Statistics 500 - Homework #5, Fall 2020
Due: by noon Friday, 09/25/2020

Reading Assignment: Statistical Sleuth, Chapters 5 and 6. Review the conceptual exercises at the end of Chapter 5. Solutions to the conceptual exercises are given at the end of Chapter 5.

1. Suppose that we have an experiment for comparison of four treatments (r = 4) groups and we have three independent observations (n_i = 3) from each treatment.

a. For this problem, consider the cell means model.

   (i) Write the four components of the linear model: $\boldsymbol{Y}, X, \boldsymbol{\beta}, \boldsymbol{\varepsilon}$.

   (ii) Show the calculation of the estimate of the parameter vector $\boldsymbol{\beta}$.

b. For this problem, consider the effects model with $\sum_{i=1}^{4} \alpha_i = 0$.

   (i) Write the four components of the linear model: $\boldsymbol{Y}, X, \boldsymbol{\beta}, \boldsymbol{\varepsilon}$.

   (ii) Show the calculation of the estimate of the parameter vector $\boldsymbol{\beta}$.

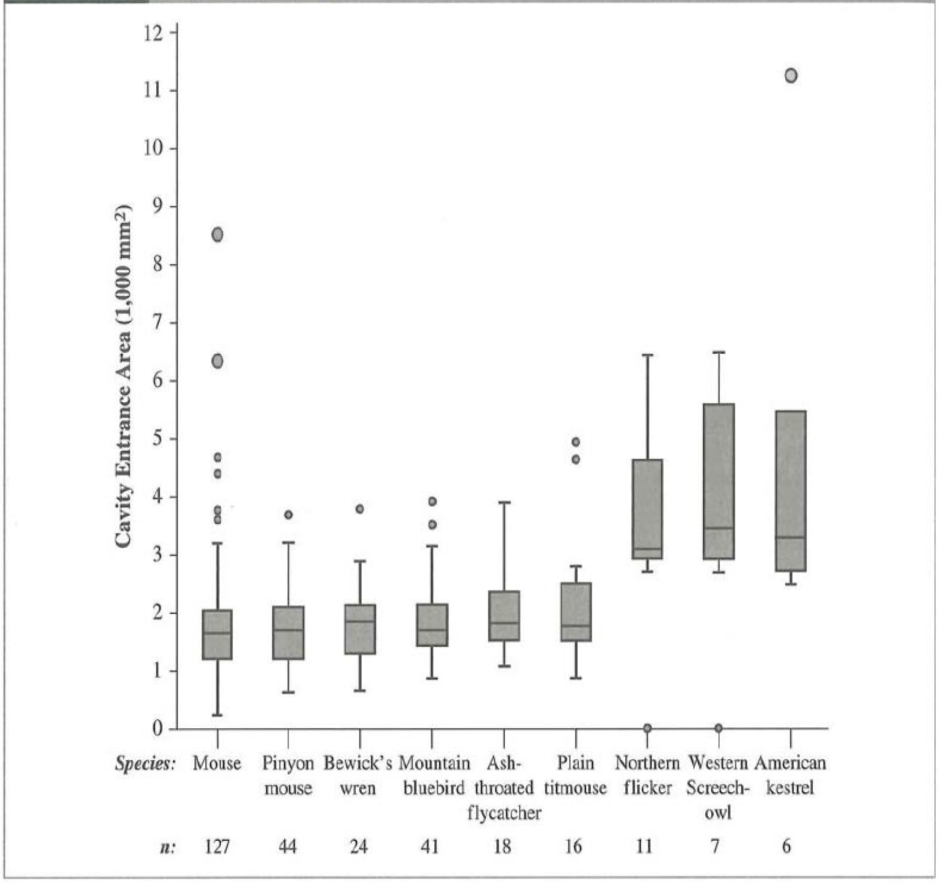
2. Exercise 19 (page 143) at the end of chapter 5 in the *Statistical Sleuth*, describes an observational study of competition among species of birds and rodents for nesting sites in cavities of rocks and trees (data from Donald Youkey, Oregon State University Department of Fisheries and Wildlife). The areas of entrances to nesting cavities were measured for 294 nesting sites for nine common species of birds and rodents in Oregon. The box plots from Display 5.22 on page 144 of the Sleuth are shown on the next page.

   The box plots suggest that variation in areas of cavity entrances is greater for species for which the average areas of entrances are larger. There is also some indication that the distributions of entrance areas are right skewed for most species. This suggests the need for a transformation of the data to promote symmetry and reduce the relationship between the means and variances of entrance areas for the nine species. On a logarithmic scale, the variances are more nearly the same for the nine species and scenes of the distributions of entrance sizes is greatly reduced. Consequently, the following analysis is done with the natural logarithms of the cavity entrance areas. Sample sizes, sample means and sample standard deviations are shown below with samples means and sample standard deviations computed from the natural logarithms of the observed areas of cavity entrances. There is no file with the actual data. You will need to answer the following questions using only the values of the summary statistics show in the table.

Summary Statistics for Natural Logarithm of Areas of Nesting Cavity Entrances

Species	*Sample Size (n)*	*Sample Mean Log(1000 mm^2)*	*Sample Std. Dev. Log(1000 mm2)*
Mouse	127	7.347	0.4979
Pinyon mouse	44	7.369	0.4235
Bewick's wren	24	7.428	0.3955
Mountain bluebird	41	7.487	0.3181
Ash-throated flycatcher	18	7.563	0.3111
Plain titmouse	16	7.568	0.4649
Northern flicker	11	8.214	0.2963
Western Screech-owl	7	8.272	0.3242
American kestrel	6	8.297	0.5842

DISPLAY 5.22 Box plots for areas of entrances to cavities used by different species



a. Compute the pooled estimate of variance for the log-transformed data.

b. Construct an analysis of variance (ANOVA) table for the log transformed data.

c. Perform an F-test of the null hypothesis that means of the natural logarithm of cavity entrance areas are the same for all nine species. Report the value of your F-statistics, its degrees of freedom, and the corresponding p-value.

d. State your conclusion in the context of the study.

e. Show that the model sum of squares $SS_{model} = \sum_{i=1}^{r} n_i(\bar{Y}_{i.} - \bar{Y}_{..})^2$ can be expressed as $SS_{model} = \sum_{i=1}^{r} n_i \bar{Y}_{i.}^{\ 2} - N\,\bar{Y}_{..}^{\ 2}$, where N is the total sample size.

3. Researchers were interested in the effect of a dietary supplement on weight gain in hogs. A total of 12 hogs were used for the experiment. Researchers randomized the 12 hogs so that 3 hogs were assigned to each of 4 treatment groups. Each treatment involved adding a certain amount of the dietary supplement to the daily feed given to the hogs. The 4 amounts considered were 0g, 20g, 40g and 60g of the supplement per kg of feed. The amount of weight gained by each hog was measured in kilograms after 6 weeks. The observed mean weight gains in pounds are provided in the following table.

Amount of Supplement	0	20	40	60
Number of Hogs	3	3	3	3
Mean Weight Gain	22	25	29	32

a. Fill in the missing entries in the following ANOVA table.

Source of Variation	DF	Sum of Squares	Mean Square
Model		174	
Error			28
Corrected Total			

b. Use the entries in your ANOVA table to conduct a test of whether there is a difference among the four group means. Provide the null and alternative hypotheses, calculate the value for the test statistic, give its degrees of freedom, provide the p-value, and your conclusion within the context of this study.

c. For the effects model $Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$, where i= 1, 2, 3, 4 and j =1,2,3, give the parameter vector β when we use the baseline constraint $\alpha_4 = 0$ and give the corresponding design matrix X.

d. With regard to the mean weight gains for the four treatments, what does α_2 represent when the we use the baseline constraint $\alpha_4 = 0$?

4. In 1879, A. A. Michelson made 100 determinations of the velocity of light in air. The data used in this analysis were reported by Stigler (1977, Annals of Statistics, 5:4, 1075). The numbers are in km/sec, and have had 299,000 subtracted from them. The currently accepted "true" velocity of light in a vacuum is 299,792.5 km/sec, but the velocity of light in air could be slower. The data for this exercise were modified by Stigler to correct for overall bias in Michelson's measurement for the speed of light in air. The measurements are grouped into five trials with 20 determinations for each trial. Since each determination was an attempt to measure the same "true" value of the speed of light in air, one might expect that the population means of possible measurements should be the same for all five trials. However, adjustments to the equipment or method for measuring the speed of light in air may have been made between trials, and these may cause the mean values to different across trials. (Note that although one can argue that the trails are random effects, for this problem, let's treat trails as fixed effects.) SAS code that contains the data is posted in the file **lightspeed_data.sas**.

   a. Construct an ANOVA table for these data.
   b. Report the value of the F-statistic for testing the null hypothesis of equal means for the five trials: $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$. Using a Type I error level of 0.05, state your conclusion in the context of the study.

5. Consider the following functions of the population means:

   (i) $\mu_1 + 3\mu_2 - 4\mu_3$

   (ii) $0.3\mu_1 + 0.4\mu_2 + 0.1\mu_3 + 0.1\mu_4 + 0.1\mu_5$

   (iii) $\dfrac{\mu_1 + \mu_2 + \mu_3}{3} - \dfrac{\mu_4 + \mu_5}{2}$

   (iv) $\mu_3 - \dfrac{\mu_4 \times \mu_5}{2}$

   a. Which of the functions are contrasts? Justify your answer.
   b. Are the contrasts that you identified from part (a) orthogonal to each other? Use sample size from problem 4.
   c. Using the data from problem 4, construct a 95 percent confidence interval for each of the functions that you determined to be a contrast.