Chapter 9
Kendall-Tau Regression for Ordinal Data

1 General Purpose

Linear regressions (Chaps. 8 and 10) are adequate for outcomes with continuous data, otherwise called scale data. Continuous data have a stepping pattern, where the steps have equal intervals. If, in a regression model, the outcome data have a stepping pattern, but the intervals are not equal, then the term ordinal data is more appropriate for such data, and regression testing of ranks is more appropriate. The data need to be tested according to the magnitude of their rank numbers. This chapter is to assess how rank testing of regression models performs as compared to traditional linear regression.

2 Schematic Overview of Type of Data File

Rank number exposure	Rank number outcome
.	.
.	.
.	.
.	.
.	.
.	.
.	.

T.J. Cleophas, A.H. Zwinderman, *Clinical Data Analysis on a Pocket Calculator*,
DOI 10.1007/978-3-319-27104-0_9

3 Primary Scientific Question

Is rank testing of linear by linear data adequately sensitive for testing linear data where the order of data may be more important than the magnitude itself. The latter type of data is usually called ordinal data.

4 Data Example

In a short stay hospital the numbers hospitalization days were used to predict the numbers of medical complications. It was assumed, that, the longer the stay, the more risk of multiple complications.

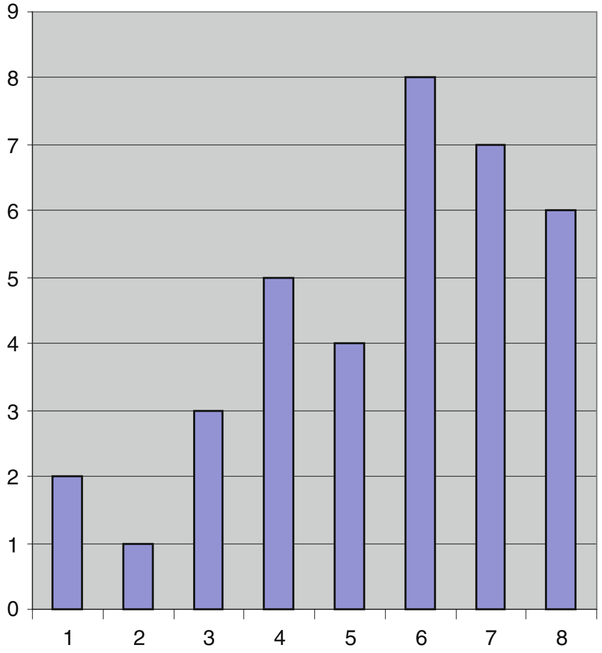
Patient no.	1	2	3	4	5	6	7	8
Days in hospital	8	9	10	2	3	4	5	16
Numbers if complications	5	20	16	2	1	3	6	12

A traditional linear regression of the above data will produce a correlation coefficient of 0.695 with a p-value of 0.056, which is not statistically significant.

From the above 8 patients the underneath rank numbers different in magnitude can be obtained:

Rank numbers of days in hospital	5	6	7	1	2	3	4	8
Rank numbers of complications	4	8	7	2	1	3	5	6

Rank numbers of days in hospital in ascending order	1	2	3	4	5	6	7	8
Corresponding rank numbers of complications	2	1	3	5	4	8	7	6



The above graph (with days in hospital on the x-axis and numbers of complications on the y-axis), shows, that, although the numbers of complications tend to increase with the numbers of hospital days their relationship is far from linear. However, the data are not continuous, but ranks, and rank testing is the appropriate analysis.

5 Rank Correlation Testing

We wish to know, whether the rank correlation between days in hospital and numbers of complication is statistically significant.

1	2	3	4	5	6	7	8
2	1	3	5	4	8	7	6

In the above second row right from 2 we have 6 values larger than 2,
right from 1 we have also 6 values larger than 1,
right from 3 we have also 5 values larger than 3,
....
....

If we add-up all these values, we will end up with a value of 23.

In the above second row right from 2 we have 1 value smaller than 2,
right from 1 we have also 0 values smaller than 1,
right from 3 we have also 0 values smaller than 3,
....
....

If we add-up all these values, we will end up with a value of 5.

The rank correlation coefficient Tau $= (23–5) / [½\ n\ (n - 1)]$
$= 0.64.$

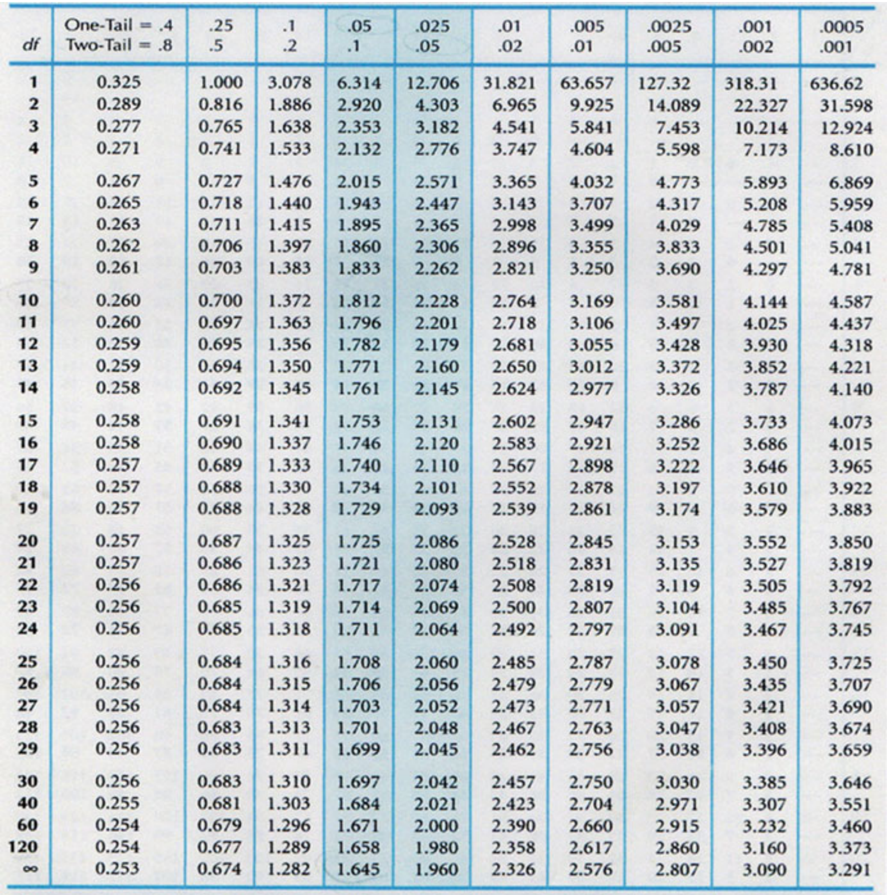
As Tau runs from 0 to 1 (or rather −1 to 1), a value of 0, 64 indicates a pretty strong correlation coefficient. However, in order to test whether this correlation coefficient is significantly different from 0, we need to test it against its standard error, using a z-test (see also Chap. 37).

$z = (|\text{Tau}| - 1) / \sqrt{}\ [n\ (n - 1)\ (2n + 5) / 18]$
$z = 17 / 8.08$
$z = 2.10$

Z-values are equal to t-values with ∞ degrees of freedom, and can be found in the bottom row of the t-table.

6 T-Table

Our above z-value, 2.10. is >1.960. According to the underneath t-table (bottom row) a z-value >1.960, corresponding with a two-tail p-value of <0.05 (look right up at the 2nd upper area under the curve row), indicates, that a significant correlation between the x- and y-variable. The days in hospital is closer to numbers of complications than could happen by chance at p-value slightly <0.05, and the association between the two variables is, thus, statistically significant.



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.

7 Conclusion

The Kendall-Tau regression assesses just like the traditional linear regression the level of linear relationships between two variables. However, Kendall-Tau provides a slightly less sensitive result. The traditional linear model (as described in the Chaps. 8 and 10) of the rank data produces an r-value of 0.857 and a p-value of 0.007. Why so? Unlike traditional linear regression, Kendall-Tau takes into account that the intervals between the rank values may not be identical. Also, just like in the non-parametric tests for comparing treatment groups and treatment modalities (Chaps. 6, 7, and 34), if an analysis method takes into account more, it will usually produce less spectacular _results. More in general, if you account more, you will prove less.

8 Note

More background, theoretical and mathematical information of rank testing are given in the Chaps 5, 6, and 7 of this work, and in the Chaps. 1, 2, 4, 9, 13, SPSS for starters and 2nd levelers 2nd edition, Springer Heidelberg Germany, 2015, from the same authors.