STAT6123 Generalised Linear Modelling
Supplementary Assignment

- This assignment is worth 50% of the overall mark for STAT6123.

- The deadline for submission is **16.00 on Friday 19 August 2022**.

- Standard University policies and procedures will be followed for late submission, extensions and academic integrity (see the Module Outline for details).

- Submission is via Blackboard.

  - You should submit a report containing your answers in a file called `sup-report-ID.pdf`, where *ID* is your student ID number, for example `sup-report-1234567.pdf`. In the STAT6123 Supplementary Assignment folder, click on STAT6123 Supplementary report submission to submit your report. Please enter this file name as the Submission Title.

  - You should not include R code used in your analysis in your report, but you must submit a separate R script via Blackboard containing your code called `sup-code-ID.R`, for example `sup-code-1234567.R`. This code should reproduce the results contained in your report. Please rename and use the R template `sup-code-yyy.R` provided. In the STAT6123 Assignments folder, click on STAT6123 Supplementary code submission to submit your code.

- Please also email your R script to n.tzavidis@soton.ac.uk.

- The page limits given below for each task are strict and is easily sufficient to receive full credit. Any pages beyond the limits will not be marked.

Task 1 [Total 65 marks, max. 9 pages]

You are employed by the Government Statistical Service of your country and the Head of the Methodology team has tasked you with finding the best possible model for predicting turnover in the current year for businesses in the country for small enterprises. The dataset you have available (`turnover.txt`) comes from a structural business survey. It contains data from 5071 firms on the following variables:

`Turnover2`	Turnover in the current year
`Turnover1`	Turnover in the previous year
`Size`	The number of employees working for the firm. This variable consists of 5 classes: 1 if the firm has one employee, 2 if the firm has 2-4 employees, 3 if the firm has 5-9 employees, 4 if the firm has 10-19 and 5 if the firm has 20-49 employees
`Industry`	A code indicating the industry sector a firm belongs to

1. Present informative descriptive analysis. Justify why the descriptive analysis you present is relevant.

   [12 marks]

2. Model current turnover. Answer the following questions:

   (a) Use the original data and a regression model to describe the relationship between turnover in the current year and turnover in the previous year. Is turnover in the previous year a good predictor of turnover in the current year? Assess the fit of your model and justify your conclusion.

   [6 marks]

   (b) Use appropriate tools and present analysis to assess the validity of (i) the normality assumption and (ii) the assumption of constant variance.

   [6 marks]

   (c) Depending on the results for Part (b), decide whether you need to amend your model. If you decide to do so, discuss how you modified the model and present your analysis. Present diagnostic analyses that compares the modified model to the model from Part (b) and discuss whether the model fit has improved and in what way.

   [9 marks]

   (d) Now that you have decided the best possible model for turnover in the current year as a function of turnover in the previous year, decide whether the size of the firm is a significant predictor of turnover. Justify your decision to whether or not to include the size variable in your model. Briefly discuss the differences between the current turnover for firms with different number of employees.

   [6 marks]

   (e) Decide whether there are significant differences in the current turnover for firms that operate in different industries. Justify your conclusion.

   [3 marks]

(f) Finally, add to your model any term(s) that may improve the its fit.

[6 marks]

3. Summarise the results of your analysis in a separate section. This needs to be written concisely and using non-technical language so that it can be communicated by the National Statistician of your country to the Department for Business, Innovation and Skills.

[12 marks]

4. Up to 5 marks will be allocated for general presentation of the results in the report. Your report should start with Question 1. For Question 2 please use separate, appropriately numbered, subsections for answering Parts (a) to (f). Your report should conclude with the executive summary that responds to Question 3. Please be concise, answer the question you are asked to, justify the analysis tools you use for answering each question and avoid presenting analysis that is not relevant.

[5 marks]

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

Task 2 [Total 35 marks, max. 1 page]

The sinking of the RMS Titanic is one of the deadliest maritime disasters in history. Over 1500 people died as a consequence of a collision with an iceberg. The data file for this part of the assessment is called `titanic.csv` and contains 2099 records on the following variables:

gender	(1 = male, 0 = female)
fare	Amount paid for the ticket
survived	(1 = yes, 0 = no)

Using the theoretical results in the lecture notes and the Fisher-scoring algorithm, write your own R code to estimate a model to study the impact of gender and fare on the probability of survival. **You are not allowed to use existing R functions for estimating the model.** However, you are allowed to use other R functions, for example, those required for matrix algebra computations.

You need to (a) submit reproducible R code using the R template for this task which will be used to replicate your answers and (b) include the answers to the tasks below in your report.

Use your code to <u>compute and present</u>:

1. Point estimates of the model parameters. [10 marks]

2. Estimates of the standard errors of the model parameters. [5 marks]

3. Wald statistics and p-values for hypothesis testing. [5 marks]

4. The value of the model deviance. [5 marks]

5. Compare the empty model to the model that controls for the effect of `gender` and `fare`. Decide whether adding `gender` and `fare` as predictors improves the fit of the model. Present the results. [5 marks]

6. Up to 5 marks will be allocated for presenting a well-structured code. [5 marks]