

Cuestiones Seminarios 4 y 5

Bioestadística

06/05/2020

Cuestiones a responder en el cuaderno de campo

Observaciones y medición - Distribuciones de probabilidad

Consideremos la tabla de datos siguientes que contiene el peso y la altura de una muestra de 45 individuos de una población

	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

Cuestión a responder en el cuaderno de campo:

Vamos a estudiar la distribución de la variable peso para determinar si podemos asumir que sigue una distribución normal para ello calculamos su media

$$\mathbb{E}(\text{peso}) = 65.0888889 \text{ kg},$$

y su varianza

$$\mathbb{E}((\text{peso} - 65.0888889)^2) = 80.1737374 \text{ kg}^2.$$

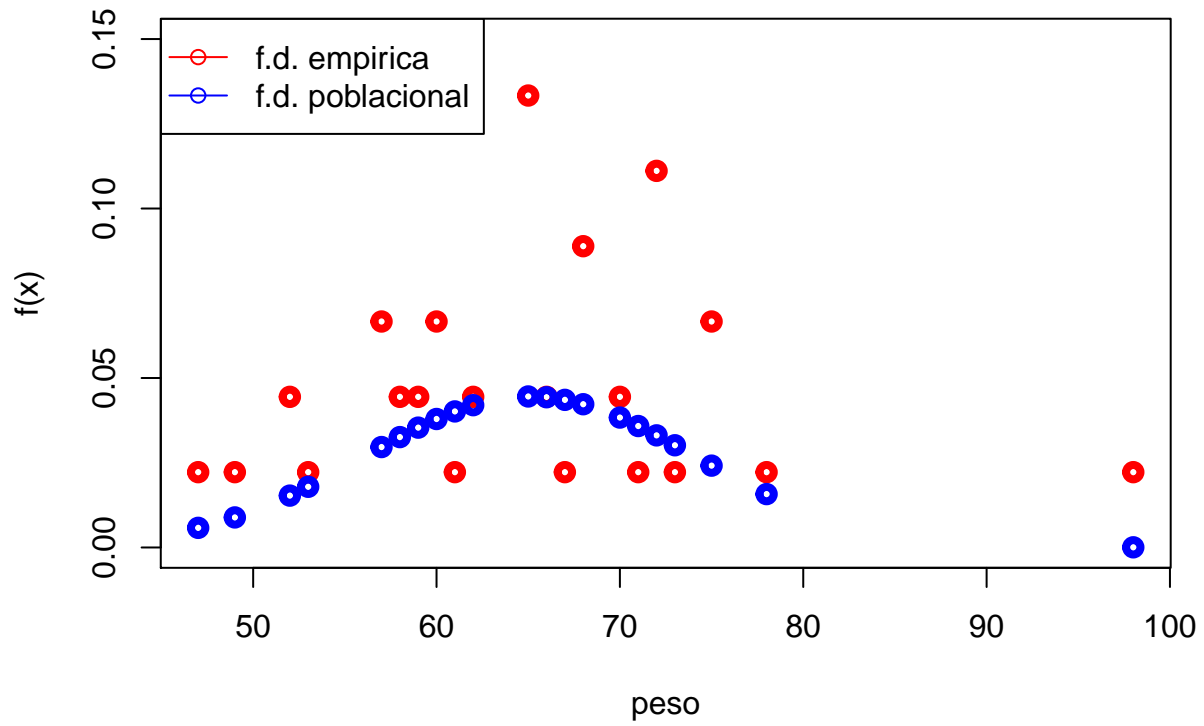
Podemos entonces suponer que la variable peso sigue una distribución normal de media 65.0888889 kg y desviación típica 8.9539789 kg. Tenemos que tener en cuenta que la variable peso toma los valores siguientes en muestra:

$$\text{peso}(\Omega_n) = \{47, 49, 52, 53, 57, 58, 59, 60, 61, 62, 65, 66, 67, 68, 70, 71, 72, 73, 75, 78, 98\}$$

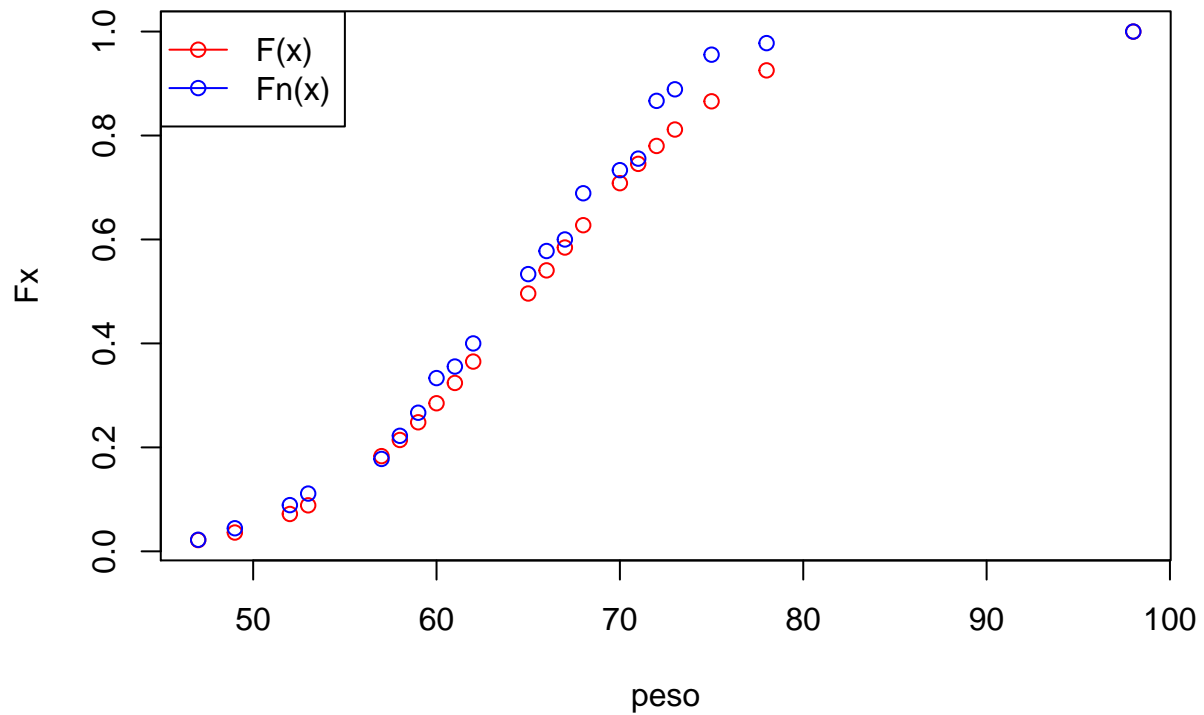
Para cada uno de estos 21 valores calcularemos su frecuencia relativa empírica $f_{\text{altura}}(y_i|\Omega_n)$, su frecuencia relativa poblacional $f_{\text{altura}}(y_i)$ y del mismo modo los valores para función de distribución empírica $F_n(y_i)$ y la correspondiente poblacional $F(y_i)$ para y_i tomando valores en el conjunto

$$\text{peso}(\Omega_n) = \{47, 49, 52, 53, 57, 58, 59, 60, 61, 62, 65, 66, 67, 68, 70, 71, 72, 73, 75, 78, 98\}$$

Comparamos en primer lugar las funciones de densidad de la normal f_{peso} y la función de densidad empírica $f_{\text{peso}}(\cdot|\Omega_n)$ sobre los valores de $\text{peso}(\Omega_n)$ obteniendo la gráfica siguiente:



Ahora compararemos las funciones de distribución empírica y poblacional



Finalmente, podemos calcular en ambos casos los errores relativos medios cometidos:

$$\frac{1}{21} \sum_{i=1}^{21} \frac{|F(y_i) - F_n(y_i)|}{F(y_i)} = 0.0920817$$

entre las funciones de distribución y

$$\frac{1}{21} \sum_{i=1}^{21} \frac{|f(y_i) - f(y_i|\Omega_n)|}{f(y_i)} = 21.2208177$$

entre las funciones de densidad.

- A la vista de los dos gráficos anteriores ¿tenemos suficientes argumentos para justificar que la variable peso sigue una distribución normal?
- ¿Qué conclusiones podemos obtener de este ejemplo práctico?