

**Probabilidades y Estadística (C)****Ejercicio**

Se compararon tres dietas respecto al control de azúcar en la sangre en pacientes diabéticos. En el archivo estad\_descriptiva.txt se encuentran los valores de glucosa para las tres dietas consideradas (A, B, C), que contienen las lecturas de glucosa en la sangre de los pacientes. Es deseable que el paciente tenga valores entre 80 - 110 mg/dl.

- Para cada una de las tres dietas calcule medidas de centralidad: la media, la mediana, la media  $\alpha$ -podada para  $\alpha = 0.1, 0.2$ . Para cada dieta compare los valores obtenidos de las cuatro medidas de posición. Mirando las medidas resumen ¿qué dieta no parece ser muy saludable?
- Calcule medidas de dispersión: el desvío estándar, la distancia intercuartiles y la mad de cada una de las dietas. Compare los valores de dispersión obtenidos. ¿Cuál de las dietas parece ser la más estable?
- Obtenga los percentiles 10, 25, 50, 75 y 90. Compare los valores de los percentiles obtenidos entre las distintas dietas. Calcule la distancia intercuartil y el rango de las distancia dietas.
- Construya diagramas de tallo-hoja e histogramas que permitan visualizar los valores de glucosa para cada dieta. Compare la distribución de glucosa. ¿Alguna de ellas parece bimodal? ¿En alguna de ellas parece haber valores alejados? ¿Las dietas mantienen a los pacientes en los valores deseados? ¿La distribución de glucosa es asimétrica en alguno de los grupos? ¿En algún caso el ajuste normal parece razonable?
- Grafique los box-plots correspondientes. ¿Cómo se compara la información que dan estos gráficos con la obtenida con los histogramas? En base a los gráficos obtenidos, discuta simetría, presencia de outliers y compare dispersiones nuevamente.
- Grafique los q-q plots correspondientes. ¿En algún caso el ajuste normal parece razonable?
- En base a todo el análisis anterior, ¿cuál le parece la dieta más aconsejable?

Datos: Todas las muestras son de tamaño 100.

```
a) >summary(DietaA)
  Min. 1st Qu.  Median  Mean 3rd Qu.  Max.
 68.00  90.75  99.00  98.68 105.20 121.00
> summary(DietaB)
  Min. 1st Qu.  Median  Mean 3rd Qu.  Max.
 68.00  79.00  94.50  94.46 110.20 122.00
> summary(DietaC)
  Min. 1st Qu.  Median  Mean 3rd Qu.  Max.
 68.00  70.00  73.00  77.29  81.00 121.00
> mean(gluco,trim=0.1)
dietaa dietab dietac
98.525 94.500 75.300
> mean(gluco,trim=0.2)
dietaa dietab dietac
98.53333 94.40000 74.50000
```

b) Varianza muestral

```
> var(DietaA)
[1] 103.7147
> var(DietaB)
[1] 280.4125
```

```
> var(DietaC)
[1] 112.006
#Desvío Muestral: raíz cuadrada de la varianza muestral
> sqrt(var(DietaA))
[1] 10.18404
> sqrt(var(DietaB))
[1] 16.74552
> sqrt(var(DietaC))
[1] 10.58329
#Distancia Intercuartos
##¿Cómo hacemos para obtener el primer cuarto?
> median(sort(DietaA)[1:50])
[1] 90.5
> median(sort(DietaB)[1:50])
[1] 79
> median(sort(DietaC)[1:50])
[1] 70
> ##¿Y el tercer cuarto?
> median(sort(DietaA)[51:100])
[1] 105.5
> median(sort(DietaB)[51:100])
[1] 110.5
> median(sort(DietaC)[51:100])
[1] 81
> ##Luego, las distancias intercuartos de cada uno son:
> median(sort(DietaA)[51:100])-median(sort(DietaA)[1:50])
[1] 15
> median(sort(DietaB)[51:100])-median(sort(DietaB)[1:50])
[1] 31.5
> median(sort(DietaC)[51:100])-median(sort(DietaC)[1:50])
[1] 11
> #Mad
> mad(DietaA)
[1] 11.1195
> mad(DietaB)
[1] 22.9803
> mad(DietaC)
[1] 5.9304

c) > quantile(DietaA) #Da los percentiles 0% (mínimo), 25%, 50%, 75% y 100%(máximo)
  0%  25%  50%  75% 100%
68.00 90.75 99.00 105.25 121.00
> quantile(DietaB)
  0%  25%  50%  75% 100%
68.00 79.00 94.50 110.25 122.00
> quantile(DietaC)
  0%  25%  50%  75% 100%
 68  70  73  81 121
#Percentiles 10% y 90%
> quantile(DietaA,0.10)
```

```

10%
86
> quantile(DietaB,0.10)
10%
75.8
> quantile(DietaC,0.10)
10%
68
> quantile(DietaA,0.90)
90%
111.1
> quantile(DietaB,0.90)
90%
115
> quantile(DietaC,0.90)
90%
90
> #Observación: Para obtener las distancias intercuartiles podemos hacer:
> quantile(DietaA,0.75)-quantile(DietaA,0.25)
75%
14.5
> quantile(DietaB,0.75)-quantile(DietaB,0.25)
75%
31.25
> quantile(DietaC,0.75)-quantile(DietaC,0.25)
75%
11
> #Rango muestral
> range(DietaA)
[1] 68 121
> range(DietaB)
[1] 68 122
> range(DietaC)
[1] 68 121

```

```
d) >stem(DietaA)
```

The decimal point is 1 digit(s) to the right of the |

```

6 | 8
7 |
8 | 011456666667777888899
9 | 0001122334555556666777788999999
10 | 000001122222444445555668888999
11 | 000111234566899
12 | 1

```

```
> stem(DietaB)
```

The decimal point is 1 digit(s) to the right of the |

```

6 | 889
7 | 002334466666777788899999
8 | 00000001222233333346678
9 |
10 | 1124445566666678899999
11 | 00011111122223445555778899
12 | 2
> stem(DietaC)

```

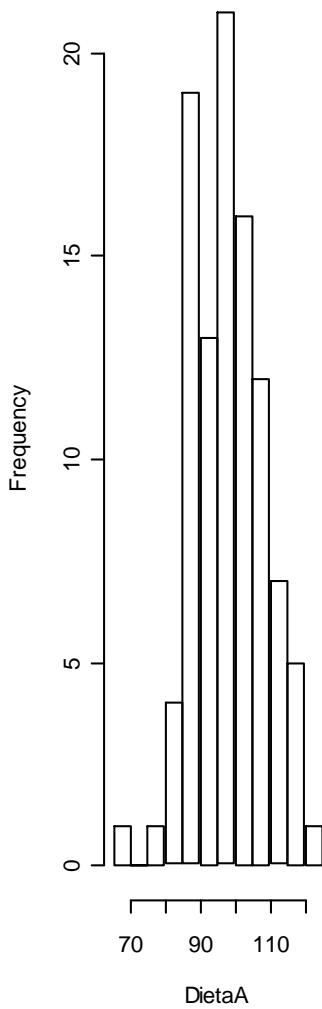
The decimal point is 1 digit(s) to the right of the |

```

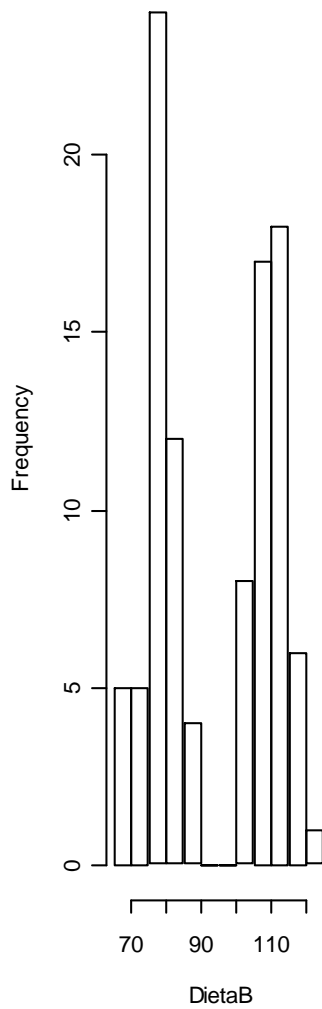
6 | 88888888888999999999
7 | 0000000000111111222222233333334444555567889999
8 | 0000111222344556679
9 | 000455
10 | 1379
11 | 0
12 | 1

```

**Histogram of DietaA**



**Histogram of DietaB**



**Histogram of DietaC**

