Test of Hypotheses for Means and Proportions

Null and alternative hypotheses, test statistic, type I and II errors, significance level, p-value

Dr.Mamatha.H.R

Professor

Department of Computer Science and Engineering

PES University

Bangalore

Course material created using various Internet resources and text book

a sample of 50 microdrills had an average lifetime of

$X = 12.68$ holes drilled and a standard deviation of $s = 6.83$.

Let us assume that the main question is whether or not the population mean lifetime μ is greater than 11.

We address this question by examining the value of the sample mean X.

We see that $X > 11$,but because of the uncertainty in X, this does not guarantee that $\mu > 11$.

We would like to know just how certain we can be that $\mu > 11$

. A confidence interval is not quite what we need. a 95% confidence interval for the population mean μ was computed to be *(*10.79, 14.57*)*.

This tells us that we are 95% confident that μ is between 10.79 and 14.57.

But the confidence interval does not directly tell us how confident we can be that $\mu > 11$.

The statement "$\mu > 11$" is a **hypothesis** about the population mean μ.

To determine just how certain we can be that a hypothesis such as this is true, we must perform a **hypothesis test**.

A hypothesis test produces a number between 0 and 1 that measures the degree of certainty we may have in the truth of a hypothesis about a quantity such as a population mean or proportion.

hypothesis tests are closely related to confidence intervals.

In general, whenever a confidence interval can be computed, a hypothesis test can also be performed, and vice versa.

Introduction

- Setting up and testing hypotheses is an essential part of statistical inference. In order to formulate such a test, usually some theory has been put forward, either because it is believed to be true or because it is to be used as a basis for argument, but has not been proved.
- Hypothesis testing refers to the process of using statistical analysis to determine if the differences between observed and hypothesized values are due to random chance or to true differences in the samples.
  - Statistical tests separate significant effects from mere luck or random chance.
  - All hypothesis tests have unavoidable, but quantifiable, risks of making the wrong conclusion.

Introduction

- Suppose that a pharmaceutical company is concerned that the mean potency μ of an antibiotic meet the minimum government potency standards. They need to decide between two possibilities:
  - **The mean potency μ does not exceed the required minimum potency.**
  - **The mean potency μ exceeds the required minimum potency.**
- This is an example of a **test of hypothesis.**

Introduction

- Similar to a courtroom trial. In trying a person for a crime, the jury needs to decide between one of two possibilities:
  - **The person is guilty.**
  - **The person is innocent.**
- To begin with, the person is assumed innocent.
- The prosecutor presents evidence, trying to convince the jury to reject the original assumption of innocence, and conclude that the person is guilty.

Five Steps of a Statistical Test

A statistical test of hypothesis consist of five steps

1. Specify statistical hypothesis which include a **null hypothesis** H_0 and a **alternative hypothesis** H_a
2. Identify and calculate **test statistic**
3. Identify distribution and find **p-value**
4. Make a **decision** to reject or not to reject the null hypothesis
5. State conclusion

Null and Alternative Hypothesis

The null hypothesis, H_0:

- The hypothesis we wish to falsify
- Assumed to be true until we can prove otherwise.

The alternative hypothesis, H_a:

- The hypothesis we wish to prove to be true

Court trial:	**Pharmaceuticals:**
H_0: innocent	**H_0: μ does not exceeds required potency**
H_a: guilty	**H_a: μ exceeds required potency**

Examples of Hypotheses

You would like to determine if the diameters of the ball bearings you produce have a mean of 6.5 cm.

$H_0: \mu = 6.5$

$H_a: \mu \neq 6.5$

(Two-sided or two tailed alternative)

Examples of Hypotheses

Do the "16 ounce" cans of peaches meet the claim on the label (on the average)?

Notice, the real concern would be selling the consumer less than 16 ounces of peaches.

H_0: $\mu \geq 16$

H_a: $\mu < 16$

One-sided or one-tailed alternative

Comments on Setting up Hypothesis

- The null hypothesis **must** contain the equal sign.

  This is absolutely necessary because the distribution of test statistic requires the null hypothesis to be assumed to be true and the value attached to the equal sign is then the value assumed to be true.

- The alternate hypothesis should be what you are really attempting to show to be true.

  This is not always possible.

There are two possible **decisions**: reject or fail to reject the null hypothesis. Note we say “fail to reject” or “not to reject” rather than “accept” the null hypothesis.

Two Types of Errors

There are two types of errors which can occur in a statistical test:

- **Type I error**: reject the null hypothesis when it is true
- **Type II error**: fail to reject the null hypothesis when it is false

Actual Fact / Jury's Decision	Guilty	Innocent
Guilty	Correct	Error
Innocent	Error	Correct

Actual Fact / Your Decision	H_0 true	H_0 false
Fail to reject H_0	Correct	Type II Error
Reject H_0	Type I Error	Correct

Error Analogy

Consider a medical test where the hypotheses are equivalent to

H_0: the patient has a specific disease

H_a: the patient doesn't have the disease

Then,

Type I error is equivalent to a false negative

(I.e., Saying the patient does not have the disease when in fact, he does.)

Type II error is equivalent to a false positive

(I.e., Saying the patient has the disease when, in fact, he does not.)

Two Types of Errors

Define:

α = P(Type I error) = P(reject H_0 when H_0 is true)

β =P(Type II error) = P(fail to reject H_0 when H_0 is false)

We want to keep the both α and β as small as possible. The value of α is controlled by the experimenter and is called the **significance level**.

Generally, with everything else held constant, decreasing one type of error causes the other to increase.

Balance Between α and β

- The only way to decrease both types of error simultaneously is to increase the sample size.
- No matter what decision is reached, there is always the risk of one of these errors.
- Balance: identify the largest **significance level** α as the maximum tolerable risk you want to have of making a type I error. Employ a test procedure that makes type II error β as small as possible while maintaining type I error smaller than the given significance level α.

Test Statistic

• A **test statistic** is a quantity calculated from sample of data. Its value is used to decide whether or not the null hypothesis should be rejected.

• The choice of a test statistic will depend on the assumed probability model and the hypotheses under question. We will learn specific test statistics later.

• We then find sampling distribution of the **test statistic** and calculate the probability of rejecting the null hypothesis (**type I error**) if it is in fact true. This probability is called the p-value

P-value

• The p-**value** is a measure of inconsistency between the hypothesized value under the null hypothesis and the observed sample.

• The p-value is the probability, assuming that H_0 is true, of obtaining a test statistic value at least as inconsistent with H_0 as what actually resulted.

• It measures whether the test statistic is **likely** or **unlikely**, assuming H_0 is true. Small p-values suggest that the null hypothesis is unlikely to be true. The smaller it is, the more convincing is the rejection of the null hypothesis. It indicates the strength of evidence for rejecting the null hypothesis H_0

Decision

A decision as to whether H_0 should be rejected results from comparing the p-value to the chosen significance level α:

- H_0 should be rejected if p-value $\leq \alpha$.
- H_0 should not be rejected if p-value $> \alpha$.

When p-value$>\alpha$, state "fail to reject H_0" or "not to reject" rather than "accepting H_0". Write "there is insufficient evidence to reject H_0".

Another way to make decision is to use **critical value** and **rejection region**.

Five Steps of a Statistical Test

A statistical test of hypothesis consist of five steps

1. Specify the **null hypothesis** H_0 and **alternative hypothesis** H_a in terms of population parameters
2. Identify and calculate **test statistic**
3. Identify distribution and find **p-value**
4. Compare p-value with the given significance level and decide if to reject the null hypothesis
5. State conclusion

Large Sample Test for Population Mean

Step 1: Specify the null and alternative hypothesis

- $\mathbf{H_0: \mu = \mu_0}$ **versus** $\mathbf{H_a: \mu \neq \mu_0}$ (two-sided test)
- $\mathbf{H_0: \mu = \mu_0}$ **versus** $\mathbf{H_a: \mu > \mu_0}$ (one-sided test)
- $\mathbf{H_0: \mu = \mu_0}$ **versus** $\mathbf{H_a: \mu < \mu_0}$ (one-sided test)

Step 2: Test statistic for large sample ($n \geq 30$)

$$z = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$$

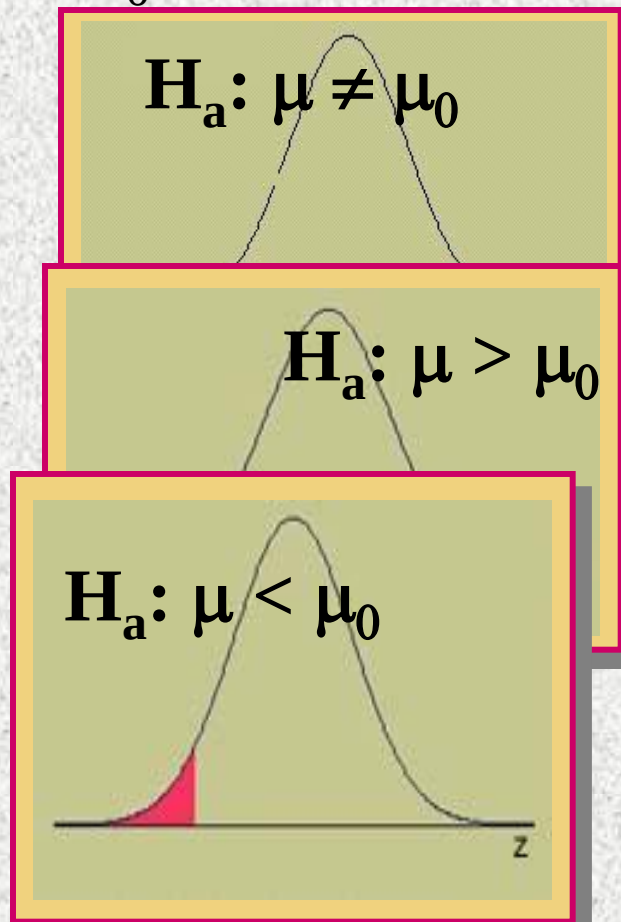
where n, $\bar{x}$ and s are sample size, mean and standard deviation

Intuition of the Test Statistic

If H_0 is true, the value of $\bar{x}$ should be close to μ_0, and z will be close to 0. If H_0 is false, $\bar{x}$ will be much larger or smaller than μ_0, and z will be much larger or smaller than 0, indicating that we should reject H_0. Thus

- z is much larger or smaller than 0 provides evidence against H_0
- z is much larger than 0 provides evidence against H_0
- z is much smaller than 0 provides evidence against H_0

How much larger (or smaller) is large (small) enough?



Large Sample Test for Population Mean

Step 3: When n is large, the sampling distribution of z will be approximately standard normal under H_0.

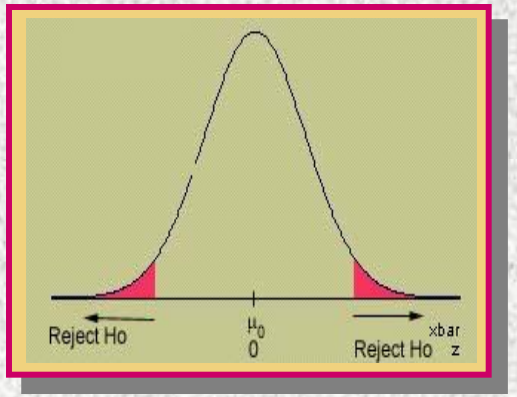
Compute sample statistic

$$z^* = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

z is defined for any possible sample. Thus it is a random variable which can take many different values and the sampling distribution tells us the chance of each value. z* is computed from the given data, thus a fixed number.

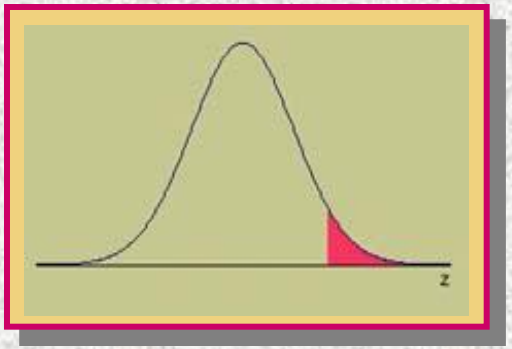
Large Sample Test for Population Mean

– $\mathbf{H_a: \mu \neq \mu_0}$ (two-sided test)



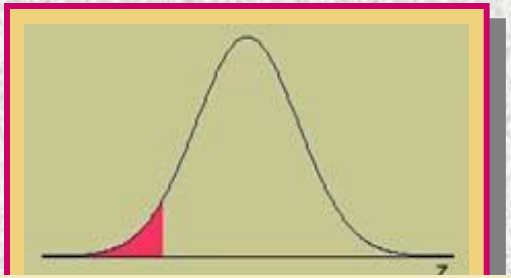
$$\text{p-value} = P(z < -|z^*|) + P(z > |z^*|)$$
$$= 2P(z > |z^*|)$$

– $\mathbf{H_a: \mu > \mu_0}$ (one-sided test)



$$\text{p-value} = P(z > z^*)$$

– $\mathbf{H_a: \mu < \mu_0}$ (one-sided test)



$$\text{p-value} = P(z < z^*)$$

P(z>|z*|), P(z>z*) and P(z<z*) can be found from the normal table

Example

The daily yield for a chemical plant has averaged 880 tons for several years. The quality control manager wants to know if this average has changed. She randomly selects 50 days and records an average yield of 871 tons with a standard deviation of 21 tons. Conduct the test using α=.05.

$$H_0 : \mu = 880 \text{ vs } H_a : \mu \neq 880$$

Test statistic:

$$\mu_0 = 880, n = 50, \bar{x} = 871, s = 21$$

$$z^* \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{871 - 880}{21/\sqrt{50}} = -3.03$$

Example

p - value : this is a two - sided test

$$p - \text{value} = 2 \times P(z > 3.03) = 2(.0012) = .0024$$

Decision: since p-value<α, we reject the hypothesis that μ=880.

Conclusion: the average yield has changed and the change is **statistically significant** at level .05.

In fact, the p-value tells us more: the null hypothesis is very unlikely to be true. If the significance level is set to be any value greater or equal to .0024, we would still reject the null hypothesis. Thus, another interpretation of the p-value is the **smallest** level of significance at which H_0 would be rejected, and p-value is also called the **observed significance level**.

Example

A homeowner randomly samples 64 homes similar to her own and finds that the average selling price is $252,000 with a standard deviation of $15,000. Is this sufficient evidence to conclude that the average selling price is greater than $250,000? Use $\alpha = .01$.

$$H_0 : \mu = 250{,}000 \quad \text{vs} \quad H_a : \mu > 250{,}000$$

Test statistic:

$$z^* \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{252{,}000 - 250{,}000}{15{,}000/\sqrt{64}} = 1.07$$

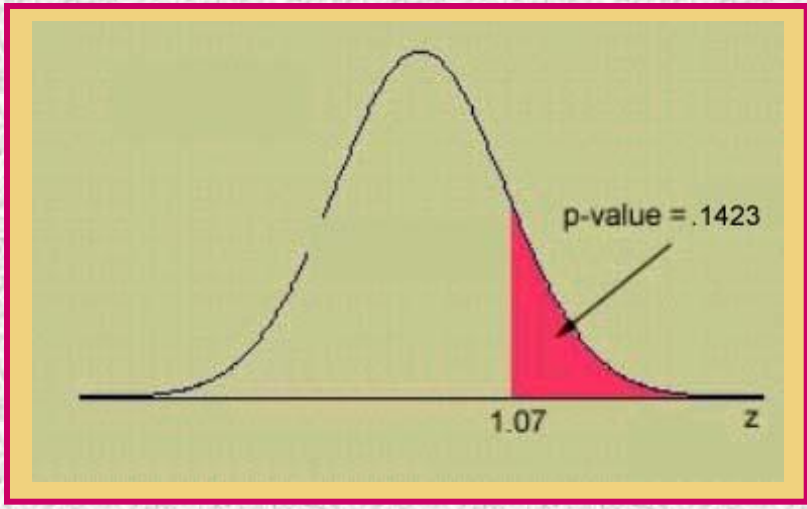
Example

p - value :

this is a one - sided test

$p - \text{value} = P(z > 1.07)$

$= .5 - .3577 = .1423$



Decision: since the p-value is greater than $\alpha = .01$, H_0 is not rejected.

Conclusion: there is insufficient evidence to indicate that the average selling price is greater than $250,000.

Summary

Let $X_1, \ldots, X_n$ be a *large* (e.g., $n > 30$) sample from a population with mean μ and standard deviation σ.
To test a null hypothesis of the form $H_0 : \mu \leq \mu_0$, $H_0 : \mu \geq \mu_0$, or $H_0 : \mu = \mu_0$:

- Compute the z-score: $z = \dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$.
  If σ is unknown it may be approximated with s.
- Compute the P-value. The P-value is an area under the normal curve, which depends on the alternate hypothesis as follows:

Alternate Hypothesis	P-value
$H_1 : \mu > \mu_0$	Area to the right of z
$H_1 : \mu < \mu_0$	Area to the left of z
$H_1 : \mu \neq \mu_0$	Sum of the areas in the tails cut off by z and $-z$

Summary

- The smaller the P-value, the more certain we can be that H_0 is false.
- The larger the P-value, the more plausible H_0 becomes, but we can never be certain that H_0 is true.
- A rule of thumb suggests to reject H_0 whenever $P \leq 0.05$. While this rule is convenient, it has no scientific basis.

Statistical Significance

- Whenever the *P*-value is less than a particular threshold, the result is said to be "statistically significant" at that level.
- for example, if $P \leq 0.05$, the result is statistically significant at the 5% level;
- if $P \leq 0.01$, the result is statistically significant at the 1% level.
- If a result is statistically significant at the $100\alpha\%$ level, we can also say that the null hypothesis is "rejected at level $100\alpha\%$."

Example

A hypothesis test is performed of the null hypothesis $H0 : \mu = 0$. The P-value turns out to be 0.03. Is the result statistically significant at the 10% level? The 5% level ? The 1% level? Is the null hypothesis rejected at the 10% level? The 5% level? The 1% level?

Solution

The result is statistically significant at any level greater than or equal to 3%.

Thus it is statistically significant at the 10% and 5% levels, but not at the 1% level.

Similarly, we can reject the null hypothesis at any level greater than or equal to 3%, so H_0 is rejected at the 10% and 5% levels, but not at the 1% level.

Summary

Let α be any value between 0 and 1. Then, if $P \leq \alpha$,

- The result of the test is said to be statistically significant at the $100\alpha\%$ level.
- The null hypothesis is rejected at the $100\alpha\%$ level.
- When reporting the result of a hypothesis test, report the P-value, rather than just comparing it to 5% or 1%.

a result was reported as "statistically significant at the 5% level" or "statistically significant ($P < 0.05$)."

This is a poor practice, for three reasons.

First, it provides no way to tell whether the *P*-value was just barely less than 0.05, or whether it was a lot less.

Second, it implies that there is a big difference between a *P*-value just under 0.05 and one just above 0.05, when in fact there is little difference.

Third, a report like this does not allow readers to decide for themselves whether the *P*-value is small enough to reject the null hypothesis.

If a reader believes that the null hypothesis should not be rejected unless $P < 0.01$, then reporting only that $P < 0.05$ does not allow that reader to decide whether or not to reject *H*0.

The *P*-value Is Not the Probability That *H*0 Is True

Since the *P*-value is a probability, and since small *P*-values indicate that *H*0 is unlikely to be true, it is tempting to think that the *P*-value represents the probability that *H*0 is true.

The null hypothesis, either is true or is not true. The truth or falsehood of *H*0 cannot be changed by repeating the experiment.

It is therefore not correct to discuss the "probability" that *H*0 is true.

Choose *H*0 to Answer the Right Question

When performing a hypothesis test, it is important to choose *H*0 and *H*1 appropriately so that the result of the test can be useful in forming a conclusion.

Specifications for a water pipe call for a mean breaking strength μ of more than 2000 lb per linear foot. Engineers will perform a hypothesis test to decide whether or not to use a certain kind of pipe. They will select a random sample of 1 ft sections of pipe, measure their breaking strengths, and perform a hypothesis test. The pipe will not be used unless the engineers can conclude that $\mu > 2000$. Assume they test $H0 : \mu \leq 2000$ versus $H1 : \mu > 2000$. Will the engineers decide to use the pipe if $H0$ is rejected? What if $H0$ is not rejected?

Solution

If $H0$ is rejected, the engineers will conclude that $\mu > 2000$, and they will use the pipe. If $H0$ is not rejected, the engineers will conclude that μ *might* be less than or

equal to 2000, and they will not use the pipe.

In this Example, the engineers' action with regard to using the pipe will differ depending on whether $H0$ is rejected or not. This is therefore a useful test to perform,and $H0$ and $H1$ have been specified correctly.

In Example , assume the engineers test $H0 : \mu \geq 2000$ versus $H1 : \mu < 2000$.Will the engineers decide to use the pipe if $H0$ is rejected? What if $H0$ is not rejected?

Solution

If $H0$ is rejected, the engineers will conclude that$\mu < 2000$, and they will not use the pipe. If $H0$ is not rejected, the engineers will conclude that μ *might* be greater than or equal to 2000, but that it also might not be. So again, they won't use the pipe.

In Example , the engineers' action with regard to using the pipe will be the same—they won't use it—whether or not $H0$ is rejected. There is no point in performing this test. The hypotheses $H0$ and $H1$ have not been specified correctly

Final note: In a one-tailed test, the equality always goes with the null hypothesis.

Thus if $\mu 0$ is the point that divides $H0$ from $H1$, we may have $H0 : \mu \leq \mu 0$ or $H0 : \mu \geq \mu 0$,but never $H0 : \mu < \mu 0$ or $H0 : \mu > \mu 0$

when defining the null distribution, we represent $H0$ with the value of μ closest to $H1$.

Without the equality, there is no value specified by $H0$ that is the closest to $H1$. Therefore the equality must go with $H0$.

Statistical Significance Is Not the Same as Practical Significance

Sometimes statistically significant results do not have any scientific or practical importance.

Assume that a process used to manufacture synthetic fibers is known to produce fibers with a mean breaking strength of 50 N. A new process, which would require considerable retooling to implement, has been developed. In a sample of 1000 fibers produced by this new method, the average breaking strength was 50.1 N, and the standard deviation was 1 N. Can we conclude that the new process produces fibers with greater mean breaking strength?

let μ be the mean breaking strength of fibers produced by the new process. We need to test $H0 : \mu \leq 50$ versus $H1 : \mu > 50$.

In this way, if we reject $H0$, we will conclude that the new process is better. Under $H0$, the sample mean X has a normal distribution with mean 50 and standard deviation $1/\sqrt{1000} = 0.0316$.

The z-score is

$z = 50.1 - 50/0.0316$

$= 3.16$

The P-value is 0.0008. This is very strong evidence against $H0$.

The new process produces fibers with a greater mean breaking strength.

What practical conclusion should be drawn from this result?

On the basis of the hypothesis test, we are quite sure that the new process is better. Would it be worthwhile

to implement the new process?

Probably not.

The reason is that the difference between

the old and new processes, although highly statistically significant, amounts to only 0.1 N.

It is unlikely that this difference is large enough to matter.

a result can be statistically significant without being large enough to be of practical importance.

A difference is statistically significant when it is large compared to its standard deviation. In the example, a difference of 0.1 N was statistically significant because the standard deviation was only 0.0316 N.

When the standard deviation is very small, even a small difference can be statistically significant.

The *P*-value does not measure practical significance.

What it does measure is the degree of confidence we can have that the true value is really different from the value specified by the null hypothesis.

When the *P*-value is small, then we can be confident that the true value is really different

The Relationship Between Hypothesis Tests and Confidence Intervals

In a hypothesis test for a population mean μ, we specify a particular value of μ (the null hypothesis) and determine whether that value is plausible.

In contrast, a confidence interval for a population mean μ can be thought of as the collection of all values for μ that meet a certain criterion of plausibility, specified by the confidence level $100(1-\alpha)\%$.

the values contained within a two-sided level $100(1-\alpha)\%$ confidence interval for a population mean μ are precisely those values for which the P-value of a two-tailed hypothesis test will be greater than α.

The sample mean lifetime of 50 microdrills was $X = 12.68$ holes drilled and the standard deviation was $s = 6.83$.

Setting α to 0.05 (5%), the 95% confidence interval for the population mean lifetime μ was computed to be $(10.79, 14.57)$.

Suppose we wanted to test the hypothesis that μ was equal to one of the endpoints of the confidence interval.

For example, consider testing $H0 : \mu = 10.79$ versus $H1 : \mu \neq 10.79$.

Under $H0$, the observed value $X = 12.68$ comes from a normal distribution with mean 10.79 and standard deviation

$6.83/\sqrt{50} = 0.9659$.

The z-score is $(12.68 - 10.79)/0.9659 = 1.96$. Since $H0$ specifies that μ is *equal* to 10.79, both tails contribute to the P-value, which is 0.05, and thus equal to α

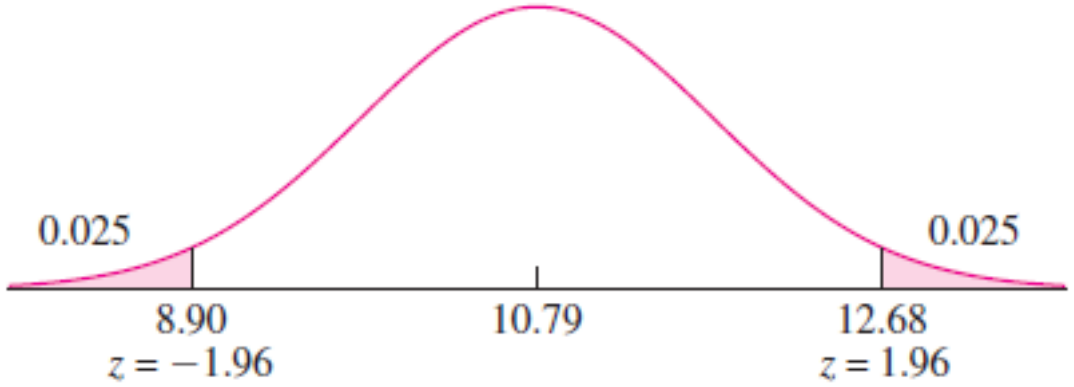


FIGURE 6.4 The sample mean $\overline{X}$ is equal to 12.68. Since 10.79 is an endpoint of a 95% confidence interval based on $\overline{X} = 12.68$, the P-value for testing $H_0 : \mu = 10.79$ is equal to 0.05.

consider testing the hypothesis $H0: \mu = 14.57$ versus $H1: \mu \neq 14.57$, where 14.57 is the other endpoint of the confidence interval. This time we will obtain

$z = (12.68 - 14.57)/0.9659 = -1.96$, and again the P-value is 0.05.

It is easy to check that if we choose any value $\mu 0$ in the interval $(10.79, 14.57)$ and test $H0: \mu = \mu 0$ versus $H1: \mu = \mu 0$, the P-value will be greater than 0.05.

On the other hand, if we choose $\mu 0 < 10.79$ or $\mu 0 > 14.57$, the P-value will be less than 0.05.

Thus the 95% confidence interval consists of precisely those values of μ whose P-values are greater than 0.05 in a hypothesis test.

In this sense, the confidence interval contains all the values that are plausible for the population mean μ.

a one-sided level $100(1-\alpha)\%$ confidence interval consists of all the values for which the *P*-value in a one-tailed test would be greater than α.

For example, with $X = 12.68$, $s = 6.83$, and $n = 50$, the 95% lower confidence bound for the lifetime of the drills is 11.09.

If $\mu 0 > 11.09$, then the *P*-value for testing $H0 : \mu \leq \mu 0$

will be greater than 0.05. Similarly, the 95% upper confidence bound for the lifetimes of the drills is 14.27. If $\mu 0 < 14.27$, then the *P*-value for testing $H0 : \mu \geq \mu 0$ will be greater than 0.05.

Tests for a Population Proportion

A supplier of semiconductor wafers claims that of all the wafers he supplies, no more than 10% are defective. A sample of 400 wafers is tested, and 50 of them, or 12.5%, are defective. Can we conclude that the claim is false?

from the Central Limit Theorem, since the sample size is large, that

$$P \sim N(p, \; p(1-p)/n)$$

The null and alternate hypotheses are

$H0 : p \leq 0.1$ versus $H1 : p > 0.1$

To perform the hypothesis test, we assume H_0 to be true and take $p = 0.1$. Substituting $p = 0.1$ and $n = 400$ in expression (6.1) yields the null distribution of $\hat{p}$:

$$\hat{p} \sim N(0.1,\ 2.25 \times 10^{-4})$$

The standard deviation of $\hat{p}$ is $\sigma_{\hat{p}} = \sqrt{2.25 \times 10^{-4}} = 0.015$. The observed value of $\hat{p}$ is $50/400 = 0.125$. The z-score of $\hat{p}$ is

$$z = \frac{0.125 - 0.100}{0.015} = 1.67$$

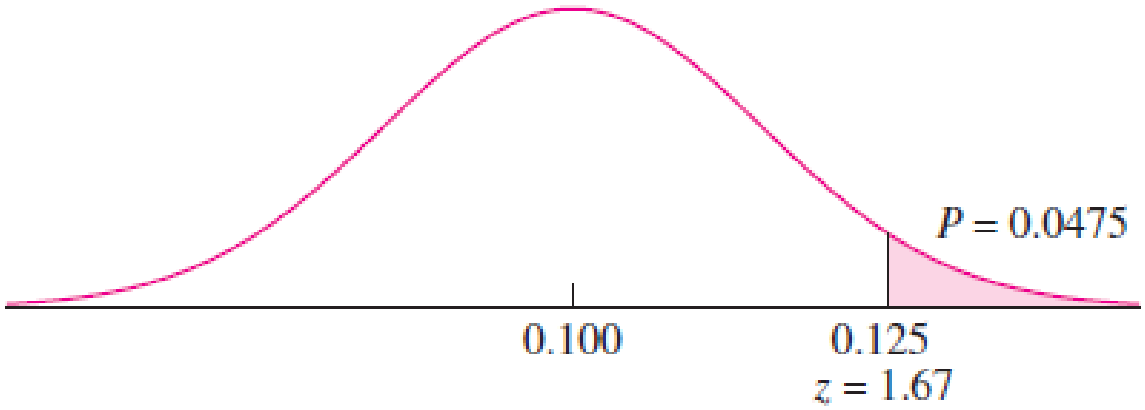


FIGURE 6.5 The null distribution of $\hat{p}$ is $N(0.1, 0.015^2)$. Thus if H_0 is true, the probability that $\hat{p}$ takes on a value as extreme as or more extreme than the observed value of 0.125 is 0.0475. This is the P-value.

The article "Refinement of Gravimetric Geoid Using GPS and Leveling Data" (W. Thurston, *Journal of Surveying Engineering*, 2000:27–56) presents a method for measuring orthometric heights above sea level. For a sample of 1225 baselines, 926 gave results that were within the class C spirit leveling tolerance limits. Can we conclude that this method produces results within the tolerance limits more than 75% of the time?

Solution

Let p denote the probability that the method produces a result within the tolerance limits. The null and alternate hypotheses are

$$H_0 : p \leq 0.75 \quad \text{versus} \quad H_1 : p > 0.75$$

The sample proportion is $\widehat{p} = 926/1225 = 0.7559$. Under the null hypothesis, $\widehat{p}$ is normally distributed with mean 0.75 and standard deviation $\sqrt{(0.75)(1-0.75)/1225} = 0.0124$. The z-score is

$$z = \frac{0.7559 - 0.7500}{0.0124} = 0.48$$

The P-value is 0.3156 (see Figure 6.6). We cannot conclude that the method produces good results more than 75% of the time.

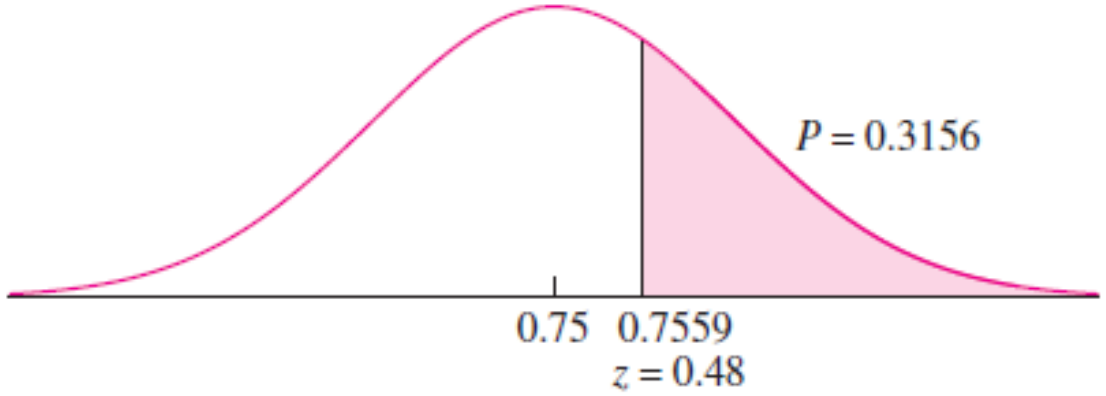


FIGURE 6.6 The null distribution of $\hat{p}$ is $N(0.75,\ 0.0124^2)$. Thus if H_0 is true, the probability that $\hat{p}$ takes on a value as extreme as or more extreme than the observed value of 0.7559 is 0.3156. This is the P-value.

Summary

Let X be the number of successes in n independent Bernoulli trials, each with success probability p; in other words, let $X \sim \text{Bin}(n,\ p)$.

To test a null hypothesis of the form $H_0 : p \leq p_0$, $H_0 : p \geq p_0$, or $H_0 : p = p_0$, assuming that both np_0 and $n(1 - p_0)$ are greater than 10:

- Compute the z-score: $z = \dfrac{\widehat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}}$.
- Compute the P-value. The P-value is an area under the normal curve, which depends on the alternate hypothesis as follows:

Alternate Hypothesis	P-value
$H_1 : p > p_0$	Area to the right of z
$H_1 : p < p_0$	Area to the left of z
$H_1 : p \neq p_0$	Sum of the areas in the tails cut off by z and $-z$