

Project 8: One-way ANOVA

Statistics 401: Fall 2016

Due Tuesday, December 6

Turn in your answers in a type-written report. Number the problems appropriately.

1. In Montana, the general season for deer and elk always ends on the Sunday after Thanksgiving. On Nov 26, the *Bozeman Chronicle* reported that “hunters in southwestern Montana have been having a tough year as weather conditions allow the states marquee big game animals to stay in the high country for longer” (if you are interested, check out the full story at http://www.bozemandailychronicle.com/news/environment/hunter-success-below-average-going-into-final-weekend/article_c620225a-0a8c-5ec2-ab5a-4016f54d4b18.html). Every hunting season, Montana sets up hunting check stations to count the number of elk and deer harvested by passing hunters. The following table shows the (fictitious) number of harvested deer and elk at 8 different check stations throughout Montana in 2016, and also shows counts of deer and elk at these same check stations in 2015 (when Montana hunters had a record high harvest).

Station	2016	2015
1	54	68
2	75	100
3	53	71
4	56	74
5	74	100
6	71	83

- (a) Do the data support the hypothesis that this year’s deer and elk harvest is less than last year’s harvest? Perform the test BY HAND at a significance level of $\alpha = 0.05$. Include ALL 6 hypothesis test steps including a conclusion in terms of the problem.
 - (b) Use R’s `t.test` function to verify the results of your test. Include the R-code and R-output in the Appendix of your report.
2. Recall the article *In Praise of Gratitude* reported by Harvard Medical School in November of 2011 from problem #2 in project #2 that described how researchers Emmons and McCullough applied an experiment to university students enrolled in class to test their hypothesis that gratitude is the key to happiness. Three groups kept daily journals. The first group kept a “Gratitude Journal”, where the subjects were instructed to write about something that they were grateful for in each entry. The second group kept a “Stress Journal,” where the subjects were instructed to focus on something that they were stressed about in each entry. The last group kept a “Regular Journal” where the subjects were not instructed to focus on anything in particular. The average amount of sleep (in hours) for each group was compared.

The “gratitude.txt” data file on the STAT401 website shows hypothetical data from this experiment. Use R to complete this problem. Attach all R commands and R output used in an appendix. Label all necessary Figures and Tables and refer to these figures and tables from the text of your report.

- (a) Assuming that the experiment was a completely randomized design, how must the experiment have been conducted?
- (b) Give the value of $x_{2,13}$ and explain what this value is in terms of the problem.

- (c) Construct side-by-side boxplots for a visual comparison of the groups (see Chapter 1.6-1.7 notes). Include a figure of the plot in your report.
- (d) Give the null and alternative hypotheses to be tested by a one-way ANOVA.
- (e) Fit the ANOVA model. Include a table that displays the one-way ANOVA table as in the Chapter 5.5 notes.
- (f) From the ANOVA table, find the degrees of freedom DFG and DFE by hand. Show your work!
- (g) From the ANOVA table, find the values of MSG and MSE by hand. Show your work!
- (h) Explain what MSG and MSE are in terms of the problem.
- (i) Check the following two assumptions.
- i. Does the evidence suggest that the data for each group are not normal? Include the a normal probability plot, a smoothed histogram of the studentized residuals, and the correlation test of the studentized residuals to justify your answer. Include the plots in a figure in your report.
 - ii. Check the rest of the assumptions.
 - iii. Does the assumption of constant variance hold? Why or why not?
- (j) Give the distribution of the F statistic assuming that H_0 is true.
- (k) What is your decision regarding H_0 ?
- (l) Assume that the individuals in the experiment are not from a random sample. Second, assume that that the experiment was a CRD. Is it appropriate to make a cause and effect conclusion? Why? To what population can these results be inferred? Give a conclusion in terms of the problem, and mention cause-and-effect and the appropriate population.
3. (a) Is it appropriate to conduct a follow-up test for the gratitude ANOVA? Why or why not?
- (b) If it is appropriate, compute 95% Tukey confidence intervals for the pairwise differences between means. Provide a table of the results of the simultaneous confidence intervals as in the Chapter 15 notes.
- (c) Summarize the results of this study (i.e. give a conclusion). In particular, is there a treatment group which has significantly the longest night's sleep on average? Justify your answer.
- (d) Fill in the following table, label it, and include it in your report.

parameter	estimate	Explanation of the parameter in English in terms of the problem
Estimate for μ_3		
Estimate of σ		