

FES758b / S&DS 363 / S&DS 563
Multivariate Statistics
Homework #5 - MANOVA / GLM
Due : Monday, 4/8/19 on CANVAS (by midnight)

HOMEWORK WORKED OUT FOR SAMPLE DATASET
The example below is JUST FOR YOUR PRACTICE.
NOTHING TO TURN IN HERE!

DANIELA.csv contain data collected by Daniela Cusack from three plantations. Each plantation was divided into areas with homogenous overstory tree species of six types. In class we looked at factors which predicted the number of individual saplings in each of three height classes. The saplings were also classified in terms of the dispersal mechanism associated with that species. The three dispersal mechanisms (the response variables) were birds, mammals, or other (includes wind, water, bats, gravity).

.....

SAS Results

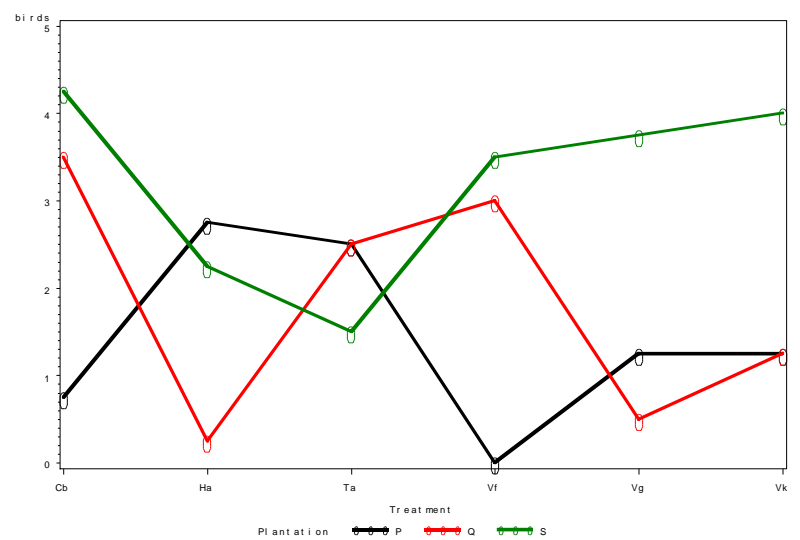
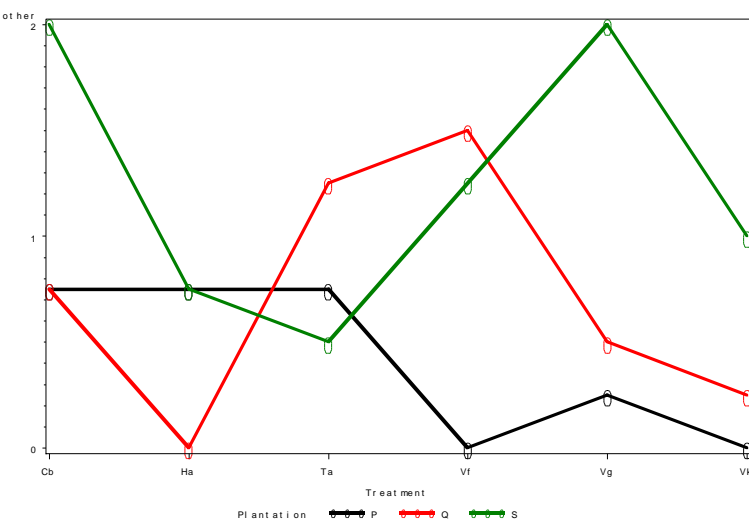
1. Look at interaction plots between plantation and overstory species for each of the dispersal mechanisms. Discuss what you see.

Here is SAS code for this :

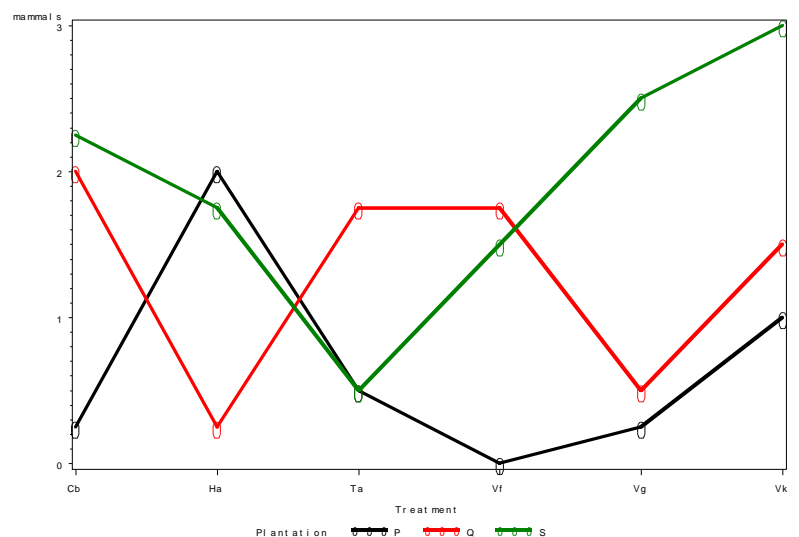
```
*MUST SORT DATA TO GET INTERACTION PLOTS!!!;
PROC SORT DATA=IN.overstory; BY treatment plantation;
PROC MEANS DATA=IN.overstory;
  VAR mammals birds other;
  BY treatment plantation;
  OUTPUT OUT=overMEAN;
RUN;

DATA OVERMEAN; SET OVERMEAN; IF _STAT_='MEAN';
RUN;
SYMBOL1 VALUE=0 C=BLACK H=2 W=5 I=JOIN;
SYMBOL2 VALUE=0 C=RED H=2 W=5 I=JOIN;
SYMBOL3 VALUE=0 C=GREEN H=2 W=5 I=JOIN;
PROC GPLOT DATA=overmean;
  PLOT mammals*treatment=plantation;
  PLOT birds*treatment=plantation;
  PLOT other*treatment=plantation;
RUN;
```

Here are results :



These plots suggest that there may be an interaction between plantation and overstory species (i.e. treatment). Also suggests that there may not be much of a treatment effect (i.e. overstory type doesn't affect dispersal rate). Also seems like overall that the S plantation has generally higher rates than the other plantations.



- Run MANOVA for these two categorical factors. Discuss your results, both univariate and multivariate.

Here is Code and results for the GLM model (i.e. two-way MANOVA) :

```
proc glm data=in.overstory;
class treatment plantation;
model mammals birds other=plantation treatment plantation*treatment /
solution;
manova h=plantation treatment plantation*treatment;
run;
```

Results follow (lots of output).

Univariate Results : For mammals, there are differences between plantations, there are no overall observed differences due to overstory treatment effect; however, there is evidence on an interaction effect.

The individual coefficients suggest that for overstory species, the TA species if different from VK (and perhaps other) species – might want to test this as an indicator variable.

Similar results are observed for birds. No TA species effect thing observed for ‘other’ dispersal mechanisms.

Multivariate Results : Overall, there are differences between Plantations (all multivariate statistics are significant). For overstory species, only Roy’s Greatest Root is significant, which suggest that there is a single direction in multivariate space that shows differences between overstory treatment groups (since Roy’s Greatest Root only tests the first eigenvalue which is associated with the direction of maximum discrimination’). Most of the multivariate tests suggest there is an interaction effect between Plantation and overstory Treatment.

Dependent Variable: mammals mammals

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	17	55.6250000	3.2720588	2.79	0.0021
Error	54	63.2500000	1.1712963		
Corrected Total	71	118.8750000			

R-Square	Coeff Var	Root MSE	mammals Mean
0.467928	83.78821	1.082264	1.291667

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Plantation	2	18.75000000	9.37500000	8.00	0.0009
Treatment	5	6.79166667	1.35833333	1.16	0.3410
Treatment*Plantation	10	30.08333333	3.00833333	2.57	0.0128

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Plantation	2	18.75000000	9.37500000	8.00	0.0009
Treatment	5	6.79166667	1.35833333	1.16	0.3410
Treatment*Plantation	10	30.08333333	3.00833333	2.57	0.0128

Parameter		Estimate	Standard Error	t Value	Pr > t
Intercept		3.000000000 B	0.54113221	5.54	<.0001
Plantation	P	-2.000000000 B	0.76527652	-2.61	0.0116
Plantation	Q	-1.500000000 B	0.76527652	-1.96	0.0552
Plantation	S	0.000000000 B	.	.	.
Treatment	Cb	-0.750000000 B	0.76527652	-0.98	0.3314
Treatment	Ha	-1.250000000 B	0.76527652	-1.63	0.1082
Treatment	Ta	-2.500000000 B	0.76527652	-3.27	0.0019
Treatment	Vf	-1.500000000 B	0.76527652	-1.96	0.0552

Treatment	Vg	-0.500000000	B	0.76527652	-0.65	0.5163
Treatment	Vk	0.000000000	B	.	.	.
Treatment*Plantation	Cb P	0.000000000	B	1.08226443	0.00	1.0000
Treatment*Plantation	Cb Q	1.250000000	B	1.08226443	1.15	0.2532
Treatment*Plantation	Cb S	0.000000000	B	.	.	.
Treatment*Plantation	Ha P	2.250000000	B	1.08226443	2.08	0.0424
Treatment*Plantation	Ha Q	0.000000000	B	1.08226443	0.00	1.0000
Treatment*Plantation	Ha S	0.000000000	B	.	.	.
Treatment*Plantation	Ta P	2.000000000	B	1.08226443	1.85	0.0701

Dependent Variable: mammals mammals

Parameter	Estimate	Standard Error	t Value	Pr > t
Treatment*Plantation Ta Q	2.750000000 B	1.08226443	2.54	0.0140
Treatment*Plantation Ta S	0.000000000 B	.	.	.
Treatment*Plantation Vf P	0.500000000 B	1.08226443	0.46	0.6459
Treatment*Plantation Vf Q	1.750000000 B	1.08226443	1.62	0.1117
Treatment*Plantation Vf S	0.000000000 B	.	.	.
Treatment*Plantation Vg P	-0.250000000 B	1.08226443	-0.23	0.8182
Treatment*Plantation Vg Q	-0.500000000 B	1.08226443	-0.46	0.6459
Treatment*Plantation Vg S	0.000000000 B	.	.	.
Treatment*Plantation Vk P	0.000000000 B	.	.	.
Treatment*Plantation Vk Q	0.000000000 B	.	.	.
Treatment*Plantation Vk S	0.000000000 B	.	.	.

NOTE: The X'X matrix has been found to be singular, and a generalized inverse was used to solve the normal equations. Terms whose estimates are followed by the letter 'B' are not uniquely estimable.

Dependent Variable: birds birds

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	17	124.5694444	7.3276144	3.07	0.0009
Error	54	128.7500000	2.3842593		
Corrected Total	71	253.3194444			

R-Square	Coeff Var	Root MSE	birds Mean
0.491748	71.72615	1.544105	2.152778

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Plantation	2	42.19444444	21.09722222	8.85	0.0005
Treatment	5	8.73611111	1.74722222	0.73	0.6020
Treatment*Plantation	10	73.63888889	7.36388889	3.09	0.0036

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Plantation	2	42.19444444	21.09722222	8.85	0.0005

Treatment	5	8.73611111	1.74722222	0.73	0.6020
Treatment*Plantation	10	73.63888889	7.36388889	3.09	0.0036

Parameter		Estimate	Standard Error	t Value	Pr > t
Intercept		4.000000000 B	0.77205234	5.18	<.0001
Plantation	P	-2.750000000 B	1.09184689	-2.52	0.0148
Plantation	Q	-2.750000000 B	1.09184689	-2.52	0.0148
Plantation	S	0.000000000 B	.	.	.
Treatment	Cb	0.250000000 B	1.09184689	0.23	0.8198
Treatment	Ha	-1.750000000 B	1.09184689	-1.60	0.1148
Treatment	Ta	-2.500000000 B	1.09184689	-2.29	0.0260
Treatment	Vf	-0.500000000 B	1.09184689	-0.46	0.6488
Treatment	Vg	-0.250000000 B	1.09184689	-0.23	0.8198
Treatment	Vk	0.000000000 B	.	.	.
Treatment*Plantation	Cb P	-0.750000000 B	1.54410468	-0.49	0.6291
Treatment*Plantation	Cb Q	2.000000000 B	1.54410468	1.30	0.2007
Treatment*Plantation	Cb S	0.000000000 B	.	.	.
Treatment*Plantation	Ha P	3.250000000 B	1.54410468	2.10	0.0400
Treatment*Plantation	Ha Q	0.750000000 B	1.54410468	0.49	0.6291
Treatment*Plantation	Ha S	0.000000000 B	.	.	.
Treatment*Plantation	Ta P	3.750000000 B	1.54410468	2.43	0.0185

Dependent Variable: birds birds

Parameter		Estimate	Standard Error	t Value	Pr > t
Treatment*Plantation	Ta Q	3.750000000 B	1.54410468	2.43	0.0185
Treatment*Plantation	Ta S	0.000000000 B	.	.	.
Treatment*Plantation	Vf P	-0.750000000 B	1.54410468	-0.49	0.6291
Treatment*Plantation	Vf Q	2.250000000 B	1.54410468	1.46	0.1509
Treatment*Plantation	Vf S	0.000000000 B	.	.	.
Treatment*Plantation	Vg P	0.250000000 B	1.54410468	0.16	0.8720
Treatment*Plantation	Vg Q	-0.500000000 B	1.54410468	-0.32	0.7473
Treatment*Plantation	Vg S	0.000000000 B	.	.	.
Treatment*Plantation	Vk P	0.000000000 B	.	.	.
Treatment*Plantation	Vk Q	0.000000000 B	.	.	.
Treatment*Plantation	Vk S	0.000000000 B	.	.	.

NOTE: The X'X matrix has been found to be singular, and a generalized inverse was used to solve the normal equations. Terms whose estimates are followed by the letter 'B' are not uniquely estimable.

Dependent Variable: other

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	17	26.12500000	1.53676471	2.46	0.0063
Error	54	33.75000000	0.62500000		
Corrected Total	71	59.87500000			

R-Square	Coeff Var	Root MSE	other Mean
0.436326	99.86140	0.790569	0.791667

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Plantation	2	8.58333333	4.29166667	6.87	0.0022
Treatment	5	4.79166667	0.95833333	1.53	0.1950
Treatment*Plantation	10	12.75000000	1.27500000	2.04	0.0466

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Plantation	2	8.58333333	4.29166667	6.87	0.0022
Treatment	5	4.79166667	0.95833333	1.53	0.1950
Treatment*Plantation	10	12.75000000	1.27500000	2.04	0.0466

Parameter		Estimate	Standard Error	t Value	Pr > t
Intercept		1.000000000 B	0.39528471	2.53	0.0144
Plantation	P	-1.000000000 B	0.55901699	-1.79	0.0792
Plantation	Q	-0.750000000 B	0.55901699	-1.34	0.1853
Plantation	S	0.000000000 B	.	.	.
Treatment	Cb	1.000000000 B	0.55901699	1.79	0.0792
Treatment	Ha	-0.250000000 B	0.55901699	-0.45	0.6565
Treatment	Ta	-0.500000000 B	0.55901699	-0.89	0.3751
Treatment	Vf	0.250000000 B	0.55901699	0.45	0.6565
Treatment	Vg	1.000000000 B	0.55901699	1.79	0.0792
Treatment	Vk	0.000000000 B	.	.	.
Treatment*Plantation	Cb P	-0.250000000 B	0.79056942	-0.32	0.7530
Treatment*Plantation	Cb Q	-0.500000000 B	0.79056942	-0.63	0.5298
Treatment*Plantation	Cb S	0.000000000 B	.	.	.
Treatment*Plantation	Ha P	1.000000000 B	0.79056942	1.26	0.2113
Treatment*Plantation	Ha Q	0.000000000 B	0.79056942	0.00	1.0000
Treatment*Plantation	Ha S	0.000000000 B	.	.	.
Treatment*Plantation	Ta P	1.250000000 B	0.79056942	1.58	0.1197

Dependent Variable: other

Parameter		Estimate	Standard Error	t Value	Pr > t
Treatment*Plantation	Ta Q	1.500000000 B	0.79056942	1.90	0.0631
Treatment*Plantation	Ta S	0.000000000 B	.	.	.
Treatment*Plantation	Vf P	-0.250000000 B	0.79056942	-0.32	0.7530
Treatment*Plantation	Vf Q	1.000000000 B	0.79056942	1.26	0.2113

Treatment*Plantation Vf S	0.000000000 B	.	.	.
Treatment*Plantation Vg P	-0.750000000 B	0.79056942	-0.95	0.3470
Treatment*Plantation Vg Q	-0.750000000 B	0.79056942	-0.95	0.3470
Treatment*Plantation Vg S	0.000000000 B	.	.	.
Treatment*Plantation Vk P	0.000000000 B	.	.	.
Treatment*Plantation Vk Q	0.000000000 B	.	.	.
Treatment*Plantation Vk S	0.000000000 B	.	.	.

NOTE: The X'X matrix has been found to be singular, and a generalized inverse was used to solve the normal equations. Terms whose estimates are followed by the letter 'B' are not uniquely estimable.

The GLM Procedure
Multivariate Analysis of Variance

Characteristic Roots and Vectors of: E Inverse * H, where
H = Type III SSCP Matrix for Plantation
E = Error SSCP Matrix

Characteristic Root	Percent	Characteristic Vector mammals	V'EV=1 birds	other
0.65908195	93.98	0.02545218	0.05607267	0.11657032
0.04220450	6.02	-0.16462472	0.10087871	0.01551757
0.00000000	0.00	-0.03795726	-0.03382330	0.12965599

MANOVA Test Criteria and F Approximations for the Hypothesis of No Overall Plantation Effect
H = Type III SSCP Matrix for Plantation
E = Error SSCP Matrix

S=2 M=0 N=25

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.57833466	5.46	6	104	<.0001
Pillai's Trace	0.43775243	4.95	6	106	0.0002
Hotelling-Lawley Trace	0.70128645	6.02	6	67.584	<.0001
Roy's Greatest Root	0.65908195	11.64	3	53	<.0001

NOTE: F Statistic for Roy's Greatest Root is an upper bound.
NOTE: F Statistic for Wilks' Lambda is exact.

Characteristic Roots and Vectors of: E Inverse * H, where
H = Type III SSCP Matrix for Treatment
E = Error SSCP Matrix

Characteristic Root	Percent	Characteristic Vector mammals	V'EV=1 birds	other
0.37140765	80.86	-0.12906445	0.08720092	0.12616486
0.07431415	16.18	0.09056686	0.02757895	0.04952936
0.01360850	2.96	-0.06580182	0.07810298	-0.11076723

MANOVA Test Criteria and F Approximations for the Hypothesis of No Overall Treatment Effect
H = Type III SSCP Matrix for Treatment
E = Error SSCP Matrix

	S=3	M=0.5	N=25			
Statistic	Value	F Value	Num DF	Den DF	Pr > F	
Wilks' Lambda	0.66962536	1.50	15	143.95	0.1122	
Pillai's Trace	0.35342158	1.44	15	162	0.1335	
Hotelling-Lawley Trace	0.45933030	1.56	15	93.113	0.0991	
Roy's Greatest Root	0.37140765	4.01	5	54	0.0036	

NOTE: F Statistic for Roy's Greatest Root is an upper bound.

Characteristic Roots and Vectors of: E Inverse * H, where
H = Type III SSCP Matrix for Treatment*Plantation
E = Error SSCP Matrix

Characteristic Root	Percent	Characteristic Vector mammals	V'EV=1 birds	other
0.96430157	83.05	0.02260670	0.06053308	0.11038347
0.12388040	10.67	-0.14514093	0.10390370	-0.03122217
0.07292461	6.28	-0.08725202	0.00210760	0.13221488

MANOVA Test Criteria and F Approximations for the Hypothesis
of No Overall Treatment*Plantation Effect
H = Type III SSCP Matrix for Treatment*Plantation
E = Error SSCP Matrix

	S=3	M=3	N=25			
Statistic	Value	F Value	Num DF	Den DF	Pr > F	
Wilks' Lambda	0.42218474	1.75	30	153.31	0.0158	
Pillai's Trace	0.66910687	1.55	30	162	0.0449	
Hotelling-Lawley Trace	1.16110658	1.97	30	111.02	0.0059	
Roy's Greatest Root	0.96430157	5.21	10	54	<.0001	

NOTE: F Statistic for Roy's Greatest Root is an upper bound.

3. Perform (multivariate) contrasts to compare levels of a particular factor or combinations of factors. Discuss your results.

*To do this, we have to analyze this as a **ONE WAY MANOVA** with a combination treatment effect. The code below creates a new variable called **trtcombine** which just combines plantation code and treatment code.*

There are LOTS of possibilities here. I contrasted Plantations P and Q, and P and S. I contrasted overstory species Vf vs. Vg. Finally, I looked at the interaction between Vf and Vg between plots P and Q.

```
data in.overstory; set in.overstory;
trtcombine=trim(trim(plantation) || trim(treatment));
run;
```


To run the actual model with contrasts use

```
proc sort data=in.overstory; by trtcombine;
proc glm data=in.overstory;
class trtcombine;
model mammals birds other=trtcombine;
contrast 'P vs Q' trtcombine 1 1 1 1 1 1 -1 -1 -1 -1 -1 -1 0 0 0 0 0 0;
contrast 'P vs S' trtcombine 1 1 1 1 1 1 0 0 0 0 0 0 -1 -1 -1 -1 -1 -1;
contrast 'Vf vs Vg' trtcombine 0 0 0 1 0 -1 0 0 0 0 1 0 -1 0 0 0 0 1 0 -1 0;
contrast 'Vf vs Vg PQ interaction' trtcombine 0 0 0 1 0 -1 0 0 0 0 -1 0 0 0 0 -1 0 1 0 0
0 0 0 0 0 0;
manova h=trtcombine;
run;
```

Lots of output, I leave interpretation to you.

4. Daniela also measured the amount of light and the density of forest litter on each plot. Fit a model that includes these covariates as predictors of the number of saplings associated with each dispersal mechanism.

Use PROC PLOT to make plots to check for linearity. For Daniela's data, we add the possible covariate effects of light and litter. The SAS code is

```
PROC GLM DATA=IN.OVERSTORY;
CLASS TREATMENT PLANTATION;
MODEL MAMMALS BIRDS OTHER=TREATMENT PLANTATION TREATMENT*PLANTATION
LIGHT LITTER / SOLUTION;
MANOVA H=TREATMENT PLANTATION TREATMENT*PLANTATION LIGHT LITTER;
RUN;
```

The output shows that neither appears to have a significant univariate or multivariate effect on dispersion mechanism counts.

MANOVA Test Criteria and Exact F Statistics for the Hypothesis of No Overall Light Effect
H = Type III SSCP Matrix for Light
E = Error SSCP Matrix

	S=1	M=0.5	N=22			
Statistic	Value	F Value	Num DF	Den DF	Pr > F	
Wilks' Lambda	0.91994901	1.33	3	46	0.2748	
Pillai's Trace	0.08005099	1.33	3	46	0.2748	
Hotelling-Lawley Trace	0.08701677	1.33	3	46	0.2748	
Roy's Greatest Root	0.08701677	1.33	3	46	0.2748	

MANOVA Test Criteria and Exact F Statistics for the Hypothesis of No Overall Litter Effect
H = Type III SSCP Matrix for Litter
E = Error SSCP Matrix
S=1 M=0.5 N=22

Statistic	Value	F Value	Num DF	Den DF	Pr > F
-----------	-------	---------	--------	--------	--------

Wilks' Lambda	0.99465782	0.08	3	46	0.9693
Pillai's Trace	0.00534218	0.08	3	46	0.9693
Hotelling-Lawley Trace	0.00537088	0.08	3	46	0.9693
Roy's Greatest Root	0.00537088	0.08	3	46	0.9693

5. Check model assumptions by making a chi-square quantile plot of the residuals. Modify your model as appropriate based on your findings.

I've already discussed how to make chi-square quantile plots. You just need the residuals. The SAS code simply adds an OUTPUT option to PROC GLM

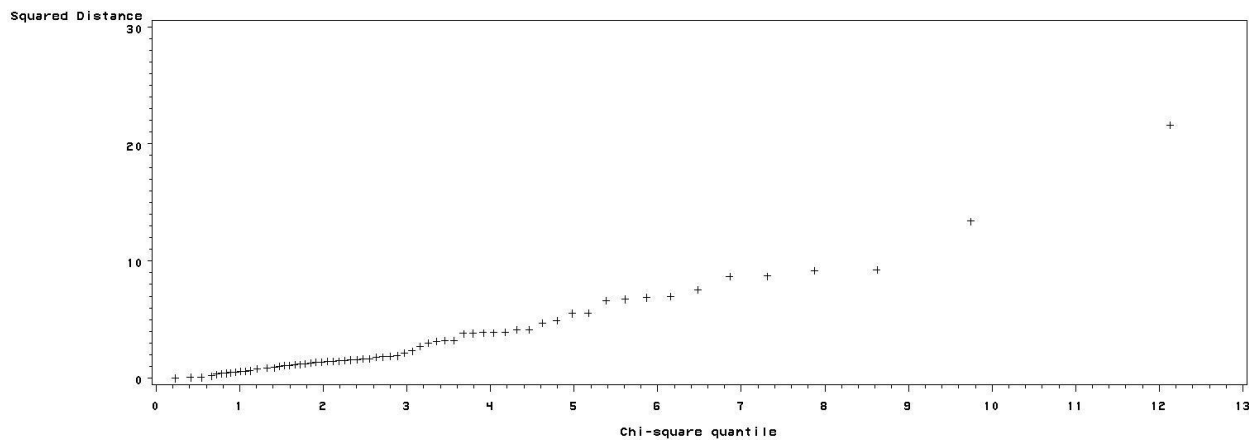
```
PROC GLM DATA=IN.OVERSTORY ;
CLASS TREATMENT PLANTATION;
MODEL MAMMALS BIRDS OTHER=TREATMENT PLANTATION TREATMENT*PLANTATION;
MANOVA H=TREATMENT PLANTATION TREATMENT*PLANTATION;
OUTPUT OUT=OUTSTAT RESIDUAL=RESIDUALA RESIDUALB RESIDUALC;
RUN;

%INCLUDE "C:\Documents and Settings\jon\My Documents\Classes\Multivariate
Articles Programs Resources\Software Programs\SAS programs\MULTNORM.SAS";

%MULTNORM(VAR= RESIDUALA RESIDUALB RESIDUALC, DATA=OUTSTAT)
```

The resulting plot looks good – no evidence of serious departure from multivariate normality.

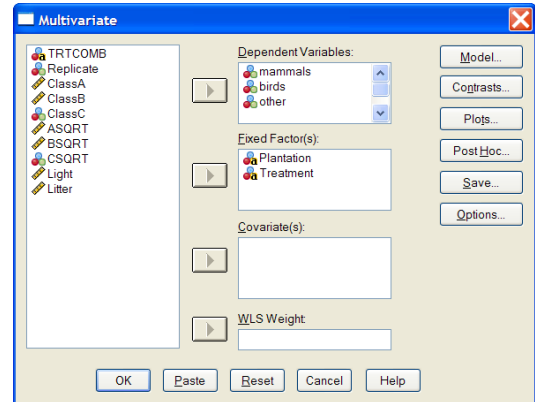
MULTNORM macro: Chi-square Q-Q plot



SPSS Results

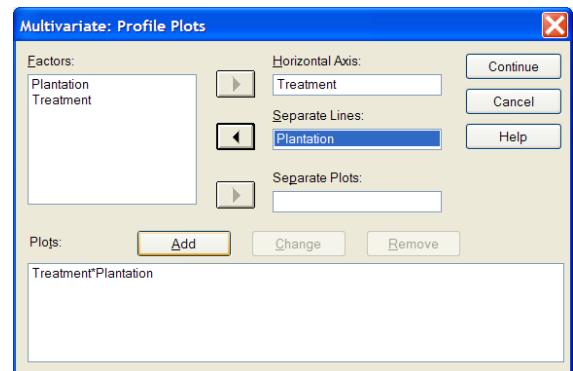
1. Look at interaction plots between plantation and overstory species for each of the dispersal mechanisms. Discuss what you see.

To make interaction plots in SPSS (called **Profile Plots**), use Analyze → Generalized Linear Model → Multivariate. This will give you the plots **and** the model needed for question two. Indicate that mammals, birds and other the Dependent Variables and that Plantation and Treatment are 'Fixed Factors'.



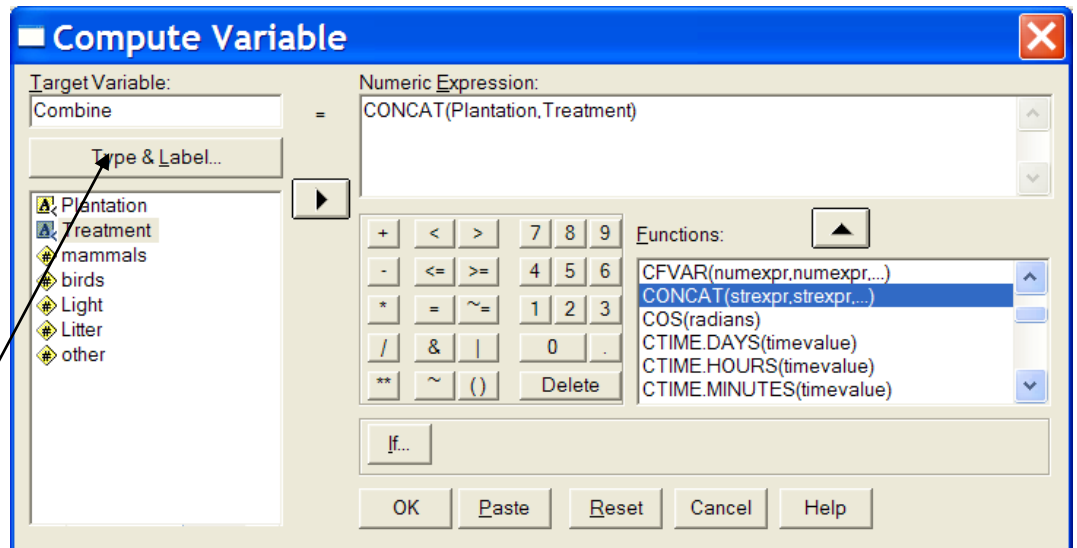
Click on Plots, indicate that you'd like Treatment on the Horizontal Axis and Plantation as Separate Lines, then click **ADD**.

See SAS section for interpretation of results and similar plots.



2. Run MANOVA for these two categorical factors. Discuss your results, both univariate and multivariate.
3. Perform (multivariate) contrasts to compare levels of a particular factor or combinations of factors. Discuss your results.

To make new combination treatment variable, can do manually in EXCEL, or in SPSS use TRANSFORM → COMPUTE and use the Concatenate function : Make sure you click here and indicate that the data type is STRING



To run the contrasts, click on the CONTRASTS button in SPSS when running Analyze → GLM.

However, this only gives UNIVARIATE contrasts. No multivariate contrasts available that I know of – sorry!

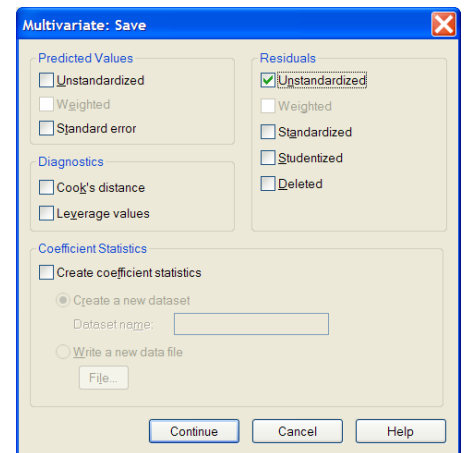
4. Daniela also measured the amount of light and the density of forest litter on each plot. Fit a model that includes these covariates as predictors of the number of saplings associated with each dispersal mechanism.

In SPSS, just enter LIGHT and LITTER in the COVARIATES box in Analyze → General Linear Models. Results are discussed in SAS section.

5. Check model assumptions by making a chi-square quantile plot of the residuals. Modify your model as appropriate based on your findings.

I've already discussed how to make chi-square quantile plots in SPSS (see notes at the very end of Principle Components Analysis).

In SPSS, when using Analyze → General Linear Models → Multivariate, click on SAVE and then choose UNSTANDARDIZED RESIDUALS



R Results (code only – for results see SAS section)

1. Look at interaction plots between plantation and overstory species for each of the dispersal mechanisms. Discuss what you see.

Here is the R code : See SAS section above for results and interpretation.

```
#get the data
daniela=read.csv("http://www.reuningscherer.net/stat660/data/Daniela.csv", header=T)

#make an interaction plots

#this statement makes 4 plots per page
par(mfrow=c(2,2))

#this makes the plots
interaction.plot(daniela$Treatment,daniela$Plantation,daniela$mammals,
  lwd=3,col=c("red","blue","black"),xlab="Species",main="Interaction
Plot for Mammals")
interaction.plot(daniela$Treatment,daniela$Plantation,daniela$birds,
  lwd=3,col=c("red","blue","black"),xlab="Species",main="Interaction
Plot for Birds")
interaction.plot(daniela$Treatment,daniela$Plantation,daniela$other,
  lwd=3,col=c("red","blue","black"),xlab="Species",main="Interaction
Plot for Other")
```

2. Run MANOVA for these two categorical factors. Discuss your results, both univariate and multivariate.

```
#fit linear model
mod1=manova(as.matrix(daniela[,8:10])~daniela$Treatment +
  daniela$Plantation +daniela$Plantation*daniela$Treatment)

#get univariate results
summary.aov(mod1)

#get multivariate results
summary.manova(mod1)
summary.manova(mod1,test="Wilks")
```

3. Perform (multivariate) contrasts to compare levels of a particular factor or combinations of factors. Discuss your results.

I'll simply say here that contrasts in R are difficult and at present not pleasant. You can see the comments I've made in the notes about contrasts – but for the moment, even I'm giving up in frustration. SAS is much better for this.

Consider it major extra credit if you get this to work in R.

See website link in the notes for some sense of how to fit multivariate contrasts.

4. Daniela also measured the amount of light and the density of forest litter on each plot. Fit a model that includes these covariates as predictors of the number of saplings associated with each dispersal mechanism.

```
#remove missing values
daniela2=daniela[complete.cases(daniela[,1:15]),1:15]

#fit linear model
mod1=lm(as.matrix(daniela2[,8:10])~daniela2$Treatment*daniela2$Plantation + daniela2$Light + daniela2$Litter)

#get univariate results
summary.lm(mod1)
```

NOTE : I haven't been able to get R to give the correct Multivariate Output Yet – I'll let you know if I figure this out!

5. Check model assumptions by making a chi-square quantile plot of the residuals. Modify your model as appropriate based on your findings.

I've already discussed how to make chi-square quantile plots in R using the code online in CSQplot.R. You just need the residuals which are contained in mod1\$residuals.

HOMEWORK ASSIGNMENT

PLEASE turn in the following answers for YOUR DATASET!
If MANOVA/GLM is not appropriate for your data, use one the
loaner dataset described at the end of the assignment.

List your Name(s – if a group) and a short reminder
description of your dataset.

You will need to modify the specifics of the assignment below somewhat to be appropriate for your data (i.e. perhaps you'll do one-way MANOVA, etc). Use your judgment or check with TA's/JDRS. Also, not all of the software programs discussed will support all the questions below. Modify as appropriate.

This homework assumes that you have

- *Two or more continuous response variables (this is a REQUIREMENT)*
 - *Two (or more) categorical predictors. One only is ok.*
 - *One or more additional continuous predictors (if none, that's ok)*
1. For two categorical factors, make interaction plots for each of your response variables. Discuss what you see. If you only have one categorical factor, then provide some plot/discussion of which means are different for which response variables (i.e. boxplots)
 2. Run Two-Way MANOVA for these two categorical factors. Discuss your results, both univariate and multivariate. If only one categorical predictor, do one-way MANOVA.
 3. Perform (multivariate) contrasts to compare levels of a particular factor or combinations of factors. Discuss your results. If you're using SPSS/R, you won't be able to do the multivariate contrasts.
 4. If applicable, add a continuous variable to your model and fit as a multiple-response generalized linear model. Before you fit the model, make some plots to see if there are linear relationships between your covariates and your responses.
 5. Check model assumptions by making a chi-square quantile plot of the residuals. Modify your model as appropriate based on your findings.

BONUS – use some sort of multivariate correction procedure to modify your p-values as appropriate.

CODING HM – As an alternative to this assignment, research/code/figure out how to do multivariate contrasts in R. Write a nice function, turn this in! If you do this, no need to actually do MANOVA, etc.

LOANER DATASET

(if MANOVA is not appropriate for your data)

OHIOCRIMEHM.sas7bcat and OHIOCRIMEHM.xls both contain information on attitudes toward crime and criminals obtained in 1995 in a central Ohio survey. For this analysis, the following variables are included :

DEPENDENT VARIABLES (measured on a 1-7 scale, 1=strongly disagree, 7=strongly agree)

- V10. During the summer, the state government should help provide jobs for inner-city youths from poor families.
- V12. Develop recreation programs, like midnight basketball, so that youths will have something to do instead of wandering the streets at night.
- V16. Instead of arresting people who use drugs, get them into treatment programs that can help get them off drugs.
- V19. Provide job training and apprenticeship programs so that all youths will have the opportunity to get good paying jobs as adults.
- V23. Provide help to families and their children as soon as a child shows signs that he or she might later get into trouble with the law.

INDEPENDENT VARIABLES (CONTINUOUS)

- V64 It is important to hate the sin but love the sinner.
- V67 I would describe myself as very religious.

(both of above measured on 7 point scale)

V86 How would you rate your own political views (scale is 1 to 9, 1 (extremely liberal), 5 (moderate) and 9 (extremely conservative))

V87 Income (1 -6 scale, 1=less than \$15,000, 6 =\$75,000 or more)

INDEPENDENT VARIABLES (CATEGORICAL)

V70 Gender (0=female, 1=male)

V71 Race (1=white, 2=black, 3=other)

V72 Education (1 = never went to high school, 2 = went to high school but did not graduate, 3 = graduated from high school, 4 = finished one year of college (or post-high school training), 5 = finished two years of college, 6 = finished three years of college, 7 = graduated from college, 8 = finished one or more years of graduate school)

Education could be categorical or continuous

V73 In the last 12 months, has someone broken into your house? (1=yes,2=no)