

Chapter 18 – Comparing two population means – Independent samples

In chapter 17, we considered the case of Matched pairs – dependent samples

Review: One-sample and Matched Pairs Problems

One-sample problems
 Inference about one population
 One SRS
 One measurement on each individual
 C.I.: $\bar{x} \pm t^* \frac{s}{\sqrt{n}}$ Test Statistic: $t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$ $df = n - 1$

Matched pairs problems
 Inference about one population of pairs
 One SRS of paired individuals with one measurement on each individual within the pair

OR
 Inference about one population
 One SRS with two measurements on each individual, randomizing if possible (e.g., before and after)

Compute differences between observed values in each pair.
 C.I.: $\bar{d} \pm t^* \frac{s_d}{\sqrt{n}}$ Test Statistic: $t = \frac{\bar{d} - 0}{s_d/\sqrt{n}}$ $df = n - 1$

B: Before
A: After

In this chapter 18, we compare two population means from independent samples

Two-Sample Problems

Compare **two populations** or **two treatments**

Two Populations (Survey)
 Take two separate SRS's. (Stratified sample with 2 strata)
 One measurement on each individual.

Sample 1 from Population 1 Sample 2 from Population 2

Two Treatments (Experiment)
 Randomly divide individuals into two treatment groups.
 One measurement on each individual.

Random Allocation → Treatment 1
 Random Allocation → Treatment 2

Examples of Two-Sample Problems

Compare mean weight of filling of double stuffed with mean weight of filling of regular oreos.

$n_{\text{Regular Oreo}} = 49$
 $n_{\text{Double stuffed}} = 52$

Fail to reject H_0

Notation for Comparing Two Population Means

	One population		One Sample	
Mean	μ		\bar{x}	
Standard deviation	σ		S	
	Two populations		Two Samples	
Means	μ_1	μ_2	\bar{x}_1	\bar{x}_2
Standard deviations	σ_1	σ_2	S ₁	S ₂

How should we combine two parameters to get one for comparison?

Sum	$\mu_1 + \mu_2?$	☹	Not informative
Difference	$\mu_1 - \mu_2?$	☺	Informative and mathematically nice
Product	$\mu_1 \cdot \mu_2?$	☹	Not informative
Ratio	$\mu_1 / \mu_2?$	☹	Informative but not mathematically nice

What is the corresponding statistic? $\bar{x}_1 - \bar{x}_2$

Notation for Comparing Two Population Means

	One population		One Sample	
Mean	μ		\bar{x}	
Standard deviation	σ		S	
	Two populations		Two Samples	
Means	μ_1	μ_2	\bar{x}_1	\bar{x}_2
Standard deviations	σ_1	σ_2	S ₁	S ₂

How should we combine two parameters to get one for comparison?

$\mu_1 - \mu_2$ In words: Difference between two population means

$\bar{x}_1 - \bar{x}_2$ In words: Difference between two sample means

What is the corresponding statistic? $\bar{x}_1 - \bar{x}_2$

Two-Sample Problems

Compare **two populations** or **two treatments**

Two Populations (Surveys)

Take two separate SRS's from each of two distinct populations.

Measure same variable on individuals in both samples.

Perform test of hypothesis on $H_0: \mu_1 - \mu_2 = 0$.

If significant, compute a confidence interval to estimate $\mu_1 - \mu_2$.

Two Treatments (Experiments)

Randomly divide individuals into two groups.

Apply different treatment to each group.

Measure same variable on individuals in both treatment groups.

Perform a test of hypothesis on $H_0: \mu_{T1} - \mu_{T2} = 0$.

T1 = Treatment 1 and T2 = Treatment 2

Steps for Two-Sample t Test of Significance

Conservative Method (for use without software)

Step 1 STATE: Describe the problem.

Step 2 PLAN: Recognize need for comparing two means.
Specify H_0 and H_a in terms of μ_1 and μ_2 ; choose α .

Step 3 SOLVE: Collect data--random allocation or 2 independent SRS's.
Plot both data sets; compute \bar{x}_1 , \bar{x}_2 , s_1 , and s_2 .
Check: data collection ok and no outliers if $n_1 + n_2 < 40$.

Calculate test statistic: $t = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$ $df = \min(n_1 - 1, n_2 - 1)$

Obtain P -value from t table.

Step 4 CONCLUDE: Draw conclusions in context.

Using software is recommended to get more accurate P-values.

We will use the calculator to test hypothesis or construct confidence intervals for two population means

Randomized Design Example (Two Treatments)

Conservative Method (for use without software)

Step 1 STATE the problem and identify essential information.

A pharmaceutical company is conducting pre-clinical trials of an experimental anti-depressant drug. Since several subjects are complaining of dryness, a technician is assigned to investigate with 20 rats. She plans to randomly allocate 10 rats to receive the drug injection and 10 rats to receive a placebo injection and measure their water intake during the next 24 hour period.

Using $\alpha = 0.05$, the research question is:

"Does the anti-depressant cause an increase in water consumption?"

Step 2 PLAN: Need two-sample t -test for means.

Specify H_0 and H_a ; choose α .

$H_0: \mu_D - \mu_P = 0$ versus $H_a: \mu_D - \mu_P > 0$

Type of Alternative:	greater than
Null value ($\mu_D - \mu_P$):	0
n_D :	10
n_P :	10
α :	0.05

Randomized Design Example (Two Treatments)

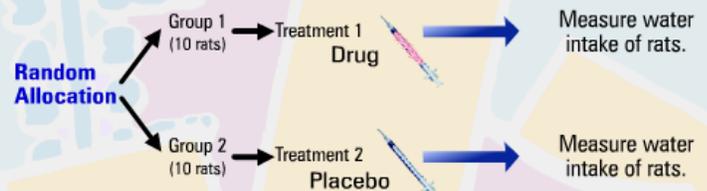
Conservative Method (for use without software)

Step 3 SOLVE: Collect data by conducting experiment.

Randomized Design Example (Two Treatments)

Conservative Method (for use without software)

Step 3 SOLVE: Collect data by conducting experiment.



Randomized Design Example (Two Treatments)

Conservative Method (for use without software)

Step 3 SOLVE: Plot each data set, compute \bar{x}_D , \bar{x}_P and s_D and s_P .

Drug	Placebo
8.2	8.5
9.0	7.2
8.8	7.7
8.3	8.0
7.5	8.6
9.2	7.8
9.9	8.7
7.8	7.8
8.4	8.0
7.7	7.0

	Drug	Placebo
\bar{x}	8.48	7.93
s	0.750	0.564
n	10	10

Side-By-Side Stemplot	
Leaf unit = 0.1	
Drug	Placebo
	7 0 2
5 7 8	7 7 8 8
2 3 4	8 0 0
8	8 5 6 7
0 2	9
9	9

Randomized Design Example (Two Treatments)

Conservative Method (for use without software)

Step 3 SOLVE: (cont.) Check conditions

Rats randomly assigned to treatments;
no outliers or strong skewness in either
data set; use of t procedure is ok.

Randomized Design Example (Two Treatments)

Conservative Method (for use without software)

Step 3 SOLVE: (cont.) Calculate test statistic

$$\bar{x}_D - \bar{x}_P = 8.48 - 7.93 = 0.55$$

$$t = \frac{(\bar{x}_D - \bar{x}_P) - 0}{\sqrt{\frac{s_D^2}{n_D} + \frac{s_P^2}{n_P}}} = \frac{(8.48 - 7.93) - 0}{\sqrt{\frac{0.750^2}{10} + \frac{0.564^2}{10}}} = \frac{0.55}{0.297} = 1.85$$

	Drug	Placebo
\bar{x}	8.48	7.93
s	0.750	0.564
n	10	10

Randomized Design Example (Two Treatments)

Conservative Method (for use without software)

Step 3 SOLVE: (cont.) Find P -value for $t=1.85$.

$$df = \min(10 - 1, 10 - 1) = \min(9, 9)$$

$$.05 > P > .025$$

9	0.703	0.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
One-sided P	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
	1.833 < 1.85 < 2.262											

Randomized Design Example (Two Treatments)

Conservative Method (for use without software)

Step 3 SOLVE: (cont.) Find P -value for $t=1.85$.

$$df = \min(10 - 1, 10 - 1) = \min(9, 9)$$

Step 4 CONCLUDE: Draw conclusions in context.

$$P\text{-value: } .025 < P < .05 = \alpha \text{ (level of significance)}$$

Sufficient evidence to reject the null hypothesis at $\alpha = 0.05$

Conclusion in context: At $\alpha = 0.05$, the average water intake for the rats in the drug group is significantly greater than the average water intake for the rats in the placebo group.

Valid experiment: The anti-depressant drug causes an increase in thirst in rats.

90% Confidence Interval for $\mu_1 - \mu_2$ **Conservative Method (for use without software)**

$$\bar{x}_1 - \bar{x}_2 \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \rightarrow 8.48 - 7.93 \pm 1.833 \sqrt{\frac{0.750^2}{10} + \frac{0.564^2}{10}}$$

$$\text{Confidence level} = 90\% \rightarrow 0.55 \pm 1.833 (0.297)$$

$$\text{df} = \min(10 - 1, 10 - 1) = 9 \rightarrow (0.006, 1.094)$$

	Drug	Placebo
\bar{x}	8.48	7.93
s	0.750	0.564
n	10	10

90% Confidence Interval for $\mu_1 - \mu_2$ **Conservative Method (for use without software)**

$$\bar{x}_1 - \bar{x}_2 \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \rightarrow 8.48 - 7.93 \pm 1.833 \sqrt{\frac{0.750^2}{10} + \frac{0.564^2}{10}}$$

$$\text{Confidence level} = 90\% \rightarrow 0.55 \pm 1.833 (0.297)$$

$$\text{df} = \min(10 - 1, 10 - 1) = 9 \rightarrow (0.006, 1.094)$$

	Drug	Placebo
\bar{x}	8.48	7.93
s	0.750	0.564
n	10	10

Interpretation in context: The difference between the true mean water intake of rats given the drug and the true mean water intake of rats given the placebo is somewhere between 0.006 and 1.094 ml with 90% confidence.

Note: The confidence interval does not include 0; hence $\mu_D \neq \mu_P$, confirming our conclusion from the test of significance.

Increase in water intake

90% Confidence Interval for $\mu_1 - \mu_2$ **Conservative Method (for use without software)**

$$\bar{x}_1 - \bar{x}_2 \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \rightarrow 8.48 - 7.93 \pm 1.833 \sqrt{\frac{0.750^2}{10} + \frac{0.564^2}{10}}$$

$$\text{Confidence level} = 90\% \rightarrow 0.55 \pm 1.833 (0.297)$$

$$\text{df} = \min(10 - 1, 10 - 1) = 9 \rightarrow (0.006, 1.094)$$

	Drug	Placebo
\bar{x}	8.48	7.93
s	0.750	0.564
n	10	10

Interpretation in context: The difference between the true mean water intake of rats given the drug and the true mean water intake of rats given the placebo is somewhere between 0.006 and 1.094 ml with 90% confidence.

Note: The confidence interval does not include 0; hence $\mu_D \neq \mu_P$, confirming our conclusion from the test of significance.