ECON 256 – Correlations Worksheet Professor Jack Rossbach

Problem 1 For this exercise, we will be computing the correlation between two data series

Data Series	Observation 1	Observation 2	Observation 3
X	15	15	30
Y	10	25	4

Q1.1) Compute the mean (average) value of both series. There are three observations so $n = 3$ in the functions below. x_i is the ith observation of X.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \; ; \;\; \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

$$\bar{x} = \frac{1}{3}(15 + 15 + 30) = \frac{1}{3}(60) = \mathbf{20}$$
$$\bar{y} = \frac{1}{3}(10 + 25 + 4) = \frac{1}{3}(39) = \mathbf{13}$$

Q1.2) Compute the Covariance between X and Y. The covariance is defined as

$$\text{Cov}(X, Y) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

$$\text{Cov}(X, Y) = \frac{1}{3-1}\big((15 - \mathbf{20})(10 - \mathbf{13}) + (15 - \mathbf{20})(25 - \mathbf{13}) + (30 - \mathbf{20})(4 - \mathbf{13})\big)$$
$$= \frac{1}{2}\big((-5)(-3) + (-5)(12) + (10)(-9)\big)$$
$$= \frac{1}{2}(15 - 60 - 90) = \frac{1}{2}(-135) = \mathbf{-67.5}$$

Q1.3) The variance of X is defined as $\text{Var}[X] = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$. If we compute it, we find $\text{Var}[X] = 75$. Similarly, $\text{Var}[Y] = 117$. Using the formula below, compute the correlation, r, between X and Y.

$$r = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}[X]} \times \sqrt{\text{Var}[Y]}}$$

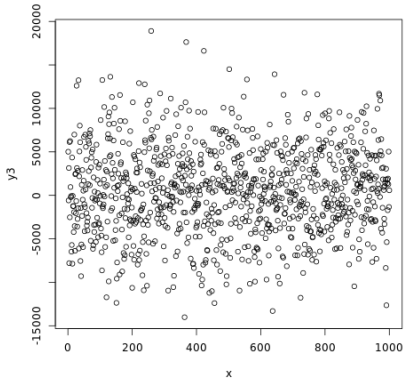
$$r = \frac{-67.5}{\sqrt{75} \times \sqrt{117}} = \frac{-67.5}{8.66 \times 10.82} = \frac{-67.5}{93.70} = \mathbf{-0.72}$$

See Last Page for Additional Discusson on the Formulas for Correlation and Covariance

Problem 2

Look at the following graphs. Classify each graph as having a **Positive correlation** or **Negative correlation**, and whether the correlation is strong (close to 1 or minus -1), medium strength (closer to 0.5 or -0.5), or weak (close to 0). If the correlation is weak, you may not be able to tell whether it is positive or negative, but you can still guess.

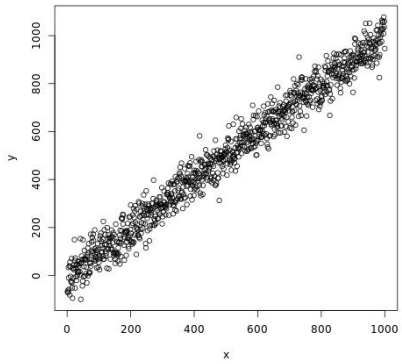
Q2.1)



Is Correlation Positive or Negative?
Correlation is **Positive** (Barely, so it's fine if couldn't tell)

Is Correlation Strong, Medium, or Weak?
Weak. Correlation is $r = 0.05$ which is close to zero

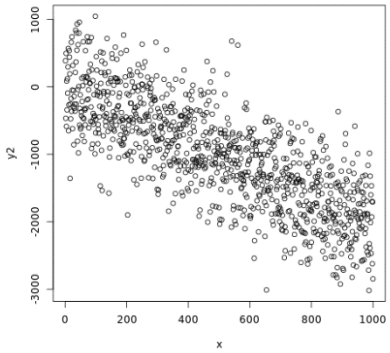
Q2.2)



Is Correlation Positive or Negative?
Correlation is **Positive**

Is Correlation Strong, Medium, or Weak?
Strong. Correlation is $r = 0.98$ which is close to one

Q2.3)

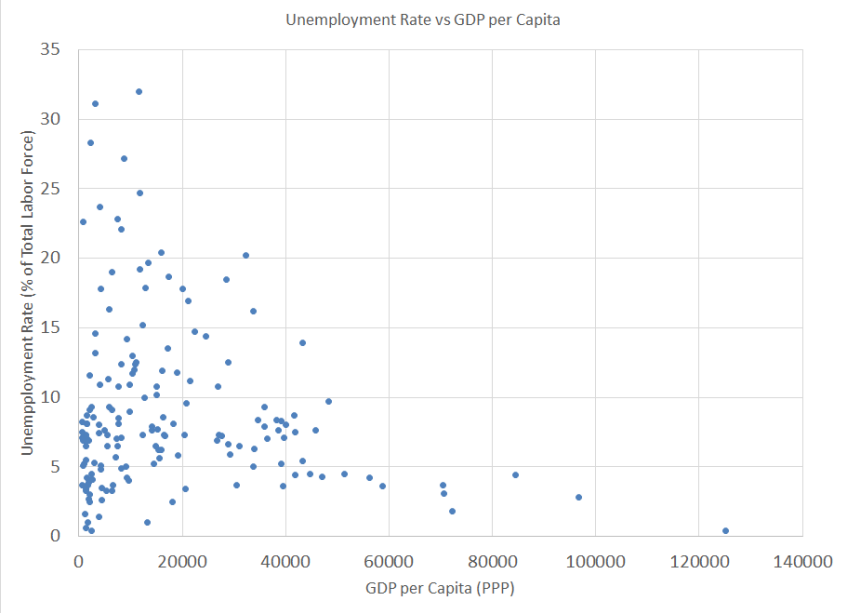


Is Correlation Positive or Negative?
Correlation is **Negative**

Is Correlation Strong, Medium, or Weak?
Either **Medium Strength** or **Strong** is acceptable.
Correlation is $r = -0.75$ which is right on the boundary between being Strong or Medium strength.

Problem 3

Q3.1) Below is the Unemployment Rate vs GDP per Capita (PPP) in 2010 for all countries with data.



Does the correlation look Positive or Negative?

Negative (but it's hard to tell)

Does it look like a strong or weak correlation?

Weak

(medium strength wouldn't be marked incorrectly. It's a tough graph to tell just by looking)

Q3.2) Using the following summary statistics, compute the correlation between GDP per Capita and Unemployment for the World. I'll refer to the first series as GDP and the second as Unemp.

$$\text{Cov}[\text{GDP}, \text{Unemp}] = -19739.6$$
$$\text{SD}[\text{GDP}] = \sqrt{\text{Var}[\text{GDP}]} = 19704.8$$
$$\text{SD}[\text{Unemp}] = \sqrt{\text{Var}[\text{Unemp}]} = 6.05$$

What is the correlation, r, between GDP per Capita and the Unemployment rate. Is the correlation expected or unexpected?

$$r = \frac{\text{Cov}(\text{GDP}, \text{Unemp})}{\sqrt{\text{Var}[\text{GDP}]} \times \sqrt{\text{Var}[\text{Unemp}]}} = \frac{-19739.6}{19704.8 \times 6.05} = \frac{-19739.6}{119214.0} = \mathbf{-0.165}$$

This means there is a **negative correlation** between GDP per capita and Unemployment. This is as most people expect, richer countries have less unemployed people because the labor market functions better so people can get jobs. Also, countries with less unemployment are richer, since more people are working and producing goods and services.

Note also that the correlation is pretty weak. When we compute the correlation separately for rich and poor countries, we find a positive correlation for one group and a negative correlation for the other group. When we combine the rich and poor countries, it weakens the overall correlation.

Q3.3) Let's now compute the correlation between GDP per Capita and the Unemployment Rate separately for countries that have GDP per capita's less than $10,000/person and countries that have GDP per capita's over that. We'll refer to the first set as Poor countries and the second as Rich countries.

For Poor Countires, we have the following statistics.

$$\text{Cov}[\text{GDP}, \text{Unemp}] = 4153$$
$$\text{SD}[\text{GDP}] = \sqrt{\text{Var}[\text{GDP}]} = 2823.9$$
$$\text{SD}[\text{Unemp}] = \sqrt{\text{Var}[\text{Unemp}]} = 6.46$$

What is the correlation, r, between GDP per Capita and the Unemployment rate for the poor countries? How might you explain this?

$$r = \frac{\text{Cov}(\text{GDP}, \text{Unemp})}{\sqrt{\text{Var}[\text{GDP}]} \times \sqrt{\text{Var}[\text{Unemp}]}} = \frac{4153}{2823.9 \times 6.46} = \frac{4153}{18242.39} = \mathbf{0.228}$$

This means there is a **positive correlation** between GDP per capita and Unemployment for poor countries. This is surprising, it means that poor countries with higher GDP per Capita have more unemployed people (as a fraction of the labor force) on average.

The causality is not clear, and the correlation is not particularly strong, but one possible contributing factor is that if a country is very poor, it may lack any sort of safety net for unemployed people. In this case, if you are unemployed you may starve to death or otherwise not be able to survive. Since there is no safety net, most people will either work or die, which means less people will be unemployed compared to countries that are slightly richer and able to provide social safety nets such as unemployment and welfare benefits.

Q3.4) For Rich Countires, we have the following statistics.

$$\text{Cov}[\text{GDP}, \text{Unemp}] = -54872.8$$
$$\text{SD}[\text{GDP}] = \sqrt{\text{Var}[\text{GDP}]} = 20420.5$$
$$\text{SD}[\text{Unemp}] = \sqrt{\text{Var}[\text{Unemp}]} = 5.63$$

What is the correlation, r, between GDP per Capita and the Unemployment rate for the rich countries?

$$r = \frac{\text{Cov}(\text{GDP}, \text{Unemp})}{\sqrt{\text{Var}[\text{GDP}]} \times \sqrt{\text{Var}[\text{Unemp}]}} = \frac{-54872.8}{20420.5 \times 5.63} = \frac{-54872.8}{114967.4} = \mathbf{-0.477}$$

For Rich countries, we are back to having **negative correlation** between unemployment and GDP as expected. Although there are differences across countries in how generous the safety nets are, unlike poor countries, every rich country has a social safety net of some sort so that unemployed people and people that lose their jobs generally do not die.

Additional Discussion on Formulas for Correlation, Covariance, and Variance

In the week 8 slides, we give the formula for the correlation between X and Y as either:

$$r = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}[X]} \times \sqrt{\text{Var}[Y]}}$$

Or

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2} \times \sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

The reason these are the same formulas are because the definitions for the sample covariance and variance (we will always use the sample statistics in this class, and not the population statistics)

$$\text{Cov}(X,Y) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

$$\text{Var}(X) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$\text{Var}(Y) = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$$

Therefore

$$\frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}[X]} \times \sqrt{\text{Var}[Y]}} = \frac{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} \times \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

$$= \frac{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\left(\frac{1}{n-1}\right)^2} \times \sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2} \times \sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

$$= \frac{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\left(\frac{1}{n-1}\right) \times \sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2} \times \sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

$$= \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2} \times \sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

So the formulas are the same.