PRACTICA 6: VARIABLES ALEATORIAS.
Curso 2017/18
Grado en biologa sanitaria. Universidad de Alcalá

1. En una binomial $B(7, 1/4)$, calcula las siguientes probabilidades. Se recomienda resolver el ejercicio con las funciones `dbinom` (probabilidad puntual, $P(X = 3) = dbinom(3, size = 7, prob = 1/4)$) y `pbinom` (probabilidad acumulada, $P(X \leq 3) = pbinom(3, size = 7, prob = 1/4)$). Puedes leer la sección 1 del tutorial05 (visita www.postdata-statistics.com).

   (a) $P(X \leq 4)$.
   Con `dbinom` se calcula

   $$P(X \leq 4) = P(X = 0) + P(X = 1) + \cdots = P(X = 4)$$

   ```
   sum(dbinom(0:4, size = 7, prob = 1/4))
   ## [1] 0.9871216
   ```

   Con `pbinom` el cálculo es, directamente

   ```
   pbinom(4, size = 7, prob = 1/4)
   ## [1] 0.9871216
   ```

   (b) $P(X < 3)$.
   (c) $P(X \geq 2)$.
   (d) $P(X > 3)$.
   (e) $P(2 \leq X \leq 4)$ (o, lo que es lo mismo, $P((X \geq 2) \cap (X \leq 4))$).
   (f) $P(2 < X < 4)$.
   (g) $P(2 \leq X < 4)$.
   (h) $P(2 < X \leq 4)$.
   (i) $P(X < 2|X \leq 5)$

2. En los siguientes ejercicios sea X una variable aleatoria de tipo $N(5, 3)$. Calcula las probabilidades y valores que se indican. Puedes leer la sección 5 del tutorial05 (visita www.postdata-statistics.com).

   (a) Calcula $P(X > 6)$, $P(X > 7)$ y, finalmente, $P(X > 10)$. ¿Qué observas?
   (b) Calcula $P(X < 4)$, $P(X < 3)$ y, finalmente, $P(X < 0)$. ¿Ves alguna relación con los valores del anterior apartado? Para entenderlo, puedes hacer un dibujo de la normal con la que estás trabanjando

(c) $P(4.5 < x < 5.5)$, $P(2 < X < 8)$ y, finalmente, $P(0 < X < 10)$.

(d) Los valores k_1 y k_2 tales que $P(X < k_1) = 0.90$ y $P(X < k_2) = 0.95$.

(e) Los valores k_1 y k_2 tales que $P(X > k_1) = 0.1$ y $P(X > k_2) = 0.05$. ¿Ves alguna relación con los valores del anterior apartado?

(f) $P(X > 2|X \leq 6)$

Ahora, con la misma variable X, responde a estas preguntas sin usar el ordenador:

(g) El valor $P(X < 7)$ ¿es mayor o menor que $\frac{1}{2}$?

(h) El valor $P(X > 8)$ ¿es mayor o menor que $\frac{1}{2}$?

(i) ¿Cuál de estos dos valores es más grande, $P(X > 4)$ o $P(X < 5.5)$?

(j) El valor k tal que $P(X > k) = 0.6$ ¿es mayor o menor que 5?

(k) El valor k tal que $P(X < k) = 0.1$ ¿es mayor o menor que 5? Soluciones en la página 6.

3. 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.

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

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

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

   (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?

   ```
   ## [1] 0.1816086
   ```

   (d) Calcula las anteriores probabilidades si la desviación típica de la cantidad de producto que absorve la zanahoria es 0.5 en lugar de 2.5. ¿qué observas?

      i. Probabilidad de que una zanahoria seleccionada al azar haya sido contaminada en el proceso de fumigación.

      ```
      ## [1] 0.2118554
      ```

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

```
## [1] 0.287164
```

iii. 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?

```
## [1] 0.4126567
```

4. La concentración (en ppm) de cierta sustancia en el aire 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 8.

5. Un examen puntua sobre 10 puntos. Se sabe que:

   - El 35% de la población examinada obtuvo una nota superior a 6;
   - El 25% entre 4 y 6, y
   - El 40% inferior a 4.

   Suponiendo que las notas siguen una distribución normal, halla la nota media y la desviación típica. ¿Qué porcentaje de la población tiene una nota que se diferencie de la media en menos de dos unidades? Soluciones en la página 9.

6. 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 distribucióon 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 distribucióon 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

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

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 11.

7. 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 11.

8. Dispones de dos medios de cultivo celular, A y B. Los fabricantes aseguran que la concentración de nutrientes es:

   - en el cultivo A, en promedio, de 10gr/dm^3, con una desviación típica de 0.5gr/dm^3
   - en el cultivo B, en promedio, de 12gr/dm^3, con una desviación típica de 0.6gr/dm^3

   Si mezclas 3dm^3 de A con 5dm^3 de B, determina la concentración esperada y su desviación típica. Soluciones en la página 12.

9. Considera la v.a. continua $X =$ "N de horas de espera en el servicio de urgencias". A partir de una muestra grande se determina que

Minutos espera	N individuos
$[0, 60)$	25
$[60, 120)$	35
$[120, 180)$	20
$[180, 240)$	15
$[240, 300)$	3
$[300, 360)$	2

   (a) Usa las frecuencias relativas para calcular la probabilidad de esperar entre 0 y 60 minutos, entre 60 y 120,...

(b) Calcula la probabilidad de esperar entre 60 y 180 minutos.

(c) Calcula la probabilidad de esperar entre 60 y 150 minutos.

Soluciones en la página 12.

1. 1, pág. 1

   Fijamos los valores de n y p en este ejercicio:

```
n = 7
p = 1/4
```

   y empezamos con los cálculos.

   (a) Es directo:

```
pbinom(4, size = n, prob = p)

## [1] 0.9871216
```

   (b) $P(X < 3) = P(X \leq 2)$, luego:

```
pbinom(2, size = n, prob = p)

## [1] 0.7564087
```

   (c) $P(X \geq 2) = 1 - P(X < 2) = 1 - P(X \leq 1)$. Cuidado en el segundo paso!

```
1- pbinom(1, size = n, prob = p)

## [1] 0.5550537
```

   De otra manera, usando `dbinom` y sumando:

```
sum(dbinom(2:n, size = n, prob = p))

## [1] 0.5550537
```

   (d) $P(X > 3) = 1 - P(X \leq 3)$

```
1- pbinom(3, size = n, prob = p)

## [1] 0.07055664
```

   (e) Podemos usar que $P(2 \leq X \leq 4) = P(X \leq 4) - P(x \leq 1)$

```
pbinom(4, size = n, prob = p) - pbinom(1, size = n, prob = p)

## [1] 0.5421753
```

   Compruébalo sumando valores de `dbinom`.

(f) Ahora usamos $P(2 < X < 4) = P(X \leq 3) - P(x \leq 2) = P(X = 3)$

```
pbinom(3, size = n, prob = p) - pbinom(2, size = n, prob = p)

## [1] 0.1730347

dbinom(3, size = n, prob = p)

## [1] 0.1730347
```

El resultado es el mismo, claro.

(g) $P(2 \leq X < 4) = P(X \leq 3) - P(x \leq 1)$

```
pbinom(3, size = n, prob = p) - pbinom(1, size = n, prob = p)

## [1] 0.4844971
```

(h) $P(2 < X \leq 4) = P(X \leq 4) - P(x \leq 2)$

```
pbinom(4, size = n, prob = p) - pbinom(2, size = n, prob = p)

## [1] 0.2307129
```

(i) $P(X < 2 | X \leq 5) = \frac{P((X<2)\cap(X\leq 5))}{p(x\leq 5)} = \frac{P(X<2)}{p(x\leq 5)}$

```
pbinom(1, size = n, prob = p)/pbinom(5, size = n, prob = p)

## [1] 0.4455446
```

2. 2, pág. 1

Todas las respuestas son aproximadas. Definir los parámetros

```
mu = 5
sigma = 3
```

(a) $P(X > 6) = 1 - P(X < 6)$

```
1-pnorm(6, mean = mu, sd =sigma)

## [1] 0.3694413
```

También sepuede hacer el cálculo "directo": observa el uso del argumento `lower.tail` (su valor por defecto es `FALSE`)

```
1 - pnorm(7, mean = mu, sd =sigma)

## [1] 0.2524925

pnorm(7, mean = mu, sd =sigma, lower.tail = FALSE)

## [1] 0.2524925
```

```r
pnorm(10, mean = mu, sd =sigma, lower.tail = FALSE)
```

```
## [1] 0.04779035
```

A medida que nos desplazamos hacia la derecha los valores van siendo más y más pequeños.

(b) En este caso:

```r
pnorm(4, mean = mu, sd =sigma)
```

```
## [1] 0.3694413
```

```r
pnorm(3, mean = mu, sd =sigma)
```

```
## [1] 0.2524925
```

```r
pnorm(0, mean = mu, sd =sigma)
```

```
## [1] 0.04779035
```

Las respuestas son las mismas del anterior apartado, por la simetría de la curva normal respecto de la media, que está en $\mu = 5$. Por ejemplo, los valores 3 y 7 están a la misma distancia, a izquierda y derecha de μ, respectivamente, y por eso

$$P(X < 3) = P(X > 7) = 0.2525$$

(c) El cálculo es

```r
pnorm(5.5, mean = mu, sd = sigma)-pnorm(4.5, mean = mu, sd = sigma)
```

```
## [1] 0.1323677
```

Observa que también se puede calcular así

```r
(pnorm(8, mean = mu, sd = sigma)-0.5)*2
```

```
## [1] 0.6826895
```

```r
(pnorm(10, mean = mu, sd = sigma)-0.5)*2
```

```
## [1] 0.9044193
```

(d) Se trata de los percentiles 90 y 95, y la sintaxis vertorial permite calcularlos de una vez

```r
qnorm(c(0.9, 0.95), mean = mu, sd = sigma)
```

```
## [1] 8.844655 9.934561
```

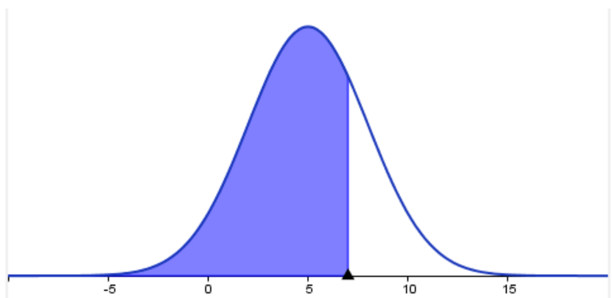
(e) Los mismos valores de k_1 y k_2, de nuevo por la simetría de la curva normal.

(f) $P(X > 2|X \leq 6) = \frac{P((X>2)\cap(X\leq6))}{p(x\leq6)}$

```
numerador2 = pnorm(6, mean = mu, sd = sigma)-
             pnorm(2, mean = mu, sd = sigma)
denominador2 = pnorm(6, mean = mu, sd = sigma)
numerador2/denominador2
```

```
## [1] 0.7483894
```

(g) Es mayor que 1/2. Se tiene $P(X < 7) = 0.7475$. Este apartado y el siguiente se resuelven teniendo en cuenta que el área de cada una de las dos mitades de la curva es $\frac{1}{2}$, observando si el punto que hemos tomado está a la derecha o la izquierda de μ, y si la probabilidad que calculamos incluye todos los valores mayores o menores. Por ejemplo, para la primera pregunta tenemos que pensar en un dibujo como este:



con el que resulta evidente que la respuesta es mayor que $\frac{1}{2}$.

(h) Es menor que 1/2. Se tiene $P(X > 8) = 0.1587$.

(i) El valor $P(X > 4)$ es más grande. Se tiene $P(X > 4) = 0.6306$, mientras que $P(X < 5.5) = 0.5662$. En este caso la respuesta es fácil de ver porque el valor 4 está más lejos de $\mu = 5$ que 5.5.

(j) Las ideas para este apartado y el siguiente son las mismas, pero ahora tenemos que pensar en los valores del eje x, en vez de pensar en las probabilidades (áreas) que definen esos valores. El valor tiene que ser menor que 5 (de otra manera, la probabilidad sería menor que $\frac{1}{2}$). Se obtiene $k = 4.24$

(k) El valor tiene que ser menor que 5. Se obtiene $k = 1.156$.

3. 4, pág. 3 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 al cola izquierda (es decir, cosas como $P(X < 60)$) se puede calcular la probabilidad del complementario: $P(X > 60) = 1 - P(X > 60)$

```r
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

```r
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(70, mean = 55, sd = 10)
numerador/denominador
```

```
## [1] 0.2475851
```

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 distribucin:

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

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

4. pág. 3 El enunciado indica que X=”nota sobre 10 puntos” se comporta como $X \sim B(\mu, \sigma)$ y que

- $P(X > 6) = 0.35$
- $P(4 < X < 6) = 0.25$
- $P(X < 4) = 0.4$

Para determinar μ y σ deben aparecer en alguna expresión relacionada con los datos del problema. Una opción es estandatizar: por un lado se involucra a μ y a σ y, por otro lado, con una normal $N(0, 1)$ podemos hacer cálculos, ya que conoecmos su media y su desviación estándar. Como, además, hay dos incógnitas, son necesarias dos ecuaciones. Usaremos las expresiones más sencillas:

- $0.4 = P(X < 4) = P\left(\frac{X-\mu}{\sigma} < \frac{4-mu}{\sigma}\right)$
- $0.35 = P(X > 6) = P\left(\frac{X-\mu}{\sigma} > \frac{6-mu}{\sigma}\right)$. Esto es equivalente, por simetría, a $0.65 = P(X < 6) = P\left(\frac{X-\mu}{\sigma} < \frac{6-\mu}{\sigma}\right)$.

A poco que lo pienses, lo que dicen las expresiones anteriores es que $\frac{4-\mu}{\sigma}$ y $\frac{6-\mu}{\sigma}$ son, respectivamente, los percentiles 40 y 65 de una normal $N(0, 1)$, que calculamos a continuación con cuatro cifras significativas

```
(percentiles = signif(qnorm(c(.25, .65)), digits = 4))

## [1] -0.6745  0.3853
```

es decir:

$$-0.6745 = \frac{4-\mu}{\sigma} \qquad 0.3853 = \frac{6-\mu}{\sigma}$$

de donde se tiene

$$\left\{ \begin{array}{l} \mu - 0.6745\sigma = 4 \\ \mu + 0.3853\sigma = 6 \end{array} \right.$$

Restar a la segunda ecuación la primera

$$\left\{ \begin{array}{l} \mu - 0.6745\sigma = 4 \\ 1.0598\sigma = 4 \end{array} \right. \Leftrightarrow \left\{ \begin{array}{l} \mu - 0.6745\sigma = 4 \\ 1.0598\sigma = 2 \end{array} \right.$$

de donde se obtiene que σ vale

```
## [1] 1.887149
```

y μ vale

```
## [1] 5.272882
```

5. pág. 3 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:

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

Adems, 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 distribucin 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 distribucin 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 funcin distribucin 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

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

## [1] 0.1181654
```

6. pág. 4 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) $P(X < 180)$

```
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 $P(X \leq 190)$

```
(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 sĺo 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

```
6350 * prob190

## [1] 901.6263
```

es decir,

```
floor(6350 * prob190)

## [1] 901
```

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

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

## [1] 173.9441
```

7. pág. 4

La idea es que la mezcla está hecha a partir de medios de cultivo de los que conocemos la concentración media y la varianza y, además, sabemos qué cantidad se ha utilizado de cada medio. La variable aleatoria X = "cantidad de nutrientes en la mezcla" se define como

$$X = 10\frac{gr}{dm^3} * 3dm^3 + 12\frac{gr}{dm^3} * 5dm^3$$

8. pág. 4