

IES302 2011/2 Part II.5 Dr.Prapun

Example 14.9. Evaluation of Teaching Techniques: See Figure 24.

ABSTRACT: THIS STUDY TESTS THE EFFECT OF HOMEWORK COLLECTION AND QUIZZES ON EXAM SCORES.

The hypothesis for this study is that an instructor can improve a student's performance (exam scores) through influencing the student's perceived effort-reward probability. An instructor accomplishes this by assigning tasks (teaching techniques) which are a part of a student's grade and are perceived by the student as a means of improving his or her grade in the

class. The student is motivated to increase effort to complete those tasks which should also improve understanding of course material. The expected final result is improved exam scores. The null hypothesis for this study is:

H_0 : Teaching techniques have no significant effect on students' exam scores. . . .

Source: "Evaluation of Teaching Techniques" by David R. Vruwink and Janon R. Otto, published in *The Accounting Review*, Vol. LXII, No. 2, April 1987. Reprinted by permission.

Figure 24: Evaluation of Teaching Techniques

14.10. Strong vs. Weak Conclusions

- (a) Since the analyst can directly control the probability of wrongly rejecting H_0 , we always think of rejection of the null hypothesis H_0 as a **strong** conclusion.
- (b) It is customary to think of the decision to accept H_0 as a **weak** conclusion, unless we know that β is acceptably small. Therefore, rather than saying we "accept H_0 ", we prefer the terminology "fail to reject H_0 ."

14.11. Interpretation of “failing to reject H_0 ”:

- (a) We have not found “sufficiently strong” evidence “to reject H_0 ” (or “in support of H_1 ” or “to make a strong statement”) at the level of significance being used for the test.
- (b) Does not necessarily mean that H_0 is true.
- (c) It may simply mean that more data are required to reach a strong conclusion.

Example 14.12. Useful analog between hypothesis testing and a jury trial: In a trial the defendant is assumed innocent (this is like assuming the null hypothesis to be true). If strong evidence is found to the contrary, the defendant is declared to be guilty (we reject the null hypothesis). If there is insufficient evidence the defendant is declared to be not guilty. This is not the same as proving the defendant innocent and so, like failing to reject the null hypothesis, it is a weak conclusion.

14.13. Reporting the results of a hypothesis test

- (a) ***Fixed significance level testing:*** State that the null hypothesis was or was not rejected at a specified α -value.
 - For example, we can say that $H_0 : \mu = 50$ was rejected at the 0.05 level of significance.
 - May be inadequate because
 - (i) It gives the decision maker no idea about whether the computed value of the test statistic was just barely in the rejection region or whether it was very far into this region.
 - (ii) It imposes the predefined α on other users of the information. Some decision makers might be uncomfortable with the risks implied by the chosen α .
- (b) ***P-value (probability-value) approach***
 - Adopted widely in practice.

- Determine the exact level of significance associated with the calculated value of the test statistic.
- Gained popularity in recent years, largely as a result of the convenience and the “number-crunching” ability of the computer.

Definition 14.14. The *P-value* is the smallest α that would lead to rejection of H_0 with the given data.

- For a given set of data, the *P-value* is sometimes referred to as the *observed* level of significance.

14.15. Connection between Hypothesis Tests and Confidence Intervals: If $[\ell, u]$ is a confidence interval for the parameter μ , the test of size α of the hypothesis

$$\begin{aligned} H_0 : \mu &= \mu_0 \\ H_1 : \mu &\neq \mu_0 \end{aligned}$$

will lead to rejection of H_0 if and only if μ_0 is **not** in the $100(1-\alpha)\%$ CI $[\ell, u]$.

Although hypothesis tests and CIs are equivalent procedures insofar as decision making or inference about μ is concerned, each provides somewhat different insights. For instance, the confidence interval provides a range of likely values for μ at a stated confidence level, whereas hypothesis testing is an easy framework for displaying the *risk levels* such as the *P-value* associated with a specific decision.

14.1 Two-Tail Testing of a Mean, Population Variance Known

Suppose that we wish to test the hypotheses

$$\begin{aligned} H_0 : \mu &= \mu_0 \\ H_1 : \mu &\neq \mu_0 \end{aligned}$$

where μ_0 is a specified constant. We have a random sample X_1, X_2, \dots, X_n from a normal population (or non-normal population but $n \geq 30$).

Recall that

Situations can occur where the population mean is unknown but past experience has provided us with a trustworthy value for the population standard deviation. Although this possibility is more likely in an industrial production setting, it can sometimes apply to employees, consumers, or other nonmechanical entities.

14.16. It is usually more convenient to standardize the sample mean and use a test statistic based on the standard normal distribution. That is, the test procedure for $H_0 : \mu = \mu_0$ uses the *test statistic*²⁴, *z-test* :

The value of the calculated test statistic is used in conjunction with a *decision rule* to determine either “reject H_0 ” or “fail to reject H_0 ”. This decision rule must be established prior to collecting the data; it specifies how you will reach the decision.

14.17. To complete a hypothesis test, you will need to write a conclusion that carefully describes the meaning of the decision relative to the intent of the hypothesis test.

- (a) If the decision is “reject H_0 ”, then the conclusion should be worded something like, “There is sufficient evidence at the α level of significance to show that ... [the meaning of the alternative hypothesis].”
- (b) If the decision is “fail to reject H_0 ”, then the conclusion should be worded something like, “There is not sufficient evidence at the α level of significance to show that . . . [the meaning of the alternative hypothesis].”

²⁴Test statistic is a random variable whose value is calculated from the sample data and is used in making the decision “reject H_0 ” or “fail to reject H_0 ”.

Example 14.18. When a robot welder is in adjustment, its mean time to perform its task is 1.3250 minutes. Past experience has found the standard deviation of the cycle time to be 0.0396 minutes. An incorrect mean operating time can disrupt the efficiency of other activities along the production line. For a recent random sample of 80 jobs, the mean cycle time for the welder was 1.3229 minutes. Does the machine appear to be in need of adjustment?

- (a) Formulate the Null and Alternative Hypotheses:
- (b) Select the Significance Level: The significance level used will be $\alpha = 0.05$.
If the machine is running properly, there is only a 0.05 probability of our making the mistake of concluding that it requires adjustment.
- (c) Identify Critical Values for the Test Statistic and State the Decision Rule: The population standard deviation (σ) is known and the sample size is large, so the normal distribution is appropriate and the test statistic will be Z_0 , calculated as
- (d) Identify Critical Values for the Test Statistic and State the Decision Rule:

The decision rule can be stated as “Reject H_0 if calculated $Z_0 < -1.96$ or $> +1.96$, otherwise do not reject.”

- (e) Compare Calculated and Critical Values and Reach a Conclusion for the Null Hypothesis: The calculated value, $z_0 = -0.47$, falls within the non-rejection region. Therefore, At the 0.05 level of significance, H_0 cannot be rejected.
- (f) Make the Related Business Decision: Based on these results, the robot welder is not in need of adjustment. The difference between the hypothesized population mean, $\mu_0 = 1.3250$ minutes, and the observed sample mean, $\bar{x} = 1.3229$, is judged to have been merely the result of chance variation.
- If we had used the sample to construct a 95% confidence interval for μ , the interval would have been from 1.3142 to 1.3316 minutes.

Notice that the hypothesized value, $\mu_0 = 1.3250$ minutes, falls within the 95% confidence interval that is, the confidence interval tells us that μ could be 1.3250 minutes.

14.19. To test the null hypothesis using the P -value approach, we first identify the most extreme critical value that the test statistic would be capable of exceeding. This is equivalent to your jumping as high as you can with no bar in place, then having the judges tell you how high you would have cleared if there had been a crossbar.

Then, we find the value of α corresponding to the extreme critical value. This is the same as the probability of observing a value of the sample mean \bar{X} that is at least as extreme as \bar{x} , given that H_0 is true.

14.20. For the two-sided H_1 , the P -value is

$$P = 2(1 - \Phi(|z_0|)).$$

Example 14.21. Continue from Example 14.18, we have $z_0 = -0.47$

14.22. To make a strong conclusion (rejection of H_0), the P -value should be “small”. See Figure 25.

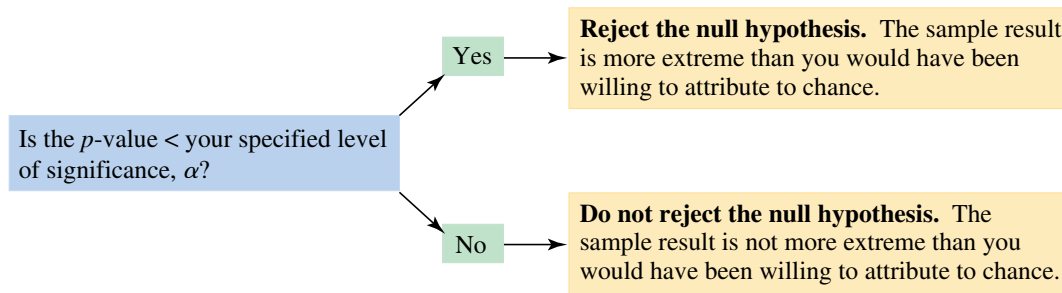


Figure 25: Interpreting the P -value

14.2 Two-Tail Testing of a Mean, Population Variance Unknown

The true standard deviation of a population will usually be unknown. In which case, we cannot directly talk about

$$Z_0 = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}.$$

As in Section 13.2, when σ is unknown, a logical procedure is to replace σ with the sample standard deviation S . The random

variable Z_0 now becomes

$$T_0 = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}.$$

14.23. Test statistic, t -test for a sample mean:

$$T_0 = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}.$$

If H_0 is true, T_0 has a t distribution with $n - 1$ degrees of freedom. When we know the distribution of the test statistic when H_0 is true (this is often called the **reference distribution** or the **null distribution**), we can calculate the P -value from this distribution, or, if we use a fixed significance level approach, we can locate the critical region to control the type I error probability at the desired level.

14.24. Statistics software packages calculate and display P -values. However, in working problems by hand, it is useful to be able to find the P -value for a t -test. Because the t -table in Figure 23 contains only 10 critical values for each t distribution, determining the exact P -value from this table is usually impossible. However, we can use the the t -table in Figure 23 to find lower and upper bounds on the P -value.

14.3 Type I and Type II Errors: A Revisit

14.25. The statistician's job is thus to "balance" the three values of α , β , and n to achieve an acceptable testing situation.

14.26. α vs. β : Type I and type II errors are related. A decrease in the probability of one type of error results in an increase in the probability of the other, provided that the sample size n does not change.

Definition 14.27. The **power** of a statistical test is the probability of *correctly* rejecting H_0 when H_1 is true.

$$= 1 - \beta$$

= probability of rejecting a false H_0 .

- Depend on the true value of the population mean μ , a quantity that we do not know. So, we calculate it at each possible value of the true mean μ .
 - The **power curve** of the test is the plot of μ vs. $1 - \beta$.
 - The **operating characteristic (OC) curve** is the plot of μ vs. β .

14.28. Power curve construction: Assume different population mean values at which H_0 would be false, then determine the probability that an observed sample mean would fall into a rejection region originally specified by the decision rule of the test.

14.29. In two-tail tests, the power curve $1 - \beta$ will have a zero value when the assumed population mean μ equals the hypothesized value μ_0 , then will increase toward 1.0 in both directions from that assumed value for the mean.

- In appearance, it will somewhat resemble an upside-down normal curve.
- On the other hand, the OC curve (β) increases as the (assumed) true value of the parameter (μ) approaches the value μ_0 hypothesized in H_0 . The value of β decreases as the difference between the assumed true mean μ and the hypothesized value μ_0 increases.

14.30. For a fixed decision rule (same rejection and non-rejection regions), we can decrease both α and β by using a larger sample size n .

14.31. If a test is carried out at a specified significance level (e.g., $\alpha = 0.05$), using a larger sample size will change the decision rule but will not change α . This is because α has been decided upon in advance. However, in this situation the larger sample size will reduce the value of β .

Summary: an increase in sample size n reduces β , provided that α is held constant.

15 Simple Linear Regression

Many problems in engineering and the sciences involve a study or analysis of the relationship between two or more variables. There are many situations where the relationship between variables is not deterministic.

Definition 15.1. The collection of statistical tools that are used to model and explore relationships between variables that are related in a nondeterministic manner is called *regression*²⁵ *analysis*.

In this section we present the *simple* situation where there is only one independent or *regressor predictor variable* x and the relationship with the *response variable* y is assumed to be linear.

Definition 15.2. *Simple linear regression model:*

$$Y = \beta_0 + \beta_1 x + \varepsilon.$$

- (a) ε is a random error with mean zero and (unknown) variance σ^2 . The random errors corresponding to different observations are also assumed to be uncorrelated random variables.
- (b) The mean of the random variable Y is related to x by the following straight-line relationship:

$$\mathbb{E}[Y|x] = \mu_{Y|x} = \beta_0 + \beta_1 x.$$

- The slope and intercept of the line are called *regression coefficients*.
- (c) For a fixed value of x the actual value of Y is determined by the mean value function (the linear model) plus a random error term ε .

²⁵Historical Note: Sir Francis Galton first used the term regression analysis in a study of the heights of fathers (x) and sons (y). Galton fit a least squares line and used it to predict the son's height from the father's height. He found that if a father's height was above average, the son's height would also be above average, but not by as much as the father's height was. A similar effect was observed for below average heights. That is, the son's height "regressed" toward the average. Consequently, Galton referred to the least squares line as a *regression line*.

(d) The variance of Y given x is

15.3. In most real-world problems, the values of the intercept and slope (β_0, β_1) and the error variance σ^2 will not be known, and they must be estimated from sample data. Then this fitted regression equation or model is typically used in prediction of future observations of Y , or for estimating the mean response at a particular level of x .

15.4. Suppose that we have n pairs of observations

$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n).$$

Figure 26 shows a typical scatter plot of observed data and a candidate for the estimated regression line. The estimates of β_0 and β_1 should result in a line that is (in some sense) a “best fit” to the data.

The German scientist Karl Gauss (1777-1855) proposed estimating the parameters β_0 and β_1 to minimize the sum of the squares of the vertical deviations in Figure 26.

We call this criterion for estimating the regression coefficients the method of **least squares**.

Theorem 15.5. The least squares estimates of the intercept and slope in the simple linear regression model are

$$\hat{\beta}_1 = \frac{\overline{xy} - (\bar{x})(\bar{y})}{\overline{x^2} - (\bar{x})^2} = \frac{\sum_{i=1}^n x_i y_i - \frac{\left(\sum_{i=1}^n x_i\right)\left(\sum_{i=1}^n y_i\right)}{n}}{\sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i\right)^2}{n}}$$

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

- The line $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ is called the **fitted** or **estimated regression line**.
- The error in the fit of the model to the i th observation y_i is given by $e_i = y_i - \hat{y}_i$. This is called the **residual**.

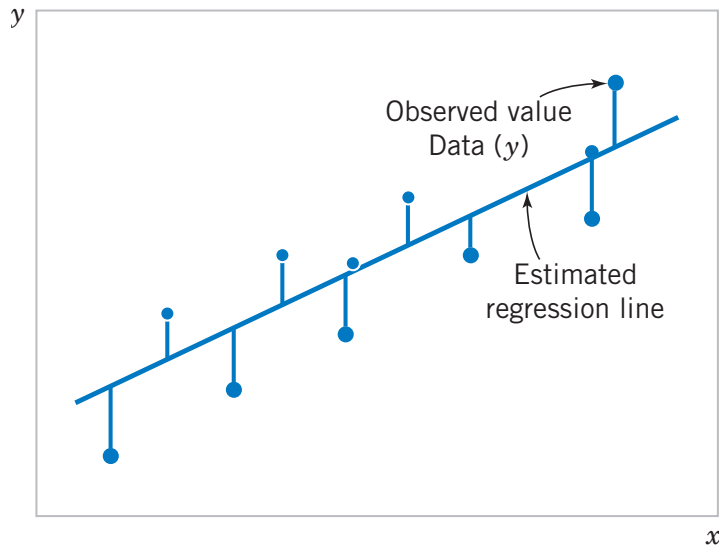


Figure 26: Deviations of the data from the estimated regression model.

15.6. Note that

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$$

where

$$S_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^n x_i y_i - \frac{\left(\sum_{i=1}^n x_i\right) \left(\sum_{i=1}^n y_i\right)}{n}$$

and

$$S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i\right)^2}{n} = \sum_{i=1}^n x_i^2 - n(\bar{x})^2$$

15.7. The residuals $e_i = y_i - \hat{y}_i$ are used to obtain an estimate of σ^2 . The sum of squares of the residuals, often called the **error sum of squares**, is

$$SS_E = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = SS_T - \hat{\beta}_1 S_{xy}$$

where

$$SS_T = S_{yy} = \sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n y_i^2 - \frac{\left(\sum_{i=1}^n y_i\right)^2}{n} = \sum_{i=1}^n y_i^2 - n(\bar{y})^2$$

Theorem 15.8. The expected value of SS_E is

$$\mathbb{E}[SS_E] = (n - 2)\sigma^2.$$

Therefore an unbiased estimator of σ^2 is

$$\hat{\sigma}^2 = \frac{SS_E}{n - 2}.$$