

Modelo Lineal Generalizado

Trabajo Práctico 2

1. La siguiente tabla muestra el número de muertes debidas a caídas accidentales para cada mes del año 1970.

Mes	Número de Muertes
Enero	1668
Febrero	1407
Marzo	1370
Abril	1309
Mayo	1341
Junio	1338
Julio	1406
Agosto	1446
Septiembre	1332
Octubre	1363
Noviembre	1410
Diciembre	1526

Cuadro 1: Valores Observados

- a) ¿Hay alguna evidencia de que la distribución no es uniforme a lo largo del tiempo? Justifique.
 - b) ¿Parece haber algún patrón estacional?
2. De acuerdo con la *Ley de Hardy–Weinberg* si una secuencia de genes está en equilibrio, los genotipos AA, Aa y aa aparecen en la población con una frecuencia $(1 - \theta)^2$, $2\theta(1 - \theta)$ y θ^2 , respectivamente. En una muestra de la población china de Hong Kong tomada en el año 1937 los tipos de sangre ocurrieron con la frecuencia que muestra la tabla

	Tipo Sanguíneo			Total
	M	MN	N	
Frecuencia	342	500	187	1029

Cuadro 2: Valores Observados

donde M y N son eritrocitos antígenos.

- a) Usando la *Ley de Hardy–Weinberg* estime las probabilidades de cada celda. Calcule la varianza asintótica de θ utilizando la teoría vista en clase. Halle un intervalo de confianza de nivel asintótico 0.95 para θ . Justifique.
- b) A partir de a) calcule la varianza asintótica de los estimadores propuestos para las probabilidades de las casillas. ¿Cómo estimaría estas varianzas?
- c) Testee el ajuste del modelo de Hardy–Weinberg a los datos obtenidos usando los estadísticos de Pearson y G^2 . ¿Parece que este modelo ajusta a los datos? ¿Son confiables las aproximaciones χ^2 en este ejemplo? ¿Por qué ?
3. Bliss y Fisher (1953) reportan los datos de recuento de hembras adultas de ácaros rojos en 6 plantas de manzana de una quinta que fue fumigada. De cada árbol se recolectaron 25 hojas y en cada hoja se contó cuantas hembras había. Se pensó que la distribución del número de hembras por hoja podría ajustarse con una distribución Poisson. La siguiente tabla muestra los datos obtenidos

Número de Hembras por Hojas	Cantidad observada
0	70
1	38
2	17
3	10
4	9
5	3
6	2
7	1
8+	0

Cuadro 3: Valores Observados

Testee mediante el estadístico de Pearson el ajuste de una distribución Poisson a los datos obtenidos luego de haber fusionado las últimas tres casillas. ¿Parece razonable el ajuste? ¿Por qué ? Calcule los residuos de Pearson.

4. En un estudio ecológico en el que se investigó los hábitos alimentarios de pájaros, se contó el número de saltos entre vuelos de distintos pájaros.

La siguiente tabla contiene los datos obtenidos. Se sospecha que los datos tienen una distribución geométrica. ¿Parece razonable esta sospecha? Justifique su respuesta. ¿Es necesario unir algunas casillas?

5. En un estudio demográfico Kiser y Schaefer (1949) reportaron la siguiente tabla sobre mujeres que al menos se casaron una vez.

Se desea saber si el nivel de educación y el estado civil están asociados. Realice un test para decidir si las dos variables son independientes. ¿Cuál es su decisión?

Número de Saltos	Frecuencia
1	48
2	31
3	20
4	9
5	6
6	5
7	4
8	2
9	1
10	1
11	2
12	1

Cuadro 4: Ejercicio 4. Valores Observados

Educación	Casada una vez	Casada más de una vez	Total
Universitaria	550	61	611
No Universitaria	681	144	825
Total	1231	205	1436

Cuadro 5: Kiser y Schaefer (1949)

6. **Datos Apareados** Vianna, Greenwald y Davies (1971) recolectaron datos a fin de comparar los porcentajes de amigdalectomías en un grupo de pacientes con *Mal de Hodgkin* y un grupo control comparable, los datos se muestran en la siguiente tabla.

	Amigdalectomía	No Amigdalectomía
Hodgkin	67	34
Control	43	64

Cuadro 6: Vianna, Greenwald y Davies (1971)

La tabla muestra que el 66% de los pacientes que sufren de Mal de Hodgkin ha tenido una amigdalectomía comparada con el 40% del grupo control. El test de homogeneidad da un estadístico igual a 14.26 con 1 grado de libertad da altamente significativo. Los investigadores dedujeron que de alguna manera las amígdalas tienen un efecto protector contra el Mal de Hodgkin.

Johnson and Johnson (1972) seleccionaron 85 pacientes que tenían un pariente sano y con una edad de a lo sumo 5 años de diferencia. La tabla obtenida fue la siguiente.

Los investigadores calcularon un estadístico de Pearson que dio 1.53, que no resultó significativo, lo que parecía contradecir los resultados de Vianna, Greenwald y Davies (1971).

	Amigdalectomía	No Amigdalectomía
Hodgkin	41	44
Control	33	52

Cuadro 7: Johnson and Johnson (1972)

- a) ¿Qué se podría criticar a este último análisis?
b) La información de la tabla anterior puede reescribirse como

Paciente	Pariente	
	No Amigdalectomía	Amigdalectomía
No Amigdalectomía	37	7
Amigdalectomía	15	26

Cuadro 8: Johnson and Johnson (1972)

Para esta nueva tabla la hipótesis nula apropiada parece ser que la probabilidad de Amigdalectomía y de No Amigdalectomía son las mismas para los pacientes y sus parientes. Muestre que en este caso la hipótesis nula a testear sería:

$$H_o : \pi_{12} = \pi_{21}$$

- c) A partir del Ejercicio 3 de la Práctica 1 derive un test apropiado para testear H_o y pruebe que el estadístico es

$$\chi^2 = \frac{(n_{12} - n_{21})^2}{n_{12} + n_{21}}$$

¿Cuántos grados de libertad tiene?

Este test se conoce como **test de Mc Nemar**.

- d) Calcule el estadístico del test de Mc Nemar para la tabla dada. ¿Cuál es el p-valor? ¿A qué conclusión se llega? ¿Es coherente con las conclusiones de Johnson and Johnson (1972)?

7. Para los datos de Everitt (1977) dados en el Ejercicio 5 de la Práctica 1 testee la hipótesis de independencia. Especifique los grados de libertad y el p-valor ¿Cuál es su conclusión?

- a) Compute los residuos a partir de a).
b) Usando los resultados de la Práctica 1, testee la hipótesis nula de que la variable Conducta en clase es independiente de las otras dos. ¿A qué conclusiones arriba?

8. a) Para los datos del Ejercicio 6 de la Práctica 1 testee la hipótesis de independencia y compute los residuos.

- b) Compute los residuos a partir de a). ¿Qué impresión tiene?

- c) Usando los resultados de la Práctica 1, testee la hipótesis nula de que las variables Nivel de Colesterol y Nivel de Presión son independientes dado Tipo de Personalidad. ¿A qué conclusiones arriba?

9. Consideremos la siguiente tabla

n_{11}	n_{12}	n_{1+}
n_{21}	n_{22}	n_{2+}
n_{+1}	n_{+2}	n

Cuadro 9: Marginales Fijas

- a) Supongamos que todas las marginales n_{i+} y n_{+j} son fijas. Pruebe que el odd ratio $\theta = \frac{n_{11}n_{22}}{n_{12}n_{21}}$ es una función creciente de n_{11} .
- b) Supongamos ahora que el muestreo es producto multinomial, es decir que $n_{11} \sim Bi(n_{1+}, \pi_{1|1})$ y $n_{21} \sim Bi(n_{2+}, \pi_{1|2})$ son independientes. Pruebe que bajo el supuesto de homogeneidad, $n_{11}|_{n_{11}+n_{21}=s} = t$ tiene distribución hipergeométrica. ¿De qué parámetros?
- c) En un campo costero hay 30 pinos, 15 en la costa y otros 15 en el interior del mismo. De los 30 pinos, 4 están secos y los restantes están saludables. Si los 4 pinos secos están sobre la costa, ¿hay evidencia de que el viento o algún otro factor sobre la zona costera afecte a los pinos o la diferencia observada es atribuible meramente al azar?

A través del test de exacto de Fisher testee si hay diferencia entre los dos grupos.