CHAPTER 3

Examples from practical research

This chapter discusses examples where the original data was freely available. These examples will be used throughout this manual in order to illustrate how different effect size indices are calculated and interpreted.

3.1 Example A : Hypnotherapeutic ego strengthening (HES) for male coronary artery bypass patients (de Klerk et.al., 2004)

From the population of white, male, Afrikaans speaking patients who received their first coronary artery bypass surgery at a given hospital, 50 are randomly assigned to equally sized control and experimental groups. No statistically significant differences between the two groups is found when considering the following variables: the biographical variables, age, education level, martial status, cardiac history, family health, social background and lifestyle stress. Before the operation, HES is applied to the patients in the experimental group (EG) and after the operation it is disengaged, while the control group (CG) receives no such treatment. The BECK-depression inventory (BDI) and "Profile of Mood States" (POMS) questionnaire are filled in before the therapy, immediately after the therapy and again during a follow-up session six weeks after the therapy. The BDI's values which vary between 0 and 13 indicate minimal depression, values between 14 and 19 indicate moderate depression, while values between 29 and 63 indicate severe depression. Only the angst and depression scales of POMS are used, denoted by POMS_A and POMS_D respectively. Table A.1 shows the means and standard deviations (SD) of the EG and CG for all three scales for before the therapy, after the therapy and during the follow-up. The STATISTICA or EXCEL data set 'data_A' is available on the manual's web page.

Table A.1

Means and SD's of Example A

Group	Questionnaire	Before	After	Follow-up
	BDI	13,04(6,14)*	8,72(6,82)	6,76(4,94)
EG	POMS_D	18,00(12,26)	8,08(9,84)	6,32(7,84)
(n=25)	POMS_A	22,24(6,13)	15,20(4,40)	11,60(4,11)
	BDI	11,56(5,85)	15,56(6,95)	16,36(9,55)
KG	POMS_D	13,68(9,87)	14,72(12,73)	13,96(12,95)
(n=25)	POM_A	19,72(6,15)	20,28(6,39)	17,16(5,74)

* values in parentheses are the SD's

The purpose was to determine whether or not the application of HES caused a lowering in depression and angst in the experimental group, but not in the control group.

A two-way analysis of variance (ANOVA) model with the test opportunity as repeated (dependent) measurements is given in Table A.2 for BDI in the experimental group.

Table A.2

2-way ANOVA with repeated measurements over tests for BDI

Source of variation	df	SS	Mean SS	F	P
Between tests	2	516,19	258,09	10,47	0,00017
Within tests:	72	2595,56	36,05		
Patients	24	1425,41	59,39	2,41	0,00470
Patient × tests	48	1183,15	24,65		

3.2 Example B: The personality preferences of lecturers and students (Rothmann et.al., 2000b)

The personality preferences of lecturers and students at a university's faculty for Economic and Management sciences is determined by the application of the "Myers-Briggs Type Indicator" (MBTI) questionnaire. All of the lecturers, second year and third year students who were willing to take part in the study completed a questionnaire.
Altogether 282 individuals took part in the study, of which 28 were lecturers, 121 were male students and 133 were female students.

The following aspects of personality preference are measured by the MBTI:

- The manner in which people choose to interact with others: **E**xtraversion (E) – **I**ntroversion (I).
- The manner in which people choose to focus on data and its collection: **S**ensing (S) – I**N**tuition (N).
- The manner in which people choose to process data and make decisions:: **T**hinking (T) – **F**eeling (F).
- The manner in which people choose to organise themselves: **J**udging (J) – **P**erceiving (P).

These 4 aspects could also be measured on a continuous scale centred on the value 100, so that the larger the observation, the greater that person's preference for the second option (for example, the second option for Extraversion – Introversion is I) etc. Table B.1 provides the means and SD's of the students and lecturers for each of the abovementioned aspects:

Table B.1

Means and SD of continuous preference score

Aspect	Student (N=254)	Lecturer (N=28)
E/I	94,58(25,15)	107,64(25,06)
S/N	86,65(20,58)	84,57(27,60)
T/F	86,79(21,66)	82,64(22,47)
J/P	91,08(28,60)	70,07(25,93)

The purpose was to compare the students' and lecturers' preferences in order to make suggestions for the development of both groups if there is no correspondence between them.

Based on the groupings formed by the MBTI on each of the groups, there are a total of 16 (i.e., 2^4) preference types that can be obtained (e.g., ESTJ, ISTP, etc.). From these types there are four temperament types, SJ, SP, NT and NF, which can be obtain if we only look at the S/N and J/P aspects. Table B.2 provides the frequencies of male students, female students and lecturers which fall within each of these types:

Table B.2

Frequency of temperament types

Type	Male students	Female students	Lecturers
SJ	57(47,1%)	79(59,4%)	20(71,4%)
SP	29(24,0%)	23(17,3%)	0(0,0%)

NT	23(19,0%)	19(14,3%)	5(17,9%)
NF	12(9,9%)	12(9,0%)	3(10,7%)
Total	121(100%)	133(100%)	28(100%)

The purpose here is to determine both the study and lecturing style profiles of students and lecturers respectively, and also to determine the degree of correspondence between the two.

In this example we do not draw a probability sample from a population in such a way that the results from the sample could be generalized to the population. Even though the 254 students were obtained from a population of 758 second and third year students, the sample obtained is not necessarily representative of all of the students and so it should rather be seen as the *study population* that we want to study. The same is true for the 28 lecturers of the 37 available lecturers. In the practical application of effect size indices then, in Example B, only population indices will be calculated.

3.3 Example C: Smoking and the risk of coronary heart disease (Kline, 2004a: 155-156)

In this example by Glass & Hopkins, 120 employees with a coronary heart disease of a large company are paired (using biographical properties like gender, age, number of years in service, etc.) with 120 employees without the disease. We then have two groups that, apart from the heart disease, are essentially the same. Everyone in the study is classified in one of four smoking categories: non-smoking, less than one pack of cigarettes a day, one pack of cigarettes a day and more than one pack of cigarettes a day. The observed numbers are as follows:

Table C.1

Two-way frequency table of the level of smoking vs. occurrence of coronary heart disease

		Coronary heart disease		
		Yes	No	Total
Level of Smoking	Non	42	61	103
	<1	19	23	42
	1	39	25	64
	>1	20	11	31
	Total	120	120	240

The purpose of the study was to determine whether or not smoking increased the risk of coronary heart disease.

3.4 Example D: Self respect of three communities in North-east Australia (Smithson, 2000:246)

In the late 1980's a random sample of aboriginals and non-aboriginals living in medium sized urban areas in North-east Australia, was drawn. A sample of aboriginals living in rural areas was also drawn. Self-respect is measured using the Rosenberg Self-Esteem Scale (a scale ranging from 6 to 15).

Table D.1

Descriptive statistics of the self-respect data

Group		n	$\bar{x}$	s
Non-Aboriginal, urban	(1)	159	13,138	1,348
Aboriginal, urban	(2)	94	12,660	1,332
Aboriginal, rural	(3)	191	12,257	1,455
		444		

Table D.2

ANOVA-table of the self-respect data

Source of variation	df	SS	Mean SS	F	p
Group	2	67,47	33,74	17,41	<0,0001
Error	441	854,49	1,94		
Total	443	921,96			

The results indicate that there is a highly significant statistical difference between the mean self-respect values of the three groups (p<0,001). However, primary interest lies in trying to determine if the differences between the means is practically significant.

(The STATISTICA data set named ESTEEM can be downloaded from this manual's web page)

3.5 Example E: Cholesterol and Blood pressure of heart patients (Smithson, 2000: dataset HEART)

The cholesterol levels of male (Gender = 1) and female (Gender = 2) heart patients who do not exercise (oef = 0), exercise infrequently (oef = 1), exercise moderately (oef = 2) and exercise frequently (oef = 3) are obtained at the beginning of the study (chol_0), after one year (chol_1), after two years, three years and four years (chol_2-chol_4). The systolic (sist) and diastolic (diast) blood

pressure of the patients are also obtained. Table E.1 displays the descriptive statistics per gender and exercise group.

Table E.1:
Descriptive statistics for patients per gender and level of exercise.

GENDER	OEF	n	CHOL_0 Mean	CHOL_0 Std.Dev.	CHOL_1 Mean	CHOL_1 Std.Dev.	CHOL_2 Mean	CHOL_2 Std.Dev.	CHOL_3 Mean	CHOL_3 Std.Dev.	CHOL_4 Mean	CHOL_4 Std.Dev.	SIST Mean	SIST Std.Dev.	DIAST Mean	DIAST Std.Dev.
1	Total	12	180.67	30.10	190.17	37.23	201.17	42.81	215.17	55.20	221.25	61.31	110.50	10.34	73.83	11.00
	0	5	180.20	31.38	192.40	42.91	205.00	45.37	215.20	60.10	226.00	68.61	114.80	9.86	80.40	10.53
	1	4	185.00	44.05	192.00	50.68	205.00	61.63	221.50	77.93	222.50	83.48	104.50	10.88	67.50	3.79
	2	2	173.50	4.95	183.50	4.95	190.50	0.71	200.00	14.14	197.50	3.54	109.00	12.73	67.00	18.38
	3	1	180.00	0.00	185.00	0.00	188.00	0.00	220.00	0.00	240.00	0.00	116.00	0.00	80.00	0.00
2	Total	38	171.08	26.65	178.95	28.16	182.45	28.06	184.24	30.22	191.00	34.76	121.32	7.36	76.16	7.80
	0	14	178.00	17.51	184.57	19.53	189.50	17.25	190.71	21.06	195.57	27.12	120.14	7.25	74.71	8.36
	1	16	167.69	29.77	179.75	32.56	182.94	31.65	182.88	34.05	192.75	38.02	120.25	7.90	78.63	5.50
	2	7	170.86	32.77	172.00	31.24	173.43	34.36	180.14	37.15	183.86	42.80	124.86	5.52	76.00	8.49
	3	1	130.00	0.00	136.00	0.00	139.00	0.00	144.00	0.00	149.00	0.00	130.00	0.00	58.00	0.00
	All Groups	50	173.38	27.51	181.64	30.55	186.94	32.73	191.66	39.39	198.26	43.89	118.72	9.31	75.60	8.61

	OEF	n	CHOL_0 Mean	CHOL_0 Std.Dev.	CHOL_1 Mean	CHOL_1 Std.Dev.	CHOL_2 Mean	CHOL_2 Std.Dev.	CHOL_3 Mean	CHOL_3 Std.Dev.	CHOL_4 Mean	CHOL_4 Std.Dev.	SYST Mean	SYST Std.Dev.	DIAST Mean	DIAST Std.Dev.
Both Genders	0	19	178.58	21.01	186.63	26.41	193.58	26.86	197.16	35.29	203.58	42.03	118.74	8.09	76.21	9.04
	1	20	171.15	32.51	182.20	35.60	187.35	38.37	190.60	46.10	198.70	48.90	117.10	10.47	76.40	6.85
	2	9	171.44	28.46	174.56	27.58	177.22	30.70	184.56	33.72	186.89	37.57	121.33	9.59	74.00	10.58
	3	2	155.00	35.36	160.50	34.65	163.50	34.65	182.00	53.74	194.50	64.35	123.00	9.90	69.00	15.56

The cholesterol measurements per patient are time-dependent. This means that the means can be compared and the time effect can be ascertained. In addition, the genders and levels of exercise can also be compared.

3.6 Example F: Serum-cholesterol of men within activity groups

From a probability sample drawn from two South African industrial regions, the total serum-cholesterol, HDL and LDL cholesterol, triglyceride and physical activities of 1472 males, aged between 10 and 64, is measured. Based on an activity index which was determined beforehand, the men are divided into three activity groups: Low activity, Moderate activity and High activity. Table F.1 displays the descriptive statistics per activity group of the data.

To establish whether or not physical activity has an influence on cholesterol, an analysis of covariance with total serum-cholesterol as the response is conducted with the activity group grouping variable, while correcting for age. Table F.2 displays the analysis of covariance results, while Table F.3 displays the modified means of the cholesterol values per group after controlling for age. Table F.4 shows the analysis of variance results when one does not control for age.

Table F.1:Descriptive statistics for male activity groups

Act_grp	Mean age	N Age	Std.Dev. Age	Mean S_CHO	N S_CHO	Std.Dev. S_CHO	Mean S_TRI	N S_TRI	Std.Dev. S_TRI	Mean HDL_C	N HDL_C	Std.Dev. HDL_C	Mean LDL_C	N LDL_C	Std.Dev. LDL_C
1	35.9	728	14.5	552.5	728	166.9	227.4	702	180.1	99.4	704	23.9	373.1	704	117.8
2	24.3	237	14.4	509.7	237	134.2	154.3	227	86.2	108.5	234	27.4	337.3	234	108.4
3	23.0	468	12.5	482.0	468	123.1	152.5	444	98.8	108.2	457	25.2	315.2	457	88.7
All Groups	29.8	1433	15.2	522.4	1433	151.9	191.1	1373	149.5	103.8	1395	25.3	348.1	1395	110.6

Table F.2: Analysis of Covariance on S_CHOL corrected for age

Source of variation	SS	df	Mean SS	F	p
Age	4392610	1	4392610	231.04	0.000
Act_grp	117750	2	58875	3.10	0.046
Error	27168391	1429	19012		
Total	31678752	1432			

Table F.3: Modified means of age

Act_grp	N	Mod. Mean S_CHO	Std.error S_CHO
1	728	528.0	5.4
2	237	531.6	9.1
3	468	509.0	6.6

Table F.4: Analysis of variance on S_CHOL

Source of variation	SS	df	Gem SS	F	p
Act_grp	1461712	2	730856	33.11	0.000000
Error	31561001	1430	22071		
Total	33022714	1432			

3.7 Example G: Questionnaire concerning the advantages of imported motor vehicles

(Statsoft, Inc, 2004: Example dataset '10 item')

A questionnaire containing 10 items with statements such as "imported motor vehicles all look the same" or "imported motor vehicles do not have any personality", is given to 100 Americans to complete. For each statement a response must be given on a 9 point scale indicating whether they agree (9 indicates they agree completely) or disagree (1 indicates that they completely disagree) with the statement.

Table G.1 shows the ANOVA table obtained from STATISTICA's 'Reliability/Item analysis', while Table G.2 displays the inter-item correlation. Further, Table G.3 is a summary of the mean and variance of each item, while Table G.4 is the ANOVA table obtained when "poor" items (items 5 and 6) are removed.

Interest lies in determining how reliable the items in this questionnaire are.

The STATISTICA or EXCEL data set '10item' is available on the manual's web page.

Table G.1: ANOVA with items as random effect

Source of variation	Sum of squares	df	Mean Sum v squares	F	p
Between people	676.179	99	6.830		
Within people	1263.500	900	1.404		
Between Items	11.769	9	1.308	0.931	0.497
Error	1251.731	891	1.405		
Total	1939.679	999			

Mean Inter-item-correlation 0.2863

Table G.3: Means and variances

	Mean	Variance
ITEM1	4.50000	2.09091
ITEM2	4.74000	1.58828
ITEM3	4.70000	1.82828
ITEM4	4.48000	1.74707
ITEM5	4.59000	2.18374
ITEM6	4.55000	2.18939
ITEM7	4.65000	1.86616
ITEM8	4.78000	1.95111
ITEM9	4.67000	2.02131
ITEM10	4.45000	2.00758
SUM	46.11000	68.30091

Sum of variances= 19.47384

Alpha= 0.794313

Table G.2: Inter-item correlations

	ITEM1	ITEM2	ITEM3	ITEM4	ITEM5	ITEM6	ITEM7	ITEM8	ITEM9	ITEM10
ITEM1	1.000	0.576	0.491	0.428	0.040	0.139	0.539	0.375	0.396	0.500
ITEM2	0.576	1.000	0.464	0.361	-0.025	0.115	0.569	0.552	0.442	0.468
ITEM3	0.491	0.464	1.000	0.364	0.110	0.048	0.468	0.377	0.353	0.293
ITEM4	0.428	0.361	0.364	1.000	0.055	-0.043	0.267	0.353	0.418	0.369
ITEM5	0.040	-0.025	0.110	0.055	1.000	0.026	0.008	0.132	-0.036	0.007
ITEM6	0.139	0.115	0.048	-0.043	0.026	1.000	0.141	0.098	0.068	0.088
ITEM7	0.539	0.569	0.468	0.267	0.008	0.141	1.000	0.441	0.288	0.432
ITEM8	0.375	0.552	0.377	0.353	0.132	0.098	0.441	1.000	0.411	0.515
ITEM9	0.396	0.442	0.353	0.418	-0.036	0.068	0.288	0.411	1.000	0.405
ITEM10	0.500	0.468	0.293	0.369	0.007	0.088	0.432	0.515	0.405	1.000

Table G.4: ANOVA with 8 items as random effects

Source of variation	Sum of squares	df	Mean Sum v squares	F	p
Between people	741.864	99	7.494		
Within people	764.375	700	1.092		
Between Items	11.269	7	1.610	1.481	0.171
Error	753.106	693	1.087		
Total	1506.239	799			