Chapter 9
Typology of Medical Data (51 Patients)

General Purpose

Apart from histograms (see Chap. 1, Statistics applied to clinical studies 5th edition, "Hypotheses, data, stratification". pp 1–14, Springer Heidelberg Germany, 2012), and Q-Q plots (Chap. 42 of current work), the typology of data and frequency procedures (to be reviewed in the Chaps. 10, and 11 of the current work) are a good way to start looking at your data. First, we will address the typology of the data.

Nominal Data

Nominal data are discrete data without a stepping pattern, like genders, age classes, family names. They can be assessed with pie charts, frequency tables and bar charts.

Ordinal Data

Ordinal data are also discrete data, however, with a stepping pattern, like severity scores, intelligence levels, physical strength scores. They are usually assessed with frequency tables and bar charts.

Scale Data

Scale data also have a stepping pattern, but, unlike ordinal data, they have steps with equal intervals. With small steps they are called continuous data. They are sometimes called quantitative data, while nominal and ordinal data are traditionally called qualitative data. The scale data are assessed with summary tables and histograms.

The typology of the data values become particularly important when it comes to statistical analyses. E.g., means and standard deviations makes no sense with nominal data. The problem with ordinal data is that the steps are usually not equal, like with scale data. With ordinal data you will usually have a mix-up of larger and smaller steps. This biases the outcome if you use a scale data test for their analysis.

The Chap. 37 of the current book, entitled "Ordinal scaling for clinical scores with inconsistent intervals", shows how this problem can mathematically be largely solved by complementary log-log transformations.

Primary Scientific Question

In econometrics and marketing research (Foroni, Econometric models for mixed-frequency data, edited by European University of Economics, Florence, 2012), frequency procedures are routinely used for the assessment of nominal, ordinal and scale data. Can they also be adequately applied for assessing medical data?

Example

The patients of an internist's outpatient clinic are reviewed.

nominal variable	ordinal variable	scale variable
agegroup	severity	time
2,00	2,00	2,50
2,00	1,00	6,00
2,00	1,00	2,50
1,00	3,00	2,00
1,00	1,00	5,00
2,00	1,00	4,00
2,00	3,00	,50
1,00	1,00	2,50
2,00	3,00	4,00
2,00	2,00	1,50

agegroup: 1 = senior, 2 = adult, 3 = adolescent, 4 = child
severity: complaint severity scores 1–4
time: consulting time (minutes)

The first 10 patients are in the above table. The entire file (51 patients) is entitled "frequencies", and is available at extras.springer.com. We will start by opening the file in SPSS statistical software.

Example   55

Nominal Variable

Command:

click Analyze....Descriptive Statistics....Frequencies....Variable(s): enter agegroups....
mark Display frequency tables....click Charts....click Pie charts....click OK.

The underneath pie chart shows that seniors and adults predominate and that children are just a small portion of the outpatient clinic population.
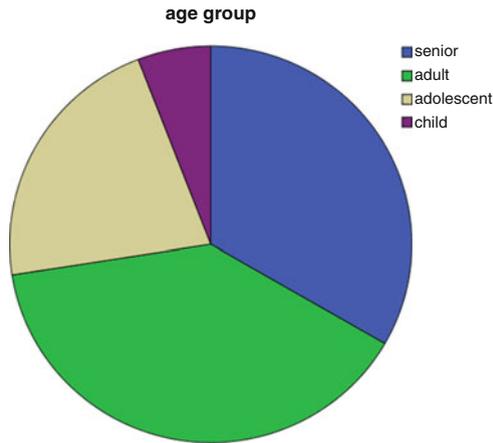
The frequency table shows precise frequencies of the nominal categories.

Age group		Frequency	Percent	Valid percent	Cumulative percent
Valid	Senior	17	33,3	33,3	33,3
	Adult	20	39,2	39,2	72,5
	Adolescent	11	21,6	21,6	94,1
	Child	3	5,9	5,9	100,0
	Total	51	100,0	100,0	

If you wish, you could present your data in the form of descending or ascending frequencies.

Command:

click Analyze....Descriptive Statistics....Frequencies....Variable(s): enter agegroups
....mark Display frequency tables....click Charts....click Bar charts....click Continue
....click Format....click Descending counts....click Continue.......click OK.

The underneath graph is in the output sheet. It shows an ordered bar chart with adults as largest category.
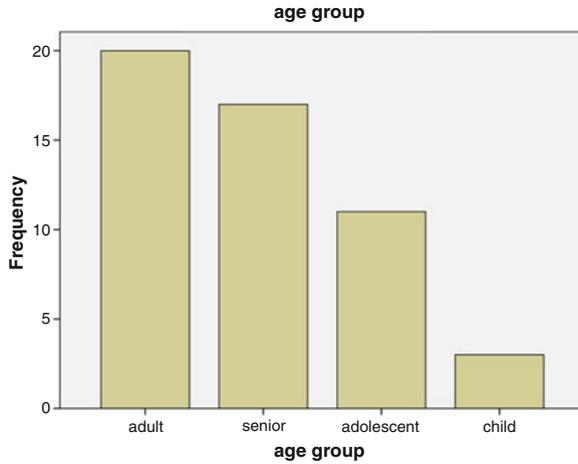
Ordinal Variable

Command:

click Analyze....Descriptive Statistics....Frequencies....Variable(s): enter severity .... mark Display frequency tables....click Charts....click Bar charts....click Continue .... click Format....click Ascending counts....click Continue.......click OK.

According to the severity score count the underneath graph shows the percentages of patients. Most of them are in the score one category, least of them in the score five category.

Example   57

complaint severity score

complaint severity score
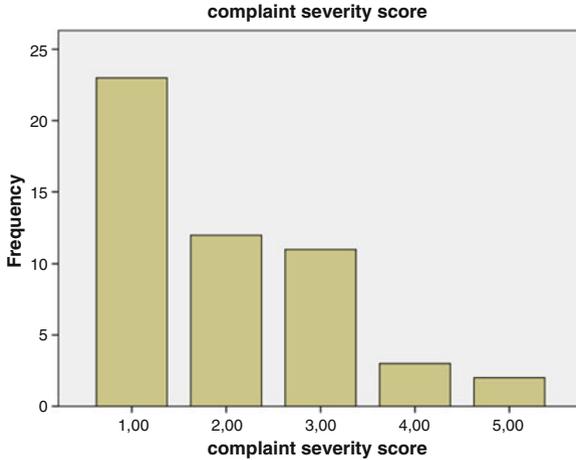
The table gives the precise numbers of patients in each category as well as the percentages. If we have missing values, the valid percent column will give the adjusted percentages, while the cumulative percentage gives the categories one and two, one and two and three etc. percentages.

Complaint severity score

|		Frequency	Percent	Valid percent	Cumulative percent
Valid	1,00	23	45,1	45,1	45,1
	2,00	12	23,5	23,5	68,6
	3,00	11	21,6	21,6	90,2
	4,00	3	5,9	5,9	96,1
	5,00	2	3,9	3,9	100,0
	Total	51	100,0	100,0	

Scale Variable

Command:

click Analyze....Descriptive Statistics....Frequencies....Variable(s): enter time .... remove mark from "Display frequency tables"....click Statistics....mark Quartiles.... Std.deviations....Minimum....Maximum....   Mean....Median   ....   Skewness.... Kurtosis....click Continue....then click Charts....Histograms…mark Show normal curve on histogram....click Continue....click OK.

The statistics table tells us that the consulting time is 3,42 min on average, and 50 % of the consults are between 2 and 4 min. The most extreme consults took 0,5 and 15,0 min.

Statistics		
consulting time (min)		
N	Valid	51
	Missing	0
Mean		3,4216
Median		2,5000
Std. Deviation		2,99395
Skewness		2,326
Std. Error of Skewness		,333
Kurtosis		5,854
Std. Error of Kurtosis		,656
Minimum		,50
Maximum		15,00
Percentiles	25	2,0000
	50	2,5000
	75	4,0000

The histogram shows the frequency distribution of the data and suggests skewness to the right. Most of the consults took as little as less than 5 min but some took no less than 5–15 min. This means that the data are not very symmetric and means, and that standard deviation are not very accurate to summarize these data.

Histogram

Mean = 3,42
Std. Dev. = 2,994
N = 51
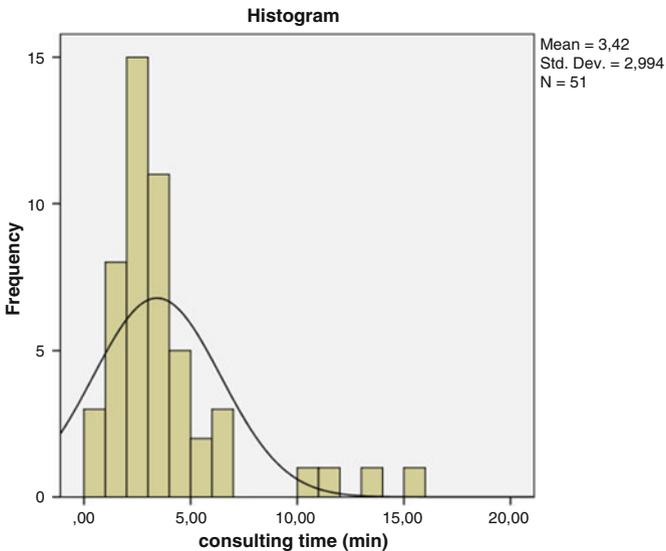
Indeed, a significant level of skewness to the right is in the data, because $2{,}326/0{,}333 = 6{,}985$ is much larger than 1,96 (see the above table). We will try and

use a logarithmic transformation of these skewed data. because this often "normalizes" the skewness.

Command:

click Transform....Compute Variable....type logtime in Target Variable....type ln(time) in Numeric Expression....click OK.

In the main screen it can be observed that SPSS now has produced a novel variable entitled "logtime". We will perform the scale variable analysis again, and replace the variable "time" with "logtime".

Command:

click Analyze....Descriptive Statistics....Frequencies....Variable(s): enter logtime.... click Charts....Histograms…mark Show normal curve on histogram....click Continue....click OK.
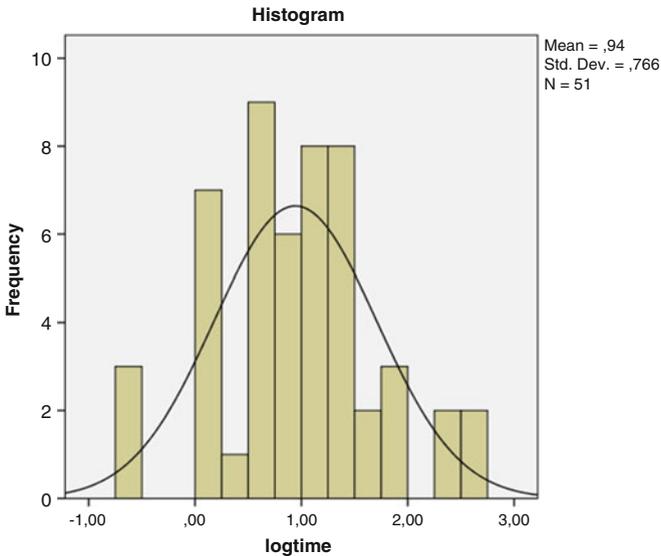
In the output sheets the underneath graph is shown. The data distribution looks less skewed and much closer to a normal distribution now. The logtime data can now be used for data analysis using normal statistical tests.

Conclusion

Data can be classified as nominal, ordinal and scale. For each type frequencies and frequency distributions can readily be calculated, and they enable an unbiased view of their patterns. Nominal data have no mean value. Ordinal data are tricky, because,

although they have a stepping pattern, they offer a mix-up of larger and smaller steps. Ordinal regression can largely adjust this irregularity. Skewed scale data often benefit from log-data transformations.

Note

More background, theoretical and mathematical information of ordinal data is given in Chap. 37 of the current book, entitled "Ordinal scaling for clinical scores with inconsistent intervals".