GALATEA
Manual de la Interfaz gráfica del Usuario

Galatea Team

Enero 2024

Contents

Chapter 1

STAT Menu

1.1 Pearson's chi-squared test

Es una prueba no paramétrica que se usa tanto para distribuciones discretas como continuas. Se formulan las siguientes hipótesis:

- H_o: la variable X tiene distribución de probabilidad $f(x)$ con los parámetros específicos $\lambda_1, \lambda_2, \ldots$

- H_1: X tiene otra distribución de probabilidad.

Por ejemplo, la distribución de probabilidad que se supone en la hipótesis nula, esto es en H_o, puede ser una distribución normal cuyos parámetros serían la media μ y la desviación estándar σ. La hipótesis nula se evalúa con un cierto nivel de significación α, esto es, una medida del error que se cometería al rechazarla siendo cierta. Por lo general este nivel se establece en $\alpha = 0.01$, 0.05 o 0.10; indicando que el resultado de la prueba en fiable en un 1%, 5% o 10% respectivamente.

Para aplicarla los datos deben estar agrupados en frecuencias y la muestra tiene que ser lo suficientemente grande para que sea válida. Usualmente se establece que si en los datos agrupados aparece una frecuencia con valor menor a 5, esta barra del histograma no se use. Si hay más de una columna con frecuencia menor a 5, entonces deben combinarse en una para obtener una frecuencia con valor numérico mayor a 5.

El estadístico χ^2 viene dado por

$$\chi^2 = \sum_{i=1}^{N} \frac{(O_i - E_i)^2}{E_i} \ , \tag{1.1}$$

donde N es el número del grupos del histograma de frecuencia, O_i es la frecuencia observada del grupo i y E_i es el valor esperado para el grupo i si su distribución de probabilidades fuese la supuesta en la hipótesis nula.

Para aceptar o rechazar la hipótesis nula se compara el valor obtenido en la ecuación (**??**) con un valor crítico tabulado $\chi_{\alpha;k}$. Aceptando la hipótesis nula si

$$\chi^2 < \chi^2_{k,p} \ , \tag{1.2}$$

donde $k = N - 1$ es el grado de libertad y $p = 1 - \alpha$.

1.1.1 Ejemplo 1

El número de fallas por semana que sufre un equipo durante 36 semanas de trabajo es la siguiente:

i	1	2	3	4	5
Nº fallas por semana x	0	1	2	3	4 o más
Nº de semanas con O_i fallas	6	8	10	6	6

¿La muestra de datos se ajusta a una distribución de Poisson con media λ, con un nivel de significación de $\alpha = 5\%$?

Para responder la pregunta se debe realizar una prueba de bondad de ajuste donde las hipótesis deben ser: H_0 Los datos se ajustan a la distribución de Poisson contra :H_1 Los datos no se ajustan a la distribución de Poisson.

Comencemos por calcular el promedio de fallas por semana, que es nuestro estimador para la media λ,

$$\lambda = \frac{0 \times 6 + 1 \times 8 + 2 \times 10 + 3 \times 6 + 4 \times 6}{6 + 8 + 10 + 6 + 6} = \frac{70}{36} = 1.94 \text{ falla/semana}$$

Ahora calculamos los valores esperados E_i a partir de

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!} \ ,$$

de donde obtenemos la tabla

X	0	1	2	3	4 o más
O_i	6	8	10	6	6
E_i	5.1504	10.0147	9.7365	6.3107	4.7878
$\frac{(O_i-E_i)^2}{E_i}$	0.1401	0.4053	0.0071	0.0153	0.3069

Con los datos tabulados calculamos

$$\chi^2 = \sum_{i=1}^{5} \frac{(O_i - E_i)^2}{E_i} = 0.8748 \ .$$

También, buscamos el valor de $\chi^2_{N-m-1,1-\alpha}$, donde $N = 5$ es número de categorías en que se agrupan los datos, $m = 1$ es el número de parámetros estimados en la distribución teórica y $N - m - 1 = 3$ son los grados de libertad de la distribución χ^2. En nuestro caso es

$$\chi^2_{3,0.95} = 7.81 \ ,$$

como $0.8901 < 7.81$ entonces se acepta la hipótesis nula H_0 con un nivel de significación del 5%, es decir, que se acepta que la muestra de datos correspondientes al número de fallas por semana se ajusta a la distribución de Poisson.

1.1.2 Ejemplo 2: Atención al cliente

Repitiendo el procedimiento del ejemplo **??**, haremos la prueba χ^2 de bondad de ajuste para el histograma de tiempo entre llegadas (ver Fig. **??**a) que se obtuvo en el ejemplo **??**.

Comencemos calculando la media y la varianza muestral, dadas por

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{N} O_i \frac{a_{i+1} - a_i}{2} = 2.51 \text{ min} \qquad \text{y} \qquad s^2 = \frac{1}{n-1}\sum_{i=1}^{N} O_i \left(\frac{a_{i+1} - a_i}{2} - \bar{x}\right)^2 = 6.22 \text{ min}^2 ,$$

respectivamente, donde $n = 10000$ es el tamaño de la muestra, $N = 100$ es el número de intervalos o grupos del histograma, O_i es la frecuencia del i-ésimo intervalo, a_{i+1} y a_i son los extremos superior e inferior de ese intervalo, por lo tanto, $(a_{i+1} - a_i)/2$ es su punto medio.

Sabiendo la naturaleza del fenómeno, observando la forma del histograma de frecuencia y viendo que la desviación estándar muestral, $s = 2.54$ min, es aproximadamente igual a la media muestral, $\bar{x} = 2.51$ min, cabe preguntarse si los tiempos entre llegadas de personas al punto de atención al público siguen una distribución exponencial con media $\lambda^{-1} = 2.51$ min, cuyas funciones de densidad y de distribución viene dadas por

$$f(x,\lambda) = \lambda e^{-\lambda x} \qquad \text{y} \qquad F(x,\lambda) = 1 - e^{-\lambda x}$$

respectivamente.

Para responder a la pregunta hagamos la prueba χ^2 de bondad de ajuste con un nivel de significación de $\alpha = 1\%$. Los valores esperado en cada intervalo vienen dados por

$$\begin{aligned} E_i = & \; P(a_i < x \leq a_{i+1}) = F(a_{i+1},\lambda) - F(a_i,\lambda) \\ = & \; (1 - e^{-\lambda a_{i+1}}) - (1 - e^{-\lambda a_i}) = e^{-\lambda a_i} - e^{-\lambda a_{i+1}} , \end{aligned}$$

De estos datos obtenemos

$$\chi^2 = \sum_{i=1}^{N} \frac{(O_i - E_i)^2}{E_i} = 105.4052 .$$

En este caso tenemos $N = 100$ intervalos, $m = 1$ parámetro estimado en la distribución teórica y, por lo tanto $N - m - 1 = 98$ son los grados de libertad de la distribución χ^2 que tenemos que usar con un nivel de significancia $\alpha = 1\%$.

$$\chi^2_{0.99,98} = 133.4756 ,$$

como 105.4052 < 133.4756 podemos decir que se acepta, con un nivel de significancia de $\alpha = 1\%$, la hipótesis nula H_0: que los tiempos entre llegada se distribuyen según una exponencial con media $\lambda^{-1} = 2.51$ min.

1.1.3 Ejemplo 3: Atención al cliente

Mediante una prueba χ^2 de bondad de ajuste determine si el tipo de servicio de solicitado por los clientes del ejemplo **??** (ver Fig. **??**b) se ajusta a una distribución donde el 5% de los clientes solicitan una trámite de apertura, 80% hacen una consulta y el 15% restante vienen al punto por un reclamo. Note que en este caso no es necesario estimar ningún parámetro (m=0).