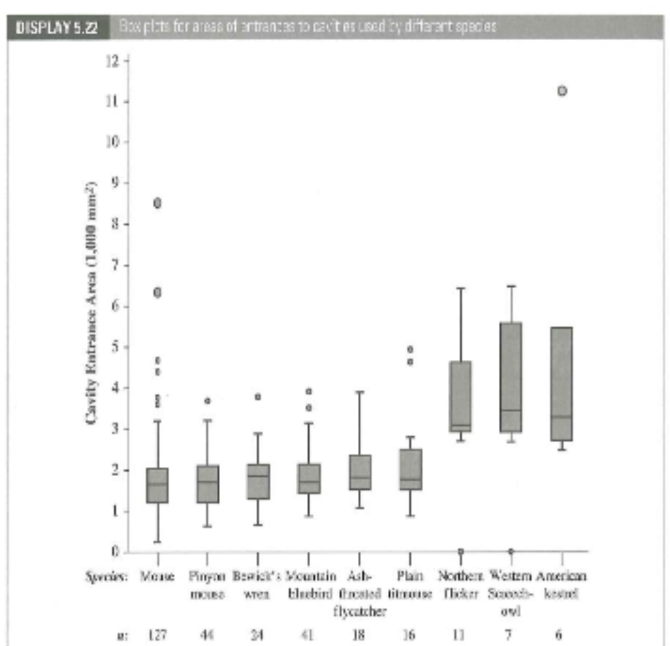
Directions: Type or clearly handwrite your solutions to each of the following exercises. Partial credit cannot be given unless all work is shown. You may work in groups provided that each person takes responsibility for understanding and writing out the solutions. Additionally, you must give proper credit to your collaborators by providing their names on the line below (if you worked alone, write "No Collaborators"):

____

1. [**+30**]**:** 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 boxplots from Display 5.22 on page 144 are shown below.



The boxplots 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.

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

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.

(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 the F-statistics, its degrees of freedom, and the corresponding p-value.

(d) Interpret the test result in the context of the study.

2. **[+40]:** 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	Degrees of Freedom	Sum of Squares	Mean Square
Model		174	
Error			28
Total			

(b) For the cell means model $Y_{ij} = \mu_i + \epsilon_{ij}$, where $i = 1, 2, 3, 4$ and $j = 1, 2, 3$, what is the parameter vector $\boldsymbol{\beta}$, the corresponding design matrix $\mathbf{X}$, and the estimated $\hat{\boldsymbol{\beta}}$ vector?

(c) For the effects model $Y_{ij} = \mu + \alpha_i + \epsilon_{ij}$, where $i = 1, 2, 3, 4$ and $j = 1, 2, 3$, using the baseline constraint $\alpha_4 = 0$, what is the parameter vector $\boldsymbol{\beta}$, the corresponding design matrix $\mathbf{X}$, and the estimated $\hat{\boldsymbol{\beta}}$ vector?

(d) For the model and constraint in part (c), what does α_2 represent in the context of the study?

(e) For the effects model $Y_{ij} = \mu + \alpha_i + \epsilon_{ij}$, where $i = 1, 2, 3, 4$ and $j = 1, 2, 3$, using the sum-to-zero constraint $\sum_{i=1}^{r} \alpha_i = 0$, what is the parameter vector $\boldsymbol{\beta}$, the corresponding design matrix $\mathbf{X}$, and the estimated $\hat{\boldsymbol{\beta}}$ vector?

(f) For the model and constraint in part (e), what does α_2 represent in the context of the study?

(g) For the models and/or constraints given in parts (b), (c), and (e), the ANOVA table will be the same provided that $\mathbf{X}\hat{\boldsymbol{\beta}}$ is the same. Verify that this is true for all 3 models/constraints.

3. [**+30**]**:** The data table below gives the fuel economy of 8 different vehicles of the same car class from four different car companies.

Mercedes	Buick	Peugeot	Volkswagen
23.8	18.6	23.9	40.1
24.7	18.9	23.3	40.1
25.1	19.2	24.1	40.5
24.7	19.2	23.3	40.4
24.5	19.3	24.0	40.9
24.8	19.7	24.7	41.8
25.0	19.7	25.0	42.0
25.0	19.8	25.3	43.0

The data can also be found in the `car.csv` file posted in Canvas.

(a) Construct an ANOVA table for these data in R using the `aov()` function and provide a screenshot of the output.

(b) Examine the associated parameter estimates using the `$coefficients` variable as described during the lab. Which model constraint does R use?

(c) Report the value of the F-statistic and the corresponding p-value for testing the null hypothesis of equal means for the four car companies: $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$. Interpret the test result in the context of the study.

(d) Check the ANOVA assumptions in R and include screenshots of any relevant figures. Are any of the assumptions not satisfied for this study?

(e) Perform a Kruskal-Wallis non-parameteric test in R for the equality of the fuel economy distributions for the four car companies. For the test, report the null and alternative hypotheses, test statistic, p-value, and interpret the test result in the context of the study.

Total: 100 points **# correct:** ____________ **%:** ____________