Q6. In class, we talked about how to compute the sample mean of a variable X,

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

and sample covariance of two variables X and Y,

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

Recall, the sample variance of X is

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

It can be very helpful to understand some basic properties of these statistics. If you want to write your calculations on a piece of paper, take a photo, and upload that to your GitHub repo, that's probably easiest.

We're going to look at **linear transformations** of X, $Y = a + bX$. So we take each value of X, x_i, and transform it as $y_i = a + bx_i$.

1. Show that $m(a+bX) = a + b\times m(X)$.
2. Show that $\text{cov}(X,X) = s^2$.
3. Show that $\text{cov}(X, a+bY) = b\times\text{cov}(X,Y)$
4. Show that $\text{cov}(a+bX, a+bY) = b^2\text{cov}(X,Y)$. Notice, this also means that $\text{cov}(bX,bX) = b^2 s^2$.
5. Suppose $b > 0$ and let the median of X be $\text{med}(X)$. Is it true that the median of $a + bX$ is equal to $a + b\times\text{med}(X)$? Is the IQR of $a+bX$ equal to $a + b\times\text{IQR}(X)$?
6. Show by example that the means of X^2 and $\sqrt{X}$ are generally not $(m(X))^2$ and $\sqrt{m(X)}$. So, the results we derived above really depend on the linearity of the transformation $Y = a + bX$, and transformations like $Y = X^2$ or $Y = \sqrt{X}$ will not behave in a similar way.

1) Transformed variable

$Y = a + bX \rightarrow y_i = a + bx_i$

Sample mean of Y

$$m(Y) = \frac{1}{N}\sum_{i=1}^{N} y_i$$

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

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

$$= \frac{1}{N}\left(aN + b\sum_{i=1}^{N}x_i\right)$$

$$= a + b\frac{1}{N}\sum_{i=1}^{N}x_i$$

$$\boxed{m(Y) = a + b\,m(X)}$$

2) $Y = X$ in covariance formula

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

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

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

3) Covariance formula — linear transformation — Sample mean

$$\text{cov}(X,Z) = \frac{1}{N}\sum_{i=1}^{N}(x_i - m(X))(z_i - m(Z)) \qquad Z = a + bY \rightarrow z_i = a + by_i \qquad m(Z) = a + bm(Y)$$

$$\text{cov}(X, a+bY) = \frac{1}{N}\sum_{i=1}^{N}(x_i - m(X))((a+by_i) - m(a+bY))$$

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

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

$$\text{so} \quad \text{cov}(X, a+bY) = b\cdot\text{cov}(X,Y)$$

4) Substitute bX into covariance formula

$$\text{cov}(bX,bX) = \frac{1}{N}\sum_{i=1}^{N}((bx_i) - m(bX))((bx_i) - m(bX))$$

From part 1: $m(bX) \rightarrow bm(X)$

$$\text{cov}(bX,bX) = \frac{1}{N}\sum_{i=1}^{N}(bx_i - bm(X))(bx_i - bm(X))$$

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

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

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

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

5) Given: $Y = a + bX$, $b > 0$

Median

half of the $x_i \le med(x)$, half of the $x_i \ge med(x)$

since $b > 0$ order is the same

$x_i \le med(x) \rightarrow a + bx_i \le a + b\,med(x)$

$x_i \ge med(x) \rightarrow a + bx_i \ge a + b\,med(x)$

so median of Y is: $med(Y) = med(a + bX) = a + b\,med(x)$

IQR

$IQR = Q3_x - Q1_x$ ↗ 75% of the $x_i \le Q3_x$ ↗ 25% of the $x_i \le Q1_x$

linear transformation

if $x_i \le Q1_x$, then $a + bx_i \le a + bQ1_x$

if $x_i \le Q3_x$, then $a + bx_i \le a + bQ3_x$

so exactly the same 25% of observations that were below $Q1_x$ in X are below $a + bQ1_x$ in Y

$Q1_Y = a + bQ1_x$, $Q3_Y = a + bQ3_x$

compute IQR

$IQR(Y) = Q3_Y - Q1_Y = (a + bQ3x) - (a + bQ3x) - (a + bQ1_x) = b(Q3x - Q1_x) = bIQR(x)$

↳ $IQR(a + bX) = bIQR(x) \neq a + bIQR(X)$

6) Example mean

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

mean of x^2

$x^2 = \{1^2, 3^2\} = \{1, 9\}$

$m(x^2) = \frac{1+9}{2} = 5$ → $m(x^2) \neq (m(x))^2$

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

mean of $\sqrt{x}$

$\sqrt{x} = \{\sqrt{1}, \sqrt{3}\}$

$m(\sqrt{x}) = \frac{\sqrt{1} + \sqrt{3}}{2}$ → $m(\sqrt{x}) \neq \sqrt{m(x)}$

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