10.3 Two Population Means: σs NOT equal

GOALS:
1. Compare 2 sample means when the population standard deviations are not equal.
2. Use the distribution of the difference between the means to evaluate the samples.
3. Arrive at a conclusion: are the means from the same population or are they different?

Study Ch. 10.3, 67-70 all, 73-77 (no CI), 81, 83
[# 61-67, 71, 73]

Which to use, pooled or unpooled? JB Statistics

Class Notes: Prof. G. Battaly, Westchester Community College, NY

Statistics Home Page Class Notes Homework

10.3 Two Population Means: σs NOT equal

Cholesterol levels are measured for 28 randomly selected heart attack patients (2 days after their attacks) and 30 randomly selected other hospital patients who did not have a heart attack. The response is quantitative so we compare means. It is thought that cholesterol levels will be higher for the heart attack patients.

For the 28 heart attack patients, the mean cholesterol level was 253.9 with a standard deviation of 47.7. For the 30 other hospital patients who did not have a heart attack, the mean cholesterol level was 193.1 with a standard deviation of 22.3.

Assuming nd and a significance level of 5%, are the cholesterol levels for the heart attack patients higher than those for the other patients?

What kind of problem is this?
How is it similar to other problems we have done?
How is it different?

rt tailed test, non-pooled
t=6.147, p=1.85E-7,
df=37.68

What kind of problem is this? hypothesis test
How is it similar to other problems we have done? 2 samples - like pooled-t test
How is it different? not pooled-t because it does not meet assumption that $\sigma_1 = \sigma_2$

Class Notes: Prof. G. Battaly, Westchester Community College, NY

Statistics Home Page Class Notes Homework

10.3 Two Population Means: σs NOT equal

If we assume that the population standard deviations are equal, but they are not, how poor an outcome would we expect?

If we find a 95% confidence interval:

% of intervals that contain $\mu_1 - \mu_2$:

assuming equal standard deviations and $\sigma_1 = 10$ $\sigma_1 = 10$

n_1	n_2	$\sigma_2 = 2$	$\sigma_2 = 5$	$\sigma_2 = 10$	$\sigma_2 = 20$	$\sigma_2 = 50$
10	10	93.7	94.5	94.9	94.8	93.7
10	20	82.9	88.7	95.0	98.2	98.9
10	40	68.5	81.8	95.0	99.5	99.9

table from JB Statistics

1. Slightly less confidence if same sample sizes.
2. If different sample sizes:
   a) a **BIG** decrease in confidence if the larger sample has a smaller standard deviation.
   b) an increase in confidence if the larger sample has a larger standard deviation.

Conclude: Better to use Non-pooled if there is a question about equal population standard deviations; ie: not sure that $\sigma_1 = \sigma_2$ especially if the sample sizes are different.

Class Notes: Prof. G. Battaly, Westchester Community College, NY

Statistics Home Page Class Notes Homework

10.3 Two Population Means: σs NOT equal

NonPooled - t Test

Assumptions: 1. SRS, 2. Independent samples 3. ND or large samples 4. $\neq$ σ's

Step 1: H_0: $\mu_1 = \mu_2$

H_a: $\mu_1 \neq \mu_2$ or $\mu_1 < \mu_2$ or $\mu_1 > \mu_2$

Step 2: Decide α and Sketch

Step 3: Compute test statistic: $t = \dfrac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$

standard error

$\sigma_{\bar{x}} = \sigma/\sqrt{n}$ | $s/\sqrt{n}$

$\dfrac{\sqrt{s^2}}{\sqrt{n}}$

$\sqrt{\dfrac{s^2}{n}}$

Step 4: Find CV(s) OR Find p-value

Table IV using *df, Δ* :

round down to the nearest integer

$$\Delta = \frac{\left[\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}\right]^2}{\dfrac{(s_1^2/n_1)^2}{n_1 - 1} + \dfrac{(s_2^2/n_2)^2}{n_2 - 1}}$$

df different because estimating two different standard deviations as well as estimating the means for testing (Welch's pooled degrees of freedom.)

Step 5: Decide whether to reject H_0 or not $p \,?\, \alpha$

Step 6: Verbal interpretation

Class Notes: Prof. G. Battaly, Westchester Community College, NY

Statistics Home Page Class Notes Homework

10.3 Two Population Means: σs NOT equal

non-pooled

Pooled or Non-pooled Variances: = ? or Not ?

G: Seasgrass taller, thicker in Corpus Christy Bay than in Lower Laguna Madre. To compare ammonium concentrations, collected random samples. A normal probability plot of the Laguna Madre data was approximately a straight line.

At the 1% significance level, is the mean ammonium concentration in Corpus Christy Bay larger than that in Lower Laguna Madre?

$$\bar{x}_c = 115.1, \quad s_c = 79.4, \quad n_c = 51$$
$$\bar{x}_L = 24.3, \quad s_L = 10.5, \quad n_L = 19$$

Approach:
What kind of problem is this? *"At 1% s.l..." -> hypothesis test*
Which procedure?
2 samples -> either pooled-t or non-pooled-t *(at this point)*
Need nd for both: n_C large, NPP for n_L
Need $\sigma_1 = \sigma_2$ for pooled-t but 79.4 ≉ 10.5 -> NON-pooled-t

Calculator:2 Sample t Test
STAT/TESTS
2-SampTTest
Stats/
mean1: ___
s1: ___
n1: ___
mean2: ___
s2: ___
n2: ___
μ1: ≠μ2 <μ2 >μ2
Pooled: No Yes
calculate

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$\Delta = \frac{\left[\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right]^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}}$$

- If sample sizes are very different and s_1 and s_2 are different, **do not use** the pooled procedure.

Class Notes: Prof. G. Battaly, Westchester Community College, NY

Statistics Home Page Class Notes Homework

10.3 Two Population Means: σs NOT equal

non-pooled

G: Seasgrass taller, thicker in Corpus Christy Bay than in Lower Laguna Madre. To compare ammonium concentrations, collected random samples. A normal probability plot of the Laguna Madre data was approximately a straight line.

At the 1% significance level, is the mean ammonium concentration in Corpus Christy Bay larger than that in Lower Laguna Madre?

$$\bar{x}_c = 115.1, \quad s_c = 79.4, \quad n_c = 51$$
$$\bar{x}_L = 24.3, \quad s_L = 10.5, \quad n_L = 19$$

Calculator:2 Sample t Tes

STAT/TESTS
2-SampTTest
Stats/
mean1: ___
s1: ___
n1: ___
mean2: ___
s2: ___
n2: ___
μ1: ≠μ2 <μ2 >μ2
Pooled: No Yes
calculate

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$\Delta = \frac{\left[\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right]^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}}$$

Class Notes: Prof. G. Battaly, Westchester Community College, NY

Statistics Home Page Class Notes Homework

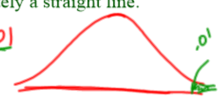
10.3 Two Population Means: σs NOT equal non-pooled

G: Seasgrass taller, thicker in Corpus Christy Bay than in Lower Laguna Madre. To compare ammonium concentrations, collected random samples. A normal probability plot of the Laguna Madre data was approximately a straight line.

Independent

At the 1% significance level, is the mean ammonium concentration in Corpus Christy Bay larger than that in Lower Laguna Madre?

$\alpha = 0.01$



Non-pooled-t

②

$\bar{x}_c = 115.1, \quad s_c = 79.4, \quad n_c = 51$

$\bar{x}_L = 24.3, \quad s_L = 10.5, \quad n_L = 19$

① $H_0: \mu_c = \mu_L$

$H_a: \mu_c > \mu_L$

③ $t = \dfrac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}} = \dfrac{115.1 - 24.3}{\sqrt{\dfrac{79.4^2}{51} + \dfrac{10.5^2}{19}}} = 7.182$

$\Delta = \dfrac{\left[\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}\right]^2}{\dfrac{(s_1^2/n_1)^2}{n_1 - 1} + \dfrac{(s_2^2/n_2)^2}{n_2 - 1}} = 54.47$

$54 = df$

④ $p = 4.98(10^{-11})$
$= 0.0000$

⑤ $p = 0.0000 < 0.01 = \alpha \therefore$ rej. H_0

⑥ There is very strong evidence.
Concl. Concentration of ammonium at C is greater than at L.

Class Notes: Prof. G. Battaly, Westchester Community College, NY

Statistics Home Page Class Notes Homework

10.3 Two Population Means: σs NOT equal

students set up

G: Efforts to reduce disabilities of bus drivers by improving bus routes (intervention) to alleviate stress. Heart rates of drivers on improved routes were compared to that of drivers on normal routes (control).

G: Independent - diff. drivers, srs, nd

At the 5% significance level, does the intervention reduce the mean heart rate in bus drivers?

$\bar{x}_I = 67.90, \quad s_I = 5.49, \quad n_I = 10$

$\bar{x}_c = 66.81, \quad s_c = 9.04, \quad n_c = 31$

Calculator: 2 Sample t Test
STAT/TESTS
2-SampTTest
Stats/
mean1: ___
s1: ___
n1: ___
mean2: ___
s2: ___
n2: ___
μ1: ≠μ2 <μ2 >μ2
Pooled: No Yes
calculate

Approach:
What kind of problem is this? *"At 1% s.l..." -> hypothesis test*
Which procedure?
2 samples -> either pooled-t or non-pooled-t (at this point)
Need nd for both: Given
Need $\sigma_1 = \sigma_2$ for pooled-t but $n_c > n_I$ and $s_c > s_I$ so pooled would be conservative; use NON-pooled

$\Delta = \dfrac{\left[\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}\right]^2}{\dfrac{(s_1^2/n_1)^2}{n_1 - 1} + \dfrac{(s_2^2/n_2)^2}{n_2 - 1}}$

Class Notes: Prof. G. Battaly, Westchester Community College, NY

Statistics Home Page Class Notes Homework

10.3 Two Population Means: σs NOT equal

students set up

G: Efforts to reduce disabilities of bus drivers by improving bus routes (intervention) to alleviate stress. Heart rates of drivers on improved routes were compared to that of drivers on normal routes (control). (Independent - diff. drivers, srs, nd)

At the 5% significance level, does the intervention reduce the mean heart rate in bus drivers?

$\bar{x}_I = 67.90, \quad s_I = 5.49, \quad n_I = 10$

$\bar{x}_c = 66.81, \quad s_c = 9.04, \quad n_c = 31$

① $H_0: \mu_I = \mu_c$ $\quad H_a: \mu_I < \mu_c$ $\quad$ ② $\alpha = 0.05$

③ $$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{67.90 - 66.81}{\sqrt{\frac{5.49^2}{10} + \frac{9.04^2}{31}}} = 0.459$$

$$\Delta = \frac{\left[\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right]^2}{\left[\frac{s_1^2}{n_1}\right]^2 + \left[\frac{s_2^2}{n_2}\right]^2} = \frac{\left(\frac{5.49^2}{10} + \frac{9.04^2}{31}\right)^2}{\left(\frac{5.49^2}{10}\right)^2 + \left(\frac{9.04^2}{31}\right)^2} = 25.74$$

$df = 25$

④ $p = 0.675$

⑤ $p = 0.675 > \alpha = 0.05$ Do NOT rej. H_0.

⑥ Conclude that data is insufficient to support the intervention as a means of reducing mean heart rate of bus drivers.

$$\Delta = \frac{\left[\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right]^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}}$$

Class Notes: Prof. G. Battaly, Westchester Community College, NY

Statistics Home Page Class Notes Homework

10.3 Two Population Means: σs NOT equal

Cholesterol levels are measured for 28 heart attack patients (2 days after their attacks) and 30 other hospital patients who did not have a heart attack. The response is quantitative so we compare means. It is thought that cholesterol levels will be higher for the heart attack patients.

For the 28 heart attack patients, the mean cholesterol level was 253.9 with a standard deviation of 47.7.
For the 30 other hospital patients who did not have a heart attack, the mean cholesterol level was 193.1 with a standard deviation of 22.3.

Assuming nd and a significance level of 5%, are the cholesterol levels for the heart attack patients higher than those for the other patients?

rt tailed test, non-pooled
t=6.147, p=1.85E-7,
df=37.68

Class Notes: Prof. G. Battaly, Westchester Community College, NY

Statistics Home Page Class Notes Homework

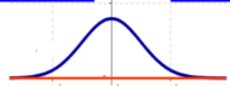
10.3 Two Population Means: σs NOT equal

Cholesterol levels are measured for 28 heart attack patients (2 days after their attacks) and 30 other hospital patients who did not have a heart attack. The response is quantitative so we compare means. It is thought that cholesterol levels will be higher for the heart attack patients.

For the 28 heart attack patients, the mean cholesterol level was 253.9 with a standard deviation of 47.7. For the 30 other hospital patients who did not have a heart attack, the mean cholesterol level was 193.1 with a standard deviation of 22.3.

Assuming nd and a significance level of 5%, **are the cholesterol levels for the heart attack patients higher than those for the other patients**?

H_o: ______ H_a: ______



$\bar{x}_1 = 253.9 \quad s_1 = 47.7 \quad n_1 = 28$

$\bar{x}_2 = 193.1 \quad s_2 = 22.3 \quad n_2 = 30$

Use non-pooled t-test:

because $s_1 > 2\, s_2$.

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$\Delta = \frac{\left[\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right]^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}}$$

Class Notes: Prof. G. Battaly, Westchester Community College, NY

Statistics Home Page Class Notes Homework

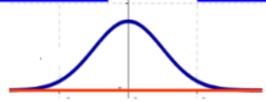
10.3 Two Population Means: σs NOT equal

Cholesterol levels are measured for 28 heart attack patients (2 days after their attacks) and 30 other hospital patients who did not have a heart attack. The response is quantitative so we compare means. It is thought that cholesterol levels will be higher for the heart attack patients.

For the 28 heart attack patients, the mean cholesterol level was 253.9 with a standard deviation of 47.7. For the 30 other hospital patients who did not have a heart attack, the mean cholesterol level was 193.1 with a standard deviation of 22.3.

Assuming nd and a significance level of 5%, **are the cholesterol levels for the heart attack patients higher than those for the other patients**?

H_o: ______ H_a: ______



$\bar{x}_1 = 253.9 \quad s_1 = 47.7 \quad n_1 = 28$

$\bar{x}_2 = 193.1 \quad s_2 = 22.3 \quad n_2 = 30$

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{253.9 - 193.1}{\sqrt{\frac{47.7^2}{28} + \frac{22.3^2}{30}}} = 6.147 \qquad df = 37.7 \sim 37$$

$p = 1.85(10^{-7}) = 0.0000$

$p = 0.0000 < 0.05 = \alpha \quad \therefore \text{rej } H_o$

Conclude: cholesterol levels of heart attach patients is higher than patient w/o heart attacks.

Class Notes: Prof. G. Battaly, Westchester Community College, NY

Statistics Home Page Class Notes Homework

10.3 Two Population Means: σs NOT equal class

EXTRAS

G: independent SRS, n.d. F: Use non-pooled t test:

a) 2-tailed α=0.05

G: $\bar{x}_1 = 15$, $s_1 = 2$, $n_1 = 15$
$\bar{x}_2 = 12$, $s_2 = 5$, $n_2 = 15$

Calculator: 2 Sample t Test
STAT/TESTS
2-SampTTest
Stats/
mean1: ___
s1: ___
n1: ___
mean2: ___
s2: ___
n2: ___
μ1: ≠μ2 <μ2 >μ2
Pooled: No Yes
calculate

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$\Delta = \frac{\left[\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right]^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}}$$

10.3 Two Population Means: σs NOT equal class

EXTRAS

G: independent SRS, n.d. F: Use non-pooled t test:

a) 2-tailed α=0.05

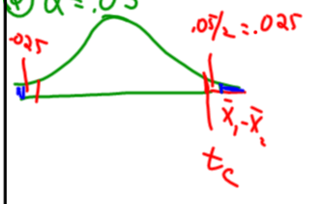
G: $\bar{x}_1 = 15$, $s_1 = 2$, $n_1 = 15$
$\bar{x}_2 = 12$, $s_2 = 5$, $n_2 = 15$

3 $$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{15-12}{\sqrt{\frac{2^2}{15} + \frac{5^2}{15}}} = 2.158$$

1 $H_0: \mu_1 = \mu_2$

$H_a: \mu_1 \neq \mu_2$

2 $\alpha = .05$



4 $p = 0.0444$

5 $p = 0.0444 < \alpha = 0.05$

rej. Ho.

6 evid. is suff. to rej. hyp. that means are eq.

$$\Delta = \frac{\left[\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right]^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}} = df = 18.37 = \underline{18}$$

(0.083, 5.917)

10.3 Two Population Means: σs NOT equal

students

EXTRAS

G: independent SRS, n.d. F: Use non-pooled t test:

a) left-tailed α=0.05 (NOTE: can also use pooled-t because std dev not very different)

G: $\bar{x}_1 = 20$, $s_1 = 4$, $n_1 = 10$
$\bar{x}_2 = 23$, $s_2 = 5$, $n_2 = 15$

Calculator: 2 Sample t Test
STAT/TESTS
2-SampTTest
Stats/
mean1: ___
s1: ___
n1: ___
mean2: ___
s2: ___
n2: ___
μ1: ≠μ2 <μ2 >μ2
Pooled: No Yes
calculate

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$\Delta = \frac{\left[\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right]^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}}$$

10.3 Two Population Means: σs NOT equal

students

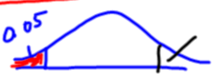
EXTRAS

G: independent SRS, n.d. F: Use non-pooled t test:

a) left-tailed α=0.05

① $H_0: \mu_1 = \mu_2$ $H_a: \mu_1 < \mu_2$

② $\alpha = 0.05$



G: $\bar{x}_1 = 20$, $s_1 = 4$, $n_1 = 10$
$\bar{x}_2 = 23$, $s_2 = 5$, $n_2 = 15$

③ $$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{20 - 23}{\sqrt{\frac{4^2}{10} + \frac{5^2}{15}}} = -1.660$$

df: 22.0999
df = 22

④ $p = 0.055$

⑤ $p = 0.055 > \alpha = 0.05$

⑥ Do NOT rej. Ho.

⑦ Data is not suff to claim that $\mu_1 < \mu_2$

$$\Delta = \frac{\left[\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right]^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}}$$

10.3 Two Population Means: σs NOT equal

EXTRAS

Review: Number of young per litter. 24 Florida, 44 Virginia.

At the 1% sign. level, do data provide sufficient evidence to conclude that litter size in Florida is less than that in Virginia?

G: ~ nd, srs, independent, σ's =

G: FL $\bar{x}_1 = 5.46$, $s_1 = 1.59$, $n_1 = 24$

VA $\bar{x} = 7.59$, $s_2 = 2.68$, $n_2 = 44$

Calculator:2 Sample t Test
STAT/TESTS
2-SampTTest
Stats/
mean1: ___
s1: ___
n1: ___
mean2: ___
s2: ___
n2: ___
μ1: ≠μ2 <μ2 >μ2
Pooled: No Yes
calculate

Class Notes: Prof. G. Battaly, Westchester Community College, NY

Statistics Home Page Class Notes Homework