STAT 100 Lecture 24: Comparison of Two Populations Part 2: Matched Pair Comparisons

Nate Strawn

November 14

Last Time...

1. Independent Samples from Two Populations
   - Large Sample Confidence Intervals and Hypothesis Tests
   - Small Sample Confidence Intervals and Hypothesis Tests when $\sigma_1 = \sigma_2 = \sigma$

Today's Agenda

1. Independent Samples from Two Populations
   - Small Sample Inferences when $\sigma_1 \neq \sigma_2$
2. Matched Pairs Design
3. Examples

Last time, we considered the case there $\sigma_1 = \sigma_2$.

Small Sample Inferences for $\mu_1 - \mu_2$ When the Populations Are Normal But σ_1 and σ_2 Are Not Assumed to Be Equal

A $100(1-\alpha)\%$ conservative confidence interval for $\mu_1 - \mu_2$ is given by

$$\overline{X} - \overline{Y} \pm t_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

where $t_{\alpha/2}$ denotes the upper $\alpha/2$ point of the t distribution with d.f. $=$ smaller of $n_1 - 1$ and $n_2 - 1$.

The null hypothesis $H_0 : \mu_1 - \mu_2 = \delta_0$ is tested using the test statistic

$$T^* = \frac{\overline{X} - \overline{Y} - \delta_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \qquad \text{d.f.} = \text{smaller of } n_1 - 1 \text{ and } n_2 - 1$$

Example: A 95% Confidence Interval for $\mu_1 - \mu_2$

Example

Suppose $n_1 = 25$ and $n_2 = 36$ are the number of exam grades from Section 1 and Section 2 (for the same exam) respectively, and suppose the following summary statistics are obtained:

	Section 1	*Section 2*
$\overline{x}$	*68.2*	*64.5*
S	*10.4*	*8.9*

1. *Compute a 95% confidence interval for the difference in average exam grades, $\mu_1 - \mu_2$.*
2. *Test the One-Sided hypothesis $H_0 : \mu_1 - \mu_2 = 1$ versus $H_1 : \mu_1 - \mu_2 > 1$ at the $\alpha = 0.05$ level.*

Matching Pairs

Matched Pairs Design

Matched pair	Experimental units	
1	2	1
2	1	2
3	1	2
⋮	⋮	⋮
n	2	1

Units in each pair are alike, whereas units in different pairs may be dissimilar. In each pair, a unit is chosen at random to receive treatment 1, the other unit receives treatment 2.

Data Arising from Matched Pairs Design

Structure of Data for a Matched Pair Comparison

Pair	Treatment 1	Treatment 2	Difference
1	X_1	Y_1	$D_1 = X_1 - Y_1$
2	X_2	Y_2	$D_2 = X_2 - Y_2$
.	.	.	.
.	.	.	.
.	.	.	.
n	X_n	Y_n	$D_n = X_n - Y_n$

The differences $D_1, D_2, \ldots, D_n$ are a random sample.
Summary statistics:

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

Inferences About Large Samples of Matched Pairs

When n is large, the Central Limit Theorem applies to the differences $D_1, \ldots, D_n$. Thus,

$$Z = \frac{\overline{D} - \delta}{S_D/\sqrt{n}}$$

is approximately distributed as $N(0,1)$. We can now apply our earlier techniques to get confidence intervals and hypothesis tests concerning the mean of D, δ.

Example: Confidence Interval and Hypothesis Test for δ

Example

Consider the following summary statistics for data arising from a matched pairs design:

	n	*mean*	*sd*
differences	*64*	*5.2*	*2.1*

1. *Compute a 99% confidence interval for* δ.
2. *Perform a hypothesis test for* $H_0 : \delta = 3$ *versus* $H_1 : \delta > 3$ *at the* $\alpha = 0.05$ *level.*

Inferences about Small Samples of Matched Pairs

Small Samples Inferences about the Mean Difference δ

Assume that the differences $D_i = X_i - Y_i$ are a random sample from an $N(\delta, \sigma_D)$ distribution. Let

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

Then:

1. A $100(1 - \alpha)\%$ confidence interval for δ is given by

$$\left(\overline{D} - t_{\alpha/2} \frac{S_D}{\sqrt{n}}, \quad \overline{D} + t_{\alpha/2} \frac{S_D}{\sqrt{n}} \right)$$

   where $t_{\alpha/2}$ is based on $n - 1$ degrees of freedom.

2. A test of $H_0 : \delta = \delta_0$ is based on the test statistic

$$T = \frac{\overline{D} - \delta_0}{S_D / \sqrt{n}} \qquad \text{d.f.} = n - 1$$

Example: Confidence Interval and Hypothesis Test for δ

Example

Consider the following summary statistics for data arising from a matched pairs design:

	n	*mean*	*sd*
differences	*16*	*5.2*	*2.1*

1. *Compute a 95% confidence interval for* δ.
2. *Perform a hypothesis test for* $H_0 : \delta = 1$ *versus* $H_1 : \delta > 1$ *at the* $\alpha = 0.25$ *level.*

For Next Time

- Read Section 10.4 from Johnson and Bhattacharyya
- Group Problems:

Group	1	2	3	4	5
Problem	10.33	10.35	10.37	10.39	10.41
Group	1	2	3	4	5
Problem	10.33	10.35	10.37	10.39	10.41