Multivariate Analysis of Variance

Today . . .

- MANOVA—finding linear combinations that best separate groups (hey, wait a minute, didn't we just . . .)
- Where does MANOVA fit among the other ways that a data set could be analyzed?
- How does MANOVA extend univariate approaches?

To explore these questions, consider the ways that the following data could be analyzed:

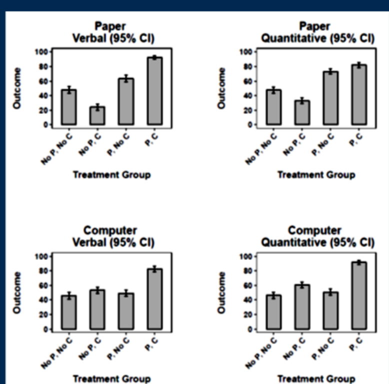
Participants are randomly assigned to one of four groups that get different kinds of training for taking tests: no training, standard paper-pencil training, computer-based training, and both standard training and computer-based training.

Following training, all participants complete a test of verbal ability and a test of quantitative ability in standard paper-and-pencil format and in computer format (all counterbalanced).

Within-Subjects

Between-Groups

	Standard Measure of Verbal Ability	Standard Measure of Quantitative Ability	Computer Measure of Verbal Ability	Computer Measure of Quantitative Ability
No Training				
Standard Training Only				
Computer Training Only				
Both				



The groups appear to be different in their patterns of outcomes.

Within-Subjects

Between-Groups

	Standard Measure of Verbal Ability	Standard Measure of Quantitative Ability	Computer Measure of Verbal Ability	Computer Measure of Quantitative Ability
No Training				
Standard Training Only				
Computer Training Only				
Both				

With 16 means, there are 15 degrees of freedom and a lot of choices for partitioning this variance.

	Standard Measure of Verbal Ability	Standard Measure of Quantitative Ability	Computer Measure of Verbal Ability	Computer Measure of Quantitative Ability
No Training				
Standard Training Only				
Computer Training Only				
Both				

	No Computer Training	Computer Training
No Standard Training		
Standard Training		

There is an inherent factorial structure to the groups. This structure can be incorporated into the analysis.

Questions

Groups

	Computer Training?	Standard Training?	Both or None?
No Training	No	No	Yes
Standard Training	No	Yes	No
Computer Training	Yes	No	No
Standard and Computer Training	Yes	Yes	Yes

Questions

Groups

	Computer Training?	Standard Training?	Both or None?
No Training	-1	-1	1
Standard Training	-1	1	-1
Computer Training	1	-1	-1
Standard and Computer Training	1	1	1

		Questions		
		Computer Training?	Standard Training?	Both or None?
Groups	No Training	-1	-1	1
	Standard Training	-1	1	-1
	Computer Training	1	-1	-1
	Standard and Computer Training	1	1	1

These contrast codes create three separate two-group comparisons. They are the standard main effects and interaction from ANOVA.

Group 1	Group 2	Group 3	. . .	Group G
M_1 SD_1	M_2 SD_2	M_3 SD_4		M_G SD_G

$$MS_{\text{Within Groups}} = \frac{SS_{\text{Within Groups}}}{N-k}$$

$$MS_{\text{Between Groups}} = \frac{SS_{\text{Between Groups}}}{k-1}$$

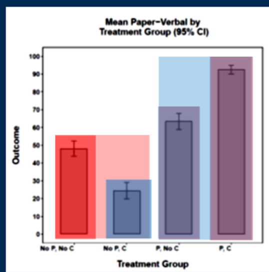
$$F = \frac{MS_{\text{Between Groups}}}{MS_{\text{Within Groups}}}$$

$$\hat{\sigma}^2_{\bar{Y}} = \frac{\hat{\sigma}^2_Y}{N}$$

$$N\hat{\sigma}^2_{\bar{Y}} = \hat{\sigma}^2_Y$$

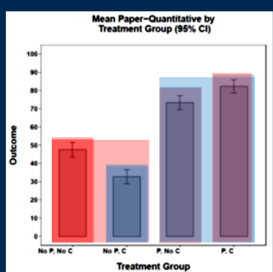
	No Computer Training	Computer Training	
No Standard Training	M_{11}	M_{12}	$M_{1.}$
Standard Training	M_{21}	M_{22}	$M_{2.}$
	$M_{.1}$	$M_{.2}$	

The same principles hold for factorial designs. The *F* statistic is the ratio of two estimates of random variability under the null hypothesis, with the numerator an estimate based on variability of means. The variability of means could be based on marginal means.



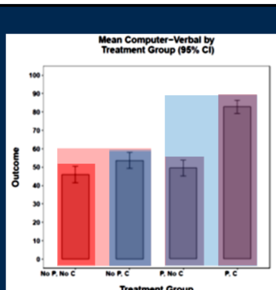
	No Computer Training	Computer Training	
No Standard Training	M_{11}	M_{12}	$M_{1.}$
Standard Training	M_{21}	M_{22}	$M_{2.}$
	$M_{.1}$	$M_{.2}$	

```
AOV_5 <- aov(P_Verbal ~ Tx_P * Tx_C, data = Skills_Trimmed)
summary(AOV_5)

##             Df Sum Sq Mean Sq F value Pr(>F)
## Tx_P         1  43991   43991  456.58 <2e-16
## Tx_C         1    150     150    1.55   0.22
## Tx_P:Tx_C    1  17267   17267  179.21 <2e-16
## Residuals   95   9153      96
```



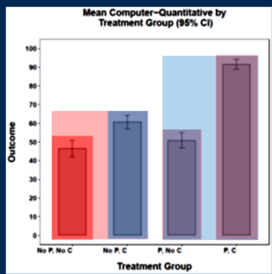
	No Computer Training	Computer Training	
No Standard Training	M_{11}	M_{12}	$M_{1.}$
Standard Training	M_{21}	M_{22}	$M_{2.}$
	$M_{.1}$	$M_{.2}$	

```
AOV_6 <- aov(P_Quant ~ Tx_P * Tx_C, data = Skills_Trimmed)
summary(AOV_6)

##             Df Sum Sq Mean Sq F value  Pr(>F)
## Tx_P         1  34759   34759  405.53 < 2e-16
## Tx_C         1    237     237    2.76     0.1
## Tx_P:Tx_C    1   3423    3423   39.94 8.4e-09
## Residuals   95   8143      86
```



	No Computer Training	Computer Training	
No Standard Training	M_{11}	M_{12}	$M_{1.}$
Standard Training	M_{21}	M_{22}	$M_{2.}$
	$M_{.1}$	$M_{.2}$	

```
AOV_7 <- aov(C_Verbal ~ Tx_P * Tx_C, data = Skills_Trimmed)
summary(AOV_7)

##             Df Sum Sq Mean Sq F value  Pr(>F)
## Tx_P         1   6833    6833    67.3 1.1e-12
## Tx_C         1  10201   10201   100.5 < 2e-16
## Tx_P:Tx_C    1   4049    4049    39.9 8.6e-09
## Residuals   95   9644     102
```



	No Computer Training	Computer Training	
No Standard Training	M_{11}	M_{12}	$M_{1.}$
Standard Training	M_{21}	M_{22}	$M_{2.}$
	$M_{.1}$	$M_{.2}$	

```
AOV_8 <- aov(C_Quant ~ Tx_P * Tx_C, data = Skills_Trimmed)
summary(AOV_8)

##             Df Sum Sq Mean Sq F value  Pr(>F)
## Tx_P         1   8099    8099    98.5 2.4e-16
## Tx_C         1  18593   18593   226.1 < 2e-16
## Tx_P:Tx_C    1   4324    4324    52.6 1.1e-10
## Residuals   95   7813      82
```

	Verbal Ability	Quantitative Ability
Standard Measure		
Computer Measure		

	Standard Measure of Verbal Ability	Standard Measure of Quantitative Ability	Computer Measure of Verbal Ability	Computer Measure of Quantitative Ability
No Training				
Standard Training Only				
Computer Training Only				
Both				

The outcome measures also have an inherent structure that can be built into the analysis.

Question

Measure	Computer Measure?	Quantitative Ability?	Both or Neither?
Standard Measure of Verbal	No	No	Yes
Standard Measure of Quantitative	No	Yes	No
Computer Measure of Verbal	Yes	No	No
Computer Measure of Quantitative	Yes	Yes	Yes

Question

Measure	Computer Measure?	Quantitative Ability?	Both or Neither?
Standard Measure of Verbal	-1	-1	1
Standard Measure of Quantitative	-1	1	-1
Computer Measure of Verbal	1	-1	-1
Computer Measure of Quantitative	1	1	1

Question

Measure	Computer Measure?	Quantitative Ability?	Both or Neither?
Standard Measure of Verbal	-1	-1	1
Standard Measure of Quantitative	-1	1	-1
Computer Measure of Verbal	1	-1	-1
Computer Measure of Quantitative	1	1	1

The contrasts create three new linear combinations of means. They represent the standard main effects and interaction from a repeated measures ANOVA. We can also simply sum the measures.

	No Computer Training	Computer Training
No Standard Training		
Standard Training		

X

	Verbal Ability	Quantitative Ability
Standard Measure		
Computer Measure		

In combination, the two smaller designs create all main effects and interactions in the larger 2 x 2 x 2 x 2 design.

Each effect is a single degree of freedom—15 in all (4 main, 6 two-way, 4 three-way, 1 four-way).

	Verbal Ability	Quantitative Ability	
Standard Measure	M_{11}	M_{12}	$M_{1.}$
Computer Measure	M_{21}	M_{22}	$M_{2.}$
	$M_{.1}$	$M_{.2}$	

Note that when the linear combination has weights that sum to 0, the intercept is a meaningful test (a repeated measures main effect).

```
# Mode: The difference between paper and computer measures.
AOV_14 <- aov(I(P_Verbal + P_Quant - C_Verbal - C_Quant) ~ Tx_P_NF +
    Tx_C_NF + I(Tx_P_NF * Tx_C_NF), data = Skills_Trimmed)
summary(AOV_14, intercept = TRUE)

##                      Df Sum Sq Mean Sq F value   Pr(>F)
## (Intercept)           1   2086    2086     9.9   0.0022
## Tx_P_NF               1  49961   49961   237.1  < 2e-16
## Tx_C_NF               1  57851   57851   274.5  < 2e-16
## I(Tx_P_NF * Tx_C_NF)  1   3663    3663    17.4 0.000068
## Residuals            95  20020     211
```

	Verbal Ability	Quantitative Ability	
Standard Measure	M_{11}	M_{12}	$M_{1.}$
Computer Measure	M_{21}	M_{22}	$M_{2.}$
	$M_{.1}$	$M_{.2}$	

Now the linear combination is a test of verbal versus quantitative abilities.

```
# Domain: The difference between verbal and quantitative measures.
AOV_15 <- aov(I(P_Verbal - P_Quant + C_Verbal - C_Quant) ~ Tx_P_NF +
    Tx_C_NF + I(Tx_P_NF * Tx_C_NF), data = Skills_Trimmed)
summary(AOV_15, intercept = TRUE)

##                      Df Sum Sq Mean Sq F value  Pr(>F)
## (Intercept)           1   4125    4125   63.55 3.5e-12
## Tx_P_NF               1    255     255    3.93    0.05
## Tx_C_NF               1     60      60    0.92    0.34
## I(Tx_P_NF * Tx_C_NF)  1   5009    5009   77.17 6.5e-14
## Residuals            95   6166      65
```

	Verbal Ability	Quantitative Ability	
Standard Measure	M_{11}	M_{12}	$M_{1.}$
Computer Measure	M_{21}	M_{22}	$M_{2.}$
	$M_{.1}$	$M_{.2}$	

The outcome linear combination is the interaction in the within-subjects part of the design. All effects in this table are interactions.

```
# Mode x Domain interaction
AOV_16 <- aov(I(P_Verbal - P_Quant - C_Verbal + C_Quant) ~ Tx_P_NF +
    Tx_C_NF + I(Tx_P_NF * Tx_C_NF), data = Skills_Trimmed)
summary(AOV_16, intercept = TRUE)

##                      Df Sum Sq Mean Sq F value  Pr(>F)
## (Intercept)           1    786     786    10.7 0.00149
## Tx_P_NF               1    939     939    12.8 0.00055
## Tx_C_NF               1   3966    3966    54.0 6.9e-11
## I(Tx_P_NF * Tx_C_NF)  1   5627    5627    76.7 7.5e-14
## Residuals            95   6974      73
```

```r
# Sum over repeated measures.  This produces the between-subjects
# part of the design.
AOV_13 <- aov(I(P_Verbal + P_Quant + C_Verbal + C_Quant) ~ Tx_P_NF +
    Tx_C_NF + I(Tx_P_NF * Tx_C_NF), data = Skills_Trimmed)
summary(AOV_13, intercept = TRUE)

##                      Df  Sum Sq Mean Sq F value  Pr(>F)
## (Intercept)           1 5502086 5502086  4938.0 < 2e-16
## Tx_P_NF               1  323576  323576   290.4 < 2e-16
## Tx_C_NF               1   54846   54846    49.2 3.3e-10
## I(Tx_P_NF * Tx_C_NF)  1  101954  101954    91.5 1.4e-15
## Residuals            95  105852    1114
```

When the outcome measures are summed, the analysis becomes the between-groups part of the design. In this case, the intercept test is not meaningful. It merely tests if the grand mean is different from 0.

In a factorial, A x B, design, the sums of squares methods are:

- Type I SS:
  - SS(A)
  - SS(B|A)
  - SS(AxB|A,B)
- Type II SS:
  - SS(A|B)
  - SS(B|A)
  - SS(AxB|A,B)
- Type III SS
  - SS(A|B,AxB)
  - SS(B|A,AxB)
  - SS(AxB|A,B)

Different functions in R will use different defaults. If the design is balanced, it won't matter, but if the predictors are correlated, then this must be given some careful attention.

By constructing single degree of freedom effects, we have completely defined the problem in a way with no freedom to explore the data.

The group distinctions are locked into traditional main effects and the interaction—each a simple comparison of one "group" to another "group."

The linear combinations on the outcome side are likewise restrictive, comprising simple differences among collections of means.

In a way, this a very simple and very restrictive discriminant analysis.

The factorial design imposes structure on the data.

How might we relax the restrictions on the dependent variable side to allow an alternative discriminant analysis?

How might we relax the restrictions on the independent variable side to allow an alternative discriminant analysis?

What do we lose by relaxing the restrictions imposed by the factorial designs?

By relaxing the constraints imposed by both factorial designs, we arrive at a traditional discriminant analysis: "Are there linear combinations of the outcome variables that produce maximum separation of the groups?"

Both the traditional ANOVA and the discriminant analyses operate on linear combinations. They differ in whether the weights are imposed (guided by theory or design) or derived (identified by the analysis). Both are "best" in an important way.

How might the two approaches be blended?

One difference between MANOVA and discriminant analysis is that the groups to be discriminated in MANOVA are often under experimental control.

A second (trivial) difference is that the groups are usually labeled as the independent variable rather than the outcome.

The goal, however, is the same: Find independent linear combinations of variables that best separate the groups (however they are defined).

MANOVA has traditionally relied on a broader number of statistical tests for determining the significance of the linear combinations and for determining the nature of the discrimination that is identified.

Four common tests of significance represent generalizations of the univariate approach to significance testing.

In the univariate model, an F test gauges the amount of between-groups variability to within-groups variability.

Group 1	Group 2	Group 3	. . .	Group G
M_1 SD_1	M_2 SD_2	M_3 SD_4		M_G SD_G

$$MS_{\text{Within Groups}} = \frac{SS_{\text{Within Groups}}}{N-k}$$

$$MS_{\text{Between Groups}} = \frac{SS_{\text{Between Groups}}}{k-1}$$

$$F = \frac{MS_{\text{Between Groups}}}{MS_{\text{Within Groups}}}$$

$$\hat{\sigma}^2_{\bar{Y}} = \frac{\hat{\sigma}^2_Y}{N}$$

$$N\hat{\sigma}^2_{\bar{Y}} = \hat{\sigma}^2_Y$$

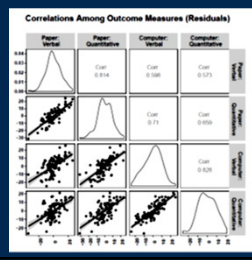
In the multivariate model, the single sums of squares are replaced by matrices of sums of squares and cross-products for the effect or hypothesis (**H)** and for error (**E**). Different combinations of these matrices define the four most common tests of significance:

- Roy's test is a function of the largest eigenvalue of HE^{-1}
- Lawley and Hotelling's Trace = $tr(HE^{-1})$
- Pillai's test is a function of $tr(H(H+E)^{-1})$
- Wilk's likelihood test is based on $\Lambda = |E|/|H+E|$

The latter three usually agree because they use much the same information.

There are two other problems that MANOVA can help solve:

- The multiple ANOVAs approach, with each DV analyzed separately, produces a large number of significance tests. These tests will not produce independent assessments if the measures are correlated. This approach can also produce a large number of significance tests (and inflate the Type I error rate).



- The multiple outcomes may not be in the same metric, eliminating the repeated measures option. Even if in the same metric, the comparisons may not make conceptual sense. MANOVA provides additional interesting ways to view the data that overcome this “metric” problem.

Next time . . .

Example and extending the interpretations from discriminant analysis