Mathematics for Computer Science and Engineering

Confidence Intervals: Paired data

Mamatha H. R.
Professor
Department of Computer Science and Engineering
PES University, Bangalore

Topics to be covered

- **5.7 - Confidence Intervals with Paired Data**

Textbook1: Chapter 5, section 5.7

Confidence intervals for paired data

- We have seen how to compare two independent samples
- Now **we will see how to compare two samples that are paired ▯**
- In other words **the two samples are not independent, Y1 and Y2 are linked in some way, usually by a direct relationship ▯**
- *For example*,
    - measure the weight of subjects before and after a six month diet.
    - A tire manufacturer wishes to compare the tread wear of tires made of a new material with that of tires made of a conventional material.

Confidence intervals for paired data

- A **paired t-test is used to compare two population means** where you have two samples in which observations in one sample can be paired with observations in the other sample.

- Examples of where this might occur are:
  - **Before-and-after observations on the same subjects**
  - (e.g. students' diagnostic test results before and after a particular module or course).

- A comparison of two different methods of measurement or two different treatments where the measurements/treatments are applied to the same subjects.

Example

Let x = test score before the module, y = test score after the module .To find that the true mean difference is zero, the procedure is as follows:

1. Calculate the difference ($d_i = y_i - x_i$) between the two observations on each pair, making sure you distinguish between positive and negative differences.

2. Calculate the mean difference, $\bar{d}$.

3. Calculate the standard deviation of the differences, s_d, and use this to calculate the standard error of the mean difference, $SE(\bar{d}) = \frac{s_d}{\sqrt{n}}$

4. Calculate the t-statistic, which is given by $T = \frac{\bar{d}}{SE(\bar{d})}$. Under the null hypothesis, this statistic follows a t-distribution with $n-1$ degrees of freedom.

5. Use tables of the t-distribution to compare your value for T to the t_{n-1} distribution. This will give the p-value for the paired t-test.

To Remember

For this test to be valid the differences only need to be approximately normally distributed. Therefore, it would not be advisable to use a paired t-test where there were any extreme outliers.

Example

A tire manufacturer wishes to **compare the tread wear of tires made of a new material with that of tires made of a conventional material. One tire of each type is placed on each front wheel of each of 10 front-wheel-drive automobiles. The choice as to which type of tire goes on the right wheel and which goes on the left is made with the flip of a coin.** Each car is driven for 40,000 miles, then the tires are removed, and the depth of the tread on each is measured.
We wish to find a 95% confidence interval for the mean difference in tread wear between old and new materials in a way that takes advantage of the reduced variability produced by the paired design.

	Car									
	1	2	3	4	5	6	7	8	9	10
New material	4.35	5.00	4.21	5.03	5.71	4.61	4.70	6.03	3.80	4.70
Old material	4.19	4.62	4.04	4.72	5.52	4.26	4.27	6.24	3.46	4.50
Difference	0.16	0.38	0.17	0.31	0.19	0.35	0.43	−0.21	0.34	0.20

Example

To put this into statistical notation, let $(X1, Y1), \ldots, (X10, Y10)$ *be the pairs.*
(Xi -> tread on the tire made from new material on the i^{th} car) and
(Yi -> tread on the tire made from old material on the i^{th} car.)

Let $D_i = X_i - Y_i$ represent the difference between treads for tires on the i^{th} car. Let μ_X and μ_Y represent the population means for X and Y, respectively.
We wish to find a 95% confidence interval for the difference $\mu_X - \mu_Y$.
Let μ_D represent the population mean of the differences.
Then $\mu_D = \mu_X - \mu_Y$.

It follows that a confidence interval for μ_D will also be a confidence interval for $\mu_X - \mu_Y$.
Since the sample $D1, \ldots, D10$ is a random sample from a population with mean μ_D, we can use one-sample methods to find confidence intervals for μ_D.
In this example, since the sample size is small, we use the Student's t method.

Textbook: Page 371

Example

$$\overline{D} = 0.232 \qquad s_D = 0.183$$

The sample size is 10, so there are nine degrees of freedom. The appropriate t value is $t_{9,.025} = 2.262$.

The confidence interval

therefore $0.232 \pm (2.262)(0.183)/\sqrt{10}$, or $(0.101,\ 0.363)$.

Textbook: Page 371

Confidence interval for the true mean difference

Summary

Let $D_1, \ldots, D_n$ be a *small* random sample ($n \leq 30$) of differences of pairs. If the population of differences is approximately normal, then a level $100(1-\alpha)\%$ confidence interval for the mean difference μ_D is given by

$$\overline{D} \pm t_{n-1,\alpha/2}\frac{s_D}{\sqrt{n}} \tag{5.24}$$

where s_D is the sample standard deviation of $D_1, \ldots, D_n$. Note that this interval is the same as that given by expression (5.9).

If the sample size is large, a level $100(1-\alpha)\%$ confidence interval for the mean difference μ_D is given by

$$\overline{D} \pm z_{\alpha/2}\sigma_{\overline{D}} \tag{5.25}$$

In practice $\sigma_{\overline{D}}$ is approximated with $s_D/\sqrt{n}$. Note that this interval is the same as that given by expression (5.1).

Textbook: Page 372

References

- “Statistics for Engineers and Scientists”, William Navidi, McGraw Hill Education, India, 4th Edition, 2015.

Dr. Mamatha. H.R
Professor
Department of Computer Science and Engineering
PES University, Bangalore