Chapter 8

Tests of Hypotheses Based on a Single Sample

Jing Zhang
Department of Mathematics and Statistics
DALHOUSIE UNIVERSITY

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

Hypotheses

- The null hypothesis 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.
- The alternative to the null hypothesis $H_0: \theta = \theta_0$ will look like one of the following three assertions:

  1. $H_a: \theta > \theta_0$ 2. $H_a: \theta < \theta_0$ 3. $H_a: \theta \neq \theta_0$

Test Procedures

- Step 1: state H_0 and H_a for the problem
- Step 2: calculate the test statistic by using the sample
- Step 3: find the P-value
- Step 4: make the decision according to the decision rule
- Step 5: draw the conclusion

- A test statistic is a function of the sample data used as a basis for deciding whether H_0 should be rejected. The selected test statistic should discriminate effectively between the two hypotheses. That is, values of the statistic that tend to result when H_0 is true should be quite different from those typically observed when H_0 is not true.
- 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 data.

- A conclusion is reached in a hypothesis testing analysis by selecting a number α, called a significance level (alternatively, level of significance) of the test, that is reasonably close to 0. The significance levels used most frequently in practice are (in order) $\alpha = 0.1, 0.05, 0.01$.
- Decision Rule: H_0 will be rejected in favor H_a if P-value $\leq \alpha$, whereas H_0 will not be rejected (still considered to be plausible) if P-value $> \alpha$.

DALHOUSIE UNIVERSITY

P-value:

- The definition of P-value is obviously somewhat complicated. Here are some important points:

1. The P-value is a probability.
2. This probability is calculated assuming that the null hypothesis is true.
3. To determine the P-value, we must first decide which value of the test statistic are at least as contradictory to H_0 as the value obtained from our sample.
4. The smaller the P-value, the stronger is the evidence against H_0 and in favor of H_a.
5. The P-value is **not** the probability that the null hypothesis is true or that is false, nor is it the probability that an erroneous conclusion is reached.

Errors in Hypothesis Testing

There are two different types of errors that might be made in the course of a statistical hypothesis testing analysis.

- 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 it is false.

- The test procedure that rejects H_0 if P-value $\leq \alpha$ and otherwise does not reject H_0 has $P(Type\ I\ error) = \alpha$. That is, the significance level employed in the test procedure is the probability of a type I error.
- Let β denote the probability of committing a type II error.
- Suppose an experiment or sampling procedure is selected, a sample size is specified, and a test statistic is chosen. Then increasing the significance level α, i,.e., employing a larger type I error probability, results in a smaller value of β for any particular parameter value consistent with H_a.

Z Tests for Hypotheses about a Population Mean

Case I : A Normal Population Distribution with Known σ

Null hypothesis: $H_0: \mu = \mu_0$

Test statistic: $Z_{test} = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$

Alternative Hypothesis and P-value

$H_a: \mu > \mu_0 \quad P - value = P(Z \geq Z_{test})$

$H_a: \mu < \mu_0 \quad P - value = P(Z \leq Z_{test})$

$H_a: \mu \neq \mu_0 \quad P - value = 2 \times P(Z \geq |Z_{test}|)$

Case II: Large-Sample Test (without requiring either a normal distribution or known σ)

Test statistic: $Z_{test} = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$

Example 8.1

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, yield a sample average activation temperature of 131.08. If the distribution of activation temperature is normal with standard deviation 1.5, does the data contradict the manufacturer’s claim at significance level $\alpha=0.01$?

Solution:

Step 1: $H_0: \mu = 130$

$H_a: \mu \neq 130$

Step 2: Test statistic: $Z_{test} = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{131.08 - 130}{1.5/\sqrt{9}} = 2.16$

Step 3: Find the P-value:

$P - value = 2 \times P(Z \geq |Z_{test}|) = 2 \times P(Z \geq |2.16|)$

$= 2 \times P(Z \geq 2.16) = 2 \times \left(1 - \Phi(2.16)\right) = 2 \times (0.0154) = 0.0308$

Step 4: Make the decision:

Since P-value = 0.0308 $>$ 0.01(α), H_0 cannot be rejected at significance level 0.01.

Step 5: Draw the conclusion:

The data does not give strong support to the claim that the true average differs from the design value of 130.

Decision Rules:

- P-value approach:

If P-value $\leq \alpha$, reject H_0.

$H_a: \mu > \mu_0 \quad P - value = P(Z \geq Z_{test})$

$H_a: \mu < \mu_0 \quad P - value = P(Z \leq Z_{test})$

$H_a: \mu \neq \mu_0 \quad P - value = 2 \times P(Z \geq |Z_{test}|)$

- Critical Value approach:

$H_a: \mu > \mu_0$ If $Z_{test} \geq Z_{\alpha}$, then reject H_0.

$H_a: \mu < \mu_0$ If $Z_{test} \leq -Z_{\alpha}$, then reject H_0.

$H_a: \mu \neq \mu_0$ If $Z_{test} \geq +Z_{\frac{\alpha}{2}}$ or $Z_{test} \leq -Z_{\frac{\alpha}{2}}$, then reject H_0.

Example 8.1

$H_0: \mu = 130$

$H_a: \mu \neq 130$

Test statistic: $Z_{test} = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{131.08 - 130}{1.5/\sqrt{9}} = 2.16$

Decision Rule:

If $Z_{test} \geq +Z_{\frac{\alpha}{2}}$ or $Z_{test} \leq -Z_{\frac{\alpha}{2}}$, then reject H_0.

$Z_{\alpha/2} = Z_{0.01/2} = Z_{0.005} = 2.576$.

Make the decision:

Since $Z_{test} < +2.576$, then don't reject H_0.

Draw the conclusion:

The data does not give strong support to the claim that the true average differs from the design value of 130.

β, the Probability of a type II error

The Z tests with known σ are among the few in statistics for which there are simple formulas available for β, the probability of a type II error.

$H_a\!: \mu > \mu_0$ $\qquad$ $\beta(\mu') = \Phi(Z_\alpha + \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}})$

$H_a\!: \mu < \mu_0$ $\qquad$ $\beta(\mu') = 1 - \Phi(-Z_\alpha + \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}})$

$H_a\!: \mu \neq \mu_0$ $\qquad$ $\beta(\mu') = \Phi\left(Z_{\frac{\alpha}{2}} + \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}}\right) - \Phi\left(-Z_{\frac{\alpha}{2}} + \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}}\right)$

Where $\Phi(z) = \textit{the standard normal cdf}$ and μ'= the true population mean

Sample Size Determination

The sample size n for which a level α test also has $\beta(\mu') = \beta$ at the alternative value μ' is

$n = [\frac{\sigma(Z_\alpha + Z_\beta)}{\mu_0 - \mu'}]^2$ for a one-tailed (upper or lower) test.

$n = [\frac{\sigma(Z_{\alpha/2} + Z_\beta)}{\mu_0 - \mu'}]^2$ for a two-tailed test (an approximate solution)

Example 8.2

Let μ denote the true average tread life of a certain type of tire. Consider testing $H_0: \mu = 30{,}000$ versus $H_a: \mu > 30{,}000$ based on a sample of size n = 16 from a normal population distribution with $\sigma = 1500$.

a) Find the probability of type II error when $\mu = 31{,}000$ with $\alpha = 0.01$.

b) Find the sample size n for which a level 0.01 test also have $\beta(31{,}000) = 0.1$.

Solutions:

a) For the upper tailed test,

$$\beta(\mu') = \Phi(Z_\alpha + \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}})$$

$$\mu_0 = 30{,}000\ \mu' = 31{,}000\ and\ \sigma = 1500.$$

$$Z_\alpha = Z_{0.01} = 2.33$$

$$\beta(31000) = \Phi\left(2.33 + \frac{30{,}000 - 31{,}000}{\frac{1500}{\sqrt{16}}}\right) = \Phi(-0.34) = 0.3669$$

b) $\beta(31{,}000) = 0.1$ and $Z_{0.1} = 1.28$ (Found this value from Z table).

$$n = [\frac{1500(2.33 + 1.28)}{30{,}000 - 31{,}000}]^2 = (-5.42)^2 = 29.32$$

The sample size must be an integer, so n = 30 tires should be used.

The One-Sample T test

When n is small, the Central Limit Theorem (CLT) can no longer be invoked to justify the use of a large-sample test.

The One-Sample T test

Null hypothesis: $H_0: \mu = \mu_0$

Test statistic: $t_{test} = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$ t_{test} *has a t distribution with* $n-1$ *df*.

Alternative Hypothesis and P-value

$H_a: \mu > \mu_0 \quad P-value = P(t_{n-1} \geq t_{test})$

$H_a: \mu < \mu_0 \quad P-value = P(t_{n-1} \leq t_{test})$

$H_a: \mu \neq \mu_0 \quad P-value = 2 \times P(t_{n-1} \geq |t_{test}|)$

Decision Rules:

- P-value approach:

If P-value $\leq \alpha$, reject H_0.

$H_a: \mu > \mu_0 \quad P - value = P(t_{n-1} \geq t_{test})$

$H_a: \mu < \mu_0 \quad P - value = P(t_{n-1} \leq t_{test})$

$H_a: \mu \neq \mu_0 \quad P - value = 2 \times P(t_{n-1} \geq |t_{test}|)$

- Critical Value approach:

$H_a: \mu > \mu_0$ If $t_{test} \geq t_\alpha$, then reject H_0.

$H_a: \mu < \mu_0$ If $t_{test} \leq -t_\alpha$, then reject H_0.

$H_a: \mu \neq \mu_0$ If $t_{test} \geq +t_{\frac{\alpha}{2}}$ or $t_{test} \leq -t_{\frac{\alpha}{2}}$, then reject H_0.

df = n-1

P-value and the T Distribution

- The format of the t distribution table provided in most statistics textbooks does not have sufficient detail to determine the exact P-value for a hypothesis test.
- However, we can still use the t distribution table to identify a range for the P-value.
- An advantage of computer software packages is that the computer output will provide the P-value for the t distribution.

Example 8.3

A State Highway Patrol periodically samples vehicle speeds at various locations on a particular roadway. The sample of vehicle speeds is used to test the hypothesis

$$H_0\colon \mu = 65.$$
$$H_a\colon \mu > 65.$$

The locations where H_0 is rejected are deemed the best locations for radar traps. At Location F, a sample of 16 vehicles shows a mean speed of 67.4 mph with a standard deviation of 4.2 mph. Use α=0.05 to test he hypothesis.

Solution:

Step 1:

H_0: $\mu = 65$.

H_a: $\mu > 65$.

Step 2: Test statistic: $t_{test} = \frac{\bar{X}-\mu_0}{s/\sqrt{n}} = \frac{67.4-65}{4.2/\sqrt{16}} = 2.286$

Step 3: Find the P-value

$$P - value = P(t_{n-1} \geq t_{test}) = P(t_{15} \geq 2.286) = (0.01{,}0.025)$$

For t = 2.286, P-value must be less than 0.025(for t=2.131) and greater than 0.01 (for t=2.602).

Step 4: Make the decision:

P-value $\leq \alpha(0.05)$, we reject H_0.

Step 5: Draw the conclusion:

We have enough evidence to conclude that the mean speed of vehicles at Location F is greater than 65 mph.

Tests Concerning a Population Proportion

Null hypothesis: $H_0\colon p = p_0$

Test statistic: $Z_{test} = \dfrac{\bar{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$

- P-value approach (If P-value $\leq \alpha$, reject H_0)

$H_a\colon p > p_0 \quad P - value = P(Z \geq Z_{test})$

$H_a\colon p < p_0 \quad P - value = P(Z \leq Z_{test})$

$H_a\colon p \neq p_0 \quad P - value = 2 \times P(Z \geq |Z_{test}|)$

- Critical Value approach:

$H_a\colon p > p_0$ If $Z_{test} \geq Z_\alpha$, then reject H_0.

$H_a\colon p < p_0$ If $Z_{test} \leq -Z_\alpha$, then reject H_0.

$H_a\colon p \neq p_0$ If $Z_{test} \geq +Z_{\frac{\alpha}{2}}\ or\ Z_{test} \leq -Z_{\frac{\alpha}{2}}$, then reject H_0

β, the Probability of a type II error

$H_a\!: p > p_0$

$$\beta(p') = \Phi[\frac{p_0 - p' + Z_\alpha\sqrt{p_0(1-p_0)/n}}{\sqrt{p'(1-p')/n}}]$$

$H_a\!: p < p_0$

$$\beta(p') = 1 - \Phi[\frac{p_0 - p' - Z_\alpha\sqrt{p_0(1-p_0)/n}}{\sqrt{p'(1-p')/n}}]$$

$H_a\!: p \neq p_0$

$$\beta(p') = \Phi\left[\frac{p_0 - p' + Z_{\frac{\alpha}{2}}\sqrt{\frac{p_0(1-p_0)}{n}}}{\sqrt{\frac{p'(1-p')}{n}}}\right] - \Phi\left[\frac{p_0 - p' - Z_{\frac{\alpha}{2}}\sqrt{\frac{p_0(1-p_0)}{n}}}{\sqrt{\frac{p'(1-p')}{n}}}\right]$$

Sample Size Determination

The sample size n for which the level α test also satisfies $\beta(p') = \beta$ is

$$n = \left[\frac{Z_\alpha \sqrt{p_0(1-p_0)} + Z_\beta \sqrt{p'(1-p')}}{p' - p_0} \right]^2$$ for one-tailed test

$$n = \left[\frac{Z_{\alpha/2} \sqrt{p_0(1-p_0)} + Z_\beta \sqrt{p'(1-p')}}{p' - p_0} \right]^2$$ for two-tailed test

Example 8.4

Student use of cell phones during class is perceived by many faculty to be an annoying but perhaps harmless distraction. However, the use of a phone to text during an exam is a serious breach of conduct. The article "The Use and Abuse of Cell Phones and Text Messaging During Class: A Survey of College Students" reported that 27 out of the 267 students in a sample admitted to doing this. Can it be concluded at significance level 0.001 that more than 5% of all students in the population sampled had texted during an exam?

Solution:

Step 1: $H_0: p = 0.05$

$H_a: p > 0.05$

Step 2: Test statistic:

$$Z_{test} = \frac{\bar{p} - p_0}{\sqrt{p_0(1 - p_0)/_n}} = \frac{\frac{27}{267} - 0.05}{\sqrt{0.05(1 - 0.05)/267}} = 3.84$$

Step 3: The P-value for this upper-tailed test is

$P - value = P(Z \geq Z_{test}) = P(Z > 3.84) = 1 - \Phi(3.84) < 1 - \Phi(3.49)$
$= 0.0003$

Step 4: Make the decision:

P-value < 0.0003 < $\alpha(0.001)$, therefore reject H_0.

Step 5: Draw the conclusion:

We have enough evidence to conclude that the population percentage of students who text during an exam exceeds 5% is very compelling.

Example 8.5

A package-delivery service advertises that at least 90% of all packages brought to its office by 9 am for delivery in the same city are delivered by noon that day. Let p denote the true proportion of such packages that are delivered as advertised and consider that hypotheses $H_0: p = 0.9$ versus $H_a: p < 0.9$. If only 80% of the packages are delivered as advertised, how likely is it that a level 0.01 test based on n = 225 packages will detect such a departure from H_0? What should the sample size be to ensure that $\beta(0.8) = 0.01$?

Solution:

With $\alpha = 0.01, p_0 = 0.9, p' = 0.8, n = 225\ and\ Z_{0.01} = 2.33$

$$\beta(p') = 1 - \Phi[\frac{p_0 - p' - Z_\alpha\sqrt{p_0(1-p_0)/n}}{\sqrt{p'(1-p')/n}}]$$

$$\beta(0.8) = 1 - \Phi\left[\frac{0.9 - 0.8 - 2.33\sqrt{\frac{0.9(1-0.9)}{225}}}{\sqrt{\frac{0.8(1-0.8)}{225}}}\right]$$

$$= 1 - \Phi(2.00) = 0.0228$$

Thus the probability that H_0will be rejected using the test when p=0.8 is 0.9772; roughly 98% of all samples will result in correct rejection of H_0.

- Using $Z_\alpha = Z_\beta = Z_{0.01} = 2.33$ in the sample size formula yields

$$n = \left[\frac{Z_\alpha\sqrt{p_0(1-p_0)} + Z_\beta\sqrt{p'(1-p')}}{p'-p_0}\right]^2$$

$$n = [\frac{2.33\sqrt{0.9(1-0.9)} + 2.33\sqrt{0.8(1-0.8)}}{0.8-0.9}]^2 \approx 266$$

The Power of the Test

1-β can be expressed as the **Power** of the test. A small value of β (close to 0) is equivalent to large power (near 1). A *powerful* test is one that has high power and therefore good ability to detect when the null hypothesis is false

The Relationship between Confidence Intervals and hypothesis Tests

Let $(\hat{\theta}_L, \hat{\theta}_U)$ be a confidence interval for θ with confidence level $100(1-\alpha)\%$. Then a test of $H_0\colon \theta = \theta_0$ versus $H_a\colon \theta \neq \theta_0$ with significance level a rejects the null hypothesis if the null value θ_0 is not included in the CI and does not reject H_0 if the null value does lie in the CI.