THE GENERAL LINEAR MODEL AND UNBALANCED DESIGNS

ERSH 8310

Keppel and Wickens Chapter 14

Today's Class

- The General Linear Model
- The Two-Factor Analysis
- Averaging of Groups and Individuals
- Contrasts and Other Analytical Analyses

The General Linear Model

The General Linear Model

- The ideal for most research designs is to have an equal number of subjects in each group.
- However, unequal sample sizes are often the reality, even in studies that were planned to have equal samples.
- The use of classification factors (e.g., gender, college major, race, occupation, etc.) whose levels can only be determined after the sampling has occurred almost always leads to unequal groups.

The General Linear Model

- To analyze experiments with unequal samples, the statistical procedures of the analysis of variance must be given in a more general form.
- Statisticians have observed that most varieties of the analysis of variance can be expressed in a common way and that this representation also includes related techniques such as multiple regression.
- This approach is known as the general linear model.

A Numerical Example

- The one-way data presented in Section 3.5 will be re-analyzed (see Table 14.1).

***Table 14.1:** Data from three groups (from Table 3.7, p. 57) with descriptions fitted under the null hypothesis (first block of three columns) and its alternative (second block).*

Group	Y_{ij}	$\bar{Y}_T$	$Y_{ij} - \bar{Y}_T$	$(Y_{ij} - \bar{Y}_T)^2$	$\bar{Y}_j$	$Y_{ij} - \bar{Y}_j$	$(Y_{ij} - \bar{Y}_j)^2$
		Null hypothesis true			Null hypothesis false		
a_1	5	7.77	−2.77	7.67	7.00	−2.00	4.00
	9	7.77	1.23	1.51	7.00	2.00	4.00
	7	7.77	−0.77	0.59	7.00	0.00	0.00
a_2	12	7.77	4.23	17.90	10.00	2.00	4.00
	10	7.77	2.23	4.97	10.00	0.00	0.00
	10	7.77	2.23	4.97	10.00	0.00	0.00
	8	7.77	0.23	0.05	10.00	−2.00	4.00
	11	7.77	3.23	10.43	10.00	1.00	1.00
	9	7.77	1.23	1.51	10.00	−1.00	1.00
a_3	3	7.77	−4.77	22.75	5.00	−2.00	4.00
	6	7.77	−1.77	3.13	5.00	1.00	1.00
	5	7.77	−2.77	7.67	5.00	0.00	0.00
	6	7.77	−1.77	3.13	5.00	1.00	1.00
Sum of squared deviations:				86.28			24.00

A Numerical Example

- When the null hypothesis is true, the unexplained sum of squares is

$$SS_{unexp}^{H_0} = \sum (Y_{ij} - \bar{Y}_T)^2.$$

- When the alternative hypothesis is true, the unexplained sum of squares is

$$SS_{unexp}^{H_1} = \sum_{ij} (Y_{ij} - \bar{Y}_j)^2.$$

Variability

- The variability produced by the treatment effects can be measured by taking the difference between the two unexplained sums of squares:

$$
\begin{aligned}
SS_A &= SS_{unexp}^{H_0} - SS_{unexp}^{H_1} \\
&= (\text{treatment effects} + \text{error}) - (\text{error})
\end{aligned}
$$

Additional Notes

- Note that:

$$SS_{error} = SS_{unexp}^{H_1}.$$

Degrees of Freedom

- The degrees of freedom are equal to the different values used in the predictions (e.g., one for the grand mean μ_T in the null hypothesis, and three for the $a = 3$ group means μ_i in the alternative hypothesis).
- The degrees of freedom for the effect equal the difference between the two values,

$$df_A = df_{unexp}{}^{H_0} - df_{unexp}{}^{H_1}.$$

F-Ratio

- We can obtain the F ratio using:

$$F = \frac{MS_A}{MS_{error}},$$

- where
  - $MS_A = SS_A / df_A$
  - $MS_{error} = SS_{error} / df_{error}$

The Linear Model

- For the one-way design, the linear model is:

$$Y_{ij} = \mu_T + \alpha_j + E_{ij}$$

  with Σ_i $a_i = 0$.
- Note that the above is the alternative-hypothesis model.
- The null-hypothesis model is

$$Y_{ij} = \mu_T + E_{ij}.$$

- The two models form a hierarchical pair, in which the null-hypothesis model is a special case of the other.

Testing by Model Comparison

- The general linear model uses a hierarchical pair of linear models that differ in whether the effect to be tested is included or not.
- The sum of squared deviations has the form,

$$SS_{unexp} = \sum_{ij} [(\text{data}) - (\text{fitted value})]^2 = \sum_{ij} (Y_{ij} - \hat{Y}_{ij})^2.$$

- In the one-way analysis, $[{}^\wedge Y]_{ij} = [{}^{-}Y]_T$ for the null hypothesis model and $[{}^\wedge Y]_{ij} = [{}^{-}Y]_i$ for the alternative-hypothesis model.

Testing by Model Comparison

- The difference between two sums of squares is the sum of squares for the test of a null hypothesis,

$$SS_{effect} = SS_{unexp}^{H_0} - SS_{unexp}^{H_1}.$$

- The degre … lained variability are calculated via

$$df_{unexp} = (\text{observations}) - (\text{parameters}).$$

Testing by Model Comparison

- For the data in Table 14.1, the null-hypothesis model has only one parameter and the alternative-hypothesis model has three parameters (i.e., μ_T and two α_i due to the constraint that $\Sigma_i\ a_i = 0$).
- The degrees of freedom for an effect are calculated by:

$$df_{effect} = df_{unexp}^{H_0} - df_{unexp}^{H_1}.$$

F-Ratio

- We can obtain the F ratio using:

$$F = \frac{MS_{effect}}{MS_{error}},$$

- where
  - $MS_{effect} = SS_{effect} / df_{effect}$
  - $MS_{error} = SS_{error} / df_{error}$.

The Two-Factor Analysis

The Two-Factor Analysis

- Table 14.2 contains data for the three treatment levels for factor A (i.e., the first session a_1, the second session a_2, and the third session a_3) and the two levels for factor B (i.e., men b_1, and women b_2).

The General Linear Model in a Factorial Design

- The standard two-factor linear model is:

$$Y_{ijk} = \mu_T + \alpha_j + \beta_k + (\alpha\beta)_{jk} + E_{ijk}.$$

- It is sometimes called the full linear model or the general alternative-hypothesis model (i.e., H_1).
- We may create the three null-hypothesis models by deleting different parameters from this model.

Two-Factor Null Hypothesis Models

$$Y_{ijk} = \mu_T + \beta_k + (\alpha\beta)_{jk} + E_{ijk},$$

$$Y_{ijk} = \mu_T + \alpha_j + (\alpha\beta)_{jk} + E_{ijk},$$

$$Y_{ijk} = \mu_T + \alpha_j + \beta_k + E_{ijk}.$$

No Differences Model

- Note that the model of no differences among groups is:

$$Y_{ijk} = \mu_T + E_{ijk},$$

- And its sum of squares is:

$$SS_{total} = \sum_{ijk} (Y_{ijk} - \mu_T)^2$$

- With the degrees of freedom N-1, where N is the total number of observations.

Sums of Squares

- Using the null-hypothesis models, we can obtain the sums of squares for the main effects and interaction as:

$$SS_A = SS_{unexp}^{H_0(A)} - SS_{unexp}^{H_1},$$

$$SS_B = SS_{unexp}^{H_0(B)} - SS_{unexp}^{H_1},$$

and

$$SS_{A\times B} = SS_{unexp}^{H_0(A\times B)} - SS_{unexp}^{H_1}.$$

- res. The corresponding degrees of freedoms are (a-1), (b-1), and (a-1)(b-1).

Nonorthogonality of the Effects

- The unequal group sizes cause the A, B, and $A \times B$ effects to be partially confounded with each other.
- This nonorthogonality is an intrinsic characteristic of unbalanced factorial designs, and it is why they are also called nonorthogonal or unbalanced designs.
- The additive sums of squares can be obtained using the Type I sums of squares (i.e., hierarchical sums of squares).

Averaging of Groups and Individuals

Averaging of Groups and Individuals

- The presence of unequal sample sizes presents a dilemma whenever we combine data from two or more groups.
- The averages can be obtained either score-based (i.e., considering different sample sizes) or group-based computation.
- The group-based averages are often identified as estimated means, adjusted means, or least-squares means.
- These two sets of means have different interpretations.

Averaging of Groups and Individuals

- The Type III and Type I sums of squares are identical when the sample sizes are equal.
- Unless the conclusions depend on the size of the samples, the Type III sums of squares should be used with unequal sample sizes.
- You should interpret them using the marginal averages based on the cell means (i.e., the estimated, adjusted, or least-squares means).

Contrasts and Other Analytical Analyses

Contrasts and Other Analytical Analyses

- We can analyze the first three columns of Table 14.2 using SPSS and obtain $MS_{A\ at\ b1}$.
- The F ratio can be obtained with the MS_{error} from the complete analysis.

Contrasts and Other Analytical Analyses

We may test a single-df effect with

$$\hat{\psi} = \sum_{jk} c_{jk} \bar{Y}_{jk} ,$$

where $\sum_{jk} c_{jk} = 0$, by

$$t_{\psi} = \frac{\hat{\psi}}{s_{\psi}}$$

with

$$s_{\psi}^2 = MS_{error} \sum_{jk} \frac{c_{jk}^2}{n_{jk}} .$$

Sensitivity to Assumptions

- Unequal samples do increase the sensitivity of analyses to heterogeneity of variance.
- Try to design your study so that the sizes of the groups are the same.

Final Thought

- Today's class delved into what to do following the overall two-way analysis.
- Interaction effects were the focus of the discussion.

- Note that most everything discussed today came in the presence of the possible interaction between independent variables.
- Pretty much anything that you can do in a one-way ANOVA can be accomplished for interactions.