

Problem Set #3: Multiple Regression and ANOVA

(1) (10 pts) Camdilla was a student in Tim Karel's lab studying rodent populations in the CA desert. She wanted to figure out what determines the density of populations of the Desert Kangaroo Rat. She suspects that food, shelter, and predators play a role. She estimates the density of rats in 50, one-hectare plots. She estimates shrub cover (i.e. amount of shelter), annual seed production (i.e. amount of food), and snake density (i.e. predator threat). She tried three separate linear regressions to test the individual effects of shelter, food, and predators on the density of rats. However, she also tried multiple regression to see if it reveals more on the ecology of this system. Show findings from the three separate linear regressions. What are the results and how do they differ?

Rats x Shelter: The r and r squared (79%) show us a positive correlation between the rats and shelter. Both parametric p -values <0.05 , this relationship is statistically significant.
 $r = 0.8931335$
 $r\text{-square} = 0.7976875$

Rats x Food: The r and r squared (75%) show us a positive correlation between the rats and shelter. Both parametric p -values <0.05 , this relationship is statistically significant.
 $r = 0.8675806$
 $r\text{-square} = 0.7526961$

Rats x Predation: The r and r squared (5%) show us a weak negative correlation between the rats and shelter. The two-tailed parametric p -value shows it's not statistically significant but the 1 tailed does.
 $r = -0.2421115$
 $r\text{-square} = 0.05861797$
2-tailed $p = 0.09026689$
1-tailed $p = 0.04513345$

Multiple Regression: The best model for this was rat density x predation x food and the second best was with all of the variables. Because we are interested in an ecosystem where all variables are at play I chose to report results from the model with all variables.

Results: Show a statistically significant relationship between the variables and rat density per quadrant. All p -values were <0.001 with food being the strongest predictor of density. It had a p -value of 2.2×10^{-16} . Food also had an F -value of 222.257 in comparison to shelter and predation which had F s around 13, confirming the significance of the variation.

Discussion: The multiple regression shows us that food has a stronger correlation with rat density than shelter which was the top determining factor in the linear regressions.

(2) (5 pts) A plant ecologist wishes to test the hypothesis that the height of white sage depends on properties of the soil. She gathers soil from three sites (a sand dune, a site in the SMMountains, and chaparral) and also used, sterilized potting soil. The soils vary in nutrients, grain size, and many other characteristics. She plants five individual plants in each soil type in separate pots in the greenhouse. After four months, she measures the height of each plant (in cm). Do an ANOVA by

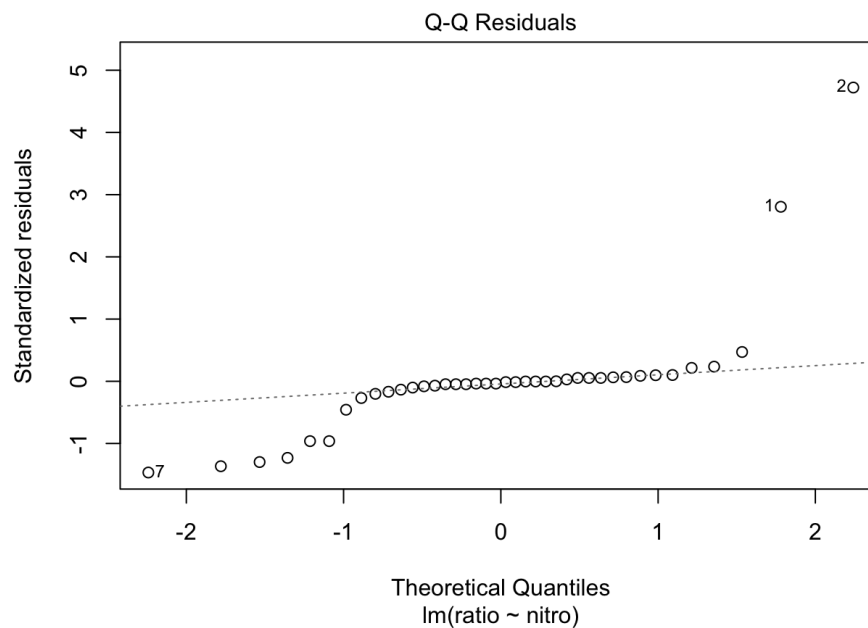
“hand”, using Excel, (do not use R) to test the hypothesis that these four soil types affect the size of coastal sage scrub. Assume that the data are normal and homoscedastic, so you don’t need to test for these. Write a single sentence describing the results (include F, df, and p).

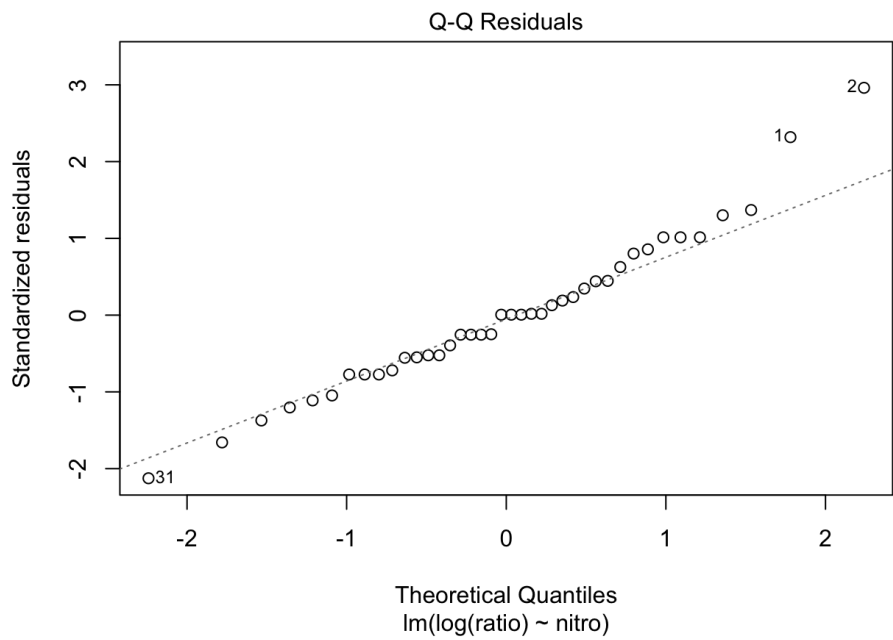
Sterile Soil	Sand Dune	Mountains	Chaparral
11.3	11.1	9.9	12.9
11.4	12.3	10.7	13.3
9.7	11.5	11.5	13.1
10.5	13.1	11.8	14.7
10.1	12.5	10.9	14.1

The one way ANOVA shows a statistically significant effect of soil type on plant, confirmed by an F- value of 16.01 at 3 degrees of freedom and a p-value of 0.00004.

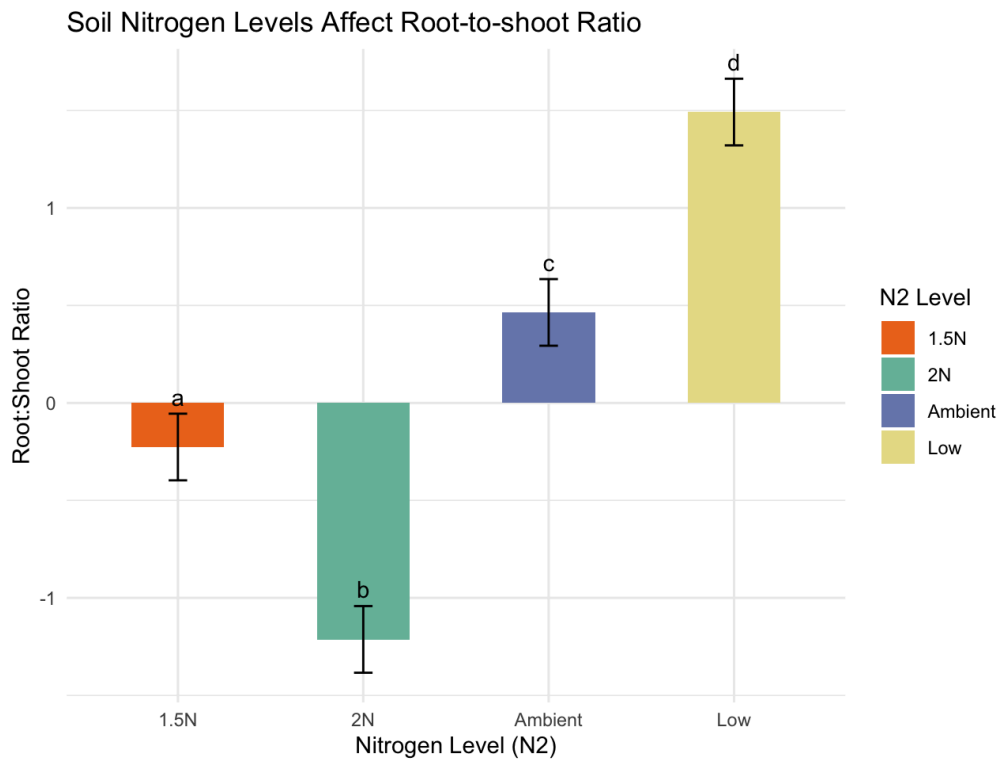
(3) (10 pts) Anna is interested in whether the nutrient content of soils influences the root:shoot ratio of the plants growing in it. She tests the null hypothesis of no difference in ratios under four different soil nutrient conditions that she has manipulated. Do the data meet the ANOVA assumptions? Is transformation required? Show diagnostic plots.

Non normal data ([qqplot 1](#)). Used log transformation on rootshoot numbers ([qqplot 2](#)).





Use ANOVA to determine whether N₂-level soil significantly affects the root-shoot ratio of this plant. Use Tukey HSD to determine which treatments differ from each other. Interpret the data and present the ANOVA results in a sentence. Provide a publication-quality graph that shows the results of the Tukey tests with letters above different bars.



The one-way ANOVA shows a statistically significant effect of nitrogen level on root-shoot ratio, confirmed by an F- value of 44.474 at 3 degrees of freedom and a p-value of 3.405e-12.

(4a) (7 pts) Granivorous desert rodents may influence plant densities by consuming seeds of plants. To test this hypothesis, plots surrounding individual plants were established as unmanipulated controls, rodent exclosures (using fine mesh wire fences buried 2 feet in the sand and 3 feet above ground), and a fence control treatment. The fence control had the same fencing as the exclosures, but holes in the side still allowed rodents to enter. The data in the file "seedsrodents" are the seed mass (in grams) of individual plants in the different treatments. Use planned comparisons (*a priori* contrasts) to test two hypotheses: (1) there is no experimental artifact of using exclosures; and (2) rodents reduce the abundance of seeds (as estimated by mass of seeds). Give the coefficients that you used for each treatment for each contrast. Carry out the analyses and interpret the data in the context of the contrasts you designed above.

Exclusion effect $c_1 \leftarrow c(1,0,-1)$. Rodent Effect $c_2 \leftarrow c(1,-2,1)$. The split summary of my model shows that Exclusion, Rodents and general Treatment all have a statistically significant (p-value < 0.05) effect on the abundance of seeds. Exclusion was the most significant with an F- value of 153.44 compared to 13.69 for rodents and 83.57 for overall treatment.

(4b) (3 pts) Just so you can see how the results can differ, also do an ANOVA, followed by Tukey HSD post-hoc tests to make all possible pairwise comparisons of the three treatment means. Based on this *post-hoc* approach, would your interpretation of the experiment differ at all from that obtained by the planned comparison approach?

No it does not

ANOVA: Shows the same result as overall treatment in the previous step. No difference.

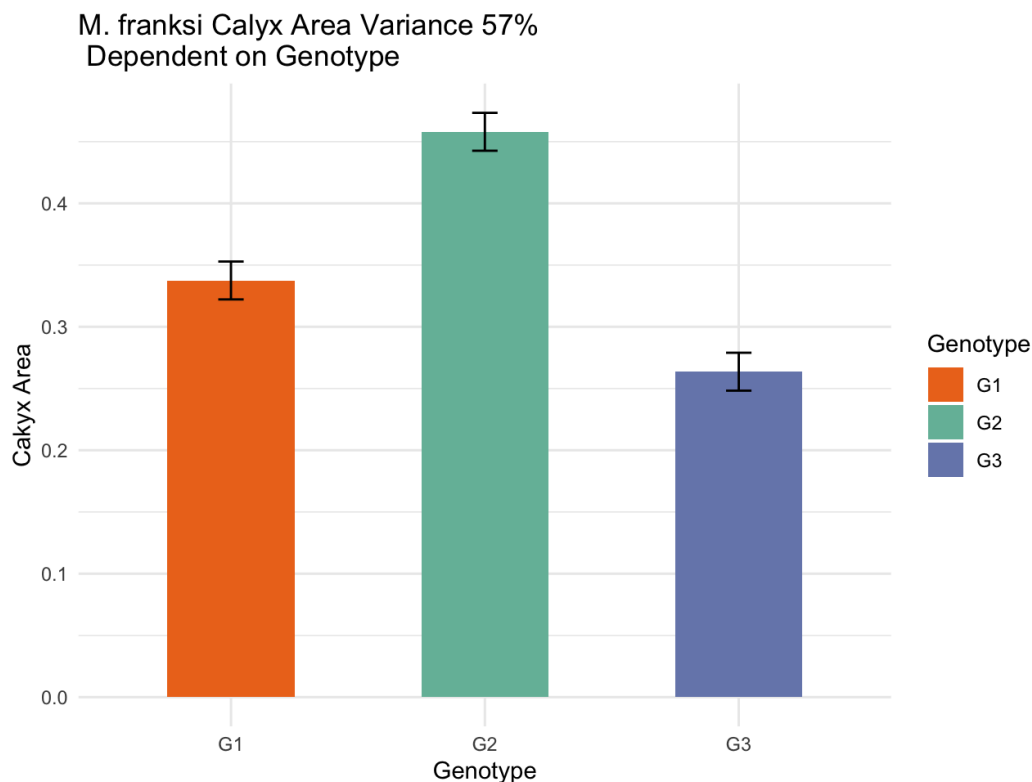
Tukey: Shows the same result as the previous step with emphasis on Exclusion having the strongest effect on seed mass.

(5) (7 pts) Carla studied morphological variation in the Caribbean coral *Montastraea franksi*. She wanted to determine whether colonies of this coral differ in morphology. A morphological feature that can vary within coral species is the calyx – the portion of the skeleton where the polyp lies. Carla's null hypotheses were that calyx area does not differ either (1) within or (2) among colonies of *M. franksi*.

To test her hypothesis, she randomly chose 3 colonies (genets), which she broke into 3 pieces each (ramets). All of the ramets from each colony were placed out in the field and allowed to grow, thus making them independent from one another. However, ramets of the same genet are genetically related, so not fully independent. After 8 months, the ramets were collected for morphometric analysis. She took 10 measurements of calyx area from each ramet. The results are in the csv file "montastraea". Use a nested model to quantify where most phenotypic variation in calyx area lies in populations of *M. franksi* using variance components. What are the percentage variances explained by Genets and Ramets?

Ramets 0%, Genets 57.04%

Make a publication-quality graph with a legend that shows differences among genets.



(6) (8 pts) Ananda conducted an experiment to test the effects of temperature and CO₂ (pCO₂) on growth rates of recently settled corals to evaluate the possible effects of these two stressors on a species of coral. She used two levels of temperature (ambient and high) and two levels of pCO₂ (ambient and high) in a crossed (orthogonal) design. For each of the 4 treatment combinations, she had 4 replicate tanks, each containing 10 recently settled corals. She measures the growth in basal area (mm²) of each coral from digital photographs taken at the start of the experiment and at the end. To avoid pseudoreplication, she averages growth among all corals in each tank and treats that average as a replicate. The data are in the csv file “temp_CO2”.

As always, evaluate whether the data meet the assumptions of the test and consider transforming them if they don't.

Make a publication quality bar graph (with legend) that summarizes the results.

Conduct a two-way ANOVA with the factors temperature and pCO₂. Present and interpret the statistical results in light of the biological questions.

ANOVA: Shows that temperature has a significant effect on coral basal growth. CO₂ alone does not. The F- value for Temp alone was 6.9761 compared to 0.0383 for CO₂ and 0.9569 for the combination. Temp had a p-value of 0.02152.

