

PRIMER PARCIAL
SÁBADO 30 DE ABRIL DE 2016.

Número de Parcial	Cédula	Nombre y Apellido

PARA USO DOCENTE.				
Ej. 1	Ej. 2	Ej. 3	Ej.4	TOTAL

Ejercicio 1. [8 puntos]

Un pueblo de 2000 habitantes sufre regularmente brotes epidémicos de una cierta enfermedad. De la experiencia de brotes anteriores se sabe que durante un brote la probabilidad de que un habitante deba ser internado es 0.1. Se supone que el pueblo se encuentra bajo un brote epidémico de la enfermedad.

1. Estimar la probabilidad de que exactamente 200 habitantes deban ser internados.
2. Estimar la probabilidad de que deban ser internados entre 185 y 215 habitantes.
3. Sea s el número de camas disponibles en el hospital del pueblo. ¿Cuál es el mínimo valor de s tal que la probabilidad de que se deba internar a más de s habitantes sea menor que 0.1?

Ejercicio 1. Solución

Sea X la cantidad de habitantes que deben ser internados durante un brote epidémico. La variable aleatoria X tiene distribución Binomial con parámetros $n = 2000$ y $p = 0.1$. Observando que el parámetro n es grande, podemos utilizar la aproximación Normal de la Binomial. Sean $\mu = np = 2000 \times 0.1 = 200$ y $\sigma = \sqrt{np(1-p)} = 13.42$.

1. $P(X = 200) \approx \frac{1}{\sigma} \varphi\left(\frac{200-\mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi}13.42} e^{-0} = 0.0297$.
- 2.

$$\begin{aligned}
 P(185 \leq X \leq 215) &\approx \Phi\left(\frac{215-\mu}{\sigma}\right) - \Phi\left(\frac{185-\mu}{\sigma}\right) = \Phi(1.12) - \Phi(-1.12) \\
 &= 2\Phi(1.12) - 1 = 2 \times 0.8686 - 1 = 0.7372
 \end{aligned}$$

3. Se busca el mínimo valor de s tal que $P(X > s) \leq 0.1$, es decir $P(X \leq s) \geq 0.9$. Nuevamente aproximando, se tiene que $P(X \leq s) \approx \Phi\left(\frac{s-\mu}{\sigma}\right)$. Por lo tanto vamos a buscar en la tabla de la $\mathcal{N}(0, 1)$ el menor valor z tal que $\Phi(z) \geq 0.9$. De la tabla se obtiene que $z = 1.29$. Resulta entonces que $s = \sigma \times z + \mu = 217.3$. Por lo tanto el mínimo número de camas debe ser $s = 218$.

Ejercicio 2. [10 puntos]

El cartero de una gran ciudad dice ser capaz de identificar el género del remitente a partir de su caligrafía. Un colega escéptico le propone el siguiente test: toma 10 cartas de las cuáles sabe que en 5 el remitente es mujer y en 5 el remitente es un hombre y le pide que indique cuáles fueron las 5 cartas escritas por una mujer. El cartero acierta en 4 de las 5 cartas.

1. El colega no está convencido e insiste en que este resultado es producto del azar. Se supone entonces que el cartero elige las 5 cartas **al azar**:
 - (a) Sea X la variable aleatoria que cuenta el número de aciertos del cartero. Indicar \mathcal{R}_X el recorrido de X . Hallar y graficar $p_X(x)$ la función de probabilidad puntual de X .
 - (b) Hallar la probabilidad de que el número de aciertos sea mayor o igual que 4.
 - (c) ¿Usted está de acuerdo con el escepticismo del colega? Justifique su respuesta.
2. Para convencer a su colega, el cartero propone repetir el experimento pero con 100 cartas (50 escritas por una mujer y 50 por un hombre). En este caso el cartero acierta en 32 cartas. En la Figura 1 se indica algunos valores de la función de probabilidad puntual de X en este caso que pueden ser de utilidad. ¿Cambia esto su opinión con respecto al escepticismo del colega? Justifique su respuesta.

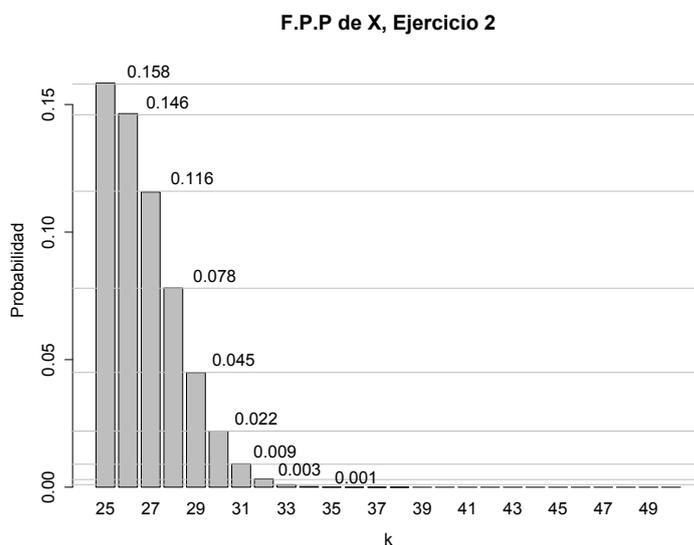


Figure 1: Esta figura muestra la función de probabilidad puntual de X - para algunos valores significativos - del Ejercicio 2, parte 2. Los valores no indicados se estiman como iguales a cero.

Ejercicio 2. Solución

1. (a) El número de aciertos X del cartero se modela como $X \sim \mathcal{H}(10, 5, 5)$ Hipergeométrica de parámetros $N = 10$, $k = 5$ y $n = 5$; pues se eligen 5 cartas, de un total de 10 de las cuáles 5 fueron escritas por una mujer y claramente no puede haber repeticiones. Por lo tanto $\mathcal{R}_X = \{0, 1, 2, 3, 4, 5\}$ y $p_X(k) = P(X = k) = \frac{C_k^5 C_{5-k}^5}{C_{10}^5} = \frac{(C_k^5)^2}{C_{10}^5}$ (ver Fig. ??). Haciendo los cálculos se obtiene que:

$$p_X(0) = p_X(5) = \frac{1}{252} = 0.004$$

$$p_X(1) = p_X(4) = \frac{25}{252} = 0.0992$$

$$p_X(2) = p_X(3) = \frac{100}{252} = 0.397$$

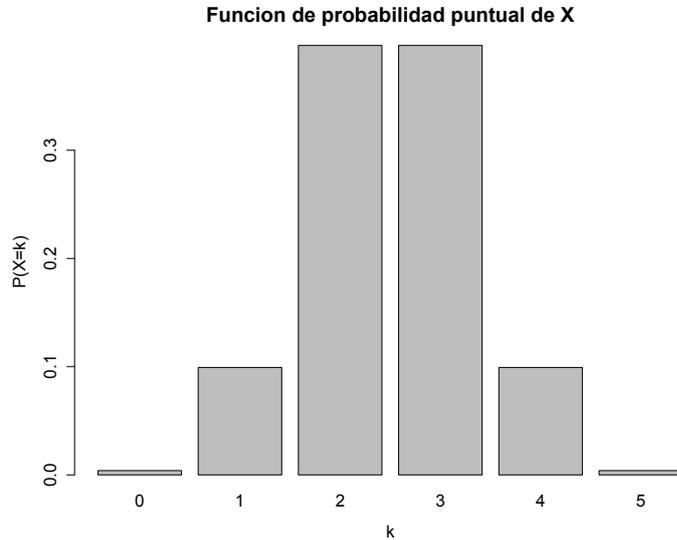


Figure 2: Esta figura muestra la función de probabilidad puntual de $X \sim \mathcal{H}(10, 5, 5)$.

(b) $P(X \geq 4) = P(X = 4) + P(X = 5) = 0.0992 + 0.004 = 0.1032$.

(c) La parte anterior nos dice que si el cartero eligiera al azar entonces tiene una probabilidad de 0.1 de haber acertado en 4 o más cartas. Esto puede interpretarse como que el cartero acertaría 1 de cada 10 veces que intentara identificar a las cartas cuyo remitente es una mujer. Si considera que esta probabilidad es significativa, entonces usted está de acuerdo con el escepticismo del colega. Sin embargo si cree que esta probabilidad es pequeña entonces no está de acuerdo con el colega y cree que hay evidencia para sostener la capacidad del cartero de reconocer el género del remitente.

2. Nuevamente tenemos que si el cartero elige al azar entonces $X \sim \mathcal{H}(100, 50, 50)$ tiene distribución Hipergeométrica de parámetros $N = 100$, $k = 50$ y $n = 50$. Al igual que en la parte anterior vamos a calcular la probabilidad de que haya acertado más de 32 cartas, esto es $P(X \geq 32) = \sum_{k=32}^{50} p_X(k)$.

Utilizando la Figura 1 obtenemos que $P(X \geq 32) = 0.003 + 0.001 = 0.004$. Esto es, apenas 4 de 100 veces que se repitiera este experimento, el cartero acertaría 32 o más veces. Si usted aún no cree en la capacidad del cartero, entonces es un escéptico incurable.

Ejercicio 3. [12 puntos]

La siguiente tabla muestra los datos correspondientes a un estudio realizado con el objetivo de comprender la relación entre el diámetro de los granos de arena (D en mm) y la pendiente de profundidad del agua (P en grados) para $n = 9$ playas.

i	1	2	3	4	5	6	7	8	9
D	0.170	0.190	0.220	0.235	0.235	0.300	0.350	0.420	0.850
P	0.63	0.70	0.82	0.88	1.15	1.50	4.40	7.30	11.30

Algunos datos que pueden ser de utilidad en el ejercicio: $\sum_{i=1}^9 (D_i - \bar{D})^2 = 0.355$ y $\sum_{i=1}^9 (P_i - \bar{P})^2 = 108.28$.

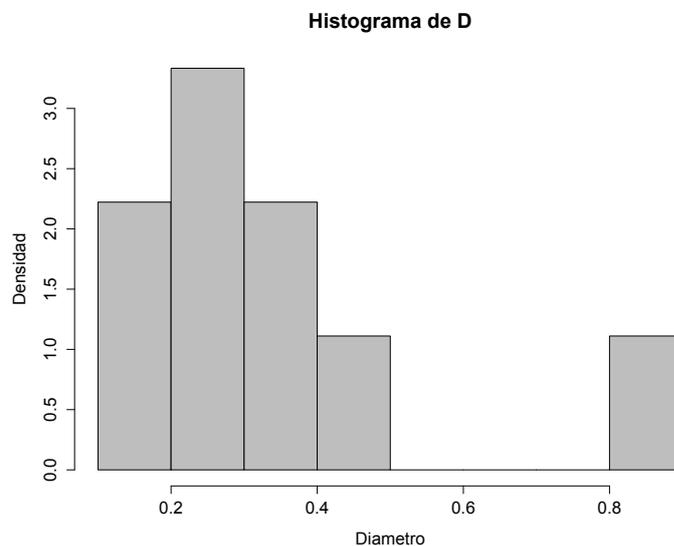
1. Realizar un histograma de D con ocho intervalos que cubran el rango de 0.1 mm a 0.9 mm.
2. Realizar un diagrama de cajas (boxplot) para D (indicar en el diagrama todos los valores numéricos que son necesarios para construirlo). Estudiar la simetría de la distribución de D . Indicar si existen datos atípicos.
3. (a) Realizar y describir un diagrama de dispersión para (D, P) .
(b) Calcular el coeficiente de correlación r . Indicar si la correlación es débil, moderada, o fuerte.
4. Explicar qué porcentaje de la variación de P es explicada por la variación de D .
5. Hallar la ecuación de la recta de regresión. Calcular el error cuadrático medio de predicción.
6. Si el diámetro de los granos de arena de una playa es mayor que el de otra en 0.1 mm, ¿cuánto es la diferencia pronosticada entre las respectivas pendientes?

Ejercicio 2. Solución

1. Para hacer el histograma de D dividimos el intervalo $[0.1, 0.9]$ en ocho intervalos iguales de longitud $h = 0.1$ mm. Luego para cada subintervalo calculamos la densidad del mismo, es decir la frecuencia relativa dividido h . Hacemos una tabla

Intervalo	0.1 – 0.2	0.2 – 0.3	0.3 – 0.4	0.4 – 0.5	0.5 – 0.6	0.6 – 0.7	0.7 – 0.8	0.8 – 0.9
Frec. Rel.	2/9	3/9	2/9	1/9	0	0	0	1/9
Densidad	20/9	30/9	20/9	10/9	0	0	0	10/9

Luego solo resta graficar las densidades con una barra para cada intervalo.



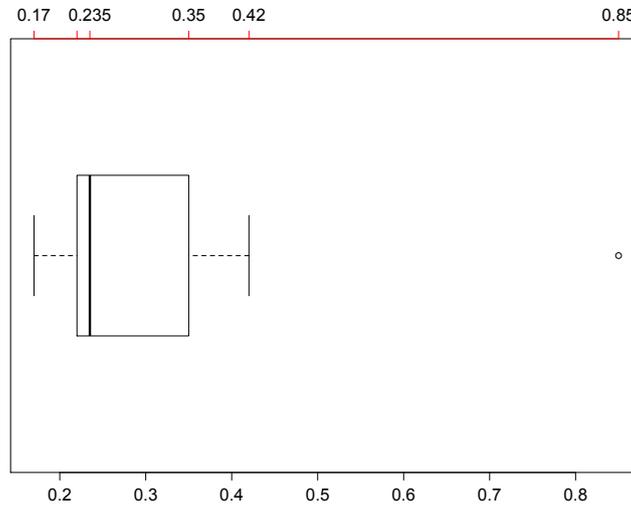
2. Para hacer un boxplot de D debemos primero calcular :

$$\frac{L_i \quad Q_1 \quad M \quad Q_3 \quad L_s}{0.025 \quad 0.22 \quad 0.235 \quad 0.35 \quad 0.545}$$

De aquí vemos que los extremos de los brazos del boxplot deben ser

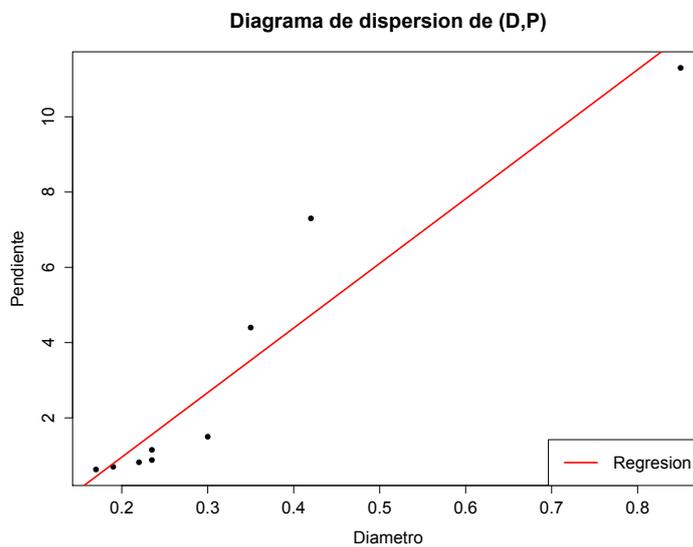
$$\min_{D_i \geq L_i} D_i = 0.17 \text{ y } \max_{D_i \leq L_s} D_i = 0.42.$$

El boxplot queda entonces:



Vemos que $Q_3 - M > M - Q_1$, por lo que la distribución de D es notoriamente asimétrica hacia la derecha. Además, el dato $D_9 = 0.85$ es un dato atípico.

3. (a) Para hacer el diagrama de dispersión debemos graficar los pares (D_i, P_i) para $i = 1, \dots, 9$. El resultado es:



Del diagrama vemos que existe una asociación fuerte entre P y D . Además, la asociación es positiva, ya que P parece aumentar cuando D aumenta. Vemos también que el par $(0.85, 11.3)$ es atípico y le da fuerza a la asociación.

(b) El coeficiente de correlación se calcula usando la fórmula

$$r = \frac{\text{cov}(D, P)}{s_D s_P}.$$

Una forma práctica consiste en agrupar los cálculos en una tabla:

	D_i	\bar{D}	$D_i - \bar{D}$	$(D_i - \bar{D})^2$	P_i	\bar{P}	$P_i - \bar{P}$	$(P_i - \bar{P})^2$	$(D_i - \bar{D})(P_i - \bar{P})$
	0.170	0.33	-0.160	0.0256	0.63	3.1867	-2.5567	6.5365	0.4091
	0.190	0.33	-0.140	0.0196	0.70	3.1867	-2.4867	6.1835	0.3481
	0.220	0.33	-0.110	0.0121	0.82	3.1867	-2.3667	5.6011	0.2603
	0.235	0.33	-0.095	0.0090	0.88	3.1867	-2.3067	5.3207	0.2191
	0.235	0.33	-0.095	0.0090	1.15	3.1867	-2.0367	4.1480	0.1935
	0.300	0.33	-0.030	0.0009	1.50	3.1867	-1.6867	2.8448	0.0506
	0.350	0.33	0.020	0.0004	4.40	3.1867	1.2133	1.4722	0.0243
	0.420	0.33	0.090	0.0081	7.30	3.1867	4.1133	16.9195	0.3702
	0.850	0.33	0.520	0.2704	11.30	3.1867	8.1133	65.8262	4.2189
Suma	2.97			0.3551	28.68			114.8525	6.0941

Luego, el coeficiente de correlación es

$$r = \frac{6.0941}{\sqrt{0.3551}\sqrt{114.8525}} = 0.9542.$$

Vemos que la correlación es muy fuerte.

Observación: En esta parte del ejercicio hay un error en la letra. El valor correcto de $\sum_{i=1}^9 (P_i - \bar{P})^2$ es 114.85 y no 108.28. Usando el dato dado por la letra el valor de r que se obtiene es 0.9827. Ambos valores de r serán considerados correctos en la corrección.

4. El porcentaje de la variación de P que es explicada por la variación de D , es por definición

$$100 \times \frac{\text{Variación explicada}}{\text{Variación de } P} = 100 \times \frac{s_{\text{reg}}^2}{s_P^2}.$$

Del teórico, podemos calcularla usando la fórmula $s_{\text{reg}}^2/s_P^2 = r^2$. Luego

$$\% \text{ de variación explicada} = 91\%.$$

5. La ecuación de la recta de regresión es $y = \frac{rs_P}{s_D} (x - \bar{D}) + \bar{P}$. En nuestro caso

$$y = 17.16 \times (x - 0.33) + 3.19$$

Aquí, hemos redondeado a dos cifras significativas para simplificar la fórmula.

El error cuadrático medio de predicción, se puede calcular usando la fórmula

$$\text{E.C.M.} = \sqrt{1 - r^2} \times s_P.$$

En nuestro caso obtenemos $\text{E.C.M.} = 1.134$.

6. Si la diferencia en diámetro de los granos de arena de las playas $\Delta D = 0.1$ mm, entonces podemos obtener ΔP de la siguiente forma:

$$\Delta P = \frac{rs_P}{s_D} \times \Delta D = 17.16 \times 0.1 = 1.716 \text{ grados.}$$

Ejercicio 4. [10 puntos]

1. Sea X una variable aleatoria Poisson de parámetro $\lambda > 0$. Hallar la(s) moda(s) de X discutiendo según λ . Recordar que θ es moda de X si $P(X = \theta) \geq P(X = k) \forall k \in \mathcal{R}_X$.
2. Una pequeña empresa de alquiler de bicicletas tiene un stock de 12 bicicletas. La experiencia indica que la demanda diaria de bicicletas (D) es una variable aleatoria Poisson con parámetro $\lambda = 10$ (en la tabla al final del ejercicio se presentan las probabilidades puntuales significativas de D)
 - (a) Hallar la probabilidad de que un día la demanda supere al stock.
 - (b) Hallar la mayor moda θ de D y calcular la probabilidad de que un día la demanda supere el valor θ .
 - (c) Hallar el mínimo número de bicicletas en stock necesarias para que la probabilidad de que la demanda supere al stock sea menor que 0.1.
 - (d) Se considera una semana de 6 días laborables:
 - i. Hallar la probabilidad de que ningún día de la semana la demanda supere al stock.
 - ii. Hallar la probabilidad de que al menos dos días de la semana la demanda supere al stock.

k	0	1	2	3	4	5	6	7	8	9	10	11
$P(D = k)$	0.000	0.000	0.002	0.008	0.019	0.038	0.063	0.090	0.113	0.125	0.125	0.114
k	12	13	14	15	16	17	18	19	20	21	22	23
$P(D = k)$	0.095	0.073	0.052	0.035	0.022	0.013	0.007	0.004	0.002	0.001	0.000	0.000

Ejercicio 4. Solución

1. Hay que estudiar cociente $\frac{P(X=k+1)}{P(X=k)}$. Si k es moda de X , entonces el cociente es menor que 1. Usando la función de probabilidad se tiene que $\frac{P(X=k+1)}{P(X=k)} = \frac{\lambda}{k+1}$. Por lo tanto, si $\lambda \in \mathbb{N}$ entonces $k = \lambda$ y $k = \lambda - 1$ son modas de X , mientras que si $\lambda \notin \mathbb{N}$ entonces la única moda es $k = \lfloor \lambda \rfloor$.
2. Tenemos que la demanda D se comporta como una v.a. Poisson de parámetro $\lambda = 10$. En la tabla al final del ejercicio tenemos algunas probabilidades puntuales de esta distribución que pueden ser de utilidad.
 - (a) $P(D > 12) = 0.073 + 0.052 + 0.035 + 0.022 + 0.013 + 0.007 + 0.004 + 0.002 + 0.001 = 0.209$.
 - (b) De acuerdo a la parte anterior (o la tabla), las modas son $\theta_1 = 9$ y $\theta_2 = 10$. Por lo tanto la mayor moda es $\theta = 10$. Entonces

$$P(D > 10) = P(D = 11) + P(D = 12) + P(D > 12) = 0.114 + 0.095 + 0.2 = 0.4180.$$
 - (c) Buscamos entonces el menor valor de k tal que $P(D > k) < 0.1$. De la tabla tenemos que $k = 14$. Se puede ver sumando desde los últimos valores de la tabla y hallando el primer valor tal que la probabilidad acumulada es mayor a 0.1; o sumar desde el primero de los valores y hallar el primer valor tal que la probabilidad acumulada es mayor a 0.9.
 - (d) Sea Y la variable aleatoria que cuenta el número de días de la semana en que la demanda supera al stock. Resulta entonces Y es una variable aleatoria Binomial con parámetros $n = 6$ y $p = P(D > 12) = 0.209$.
 - i. $P(Y = 0) = (1 - p)^6 = 0.245$.
 - ii.

$$\begin{aligned}
 P(Y \geq 2) &= 1 - P(Y, 2) = 1 - [P(Y = 0) + P(Y = 1)] \\
 &= 1 - [0.245 + 6 \times 0.209 \times 0.791^5] \\
 &= 1 - [0.245 + 0.3883] = 0.367.
 \end{aligned}$$