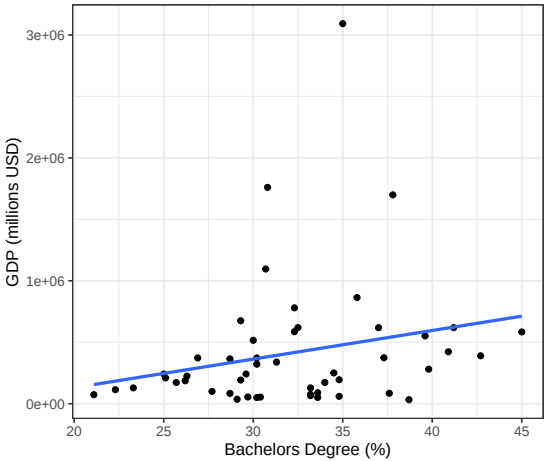
Today's Agenda

Extending the OLS Regression using Dataset 1

1. Dummy predictors
2. Categorical predictors
3. Transforming the variables
4. Transforming the model

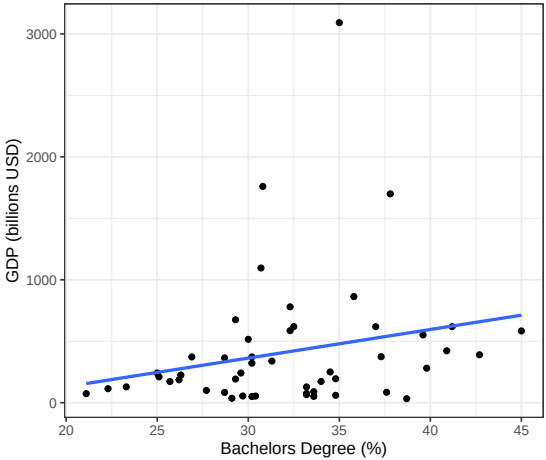
Transforming the Variables

	GDP (millions)
Bachelors (%)	23,271.42
	(14,124.28)
Intercept	−335,020.10
	(460,391.60)
Observations	50
Adjusted R^2	0.03
Residual Std. Error	528,114.80 (df = 48)
F Statistic	2.71 (df = 1; 48)
Note:	*p<0.05



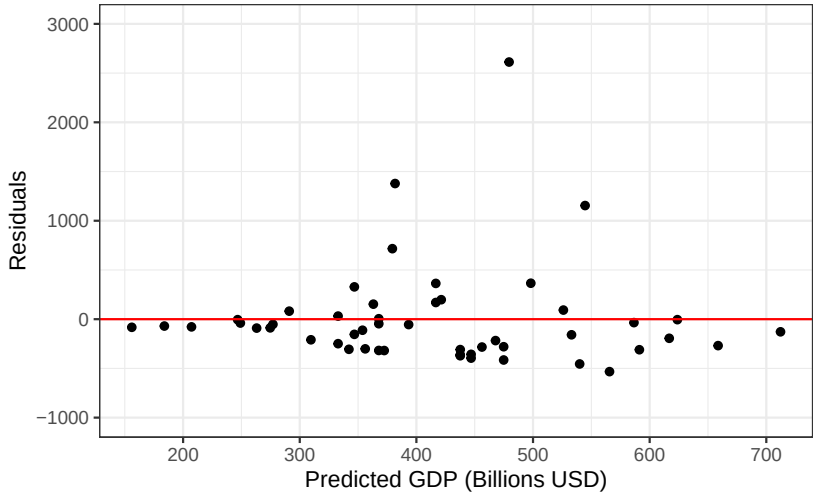
Transforming the Variables

	GDP (billions)
Bachelors (%)	23.27
	(14.12)
Intercept	−335.02
	(460.39)
Observations	50
Adjusted R^2	0.03
Residual Std. Error	528.11 (df = 48)
F Statistic	2.71 (df = 1; 48)
Note:	*p<0.05

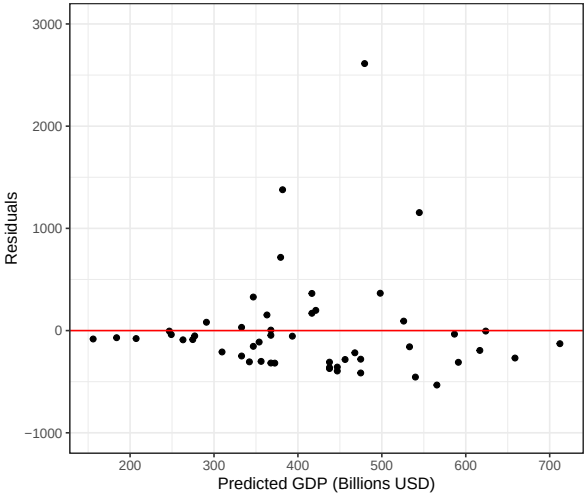


	GDP (millions)	(billions)	(10 billion)	(100 billion)
	(1)	(2)	(3)	(4)
Bachelors (%)	23,271.42	23.27	2.33	0.23
	(14,124.28)	(14.12)	(1.41)	(0.14)
Intercept	−335,020.10	−335.02	−33.50	−3.35
	(460,391.60)	(460.39)	(46.04)	(4.60)
Observations	50	50	50	50
Adjusted R^2	0.03	0.03	0.03	0.03
Residual Std. Error (df = 48)	528,114.80	528.11	52.81	5.28
F Statistic (df = 1; 48)	2.71	2.71	2.71	2.71
Note:				*p<0.05

Transformation 1: Shift the Decimal Point



Transformation 1: Shift the Decimal Point



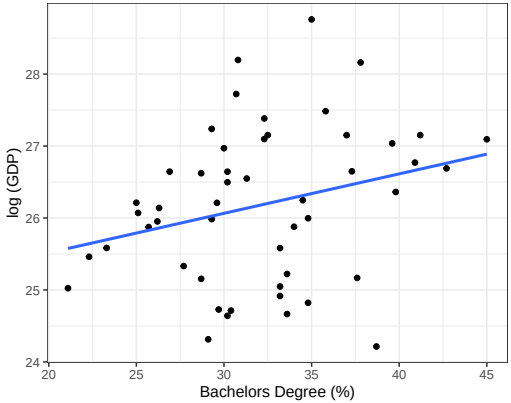
36 / 50 (72 %) are below the zero line.

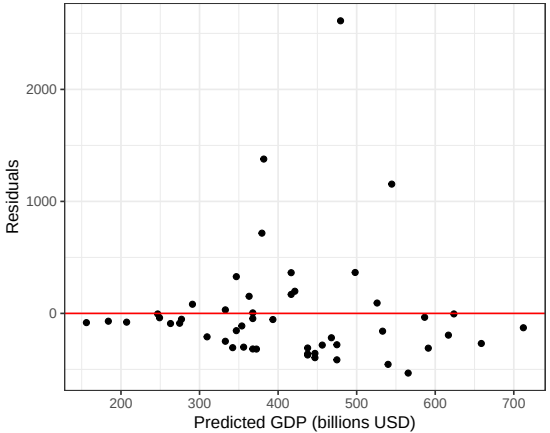
Aim is 50%-50%

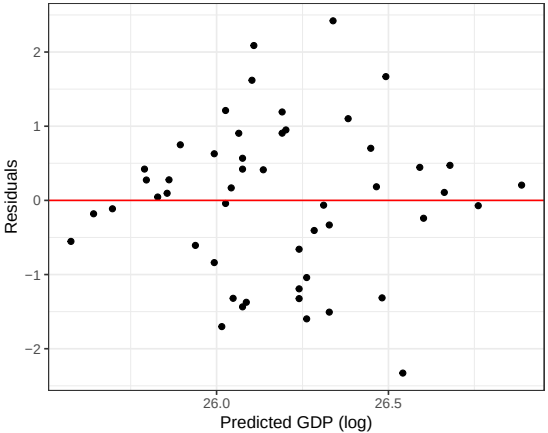
Transformation 2: Natural Logarithms

1. Create a new variable: "log GDP"
   - Excel function for the natural log is = LN()
   - GDP in Dataset 1 is in millions so you have to convert it to dollars first!
   - = LN(GDP * 1e6)
2. Practice: Regress log GDP on bachelors

	(billions)	(log)
	(1)	(2)
Bachelors (%)	23.27	0.05
	(14.12)	(0.03)
Intercept	−335.02	24.42*
	(460.39)	(0.91)
Observations	50	50
Adjusted R^2	0.03	0.06
Residual Std. Error (df = 48)	528.11	1.04
F Statistic (df = 1; 48)	2.71	3.86
Note:		*p<0.05



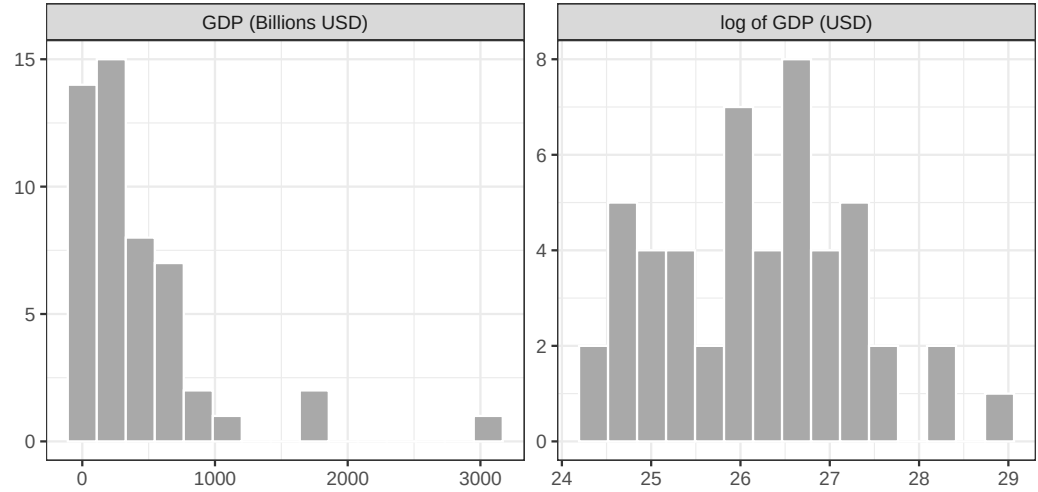




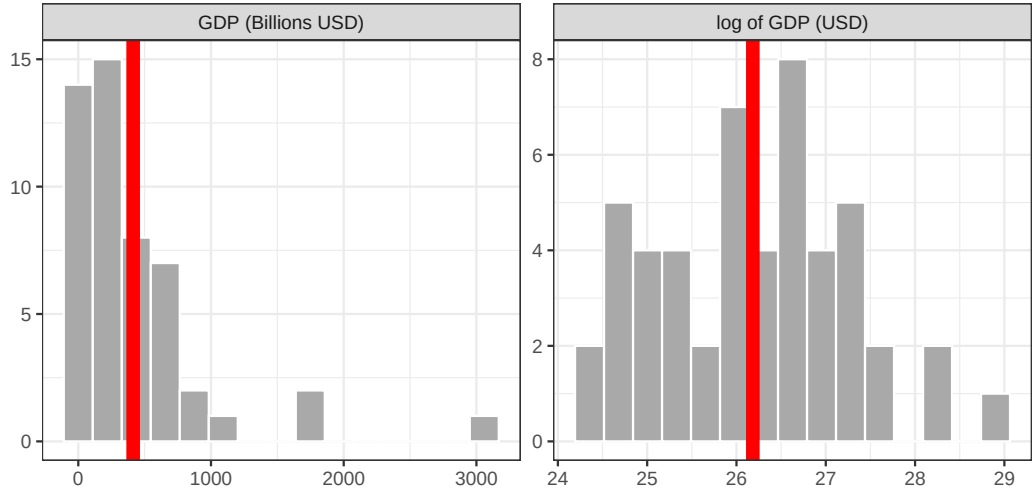
Above the line = 14 (28%)
Below the line = 36 (72%)

Above the line = 27 (54%)
Below the line = 23 (46%)

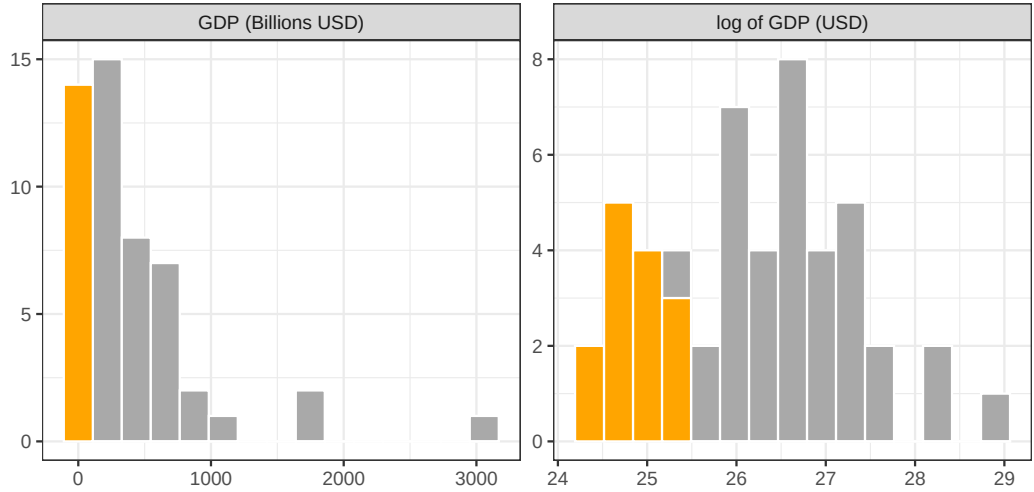
Transformation 2: Natural Logarithms



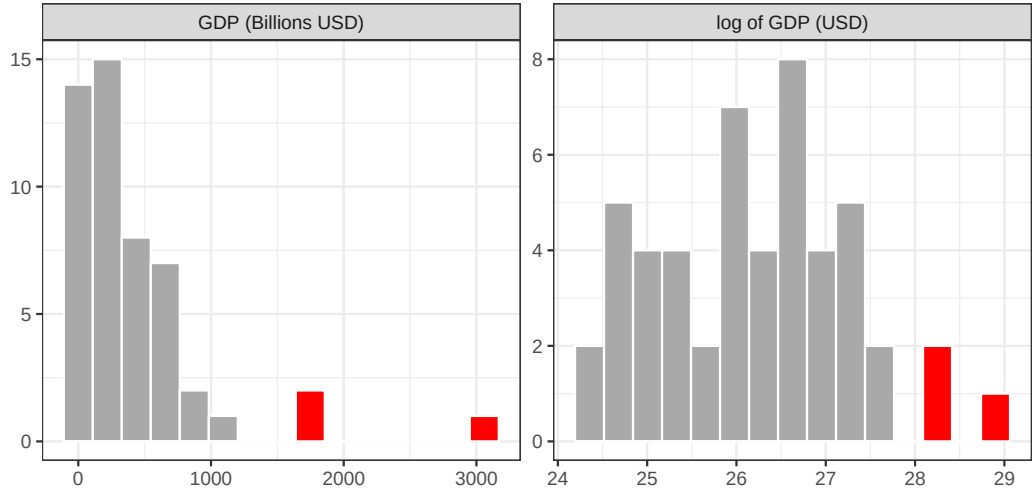
Transformation 2: Natural Logarithms



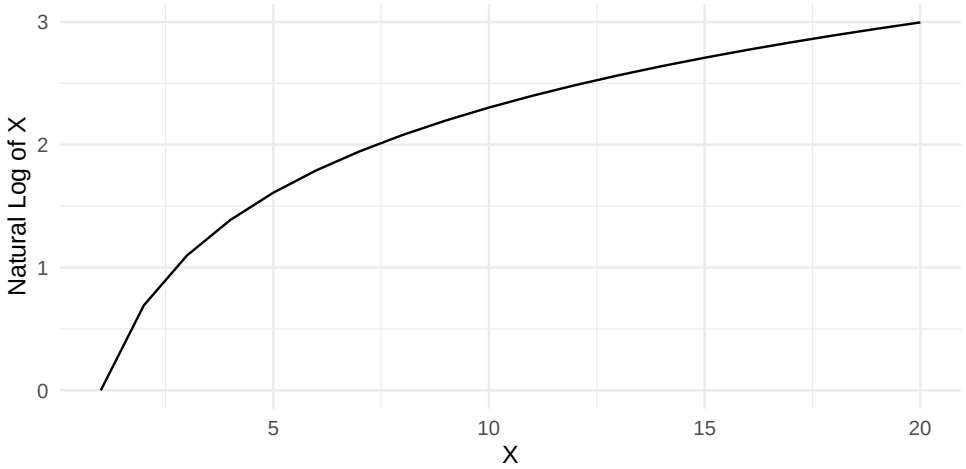
Transformation 2: Natural Logarithms



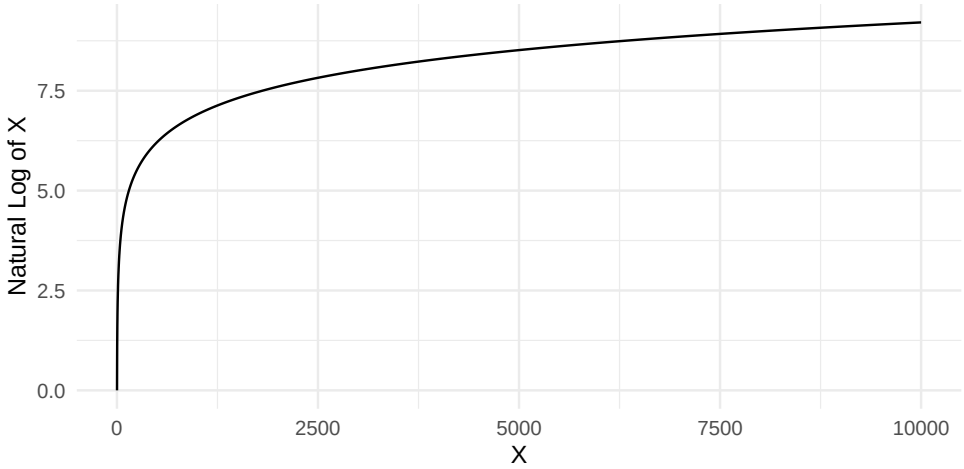
Transformation 2: Natural Logarithms



Transformation 2: Natural Logarithms



Transformation 2: Natural Logarithms



Transformation 2: Natural Logarithms

The natural log scale = multiplying by e

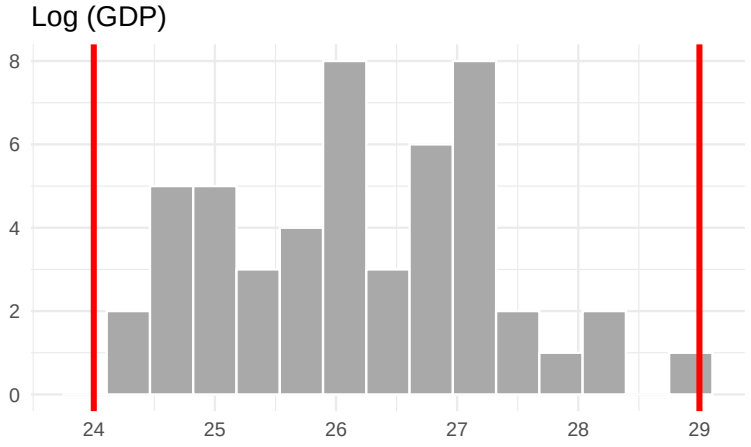
- *e* is Euler's Number (2.718282...)
- Typically written as log_e X or ln X

Transform back to linear scale using eX

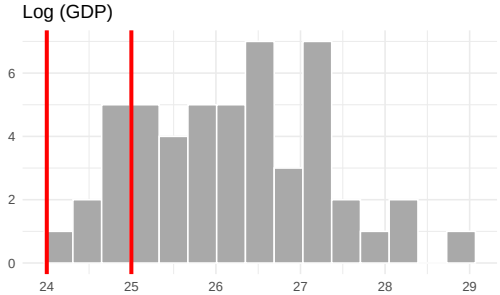
Transformation 2: Natural Logarithms

GDP for State X = \$5,000 or 8.5 (ln)

- log_e 5,000 $\approx$ 8.5
- $e^{8.5} \approx$ 5,000
- Excel function is = EXP()



- e^{24} is approximately \$26 billion
- e^{29} is approximately \$3.9 trillion



- $e^{24} \approx$ \$26 billion
- $e^{25} \approx$ \$72 billion

One unit on the LN scale:

- value x e

General rule of thumb:

- value x 3

Making Point Estimates

	(log GDP)
Bachelors (%)	0.05*
	(0.03)
Constant	24.42*
	(0.91)
Observations	50
Adjusted R^2	0.06
Residual Std. Error	1.04 (df = 48)
F Statistic	3.86* (df = 1; 48)
Note:	*p<0.06

ln (Outcome) = Intercept + Coefficient * (Predictor)

Making Point Estimates

	(log GDP)
Bachelors (%)	0.05*
	(0.03)
Constant	24.42*
	(0.91)
Observations	50
Adjusted R^2	0.06
Residual Std. Error	1.04 (df = 48)
F Statistic	3.86* (df = 1; 48)
Note:	*p<0.06

ln (GDP) = 24.42 + 0.05 * (Bachelors)

Making Point Estimates

	(log GDP)
Bachelors (%)	0.05*
	(0.03)
Constant	24.42*
	(0.91)
Observations	50
Adjusted R^2	0.06
Residual Std. Error	1.04 (df = 48)
F Statistic	3.86* (df = 1; 48)
Note:	*p<0.06

ln (GDP) = 24.42 + 0.05 * 32.16 = 26.03

Making Point Estimates

	(log GDP)
Bachelors (%)	0.05*
	(0.03)
Constant	24.42*
	(0.91)
Observations	50
Adjusted R^2	0.06
Residual Std. Error	1.04 (df = 48)
F Statistic	3.86* (df = 1; 48)
Note:	*p<0.06

GDP = 26.03 (ln) = $e^{26.03}$ = 201.7 Billion USD

Making Point Estimates

$$\text{ln(Outcome)} = \text{Intercept} + \text{Coefficient} * (\text{Predictor})$$

$$\text{Outcome} = e^{Intercept + Coefficient * (Predictor)}$$

Practice

Dataset 2: Do wealthier countries live longer?

- Regress life expectancy on GDP
- Regress life expectancy on log(GDP)

	Life Expectancy	
	(1)	(2)
GDP	0.00*	
	(0.00)	
log(GDP)		1.65*
		(0.22)
Constant	72.19*	32.07*
	(0.57)	(5.50)
Observations	173	173
Adjusted R^2	0.03	0.24
Residual Std. Error (df = 171)	7.29	6.47
F Statistic (df = 1; 171)	6.53*	54.61*
Note:		*p<0.05