Chapter 2

ANOVA Approaches to Longitudinal Data

There are two classical approaches to the analysis of longitudinal data, the first, called variously univariate mixed-model, split-plot, or repeated-measures ANOVA, and the second based on multivariate ANOVA (MANOVA). Both models assume interval measurement and normally distributed errors that are homogeneous across groups. In some cases, normality and homogeneity of variance can be brought about through transformation (*e.g.*, natural log transformation). For both models, the primary focus is on comparison of group means, and neither model is informative about individual growth curves

(*i.e.*, subject-specific trends). Furthermore, the time-points are assumed to be fixed across subjects (either evenly or unevenly spaced) and are treated as a classification variable in the ANOVA or MANOVA model. This precludes analysis of unbalanced designs in which different subjects are measured on different occasions. Both models are based on least squares estimation and are therefore adversely affected by outliers and missing data. While the ANOVA model can handle some missing data (*i.e.*, there are methods for unbalanced ANOVA), the MANOVA model cannot handle any missing data. In terms of the variance-covariance structure for the responses ($\boldsymbol{y}_i$), the ANOVA model assumes compound symmetry (*i.e.*, equal variances and covariances over time), whereas the MANOVA model makes no assumption regarding the specific form of the variance-covariance structure. While this is an important advantage of MANOVA over ANOVA, it is tempered by the larger limitation of requiring complete data for all subjects in the MANOVA model. As such, application of MANOVA must follow deletion of all subjects without complete data, which is essentially a completer analysis, and is prone to substantial bias in that the composition of subjects that complete the study can be quite different from the composition of the subjects at the time of randomization.

In the following sections of this chapter, we describe the univariate ANOVA

model for within-subject designs in detail. In chapter 3, we present the alternative multivariate approach. While these approaches are no longer recommended for routine application (if at all), they are important in that they fix ideas for the development of the more modern and advanced methods that are the primary focus of this book.

2.1 Single-sample Repeated Measures ANOVA

In the single-sample case, the model is referred to as the randomized block ANOVA. In this case, we have no intervention or group effects, but are simply using the model to characterize rates of change over time. In the multiple-sample case, the model is referred to as a split-blocks or split-plot ANOVA. In the social and behavioral sciences literature, the model is often referred to as a "Repeated Measures ANOVA." Returning to the single-sample case, where we have $i = 1, \ldots, N$ subjects and $j = 1, \ldots, n$ measurement occasions, the randomized blocks ANOVA is given by the linear model,

$$y_{ij} = \mu + \pi_i + \tau_j + e_{ij} \ , \tag{2.1}$$

where

- μ = the grand mean,

- π_i = the individual difference component for subject i, which is assumed to be constant over time,
- τ_j = the effect of time, assumed to be the same for all subjects,
- and e_{ij} = the error for subject i on occasion j

For the purpose of parameter estimation and hypothesis testing, we assume that the random components are distributed

- $\pi_i \sim N(0, \sigma_\pi^2)$ between-subjects variance
- $e_{ij} \sim N(0, \sigma_e^2)$ within-subjects variance

2.1.1 Data layout

Consider the following design:

	timepoint			
subject	1	2	...	n
1	y_{11}	y_{12}	...	y_{1n}
2	y_{21}	y_{22}	...	y_{2n}
.	.	.	...	.
.	.	.	...	.
N	y_{N1}	y_{N2}	...	y_{Nn}

In this design there is one observation per cell. The design is similar to a randomized block design with subjects as blocks. In the simple case of $n = 2$, it is identical to paired-t test. In terms of model assumptions, we assume that

$$\sum_{j=1}^{n} \tau_j = 0$$

$$E(y_{ij}) = \mu + \tau_j$$

$$V(y_{ij}) = V(\mu + \tau_j + \pi_i + e_{ij}) = \sigma_\pi^2 + \sigma_e^2$$

$$C(y_{ij}, y_{i'j}) = 0 \ \ for \ \ i \neq i'$$

$$C(y_{ij}, y_{ij'}) = \sigma_\pi^2 \ \ for \ \ j \neq j'$$

The magnitude of the within-subject association is described by the intra-class correlation.

$$Corr(y_{ij}, y_{ij'}) = \frac{\sigma_\pi^2}{\sigma_\pi^2 + \sigma_e^2} \ . \tag{2.2}$$

The intra-class correlation is the same correlation for all longitudinal pairs and represents the average correlation of y from any two timepoints. The intra-class correlation ranges from 0 to 1, so it can be interpreted as a proportion of

the total variance that is attributable to subjects. Given these assumptions, the variance covariance matrix has the "compound symmetric" structure:

$$\Sigma_{\boldsymbol{y}_i} = \begin{bmatrix} \sigma_e^2 + \sigma_\pi^2 & \sigma_\pi^2 & \sigma_\pi^2 & \dots & \sigma_\pi^2 \\ \sigma_\pi^2 & \sigma_e^2 + \sigma_\pi^2 & \sigma_\pi^2 & \dots & \sigma_\pi^2 \\ \dots & \dots & \dots & \dots & \dots \\ \sigma_\pi^2 & \dots & \sigma_\pi^2 & \sigma_e^2 + \sigma_\pi^2 & \sigma_\pi^2 \\ \sigma_\pi^2 & \sigma_\pi^2 & \dots & \sigma_\pi^2 & \sigma_e^2 + \sigma_\pi^2 \end{bmatrix} \quad (2.3)$$

where the variance is homogeneous across time: $\sigma_e^2 + \sigma_\pi^2$, the covariances are homogeneous across time: σ_π^2, and the correlation is given by $\sigma_\pi^2/(\sigma_e^2 + \sigma_\pi^2)$. Unfortunately, compound symmetry is not very realistic for longitudinal data. First, variances often change over time, where subjects are generally more similar at the start of the trial than at the end of the trial where some have responded to treatment and others have not. Second, covariances close in time are usually greater than covariances that are further separated in time.

In the balanced case (*i.e.*, no missing data and all subjects measured on the same occasions), the ANOVA table for a model with random subjects and

fixed time is of the following form:

Source	df	SS	MS	E(MS)
Subjects	$N-1$	$\mathrm{SS}_S = n\sum_{i=1}^{N}(\bar{y}_{i.} - \bar{y}_{..})^2$	$\frac{\mathrm{SS}_S}{N-1}$	$\sigma_e^2 + n\sigma_\pi^2$
Time	$n-1$	$\mathrm{SS}_T = N\sum_{j=1}^{n}(\bar{y}_{.j} - \bar{y}_{..})^2$	$\frac{\mathrm{SS}_T}{n-1}$	$\sigma_e^2 +$ $N\sum(\tau_j - \tau_.)^2$
Residual	$(N-1)$ $\times(n-1)$	$\mathrm{SS}_R = \sum_{i=1}^{N}\sum_{j=1}^{n} y_{ij}$ $-\bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2$	$\frac{\mathrm{SS}_R}{(N-1)(n-1)}$	σ_e^2
total	$Nn-1$	$\mathrm{SS}_y = \sum_{i=1}^{N}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{..})^2$		

where

- $\bar{y}_{..}$ = grand mean (averaged over time and subjects)
- $\bar{y}_{i.}$ = subject mean ($i = 1, \ldots, N$)
- $\bar{y}_{.j}$ = timepoint mean ($j = 1, \ldots, n$)

Tests of hypothesis are constructed as follows:

$$H_S: \qquad \sigma_\pi^2 = 0 \qquad F_S = \frac{\mathrm{MS}_S}{\mathrm{MS}_R} \overset{H_S}{\sim} F_{N-1,(N-1)(n-1)}$$

$$H_T: \quad \tau_1 = \tau_2 = ... = \tau_n = 0 \qquad F_T = \frac{\mathrm{MS}_T}{\mathrm{MS}_R} \overset{H_T}{\sim} F_{n-1,(N-1)(n-1)} \tag{2.4}$$

The intra-class correlation for this design describes the relative magnitude of σ_π^2. In general, we assume that $\sigma_\pi^2 > 0$, and we want to quantify the magnitude of the within-subject effect as,

$$r = \frac{\hat{\sigma}_\pi^2}{\hat{\sigma}_\pi^2 + \hat{\sigma}_e^2} \ . \tag{2.5}$$

where

$$\hat{\sigma}_\pi^2 \ = \ (\ \mathrm{MS}_S - \ \mathrm{MS}_R) \ / \ n \tag{2.6}$$

and

$$\hat{\sigma}_e^2 \ = \ \mathrm{MS}_R \tag{2.7}$$

When $\mathrm{MS}_S \leq \mathrm{MS}_R$, then $\hat{\sigma}_\pi^2 = 0$. Notice that the ICC is akin to an R^2 statistic; it represents the proportion of (unexplained) variation due to subjects. As subjects' data are highly correlated, r becomes large. The value of r can

vary depending on model covariates. As more and more of the within-subject variability is explained by model covariates, the value of r decreases.

The overall test of the null hypothesis of no difference over time,

$$H_T : \tau_1 = \tau_2 = ... = \tau_n = 0 \tag{2.8}$$

is a very global test. We can construct contrasts to address more specific time-related comparisons as follows:

$$L_{j'} \;=\; \sum_{j=1}^{n} c_{j'j}\; \bar{y}_{.j} \qquad j' = 1, \ldots, n-1 \qquad \sum_{j=1}^{n} c_{j'j} = 0 \tag{2.9}$$

then

$$\mathrm{MS}_{j'} \;=\; \mathrm{SS}_{j'} \;=\; \frac{N L_{j'}^2}{\sum_{j=1}^{n} c_{j'j}^2} \tag{2.10}$$

and

$$H_{j'} : L_{j'} = 0 \qquad F_{j'} = \frac{\mathrm{MS}_{j'}}{\mathrm{MS}_R} \overset{H_{j'}}{\sim} F_{1,(N-1)(n-1)}$$

$$t_{j'} = \frac{L_{j'}}{\sqrt{\mathrm{MS}_R \left[\sum_{j=1}^{n} \frac{c_{j'j}^2}{N}\right]}} \overset{H_{j'}}{\sim} t_{(N-1)(n-1)}$$

If the set of $n-1$ contrasts are orthogonal, then

$$\mathrm{SS}_T = \sum_{j'=1}^{n-1} \mathrm{SS}_{j'} \ , \tag{2.11}$$

and we have an independent partitioning of the variation due to time.

Trend Analysis - Orthogonal Polynomials

One approach to testing specific time-related contrasts is to characterize $n-1$ time effects as $n-1$ orthogonal polynomials (see Bock (1975), Pearson & Hartley (1970), Draper & Smith (1981), and Fleiss (1986)). For example, the $n-1 \ \times \ n$ contrast matrix for $n=4$ is

$$\boldsymbol{C} \ = \begin{array}{c} \left[\begin{array}{rrrrcr} -3 & -1 & 1 & 3 & \vdots & \div\sqrt{20} \\ 1 & -1 & -1 & 1 & \vdots & \div\sqrt{4} \\ -1 & 3 & -3 & 1 & \vdots & \div\sqrt{20} \end{array}\right] \end{array} \begin{array}{l} \text{linear} \\ \text{quad} \\ \text{cubic} \end{array} \tag{2.12}$$

for linear, quadratic, and cubic trend components. Some advantages of orthogonal polynomial are that:

- they are orthogonal (and orthonormal with $\div$),
- they can be generalized for unequally spaced timepoints,
- they are useful for determining "degree" of change across time,
- they can specify fewer than $n-1$ contrasts.

Change Relative to Baseline

In some cases, we are less concerned with the form of the growth curves, and more concerned about testing whether any change has occurred whatsoever. In this case, we can construct contrasts for each time-point relative to the first time-point, presumably baseline, as follows:

$$\boldsymbol{C} = \begin{bmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix} \tag{2.13}$$

These contrasts are for T2-T1, T3-T1, T4-T1, respectively. These contrasts are,

- not orthogonal,
- are sometimes called simple contrasts,
- can use any cell as the reference cell (*e.g.*, the last time-point).

Consecutive Time Comparisons - Profile Contrasts

In other cases, we may have interest in determining whether each consecutive time-point is significantly different from the immediately previous time-point.

These contrasts are sometimes referred to as "profile contrasts," and are constructed as follows:

$$\boldsymbol{C} = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix} \tag{2.14}$$

to test T2-T1, T3-T2, T4-T3, respectively. Profile contrasts are:

- not orthogonal,
- useful for identifying when change begins (and ends),

Contrasting Time-point to Mean of Subsequent Time-points - Helmert Contrasts

When interest is in contrasting each time-point to the mean of all subsequent time-points, we can construct Helmert contrasts (Bock, 1975) as follows:

$$\boldsymbol{C} = \begin{bmatrix} 1 & -1/3 & -1/3 & -1/3 \\ 0 & 1 & -1/2 & -1/2 \\ 0 & 0 & 1 & -1 \end{bmatrix} \tag{2.15}$$

for T1 versus average of T2, T3, T4; T2 versus average of T3, T4; T3 versus T4. Helmert contrasts are:

- orthogonal,
- and are useful for "ordered" tests.

Contrasting Each Time-point to the Mean of Others - Deviation Contrasts

When there is interest in contrasting each time-point to the mean of all other time-points, deviation contrasts can be computed as:

$$\boldsymbol{C} = \begin{bmatrix} 1 & -1/3 & -1/3 & -1/3 \\ -1/3 & 1 & -1/3 & -1/3 \\ -1/3 & -1/3 & 1 & -1/3 \end{bmatrix} \tag{2.16}$$

for T1 versus the average of T2, T3, T4; T2 versus the average of T1, T3, T4; and T3 versus the average of T1, T2, T4. Deviation contrasts are:

- not orthogonal,
- are useful for situations in which there is "vague prior knowledge",
- and the omitted time-point can be other than last.

Multiple Comparisons

Although $n-1$ multiple comparisons can often be specified a-priori, we are nevertheless faced with making multiple comparisons that will potentially lead

to elevated experiment-wise Type I error rates. The most conservative solution is to use the so-called Bonferroni corrected α level, which is given as

$$\alpha^* = \alpha/(n-1)\ . \tag{2.17}$$

A far less conservative alternative is to apply the so-called Fisher protected test logic, in which each individual test is conducted at the α level, but the individual tests are only applied when the global test $H_T : \tau_1 = \tau_2 = ... = \tau_n = 0$ is rejected. For the special case of orthogonal polynomials, we can start with the highest-order polynomial and eliminate each degree polynomial until we encounter the first significant one. Hummel & Sligo (1971) support Fisher protected test if n is not too large. Also, from Rosner (1995, p. 319):

> "If a few linear contrasts, which have been specified in advance, are to be tested, then it may not be necessary to use a multiple-comparisons procedure, since if such procedures are used, there will be less power to detect differences for linear contrasts whose means are truly different from zero. Conversely, if many contrasts are to be tested, which have not been specified before looking at the data, then multiple-comparisons procedures may be useful in protecting against declaring too many significant differences."

2.2 Multiple-sample Repeated Measures ANOVA

In the multiple-sample case, the ANOVA model for repeated measurements is referred to as a "split-plots" ANOVA model. An example data layout is:

		timepoint			
group	subject	1	2	...	n
1	1	y_{111}	y_{112}	...	y_{11n}
1	2	y_{121}	y_{122}	...	y_{12n}
1	.	.	.	...	.
1	.	.	.	...	.
1	N_1	y_{1N_11}	y_{1N_12}	...	y_{1N_1n}
.	.	.	.	...	.
.	.	.	.	...	.
.	.	.	.	...	.
.	.	.	.	...	.
s	1	y_{s11}	y_{s12}	...	y_{s1n}
s	2	y_{s21}	y_{s22}	...	y_{s2n}
s	.	.	.	...	.
s	.	.	.	...	.
s	N_s	y_{sN_s1}	y_{sN_s2}	...	y_{sN_sn}

where

- $h = 1, \ldots, s$ groups
- $i = 1, \ldots, N_h$ subjects in group h (with $N = \sum_{h=1}^{s} N_h$)
- $j = 1, \ldots, n$ timepoints

The ANOVA model is:

$$y_{hij} = \mu + \gamma_h + \tau_j + (\gamma\tau)_{hj} + \pi_{i(h)} + e_{hij} \tag{2.18}$$

where

- μ = grand mean
- γ_h = effect of group h $(\sum_h \gamma_h = 0)$
- τ_j = effect of time j $(\sum_j \tau_j = 0)$
- $(\gamma\tau)_{hj}$ = interaction effect of time j and group h

  $[\sum_h \sum_j (\gamma\tau)_{hj} \ = \ 0]$
- $\pi_{i(h)}$ = individual difference component for subject i in group h
- e_{hij} = error for subject i in group h at time j

Distributional assumptions include

- $\pi_{i(h)} \sim N(0, \sigma_\pi^2)$

- $e_{hij} \sim N(0, \sigma_e^2)$

which imply a compound symmetry structure for $V(\boldsymbol{y}_i)$. In this design, subjects are considered random effects and group and time are considered fixed effects. The data are assumed to be balanced in terms of n (*i.e.*, time-points),

but not in terms N_h, (*i.e.*, group sample sizes). The ANOVA table is as follows:

Source	df	SS	MS	E(MS)
Group	$s-1$	$\mathrm{SS}_G = n\sum_{h=1}^{s} N_h\,(\bar{y}_{h..} - \bar{y}_{...})^2$	$\frac{\mathrm{SS}_G}{s-1}$	$\sigma_e^2 + n\sigma_\pi^2 + D_G$
Time	$n-1$	$\mathrm{SS}_T = N\sum_{j=1}^{n}(\bar{y}_{..j} - \bar{y}_{...})^2$	$\frac{\mathrm{SS}_T}{n-1}$	$\sigma_e^2 + D_T$
Group × Time	$(s-1)$ $\times(n-1)$	$\mathrm{SS}_{GT} = \sum_{h=1}^{s}\sum_{j=1}^{n} N_h\,(\bar{y}_{h.j} - \bar{y}_{h..} - \bar{y}_{..j} + \bar{y}_{...})^2$	$\frac{\mathrm{SS}_{GT}}{(s-1)(n-1)}$	$\sigma_e^2 + D_{GT}$
Subjects in Grps	$N-s$	$\mathrm{SS}_{S(G)} = n\sum_{h=1}^{s}\sum_{i=1}^{N_h}(\bar{y}_{hi.} - \bar{y}_{h..})^2$	$\frac{\mathrm{SS}_{S(G)}}{N-s}$	$\sigma_e^2 + n\sigma_\pi^2$
Residual	$(N-s)$ $\times(n-1)$	$\mathrm{SS}_R = \sum_{h=1}^{s}\sum_{i=1}^{N_h}\sum_{j=1}^{n}(y_{hij} - \bar{y}_{h.j} - \bar{y}_{hi.} + \bar{y}_{h..})^2$	$\frac{\mathrm{SS}_R}{(N-s)(n-1)}$	σ_e^2
total	$Nn-1$	$\mathrm{SS}_y = \sum_{h=1}^{s}\sum_{i=1}^{N_h}\sum_{j=1}^{n}(y_{hij} - \bar{y}_{...})^2$		

where

- $h = 1, \ldots, s$ groups
- $i = 1, \ldots, N_h$ subjects in group h (with $N = \sum_{h=1}^{s} N_h$)
- $j = 1, \ldots, n$ timepoints

and D_G, D_T, and D_{GT} represent differences among groups, timepoints, and group $\times$ time interaction, respectively.

2.2.1 Testing for Group by Time interaction

The group by time interaction, which is typically the test of primary interest, is constructed as:

$$H_{GT} : D_{GT} = 0 \qquad F_{GT} = \frac{\mathrm{MS}_{GT}}{\mathrm{MS}_R} \overset{H_{GT}}{\sim} F_{(s-1)(n-1),(N-s)(n-1)} \ . \tag{2.19}$$

If rejected,

- group differences are not the same across time,
- group curves across time are not parallel,
- group and time effects are confounded with the interaction and cannot be separately tested (or estimated).

If accepted,

$$H_T : \tau_1 = \tau_2 = \ldots = \tau_n = 0 \qquad F_T = \frac{\mathrm{MS}_T}{\mathrm{MS}_R} \overset{H_T}{\sim} F_{n-1,(N-s)(n-1)} \tag{2.20}$$

$$H_G : \gamma_1 = \gamma_2 = \ldots = \gamma_s = 0 \qquad F_G = \frac{\mathrm{MS}_G}{\mathrm{MS}_{S(G)}} \overset{H_G}{\sim} F_{s-1,N-s} \tag{2.21}$$

H_T and H_G are separately and independently testable

2.2.2 Testing for Subject effect

To test for the significance of random subject effects, we construct the following statistic:

$$H_{S(G)} : \sigma_\pi^2 = 0 \qquad F_{S(G)} = \frac{\mathrm{MS}_{S(G)}}{\mathrm{MS}_R} \overset{H_S(G)}{\sim} F_{N-s,(N-s)(n-1)} \ . \tag{2.22}$$

In general, we assume that $\sigma_\pi^2 > 0$, and estimate the intra-class correlation as

$$r = \hat{\sigma}_\pi^2 / (\hat{\sigma}_\pi^2 + \hat{\sigma}_e^2) \tag{2.23}$$

The intra-class correlation represents the proportion of (unexplained) variation due to subjects.

2.2.3 Contrasts for time effects

As in the single-group case, it is advantageous to characterize time and group by time effects using time-related contrasts. These include the same sets of contrasts described earlier for single-group designs, namely, orthogonal polynomials, profile contrasts, Helmert contrasts *etc.*, As before, it may be necessary to consider the multiple comparisons issue and utilize Fisher protected tests or Bonferroni correction for the $(n-1)(s-1)$ group by time contrasts (and for the $n-1$ time contrasts).

Orthogonal Polynomial Partition of SS

An example orthogonal polynomial decomposition of the time effect for a four time-point design is given by

$$\boldsymbol{C}_4 = \begin{array}{c} \left[\begin{array}{rrrrcr} -3 & -1 & 1 & 3 & \vdots & \div\sqrt{20} \\ 1 & -1 & -1 & 1 & \vdots & \div\sqrt{4} \\ -1 & 3 & -3 & 1 & \vdots & \div\sqrt{20} \end{array}\right] \end{array} \begin{array}{l} \boldsymbol{c}_1 \\ \boldsymbol{c}_2 \\ \boldsymbol{c}_3\ . \end{array} \tag{2.24}$$

The decomposition of the Time sum of squares is given by:

Time	df	SS
linear	1	$\mathrm{SS}_{T_1} \;=\; N\,\boldsymbol{c}_1\,\bar{\boldsymbol{y}}_{..j}\,\bar{\boldsymbol{y}}'_{..j}\,\boldsymbol{c}'_1$
quadratic	1	$\mathrm{SS}_{T_2} \;=\; N\,\boldsymbol{c}_2\,\bar{\boldsymbol{y}}_{..j}\,\bar{\boldsymbol{y}}'_{..j}\,\boldsymbol{c}'_2$
.	.	.
$(n-1)$th	1	$\mathrm{SS}_{T_{n-1}} \;=\; N\,\boldsymbol{c}_{n-1}\,\bar{\boldsymbol{y}}_{..j}\,\bar{\boldsymbol{y}}'_{..j}\,\boldsymbol{c}'_{n-1}$
	$n-1$	SS_T

where $\bar{\boldsymbol{y}}_{..j} = n \times 1$ is the vector of timepoint means, and $\boldsymbol{c}_j = 1 \times n$ is the vector of contrasts of order j. Corresponding F-statistics for each trend component are:

$$
\begin{aligned}
F_{T_{n-1}} &= \mathrm{SS}_{T_{n-1}} \,/\, \mathrm{MS}_R \\
&\;\;.\quad .\quad . \\
F_{T_1} &= \mathrm{SS}_{T_1} \,/\, \mathrm{MS}_R\,.
\end{aligned}
$$

Typically, we determine the polynomial of least degree by working backwards. Similarly, the Group by Time sum of squares is decomposed as:

G × T	df	SS
linear	$s-1$	$\mathrm{SS}_{GT_1} = \sum_h N_h \, \boldsymbol{c}_1 \, \bar{\boldsymbol{y}}_{h.j} \, \bar{\boldsymbol{y}}'_{h.j} \, \boldsymbol{c}'_1 - \mathrm{SS}_{T_1}$
quadratic	$s-1$	$\mathrm{SS}_{GT_2} = \sum_h N_h \, \boldsymbol{c}_2 \, \bar{\boldsymbol{y}}_{h.j} \, \bar{\boldsymbol{y}}'_{h.j} \, \boldsymbol{c}'_2 - \mathrm{SS}_{T_2}$
.	.	.
$(n-1)$th	$s-1$	$\mathrm{SS}_{GT_{n-1}} = \sum_h N_h \, \boldsymbol{c}_{n-1} \, \bar{\boldsymbol{y}}_{h.j} \, \bar{\boldsymbol{y}}'_{h.j} \, \boldsymbol{c}'_{n-1} - \mathrm{SS}_{T_{n-1}}$
	$(s-1)$ $\times(n-1)$	SS_{GT}

where $\bar{\boldsymbol{y}}_{h.j} = n \times 1$ is the vector of timepoint means for group h, and $\boldsymbol{c}_j = 1 \times n$ is the vector of contrasts of order j. F-statistics are given by:

$$
\begin{aligned}
F_{GT_{n-1}} &= \frac{\mathrm{SS}_{GT_{n-1}} \,/\, (s-1)}{\mathrm{MS}_R} \\
&\;\;\;\;\;\; . \quad . \quad . \\
&\;\;\;\;\;\; . \quad . \quad . \\
F_{GT_1} &= \frac{\mathrm{SS}_{GT_1} \,/\, (s-1)}{\mathrm{MS}_R}
\end{aligned}
$$

Again, the order of the polynomial of least degree for the Group by Time interaction is determined working backwards.

2.2.4 Compound Symmetry and Sphericity

For both univariate RB and SP ANOVA models

$$V(\boldsymbol{y}_i) \;=\; \sigma_\pi^2 \mathbf{1}_n \mathbf{1}_n' \;+\; \sigma_e^2 \boldsymbol{I}_n \tag{2.25}$$

Compound symmetry (CS) structure is given by

$$\begin{aligned} V(y_{ij}) &= \sigma_\pi^2 + \sigma_e^2 \qquad \forall\, j \\ C(y_{ij}, y_{ij'}) &= \sigma_\pi^2 \qquad \forall\, j \text{ and } j' \;\; (j \neq j') \,. \end{aligned} \tag{2.26}$$

The correlation between responses across time-points is given by the intra-class correlation

$$\rho(y_{ij}, y_{ij'}) \;=\; \sigma_\pi^2 \,/\, (\sigma_\pi^2 + \sigma_e^2) \,. \tag{2.27}$$

The CS assumption is highly restrictive, and often unrealistic (especially as n gets large). CS is a special case of the more general situation, sphericity, under which F-tests for time-related terms from ANOVA models are valid. If sphericity holds, then F-tests are valid, otherwise, if sphericity doesn't hold, then F-tests are generally too liberal.

Sphericity

Sphericity, sometimes called circularity can be expressed in different ways. The most general is that all variances of all pair-wise differences between variables are equal

$$\begin{aligned} V(y_{ij} - y_{ij'}) &= V(y_{ij}) + V(y_{ij'}) - 2C(y_{ij}, y_{ij'}) \\ &= \text{ constant } \forall\, j \text{ and } j' \ . \end{aligned}$$

CS is a special case of sphericity, *i.e.*,

$$\begin{aligned} V(y_{ij}) &= \sigma_\pi^2 + \sigma_e^2 \qquad \text{constant} \\ C(y_{ij}, y_{ij'}) &= \sigma_\pi^2 \qquad\qquad \text{constant} \ . \end{aligned} \tag{2.28}$$

More generally, Crowder and Hand (1990, page 50) note:

> "MS ratios derived by the univariate approach follow exact F-distributions if and only if the covariance matrix of the orthonormal contrasts has equal variances and zero covariances."

This statement is equivalent to:

$$\underset{(n-1)\times n}{\boldsymbol{C}} \quad \underset{n\times n}{V(\boldsymbol{y}_i)} \quad \underset{(n-1)\times n}{\boldsymbol{C}'} \quad = \quad \text{constant} \quad \underset{(n-1)\times(n-1)}{\boldsymbol{I}_{t-1}} \tag{2.29}$$

where $\boldsymbol{C}$ is a matrix of orthonormal polynomials.

As noted by Mauchly (1940), sphericity tests have low statistical power for small sample sizes. For large samples, the test is likely to be significant even though the effect on the F-test may be negligible. Sphericity tests are sensitive to departures from normality and to the presence of outliers. As such, sphericity tests should be used as a guide and not as a rule.

If the sphericity assumption is rejected, or deemed implausible, one can use a multivariate repeated measures analysis (MANOVA), which allows for general $V(\boldsymbol{y}_i)$ but does not allow for any missing data across time. Other classical statistical alternatives include adjusted univariate F-tests as described by Greenhouse and Geisser (1959), and Huynh and Feldt (1976), both of which are generally overly conservative.

2.2.5 Running Univariate Repeated-measures ANOVA via Multivariate Setup

The univariate problem is described as:

- univariate setup:

group	subject	time	y
1	1	1	y_{111}
1	1	2	y_{112}
1	1	3	y_{113}

whereas the multivariate problem can be described as:

- multivariate setup:

group	subject	y1	y2	y3
1	1	y_{111}	y_{112}	y_{113}

The differences between the two approaches are:

- Subjects with missing data on any timepoint are deleted from the analysis (unlike in the univariate setup).
- For the time-related contrasts, though SS and MS are the same, the F-tests are not.
- We only get a test of the sphericity assumption via the multivariate setup.
- The denominator degrees of freedom are greater for the univariate repeated-measures tests, which generally results in greater power for the univariate tests.

- The sphericity test is for the pooled subjects-within-groups variance-covariance matrix.

- We get different results from sphericity tests if the contrast matrix $\boldsymbol{C}$ is not orthonormal.

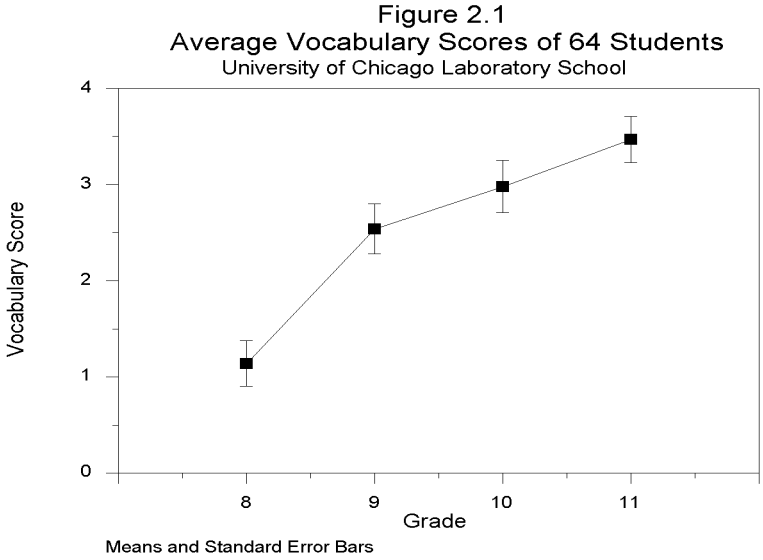
- The overall F-tests of main effects and interactions do not change if different contrasts are used.

In the following chapter we discuss the MANOVA approach to analysis of longitudinal data in detail.

2.3 Illustration

Bock (1975) presents data on vocabulary growth measured in a cohort of 64 students at the University of Chicago Laboratory school. The longitudinal data consist of repeated measurements of the vocabulary section of the cooperative reading test (Davis, 1950). Alternate forms of the test were administered in eighth through eleventh grade. Since this age range marks the period of time that physical growth begins to decelerate, he hypothesized that a similar deceleration might be observed in the acquisition of new vocabulary as well.

Figure 2.1 displays a plot of the average scores versus grade and visually suggests that the rate of change is indeed decelerating.



Summary statistics (means, standard deviations, and correlations) are presented in Table 2.1.

Table 2.1

Means, Standard Deviations, and Correlations

for the Vocabulary-Growth Data

Grade	Mean	SD	Correlations			
8	1.137	1.889	1.000			
9	2.542	2.085	.810	1.000		
10	2.988	2.169	.868	.785	1.000	
11	3.472	1.925	.785	.757	.811	1.000

Inspection of Table 2.1, reveals that the variances and covariances are reasonably homogeneous over time, supporting the assumption of compound symmetry underlying the mixed-model ANOVA.

This example can be used to nicely illustrate the simplest case of a one-sample mixed-model ANOVA. In the model, subject represents a random effect, and time represents a fixed effect. The ANOVA is illustrated in Table 2.2, which presents the ANOVA results.

Table 2.2

Mixed-Model ANOVA Results

for the Vocabulary-Growth Data

Source	df	SS	MS	F	p
Subjects	63	$\mathrm{SS}_S = 873.60$	13.87	16.91	.0001
Grade (*i.e.*, "Time")	3	$\mathrm{SS}_T = 194.34$	64.78	79.02	.0001
Residual	189	$\mathrm{SS}_R = 154.94$	0.82		
Total	255	$\mathrm{SS}_y = 1{,}222.88$			

The estimate of the error variance, which equals MSR, is obtained as:

$$\hat{\sigma}_e^2 = \frac{\mathrm{SS}_R}{(N-1)(n-1)} = \frac{154.94}{189} = 0.82$$

The estimate of the subject variance is gotten as:

$$\hat{\sigma}_\pi^2 = \frac{\mathrm{MS}_S - \mathrm{MS}_R}{n} = \frac{13.87 - 0.82}{4} = 3.26\ ,$$

and so the intraclass correlation equals

$$r = \frac{\hat{\sigma}_\pi^2}{\hat{\sigma}_\pi^2 + \hat{\sigma}_e^2} = \frac{3.26}{3.26 + 0.82} = .80\ .$$

Thus, as one would expect, there is a tremendous effect of subjects on their vocabulary scores. 80% of the variation in vocabulary, that is not explained by grade, is attributable to subjects.

In terms of the grade effect, the ANOVA table reveals that we must reject the null hypothesis of no grade effect. This is clearly supported by Figure 2.1 which shows that vocabulary generally increases with grade. However, to obtain a more sensitive analysis, we can examine the significance of the individual polynomial terms based on the order-4 orthogonal polynomial matrix:

$$\boldsymbol{C} = \begin{bmatrix} -.67082 & -.22361 & .22361 & .67082 \\ .5 & -.5 & -.5 & .5 \\ -.22361 & .67082 & -.67082 & .22361 \end{bmatrix}.$$

Premultiplying the 4×1 vector of column means

$$\bar{\boldsymbol{y}} = \begin{bmatrix} 1.14 \\ 2.54 \\ 2.99 \\ 3.47 \end{bmatrix},$$

by $\boldsymbol{C}$, we obtain the following orthogonal estimates:

$$\text{Linear } = 1.67, \quad \text{Quadratic } = -0.46, \quad \text{Cubic } = 0.22.$$

Note that the squares of these estimates multiplied by 64 (*i.e.*, the number of subjects) are the numerators of the F-ratios for linear, quadratic, and cubic trends shown in Table 2.3. The denominator is $\hat{\sigma}_e^2 = 0.82$ and the denominator degrees of freedom equals 189 for these F-ratios. Clearly, the positive linear trend is highly significant. However, the significant quadratic term, coupled with its negative sign, reveals that the observed deceleration in the growth rate is statistically significant. Finally, the cubic term is only marginally significant suggesting that the deceleration reverses to some extent with increasing age (see Figure 2.1). Examining the trend estimates, we see that the cubic estimate also pales in comparison with the dominant linear and moderate quadratic trend components: there is clearly diminishing importance as the order of the polynomial is increased.

Table 2.3

Orthogonal Polynomial Decomposition of the Grade Effect

Source	df	SS	F	p
Grade				
Linear	1	$SS_{T_1} = 177.58$	216.63	.0001
Quadratic	1	$SS_{T_2} = 13.58$	16.56	.0001
Cubic	1	$SS_{T_3} = 3.17$	3.86	.051

Taken together, these results indicate a decelerating positive trend across grade, supporting the notion that vocabulary acquisition is slowing down as students approach maturity.

2.4 Summary

In summary, the ANOVA approach to analysis of longitudinal data represents a well understood and well developed statistical methodology. In addition, there is considerable available computer software for ANOVA computation. The results are based on relatively simple and non-iterative calculations. By contrast, the ANOVA model for repeated measurements assumes sphericity, which is unrealistic for most applications where variances tend to increase with time and correlation decreases with increasing intervals in time. Other

limitations include limited treatment of missing data, and the requirement that all subjects are measured at on the same occasions.