Universidade Vigo

Departamento de Estatística e Investigación Operativa

E.Superior de Enxeñería Informática
Edificio Politécnico
Campus de Ourense
E-32004 Ourense

Tel. 986 387 000
http://esei.uvigo.es

Estatística

Apelidos:	Nome:	DNI:

1. (*3 puntos*) El archivo adjunto *datos.metais.txt*[1], contiene datos de contaminación de diferentes metales pesados (en ppm) en diferentes localizaciones geográficas en formato UTM (columnas X,Y). Asociado a cada localización aparece la denominación del lugar en la recogida del dato. Para este conjunto de datos:

   a) Clasificar estadísticamente las variables *Zona, X, Y, Metal1, Metal2*.

   b) Para la v. Metal1, los valores mayores de 200 son producto de un error de cálculo y deben ser divididos por 10. Dar la distribución completa de frecuencias agrupando la distribución en intervalos con puntos de corte (0,10,20,30,40,50,200).

   c) Calcular la media muestral y la mediana con la variable agrupada y sin agrupar.

   d) Dar un resumen numérico completo de la variable Metal1 para la zona *Llodero*. Comenta los resultados obtenidos

   e) Compara con un diagrama de cajas los valores de la v. Metal1 en las zonas *Llodero, ALV, Praia*. Extrae conclusiones.

2. *(2 puntos)* Consideremos un sistema electrónico que consta de diez componentes que funcionan independientemente teniendo cada uno una probabilidad de fallo de 0.05.

   a) Calcula la fiabilidad del sistema (probabilidad de que el sistema funcione correctamente). El sistema funciona correctamente si funcionan todas sus componentes.

   Solución: Sea X='Número de componentes del sistema que no funcionan'$\sim Bi\,(10, p = 0.05)$.

   $$P\,(X = 0) = \binom{10}{0} p^0\,(1-p)^{10} = 0.95^{10} = 0.5987369 \tag{1}$$

   □

   b) Si para aumentar la fiabilidad del sistema, se conectan en paralelo dos sistemas iguales al descrito, calcula la fiabilidad del nuevo sistema.

   Solución: Sean A_1, A_2 la representación de que funciona el 1er y 2º sistema, respectivamente. La probabilidad pedida es

   $$P\,(A_1 \cup A_2) = P\,(A_1) + P\,(A_2) - P\,(A_1 \cap A_2) = 0.5987369 + 0.5987369 - 0.5987369 * 0.5987369 = 0.8389879 \tag{2}$$

   □

3. *(2.5 puntos)* El lenguaje de programación *gfortran* dispone de una rutina *rand()* que devuelve números pseudoaleatorios entre 0 y 1, representando una v.a. continua con función de densidad:

   $$f(x) = \begin{cases} 1 \text{ cuando } 0 < x < 1 \\ 0 \text{ en otro caso} \end{cases}$$

   a) Calcular la $P(X > 0.75 \cup X \leq 0.25)$.

   Solución: punif(0.75,min=0,max=1) - punif(0.25,min=0,max=1)=0.5 □

   b) Se desea obtener números enteros de 1 al 10 y se define para ello la siguiente v.a. $Y = Parte\_Entera(10 * X) + 1$. Dar los posibles valores de la variable Y y la probabilidad de cada uno de ellos.

[1] Descargar desde la url *https://dl.dropboxusercontent.com/u/29008031/datos.metais.txt*

E.Superior de Enxeñería Informática
Edificio Politécnico
Campus de Ourense
E-32004 Ourense

Tel. 986 387 000
http://esei.uvigo.es

Solución: La variable Y toma valores enteros (por definición) desde 1 hasta 10. Sus probabilidades son:

$$P(Y = 1) = P(0 < X < 1) = \frac{1}{10},$$

$$P(Y = i) = P(i - 1 \leq X < i) = \frac{1}{10}, \text{para } i = 2, \ldots, 10$$

$\square$

4. *(2.5 puntos)* Desde un laboratorio con 20 terminales se imprime en una misma impresora. La tasa de envío de documentos a imprimir desde un terminal es una $Poisson(\lambda = 3)$ por unidad de tiempo. Suponiendo independencia, responder razonadamente:

a) ¿Cuál es la probabilidad de que un terminal de cualquiera de los 20 envíe más de 2 trabajos?

Solución: Sea X_1=*nº de documentos enviados desde un terminal*$\sim Pois(\lambda = 3)$
$P(X_1 > 2) = 1 - ppois(2, 3) = 0.5768099$
Sea Y=*n de terminales que envían más de 2 trabajos* $\sim Bi(20, 0.5768099)$
$P(Y = 1) = dbinom(1, size = 20, prob = 0.5768099) = \binom{20}{1} 0.5768099^1 \, (1 - 0.5768099)^{19} = 9.252033e - 07$

$\square$

b) ¿Y si fuesen 100 terminales? Responder usando y sin usar la aproximación normal con la corrección de continuidad.

Solución: Sea Y=*nº de terminales de 100 que envían más de 2 trabajos*$\sim Binomial(n = 100, p = 0.5768099)$
P(Y=1)= dbinom(1,100,0.5768099) = 6.138286e-36
Ahora $Y \sim Binomial(n = 100, p = 0.5768099) \approx N(100 * 0.5768099, \sqrt{100 * 0.5768099 * (1 - 0.5768099)}$, por lo tanto:
$P(Y = 1) \approx P(1 - 0.5 \leq \tilde{Y} < 1 + 0.5) = pnorm(1.5, 100 * 0.5768099, sqrt(100 * 0.5768099 * (1 - 0.5768099))) - pnorm(0.5, 100 * 0.5768099, sqrt(100 * 0.5768099 * (1 - 0.5768099))) = 2.629516e - 30$

$\square$

Universidade Vigo

Departamento de Estatística e Investigación Operativa

E.Superior de Enxeñería Informática
Edificio Politécnico
Campus de Ourense
E-32004 Ourense

Tel. 986 387 000
http://esei.uvigo.es

Estatística - Variables Aleatorias

Apelidos: **Nome:** **DNI:**

1. *(3 puntos)* La función de densidad de la v.a. X de Epanechnikov absolutamente continua es:

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

a) Determinar k para que sea función de densidad. Obtener la función de distribución.

Solución: $P(X \in \Re) = 1$
$P(X \in \Re) = \int_{-\infty}^{\infty} f(x)\,dx = \int_{-1}^{1} k\left(1-x^2\right)dx = k\left(x - \frac{1}{3}x^3\right)\Big|_{-1}^{1} = k\left(1-\frac{1}{3}\right) - k\left(-1+\frac{1}{3}\right) = k\left(1-\frac{1}{3}+1-\frac{1}{3}\right) =$
$k\left(2-\frac{2}{3}\right) = k\frac{4}{3}$,
por lo tanto, $\frac{4k}{3} = 1 \Leftrightarrow k = \frac{3}{4}$ □

b) Calcular dos medidas de tendencia central distintas.

Solución: $\mu = E(X) = \int_{-\infty}^{\infty} xf(x)\,dx = \frac{3}{4}\int_{-1}^{1} x\left(1-x^2\right)dt = 0$
Mediana=0, porque la función de densidad es simétrica con respecto al 0, i.e. $f(x) = f(-x)$. □

c) Calcular la probabilidad de que la v. X tome valores en el intervalo (-0.75,0.5) condicionado a que X toma valores negativos.

Solución: $P(X \in (-0.75, 0.5) | X < 0) = \frac{P(X\in(-0.75,0.5)\cap X<0)}{P(X<0)} == \frac{P(X\in(-0.75,0))}{P(X<0)} = \frac{\int_{-0.75}^{0} f(x)dx}{0.5} = \frac{1}{2}\int_{-0.75}^{0}\frac{3}{4}\left(1-x^2\right)dx$
$\frac{3}{8}\left(x-\frac{1}{3}x^3\right)\Big|_{-0.75}^{0} = \frac{3}{8}\left(0+0.75-\frac{1}{3}0.75^3\right) = \frac{3}{8}\left(0.75 - \frac{0.421875}{3} = 0.2285156\right)$ □

2. *(3 puntos)* Si $X \backsim N(0.5, 2)$, calcula,

a) $P(-0.5 < X < 2) =$

Solución: $P(-0.5 < X < 2) = P(X < 2) - P(X < -0.5) = P\left(N(0,1) < \frac{2-0.5}{2}\right) - P\left(N(0,1) < \frac{-0.5-0.5}{2}\right) =$
$P(N(0,1) < 0.75) - P(N(0,1) < -0.5) = (1 - 0.2266) - P(N(0,1) > 0.5) = 0.7734 - 0.3085 = 0.4649.$
□

b) El valor de a que verifica $P(X > a) = 0.95$. Llega al valor de a usando las funciones de R relacionadas con la Normal con y sin los argumentos *mean=0,sd=1*.

Solución:

El cuantil 0.05, $x_{0.05}$, verifica que: $P(X < x_{0.05}) = 0.05$, y aplicando el complementario tenemos el valor de a. Tipificando y resolviendo obtenemos que:
$P\left(N(0,1) < \frac{x_{0.05}-0.50}{2}\right) = 0.05$. El cuantil de la normal 0.5 es negativo porque su probabilidad asociada es menor que 0.5. Aplicando simetría con respecto al cero y buscando en las tablas el valor 0.05 obtenemos que:
$-\frac{x_{0.05}-0.50}{2} = 1.645$, resolviendo, $x_{0.05} = 0.50 - 2 * 1.645 = -2.79$. □

3. *(4 puntos)* Sea X una variable aleatoria distribuida normalmente con media 50 y desviación estándar 4, calcula:

a) $P(X < 40)$, y el intervalo más pequeño que contenga 0.9 de probabilidad.

b) Se repite de forma independiente 100 veces la v. aleatoria X. ¿Cuál es la probabilidad de que aparezca 60 veces o más el suceso $\{X > 40\}$? Responder usando la aproximación normal con la corrección de continuidad.

E.Superior de Enxeñería Informática
Edificio Politécnico
Campus de Ourense
E-32004 Ourense

Tel. 986 387 000
http://esei.uvigo.es

a) $P(X < 40) = P\left(\frac{X-50}{4} < \frac{40-50}{4}\right) = P(Z < -2.5) = P(Z > 2.5) = 0.00621$ con $Z \approx N(0,1)$,
y $P(20 < X < 50) = P(X < 50) - P(X < 20) = P\left(\frac{X-50}{4} < \frac{50-50}{4}\right) - P\left(\frac{X-50}{4} < \frac{20-50}{4}\right) = P(Z < 0) - P(Z < -7.5) \approx 0.5 - 0.$

b) $P(a < X < b) = P\left(\frac{a-50}{4} < \frac{X-50}{4} < \frac{b-50}{4}\right) = 0.9$ como la f. de densidad de la distribución normal es simétrica y unimodal con respecto al 0 y creciente hacia el 0, entonces el límite inferior y el superior del intervalo tienen que distar del 0 la misma cantidad, es decir, $\frac{b-50}{4} - 0 = 0 - \frac{a-50}{4}$ y el área está concentrada en torno al cero.
$P\left(Z < \frac{b-50}{4}\right) = 0.95$ es decir el cuantil 0.95 de la distribución $N(0,1)$, $z_{0.95} = 1.645$ y por lo tanto $\frac{b-50}{4} = 1.645$, $b = 50 + 4 * 1.645 = 56.58$, y $a = 43.42$.