PRACTICA 6: VARIABLES ALEATORIAS. Curso 2023/24.

Grado en biología sanitaria. Universidad de Alcalá

1. La probabilidad de que la piel de un guisante sea rugosa es de 3/4 y de que sea lisa es 1/4. Si dispones de 10 semillas, calcula las siguientes probabilidades:

   (a) Que 3 de ellas sean rugosas.

   (b) Que haya 3 o más y 5 o menos lisas.

   Soluciones en la página 3.

2. Por experiencia, se sabe que cada semana se desechan 3 muestras de tejido por mala manipulación. Calcular la probabilidad de

   (a) Que una semana se desechen 2.

   (b) Que una semana se desechen entre 2 y 5.

   (c) Que en dos semanas se desechen menos de 4.

   (d) Que en dos semanas se desechen 4 o más.

   Soluciones en la página 3.

3. La concentración (en ppm) de cierta sustancia en el plasma sanguineo se modeliza con una distribución normal de media 55 y desviación típica 10.

   (a) Calcula la probabilidad de que haya más de 60ppm.

   (b) Has tomado una muestra en una zona en la que sabes que la concentración es, al menos, de 75ppm, ¿cuál es la probabilidad de en la muestra haya una concentración de entre 70ppm y 80ppm?

   (c) Si se decide seleccionar, de entre todas las muestras posibles, el 10% con mayor concentración. ¿Cuál es la concentración mínima para que una muestra pueda ser seleccionada?

   Soluciones en la página 4.

4. Se fumiga una plantación de zanahorias con un producto tóxico. Se sabe que la cantidad de producto que absorbe una zanahoria (en mg) es una variable aleatoria con distribución normal de media 4 y desviación típica 1.5. Se considera que una zanahoria está contaminada si ha absorbido más de 6 mg del producto tóxico:

   (a) Calcula la probabilidad de que una zanahoria seleccionada al azar haya sido contaminada en el proceso de fumigación.

   (b) Si se seleccionan al azar 5 zanahorias, ¿cuál es la probabilidad de que al menos dos de ellas estén contaminadas?

   (c) Si se seleccionan al azar 5 zanahorias, ¿cuál es la probabilidad de que al menos dos de ellas estén contaminadas, sabiendo que hay al menos 1 contaminada?
   Soluciones en la página 5.

5. Dada una variable aleatoria T que sigue una distribución t de Student con 12 grados de libertad, usa las funciones pt() (de distribución) y qt() (cuantiles) para calcular

(a) $P(T < 2)$
(b) $P(-1 < T < 1.5)$
(c) $P(T > 0.5)$
(d) El valor t^* de T tal que $P(T < t^*) = 0.90$
(e) El valor t^* de T tal que $P(T > t^*) = 0.40$

Soluciones en página 6

6. Dada una variable aleatoria W que sigue una distribución Chi cuadrado con 10 grados de libertad, usa las funciones pchisq() (de distribución) y qchisq() (cuantiles) para calcular

   (a) $P(W < 2)$
   (b) $P(1 < W < 4)$
   (c) $P(W > 3.5)$
   (d) El valor w^* de T tal que $P(W < w^*) = 0.87$
   (e) El valor w^* de T tal que $P(W > w^*) = 0.45$

   Soluciones en página 6

7. La función de densidad radial de probabilidad radial (distancia del electrón al núcleo, en picómetros, pm) en el estado fundamental (el de mínima energía) viene dado la expresión

$$f(r) = \frac{4}{a_0^3} r^2 e^{-2r/a_0}$$

   donde a_0 es el radio de Bohr (por ejemplo, para el hidrógeno vale aproximadamente 52.9 pm). Se pide, para el átomo del hidrógeno:

   - Comprueba que f es una función de densidad de probabilidad.
   - Calcula la probabilidad de encontrar el electrón a una distancia de entre 20pm y 100pm del núcleo.

   Soluciones en la página 7. Recuerda que tienes Wolframalpha para que haga las integrales por ti.

8. Una fábrica tiene tres máquinas en funcionamiento, para producir cierto tipo de piezas. Sabemos que:

   - La máquina M1 produce el 20% de las piezas. El peso (en gramos) de las piezas producidas por esta máaquina sigue una distribución uniforme en el intervalo $(15, 25)$.
   - La máquina M2 produce el 45% de las piezas. El peso (en gramos) de las piezas producidas por esta máquina sigue una distribución normal de tipo N(20, 1.5).
   - La máquina M3 produce el resto de las piezas. El peso (en gramos) de las piezas producidas por esta máquina sigue una distribución continua cuya densidad viene dada por la función

$$f(x) = \begin{cases} \frac{3}{500}(x-15)(25-x) & si\ 15 \leq x \leq 25 \\ 0 & en\ otro\ caso \end{cases}$$

Una pieza se considera defectuosa si su peso es mayor de 23 gramos. Sabiendo que una pieza ha resultado defectuosa, ¿cuál es la probabilidad de que proceda de la máquina M2? Soluciones en la página 8.

9. Siete mil corredores participan en una carrera local en la que se pueden clasificar para correr la maratón de la ciudad de New York si completan los 42 km en un tiempo inferior a 3 horas y 10 minutos. Solamente 6350 corredores han terminado la carrera. Si los tiempos empleados en completar los 42 km se distribuyen normalmente con una media de 3 horas 40 minutos y una desviación típica de 28 minutos, se pide:

   (a) ¿Cuál es la probabilidad de que un corredor elegido al azar haya empleado menos de 3 horas en completar la carrera?

   (b) ¿Cuántos corredores se han clasificado para la maratón de Nueva York?

   (c) Si se clasifican más de 800 corredores, la organización aplicará como criterio de selección para acudir a la maratón de Nueva York haber conseguido un tiempo que esté incluido dentro del 5% de los mejores tiempos. ¿Qué tiempo máximo se debe emplear en la carrera local para ser seleccionado para la maratón de Nueva York con este criterio?

   Soluciones en la página 9.

Soluciones.

1. La variable aleatoria número de semillas con la piel rugosa es una $B(10, 3/4)$ y si tiene la piel lisa es una $B(10, 1/4)$. Entonces

   (a)
   ```
   dbinom(3, size = 10, prob = 3/4)

   ## [1] 0.003089905
   ```

   (b) Lo caluclamos de dos formas, con la función de densidad
   ```
   sum(dbinom(3:5, size = 10, prob = 1/4))

   ## [1] 0.4546795
   ```
   o bien con la función de distribución
   ```
   pbinom(5, size = 10, prob = 1/4) - pbinom(2, size = 10, prob = 1/4)

   ## [1] 0.4546795
   ```

   Soluciones en la página 3.

2. Pág 1 La variable X = número de muestras desechadas por semana essigue una distribución de Poisson de parámetro $\lambda = 3$. Entonces

   (a) Hay que calcular $P(X = 2)$
   ```
   dpois(x = 2, lambda = 3)

   ## [1] 0.2240418
   ```

(b) Se pide $P(2 \leq X \leq 5)$.

```
sum(dpois(2:5, lambda = 3))
## [1] 0.7169338
```

o, lo que es lo mismo,

```
ppois(5, lambda = 3)- ppois(1, lambda = 3)
## [1] 0.7169338
```

(c) Ahora, como es cada dos semanas, tenemos $\lambda = 3 * 2 = 6$, y se pide $P(X < 4)$

```
ppois(3, lambda = 6)
## [1] 0.1512039
```

(d) En este caso es $P(X \geq 4)$

```
1 - ppois(3, lambda = 6)
## [1] 0.8487961
```

3. Pág. 1 La cantidad de contaminante X sigue una $N(\mu = 55, \sigma = 10)$

(a) La probabilidad pedida es $P(X > 60)$, que se puede calcular de varias formas diferentes con la función `pnorm`. Como se trata de la cola derecha de la distribución y `pnorm` calcula probabilidades con la cola izquierda (es decir, cosas como $P(X < 60)$) se puede calcular la probabilidad del complementario: $P(X > 60) = 1 - P(X < 60)$

```
1-pnorm(60, mean = 55, sd = 10)
## [1] 0.3085375
```

si quieres hacer saber a `qnorm` que en realidad te interesa la cola derecha, puedes utilizar directamente

```
pnorm(60, mean = 55, sd = 10, lower.tail = FALSE)
## [1] 0.3085375
```

donde `lower.tail = FALSE` le indica a la función `qnorm` que te interesan las probabilidades a la derecha de $X = 60$, y no a su izquierda.

(b) Se trata de una probabilidad condicionada: $P(70 < X < 80|X \geq 75)$. Con la definición de probabilidad condicionada

$$P(70 < X < 80|X \geq 75) = \frac{P((70 < X < 80) \cap (X > 75))}{P(X \geq 75)}$$

los números que están entre 70 y 80 y, a la vez, son mayores que 75 son los números que están entre 75 y 80. Por tanto, la probabilidad pedida es

```r
numerador = pnorm(80, mean = 55, sd = 10) -
            pnorm(75, mean = 55, sd = 10)
denominador = 1-pnorm(75, mean = 55, sd = 10)
numerador/denominador
```

```
## [1] 0.7270493
```

Se usa esa sintaxis (numerador y denominador) para que quepa todo en la página.

(c) En este caso queremos calcular el valor de la concentración de contaminante que deja por encima de sí el 10% de las muestras (y, por debajo, el 90%). Esa cantidad es el percentil 90 y se calcula así:

```r
qnorm(0.9, mean = 55, sd = 10)
```

```
## [1] 67.81552
```

Observa que, de nuevo, existe la posibilidad de usar la cola derecha de la distribución:

```r
qnorm(0.1, mean = 55, sd = 10, lower.tail = FALSE)
```

```
## [1] 67.81552
```

4. Pág. 1

(a) La variable X = cantidad de contaminante absorbida por la zanahoria es

$$X \sim N(4, 1.5)$$

La probabilidad de la zanahoria esté contaminada es $P(X > 6)$

```r
(p = 1 - pnorm(6, mean = 4, sd = 1.5))
```

```
## [1] 0.09121122
```

o bien calcular directamente la cola derecha de la distribución (argumento lower.tail = FALSE)

```r
(p = pnorm(6, mean = 4, sd = 1.5, lower.tail = FALSE))
```

```
## [1] 0.09121122
```

(b) Ahora, la pvariable Y = número de zanahorias contaminadas de entre las 5 de la muestra es

$$Y \sim B(5, p)$$

donde p es la probabilidad de estar contaminada, calculada en el apartado anterior

```r
pbinom(1, size = 5, prob = p, lower.tail = FALSE)
```

```
## [1] 0.06903121
```

(c) Es una probabilidad condicionada

```
pbinom(1, size = 5, prob = p,
   lower.tail = FALSE)/pbinom(0, size = 5, prob = p,
                              lower.tail = FALSE)

## [1] 0.1816086
```

5. Pág. 1

   (a) $P(T < 2)$

   ```
   pt(2, df = 12)

   [1] 0.9656725
   ```

   (b) $P(-1 < T < 1.5)$

   ```
   pt(1.5, df = 12) - pt(-1, df = 12)

   [1] 0.7517467
   ```

   (c) $P(T > 0.5)$

   ```
   1-pt(0.5, df = 12)

   [1] 0.3130587
   ```

   o, alternativamente,

   ```
   pt(0.5, df = 12, lower.tail = F)

   [1] 0.3130587
   ```

   (d) El valor t^* de T tal que $P(T < t^*) = 0.90$

   ```
   qt(0.9, df = 12)

   [1] 1.356217
   ```

   (e) El valor t^* de T tal que $P(T > t^*) = 0.40$. Hay dos formas de calcularlo, o bien teniendo en cuenta que $P(T > t^*) = 0.40$ equivale a $P(T < t^*) = 0.60$

   ```
   qt(0.6, df = 12)

   [1] 0.2590327
   ```

   o bien mirando la cola derecha de la distribución

   ```
   qt(0.4, df = 12, lower.tail = FALSE)

   [1] 0.2590327
   ```

6. Pág. 2

(a) $P(W < 2)$

```
pchisq(2, df = 10)

[1] 0.003659847
```

(b) $P(1 < W < 4)$

```
pchisq(4, df = 10) - pchisq(1, df = 10)

[1] 0.0524809
```

(c) $P(W > 3.5)$

```
1-pchisq(3.5, df = 10)

[1] 0.9670984
```

o, alternativamente,

```
pchisq(3.5, df = 10, lower.tail = F)

[1] 0.9670984
```

(d) El valor w^* de T tal que $P(W < w^*) = 0.87$

```
qchisq(0.87, df = 10)

[1] 15.05693
```

(e) El valor w^* de T tal que $P(W > w^*) = 0.45$. Hay dos formas de calcularlo, o bien teniendo en cuenta que $P(W > w^*) = 0.45$ equivale a $P(W < w^*) = 0.55$

```
qchisq(0.55, df = 10)

[1] 9.892216
```

o bien mirando la cola derecha de la distribución

```
qt(0.45, df = 10, lower.tail = FALSE)

[1] 0.1288902
```

7. Pág. 2

- La función $f(r)$ ya es positiva (porque es una exponencial o cero). Hay que comprobar que

$$\int_0^\infty \frac{4}{a_0^3} r^2 e^{-2r/a_0} dr = 1$$

Para ello, teclea en www.wolframalpha.com

```
integrate 4/(52.9)^3 r^2*exp(-2* r /52.9) dr from 0 to infinity
```

y obtendrás la solución.

- De la misma manera, la probabilidad pedida es

$$\int_{20}^{100} \frac{4}{a_0^3} r^2 e^{-2r/a_0} dr$$

Para ello, teclea en www.wolframalpha.com

`integrate 4/(52.9)^3 r^2*exp(-2* r /52.9) dr from 20 to 100`

para obtener que la probabilidad vale 0.6866

Soluciones en la página 7. Recuerda que tienes Wolframalpha para que haga las integrales por ti.

8. Pág. 2 Si llamamos M1, M2 y M3 a los eventos "la pieza ha sido producida en M1, M2 y M3", espectivamamente, y "D" a "la pieza es defectuosa, lo que se pide es

$$P(M2|D)$$

Conocemos las probabilidades de que la pieza haya sido producida en M1, M2 y M3:

$$P(M1) = 0.2 \qquad P(M2) = 0.45 \qquad P(M3) = 0.35$$

AdemÃ¡s, podemos calcular

$$P(D|M1) \qquad P(D|M2) \qquad P(D|M3)$$

por lo que el problema se resuelve con el teorema de Bayes. Para $P(D|M1)$ hay que calcular $P(X > 23)$ (X es el peso de la pieza) con una distribución uniforme en el intervalo (15, 25)

```
# P(D|M1)
pDM1 = 1 - punif(23, min = 15, max = 25)
```

Para $P(D|M2)$ hay que calcular $P(X > 23)$ con una distribución normal $N(20, 1.5)$

```
# P(D|M2)
pDM2 = 1 - pnorm(23, mean = 20, sd = 1.5)
```

Para $P(D|M3)$ hay que calcular $P(X > 23)$ con la función distribución de masa $f(x)$. Hay que usar, por ejemplo, Wolframalpha:

(a) Ve a http://www.wolframalpha.com
(b) Copia y pega en la barra de búsqueda
`integrate 3/500 (x - 15)(25 - x) from 23 to 25`

El resultado es 13/125, que guardamos en una variable para usar posteriormente

```
pDM3 = 13/125
```

Ahora, aplicando la fórmula del teorema de Bayes se tiene

```r
0.45*pDM2 / (0.2*pDM1 + 0.45*pDM2 + 0.35*pDM3)
```

```
## [1] 0.1181654
```

9. Pág. 3 La variable X = ”tiempo empleado en completar el recorrido” es una normal con media 3 horas y 40 minutos y desviación típica 28 minutos. Lo primero es traducir todo a minutos (o a horas) para que las unidades sean las mismas; usaremos minutos: $X \sim N(220, 28)$

   (a) $P(X < 180)$

   ```r
   pnorm(180, mean = 220, sd = 28)
   ```

   ```
   ## [1] 0.07656373
   ```

   (b) cada corredor debe emplear menos de 190 minutos, por lo que la probabilidad de que necesite, como mucho, ese tiempo es $P(X \leq 190)$

   ```r
   (prob190 = pnorm(190, mean = 220, sd = 28))
   ```

   ```
   ## [1] 0.1419884
   ```

   Nos interesa la variable Y = ”número de corredores que termina la maratón en menos de 190 minutos de entre los 6350 que acabaron la carrera”. Si lo piensas, Y es una variable binomial (cada corredor sólo puede llegar a tiempo o no y corren de forma independiente) de parámetros $n = 6350$ y $p = P(X \leq 190)$. El número esperado de éxitos es np, es decir

   ```r
   6350 * prob190
   ```

   ```
   ## [1] 901.6263
   ```

   es decir,

   ```r
   floor(6350 * prob190)
   ```

   ```
   ## [1] 901
   ```

   (c) Los mejores tiempos son los más rápidos, es decir, se trata del pecentil 5 (en minutos):

   ```r
   qnorm(0.05, mean = 220, sd= 28)
   ```

   ```
   ## [1] 173.9441
   ```