

QERM 598 - HW 3
Due January 30, 2008
Eli Gurarie

Introduction to ANOVA

1. The following set of problems refers to the lecture notes, relying in particular on the “Math Facts” presented in the preface.
 - (a) Discuss briefly the advantages and disadvantages of the two types of plots used to present the data on page 4.
 - (b) Show how equation (6) leads to equation (7).
 - (c) Derive expression (9).
 - (d) Use simulations in R to demonstrate the relationship between iid standard normal variables and the F -distribution.
2. Get two samples of data (either from a previous homework, the internets, or generate your own) and perform a two-sided t -test and a single-factor ANOVA on the data and present the results. The p -values should be identical. Explain how these two tests are mathematically equivalent.
3. Analysis of birth timing for Steller sealions on rookeries in Russia: In 2005, researchers situated on several reproductive sea-lion rookeries in Russia (Figure 1) observed the birth of pups (Figure 2) of forty randomly selected sealions on each of six reproductive rookeries. There is some interest as to whether the birth timing of the sealions is different across these rookeries, which span about 1000 km in latitude. The data is located in file “SeaLionBirth.dat”.
 - (a) Create a boxplot of the results of birth-timing. Does it look from the boxplot that island is an important factor is predicting birth timing?
 - (b) Formulate two statistical models that predict birth timing of sealions, one that excludes the effect of island and one that includes the effect of island. Identify the number of parameters in each model. Make sure to specify the indices and their range of values. What are a , n , and N ?
 - (c) Obtain the group means and sample variances for the six islands.
 - (d) Obtain the Total Sum of Squares, the Error Sum of Squares and the Treatment Sum of Squares.
 - (e) Obtain the Mean Square Error and Treatment effects.
 - (f) Obtain an statistic that will allow you to test the hypothesis of significant island effect. What is the distribution of this statistic under the null hypothesis? Obtain a p -value from comparing this statistic to a null-distribution. Plot this distribution, and show the value of the statistic on your plot.

- (g) Fill out the components of the single-factor fixed-effects ANOVA table.
- (h) Obtain an ANOVA table in one or two lines using the “`lm`” and “`anova`” commands.
- (i) Show some diagnostic plots to assess the validity of the ANOVA assumptions.
- (j) What are your conclusions regarding the effect of Island on the birth timing of Steller sea lions in Russia?
- (k) Extra (virtual) Credit: Read about Tukey’s honestly significant differences test in an applied statistics text, run a multiple pairwise comparison using the Tukey’s HSD “by hand” and compare it to R’s output using the “`TukeyHSD`” function. Formulate your anova table using “`aov`” rather than “`lm`” and “`anova`”. Can you make more nuanced conclusions based on this test?

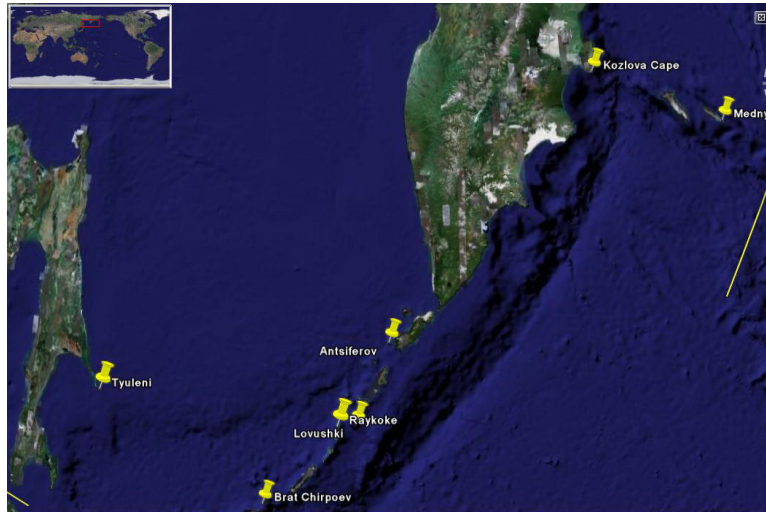


Figure 1: Locations of seven major Steller sea lion rookeries in Eastern Russia



Figure 2: Steller sea lion female giving birth.