Distribuciones de probabilidad

Antonio Falcó

Seminario 5

Resumen del Seminario 4

- Sea Ω una población experimental y $\mathbb{P}$ una probabilidad definida sobre la misma.
- Sea X una variable aleatoria definida sobre Ω.
- Sea F la función de distribución para X en $\mathbb{P}$.

Con los datos obtenidos mediante observaciones de una variable de interés X sobre una muestra Ω_n de la población experimental Ω construimos la función de distribución empírica

$$F_n(x) = \mathbb{P}(x \leq X|\Omega_n) \approx \mathbb{P}(x \leq X) = F(x),$$

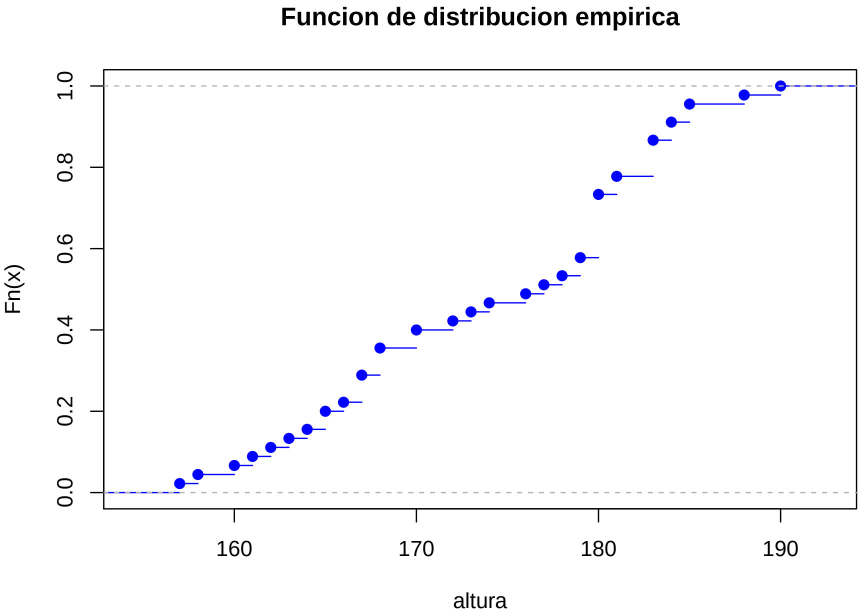
como aproximación de la función de distribución poblacional $F(x) = F(X \leq x)$.

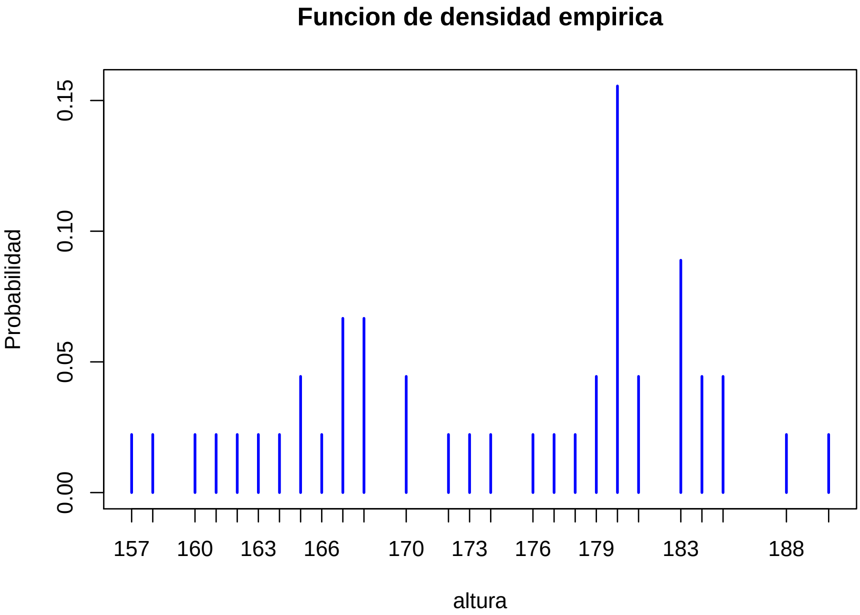
Empleamos las observaciones obtenidas de una muestra Ω_1 de la talla y el peso de 45 individuos de una población Ω :

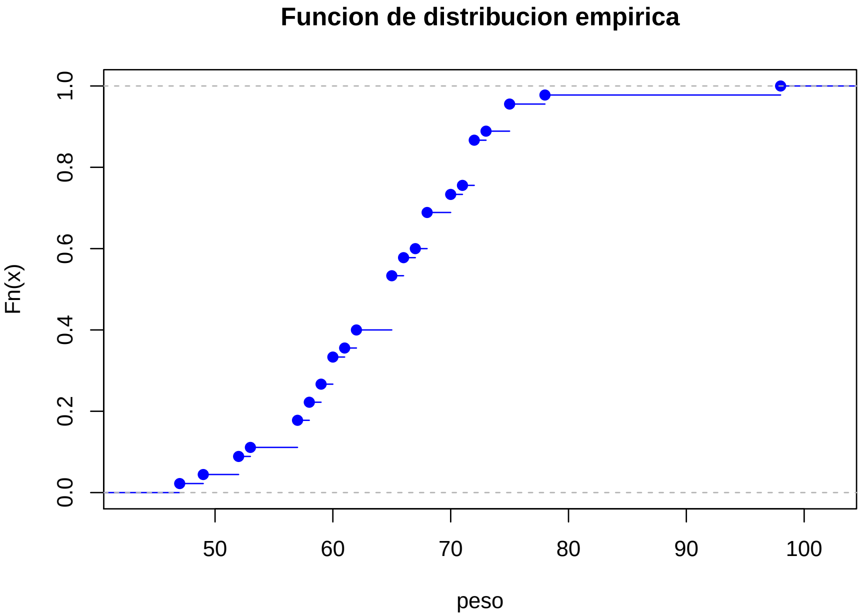
	altura	peso
1	180	70
2	177	57
3	180	60
4	180	66
5	183	62
6	184	68
7	185	65
8	184	72
9	174	65
10	180	72
11	168	52
12	180	75
13	183	75
14	181	68
15	180	65

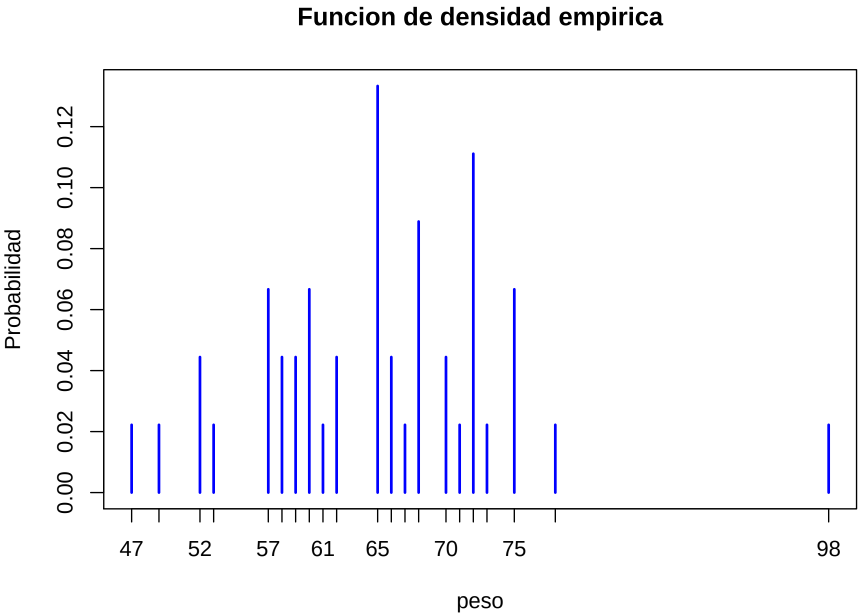
	altura	peso
16	190	66
17	183	78
18	167	60
19	181	67
20	179	98
21	173	75
22	170	68
23	170	59
24	183	72
25	179	73
26	180	72
27	188	70
28	176	65
29	178	72
30	185	71

	altura	peso
31	168	52
32	157	47
33	167	53
34	168	57
35	163	65
36	167	60
37	166	68
38	164	49
39	172	57
40	165	59
41	158	62
42	161	65
43	160	61
44	162	58
45	165	58


Funcion de distribucion empirica
Fn(x)
0.0
0.2
0.4
0.6
0.8
1.0
160
170
180
190
altura


Funcion de densidad empirica
Probabilidad
0.00
0.05
0.10
0.15
157
160
163
166
170
173
176
179
183
188
altura


Funcion de distribucion empirica
Fn(x)
0.0
0.2
0.4
0.6
0.8
1.0
50
60
70
80
90
100
peso


Funcion de densidad empirica
Probabilidad
0.00
0.02
0.04
0.06
0.08
0.10
0.12
47
52
57
61
65
70
75
98
peso

Relación entre función de distribución y función de densidad

Conocemos que por ejemplo,

$$\mathbb{P}(\texttt{altura} = 183|\Omega_n) = 0.0888889.$$

Sin embargo si consideramos que la variable `altura` es una variable aleatoria continua (por ejemplo, con una distribución normal) entonces la teoría nos dice que

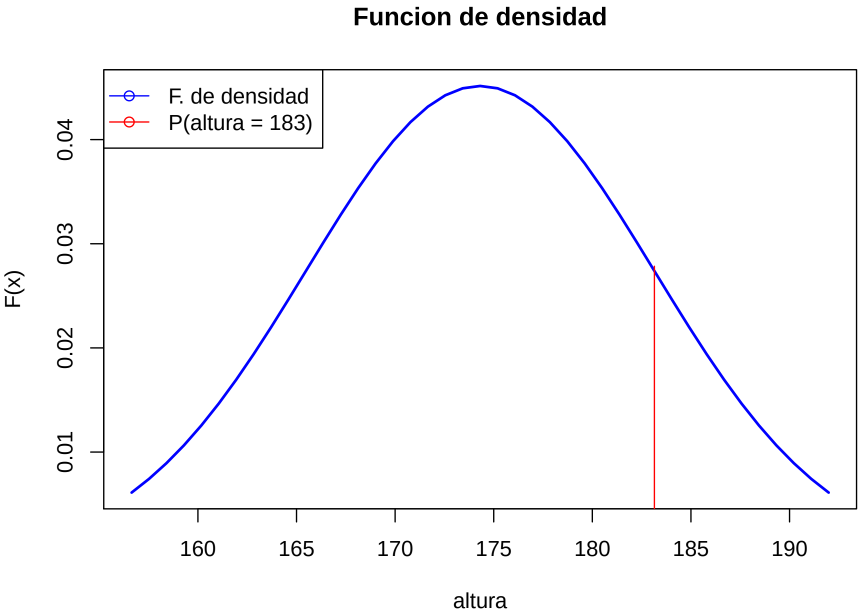
$$\mathbb{P}(\texttt{altura} = 183) = 0.$$

Esto se debe a que si $X \sim \mathcal{N}(\mu, \sigma)$ entonces

$$\mathbb{P}(a < X \leq b) = \text{ área debajo de la curva entre } a \text{ y } b \text{ },$$

en consecuencia

$$\mathbb{P}(X = x) = \text{ área de un segmento } = 0.$$


Funcion de densidad
F. de densidad
P(altura = 183)
0.01
0.02
0.03
0.04
F(x)
160
165
170
175
180
185
190
altura

Cuestión

¿Qué relación existe entre la función de densidad y su aproximación la función de densidad empírica? ¿Qué es mejor para caracterizar un estudio estadístico la función de distribución empírica o la función de densidad empírica?

Cuestión científica

¿Frecuencias acumuladas o Frecuencias relativas? Como implementar **técnicas de aprendizaje estadístico** en la práctica.

La función de distribución y la función de densidad I

Si X sigue una distribución $\mathcal{N}(\mu, \sigma)$ entonces la tabla de la distribución normal se calcula empleando la expresión de su función de distribución:

$$F(x; \mu, \sigma) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} \, dy.$$

Entonces, se cumple (Teorema fundamental del cálculo)

$$\frac{d}{dx} F(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

A la función

$$f(y; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} = f_X(y)$$

(la campana de Gauss) se le llama **función de densidad de la distribución normal** $\mathcal{N}(\mu, \sigma)$.

La función de distribución y la función de densidad II

Si X sigue una distribución binomia $\mathcal{B}(n,p)$ entonces la tabla de la distribución binomial se calcula empleando la expresión de su función de distribución:

$$F(x;n,p) = \mathbb{P}(X \leq x) = \sum_{y \leq x} \binom{n}{y} p^y (1-p)^{n-y}$$

donde y toma exclusivamente los valores $0, 1, 2, \ldots, n$. Entonces a la función

$$f(y;n,p) = \mathbb{P}(X = y) = \binom{n}{y} p^y (1-p)^{n-y} = f_X(y)$$

se le llama **función de densidad de la distribución binomial** $\mathcal{B}(n,p)$.

Una versión intuitiva de estas expresiones matemáticas

La función de distribución de una variable continua se puede representar como una suma acumulada:

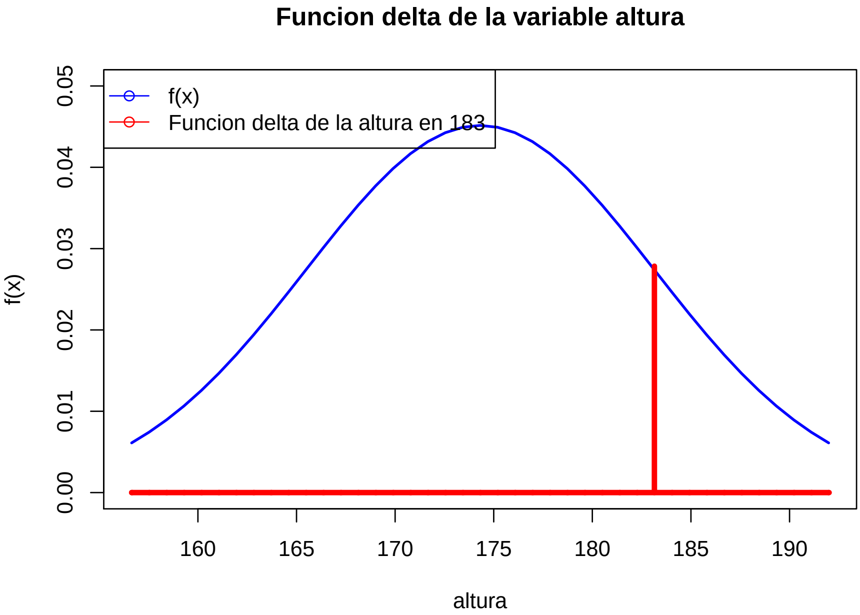
$$F(x) = \mathbb{P}(X \leq x) = \sum_{y \leq x} f_X(y) = \begin{array}{l} \text{suma de todas las frecuencias} \\ \text{relativas del suceso } \{X = y\} \\ \text{para valores de } y \text{ no superiores a } x, \end{array}$$

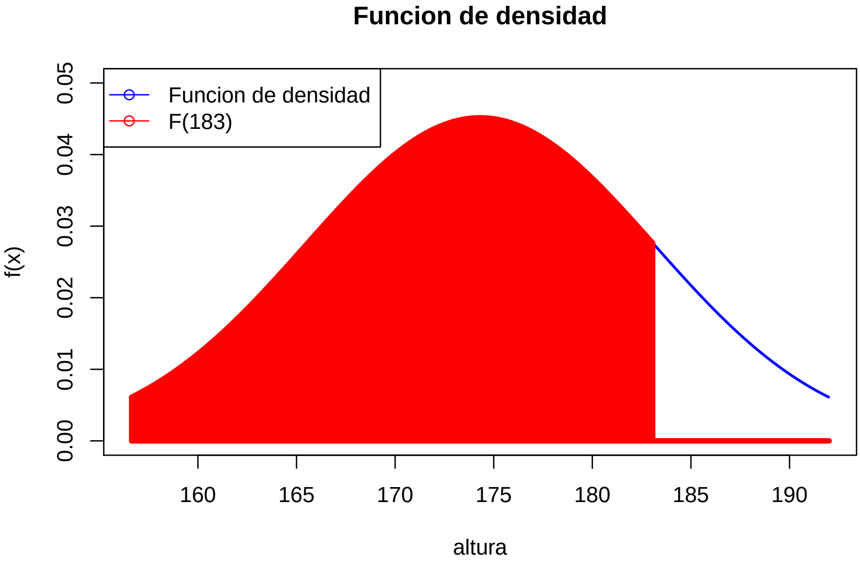
(cuando la variable es continua $\sum_{y \leq x} = \int_{-\infty}^{x} dx$) donde utilizamos para escribirla la notación:

$$f_X(y) := \textbf{ frecuencia relativa del suceso } \{X = y\}\ .$$

Cada frecuencia relativa $f_X(y)$ la podemos representar mediante una 'función' llamada

$$\textbf{[delta de } f_X(y)\textbf{]}(x) = \begin{cases} f_X(y) & \text{si} \quad x = y \\ 0 & \text{si} \quad x \neq y. \end{cases}$$

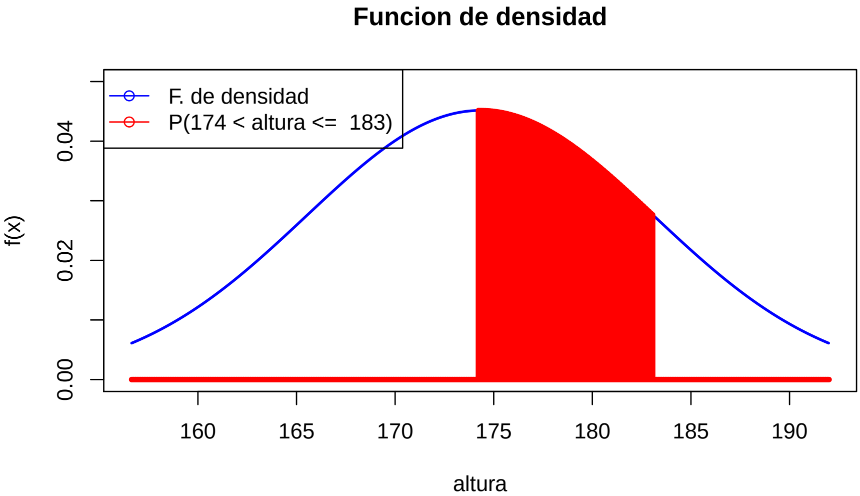

Funcion delta de la variable altura
f(x)
Funcion delta de la altura en 183
0.00
0.01
0.02
0.03
0.04
0.05
f(x)
160
165
170
175
180
185
190
altura



$F(183)$ **se calcula como la suma de todas las funciones deltas de la altura, para alturas no superiores a 183 cm**, en color rojo en la gráfica.

El concepto de área y probabilidad

$$\mathbb{P}(a < X \leq b) = \sum_{a<y\leq b} f_X(y) = \int_a^b f_X(y)dy.$$



Discusión

- Esta notación introduce coherencia en la definición de observación X asociada a una población experimental Ω dotada de una probabilidad $\mathbb{P}$.
- Si suponemos que X tiene una función de distribución F y una función de densidad f que cumple

$$f_X(y) = \text{ frecuencia relativa del suceso } \{X = y\}$$

entonces

$$F(x) = \mathbb{P}(X \leq x) = \sum_{y \leq x} f_X(y)$$

- Esta formulación es válida sea la variable X discreta o continua, y es coherente con la toma de una muestra Ω_n, entonces

$$F_n(x) = \mathbb{P}(X \leq x|\Omega_n) = \sum_{y \leq x} f_X(y|\Omega_n), \text{ de forma que}$$

$$F(x) = F_n(x) + \sum_{y \leq x} f_X(y|\overline{\Omega}_n) = F_n(x) + \text{Error en la muestra } \Omega_n$$

Ejemplo

Si tomamos la variable `altura` sobre la muestra de 45-individuos entonces

$$f_{\texttt{altura}}(y|\Omega_n) = \text{ frecuencia relativa de } y \text{ en } \Omega_n$$

los posibles valores para y y su correspondiente frecuencia relativa $f_{\texttt{altura}}(y|\Omega_n)$ sobre la muestra son

```
table(altura)/length(altura)

## altura
##        157        158        160        161        162        163        164
## 0.02222222 0.02222222 0.02222222 0.02222222 0.02222222 0.02222222 0.02222222
##        165        166        167        168        170        172        173
## 0.04444444 0.02222222 0.06666667 0.06666667 0.04444444 0.02222222 0.02222222
##        174        176        177        178        179        180        181
## 0.02222222 0.02222222 0.02222222 0.02222222 0.04444444 0.15555556 0.04444444
##        183        184        185        188        190
## 0.08888889 0.04444444 0.04444444 0.02222222 0.02222222
```

en particular para $y = 157$ tenemos $f_{\texttt{altura}}(157|\Omega_n) = 0.0222222$.

Definición

- Para estudiar una medida X (variable aleatoria) definida sobre una población experimental Ω con una probabilidad $\mathbb{P}$,
- Consideramos las frecuencias relativas de X sobre Ω :

$$\{f_X(x) : x = X(\omega) \text{ para } \omega \text{ individuo de } \Omega\}$$

  - Si X es una variable aleatoria discreta, entonces $f_X(x)$ es una función continua.
  - Si X es una variable discreta entonces X toma valores en un conjunto discreto de números distintos (las observaciones numéricas de la variable X)

$$x_1, x_2, \ldots, x_n, \ldots$$

    de forma que cada uno de ellos tiene una frecuencia relativa

$$f_X(x_1), f_X(x_2), \ldots, f_X(x_n), \ldots.$$

    En este caso, se suele escribir $p_i = \mathbb{P}(X = x_i) = f_X(x_i)$.
- Se cumple que $\sum_{x \in X(\Omega)} f_X(x) = 1 = \int_{-\infty}^{\infty} f_X(x)dx$.

Definición (media o valor esperado)

- Si X es una variable aleatoria definida sobre Ω con una función de densidad $f_X(x)$ entonces se define la **media o valor esperado de** X como

$$\mu = \mathbb{E}(X) = \sum_{x \in X(\Omega)} x \cdot f_X(x)$$

- Si X es una variable aleatoria definida sobre Ω con una función de densidad $f_X(x)$ y λ es un escalar, entonces se define la media de λX como

$$\mathbb{E}(\lambda X) = \sum_{x \in X(\Omega)} \lambda x \cdot f_X(x) = \lambda \sum_{x \in X(\Omega)} x \cdot f_X(x) = \lambda \mathbb{E}(X)$$

- Si X, Y son dos variable aleatorias definidas sobre Ω con funciónes de densidad $f_X(x)$ y $f_Y(y)$ respectivamente, entonces se define la media de $X + Y$ como

$$\mathbb{E}(X + Y) = \sum_{x \in X(\Omega)} x \cdot f_X(x) + \sum_{y \in Y(\Omega)} y \cdot f_Y(y)$$

Definición (varianza)

Si X es una variable aleatoria definida sobre Ω con una función de densidad $f_X(x)$ y media $\mu = \mathbb{E}(X) = \sum_{x \in X(\Omega)} x f_X(x)$, entonces se define la **varianza** de X como

$$\sigma^2 = \mathrm{Var}(X) = \mathbb{E}((X-\mu)^2) = \sum_{x \in X(\Omega)} (x-\mu)^2 \cdot f_X(x)$$

Recordemos la fórmula $(x-\mu)^2 = x^2 - 2\mu x + \mu^2$, entonces

$$\begin{aligned}
\mathrm{Var}(X) = \mathbb{E}((X-\mu)^2) &= \sum_{x \in X(\Omega)} (x-\mu)^2 \cdot f_X(x) \\
&= \sum_{x \in X(\Omega)} (x^2 - 2\mu x + \mu^2) f_X(x) \\
&= \sum_{x \in X(\Omega)} x^2 f_X(x) - 2\mu \sum_{x \in X(\Omega)} x f_X(x) + \mu^2 \sum_{x \in X(\Omega)} f_X(x) \\
&= \sum_{x \in X(\Omega)} x^2 f_X(x) - \mu^2 = \mathbb{E}(X^2) - \mu^2.
\end{aligned}$$

$\sigma = \sqrt{\mathrm{Var}(X)}$ se le llama **desviación típica**.

Ejemplos

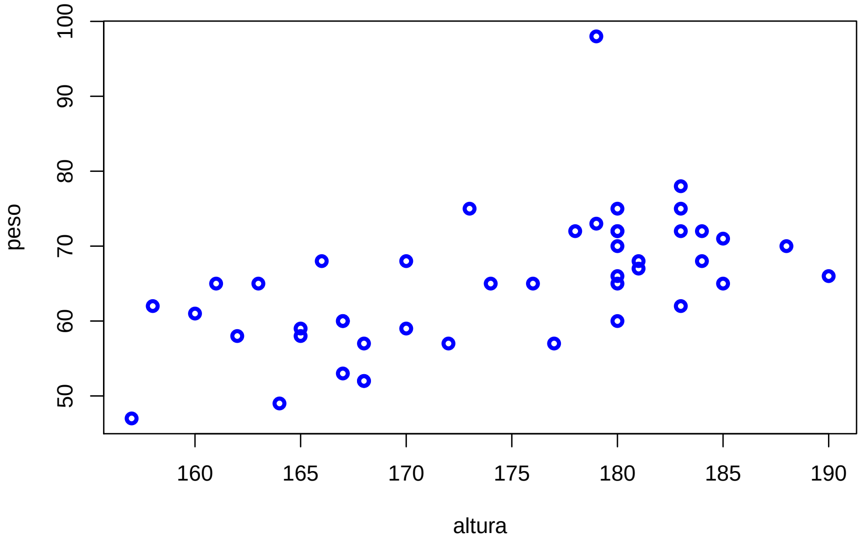
- $\mathbb{E}(\texttt{altura}) = \texttt{mean(altura)} = 174.3111111$ cm,
- $\mathrm{Var}(\texttt{altura}) = \texttt{var(altura)} = 78.0828283$ cm^2,
- $\sqrt{\mathrm{Var}(\texttt{altura})} = \texttt{sd(altura)} = 8.8364489$ cm,
- $\mathbb{E}(\texttt{peso}) = \texttt{mean(peso)} = 65.0888889$ kg,
- $\mathrm{Var}(\texttt{peso}) = \texttt{var(peso)} = 80.1737374$, kg^2.
- $\sqrt{\mathrm{Var}(\texttt{peso})} = \texttt{sd(peso)} = 8.9539789$ kg,

Cuestión científica

Las variables altura y peso están definidas sobre la misma población experimental, podemos entonces representar las dos variables de forma conjunta para cada individuo de la poblacion:

$$\omega \mapsto (\texttt{altura}(\omega), \texttt{peso}(\omega)).$$

¿Podemos afirmar que a mayor altura los individuos de la población tienen un mayor peso?


peso
100
90
80
70
60
50
160
165
170
175
180
185
190
altura