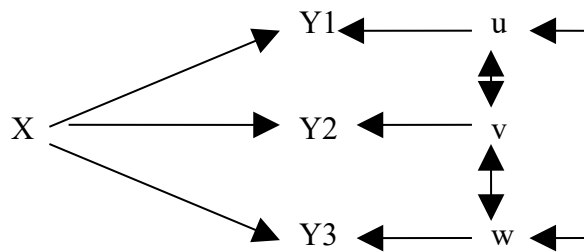


Brief Overview of Manova

Richard Williams, University of Notre Dame, <https://academicweb.nd.edu/~rwilliam/>
Last revised April 6, 2015

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

$$\begin{aligned} H_0: \quad & \mu_{11} = \mu_{21} \\ & \mu_{12} = \mu_{22} \\ & \mu_{13} = \mu_{23} \end{aligned}$$

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. X = black. The Y variables are income, education and job experience. The `manova` command provides various global statistics for testing whether the black and white means differ on any of the three dependent variables (when there are only two groups, as is the case here, all the tests are equivalent to Hotelling’s T-squared, which tests whether a set of means is equal between two groups). These global tests show that there are significant racial differences on at least one of the dependent variables. The `mvreg` command gives us the actual coefficient estimates along with tests for each dependent variable separately. It is clear that there are significant racial differences on all three of the DVs.

```
. use https://academicweb.nd.edu/~rwilliam/statafiles/blwh.dta, clear
. manova income educ jobexp = i.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

```
. mvreg
```

Equation	Obs	Parms	RMSE	"R-sq"	F	P
income	500	2	7.768778	0.2520	167.7605	0.0000
educ	500	2	3.698475	0.1385	80.066	0.0000
jobexp	500	2	4.931661	0.0526	27.66301	0.0000

	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
income						
1.black	-11.25	.8685758	-12.95	0.000	-12.95652	-9.543475
_cons	30.04	.3884389	77.34	0.000	29.27682	30.80318
educ						
1.black	-3.7	.413502	-8.95	0.000	-4.512424	-2.887576
_cons	13.9	.1849237	75.17	0.000	13.53667	14.26333
jobexp						
1.black	-2.9	.5513765	-5.26	0.000	-3.983311	-1.816689
_cons	14.1	.2465831	57.18	0.000	13.61553	14.58447

You can also use the `sem` command to estimate a MANOVA model. The `covstructure` option allows the residuals for the three dependent variables to be freely correlated.

```
. sem black -> income educ jobexp, covstructure(e._En, unstructured)
```

Endogenous variables

Observed: income educ jobexp

Exogenous variables

Observed: black

