EDA Assignment

Phillip Bonsu

Q1a.

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

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

Q1b.

$$Cov(x, a+bY) = bCov(x,Y)$$

$$Cov(x, a+by) = \frac{1}{N}\sum_{i=1}^{N}(x_i - m(x))((a+by_i) - m(a+by))$$

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

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

$$= b\cdot\frac{1}{N}\sum_{i=1}^{N}(x_i - m(x))(y_i - m(y)) = bCov(x,y)$$

Q1c.

$Cov(a+bx, a+bx) = b^2 Cov(x,x)$ and $Cov(x,x) = S^2$

$X = Y$ $\quad Cov(x, a+bx) = bCov(x,x)$

$a = a+bx$ $\quad Cov(a+bx, a+bx) = bCov(a+bx, x) = b(bCov(x,x)) = b^2 Cov(x,x)$

def $U = V = x$

$$Cov(x,x) = \frac{1}{N}\sum_{i=1}^{N}(x_i - m(x))^2 = S^2$$

Q1d.

Median: If g is non-decreasing the data's order will not change. $g(x) = g(median(x))$. When N is odd it is exact and when N is defined as a middle observation it is even.

Quartile: Similar to the median. Saved by a non-decreasing transform

$$Q_{g(x)}(p) = g(Q_x(p))$$

IQR: $IQR(g(x)) = g(Q_3(x)) - g(Q_1(x))$

unless g is linear, it won't generally be equivalent to a constant multiple of the original IQR.

Range: $(g(x)) = g(\max x) - g(\min x)$

Q1e. No, $m(g(x))$ is not equal to $g(m(x))$ for nonlinear g.
Per Jensen's inequality, if g is convex then $m(g(x)) \geq g(m(x))$ and if concave, then $m(g(x)) \leq g(m(x))$. Therefore, equality is only true when linear.