

AMATYC
24th Annual Conference
Portland, Oregon
1998

MINITAB-12
Teaching Statistics With Style

Thursday, November 5, 1998
2:00 - 3:00 p.m.

Brian E. Smith
McGill University
Montreal, Quebec, Canada
Email: smithb@management.mcgill.ca

MINITAB - Release 12

MINITAB is well established as a statistical package that combines ease-of-use with powerful computational algorithms. MINITAB has been adopted as the statistical software of choice in many colleges and universities because of its user-friendliness, and the ability to teach statistics using MINITAB, without having to spend a great deal of time instructing students in the use of the software. The graphic user interface is intuitive and a brief introduction is sufficient to get students started.

MINITAB, Release 12, has added several new features, which further enhance its usefulness as a teaching tool. In this introduction to Release 12 the focus is on those new features which are likely to be of interest to instructors and students in an introductory college course in statistics.

New in Release 12

General

- Documentation helps you analyze your data using MINITAB and interpret your results.
- Dynamic worksheet size

Data and File Management

- Multiple worksheets
- Specify order to process text categories by MINITAB commands
- Create new worksheets easily with new, easy data subsetting
- Font control in data window
- Save graphs, worksheets, and session output in one file

Basic Statistics

- Store descriptive statistics easily
- Paired t-test and t-interval
- Estimation and tests for 1 and 2 proportion problems
- P-values for correlation

Analysis of Variance

- Unbalanced nested design for analyzing a wider range of experimental designs
- Analyze fully nested designs
- Multiple comparisons
- Expected mean squares, approximate F tests

Statistical Process Control

- Use normal or non-normal data with variable control charts
- Historical charts which estimate central limits and center line independently for different groups on the same chart
- Process capability analysis for normal and non-normal data, and attributes data
- Shainin Multi-Vari charts for graphical analysis of variance
- Symmetry plot

Reliability and Survival Analysis

- Parametric distribution analysis using eight common lifetime distributions
- Nonparametric distribution analysis, including Kaplan-Meier, Actuarial, and Turnbull estimates
- Exact failure, right, left, and interval censored data
- Accelerated life testing to perform a simple regression with one predictor
- Regression with life data to investigate the relationship between failure time and one or more predictors
- Probit analysis to estimate a stress distribution

Designed Experiments

- Make and analyze general full factorial designs
- Analyze factorial designs with botched runs
- Multiple response optimization: numerical, interactive graph, and overlaid contour plot
- Easily modify or update existing experimental designs created in MINITAB
- Create user-specified designs

Sample Size and Power

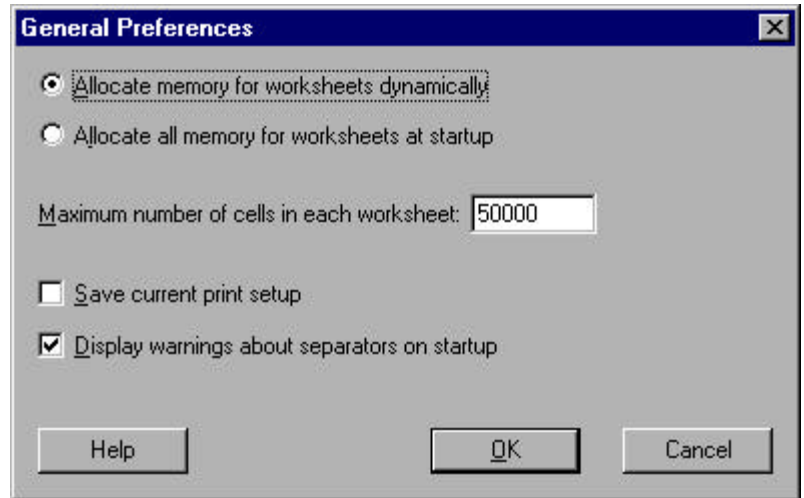
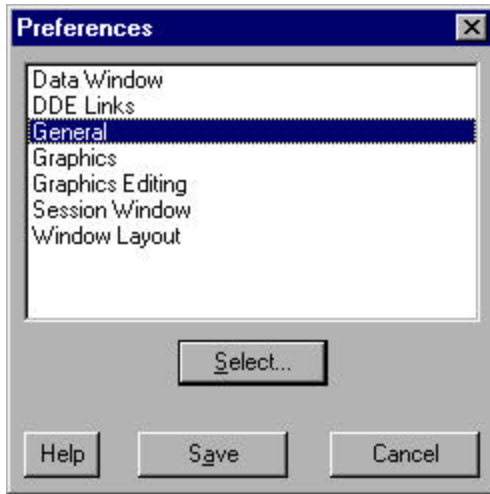
- 1-sample Z, t, and proportion
- 2-sample t and proportion
- One-way ANOVA, 2-level factorial designs, and Plackett-Burman designs

Graphics

- High-resolution dotplot
- OLE: edit MINITAB graphs in other applications
- Save graphs in TIFF, JPEG, PNG, and bitmap file formats

DYNAMIC WORKSHEET SIZE

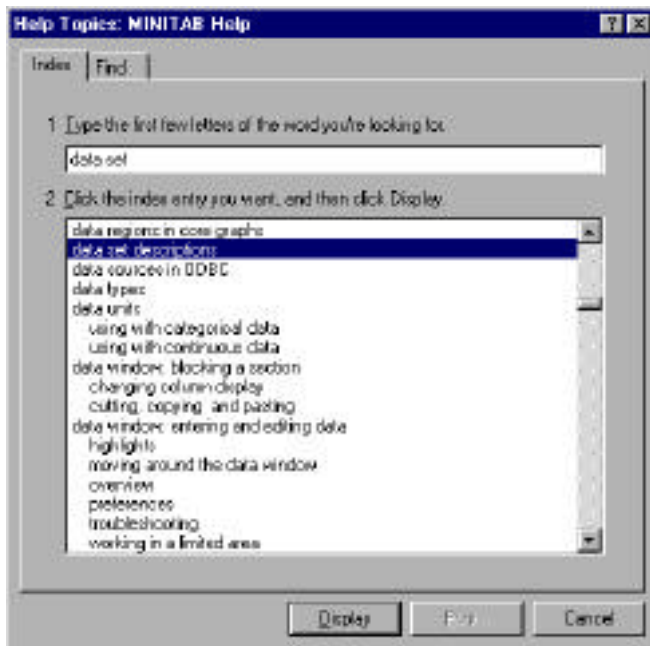
A new feature of MINITAB 12 is the ability to dynamically allocate worksheet size so that less memory is used when working with smaller worksheets. To see the options for memory allocation use **Edit** → **Preferences** → **General** to obtain the dialog boxes shown below:



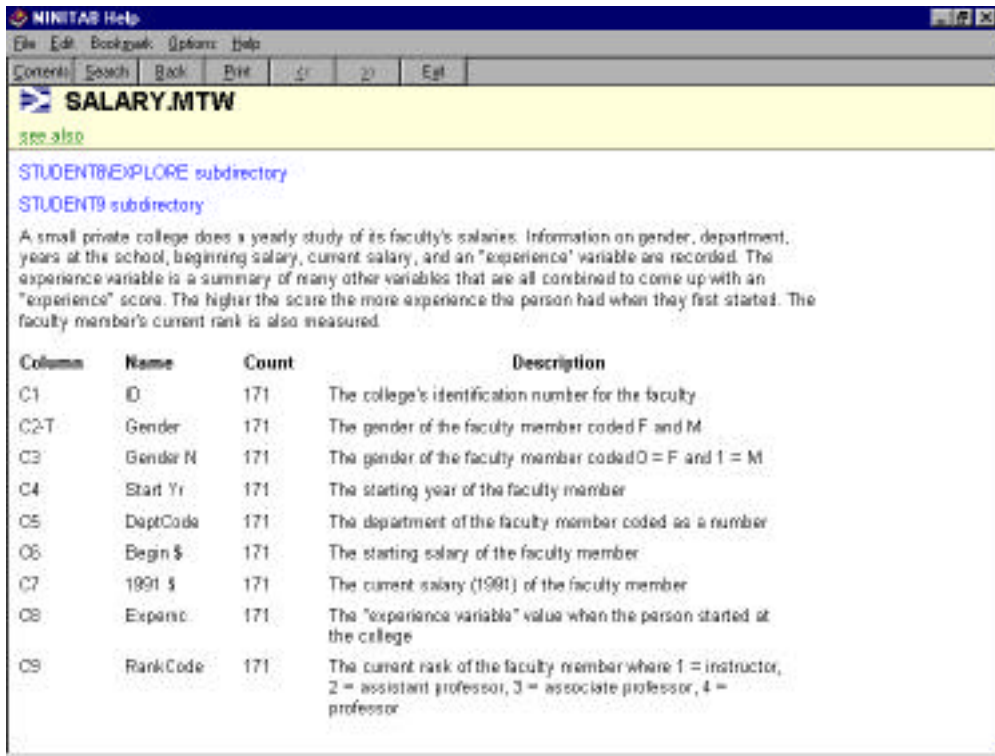
The default is 100,000 cells, but may be smaller if you do not have enough available memory. In the second box above the worksheet size is changed to 50,000 cells.

DATA SET DESCRIPTIONS

MINITAB includes a useful collection of data sets. A helpful new feature is data set descriptions for each data set. Click on **HELP** and then type data set descriptions:



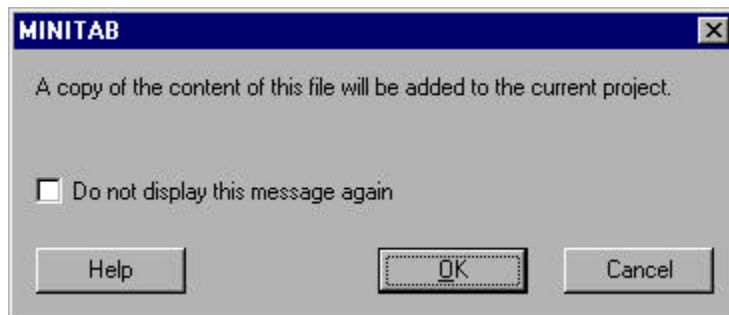
Scrolling down in the second screen above reveals the names of the more than 200 data sets in alphabetical order. For example, the data set SALARY is described below:



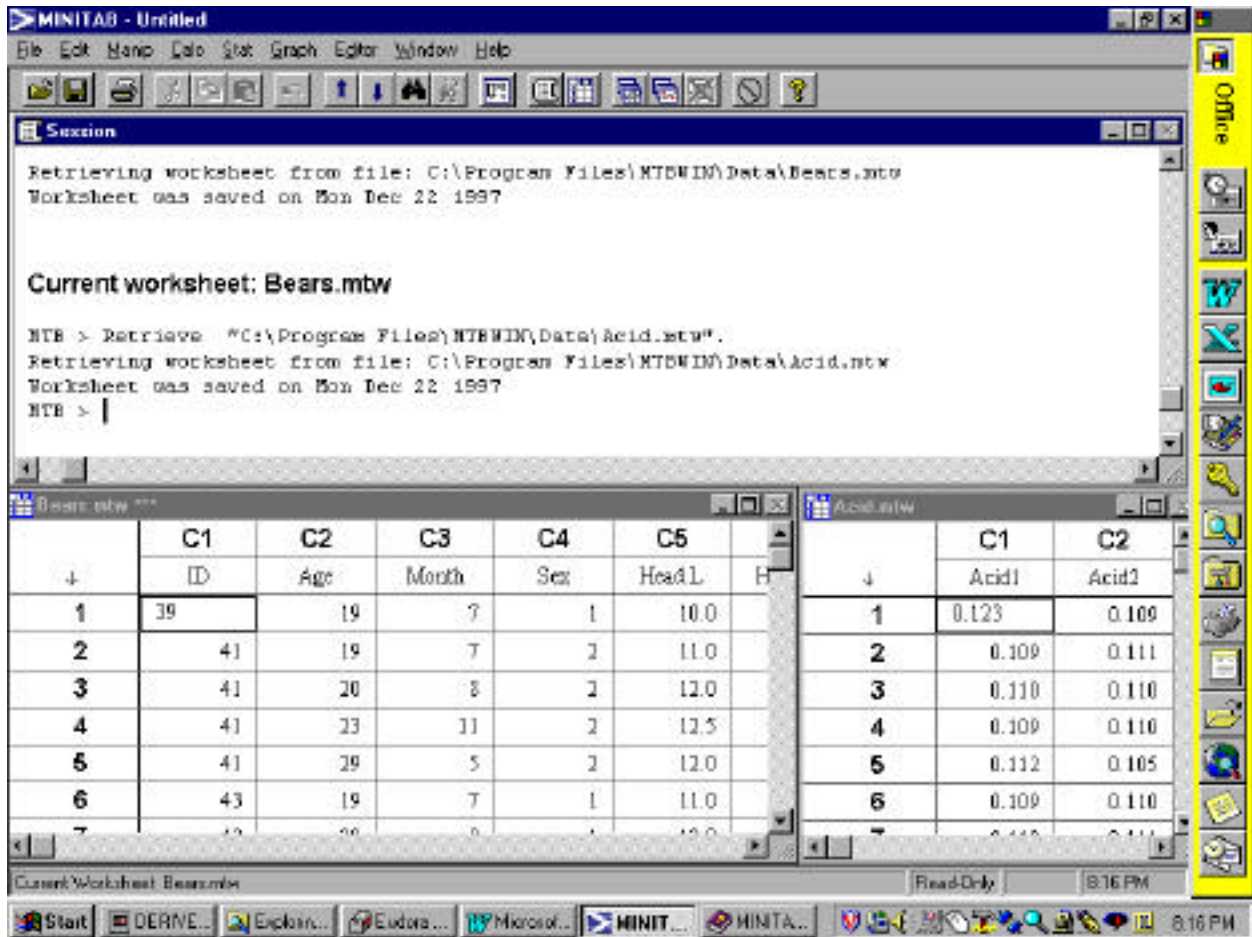
PROJECTS

A MINITAB project includes all of your work – multiple worksheets, text output from commands, graphs etc.

When you open a worksheet the following dialog box appears:

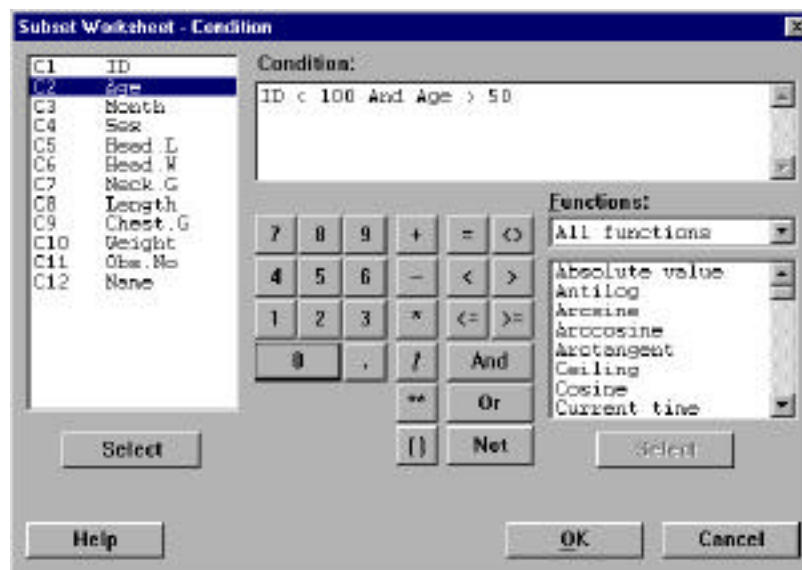


You can add as many worksheets as you wish to a project. Thus, if you want to add the data sets “acid” and “bears” to the same project you can open both worksheets and position them side by side on the screen:



DATA SUBSETTING

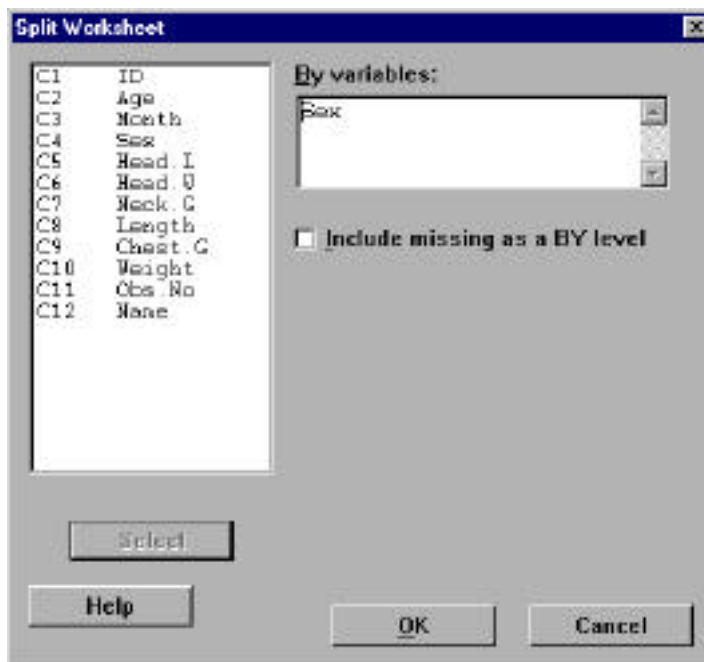
Create a new worksheet by selecting a subset of the data in an existing worksheet. For example, to create a worksheet consisting of a subset of the BEARS data set which includes rows satisfying the condition “ID < 100 and AGE > 50” select **Manip** → **Subset Worksheet** and click on **Condition** to complete the following box:



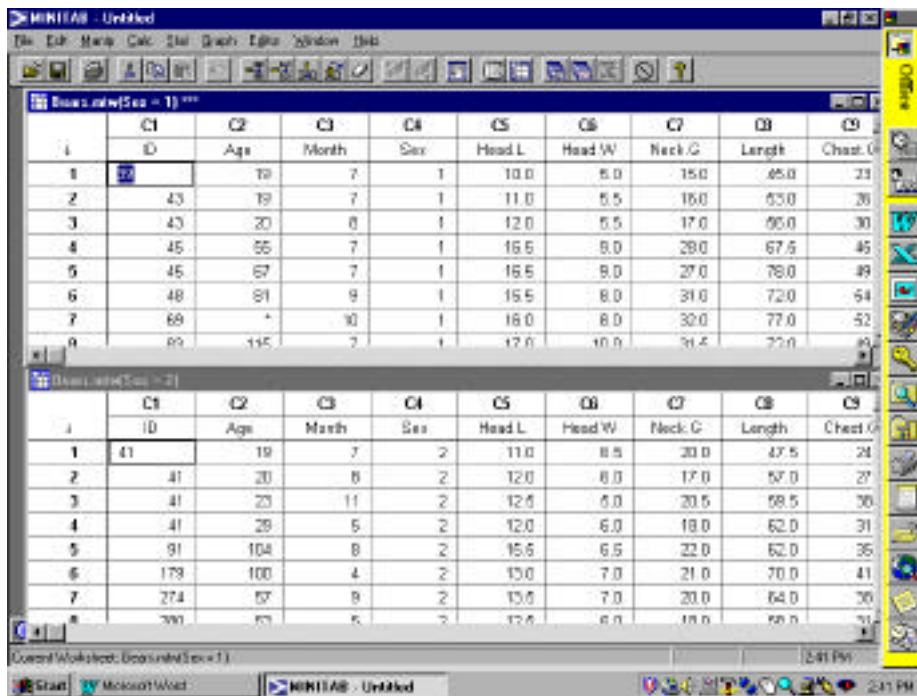
The result is the following worksheet:

	C1	C2	C3	C4	C5	C6	C7	C8	C9
i	ID	Age	Month	Sex	Head L	Head W	Neck G	Length	Chest G
1	45	55	7	1	16.5	9.0	28.0	67.5	45.0
2	45	67	7	1	16.5	9.0	27.0	78.0	49.0
3	48	81	9	1	15.5	8.0	31.0	72.0	54.0
4	83	115	7	1	17.0	10.0	31.5	72.0	49.0
5	83	117	9	1	15.5	7.5	32.0	75.0	54.5
6	83	124	4	1	17.5	8.0	32.0	75.0	55.0
7	83	140	8	1	15.0	9.0	33.0	75.0	49.0
8	91	104	8	2	15.5	6.5	22.0	62.0	35.0
9									
10									
11									
12									
13									
14									
15									
16									
17									
18									
19									
20									

The worksheet may also be split into two or more new worksheets by using the Manip → Split Worksheet command. For example if we wish to split the **BEAR** worksheet into two new worksheets by sex we would complete the following dialog box:



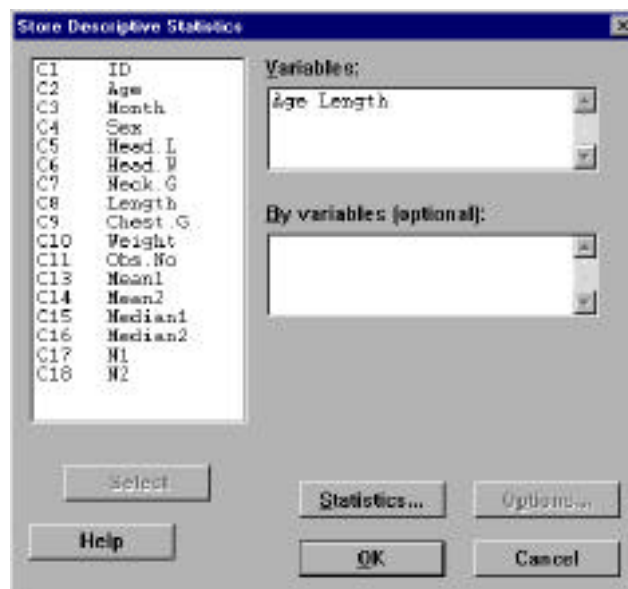
The result is the creation of two new worksheets (while retaining the original) as follows:



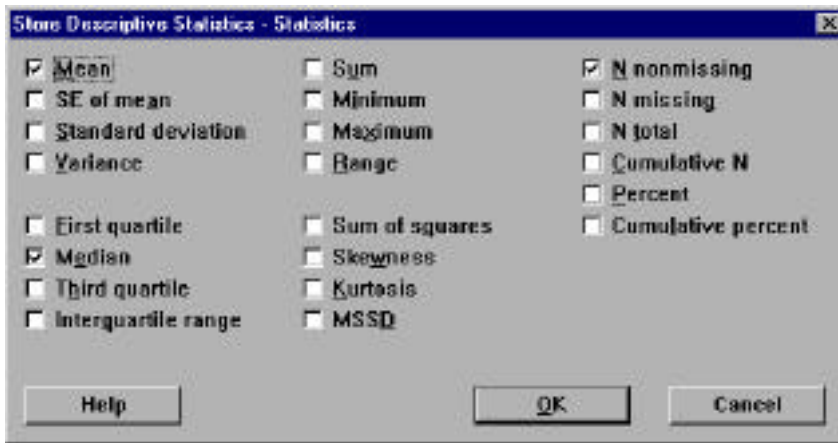
STORING DESCRIPTIVE STATISTICS

You can store descriptive statistics for each column, or for subsets within a column.

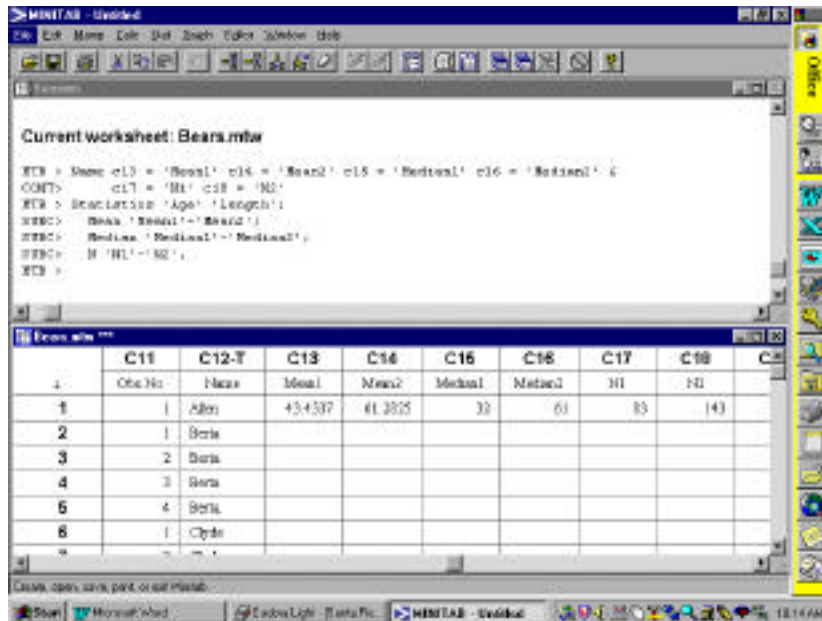
Choose **Stat** → **Basic Statistics** → **Store Descriptive Statistics** and select the variables for which you wish to store basic statistics. In the screen below we have selected the variables Age and Length.



Next click on **Statistics** and select the statistics you wish to store.

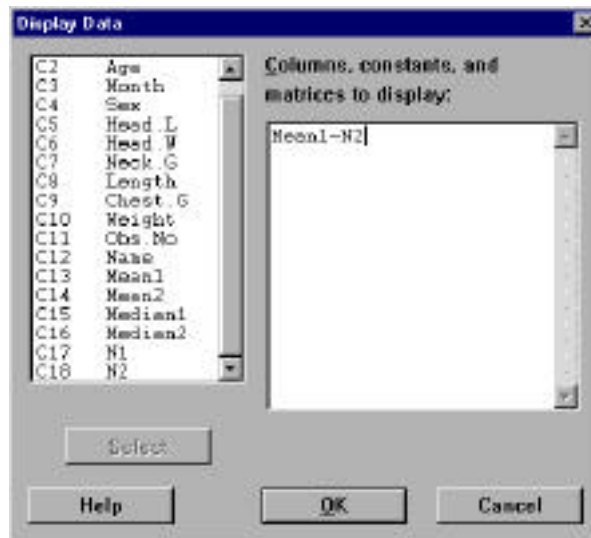


Click on **OK** to return to the previous dialog box and then click on **OK** again. The result is shown in the session window below:



Note that the mean, median and N (sample size) statistics are stored in the first available empty columns (in this case C13-C18) and that MINITAB automatically names the columns Mean1, Mean2, Median1, Median2, N1, and N2.

The stored constants can be displayed in the session window by using **Manip → Display Data**



Click on **OK** to obtain the result:

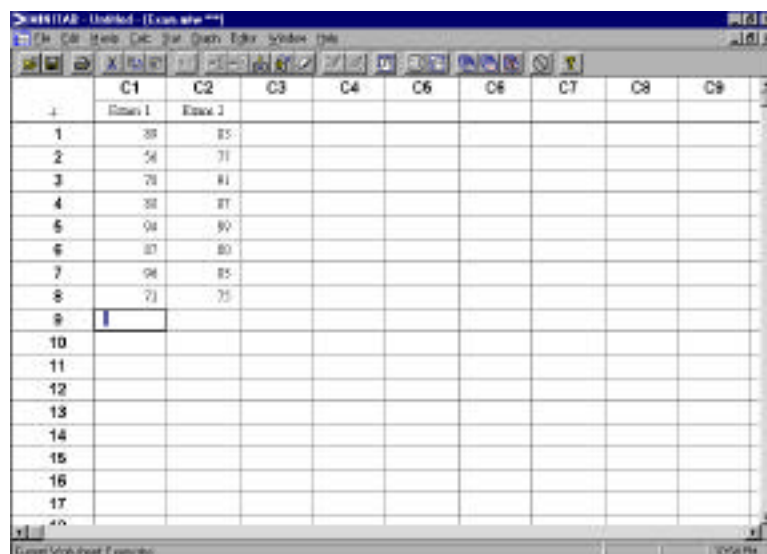
```
MTB > Print 'Mean1'-'N2'.
```

Data Display

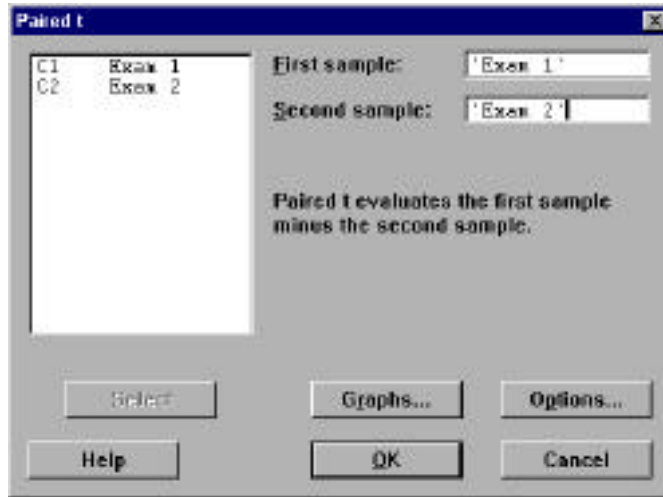
Row	Mean1	Mean2	Median1	Median2	N1	N2
1	43.4337	61.2825	32	61	83	143

PAIRED T-TEST AND T-INTERVAL

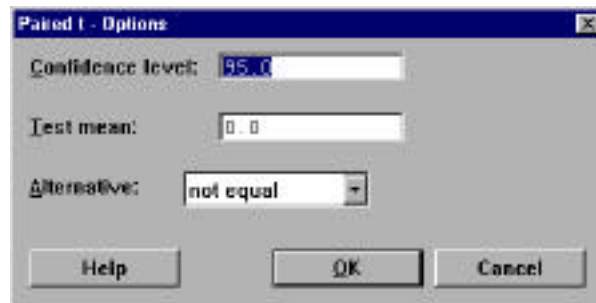
MINITAB 12 supports matched sample (paired) t-tests. Consider the worksheet EXAM.MTW located in the **Student1** directory.



To test the hypothesis that the scores in Exam1 and Exam2 are the same at the 5% level of significance, choose **Stat** → **Basic Statistics** → **Paired t...** and select Exam1 and Exam2 as **First sample** and **Second sample**, respectively.



Next select the **Options** as shown below:



Click on **OK** to obtain the following results in the session window:

Paired T for Exam 1 - Exam 2

	N	Mean	StDev	SE Mean
Exam 1	8	82.50	13.31	4.71
Exam 2	8	84.62	7.82	2.76
Difference	8	-2.12	10.79	3.81

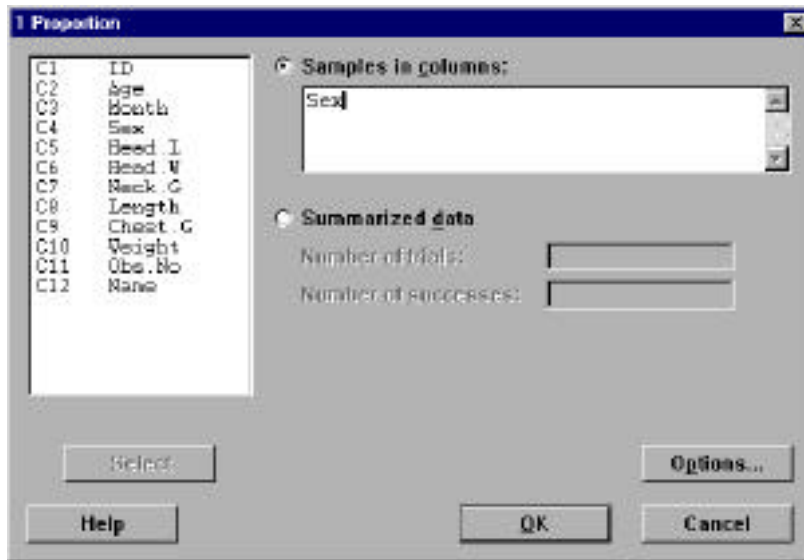
95% CI for mean difference: (-11.15, 6.90)

T-Test of mean difference = 0 (vs not = 0): T-Value = -0.56 P-Value = 0.595

We see that the 95% confidence interval for the mean difference is given by $-11.15 \leq \mu_d \leq 6.90$ and the p-value for the two-sided test of hypothesis is 0.595. We conclude that there is no significant difference in mean test scores for the two examinations, based on the sample of 8 subjects.

ESTIMATION AND TESTS FOR 1 AND 2 PROPORTION PROBLEMS

The commands **Stat** → **Basic Statistics** → **1 Proportion** invoke the following dialog box. The file **BEARS.MTW** is used for the illustration and column C4 (Sex) is selected. This column contains only 1s and 2s, where 1 = male and 2 = Female.



Click on **OK** to obtain the following results in the session window:

Test and Confidence Interval for One Proportion

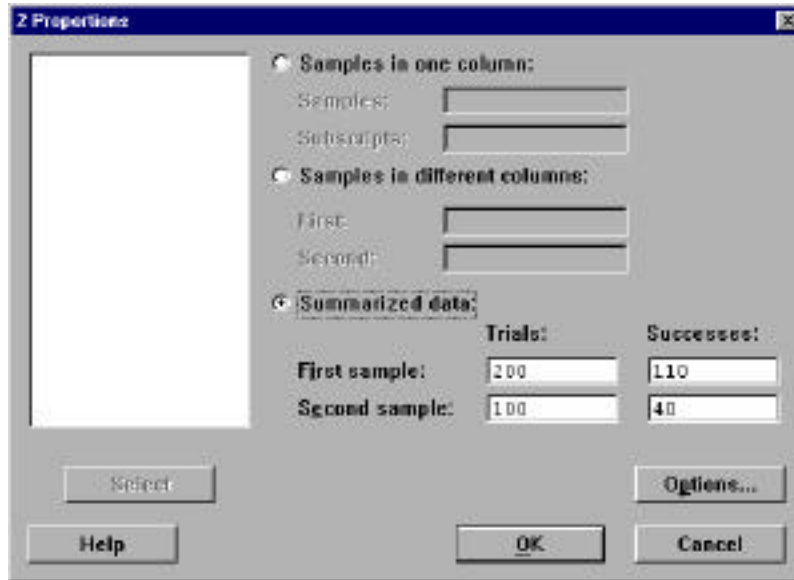
Test of $p = 0.5$ vs $p \text{ not } = 0.5$

Success = 2

Exact					
Variable	X	N	Sample p	95.0 % CI	P-Value
Sex	44	143	0.307692	(0.233280, 0.390269)	0.000

MINITAB automatically defines Success as the higher number, so that in this case Success = 2 (Female). We see that there are 44 female bears in the sample of 143 bears, for a sample proportion of 0.307692. The confidence interval indicates that we are 95% confident that the proportion of female bears lies between 23.33% and 39.03%. The p-value shows us that the population proportion of female bears is significantly different from 0.5 (50%).

Inference concerning two population proportions may be performed by selecting the sequence **Stat** → **Basic Statistics** → **2 Proportions**. In the resulting dialog box below we have chosen to perform a test of hypothesis for the difference between population proportions for summarized data, where the first population has a sample of size 200 with 110 successes and the second population is represented by a sample of size 100 with 40 successes.



The results are as follows:

Test and Confidence Interval for Two Proportions

Sample	X	N	Sample p
1	110	200	0.550000
2	40	100	0.400000

Estimate for $p(1) - p(2)$: 0.15

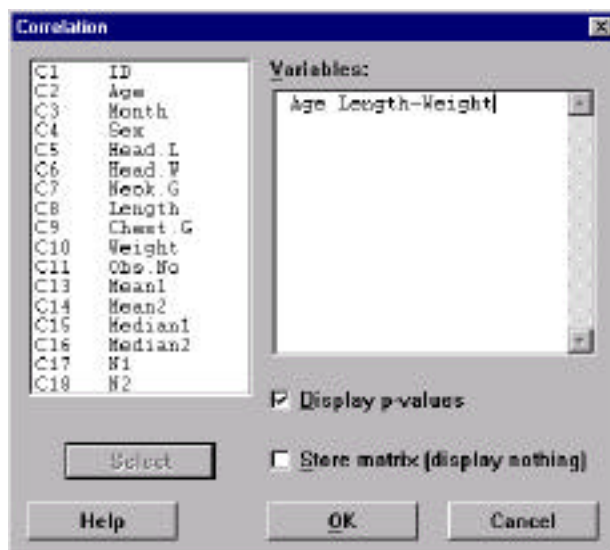
95% CI for $p(1) - p(2)$: (0.0317913, 0.268209)

Test for $p(1)-p(2)=0$ (vs not = 0): $Z = 2.49$ P-Value = 0.013

We see that the test is significant with a p-value of 0.013.

P-VALUES FOR CORRELATION

Select **Stat** → **Basic Statistics** → **Correlation** to obtain the following dialog box.



In this dialog box the BEARS.mtw worksheet is active and the variables *Age*, *Length* (the height of the bear), *Chest.G* (chest girth) and *Weight* have been selected. In order to display the correlation matrix for these four variables, click on **OK** to display the following results. Note that the p-values are shown directly beneath the correlation coefficients. In this example all of the p-values are equal to 0.000, indicating that all of the correlation coefficients are significant.

MTB > Correlation 'Age' 'Length' 'Chest.G' 'Weight'.

Correlations (Pearson)

	Age	Length	Chest.G
Length	0.691 0.000		
Chest.G	0.734 0.000	0.889 0.000	
Weight	0.774 0.000	0.875 0.000	0.966 0.000

Cell Contents: Correlation
P-Value

SAMPLE SIZE AND POWER CALCULATIONS

The commands **Stat → Power and Sample Size**

allow you to calculate power and sample size for the following procedures:

- 1-Sample Z
- 1-Sample t
- 2-Sample t
- 1 Proportion
- 2 Proportions
- One-Way ANOVA
- 2-Level Factorial Design
- Plackett-Burman Design

Example—Determining sample size for a 1-sample Z-test

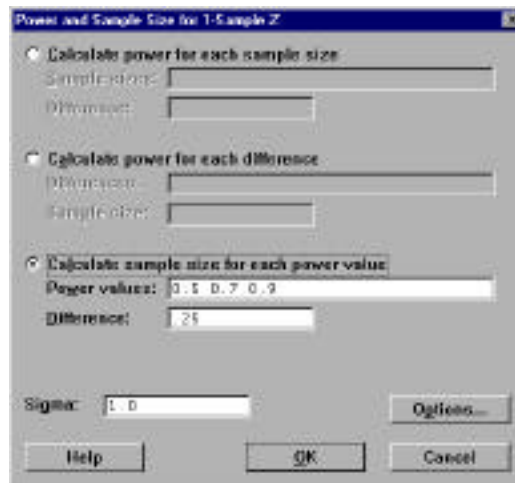
A machine is supposed to fill containers with 32 oz of soda. The quality control department requires that the volume cannot vary more than 0.25 oz per container. The filling machine tolerances are set so that the process standard deviation is 1. How many samples must be taken to estimate the mean container volume at a confidence level of 95% ($\alpha = .05$) for power values of 0.5, 0.7, and 0.9?

Solution

Choose **Stat** → **Power and Sample Size** → **1-Sample Z** and you will see the dialog box below. Observe that there are three different calculations that can be performed:

- Calculate power for each sample size
- Calculate power for each difference
- Calculate sample size for each power value

In the dialog box, the choice **Calculate sample size for each power value** is selected, and the power values of 0.5, 0.7, and 0.9 are entered in the **Power values** box. The value of **Sigma** is set at 1.0 (the default value).



Click on **OK** to obtain the following results:

Power and Sample Size

1-Sample Z Test

Testing mean = null (versus not = null)
Calculating power for mean = null + 0.25
Alpha = 0.05 Sigma = 1

Sample Size	Target Power	Actual Power
62	0.5000	0.5034
99	0.7000	0.7011
169	0.9000	0.9015

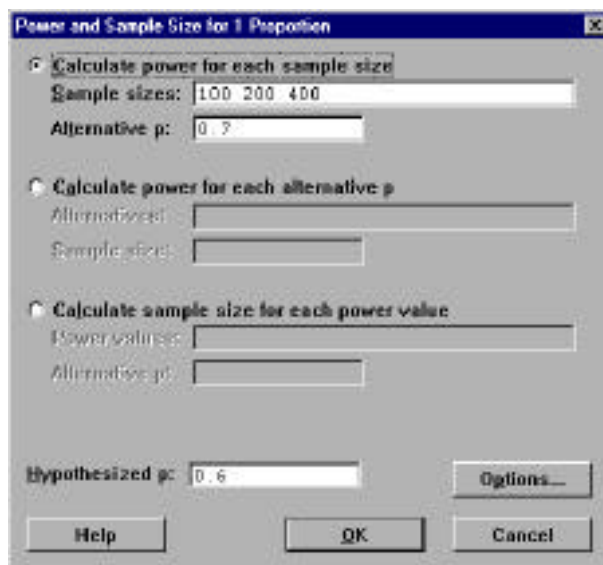
Note that MINITAB displays the sample size required to obtain the requested power values. Because the target power values would result in non-integer sample sizes, MINITAB displays the power (Actual Power) of the test to detect differences in volume greater than 0.5 oz using the next integer value for sample size. For example, if you take a sample of size 99 containers, the power for the test will be .7011.

Example— Determining sample size for tests of proportions

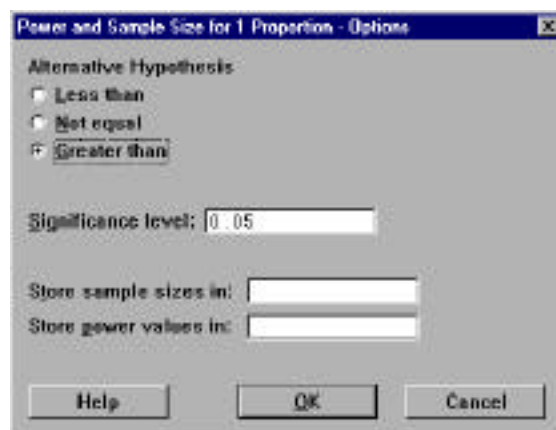
In an educational research project you want to test the hypothesis that the proportion of high-school teachers who express job satisfaction has increased by at least 10% since the latest educational reforms took effect. Previous studies have shown 60% job satisfaction. Thus the test of hypothesis is $H_0: p = 0.6$ vs $H_1: p > 0.6$. Assume a 5% level of significance.

Solution

Choose **Stat** → **Power and Sample Size** → **1-Proportion**.



Next select the **Options** subdialog box and select an upper tail test:



Clicking on **OK** in both boxes we get the following results:

Power and Sample Size

Test for One Proportion

Testing proportion = 0.6 (versus > 0.6)

Calculating power for proportion = 0.7

Alpha = 0.05 Difference = 0.1

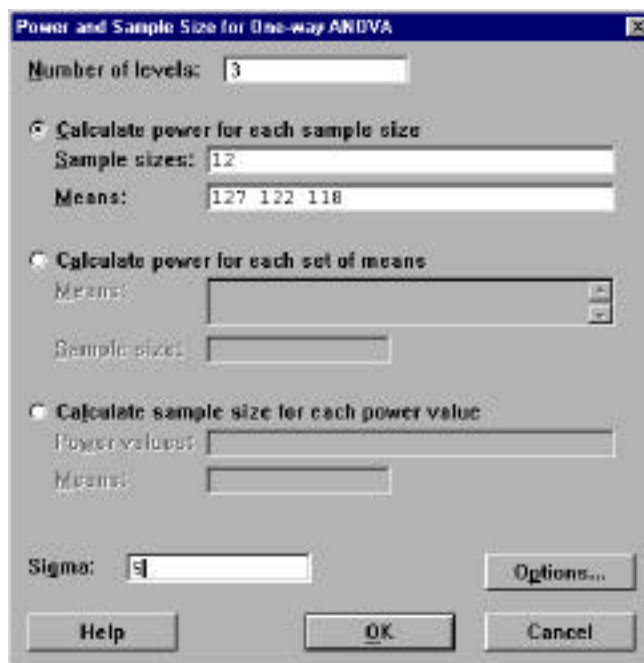
Sample

Size	Power
100	0.6641
200	0.9079
400	0.9954

We conclude that if you select a random sample of 100 high-school teachers, you will have a probability of 66.41% of detecting a 10% increase in the proportion of teachers who express job satisfaction. Similarly sample sizes of 200 and 400 will result in probabilities of 90.79 and 99.59%, respectively, of detecting a 10% increase.

Power and Sample Size for One-way ANOVA

The following dialog box shows how to compute the power for an analysis of variance with 3 treatments (number of levels) each represented by a sample of size 12. In this analysis all sample sizes must be equal. The sample means are shown to be 127, 122, and 118, respectively. The population standard deviation estimate is taken to be $\sigma = 5$.



The results are displayed below:

Power and Sample Size

One-way ANOVA

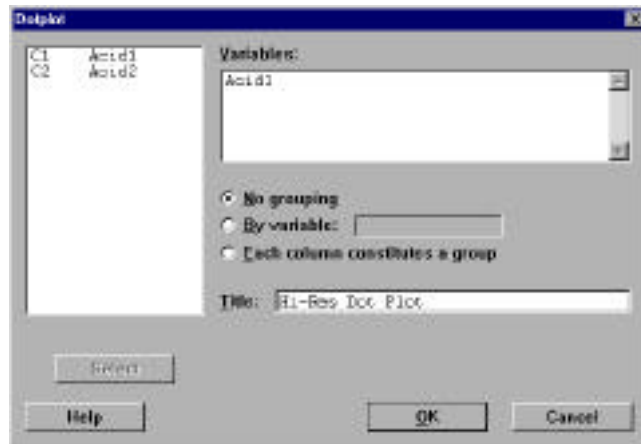
Sigma = 5 Alpha = 0.05 Number of Levels = 3
Corrected Sum of Squares of Means = 40.6667
Means = 127, 122, 118

Sample

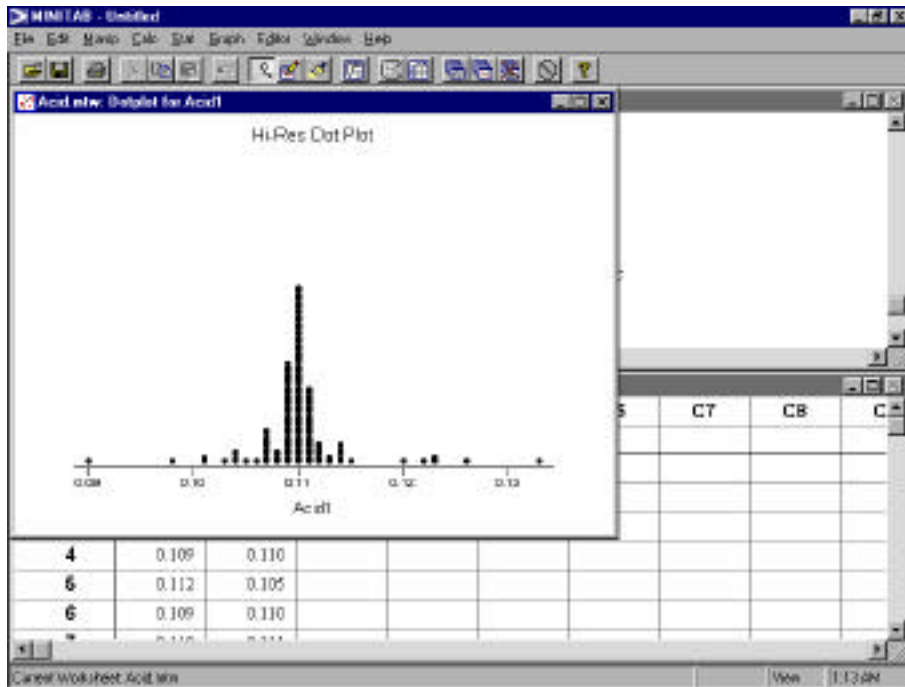
Size	Power
12	0.9729

HIGH RESOLUTION DOT PLOT

Select **Graph** → **Dotplot** to obtain the dialog box. The worksheet **Acid.mtw** is open and the variable Acid1 is selected.



Clicking on **OK** results in the following high-resolution dotplot:



CONCLUSION

This ends our presentation of MINITAB 12. We have emphasized several of the features which are new to this release and which further enhance MINITAB as an excellent choice for teaching statistics.