



## Lab-test-practice - lab test

Statistics (University of Melbourne)

# MAST20005/MAST90058: Computer Lab Practice Test

**Name:**

**Student Number:**

**Tutor's Name:**

**Tutorial Time:**

**Test duration:** 45 minutes

**This test contains 2 questions.** The total number of marks available is 20. The number of marks available for each question is indicated.

## **Authorised Materials:**

- You may bring one double-sided A4 sheet of handwritten notes.
- You may only use the lab machines. You may NOT use your own computers, phones, calculators or other electronic devices.

## **Instructions to Students:**

- Please write your answers on this question paper and submit it at the conclusion of the test.
- Marks will be awarded for answers only. There are no marks for giving the computer commands.
- All data files are available from the shared folder in the computer labs. If you have problems accessing them, please contact a supervisor as soon as possible.
- You are expected to carry out calculations in R or RStudio. If you use other software packages you may not receive any marks for your work.
- You are only allowed to open R or RStudio on the lab machines, and access files from the shared data folder. Accessing the web or other resources during the test will be treated as academic misconduct.

**Question 1 (10 marks)** This question refers to the data in the file `marks.csv`, which shows the final marks obtained by several students for two subjects taken across two semesters. We are interested in exploring the relationship of the marks between the two semesters.

- (a) Give an estimate of the correlation between the two sets of marks.
  
  
  
  
  
  
  
  
  
  
- (b) Fit a simple linear regression model that predicts the marks in semester 2 from those in semester 1. Use the usual parameterisation,  $Y = \alpha + \beta x$ . Give point estimates of all three parameters in the model.
  
  
  
  
  
  
  
  
  
  
- (c) Give 95% confidence intervals for the regression coefficients.
  
  
  
  
  
  
  
  
  
  
- (d) Another student achieved a mark of 75 in semester 1 and was interested in what the model predicts their mark will be in semester 2. Give the mean predicted mark and an 80% prediction interval for this student.
  
  
  
  
  
  
  
  
  
  
- (e) Another student, different to the ones already mentioned, achieved a mark of 70 in semester 2. Give the mean predicted mark and an 80% prediction interval for their mark in semester 1.

**Question 2 (10 marks)** This question refers to the data in the file `yields.txt`, which shows production yields of two different machines ( $A$  and  $B$ ) in a factory, based on independent measurements. Let  $\text{var}(A) = \sigma_A^2$  and  $\text{var}(B) = \sigma_B^2$ .

- (a) Report the 5-number summary statistics (minimum, 1st quartile, median, 3rd quartile, maximum) for the yields from machine  $A$ .
- (b) Use an appropriate t-test to compare the yields of the two machines, assuming that  $\sigma_A = \sigma_B$ . Report the p-value for your test. Using a 5% significance level, state the conclusion from the test.
- (c) Use an appropriate distribution-free test to compare the yields of the two machines. Report the p-value for your test. Using a 5% significance level, state the conclusion from the test.
- (d) Did you get similar results from both of the above tests? Explain why / why not.
- (e) The supplier of machine  $A$  claimed that under normal operation it should produce an average yield of 40. You are doubtful and wish to gather enough evidence to show it is less than this. You know that only a hypothesis test will convince them. Assume the population distribution is normally distributed, has standard deviation equal to the sample standard deviation from these data (from machine  $A$ ). Using an appropriate test with significance level 5%, what is the power of the test if the true mean were 35 and we had a sample size of 20 observations from machine  $A$ ?