

Guía Chi-cuadrado

Problema 1

Se considera la variable X el número de personas que se presentan a la boletería de una estación de metro durante un intervalo de 10 minutos. Se mide esta variable sobre 100 periodos de 10 minutos y se obtuvieron los datos siguientes:

X	3	4	5	6	7	8	9	10	11	12	13	14
f	7	9	14	15	17	13	8	6	5	3	2	1

donde f es el número de periodos de 10 minutos para lo cual se encontro X personas en la boletería.

3.1 Calcule el número promedio de personas que se presentan a la boletería en un intervalo de 10 minutos.

3.2 Bajo el supuesto que X sigue una distribución de Poisson, encuentre el estimador de máxima verosimilitud del parámetro λ (Muestre su resultado).

3.3 Haga un test χ^2 para justificar el supuesto que X sigue una distribución de Poisson.

Dé el p-valor y concluye (Necesitan usar las tablas de las distribuciones de Poisson y tablas χ^2).

Solución:

3.1 El promedio se calcula como: $\sum f_i X_i / n = 7.03$

3.2 El EMV de λ es $\hat{\lambda} = \bar{x} = 7.03$

3.3 El $\chi^2 = \sum \frac{(f_i - 100p_i)^2}{100p_i} = 1.94$, el p-valor: $P(\chi_{10}^2 > 1.94) = 0.99$

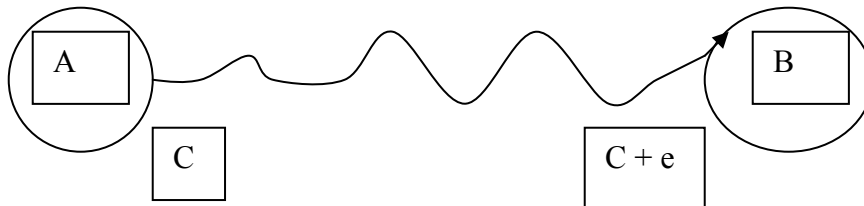
No se rechaza la hipótesis nula que X sigue una distribución de Poisson.

Problema 2

Una señal es enviada de un lugar A a otro B con un valor exacto C. Pero la señal llega a B con valores aleatorios en torno a C, a veces mayores y a veces menores. La explicación es que el ruido que se genera en la transmisión de la señal (e) es de carácter aleatorio, y si no se considera el valor de C, entonces se puede decir que para varios experimentos el valor que toma el ruido es tal que permite construir la siguiente tabla de frecuencias:

clase	frecuencia
$(-\infty; -0,09]$	1
$(-0,09; -0,07]$	0
$(-0,07; -0,05]$	2
$(-0,05; -0,03]$	12
$(-0,03; -0,01]$	8
$(-0,01; 0,01]$	18
$(0,01; 0,03]$	8
$(0,03; 0,05]$	6
$(0,05; 0,07]$	1
$(0,07; \infty)$	1

Estudios anteriores han determinado que la media del ruido es 0 y su desviación típica se puede **estimar** en 0,0314. Contraste la hipótesis de que el ruido sigue una distribución Normal usando los parámetros antes mencionados.



Solución:

El valor de n se obtiene al sumar todas las frecuencias m_i , y en este caso $n=57$. La región de rechazo es de la forma $W = \{Q > c\}$.

El valor de los p_i (probabilidad de caer en el intervalo i) se obtiene haciendo uso de la tabla Normal(0,1), calculando la probabilidad de que un valor cualquiera (llamémoslo e_i para los valores del ruido en el intervalo i) esté en el intervalo de la clase.

Ejemplo:

$$P_2 = P(-0,09 < e_i \leq 0,07) = P\left(\frac{-0,09 - \mu}{S} < \frac{e_i - \mu}{S} \leq \frac{0,07 - \mu}{S} \right)$$

$$= P(-2,866 < Z \leq -2,229)$$

Aproximando la segunda cifra decimal:

$$= P(z > 2,23) - P(Z > 2,87) = 0,0129 - 0,0021 = 0,011 \text{ (de las tablas).}$$

Así, se obtienen los p_i para cada intervalo.

Se construye la siguiente tabla:

Clase	Frecuencia(m_i)	$n \cdot p_i$	$\frac{(m_i - np_i)^2}{np_i}$
$(-\infty; -0,09]$	1	0,118	6,565
$(-0,09; -0,07]$	0	0,617	0,617
$(-0,07; -0,05]$	2	20,437	0,078
$(-0,05; -0,03]$	12	6,5	4,654
$(-0,03; -0,01]$	8	11,707	1,174
$(-0,01; 0,01]$	18	14,243	0,991
$(0,01; 0,03]$	8	11,707	1,174
$(0,03; 0,05]$	6	6,5	0,038
$(0,05; 0,07]$	1	2,437	0,847
$(0,07; \infty)$	1	0,735	0,095

Con esto, el valor de $Q = \sum_{i=1}^8 \frac{(m_i - np_i)^2}{np_i}$ es 16,234. Como $Q \rightarrow \chi^2_{q-1-l}$

con $q = n^\circ$ de intervalos, $l = n^\circ$ de estimaciones hechas, que en este caso está dado por la restricción $\sum m_i = n$. Por lo tanto, son $10-1-1=8$ grados de libertad.

Mirando la tabla para la χ^2 , obtenemos que

$$P(\chi^2_8 > c) = 0,05 \rightarrow c = 15,51.$$

Como $Q > 15,51$ entonces cae en la región de rechazo y por lo tanto se rechaza la hipótesis de que el ruido siga una distribución normal de media 0 y desviación típica estimada en 0,0314.

Problema 3

Se quiere probar si hay una diferencia de ingreso entre hombres y mujeres médicos. Para lo cual se hizo una encuesta a 200 médicos al azar y de manera independiente obteniéndose la siguiente información:

	Ingreso Bajo	Ingreso Alto	Total
Hombres	20	100	120
Mujeres	70	10	80
Total	90	110	200

Estudie la independencia entre las variables Sexo e ingreso

Solución:

De la tabla podemos ver que

- Probabilidad de ser Hombre $= P_H = \frac{120}{200}$, Probabilidad de ser Mujer $= P_M = \frac{80}{200}$
- Probabilidad de Ingreso Bajo $= P_B = \frac{90}{200}$, Probabilidad de Ingreso Alto $= P_A = \frac{110}{200}$

Entonces, si ambas variables son independientes debe cumplirse que:

- $P_{HB} = P_H P_B = \frac{120}{200} * \frac{90}{200} = 0.27$ (probabilidad que sea hombre y de ingreso bajo)
- $P_{HA} = P_H P_A = \frac{120}{200} * \frac{110}{200} = 0.33$ (probabilidad que sea hombre y de ingreso alto)
- $P_{MB} = P_M P_B = \frac{80}{200} * \frac{90}{200} = 0.18$ (probabilidad que sea mujer y de ingreso bajo)
- $P_{MA} = P_M P_A = \frac{80}{200} * \frac{110}{200} = 0.22$ (probabilidad que sea mujer y de ingreso alto)

Tabla de valores teóricos $n\hat{p}_i\hat{p}_j$

	Ingreso Bajo	Ingreso Alto	Total
Hombres	54	66	120
Mujeres	36	44	80
Total	90	110	200

$$\Rightarrow Q = \sum_{i,j} \frac{(M_{ij} - n\hat{p}_i\hat{p}_j)^2}{n\hat{p}_i\hat{p}_j} = \frac{(20-54)^2}{54} + \frac{(100-66)^2}{66} + \frac{(70-36)^2}{36} + \frac{(10-44)^2}{44} = 97.3$$

$$\Rightarrow Q \sim \chi^2_{(p-1)(q-1)} = \chi^2_{(2-1)(2-1)} = \chi^2_{1*1} = \chi^2_1 \quad y$$

$P(\chi^2_1 > 97.3) < 0.05 \Rightarrow$ se rechaza H_0 y las variables Sexo e Ingreso no son independientes.

Problema 4

6.1 Se considera una muestra de 1422 gatitos de 3 razas diferentes R_1 , R_2 y R_3 y 4 atributos diferentes del animal A_1 , A_2 , A_3 y A_4 . Se considera en la tabla 1 las frecuencias de gatitos por raza/atributo. ¿Cómo se llama la tabla 1? Busque mediante un test de hipótesis si existe o no una preferencia de ciertas razas por cierto atributo. Comente.

Tabla 1

	R1	R2	R3	Total
A1	604	42	0	646
A2	302	21	0	323
A3	15	418	0	433
A4	0	0	20	20
Total	921	481	20	1422

6.2 Se construye la tabla 2 sumando las dos primeras filas. Repita el test de la pregunta anterior. Compare los resultados y justifíquelos.

Tabla 2

	R1	R2	R3	Total
A1+A2	906	63	0	969
A3	15	418	0	433
A4	0	0	20	20
Total	921	481	20	1422

Solución:

6.1 La tabla 1 se llama tabla de contingencia. Un test de hipótesis es el de χ^2 de contingencia que mide la distancia entre los datos de frecuencias observados con los de la hipótesis nula de independencia entre raza y atributo. Los datos bajo la hipótesis de independencia entre raza y atributo se obtienen a partir de los márgenes de la

tabla1: $m_{ij} = \frac{n_{i.} \cdot n_{.j}}{n}$

	R_1	R_2	R_3	Total
A_1	418.40	218.51	9.09	646
A_2	209.20	109.26	4.54	323
A_3	280.45	146.46	6.09	433
A_4	12.95	6.77	0.28	20
Total	921	481	20	1422

El estadístico del test es entonces: $Q = \sum_{ij} \frac{(n_{ij} - m_{ij})^2}{m_{ij}}$, que bajo la hipótesis de independencia tiene una distribución aproximada de χ^2 con 6 grados de libertad ($6=(4-1)*(3-1)$).

$\frac{(n_{ij} - m_{ij})^2}{m_{ij}}$	R_1	R_2	R_3
A_1	82,3302516	142,586094	9,08579466
A_2	41,1651258	71,2930472	4,54289733
A_3	251,247444	503,406449	6,09001406
A_4	12,9535865	6,76511955	1382,28129

De la tabla anterior *observamos que* Q toma el valor de **2513.7**.

El pvalor del test se obtiene por: $P(\chi^2_6 > 2513.7)$ que es casi nulo (se puede ver con las tablas que es menor que 1%).

Por tanto para H_0 : El atributo es independiente de la raza, se desecha la independencia. Si se observa que la raza R_3 toma solamente el atributo A_4 y que la raza R_1 prefiere los atributos A_1 y A_2 y la raza R_2 prefiere el atributo A_3 no sorprende este resultado.

6.2 La tabla bajo la hipótesis de independencia:

	R_1	R_2	R_3	Total
A_1+A_2	627.60	327.77	13.63	969
A_3	280.45	146.46	6.09	433
A_5	12.95	6.77	0.28	20
Total	921	481	20	1422

Obtenemos el mismo valor del estadístico $Q=2513.7$. Esto no debería sorprender ya que se sumaron dos filas que eran similares (proporcionales). La diferencia se encuentran en los grados de libertad, que son solamente 4. El p-valor es nulo también.