What statistical analysis should I use? Statistical analyses using SPSS

Introduction

This page shows how to perform a number of statistical tests using SPSS. Each section gives a brief description of the aim of the statistical test, when it is used, an example showing the SPSS commands and SPSS (often abbreviated) output with a brief interpretation of the output. You can see the page Choosing the Correct Statistical Test for a table that shows an overview of when each test is appropriate to use. In deciding which test is appropriate to use, it is important to consider the type of variables that you have (i.e., whether your variables are categorical, ordinal or interval and whether they are normally distributed), see What is the difference between categorical, ordinal and interval variables? for more information on this.

About the hsb data file

Most of the examples in this page will use a data file called **hsb2,** high school and beyond. This data file contains 200 observations from a sample of high school students with demographic information about the students, such as their gender (**female**), socio-economic status (**ses**) and ethnic background (**race**). It also contains a number of scores on standardized tests, including tests of reading (**read**), writing (**write**), mathematics (**math**) and social studies (**socst**). You can get the hsb data file by clicking on hsb2.

One sample t-test

A one sample t-test allows us to test whether a sample mean (of a normally distributed interval variable) significantly differs from a hypothesized value. For example, using the hsb2 data file, say we wish to test whether the average writing score (**write**) differs significantly from 50. We can do this as shown below.

```
t-test
 /testval = 50
 /variable = write.
```

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
writing score	200	52.7750	9.47859	.67024

One-Sample Test

	Test Value = 50					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
writing score	4.140	199	.000	2.7750	1.4533	4.0967

The mean of the variable **write** for this particular sample of students is 52.775, which is statistically significantly different from the test value of 50. We would conclude that this group of students has a significantly higher mean on the writing test than 50.

One sample median test

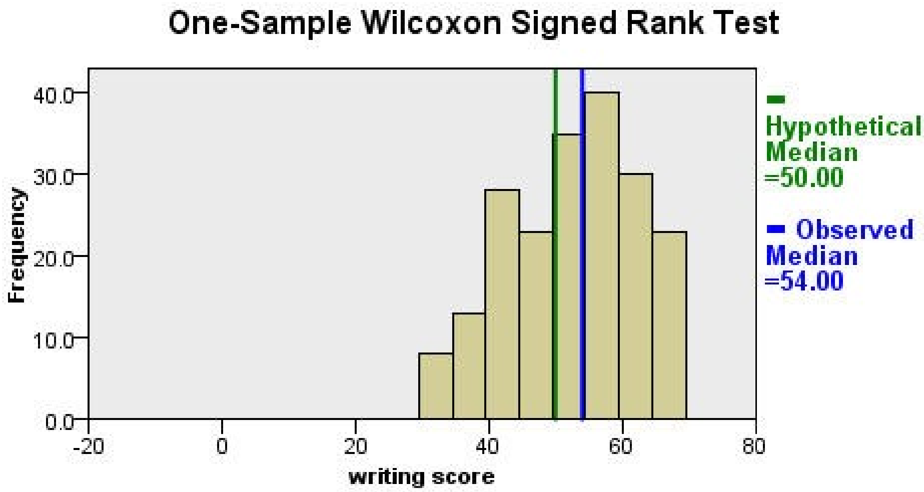
A one sample median test allows us to test whether a sample median differs significantly from a hypothesized value. We will use the same variable, **write**, as we did in the one sample t-test example above, but we do not need to assume that it is interval and normally distributed (we only need to assume that **write** is an ordinal variable).

```
nptests
/onesample test (write) wilcoxon(testvalue = 50).
```

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The median of writing score equals 50.00.	One-Sample Wilcoxon Signed Rank Test	.000	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.



Total N	200
Test Statistic	13,177.000
Standard Error	806.235
Standardized Test Statistic	4.126
Asymptotic Sig. (2-sided test)	.000

Binomial test

A one sample binomial test allows us to test whether the proportion of successes on a two-level categorical dependent variable significantly differs from a hypothesized value. For example, using the hsb2 data file, say we wish to test whether the proportion of females (**female**) differs significantly from 50%, i.e., from .5. We can do this as shown below.

```
npar tests
 /binomial (.5) = female.
```

Binomial Test

		Category	N	Observed Prop.	Test Prop.	Asymp. Sig. (2-tailed)
FEMALE	Group 1	male	91	.46	.50	.229[a]
	Group 2	female	109	.54		
	Total		200	1.00		

a. Based on Z Approximation.

The results indicate that there is no statistically significant difference (p = .229). In other words, the proportion of females in this sample does not significantly differ from the hypothesized value of 50%.

Chi-square goodness of fit

A chi-square goodness of fit test allows us to test whether the observed proportions for a categorical variable differ from hypothesized proportions. For example, let's suppose that we believe that the general population consists of 10% Hispanic, 10% Asian, 10% African American and 70% White folks. We want to test whether the observed proportions from our sample differ significantly from these hypothesized proportions.

```
npar test
 /chisquare = race
 /expected = 10 10 10 70.
```

RACE

	Observed N	Expected N	Residual
hispanic	24	20.0	4.0
asian	11	20.0	-9.0
african-amer	20	20.0	.0
white	145	140.0	5.0
Total	200		

Test Statistics

	RACE
Chi-Square[a]	5.029
df	3
Asymp. Sig.	.170

a. 0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 20.0.

These results show that racial composition in our sample does not differ significantly from the hypothesized values that we supplied (chi-square with three degrees of freedom = 5.029, p = .170).

Two independent samples t-test

An independent samples t-test is used when you want to compare the means of a normally distributed interval dependent variable for two independent groups. For example, using the hsb2 data file, say we wish to test whether the mean for **write** is the same for males and females.

```
t-test groups = female(0 1)
 /variables = write.
```

Group Statistics

	female	N	Mean	Std. Deviation	Std. Error Mean
writing score	male	91	50.1209	10.30516	1.08027
	female	109	54.9908	8.13372	.77907

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
writing score	Equal variances assumed	11.133	.001	-3.734	198	.000	-4.8699	1.30419	-7.44183	-2.29806
	Equal variances not assumed			-3.656	169.707	.000	-4.8699	1.33189	-7.49916	-2.24073

Because the standard deviations for the two groups are similar (10.3 and 8.1), we will use the “equal variances assumed” test. The results indicate that there is a statistically significant difference between the mean writing score for males and females (t = -3.734, p = .000). In other words, females have a statistically significantly higher mean score on writing (54.99) than males (50.12).

See also

- SPSS Learning Module: An overview of statistical tests in SPSS

Wilcoxon-Mann-Whitney test

The Wilcoxon-Mann-Whitney test is a non-parametric analog to the independent samples t-test and can be used when you do not assume that the dependent variable is a normally distributed interval variable (you only assume that the variable is at least ordinal). You will notice that the SPSS syntax for the Wilcoxon-Mann-Whitney test is almost identical to that of the independent samples t-test. We will use the same data file (the hsb2 data file) and the same variables in this example as we did in the independent t-test example above and will not assume that **write**, our dependent variable, is normally distributed.

```
npar test
 /m-w = write by female(0 1).
```

Test Statistics[a]

	writing score
Mann-Whitney U	3606.000
Wilcoxon W	7792.000
Z	-3.329
Asymp. Sig. (2-tailed)	.001

a. Grouping Variable: FEMALE

The results suggest that there is a statistically significant difference between the underlying distributions of the **write** scores of males and the **write** scores of females (z = -3.329, p = 0.001).

See also

- FAQ: Why is the Mann-Whitney significant when the medians are equal?

Chi-square test

A chi-square test is used when you want to see if there is a relationship between two categorical variables. In SPSS, the **chisq** option is used on the **statistics** subcommand of the **crosstabs** command to obtain the test statistic and its associated p-value. Using the hsb2 data file, let's see if there is a relationship between the type of school attended (**schtyp**) and students' gender (**female**). Remember that the chi-square test assumes that the expected value for each cell is five or higher. This assumption is easily met in the examples below. However, if this assumption is not met in your data, please see the section on Fisher's exact test below.

```
crosstabs
 /tables = schtyp by female
 /statistic = chisq.
```

type of school * FEMALE Crosstabulation

Count

		FEMALE		Total
		male	female	
type of school	public	77	91	168
	private	14	18	32
Total		91	109	200

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	.047[b]	1	.828		
Continuity Correction[a]	.001	1	.981		
Likelihood Ratio	.047	1	.828		
Fisher's Exact Test				.849	.492
Linear-by-Linear Association	.047	1	.829		
N of Valid Cases	200				

a. Computed only for a 2x2 table

b. 0 cells (.0%) have expected count less than 5. The minimum expected count is 14.56.

These results indicate that there is no statistically significant relationship between the type of school attended and gender (chi-square with one degree of freedom = 0.047, p = 0.828).

Let's look at another example, this time looking at the linear relationship between gender (**female**) and socio-economic status (**ses**). The point of this example is that one (or both) variables may have more than two levels, and that the variables do not have to have the same number of levels. In this example, **female** has two levels (male and female) and **ses** has three levels (low, medium and high).

```
crosstabs
 /tables = female by ses
 /statistic = chisq.
```

FEMALE * SES Crosstabulation

Count

		SES			Total
		low	middle	high	
FEMALE	male	15	47	29	91
	female	32	48	29	109
Total		47	95	58	200

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	4.577[a]	2	.101
Likelihood Ratio	4.679	2	.096
Linear-by-Linear Association	3.110	1	.078
N of Valid Cases	200		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 21.39.

Again we find that there is no statistically significant relationship between the variables (chi-square with two degrees of freedom = 4.577, p = 0.101).

See also

- SPSS Learning Module: An Overview of Statistical Tests in SPSS

Fisher's exact test

The Fisher's exact test is used when you want to conduct a chi-square test but one or more of your cells has an expected frequency of five or less. Remember that the chi-square test assumes that each cell has an expected frequency of five or more, but the Fisher's exact test has no such assumption and can be used regardless of how small the expected frequency is. In SPSS unless you have the SPSS Exact Test Module, you can only perform a Fisher's exact test on a 2×2 table, and these results are presented by default. Please see the results from the chi squared example above.

One-way ANOVA

A one-way analysis of variance (ANOVA) is used when you have a categorical independent variable (with two or more categories) and a normally distributed interval dependent variable and you wish to test for differences in the means of the dependent variable broken down by the levels of the independent variable. For example, using the hsb2 data file, say we wish to test whether the mean of **write** differs between the three program types (**prog**). The command for this test would be:

```
oneway write by prog.
```

ANOVA

writing score

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	3175.698	2	1587.849	21.275	.000
Within Groups	14703.177	197	74.635		
Total	17878.875	199			

The mean of the dependent variable differs significantly among the levels of program type. However, we do not know if the difference is between only two of the levels or all three of the levels. (The F test for the **Model** is the same as the F test for **prog** because **prog** was the only variable entered into the model. If other variables had also been entered, the F test for the **Model** would have been different from **prog**.) To see the mean of **write** for each level of program type,

```
means tables = write by prog.
```

Report

writing score

type of program	Mean	N	Std. Deviation
general	51.3333	45	9.39778
academic	56.2571	105	7.94334
vocation	46.7600	50	9.31875
Total	52.7750	200	9.47859

From this we can see that the students in the academic program have the highest mean writing score, while students in the vocational program have the lowest.

See also

- SPSS Textbook Examples: Design and Analysis, Chapter 7
- SPSS Textbook Examples: Applied Regression Analysis, Chapter 8
- SPSS FAQ: How can I do ANOVA contrasts in SPSS?
- SPSS Library: Understanding and Interpreting Parameter Estimates in Regression and ANOVA

Kruskal Wallis test

The Kruskal Wallis test is used when you have one independent variable with two or more levels and an ordinal dependent variable. In other words, it is the non-parametric version of ANOVA and a generalized form of the Mann-Whitney test method since it permits two or more groups. We will use the same data file as the one way ANOVA example above (the hsb2 data file) and the same variables as in the example above, but we will not assume that **write** is a normally distributed interval variable.

```
npar tests
 /k-w = write by prog (1,3).
```

Ranks

	type of program	N	Mean Rank
writing score	general	45	90.64
	academic	105	121.56
	vocation	50	65.14
	Total	200	

Test Statistics[a,b]

	writing score
Chi-Square	34.045
df	2
Asymp. Sig.	.000

a. Kruskal Wallis Test

b. Grouping Variable: type of program

If some of the scores receive tied ranks, then a correction factor is used, yielding a slightly different value of chi-squared. With or without ties, the results indicate that there is a statistically significant difference among the three type of programs.

Paired t-test

A paired (samples) t-test is used when you have two related observations (i.e., two observations per subject) and you want to see if the means on these two normally distributed interval variables differ from one another. For example, using the hsb2 data file we will test whether the mean of **read** is equal to the mean of **write**.

```
t-test pairs = read with write (paired).
```

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	reading score	52.2300	200	10.25294	.72499
	writing score	52.7750	200	9.47859	.67024

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	reading score - writing score	-.5450	8.88667	.62838	-1.7841	.6941	-.867	199	.387

These results indicate that the mean of **read** is not statistically significantly different from the mean of **write** (t = -0.867, p = 0.387).

Wilcoxon signed rank sum test

The Wilcoxon signed rank sum test is the non-parametric version of a paired samples t-test. You use the Wilcoxon signed rank sum test when you do not wish to assume that the difference between the two variables is interval and normally distributed (but you do assume the difference is ordinal). We will use the same example as above, but we will not assume that the difference between **read** and **write** is interval and normally distributed.

```
npar test
 /wilcoxon = write with read (paired).
```

Ranks

		N	Mean Rank	Sum of Ranks
reading score - writing score	Negative Ranks	97[a]	95.47	9261.00
	Positive Ranks	88[b]	90.27	7944.00
	Ties	15[c]		
	Total	200		

a. reading score < writing score

b. reading score > writing score

c. writing score = reading score

Test Statistics[b]

	reading score - writing score
Z	-.903[a]
Asymp. Sig. (2-tailed)	.366

a. Based on positive ranks.

b. Wilcoxon Signed Ranks Test

The results suggest that there is not a statistically significant difference between **read** and **write**.

If you believe the differences between **read** and **write** were not ordinal but could merely be classified as positive and negative, then you may want to consider a sign test in lieu of sign rank test. Again, we will use the same variables in this example and assume that this difference is not ordinal.

```
npar test
 /sign = read with write (paired).
```

Frequencies

		N
writing score - reading score	Negative Differences[a]	88
	Positive Differences[b]	97
	Ties[c]	15
	Total	200

a. writing score < reading score

b. writing score > reading score

c. reading score = writing score

Test Statistics[a]

	writing score - reading score
Z	-.588
Asymp. Sig. (2-tailed)	.556

a. Sign Test

We conclude that no statistically significant difference was found (p=.556).

McNemar test

You would perform McNemar's test if you were interested in the marginal frequencies of two binary outcomes. These binary outcomes may be the same outcome variable on matched pairs (like a case-control study) or two outcome variables from a single group. Continuing with the hsb2 dataset used in several above examples, let us create two binary outcomes in our dataset: **himath** and **hiread**. These outcomes can be considered in a two-way contingency table. The null hypothesis is that the proportion of students in the **himath** group is the same as the proportion of students in **hiread** group (i.e., that the contingency table is symmetric).

```
compute himath = (math>60).
compute hiread = (read>60).
execute.

crosstabs
  /tables=himath  BY hiread
  /statistic=mcnemar
  /cells=count.
```

himath * hiread Crosstabulation

Count

		hiread		
		.00	1.00	Total
himath	.00	135	21	156
	1.00	18	26	44
Total		153	47	200

Chi-Square Tests

	Value	Exact Sig. (2-sided)
McNemar Test		.749[a]
N of Valid Cases	200	

a. Binomial distribution used.

McNemar's chi-square statistic suggests that there is not a statistically significant difference in the proportion of students in the **himath** group and the proportion of students in the **hiread** group.

One-way repeated measures ANOVA

You would perform a one-way repeated measures analysis of variance if you had one categorical independent variable and a normally distributed interval dependent variable that was repeated at least twice for each subject. This is the equivalent of the paired samples t-test, but allows for two or more levels of the categorical variable. This tests whether the mean of the dependent variable differs by the categorical variable. We have an example data set called rb4wide, which is used in Kirk's book Experimental Design. In this data set, **y** is the dependent variable, **a** is the repeated measure and **s** is the variable that indicates the subject number.

```
glm y1 y2 y3 y4
 /wsfactor a(4).
```

Within-Subjects Factors

Measure: MEASURE_1

A	Dependent Variable
1	Y1
2	Y2
3	Y3
4	Y4

Multivariate Tests[b]

Effect		Value	F	Hypothesis df	Error df	Sig.
A	Pillai's Trace	.754	5.114[a]	3.000	5.000	.055
	Wilks' Lambda	.246	5.114[a]	3.000	5.000	.055
	Hotelling's Trace	3.068	5.114[a]	3.000	5.000	.055
	Roy's Largest Root	3.068	5.114[a]	3.000	5.000	.055

a. Exact statistic

b.

Design: Intercept

Within Subjects Design: A

Mauchly's Test of Sphericity[b]

Measure: MEASURE_1

					Epsilon[a]		
Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Greenhouse-Geisser	Huynh-Feldt	Lower-bound
A	.339	6.187	5	.295	.620	.834	.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: Intercept

Within Subjects Design: A

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
A	Sphericity Assumed	49.000	3	16.333	11.627	.000
	Greenhouse-Geisser	49.000	1.859	26.365	11.627	.001
	Huynh-Feldt	49.000	2.503	19.578	11.627	.000
	Lower-bound	49.000	1.000	49.000	11.627	.011
Error(A)	Sphericity Assumed	29.500	21	1.405		
	Greenhouse-Geisser	29.500	13.010	2.268		
	Huynh-Feldt	29.500	17.520	1.684		
	Lower-bound	29.500	7.000	4.214		

Tests of Within-Subjects Contrasts

Measure: MEASURE_1

Source	A	Type III Sum of Squares	df	Mean Square	F	Sig.
A	Linear	44.100	1	44.100	19.294	.003
	Quadratic	4.500	1	4.500	3.150	.119
	Cubic	.400	1	.400	.800	.401
Error(A)	Linear	16.000	7	2.286		
	Quadratic	10.000	7	1.429		
	Cubic	3.500	7	.500		

Tests of Between-Subjects Effects

Measure: MEASURE_1
Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	578.000	1	578.000	128.444	.000
Error	31.500	7	4.500		

You will notice that this output gives four different p-values. The output labeled "sphericity assumed" is the p-value (0.000) that you would get if you assumed compound symmetry in the variance-covariance matrix. Because that assumption is often not valid, the three other p-values offer various corrections (the Huynh-Feldt, H-F, Greenhouse-Geisser, G-G and Lower-bound). No matter which p-value you use, our results indicate that we have a statistically significant effect of **a** at the .05 level.

See also

- [SPSS Textbook Examples from Design and Analysis: Chapter 16]
- [SPSS Library: Advanced Issues in Using and Understanding SPSS MANOVA]
- [SPSS Code Fragment: Repeated Measures ANOVA]

Repeated measures logistic regression

If you have a binary outcome measured repeatedly for each subject and you wish to run a logistic regression that accounts for the effect of multiple measures from single subjects, you can perform a repeated measures logistic regression. In SPSS, this can be done using the **GENLIN** command and indicating binomial as the probability distribution and logit as the link function to be used in the model. The exercise data file contains 3 pulse measurements from each of 30 people assigned to 2 different diet regiments and 3 different exercise regiments. If we define a "high" pulse as being over 100, we can then predict the probability of a high pulse using diet regiment.

```
GET FILE='C:mydatahttps://stats.idre.ucla.edu/wp-content
/uploads/2016/02/exercise.sav'.

GENLIN highpulse (REFERENCE=LAST)
  BY diet (order = DESCENDING)
/MODEL diet
  DISTRIBUTION=BINOMIAL
  LINK=LOGIT
/REPEATED SUBJECT=id CORRTYPE = EXCHANGEABLE.
```

Tests of Model Effects

Source	Type III		
	Wald Chi-Square	df	Sig.

(Intercept)	8.437	1	.004
diet	1.562	1	.211

Dependent Variable: highpulse
Model: (Intercept), diet

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test		
			Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	1.253	.4328	.404	2.101	8.377	1	.004
[diet=2]	-.754	.6031	-1.936	.428	1.562	1	.211
[diet=1]	0[a]	.	.	.	.	.	.
(Scale)	1						

Dependent Variable: highpulse
Model: (Intercept), diet

a. Set to zero because this parameter is redundant.

These results indicate that **diet** is not statistically significant (Wald Chi-Square = 1.562, p = 0.211).

Factorial ANOVA

A factorial ANOVA has two or more categorical independent variables (either with or without the interactions) and a single normally distributed interval dependent variable. For example, using the hsb2 data file we will look at writing scores (**write**) as the dependent variable and gender (**female**) and socio-economic status (**ses**) as independent variables, and we will include an interaction of **female** by **ses**. Note that in SPSS, you do not need to have the interaction term(s) in your data set. Rather, you can have SPSS create it/them temporarily by placing an asterisk between the variables that will make up the interaction term(s).

```
glm write by female ses.
```

Tests of Between-Subjects Effects

Dependent Variable: writing score

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	2278.244[a]	5	455.649	5.666	.000
Intercept	473967.467	1	473967.467	5893.972	.000
FEMALE	1334.493	1	1334.493	16.595	.000
SES	1063.253	2	531.626	6.611	.002
FEMALE * SES	21.431	2	10.715	.133	.875
Error	15600.631	194	80.416		
Total	574919.000	200			
Corrected Total	17878.875	199			

a. R Squared = .127 (Adjusted R Squared = .105)

These results indicate that the overall model is statistically significant (F = 5.666, p = 0.00). The variables **female** and **ses** are also statistically significant (F = 16.595, p = 0.000 and F = 6.611, p = 0.002, respectively). However, that interaction between **female** and **ses** is not statistically significant (F = 0.133, p = 0.875).

See also

- SPSS Textbook Examples from Design and Analysis: Chapter 10
- SPSS FAQ: How can I do tests of simple main effects in SPSS?
- SPSS FAQ: How do I plot ANOVA cell means in SPSS?
- SPSS Library: An Overview of SPSS GLM

Friedman test

You perform a Friedman test when you have one within-subjects independent variable with two or more levels and a dependent variable that is not interval and normally distributed (but at least ordinal). We will use this test to determine if there is a difference in the reading, writing and math scores. The null hypothesis in this test is that the distribution of the ranks of each type of score (i.e., reading, writing and math) are the same. To conduct a Friedman test, the data need to be in a long format. SPSS handles this for you, but in other statistical packages you will have to reshape the data before you can conduct this test.

```
npar tests
 /friedman = read write math.
```

Ranks

	Mean Rank
reading score	1.96
writing score	2.04
math score	2.01

Test Statistics[a]

N	200
Chi-Square	.645
df	2
Asymp. Sig.	.724

a. Friedman Test

Friedman's chi-square has a value of 0.645 and a p-value of 0.724 and is not statistically significant. Hence, there is no evidence that the distributions of the three types of scores are different.

Ordered logistic regression

Ordered logistic regression is used when the dependent variable is ordered, but not continuous. For example, using the hsb2 data file we will create an ordered variable called **write3**. This variable will have the values 1, 2 and 3, indicating a low, medium or high writing score. We do not generally recommend categorizing a continuous variable in this way; we are simply creating a variable to use for this example. We will use gender (**female**), reading score (**read**) and social studies score (**socst**) as predictor variables in this model. We will use a logit link and on the **print** subcommand we have requested the parameter estimates, the (model) summary statistics and the test of the parallel lines assumption.

```
if write ge 30 and write le 48  write3 = 1.
if write ge 49 and write le 57  write3 = 2.
if write ge 58 and write le 70  write3 = 3.
execute.

plum write3 with female read socst
/link = logit
/print = parameter summary tparallel.
```

Case Processing Summary

		N	Marginal Percentage
write3	1.00	61	30.5%
	2.00	61	30.5%
	3.00	78	39.0%
Valid		200	100.0%
Missing		0	
Total		200	

Model Fitting Information

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	376.226			
Final	252.151	124.075	3	.000

Link function: Logit.

Pseudo R-Square

Cox and Snell	.462
Nagelkerke	.521
McFadden	.284

Link function: Logit.

Parameter Estimates

		Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
Threshold	[write3 = 1.00]	9.704	1.203	65.109	1	.000	7.347	12.061
	[write3 = 2.00]	11.800	1.312	80.868	1	.000	9.228	14.372
Location	female	1.285	.322	15.887	1	.000	.653	1.918
	read	.118	.022	29.867	1	.000	.076	.160
	socst	.080	.019	17.781	1	.000	.043	.117

Link function: Logit.

Test of Parallel Lines[a]

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Null Hypothesis	252.151			
General	250.104	2.047	3	.563

The null hypothesis states that the location parameters (slope coefficients) are the same across response categories.

a. Link function: Logit.

The results indicate that the overall model is statistically significant (p < .000), as are each of the predictor variables (p < .000). There are two thresholds for this model because there are three levels of the outcome variable. We also see that the test of the proportional odds assumption is non-significant (p = .563). One of the assumptions underlying ordinal logistic (and ordinal probit) regression is that the relationship between each pair of outcome groups is the same. In other words, ordinal logistic regression assumes that the coefficients that describe the relationship between, say, the lowest versus all higher categories of the response variable are the same as those that describe the relationship between the next lowest category and all higher categories, etc. This is called the proportional odds assumption or the parallel regression assumption. Because the relationship between all pairs of groups is the same, there is only one set of coefficients (only one model). If this was not the case, we would need different models (such as a generalized ordered logit model) to describe the relationship between each pair of outcome groups.

See also

- SPSS Data Analysis Examples: Ordered logistic regression
- SPSS Annotated Output: Ordinal Logistic Regression

Factorial logistic regression

A factorial logistic regression is used when you have two or more categorical independent variables but a dichotomous dependent variable. For example, using the hsb2 data file we will use **female** as our dependent variable, because it is the only dichotomous variable in our data set; certainly not because it common practice to use gender as an outcome variable. We will use type of program (**prog**) and school type (**schtyp**) as our predictor variables. Because **prog** is a categorical variable (it has three levels), we need to create dummy codes for it. SPSS will do this for you by making dummy codes for all variables listed after the keyword **with**. SPSS will also create the interaction term; simply list the two variables that will make up the interaction separated by the keyword **by**.

```
logistic regression female with prog schtyp prog by schtyp
 /contrast(prog) = indicator(1).
```

Dependent Variable Encoding

Original Value	Internal Value
male	0
female	1

Categorical Variables Codings

		Frequency	Parameter coding (1)	Parameter coding (2)
type of program	general	45	.000	.000
	academic	105	1.000	.000
	vocation	50	.000	1.000

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	3.147	5	.677
	Block	3.147	5	.677
	Model	3.147	5	.677

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	272.490	.016	.021

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	PROG			2.595	2	.273	
	PROG(1)	2.258	1.407	2.578	1	.108	9.568
	PROG(2)	2.046	1.986	1.061	1	.303	7.737
	SCHTYP	1.661	1.141	2.117	1	.146	5.262
	PROG * SCHTYP			2.474	2	.290	
	PROG(1) by SCHTYP	-1.934	1.233	2.461	1	.117	.145
	PROG(2) by SCHTYP	-1.828	1.840	.986	1	.321	.161
	Constant	-1.712	1.269	1.820	1	.177	.181

a. Variable(s) entered on step 1: PROG, SCHTYP, PROG * SCHTYP .

The results indicate that the overall model is not statistically significant (LR chi2 = 3.147, p = 0.677). Furthermore, none of the coefficients are statistically significant either. This shows that the overall effect of **prog** is not significant.

See also

- Annotated output for logistic regression

Correlation

A correlation is useful when you want to see the relationship between two (or more) normally distributed interval variables. For example, using the hsb2 data file we can run a correlation between two continuous variables, **read** and **write**.

```
correlations
 /variables = read write.
```

Correlations

		reading score	writing score
reading score	Pearson Correlation	1	.597
	Sig. (2-tailed)	.	.000
	N	200	200
writing score	Pearson Correlation	.597	1
	Sig. (2-tailed)	.000	.
	N	200	200

In the second example, we will run a correlation between a dichotomous variable, **female**, and a continuous variable, **write**. Although it is assumed that the variables are interval and normally distributed, we can include dummy variables when performing correlations.

```
correlations
 /variables =  female write.
```

Correlations

		FEMALE	writing score
FEMALE	Pearson Correlation	1	.256
	Sig. (2-tailed)	.	.000
	N	200	200
writing score	Pearson Correlation	.256	1
	Sig. (2-tailed)	.000	.
	N	200	200

In the first example above, we see that the correlation between **read** and **write** is 0.597. By squaring the correlation and then multiplying by 100, you can determine what percentage of the variability is shared. Let's round 0.597 to be 0.6, which when squared would be .36, multiplied by 100 would be 36%. Hence **read** shares about 36% of its variability with **write**. In the output for the second example, we can see the correlation between **write** and **female** is 0.256. Squaring this number yields .065536, meaning that **female** shares approximately 6.5% of its variability with **write**.

See also

- Annotated output for correlation
- SPSS Learning Module: An Overview of Statistical Tests in SPSS
- SPSS FAQ: How can I analyze my data by categories?
- Missing Data in SPSS

Simple linear regression

Simple linear regression allows us to look at the linear relationship between one normally distributed interval predictor and one normally distributed interval outcome variable. For example, using the hsb2 data file, say we wish to look at the relationship between writing scores (**write**) and reading scores (**read**); in other words, predicting **write** from **read**.

```
regression variables = write read
 /dependent = write
 /method = enter.
```

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.597[a]	.356	.353	7.62487

a. Predictors: (Constant), reading score

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	6367.421	1	6367.421	109.521	.000[a]
	Residual	11511.454	198	58.139		
	Total	17878.875	199			

a. Predictors: (Constant), reading score

b. Dependent Variable: writing score

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	23.959	2.806		8.539	.000
	reading score	.552	.053	.597	10.465	.000

a. Dependent Variable: writing score

We see that the relationship between **write** and **read** is positive (.552) and based on the t-value (10.47) and p-value (0.000), we would conclude this relationship is statistically significant. Hence, we would say there is a statistically significant positive linear relationship between reading and writing.

See also

- Regression With SPSS: Chapter 1 – Simple and Multiple Regression
- Annotated output for regression
- SPSS Textbook Examples: Introduction to the Practice of Statistics, Chapter 10
- SPSS Textbook Examples: Regression with Graphics, Chapter 2
- SPSS Textbook Examples: Applied Regression Analysis, Chapter 5

Non-parametric correlation

A Spearman correlation is used when one or both of the variables are not assumed to be normally distributed and interval (but are assumed to be ordinal). The values of the variables are converted in ranks and then correlated. In our example, we will look for a relationship between **read** and **write**. We will not assume that both of these variables are normal and interval.

```
nonpar corr
 /variables = read write
 /print = spearman.
```

Correlations

			reading score	writing score
Spearman's rho	reading score	Correlation Coefficient	1.000	.617
		Sig. (2-tailed)	.	.000
		N	200	200
	writing score	Correlation Coefficient	.617	1.000
		Sig. (2-tailed)	.000	.
		N	200	200

The results suggest that the relationship between **read** and **write** (rho = 0.617, p = 0.000) is statistically significant.

Simple logistic regression

Logistic regression assumes that the outcome variable is binary (i.e., coded as 0 and 1). We have only one variable in the hsb2 data file that is coded 0 and

1, and that is **female**. We understand that **female** is a silly outcome variable (it would make more sense to use it as a predictor variable), but we can use **female** as the outcome variable to illustrate how the code for this command is structured and how to interpret the output. The first variable listed after the **logistic** command is the outcome (or dependent) variable, and all of the rest of the variables are predictor (or independent) variables. In our example, **female** will be the outcome variable, and **read** will be the predictor variable. As with OLS regression, the predictor variables must be either dichotomous or continuous; they cannot be categorical.

```
logistic regression female with read.
```

Dependent Variable Encoding

Original Value	Internal Value
male	0
female	1

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	.564	1	.453
	Block	.564	1	.453
	Model	.564	1	.453

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	275.073	.003	.004

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	READ	-.010	.014	.562	1	.453	.990
	Constant	.726	.742	.958	1	.328	2.067

a. Variable(s) entered on step 1: READ.

The results indicate that reading score (**read**) is not a statistically significant predictor of gender (i.e., being female), Wald = .562, p = 0.453. Likewise, the test of the overall model is not statistically significant, LR chi-squared – 0.56, p = 0.453.

See also

- [Annotated output for logistic regression]
- [SPSS Library: What kind of contrasts are these?]

Multiple regression

Multiple regression is very similar to simple regression, except that in multiple regression you have more than one predictor variable in the equation. For example, using the hsb2 data file we will predict writing score from gender (**female**), reading, math, science and social studies (**socst**) scores.

```
regression variable = write female read math science socst
 /dependent = write
 /method = enter.
```

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.776[a]	.602	.591	6.05897

a. Predictors: (Constant), social studies score, FEMALE, science score, math score, reading score

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	10756.924	5	2151.385	58.603	.000[a]
	Residual	7121.951	194	36.711		
	Total	17878.875	199			

a. Predictors: (Constant), social studies score, FEMALE, science score, math score, reading score

b. Dependent Variable: writing score

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	6.139	2.808		2.186	.030
	FEMALE	5.493	.875	.289	6.274	.000
	reading score	.125	.065	.136	1.931	.055

math score	.238	.067	.235	3.547	.000
science score	.242	.061	.253	3.986	.000
social studies score	.229	.053	.260	4.339	.000

a. Dependent Variable: writing score

The results indicate that the overall model is statistically significant (F = 58.60, p = 0.000). Furthermore, all of the predictor variables are statistically significant except for **read**.

See also

- Regression with SPSS: Chapter 1 – Simple and Multiple Regression
- Annotated output for regression
- SPSS Frequently Asked Questions
- SPSS Textbook Examples: Regression with Graphics, Chapter 3
- SPSS Textbook Examples: Applied Regression Analysis

Analysis of covariance

Analysis of covariance is like ANOVA, except in addition to the categorical predictors you also have continuous predictors as well. For example, the one way ANOVA example used **write** as the dependent variable and **prog** as the independent variable. Let's add **read** as a continuous variable to this model, as shown below.

```
glm write with read by prog.
```

Tests of Between-Subjects Effects

Dependent Variable: writing score

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	7017.681[a]	3	2339.227	42.213	.000
Intercept	4867.964	1	4867.964	87.847	.000
READ	3841.983	1	3841.983	69.332	.000
PROG	650.260	2	325.130	5.867	.003
Error	10861.194	196	55.414		
Total	574919.000	200			
Corrected Total	17878.875	199			

a. R Squared = .393 (Adjusted R Squared = .383)

The results indicate that even after adjusting for reading score (**read**), writing scores still significantly differ by program type (**prog**), F = 5.867, p = 0.003.

See also

- SPSS Textbook Examples from Design and Analysis: Chapter 14
- SPSS Library: An Overview of SPSS GLM
- SPSS Library: How do I handle interactions of continuous and categorical variables?

Multiple logistic regression

Multiple logistic regression is like simple logistic regression, except that there are two or more predictors. The predictors can be interval variables or dummy variables, but cannot be categorical variables. If you have categorical predictors, they should be coded into one or more dummy variables. We have only one variable in our data set that is coded 0 and 1, and that is **female**. We understand that **female** is a silly outcome variable (it would make more sense to use it as a predictor variable), but we can use **female** as the outcome variable to illustrate how the code for this command is structured and how to interpret the output. The first variable listed after the **logistic regression** command is the outcome (or dependent) variable, and all of the rest of the variables are predictor (or independent) variables (listed after the keyword **with**). In our example, **female** will be the outcome variable, and **read** and **write** will be the predictor variables.

```
logistic regression female with read write.
```

Dependent Variable Encoding

Original Value	Internal Value
male	0
female	1

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	27.819	2	.000
	Block	27.819	2	.000
	Model	27.819	2	.000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	247.818	.130	.174

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	READ	-.071	.020	13.125	1	.000	.931
	WRITE	.106	.022	23.075	1	.000	1.112
	Constant	-1.706	.923	3.414	1	.065	.182

a. Variable(s) entered on step 1: READ, WRITE.

These results show that both **read** and **write** are significant predictors of **female**.

See also

- Annotated output for logistic regression

- SPSS Textbook Examples: Applied Logistic Regression, Chapter 2
- SPSS Code Fragments: Graphing Results in Logistic Regression

Discriminant analysis

Discriminant analysis is used when you have one or more normally distributed interval independent variables and a categorical dependent variable. It is a multivariate technique that considers the latent dimensions in the independent variables for predicting group membership in the categorical dependent variable. For example, using the hsb2 data file, say we wish to use **read**, **write** and **math** scores to predict the type of program a student belongs to (**prog**).

```
discriminate groups = prog(1, 3)
 /variables = read write math.
```

Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	.356[a]	98.7	98.7	.513
2	.005[a]	1.3	100.0	.067

a. First 2 canonical discriminant functions were used in the analysis.

Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1 through 2	.734	60.619	6	.000
2	.995	.888	2	.641

Standardized Canonical Discriminant Function Coefficients

	Function	
	1	2
reading score	.273	-.410
writing score	.331	1.183
math score	.582	-.656

Structure Matrix

	Function	
	1	2
math score	.913*	-.272
reading score	.778*	-.184

writing score	.775*	.630

Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions
Variables ordered by absolute size of correlation within function.
*. Largest absolute correlation between each variable and any discriminant function

Functions at Group Centroids

	Function	
type of program	1	2
general	-.312	.119
academic	.536	-1.97E-02
vocation	-.844	-6.58E-02

Unstandardized canonical discriminant functions evaluated at group means

Clearly, the SPSS output for this procedure is quite lengthy, and it is beyond the scope of this page to explain all of it. However, the main point is that two canonical variables are identified by the analysis, the first of which seems to be more related to program type than the second.

See also

- discriminant function analysis
- SPSS Library: A History of SPSS Statistical Features

One-way MANOVA

MANOVA (multivariate analysis of variance) is like ANOVA, except that there are two or more dependent variables. In a one-way MANOVA, there is one categorical independent variable and two or more dependent variables. For example, using the hsb2 data file, say we wish to examine the differences in **read**, **write** and **math** broken down by program type (**prog**).

```
glm read write math by prog.
```

Multivariate Tests[c]

Effect		Value	F	Hypothesis df	Error df	Sig.
Intercept	Pillai's Trace	.978	2883.051[a]	3.000	195.000	.000
	Wilks' Lambda	.022	2883.051[a]	3.000	195.000	.000

	Hotelling's Trace	44.355	2883.051[a]	3.000	195.000	.000
	Roy's Largest Root	44.355	2883.051[a]	3.000	195.000	.000
PROG	Pillai's Trace	.267	10.075	6.000	392.000	.000
	Wilks' Lambda	.734	10.870[a]	6.000	390.000	.000
	Hotelling's Trace	.361	11.667	6.000	388.000	.000
	Roy's Largest Root	.356	23.277[b]	3.000	196.000	.000

a. Exact statistic

b. The statistic is an upper bound on F that yields a lower bound on the significance level.

c. Design: Intercept+PROG

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	reading score	3716.861[a]	2	1858.431	21.282	.000
	writing score	3175.698[a]	2	1587.849	21.275	.000
	math score	4002.104[b]	2	2001.052	29.279	.000
Intercept	reading score	447178.672	1	447178.672	5120.994	.000
	writing score	460403.797	1	460403.797	6168.704	.000
	math score	453421.258	1	453421.258	6634.435	.000
PROG	reading score	3716.861	2	1858.431	21.282	.000
	writing score	3175.698	2	1587.849	21.275	.000
	math score	4002.104	2	2001.052	29.279	.000
Error	reading score	17202.559	197	87.323		
	writing score	14703.177	197	74.635		
	math score	13463.691	197	68.344		
Total	reading score	566514.000	200			
	writing score	574919.000	200			
	math score	571765.000	200			
Corrected Total	reading score	20919.420	199			
	writing score	17878.875	199			
	math score	17465.795	199			

a. R Squared = .178 (Adjusted R Squared = .169)

b. R Squared = .229 (Adjusted R Squared = .221)

The students in the different programs differ in their joint distribution of **read**, **write** and **math**.

See also

- SPSS Library: Advanced Issues in Using and Understanding SPSS MANOVA
- GLM: MANOVA and MANCOVA
- SPSS Library: MANOVA and GLM

Multivariate multiple regression

Multivariate multiple regression is used when you have two or more dependent variables that are to be predicted from two or more independent variables. In our example using the hsb2 data file, we will predict **write** and **read** from **female**, **math**, **science** and social studies (**socst**) scores.

```
glm write read with female math science socst.
```

Multivariate Tests[b]

Effect		Value	F	Hypothesis df	Error df	Sig.
Intercept	Pillai's Trace	.030	3.019[a]	2.000	194.000	.051
	Wilks' Lambda	.970	3.019[a]	2.000	194.000	.051
	Hotelling's Trace	.031	3.019[a]	2.000	194.000	.051
	Roy's Largest Root	.031	3.019[a]	2.000	194.000	.051
FEMALE	Pillai's Trace	.170	19.851[a]	2.000	194.000	.000
	Wilks' Lambda	.830	19.851[a]	2.000	194.000	.000
	Hotelling's Trace	.205	19.851[a]	2.000	194.000	.000
	Roy's Largest Root	.205	19.851[a]	2.000	194.000	.000
MATH	Pillai's Trace	.160	18.467[a]	2.000	194.000	.000
	Wilks' Lambda	.840	18.467[a]	2.000	194.000	.000
	Hotelling's Trace	.190	18.467[a]	2.000	194.000	.000
	Roy's Largest Root	.190	18.467[a]	2.000	194.000	.000
SCIENCE	Pillai's Trace	.166	19.366[a]	2.000	194.000	.000
	Wilks' Lambda	.834	19.366[a]	2.000	194.000	.000
	Hotelling's Trace	.200	19.366[a]	2.000	194.000	.000
	Roy's Largest Root	.200	19.366[a]	2.000	194.000	.000
SOCST	Pillai's Trace	.221	27.466[a]	2.000	194.000	.000
	Wilks' Lambda	.779	27.466[a]	2.000	194.000	.000
	Hotelling's Trace	.283	27.466[a]	2.000	194.000	.000
	Roy's Largest Root	.283	27.466[a]	2.000	194.000	.000

a. Exact statistic

b. Design: Intercept+FEMALE+MATH+SCIENCE+SOCST

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	writing score	10620.092[a]	4	2655.023	71.325	.000
	reading score	12219.658[b]	4	3054.915	68.474	.000
Intercept	writing score	202.117	1	202.117	5.430	.021
	reading score	55.107	1	55.107	1.235	.268
female	writing score	1413.528	1	1413.528	37.973	.000
	reading score	12.605	1	12.605	.283	.596
math	writing score	714.867	1	714.867	19.204	.000
	reading score	1025.673	1	1025.673	22.990	.000
science	writing score	857.882	1	857.882	23.046	.000
	reading score	946.955	1	946.955	21.225	.000

socst	writing score	1105.653	1	1105.653	29.702	.000
	reading score	1475.810	1	1475.810	33.079	.000
Error	writing score	7258.783	195	37.225		
	reading score	8699.762	195	44.614		
Total	writing score	574919.000	200			
	reading score	566514.000	200			
Corrected Total	writing score	17878.875	199			
	reading score	20919.420	199			

a. R Squared = .594 (Adjusted R Squared = .586)

b. R Squared = .584 (Adjusted R Squared = .576)

These results show that all of the variables in the model have a statistically significant relationship with the joint distribution of **write** and **read**.

Canonical correlation

Canonical correlation is a multivariate technique used to examine the relationship between two groups of variables. For each set of variables, it creates latent variables and looks at the relationships among the latent variables. It assumes that all variables in the model are interval and normally distributed. SPSS requires that each of the two groups of variables be separated by the keyword **with**. There need not be an equal number of variables in the two groups (before and after the **with**).

```
manova read write with math science
 /discrim.
```

```
* * * * * * A n a l y s i s   o f   V a r i a n c e --
design   1 * * * * * *

 EFFECT .. WITHIN CELLS Regression
 Multivariate Tests of Significance (S = 2, M = -1/2, N =
97 )

 Test Name         Value  Approx. F Hypoth. DF   Error DF
Sig. of F
```

```
 Pillais           .59783   41.99694       4.00     394.00
.000
 Hotellings       1.48369   72.32964       4.00     390.00
.000
 Wilks             .40249   56.47060       4.00     392.00
.000
 Roys              .59728
 Note.. F statistic for WILKS' Lambda is exact.

 - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
- - - - - - - - -
 EFFECT .. WITHIN CELLS Regression (Cont.)
 Univariate F-tests with (2,197) D. F.

 Variable    Sq. Mul. R  Adj. R-sq.  Hypoth. MS    Error
MS           F

 READ             .51356      .50862  5371.66966
51.65523   103.99081
 WRITE            .43565      .42992  3894.42594
51.21839    76.03569

 Variable     Sig. of F

 READ              .000
 WRITE             .000

 - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
- - - - - - - - -
```

```
 Raw canonical coefficients for DEPENDENT variables
           Function No.

 Variable            1

 READ             .063
 WRITE            .049

 - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
- - - - - - - -
 Standardized canonical coefficients for DEPENDENT
variables
           Function No.

 Variable            1

 READ             .649
 WRITE            .467

* * * * * * A n a l y s i s   o f   V a r i a n c e --
design   1 * * * * * *

 Correlations between DEPENDENT and canonical variables
           Function No.

 Variable            1

 READ             .927
 WRITE            .854
```

```
- - - - - - - - - - - - - - - - - - - - - - - - - - - - -
- - - - - - - -
 Variance in dependent variables explained by canonical
variables

 CAN. VAR.  Pct Var DE Cum Pct DE Pct Var CO Cum Pct CO

        1       79.441     79.441     47.449     47.449

- - - - - - - - - - - - - - - - - - - - - - - - - - - - -
- - - - - - - -
 Raw canonical coefficients for COVARIATES
           Function No.

 COVARIATE            1

 MATH              .067
 SCIENCE           .048

- - - - - - - - - - - - - - - - - - - - - - - - - - - - -
- - - - - - - -
 Standardized canonical coefficients for COVARIATES
           CAN. VAR.

 COVARIATE            1

 MATH              .628
 SCIENCE           .478

- - - - - - - - - - - - - - - - - - - - - - - - - - - - -
```

```
- - - - - - - -
 Correlations between COVARIATES and canonical variables
           CAN. VAR.

 Covariate           1

 MATH             .929
 SCIENCE          .873

* * * * * * A n a l y s i s   o f   V a r i a n c e --
design   1 * * * * * *

 Variance in covariates explained by canonical variables

 CAN. VAR.  Pct Var DE Cum Pct DE Pct Var CO Cum Pct CO

       1       48.544     48.544     81.275     81.275

 - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
- - - - - - - -
 Regression analysis for WITHIN CELLS error term
 --- Individual Univariate .9500 confidence intervals
 Dependent variable .. READ              reading score

 COVARIATE            B        Beta   Std. Err.
t-Value   Sig. of t

 MATH             .48129      .43977        .070
6.868        .000
 SCIENCE          .36532      .35278        .066
```

```
5.509        .000

 COVARIATE   Lower -95%  CL- Upper

 MATH              .343        .619
 SCIENCE           .235        .496
 Dependent variable .. WRITE             writing score

 COVARIATE             B        Beta   Std. Err.
t-Value   Sig. of t

 MATH            .43290      .42787        .070
6.203        .000
 SCIENCE         .28775      .30057        .066
4.358        .000

 COVARIATE   Lower -95%  CL- Upper

 MATH              .295        .571
 SCIENCE           .158        .418

 - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
- - - - - - - - -

* * * * * * A n a l y s i s   o f   V a r i a n c e --
design   1 * * * * * *

 EFFECT .. CONSTANT
 Multivariate Tests of Significance (S = 1, M = 0, N = 97
)
```

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

 Pillais           .11544   12.78959       2.00     196.00
.000
 Hotellings        .13051   12.78959       2.00     196.00
.000
 Wilks             .88456   12.78959       2.00     196.00
.000
 Roys              .11544
 Note.. F statistics are exact.

 - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
- - - - - - - - -
 EFFECT .. CONSTANT (Cont.)
 Univariate F-tests with (1,197) D. F.

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

 READ        336.96220 10176.0807  336.96220   51.65523
6.52329       .011
 WRITE      1209.88188 10090.0231 1209.88188   51.21839
23.62202       .000

 - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
- - - - - - - - -
 EFFECT .. CONSTANT (Cont.)
 Raw discriminant function coefficients
```

```
          Function No.

 Variable            1

 READ             .041
 WRITE            .124

 - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
- - - - - - - -
 Standardized discriminant function coefficients
          Function No.

 Variable            1

 READ             .293
 WRITE            .889

 - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
- - - - - - - -
 Estimates of effects for canonical variables
          Canonical Variable

  Parameter          1

       1         2.196

* * * * * * A n a l y s i s   o f   V a r i a n c e --
design   1 * * * * * *

 EFFECT .. CONSTANT (Cont.)
```

```
Correlations between DEPENDENT and canonical variables
          Canonical Variable

Variable            1

READ             .504
WRITE            .959

- - - - - - - - - - - - - - - - - - - - - - - - - - - - -
- - - - - - - -
```

The output above shows the linear combinations corresponding to the first canonical correlation. At the bottom of the output are the two canonical correlations. These results indicate that the first canonical correlation is .7728. The F-test in this output tests the hypothesis that the first canonical correlation is equal to zero. Clearly, F = 56.4706 is statistically significant. However, the second canonical correlation of .0235 is not statistically significantly different from zero (F = 0.1087, p = 0.7420).

Factor analysis

Factor analysis is a form of exploratory multivariate analysis that is used to either reduce the number of variables in a model or to detect relationships among variables. All variables involved in the factor analysis need to be interval and are assumed to be normally distributed. The goal of the analysis is to try to identify factors which underlie the variables. There may be fewer factors than variables, but there may not be more factors than variables. For our example using the hsb2 data file, let's suppose that we think that there are some common factors underlying the various test scores. We will include subcommands for varimax rotation and a plot of the eigenvalues. We will use a principal components extraction and will retain two factors. (Using these options will make our results compatible with those from SAS and Stata and are not necessarily the options that you will want to use.)

```
factor
  /variables read write math science socst
  /criteria factors(2)
  /extraction pc
  /rotation varimax
  /plot eigen.
```

Communalities

	Initial	Extraction
reading score	1.000	.736
writing score	1.000	.704
math score	1.000	.750
science score	1.000	.849
social studies score	1.000	.900

Extraction Method: Principal Component Analysis.

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.381	67.616	67.616	3.381	67.616	67.616	2.113	42.267	42.267
2	.557	11.148	78.764	.557	11.148	78.764	1.825	36.497	78.764
3	.407	8.136	86.900						
4	.356	7.123	94.023						
5	.299	5.977	100.000						

Extraction Method: Principal Component Analysis.

Component Matrix[a]

	Component

	Component	
	1	2
reading score	.858	-2.04E-02
writing score	.824	.155
math score	.844	-.195
science score	.801	-.456
social studies score	.783	.536

Extraction Method: Principal Component Analysis.

a. 2 components extracted.

Rotated Component Matrix[a]

	Component	
	1	2
reading score	.650	.559
writing score	.508	.667
math score	.757	.421
science score	.900	.198
social studies score	.222	.922

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 3 iterations.

Component Transformation Matrix

Component	1	2
1	.742	.670
2	-.670	.742

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

Communality (which is the opposite of uniqueness) is the proportion of variance of the variable (i.e., **read**) that is accounted for by all of the factors taken together, and a very low communality can indicate that a variable may not belong with any of the factors. The scree plot may be useful in determining how many factors to retain. From the component matrix table, we can see that all five of the test scores load onto the first factor, while all five tend to load not so heavily on the second factor. The purpose of rotating the factors is to get the variables to load either very high or very low on each factor. In this example, because all of the variables loaded onto factor 1 and not on factor 2, the rotation did not aid in the interpretation. Instead, it made the results even more difficult to interpret.

See also

- SPSS FAQ: What does Cronbach’s alpha mean?