PEC-3 Primavera 2023

UOC

Os puede ser útil consultar el siguiente material:

1. Reto 2. Probabilidad y variables aleatorias: Módulo Probabilidad y variables aleatorias.
2. Reto 2. Probabilidad y variables aleatorias: Variables aleatorias. Actividades Resueltas.
3. Reto 2. Probabilidad y variables aleatorias: Distribuciones de probabilidad e inferencia estadística con R.

NOMBRE:

PEC3

Muy importante:

En todas las preguntas, tenéis que indicar qué tipo de variable aleatoria usáis y plantear la probabilidad que se pide, como por ejemplo $P(X = 4), P(X \leq 3)$, etc, antes de poner las instrucciones y hacer los cálculos con R. Procurad usar las funciones: dnorm, pnorm, dbinom, . . .

Toda respuesta que no lleve todas estas explicaciones, con los comentarios pertinentes, será considerada incorrecta aunque se dé un resultado numérico correcto.

Pregunta 1 (30%)

En una empresa de venta online, el 80% de los paquetes enviados en un día llegan a su destino sin problemas. En un determinado día, se envían 250 paquetes.

a) ¿Cuál es la probabilidad de que exactamente 50 paquetes tengan problemas en llegar a su destino? (10%)
b) ¿Cuál es la probabilidad de que 69 o menos paquetes tengan problemas en llegar a su destino? (10%)

c) Si sabemos que 41 o más paquetes han tenido problemas en llegar a su destino, ¿cuál es la probabilidad de que 60 paquetes o menos hayan tenido problemas en llegar a su destino? (10%)

Solución:

Sea X="número de paquetes que han tenido problemas en llegar a su destino". Esta variable aleatoria se distribuye según una distribución binomial de parámetros $n = 250$ y $p = 1-0.8 = 0.2$. Su función de masa de probabilidad viene dada por:

$$P(X = k) = \binom{250}{k} \cdot 0.2^k \cdot 0.8^{250-k}, \qquad k = 0, 1, 2, ..., 250$$

a) Nos piden $P(X = 50)$, que podemos calcular de estas dos maneras:

- $P(X = 50) =$

```
dbinom(50, size = 250, prob = 0.20)
```

```
## [1] 0.06296802
```

- $P(X = 50) = P(X \leq 50) - P(X \leq 49) =$

```
pbinom(50, size=250, prob=0.20,lower.tail=TRUE)-
  pbinom(49, size=250, prob=0.20,lower.tail=TRUE)
```

```
## [1] 0.06296802
```

b) Ahora nos piden $P(X \leq 69)$, que podemos calcular de estas dos maneras:

- $P(X \leq 69) = P(X = 0) + P(X = 1) + P(X = 2) + ... + P(X = 69) =$

```
sum(dbinom(0:69, size = 250, prob = 0.20))
```

```
## [1] 0.9985135
```

- $P(X \leq 69) =$

```r
pbinom(69, size=250, prob=0.20,lower.tail=TRUE)
```

```
## [1] 0.9985135
```

c) La probabilidad que nos piden es una probabilidad condicionada, por lo que para calcularla aplicaremos la fórmula de la probabilidad condicionada;

$$P(X \leq 60|X \geq 41) = \frac{P(41 \leq X \leq 60)}{P(X \geq 41)} = \frac{P(X \leq 60) - P(X \leq 40)}{1 - P(X \leq 40)} =$$

$$= \frac{0.8853448}{0.9363164} = 0.9455616$$

```r
0.8853448/0.9363164
```

```
## [1] 0.9455616
```

donde estas probabilidades se han calculado con R de las siguientes maneras:

- $P(41 \leq X \leq 60) = P(X = 41) + ... + P(X = 60) =$

```r
sum(dbinom(41:60, size = 250, prob = 0.20))
```

```
## [1] 0.8853448
```

también: $P(41 <= X <= 60) = P(X <= 60) - P(X <= 40) =$

```r
pbinom(60, size=250, prob=0.20,lower.tail=TRUE)-
pbinom(40, size=250, prob=0.20,lower.tail=TRUE)
```

```
## [1] 0.8853448
```

- $1 - P(X \leq 40) =$

```r
1-sum(dbinom(0:40, size = 250, prob = 0.20))
```

```
## [1] 0.9363164
```

```
1-pbinom(40, size=250, prob=0.20,lower.tail=TRUE)
```

```
## [1] 0.9363164
```

Pregunta 2 (35%)

a) El número de solicitudes en 1 minuto que recibe una página web sigue una distribución de Poisson con una media de 200 solicitudes. ¿Cuál es la probabilidad de que en 3 minutos reciba un número de solicitudes comprendido entre 540 (donde 540 está incluido) y 650 (donde 650 está incluido)? (15%)

b) Estamos analizando el rendimiento de un algoritmo de búsqueda en una base de datos y sabemos que la probabilidad de que una búsqueda produzca el resultado deseado es de 0.65. Si se realizan diversas búsquedas de forma independiente, ¿cuál es la probabilidad de que la primera búsqueda con el resultado deseado sea antes de la quinta búsqueda? Observación: Si Y es una variable aleatoria geométrica de parámetro p para calcular $P(Y = k)$ usaremos en R la función $dgeom(x = k - 1, prop = p)$. Fijaos que en la función tenemos que poner $k - 1$. (20%)

Solución:

a) Si en 1 minuto se reciben en media 200 solicitudes, en 3 minutos se recibirán en media $200 \cdot 3 = 600$ solicitudes. Sea X el número de solicitudes que se reciben en 3 minutos. Esta variable sigue una distribución de Poisson de parámetro $\lambda = 600$, cuya función de masa de probabilidad viene dada por:

$$P(X = k) = \frac{e^{-600} \cdot 600^k}{k!}, \qquad k = 0, 1, 2, ...$$

Nos piden la probabilidad: $P(540 \leq X \leq 650)$ que podemos calcular con R de las siguientes formas:

- $P(540 \leq X \leq 650) = P(X = 540) + P(X = 541) + ... + P(X = 650) =$

```
sum(dpois(540:650, lambda=600))
```

```
## [1] 0.973246
```

- $P(540 \leq X \leq 650) = P(X \leq 650) - P(X \leq 539) =$

```
ppois(650, lambda=600,lower.tail=TRUE)-
  ppois(539, lambda=600,lower.tail=TRUE)
```

```
## [1] 0.973246
```

b) Sea Y la variable aleatoria que cuenta el número de búsquedas hasta la primera búsqueda con el resultado deseado. Esta variable Y sigue una distribución geométrica con parámetro $p = 0.65$ y cuya función de masa de probabilidad viene dada por:

$$P(Y = k) = (1 - 0.65)^{k-1} \cdot 0.65, \qquad k = 1, 2, ...$$

Nos piden la probabilidad:

$$P(Y < 5) = P(Y \leq 4) = \sum_{k=1}^{4} (1 - 0.65)^{k-1} \cdot 0.65$$

que podemos calcular con R mediante las siguientes instrucciones:

- $P(Y < 5) = P(Y \leq 4) = P(Y = 1) + P(Y = 2) + P(Y = 3) + P(Y = 4) =$

```
sum(dgeom(0:3, prob=0.65))
```

```
## [1] 0.9849937
```

- $P(Y < 5) = P(Y \leq 4) =$

```
pgeom(3, prob=0.65)
```

```
## [1] 0.9849938
```

Pregunta 3 (35%)

Se sabe que el tiempo de respuesta, X (en segundos) de un servidor web sigue una distribución normal con una desviación típica de 0.75 segundos. Se sabe que:

- La probabilidad de que el tiempo de respuesta sea menor o igual a 2 segundos es de 0.3203692.

Se pide:

a) ¿Cuál es la media (μ) de los tiempos de respuesta del servidor web? (15%)
b) ¿Qué porcentaje del tiempo de respuesta del servidor web se encuentran entre 1.25 y 2.75 segundos? (10%)
c) Encontrad el tiempo t tal que el 90% de les peticiones duren menos de t segundos. (10%)

Solución:

a) La probabilidad dada por el enunciado, la podemos expresar como $P(X \leq 2) = 0.3203692$. Tipificando, tenemos que:

$$P\left(\frac{X-\mu}{0.75} \leq \frac{2-\mu}{0.75}\right) = 0.3203692$$

Utilizando la instrucción de R de la probabilidad acumulada inversa, tenemos que:

```
qnorm(0.3203692,mean=0,sd=1,lower.tail=TRUE)
```

```
## [1] -0.4666666
```

Por tanto,

$$\frac{2-\mu}{\sigma} = -0.4666667$$

Podemos encontrar el valor de μ: $\mu = 2 + 0.466667 \cdot 0.75 = 2.35$ segundos.

```
2+0.466667*0.75
```

```
## [1] 2.35
```

b) Nos piden la probabilidad:

$P(1.25 < X < 2.75) = P(X \leq 2.75) - P(X \leq 1.25) = 0.7030986 - 0.07123338 = 0.6318652$

Donde hemos utilizado la instrucción de R que proporciona la probabilidad acumulada:

```
pnorm(2.75, mean =2.35 , sd = 0.75,lower.tail = TRUE)
```

```
## [1] 0.7030986
```

```
pnorm(1.25,  mean= 2.35, sd = 0.75,lower.tail = TRUE)
```

```
## [1] 0.07123338
```

```
pnorm(2.75, mean =2.35 , sd = 0.75,lower.tail = TRUE)-
pnorm(1.25,  mean= 2.35, sd = 0.75,lower.tail = TRUE)
```

```
## [1] 0.6318652
```

Por lo tanto, aproximadamente el 63.19% de los tiempos de respuesta del servidor web se encuentran entre 1.25 y 2.75 segundos.

c) Podemos obtener el tiempo t fácilmente utilizando la función de R de la probabilidad inversa:

```
qnorm(0.90,mean=2.35,sd=0.75,lower.tail=TRUE)
```

```
## [1] 3.311164
```

Así pues, $t_{0.90} = 3.31$ segundos.