# Exploratory Data Analysis: Q1

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

the sample mean is $m(X) = \frac{1}{n}\sum_{i=1}^{n} X_i$

Consider $m(a+bX)$

$m(a+bX) = \frac{1}{n}\sum_{i=1}^{n}(a+bX_i) = \frac{1}{n}\sum_{i=1}^{n} a + \frac{1}{n}\sum_{i=1}^{n} bX_i$

$\frac{1}{n}\sum_{i=1}^{n} a = a$

$\frac{1}{n}\sum_{i=1}^{n} bX_i = b \cdot \frac{1}{n}\sum_{i=1}^{n} X_i = b \cdot m(X)$

Thus, $m(a+bX) = a + b\,m(X)$

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

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

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

↳ $a + bm(Y)$ - we found in #1

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

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

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

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

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

$cov(a+bX, a+bX) = \underbrace{b \cdot cov(X, a+bX)}_{\text{using conclusion from question \#2}} = \underbrace{b \cdot b \cdot cov(X,X)}_{\text{using conclusion from question \#2}} = b^2 cov(X,X)$

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

&, by definition, $cov(X,X) = s^2$, thus, $b^2 cov(X,X) = b^2 s^2$

4. Is a non-decreasing transformation of the median the median of the transformed variable?

Not necessarily - the median is the middle value of an ordered dataset. If we apply a non-decreasing function to the dataset, the order of the data remains unchanged (so, the value that is the median will produce the median of the transformed data, as the order does not change), but the value of the median is not guarenteed to be the value of the original dataset's median.

Does your answer apply to any quantile? The IQR? The range?

Yes, this does apply to all the above. The Quantiles of a dataset when applied a non-decreasing transformation can be found through applying the transformation to the values corresponding to the original dataset's quartiles, as quartiles are calculated in association to the ordering of elements. Like median, though, the values of the quartiles may not be the same between the pre- and post-transformation dataset. The IQR is calculated as $Q_3 - Q_1$. As quartiles are ordered, the values from the original dataset's $Q_1$ & $Q_3$ will produce the value of the transformed dataset's $Q_3$ & $Q_1$ values, but the IQR value will not necessarily be same between the pre- and post-transformation dataset. Likewise, the range is calculated by max(dataset) - min(dataset), and - again, as this is part of the order - the values from the original dataset's max & min will produce the value of the transformed dataset's max & min, but the range value will not necessarily be same between the pre- and post-transformation dataset.

5. No, it is not always true that $g(m(x)) = m(g(x))$. Take the dataset $X = \{1,2,3\}$ $m(X) = 2$. Take $g(y) = y^2$. $g(m(x)) = 4$. $g(X) = \{1,4,9\}$. $m(g(x)) = 14/3$ $14/3 \neq 2$. Thus the sample mean of a transformation, $m(g(x)) \neq$ the transformation of the sample mean, as there exists a counter example.