**Tarea #2 Análisis de Datos Categóricos Fecha de entrega: mayo 28/mayo/2020**

1. ¿Cuáles son las principales diferencias entre un modelo loglineal y un modelo de regresión Poisson? **(1 punto)**
2. ¿Por qué el modelo loglineal y el de regresión Poisson son modelos lineales generalizados? **(1 punto)**
3. ¿En qué situaciones resulta conveniente especificar como distribución del componente aleatorio a una binomial negativa? **(1 punto)**
4. Explica en qué consiste un modelo ZIP **(1 punto)**
5. ¿En qué situaciones resulta conveniente modelar tasas en lugar de conteos? Escribe un ejemplo de una de estas situaciones. **(1 punto)**
6. **(2 puntos)** En el archivo covid\_sexo.csv se encuentran contenidos por sexo (0=mujer, 1=hombre) los casos positivos para covid-19 y el número de pacientes de estos casos que fueron hospitalizados para cada una de las delegaciones de la CDMX. Los pacientes no hospitalizados son los que se consideran ambulatorios. Este archivo se construyó con los datos del gobierno de México recuperados el 7 de mayo en https://datos.gob.mx/busca/dataset/informacion-referente-a-casos-covid-19-en-mexico
7. Con estos datos, ¿Es posible afirmar que a los hombres enfermos de covid-19 los hospitalizan más? Especifica un modelo en R que te permita responder esa pregunta. Escribe el modelo con los valores de los coeficientes e indica cuales son los componentes del modelo. Sí hay alguna variable categórica nominal, explica el significado de los valores que ésta puede tomar.
8. Interpreta todos los coeficientes de tu modelo
9. **(1.5 puntos)** La siguiente tabla ha sido realizada con los datos colectados hasta el 7 de mayo del 2020 para pacientes positivos para covid-19 en la CDMX. Se ha registrado si fueron pacientes hospitalizados (si, no), el sexo del paciente (hombre, mujer) y sí tenían diabetes (si, no). En el SalidaRTarea2Categoricos2020.pdf se han especificado los 7 modelos y nombrado con las letras A-F.

|  |  |
| --- | --- |
| Diabetes No | Diabetes Si |
|  | Hombre  | Mujer | Hombre | Mujer |
| Hospitalizado No | 2476 | 2114 | 307 | 246 |
| Hospitalizado Si | 1326 | 626 | 512 | 293 |

1. Identifica qué tipo de independencia se presenta en cada uno de los modelos especificados.
2. Haz una tabla con las devianzas, grados de libertad, AIC, términos incluidos en cada modelo ajustado y tipo de independencia

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Mo delo | Devianza | Grados de libertad | AIC | Modelo ajustado | Tipo de Independencia/Asociación |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

1. Elige al mejor modelo justificando la elección en el AIC, devianzas, grados de libertad y parsimonia.
2. ¿El *mosaicplot* que elegiste apoya tu elección? Justifica tu respuesta mostrando los *mosaicplots* de los datos originales y de tu modelo de interés.
3. **(1.5 puntos)** Los datos bacalao.csv corresponden a número de bacalaos capturados en diferentes estaciones de pesca. Para cada estación, se provee del ID de la estación, la temperatura promedio del agua (°F), la latitud, la longitud y la profundidad máxima. Utilizando como variable dependiente el número de peces capturados (conteos) y como variables explicativas el resto de las variables, se han generado dos modelos: modelo 1 y modelo 2 especificados en el archivo SalidaRTarea2Categoricos2020.pdf

**a)** Escribe el modelo 2 con los valores de los coeficientes, indicando claramente los componentes aleatorio, sistemático y la función liga.

**b)** Interpreta los coeficientes del modelo 2.

**c)** Escribe las hipótesis nula ( ) y alternativa ( ) del Análisis de Devianza que se muestra en la salida de R.

**c)** Escribe la conclusión de la prueba de Análisis de Devianza para la que escribiste las hipótesis, incluyendo en tu conclusión si vale o no la pena la inclusión de “*Depth*”(profundidad máxima) en el modelo.