STAT355 - Probability & Statistics
Chapter 8: Tests of Hypotheses Based on a Single Sample

Fall 2011

Chapter 8: Tests of Hypotheses Based on a Single Sample

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.

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 $\mu_1 - \mu_2 > 5$.

Hypotheses and Test Procedures

In any hypothesis-testing problem, there are two contradictory hypotheses under consideration. One hypothesis might be the claim $\mu = .75$ and the other $\mu \neq 0.75$, or the two contradictory statements might be $p \geq 0.10$ and $p < 0.10$.

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

- Analogy in a criminal trial.

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.

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

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.

Test Procedures - Example

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.

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 $p < .10$. Thus the alternative hypothesis is $H_a : p < .10$.

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).

Test Procedures - Example

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 $p \neq 0.10$, 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 $H_a : p < .10$.

A test of H_0 versus H_a 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.

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.

Test Procedures - Example

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.

Errors in Hypothesis Testing

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

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.

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

- 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.

Errors in Hypothesis Testing - Example

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 $H_a : p > .25$.

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

Intuitively, H_0 should be rejected if a substantial number of the crashes show no damage.

Errors in Hypothesis Testing - Example

Consider the following test procedure:
Test statistic: $X =$ the number of crashes with no visible damage

Rejection region: $R_8 = \{8, 9, 10, ..., 19, 20\}$; that is, reject H_0 if $x \geq 8$, where x is the observed value of the test statistic.
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 $p = .25$. Then

$$
\begin{aligned}
\alpha &= P(\text{type I error}) \\
&= P(H_0 \text{ is rejected when it is true}) \\
&= P(X \geq 8 \text{ when } X \;\; Bin(n = 20, p = 0.25)) \\
&= 1 - B(7; 20, .25) = 0.102
\end{aligned}
$$

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 .25.
For example,

$$\begin{aligned}
\beta(0.3) &= P(\text{type II error when } p = 0.3) \\
&= P(H_0 \text{ is not rejected when it is false because } p = 0.3) \\
&= P(X \leq 7 \text{ when } X \ Bin(20, .3)) \\
&= B(7; 20, .3) = 0.772
\end{aligned}$$

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.

- 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 α 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 0.10, 0.05, and 0.01, 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.

- The corresponding test procedure is called a level α 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.

Errors in Hypothesis Testing - Example

Suppose that a cigarette manufacturer claims that the average nicotine content μ of brand B cigarette is at most 1.5mg. The objective is to test $H_0 : \mu = 1.5$ versus $H_a : \mu > 1.5$ based on a random sample $X_1, X_2, ..., X_{32}$ of nicotine content.

- Suppose the distribution of nicotine content is known to be normal with $\sigma = 0.20$. Then $\bar{X} \sim N(\mu_X = \mu, \sigma_X = \sigma/\sqrt{32}) = N(\mu, 0.0354)$
- Test statistic: $Z = \frac{\bar{X} - 1.5}{0.0354}$
- Z expresses the distance between X and its expected value when H_0 is true as some number of standard deviations.

Errors in Hypothesis Testing - Example

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

- That is, the form of the rejection region is $z \geq c$. Let's now determine c so that $\alpha = 0.05$. When H_0 is true, Z has a standard normal distribution. Thus

$$\begin{aligned} \alpha &= P(\text{ type I error }) = P(\text{ rejecting } H_0 \text{ when } H_0 \text{ is true}) \\ &= P(Z \geq c \text{ when } Z \sim N(0,1)) \end{aligned}$$

We find $c = z_{0.05} = 1.645$.

Exercise (8.1) 13

Let $X_1, \ldots, X_n$ denote a random sample from a normal population distribution with a known value of σ.

1. For testing the hypotheses $H_0 : \mu = \mu_0$ versus $H_a : \mu > \mu_0$ (where μ_0 is fixed), show that the test with test statistic $\bar{X}$ and rejection region $\bar{x} \geq \mu_0 + 2.33\sigma/\sqrt{n}$ has significance level $\alpha = 0.01$.
2. Suppose the procedure of part (1.) is used to test $H_0 : \mu \leq \mu_0$ versus $H_a : \mu > \mu_0$. If $\mu_0 = 100, n = 25$, and $\sigma = 5$, what is the probability of type I error when $\mu = 99$? When $\mu = 98$? In general, what can be said about the probability of a type I error when the actual value of μ is less than μ_0? Verify your assertion.

Tests About a Population Mean

We will focus on three cases:

1. A Normal Population with Known σ
2. A Normal Population with Unknown σ with Large Sample
3. A Normal Population with Unknown σ with Small Sample

- In all cases, we assume that $X_1, \ldots, X_n$ represent a random sample of size n from the normal population.
- The null hypothesis in all three cases will state that μ has a particular numerical value, the null value, which we will denote by μ_0.
- Although the assumption that the value of σ is known is rarely met in practice, this case provides a good starting point because of the ease with which general procedures and their properties can be developed.

Case I: A Normal Population with Known σ

When $H_0 : \mu = \mu_0$ is true,

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0, 1)$$

Test statistic value:

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

The following are the rejection regions:

Alternative Hypothesis	Rejection Region for Level α Test
$H_a : \mu > \mu_0$	$z \geq z_\alpha$ (upper-tailed test)
$H_a : \mu < \mu_0$	$z \leq -z_\alpha$ (lower-tailed test)
$H_a : \mu \neq \mu_0$	either $z \geq z_{\alpha/2}$ or $z \leq -z_{\alpha/2}$ (two-tailed test)

Case I: A Normal Population with Known σ

Use of the following sequence of steps is recommended when testing hypotheses about a parameter.

1. Identify the parameter of interest and describe it in the context of the problem situation.
2. Determine the null value and state the null hypothesis.
3. State the appropriate alternative hypothesis.
4. Give the formula for the computed value of the test statistic.
5. State the rejection region for the selected significance level α.
6. 6. Compute any necessary sample quantities, substitute into the formula for the test statistic value, and compute that value.
7. 7. Decide whether H_0 should be rejected, and state this conclusion in the problem context.

Remark: The formulation of hypotheses (Steps 2 and 3) should be done before examining the data.

Case I: A Normal Population with Known σ - Example

A manufacturer of sprinkler systems used for fire protection in office buildings claims that the true average system-activation temperature is 130°.
A sample of $n = 9$ systems, when tested, yields a sample average activation temperature of $131.08^{\circ}F$. If the distribution of activation times is normal with standard deviation $1.5^{\circ}F$, does the data contradict the manufacturers claim at significance level $\alpha = 0.01$?

Case I: A Normal Population with Known σ - Example

A manufacturer of sprinkler systems used for fire protection in office buildings claims that the true average system-activation temperature is 130°. A sample of $n = 9$ systems, when tested, yields a sample average activation temperature of $131.08^\circ F$. If the distribution of activation times is normal with standard deviation $1.5^\circ F$, does the data contradict the manufacturers claim at significance level $\alpha = 0.01$?

1. Parameter of interest: $\mu =$ true average activation temperature.
2. Null hypothesis: $H_0 : \mu = 130$ (null value $= \mu_0 = 130$).
3. Alternative hypothesis: $H_a : \mu \neq 130$ (a departure from the claimed value in either direction is of concern).
4. Test statistic value: $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{\bar{x} - 130}{1.5/\sqrt{n}}$.
5. We reject H_0 if either $z \geq 2.58$ or $z \leq -2.58$.
6. Substituting $n = 9$ and $\bar{x} = 131.08$, we get $z = 2.16$
7. z does not fall in the rejection region. Therefore, H_0 cannot be rejected. The data does not give strong evidence to claim that the true average differs from the design value of 130.

Case II: A Normal Population with unknown σ and Large Sample

When $H_0 : \mu = \mu_0$ is true and n is large enough,

$$Z = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim N(0,1)$$

where S is the sample variance. Test procedure follows the previous case.

Case III: A Normal Population with unknown σ and Small Sample

Knowledge of the test statistics distribution when H_0 is true (the "null distribution") allows us to construct a rejection region for which the type I error probability is controlled at the desired level.

In this case, $T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim N(0,1)$ has a t distribution with $n-1$ degrees of freedom.

We will use of the upper-tail t critical value $t_{\alpha,n-1}$ to specify the rejection region $t \geq t_{\alpha,n-1}$.

$$\begin{aligned} P(\text{type I error}) &= P(H_0 \text{ is rejected when it is true}) \\ &= P(T \geq t_{\alpha,n-1} \text{ when } T \sim t_{n-1}) \\ &= \alpha \end{aligned}$$

Case III: A Normal Population with unknown σ and Small Sample

When $H_0 : \mu = \mu_0$ is true,

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t_{n-1}$$

Test statistic value:

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

The following are the rejection regions:

Alternative Hypothesis	Rejection Region for Level α Test
$H_a : \mu > \mu_0$	$t \geq t_{\alpha,n-1}$
$H_a : \mu < \mu_0$	$t \leq -t_{\alpha,n-1}$
$H_a : \mu \neq \mu_0$	either $t \geq t_{\alpha/2,n-1}$ or $t \leq -t_{\alpha/2,n-1}$

Exercise (8.2) 19

The melting point of each of 16 samples of a certain brand of hydrogenated vegetable oil was determined, resulting in $\bar{x} = 94.32$. Assuming that the distribution of the melting point is normal with $\sigma = 1.20$,

1. Test $H_0 : \mu = 95$ versus $H_a : \mu \neq 95$ using a two-tailed level 0.01 test.
2. If a level 0.01 test is used, what is $\beta(94)$, the probability of a type II error when $\mu = 94$?
3. What value of n is necessary to ensure that $\beta(94) = 0.1$ when $\alpha = 0.01$?

Tests Concerning a Population Proportion

Let p denote the proportion of individuals or objects in a population who possess a specified property (e.g., cars with manual transmissions or smokers who smoke a filter cigarette).

- If an individual or object with the property is labelled a success (S), then p is the population proportion of successes.

- Tests concerning p will be based on a random sample of size n from the population. Provided that n is small relative to the population size, X (the number of S's in the sample) has (approximately) a binomial distribution.

- Furthermore, if n itself is large [$np \geq 10$ and $n(1-p) \geq 10$], both X and the estimator $\hat{p} = X/n$ are approximately normally distributed.

- We first consider large-sample tests based on this latter fact and then turn to the small sample case that directly uses the binomial distribution.

Large-Sample Tests on Proportion

The estimator $\hat{p} = X/n$ is unbiased $[E(\hat{p}) = p]$, has approximately a normal distribution, and its standard deviation is $\sigma_{\hat{p}}\sqrt{p(1-p)/n}$.

- When H_0 is true, $E(\hat{p}) = p_0$ and $\sigma_{\hat{p}} = \sqrt{p_0(1-p_0)/n}$, so $\sigma_{\hat{p}}$ does not involve any unknown parameters.

It then follows that when n is large and H_0 is true, the test statistic

$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$$

has approximately a standard normal distribution.

Large-Sample Tests on Proportion

If the alternative hypothesis is $Ha : p > p_0$ and the upper-tailed rejection region $z \geq z_\alpha$ is used, then

$$\begin{aligned} P(\text{type I error}) &= P(H_0 \text{ is rejected when it is true}) \\ &= P(Z \geq z_\alpha \text{ when } Z \sim N(0,1)) \\ &\approx \alpha \end{aligned}$$

Thus the desired level of significance α is attained by using the critical value that captures area α in the upper tail of the z curve.

Large-Sample Tests on Proportion

Rejection regions for the other two alternative hypotheses, lower-tailed for $H_a : p < p_0$ and two-tailed for $H_a : p \neq p_0$, are justified in an analogous manner.

Null hypothesis: $H_0 : p = p_0$

Test statistic value: $Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$

The following are the rejection regions:

Alternative Hypothesis	Rejection Region for Level α Test
$H_a : p > p_0$	$z \geq z_\alpha$ (upper-tailed)
$H_a : p < p_0$	$z \leq -z_\alpha$ (lower-tailed)
$H_a : p \neq p_0$	either $z \geq z_{\alpha/2}$ or
	either $z \leq -z_{\alpha/2}$ (two-tailed)

These procedures are valid provided that $np_0 \geq 10$ and $n(1 - p_0) \geq 10$.

Large-Sample Tests on Proportion - Example

Natural cork in wine bottles is subject to deterioration, and as a result wine in such bottles may experience contamination.

An article reported that, in a tasting of commercial chardonnays, 16 of 91 bottles were considered spoiled to some extent by cork-associated characteristics.

Does this data provide strong evidence for concluding that more than 15% of all such bottles are contaminated in this way?

Let's carry out a test of hypotheses using a significance level of $\alpha = 0.10$.

Large-Sample Tests on Proportion - Example

1. $p =$ the true proportion of all commercial chardonnay bottles considered spoiled to some extent by cork-associated characteristics.
2. The null hypothesis is $H_0 : p = 0.15$.
3. The alternative hypothesis is $H_a : p > 0.15$, the assertion that the population percentage exceeds 15%.
4. Since $np_0 = 91(.15) = 13.65 > 10$ and $n(1 - p_0) = 91(.85) = 77.35 > 10$, the large-sample z test can be used. The test statistic value is $z = (\hat{p} - 0.15)/\sqrt{(0.15)(0.85)/n}$
5. The form of H_a implies that an upper-tailed test is appropriate: Reject H_0 if $z \geq z_{0.10} = 1.28$.
6. From data, $n = 91$ and $\hat{p} = 16/91 = 0.1758$, from which $z = (0.1758 - 0.15)/\sqrt{(0.15)(0.85)/91} = 0.69$
7. Since $0.69 < 1.28$, z is not in the rejection region. At significance level 0.10, the null hypothesis cannot be rejected.

Small-Sample Tests

Test procedures when the sample size n is small are based directly on the binomial distribution rather than the normal approximation.

Consider the alternative hypothesis $H_a : p > p_0$ and again let X be the number of successes in the sample.

Then X is the test statistic, and the upper-tailed rejection region has the form $x \geq c$. When H_0 is true, X has a binomial distribution with parameters n and p_0, so

$$
\begin{aligned}
P(\text{type I error}) &= P(H_0 \text{ is rejected when it is true}) \\
&= P(X \geq c \text{ when } X \sim Bin(n, p_0)) \\
&= 1 - P(X \leq c - 1 \text{ when } X \sim Bin(n, p_0)) \\
&= 1 - B(c - 1; n, p_0)
\end{aligned}
$$

Small-Sample Tests on Proportion

- It is usually not possible to find a value of c for which $P(\text{type I error})$ is exactly the desired significance level α. Instead, the largest rejection region of the form $\{c, c+1, \ldots, n\}$ satisfying $1 - Bin(c-1 : n, p_0) \leq \alpha$ is used.

Let p' denote an alternative value of $p (p' > p_0)$. When $p = p', X \sim Bin(n, p')$, so

$$
\begin{aligned}
\beta(p') &= P(\text{type II error when } p = p') \\
&= P(X < c \text{ when } X \sim Bin(n, p')) \\
&= B(c-1; n, p')
\end{aligned}
$$

Exercise (8.3) 39

A random sample of 150 recent donations at a certain blood bank reveals that 82 were type A blood. Does this suggest that the actual percentage of type A donations differs from 40%, the percentage of the population having type A blood?
Carry out a test of the appropriate hypotheses using a significance level of 0.01. Would your conclusion have been different if a significance level of 0.05 had been used?

P-Values

It is an alternative way of reaching a conclusion in a hypothesis testing analysis. One advantage is that the P-value provides an intuitive measure of the strength of evidence in the data against H_0.

Definition

The P-value is the probability, calculated assuming that the null hypothesis is true, of obtaining a value of the test statistic at least as contradictory to H_0 as the value calculated from the available sample.

Key points of the definition

- The P-value is a probability.
- This probability is calculated assuming that H_0 is true.
- Beware: The P-value is not the probability that H_0 is true, nor is it an error probability!

P-Values - Example

Urban storm water can be contaminated by many sources, including discarded batteries. When ruptured, these batteries release metals of environmental significance. A sample of 51 Panasonic AAA batteries gave a sample mean zinc mass of $2.06g$ and a sample standard deviation of $0.141g$. Does this data provide compelling evidence for concluding that the population mean zinc mass exceeds $2.0g$?

With μ denoting the true average zinc mass for such batteries, the relevant hypotheses are
$H_0 : \mu = 2.0$ versus $H_a : \mu > 2.0$.

The sample size is large enough to use CLT. The test statistic value is
$z = \frac{\bar{x}-2.0}{s/\sqrt{n}} = \frac{2.06-2.0}{0.141/\sqrt{51}} = 3.04$

The P-value is
$P(Z \geq 3.04 \text{ when } \mu = 2.0) = 1 - P(Z < 3.04 \text{ when } \mu = 2.0) = 0.0012$

P-Values - Example

More generally, the smaller the P-value, the more evidence there is in the sample data against the null hypothesis and for the alternative hypothesis. That is, H_0 should be rejected in favor of H_a when the P-value is sufficiently small.

Decision rule based on the P-value

Select a significance level α. Then

$$\begin{aligned} \text{reject } H_0 \text{ if } P\text{-value} &\leq \alpha \\ \text{do not reject } H_0 \text{ if } P\text{-value} &> \alpha \end{aligned}$$

P-Values - Example

Thus if the P-value exceeds the chosen significance level, the null hypothesis cannot be rejected at that level.

But if the P-value is equal to or less than α, then there is enough evidence to justify rejecting H_0.

In Example 14, we calculated P-value $= 0.0012$. Then using a significance level of 0.01, we would reject the null hypothesis in favor of the alternative hypothesis because $0.0012 < 0.01$.

However, suppose we select a significance level of only 0.001, which requires more substantial evidence from the data before H_0 can be rejected. In this case we would not reject H_0 because $0.0012 > 0.001$.

Exercise (8.4) 52

The paint used to make lines on roads must reflect enough light to be clearly visible at night. Let μ denote the true average reflectometer reading for a new type of paint under consideration. A test of $H_0 : \mu = 20$ versus $H_a : \mu > 20$ will be based on a random sample of size n from a normal population distribution. What conclusion is appropriate in each of the following situations?

1. $n = 15, t = 3.2, \alpha = 0.05$
2. $n = 9, t = 1.8, \alpha = 0.01$
3. $n = 24, t = -0.2$