## Question 6 - Assignment Wrangling & EDA

### Question 6.1

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

Let $y_i = a + bx_i$:

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

Split with rules of summation:

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

Since $\sum a = Na$, pull constant $b$ out:

$$m(a+bX) = \frac{Na}{N} + b\left(\frac{1}{N}\sum x_i\right)$$

Substitute definition of $m(X)$ back in:

$$m(a+bX) = a + b \cdot m(X)$$

Question 6.2

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

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

Substitute X for Y in the covariance formula:

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

$$\text{cov}(X,X) = \frac{1}{N}\sum_{i=1}^{N}(x_i - m(X))^2$$

This result is the exact same as the formula for $s^2$

Question 6.3

Property used: $m(a+bY) = a + b \cdot m(Y)$ ← from Question 6.1

$$cov(X, a+bY) = \frac{1}{N} \sum_i (x_i - m(X))([a+by_i] - m(a+bY))$$

Substitute $m(a+bY)$ with $a + b \cdot m(Y)$:

$$cov(X, a+bY) = \frac{1}{N} \sum_i (x_i - m(X))(a + by_i - a - b \cdot m(Y))$$

The $a$ terms cancel out:

$$cov(X, a+bY) = \frac{1}{N} \sum_i (x_i - m(X))(b(y_i - m(Y)))$$

Pull $b$ out of summation:

$$cov(X, a+bY) = b \cdot cov(X, Y)$$

Question 6.4

Property used: Answer from Question 6.3 & symmetry of covariance

Substitution; constants a cancel out and b factors out:

$$\text{cov}(a+bX, a+bY) = \frac{1}{1}\sum (b(x_i - m(X)) \cdot b(y_i - m(Y)))$$

Factor out both b terms:

$$\text{cov}(a+bX, a+bY) = b^2 \cdot \text{cov}(X, Y)$$

Since it is shown in Question 6.2 that:

$\text{cov}(X, X) = s^2$, it's proven that $\text{cov}(bX, bX) = b^2 s^2$.

Question 6.5

Let data be $x_1, \dots, x_N$. Transform each point:

$$y_i = a + bx_i$$

Median:

Because $b > 0$, the transformation is increasing, so order stays the same.

So, the middle values of $Y$ come from the middle values of $X$.

- If the median of $X$ is $med(X)$, then the median of $Y$ is

$$med(Y) = a + b\,med(X)$$

Therefore it is true that:

$$\boxed{med(a + bX) = a + b\,med(X)}$$

Question 6.5 (continued)

IQR:

Since:

$$IQR(X) = Q_3(X) - Q_1(X)$$

quartiles transform the same way:

$$Q_1(Y) = a + bQ_1(X), \quad Q_3(Y) = a + bQ_3(X)$$

Then:

$$IQR(Y) = Q_3(Y) - Q_1(Y)$$
$$= (a + bQ_3(X)) - (a + bQ_1(X))$$
$$= b(Q_3(X) - Q_1(X)) = b\,IQR(X)$$

Therefore, it is true that:

$$\boxed{IQR(a + bX) = b\,IQR(X)}$$

Question 6.6

Data values: Squaring

$$X = \{1, 3\}, \; N = 2$$

Find mean of X:

$$m(X) = \frac{1+3}{2} = 2$$

$$(m(X))^2 = 2^2 = 4$$

square first, then take mean:

$$X^2 = \{1^2, 3^2\} = \{1, 9\}$$

$$m(X^2) = \frac{1+9}{2} = 5$$

$$\boxed{m(X^2) = 5 \neq 4 = (m(X))^2}$$

Question 6.6 (continued)

Square root

Data values:

$$X = \{1, 9\}, N = 2$$

Mean first:

$$m(X) = \frac{1+9}{2} = 5$$

$$\sqrt{m(X)} = \sqrt{5}$$

Take square root first:

$$\sqrt{X} = \{\sqrt{1}, \sqrt{9}\} = \{1, 3\}$$

$$m(\sqrt{X}) = \frac{1+3}{2} = 2$$

$$\boxed{m(\sqrt{X}) = 2 \neq \sqrt{5} = \sqrt{m(X)}}$$

These examples show that for nonlinear transformations like $X^2$ and $\sqrt{X}$

$$m(X^2) \neq (m(X))^2, \quad m(\sqrt{X}) \neq \sqrt{m(X)}$$