Regression

Regression is a procedure where we can examine the effects of more than one independent variable on a single dependent variable.

When we discussed correlation we talked about fitting a band around a cluster of data point. In regression we are fitting line through the data point.

\[ Y = bX + a \]

The statistical technique for finding the best fitting straight line for a set of data is called regression.

We can create a statistical model

Dependent = independent\textsubscript{1} + independent\textsubscript{2} + ..... independent\textsubscript{n}

Attitudes on the death penalty = city size + gender + race

Hence attitudes on the death penalty is a function of city size, gender and race. The independent variable interact among themselves, and all three factors impact the dependent variable. Hence, if you were take a variable out like, gender, and replace it with another variable, such as religiousness, then the estimates would change.

Attitudes on the death penalty = city size + religiousness + race

We can explain how more than one factor impacts the dependent variable. How the various independent variables interact with each other and which variable has the most impact.

Controlling for the effect of one variable on another.
Example: if we are interested in gender, but we control for region and education. This would tell us the effect of gender on income when we have taken into account region and education.

Regression will give us the same type of coefficient and significance level when we examined bivariate correlations.

For now we are only concerned with the direction of the relationship (which allows us to interpret the relationship between variables) and whether the relationship is a true (significant relevant) relationship.

We should be concerned about spurious relationships (relationships that are related but meaningless, example; whether I watch MSU basketball on TV and their probability of winning).

A theory specifies the relationship between variables

As we will see changing variables within a regression model impact the over model. Playing the
game of trying various combination of variables until you find significant relationship (not legitimate). A theory will tell us what variables we need to use. But still the game of transformation and using different procedures (which is legitimate).

Multicollinearity – Regression assumes each independent variable is independent from each other, i.e., your independent variables should not be related to each other. For example, education and income are closely related so they should not be used together in a regression analysis. Party ID and Ideology, etc.

Assumption of Linearity. Might have curvilinear relationships which you should identify. Example: age and income.

The T-statistic can tell you which of several variables has the most impact on the dependent variable. The variable with the highest t-value.

Again we will be using our unrecoded variables if possible.

Let’s look at Income at our dependent variable

Independent variables:
population size
education
gender

Profile: People with higher levels of income tend to be male and have higher levels of education. Population size is not significant. Education has the most impact on predicted income.

This is not people with higher income tend to be males with higher levels of education from larger cities. Each item is separate.

Now replace population size with recoded married variable.

Profile: People with higher levels of income tend to be male, have higher levels of education, and are married. Whether you are married or not has the most impact on predicted income.

Dependent Variable: Health
Independent variables:
Income
Age
Attend

Profile: People who are healthier tend to attend a religious service, have higher levels of income, and are younger. Income has the most impact on predicting health.

Dependent variable: How happy are you?
Independent variables:
Age
income
Health.

Profile: People who are happier tend to have higher incomes and are relatively heathy. Age is insignificant. Health has the most impact on predicting general happiness.

Replace Age with watch TV.

Dependent variable: Is life exciting or dull
education
health
pray

What is profile?