REGRESSION ANALYSIS

A project sponsored by the U.S. Economic Development Administration
& Western Carolina University's Public Policy Institute

Regression Analysis: What is It?

- Simply put, it is a way to determine whether and how much variables are related to one another
- Imagine you are plotting the relationship between two vairiables with the independent variable on the x axis and the dependent variable on the y axis. Now imagine that you were going to draw the straight line that lies the closest to all of those data points. Regression analysis draws that line for you.
- Linear (straight line) bivariate regression uses one independent variable, multivariate compares allows you to determine the relative influence of multiple independent variables simultaneously.

Regression: What can it do?

- Determine if one or more variables are related
- Analyze strength of relationships
- Used for determining profits, accumulation over time, effects of an incidence on another, and on and on
- Must be sure data are measured at the same time and on the same units.

Basic Regression Analysis: Key Terms

Test of statistical significance

- Demonstrates the probability that a relationship found in a set of sample data occurred by chance.

Measure of association

- Shows how strong the relationship is between the independent and dependent variables.

Y intercept

- This is where the baseline of the regression starts. The value of Y when X=0.

Basic Regression Analysis: Key Terms

Regression coefficient

- The slope, it displays how steep or flat the regression line is. If it steep it shows a relationship and flat, it does not.

R^2

- This tells how much variation in the dependent variable is explained by knowing all of the independent variables. R^2 vary from 0 to 1 where a 1 indicates that you can explain all of the variation in the dependent variable by knowing the independent variables. A .2 would indicate that you can explain 20% of the variation, and so on.

Regression Analysis: Regression Equation

Y=a+bX

- a=the constant or Y intercept
- b=the regression coefficient or slope
- Y=predicted value of Y, the dependent variable
- X=the independent variable

Slope

- Starting at the Y intercept (farthest point to the left), for each one unit increase in the X variable, you have a ___ unit increase in the Y variable.

Regression Analysis: Interpreting Coefficients

- The statistical significance of the regression coefficient can be tested by coming up with the t-statistic
- t=b/se
- Excel will give you a t statistic and an associated significance level

Regression Analysis: Interpreting R^2

- R^2 = a number describing how much variance in the dependent variable is explained by all of the independent variables taken together.
- Varies from 0-1 but there is no one "good R^2".
- Tells us how much better can we predict the dependent variable by knowing the independent variable than by not knowing the independent variable
- If you had a model with an R^2 of .55, that means that 55% of the variation in the dependent variable is explained by the independent variable(s)

Regression Analysis: An Example

- You want to know how unemployment affects crime so you gather data by state rates in both
- H_0: Unemployment has no effect on crime by state.
- H_1: As the rate of unemployment goes up by state, so does the rate of indexed crime.

Regression: Crime Example

State Unemployment Rate*

- AK-6.2
- AL-4.1
- AR-4.2
- AZ-4.0
- CA-4.9
- CO-2.7
- CT-2.3

…

X variable

*2000 National Bureau of Labor Statistics

State Index Crime Rate**

- AK-4.2
- AL-4.5
- AR-4.1
- AZ-5.8
- CA-3.7
- CO-4.0
- CT-3.2

…

Y variable

**2000 FBI Crime Data per 100

Regression: Step by Step In Excel

1. Put your data in excel
2. Click on Tools→ Data Analysis→ Regression→ OK

**If you do not have the add-on pack installed it is easy to get…click microsoft office button*

3. Enter Input Y Range: (Click on Range of the 50 "Crime per 100K" cell data)
4. Enter Input X Range: (Click on Range of 50 "Unemployment Rate" cell data)

**Your data will be displayed in a separate tab*

Regression: The output

SUMMARY OUTPUT

Regression Statistics

Multiple R	0.3306123
R Square	0.109304493
Adjusted R Square	0.090748337
Standard Error	0.900595386
Observations	50

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*

	Coefficients	*Standard Err*	*t Stat*	*P-value*
Intercept	2.661074447	.5609107	4.744204	1.92E-05
X Variable	0.344897907	.142107	2.427029	0.019026

1. R Square=.109, so 11% of the variation in crime can be predicted by the unemployment rate
2. Y Intercept (or Crime rate) and X Variable (Unemployment Rate-For every 2.7 crimes (per 100 people) the unemployment rate will be .34% higher. In other words, after calculation, for every 111 crimes, the unemployment rate jumps 1%.
3. P Value is .019, or 1.9% this happened by chance.

Regression: Another Example

- Lets say we use the welfare expenditures and see if that plays a role in correlation with crime rate*…

(Welfare in Millions)

AK-$3,485

AL-$803

AR-$2,593

AZ-$1,987

CA-$18,510

CO-$2,211

CT-$3,020

…

*US Census Bureau 2000

Regression: Welfare Example

SUMMARY OUTPUT

Regression Statistics

Multiple R	0.196991041
R Square	0.03880547
Adjusted R Square	0.018780584
Standard Error	0.935558033
Observations	50

ANOVA	*df*	*SS*	*MS*	*F*	*Significance F*

	Coefficients	*Standard Err*	*t Stat*	*P-value*
Intercept	3.81213	.182375	20.90276	9.43E-26
X Variable	04.3954	3.16E-05	1.392071	0.170317

1. R Square=.038, so only 4% of the variation in crime can be predicted by the welfare expenditures

2. Y Intercept (or Crime rate) and X Variable (Welfare Expense)-For every Million dollars spent on welfare, the crime rate increases less than .04%.

3. P Value is .17, or 17% chance this happened by chance. **So it is not statisically significant.**

Multiple Regression

- What happens if we decide that we want to use two variables, controlling for one another?
- Let's use both unemployment rates and welfare expenditures and their effect on crime…
- Use the same steps
  1. Put your data in excel
  2. Click on Tools→ Data Analysis→ Regression→ OK
  3. Enter Input Y Range: (Click on Range of the 50 "Crime per 100K" cell data)
  4. Enter Input X Range: (Click on Range of 50 "Unemployment Rate cell data and welfare expenditures data)

Multiple Regression Results

SUMMARY OUTPUT

Regression Statistics

Multiple R	0.341997
R Square	**0.116962**
Adjusted R Square	0.079386
Standard Error	0.906205
Observations	50

ANOVA *df SS MS F Significance F*

	Coefficients	*Standard Err*	*t Stat*	*P-value*
Intercept	2.708408	0.56925	4.7578	1.9E-05
Unemploy	**0.311048**	0.15251	2.0396	**0.04704**
Welfare	2.08E-05	3.26E-05	0.63840	0.52630

1. R Square=.117 so nearly 12% of the reason for crime can be predicted by the unemployment and welfare expenditures

2.

3. P Value is .047, or less than 5% chance this happened by chance. **So it is statistically significant!**

Regression Summary

- Regression is a tool to determine whether one or more independent variable(s) influences the variation in a dependent variable.
- Can be used for many purposes in assessing the impacts of economic development activities.