Universidade Vigo

Departamento de Estatística
e Investigación Operativa

ESE Informática
Grao en Informática
Campus de Ourense

esei.uvigo.es
Fecha: 26/11/18

Estadística Descriptiva[1]

Apellidos:	**Nombre:**	**DNI:**

1. *(6.5 puntos)* El archivo *CitiesQualityOfLiving.RData*[2], contiene datos referentes a la calidad de vida en 170 ciudades del mundo. Concretamente, continen las variables

   - *city_name*: Nombre de la ciudad.
   - *DBpedia_URL*: Dirección del portal DBpedia.
   - *rating*: Puntuación obtenida.

   (a) *(0.5 puntos)* Clasificar estadísticamente las variables del archivo.

   (b) *(2.0 puntos)* Da la distribución de frecuencias completa de la variable *rating* en grupos que representen a las ciudades con niveles de vida *bajo* <*50*, $50 \leq medio < 96$ y $96 \leq alto$. Representa gráficamente y adecuamente la variable agrupada.

   (c) *(2.0 puntos)* Resume numéricamente la variable *rating* por grupo. ¿Es justificable que el coeficiente de variación disminuya por grupos al aumentar el nivel de vida?

   (d) *(2.0 puntos)* Agrupa nuevamente la variable *rating* en tres grupos *bajo, medio y alto* pero que cada grupo tenga el mismo número de ciudades.

2. *(3.5 puntos)* Dada la siguiente tabla de Estadísticos resumen, obtén la misma tabla de Estadísticos pero con la variable Tipificada. Justifica todos los valores.

```
 mean       sd     IQR        cv skewness  kurtosis   0% 25%    50%    75%  100%  n
6.565 2.893134 3.0525 0.4406907 1.166989 0.4408272 3.23 4.5 5.685 7.5525 13.42 20
```

[1]No se considerarán las respuestas contestadas sin su correspondiente desarrollo.
[2]Descargar desde la url *https://www.dropbox.com/s/5t38ykuxpj22trw/CitiesQualityOfLiving.RData?dl=0*

Variables Aleatorias[3]

Apellidos: **Nombre:** **DNI:**

1. *(3.0 puntos)* Un número se elige aleatoriamente al azar dado por la función de densidad:

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

   (a) *(1.5 puntos)* Determinar k para que sea función de densidad y la varianza.

   Solución:
   - $P(X \in \Re) = 1$
     $P(X \in \Re) = \int_{-\infty}^{\infty} f(x)\,dx = \int_{-1}^{1} \frac{1}{k} dx = \frac{1}{k}\, x]_{-1}^{1} = \frac{1}{k}(1+1) = \frac{2}{k}$,
     Por lo tanto, $\frac{2}{k} = 1 \Leftrightarrow k = 2$
   - $\mu = E(X) = \int_{-\infty}^{\infty} x f(x)\,dx = \int_{-1}^{1} \frac{x}{2} dt = \left.\frac{x^2}{4}\right]_{-1}^{1} = \frac{1}{4} - \frac{1}{4} = 0$
   - $\sigma^2 = Var(x) = E(X^2) - E(X)^2 = \int_{-\infty}^{\infty} x^2 f(x)\,dx - 0 = \int_{-1}^{1} \frac{x^2}{2} dx = \left.\frac{x^3}{6}\right]_{-1}^{1} = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$

   □

   (b) *(1.5 puntos)* Probabilidad de que el primer dígito decimal sea 1. Probabilidad de que el primer dígito decimal de su raíz cuadrada sea 3.

   Solución:
   - $P(X = ?.1??) = P(X \in (-0.2, -0.1] \cup [0.1, 0.2)) = \int_{-0.2}^{-0.1} f(x)\,dx + \int_{0.1}^{0.2} f(x)\,dx = 2 \times \int_{0.1}^{0.2} \frac{1}{2} dx = 2 \times \frac{1}{2}\, x]_{0.1}^{0.2} = \frac{2}{2}(0.2 - 0.1) = 0.1$
   - $P\left(Trunc\left(10\sqrt{X}\right) = 3\right) = P\left(10\sqrt{X} \in [3,4)\right) = P\left(\sqrt{X} \in [0.3, 0.4)\right) = P(X \in [0.09, 0.16)) = \int_{0.09}^{0.16} f(x)\,dx = \int_{0.09}^{0.16} \frac{1}{2} dx = \frac{1}{2}\, x]_{0.09}^{0.16} = \frac{1}{2}(0.16 - 0.09) = \frac{0.07}{2} = 0.035.$

   □

2. *(7.0 puntos)* El número de bits fallidos de escritura en un disco duro de 100Gb sigue una distribución de Poisson de media de 2.

   (a) *(1.5 puntos)* ¿Cuál es la probabilidad de que en un disco de 500Gb se encuentren menos de 8 errores?

   Solución: Definimos $S = X_1 + \ldots + X_5$, con X_i la v. que representa número de fallos en el trozo de disco duro de 100Gb dado por $[(i-1)*100, i*100)$Gb. $X_i \sim Pois(\lambda = 2)$, como son intervalos disjuntos las v. son independientes, entonces $S \sim Pois(\lambda = 2*5 = 10)$
   $P(S < 8) = P(S <= 7) = ppois(7, lambda = 10) = 0.2202206$ □

   (b) *(2.5 puntos)* Si dividimos el disco de 500Gb en 5 zonas de 100Gb, ¿cuál es la probabilidad de que en al menos tres zonas se tengan menos de 3 errores?

   Solución: Definimos $Y = I_{\{X_1<3\}} + I_{\{X_2<3\}} + \ldots + I_{\{X_5<3\}}$, es decir, contamos el número de éxitos $(X < 3)$ en 5 (pruebas). Como X_i son independientes, tenemos pruebas independientes. Dado que $P(X_i < 3) = P(X_i \leq 2) = ppois(2, lambda = 2) = 0.6766764$, tenemos que:

$$Y \sim Bi(5, 0.6766764)$$

   y

$$P(Y \geq 3) = 1 - P(Y \leq 2) = 1 - pbinom(2, size = 5, prob = 0.6766764) = 0.8047267$$

   □

[3]No se considerarán las respuestas contestadas sin su correspondiente desarrollo.

(c) *(3.0 puntos)* Consideremos discos duros de 10Tb, ¿cuál es la probabilidad de que se observen más de 190 errores? Dar la probabilidad exacta y la probabilidad aproximada por la distribución normal teniendo en cuenta que $Var(Pois(\lambda)) = \lambda$.

Solución: Sea S= número de errores en 10Tb, entónces siguiendo el apartado a) $S = X_1 + \ldots + X_{100} \sim Pois(\lambda = 2 * 100)$

- (Probabilidad Exacta) $P(S > 190) = 1 - ppois(190, lambda = 200) = 0.7470674$
- (Probabilidad Aprox.) Usando el T.C.L. $S \approx \tilde{S} \sim N(\lambda, \sqrt{\lambda}) = N(200, \sqrt{200})$, por lo tanto:

$$P(S > 190) \approx 1 - P(\tilde{S} \leq 190 + 0.5) = 1 - pnorm(190.5, mean = 200, sd = sqrt(200)) = 0.749129$$

□