**COGS 536 Fall 2018**

**Homework 1**

**Due: November 12, Monday, 23:59**

1. Consider the description of a study given below.

Suppose that you are interested in how well adults learn in either passive or active learning environments. Each participant is required to learn two different mazes. One maze is learnt by walking around the maze (active condition) while the other is learnt by watching someone else walk around it (passive condition). You measure how long it takes the participant to learn the maze (measured in minutes). A total of 12 participants take part in your study. The order of the two conditions is randomized.

1. What is the research question pursued in this study?

How active and passive learning environments effect adults learning time while they are learning a maze ?

1. Identify the research design and the independent/dependent variables involved.

Dependent variable: Time to learn Active. Time to learn Passive

Independent variable: Gender, Age, Maze Order, (participant count)

1. State the specific hypothesis that can be tested with this design.

Learning time of maze for active learning is %10 shorter than passive learning time

1. For each variable state what would be the appropriate level of measurement.

Learning time for active learning

learning time for passive learning

2. Suppose that you conducted the above study with a colleague. Participants were undergraduate students and were allowed up to a maximum of 30 minutes to learn each maze. Your colleague organized the data set and saved it in SPSS as maze.sav (provided in the HW1 folder). Do the following on the maze.spss file.

1. Check the Variable View of the data file carefully. Make sure that the variables have been set up correctly. Correct any errors you find.

|  |  |
| --- | --- |
| Variable Name | Modifications |
| Sex | * Measure should be Nominal. (“Measure column”) * A description can be added into “Label” column. |
| Maze Order | * Measure should be Ordinal |

1. Use either the Explore or Frequencies command to check the data – look to see whether there are any values outside the expected ranges.

Values outside the expected ranges:

* Sex: 12
* Age in Years:226
* Time to Learn Active : 97.90
* Order of Mazes:3

Variables containing missing values:

* age in years
* time to learn active
* time to learn passive
* order of mazes

**Frequencies**

|  |  |  |
| --- | --- | --- |
| **Notes** | | |
| Output Created | | 11-NOV-2018 19:41:29 |
| Comments | |  |
| Input | Data | /home/ayse/Desktop/MedicalInformatics/536/Homework1/maze.sav |
| Active Dataset | DataSet1 |
| Filter | <none> |
| Weight | <none> |
| Split File | <none> |
| N of Rows in Working Data File | 13 |
| Missing Value Handling | Definition of Missing | User-defined missing values are treated as missing. |
| Cases Used | Statistics are based on all cases with valid data. |
| Syntax | | FREQUENCIES VARIABLES=Sex Age ActiveMazeTime PassiveMazeTime MazeOrder  /STATISTICS=STDDEV VARIANCE RANGE MINIMUM MAXIMUM SEMEAN MEAN MEDIAN MODE SUM  /ORDER=ANALYSIS. |
| Resources | Processor Time | 00:00:00.00 |
| Elapsed Time | 00:00:00.00 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Statistics** | | | | | | |
|  | | Gender | Age in Years | Time to learn Active | Time to Learn Passive | Order of Mazes |
| N | Valid | 13 | 12 | 12 | 12 | 12 |
| Missing | 0 | 1 | 1 | 1 | 1 |
| Mean | | 2.46 | 36.83 | 16.0925 | 10.0617 | 1.67 |
| Std. Error of Mean | | .806 | 17.307 | 7.48021 | .88184 | .188 |
| Median | | 2.00 | 20.00 | 8.6000 | 10.1000 | 2.00 |
| Mode | | 2 | 19 | 5.34a | 6.16a | 2 |
| Std. Deviation | | 2.904 | 59.953 | 25.91222 | 3.05478 | .651 |
| Variance | | 8.436 | 3594.333 | 671.443 | 9.332 | .424 |
| Range | | 11 | 224 | 92.56 | 9.28 | 2 |
| Minimum | | 1 | 2 | 5.34 | 6.16 | 1 |
| Maximum | | 12 | 226 | 97.90 | 15.44 | 3 |
| Sum | | 32 | 442 | 193.11 | 120.74 | 20 |

**Frequency Table**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Gender** | | | | | |
|  | | Frequency | Percent | Valid Percent | Cumulative Percent |
| Valid | Male | 4 | 30.8 | 30.8 | 30.8 |
| Female | 8 | 61.5 | 61.5 | 92.3 |
| 12 | 1 | 7.7 | 7.7 | 100.0 |
| Total | 13 | 100.0 | 100.0 |  |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Age in Years** | | | | | | | | | |
|  | | | Frequency | | Percent | | Valid Percent | | Cumulative Percent |
| Valid | 2 | 1 | | 7.7 | | 8.3 | | 8.3 | |
| 18 | 1 | | 7.7 | | 8.3 | | 16.7 | |
| 19 | 3 | | 23.1 | | 25.0 | | 41.7 | |
| 20 | 2 | | 15.4 | | 16.7 | | 58.3 | |
| 21 | 2 | | 15.4 | | 16.7 | | 75.0 | |
| 25 | 1 | | 7.7 | | 8.3 | | 83.3 | |
| 32 | 1 | | 7.7 | | 8.3 | | 91.7 | |
| 226 | 1 | | 7.7 | | 8.3 | | 100.0 | |
| Total | 12 | | 92.3 | | 100.0 | |  | |
| Missing | System | 1 | | 7.7 | |  | |  | |
| Total | | | 13 | | 100.0 | |  | |  |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Time to learn Active** | | | | | | | | | |
|  | | | Frequency | | Percent | | Valid Percent | | Cumulative Percent |
| Valid | 5.34 | 1 | | 7.7 | | 8.3 | | 8.3 | |
| 5.43 | 1 | | 7.7 | | 8.3 | | 16.7 | |
| 5.45 | 1 | | 7.7 | | 8.3 | | 25.0 | |
| 6.88 | 1 | | 7.7 | | 8.3 | | 33.3 | |
| 7.44 | 1 | | 7.7 | | 8.3 | | 41.7 | |
| 8.25 | 1 | | 7.7 | | 8.3 | | 50.0 | |
| 8.95 | 1 | | 7.7 | | 8.3 | | 58.3 | |
| 9.67 | 1 | | 7.7 | | 8.3 | | 66.7 | |
| 12.20 | 1 | | 7.7 | | 8.3 | | 75.0 | |
| 12.30 | 1 | | 7.7 | | 8.3 | | 83.3 | |
| 13.30 | 1 | | 7.7 | | 8.3 | | 91.7 | |
| 97.90 | 1 | | 7.7 | | 8.3 | | 100.0 | |
| Total | 12 | | 92.3 | | 100.0 | |  | |
| Missing | System | 1 | | 7.7 | |  | |  | |
| Total | | | 13 | | 100.0 | |  | |  |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Time to Learn Passive** | | | | | | | | | |
|  | | | Frequency | | Percent | | Valid Percent | | Cumulative Percent |
| Valid | 6.16 | 1 | | 7.7 | | 8.3 | | 8.3 | |
| 6.66 | 1 | | 7.7 | | 8.3 | | 16.7 | |
| 7.23 | 1 | | 7.7 | | 8.3 | | 25.0 | |
| 7.33 | 1 | | 7.7 | | 8.3 | | 33.3 | |
| 8.33 | 1 | | 7.7 | | 8.3 | | 41.7 | |
| 10.00 | 1 | | 7.7 | | 8.3 | | 50.0 | |
| 10.20 | 1 | | 7.7 | | 8.3 | | 58.3 | |
| 10.33 | 1 | | 7.7 | | 8.3 | | 66.7 | |
| 12.36 | 1 | | 7.7 | | 8.3 | | 75.0 | |
| 12.50 | 1 | | 7.7 | | 8.3 | | 83.3 | |
| 14.20 | 1 | | 7.7 | | 8.3 | | 91.7 | |
| 15.44 | 1 | | 7.7 | | 8.3 | | 100.0 | |
| Total | 12 | | 92.3 | | 100.0 | |  | |
| Missing | System | 1 | | 7.7 | |  | |  | |
| Total | | | 13 | | 100.0 | |  | |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Order of Mazes** | | | | | |
|  | | Frequency | Percent | Valid Percent | Cumulative Percent |
| Valid | Active First | 5 | 38.5 | 41.7 | 41.7 |
| Passive First | 6 | 46.2 | 50.0 | 91.7 |
| 3 | 1 | 7.7 | 8.3 | 100.0 |
| Total | 12 | 92.3 | 100.0 |  |
| Missing | System | 1 | 7.7 |  |  |
| Total | | 13 | 100.0 |  |  |

1. Explain how you would correct the errors you find.

|  |  |
| --- | --- |
| Variable Name | Modifications |
| Sex | * 12 could be defined as missing value. (“Missing column”) |
| Age | * 226 (outlier value) could be defined as missing value.(“Missing column”) |
| Active Maze Time | * 97.90 (outlier value) could be defined as missing value.(“Missing column”). |
| Maze Order | * 3 should be defined as missing value |
|  | * Line 13 could be deleted as it contains missing values for the columns Age, ActiveMazeTime, PassiveMazeTime, MazeOrder |

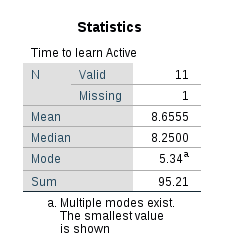
1. Save the corrected data file and submit it along with your answers to homework questions.

**EKRAN GORUNTULERINI YAPISTIR**

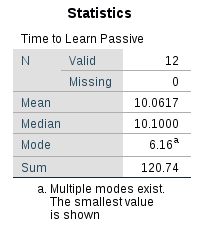
3. Using your corrected data file from question 2 (above), use SPSS to:

1. Find the mean time to learn the maze under active and passive conditions.

Mean for Active Learning Time : 8.6555

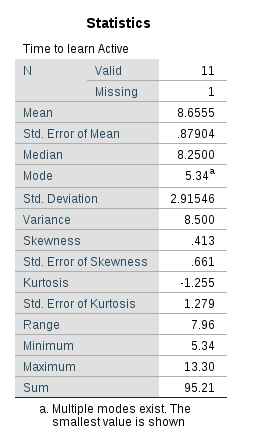


Mean for Passive Learning Time : 10.0617

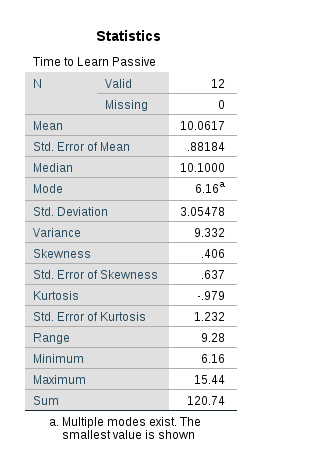


1. Report the remaining descriptive statistics for each condition.

Other descriptive statistics for Time to learn Active:

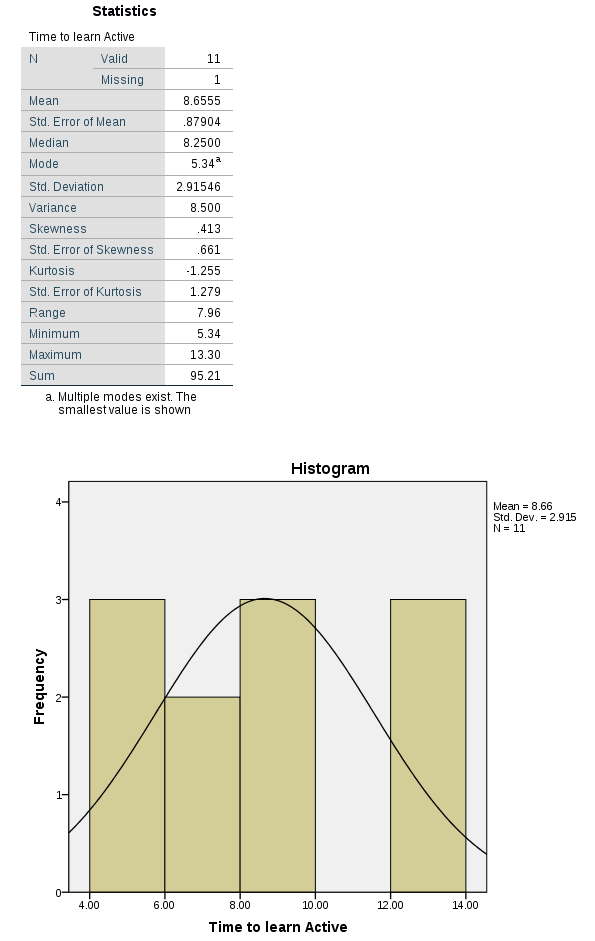


Other descriptive statistics for Time to learn Passive:



1. Provide separate histograms for the dependent variable(s) in this study. Comment whether the distribution exhibits skew and kurtosis.

**Histogram for Time to learn Active :**



Skewness : 0.413

Std. Error of Skewness : 0.661

if skewness / std. Error of skewness > 1.96, we can say skewness is statistically significant.

0.413 / 0.661 = 0.6248 => 0.6248 < 1.96 so; **skewness is not significant, distribution does not exhibit skew.**

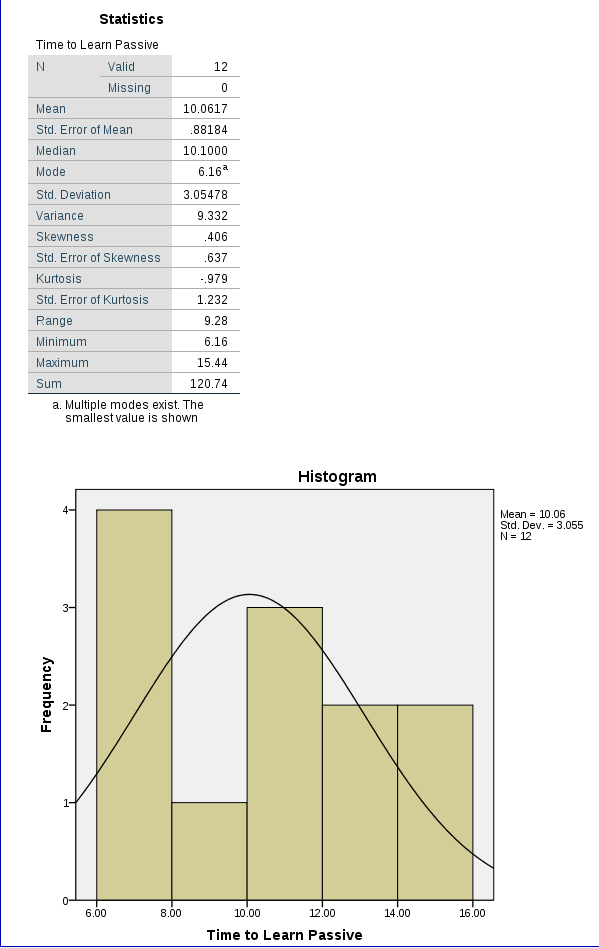
Kurtosis : -1.255

Std. Error of Kurtosis : 1.279

if Kurtosis / std. Error of Kurtosis > 1.96, we can say Kurtosis is statistically significant.

-1.255 / 1.279 = -0.9812 => -0.9812 > -1.96 so; **Kurtosis is significant, distribution exhibits kurtosis.**

**Histogram for Time to learn Passive :**



Skewness : .406

Std. Error of Skewness : .637

if skewness / std. Error of skewness > 1.96, we can say skewness is statistically significant.

.406 / .637 = .6374 => .6374 < 1.96 so; **skewness is not significant, distribution doesn’t exhibit skew.**

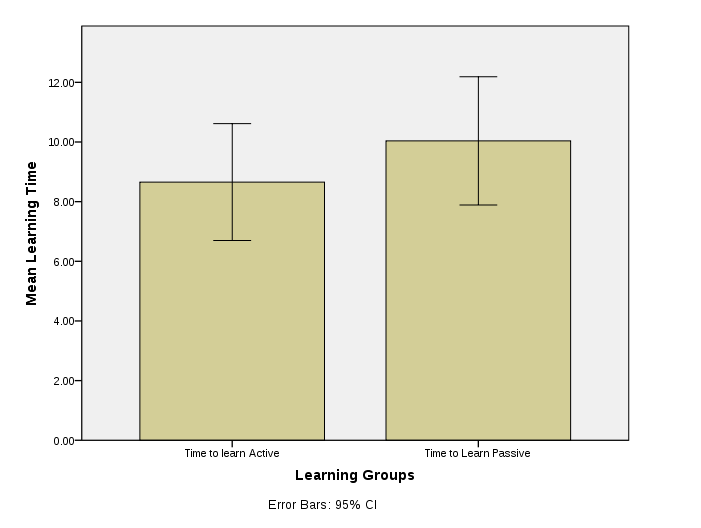
Kurtosis : -0.979

Std. Error of Kurtosis : 1.232

if Kurtosis / std. Error of Kurtosis > 1.96, we can say Kurtosis is statistically significant.

-0.979 / 1.232 = => -0.7946 > -1.96 so; **Kurtosis is significant, distribution exhibits kurtosis.**

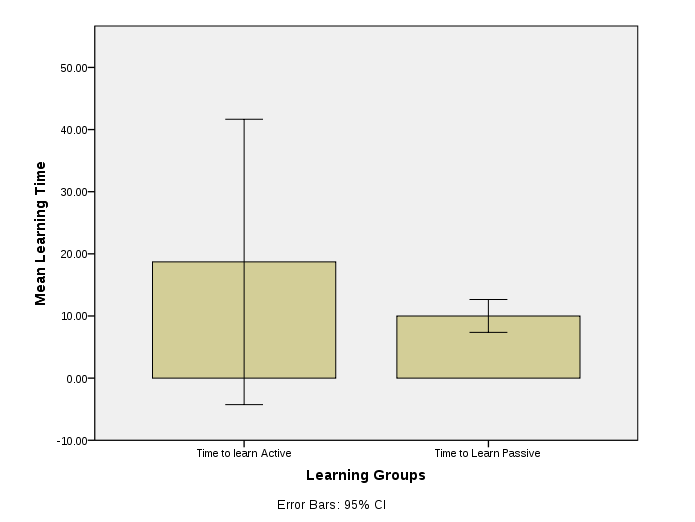
1. Draw a bar chart showing the mean values of each condition with confidence intervals .



1. Label the axes of the chart appropriately. What does this bar chart indicate regarding the outcome of the study?

According to the bar graph, learning time is more in Passive Learning(bar2). To get more reliable conclusions from the grap, we should think on the Error Bars also. As there is not too much variability around the “mean” in our error bars, we can say our graph is reliable and there is nearly no outlier in our data(we eliminated our outliers in second question).

Bar Graph before correcting(eliminating missing values) the data can be seen below:



Comments on the bar chart of the previous data(Before correcting missing values): According to the bar graph, learning time is more in Active Learning(bar1). But to make a more reliable conclusion on the graph, we should think on the Error Bars. When we look at the error bars, we see that; there is much more variability around the mean of Active Learning. The reason for that is the outlier in the data.