10.1 Sampling Distribution of Differences between 2 Independent sample means
10.2 Two Population Means: =σ's ***Pooled t-test***
10.3 Two Population Means: σs NOT equal ***Non-Pooled t-test***

GOALS:
1. Consider how two samples can be compared to determine if they are the same or different or come from the same or different populations.
2. Consider the distribution of the difference of sample means, including: - the mean of the difference
- the standard deviation of the difference

3. Use the Pooled-t Test to compare sample means when $\sigma_1 = \sigma_2$

4. Use the Non-Pooled t Test to compare means when $\sigma_1 \neq \sigma_2$

Read Ch. 10.1, Study Key Fact 10.1
Study Ch. 10.2, # 33-43, 48, 49
Study Ch. 10.3, 67-70 all, 73-77(no CI), 81, 83

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

Statistics Home Page Class Notes

10.1 Sampling Distribution of Differences between 2 independent sample means

G: * random samples of 30 males and 30 females
* tested frame of reference, pointed S, error recorded
* table of pointing errors, in degrees, with:

MALE					FEMALE				
13	68	60	22	30	14	78	18	32	80
130	18	5	70	8	8	69	35	35	91
39	3	9	58	20	20	111	111	12	68
33	11	59	3	67	3	3	109	27	66
10	38	5	167	26	138	128	36	8	176
13	23	86	15	19	122	31	27	3	15

$\bar{X}_M = 37.6$ $\bar{X}_F = 55.8$

$S_M = 38.5$ $S_F = 48.3$

<- New is 2 sample test. Previous (below) is 1-sample test.

G: BTUs consumed/household/year in US:
$\mu = 92.2$ mill BTU, n.d., $\sigma = 15$ mill BTU
n = 20 household in West US

104	84	72	95	69
80	78	74	76	81
82	61	94	65	100
70	65	83	76	84

F: Do households in the West US use a different amount of energy?

F: At the 1% significance level, do the data provide sufficient evidence to conclude that, on average, males have a better sense of direction and, in particular, a better frame of reference than females?

from 10.2

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

Statistics Home Page Class Notes

10.1 Sampling Distribution of Differences between 2 independent sample means

How do we start?

Know that mean for males is ~ nd for large samples.
Know that mean for females is ~ nd for large samples.

How do we compare them?

$\overline{X}_M = 37.6$ $\overline{X}_F = 55.8$

$S_M = 38.5$ $S_F = 48.3$

We see that: mean for males < mean for females

or $\overline{X}_M < \overline{X}_F$

or $\overline{X}_M - \overline{X}_F < 0$

$\overline{X}_1 = \overline{X}_2$ $\overline{X}_1 > \overline{X}_2$

$\overline{X}_1 - \overline{X}_2 = 0$ $\overline{X}_1 - \overline{X}_2 > 0$

$\overline{X}_1 - \overline{X}_2$ $\overline{X}_1 - \overline{X}_2$

We need to examine the distribution of

$\overline{X}_M - \overline{X}_F$

In general: the distribution of $\overline{X}_1 - \overline{X}_2$

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

Statistics Home Page Class Notes

10.1 Sampling Distribution of Differences between 2 independent sample means

the distribution of $\overline{X}_1 - \overline{X}_2$

If x is ~ nd on each of the populations 1 and 2, then $\overline{x}_1 - \overline{x}_2$ is ~nd and:

$$\mu_{\overline{x}_1 - \overline{x}_2} = \mu_1 - \mu_2$$

$$\sigma_{\overline{x}_1 - \overline{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

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

Statistics Home Page Class Notes

10.1 Sampling Distribution of Differences between 2 independent sample means

the distribution of $\overline{X}_1 - \overline{X}_2$

If x is ~ nd on each of the populations 1 and 2, then $\overline{X}_1 - \overline{X}_2$ is ~nd and:

$$\mu_{\overline{X}_1 - \overline{X}_2} = \mu_1 - \mu_2$$

$$\sigma_{\overline{X}_1 - \overline{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

comparison

Mean of the Sample Mean

$$\mu_{\overline{x}} = \mu$$

Standard Deviation of the Sample Mean

$$\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}}$$

Standard Error (of the Mean)

$$\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}} = \sqrt{\frac{\sigma^2}{n}}$$

$$\sigma = \sqrt{\sigma^2}$$

$$\sqrt{\frac{a}{b}} = \frac{\sqrt{a}}{\sqrt{b}}$$

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

Statistics Home Page Class Notes

10.1 Sampling Distribution of Differences between 2 independent sample means

When σ is known, can standardize to SNC,

But **rarely have a known σ, so will not consider a z test**

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

$$z = \frac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

No, z test

Instead, either of two *t*-tests: **σ_1, σ_2 unknown**

Yes, 2 t - tests

- $\sigma_1 = \sigma_2$ --> pooled ***t***

  s_1 and s_2 are estimates of same σ

- $\sigma_1 \neq \sigma_2$ --> non-pooled ***t***

  s_1 and s_2 are not known to estimate same σ

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

Statistics Home Page Class Notes

10.2 Two Population Means: = σ's

Instead, either of two *t*-tests:

- $\sigma_1 = \sigma_2$ --> pooled ***t***

  s_1 and s_2 are estimates of same σ

Using s_1 and s_2 as estimates of same σ, s_p is computed by weighting each sample s by the size of the sample it represents.

$$s_p = s_{pooled} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

$$df = n_1 + n_2 - 2$$

No need to memorize. Use textbook, ***pooled-t test***, and copy formula down in your solution.

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

Statistics Home Page Class Notes Homework

10.2 Two Population Means: = σ's

$\sigma_1 \neq \sigma_2$ Non-Pooled t Test

s_1 and s_2 are estimates of σ_1 and σ_2, so there are actually 4 variables: s_1, s_2 $\bar{X}_1$ $\bar{X}_2$

This requires a different way to compute the **Degrees of Freedom**

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

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

Statistics Home Page Class Notes Homework

10.2 Two Population Means: = σ's

Pooled - t Test

Assumptions: 1. SRS, 2. Independent samples
3. ND or large samples 4. = σ's 🙂

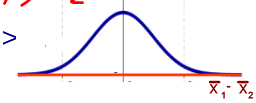
NonPooled - t Test

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

Step 1: H_0: $\mu_1 = \mu_2$ $\mu_1 - \mu_2 = 0$

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

$\mu_1 - \mu_2 \neq 0$
$\mu_1 - \mu_2 < 0$
$\mu_1 - \mu_2 > 0$



Step 2: Decide α and sketch

Step 3: Compute test statistic: $t_T = \dfrac{(\bar{x}_1 - \bar{x}_2)}{s_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$

where

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

$df = n_1 + n_2 - 2$

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

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

Step 4: Find P-value

Step 5: Decide whether to reject H_0 or not: Reject H_o if $p \leq \alpha$

Step 6: Verbal interpretation NOT Reject H_o if $p > \alpha$

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

Statistics Home Page Class Notes Homework

10.2 Two Population Means: = σ's

G: * random samples of 30 males and 30 females
* tested frame of reference, pointed S, error recorded
* table of pointing errors, in degrees, with:

MALE					FEMALE				
13	68	60	22	30	14	78	18	32	80
130	18	5	70	8	8	69	35	35	91
39	3	9	58	20	20	111	111	12	68
33	11	59	3	67	3	3	109	27	66
10	38	5	167	26	138	128	36	8	176
13	23	86	15	19	122	31	27	3	15

$\bar{x}_M = 37.6$ $\bar{x}_F = 55.8$

$s_M = 38.5$ $s_F = 48.3$

F: At the 1% significance level, do the data provide sufficient evidence to conclude that, on average, males have a better sense of direction and, in particular, a better frame of reference than females?

Pooled? or Non-Pooled?

Pooled -

Assumptions: 1. S
3. ND or large

Step 1: H_0: $\mu_1 = \mu$
H_a: $\mu_1 \neq \mu$

Step 2: Decide α ar

Step 3: Compute tes

where

$s_p = \sqrt{\dfrac{(n_1 - 1)s_1^2}{n_1 +}}$

Step 4: Find P-valu

Step 5: Decide whe

Step 6: Verbal inter

test

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

Statistics Home Page Class Notes Homework

10.2 Two Population Means: = σ's

3: ND or large samples 4: = σ's

G: * random samples of 30 males and 30 females
* tested frame of reference, pointed S, error recorded
* table of pointing errors, in degrees, with:

MALE					FEMALE				
13	68	60	22	30	14	78	18	32	80
130	18	5	70	8	8	69	35	35	91
39	3	9	58	20	20	111	111	12	68
33	11	59	3	67	3	3	109	27	66
10	38	5	167	26	138	128	36	8	176
13	23	86	15	19	122	31	27	3	15

df = 58

$\bar{X}_M = 37.6$ $s_M = 38.5$

$\bar{X}_F = 55.8$ $s_F = 48.3$

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: $t = \dfrac{(\bar{x}_1 - \bar{x}_2)}{s_p\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}$

where $s_p = \sqrt{\dfrac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}$

Step 4: Find P-value OR Find CV(s)
df = $n_1 + n_2 - 2$ Table IV

Step 5: Decide whether to reject H_0 or not

Step 6: Verbal interpretation

F: At the 1% significance level, do the data provide sufficient evidence to conclude that, on average, males have a better sense of direction and, in particular, a better frame of reference than females?

σ = ?
$\dfrac{48.3}{38.5} \approx 1.25$
not very different

① $H_0: \mu_M = \mu_F$ $H_a: \mu_M < \mu_F$

② $\alpha = 0.01$

.01

$t_c = -2.39$ -1.614 P-.056

③ $t = \dfrac{\bar{x}_M - \bar{x}_F}{s\sqrt{\frac{1}{n_M}+\frac{1}{n_F}}} = \dfrac{37.6 - 55.8}{43.676\sqrt{\frac{1}{30}+\frac{1}{30}}} = -1.614$

$s_p = \sqrt{\dfrac{(n_M-1)s_M^2 + (n_F-1)s_F^2}{n_M+n_F-2}} = \sqrt{\dfrac{29(38.5)^2 + 29(48.3)^2}{58}} = 43.676$

④ $p = 0.056$

⑤ $p = 0.056 > .01 = \alpha$
∴ Do NOT rej. H_0.

⑥ At the 1% s.l., males do not have a better frame of reference than females.

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

Statistics Home Page Class Notes Homework

10.3 Two Population Means: σs NOT equal

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?

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

Pooled? or Non-Pooled?

s2:
n2:
μ1: ≠μ2 <μ2 >
Pooled: No Ye
calculate

non-pool

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

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

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

Statistics Home Page Class Notes Homework

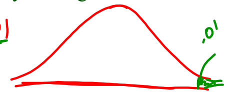
10.3 Two Population Means: σs NOT equal non-pooled

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

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

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

① $H_0: \mu_c = \mu_L$

$H_a: \mu_c > \mu_L$

③ $$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{115.1 - 24.3}{\sqrt{\frac{79.4^2}{51} + \frac{10.5^2}{19}}} = 7.982$$

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

$54 = df$

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.

10.2 Two Population Means: = σ's

Researchers are investigating how the amount of protein in the diet relates to weight gain. They randomly select 19 female rats, and consider their gain in weight between 28 and 84 days after birth. 12 were fed a high protein diet and 7 were fed a low protein diet. At the 95% confidence level, does the high protein diet relate to a higher weight gain?

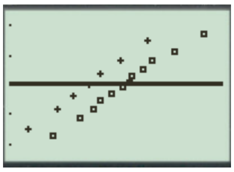
High protein	134	146	104	119	124	161	107	83	113	129	97	123
Low protein	70	118	101	85	107	132	94					

10.2 Two Population Means: = σ's

Researchers are investigating how the amount of protein in the diet relates to weight gain. They randomly select 19 female rats, and consider their gain in weight between 28 and 84 days after birth. 12 were fed a high protein diet and 7 were fed a low protein diet. At the 95% confidence level, does the high protein diet relate to a higher weight gain?

High protein	134	146	104	119	124	161	107	83	113	129	97	123
Low protein	70	118	101	85	107	132	94					

Assumpt: srs, nd?, $\sigma_1 = \sigma_2$? NPP first, confirm nd and $\sigma_1 = \sigma_2$, then finish



~nd ✓

✓ $\sigma_1 = \sigma_2$

```
1-Var Stats
 x̄=120
 Σx=1440
 Σx²=177832
 Sx=21.38818705
 σx=20.47763007
↓n=12
```

High Protein

```
1-Var Stats
 x̄=101
 Σx=707
 Σx²=73959
 Sx=20.62361106
 σx=19.09375365
↓n=7
```

Low Protein

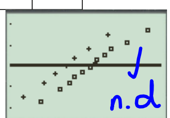
Can use pooled -t procedure.

10.2 Two Population Means: = σ's

Researchers are investigating how the amount of protein in the diet relates to weight gain. They randomly select 19 female rats, and consider their gain in weight between 28 and 84 days after birth. 12 were fed a high protein diet and 7 were fed a low protein diet. At the 95% confidence level, does the high protein diet relate to a higher weight gain?

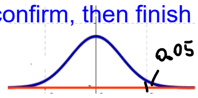
High protein	134	146	104	119	124	161	107	83	113	129	97	123
Low protein	70	118	101	85	107	132	94					

Assumpt: srs, nd?, $\sigma_1 = \sigma_2$? NPP first, confirm, then finish



$H_0: \mu_1 = \mu_2$ $\alpha = 1 - .95 = .05$

$H_a: \mu_1 > \mu_2$



$df = n_1 + n_2 - 2 = 17$

```
2-SampTTest
 Inpt:Data Stats
 List1:L1
 List2:L2
 Freq1:1
 Freq2:1
 μ1:≠μ2 <μ2 >μ2
↓Pooled:No Yes
```

Can avoid 1-variable stats and go straight to 2-sample t-test
BUT, be sure to check s_1 and s_2 before proceeding.

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{120-101}{21.121\sqrt{\frac{1}{12}+\frac{1}{7}}} = 1.891$$

```
2-SampTTest
 μ1>μ2
 t=1.891436397
 p=.0378650634
 df=17
 x̄1=120
↓x̄2=101
↑Sx1=21.3881871
 Sx2=20.6236111
 Sxp=21.1214973
 n1=12
 n2=7
```

$\sigma_1 = \sigma_2$

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

$$= \sqrt{\frac{11(21.39)^2 + 6(20.62)^2}{12+7-2}}$$

$= 21.121$

$p = 0.038 < 0.05 = \alpha$

reject H_0

Conclude: The high protein diet relates to increased weight gain.

(Note: Cannot reject if a 2-tailed test: p= 2(0.038) = 0.0757 > 0.05)

10.3 Two Population Means: σs NOT equal

randomly select who answers

True or False

True or False:

1. pooled-t requires independent samples, but non-pooled t does not.
2. pooled-t requires that the standard deviations can be assumed to be equal, but non-pooled t does not.
3. the non-pooled t test requires degrees of freedom but the pooled t test does not.
4. the pooled t test pools the sample means
5. reject the null hypothesis when $p > \alpha$

1. F, 2. T, 3. F, 4. F, 5. F

1. False. Both require independent samples.
2. True
3. False. All t tests require df.
4. False. The pooled t test pools the sample standard deviations.
5. False. Reject the null hypothesis whenever $p \le \alpha$

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

Statistics Home Page Class Notes Homework