Chapter 14

Within-Subjects Designs

ANOVA must be modified to take correlated errors into account when multiple measurements are made for each subject.

14.1 Overview of within-subjects designs

Any categorical explanatory variable for which each subject experiences all of the levels is called a **within-subjects factor**. (Or sometimes a subject may experience several, but not all levels.) These levels could be different "treatments", or they may be different measurements for the same treatment (e.g., height and weight as outcomes for each subject), or they may be repetitions of the same outcome over time (or space) for each subject. In the broad sense, the term **repeated measure** is a synonym for a within-subject factor, although often the term repeated measures analysis is used in a narrower sense to indicate the specific set of analyses discussed in Section 14.5.

In contrast to a within-subjects factor, any factor for which each subject experiences only one of the levels is a **between-subjects factor**. Any experiment that has at least one within-subjects factor is said to use a **within-subjects design**, while an experiment that uses only between-subjects factor(s) is called a **between-subjects design**. Often the term **mixed design** or **mixed within- and between-subjects design** is used when there is at least one within-subjects factor and at least one between-subjects factor in the same experiment. (Be careful to distinguish this from the so-called mixed models of chapter 15.) All of the

experiments discussed in the preceding chapters are between-subjects designs.

Please do not confuse the terms between-groups and within-groups with the terms between-subjects and within-subjects. The first two terms, which we first encountered in the ANOVA chapter, are names of specific SS and MS components and are named because of how we define the deviations that are summed and squared to compute SS. In contrast, the terms between-subjects and within-subjects refer to experimental designs that either do not or do make multiple measurements on each subject.

When a within-subjects factor is used in an experiment, new methods are needed that do not make the assumption of no correlation (or, somewhat more strongly, independence) of errors for the multiple measurements made on the same subject. (See section 6.2.8 to review the independent errors assumption.)

Why would we want to make multiple measurements on the same subjects? There are two basic reasons. First, our primary interest may be to study the change of an outcome over time, e.g., a learning effect. Second, studying multiple outcomes for each subject allows each subject to be his or her own "control", i.e., we can effectively remove subject-to-subject variation from our investigation of the relative effects of different treatments. This reduced variability directly increases power, often dramatically. We may use this increased power directly, or we may use it indirectly to allow a reduction in the number of subjects studied.

These are very important advantages to using within-subjects designs, and such designs are widely used. The major reasons for not using within-subjects designs are when it is impossible to give multiple treatments to a single subject or because of concern about confounding. An example of a case where a within-subjects design is impossible is a study of surgery vs. drug treatment for a disease; subjects generally would receive one or the other treatment, not both.

The confounding problem of within-subjects designs is an important concern. Consider the case of three kinds of hints for solving a logic problem. Let's take the time till solution as the outcome measure. If each subject first sees problem 1 with hint 1, then problem 2 with hint 2, then problem 3 with hint 3, then we will probably have two major difficulties. First, the effects of the hints **carry-over** from each trial to the next. The truth is that problem 2 is solved when the subject has been exposed to two hints, and problem 3 when the subject has been exposed to all three hints. The effect of hint type (the main focus of inference) is *confounded* with the cumulative effects of prior hints.

The carry-over effect is generally dealt with by allowing sufficient time between trials to "wash out" the effects of previous trials. That is often quite effective, e.g., when the treatments are drugs, and we can wait until the previous drug leaves the system before studying the next drug. But in cases such as the hint study, this approach may not be effective or may take too much time.

The other, partially overlapping, source of confounding is the fact that when testing hint 2, the subject has already had practice with problem 1, and when testing hint three she has already had practice with problems 1 and 2. This is the **learning effect**.

The learning effect can be dealt with effectively by using **counterbalancing**. The carryover effect is also partially corrected by counterbalancing. Counterbalancing in this experiment could take the form of collecting subjects in groups of six, then randomizing the group to all possible orderings of the hints (123, 132, 213, 231, 312, 321). Then, because each hint is evenly tested at all points along the learning curve, any learning effects would "balance out" across the three hint types, removing the confounding. (It would probably also be a good idea to randomize the order of the problem presentation in this study.)

You need to know how to distinguish within-subjects from between-subjects factors. Within-subjects designs have the advantages of more power and allow observation of change over time. The main disadvantage is possible confounding, which can often be overcome by using counterbalancing.

14.2 Multivariate distributions

Some of the analyses in this chapter require you to think about **multivariate distributions**. Up to this point, we have dealt with outcomes that, among all subjects that have the same given combination of explanatory variables, are assumed to follow the (univariate) Normal distribution. The mean and variance, along with the standard bell-shape characterize the kinds of outcome values that we expect to see. Switching from the population to the sample, we can put the value of the outcome on the x-axis of a plot and the relative frequency of that

value on the y-axis to get a histogram that shows which values are most likely and from which we can visualize how likely a range of values is.

To represent the outcomes of two treatments for each subject, we need a so-called, bivariate distribution. To produce a graphical representation of a bivariate distribution, we use the two axes (say, y1 and y2) on a sheet of paper for the two different outcome values, and therefore each pair of outcomes corresponds to a point on the paper with y1 equal to the first outcome and y2 equal to the second outcome. Then the third dimension (coming up out of the paper) represents how likely each combination of outcome is. For a bivariate Normal distribution, this is like a real bell sitting on the paper (rather than the silhouette of a bell that we have been using so far).

Using an analogy between a bivariate distribution and a mountain peak, we can represent a bivariate distribution in 2-dimensions using a figure corresponding to a topographic map. Figure 14.1 shows the center and the contours of one particular bivariate Normal distribution. This distribution has a negative correlation between the two values for each subject, so the distribution is more like a bell squished along a diagonal line from the upper left to the lower right. If we have no correlation between the two values for each subject, we get a nice round bell. You can see that an outcome like $Y_1 = 2,\ Y_2 = 6$ is fairly likely, while one like $Y_1 = 6,\ Y_2 = 2$ is quite unlikely. (By the way, bivariate distributions can have shapes other than Normal.)

The idea of the bivariate distribution can easily be extended to more than two dimensions, but is of course much harder to visualize. A multivariate distribution with k-dimensions has a k-length vector (ordered set of numbers) representing its mean. It also has a $k \times k$ dimensional matrix (rectangular array of numbers) representing the variances of the individual variables, and all of the paired covariances (see section 3.6.1).

For example a 3-dimensional multivariate distribution representing the outcomes of three treatments in a within-subjects experiment would be characterized by a mean vector, e.g.,

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix},$$

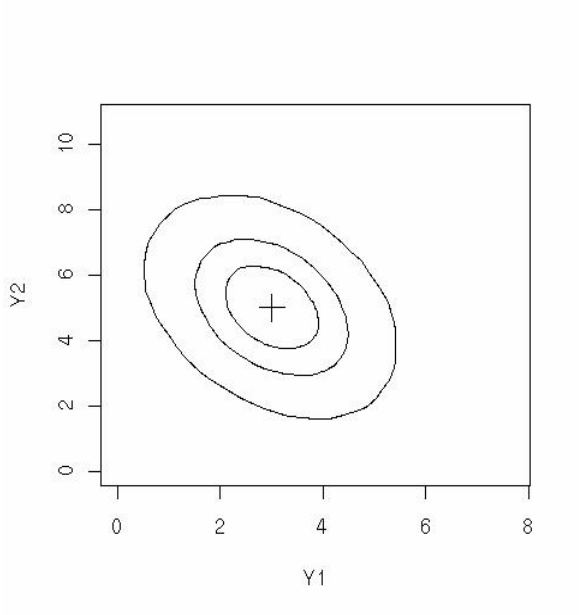

Figure 14.1: Contours enclosing 1/3, 2/3 and 95% of a bivariate Normal distribution with a negative covariance.

and a variance-covariance matrix, e.g.,

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \gamma_{1,2} & \gamma_{1,3} \\ \gamma_{1,2} & \sigma_2^2 & \gamma_{2,3} \\ \gamma_{1,3} & \gamma_{2,3} & \sigma_3^2 \end{bmatrix}.$$

Here we are using $\gamma_{i,j}$ to represent the covariance of variable Y_i with Y_j.

Sometimes, as an alternative to a variance-covariance matrix, people use a variance vector, e.g.,

$$\sigma^2 = \begin{bmatrix} \sigma_1^2 \\ \sigma_2^2 \\ \sigma_3^2 \end{bmatrix},$$

and a correlation matrix, e.g.,

$$\text{Corr} = \begin{bmatrix} 1 & \rho_{1,2} & \rho_{1,3} \\ \rho_{1,2} & 1 & \rho_{2,3} \\ \rho_{1,3} & \rho_{2,3} & 1 \end{bmatrix}.$$

Here we are using $\rho_{i,j}$ to represent the correlation of variable Y_i with Y_j.

If the distribution is also Normal, we could write the distribution as $Y \sim N(\mu, \Sigma)$.

14.3 Example and alternate approaches

Consider an example related to the disease osteoarthritis. (This comes from the OzDASL web site, OzDASL. For educational purposes, I slightly altered the data, which can be found in both the tall and wide formats on the data web page of this book: osteoTall.sav and osteoWide.sav.) Osteoarthritis is a mechanical degeneration of joint surfaces causing pain, swelling and loss of joint function in one or more joints. Physiotherapists treat the affected joints to increase the range of movement (ROM). In this study 10 subjects were each given a trial of therapy with two treatments, TENS (an electric nerve stimulation) and short wave diathermy (a heat treatment), plus control.

We cannot perform ordinary (between-subjects) one-way ANOVA for this experiment because each subject was exposed to all three treatments, so the errors (ROM outcomes for a given subject for all three treatments minus the population means of outcome for those treatment) are almost surely correlated, rather than independent. Possible appropriate analyses fall into four categories.

1. Response simplification: e.g. call the difference of two of the measurements on each subject the response, and use standard techniques. If the within-subjects factor is the only factor, an appropriate test is a one-sample t-test for the difference outcome, with the null hypothesis being a zero mean difference. In cases where the within-subjects factor is repetition of the same measurement over time or space and there is a second, between subjects-factor, the effects of the between subjects factor on the outcome can be studied by taking the mean of all of the outcomes for each subject and using standard, between-subjects one-way ANOVA. This approach does not fully utilize the available information. Often it cannot answer some interesting questions.

2. Treat the several responses on one subject as a single "multivariate" response and model the correlation between the components of that response. The main statistics are now matrices rather than individual numbers. This

approach corresponds to results labeled "multivariate" under "repeated measures ANOVA" for most statistical packages.

3. Treat each response as a separate (univariate) observation, and treat "subject" as a (random) blocking factor. This corresponds to within-subjects ANOVA with subject included as a random factor and with no interaction in the model. It also corresponds to the "univariate" output under "repeated measures". In this form, there are assumptions about the nature of the within-subject correlation that are not met fairly frequently. To use the univariate approach when its assumptions are not met, it is common to use some approximate correction (to the degrees of freedom) to compensate for a shifted null sampling distribution.

4. Treat each measurement as univariate, but explicitly model the correlations. This is a more modern univariate approach called "mixed models" that subsumes a variety of models in a single unified approach, is very flexible in modeling correlations, and often has improved interpretability. As opposed to "classical repeated measures analysis" (approaches 2 and 3), mixed models can accommodate missing data as oppposed to dropping all data from every subject who is missing one or more measurements), and it accommodates unequal and/or irregular spacing of repeated measurements. Mixed models can also be extended to non-normal outcomes. (See chapter 15.)

14.4 Paired t-test

The paired t-test uses response simplification to handle the correlated errors. It only works with two treatments, so we will ignore the diathermy treatment in our osteoarthritis example for this section. The simplification here is to compute the difference between the two outcomes for each subject. Then there is only one "outcome" for each subject, and there is no longer any concern about correlated errors. (The subtraction is part of the paired t-test, so you don't need to do it yourself.)

In SPSS, the paired t-test requires the "wide" form of data in the spreadsheet rather than the "tall" form we have used up until now. The tall form has one outcome per row, so it has many rows. The wide form has one subject per row with two or more outcomes per row (necessitating two or more outcome columns).

The paired t-test uses a one-sample t-test on the single column of computed differences. Although we have not yet discussed the one-sample t-test, it is a straightforward extension of other t-tests like the independent-sample t-test of Chapter 6 or the one for regression coefficients in Chapter 9. We have an estimate of the difference in outcome between the two treatments in the form of the mean of the difference column. We can compute the standard error for that difference (which is the square root of the variance of the difference column divided by the number of subjects). Then we can construct the t-statistic as the estimate divided by the SE of the estimate, and under the null hypothesis that the population mean difference is zero, this will follow a t-distribution with $n - 1$ df, where n is the number of subjects.

The results from SPSS for comparing control to TENS ROM is shown in table 14.1. The table tells us that the best point estimate of the difference in population means for ROM between control and TENS is 17.70 with control being higher (because the direction of the subtraction is listed as control minus TENS). The uncertainty in this estimate due to random sampling variation is 7.256 on the standard deviation scale. (This was calculated based on the sample size of 10 and the observed standard deviation of 22.945 for the observed sample.) We are 95% confident that the true reduction in ROM caused by TENS relative to the control is between 1.3 and 34.1, so it may be very small or rather large. The t-statistic of 2.439 will follow the t-distribution with 9 df if the null hypothesis is true and the assumptions are met. This leads to a p-value of 0.037, so we reject the null hypothesis and conclude that TENS reduces range of motion.

For comparison, the incorrect, between-subjects one-way ANOVA analysis of these data gives a p-value of 0.123, leading to the (probably) incorrect conclusion that the two treatments both have the same population mean of ROM. For future discussion we note that the within-groups SS for this incorrect analysis is 10748.5 with 18 df.

For educational purposes, it is worth noting that it is possible to get the same correct results in this case (or other one-factor within-subjects experiments) by performing a two-way ANOVA in which "subject" is the other factor (besides treatment). Before looking at the results we need to note several important facts.

Paired Differences							
			95% Confidence Interval of the Difference				
Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	Sig. (2-tailed)
17.700	22.945	7.256	1.286	34.114	2.439	9	0.037

Table 14.1: Paired t-test for control-TENS ROM in the osteoarthritis experiment.

There is an important concept relating to the repeatability of levels of a factor. A factor is said to be a **fixed factor** if the levels used are the same levels you would use if you repeated the experiment. Treatments are generally fixed factors. A factor is said to be a **random factor** if a different set of levels would be used if you repeated the experiment. Subject is a random factor because if you would repeat the experiment, you would use a different set of subjects. Certain types of blocking factors are also random factors.

The reason that we want to use subject as a factor is that it is reasonable to consider that some subjects will have a high outcome for all treatments and others a low outcome for all treatments. Then it may be true that the errors relative to the overall subject mean are uncorrelated across the k treatments given to a single subject. But if we use both treatment and subject as factors, then each combination of treatment and subject has only one outcome. In this case, we have zero degrees of freedom for the within-subjects (error) SS. The usual solution is to use the interaction MS in place of the error MS in forming the F test for the treatment effect. (In SPSS it is equivalent to fit a model without an interaction.) Based on the formula for expected MS of an interaction (see section 12.4), we can see that the interaction MS is equal to the error MS if there is no interaction and larger otherwise. Therefore if the assumption of no interaction is correct (i.e,. treatment effects are similar for all subjects) then we get the "correct" p-value, and if there really is an interaction, we get too small of an F value (too large of a p-value), so the test is conservative, which means that it may give excess Type 2 errors, but won't give excess Type 1 errors.

The two-way ANOVA results are shown in table 14.2. Although we normally ignore the intercept, it is included here to demonstrate the idea that in within-subjects ANOVA (and other cases called nested ANOVA) the denominator of the F-statistic, which is labeled "error", can be different for different numerators (which

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	Hypothesis	173166.05	1	173166.05	185.99	<0.0005
	Error	8379.45	9	931.05		
rx	Hypothesis	1566.45	1	1566.45	5.951	0.035
	Error	2369.05	9	263.23		
subject	Hypothesis	8379.45	9	931.05	3.537	0.037
	Error	2369.05	9	263.23		

Table 14.2: Two-way ANOVA results for the osteoarthritis experiment.

correspond to the different null hypotheses). The null hypothesis of main interest here is that the three treatment population means are equal, and that is tested and rejected on the line called "rx". The null hypothesis for the random subject effect is that the population variance of the subject-to-subject means (of all three treatments) is zero.

The key observation from this table is that the treatment (rx) SS and MS corresponds to the between-groups SS and MS in the incorrect one-way ANOVA, while the sum of the subject SS and error SS is 10748.5, which is the within-groups SS for the incorrect one-way ANOVA. This is a decomposition of the four sources of error (see Section 8.5) that contribute to σ^2, which is estimated by SS_{within} in the one-way ANOVA. In this two-way ANOVA the subject-to-subject variability is estimated to be 931.05, and the remaining three sources contribute 263.23 (on the variance scale). This smaller three-source error MS is the denominator for the numerator (rx) MS for the F-statistic of the treatment effect. Therefore we get a larger F-statistic and more power when we use a within-subjects design.

How do we know which error terms to use for which F-tests? That requires more mathematical statistics than we cover in this course, but SPSS will produce an EMS table, and it is easy to use that table to figure out which ratios are 1.0 when the null hypotheses are true.

It is worth mentioning that in SPSS a one-way within-subjects ANOVA can be analyzed either as a two-way ANOVA with subjects as a random factor (or even as a fixed factor if a no-interaction model is selected) or as a repeated measures analysis (see next section). The p-value for the overall null hypothesis, that the population outcome means are equal for all levels of the factor, is the same for

each analysis, although which auxiliary statistics are produced differs.

A two-level one-way within-subjects experiment can equivalently be analyzed by a paired t-test or a two-way ANOVA with a random subject factor. The latter also applies to more than two levels. The extra power comes from mathematically removing the subject-to-subject component of the underlying variance (σ^2).

14.5 One-way Repeated Measures Analysis

Although repeated measures analysis is a very general term for any study in which multiple measurements are made on the same subject, there is a narrow sense of repeated measures analysis which is discussed in this section and the next section. This is a set of specific analysis methods commonly used in social sciences, but less commonly in other fields where alternatives such as mixed models tends to be used.

This narrow-sense repeated measures analysis is what you get if you choose "General Linear Model / Repeated Measures" in SPSS. It includes the second and third approaches of our list of approaches given in the introduction to this chapter. The various sections of the output are labeled univariate or multivariate to distinguish which type of analysis is shown.

This section discusses the k-level ($k \geq 2$) one-way within-subjects ANOVA using repeated measures in the narrow sense. The next section discusses the mixed within/between subjects two-way ANOVA.

First we need to look at the assumptions of repeated measures analysis. One-way repeated measures analyses assume a Normal distribution of the outcome for each level of the within-subjects factor. The errors are assumed to be uncorrelated between subjects. Within a subject the multiple measurements are assumed to be correlated. For the univariate analyses, the assumption is that a technical condition called **sphericity** is met. Although the technical condition is difficult to understand, there is a simpler condition that is nearly equivalent: compound symmetry. **Compound symmetry** indicates that all of the variances are equal and all of the covariances (and correlations) are equal. This variance-covariance

pattern is seen fairly often when there are several different treatments, but is unlikely when there are multiple measurements over time, in which case adjacent times are usually more highly correlated than distant times.

In contrast, the multivariate portions of repeated measures analysis output are based on an unconstrained variance-covariance pattern. Essentially, all of the variances and covariances are estimated from the data, which allows accommodation of a wider variety of variance-covariance structures, but loses some power, particularly when the sample size is small, due to "using up" some of the data and degrees of freedom for estimating a more complex variance-covariance structure.

Because the univariate analysis requires the assumption of sphericity, it is customary to first examine the Mauchly's test of sphericity. Like other tests of assumptions (e.g., Levene's test of equal variance), the null hypothesis is that there is no assumption violation (here, that the variance-covariance structure is consistent with sphericity), so a large (>0.05) p-value is good, indicating no problem with the assumption. Unfortunately, the sphericity test is not very reliable, being often of low power and also overly sensitive to mild violations of the Normality assumption. It is worth knowing that the sphericity assumption cannot be violated with $k = 2$ levels of treatment (because there is only a single covariance between the two measures, so there is nothing for it to be possible unequal to), and therefore Mauchly's test is inapplicable and not calculated when there are only two levels of treatment.

The basic overall univariate test of equality of population means for the within-subjects factor is labeled "Tests of Within-Subjects Effects" in SPSS and is shown in table 14.3. If we accept the sphericity assumption, e.g., because the test of sphericity is non-significant, then we use the first line of the treatment section and the first line of the error section. In this case F=MS_{between} divided by MS_{within}=1080.9/272.4=3.97. The p-value is based on the F-distribution with 2 and 18 df. (This F and p-value are exactly the same as the two-way ANOVA with subject as a random factor.)

If the sphericity assumption is violated, then one of the other, corrected lines of the Tests of Within-Subjects Effects table is used. There is some controversy about when to use which correction, but generally it is safe to go with the Huynh-Feldt correction.

The alternative, multivariate analysis, labeled "Multivariate Tests" in SPSS is shown in table 14.4. The multivariate tests are tests of the same overall null hypothesis (that all of the treatment population means are equal) as was used for

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
rx	Sphericity Assumed	2161.8	2	1080.9	3.967	.037
	Greenhouse-Geisser	2161.8	1.848	1169.7	3.967	.042
	Huynh-Feldt	2161.8	2.000	1080.9	3.967	.042
	Lower-bound	2161.8	1.000	1169.7	3.967	.042
Error(rx)	Sphericity Assumed	4904.2	18	272.4		
	Greenhouse-Geisser	4904.2	16.633	294.8		
	Huynh-Feldt	4904.2	18,000	272.4		
	Lower-bound	4904.2	9.000	544.9		

Table 14.3: Tests of Within-Subjects Effects for the osteoarthritis experiment.

the univariate analysis.

The approach for the multivariate analysis is to first construct a set of $k-1$ orthogonal contrasts. (The main effect and interaction p-values are the same for every set of orthogonal contrasts.) Then SS are computed for each contrast in the usual way, and also "sum of cross-products" are also formed for pairs of contrasts. These numbers are put into a $k-1$ by $k-1$ matrix called the SSCP (sums of squares and cross products) matrix. In addition to the (within-subjects) treatment SSCP matrix, an error SSCP matrix is constructed analogous to computation of error SS. The ratio of these matrices is a matrix with F-values on the diagonal and ratios of treatment to error cross-products off the diagonal. We need to make a single F statistic from this matrix to get a p-value to test the overall null hypothesis. Four methods are provided for reducing the ratio matrix to a single F value. These are called Pillai's Trace, Wilk's Lambda, Hotelling's Trace, and Roy's Largest Root. There is a fairly extensive, difficult-to-understand literature comparing these methods, but it most cases they give similar p-values.

The decision to reject or retain the overall null hypothesis of equal population outcome means for all levels of the within-subjects factor is made by looking at

Effect		Value	F	Hypothesis df	Error df	Sig.
modality	Pillai's Trace	0.549	4.878	2	8	0.041
	Wilk's Lambda	0.451	4.878	2	8	0.041
	Hotelling's Trace	1.220	4.878	2	8	0.041
	Roy's Largest Root	1.220	4.878	2	8	0.041

Table 14.4: Multivariate Tests for the osteoarthritis experiment.

the p-value for one of the four F-values computed by SPSS. I recommend that you use "Pillai's trace". The thing you should not do is pick the line that gives the answer you want! In a one-way within-subjects ANOVA, the four F-values will always agree, while in more complex designs they will disagree to some extent.

Which approach should we use, univariate or multivariate? Luckily, they agree most of the time. When they disagree, it could be because the univariate approach is somewhat more powerful, particularly for small studies, and is thus preferred. Or it could be that the correction is insufficient in the case of far deviation from sphericity, in which case the multivariate test is preferred as more robust. In general, you should at least look for outliers or mistakes if there is a disagreement.

An additional section of the repeated measures analysis shows the planned contrasts and is labeled "Tests of Within-Subjects Contrasts". This section is the same for both the univariate and multivariate approaches. It gives a p-value for each planned contrast. The default contrast set is "polynomial" which is generally only appropriate for a moderately large number of levels of a factor representing repeated measures of the same measurement over time. In most circumstances, you will want to change the contrast type to simple (baseline against each other level) or repeated (comparing adjacent levels).

It is worth noting that post-hoc comparisons are available for the within-subjects factor under Options by selecting the factor in the Estimated Marginal Means box and then by checking the "compare main effects" box and choosing Bonferroni as the method.

14.6 Mixed between/within-subjects designs

One of the most common designs used in psychology experiments is a two-factor ANOVA, where one factor is varied between subjects and the other within subjects. The analysis of this type of experiment is a straightforward combination of the analysis of two-way between subjects ANOVA and the concepts of within-subject analysis from the previous section.

The interaction between a within- and a between-subjects factor shows up in the within-subjects section of the repeated measures analysis. As usual, the interaction should be examined first. If the interaction is significant, then (changes in) both factors affect the outcome, regardless of the p-values for the main effects. Simple effects contrasts in a mixed design are not straightforward, and are not available in SPSS. A profile plot is a good summary of the results. Alternatively, it is common to run separate one-way ANOVA analyses for each level of one factor, possibly using planned and/or post-hoc testing. In this case we test the simple effects hypotheses about the effects of differences in level of one factor at fixed levels of the other factor, as is appropriate in the case of interaction. Note that, depending on which factor is restricted to a single level for these analyses, the appropriate ANOVA could be either within-subjects or between-subjects.

If an interaction is not significant, we usually remove it from the model, but that is not possible in repeated measures ANOVA. You should ignore it and interpret both "main effects" overall null hypothesis as equal means for all levels of one factor averaging over (or ignoring) the other factor. For the within-subjects factor, either the univariate or multivariate tests can be used.

There is a separate results section for the overall null hypothesis for the between subjects factor. Because this section compares means between levels of the between-subjects factor, and those means are reductions of the various levels of the within-subjects factor to a single number, there is no concern about correlated errors, and there is only a single univariate test of the overall null hypothesis.

For each factor you may select a set of planned contrasts (assuming that there are more than two levels and that the overall null hypothesis is rejected). Finally, post-hoc tests are available for the between-subjects factor, and either the Tukey or Dunnett test is usually appropriate (where Dunnett is used only if there is no interest in comparisons other than to the control level). For the within-subjects factor the Bonferroni test is available with Estimated Marginal Means.

Repeated measures analysis is appropriate when one (or more) factors is a within-subjects factor. Usually univariate and multivariate tests agree for the overall null hypothesis for the within-subjects factor or any interaction involving a within-subjects factor. Planned (main effects) contrasts are appropriate for both factors if there is no significant interaction. Post-hoc comparisons can also be performed.

14.6.1 Repeated Measures in SPSS

To perform a repeated measures analysis in SPSS, use the menu item "Analyze / General Linear Model / Repeated Measures." The example uses the data in circleWide.sav. This is in the "wide" format with a separate column for each level of the repeated factor.

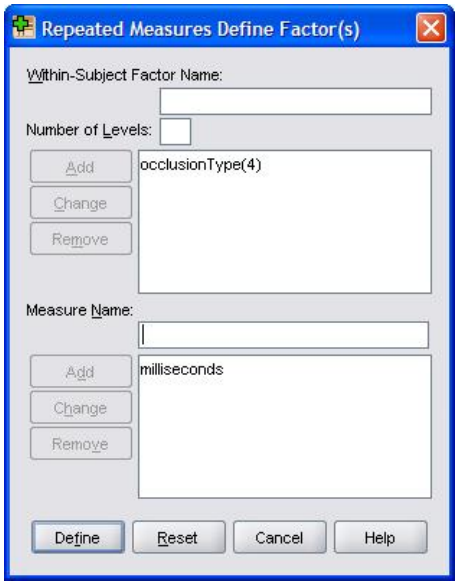

Figure 14.2: SPSS Repeated Measures Define Factor(s) dialog box.

Unlike other analyses in SPSS, there is a dialog box that you must fill out before seeing the main analysis dialog box. This is called the "Repeated Measures Define Factor(s)" dialog box as shown in Figure 14.2. Under "Within-Subject Factor

Name" you should enter a (new) name that describes what is different among the levels of your within-subjects factor. Then enter the "Number of Levels", and click Add. In a more complex design you need to do this for each within-subject factor. Then, although not required, it is a very good idea to enter a "Measure Name", which should describe what is measured at each level of the within-subject factor. Either a term like "time" or units like "milliseconds" is appropriate for this box. Click the "Define" button to continue.

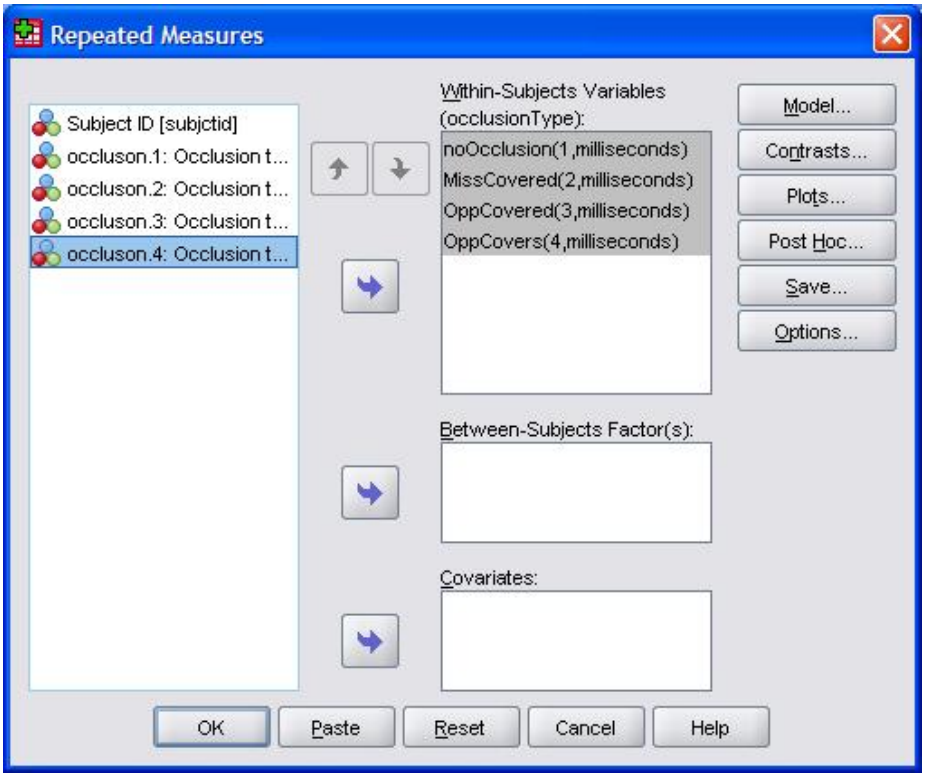

Figure 14.3: SPSS Repeated Measures dialog box.

Next you will see the Repeated Measures dialog box. On the left is a list of all variables, at top right right is the "Within-Subjects Variables" box with lines for each of the levels of the within-subjects variables you defined previously. You should move the k outcome variables corresponding to the k levels of the within-subjects factor into the "Within-Subjects Variables" box, either one at a time or all together. The result looks something like Figure 14.3. Now enter the between-subjects factor, if any. Then use the model button to remove the interaction if

desired, for a two-way ANOVA. Usually you will want to use the contrasts button to change the within-subjects contrast type from the default "polynomial" type to either "repeated" or "simple". If you want to do post-hoc testing for the between-subjects factor, use the Post-Hoc button. Usually you will want to use the options button to display means for the levels of the factor(s). Finally click OK to get your results.