COLECCIÓN DE CUESTIONES TEÓRICAS

1.t. Una variable aleatoria χ^2 tiene 10 grados de libertad. Hallar la media, la varianza y la probabilidad de que dicha variable aleatoria sea mayor que 9,342.

Solución:

La media y varianza de la χ^2 de Pearson: $\mu = 10$ $\sigma^2 = 2 \,.\, 10 = 20$

$$P\left[\chi^2_{10} > 9{,}342\right] = 0{,}5$$

2.t. Se consideran dos variables aleatorias independientes X e Y. La variable X tiene una distribución normal N(0,1). La variable Y tiene una distribución χ^2 con 4 grados de libertad.

Hallar en $P\left[\dfrac{2X}{\sqrt{Y}} \le m\right] = 0{,}05$ el valor de m

Solución: $X \sim N(0\ ;1)$ $Y \sim \chi^2_4$ $t_n = \dfrac{z}{\sqrt{\dfrac{1}{n}\chi^2_n}}$

con lo cual, $t_4 = \dfrac{z}{\sqrt{\dfrac{1}{4}\chi^2_4}} = \dfrac{z}{\dfrac{1}{2}\sqrt{\chi^2_4}} = \dfrac{2z}{\sqrt{\chi^2_4}}$

$$P\left[\frac{2X}{\sqrt{Y}} \le m\right] = P\left[t_4 \le m\right] = 0{,}05 = P\left[t_4 \ge -m\right] = 0{,}05 \quad \mapsto \quad m = -2{,}132$$

3.t. ¿Cuándo se deberá utilizar un contraste de independencia y cuando uno de homogeneidad?

Solución:

Contraste de independencia: cuando se trata de contrastar si existe dependencia entre dos características de la misma población.
Contraste de homogeneidad: cuando se desea contrastar si dos o más muestras proceden de la misma población.

4.t. Señale qué características pueden considerarse esenciales en el planteamiento de un contraste paramétrico.

- ☐ Formulación de las hipótesis nula H_0 y alternativa H_1 en términos estadísticos. Ambas hipótesis deben ser mutuamente excluyentes.
- ☐ Determinación del test estadístico o estadístico de prueba apropiado
- ☐ Selección del nivel de significación α
- ☐ Determinación de la región crítica.
- ☐ Selección aleatoria de la muestra
- ☐ Establecimiento de la regla de decisión y su interpretación.

Solución:

- ☒ Formulación de las hipótesis nula H_0 y alternativa H_1 en términos estadísticos. Ambas hipótesis deben ser mutuamente excluyentes.
- ☒ Determinación del test estadístico o estadístico de prueba apropiado
- ☒ Selección del nivel de significación α
- ☒ Determinación de la región crítica.
- ☒ Selección aleatoria de la muestra
- ☒ Establecimiento de la regla de decisión y su interpretación.

5.t. Concepto de nivel de significación y potencia de un contraste. Relación entre ambos.

Solución:

El nivel de significación α de un contraste es la probabilidad de cometer error de tipo I, siendo la probabilidad de rechazar la hipótesis nula, siendo ésta cierta. También se denomina tamaño de la región crítica (o de rechazo) ya que la probabilidad de que el estimador pertenezca a la región crítica es precisamente α.

La potencia del contraste es la probabilidad de rechazar la hipótesis nula. La probabilidad de cometer un error de tipo II se denota por β, siendo la probabilidad de aceptar la hipótesis nula, siendo falsa. En consecuencia:

$$\text{Potencia del contraste} = \begin{cases} \alpha & \text{si } H_0 \text{ cierta} \\ 1-\beta & \text{si } H_0 \text{ falsa} \end{cases}$$

Para un tamaño muestral n fijo, si α aumenta, entonces β disminuye y, por lo tanto, $1-\beta$ aumenta, y viceversa.

6.t. Error cuadrático medio de un estimador: concepto. ¿Para qué se utiliza?.

Solución:

Sea $\hat{\theta}$ el estimador de un parámetro poblacional θ. Se define el error cuadrático medio como el valor de $E\left[(\hat{\theta} - \theta)^2\right]$:

$$ECM(\hat{\theta}) = E\left[(\hat{\theta} - \theta)^2\right]$$

Si al valor $E\left[(\hat{\theta} - \theta)^2\right]$ se suma y se resta $E(\hat{\theta})$:

$$E\left[(\hat{\theta} - E(\hat{\theta}) + E(\hat{\theta}) - \theta)^2\right] = E\left[\left(\hat{\theta} - E(\hat{\theta})\right)^2\right] + E\left[\left(E(\hat{\theta}) - \theta\right)^2\right] + 2\,E\left[\left[\overbrace{(\hat{\theta} - E(\hat{\theta})}^{=0}\right].\left[E(\hat{\theta}) - \theta)\right]\right] =$$

$$= Var\left[\hat{\theta}\right] + \left(E(\hat{\theta}) - \theta\right)^2$$

El valor $b(\hat{\theta}) = E(\hat{\theta}) - \theta$ se denota como sesgo de $\hat{\theta}$, quedando: $ECM(\hat{\theta}) = Var(\hat{\theta}) + \left[b(\hat{\theta})\right]^2$

En el problema de la estimación puntual interesa que el error cuadrático medio sea lo menor posible, se consigue cuanto menor sean la varianza de $\hat{\theta}$ y el valor absoluto del sesgo $b(\hat{\theta})$.
Si el estimador es insesgado $\left[b(\hat{\theta}) = 0\right]$ el error cuadrático medio $ECM(\hat{\theta}) = Var(\hat{\theta})$.

7.t. ¿Cuál es el objetivo de la estimación por intervalos de confianza? Razone la respuesta.

Solución:

Es establecer un intervalo de poca amplitud y alta probabilidad (coeficiente de confianza) de modo que en su interior se encuentre un determinado parámetro de la distribución de la variable aleatoria.

8.t. Se supone que la rentabilidad de un producto ofrecido por una entidad bancaria es una variable aleatoria que tiene como función de densidad:

$$f(x) = \begin{cases} k(x^2 - 1) & \text{si } 0 \le x \le 2 \\ 0 & \text{en otro caso} \end{cases}$$

Hallar el valor de k para que $f(x)$ sea función de densidad

Solución:

El ejercicio carece de sentido, $f(x)$ no puede ser función de densidad para ningún valor de k:

$$\begin{cases} k < 0 & f(x) < 0 \quad \text{en el intervalo } [1,\, 2) \\ k = 0 & f(x) = 0 \quad \forall x \in R \\ k > 0 & f(x) < 0 \quad \text{en el intervalo } [0,\, 1) \end{cases}$$

Una función de densidad $f(x)$ es una función real no negativa.

COLECCIÓN DE EJERCICIOS PRÁCTICOS

1. En una empresa de construcción el número mensual medio de días de baja de los trabajadores debidas a accidentes laborales por obra, se distribuye según una Poisson con varianza 15,6. Calcular la probabilidad de que, en cuatro obras independientes, se acumulen más de 68 días, en total, de baja en el periodo de un mes.

Solución:

En una distribución de Poisson $\sigma^2 = \lambda = 15{,}6$, si

$$x_i : P(\lambda = 15{,}6) \quad a \quad \sum_{i=1}^{4} x_i : P(4 \,.\, 15{,}6 = 62{,}4) \quad ; \; N(\lambda = 62{,}4 \,; \sqrt{62{,}4} = 7{,}9)$$

Efectuando la corrección de continuidad:

$$P\big[X > 68{,}5\big] = P\left[\frac{X - 62{,}4}{7{,}9} > \frac{68{,}5 - 62{,}4}{7{,}9}\right] = P\left[z > \frac{68{,}5 - 62{,}4}{7{,}9}\right] = P\big[z > 0{,}77\big] = 0{,}2206$$

2. Una cadena comercial que tiene 81 establecimientos por Europa ha comprobado que el volumen de ventas mensuales de cada uno, en millones de euros, se ajusta a la función de densidad $f(x) = \dfrac{x}{50}$ para $0 \le x \le 10$. Suponiendo que las ventas de los distintos establecimientos son independientes, se desea conocer:

a) La probabilidad de que la venta total de la cadena en un mes determinado sea mayor de 550 millones de euros.

b) El volumen de venta total que es superado con una probabilidad del 30%.

Solución:

a) Sea $X_i =$'millones de euros de venta mensuales en el establecimiento i-ésimo'

$$\alpha_1 = \mu = \int_0^{10} x \,.\, f(x)\, dx = \int_0^{10} x \,.\, \frac{x}{50}\, dx = \int_0^{10} \frac{x^2}{50}\, dx = \frac{1}{50} \left.\frac{x^3}{3}\right|_0^{10} = \frac{10^3}{3 \,.\, 50} = \frac{20}{3}$$

$$\alpha_2 = \int_0^{10} x^2 \,.\, f(x)\, dx = \int_0^{10} x^2 \,.\, \frac{x}{50}\, dx = \int_0^{10} \frac{x^3}{50}\, dx = \frac{1}{50} \left.\frac{x^4}{4}\right|_0^{10} = \frac{10000}{200} = 50$$

$$\sigma^2 = \alpha_2 - \alpha_1^2 = 50 - \left(\frac{20}{3}\right)^2 = \frac{50}{9} \quad a \quad \sigma = \sqrt{\frac{50}{9}} = \frac{5\sqrt{2}}{3}$$

La variable $X = \sum_{i=1}^{81} X_i : N\left[81\mu \; ; \; \sqrt{81}\ \sigma\right] \equiv N\left[540 \; ; \; 15\sqrt{2}\right]$

$$P\left[X > 550\right] = P\left[\frac{X-540}{15\sqrt{2}} > \frac{550-540}{15\sqrt{2}}\right] = P\left[z > \frac{10}{15\sqrt{2}}\right] = P\left[z > 0{,}47\right] = 0{,}3192$$

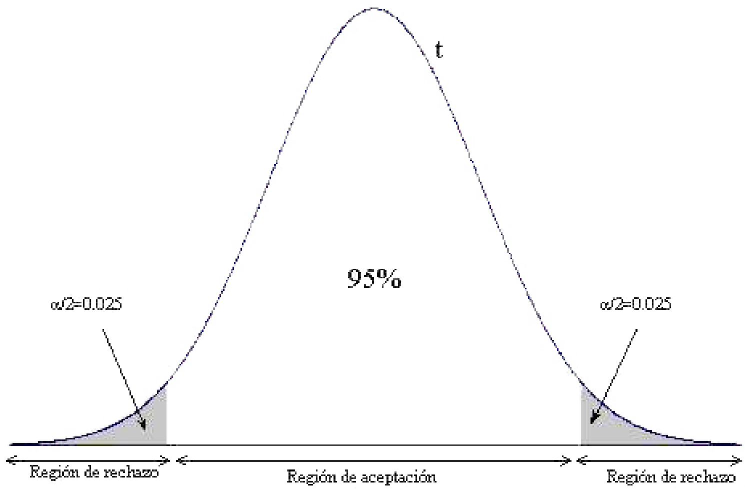
b) $P\left[X > k\right] = P\left[\frac{X-540}{15\sqrt{2}} > \frac{k-540}{15\sqrt{2}}\right] = P\left[z > \frac{k-540}{15\sqrt{2}}\right] = 0{,}3 \quad \mapsto \quad \frac{k-540}{15\sqrt{2}} = 0{,}52$

$$k - 540 = 0{,}52 \,.\, 15\sqrt{2} \quad \mapsto \quad k = 540 + 0{,}52 \,.\, 15\sqrt{2} \simeq 551 \text{ millones de euros}$$

3. Se sabe que las pérdidas anuales de un determinado grupo de empresas se distribuyen según una ley normal de media 6000 euros. Se toma una muestra de diez empresas y se obtiene una varianza muestral de 250 (euros^2).
Calcule los valores a y b para los que se verifica: P (a $\leq \overline{x} \leq$ b) = 0,950.

Solución:

La variable $\dfrac{\overline{x} - \mu}{\sigma_x^2 \big/ \sqrt{n-1}} \sim t_{n-1}$



Para 9 grados de libertad, de las tablas se obtiene: $P\left[-2{,}262 \leq t_9 \leq 2{,}262\right] = 0{,}950$

En consecuencia, $-2{,}262 \leq \dfrac{\overline{x} - 6000}{\sqrt{250/9}} \leq 2{,}262 \quad \mapsto \quad -2{,}262 \;.\; 5{,}270 \leq \overline{x} - 6000 \leq 2{,}262 \;.\; 5{,}270$

$a = 5888{,}07 \leq \overline{x} \leq b = 6011{,}92$

4. La rentabilidad de un producto ofrecido por una entidad bancaria es una variable aleatoria continua con distribución uniforme de media 5% y recorrido de un 6%. Se pide:

a) Función de densidad y de distribución de la variable aleatoria

b) Probabilidad de que la rentabilidad sea inferior al 5%

c) Probabilidad de que la rentabilidad sea exactamente un 7%

d) Varianza de la rentabilidad del producto

Solución:

a) Se tiene que $\begin{cases} \dfrac{a+b}{2}=5 \\ b-a=6 \end{cases} \mapsto \quad a=2 \quad b=8$

Función de densidad $f(x)=\begin{cases} \dfrac{1}{6} & 2\le x\le 8 \\ 0 & \text{en el resto} \end{cases}$ Función de distribución $F(x)=\begin{cases} 0 & x<2 \\ \dfrac{x-2}{6} & 2\le x<8 \\ 1 & x\ge 8 \end{cases}$

b) $P\left[X<5\right]=F\left[5\right]=\dfrac{5-2}{6}=\dfrac{1}{2}$

c) $P\left[X=7\right]=0$ por tratarse de una distribución continua.

d) $Var(X)=\dfrac{(8-2)^2}{12}=3$

5. Una agencia de viajes supone que el 70% de la población ha salido al extranjero de vacaciones en alguna ocasión. Con objeto de obtener una información más precisa toma una muestra aleatoria simple.

a) Obtener la distribución muestral del estadístico correspondiente para una muestra aleatoria de 90 personas y de 250 personas.

b) Gráfica de las funciones de densidad del estadístico correspondiente para ambas muestras.

c) Explica con qué muestra se obtiene mayor potencia en el contraste.

Solución:

a) La hipótesis de la agencia de viajes supone que la proporción poblacional de personas que han salido al extranjero en alguna ocasión es 0,7. Para obtener una información más precisa del parámetro poblacional (es decir, para contrastar dicha hipótesis) se toma el estadístico proporción muestral.

La proporción muestral $\hat{p} \sim N\left[p\,,\ \sqrt{\dfrac{p\,.\,q}{n}}\right]$

- Para la muestra $n = 90 \;\mapsto\; \hat{p} \sim N\left[0,7;\ \sqrt{\dfrac{0,7\,.\,0,3}{90}}\right] \equiv N\left[0,7\ ;\ 0,048\right]$

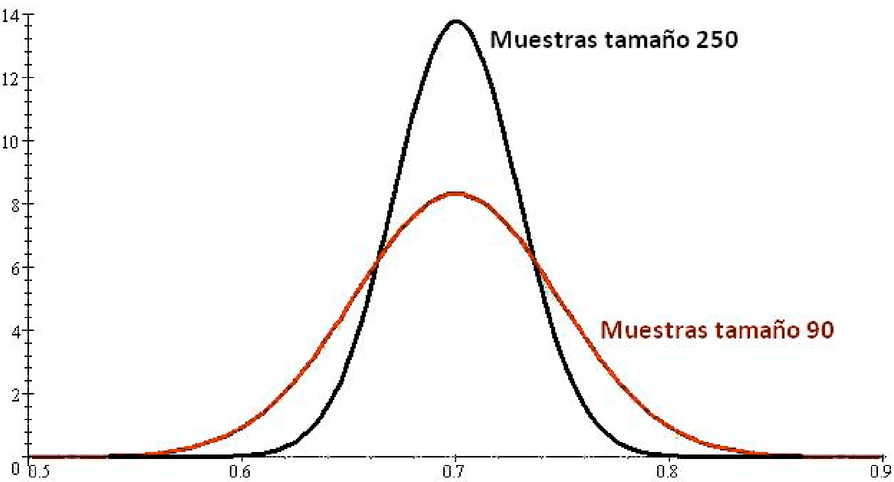
- Para la muestra $n = 250 \;\mapsto\; \hat{p} \sim N\left[0,7;\ \sqrt{\dfrac{0,7\,.\,0,3}{250}}\right] \equiv N\left[0,7\ ;\ 0,029\right]$

b) La funciones de densidad son dos campanas de Gauss, con máximo en el punto de abscisa 0,7 y ordenada $\dfrac{1}{\sigma\,.\,\sqrt{2\pi}}$.

- Para la muestra $n = 90 \;\mapsto\; \left[0,7\ ;\ \dfrac{1}{0,048\,.\,\sqrt{2\pi}}\right] \equiv \left[0,7\ ;\ 8,31\right]$

- Para la muestra $n = 250 \;\mapsto\; \left[0,7\ ;\ \dfrac{1}{0,029\,.\,\sqrt{2\pi}}\right] \equiv \left[0,7\ ;\ 13,76\right]$

Los puntos de inflexión tienen abscisas, en cada caso, en $\left[0,7 \pm \sigma\right]$



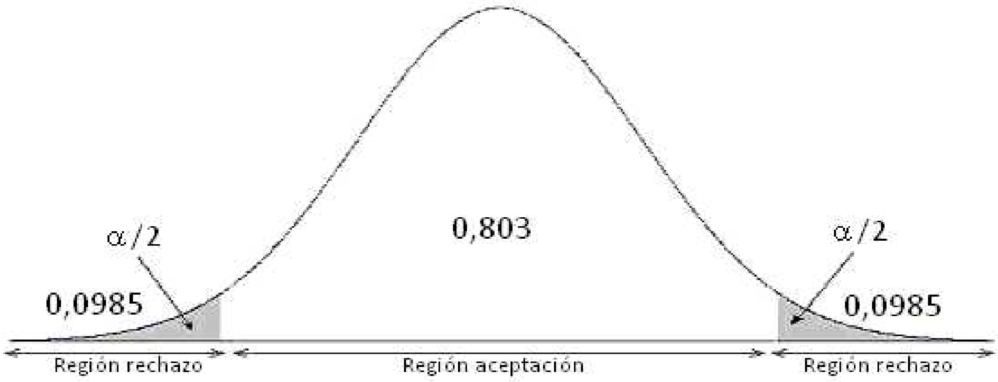
c) Para un mismo nivel de significación, el tamaño de la región de aceptación es menor para muestras de tamaño 250 y en consecuencia más improbable cometer un error tipo II, es decir, el contraste efectuado con muestras de tamaño 250 tiene mayor potencia que el efectuado con muestras de tamaño 90.

6. Una variable aleatoria sigue una distribución normal con una varianza de 2. Seleccionada una muestra aleatoria simple $(X_1,\ X_2,\ \cdots\ ,X_n)$ con media $\overline{X}$.
Hallar el tamaño muestral necesario para que se verifique $P\left[\mu-0,30<\overline{X}<\mu+0,30\right]=0,803$

Solución:

La media muestral $\overline{X}\sim N\left[\mu,\ \dfrac{\sqrt{2}}{\sqrt{n}}\right]$

$$P\left[\mu-0,30<\overline{X}<\mu+0,30\right]=P\left[\frac{\mu-0,30-\mu}{\dfrac{\sqrt{2}}{\sqrt{n}}}<\frac{\overline{X}-\mu}{\dfrac{\sqrt{2}}{\sqrt{n}}}<\frac{\mu+0,30-\mu}{\dfrac{\sqrt{2}}{\sqrt{n}}}\right]=P\left[\frac{-0,30\,.\sqrt{n}}{\sqrt{2}}<z<\frac{0,30\,.\sqrt{n}}{\sqrt{2}}\right]=0,803$$



$$P\left[z>\frac{0,30\,.\sqrt{n}}{\sqrt{2}}\right]=0,0985\ \mapsto\ \frac{0,30\,.\sqrt{n}}{\sqrt{2}}=1,29\ \mapsto\ \sqrt{n}=\frac{1,29\,.\sqrt{2}}{0,30}\ \mapsto\ n=\left[\frac{1,29\,.\sqrt{2}}{0,30}\right]^2=36,98$$

Se toma $n=37$

7. Comprobar si la distribución de baterías de la tabla adjunta se distribuyen según una ley normal $N(3,5;\ 0,7)$ a un nivel 0,05

Intervalos	1,45 – 1,95	1,95 – 2,45	2,45 – 2,95	2,95 – 3,45	3,45 – 3,95	3,95 – 4,45	4,45 – 4,95
Frecuencia	2	1	4	15	10	5	3

Solución:

Se plantean las hipótesis:

H_0: Los datos provienen de la distribución normal $N(3,5;\ 0,7)$

H_1: Los datos no provienen de la distribución normal $N(3,5;\ 0,7)$

En este ejercicio se conoce la media y desviación típica de la población, por lo que no se tiene que estimar. En caso de que no se conocieran, se calculan con los datos recogidos $(x_i\,,\ n_i)$ la media $\mu = 3,4125$ y la desviación típica $\sigma = 0,6881$, teniendo en cuenta que en el estadístico χ^2_{k-p-1} el parámetro $p = 2$ ya que se habían calculado tres parámetros $(\mu\,,\ \sigma)$

Intervalos	x_i	n_i	p_i	$e_i = n\,.\,p_i$
1,45 – 1,95	1,7	2	0,0119	0,476
1,95 – 2,45	2,2	1	0,0532	2,128
2,45 – 2,95	2,7	4	0,148	5,92
2,95 – 3,45	3,2	15	0,2573	10,292
3,45 – 3,95	3,7	10	0,2668	10,672
3,95 – 4,45	4,2	5	0,1742	6,968
4,45 – 4,95	4,7	3	0,067	2,68
		40		

Es necesario que las frecuencias esperadas de las distintas modalidades no sea inferior a 5, $e_i \geq 5\ \ \forall i$, teniendo que agrupar modalidades contiguas en una sola hasta lograrlo.

Mediante la tabla normal se hallan las probabilidades de cada uno de los intervalos

$$P[1,45 \leq x < 1,95] = P\left[\frac{1,45-3,5}{0,7} \leq \frac{x-3,5}{0,7} < \frac{1,95-3,5}{0,7}\right] = P[-2,92 \leq x < -2,21] = 0,0136 - 0,0017 = 0,0119$$

$$P[1,95 \leq x < 2,45] = P\left[\frac{1,95-3,5}{0,7} \leq \frac{x-3,5}{0,7} < \frac{2,45-3,5}{0,7}\right] = P[-2,21 \leq x < -1,5] = 0,0668 - 0,0136 = 0,0532$$

$$P[2,45 \leq x < 2,95] = P\left[\frac{2,45-3,5}{0,7} \leq \frac{x-3,5}{0,7} < \frac{2,95-3,5}{0,7}\right] = P[-1,5 \leq x < -0,79] = 0,2148 - 0,0668 = 0,148$$

$$P[2,95 \leq x < 3,45] = P\left[\frac{2,95-3,5}{0,7} \leq \frac{x-3,5}{0,7} < \frac{3,45-3,5}{0,7}\right] = P[-0,79 \leq x < -0,07] = 0,4721 - 0,2148 = 0,2573$$

$$P[3,45 \leq x < 3,95] = P\left[\frac{3,45-3,5}{0,7} \leq \frac{x-3,5}{0,7} < \frac{3,95-3,5}{0,7}\right] = P[-0,07 \leq x < 0,64] = 0,5279 - 0,2611 = 0,2668$$

$$P[3,95 \le x < 4,45] = P\left[\frac{3,95-3,5}{0,7} \le \frac{x-3,5}{0,7} < \frac{4,45-3,5}{0,7}\right] = P[0,64 \le x < 1,36] = 0,2611 - 0,0869 = 0,1742$$

$$P[4,45 \le x < 4,95] = P\left[\frac{4,45-3,5}{0,7} \le \frac{x-3,5}{0,7} < \frac{4,95-3,5}{0,7}\right] = P[1,36 \le x < 2,07] = 0,0869 - 0,0192 = 0,067$$

Agrupando modalidades contiguas para conseguir que cada una de ellas tenga una frecuencia esperada $e_i \ge 5 \ \forall i$, se tiene:

Intervalos	n_i	$e_i = n\,.\,p_i$	n_i^2	$\frac{n_i^2}{e_i}$
1,45 – 2,95	7	8,524	49	5,7485
2,95 – 3,45	15	10,292	225	21,8616
3,45 – 3,95	10	10,672	100	9,3703
3,95 – 4,95	8	9,648	64	6,6335
	40			43,6139

Los grados de libertad son $[k-p-1] = 4-0-1 = 3$, en este caso como ya se ha indicado anteriormente no ha sido necesario calcular los parámetros $(\mu,\ \sigma)$. En caso de que no se hubieran conocido, los grados de libertad serían: $[k-p-1] = 4-2-1 = 1$

El estadístico observado: $\chi^2_{4-0-1} = \sum_{i=1}^{4} \frac{n_i^2}{e_i} - n = \frac{49}{8,524} + \frac{225}{10,292} + \frac{100}{10,672} + \frac{64}{9,648} - 40 = 3,6139$

El estadístico teórico: $\chi^2_{0,05;\,3} = 7,815$

Como $\chi^2_3 = 3,6139 < \chi^2_{0,05;\,3} = 7,815$ se acepta la hipótesis nula, concluyendo que los datos proceden de la población normal $N(3,5;\ 0,7)$ a un nivel de significación de 0,05

8. El número de defectos en las tarjetas de circuito impreso sigue una distribución Poisson. Se reúne una muestra aleatoria de 60 tarjetas de circuito impreso y se observa el número de defectos. Los resultados obtenidos son los siguientes:

Número de defectos	0	1	2	3 o más
Frecuencia observada	32	15	9	4

¿Muestran los datos suficiente evidencia para decir que provienen de una distribución de Poisson con una fiabilidad del 95%?

Solución:

H_0: Los datos provienen de una distribución de Poisson $P(\lambda = \overline{x})$

H_1: Los datos no provienen de una distribución de Poisson $P(\lambda = \overline{x})$

La distribución de Poisson se caracteriza porque sólo depende de un parámetro λ que coincide con la media. En este sentido, para ajustar la distribución observada a la de Poisson es necesario calcular la media. Denotando por x = ' número de defectos en tarjetas impresas' y n_i = ' frecuencia observada', se tiene:

x_i	0	1	2	3 o más
n_i	32	15	9	4
$x_i \cdot n_i$	0	15	18	12

$$\overline{x} = \frac{\sum x_i \cdot n_i}{n} = \frac{45}{60} = 0,75 \quad \mapsto \quad \lambda = 0,75 \quad \mapsto \quad P(x=k) = \frac{0,75^k}{k!}\, e^{-0,75} \quad k = 0, 1, 2, 3$$

$$P(x=0) = \frac{0,75^0}{0!}\, e^{-0,75} = 0,472 \qquad P(x=1) = \frac{0,75}{1!}\, e^{-0,75} = 0,354$$

$$P(x=2) = \frac{0,75^2}{2!}\, e^{-0,75} = 0,133 \qquad P(x \geq 3) = 1 - P(x<3) = 1 - \left[0,472 + 0,354 + 0,133\right] = 0,041$$

x_i	n_i	p_i	$e_i = n \cdot p_i$
0	32	0,472	28,32
1	15	0,354	21,24
2	9	0,133	7,98
3 o más	4	0,041	2,46
	60		

Es necesario que las frecuencias esperadas de las distintas modalidades $e_i \geq 5 \ \forall i$, teniendo que agrupar las dos últimas modalidades.

x_i	n_i	$e_i = n \cdot p_i$	n_i^2	$\frac{n_i^2}{e_i}$
0	32	28,32	1024	36,1582
1	15	21,24	225	10,5932
2 o más	13	10,44	169	16,1877
	60			62,9392

Los grados de libertad son $\left[k - p - 1\right] = 3 - 1 - 1 = 1$

El estadístico observado: $\chi^2_{3-1-1} = \sum_{i=1}^{3} \frac{n_i^2}{e_i} - n = \frac{1024}{28,32} + \frac{225}{21,24} + \frac{169}{10,44} - 60 = 2,9392$

El estadístico teórico: $\chi^2_{0,05;\,1} = 3,841$

Como $\chi^2_1 = 2,9392 < \chi^2_{0,05;\,1} = 3,841$ se acepta la hipótesis nula, concluyendo que los datos proceden de una población de Poisson de parámetro $\lambda = 0,75$, a un nivel de significación de 0,05

9. Se dan tres sucesos aleatorios A, B y C, independientes dos a dos, los cuales, sin embargo, no pueden ocurrir simultáneamente. Suponiendo que todos ellos tienen igual probabilidad p, calcular el valor de p que hace máxima la probabilidad de ocurrencia de al menos uno de los sucesos.

Solución:

Por la hipótesis de partida:

$$P(A\cap B)=P(A).P(B)=p^2 \quad P(A\cap C)=P(A).P(C)=p^2 \quad P(B\cap C)=P(B).P(C)=p^2 \quad P(A\cap B\cap C)=0$$

$$P(A\cup B\cup C) = P(A) + P(B) + P(C) - P(A\cap B) - P(A\cap C) - P(B\cap C) + P(A\cap B\cap C) = 3p - 3p^2$$

Denotando por $f(p)=3p - 3p^2 \quad \mapsto \quad f'(p)=3 - 6p = 0 \quad \mapsto \quad p=\dfrac{1}{2}$

$$P(A\cup B\cup C) = \frac{3}{2}-\frac{3}{4}=\frac{3}{4}$$

10. A un puesto aduanero llega periódicamente misiones diplomáticas (constituidas por 10 miembros) procedentes de un determinado país. El mencionado país es un gran productor de marihuana, circunstancia, que de vez en cuando, es aprovechada por sus misiones diplomáticas para introducir algún cargamento en el país que visitan, siendo la forma de hacerlo el que dos de cada diez miembros lleven en sus maletas la hierba.
Los aduaneros tienen ya información, pero, para no producir incidentes diplomáticos, se limitan a inspeccionar dos de cada diez maletas dejando pasar a la misión si en las maletas inspeccionadas no encuentran droga. Su experiencia les dice además que el 10% de las misiones portan drogas.
Si una misión inspeccionada no arroja resultado positivo, ¿cuál es la probabilidad de que realmente dicha misión no lleve droga alguna?

Solución:

Sean los sucesos:

$A=$'La Delegación porta droga' $\qquad$ $B=$'La misión inspeccionada tiene resultado negativo'

$$P\left[A\right]=0,1 \quad P\left[\overline{A}\right]=0,9$$

$$P\left[B/A\right]=\frac{\dbinom{8}{2}\dbinom{2}{0}}{\dbinom{10}{2}}=\frac{\dfrac{8!}{2!\,6!}}{\dfrac{10!}{2!\,8!}}=\frac{28}{45}=0,6222 \qquad\qquad P\left[B/\overline{A}\right]=1$$

Aplicando el teorema de Bayes:

$$P\left[\overline{A}/B\right]=\frac{P\left[\overline{A}\right].P\left[B/\overline{A}\right]}{P\left[\overline{A}\right].P\left[B/\overline{A}\right]+P\left[A\right].P\left[B/A\right]}=\frac{0,9\,.\,1}{0,9\,.\,1+0,1\,.\,0,6222}=0,9353$$

11. Un mayorista tiene 200 clientes clasificados en la siguiente tabla según si realizan pedidos regularmente o de forma esporádica y según si efectúan el pago al contado o a través de créditos:

	Forma de pago	
Tipo de pedido	Al contado	Crédito
Regular	10	15
Esporádico	20	155

En el marco de una campaña publicitaria, el mayorista decide sortear un viaje entre sus clientes eligiendo uno de ellos al azar.

a) ¿Cuál es la probabilidad de que el cliente elegido al azar realice pedidos de forma regular o bien utilice créditos para efectuar sus pagos?

b) Calcule la probabilidad de que el cliente afortunando con el viaje realice pedidos regularmente si sabemos que el elegido efectúa sus pagos mediante créditos.

c) Calcule la probabilidad de que el cliente afortunado con el viaje realice los pagos mediante crédito si sabemos que realiza pedidos regularmente.

d) ¿Son independientes los sucesos 'comprar a crédito' y 'comprar regularmente'?

Solución:

	Forma de pago		
Tipo de pedido	Al contado	Crédito	
Regular	10	15	25
Esporádico	20	155	175
	30	170	200

Sean los sucesos:

R = 'un cliente realiza pedidos regularmente'

C = 'un cliente efectúa los pagos mediante créditos'

a) $P(R) = \dfrac{25}{200} = 0,125 \qquad P(C) = \dfrac{170}{200} = 0,85 \qquad P(R \cap C) = \dfrac{15}{200} = 0,075$

$$P(R \cup C) = P(R) + P(C) - P(R \cap C) = 0,125 + 0,85 - 0,075 = 0,9$$

b) $P(R / C) = \dfrac{P(R \cap C)}{P(C)} = \dfrac{0,075}{0,85} = 0,088$ o bien $P(R / C) = \dfrac{15}{170} = 0,088$

c) $P(C / R) = \dfrac{P(C \cap R)}{P(R)} = \dfrac{0,075}{0,125} = 0,6$ o bien $P(R / C) = \dfrac{15}{25} = 0,6$

d) No son independientes los sucesos puesto que $P(R) = 0,125 \neq P(R / C) = 0,088$

12. Las compañías de seguros de automóviles suelen penalizar en sus primas a los conductores más jóvenes, con el criterio que éstos son más propensos a tener un mayor número de accidentes. En base a la tabla adjunta, con un nivel de significación del 5%, contrastar si el número de accidentes es independiente de la edad del conductor.

Edad del conductor	Número de accidentes al año				
	0	1	2	3	4
25 o menos	10	10	20	40	70
26 - 35	20	10	15	20	30
más de 36	60	50	30	10	5

Solución:

Hipótesis nula H_0: 'El número de accidentes sufridos por los conductores no depende de la edad del conductor'

Se acepta H_0:

$$\chi^2_{(k-1).(m-1)} = \sum_{i=1}^{k}\sum_{j=1}^{m}\frac{(n_{ij}-e_{ij})^2}{e_{ij}} = \sum_{i=1}^{k}\sum_{j=1}^{m}\frac{n_{ij}^2}{e_{ij}} - n < \chi^2_{\alpha;\ (k-1).(m-1)}$$

Región de rechazo de la hipótesis nula: $R_{rechazo} = \left\{ \chi^2_{(k-1).(m-1)} \geq \chi^2_{\alpha;\ (k-1).(m-1)} \right\}$

Se forma la tabla de contingencia 3 x 5 donde cada frecuencia observada $(n_{ij})_{i=1,\cdots,k\ ;\ j=1,\cdots,m}$ tiene una frecuencia teórica o esperada en caso de independencia $e_{ij} = \dfrac{n_{i\bullet} \times n_{\bullet j}}{n}$

Edad del conductor	Número de accidentes por año: 0	1	2	3	4	$\sum_{j=1}^{m} n_{i\bullet}$
25 o menos	10 $e_{11}=33,75$	10 $e_{12}=26,25$	20 $e_{13}=24,37$	40 $e_{14}=26,25$	70 $e_{15}=39,37$	150 (150)
26 - 35	20 $e_{21}=21,37$	10 $e_{22}=16,62$	15 $e_{23}=15,44$	20 $e_{24}=16,62$	30 $e_{25}=24,94$	95 (95)
más de 36	60 $e_{31}=34,87$	50 $e_{32}=27,12$	30 $e_{33}=25,19$	10 $e_{34}=27,12$	5 $e_{35}=40,69$	155 (155)
$\sum_{i=1}^{k} n_{\bullet j}$	90	70	65	70	105	400

$e_{11} = \dfrac{150\,.\,90}{400} = 33,75$ $\quad e_{12} = \dfrac{150\,.\,70}{400} = 26,25$ $\quad e_{13} = \dfrac{150\,.\,65}{400} = 24,37$ $\quad e_{14} = \dfrac{150\,.\,70}{400} = 26,25$ $\quad e_{15} = \dfrac{150\,.\,105}{400} = 39,37$

$e_{21} = \dfrac{95\,.\,90}{400} = 21,37$ $\quad e_{22} = \dfrac{95\,.\,70}{400} = 16,62$ $\quad e_{23} = \dfrac{95\,.\,65}{400} = 15,44$ $\quad e_{24} = \dfrac{95\,.\,70}{400} = 16,62$ $\quad e_{25} = \dfrac{95\,.\,105}{400} = 24,94$

$e_{31} = \dfrac{155\,.\,90}{400} = 34,87$ $\quad e_{32} = \dfrac{155\,.\,70}{400} = 27,12$ $\quad e_{33} = \dfrac{155\,.\,65}{400} = 25,19$ $\quad e_{34} = \dfrac{155\,.\,70}{400} = 27,12$ $\quad e_{35} = \dfrac{155\,.\,105}{400} = 40,69$

Estadístico observado: $\chi^2_{(3-1)\,.\,(5-1)} = \chi^2_8 = \sum_{i=1}^{3}\sum_{j=1}^{5}\frac{(n_{ij} - e_{ij})^2}{e_{ij}} = \sum_{i=1}^{3}\sum_{j=1}^{5}\frac{n_{ij}^2}{e_{ij}} - n =$

$$= \left(\frac{10^2}{33,75} + \frac{10^2}{26,25} + \frac{20^2}{24,37} + \frac{40^2}{26,25} + \frac{70^2}{39,37}\right) + \left(\frac{20^2}{21,37} + \frac{10^2}{16,62} + \frac{15^2}{15,44} + \frac{20^2}{16,62} + \frac{30^2}{24,94}\right) +$$

$$+ \left(\frac{60^2}{34,87} + \frac{50^2}{27,12} + \frac{30^2}{25,19} + \frac{10^2}{27,12} + \frac{5^2}{40,69}\right) - 400 = 143,51$$

Estadístico teórico: $\chi^2_{0,05\,;\,(3-1)\,.\,(5-1)} = \chi^2_{0,05\,;8} = 15,507$

Como $\chi^2_8 = 143,51 > 15,507 = \chi^2_{0,05\,;\,8}$ se rechaza la hipótesis nula de independencia entre la edad del conductor y el número de accidentes.
En consecuencia, la edad influye significativamente en el número de accidentes al año.