CHULALONGKORN UNIVERSITY
FACULTY OF ECONOMICS

2946653 Research Methods in Labour Economics
and Human Resource Management

EXERCISE 1

Instruction

1. Students are encouraged to work in pairs.
2. Students should hand in your script by the end of the lecture. No grading for late assignment.
3. To show that you have done the computation by yourselves, you have to copy the print out from EView or MS Excel into your script. Photocopy is unacceptable.

Question 1: Basic Concepts

One of the simplest ways to see the relationship between any two economic variables is to find the correlation coefficient, ρ. Before we discuss how to compute ρ, let's refresh some statistical concepts first. For a variable X and Y with N observations each,

$$\text{Sample Mean:} \qquad \overline{X} = \sum_{i=1}^{N} \frac{X_i}{N}$$

$$\text{Sample Variance} \quad : \quad \widehat{\sigma}_X^2 = \sum_{i=1}^{N} \frac{\left(X_i - \overline{X}\right)^2}{N-1}$$

$$\text{Standard Deviation} \quad : \quad \widehat{\sigma}_X = \sqrt{\widehat{\sigma}_X^2} = \sqrt{\sum_{i=1}^{N} \frac{\left(X_i - \overline{X}\right)^2}{N-1}}$$

$$\text{Sample Covariance} \quad : \quad \widehat{\sigma}_{XY} = \sum_{i=1}^{N} \frac{\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{N-1}$$

The (sample) correlation coefficient between X and Y , denoted by $\widehat{\rho}$, is computed by the following formula

$$\widehat{\rho} = \frac{\widehat{\sigma}_{XY}}{\widehat{\sigma}_X \widehat{\sigma}_Y}$$

Notice that $-1 \leq \widehat{\rho} \leq 1$. Also, if X and Y are independent, then $\widehat{\rho} = 0$ but the reverse is not always true.

Consider DATA SETS 1, there are four sets of data on the average weekly earnings of production workers from selective industries of the United States.

1. From any industry of your choice, compute the mean and the variance of each year.
2. Again, from any industry of your choice (must be differed from the above question), pick any two years and compute the covariance.
3. Pick any two industries then select a year of your choice, plot the scatter diagram, compute the variances, the covariance and the correlation coefficient.
4. Provide an interpretation to the correlation coefficient and the scatter diagram you have just computed.

Question 2: Hypothesis Testing

The correlation coefficient you have just computed in Question 1 is called the sample correlation coefficient or $\widehat{\rho}$. Since we do not know the population correlation coefficient, ρ, we need to make sure that our computed statistic is a good representative of the population. We therefore consider the following null and alternative hypotheses:

$$\begin{aligned} H_0 &: \rho = 0 \\ H_1 &: \rho \neq 0 \end{aligned}$$

We want to test whether the population correlation coefficient is significantly different from zero ,i.e., the two sets of data of your choice in Question 1 do not have any correlation. Note that when ρ is zero, $\widehat{\rho}$ has a normal distribution,

$$\widehat{\rho} \sim \mathcal{N}\left(0, \frac{1-\widehat{\rho}^2}{N-2}\right)$$

To test the null hypothesis, we use the t ratio from the following formula

$$t_{\widehat{\rho}} = \frac{\widehat{\rho} - \rho}{\widehat{\sigma}_{\widehat{\rho}}}$$

with $N-2$ degrees of freedom where N is the number of observations. We then compare $t_{\widehat{\rho}}$ with its critical value at a $\alpha\%$ significance level which can be obtained from the t-statistic table.

1. From the correlation coefficient you found in Question 1.4, find $t_{\widehat{\rho}}$.
2. Select a significant level α then find the associated critical value of t ratio.
3. Test the above null hypothesis against the alternative one.
4. Instead of doing the two-tailed test, consider the following alternative hypothesis,

   $$H_1 : \rho > 0$$

   Using the same significant level as (4), find $t_{\widehat{\rho}}$, its associated critical value and test the null hypothesis against the above alternative.

Question 3: Spearman's Rank Correlation Coefficient

The correlation coefficient we have studied in Question 1 and 2 are for quantitative data. What should we do if we want to find a correlation between qualitative variables such as preferences (like the most - hate it), education performance (top of the class - ranked last), or working environment (very good - extremely bad)? Spearman's rank correlation coefficient can solve this problem. Let d = difference in the ranks assigned to the same individual of phenomenon and N = numbers of individuals or phenomena ranked, Spearman's rank correlation coefficient is computed by

$$\widehat{\rho}_s = 1 - \frac{6\sum_{i=1}^{N} d^2}{N\left(N^2-1\right)}$$

For example, suppose a doctor collects the data on weight and height from 5 of his patients. Let X_i represents the weight (Kgs) of an individual i and Y_i represents the height (cms) of an individual i. The rank of X_i and Y_i and d are as follow:

Individual	X_i	Y_i	Rank of X_i	Rank of Y_i	d_i
1	60.5	189	4	5	-1
2	55	165	2	1	1
3	72	175	5	4	1
4	45	166	1	2	-1
5	59	172	3	3	0

Note that the rank of X_i and Y_i can be either ascending or descending orders.

The hypothesis testing in this case is different from the previous two questions. For Spearman's rank correlation coefficient, if the number of observations is more than 30, it has a normal distribution

$$\widehat{\rho}_s \sim \mathcal{N}\left(0, \frac{1}{N-1}\right)$$

and we use the Z value to test the null hypothesis against the alternative hypothesis,

$$Z_{\widehat{\rho}_s} = \frac{\widehat{\rho}_s - \rho_s}{\widehat{\sigma}_{\widehat{\rho}_s}}$$

For $N < 30$, we can compare $\widehat{\rho}_s$ directly with its critical value from the computed table.

Your task is to find 10 of your classmates and ask them their undergraduate and their current GPAXs. Then, compute Spearman's rank correlation coefficient. You also have to test the following null hypothesis $H_0 : \rho_s = 0$ against $H_1 : \rho_s \neq 0$ and $H_2 : \rho_s > 0$. Use the following critical value of Spearman's rank correlation coefficient. To prevent you from making up the data, you have to submit the raw data including your classmates' names. Also, discuss your finding.

(Hint: For this question, MS Excel might be more useful than EView.)

one-tailed	$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.01$
two-tailed	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.02$
critical value	0.564	0.648	0.745

Question 4: Your Own Experiment

Choose two economic variables (which should have an economic relationship) and find the data on the internet. They must contain at least 10 observations.

1. State your data source(s).
2. Plot the scatter diagram of these two variables.
3. Compute the variances, covariance and the correlation coefficient
4. Test the null hypothesis $H_0 : \rho = 0$ against $H_1 : \rho \neq 0$.
5. Conduct the one-tailed hypothesis testing with the same significant level as the previous question.