

Quantitative Methods in Political Science: Linear Regression I

Thomas Gschwend | Domantas Undzėnas | Muhammad Muhammad | David Grundmanns
Week 4 - 24 September 2025

Roadmap

- Understand and model stochastic processes
- Understand statistical inference
- Implement it mathematically and learn how to estimate it
 - OLS
 - Maximum Likelihood
- Implement it using software
 - R
 - Basic programming skills

The Linear Regression Model

Estimation

The OLS Approach

Deriving the OLS Estimator

OLS Regression in Practice

Regression Diagnostics

Transformation and Nonlinearity

Statistical Inference for Linear Models

Assumptions

Standard Errors and Confidence Intervals

The Linear Regression Model

- Regression analysis examines the relationship between a dependent variable, Y , and one or more independent variables, X_1, \dots, X_k .
- The **dependent variable** is the quantity we want to explain.
 - **Examples:** party vote shares, level of corruption, income, democratization.
- The **independent variable** is the quantity that we use to explain variation in the dependent variable.
 - **Examples:** Economic, political, institutional, or demographic variables.

- Linear model: $Y = a + bX$.
 - a is the **intercept**: value of Y when X is zero.
 - b is the **slope**: change in Y for a one-unit increase in X .
- This model implies a **perfect linear relationship**.
- In actual research, this is never the case, so we need to add an **error term**:

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

- The error term (or disturbance), ϵ , represents **unobserved** factors other than X that affect Y .

Example

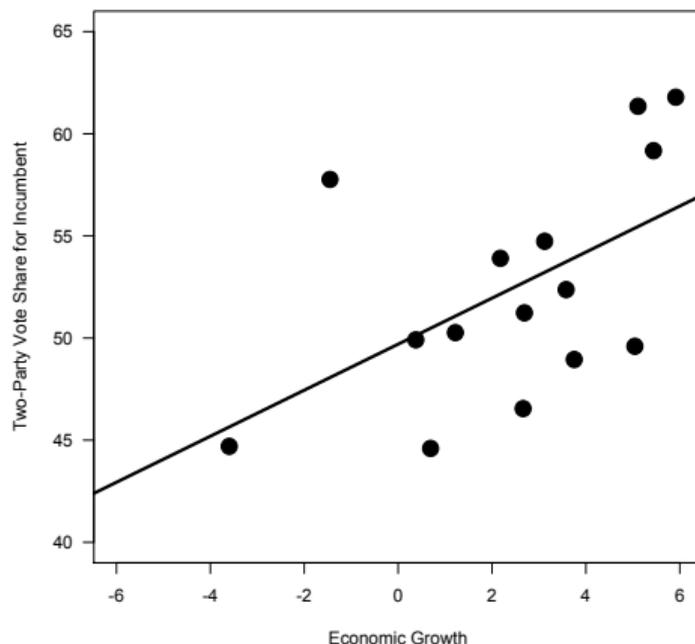
- Let's re-examine the relationship between economic growth *prior* to an election and the vote share received by the incumbent presidential party in the US.
- Let $i = 1, \dots, n$ denote observations with total sample size n .
- For each observation i , we can write:

$$\text{VoteShare}_i = \beta_0 + \beta_1 \text{Growth}_i + \epsilon_i$$

What is the relationship between X and Y?

$$\text{VoteShare}_i = \beta_0 + \beta_1 \text{Growth}_i + \epsilon_i$$

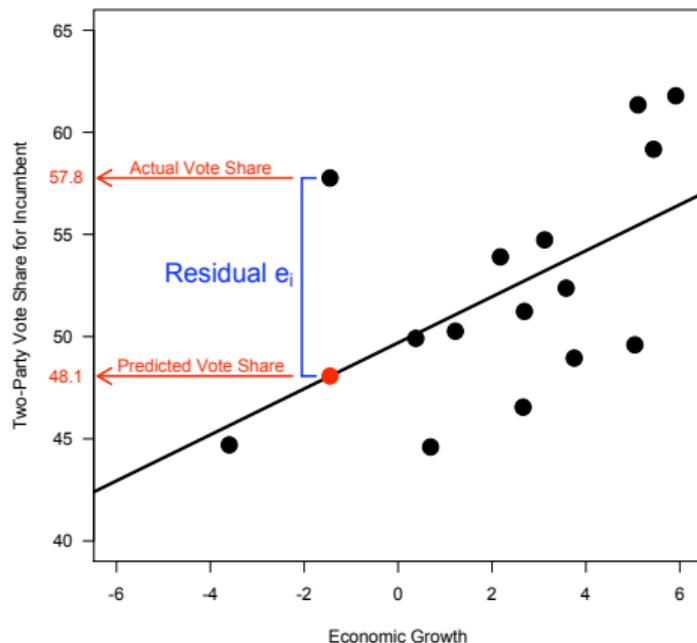
US Presidential Elections (1948-2004)



Residuals: How bad is our model prediction?

$$\text{RESIDUAL} = \text{ACTUAL} - \text{PREDICTED}$$

US Presidential Elections (1948-2004)



Estimation

How to Pick the Best Line? The OLS Approach

- OLS is short for “Ordinary Least Squares”.
- The best line is the line that **minimizes** the **sum of squared residuals** (SSR).
- Residuals are vertical deviations from the line (the observed fitting errors):

$$e_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)$$

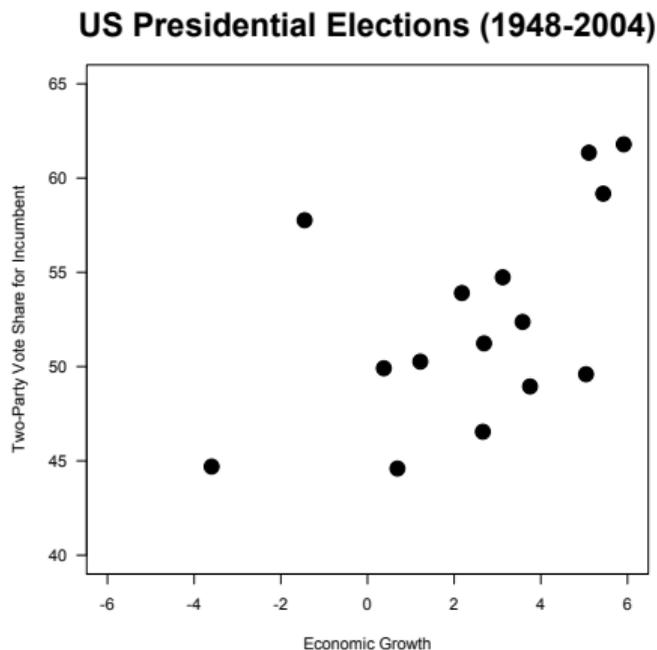
$$e_i = y_i - \hat{y}_i$$

- Conceptually, we minimize $e_i^2 = \sum_{i=1}^n (\text{ACTUAL}_i - \text{PREDICTED}_i)^2 = \sum_{i=1}^n (\text{RESIDUAL}_i)^2$.
- Mathematically, we solve the following optimization problem (“Least Squares”):

$$\min_{\hat{\beta}_0, \hat{\beta}_1} \sum_{i=1}^n e_i^2 \quad \Leftrightarrow \quad \min_{\hat{\beta}_0, \hat{\beta}_1} \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

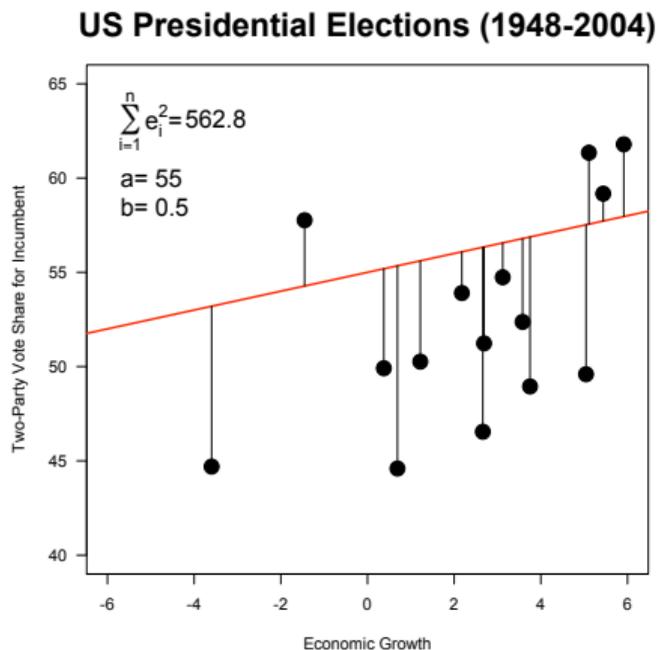
How to Pick the Best Line? The OLS Approach

We minimize the sum of squared residuals: $\min_{\hat{\beta}_0, \hat{\beta}_1} \sum_{i=1}^n e_i^2$.



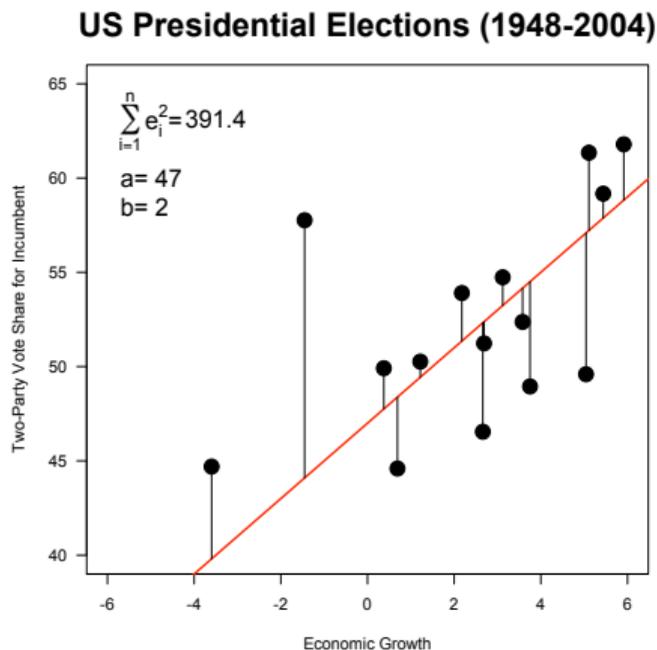
How to Pick the Best Line? The OLS Approach

We minimize the sum of squared residuals: $\min_{\hat{\beta}_0, \hat{\beta}_1} \sum_{i=1}^n e_i^2$.



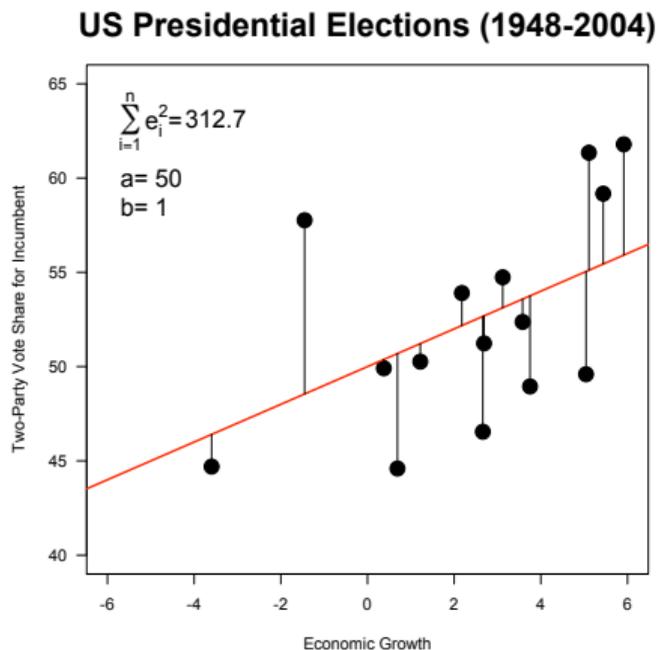
How to Pick the Best Line? The OLS Approach

We minimize the sum of squared residuals: $\min_{\hat{\beta}_0, \hat{\beta}_1} \sum_{i=1}^n e_i^2$.



How to Pick the Best Line? The OLS Approach

We minimize the sum of squared residuals: $\min_{\hat{\beta}_0, \hat{\beta}_1} \sum_{i=1}^n e_i^2$.



How to Pick the Best Line? The OLS Approach

- How to pick the best line? Get the **best slope** and **best intercept** using differential calculus.

- For $\text{Var}(x) \neq 0$, the **slope coefficient** $\hat{\beta}_1$ is given by

$$\hat{\beta}_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} = \frac{\text{Cov}(x, y)}{\text{Var}(x)}.$$

- The **intercept coefficient** $\hat{\beta}_0$ is given by

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad \text{where } \bar{y} = \sum_{i=1}^n \frac{y_i}{n} \text{ and } \bar{x} = \sum_{i=1}^n \frac{x_i}{n}.$$

- An estimator is **unbiased** if its expected value, $E(\hat{\theta})$, is **identical** to the population value, θ .
- The OLS estimator is **best** in the sense that it has the **lowest variance** among all unbiased estimators.
- The OLS estimator is said to be the **Best Linear Unbiased Estimator** (BLUE).

Digression: Getting the Coefficients

- Minimizing the sum of squared residuals implies the following **minimization problem**:

$$\min_{\hat{\beta}_0, \hat{\beta}_1} \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

- Then, we have the following two **partial derivatives**:

$$\frac{\partial \sum e_i^2}{\partial \hat{\beta}_0} = 2 \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)(-1)$$

$$\frac{\partial \sum e_i^2}{\partial \hat{\beta}_1} = 2 \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)(-x_i)$$

Digression: Getting the Coefficients ($\hat{\beta}_0$)

- With the above partial derivative, the **first-order condition** for the **first equation** is given as:

$$(-2) \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$$

- This allows us to derive an expression for $\hat{\beta}_0$.

$$(-2) \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$$

$$\sum_{i=1}^n y_i - n\hat{\beta}_0 - \hat{\beta}_1 \sum_{i=1}^n x_i = 0$$

$$\frac{\sum_{i=1}^n y_i}{n} - \frac{\hat{\beta}_1 \sum_{i=1}^n x_i}{n} = \hat{\beta}_0$$

$$\bar{y} - \hat{\beta}_1 \bar{x} = \hat{\beta}_0$$

Digression: Getting the Coefficients ($\hat{\beta}_1$)

- With the above partial derivative, the **first-order condition** for the **second equation** is given as:

$$(-2) \left(\sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)(x_i) \right) = 0$$

- This allows us to derive an expression for $\hat{\beta}_1$.

$$(-2) \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)(x_i) = 0$$

$$\sum_{i=1}^n (y_i - \bar{y} + \hat{\beta}_1 \bar{x} - \hat{\beta}_1 x_i)(x_i) = 0$$

$$\sum_{i=1}^n (y_i - \bar{y} - \hat{\beta}_1 (x_i - \bar{x}))(x_i) = 0$$

$$\sum_{i=1}^n (x_i)(y_i - \bar{y}) = \hat{\beta}_1 \sum_{i=1}^n (x_i)(x_i - \bar{x})$$

Digression: Getting the Coefficients ($\hat{\beta}_1$)

- Almost there:

$$\begin{aligned}\sum_{i=1}^n (x_i)(y_i - \bar{y}) &= \hat{\beta}_1 \sum_{i=1}^n (x_i)(x_i - \bar{x}) \\ \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) &= \hat{\beta}_1 \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x}) \\ \hat{\beta}_1 &= \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})} \\ \hat{\beta}_1 &= \frac{\text{Cov}(x, y)}{\text{Var}(x)}\end{aligned}$$

- This is exactly what we wanted to show.

OLS Regression in Practice

Least Squares Regression: Interpretation

- The **regression line** is: $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$.
- Interpreting the **slope coefficient** $\hat{\beta}_1$: *On average*, a one-unit increase in X produces a $\hat{\beta}_1$ unit increase in Y .
- More generally, the so-called **marginal effect** of an infinitesimal change in X on Y , i.e., $\frac{\partial \hat{Y}}{\partial X} = \hat{\beta}_1$, is constant (independent of X) in case of OLS.
- The **predicted value** for X is \hat{Y} .
- Interpreting the **intercept coefficient**: When X is zero, the predicted value for \hat{Y} is $\hat{\beta}_0$. Note that this may **not** be a meaningful quantity.
 - We will see next week that this is particularly important for **interaction effects**.
- The regression line **always** passes through two points:
 - Point 1: $(x_i = 0, y_i = \hat{\beta}_0)$. **Why?**
 - Point 2: $(x_i = \bar{x}, y_i = \bar{y})$. **Why?**

Least Squares Regression: Example

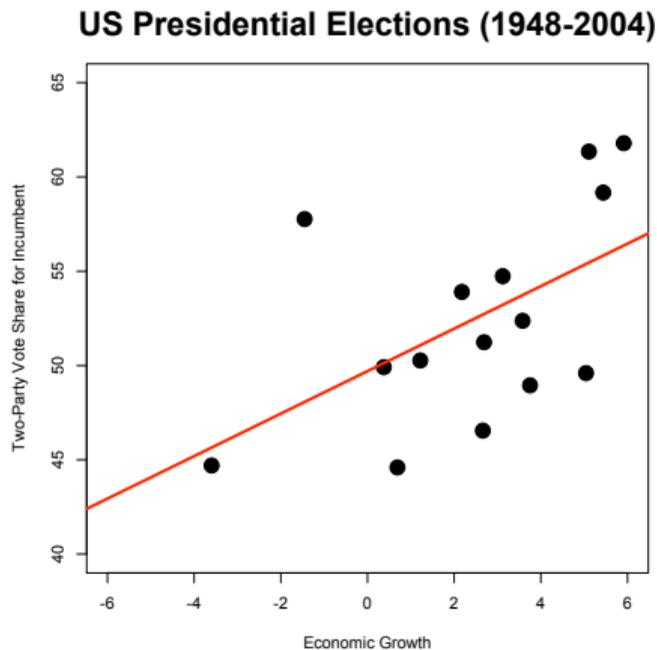
Year	VoteShare	Growth	$(y_i - \bar{y})$	$(x_i - \bar{x})$
1948	52.37	3.579	-0.088	1.131
1952	44.595	.691	-7.863	-1.757
1956	57.764	-1.451	5.306	-3.899
1960	49.913	.377	-2.545	-2.071
1964	61.344	5.109	8.886	2.661
1968	49.596	5.043	-2.862	2.595
1972	61.789	5.914	9.331	3.466
1976	48.948	3.751	-3.510	1.303
1980	44.697	-3.597	-7.761	-6.045
1984	59.17	5.440	6.712	2.992
1988	53.902	2.178	1.444	-0.270
1992	46.545	2.662	-5.913	0.214
1996	54.736	3.121	2.278	0.673
2000	50.265	1.219	-2.193	-1.229
2004	51.233	2.690	-1.225	0.242
	$\bar{y} = 52.4578$	$\bar{x} = 2.4484$		

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{111.559}{99.0181} = 1.127$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 52.4578 - 1.127 \cdot 2.4484 = 49.699$$

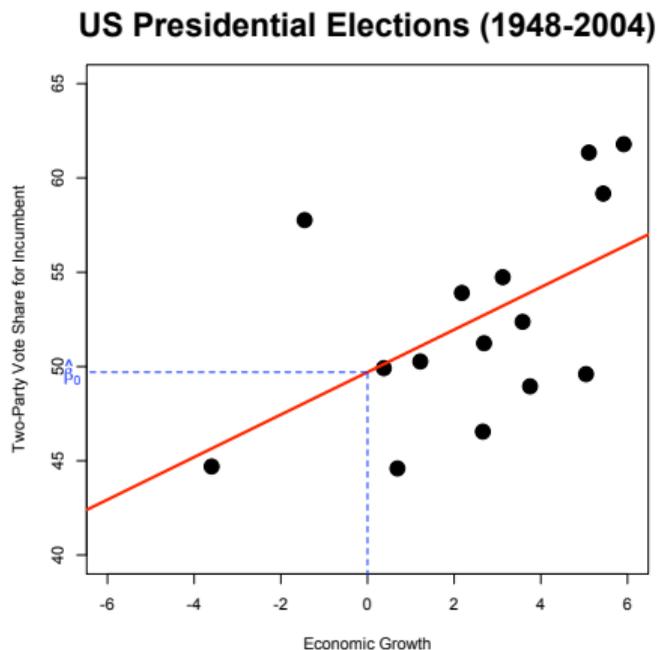
Least Squares Regression: Example

- OLS model estimation: $\widehat{VoteShare}_i = 49.699 + 1.127 \cdot Growth_i$
- The sum of squared residuals is minimized at $\sum_{i=1}^n e_i^2 = 311.1486$



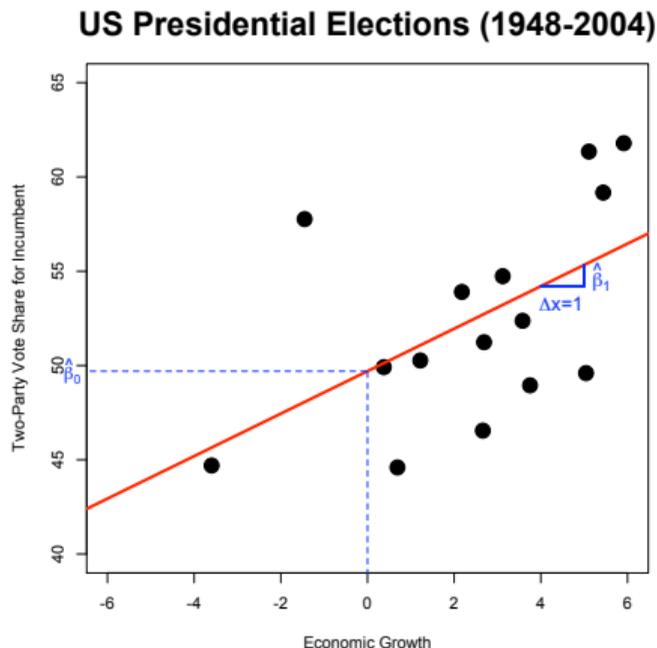
Least Squares Regression: Example

- OLS model estimation: $\widehat{VoteShare}_i = 49.699 + 1.127 \cdot Growth_i$
- The sum of squared residuals is minimized at $\sum_{i=1}^n e_i^2 = 311.1486$



Least Squares Regression: Example

- OLS model estimation: $\widehat{VoteShare}_i = 49.699 + 1.127 \cdot Growth_i$
- The sum of squared residuals is minimized at $\sum_{i=1}^n e_i^2 = 311.1486$

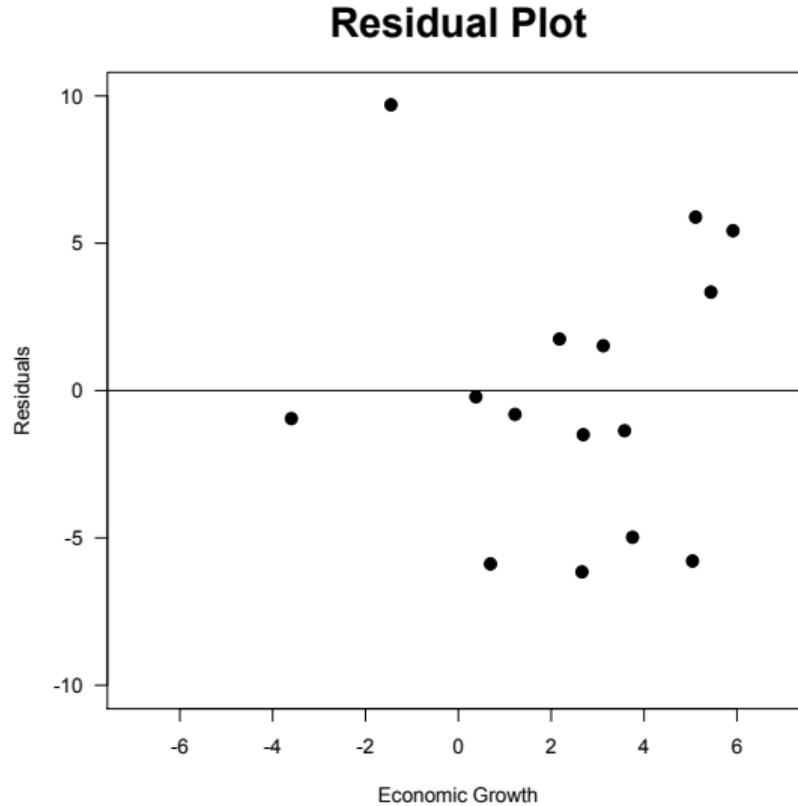


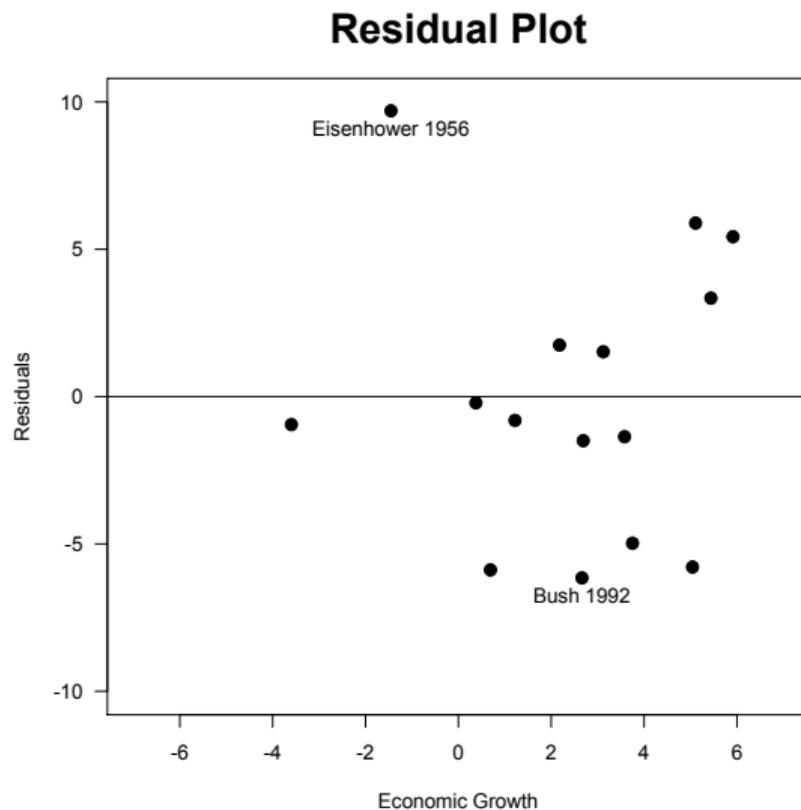
Regression Diagnostics

Regression Diagnostics: Residual Analysis ($y_i - \hat{y}_i$)

- Patterns in residuals signal that **systematic influences** on Y still have not been captured by our model, or that our model misrepresents the data, or that errors do not have a **constant variance**.
- The residual plot is a **diagnostic plot** as it helps us to detect patterns in the residuals.
- A **residual plot** is a scatterplot of the regression residuals against the explanatory variable X or the predicted values \hat{Y} .
- **Punchline:** Residual patterns diagnose model shortcomings.
- Ideally, residuals plots should look as if the pattern was generated by pure chance.
- By construction (first-order condition of $\hat{\beta}_0$ and $\hat{\beta}_1$), OLS residuals sum to zero:

$$\sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = \sum_{i=1}^n (y_i - \hat{y}_i) = \sum_{i=1}^n e_i = 0.$$





- How well does our model explain the variation in the dependent variable?
- Disaggregate the variance of Y into that part that we have explained by X and that part that we have not explained.
 - **Explained sum of squares** (ESS) = $\sum_{i=1}^n (\hat{y}_i - \bar{y})^2$
 - **Residual sum of squares** (RSS) = $\sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n e_i^2$
 - **Total sum of squares** (TSS) = $\sum_{i=1}^n (y_i - \bar{y})^2$
- **Remember:** TSS=RSS+ESS

Regression Diagnostics: Goodness-of-Fit

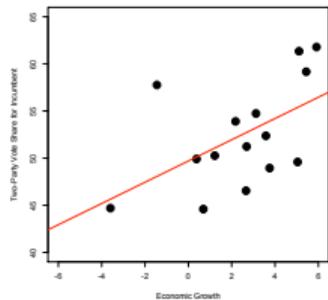
$$\frac{\text{Explained Variance}}{\text{Total Variance}} = 1 - \frac{\text{Unexplained Variance}}{\text{Total Variance}} = \text{Goodness-of-fit}$$

$$\frac{ESS}{TSS} = \frac{TSS - RSS}{TSS} = 1 - \frac{RSS}{TSS} = R^2$$

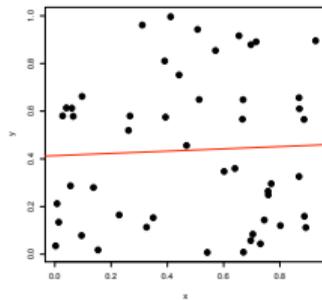
- Interpretation: Proportion of the total variance **explained by the fitted model**.
- The goodness-of-fit measure is bounded: $0 \leq R^2 \leq 1$
- For a bivariate linear regression model, R^2 is identical to the squared Pearson's r correlation coefficient of x and y .
- Note the following two caveats:
 - R^2 is **not resistant** to outliers. One outlier can distort the value.
 - R^2 **increases** with the addition of more explanatory variables.

Regression Diagnostics: Goodness-of-Fit

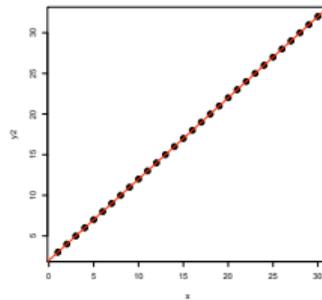
$$R^2 = 0.288$$



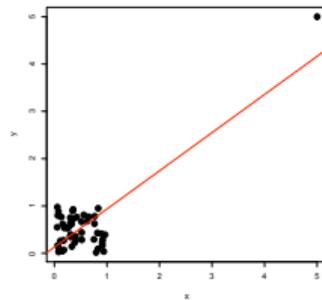
$$R^2 = 0.002$$



$$R^2 = 1$$



$$R^2 = 0.664$$



Regression Diagnostics: Goodness-of-Fit

- R^2 **increases** the more explanatory variables we add.
- This is due to the fact that the sum of squared residuals never goes up as more variables are added.
- The **adjusted R^2** , therefore, imposes a penalty for adding independent variables.
- If an independent variable is added to a regression, the RSS falls, but so do the **degrees of freedom** in the regression model.
- While the regular R^2 is bounded in $[0, 1]$, the adjusted R^2 even can become negative. This indicates a **bad model fit**.
- The adjusted R^2 for sample size, n , and k independent variables is defined as

$$\text{Adj. } R^2 = 1 - (1 - R^2) \left(\frac{n-1}{n-k-1} \right).$$

- Example: US Presidential Election Data.

$$R^2 = 0.808 \text{ and } \text{Adj. } R^2 = 1 - (1 - 0.808) \left(\frac{15-1}{15-1-1} \right) = 0.793$$

Transformation and Nonlinearity

What does linear in Linear Regression actually mean?

- The linearity assumption refers to **linearity in parameters** only.
- This allows for **nonlinearities** in variables, e.g., \sqrt{x} , $\log(x)$.
- Suppose you want to model the following:

$$Y = \beta_0 X^{\beta_1} \cdot \epsilon$$

Then, $\log(Y) = \log(\beta_0) + \beta_1 \log(X) + \log(\epsilon)$

Which is $\tilde{Y} = \tilde{\beta}_0 + \tilde{\beta}_1 \tilde{X} + \tilde{\epsilon}$

and can be estimated via OLS.

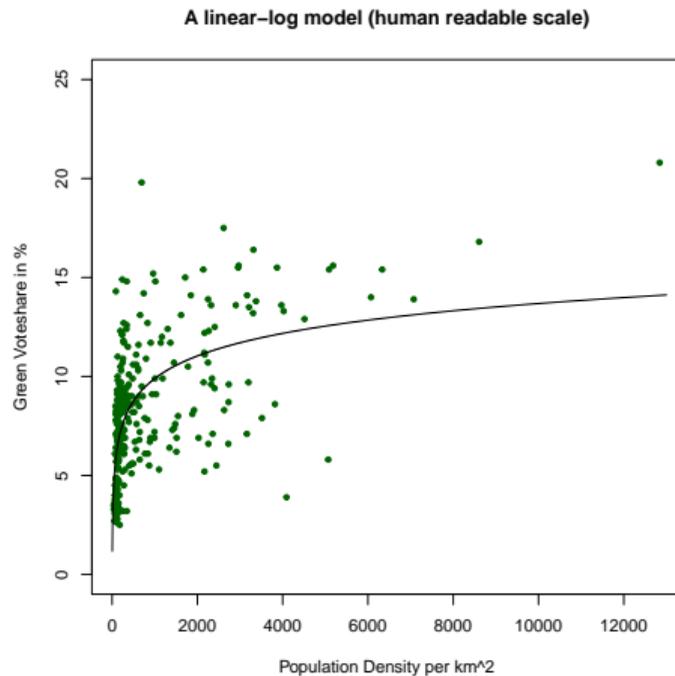
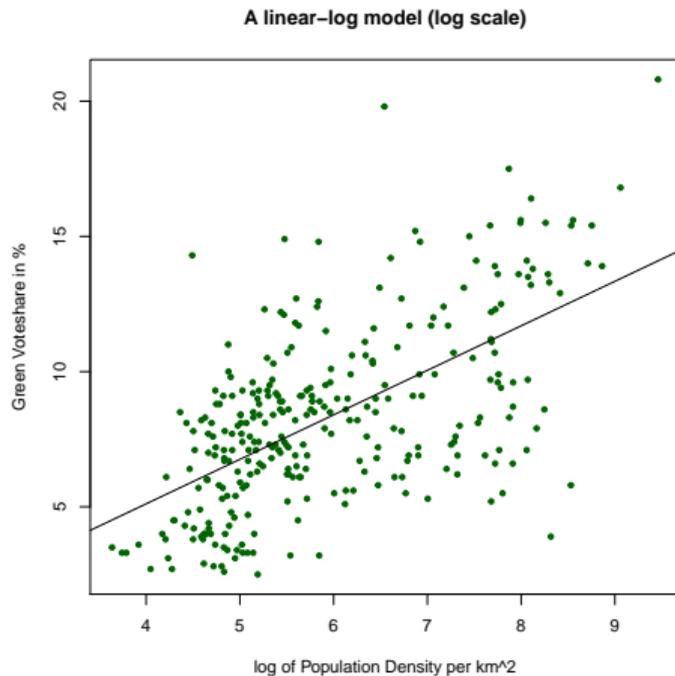
- Interpretation of the estimates $\hat{\beta}_1$ (and $\hat{\beta}_0$ respectively)?

Logarithmic Transformation of Variables

- In applied work, you will sometimes encounter a dependent variable or a covariate that is **logarithmically transformed**:
 - Log-transformed covariates makes sense, if we theoretically expect an nonlinear decreasing impact of X on Y , e.g., the effect diminishes if X increases (e.g., GDP, district magnitude, pop. density).
 - Log-transformed dependent variable makes sense, if the residuals are not nearly normally distributed (e.g., rightly skewed) and inclusion of further variables did not help (to fix it afterwards).
- Our goal to make the relationship between two variables more linear through transforming (one or both of) them.

Logarithmic Transformation of Variables

- Interpretation of coefficient changes with respect to untransformed variables.



Statistical Inference for Linear Models

Classical OLS Assumptions

Suppose we have the following bivariate linear model:

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

We need **two** assumptions to derive **unbiased** regression coefficients, $\hat{\beta}_0$ and $\hat{\beta}_1$.

- A1: An almost trivial assumption is that coefficients (i.e., parameters) are **linear**.
- A2: We make a **zero conditional mean** assumption:

$$E(\epsilon_j | X) = 0$$

- These assumptions are sufficient to estimate **unbiased coefficients**, $\hat{\beta}_0$ and $\hat{\beta}_1$, with OLS.

Classical OLS Assumptions

To also estimate the **variance** of the coefficients, we need to make additional assumptions.

- A3: We assume constant variance, which is known as **homoskedasticity**, regardless of the values of X :

$$\text{Var}(\epsilon_i | X) = \sigma^2$$

- A4: We assume **no correlation** among any pair of error terms:

$$\text{Cov}(\epsilon_i, \epsilon_j | X_i, X_j) = 0 \quad \forall i \neq j$$

- A5: We assume **normality** of the error term:

$$\epsilon_i | X \sim \mathcal{N}(0, \sigma^2)$$

Standard Errors for Regression Coefficients

- If A3, the **zero conditional mean assumption**, i.e., $E(\epsilon_i | X) = 0$, holds, we get

$$E(\hat{\beta}_0) = \beta_0$$

$$E(\hat{\beta}_1) = \beta_1.$$

- Assuming A5, **normally distributed errors**, $\epsilon | X \sim \mathcal{N}(0, \sigma^2)$, the OLS coefficients themselves are normally distributed.

$$\hat{\beta}_0 \sim \mathcal{N}(\beta_0, \text{Var}(\hat{\beta}_0))$$

$$\hat{\beta}_1 \sim \mathcal{N}(\beta_1, \text{Var}(\hat{\beta}_1))$$

- This allows us to calculate **standard errors** based on normal approximation.

Standard Errors for Regression Coefficients

- Variance and standard error for our estimated **slope coefficient**, $\hat{\beta}_1$, is:

$$\text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$
$$\text{SE}(\hat{\beta}_1) = \sqrt{\frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}$$

- Variance and standard error for our estimated **intercept coefficient**, $\hat{\beta}_0$, is:

$$\text{Var}(\hat{\beta}_0) = \frac{\sigma^2 \frac{\sum_{i=1}^n x_i^2}{n}}{\sum_{i=1}^n (x_i - \bar{x})^2}$$
$$\text{SE}(\hat{\beta}_0) = \sqrt{\frac{\sigma^2 \frac{\sum_{i=1}^n x_i^2}{n}}{\sum_{i=1}^n (x_i - \bar{x})^2}}$$

- Before we can estimate standard errors, we need to estimate $\hat{\sigma}^2$ because we do not observe σ^2 .

Standard Errors for Regression Coefficients

- While the regression error variance, σ^2 , is **inherently unobservable**, we can fortunately estimate it from the model residuals e_i .
- In the **bivariate** model (because we have **two optimization conditions**) an unbiased estimator for the error variance (aka *residual variance*) is given as

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n e_i^2}{n - 2}.$$

- This can be generalized to get an unbiased estimator in a **multiple regression model** with k independent variables and one intercept (i.e., $k + 1$ parameters):

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n e_i^2}{n - (k + 1)} = \frac{\sum_{i=1}^n e_i^2}{n - k - 1}.$$

- Thus, we get the **root mean squared error (RMSE)** aka **standard error of the estimate** (*How far is the model off on average?*) of Y as:

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^n e_i^2}{n - k - 1}}.$$