Siwen Liao (zfc9nu)

DS 3001 EDA Assignment Question 1

① We want to show that $m(a+bX) = a+b \times m(X)$.

Formula: $m(X) = \frac{1}{N}\sum_{i=1}^{N} x_i$

$$m(a+bX) = \frac{1}{N}\sum_{i=1}^{N}(a+bx_i)$$

$$= \frac{1}{N}\left(\sum_{i=1}^{N} a + \sum_{i=1}^{N} bx_i\right) = \frac{1}{N}\left(\sum_{i=1}^{N} a + b\sum_{i=1}^{N} x_i\right)$$

$$= \frac{1}{N}\sum_{i=1}^{N} a + b\cdot\frac{1}{N}\sum_{i=1}^{N} x_i$$

$$= \frac{1}{N}(N\cdot a) + b\cdot m(X) = \boxed{a+b\cdot m(X)}$$

② We want to show that $\text{cov}(X, a+bY) = b \times \text{cov}(X,Y)$.

Formula: $\text{cov}(X,Y) = \frac{1}{N}\sum_{i=1}^{N}(x_i - m(X))(y_i - m(Y))$

$$\text{cov}(X, a+bY) = \frac{1}{N}\sum_{i=1}^{N}(x_i - m(X))((a+by_i) - m(a+bY))$$

$$= \frac{1}{N}\sum_{i=1}^{N}(x_i - m(X))(a+by_i - (a+b\cdot m(Y)))$$

$$= \frac{1}{N}\sum_{i=1}^{N}(x_i - m(X))(a-a+by_i - b\cdot m(Y))$$

$$= \frac{1}{N}\sum_{i=1}^{N}(x_i - m(X))(b(y_i - m(Y)))$$

$$= b\cdot\frac{1}{N}\sum_{i=1}^{N}(x_i - m(X))(y_i - m(Y)) = \boxed{b\cdot \text{cov}(X,Y)}$$

③ We want to show that

$$\mathrm{cov}(a+b\mathbb{X}, a+b\mathbb{X}) = b^2 \mathrm{cov}(\mathbb{X}, \mathbb{X})$$

and $\mathrm{cov}(\mathbb{X}, \mathbb{X}) = s^2$.

Formulas: $\mathrm{cov}(\mathbb{X}, Y) = \frac{1}{N}\sum_{i=1}^{N}(x_i - m(\mathbb{X}))(y_i - m(Y))$

$$s^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - m(\mathbb{X}))^2$$

$$\begin{aligned}
&\mathrm{cov}(a+b\mathbb{X}, a+b\mathbb{X}) \\
&= \frac{1}{N}\sum_{i=1}^{N}(a+bx_i - m(a+b\mathbb{X}))(a+bx_i - m(a+b\mathbb{X})) \\
&= \frac{1}{N}\sum_{i=1}^{N}(a+bx_i - (a+b\cdot m(\mathbb{X})))^2 \\
&= \frac{1}{N}\sum_{i=1}^{N}(a+bx_i - a - b\cdot m(\mathbb{X}))^2 \\
&= \frac{1}{N}\sum_{i=1}^{N}(a-a+bx_i - b\cdot m(\mathbb{X}))^2 \\
&= \frac{1}{N}\sum_{i=1}^{N}(b(x_i - m(\mathbb{X})))^2 \\
&= \frac{1}{N}\sum_{i=1}^{N}b^2(x_i - m(\mathbb{X}))^2 \\
&= b^2\cdot\frac{1}{N}\sum_{i=1}^{N}(x_i - m(\mathbb{X}))(x_i - m(\mathbb{X})) = \boxed{b^2 \mathrm{cov}(\mathbb{X}, \mathbb{X})}
\end{aligned}$$

$$\begin{aligned}
\mathrm{cov}(\mathbb{X}, \mathbb{X}) &= \frac{1}{N}\sum_{i=1}^{N}(x_i - m(\mathbb{X}))(x_i - m(\mathbb{X})) \\
&= \frac{1}{N}\sum_{i=1}^{N}(x_i - m(\mathbb{X}))^2 = s^2
\end{aligned}$$

④ Non-decreasing transformations:

if $x \geq x'$, then $g(x) \geq g(x')$

Median: Yes, a non-decreasing transformation of the median is the median of the transformed variable.

Taking the 2 cases we have, $2+5X$ would just scale and shift all values, so the original median would still be the median.

$\text{arcsinh}(X)$ scales down extreme values and does not change the order of the data.

In general, non-decreasing transformations do not change the order of the data, so the median will remain the median of the transformed values.

Any quantile: Since non-decreasing transformations do not change the order of the data, any quantile will remain the same quantile of the transformed values.

IQR and Range: These will <u>not</u> remain the same unless the non-decreasing transformation is a linear shift. (causing the differences to remain the same). For example, for $2+X$, the IQR and range remain the same.

For $2+5X$, the IQR and range need to be multiplied by 5.

For $\text{arcsinh}(x)$, the transformation is non-linear so the differences between each value will change non-linearly, causing the IQR and range to change.

⑤ non-decreasing transformation $g()$

$$m(g(X)) = g(m(X))$$

No, for example, log is a non-decreasing transformation/function.

Let $X = \{1, 10\}$.

$$m(\log(X)) = \frac{\log(1) + \log(10)}{2} = 0.5$$

$$\log(m(X)) = \log\left(\frac{1+10}{2}\right) = \log\left(\frac{11}{2}\right) \approx 0.74$$

$$m(\log(X)) \neq \log(m(X))$$

So, $m(g(X)) = g(m(X))$ is not always true.