

EDA & Wrangling: Question 6

Zoe Tankersley

1) $m(a+bx) = a+b(m(x))$

$$\begin{aligned} m(a+bx) &= \frac{1}{N} \sum_{i=1}^N (a+bx) \\ &= \frac{1}{N} \sum_{i=1}^N a + \sum_{i=1}^N bx \\ &= \frac{1}{N} \left(Na + b \sum_{i=1}^N x \right) \\ &= a + b \frac{1}{N} \sum_{i=1}^N x \end{aligned}$$

$m(a+bx) = a + b m(x)$

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) $\text{cov}(x, a+bx) = b \text{cov}(x, y)$

$$\begin{aligned} \text{cov}(x, a+bx) &= \frac{1}{N} \sum_{i=1}^N (x_i - m(x)) ((a+bx_i) - m(a+bx)) \\ &= \frac{1}{N} \sum_{i=1}^N (x_i - m(x)) (a+bx_i - a - b(m(y))) \\ &= \frac{1}{N} \sum_{i=1}^N (x_i - m(x)) (bx_i - b(m(y))) \\ &= \frac{1}{N} \sum_{i=1}^N (x_i - m(x)) b(y_i - m(y)) \end{aligned}$$

$\text{cov}(x, a+bx) = b \frac{1}{N} \sum_{i=1}^N (x_i - m(x))(y_i - m(y)) = b \text{cov}(x, y)$

4) $\text{cov}(a+bx, a+bx) = b^2 \text{cov}(x, y)$

$$\text{cov}(a+bx, a+bx) = \frac{1}{N} \sum_{i=1}^N b(x_i - m(x)) b(y_i - m(y))$$

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 $\boxed{\text{Yes, } \text{med}(a+bx) = a + b \times \text{med}(x)}$

$$\begin{aligned} \text{IQR}(a+bx) &= (a+bQ_3(x)) - (a+bQ_1(x)) \\ &= b(Q_3(x) - Q_1(x)) \\ &= b\text{IQR}(x) \end{aligned}$$

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

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

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

$$m(x^2) = (1^2 + 2^2 + 3^2)/3 = \frac{14}{3}$$

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

$$m\sqrt{x} \neq (\sqrt{1} + \sqrt{2} + \sqrt{3})/3$$

Since $4 \neq \frac{14}{3}$ 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.