
RES5115: RESEARCH PREPARATION: PRINCIPLES AND APPROACHES**QUANTITATIVE ASSIGNMENT****DUE DATE: Sunday 31st of October 2021 before midnight****TASK 1 (20% of marks for unit)**

The manager of a retail store wishes to know if there is a difference in sales performance for staff attending sales training courses and if so, which course provides the best performance. The data file “**salesperformance.sav**” contains sales data for 3 groups of employees with different sales training. Using the data file provided, test the following hypothesis that there is no difference in sales performance:

$$H_0: \mu_1 = \mu_2 = \mu_3$$

against the alternative hypothesis that some groups performed better than others

$$H_1: \text{at least one } \mu_i \neq \mu_j$$

Firstly, you will need to display the data for each treatment graphically to look at the distributions. This can be done in a single diagram (e.g. box plots). Describe the distributions and presence of any extreme values.

Secondly, test the assumptions prior to analysis. If the homogeneity of variance assumption is not met (significance ≤ 0.05) you will need to transform the dependent variable. You should try $y^* = \sqrt{y}$, $y^* = \ln(y)$ and the Box Cox transformation $y^*(\lambda) = \frac{y^\lambda - 1}{\lambda}$, $\lambda \neq 0$ with various values of λ such that $-3 \leq \lambda \leq 3$ using increments of 0.5. Examples using the Box Cox transformation can be found on the internet

<https://www.statisticshowto.datasciencecentral.com/box-cox-transformation/>

Can the assumptions be satisfied using a transformation (if required)?

Thirdly, analyse the raw data or transformed data (if the transformation is successful) to test whether there is sufficient evidence to reject the null hypothesis H_0 . If the results are significant perform a post hoc analysis to determine which means are significantly different to each other. Present your methods, results and findings in the form of a short scientific report. See page 3.

TASK 2 (20% of marks for Unit)

A lecturer wishes to predict students' exam results based on their performances in 8 assignments (quizzes) and 2 tests conducted during the semester. Can students' exam marks be accurately predicted? Which combination of variables provides the best prediction of exam marks? What is the equation to predict exam marks? How much of the variability can be explained by the model?

You should look at the relationships between the independent variables and the dependent variable graphically using scatter plots to help your understanding (there is no need to include these graphs in your report for this task).

Develop a model to predict exam marks from the quiz and test scores. To obtain a parsimonious model only include independent predictor variables where the coefficients are significant. Try various combinations of independent variables by removing variables when the coefficients are not significant at the 0.05 level. You may also use the SPSS "stepwise" method option (to be explained in Module 7 workshop). Make sure you justify your final model selection based on model simplicity and R-square. How much of the variability can be explained by the model?

For your final model save the unstandardized residuals (actual exam score – predicted exam score). Compare the predicted exam marks with actual exam marks using a histogram and Q-Q plot of residuals. Perform a Shapiro- Wilk test of normality on the residuals. What are the assumptions about the distribution of the residuals? What is the range of values for the residuals? Can the residuals be used to identify students where there is a possible breach of academic integrity? Explain.

I am expecting every student to have a different solution. There is no right or wrong answer to this task. Instead, I will be looking at your approach to this problem and your interpretation of the results.

The variables in the "assessments.sav" data file are as follows:

Tst_15Total – score out of 15 for test 1
Tst2_15Total – score out of 15 for test 2
Exam_50Total – Exam mark out of 50
Quiz1TotalPts14 – Quiz1 points out of 14
.
.
Quiz8TotalPts24 – Quiz8 points out of 24

For this task you will need to take a unique sample of the data as follows:

Open the data file in SPSS

Go to **Transform > Random Number Generators**

Select "**Set Starting Point**" and "**Fixed Value**" then enter your student number and click "**OK**".

Go to **Data > Select Cases**

Select: "**Random sample of cases**" and click on "**Sample**".

Set the sample to “**Exactly 150 cases from the first 182 cases**”

Click on: “**Continue**”

Click on: “**OK**”

Save your data.

If you go to the “data window” you will see that cases have been randomly excluded providing every student with a unique set of data for this task. Also be aware the file does contain missing values where students have not completed the assessments.

Present your methods, results and findings in the form of a short scientific report. See below.

Report Preparation

For each task write a short scientific report (4 pages maximum) presenting your results and findings. I will be looking mainly at the methods and results sections.

Your report should have the following structure:

Introduction and Objectives

States the purpose of the investigation.

Methods

State the procedure and applications used. State the hypothesis to be tested and the statistical assumptions.

Analysis and Results

Presentation of the results. Include description of the results.

Discussion

Analysis, interpretation and evaluation of the results. Were the results statistically significant? Were the statistical assumptions met?

Conclusions

What conclusions emerged from the findings and any recommendations.

References

Provide at least 3 references (e.g. Box Cox, SPSS) that confirm to the [APA7 standard](#) for each task. Ensure all references are cited in the text of your report. The reference may be a website.