**Problema 12:**

En una clase de 40 estudiantes, se hace una encuesta sobre el número de llamadas por teléfono móvil que hacen al día. Los resultados obtenidos son los siguientes:

| Número de llamadas | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| Frecuencia | 12 | 14 | 7 | 4 | 2 | 1 |

Se pide:

(a) Contrastar si el número de llamadas sigue una distribución de Poisson. ($\alpha = 0.05$).

**i) Planteamiento:** $\begin{cases} H_0 : X \sim P(\lambda) \\ H_1 : X \nsim P(\lambda) \end{cases}$

**ii) Elección de un estadístico *D*** como medida de discrepancia.

Como la variable es **discreta** y $\boldsymbol{n \geq 30}$, el caso es propicio para utilizar el **test de la Chi-cuadrado**.

**iii) Elección del criterio de rechazo**

Nos dan $\boldsymbol{\alpha = 0.05}$ y estudiaremos los dos criterios de rechazo: región crítica y p-valor.

**iv) Calcular el valor de $D$** a partir de la muestra y **aplicar el criterio de decisión**.

A partir de los datos de la muestra, construimos la tabla de frecuencias ($n_i$) y de frecuencias esperadas ($np_i$).

**Para hallar las frecuencias esperadas**, hemos de obtener los valores $p_i$ (probabilidad de cada valor o clase suponiendo $H_0$ cierta). Como $H_0$: $X \sim P(\lambda)$, para calcular dichas probabilidades hemos de tener un valor para $\lambda$.

Para ello, estimamos $\lambda \approx \overline{X}$ (que es el estimador tanto por el método de los momentos como por máxima verosimilitud). En este caso $\lambda \approx 1.25$.

Por tanto, utilizando la fórmula de la probabilidad de la Poisson: $p_i = P(X = x_i) = e^{-1.25} \dfrac{1.25^{x_i}}{x_i!}$ [1]

Las frecuencias esperadas se calculan como $n \cdot p_i = 40 \cdot p_i$ ($n=40$ es el tamaño de la muestra, que se obtiene sumando todas las frecuencias).

Tenemos la siguiente **tabla**, en donde tenemos que incluir **todos los valores** que la variable puede tomar, aunque no los haya tomado en la muestra (por eso añadimos la clase de los valores >5):

| Valores | ni: Frecuencia | pi: Probabilidad $H_0$ cierta | n*pi: Frecuencia esperada |
|---|---|---|---|
| 0 | 12 | 0,2865 | 11,46 |
| 1 | 14 | 0,3581 | 14,324 |
| 2 | 7 | 0,2238 | 8,952 |
| 3 | 4 | 0,0932 | 3,728 |
| 4 | 2 | 0,0291 | 1,164 |
| 5 | 1 | 0,0073 | 0,292 |
| >5 | 0 | 0,0018 | 0,072 |

[1] En este caso no se utilizan tablas, pues no las hay para $\lambda$=1.25

En **rojo** hemos señalado los valores cuya **frecuencia esperada** (***np$_i$***) es **menor que 5**, y hemos de construir una nueva tabla en donde agrupamos los valores de forma que se cumpla la condición necesaria para asegurar que el estadístico ***D*** sigue una distribución Chi-Cuadrado: ***np$_i$* ≥5**, para todo *i*

**Agrupamos adecuadamente los datos** y obtenemos:

| Valores | ni: Frecuencia | pi: Probabilidad H0 cierta | npi: Frecuencia esperada | di |
|---|---|---|---|---|
| 0 | 12 | 0,2865 | 11,46 | 0,0254 |
| 1 | 14 | 0,3581 | 14,324 | 0,0073 |
| 2 | 7 | 0,2238 | 8,952 | 0,4256 |
| [3, inf) | 7 | 0,1314 | 5,256 | 0,5787 |
| | | | | |
| | | | d= | 1,0371 |

En la última columna, en azul, hemos calculado el valor del estadístico $D = \sum_{i=1}^{k} \underbrace{\frac{(n_i - n\,p_i)^2}{n\,p_i}}_{Di}$, a partir de los datos de la muestra y obtenemos **d=1.0371**.

Ahora, ya podemos asegurar que $D \approx \chi^2_{k-r-1}$.

En este caso:

- ***k=4***, pues es el número de clases en las que finalmente agrupamos
- ***r=1***, pues hemos obtenido de la muestra una estimación del parámetro $\lambda$ (**un** parámetro).

Por tanto, si $H_0$ es cierta, se tiene que $D \approx \chi^2_2$.

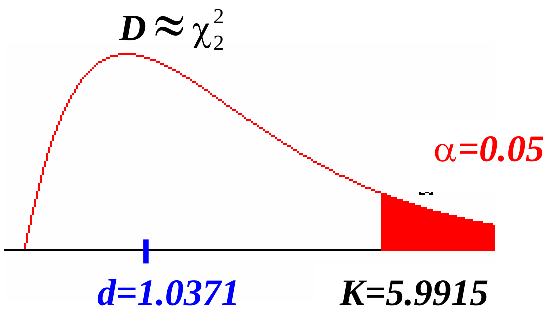
- **Criterio región crítica:**

Como $\alpha = 0.05$, buscamos un valor $K \in \mathbb{R}$, tal que:

$$P\left(D > K \big/ H_0\ cierta\right) = P\left(\chi^2_2 > K\right) = 0.05$$

$$\underset{TABLAS}{\Rightarrow} K = 5.9915$$

Como **d=1.0371 < *K*=5.9915, no rechazamos H$_0$** (es aceptable suponer que la variable sigue un modelo de Poisson)



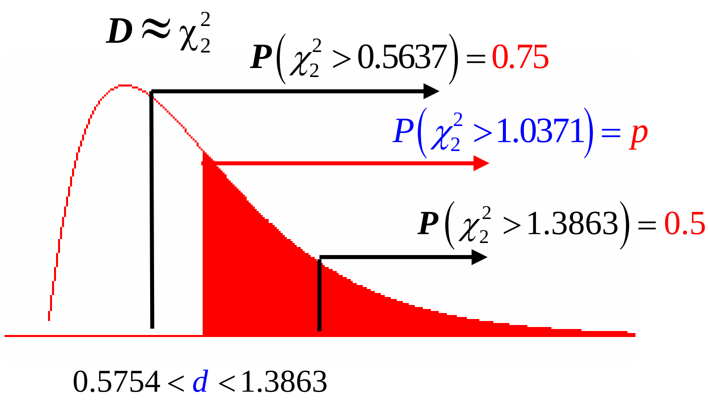

- **Criterio del p-valor:**

Calculamos el p-valor, según su definición:

$$p = P\left(D > d \big/ H_0\ cierta\right) = P\left(\chi^2_2 > 1.0371\right)$$

A partir de las **tablas**, también podemos obtener información sobre el p-valor.

Buscamos en las tablas los valores entre los que esté el valor obtenido ***d***, y calculamos la probabilidad que dejan a su **derecha** (las tablas dan el valor que dejan a su izquierda).



Distribución CHI cuadrado acumulada $\chi^2_k$ 1-p Valores con redondeo en el 5º decimal

Valores x tales que $P(X_k \le x)=p$ k es el nº de grados de libertad

| k \ p | 0,005 | 0,01 | 0,025 | 0,05 | 0,1 | 0,25 | 0,5 | 0,75 | 0,9 | 0,95 | 0,975 | 0,99 | 0,995 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0,0000 | 0,0002 | 0,0010 | 0,0039 | 0,0158 | 0,1015 | 0,4549 | 1,3233 | 2,7055 | 3,8415 | 5,0239 | 6,6349 | 7,8794 |
| 2 | 0,0100 | 0,0201 | 0,0506 | 0,1026 | 0,2107 | 0,5754 | 1,3863 | 2,7726 | 4,6052 | 5,9915 | 7,3778 | 9,2104 | 10,5965 |
| 3 | 0,0717 | 0,1148 | 0,2158 | 0,3518 | 0,5844 | 1,2125 | 2,3660 | 4,1083 | 6,2514 | 7,8147 | 9,3484 | 11,3449 | 12,8381 |
| 4 | 0,2070 | 0,2971 | 0,4844 | 0,7107 | 1,0636 | 1,9226 | 3,3567 | 5,3853 | 7,7794 | 9,4877 | 11,1433 | 13,2767 | 14,8602 |

d=1.0371

El p-valor estará comprendido entre dichas probabilidades. En este caso: **0.5 <** ***p*** < 0.75.

Como $p>\alpha=0.05$, **no rechazamos $H_0$ .**

**Conclusión:** Puede considerarse que el *número de llamadas* sigue una distribución de Poisson.

(b) Contrastar si sigue una distribución de Poisson de media 1 con $\alpha$ = 0.05. Obtener el p-valor del contraste.

**i) Planteamiento:** $\begin{cases} H_0 : X \sim P(1) \\ H_1 : X \nsim P(1) \end{cases}$

**ii)** y **iii)** igual que antes.

**iv) Calcular el valor de $D$** a partir de la muestra y **aplicar el criterio de decisión**.

A partir de los datos de la muestra, construimos la tabla de frecuencias ($n_i$) y de frecuencias esperadas ($np_i$), pero en este caso no hay que hacer ninguna estimación de $\lambda$. Además, podemos obtener los valores de $p_i$ a partir de las tablas.

| Valores | ni: Frecuencia | pi: Probabilidad H0 cierta | npi: Frecuencia esperada |
|---|---|---|---|
| 0 | 12 | 0,3679 | 14,715 |
| 1 | 14 | 0,3679 | 14,715 |
| 2 | 7 | 0,1839 | 7,358 |
| 3 | 4 | 0,0613 | 2,453 |
| 4 | 2 | 0,0153 | 0,613 |
| 5 | 1 | 0,0031 | 0,123 |
| >5 | 0 | 0,0006 | 0,024 |

En **rojo** señalamos los valores cuya **frecuencia esperada** (***$np_i$***) es **menor que 5**, y hemos de construir una nueva tabla en donde agrupamos los valores de forma que se cumpla la condición necesaria para asegurar que el estadístico ***D*** sigue una distribución Chi-Cuadrado: **$np_i \geq 5$**, para todo *i*

**Agrupamos adecuadamente los datos** y obtenemos:

| Valores | ni: Frecuencia | pi: Probabilidad H0 cierta | npi: Frecuencia esperada | di |
|---|---|---|---|---|
| 0 | 12 | 0,3679 | 14,716 | 0,5013 |
| 1 | 14 | 0,3679 | 14,716 | 0,0348 |
| [2, inf) | 14 | 0,264241118 | 10,56964471 | 1,1133 |
| | | | d= | 1,6494 |

Ahora, ya podemos asegurar que $D \approx \chi^2_{k-r-1}$.

En este caso:

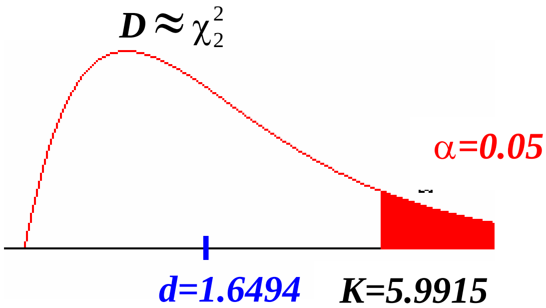
- o $k=3$, pues es el número de clases en las que finalmente agrupamos
- o $r=0$, pues **no** hemos tenido que estimar el parámetro $\lambda$.

Por tanto, si $H_0$ es cierta, se tiene que $D \approx \chi^2_2$.

- **Criterio región crítica:**

Como $\alpha=0.05$ y $D \approx \chi^2_2$, tenemos la misma región critica que antes.

Como **d=1.6494 < *K*=5.9915, no rechazamos $H_0$** (es aceptable suponer que la variable sigue un modelo P(1))



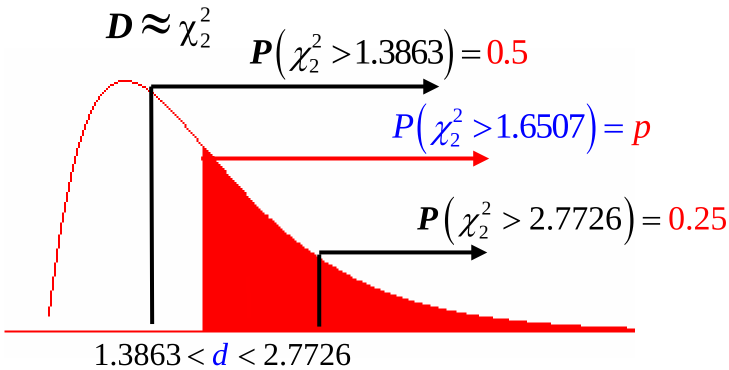

- **Criterio del p-valor:**

El p-valor sí va a ser diferente:

$$p = P\left(\chi^2_2 > 1.6507\right)$$

A partir de las **tablas**, también podemos obtener información sobre el p-valor. En este caso: **0.25 <** $p$ < 0.5.[2]

Como $p > \alpha = 0.05$, **no rechazamos $H_0$ .**




Distribución CHI cuadrado acumulada $\chi^2_k$ 1-p Valores con redondeo en el 5º decimal

Valores x tales que $P(X_k \le x) = p$ k es el nº de grados de libertad

| k \| p | 0,005 | 0,01 | 0,025 | 0,05 | 0,1 | 0,25 | 0,5 | 0,75 | 0,9 | 0,95 | 0,975 | 0,99 | 0,995 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0,0000 | 0,0002 | 0,0010 | 0,0039 | 0,0158 | 0,1015 | 0,4549 | 1,3233 | 2,7055 | 3,8415 | 5,0239 | 6,6349 | 7,8794 |
| 2 | 0,0100 | 0,0201 | 0,0506 | 0,1026 | 0,2107 | 0,5754 | 1,3863 | 2,7726 | 4,6052 | 5,9915 | 7,3778 | 9,2104 | 10,5965 |
| 3 | 0,0717 | 0,1148 | 0,2158 | 0,3518 | 0,5844 | 1,2125 | 2,3660 | 4,1083 | 6,2514 | 7,8147 | 9,3484 | 11,3449 | 12,8381 |
| 4 | 0,2070 | 0,2971 | 0,4844 | 0,7107 | 1,0636 | 1,9226 | 3,3567 | 5,3853 | 7,7794 | 9,4877 | 11,1433 | 13,2767 | 14,8602 |

d=1.6507

**Conclusión:** Puede considerarse que el *número de llamadas* sigue una distribución P(1).

[2] Con Statgraphics obtenemos $P\left(\chi^2_2 > 1.6507\right) = 0.44$, que permite aceptar con cierta confianza que X puede seguir una P(1).