

Bowerman, O'Connell, Aitken Schermer, & Adcock, Business Statistics in Practice, Canadian edition

## **Online Learning Centre**

### **Technology Step-by-Step - SPSS**

SPSS is a statistical software package originally developed to analyze social science research data.

"SPSS Inc. is the world's leading provider of predictive analytics software and solutions. The company's predictive analytics technology connects data to effective action by drawing reliable conclusions about current conditions and future events."

The current version is SPSS 16.0 and is available for Microsoft® Windows XP (32-bit versions) or Vista™, Apple© Mac OS X 10.4 (Tiger™) or 10.5 (Leopard™) and Linux® operating systems.

Procedures explained utilize SPSS 14. The author's operating system is Windows XP Professional

The described SPSS procedures will follow topics in the textbook. *Note:* Procedures shown are representative, not comprehensive. Refer to the SPSS user guide for a more in-depth description of statistical methods.

SPSS, Inc. website is [www.spss.com](http://www.spss.com).

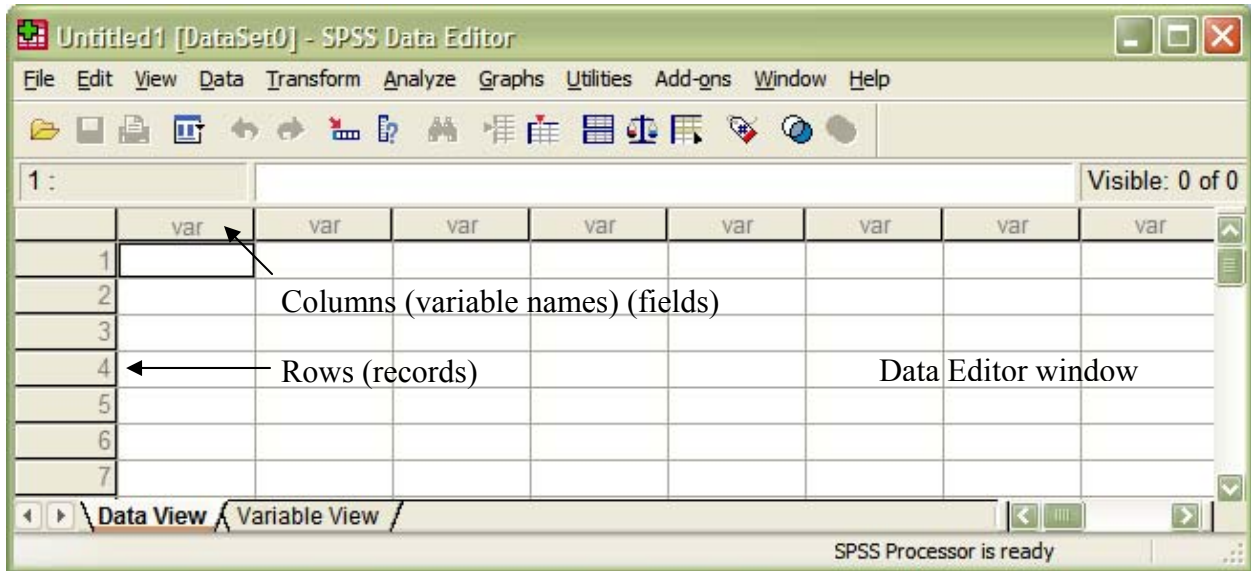
---

## SPSS

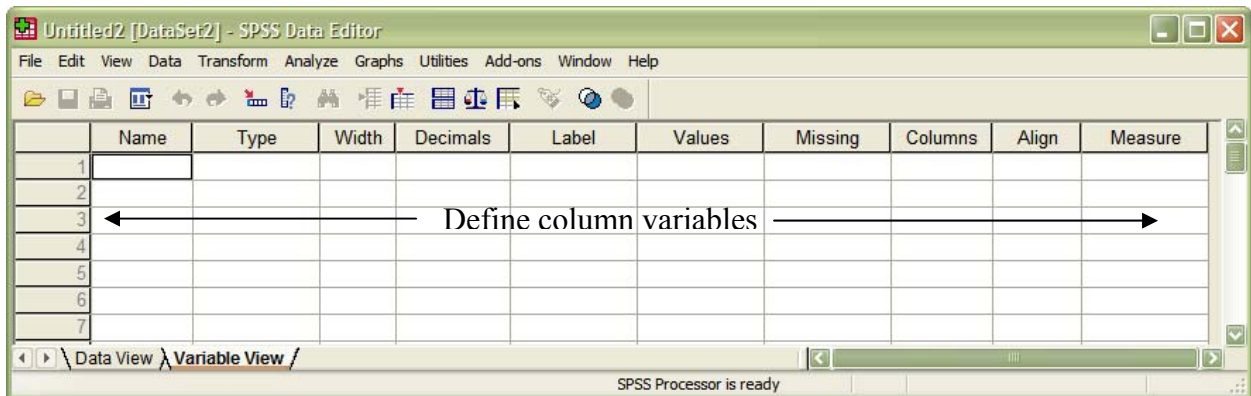
### Starting SPSS

In Windows XP, click Start > Programs > SPSS for Windows > SPSS for Windows

### Opening Screen (Data View)



### Variable View



### Entering Data

In Data View, input data values in columns under variable names (defined fields).

In Variable View, name variables (column headings, or fields), define variables including data type (i.e. string, date, or numeric), measurement level (nominal, ordinal, or scale) and value (descriptive) labels.

The following procedures assume that data variable columns (fields) have been defined and data has been input in data view in rows (records).

## Save a Data File

Click  or File > Save to desired folder. (Default filename extension is .SAV)

## Open a Data File

Click  or File > Open > Open Data... from saved file folder.

## SPSS Commands

Menus or Toolbar buttons may be used to access available commands depending on the desired procedure. SPSS commands may also be entered and run manually by selecting File > New > Syntax to open the Sysntax window. Procedures shown only use menu commands. Results of commands are displayed in the Output – SPSS Viewer window.

### Categorical Frequency Table (Qualitative or Discrete Variable)

Select Analyze > Descriptive Statistics > Frequencies. In Descriptives, select appropriate (qualitative or discrete data) label or name in left source list, click ► to move variable to right Variable(s) list. Click OK to display output in SPSS Viewer including Frequency, Percent, and Cumulative Percent.

### Grouped Frequency Table (Quantitative Variable)

(Manual Coding) In Variable View, define a column variable as numeric and under Values, define classes (i.e. Value: 1, 2 and 3, Labels: Under 20, 20-50, Over 50, respectively). Enter variable name (i.e. AgeGroup) and appropriate label In Data View, code each data value (i.e. 0, 1, or 2) that fits defined variable (i.e. Under 20, 20-50, Over 50) under column label (i.e. AgeGroup). *Note:* Click View > Value Labels to display labels instead of value code in Data View.

(Automatic Coding) Select Transform > Visual Binning... In Visual Binning, select appropriate (quantitative data) variable label or name in left source list, click ► to move variable to right Variables to bin list. Click Continue. Under Scanned Variable List: Variable, select variable name. With Current variable name displayed, type desired new variable name (i.e. AgeGroup) in Binned variable box. Select Make Cutpoints. Under Equal Width Intervals | First Cutpoint Location, type upper class limit of lowest bin. Under Number of Cutpoints, type desired number of classes (or bins). Under width, type desired class (bin) width. Click Apply. Click Make labels to automatically generate labels. Click OK, then OK at prompt. Output will be displayed in SPSS Viewer. Note value labels (i.e. 1,2, 3..) beside classes (bins). In Data View, a new column will be created with coded value labels for each class (bin).

To display frequency table, select Analyze > Descriptive Statistics > Frequencies. In Descriptives, select variable label or name containing class (bin) coding in left source list, click ► to move variable to right Variable(s) list. Click OK to display output in SPSS Viewer including classes (bins) with frequency, percent, and cumulative percent.

## Histograms and Other Charts

To display histogram, select Analyze > Descriptive Statistics > Frequencies. In Descriptives, select variable label or name in left source list, click ► to move variable to right Variable(s) list. Click Charts... > select Histograms (or desired chart). Click Continue > Click OK to display chart output in SPSS Viewer.

Alternatively, select Graphs > Chart Builder (read pop-up window) > click Gallery > select Histogram (or desired chart) in Choose From: list > drag Simple Histogram (or other chart subtype) icon onto canvas > drag desired variables to the x-axis and y-axis drop zones. Click OK to display chart in Output: SPSS Viewer.

## **Descriptive Measures**

Define column variable and input data. Select Analyze > Descriptive Statistics > Descriptives. In Descriptives, select variable label or name in left source list, click ► to move variable to right Variable(s) list. Click Options and select desired descriptive measures. Click Continue > OK to display output in SPSS Viewer.

## **Measures of Position**

Select Analyze > Descriptive Statistics > Descriptives... > In Descriptives, select one or more variable labels or names in left source list, click ► to move variable to right Variable(s) list. > check Save standardized values as variables. > click OK. In Data View, new column(s) or fields are created containing z-scores of selected variable labels with labels prefaced with a z. (i.e. Age → zAge).

## **Relative Frequency (Percentage) Probabilities**

Define column variables and input data. Select Transform > Compute variable... In dialogue box, under Target Variable: type new variable name (i.e. RelFreq). In source list, click ► to move frequency variable to Numeric Expression: box. Type /n (where n is the sum of the frequency column). Click OK to display results in new variable column.

Another method is to repeat the procedure for Categorical Frequency Table using discrete data. In SPSS Viewer, Percent column represents relative frequency probabilities.

## **Contingency Table**

Enter data in two columns and define column variables. Select Analyze > Descriptive Statistics > Crosstabs.... In Crosstabs, select variable label or name in left source list, click ► to move variable to Row(s):. Repeat to move 2<sup>nd</sup> variable from source list to Column(s):. Optional: Click Cells... and select Percentages | Row, Column and Total. Click Continue. Click OK to display output in SPSS Viewer.

## **Factorials, Permutations, and Combinations**

SPSS does not have built-in functions for Factorials, Permutations, and Combinations.

## **Binomial Probability Distribution**

Define column variables and input data for X (i.e. 0, 1, 2...n). Select Transform > Compute variable... In dialogue box, under Target Variable: type new variable name (i.e. ProbX). In Numeric Expression, under Function group: | PDF & Noncentral PDF, select under Functions and Special Variables | Pdf.Binom,. Click ▲ to move function to Numeric Expression: box. Replace PDF.BINOM(?,?,?) with values for PDF.BINOM(X,N,p), (where X is column variable name for X-values, N is number of trials, p is probability of success for binomial). Click OK to display binomial probabilities in new variable column.

If cumulative probabilities are desired, use new variable name (i.e. CumProbX) and select CDF & Noncentral CDF, select under Functions and Special Variables | Cdf.Binom using same format as Pdf.Binom.

### **Graph Binomial Distribution**

Create a binomial probability distribution in Data View. Select Graphs > Chart Builder (read, then click OK for displayed dialogue) > Under Choose from: select Histogram | drag Simple Histogram in into chart preview window. From Variable list: drag variable name (i.e. X) to X-Axis? and drag variable name (i.e. ProbX) to Y-Axis? in preview window. Click OK to display chart in Output | SPSS Viewer.

To create a line chart, select Line | Simple type and use either variable for probability of X or cumulative probability of X as desired.

### **Poisson Probability Distribution**

Define column variables and input data for X (i.e. 0, 1, 2...n). Select Transform > Compute variable... In dialogue box, under Target Variable: type new variable name (i.e. ProbX). In Numeric Expression, under Function group: | PDF & Noncentral PDF, select under Functions and Special Variables | Pdf.Poisson. Click ▲ to move function to Numeric Expression: box. Replace PDF.POISSION(?,?) with values for PDF.POISSION(X, mean), (where X is column variable name for X-values, and mean is the distribution mean). Click OK to display poisson probabilities in new variable column.

If cumulative probabilities are desired, use new variable name (i.e. CumProbX) and select CDF & Noncentral CDF, select under Functions and Special Variables | Cdf.Poisson using same format as Pdf.Poisson.

### **Standard Normal Distribution**

Refer to Measures of Position for z-score conversions of numeric variables.

For cumulative probability distributions, define column variables and input data (X values). Select Transform > Compute variable... In dialogue box, under Target Variable: type new variable name (i.e. AreaLEFT). In Numeric Expression, under Function group: | PDF & Noncentral PDF, select under Functions and Special Variables | Cdf.Normal,. Click ▲ to move function to Numeric Expression: box. Replace CDF.NORMAL(?,?,?) with values for CDF.NORMAL(X, mean, standard deviation), (where X is column variable name for X-values, mean of X, standard deviation of X). Click OK to display probabilities, cumulative areas, in new variable column.

*Note:* If column variables represent z scores, use mean = 0 and standard deviation = 1 in expression. Alternatively, the CDFNORM function assumes mean = 0 and standard deviation = 1 and returns same value for area to the left.

*Note:* To calculate area to the right of value, use 1- Cdf.Normal(value, mean, standard deviation) in Expression: box. To calculate area between, input in Expression: box, Cdf.Normal(rightmost value, mean, standard deviation) – Cdf.Normal(leftmost value, mean, standard deviation).

### **Determining Normality**

Graphical Method → define column variables and input data. Select Analyze > Descriptive Statistics > Frequencies. Select desired variable name. Click Charts... > select Histograms | check With normal curve. Click Continue > Click OK to display chart output in SPSS Viewer. \*Check for a bell-shaped curve on histogram.

Normal Probability Plot Method → define column variables and input data. Select Analyze > Descriptive Statistics > Explore. Select desired variable name. Under Display: check Both or Plots (as desired). Click Plots... > under Descriptive: check Histogram. Check Normality plots with tests. Click Continue > click OK to display chart output in SPSS Viewer. \*Check if histogram is bell-shaped, check Normal Q-Q Plot for linearity, check Detrended Q-Q Plot for equal dispersion of data points above and below line. For Boxplot, check for symmetry and outliers.

### Confidence Interval Estimate

Normal  $z$  Interval (means) → define column variables and input data. Select Analyze > Descriptive Statistics > Explore... From source list, select desired variable name and move to Dependent List: box. Under Display, select Both or Statistics (as desired). Click Statistics... > Descriptives, | Confidence Interval for Mean: {type desired level of confidence}. Click Continue > OK to display Output in SPSS Viewer including confidence interval for mean (lower and upper bound).

$t$  Interval (means) → define column variables and input data. Select Analyze > Compare Means > One-sample T test... From source list, select desired variable name and move to Test Variable(s): box. Click Options... and input desired Confidence Interval:. Click Continue > OK to display Output in SPSS Viewer including confidence interval of the difference (lower and upper).

Normal  $z$  Interval (proportions) → SPSS does not have a menu procedure for confidence interval estimates for proportion, although there is a procedure for two variable ratios. Command syntax may be used to manually calculate the lower and upper confidence limits.

*Workaround:* Define 7 variables with appropriate names and decimals (i.e.  $n$ ,  $x$ ,  $\text{phat}$ ,  $\text{sterr}$ ,  $z$ , lower, upper). Input data (i.e.  $n$ ,  $x$ ). Select Transform > Compute variable... Input appropriate Target variable: name and use source list variable and Function group | Functions and Special Variables: to create numeric expressions. Click OK twice after creating each expression, then repeat for each variable.

$\text{phat} = x / n$

$z = \text{Inverse DF} \mid \text{Idf.Normal}(\text{prob}, \text{mean}, \text{stdev})$ , where *prob* is cumulative area to left of upper  $z$ -score,  $\text{mean} = 0$ ,  $\text{stdev} = 1$ .

$\text{sterr} = z * \text{Arithmetic} \mid \text{Sqrt}(\text{phat} * (1 - \text{phat}) / n)$

lower =  $\text{phat} - \text{sterr}$

upper =  $\text{phat} + \text{sterr}$

Variance → SPSS does not have a menu procedure confidence interval estimate for variances and standard deviations.

### Hypothesis Tests

$t$  Test (means) → define column variables and input data. Select Analyze > Compare Means > One-sample T test... From source list, select desired variable name and move to Test Variable(s): box. In Test Value: field, type hypothesized value for comparison. . Click Options... and input desired Confidence Interval:. *Note:* Confidence level =  $1 - \text{significance level}$ . Click Continue > OK to display Output in SPSS Viewer including Sig (2-tailed) or  $P$ -value and  $t$ -value under T-test | One-Sample Test.

## Hypothesis Test – Difference Between Two Means

*t* Test (independent samples) → define column variables and input data. Select Analyze > Compare Means > Independent-Samples T test... From source list, select and move desired variable name to be used as Test Variable(s): and Grouping Variable: respectively. For numeric grouping variables, click Define Groups... and input values to define the groups. Click Options... input desired Confidence Interval:. *Note:* Confidence level = 1 – significance level. Click Continue > OK to display Output in SPSS Viewer including Sig (2-tailed) or *P*-value and *t*-value under T-test | Independent-Samples Test.

*t* Test (dependent samples) → define column variables and input data. Select Analyze > Compare Means > Paired-sample T test... From source list, select two desired variable names so that both variables appear under Current Selections and move to Paired Variables: window. Click Options... input desired value for Confidence Interval:. *Note:* Confidence level = 1 – significance level. Click Continue > OK to display Output in SPSS Viewer including Sig (2-tailed) or *P*-value and *t*-value under T-test | Paired Samples Test.

## Hypothesis Test – Difference Between Two Variances

Refer to output in test Between Two Means.

## Hypothesis Test - Correlation and Regression

Scatter Plot → In data editor, define two variables (i.e. *x*, *y*) and input paired data with headings. Select Graphs > Chart Builder (read pop-up window) > click Gallery > select Scatter/Dot in Choose From: list > drag Simple Scatter icon onto canvas > drag required scale variables to the *x*-axis and *y*-axis drop zone on canvas. Click OK to display scatterplot in Output: SPSS Viewer.

Correlation → In data editor, define two variables (i.e. *x*, *y*) and input paired data with headings. Select Analyze > Correlate > Bivariate... Move two variables (*x*, *y*) from source list to Variables: window. Under Correlation Coefficients, check Pearson. Under Test of Significance, check Two-tailed or One-tailed (as required by hypothesis). Click OK to display Output in SPSS Viewer including Sig, (2-tailed or 1-tailed) which is the *P*-value for the test.

Regression Equation → In data editor, define two variables (i.e. *x*, *y*) and input paired data with headings. Select Analyze > Regression > Linear... In Linear Regression dialogue box, move appropriate variable names from source list to Dependent: box and Independent(s): box, respectively. Click OK to display output in SPSS Viewer including Coefficients table, Sig (*P*-value), and values intersecting Understandardized Coefficients and Models: B and Constants (*y*-intercept) | B and *x*<sub>1</sub> (slope).

Multiple Regression → In data editor, define dependent and independent variables and input data with headings. Select Analyze > Regression > Linear... In Linear Regression dialogue box, select preferred variable name from source list to Dependent: field and other variables to Independent(s): field. Click OK to display Output in SPSS Viewer including Sig, (*P*-value) under ANOVA and Coefficients and which is the *P*-value for the test.

## Chi-Square Hypothesis Test

$\chi^2$  Goodness-of-Fit → Input observed frequencies in data editor and define and name variable. Select Analyze > Nonparametric Tests > Chi-Square... In Chi-Square Tests dialogue box, move the desired variable name from the source list to the Test Variable List:. Under Expected Values, select either All categories equal or if categories are difference, and input appropriate Values. Click OK to display Output in SPSS Viewer including observed and expected frequencies,  $\chi^2$  test statistic and Asymp.Sig or *P*-value.

$\chi^2$  Independence Test or Contingency Table Test → Input observed frequencies in data editor and define and name variables. Select Analyze > Descriptive Statistics > Crosstabs... In Crosstabs dialogue box, move one variable name from the source list to the Row(s): window and another variable name to the Column(s): window. Click Statistics... In Crosstabs: Statistics dialogue, select Chi-square > Continue. Click Cells... In Crosstabs: Cell Display, select desired cell display (i.e. Counts: Observed, Expected; Percentages: Row, Column, Total). Click Continue > OK to display Output in SPSS Viewer including Crosstabulation table and Chi-Square Tests.

### **One-Way Analysis of Variance**

Define column variables and input data. Select Analyze > Compare Means > One-Way ANOVA... In dialogue box, select one variable and move to Dependent List: field and select 2<sup>nd</sup> variable and move to Factor: field. Click Post Hoc... in dialogue box, select desired test and significance level. Click Continue. Click Options and select desired Statistics and Plot. Click Continue > OK to display Output results in SPSS Viewer including the Sig (*P*-value) and *F* test value and Means Plot (if selected).

### **Two-Way Analysis of Variance**

Define column variables and input data into 3 columns. Select Analyze > General Linear Model > Univariate ... In Univariate dialogue box, select the response variable and move to Dependent Variable: field and select the row and column factor variables and move to Fixed Factor(s): field. Click OK to display Output results in SPSS Viewer including the Sig (*P* values) and *F* test values under Test of Between-Subjects Effects.

### **Signs Test**

Define column variables and input data. Select Analyze > Nonparametric Tests > 2-Related-Samples... In dialogue box, move two variable names (check both variables are shown under Current Selections) from source list to Test Pair(s) List:. Under Test Type, click Sign. Click OK to display Output in SPSS Viewer including Exact Sig (2-tailed) or *P*-value under Test Statistics.

### **The Paired-Sample T Test**

Define column variables and input data. Select Analyze > Compare Means > Paired Samples T-Test... In source list move appropriate variable names (check both variables are shown under Current Selections) to Paired Variables: field. Select 2<sup>nd</sup> paired variables. Click OK to display Output in SPSS Viewer including Correlation, and Sig., (*P*-value) and *t*-value.

### **Wilcoxon Rank Sum Test (Mann-Whitney)**

Define column variables and input data. Select Analyze > Nonparametric Tests > 2 Independent Samples... In dialogue box, move required variable from source list to Test Variable List:. Under Test Type, select Mann-Whitney U. Under Grouping Variable:, click Define Groups... In dialogue box, input appropriate values in Group 1: and Group 2: . Click Continue > OK to display Output in SPSS Viewer including *Z* and Asymp.Sig (2-tailed) or *P*-value.

### **Wilcoxon Signed-Rank Test**

Define column variables and input data. Select Analyze > Nonparametric Tests > 2-Related-Samples... In dialogue box, move two variable names (check both variables are shown under Current Selections)



from source list to Test Pair(s) List:. Under Test Type, click Wilcoxon. Click OK to display Output in SPSS Viewer including Exact Sig (2-tailed) or *P*-value under Test Statistics.

### **Kruskal-Wallis Test**

Define column variables and input data. Select Analyze > Nonparametric Tests > K Independent Samples. In dialogue box source list move desired variable to Test Variable List: window. Click Define Range... input appropriate values under Range for Grouping Variables: Minimum and Maximum. Click Continue > OK to display Output in SPSS Viewer including test statistic Chi-Square and Asymp.Sig or *P*-value.

### **Runs Test for Randomness**

Define column variables and input data. Select Analyze > Select Nonparametric Test > Runs... In Dialogue Box, select variable name in source list and move to Test Variable List:. Select appropriate Cut Point in dialogue. Click Options... select Statistics | Quartiles. Click Continue > OK to display Output in SPSS Viewer including Asymp.Sig (2-tailed) or *P*-value.