Course: Mathematical statistics

Week 14: Test of hypothesis as regression coefficient

Lecturer: Nagulama Moses

Kumi University

June 3, 2025

Outline

What is Hypothesis Testing in Regression?

In regression analysis, we are often interested in making inferences about the model parameters, specifically the slope and intercept.

This includes hypothesis testing and constructing confidence intervals to determine whether these parameters significantly differ from specified values, such as zero.

To carry out these statistical procedures, we must make an important assumption

Therefore, the complete set of assumptions for the classical linear regression model includes:

- The errors ε_i are normally and independently distributed,
- Each with mean 0 and constant variance σ^2.

Why is it Important?

- Helps us identify meaningful predictors.
- Ensures that models are not over fitted with non-informative variables.
- Supports decision-making in predictive analytics.

Hypothesis Testing on the Slope

Objective: Test whether the slope β_1 equals a specified constant β_{10}.

Hypotheses:

$$H_0 : \beta_1 = \beta_{10} \quad \text{vs.} \quad H_1 : \beta_1 \neq \beta_{10}$$

Assumptions:

- Errors $\varepsilon_i \sim$ (Normally and Independently Distributed)NID$(0, \sigma^2)$
- Observations $y_i \sim \text{NID}(\beta_0 + \beta_1 x_i, \sigma^2)$

Distribution of the Estimator

- $\hat{\beta}_1$ is normally distributed:

$$\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{S_{xx}}\right)$$

- If σ^2 were known, the test statistic would be:

$$Z_0 = \frac{\hat{\beta}_1 - \beta_{10}}{\sqrt{\sigma^2/S_{xx}}} \sim N(0,1)$$

- But σ^2 is unknown and we estimate using:

$$\text{MS}_{\text{Res}(MeanSquareResidual)} = \frac{1}{n-2}\sum(y_i - \hat{y}_i)^2$$

t-Test Statistic for the Slope

t-Statistic

$$t_0 = \frac{\hat{\beta}_1 - \beta_{10}}{\sqrt{\mathrm{MS}_{\mathrm{Res}}/S_{xx}}}$$

Distribution

$$t_0 \sim t_{n-2} \quad \text{under } H_0$$

Decision Rule

$$\text{Reject } H_0 \text{ if } |t_0| > t_{\alpha/2,\, n-2}$$

Alternative; Use P-value: reject if P-value $< \alpha$

Testing the Intercept

Hypotheses

$$H_0 : \beta_0 = \beta_{00} \quad \text{vs.} \quad H_1 : \beta_0 \neq \beta_{00}$$

t-Statistic

$$t_0 = \frac{\hat{\beta}_0 - \beta_{00}}{\text{se}(\hat{\beta}_0)}$$

Standard Error

$$\text{se}(\hat{\beta}_0) = \sqrt{\text{MS}_{\text{Res}} \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right)}$$

Decision Rule

$$\text{Reject } H_0 \text{ if } |t_0| > t_{\alpha/2,\, n-2}$$

Therefore;

- Hypothesis tests for slope/intercept help assess model significance.
- Use t-statistics when σ^2 is unknown.
- Always check assumptions: normality, independence, and constant variance.
- Confidence intervals can be constructed similarly.

Example 1: Testing the Slope

A researcher fits a simple linear regression model to study the relationship between hours studied (x) and exam score (y). The estimated slope is $\hat{\beta}_1 = 2.5$ with a standard error of 0.8. Assume errors are normally distributed and sample size $n = 20$.

Test the hypothesis:

$$H_0 : \beta_1 = 0 \quad \text{vs.} \quad H_1 : \beta_1 \neq 0$$

at the 5% significance level.

Calculate the test statistic

$$t_0 = \frac{\hat{\beta}_1 - \beta_{10}}{\text{se}(\hat{\beta}_1)} = \frac{2.5 - 0}{0.8} = 3.125$$

Degrees of freedom:

$df = n - 2 = 20 - 2 = 18.$

Find the critical value

From t-distribution table, for $\alpha = 0.05$ (two-tailed), $t_{0.025,18} \approx 2.101$.

Decision
Since $|t_0| = 3.125 > 2.101$, we reject H_0.

Conclusion: There is sufficient evidence at the 5% level to conclude the slope is significantly different from zero.

Example 2: Testing the Intercept

A linear regression model estimates the intercept $\hat{\beta}_0 = 10$ with a standard error of 3. The sample size is $n = 15$.

Test whether the intercept differs significantly from zero at the 1% significance level.

1. Calculate the test statistic.
2. State your conclusion.

Solution

Calculate the test statistic

$$t_0 = \frac{\hat{\beta}_0 - 0}{\text{se}(\hat{\beta}_0)} = \frac{10 - 0}{3} = 3.33$$

Degrees of freedom: $df = n - 2 = 15 - 2 = 13$.

Find the critical value

From t-distribution table, for $\alpha = 0.01$ (two-tailed), $t_{0.005,13} \approx 3.012$.

Decision

Since $|t_0| = 3.33 > 3.012$, we reject H_0.

Conclusion: There is sufficient evidence at the 1% level to conclude that the intercept differs significantly from zero.

Testing Significance of Regression

One of the most important special cases of hypothesis testing in simple linear regression concerns whether the predictor variable x has any linear relationship with the response variable y. This corresponds to testing the

null hypothesis:

$$H_0 : \beta_1 = 0 \quad \text{versus} \quad H_1 : \beta_1 \neq 0$$

This test is commonly referred to as **testing the significance of regression**.

Interpretation of the Hypotheses

Failing to reject $H_0 : \beta_1 = 0$ means there is no evidence of a linear relationship between x and y. This situation may arise because:

- x has little or no explanatory power for the variation in y. In this case, the best prediction for y is simply the mean $\hat{y} = \bar{y}$.

- Alternatively, the relationship between x and y may not be linear; there could be a non-linear association the simple linear model fails to capture.

Thus, failing to reject H_0 is equivalent to stating there is no **linear** association between x and y.

Interpretation of the Hypotheses (cont.)

Rejecting $H_0 : \beta_1 = 0$ implies that x significantly explains the variability in y. This means:

- There is evidence of a linear effect of x on y, supporting the adequacy of the simple linear regression model.
- However, this does not rule out that a more complex model, such as one including polynomial terms or other transformations of x, might better fit the data.

The Test Procedure

The significance test for $H_0 : \beta_1 = 0$ can be conducted using the t-statistic:

$$t_0 = \frac{\hat{\beta}_1 - 0}{\mathrm{se}(\hat{\beta}_1)} = \frac{\hat{\beta}_1}{\mathrm{se}(\hat{\beta}_1)}$$

where $\hat{\beta}_1$ is the estimated slope, and $\mathrm{se}(\hat{\beta}_1)$ is the standard error of $\hat{\beta}_1$.
The test decision rule at significance level α is:

$$\text{Reject } H_0 \quad \text{if} \quad |t_0| > t_{\alpha/2,\, n-2}$$

where $t_{\alpha/2,\, n-2}$ is the critical value from the Student's t-distribution with $n-2$ degrees of freedom.

Example 1: Testing the Slope

A researcher fits a simple linear regression model relating hours studied (x) to exam score (y). The estimated slope is $\hat{\beta}_1 = 2.5$ with a standard error of 0.8. The sample size is $n = 20$.
Test the hypothesis:

$$H_0 : \beta_1 = 0 \quad \text{vs} \quad H_1 : \beta_1 \neq 0$$

at the 5% significance level.

Solution 1

Calculate the test statistic

$$t_0 = \frac{\hat{\beta}_1 - 0}{\text{se}(\hat{\beta}_1)} = \frac{2.5}{0.8} = 3.125$$

Find critical value Degrees of freedom: $n - 2 = 18$

At $\alpha = 0.05$, two-tailed test, critical value $t_{0.025,18} \approx 2.101$.

Decision

$$|t_0| = 3.125 > 2.101 \implies \text{Reject } H_0$$

There is sufficient evidence at the 5% level to conclude that hours studied has a significant linear effect on exam score.

Example 2: Testing the Intercept

In a linear regression, the estimated intercept is $\hat{\beta}_0 = 10$ with standard error 3. The sample size is $n = 15$.

Test:

$$H_0 : \beta_0 = 0 \quad \text{vs} \quad H_1 : \beta_0 \neq 0$$

at the 1% significance level.

Solution

Calculate test statistic

$$t_0 = \frac{\hat{\beta}_0 - 0}{\text{se}(\hat{\beta}_0)} = \frac{10}{3} = 3.33$$

Degrees of freedom $n - 2 = 13$

Critical value at $\alpha = 0.01$ **two-tailed** $t_{0.005,13} \approx 3.012$

Step 4: Decision

$$|t_0| = 3.33 > 3.012 \implies \text{Reject } H_0$$

The intercept is significantly different from zero at the 1% level.

Example 3

Suppose a regression analysis estimates $\hat{\beta}_1 = 0.4$ with $\text{se}(\hat{\beta}_1) = 0.5$, and $n = 25$.

Test:

$$H_0 : \beta_1 = 0 \quad \text{vs} \quad H_1 : \beta_1 \neq 0$$

at $\alpha = 0.05$.

Solution

Test statistic

$$t_0 = \frac{0.4}{0.5} = 0.8$$

Degrees of freedom $25 - 2 = 23$
Critical value $t_{0.025,23} \approx 2.069$

$$|0.8| < 2.069 \implies \text{Fail to reject } H_0$$

There is no significant evidence at the 5% level to conclude a linear relationship between x and y.

Example 4: Two-sided Test with Known Critical Value

A regression slope is estimated as $\hat{\beta}_1 = -1.2$, with standard error 0.4 and $n = 18$.
Test $H_0 : \beta_1 = 0$ against $H_1 : \beta_1 \neq 0$ at $\alpha = 0.05$.
Given $t_{0.025,16} = 2.12$, determine the test statistic and conclusion.

Solution

Compute test statistic

$$t_0 = \frac{-1.2}{0.4} = -3.0$$

Compare to critical value

$$|t_0| = 3.0 > 2.12$$

Reject H_0. There is sufficient evidence at the 5% level that the slope is different from zero, indicating a significant linear relationship between x and y.

Example: Simple Linear Regression Test

Given the following data:

- X (Hours studied): 1, 2, 3, 4, 5
- Y (Test score): 2, 4, 5, 4, 5

1. Compute the estimated regression coefficient $\hat{\beta}_1$
2. Test the hypothesis:

$$H_0 : \beta_1 = 0 \quad \text{vs} \quad H_1 : \beta_1 \neq 0$$

using a 5% significance level.

Solution: Compute Estimates

Let the regression model be $Y = \beta_0 + \beta_1 X + \varepsilon$

Given:

$$\bar{X} = \frac{1+2+3+4+5}{5} = 3, \quad \bar{Y} = \frac{2+4+5+4+5}{5} = 4$$

Compute $\hat{\beta}_1$:

$$\hat{\beta}_1 = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(X_i - \bar{X})^2}$$

$$\Rightarrow \hat{\beta}_1 = \frac{(1-3)(2-4) + (2-3)(4-4) + (3-3)(5-4) + (4-3)(4-4) +}{(1-3)^2 + (2-3)^2 + (3-3)^2 + (4-3)^2 + (5-3)^2}$$

$$\Rightarrow \hat{\beta}_1 = \frac{(-2)(-2) + (-1)(0) + (0)(1) + (1)(0) + (2)(1)}{4 + 1 + 0 + 1 + 4}$$

$$= \frac{4 + 0 + 0 + 0 + 2}{10} = \frac{6}{10} = 0.6$$

Compute Intercept and Residuals

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} = 4 - 0.6(3) = 4 - 1.8 = 2.2$$

The estimated regression equation is:

$$\hat{Y} = 2.2 + 0.6X$$

Now compute the residuals and residual sum of squares (RSS):

$$\text{RSS} = \sum (Y_i - \hat{Y}_i)^2$$

Predicted $\hat{Y}_i$ values:

$$\hat{Y} = [2.8,\ 3.4,\ 4.0,\ 4.6,\ 5.2]$$

Actual

$$(Y = [2,\ 4,\ 5,\ 4,\ 5]) \Rightarrow \text{Residuals} = [-0.8,\ 0.6,\ 1.0,\ -0.6,\ -0.2]$$

$$\text{RSS} = (-0.8)^2 + (0.6)^2 + (1)^2 + (-0.6)^2 + (-0.2)^2$$
$$= 0.64 + 0.36 + 1 + 0.36 + 0.04 = 2.4$$

Standard error of $\hat{\beta}_1$: $\quad SE(\hat{\beta}_1) = \sqrt{\frac{RSS}{n-2}} \div \sqrt{\sum(X_i - \bar{X})^2}$

$$= \sqrt{\frac{2.4}{3}} \div \sqrt{10} = \sqrt{0.8}/\sqrt{10} \approx 0.894/3.16 \approx 0.283$$

Hypothesis Test

$$t_0 = \frac{\hat{\beta}_1 - 0}{SE(\hat{\beta}_1)} = \frac{0.6}{0.283} \approx 2.12$$

Degrees of freedom $= n - 2 = 5 - 2 = 3$
From t-distribution table, for $\alpha = 0.05$, two-tailed test:

$$t_{0.025,3} \approx 3.182$$

Decision:

$$|t_0| = 2.12 < 3.182 \Rightarrow \text{Fail to reject } H_0$$

There is insufficient evidence at the 5% level to conclude that hours studied has a significant linear effect on test score.

References

- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to Linear Regression Analysis*.
- Kutner, M. H., Nachtsheim, C. J., & Neter, J. (2004). *Applied Linear Regression Models*.

Thank You!