Inference for Matched Pairs and Two-Sample Data

Will Landau

Iowa State University

Apr 4, 2013

Outline

Matched Pairs

Two-Sample Inference: Large Samples

Matched pairs

- A **matched pairs** dataset is for which measurements naturally group into pairs.
- Examples:
  - Practice SAT scores before and after a prep course.
  - Severity of a disease before and after a treatment.
  - Leading edge measurement and trailing edge measurement for each workpiece in a sample.
  - Your height and the height of your friend, measured once each year for several years.
  - Bug bites on on right arm and bug bites on left arm (one has repellent and the other doesn't).

Example: fuel economy

- Twelve cars were equipped with radial tires and driven over a test course.
- Then the same 12 cars (with the same drivers) were equipped with regular belted tires and driven over the same course.
- After each run, the cars gas economy (in km/l) was measured.

	1	2	3	4	5	6
Radial	4.2	4.7	6.6	7.0	6.7	4.5
Belted	4.1	4.9	6.2	6.9	6.8	4.4
	7	8	9	10	11	12
Radial	5.7	6.0	7.4	4.9	6.1	5.2
Belted	5.7	5.8	6.9	4.7	6.0	4.9

- Using significance level $\alpha = 0.05$ and the method of critical values, test for a difference in fuel economy between the radial tires and belted tires.
- Construct a 95% confidence interval for true mean difference due to tire type.

Example: fuel economy

- First, calculate the differences (radial - belted):

	1	2	3	4	5	6
Radial	4.2	4.7	6.6	7.0	6.7	4.5
Belted	4.1	4.9	6.2	6.9	6.8	4.4
Difference	0.1	-0.2	0.4	0.1	-0.1	0.1

	7	8	9	10	11	12
Radial	5.7	6.0	7.4	4.9	6.1	5.2
Belted	5.7	5.8	6.9	4.7	6.0	4.9
Difference	0	0.2	0.5	0.2	0.1	0.3

- $\overline{d} = 0.142$, $s_d = 0.198$

Example: fuel economy

1. $H_0 : \mu_d = 0$, $H_a : \mu_d \neq 0$
2. $\alpha = 0.05$
3. I use the test statistic:

$$K = \frac{\overline{d} - 0}{s_d/\sqrt{n}}$$

which has a $t_{n-1} = t_{11}$ distribution, assuming:

- H_0 is true.
- $d_1, \ldots, d_{12}$ were independent draws from $N(\mu_d, \sigma_d^2)$
- I will reject H_0 if $|K| > |t_{11,1-\alpha/2}| = t_{11,0.975} = 2.20$

4. The moment of truth:

$$K = \frac{0.142}{0.198/\sqrt{12}} = 2.48$$

5. With $K = 2.48 > 2.20$, I reject H_0.
6. There is enough evidence to conclude that the fuel economy differs between radial tires and belted tires.

Example: fuel economy

- The two-sided 95% confidence interval for the true mean fuel economy difference is:

$$
\begin{aligned}
&= \left(\overline{d} - t_{11,1-\alpha/2}\frac{s_d}{\sqrt{n}},\ \overline{d} - t_{11,1-\alpha/2}\frac{s_d}{\sqrt{n}}\right) \\
&= \left(0.142 - t_{11,0.975}\frac{0.198}{\sqrt{12}},\ 0.142 + t_{11,0.975}\frac{0.198}{\sqrt{12}}\right) \\
&= (0.142 - 2.20 \cdot 0.057,\ 0.142 + 2.20 \cdot 0.057) \\
&= (0.0166,\ 0.2674)
\end{aligned}
$$

- We're 95% confident that for the car type studied, radial tires get between 0.0166 km/l and 0.2674 km/l more in fuel economy than belted tires.

Your Turn: wood product

- Consider the operation of an end-cut router in the manufacture of a company's wood product.
- Both a leading-edge and a trailing-edge measurement were made on each wooden piece to come off the router.

Leading-Edge and Trailing-Edge Dimensions for Five Workpieces

Piece	Leading-Edge Measurement (in.)	Trailing-Edge Measurement (in.)
1	.168	.169
2	.170	.168
3	.165	.168
4	.165	.168
5	.170	.169

- Is the leading edge measurement different from the trailing edge measurement for a typical wood piece? Do a hypothesis test at $\alpha = 0.05$ to find out.
- Make a two-sided 95% confidence interval for the true mean of the difference between the measurements.

Answers: wood product

- Take paired differences (leading edge - trailing edge).

Piece	d = Difference in Dimensions (in.)	
1	$-.001$	$(= .168 - .169)$
2	$.002$	$(= .170 - .168)$
3	$-.003$	$(= .165 - .168)$
4	$-.003$	$(= .165 - .168)$
5	$.001$	$(= .170 - .169)$

- The sample mean is $\overline{d} = -8 \times 10^{-4}$, and the sample standard deviation is $s_d = 0.0023$.
- Let μ_d be the true mean of the differences.

Answers: wood product

1. $H_0 : \mu_d = 0$, $H_a : \mu_d \neq 0$.
2. $\alpha = 0.05$, $n = 5$.
3. Since σ_d is unknown, I use the test statistic:

$$K = \frac{\overline{d} - 0}{s_d/\sqrt{n}}$$

   - Assume $d_1, \ldots, d_5 \sim N(\mu_d, \sigma_d^2)$
   - $K \sim t_{n-1} = t_4$.
   - Reject H_0 if $|K| > |t_{4,\ 1-\alpha/2}|$
4. The moment of truth:

$$K = \frac{-8 \times 10^{-4} - 0}{0.0023/\sqrt{5}} = -0.78$$

$$t_{4,1-\alpha/2} = t_{4,1-0.05/2} = t_{4,0.975} = 2.78$$

5. Since $|K| = 0.78 \ngtr 2.78 = t_{4,0.975}$, I fail to reject H_0.
6. There is not enough evidence to conclude that the leading edge measurements differ significantly from the trailing edge measurements.

Answers: wood product

- I can make a two-sided 95% confidence interval for μ_d in the usual way:

$$
\begin{aligned}
&\left(\overline{d} - t_{4,\ 1-\alpha/2} \cdot \frac{s}{\sqrt{n}},\ \overline{d} + t_{4,\ 1-\alpha/2} \cdot \frac{s}{\sqrt{n}}\right) \\
&= \left(-8 \times 10^{-4} - t_{4,0.975} \cdot \frac{0.0023}{\sqrt{5}},\ -8 \times 10^{-4} + t_{4,0.975} \cdot \frac{0.0023}{\sqrt{5}}\right) \\
&= \left(-8 \times 10^{-4} - 2.78 \cdot 0.0010,\ -8 \times 10^{-4} + 2.78 \cdot 0.0010\right) \\
&= (-0.00358, 0.00198)
\end{aligned}
$$

- We are 95% confident that the true mean difference between leading edge and trailing edge measurements is between -0.00358 in and 0.001298 in.

Outline

Matched Pairs

Two-Sample Inference: Large Samples

Two-sample inference

- Comparing the means of two distinct populations without pairing up individual measurements.
- 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:

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

  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, and the sample 2 points are iid with mean μ_2 and variance σ_2^2.
- 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:

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

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

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

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

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

Example: packing weights

- A company research effort involved finding a workable geometry for molded pieces of a solid.
- One comparison made was between the weight (in grams) of molded pieces of a particular geometry that could be poured into a standard container, and the weight of irregularly shaped pieces (obtained through crushing), that could be poured into the same container.
- $n_1 = 24$ crushed pieces and $n_2 = 24$ molded pieces were made and weighed.
- μ_1 is the true mean packing weight of the crushed pieces, and μ_2 is the true mean packing weight of the molded pieces.
- I want to formally test the claim that the crushed weights are greater than the molded weights.

Example: packing weights

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

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

- n_1 and n_2 are each < 25, but each sample is normally distributed enough to flex that rule and allow $n_1 = n_2 = 24$.
- Assume the crushed weights are iid (μ_1, σ_1^2).
- Assume the molded weights are iid (μ_2, σ_2^2).
- $K \sim N(0, 1)$ under the null hypothesis.

Example: packing weights

4. The moment of truth:

$$K = \frac{(\overline{x}_1 - \overline{x}_2) - 0}{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \frac{179.55 - 132.97 - 0}{\sqrt{\frac{(8.34)^2}{24} + \frac{(9.31)^2}{24}}} = 18.3$$

$$\begin{aligned} \text{p-value} &= P(Z > K) = 1 - \Phi(K) = 1 - \Phi(18.3) \\ &= 4 \times 10^{-75} \end{aligned}$$

5. With a p-value of $4 \times 10^{-75} < \alpha$, we reject H_0 in favor of H_a.
6. There is overwhelming evidence that more crushed solid material by weight can be poured into the container than molded solid material.

Example: packing weights

- The analogous lower 95% confidence interval for $\mu_1 - \mu_2$ is:

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

$$= \left((179.55 - 132.97) - z_{0.95} \sqrt{\frac{(8.34)^2}{24} + \frac{(9.31)^2}{24}},\ \infty \right)$$

$$= (46.58 - 1.64 \cdot 2.55,\ \infty)$$

$$= (42.40,\ \infty)$$

- We're 95% confident that the true mean packing weight of crushed solids is at least 42.40 g greater than that of the molded solids.

Example: packing weights

Molded		Crushed
7.9	11	
4.5, 3.6, 1.2	12	
9.8, 8.9, 7.9, 7.1, 6.1, 5.7, 5.1	12	
2.3, 1.3, 0.0	13	
8.0, 7.0, 6.5, 6.3, 6.2	13	
2.2, 0.1	14	
	14	
2.1, 1.2, 0.2	15	
	15	
	16	1.8
	16	5.8, 9.6
	17	1.3, 2.0, 2.4, 3.3, 3.4, 3.7
	17	6.6, 9.8
	18	0.2, 0.9, 3.3, 3.8, 4.9
	18	5.5, 6.5, 7.1, 7.3, 9.1, 9.8
	19	0.0, 1.0
	19	

Your turn: anchor bolts

- An experiment carried out to study various characteristics of anchor bolts resulted in 78 observations on shear strength (kip) of 3/8-in. diameter bolts and 88 observations on strength of 1/2-in. diameter bolts.

Variable	N	Mean	Median	TrMean	StDev	SEMean
diam 3/8	78	4.250	4.230	4.238	1.300	0.147

Variable	Min	Max	Q1	Q3
diam 3/8	1.634	7.327	3.389	5.075

Variable	N	Mean	Median	TrMean	StDev	SEMean
diam 1/2	88	7.140	7.113	7.150	1.680	0.179

Variable	Min	Max	Q1	Q3
diam 1/2	2.450	11.343	5.965	8.447

- Let Sample 1 be the 1/2 in diameter bolts and Sample 2 be the 3/8 in diameter bolts.
- Using a significance level of $\alpha = 0.01$, find out if the 1/2 in bolts are more than 2 kip stronger (in shear strength) than the 3/8 in bolts.
- Calculate and interpret the appropriate 99% confidence interval to support the analysis.

Answers: anchor bolts

- $n_1 = 88, \ n_2 = 78.$
- $\overline{x}_1 = 7.14, \ \overline{x}_2 = 4.25$
- $s_1 = 1.68, \ s_2 = 1.3$

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

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

   - Assume:
     - H_0 is true.
     - Sample 1 points are drawn from iid (μ_1, σ_1^2) distributions.
     - Sample 2 points are drawn from iid (μ_2, σ_2^2) distributions.
   - Then, $K \sim N(0, 1)$

Answers: anchor bolts

4. The moment of truth:

$$K = \frac{(\overline{x}_1 - \overline{x}_2) - 2)}{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \frac{(7.14 - 4.25) - 2}{\sqrt{\frac{(1.68)^2}{88} + \frac{(1.3)^2}{78}}} = 3.84$$

$$\begin{aligned} \text{p-value} &= P(Z > K) = 1 - P(Z \leq K) = 1 - P(Z \leq 3.84) \\ &= 1 - \Phi(3.84) \approx 0 \end{aligned}$$

5. With a p-value $\approx 0 < \alpha = 0.01$, we reject H_0 in favor of H_a.
6. There is overwhelming evidence that the $1/2$ in anchor bolts are more than 2 kip stronger in shear strength than the $3/8$ in bolts.

Answers: anchor bolts

- I use a lower confidence interval for $\mu_1 - \mu_2$:

$$\left((\overline{x_1} - \overline{x}_2) - z_{1-\alpha}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}},\ \infty \right)$$
$$= \left((7.14 - 4.25) - z_{0.99} \cdot \sqrt{\frac{1.68^2}{88} + \frac{1.3^2}{78}},\ \infty \right)$$
$$= (2.89 - 2.33 \cdot 0.232,\ \infty)$$
$$= (2.35,\ \infty)$$

- We're 99% confident that the true mean shear strength of the 1/2 in anchor bolts is at least 2.35 kip more than the true mean shear strength of the 3/8 in anchor bolts.