MANOVA: Multivariate Analysis of Variance

or Multiple Analysis of Variance

D. Gordon E. Robertson, PhD,
School of Human Kinetics,
University of Ottawa

Review of ANOVA: Univariate Analysis of Variance

- Simple ANOVAs are used to investigate whether or not there is a difference in the scores of a **single** dependent variable (DV) that is due to membership in a group in comparison to two or more groups. I.e., are there significant differences between three or more independent groups based on a single independent variable.

- The independent variable (IV) is a nominal quantity; the dependent variable should be an ordinal or ratio/interval quantity.

Example:
Stair descent study

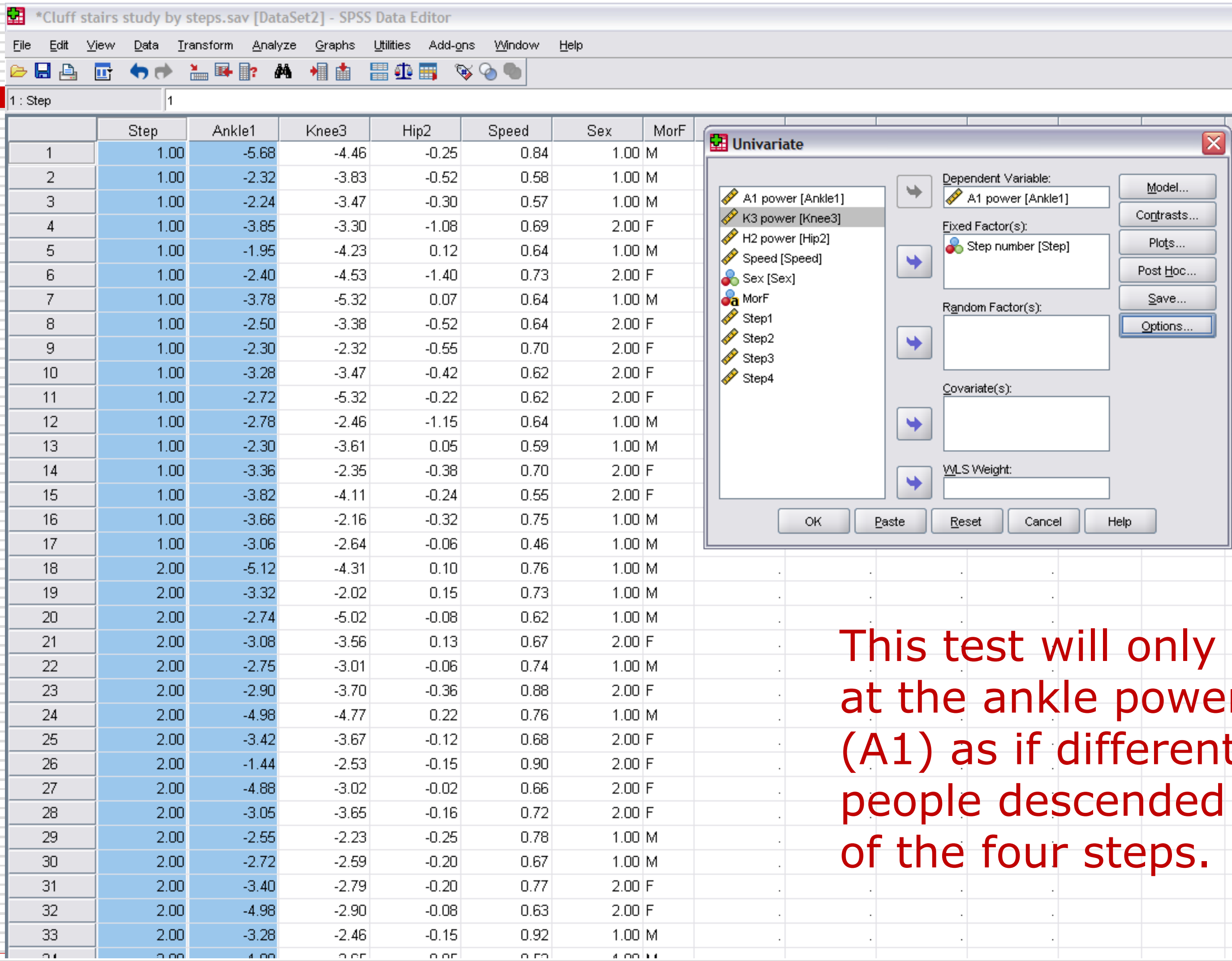
Cluff and Robertson (2011) Gait & Posture. 33:423-8.

- Seventeen subjects (9 males, 8 females)
- Descend a f… for last step… analyzed.
- Three meas… recorded, o… plantar flex… extensors, and one by the hip flexors

Design

- Dependent variable(s):
  - A1 – peak ankle plantar flexor, negative power
  - K3 – peak knee extensor, negative power
  - H2 – peak hip flexor, positive power
- Independent variables:
  - Steps (1 to 4)
  - Sex (M or F)
- Designs to be tested:
  - One-way factorial
  - One-way repeated-measures
  - Two-way repeated-measures MANOVA (step, joint)
  - Three-way mixed MANOVA (by adding sex)

One-way Factorial ANOVA



	Step	Ankle1	Knee3	Hip2	Speed	Sex	MorF
1	1.00	-5.68	-4.46	-0.25	0.84	1.00	M
2	1.00	-2.32	-3.83	-0.52	0.58	1.00	M
3	1.00	-2.24	-3.47	-0.30	0.57	1.00	M
4	1.00	-3.85	-3.30	-1.08	0.69	2.00	F
5	1.00	-1.95	-4.23	0.12	0.64	1.00	M
6	1.00	-2.40	-4.53	-1.40	0.73	2.00	F
7	1.00	-3.78	-5.32	0.07	0.64	1.00	M
8	1.00	-2.50	-3.38	-0.52	0.64	2.00	F
9	1.00	-2.30	-2.32	-0.55	0.70	2.00	F
10	1.00	-3.28	-3.47	-0.42	0.62	2.00	F
11	1.00	-2.72	-5.32	-0.22	0.62	2.00	F
12	1.00	-2.78	-2.46	-1.15	0.64	1.00	M
13	1.00	-2.30	-3.61	0.05	0.59	1.00	M
14	1.00	-3.36	-2.35	-0.38	0.70	2.00	F
15	1.00	-3.82	-4.11	-0.24	0.55	2.00	F
16	1.00	-3.66	-2.16	-0.32	0.75	1.00	M
17	1.00	-3.06	-2.64	-0.06	0.46	1.00	M
18	2.00	-5.12	-4.31	0.10	0.76	1.00	M
19	2.00	-3.32	-2.02	0.15	0.73	1.00	M
20	2.00	-2.74	-5.02	-0.08	0.62	1.00	M
21	2.00	-3.08	-3.56	0.13	0.67	2.00	F
22	2.00	-2.75	-3.01	-0.06	0.74	1.00	M
23	2.00	-2.90	-3.70	-0.36	0.88	2.00	F
24	2.00	-4.98	-4.77	0.22	0.76	1.00	M
25	2.00	-3.42	-3.67	-0.12	0.68	2.00	F
26	2.00	-1.44	-2.53	-0.15	0.90	2.00	F
27	2.00	-4.88	-3.02	-0.02	0.66	2.00	F
28	2.00	-3.05	-3.65	-0.16	0.72	2.00	F
29	2.00	-2.55	-2.23	-0.25	0.78	1.00	M
30	2.00	-2.72	-2.59	-0.20	0.67	1.00	M
31	2.00	-3.40	-2.79	-0.20	0.77	2.00	F
32	2.00	-4.98	-2.90	-0.08	0.63	2.00	F
33	2.00	-3.28	-2.46	-0.15	0.92	1.00	M

This test will only look at the ankle powers (A1) as if different people descended each of the four steps.

Review of ANOVA

- Before examining the F-statistic, check that the ANOVA meets the homogeneity of variance test. SPSS uses Levene's test.

Levene's Test of Equality of Error Variances[a]

Dependent Variable:A1 power

F	df1	df2	Sig.
.819	3	64	.488

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: + Step

- Since significance (i.e., 0.488) is greater than $\alpha = 0.05$, the null hypothesis is rejected so the variances across groups are assumed to be equal.

Basic ANOVA Output

Tests of Between-Subjects Effects
Dependent Variable:A1 power

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power
Cor. model	17.422	3	5.807	5.209	.003	.196	.911
	899.794	1	899.794	807.109	.000	.927	1.000
Step	17.422	3	5.807	5.209	.003	.196	.911
Error	71.350	64	1.115				
Total	988.566	68					
Corrected Total		88.772	67				

The IV, "Step"

A B C D

Important information in the ANOVA output:

A. The F ratio
B. Significance of that F ratio
C. Partial eta squared (estimate of the "effect size" attributable to between-group differences
D. Power to detect the effect (0.911 is very powerful)

Post hoc test of ANOVA

- Since there is a significant F we may do *post hoc* testing. If not significant this CANNOT BE DONE.
- The ANOVA only shows that "at least one" group was different from the others. But which one?
- Since we are going to test all possible pairs the Scheffé test is recommended. It is also the most conservative.

The results show that steps 1 and 2 are significantly different from step 4 but step 3 is not different from any of the other steps.

Multiple Comparisons

A1 power
Scheffe

(I) Step number	(J) Step number	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
1	2	.2664	.36216	.909	-.7735	1.3062
	3	.7135	.36216	.284	-.3263	1.7534
	4	1.3365*	.36216	.006	.2966	2.3763
2	1	-.2664	.36216	.909	-1.3062	.7735
	3	.4472	.36216	.678	-.5927	1.4870
	4	1.0701*	.36216	.041	.0302	2.1100
3	1	-.7135	.36216	.284	-1.7534	.3263
	2	-.4472	.36216	.678	-1.4870	.5927
	4	.6229	.36216	.405	-.4169	1.6628
4	1	-1.3365*	.36216	.006	-2.3763	-.2966
	2	-1.0701*	.36216	.041	-2.1100	-.0302
	3	-.6229	.36216	.405	-1.6628	.4169

Based on observed means.
The error term is Mean Square(Error) = 1.115.

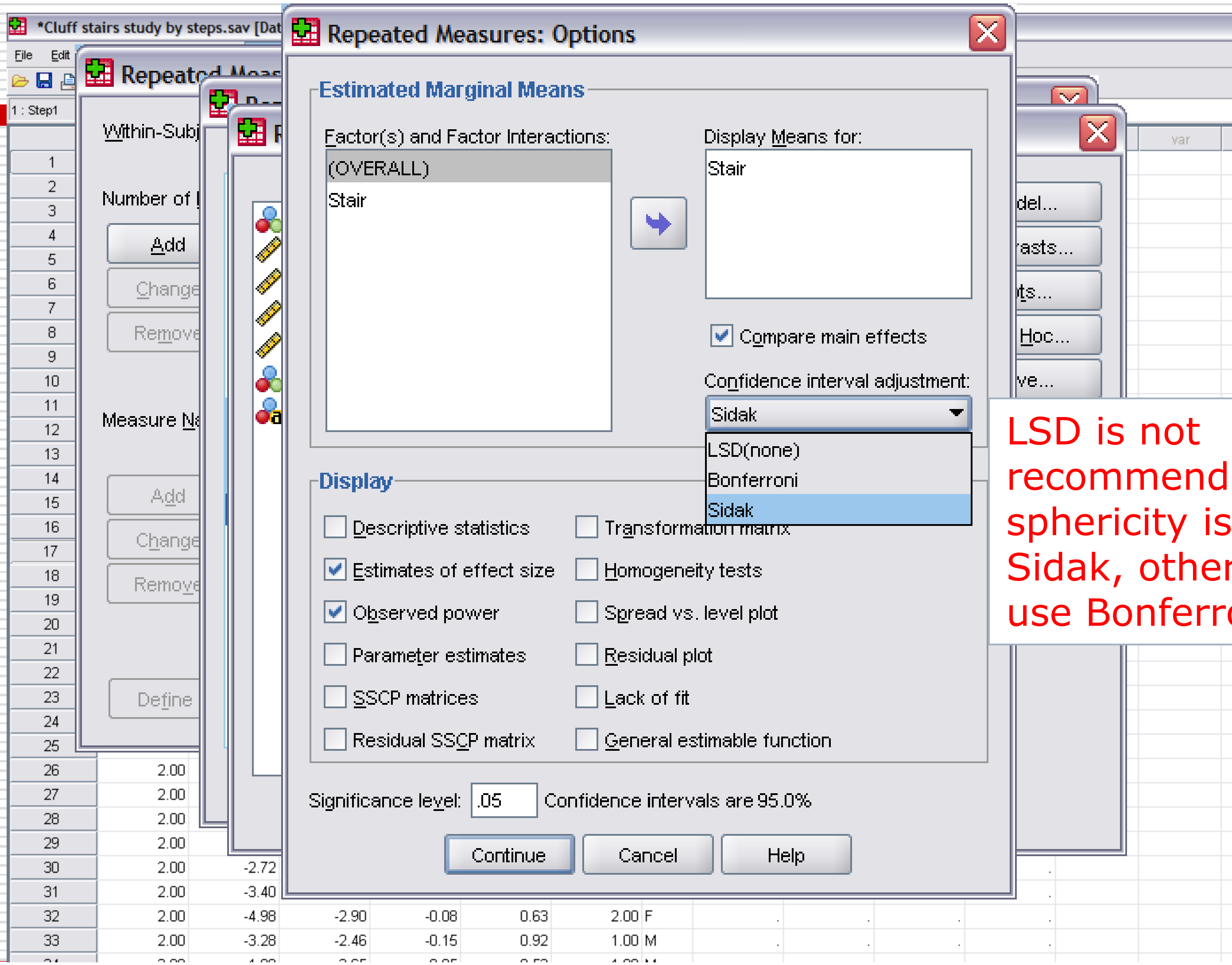
*. The mean difference is significant at the .05 level.

One-way Repeated-measures ANOVA

Now repeat the analysis with the same data but recognizing that the same people were used, i.e., it was a repeated-measures design.

First, repeated-measures requires the homogeneity of covariance's, i.e., the variances of the differences between groups are equal. This is called “sphericity”.

Repeated-measures ANOVA



LSD is not recommended, if sphericity is OK use Sidak, otherwise use Bonferroni.

Repeated-measures ANOVA: Test for sphericity

SPSS uses Mauchly's Test of Sphericity.

Mauchly's Test of Sphericity[b]

Measure:MEASURE_1

Within Subject...	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[a]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Stair	.585	7.893	5	.163	.723	.841	.333

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

b. Design:
Within Subjects Design: Stair

Since the p-value (Sig. = 0.163) is greater than $\alpha = 0.05$, we reject the null hypothesis that covariances are unequal and can "assume sphericity".

Repeated-measures ANOVA: Results

SPSS shows results for four different assumptions. We can choose the first.

Tests of Within-Subjects Effects

Measure:MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Stair	Sphericity Assumed	17.422	3	5.807	27.706	.000
	Greenhouse-Geisser	17.422	2.170	8.029	27.706	.000
	Huynh-Feldt	17.422	2.523	6.907	27.706	.000
	Lower-bound	17.422	1.000	17.422	27.706	.000
Error(Stair)	Sphericity Assumed	10.061	48	.210		
	Greenhouse-Geisser	10.061	34.717	.290		
	Huynh-Feldt	10.061	40.360	.249		
	Lower-bound	10.061	16.000	.629		

Since the p-value (Sig. = .000) is less than α = 0.05, the null hypothesis is rejected and conclude there is a significant difference across stair steps. Note, a p-value of 0.000 is written $p<0.0005$.

Repeated-measures ANOVA: Test for best fit

SPSS shows results of fitting polynomials from linear to degree *k*-1.

Tests of Within-Subjects Contrasts

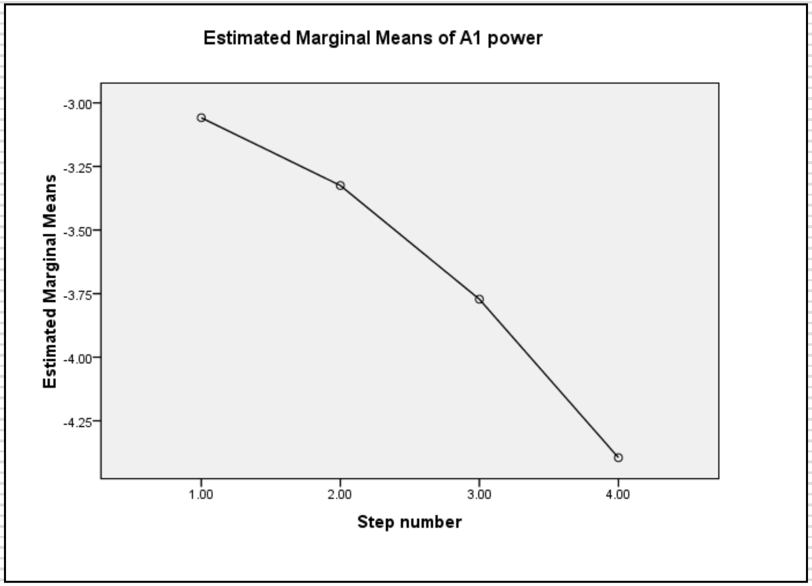
Measure:MEASURE_1

Source	Stair	Type III Sum of Squares	df	Mean Square	F	Sig.
Stair	Linear	16.882	1	16.882	71.498	.000
	Quadratic	.540	1	.540	3.207	.092
	Cubic	2.175E-5	1	2.175E-5	.000	.992
Error(Stair)	Linear	3.778	16	.236		
	Quadratic	2.696	16	.169		
	Cubic	3.587	16	.224		

Since there are only 4 steps, SPSS only tests to a cubic (3rd degree) fit. In this example a linear fit was best. Note, this statistic makes no sense if the DV is not ordered, such as time, age, or date.

Repeated-measures ANOVA: Plot of marginal means

SPSS can plot the group means. This plot shows the means for each step.



Looks like a linear increase in A1 power as people descend the stairs. Note, A1 is a <u>negative</u> power.

Repeated-measures ANOVA: *Post hoc* tests

- Since there is a significant F we can do *post hoc* testing. If not significant this step IS NOT DONE.
- We will use the **Sidak** *post hoc* test. **Bonferroni** is too conservative. Choose from the **Options…** menu, NOT the **Post Hoc…** menu!

The results now show that steps 1 and 2 are not significantly different for each other but are different from 3 and 4 and steps 3 and 4 are different from all the other steps. This is a better result than the factorial ANOVA.

Pairwise Comparisons

Measure:MEASURE_1

(I) Stair	(J) Stair	Mean Difference (I-J)	Std. Error	Sig.a	95% Confidence Interval for Differencea	
					Lower Bound	Upper Bound
1	2	.266	.197	.730	-.326	.858
	3	.714*	.144	.001	.282	1.145
	4	1.336*	.192	.000	.762	1.911
2	1	-.266	.197	.730	-.858	.326
	3	.447*	.132	.023	.051	.843
	4	1.070*	.121	.000	.708	1.432
3	1	-.714*	.144	.001	-1.145	-.282
	2	-.447*	.132	.023	-.843	-.051
	4	.623*	.140	.002	.204	1.042
4	1	-1.336*	.192	.000	-1.911	-.762
	2	-1.070*	.121	.000	-1.432	-.708
	3	-.623*	.140	.002	-1.042	-.204

Based on estimated marginal means

a. Adjustment for multiple comparisons: Sidak.

*. The mean difference is significant at the .05 level.

MANOVA: Example: Repeated-measures

The MANOVA or multivariate analysis of variance tests the hypothesis that one or more independent variables, or factors, have an effect on a set of two or more dependent variables.

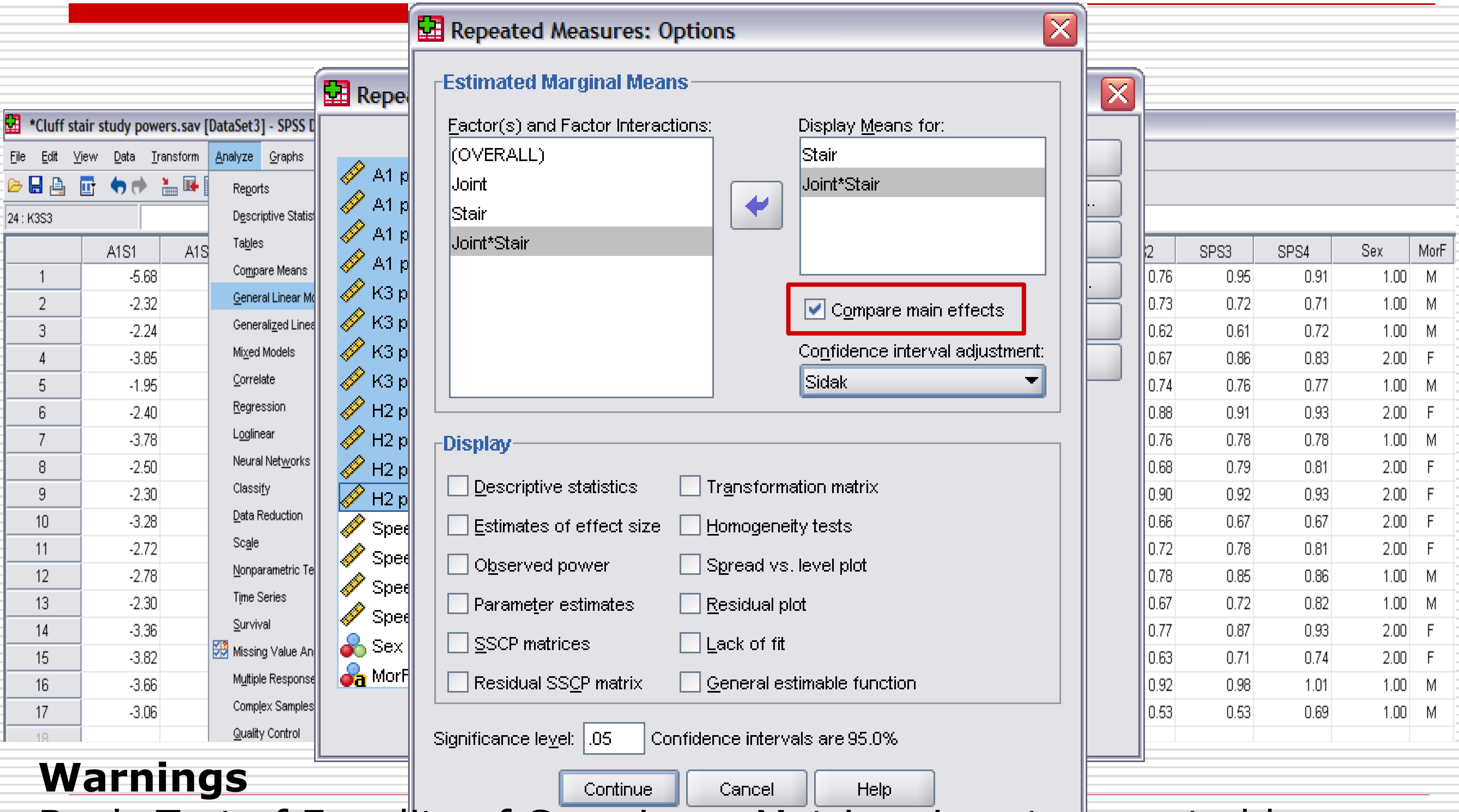
The present data must be organized as a repeated-measures ANOVA with three dependent variables, one for ankle powers, one for knee powers, and one for hip powers. The factor "stair" (step number) is also a repeated-measure since the same subjects descended each step.

This is now a **repeated-measures MANOVA**.

Assumptions for MANOVAs

1. The **two or more dependent variables** should be measured at the **interval** or **ratio level** (i.e., they are **continuous**). I.e., A1, K3, H2.
2. Your **independent variable** should consist of **two or more nominal or categorical**, **independent groups**.
3. You should have **independence of observations**, which means that there is no relationship between the observations in each group or between the groups themselves. For example, there must be different participants in each group with no participant being in more than one group. This is more of a study design issue than something you can test for, but it is an important assumption of the one-way MANOVA.
4. There should be **multivariate normality**. However, in practice, it is not uncommon to simply check that your **dependent variables** are **approximately normally distributed for each category of the independent variable**.
5. There needs to be **homogeneity of variances** (i.e., **equality of variances between the independent groups**).

Each column is a repeated measure. Each row is a subject.



Warnings

Box's Test of Equality of Covariance Matrices is not computed because there are fewer than two nonsingular cell covariance matrices.

MANOVA: Multivariate tests

Multivariate Tests[c]

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Joint	Pillai's Trace	.950	141.389[a]	2.000	15.000	.000	.950	282.778	1.000
	Wilks' Lambda	.050	141.389[a]	2.000	15.000	.000	.950	282.778	1.000
	Hotelling's Trace	18.852	141.389[a]	2.000	15.000	.000	.950	282.778	1.000
	Roy's Largest Root	18.852	141.389[a]	2.000	15.000	.000	.950	282.778	1.000
Stair	Pillai's Trace	.743	13.492[a]	3.000	14.000	.000	.743	40.476	.998
	Wilks' Lambda	.257	13.492[a]	3.000	14.000	.000	.743	40.476	.998
	Hotelling's Trace	2.891	13.492[a]	3.000	14.000	.000	.743	40.476	.998
	Roy's Largest Root	2.891	13.492[a]	3.000	14.000	.000	.743	40.476	.998
Joint * Stair	Pillai's Trace	.862	11.442[a]	6.000	11.000	.000	.862	68.654	.999
	Wilks' Lambda	.138	11.442[a]	6.000	11.000	.000	.862	68.654	.999
	Hotelling's Trace	6.241	11.442[a]	6.000	11.000	.000	.862	68.654	.999
	Roy's Largest Root	6.241	11.442[a]	6.000	11.000	.000	.862	68.654	.999

a. Exact statistic

b. Computed using alpha = .05

c. Design:
Within Subjects Design: Joint + Stair + Joint * Stair

The both main effects (Joint & Stair) and the interaction (Joint*Stair) are significant, with $p<0.0005$.

MANOVA: Test for sphericity

Mauchly's Test of Sphericity[b]

Measure:MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[a] Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Joint	.928	1.129	2	.569	.932	1.000	.500
Stair	.955	.670	5	.985	.970	1.000	.333
Joint * Stair	.036	45.916	20	.001	.435	.527	.167

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

b. Design:
Within Subjects Design: Joint + Stair + Joint * Stair

Main effects pass sphericity test but interaction does not.

MANOVA: Within-subjects tests

Tests of Within-Subjects Effects

Measure:MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Joint	Sphericity Assumed	502.882	2	251.441	140.983	.000	.898	281.967	1.000
	[illegible]	[illegible]	65	269.670	140.983	.000	.898	262.907	1.000
	[illegible]	[illegible]	00	251.441	140.983	.000	.898	281.967	1.000
	[illegible]	[illegible]	00	502.882	140.983	.000	.898	140.983	1.000
Erro	[illegible]	[illegible]	32	1.783					
	[illegible]	[illegible]	37	1.913					
	[illegible]	[illegible]	00	1.783					
	[illegible]	[illegible]	00	3.567					
Stair	Sphericity Assumed	7.090	3	2.363	14.239	.000	.471	42.716	1.000
	[illegible]	[illegible]	0	2.437	14.239	.000	.471	41.429	1.000
	[illegible]	[illegible]	0	2.363	14.239	.000	.471	42.716	1.000
	[illegible]	[illegible]	0	7.090	14.239	.002	.471	14.239	.943
Error	[illegible]	[illegible]	8	.166					
	[illegible]	[illegible]	4	.171					
	[illegible]	[illegible]	0	.166					
	Lower-bound	7.967	16.000	.498					
Joint * Stair	Sphericity Assumed	12.785	6	2.131	12.327	.000	.435	73.964	1.000
	[illegible]	[illegible]	0	4.899	12.327	.000	.435	32.173	.999
	[illegible]	[illegible]	4	4.041	12.327	.000	.435	39.005	1.000
	[illegible]	[illegible]	0	12.785	12.327	.003	.435	12.327	.909
Error	[illegible]	[illegible]	6	.173					
	[illegible]	[illegible]	8	.397					
	[illegible]	[illegible]	6	.328					
	[illegible]	[illegible]	0	1.037					

a. [illegible]

Main effects and interaction are all significant. Most are p<0.0005 (i.e., .000)

Partial Eta Squares indicate effect sizes. Most are relatively high.

The Observed Powers were all greater than 80% (.800) therefore sample size was adequate.

MANOVA:
Best fit tests

Tests of Within-Subjects Contrasts

Measure:MEASURE_1

Source	J...	Stair	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Joint	Linear		397.510	1	397.510	182.154	.000	.919	182.154	1.000
	Quadratic		105.372	1	105.372	76.098	.000	.826	76.098	1.000
Error(Joint)	Linear		34.916	16	2.182					
	Quadratic		22.155	16	1.385					
Stair		Linear	4.598	1	4.598	34.050	.000	.680	34.050	1.000
		Quadratic	2.113	1	2.113	10.404	.005	.394	10.404	.857
		Cubic	.379	1	.379	2.372	.143	.129	2.372	.305
Error(Stair)		Linear	2.161	16	.135					
		Quadratic	3.249	16	.203					
		Cubic	2.557	16	.160					
Joint * Stair	Linear	Linear	10.655	1	10.655	70.118	.000	.814	70.118	1.000
		Quadratic	.002	1	.002	.026	.875	.002	.026	.053
		Cubic	.102	1	.102	.659	.429	.040	.659	.119
	Quadratic	Linear	1.899	1	1.899	7.202	.016	.310	7.202	.712
		Quadratic	.033	1	.033	.160	.694	.010	.160	.066
		Cubic	.094	1	.094	.491	.494	.030	.491	.101
Error(Joint*Stair)	Linear	Linear	2.431	16	.152					
		Quadratic	1.096	16	.069					
		Cubic	2.476	16	.155					
	Quadratic	Linear	4.219	16	.264					
		Quadratic	3.313	16	.207					
		Cubic	3.059	16	.191					

a. Computed using alpha = .05

MANOVA: Example
Mixed with repeated-measures

- Add sex (MorF) as a between-subjects variable.

Tests of Between-Subjects Effects

Measure:MEASURE_1
Transformed Variable:Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
	1209.211	1	1209.211	410.902	.000	.965	410.902	1.000
MorF	.426	1	.426	.145	.709	.010	.145	.065
Error	44.142	15	2.943					

a. Computed using alpha = .05

- In this case there was no significant difference between sexes, p=0.709.

MANOVA:
Post hoc tests

-Step 1 differs from 4

-Step 2 differs from 3 and 4

-Step 3 differs from 2 and 4

-Step 4 differs from 1 and 2.

Pairwise Comparisons

Measure:MEASURE_1

(I) Stair	(J) Stair	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a] Lower Bound	Upper Bound
1	2	-.147	.084	.471	-.402	.108
	3	.104	.085	.801	-.151	.360
	4	.363*	.076	.001	.134	.593
2	1	.147	.084	.471	-.108	.402
	3	.252*	.081	.042	.007	.496
	4	.510*	.085	.000	.253	.768
3	1	-.104	.085	.801	-.360	.151
	2	-.252*	.081	.042	-.496	-.007
	4	.259	.089	.062	-.010	.527
4	1	-.363*	.076	.001	-.593	-.134
	2	-.510*	.085	.000	-.768	-.253
	3	-.259	.089	.062	-.527	.010

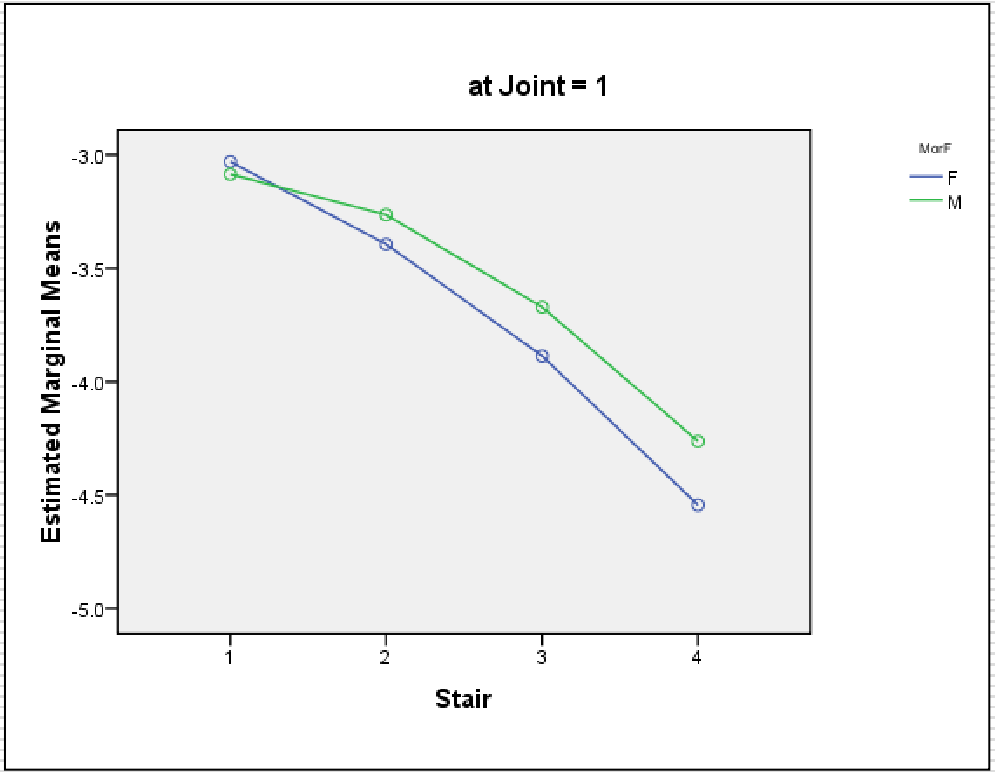
Based on estimated marginal means

a. Adjustment for multiple comparisons: Sidak.

*. The mean difference is significant at the .05 level.

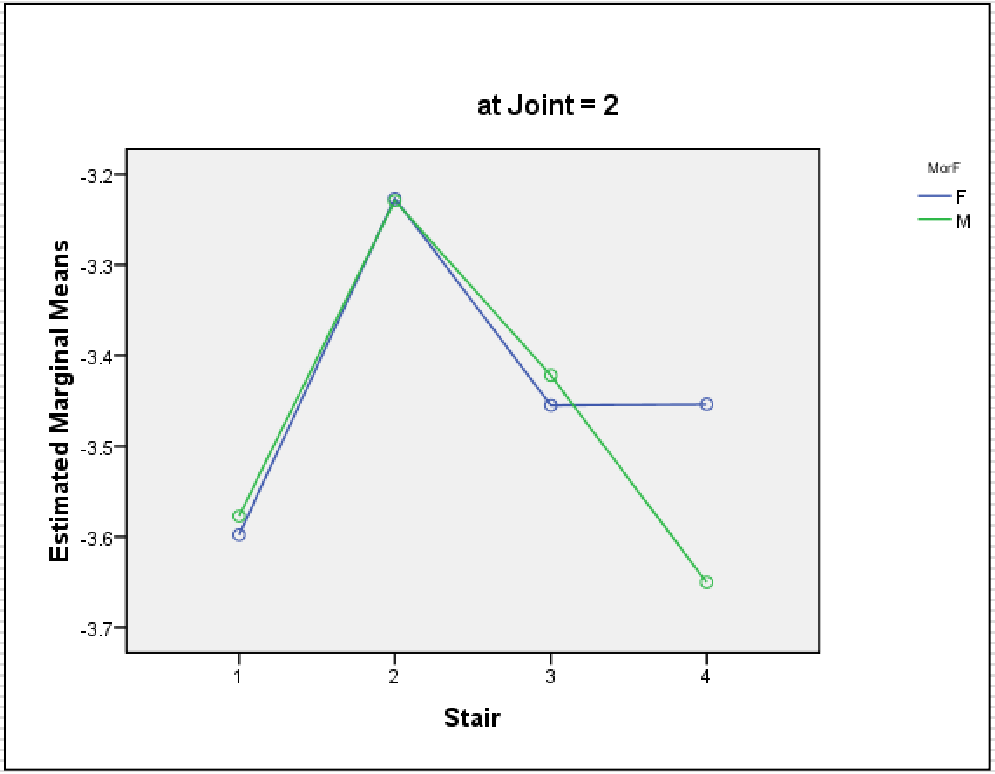
MANOVA: Plots: ankle plantar flexors

no sex difference, but powers do increase (negatively) during descent



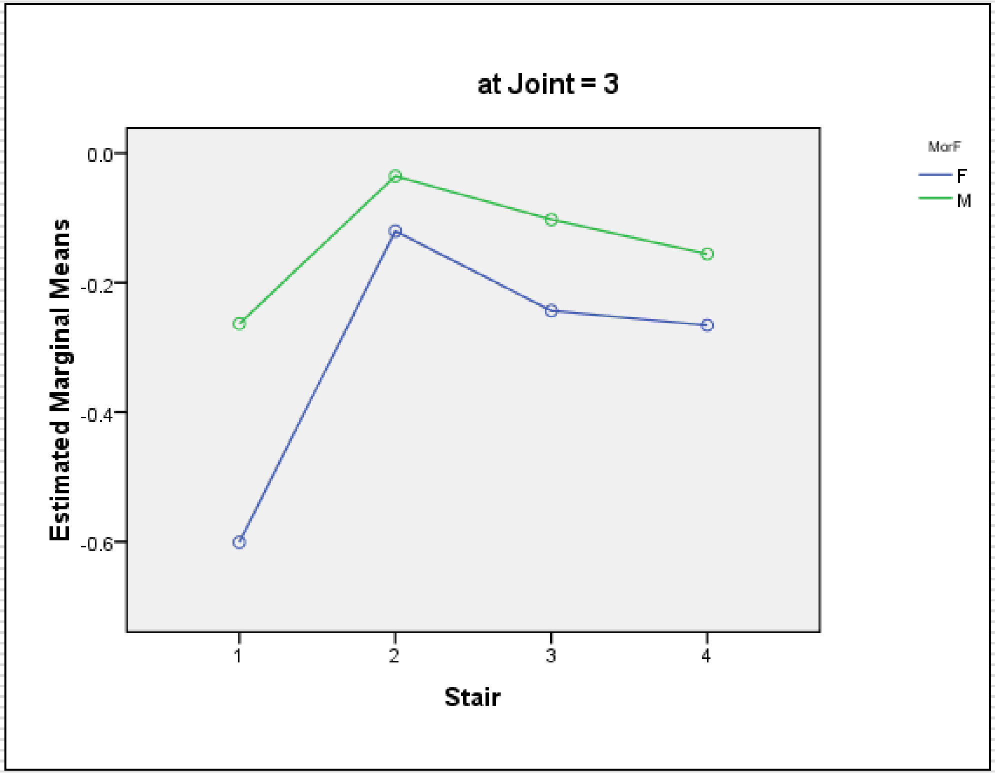
MANOVA: Plots: knee extensors

no sex difference, but no change linearly across steps, notice first step is quite high (negatively)



MANOVA: Plots: hip flexors

No sex difference, but powers were reduced except for first step, because of starting from rest (i.e., higher centre)



Why Should You Do a MANOVA?

You do a MANOVA instead of a series of one-at-a-time ANOVAs for two reasons:

- To reduce the experiment-wise level of Type I error. E.g., 8 *F* tests at $\alpha=0.05$ each means the experiment-wise probability of a Type I error (rejecting the null hypothesis when it is true) is 40%! The so-called overall test or omnibus test protects against this inflated error probability only when the null hypothesis is true. If a significant multivariate test occurs you may perform a series of ANOVAs on the individual variables. Bonferroni adjustments to the error rates provide additional "protection" from inflating Type I errors.

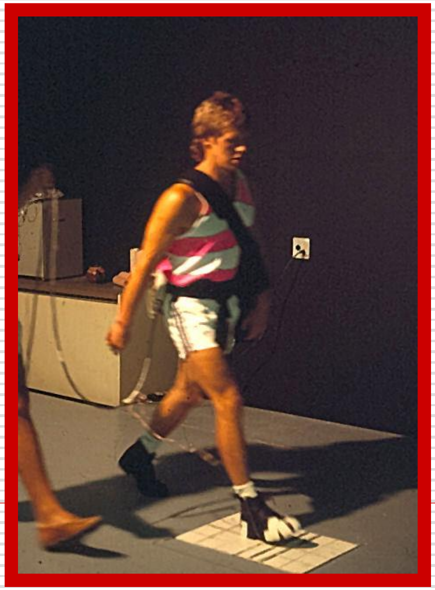
Why Should You Do a MANOVA?

Second reason:

- Individual ANOVAs may not produce a significant main effect on the DV, but in combination they might. This suggests that the variables are more meaningful taken together than considered separately since MANOVA takes into account the intercorrelations among the DVs.

Lab

- Analyze the data from the mailbag study.
  - File: Mailbag study with repeated-measures.sav
- Males and females walked without a mailbag, with a side-satchel mailbag, a side-satchel with a trap-strap, a Bigg's mailbag, and a front pack mailbag.
- Integrated EMGs from both left and right erector spinae were collected.


Lab questions

- What are the independent and dependent variables?
- What is the name of this design?
- Was there a difference between the sexes?
- Were there significant differences between the mail bags?
- Which condition(s) (other than no bag) had the least EMG activity?

Steps for Running a MANOVA

1. Choose **Multivariate…** or **Repeated measures…** from the **General Linear Model** item under **Analysis**
2. Move the dependent variables to the **Dependent Variables** box (MANOVA) or create repeated-measures variables
3. Move the grouping variable(s) (independent nominal variables) into the **Fixed Factors(s)** box
4. Click on the **Options…** button
5. Check options such as **Observed power**, **Estimates of effect size**, **Homogeneity tests** as needed
6. Move the grouping variable(s) (Fixed factors) to the **Display Means for** box
7. Check the **Compare main effects** box if you want to do *post hoc* testing

Steps for Running a MANOVA

8. Select the **Sidak** item from the drop down menu
9. Press **Continue** button
10. Press the **Model...** button
11. Usually select the **Full factorial** option
12. Press **Continue** button
13. Press the **Plots...** button and then add whatever plots are of interest. Usually put the independent variable of the horizontal axis.
14. Press **Continue** button
15. Press the **OK** button to start the analysis

Interpreting the Output

1. First check that data pass **Box's Test of Equality of Covariance Matrices**. Sig. value must be greater than 0.001.
2. From the **Multivariate Tests** results check the fixed factor for significance. Ignore the Intercept row.
3. Usually look at the row labelled **Wilks' Lambda**. Pillai's Trace is the most powerful procedure. Roy's Largest Root should be ignored if the other three rows are not significant. If Pillai's Trace and Hotelling Trace values are close to each other the effect of the variable is weak.
4. Check the results of **Levene's Test of Equality of Error Variances**.
5. Next, look at the univariate ANOVA results in the table labelled **Test of Between-Subjects Effects**.
6. If you are doing a two-way or higher ANOVA start with the interactions. If significant these are the most important results.
7. Next, look at the main effects of the individual dependent variables.
8. Finally, if any main effect is significant use ***post hoc* testing** to determine where the difference(s) lie.