Basic Analyses Overview

Data Set Overview

- Choosing the Correct Statistical Test -> http://www.ats.ucla.edu/stat/mult_pkg/whatstat/choosestat.html
- Overview of Data Analyses -> http://www.ats.ucla.edu/stat/spss/whatstat/whatstat.htm#hsb
- Data Set → Hsb2 (click to link to data set from above site)

Analyses Overview

- Correlation
- One-Sample T-Test
- Two Independent Samples T-Test
- Chi-Square (two examples provided)
- One-Way Anova

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**.

Example Output

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

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.

Example Output

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.

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.

Example T-Test Output

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

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).

Chi-Square

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.

Example Output #1

type of school * FEMALE Crosstabulation

Count

		FEMALE		
		male	female	Total
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).

Chi-Square Example 2

- 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).

Example Output #2

FEMALE * SES Crosstabulation

Count

		SES			
		low	middle	high	Total
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)

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**).

Example Output

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**.)

Continued

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.