

Práctica 2. Exploración de datos II. Soluciones.

Universidad de Alcalá. Curso 2023-24.

Estadística (650008). Grado en Biología sanitaria.

Actualizado: 2023-10-05

Introducción

Trabajaremos con una tabla de datos procedente de una muestra de 1000 mujeres que participaron en un estudio sobre osteoporosis. El fichero contiene algunas variables auxiliares en las columnas iniciales, pero nosotros nos vamos a fijar en estas.

- edad (en años).
- peso (en kg).
- talla (altura en cm).
- imc (índice de masa corporal)
- bua (resultado de la exploración densitométrica)
- clasific (normal / osteopenia / osteoporosis)
- menarqui (edad primera menstruación, en años)
- edad_menop (edad inicio menopausia, en años)
- menopausia (sí, no)
- tipo de menopausia
- nivel educativo

Organiza tu entorno de trabajo

Empieza por descargar la tabla de datos desde aquí y guardarlos en la variable `osteo` (para que coincida con las soluciones). Si no recuerdas como hacerlo, se explica en los vídeos de trabajo previo a esta práctica.

Ejercicio 1

Calcula el peso medio de las 100 primeras mujeres de la tabla.

Literalmente:

```
mean(osteo$peso[1:100])
```

```
## [1] 67.178
```

Observa que, desglosando lael comando anterior

```
# Valores del 1 al 100,  
1:100
```

```
## [1] 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18  
## [19] 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36  
## [37] 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54  
## [55] 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72  
## [73] 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90  
## [91] 91 92 93 94 95 96 97 98 99 100
```

```
# para indicar las posiciones dentro del vector osteo$peso.
```

```
# Los pesos de las 100 primeras entradas de la tabla son  
osteo$peso[1:100]
```

```
## [1] 70.0 53.0 64.0 78.0 56.0 63.5 86.0 61.5 60.5 64.0 70.3 74.4  
## [13] 55.5 89.0 50.6 71.4 78.0 72.0 68.0 75.0 66.5 101.0 66.5 70.0  
## [25] 60.1 67.0 67.0 70.5 66.5 58.5 60.0 70.0 68.0 63.0 64.0 75.0  
## [37] 74.0 64.0 71.0 67.0 63.0 60.0 53.0 55.0 77.0 57.0 82.0 75.0  
## [49] 58.0 68.5 68.2 57.5 72.8 77.0 74.0 60.0 57.0 59.5 60.5 66.0  
## [61] 58.8 63.7 81.0 70.0 54.5 85.0 74.0 46.0 79.0 58.0 62.0 69.0  
## [73] 46.5 77.5 69.0 65.5 75.5 68.0 75.0 67.5 67.5 66.5 64.5 70.0  
## [85] 74.0 55.0 74.5 53.0 68.0 80.0 59.5 80.0 75.0 58.0 74.0 82.5  
## [97] 55.0 56.0 77.0 51.0
```

```
# y su media se calcula con  
mean(osteo$peso[1:100])
```

```
## [1] 67.178
```

Ejercicio 2

Calcula el peso medio de las 100 primeras mujeres con estudios SUPERIORES

Ahora la selección de los valores de los que se calcula la media no se hace por la posición que ocupan en la tabla, sino con una condición

```
mean(osteo$peso[osteo$nivel_ed == "SUPERIORES"])
```

```
## [1] 62.71429
```

Ejercicio 3

Calcula la desviación típica (muestral) del imc de las mujeres con edad menor o igual que 50 años.

Puedes trabajar con los datos como si se tratara de un vector

```
sd(osteo$imc[osteo$edad <= 50])
```

```
## [1] 4.842197
```

O seleccionarlos en la tabla

```
sd(osteo[osteo$edad <= 50 , c("imc")] )
```

```
## [1] 4.842197
```

¿Ves la diferencia en la sintaxis?

Ejercicio 4

Calcula la media de la variable peso para el 50% de las mujeres con menor imc Análogo al anterior, pero seleccionando sólo las filas en las que el peso es mayor que la mediana del imc

```
mean(osteo[osteo$imc < median(osteo$imc), "peso"])
```

```
## [1] 61.14036
```

Ejercicio 5

Visualiza la talla y el imc de las mujeres que ocupan las posiciones 100 a la 105 y de la 120 a la 123.

Hay varias estrategias posibles. Puedes calcular por un lado las tallas y por otro los imcs, como si se tratara de dos vectores

```
osteo$talla[c(100:105, 120:123)]
```

```
## [1] 158 150 159 166 153 155 158 152 156 160
```

```
osteo$imc[c(100:105, 120:123)]
```

```
## [1] 20.43 23.11 34.61 21.41 30.76 31.63 33.25 27.27 27.53 24.80
```

También puedes crear una subtabla seleccionando filas y columnas en la tabla osteo

```
# sintaxis  
# tabla[filas, columnas]  
osteo[c(100:105, 120:123) , c("talla","imc")]
```

```
##      talla  imc  
## 100    158 20.43  
## 101    150 23.11  
## 102    159 34.61  
## 103    166 21.41  
## 104    153 30.76  
## 105    155 31.63  
## 120    158 33.25  
## 121    152 27.27  
## 122    156 27.53  
## 123    160 24.80
```

Te puedes referir a las columnas por la posición que ocupan en la tabla (son la 7ª y la 8ª)

```
# sintaxis  
# tabla[filas, columnas]  
osteo[c(100:105, 120:123) , c(7, 8)]
```

```
##      talla  imc  
## 100    158 20.43  
## 101    150 23.11  
## 102    159 34.61  
## 103    166 21.41  
## 104    153 30.76  
## 105    155 31.63  
## 120    158 33.25  
## 121    152 27.27  
## 122    156 27.53  
## 123    160 24.80
```

Ejercicio 6

Visualiza la tabla con las columnas imc, talla y nivel_ed de las mujeres con menos de 48 años.

Hay varias estrategias posibles. Está claro qué columnas hay que incluir. Ahora la condición se refiere qué filas aparecerán en la tabla

```
# sintaxis
# tabla[filas, columnas]
osteo[osteo$edad < 48 , c("imc", "talla", "nivel_ed")]
```

##	imc	talla	nivel_ed
## 2	22.94	152.0	SECUNDARIOS
## 3	25.64	158.0	PRIMARIOS
## 5	22.72	157.0	PRIMARIOS
## 6	21.97	170.0	SECUNDARIOS
## 46	24.03	154.0	SUPERIORES
## 57	23.73	155.0	SECUNDARIOS
## 59	21.44	168.0	PRIMARIOS
## 60	24.54	164.0	PRIMARIOS
## 61	24.79	154.0	PRIMARIOS
## 68	21.29	147.0	SECUNDARIOS
## 71	21.45	170.0	SUPERIORES
## 72	28.35	156.0	SECUNDARIOS
## 73	18.16	160.0	SUPERIORES
## 100	20.43	158.0	PRIMARIOS
## 103	21.41	166.0	PRIMARIOS
## 104	30.76	153.0	PRIMARIOS
## 107	22.07	160.0	PRIMARIOS
## 115	21.45	163.0	SECUNDARIOS
## 116	24.34	151.0	SUPERIORES
## 117	26.58	164.0	SECUNDARIOS
## 119	35.42	154.0	PRIMARIOS
## 123	24.80	160.0	SECUNDARIOS
## 125	25.08	161.0	PRIMARIOS
## 129	23.05	160.0	PRIMARIOS
## 132	35.44	162.0	PRIMARIOS
## 140	21.25	146.0	PRIMARIOS
## 150	20.78	152.0	SUPERIORES
## 153	32.32	165.0	PRIMARIOS
## 161	25.56	157.0	SECUNDARIOS
## 164	31.98	149.0	PRIMARIOS
## 167	21.78	156.0	SUPERIORES
## 170	28.65	147.0	PRIMARIOS
## 171	25.00	159.0	SECUNDARIOS
## 189	24.31	171.5	PRIMARIOS
## 193	25.53	162.0	PRIMARIOS
## 196	25.71	151.5	PRIMARIOS
## 197	26.99	170.0	PRIMARIOS
## 229	22.66	160.0	SUPERIORES
## 235	27.44	158.0	PRIMARIOS
## 244	32.46	156.0	PRIMARIOS
## 249	31.84	154.0	PRIMARIOS
## 251	33.51	155.0	PRIMARIOS
## 259	31.88	147.0	PRIMARIOS
## 267	22.49	164.0	SECUNDARIOS
## 268	28.64	169.0	PRIMARIOS
## 271	32.45	158.0	PRIMARIOS
## 278	24.88	161.0	SIN ESTUDIOS
## 282	23.73	155.0	SUPERIORES
## 286	24.99	153.0	PRIMARIOS

## 288	27.05	162.0	PRIMARIOS
## 298	22.23	167.0	PRIMARIOS
## 299	25.72	162.0	PRIMARIOS
## 300	24.97	155.0	PRIMARIOS
## 307	25.78	150.0	PRIMARIOS SIN FINALIZAR
## 310	30.44	158.0	PRIMARIOS
## 314	25.89	156.0	PRIMARIOS
## 323	31.62	163.0	PRIMARIOS
## 330	24.24	152.0	PRIMARIOS
## 332	40.25	152.0	PRIMARIOS
## 334	25.35	159.5	SECUNDARIOS
## 342	28.04	158.0	PRIMARIOS
## 354	35.34	156.0	PRIMARIOS
## 356	23.01	160.0	PRIMARIOS
## 359	25.85	161.0	PRIMARIOS
## 363	23.15	161.0	SECUNDARIOS
## 374	35.56	150.0	PRIMARIOS
## 388	28.19	153.0	SUPERIORES
## 389	25.39	173.0	SUPERIORES
## 391	21.08	154.0	SUPERIORES
## 396	23.73	159.0	PRIMARIOS
## 400	27.51	164.0	SECUNDARIOS
## 401	20.70	158.5	SUPERIORES
## 404	27.47	155.0	PRIMARIOS
## 410	25.09	149.0	SIN ESTUDIOS
## 417	24.35	153.0	SECUNDARIOS
## 420	28.01	153.5	SECUNDARIOS
## 429	28.46	154.0	PRIMARIOS SIN FINALIZAR
## 431	24.15	161.5	PRIMARIOS
## 432	25.22	149.0	PRIMARIOS
## 445	25.59	160.0	SECUNDARIOS
## 447	33.46	164.0	PRIMARIOS
## 452	29.92	171.0	SECUNDARIOS
## 455	24.67	152.0	PRIMARIOS
## 460	20.69	166.0	PRIMARIOS
## 480	34.60	153.0	SECUNDARIOS
## 482	25.58	152.5	PRIMARIOS
## 499	25.00	162.0	SECUNDARIOS
## 509	31.81	152.0	PRIMARIOS
## 519	27.78	161.0	SECUNDARIOS
## 528	23.29	139.0	PRIMARIOS
## 529	34.38	160.0	PRIMARIOS SIN FINALIZAR
## 532	24.96	162.0	PRIMARIOS
## 534	25.63	153.0	SUPERIORES
## 538	22.23	158.0	SECUNDARIOS
## 544	20.07	170.0	PRIMARIOS
## 545	24.25	154.0	PRIMARIOS SIN FINALIZAR
## 546	28.30	155.0	PRIMARIOS
## 547	21.78	156.0	SECUNDARIOS
## 548	21.41	166.0	SECUNDARIOS
## 549	27.06	155.0	PRIMARIOS SIN FINALIZAR
## 550	22.63	165.0	SECUNDARIOS
## 551	17.21	167.0	SECUNDARIOS
## 553	21.08	154.0	SUPERIORES

##	555	24.03	154.0	PRIMARIOS SIN FINALIZAR
##	558	21.55	171.0	SUPERIORES
##	559	21.91	162.0	SECUNDARIOS
##	562	29.14	154.0	SECUNDARIOS
##	563	23.31	155.0	PRIMARIOS
##	564	24.01	162.0	SECUNDARIOS
##	566	23.73	159.0	SUPERIORES
##	570	27.73	160.0	PRIMARIOS
##	572	26.91	156.0	SECUNDARIOS
##	574	26.57	157.0	SECUNDARIOS
##	577	22.15	159.0	SECUNDARIOS
##	578	22.94	152.0	SECUNDARIOS
##	580	23.83	156.0	PRIMARIOS
##	584	20.96	156.0	SECUNDARIOS
##	585	31.64	157.0	PRIMARIOS SIN FINALIZAR
##	586	25.65	164.0	PRIMARIOS
##	591	22.60	156.0	SECUNDARIOS
##	592	25.78	150.0	PRIMARIOS
##	593	29.38	156.0	SECUNDARIOS
##	594	21.44	153.0	PRIMARIOS
##	595	26.42	163.0	SECUNDARIOS
##	597	27.43	162.0	PRIMARIOS
##	599	25.80	146.0	PRIMARIOS
##	600	27.77	164.0	PRIMARIOS
##	601	22.50	166.0	PRIMARIOS
##	602	20.58	162.0	SECUNDARIOS
##	604	23.94	157.0	SECUNDARIOS
##	606	28.89	157.0	PRIMARIOS
##	608	29.90	153.0	PRIMARIOS
##	609	23.90	151.0	PRIMARIOS
##	611	29.24	158.0	SECUNDARIOS
##	612	40.04	168.0	PRIMARIOS
##	614	24.35	155.0	PRIMARIOS
##	620	48.39	148.0	PRIMARIOS
##	621	23.19	154.0	SECUNDARIOS
##	629	21.52	165.0	PRIMARIOS
##	635	25.24	158.0	PRIMARIOS
##	642	24.74	167.0	PRIMARIOS
##	643	26.49	153.0	SIN ESTUDIOS
##	650	17.58	160.0	SECUNDARIOS
##	651	26.50	159.0	SECUNDARIOS
##	653	19.15	171.0	SECUNDARIOS
##	655	25.00	160.0	SECUNDARIOS
##	659	21.30	165.0	SUPERIORES
##	661	22.94	152.0	PRIMARIOS SIN FINALIZAR
##	671	22.67	150.0	PRIMARIOS
##	673	21.17	149.0	SECUNDARIOS
##	677	21.63	158.0	PRIMARIOS SIN FINALIZAR
##	679	26.50	171.0	SECUNDARIOS
##	682	25.32	159.0	PRIMARIOS
##	683	22.04	165.0	SECUNDARIOS
##	684	21.30	157.0	SUPERIORES
##	704	23.42	156.0	PRIMARIOS
##	708	27.69	159.0	PRIMARIOS

## 712	24.35	153.0	SECUNDARIOS
## 724	25.92	147.0	SIN ESTUDIOS
## 725	21.97	172.0	SECUNDARIOS
## 734	25.64	158.0	SECUNDARIOS
## 736	21.91	157.0	SECUNDARIOS
## 740	33.69	156.0	PRIMARIOS
## 748	26.70	159.0	PRIMARIOS
## 749	21.60	161.0	SECUNDARIOS
## 750	30.99	154.0	PRIMARIOS
## 755	28.04	163.0	PRIMARIOS
## 757	27.58	161.0	PRIMARIOS
## 762	30.38	159.0	PRIMARIOS
## 780	31.22	155.0	SIN ESTUDIOS
## 785	26.64	155.0	SECUNDARIOS
## 787	20.09	148.0	SECUNDARIOS
## 788	34.65	151.0	SECUNDARIOS
## 789	25.10	152.0	SECUNDARIOS
## 792	26.17	157.0	PRIMARIOS
## 795	22.06	169.0	SUPERIORES
## 805	35.80	155.0	PRIMARIOS SIN FINALIZAR
## 809	23.34	163.0	PRIMARIOS
## 814	25.97	152.0	PRIMARIOS SIN FINALIZAR
## 815	32.78	161.5	PRIMARIOS
## 818	40.37	155.0	PRIMARIOS
## 821	38.71	155.0	SUPERIORES
## 830	22.86	162.0	PRIMARIOS
## 836	25.84	164.0	PRIMARIOS
## 838	25.03	163.0	SECUNDARIOS
## 841	20.32	163.0	SECUNDARIOS
## 843	20.54	148.0	SECUNDARIOS
## 845	24.17	164.0	SECUNDARIOS
## 846	29.52	154.0	PRIMARIOS
## 850	23.15	144.0	PRIMARIOS
## 854	24.24	165.0	PRIMARIOS
## 855	26.22	155.0	PRIMARIOS
## 859	25.78	150.0	PRIMARIOS
## 863	30.67	162.0	SECUNDARIOS
## 870	25.35	172.0	PRIMARIOS
## 872	27.14	164.0	PRIMARIOS
## 874	24.52	159.0	PRIMARIOS
## 878	24.24	165.0	PRIMARIOS
## 879	27.51	164.0	SECUNDARIOS
## 880	21.09	160.0	SECUNDARIOS
## 881	32.16	159.0	PRIMARIOS
## 884	24.46	163.0	SECUNDARIOS
## 893	29.86	159.0	PRIMARIOS SIN FINALIZAR
## 900	35.56	150.0	PRIMARIOS
## 908	31.63	155.0	SECUNDARIOS
## 909	22.03	159.0	SECUNDARIOS
## 911	33.74	143.0	PRIMARIOS
## 912	29.90	153.0	PRIMARIOS
## 916	32.21	161.0	PRIMARIOS SIN FINALIZAR
## 921	30.24	150.5	PRIMARIOS
## 922	32.30	166.0	PRIMARIOS

```

## 923 29.30 160.0      PRIMARIOS
## 924 22.17 157.5      SECUNDARIOS
## 927 22.76 161.0      SECUNDARIOS
## 928 21.75 147.0      PRIMARIOS
## 929 32.46 156.0      SECUNDARIOS
## 930 20.31 160.0      SECUNDARIOS
## 932 23.44 160.0      PRIMARIOS
## 936 25.24 158.0      SECUNDARIOS
## 937 30.58 150.0      PRIMARIOS
## 938 36.40 151.0      PRIMARIOS
## 940 21.27 166.0      PRIMARIOS
## 941 28.00 150.0      PRIMARIOS
## 943 25.63 153.0      PRIMARIOS
## 944 22.03 158.0      PRIMARIOS
## 953 23.80 164.0      SUPERIORES
## 957 32.00 150.0 PRIMARIOS SIN FINALIZAR
## 958 26.78 157.0      SECUNDARIOS
## 959 24.86 169.0      PRIMARIOS
## 960 25.95 170.0      PRIMARIOS
## 961 21.34 162.0      PRIMARIOS
## 962 24.00 150.0      PRIMARIOS
## 964 22.43 158.0      SUPERIORES
## 967 22.95 167.0      SECUNDARIOS
## 973 29.14 155.0      SECUNDARIOS
## 975 21.88 160.0      PRIMARIOS
## 977 28.30 155.0      SECUNDARIOS
## 979 27.55 165.0      PRIMARIOS

```

Ejercicio 7

Calcula la media y la cuasidesviación típica muestral de la variable edad para las mujeres con y sin menopausia. ¿A qué conclusión llegarías?

Usaremos las funciones mean y sd junto con aggregate

```
aggregate(osteo$edad ~ osteo$menop, FUN = mean)
```

```

## osteo$menop osteo$edad
## 1          NO    47.31353
## 2          SI    56.07604

```

Las medias difieren en casi 10 años.

Otra posibilidad es

```
mean(osteo$edad[osteo$menop == "SI"])
```

```
## [1] 56.07604
```

```
mean(osteo$edad[osteo$menop != "SI"])
```

```
## [1] 47.31353
```

```
mean(osteo$edad[osteo$menop == "NO"])
```

```
## [1] 47.31353
```

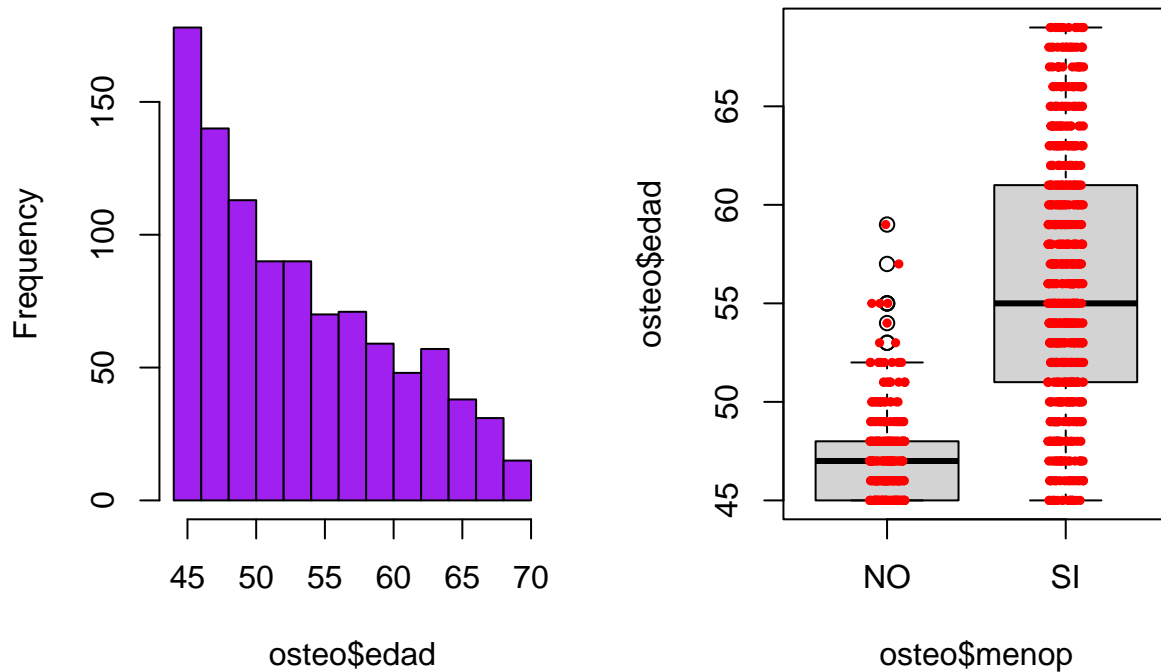
Observa que hemos usado != para expresar diferente de "SI" Podemos visualizar los datos


```

par(mfrow = c(1,2))
hist(osteo$edad, col = "purple")
boxplot(osteo$edad ~ osteo$menop, main = "Edad en función de tener o no la menop")
stripchart(osteo$edad ~ osteo$menop, add = TRUE, method = "jitter", pch = 19, col = "red", cex = .5, ve

```

Histogram of osteo\$edad Edad en función de tener o no la menop



```

par(mfrow = c(1,1))

```

y calcular las dispersiones

```

aggregate(osteo$edad ~ osteo$menop, FUN = sd)

```

```

##   osteo$menop osteo$edad
## 1          NO    2.326933
## 2          SI    6.370163

```

para constatar que, además de tener una media mayor, la dispersión respecto de las medias también es superior en la edad de las mujeres que tienen menopausia.

Ejercicio 8

Calcula la media de la variable peso en función de la variable tipo_men.

Podemos usar la función aggregate:

```

aggregate(osteo$peso ~ osteo$tipo_men, FUN = mean)

```

```

##           osteo$tipo_men osteo$peso
## 1                   AMBAS    68.93291
## 2          HISTERECTOMIA    72.03492

```

```
## 3          NATURAL    70.08915
## 4 NO MENOPAUSIA/NO CONSTA 66.80066
## 5          OVARIECTOMIA 69.98182
```

- Las medias varían entre 66.8 y 72.

Ejercicio 9

Analiza los cuartiles de la variable `peso` en función de la variable `tipo_men`. Interpreta los resultados.

Podemos usar la función `aggregate`:

```
aggregate(osteo$peso ~ osteo$tipo_men, FUN = quantile)
```

```
##          osteo$tipo_men osteo$peso.0% osteo$peso.25% osteo$peso.50%
## 1          AMBAS          45.000          60.500          68.000
## 2  HISTERECTOMIA          46.000          63.100          70.000
## 3          NATURAL          45.000          62.500          69.000
## 4 NO MENOPAUSIA/NO CONSTA 44.000          58.000          65.000
## 5          OVARIECTOMIA 64.000          67.250          69.000
## osteo$peso.75% osteo$peso.100%
## 1          74.000          101.000
## 2          80.000          113.000
## 3          75.625          123.500
## 4          73.500          106.000
## 5          72.150           80.000
```

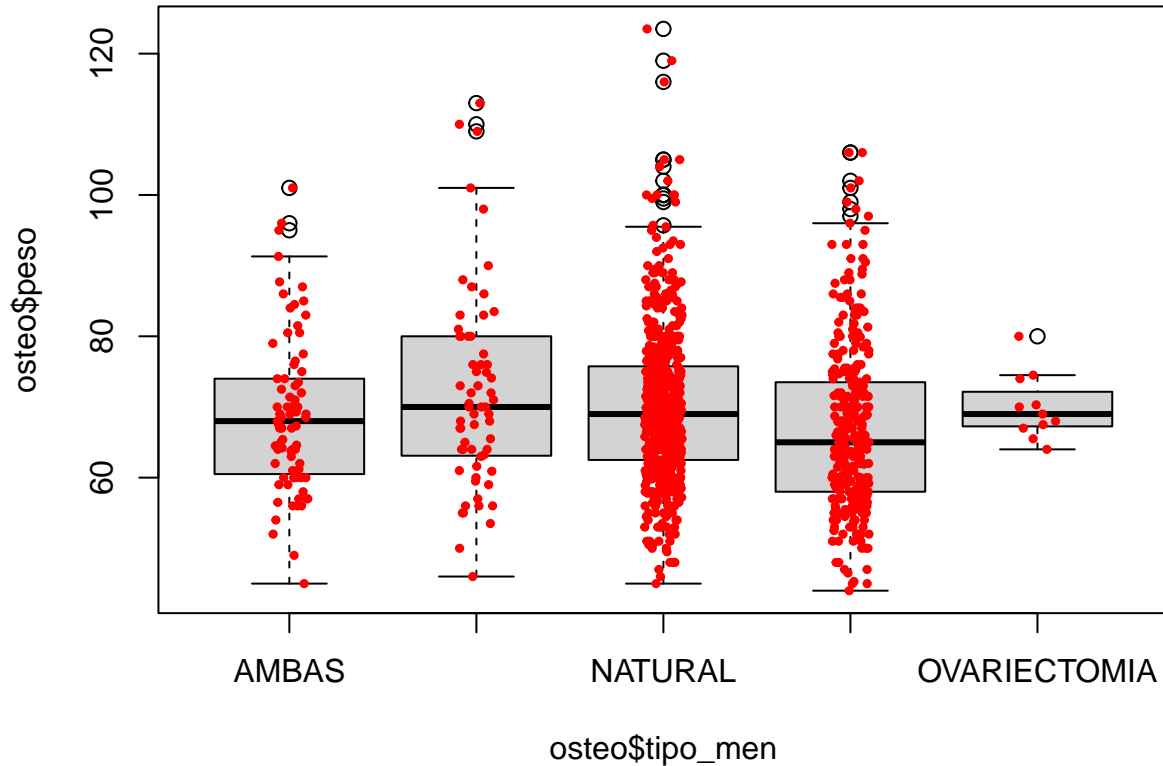
- El peso mínimo parece ser similar, excepto para el nivel `ovariectomia`, que es sensiblemente mayor que el resto.
- El peso máximo del nivel `ovariectomia` que es sensiblemente menor que el resto. No obstante, para el resto de niveles el pesmáximo no es homogéneo.

Ejercicio 10

Utiliza `boxplots` para abordar gráficamente la cuestión anterior. Basta con hacer

```
boxplot(osteo$peso ~ osteo$tipo_men, horizontal = F, main = "Peso en función del tipo de menop")
stripchart(osteo$peso ~ osteo$tipo_men, add = TRUE, method = "jitter", pch = 19, col = "red", cex = .5,
```

Peso en función del tipo de menop



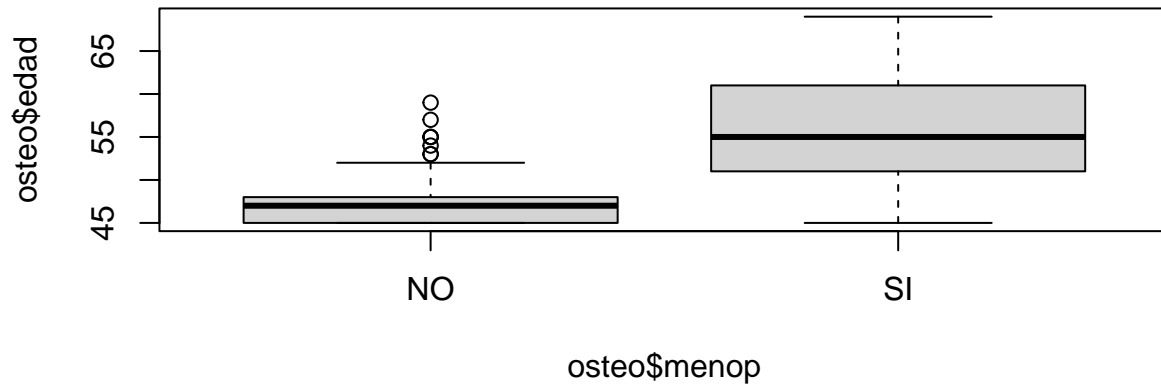
y observamos gráficamente lo que apuntábamos en el apartado anterior.

Ejercicio 11

Ejercicio guiado. Localiza los valores atípicos de la edad para cada nivel del factor menop, eliminalos y genera los boxplot correspondientes para concluir si las dispersiones de la edad en cada nivel del factor son o no comparables.

Para localizar los valores atípicos tenemos que acceder a la información que genera R para construir el boxplot. Esto se hace guardando todo el boxplot en una variable, que llamaremos, por ejemplo, bp:

```
bp = boxplot(osteopeso ~ osteotipo_men)
```



Fíjate en que, en realidad, hemos generado 2 boxplots, uno para las mujeres que tienen la menopausia y otro para las que no. Puedes ver en el marco superior derecho *environment* que hay una nueva variable, `bp` que es una lista de 9 elementos. Uno de ellos contiene el valor que tienen los datos atípicos:

```
bp$out
```

```
## [1] 57 59 55 53 53 55 54 55
```

Lo siguiente es localizarlos en la tabla, para eso hay que usar la función `which`. Observa el siguiente código:

```
which(osteo$edad %in% bp$out)
```

```
## [1] 1 4 10 12 20 22 30 36 38 39 40 41 43 47 48 50 52 58
## [19] 63 80 84 89 92 94 95 96 97 102 105 106 112 120 122 124 126 128
## [37] 131 133 135 137 138 139 147 154 160 162 169 185 192 194 195 199 202 204
## [55] 205 206 208 210 212 213 215 219 221 222 223 227 232 238 239 240 246 256
## [73] 257 258 261 262 263 281 283 285 290 292 293 296 316 319 326 327 328 331
## [91] 338 341 343 344 345 349 351 352 353 360 362 364 368 370 372 373 378 380
## [109] 382 383 384 397 399 403 411 413 416 418 425 426 428 433 434 435 437 439
## [127] 440 441 450 456 461 474 475 478 487 501 506 515 522 523 526 530 533 541
## [145] 542 554 617 622 623 624 628 631 636 638 647 654 663 668 669 670 672 676
## [163] 678 680 687 694 698 705 709 711 722 723 726 728 733 735 760 761 767 768
## [181] 770 774 782 783 790 791 796 797 798 799 810 811 812 813 816 819 831 832
## [199] 833 834 835 864 890 898 981 998
```

Su traducción simultánea sería, literalmente: qué (`which`) mujeres tienen una edad (`osteo$edad`) que está entre (`%in%`) los valores atípicos de los boxplots (`bp$out`). Si te parece una sintaxis rara, te tenemos que dar la razón. Recuerda que no tienes que memorizarla; sólo que saber dónde encontrarla.

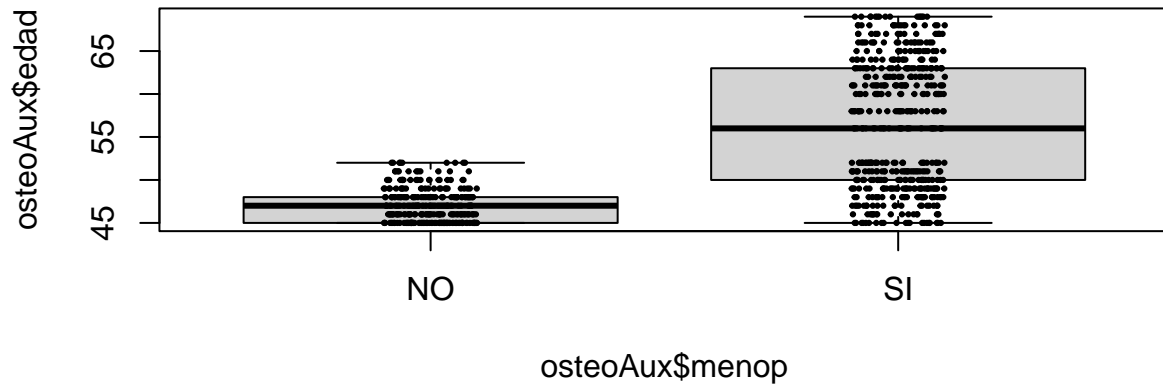
```
# crear nueva tabla "osteoAux" de la que se eliminan las filas de "osteo" que continen outliers
osteoAux = osteo[-which(osteo$edad %in% bp$out), ]
```

```
# representar los boxplots de la nueva tabla
```

```
boxplot(osteoAux$edad ~ osteoAux$menop)
```

```
stripchart(osteoAux$edad ~ osteoAux$menop,
```

```
add = T, vertical = T, method = "jitter", pch = 19, cex = .3)
```



Aunque no es el caso, podrían aparecer nuevos outliers que, de ninguna manera, podrían ser ya considerados como tales. Sería sólo un efecto secundario de haber eliminado la primera tanda de atípicos. Recuerda que no siempre es preciso eliminar los valores atípicos. Su presencia puede ser una pista inestimable en nuestro análisis.