Similar to simple linear, but you have multiple independent variables that effect the dependent variable.

Assumptions of linear regression

1-Linearity

2-Homoscedasticity

3-Multivariate normality

4-Independence of errors

5-Lack of Multicollinearity

You need to check and make sure these are true before making the model.

For categorical variabvles, you need to create dummy variables. For each category, you need to create a new column.

You do not want to use all of your dummy variable columns in your regression model.

**Dummy variable trap**

You cannot have the constant and both dummy variables in the mode at the same time

Always omit one of the dummy variables

You also need to do this if you have multiple categories that need dummy variables

**Building models.**

Garbage in garbage out. = by putting all the data into the model, it tends to cloud or complicate the results.

You will then have to explain all these variables. If you are using a ton of models, this might be very difficult and long to do.

5 methods to build models.

1-all in

2-backward elimination

3-forward selection

4-Bidirectional elimination

5-Score comparison

These are often referred to as stepwise regression.

Cases to use each

1-You have prior knowledge that all these variables influence the dependent variable.

b-you have to or don’t have another choice.

c- you may try this first, before moving to backward elimination

-Backward elimination

1-You select a significant level to stay in the model.

2-Fit the full model will all possible predictor.

3-Consider the predictor with the highest p value, if the P value is greater than the significant level, got to step four. 4- remove the predictor with the highest P value

5-fit model without this variable.

You do this and remove all variables where the P value is greater than the significance level.

**Make sure to fit the model after every removal.**

**Forward selection**

1-select significance level

2-Fit all simple regression models y~xn select the one with the lowest P value

3-Keep the previous variable and fit all possible models with one extra predictor added to the ones you already have

4-Consider the predictor with the lowest P -value. If p< significance level, go to step 3.. other wise end. Keep the previous model, because the variable you just added does not meet the significance test and is therefore insignificant

**Bidirectional elimination**

1-Select a significance level to enter and to stay in the model SLenter = 05, SLstay= .05

2- Perform the next step of forward selection(new variables must have: P< SLenter to enter)

3-Perform all steps of backward elimination(old variables must have p<slstay to stay)

4-No new variables can enter, and no old variables can exit

Model is ready

**All possible models**

1Select a criterion of goodness of fit(eg Akaike criterion)

2Construct all possible regression models 2n – 1 total combinations

3-Select the one with the best criterion

The model is ready