Chapter 6
Paired Continuous Data (Paired T-Test, Wilcoxon Signed Rank Test)

1 General Purpose

Studies where two outcomes in one patient are compared with one another, are often called crossover studies, and the observations are called paired observations.

As paired observations are usually more similar than unpaired observations, special tests are required in order to adjust for a positive correlation between the paired observations.

2 Schematic Overview of Type of Data File

Outcome 1	Outcome 2
.	.
.	.
.	.
.	.
.	.
.	.
.	.
.	.
.	.

T.J. Cleophas, A.H. Zwinderman, *Clinical Data Analysis on a Pocket Calculator*,
DOI 10.1007/978-3-319-27104-0_6

3 Primary Scientific Question

Is the first outcome significantly different from second one.

4 Data Example

The underneath study assesses whether some sleeping pill is more efficacious than a placebo. The hours of sleep is the outcome value.

Patient number	Outcome 1	Outcome 2	Individual Differences
1	6.0	5.1	0.9
2	7.1	8.0	−0.9
3	8.1	3.8	4.3
4	7.5	4.7	3.1
5	6.4	5.2	1.2
6	7.9	5.4	2.5
7	6.8	4.3	2.5
8	6.6	6.0	0.6
9	7.3	3.7	3.8
10	5.6	6.2	−0.6

Outcome = hours of sleep after treatment

5 Analysis: Paired T-Test

Rows may be more convenient than columns, if you use a pocket calculator, because you read the data like you read the lines of a textbook. Two rows of observations in 10 persons are given underneath:

Observations 1:
6.0, 7.1, 8.1, 7.5, 6.4, 7.9, 6.8, 6.6, 7.3, 5.6

Observations 2:
5.1, 8.0, 3.8, 4.4, 5.2, 5.4, 4.3, 6.0, 3.7, 6.2

Individual differences:
0.9, −0.9, 4.3, 3.1, 1.2, 2.5, 2.5, 0.6, 3.8, −0.6

A. not significant
B. $0.05 < p < 0.10$
C. $P < 0.05$
D. $P < 0.01$

Is there a significant difference between the observations 1 and 2, and which level of significance is correct (P = p-value, SD = standard deviation) ?

Mean difference	=	1.59

SDs are calculated as demonstrated in the Chap. 1.

SD of mean difference	=	$\sqrt{(SD_1^2 + SD_2^2)}$
	=	1.789 (SD = standard deviation)
SE = SD/√10	=	0.567 (SE = standard error)
t = 1.59/0.567	=	2.81

We have here (10–1) = 9 degrees of freedom, because we have 10 patients and 1 group of patients. According to the underneath t-table the p-value equals < 0.05, and we can conclude that a significant difference between the two observations is in the data: the values of row 1 are significantly higher than those of row 2. The answer C is correct.

6 T-Table

The t-table has a left-end column giving degrees of freedom (≈ sample sizes), and two top rows with p-values (areas under the curve = p-values), one-tail meaning that only one end of the curve, two-tail meaning that both ends are assessed simultaneously. The t-table is, furthermore, full of t-values, that, with ∞ degrees of freedom, are equal to z-values (Chap. 36). The t-values are to be understood as mean results of studies, but not expressed in mmol/l, kilograms, but in so-called SEM-units (Standard error of the mean units), that are obtained by dividing your mean result by its own standard error. With many degrees of freedom (large samples) the curve will be a little bit narrower, and more in agreement with nature.

df	One-Tail = .4	.25	.1	.05	.025	.01	.005	.0025	.001	.0005
	Two-Tail = .8	.5	.2	.1	.05	.02	.01	.005	.002	.001
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657	127.32	318.31	636.62
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925	14.089	22.327	31.598
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841	7.453	10.214	12.924
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.265	0.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.260	0.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.259	0.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.258	0.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.258	0.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.257	0.689	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.257	0.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.257	0.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.257	0.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	0.257	0.686	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	0.256	0.686	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	0.256	0.685	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	0.256	0.685	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	0.256	0.684	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	0.256	0.684	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	0.256	0.684	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	0.256	0.683	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	0.256	0.683	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	0.256	0.683	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	0.255	0.681	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
60	0.254	0.679	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
120	0.254	0.677	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
∞	0.253	0.674	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291

A t-value of 2.81 with 9 degrees of freedom indicates, that we will need the 9th row of the t-values. The upper row of the table gives the area under the curve of the Gaussian-like t-distribution. The t-value 2.81 is left from 3.250, and right from 2.262. Now look right up at the upper row: we are right from 0.05. The p-value is < 0.05.

7 Alternative Analysis: Wilcoxon Signed Rank Test

The t-tests as reviewed in the previous section are suitable for studies with Gaussian-like data distributions. However, if there are outliers, then the t-tests will not be adequately sensitive, and non-parametric tests will have to be applied. We should add that non-parametric tests are also adequate for testing normally distributed data. And, so, these tests are, actually, universal, and are, therefore, absolutely recommended.

Calculate the p-value using the Wilcoxon signed rank test.

Observations 1:
6.0, 7.1, 8.1, 7.5, 6.4, 7.9, 6.8, 6.6, 7.3, 5.6

Observations 2:
5.1, 8.0, 3.8, 4.4, 5.2, 5.4, 4.3, 6.0, 3.7, 6.2

Individual differences:
0.9, −0.9, 4.3, 3.1, 1.2, 2.5, 2.5, 0.6, 3.6, −0.6

Rank numbers:
3.5, 3.5, 10, 7, 5, 8, 6, 2, 9, 1

A. not significant
B. $0.05 < p < 0.10$
C. $p < 0.05$
D. $p < 0.01$

Is there a significant difference between observations 1 and 2? Which significance level is correct?

The individual differences are given a rank number dependent on their magnitude of difference. If two differences are identical, and if they have for example the rank numbers 3 and 4, then an average rank number is given to both of them, which means 3.5 and 3.5. Next, all positive and all negative rank numbers have to be added up separately. We will find 4.5 and 50.5. According to the Wilcoxon table underneath, with 10 numbers of pairs, the smaller one of the two add-up numbers must be smaller than 8 in order to be able to speak of a p-value < 0.05. This is true in our example.

8 Wilcoxon Test Table

Wilcoxon Test Table

Number of pairs	$P < 0.05$	$P < 0.01$
7	2	0
8	2	0
9	6	2
10	8	3
11	11	5
12	14	7
13	17	10
14	21	13
15	25	16
16	30	19

The first column gives the numbers of pairs in your paired data file. Rank numbers of positive and negative differences are separately added up. The second and third columns give the smaller one of the two add-up sums required for statistical significance.

As demonstrated in the above table, also according to the non-parametric Wilcoxon's test the outcome one is significantly larger than the outcome two. The p-value of difference here equals $p < 0.05$. We should add that non-parametric tests take into account more than the t-test, namely, that Non-gaussian-like data are accounted for. If you account more, then you will prove less. That's why the p-value may be somewhat larger.

9 Conclusion

The significant effects indicate that the null-hypothesis of no difference between the two outcome can be rejected. The treatment 1 performs better than the treatment 2. It may be prudent to use the non-parametric tests, if normality is doubtful like in the current small data example given. Paired t-tests and Wilcoxon signed rank tests need, just like multivariate data, more than a single outcome variable. However, they cannot assess the effect of predictors on the outcomes, because they do not allow for predictor variables. They can only test the significance of difference between the outcomes.

10 Note

The theories of null-hypotheses and frequency distributions and additional examples of paired t-tests and Wilcoxon signed rank tests are reviewed in Statistics applied to clinical studies 5th edition, Chaps. 1 and 2, Springer Heidelberg Germany, 2012, from the same authors.