

Tarea 1. Análisis de datos categóricos.

Fecha de entrega: 19/marzo/2020

1. En Estados Unidos se levantó una encuesta para evaluar si la religiosidad de las personas estaba asociada con su preferencia política. Se eligieron 350 personas al azar. Entre las preguntas que se les realizaron estuvieron: ¿Cree usted en el infierno? y ¿Con qué preferencia política se identifica?

Los resultados se han resumido en la siguiente tabla de contingencia

Tabla 1. Preferencia política*Creencia en el infierno

| | Si creo | No creo | No sé |
|-------------|---------|---------|-------|
| Republicano | 95 | 33 | 10 |
| Sin partido | 30 | 20 | 6 |
| Demócrata | 80 | 66 | 10 |

1.1 ¿A qué tipo de esquema de muestreo pertenece este problema? Justifiquen su respuesta

1.2 ¿Es un modelo de independencia o de homogeneidad? Justifiquen su respuesta

1.3 Establezcan las hipótesis de la prueba de Ji cuadrada

1.4 Realicen “a mano” la prueba de Ji cuadrada con y sin corrección de Yates. Concluyan.

1.5 Corrobores el resultado en R. Muestran el resultado.

1.6 Realicen el *mosaicplot* correspondiente. Descríbanlo.

2. Lean el artículo de Tapia, José A. y Nieto, F. Javier. “Razón de posibilidades: una propuesta de traducción de la expresión odds ratio”. Salud Pública de México. Julio-Agosto, 1993. Vol. 35 No.4 Pág. 419-424. Elaboren un resumen en media cuartilla.

3. En el artículo Clopper, C. J. & Pearson, E. S. 1934. **The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial** *Biometrika*, Vol. 26, No. 4, vienen las tablas para calcular un intervalo exacto de confianza para p. Utilicen esas tablas para calcular un intervalo al 95% y otro al 99% para una muestra tamaño $n=50$ y $x=9$. Muestran las tablas y señalen los valores calculados en ellas.

4. En una escuela primaria se le pregunta a los niños sí les gusta leer. Se seleccionan 15 niños a los que sí les gusta leer y 15 niños a los que no les gusta leer. A su vez, se les pregunta sí les gustan las matemáticas. Los resultados se encuentran resumidos en la siguiente tabla:

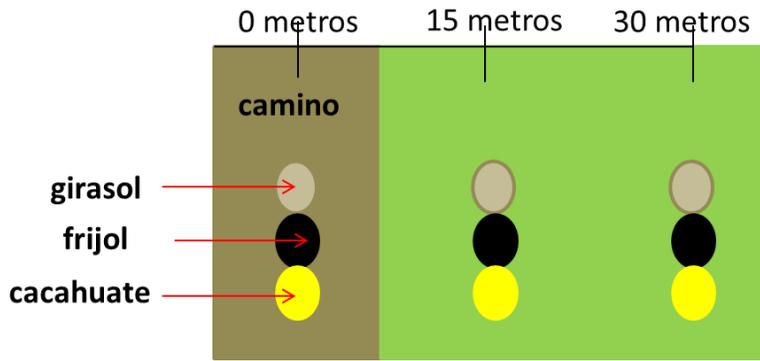
| | Si le gusta leer | No le gusta leer |
|------------------------------|------------------|------------------|
| Si le gustan las matemáticas | 12 | 2 |
| No le gustan las matemáticas | 3 | 13 |

4.1 ¿A qué tipo de esquema de muestreo pertenece este problema? Justifiquen su respuesta

4.2 ¿Es un modelo de independencia o de homogeneidad? Justifiquen su respuesta

4.3 Realiza la prueba exacta de Fisher para estos datos. Interpretala.

5. Los datos contenidos en el archivo **semillas_bosque.csv** corresponden a un experimento que llevaron a cabo unos biólogos. Sobre un camino en el bosque (0 metros), colocaron 50 semillas de cacahuete, 50 semillas de girasol y 50 semillas de frijol. Hicieron lo mismo a 15 y 30 metros del camino, como se esquematiza a continuación. Al día siguiente, contaron el número de semillas que habían quedado (semillas.presentes) y el número de semillas que se habían llevado los animales del bosque (semillas.removidas).



5.1 Importa los datos a R y especifica los siguientes modelos:

```
datos<-read.csv (.../semillas_bosque.csv)
```

```
modelo1<-glm(cbind(semillas.removidas,semillas.presentes)~especie,binomial(link="logit"), data=datos)
```

```
modelo2<-glm(cbind(semillas.removidas,semillas.presentes)~ distancia.camino.m,binomial(link="logit"), data=datos)
```

```
modelo3<-glm(cbind(semillas.removidas,semillas.presentes)~especie+distancia.camino.m,binomial(link="logit"), data=datos)
```

p= probabilidad de que la semilla sea removida

5.2 Escribe la fórmula completa de cada uno de estos modelos

5.3 Ejecuta la instrucción `anova(modelo3, test="Chisq")`. Explícala, indicando en un esquema el valor de χ^2 crítico, devianza, p-value y alfa para cada una de las comparaciones de la tabla. (Nota: Los esquemas pueden ser realizados a mano, en R o de otra manera. Sólo interesa que sean claros y legibles, y que se note que han entendido la utilidad de los valores en la tabla de análisis de devianza.)

5.4 ¿Cuál es el mejor modelo y porque? Utiliza la devianza, el AIC, la significancia de los coeficientes y el análisis de devianza que ejecutaste en el punto 3 para justificar tu respuesta

5.5 ¿Qué medida conviene calcular para evaluar la bondad de ajuste en datos agrupados y cuál conviene para datos desagrupados?

5.6 ¿Estos son datos agrupados o desagrupados? Evalúa la bondad de ajuste en el modelo que elegiste. Explica.

5.7 Interpreta los coeficientes del modelo 3

5.8 Utilizando los coeficientes del modelo 3, ¿Cuál será la probabilidad de que una semilla de frijol que se coloca a 10 m del camino sea removida?

5.9 Calcula las probabilidades para cada una de las distancias y las semillas evaluadas en este experimento.

5.10 Realiza un gráfico que muestre las probabilidades calculadas en el punto anterior.

