Hypothesis Tests & Confidence Intervals

Dr. Patrick Toche

Textbook:

James H. Stock and Mark W. Watson, *Introduction to Econometrics*, 4th Edition, Pearson.

Other references:

Joshua D. Angrist and Jörn-Steffen Pischke, *Mostly Harmless Econometrics: An Empiricist's Companion*, 1st Edition, Princeton University Press.
Jeffrey M. Wooldridge, *Introductory Econometrics: A Modern Approach*, 7th Edition, Cengage Learning.

The textbook comes with online resources and study guides. Other references will be given from time to time.

In this lesson you will learn ...

- to test hypotheses about the population regression coefficients
- standard errors and the regression equation
- two-sided versus one-sided hypotheses
- tests about population slope versus population intercept
- confidence intervals for regression coefficients
- regression with binary independent variables
- heteroskedasticity and homoskedasticity

Recall: Hypothesis Test About Sample Mean

- **Two-Sided Test About μ**

$$H_0: \ \mathsf{E}[Y] = \mu_0$$
$$H_1: \ \mathsf{E}[Y] \neq \mu_0$$

- **Step 1:** Compute SE$(\overline{Y})$
- **Step 2:** Compute t-statistic

$$t_0 = \frac{\overline{Y} - \mu_0}{\text{SE}(\overline{Y})}$$

- **Step 3 | α Variant:** Set significance level α and compute the critical t value for that level:

$$|t_0| > t_{\alpha/2} \implies \text{Reject } H_0$$

- **Step 3 | p Variant:** Compute p-value for a two-sided test

$$p\text{-value} = 2\Phi(-|t_0|) \to \text{small} \implies \text{Reject } H_0$$

The challenge is deciding whether values like p-value $\approx 5\%$ are "small" for your purpose.

Testing Hypotheses About Regression Coefficients

- **Two-Sided Test About β_1**

$$H_0: \beta_1 = \beta_{1,0}$$
$$H_1: \beta_1 \neq \beta_{1,0}$$

- **Step 1:** Compute the standard error

$$\boxed{\text{SE}(\hat{\beta}_1)}$$

- **Step 2:** Compute the t-statistic

$$\boxed{t_0 = \frac{\hat{\beta}_1 - \beta_{1,0}}{\text{SE}(\hat{\beta}_1)}}$$

- **Step 3:** Compute the p-value

$$\boxed{p\text{-value} = 2\Phi(-|t_0|) \to \text{Is } p\text{-value small?}}$$

p-value: Probability of sampling a value $\hat{\beta}_1$ at least as far from $\beta_{1,0}$ as our sample $\hat{\beta}_1$ actually is.

Regression Equations Reporting

- Regression results report the standard errors associated with each coefficient estimate. They are usually reported in parentheses below each coefficient.

$$\widehat{TestScore} = \underset{(10.4)}{698.9} - \underset{(0.52)}{2.28} \times STR, \quad R^2 = 0.051, \quad SER = 18.6$$

The above report is equivalent to:

$$\widehat{TestScore} = \beta_0 + \beta_1 \times STR$$
$$\hat{\beta}_0 = 698.9$$
$$\mathsf{SE}(\hat{\beta}_0) = 10.4$$
$$\hat{\beta}_1 = -2.28$$
$$\mathsf{SE}(\hat{\beta}_1) = 0.52$$

Testing Hypotheses

- A very common desire is to test the significance of the regression coefficients:

$$H_0 : \beta_1 = 0$$
$$H_1 : \beta_1 \neq 0$$

- **Step 1:** Read the standard error from the regression output.
- **Step 2:** Compute the t-statistic under the null:

$$t_0 = \frac{-2.28}{0.52} = -4.38$$

- **Step 3 | α Variant:** Let $\alpha = 0.05$ – a good criterion for the social sciences, not so much for medical research! Compute the critical value or read it from a probability table. Since the sample size is large, we can approximate the Student-t distribution with the standard normal distribution:

$$t_{\alpha/2} \approx 1.96$$

- Because $|t_0 > t_{\alpha/2}|$, we reject the null hypothesis in favor of the two-sided alternative at the $\alpha = 5\%$ significance level.

Critical Values: Standard Normal Distribution — R code

- Compute the critical z-value for a two-sided test:

```
alpha = 0.05
qnorm(1-alpha/2)
## 1.959964
```

- Compute the critical z-value for a one-sided test:

```
alpha = 0.05
qnorm(1-alpha)
## 1.644854
```

- What do these compute?

```
qnorm(0.05)
## -1.644854

qnorm(0.05/2)
## -1.959964
```

Critical Values: Student-t Distribution — R code

- Compute the critical t-value for a two-sided test, with 10 degrees of freedom:

```
alpha = 0.05
qt(1-alpha/2, df=10)
## 2.228139
```

- Compute the critical t-value for a one-sided test, with 10 degrees of freedom:

```
alpha = 0.05
qt(1-alpha, df=10)
## 1.812461
```

- Note how the critical t-value is larger than the critical z-value.

p-Values: Standard Normal Distribution R code

- Compute the p-value for a two-sided test:

```
2 * pnorm(-4.38)
## 1.186793e-05
```

- Compute the p-value for a one-sided test:

```
pnorm(-4.38)
## 5.933965e-06
```

- A p-value smaller than (say) 0.05 suggests that if the null hypothesis is true, then our sample estimate is at least two standard deviations away from the hypothesized mean.
- A p-value smaller than (say) 0.05 provides evidence against the null hypothesis.
- A p-value smaller than (say) 0.0001 provides even stronger evidence against the null hypothesis.
- Beware: This inference is valid if the estimated model satisfies all the conditions needed for inference. If one of these assumptions is violated, the p-value may no longer provide adequate guidance.
- Note how the p-value for the one-sided test is smaller than for the two-sided test.

p-Values: Student-t Distribution R code

- Compute the p-value for a two-sided test, with 10 degrees of freedom:

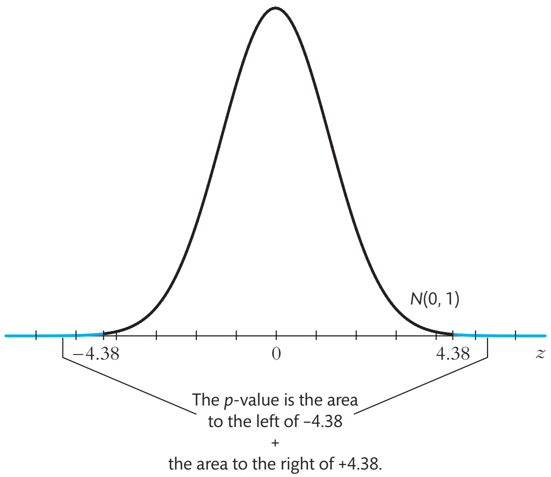
```
2 * pt(-4.38, df=10)
## 0.0013774
```

- Compute the p-value for a one-sided test, with 10 degrees of freedom:

```
pt(-4.38, df=10)
## 0.0006887001
```

- Note how the p-value for the Student-t distribution is larger than for the standard Normal distribution.

Understand the p-Value



The p-value of a two-sided test for $t_0 = -4.38$ is about 0.00001.

Confidence Interval for Slope Coefficient

- **Confidence Interval for β_1:**
  An interval that contains the true value of β_1 with a given probability.

$$\boxed{\hat{\beta}_1 \pm t_{\alpha/2} \times \text{SE}(\beta_1)}$$

- The probability is related to the significance level: $P = 1 - \alpha$.
- Popular values are 90%, 95%, and 99%.
- For $\alpha = 0.05$, the true value of β_1 is contained in 95% of all possible samples.
- Confidence Interval for β_1 in the regression of $TestScore$ on STR:

$$\beta_1 \in (-2.28 \pm 1.96 \times 0.52)$$
$$\implies -3.30 < \beta_1 < -1.26$$

for $t_{0.05} \approx 1.96$.

Confidence Interval for Predicted Change

- **Confidence Interval for predicted effect of a change in X:**

$$\left[\hat{\beta}_1 \pm t_{\alpha/2} \times \text{SE}(\beta_1)\right] \times \Delta X$$

- Confidence Interval for $\beta_1 \Delta X$ in the regression of $TestScore$ on STR, with $t_{0.05} \approx 1.96$:

$$-2.28\Delta X - 1.96 \times 0.52\Delta X < \beta_1 \Delta X < -2.28\Delta X + 1.96 \times 0.52\Delta X$$
$$-3.30\Delta X < \beta_1 \Delta X < -1.26\Delta X$$

Thus, for $\Delta X = -2$, the confidence interval for the predicted change is:

$$6.60 < \beta_1 \Delta X < 2.52$$

Regression when X is a Binary Variable

- **Binary Variable:**
A discrete variable that can take on only two possible values, e.g. 0 and 1.
- Examples: Male Vs Female. Boom Vs Recession. Employed Vs Unemployed. Democrat Vs Republican.
- Also called an indicator variable and/or a dummy variable.
- **Categorical Variable:**
A generization to several states. Example: African, American, Asian, European. Blood types: A, B, AB, O. Vaccination Status: Non-vaccinated, One dose, Two doses, Three doses.
- Also called a dichotomous variable.
- In regression analysis, the presence of categorical variables changes the interpretation of the regression results, but does not change the computation of regression coefficients.

Interpreting Regression Coefficients

- Let STR_i denote the student-teacher ratio in district i. Let $D_i \in \{0, 1\}$ according to:

$$D_i = \begin{cases} 1 \text{ if } STR_i & < 20 \\ 0 \text{ if } STR_i & \geq 20 \end{cases}$$

- The population regression with D_i as the regressor is:

$$Y_i = \beta_0 + \beta_1 D_i + u_i$$

and is equivalent to:

$$Y_i = \beta_0 + u_i \text{ if } D_i = 0$$
$$Y_i = \beta_0 + \beta_1 + u_i \text{ if } D_i = 1$$

which implies $\text{E}[Y_i | D_i = 1] = \beta_0 + \beta_1$.

- Because β_1 is the difference in the population means, the OLS estimator $\hat{\beta}_1$ is the difference between the sample averages of Y_i in the two groups.

Hypothesis Tests & Confidence Intervals

- The null hypothesis that the two population means are the same can be tested against the alternative hypothesis that they differ by testing the null hypothesis $\beta_1 = 0$ against the alternative $\beta_1 \neq 0$.
- Example:In the regression of the test score against the student–teacher ratio binary variable D_i.

$$\underset{}{\widehat{TestScore}} = \underset{(1.3)}{650.0} + \underset{(1.8)}{7.4} \times D, \quad R^2 = 0.037, \quad SER = 18.7$$

- The average test score for the sub-sample with student–teacher ratios greater than or equal to 20 ($D = 0$) is 650.0, and the average test score for the other sub-sample ($D = 1$) is $650.0 + 7.4 = 657.4$.
- The difference between the sample average test scores for the two groups is 7.4.
- Is the difference in the population mean test scores in the two groups statistically significantly different from 0 at the 5% level?

$$t = 7.4/1.8 = 4.04 > 1.96$$

The null can be rejected at the 5% level.

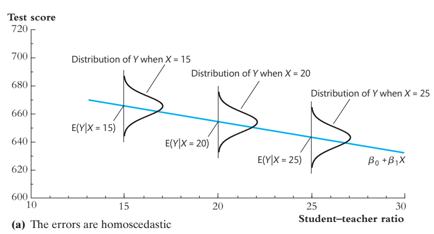
Homoskedasticity and Heteroskedasticity

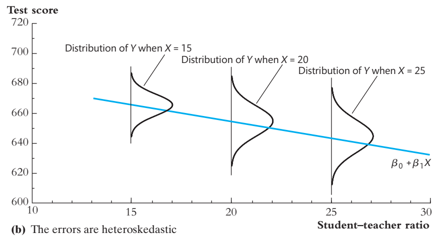
- The error term u_i is homoskedastic if the variance of the conditional distribution of u_i given X_i is constant for $i = 1, \ldots, n$ and in particular does not depend on X_i. Otherwise, the error term is heteroskedastic.
- Whether the errors are homoskedastic or heteroskedastic, the OLS estimator is unbiased, consistent, and asymptotically normal.
- Economic theory rarely gives any reason to believe that the errors are homoskedastic — It is prudent to assume that the errors might be heteroskedastic. Many software programs report homoskedasticity- only standard errors as their default setting.
- If the error term is homoskedastic, the formulas for the variances of $\hat{\beta}_0$ and $\hat{\beta}_1$ simplify:

$$\sigma^2_{\hat{\beta}_0} = \frac{\frac{1}{n} \cdot \sigma_u^2 \cdot \frac{1}{n} \sum_{i=1}^n X_i^2}{\frac{1}{n} \sum_{i=1}^n (X_i - \overline{X})^2}$$

$$\sigma^2_{\hat{\beta}_1} = \frac{\frac{1}{n} \sigma_u^2}{\frac{1}{n} \sum_{i=1}^n (X_i - \overline{X})^2}$$

Homoskedasticity and Heteroskedasticity





The spread of these distributions does not depend on x.

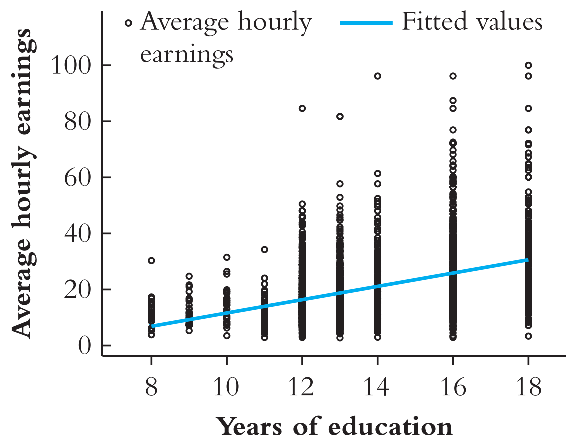
These become more spread out for larger class sizes.

Heteroskedasticity: Application

- Workers with more education have higher earnings than workers with less education. On average, hourly earnings increase by $\$2.37$ for each additional year of education.
- The spread of the distribution of earnings increases with the years of education. While some workers with many years of education have low-paying jobs, very few workers with low levels of education have high-paying jobs. For workers with 10 years of education, the standard deviation of the residuals is $\$6.31$; for workers with a high school diploma, it is $\$8.54$; and for workers with a college degree, $\$13.55$.
- Not all college graduates will be earning $\$75$ per hour by age 29, but some will, but workers with only 10 years of education have no shot at those jobs.

$$\widehat{Earnings} = \underset{(1.36)}{-12.12} + \underset{(0.10)}{2.37} \times Education, \quad R^2 = 0.185, \quad SER = 11.24$$

Heteroskedasticity: Application



Hourly Earnings and Years of Education for 2731 full-time 29- to 30-year-old workers in the United States, 2015.

The Gauss–Markov Theorem

- **BLUE:** Under the "Gauss–Markov conditions", the OLS estimator $\hat{\beta}_1$ has the smallest conditional variance given $X_1, \ldots, X_n$ — is the Best — of all Linear conditionally Unbiased Estimators of β_1.
- Drawbacks:
  1. If the error term is heteroskedastic, OLS is no longer the efficient linear conditionally unbiased estimator.
  2. There are other candidate estimators that are not linear and conditionally unbiased.
- **Weighted Least Squares (WLS):** If the conditional variance of u_i given X_i is known up to a constant factor of proportionality, then it is possible to construct an estimator that has a smaller variance than the OLS estimator.
- WLS: weighs the ith observation by the inverse of the square root of the conditional variance of ui given X_i.
- The practical problem with weighted least squares is that you must know how.
- **Least Absolute Deviations (LAD):** The LAD estimator is robust to large outliers.

Summary

- Hypothesis testing for regression coefficients is analogous to hypothesis testing for the population mean: Use the t-statistic to calculate the p-values and either accept or reject the null hypothesis.
- Like a confidence interval for the population mean, a 95% confidence interval for a regression coefficient is computed as the estimator $\pm 1.96\times$ standard errors.
- When X is binary, the regression model can be used to estimate and test hypotheses about the difference between the population means of the "$X = 0$" group and the "$X = 1$" group.
- In general, the error u_i is heteroskedastic; that is, the variance of u_i at a given value of X, var$(u_i|X_i = x)$, depends on x.
- The error is homoskedastic if var $u_i|X_i = x$ is constant. Heteroskedasticity-robust standard errors produce valid statistical inference.
- If the three least squares assumption hold *and* if the regression errors are homoskedastic, then the Gauss–Markov theorem implies that the OLS estimator is BLUE.
- If the three least squares assumptions hold, if the regression errors are homoskedastic, and if the regression errors are normally distributed, then the OLS t-statistic computed using homoskedasticity-only standard errors has a Student-t distribution under the null.

Problems and Applications

Stock & Watson, Introduction (4th), Chapter 5, Review Question 1.

Outline the procedures for computing the p-value of a two-sided test of $H0\colon\ \mu_Y = 0$ using an i.i.d. set of observations $Yi, i = 1, \ldots, n$. Outline the regression model using an i.i.d. set of observations $Y_i, X_i, i = 1, \ldots, n$.

Stock & Watson, Introduction (4th), Chapter 5, Review Question 3.

Define homoskedasticity and heteroskedasticity. Provide a hypothetical empirical example in which you think the errors would be heteroskedastic, and explain your reasoning.

Wooldridge, Introduction (7th), Chapter 8, Problem 1.

Which of the following are consequences of heteroskedasticity?

1. The OLS estimators, $\hat{\beta}_i$, are inconsistent.
2. The OLS estimators are no longer BLUE.

Problems and Applications

Stock & Watson, Introduction (4th), Chapter 5, Exercise 1.

A researcher, using data on class size (CS) and average test scores from 100 third-grade classes, estimates the OLS regression:

$$\widehat{TestScore} = \underset{(20.4)}{520.4} - \underset{(2.21)}{5.82} \times CS, \quad R^2 = 0.08, \quad SER = 11.5$$

1. Construct a 95% confidence interval for β_1, the regression slope coefficient.
2. Calculate the p-value for the two-sided test of the null hypothesis $H_0\colon \beta_1 = 0$. Do you reject the null hypothesis at the 5% level? At the 1% level?
3. Calculate the p-value for the two-sided test of the null hypothesis $H_0\colon \beta_1 = -5.6$. Without doing any additional calculations, determine whether -5.6 is contained in the 95% confidence interval for β_1.
4. Construct a 99% confidence interval for β_0.

Problems and Applications

Stock & Watson, Introduction (4th), Chapter 5, Exercise 3.

Suppose a random sample of 200 20-year-old men is selected from a population and their heights and weights are recorded. A regression of weight on height yields

$$\widehat{Weight} = \underset{(2.15)}{-99.41} + \underset{(0.31)}{3.94} \times Height, \quad R^2 = 0.81, \quad SER = 10.2$$

where $Weight$ is measured in pounds and $Height$ is measured in inches. Two of your classmates differ in height by 1.5 inches. Construct a 99% confidence interval for the difference in their weights.

Keywords

null hypothesis two-sided alternative hypothesis standard error of regression slope t-statistic p-value confidence interval for regression slope confidence level indicator variable dummy variable homoskedasticity heteroskedasticity robust standard error Gauss–Markov theorem best linear unbiased estimator (BLUE) weighted least squares (WLS)

Problems and Applications

Stock & Watson, Introduction (4th), Chapter 5, Exercise 5.

In the 1980s, Tennessee conducted an experiment in which kindergarten students were randomly assigned to "regular" and "small" classes and given standardized tests at the end of the year. (Regular classes contained approximately 24 students, and small classes contained approximately 15 students.) Suppose, in the population, the standardized tests have a mean score of 925 points and a standard deviation of 75 points. Let $SmallClass$ denote a binary variable equal to 1 if the student is assigned to a small class and equal to 0 otherwise. A regression of $TestScore$ on $SmallClass$ yields

$$\widehat{TestScore} = \underset{(1.6)}{-918.0} + \underset{(2.5)}{13.9} \times SmallClass, \quad R^2 = 0.01, \quad SER = 74.6$$

1. Do small classes improve test scores? By how much? Is the effect large? Explain.
2. Is the estimated effect of class size on test scores statistically significant? Carry out a test at the 5% level.
3. Construct a 99% confidence interval for the effect of $SmallClass$ on $TestScore$.
4. Does least squares assumption 1 plausibly hold for this regression? Explain.