CHAPTER 6

Comparison of more than two groups of observations

In Chapter 4 effect size indices are discussed for the differences between the means of two groups of measurements (both independent and dependent). In the case where an experimental design or study consists of more than two groups of measurements then there are two types of effects which can be of interest and for which effect size indices serve as measures. The first is the so called *omnibus effect* which attempts to determine if at least two of the groups' measurements' means differ and also the extent of this difference. The second is the *contrast effect* which compares the means of the measurements of specific groups or combinations of groups.

Example 6.1:

In Example A of Chapter 3 we found that there were 3 groups of measurements: before tests, after tests and follow-up tests. An omnibus effect allows us to establish if there are any differences between, for example, the mean BDI-measurements of the before, after and follow-up tests. A contrast could be something like, $\overline{x}_B - \overline{x}_A$, i.e., the difference between the means of the before test and the after test. Another example of a contrast is $\overline{x}_B - \frac{1}{2}\left(\overline{x}_A + \overline{x}_F\right)$, where $\overline{x}_F$ is the mean of the follow-up test. In this case the average of the means of the after and follow-up tests is compared to the before test's mean. □

Note:

Note that in the above discussion the term *group of measurements* is a more general term than, say, a group of people, and it can refer to a population or sample of these elements. Groups of measurements can be *independent,* which means that the measurements on the groups/populations/samples of elements are separate from one another. We make this distinction because we may find that the measurements can be dependent measurements on the same element in each population or sample. Example 6.1 illustrates this, because the before, after and follow-up measurements are each conducted on a single heart bypass patient and are thus dependent.

6.1 Indices for omnibus effects for independent measurements

An obvious extension of Cohen's δ to more than two groups, would be (Cohen, 1969, 1977, 1988):

$$\delta_{omn} = \frac{\mu_{max} - \mu_{min}}{\sigma} \quad , \tag{6.1}$$

where μ_{max} and μ_{min} are the largest and smallest means of the groups respectively and σ is the common SD of all of the groups.

With one-way analysis of variance (ANOVA) in mind, Cohen suggests the index f, defined as:

$$f = \frac{\sigma_\mu}{\sigma} \quad , \tag{6.2}$$

where

$$\sigma_\mu^2 = \frac{1}{k}\sum_{i=1}^{k}\left(\mu_\mathrm{i} - \mu\right)^2 , \tag{6.3}$$

is the variance of the $\mu_i's$,

with k the number of groups;

μ_i the mean of the $i-$th group;

μ the mean of all of the $\mu_i's$,

and assuming that all of the groups are of equal size.

Suppose that σ_t^2 is the *total variance* of all of the measurements over the groups, then for equal group sizes and equal variances we have that:

$$\sigma_t^2 = \sigma^2 + \sigma_\mu^2 \tag{6.4}$$

A sensible effect size index can then be the proportion of the total variance, which can be attributed to σ_t^2:

$$\eta^2 = \frac{\sigma_\mu^2}{\sigma^2 + \sigma_\mu^2} \quad . \tag{6.5}$$

This is Pearson's eta-squared and is related to f (from (6.2) and (6.4)) as follows:

$$\eta^2 = \frac{f^2}{1 + f^2} \tag{6.6}$$

The index η^2 , and its estimators, has found greater utility as an omnibus effect size index in practical situations than either δ_{omn} or f . Consequently, we will not concentrate any further these other indices.

6.1.1 Estimation of η^2:

Some notation regarding the one-way ANOVA will first be provided before we continue any further (for more details consult Steyn et al., 1998: 511-513).

SS_G : between groups sum of squares,

SS_E : within groups (error) sum of squares,

SS_{tot} : total sum of squares.

If the measurements are obtained from random sampling, then a biased estimator for η^2 is:

$$\tilde{\eta}^2 = \frac{SS_G}{SS_{tot}} \quad , \tag{6.7}$$

which underestimates η^2. This bias is expressed by Fowler (1985) as:

$$(1-\eta^2)\left[k-(1-\eta^2)(1+2\eta^2)\right]/n \, .$$

Hays's estimator $\hat{\omega}^2$ (see Fidler & Thompson, 2001: 585) is a modification of $\tilde{\eta}^2$ and is given by:

$$\hat{\omega}^2 = \frac{SS_G - (k-1) SS_E / (n-k)}{SS_{tot} + SS_E / (n-k)}, \tag{6.8}$$

where n is the total number of measurements in all of the k groups.

Notes:

1) While $\hat{\omega}^2$ is a modified estimator used to limit the influence of the bias, an unfortunate downside is that this estimator can now take on *negative* values. This can occur if the following variance ratio is less than 1, i.e.,

$$F = \frac{SS_G / (k-1)}{SS_F / (n-k)} < 1, \tag{6.9}$$

which is usually only the case if the hypothesis $H_0 : \mu_1 = \mu_2 = ... = \mu_k$ can not be rejected, which is, in turn, linked to small values of η^2. In these cases we let $\hat{\omega}^2 = 0$, because $\eta^2 > 0$ can not be estimated by a negative value.

2) The estimator $\hat{\omega}^2$ is the ANOVA analogue of R_a^2, the adjusted R^2 defined in paragraph 5.2.5, while $\hat{\omega}^2$ described in paragraph 5.3.2 is a special case with $k = 2$.

3) We can express $\hat{\omega}^2$ in terms of the variance ratio, F, as

$$\hat{\omega}^2 = \frac{F-1}{F + \frac{n-k}{k-1}}, \tag{6.10}$$

which is generalized in equation (5.39) in paragraph 5.3.3.

4) According to Carroll & Nordholm (1975) $\hat{\omega}^2$ is an estimator for

$$\omega^2 = \frac{(1/n)\sum_i n_i(\mu_i - \mu)^2}{\sigma^2 + (1/n)\sum_i n_i(\mu_i - \mu)^2} \;, \qquad (6.11)$$

where $\frac{n_i}{n} = \frac{N_i}{N}$, i.e. that n_i is eweredig is aan die populasiegrootte N_i. Let op dat dit na η^2 herlei as die n_i's gelyk is (en dus die N_i's ook gelyk).

5) Strictly speaking, $\hat{\omega}^2$ is not unbiased for η^2, but the following estimator is (Hays, 1973:486):

$$\hat{\eta}^2 = \frac{(n-k-2)F/(n-k)-1}{(n-k-2)F/(n-k)+(n-k)/(k-1)} \qquad (6.12)$$

Example 6.2:

Consider Example D in Chapter 3. Here $k=3$, $n=444$, $SS_G = 67,47$, $SS_E = 854,49$, $SS_{tot} = 921,96$ and $F = 17,41$.

$$\tilde{\eta}^2 = \frac{67,74}{921,96} = 0,0735$$

$$\hat{\omega}^2 = \frac{67,74 - 2\times 854,49/441}{921,96 + 854,49/441} = \frac{67,74 - 3,88}{921,96 + 1,94} = \frac{63,86}{923,90}$$

$$= 0,0690 .$$

This quantity is easier to calculate if we use equation (6.10):

$$\hat{\omega}^2 = \frac{17,41 - 1}{17,41 + \frac{441}{2}} = \frac{16,41}{237,91} = 0,0690$$

From equation (6.11) the estimator is:

$$\hat{\eta}^2 = \frac{439 \times 17,41/441 - 1}{439 \times 17,41/441 + 441/2} = \frac{17,33 - 1}{17,33 + 220,5} = 0,0687 .$$

Clearly $\tilde{\eta}^2$ produces a slightly higher value since it over estimates η^2. The estimated values obtained from $\hat{\omega}^2$ and $\hat{\eta}^2$ are, for practical purposes, the same.

□

6.1.2 Confidence intervals for η^2

Under the assumption that random samples are drawn from normal populations, Fowler (1985) provides an approximate $100(1-\alpha)\%$ *CI* for η^2 derived from the Laubscher-approximatione of the non-central F distribution. Fowler showed, through the use of simulation studies, that these intervals yield the correct coverage probabilities for 2, 4 or 8 groups each with 5, 10 or 20 observations per group. The variance-ratio F then follows a non-central F - distribution with $k-1$ and *n* - *k* degrees of freedom and non-centrality parameter $ncp_F = n\eta^2 / \left(1-\eta^2\right)$. As before, a *CI* for ncp will first be constructed. The boundaries of this interval are:

$$ncp_{FL} = \tfrac{1}{2}\left[wx + z^2_{\alpha/2}\left(x+c\right) - 2\left(k-1\right) + c \right] - z_{\alpha/2}\sqrt{wx\left(x+c\right)}$$

and (6.13)

$$ncp_{FU} = \tfrac{1}{2}\left[wx + z^2_{\alpha/2}\left(x+c\right) - 2\left(k-1\right) + c \right] + z_{\alpha/2}\sqrt{wx\left(x+c\right)}$$

where $w = 2\left(n-k\right) - 1$

$x = \left(k-1\right)F / \left(n-k\right)$

$c = \left(k-1+2nx\right) / \left(k-1+nx\right)$.

The approximate *CI*-boundaries for η^2 then follow from the definition of ncp_F and are given by:

$$\eta^2_{L,approx} = ncp_{FL} / \left(ncp_{FL} + n\right)$$

and (6.14)

$$\eta^2_{U,approx} = ncp_{FU} / \left(ncp_{FU} + n\right)$$

One can also construct an exact $100(1-\pm)\%$ *CI,* $\left(\eta^2_L;\eta^2_U\right)$, by making use of the SAS-program VI_R2 which was already discussed in paragraph 5.2.5. The inputs for this function are $u = k-1$, n and F. The program also calculates the estimators $\tilde{\eta}^2$ and $\hat{\eta}^2$ (as $R2$ and $R2a$).

Example 6.3:

Consider Example 6.2. To determine the 95% *CI* for η^2, first calculate

$$w = 2\times 441 - 1 = 883$$

$$x = 2\times 17,41/441 = 0,079$$

$$c = \left(3 - 1 + 2\times 444\times 0,079\right)/\left(3 - 1 + 444\times 0,079\right)$$

$$=72,113/37,057 = 2,057$$

$$ncp_{FL} = \tfrac{1}{2}\left[883\times 0,079 + 1,96^2\left(0,079 + 2,057\right) - 2\times 2 + 2,057\right]$$

$$-1,96\sqrt{883\times 0,079\left(0,079 + 2,057\right)}$$

$$= \tfrac{1}{2}\left[69,757 + 8,206 - 6,057\right] - 23,925$$

$$= 35,953 - 23,925$$

$$= 12,028$$

$$ncp_{FU} = 35,953 + 23,925$$

$$= 59,878$$

$$\eta^2_{L,approx} = 12,028/\left(12,028 + 444\right) = 0,026$$

$$\eta^2_{U,approx} = 59,878/\left(59,878 + 444\right) = 0,119\,.$$

The exact 95% *CI* for ncp_F obtained from the $R2\_VI$ program is $(14,387\,;61,194)$, while the *CI* for η^2 is: $(0,032\,;0,121)$.

It would appear that the exact interval produces slightly higher values for both the lower and upper bounds of the interval. □

6.1.3 Comparing more than two proportions

As in the case with the differences in proportions discussed in paragraph 5.4.4, proportions can be treated as means and the effect sizes can be calculated as in previous paragraphs. Let

$$y_{ij} = \begin{cases} 1 \text{ if population } i'^{s} j\text{-th element is positive} \\ 0 \text{ otherwise} \end{cases}$$

and

$$Y_{ij} = \begin{cases} 1 \text{ if the sample from population } i' \, s \; j\text{-th element is positive} \\ 0 \text{ otherwise} \end{cases}.$$

Let N_i and n_i denote the population and sample sizes respectively, and the analysis of variance (ANOVA) tables for y_{ij} and Y_{ij} are given by (a) and (b) below (D'Agostino, 1972):

	Source of variation	Sum of Squares	Degrees of freedom	Mean Sum of Squares
a)	Between populations	$\sum_i N_i(\pi_i - \bar{\pi})^2 = a$	-	-
	Within populations	$\sum_i N_i \pi_i (1-\pi_i) \; = b$	-	-
	Total	$N\bar{\pi}(1-\bar{\pi}) \; = Nc$	-	-

b)	Between samples	$\sum_i n_i(p_i - \overline{p})^2 = A$	$k-1$	$\frac{A}{k-1}$
	Within samples	$\sum_i n_i p_i(1-p_i) \ = B$	$n-k$	$\frac{B}{n-k}$
	Total	$n\overline{p}(1-\overline{p}) \ = nC$	$n-1$	

Here $\pi_i = \frac{1}{N_i}\sum_j y_{ij}$, denotes the *i*-th population proportion, while $p_i = \frac{1}{n_i}\sum_j Y_{ij}$, denotes the *i*-th sample proportion. Further, let $\overline{\pi} = \frac{1}{N}\sum_i N_i\pi_i$ and $\overline{p} = \frac{1}{n}\sum_i n_i p_i$, where $N = \sum_i N_i$ and $n = \sum_i n_i$.

From (6.7), the "weighted" effect size for the equality of *k* unequal sized populations' proportions is:

$$\eta^2_{pw} = \frac{a}{Nc} = \frac{\sum_i N_i(\pi_i - \overline{\pi})^2}{N\overline{\pi}(1-\overline{\pi})} \qquad (6.15)$$

If $N_1 = N_2 = ... = N_k$, then the unweighted value becomes

$$\eta^2_p = \frac{\frac{1}{k}\sum_i(\pi_i - \overline{\pi})^2}{\overline{\pi}(1-\overline{\pi})}, \qquad (6.16)$$

which is the proportion analogue of η^2 in (6.5).

The sample effect size is, according to the ANOVA table (b), given by

$$\tilde{\eta}_P^2 = \frac{A}{nC} = \frac{\sum_i n_i (p_i - \overline{p})^2}{n\overline{p}(1-\overline{p})} \tag{6.17}$$

This is the analogue of (6.7) and ${\eta_p}^2$ is overestimated.

The ratio of variances obtained from ANOVA table (b) in terms of proportions becomes:

$$F_P = \frac{\frac{A}{k-1}}{\frac{B}{n-k}} = \frac{\sum_i \frac{n_i (p_i - \overline{p})^2}{k-1}}{\sum_i \frac{n_i p_i (1-p_i)}{n-k}} \; . \tag{6.18}$$

For large samples F_p has an approximate non-central $F_{k-1,n-k}$-distribution with non-centrality $\frac{n\eta_p^2}{1-\eta_p^2}$. Consequently it is an unbiased estimator for ${\eta_p}^2$ from (6.12):

$$\hat{\eta}_p^2 = \frac{\frac{(n-k-2)F_p}{n-k} - 1}{\frac{(n-k-2)F_p}{n-k} + \frac{n-k}{k-1}} \; . \tag{6.19}$$

Remarks:

The k populations' proportions can be obtained from the following 2 x k - contingency table:

	Populations		
	1	2	k
Positive	$N_1\pi_1$	$N_2\pi_2$	$N_k\pi_k$
Negative	$N_1(1-\pi_1)$	$N_2(1-\pi_2)$	$N_k(1-\pi_k)$

According to D'Agostino (1972) the chi-squared statistics then becomes

$$x^2 = \frac{a}{c} = \frac{\sum_i N_i(\pi_i - \overline{\pi})^2}{\overline{\pi}(1-\overline{\pi})} = N\eta_{pw}^2 . \qquad (6.20)$$

The sample analogue is then:

$$X^2 = \frac{A}{C} = \frac{\sum_i n_i(p_i - \overline{p})^2}{\overline{p}(1-\overline{p})} = n\tilde{\eta}_p^2 . \qquad (6.21)$$

From these results it follows that η_{pw}^2 and $\tilde{\eta}_p^2$ are the same as w^2 for a (2 x k)-contingency table. In the case where equal population sizes are used, then $\hat{\eta}_p^2$ can be used as an unbiased estimator for w^2.

Example 6.4 (Cohen, 1969: 219):

The following proportional grouping is made with respect to the gender and political preference in America:

	Democrat	Republican	Independent	Gender proportions
Male	0,22	0,35	0,03	0,60
Female	0,23	0,10	0,07	0,40
Preference proportions	0,45	0,45	0,10	1,00

In order to calculate the effect sizes η_{pw}^2, let π_1, π_2 and π_3 denote the proportion of men affiliated with each party and, without loss of generality, let $N_1 = 45$, $N_2 = 45$, $N_3 = 10$ and $N = 100$.

Now, because $\pi_1 = \frac{0,22}{0,45} = 0,489$, $\pi_2 = \frac{0,35}{0,45} = 0,778$ and $\pi_3 = 0,3$, we have that

$\overline{\pi} = 0,45 \times 0,489 + 0,45 \times 0,778 + 0,10 \times 0,30 = 0,6$ (= proportion men).

$$\eta_{pw}^2 = \frac{45(0,489-0,6)^2 + 45(0,778-0,6)^2 + 10(0,3-0,6)^2}{100 \times 0,6 \times 0,4}$$

$= \frac{2,88}{24} = 0,12$,

while $w = \sqrt{\eta^2_{pw}} = 0,346$.

Note that if equal weights (population sizes) are used for each party then we would have $\pi = \frac{1}{3}(0,489 + 0,778 + 0,3) = 0,522,$ so that

$$\eta^2_p = \frac{(0,489 - 0,522)^2 + (0,778 - 0,522)^2 + (0,3 - 0,522)^2}{0,522 \times 0,478}$$

$= \frac{0,116}{0,250} = 0,465$, $w = 0,682$.

.

This differs greatly from the value of $\sqrt{\eta^2_{pw}}$.

6.1.4 Guideline values for the omnibus effect η^2

Cohen (1969, 1977, 1988)'s finds that from δ_{omn} in (6.1) it follows that, if $k = 2$, then $\delta_{omn} = \delta$ (defined in Chapter 4) and $\sigma_\mu = \frac{1}{2}(\mu_1 - \mu_2)$ so that

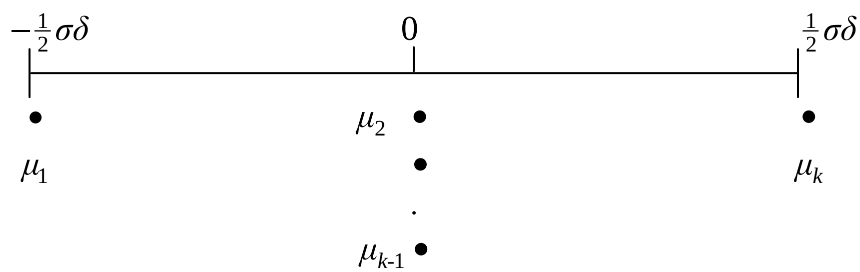
$$f = \tfrac{1}{2}\delta . \qquad (6.22)$$

This relationship allows us to state that the values $0,1$; $0,25$ and $0,4$ can be seen as small, medium and large effects for f . The values are derived from $\delta = 0,2$; $0,5$ and $0,8$. By making use of the relationships between η^2 and f in (6.6) Cohen offers the following convenient guideline values for η^2:

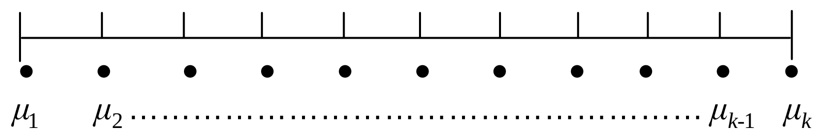
- Small effect : $\eta^2 = 0,01$
- Medium effect : $\eta^2 = 0,06$
- Large effect : $\eta^2 = 0,14$.

These values were only obtained for the proportion variance of population membership in the case involving only 2 populations. Cohen attempts to provide a motivation for why it might also be applicable for $k > 2$ populations. He considers three different patterns which describe the variation of the means $\mu_1, \mu_2, ..., \mu_k$ over the interval $\left(-\frac{1}{2}\sigma\delta;\ \frac{1}{2}\sigma\delta\right)$:

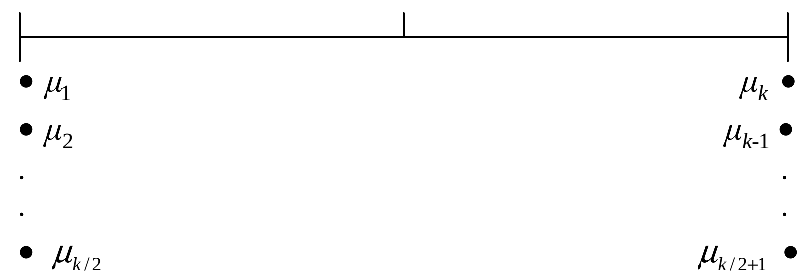
Pattern 1:



Pattern 2:



Pattern 3:



Note that pattern 1 reflects the smallest variation of μ's, and that pattern 3 reflects the largest (assuming that k is an even number, so that half of the μ's lie each of the two end points of the interval). Pattern 2 displays the situation where μ's are equally spaced, meaning that the variation in this case lies midway between the variation found in patterns 1 and 3.

For each of the patterns the quantity f (and consequently η^2 as well) can be expressed as a different function of δ and k. Table 6.1 lists these different expressions, as well as expressions of δ in terms of f. Examples of these expressions are also provided for $k=2$ and 8:

Table 6.1

	f			δ		
Pattern	general	$k=2$	$k=8$	general	$k=2$	$k=8$
1	$\delta\sqrt{\frac{1}{2k}}$	$0,5\delta$	$0,25\delta$	$f\sqrt{2k}$	$2f$	$4f$
2	$\frac{\delta}{2}\sqrt{\frac{k+1}{3(k-1)}}$	$0,5\delta$	$0,33\delta$	$2f\sqrt{\frac{3(k-1)}{k+1}}$	$2f$	$3,06f$
3	$\frac{1}{2}\delta$	$0,5\delta$	$0,5\delta$	$2f$	$2f$	$2f$

Note that for $k=2$ the expressions for each of the three patterns are identical and δ is the standardized difference in means as discussed in Chapter 4.

6.1.5 Motivation for the guideline values proposed by Cohen

1. *Small effect* $\left(f=0,1\ ;\ \eta^2=0,01\right)$**:** In this case the SD of the means is one tenth of the SD of the original measurements within the population. For pattern 1 and $k=2$. This agrees with the value $\delta=0,2$, while for $k=8$ it agrees with $\delta=0,4$. For $k=2$ we get $\delta=0,306$ for pattern 2 and $\delta=0,2$ for pattern 3. Pattern 3, which has the largest variation in the μ's of the three patterns, produces the same value at $k=8$ and $k=2$, i.e., $\delta=0,2$, which is considered a small effect. When the variation is smallest (as in pattern 1) the value of δ can become much larger, e.g., $\delta=0,4$ when $k=8$. This result follows from the fact that only two mean values lie away from the rest of the mean values and provision must be made for the possibility of widening the interval on which the

μ's vary. In terms of η^2, this means that the proportion variance ascribed to population membership is only 1%.

2. *Medium effect* $\left(f = 0,25 \;;\; \eta^2 = 0,06 \right)$: For $k = 2$, and pattern 3, this agrees with $\delta = 0,5$, which was previously considered a medium effect (see paragraph 4.5). In the case where there is a small degree of variation, as in pattern 1, we have $\delta = 1,0$ for $k = 8$. This implies that the extreme means differ by one SD. The proportion variation which can be attributed to population membership is now 6%. Cohen uses the example of mean IQ's of 7 groups, each consisting of a different profession, which have $\sigma = 12$ and are equally distributed on the interval 98-107. In this example a value of $f = 0,25$ is obtained.

3. *Large effect* $\left(f = 0,4 \;;\; \eta^2 = 0,14 \right)$: For $k = 2$, and pattern 3, this agrees with $\delta = 0,8$, which was previously considered a large effect. For pattern 1 this means that, if $k = 8$, the two extreme means differ by 1,6 SD's from one another. The proportion variance η^2 is now 14%. In the example involving the 7 groups of professions and their mean IQ's described above, the IQ's must vary between 98 and 112 in order to produce a value of $\delta = 0,4$.

6.2 Indices for omnibus effects for dependent measurements

Consider Example B from Chapter 3 where the three dependent measurements of the before, after and follow-up tests are recorded for each person within the control and experimental groups. If we are interested in treating the three tests as dependent groups when we compare them to one another, then an ANOVA with repeated measurements over the test opportunities should be conducted. For the independent groups in the previous paragraph we only had two sources of variation: between groups and within groups. With dependent measurements within groups, the within group variation can be further sub-divided into a between person variation (or more generally, a “subject” variation) and a person within groups variation, i.e. the person $\times$ group interaction, which is now considered to be the error variation.

Let σ_p^2 be the variance of the person (or subject) effect and σ_e^2 be the new error variance, the original error variance becomes

$$\sigma^2 = \sigma_p^2 + \sigma_e^2 \quad (6.23)$$

and the total variance in (6.4) can now be expressed as:

$$\sigma_t^2 = \sigma_\mu^2 + \sigma_p^2 + \sigma_e^2 \quad (6.24)$$

Similarly, the sum of squares, as discussed in paragraph 6.1.1, can now be defined as

$$SS_F = SS_P + SS_e \quad (6.25)$$

where SS_P: between person (subjects) sum of squares

SS_e: person within groups (error) sum of squares,

so that:

$$SS_{tot} = SS_G + SS_P + SS_e \quad (6.26)$$

As in paragraph 5.2.2 the partial η^2 can also be obtained (like the partial R^2) where one can remove the influence of variables which are not of primary interest. When we used independent measurements σ_μ^2 could be divided by the variance σ_t^2, as was shown in (6.4). Now, however, we divide by $\sigma_\mu^2 + \sigma_e^2$ instead of σ_t^2 as in (6.16) because σ_p^2 has nothing to do with the error variation. Therefore

$$partial\ \eta^2 = \frac{\sigma_\mu^2}{\sigma_\mu^2 + \sigma_e^2} \quad , \quad (6.27)$$

which, for dependent measurements, controls for the person effect which we are not interested in at present. The partial η^2 can thus be used as an *omnibus effect size index* in this case and it can be estimated by (Kline 2004a: Table 6.8) substituting the estimators of σ_μ^2 and σ_e^2 into (6.27):

$$partial\ \hat{\eta}^2 = \frac{meanSS_G - meanSS_e}{meanSS_G + \left(\frac{kn}{k-1} - 1\right) meanSS_e}, \qquad (6.28)$$

where $meanSS_G = SS_G / (k-1)$ and $meanSS_e = SS_e / (n-1)(k-1)$. Note that n now denotes the number of persons (subjects) and is no longer the total number of measurements as in independent measurement studies.

Example 6.5:

Consider Example A of Chapter 3. Here we find that there are 3 dependent measurements per person, meaning that there are $k = 3$ dependent groups. From Table A.2 it follows that $n = 25$ (only the experimental group) $meanSS_G = 516{,}19 / 2$ and $meanSS_e = 1183{,}15/(2 \times 24)$.

The estimation of partial $\hat{\eta}^2$:

$$partial\ \hat{\eta}^2 = \frac{258{,}10 - 24{,}65}{258{,}10 + 36{,}5 \times 24{,}65} = \frac{233{,}45}{1157{,}78} = 0{,}202.$$

This means that the proportion of the total variance, controlling for the person effect, which can be attributed to the tests is $0{,}202$. This indicates a large effect.

□

According to Kline(2004a: 191) there are, at present, no known computer packages which are capable of producing confidence intervals for the omnibus effect size indices *partial* η_G^2 for dependent groups. Therefore we will not provide a *CI* in this case.

6.2.1 Intra-class correlation coefficient:

We assume, for the effect size index partial η^2 and its estimator, that the group effect is a *fixed effect* and is controlled for the person effect. By fixed effect we mean that the treatments or tests used are chosen to be a fixed set of treatments, e.g., the before, after and follow-up tests Example A, Chapter 3. On the other hand, if the group effect is random, and the person effect is important, then

$$partial\ \hat{\eta}^2 = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_e^2} \quad , \qquad (6.29)$$

is an index which controls for the groups effect. This index is called the *intra-class correlation coefficient* ρ_I (Bartko,1966) and is estimated by (Kline, 2004a: Table 6.8):

$$\hat{\rho}_I = \frac{MS_p - MS_e}{MS_p - (k-1)MS_e} = \frac{F_p - 1}{F_p - k + 1} \qquad (6.30)$$

where $MS_p = mean\ SS_G$,

$MS_e = mean\ SS_e$,

from the one-way ANOVA model with person as the group factor and the variance ratio is $F_p = MS_p / MS_e$. An application of this is if there are k test items measured on each of the n persons and the items are considered to be a random sample from a population of test items which all measure the same property of a person (like, for example, a section in an interest test where 15 items in the test all measure the same interest). This index represents the proportion of the total variance which is ascribed to the variance between the persons, but it is also the joint correlation between any two items. This correlation is also known as the *reliability* of any of the k items (see, for example, Bartko 1966; Shrout & Fleiss, 1979).

Bartko (1966) shows that ρ_I should not only be considered as a proportion variation (as in equation (6.30)) but is, in actuality, a correlation. This follows because

$$\sigma_p^2 = Cov\left(x_i, x_j\right) \quad \text{and}$$

$$\sigma_p^2 + \sigma_e^2 = Var\left(x_i\right) = Var\left(x_j\right) \quad \text{, so that}$$

$$\rho_I = \frac{Cov\left(x_i, x_j\right)}{\sqrt{Var\left(x_i\right)}\sqrt{Var\left(x_j\right)}} = Cor\left(x_i, x_j\right) \quad ,$$

where x_i and x_j are the i^{th} and j^{th} measurements on the persons. Therefore, ρ_I can also be estimated by the mean inter-item correlation. If ρ_I is considered to be an effect size index, we can use the same guideline values which were used for the Pearson product moment correlation coefficient ρ or its sample analogue r.

Thus take $\rho_I = 0{,}1$: small effect

$\rho_I = 0{,}3$: medium effect

$\rho_I = 0{,}5$: large effect.

Clark & Watson (1995) recommend that the mean inter-item correlation should lie between $0{,}15$ and $0{,}5$, but they also accept that it will also depend on the underlying construct which must be measured by the items. This interval roughly agrees with the interval $0{,}1$ - $0{,}5$ of guideline values which were recommended. Clark & Watson show that it is evidently important to investigate the values of the individual inter-item correlations. These values should also lie within the interval $0{,}15$ - $0{,}5$ and should also be reasonable homogenous: in their, rather eloquent, wording: “the inter-correlation matrix should appear as a calm but insistent sea or small but highly similar correlations”. This condition ensures that

$\hat{\rho}_I$ provides an estimator for any inter-item correlation, which should all be equal to one another.

6.2.2 Cronbach alpha coefficient

While ρ_I is the reliability of a single item, it is occasionally also important to determine the reliability of the mean or sum of k items. If one assumes that the items have equal reliability ρ_I and the items all have the same variance, then reliability of the mean over k measurements can be obtained from ρ_I by applying the Spearman-Brown-formula (see Steyn, 2004: 10):

$$\rho_{xx}^{(k)} = \frac{k\rho_I}{(k-1)\rho_I + 1} \quad , \tag{6.31}$$

and it is estimated by the *Cronbach* $\boldsymbol{\alpha}$ *- coefficient*

$$\alpha = \frac{k}{k-1}\left[1 - \frac{\sum_{i=1}^{k} Var(x_i)}{Var\left(\sum_{i=1}^{k} x_i\right)}\right] \quad , \tag{6.32}$$

or by

$$\hat{\rho}_{xx}^{(k)} = \frac{k\hat{\rho}_I}{(k-1)\hat{\rho}_I + 1} \ . \tag{6.33}$$

Note that while $\hat{\rho}_{xx}^{(k)}$ is based on the results of an ANOVA or the inter-item correlations, only the item variances and the variance of the sum of the items are used in the calculation of α.

From (6.31), guideline values for $\rho_{xx}^{(k)}$ (and also $\hat{\rho}_{xx}^{(k)}$, α) can be determined if we accept those proposed for ρ_I. The Cronbach alpha is a function of the number of items, k, and Table 6.1 can be used as a guideline to gauge the size of the indices. From Table 6.1 it follows that if, for example, $\hat{\rho}_I$ is as small as $0{,}1$ then the value of α is $0{,}69$ when $k = 20$, but it is $0{,}18$ if $k = 2$. With large values of $\hat{\rho}_I$, then α is large even when there are very few items.

Table 6.2: Cronbach-alpha values

		Number of items						
Effect	Intra-class-correlation	2	3	4	5	10	20	50
Small	0.1	0.18	0.25	0.31	0.36	0.53	0.69	0.85
	0.2	0.33	0.43	0.50	0.56	0.71	0.83	0.93
Medium	0.3	0.46	0.56	0.63	0.68	0.81	0.90	0.96
	0.4	0.57	0.67	0.73	0.77	0.87	0.93	0.97
Large	0.5	0.67	0.75	0.80	0.83	0.91	0.95	0.98

Example 6.6:

Consider Example G discussed in Chapter 3. The items in this example can be considered to be a random effect, this means that the estimation of the partial η_p^2 is given by the intra-class correlation coefficient:

$$\hat{\rho}_I = \frac{6{,}830 - 1{,}404}{6{,}830 + 9 \times 1{,}404} = 0{,}279.$$

This provides an estimation of the joint correlation between any two items, and the value is interpreted as a medium effect. Note that the mean of all the inter-item correlations in Table G.2 (the $r = 1$ is included on the diagonal) is $0,286$, which does not differ greatly from $\hat{\rho}_I$.

The reliability of the mean (or sum) of the 10 items is estimated by:

$$\hat{\rho}_{xx}^{(k)} = \frac{10 \times 0,279}{9 \times 0,279 + 1} = 0,795,$$

while the Cronbach-α value obtained from Table G.3 is:

$$\alpha = \frac{10}{9}\left(1 - \frac{19,47}{68,30}\right) = 0,794,$$

which are, for practical purposes, the same values. From Table 6.2 for k = 10, it would appear that this is a medium effect.

Despite the fact that the mean inter-item correlation obtained for Table G.2 is $0,2863$, the inter-correlations are actually quite different. In particular, items 5 and 6 which produce small and even negative correlations. This is an indication that these items do not belong to the underlying construct which we would like these items to be measuring and should rather be removed. Without items 5 and 6 the mean inter-item correlation is now $0,429$ and $\alpha = 0,855$. Therefore, the reliability of one item and that of the remaining 8 items both tend toward large effects.

□

6.2.3 Confidence intervals for ρ_I and $\rho_{xx}^{(k)}$

Under the assumption of normality of the items, it follows that $F_p = MS_p / MS_F$ has an F-distribution with $n(k-1)$ and $n-1$ degrees of freedom. Consequently, the $(1-\alpha)100\%$ *CI* for the variance ratio σ_p^2 / σ_e^2's boundaries (see Shrout & Fleiss, 1979) is:

$$F_L = F_p / F_{\alpha/2}\left(n-1; n(k\text{-}1)\right)$$

and

$$F_U = F_p \cdot F_{\alpha/2}\left(n(k\text{-}1); n\text{-}1\right),$$

so that the approximate *CI* for ρ_I is:

$$\left(\rho_{I,L}; \rho_{I,U}\right) = \left(\frac{F_L - 1}{F_L + k - 1}; \frac{F_U - 1}{F_U + k - 1}\right). \tag{6.34}$$

By applying the Spearman-Brown formula, the approximate $(1-\alpha)100\%$ *CI* for $\rho_{xx}^{(k)}$ is:

$$\left(\frac{k\rho_{I,L}}{(k-1)\rho_{I,L} + 1}; \frac{k\rho_{I,U}}{(k-1)\rho_{I,U} + 1}\right) \tag{6.35}$$

Example 6.7:

Consider Example 6.6 where items 5 and 6 have been removed, so that $k = 8$, and, from Table G.4 it follows that $MS_p = 7,494,$ $MS_E = 1,092$ so that $F_p = 7,494/1,092 = 6,863$. For a 95% *CI* : $F_{0,025}(700;99) = 1,372$ and $F_{0,025}(99;700) = 1,326$, so that $F_L = 6,863/1,326 = 5,176$ and $F_U = 6,863 \times 1,372 = 9,416.$

$$\left(\rho_{I,L}; \rho_{I,U}\right) = \left(\frac{5,176\text{ - }1}{5,176 + 8\text{ - }1}; \frac{9,416\text{ - }1}{9,416 + 8\text{ - }1}\right) = (0,343; 0,513)$$

For $\rho_{xx}^{(8)}$ the *CI* is:

$$\left(\frac{8 \times 0,343}{7 \times 0,343 + 1}; \frac{8 \times 0,513}{7 \times 0,513 + 1}\right) = (0,807; 0,894).$$

The value ρ_I can thus be as small as $0,34$, which is a medium effect, but it can also be as large as $0,51$, which is a large effect. The reliability of the mean of the 8 items can be considered to be medium to large according to the guidelines in Table 6.1.

6.2.4 Limits of agreement and reliability coefficients (*Yi* et.al, 2008)

Suppose that *k* different measurement instruments, that are supposed to measure the same trait, are applied to the elements of a random sample of *n* individuals or objects. The following question can now be posed: To what degree do the different measurement instruments agree with one another?
The following model describes the measurements Y_{ij} obtained:

$$Y_{ij} = \mu + a_i + e_{ij}\,, \qquad (6.36)$$

where Y_{ij} denotes the measurement of the *j*-th measurement instrument on the *i*-th object, μ denotes the corresponding mean measurement, and a_i is a random effect that varies for each object with a variance of σ_B^2 (called the between object variation). The error term e_{ij} has a variance of σ_w^2, and it is called the within object variation. The means of both a_i and e_{ij} are assumed to be 0.

- Two measurement instruments (k = 2):

Let the difference between two measurements on object *i* be given by:

$$D_i = Y_{i1} - Y_{i2} = e_{i1} - e_{i2}, \text{ such that}$$

$$E(D_i) = 0 \text{ and } Var(D_i) = 2\sigma_w^2.$$

Let $\bar{D} = \frac{1}{n}\sum_{i=1}^{n} D_i$ and $S_D = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(D_i - \bar{D}\right)^2}$ denote the sample mean difference and its standard deviation (SD) respectively. The interval $\bar{D} \pm 1{,}96 S_D$ represents the estimated interval where 95% of the differences in the population from which the sample of objects originated, should fall. This interval is known as the *limits of agreement*. If the scientifically acceptable difference lies outside of these limits then it indicates agreement between the two measurement instruments for the given data set.

If σ_w is estimated from the data as $\hat{\sigma}_w$, then the *repeatability coefficient* is defined as:

$$h = 1{,}96\sqrt{2}\hat{\sigma}_w = 1{,}96 S_D, \qquad (6.37)$$

which is the upper limit of agreement if $\bar{D} = 0$.

- More than two measurement instruments (k > 2):

Here we find that the limits of agreement can no longer be used. However, we can still make use of the repeatability coefficient:

$$h = 1{,}96\sqrt{2}\hat{\sigma}_w ,$$

with $\hat{\sigma}_w^2$ the error mean squares from an repeated measures analysis of variance (ANOVA):

$$\hat{\sigma}_w^2 = SSW / \left[n(k-1)\right] = \sum_{i=1}^{n}\sum_{j=1}^{k}\left(Y_{ij} - \bar{Y}_i\right)^2 / \left[n(k-1)\right], \quad (6.38)$$

where $\bar{Y}_i = \frac{1}{k}\sum_{j=1}^{k} Y_{ij}$.

Motivation: SSW can also be written as:

$$SSW = \sum_{i=1}^{n}\sum_{\substack{j-1 \\ j\neq j'}}^{k}\sum_{j'=1}^{k}\left(Y_{ij} - Y_{ij'}\right)^2 / k,$$

where $Y_{ij} - Y_{ij'}$ are the $k(k-1)/2$ paired differences for object i. Let $\bar{D}_i^2$ denote the mean of the squared differences of the i-th object, then

$$SSW = \sum_{i=1}^{n} \left(\frac{k(k-1)}{2} \right) \bar{D}_i^2 / k$$

$$= \sum_{i=1}^{n} (k-1) \bar{D}_i^2 / 2 \ ,$$

thus $\hat{\sigma}_w^2 = \sum_{i=1}^{n} \bar{D}_i^2 / 2n$, which denotes the mean of the squared differences between paired measurements.

If one assumes that $E(Y_{ij} - Y_{ij'}) = 0$ and that the differences are normally distributed with variance $2\sigma_w^2$, then for the difference D:

$$P\left(0 < |D| < 1{,}96\sqrt{2}\hat{\sigma}_w\right)$$

$$= P\left(-1{,}96\sqrt{2}\hat{\sigma}_w < D < 1{,}96\sqrt{2}\hat{\sigma}_w\right)$$

$$\doteq 0{,}95. \quad (6.39)$$

This means that the *absolute* pairwise differences between the measurement instruments lie between 0 and $1{,}96\sqrt{2}\hat{\sigma}_w$ with probability 0,95. If the repeatability coefficient is smaller than the scientifically acceptable absolute difference then it indicates a strong agreement.

Under the assumption of normality of the error terms e_{ij}, one finds that $\hat{\sigma}_w^2$ follows a Chi-squared distribution with $n(k-1)$ degrees of freedom. This means that the *upper bound of the 95% confidence interval for the population repeatability coefficient* is: $B_h = \sqrt{\frac{(1{,}96)^2 \, 2n(k-1)\hat{\sigma}_w^2}{\chi^2_{\alpha,n(k-1)}}}$, (6.40)

and with k = 2: $B_h = \frac{1{,}96\sqrt{n}S_D}{\sqrt{\chi^2_{\alpha,n}}}$, (6.41)

where $\chi^2_{\alpha,\nu}$ is the 100α-th percentile of a Chi-squared distribution with ν degrees of freedom.

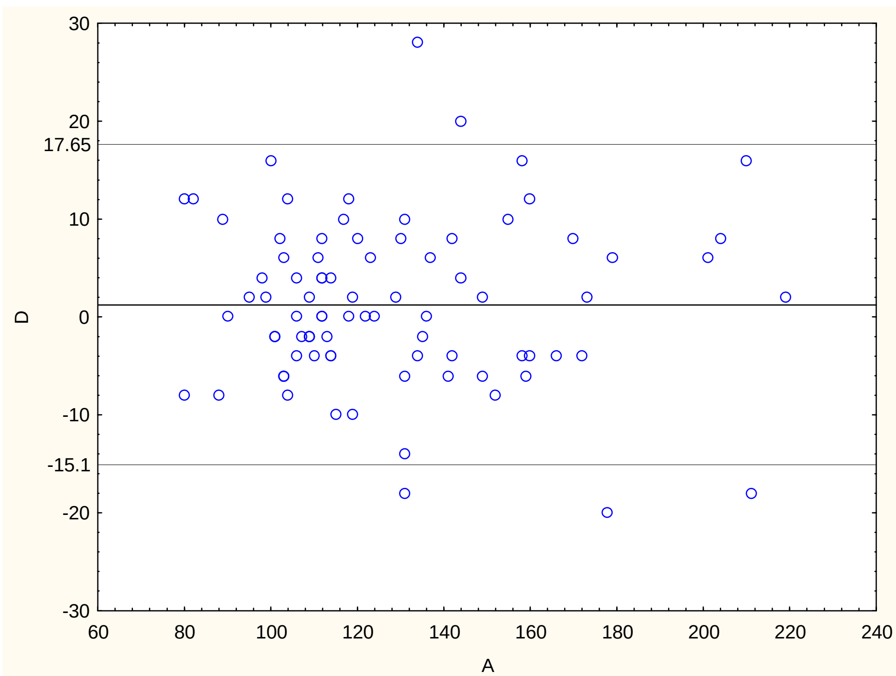
If B_h is smaller than the scientifically accepted absolute difference between the measurement instruments, then there is a statistically significant agreement at a 5% significance level.

Example 6.8 (Yi et al., 2008, Example2):

The blood pressures of 85 people are measured 3 times by an observer (Table II, Yi et al.)

a) First 2 measurements (J1 and J2):

Bland-Altman plot (Bland & Altman, 1986) of $D = J1 - J2$ and $A = \frac{1}{2}(J1 + J2)$:



The solid horizontal lines display the bias and limits of agreement:

$\bar{D} = 1.25$ *and* $\bar{D} \pm 1,96 S_D = 1,25 \pm 1,96 \times 8,366 = (-15,1;17,65)$. Some differences lie outside these limits.

$h = 1,96 \times 8,366 = 16,4$, with upper 95% confidence bound

$$B_h = \frac{1,96\sqrt{85 \times 8,366}}{\sqrt{64,75}} = 18,79.$$

The reliability coefficient h can thus be as large as 18,79 with probability 95%.

If the scientifically acceptable difference is 20, then there is an adequate agreement.

b) For all three measurements (J1 – J3):

Here we have that $\hat{\sigma}_w^2 = 37,408$, $SSW = 6359$, $h = 1,96\sqrt{2 \times 37,408} = 16,95$ and

$B_h = \dfrac{1,96\sqrt{2 \times 6359}}{140,85} = 18,62$.

□

6.3 Indices for contrast-effects

A *contrast of population means* $\mu_1,....\mu_k$ is defined (Kline, 2004a: 164) as:

$$\psi = c_1\mu_1 + c_2\mu_2 + ... + c_k\mu_k = \sum_{i=1}^{k} c_i\mu_i, \tag{6.42}$$

where c_1, c_2 …, c_k are the *contrast weights*, such that

$$c_1 + c_2 + \ldots + c_k = \sum_{i=1}^{k} c_i = 0. \tag{6.43}$$

Similarly, the contrast of the sample means $\overline{x}_1, \overline{x}_2, ..., \overline{x}_k$ is an estimator for ψ:

$$\hat{\psi} = c_1\overline{x}_1 + c_2\overline{x}_2 + ... + c_k\overline{x}_k = \sum_{i=1}^{k} c_i\overline{x}_i. \tag{6.44}$$

Example 6.9:

Consider Example 6.2 where the 3 populations were “non-aboriginals, urban” (population 1), “aboriginals, urban” (population 2) and “aboriginals, rural” (population 3). In Example 6.2 the omnibus effect η^2 is estimated as 0,069. However, we are actually interested in determining the following differences:

(a) $\mu_1 - \frac{1}{2}(\mu_2 + \mu_3)$, i.e., the non-aboriginals vs. aboriginals;

(b) $\mu_2 - \mu_3$, i.e., urban vs. rural aboriginals;

(c) $\mu_1 - \mu_2$, i.e., non-aboriginals vs. aboriginals in urban areas.

The contrast weights are now:

Contrast	c_1	c_2	c_3	Total
(a)	1	-½	-½	0
(b)	0	1	-1	0
(c)	1	-1	0	0

The contrasts with their estimators are thus:

(a)
$$\psi = (1)\mu_1 + \left(-\frac{1}{2}\right)\mu_2 + \left(-\frac{1}{2}\right)\mu_3 = \mu_1 - \frac{1}{2}\mu_2 - \frac{1}{2}\mu_3$$

and

$$\hat{\psi} = \bar{x}_1 - \frac{1}{2}\bar{x}_2 - \frac{1}{2}\bar{x}_3 \quad ;$$

(b)
$$\psi = (0)\mu_1 + (1)\mu_2 + (-1)\mu_3 = \mu_2 - \mu_3$$

and

$$\hat{\psi} = \bar{x}_2 - \bar{x}_3 \quad ;$$

(c)
$$\psi = (1)\mu_1 + (-1)\mu_2 + (0)\mu_3 = \mu_1 - \mu_2$$

and

$$\hat{\psi} = \bar{x}_1 - \bar{x}_2$$ □

When $k = 2$, then $\mu_1 - \mu_2$ is also a contrast. The effect size index used to compare μ_1 and μ_2, was the standardized difference:

$$\delta = \frac{\mu_1 - \mu_2}{\sigma^*} \quad ,$$

where σ^* is a standard deviation which can be defined in a number of different ways (see Chapter 4's introduction).

Similarly, an effect size can be defined by a *standardized contrast*:

$$\delta_\psi = \frac{\psi}{\sigma^*} \quad \text{and} \quad \hat{\delta}_\psi = \frac{\hat{\psi}}{\hat{\sigma}^*} \tag{6.45}$$

where $\hat{\delta}_\psi$ is the estimator for δ_ψ.

6.3.1 Choices of σ^* and $\hat{\sigma}^*$

Kline (2004a: 172) provides the following three choices:

1) Choose one population's SD (or sample's SD), usually the control or reference group, and then set $\sigma^* = \sigma_c$, where σ_c is the chosen population's SD. An estimator for σ^* is $\hat{\sigma}^* = s_c$, where s_c is the sample SD of the chosen group. This produces effect size indices which are similar to Glass's Δ (see paragraph 4.2).

2) Assume that the populations which are involved in the contrast all have the same SD, namely σ, and estimate $\sigma^* = \sigma$ by the pooled SD, $s_{p,\psi}$, where

$$s_{p,\psi}^2 = \frac{\left(n_{i_1} - 1\right)s_{i_1}^2 + \left(n_{i_2} - 1\right)s_{i2}^2 + \ldots + \left(n_{i_m} - 1\right)s_{im}^2}{n_{i_1} + n_{i_2} + \ldots + n_{i_m} - m} \quad , \tag{6.46}$$

where n_{i_j} and $s_{i_j}^2$ are the sample sizes and variances from population i_j, $j = 1, \ldots, m,$ and m is the number of populations involved in the contrast. For example, for contrast (c) in Example 6.7 it is

$$s_{p,\psi}^2 = \frac{\left(n_1 - 1\right)s_1^2 + \left(n_2 - 1\right)s_2^2}{n_1 + n_2 - 2},$$

which is the same as s_p^2 in (4.4).

3) Assume that all k populations have the same variance and estimate $\sigma^* = \sigma$ by the pooled SD of all of the populations, $s_{p,\psi}$, where $m = k$ as in (6.46). In Example 6.8 we have that

$$s_{p,\psi}^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + (n_3 - 1)s_3^2}{n_1 + n_2 + n_3 - 3} .$$

In addition to the three choices given by Kline, we can also include the following:

4) If we do not assume equal population SD's, then let σ^* be defined as the maximum of the SD's involved in the contrast:

$$\sigma_{max} = max\left(\sigma_{i_1}, \sigma_{i2}, ..., \sigma_{i_m}\right), \qquad (6.47)$$

where i_j is the index of the $j-th$ population involved. An obvious estimator is

$$s_{max} = max\left(s_{i_1}, s_{i2}, ..., s_{i_m}\right) \qquad (6.48)$$

5) If no assumption regarding equal populations are made and the population sizes are known, then define σ^* as σ_w, where

$$\sigma_w^2 = w_{i_1}\sigma_{i_1}^2 + w_{i_2}\sigma_{i_2}^2 + ... + w_{i_m}\sigma_{i_m}^2 , \qquad (6.49)$$

with estimator

$$s_w^2 = w_{i_1}s_{i_1}^2 + w_{i_2}s_{i_2}^2 + ... + w_{i_m}s_{i_m}^2 \qquad (6.50)$$

where

$$w_{i_1} = N_{i_1} / N_m , w_{i_2} = N_{i_2} / N_m , ..., w_{i_m} = N_{i_m} / N_m$$

and

$$N_m = N_{i_1} + N_{i_2} + ... + N_{i_m} \quad .$$

Example 6.10

In Chapter 3's Example B, the students can be further sub-divided into males and females. The following descriptive statistics for E/I accompany Table B.1:

Men			Women			Lecturers		
n_1	μ_1	σ_1	n_2	μ_2	σ_2	n_3	μ_3	σ_3
121	93,69	24,68	133	95,39	25,15	28	107,64	25,06

Suppose that (a) the students must be compared to the lecturers and (b) the men must be compared to the women.

(a)
$$\psi = \tfrac{1}{2}(\mu_1 + \mu_2) - \mu_3 = \tfrac{1}{2}(93{,}69 + 95{,}39) - 107{,}64$$
$$= -13{,}1$$

The effect sizes are then calculated for each choice of σ^*:

1) Assume that the lecturers are chosen as the reference population, then

$$\sigma^* = \sigma_3 = 25{,}06$$
$$13{,}1/25/06 = -0{,}523.$$

the population variances are known and unequal and all three populations are involved in the contrast, thus, let:

4)
$$\sigma^* = \sigma_{max} = \sigma_2 = 25{,}15$$
$$\delta_\psi = -13{,}1/25{,}15 = -0{,}521$$

5) $\sigma^* = \sigma_w$, with

$$\sigma_{2w}^2 = \frac{121 \times 24{,}68^2 + 133 \times 25{,}15^2 + 28 \times 25{,}06^2}{121 + 133 + 28}$$
$$= \frac{175410{,}98}{282} = 622{,}02,$$

so that

$$\delta_\psi = -13{,}1/\sqrt{622{,}02} = -0{,}525.$$

(b) $\psi = \mu_1 - \mu_2 = 93{,}69 - 95{,}39 = -1{,}7$

1) Using women as the reference point we have, $\sigma^* = \sigma_2 ,= 25{,}15$, thus

$$\delta_\psi = \frac{-1{,}7}{25{,}15} = -0{,}0676$$

Seeing as only the first 2 populations are involved and the variances are known:

4) $\sigma^* = \sigma_{max} = \sigma_2$

so that

$\delta_\psi = -0{,}0676$ as in 1.

5)
$$\sigma_w^2 = \frac{121 \times 24{,}68^2 + 133 \times 25{,}15^2}{121 + 133} = \frac{157826{,}88}{254}$$

$$= 621{,}37$$

$$\delta_\psi = -1{,}7 / \sqrt{621{,}37} = -0{,}0682.$$ □

6.3.2 Guideline values for effect sizes of contrasts

Contrasts are, in many cases, simply the difference between two means. Further, it can also be the difference between the average of group means. In Example 6.7 the contrasts (b) and (c) are both the differences between 2 means, while the contrast in (a) is the difference between the average of the two means, μ_2 and μ_3, and the mean μ_1.

The contrasts are then converted into effect size indices by standardizing them. This involves dividing the contrast by a standard deviation. They now have the same form as the standardized differences (e.g., δ) in Chapter 4. The same guideline values that were used for δ can thus also be used here:

- Small effect: $\delta_\psi = 0{,}2$

- Medium effect: $\delta_\psi = 0,5$
- Large effect: $\delta_\psi = 0,8$

In part (a) of Example 6.9, all of the different variations of the effect size indices based on the contrasts show a medium effect $(-0,521; -0,521$ and $-0,525)$, while the results from part (b) of Example 6.9 all show small effects $(-0,0676; -0,0676$ and $-0,0682)$.

6.3.3 Contrasts for dependent measurements:

When more than 2 measurements are made on each person or subject, then the means of these measurements (or combinations of these means) can also be compared by using contrasts. Example 6.1 provides some examples of these types of contrasts for the sample means of before test, after test and follow-up test results obtained from patients. Kline (2004a: 174) recommends two methods for obtaining contrasts in these situations:

1. Use σ^* and $\hat{\sigma}^*$ as in the previous paragraph. In this case we standardize ψ by dividing it by a SD which is on the same scale as he original measurements. This method does not consider the correlations which exist between the measurements.

2. Standardize ψ and $\hat{\psi}$ by dividing by $\sigma_{D\psi}$ or $s_{D\psi}$, which is the SD of the contrast differences within a person. In Example 6.1 the contrast $\overline{x}_V - \frac{1}{2}(\overline{x}_N + \overline{x}_O)$ is dividied by the SD of $D = x_V - \frac{1}{2}(x_N + x_O)$, which was calculated as the before test measurement minus the average of the after test and follow-up test measurements for each patient.

The effect size is now, as in paragraph 4.4, standardized using a different unit to the one used by the original measurements, namely, the units of the contrast difference. Depending on the inter-correlations between the measurements, it is possible that $\sigma_{D\psi}$ and σ^* can differ greatly.

Example 6.11

(a) Consider Example A found in Chapter 3. Suppose that we need to estimate the effect size of the contrast $\mu_V - \frac{1}{2}(\mu_N + \mu_O)$ for the experimental group. The sample's descriptive statistics can be found in Table A.1. We can then calculate $\hat{\psi}$ and $s_{D\psi}$ for BDI-values of the experimental group (EG):

n	$\hat{\psi}$	s_{D_ψ}
25	5,3	7,25

Different choices of $\hat{\sigma}^*$ will then produce the following effect sizes for ψ :

1) $\hat{\sigma}^* = s_v$ (The SD of the before test is used as reference):

$$\hat{\delta}_\psi = \frac{5{,}3}{6{,}14} = 0{,}86$$

2) $\hat{\sigma}^* = s_p$, the common SD, with $n_1 = n_2 = n_3$, so that:

$$s_p^2 = \frac{1}{3}\left(6{,}14^2 + 6{,}82^2 + 4{,}94^2\right) = 36{,}21, \text{ thus}$$

$$\hat{\delta}_\psi = \frac{5{,}3}{\sqrt{36{,}21}} = 0{,}88$$

3) $\hat{\sigma}^* = s_{max} = s_N = 6{,}82$, so that

$$\hat{\delta}_\psi = \frac{5{,}3}{6{,}82} = 0{,}78$$

These estimators use SD values that are measured in the same units as each of the before, after and follow-up test measurements. According to the guideline values proposed by Cohen, these are all large effects.

4) If, on the other hand, we make use of $\hat{\sigma}^* = s_{D\psi} = 7,25$,

we obtain $\hat{\delta}_\psi = \frac{5,3}{7,25} = 0,73$, which is somewhat smaller, but is standardized using the units of the contrast differences.

(b) Consider Example E, Chapter 3 and use the 5 cholesterol values of the heart patients who did not exercise to determine the effects of the following contrasts $\psi_1 = \mu_1 - \mu_0$, $\psi_2 = \mu_2 - \mu_1$ and $\psi_3 = \mu_4 - \mu_0$.

The following information (which can be used to supplement the contents of Table E.1) can be obtained from the data:

n	$\hat{\psi}_1$	$s_{\psi 1}$	$\hat{\psi}_2$	$s_{\psi 2}$	$\hat{\psi}_3$	$s_{\psi 3}$
19	8,05	9,24	6,95	8,22	25,0	28,91

1) Use the initial cholesterol values as the reference point, then

$$\hat{\delta}\psi_1 = \frac{8,05}{21,01} = 0,383\ ;\ \hat{\delta}\psi_2 = \frac{6,95}{21,01} = 0,331\ ;\ \hat{\delta}\psi_3 = \frac{25}{21,01} = 1,190$$

2) Do not assume that the SD's of the population cholesterol values over the years are equal. Use the maximum SD of the groups involved in each particular contrast:

$$\hat{\delta}\psi_1 = \frac{8,05}{26,41} = 0,305\ ;\ \hat{\delta}\psi_2 = \frac{6,95}{26,86} = 0,259\ ;\ \hat{\delta}\psi_3 = \frac{25}{42,03} = 0,595$$

3) If we want to obtain the effect sizes in terms of the contrast difference units, then:

$$\hat{\delta}\psi_1 = \frac{8,05}{9,24} = 0,871\ ;\ \hat{\delta}\psi_2 = \frac{6,95}{8,22} = 0,845\ ;\ \hat{\delta}\psi_3 = \frac{25}{28,91} = 0,865.$$

Note how the values of the effect size indices vary depending on which of the methods is used. In terms of the original measurements' units, ψ_1 and ψ_2 both have small effects and ψ_3 has a large to medium effect. However, when we use the same units as the contrast differences, then all three contrasts appear to have large effects .

□

6.3.4 Confidence intervals for $\delta_\psi = \psi / \sigma^*$ (independent samples)

An approximate $(1-\alpha)100\%$ *CI* for δ_ψ is obtained by dividing the boundaries of the *CI* for ψ by the estimator $s_{p,\psi}$ for σ^* as in (6.31).

The $(1-\alpha)100\%$ *CI* for ψ for the situation where we assume that the m populations involved in the contrast all have the *same SD,* σ, and are normally distributed, is:

$$\left(\psi_0 ; \psi_B\right) = \hat{\psi} \pm t_{\alpha/2}\left(n_m - m\right) s_{\hat{\psi}} , \qquad (6.51)$$

where

$$s_{\hat{\psi}}^2 = \left(\frac{c_1^2}{n_1} + \frac{c_2^2}{n_2} + \ldots + \frac{c_k^2}{n_k}\right) s_{p,\psi}^2 \qquad (6.52)$$

and

$t_{\alpha/2}\left(n_m - m\right)$ is the $\left(1-\overline{\alpha}\right)$-th percentile of the *t*-distribution with $n_m - m$ degrees of freedom and $n_m = n_{i_1} + n_{i_2} + \ldots + n_{i_m}$ is the total of the m samples involved in the contrast, ψ (Note: The expression given for $s_{\hat{\psi}}$ in Kline, 2004a:175 is incorrect).

The approximate $(1-\alpha)100\%$ *CI* for $s_\psi = \dfrac{\psi}{\sigma^*}$ is then:

$$\left(\frac{\psi_L}{\hat{\sigma}^*} \; ; \; \frac{\psi_U}{\hat{\sigma}^*}\right). \qquad (6.53)$$

Notes:

1. If n_m is large, then the assumption of normally distributed populations is no longer necessary and we substitute $t_{\alpha/2}(n_m - m)$ with $z_{\alpha/2}$, the normal $(1-\overline{\alpha})$ percentile.
2. If $s_{p,\psi}$ is based on all k samples, then define: $n_m = n = n_1 + n_2 + ... + n_k$.

As before, it is possible to obtain an *exact CI* for δ_ψ using the same assumptions described above. The following steps can be used to create this interval: (Kline, 2004a: 177-179):

1. Calculate

$$t_{\hat{\psi}} = \frac{\hat{\psi}}{s_{\hat{\psi}}} \quad , \tag{6.54}$$

   where $s_{\hat{\psi}}$ is given in equation (6.52).

2. Obtain, through the use of the SAS-program (see the webpage of this manual), the *VI_delta_contrast,* which calculates a $(1-\alpha)100\%$ *CI* for ncp_ψ $(ncp_{\psi,L}; ncp_{\psi,U})$, and where

$$ncp_\psi = \delta_\psi / \sqrt{\sum_{i=1}^{k} \frac{c_1^2}{n_i}} \quad , \tag{6.55}$$

   The non-centrality parameter comes from a non-central *t*-distribution with $n_m - m$ degrees of freedom. The inputs for this program are $t_{\hat{\psi}}, c_1, c_2$, …, c_k, n_1, n_2, …, n_k and α (for some theortetical background, please see Appendix A).

3. The boundaries for the *CI* for δ_ψ can be expressed as:

$$\delta_{\psi,L} = ncp_{\psi,L} \sqrt{\sum_{i=1}^{k} \frac{c_1^2}{n_i}}$$

and (6.56)

$$\delta_{\psi,U} = ncp_{\psi,U}\sqrt{\sum_{i=1}^{k}\frac{c_1^2}{n_i}} \,.$$

Note that while the approximate *CI* for δ_ψ can be used for any choice of σ^* (i.e., σ, σ_{max} or any of $\sigma_1, \sigma_2,...,\sigma_k$), when using the estimator $\hat{\sigma}^*$ or σ^* the exact *CI* can only be obtained for $\delta_\psi = \psi / \sigma$ (i.e., where we assume equal SD's for the populations).

The computer program *PSY* by Bird et al. (2000) can, from a given set of data, construct confidence intervals using contrasts and also standardize these contrasts using $s_{p,\psi}$ (which is based on all of the groups). Standardized *CI*'s can also be obtained from this program. It thus produces the effect size indices given in equation (6.31) with $\hat{\sigma}^* = s_{p,\psi}$ and their *CI*'s, as given in equation (6.51). These programs are available on the web page of this manual.

Example 6.12:

Consider Example 6.9's contrast

(a) $\psi=\mu_1 - \frac{1}{2}(\mu_2 + \mu_3)$ so, from Table D.1, it follows that:

$$\begin{aligned}\hat{\psi} &= 13{,}138 - \frac{1}{2}(12{,}66 + 12{,}57)\\ &= 0{,}523\end{aligned}$$

From Table D.2 we have that $s_{p,\psi}^2 = 1{,}94$ and further

$$\begin{aligned}s_{\hat{\psi}}^2 = 1{,}94\left(\frac{1}{159} + \frac{\left(-\frac{1}{2}\right)^2}{94} + \frac{\left(-\frac{1}{2}\right)^2}{191}\right) &= 1{,}94(0{,}0063 + 0{,}0027 + 0{,}0013)\\ &= 0{,}0199\end{aligned}$$

Thus, $\hat{\delta}\psi = \hat{\psi} / s_{p,\psi} = \frac{0{,}523}{1{,}94} = 0{,}375$.

The boundaries of the 95% *CI* for ψ are now $0,523 \pm 1,96\sqrt{0,0199} = 0,523 \pm 0,276 = (0,247\ ;\ 0,799)$.

Subsequently, the approximate 95% *CI* for δ_ψ is then:

$$\left(0,247 / \sqrt{1,94}\ ;\ 0,799 / \sqrt{1,94}\right) = (0,177\ ;\ 0,573).$$

The exact 95% *CI* for δ_ψ can be calculated using the program *VI_delta-contrast* by using the following input values: $c_1 = 1$, $c_2 = -\frac{1}{2}$, $c_3 = -\frac{1}{2}$, $n_1 = 159$, $n_2 = 94$ and $n_3 = 191$; $t_{\hat{\psi}} = \hat{\psi} / s_{\hat{\psi}} = 0,523 / \sqrt{0,0199} = 3,707$. Then, for ncp_ψ, the interval obtained is $(1,730\ ; 5,680)$ and for $\delta_\psi : (0,175\ ;\ 0,576)$, which is very similar to the approximate interval. These confidence boundaries indicate that the contrasts can have a small to medium effect.

□

6.3.5 Confidence intervals for $\delta_\psi = \psi / \sigma^*$ (dependent samples)

The $(1-\alpha)100\%$ *CI* for δ_ψ can be determined using arguments similar to those discussed in the previous paragraph, except that the standard error of $\hat{\psi}$, i.e., $s_{\hat{\psi}}$, is now given by

$$s_{\hat{\psi}}^2 = \frac{1}{n}\left(c_1^2 + c_2^2 + \ldots + c_k^2\right) s_{D_\psi}^2 \quad , \tag{6.57}$$

where n is the number of persons or subjects and s_{D_ψ} is the SD of

$$D_\psi = c_1 x_1 + c_2 x_2 + \ldots + c_k x_k , \tag{6.58}$$

where $x_1, x_2, \ldots, x_k$ are the dependent measurements per person (or subject).

Thus, the $(1-\alpha)100\%$ *CI* for ψ is:

$$(\psi_L, \psi_U) = \hat{\psi} \pm t_{\alpha/2}(n-1) s_\psi^2. \tag{6.59}$$

The approximate $(1-\alpha)100\%$ *CI* for ψ/σ^* is once again given by (6.39), i.e., by dividing the boundaries for ψ by $\hat{\sigma}^*$.

The exact *CI* for $\psi/\sigma_{D\psi}$ (which are standardized with the SD of the contrast differences) can be constructed in much the same way as was done in paragraph 4.4.1. In this case we have $t_{\hat{\psi}} = \hat{\psi}/s_{\hat{\psi}}$, $ncp_{\psi} = \sqrt{n}\delta_{\psi} / \sqrt{\sum_{i=1}^{k} c_i^2}$, and the CI for ncp_{ψ} is $\left(ncp_{\psi,L}\ ;\ ncp_{\psi,U}\right)$. It follows that

$$\delta_{\psi,L} = ncp_{\psi,L} \sum_{i=1}^{k} c_i^2 / \sqrt{n}$$

and (6.60)

$$\delta_{\psi,U} = ncp_{\psi,U} \sum_{i=1}^{k} c_i^2 / \sqrt{n}$$

The program *VI_delta_contrastD* can be used here. It is available op the web page.

Example 6.13:

In Example 6.10, the effect size of the contrast $\psi = \mu_V - \frac{1}{2}(\mu_N + \mu_O)$ is obtained in four different ways. In this case $n = 25$, $c_1 = 1$, $c_2 = -\frac{1}{2}$, $c_3 = -\frac{1}{2}$, $\hat{\psi} = 5{,}3$ and $s_{D_\psi} = 7{,}25$. The 95% *CI* for ψ is thus:

$$\begin{aligned}(\psi_O, \psi_B) &= 5{,}3 \pm t_{0{,}025}(24)\sqrt{1^2 + \left(-\frac{1}{2}\right)^2 + \left(-\frac{1}{2}\right)^2} \cdot 7{,}25/\sqrt{25} \\ &= 5{,}3 \pm 2{,}064\sqrt{1{,}5} \times 7{,}25/5 \\ &= 5{,}3 \pm 3{,}665 \\ &= (1{,}635\ ;\ 8{,}965)\end{aligned}$$

The approximate *CI* for δ_ψ, if $\sigma^* = \sigma_V$, is then: $(1{,}635/6{,}14\ ;\ 8{,}965/6{,}14) = (0{,}266\ ;\ 1{,}460)$, which means that δ_ψ can vary between a small effect to a large effect with 95% probability. Similarly, *CI*'s can be obtained when we choose $\hat{\sigma}^* = s_p$ or $\hat{\sigma}^* = s_{max}$. In the situation where one standardizes using the units of the contrast difference, then the following is approximately true:

$(1{,}635/7{,}25\ ;\ 8{,}965/7{,}25) = (0{,}226\ ;\ 1{,}237)$, which can be confirmed with the exact 95% *CI*:

$t_{\hat{\psi}} = 5{,}3/\left(\sqrt{1{,}5}\times\ 7{,}25/5\right) = 2{,}984$, $\left(ncp_{\psi,L}\ ;\ ncp_{\psi,U}\right) = (0{,}826\ ;\ 5{,}090)$, from which it follows that: $\left(\delta_{\psi,L}\ ;\ \delta_{\psi,U}\right) = (0{,}202\ ;\ 1{,}247)$.

This effect size varies between a small and large effect.

6.4 Comparing independent groups after controlling for a covariate

In an experiment where a group of subjects are randomly assigned to treatment groups, it ensures (to a certain degree) that these groups are similar with respect to the factors which are beyond the control of the experimenter. Suppose that a researcher wishes to compare three methods of teaching mathematics and that students are randomly divided into three groups. One can then assume that, if there are differences between the groups' mathematical proficiency, then it implies that these differences are attributed to the teaching methods and not the unobserved factors like age, IQ or socio-economic status.

Unfortunately, it is not always possible to randomly divide subjects into the groups that we wish to compare as described above. Consider the example provided in Chapter 3's Example F. In this situation the men in the sample are divided into the three activity groups based on a questionnaire concerned with their physical activities. If one wants to compare the means of the total serum cholesterol (S_CHOL), then it is possible that other factors can play a role in making it seem as though these means are not all the same. These uncontrolled factors, like, for example, age, could be highly correlated with cholesterol and as such could influence the cholesterol values. If, for example, the low activity group's age is higher than the other groups, then one might find that high cholesterol values are a result of the age of the individuals and not the fact that they are in the low activity group. It is thus important to control for age. This can be done by making use of *Analysis of Covariance (ANCOVA).*

On top of the assumptions made in an ANOVA, two extra assumptions are included now: (Kline, 2004a: 192):

(a) There is homogeneity of linear regressions of the dependent variable y on the covariate x. This means that each of the populations have a common regression coefficient β.

(b) The covariate, x, is *measured without error*.

An ANCOVA entails that, under the abovementioned assumptions, the dependent variable is corrected for x by using the following expression

$$y' = y - \beta\left(x - \mu_x\right) \quad . \tag{6.61}$$

The method then proceeds by performing a simple ANOVA on the variable y'.

Suppose that in the i-th population the mean of the corrected y_i values (i.e., the mean of the y_i' values) is given by μ_i' with common SD $\sigma_{y.x}$. The variance of the $\mu_i's$ is then

$$\sigma_\mu^2 = \frac{1}{k}\sum_{i=1}^{k}\left(\mu_i' - \mu'\right)^2$$

(similarly to (6.3)), so that the proportion of the total variance attributed to population membership, after controlling for x, is:

$$\eta_{y.x}^2 = \frac{\sigma_{\mu'}^2}{\sigma_{\mu'}^2 + \sigma_{y.x}^2} \tag{6.62}$$

Since $\sigma_{\mu'}^2$ and $\sigma_{y.x}^2$ are the variances of the $\mu_i\,'$ and $y_i\,'$ respectively (similar to the expressions for μ_i and y_i in paragraph 6.1), the expression $\eta_{y.x}^2$ can be treated as an omnibus effect size index for comparing the k population-means after controlling for the covariate x (see Cohen, 1977: 379).

Estimation of $\eta_{y.x}^2$ involves using the same effect size indices based on the sum of squares which were used in ANOVA. Note that there is an additional source of variation which is attributed to the covariate x, i.e., SS_x is defined such that

$$SS_{tot} = SS_G + SS_x + SS_E \ , \tag{6.63}$$

where SS_G and SS_E will both be smaller than the quantities given in paragraph 6.1.1. The approximate index for an omnibus effect is then:

$$\hat{\eta}_{y.x}^2 = \frac{SS_G}{SS_G + SS_E} \ , \tag{6.64}$$

where SS_G and SS_E are called the group and error sums of squares in an ANCOVA respectively.

Example 6.14:

Consider Example F, Chapter 3. Table F.2 shows the results of the ANCOVA with activity group as the grouping variable and age as the covariate, performed for S_CHOL. For age (x) we have that $F = 231{,}04\,(p < 0{,}0001)$, which means it has a highly significant influence on S_CHOL. Further, the activity groups significantly differ with respect to the mean corrected S-CHOL values (y'), because $F = 3{,}1\ (p = 0{,}046)$. We also find that $SS_G = 117750,\ SS_x = 4392610$ and $SS_F = 27168391,$ so that

$$\hat{\eta}^2_{y.x} = \frac{117750}{117750 + 27168391} = 0,0043,$$

which indicates a very small effect.

Note that the corresponding estimator obtained from the ANOVA where we do not control for age, is, from Table F.4:

$$\hat{\eta}^2_{y.x} = \frac{1461712}{33022714} = 0,044,$$

which is 10 times greater than $\hat{\eta}^2_{y.x}$ and suggests a medium effect. It would have been dangerous to judge the differences in the mean S-CHOL values over the activity groups as having a medium effect. When we control for age these differences disappear. Also note that in ANOVA the values of both SS_G and SS_F increase. □

6.4.1 Contrasts in analysis of covariance

After one controls for the covariate x, the contrasts have the form:

$$\psi' = c_1\mu_1' + c_2\mu_2' + ... + c_k\mu_k'$$

which can be estimated by

$$\hat{\psi}' = c_1\overline{y}_1' + c_2\overline{y}_2' + ... + c_k\overline{y}_k', \quad (6.65)$$

where

$$\overline{y}_i' = \overline{y}_i - b\left(\overline{x}_i - \overline{x}\right),\ i = 1,...,k \quad (6.66)$$

where b is the common estimated regression coefficient for all the groups' linear regressions of y on x ; and $\overline{x}$ is the mean of x over all the groups.

From the computer output of an ANCOVA, one can obtain $\overline{y}_i'$, the adjusted means or also least squares means ('LS means'), although it is usually not necessary to calculate $\overline{y}_i'$ using (6.66).

The effect size index is then

$$\delta_{\psi'} = \frac{\psi'}{\sigma^*} \quad ,$$

where σ^* is chosen as before (see paragraph 6.3.1), i.e., using the SD's of the populations's adjusted y' -values.

When estimating $\delta_{\psi'}$ one can, according to Olejnik & Algina (2000: 254), standardize the contrast $\hat{\psi}'$ as before, but then $\hat{\sigma}^*$ must be based on a SD of the sample's adjusted values. When we assume that the SD's of the adjusted values are equal for each group, then the easiest choice is $\hat{\sigma}^* = \sqrt{meanSS_E}$, where $meanSS_E$ is the mean error-sum of squares of the ANCOVA. Therefore, an estimator of the effect size index is:

$$\hat{\delta}_{\psi'} = \hat{\psi}' / \sqrt{meanSS_E} \quad . \qquad (6.67)$$

Example 6.15:
Consider Example F of Chapter 3. Suppose that we want to compare the three activity groups in a pairwise fashion, the contrasts are then the comparisons $\psi_1' = \mu_1' - \mu_2'$; $\psi_2' = \mu_1' - \mu_3'$ and $\psi_3' = \mu_2' - \mu_3'$ after we have controlled for age. From Table F.3 we obtain the adjusted $\bar{y}_i'$ values while the mean error sum of squares 19012 is obtained from Table F.2. The effect size indices are then:

$$\hat{\delta}_{\psi_1'} = \frac{528 - 531,6}{\sqrt{19012}} = -0,026$$

$$\hat{\delta}_{\psi_2'} = \frac{528 - 509}{\sqrt{19012}} = 0,138$$

$$\hat{\delta}_{\psi_3'} = \frac{531,6 - 509}{\sqrt{19012}} = 0,164 \, .$$

These all show a small effect which means that, while they are statistically significant, the differences between the adjusted means are not important differences. Even the contrast $\psi_2 = \mu_1 - \mu_3$ which was not controlled for age has the following effect size index value (obtained from Tables F.1 and F.4):

$$\hat{\delta}_{\psi_2} = \frac{552,5 - 482,0}{\sqrt{22071}} = 0,475 ,$$

which is a medium effect, even though the F-test was highly significant. □

6.4.2 Confidence intervals of effect size indices after controlling for a covariate

Since the values of $y_i\,'$ after controlling for a covariate can be used to determine omnibus and contrast effects to compare independent groups (as in paragraphs 6.1 and 6.3), one can also create $(1\text{-}\alpha)100\%$ *CI*'s as was done in 6.1.2 and 6.3.4. The approximate and exact *CI*'s for $\eta^2_{y.x}$ are determined according to equations (6.13), (6.14) and the SAS-program **VI_R2**, but we now use the variance ratio of the ANCOVA with $k-1$ and $n-k-1$ degrees of freedom.

For ψ' the $(1\text{-}\alpha)100\%$ can be determined using equations (6.45) and (6.46) where $s^2_{p,\psi} = meanSS_E$ from the ANCOVA with $n-k-1$ degrees of freedom instead of $n_m - m$. Subsequently, a *CI* for $\delta_{\psi'}$ can be obtained from equation (6.53) with $\hat{\sigma}^* = \sqrt{meanSS_E}$.

The steps in paragraph 6.3.4 are followed to obtain the exact *CI* for calculating $\delta_{\psi'}$ ‘s, where

$$t_\psi\,' = \frac{\hat{\psi}\,'}{s_{\hat{\psi}}\,'} \quad ,$$

with $s_{\hat{\psi}}\,'$ as in (6.31) with $s^2_{p,\psi} = meanSS_E$ from the ANCOVA and $n-k-1$ degrees of freedom.

Note that, when determining the above *CI*'s, we make the assumption of homogeneity of variances (i.e., $\sigma^* = \sigma$).

Example 6.16:

In Example 6.14 is $\eta^2_{y.x}$ estimated by $\hat{\eta}^2_{y.x} = 0,0043$. Use (6.12) and Table F.2, then it follows that

$$F = 3,1 \;\; ; \;\; w = 2\times 1429 - 1 = 2857 \;\;,\;\; x = 2\times 3,1/1429 = 0,0043$$

$$\mathrm{C} = \left(2 + 2\times 1433\times 0,0043\right) / \left(2 + 1433\times 0,0043\right) = 1,755 .$$

A 95% *CI* for $ncp_F = n\eta^2_{y.x} / \left(1 - \eta^2_{y.x}\right)$ by making use of (6.12) is:

$$ncp_{FL} = \frac{1}{2}\left[2857\times 0,0043 + 1,96^2\left(0,0043 + 1,755\right) - 4 + 1,755\right]$$
$$-1,96\sqrt{2857\times 0,0043\left(0,0043 + 1,755\right)}$$
$$= 8,399 - 9,112 = -0,713 ,$$
$$nsp_{FU} = 8,399 + 9,112 = 17,511$$

However, since ncp_{FL} is negative, we take $ncp_{FU} = 0$, so that, using (6.13):

$$\eta^2_{y.x,L,ben} = 0$$
$$\eta^2_{y.x,U,ben} = 17,511/\left(17,511 + 1433\right) = 0,0121 .$$

The 95% *CI* for $\eta^2_{y.x}$ is thus $\left(0;0,012\right)$, which is confirmed by the exact *CI*'s upper bound $0,01274$, but the value of the lower bound can not be calculated. The exact *CI* for the effect size index, when one does not control for age, is $\left(0,025\,;\,0,066\right)$, which is a small to medium effect. The controlled case, however, shows only a small effect.

The contrast $\psi_2{}' = \mu_1{}' - \mu_3{}'$ in Example 6.15 is estimated by $\hat{\psi}_2{}' = 19$ and

$$S_{\hat{\psi}}{}' = \sqrt{\left(\frac{1^2}{728} + \frac{(-1)^2}{468}\right)19012} = 8,169 \text{, thus}$$

$$t_{\psi_2}{}' = 19/8,169 = 2,326 .$$

A 95% *CI* for ψ_2' is: $19 \pm 1{,}96 \times 8{,}169 = 19 \pm 16{,}01$

$$= (2{,}99\ ;\ 35{,}01),$$

so that an approximate 95% *CI* for δ_{ψ_2}' is :

$$\left(\frac{2{,}99}{\sqrt{19012}}\ ;\ \frac{35{,}01}{\sqrt{19012}} \right) = (0{,}022\,; 0{,}254)\ .$$

By using the program **VI_delta_contrast**, we find practically the same interval.

□

6.4.3 More than one covariate

Suppose that now we control for ℓ covariates, so that, similarly to (6.66), we have:

$$y' = y - \beta_1(x_1 - \mu_{x_1}) - \beta_2(x_2 - \mu_{x_2}) - ... - \beta_\ell(x_\ell - \mu_{x_\ell}) \qquad (6.68)$$

where y' are the adjusted y values. Similarly to (6.63) we obtain

$$SS_{tot} = SS_G + SS_{x_1} + SS_{x_2} + ... + SS_{x_\ell} + SS_E\ , \qquad (6.69)$$

so that SS_G and SS_E are the sums of squares which remain after $x_1,...,x_\ell$ are removed. The omnibus effect and its estimator in (6.62) and (6.64) thus remain unchanged, but now σ_μ^2 and $\sigma_{y.x}^2$ have a new interpretation and SS_G and SS_E are obtained from an ANCOVA with covariates $x_1, x_2,...,x_\ell$.

When Calculating *CI*'s, the degrees of freedom of $s_{p,\psi'}^2$ are $n-k-1$ and still calculated using $meanSS_E$.