Chapter 27
Robust Tests for Imperfect Data

1 General Purpose

Robust tests are wonderful for imperfect data, because they often produce significant results, when standard tests don't. They may be able to handle major outliers in data files without largely changing the overall test results.

2 Schematic Overview of Type of Data File

Outcome
.
.
.
.
.
.
.
.
.

T.J. Cleophas, A.H. Zwinderman, *Clinical Data Analysis on a Pocket Calculator*,
DOI 10.1007/978-3-319-27104-0_27

3 Primary Scientific Question

With data imperfect due to major outliers, can robust tests provide significant effects, if traditional tests don't.

4 Data Example

Frailty score-improvements after physiotherapy of 33 patients are measured in a study. The data are in the second column of underneath data.

Patient	Score-improvement	Deviation from median	Trimmed data	Winsorized data
1	−8.00	11		−1.00
2	−8.00	11		−1.00
3	−8.00	7		−1.00
4	−4.00	7		−1.00
5	−4.00	7		−1.00
6	−4.00	7		−1.00
7	−4.00	7		−1.00
8	−1.00	4	−1.00	−1.00
9	0.00	3	0.00	0.00
10	0.00	3	0.00	0.00
11	0.00	3	0.00	0.00
12	1.00	2	1.00	1.00
13	1.00	2	1.00	1.00
14	2.00	1	2.00	2.00
15	2.00	1	2.00	2.00
16	2.00	1	2.00	2.00
17	3.00	median	3.00	3.00
18	3.00	0	3.00	3.00
19	3.00	0	3.00	3.00
20	3.00	0	3.00	3.00
21	4.00	1	4.00	4.00
22	4.00	1	4.00	4.00
23	4.00	1	4.00	4.00
24	4.00	1	4.00	4.00
25	5.00	2	5.00	5.00
26	5.00	2	5.00	5.00
27	5.00	2		5.00
28	5.00	2		5.00
29	6.00	3		5.00
30	6.00	3		5.00
31	6.00	3		5.00
32	7.00	4		5.00
33	8.00	5		5.00

The data suggest the presence of some central tendency: the values 3.00 and 5.00 are observed more frequently than the rest. However, the one sample t-test shows a mean difference from zero of 1.45 scores with a p-value of 0.067. Thus, not statistically significant.

5 T-Test for Medians and Median Absolute Deviations (MADs)

Underneath are descriptives of the above data that are appropriate for robust testing are given.

Mean	1.455
standard deviation	4.409
standard error	0.768
mean after replacing outcome 1st patient with 0.00	1.697
mean after replacing outcome first 3 patients with 0.00	2.182
median	3.000
MAD	2.500
mean of the Winsorized data	1.364
standard deviation of the Winsorized data	3.880

MAD = median absolute deviation = the median value of the sorted deviations from the median of a data file.

If the mean does not accurately reflect the central tendency of the data e.g. in case of outliers (highly unusual values), then the median (value in the middle) or the mode (value most frequently observed) may be a better alternative to summarizing the data and making predictions from them.

$$\text{Median} = 3.00$$

The above example shows in the third column the deviations from the median, and the table gives the median of the deviations from median (MAD = median absolute deviation).

$$\text{MAD} = 2.50$$

If we assume, that the data, though imperfect, are from a normal distribution, then the standard deviation of this normal distribution can be approximated from the equation

$$\text{standard deviation}_{\text{median}} = 1.426 \times \text{MAD} = 3.565$$

$$\text{standard error}_{\text{median}} = 3.565/\sqrt{n} = 3.565/\sqrt{33} = 0.6206$$

A t-test is, subsequently, performed, and produces a very significant effect: physiotherapy is really helpful.

$$t = \text{median}/\text{standard error}_{\text{median}} = 3.00/0.6206 = 4.834.$$

The underneath t-table is used for determining the p-value. For 32 (=33 – 1) degrees of freedom, and a t-value >3.646 means a p-value of <0.001 (two-tail).

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

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.

6 T-Test for Winsorized Variances

The terminology comes from Winsor’s principle: all observed distributions are Gaussian in the middle. First, we have to trim the data, e.g., by 20 % on either side (see data file, fourth column). Then, we have to fill up their trimmed values with Winsorized scores, which are the smallest and largest untrimmed scores (data file, fifth column). The mean is, then, calculated, as well as the standard deviation and standard error, and a t-test is performed for null hypothesis testing.

Winsorized mean = 1.364
Winsorized standard deviation = 3.880
Winsorized standard error = $3.880 / \sqrt{n}$
= $3.880 / \sqrt{33}$
= 0.675

t-test
t = Winsorized mean / Winsorized standard error
t = 2.021

With 32 (=33−1) degrees of freedom and a t-value > 2.2042 means a p-value < 0.05 (two-sided).

$$\text{p-value} < 0.05$$

7 Mood’s Test (One Sample Wilcoxon’s Test)

	−8.00	−8.00	−8.00	−4.00	−4.00	−4.00	−1.00	0.00	0.00	0.00	1.00	….
−8.00		−8.00	−8.00	−6.00	−6.00	−6.00	−4.50	−4.00	−4.00	−4.00	−3.50	
−8.00			−8.00	−6.00	−6.00	−6.00	−4.50	−4.00	−4.00	−4.00	−3.50	
−8.00				−6.00	−6.00	−6.00	−4.50	−4.00	−4.00	−4.00	−3.50	
−4.00					−4.00	−4.00	−2.50	−2.00	−2.00	−2.00	−1.50	
−4.00						−4.00	−2.50	−2.00	−2.00	−2.00	−1.50	
−4.00							−2.50	−2.00	−2.00	−2.00	−1.50	
−1.00								−0.50	−0.50	−0.50	0.00	
0.00									0.00	0.00	0.50	
0.00										0.00	0.50	
0.00											0.50	
1.00												
….												
….												

The Mood's test is, sometimes, called the one sample Wilcoxon's test. The above table shows how it works. Paired averages [(vertical value + horizontal value)/2] are calculated. If the data are equally distributed around an average of 0, then we will have half of the average being positive, half negative.

We observe 1122 paired averages,
1122/2 = 561 should be positive,
349 positive paired averages are found.
A chi-square test is performed
chi-square value = (Observed − expected numbers)^2/Expected numbers
chi-square value = $(349–561)^2$ / 349 = 128.729
$p < 0.001$ with 1 degree of freedom

The underneath chi-square table has an upper row with areas under the curve, a left-end column with degrees of freedom, and a whole lot of chi-square values. It shows that, for 1 degrees of freedom, and chi-square values >10.827, we will find a p-value <0.001.

Chi-squared distribution

	Two-tailed *P*-value			
df	0.10	0.05	0.01	0.001
1	2.706	3.841	6.635	10.827
2	4.605	5.991	9.210	13.815
3	6.251	7.851	11.345	16.266
4	7.779	9.488	13.277	18.466
5	9.236	11.070	15.086	20.515
6	10.645	12.592	16.812	22.457
7	12.017	14.067	18.475	24.321
8	13.362	15.507	20.090	26.124
9	14.684	16.919	21.666	27.877
10	15.987	18.307	23.209	29.588
11	17.275	19.675	24.725	31.264
12	18.549	21.026	26.217	32.909
13	19.812	22.362	27.688	34.527
14	21.064	23.685	29.141	36.124
15	22.307	24.996	30.578	37.698
16	23.542	26.296	32.000	39.252
17	24.769	27.587	33.409	40.791
18	25.989	28.869	34.805	42.312
19	27.204	30.144	36.191	43.819
20	28.412	31.410	37.566	45.314
21	29.615	32.671	38.932	46.796
22	30.813	33.924	40.289	48.268
23	32.007	35.172	41.638	49.728

(continued)

	Two-tailed *P*-value			
df	0.10	0.05	0.01	0.001
24	33.196	36.415	42.980	51.179
25	34.382	37.652	44.314	52.619
26	35.536	38.885	45.642	54.051
27	36.741	40.113	46.963	55.475
28	37.916	41.337	48.278	56.892
29	39.087	42.557	49.588	58.301
30	40.256	43.773	50.892	59.702
40	51.805	55.758	63.691	73.403
50	63.167	67.505	76.154	86.660
60	74.397	79.082	88.379	99.608
70	85.527	90.531	100.43	112.32
80	96.578	101.88	112.33	124.84
90	107.57	113.15	124.12	137.21
100	118.50	124.34	135.81	149.45

8 Conclusion

The above three robust tests produced p-values of <0.001, <0.05, and <0.001, while the one sample t-test was not statistically significant. Robust tests are wonderful for imperfect data, because they often produce significant results, when standard tests don’t.

9 Note

More background, theoretical and mathematical information of robust tests is given in SPSS for Starters part 2, Chap. 20, Springer Heidelberg Germany, 2012, from the same authors.