EDA Q1:

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

$$= \frac{(a+bX_1)+(a+bX_2)+\dots+(a+bX_n)}{N}$$

$$= \frac{a+bX_1+a+bX_2+\dots+a+bX_n}{N}$$

$$= \frac{b(x_1+x_2+\dots+x_n)+aN}{N}$$

$$= \frac{b(x_1+x_2+\dots+x_n)}{N} + \frac{aN}{N}$$

$$= b\frac{(x_1+x_2+\dots+x_n)}{N} + a$$

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

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

2. $$cov(X,a+bY) = \frac{1}{N}\sum_{i=1}^{N}(x_i - m(X))(y_i - m(Y))$$

$$= \frac{1}{N}\sum_{i=1}^{N}(X_i - m(X))((a+bY_i) - m(a+bY))$$

$$= \frac{1}{N}(x_1 - \frac{1}{N}\sum_{i=1}^{N}x_i)(a+bY_1 - a+b\cdot m(Y)) +$$
$$(x_2 - \frac{1}{N}\sum_{i=1}^{N}x_i)((a+bY_2) - a+b\cdot m(Y)) + \dots +$$
$$(x_n - m(X))((a+bY_n) - a+b\cdot m(Y))$$

$$= \cancel{\frac{1}{N}(X_n}$$

$$= \frac{1}{N}\sum_{i=1}^{n}(X_i - m(X))((a+bY_i) - (a+b\,m(Y)))$$

$$= \frac{1}{N}\sum_{i=1}^{n}(x_i - m(X))(a+bY_i - a - b\cdot m(Y))$$

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

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

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

3. $\text{cov}(a+bX, a+bX) = b^2 \text{cov}(X,X)$

$$\text{cov}(a+bX, a+bX) = \frac{1}{N}\sum_{i=1}^{N} (\cancel{X_i}(a+bX_i) - (a+bm(X)))((a+bX_i) - (a+bm(X)))$$

$$= \frac{1}{N}\sum_{i=1}^{N} (a+bX_i - a - bm(X))(a+bX_i - a - bm(X))$$

$$= \frac{1}{N}\sum_{i=1}^{N} (bX_i - bm(X))(bX_i - b\cdot m(X))$$

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

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

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

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

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

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

$$= S^2 \quad \text{(sample co-variance)}$$

4. $P(X \leq m) \geq 0.5$ and $P(X \geq m) \geq 0.5$

$P(g(X) \leq g(m)) = P(X \leq m) \geq 0.5$

↳ • if g is non-decreasing, $X \leq m \iff g(X) < g(m)$

$P(g(X) \geq g(m)) = P(X \geq m) \geq 0.5$

hence $g(m)$ is also median of $g(X)$

this applies to all quantiles in $q \in (0,1)$

if Q is the quantile of X, it suggests that at least q% of the X values are at or below Q : $g(X) \leq g(Q)$

hence $P(g(X) \leq g(Q)) = P(X \leq Q) \geq q$ (q-quantile)

$P(g(X) \geq g(Q)) = P(X \geq Q) \geq 1-q$

hence it works the same for all quantiles.

4. $IQR(X) = Q_{0.75}(X) - Q_{0.25}(X)$

transformation: $IQR(g(X)) = Q_{0.75}(g(X)) - Q_{0.25}(g(X))$

$IQR(g(X)) = g(Q_{0.75}(X)) - g(Q_{0.25}(X))$

the two equations are not the same, as the 75^{th} Quantile of $g(X)$ is different than the $g(X)$ of the 75^{th} quantile of X.

Range: $Max(X) - Min(X)$

$Range(g(X)) = Max(g(X)) - Min(g(X))$

$Range(g(X)) = g(max(X)) - g(min(X))$

the two equations here are also not equal as the maximum and minimum value of $g(X)$ is different than the $g(X)$ of the maximum and minimum value of X.

Therefore IQR and Range doesn't apply to this condition.

5. $P(X \leq m(X)) \geq 0.5$ and $P(X \geq m(X)) \geq 0.5$

if g is non-decreasing: $X \leq m(X) = g(X) \leq g(m(X))$

$P(g(X) \leq g(m(X))) = P(X \leq m(X)) \geq 0.5$

$P(g(X) \geq g(m(X))) = P(X \geq m(X)) \geq 0.5$

hence $g(m(X))$ satisfies the medvan property for $g(X)$.