

Chapter 16 - Analyses of Variance and Covariance as General Linear Models

16.1 Eye fixations per line of text for poor, average, and good readers:

a. Design matrix, using only the first subject in each group:

$$X = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{bmatrix}$$

b. Computer exercise:

$$R^2 = .608 \quad SS_{\text{reg}} = 57.7333 \quad SS_{\text{residual}} = 37.2000$$

c. Analysis of variance:

$$\bar{X}_1 = 8.2000 \quad \bar{X}_2 = 5.6 \quad \bar{X}_3 = 3.4 \quad \bar{X}_\cdot = 5.733$$

$$n_1 = 5 \quad n_2 = 5 \quad n_3 = 5 \quad N = 15 \quad \Sigma X = 86 \quad \Sigma X^2 = 588$$

$$SS_{\text{total}} = \Sigma X^2 - \frac{(\Sigma X)^2}{N} = 588 - \frac{86^2}{15} = 94.933$$

$$SS_{\text{group}} = n \Sigma (\bar{X}_j - \bar{X}_\cdot)^2 = 5[(8.2000 - 5.733)^2 + (5.6 - 5.733)^2 + (3.4 - 5.733)^2] \\ = 57.733$$

$$SS_{\text{error}} = SS_{\text{total}} - SS_{\text{group}} = 94.933 - 57.733 = 37.200$$

Source	df	SS	MS	F
Group	2	57.733	28.867	9.312*
Error	12	37.200	3.100	
Total	14	94.933		

$$*p < .05 \quad [F_{.05(2,12)} = 3.89]$$

16.3 Data from Exercise 16.1, modified to make unequal ns:

$$R^2 = .624 \quad SS_{\text{reg}} = 79.0095 \quad SS_{\text{residual}} = 47.6571$$

Analysis of variance:

$$\bar{X}_1 = 8.2000 \quad \bar{X}_2 = 5.8571 \quad \bar{X}_3 = 3.3333 \quad \bar{X}_\cdot = 5.7968$$

$$n_1 = 5 \quad n_2 = 7 \quad n_3 = 9 \quad N = 21 \quad \Sigma X = 112 \quad \Sigma X^2 = 724$$

$$SS_{total} = \Sigma X^2 - \frac{(\Sigma X)^2}{N} = 724 - \frac{112^2}{21} = 126.6666$$

$$SS_{group} = \Sigma n_j (\bar{X}_j - \bar{X})^2 = 5[(8.2000 - 5.7968)^2] + 7[(5.8571 - 5.7968)^2] + 9[(3.3333 - 5.7968)^2]$$

$$= 79.0095$$

$$SS_{error} = SS_{total} - SS_{group} = 126.6666 - 79.0095 = 47.6571$$

Source	df	SS	MS	F
Group	2	79.0095	39.5048	14.92*
Error	18	47.6571	2.6476	
Total	20	126.6666		

$$*p < .05 \quad [F_{.05(2,18)} = 3.55]$$

16.5 Relationship between Gender, SES, and Locus of Control:

a. Analysis of Variance:

		SES			Mean
		Low	Average	High	
Gender	Male	12.25	14.25	17.25	14.583
	Female	8.25	12.25	16.25	12.250
	Mean	10.25	13.25	16.75	13.417

$$\Sigma X = 644 \quad \Sigma X^2 = 9418 \quad n = 8 \quad N = 48$$

$$SS_{total} = \Sigma X^2 - \frac{(\Sigma X)^2}{N} = 9418 - \frac{644^2}{48} = 777.6667$$

$$SS_{gender} = sn \Sigma (\bar{X}_i - \bar{X}_{..})^2 = 3(8)[(14.583 - 13.417)^2 + (12.250 - 13.417)^2]$$

$$= 65.333$$

$$SS_{SES} = gn \Sigma (\bar{X}_{.j} - \bar{X}_{..})^2 = 2(8)[(10.25 - 13.417)^2 + (13.25 - 13.417)^2 + (16.75 - 13.417)^2]$$

$$= 338.6667$$

$$SS_{cells} = n \Sigma (\bar{X}_{ij} - \bar{X}_{..})^2 = 8[(12.25 - 13.417)^2 + \dots + (16.25 - 13.417)^2] = 422.6667$$

$$SS_{GS} = SS_{cells} - SS_{gender} - SS_{SES} = 422.6667 - 65.3333 - 338.6667 = 18.6667$$

$$SS_{error} = SS_{total} - SS_{cells} = 777.6667 - 422.6667 = 355.0000$$

Source	df	SS	MS	F
Gender	1	65.333	65.333	7.730*

SES	2	338.667	169.333	20.034*
<i>G</i> x <i>S</i>	2	18.667	9.333	1.104
Error	42	355.000	8.452	
Total	47	777.667		

$$*p < .05 \quad [F_{.05(1,42)} = 4.08; F_{.05(2,42)} = 3.23]$$

b. ANOVA summary table constructed from sums of squares calculated from design matrix:

$$SS_G = SS_{\text{reg}(\alpha, \beta, \alpha\beta)} - SS_{\text{reg}(\beta, \alpha\beta)} = 422.6667 - 357.3333 = 65.333$$

$$SS_S = SS_{\text{reg}(\alpha, \beta, \alpha\beta)} - SS_{\text{reg}(\alpha, \alpha\beta)} = 422.6667 - 84.0000 = 338.667$$

$$SS_{GS} = SS_{\text{reg}(\alpha, \beta, \alpha\beta)} - SS_{\text{reg}(\alpha, \beta)} = 422.6667 - 404.000 = 18.667$$

$$SS_{\text{total}} = SS_Y = 777.667$$

The summary table is exactly the same as in part a (above).

16.7 The data from Exercise 16.5 modified to make unequal *ns*:

$$SS_{\text{error}} = SS_Y - SS_{\text{reg}(\alpha, \beta, \alpha\beta)} = 750.1951 - 458.7285 = 291.467$$

$$SS_G = SS_{\text{reg}(\alpha, \beta, \alpha\beta)} - SS_{\text{reg}(\beta, \alpha\beta)} = 458.7285 - 398.7135 = 60.015$$

$$SS_S = SS_{\text{reg}(\alpha, \beta, \alpha\beta)} - SS_{\text{reg}(\alpha, \alpha\beta)} = 458.7285 - 112.3392 = 346.389$$

$$SS_{GS} = SS_{\text{reg}(\alpha, \beta, \alpha\beta)} - SS_{\text{reg}(\alpha, \beta)} = 458.7285 - 437.6338 = 21.095$$

Source	<i>df</i>	SS	MS	<i>F</i>
Gender	1	60.015	60.015	7.21*
SES	2	346.389	173.195	20.80*
<i>G</i> x <i>S</i>	2	21.095	10.547	1.27
Error	35	291.467	8.328	
Total	40			

$$*p < .05 \quad [F_{.05(1,35)} = 4.12; F_{.05(2,35)} = 3.27]$$

16.9 Model from data in Exercise 16.5:

$$1.1667A_1 - 3.1667B_1 - 0.1667B_2 + 0.8333AB_{11} - 0.1667AB_{12} + 13.4167$$

Means:

	SES (<i>B</i>)		
	Low	Avg	High

Gender (A)	Male	12.25	14.25	17.25	14.583
	Female	8.25	12.25	16.25	12.250
		10.25	13.25	16.75	13.417

$$\hat{\mu} = \bar{X}_{..} = 13.4167 = b_0 = \text{intercept}$$

$$\hat{\alpha}_1 = \bar{A}_1 - \bar{X}_{..} = 14.583 - 13.4167 = 1.1667 = b_1$$

$$\hat{\beta}_1 = \bar{B}_1 - \bar{X}_{..} = 10.25 - 13.4167 = -3.1667 = b_2$$

$$\hat{\beta}_2 = \bar{B}_2 - \bar{X}_{..} = 13.25 - 13.4167 = -0.1667 = b_3$$

$$\hat{\alpha}\hat{\beta}_{11} = \overline{AB}_{11} - \bar{A}_1 - \bar{B}_1 + \bar{X}_{..} = 12.25 - 14.583 - 10.25 + 13.4167 = 0.8337 = b_4$$

$$\hat{\alpha}\hat{\beta}_{12} = \overline{AB}_{12} - \bar{A}_1 - \bar{B}_2 + \bar{X}_{..} = 14.25 - 14.583 - 13.250 + 13.4167 = -0.1667 = b_5$$

16.11 Does Method III really deal with unweighted means?

Means:

	B1	B2	weighted	unweighted
A1	4	10	8.5	7.0
A2	10	4	8.0	7.0
weighted	8.0	8.5	8.29	
unweighted	7.0	7.0		7.0

The full model produced by Method 1: $\hat{Y} = 0.0A_1 + 0.0B_1 - 3.0AB_{11} + 7.0$

Effects calculated on weighted means:

$$\hat{\mu} = \bar{X}_{..} = 8.29 \neq b_0 = \text{intercept}$$

$$\hat{\alpha}_1 = \bar{A}_1 - \bar{X}_{..} = 8.5 - 8.29 = .21 \neq b_1$$

$$\hat{\beta}_1 = \bar{B}_1 - \bar{X}_{..} = 8.0 - 8.29 = -.29 \neq b_2$$

$$\hat{\alpha}\hat{\beta}_{11} = \overline{AB}_{11} - \bar{A}_1 - \bar{B}_1 + \bar{X}_{..} = 4.00 - 8.5 - 8.0 + 8.29 = -4.21 \neq b_3$$

Effects calculated on unweighted means:

$$\hat{\mu} = \bar{X}_{..} = 7.00 = b_0 = \text{intercept}$$

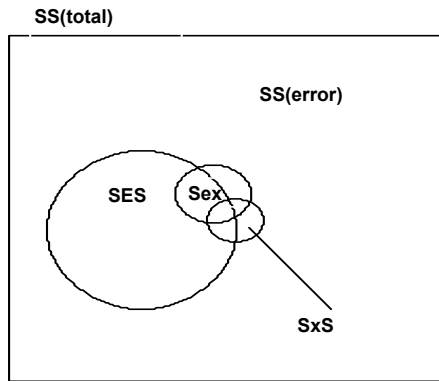
$$\hat{\alpha}_1 = \bar{A}_1 - \bar{X}_{..} = 7.0 - 7.0 = 0.0 = b_1$$

$$\hat{\beta}_1 = \bar{B}_1 - \bar{X}_{..} = 7.0 - 7.0 = 0.0 = b_2$$

$$\hat{\alpha}\hat{\beta}_{11} = \overline{AB}_{11} - \bar{A}_1 - \bar{B}_1 + \bar{X}_{..} = 4.00 - 7.0 - 7.0 + 7.0 = -3.0 = b_3$$

These coefficients found by the model clearly reflect the effects computed on unweighted means. Alternately, carrying out the complete analysis leads to $SS_A = SS_B = 0.00$, again reflecting equality of unweighted means.

16.13 Venn diagram representing the sums of squares in Exercise 16.7:



16.15 Energy consumption of families:

a. Design matrix, using only the first entry in each group for illustration purposes:

$$X = \begin{bmatrix} 1 & 0 & 58 & 75 \\ \dots & \dots & \dots & \dots \\ 0 & 1 & 60 & 70 \\ \dots & \dots & \dots & \dots \\ -1 & -1 & 75 & 80 \end{bmatrix}$$

b. Analysis of covariance:

$$SS_{\text{reg}(\alpha, \text{cov}, \alpha c)} = 2424.6202$$

$$SS_{\text{reg}(\alpha, \text{cov})} = 2369.2112$$

$$SS_{\text{residual}} = 246.5221 = SS_{\text{error}}$$

There is not a significant decrement in SS_{reg} and thus we can continue to assume homogeneity of regression.

$$SS_{\text{reg}(\alpha)} = 1118.5333$$

$$SS_{\text{cov}} = SS_{\text{reg}(\alpha, \text{cov})} - SS_{\text{reg}(\alpha)} = 2369.2112 - 1118.5333 = 1250.6779$$

$$SS_{\text{reg(cov)}} = 1716.2884$$

$$SS_A = SS_{\text{reg}(\alpha, \text{cov})} - SS_{\text{reg(cov)}} = 2369.2112 - 1716.2884 = 652.9228$$

Source	df	SS	MS	F
Covariate	1	1250.6779	1250.6779	55.81*
A (Group)	2	652.9228	326.4614	14.57*
Error	11	246.5221	22.4111	
Total	14	2615.7333		

$$* p < .05 \quad [F_{.05(1,11)} = 4.84; F_{.05(2,11)} = 3.98]$$

16.17 Adjusted means for the data in Exercise 16.16:

(The order of the means may differ depending on how you code the group membership and how the software sets up its design matrix. But the numerical values should agree.)

$$\hat{Y} = -7.9099A_1 + 0.8786A_2 - 2.4022B_1 + 0.5667AB_{11} + 0.1311AB_{21} + 0.7260C + 6.3740$$

$$\begin{aligned} \hat{Y}_{11} &= -7.9099(1) + 0.8786(0) - 2.4022(1) + 0.5667(1) + 0.1311(0) \\ &\quad + 0.7260(61.3333) + 6.3740 = 41.1566 \end{aligned}$$

$$\begin{aligned} \hat{Y}_{12} &= -7.9099(1) + 0.8786(0) - 2.4022(-1) + 0.5667(-1) + 0.1311(0) \\ &\quad + 0.7260(61.3333) + 6.3740 = 44.8276 \end{aligned}$$

$$\begin{aligned} \hat{Y}_{21} &= -7.9099(0) + 0.8786(1) - 2.4022(1) + 0.5667(0) + 0.1311(1) \\ &\quad + 0.7260(61.3333) + 6.3740 = 49.5095 \end{aligned}$$

$$\begin{aligned} \hat{Y}_{22} &= -7.9099(0) + 0.8786(1) - 2.4022(-1) + 0.5667(0) + 0.1311(-1) \\ &\quad + 0.7260(61.3333) + 6.3740 = 54.0517 \end{aligned}$$

$$\begin{aligned} \hat{Y}_{31} &= -7.9099(-1) + 0.8786(-1) - 2.4022(1) + 0.5667(-1) + 0.1311(-1) \\ &\quad + 0.7260(61.3333) + 6.3740 = 54.8333 \end{aligned}$$

$$\begin{aligned} \hat{Y}_{32} &= -7.9099(-1) + 0.8786(-1) - 2.4022(-1) + 0.5667(1) + 0.1311(1) \\ &\quad + 0.7260(61.3333) + 6.3740 = 61.0333 \end{aligned}$$

(We enter 61.3333 for the covariate in each case, because we want to estimate what the cell means would be if the observations in those cells were always at the mean of the covariate.)

16.19 Computer exercise.

16.21 Analysis of GSIT in Mireault.dat:

Tests of Between-Subjects Effects

Dependent Variable: GSIT

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1216.924 ^a	5	243.385	2.923	.013
Intercept	1094707.516	1	1094707.516	13146.193	.000
GENDER	652.727	1	652.727	7.839	.005
GROUP	98.343	2	49.172	.590	.555
GENDER * GROUP	419.722	2	209.861	2.520	.082
Error	30727.305	369	83.272		
Total	1475553.000	375			
Corrected Total	31944.229	374			

a. R Squared = .038 (Adjusted R Squared = .025)

Estimated Marginal Means

GENDER * GROUP

Dependent Variable: GSIT

GENDER	GROUP	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
Male	1	62.367	1.304	59.804	64.931
	2	64.676	1.107	62.500	66.853
	3	63.826	1.903	60.084	67.568
Female	1	62.535	.984	60.600	64.470
	2	60.708	.858	59.020	62.396
	3	58.528	1.521	55.537	61.518

16.23 Analysis of variance on the covariate from Exercise 16.22.

The following is abbreviated SAS output.

General Linear Models Procedure

Dependent Variable: YEARCHOLL

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	13.3477645	2.6695529	2.15	0.0600
Error	292	363.0012288	1.2431549		
Corrected Total	297	376.3489933			

	R-Square	C.V.	Root MSE	YEARCOLL Mean	
	0.035466	41.53258	1.11497	2.6845638	
Source	DF	Type III SS	Mean Square	F Value	Pr > F
GENDER	1	5.95006299	5.95006299	4.79	0.0295
GROUP	2	0.78070431	0.39035216	0.31	0.7308
GENDER*GROUP	2	2.96272310	1.48136155	1.19	0.3052

GENDER	GROUP	YEARCOLL LSMEAN
1	1	2.27906977
1	2	2.53225806
1	3	2.68421053
2	1	2.88888889
2	2	2.85000000
2	3	2.70967742

These data reveal a significant difference between males and females in terms of YearColl. Females are slightly ahead of males. If the first year of college is in fact more stressful than later years, this could account for some of the difference we found in Exercise 16.21.