

R Intermediate Short Course Bonus Material

Questions and Solutions

(Self Study)

The University of the West Indies, St Augustine

July 2022 Offering
(online)

Question 1

In an experiment to compare the proficiency of two factories the number of defective light bulbs manufactured were recorded. It was found that in factory 1, 2 out of 48 light bulbs were defective, while in factory 2, 4 out of 52 lights bulbs were defective.

Test at a 6% level of significance whether the proportion of defective light bulbs produced in factory 1, is less than the proportion of defective light bulbs produced in factory 2.

Assume the Yates correction for continuity to be true.

Question 2

The CEO of a company suspects that his female employees are more likely to be absent from work than males. After some research, he decided to construct a contingency table using past attendance records. The resulting contingency table is as follows:

	Absent	Not absent
Male	56	80
Female	23	41

Perform an independence test at a 5% level of significance to determine if there is a significant relationship between absenteeism and gender. Assume that Yates correction for continuity is false.

Question 3

The *wtloss* built-in data set in R shows the variation in weight of one patient in a weight rehabilitation program. Using this data, conduct a correlation test at $\alpha = 0.02$ to determine if there is a significant relationship between weight and time (or days).

Recall that the method of testing used depends on the normality of the variables of interest.

Question 4

A coach wanted to examine the total scores of opening batsmen on 3 different pitches after a test match series. The following table shows the scores:

Pitch 1	643	655	702
Pitch 2	469	427	525
Pitch 3	484	456	402

- (i) Use the Kruskal-Wallis test for data given above to determine if there are any differences in the mean scores of batsmen at $\alpha = 0.10$.
- (ii) Make comparisons between the pitches using Tukey's comparison test. What conclusions can you draw?

Question 5

The *sugar.txt* data set shows the sugar content of beet roots grown under 3 treatments of nitrogen: 50 lb, 75 lb and 100 lb per acre. It was thought that soil could also affect sugar content so 5 blocking factors were introduced to indicate plots of land.

Construct an appropriate ANOVA table for the data and use it to test if nitrogen treatment affects the mean sugar content in beet roots at a 1% level of significance.

See course website for the download link to *sugar.txt* data set.

Question 6

The table below shows the pH level, x and arsenic percentage removed, y from 10 samples of water.

x	7.01	7.11	7.12	7.24	7.94	7.94	8.04	8.05	8.07	8.90
y	60	67	66	52	50	45	52	48	40	23

- (i) Determine a fitted regression model for the data.
- (ii) Test the hypothesis $H_0: \beta_1 = 0$ versus $H_1: \beta_1 \neq 0$ at a 5% level of significance, where β_1 is the population slope coefficient.
- (iii) Construct an ANOVA table for the simple linear regression.
- (iv) Conduct tests to determine if the residuals of the regression follow the 3 assumptions: normality, constant variance and independence.