Universida de Vigo

Departamento de Estatística e Investigación Operativa

Fac. CC. Económicas
e Empresariais
Campus de Vigo
E-36310 Vigo

Tel. 986 812 440
Fax 986 812 401
webs.uvigo.es/depc05
depc05@uvigo.es

Estadística Descriptiva

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

1. *(4 puntos)* En la tabla siguiente se muestra la distribución de frecuencias de las variables *PC* si el hogar dispone de ordenador y *Nº Hab*, número de habitantes por hogar de un estudio a hogares Ourensanos sobre brecha digital:

PC /Nº Hab	1	2	3	4	5	6
Si	11	69	92	60	15	0
Non	57	134	22	7	0	1

Solución:
```
matriz <- matrix(c(11,57,69,134,92,22,60,7,15,0,0,1),ncol=6)
colnames(matriz) <- 1:6
rownames(matriz) <- c("Si","Non")
```

□

a) Calcular, la distribución marginal y el valor medio del número de habitantes por hogar. Dar el porcentaje de hogares que son catalogadas como familias numerosas (≥ 3 habitantes por hogar).

Solución:

```
> m2 <- addmargins(matriz)[3,]   # Marxinal de Nº hab
> m2
  1   2   3   4   5   6 Sum
 68 203 114  67  15   1 468
>   sum((1:6)*m2[-7])/m2[7]   # media da marxinal Nº hab
     Sum
2.489316
>   sum(m2[3:6]*100)/m2[7]    # porcentaxe de habitantes por fogar con >= 3 habitantes
     Sum
42.09402
```

□

b) ¿Que número de habitantes se corresponde con el menor valor para el 70 % de los hogares con más habitantes?

Solución:

```
>   x <- rep( 1:6, m2[-7])    # construo o vector de tamaño 468
>   quantile(x,0.30)          # Cuantil 0.3
30%
  2
```

□

c) Distribución de frecuencias relativas del par (*PC, Nº de habitantes por hogar con familias numerosas*).

Solución:

Fac. CC. Económicas
e Empresariais
Campus de Vigo
E-36310 Vigo

Tel. 986 812 440
Fax 986 812 401
webs.uvigo.es/depc05
depc05@uvigo.es

```
>   matriz2 <- matriz[1:2,3:6]
>   matriz2/sum(matriz2)
            3          4          5           6
Si  0.4670051 0.30456853 0.07614213 0.000000000
Non 0.1116751 0.03553299 0.00000000 0.005076142
```

□

2. *(6 puntos)* El archivo *tempo.csv* que está en *dropbox* contiene información sobre el tiempo dedicado a la materia Estadística cada semana. Contiene las variables:
temp: Tiempo dedicado a la materia Estadística. Expresado en minutos.
semana: número de semana desde el comienzo de curso.
grupo: grupo en donde se recogieron los datos.

Solución:

```
datos <- read.csv('http://dl.dropbox.com/u/29008031/tempos.csv',sep=';',dec='.')
```

□

a) Describe numéricamente y gráficamente las v. *temp* y *semana*. Interpreta los valores dados. ¿Cuál es el tiempo mínimo dedicado a la materia del 80 % de los alumnos que más tiempo dedican?.

Solución: La variable *temp* es continua por lo tanto su representación gráfica es a través de un histograma

```
# descripción numérica
summary(datos$temp); sd(datos$temp)
hist(datos$temp,freq=FALSE) # eje Y la densidad de frecuencia
```

Interpretación de los valores y la gráfica ...
Cuantil 0.20 de la distribución continua *tempo*:

```
quantile(datos$temp,0.2)
20%
  60
```

□

La variable *semana* es discreta porque toma valores $1, 2, \ldots$ lo tanto su representación gráfica es a través de un diagrama de barras

```
# descripción numérica
summary(datos$semana); sd(datos$semana)
barplot(table(datos$semana)/sum(table(datos$semana))) #frecuencia relativa
```

Interpretación de los valores y la gráfica ...

b) Agrupa la variable *temp* en intervalos de longitud 1 hora. Da la frecuencia relativa por minuto de estudio y por intervalo.

Solución:

Departamento de Estatística e Investigación Operativa
Fac. CC. Económicas e Empresariais
Campus de Vigo
E-36310 Vigo
Tel. 986 812 440
Fax 986 812 401
webs.uvigo.es/depc05
depc05@uvigo.es

```
temp2 <- cut(datos$temp, breaks= seq(0,max(datos$temp),by=60),include.lowest=TRUE)
 table(temp2)/sum(table(temp2))    # frecuencia relativa por intervalo
 table(temp2)/sum(table(temp2))/60    # frecuencia relativa por minuto
```

□

c) Calcula el tiempo medio por grupo. También los tiempos medios durante los dos primeros meses (semanas del 1 al 8 inclusives) de clase y el resto.

Solución:

```
    by(datos$temp, datos$grupo, mean)    # medias por grupo
     datos$grupo: EST-1
[1] 199.7143
------------------------------------------
datos$grupo: EST-2
[1] 160.7031
------------------------------------------
datos$grupo: EST-3
[1] 175.9211
------------------------------------------
datos$grupo: EST-4
[1] 107.9787
------------------------------------------
datos$grupo: EST-5
[1] 116.4706

    mean(datos$temp[datos$semana<=8])     # media dos 2 primeiros meses <-> 8 semanas
     mean(datos$temp[datos$semana>8])     # media dos meses 3-adiante
```

□

Nota: para abrir el achivo usar desde R ejecutar la instrucción
datos <- read.csv('http://dl.dropbox.com/u/29008031/tempos.csv',sep=';',dec='.')

UniversidadeVigo

Departamento
de Estatística
e Investigación
Operativa

Fac. CC. Económicas
e Empresariais
Campus de Vigo
E-36310 Vigo

Tel. 986 812 440
Fax 986 812 401
webs.uvigo.es/depc05
depc05@uvigo.es

Calculo de Probabilidades y Variables aleatorias

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

Para facilitar la corrección de la prueba, marque en la tabla la letra de la respuesta que considere correcta para cada una de las cuestiones propuestas[1].

Pregunta 1	a	b	c
Pregunta 2	a	b	c

1. *(2 puntos)* La duración, en días, de los focos fabricados por una empresa es una variable con densidad $f(x) = \frac{k}{x^3}$ si $x > 1$ (y 0 en el otro caso). El valor de k para que la función esté bien definida es:

   a) $k = 2$

   b) $k > 0$.

   c) $k = 1$

   Solución: $k = 2$
   $k \int_1^\infty x^{-3}dx = \frac{-k}{2}x^{-2}|_1^\infty = \frac{k}{2}$
   Como tiene que ser igual a 1, entonces $k = 2$

2. *(2 puntos)* En una prueba de 10 preguntas de múltiple opción donde cada pregunta tiene 4 opciones y sólo una es correcta. Si la elección de la respuesta es aleatoria, la probabilidad de aprobar (con 3 decimales) es:

   a) 0.5

   b) 0.078

   c) 0.058

```
Solución correcta:
b) sum( dbinom( 5:10, size=10,prob=.25))=0.0781269
```

1. *(3 puntos)* Una caja contiene 4 monedas con una cruz en cada lado, 3 monedas con una cara en cada lado y 2 monedas legales. Si se selecciona al azar una de estas nueve monedas y se lanza una vez.

   a) ¿Cuál es la probabilidad de obtener cara?.

   Solución:
   Aplicando probabilidades totales descomponemos el suceso *Cara* en función de a que tipo de moneda se obtuvo *G1=4 monedas sin cara*,*G2=3 monedas sin cruz*,*G3=2 monedas legales*
   $P(C) = P(C/G1)P(G1) + P(C/G2)P(G2) + P(C/G3)P(G3) = 0 * 4/9 + 1 * 3/9 + 1/2 * 2/9 = 8/18 = 4/9 = 0.4444$
   O si consideramos que hay 9 monedas con 8 caras y 10 cruces–> 8/18 □

   b) Si se obtiene cara, ¿cuál es la probabilidad de que la moneda sea legal?.

   Solución:
   Aplicando la regla de Bayes:
   $P(G3/C) = \frac{P(C/G3)*P(G3)}{P(C)} = \frac{(1/2)*(2/9)}{8/18} = 1/4 = 0.25$ □

2. *(3 puntos)* Los errores de cálculo de un determinado proceso siguen una distribución $U\left(-1 \times 10^{-6}, 1 \times 10^{-6}\right)$. Se considera un error inadmisible si el error supera en valor absoluto el valor de 0.75×10^{-6}. Calcular:

   a) La probabilidad de error inadmisible.

   Solución:
   Sea X= *error* $\sim U\left(-1 \times 10^{-6}, 1 \times 10^{6}\right)$,
   P(*error inadmisible*)= $P(X < -0.75 \times 10^{-6}) + P(X > 0.75 \times 10^{-6}) = 2 * P(X < -0.75 \times 10^{-6}) = 2 * punif(-0.75 * 10^{-6}, min = -10^{-6}, max = 10^{-6}) = 0.25$ □

[1] 1 respuestas incorrectas penalizan un tercio de una respuesta correcta. Las preguntas en blanco no penalizan. No se considerarán las respuestas contestadas sin su correspondiente desarrollo.

Universidade Vigo

Departamento de Estatística e Investigación Operativa

Fac. CC. Económicas
e Empresariais
Campus de Vigo
E-36310 Vigo

Tel. 986 812 440
Fax 986 812 401
webs.uvigo.es/depc05
depc05@uvigo.es

b) Si se ejecutan 1000 cálculos independientes del mismo proceso. ¿Cuál es la probabilidad de que el nº de errores este entre 225 y 275?. Dar los valores teóricos y los aproximados por la distribución normal con corrección de continuidad.

Solución:
Sea Y=*nº de operaciones erróneas* $\sim Bi\,(n = 1000, p = 0.25) \approx N\left(1000 \times 0.25, \sqrt{1000 \times 0.25 \times 0.75}\right)$ aproximación válida porque $n > 30$ y $np(1-p) = 187.5 > 5$

$$\begin{aligned} P\,(225 < Y < 275) = \sum_{i=226}^{274} \binom{1000}{i} 0.25^i (1-0.25)^{1000-i} \approx \\ P\,(226 - 0.5 < N(250, 13.693)) < 274 + 0.5) = \\ P\,(N(250, 13.693)) < 274 + 0.5) - P\,(N(250, 13.693)) < 226 - 0.5) = \\ pnorm(274.5, 250, 13.693) - pnorm(225.5, 250, 13.693) = 0.926423 \end{aligned}$$

Sin aproximación a la normal tenemos $sum(dbinom(226 : 274, size = 1000, prob = 0.25)) = 0.9265236$ □