

R can be intimidating for a many who are accustomed to programs such as SPSS or Excel or even STATA. This is because you have to tell it (R) almost everything to do. Additionally, some begin to use R for statistical analysis without a good idea of the types of questions different techniques are designed to address - such as using a linear regression to model data that is discrete (0,1). Below I have included some basic R Commands for the beginning Analyst along with the questions they are designed to answer, i.e. t-tests to linear regression. I do not deal a lot in the assumptions of the techniques but do provide brief commentary where absolutely necessary.

The following require MASS, foreign, polycor, and psych packages, e.g. require(MASS)

Also, quick means of loading .csv data files using the "foreign" library: data<-read.csv(file="~/somedata.csv")

Did you do something to one group but not the other, e.g. is this an experimental and control group?

If **Yes** then **paired t-test**: t.test(y1,y2,paired=TRUE) # where y1 & y2 are numeric vectors

If **No** then **independent samples t-test**: t.test(y1,y2) # if you have more than one group, an ANOVA is preferred

Do you want to know if the difference is based on some classification such as belonging to a certain group or prior performance?

One Way ANOVA: fit <- aov(y ~ A, data=mydata) # fit is the name of the model, you will want to name each model differently. Also note, ANOVA gives you a variable F indicating whether there is a difference, but not the direction of the difference.

1. Fixed-effects models which assume that data from normal populations that differ in their means allows the estimation of the range of response that any treatments towards them will generate.
2. Random-effects models which assume that data from a constrained hierarchy of different populations are sampled with different factor levels.
3. Mixed-effects models which describe the situations where both fixed and random effects are present.

Do you want to know the direction of a continuous variable in relation to a dependent variable for two different groups?

ANCOVA: fit <- aov(y ~ A + x, data=mydata) # ANCOVA interpretations depend largely on the data that are entered. If you get to the point of having to use an ANCOVA you may as well use a regression because the effects are more easily interpreted and have the same assumptions. You can test these assumptions in regression though, but the tests for ANCOVA are less accepted.

Do you want to know if there is a directional relationship between two continuous variables?

Correlation: `cor(x,y, method="pearson")` or for an entire dataset `cor(mydata, use="complete.obs", method="pearson")` #there are other types of correlations that deal with ranking of strength but for continuous variables use Pearson

Do you want to know if there is a relationship between two ordinal variables?

Polychoric Correlation: `fit<-hetcor(mydata)$correlations` #data must be loaded first. This is different than standard correlation which assumes continuous variables, i.e. the space between 0 and 1 or 1 and 2 has meaning or can be known.

Do you want to know the effect of one variable on a single continuous variable?

Ordinary Least Squares Regression: `fit<-lm(y~x1+x2+x3, data=mydata)` # Only holds if assumptions are true, i.e. linearity in the coefficients, no autocorrelation, no heteroscedasticity. There are number of tests you can use to test these assumptions but if you are not familiar with them it will be too much to cover here. Let it suffice to say that you can use the `plot` command to examine how well your model fits the data to at least get an idea about these assumptions, e.g. `plot(fit)`