EDA assignment

Q1

$*m(x)$ = sample mean of X

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

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

plug in: $m(a+bX) = \frac{1}{N}\sum_{i=1}^{N}(a+bX_i)$

$m(a+bX) = \frac{1}{N}\left(\underbrace{\sum_{i=1}^{N} a}_{=N\cdot a} + \underbrace{\sum_{i=1}^{N} bx_i}_{=b\sum_{i=1}^{N} x_i}\right) = (m(a) + m(bx))$

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

$m(a+bX) = a + b \cdot \underbrace{\frac{1}{N}\sum_{i=1}^{N} x_i}_{=\text{sample mean formula} = m(x)}$

$\therefore$

$$\boxed{m(a+bX) = a + b \cdot m(x)}$$

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

covariance formula:

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

plug in: $cov(x, a+bY) = \frac{1}{N}\sum_{i=1}^{N}(x_i - m(x))(a + by_i - \underbrace{m(a+by)}_{=a+b\cdot m(y)})$

$a + by_i - a - b\cdot m(y)$

$cov(x, a+by) = \frac{1}{N}\sum_{i=1}^{N}(x_i - m(x))\underbrace{(by_i - bm(y))}_{=b(y_i - m(y))}$

$cov(x, a+by) = b \cdot \underbrace{\frac{1}{N}\sum_{i=1}^{N}(x_i - m(x))(y_i - m(Y))}_{=\text{covariance formula}}$

$\therefore$

$$\boxed{cov(x, a+by) = b \cdot cov(x,y)}$$

3. $cov(a+bX, a+bX) = b^2 cov(X,X)$

plug in to covariance formula:

$$cov(a+bX, a+bX) = \frac{1}{N}\sum_{i=1}^{N}((a+bx_i) - \overbrace{m(a+bX)}^{= a+b\cdot m(x)})((a+bx_i) - m(a+bx))$$

$$= \frac{1}{N}\sum_{i=1}^{N}((a+bx_i) - (a+b\cdot m(x))((a+bx_i) - (a+b\cdot m(x))$$

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

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

$$= b^2 \cdot \underbrace{\frac{1}{N}\sum_{i=1}^{N}(x_i - m(x))^2}$$

= sample variance formula (s^2) AND $cov(X,X)$

$\therefore$

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

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

transformation of median = median of transformed variable?

4. non-decr. transformation preserves order. yes, everywhere, b/c a non-decr. transformation preserves order. since the median is just the middle value, the transformed median is the same as the median of the transformed variable. This preserves quantiles as well

5. No, not always. $m(a+bX) = a+b\cdot m(X)$ applies only to linear functions (like $y = a+bx$). For Non-linear transformations for $g()$, $m(g(x)) \neq g(m(x))$ b/c the mean is dependent on variance as well as order (unlike median). For example, if $g(x) = x^2$.

ex: $X = 1, 2, 3$ $\quad g(x) = x^2 = 1, 4, 9$

$m = \frac{6}{3} = 2 \quad \neq \quad m = \frac{1+4+9}{3} = \frac{14}{3}$