COMP257 2017 Quiz 1

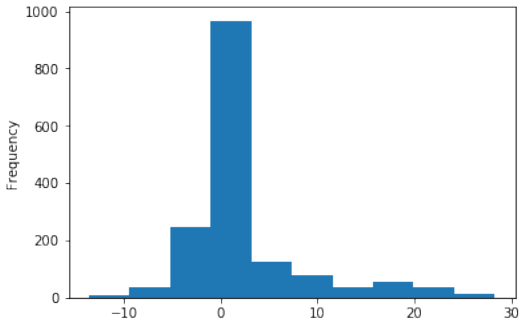
Answer the questions in the spaces provided.
This exam has 7 questions, for a total of 85 points.

Name: _____

Student ID: _____

1. (15 points) In the CRIPS-DM methodology, there is a back-link between the Evaluation phase and the first phase of the process, Business Understanding. Explain what the Evaluation stage should involve and how it might inform the next iteration through Business Understanding.

2. (10 points) The following plot shows the distribution of some data. How would you describe this distribution? Point out any significant features that you observe.



3. (15 points) I'm looking at some data from a drug trial, trying to find out whether the people treated with the drug had different outcomes from those that have not been treated. The null hypothesis is that the two groups have the same outcomess. I perform a *Fisher Test* on the data and get a *p-value* of 0.0322.

   Explain what the *p-value* means in this context. What is the answer to my question (is the drug effective) given the *p-value* of 0.0322?

4. (10 points) When the current Same Sex Marriage postal survey was announced, there were some concerns voiced that the survey would be biased because young people are not familiar with using the postal system. Assuming this to be true (young people don't know how to post letters) explain why this might lead to bias in the result of the survey. Under what circumstances would this not lead to bias?

   (Note this is not a political question – it's about surveys.)

5. (10 points) Explain what the terms in the standard static linear regression model equation below mean?

$$y_i = b_0 + b_1 x + e$$

6. (10 points) Draw a graph showing the simplest representation of the terms in $y_i = b_0 + b_1 x + e$.

7. (15 points) Explain the steps you would take in constructing a static linear regression model with two variables taking into account any data anomalies?