

REGRESSION 101

August 31, 2020

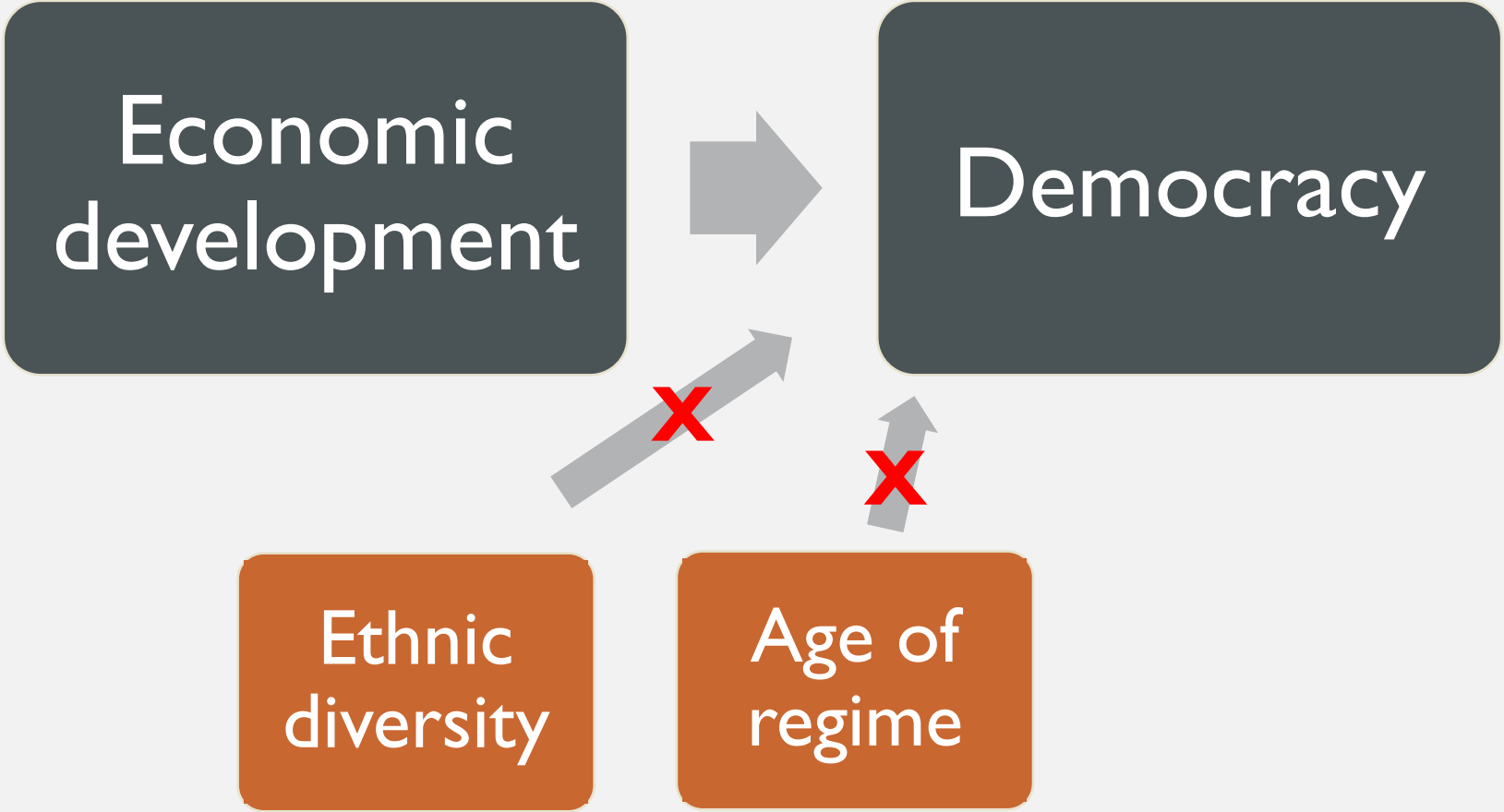
Carrie Coberly

GOALS

- Learn the purpose of using regression in political science
- Understand the concept of statistical significance

WHY REGRESSION?

- Regression analysis uses math to hold other variables constant while we compare two variables to each other.
- Allows us to eliminate some alternate explanations and get closer to identifying a relationship between cause and effect.



Democracy

Econ. Devp.

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

Diversity

Age

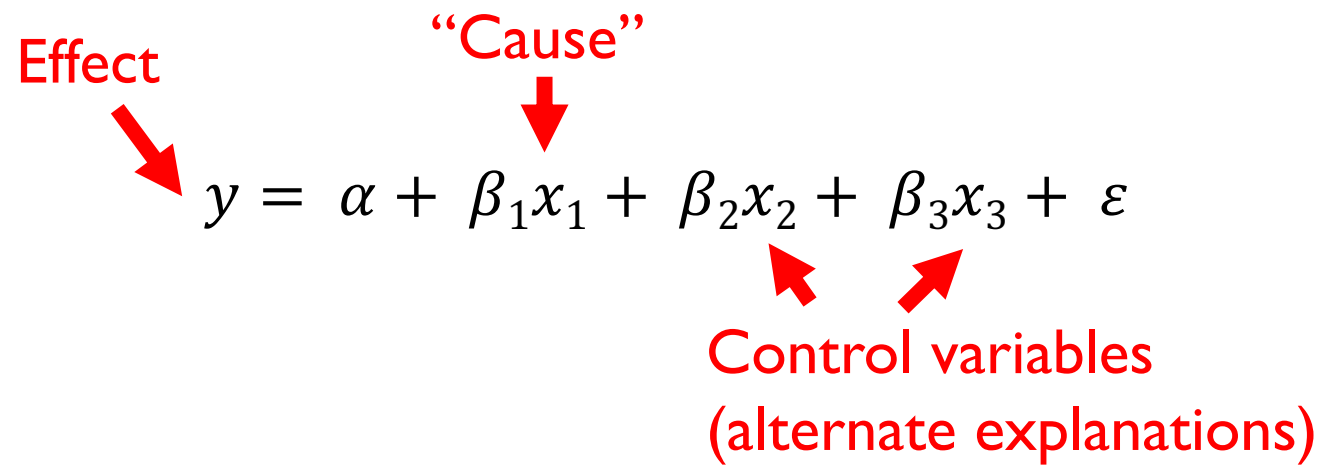
The diagram illustrates a linear regression model. The dependent variable is y . The independent variables are x_1 , x_2 , and x_3 . The coefficients are α , β_1 , β_2 , and β_3 . The error term is ε . Red arrows indicate the following mappings: Democracy points to β_1 , Econ. Devp. points to β_2 , Diversity points to β_2 , and Age points to β_3 .

Effect

“Cause”

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

Control variables
(alternate explanations)



Size/direction of
effect (coefficient)

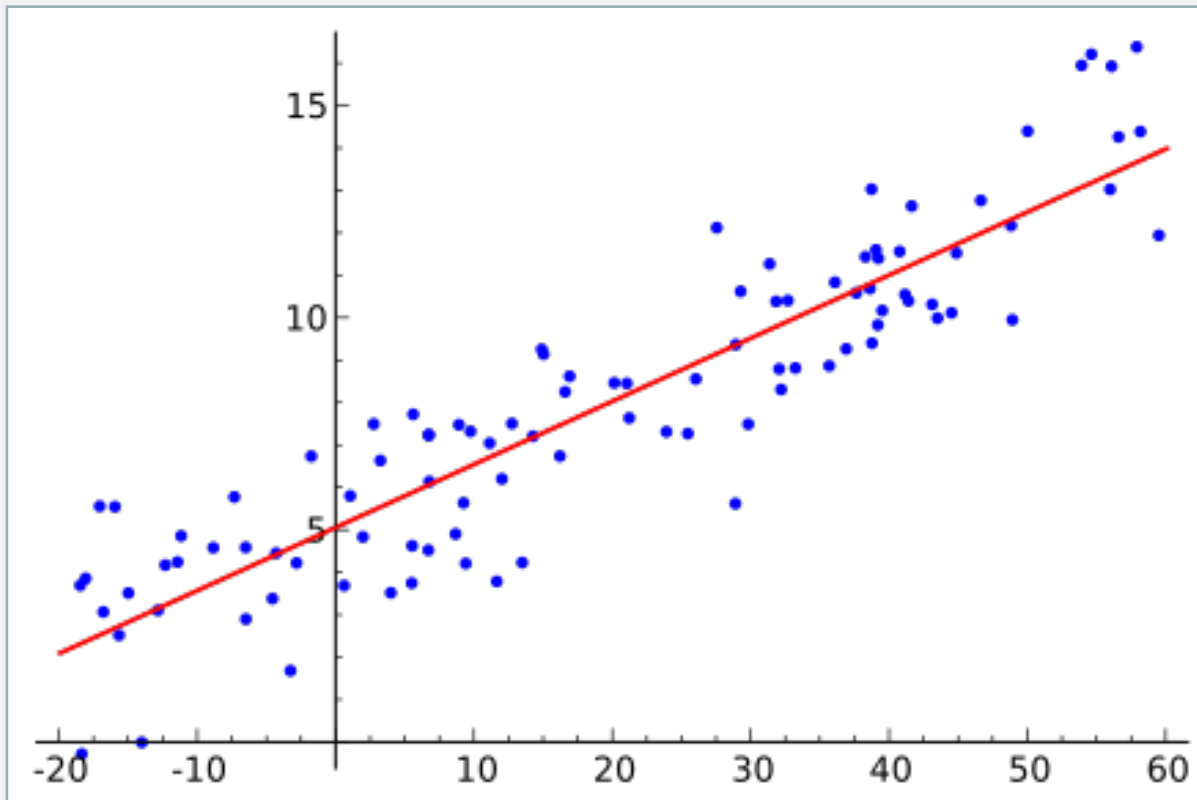


Error



$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

LINEAR REGRESSION



Draws the best fit line for your data:

$$y = \alpha + \beta x + \varepsilon$$

y = dependent variable

α = intercept/constant

β = coefficient/slope of the line

x = independent variable

ε = error

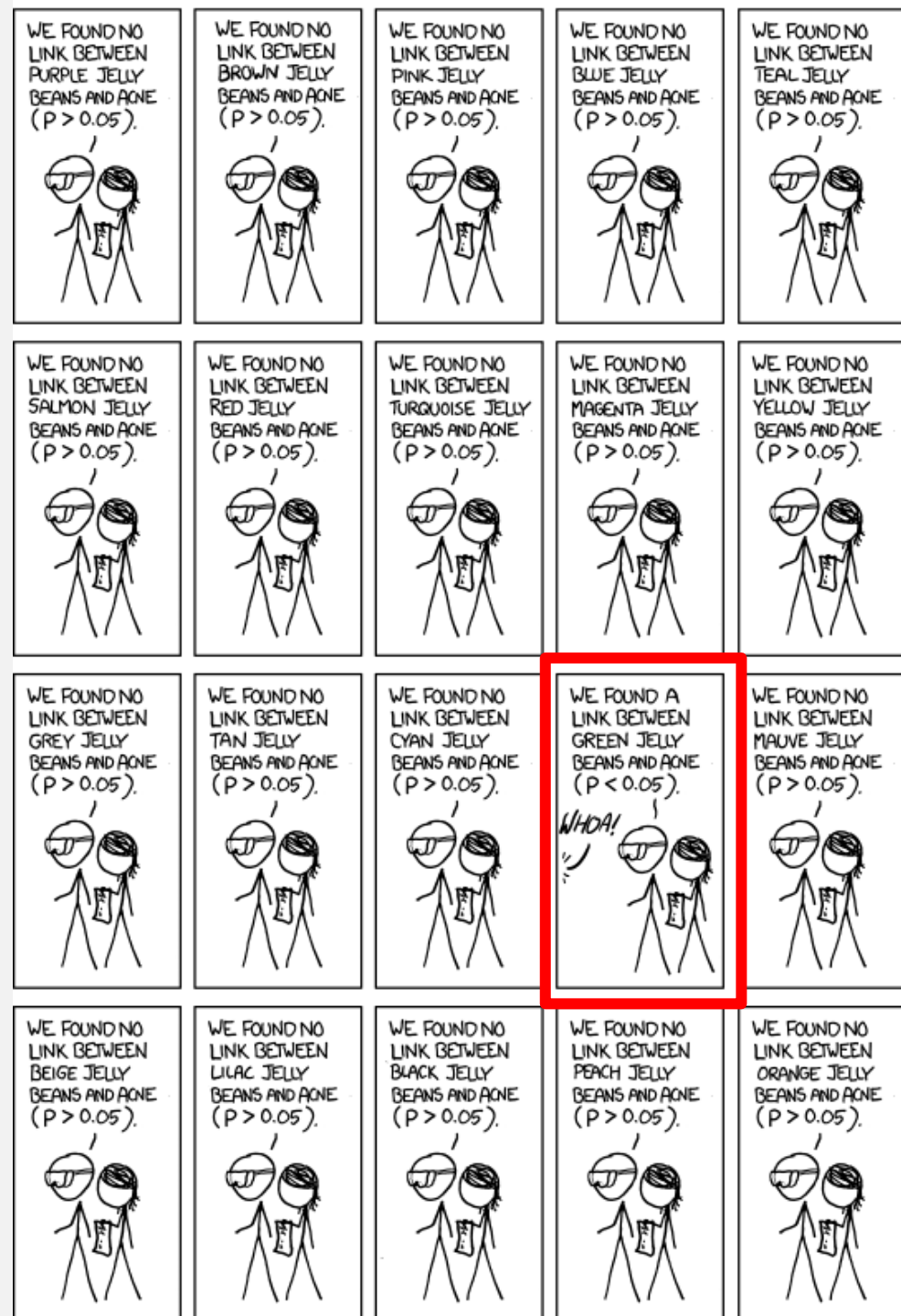
INTERPRETING COEFFICIENTS

- What is the *sign* of the coefficient (+ or -)
- Positive coefficient show positive relationships (as interstate war increases, state development increases)
- Negative coefficients show negative relationships (as civil war increases, state development decreases)

STATISTICAL SIGNIFICANCE

- *Can we tell the difference between our result and zero*
 - Point estimates (coefficients) are statistically significant if you can distinguish them from zero.
 - Differences are statistically significant if you can distinguish the two point estimates from each other.
- *We talk about certainty at a 95% confidence level*

- 95% confidence means there is a 1 in 20 chance (5%) the research is wrong – that because of error in the data or your model, you accidentally got a result when there was none.



REPORTING ERROR

- **Standard error**
 - If the (absolute value of the) coefficient is more than (approximately) two times the standard error, it is statistically significant
- **Confidence interval**
 - If the confidence interval does NOT include 0, it is statistically significant
- **p-value**
 - $p \leq 0.05$ is statistically significant

FOR MORE
HELP...

NEW YORK TIMES BESTSELLER

naked statistics

STRIPPING THE DREAD FROM THE DATA



"Brilliant, funny
... the best math teacher
you never had."
—*San Francisco Chronicle*

charles wheelan

AUTHOR OF NAKED ECONOMICS

REGRESSION TABLE

TABLE 3
REGRESSIONS OF FREEDOM HOUSE SCORES ON
HYPOTHESIZED DETERMINANTS^a

<i>Variable</i>	<i>Model 1</i>	<i>Model 2</i>	<i>Model 3</i>	<i>Model 4</i>	<i>Model 5</i>
Constant	0.17 (0.84)	-0.15 (0.70)	0.27 (0.81)	-0.15 (0.60)	0.19 (0.62)
Islamic religious tradition	-1.24*** (0.27)	-1.27*** (0.27)	-1.26*** (0.27)	-1.34*** (0.27)	-1.68*** (0.27)
Economic development	1.40*** (0.21)	1.48*** (0.19)	1.40*** (0.20)	1.50*** (0.17)	1.39*** (0.17)
Sociocultural division	-0.32 (0.43)		-0.30 (0.42)		
Economic performance	0.07 (0.05)	0.06 (0.04)	0.06 (0.04)		
British colonial heritage	0.25 (0.30)	0.18 (0.30)			
Communist heritage	0.20 (0.27)				
OPEC membership	-1.36** (0.46)	-1.46** (0.45)	-1.42** (0.46)	-1.53** (0.48)	
Adj. R ²	.55	.55	.55	.55	.51
<i>N</i>	149	149	149	149	149

*p<0.05; **p<0.01; ***p<0.001

^aEntries in this table and all others are unstandardized regression coefficients with White-corrected robust standard errors in parentheses.

INTERPRETING COEFFICIENTS

- A one **unit** increase in **X** is associated with a **coefficient unit** increase/decrease in **Y**, holding **all other variables** constant. These results are statistically significant at the 95% confidence level.
- *For example:* A \$10,000 increase in GDP per capita is associated with a 0.5 point increase in democracy score, holding ethnic fractionalization constant. These results are statistically significant at the 95% confidence level.