underline{e}^T \hat{\underline{Y}} \quad \text{Con } \hat{\underline{Y}} = (\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n)^T$$

- Utilizando lo anterior verifique que:

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$$

**Ejercicio 4.5.17.** Encuentre la ecuación de la parábola  $f(x) = \beta_0 + \beta_1x + \beta_2x^2$  que pasa por los puntos:

$$(0, 1), (1, 6), (2, 17)$$

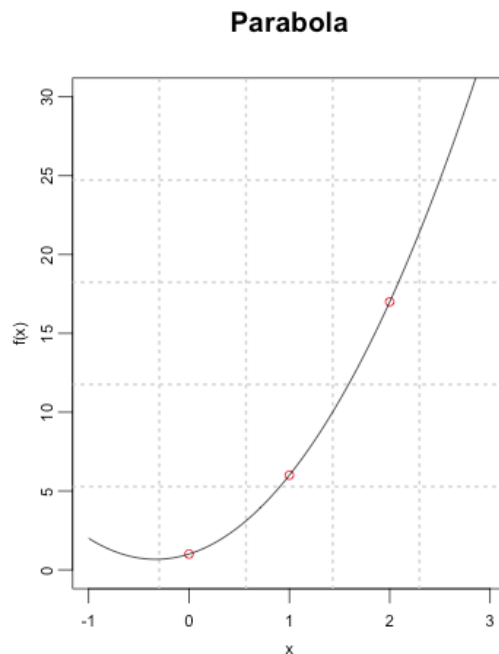


Figura 4.39: Ajuste de Parabola

**Ejercicio 4.5.18.** La tabla *FootballLeague.csv* contiene los datos sobre el desempeño de los equipos de la liga nacional de fútbol de E.U.A. durante 1976.

- Encuentre los estimadores máximo verosímiles de los parámetros del modelo lineal múltiple que relaciona el número de juegos ganados con
  - Yardas por aire del equipo ( $x_2$ )
  - El porcentaje de Yardas por Tierra ( $x_7$ )
  - Las Yardas por tierra del contrario ( $x_8$ )

$$y_i = \beta_0 + \beta_1x_2 + \beta_2x_7 + \beta_3x_8 + \varepsilon_i \quad \varepsilon_i \sim N(0, \sigma^2)$$

- Repita el inciso anterior pero ajustando el modelo:

$$y_i = \beta_1 \log(x_2) + \beta_2 \log(x_7) + \varepsilon_i \quad \varepsilon_i \sim N(0, \sigma^2)$$

- ¿Qué modelo tiene una menor suma de cuadrados del error?

**Ejercicio 4.5.19.** Se hizo un ajuste lineal múltiple con intercepto, complete la siguiente tabla ANOVA:

`anova(lm(y ~ x))`

*Analysis of Variance Table*

*Response: y*

<i>Fnte. de Var.</i>	<i>Df.</i>	<i>Sum Sq</i>	<i>Mean Sq</i>	<i>F value</i>	<i>Pr(&gt;F)</i>
<i>x</i>	<i>3</i>		<i>1600.81</i>		<i>&lt;2.2e-16 ***</i>
<i>Residuals</i>	<i>36</i>	<i>146.9</i>			

*Responda lo siguiente*

- ¿Con cuántas observaciones se hizo el ajuste?
- ¿Con cuántas variables se hizo el ajuste?
- Tomando  $\alpha = 0.01$ . ¿Rechazaría la hipótesis  $H_0$ ?
- Proporcione el valor que toma la estimación insesgada para  $\sigma^2$
- Proporcione el valor que toma la estimación por máxima verosimilitud para  $\sigma^2$
- Construya un intervalo de confianza al 95 % para  $\sigma^2$
- ¿Cuanto vale  $S_{yy} := \sum_{i=1}^n (y_i - \bar{y})^2$ ?
- ¿Qué porcentaje de la variabilidad es explicada por el modelo?

**Ejercicio 4.5.20.** La tabla `empleo.csv` contiene información sobre datos de desempleo de un país. Las variables que contiene son:

- $X_1$ : Percentage price deflation
- $X_2$ : GNP in millions of dollars
- $X_3$ : Number of unemployed in thousands
- $X_4$ : Number of people employed by the military

- $X_5$ : Number of people over 14
- $X_6$ : Year
- $Y$ : People employed.

Se desea hacer un análisis de regresión que ayude a explicar el número de personas empleadas con base en las distintas variables de la forma:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \beta_6 x_{i6} + \varepsilon_i \quad \varepsilon_i \sim N(0, \sigma^2)$$

- Construya la tabla ANOVA y haga la prueba de significancia de la regresión.

$$H_0 : \beta_1 = \dots = \beta_6 = 0 \quad \text{vs} \quad H_1 : \beta_j \neq 0 \text{ para algún } j \in \{1, \dots, 6\}$$

¿ Si  $\alpha = 0.05$ , rechazaría  $H_0$ ?

- Calcule  $R^2$  y  $R^2$  ajustado
- Encuentre los intervalos al 98 % de confianza para cada uno de los parámetros de este modelo.
- Lleve a cabo la prueba de hipótesis:

$$H_0 : \beta_i = 0 \quad \text{vs} \quad H_1 : \beta_i \neq 0 \quad \text{con} \quad i \in \{1, \dots, 6\}$$

A un  $\alpha = 0.03$  qué pruebas no son rechazadas e interprete.

- Lleve a cabo la prueba de hipótesis:

$$H_0 : \beta_i = 2 \quad \text{vs} \quad H_1 : \beta_i > 2 \quad \text{con} \quad i \in \{1, \dots, 6\}$$

A un  $\alpha = 0.03$  qué pruebas no son rechazadas

- Contraste la prueba de hipótesis:

$$H_0 : \beta_1 = \beta_2 = \beta_5 = 0 \quad H_1 : \beta_j \neq 0 \text{ para algún } j \in \{1, 2, 5\}$$

**Ejercicio 4.5.21.** La tabla *FootballLeague.csv* contiene los datos sobre el desempeño de los equipos de la liga nacional de fútbol de E.U.A. durante 1976.

- Ajuste un modelo lineal múltiple que relaciona el número de juegos ganados con
  - Yardas por aire del equipo ( $x_2$ )
  - El porcentaje de Yardas por Tierra ( $x_7$ )

- *Las Yards por tierra del contrario* ( $x_8$ )

$$y_i = \beta_0 + \beta_1 x_2 + \beta_2 x_7 + \beta_3 x_8 + \varepsilon_i \quad \varepsilon_i \sim N(0, \sigma^2)$$

- *Contraste la prueba de hipótesis:*

$$H_0 : \beta_1 = \beta_3 = 0 \quad H_1 : \beta_1 \neq 0 \text{ o } \beta_3 \neq 0$$

¿Con  $\alpha = 0.05$ , rechazaría  $H_0$ ?

- *Gráfique la región al 95 % de confianza para los parámetros  $\beta_1$  y  $\beta_3$ , ¿El  $(0,0)$  esta contenido en la región de confianza?*
- *Calcule el vector de residuales  $\underline{e} = (\hat{e}_1, \hat{e}_2, \dots, \hat{e}_n)$ , donde  $e_i = y_i - \hat{y}_i$*
- *Valide el modelo:*
  - *Normalidad de los residuales (Q-Q Plot, Histogramas)*
  - *Homocedasticidad de los residuales (Grafico, Prueba Levene, Prueba Bartlett)*
  - *Independencia de los Residuales (Durbin-Watson, Prueba de Rachas)*
  - *Colinealidad de la matriz Diseño*
  - *Identifique Outliers y observaciones influyentes.*
- *Encuentre los intervalos al 98 % de confianza para cada uno de los parámetros  $\beta_0, \beta_1, \beta_2, \beta_3$*
- *Encuentre el intervalo al 92 % de confianza para la respuesta media de número juegos ganados cuando  $x_2 = 2300$ ,  $x_7 = 56$  y  $x_8 = 2100$  (Verifique si estamos en zona de intrapolación o extrapolación)*
- *Encuentre el intervalo al 93 % de confianza para el número de juegos ganados (nueva observación) cuando  $x_2 = 2100$ ,  $x_7 = 60$  y  $x_8 = 2000$  (Verifique si estamos en zona de intrapolación o extrapolación)*

## 4.6. Modelos Lineales Generalizados

Los modelos lineales generalizados contienen un componente aleatorio en cada una de sus  $n$  observaciones independientes en una variable respuesta que pertenece a la *familia exponencial natural*

$$\underline{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}_{n \times 1}$$

$$f_{Y_i}(y_i; \theta_i) = a(\theta_i) b(y_i) \exp(y_i Q(\theta_i)) \quad \forall i \in \{1, 2, \dots, n\} \quad (4.24)$$

Donde  $Q(\theta_i)$  es un parámetro natural.

En su forma más general se incluye un parámetro de dispersión  $\phi$ , entonces la función de densidad queda de la siguiente manera:

$$f_{Y_i}(y_i; \theta_i; \phi) = \exp \left[ \frac{(y_i \theta_i - b(\theta_i))}{a(\phi)} + c(y_i, \phi) \right] \quad \forall i \in \{1, 2, \dots, n\} \quad (4.25)$$

Estas dos ecuaciones son equivalentes si en (4.24) se hacen los siguientes cambios de variables:

- $a(\theta_i) = \exp \left[ \frac{-b(\theta_i)}{a(\phi)} \right]$
- $b(y_i) = \exp [c(y_i, \phi)]$
- $Q(\theta_i) = \frac{\theta_i}{a(\phi)}$

Este tipo de modelos incluyen un componente sistémico que relaciona a la variable respuesta en función de la media de dicha variable en forma lineal.

Sea  $X_{ij}$  el valor del predictor  $j$  (con  $j \in \{1, 2, \dots, p\}$ ) para el sujeto  $i$  (con  $i \in \{1, 2, \dots, n\}$ ), entonces el modelo lineal general puede representarse de la siguiente manera:

$$g(\mu_i) = \sum_{j=1}^p \beta_j X_{ij} \quad \text{con } \mu_i = \mathbb{E}(Y_i) \quad \forall i \in \{1, 2, \dots, n\}$$

Usualmente  $X_{i1} = 1 \quad \forall i \in \{1, 2, \dots, n\}$  y es conocido como el intercepto.

La función  $g(\cdot)$  es conocida como la función liga y conecta al componente aleatorio con el sistémico, además  $g(\cdot)$  es diferenciable y monótona.

### 4.6.1. Estimación de los parámetros

El objetivo de la estimación de los parámetros es poder manipular e interpretar mejor la información que arroja el modelo para posteriormente dar conclusiones generales sobre los datos y dar predicciones acordes al comportamiento de estos.

#### Ecuaciones de Verosimilitud

Sea  $L_i = \ln(f_{Y_i}(y_i; \theta_i; \phi))$  la contribución de  $Y_i$  sobre la verosimilitud, entonces la *log-verosimilitud* es:

$$L = \ln \left( \prod_{i=1}^n f_{Y_i}(y_i; \theta_i; \phi) \right) = \sum_{i=1}^n \ln(f_{Y_i}(y_i; \theta_i; \phi)) = \sum_{i=1}^n L_i \quad (4.26)$$

Donde:  $L_i = \frac{(y_i \theta_i - b(\theta_i))}{a(\phi)} + c(y_i, \phi)$



Bajo condiciones de regularidad satisfactorias para los miembros de la familia exponencial se tiene:

- $\mathbb{E} \left( \frac{\partial L_i}{\partial \theta_i} \right) = 0$
- $\mathbb{E} \left( -\frac{\partial^2 L_i}{\partial \theta_i^2} \right) = -\mathbb{E} \left( \left( \frac{\partial L_i}{\partial \theta_i} \right)^2 \right)$

Entonces las ecuaciones de verosimilitud para el modelo lineal general son:

$$\frac{\partial L_i}{\partial \beta_j} = \frac{\partial L_i}{\partial \theta_i} \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j} \quad (4.27)$$

Con:  $\eta_i = \beta_j X_{ij}$ ,  $\frac{\partial L_i}{\partial \theta_i} = \frac{(Y_i - \mu_i)}{a(\phi)}$ ,  $\frac{\partial \theta_i}{\partial \mu_i} = \frac{a(\phi)}{Var(Y_i)}$

Por lo tanto las ecuaciones de verosimilitud quedan expresadas de la siguiente manera:

$$\frac{\partial L_i}{\partial \beta_j} = \frac{(Y_i - \mu_i)}{a(\phi)} \frac{a(\phi)}{Var(Y_i)} \frac{\partial \mu_i}{\partial \eta_i} X_{ij} = \frac{(Y_i - \mu_i)}{Var(Y_i)} X_{ij} \frac{\partial \mu_i}{\partial \eta_i} \quad \forall i \in \{1, 2, \dots, n\} \quad (4.28)$$

Podemos notar que las ecuaciones de verosimilitud dependen de la distribución de  $Y_i$  sólo a través de su media ( $\mu_i$ ) y su varianza ( $Var(Y_i)$ ). En algunas distribuciones hay una relación entre la media y la varianza ( $Var(Y_i) = h(Var(\mu_i))$ ), de hecho en los miembros de la familia exponencial, la media y la varianza caracterizan a la distribución, aunque la estrecha relación entre la media y la varianza puede generar problemas de sobre dispersión.

### 4.6.2. Ajuste de un modelo lineal generalizado

Como los estimadores de  $\underline{\beta}$  por máxima verosimilitud no siempre son lineales, entonces se ocupan métodos iterativos como son:

- Newton-Raphson.
- Método de Fisher.

### Prueba de hipótesis

La idea es poder comparar dos modelos que tienen la misma distribución subyacente y tienen la misma función Link, para saber cuál es modelo que da un mejor ajuste, empezaremos a comparar modelos anidados, es decir, dados dos modelos  $M_0$  y  $M_1$  tales que todos los parámetros de  $M_0$  están en  $M_1$  más otros parámetros, entonces si la significancia de  $M_0$  es similar o igual a la significancia de  $M_1$  nos quedaremos con el modelo más sencillo, eso nos diría que los demás parámetros no aportan mucho en la explicación total, en cambio si la significancia de  $M_0$  es inferior a la significancia de  $M_1$  entonces nos quedamos con el modelo más general, lo que nos está diciendo es que quitarle parámetros a nuestro ajuste nos quita poder de estimación y futura predicción. Analizaremos el caso de la devianza residual.

**Cociente de Verosimilitud** Consideremos al modelo saturado (es decir el modelo que tiene el máximo número de parámetros  $m$ ) como  $M_s$  y un modelo  $M_2$  (con  $q$  parámetros, con  $q < m$ ). En el modelo saturado los  $m$  parámetros ajustan perfectamente a los datos, es decir, en el modelo saturado considera la estimación como algo determinista, no deja nada a la aleatoriedad, visto de otra forma, el modelo saturado deja toda la variabilidad a la parte sistemica y nada a la parte aleatorio y el modelo  $M_2$  sí considera una parte aleatoria, además la idea será reducir o simplificar la información, la representación mínima de  $M_2$  es el modelo nulo (este modelo sólo considera un parámetro para todas las observaciones).

Consideremos a  $\hat{\theta}_s$  como el estimador de todos los parámetros del modelo saturado, entonces  $L(\hat{\theta}_s, y)$  la verosimilitud evaluada en el estimador y sea  $\hat{\theta}_2$  el estimador de los parámetros del modelo reducido, por lo que  $L(\hat{\theta}_2, y)$  la verosimilitud evaluada en el estimador. El cociente de verosimilitudes sería:

$$\lambda = \frac{L(\hat{\theta}_s, y)}{L(\hat{\theta}_2, y)}$$

Aplicamos el logaritmo para obtener la diferencia de log-verosimilitudes con lo que tenemos:

$$\log(\lambda) = \ell(\hat{\theta}_s, y) - \ell(\hat{\theta}_2, y)$$

Si  $\log(\lambda)$  es muy grande, entonces la estimación por menos parámetros es mala, en cambio si  $\log(\lambda)$  es pequeño o cercano a 0 quiere decir que el ajuste por menos parámetros es bueno.

La devianza es una medida que fue inducida por Nelder y Wedderburn en 1972, la cual se define de manera general como:

$$D = 2 \left[ \ell(\hat{\theta}_s, y) - \ell(\hat{\theta}_2, y) \right]$$

$$D = 2 \sum_{i=1}^n a^{-1}(\phi) \left[ y_i (\hat{\theta}_{si} - \hat{\theta}_i) - b(\hat{\theta}_{si}) + b(\hat{\theta}_i) \right]$$

Sólo nos faltaría determinar la distribución asociada a  $D$  para decidir si el modelo reducido hace un buen ajuste o no.

$$D \sim \chi_{(n-q, \nu)}^2 \quad \text{con } \nu = 2 \left[ \ell(\hat{\beta}_s, y) - \ell(\hat{\beta}_2, y) \right]$$

Si  $\nu \approx 0$  entonces  $D$  se distribuye aproximadamente:

$$D \sim \chi_{(n-q)}^2$$

Recordemos que  $H_0$  sería el modelo reducido ajusta bien contra  $H_1$  sería el modelo reducido no hace un buen ajuste, por lo que rechazaríamos la hipótesis nula si  $D > \chi_{(m-q)}^{2(1-\alpha)}$ , si rechazamos

la hipótesis nula, quiere decir que el modelo saturado ajusta mejor que el modelo reducido, pero eso no nos garantiza que el ajuste sea correcto o bueno.

### 4.6.3. Ejemplos de modelos lineales generalizados

Como lo mencionamos antes, la distribución de la variable aleatoria  $Y_i$  va a depender del tipo de ajuste a aplicar

- Regresión Logística ( $Y_i \sim \text{Bernoulli}(p_i)$ ).
- Regresión Poisson ( $Y_i \sim \text{Poisson}(\mu_i)$ ).
- Modelo Log-lineal.
  - Multinomial (conteos fijos)
  - Poisson (conteos aleatorios) ( $Y_i \sim \text{Poisson}(n(i))$ ).

#### Regresión Logística

Este modelo es muy útil cuando las variables son binarias (0 ó 1) y sirve para poder clasificar a los sujetos dependiendo de sus características medidas.

**Modelo** Sean  $Y_1, \dots, Y_n$  las variables respuesta binarias, con  $p$  variables explicativas. La distribución asociada a  $Y_i$  es  $\text{Bernoulli}(p_i)$  y la función liga es la siguiente:

$$\log\left(\frac{\mathbb{E}(Y_i)}{1 - \mathbb{E}(Y_i)}\right) = \log\left(\frac{p_i}{1 - p_i}\right) = \beta_1 + \sum_{j=2}^p \beta_j X_{ij}$$

**Prueba de Wald** Esta prueba nos ayuda a ver la hipótesis nula  $H_0 : \beta_i = 0$  contra  $H_1 : \beta_i \neq 0$ , esta prueba no usa el cociente de verosimilitudes, esta se constreya apartir de aproximaciones distribucionales de los parámetros estimados, entonces el estadístico de *Wald* para cada parámetro es:

$$Z = \frac{\hat{\beta}_i}{\sqrt{\text{Var}(\hat{\beta}_i)}} \sim_{\text{aprox.}} N(0, 1)$$

Donde  $\hat{\beta}_i$  es el estimador máximo verosimil de  $\beta_i$ , entonces se rechaza la hipótesis nula si  $Z > Z^{1-\frac{\alpha}{2}}$  (donde  $Z^{1-\frac{\alpha}{2}}$  es el cuantil  $1 - \frac{\alpha}{2}$  de una distribución  $N(0, 1)$ ), o también podemos rechazar  $H_0$  si  $p\text{-value} < \alpha$ , casi siempre  $\alpha = 0.05$ , rechazar esta prueba nos estaría indicando que  $\beta_i$  es significativa para el modelo, es decir la variable asociada sí estaría explicando a nuestra variable respuesta.

**Devianza** Esta es una medida de asociación que sirve para ver que tan bueno es el ajuste de un modelo reducido con respecto al modelo saturado (todas las variables e iteraciones posibles). Para este modelo se usa la siguiente transformación:

$$\text{Sea } Z_i = \sum_{j=1}^{n_i} Y_j, \quad Z_i \sim \text{Bin}(n_i, p_i)$$

Entonces la devianza residual quedaría expresada como:

$$D = 2 \sum_{i=1}^n Z_i \ln \left( \frac{Z_i}{n_i \hat{p}_i} \right) + 2 \sum_{i=1}^n (n_i - Z_i) \ln \left( \frac{n_i - Z_i}{n_i (1 - \hat{p}_i)} \right)$$

En este caso la hipótesis nula se interpretaría como:

$H_0$  : Todos los coeficientes que están en el modelo saturado y no en el modelo reducido son cero.

vs

$H_1$  : Algun coeficiente que está en el modelo saturado y no en el modelo reducido no es cero.

Se rechaza la hipótesis nula si:  $D > \chi_{(n-q)}^{2(1-\alpha)}$  donde  $n$  es el número de observaciones y  $q$  el número de parámetros del modelo reducido.

**Prueba de Hosmer-Lemeshow** Esta prueba se desarrollo para los modelos de regresión logística, fue diseñada en el años 2000 por *Hosmer* y *Lemeshow*, la prueba es la siguiente:

$H_0$  : El ajuste de nuestros datos es pertinente con los parametros establecidos.

vs

$H_1$  : El ajuste está mal planteado con los parametros dados.

Basicamente se ve la eficiencia del ajuste, es decir se evalúa la diferencia entre el valor observado y el ajustado, el estadístico  $X^2$  se define como:

$$X^2 = \sum_{i=1}^n \frac{Z_i - n_i \hat{p}_i}{n_i \hat{p}_i (1 - \hat{p}_i)}$$

Donde  $X^2$  tiene una distribución  $\chi_{q-2}^2$ . Entonces se rechaza  $H_0$  si el estadístico  $X^2 > \chi_{q-2}^{2(1-\alpha)}$  con  $q$  el número de parámetros con los que se hizo el ajuste (incluido el intercepto).

**Pseudo  $R^2$**  Esta medida nos da una idea de que porcentaje de variabilidad total es explicada por nuestro modelo planteado, se calcula de la siguiente manera:

$$R_{pseudo}^2 = \frac{D_0 - D}{D_0}$$

Donde  $D_0$  es la devianza nula (devianza del modelo ajustado sólo con el intercepto) y  $D$  es la devianza del modelo reducido que calculamos anteriorme. Esta es una medida aproximada que nos dice que tan bueno es el ajuste propuesto con los parámetros seleccionados.

**Ejemplo 4.6.1.** *Los datos en la base ejrl.sav corresponden a 700 clientes a los que se les negó o no se les negó un crédito, además a cada cliente se le midieron distintas variables socioeconómicas, algunas de las cuales influyeron en esa decisión, la base contiene los siguientes campos:*

- *Age: Edad medida en años cumplidos.*
- *Ed: Nivel educativo.*
  - *1: No termino el High – school.*
  - *2: Se gradúo del High – school.*
  - *3: Algún tipo de educación superior.*
  - *4: Término la educación superior.*
  - *5: Posgrado finalizado.*
- *Employ: Número de años en el empleo actual.*
- *Addrres: Número de años viviendo en la dirección actual.*
- *Income: Ingresos del hogar medidos en miles de dólares.*
- *Debtinc: Tasa de deuda contra ingresos por 100.*
- *Creddebt: Deudas con las tarjetas de crédito medido ed miles de dólares.*
- *Othdebt: Otras deudas en miles de dólares.*
- *Default: Crédito negado anteriormente.*
  - *0: No*
  - *1: Sí*

*La finalidad de un nuevo análisis es determinar una regla de decisión que permita establecer para nuevos clientes si se les otorga o no el crédito.*

```

#Cargamos librerias
library(foreign)
library(MASS) #Para la funci\on stepAIC, que hace el m\etodo de seleccion Backward
library(ResourceSelection) #Para poder hacer la prueba de Hosmer-Lemeshow

## ResourceSelection 0.3-2 2017-02-28

#Cargaamos la base de datos
base = read.spss("ejrl.sav",to.data.frame=TRUE)

## re-encoding from CP1252

#Hacemos las variables explicativas como factores
base$ed<-factor(base$ed)
base$default<-factor(base$default)
#Liberamos las variables de la base de datos
attach(base)

#Ajustamos un modelo inicial que contenga a todas las variables explicativas
model1<-glm(default~ age+ed+employ+address+income
             +debtinc+creddebt+othdebt,binomial)

#Observamos los resultados
summary(model1)

##
## Call:
## glm(formula = default ~ age + ed + employ + address + income +
##      debtinc + creddebt + othdebt, family = binomial)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.4321  -0.6455  -0.2896   0.2802   3.0250
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -1.590424   0.605329  -2.627   0.0086 **
## age             0.035497   0.017586   2.018   0.0435 *
## edHigh school degree  0.307643   0.251628   1.223   0.2215
## edSome college  0.352652   0.339949   1.037   0.2996
## edCollege degree -0.085158   0.472917  -0.180   0.8571
## edPost-undergraduate degree  0.876191   1.293810   0.677   0.4983
## employ        -0.260676   0.033407  -7.803 6.04e-15 ***
## address       -0.105429   0.023263  -4.532 5.84e-06 ***
## income        -0.007823   0.007785  -1.005   0.3149
## debtinc        0.070704   0.030596   2.311   0.0208 *
## creddebt       0.624980   0.112917   5.535 3.11e-08 ***
## othdebt        0.052999   0.078467   0.675   0.4994

```

```

## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 804.36  on 699  degrees of freedom
## Residual deviance: 549.56  on 688  degrees of freedom
## AIC: 573.56
##
## Number of Fisher Scoring iterations: 6

#Aparentemente la variable de nivel educativo, ingreso del hogar y otro tipo de deuda
#no son significativas para determinar si a una persona se le otorga o no un cr\edito.

#Usaremos el m\etodo Backward para determinar cuales de estas variables son significativas
modelo <- stepAIC(model1, direction=c("both"))

## Start:  AIC=573.56
## default ~ age + ed + employ + address + income + debtinc + creddebt +
##      othdebt
##
##           Df Deviance   AIC
## - ed       4   552.21 568.21
## - othdebt   1   550.02 572.02
## - income   1   550.50 572.50
## <none>     549.56 573.56
## - age      1   553.61 575.61
## - debtinc  1   554.90 576.90
## - address  1   571.85 593.85
## - creddebt 1   594.07 616.07
## - employ   1   634.47 656.47
##
## Step:  AIC=568.21
## default ~ age + employ + address + income + debtinc + creddebt +
##      othdebt
##
##           Df Deviance   AIC
## - othdebt   1   553.02 567.02
## - income    1   553.02 567.02
## <none>     552.21 568.21
## - age       1   555.95 569.95
## - debtinc   1   556.80 570.80
## + ed        4   549.56 573.56
## - address   1   573.92 587.92
## - creddebt  1   597.37 611.37
## - employ    1   652.19 666.19

```

```

##
## Step: AIC=567.02
## default ~ age + employ + address + income + debtinc + creddebt
##
##           Df Deviance   AIC
## - income   1   553.18 565.18
## <none>           553.02 567.02
## + othdebt   1   552.21 568.21
## - age       1   556.72 568.72
## + ed        4   550.02 572.02
## - debtinc   1   571.88 583.88
## - address   1   574.46 586.46
## - creddebt  1   597.93 609.93
## - employ    1   653.02 665.02
##
## Step: AIC=565.18
## default ~ age + employ + address + debtinc + creddebt
##
##           Df Deviance   AIC
## <none>           553.18 565.18
## - age       1   556.73 566.73
## + income    1   553.02 567.02
## + othdebt   1   553.02 567.02
## + ed        4   550.51 570.51
## - address   1   575.00 585.00
## - debtinc   1   576.98 586.98
## - creddebt  1   618.56 628.56
## - employ    1   661.59 671.59

#Observamos los pasos que siguió para asignar el modelo, este método se basa
#en la minimización del AIC, medida que nos da una referencia sobre el ajuste
#mientras más pequeño sea el AIC será mejor el ajuste.
#Observamos los nuevos resultados
summary(modelo)

##
## Call:
## glm(formula = default ~ age + employ + address + debtinc + creddebt,
##      family = binomial)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.3555  -0.6521  -0.2949   0.2592   2.9132
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)

```



```

## (Intercept) -1.63128    0.51268   -3.182   0.00146 **
## age          0.03256    0.01717    1.896   0.05799 .
## employ      -0.26076    0.03011   -8.662 < 2e-16 ***
## address     -0.10365    0.02309   -4.490 7.13e-06 ***
## debtinc     0.08926    0.01855    4.813 1.49e-06 ***
## creddebt    0.57265    0.08723    6.565 5.20e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 804.36 on 699 degrees of freedom
## Residual deviance: 553.18 on 694 degrees of freedom
## AIC: 565.18
##
## Number of Fisher Scoring iterations: 6

```

*#En este modelo las variables parecen ser significativas, aunque la variable edad es significativa a un  $\alpha=0.1$ , pero con un  $\alpha=0.05$  dejaría de ser significativa*

*#Usaremos la Devianza residual para corroborar si el último modelo obtenido es adecuado  
#Calculamos el cuantil 0.95 de una distribución ji-cuadrada con 694 grados de libertad  
#Los grados de libertad son  $n=700$  observaciones menos el número de parámetros en el ajuste  
#es decir 5 variables más el intercepto, es decir 6 parámetros*

```

qchisq(0.95,694)

```

```
## [1] 756.3962
```

*#Como  $D=553.18 < 756.3962$  entonces no se rechaza la hipótesis nula, es decir el modelo reducido hace un buen ajuste, así mismo podríamos decir que los coeficientes de los demás parámetros que no consideramos son cero.*

*#Ahora, analizaremos la devianza nula, la hipótesis nula es la misma que en el caso anterior, es decir, queremos ver si todos los parámetros son cero, a excepción de  $\beta_1$  que es el intercepto, alternativamente si aceptamos la hipótesis nula estaríamos diciendo que el modelo saturado y el modelo con tan solo el intercepto son similares y que no tendría sentido práctico ajustar con variables explicativas a nuestra variable respuesta, bastaría solo ajustar con una constante.*

```
1-pchisq(804.36,699)
```

```
## [1] 0.003413233
```

*#como el p-value de esta prueba es  $< 0.05$  se rechaza la hipótesis nula.  
#Es decir, nuestro modelo sí debe de tener variables explicativas, estas serán  
#las que se encontraron vía Backward (intercepto, edad, años en el mismo trabajo,  
#años viviendo en la misma casa, tasa de deuda contra ingresos y deuda de crédito)*

```

#esas variables parecen explicar bien a nuestra variable respuesta que indica
#negarle o no el cr\ 'edito.

#Otra prueba de bondad de ajuste es la de Hosmer-Lemeshow, cuya hip\ 'otesis
#nula es que el modelo ajusta bien a los datos. Esta es una versi\ 'on
#mejorada de una prueba ji-cuadrada, hecha anteriormente variable por variable.
hoslem.test(modelo$y,fitted(modelo),6)

##
## Hosmer and Lemeshow goodness of fit (GOF) test
##
## data: modelo$y, fitted(modelo)
## X-squared = 4.8135, df = 4, p-value = 0.307

#Como el p-value asociado es 0.307>0.05 no rechazamos la hip\ 'otesis nula y
#se comprueba que el modelo ajusta a los datos.

#El pseudo R^2 puede calcularse como:
pseudoR<-(modelo$null.deviance-deviance(modelo))/modelo$null.deviance
pseudoR

## [1] 0.3122816

#Vale 0.3122816, este valor nos sirve como referencia para ver que tan bueno
#es el ajuste, este modelo "explica" m\ 'as del 30% de la variabilidad total
#de nuestros datos originales.

#As\ 'i mismo podemos calcular la exponencial de los coeficientes, que es
#lo que se interpreta
exp(coefficients(modelo))

## (Intercept)      age      employ      address      debtinc      creddebt
## 0.1956787  1.0330933  0.7704649  0.9015443  1.0933599  1.7729658

#Los intervalos de confianza
confint(modelo)

## Waiting for profiling to be done...

##              2.5 %      97.5 %
## (Intercept) -2.643050697 -0.62921296
## age         -0.001294816  0.06622749
## employ      -0.322259213 -0.20406310
## address     -0.149967314 -0.05928079
## debtinc     0.053210015  0.12605289
## creddebt    0.411887700  0.75328617

```

```

#como era de esperarse la \ 'unica variable que contiene al 0 dentro de su intervalo
#de confianza es la variable edad (esto por el p-valor asociado anteriormente)
exp(confint(modelo))

## Waiting for profiling to be done...

##           2.5 %    97.5 %
## (Intercept) 0.0711439 0.5330111
## age         0.9987060 1.0684698
## employ      0.7245104 0.8154109
## address     0.8607361 0.9424421
## debtinc     1.0546511 1.1343422
## creddebt    1.5096649 2.1239683

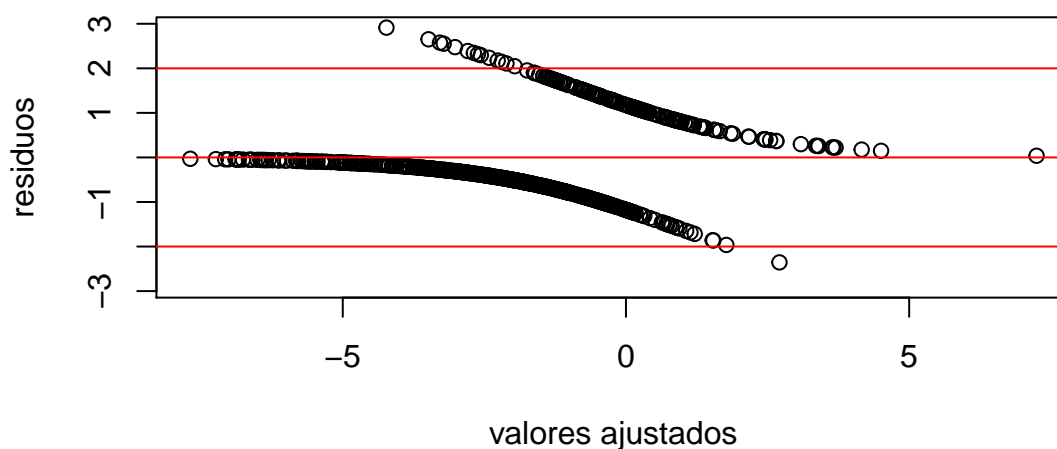
#Los valores ajustados
base$vajustados<-fitted(modelo)

#Analisis de residuos
p=predict(modelo) #valores de los logits ajustados
r=residuals(modelo)

plot(p,r,xlab="valores ajustados",ylab="residuos",ylim=c(-1,1)*max(abs(r)))
abline(0,0,col="red");abline(2,0,col="red");abline(-2,0,col="red")
title("Valores ajustados frente a residuos",col.main="red")

```

## Valores ajustados frente a residuos



```

#donde podemos ver que son pocas variables donde son significativos,
#es decir mayores que el valor absoluto de 2

#Para predecir

```

```

#Establecemos el punto de corte en p-gorro=0.5, entonces asignamos al
#grupo NO (cuando no se le neg\ 'o el cr\ 'edito) si p-gorro<0.5 (donde p es la
#probabilidad de \ 'exito (en este caso el \ 'exito es que se le negar\ 'a el cr\ 'edito
#anteriormente) y en caso contrario se asigno al grupo YES
base$grupo[base$vajustados < 0.5] <- "NO"
base$grupo[base$vajustados >= 0.5] <- "YES"
tablaclasif<-table(base$default,base$grupo)
tablaclasif

##
##      NO YES
## No  476  41
## Yes  89  94

#0 en porcentajes por renglones tenemos
prop.table(tablaclasif,1)

##
##      NO      YES
## No  0.92069632 0.07930368
## Yes 0.48633880 0.51366120

#Entonces podemos observar que la tasa de no negaci\ 'on del cr\ 'edito parece correcta
#ya que es superior al 90% por lo que se deduce que no se niegan los cr\ 'editos de
#manera oportuna, en cambio podemos ver que la tasa de negaci\ 'on de cr\ 'edito es
#apenas superior al 50%, entonces se niegan cr\ 'editos de manera poco eficiente
#a personas que tal vez no se les deber\ 'ia negar.
#Lo que nos hubiera gustado es que la clasificaci\ 'on fuera mucho mejor, pero
#como s\ 'olo explicamos m\ 'as de 30% de la variabilidad total de los datos,
#provoca que la clasificaci\ 'on no sea tan buena a favor de la negaci\ 'on del cr\ 'edito.

#predicci\ 'on de valores nuevos
#supongamos que llega un nuevo cliente tiene 37 años, que no termin\ 'o la
#preparatoria, tiene 20 años en el mismo empleo, 13 en la misma direcci\ 'on,
#un ingreso (en miles) de 41, una raz\ 'on de deuda e ingreso de 12.9, una
#deuda en su tarjeta de cr\ 'edito (en miles) de 0.9 y en otras deudas de 4.49
#la pregunta ser\ 'ia ¿se le niega o no el cr\ 'edito?
age<-37
employ<-20
address<-13
debtinc<-12.9
creddebt<-0.9
ndatos1=data.frame(age,employ,address,debtinc,creddebt)
#datos con valores en las variables independientes
predict(modelo, newdata = ndatos1,type = "response")

```

```
##          1
## 0.004856852
```

*#0.004856852 de probabilidad de que se le otorguen el credito, por lo tanto no se le dar\`ia el cr\`edito.*

## Regresión Poisson

Este tipo de regresión no esta asociada a una tabla de contingencia como en el caso de los modelos *log – lineales*, aquí las variables explicativas son continuas y categóricas. Este tipo de modelo es excelente para poder interpretar y predecir tasas como variables respuesta.

**Modelo** Sean  $Y_1, \dots, Y_n$  las variables respuesta que hacen referencia a conteos, con  $p$  variables explicativas (Incluyendo la constante). La distribución asociada a  $Y_i$  es *poisson* y la función liga es la siguiente:

$$\log(\mathbb{E}(Y_i)) = \beta_1 + \sum_{j=2}^p \beta_j X_{ij}$$

**Devianza** Esta es una medida de asociación que sirve para ver que tan bueno es el ajuste del modelo reducido con respecto al modelo saturado (todas las variables e iteraciones posibles). Para este modelo se ocupa la siguiente formula:

$$D = 2 \sum_{\{i|n(i)>0\}} n(i) \ln \left( \frac{n(i)}{\hat{m}(i)} \right)$$

con  $n(i)$  el número de observaciones sobre la variable  $i$

**Ejemplo 4.6.2.** *Ajuste un modelo de regresión Poisson con los datos de la base basecancermodelopoisson.csv, en este caso se busca modelar el número de pacientes con cáncer en una población, donde se buscan encontrar las variables que podrian determinar esta enfermedad.*

```
#Lee los archivos .csv y los almacena

basecancer<-read.table("base cancer modelo poisson.csv",sep="," ,header=TRUE)

#Cargar la librer\`ias

library(nortest) #Para la prueba de normalidad

#Nombres de las variables

names(basecancer)
```

```

## [1] "POBLA" "ANALFABE" "PRIMARIA" "DRENAJE" "ELECTRIC" "AGUAPOTA"
## [7] "HACINAMI" "PISOTIER" "MENOSDES" "DOSSALAR" "INDICE" "EDAD"
## [13] "SEXO" "AGRUPADO" "OBSERVAD" "POBEDADS" "POBNACSE" "GRADO"
## [19] "LOG_POBE" "STDPEARS"

#Convertir las variables EDAD (gpos de edad), SEXO, AGRUPADO (años
#agrupados), GRADO (grado de marginaci\on localidad a la que pertenece
#en grupos) en factores. La \unica variable continua que incluye el modelo
#es ANALFABE (% de analfabetismo del municipio a la que pertenecen
#un grupo de individuos con el mismo valor en sus variables explicativas)

basecancer$EDAD<-factor(basecancer$EDAD)
basecancer$SEXO<-factor(basecancer$SEXO)
basecancer$AGRUPADO<-factor(basecancer$AGRUPADO)
basecancer$GRADO<-factor(basecancer$GRADO)

#Fijamos la base de datos

attach(basecancer)

#Ajustamos el modelo con las variables EDAD, SEXO, AGRUPADO, GRADO,
#ANALFABE y la interacci\on GRADO*SEXO. La respuesta es OBSERVAD
#(defunciones totales para cada combinaci\on de valores de las variables
#explicativas en cada municipio) y offset dado por el logaritmo natural
#de la pob. total por edad y sexo para ese municipio (variable LOG_POBE).

model1<-glm(OBSERVAD~EDAD+SEXO+AGRUPADO+GRADO+ANALFABE+GRADO:SEXO+offset(LOG_POBE),poisson)
summary(model1)

##
## Call:
## glm(formula = OBSERVAD ~ EDAD + SEXO + AGRUPADO + GRADO + ANALFABE +
## GRADO:SEXO + offset(LOG_POBE), family = poisson)
##
## Deviance Residuals:
## Min 1Q Median 3Q Max
## -5.8049 -0.6497 -0.3467 -0.1416 17.4369
##
## Coefficients:
## Estimate Std. Error z value Pr(>|z|)
## (Intercept) -10.389789 0.028428 -365.477 < 2e-16 ***
## EDAD2 0.317232 0.027414 11.572 < 2e-16 ***
## EDAD3 0.229304 0.027182 8.436 < 2e-16 ***
## EDAD4 0.288840 0.027165 10.633 < 2e-16 ***
## EDAD5 0.573413 0.026743 21.442 < 2e-16 ***
## SEXO2 -0.205175 0.014705 -13.953 < 2e-16 ***

```

```
## AGRUPADO4      0.087490    0.013869     6.309 2.82e-10 ***
## AGRUPADO5      0.161983    0.013779    11.756 < 2e-16 ***
## AGRUPADO6      0.199101    0.013716    14.517 < 2e-16 ***
## GRADO2         0.347787    0.019194    18.119 < 2e-16 ***
## GRADO3         0.542025    0.022090    24.537 < 2e-16 ***
## GRADO4         0.549323    0.027877    19.705 < 2e-16 ***
## GRADO5         0.311650    0.050871     6.126 9.00e-10 ***
## ANALFABE      -0.008952    0.001341    -6.678 2.43e-11 ***
## SEXO2:GRADO2  -0.005678    0.027546    -0.206  0.837
## SEXO2:GRADO3  -0.012978    0.027952    -0.464  0.642
## SEXO2:GRADO4   0.041565    0.026914     1.544  0.123
## SEXO2:GRADO5   0.044534    0.046876     0.950  0.342
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## (Dispersion parameter for poisson family taken to be 1)
##
```

```
## Null deviance: 104732 on 98138 degrees of freedom
## Residual deviance: 101335 on 98121 degrees of freedom
## (21 observations deleted due to missingness)
## AIC: 150349
##
```

```
## Number of Fisher Scoring iterations: 6
```

*#La devianza residual es pequeña (101,335) con respecto a los grados de libertad (98121) (inferior a dos veces los gl) lo cual indica intuitivamente que no se rechaza que el modelo ajusta bien a los datos. Formalmente calculamos el p-value como*

```
1-pchisq(101335,98121)
```

```
## [1] 3.511635e-13
```

*#El cual es casi cero así que como es menor que cualquier nivel de significancia alfa usual rechazar la hipótesis nula de que el modelo ajusta bien a los datos. Sin embargo, aquí el problema es que el tamaño de la muestra es demasiado grande, así que los grados de libertad también en lo son. Así que tal vez es mejor no utilizar la prueba correspondiente que la devianza residual que compara con el modelo saturado. También viene la devianza nula, la cual compara al modelo con un modelo con solo término constante, así que la hipótesis nula es que el modelo sin variables explicativas y solo con término cte. es similar. Aquí me parece bien lo que convendría es comparar los modelos con solo constante y con todas las variables.*

```
modelcte<-glm(OBSERVAD~1+offset(LOG_POBE),poisson)
summary(modelcte)
```

```

##
## Call:
## glm(formula = OBSERVAD ~ 1 + offset(LOG_POBE), family = poisson)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q        Max
## -5.9146  -0.6276  -0.3429  -0.1431   18.6256
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -9.870152   0.004788  -2061  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 104732  on 98138  degrees of freedom
## Residual deviance: 104732  on 98138  degrees of freedom
## (21 observations deleted due to missingness)
## AIC: 153711
##
## Number of Fisher Scoring iterations: 6

anova(modelcte,model1, test="Chi")

## Analysis of Deviance Table
##
## Model 1: OBSERVAD ~ 1 + offset(LOG_POBE)
## Model 2: OBSERVAD ~ EDAD + SEXO + AGRUPADO + GRADO + ANALFABE + GRADO:SEXO +
##      offset(LOG_POBE)
##   Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
## 1      98138      104732
## 2      98121      101335 17   3396.9 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

#El p-valor indica rechazar la hip\otesis nula de que el modelo con solo
#la constante y el modelo con esta y todas las variables explicativas
#sean similares. O sea vale la pena agregar las variables.

summary(model1)

##
## Call:
## glm(formula = OBSERVAD ~ EDAD + SEXO + AGRUPADO + GRADO + ANALFABE +
##      GRADO:SEXO + offset(LOG_POBE), family = poisson)

```



```

##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -5.8049  -0.6497  -0.3467  -0.1416  17.4369
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -10.389789  0.028428 -365.477 < 2e-16 ***
## EDAD2        0.317232  0.027414  11.572 < 2e-16 ***
## EDAD3        0.229304  0.027182   8.436 < 2e-16 ***
## EDAD4        0.288840  0.027165  10.633 < 2e-16 ***
## EDAD5        0.573413  0.026743  21.442 < 2e-16 ***
## SEX02       -0.205175  0.014705 -13.953 < 2e-16 ***
## AGRUPADO4    0.087490  0.013869   6.309 2.82e-10 ***
## AGRUPADO5    0.161983  0.013779  11.756 < 2e-16 ***
## AGRUPADO6    0.199101  0.013716  14.517 < 2e-16 ***
## GRADO2       0.347787  0.019194  18.119 < 2e-16 ***
## GRADO3       0.542025  0.022090  24.537 < 2e-16 ***
## GRADO4       0.549323  0.027877  19.705 < 2e-16 ***
## GRADO5       0.311650  0.050871   6.126 9.00e-10 ***
## ANALFABE    -0.008952  0.001341  -6.678 2.43e-11 ***
## SEX02:GRADO2 -0.005678  0.027546  -0.206  0.837
## SEX02:GRADO3 -0.012978  0.027952  -0.464  0.642
## SEX02:GRADO4  0.041565  0.026914   1.544  0.123
## SEX02:GRADO5  0.044534  0.046876   0.950  0.342
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 104732  on 98138  degrees of freedom
## Residual deviance: 101335  on 98121  degrees of freedom
##      (21 observations deleted due to missingness)
## AIC: 150349
##
## Number of Fisher Scoring iterations: 6

#Vienen los estimadores de los par\ 'ametros y vemos como ninguno de los
#par\ 'ametros asociados a las interacciones son significativos as\ 'i que
#quitamos la interacci\ 'on del modelo.

#Ajustamos el modelo sin la interacci\ 'on

model2<-glm(OBSERVAD~EDAD+SEXO+AGRUPADO+GRADO+ANALFABE+offset(LOG_POBE),poisson)
summary(model2)

```

```

##
## Call:
## glm(formula = OBSERVAD ~ EDAD + SEXO + AGRUPADO + GRADO + ANALFABE +
##      offset(LOG_POBE), family = poisson)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q        Max
## -5.7941  -0.6503  -0.3469  -0.1415   17.4548
##
## Coefficients:
##              Estimate Std. Error  z value Pr(>|z|)
## (Intercept) -10.392656   0.028022 -370.876 < 2e-16 ***
## EDAD2         0.317240   0.027414  11.572 < 2e-16 ***
## EDAD3         0.229307   0.027182   8.436 < 2e-16 ***
## EDAD4         0.288821   0.027165  10.632 < 2e-16 ***
## EDAD5         0.573383   0.026743  21.441 < 2e-16 ***
## SEXO2        -0.198586   0.009667 -20.542 < 2e-16 ***
## AGRUPADO4     0.087503   0.013869   6.309 2.80e-10 ***
## AGRUPADO5     0.162004   0.013779  11.757 < 2e-16 ***
## AGRUPADO6     0.199145   0.013715  14.520 < 2e-16 ***
## GRADO2        0.345311   0.015082  22.895 < 2e-16 ***
## GRADO3        0.536358   0.018521  28.960 < 2e-16 ***
## GRADO4        0.567418   0.025258  22.465 < 2e-16 ***
## GRADO5        0.331043   0.046511   7.118 1.10e-12 ***
## ANALFABE     -0.008949   0.001341  -6.675 2.47e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 104732  on 98138  degrees of freedom
## Residual deviance: 101339  on 98125  degrees of freedom
##      (21 observations deleted due to missingness)
## AIC: 150345
##
## Number of Fisher Scoring iterations: 6

```

*#Una prueba formal para ver si la interacci\on es significativa consiste en comparara el model1 (con la interacci\on) con el model 2 (sin la interacci\on) a trav\es de los cocientes de verosimilitud correspondientes, esto es a trav\es de la diferencia de devianza entre uno y otro modelo. #La hip\otesis nula ser\ia entonces que el modelo sin interacci\on es similar al modelo con la interacci\on, en cuyo caso convendr\ia quitar la interacci\on porque no aporta nada al modelo, y entonces lo que queremos es #no rechazar esta prueba (que el p valor supere a alfa). As\i que HO la podemos #pensar como que no hay efecto de la interacci\on contra la alternativa de que s\i hay efecto.*

```

#En este caso pvalue=0.3687>0.05 as\ 'i que no rechazamos la hip\ 'otesis nula
#y no hay efecto de la interacci\ 'on, as\ 'i que hay que quitarla. Ver que
#el p valor coincide con el pvalor que daba SPSS para el estad\ 'istico
#de Wald (0.368)

anova(model1,model2, test="Chi")

## Analysis of Deviance Table
##
## Model 1: OBSERVAD ~ EDAD + SEXO + AGRUPADO + GRADO + ANALFABE + GRADO:SEXO +
##   offset(LOG_POBE)
## Model 2: OBSERVAD ~ EDAD + SEXO + AGRUPADO + GRADO + ANALFABE + offset(LOG_POBE)
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1      98121      101335
## 2      98125      101339 -4  -4.2855   0.3687

#Vemos cu\ 'antos son los niveles de cada variable

levels(EDAD)

## [1] "1" "2" "3" "4" "5"

levels(SEXO)

## [1] "1" "2"

levels(AGRUPADO)

## [1] "3" "4" "5" "6"

levels(GRADO)

## [1] "1" "2" "3" "4" "5"

#Y en el modelo usamos la \ 'ultima categor\ 'ia (en n\ 'umero no en valor) en
#cada caso como referencia por ej en AGRUPADO es la 4

model3<-glm(OBSERVAD~relevel(EDAD,5)+relevel(SEXO,2)+relevel(AGRUPADO,4)+
            relevel(GRADO,5)+ANALFABE+offset(LOG_POBE),poisson)

summary(model3)

##
## Call:
## glm(formula = OBSERVAD ~ relevel(EDAD, 5) + relevel(SEXO, 2) +
##   relevel(AGRUPADO, 4) + relevel(GRADO, 5) + ANALFABE + offset(LOG_POBE),
##   family = poisson)
##

```

```

## Deviance Residuals:
##      Min        1Q    Median        3Q        Max
## -5.7941  -0.6503  -0.3469  -0.1415  17.4548
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -9.487671   0.051757 -183.313 < 2e-16 ***
## relevel(EDAD, 5)1  -0.573383   0.026743  -21.441 < 2e-16 ***
## relevel(EDAD, 5)2  -0.256143   0.013763  -18.611 < 2e-16 ***
## relevel(EDAD, 5)3  -0.344076   0.013285  -25.899 < 2e-16 ***
## relevel(EDAD, 5)4  -0.284563   0.013256  -21.467 < 2e-16 ***
## relevel(SEXO, 2)1   0.198586   0.009667   20.542 < 2e-16 ***
## relevel(AGRUPADO, 4)3 -0.199145   0.013715  -14.520 < 2e-16 ***
## relevel(AGRUPADO, 4)4 -0.111643   0.013361   -8.356 < 2e-16 ***
## relevel(AGRUPADO, 4)5 -0.037141   0.013262   -2.801  0.0051 **
## relevel(GRADO, 5)1  -0.331043   0.046511   -7.118 1.10e-12 ***
## relevel(GRADO, 5)2   0.014268   0.042035    0.339  0.7343
## relevel(GRADO, 5)3   0.205316   0.037546    5.468 4.54e-08 ***
## relevel(GRADO, 5)4   0.236375   0.031128    7.594 3.11e-14 ***
## ANALFABE           -0.008949   0.001341   -6.675 2.47e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 104732  on 98138  degrees of freedom
## Residual deviance: 101339  on 98125  degrees of freedom
##      (21 observations deleted due to missingness)
## AIC: 150345
##
## Number of Fisher Scoring iterations: 6

```

*#En este caso s\ 'i se pueden interpretar los par\ 'ametros o su exponencial*

```
exp(coefficients(model3))
```

```

##      (Intercept)      relevel(EDAD, 5)1      relevel(EDAD, 5)2
##      0.0000757804      0.5636152811      0.7740311294
##      relevel(EDAD, 5)3      relevel(EDAD, 5)4      relevel(SEXO, 2)1
##      0.7088748830      0.7523432880      1.2196770079
## relevel(AGRUPADO, 4)3 relevel(AGRUPADO, 4)4 relevel(AGRUPADO, 4)5
##      0.8194308208      0.8943637485      0.9635402654
##      relevel(GRADO, 5)1      relevel(GRADO, 5)2      relevel(GRADO, 5)3
##      0.7181744119      1.0143699597      1.2279124253
##      relevel(GRADO, 5)4      ANALFABE
##      1.2666490383      0.9910913188

```

```

#Por ejemplo el coef. estimado del grupo de edad 1 indica que en el
#grupo de edad inferior la tasa de defunci'on por cancer es casi la mitad
#respecto a la tasa de defunci'on que se tiene en el \ultimo gpo de edad.
#Y as'i para los otros t'erminos. No hay sobredispersi'on porque el
#cociente entre la devianza residual y sus gl es 101339/98125=1.032.
#Esto es casi uno

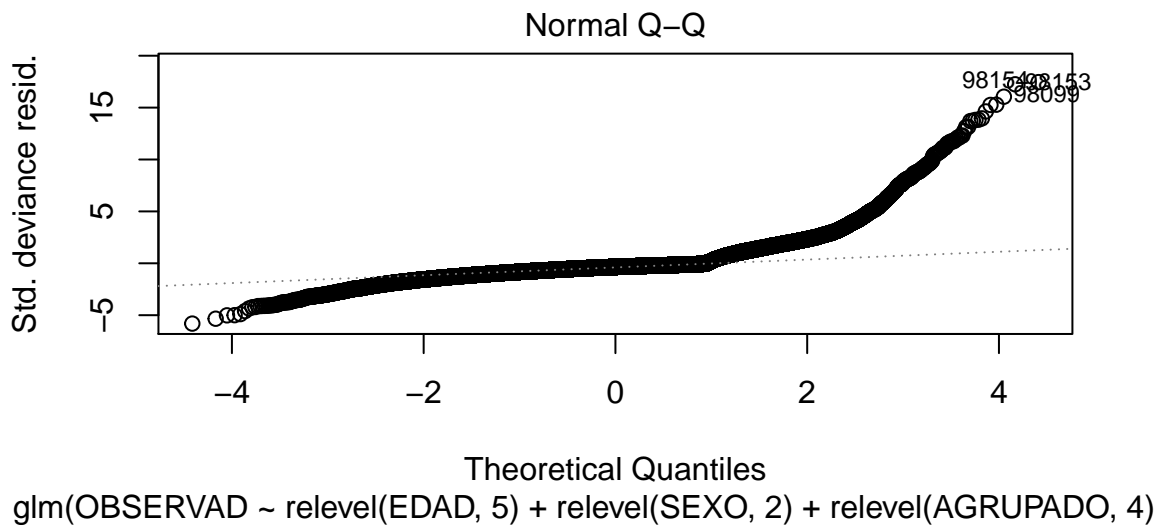
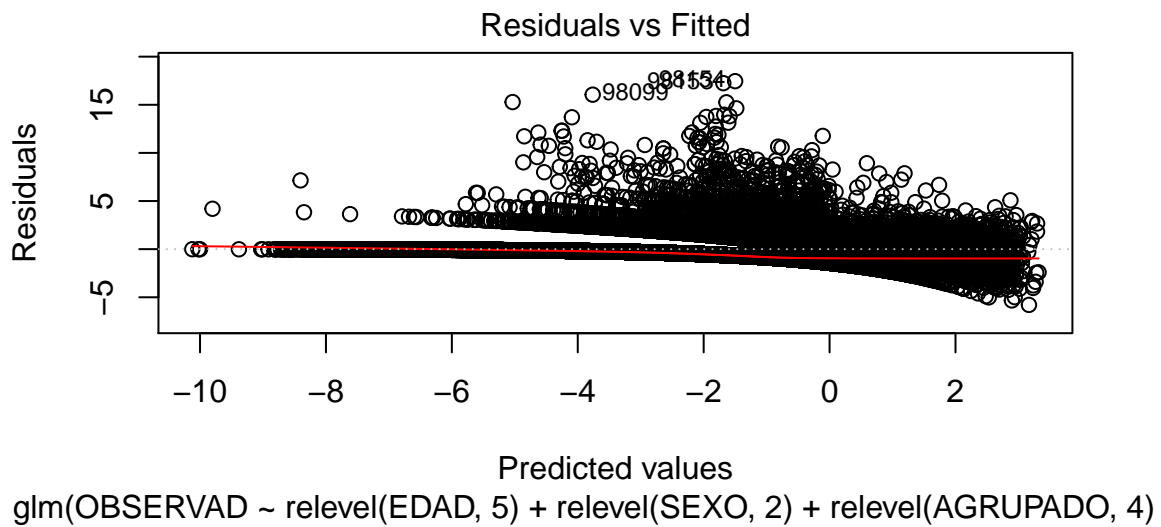
```

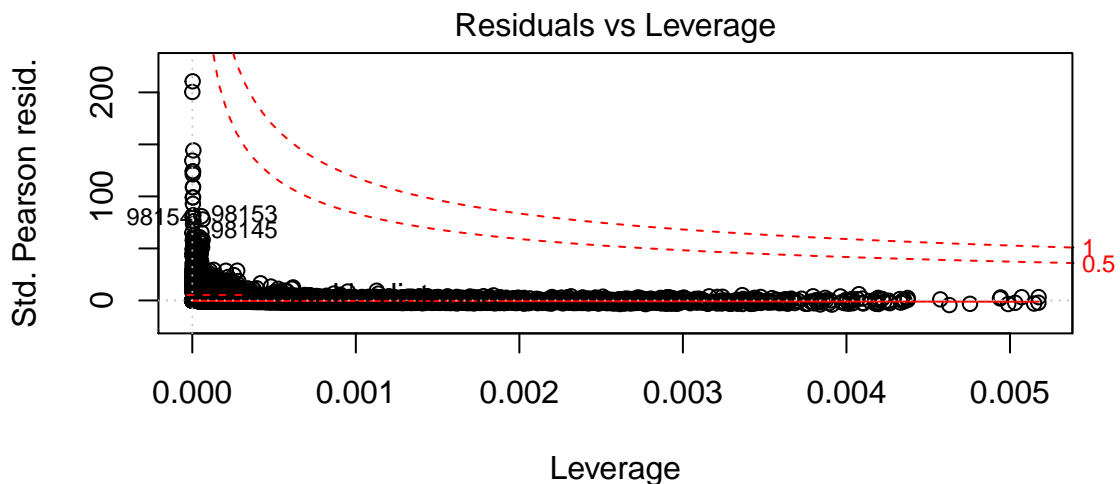
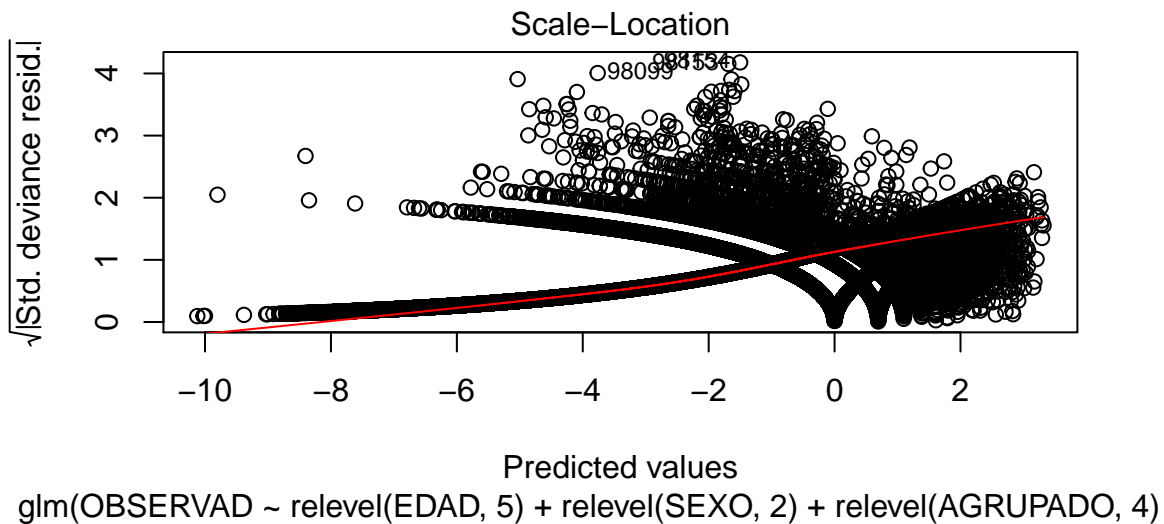
```

#Podemos verificar qu'e tan bien ajusta el modelo o si hay que hacer
#alg'un cambio con la gr'afica en papel normal de los residuales y la
#gr'afica de los valores predecidos contra residuales:

```

```
plot(model3)
```





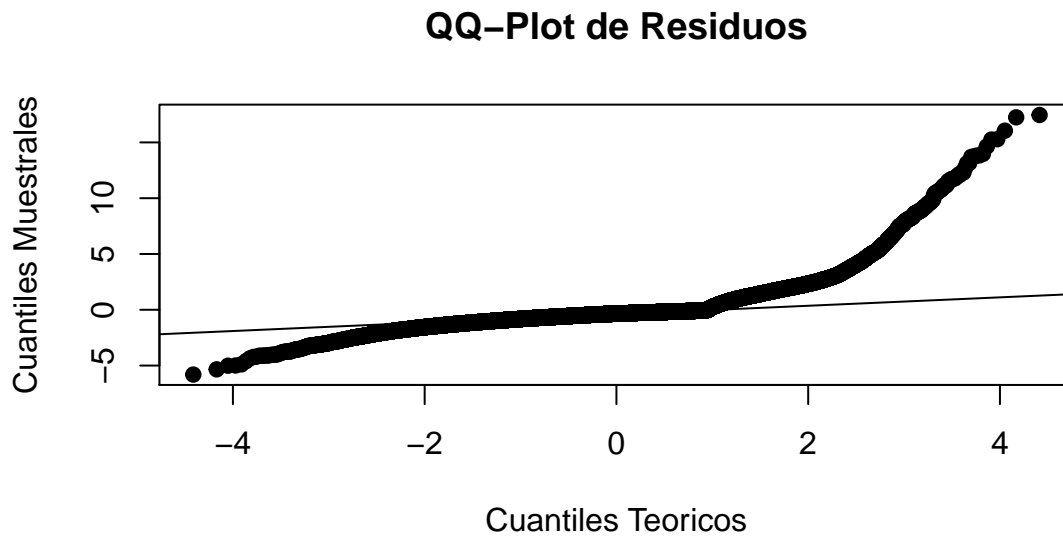
*#No se satisface normalidad y hay estructuras raras en la gr\afica que  
 #se usa para verificar varianza constante (valores predecidos contra  
 #residuales) as\i que tal vez pudiera ser conveniente cambiar la funci\on  
 #liga. Sin embargo, en este caso en particular la problem\atica es m\as  
 #probable que se resuelva al hacer un suavizamiento al considerar que  
 #al tratarse de una enfermedad rara (es un c\ancer espec\ifico) hay muchos  
 #valores que son muy peque\os en los datos.*

*#Por supuesto que otra posibilidad era tratar a los residuales por separado*

```
resestandar<-rstandard(model3)
```

```
##Gr\'afica probabilidad normal (QQ plot)

qqnorm(resestandar,main="QQ-Plot de Residuos", pch=19,
       xlab="Cuantiles Teoricos", ylab="Cuantiles Muestrales")
qqline(resestandar)
```



```
##Formalmente usar la prueba Anderson-Darling, la cual indica rechazar normalidad

ad.test(resestandar)

##
## Anderson-Darling normality test
##
## data: resestandar
## A = 6872.3, p-value < 2.2e-16

##Es m\as que claro que nuestros datos no son normales pues tiene un
#p valor muy reducido
```

**Ejercicio 4.6.1.** Una compañía entrevista a 230 clientes potenciales y registra su edad (categorizada), género (0 es hombre, 1 es mujer), estado civil (0 es soltero, 1 es casado) y si su estilo de vida es activo (1) o no (0) para determinar que tipo de desayuno prefieren entre dos opciones (0 es omelet y 1 es cereal) que se les dan a probar. Los datos se encuentran en el archivo `cereal.csv`.

- Se decidió introducir en el modelo las variables correspondientes a edad la cual esta categorizada en 4 grupos (1 menores de 31 años, 2 entre 31 y 45 años, 3 entre 46 y 60 años, 4

mayores de 60 años), estados civil, estilo de vida y género para tratar de explicar que desayuno preferieren. ¿Son todas las variables significativas? Justifique porque sí o porque no y si alguna no lo fuera elimínela del modelo y vuelva a ajustarlo.

- Para el modelo seleccionado anteriormente haga un análisis completo del ajuste. Esto quiere decir comparar el modelo ajustado con el que sólo tiene constantes, calcular la devianza residual y estadístico  $X^2$  e inferir sobre el ajuste del modelo, calcular el pseudo  $R^2$  e interpretar, ver si las variables son significativas, interpretar los valores estimados bajo el modelo y comentar acerca de la clasificación que se logra a través del mismo.
- Para el modelo seleccionado proporcione la probabilidad estimada en cada valor de la variable respuesta para un sujeto con más de 60 años de edad, mujer, soltera e inactiva ¿A qué categoría se le asignaría y por qué?

**Ejercicio 4.6.2.** La base de de datos *fertilidad.csv* contiene la información de 100 voluntarios que dieron una muestra de semen la cual fue analizada de acuerdo al criterio WHO 2010. La concentración de esperma está relacionada con aspectos socioeconómicos, estado de salud, estilo de vida y factores embrionales del donante, la base contiene las siguientes variables:

- Estación del año en la que se realizó el análisis.
  - -1: Invierno
  - -0.33: Primavera
  - 0.33: Verano
  - 1: Otoño
- Edad en la que se realizó el estudio. (0.5 es 18 años y 1 es 36 años, las edades intermedias se encuentran definidas por los decimales intermedios)
- Enfermedades infantiles como: varisela, polio, sarampión o paperas)
  - 0: No
  - 1: Sí
- Accidente o trauma serio
  - 0: No
  - 1: Sí
- Intervención quirúrgica
  - 0: No



- 1: Sí
- *Fiebre alta en el último año*
  - 0: No
  - 1: Sí
- *Consumo de alcohol*
  - 0.2: Casi nunca o nunca
  - 0.4: Una vez por semana
  - 0.6: Varias veces por semana
  - 0.8: Todos los días
  - 1: Varias veces al día
- *Fuma*
  - -1: Nunca
  - 0: Ocasionalmente
  - 1: Todos los días
- *Número de horas sentado por día.*
- *Diagnostico*
  - N: Normal
  - O: Alterado
- *Se decidió introducir en el modelo con las variables 9 correspondientes para tratar de explicar el diagnóstico final. ¿Son todas las variables significativas? Justifique porque sí o porque no y si alguna no lo fuera elimínela del modelo y vuelva a ajustarlo.*
- *Para el modelo seleccionado anteriormente haga un análisis completo del ajuste. Esto quiere decir comparar el modelo ajustado con el que sólo tiene constantes, calcular la devianza residual y estadístico  $X^2$  e inferir sobre el ajuste del modelo, calcular el pseudo  $R^2$  e interpretar, ver si las variables son significativas, interpretar los valores estimados bajo el modelo y comentar acerca de la clasificación que se logra a través del mismo.*

**Ejercicio 4.6.3.** La base de datos *poisson\_sim.csv* contiene información sobre el número de premios ganados por estudiantes de preparatoria, esta es la variable explicativa que se encuentra en el campo *num<sub>a</sub>wards*, las variables explicativas que se proponen son *prog* que tiene información

sobre el tipo de programa al cual se inscribió el estudiante (vocacional, general, académico) y math que tiene el puntaje de su examen final de matemáticas.

Realice un ajuste con la regresión poisson, para determinar los parámetros de ajuste, construya el modelo saturado y utilice el método backward para encontrar el mejor ajuste. Interprete los resultados del ajuste.

**Ejercicio 4.6.4.** Con la base de datos crab.txt Realice un ajuste con la regresión poisson, para determinar los parámetros de ajuste, construya el modelo saturado y utilice el método backward para encontrar el mejor ajuste. Interprete los resultados del ajuste.

La base se construyó a partir del estudio que investigó los factores que afectan si el cangrejo hembra tenía otros machos, llamados satélites, que residían cerca de ella. Las variables explicativas que se cree que afectan esto incluyen el color del cangrejo hembra ( $C$ ), la condición de la columna vertebral ( $S$ ), el peso ( $Wt$ ) y el ancho del caparazón ( $W$ ). El resultado de la respuesta para cada cangrejo hembra es su número de satélites ( $Sa$ ). Hay 173 mujeres en este estudio.

## 4.7. Otros Enfoques de estimación

### 4.7.1. El enfoque bayesiano

Idea: diseñar una Teoría Estadística, basada en una pequeña serie de principios básicos, que nos permita *estructurar* la solución a cualquier problema de inferencia.

La vía: la Teoría de la Decisión

¿Para qué una Teoría Estadística?

- Para darle a la Estadística una estructura coherente
- Porque con otros enfoques pueden presentarse casos en los que: (i) no hay una solución razonable; (ii) se presentan paradojas.

**Teorema de Bayes.** Dados dos eventos  $A$  y  $B$  tales que  $\Pr(B) > 0$ ,

$$\Pr(A|B) = \frac{\Pr(B|A) \Pr(A)}{\Pr(B)}.$$

Si  $\{A_i : i = 1, 2, \dots, M\}$  es un conjunto exhaustivo de eventos mutuamente excluyentes, entonces

$$\Pr(A_i|B) = \frac{\Pr(B|A_i) \Pr(A_i)}{\sum_{j=1}^M \Pr(B|A_j) \Pr(A_j)}.$$

- El uso del Teorema de Bayes en pruebas de diagnóstico es bastante común y no causa controversias.

- Mucho más controversial es el uso del Teorema de Bayes en análisis estadísticos generales, en los que los parámetros son las cantidades desconocidas de interés y por lo tanto se requiere especificar probabilidades sobre sus valores.

### *Diferencias*

- *Inferencia estadística tradicional*: ¿Qué nos dicen los datos  $X$  acerca del parámetro  $\theta$ ?  
(*Ignora toda evidencia externa*)
- *Inferencia bayesiana*: ¿Cómo cambian nuestros juicios originales acerca del valor de la cantidad desconocida  $\theta$  a la luz de los datos  $X$ ?  
(*Puede tomar en cuenta cualquier evidencia externa*)

En general tenemos:

- Datos,  $X$ ; y
- Cantidades desconocidas,  $\theta$ , cuyo valor nos interesa.

Las cantidades desconocidas descritas por  $\theta$  pueden ser: parámetros del modelo, observaciones faltantes, mediciones que no podemos observar directamente o con suficiente precisión, etc.

Como estadísticos, postulamos un modelo de probabilidad

$$p(x|\theta)$$

Desde el punto de vista bayesiano, además,

- $\theta$  debe tener una *distribución de probabilidad*,  $p(\theta)$ , que refleje nuestra incertidumbre *inicial* acerca de su valor.
- $X$  es conocido, así que debemos condicionar en su valor observado,  $x$ .

Por lo tanto, nuestro conocimiento acerca del valor de  $\theta$  queda descrito a través de su *distribución final*

$$p(\theta|x)$$

El Teorema de Bayes nos dice cómo encontrarla:

$$p(\theta|x) = \frac{p(\theta) p(x|\theta)}{\int p(\theta) p(x|\theta) d\theta}$$

\*

El Teorema de Bayes es la clave del *proceso de aprendizaje*.

Al hacer inferencias sobre un parámetro  $\theta$ , generalmente se cuenta con algún tipo de información (juicios, creencias) acerca de su valor, incluso antes de observar los datos.

El precio adicional que hay que pagar es la especificación de una distribución de probabilidad sobre  $\theta$  que describa la información que se tiene sobre su valor.

Cabe mencionar que los procedimientos clásicos también se basan (implícitamente) en apreciaciones subjetivas (¿Por qué un modelo normal?, ¿Por qué  $\alpha = 0.05$ ?)

Los cuatro pasos a seguir dentro del enfoque bayesiano:

1. Especificación de un modelo muestral,  $p(x|\theta)$
2. Especificación de una distribución inicial,  $p(\theta)$
3. Cálculo de la distribución final,  $p(\theta|x)$ , vía el Teorema de Bayes
4. Resumen de la información contenida en  $p(\theta|x)$  para hacer inferencias sobre las cantidades de interés (parámetros, observaciones futuras, etc.)

El problema de elegir un modelo para describir el proceso que generó los datos es esencialmente el mismo que desde el punto de vista clásico. El modelo elegido dependerá del problema en turno y del propósito del análisis.

En ocasiones, la forma en la que se obtuvieron los datos puede sugerir modelos apropiados *como punto de partida* (e.g., muestreo binomial, conteos Poisson).

Con frecuencia, el modelo refleja una hipótesis cuya plausibilidad es verificada posteriormente en el contexto de los datos (e.g.,  $Y$  y  $X$  se relacionan linealmente entre sí).

“Todos los modelos son incorrectos, pero algunos modelos son más útiles que otros.” (*George E.P. Box*)

1. *Distribución inicial* Este es un aspecto fundamental del enfoque bayesiano. El análisis es subjetivo dado que depende del conocimiento que el investigador tiene antes de observar los datos (y que describe a través de *su* distribución inicial). Sin embargo, si la distribución

inicial es razonable, su efecto sobre las inferencias disminuye conforme se tienen más datos. En ocasiones tenemos una idea vaga de la forma que debería tener la distribución inicial. Tal vez incluso somos capaces de asignar valores, por ejemplo, a su media y su varianza, pero no podemos ser más precisos. En estos casos es común usar una distribución inicial consistente con nuestra información pero cuya forma sea conveniente, e.g. tal que dé lugar a análisis más sencillos. (→ Familias conjugadas)

En otros casos puede considerarse que no se tiene información inicial sobre el valor del parámetro (o, por algún motivo, no es deseable incluir nuestra información inicial en el análisis).

En estas situaciones nos gustaría poder utilizar una distribución inicial que refleje nuestra ignorancia acerca del valor del parámetro.

En términos generales siempre es posible encontrar este tipo de distribuciones iniciales *no-informativas*.

Sin embargo, excepto en modelos relativamente simples, esta labor es complicada y no está exenta de problemas.

2. *Distribución final* En términos de variables aleatorias, el Teorema de Bayes toma la forma

$$p(\theta|x) = \frac{p(\theta)p(x|\theta)}{\int p(\tilde{\theta})p(x|\tilde{\theta})d\tilde{\theta}}.$$

El denominador,  $p(x) = \int p(\tilde{\theta})p(x|\tilde{\theta})d\tilde{\theta}$ , no depende de  $\theta$ , por lo que es común escribir

$$p(\theta|x) \propto p(\theta)p(x|\theta).$$

En la práctica, el cálculo de la distribución final puede ser un asunto complicado, especialmente si la dimensión del parámetro no es pequeña.

Sin embargo, para ciertas combinaciones de distribuciones iniciales y verosimilitudes es posible simplificar el análisis. (→ Familias conjugadas)

En otros casos se requieren aproximaciones analíticas y/o técnicas computacionales relativamente sofisticadas. (→ ¡Sesión de mañana!)

### 3. *Inferencia*

El enfoque bayesiano proporciona inferencias más completas en el sentido de que toda la información disponible sobre el valor de  $\theta$  queda representada a través de la distribución final.

Es decir, desde el punto de vista bayesiano, el problema de inferencia se reduce a encontrar  $p(\theta|x)$ : la distribución final *es* la inferencia.

**La única receta de la Inferencia Bayesiana.** . . . .consiste en encontrar la distribución condicional de todas aquellas cantidades de interés cuyo valor desconocemos dado el valor conocido de las variables observadas.

Por supuesto, en la práctica generalmente es deseable resumir este tipo de inferencias en la forma de una estimación puntual, una estimación por intervalo, una prueba de hipótesis, etc.

### *Robustez*

- En Estadística, independientemente del enfoque que se utilice, es importante entender hasta qué punto el modelo usado es robusto antes posibles violaciones a los supuestos.
- Lo anterior también es cierto dentro del enfoque bayesiano en lo que se refiere a la especificación de la distribución inicial.
- En ocasiones el modelo es tal que las inferencias *no* se modifican sustancialmente ante cambios moderados en la distribución final. Esto ocurre, por ejemplo, cuando el tamaño de la muestra es suficientemente grande.
- En otros casos, sin embargo, puede ocurrir que incluso cambios aparentemente insignificantes en la distribución inicial produzcan inferencias completamente distintas.

Algunos autores sugieren que, en la práctica, es conveniente comparar los resultados de los análisis derivados de por lo menos tres distribuciones iniciales distintas:

- Una distribución inicial no-informativa
- Una distribución inicial (tentativa) que refleje los aspectos más importantes nuestra información inicial
- Una distribución inicial (tal vez artificialmente) más informativa

La idea es que, si las inferencias no son muy distintas en cada uno de estos casos, el análisis (dados los datos observados) será relativamente robusto en lo que se refiere a la elección de la distribución inicial. No será necesario entonces preocuparse demasiado por especificar una distribución inicial con mucha precisión.

En caso contrario, es importante hacer el esfuerzo necesario para especificar una distribución que refleje genuinamente nuestra información inicial.

*Un ejemplo simple de inferencia bayesiana (distribución Binomial)*

- Datos:  $x$  éxitos en  $n$  ensayos independientes, cada uno con probabilidad de éxito  $\theta$ .

Por ejemplo,  $\theta$  puede representar la tasa de respuesta ante cierta dosis de una sustancia tóxica, y  $x$  el número de individuos, de un total de  $n$  expuestos, que presentan efectos adversos.

- Función de verosimilitud:

$$p(x|\theta) = \text{Bin}(x|\theta; n) = \binom{n}{x} \theta^x (1 - \theta)^{n-x} \propto \theta^x (1 - \theta)^{n-x}$$

- Distribución inicial:

$$p(\theta) = \text{Beta}(\theta|a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1} (1 - \theta)^{b-1} \propto \theta^{a-1} (1 - \theta)^{b-1}$$

- Distribución final:

$$\begin{aligned} p(\theta|x) &\propto p(\theta) p(x|\theta) \\ &\propto \theta^{x+a-1} (1 - \theta)^{n-x+b-1} \\ &\propto \text{Beta}(\theta|x + a, n - x + b) \end{aligned}$$

Notemos que tanto la distribución inicial como la final son Beta.

En este caso se dice que la familia de distribuciones Beta es *conjugada* para el modelo Binomial.

Supongamos que, dada la información inicial disponible, se determina que  $E(\theta) = 0.40$  y que  $\Pr(\theta > 0.54) = 0.10$ . Esto implica que  $a = 9.2$  y  $b = 13.8$ .

Interpretación: esta información inicial es equivalente a la de una muestra de tamaño  $a+b = 23$  en la que se obtuvieron  $a = 9.2$  éxitos.

Para la distribución  $\text{Beta}(a, b)$  se sabe que la media está dada por  $m = a/(a+b)$  y la varianza por  $s^2 = m(1-m)/(a+b+1)$ .

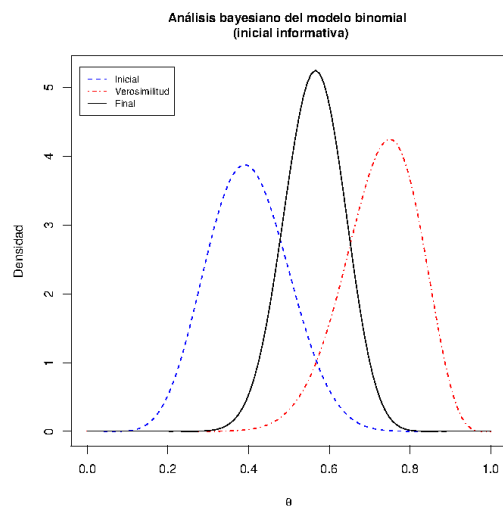
Entonces, *a priori*, la media de  $\theta$  es  $m = 0.40$  y la desviación estándar es  $s = 0.1$ .

Supongamos ahora que, al realizar un experimento con  $n = 20$  individuos expuestos, observamos  $x = 15$  individuos afectados.

Desglose de la información	Inicial	Datos	Final
Éxitos	9.2	15	24.2
Fracasos	13.8	5	18.8
Total	23	20	43

La media y la desviación estándar de la distribución final de  $\theta$  están dadas por  $E(\theta|x) = 0.563$  y  $\text{sd}(\theta|x) = 0.075$ , respectivamente.

Notemos que  $\Pr(\theta > 0.54|x) = 0.62$



### Caso no informativo

Supongamos que no se tiene o no se desea utilizar la información inicial.

Esto se puede especificar a través de una distribución inicial uniforme, lo que implica que  $a = b = 1$ .

En este caso, con  $x = 15$  individuos afectados de un total de  $n = 20$  individuos expuestos, tenemos:

Desglose de la información	Inicial	Datos	Final
Éxitos	1	15	16
Fracasos	1	5	6
Total	2	20	22

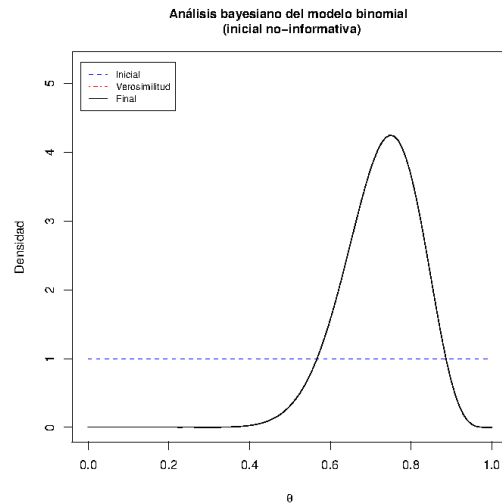
La media y la desviación estándar de la distribución final de  $\theta$  están dadas por  $E(\theta|x) = 0.727$  y  $sd(\theta|x) = 0.093$ , respectivamente.

Por otro lado, la *moda* de la distribución final es igual a 0.75, valor que coincide con el *estimador de máxima verosimilitud* para  $\theta$  en este caso.

Cabe hacer notar que en este caso  $\Pr(\theta > 0.54|x) = 0.97$

Supongamos ahora que estamos interesados en probar la hipótesis  $H_0 : \theta \leq 0.40$ . Entonces, la probabilidad  $\Pr(\theta \leq 0.40|x) = 0.0008$  puede usarse para determinar que los datos *no* apoyan esta hipótesis nula.





## Predicción

Hasta el momento sólo hemos discutido el problema de hacer inferencias acerca del valor desconocido de un parámetro.

En muchas situaciones, sin embargo, el propósito de formular un modelo estadístico es hacer *predicciones* sobre el valor de una o más observaciones futuras.

Este problema se resuelve de manera más elegante desde el punto de vista bayesiano que desde el punto de vista clásico.

Al hacer inferencias predictivas sobre el valor de una observación futura con base en un modelo, deben tomarse en cuenta dos fuentes de incertidumbre:

- Incertidumbre sobre el valor del parámetro (sobre el cual se pueden hacer inferencias con base en la distribución final).
- Incertidumbre por el hecho de que cualquier observación futura es aleatoria en sí misma (aún si conociéramos el verdadero valor del parámetro, no podríamos predecir con certeza el valor de una observación futura).

Dentro del enfoque clásico de la Estadística, es común ajustar el modelo con base en los datos (obteniendo un estimador puntual  $\hat{\theta}$ ), y entonces hacer predicciones con base en el modelo  $p(x|\hat{\theta})$  como si éste fuera el modelo verdadero.

De esta manera, se ignora completamente la primera fuente de incertidumbre, lo que produce predicciones que aparentan ser más precisas de lo que realmente son.

En contraste, el enfoque bayesiano toma en cuenta las dos fuentes de incertidumbre de manera natural.

### *Distribución predictiva*

Supongamos que tenemos una muestra observada  $\mathbf{x} = (x_1, \dots, x_n)'$  de  $p(x|\theta)$  y que se desea hacer inferencias acerca del valor futuro de  $Y = X_{n+1}$ .

Dada una distribución inicial  $p(\theta)$ , el Teorema de Bayes produce la distribución final  $p(\theta|\mathbf{x})$ .

Siguiendo la “*única receta de la inferencia bayesiana*”, debemos entonces encontrar la distribución condicional de  $Y$  dado el valor observado de  $\mathbf{x}$ .

Dicha distribución está dada por

$$\begin{aligned} p(y|\mathbf{x}) &= \int p(y|\theta, \mathbf{x})p(\theta|\mathbf{x}) d\theta \\ &= \int p(y|\theta)p(\theta|\mathbf{x}) d\theta \\ &= E_{p(\theta|\mathbf{x})}[p(y|\theta)] \end{aligned}$$

y se conoce como la *distribución predictiva (final)*.

*Continuación del ejemplo (distribución Binomial)*

Supongamos que estamos considerando detener el estudio si por lo menos 25 de 40 nuevos individuos tratados presentan efectos adversos. Con base en la información disponible, ¿Cuál es la probabilidad de que detengamos el estudio?

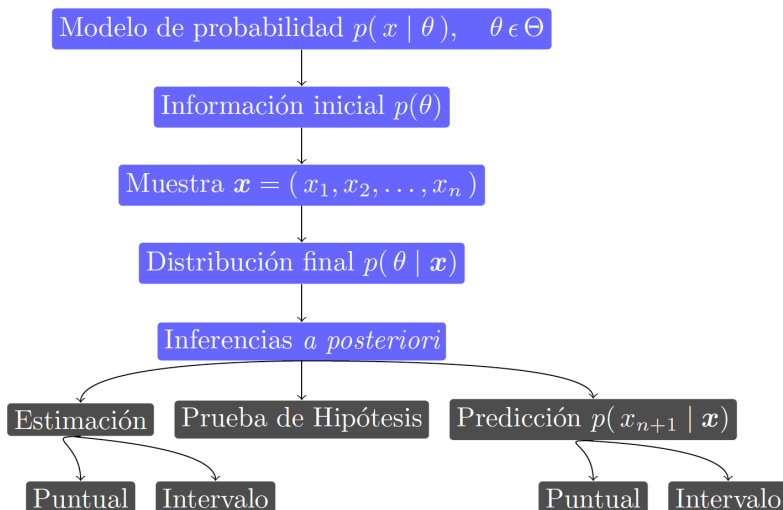
Estamos considerando observar  $n^*$  ensayos adicionales y nos interesa predecir el número de ‘éxitos’,  $X^*$ , en esos  $n^*$  ensayos.

La distribución predictiva (final) es *Binomial-Beta*:

$$p(x^*|x) = \binom{n^*}{x^*} \frac{\Gamma(n+a+b)\Gamma(x^*+x+a)\Gamma(n^*-x^*+n-x+b)}{\Gamma(x+a)\Gamma(n-x+b)\Gamma(n^*+n+a+b)}.$$

Esta distribución tiene media  $E(X^*|x) = 22.5$  y desviación estándar  $sd(X^*|x) = 4.3$ . Además, es tal que  $\Pr(X^* \geq 25|x) = 0.33$ .

*Recapitulando...*



*A manera de conclusión...* Los siguientes tres aspectos fundamentales caracterizan al enfoque bayesiano:

Información inicial: cada problema es único y tiene su propio contexto, del cual se deriva la información inicial sobre el parámetro (o cualquier otra característica) de interés.

Probabilidad subjetiva: se reconoce explícitamente que toda asignación de probabilidades es subjetiva (i.e., dependen del estado de información del individuo que las asigna). No pretende ser un enfoque “objetivo”.

Coherencia interna: al considerar a  $\theta$  como aleatorio, los métodos bayesianos de inferencia se desarrollan de manera natural a partir de la teoría de la probabilidad y por lo tanto no presentan contradicciones internas.