8 Tests of Hypotheses Based on a Single Sample

8.1 Hypotheses and Test Procedures

μ อยู่ในช่วง CI รึเปล่า ?

Hypotheses and Test Procedures

A **statistical hypothesis,** or just *hypothesis*, is a claim or assertion either about the value of a single parameter (population characteristic or characteristic of a probability distribution), about the values of several parameters, or about the form of an entire probability distribution.

One example of a hypothesis is the claim $\mu = 0.75$, where μ is the true average inside diameter of a certain type of PVC pipe.

Another example is the statement $p < 0.10$, where p is the proportion of defective circuit boards among all circuit boards produced by a certain manufacturer.

Hypotheses and Test Procedures

If μ_1 and μ_2 denote the true average breaking strengths of two different types of twine, one hypothesis is the assertion that $\mu_1 - \mu_2 = 0$, and another is the statement $\mu_1 - \mu_2 > 5$.

Yet another example of a hypothesis is the assertion that the stopping distance under particular conditions has a normal distribution.

In any hypothesis-testing problem, there are two contradictory hypotheses under consideration.
One hypothesis might be the claim $\mu = 0.75$ and
the other $\mu \neq 0.75$, or the two contradictory statements might be $p \geq .10$ and $p < 0.10$.

Hypotheses and Test Procedures

The objective is to decide, based on sample information, which of the two hypotheses is correct.

There is a familiar analogy to this in a criminal trial. One claim is the assertion that the accused individual is innocent.

In the U.S. judicial system, this is the claim that is initially believed to be true. Only in the face of strong evidence to the contrary should the jury reject this claim in favor of the alternative assertion that the accused is guilty.

Hypotheses and Test Procedures

In this sense, the claim of innocence is the favored or protected hypothesis, and the burden of proof is placed on those who believe in the alternative claim.

Similarly, in testing statistical hypotheses, the problem will be formulated so that one of the claims is initially favored.

This initially favored claim will not be rejected in favor of the alternative claim unless sample evidence contradicts it and provides strong support for the alternative assertion.

Hypotheses and Test Procedures

Definition

The **null hypothesis,** denoted by H_0, is the claim that is initially assumed to be true (the "prior belief" claim).

The **alternative hypothesis,** denoted by H_a, is the assertion that is contradictory to H_0.The null hypothesis will be rejected in favor of the alternative hypothesis only if sample evidence suggests that H_0 is false.

If the sample does not strongly contradict H_0, we will continue to believe in the plausibility of the null hypothesis. The two possible conclusions from a hypothesis-testing analysis are then *reject* H_0 or *fail to reject* H_0.

Hypotheses and Test Procedures

A **test of hypotheses** is a method for using sample data to decide whether the null hypothesis should be rejected.

Thus we might test H_0: $\mu = 0.75$ against the alternative H_a: $\mu \neq 0.75$. Only if sample data strongly suggests that μ is something other than 0.75 should the null hypothesis be rejected.

In the absence of such evidence, H_0 should not be rejected, since it is still quite plausible.

Hypotheses and Test Procedures

Sometimes an investigator does not want to accept a particular assertion unless and until data can provide strong support for the assertion.

As an example, suppose a company is considering putting a new type of coating on bearings that it produces.

The true average wear life with the current coating is known to be 1000 hours. With μ denoting the true average life for the new coating, the company would not want to make a change unless evidence strongly suggested that μ exceeds 1000.

Hypotheses and Test Procedures

An appropriate problem formulation would involve testing H_0: $\mu = 1000$ against H_a: $\mu > 1000$.

The conclusion that a change is justified is identified with H_a, and it would take conclusive evidence to justify rejecting H_0 and switching to the new coating.

Scientific research often involves trying to decide whether a current theory should be replaced by a more plausible and satisfactory explanation of the phenomenon under investigation.

Hypotheses and Test Procedures

A conservative approach is to identify the current theory with H_0 and the researcher's alternative explanation with H_a.

Rejection of the current theory will then occur only when evidence is much more consistent with the new theory.

In many situations, H_a is referred to as the "researcher's hypothesis," since it is the claim that the researcher would really like to validate.

Hypotheses and Test Procedures

The word *null* means “of no value, effect, or consequence,” which suggests that H_0 should be identified with the hypothesis of no change (from current opinion), no difference, no improvement, and so on.

Suppose, for example, that 10% of all circuit boards produced by a certain manufacturer during a recent period were defective.

An engineer has suggested a change in the production process in the belief that it will result in a reduced defective rate.

Hypotheses and Test Procedures

Let p denote the true proportion of defective boards resulting from the changed process.

Then the research hypothesis, on which the burden of proof is placed, is the assertion that $p < 0.10$. Thus the alternative hypothesis is H_a: $p < 0.10$.

In our treatment of hypothesis testing, H_0 will generally be stated as an equality claim. If θ denotes the parameter of interest, the null hypothesis will have the form H_0: $\theta = \theta_0$, where θ_0 is a specified number called the *null value* of the parameter (value claimed for θ by the null hypothesis).

Hypotheses and Test Procedures

As an example, consider the circuit board situation just discussed. The suggested alternative hypothesis was H_a: $p < 0.10$, the claim that the defective rate is reduced by the process modification.

A natural choice of H_0 in this situation is the claim that $p \geq 0.10$, according to which the new process is either no better *or* worse than the one currently used.

We will instead consider H_0: $p = 0.10$ versus H_a: $p < 0.10$.

Hypotheses and Test Procedures

The rationale for using this simplified null hypothesis is that any reasonable decision procedure for deciding between H_0: $p = 0.10$ and H_a: $p < 0.10$ will also be reasonable for deciding between the claim that $p \geq 0.10$ and H_a.

The use of a simplified H_0 is preferred because it has certain technical benefits, which will be apparent shortly.

Hypotheses and Test Procedures

The alternative to the null hypothesis H_a: $\theta = \theta_0$ will look like one of the following three assertions:

1. H_a: $\theta > \theta_0$
(in which case the implicit null hypothesis is $\theta \leq \theta_0$),

2. H_a: $\theta < \theta_0$
(in which case the implicit null hypothesis is $\theta \geq \theta_0$), or

3. H_a: $\theta \neq \theta_0$

Hypotheses and Test Procedures

For example, let σ denote the standard deviation of the distribution of inside diameters (inches) for a certain type of metal sleeve.

If the decision was made to use the sleeve unless sample evidence conclusively demonstrated that $\sigma > 0.001$, the appropriate hypotheses would be H_0: $\sigma = 0.001$. versus H_a: $\sigma > 0.001$.

The number θ_0 that appears in both H_0 and H_a (separates the alternative from the null) is called the **null value.**

Test Procedures

Test Procedures

A test procedure is a rule, based on sample data, for deciding whether to reject H_0.

A test of H_0: $p = .10$ versus H_a: $p < .10$ in the circuit board problem might be based on examining a random sample of $n = 200$ boards.

Let X denote the number of defective boards in the sample, a binomial random variable; x represents the observed value of X.

Test Procedures

If H_0 is true, $E(X) = np = 200(.10) = 20$, whereas we can expect fewer than 20 defective boards if H_a is true.

A value x just a bit below 20 does not strongly contradict H_0, so it is reasonable to reject H_0 only if x is substantially less than 20.

One such test procedure is to reject H_0 if $x \leq 15$ and not reject H_0 otherwise.

Test Procedures

This procedure has two constituents:

(1) a *test statistic*, or function of the sample data used to make a decision, and

(2) a *rejection region* consisting of those x values for which H_0 will be rejected in favor of H_a.

For the rule just suggested, the rejection region consists of $x = 0, 1, 2, \ldots,$ and 15.

H_0 will not be rejected if $x = 16, 17, \ldots, 199$, or 200.

Test Procedures

A test procedure is specified by the following:

1. A **test statistic,** a function of the sample data on which the decision (reject H_0 or do not reject H_0) is to be based

2. A **rejection region,** the set of all test statistic values for which H_0 will be rejected

The null hypothesis will then be rejected if and only if the observed or computed test statistic value falls in the rejection region.

Errors in Hypothesis Testing

Errors in Hypothesis Testing

The basis for choosing a particular rejection region lies in consideration of the errors that one might be faced with in drawing a conclusion.

Consider the rejection region $x \leq 15$ in the circuit board problem. Even when H_0: $p = .10$ is true, it might happen that an unusual sample results in $x = 13$, so that H_0 is erroneously rejected.

On the other hand, even when H_a: $p < .10$ is true, an unusual sample might yield $x = 20$, in which case H_0 would not be rejected—again an incorrect conclusion.

Errors in Hypothesis Testing

Thus it is possible that H_0 may be rejected when it is true or that H_0 may not be rejected when it is false.

These possible errors are not consequences of a foolishly chosen rejection region.

Either error might result when the region $x \leq 14$ is employed, or indeed when any other sensible region is used.

Errors in Hypothesis Testing

Definition

A **type I error** consists of rejecting the null hypothesis H_0 when it is true.

A **type II error** involves not rejecting H_0 when H_0 is false.

In the nicotine scenario, a type I error consists of rejecting the manufacturer's claim that $\mu = 1.5$ when it is actually true.

If the rejection region $\bar{x} \geq 1.6$ is employed, it might happen that $\bar{x} = 1.63$ even when $\mu = 1.5$, resulting in a type I error.

Errors in Hypothesis Testing

Alternatively, it may be that H_0 is false and yet $\bar{x} = 1.52$ is observed, leading to H_0 not being rejected (a type II error).

In the best of all possible worlds, test procedures for which neither type of error is possible could be developed.

However, this ideal can be achieved only by basing a decision on an examination of the entire population. The difficulty with using a procedure based on sample data is that because of sampling variability, an unrepresentative sample may result, e.g., a value of $\overline{X}$ that is far from μ or a value of $\hat{p}$ that differs considerably from p.

Errors in Hypothesis Testing

Instead of demanding error-free procedures, we must seek procedures for which either type of error is unlikely to occur.

That is, a good procedure is one for which the probability of making either type of error is small.

The choice of a particular rejection region cutoff value fixes the probabilities of type I and type II errors.

Errors in Hypothesis Testing

These error probabilities are traditionally denoted by α and β, respectively.

Because H_0 specifies a unique value of the parameter, there is a single value of α.

However, there is a different value of β for each value of the parameter consistent with H_a.

Example 1

A certain type of automobile is known to sustain no visible damage 25% of the time in 10-mph crash tests. A modified bumper design has been proposed in an effort to increase this percentage.

Let p denote the proportion of all 10-mph crashes with this new bumper that result in no visible damage.

The hypotheses to be tested are H_0: $p = 0.25$ (no improvement) versus H_a: $p > 0.25$.

The test will be based on an experiment involving $n = 20$ independent crashes with prototypes of the new design.

Example 1

cont'd

Intuitively, H_0 should be rejected if a substantial number of the crashes show no damage. Consider the following test procedure:

Test statistic: X = the number of crashes with no visible damage

Rejection region: $R_8 = \{8, 9, 10, \ldots, 19, 20\}$; that is, reject H_0

if $x \geq 8$, where x is the observed value of the test statistic.

Example 1

cont'd

This rejection region is called *upper-tailed* because it consists only of large values of the test statistic.

When H_0 is true, X has a binomial probability distribution with n = 20 and p = 0.25. Then

$$\begin{aligned}\alpha &= P(\text{type I error}) = P(H_0 \text{ is rejected when it is true}) \\ &= P(X \geq 8 \text{ when } X \sim \text{Bin}(20, 0.25)) \\ &= 1 - B(7; 20, 0.25) \\ &= 1 - .898 \\ &= .102\end{aligned}$$

Example 1

cont'd

That is, when H_0 is actually true, roughly 10% of all experiments consisting of 20 crashes would result in H_0 being incorrectly rejected (a type I error).

In contrast to α, there is not a single β. Instead, there is a different β for each different p that exceeds 0.25.

Thus there is a value of β for $p = 0.3$ (in which case $X \sim \text{Bin}(20, 0.3)$), another value of β for $p = 0.5$, and so on.

Example 1

cont'd

For example,

$\beta(.3) = P(\text{type II error when } p = 0.3)$

$= P(H_0 \text{ is not rejected when it is false because } p = 0.3)$

$= P(X \leq 7 \text{ when } X \sim \text{Bin}(20, 0.3))$

$= B(7; 20, 0.3) = .772$

When p is actually 0.3 rather than 0.25 (a "small" departure from H_0), roughly 77% of all experiments of this type would result in H_0 being incorrectly not rejected!

Example 1

cont'd

The accompanying table displays β for selected values of p (each calculated for the rejection region R_8).

p	.3	.4	.5	.6	.7	.8
$\beta(p)$	.772	.416	.132	.021	.001	.000

Clearly, β decreases as the value of p moves farther to the right of the null value .25.

Intuitively, the greater the departure from H_0, the less likely it is that such a departure will not be detected.

Example 1

cont'd

The proposed test procedure is still reasonable for testing the more realistic null hypothesis that $p \leq .25$.

In this case, there is no longer a single α, but instead there is an α for each p that is at most .25: $\alpha(.25)$, $\alpha(.23)$, $\alpha(.20)$, $\alpha(.15)$, and so on. It is easily verified, though, that $\alpha(p) < \alpha(.25) = .102$ if $p < .25$.

That is, the largest value of α occurs for the boundary value .25 between H_0 and H_a.

Thus if α is small for the simplified null hypothesis, it will also be as small as or smaller for the more realistic H_0.

Errors in Hypothesis Testing

Proposition

Suppose an experiment and a sample size are fixed and a test statistic is chosen.

Then decreasing the size of the rejection region to obtain a smaller value of α results in a larger value of β for any particular parameter value consistent with H_a.

Errors in Hypothesis Testing

This proposition says that once the test statistic and n are fixed, there is no rejection region that will simultaneously make both α and all β's small.

A region must be chosen to effect a compromise between α and β.

Because of the suggested guidelines for specifying H_0 and H_a, a type I error is usually more serious than a type II error (this can always be achieved by proper choice of the hypotheses).

Errors in Hypothesis Testing

The approach adhered to by most statistical practitioners is then to specify the largest value of a that can be tolerated and find a rejection region having that value of α rather than anything smaller.

This makes β as small as possible subject to the bound on α. The resulting value of α is often referred to as the **significance level** of the test.

Traditional levels of significance are .10, .05, and .01, though the level in any particular problem will depend on the seriousness of a type I error—the more serious this error, the smaller should be the significance level.

Errors in Hypothesis Testing

The corresponding test procedure is called a **level $\boldsymbol{\alpha}$ test** (e.g., a level .05 test or a level .01 test).

A test with significance level α is one for which the type I error probability is controlled at the specified level.

Example 5

Again let μ denote the true average nicotine content of brand B cigarettes. The objective is to test H_o: $\mu = 1.5$ versus H_a: $\mu > 1.5$ based on a random sample $X_1, X_2, \ldots, X_{32}$ of nicotine content.

Suppose the distribution of nicotine content is known to be normal with $\sigma = .20$.

Then $\bar{X}$ is normally distributed with mean value $\mu_{\bar{x}} = \mu$ and standard deviation $\sigma_{\bar{x}} = .20/\sqrt{32} = .0354$.

Example 5

cont'd

Rather than use $\overline{X}$ itself as the test statistic, let's standardize $\overline{X}$, assuming that H_0 is true.

Test statistic: $Z = \dfrac{\overline{X} - 1.5}{\sigma/\sqrt{n}} = \dfrac{\overline{X} - 1.5}{.0354}$

Z expresses the distance between $\overline{X}$ and its expected value when H_0 is true as some number of standard deviations.

For example, $z = 3$ results from an $\overline{x}$ that is 3 standard deviations larger than we would have expected it to be were H_0 true. Rejecting H_0 when $\overline{x}$ "considerably" exceeds 1.5 is equivalent to rejecting H_0 when z "considerably" exceeds 0.

Example 5

cont'd

That is, the form of the rejection region is $z \geq c$. Let's now determine c so that $\alpha = 0.5$. When H_0 is true, Z has a standard normal distribution. Thus

$$\alpha = P(\text{type I error}) = P(\text{rejecting } H_0 \text{ when } H_0 \text{ is true})$$

$$= P(Z \geq c \text{ when } Z \sim N(0, 1))$$

The value c must capture upper-tail area .05 under the z curve. So, directly from Appendix Table A.3,

$$C = z_{.05} = 1.645.$$

Example 5

cont'd

Notice that $z \geq 1.645$ is equivalent to $\bar{x} - 1.5 \geq (.0354)(1.645)$, that is, $\bar{x} \geq 1.56$. Then β involves the probability that $\bar{X}$ **<** 1.56 and can be calculated for any μ greater than 1.5.