Sample Exam Question

**Section A**

**An analysis of Crab Body Dimensions.**

Male fiddler crabs have **one large claw** (used to attract females) and **one small claw**. These are classified as their **major** and **minor** claws respectively. Major claws can be either on the right or left.

The data in file crab.csv contains the following measurements taken from male crabs that have been sampled from a large crab colony.

**crab.id** an ID number for the measurement

**body.size** the size of the crabs body in cm

**major.claw** the size of the crabs major claw in cm

**minor.claw** the size of the crabs minor claw in cm

**claw.ratio** the ratio of the claw sizes (major/minor)

**major.side** which side the major claw is on (left or right)

We want to explore the data, and in particular be able to determine an expected body size given the values of the other measurements.

Q1i) Is this an example of constructing a model for inference or prediction?

(1 mark)

*inference / prediction* (delete as appropriate)

ii) Download the file crab.csv and write R code to load the data from the file into your environment.

(1 mark)

iii) How many rows are in the dataset?

(1 mark)

iv) What is the mean body size of the crabs sampled?

(1 mark)

v) What is the median size of the major claws recorded?

(1 mark)

vi) What percentage of the crabs in the dataset had their major claw on the right hand side?

(1 mark)

6 marks

Q2i) Write code that deletes the ID column from the loaded data.

(1 mark)

ii) Draw a scatter plot showing body size against major claw size.

Ensure that the axes are labelled as **body size (cm)** and **major claw size (cm)** and the plot has the title **Body measurements of crabs**.

(4 mark)

iii) Construct a linear model to fit **body size** using **major claw size**.

(1 mark)

iv) Discuss how we can interpret the value of the fitted coefficient for **major claw size**.

(1 marks)

v) Use the result of the fit to predict the mean body size of crabs that have major claw size equal to 10cm.

(1 mark)

/8 marks

Total 14

**Crabs are able to regrow a lost claws. It is proposed that this may lead to outliers as regrowing claws will not be at their final size.**

3i) On the scatter plot you have created there is one point (with claw size around 7cm) that appears to be an outlier to the fit .

Identify the row index for this outlier and use this to find its exact body size and claw size.

(2 marks)

Row index: Body size: Claw size:

ii) Is the position of this outlier consistent with the hypothesis that this crab is in the process of regrowing a lost claw? Explain your answer.

(2 mark)

4i) Calculate the coefficient of Pearson’s correlation r between the measurements of major and minor claw sizes.

(1 mark)

ii) Explain why including both these variables in a linear regression model to predict body size could be problematic.

(1 mark)

iii) It is proposed instead to use claw ratio in the model to avoid such problems.

Build a linear model of **body size** fitted against the two predictors **major claw size** and **claw ratio**.

(1 mark)

iv) Examine the information on the fit result. Look at the reported F-statistic.

What does this reported F-statistic relate to?

(1 mark)

v) In this case what does the result tell us? Explain your answer.

(1 mark)

vii) Construct a linear regression model that adds **major side** as a third predictor.

(1 mark)

viii) Is there evidence that including this predictor improves the model? Explain your answer.

(2 marks)

**Section (B)**

*Start a new R file for your answer to this question.*

Load the following file into R: starbucks\_data.csv

Write R code to find:

B1. Total Sales for each day recorded in the dataset.

(1 mark)

B2. The highest number of frappucinos sold in a single day.

(1 mark)

B3. The mean value of sales on Mondays

(1 mark)

B4. Perform a t-test to test the hypothesis that daily sales differ between a Monday and a Friday.

(1 mark)

B5. Interpret the result against the hypothesis based on a significance level of 0.05.

(1 mark)