Chapter 9: Two Paired Samples Tests

Ott & Longnecker Sections: 6.4 (Sign Test not covered in O&L)

Key Concepts: Paired samples vs. independent samples, testing the differences.

In the previous section, our data consisted of samples drawn independently from two populations, and we saw several procedures for testing the difference between measures of central location based on those samples. Here we will concern ourselves with comparing measures of central location of two populations when the two samples are in fact paired (and hence, dependent).

1 The Paired T-Test (Paired Observations, Normal Differences)

1.1 Example

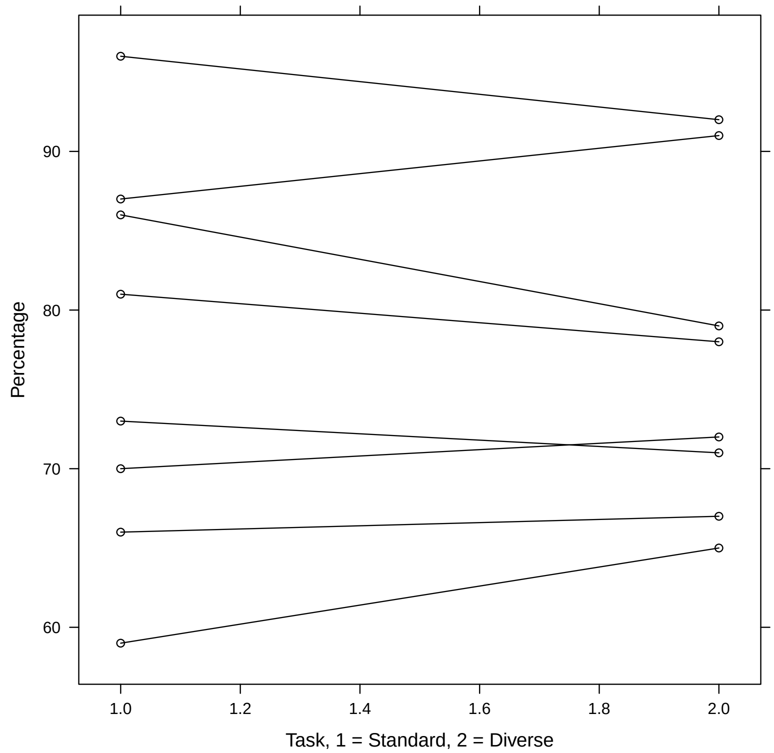
Musculoskeletal disorders of the neck and shoulders are common in office workers because of repetitive tasks. Long periods of upper-arm elevation above 30 degrees have been shown to be related to disorders. It was thought that varying working conditions over the course of the day could alleviate some of these problems. Eight office workers were randomly selected. They were observed for one work day under the standard conditions, and the percentage of time that their dominant upper-arm was below 30 degrees was recorded. The next day, these same individuals had their work diversified, and again were observed. The results are below. *Don't give differences yet*:

Participant	1	2	3	4	5	6	7	8
Diverse	78	91	79	65	67	72	71	96
Standard	81	87	86	59	66	70	73	92

We'd like to know whether the diversification changes the percentage of time where arm angle is below 30 degrees. So, we want to test:

$H_0 : \mu_{diverse} - \mu_{standard} = 0$
vs.
$H_A : \mu_{diverse} - \mu_{standard} \neq 0$

As always, we should start with some graphing. This data is clearly paired, since we are taking two measurements on each individual. We need to use a graphing method that captures the pairing. One way is a graph with the dots connected for each individual:

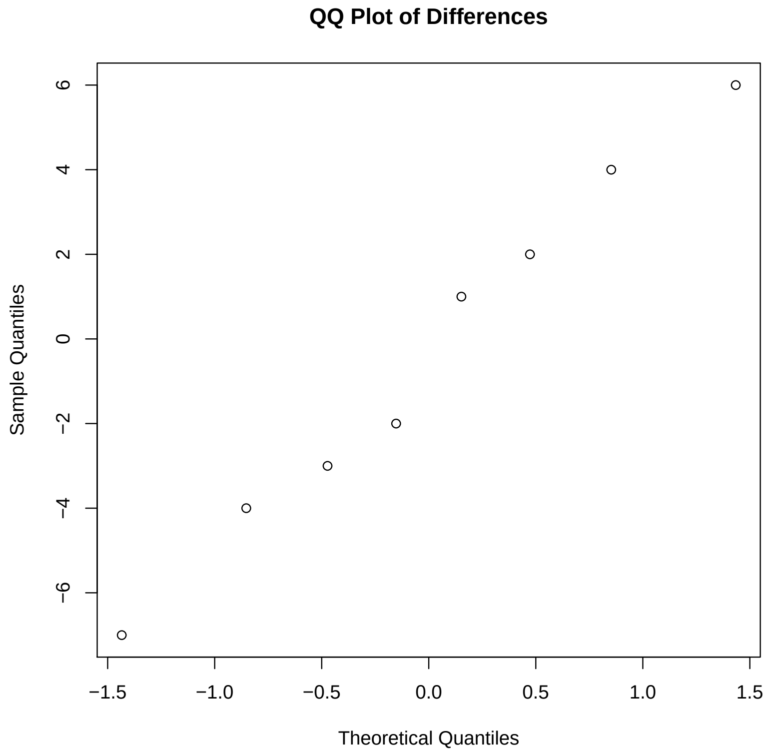


It seems that some go up, some go down, and some stay the same. It's unlikely that we will see much difference. We might be tempted to use a T-test here, but remember that the T-tests we've seen so far assume that all the data is independent between and within groups, and we know that the groups are dependent because of the pairing. There are a variety of ways to handle this dependence, but a simple way is to simply remove it by taking the differences for each pair. The data is then the change from standard to diverse, rather than the raw percentage. This fixes the problem because the workers are independent, and now we have only one observation (the change) for each one. Now, we can run one sample tests on the *differences*.

So, to do the paired T-test, the first step is to take the differences. I took them diverse - standard, but either way is fine, as long as you stay consistent. The raw differences are:

Differences: -3, 4, -7, 6, 1, 2, -2, 4.

If we want to use a one-sample T-test, we need to check normality *of the differences*:



This looks quite straight, so we can safely assume normality. Now that we are considering differences, we could change our hypotheses slightly to reflect this. If we let μ_d be the true mean difference, then we want to test:

$H_0 : \mu_d = 0$
vs.
$H_A : \mu_d \neq 0$

(Usually we are testing $\mu_d = 0$ because this is more natural, but theoretically we could test for any specified mean difference.) If we let d_i denote the observed differences, we might define $\bar{d}$ to be the sample mean of the differences and s_d to be the sample standard deviation of the differences. Then our t-statistic is:

$$t_{obs} = \frac{\bar{d}-0}{\frac{s_d}{\sqrt{n}}},$$

where n is the number of *pairs*. This statistic is distributed as a t_{n-1}, the same as it was in the one-sample case. For this data we get $\bar{d} = 0.63$ and $s_d = 4.34$, so our statistic is:

$$t_{obs} = \frac{0.63-0}{\frac{4.34}{\sqrt{8}}} = 0.41.$$

We compare to a T_7 distribution and find from the table (using the closest value we have, $t_{obs} = 0.4$) that the p-value is about 0.702 (don't forget that the alternative is two-sided).

Using R, the p-value is close to 0.696. Thus we would not reject. It seems that diversification of work has little effect on arm angle.

The paired T-test requires that the differences between the values for each pair are normally distributed. Note that the individual data points need not necessarily be normally distributed - only the differences.

As far as when to use an independent samples test vs. when to pair: if the samples are obviously independent, use the independent samples test. If the data are obviously paired, use the paired test. The data should inform the method of analysis: remember, the goal is to make the correct inference, not get a small p-value! However, it is sometimes the case that the design of the experiment is ambiguous. Some things to keep in mind are (1) if the data are positively correlated (as you'd expect with truly paired data), then the standard error of the differences will be smaller than if you treated them independently; (2) on the other hand, the independent samples test has more degrees of freedom.

1.2 Recap of paired T-test

The data consists of paired observations. Let:

$$X_{1,i} = \text{a data point from population 1}$$
$$\mu_1 = \text{true mean of population 1}$$
$$X_{2,i} = \text{a data point from population 2}$$
$$\mu_2 = \text{true mean of population 2}$$
$$D_i = X_{1,i} - X_{2,i} = \text{the difference for pair } i$$
$$n = \text{number of pairs}$$
$$\sigma_D^2 = \text{true variance of the differences}$$

We wish to test: $H_0 : \mu_1 - \mu_2 = \delta$ vs. $H_A : \mu_1 - \mu_2 \neq \delta$

Good graphs for exploring the data include dotplots where the observations for each pair are connected by lines, and QQ plots of the differences to check normality.

If based on our prior knowledge and after exploring the data we are willing to assume:

- All of the *pairs* are indepdendent
- The differences follow a normal distribution

Then the test statistic is:

$$t = \frac{\bar{D} - \delta}{\frac{S_D}{\sqrt{n}}}$$

Where $\bar{D}$ is the mean of the differences and S_D is the sample standard deviation of the differences.
In the end, compare to a T distribution on $n - 1$ degrees of freedom.

2 The Sign Test

Another way to relax the normality assumption is with a version of the sign test. Remember, the sign test is a test about the median. In the paired case, this implies that the test will be about whether the median of the differences is equal to a specified value (usually zero). However, as before, if the distribution of the differences is symmetric, it is also a test about the mean difference. To do the sign test in a paired situation, count the number of pairs where the difference is positive. Then perform the sign test as usual.

2.1 Sign test in two paired samples

- **Assumption.** $D_1, \ldots, D_n$ are an i.i.d. sample.
- **Hypothesis.** Let m be the population *median*. We wish to test $H_0 : m = 0$.
- **Test statistic.** $B = \#$ of data values greater than 0 (Ignore values tied with 0.)

  Note that if H_0 is true, $B \sim \text{Binomial}(n^*, 0.5)$, where n^* is the number of data points not equal to 0(remember, these values are removed from the data set).
- **P-value**. Let b be the observed number of data points greater than 0. If:
  - $H_A : m > 0$: calculate $P(B \geq b) = P(B = b) + P(B = b + 1) + \ldots + P(B = n^*)$.
  - $H_A : m < 0$: calculate $P(B \leq b) = P(B = b) + P(B = b - 1) + \ldots + P(B = 1) + P(B = 0)$.
  - $H_A : m \neq 0$: calculate $P(B \geq b)$ and $P(B \leq b)$. Then, take the smaller of these two values, and double that.

Let's again use our office data. Though the sign test does not require normality, it still works when the differences are normal. In our example there were 5 pairs where the difference was positive, the p-value works out to 0.73. Again, we would not reject.