

IS/AAE374 Discussion Section  
Analyzing Complex Relationships using Regression

### Regression Analysis

- We can gain a great deal of insight from regression analysis and it is extensively used not only in statistics and economics but also politics, sociology, and baseball. Today we'll cover the basics of regression analysis so that you can calculate regressions and read tables with regression results.
- A regression line summarizes the relationship between the two variables (Y and X), but only when we are looking for one variable to explain or predict the other variable.

$$Y = \beta_0 + \beta_1 X + e$$

- Y is called a dependent variable, and X is an independent or explanatory variable. We are trying to estimate Y given our values of X.
- With multiple explanatory variables,
$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + e$$
- By minimizing the distance between the actual observations of X and Y, we can estimate the parameters and predict Y for a given X.

### Interpretation

- Goodness of fit
  - R-squared term (coefficient of determination): ranges from 0 to 1 and tells the extent to which the dependent (Y) variable can be explained by the independent (X) variable.
    - Remember r? Great!
  - These values are often used to determine the strength of a model as a whole
- T-tests
  - T-tests are used to determine whether individual parameters are significant predictors of the dependent variable.
    - Remember the significance levels? All of them?
  - Economic significance vs. statistical significance

### Exercise 1 - Growth Rate ("GR" in i\_love\_economics.xls)

$$GrowthRate = \beta_0 + \beta_1 \times InitialGDP + e$$

- Estimate the  $\beta$ 's, and judge if they are statistically significant.<sup>1</sup>

---

<sup>1</sup> To run a regression in Excel you must install the Analysis Toolpak in the Excel add-in options (we already installed it when we used the histograms). Then, go to Data Analysis, and select regression. Key results include the coefficients and the p-values (recall that p-values reveal the statistical significance of each coefficient).

**Exercise 2 – Technological Parameters** (“TFPG” in i\_love\_economics.xls)

$$y = Ak^{\beta_1} s^{\beta_2} \quad \text{or} \quad \log y = \log A + \beta_1 \log k + \beta_2 \log s$$

- Estimate the  $\beta$ 's in both 1965 and 1990.

**Exercise 3 - Growth Rate with multiple explanatory variables** (“GR” in i\_love\_economics.xls)

$$\text{GrowthRate} = \beta_0 + \beta_1 \times \text{InitialGDP} + \beta_2 \times \text{Education} + \beta_3 \times \text{Health} + e$$

- Estimate the  $\beta$ 's. Are they statistically significant?

**Examples** (Wright and Birdsall)

TABLE 4—REGRESSIONS FOR MANUFACTURED NET EXPORTS OF THE UNITED STATES, 1879–1940

	Constant	Capital/ Labor	Natural Resource Coefficient	Average Wage	Percent Women and Children	R <sup>2</sup>
1879	−3127 (0.68)	2092** (2.24)	−10830 (0.74)	−1853 (0.27)		0.079
	−228 (0.06)	1725* (1.77)	−12690 (0.83)		−156 (1.53)	0.103
1899	−4068 (0.66)	3729* (1.73)	−4324 (0.11)	−802 (0.07)		0.075
	1735 (0.28)	3140 (1.46)	−8727 (0.21)		−255** (2.02)	0.093
1909	−8965 (0.92)	2648 (1.17)	46950 (1.17)	959 (0.06)		0.146
	260 (0.04)	1810 (0.75)	44154 (0.99)		−380** (2.25)	0.193
1914	−21041** (2.56)	1600 (0.53)	103103* (1.71)	28468** (2.12)		0.261
	216 (0.02)	1038 (0.33)	98271* (1.55)		−329* (1.93)	0.275
1928	−21067 (1.20)	5040 (0.83)	112264** (2.19)	18856 (0.52)		0.143
	−4342 (0.17)	4413 (0.67)	107406** (2.01)		−333 (0.87)	0.149
1940	−31898 (1.13)	−1862 (0.42)	126449** (2.22)	85642 (1.38)		0.085
	23714 (1.24)	−2750 (0.58)	117138** (2.11)		−629* (1.79)	0.077

Notes: Method of estimation is ordinary least-squares, *t*-ratios (in parentheses) adjusted for heteroscedasticity following procedure of White (1980). \*Denotes statistical significance at the 5 percent confidence level; \*\*denotes the 1 percent confidence level. There are 64 nonzero observations in 1879, 83 in 1899, and 96 in the remaining years.

Table 2. *Determinants of GDP Growth in a Cross-Economy Sample, 1960–85*

Variable	Basic regression	Basic regression with manufactured exports	Basic regression with income share	Basic regression with income share, excluding education variables
Per capita GDP, 1960 (thousands of 1980 dollars)	−0.0075 (−6.25)	−0.0069 (−5.135)	−0.0075 (−4.730)	−0.0020 (−1.801)
Primary school enrollment rate, 1960	0.0250 (4.464)	0.0271 (4.532)	0.0243 (3.024)	
Secondary school enrollment rate, 1960	0.0305 (3.861)	0.0262 (1.723)	0.0366 (2.427)	
Ratio of manufactured exports to GDP, 1965		0.0007 (1.539)		
Secondary school enrollment– export interaction, 1960 <sup>a</sup>		0.0005 (0.324)		
Income share ratio <sup>b</sup>			−0.0013 (−1.897)	−0.0018 (−2.406)
Government consumption share of GDP, 1970–85 <sup>c</sup>	−0.1190 (−4.250)	−0.0566 (−2.419)	−0.1229 (−4.380)	−0.1495 (−4.760)
Average annual number of revolutions, 1960–85	−0.0195 (−3.095)	−0.0168 (−1.987)	−0.0176 (−1.867)	−0.0268 (−2.493)
Average annual number of assassinations, 1960–85	−0.0333 (−2.148)	−0.0024 (−0.738)	−0.0055 (−1.526)	−0.0067 (−1.626)
Absolute deviation in investment deflator, 1960 <sup>d</sup>	−0.0143 (−2.698)	−0.0139 (−2.136)	−0.0086 (−1.093)	−0.0176 (−2.012)
Constant	0.0302	0.0202	0.0418	0.0706
R <sup>2</sup>	0.56	0.57	0.54	0.36
Number of observations	98	100	74	74