Reminder TD_3

Marion Brouard, Pauline Leveneur

October 27, 2025

- **Objective of the review:**
  1. Review some asymptotic properties of estimators
  2. Learn how to perform and interpret tests:
     - t-test
     - F-test
  3. Learn to interpret a p-value
  4. Concept of homoskedasticity

1 Asymptotic Properties

- **Theorem 1 (Chapter 2):**

If $(Y_i, X_i)_{i=1,...,n}$ are iid with the same distribution as (Y,X), $E(||X||^2) < \infty$, $E(\varepsilon^2||X||^2) < \infty$ and $E(XX')$ is invertible, then:

$$\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{d} \mathcal{N}(0, E(XX')^{-1}E(\varepsilon^2 XX')E(XX')^{-1}) \tag{1}$$

i.e., $\hat{\beta}_j$ is asymptotically normal. The larger n is, the closer the distribution of the estimator approaches the normal distribution (*Question 3*). (Note: If the estimator meets the Gauss-Markov assumptions and the error is normally distributed, then the estimator follows a normal distribution conditional on the covariates[1])

In *Question 2*, we show that the estimator is not asymptotically normal if ε has no expectation. (Example: Cauchy distribution)

- **Asymptotic Variance:**

In the case of simple regression, if the true model is homoskedastic (i.e. $V(\varepsilon|X) = \sigma^2$ or in the lecture $E(\varepsilon^2 XX') = E(\varepsilon^2)E(XX')$), the asymptotic variance can be written as:

$$\boxed{V_a(\hat{\beta}_j) = \frac{\sigma^2}{V(X_j)}} \tag{2}$$

[1] Reminder: The OLS estimator exists and is unbiased if we have:

- A random sample
- A theoretical model written as $y = \beta_0 + \beta_1 x_{i1} + ... + \beta_p x_{ip} + \varepsilon$ (linearity in parameters)
- X is of full rank (no multicollinearity)
- Strong exogeneity $E(\varepsilon|X) = 0$

If we also add the homoskedasticity condition (i.e. $V(\varepsilon|X) = \sigma^2$), then we meet the Gauss-Markov assumptions, concluding that the OLS estimator is BLUE (Best Linear Unbiased Estimator)

Thus, the variance of $\hat{\beta}_j$ (resp. the precision of $\hat{\beta}_j$) increases (resp. decreases) when the error variance increases and decreases (resp. increases) when the variance of X_j increases. *(Question 4)*

2 Hypothesis Testing

If the assumptions of Theorem 1 hold, and $E(||X^4||) < \infty$ and $E(|\varepsilon| \times ||X||^3) < \infty$, then the tests in this section are consistent. *(Cf Proposition 4 Ch2)*

2.1 Test for One Parameter

cf. Chapter 2, slide 17

We want to know if, all other things being equal, x_j has a causal effect on y. So, we construct a test to see if β_j is equal to 0.

We define the **null hypothesis**

$$H_0 : \beta_j = 0 \tag{3}$$

Note that we are interested in the true parameter, so the hypothesis is about the "population parameter" β_j, not the estimator.

We choose the **level of the test**. We construct a test to reject H_0 at a significance level α. (i.e., the probability of rejecting H_0 when it is true. For example, if $\alpha = 0.05$, we take a 5% risk of wrongly rejecting H_0.)

We construct the **t-statistic** or **t-ratio** as follows:

$$\boxed{t_{\hat{\beta}_j} \equiv \frac{\hat{\beta}_j}{se(\hat{\beta}_j)}} \tag{4}$$

Reminder: This test is constructed from the assumptions of Property 4, under which:

$$\frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \sim t_{n-k-1} \qquad \text{and} \qquad \frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \xrightarrow{d} \mathcal{N}(0,1)$$

A sufficiently distant value of $t_{\hat{\beta}_j}$ from 0 will allow rejecting the null hypothesis. The rejection rule depends on the **alternative hypothesis** H_1. The left-hand side will enable us to build inference tests when estimating on a finite sample. The right-hand side result will be useful for building asymptotic tests. Asymptotically, both tests are equivalent but the finite sample test requires that the error terms are normally distributed.

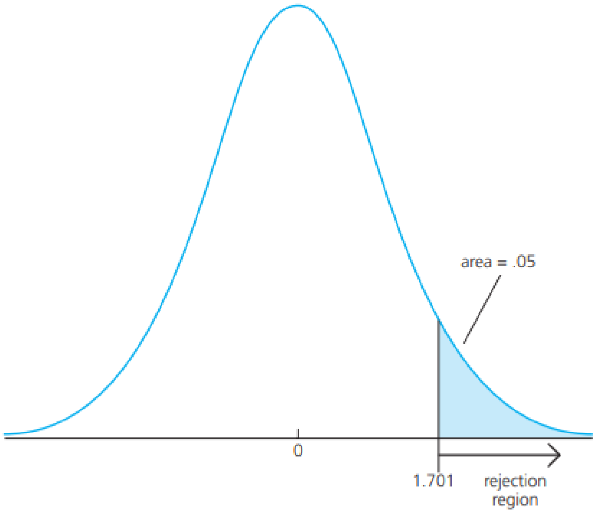
- **One-sided Test:** $H_1 : \beta_j > 0$

Under $\beta_j > 0$, we look for a sufficiently high value of $t_{\hat{\beta}_j}$ to reject H_0 in favor of H_1. (Indeed, a negative value far from 0 does not provide evidence for H_1.)

$$\boxed{\underline{\text{Rejection Zone:}} \quad t_{\hat{\beta}_j} > q_{1-\alpha}}$$

Where q is the critical value from a Student's t-distribution, depending on degrees of freedom (n-k-1, where k is the number of slope parameters). *Alternatively, the quantile of order 1 - α of a standard normal distribution if n is large since a Student's t-distribution can be approximated by a standard normal distribution when n is large.*

Example: *(Source, Wooldridge)* When df=28 and $\alpha = 0.05$, q=1.701.



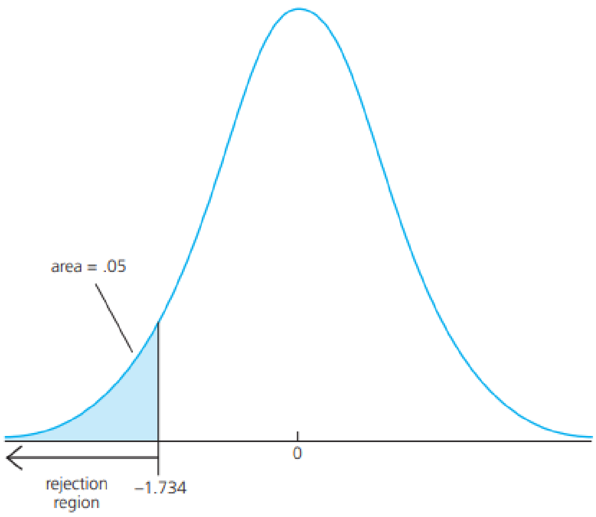
We reject H_0 in favor of H_1 at the 5% level if $t_{\hat{\beta}_j} > 1.701$.
Note: The larger α is, the smaller $q_{1-\alpha}$ (the larger the rejection zone), and thus we reject H_0 more often by mistake. Therefore, if we reject at the 5% level, we also reject at 10%.

- **One-sided Test:** $H_1 : \beta_j < 0$

Similar to the previous hypothesis, except we reject if $t_{\hat{\beta}_j}$ is sufficiently negative.

$$\boxed{\text{Rejection Zone:} \quad t_{\hat{\beta}_j} < -q_{1-\alpha}}$$

Example: *(Source, Wooldridge)* When df=18 and $\alpha = 0.05$, q=1.734.



We reject H_0 in favor of H_1 at the 5% level if $t_{\hat{\beta}_j} < -1.734$.

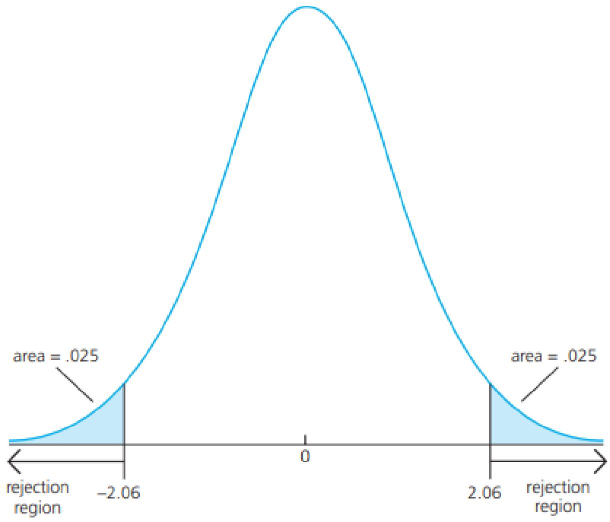
- **Two-sided test:** $H_1 : \beta_j \neq 0$

To reject H_0 in favor of H_1, the value of $t_{\hat{\beta}_j}$ must be far from 0, either positively or negatively.

$$\boxed{\text{Rejection Zone:} \quad |t_{\hat{\beta}_j}| > q_{1-\frac{\alpha}{2}}}$$

where $q_{1-\frac{\alpha}{2}}$ is the quantile at level $1 - \frac{\alpha}{2}$ of a standard normal distribution when doing asymptotic tests. One needs to compare to the two corresponding critical values of a Student distribution when doing inference in finite sample. If $\alpha = 5\%$, then the critical value is found at the 97.5th percentile (1 - $\frac{\alpha}{2}$) of the t-distribution.

Example: *(Source: Wooldridge)* When df=25 and $\alpha = 0.05$, the critical value $q = 2.060$.



We reject H_0 in favor of H_1 if $t_{\hat{\beta}_j} < -2.06$ or if $t_{\hat{\beta}_j} > 2.06$. Note once again that the figure is doing inference on finite sample. The threshold would be at 1.96 for asymptotic tests. If H_0 is rejected at $\alpha = 0.05$, we say that the coefficient is significant at the 5% level.

- **Special case for the null hypothesis** $H_0 : \beta_j = a$, where a is a real number.

The **t-statistic** is written as:

$$\boxed{t_{\hat{\beta}_j} \equiv \frac{\hat{\beta}_j - a}{se(\hat{\beta}_j)}} \tag{5}$$

The rejection zones and test levels remain the same as in previous cases.

2.2 Test on Multiple Parameters

See Chapter 2, slides 24-25

We are now interested in joint hypothesis tests (Fisher's test). For example, we want to test whether a group of variables has an effect on the dependent variable.
Note: The Fisher test is equivalent to the Student t-test in the case where we perform a hypothesis test on only one parameter. However, doing two simple tests does not amount to a joint test. One needs to take into account the correlation between the estimators of a multivariate regression.

We define the **null hypothesis**:

$$H_0 : R\beta = b \tag{6}$$

where R is a matrix of size $r \times k$, where r is the number of equalities to test and k is the number of parameters, and b is a given vector.

Example: Consider the model $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \varepsilon$

- We want to test $\beta_1 = \beta_2$ ($r = 1$), in this case, $R = \begin{pmatrix} 0 & 1 & -1 & 0 & 0 \end{pmatrix}$ and $b = 0$, meaning $R\beta = b \iff \beta_1 - \beta_2 = 0 \iff \beta_1 = \beta_2$.
- We want to test $\beta_1 = \beta_2 = 0$ ($r = 2$), in this case, $R = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}$ and $b = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, meaning $R\beta = b \iff b = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$.

The alternative hypothesis is

$$H_1 : R\beta_0 \neq b \tag{7}$$

We construct the **F-statistic** as follows:

$$\boxed{F = \frac{n(R\hat{\beta} - b)'(R\hat{V}_a R')^{-1}(R\hat{\beta} - b)}{r}} \tag{8}$$

To reject H_0, we define the rejection region as follows:

$$\boxed{\underline{\text{Rejection Region:}} \quad F > \frac{q_{1-\alpha}(r)}{r}}$$

where $q_{1-\alpha}(r)$ is the quantile of order $1-\alpha$ from a $\chi^2(r)$ distribution. The intuition of a χ distribution comes from the multiplication of the two asymptotically normal estimators in equation (8) -and using Cochran theorem-. However, in finite samples, one must compare the F-statistic to the critical value of an F-distribution.

If H_0 is rejected, we say that the variables are "jointly significant."

- **Summary of Interpretation:**
  - The higher the t-statistic, the less likely we are to wrongly reject H_0. Under the null hypothesis of the coefficient, $t_{\hat{\beta}_j} \equiv \frac{\hat{\beta}_j}{se(\hat{\beta}_j)}$, and, all else being equal, if $se(\hat{\beta}_j)$ decreases, the precision of the estimator increases: we are less likely to make a wrong rejection. If $se(\hat{\beta}_j)$ decreases, the t-statistic increases!
  - If we reject H_0 at a 5% significance level ($\alpha = 0.05$), we will also reject the null hypothesis at a 10% significance level, but not necessarily at the 1% level.

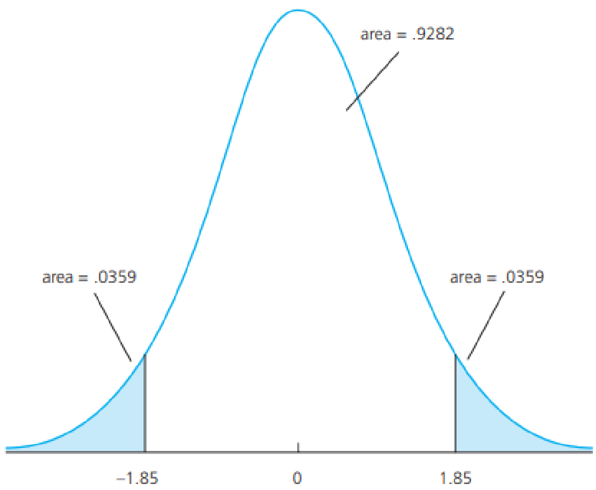
3 p-value

The ***p*-value** is the smallest significance level at which the null hypothesis would be rejected, given the test statistic. The p-value is calculated as the probability of observing a value even more extreme than the test statistic (t-statistic). As a probability, the p-value is between 0 and 1.

In the case of a simple hypothesis and a two-sided test $H_0 : \beta_j = 0$ vs. $H_0 : \beta_j \neq 0$, then the p-value $= P(|T| > |t_{\hat{\beta}_j}|)$, where T follows a Student's t-distribution with $n - k - 1$ degrees of freedom (or a standard normal distribution when n is large).

We reject the null hypothesis at a significance level of $(100 \times \alpha)\%$ **if the *p*-value** $< \alpha$. For example, we reject H_0 at a 5% significance level if the p-value < 0.05.
Note: If we reject H_0 at a 5% level, we also reject H_0 at a 10% level but not necessarily at a 1% level.

<u>Example</u>: *(Source: Wooldridge)* When $df = 40$ and $t = 1.85$.



In this case, the p-value $= P(|T| > t) = P(|T| > 1.85) = 2P(T > 1.85) = 2 \times 0.0359 = 0.0718$. The coefficient is significant at the 10% level but not at the 5% level.

In R, the p-value displayed in regression tables is associated with the hypotheses $H_0 : \beta_j = 0$ vs. $H_0 : \beta_j \neq 0$. If the p-value $< \alpha$, you can say that the coefficient β_j is significant (i.e., significantly different from 0, meaning that the variable x_j has an effect on the dependent variable).

- **Summary of Interpretation:**
  - The smaller the p-value, the less likely we are to reject the null hypothesis incorrectly.
  - In R, the p-value corresponds to a two-sided test for the nullity of the coefficient. We are looking to see if the associated regressor has a causal effect on the dependent variable. If the p-value is below 0.05, we say the coefficient is significant at the 5% level.
  - In practice, in R, we use the p-value more than the t-test to interpret the significance of a coefficient, as the p-value does not require knowledge of a critical value.

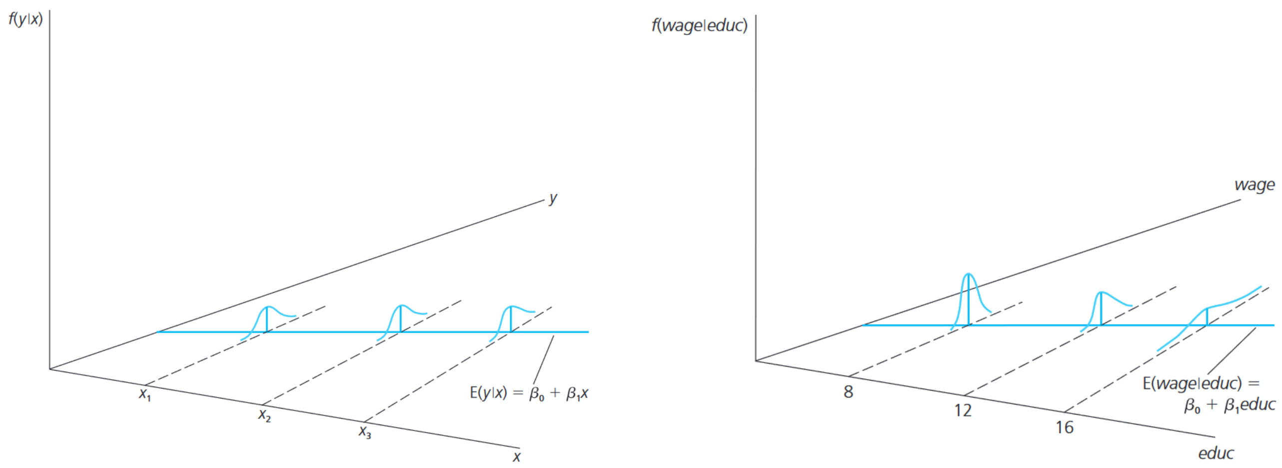
4 Homoskedasticity

Homoskedasticity: The error term ε has the same variance regardless of the value of the regressors X, i.e., $V(\varepsilon|X) = \sigma^2$. (For heteroskedasticity: $V(\varepsilon|X) = \sigma_i^2$). In the lecture, we also use a weaker definition of homoskedasticity: $E(\varepsilon^2 XX') = E(\varepsilon^2)E(XX') = \sigma^2 E(XX')$.
(*For example, the variance of salary conditional on the level of education is not constant: For a low level of education, the salary range is narrow. In contrast, for a higher level of education, the salary range is wider.*) (Q_3)

Figure 1: Homoskedasticity VS Heteroskedasticity

(a) Homoskedasticity

(b) Heteroskedasticity



Source: Wooldridge, "Introductory Econometrics"

Estimation: Using `lm_robust` in R for regression allows for robust estimators against heteroskedasticity of residuals (this means that inference will be valid even if the residuals are heteroskedastic). The `lm` function reports standard errors as if the residuals are homoskedastic. In practice, the variance-covariance matrix is calculated differently. **Adding `lm_robust` does not change the estimated coefficients but changes the standard errors of the coefficients** (standard errors become larger => associated p-values increase => coefficients become less significant).

In short, if we use `lm_robust` and the residuals are homoskedastic, inference is valid, but if the residuals are heteroskedastic, inference is not valid. However, if we use `lm_robust`, inference is valid in both cases. Therefore, it is **recommended to always use `lm_robust`** when unsure if the residuals are homoskedastic.

Alternative Code: In *R*, the `coeftest` function with the `vcov=vcovHC()` option from the `sandwich` package can be used to display regression results obtained from `lm` with corrected standard errors.