9.07 Introduction to Statistics for Brain and Cognitive Sciences
Emery N. Brown

Lecture 16: Analysis of Variance

I. Objectives

Understand analysis of variance as a special case of the linear model.

Understand the one-way and two-way ANOVA models.

II. Analysis of Variance

The analysis of variance is a central part of modern statistical theory for linear models and experimental design. It represents another important contribution of Fisher to statistical theory.

A. Motivation

To motivate the analysis of variance framework, we consider the following example.

Example 16.1 Age and Drug-Related Effects on Cognitive Performance. A new drug designed to enhance cognitive performance is ready for testing in animals. Suppose we have a group of young and a group of old rats for the test sets. Each group contains 12 animals and each group is divided into three subgroups, A, B and C. For each group, subgroup A is a control group, Subgroup B receives dose level one of the new drug and subgroup C receives dose level two. Dose level two is twice the dose of dose level one in mg/kg. The average execution time in minutes of a previously well-learned binary choice task for each animal in each group is measured over 3 repetitions. The results from the experiment are reported in Table 16.1. Is there a dose dependent effect of the drug on performance? Is the performance effect different for different age groups?

		Drug Level		
		Control	Dose One	Dose Two
Age	Young	56, 62, 57, 72	64, 34, 64, 41	33, 37, 40, 16
	Old	62, 72, 61, 91	64, 48, 34, 63	17, 21, 49, 54

Table 16.1 Execution Times of Age and Drug-Related Effects on Performance

A boxplot of the data by drug level is shown in Fig. 16.1. There does seem to be an apparent increase in performance (decrease in execution time) as a function of drug level.

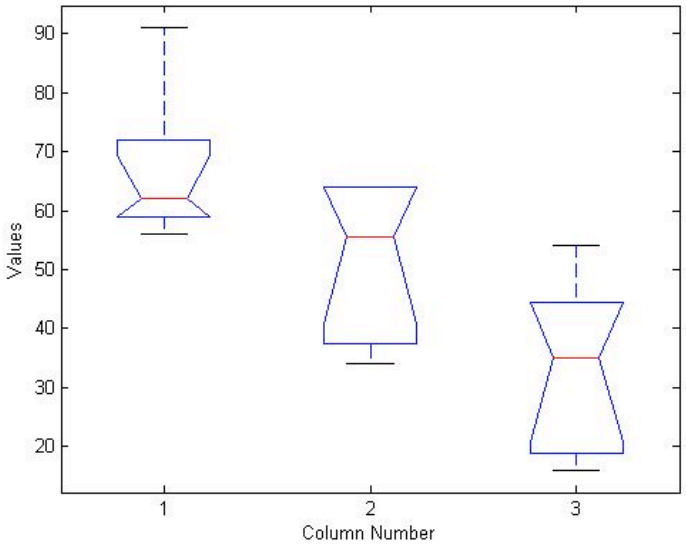

Figure 16.1. Box plots of the performance data by drug dose level.

B. Analysis of Variance Model

To formulate a statistical framework for **Example 16.1**, we first assume that there is no age effect. We can then view the problem as a special case of the simple regression model in which the regressor or covariate has three levels: control, dose level one and dose level two. We then ask, can the variance in the performance data be explained by taking account of drug level in the analysis? This is an **analysis of variance (ANOVA)**. In the regression problem, we studied how the variance in the data could be explained by the regressors and we summarized the results in an ANOVA table. Now, in ANOVA, we consider the case in which the regressors have discrete levels.

C. One-Way Analysis of Variance

We consider data y_{ij} divided into $i = 1,\ldots,I$ groups and having $j = 1,\ldots,J_i$ subjects per group. We model the data as

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \tag{16.1}$$

where μ is the mean level, α_i is the effect of the i^{th} level and ε_{ij} is random error. We assume

i) $E[y_{ij} \mid \text{effect } i] = \mu + \alpha_i$.

ii) The effect levels are fixed and unknown.

iii) The ε_{ij} are independent Gaussian random variables with mean zero and variance σ^2.

The model in Eq. 16.1 is termed the **one-way ANOVA model**. To develop a regression formulation of the one-way ANOVA model, let

$$\beta_i = \mu + \alpha_i \tag{16.2}$$

and take

$$Y^T = (y_{11}, y_{12}, ..., y_{1J_1}; y_{21}, y_{22}, ..., y_{2J_2}, ..., y_{I1}, y_{I2}, ..., y_{IJ_I}) \tag{16.3}$$

and define

$$X = \begin{array}{c} \begin{matrix} X_1 & X_2 & X_3 & ... & X_I \end{matrix} \\ \begin{bmatrix} 1 & 0 & 0 & ... & 0 \\ 1 & 0 & 0 & ... & 0 \\ ... & ... & ... & ... & ... \\ 1 & 0 & 0 & ... & 0 \\ -- & -- & -- & -- & -- \\ 0 & 1 & 0 & ... & 0 \\ 0 & 1 & 0 & ... & 0 \\ ... & ... & ... & ... & ... \\ 0 & 1 & 0 & ... & 0 \\ -- & -- & -- & -- & -- \\ ... & ... & ... & ... & ... \\ -- & -- & -- & -- & -- \\ 0 & 0 & 0 & ... & 1 \\ 0 & 0 & 0 & ... & 1 \\ ... & ... & ... & ... & ... \\ 0 & 0 & 0 & ... & 1 \end{bmatrix} \end{array} \tag{16.4}$$

$\beta' = (\beta_1, ..., \beta_I)$, where X_i is a 1 if the subject is in group i and 0 otherwise, for $i = 1, ..., I$. We obtain as in the case of multiple regression, the linear model

$$Y = X\beta + \varepsilon \tag{16.5}$$

where

$$\varepsilon^T = (\varepsilon_{11}, \varepsilon_{12}, ..., \varepsilon_{1J_1}; \varepsilon_{21}, \varepsilon_{22}, ..., \varepsilon_{2J_2}; ,,,; \varepsilon_{I1}, \varepsilon_{I2}, ..., \varepsilon_{IJ_I}). \tag{16.6}$$

Proceeding as in the case of the regression models, it follows that the likelihood is

$$L(\beta \mid Y) = (\frac{1}{2\pi\sigma^2})^{I\sum_{i=1}^{I} J_i} \exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{I}\sum_{j=1}^{J_I}(Y_{ij}-\beta_i)^2\} \tag{16.7}$$

and the log likelihood is

$$\begin{aligned}\log L(\beta \mid Y) &= -I\sum_{i=1}^{I} J_i \log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{I}\sum_{j=1}^{J_i}(Y_{ij}-\beta_i)^2 \\ &= -I\sum_{i=1}^{I} J_i \log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(Y-X\beta)^T(Y-X\beta).\end{aligned} \tag{16.8}$$

If we differentiate the log likelihood with respect to $\beta,$ and set the derivative equal to zero, we obtain the normal equations

$$(X^T X)\beta = X^T Y \tag{16.9}$$

where

$$(X^T X) = \begin{bmatrix} J_1 & & & \\ & J_2 & & \\ & & & \\ & & & J_I \end{bmatrix} \quad X^T Y = \begin{bmatrix} J_1 & \bar{Y}_1 \\ J_2 & \bar{Y}_1 \\ & \\ & \\ J_I & \bar{Y}_I \end{bmatrix} \tag{16.10}$$

The inverse of $X^T X$ is simply the diagonal matrix with i^{th} diagonal element $J_i^{-1}.$ It follows that the maximum likelihood estimates of the β_i are

$$\beta_i = \bar{Y}_i = J_i^{-1}\sum_{j=1}^{J_i} Y_{ij} \tag{16.11}$$

for $i = 1,...,I$ and that upon differentiating with respect to σ^2 we obtain, just as in the case of the multiple regression model,

$$\begin{aligned}\hat{\sigma}^2_{ML} &= (I\sum_{j=1}^{I} J_i)^{-1}(Y-X\hat{\beta})^T(Y-X\hat{\beta}) \\ &= (I\sum_{j=1}^{I} J_i)^{-1}\sum_{i=1}^{I}\sum_{j=1}^{J_i}(Y_{ij}-\hat{\beta}_i)^2.\end{aligned} \tag{16.12}$$

Again, we will use an unbiased estimate of the residual mean-squared error instead of the maximum likelihood estimate in Eq. 16.12. The Pythagorean relation for the One-Way Analysis of Variance model expressed in the ANOVA table is

Source	Sum of Squares	Degrees of Freedom	Mean Square
Between Groups	$\sum_{i=1}^{I} J_i(\bar{Y}_i - \bar{Y})^2$	$I-1$	MS_B^2
Within Groups	$\sum_{i=1}^{I}\sum_{j=1}^{J}(Y_{ij} - \bar{Y}_i)^2$	$\sum_{i=1}^{I}(J_i - 1) = n - I$	MS_w^2
Mean	$n\bar{Y}^2$	1	
Total	$\sum_{i=1}^{I}\sum_{j=1}^{J} Y_{ij}^2$	$\sum_{i=1}^{I} J_i = n$	

where $\bar{Y} = (\sum_{i=1}^{I} J_i)^{-1} \sum_{i=1}^{I} J_i \bar{Y}_i.$ Notice that the Total Sum of Squares about the mean $TSS = \sum_{i=1}^{I}\sum_{j=1}^{J} Y_{ij}^2 - n\bar{Y}^2$.

Remark 16.1. The F- statistic to test the null hypothesis $H_0 : \beta_1 = \beta_2 = ... = \beta_I = 0$ is

$$F = \frac{MS_B^2}{MS_w^2} \tag{16.13}$$

which we compare with the $1-\delta$ quantile of the F- distribution on $I-1$ and $n-I$ degrees of freedom.

Remark 16.2. The between groups sum of squares is the analog of the explained or regression sum of squares and the within groups sum of squares is the analog of the residual or error sum of squares for the simple and multiple regression problems. Notice that first three rows in ANOVA table are orthogonal.

Remark 16.3. If we return to the original model

$$E[y_{ij}] = \mu + \alpha_i = \beta_i \tag{16.14}$$

it follows that

$$\hat{\alpha}_i = \bar{Y}_i - \bar{Y}. \tag{16.15}$$

That is, the $\hat{\alpha}_i$'s are estimates of the deviances of the group means from the overall mean. Hence, we see that $H_0 : \beta_1 = \beta_2 = ... = \beta_I = 0$ is equivalent to the statement that the data are simply explained by the overall mean.

Example 16.1 (continued). To apply the one-way ANOVA model to this problem, we collapse Table 16.1 into 3 groups according to drug level. Hence, $I = 3$ and $J_i = 8,$ for $i = 1, 2, 3.$ We are interested first in the null hypothesis $H_0 : \beta_1 = \beta_2 = \beta_3 = 0,$ which means there is no difference in cognitive performance between the different dose levels of the drug. The ANOVA table is

Source	SS	DF	MS	F	p-value
Between Groups	$4,434.25$	2	$2,217.13$	12.52	0.0003
Within Groups	$3,717.75$	21	177.04		
Mean	$61,206.00$	1			
Total	$69,358.00$	24			

Again note that the $TSS = 8,152 = 69,358.00 - 61,206.00$ and that it has 23 degrees of freedom. Based on this analysis, we conclude that the drug does have an effect on performance.

D. Two-Way Analysis of Variance

To assess the effect of both age and drug level on performance, we require a **two-way analysis of variance model**. A two-way analysis of variance model allows us to assess the extent two which two factors may be used to describe variance in a response or independent variable. If we have a two-way classification with I rows and J columns, with K observations $Y_{ijk}, k = 1, ..., K$ in each cell, then the usual two-way fixed-effects analysis of variance model is

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk} \tag{16.16}$$

for $i = 1, ..., I,\ j = 1, ..., J,\ k = 1, ..., K,$ where μ is the mean effect, α_i is the i^{th} row effect, β_j is the j^{th} column effect and γ_{ij} is the interaction between the i^{th} row and the j^{th} column effect and ε_{ijk} are independent zero mean Gaussian random variables with mean zero and variance σ^2. In Example 16.4 $I = 2$ age groups $J = 3$ drug dose levels, there are six interaction terms and $K = 8$ in each cell (Table 16.1). We require an additional assumption to make the model parameters estimable namely,

$$\sum_{i=1}^{I} \alpha_i = \sum_{j=1}^{J} \beta_j = \sum_{i=1}^{I} \gamma_{ij} (\text{for all } j) = \sum_{j=1}^{J} \gamma_{ij} (\text{for all } i) = 0. \tag{16.17}$$

By carrying out an analysis similar to the one for the **One-Way ANOVA** it is possible to show that the ANOVA table for this model is

Source	Sum of Squares	Degrees of Freedom	Mean Square
Rows	$JK\sum_{i=1}^{I}(\bar{Y}_{i..}-\bar{Y})^2$	$I-1$	MS_R^2
Columns	$IK\sum_{j=1}^{J}(Y_{.j.}-\bar{Y})^2$	$J-1$	MS_c^2
Interaction	$K\sum_{i=1}^{I}\sum_{j=1}^{J}(Y_{ij.}-\bar{Y}_{i..}-\bar{Y}_{.j.}+\bar{Y})^2$	$(I-1)(J-1)$	MS_{RC}^2
Residual	by subtraction	$IJ(K-1)$	s^2
Mean	$IJK\bar{Y}^2$	1	
Total	$\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}Y_{ijk}^2$	IJK	

where

$$
\begin{aligned}
Y_{i..} &= (JK)^{-1}\sum_{j=1}^{J}\sum_{k=1}^{K}Y_{ijk}, \quad i^{\text{th}} \text{ row mean} \\
Y_{.j.} &= (IK)\sum_{i=1}^{I}\sum_{k=1}^{K}Y_{ijk}, \quad j^{\text{th}} \text{ column mean} \\
Y_{ij.} &= (K)^{-1}\sum_{k=1}^{K}Y_{ijk}, \quad i^{\text{th}} \text{ cell mean} \\
\bar{Y} &= (IJK)^{-1}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}Y_{ijk}, \text{ overall mean}
\end{aligned}
\tag{16.18}
$$

The standard null hypotheses and F- tests are

$$
\begin{aligned}
&H_0: \text{all } \alpha_i = 0 \quad F = MS_R^2/s^2 \quad \text{compared with } F[(I-1), IJ(K-1)] \\
&H_0: \text{all } \beta_j = 0 \quad F = MS_c^2/s^2 \quad \text{compared with } F[(J-1), IJ(K-1)] \\
&H_0: \text{all } \gamma_{ij} = 0 \quad F = MS_{Rc}^2/s^2 \quad \text{compared with } F[(I-1)(J-1), IJ(K-1)]
\end{aligned}
\tag{16.19}
$$

Example 16.1 (continued). The partially completed two-way ANOVA table for this problem is

Source	Sum of Squares	Degrees of Freedom
Rows	150.00	1
Columns	4,434.25	2
Interactions	72.75	2
Residual	3,495.00	18
Mean	61,206.00	1
Total	69,358.00	24

One of the problems in **Homework Assignment 10** is to complete this ANOVA table and determine if there is an effect on performance of age, of drug does level and if there is an interaction between age and drug dose level.

Remark 16.4. We can conduct an analysis of the residuals for this ANOVA model (See Draper and Smith, 1981).

Remark 16.5. The row, column, interaction effects and the residuals are all orthogonal. The effects are assumed to be fixed. Another formulation of the ANOVA models views either the column or row effects as random variables. In this case, we have random effects or mixed effects ANOVA models.

Remark 12.6. If we have a model in which there is a mixture of discrete and continuous explanatory variables, then the analysis of variance becomes the **analysis of covariance**. It is essentially the combination of the regression and ANOVA models. Similarly, when there are missing observations in an ANOVA table, the regression formulation is key. In these cases, the orthogonality is often lost.

Remark 16.7. If we have independent identically distributed Gaussian errors and the mean of each y_i is a nonlinear function of the corresponding x_i such as

$$y_i = g(\beta, x_i) + \varepsilon_i \tag{16.20}$$

then we still have a Gaussian likelihood $\beta.$ However, to estimate the parameter requires a nonlinear optimization routine to maximize the likelihood. Notice that in the case of the fMRI model we have

$$Y = X(\theta)\beta + \varepsilon \tag{16.21}$$

where $\theta = (\alpha, \tau, \delta)$ and α is the exponent in the gamma function, τ is the time-constant for the decline of the hemodynamic response and δ is the delay. Here, we can split the maximization

of the likelihood into linear steps and nonlinear steps. That is, if the Newton's procedure for maximizing the likelihood gives at iteration k

$$\theta^{(k)} = \theta^{(k-1)} - \nabla^2 \log L'(\theta^{(k-1)})^{-1} \nabla \log L'(\theta^{(k-1)}), \tag{16.22}$$

then the corresponding estimate of β at iteration k is

$$\hat{\beta}^{(k)} = [X^T(\theta^{(k)})X(\theta^{(k)})]X^T(\theta^{(k)})Y. \tag{16.23}$$

where $\log L'(\theta)$ is the concentrated log likelihood, i.e. the log likelihood written in terms of θ after solving for or "concentrating out" β. We did this for our maximum likelihood analysis of the gamma distribution in **Lecture 9**. Every time we estimate σ^2 by maximum likelihood for a Gaussian model we first solve for the estimate of the mean (**Lecture 9**) or the regression parameters (**Lectures 14** and **15**). This approach is not only computationally efficient, but also as $k \to \infty$ for the Newton's iteration

$$\theta^{(k)} \to \theta^{(\infty)} = \hat{\theta}_{ML} \tag{16.24}$$

and

$$\hat{\beta}_{ML} = [X^T(\hat{\theta}_{ML})X(\theta_{ML})]^{-1}X^T(\hat{\theta}_{ML})Y. \tag{16.25}$$

Remark 16.8. As we will see in **Lecture 17**, an important generalization of the linear regression models, we have studied here is the **generalized linear model (GLM)**. This will allow us to have dependent variables $(y_i\text{'s})$ in our regression models that are non-Gaussian and the computations can be carried out as iteratively reweighted least squares. Under this procedure, these models will be estimated by maximum likelihood.

Remark 16.9. The approach we have taken of viewing the simple linear regression, the multiple linear regression and the ANOVA models in a similar framework is termed the **general linear model**. It is important to appreciate the difference between this and the **generalized linear model** we will discuss in **Lecture 11**. It is also important to realize that the so-called general linear model and the **Statistical Parametric Map (SPM)** used by Carl Friston and colleagues at University College to analyze fMRI data is just a multiple linear regression model.

Remark 16.10. The time domain time-series methods we will derive in **Lecture 18**, will follow directly from writing down a linear regression model of the form

$$x_t = \alpha_0 + \sum_{j=1}^{p} \alpha_j x_{t-j} + \varepsilon_t. \tag{16.26}$$

This is an autoregression model of order p. The $x_t\text{'s}$ appear on both the left and right sides of the equation. Hence, they are no longer fixed constants.

Remark 16.11. A key problem in functional neuroimaging data analysis is that of multiple hypothesis tests. It comes about because every voxel is analyzed independently. The major

problem with multiple hypothesis tests is that there is a non-trivial finite probability of rejecting the null hypothesis. Various correction approaches such as Bonferroni corrections and false discovery rate are used.

III. Summary

We have shown that the simple linear regression, the multiple linear regression and the ANOVA models can all be derived in a unified likelihood framework. Most importantly, it provides a way of thinking about stimulus response experiments, which as we point out in the **Introductory Lecture**, is a central paradigm in neuroscience. The linear model framework is the crucial stepping stone to the advanced methods we will develop in the balance of the course.

Acknowledgments

I am grateful to Uri Eden for help with the computations and to Julie Scott for technical assistance.

References

DeGroot MH, Schervish MJ. *Probability and Statistics*, 3rd edition. Boston, MA: Addison Wesley, 2002.

Draper NR, Smith H. *Applied Regression Analyisis*, 2nd edition. New York: Wiley, 1981.

Rice JA. *Mathematical Statistics and Data Analysis*, 3rd edition. Boston, MA, 2007.

Rosner B. *Fundamentals of Biostatistics*, 6th edition. Boston, MA: Duxbury Press, 2006.

Ryan TP. *Modern Regression Methods*. New York: John Wiley & Sons, Inc., 1997.