EXAMEN-METODOS-JUNIO-RESUELTO.pdf

TejeroBulldog

Metodos Estadisticos para la Computacion

1º Grado en Ingeniería del Software

Escuela Técnica Superior de Ingeniería Informática
Universidad de Málaga

Examen ordinadio de junio RESUELTO

1. (1 punto) Considera las variables estadísticas: X = rentabilidad, en tanto por ciento, de las acciones de Macdonalds e Y = tiempo, en años, que un accionista tarda en vender. Las rectas de regresión de X e Y son $6y + 5x = 7$ y $3y + 2x = 4$, donde la variable x representa a la variable estadística X e y, a Y.

   a) ¿Cuál es la recta de regresión de Y sobre X? ¿Y la de X sobre Y?

   b) Tengo pensado mantener las acciones de Macdonalds durante 3 años. ¿Qué predicción puedes hacer sobre la rentabilidad?

   c) Sabiendo que $\sigma_Y^2 = 5$, calcula la varianza de X, la covarianza y el coeficiente de correlación lineal.

1.a) y/x es la recta con la pendiente más pequeña no el número más pequeño. Por tanto,

si tenemos $y = 7/6 - 5/6\, x$
$y = 4/3 - 2/3\, x$

la pendiente más pequeña es $-2/3$ y por tanto y/x es $3y + 2x = 4$

x/y es $6y + 5x = 7$.

Recuerda: Una recta horizontal tiene pendiente cero por tanto si estoy viendo qué pendiente es más pequeña, será la que se acerque más al cero (independientemente del signo)

1.b) la predicción se realiza en x/y no en las dos por si acaso... Sol $x = -2'2$

1.c)

De x/y que $x = \frac{7}{5} - \frac{6}{5}y$ sabemos que $b = \frac{\text{COV}(x,y)}{\text{VAR}(y)}$; $-\frac{5}{6} = \frac{\text{COV}(x,y)}{5}$; $\boxed{\text{COV}(x,y) = -6}$

De y/x que $y = \frac{4}{3} - \frac{2}{3}x$ sabemos que $b = \frac{\text{COV}(x,y)}{\text{VAR}(x)}$; $-\frac{2}{3} = \frac{\text{COV}(x,y)}{\text{VAR}(x)}$; $\boxed{\text{VAR}(x) = 9}$

$$r^2 = \frac{[\text{COV}(x,y)]^2}{\text{VAR}(x)\cdot\text{VAR}(y)} = \frac{(-6)^2}{9\cdot 5} = 0'8$$

$$r = \frac{\text{COV}(x,y)}{S_x \cdot S_y} = \frac{-6}{\sqrt{9}\cdot\sqrt{5}} = -0'8944$$

Mejor usar la fórmula de r que $r = \sqrt{0'8}$ ya que corremos el riesgo de olvidar el signo de la covarianza!

2. (1 punto) La variable bidimensional (X,Y) toma los siguientes 4 valores $(1,1)$, $(2,4)$, $(3,9)$, $(4,16)$. Calcula el Error Cuadrático Medio (MSE) para estos dos modelos lineales y determina en base a eso cuál de los dos modelos es mejor:

$$y = -6 + 6 * x \qquad y = -5 + 5 * x$$

x_i	y_j
1	1
2	4
3	9
4	16

Modelo 1 $(y = -6+6x)$

x	$f(x) = y_j^{est}$
1	0
2	6
3	12
4	18

Modelo 2 $(y = -5+5x)$

x	$f(x) = y_j^{est}$
1	0
2	5
3	10
4	15.

$\boxed{n=4}$

$$\boxed{MSE = \sum \frac{e_{ij}^2}{n}}$$

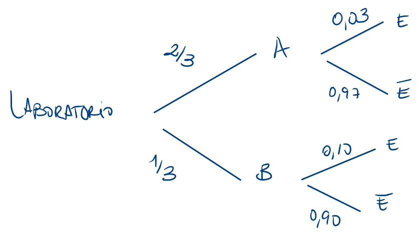
Para Modelo 1 : $MSE = \frac{1}{4}\left[(1-0)^2 + (4-6)^2 + (9-12)^2 + (16-18)^2\right] = \frac{1}{4} \cdot 18 = \underline{9/2}$

Para Modelo 2 : $MSE = \frac{1}{4}\left[(1-0)^2 + (4-5)^2 + (9-10)^2 + (16-15)^2\right] = \frac{1}{4} \cdot 4 = 1$ ← El modelo 2 presenta menos errores

3. (1,5 puntos) Un laboratorio produce dos medicamentos A y B para la misma afección. En el primero de ellos se reportan efectos secundarios en el 3 % de los casos, mientras que esta tasa aumenta a un 10 % para el medicamento B. Si se estima que de cada tres pacientes en tratamiento, dos toman el medicamento A,

a) (0,5 puntos) Calcula la probabilidad de que un paciente en tratamiento, elegido aleatoriamente, reporte efectos secundarios.

b) (1 puntos) Si se eligen aleatoriamente 40 pacientes que toman el mismo medicamento, y nos aseguran que entre ellas hay menos de dos personas con efectos secundarios, ¿cuál es la probabilidad de que estuvieran tomando el medicamento A?.



a) T.P. TOTAL

$P(E) = P(A) \cdot P(E/A) + P(B) \cdot P(E/B)$

$\frac{2}{3} \cdot 0'03 + \frac{1}{3} \cdot 0'1 = \frac{4}{75} = 0'05\hat{3}$

(b es independiente del apartado a)

b) T. Bayes $\quad P(A/E) = \dfrac{P(A) \cdot P(E/A)}{P(E)}$ (Hallar) (T.P. TOTAL) (Hay que hallar datos, antes de aplicar T. Bayes)

· Para A : ¿$P(X<2)$? $\quad X \sim B(40, 0'03)$ Como $n>30$ y $n \times p = 40 \times 0'03 = 1'2 < 5 \Rightarrow$ Paso a Poisson(λ) con $\lambda = 1'2$

$f(x) = \dfrac{e^{-\lambda} \cdot \lambda^x}{x!}$

$P(X<2) = P(X=0) + P(X=1) =$

$= e^{-1'2}\left(\dfrac{1'2^0}{0!} + \dfrac{1'2}{1!}\right) = 0'6626$

· Para B $\quad$ ¿$P(X<2)$? $\quad X \sim B(40, 0'1)$ Como $n>30$ y $np = 40 \cdot 0'1 = 4 < 5$ paso a Poisson(λ) con $\lambda = 4$

$P(X<2) = P(X=0) + P(X=1) = e^{-4}\left(\dfrac{4^0}{0!} + \dfrac{4}{1!}\right) = 0'0915$

Por tanto, $P(E) = P(A) \cdot P(E/A) + P(B) \cdot P(A/E)$ (T.P. TOTAL)

$\frac{2}{3} \cdot 0'6626 + \frac{1}{3} \, 0'0915 = 0'4722\hat{3}$

$P(A/E) = \dfrac{P(A) \cdot P(E/A)}{P(E)} = \dfrac{2/3 \cdot 0'6626}{0'47223} = 0,935413$ $\quad \boxed{93'54\,\%}$

4. (2 puntos) De cierta variable aleatoria se sabe que su mediana es 3/2, y que su función de densidad es la siguiente:

$$f(x) = \begin{cases} 0 & si \quad x \leq 0 \\ x^2 & si \quad 0 < x \leq 1 \\ a + bx & si \quad 1 < x \leq 3 \\ c & si \quad 3 < x \end{cases}$$

a) (1 punto) Determina los valores $a, b, c \in \mathbb{R}$. Determina su esperanza.

b) (1 punto) Realizamos el experimento de elegir aleatoriamente un número real X siguiendo esta distribución de probabilidad. Consideramos la variable aleatoria N = *número de puntos obtenidos* según la siguiente regla: **3 puntos si** $X \leq 1$, **2 puntos si** $1 < X \leq 2$ **y 1 punto si** $X > 2$. Calcula la función de distribución de N, determina su esperanza y su moda.

a) $\boxed{a = 1/3 \qquad b = 0 \qquad c = 0}$ Soluciones. $E(X) = 19/12$

Ecuaciones: 1ª) $f(x) = \int_{-\infty}^{\infty} f(x) = 1$; $f(x) = \cancel{\int_{-\infty}^{0} 0 \, dx} + \int_0^1 x^2 dx + \int_1^3 (a+bx) dx + \cancel{\int_3^{\infty} 0 \, dx} = 1$

2ª) $F(3/2) = 0'5$; $F(x) = \int_{-\infty}^{0} 0 \, dx + \int_0^1 x^2 dx + \int_1^{3/2} (a+bx) dx = 0'5$

Recomendación: Escribir siempre todo para no olvidar esta parte.

$E(X) = \int_{-\infty}^{\infty} x \cdot f(x) \, dx = 19/12$

Entonces $f(x) = \begin{cases} 0 & x \leq 0 \\ x^2 & 0 < x \leq 1 \\ \frac{1}{3} & 1 < x \leq 3 \\ 0 & x > 3 \end{cases}$

b) Para hacer este apartado podemos hacerlo a partir de $f(x)$ o de $F(x)$.

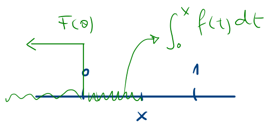
Aunque sea más largo, lo resuelvo hallando $F(x)$ antes para que tengáis más practica.

Hallar $F(x)$:

Si $x<0$

$$F(x)=\int_{-\infty}^{x} f(t)\,dt=\int_{-\infty}^{x} 0\,dt=0.$$

Si $0<x\leq 1$

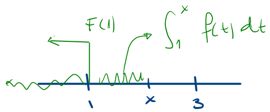


$$F(x)=F(0)+\int_0^x f(t)\,dt=$$

$$=0+\int_0^x t^2\,dt=\left.\frac{t^3}{3}\right]_0^x=\frac{x^3}{3}$$

$F(x)=0$
$F(0)=0$

Si $1<x\leq 3$



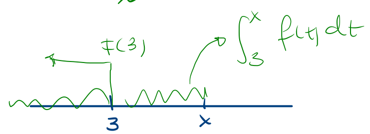
$$F(x)=F(1)+\int_1^x f(t)\,dt=$$

$$\frac{1}{3}+\int_1^x \frac{1}{3}\,dt=\frac{1}{3}+\left.\frac{1t}{3}\right]_1^x=\frac{1}{3}+\frac{x}{3}-\frac{1}{3}$$

$$=\frac{x}{3}$$

$F(x)=\frac{x^3}{3}$
$F(1)=\frac{1}{3}$

Si $x>3$



$$F(x)=F(3)+\int_3^x f(t)\,dt$$

$$F(x)=1+\int_3^x 0\,dt=1$$

$F(x)=\frac{x}{3}$
$F(3)=1$

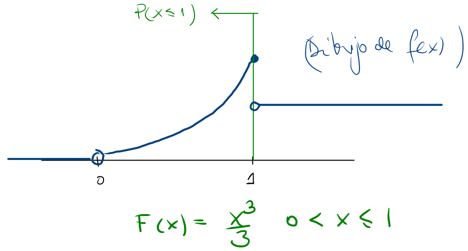
Por tanto,

$$f(x) = \begin{cases} 0 & x \leq 0 \\ x^2 & 0 < x \leq 1 \\ \frac{1}{3} & 1 < x \leq 3 \\ 0 & x > 3 \end{cases} \qquad F(x) = \begin{cases} 0 & x \leq 0 \\ \frac{x^3}{3} & 0 < x \leq 1 \\ \frac{1}{3}x & 1 < x \leq 3 \\ 0 & x \geq 3 \end{cases}$$

Apartado b)

3 puntos si $x \leq 1$

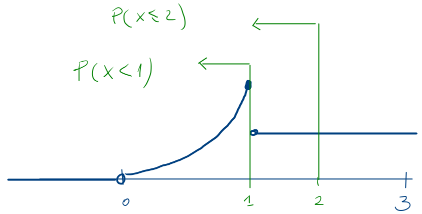
$$P(x \leq 1) = F(1) = \frac{1}{3}$$



2 puntos si $1 < x \leq 2$

$$P(x \leq 2) - P(x < 1) = F(2) - F(1) =$$

$$= \frac{2}{3} - \frac{1}{3} = \frac{1}{3}$$



$$F(x) = \begin{cases} \frac{x^3}{3} & 0 < x \leq 1 \quad \leftarrow (F(1)) \\ \frac{1}{3}x & 1 < x \leq 3 \quad \leftarrow F(2) \end{cases}$$

1 punto si $x > 2$

$$P(x > 2) = 1 - P(x < 2) = 1 - F(2) = 1 - \frac{2}{3} = \frac{1}{3}$$

$\leftarrow$ ($F(2)$ en $F(x) = \frac{1}{3}x \quad 1 < x \leq 3$)

Nota: sin necesidad de hallar $F(x)$ se podría haber hecho:

3 puntos $\int_0^1 f(x)dx = \int_0^1 x^2 dx = 1/3$

2 puntos $\int_1^2 f(x)dx = \int_1^2 1/3 \, dx = 1/3$

1 punto $\int_2^3 f(x)dx = \int_2^3 1/3 \, dx = 1/3$

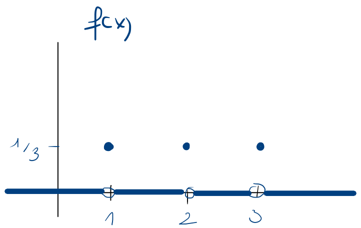
Por tanto, hemos formado 1 variable discreta tal que:

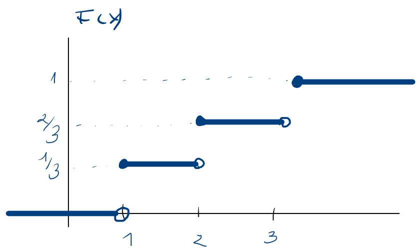
X	1 punto	2 puntos	3 puntos
f(x)	1/3	1/3	1/3
F(x)	1/3	2/3	1

$$E(x) = 3 \cdot \frac{1}{3} + 2 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}$$

Moda $= \{1, 2, 3\}$

Donde sus gráficas son:





5. (1,5 puntos) Una empresa produce dos tipos de balones esféricos (A y B). El radio de los de tipo A se distribuye según una normal de media 20 cm y varianza 16 cm^2, y los de tipo B según una normal de media 16 cm y desviación típica 3 cm. El 70 % de los balones que produce la empresa son del tipo A y el resto de tipo B.

a) (0,5 puntos) Calcular la probabilidad de que el radio de un balón de tipo A mida $3cm$ más que el radio de un balón de tipo B.

b) (0,5 puntos) Calcular la probabilidad de que el radio de un balón escogido al azar de esta empresa mida menos de 22 cm

c) (0,5 puntos) Control de calidad escoge 10 balones de la empresa. ¿Cuál es la probabilidad de que al menos 2 de ellos midan más de 22 cm de radio?

0,7 — A $\quad N(\mu,\sigma) \quad \mu_A = 20 \quad \sigma_A^2 = \quad \sigma_A =$

0,3 — B $\quad N(\mu,\sigma) \quad \mu_B = 15 \quad \sigma_B^2 = \quad \sigma_B =$

a) $P\left[(\underset{A}{x_1} - \underset{B}{x_2}) > 3\right] = P\left(Z > \dfrac{3-(\mu_1-\mu_2)}{\sqrt{\sigma_1^2+\sigma_2^2}}\right) = P\left(Z > \dfrac{3-(20-16)}{\sqrt{16+9}}\right) = P(Z > -0'2) = P(Z < 0'2) =$

$= 1 - P(Z > 0'2) = 1 - 0'4207 = 0'5793$

b) X ~ radio $\quad P(X<22)$?

T.P. TOTAL

$P(X<22) = P(A)\cdot P(X<22/A) + P(B)\cdot P(X<22/B)$

• Para $P(X<22/A) = P\left(Z_A < \dfrac{22-20}{4}\right) = P(Z_A < 0'5) = 1 - P(Z_A > 0'5) = 1 - 0'3085 = 0'6915$

• Para $P(X<22/B) = P\left(Z_B < \dfrac{22-16}{3}\right) = P(Z_B < 2) = 1 - P(Z_B > 2) = 1 - 0'0228 = 0'9772$

Por tanto, $P(X<22) = 0{,}7 \cdot 0'6915 + 0'3 \cdot 0'9772 = \boxed{0'77721}$ $\quad$ 77'72 %

c) X ~ radio balones ; $\quad$ ¿$P(X \geq 2)$?

$X \sim B(n,p)$ donde p es $P(X>22) = 1 - P(X<22) = 0'22279$

$X \to B(10, 0'22279)$

$P(X\geq 2) = 1 - P(X<2) = 1 - \left[P(X=0) + P(X=1)\right] = 1 - \left[\binom{10}{0} 0'22279^0 (1-0'22279)^{10} + \binom{10}{1} 0'22279 (1-0'22279)^9\right] = 0'6890$

68'90 %

AMPLIO EL ENUNCIADO: !

E.) Sabiendo que un balón mide menos de 22 cm ¿ De qué tipo es más probable que sea?

T. Bayes (2 veces)

$$P(A/x<22) = \frac{P(A)\cdot P(x<22/A)}{P(x<22)} = \frac{0'7 \cdot 0'6564}{0,75581} = \boxed{0'6079}$$

$$P(B/x<22) = \frac{P(B)\cdot P(x<22/B)}{P(x<22)} = \frac{0'3 \cdot 0'9901}{0,75581} = \boxed{0,3929}$$

Es más probable que sea del tipo A.

f.) Calcular la probabilidad de que el radio de un balón de tipo B mida 3 cm más que el radio de un balón tipo A (pregunta lo mismo que en el aparado a pero al revés!)

$$a)\ P[(x_2-x_1)>3] = P\left(Z > \frac{3-(\mu_2-\mu_1)}{\sqrt{\sigma_1^2+\sigma_2^2}}\right) = P\left(Z > \frac{3-(16-20)}{\sqrt{16+9}}\right) = P(Z>1'4) = 0'0808$$

! Nota, no importa que ese valor sea negativo (Es lo que marca el enunciado "(x_2-x_1)")

g.) Calcular la probabilidad de que la diferencia de la variable radios sea mayor que 3

Tenemos que considerar que $P(x_1-x_2)>3$ ó que $P(x_2-x_1)>3$.

Por tanto tenemos que calcular $[P(x_1-x_2) \cup (P(x_2-x_1)]>3$.

observemos que estas probabilidades están ya calculadas en el apartado a) y f) por tanto, sólo debemos sumarlas.

$$0'5793 + 0'0808 = 0'6601 \qquad 66'01\ \%$$

h.) Se hace el envio de 80 unidades de balones esféricos solo y exclusivamente de tipo A. Calcula la probabilidad de que la media de la variable radios sea menor de 19cm

Datos $\quad x \rightarrow$ radio

$X \sim N(20,4) \qquad \mu_A = 20 \qquad \sigma_A^2 = 16 \qquad \sigma_A = 4.$

Observación: $80 > 30$ (muestra grande) $\rightarrow$ T. C. Límite $\quad \bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$

¿ $P(\bar{x}<19)$? (Debemos tipificar teniendo en cuenta el T. C. Límite).

$$P(\bar{x}<19) = P\left(Z < \frac{19-20}{4/\sqrt{80}}\right) = P(Z<-2'23) = P(Z>2'23) = 0'0129 \qquad 1'2\%$$

6. (1,5 puntos) Se han medido la altura de 50 chimpancés de una especie A y se ha obtenido una media de $100cm$ y una varianza de $98cm^2$ y la altura de 55 chimpancés de una especie B y se ha obtenido una media de $103cm$ y una varianza de $118.8cm^2$.

a) (1 punto) Hallar un intervalo de confianza al 95 % para la diferencia de medias entre las dos alturas, bajo la hipótesis de normalidad de los datos.

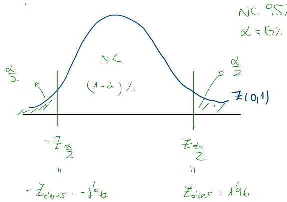
b) (0,5 puntos) A la vista de lo anterior, ¿podríamos afirmar con esa confianza que la media de altura de la especie B es mayor que la media de altura de la especie A?

Intervalo de confianza para la diferencia de medias $(\mu_1 - \mu_2)$ de dos distribuciones normales $N(\mu_1, \sigma_1)$ y $N(\mu_2, \sigma_2)$

Población

Varianzas	Muestras	Varianzas	Intervalo
Conocidas			~~$I = \left[(\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right]$~~ NO
	grandes $n_1+n_2>30$ $n_1 \simeq n_2$		$I = \left[(\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}\right]$
Desconocidas	Pequeñas $n_1+n_2 \leq 30$	Iguales	$I = \left[(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2, n_1+n_2-2} \cdot s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right]$
		Distintas	$I = \left[(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2, f}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}\right]$

NC 95%
$\alpha = 5\%$



$X \rightsquigarrow$ altura chimpancés

Población 1		Población 2
$X \rightsquigarrow N(\mu_1, \sigma_1)$		$X \rightsquigarrow N(\mu_2, \sigma_2)$
Muestra A	$\leftarrow n_{TOTAL} = 105 > 30 \rightarrow$	Muestra B
$n_A = 50$		$n_B = 55$
$\bar{x}_A = 100$		$\bar{x}_B = 103$
$S_A^2 = 98$		$S_B^2 = 118'8$
$\boxed{\hat{S}_A^2 = 100 \text{ Cuasivarianza!}}$		$\boxed{\hat{S}_B^2 = 121 \text{ Cuasivarianza!}}$

Recuerda $\hat{S}^2 = \frac{n}{n-1} S^2$

$\hat{S}^2 = \frac{50}{49} \cdot 98 = 100.$

a) ¿$IC_{95\%}$ para la diferencia de medias?

Podemos adoptar $IC_{95\%}$ para $\mu_1 - \mu_2$ ó $IC_{95\%}$ para $\mu_2 - \mu_1$ ya que el enunciado no especifica. Realizaremos el ejercicio de esta forma (Habrá que tener en cuenta que las interpretaciones también cambian)

RECUERDA:

Si IC para $\mu_1 - \mu_2$ ó $p_1 - p_2$

Si $[\ 0\ ]$ $\mu_1 = \mu_2$

Si $0\ [\quad]$ $\mu_1 > \mu_2$

Si $[\quad]\ 0$ $\mu_1 < \mu_2$

Si IC para $\mu_2 - \mu_1$ ó $p_2 - p_1$ entonces

Si $[\ 0\ ]$ $\mu_1 = \mu_2$

Si $0\ [\quad]$ $\mu_1 < \mu_2$

Si $[\quad]\ 0$ $\mu_1 > \mu_2$

$$IC_{95\%} \text{ para } \mu_2 - \mu_1 = \left[(\bar{x}_2 - \bar{x}_1) \pm Z_{\frac{\alpha}{2}}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}\right]$$

$$= \left[(103 - 100) \pm 1'96\sqrt{\frac{100}{50} + \frac{121}{55}}\right] = \left[-1'01 \ , \ 7'01\right]$$

b) Interpretación: como el cero está dentro del intervalo podemos afirmar con nivel de confianza del 95% que no existen diferencias significativas entre las medias poblacionales.

7. (1,5 puntos) Una empresa de estudios de mercado lleva a cabo un sondeo en Madrid y Barcelona. Se pregunta a residentes de ambas ciudades por su marca de smartphone favorito, entre 3 conocidas marcas: Huawei, Samsung y Xiaomi. Este es el resultado:

*Ciudad**Marca*	Huawei	Samsung	Xiaomi
Madrid	150	210	110
Barcelona	200	200	130

Proporciona una de estas dos informaciones (**sólo una de ellas, la que prefieras**):

a) Contrasta con un nivel de significación del 2,5 % si la elección de marca preferida de smartphone está relacionada con la ciudad en la que residen los encuestados.

$X \sim$ ciudad

$Y \sim$ marca smartphone

Contraste:

H_0: X e Y son variables independientes

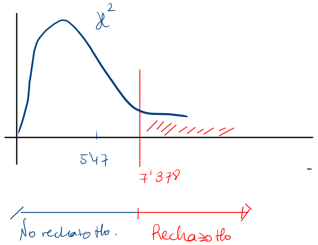
H_1: X e Y no son variables independientes.

Tabla de frecuencias esperadas en caso de que H_0 sea cierta:

164'5	192'7	112'8
185'5	217'3	127'2

- Estadístico de prueba $\hat{\chi}^2_0 = \dfrac{\sum (o_i - e_i)^2}{e_i}$; $\hat{\chi}^2_0 = \dfrac{(150-164'5)^2}{164'5} + \dots + \dfrac{(130-127'2)^2}{127'2} = 5,47$

- Valor crítico $\chi^2_{\alpha,(k-1)(m-1)} = \chi^2_{0'025,2} = 7'378$



$\hat{\chi}^2_0 < \chi^2_{0'025,2}$ por tanto no rechazo H_0. Asumo que las variables ciudad y marca tienen independencia con $\alpha = 2'5\%$.