1) La variable categórica "city" decido no describirla ya que no es relevante para el análisis, ya que sus datos no se repiten y su única relación es que son Ciudades de Colombia.

Pero voy a calcular todos los datos a la variable númerica Cuantitativa de la tabla (GDP (USD Billion)); para los demas columnas se realizara en el notebook.

*Mediana

• Ordenar:

0,6; 0,7; 0,8; 0,9; 1; 1,1; 1,2; 1,3; 1,5; 1.7, 1.8, 2, 2.1, 2.3, $\boxed{2.5,\ 2.8}$, 3.0, 3.2, 3.5, 3.8, 4, 4.8, 5.1, 6.2, 7.3, 10.5, 16.8, 22.4, 44.1, 103.5

Total datos => 30

$$\Rightarrow \frac{2.8+2.5}{2} \Rightarrow 2,65$$

*Media $\frac{\Sigma \text{datos}}{N}$

$$\frac{262,5}{30} = 8,75$$

*Moda. Ningún dato se repite,

*Desviación estandar=>

Como la tabla no dice si es muestra o población, Pero no estan todos las ciudades, entonces

$$S = \sqrt{\frac{1}{n-1} \Sigma (x_i - \bar{x})^2}$$

$S=\sqrt{\frac{1}{29}\sum(xi-8{,}75)^2}$

$S=\sqrt{\frac{1}{29}\left[(0{,}6-8{,}75)^2+(0{,}7-8{,}75)^2+(0{,}8-8{,}75)^2+\dots+(103{,}5-8{,}75)^2\right]}$

$S=19{,}91$

1,2) Como ya tengo los datos ordenados de la columna GDP(USD Billon) en el paso anterior, realizare el boxplot sobre esta columna, los demás se haran en el notebook.

$Q_1=$ Mediana de la primera mitad de datos.

$Q_2=$ Mediana

$Q_3=$ Mediana de la segunda mitad

$Q_1 \Rightarrow$ 0,6, 0.7, 0.8, 0.9, 1, 1.1, $\boxed{1.2,\ 1.3}$ 1.5, 1.7, 1.8, 2, 2.1, 2.3

$Q_1 \Rightarrow \frac{1.2+1.3}{2} \Rightarrow 1{,}25$

$Q_2 \Rightarrow 2{,}65$

$Q_3 \Rightarrow$ 3, 3.2, 3.5, 3.8, 4, 4.8, $\boxed{5.1,\ 6.2}$ 7.3, 10.5, 16.8, 22.4, 44.1, 103.5

$Q_3 \Rightarrow \frac{5.1+6{,}2}{2} \Rightarrow 5{,}65$

$X_{min} \Rightarrow 0{,}6$ $X_{max}= 103{,}5$

Limite inferior =>

$Q_1 - 1,5 * IQR$

$1,25 - 1,5 * 4,4 \Rightarrow -5,35$

$IQR = Q_3 - Q_1$

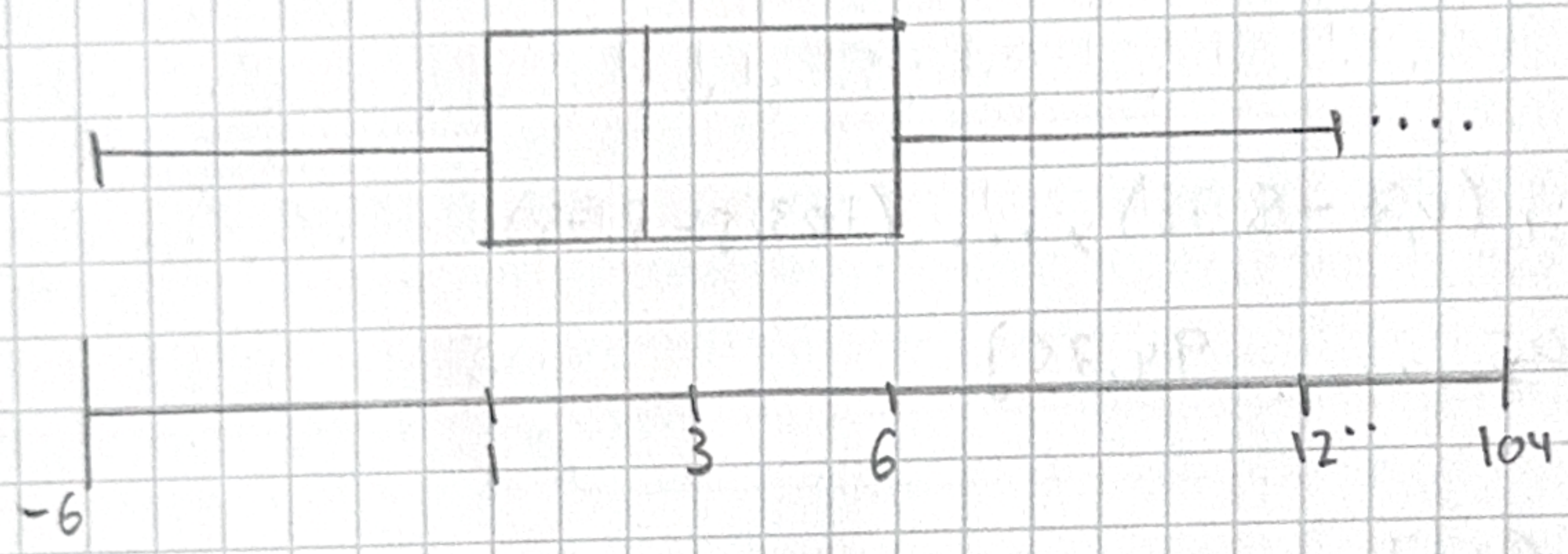
$IQR = 5.65 - 1,25$

$IQR = 4,4$

Limite Superior

$Q_3 + 1,5 * IQR$

$5.65 + 1,5 * 4,4 = 12,25$



1,4 voy a correlacionar GDP con Population

→ Como ya tengo S y $\bar{X}$ de la columna GDP y partimos de que la columna Population será

$\bar{y} = \frac{\sum y_i}{n}$

$S_y = \sqrt{\frac{\sum(y_i - \bar{y})^2}{n-1}}$

$\bar{X} = 8,75$

$\bar{y} = 0,73$

$$\operatorname{Cor}(x,y) = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2}}$$

$\sum(x_i - \bar{x})^2 \Rightarrow (0{,}6 - 8{,}75)^2 + (0{,}7 - 8{,}75)^2 + \ldots + (103{,}5 - 8{,}75)^2$

$\sum(x_i - \bar{x})^2 \Rightarrow 11500{,}955$

$\sum(y_i - \bar{y})^2 = (0{,}01 - 0{,}73)^2 + (0{,}01 - 0{,}73)^2 + \ldots + (7{,}18 - 0{,}73)^2$

$\sum(y_i - \bar{y})^2 = 53{,}054$

$(x_i - \bar{x}) \Rightarrow [(0{,}6 - 8{,}75), (0{,}7 - 8{,}75), \ldots (103{,}5 - 8{,}75)$

$\Rightarrow [-8{,}15, -8{,}05, \ldots, 94{,}75]$

$(y_i - \bar{y}) = [-0{,}72, \ldots, 6{,}45]$

$\sum(x_i - \bar{x})(y_i - \bar{y}) \Rightarrow 773{,}73$

$$\operatorname{Cor}(x,y) = \frac{773{,}73}{\sqrt{11500.955} \cdot \sqrt{53{,}054}}$$

$\operatorname{Cor}(x,y) = 0{,}990520$

Esta correlación entre GDP y Population (x,y) respectivamente aumentan siempre y cuando la otra aumente. y es una Función fuerte.

1,3 Covarianza entre los mismas dos variables (GDP y Population)

$\Sigma(xi-\bar{x}) \Rightarrow [(0,6-8,75)+(0,7-8,75)+\ldots+(103.5-8,75)] \Rightarrow$

$\Sigma(yi-\bar{y}) \Rightarrow [(0,01-0,73)+(0,01-0,73)+\ldots+(7.18-0,73)$

$Cov(x_y) = \dfrac{\Sigma(xi-\bar{x})*(yi-\bar{y})}{N} \Rightarrow \dfrac{773.73}{30} = 25,79$ ó $\dfrac{773,73}{29} = 26,6803$