

**Assignment #11 — Due Friday, December 9 by 4:00 P.M.**

Turn in homework to your TA's mailbox using this sheet as the cover page.

Fill in your name and also circle the *lecture section in which you are registered* and circle the *discussion section you expect to attend* to pick up this assignment.

**Name:**

**Lecture 1 (Hanlon).**      **311:** Tu 1:00 - 2:15pm      **312:** Th 8:00 - 9:15am      **313:** We 1:00 - 2:15pm

**Lecture 2 (Larget).**      **321:** Tu 1:00 - 2:15pm      **322:** We 2:30 - 3:45pm      **323:** We 1:00 - 2:15pm

Please answer the following questions.

1. Independent samples of sizes 10, 15, 25, 20, and 20 are collected. Complete the ANOVA table for this problem.

Source	df	SS	MS	F-statistic	P-value
groups					
error		250.5			
Total		689.6			

2. Complete the ANOVA table for this problem based on the following summary information.

Group	A	B	C	D
mean	53.5	49.7	60.2	59.2
SD	3.21	3.14	4.79	2.64
n	10	6	7	11

3. Independent samples of sizes 10, 15, 20, 20, and 25 are taken. The sample variance of the combined sample of 90 observations is 9.0, and the five sample variances are respectively 4.4, 6.2, 3.7, 5.1, and 3.8. Complete an ANOVA table for this data and use the R function `pf()` to find the p-value. *Hint: use sample size information to find the degrees of freedom, the combined sample variance to find the total sum of squares, the separate sample variances to find the error sum of squares (or residual sum of squares), and then use relationships in the table to complete it.*
4. Use the data set from p.426, Problem 18. But answer these questions.
  - (a) Using the summary of the data provided in the text, complete the ANOVA table for the problem.
  - (b) Compute  $R^2$  for this analysis.
  - (c) Compute a 95% confidence interval for the mean difference between Strain A and Strain B.
  - (d) Use ANOVA to test the equality of means among the four strains. Interpret the results within the context of the problem.

5. Read the background from problem 22 on page 427. The data for this problem is in the file `pinecones.csv`.
- Plot the data with a dotplot for each sample on a common scale. Comment on features that may affect the validity of an ANOVA analysis. Is there strong skewness in the samples? Is the spread greatly different from sample to sample?
  - Use statistical software to fit an ANOVA table for this data.
  - From the table, give a point estimate for the common standard deviation of pine cone masses.
  - Using  $t$ -methods and without concern for simultaneous inference (or multiple testing), find 95% confidence intervals for the differences between mean cone size between:
    - Island, squirrels absent & Island, squirrels present;
    - Mainland, squirrels present & Island, squirrels present.
  - If one found simultaneous 95% confidence intervals using Tukey's method, would they be wider, the same size, or shorter than the intervals in the previous part? Briefly explain.

If you use R for these calculations, here is some guidance.

```
# To read in data, use read.csv()
> pine = read.csv("pinecones.csv")
> str(pine)
# To graph the data, load the lattice library.
# We will sort the factor habitat so that the Island, squirrels present group comes first,
# as this is the group to be directly compared to the others.
# As the mean pine cone mass is smallest for this group,
# we can use the reorder() function from lattice.
> library(lattice)
> pine$habitat = reorder(pine$habitat,pine$conemass)
# Here are two graphs to do the same thing, but to label the groups differently.
# Use xyplot() with group as the y-variable and conemass as the x-variable
> plot( xyplot(habitat ~ conemass, data=pine) )
# Use dotplot and the | syntax to split conemass by the group habitat in different panels.
# Also use layout to create an array of 1 row and 3 columns.
>
# Fit an ANOVA model
> pine.lm = lm(conemass ~ habitat, data=pine)
> anova(pine.lm)
# Use summary to show estimates for some differences in group means with SEs.
> summary(pine.lm)
# Verify that the SEs agree with textbook formulas
# These first lines calculate SSE and MSE using residuals()
> sse = sum(residuals(pine.lm)^2)
> mse = sse / pine.lm$df.residual
> sqrt(mse) * sqrt( (1/6) + (1/5) )
> sqrt(mse) * sqrt( (1/5) + (1/5) )
# Here is the t multiplier for 95% confidence intervals
# using the residuals (error) degrees of freedom
> t.crit = qt(0.975,pine.lm$df.residual)
> t.crit
```