STT 6300: Final Exam

December 4, 2017

Note: The final is due by the end of class on 12/06/2017 (you may drop it off anytime during class on that day). The completed final must be delivered in person (email is not acceptable). Since this is an exam, working together is prohibited, but you are welcome to use the book, course notes, etc. Good luck!

Question 1 (10 points) Suppose we want to compare fasting serum-cholesterol levels among recent Asian immigrants to the United States with typical levels found in the general U.S. population. Suppose we assume cholesterol levels in women ages 21–40 in the United States are approximately normally distributed with mean 190 mg/dL. It is unknown whether cholesterol levels among recent Asian immigrants are higher or lower than those in the general U.S. population. Blood tests are performed on 100 female Asian immigrants ages 21–40, and the mean level ($\bar{x}$) is 181.52 mg/dL with standard deviation = 40 mg/dL. Assuming that levels among recent female Asian immigrants are normally distributed with unknown mean μ, answer the following:

1. What is the population of interest?

2. Using $\alpha = 0.05$, test the null hypothesis $H_0 : \mu = 190$ vs. the alternative hypothesis $H_1 : \mu \neq 190$. Interpret your results using 2–3 complete sentences.

3. Compute a 95% confidence interval for the population mean μ and interpret the results.

Question 2 (15 points) Suppose a sample of eight 35- to 39-year-old nonpregnant, premenopausal OC users who work in a company and have a mean systolic blood pressure (SBP) of 13.86 mm Hg and sample standard deviation of 15.34 mm Hg are identified. A sample of 21 nonpregnant, premenopausal, non-OC users in the same age group are similarly identified who have a mean SBP of 127.44 mm Hg and sample standard deviation of 18.23 mm Hg. Suppose that SBP is normally distributed in the first group with mean μ_1 and variance σ_1^2 and in the second group with mean μ_2 and variance σ_2^2. We wish to test the hypothesis $H_0 : \mu_1 = \mu_2$ vs. $H_1 : \mu_1 \neq \mu_2$. Using $\alpha = 0.05$ and assuming equal variances (i.e., $\sigma_1^2 = \sigma_2^2 = \sigma^2$), answer the following:

1. What is the pooled estimate of the variance from the two independent samples?
2. What is the value of the test statistic t?
3. What is the rejection region for the test?
4. Based on the previous two parts, what is your conclusion regarding H_0?
5. Compute a 95% confidence interval for the true difference $\mu_1 - \mu_2$.
6. Summarize the results using 2–3 complete sentences.

Question 3 (15 points) Run the R script `hematology.R` and answer the following questions:

1. Based on a scatterplot, does a simple linear regression model seem reasonable for these data? Explain your answer using 2–3 complete sentences.

2. Fit a simple linear regression model and write out the estimated equation for the line.

3. Using $\alpha = 0.05$, test the hypothesis $H_0 : \beta_1 = 0$ vs. $H_1 : \beta_1 \neq 0$. Interpret the results of the test in the context of this problem using 2–3 complete sentences.

4. Based on residual diagnostic plots, does it look as though these data satisfy the assumptions of the simple linear regression model? Explain your answer using 2–3 complete sentences.

5. Compute a 95% confidence interval for the slope β_1 and interpret the results.

6. Summarize the results of the analysis using 2–3 complete sentences.

Question 4 (15 points) 22 bypass-patients were randomly divided into 3 treatment groups A, B, and C (different respiration). Interest was in whether or not the average amount of folic acid in red blood cells after 24 hours differs between the three treatment groups. Using the R script `amess.R`, answer the following questions:

1. From graphical inspection, does there appear to be a difference between the means of the three treatment groups?
2. Write out the null and alternative hypotheses.
3. Using $\alpha = 0.05$, test the null hypothesis and state your conclusion based on the p-value of the test.
4. Using Tukey's HSD method, which (if any) of the pairwise comparisons appear to be significant? In other words, which pairs of treatment means appear to be significantly different at the $\alpha = 0.05$ level? Interpret the results using 2–3 complete sentences.
5. Based on residual diagnostic plots, does it look as though these data satisfy the assumptions of the one-way ANOVA model? Explain your answer using 2–3 complete sentences.