

Práctica: Estadística Descriptiva

1. Considere los datos del archivo **datos1.txt**. Efectuar un histograma sobre el mismo con los siguiente criterios:

- (a) Con el esquema por defecto de R.
- (b) Con 1 corte.
- (c) Con 10 cortes.
- (d) Con 100 cortes.

Qué se puede observar de este análisis? Comparar con los datos de **datos2.txt**. Realizar un histograma y un boxplot, diagnosticar si presenta algún problema y trate de solucionarlo.

2. Consideremos X_1, \dots, X_n mediciones. Sean \bar{X} y \tilde{X} la media y la mediana respectivamente.

- (a) Supongamos que a cada medición se le suma una constante c , cómo se verán afectadas la media y la mediana?
- (b) Supongamos que a cada medición se la multiplica por una constante c , cómo se verán afectadas la media y la mediana?
- (c) Sea s_X^2 la varianza muestral correspondiente a la muestra. Demostrar que
 - i. $s_X^2 = \sum_{i=1}^n X_i^2 - \bar{X}^2$.
 - ii. Si $Y_i = X_i + c$, con c constante, pruebe entonces que $S_Y^2 = S_X^2$.
 - iii. Si $Y_i = cX_i$, con c constante, pruebe entonces que $S_Y^2 = c^2 S_X^2$.
- (d) Para qué valores de μ se minimiza $\sum_{i=1}^n (X_i - \mu)^2$ (sug. derive con respecto a μ).

3. La idea de este ejercicio es estudiar el comportamiento de las distintas gráficas ante diversas situaciones simuladas.

- (a) Genere, utilizando el comando **rnorm(1000)**, 1000 elementos provenientes de una normal estándar, guarde los valores en una variable x . Obtenga un histograma, un boxplot y un qqnorm de x .
- (b) Repita el análisis anterior ante las siguientes situaciones:
 - i. Binomial: **rbinom(1000,10,0.4)**.
 - ii. Chi-cuadrado: **rchisq(1000,50)**.
 - iii. F de Snedecor: **rf(1000,90,40)**.
 - iv. Gamma: **rgamma(1000,0.7)**.

4. El archivo **graduados.txt** contiene los promedios obtenidos en su carrera de gradode 30 inscriptos en el programa de postgrado del Departamento de Ingeniería Industrial de la Universidad de Berkeley.

- (a) Calcular la media, la mediana muestral y la media 10%-podada.
- (b) Calcular el desvío estándar, la distancia intercuartil y la MAD.

- (c) Realizar un boxplot sobre este conjunto de datos. Cuáles son las características más sobresalientes? Cómo relaciona lo observado en el boxplot con lo obtenido en los incisos anteriores?
 - (d) Es razonable suponer normalidad de los datos? Con qué gráfico o herramienta lo podría verificar?
5. En el archivo **ciudades.txt** obtendrá los valores de población, en cientos de miles, de las 10 ciudades más pobladas de 4 países en 1967.
- (a) Construir un boxplot paralelo con los cuatro grupos de datos y determinar puntos extremos en cada uno de ellos.
 - (b) Comparar sus centros, sus dispersiones y su simetría. Cuál es el país más homogéneamente poblado?
6. Con el fin de determinar cuál sería un mejor suplemente dietario, se realizó una comparación de la retención de dos formas de hierro: Fe^{2+} y Fe^{3+} . Los investigadores dividieron aleatoriamente a 36 ratas en dos grupos de igual número. A un grupo se le suministró en forma oral una concentración de 1.2 millimolar de Fe^{2+} y al otro grupo se le suministró la misma concentración de Fe^{3+} . Al cabo de cierto tiempo se realizó un conteo en cada rata para determinar el porcentaje de hierro retenido. Estos datos se encuentran en el archivo **hierro.txt**.
- (a) Realizar los boxplots, los histogramas y los qq-plots de los porcentajes obtenidos para cada grupo. En base a estos gráficos, es razonable suponer que cada uno de los conjuntos de datos provienen de una distribución normal?
 - (b) En una segunda etapa, los investigadores transformaron los datos aplicando la función logaritmo natural a cada una de las observaciones. Repetir el análisis anterior sobre los datos transformados.
7. El archivo **cpu.txt** contiene los tiempos de CPU (en segundos) correspondientes a 1000 trabajos enviados por una consultora. Para este conjunto de de datos:
- (a) Calcular la media muestral, la mediana muestral y la media α -podada con $\alpha = 0.1$.
 - (b) Calcular el desvío estándar muestral y la distancia intercuartil.
 - (c) Realizar un histograma y un boxplot, determinar características sobresalientes y eventuales outliers.
 - (d) Supondría usted que los datos tienen una distribución normal?
 - (e) Qué medida de posición considera más apropiada para describir el centro de los datos?