MANOVA - Equations

Lecture 11

Psy524

Andrew Ainsworth

Data design for MANOVA

IV1	IV2	DV1	DV2	DVM
1	1	S01(11)	S01(11)	S01(11)
		S02(11)	S02(11)	S02(11)
		S03(11)	S03(11)	S03(11)
	2	S04(12)	S04(12)	S04(12)
		S05(12)	S05(12)	S05(12)
		S06(12)	S06(12)	S06(12)
2	1	S07(21)	S07(21)	S07(21)
		S08(21)	S08(21)	S08(21)
		S09(21)	S09(21)	S09(21)
	2	S10(22)	S10(22)	S10(22)
		S11(22)	S11(22)	S11(22)
		S12(22)	S12(22)	S12(22)

Data design for MANOVA

	Mild			Moderate			Severe		
	WRAT-R	*WRAT-A*	*(IQ)*	*WRAT-R*	*WRAT-A*	*(IQ)*	*WRAT-R*	*WRAT-A*	*(IQ)*
	115	108	(110)	100	105	(115)	89	78	(99)
Treatment	98	105	(102)	105	95	(98)	100	85	(102)
	107	98	(100)	95	98	(100)	90	95	(100)
	90	92	(108)	70	80	(100)	65	62	(101
Control	85	95	(115)	85	68	(99)	80	70	(95)
	80	81	(95)	78	82	(105)	72	73	(102)

Steps to MANOVA

- MANOVA is a multivariate generalization of ANOVA, so there are analogous parts to the simpler ANOVA equations

Steps to MANOVA

- ANOVA –

$$\sum_i \sum_j (Y_{ij} - GM_{(y)})^2 = n \sum_j \left(\bar{Y}_j - GM_{(y)} \right)^2 + \sum_i \sum_j (Y_{ij} - \bar{Y}_j)^2$$

$$SS_{Total(y)} = SS_{bg(y)} + SS_{wg(y)}$$

Steps to MANOVA

- When you have more than one IV the interaction looks something like this:

$$n_{km}\sum_{k}\sum_{m}(DT_{km}-GM_{(DT)})^2 = n_k\sum_{k}\left(D_k - GM_{(D)}\right)^2 + n_m\sum_{m}\left(T_m - GM_{(T)}\right)^2 +$$

$$\left[n_{km}\sum_{k}\sum_{m}\left(DT_{km}-GM_{(DT)}\right)^2 - n_k\sum_{k}\left(D_k - GM_{(D)}\right)^2 - n_m\sum_{m}\left(T_m - GM_{(T)}\right)^2\right]$$

$$SS_{bg} = SS_D + SS_T + SS_{DT}$$

Steps to MANOVA

- The full factorial design is:

$$\sum_i \sum_k \sum_m (Y_{ikm} - GM_{(ikm)})^2 = n_k \sum_k \left(D_k - GM_{(D)} \right)^2 + n_m \sum_m \left(T_m - GM_{(T)} \right)^2 +$$

$$\left[n_{km} \sum_k \sum_m \left(DT_{km} - GM_{(DT)} \right)^2 - n_k \sum_k \left(D_k - GM_{(D)} \right)^2 - n_m \sum_m \left(T_m - GM_{(T)} \right)^2 \right]$$

$$+ \sum_i \sum_k \sum_m (Y_{ikm} - DT_{km})^2$$

Steps to MANOVA

- MANOVA - You need to think in terms of matrices
  - Each subject now has multiple scores, there is a matrix of responses in each cell
  - Matrices of difference scores are calculated and the matrix squared
  - When the squared differences are summed you get a sum-of-squares-and-cross-products-matrix (**S**) which is the matrix counterpart to the sums of squares.
  - The determinants of the various **S** matrices are found and ratios between them are used to test hypotheses about the effects of the IVs on linear combination(s) of the DVs
  - In MANCOVA the **S** matrices are adjusted for by one or more covariates

Matrix Equations

- If you take the three subjects in the treatment/mild disability cell:

$$Y_{i11} = \begin{bmatrix} 115 \\ 108 \end{bmatrix} \begin{bmatrix} 98 \\ 105 \end{bmatrix} \begin{bmatrix} 107 \\ 98 \end{bmatrix}$$

Matrix Equations

- You can get the means for disability by averaging over subjects and treatments

$$D_1 = \begin{bmatrix} 95.83 \\ 96.50 \end{bmatrix} D_2 = \begin{bmatrix} 88.83 \\ 88.00 \end{bmatrix} D_3 = \begin{bmatrix} 82.67 \\ 77.17 \end{bmatrix}$$

Matrix Equations

- Means for treatment by averaging over subjects and disabilities

$$T_1 = \begin{bmatrix} 99.89 \\ 96.33 \end{bmatrix} T_2 = \begin{bmatrix} 78.33 \\ 78.11 \end{bmatrix}$$

Matrix Equations

- The grand mean is found by averaging over subjects, disabilities and treatments.

$$GM = \begin{bmatrix} 89.11 \\ 87.22 \end{bmatrix}$$

Matrix Equations

- Differences are found by subtracting the matrices, for the first child in the mild/treatment group:

$$(Y_{111} - GM) = \begin{bmatrix} 115 \\ 108 \end{bmatrix} - \begin{bmatrix} 89.11 \\ 87.22 \end{bmatrix} = \begin{bmatrix} 25.89 \\ 20.75 \end{bmatrix}$$

Matrix Equations

- Instead of squaring the matrices you simply multiply the matrix by its transpose, for the first child in the mild/treatment group:

$$(Y_{111} - GM)(Y_{111} - GM)' = \begin{bmatrix} 25.89 \\ 20.75 \end{bmatrix} \begin{bmatrix} 25.89 & 20.75 \end{bmatrix} = \begin{bmatrix} 670.29 & 537.99 \\ 537.99 & 431.81 \end{bmatrix}$$

Matrix Equations

- This is done on the whole data set at the same time, reorganaizing the data to get the appropriate S matrices. The full break of sums of squares is:

$$\sum_i \sum_k \sum_m (Y_{ikm} - GM)(Y_{ikm} - GM)' = n_k \sum_k (D_k - GM)(D_k - GM)'$$

$$+ n_m \sum_m (T_m - GM)(T_m - GM)' + [n_{km} \sum_{km} (DT_{km} - GM)(DT_{km} - GM)'$$

$$- n_k \sum_k (D_k - GM)(D_k - GM)' - n_m \sum_m (T_m - GM)(T_m - GM)']$$

$$+ \sum_i \sum_k \sum_m (Y_{ikm} - DT_{km})(Y_{ikm} - DT_{km})'$$

Matrix Equations

- If you go through this for every combination in the data you will get four S matrices (not including the S total matrix):

$$S_D = \begin{bmatrix} 570.29 & 761.72 \\ 761.72 & 1126.78 \end{bmatrix} S_T = \begin{bmatrix} 2090.89 & 1767.56 \\ 1767.56 & 1494.22 \end{bmatrix}$$

$$S_{DT} = \begin{bmatrix} 2.11 & 5.28 \\ 5.28 & 52.78 \end{bmatrix} S_{s/DT} = \begin{bmatrix} 544.00 & 31.00 \\ 31.00 & 539.33 \end{bmatrix}$$

Test Statistic – Wilk’s Lambda

- Once the S matrices are found the diagonals represent the sums of squares and the off diagonals are the cross products
- The determinant of each matrix represents the generalized variance for that matrix
- Using the determinants we can test for significance, there are many ways to this and one of the most is Wilk’s Lambda

Test Statistic – Wilk's Lambda

$$\Lambda = \frac{\left|S_{error}\right|}{\left|S_{effect} + S_{error}\right|}$$

- this can be seen as the percent of non-overlap between the effect and the DVs.

Test Statistic – Wilk's Lambda

- For the interaction this would be:

$$\left|S_{s/DT}\right| = 292436.52$$

$$S_{DT} + S_{s/DT} = \begin{bmatrix} 2.11 & 5.28 \\ 5.28 & 52.78 \end{bmatrix} + \begin{bmatrix} 544.00 & 31.00 \\ 31.00 & 539.33 \end{bmatrix} = \begin{bmatrix} 546.11 & 36.28 \\ 36.28 & 529.11 \end{bmatrix}$$

$$\left|S_{DT} + S_{s/DT}\right| = 322040.95$$

$$\Lambda = \frac{292436.52}{322040.95} = .908068$$

Test Statistic – Wilk's Lambda

- Approximate Multivariate F for Wilk's Lambda is

$$F(df_1, df_2) = \left(\frac{1-y}{y}\right)\left(\frac{df_2}{df_1}\right)$$

$$where\ y = \Lambda^{1/s},\ s = \sqrt{\frac{p^2(df_{effect})^2 - 4}{p^2 + (df_{effect})^2 - 5}},$$

$$p = number\ of\ DVs,\ df_1 = p(df_{effect})$$

$$df_2 = s\left[(df_{error}) - \frac{p - df_{effect} + 1}{2}\right] - \left[\frac{p(df_{effect}) - 2}{2}\right]$$

Test Statistic – Wilk's Lambda

- So in the example for the interaction:

$$s = \sqrt{\frac{(2)^2(2)^2 - 4}{(2)^2 + (2)^2 - 5}} = 2$$

$$y = .908068^{1/2} = .952926$$

$$df_1 = 2(2) = 4$$

$$df_{error} = dt(n-1) = 3(2)(3-1) = 12$$

$$df_2 = 2\left[12 - \frac{2-2+1}{2}\right] - \left[\frac{2(2)-2}{2}\right] = 22$$

$$Approximate\ \text{F}(4{,}22) = \left(\frac{.047074}{.952926}\right)\left(\frac{22}{4}\right) = 0.2717$$

Eta Squared

$$\eta^2 = 1 - \Lambda$$

Partial Eta Squared

$$\eta^2 = 1 - \Lambda^{1/s}$$