**Maestría en Probabilidad y Estadística**

**Modelos Estadísticos I**

**Examen Parcial #3**

**Sabado, 29 de mayo de 2021**

**Nombre:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Problema 1.-Regresión Logística.-**  El número de muertes por leucemia y otros canceres entre los sobrevivientes de la bomba atómica de Hiroshima se presentan en la siguiente tabla, clasificados por la dosis de radiación recibida. Los datos se refieren a muertes ocurridas durante el período 1950-1959, entre sobrevivientes que tenían 6edades de 25 a 64 años en 1950.

|  |
| --- |
| leucemia o\_cancer total dosis prop.leucemia dosis2 dosis3  1 13 378 391 0 0.03324808 0 0  2 5 200 205 1 0.02439024 1 1  3 5 151 156 2 0.03205128 4 8  4 3 47 50 3 0.06000000 9 27  5 4 31 35 4 0.11428571 16 64  6 18 33 51 5 0.35294118 25 125 |

Haga lo siguiente: …

a)Considere el Modelo 1, que tiene a la dosis únicamente como variable explicativa. Comente el ajuste de este modelo a los datos.

b) Considere el modelo 2, que tiene, además de la variable explicativa original (dosis), un término cuadrático (dosis2 ) dos variables explicativas. Agregamos el término cuadrático por recomendación de un investigador que conoció este estudio. Comente el ajuste de este modelo a los datos.

c) Consideramos además el Modelo 3, que tiene tres variables explicativas dadas por los términos polinomiales de la dosis, de ordenes 1,2 y 3 (dosis,dosis2 ,y dosis3 ). Comente el ajuste del modelo a los datos.

d) Comparando estos modelos y los derivados de ellos, ¿Cuál modelo tiene un mejor ajuste a los datos?. Argumente sus respuestas.

**Problema 2.-Ligas identidad y logaritmo**

Se realizó un estudio sobre un tipo de evento cuya ocurrencia se ve influenciada por una variable X. Durante nueve días de realizó el conteo del número de ocurrencias del evento estudiado, manteniendo el valor de la variable X (continua) en tres valores diferentes, -1, 0 y 1.

En la siguiente tabla se muestran los conteos que resultaron en el estudio, así como los valores correspondientes de la variable de influencia.

Analice los datos, considerando un modelo de regresión Poisson.

Tabla.- Datos del Problema 3.

|  |
| --- |
|  |
| 1 2 -1 4 7 0 7 10 1  2 3 -1 5 8 0 8 12 1  3 6 0 6 9 0 9 15 1 |

Haga lo siguiente:

1. haga un análisis considerando la liga natural, logarítmica. Comente los resultados.
2. Haga un análisis considerando la liga identidad. Comente los resultados.
3. Compare los resultados obtenidos en los incisos (a) y (b).
4. ¿Los datos muestran evidencia a favor de uno de los dos modelos (liga logarítmica o liga identidad?)

**Problema 3.-**Una forma simple de ilustrar analíticamente la sobredispersión en el caso de regresión logística, es suponer algún tipo de variabilidad en el parámetro binomial . Sea la variable aleatoria que reemplaza el parámetro binomial conocido, donde tiene una distribución y varianza Entonces, si Y es la variable aleatoria binomial, se cumple que su varianza es mayor que la cantidad dada por la distribución binomial simple, esto es,

.

Sugerencia. Utilizar, y además,

**Problemas 1.-**Considere una muestra aleatoria con la distribución exponencial

.

Obtenga la deviance comparando el modelo máximal con diferentes valores de para cada y el modelo con , para toda .

**Problema 2.**-Sea el valor máximo de la función logverosimilitud para el modelo minimal con predictor lineal y sea lineal .

1. Pruebe que la estadística ji-cuadrada es, , donde es la deviance para el modelo minimal y es la deviance para el modelo más general.
2. Deducir que si entonces tiene la distribución Ji-cuadrada central con grados de libertad.

**Problema 3.**-El número de muertes por leucemia y otros tipos de cáncer entre supervivientes de la bomba atómica de Hiroshima se muestran en la Tabla 1, clasificados por la dosis de radiación recibida. Los datos se refieren a muertes durante el periodo 1950-1959, entre supervivientes cuyas edades estaban entre 25 y 64 años en 1950. Obtener un modelo adecuado para describir la relación dosis –respuesta entre la radiación y las tasas de mortalidad proporcionales para leucemia.

Tabla 1. Muertes por leucemia y otros tipos de canceres, clasificadas por las dosis de radiación recibida de la bomba atómica de Hiroshima.

|  |  |
| --- | --- |
| Muertes | Dosis de radiación (rads)  0 1-9 10-49 50-99 100-199 200+ |
| Leucemia  Otros canceres | 13 5 5 3 4 18  378 200 151 47 31 33 |
| Total | 391 205 156 50 35 51 |

**Problema 4.**-En la clase vimos el procedimiento iterativo IRLS de estimación de los parámetros de los modelos lineales generalizados, para el modelo de regresión logística. En este caso, para obtener el algoritmo de estimación, se consideraban las siguientes relaciones:

, , , .

Haga lo siguiente:

1. Haga el desarrollo de las relaciones anteriores para el modelo de regresión Poisson.
2. Obtenga el algoritmo correspondiente a la regresión Poison.
3. Ilustre lo anterior con el siguiente ejemplo.
4. En la siguiente tabla se muestran los números de casos de AIDS que se presentaron en los años señalados, en Bélgica a partir de 1981. Se asume que estos datos siguen un modelo de regresión de Poisson, donde la variable explicativa son los años y la respuestas los números de casos por año.

**Tabla.-**Números de casos de AIDS que ocurrieron anualmente de 1981 a 1993 en Bélgica.

|  |
| --- |
| Año 1981 1982 1983 1984 1985 1986 1987 1988 1989 1990 1991 1992 1993  Casos 12 14 33 50 67 74 123 141 165 204 253 246 240 |

**Problema 5.**-En un estudio médico se cuentan los números de pólipos en pacientes que padecen poliposis adenomatosa familiar, después de recibir un tratamiento con un nuevo medicamento, o de un placebo. En la Tabla XX se presentan los datos asociados al estudio.

Haga lo siguiente

1. Grafique los datos y coméntelos
2. Encuentre un modelo MLG Poisson adecuado para modelar los datos y justifique que hay sobredispersion.
3. Ajuste un modelo cuasi-Poisson a los datos.
4. Ajuste un MLG binomial negativo a los datos.
5. Elija un modelo final y comente su elección.

**Tabla**.-Número de pólipos en los grupos tratamiento y placebo del estudio del Problema 5.

|  |  |
| --- | --- |
| Grupo Tratamiento | Grupo Placebo |
| Número Edad Número Edad  1 22 17 22  1 23 25 17  2 16 33 23  3 23  3 23  4 42 | Número Edad Número Edad  7 34 44 19  10 30 46 22  15 50 50 34  28 18 61 13  28 22 63 20  40 27 |