Classical Normal Linear Regression Model

Jamie Monogan

University of Georgia

Intermediate Political Methodology
Objectives

By the end of this meeting, participants should be able to:

- Explain why the distributional assumption about the disturbance term is important for inference.
- Explain the implications of the normality assumption for disturbances and the reasons we make this assumption.
- Name the properties of OLS estimators under the normality assumption.
The Probability Distribution of $u_i$ Matters

We can define the estimator of $\hat{\beta}_2$ as:

$$\hat{\beta}_2 = \sum k_i Y_i$$

where

$$k_i = \frac{X_i - \bar{X}}{\sum(X_i - \bar{X})^2}.$$

Our population regression model is $Y_i = \beta_1 + \beta_2 X_i + u_i$, so by substitution:

$$\hat{\beta}_2 = \sum k_i(\beta_1 + \beta_2 X_i + u_i).$$

Regressors are assumed fixed, as are population parameters. Hence, $u_i$ is the only random quantity in this equation. Hence, the assumption we make about the probability distribution of $u_i$ will affect the sampling distribution of $\hat{\beta}_2$. 
The Normality Assumption for $u_i$

We commonly assume that the disturbances have a Normal (or Gaussian) distribution:

- **Mean:** $E(u_i) = 0$
- **Variance:** $E[u_i - E(u_i)]^2 = E(u_i^2) = \sigma^2$
- **Covariance:** $cov(u_i, u_j): E\{[u_i - E(u_i)][u_j - E(u_j)]\} = E(u_iu_j) = 0 \text{ for } i \neq j$

More compactly, we say: $u_i \sim \mathcal{N}(0, \sigma^2)$. Our third assumption implies that each disturbance is independent of all of the others. Thus, we say that the disturbances are independently and identically distributed: $u_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$ or $u_i \sim \text{NID}(0, \sigma^2)$. 
Recall: The Normal (or Gaussian) Distribution

Properties of the Distribution

- $X \sim \mathcal{N}(\mu, \sigma^2)$
- $x \in \mathbb{R}$
- $\mu \in \mathbb{R}$
- $\sigma > 0$
- $p(x) = \frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
- Common special case: $X \sim \mathcal{N}(0, 1)$, “standard normal”
- $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$
- $\Phi(x) =$ standard normal CDF
Normal Distribution
Probability Density Function and Cumulative Distribution Function

$X \sim N(0,1)$, PDF

$X \sim N(0,1)$, CDF
1. By the **central limit theorem** (CLT), the sum of a large number of independent and identically distributed random variables converges to a Normal distribution.

2. Even with a small number of variables or variables that are not strictly independent, the sum may still have a Normal distribution.

3. Ease of inference: The sum of normally distributed variables has a Normal distribution. Hence, $\hat{\beta}_1$ and $\hat{\beta}_2$ each have a Normal sampling distribution.

4. Many natural phenomena follow the Normal distribution, it is well known, and it is relatively simple with only two parameters.

5. With a small sample size, the Normality assumption allows us to use $t$, $F$, and $\chi^2$ tests in regression models.

6. In large data sets, deviations from the Normal assumption are not so critical.
Properties of OLS under the Normality Assumption

Remember: Items #1 & 2 came last week without Normality.

1. Unbiased.
2. Efficient.
3. Consistent.

4. \( \hat{\beta}_1 \sim \mathcal{N}(\beta_1, \sigma_{\hat{\beta}_1}^2) \) and \( Z = \frac{\hat{\beta}_1 - \beta_1}{\sigma_{\hat{\beta}_1}} \) is distributed \( Z \sim \mathcal{N}(0, 1) \).

5. \( \hat{\beta}_2 \sim \mathcal{N}(\beta_2, \sigma_{\hat{\beta}_2}^2) \) and \( Z = \frac{\hat{\beta}_2 - \beta_2}{\sigma_{\hat{\beta}_2}} \) is distributed \( Z \sim \mathcal{N}(0, 1) \).

6. \( \frac{(n - 2)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-2} \)

7. \((\hat{\beta}_1, \hat{\beta}_2)\) are distributed independently of \( \hat{\sigma}^2 \).

8. \( \hat{\beta}_1 \) and \( \hat{\beta}_2 \) are the best unbiased estimators (BUE).

Heavy emphasis on sampling distributions here.
Gauss-Markov Assumptions in Verbal and Scalar Form

1. Correct specification ($E\{u_i|X_i\} = 0$ for $i = 1, \ldots, n$)
2. Nonstochastic regressors ($\{u_1, \ldots, u_n\}$ and $\{X_1, \ldots, X_n\}$ are independent for $i = 1, \ldots, n$)
3. Error properties
   a. Homoscedastic ($\text{var}\{u_i\} = \sigma^2$ for $i = 1, \ldots, n$)
   b. Independent ($\text{cov}\{u_i, u_j\} = 0$ for $i, j = 1, \ldots, n$ where $i \neq j$)
   c. Normally distributed ($u_i \sim \mathcal{N}(0, \sigma^2)$)

Assumption Sets

- The “weak set” includes all but 3(c): sufficient for properties of $\beta$ (and proof of Gauss-Markov)
- The “strong set” includes 3(c): sufficient for inference about $\beta$
Gauss-Markov Theorem

- Under the “weak set” assumptions OLS is BLUE
  - (B)est (i.e., efficient)
  - (L)inear
  - (U)nbiased
  - (E)stimator

- Under the “strong set,” we are poised for drawing inferences about the population with confidence intervals and hypothesis testing.
For Next Time

- Look at Equation 4.3.6. Is \( \sigma \hat{\beta}_2 \) a known quantity? Why or why not?
- What would it mean if we instead wrote the equation \( T = \frac{\hat{\beta}_2 - \beta_2}{\hat{\sigma}_{\hat{\beta}_2}} \)?

  How does \( T \) differ substantively from Equation 4.3.6?

- The quantity \( \hat{\sigma}_{\hat{\beta}_2} \) has a \( \chi^2 \) distribution with \( n - 2 \) degrees of freedom. Thinking back to last semester, what distribution does \( T \) have? Why?

- Look at the regression analysis you did by hand of the 10 Georgia congressional elections in 2010, and incorporate your estimate of \( \hat{\beta}_2 \) and \( \hat{\sigma}_{\hat{\beta}_2} \) into this question. Suppose you believe that the better Obama did in a district the worse the Republican congressional candidate did. That means your hypothesis is:
  \[ H_0 : \beta_2 = 0, \quad H_A : \beta_2 < 0. \]

  In the context of testing this hypothesis, what value will \( T \) take on? (Show me the value of each component of \( T \), then calculate a final value.) What is the distribution of \( T \)?