

ECO 1164

Research Methods and Statistics 2

Spring 2019 – 2020

ASSIGNMENT

Submission Date: 23 May 2020, 8:30

RULES

- The weight of this assignment is **35%** of your term grade.
- You need to submit **2 documents**:
 1. **Answer sheet**: You need to write your answers that you have found in R Studio in a different sheet. See the example answer sheet in the attachment.
 - Your answer sheet can be prepared in Word, which is provided as example answer sheet.
 - Or, you can write your answers on a paper. In this case, you need to use a scanner application in order to transform your hard copy homework into a pdf format and submit a unique document.
 - Possible scan applications: Scannable, CamScanner, Microsoft Office Lens| PDF Scan
 2. **R script**: Write commands for each question on the same R script. Leave blank lines between each question and write the number of the question at the beginning of each question. See the example R script in the attachment.
- **Answer sheets without R scripts can not be graded.**
- **Use your own wordings in your answers, otherwise the score of your answer is going to be 0 (ZERO).**
- **No submissions are going to be accepted after the deadline.**

PART 1: Descriptive Statistics (75 pts)

Answer all of the questions in Part 1 by using “wage2_subset.rds” data. Find “wage2_subset.rds” data in the attachment.

Data: “wage2_subset.rds” data is a subset of “wage2” data provided in Wooldridge package. Definition for each variable in the data set is provided below:

- **wage:** monthly earnings
- **hours:** average weekly hours
- **IQ:** IQ score
- **educ:** years of education
- **exper:** years of work experience
- **age:** age in years
- **married:** =1 if married
- **urban:** =1 if live in a metropolitan area
- **meduc:** mother's education
- **feduc:** father's education

Questions:

1. **(5pts)** State the number of variables and observations available in the data set.
2. **(5pts)** Classify all variables with respect to measurement levels (Hint: First determine whether a variable is quantitative or qualitative. Then, determine whether it is nominal or ordinal if the variable is qualitative and interval or rational if it is quantitative.) Provide your answer in a table format.

Example of a table format:

| | Qualitative | | Quantitative | |
|---------------|-------------|---------|--------------|----------|
| Variable name | Nominal | Ordinal | Interval | Rational |
| Variable 1 | X | | | |
| Variable 2 | | | | X |

3. **(15pts)** Calculate measures of location and variability provided in the table for **all quantitative variables EXCEPT “hours”, “meduc” and “feduc”**. Provide your answer in a table format and round all measures to two decimals places.

Example of a table format:

| Variable name | Mean | Median | Mode | 25 th Percentile | 3 rd Quartile | Range | IQR | Var | Std. Dev. | Coeff. of Var. |
|---------------|------|--------|------|-----------------------------|--------------------------|-------|-----|-----|-----------|----------------|
| Variable 1 | | | | | | | | | | |
| Variable 2 | | | | | | | | | | |

4. **(9pts)** Draw a table showing frequency, relative frequency and percentage relative frequency of education. Set classes as primary school (if $8 \leq \text{educ} < 12$), high school (if $12 \leq \text{educ} < 16$), university/college (if $16 \leq \text{educ} < 18$) and graduate school (if $\text{educ} \geq 18$).
5. **(7pts)** Answer following questions by using cross-tabulations.
 - a. Provide percentage shares of married-workers living in urban and non-urban area and compare it with single-workers counterparts.
 - b. Provide education level shares of workers living in urban and non-urban area. Are education level distributions of workers living in urban area and non-urban area similar? Discuss your results.

6. **(6pts)** Draw a histogram for wage. Discuss the skewness of the distribution by
- using the skewness number,
 - using the shape of the histogram,
 - comparing the mean and the median.
7. **(10pts)** Draw a horizontal box plot for wage, showing five-number summary on the x-axis.
- Is the shape of distribution provided by the box-plot consistent with the one provided by its histogram?
 - According to box-plot, is there any outlier?
 - Check whether there is an outlier by using z-score.
8. **(12pts)** Discuss the strength and direction of the relationship between variables provided below by using a single measure and draw their scatter plot.
- wage and educ
 - educ and meduc
 - meduc and feduc
9. **(6pts)** If wage is measured as hourly wage how do the strength and the direction of the relation between wage and educ change? Assume that there are 4 weeks in each month. Calculate the covariance and correlation in this new case. Discuss the difference between these two measures using your results.

PART 2: Probability and Discrete Distribution (25 pts)

10. **(6pts)** Bahçeşehir University wishes to evaluate the effectiveness of the university promotion days in university choice of Istanbul based students. 75% of all potential students in Istanbul were reached along the university promotion days. 28% of those contacted chose to study at Bahçeşehir University while 8% of students who chose to study at Bahçeşehir University are not contacted during promotion days.
- Draw a tree diagram (with probabilities on each branch and all outcome probabilities) to represent the two-step experiment.
 - What is the probability that a student who chooses to study at Bahçeşehir University has not been contacted during the university promotion days?
11. **(3 pts)** An investor is considering three strategies for a \$1,000 investment. The probable returns are estimated as follows:
- Strategy 1:** A profit of \$10,000 with probability 0.15 and a loss of \$1,000 with probability 0.85

Strategy 2: A profit of \$1,000 with probability 0.50, a profit of \$500 with probability 0.30, and a loss of \$500 with probability 0.20

Strategy 3: A certain profit of \$400

Which strategy has the highest expected profit? Explain why you would or would not advise the investor to adopt this strategy.

12. **(10pts)** Bahcesehir University has started a new shuttle rides in between Besiktas South Campus and Goztepe Kampus. Records were kept on the numbers of students using this new shuttle ride from Goztepe to Besiktas during the early-morning weekday service. The accompanying table shows proportions over all weekdays.

| | | | | | | | | |
|--------------------|------|------|------|------|------|------|------|------|
| Number of students | 20 | 25 | 30 | 35 | 40 | 45 | 50 | 55 |
| Proportion | 0.23 | 0.31 | 0.19 | 0.12 | 0.08 | 0.04 | 0.02 | 0.01 |

- Graph the probability distribution.
 - Calculate and graph the cumulative probability distribution.
 - What is the probability that on a randomly chosen weekday at least 35 students will take the shuttle from Goztepe to Besiktas?
 - Two weekdays are chosen at random. What is the probability that on both of these days there will be fewer than 30 students who take the shuttle from Goztepe to Besiktas?
 - Find the mean and standard deviation of the number of students using this new shuttle ride from Goztepe to Besiktas.
 - If the cost of a ride is 7TRY, find the mean and standard deviation of the total payments of students from this new shuttle ride from Goztepe to Besiktas on a weekday.
13. **(6pts)** A campus finance officer finds that, for all parking tickets issued, fines are paid for 78% of the tickets. The fine is 2 TRY. In the most recent week, 620 parking tickets have been issued.
- What is the random variable in this experiment?
 - What is the distribution of the random variable? Why?
 - Given the distribution, state what do numbers 78% and 620 correspond to?
 - Find the mean and standard deviation of the number of these tickets for which the fines will be paid.
 - Find the mean and standard deviation of the amount of money that will be obtained from the payment of these fines.