Chapter 13

GLM: Multiple dependent variables

13.1 Introduction

Gene expression is a major interest in neuroscience. Suppose that research group interested in the expression of a gene assigns 10 rats to a control (i.e., vehicle) condition and 10 to a treatment condition that administers a substance hypothesized to influence that gene’s transcription. The group assays for the amount of the gene-specific mRNA in two different sections of the brain. A common method of analysis would be to compute one t test for the Control and Treatment groups for the mRNA assay in three different brain regions. A common method of presenting such data would be a bar chart of means with error bars. Such a graph is depicted in Figure 13.1 for simulated data. The difference between means for Area 1 is not significant ($t(17.8) = 1.44, p = 0.17$). Neither is the different for Area 2 ($t(16.6) = -1.40, p = 0.18$) nor the difference between Treatments and Controls for Area 3 ($t(17.7) = 1.60, p = 0.13$). Would you conclude that the treatment substance has no effect on gene expression in these two regions?

Take these data to your favorite biostatistician and ask the same question. The answer will be, “Dunno.”

The major statistical issue in these data is that the same Control rat assayed for brain area 1 was also assayed for brain areas 2 and 3. Same for any Treatment rat. The structure of the data set is depicted in Table 13.1 for the first six rats in the data set. In Figure 13.1 there are three dependent variables per rat–Area1, Area2, and Area3. When more than one dependent variable is analyzed, then it is necessary to ask, “how highly are the dependent variables correlated?” In the present case, the correlation between mRNA values in the three area is around 0.50. This is sufficiently greater than 0 that an appropriate statistical analysis must take those correlations into account.

This chapter deals with the analysis of multiple correlated dependent vari-
This chapter will speak of the multivariate analysis of variance (MANOVA). This should really be called the multivariate general linear model, but the term MANOVA has become so entrenched that it is silly to try to alter the terminology now. Although MANOVA is discussed, you will not learn the logic behind the procedure. MANOVA is a part of multivariate statistics and requires statistical knowledge that is too advanced for this text.

Instead, we concentrate on the repeated measures design. This occurs when there are several dependent variables, all of which have something in common in terms of their measurement. The classic cases in neuroscience will be the same response measured over time or the same assay performed on material from different sections of the brain.\footnote{MANOVA is more flexible. In principle, it could be used to analyze different responses over times or assays for different substances in the same or different brain regions.}
Table 13.1: First six observations in the mRNA expression data set.

<table>
<thead>
<tr>
<th>Rat</th>
<th>Group</th>
<th>Area1</th>
<th>Area2</th>
<th>Area3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Control</td>
<td>8.58</td>
<td>15.31</td>
<td>9.13</td>
</tr>
<tr>
<td>2</td>
<td>Control</td>
<td>8.18</td>
<td>15.15</td>
<td>11.50</td>
</tr>
<tr>
<td>3</td>
<td>Control</td>
<td>11.70</td>
<td>16.91</td>
<td>15.36</td>
</tr>
<tr>
<td>4</td>
<td>Treatment</td>
<td>8.16</td>
<td>15.60</td>
<td>13.34</td>
</tr>
<tr>
<td>5</td>
<td>Treatment</td>
<td>8.06</td>
<td>15.78</td>
<td>9.51</td>
</tr>
<tr>
<td>6</td>
<td>Treatment</td>
<td>7.72</td>
<td>15.01</td>
<td>9.32</td>
</tr>
</tbody>
</table>

### 13.2 Correlated dependent variables

#### 13.2.1 The error covariance matrix

In a ordinary GLM, there is a single dependent variable, and the prediction errors have a mean of 0 and a variance that can be computed after the GLM is fitted. When there are multiple dependent variables, there will be prediction errors for each of the dependent variables. In the example above, there are two dependent variables, Area1 and Area2, so there will be prediction errors for Area1 and also prediction errors for Area2. After the model has been fitted to both variables, we can calculate the prediction errors. Call them $E_1$ and $E_2$. The equations for these are

$$ E_1 = Y_1 - \hat{Y}_1 $$

and

$$ E_2 = Y_2 - \hat{Y}_2 $$

Given a model that generates predicted values, we can actually arrive at numbers for $Y_1$ and $\hat{Y}_2$. Subtracting these numbers from the observed values of the dependent variables, $Y_1$ and $Y_2$ gives actual numbers for $E_1$ and $E_2$.

We can now calculate the means of $E_1$ and $E_2$. They will be 0. We can also calculate the variances of $E_1$ and $E_2$. Because, however, there are two variables, we can also calculate the covariance and the correlation between $E_1$ and $E_2$.

The covariance matrix between $E_1$ and $E_2$ is called the error covariance matrix, and it is at the heart of the analysis of multiple dependent variables. In the present case, we can designate this matrix as $\Sigma_E$ and it will have the form

$$ \Sigma_E = \begin{pmatrix} \sigma^2_{E_1} & \text{cov}(E_1, E_2) \\ \text{cov}(E_1, E_2) & \sigma^2_{E_2} \end{pmatrix} $$

(13.3)

If there were four dependent variables, then $\Sigma_E$ would be a four by four matrix with the variance of the errors for each of the four variables on the diagonal and the covariance for each pair of variables on the appropriate off diagonal element.

In Equation 13.3, the key statistic is $\text{cov}(E_1, E_2)$. As stated in the introduction, when this quantity is 0, then there is no need for the fancy methods.
described in this chapter. Just perform individual GLMs in each dependent variable and correct for multiple testing.

When variables are correlated, visual inspection of means and error bars can be misleading.

When, however, cov($E_1, E_2$) is not equal to 0, then it is necessary to use the techniques described in this chapter. Why? Because visual inspection of means and error bars can be misleading when the dependent variables are correlated. The next section explains why.

13.2.2 Why correlated error is important

Let’s begin with a thought experiment. We have a control and treatment group and assays in a dozen different areas of the brain. In reality there is no treatment effect in any area but by dumb luck, the treatment group has a higher mean than the controls in the Area 1. The difference is not statistically significant, but on a graph, it appears interesting. Suppose that the correlation between levels in Area 1 and Area 2 is 1.0. This necessitates that the Treatment mean will also be higher than the Control mean in Area 2, and, relative to the standard deviation in Area 2, by the same amount as the difference in Area 1. In other words, the mean difference in Area 2 looks just as interesting as the one in Area 1 although it too does not reach significance.

Area 3 is also correlated perfectly with Areas 1 and 2 as are the remaining nine brain areas. A plot of all means would resemble that in Figure 13.2. Anyone unfamiliar with the thought experiment would probably interpret the results in Figure 13.2 as follows. “Wow! In all 12 areas, the treatment group is greater than the control group. Even though there is no statistical significance in an area, the overall patterning cannot be due to chance. There must be a treatment effect.” The reality, a chance difference occurred and it looks “regular” only because of the perfect correlation among the brain areas.

Of course, in the real world one would never observe a perfect correlation among the 12 brain areas. The principle, however, remains the same. Let’s take a simple example. Height and weight are correlated. On average, taller people weigh more than shorter people. If we took two random samples of people and by dumb luck one was noticeably—but not statistically significantly—taller than the other, what would you expect their mean weights to be? The best bet is that the taller group would weigh more than the shorter group. The taller group would probably have a larger head circumference, longer femurs and larger big toe nails. The general point is this—if the slings and arrows of outrageous fortune lead one group to differ from another on variable X, the the two groups should be expected to differ on all variables highly correlated with X.

This example illustrates what expected means should look like when variables are positively correlated. Groups higher on variable X are also expected
to be higher on variables correlated with X. When the study design involves sampling from the general population, then this pattern is to be expected.

In experimental research, however, the correlations among dependent variables may hide important relationships. Re-examine Figure 13.1. Under the (correct) model that absent any experimental manipulations mRNA expression is positively correlated in the three areas, then higher means for Controls in Areas 1 and 3 might be expected, if by chance, the Controls happened to have a higher mean on one of the two variables. The problem comes with Area 2. According to the “sampling error/by chance” model, Controls should also have a higher mean than the Treatment group in this assay. They do not. Is this just a chance aberration or is it something of substance? Here, only the statistics can tell the difference. There are too many probabilities for the human mind to consider at once.

### 13.2.2.1 A really gnarly example

Focus on brain areas 1 and 2 in Figure 13.1 and let us pursue the likelihood of observing these two means under different values of the correlation between mRNA expression in Area 1 and Area 2. Figure 13.3 illustrates this relationship under four different values of the correlation. In this figure, the blue dot is the coordinate for the control means for Area 1 and Area 2 in Figure 13.1 and the
Figure 13.3: mRNA expression means as a function of the correlation coefficient in brain areas 1 and 2.
red square is the coordinate for the Treatment means in these two areas. Note that these means are the same in all four quadrants, i.e., the blue dot and the red square do not change. Neither do the shapes and sizes of the two gray boxes on the upper left and lower right of the four figures. What changes is the

Concentrate on the upper left pane where $\rho = 0$. The small black dots are the result of 10,000 simulated means based on the data from 13.1 under the assumption that the correlation between the rat’s mRNA expression in Area 1 and the mRNA expression in Area 2 is 0. The solid line is the regression line for these 10,000 simulated means. The upper left hand gray box gives the area of observing sampled means more extreme than the Treatment means. The lower right hand gray box gives the area for observing samples means more extreme than the Control means. An analogue of a $p$ value for the data in the quadrant where $\rho = 0$ is to count up all the points falling within the two gray boxes and divide by 10,000.

Clearly, when $\rho = 0$, a large number of dots falls into the two gray boxes, giving a high $p$ value. Your favorite biostatistician would say “if the correlation between mRNA levels in the two areas is 0, then the data on the first two brain areas in Figure 13.1 suggest that the experimental treatment has no effect on the expression of this gene.”

Compare this with the lower right-hand quadrant of Figure 13.3. Here, the correlation in expression in Area 1 and Area 2 is 0.75. The percentage of small black dots that fall into the “by chance” gray boxes is less than 5%. Hence, the probability of observing the blue dot mean and the red square mean is less than 0.05. If the correlation is this large, then the two observed means are unlikely to have been observed simply by chance.

It takes mental practice to envision the sampling distributions of means when the variables are correlated, but a familiar example can help. Height is correlated with weight, and men are taller on average than women. Hence, we expect that in any random sample of people, the average weight of men would exceed the average for women. Now re-examine Figure 13.1, substitute “men” for Controls and “women” for Treatment. Let “height” be brain area 1, and “weight” be brain area 2. The height data are hardly surprising—men are taller than women. But the weight data demand explanation because the women weigh more than the men! We would automatically suspect something funny such as nonrandom sampling.

We mammals have nervous systems highly developed to be sensitive to learning. If you can experience, cogitate on, and arrive at logical conclusions about many different sampling schemes and experimental effects on a system of correlated variables, then you are a better mammal than I am. The only way that I can deal with such complexity is to let the statistics sort out the issues.

### 13.3 Response profiles

When there are multiple dependent variables and the GLM model has groups, the starting point and the ending point for analysis is a table or plot of means of
the dependent variables. Hence, the very first task in any analysis is to calculate group means and either make a table or a graph of them. These plots are often called profiles.\textsuperscript{2}

Imagine the plot of means (in which the means are connected by lines) for a single group as a potentially mountainous landscape with peaks and valleys. This cartoonish landscape is called a “profile.” Some profiles can be very jagged with high peaks and deep valleys. Others may be relatively flat. Still others might resemble an incline with bumps here and there.

The profile has two characteristics—level and shape. Level is analogous to overall elevation and can be measured as the average of the group’s means on the dependent variables. Shape is the configuration of the peaks and valleys.

In analyzing multiple dependent variables, always compare the profiles for the groups ask yourself whether they differ in level only, in shape only, or in both level and shape. Figure 13.4 gives different types of profiles on three dependent variables for two groups.

Panels (A) and (B) illustrate profiles that differ only in level but not shape. The two are “parallel” in a general sense. If one moved Group 2’s profile up, only could perfectly overlap it with Group 1’s profile.

Panels (C) and (D) depict profiles that do not differ in level, but do differ in shape. The average of the three means for Group 1 equals the average for Group 2. Hence, overall average “elevation” is the same in both groups. The configural shapes of the profiles, however, differ in the two groups.

Finally, panels (E) and (F) illustrate profiles that differ in both level (or elevation) and shape.

Differences in profile elevation shape are test by the interaction between the grouping variable and the dependent variables. The main effect of the grouping variable tests for differences in overall elevation. This will become clear later when we deal with an example. First, however, we must discuss the two types of repeated measures analysis.

13.4 Types of repeated measures analysis

There are two major statistical techniques to perform repeated measures analysis\textsuperscript{3}. The first will be called “classical” or “traditional” repeated measures. This is an extension of the GLM that was already discussed to multiple dependent variables. There are four major characteristics of the classic approach. First, any observation with a missing value on any of the independent or dependent variables is completely eliminated from the statistical analysis. Second, the estimation of parameters and statistical tests use the least squares criterion (see Section X.X). Third, the error covariance matrix (described below) is assumed

\textsuperscript{2}There are advanced statistical techniques for analyzing profiles. See Tabachnick & Fidel (20X.X) for further details.

\textsuperscript{3}Technically, there are three, the third being generalized least squares. The difference, however, between generalized least squares and the mixed model approach is not great and the current generalized least squares software is not as widely available as mixed model software.
Figure 13.4: Examples of response profiles that differ in elevation (level) and shape.
to have a certain form or pattern. Finally, you have limited control over the models that can be fitted to the data.

The second approach is to use "mixed models." The whole mixed model statistical approach is a very big circle that encompasses the smaller circle of using mixed models for repeated measures analysis. The big circle is beyond the scope of this text. It is useful for observational and epidemiological problems in neuroscience, but is not relevant for laboratory experiments.

The four major characteristics of the mixed model approach to repeated measures analysis are the same as the four characteristics for classic repeated measures analysis (insert a "not" in each of the four items). First, observations with a missing value on one or more of the independent and dependent variables will be included in the analysis. Second, estimation and statistical tests are not based on least squares; they are performed using maximum likelihood. Third, no form or pattern is assumed for the error covariance matrix, although you may specify one. And finally, you have complete control over the models that can be fitted to the data.

You may rightly ask why traditional repeated measures analysis is even considered given the obvious advantage in flexibility of mixed models. The major reason is that most of neuroscience has not kept pace with advances in quantitative analysis. (Highly technical areas like imaging and genomics are an exception. Here, it takes professional biostatisticians to develop algorithms for data reduction and analysis). When a design has repeated measures, many neuroscientists do not even recognize that fact, let alone perform the appropriate analysis. Most of those who recognize the repeated measures design of their experiment are cognizant of only the traditional approach. Hence, if you submit a paper using the state of the art statistical techniques, be prepared to encounter resistance. Editors and reviewers who have not kept up with the literature often regard a "mixed model," something that biostatisticians consider standard fare, as suspicious and untested, even magical. Prepare yourself for rejection.

Hence, we will treat traditional repeated measures here, even though the mixed model approach has decided advantages. Moving the field of neuroscience away from misleading visual interpretation of means and error bars towards a more scientific approach to correlated dependent variables is, at least, a minor victory.

### 13.5 Choosing the type of analysis

Figure 13.5 presents a flowchart for determining the type of analysis to perform. In using this flowchart, it is assumed that you have already determined that the dependent variables are indeed correlated.

The MANOVA will always be a correct choice, but not necessarily the best choice. The problem is that the MANOVA, on average, has less statistical power than a repeated measures analysis. Hence, if you can meet the assumptions of repeated measures, then that will be a better choice.

A second major principle to keep in mind is that good traditional repeated
measures software will also perform a MANOVA and test the assumptions of the repeated measures model. You may have to choose some options to perform these analyses, but they should always be available. Thus, you can always perform the repeated measures analysis and examine the output. If the assumptions are not met, then you can interpret and report the MANOVA results without performing a separate analysis. If the assumptions are met, then you should interpret the repeated measures statistics.

A final consideration is the model. Mixed model software permits more flexibility in specifying the model than traditional repeated measures. Traditional repeated measures software automatically fits interaction terms involving what are called within-subject effects and other terms in the model. These interactions do not have to be modeled using mixed model software.

The following sections explain each decision point in the flowchart. You will note an attitude of “just go ahead and fit the classic repeated measures model.” That is no mistake. When you fit the classic model and use the appropriate options, then the results will give two sets of statistics. The first is the MANOVA set of statistics. These are always valid but may not be the optimal statistics. The second are the classic repeated measures statistics. If the statistical tests indicate that the assumptions of the traditional repeated measures statistics are not valid, you can always interpret the MANOVA statistics.

13.5.1 Missing values less than 5%?

There have been whole books written on how to deal with missing values but most of these esoteric applications are more suitable for large observational data sets than for experimental neuroscience.

If an observation has even one missing value on all of the independent and dependent variables, then traditional repeated measures completely ignores that observation. The non missing values do not enter into the analysis. Mixed models, on the other hand, will use all of the available data. Hence, if your data have a number of missing values, use mixed models. If the number is very small replace the value with the overall mean of the variable.

The 5% rule is completely arbitrary. One would really like to see a large analysis of many different data sets in neuroscience that compares the result from the analysis of a complete data set to those results in which data points were selected at random and replaced by the mean. I am not aware of any such study, so there really no empirical guidelines. The 5% is a clinical guesstimate based on teaching statistics to graduate neuroscience students and giving them data sets with missing values.

The small sample sizes in neuroscience should encourage a conservative approach in dealing with missing values. Suppose that you have a missing value for a time point in a control animal. Do you replace that with the overall mean or with the mean of the controls? Replacing with the overall mean is conservative and will diminish statistical power (i.e., make it harder to detect an effect

\footnote{Within-subjects and between-subjects effects are discussed later in Section X.X}
Figure 13.5: Flowchart for the analysis of multiple dependent variables.
that is truly present). Replacing it with the control mean is liberal and will increase the probability of a false positive. If the sample size for a group were 30 or more, the effect would be very small. When the group sample size is 4, however, the effect could give misleading results.

If you have any uncertainty, follow the mixed model approach. It is always justified.

13.5.2 Automatic model

The term “automatic model” is not standard. It is used here to denote the fact that classic repeated measures will automatically fit interaction terms. To understand this, recall the mRNA expression example in Section 13.1. The independent variable is Group, a GLM factor with two levels—Control and Treatment. The three dependent variables can be considered another GLM factor. Call it Area and note that it has three levels—Area1, Area2, and Area3. A classic repeated measures analysis will predict the mRNA density as a function of Group, Area, and the (Group × Area) interaction. You do not have a choice about the interaction term. It will be fitted and you cannot drop it. This is an “automatic model.”

With the mixed model approach, you have complete control over the model. You can drop the interaction term if you wish. Hence, the mixed model gives you more flexibility over the model than classic repeated measures.

13.5.3 Are the variables cognate?

The next step in analyzing multiple dependent variables is to ask yourself whether the variables are similar in terms of what they measure. If the answer is “yes,” then the variables are cognate. Examples would include the same response measured over time or an assay for the same substance performed in different areas of the brain. In some cases, the term “cognate” could refer to different assays of the same generic class of substances. For example, assays of cortisol, corticotropin-releasing hormone (CRH), and adrenocorticotropic hormone (ACTH) may be considered cognate in the context of studies about activation of the hypothalamic-pituitary-adrenal (HPA) axis.

If you are in a quandary and suspect that your type of data falls into a gray area, go with “yes” in the flowchart. After all, if you do the classic repeated measures analysis and use the options to output the MANOVA, you can always interpret those statistics.

13.5.4 Is the error covariance matrix patterned?

Technically, you do not have to ask this question. It is included here only to alert you to one of the main assumptions of classic repeated measures analysis,
namely that the error covariance matrix has a Type H pattern. The pattern is discussed later. Here, let’s explain why you do not need to answer this question.

The computation of an error covariance matrix is laborious. If you fit a classic repeated measures analysis according to the algorithm described later, it will calculate the error covariance matrix and test whether or not it meets the Type H assumption. Hence, do not worry about the pattern. Just go ahead and run the analysis but make certain to examine the error covariance matrix and the test of whether or not it meets the Type H assumption.

The most common Type H pattern is called compound symmetry. In compound symmetry, the diagonals of the error covariance matrix (i.e., the variances) are all the same and the off diagonals (i.e., the covariances) are all the same. Let $\Sigma_E$ denote the error covariance matrix, $\sigma^2_t$, the variance of a variable, and $\rho$, the correlation between a pair of variables. A matrix with compound symmetry has the following algebraic form

$$\Sigma_E = \begin{pmatrix} \sigma^2 & \rho \sigma^2 & \rho \sigma^2 \\ \rho \sigma^2 & \sigma^2 & \rho \sigma^2 \\ \rho \sigma^2 & \rho \sigma^2 & \sigma^2 \end{pmatrix} \quad (13.4)$$

Many error covariance matrices that do not meet compound symmetry, however, can come close enough to a Type H pattern to be used in classic repeated measures.

### 13.6 Classic repeated measures

In terms of logic, a classic repeated measures GLM is no different than an ordinary GLM. The differences are in two mechanical areas: (1) the mechanics of estimating parameters and testing for significance are different; and (2) the output from the statistical procedure is confusing because it follows old “hand calculation” algorithms and also involves many advanced statistical quantities.

The following algorithm let you perform and interpret classic repeated measures. It will not, however, let you understand all of it. Full understanding requires expertise in matrix algebra and multivariate statistics—two topics beyond the scope of this text. Do not worry much about the “black box” of the statistics that are not discussed here. One can be very expert in using a computer without complete understanding of the physics behind the hardware.

The algorithm for a classic repeated measures design in given in Table 13.2. The next sections describe each step of the algorithm.

#### 13.6.1 Write a traditional GLM table for the design

Revisit the mRNA expression data. The scientific questions are: (1) does the treatment alter mRNA expression; (2) does gene expression differ over the three brain regions; and (2) is the treatment effect the same across the three brain

---

6A Type H pattern is given in the the advanced Section 13.8.
Table 13.2: Algorithm for classic repeated measures.

<table>
<thead>
<tr>
<th>Step</th>
<th>Task:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Write a traditional GLM table for the design.</td>
</tr>
<tr>
<td>2</td>
<td>Edit the columns of the GLM table.</td>
</tr>
<tr>
<td>3</td>
<td>Determine the between-subjects and the within-subjects effects.</td>
</tr>
<tr>
<td>4</td>
<td>Fit the GLM model.</td>
</tr>
<tr>
<td>5</td>
<td>Copy relevant quantities from the output into the summary table.</td>
</tr>
<tr>
<td>6</td>
<td>Examine the summary table for consistent results</td>
</tr>
<tr>
<td>7</td>
<td>If necessary, revise the GLM model and start over</td>
</tr>
</tbody>
</table>

Table 13.3: Logical ANOVA table of effects for the mRNA expression data ignoring repeated measures.

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>$F_{obs}$</th>
<th>$p(F &gt; F_{obs})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Area</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group $\times$ Area</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

regions? There are three variables in this logic. The dependent variable, of course, is the density of the mRNA assay. The two predictor variables are Group (control versus treatment) and Area of the brain. According to logic and what we have already learned, if we completely ignored repeated measures we would write the model as

$$mRNA = \beta_0 + \beta_1 Group + f(Area) + f(\text{Group} \times \text{Area})$$

Note that the notation $f(Area)$ and $f(\text{Group} \times \text{Area})$ is used because Area is being treated as a qualitative variable. The GLM or ANOVA table of effects would have the generic form given in Table 13.3.

You should write down a table like this for your problem. Although 13.3 includes the columns labeled “df” through “$p(F > F_{obs})$,” you only need to write down the column labeled “Source,” “$F_{obs}$”, and “$p(F > F_{obs})$.”

### 13.6.2 Add columns to the table

The next step in performing a traditional repeated measures analysis is to add columns. This starts with the Source column from Table 13.3 and then adds a variable called Type. (We discuss that later). It drops the columns for df, SS, and MS but keeps those for the observed $F$ statistic ($F_{obs}$), and the probability value for the $F$ statistic (column $p(F > F_{obs})$ in Table 13.3, but renamed $p(F_{obs})$ in Table 13.4 to save space). Add two more columns for two types of $p$ values that are adjusted to account for the discrepancies in assumptions about the
error covariance matrix. Label these columns $p$(GG) for the Greenhouse-Geisser correction and $p$(HF) for the Huynh-Feldt correction.\footnote{An update of the Huynh-Feldt correction may also be called the Huynh-Feldt-Lecoutre correction which may be output in some software.}

Finally, add two columns with the respective labels “$p$(Wilks)” and “$p$(Pillai).” The columns will be for statistics for some of the MANOVA effects.

A complete table of this form is shown in Table 13.4.

\subsection*{13.6.3 Determine the between-subjects and within-subjects effects.}

Here, we enter a B or a W under the column labeled “Type” for each effect in the model. A “B” stands for between-subjects effects and a “W” stands for within-subjects effects. A between-subjects effect is any variable in the model that differentiates the rows of the data set. Variable Group is a between-subjects effect because it differentiates the rows into two types–Control and Treatment.

A within-subjects effect is any variable in the model that differentiates the columns of the data set. Variable Area is a within-subjects effect because it differentiates among the columns called Area1, Area2, and Area3. Finally, any interaction that involves a within-subjects effect is also a within-subjects effect. Hence the interaction between Group and Area is treated as a within-subjects effect.

Some of these statistics are not relevant for the between-subjects effects.

Hence, put an “X” in the columns labeled “$p$(GG)” through “$p$(Pillai)” for all the between-subjects effects.

The trick now is to run a repeated measures analysis and copy the relevant items from the output into Table 13.4. Hold onto your seat. This part gets pretty ugly.

\subsection*{13.6.4 Fit the GLM model}

Here, software makes a big difference. Use either SAS or SPSS. They have all of the options required to perform the analysis with the options of testing the assumptions of the repeated measures. You can perform a classic repeated measures analysis in R, but you must be very careful. There are some very simple R scripts in books and on the web. Some of them make a number of hidden assumptions which, if violated, can give misleading results.

Table 13.5 gives the SAS code for performing a repeated measures analysis on the mRNA expression data. The PROC GLM statement starts off like a typical GLM with the expectation that in the MODEL statement there are three dependent variables instead of one. The next statement is a MANOVA statement that tests the null hypothesis that there is no overall difference between the control and treatment groups on the means of the three dependent variables (the H = Group argument to the MANOVA statement).

If your data have more than one between-subjects effects, then include a separate MANOVA statement for each between-subjects effect. For example,
Table 13.4: Summary table for a classic repeated measures analysis on the mRNA expression data.

<table>
<thead>
<tr>
<th>Source</th>
<th>Type</th>
<th>$F_{obs}$</th>
<th>$p(F_{obs})$</th>
<th>$p(GG)$</th>
<th>$p(HF)$</th>
<th>$p(\text{Wilks})$</th>
<th>$p(\text{Pillai})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Area</td>
<td>W</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group $\times$ Area</td>
<td>W</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 13.5: SAS code for a classic repeated measures analysis on the mRNA data.

```sas
PROC GLM DATA=qmin13.mRNAdata;
  CLASS Group;
  MODEL Area1 Area2 Area3 = Group;
  * Put a separate MANOVA statement here, one for each between subjects effect.
  If you have multiple effects, you only need to specify the PRINTE option in the first one;
  MANOVA H=Group / PRINTE;
  * Include the PRINTE option on the REPEATED statement to get Mauchly’s test of the assumption that the error covariance matrix is Type H;
  REPEATED BrainArea 3 CONTRAST / PRINTE SUMMARY;
RUN;
QUIT;
```

Suppose the mRNA data set had a GLM factor of Age with levels Old and Young. Age is a between-subjects effect because it differentiates the rows from one another. We would include the statement “MANOVA H=Age” to test the Age effect. The interaction between Group and Age would also be a between-subjects factor because both Group and Age are themselves between-subjects effects. Hence, we would also include the statement “MANOVA H=Group*Age” to perform a MANOVA for the interaction.

### 13.6.5 Copy the relevant quantities into the summary table

We not wade through a large amount of output, much of which can be ignored, to find the relevant quantities to copy into our summary table, i.e., Table 13.4. Figures 13.6 through 13.9 present the selected portions of the output that are important to examine.

#### 13.6.5.1 PROC GLM output, section (1)

The first part of the output from PROC GLM is list what is called the “univariate GLMs.” Because there is only only predictor variable on the right hand side of the MODEL statement in Table 13.5, the significance of the overall ANOVA table is the same as the significance for Group. Hence, only the overall ANOVA table is presented. Note that none of the univariate GLMs is significant. A researcher who is insensitive to the fact that scores in the brain areas are correlated could very easily look at the output in Figure 13.6 and mistakenly conclude that the treatment has little influence on mRNA expression.
13.6.5.2 PROC GLM output, section (2)

This part begins the output from the MANOVA statement. Because the option PRINTE was given on the MANOVA statement, PROC GLM prints the error matrix. Technically, this is a fancy type of matrix called a sum of squares and cross products (SSCP) matrix, but if we divided each of the numbers in this matrix by the degrees of freedom, we would have the error covariance matrix. Note that the entry for Area3 is over four times that for Area1. This cautions us that the variances of mRNA expression in the brain areas differs.

The really important matrix is labeled the “Partial Correlation Coefficients from the Error SSCP Matrix.” Ignore the word “partial.” This is the correlation matrix for the error covariance matrix. Always inspect the patterning of these correlations. They can tell you a lot about how individual differences in one dependent variable predict individual differences in another dependent variable. If the correlations are all low and hardly different from 0, then there is no need to perform repeated measures. You can interpret the univariate GLM from Figure 13.6 and ignore the rest of the output.

In the present example, the three correlations are moderate and close to being equal. This means that rats with high expression in any one area tend to have high expression in the other two areas.

13.6.5.3 PROC GLM output, section (3)

This section gives the results of the MANOVA for the between-subjects factor Group. The null hypothesis here is that the (3 by 1) vector of means for each group is being pulled out of the same “hat.” The hat in this case does not contain individual means. Rather, it contains a large number of slips of paper. On each slip three means are written, one for Area 1, one for Area 2, and one for Area 3. The null hypothesis is that the only differences among the (3 by 1) vectors of observed means for the Control and the Treatment groups is due to sampling error.

Traditional multivariate analysis prints four statistics: Wilks’ \( \lambda \), Pillai’s trace, the Hotelling-Lawley trace, and Roy’s greatest root. Again, the nature and properties of these statistics are topics beyond our scope of discussion. Wilks’ \( \lambda \) is the statistic most often used presented, although Pillai’s trace is also encountered. Hence, we concentrate on only those two.

To assess significance, these statistics are converted to \( F \) statistics and the significance value of the \( F \) is used. In some cases, the present case being one of them, the conversion is mathematically exact. That is the reason why the \( F \) statistic for all four multivariate statistics is the same. In other cases, the conversion is an approximation and the \( F \) statistics can vary. The typical way of reporting the results is to present the multivariate statistic, the \( F \), and the \( p \) value for \( F \). For example, “Wilks’ \( \lambda \) = 0.61, \( F(3, 16) = 3.34, p = .046 \).”

Contrast the multivariate results with those for the three univariate GLMs in Figure 13.6. In the multivariate case, there is a strong suggestion that there are indeed differences in mRNA expression. The multivariate test examines the
observed means in Figure 13.1 and asks the omnibus question of whether the means of the Controls differ from those of the Treatment group in level and/or shape.

The reason why the multivariate statistic found significance is that there is a different in the profile patterning to the means in the two groups. (This will become clearer when we examine other results later in the output). The partial correlations from section (2) of the output suggest that a group that, by dumb luck, is high on expression in one area should also be high in the other two. Instead, Figure 13.1 demonstrates that the treatment group is higher, not lower, than the controls in Area 2. Multivariate statistics are sensitive to such differences in patterns. Univariate statistics focus on one variable and ignore the others. Hence, they will not always be able to detect a difference in pattern.

13.6.5.4 PROC GLM output, section (4)

The actual repeated measures analysis generated from the REPEATED statement in Table 13.5 begins here. The first section gives the design for the within-subjects factor. There is only one factor in this analysis—it was called BrainArea in the REPEATED statement in Table 13.5—and it has three levels. Hence, the dependent variables for repeated measures factor BrainArea are, in order, Area1, Area2, and Area3. It is possible to have more complicated within-subjects designs, so it is crucial to examine the output in this section to make certain that you have specified the correct design in the REPEATED statement.

13.6.5.5 PROC GLM output, section (5)

This section outputs what are called “sphericity tests.” The key one is the one performed on the “Orthogonal Components” because this tests whether the assumption that the error covariance matrix has a Type H pattern is reasonable. If the $p$ value for the $\chi^2$ is low, then we reject the hypothesis that the pattern in Type H. In the example, $\chi^2 = 3.83$ and $p = 0.147$. Hence, we conclude that the error covariance matrix is reasonably close to a Type H matrix. This permits us to interpret the results of the classic repeated measures analysis.

13.6.5.6 PROC GLM output, section (6)

Section (6) gives the MANOVA results for the within-subjects factors. As before, record only $p$ values for Wilks’ $\lambda$ and Pillai’s trace. The effect for BrainArea is highly significant ($F(2, 17) = 314.57, p < .0001$). This means that controlling for mean differences between the Control and Treatment Groups, the means for the three different brain areas are not being sampled from the same “hat” of means.

Whether differences in mRNA density among the three areas is of substantive importance is a very different matter from its statistical significance. Here, the result must be placed into the context given by what is already known
Figure 13.6: Output from PROC GLM for the repeated measures analysis of the mRNA expression data, part 1.

<table>
<thead>
<tr>
<th>Dependent Variable: Areal</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source</td>
<td>DF</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model</td>
<td>1</td>
<td>2.56328000</td>
<td>2.07</td>
<td>0.1675</td>
</tr>
<tr>
<td>Corrected Total</td>
<td>18</td>
<td>22.30170000</td>
<td>1.2389333</td>
<td></td>
</tr>
<tr>
<td>R-Square</td>
<td>0.103088</td>
<td>12.98980</td>
<td>1.113096</td>
<td>8.569000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dependent Variable: Area2</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source</td>
<td>DF</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model</td>
<td>1</td>
<td>3.96050000</td>
<td>1.95</td>
<td>0.1796</td>
</tr>
<tr>
<td>Error</td>
<td>18</td>
<td>36.55942000</td>
<td>2.03107889</td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>19</td>
<td>40.51992000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R-Square</td>
<td>0.097742</td>
<td>9.087867</td>
<td>1.425159</td>
<td>15.68200</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dependent Variable: Area3</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source</td>
<td>DF</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model</td>
<td>1</td>
<td>11.17512500</td>
<td>2.57</td>
<td>0.1262</td>
</tr>
<tr>
<td>Error</td>
<td>18</td>
<td>78.23437000</td>
<td>4.34635389</td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>19</td>
<td>89.40949500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R-Square</td>
<td>0.124988</td>
<td>18.48957</td>
<td>2.084791</td>
<td>11.27550</td>
</tr>
</tbody>
</table>
Figure 13.7: Output from PROC GLM for the repeated measures analysis of the mRNA expression data, part 2.

The GLM Procedure
Multivariate Analysis of Variance

\[ E = \text{Error SSCP Matrix} \]

\[
\begin{array}{ccc}
\text{Area1} & \text{Area2} & \text{Area3} \\
\text{Area1} & 22.3017 & 15.65304 & 24.78371 \\
\text{Area2} & 15.65304 & 36.55942 & 25.37513 \\
\text{Area3} & 24.78371 & 25.37513 & 78.23437 \\
\end{array}
\]

Partial Correlation Coefficients from the
Error SSCP Matrix / Prob > |r|

\[
\begin{array}{ccc}
\text{DF = 18} & \text{Area1} & \text{Area2} & \text{Area3} \\
\text{Area1} & 1.000000 & 0.548189 & 0.593333 \\
& & 0.0151 & 0.0074 \\
\text{Area2} & 0.548189 & 1.000000 & 0.474471 \\
& & 0.0151 & 0.0401 \\
\text{Area3} & 0.593333 & 0.474471 & 1.000000 \\
& & 0.0074 & 0.0401 \\
\end{array}
\]

MANOVA Test Criteria and Exact F Statistics for the Hypothesis
of No Overall [Group] Effect

\[ H = \text{Type III SSCP Matrix for Group} \]
\[ E = \text{Error SSCP Matrix} \]

\[
S=1 \quad M=0.5 \quad N=7
\]

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Value</th>
<th>F Value</th>
<th>Num DF</th>
<th>Den DF</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wilks' Lambda</td>
<td>0.61446143</td>
<td>3.35</td>
<td>3</td>
<td>16</td>
<td>0.0456</td>
</tr>
<tr>
<td>Pillai's Trace</td>
<td>0.38553857</td>
<td>3.35</td>
<td>3</td>
<td>16</td>
<td>0.0456</td>
</tr>
<tr>
<td>Hotelling-Lawley Trace</td>
<td>0.62744146</td>
<td>3.35</td>
<td>3</td>
<td>16</td>
<td>0.0456</td>
</tr>
<tr>
<td>Roy's Greatest Root</td>
<td>0.62744146</td>
<td>3.35</td>
<td>3</td>
<td>16</td>
<td>0.0456</td>
</tr>
</tbody>
</table>
Figure 13.8: Output from PROC GLM for the repeated measures analysis of the mRNA expression data, part 3.

The GLM Procedure
Repeated Measures Analysis of Variance

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Area1</th>
<th>Area2</th>
<th>Area3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level of BrainArea</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

Regressed Means Level Information

<table>
<thead>
<tr>
<th>Variables</th>
<th>DF</th>
<th>Value</th>
<th>Pr &gt;</th>
<th>Value</th>
<th>Pr &gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transformed Variates</td>
<td>2</td>
<td>7.014</td>
<td>&lt;.001</td>
<td>15.256</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Orthogonal Components</td>
<td>2</td>
<td>5.462</td>
<td>&lt;.001</td>
<td>2.597</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

Sphericity Tests

MANOVA Test Criteria and Exact F Statistics for the Hypothesis of no BrainArea Effect

\[ H = \text{Type III SSCP Matrix for BrainArea} \]
\[ E = \text{Error SSCP Matrix} \]
\[ S = 1, M = 0, N = 7.5 \]

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Value</th>
<th>F Value</th>
<th>Num DF</th>
<th>Den DF</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wilks' Lambda</td>
<td>0.02630981</td>
<td>314.57</td>
<td>2</td>
<td>17</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Pillai's Trace</td>
<td>0.97369019</td>
<td>314.57</td>
<td>2</td>
<td>17</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Hotelling-Lawley Trace</td>
<td>37.00863978</td>
<td>314.57</td>
<td>2</td>
<td>17</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Roy's Greatest Root</td>
<td>37.00863978</td>
<td>314.57</td>
<td>2</td>
<td>17</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

MANOVA Test Criteria and Exact F Statistics for the Hypothesis of no BrainArea*Group Effect

\[ H = \text{Type III SSCP Matrix for BrainArea*Group} \]
\[ E = \text{Error SSCP Matrix} \]
\[ S = 1, M = 0, N = 7.5 \]

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Value</th>
<th>F Value</th>
<th>Num DF</th>
<th>Den DF</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wilks' Lambda</td>
<td>0.61919724</td>
<td>5.23</td>
<td>2</td>
<td>17</td>
<td>0.0170</td>
</tr>
<tr>
<td>Pillai's Trace</td>
<td>0.38080276</td>
<td>5.23</td>
<td>2</td>
<td>17</td>
<td>0.0170</td>
</tr>
<tr>
<td>Hotelling-Lawley Trace</td>
<td>0.61499427</td>
<td>5.23</td>
<td>2</td>
<td>17</td>
<td>0.0170</td>
</tr>
<tr>
<td>Roy's Greatest Root</td>
<td>0.61499427</td>
<td>5.23</td>
<td>2</td>
<td>17</td>
<td>0.0170</td>
</tr>
</tbody>
</table>
Figure 13.9: Output from PROC GLM for the repeated measures analysis of the mRNA expression data, part 4.

The GLM Procedure
Repeated Measures Analysis of Variance
Tests of Hypotheses for Between Subjects Effects

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type III SS</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>1</td>
<td>2.90840167</td>
<td>2.90840167</td>
<td>0.58</td>
<td>0.4545</td>
</tr>
<tr>
<td>Error</td>
<td>18</td>
<td>89.57308333</td>
<td>4.97628241</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

--- <Page> -------------------------------------------

The GLM Procedure
Repeated Measures Analysis of Variance
Univariate Tests of Hypotheses for Within Subject Effects

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type III SS</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
<th>G - G</th>
<th>H-F-L</th>
</tr>
</thead>
<tbody>
<tr>
<td>BrainArea</td>
<td>2</td>
<td>515.5810233</td>
<td>257.7905117</td>
<td>195.29</td>
<td>&lt;.0001</td>
<td>&lt;.0001</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>BrainArea*Group</td>
<td>2</td>
<td>14.7905033</td>
<td>7.3952517</td>
<td>5.60</td>
<td>0.0076</td>
<td>0.0120</td>
<td>0.0098</td>
</tr>
<tr>
<td>Error(BrainArea)</td>
<td>36</td>
<td>47.5224067</td>
<td>1.3200669</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Greenhouse-Geisser Epsilon 0.8321
Huynh-Feldt-Lecoutre Epsilon 0.9066
about gene expression in these areas. The same logic applies to other repeated measures designs. In clinical trials, patients tend to get better, so a repeated measures finding that there is significant decrease in symptoms over time is hardly surprising.

The test for the interaction between Group and BrainArea is significant (Wilks' $\lambda = 0.62$, Pillai's trace $= 0.38$, $F(2, 17) = 5.23$, $p = 0.17$). This tells us that there are significant profile differences in the patterning of the means between the Control and the Treatment groups. From Figure X.X, the profile of means for the Treatment group is more “pointed” than the one for Controls.

Given that the MANOVA results are significant, one could legitimately skip the rest of the output and just rely on the multivariate statistics. Significant MANOVA results coupled with the fact that the error covariance matrix does not significantly depart from being a type H matrix mean that the traditional repeated measures analysis will also give significant results for the within-subjects effects in the model.

13.6.5.7 PROC GLM output, section (7)

The final section of output gives the results of the traditional or classic repeated measures analysis. The output is divided into two sections, the first dealing with the between-subjects effects and the second presenting the within-subjects effects.

There is only one within-subject effect, Group, and it is not significant ($F(1, 18) = 0.58$, $p = 0.455$). Once again, to interpret this effect we must always examine the observed means. Between-subjects effects always test for differences in profile level, not profile shape. Hence, the null hypothesis for Group is that the average of the three means for the Control group is within sampling error of the three means of the Treatment group.

You may rightly question why Group is not significant here whereas it was significant in the MANOVA. In this part of the output the hypothesis refers only to profile level. The MANOVA, however, tests for differences in level and/or shape. The fact that the classic repeated measures found an insignificant for Group tells us that there is no difference in overall level. The significant MANOVA, however, tells us that the significant difference between groups is entirely due to differences in profile shape.

The results for the within-subjects effects are listed as the “Univariate Tests.” Here, we see an ANOVA table of effects with the addition of two extra columns giving “adjusted” $p$ values for the observed $F$ statistic. The two adjustments are made to account for discrepancies between the actual error covariance matrix and its assumed form. The first adjustment, given in the column labeled “G - G” in the output is the Greenhouse-Geisser adjustment. The second, labeled “H-F-L” is the Huynh-Feldt-Lecoutre adjustment, most often called just the Huynh-Feldt or H-F adjustment. Given that the assumption of the form of the error covariance matrix was robust, we should not expect these values to differ much from the ordinary $p$ value for the $F$. And they don’t. You should copy these $p$ values into the appropriate columns of Table 13.5.
CHAPTER 13. GLM: MULTIPLE DEPENDENT VARIABLES

The final section of the output prints what are called the epsilon ($\epsilon$) statistics for the Greenhouse-Geisser and the Huynh-Feldt-Lecoutre adjustments. If the error covariance matrix was a perfect Type H matrix, then these values would be 1.0 and the adjusted $p$ values would equal the regular $p$ value.

13.6.6 Examine the table

Table 13.6 gives the results of placing the numbers from the output into Table 13.4. It is now time to look at patterns in the data. We have already talked of the patterns above, but repetition aids learning, so let's do it again.

First examine the between-subjects effects. Recall that the $F$ and its $p$ value from the classic repeated measures tests for only differences in profile level. The MANOVA results--$p$(Wilks) and $p$(Pillai)--on the other hand test for differences in both profile level and shape. Also, the $F$ and $p$ from the classic repeated measures is often more powerful than the MANOVA for detecting differences in level. A comparison of the traditional repeated measures statistics to the MANOVA for a between-subjects variable can give insight into the nature of the difference.

In the present case, the repeated measures $F$ for variable Group is not close to being significant. Thus, there is no evidence that the mean profiles between the Control and Treatment groups differ in overall level of mRNA density. The MANOVA, however, is significant, albeit just barely. This suggests that there may be a difference in profile shape. If there is a significant Group by Area interaction in the classic repeated measures analysis, then we will be more confident that the difference in shape is real.

Both the MANOVA and the classic repeated measures statistics for the within-subjects effects are consistent. They are all highly significant for brain Area, telling us that there are definitely mean differences among the three areas. The interaction of Group and Area is also significant in all statistics.

What happens when the significance levels are not consistent across the different levels? Recall that the classic repeated measures statistics will be more powerful than the MANOVA. Hence, if the assumption of a Type H error covariance matrix is close to being satisfied, the univariate $F$ statistics are significant, but the MANOVA $p$ values are not significant, then it is legitimate to interpret the classic statistics. If the assumption of a Type H matrix is not met, then ignore the classic $F$ and its $p$ value and examine the consistency for the Greenhouse-Geisser correction, the Huynh-Feldt correction, and the two MANOVA statistics. Finally, if there is still inconsistency, then be honest and just report that fact. Sometimes science does not give a clear black or white answer.

13.6.7 Revise the GLM model

Here, you may want to delete terms from the GLM model to arrive at a parsimonious model. Remember that you have no control over the interactions involving the within-subjects part of the design in classic repeated measures
Table 13.6: Summary table for a classic repeated measures analysis on the mRNA expression data.

<table>
<thead>
<tr>
<th>Source</th>
<th>Type</th>
<th>$F_{obs}$</th>
<th>$p(F_{obs})$</th>
<th>$p(GG)$</th>
<th>$p(HF)$</th>
<th>$p$(Wilks)</th>
<th>$p$(Pillai)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>B</td>
<td>0.58</td>
<td>0.4545</td>
<td>X</td>
<td>X</td>
<td>0.0456</td>
<td>0.0456</td>
</tr>
<tr>
<td>Area</td>
<td>W</td>
<td>195.29</td>
<td>&lt;.0001</td>
<td>&lt;.0001</td>
<td>&lt;.0001</td>
<td>&lt;.0001</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Group × Area</td>
<td>W</td>
<td>5.60</td>
<td>0.0076</td>
<td>0.0120</td>
<td>0.0098</td>
<td>0.0170</td>
<td>0.0170</td>
</tr>
</tbody>
</table>
13.7 Mixed model repeated measures

Here, I give strong recommendations on software. Use the major statistical packages: SAS, SPSS or Stata. R can perform mixed model repeated measures, but the packages that do so (lme4) are so poorly documented that even professionals have difficulty understanding the procedures. Many other packages do not even have routines for mixed model analysis or, if they do, do not have the flexibility of the major statistical programs.

Also, recall that the treatment of mixed models here is only for repeated measures. Mixed models encompass a very large circle that contains the much smaller circle of repeated measures analysis. Try to rely documentation here and other references (e.g., Wolfinger, R. D. and Chang, M. (1995), "Comparing the SAS GLM and MIXED Procedures for Repeated Measures," Proceedings of the Twentieth Annual SAS Users Group Conference) that treat only the repeated measures aspects of mixed models.

SAS, SPSS, and Stata procedures for mixed model analysis required a stacked data set. Here, the rows are no longer independent observations. Instead, there is a variable like ID number that identifies all of the data that belongs to an independent observation. There stacked data set has a single “dependent variable” and other variables that inform the program of the within-subjects effects.

In the mRNA data, the stacked data set must have a variable denoting the individual rat. We will call this Rat and it will be the integer in the data set that distinguishes one rat from another. The we will call the dependent variable “mRNA” and it will be the value of Area1, Area2, or Area3 in the data set previously displayed in Table X.X. Finally we must have a variable informing the procedure from area of the brain the value of mRNA is taken. We call this variable Area and assign it values of 1, 2 or 3 for, respectively, measurements of Area1, Area2 and Area3. In SAS, you can construct a data set using the point-and-click interface in SAS Enterprise or do it by code. SPSS also has a point and click interface. Otherwise, you can create one yourself using a little coding. Table 13.7 gives SAS code for creating a stacked data set for the mRNA expression data and Table 13.8 displays the first six records in the stacked data set.

To specify a repeated measures mixed model, start with Section X.X. This gives the basic logic of the design. Remember that the logic of the effects in the GLM remains the same—it is just that we must use fancy methods to estimate the parameters and assess statistical significance. Hence, in the mRNA data, we have the model

\[
\text{mRNA} = \beta_0 + \beta_1 \text{Group} + f(\text{Area}) + f(\text{Group} \times \text{Area})
\] (13.5)
Table 13.7: SAS code to create a stacked data set for the mRNA expression data.

```sas
DATA stacked;
  SET qmin13.mRNAdata;
  Area = 1; mRNA = Area1; OUTPUT;
  Area = 2; mRNA = Area2; OUTPUT;
  Area = 3; mRNA = Area3; OUTPUT;
  KEEP Rat Group Area mRNA;
RUN;
```

Table 13.8: The first two independent observations in a stacked data set for the mRNA expression data.

<table>
<thead>
<tr>
<th>Rat</th>
<th>Group</th>
<th>Area</th>
<th>mRNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Control</td>
<td>1</td>
<td>8.58</td>
</tr>
<tr>
<td>1</td>
<td>Control</td>
<td>2</td>
<td>15.31</td>
</tr>
<tr>
<td>1</td>
<td>Control</td>
<td>3</td>
<td>9.13</td>
</tr>
<tr>
<td>2</td>
<td>Control</td>
<td>1</td>
<td>8.18</td>
</tr>
<tr>
<td>2</td>
<td>Control</td>
<td>2</td>
<td>15.15</td>
</tr>
<tr>
<td>2</td>
<td>Control</td>
<td>3</td>
<td>11.50</td>
</tr>
</tbody>
</table>

Again read “the predicted value of mRNA equals a constant ($\beta_0$) plus a weight ($\beta_1$) times variable Group plus a function of variable Area plus a function of the interaction between variables Group and Area.” We can use the weight $\beta_1$ for Group because Group has only two levels—Control and Treatment. We use $f(\text{Area})$ and $f(\text{Group} \times \text{Area})$ because Area has three levels and, hence, more than one $\beta$ will be used for these levels.

The SAS statements to perform a repeated measures analysis using mixed models are presented in Table 13.9.

Table 13.9: SAS statements to perform a repeated measures analysis using PROC MIXED.

```sas
PROC MIXED DATA=stacked;
  CLASS Group Area;
  MODEL mRNA = Group Area Group*Area / SOLUTION;
  REPEATED Area / SUBJECT=Rat TYPE=HF R RCORR;
RUN;
```
CHAPTER 13. GLM: MULTIPLE DEPENDENT VARIABLES

13.7.1 Output from PROC MIXED

The PROC MIXED procedure in SAS gives a large amount of output, most of which can be ignored. Figures ??, 13.10, and 13.11 give the output from the PROC MIXED analysis in SAS. The order has been slightly rearranged for convenience. The following sections explain the output.

13.7.1.1 PROC MIXED output, section (1)

This part of the output described the data and some aspects of the model. Peruse it to make certain that you are analyzing the correct data set, that the dependent variable is correct, that any class (i.e., categorical) variables have the correct number of levels, that the correct form of the error covariance matrix is used (this is the line labeled “Covariance Structure”), and that the data have been read in properly. If you make a mistake in creating a stacked data set, try to catch it here instead of wasting pondering meaningless results.

13.7.1.2 PROC MIXED output, section (2)

In mixed models, parameters are calculated and statistics are estimated using what mathematicians call “numerical methods.” Numerical methods start with a reasonable set of parameter estimates, evaluate the fit of the model using that set to the data, and then employ sophisticated computer algorithms to find a better parameter set. This process continues (“iterates”) until the difference between the current parameter set and the previous one is trivial. When this happens, the algorithm is said to have “converged” on a solution.

The most important part of this output is the message “Convergence criteria met.” If you do not receive this message, then do not interpret the results. If the number of iterations is large and there is no convergence, then there is something the matter with the data or the model is not identified (i.e., it is asking too much of the data). Check the stacked data set to make certain that it is correct. If the problem is with the model, start with a simple model and with a simple structure for the covariance matrix (TYPE = CS fits the simplest structure, compound symmetry). Once you have gotten the simple model to converge, then try a slightly more complicated model. Working you way up can help to isolate problem in which the model is asking too much of the data.

13.7.1.3 PROC MIXED output, section (3)

Because the R option was given in the REPEATED statement (see Section X.X and Table X.X), the procedure prints the estimate of the error covariance matrix here. Do not confuse the R matrix with a correlation matrix. In a complete failure of intellect, early developers of mixed models violated the major precepts of human factors engineering and denoted the error covariance matrix as R. The only reason to pay attention here is if you are uncertain of the error covariance

---

8Technically, R is the error covariance matrix for fixed effects.
structure and want to compare the structure that you just modeled to another structure. Ignore the phrase “... for Subject 1.” That is meant for complicated types of mixed models that do not concern us.

The next matrix gives the error correlation matrix. You must always examine this matrix. If the correlations are close to 0, then there is no need to perform repeated measures.

When you specify a form for the covariance matrix, then this section gives the parameters for the matrix. The Huynh-Feldt form that was specified by the TYPE = HF option in the REPEATED statement (see Table X.X) has four parameters for this problem. They values need not concern us, but the number of parameters becomes important when you compare different repeated measures models.

13.7.1.4 PROC MIXED output, section (4)

This is a very important section if you are comparing different models using the more advanced techniques of model comparisons described in Section X.X. Otherwise, they can be ignored.

13.7.1.5 PROC MIXED output, section (5)

This part of the output is seriously misplaced and horrendously mislabeled. The chi square ($\chi^2$) applies to the error covariance matrix and not to the overall model. It tests whether the parameters of the covariance matrix that you specified construct a matrix that predicts better than an error covariance matrix with a the same number on each diagonal and all off diagonals equal to 0. The null hypothesis is that the only difference in predictability between the two is due to sampling error. Under this condition, the $\chi^2$ statistic should be a random selection from a $\chi^2$ distribution where the degrees of freedom equal the number of parameters in the matrix structure that you specified less 1. If the $\chi^2$ is large and the probability of observing one larger than it is very small, then the matrix you specified is a significantly better predicted than the one with the same variance and no correlations.

In many cases, including the present one, this test is not important. If you specify a model in which all of the variances are equal, then the $\chi^2$ is a test that the correlations, on average, are not 0.

13.7.1.6 PROC MIXED output, section (6)

The section is analogous to the ANOVA table of effects in a least-squares GLM. The $F$ statistics, however, are not calculated as those in the least-squares GLM and may be unreliable in small samples. Usually, when the $p$ values for the $F$s are very small, as they are in this case for Area and Group*Area, they will not change when more precise tests are used, even with small samples. Hence, we can interpret these.

The main conclusion is that there is a significant interaction between the treatment and brain area. The treatment does alter mRNA expression, but its
CHAPTER 13. GLM: MULTIPLE DEPENDENT VARIABLES

eff ects on gene regulation depends on the brain area. As always, when there is a significant interaction, we ignore the significance for the main effects. Both Group and Area are significantly important. It is just that the magnitude of the effect of one depends on the level of the other.

13.7.1.7 PROC MIXED output section (7)
The final section of output is given in Figure 13.11 and consists of the parameter estimates that we requested by using the SUMMARY option on the REPEATED statement. Your software may print these automatically. Note how SAS presents these as if the variables were dummy codes of the effects.

13.8 Patterned covariance matrices*

In a mixed models analysis, it is possible to specify the structure of the error covariance matrix. Here, we list some of the more common types of structure. Let \( \Sigma_E \) denote the error covariance matrix, \( \sigma^2 \), the variance of a variable, and \( \rho \), the correlation between a pair of variables. A completely unstructured or unpatterned matrix for three variables will have the structure

\[
\Sigma_E = \begin{pmatrix}
\sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \rho_{13}\sigma_1\sigma_3 \\
\rho_{12}\sigma_1\sigma_2 & \sigma_2^2 & \rho_{23}\sigma_2\sigma_3 \\
\rho_{13}\sigma_1\sigma_3 & \rho_{23}\sigma_2\sigma_3 & \sigma_3^2 \\
\end{pmatrix}
\] (13.6)

We have already seen a matrix with compound symmetry

\[
\Sigma_E = \begin{pmatrix}
\sigma^2 & \rho\sigma^2 & \rho\sigma^2 \\
\rho\sigma^2 & \sigma^2 & \rho\sigma^2 \\
\rho\sigma^2 & \rho\sigma^2 & \sigma^2 \\
\end{pmatrix}
\] (13.7)

A covariance matrix that is isocorrelational has off diagonal elements that are equal to a single correlation, denoted here as \( \rho \), times the product of the standard deviation of the two variables. The variances, however, may differ. Hence, its generic form is

\[
\Sigma_E = \begin{pmatrix}
\sigma_1^2 & \rho\sigma_1\sigma_2 & \rho\sigma_1\sigma_3 \\
\rho\sigma_1\sigma_2 & \sigma_2^2 & \rho\sigma_2\sigma_3 \\
\rho\sigma_1\sigma_3 & \rho\sigma_2\sigma_3 & \sigma_3^2 \\
\end{pmatrix}
\] (13.8)

Note that a matrix with compound symmetry must be isocorrelational. Sometimes the isocorrelational matrix is called a matrix with compound symmetry but with heterogeneous variances.

When the repeated measures are measures over time, the error covariance matrix may take the form of an antedependence or autoregression. The general antedependence model is depicted in Figure 13.12 for three time points. The term antedependence signifies that any time point is dependent on the values of older time points. The figure depicts the case in which the antedependence
### Model Information

- **Data Set**: WORK.STACKED
- **Dependent Variable**: mRNA
- **Covariance Structure**: Huynh-Feldt
- **Subject Effect**: Rat
- **Estimation Method**: REML
- **Residual Variance Method**: None
- **Fixed Effects SE Method**: Model-Based
- **Degrees of Freedom Method**: Between-Within

### Class Level Information

<table>
<thead>
<tr>
<th>Class</th>
<th>Levels</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>2</td>
<td>Control Treatment</td>
</tr>
<tr>
<td>Area</td>
<td>3</td>
<td>1 2 3</td>
</tr>
</tbody>
</table>

### Dimensions

- **Covariance Parameters**: 4
- **Columns in X**: 12
- **Columns in Z**: 0
- **Subjects**: 20
- **Max Obs Per Subject**: 3

### Number of Observations

- **Number of Observations Read**: 60
- **Number of Observations Used**: 60
- **Number of Observations Not Used**: 0

### Iteration History

<table>
<thead>
<tr>
<th>Iteration</th>
<th>Evaluations</th>
<th>-2 Res Log Like</th>
<th>Criterion</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>217.37232891</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>199.54069410</td>
<td>0.00852116</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>199.05541195</td>
<td>0.00055154</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>199.02639616</td>
<td>0.00000460</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>199.02616506</td>
<td>0.00000000</td>
</tr>
</tbody>
</table>

Convergence criteria met.
Figure 13.10: Output from PROC MIXED for the repeated measures analysis of the mRNA expression data, part 2.

<table>
<thead>
<tr>
<th>Estimated R Matrix for Subject 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Row</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Estimated R Correlation Matrix for Subject 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Row</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Covariance Parameter Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parm</td>
</tr>
<tr>
<td>Var(1)</td>
</tr>
<tr>
<td>Var(2)</td>
</tr>
<tr>
<td>Var(3)</td>
</tr>
<tr>
<td>HF</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fit Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>-2 Res Log Likelihood</td>
</tr>
<tr>
<td>AIC (smaller is better)</td>
</tr>
<tr>
<td>AICC (smaller is better)</td>
</tr>
<tr>
<td>BIC (smaller is better)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Null Model Likelihood Ratio Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>DF</td>
</tr>
<tr>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Type 3 Tests of Fixed Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Num</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>2</td>
</tr>
</tbody>
</table>
Figure 13.11: Output from PROC MIXED for the repeated measures analysis of the mRNA expression data, part 3.

| Effect   | Group       | Area | Estimate | Error | DF  | t Value | Pr > |t| |
|----------|-------------|------|----------|-------|-----|---------|-------|---|
| Intercept|             |      | 10.5280  | 0.6059| 18  | 17.38   | <.0001|   |
| Group    | Control     |      | 1.4950   | 0.8569| 18  | 1.74    | 0.0981|   |
| Group    | Treatment   |      | 0        |       |     |         |       |   |
| Area     | 1           |      | -2.3170  | 0.5138| 36  | -4.51   | <.0001|   |
| Area     | 2           |      | 5.5990   | 0.5138| 36  | 10.90   | <.0001|   |
| Area     | 3           |      | 0        |       |     |         |       |   |
| Group*Area| Control     | 1    | -0.7790  | 0.7267| 36  | -1.07   | 0.2908|   |
| Group*Area| Control     | 2    | -2.3850  | 0.7267| 36  | -3.28   | 0.0023|   |
| Group*Area| Control     | 3    | 0        |       |     |         |       |   |
| Group*Area| Treatment   | 1    | 0        |       |     |         |       |   |
| Group*Area| Treatment   | 2    | 0        |       |     |         |       |   |
| Group*Area| Treatment   | 3    | 0        |       |     |         |       |   |
CHAPTER 13. GLM: MULTIPLE DEPENDENT VARIABLES

Figure 13.12: An antedependence model for three time points.

\[ T_1 \xrightarrow{\rho_{12}} T_2 \xrightarrow{\rho_{23}} T_3 \]

is only on the previous time point. Hence, the correlation between Time 1 and Time 2 is \( \rho_{12} \) and the correlation between Time 2 and Time 3 is \( \rho_{23} \). The correlation between Time 1 and Time 3 is the product of these two correlations or \( \rho_{12}\rho_{23} \). Hence, the error covariance matrix has the form

\[
\Sigma_E = \begin{pmatrix}
\sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \rho_{12}\rho_{23}\sigma_1\sigma_3 \\
\rho_{12}\sigma_1\sigma_2 & \sigma_2^2 & \rho_{23}\sigma_2\sigma_3 \\
\rho_{12}\rho_{23}\sigma_1\sigma_3 & \rho_{23}\sigma_2\sigma_3 & \sigma_3^2
\end{pmatrix}
\]  

(13.9)

This equation permits heterogeneous (i.e., different variances). An antedependence model may also have a single (aka homogeneous) variance or

\[
\Sigma_E = \begin{pmatrix}
\sigma_1^2 & \rho_{12}\sigma_1^2 & \rho_{12}\rho_{23}\sigma_1^2 \\
\rho_{12}\sigma_1^2 & \sigma_2^2 & \rho_{23}\sigma_2^2 \\
\rho_{12}\rho_{23}\sigma_1^2 & \rho_{23}\sigma_2^2 & \sigma_3^2
\end{pmatrix}
\]  

(13.10)

Autoregression is a type of antedependence in which the correlation between two time points is the same. Eliminate the subscripts “12” and “23” in Figure 13.12 and the antedependence model becomes autoregressive. The correlation between Time 1 and Time 2 equals \( \rho \) as does the correlation between Time 2 and Time 3. The correlation between Time 1 and Time 3 becomes \( \rho^2 \). Hence, for heterogeneous variances,

\[
\Sigma_E = \begin{pmatrix}
\sigma_1^2 & \rho\sigma_1\sigma_2 & \rho^2\sigma_1\sigma_3 \\
\rho\sigma_1\sigma_2 & \sigma_2^2 & \rho\sigma_2\sigma_3 \\
\rho^2\sigma_1\sigma_3 & \rho\sigma_2\sigma_3 & \sigma_3^2
\end{pmatrix}
\]  

(13.11)

and for homogeneous variances

\[
\Sigma_E = \begin{pmatrix}
\sigma^2 & \rho\sigma^2 & \rho^2\sigma^2 \\
\rho\sigma^2 & \sigma^2 & \rho\sigma^2 \\
\rho^2\sigma^2 & \rho\sigma^2 & \sigma^2
\end{pmatrix}
\]  

(13.12)

For completeness, let’s look at the form of the Type H matrix of Huynh & Feldt (19X.X). This has a different parameterization. For \( k \) dependent variables, there will be \( k \) parameters, \( \theta_1, \theta_2, \ldots, \theta_k \). There is one additional parameter applied to the variances that is denoted here as \( \lambda \). The form of the Type H matrix is

\[
\Sigma_E = \begin{pmatrix}
2\theta_1 + \lambda & \theta_1 + \theta_2 & \theta_1 + \theta_3 \\
\theta_1 + \theta_2 & 2\theta_2 + \lambda & \theta_2 + \theta_3 \\
\theta_1 + \theta_3 & \theta_2 + \theta_3 & 2\theta_3 + \lambda
\end{pmatrix}
\]  

(13.13)
A final class of patterns are those that fit spatial autocorrelation. The concept is analogous to antedependence and autoregression but applied to either two or three dimensional space instead of time. Spatial autocorrelation is a topic of current research in imaging and computational neuroscience (e.g., Eglen et al., 2008; Keil, Oros-Peusquens, & Shah, 2012) and takes considerable expertise in both biostatistics and computer technology. If your data are sampled from different regions in a small enough area in the brain that you suspect that spatial autocorrelation may be an issue, then you should consult with an expert on the topic.