More on Inference for Two-Sample Data

Yifan Zhu

Iowa State University

Outline

Two-sample inference

- Comparing the means of two distinct populations with respect to the same measurement.
- Examples:
  - SAT scores of high school A vs. high school B.
  - Severity of a disease in women vs. in men.
  - Heights of New Zealanders vs. heights of Ethiopians.
  - Coefficients of friction after wear of sandpaper A vs. sandpaper B.
- Notation:

Sample	1	2
Sample size	n_1	n_2
True mean	μ_1	μ_2
Sample mean	$\overline{x}_1$	$\overline{x}_2$
True variance	σ_1^2	σ_2^2
Sample variance	s_1^2	s_2^2

$n_1 \geq 25$ and $n_2 \geq 25$, variances known

- We want to test $H_0 : \mu_1 - \mu_2 = \#$ with some alternative hypothesis
- If σ_1^2 and σ_2^2 are known, use the test statistic:

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

  which has a $N(0, 1)$ distribution if:
  - H_0 is true.
  - The sample 1 points are iid with mean μ_1 and variance σ_1^2, the sample 2 points are iid with mean μ_2 and variance σ_2^2, and the two samples are independent.
- The confidence intervals (2-sided, 1-sided upper, and 1-sided lower, respectively) for $\mu_1 - \mu_2$ are:

$$\left((\overline{x_1} - \overline{x}_2) - z_{1-\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}},\ (\overline{x_1} - \overline{x}_2) + z_{1-\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right)$$

$$\left(-\infty,\ (\overline{x_1} - \overline{x}_2) + z_{1-\alpha}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right)$$

$$\left((\overline{x_1} - \overline{x}_2) - z_{1-\alpha}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}},\ \infty \right)$$

$n_1 \geq 25$ and $n_2 \geq 25$, variances UNknown

- If σ_1^2 and σ_2^2 are UNknown, use the test statistic:

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

- and confidence intervals for $\mu_1 - \mu_2$:

$$\left((\overline{x_1} - \overline{x}_2) - z_{1-\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}},\ (\overline{x_1} - \overline{x}_2) + z_{1-\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right)$$

$$\left(-\infty,\ (\overline{x_1} - \overline{x}_2) + z_{1-\alpha}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right)$$

$$\left((\overline{x_1} - \overline{x}_2) - z_{1-\alpha}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}},\ \infty \right)$$

Outline

Two-Sample Inference: Large Samples

Two-Sample Inference: Small samples

Small samples and $\sigma_1^2 = \sigma_2^2 = \sigma^2$ (UNknown)

- Assuming $\sigma_1^2 = \sigma_2^2 = \sigma^2$, then we can use the **pooled sample variance** to estimate σ^2,

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

- A test statistic to test $H_0 : \mu_1 - \mu_2 = \#$ against some alternative is:

$$T = \frac{\overline{x}_1 - \overline{x}_2 - \#}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

- $T \sim t_{n_1+n_2-2}$ assuming:
  - H_0 is true.
  - The sample 1 points are iid $N(\mu_1, \sigma^2)$, the sample 2 points are iid $N(\mu_2, \sigma^2)$, and the sample 1 points are independent of the sample 2 points.

Small samples and $\sigma_1^2 = \sigma_2^2 = \sigma^2$ (UNknown)

- $1-\alpha$ confidence intervals (2-sided, 1-sided upper, and 1-sided lower, respectively) for $\mu_1 - \mu_2$ under these assumptions are of the form:

$$\left((\overline{x_1} - \overline{x}_2) - t_{\nu,\ 1-\alpha/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}},\ (\overline{x_1} - \overline{x}_2) + t_{\nu,\ 1-\alpha/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right)$$

$$\left(-\infty,\ (\overline{x_1} - \overline{x}_2) + t_{\nu,\ 1-\alpha} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right)$$

$$\left((\overline{x_1} - \overline{x}_2) - t_{\nu,\ 1-\alpha} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}},\ \infty\right)$$

where $\nu = n_1 + n_2 - 2$.

Example: springs

- The data of W. Armstrong on spring lifetimes (appearing in the book by Cox and Oakes) not only concern spring longevity at a 950 N/mm2 stress level but also longevity at a 900 N/mm2 stress level.

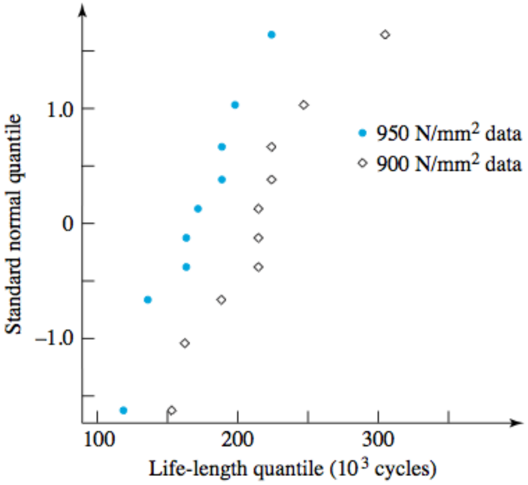
Spring Lifetimes under Two Different Levels of Stress (10^3 cycles)

950 N/mm^2 Stress	900 N/mm^2 Stress
225, 171, 198, 189, 189	216, 162, 153, 216, 225
135, 162, 135, 117, 162	216, 306, 225, 243, 189

- Let sample 1 be the 900 N/mm^2 stress group and sample 2 be the 950 N/mm^2 stress group.
- $\overline{x}_1 = 215.1, \overline{x}_2 = 168.3$.
- Let's do a hypothesis test to see if the sample 1 springs lasted significantly longer than the sample 2 springs.

Check normality and homogeneity of variances

Make a normal Q-Q plot of both sample on the same plot. If both sample look like a straight line and these two lines are almost parallel, then it is plausible that both sample are normally distributed with equal variancec.



Example: springs

1. $H_0 : \mu_1 - \mu_2 = 0$, $H_a : \mu_1 - \mu_2 > 0$.
2. $\alpha = 0.05$
3. The test statistic is:

$$T = \frac{(\overline{x}_1 - \overline{x}_2) - 0}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

- Assume:
  - H_0 is true.
  - The sample 1 spring lifetimes are iid $N(\mu_1, \sigma^2)$
  - The sample 2 spring lifetimes are iid $N(\mu_2, \sigma^2)$
  - The sample 1 spring lifetimes are independent of those of sample 2.
- Under these assumptions, $T \sim t_{n_1+n_2-2} = t_{10+10-2} = t_{18}$.
- Reject H_0 if $T > t_{18,\ 1-\alpha}$

Example: springs

$$s_1 = \sqrt{\frac{1}{n_1 - 1} \sum_i (x_{1,i} - \overline{x}_1)^2}$$

$$= \sqrt{\frac{1}{9}(225 - 215.1)^2 + (171 - 215.1)^2 + \cdots + (162 - 215.1)^2} = 42.9$$

$$s_2 = \sqrt{\frac{1}{n_2 - 1} \sum_i (x_{2,i} - \overline{x}_2)^2}$$

$$= \sqrt{\frac{1}{9}(225 - 168.3)^2 + (171 - 168.3)^2 + \cdots + (162 - 168.3)^2} = 33.1$$

$$s_p = \sqrt{\frac{(10-1)42.9^2 + (10-1)33.1^2}{10 + 10 - 2}} = 38.3$$

Example: springs

4.

$$t = \frac{(\overline{x}_1 - \overline{x}_2) - 0}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{215.1 - 168.3 - 0}{38.3 \cdot \sqrt{\frac{1}{10} + \frac{1}{10}}} = 2.7$$

$$\begin{aligned} t_{18,\ 1-\alpha} &= t_{18,\ 1-0.05} = t_{18,\ 0.95} \\ &= 1.73 \end{aligned}$$

5. With $t = 2.7 > 1.73 = t_{18,0.95}$, we reject H_0 in favor of H_a.
6. There is enough evidence to conclude that springs last longer if subjected to 900 N/mm^2 of stress than if subjected to 950 N/mm^2 of stress.

Example: springs

- A 95%, 2-sided confidence interval for the difference in lifetimes is:

$$\left((\overline{x_1} - \overline{x}_2) - t_{\nu,\ 1-\alpha/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}},\ (\overline{x_1} - \overline{x}_2) + t_{\nu,\ 1-\alpha/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right)$$

Using $t_{\nu,\ 1-\alpha/2} = t_{18,1-0.05/2} = t_{18,\ 0.975} = 2.1$:

$$\left((215.1 - 168.3) - 2.1 \cdot 38.3\sqrt{\frac{1}{10} + \frac{1}{10}}, (215.1 - 168.3) + 2.1 \cdot 38.3\sqrt{\frac{1}{10} + \frac{1}{10}}\right)$$
$$= (10.8, 82.8)$$

- We are 95% confident that the springs subjected to 900 N/mm^2 of stress last between 10.8×10^3 and 82.8×10^3 cycles longer than the springs subjected to 950 N/mm^2 of stress.

Your turn: stopping distances

- Suppose μ_1 and μ_2 are true mean stopping distances (in meters) at 50 mph for cars of a certain type equipped with two different types of breaking systems.
- Suppose $n_1 = n_2 = 6, \overline{x}_1 = 115.7, \overline{x}_2 = 129.3, s_1 = 5.08, s_2 = 5.38$.
- Use significance level 0.01 to test $H_0 : \mu_1 - \mu_2 = -10$ vs. $H_a : \mu_1 - \mu_2 < -10$.
- Construct a 2-sided 99% confidence interval for the true difference in stopping distances.

Answers: stopping distances

1. $H_0 : \mu_1 - \mu_2 = -10$, $H_a : \mu_1 - \mu_2 < -10$.
2. $\alpha = 0.01$
3. The test statistic is:

$$T = \frac{(\overline{x}_1 - \overline{x}_2) - (-10)}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

- Assume:
  - H_0 is true.
  - The sample 1 stopping distances are iid $N(\mu_1, \sigma^2)$
  - The sample 2 stopping distances are iid $N(\mu_2, \sigma^2)$
  - The sample 1 stopping distances are independent of those of sample 2.
- Under these assumptions, $T \sim t_{n_1+n_2-2} = t_{6+6-2} = t_{10}$.
- Reject H_0 if $T < t_{10,\ \alpha}$

Answers: stopping distances

- $s_1 = 5.08$, $s_2 = 5.38$.
- $$s_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}$$
$$= \sqrt{\frac{(6-1)(5.08)^2 + (6-1)(5.38)^2}{6+6-2}}$$
$$= 5.23$$

Answers: stopping distances

4.

$$t = \frac{(\overline{x}_1 - \overline{x}_2) - (-10)}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{115.7 - 129.3 + 10}{5.23 \cdot \sqrt{\frac{1}{6} + \frac{1}{6}}} = -1.19$$

$$t_{10,\ 1-\alpha} = t_{10,\ 0.99} = -2.76$$

5. With $t = -1.19 \not< -2.76 = t_{10,0.99}$, we reject H_0 in favor of H_a.
6. There is not enough evidence to conclude that the stopping distances of breaking system 1 are less than those of breaking system 2 by over 10 meters.

Answers: stopping distances

- A 99%, 2-sided confidence interval for the difference in breaking distances is:

$$\left((\overline{x_1} - \overline{x}_2) - t_{\nu,\ 1-\alpha/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}},\ (\overline{x_1} - \overline{x}_2) + t_{\nu,\ 1-\alpha/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right)$$

Using $t_{\nu,\ 1-\alpha/2} = t_{10,1-0.01/2} = t_{10,\ 0.995} = 3.17$:

$$\left((115.7 - 129.3) - 3.17 \cdot 5.23\sqrt{\frac{1}{6} + \frac{1}{6}}, (115.7 - 129.3) + 3.17 \cdot 5.23\sqrt{\frac{1}{6} + \frac{1}{6}}\right)$$
$$= (-23.17, -4.03)$$

- We are 99% confident that the true mean stopping distance of breaking system 1 is anywhere from 23.17 m to 4.03 m less than that of breaking system 2.

What if $\sigma_1^2 \neq \sigma_2^2$?

- The test statistic for testing $H_0 : \mu_1 - \mu_2 = \#$ vs. some H_a is:

$$T = \frac{\overline{x}_1 - \overline{x}_2 - \#}{\sqrt{\frac{s_2^2}{n_2} + \frac{s_1^2}{n_1}}}$$

Under the assumptions that:

  - H_0 is true.
  - The sample 1 observations are iid $N(\mu_1, \sigma_1^2)$ and the sample 2 observations are iid $N(\mu_2, \sigma_2^2)$

The test statistic has an approximate $t_{\hat{\nu}}$ distribution, where the degrees of freedom is estimated by the following special case of the Cochran-Satterthwaite approximation for linear combinations of mean squares:

$$\hat{\nu} = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{s_1^4}{(n_1-1)n_1^2} + \frac{s_2^4}{(n_2-1)n_2^2}}$$

What if $\sigma_1^2 \neq \sigma_2^2$?

- Under these assumptions, the $1-\alpha$ confidence intervals for $\mu_1 - \mu_2$ become:

$$\left((\overline{x_1} - \overline{x}_2) - t_{\widehat{\nu},\ 1-\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}},\ (\overline{x_1} - \overline{x}_2) + t_{\widehat{\nu},\ 1-\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}\right)$$

$$\left(-\infty,\ (\overline{x_1} - \overline{x}_2) + t_{\widehat{\nu},\ 1-\alpha}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}\right)$$

$$\left((\overline{x_1} - \overline{x}_2) - t_{\widehat{\nu},\ 1-\alpha}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}},\ \infty\right)$$

Example: springs

- In the springs example, σ_1^2 probably doesn't equal σ_2^2 because $s_1 = 57.9$ and $s_2 = 33.1$.
- I'll redo the hypothesis test and the confidence interval using:

$$\widehat{\nu} = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{s_1^4}{(n_1-1)n_1^2} + \frac{s_2^4}{(n_2-1)n_2^2}} = \frac{\left(\frac{57.9^2}{10} + \frac{33.1^2}{10}\right)^2}{\frac{57.9^4}{(10-1)10^2} + \frac{33.1^4}{(10-1)10^2}} = 14.3$$

Example: springs

1. $H_0 : \mu_1 - \mu_2 = 0,\ H_a : \mu_1 - \mu_2 > 0.$
2. $\alpha = 0.05$
3. The test statistic is:

$$T = \frac{(\overline{x}_1 - \overline{x}_2) - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

- Assume:
    - H_0 is true.
    - The sample 1 spring lifetimes are $N(\mu_1, \sigma_1^2)$
    - The sample 2 spring lifetimes are $N(\mu_2, \sigma_2^2)$
    - The sample 1 spring lifetimes are independent of those of sample 2.
- Under these assumptions, $T \sim t_{\widehat{\nu}} = t_{14.3}$.
- Reject H_0 if $T > t_{14.3,\ 1-\alpha}$

Example: springs

4. The moment of truth:

$$t = \frac{(\overline{x}_1 - \overline{x}_2) - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{215.1 - 168.3 - 0}{\sqrt{\frac{57.9^2}{10} + \frac{33.1^2}{10}}} = 2.22$$

$$\begin{aligned} t_{14.3,\ 1-\alpha} &= t_{14.3,\ 1-0.05} = t_{14.3,\ 0.95} \\ &= 1.76 \quad \text{(Take } \nu = 14 \text{ if you're using the t table)} \end{aligned}$$

5. With $t = 2.22 > 1.76 = t_{14.3,0.95}$, we reject H_0 in favor of H_a.
6. There is still enough evidence to conclude that springs last longer if subjected to 900 N/mm^2 of stress than if subjected to 950 N/mm^2 of stress.

Example: springs

- A 95%, 2-sided confidence interval for the difference in lifetimes is:

$$\left((\overline{x_1} - \overline{x}_2) - t_{\widehat{\nu},\ 1-\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}},\ (\overline{x_1} - \overline{x}_2) + t_{\widehat{\nu},\ 1-\alpha/2}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right)$$

Using $t_{\widehat{\nu},\ 1-\alpha/2} = t_{14.3,\ 1-0.05/2} = t_{14.3,0.975} = 2.14$:

$$\left((215.1 - 168.3) - 2.14 \cdot \sqrt{\frac{57.9^2}{10} + \frac{33.1^2}{10}}, \right.$$

$$\left. (215.1 - 168.3) + 2.14 \cdot \sqrt{\frac{57.9^2}{10} + \frac{33.1^2}{10}} \right)$$

$$= (1.67,\ 91.9)$$

- We are 95% confident that the springs subjected to 900 N/mm^2 of stress last between 1.67×10^3 and 91.1×10^3 cycles longer than the springs subjected to 950 N/mm^2 of stress.

Your turn: fabrics

- The void volume within a textile fabric affects comfort, flammability, and insulation properties. Permeability ($cm^3/cm^2/s$) of a fabric refers to the accessibility of void space to the flow of a gas or liquid.
- Consider the following data on two different types of plain-weave fabric:

Fabric Type	Sample Size	Sample Mean	Sample Standard Deviation
Cotton	10	51.71	.79
Triacetate	10	136.14	3.59

- Let Sample 1 be the triacetate fabric and Sample 2 be the cotton fabric.
- Using $\alpha = 0.05$, attempt to verify the claim that triacetate fabrics are more permeable than the cotton fabrics on average.
- Construct and interpret a two-sided 95% confidence interval for the true difference in mean permeability.

Answers: fabrics

- $n_1 = n_2 = 10$.
- $\overline{x}_1 = 136.14$, $\overline{x}_2 = 51.71$.
- $s_1 = 3.59$, $s_2 = 0.79$.
-
$$\widehat{\nu} = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{s_1^4}{(n_1-1)n_1^2} + \frac{s_2^4}{(n_2-1)n_2^2}} = \frac{\left(\frac{3.59^2}{10} + \frac{0.79^2}{10}\right)^2}{\frac{3.59^4}{(10-1)10^2} + \frac{0.79^4}{(10-1)10^2}} = 9.87$$
- If you're using the t table, round down to $\nu = 9$ to avoid unneccessary false positives.

Answers fabrics

1. $H_0 : \mu_1 - \mu_2 = 0$, $H_a : \mu_1 - \mu_2 > 0$.
2. $\alpha = 0.05$
3. The test statistic is:

$$T = \frac{(\overline{x}_1 - \overline{x}_2) - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

- Assume:
  - H_0 is true.
  - The triacetate permeabilities are $N(\mu_1, \sigma_1^2)$
  - The cotton permeabilities are $N(\mu_2, \sigma_2^2)$
  - The triacetate permeabilities are independent of the cotton permeabilities.
- Under these assumptions, $T \sim t_{\widehat{\nu}} = t_{9.87}$.
- Reject H_0 if $T > t_{9.87,\ 1-\alpha}$

Answers fabrics

4.

$$t = \frac{(\overline{x}_1 - \overline{x}_2) - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{136.14 - 51.71 - 0}{\sqrt{\frac{3.59^2}{10} + \frac{0.79^2}{10}}} = 72.63$$

$$t_{9.87,\ 1-\alpha} \approx t_{9,1-\alpha} = t_{9,\ 0.95} = 1.83$$

5. With $t = 72.63 > 1.83 = t_{9,0.95}$, we reject H_0 in favor of H_a.
6. There is overwhelming evidence to conclude that the triacetate fabrics are more permeable than the cotton fabrics.

Answers fabrics

- With $t_{\widehat{\nu},1-\alpha/2} \approx t_{9,0.975} = 2.26$, a 95%, 2-sided confidence interval for the difference in lifetimes is:

$$\left((\overline{x_1} - \overline{x}_2) - t_{\widehat{\nu},\ 1-\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}},\ (\overline{x_1} - \overline{x}_2) + t_{\widehat{\nu},\ 1-\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right)$$

$$\left((136.14 - 51.71) - 2.26 \cdot \sqrt{\frac{3.59^2}{10} + \frac{0.79^2}{10}}, \right.$$

$$\left. (136.14 - 51.71) + 2.26 \cdot \sqrt{\frac{3.59^2}{10} + \frac{0.79^2}{10}} \right)$$

$$= (81.80,\ 87.06)$$

- We are 95% confident that the permeability of the triacetate fabric exceeds that of the cotton fabric by anywhere between 81.80 $cm^3/cm^2/s$ and 87.06 $cm^3/cm^3/s$.