Appendix A: Review of the General Linear Model

The general linear model is an important tool in many fMRI data analyses. As the name "general" suggests, this model can be used for many different types of analyses, including correlations, one-sample t-tests, two-sample t-tests, analysis of variance (ANOVA), and analysis of covariance (ANCOVA). This appendix is a review of the GLM and covers parameter estimation, hypothesis testing, and model setup for these various types of analyses.

Some knowledge of matrix algebra is assumed in this section, and for a more detailed explanation of the GLM, it is recommended to read Neter et al. (1996).

A.1 Estimating GLM parameters

The GLM relates a single continuous dependent, or response, variable to one or more continuous or categorical independent variables, or predictors. The simplest model is a *simple linear regression*, which contains a single independent variable. For example, finding the relationship between the dependent variable of mental processing speed and the independent variable, age (Figure A.1). The goal is to create a model that fits the data well and since this appears to be a simple linear relationship between age and processing speed, the model is $\mathbf{Y} = \beta_0 + \beta_1 \mathbf{X_1}$, where $\mathbf{Y}$ is a vector of length T containing the processing speeds for T subjects, β_0 describes where the line crosses the y axis, β_1 is the slope of the line and $\mathbf{X_1}$ is the vector of length T containing the ages of the subjects. Note that if β_0 is omitted from the model, the fitted line will be forced to go through the origin, which typically does not follow the trend of the data very well and so intercepts are included in almost all linear models.

Notice that the data points in Figure A.1 do not lie exactly in a line. This is because $\mathbf{Y}$, the processing speeds, are random quantities that have been measured with some degree of error. To account for this, a random error term is added to the GLM model

$$\mathbf{Y} = \beta_0 + \beta_1 \mathbf{X_1} + \boldsymbol{\epsilon}$$

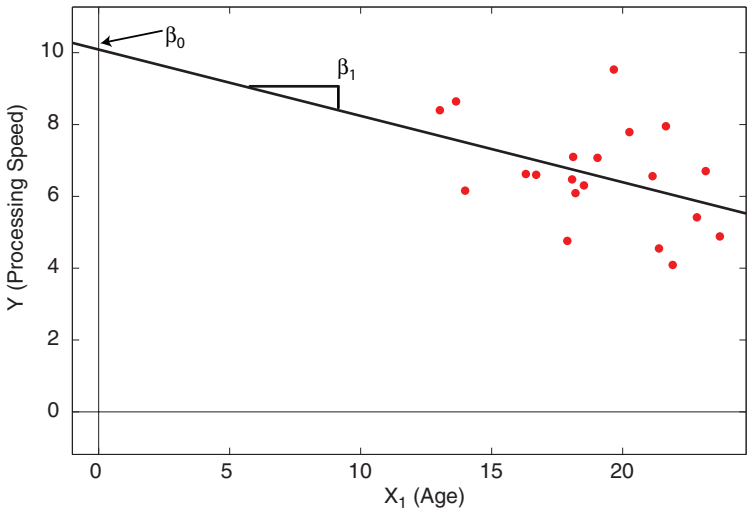

Figure A.1. Example of data and model for a simple linear regression. The intercept, β_0, is where the line crosses the *y* axis and the slope, β_1, describes the relationship between the dependent variable, mental processing speed, and the independent variable, age.

where ϵ is a random vector of length T that describes the distribution of the error between the true value of the dependent variable and the measured values that are obtained for the study. The standard assumption is that $\boldsymbol{\epsilon}$ is normally distributed such that the vector has a mean of 0 and a variance of σ^2. Further, any two elements of the error vector are assumed to be uncorrelated, $\text{Cor}(\epsilon_i, \epsilon_j) = 0$. This is typically written as $\boldsymbol{\epsilon} \sim N(0, \sigma^2 \mathbf{I})$, where N is the multivariate normal distribution and $\mathbf{I}$ is a $T \times T$ identity matrix, which only has 1s along the diagonal and 0s on the off diagonal.

The interpretation of the model follows: If we were to know the true values of β_0 and β_1, then for a given age, say 20 years old, the expected mean processing speed for this age would be $\beta_0 + \beta_1 \times 20$. If we were to collect a sample of processing speeds from 20 year olds, the distribution of the data would be normal with a mean of $\beta_0 + \beta_1 \times 20$ and a variance of σ^2. Although the mean processing speed would change for different age groups, the variance would be the same. The distribution for those with an age of 10 years would have a mean of $\beta_0 + \beta_1 \times 20$ and a variance of σ^2.

A.1.1 Simple linear regression

To find the estimates of the parameters β_0 and β_1, the method of *least squares* is used, which minimizes the squared difference between the data, $\mathbf{Y}$, and the estimates, $\hat{\mathbf{Y}} = \hat{\beta}_0 + \hat{\beta}_1 \mathbf{X}_1$. This difference is known as the residual, and this is denoted by $\mathbf{e} = \mathbf{Y} - \hat{\mathbf{Y}}$ and would be the horizontal distance between a point and the fitted line. The estimate of the error variance, σ^2, is given by $\hat{\sigma}^2 = \frac{\mathbf{e}'\mathbf{e}}{T-2}$, where T is the number

of data points. The quantity $T-2$ is the degrees of freedom for this model and consists of the amount of information going into the model (T data points) minus the number of parameters we had to estimate (two, β_0 and β_1). The line shown in Figure A.1 illustrates the least squares fit of the model to the data where, for a given age, the distribution of processing speed for that age is estimated to have a mean of $\hat{\beta}_0 + age \times \hat{\beta}_1$, and $\hat{\sigma}^2$ is the estimate of the variance of the data for that value of age.

Under the assumptions that the error has a mean of zero, constant variance and correlation of 0, the least squares estimates of the $\hat{\beta}_i$s have a nice property according to the Gauss Markov theorem (Graybill, 1961), which is that $\hat{\beta}$ is unbiased and has the smallest variance among all unbiased estimators of β. In other words if we were to repeat the experiment an infinite number of times and estimate $\hat{\beta}$ each time, the average of these $\hat{\beta}$s would be equal to the true value of β. Not only that, but the variance of the estimates of $\hat{\beta}$ is smaller than any other estimator that gives an unbiased estimate of β. When the assumptions are violated, the estimate will not have these properties, and Section A.3 will describe the methods used to handle these situations.

A.1.2 Multiple linear regression

It is possible to have multiple independent variables, $\mathbf{X_1}, \mathbf{X_2}, \ldots, \mathbf{X_p}$, in which case the GLM would be $\mathbf{Y} = \beta_0 + \beta_1 \mathbf{X_1} + \beta_2 \mathbf{X_2} + \cdots + \beta_p \mathbf{X_p} + \boldsymbol{\epsilon}$. The error term, $\boldsymbol{\epsilon}$, is distributed the same as before ($\boldsymbol{\epsilon} \sim N(0, \sigma^2 \mathbf{I})$) and each parameter, β_i, is interpreted as the effect of $\mathbf{X_i}$ controlling for all other variables in the model. So, for example, if age and gender were independent variables, the parameter estimate for age would be the relationship of age on processing speed, adjusting for gender or holding gender constant. Sometimes the parameters in a multiple linear regression are referred to as *partial regression coefficients* since they reflect the effect of one predictor controlling for all of the other predictors.

The multiple linear regression formula can be concisely expressed using matrix algebra as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \epsilon$$

where $\mathbf{X}$ is a $T \times p$ matrix with each column corresponding to an $\mathbf{X_i}$ and $\boldsymbol{\beta}$ is a column vector of length $p+1$, $\boldsymbol{\beta} = [\beta_0, \beta_1, \cdots, \beta_p]'$. The use of matrix algebra makes the derivation of $\hat{\boldsymbol{\beta}}$ easy. Since $\mathbf{X}$ isn't a square matrix, we can't solve the equation $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta}$ by premultiplying both sides by $\mathbf{X}^{-1}$, because only square matrices have inverses. Instead, if we first premultiply both sides of the equation by $\mathbf{X}'$, we have the so-called *normal equations*

$$\mathbf{X}'\mathbf{Y} = \mathbf{X}'\mathbf{X}\boldsymbol{\beta}$$

Table A.1. Two examples of rank deficient matrices

$$\begin{pmatrix} 1 & 0 & 7 \\ 1 & 0 & 7 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix} \qquad \begin{pmatrix} 1 & 0 & 2 \\ 1 & 0 & 2 \\ 1 & 1 & 5 \\ 1 & 1 & 5 \end{pmatrix}$$

It can be shown that any $\boldsymbol{\beta}$ that satisfies the normal equations minimizes the sum-of-squares of the residuals $\mathbf{e}'\mathbf{e}$, and thus it gives the least squares solution

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \tag{A.1}$$

assuming $\mathbf{X}'\mathbf{X}$ is invertible.

The estimate for σ^2 is the same as before, $\hat{\sigma}^2 = \frac{\mathbf{e}'\mathbf{e}}{T-(p+1)}$, where T is the number of rows in $\mathbf{X}$ and $p+1$ is the number of columns, resulting in $T-(p+1)$ the degrees of freedom for multiple linear regression.

In order for the inverse of $\mathbf{X}'\mathbf{X}$ to exist, $\mathbf{X}$ must have full column rank, which means no column is a linear combination of any of the other columns in the design matrix. In Table A.1, the matrix on the left-hand side is rank deficient since multiplying the first column by 7 yields the third column, and the matrix on the right is rank deficient since twice the first column plus three times the second equals the third column. If the design matrix is rank deficient there is not a unique solution for the parameter estimates. Consider the matrix on the left of Table A.1, and estimate the three corresponding parameters, β_1, β_2, β_3, when $\mathbf{Y} = [14\ 14\ 0\ 0]'$. It is easy to show that not only are $\beta_1 = 0, \beta_2 = 0$, and $\beta_3 = 2$ parameters that give an exact solution since

$$\begin{pmatrix} 1 & 0 & 7 \\ 1 & 0 & 7 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 2 \end{pmatrix} = \begin{pmatrix} 14 \\ 14 \\ 0 \\ 0 \end{pmatrix} \tag{A.2}$$

but $\beta_1 = 14, \beta_2 = 0$, and $\beta_3 = 0$ and an infinite number of other combinations will also perfectly fit the data.

A.2 Hypothesis testing

The previous section described how to obtain the estimates of the parameters $\beta_0, \beta_1, \ldots, \beta_p$ and σ^2, and this section describes how to carry out hypothesis tests on linear combinations, or contrasts, of the β_is. A row-vector of length $p+1$ is used to define the contrast to be tested. The simplest contrast tests a single parameter in the vector of parameters, $\boldsymbol{\beta}$. For example, if there were four parameters in the model, $[\beta_0, \beta_1, \beta_2, \beta_3]'$, then the contrast to test whether the first parameter, β_0 was different from 0, $H_0 : \beta_0 = 0$, would be $\mathbf{c} = [1\ 0\ 0\ 0]$, since $\mathbf{c}\boldsymbol{\beta} = \beta_0$. It is also possible to test

whether two parameters are different from each other. To test, $H_0 : \beta_2 = \beta_3$, which is equivalent to $H_0 : \beta_2 - \beta_3 = 0$, the contrast $\mathbf{c} = [0 \;\; 1 \;\; -1 \;\; 0]$ would be used. In both of these cases, the null hypothesis can be re-expressed as $H_0 : \mathbf{c}\boldsymbol{\beta} = 0$.

To test the hypothesis, the distribution of $\mathbf{c}\hat{\boldsymbol{\beta}}$ under the null assumption, that the contrast is equal to 0, must be known. It can be shown that the distribution of $\mathbf{c}\hat{\boldsymbol{\beta}}$ is normal with a mean of $\mathbf{c}\boldsymbol{\beta}$ and a variance of $\mathbf{c}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}'\sigma^2$, so under the null hypothesis, $\mathbf{c}\hat{\boldsymbol{\beta}} \sim N(0, \mathbf{c}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}'\sigma^2)$. Since we do not know the variance σ^2, we cannot use the normal distribution to carry out the hypothesis test. Instead, we use the t statistic

$$t = \frac{\mathbf{c}\hat{\boldsymbol{\beta}}}{\sqrt{\mathbf{c}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}'\hat{\sigma}^2}} \tag{A.3}$$

which, under the null, is distributed as a t-distribution with $T - (p + 1)$ degrees of freedom. A P-value for a one-sided alternative hypothesis, such as $H_A : \mathbf{c}\boldsymbol{\beta} > 0$ is given by $P(T_{T-(p+1)} \geq t)$, where $T_{N-(p+1)}$ is a random variable following a t-distribution with $T - (p + 1)$ degrees of freedom, and t is the observed test statistic. The P-value for a two-sided hypothesis test, $H_A : \mathbf{c}\boldsymbol{\beta} \neq 0$, is calculated as $P(T_{T-(p+1)} \geq |t|)$.

In addition to hypothesis testing of single contrasts using a t-statistic, one can also simultaneously test multiple contrasts using an F-test. For example, again using the model with four parameters, to test whether all of the βs are 0, $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$, one would specify a set of contrasts in the form of a matrix. Each row of the contrast corresponds to one of the four simultaneous tests, in this case that a particular β_i is 0, and looks like the following:

$$\mathbf{c} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{A.4}$$

The F-statistic is then given by

$$F = (\mathbf{c}\hat{\boldsymbol{\beta}})'[r\mathbf{c}\left(\widehat{\mathrm{Cov}}(\hat{\boldsymbol{\beta}})\right)\mathbf{c}']^{-1}(\mathbf{c}\hat{\boldsymbol{\beta}}) \tag{A.5}$$

where r is the rank of c and typically is equal to the number of rows in c. The F-statistic in Equation (A.5) is distributed as an F with r numerator and $T - (p+1)$ denominator degrees of freedom ($F_{r,T-(p+1)}$).

A.3 Correlation and heterogeneous variances

One of the important assumptions of the GLM, mentioned at the beginning of this appendix, is that the elements of the error vector, ϵ are uncorrelated, $\mathrm{Cor}(\epsilon_i, \epsilon_j) = 0$ for $i \neq j$ and that they all have the same variance, $\mathrm{Var}(\epsilon_i) = \sigma^2$ for all i. There are

many cases when this assumption is violated. For example, imagine that the dataset on age and processing speed included sets of identical twins; in this case, some individuals will be more similar than others. More relevant to fMRI, this can also occur when the dependent variable $\mathbf{Y}$ includes temporally correlated data. When this occurs, the distribution of the error is given by $\text{Cov}(\boldsymbol{\epsilon}) = \sigma^2\mathbf{V}$, where $\mathbf{V}$ is the symmetric correlation matrix and σ^2 is the varaince.

The most common solution to this problem is to *prewhiten* the data, or to remove the temporal correlation. Since a correlation matrix is symmetric and positive definite, the Cholesky decomposition can be used to find a matrix $\mathbf{K}$ such that $\mathbf{V}^{-1} = \mathbf{K}'\mathbf{K}$ (see Harville (1997) for more details on matrix decomposition). To prewhiten the data, $\mathbf{K}$ is premultiplied on both sides of the GLM to give

$$\mathbf{KY} = \mathbf{KX}\boldsymbol{\beta} + \mathbf{K}\boldsymbol{\epsilon} \tag{A.6}$$

Since the errors are now independent,

$$\text{Cov}(\mathbf{K}\epsilon) = \mathbf{K}\text{Cov}(\epsilon)\mathbf{K}' = \sigma^2\mathbf{I}$$

we can rewrite Equation (A.6) as

$$\mathbf{Y}^* = \mathbf{X}^*\boldsymbol{\beta} + \boldsymbol{\epsilon}^* \tag{A.7}$$

where $\mathbf{Y}^* = \mathbf{KY}$, $\mathbf{X}^* = \mathbf{KX}$, and $\boldsymbol{\epsilon}^* = \mathbf{K}\boldsymbol{\epsilon}$. Most important, since $\text{Cov}(\boldsymbol{\epsilon}^*) = \sigma^2\mathbf{I}$, the previously stated assumptions hold, and we can use least squares to estimate our parameters and their variances. The parameter estimates would be

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^{*'}\mathbf{X}^*)^{-1}\mathbf{X}^{*'}\mathbf{Y}^* \tag{A.8}$$

which can be also written as $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}$. The covariance of $\hat{\boldsymbol{\beta}}$ is given by

$$\widehat{\text{Cov}}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}^{*'}\mathbf{X}^*)^{-1}\hat{\sigma}^2 \tag{A.9}$$

or $\widehat{\text{Cov}}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}\mathbf{V}^{-1}\mathbf{X})^{-1}\hat{\sigma}^2$ and $\hat{\sigma}^2$ is estimated as shown earlier.

If the error terms are uncorrelated, $\text{Cor}(\boldsymbol{\epsilon}) = \mathbf{I}$, but the assumption of equal variances is violated, $\text{Var}(\epsilon_i) \neq \text{Var}(\epsilon_j)$, $i \neq j$, the variances are said to be heterogeneous, and the GLM is estimated as shown in Equations (A.8) and (A.9), with $\mathbf{K} = \text{diag}(1/\sigma_1, 1/\sigma_2, \ldots, 1/\sigma_T)$. The expression $\text{diag}(1/\sigma_1, \ldots, 1/\sigma_T)$ simply refers to a matrix with 0s on the off diagonal and $1/\sigma_1, 1/\sigma_2, \ldots, 1/\sigma_T$ along the diagonal. This is known as weighted linear regression. In both the prewhitening approach and the weighted linear regression approach, the necessary variance and covariance parameters are estimated from the data and then used to get the contrast estimates and carry out hypothesis testing.

A.4 Why "general" linear model?

The GLM is a powerful tool, since many different types of analyses can be carried out using it including: one-sample t-tests, two-sample t-tests, paired t-tests, ANOVA, and ANCOVA. The first section illustrated simple linear regression in the example where processing speed was modeled as a function of age. Figure A.2 shows some common analyses that are carried out using the GLM where the top example is the simplest model, a one-sample t-test. In this case, we have one group and are interested in testing whether the overall mean is 0. The design is simply a column of 1s and the contrast is $\mathbf{c} = [1]$.

The next design shown in Figure A.2 is a two-sample t-test, where the data either belong to group 1 (G1) or group 2 (G2). In the outcome vector, $\mathbf{Y}$, all G1 observations are at the beginning, and G2 observations follow. The design matrix has two columns, where the parameter for each column corresponds to the means for G1 and G2, respectively. The contrast shown tests whether the means of the two groups are equal, but it is also possible to test the mean of each group using the separate contrasts, $\mathbf{c} = [1\ 0]$ and $\mathbf{c} = [0\ 1]$. Note that there are alternative ways of setting up the design for a two-sample t-test, which are not illustrated in the figure. Two other examples of design matrices for the two-sample t-test are given as $\mathbf{X}_{T1}$ and $\mathbf{X}_{T2}$ in Equation (A.10).

$$\mathbf{X}_{T1} = \begin{pmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & 1 \end{pmatrix} \quad \mathbf{X}_{T2} = \begin{pmatrix} 1 & 1 \\ \vdots & \vdots \\ 1 & 1 \\ 1 & -1 \\ \vdots & \vdots \\ 1 & -1 \end{pmatrix} \tag{A.10}$$

In $\mathbf{X}_{T1}$, the first column models the mean of the *baseline* or unmodeled group mean. In this case, the mean of group 1 is not explicitly modeled and so the parameter corresponding to the first column would be the mean for group 1, and the parameter associated with the second column would be the difference in means between groups 1 and 2.

In the case of $\mathbf{X}_{T2}$ the first column corresponds to the overall mean of the data and the second column is the difference between the means of group 1 and group 2. It is often the case that there are multiple ways of setting up a design matrix, so it is important to understand what the parameters for the columns of the design correspond to. For $\mathbf{X}_{T1}$, for example, $\mathbf{X}_{T1}\boldsymbol{\beta}$ would yield the vector, $\hat{\mathbf{Y}} = [\beta_0, \beta_0, \ldots, \beta_0, \beta_0 + \beta_1, \ldots, \beta_0 + \beta_1]'$, and so it is clear to see that β_0 corresponds to the mean of group 1 and $\beta_0 + \beta_1$ is the mean for group 2, hence β_1 is the difference between the two means. Similarly, $\mathbf{X}_{T2}\boldsymbol{\beta}$ gives a value of $\beta_0 + \beta_1$ for the group 1 entries and $\beta_0 - \beta_1$ for the group 2 entries, meaning β_0 is the overall mean and β_1 is the difference in means between the two groups.

Test Description	Order of data	Xβ	Hypothesis Test
One-sample t-test. 6 observations	G_1 G_2 G_3 G_4 G_5 G_6	$\begin{pmatrix}1\\1\\1\\1\\1\\1\end{pmatrix}\begin{pmatrix}\beta_1\end{pmatrix}$	H_0: Overall mean=0 $H_0: \beta_1 = 0$ $H_0: c\beta = 0$ $c = [1]$
Two-sample t-test. 5 subjects in group 1 (G1) and 5 subjects in group 2 (G2)	$G1_1$ $G1_2$ $G1_3$ $G1_4$ $G1_5$ $G2_1$ $G2_2$ $G2_3$ $G2_4$ $G2_5$	$\begin{pmatrix}1&0\\1&0\\1&0\\1&0\\1&0\\0&1\\0&1\\0&1\\0&1\\0&1\end{pmatrix}\begin{pmatrix}\beta_{G1}\\\beta_{G2}\end{pmatrix}$	H_0: mean of G1 different from G2 $H_0: \beta_{G1} - \beta_{G2} = 0$ $H_0: c\beta = 0$ $c = [1\ \ -1]$
Paired t-test. 5 paired measures of A and B.	A_{S1} B_{S1} A_{S2} B_{S2} A_{S3} B_{S3} A_{S4} B_{S4} A_{S5} B_{S5}	$\begin{pmatrix}1&1&0&0&0&0\\-1&1&0&0&0&0\\1&0&1&0&0&0\\-1&0&1&0&0&0\\1&0&0&1&0&0\\-1&0&0&1&0&0\\1&0&0&0&1&0\\-1&0&0&0&1&0\\1&0&0&0&0&1\\-1&0&0&0&0&1\end{pmatrix}\begin{pmatrix}\beta_{diff}\\\beta_{S1}\\\beta_{S2}\\\beta_{S3}\\\beta_{S4}\\\beta_{S5}\end{pmatrix}$	H_0: A is different from B $H_0: \beta_{diff} = 0$ $H_0: c\beta = 0$ $c = [1\ 0\ 0\ 0\ 0\ 0]$
Two way ANOVA. Factor A has two levels and factor B has 3 levels. There are 2 observations for each A/B combination.	$A1B1_1$ $A1B1_2$ $A1B2_1$ $A1B2_2$ $A1B3_1$ $A1B3_2$ $A2B1_1$ $A2B1_2$ $A2B2_1$ $A2B2_2$ $A2B3_1$ $A2B3_2$	$\begin{pmatrix}1&1&1&0&1&0\\1&1&1&0&1&0\\1&1&0&1&0&1\\1&1&0&1&0&1\\1&1&-1&-1&-1&-1\\1&1&-1&-1&-1&-1\\1&-1&1&0&-1&0\\1&-1&1&0&-1&0\\1&-1&0&1&0&-1\\1&-1&0&1&0&-1\\1&-1&-1&-1&1&1\\1&-1&-1&-1&1&1\end{pmatrix}\begin{pmatrix}\beta_{mean}\\\beta_{A1}\\\beta_{B1}\\\beta_{B2}\\\beta_{A1B1}\\\beta_{A1B2}\end{pmatrix}$	F-tests for all contrasts H_0: Overall mean=0 $H_0: \beta_{mean} = 0$ $H_0: c\beta = 0$ $c = [1\ 0\ 0\ 0\ 0\ 0]$ H_0: Main A effect=0 $H_0: \beta_{A1} = 0$ $H_0: c\beta = 0$ $c = [0\ 1\ 0\ 0\ 0\ 0]$ H_0: Main B effect = 0 $H_0: \beta_{B1} = \beta_{B2} = 0$ $H_0: c\beta = 0$ $c = \begin{bmatrix}0&0&1&0&0&0\\0&0&0&1&0&0\end{bmatrix}$ H_0: A/B interaction effect=0 $H_0: \beta_{A1B1} = \beta_{A1B2} = 0$ $H_0: c\beta = 0$ $c = \begin{bmatrix}0&0&0&0&1&0\\0&0&0&0&0&1\end{bmatrix}$

Figure A.2. Examples of GLM models for popular study designs including: One-sample *t*-test, two-sample *t*-test, paired *t*-test, and two-way ANOVA. The first column describes the model, the second column describes how the data are ordered in the outcome vector, the third column shows the design matrix, and the last column illustrates the hypothesis tests and corresponding contrasts. Note, in the ANOVA example *F*-tests are used for all contrasts, whereas *t*-tests are used for the other examples.

The paired t-test is the third example in Figure A.2, where there are N groups of paired observations. For example, we could have N subjects scanned on two sessions and want to compare session 2 to session 1. In the outcome vector, $\mathbf{Y}$, observations are ordered by subject, session 1 followed by session 2. The first column of the design matrix is modeling the difference and the last N columns of the design matrix are modeling subject-specific means. By adjusting for the subject-specific means, the difference refers to the difference in the *centered* or demeaned pairs of data points. To test the paired difference, the contrast only includes the first parameter, the rest of the parameters related to the subject-specific means are considered "nuisance," since we do not typically test them, but they are necessary to include in the model to pick up extra variability due to each subject having a different mean.

The last example illustrates a two-way ANOVA, with two levels for the first factor and three levels for the second factor. There are a couple of ways to set up this model, but the one illustrated here is a *factor effects* setup and is used when the interest is in testing the typical ANOVA hypotheses of overall mean, main effects, and interaction effects. In general, the format used to create the regressors is as follows: For each factor the number of regressors is equal to one less than the number of levels for that factor. So our first factor, call it A, will have one regressor associated with it and the second factor, call it B, will have two. Each regressor is modeling the difference between a level of the factor to a baseline level. For example, the second column of X in the ANOVA panel of Figure A.2 is the regressor for factor A and takes a value of 1 for rows corresponding to level 1 of A (A1), and since the second level is the reference level, all rows corresponding to A2 are -1. The third and fourth columns are the regressors for factor B and the third level, B3, is the reference so both regressors are -1 for those corresponding rows. The third regressor compares B1 to B3, so it is 1 for level B1 and 0 for level B2. The fourth regressor compares B2 to B3 and so it is 0 for B1 and 1 for B2. The last two columns make up the interaction and are found by multiplying the second and third and second and fourth columns, respectively. All contrasts are tested using an F-test, since this is standard for ANOVA. To test the main effects, we simply include a contrast for each regressor corresponding to that factor, and to test the interaction we would include a contrast for each regressor corresponding to an interaction. The other option is to use a *cell means* approach, where we simply have six regressors, one for each of the six cells of the 2×3 ANOVA. It is an extension of the two-sample t-test model shown in Figure A.2 and is more convenient when we are interested in testing hypothesis that compare the means between cells of the ANOVA.

It should be noted that in some cases it is possible to use a linear regression model when there are repeated measures. For example, the two-sample t-test can be thought of as a one-way ANOVA with two levels and repeated measures across the levels. In a similar fashion, the two-factor ANOVA model in the bottom panel of Figure A.2 can be extended to a repeated measures case, where measures are repeated for *all* factors in the model, say a subject is studied before and after a treatment (factor A)

for three types of memory tasks (factor B). In this case, the single mean column would be broken up into separate subject means, and these columns of the design matrix would be treated as nuisance. A very important note when using the linear model for repeated measures ANOVA designs is that it *only* works in the case when the measures are *balanced* across the factor levels. So, for example, if a subject was missing measurements for the second and third levels of factor B after treatment, this linear regression approach cannot be used. In cases such as this, more complicated models and estimation strategies are necessary to achieve appropriate test statistics and hypothesis test results.