Microsoft Excel for Economic Data Analysis
Video Lecture #2 Notes

Constructing Frequency Distributions Using Excel

A frequency distribution is effectively a count of the number of observations from a particular sample that fall into different ranges. For a given sample of values of a numeric variable, we can identify to following values:

Minimum: Smallest value in the sample
Maximum: Largest value in the sample
Range: Maximum minus the Minimum
Sample Size (usually denoted n): the number of individual observations in the sample.

Ordering the data from the minimum to the maximum, first decide how many equally spaced "classes" to break the range of data into. Rough guide:

Sample Size	Number of Classes
Fewer than 50	5 – 7
50 to 100	7 – 8
101 to 500	8 – 10
501 to 1,000	10 – 11
1,000 to 5,000	11 – 14
More than 5,000	14 – 20

Next, determine the "class width."

$$w = Class\ Width = \frac{Maximum - Minimum}{Number\ of\ Classes}$$

Rounded up to insure all the data falls in a class.

Class	Value
1	Minimum + Class Width
2	Class 1 + Class Width
3	Class 2 + Class Width
…	…

Then count the number of observations that fall into each class.

Frequency distributions for samples with "extreme" values.
Quartile and Percentile order statistics.
First Quartile – Q1: 25% of the observations fall below this value
Second Quartile – Q2 or Median: 50% of the observations fall below this value
Third Quartile – Q3: 75% or the observations fall below this value

Excel function =QUARTILE.EXC(range,k)
Where k = {1,2,3}

Percentiles
P5: 5% of the observations fall below this value
P10: 10% of the observations fall below this value
Etc.

Excel function: =PERCENTILE.EXC(range,k)
Where k is any number between zero and one.

When extreme values are present in a sample, try "trimming" the sample by putting the bottom 5% – 10% of the sample in an "extreme low" class and the top 90% – 95% in an "extreme high" class to get a clearer picture of how the bulk of the sample is distributed.

Creating a category variable using IF statements

=IF(

IF(logical_test, [value_if_true], [value_if_false])
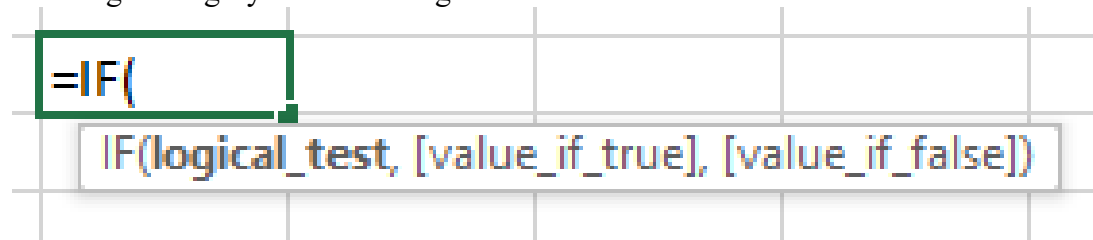
IF statements can be nested and include other embedded Excel functions.