**Assignment Brief**

This assignment requires you to carry out an investigation of a dataset. You should put your findings in a report with a maximum length of 4 sides of A4. If your report is more than 4 sides, only the first 4 sides will be marked. Your report should include an introduction, tables and charts pasted from Minitab, comments on and interpretation of outputs and results, and conclusions.

**Investigation**

‘Back in Biz Ltd’ is a company that offers online skills training to people in unemployment. The company is in the process of carrying out a review of the training packages it offers. The file ‘**MiniTab Data**’ contains data on 120 of the company’s clients who have recently taken the end of course skills assessment for one of its most popular training courses (each row of the dataset represents a different client). Basic demographic information, age and gender, is included along with key information relating to the training course and includes the following variables:

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable name** | **Description** | **Variable type** | **Coding** |
| Age | Age in years | Scale | N/A |
| Gender | Gender | Nominal | 0 = Female  1 = Male |
| Assessment\_result | Result of the final skills test | Nominal | 0 = Fail  1 = Pass |
| Duration\_days | Time in days between registering for the online training course and taking the final skills assessment | Scale | N/A |

N/A = Not applicable

1. Produce a suitable chart to explore the duration (in days) between registering for the online training course and taking the final skills assessment. Using the chart to guide and justify your choices, obtain appropriate summary statistics to describe this variable.
2. The company is interested to explore if there are gender differences in achievement. Produce a suitable table to compare the two variables ‘gender’ and ‘assessment\_result’ and carefully discuss the results. Is there any suggestion of a gender difference in achievement? Note: you may present your table directly from Minitab or use another package, such as Word or Excel, to construct the table.
3. Produce an appropriate chart to investigate if there is a relationship between age and the duration between registering for the training course and taking the online skills assessment. Calculate an appropriate statistic to describe the strength of association between these two variables, and comment as fully as you can on this association using your outputs.
4. Fit a regression model to investigate whether age is a predictor of the duration between registration and completing the online skills assessment. Clearly interpret the fitted model and discuss the goodness of fit.
5. Calculate the predicted duration between registering for the training course and taking the online skills assessment for a 55 year old. Comment on the validity of the prediction.