Click to prove you're human

Tukey kramer test in r

This tutorial is part of a series designed to accompany a course using The Analysis of Biological Data. The rest of the tutorials can be found here. This tutorial is based on topics in Chapter 15 of ABD. Use R to perform analysis of variance (ANOVA) to compare the means of multiple groups. Perform Tukey-Kramer tests to look at unplanned contrasts between all pairs of groups. Use Kruskal-Wallis tests to test for difference between groups without assumptions of normality. If you have not already done so, download the zip file containing Data, R scripts, and other resources for these tutorials. Remember to start RStudio from the “ABDLabs.Rproj” file in that folder to make these exercises work more seamlessly. 1. The European cuckoo does not look after its own eggs, but instead lays them in the nests of birds of other species. Previous studies showed that cuckoos sometimes have evolved to lay eggs that are colored similarly to the host bird species’ eggs. Is the same true of egg size – do cuckoos lay eggs similar in size to the size of the eggs of their hosts? The data file “cuckooeggs.csv” contains data on the lengths of cuckoo eggs laid in the nests of a variety of host species. Here we compare the mean size of cuckoo eggs found in the nests of different host species. Plot a multiple histogram showing cuckoo egg lengths by host species. Calculate a table that shows the mean and standard deviation of length of cuckoo eggs for each host species. Look at the graph and the table. For these data, would ANOVA be a valid method to test for differences between host species in the lengths of cuckoo eggs in their nests? Use ANOVA to test for a difference between host species in the mean size of the cuckoo eggs in their nests. What is your conclusion? Assuming that ANOVA rejected the null hypotheses of no mean differences, use a Tukey-Kramer test to decide which pairs of host species are significantly different from each other in cuckoo egg mean length. What is your conclusion? 2. The pollen of the maize (corn) plant is a source of food to larval mosquitoes of the species Anopheles arabiensis, the main vector of malaria in Ethiopia. The production of maize has increased substantially in certain areas of Ethiopia recently, and over the same time period, malaria has entered in to new areas where it was previously rare. This raises the question, is the increase of maize cultivation partly responsible for the increase in malaria? One line of evidence is to look for an association between maize production and malaria incidence at different geographically dispersed sites (Kebede et al. 2005). The data set “malaria vs maize.csv” contains information on several high-altitude sites in Ethiopia, with information about the level of cultivation of maize (low, medium or high in the variable maize_yield) and the rate of malaria per 10,000 people (incidence_rate_per_ten_thousand). Plot a multiple histogram to show the relationship between level of maize production and the incidence of malaria. ANOVA is a logical choice of method to test differences in the mean rate of malaria between sites differing in level of maize production. Calculate the standard deviation of the incidence rate for each level of maize yield. Do these data seem to conform to the assumptions of ANOVA? Describe any violations of assumptions you identify. Compute the log of the incidence rate and redraw the multiple histograms for different levels of maize yield. Calculate the standard deviation of the log incidence rate for each level of maize yield. Does the log-transformed data better meet the assumptions of ANOVA than did the untransformed data? Test for an association between maize yield and malaria incidence. 3. Animals that are infected with a pathogen often have disturbed circadian rhythms. (A circadian rhythm is an endogenous daily cycle in a behavior or physiological trait that persists in the absence of time cues.) Shirasu-Hiza et al. (2007) wanted to know whether it was possible that the circadian timing mechanism itself could have an effect on disease. To test this idea they sampled from three groups of fruit flies: one “normal”, one with a mutation in the timing gene tim01, and one group that had the tim01 mutant in a heterozygous state. They exposed these flies to a dangerous bacteria, Streptococcus pneumoniae, and measured how long the flies lived afterwards, in days. The date file “circadian mutant health.csv” shows some of their data. Plot a histogram of each of the three groups. Do these data match the assumptions of an ANOVA? Use a Kruskal-Wallis test to ask whether lifespan differs between the three groups of flies. A one-way ANOVA is used to determine whether or not there is a statistically significant difference between the means of three or more independent groups. If the overall p-value from the ANOVA table is less than some significance level, then we have sufficient evidence to say that at least one of the means of the groups is different from the others. However, this doesn’t tell us which groups are different from each other. It simply tells us that not all of the group means are equal. In order to find out exactly which groups are different from each other, we must conduct a post hoc test. One of the most commonly used post hoc tests is Tukey’s Test, which allows us to make pairwise comparisons between the means of each group while controlling for the family-wise error rate. This tutorial explains how to perform Tukey’s Test in R. Note: If one of the groups in your study is considered a control group, you should instead use Dunnett’s Test as the post-hoc test. Example: Tukey’s Test in R Step 1: Fit the ANOVA Model. The following code shows how to create a fake dataset with three groups (A, B, and C) and fit a one-way ANOVA model to the data to determine if the mean values for each group are equal: #make this example reproducible set.seed(0) #create data data