# Chapter 9

## Inferences Based on Two Samples

Jing Zhang
Department of Mathematics and Statistics
DALHOUSIE UNIVERSITY

# Z Tests and Confidence Intervals for a Difference Between Two Population Means

- Basic Assumptions:

*1)* $X_1, X_2, \dots, X_m$ is a random sample from a distribution with mean $\mu_1$ and variance $\sigma_1^2$.

*2)* $Y_1, Y_2, \dots, Y_n$ is a random sample from a distribution with mean $\mu_2$ and variance $\sigma_2^2$.

3) The X and Y samples are independent of one another

# Case 1:Test Procedures for Normal Populations with Known Variances

Null hypothesis: $H_0: \mu_1 - \mu_2 = \Delta_0$

Test statistic value: $Z_{test} = \frac{(\bar{x} - \bar{y}) - \Delta_0}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}}$

Make the decision

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

$H_a: \mu_1 - \mu_2 > \Delta_0 \qquad P - value = P(Z \geq Z_{test})$

$H_a: \mu_1 - \mu_2 < \Delta_0 \qquad P - value = P(Z \leq Z_{test})$

$H_a: \mu_1 - \mu_2 \neq \Delta_0 \qquad P - value = 2 \times P(Z \geq |Z_{test}|)$

- Critical value approach

$H_a: \mu_1 - \mu_2 > \Delta_0$ If $Z_{test} \geq Z_\alpha$, then reject $H_0$.

$H_a: \mu_1 - \mu_2 < \Delta_0$ If $Z_{test} \leq -Z_\alpha$, then reject $H_0$.

$H_a: \mu_1 - \mu_2 \neq \Delta_0$ If $Z_{test} \geq +Z_{\frac{\alpha}{2}}\ or\ Z_{test} \leq -Z_{\frac{\alpha}{2}}$ , then reject $H_0$.

# Example 9.1

Analysis of a random sample consisting of $\mathrm{m} = 20$ specimens of cold-rolled steel to determine yield strengths results in a sample average strength of $\bar{x} = 29.8\ ksi$. A second random sample of $\mathrm{n} = 25$ two-sided galvanized steel specimens gave a sample average strength of $\bar{y} = 34.7\ ksi$. Assuming that the two yield-strength distributions are normal with $\sigma_1 = 4.0$ and $\sigma_2 = 5.0$, does the data indicate that the corresponding true average yield strengths $\mu_1$ and $\mu_2$ are different? Let’s carry out a test at significance level $\alpha = 0.01$

# Solution:

$H_0: \mu_1 - \mu_2 = 0$

$H_a: \mu_1 - \mu_2 \neq 0$

$$Z_{test} = \frac{(\bar{x} - \bar{y}) - \Delta_0}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} = \frac{29.8 - 34.7 - 0}{\sqrt{\frac{4^2}{20} + \frac{5^2}{25}}} = -3.66$$

- P-value approach

$P - value = 2 \times P(Z \geq |Z_{test}|) = 2 \times P(Z \geq 3.66) \approx 2(0) = 0$

Since P-value$\approx 0 \leq 0.01(\alpha), reject\ H_0$.

- Critical value approach

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

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

Since $Z_{test} = -3.66 \leq -Z_{\frac{\alpha}{2}}\ (-2.576), reject\ H_0$.

The sample data strongly suggests that the true average yield strength for cold-rolled steel differs from that for galvanized steel.

# $\beta, \beta(\Delta') = P(type\ II\ error\ when\ \mu_1 - \mu_2 = \Delta')$

- $H_a: \mu_1 - \mu_2 > \Delta_0$

$$\beta(\Delta') = \Phi(z_\alpha - \frac{\Delta' - \Delta_0}{\sigma})$$

- $H_a: \mu_1 - \mu_2 < \Delta_0$

$$\beta(\Delta') = 1 - \Phi(-z_\alpha - \frac{\Delta' - \Delta_0}{\sigma})$$

- $H_a: \mu_1 - \mu_2 \neq \Delta_0$

$$\beta(\Delta') = \Phi\left(z_{\alpha/2} - \frac{\Delta' - \Delta_0}{\sigma}\right) - \Phi\left(-z_{\alpha/2} - \frac{\Delta' - \Delta_0}{\sigma}\right)$$

Where $\sigma = \sigma_{\bar{X}-\bar{Y}} = \sqrt{(\sigma_1^2/m) + (\sigma_2^2/n)}$

# Example 9.2 (9.1 continued)

Suppose that when $\mu_1$ and $\mu_2$ (the true average yield strengths for the two types of steel) differ by as much as 5, the probability of detecting such a departure from $H_0$ (the power of the test) should be 0.90. Does a level 0.01 test with sample size m = 20 and n = 25 satisfy this condition?

Solution: $\sigma = \sigma_{\bar{X}-\bar{Y}} = \sqrt{(\sigma_1^2/m) + (\sigma_2^2/n)} = 1.34$

$$\beta(\Delta') = \Phi\left(z_{\alpha/2} - \frac{\Delta' - \Delta_0}{\sigma}\right) - \Phi\left(-z_{\frac{\alpha}{2}} - \frac{\Delta' - \Delta_0}{\sigma}\right)$$

$$\beta(5) = \Phi\left(2.576 - \frac{5-0}{1.34}\right) - \Phi\left(-2.576 - \frac{5-0}{1.34}\right) = \Phi(-1.15) - \Phi(-6.31)$$
$$= 0.1251$$

Thus the power is 1-0.1251=0.8749. Because this is somewhat less than 0.9, slightly larger sample size should be used.

# Confidence Intervals for $\mu_1 - \mu_2$

$100(1-\alpha)\%$ confidence Interval for $\mu_1 - \mu_2$

$$(\bar{x} - \bar{y}) \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}$$

# Example 9.3 (9.1 continued)

Find the 99% CI for the true difference between $\mu_1$and $\mu_2$.

Solution:

$$(29.8 - 34.7) \pm 2.576(1.34) = -4.9 \pm 3.45$$
$$= (-8.35, -1.45)$$

We are 99% confident that the true difference is between -8.35 and -1.45.

# Case II: Large-Sample Tests and CI

The assumption of normal population distributions and known values of $\sigma_1$ and $\sigma_2$ are fortunately unnecessary when both sample sizes are sufficiently large ($m > 40 \ and \ n > 40$).

Test statistic value: $Z_{test} = \frac{(\bar{x}-\bar{y})-\Delta_0}{\sqrt{\frac{s_1^2}{m}+\frac{s_2^2}{n}}}$

$100(1-\alpha)\%$ confidence Interval for $\mu_1 - \mu_2$

$$(\bar{x}-\bar{y}) \pm z_{\alpha/2}\sqrt{\frac{s_1^2}{m}+\frac{s_2^2}{n}}$$

# The Two-Sample t Test and Confidence Interval

- Assumptions:

Both population distributions are normal, so that $X_1, X_2, \ldots X_m$ is a random sample from a normal distribution and so is $Y_1, Y_2, \ldots Y_n$ (with the X's and Y's independent of one another.

- Theorem:

$$T = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}}$$

has approximately a t distribution with df v estimated from the data by

$$v = \frac{\left(\frac{s_1^2}{m} + \frac{s_2^2}{n}\right)^2}{\frac{(s_1^2/m)^2}{m-1} + \frac{(s_2^2/n)^2}{n-1}}$$

(round v down to the nearest integer)

- The two-sample t confidence interval for $\mu_1 - \mu_2$ with confidence level $100(1-\alpha)\%$ is then

$$(\bar{x} - \bar{y}) \pm t_{\alpha/2,v} \sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}$$

- The two-sample t test for testing $H_0: \mu_1 - \mu_2 = \Delta_0$ is as follows:

Test statistic value: $t_{test} = \frac{(\bar{x}-\bar{y})-\Delta_0}{\sqrt{\frac{s_1^2}{m}+\frac{s_2^2}{n}}}$

Assumptions: Both population distributions are normal, and the two random samples are selected independently of one another.

# Example 9.2

The deterioration of many municipal pipeline networks across the country is a growing concern. One technology proposed for pipeline rehabilitation uses a flexible liner threaded through existing pipe. The article “Effect of Welding on a High-Density Polyethylene Liner” reported the following data on tensile strength (psi) of liner specimens both when a certain fusion process was used and when this process was not used. (Assume that the tensile strength distributions under the two conditions are both normal)

| No fusion | m=10 | $\bar{x} = 2902.8$ | $s_1 = 277.3$ |
|---|---|---|---|
| Fused | n=8 | $\bar{y} = 3108.1$ | $s_2 = 205.9$ |

The authors of the article stated that the fusion process increased the average tensile strength. Let’s carry out a test of hypotheses to see whether the data supports this conclusion.(use 0.05 significance level)

# Solution

Let $\mu_1$ be the true average tensile strength of specimens when the no-fusion treatment is used and $\mu_2$ denote the true average tensile strength when the fusion treatment is used.

$$H_0: \mu_1 - \mu_2 = 0$$

$$H_a: \mu_1 - \mu_2 < 0$$

$$t_{test} = \frac{(\bar{x} - \bar{y}) - \Delta_0}{\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}} = \frac{2902.8 - 3108.1 - 0}{\sqrt{\frac{277.3^2}{10} + \frac{205.9^2}{8}}} = -1.8$$

$$v = \frac{(\frac{s_1^2}{m} + \frac{s_2^2}{n})^2}{\frac{(s_1^2/m)^2}{m-1} + \frac{(s_2^2/n)^2}{n-1}} = 15.94 \approx 15$$

Decision Rule:

*Critical value approach:*

$$if\ t_{test} \leq -t_{\alpha}, then\ reject\ H_0.$$

$$t_{0.05,15} = 1.753, since\ t_{test} = -1.8 < -1.753,$$

$$we\ can\ barely\ reject\ the\ null\ hypothesis\ in\ favor\ of\ the$$

$$alternative\ hyptohesis\ at\ 0.05\ significance\ level.$$

*P-value approach:*

$$If\ p - value\ \leq \alpha, then\ reject\ H_0.$$

$$P - value = P(t \leq t_{test}) = P(t \leq -1.8) = P(t \geq 1.8) = (0.025, 0.05)$$

*Since P-value less than α, we can reject the null hypothesis.*

Conclusion:

The data support the conclusion stated in the article.

# Pooled t Procedures

Assumptions: Both population distributions are normal, two random samples are selected independently of one another, two population distributions have equal variances ($\sigma_1^2 = \sigma_2^2$)

- $100(1-\alpha)\%$ confidence interval for $\mu_1 - \mu_2$

$$(\bar{x} - \bar{y}) \pm t_{\alpha/2,v} \cdot s_p \cdot \sqrt{\frac{1}{m} + \frac{1}{n}}$$

Where $s_p$ is called the pooled sample standard deviation

$$s_p = \sqrt{\frac{(m-1)s_1^2 + (n-1)s_2^2}{m+n-2}}$$

The two-sample pooled t test for testing $H_0: \mu_1 - \mu_2 = \Delta_0$ is as follows:

$$t_{test} = \frac{(\bar{x} - \bar{y}) - \Delta_0}{s_p \cdot \sqrt{\frac{1}{m} + \frac{1}{n}}}$$

The degrees of freedom $v = m + n - 2$

# Analysis of Paired Data

There are a number of experimental situation in which there is only one set of n individuals or experimental objects, making two observations on each one results in a natural pairing of values.

Assumptions: The data consists of n independently selected pairs $(X_1, Y_1), (X_2, Y_2), \ldots (X_n, Y_n)$, with $E(X_i) = \mu_1$ and $E(Y_i) = \mu_2$. Let $D_1 = X_1 - Y_1, D_2 = X_2 - Y_2, \ldots\ D_n = X_n - Y_n$ so the $D_i's$ are the differences within pairs. The $D_i's$ are assumed to be normally distributed with mean value $\mu_D$ and variance $\sigma_D^2$.

# The Paired t Test

Null hypothesis: $H_0: \mu_D = \Delta_0 \ (Where\ \mu_D = \mu_1 - \mu_2)$

Test statistic value: $t_{test} = \frac{\bar{d} - \Delta_0}{s_D/\sqrt{n}}$

- P-value approach:

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

$H_a: \mu_D > \Delta_0 \quad P - value = P(t_{n-1} \geq t_{test})$

$H_a: \mu_D < \Delta_0 \quad P - value = P(t_{n-1} \leq t_{test})$

$H_a: \mu_D \neq \Delta_0 \quad P - value = 2 \times P(t_{n-1} \geq |t_{test}|)$

- Critical Value approach:

$H_a: \mu_D > \Delta_0$ If $t_{test} \geq t_\alpha$, then reject $H_0$.

$H_a: \mu_D < \Delta_0$ If $t_{test} \leq -t_\alpha$, then reject $H_0$.

$H_a: \mu_D \neq \Delta_0$ If $t_{test} \geq +t_{\frac{\alpha}{2}}\ or\ t_{test} \leq -\ t_{\frac{\alpha}{2}}$, then reject $H_0$.

df = n-1

# Example 9.3

Many freeways have service signs that give information on attractions, camping, lodging, food, and gas services prior to off-ramps. These signs typically do not provide information on distances. The article "Evaluation of Adding Distance Information to Free-way Specific Service Signs" reported that in one investigation, six sites along Virginia interstate high-ways where service signs are posted were selected. For each site, crash data was obtained for a three-year period before distance information was added to the service signs and for a one-year period afterward. The number of crashes per year before and after the sign changes were as follows:

| Before | 15 | 26 | 66 | 115 | 62 | 64 |
|---|---|---|---|---|---|---|
| After | 16 | 24 | 42 | 80 | 78 | 73 |

The cited article included the statement “A paired t test was performed to determine whether there was any change in the mean number of crashes before and after the addition of distance information on the signs” Carry out such a test. (α=0.05)

*Solution:*

| Before | 15 | 26 | 66 | 115 | 62 | 64 |
|---|---|---|---|---|---|---|
| After | 16 | 24 | 42 | 80 | 78 | 73 |
| Diff=B-A | -1 | 2 | 24 | 35 | -16 | -9 |

$$\bar{d} = \frac{\sum d_i}{n} = \frac{-1 + 2 + 24 + 35 + (-16) + (-9)}{6} = 5.83$$

$$s_D = \sqrt{\frac{\sum (d_i - \bar{d})^2}{n-1}} = 19.70$$

$H_0 : \mu_D = 0$

$H_a : \mu_D \neq 0$

$$t_{test} = \frac{\bar{d} - \Delta_0}{s_D / \sqrt{n}} = \frac{5.83 - 0}{19.70/\sqrt{6}} = 0.72$$

Df=n-1=5

- P-value approach:

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

$P - value = 2 \times P(t_{n-1} \geq |t_{test}|) = 2 \times P(t \geq 0.72) = 2 \times (> 0.1)$
$= (> 0.2)$

P-value >0.2 then greater than 0.05(α), therefore, don't reject $H_0$.

- Critical Value approach:

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

$t_{\frac{\alpha}{2}} = t_{0.025} = 2.571$ $and$ $t_{test} = 0.72 < +2.571, then, don't$ $reject$ $H_0$.

- Conclusion:

We don’t have enough evidence to conclude that there was a difference in the mean number of crashes before and after the addition of distance information on the signs.

# The Paired t Confidence Interval

$100(1-\alpha)\%$ confidence interval of $\mu_D$

$$\bar{d} \pm t_{\frac{\alpha}{2},n-1} \cdot \frac{s_D}{\sqrt{n}}$$

Practice Question:

Find a 95% CI for the true difference in the mean number of crashes in example 9.3.

# Inferences Concerning a Difference Between Population Proportions

A Large-Sample Test Procedure

Null hypothesis: $H_0: p_1 - p_2 = 0$

Test statistic value: $z_{test} = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{m}+\frac{1}{n})}}$

$$\hat{p} = \frac{m}{m+n}\hat{p}_1 + \frac{n}{m+n}\hat{p}_2$$

Make the decision

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

$H_a: p_1 - p_2 > 0 \qquad P - value = P(Z \geq Z_{test})$

$H_a: p_1 - p_2 < 0 \qquad P - value = P(Z \leq Z_{test})$

$H_a: p_1 - p_2 \neq 0 \qquad P - value = 2 \times P(Z \geq |Z_{test}|)$

- Critical value approach

$H_a: p_1 - p_2 > 0$ If $Z_{test} \geq Z_\alpha$, then reject $H_0$.

$H_a: p_1 - p_2 < 0$ If $Z_{test} \leq -Z_\alpha$, then reject $H_0$.

$H_a: p_1 - p_2 \neq 0$ If $Z_{test} \geq +Z_{\frac{\alpha}{2}}$ $or$ $Z_{test} \leq -Z_{\frac{\alpha}{2}}$ , then reject $H_0$.

The test can safely be used as long as $m\hat{p}_1, m(1 - \hat{p}_1), n\hat{p}_2$ $and$ $n(1 - \hat{p}_2)$ are all at least 10.

# Example 9.4

The article "Aspirin Use and Survival After Diagnosis of Colorectal Cancer" reported that of 549 study participants who regularly used aspirin after being diagnosed with colorectal cancer, there were 81 colorectal cancer-specific deaths, whereas among 730 similarly diagnosed individuals who did not subsequently use aspirin, there were 141 colorectal cancer-specific deaths. Does this data suggest that the regular use of aspirin after diagnosis will decrease the incidence rate of colorectal cancer-specific deaths? Let's test the appropriate hypotheses using a significance level of 0.05.

# Solution:

Let $p_1$ denote the true proportion of deaths for those who regularly used aspirin and $p_2$ denote the true proportion of deaths for those who did not use aspirin.

$H_0: p_1 - p_2 = 0$
$H_a: p_1 - p_2 < 0$

$$\hat{p}_1 = \frac{81}{549} = 0.1475 \quad \hat{p}_2 = \frac{141}{730} = 0.1932$$

$$\hat{p} = \frac{81 + 141}{549 + 730} = 0.1736$$

- $z_{test} = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{m}+\frac{1}{n})}} = \frac{0.1475-0.1932}{\sqrt{0.1736(1-0.1736)(\frac{1}{549}+\frac{1}{730})}} = -2.14$

- P-value approach:

The corresponding P-value for a lower-tailed z test:

$$P - value = P(Z \leq z_{test}) = P(Z \leq -2.14) = 0.0162.$$

Because 0.0162 ≤ 0.05, the null hypothesis can be rejected at significance level 0.05.

- Critical value approach:

If $Z_{test} \leq -Z_{\alpha}$, then reject $H_0$. $Z_{\alpha} = Z_{0.05} = 1.645$

-2.14≤ $-1.645$, the null hypothesis can be rejected.

We have enough evidence to conclude that the use of aspirin in these circumstances is beneficial.

# Type II Error Probabilities and Sample Size

$H_a: p_1 - p_2 > 0$

$$\beta(p_1, p_2) = \Phi[\frac{z_\alpha\sqrt{\bar{p}(1-\bar{p})(\frac{1}{m}+\frac{1}{n})} - (p_1 - p_2)}{\sigma}]$$

$H_a: p_1 - p_2 < 0$

$$\beta(p_1, p_2) = 1 - \Phi[\frac{-z_\alpha\sqrt{\bar{p}(1-\bar{p})(\frac{1}{m}+\frac{1}{n})} - (p_1 - p_2)}{\sigma}]$$

$H_a: p_1 - p_2 \neq 0$

$$\beta(p_1, p_2) = \Phi\left[\frac{z_{\frac{\alpha}{2}}\sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{m}+\frac{1}{n}\right)} - (p_1 - p_2)}{\sigma}\right] - \Phi[\frac{-z_{\frac{\alpha}{2}}\sqrt{\bar{p}(1-\bar{p})(\frac{1}{m}+\frac{1}{n})} - (p_1 - p_2)}{\sigma}]$$

Where $\bar{p} = (mp_1 + np_2)/(m + n)$ and

$$\sigma = \sqrt{\frac{p_1(1-p_1)}{m} + \frac{p_2(1-p_2)}{n}}$$

**Sample Size:**

For the case m=n, the level of $\alpha$ test has type II error probability $\beta$ at the alternative value $p_1, p_2\ with\ p_1 - p_2 = d$ when

$$n = \frac{[z_\alpha \sqrt{\frac{(p_1+p_2)(2-p_1-p_2)}{2}} + z_\beta \sqrt{p_1(1-p_1) + p_2(1-p_2}\ ]^2}{d^2}$$

For an upper- or lower-tailed test, with $\alpha/2$ replacing $\alpha$ for a two-tailed test.

DALHOUSIE UNIVERSITY

# A Large-Sample Confidence Interval

A CI for $p_1 - p_2$ with confidence level approximately $100(1-\alpha)\%$ is

$$(\hat{p}_1 - \hat{p}_2) \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{m} + \frac{\hat{p}_2(1-\hat{p}_2)}{n}}$$

This interval can safely be used as long as $m\hat{p}_1, m(1-\hat{p}_1), n\hat{p}_2$ $and$ $n(1-\hat{p}_2)$ are all at least 10.

# Example 9.5

Do teachers find their work rewarding and satisfying? The article “Work-Related Attitudes” reports the results of a survey of 395 elementary school teachers and 266 high school teachers. Of the elementary school teachers, 224 said they were very satisfied with their job, whereas 126 of the high school teacher were very satisfied with their work. Find a 95% CI for the difference between the proportion of all elementary school teachers who are very satisfied and all high school teachers who are very satisfied with their work.

# Solution:

$$\hat{p}_1 = \frac{224}{395} = 0.5671, \hat{p}_2 = \frac{126}{266} = 0.4737$$

$$(\hat{p}_1 - \hat{p}_2) \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{m} + \frac{\hat{p}_2(1-\hat{p}_2)}{n}} =$$

$$(0.5671 - 0.4737)$$

$$\pm 1.96 \sqrt{\frac{0.5671(1-0.5671)}{395} + \frac{0.4737(1-0.4737)}{266}} =$$

$$0.0934 \pm 0.0774 = (0.016, 0.1708)$$