

Adicionales Estadística Descriptiva

Probabilidad y Estadística para Computadores

Ejercicio 1:

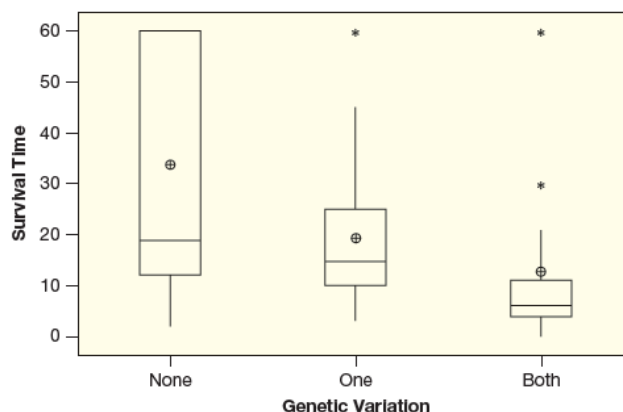
Genética y supervivencia del cáncer

Los perfiles genéticos pueden ayudar a los oncólogos a predecir el tiempo de supervivencia de los pacientes con cáncer. En un estudio, 58 científicos buscaron variaciones en dos genes que codifican proteínas que rigen la reparación del ADN en 103 pacientes con cáncer de pulmón avanzado. Se encontraron variaciones de ambos genes en 13 de los pacientes, se encontraron variaciones en solo uno de los genes en 64 de los pacientes y 26 de los pacientes no tenían ninguna variación. Los científicos compararon el tiempo de supervivencia con quimioterapia para los pacientes en cada uno de los tres grupos. (El estudio duró 60 meses).

Estamos interesados en saber si las diferencias genéticas pueden ayudarnos a predecir el tiempo de supervivencia.

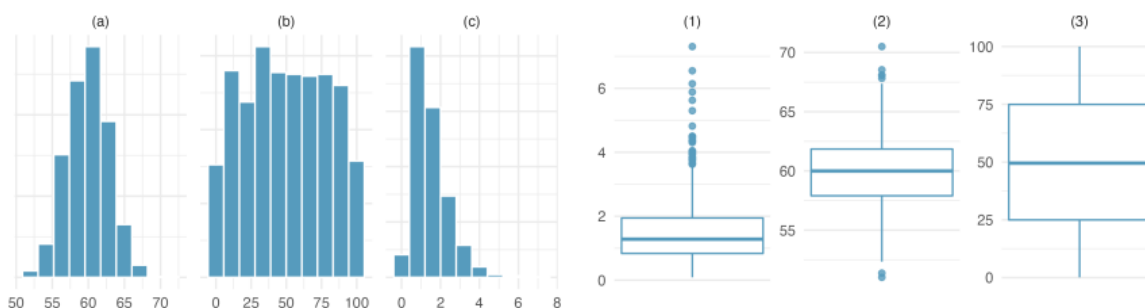
(a) ¿Cuál es la variable explicativa y cuál es la variable de respuesta? Indique si cada uno es categórico o cuantitativo.

(b) La figura muestra diagramas de caja uno al lado del otro para los tres grupos. Discuta lo que muestra el gráfico. ¿Qué conjeturas podríamos hacer sobre estas variaciones genéticas y el tiempo de supervivencia?



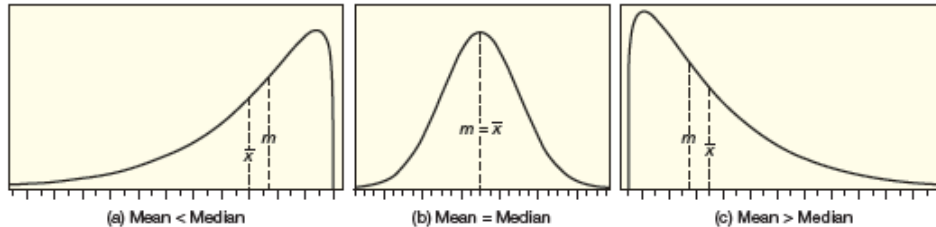
Ejercicio 2:

Describe (en palabras) la distribución de los histogramas y asócielos con los diagramas de caja (boxplot).



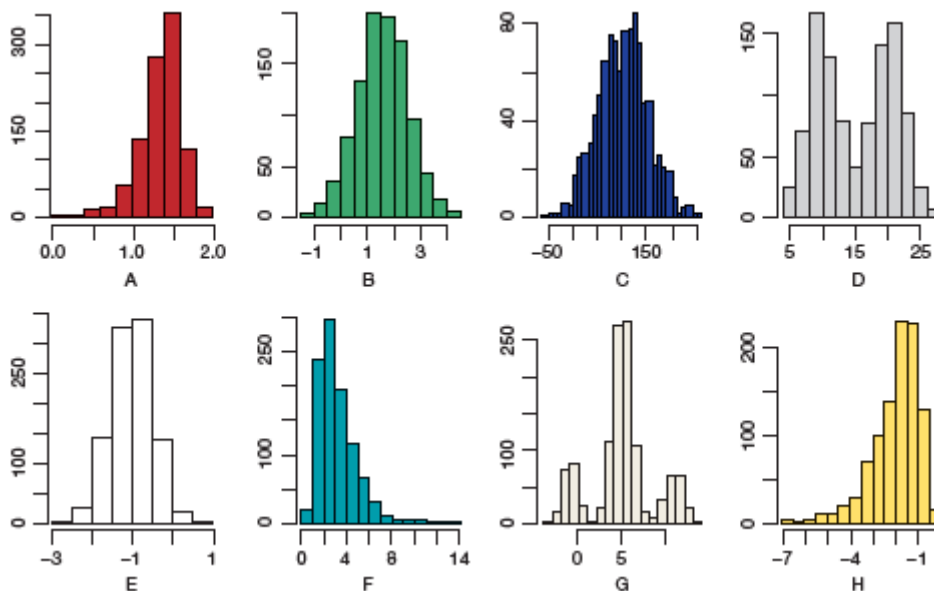
Ejercicio 3:

Dado que la mediana corta un histograma por la mitad, si un histograma es simétrico, la mediana está justo en el medio y es aproximadamente igual a la media. Si los datos están sesgados a la derecha, los valores en la cola derecha extendida elevan la media, pero tienen poco efecto en la mediana. En este caso, la media es mayor que la mediana. De manera similar, si los datos están sesgados hacia la izquierda, la media es menor que la mediana.



Dado el siguiente gráfico, contestar:

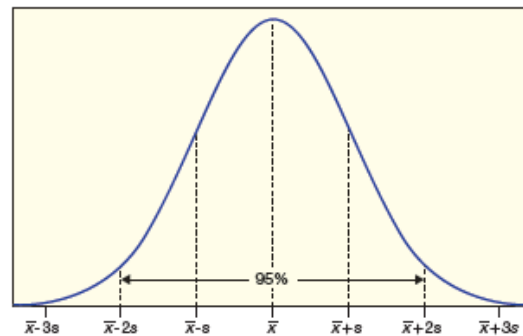
- ¿Qué histogramas están sesgados a la izquierda?
- ¿Qué histogramas están sesgados a la derecha?
- ¿Qué histogramas son aproximadamente simétricos?
- ¿Qué histogramas son aproximadamente simétricos y en forma de campana?
- Para cada uno de los indique si es probable que la media sea mayor que la mediana, menor que la mediana, o aproximadamente igual a la mediana.



Ejercicio 4: Uso de la desviación estándar: la regla del 95 %

Si una distribución de datos es aproximadamente simétrica y tiene forma de campana, aproximadamente el 95 % de los datos deben estar dentro de dos desviaciones estándar de la media. Esto significa que alrededor

del 95 % de los datos de una muestra de una distribución en forma de campana deberían estar en el intervalo de $[\bar{x} - 2s; \bar{x} + 2s]$.



Se recolectó información sobre 110 nuevos modelos de automóviles en 2020. Entre las variables proporcionadas para estos automóviles, está QtrMile, tiempo (en segundos) necesario para que un automóvil recorra un cuarto de milla desde parado.

El siguiente es un histograma de dicha variable. ¿La distribución es aproximadamente simétrica y tiene forma de campana? Usa el histograma para dar una estimación aproximada de la media y la desviación estándar de los tiempos de un cuarto de milla.

