Lecture 8

Paired Samples

In this lecture, we discuss the comparison of paired data. In the previous 2-sample tests, one of the assumptions about the two data sets we were comparing was that they were independent - sample 1 had nothing to do with sample 2. Sometimes, however, there is a natural pairing between samples. For instance, a drug and a control might both be tested on the same individual for a number of individuals. In this case, the two samples are not independent. The individual serves as a blocking factor, reducing variation and alleviating confounding factors. Of course, the design still has to be good. Whether the treatment or the control/placebo came first or second should be randomized in this example. The method of analysis is to do one-sample statistics on the differences rather than 2-sample statistics using each data set. To use parametric t-tests, the requirement is that the differences be normally distributed or the sample size be moderately large. The distribution of the separate, individual data sets is irrelevant.

This lecture will apply numerous test to the same data set. This is NOT what you should do in your projects or when given data. You should decide which test is most appropriate for your data and apply that ONE test. At any rate, if you do try multiple tests, you should only tell your reader about the most appropriate one. Of the tests below, the t-test is the most appropriate for these data because the sample size is relatively large and the data (differences) are not too skewed and the t-test is the most powerful of the three tests discussed.

For this lecture we will use the following data. The mercury levels in 25 juvenile black marlins were obtained by two methods: an old “permanganate” method and a new “selective reduction” method that allows for simultaneous estimation of inorganic and methyl mercury levels. The following data were originally from Kacprzak and Chvojka (1976) and used in Rice’s text: Mathematical Statistics and Data Analysis.

Fish	Selective Reduction(ppm)	Permanganate (ppm)	Difference (old-new)	Signed Rank
1	0.32	0.39	0.07	15.5
2	0.40	0.47	0.07	15.5
3	0.11	0.11	0	
4	0.47	0.43	-0.04	-11
5	0.32	0.42	0.10	19
6	0.35	0.30	-0.05	-13.5
7	0.32	0.43	0.11	20
8	0.63	0.98	0.35	23
9	0.5	0.86	0.36	24
10	0.60	0.79	0.19	22
11	0.38	0.33	-0.05	-13.5
12	0.46	0.45	-0.01	-2.5
13	0.20	0.22	0.02	6.5
14	0.31	0.30	-0.01	-2.5
15	0.62	0.60	-0.02	-6.5
16	0.52	0.53	0.01	2.5
17	0.77	0.85	0.08	17.5
18	0.23	0.21	-0.02	-6.5
19	0.30	0.33	0.03	9
20	0.70	0.57	-0.13	-21
21	0.41	0.43	0.02	6.5
22	0.53	0.49	-0.04	-11
23	0.19	0.20	0.01	2.5
24	0.31	0.35	0.04	11
25	0.48	0.40	-0.08	-17.5

t-test for paired data

A t-test for these data would involve a one-sample t-test on the differences. Is the mean difference zero or not? (The alternative could also be one sided, but here there is no information about whether the new method should give higher or lower mercury estimates than the old method.) Let D denote differences with data $d_1, d_2, \ldots, d_n$. Then the test statistic is

$$t = \frac{\bar{d} - 0}{\text{S.E.}(\bar{d})} \text{ where } \text{S.E.}(\bar{d}) = \frac{\text{st.dev.}(d)}{\sqrt{n}}$$

Confidence intervals for the difference are given by $\left(\bar{d} - t_{\frac{\alpha}{2},\text{d.f.}} \times \text{S.E.}(\bar{d}), \bar{d} + t_{\frac{\alpha}{2},\text{d.f.}} \times \text{S.E.}(\bar{d})\right)$

In older versions of Minitab, you create a column containing the differences, say using the menu Calc > Calculator... Then you perform one-sample t-procedures on it using the menu Stat > Basic Statistics > 1-Sample t... In newer versions, you can use Stat > Basic Statistics > Paired t...

Example 1 For the data above, a test of whether or not there is a difference in mercury measurements between the two methods is insignificant ($T = 1.74$ with 24 degrees of freedom and a p-value of 0.094.) ■

Sign test for paired data

If the median difference is zero, you would expect half the differences to be positive and half to be negative. If you had all positive differences or all negative differences for a sample of size greater than 5 or 6 or so, you might well be suspicious that there was a difference. The sign test is based on this very simple observation and it has no other assumptions whatsoever. Its disadvantage is that it completely ignores the magnitude of the differences and is not very powerful.

The mathematics behind this test is very simple. If each sign has a 50-50 chance of being +, then the probability of having k +'s in a sample of size n is

$$f(k|n) = \binom{n}{k}\left(\frac{1}{2}\right)^n$$

To get the p-value, you simply sum this quantity over the appropriate tails. For instance, if the test was one-sided and k was large, you would find $f(k|n) + f(k+1|n) + \cdots + f(n|n)$.

In Minitab, sign procedures can be found under the menu Stat > Nonparametrics > 1-Sample Sign...

Wilcoxon signed rank test for paired data

The Wilcoxon signed rank test is more powerful than the sign test because it considers the magnitude of the differences as well as the signs. However, it replaces the values of the differences with the ranks of their absolute values multiplied by their signs so it is robust against outliers and skewness in the differences. The null hypothesis is simple that the data is symmetrical about zero.

Letting W_+ denote the sum of the positive ranks, we can determine that, under the null hypothesis,

$$\mathrm{E}W_+ = \frac{n(n+1)}{4} \text{ and } \mathrm{Var}(W_+) = \frac{n(n+1)(2n+1)}{24}$$

The normalized sum is approximately normally distributed:

$$Z \approx \frac{W_+ - \mathrm{E}(W_+)}{\mathrm{Var}(W_+)}$$

In Minitab, sign procedures can be found under the menu Stat > Nonparametrics > 1-Sample Wilcoxon...

The graph!

You should always, always, always try to look at your data. In this case, a good graphic to look at is a scatterplot. All of the tests above say that there is no statistical difference between the two methods. What does the scatterplot suggest?

In Minitab, scatterplots can be created using the menu Graph > Scatterplot...

References and Readings

[1] J. Kacprzak and R. Chvojka. Determination of methyl mercury in fish by flameless atomic absorption spectroscopy and comparison with an acid digestion method for total mercury. *Journal of the Association of Official Analytical Chemists*, 59:153–157, 1976.

[2] John Rice. *Mathematical Statistics and Data Analysis*. Brooks/Cole, 3^{rd} edition, 2006.

Exercises for Lecture 8

1. –

2. –