Procedimientos estadística

Antes de cualquier ejercicio, $X \equiv$ Definición de la variable para todos los casos y en todas las modificaciones

Tema 1 Estadística descriptiva

Correlaciones e interpretación

$$R^2 = 0{,}89 \rightarrow r = -\sqrt{R^2} = -0{,}9434 \left\| \begin{array}{l} \text{si } r > 0 \rightarrow s_{xy} > 0 \rightarrow \text{Dependencia positiva} \\ \text{si } r < 0 \rightarrow s_{xy} < 0 \rightarrow \text{Dependencia negativa} \end{array} \right.$$

Si la dependencia entre X e Y es negativa

Covarianza de X e Y dada la varianza de una de las variables y R^2

$$R^2 = 0{,}89 = \frac{{s_{xy}}^2}{{s_x}^2 \cdot {s_y}^2} \rightarrow \frac{{s_{xy}}^2}{{s_x}^2 \cdot 0{,}0253} \rightarrow \left. \begin{array}{r} \dfrac{{s_{xy}}^2}{{s_x}^2 \cdot 0{,}0253} = 0{,}89 \cdot 0{,}0253 \\ \hat{a} = \dfrac{{s_{xy}}^2}{{s_x}^2} = -0{,}07 \end{array} \right\| 0{,}022517 =$$

$$= \frac{s_{xy}}{{s_x}^2} \cdot s_{xy} \rightarrow s_{xy} = -\frac{0{,}022517}{0{,}07} = -0{,}3217$$

Media y desviación típica

$$\sum x_i = 537 \rightarrow \bar{x} = \frac{537}{n} \rightarrow \frac{537}{6} = 89{,}5;\ \sum {x_i}^2 = 48107 \text{ (cada dato al cuadrado)}$$

$${s_x}^2 = \frac{n}{n-1}\left(\frac{\overline{x^2}}{n} - \bar{x}^2\right) = 9{,}1$$

$$S_x = 3{,}0166 \text{ (Desviación típica)}$$

Para la combinación de X e Y (también llamado coeficiente de correlación)

$$s_{xy} = \frac{n}{n-1}(\overline{xy} - \overline{x \cdot y})$$

Dispersión relativa

$$C_{v_x} = \frac{s_x}{\bar{x}} = \frac{3{,}0166}{89{,}5} = 0{,}0337\ ;\ C_{v_y} = \frac{s_y}{\bar{y}} = \frac{3{,}0166}{92{,}5} = 0{,}0326 \| \rightarrow C_{v_x} = \text{mayor dispersión relativa}$$

Sustitución de dato al azar por atípico

- La media disminuye al disminuir la cuantía de una observación pero la varianza aumentará pues aumenta la dispersión de datos.
- La mediana no varía ya que en un cálculo influye la posición de los datos y no su cuantía. La varianza aumentará.
- El rango intercuartílico permanece cte pues el hecho de que cambie el valor mínimo, no influye en el cálculo de Q_1 y Q_3

Determinar la a(pendiente de la recta) dadas las varianzas de X e Y $({s_x}^2, {s_y}^2)$

$$Y = aX + 42{,}92 \qquad r = -0{,}9$$

$$r = \frac{s_{xy}}{s_x \cdot s_y} \rightarrow \frac{s_{xy}}{\sqrt{3{,}91} \cdot \sqrt{2{,}23}} \rightarrow s_{xy} = -0{,}9 \cdot \sqrt{3{,}91} \cdot \sqrt{2{,}23} = -2{,}6577$$

$$\hat{a} = \frac{s_{xy}}{{s_x}^2} = \frac{-2{,}6577}{3{,}91} = -0{,}6797$$

Determinar la b (ordenada en el origen)

$$\bar{x} = 417{,}7, \bar{y} = 1160{,}6\ ; y = 2{,}52 \cdot x + b \rightarrow \hat{b} = \bar{y} - \hat{a} \cdot \bar{x} \rightarrow 1160{,}6 - 2{,}52 \cdot 417{,}7 = 107{,}996$$

Coeficiente de determinación y comentarios de bondad del ajuste

$$R^2 = \frac{{s_{xy}}^2}{{s_x}^2 \cdot {s_y}^2} \rightarrow s_{xy} = \hat{a} \cdot {s_x}^2 \rightarrow \begin{array}{c} 0{,}9 \text{ ajuste bueno} \\ < 0{,}9\ \textit{ajuste muy bueno} \end{array}$$

Translaciones (ej) $Z = X + 5$

$$\bar{Z} = \bar{x} + 5;\ {s_z}^2 = {s_x}^2; s_{xy} = \frac{1}{30}\sum_{i=1}^{30}(z_i - \bar{z})(y_i - \bar{y}) = \frac{1}{30}\sum_{i=1}^{30}\big(x_i + 5 - (\bar{x} + 5)\big)(y_i - \bar{y}) \rightarrow$$

$$\rightarrow y = c + dz \rightarrow \begin{array}{c} \hat{d} = \dfrac{s_{yz}}{{s_z}^2} = 2{,}52 \\ \hat{c} = \bar{y} - \hat{d} \cdot \bar{z} = 1160{,}6 - 2{,}52 \cdot 422{,}7 \end{array} \rightarrow y = 2{,}52 \cdot z + 95{,}396$$

Datos agrupados (% mayor de un valor) ej

$$y > 85 = \frac{10 + 5 + 3}{30} \cdot 100 = 60$$

Tema 2 Introducción a la teoría de la probabilidad

Definiciones previas de los datos

$$p(\bar{A}) = 1 - p(A);\ p(\bar{A} \cup \bar{B}) = p(\overline{A \cap B}) = 1 - p(A \cap B);\ \ p(A/B) = \frac{p(A \cap B)}{p(B)}$$

En algunos casos los parámetros vendrán dados directamente, pero no es lo frecuente, entonces debemos sacar el %, se tendrá que hacer una relación entre los 3-4 datos que se aporten (fincas, fabricas, envasadoras etc,) Sería del tipo:

$$\left. \begin{array}{l} P_A = \text{Procedimiento A} = x \rightarrow p(P_A) = 0{,}2 \\ P_B = \text{Procedimiento B} = x \rightarrow p(P_B) = 0{,}2 \\ P_C = \text{Procedimiento C} = x \rightarrow p(P_C) = 0{,}6 \end{array} \right\} 5x = 1 \rightarrow x = \frac{1}{5}$$

A continuación se define la variable (recolección, mal estado, aparición parásito, etc), es importante identificar qué significa, ejemplos:

Encabezado de ejercicios de probabilidad

$$Q \equiv \text{recolección, mal estado, aparición parásito}$$

Probabilidad de que en la variedad P_1 aparezca un determinado parásito

$$p(Q/P_1)$$

Probabilidad de que en la variedad P_2 no aparezca un determinado parásito

$$p(\bar{Q}/P_2) = 1 - p(Q/P_2)$$

Producción variedad P_3 y aparece el parásito

$$p(Q \cap P_3) \rightarrow \frac{p(Q \cap P_3)}{p(P_3)} = p(Q/P_3)$$

Producción variedad P_3 y NO aparece el parásito

$$p(\bar{Q} \cap P_3) \rightarrow \frac{p(\bar{Q} \cap P_3)}{p(P_3)} = p(\bar{Q}/P_3) \rightarrow 1 - p(Q/P_3)$$

Probabilidad de tomar uno al azar y que se dé el suceso, si fuera de que NO se dé el suceso, sería

$$P(\bar{D}) = 1 - P(D)$$

$$P(D) = \sum_{i=1}^{3} p(L_i) \cdot p(D/L_i)$$

Que se dé el suceso (si no se diera el suceso donde está la R sería $\bar{R}$) y determinar con que probabilidad se daría dependiendo de los casos posibles

$$p(F_n/R) = \frac{p(F_n) \cdot p(R/F_n)}{p(R)}; p(F_n/\bar{R}) = \frac{p(F_n) \cdot p(\bar{R}/F_n)}{p(\bar{R})} = \frac{p(F_n) \cdot \big(1 - p(R/F_n)\big)}{1 - p(R)}$$

Si elegido uno tiene el suceso y no es de P_2 (**Si se da el suceso y es de P_2 sería sin rectarle 1**)

$$p(\bar{P_2}/A) = 1 - p(P_2/A) = 1 - \frac{p(P_2 \cap A)}{p(A)} = 1 - \frac{p(P_2) \cdot p(A/P_2)}{p(A)}$$

Si es elegido uno y no tiene el suceso y no es de P_2 Aquí se aplica la ley de Morgan

$$p(\bar{A} \cap \bar{P_2}) = p(\overline{A \cup P_2}) = 1 - p(A \cup P_2) = 1 - [p(A) + p(P_2) - p(P_2) \cdot p(A/P_2)]$$

Si se da el suceso y puede ser de P_2 ó P_3 (si no se da el suceso $\bar{R}$)

$$p(P_2 \cup P_2/R) = \frac{P[(P_2 \cup P_3 \cap R)]}{p(R)} = \frac{p[p(P_2 \cap R) \cup p(P_3 \cap R)]}{p(R)} \rightarrow$$

$$\xrightarrow{\text{incompatible}} \frac{p(P_2 \cap R) + p(P_3 \cap R)}{p(R)} = \frac{p(P_2) \cdot p(P_2/R) + p(P_3) \cdot p(P_3/R)}{p(R)}$$

Si es de P_3 y no se da el suceso

$$p(P_3/\bar{R}) = \frac{p(\bar{R} \cap P_3)}{p(\bar{R})} = \frac{p(P_3) \cdot p(\bar{R}/P_3)}{p(\bar{R})} \rightarrow \frac{p(P_3) \cdot \big(1 - p(R/P_3)\big)}{1 - p(R)}$$

Que no se dé el suceso y que no sea ni de A ni de C (3 fabricas)

$$p(F_A \cap F_C/\bar{D}) = 1 - p(F_B/\bar{D}) = \frac{p(F_B) \cdot p(\bar{D}/F_B)}{p(\bar{D})} = \frac{p(F_B) \cdot p\big(1 - p(D/F_B)\big)}{1 - p(D)}$$

Se dé el suceso y no sea por F_3

$$p(R \cap \overline{F_3}) = p[R \cap (F_1 \cup F_2)] = p[(R \cap F_1) \cup (R \cap F_2)] \xrightarrow{\text{Incompatibles}}$$

$$\rightarrow p(R \cap F_1) + p(R \cap F_2) = p(F_1) \cdot p(R/F_1) + p(F_2) \cdot p(R/F_2)$$

Para pasar de Cont. a Discreta
A lo sumo $\leq +0{,}5$
Al menos $\geq -0{,}5$
Estricto mayor +0,5
Estricto menor -0,5

Tema 3-4-5 Variables aleatorias-Vectores aleatorios-Muestreo y distribuciones muéstrales

Distribución normal

<u>Definición de la variable</u>

$$X \equiv Definición\ de\ la\ variable \sim N(\mu,\sigma) \rightarrow$$
$$\rightarrow \begin{matrix} Las\ \theta(valor)\ se\ miran\ en\ la\ tabla\ normal \rightarrow x\ para\ Pr(X<x) \\ las\ Z_{valor}\ se\ miran\ en\ la\ tabla\ normal \rightarrow Pr(X<x)\ para\ x \end{matrix}$$

<u>Probabilidad dada $\mu = 255, \sigma = 10; antes\ de\ 244$</u>

$$p(x<244) = \theta\left(\frac{224-255}{10}\right) = \theta(-1{,}1) = 1-\theta(1{,}1)$$

<u>Si dan un suceso con "n" y se ha calculado previamente la probabilidad del suceso y nos darán "al menos un porcentaje (20%)"</u>, es te tipo se **modeliza como una Binomial**
Se define la nueva variable $\mathrm{Y} \equiv$ nº de lo que sea que cumple la probabilidad$\sim \mathrm{n} = 100, \mathrm{p} = 0{,}1357$

$$\begin{cases} \mathrm{np} = 100 \cdot 0{,}1357 = 13{,}57 \sim \mu \text{ esperanza} \\ \mathrm{np}\cdot(1-\mathrm{p}) = 100\cdot 0{,}1357\cdot(1-0{,}1357) = 11{,}72 \sim \sigma^2 \rightarrow \sigma = 3{,}42 \text{ varianza} \end{cases}$$

Se hace cambio de discreta a continua

$$p(Y \geq 20) \sim p(W \geq 19{,}5)$$

Y se resuelve como una normal, con los cambios establecidos

$$p(W \geq 19{,}5) = 1-\theta\left(\frac{19{,}5-13{,}57}{3{,}42}\right) = 1-\theta(1{,}73) = 1-0{,}9582$$

<u>Si se reduce la media, dada una probabilidad, cambio en la desviación típica</u>

$$\mathrm{T} \equiv \text{Nueva definición de variable} \sim N(\mu = 255-3, \sigma = ?) \rightarrow p(T<244) = 0{,}18 \rightarrow$$
$$\rightarrow 0{,}18 = \theta\left(\frac{244-252}{\sigma}\right) = \theta\left(\frac{-8}{\sigma}\right) = 1-\theta\left(\frac{8}{\sigma}\right) \rightarrow 1-0{,}18 = \theta\left(\frac{8}{\sigma}\right) \rightarrow Z_{0{,}82} = 0{,}92 \rightarrow \sigma = \frac{8}{0{,}92} = 8{,}7$$

Distribución de Poisson

Generalmente los sucesos

$$X \equiv \text{Definición de variable} \sim P_0(\lambda = \text{valor}) \rightarrow f_x(x) = e^{-\lambda}\cdot\frac{\lambda^x}{x!}, x = 0,1,2,3,4,5 \ldots$$

<u>A lo sumo uno</u>

<u>Calcular λ sabiendo la probabilidad</u>
Si la repetición del suceso es que no se dé ej *No encontrar erratas*

$$p(x=0) \rightarrow f(0) = e^{-\lambda}\cdot\frac{\lambda^0}{0!} = 0'6065 \rightarrow e^{-\lambda} = 0{,}6065 \rightarrow \ln 0{,}6065 = -\lambda \rightarrow \lambda = 0{,}5$$

$$p(x \leq 1) = e^{-0{,}5}\cdot\left(\frac{0{,}5^0}{0!}+\frac{0{,}5^1}{1!}\right) = 0{,}9098$$

<u>Probabilidad de exactamente un valor</u>
*Probabilidad de que en dos días un estudiante realice exactamente 7 visitas sabiendo λ = 5

$$p_0 = (\lambda = 5\cdot 2 = 10); p(y=7) \rightarrow f_y(y) = e^{-10}\cdot\frac{10^7}{7!}$$

<u>Cambios en la frecuencia de λ, ej. dada una semana en el encabezado, para 1 mes</u>

$$Y \equiv \text{Redefinir la variable para tiemp nuevo} \rightarrow X_1 + \cdots X_n \rightarrow P_0(\lambda = 4\cdot 0{,}5 = 2)$$

<u>Probabilidad de intervalo</u>

$$P(4 \leq x \leq 7) = e^{-\lambda}\cdot\frac{\lambda^y}{y!} \rightarrow e^{-2}\cdot\left(\frac{2^4}{4!}+\frac{2^5}{5!}+\frac{2^6}{6!}+\frac{2^7}{7!}\right)$$

Distribución Binomial

Suele ser un apartado de las demás distribuciones, cuando se dan "n" sucesos y probabilidad calculada para un tipo de suceso en apartados anteriores

Distribución exponencial (generalmente circuitos)

Para todos los casos el encabezado

$$\lambda = \frac{1}{\text{media dada}}\begin{cases} \text{Función de densidad } f_x(x) = \begin{matrix} 0, & x \leq 0 \\ \lambda\cdot e^{\lambda x}, & x>0 \end{matrix} \\ \text{Función de distribución } F_x(x) = \begin{matrix} 0, & x \leq 0 \\ 1-e^{\lambda x}, & x>0 \end{matrix} \end{cases} \rightarrow \text{Utilizamos la } F_x(x)$$

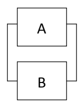
Dados dos avisos consecutivos es al menos de 10 horas, determinar la probabilidad de que no se reciba aviso alguno en las siguientes 8 horas

$F_T(t) = 1-e^{-\frac{t}{8}}, t>0$ (siempre el mayor izquierda, menor derecha)

$$P(T \geq 18/T \geq 10) = \frac{p(T \geq 18)}{p(T \geq 10)} = \frac{1-e^{-\frac{18}{8}}}{1-e^{-\frac{10}{8}}} = e^{-\frac{18}{8}+\frac{10}{8}} = e^{-1} = 0{,}3679$$

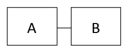
Determinar que el dispositivo funcione al menos un nº dado (200 horas), se resuelve todo el sistema, mirar cajetines derecha

Paralelo



$$p[(A \geq 200) \cup (B \geq 200)] \rightarrow$$
$$\rightarrow p(A \geq 200) + p(B \geq 200) - p(B \geq 200)\cdot p(A \geq 200)$$

Serie



$$p[(A \geq 200) \cap (B \geq 200)] \rightarrow$$
$$\rightarrow p(A \geq 200)\cdot p(B \geq 200)$$

Distribución sin conocimiento específico (Desigualdad Tchebychev)

Funciones

Función de distribución F(x) y buscar un valor indeterminado en las ecuaciones, se sustituye en la ecuación con límites

Función de densidad	Función de probabilidad	Función de distribución
Representación $f(x)$ estas llevan x en los intervalos	Representación $f(x)$ estas no llevan x	Representación $F(x)$
Para pasar a Distribución		Pasar a la función de densidad derivar
integrar	Sumatorios $\sum x_i \cdot f_i$	

Si da función de densidad y hallar la k para una función de densidad

$$f(x) \rightarrow 1 = \int_{\mathbb{R}} f(x)\,dx = k\int_0^1 (x^2-2x+1)dx \rightarrow k\left[\frac{x^3}{3}\bigg|_0^1 + x^2|_0^1 + x|_0^1\right] \rightarrow \left(\left(\frac{1^3}{3}-\frac{0^3}{3}\right)+(1^2-0^1)+(1-0)\right) \rightarrow 1 = k\cdot\frac{1}{3} \rightarrow \boxed{k=3}$$

$$F(x) = 3\left(\frac{x^3}{3}-x^2+x\right) \rightarrow x^3-3x^2+3x$$

Tema 6-7 Introducción a la teoría de la estimación-Introducción a los contrastes de hipótesis

Encabezado

<u>Tenemos que definir cuantos nºs de muestra hay, la $\bar{x} = será\ nuestra\ \mu$</u>

<u>Construir un intervalo con confianza conocida e indicar los resultados teóricos</u>
Resultados teóricos debemos siempre poner lo mismo
IC al (Intervalo de confianza dado) para μ tomamos $\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \rightarrow Z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$
*Estimación de σ si no es conocida

$$S \equiv \text{desviación típica muestral } \rightarrow S = \sqrt{\frac{1}{n-1}\cdot\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

Aplicamos el teorema de Fisher teniendo que la variable aleatoria

$$T \equiv \frac{\bar{x}-\mu}{S/\sqrt{n}} \sim t_n - 1$$

<u>Determinar el nº de la muestra dada una media, para un error y una confianza dada</u>

$$\text{Confianza} = p\left(\overbrace{-t_{n-1};\frac{\alpha}{2}}^{\text{t de student}} \leq \frac{\bar{x}-\mu}{S/\sqrt{n}} \leq t_{n-1};\frac{\alpha}{2}\right) \xrightarrow{\text{despejando } \mu}$$
$$\rightarrow p\left(\bar{x}-t_{n-1};\frac{\alpha}{2}\cdot\frac{S}{\sqrt{n}} \leq \mu \leq \bar{x}+t_{n-1};\frac{\alpha}{2}\cdot\frac{S}{\sqrt{n}}\right) \rightarrow$$
$$\rightarrow \bar{x} \pm t_{n-1};\frac{\alpha}{2}\cdot\frac{S}{\sqrt{n}} \rightarrow \text{sustituimos y dada } \bar{x} \pm \text{valor} \rightarrow (\text{valor}, \text{valor})$$

<u>Contraste de hipótesis significativamente mayor, o significativamente menor, o hipótesis nula y calcular el p-valor</u>

$$\begin{matrix} tipos\ contrastes\ de\ hipótesis \\ \begin{matrix} H_0\ \mu = valor \\ H_1\ \mu > valor \end{matrix}; \begin{matrix} H_0\ \mu = valor \\ H_1\ \mu < valor \end{matrix}; \begin{matrix} H_0\ \mu = valor \\ H_1\ \mu \neq valor \end{matrix} \rightarrow \end{matrix} \quad \begin{matrix} para > ó <\ \alpha = 1 - confianza \\ para\ hipótesis\ nula\ \frac{\alpha}{2} = 1 - confianza \end{matrix}$$

hacemos
$H_0 = \frac{\bar{x}-\mu}{S/\sqrt{n}}$, se compara con el valor de t de student para k grados de libertad y confianza dada y se dice si se puede rechazar o no el resultado
Dar intervalo de p-valor con las dos t de student más cercanas, y establecer un intervalo
<u>Determinar tamaño de la muestra</u>

$$\text{sup } \bar{x} \sim N|\mu, \sigma = \text{valor}| \rightarrow ¿\, n/Z_{1-\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}} < (\text{error})\text{al}(\text{confianza}) \rightarrow$$
$$\rightarrow n > \left(\frac{\overbrace{Z_{1-\frac{\alpha}{2}}}^{\text{valor más alto de grados de libertad}}\cdot\sigma}{\text{error}}\right) = \text{necesitamos al menos (valor que dé)}$$