1. With the data on blood types:

a. [2 marks] Construct a contingency table that describes the data from the study.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Type A | Type B | Total |
| Present | 5 | 20 | 25 |
| Absent | 35 | 40 | 75 |
| Total | 40 | 60 | 100 |

Contingency Table:

b. [2 marks] Construct a side-by-side bar chart that compares the relative frequency of Type A blood samples that contain the protein to the relative frequency of Type B blood samples that contain the protein. Beside this pair, put another pair that compares the relative frequency of Type A blood samples that do not contain the protein to the relative frequency of Type B blood samples that do not contain the protein.

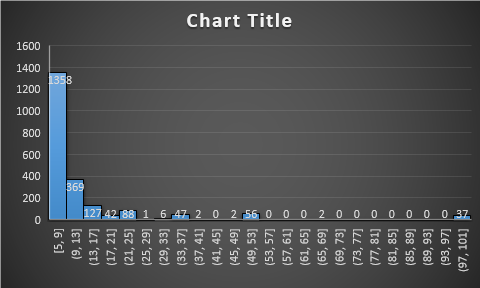
c. [2 marks] Construct a pie chart that shows the proportion of blood samples that contain the protein that are Type A vs those that are Type B. Construct a second pie chart that shows the proportion of blood samples that do not contain the protein that are Type A vs those that are Type B.

d. [2 marks] Is the presence/absence of the protein independent of blood type? Explain.

No, the presence of protein is not independent of the blood type, the blood type does affect whether or not they’re present. We can see that in the pie chart that the two different charts do not mirror each other, so the blood type must have an effect.

2. With the data on flight cancelations:

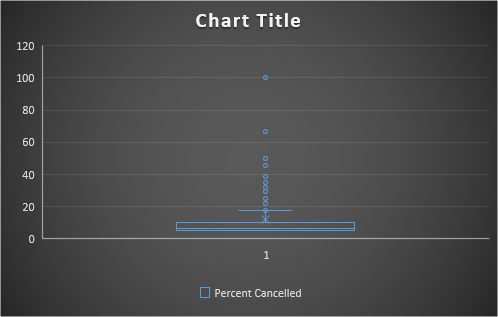
1. [2 marks] Construct a histogram for the variables Percentage Cancelled.



1. [2 marks] Determine the 5-number summary for the variable Percentage Cancelled.

| **Median** | **Range** | **Min** | **Max** | **Q1** | **Q3** |
| --- | --- | --- | --- | --- | --- |
| **6.66** | **95** | **5** | **100** | **5** | **10** |

1. [2 marks] Construct a Boxplot for the variable Percentage Cancelled.



1. [2 marks] Describe the shape, centre, and spread of the distribution. Report on the symmetry, number of modes, and any gaps or outliers. You should mention any concerns you may have about the data.

The majority of the data is in the first column, between 5 and 9%. This graph is not symmetric; it is skewed to the right. The graph is unimodal, and there are many gaps past the cluster at the beginning, leading out to the outliers at 97-100%.