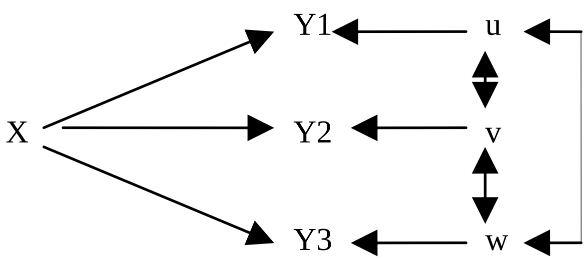
Brief Overview of Manova

In this and other handouts, we'll briefly go over some advanced techniques that can be useful when estimating complicated models. We won't discuss these in detail, but at least you'll know what to look up should you encounter such problems in your research.

MULTIPLE DEPENDENT VARIABLES: MANOVA. We are often interested in models such as the following:



In this model, there are multiple dependent variables. The IV, X, affects each of them. However, their residuals are also correlated, presumably because of the influence of other variables omitted from the model. The disturbances are connected to each other by two-sided arrows because they are assumed to be correlated, but without a specification of which disturbance is a cause and which is an effect of the others.

A common situation in which this occurs is when X is a "treatment" variable, and is coded 0-1 (subject is or is not a member of the treatment group). Why might we want to have more than one dependent variable? In many cases, researchers are not interested in a single measure of group differences. Rather, there are often several components, constructs or behaviors that might be affected by the treatment or that are useful to separate the groups.

For example, if we wanted to evaluate the effects of a training program to increase assertiveness, we might be interested in the effects of the program on (1) assertive behavior, (2) anxiety about being assertive, and (3) self-esteem.

Another example: Julie Hart's dissertation looks at the effect of a peer mediation program on conflict within schools. Her dependent variables include such things as perceptions of school safety, number of reported conflicts, time spent on discipline rather than teaching, etc.

Recall that, in ANOVA, one evaluates mean differences between groups on a single dependent variable. This can also be done with OLS regression, using dummy independent variables. With MANOVA (multivariate ANOVA) one evaluates mean differences on two or more dependent variables simultaneously.

Put another way, with ANOVA we test

H_0: $\mu_1 = \mu_2$

where the subscripts refer to the group. We would do this three times, once for each dependent variable. With Manova, we test

$$H_0: \quad \mu_{11} = \mu_{21}$$
$$\mu_{12} = \mu_{22}$$
$$\mu_{13} = \mu_{23}$$

where the first subscript refers to the group and the second subscript refers to the variable number. Alternatively, we can think of Manova as testing the hypothesis that

$$H_0: \quad \beta_{1X} = \beta_{2X} = \beta_{3X} = 0$$

This may look familiar, but actually we have never tested a hypothesis such as this. We have tested whether, for a *single* dependent variable, *one or more IVs* have zero effects. Here we are testing whether, for *different* dependent variables, the *same IV* has zero effects.

MANOVA is preferable to multiple ANOVAs (or regressions with dummy variables) because

- Multiple ANOVA/OLS runs can capitalize on chance. For example, if you have 20 dependent variables, you expect X to have a "significant" effect on one of them if $\alpha = .05$. MANOVA does a global test of whether group means differ for any of the variables.

- ANOVA/OLS ignore the intercorrelations between the IVs. Because MANOVA takes them into account, it can provide a more powerful statistical test. Manova uses more information about the data than ANOVA does.

Another common use of MANOVA is in a repeated measures design, where the same variable is measured at different points in time. For example, in Hart's dissertation, she looked at conflict levels both before and after the experimental program was introduced.

Here is a quick example, using the SPSS MANOVA program. X = Race. The Y variables are education, job experience, and income.

```
-> MANOVA
->   educ jobexp income  BY black(0 1)
->   /PRINT SIGNIF(MULT UNIV )
->   /NOPRINT PARAM(ESTIM)
->   /METHOD=UNIQUE
->   /ERROR WITHIN+RESIDUAL
->   /DESIGN  .
* * * * * * A n a l y s i s   o f   V a r i a n c e * * * * * *

 EFFECT .. BLACK
 Multivariate Tests of Significance (S = 1, M = 1/2, N = 247 )

 Test Name         Value    Exact F Hypoth. DF   Error DF  Sig. of F

 Pillais          .25972   58.00628       3.00     496.00       .000
 Hotellings       .35084   58.00628       3.00     496.00       .000
 Wilks            .74028   58.00628       3.00     496.00       .000
 Roys             .25972
 Note.. F statistics are exact.

 - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
 EFFECT .. BLACK (Cont.)
 Univariate F-tests with (1,498) D. F.

 Variable   Hypoth. SS   Error SS Hypoth. MS   Error MS          F  Sig. of F

 EDUC       1095.20000 6812.00000 1095.20000   13.67871   80.06600       .000
 JOBEXP      672.80000 12112.0000  672.80000   24.32129   27.66301       .000
 INCOME     10125.0000 30056.2500 10125.0000   60.35392  167.76045       .000
```

The first part of the results, labeled "multivariate tests of significance," gives various global statistics for testing whether black and white means differ on any of the three dependent variables. The next part, labeled "univariate F tests," examines each dependent variable separately. No matter what you look at, it is pretty clear that there are significant racial differences on all three of the DVs.

Using Stata,

```
. use http://www.nd.edu/~rwilliam/stats2/statafiles/blwh.dta, clear
. manova income educ jobexp = black, continuous(black)

                           Number of obs =     500

                           W = Wilks' lambda      L = Lawley-Hotelling trace
                           P = Pillai's trace     R = Roy's largest root

              Source |  Statistic     df   F(df1,    df2) =   F   Prob>F
          -----------+--------------------------------------------------
               black | W   0.7403      1     3.0   496.0    58.01 0.0000 e
                     | P   0.2597            3.0   496.0    58.01 0.0000 e
                     | L   0.3508            3.0   496.0    58.01 0.0000 e
                     | R   0.3508            3.0   496.0    58.01 0.0000 e
                     |--------------------------------------------------
            Residual |               498
          -----------+--------------------------------------------------
               Total |               499
          --------------------------------------------------------------
                           e = exact, a = approximate, u = upper bound on F
```

For much more detail on MANOVA, see Multivariate Analysis of Variance, by James H. Bray and Scott E. Maxwell (paper #54 in the Sage Series On Quantitative Applications In The Social Sciences.)