UNIVERSIDAD SIMÓN BOLÍVAR
DEPARTAMENTO DE CÓMPUTO CIENTÍFICO Y ESTADÍSTICA
CÁTEDRA: ESTADISTICA PARA INGENIEROS (CO-3321)

Laboratorio de Pruebas de Bondad de Ajuste.

El estadístico de prueba bajo H_0 es

$$\chi^2 = \sum_{i=1}^{k} \frac{(f_i - np_i)^2}{np_i} \sim \chi^2_{k-1-r}$$

donde r es el número de parámetros desconocidos.
k es el número de casos a considerar.
La región de rechazo es $RR = (\chi^2_{k-1-r;\alpha}, \infty)$ y
el p-valor es $1 - P(\chi^2 \leq \chi^2_{obs})$

Ejemplos:

(1) 90 personas bajan, de una en una, por una rampa que conduce a 3 puertas. Se encontraron 23 presonas que prefieren la puerta 1, 36 que prefieren la puerta 2 y 31 que prefieren la puerta 3. Al nivel de significación $\alpha = 0.05$. ¿Se puede afirmar que las personas no muestran preferencia con respecto a la elección de una puerta?.

Solución:

Tenemos $H_0 : p_i = 1/3$ para todo $i = 1, 2, 3$, contra
$H_1 : p_i \neq 1/3$, para algún $i = 1, 2, 3$
$> alpha < -0.05$
$> n < -90$
$> k < -3$
$> r < -0$
$> p < -c(1/3, 1/3, 1/3)$
$> fi < -c(23, 26, 31)$
$> chi2.obs < -sum((f_i - n * p) \wedge 2/(n * p))$
$> chi2.obs$
$[1]\, 1.2$
$> chi2.alpha < -qchisq(1 - alpha, k - 1 - r)$
$> chi2.alpha$
$[1]\, 5.991465$

Esto nos da como resultado $\chi^2_{obs} = 2.2$ y $\chi^2_{k-1-r;\alpha} = 5.991465$

Por lo tanto, la región de rechazo es $RR = (5.991465, \infty)$

Como $chi2.obs$ no pertenece a la región de rechazo los datos presentan suficientes evidencias para no rechazar H_0 al nivel de significación del 5%. En consecuencia, se puede concluir que las personas no muestran preferencias con respecto a la elección de una puerta determinada.

Usando el p-valor, es decir,

$> p.valor < -1 - pchisq(chi2.obs, k-1-r)$

$> p.valor$

[1] 0.3328711

Obtenemos como resultado $p - valor = 0.3328711$. Como $p - valor > \alpha$, los datos presentan suficientes evidencias para no rechazar H_0 al nivel de significación del 5%. En consecuencia, se puede concluir que las personas no muestran preferencias con respecto a la elección de una puerta determinada.

(2) Disponemos de una muestra de 250 equipos electrónicos, cuyos costos de reparación (en Bs.F) son los presentados en la tabla adjunta y queremos saber si los datos de esta muestra provienen de una distribución normal. Use $\alpha = 0.05$.

Costos (reparación)	**Número de equipos**
30-40	16
40-50	18
50-60	22
60-70	51
70-80	62
80-90	55
90-100	22
100-110	4

Solución:

Tenemos H_0: Los datos provienen de una distribución normal, contra:

H_1: Los datos no provienen de una distribución normal.

$> alpha < -0.05$

$> r < -2$

$> fi < c(16, 18, 22, 51, 62, 55, 22, 4)$

$> k < -length(fi)$

$> k$
[1] 8
$> n < -sum(fi)$
$> n$
[1] 250
$> yi < -c(35, 45, 55, 65, 75, 85, 95, 105)$
$> xbarra < -sum(fi * yi)/n$
$> x.barra < -rep(xbarra, k)$
$> S.cuadrado < -sum(fi * (yi - x.barra) \wedge 2)/(n - 1)$
$> S < -sqrt(S.cuadrado)$
$> pi < -pnorm(4 : 11 * 10, xbarra, S) - pnorm(3 : 10 * 10, xbarra, S)$
$> chi2.obs < -sum((fi - n * pi) \wedge 2/(n * pi))$
$> chi2.obs$
[1] 28.85710
$> chi2.alpha < -qchisq(1 - alpha, k - 1 - r)$
$> chi2.alpha$
[1] 11.07050

Obtenemos $\chi^2_{obs} = 28.85710$ y $\chi^2_{k-1-r;\alpha} = 11.07050$
Entonces, la región de rechazo es $RR = (11.07050\,,\,\infty)$
Como $\chi^2_{obs} \in RR$, los datos presentan suficientes evidencias para rechazar H_0 al nivel de significación del 5%. Por lo tanto, los datos no provienen de una distribución normal.

Por otra parte, para calcular el p-valor:
$> p.valor < -1 - pchisq(chi2.obs, k - 1 - r)$
$> p.valor$
[1]0.000025

Como el $p - valor \leq \alpha$, los datos presentan suficientes evidencias para rechazar H_0 al nivel de significación del 5%. Por lo tanto, los datos no provienen de una distribución normal.