

Práctica 4. Relación de Problemas de Cálculo de Probabilidades.

Juan de Dios Luna del Castillo, Antonio Martín Andrés



**UNIVERSIDAD
DE GRANADA**

Todo el material para el conjunto de actividades de este curso ha sido elaborado y es propiedad intelectual del grupo **BioestadísticaR** formado por:

Antonio Martín Andrés
Juan de Dios Luna del Castillo,
Pedro Femia Marzo,
Miguel Ángel Montero Alonso,
Christian José Acal González,
Pedro María Carmona Sáez,
Juan Manuel Melchor Rodríguez,
José Luis Romero Béjar,
Manuela Expósito Ruíz,
Juan Antonio Villatoro García.

Todos los integrantes del grupo han participado en todas las actividades, en su elección, construcción, correcciones o en su edición final, no obstante, en cada una de ellas, aparecerán uno o más nombres correspondientes a las personas que han tenido la máxima responsabilidad de su elaboración junto al grupo de **BioestadísticaR**.

Todos los materiales están protegidos por la Licencia Creative Commons **CC BY-NC-ND** que permite "descargar las obras y compartirlas con otras personas, siempre que se reconozca su autoría, pero no se pueden cambiar de ninguna manera ni se pueden utilizar comercialmente".

Práctica 4. Relación de Problemas de Cálculo de Probabilidades

Juan de Dios Luna del Castillo y Antonio Martín Andrés

Ejercicio 4.1

Sabiendo que la presión sistólica sigue en individuos normales una $N(127; 24)$:

- a) ¿Entre qué dos valores estará el 95% de los individuos de la población de normales?
- b) ¿Cuál es el percentil 90?

Solución con tablas y calculadora

- a) Si queremos saber dos valores entre los que esté el 95% de la población, fuera de esos dos valores estará el 5% restante. Luego $\alpha = 0.05$. Es decir deseamos un intervalo de dos colas para una $N(127; 24)$ a un nivel de error $\alpha = 0.05$. Mirando en la tabla 1 de la $N(0; 1)$, $z_{0.05} = 1.96$ y sustituyendo en la expresión

$$x \in \mu \pm z_{\alpha} \sigma$$

tendremos

$$x \in 127 \pm (1.96)(24) \rightarrow x \in (127 - (1.96)(24); 127 + (1.96)(24)) \rightarrow x \in (80.0, 174.0)$$

a un nivel de error del 5%.

- b) El percentil 90 es un máximo a un nivel de error $\alpha = 10\%$. La expresión para el Máximo sería:
 $M = \mu + z_{2\alpha} \sigma \rightarrow P_{90} = 127 + (1.282)(24) = 157.77 = 158$

Solución con R

- a) Como nos piden dos valores entre los que estará el 95% de la población eso significa que fuera de ellos estará el 5% de la población. Si además suponemos que el intervalo de valores es simétrico con respecto a 127 (la media), entonces, por debajo del límite inferior de ese intervalo está el 2.5% y por encima del límite superior estará el 2.5% y por debajo de él estará el 97.5%. Luego los dos límites que nos piden son el percentil 2.5 y el 97.5. La función *qnorm* permitirá calcular ambos percentiles como aparece a continuación:

```
percentil2.5=qnorm(0.025, 127, 24, lower.tail = TRUE)
round(percentil2.5, 0)
```

```
## [1] 80
```

```
percentil97.5=qnorm(0.975, 127, 24, lower.tail = TRUE)
round(percentil97.5, 0)
```

```
## [1] 174
```

Lo que da el resultado $x \in (80.0, 174.0)$ a un nivel de error del 5%, idéntico al calculado a mano.

b) Razonando de la misma manera para el percentil 90, obtendremos:

```
percentil90=qnorm(0.90, 127, 24, lower.tail = TRUE)
round(percentil90, 0)
```

```
## [1] 158
```

Que da $P_{90} = 158$, que es el mismo resultado obtenido antes.

Ejercicio 4.2

En las pruebas de acceso a la Facultad de la Actividad Física y el Deporte dos estudiantes obtuvieron la misma puntuación (82,3), pero en dos años distintos. En el primer año la media fue de 75,1 y la desviación típica de 10,3, mientras que en el segundo año la media fue de 81,1 y la desviación típica de 15,4. ¿Cuál de los dos alumnos tiene mejor puntuación?

Solución con R

Solo expresaremos aquí la solución con R puesto que el uso que hacemos es como si fuera una calculadora.

```
mu <- cbind(75.1, 81.1) # Se introducen las dos medias
mu # Se listan las dos medias
```

```
##      [,1] [,2]
## [1,] 75.1 81.1
```

```
sigma <- cbind(10.3, 15.4) # Se introducen las dos desviaciones típicas
sigma # Se listan las dos desviaciones típicas
```

```
##      [,1] [,2]
## [1,] 10.3 15.4
```

```
punt<- cbind(82.3, 82.3) # Se introducen las dos puntuaciones que son la misma
punt # Se listan las dos puntuaciones
```

```
##      [,1] [,2]
## [1,] 82.3 82.3
```

```
tipi=(punt-mu)/sigma # se tipifican las dos puntuaciones
round(tipi,2) # Se sacan los valores tipificados
```

```
##      [,1] [,2]
## [1,]  0.7 0.08
```

El valor tipificado para la puntuación 82.3 el primer año es de 0.70 mientras que en el segundo año sería, 0.08. El primer valor tipificado significa que 82.3 está por encima de la media 0.70 desviaciones típicas, mientras que en el segundo año está por encima de la media 0.08 desviaciones típicas lo que indica que en el primer año el 82.3 es una puntuación mayor que en el segundo.

Ejercicio 4.3

En una determinada población el número de partos al mes sigue una distribución de Poisson de media 36. ¿Entre que dos valores del número de partos se encontrarán el 80% de los meses? Si en un mes se dieran 50 partos, ¿qué se podría decir de ese mes?

Solución con R

Debido a que no hemos dado tablas de una Poisson emplearemos R para calcular las probabilidades y los intervalos que se nos piden.

- a) Partimos de que la variable número de partos en un mes, $x \rightarrow P(\lambda = 36)$. Como $\lambda = 36 > 10$ esa Poisson es aproximable a una $N(36; 6)$. Por tanto resolveremos el problema para la Normal y luego veremos que valores de la Poisson corresponden a los valores de esa Normal y cual es el error real del intervalo. Calculemos con R el intervalo de dos colas para la $N(36; 6)$ que deja en su interior al 80% de los meses. Si deja en su interior al 80%, dejará fuera al 20% y si el intervalo es simétrico con respecto a la media dejará por encima del límite superior a un 10% y por debajo del límite inferior a un 10%. Por tanto tendremos que calcular el percentil 90 de la normal y el percentil 10.

```
percentil90=qnorm(0.90, 36, 6, lower.tail = TRUE)
round(percentil90, 2)
```

```
## [1] 43.69
```

```
percentil10=qnorm(0.10, 36, 6, lower.tail = TRUE)
round(percentil10, 2)
```

```
## [1] 28.31
```

Los valores entre los que están encerrados el 80% de los meses el número de partos son 28.31 y 43.69. Si bien esos valores de una Normal, no son valores de una variable aleatoria de Poisson, que son los que íbamos buscando. Para obtener los valores de la Poisson lo primero que haremos será pasar esos valores de la Normal a valores enteros. Para asegurarse que en el intervalo está al menos un 80%, pasaremos el valor 43.69 a 44 y el valor 28.31 a 28. Pero calculemos ahora mediante la Poisson la probabilidad de que el número de partos esté entre 28 y 44, ambos inclusive.

Los resultados de hacer esos cálculos aparecen a continuación.

```
psup= ppois(44, 36, lower.tail = TRUE)
psup
```

```
## [1] 0.9181346
```

```
pinf= ppois(27, 36, lower.tail = TRUE)
pinf
```

```
## [1] 0.07358283
```

```
pinterior=psup-pinf
pinterior
```

```
## [1] 0.8445518
```

Como se ve,

$$P(28 \leq \mathcal{P} \leq 44) = P(27 < \mathcal{P} \leq 44) = P(\mathcal{P} \leq 44) - P(\mathcal{P} \leq 27) = 0.9181346 - 0.07358283 = 0.8445518$$

Luego los valores que buscábamos eran 28 y 44 y la probabilidad de que la Poisson esté entre ellos es de 0.8445518.

Ejercicio 4.4

Obtenga la función de probabilidad de la variable aleatoria ‘número de caras al lanzar 10 veces una moneda correcta (probabilidad de cara 0.5)’. A la vista de esos datos ¿Qué se puede decir de dicha distribución de probabilidad? ¿Cuánto valen los tres cuartiles de dicha variable aleatoria?

Solución con R

Debido a que no hemos dado tablas de una Binomial emplearemos R para calcular las probabilidades y los percentiles que se nos piden.

En primer lugar crearemos un vector, llamado valores, dónde residirán los valores del 0 al 10. Crearemos a continuación otro vector que denominaremos ‘probab’ que tendrá la probabilidad de una $B(10; 0.5)$ para cada una de esos valores, esas probabilidades se calcularán con la función *dbinom*. Fundiremos esos dos vectores en un *data.frame*, al que llamaré tabla que listaremos y habremos obtenido lo que se nos pide como figura en el cuadro inmediatamente inferior.

```
valores=c(0,1,2,3,4,5,6,7,8,9,10)
valores
```

```
## [1] 0 1 2 3 4 5 6 7 8 9 10
```

```
probab=dbinom(valores,10,0.5)
probab
```

```
## [1] 0.0009765625 0.0097656250 0.0439453125 0.1171875000 0.2050781250
## [6] 0.2460937500 0.2050781250 0.1171875000 0.0439453125 0.0097656250
## [11] 0.0009765625
```

```
tabla=data.frame(valores, probab)
tabla
```

```
##   valores      probab
## 1      0 0.0009765625
## 2      1 0.0097656250
## 3      2 0.0439453125
## 4      3 0.1171875000
## 5      4 0.2050781250
```

```
## 6      5 0.2460937500
## 7      6 0.2050781250
## 8      7 0.1171875000
## 9      8 0.0439453125
## 10     9 0.0097656250
## 11    10 0.0009765625
```

Observada esa tabla lo que podemos decir es que los valores a la misma distancia del 5, por encima o por debajo tienen igual probabilidad, luego esa variable aleatoria es simétrica.

Empleando *qbinom* calcularemos los tres cuartiles como parecen en la siguiente casilla.

```
p25=qbinom(0.25,10,0.5,lower.tail=TRUE)
```

```
p25
```

```
## [1] 4
```

```
p50=qbinom(0.50,10,0.5,lower.tail=TRUE)
```

```
p50
```

```
## [1] 5
```

```
p75=qbinom(0.75,10,0.5,lower.tail=TRUE)
```

```
p75
```

```
## [1] 6
```

Como se ve los tres cuartiles son 4, 5 y 6. Pero veamos cuánto vale la probabilidad de estar entre los cuartiles primero y tercero, $P(4 \leq B \leq 6) = P(B \leq 6) - P(B < 4)$. Esos cálculos los haremos con *pbinom* y aparecen en el siguiente cuadro.

```
pmenoroigual6=pbinom(6,10,0.5, lower.tail = TRUE)
```

```
pmenoroigual6
```

```
## [1] 0.828125
```

```
pmenoroigual3=pbinom(3,10,0.5, lower.tail = TRUE)
```

```
pmenoroigual3
```

```
## [1] 0.171875
```

```
pinterior=pmenoroigual6-pmenoroigual3
```

```
pinterior
```

```
## [1] 0.65625
```

Como se ve la probabilidad de estar entre 4 y 6 no es de un 50%, como ocurriría en las continuas, sino que es mayor, exactamente del 65.625%. Esto es debido a que en las discretas los percentiles se calculan asegurando al menos el porcentaje que se pide y redondeando al entero que desborda ese porcentaje.