**BAJO PESO**

Un bebé recién nacido se considera de bajo peso si al momento de nacer su peso es menor a 2500gr. Este peso es preocupante para los médicos y se desea modelar la probabilidad de que un niño nazca con bajo peso puesto que las tasas de mortalidad infantil y las tasas de defectos al nacer son altas para bebés que nacen con bajo peso. El comportamiento de la madre durante el embarazo (dieta, hábitos de fumado, cuidado prenatal) podrían alterar enormemente las probabilidades de que su hijo/a nazca con bajo peso.

Se cuenta con datos de 189 nacimientos con información relevante de la madre sobre:

* pesomadre: peso de la madre en el último período menstrual en libras.
* fuma: estatus de fumado durante el embarazo (1: Sí fumó, 0: No fumó).
* raza: 1: negra, 2: blanca, 0: otra.
* iut: presencia de irritabilidad uterina (1: Sí, 0: No).
* edad: edad de la madre en años.

1. Construya la variable respuesta a partir del peso en gramos del recién nacido para que tenga dos categorías: bajo peso (menos de 2500gr) (1) y normal (de 2500gr en adelante) (0). Obtenga el número de casos con bajo peso y normal dentro de cada tratamiento.
2. Analice si se puede considerar el efecto de la raza independientemente del fumado usando un nivel de significancia de 0.10. Cambie al modelo de tratamiento referencia.
   1. Haga este análisis sin considerar otros factores y la covariable.
   2. Repita el análisis ahora considerando los factores y la covariable. Debe primero reducir su modelo hasta llegar a un modelo adecuado donde haga supuestos sobre las interacciones. Verifique toda las interacciones dobles que incluyen factores de diseño entre sí y con otros factores o covariables.
   3. Escriba el modelo resultante.
3. Basado en el modelo resultante del punto anterior, analice si hay una relación entre el fumado de la madre y la probabilidad de que un bebé tenga bajo peso.
   1. Haga la prueba de hipótesis adecuada. Use el alpha=0.1.
   2. Encuentre el OR al comparar los bebés de madres que fuman y no fuman. Dé una interpretación.
   3. Encuentre un límite inferior para el OR con una confianza de 90%. Dé una interpretación.
   4. ¿Se puede decir que el fumado es una causa de que una mujer tenga mayor o menor probabilidad de tener un niño con bajo peso?
4. Analice si hay una relación entre la raza de la madre y la probabilidad de que un bebé tenga bajo peso.
   1. Haga la prueba de hipótesis adecuada. Use el alpha=0.1.
   2. Encuentre el OR al comparar los bebés entre cada par de razas. Dé una interpretación.
   3. Pruebe las hipótesis referente a los diferentes OR y dé una conclusión.

**PALOMITAS**

Se realizó un experimento para determinar el efecto de la humedad (alta, baja, normal) y del tiempo de cocción en microondas (1, 2, 3 minutos) sobre el porcentaje de palomitas de maíz que quedan sin explotar. Se cuenta con el número de granos que explotan (exito) y el número que no explotan (fracaso). El experimento se realizó en 6 diferentes microondas y en cada uno de ellos se hicieron los 9 tratamientos. Use los datos en el archivo “palomitas.Rdata”

1. Cree una variable que sea la proporción de granos que explotaron en cada medición y presente un gráfico de cajas (boxplot) que permita comparar el comportamiento de la variable por tiempo de cocción. A partir del gráfico comente lo que se espera de las comparaciones.
2. Verifique (mediante prueba de hipótesis) si existe un efecto del tiempo de cocción sobre la proporción de granos que explotan. Debe considerar si la humedad afecta la prueba o no.
3. Tome ahora la variable que indica si el grano explota o no como respuesta. Además tome como factor de diseño solamente el tiempo de cocción (no use para nada la humedad). Use el tiempo como factor. Escriba el modelo que se debe utilizar y explique las tres partes del modelo lineal generalizado.
4. Obtenga las estimaciones de los coeficientes del modelo y úselos para estimar la probabilidad de que una semilla explote si se usa un tiempo de 2 minutos. Escriba todos los cálculos necesarios.
5. Haga las comparaciones entre los 3 tiempos por pares. Determine dónde hay diferencias.
6. Cuantifique las comparaciones que tengan sentido. Debe dar una interpretación de sus resultados.
7. Ahora haga el análisis considerando la humedad y vea si cambian los resultados.
8. Repita el análisis usando el tiempo de forma continua.
9. Haga el análisis usando como respuesta la proporción de granos que explotan en un modelo lineal normal y compare los resultados.

**CAIDAS**

Un estudio prospectivo busca investigar los efectos de dos intervenciones en la frecuencia de caídas en personas mayores de 65 años. Se seleccionan 100 personas que superan los 65 años y se dividen aleatoriamente en dos grupos. A un grupo se le da una **intervención consistente en explicaciones en una serie de clases (trat=0)** y al otro grupo se le dan **las mismas clases y además un entrenamiento de ejercicios aeróbicos (trat=1).** Se incluyen además 3 variables de cada persona: sexo (0: mujer, 1: hombre), índice de balance e índice de fuerza. Durante un período de 6 meses se llevó un registro del **número de caídas que cada persona tuvo.**

Los datos están en el archivo “**caidas.Rdata**”

1. Represente gráficamente los datos.
2. Escriba el modelo inicial.
3. Haga el análisis completo para determinar si hay un efecto del tratamiento. En esta parte debe tomar en cuenta solo las interacciones que puedan tener sentido, así como covariables. Debe reducir su modelo usan las técnicas apropiadas. Debe verificar que usa el modelo de conteo que cumpla con los supuestos.
4. Escriba el modelo final. Aquí debe escribirlo con todos los detalles del modelo lineal generalizado y usando la simbología adecuada. Escriba el modelo en forma larga y en forma reducida. Puede decidir si usa suma nula o tratamiento referencia, pero debe especificar claramente las restricciones.
5. De una interpretación adecuada de sus resultados.

**CANCER**

Se cuenta con datos de incidencia de cáncer de pulmón en 4 ciudades danesas en el período de 1968-1971. El propósito del ejercicio es analizar si la incidencia de cáncer está relacionada con la edad.

Los datos están en el archivo “**cancer.Rdata**”

1. Observe que la variable edad es categórica pues se ha resumido en grupos de edad. Para el análisis se quiere usar la edad de forma numérica y para esto es necesario usar los puntos medios de cada intervalo. Haga la transformación de la variable edad.
2. Observe gráficamente cómo se comportan de los datos.
3. Ajuste un modelo de Poisson para el número de casos de cáncer, usando la edad como variable independiente. Debe analizar si hay sobredispersión. Determine si hay un crecimiento del número de casos asociado a la edad. Piense cuál es el papel de la ciudad.
4. Incluya la población y haga el análisis gráfico.
5. Escriba el modelo de Poisson para la tasa de incidencia de cáncer.
6. Ajuste el modelo. Determine si hay un crecimiento en la tasa de incidencia asociado con la edad. Note la diferencia en los resultados con respecto al análisis anterior.
7. Interprete el coeficiente de la edad por un aumento razonable de la edad.
8. Haga un intervalo de confianza para el coeficiente de edad e interprételo por un aumento razonable de la edad.
9. Se puede esperar una diferencia en la tasa promedio entre ciudades?
10. Escriba el modelo final (no se está pidiendo las estimaciones sino el modelo que se usa).
11. Calcule la potencia que se tiene en este experimento para ver diferencia entre ciudades de al menos un 20% en las medias de la tasa de incidencia. Hágalo usando una distribución Poisson.

BACTERIAS

Un grupo de investigadores está interesado en analizar la asociación entre la red de apoyo social y el número de bacterias intestinales en una muestra de individuos longevos en la Zona Azul de Costa Rica en la Península de Nicoya. Para esto se tomaron en cuenta las siguientes variables predictoras:

* alim: el número de personas que acompañan a comer al adulto mayor.
* relig: el número de personas que acompañan a actividades religiosas (ir a la iglesia) al adulto mayor.
* salud: el número de personas que acompañan a los servicios de salud al adulto mayor.
* conversac: el número de personas con las que mantiene conversaciones el adulto mayor.

Además, se cuenta con 6 variables respuesta (Actinobacteriota, Bacteroidota, Verrucomicrobiota, Firmicutes, Proteobacteria, Spirochaetota) las cuáles son diferentes tipos de bacterias intestinales. Cada una de esas variables se refiere al número de bacterias presentes en la muestra fecal recolectada de cada adulto mayor.

Adicionalmente, se toma en cuenta la variable Total\_reads que se refiere al total de bacterias presentes en la muestra sin discriminar por tipo de bacteria. Se añade porque el número de bacterias que presenta cada individuo es distinto.

Los datos están en el archivo “**bacterias.Rdata**”

1. Tome alguna de las variables respuesta y haga el análisis para determinar cuáles de las variables predictoras están relacionadas con el número de bacterias. Haga el análisis sin tomar en cuenta el total de bacterias y también tomándolo en cuenta.
2. Haga el análisis usando un modelo lineal con distribución condicional normal y compare los resultados.