DS EDA Q1

1. $m(a+bx) = a + b \cdot m(x)$

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

2. $$\text{cov}(x, a+bY) = b \cdot \text{cov}(x,Y) = \frac{1}{N}\sum_{i=1}^{N}(x_i - m(x))((a+by_i) - m(a+bY))$$
$$= \frac{1}{N}\sum_{i=1}^{N}(x_i - m(x))((a+by_i) - (a+bm(Y)))$$
$$= \frac{1}{N}\sum_{i=1}^{N}(x_i - m(x))\, b(y_i - m(Y))$$
$$= b \cdot \frac{1}{N}\sum_{i=1}^{N}(x_i - m(x))(y_i - m(Y)) = \boxed{b\,\text{cov}(x,Y)}$$

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

$U = a+bX$
$$\text{cov}(U,U) = \frac{1}{N}\sum_{i=1}^{N}((a+bx_i) - m(a+bX))^2$$
$$= \frac{1}{N}\sum_{i=1}^{N}((a+bx_i) - (a+bm(X)))^2$$
$$= \frac{1}{N}\sum_{i=1}^{N}(b(x_i - m(X)))^2$$
$$= b^2 \cdot \frac{1}{N}\sum_{i=1}^{N}(x_i - m(X))^2$$
$$= b^2 s^2$$

set $a=0$, $b=1$ gives $\boxed{\text{cov}(x,x) = s^2}$ so $\boxed{b^2\text{cov}(x,x)}$

4. if $x \geq x'$ then $g(x) \geq g(x')$ like $2+5x$ or $\text{arcsinh}(x)$

Median if g strictly increasing then median of $g(x)$ equals $g(\text{median}(X))$. We see this as data sorted $x_{(1)} \leq \ldots \leq x_{(N)}$, the center order statistic corresponds w/ the middle order statistic (post applying g)

- using the quantile definition (any value btw 2 middles) w/ strictly increasing g, the population quantile formula $Q_{g(x)}(P) = g(Q_x(p))$ holds

$$\frac{1}{2}\left(g(x_{(N/2)}) + g(x_{(N/2+1)})\right) \neq g\left(\frac{1}{2}(x_{(N/2)} + x_{(N/2+1)})\right)$$

Quantile

For a strictly increasing g, $Q_g(x)(p) = g(Q_x(p))$ for all $p \in (0,1)$ — for non-decreasing can yield set of valid quantiles)

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

No, not $g(IQR(x))$ unless g is linear $g(x) = a + bx$ w/ $b \geq 0$

$g(x) = 2 + 5x$ $\quad IQR(g(x)) = 5 \cdot IQR(x)$

$g(x) = \text{arcsinh}(x) \longrightarrow \text{arcsinh}(Q_3) - \text{arcsinh}(Q_1)$

Range $\text{range}(g(X)) = \max g(X) - \min g(X) = g(\max X) - g(\min X)$

$= b \cdot \text{range}(X)$ for only linear $g(x) = a + bx$ w/ $b \geq 0$

So, median & all quantiles are equivalent for increasing transformations. IQR & range can transform when you apply g to relevant quantiles/extremes, but they $\times$ scale simply unless g is linear.

5. Is always true that $m(g(X)) = g(m(X))$ for non-decreasing function $\rightarrow$ $\boxed{\text{No}}$

in general for non-linear g, $m(g(X)) \neq g(m(X))$

if g convex (non-decreasing), $m(g(X)) \geq g(m(X))$

if g concave (non-decreasing), $m(g(X)) \leq g(m(X))$

for any non-degenerate X, $\frac{1}{N} \sum e^{x_i} > e^{\frac{1}{N} \sum x_i}$