

ANOVA vs. Regression: When to draw the line

This project will focus on the difference between linear models that use categorical or continuous predictor variables. When designing experiments, scientists are often limited by the number of experimental units they can use. These practical limitations may arise due to the cost of using animal subjects, the number of tanks available for aquatic mesocosm studies, or simply the number of hours in the day. This question will focus on the optimal use of these precious experimental units to increase our “statistical power” to detect the effect of a treatment.

A experimental design question that arises when experimental units are limiting is whether to distribute the units amongst some number of replicated discrete treatment levels (ANOVA-design) or spread the experimental units along a continuous gradient of treatment levels (regression-design). In this project, we will first revisit ANOVA and regression analyses in two cases where it is clear what experimental design was chosen. We will then evaluate the ability of these two experimental designs to detect a treatment effect (i.e. their statistical power) using simulated data.

An ANOVA-design and a regression-design experiment

1) A student conducted an experiment evaluating the effect of three different new antibiotics on growth of *E. coli* in lab cultures. Using the data in `antibiotics.txt`, generate a plot that summarizes the results and test for an effect of antibiotic treatments on the growth of *E. coli* using an ANOVA-design linear model and likelihood ratio test.

2) Another student conducted an experiment evaluating the effect of sugar concentration on growth of *E. coli* in lab cultures. Using the data in `sugar.txt`, generate a plot that summarizes the results and test for an effect of sugar concentration on growth of *E. coli* using a regression-design linear model and likelihood ratio test.

A statistical power analysis

Experimentalists and statisticians will often use simulated data to help design an experiment. In these analyses, the investigator generates data that will look like the expected results of an experiment assuming some amount of variance or error. The investigator can then ask how many experimental units would be required to detect a significant effect, assuming the strength of the relationship and error assumed when simulating data. We will use this approach to ask whether ANOVA- or regression-design experiments are more powerful in detecting a continuous relationship between two variables.

Assume that an independent variable x is linearly related to a dependent variable y with a slope (β_1) of 0.4 and a y-intercept (β_0) of 10. Imagine that we have 16 experimental units that can be used to test for an effect of x on y . One could randomly distribute the experimental units along a gradient of x (regression design) or one could replicate two levels of x 12 times each and ask whether low or high levels of x generate different levels of y (t-test design). An intermediate approach would be to have 8 replicates of three levels of x or 4 replicates of six levels of x and so on (ANOVA design).

Using the relationship between x and y defined by $\beta_1 = 0.4$ and $\beta_0 = 10$, we can generate a random dataset with 24 observations of y for any experimental design of interest given some level of error in our observations. We define our error by the standard deviation (σ) of normally distributed errors that we can add to our simulated values of y . Because our results will vary a bit from one simulated data set to another, it is a good idea to adopt something called a *monte carlo* approach where we run multiple (say 10) versions of a power analysis for a given experimental design and σ .

To evaluate the relative statistical power of regression- and ANOVA-design experiments, simulate 10 random experiments with a regression design and 10 random experiments with a two-level ANOVA-design (really this is a t-test design) for each of eight values for σ (1, 2, 4, 6, 8, 12, 16, 24). Let's say you are able to generate experimental units with x between 0 and 50. Repeat this process for a four- and eight-level ANOVA design (remember you only have 24 experimental units). Use the average p-value from likelihood ratio tests across your monte carlo runs as your metric of statistical power.

How does the ANOVA- vs regression-design perform? Does the relative performance of these experimental designs depend on the number of levels in the ANOVA experiment (2, 4, vs. 8)?