dummies
A Wiley Brand

CHEAT SHEET

SPSS STATISTICS FOR DUMMIES CHEAT SHEET

From **SPSS Statistics for Dummies, 3rd Edition**

By **Keith McCormick, Jesus Salcedo, Aaron Poh**

IBM SPSS Statistics is an application that performs statistical analysis on data. In order to perform statistical analyses correctly, you need to know the level of measurement of the variables because it defines which summary statistics and graphs should be used. It also helps to know the most commonly used procedures within the Analyze menu and possible conclusions that you can reach after conducting a statistical test.

SPSS STATISTICS VARIABLES LEVEL OF MEASUREMENT

In SPSS Statistics, the level of measurement of the variables defines which summary statistics and graphs should be used. The following table provides definitions, examples, appropriate summary statistics, and graphs for the level of measurement of the variables.

	Nominal	Ordinal	Scale
Definition	Unordered categories	Ordered categories	Both interval and ratio
Examples	Gender, geographic location, job category	Satisfaction ratings, income groups, ranking of preferences	Number of purchases, cholesterol level, age

Measures of Central Tendency	Mode	Median	Median or mean
Measures of Dispersion	None	Min/max/range	Min/max/range, Standard deviation/ variance
Graph	Pie or bar	Bar	Histogram

SPSS STATISTICS CHARTS TO SHOW RELATIONSHIPS BETWEEN A PAIR OF VARIABLES

When choosing a graph to show the relationship between variables, you need to know the level of measurement of the variables. The following table shows some of the graphs that can be used to display relationships between different types of variables.

	Categorical Dependent	**Scale Dependent**
Categorical Independent	Clustered bar or paneled pie	Error bar or boxplot
Scale Independent	Error bar or boxplot	Scatter plot

SPSS STATISTICS COMMONLY USED ANALYZE MENUS

The following table provides a list of some of the most commonly used procedures within the Analyze menu of IBM SPSS Statistics, which is an application that performs statistical analysis on data.

Submenu	**Useful For . . .**

Code Book	Reports	A quick look at all your variables at once. Level of measurement automatically controls which summary statistics are displayed.
Frequencies	Descriptives	Most useful for categorical variables. You can run all of them at once. Tells you how many of each category value you have.
Descriptives	Descriptives	Easy way to get basic scale variable info like mean and median.
Explore	Descriptives	Based on a famous book, *Exploratory Data Analysis.* An effective way to look at all kinds of variables, as well as pairs of variables.
Crosstabs	Descriptives	A test to check to see if categorical variables are independent of each other or related to each other.
Means	Compare Means	Calculates subgroup means and related statistics for dependent variables within categories of one or more independent variables.
One-Sample T-Test	Compare Means	Tests whether the mean of a single variable differs from a specified value (for example a group using a new learning method compared to the school average).
Independent Samples T-Test	Compare Means	Tests whether the means for two groups differ on a continuous dependent variable (for example, females versus males on income).
Paired Samples T-Test	Compare Means	Tests whether there is a significant difference in the mean under two conditions (for example, before versus after, or standing versus sitting).
One way ANOVA	Compare Means	Tests whether the means for two or more groups differ on a continuous dependent variable (for example, drug1 versus drug2 versus drug3 on depression).
Bivariate Correlation	Correlate	Correlations determine the similarity or difference in the way two continuous variables change in value from one case (row) to another through the data.
Linear Regression	Regression	A statistical technique that is used to predict a continuous dependent variable from one or more continuous independent variables.

INTERPRETING STATISTICAL SIGNIFICANCE IN SPSS STATISTICS

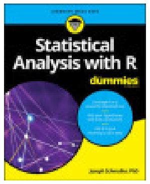
You need to know how to interpret the statistical significance when working with SPSS Statistics. When conducting a statistical test, too often people immediately jump to the conclusion that a finding “is statistically significant” or “is not statistically significant.” While that is literally true, it does not imply that there are only two conclusions to draw about a finding.

What if in the real world there is no relationship between the variables, and the test found that there was a significant relationship? In this case, you would be making an error; this type of error is called a “false positive” because you falsely conclude a positive result (think it does occur).

On the other hand, what if in the real world there is a relationship between the variables, and the test found that there was no significant relationship? In this case, you would be making an error; this type of error is called a “false negative” because you falsely conclude a negative result (think it does not occur).

In the Real World	**Statistical Test Results**	
	Not Significant (*p* > 0.5)	**Significant (*p* < 0.5)**
The two groups are not different	The null hypothesis appears true, so you conclude the groups are not significantly different.	False positive.
The two groups are different	False negative.	The null hypothesis appears false, so you conclude that the groups are significantly different.

TESTING A VARIANCE IN R



RELATED BOOK

Statistical Analysis with R For Dummies

By **Joseph Schmuller**

You might think that the function `chisq.test()` would be the best way to test a variance in R. Although base R provides this function, it's not appropriate here. Statisticians use this function to test other kinds of hypotheses.

Instead, turn to a function called `varTest`, which is in the `EnvStats` package. On the Packages tab, click Install. Then type **EnvStats** into the Install Packages dialog box and click Install. When EnvStats appears on the Packages tab, select its check box.

Before you use the test, you create a vector to hold the ten measurements:

```
FarKlempt.data2 <- c(12.43, 11.71, 14.41, 11.05, 9.53, 11.66, 9.33,11.71,14.35,13.81)
```

And now, the test:

```
varTest(FarKlempt.data2,alternative="greater",conf.level = 0.95,sigma.squared = 2.25)
```

The first argument is the data vector. The second specifies the alternative hypothesis that the true variance is greater than the hypothesized variance, the third gives the confidence level (1 – α), and the fourth is the hypothesized variance.

Running that line of code produces these results:

```
Results of Hypothesis Test


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


Null Hypothesis: variance = 2.25
```

```
Alternative Hypothesis: True variance is greater than 2.25

Test Name: Chi-Squared Test on Variance

Estimated Parameter(s): variance = 3.245299

Data: FarKlempt.data2

Test Statistic: Chi-Squared = 12.9812

Test Statistic Parameter: df = 9

P-value: 0.163459

95% Confidence Interval: LCL = 1.726327

UCL = Inf
```

Among other statistics, the output shows the chi-square (12.9812) and the p-value (0.163459). (The chi-square value in the previous section is a bit lower because of rounding.) The p-value is greater than .05. Therefore, you cannot reject the null hypothesis.

How high would chi-square (with df = 9) have to be in order to reject? Hmmm. . . .