Chapter 16
Online Analytical Procedure Cubes, a More Rapid Approach to Analyzing Frequencies (450 Patients)

General Purpose

OLAP means online analytical procedures. Cubes is a term used to indicate multidimensional datasets. OLAP cubes were first used in 1970, by SQL Express a software package for storing business data, like financial data, in an electronic warehouse, and, at the same time, turning raw data into meaningful information (business intelligence), and was initially called layered reports. Generally, financial data or production data are being summarized, and from these summaries subsummaries are computed like productions by time-periods, cities, and other subgroups. Instead of quantities of business data, quantities of health outcomes could, similarly, be analyzed. However, to date no such analyses have been performed. This chapter is to assess whether online analytical procedures can also be applied on health outcomes instead of business outcomes.

Primary Scientific Question

Can online analytical processing (OLAP) using summaries and subsummaries of clinical outcome data support traditional crosstab analyses?

Example

We will use as example the data also used in the Chaps. 10 and 11. A parallel group study of 450 patients assessed the effect of different complementary treatment modalities on qol score levels. The first 11 patients of the data file is underneath. The entire data file is entitled "qol.sav", and is in extras.springer.com.

treatment	counseling	qol	satdoctor
3	1	4	4
4	0	2	1
2	1	5	4
3	0	4	4
2	1	2	1
2	0	1	4
4	0	4	1
3	0	4	1
4	1	4	4
2	1	3	4
4	1	5	5

treatment = treatment modality (1 = cardiac fitness, 2 = physiotherapy, 3 = wellness, 4 = hydrotherapy)
counseling = counseling given (0 = no, 1 = yes)
qol = quality of life scores (1 = very low, 5 = very high)
satdoctor = satisfaction with treating doctor (1 = very low, 5 = very high)

We will start by opening the data file in SPSS.

Command:

Analyze....Reports....OLAP Cubes....Summary Variable(s): enter qol score.... Grouping Variable(s): enter treatment, counseling....click OK.

Case processing summary						
	Cases					
	Included		Excluded		Total	
	N	Percent	N	Percent	N	Percent
Qol score * treatment * counseling	450	100,0 %	0	,0 %	450	100,0 %

* Symbol of multiplication

OLAP cubes						
Treatment:Total						
Counseling:Total						
	Sum	N	Mean	Std. Deviation	% of Total Sum	% of Total N
Qol score	1,436	450	3,19	1,457	100,0 %	100,0 %

The above tables are in the output sheets. The add-up sum of all scores are given (1436), and the overall mean score of the 450 patients (3.19). Next we can slice these results into subgroups, and calculate mean scores by treatment modality. A pivoting tray is used for that purpose.

Example   97

Command:

Double-click the above table....the term Pivot is added to the menu bar....click Pivot in the menu bar....the underneath Pivoting Tray consisting of all of the variables appears....close the Pivoting Tray.
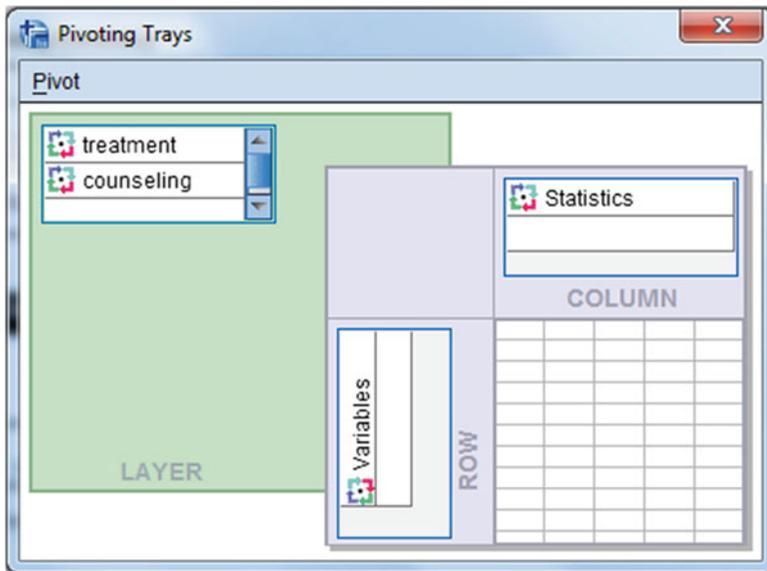
The above table now has in the upper right corner two drop boxes: treatment, and counseling. Treatment: click Total (in blue), and produce the underneath four tables with summary statistics of the four treatment modalities.

OLAP cubes

Treatment:cardiac fitness

Counseling:Total

	Sum	N	Mean	Std. Deviation	% of Total Sum	% of Total N
Qol score	387	118	3,28	1,501	26,9 %	26,2 %

OLAP cubes

Treatment:physiotherapy

Counseling:Total

	Sum	N	Mean	Std. Deviation	% of Total Sum	% of Total N
Qol score	296	100	2,96	1,449	20,6 %	22,2 %

OLAP cubes						
Treatment:wellness						
Counseling:Total						
	Sum	N	Mean	Std. Deviation	% of Total Sum	% of Total N
Qol score	332	104	3,19	1,501	23,1 %	23,1 %

OLAP cubes						
Treatment:hydrotherapy						
Counseling:Total						
	Sum	N	Mean	Std. Deviation	% of Total Sum	% of Total N
Qol score	421	128	3,29	1,381	29,3 %	28,4 %

It is easy to that the mean qol score of physiotherapy is significantly lower than that of hydrotherapy according to an unpaired t-test:

mean	Std Deviation	n	Std Error
2.96	1.449	100	0.145
3.29	1.381	128	0.122

$t = (3.29 - 2.96)/\sqrt{(0.145^2 + 0.122^2)} = 1.96$
with $(100 + 128 - 2) = 226$ degrees of freedom

This would indicate that these two mean qol scores are significantly different from one another at $p < 0.05$.

In addition to summary statistics of different treatments, we can also compute summary statistics of qol scores by counseling yes or no.

Command:

click the treatment drop box....select Total....next click the counseling drop box… first select counseling no....then select counseling yes.

The underneath tables are given.

OLAP cubes						
Treatment:Total						
Counseling:Total						
	Sum	N	Mean	Std. Deviation	% of Total Sum	% of Total N
Qol score	1,436	450	3,19	1,457	100,0 %	100,0 %

OLAP cubes						
Treatment:Total						
Counseling:No						
	Sum	N	Mean	Std. Deviation	% of Total Sum	% of Total N
Qol score	652	230	2,83	1,530	45,4 %	51,1 %

OLAP cubes						
Treatment:Total						
Counseling:Yes						
	Sum	N	Mean	Std. Deviation	% of Total Sum	% of Total N
Qol score	784	220	3,56	1,279	54,6 %	48,9 %

It is again easy to test whether the mean qol score of no-counseling is significantly lower than that of yes-counseling according to an unpaired t-test:

mean	Std Deviation	n	Std Error
2.83	1.530	230	0.101
3.56	1.279	220	0.086

$t=(3.56-2.83)/\sqrt{(0.101^2+0.086^2)}=5.49$
with $(230+220-2)=448$ degrees of freedom

This would indicate that the two means are significantly different from one another at $p<0.0001$.

Conclusion

In the current example the individual qol levels were estimated as 5 scores on a 5-points linear scale. In the Chaps. 2, 6 and 7 analyses took place by comparing frequencies of different qol scores with one another. In the OLAP cubes analysis a different approach is applied. Instead of working with frequencies of different qol scores, it works with mean scores and standard deviations. Other summary measure is also possible like sums of qol scores, medians, ranges or variances etc. Unpaired t-test can be used to test the significance of difference between various subsummaries.

Although we have to admit that the crosstab analyses and OLAP cube lead to essentially the same results, the OLAP cube procedure is faster, and few simple table are enough to tell you what is going on. Also additional statistical testing with the t-test is simple.

Note

More background, theoretical and mathematical information about the analyses of interaction matrices of frequencies are in the Chaps. 9–11. The OLAP cube is another approach with similar results, but it works more rapidly than the other methods.