Mathematics 231

Lecture 11

Liam O'Brien

Announcements

- Reading
  - Today

Today	M&M 2.3	119-121
	M&M 2.4	125-132
	Supplemental	Regression to the Mean
Next class	M&M 2.5	148-151
	M&M 2.6	154-159

Stuff To Do

- Regression to the Mean
- Transformations

Regression to the Mean

- Regression to the mean refers to the following "regression effect":

- **Extremes on one variable, say X, are less likely to be extremes on the other variable, say Y.**

Example: Regression to the Mean

- Consider a set of software companies' stock prices in 1999 and 2000.
- In 1999, the mean stock price for this set of companies was $74/share, with SD = 12.5.
- In 2000, the mean stock price for this set of companies was $74/share, with SD = 12.5.
- It seems that nothing affected the market differentially between the two years.

Example: Regression to the Mean

- On closer examination, the following surprising result emerged:
  - Companies with stock prices below average in 1999 tended to gain \$5 to \$10 per share in 2000.
  - Companies with stock prices above average in 1999 tended to lose \$5 to \$10 per share in 2000.
- Companies that were below average in 1999 showed an improvement in 2000, and vice versa.

Regression to the Mean: Why?

- Consider a company with an average stock price of $90/share in 1999. What is the prediction of the average stock price in 2000?

  Price in 2000 = a + b (price in 1999)

$$b = r\frac{s_Y}{s_X}$$

$$a = \bar{y} - b\bar{x}$$

Regression to the Mean: Why?

$$b = r\frac{s_Y}{s_X}; \qquad a = \bar{y} - b\bar{x}$$

- Suppose the correlation between 1999 prices and 2000 prices is r = 0.6.

$$b = 0.6\frac{12.5}{12.5} = 0.6$$

$$a = 74 - 0.6(74) = 29.6$$

$$\text{price in } 2000 = 29.6 + (0.6)90 = 83.6$$

Regression to the Mean: Intuition?

- Company with a 1999 stock price of $90/share. What is the predicted 2000 price?
- Consider 3 scenarios:
  1. “True” worth is $80/share but by chance the price is higher.
  2. “True” worth is $100/share but by chance the price is lower.
  3. “True” worth is $90/share; chance played no role.
- Which scenario is most/less likely?

Regression to the Mean: Other Examples

- Students who score high on the midterm tend to score high, but not as high on the final.
- A baseball player who has a spectacular rookie year tends to not perform as well his/her second year (sophomore slump).
- Tall parents tend to have children who are tall, but not as tall.

Origins of “Regression?”

- This dates back to Galton who was trying to figure out whether a child’s height could be predicted by his/her parents’ heights.
- He found that it could, but that really tall parents, tended to have children shorter than they were, and vice versa.
- He called this “reversion to mediocrity” and later changed this to “regression to mediocrity.”

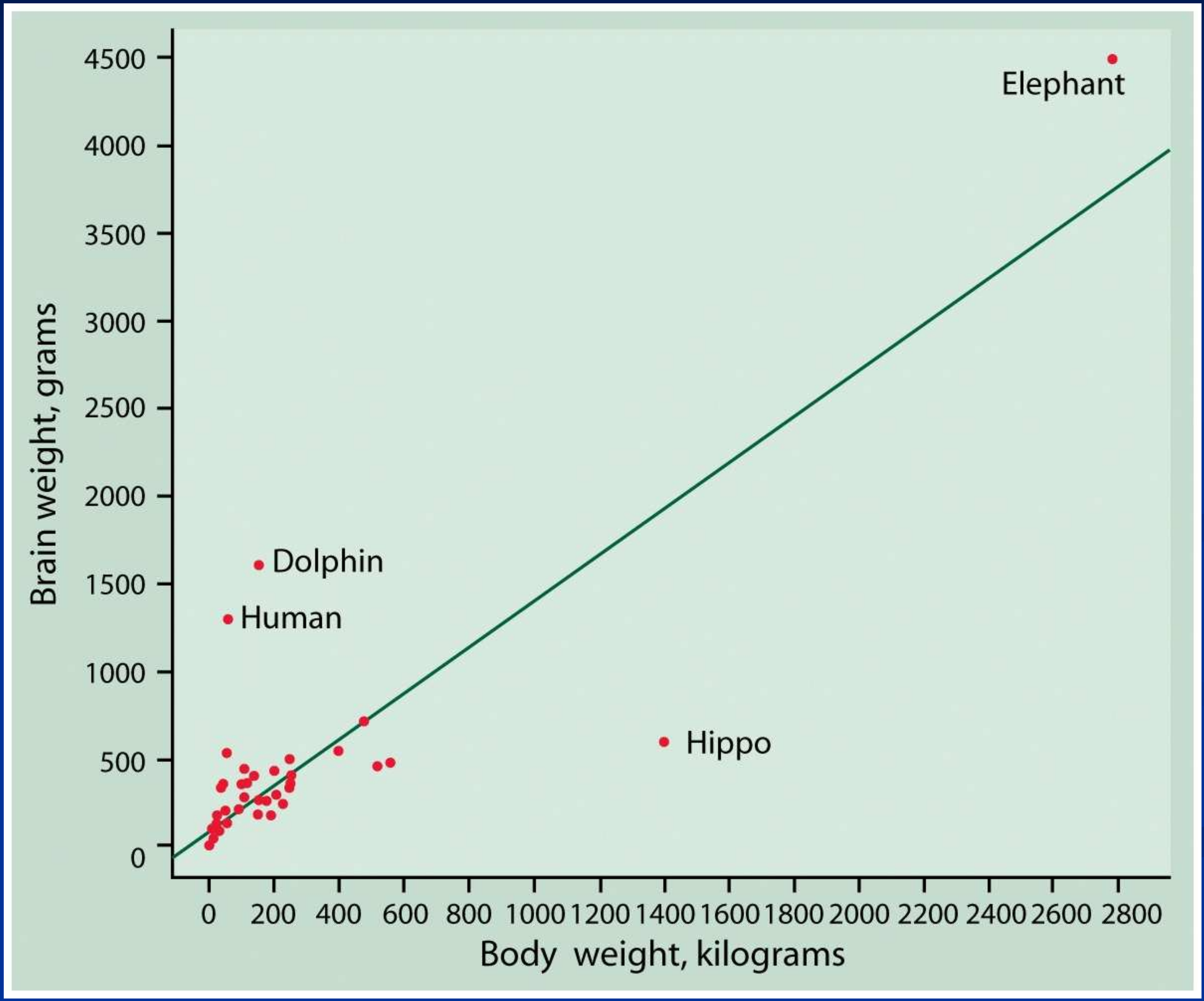
Transformations

- Transformations are useful if the regression assumptions are not met.
- Recall the assumptions:
  1. Conditional mean of Y is a linear function of X.
  2. Conditional SD of Y is constant for all X.
- We often make an additional assumption:
  3. The conditional distribution of Y is a normal distribution for any value of x.

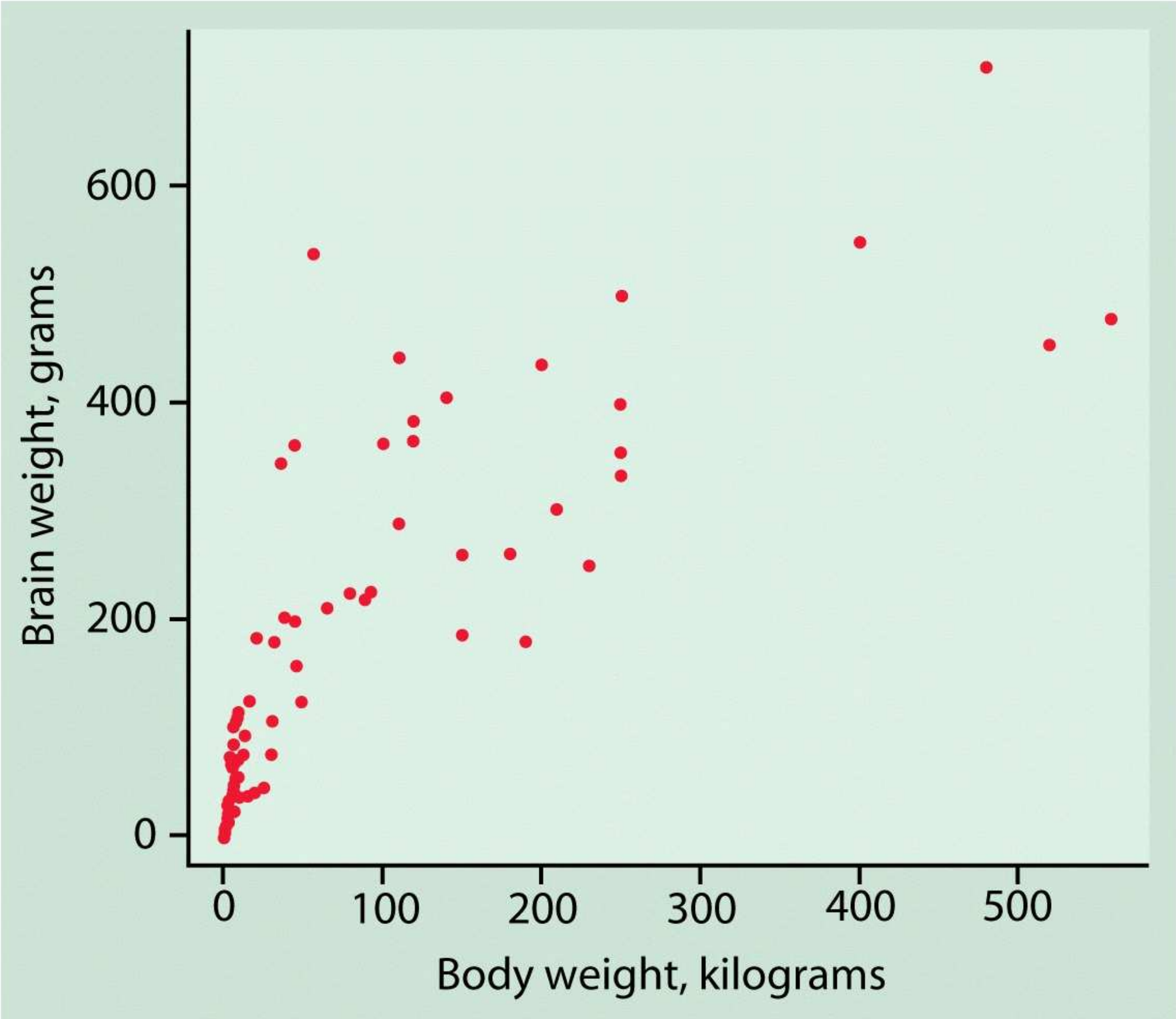
Checking Regression Assumptions

- Examine the residual plot.
- If the assumptions are not met, what can we do?
  - Pretend they are (the "ostrich" approach).
  - Consider more complex "nonlinear" models.
  - Transform data to conform to assumptions.
- What are the implications of using the ostrich approach?
- We'll consider this last option.

Example: Brain versus Body Weight



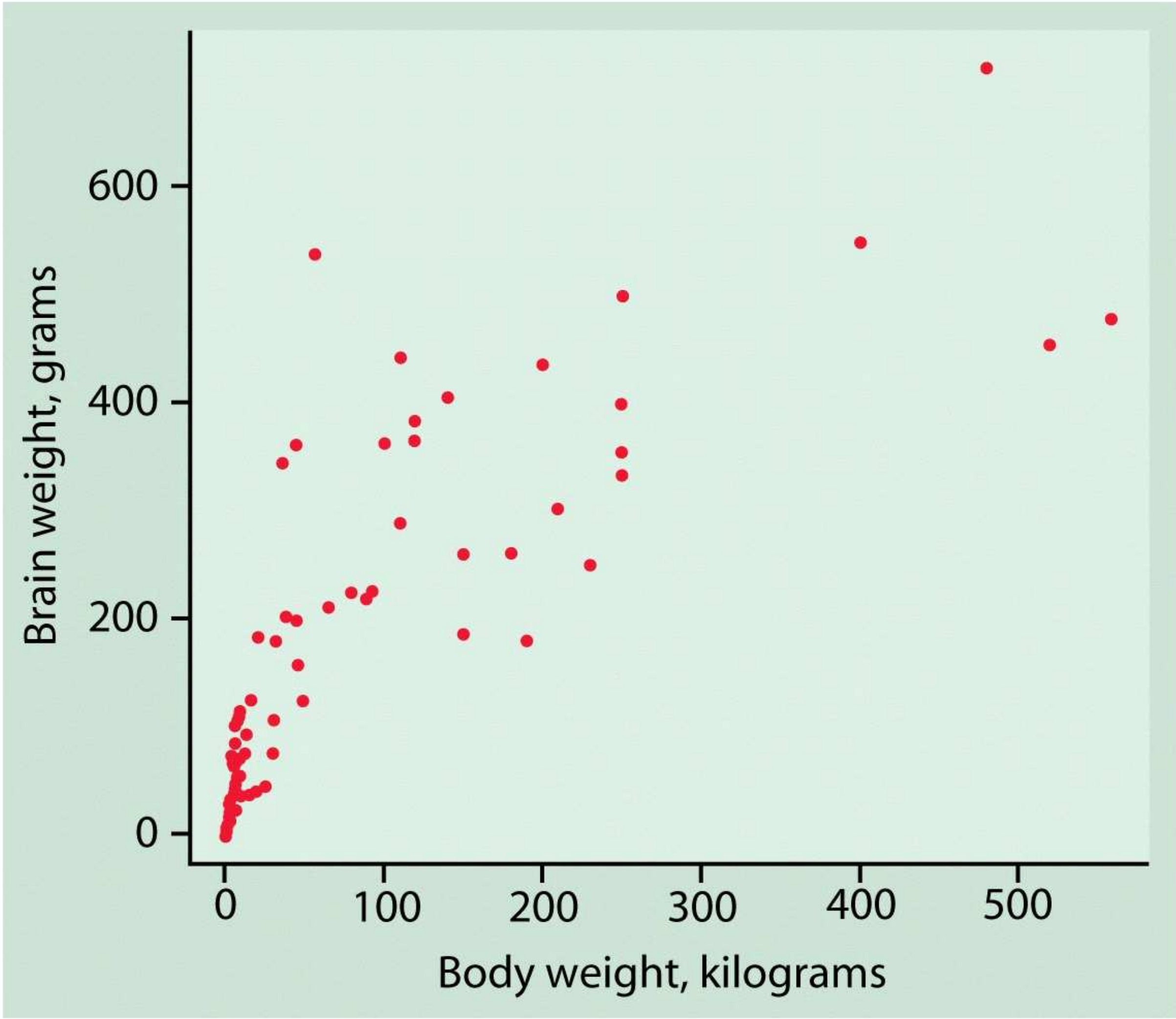
Example: Brain versus Body Weight



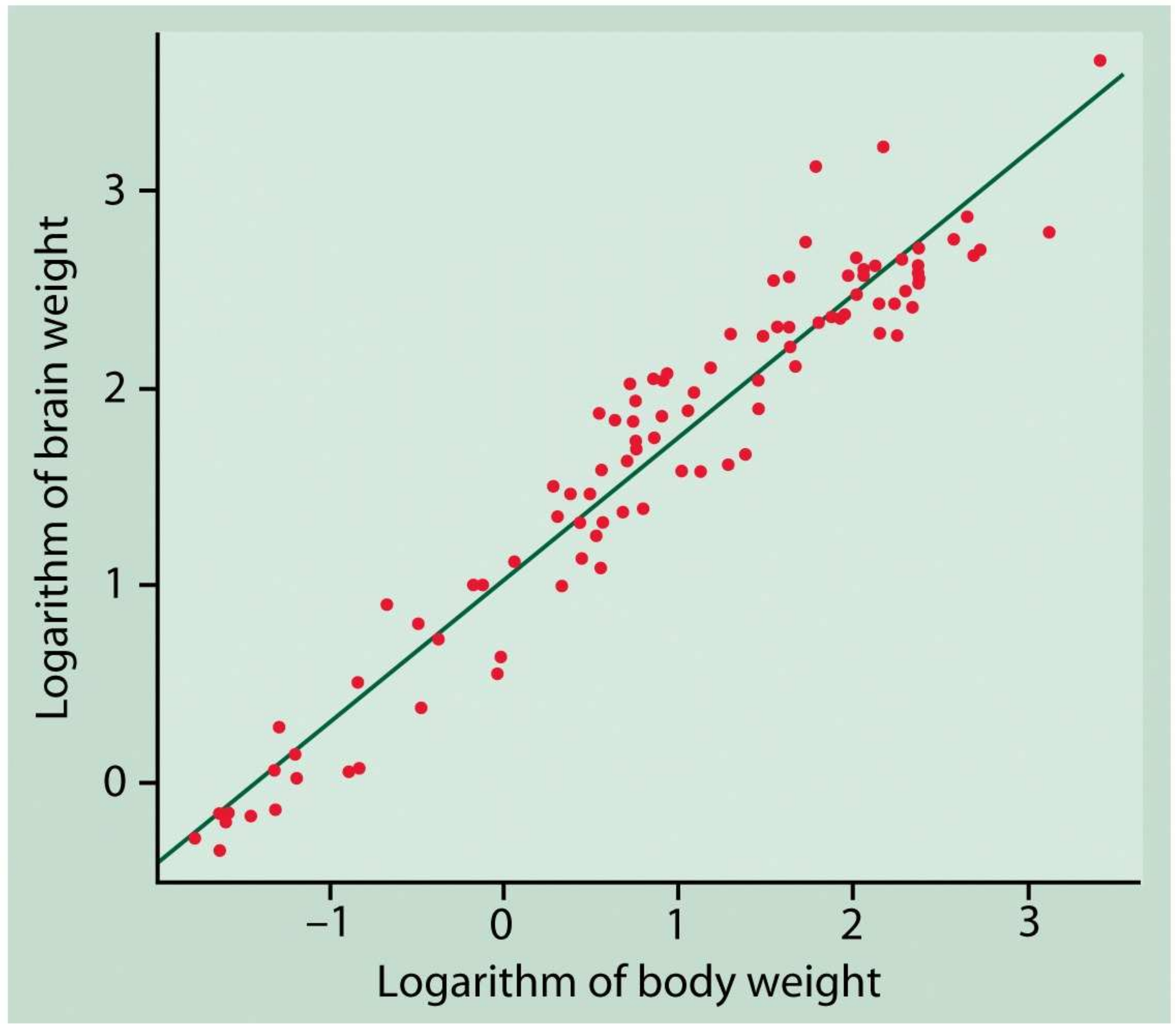
Nonlinear Transformations

- **Recall:** Earlier we discussed linear transformations; here we need nonlinear transformations.
- Nonlinear transformations can:
  - Alter the shape of distributions, making skewed distributions more symmetric.
  - Alter the conditional SD.
  - Change the form of the relationship between two variables.

Example: Brain versus Body Weight



Log Brain Versus the Log Body Weight

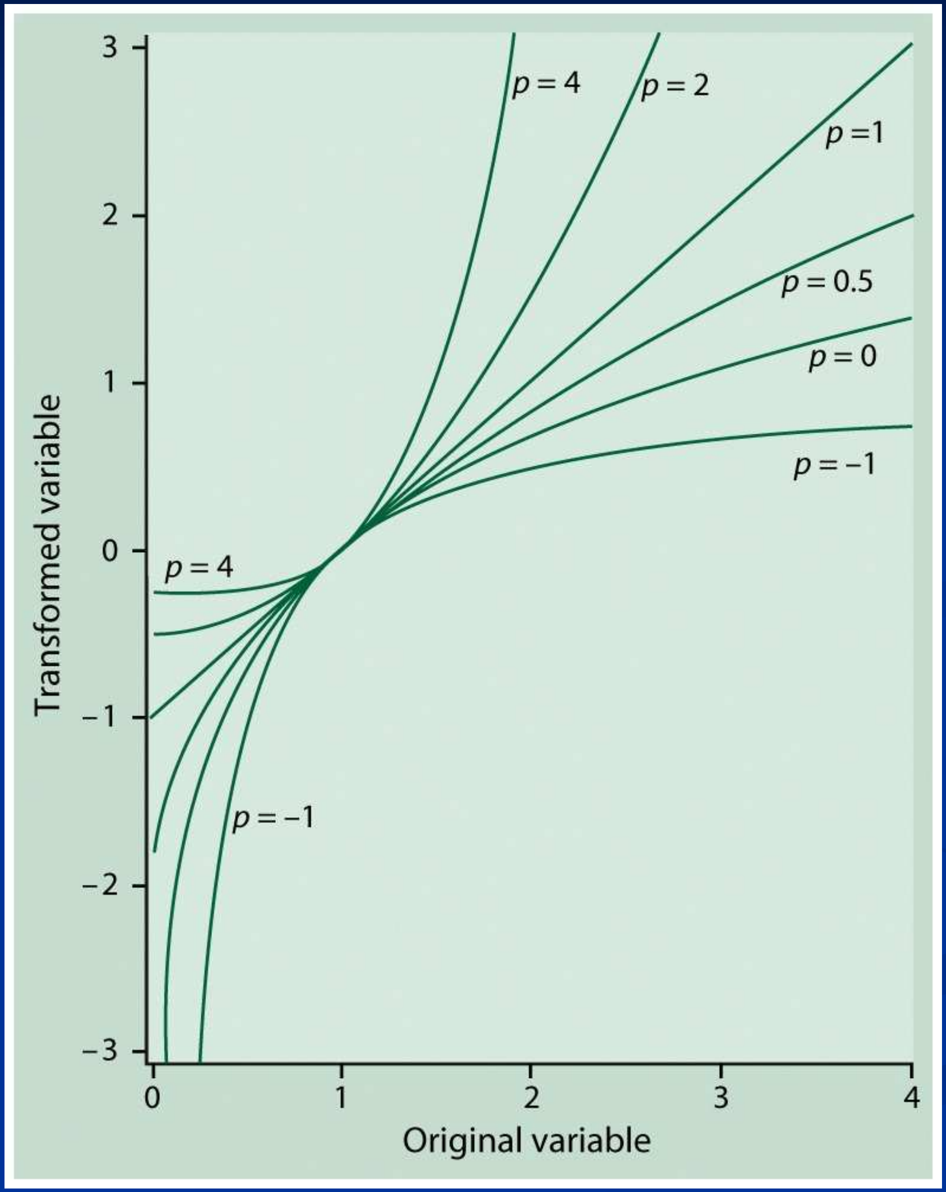


Common Nonlinear Transformations

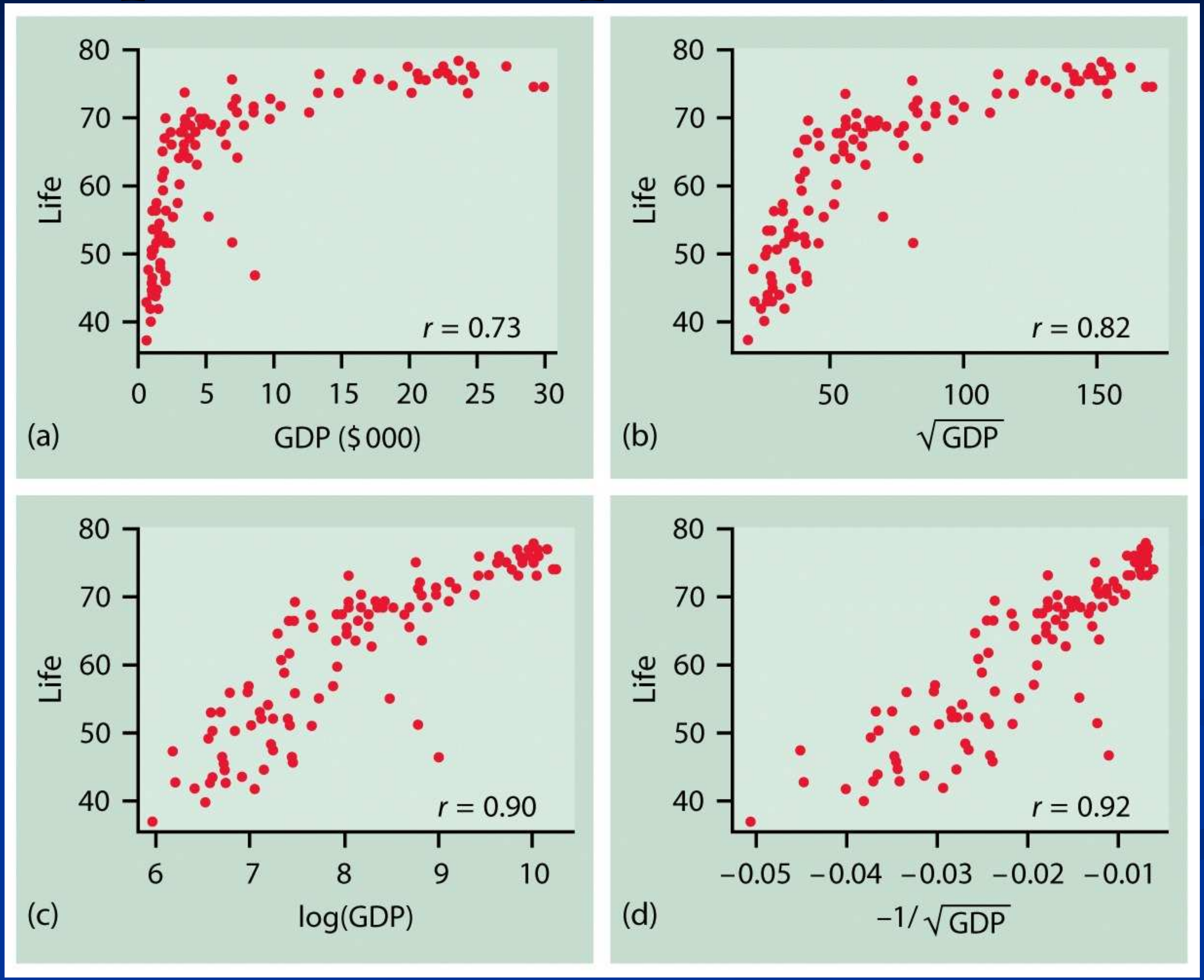
- When relationship between Y and X is not linear, consider transformations of the form Y^P and X^P, where

  p = …-3, -2, -1, -1/2, log, 1/2, 1, 2, 3…
- "Ladder of Powers" in M&M (weird and confusing)
- "Circle of Powers" (what more people use)

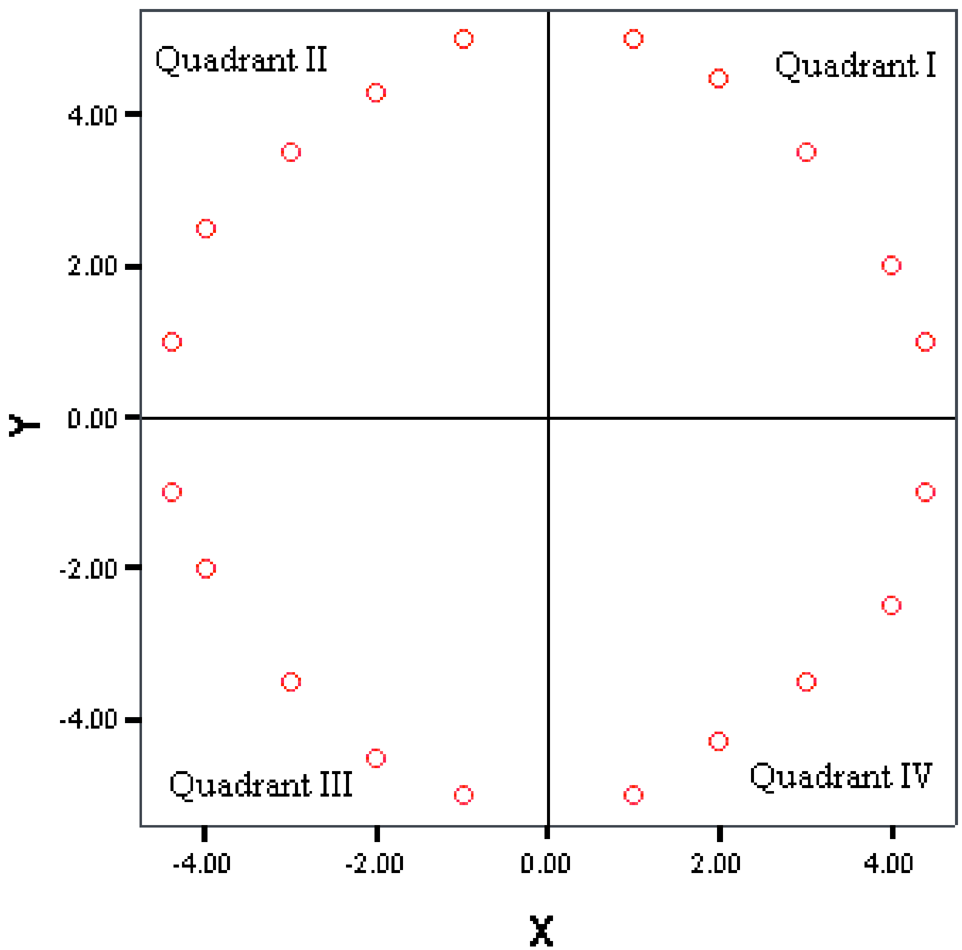
Ladder of Powers



Example: Life Expectancy and GDP



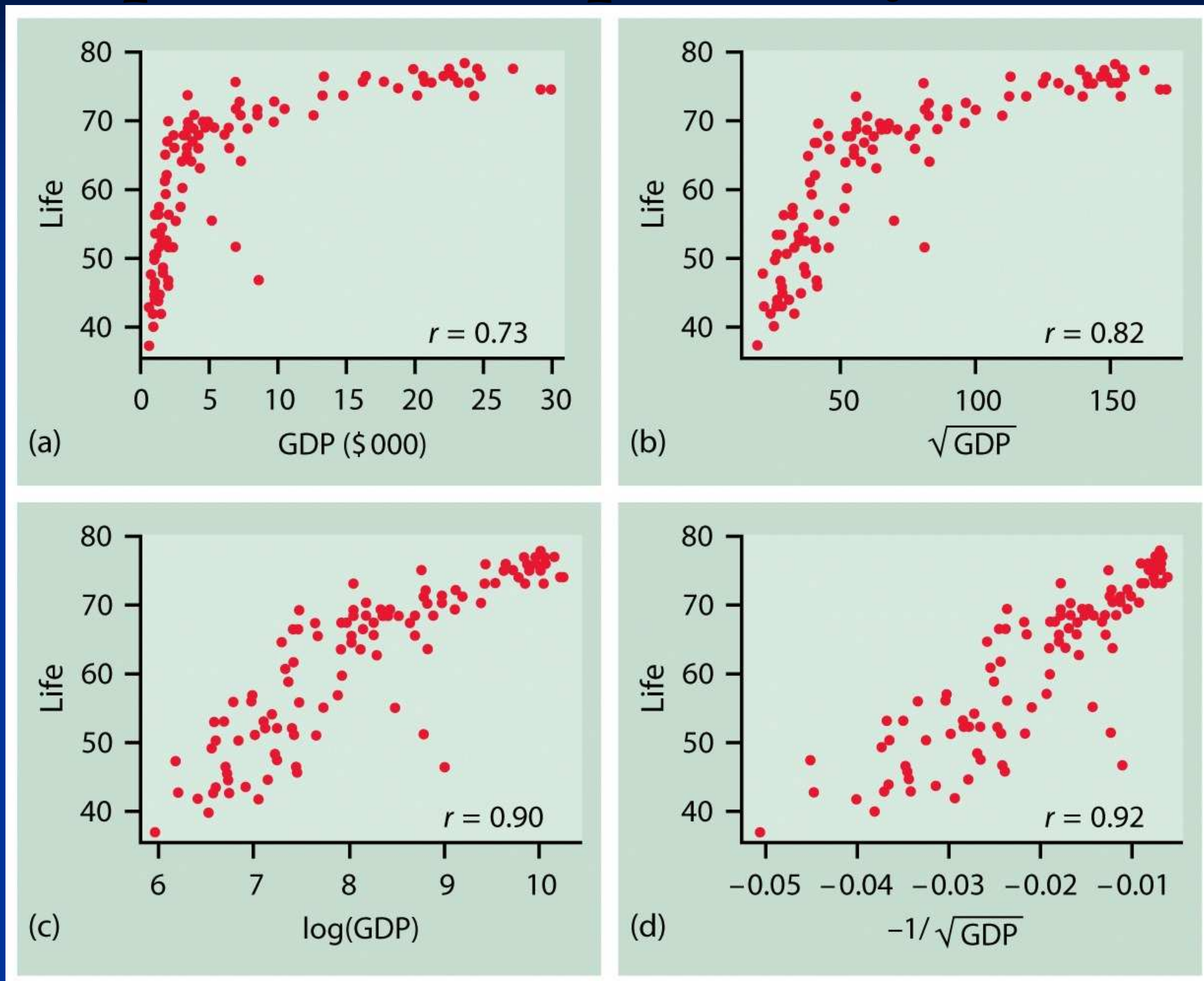
Circle of Powers



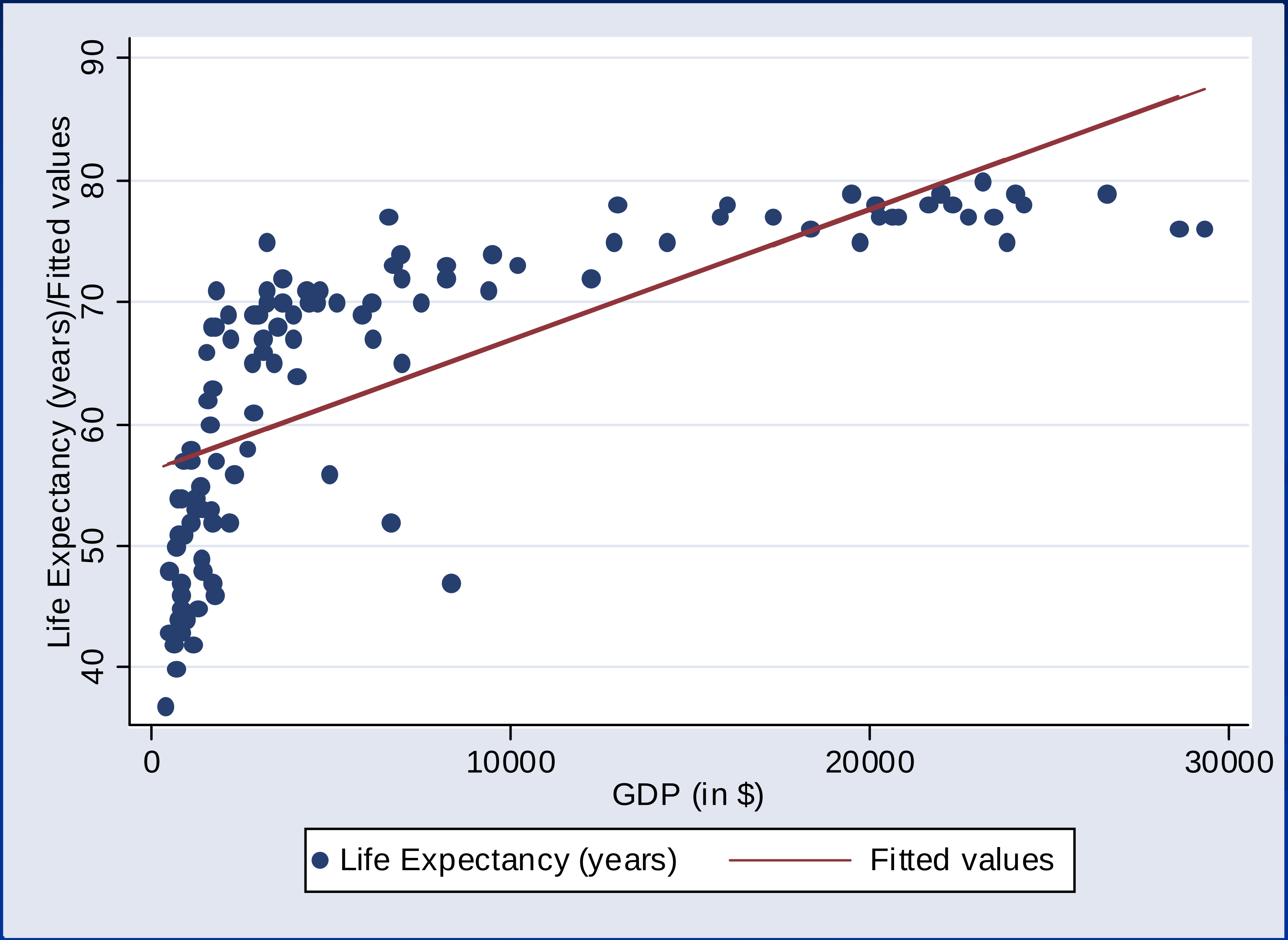
Circle of Powers

- If the pattern of the scatterplot resembles Quadrant I then we transform y up or x up, if Quadrant II then y up or x down, if Quadrant III, then y down or x down, if Quadrant IV, x up, y down. To transform up, we can square, cube ,etc.; to transform down can use square root, log, inverse, etc.

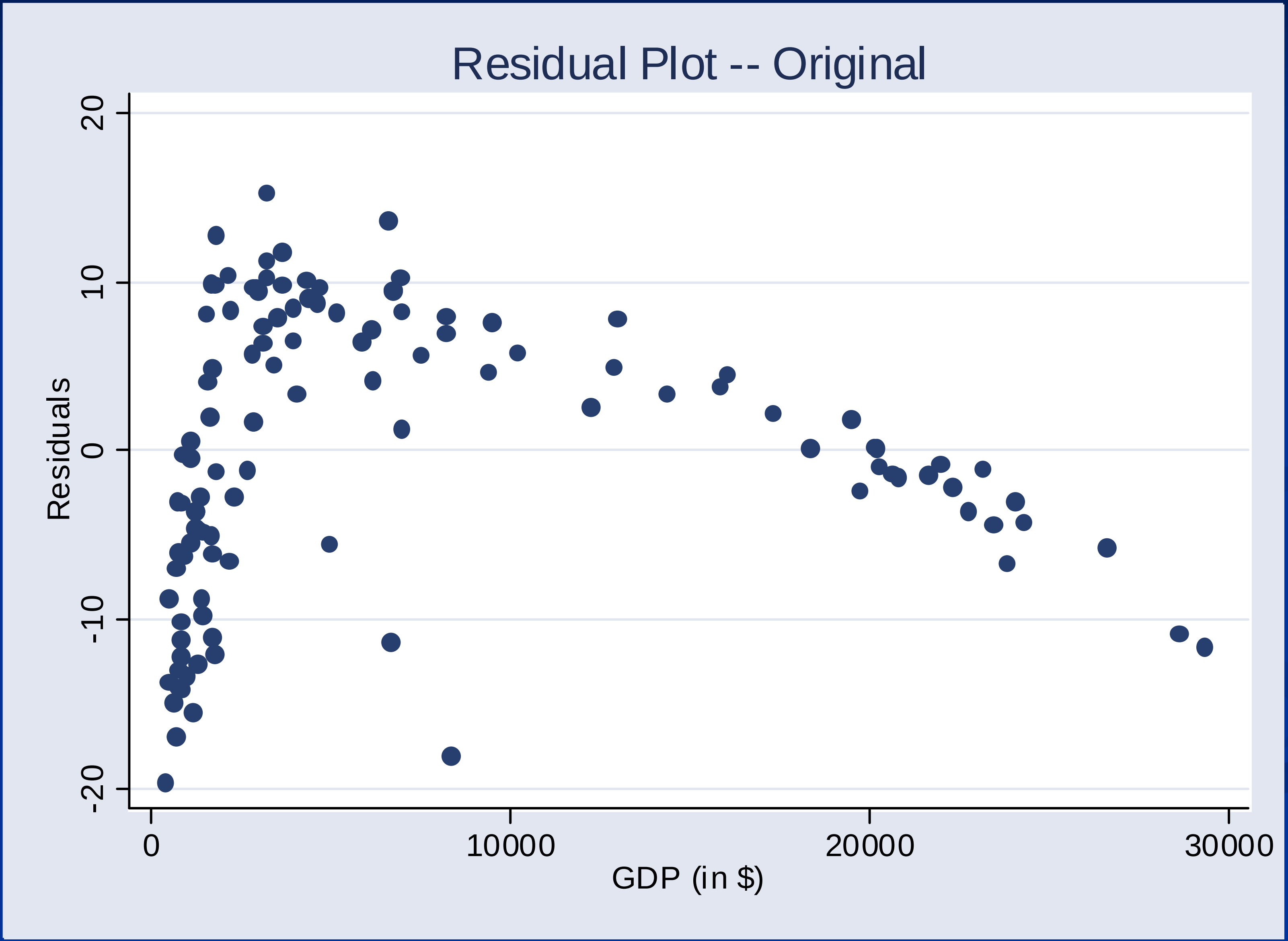
Example: Life Expectancy and GDP



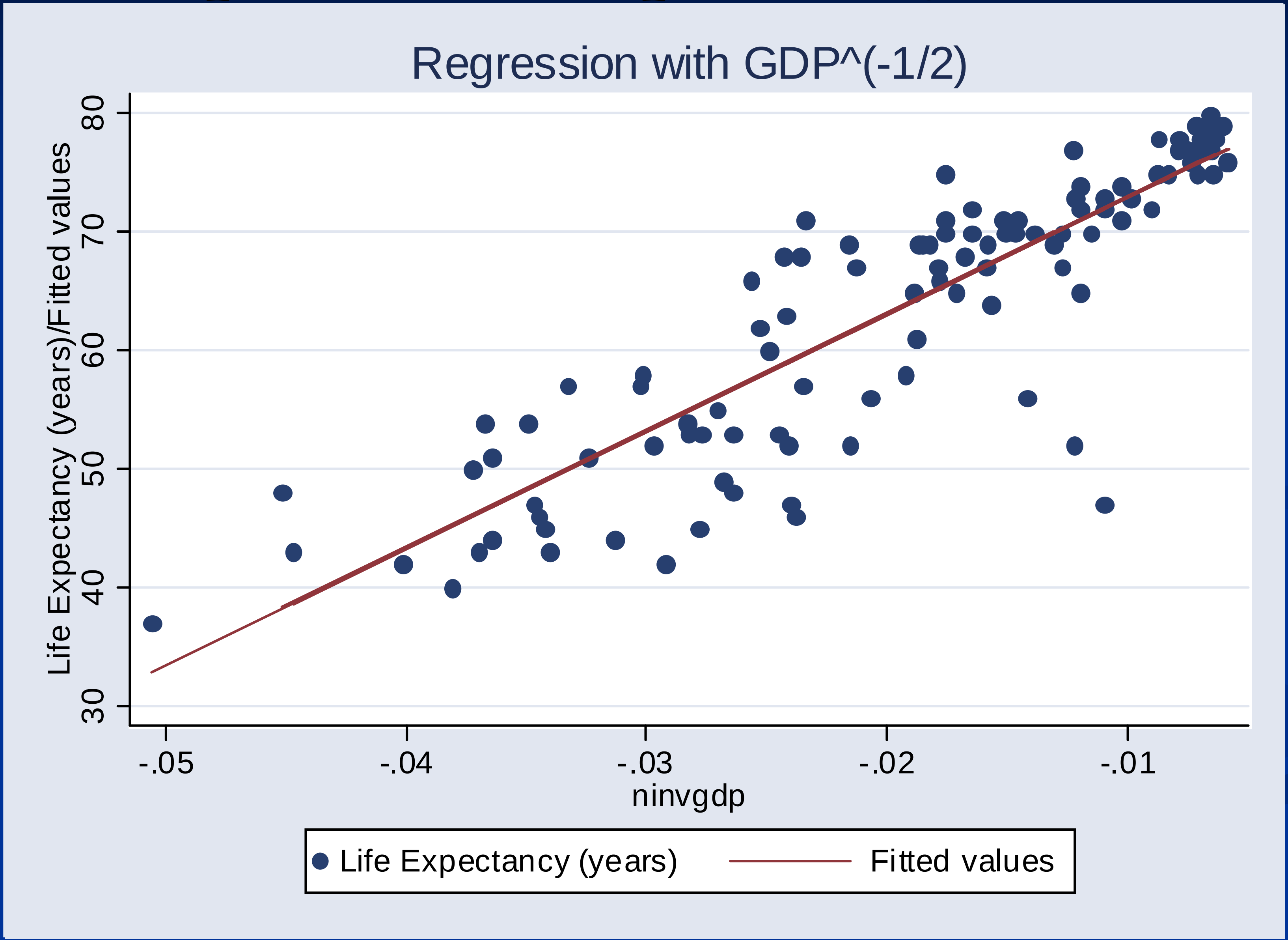
Example: Life Expectancy and GDP



Example: Life Expectancy and GDP



Example: Life Expectancy and GDP



Example: Life Expectancy and GDP