

Data 543 Data Collection

Assignment 2

Submit the pdf/html output of a well-organized RMarkdown/knitr/jupyter notebook document answering the following questions.

Due Date: Friday, December 16, 11:59pm

Total marks: 29 marks

1. Do stimulants or incentives reduce the effect of alcohol? Psychologists asked 44 male college students to memorize a list of 40 words. These students were then randomly assigned to one of four groups (11 students in each group), with each group receiving two drinks. Group A received only the alcoholic drinks; Group AC had caffeine powder dissolved in their drinks; Group AR received alcoholic drinks and also received a monetary reward for correct answers; and Group P (placebo) were told that they would receive alcohol, but instead received two non-alcoholic drinks. After receiving the drinks, the students rested for 25 minutes and then performed a ‘word completion task’. They were scored via the difference between the proportion correctly identified on one list of words and proportion incorrectly identified on a second list of words.

A simulated version of this data is provided in ‘AlcData.csv’ (it uses the summary statistics of the real study). Note that some wrangling may be necessary to answer some of the questions that eventually follow.

- (a) (1 point) What are the experimental units?
- (b) (1 point) What is the response variable?
- (c) (1 point) Which design appears to have been utilized out of these options:
‘Balanced completely randomized design’;
‘Unbalanced completely randomized design’;
‘Balanced randomized block design’;
‘Unbalanced randomized block design’.
- (d) (1 point) Are there any replicates? Explain.
- (e) (1 point) Are there any duplicates? Explain.

- (f) (5 points) Carry out a formal test for difference in means among the treatments. Do so in R, but be sure to include all important information in your answer (hypotheses, assumptions, summary output from R and note the p-value, decision and interpretation)
- (g) (5 points) Use a contrast to test if alcohol leads to a reduction in the mean number of recalled words. Again, do so in R, but include all important pieces of information in your answer. Hint: there are three treatments that receive alcohol and one that does not.
- (h) (3 points) What assumptions were necessary for ‘accuracy’ in the above tests, and does it appear they reasonably hold? Provide some plots/measures/metrics/etc to support your answer here.
2. A seed producer is considering three new varieties of tomatoes using genetic modification. They plan to field test these in the BC Interior using a commonly grown variety as a control. The trial will be for only one growing season. We will use yield, measured in kilograms per hectare, as the response variable. The producer has 30 farm partners throughout the Interior, all of whom have agreed to provide test plots of two hectare size.
- (a) (2 points) How would you set up blocking for this experiment?
- (b) (2 points) Discuss why you would use the blocks described above.
3. Consider the following ANOVA table for a randomized block design:

Source	DF	SS	MS	F	p-value
Treatments	2	20	10	5	?
Blocks	2	20	10		
Residual	4	8	2		
Total	8	48			

- (a) (1 point) Find the p-value for the F-test associated with the question mark in the above table.
- (b) (4 points) Based on the above table, complete the following ANOVA table for a completely randomized design of the same data. (In other words, imagine we had not made use of the blocking variable).

Source	DF	SS	MS	F	p-value
Treatments	?	?	?	?	?
Residual	?	?	?		
Total	?	?			

- (c) (2 points) Suppose all the required assumptions held for both the above tests, and that you (as the researcher) actually cared about whatever this ‘treatment’ is. Comment on the ramifications of the results from (a) and (b).