STATISTICAL INFERENCES (2cr)
Chapter 9 One- and Two-Sample Estimation Problems

Zhong Guan

Math, IUSB

Outline

Outline

1. 9.8 Estimating the Differences of Two Means
   - Known Variances
   - Equal But Unknown Variances
   - Unknown and Unequal Variances

2. 9.9 Paired Observations
   - Based on Paired Data

Outline

An Example

Example 1

- Comparing gas mileage between two types of engines, A and B.
- Fifty experiments were conducted using engine type A: average gas mileage is 36 mpg. Population standard deviation is $\sigma_A = 6$.
- Seventy five experiments were conducted using engine type B: average gas mileage is 42 mpg. Population standard deviation is $\sigma_B = 8$.
- Find a 95% confidence interval for the difference of the mean gas mileage, $\mu_A - \mu_B$.

An Example

Example 1

- Comparing gas mileage between two types of engines, A and B.
- Fifty experiments were conducted using engine type A: average gas mileage is 36 mpg. Population standard deviation is $\sigma_A = 6$.
- Seventy five experiments were conducted using engine type B: average gas mileage is 42 mpg. Population standard deviation is $\sigma_B = 8$.
- Find a 95% confidence interval for the difference of the mean gas mileage, $\mu_A - \mu_B$.

An Example

Example 1

- Comparing gas mileage between two types of engines, A and B.
- Fifty experiments were conducted using engine type A: average gas mileage is 36 mpg. Population standard deviation is $\sigma_A = 6$.
- Seventy five experiments were conducted using engine type B: average gas mileage is 42 mpg. Population standard deviation is $\sigma_B = 8$.
- Find a 95% confidence interval for the difference of the mean gas mileage, $\mu_A - \mu_B$.

An Example

Example 1

- Comparing gas mileage between two types of engines, A and B.
- Fifty experiments were conducted using engine type A: average gas mileage is 36 mpg. Population standard deviation is $\sigma_A = 6$.
- Seventy five experiments were conducted using engine type B: average gas mileage is 42 mpg. Population standard deviation is $\sigma_B = 8$.
- Find a 95% confidence interval for the difference of the mean gas mileage, $\mu_A - \mu_B$.

Solution of Example 1

- Point estimate of $\mu_A - \mu_B$: $\bar{X}_A - \bar{X}_B = 36 - 42 = -6$.
- 95% confidence interval for $\mu_A - \mu_B$: $\bar{X}_A - \bar{X}_B \pm E$.
- $E = ?$

$$95\% = P(\bar{X}_A - \bar{X}_B - E < \mu_A - \mu_B < \bar{X}_A - \bar{X}_B + E)$$

Equivalently

$$95\% = P(-E < (\bar{X}_A - \bar{X}_B) - (\mu_A - \mu_B) < E)$$

- By the CLT,

$$Z = \frac{(\bar{X}_A - \bar{X}_B) - (\mu_A - \mu_B)}{\sqrt{\sigma_A^2/n_A + \sigma_B^2/n_B}} \quad \text{is standard normal.}$$

Solution of Example 1

- Point estimate of $\mu_A - \mu_B$: $\bar{X}_A - \bar{X}_B = 36 - 42 = -6$.
- 95% confidence interval for $\mu_A - \mu_B$: $\bar{X}_A - \bar{X}_B \pm E$.
- $E = ?$

$$95\% = P(\bar{X}_A - \bar{X}_B - E < \mu_A - \mu_B < \bar{X}_A - \bar{X}_B + E)$$

Equivalently

$$95\% = P(-E < (\bar{X}_A - \bar{X}_B) - (\mu_A - \mu_B) < E)$$

- By the CLT,

$$Z = \frac{(\bar{X}_A - \bar{X}_B) - (\mu_A - \mu_B)}{\sqrt{\sigma_A^2/n_A + \sigma_B^2/n_B}} \quad \text{is standard normal.}$$

Solution of Example 1

- Point estimate of $\mu_A - \mu_B$: $\bar{X}_A - \bar{X}_B = 36 - 42 = -6$.
- 95% confidence interval for $\mu_A - \mu_B$: $\bar{X}_A - \bar{X}_B \pm E$.
- $E = ?$

$$95\% = P(\bar{X}_A - \bar{X}_B - E < \mu_A - \mu_B < \bar{X}_A - \bar{X}_B + E)$$

Equivalently

$$95\% = P(-E < (\bar{X}_A - \bar{X}_B) - (\mu_A - \mu_B) < E)$$

- By the CLT,

$$Z = \frac{(\bar{X}_A - \bar{X}_B) - (\mu_A - \mu_B)}{\sqrt{\sigma_A^2/n_A + \sigma_B^2/n_B}} \quad \text{is standard normal.}$$

Solution of Example 1

- Point estimate of $\mu_A - \mu_B$: $\bar{X}_A - \bar{X}_B = 36 - 42 = -6$.
- 95% confidence interval for $\mu_A - \mu_B$: $\bar{X}_A - \bar{X}_B \pm E$.
- $E = ?$

$$95\% = P(\bar{X}_A - \bar{X}_B - E < \mu_A - \mu_B < \bar{X}_A - \bar{X}_B + E)$$

Equivalently

$$95\% = P(-E < (\bar{X}_A - \bar{X}_B) - (\mu_A - \mu_B) < E)$$

- By the CLT,

$$Z = \frac{(\bar{X}_A - \bar{X}_B) - (\mu_A - \mu_B)}{\sqrt{\sigma_A^2/n_A + \sigma_B^2/n_B}} \quad \text{is standard normal.}$$

Solution of Example 1

- Point estimate of $\mu_A - \mu_B$: $\bar{X}_A - \bar{X}_B = 36 - 42 = -6$.
- 95% confidence interval for $\mu_A - \mu_B$: $\bar{X}_A - \bar{X}_B \pm E$.
- $E = ?$

$$95\% = P(\bar{X}_A - \bar{X}_B - E < \mu_A - \mu_B < \bar{X}_A - \bar{X}_B + E)$$

Equivalently

$$95\% = P(-E < (\bar{X}_A - \bar{X}_B) - (\mu_A - \mu_B) < E)$$

- By the CLT,

$$Z = \frac{(\bar{X}_A - \bar{X}_B) - (\mu_A - \mu_B)}{\sqrt{\sigma_A^2/n_A + \sigma_B^2/n_B}} \quad \text{is standard normal.}$$

Solution of Example 1

- Point estimate of $\mu_A - \mu_B$: $\bar{X}_A - \bar{X}_B = 36 - 42 = -6$.
- 95% confidence interval for $\mu_A - \mu_B$: $\bar{X}_A - \bar{X}_B \pm E$.
- $E = ?$

$$95\% = P(\bar{X}_A - \bar{X}_B - E < \mu_A - \mu_B < \bar{X}_A - \bar{X}_B + E)$$

Equivalently

$$95\% = P(-E < (\bar{X}_A - \bar{X}_B) - (\mu_A - \mu_B) < E)$$

- By the CLT,

$$Z = \frac{(\bar{X}_A - \bar{X}_B) - (\mu_A - \mu_B)}{\sqrt{\sigma_A^2/n_A + \sigma_B^2/n_B}}$$ is standard normal.

Solution of Example 1(cont.)

- Because $P(-1.96 < Z < 1.96) = 0.95$

$$\frac{E}{\sqrt{\sigma_A^2/n_A + \sigma_B^2/n_B}} = 1.96$$

$$\rightarrow \quad E = 1.96\sqrt{\sigma_A^2/n_A + \sigma_B^2/n_B}$$

$$= 1.96\sqrt{6^2/50 + 8^2/70} = 2.506$$

- a 95% confidence interval for $\mu_A - \mu_B$ is

$$[-6 - 2.506, -6 + 2.506] = [-8.506, -3.494]$$

Solution of Example 1(cont.)

- Because $P(-1.96 < Z < 1.96) = 0.95$

$$\frac{E}{\sqrt{\sigma_A^2/n_A + \sigma_B^2/n_B}} = 1.96$$

$$\rightarrow \quad E = 1.96\sqrt{\sigma_A^2/n_A + \sigma_B^2/n_B}$$

$$= 1.96\sqrt{6^2/50 + 8^2/70} = 2.506$$

- a 95% confidence interval for $\mu_A - \mu_B$ is

$$[-6 - 2.506, -6 + 2.506] = [-8.506, -3.494]$$

Solution of Example 1(cont.)

- Because $P(-1.96 < Z < 1.96) = 0.95$

$$\frac{E}{\sqrt{\sigma_A^2/n_A + \sigma_B^2/n_B}} = 1.96$$

$$\rightarrow \quad E = 1.96\sqrt{\sigma_A^2/n_A + \sigma_B^2/n_B}$$

$$= 1.96\sqrt{6^2/50 + 8^2/70} = 2.506$$

- a 95% confidence interval for $\mu_A - \mu_B$ is

$$[-6 - 2.506, -6 + 2.506] = [-8.506, -3.494]$$

Solution of Example 1(cont.)

- Because $P(-1.96 < Z < 1.96) = 0.95$

$$\frac{E}{\sqrt{\sigma_A^2/n_A + \sigma_B^2/n_B}} = 1.96$$

$$\rightarrow \quad E = 1.96\sqrt{\sigma_A^2/n_A + \sigma_B^2/n_B}$$

$$= 1.96\sqrt{6^2/50 + 8^2/70} = 2.506$$

- a 95% confidence interval for $\mu_A - \mu_B$ is

$$[-6 - 2.506, -6 + 2.506] = [-8.506, -3.494]$$

Two Normal Independent Samples

- Let $X_1, \ldots, X_m$ and $Y_1, Y_2, \ldots, Y_n$ be independent random samples of sizes m and n from $N(\mu_X, \sigma_X^2)$ and $N(\mu_Y, \sigma_Y^2)$ respectively.
- Since sample means $\overline{X}$ and $\overline{Y}$ are independent and
$$\overline{X} \sim N(\mu_X, \frac{\sigma_X^2}{m}), \quad \overline{Y} \sim N(\mu_Y, \frac{\sigma_Y^2}{n}).$$
- So
$$W = \overline{X} - \overline{Y} \sim N(\mu_X - \mu_Y, \frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n})$$
- $$P(-z_{\alpha/2} \leqslant \frac{\overline{X} - \overline{Y} - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}}} \leqslant z_{\alpha/2}) = 1 - \alpha.$$

Two Normal Independent Samples

- Let $X_1, \ldots, X_m$ and $Y_1, Y_2, \ldots, Y_n$ be independent random samples of sizes m and n from $N(\mu_X, \sigma_X^2)$ and $N(\mu_Y, \sigma_Y^2)$ respectively.
- Since sample means $\overline{X}$ and $\overline{Y}$ are independent and

$$\overline{X} \sim N(\mu_X, \frac{\sigma_X^2}{m}), \quad \overline{Y} \sim N(\mu_Y, \frac{\sigma_Y^2}{n}).$$

- So

$$W = \overline{X} - \overline{Y} \sim N(\mu_X - \mu_Y, \frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n})$$

-

$$P\left(-z_{\alpha/2} \leqslant \frac{\overline{X}-\overline{Y}-(\mu_X-\mu_Y)}{\sqrt{\frac{\sigma_X^2}{m}+\frac{\sigma_Y^2}{n}}} \leqslant z_{\alpha/2}\right) = 1-\alpha.$$

Two Normal Independent Samples

- Let $X_1, \ldots, X_m$ and $Y_1, Y_2, \ldots, Y_n$ be independent random samples of sizes m and n from $N(\mu_X, \sigma_X^2)$ and $N(\mu_Y, \sigma_Y^2)$ respectively.
- Since sample means $\overline{X}$ and $\overline{Y}$ are independent and

$$\overline{X} \sim N(\mu_X, \frac{\sigma_X^2}{m}), \quad \overline{Y} \sim N(\mu_Y, \frac{\sigma_Y^2}{n}).$$

- So

$$W = \overline{X} - \overline{Y} \sim N(\mu_X - \mu_Y, \frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n})$$

- $$P(-z_{\alpha/2} \leqslant \frac{\overline{X}-\overline{Y}-(\mu_X-\mu_Y)}{\sqrt{\frac{\sigma_X^2}{m}+\frac{\sigma_Y^2}{n}}} \leqslant z_{\alpha/2}) = 1-\alpha.$$

Two Normal Independent Samples

- Let $X_1, \ldots, X_m$ and $Y_1, Y_2, \ldots, Y_n$ be independent random samples of sizes m and n from $N(\mu_X, \sigma_X^2)$ and $N(\mu_Y, \sigma_Y^2)$ respectively.
- Since sample means $\overline{X}$ and $\overline{Y}$ are independent and

$$\overline{X} \sim N\left(\mu_X, \frac{\sigma_X^2}{m}\right), \quad \overline{Y} \sim N\left(\mu_Y, \frac{\sigma_Y^2}{n}\right).$$

- So

$$W = \overline{X} - \overline{Y} \sim N\left(\mu_X - \mu_Y, \frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}\right)$$

- $$P\left(-z_{\alpha/2} \leqslant \frac{\overline{X}-\overline{Y}-(\mu_X-\mu_Y)}{\sqrt{\frac{\sigma_X^2}{m}+\frac{\sigma_Y^2}{n}}} \leqslant z_{\alpha/2}\right) = 1 - \alpha.$$

Confidence Interval for $\mu_x - \mu_y$: σ_X^2 and σ_Y^2 known

$$P\left[(\overline{X} - \overline{Y}) - z_{\alpha/2}\sigma_W \leqslant \mu_X - \mu_Y \leqslant (\overline{X} - \overline{Y}) + z_{\alpha/2}\sigma_W\right] = 1 - \alpha,$$

where

$$\sigma_W = \sqrt{\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}}.$$

So we have a $100(1-\alpha)\%$ confidence interval for $\mu_X - \mu_Y$:

$$\left[(\overline{X} - \overline{Y}) - z_{\alpha/2}\sigma_W, \quad (\overline{X} - \overline{Y}) + z_{\alpha/2}\sigma_W\right]$$

Example 2.

If $m = 15$, $n = 8$, $\bar{x} = 70.1$, $\bar{y} = 75.3$, $\sigma_X = 60$, $\sigma_Y = 40$, and $1 - \alpha = 0.95$, then from Table A-3, $z_{\alpha/2} = z_{0.025} = 1.96$, $\bar{x} - \bar{y} = -5.2$ and

$$1.96\sigma_W = 1.96\sqrt{\frac{60^2}{15} + \frac{40^2}{8}} = 5.88.$$

Thus a 95% confidence interval for $\mu_X - \mu_Y$ is

$$[-5.2 - 5.88, -5.2 + 5.88] = [-11.08, 0.68].$$

Outline

Confidence Interval for $\mu_X - \mu_Y$: $\sigma_X^2 = \sigma_Y^2 = \sigma^2$ unknown

A $100(1-\alpha)\%$ confidence interval for $\mu_X - \mu_Y$:

$$(\overline{X} - \overline{Y}) \pm E$$

$$E = t_{\frac{\alpha}{2}}(m+n-2)S_p$$

$$S_p = \sqrt{\frac{(m-1)S_X^2+(n-1)S_Y^2}{m+n-2}\left(\frac{1}{m}+\frac{1}{n}\right)}$$

Example 3:

The following data represent the running times of films produced by two motion-picture companies.

Company	Time (minutes)						
1	102	86	98	109	92		
2	81	165	97	134	92	87	114

Assume the samples are from $N(\mu_1, \sigma^2)$ and $N(\mu_2, \sigma^2)$ with equal variances.
Give a 95% confidence interval for the difference $\mu_1 - \mu_2$ of the mean running times of the two companies.

Example 3:

The following data represent the running times of films produced by two motion-picture companies.

Company	Time (minutes)						
1	102	86	98	109	92		
2	81	165	97	134	92	87	114

Assume the samples are from $N(\mu_1, \sigma^2)$ and $N(\mu_2, \sigma^2)$ with equal variances.

Give a 95% confidence interval for the difference $\mu_1 - \mu_2$ of the mean running times of the two companies.

Example 3:

The following data represent the running times of films produced by two motion-picture companies.

Company	Time (minutes)						
1	102	86	98	109	92		
2	81	165	97	134	92	87	114

Assume the samples are from $N(\mu_1, \sigma^2)$ and $N(\mu_2, \sigma^2)$ with equal variances.
Give a 95% confidence interval for the difference $\mu_1 - \mu_2$ of the mean running times of the two companies.

Solution of Example 3:

- $m = 5$, $n = 7$, $\alpha = 0.05$;
- $\bar{X} = 97.4$, $\bar{Y} = 110$, $S_X^2 = 78.8$, $S_Y^2 = 913.33$,
- The pooled standard error

$$S_p = \sqrt{\frac{(m-1)S_X^2+(n-1)S_Y^2}{m+n-2}\left(\frac{1}{m}+\frac{1}{n}\right)}$$

$$= \sqrt{\frac{(5-1)78.8+(7-1)913.3}{5+7-2}\left(\frac{1}{5}+\frac{1}{7}\right)} = 14.10$$

- $E = t_{\frac{\alpha}{2}}(m+n-2)S_p = t_{0.025}(10)S_p = 2.228(14.10) = 31.41$
- The 95% CI:

$$(\overline{X}-\overline{Y}) \pm E = (-44.01, 18.81)$$

Solution of Example 3:

- $m = 5$, $n = 7$, $\alpha = 0.05$;
- $\bar{X} = 97.4$, $\bar{Y} = 110$, $S_X^2 = 78.8$, $S_Y^2 = 913.33$,
- The pooled standard error

$$S_p = \sqrt{\frac{(m-1)S_X^2 + (n-1)S_Y^2}{m+n-2}\left(\frac{1}{m} + \frac{1}{n}\right)}$$

$$= \sqrt{\frac{(5-1)78.8 + (7-1)913.3}{5+7-2}\left(\frac{1}{5} + \frac{1}{7}\right)} = 14.10$$

- $E = t_{\frac{\alpha}{2}}(m+n-2)S_p = t_{0.025}(10)S_p = 2.228(14.10) = 31.41$
- The 95% CI:

$$(\overline{X} - \overline{Y}) \pm E = (-44.01, 18.81)$$

Solution of Example 3:

- $m = 5$, $n = 7$, $\alpha = 0.05$;
- $\bar{X} = 97.4$, $\bar{Y} = 110$, $S_X^2 = 78.8$, $S_Y^2 = 913.33$,
- The pooled standard error

$$S_p = \sqrt{\frac{(m-1)S_X^2+(n-1)S_Y^2}{m+n-2}\left(\frac{1}{m}+\frac{1}{n}\right)}$$

$$= \sqrt{\frac{(5-1)78.8+(7-1)913.3}{5+7-2}\left(\frac{1}{5}+\frac{1}{7}\right)} = 14.10$$

- $E = t_{\frac{\alpha}{2}}(m+n-2)S_p = t_{0.025}(10)S_p = 2.228(14.10) = 31.41$
- The 95% CI:

$$(\bar{X} - \bar{Y}) \pm E = (-44.01, 18.81)$$

Solution of Example 3:

- $m = 5$, $n = 7$, $\alpha = 0.05$;
- $\bar{X} = 97.4$, $\bar{Y} = 110$, $S_X^2 = 78.8$, $S_Y^2 = 913.33$,
- The pooled standard error

$$S_p = \sqrt{\frac{(m-1)S_X^2+(n-1)S_Y^2}{m+n-2}\left(\frac{1}{m}+\frac{1}{n}\right)}$$

$$= \sqrt{\frac{(5-1)78.8+(7-1)913.3}{5+7-2}\left(\frac{1}{5}+\frac{1}{7}\right)} = 14.10$$

- $E = t_{\frac{\alpha}{2}}(m+n-2)S_p = t_{0.025}(10)S_p = 2.228(14.10) = 31.41$
- The 95% CI:

$$(\overline{X} - \overline{Y}) \pm E = (-44.01, 18.81)$$

Solution of Example 3:

- $m = 5$, $n = 7$, $\alpha = 0.05$;
- $\bar{X} = 97.4$, $\bar{Y} = 110$, $S_X^2 = 78.8$, $S_Y^2 = 913.33$,
- The pooled standard error

$$S_p = \sqrt{\frac{(m-1)S_X^2+(n-1)S_Y^2}{m+n-2}\left(\frac{1}{m}+\frac{1}{n}\right)}$$

$$= \sqrt{\frac{(5-1)78.8+(7-1)913.3}{5+7-2}\left(\frac{1}{5}+\frac{1}{7}\right)} = 14.10$$

- $E = t_{\frac{\alpha}{2}}(m+n-2)S_p = t_{0.025}(10)S_p = 2.228(14.10) = 31.41$
- The 95% CI:

$$(\bar{X}-\bar{Y}) \pm E = (-44.01, 18.81)$$

Solution of Example 3:

- $m = 5$, $n = 7$, $\alpha = 0.05$;
- $\bar{X} = 97.4$, $\bar{Y} = 110$, $S_X^2 = 78.8$, $S_Y^2 = 913.33$,
- The pooled standard error

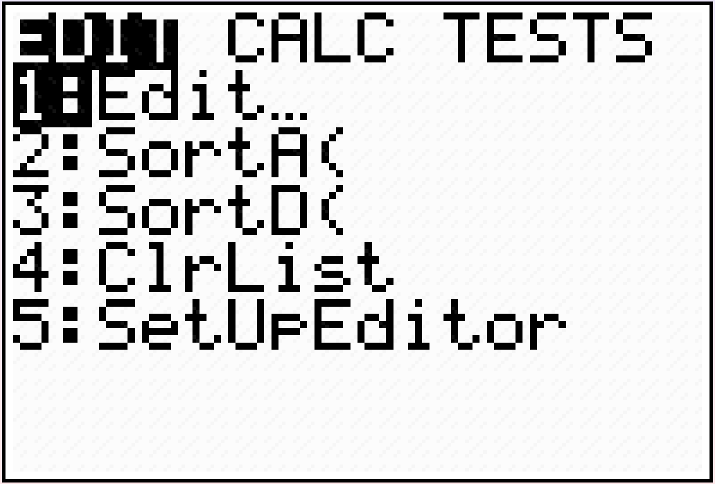
$$S_p = \sqrt{\frac{(m-1)S_X^2 + (n-1)S_Y^2}{m+n-2}\left(\frac{1}{m} + \frac{1}{n}\right)}$$

$$= \sqrt{\frac{(5-1)78.8 + (7-1)913.3}{5+7-2}\left(\frac{1}{5} + \frac{1}{7}\right)} = 14.10$$

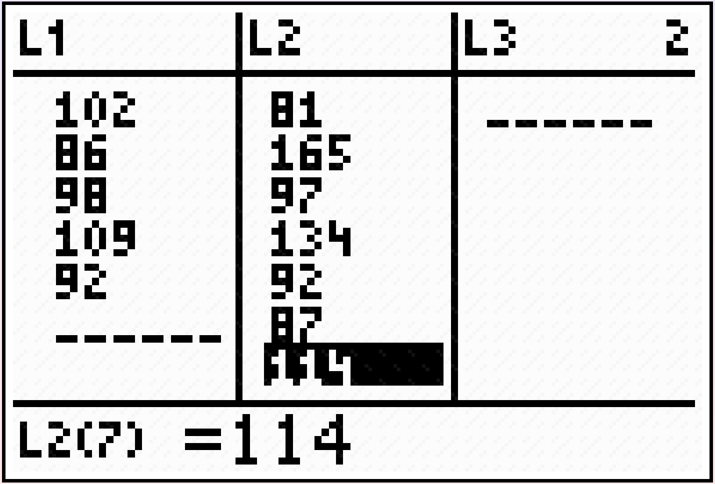
- $E = t_{\frac{\alpha}{2}}(m+n-2)S_p = t_{0.025}(10)S_p = 2.228(14.10) = 31.41$
- The 95% CI:

$$(\overline{X} - \overline{Y}) \pm E = (-44.01, 18.81)$$

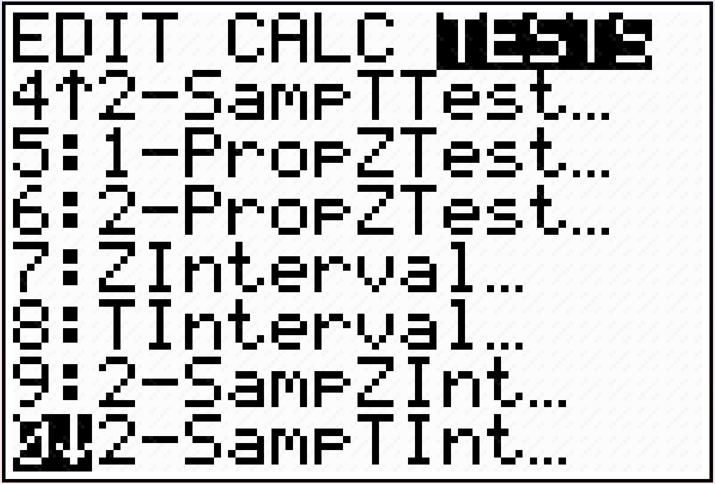
Solution of Example 3 using TI8x



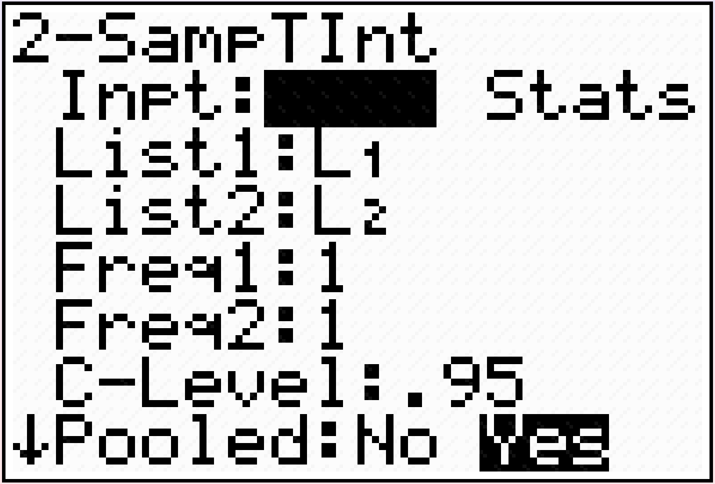
Solution of Example 3 using TI8x



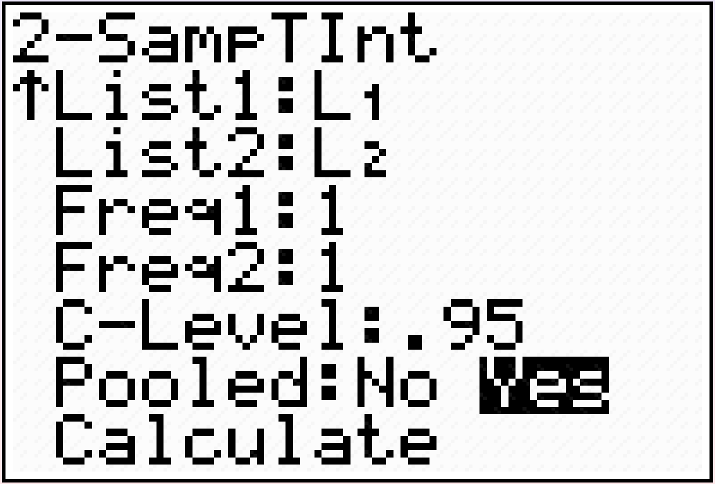
Solution of Example 3 using TI8x



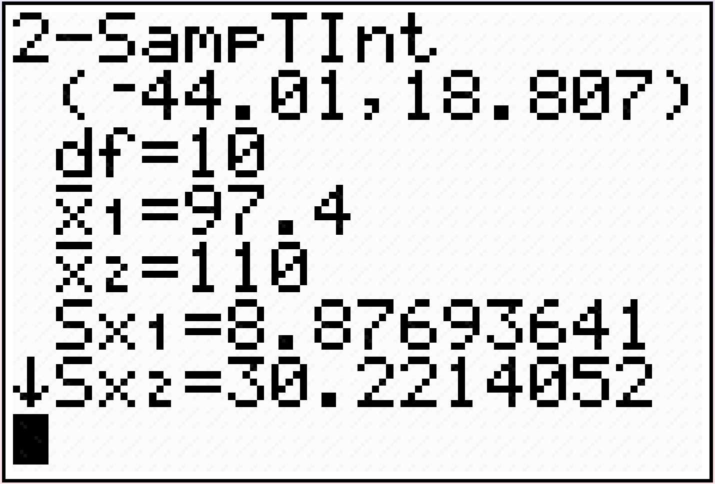
Solution of Example 3 using TI8x



Solution of Example 3 using TI8x



Solution of Example 3 using TI8x



Solution of Example 3 using Excel

Company	Times(minutes)						
1	102	86	98	109	92		
2	81	165	97	134	92	87	114

Confidence Interval: Unknown equal Variances

m	5	
n	7	
alpha	0.05	
Xbar1	=AVERAGE(C5:G5)	
Xbar2	110	
Sp	14.09584	
CI:	-44.0075	18.80748

Solution of Example 3 using Excel

Company	Times(minutes)						
1	102	86	98	109	92		
2	81	165	97	134	92	87	114

Confidence Interval: Unknown equal Variances		
m	5	
n	7	
alpha	0.05	
Xbar1	97.4	
Xbar2	=AVERAGE(C6:I6)	
Sp	14.09584	
CI:	-44.0075	18.80748

Solution of Example 3 using Excel

Company	Times(minutes)						
1	102	86	98	109	92		
2	81	165	97	134	92	87	114

Confidence Interval: Unknown equal Variances

m	5	
n	7	
alpha	0.05	
Xbar1	97.4	
Xbar2	110	
Sp	=SQRT((((C9-1)*VAR(C5:G5)+(C10-1)*VAR(C6:I6))/(C9+C10-2))*(1/C9+1/C10))	
CI:	-44.0075	18.80748

Solution of Example 3 using Excel

Company	Times(minutes)						
1	102	86	98	109	92		
2	81	165	97	134	92	87	114

Confidence Interval: Unknown equal Variances

m	5
n	7
alpha	0.05
Xbar1	97.4
Xbar2	110
Sp	14.09584
CI:	=C12-C13-TINV(C11, C9+C10-2)*C14

Solution of Example 3 using Excel

Company	Times(minutes)						
1	102	86	98	109	92		
2	81	165	97	134	92	87	114

Confidence Interval: Unknown equal Variances

m	5	
n	7	
alpha	0.05	
Xbar1	97.4	
Xbar2	110	
Sp	14.09584	
CI:	-44.0075	=C12-C13+TINV(C11, C9+C10-2)*C14

Solution of Example 3 using Excel

Company	Times(minutes)						
1	102	86	98	109	92		
2	81	165	97	134	92	87	114

Confidence Interval: Unknown equal Variances		
m	5	
n	7	
alpha	0.05	
Xbar1	97.4	
Xbar2	110	
Sp	14.09584	
CI:	-44.0075	18.80748

Outline

Unknown and Unequal Variances

(a) If $m, n \geqslant 30$, then by the C.L.T. **approximately**

$$Z = \frac{\overline{X} - \overline{Y} - (\mu_X - \mu_Y)}{\sqrt{S_X^2/m + S_Y^2/n}} \sim N(0, 1),$$

(b) If $m < 30$ or $n < 30$, then **approximately**,

$$W = \frac{\overline{X} - \overline{Y} - (\mu_X - \mu_Y)}{\sqrt{S_X^2/m + S_Y^2/n}} \sim t(\lfloor r \rfloor),$$

where $\lfloor r \rfloor$, "floor" of r, is the integer part of r and

$$\frac{1}{r} = \frac{c^2}{m-1} + \frac{(1-c)^2}{n-1}, \qquad c = \frac{S_X^2/m}{S_X^2/m + S_Y^2/n}.$$

$$r = \frac{\left(S_X^2/m + S_Y^2/n\right)^2}{\frac{1}{m-1}\left(\frac{S_X^2}{m}\right)^2 + \frac{1}{n-1}\left(\frac{S_Y^2}{n}\right)^2}.$$

Unknown and Unequal Variances

(a) If $m, n \geqslant 30$, then by the C.L.T. **approximately**

$$Z = \frac{\overline{X} - \overline{Y} - (\mu_X - \mu_Y)}{\sqrt{S_X^2/m + S_Y^2/n}} \sim N(0,1),$$

(b) If $m < 30$ or $n < 30$, then **approximately**,

$$W = \frac{\overline{X} - \overline{Y} - (\mu_X - \mu_Y)}{\sqrt{S_X^2/m + S_Y^2/n}} \sim t(\lfloor r \rfloor),$$

where $\lfloor r \rfloor$, "floor" of r, is the integer part of r and

$$\frac{1}{r} = \frac{c^2}{m-1} + \frac{(1-c)^2}{n-1}, \quad c = \frac{S_X^2/m}{S_X^2/m + S_Y^2/n}.$$

$$r = \frac{\left(S_X^2/m + S_Y^2/n\right)^2}{\frac{1}{m-1}\left(\frac{S_X^2}{m}\right)^2 + \frac{1}{n-1}\left(\frac{S_Y^2}{n}\right)^2}.$$

C.I. for $\mu_X - \mu_Y$: both σ_X^2 and σ_Y^2 unknown

(a) If $m, n \geqslant 30$, then an approximate $100(1-\alpha)\%$ confidence interval for $\mu_X - \mu_Y$ is

$$(\overline{X} - \overline{Y}) \pm z_{\alpha/2}\sqrt{\frac{S_X^2}{m} + \frac{S_Y^2}{n}}$$

(b) If $m < 30$ or $n < 30$, then a $1-\alpha$ confidence interval for $\mu_X - \mu_Y$ is

$$(\overline{X} - \overline{Y}) \pm E$$

$$E = t_{\alpha/2}(\lfloor r \rfloor)\sqrt{\frac{S_X^2}{m} + \frac{S_Y^2}{n}}$$

C.I. for $\mu_X - \mu_Y$: both σ_X^2 and σ_Y^2 unknown

(a) If $m, n \geqslant 30$, then an approximate $100(1-\alpha)\%$ confidence interval for $\mu_X - \mu_Y$ is

$$(\overline{X} - \overline{Y}) \pm z_{\alpha/2}\sqrt{\frac{S_X^2}{m} + \frac{S_Y^2}{n}}$$

(b) If $m < 30$ or $n < 30$, then a $1-\alpha$ confidence interval for $\mu_X - \mu_Y$ is

$$(\overline{X} - \overline{Y}) \pm E$$

$$E = t_{\alpha/2}(\lfloor r \rfloor)\sqrt{\frac{S_X^2}{m} + \frac{S_Y^2}{n}}$$

Example 4:

The following data represent the running times of films produced by two motion-picture companies.

Company	Time (minutes)						
1	102	86	98	109	92		
2	81	165	97	134	92	87	114

Assume the samples are from $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ with unequal variances.
Give a 95% confidence interval for the difference $\mu_1 - \mu_2$ of the mean running times of the two companies.

Example 4:

The following data represent the running times of films produced by two motion-picture companies.

Company	Time (minutes)						
1	102	86	98	109	92		
2	81	165	97	134	92	87	114

Assume the samples are from $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ with unequal variances.

Give a 95% confidence interval for the difference $\mu_1 - \mu_2$ of the mean running times of the two companies.

Example 4:

The following data represent the running times of films produced by two motion-picture companies.

Company	Time (minutes)						
1	102	86	98	109	92		
2	81	165	97	134	92	87	114

Assume the samples are from $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ with unequal variances.
Give a 95% confidence interval for the difference $\mu_1 - \mu_2$ of the mean running times of the two companies.

Solution of Example 4:

- $m = 5$, $n = 7$, $\alpha = 0.05$;
- $\bar{X} = 97.4$, $\bar{Y} = 110$, $S_X^2 = 78.8$, $S_Y^2 = 913.33$,
- $c = \frac{S_X^2/m}{S_X^2/m + S_Y^2/n} = 0.11$,
- $r = 1/\left[\frac{c^2}{m-1} + \frac{(1-c)^2}{n-1}\right] = 6.26$
- $$E = t_{.025}(6)\sqrt{\frac{78.8}{5} + \frac{913.3}{7}} = 34.49$$
- The 95% CI:
$$(\bar{X} - \bar{Y}) \pm E = (-47.09, 21.89)$$

Solution of Example 4:

- $m = 5$, $n = 7$, $\alpha = 0.05$;
- $\bar{X} = 97.4$, $\bar{Y} = 110$, $S_X^2 = 78.8$, $S_Y^2 = 913.33$,
- $c = \frac{S_X^2/m}{S_X^2/m + S_Y^2/n} = 0.11$,
- $r = 1/\left[\frac{c^2}{m-1} + \frac{(1-c)^2}{n-1}\right] = 6.26$
- $$E = t_{.025}(6)\sqrt{\frac{78.8}{5} + \frac{913.3}{7}} = 34.49$$
- The 95% CI:

$$(\overline{X} - \overline{Y}) \pm E = (-47.09, 21.89)$$

Solution of Example 4:

- $m = 5$, $n = 7$, $\alpha = 0.05$;
- $\bar{X} = 97.4$, $\bar{Y} = 110$, $S_X^2 = 78.8$, $S_Y^2 = 913.33$,
- $c = \frac{S_X^2/m}{S_X^2/m + S_Y^2/n} = 0.11$,
- $r = 1/\left[\frac{c^2}{m-1} + \frac{(1-c)^2}{n-1}\right] = 6.26$
- $$E = t_{.025}(6)\sqrt{\frac{78.8}{5} + \frac{913.3}{7}} = 34.49$$
- The 95% CI:
$$(\overline{X} - \overline{Y}) \pm E = (-47.09, 21.89)$$

Solution of Example 4:

- $m = 5$, $n = 7$, $\alpha = 0.05$;
- $\bar{X} = 97.4$, $\bar{Y} = 110$, $S_X^2 = 78.8$, $S_Y^2 = 913.33$,
- $c = \frac{S_X^2/m}{S_X^2/m + S_Y^2/n} = 0.11$,
- $r = 1/\left[\frac{c^2}{m-1} + \frac{(1-c)^2}{n-1}\right] = 6.26$
- $$E = t_{.025}(6)\sqrt{\frac{78.8}{5} + \frac{913.3}{7}} = 34.49$$
- The 95% CI:
$$(\overline{X} - \overline{Y}) \pm E = (-47.09, 21.89)$$

Solution of Example 4:

- $m = 5$, $n = 7$, $\alpha = 0.05$;
- $\bar{X} = 97.4$, $\bar{Y} = 110$, $S_X^2 = 78.8$, $S_Y^2 = 913.33$,
- $c = \frac{S_X^2/m}{S_X^2/m + S_Y^2/n} = 0.11$,
- $r = 1/\left[\frac{c^2}{m-1} + \frac{(1-c)^2}{n-1}\right] = 6.26$
-
$$E = t_{.025}(6)\sqrt{\frac{78.8}{5} + \frac{913.3}{7}} = 34.49$$
- The 95% CI:
$$(\bar{X} - \bar{Y}) \pm E = (-47.09, 21.89)$$

Solution of Example 4:

- $m = 5$, $n = 7$, $\alpha = 0.05$;
- $\bar{X} = 97.4$, $\bar{Y} = 110$, $S_X^2 = 78.8$, $S_Y^2 = 913.33$,
- $c = \frac{S_X^2/m}{S_X^2/m + S_Y^2/n} = 0.11$,
- $r = 1/\left[\frac{c^2}{m-1} + \frac{(1-c)^2}{n-1}\right] = 6.26$
- $$E = t_{.025}(6)\sqrt{\frac{78.8}{5} + \frac{913.3}{7}} = 34.49$$
- The 95% CI:

$$(\overline{X} - \overline{Y}) \pm E = (-47.09, 21.89)$$

Outline

Example 5. (Reaction Time to Red and Green)

Eight subjects were involved in an experiment was conducted to compare people's reaction times to a red light versus a green light. The reaction times in seconds were recorded as follows.

Subject	Red(X)	Green(Y)	$D = X - Y$
1	0.30	0.43	−0.13
2	0.23	0.32	−0.09
3	0.41	0.58	−0.17
4	0.53	0.46	0.07
5	0.24	0.27	−0.03
6	0.36	0.41	−0.05
7	0.38	0.38	0.00
8	0.51	0.61	−0.10

Find a 90% confidence interval for $\mu_X - \mu_Y$.

Example 5. (Reaction Time to Red and Green)

Eight subjects were involved in an experiment was conducted to compare people's reaction times to a red light versus a green light. The reaction times in seconds were recorded as follows.

Subject	Red(X)	Green(Y)	$D = X - Y$
1	0.30	0.43	-0.13
2	0.23	0.32	-0.09
3	0.41	0.58	-0.17
4	0.53	0.46	0.07
5	0.24	0.27	-0.03
6	0.36	0.41	-0.05
7	0.38	0.38	0.00
8	0.51	0.61	-0.10

Find a 90% confidence interval for $\mu_X - \mu_Y$.

The CI for $\mu_X - \mu_Y$ based Paired Data

- Let $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ be n independent pairs of (dependent) r.v.'s.
- Assume that $D_i = X_i - Y_i$, $i = 1, 2, \ldots, n$, is a random sample from $N(\mu_D, \sigma_D^2)$, $\mu_D = \mu_X - \mu_Y$.
- A $1-\alpha$ confidence interval for $\mu_X - \mu_Y$:

$$\overline{D} \pm t_{\alpha/2}(n-1)\frac{S_D}{\sqrt{n}},$$

where

$$\overline{D} = \frac{1}{n}\sum_{i=1}^{n} D_i, \quad S_D^2 = \frac{1}{n-1}\sum_{i=1}^{n}(D_i - \overline{D})^2.$$

The CI for $\mu_X - \mu_Y$ based Paired Data

- Let $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ be n independent pairs of (dependent) r.v.'s.
- Assume that $D_i = X_i - Y_i$, $i = 1, 2, \ldots, n$, is a random sample from $N(\mu_D, \sigma_D^2)$, $\mu_D = \mu_X - \mu_Y$.
- A $1-\alpha$ confidence interval for $\mu_X - \mu_Y$:

$$\overline{D} \pm t_{\alpha/2}(n-1)\frac{S_D}{\sqrt{n}},$$

where

$$\overline{D} = \frac{1}{n}\sum_{i=1}^{n} D_i, \quad S_D^2 = \frac{1}{n-1}\sum_{i=1}^{n}(D_i - \overline{D})^2.$$

The CI for $\mu_X - \mu_Y$ based Paired Data

- Let $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ be n independent pairs of (dependent) r.v.'s.
- Assume that $D_i = X_i - Y_i$, $i = 1, 2, \ldots, n$, is a random sample from $N(\mu_D, \sigma_D^2)$, $\mu_D = \mu_X - \mu_Y$.
- A $1 - \alpha$ confidence interval for $\mu_X - \mu_Y$:

$$\overline{D} \pm t_{\alpha/2}(n-1)\frac{S_D}{\sqrt{n}},$$

where

$$\overline{D} = \frac{1}{n}\sum_{i=1}^{n} D_i, \quad S_D^2 = \frac{1}{n-1}\sum_{i=1}^{n}(D_i - \overline{D})^2.$$

The CI for $\mu_X - \mu_Y$ based Paired Data

- Let $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ be n independent pairs of (dependent) r.v.'s.
- Assume that $D_i = X_i - Y_i$, $i = 1, 2, \ldots, n$, is a random sample from $N(\mu_D, \sigma_D^2)$, $\mu_D = \mu_X - \mu_Y$.
- A $1 - \alpha$ confidence interval for $\mu_X - \mu_Y$:

$$\overline{D} \pm t_{\alpha/2}(n-1)\frac{S_D}{\sqrt{n}},$$

where

$$\overline{D} = \frac{1}{n}\sum_{i=1}^{n} D_i, \quad S_D^2 = \frac{1}{n-1}\sum_{i=1}^{n}(D_i - \overline{D})^2.$$

Solution of Example 5:

- Statistics: $n = 8$, $\overline{D} = -0.0625, S_D = 0.0765$

$$t_{\alpha/2}(n-1) = t_{.05}(7) = 1.895$$

- $E = 0.0512$
- The 90% confidence interval for $\mu_X - \mu_Y$:

$$-0.0625 \pm 0.0512 = [-0.1137, -0.0113].$$

- The CI does not include 0. We conclude that people react a red light faster.

Solution of Example 5:

- Statistics: $n = 8$, $\overline{D} = -0.0625, S_D = 0.0765$

$$t_{\alpha/2}(n-1) = t_{.05}(7) = 1.895$$

- $E = 0.0512$
- The 90% confidence interval for $\mu_X - \mu_Y$:

$$-0.0625 \pm 0.0512 = [-0.1137, -0.0113].$$

- The CI does not include 0. We conclude that people react a red light faster.

Solution of Example 5:

- Statistics: $n = 8$, $\overline{D} = -0.0625, S_D = 0.0765$

$$t_{\alpha/2}(n-1) = t_{.05}(7) = 1.895$$

- $E = 0.0512$
- The 90% confidence interval for $\mu_X - \mu_Y$:

$$-0.0625 \pm 0.0512 = [-0.1137, -0.0113].$$

- The CI does not include 0. We conclude that people react a red light faster.

Solution of Example 5:

- Statistics: $n = 8$, $\overline{D} = -0.0625, S_D = 0.0765$

$$t_{\alpha/2}(n-1) = t_{.05}(7) = 1.895$$

- $E = 0.0512$
- The 90% confidence interval for $\mu_X - \mu_Y$:

$$-0.0625 \pm 0.0512 = [-0.1137, -0.0113].$$

- The CI does not include 0. We conclude that people react a red light faster.