STA255: Statistical Theory

Chapter 10: Hypothesis Testing

Summer 2017

Hypotheses

A statistical hypothesis is a statement about the one or more population parameters.

- A null hypothesis H_0 is a claim (or statement) about a population parameter that is assumed to be true until it is declared false.
  It takes the form $H_0 : \theta = \theta_0$
- The Alternative hypothesis H_1 (or H_a) a claim about a population parameter

Hypotheses

The alternative hypothesis H_1 will contain either a greater than sign (one-tailed test), a less than sign (one tailed test), or a not equal to sign (two- tailed test).

- Greater than ($>$): results if the problem says increases, improves, better, result is higher, etc.
- Less than ($<$): results if the problem says decreases, reduces, worse than, result is lower, etc.
- Not equal to ($\neq$): results if the problem says different from, no longer the same, changes, etc.

Example: One-sample t-test

An air freight company wishes to test whether or not the mean weight of parcels shipped on a particular root exceeds 10 pounds. A random sample of 49 shipping orders was examined and found to have average weight of 11 pounds. Assume that $\sigma = 2.8$ pounds. Use $\alpha = 0.05$

Hypotheses: $H_0 : \mu = 10$ vs. $H_1 : \mu > 10$

Test Statistic (TS)

The test statistic is some quantity calculated from the sample data that we have collected. It is used to determine the strength of the evidence against H_0.

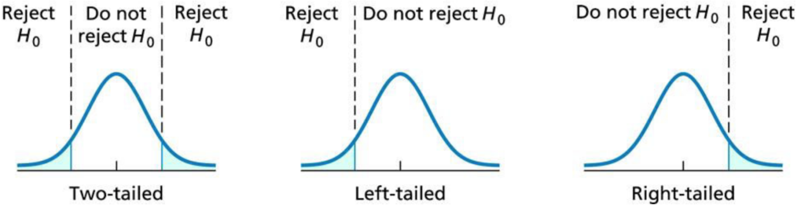
In the previous Example:

Test statistic:

$$z = \frac{\hat{\theta} - \theta}{\sigma_{\hat{\theta}}} = \frac{\bar{y} - \mu}{\sigma/\sqrt{n}} = \frac{11 - 10}{2.8/\sqrt{49}} = 2.5$$

Rejection Region (RR)

- A rejection region, the set of all test statistic values for which H_0 will be rejected (H_0 rejected if the test statistic value falls in this region.)



- The critical region is established based on α [cut- off]. α is called the level of the test.
- **In the Example:** $\alpha = 0.05$. That we reject H_0 if the computed test statistic ($z = 2.5$) is greater than 1.65.

Rejection Region (RR)

In the Example:

- **Hypotheses:** $H_0 : \mu = 10$ vs. $H_1 : \mu > 10$
- **Test statistic:** $z = \frac{\hat{\theta}-\theta}{\sigma_{\hat{\theta}}} = \frac{\bar{y}-\mu}{\sigma/\sqrt{n}} = \frac{11-10}{2.8/\sqrt{49}} = 2.5$
- **RR:** If $\alpha = 0.05$, then

  $RR = \{z : z > z_\alpha\} = \{z : z > z_{0.05}\} = \{z : z > 1.65\}$

That is, we reject H_0 if the computed test statistic $(z = 2.5)$ is greater than 1.65.

P-value

The decision to reject or fail to reject the null hypothesis is based on the p-value of the test.

- The p-value is the probability, assuming the null hypothesis is true, of observing a test statistic value as extreme or more extreme than the value observed.
- We compare the P-value with a fixed value (cut-off), called the significance level or the size of the test (α). Typical values of α used are 0.05 and 0.01.
- The smaller the p-value is, the stronger the evidence against H_0 provided by the data.

P-value

The p-value is the lowest level (of significance) at which the observed value of the test statistic is significant.

In the Example:

$$P - value = P(Z > 2.5) = 1 - (Z < 2.5) = 0.9938 = 0.0062.$$

[In R: 1-pnorm(2.5,mean=0,sd=1)]

Decision

When we carry out the test we assume H_0 is true. Hence the test will result in one of two decisions.

- Reject H_0: Hence we have sufficient evidence to conclude that the alternative hypothesis is true. Such a test is said to be significant.
- Fail to reject H_0: Hence we do not have sufficient evidence to conclude that the alternative hypothesis is true. Such a test is said to be insignificant.

Decision

Thus, the null hypothesis H0 is rejected if

(a) The calculated test statistic fall in the RR [RR approach].
OR

(b) The p-value $< \alpha$, [p-value approach].

Notes:

(1) You need to consider **ONLY** one approach to do the test.

(2) A third approach based on CI can be used, as will be seen in the next examples. But this approach is not recommended.

Decision

In the Example:

- **RR approach**: $RR = z : z > 1.65$

  Since the computed $z = 2.5 > 1.65$, we reject H_0
- **P-value approach**: since p-value $= 0.0062 < 0.05$, we reject.

Conclusion: There is a sufficient evidence to support the claim that mean weight of parcels shipped exceeds 10 pounds.

Decision

- **Note: 95% one-sided confidence bound** $(H_1 : \mu > 10)$:
  - $\bar{y} - z_\alpha \frac{\sigma}{\sqrt{n}} = 11 - 1.65 \frac{2.8}{\sqrt{49}} = 10.34$
  - Thus, the CI: $(10.34, \infty)$.
  - This interval excludes 10.
  - So we reject H0
- In general, when the hypothesized value belongs to the CI, we don't reject H_0. Otherwise, we reject H_0.

General Testing Procedure

- State the null and alternative hypothesis.
- Carry out the experiment, collect the data, verify the assumptions, and compute the value of the test statistic.
- Calculate the rejection region or the p-value.
- Make a decision on the significance of the test (reject or fail to reject H_0). Make a conclusion statement in the words of the original problem.

Example

Let's assume that we are interested in testing if the mean weight of parcels is different than 10 pounds. Use $\alpha = 0.05$.

(1) **Hypotheses:** $H_0 : \mu = 10$ *vs.* $H_1 : \mu \neq 10$

(2) **Test statistic:** $z = \frac{\bar{y} - \mu}{\sigma / \sqrt{n}} = \frac{11 - 10}{2.8/\sqrt{49}} = 2.5$

(3) **RR:** Reject H_0 if $z > 1.96$ or $z < -1.96$. Otherwise, don't reject.

(4) **Decision:** since $z = 2.5 > 1.96$, we reject H_0

(5) **Conclusion:** There is a sufficient evidence to support the claim that mean weight of parcels shipped is different than 10 pounds.

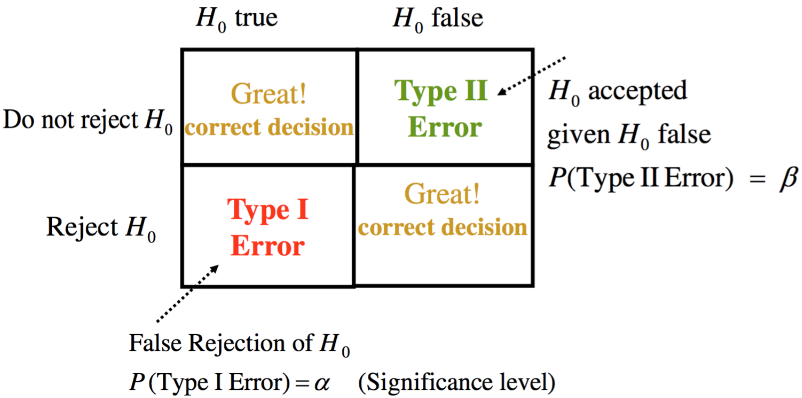
Example

Notes:

- $P - value = 2 * P_{H_0}(Z > 2.5) = 2(0.0062) = 0.0124$
- $CI : \bar{y} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = (10.216, 11.784)$.
- If we increase n considerably, then the test statistic is increased. So it become more likely to get an extreme value (and reject).

Errors

Four scenarios when making a decision based on a sample



Error Types

- A type I error occurs when the null hypothesis (H_0) is rejected when in fact H_0 is true.

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

- A type II error occurs when we fail to reject the null hypothesis (H_0) when in fact H_0 is false.

$$P(\text{Type II error}) = P(\text{Don't reject } H_0 \text{ when } H_0 \text{ is false}) = \beta$$

- The power of the test is the probability of NOT making Type II error

$$\text{Power} = 1 - \beta = P(\text{Reject } H_0 | H_0 \text{ is false})$$

Thus, the power of a test is the probability of making the correct decision if H_1 is true.

Example

Consider the previous example:

$$H_0 : \mu = 10 \ vs. \ H_1 : \mu > 10$$

If H_0 is false and the true population mean is 12. Find the power and β.

Solution:

Example

Suppose that Y is a Binomial random variable $n = 100$ and p unknown and we want to test $H_0 : p = 0.4$ against $H_1 : p > 0.4$. We reject H_0 if $Y \geq 48$.

Hint: use normal approximations for this question.

(a) Find the significance level α for this test

Solution:

Example

(b) Suppose that the true value of p is 0.55. Find the power of the test for this value of p.

Solution

Tests about Population Mean: Single Sample

- **Goal:** We hypothesize that the population mean (μ) equals some value μ_0, and state the alternative hypothesis that we wish to prove is true.
- Case I: Normal Population with known σ
  - We use one-sample z-test.
  - Not a realistic case.
- Case II: Large-Sample Test
  - Large-sample tests can be used when the sample size n is large ($n \geq 30$), but the population standard deviation σ is unknown.
  - We can use z-test or t-test z and t values are very close for large samples sizes.

Tests about Population Mean: Single Sample

Case III: Normal Population, small sample and unknown σ

- If the value of standard deviation σ is unknown, then the test statistic follows t-distribution with degrees of freedom $df = n - 1$.

- Normality must be tested before applying this procedure using, for example, normal probability plot or normality test [not required in this course].

One-sample t-test (when σ is unknown)

(1) **Hypotheses**

- $H_0 : \mu = \mu_0$
- $H_1 : \mu > \mu_0, \quad H_1 : \mu < \mu_0$ or $H_1 : \mu \neq \mu_0$

(2) **Test Statistic**

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

(3) **P-value/ Critical region**

H_1	P-value	Critical Region
$\mu > \mu_0$	$P(T_{n-1} > t)$	$t > t_{\alpha,\, n-1}$
$\mu < \mu_0$	$P(T_{n-1} < t)$	$t < -t_{\alpha,\, n-1}$
$\mu \neq \mu_0$	$2P(T_{n-1} > \lvert t \rvert)$	$t < -t_{\alpha/2,\, n-1}$ or $t > t_{\alpha/2,\, n-1}$

Example

An air freight company wishes to test whether or not the mean weight of parcels shipped on a particular root exceeds 10 pounds. A random sample of 20 shipping orders was examined and found to have average weight of 11 pounds. And s=2.8 pounds. Use $\alpha = 0.05$.

(1) **Hypotheses:** $H_0 : \mu = 10$ vs. $H_1 : \mu > 10$

(2) **Test statistic:** $(df = 20 - 1 = 19)$

$$t = \frac{11 - 10}{2.8/\sqrt{20}} = 1.597$$

Example

(3) **RR:** Reject H_0 if $t > t_{0.05} = 1.729$. Otherwise, don't reject H_0.

(4) **Decision:** $t = 1.597 \not> 1.729$, we do not reject H_0.

(5) **Conclusion:** There is no sufficient evidence to support the claim that mean weight of parcels shipped on a particular root exceeds 10 pounds.

Example

The life in hours of a battery is known to be approximately normally distributed. A random sample of 10 batteries has a mean life of 40.5 hours and a standard deviation of 1.25 hours. Is there evidence to support the claim that battery life exceeds 40 hours? Use $\alpha = 0.05$.

(1) **Hypotheses:** $H_0 : \mu = 40$ vs. $H_1 : \mu > 40$

(2) **Test statistic:** $(df = 10 - 1 = 9)$

$$t = \frac{\bar{y} - \mu_0}{\sigma/\sqrt{n}} = \frac{40.5 - 40}{1.25/\sqrt{10}} = 1.26$$

Example

(3) **RR:** Reject H_0 if $t > t_{0.05} = 1.833$. Otherwise, don't reject H_0.

(4) **Decision:** $t = 1.26 \not> 1.833$, we do not reject H_0

(5) **Conclusion:** There is no sufficient evidence to support the claim that battery life exceeds 40 hours.

Comparing Two Means

Assumptions:

(1) Suppose we have two independent simple random samples. We are interested in making statistical inferences about the difference in the population means: $\mu_1 - \mu_2$

(2)
- Either both populations are normally distributed:

$$Y_1 \ N(\mu_1, \sigma_1^2) \text{ and } Y_2 \ N(\mu_2, \sigma_2^2)$$

- The populations are possibly non-normal but both sample sizes are large enough such that the central limit theorem applies.

(3) The population standard deviations σ_1 and σ_2 are unknown (more realistic case).

Pooled Test for Comparing Two Means

When σ_1 and σ_2 are unknown and assumed equal then the statistic

$$T = \frac{(\bar{Y}_1 - \bar{Y}_2) - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

follows a t distribution with degrees of freedom (df) equal to $n_1 + n_2 - 2$ where the pooled variance s_p^2 is given by

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

2-Sample t-test: Hypotheses

We hypothesize that the difference between the population means equals some specified value D_0 (=0 in most cases) and want to test whether this value is reasonable or whether the alternative is true.

Null Hypothesis: $H_0 : \mu_1 - \mu_2 = D_0$

Alternative Hypothesis:

- $H_1 : \mu_1 - \mu_2 > D_0$
- $H_1 : \mu_1 - \mu_2 < D_0$
- $H_1 : \mu_1 - \mu_2 \neq D_0$

2-Sample pooled t-test: Test Statistic

If the assumptions are satisfied then the test statistic depends on the equality of variances

assumptions:

$$T = \frac{\bar{Y}_1 - \bar{Y}_2 - D_0}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

If we have $D_0 = 0$, then the test statistic is:

$$T = \frac{\bar{Y}_1 - \bar{Y}_2}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Example

A researcher was interested in comparing the amount of time spent browsing internet by women and by men. Independent samples of 14 women and 17 men were selected and each person was asked how many hours he or she had spent browsing the internet during the previous week. The summary statistics are as follows:

	Sample size	Mean	Standard deviation
Men	17	16.9	4.7
Women	14	11.3	4.4

Do the data provide sufficient evidence to conclude that mean time for women is less than mean time for men? Perform a pooled t-test at the 5% significance level.

Example

Solution:

Example

Example

Two different methods (A and B) have been devised to reduce the time spent in transferring materials from one location to another. Each approach is tried several times, and the times to completion (in hours) are recorded below:

Method A: 8.3 7.1 7.8 8.9 8.7 7.2 6.9 6.7
Method B: 7.8 8.1 8.3 8.5 7.6 8.6 7.4

The mean and standard deviation for each method are also calculated:
$\bar{y}_A = 7.70, \bar{y}_B = 8.04, s_A = 0.850, s_B = 0.458.$

Example

(a) Assuming that the data for each method are normally distributed with the same variance, construct a 95% confidence interval for the difference in mean times to completion between methods A and B.

Solution:

Example

(b) Is there evidence that the mean time to completion differs for the two methods? Test the null hypothesis of no difference versus the two- sided alternative using a significance level $\alpha = 0.05$. (Assume as in part (a) that the data are normally distributed with equal variances).

Solution:

Example