

# EDA & Wrangling: Question 6

Zoe Tankersley

1)  $m(a+bX) = a + b(m(X))$

$$\begin{aligned} m(a+bX) &= \frac{1}{N} \sum_{i=1}^N (a+bX_i) \\ &= \frac{1}{N} \sum_{i=1}^N a + \sum_{i=1}^N bX_i \\ &= \frac{1}{N} (Na + b \sum_{i=1}^N X_i) \\ &= a + b \frac{1}{N} \sum_{i=1}^N X_i \end{aligned}$$

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

2)  $\text{cov}(X, X) = S^2$

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

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

3)  $\text{cov}(X, a+bY) = b \text{cov}(X, Y)$

$$\begin{aligned} \text{cov}(X, a+bY) &= \frac{1}{N} \sum_{i=1}^N (X_i - m(X)) ((a+bY_i) - m(a+bY)) \\ &= \frac{1}{N} \sum_{i=1}^N (X_i - m(X)) (a+bY_i - a - b(m(Y))) \\ &= \frac{1}{N} \sum_{i=1}^N (X_i - m(X)) (bY_i - b(m(Y))) \\ &= \frac{1}{N} \sum_{i=1}^N (X_i - m(X)) b(Y_i - m(Y)) \end{aligned}$$

$$\text{cov}(X, a+bY) = b \frac{1}{N} \sum_{i=1}^N (X_i - m(X)) (Y_i - m(Y)) = b \text{cov}(X, Y)$$

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

$$\text{cov}(a+bX, a+bY) = \frac{1}{N} \sum_{i=1}^N b(X_i - m(X)) b(Y_i - m(Y))$$

← based on problem 3

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

5) If  $Y = a + bX$  and  $b > 0$

for any  $x \leq \text{med}(x)$ ,  $a + bX \leq a + b \text{med}(x)$

for any  $x \geq \text{med}(x)$ ,  $a + bX \geq a + b \text{med}(x)$

So  $\boxed{\text{yes, } \text{med}(a + bX) = a + b \cdot \text{med}(x)}$

$$\text{IQR}(a + bX) = (a + bQ_3(x)) - (a + bQ_1(x))$$

$$= b(Q_3(x) - Q_1(x))$$

$$= b \text{IQR}(x)$$

So  $\boxed{\text{IQR}(a + bX) \neq \text{IQR}(x)}$

6)  $X = [1, 2, 3]$ ,  $m(x) = 2$

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

$$m(x^2) = (1^2 + 2^2 + 3^2)/3 = 14/3$$

$$\sqrt{m(x)} = \sqrt{2}$$

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

Since  $4 \neq 14/3$  and  $\sqrt{2} \neq \frac{(1 + \sqrt{2} + \sqrt{3})}{3}$ ,

this shows that the formula  $a + bX$  only works for linear transformations, not nonlinear ones like squaring or square root.