BUILDING MATHEMATICAL MODELS

Command: **DESCRIBE**
Function: Summarize statistics for each variable listed.
Goal: To assist in inferring the nature of the population (center, spread and shape)
Conclude: Distribution is symmetric if diff. between mean and median is small compared to std.dev.
Limits: Does not calculate IQR (75percentile - 25 percentile) although data is displayed.
Does not calculate mid-IQR ((25percentile + 75percentile)/2) although data is displayed.
Does not calculate skewness or kurtosis (peakedness of shape)
Does not describe much about the shape or outliers.

```
MTB > describe 'FWSVAL' 'WKSWORK'

               N     MEAN   MEDIAN   TRMEAN    STDEV   SEMEAN
FWSVAL       152    36617    33867    34735    26414     2142
WKSWORK      152    42.56    52.00    44.51    17.47     1.42

              MIN      MAX       Q1       Q3
FWSVAL          0   129427    17070    51693
WKSWORK      0.00    52.00    40.00    52.00
```

Command: **CORRELATION**
Function: To generate a table of linear (straight-line) correlation coefficients.
Goal: To help researcher identify the most significant (linear) relationships or associations.
Approach: Generates linear correlations for all pairs of variables and shows in a triangle format.
Misuse: Correlation is meaningless for multinominal data where numbers are arbitrary.
Correlations may be meaningless for ordinal data where interals may be arbitrary.
Limits: Does not indicate the reasons for low correlations:
a. Relationship is correlated but not linearly
b. Data has two separate linear parts -- each highly correlated to a different line.
c. Data is running into edges: floors, ceilings or walls.
Action: If correlation is unexpectedly low, use PLOT to investigate possible causes.
If situation is ‘b’ or ‘c’ above, then use COPY to select relevant portion of the data.
If situation is ‘a’ above, then transform data. (Advanced topic)

```
MTB > CORR C3-C6

       FPERSONS  FRELU18   FRELU6
FRELU18   0.768
FRELU6    0.243    0.416
FPOVCUT   0.895    0.680    0.190

MTB > CORR C6 C3-C5

        FPOVCUT FPERSONS  FRELU18
FPERSONS  0.895
FRELU18   0.680    0.768
FRELU6    0.190    0.243    0.416
```

Command: **STEPWISE**
Function: Exploratory and Creative
Goal: To find the best model to explain the values of a given variable being predicted/explained.
Approach Step 1: Find best one factor model (highest correlation)
Step 2: Find best two-factor model (highest remaining or partial correlation)
Step N: Find best "N" factor model (")
Criteria: Continue until no statistically significant factor remains.
Output: Shape of model (Constant and slope for each variable)
Quality of each model (R^2 at the bottom of each column)
Action: Select best model for population.
Criteria: This population model may involve fewer predictors than best model for sample.
Reason: Diminishing marginal increase in R^2 vs. higher likelihood that relation is sample specific.
Inference: If t-value greater than 2, then coefficient is "statistically significant"
Limits: Does not identify unusual observations; does not generate best fit or residuals.
Does not generate predictions
Errors: Failure to include all relevant variables.
Mixing analytic variables and empirical variables.

```
MTB > STEPWISE 'FWSVAL' on 'FKIND' 'AGE' 'SEX' 'GRADE' 'FPERSONS'
 Stepwise regression of  FWSVAL  on  5 predictors, with N =  152

    STEP        1        2
CONSTANT    62120    21678

FKIND      -17229   -14807
T-RATIO     -7.88    -6.92

GRADE                 2704
T-RATIO               4.34

S           22286    21068
R-SQ        29.29    37.22

 More? (Yes, No, Subcommand, or Help)
SUBC> no

MTB > STEPWISE 'FWSVAL' on 'AGE' 'SEX' 'GRADE' 'FPERSONS'
Stepwise regression of  FWSVAL  on  4 predictors, with N =  152

    STEP        1        2        3
CONSTANT    20528   -22401   -36196

SEX         23514    20360    19194
T-RATIO      5.59     5.17     4.90

GRADE                 3307     3178
T-RATIO               5.13     4.98

FPERSONS                       4448
T-RATIO                        2.32

S           24110    22300    21979
R-SQ        17.24    29.67    32.14

 More? (Yes, No, Subcommand, or Help)
SUBC> no
MTB > # Even though it appears that SEX explains the most,
MTB > # that is true only for this model.
MTB > # If more single-parent familes are headed by females,
MTB > # SEX may be indicating whether the family is a single
MTB > # parent family or a two-parent family (two potential
MTB > # income earners).
```

Command: **REGRESS**
Function: Analytical -- to analyze a particular regression model.
Features: Can generate a list of unusual observations
Can generate best fits and residuals for all rows used in generating the regression line
Can generate confidence intervals and prediction intervals.
Reading: If p-value (p) for a predictor is less than alpha (α) -- user selected probability of error--, then we are $(1-\alpha)*100\%$ confident that the coefficient is "statistically significant".
If p-value (p) for F-test is less than alpha (α) -- user selected probability of error--, then we are $(1-\alpha)*100\%$ confident that the model is "statistically significant".
Limits: Does not create multiple models like STEPWISE.
Automatically assigns all the R^2 to the independent variables collectively.
Advice: First: Use STEPWISE to explore and create the best model.
Second: Use REGRESS to analyze F-test, unusual observations, errors, best fit.
Third: Use REGRESS to generate confidence intervals and prediction intervals.

```
MTB > BRIEF 1
MTB > REGRESS 'FWSVAL' 3 'SEX' 'GRADE' 'FPERSONS'

The regression equation is
FWSVAL = - 36196 + 19194 SEX + 3178 GRADE + 4448 FPERSONS

Predictor       Coef       Stdev    t-ratio        p
Constant      -36196       10648      -3.40    0.001
SEX            19194        3915       4.90    0.000
GRADE         3178.0       637.6       4.98    0.000
FPERSONS        4448        1917       2.32    0.022
s = 21979       R-sq = 32.1%     R-sq(adj) = 30.8%

Analysis of Variance
SOURCE       DF          SS          MS         F        p
Regression    3 33859960832 11286653952     23.36    0.000
Error       148 71493033984   483061056
Total       151 1.05353E+11
MTB > # Since p < .05 for F in analysis of variance (ANOVA),
MTB > #    it appears that this model is "statistically signficant".
MTB > # Since p < .05 for each coefficient, it appears that each
MTB > #    coefficient is "statistically signficant".
MTB > # Since R-sq(adj) is close to R-sq, it appears that we are
MTB > #    not overfitting our data with too many parameters.

MTB > REGRESS 'FWSVAL' 1 predictor 'WKSWORK';
SUBC> PREDICT 0;
SUBC> PREDICT 26;
SUBC> PREDICT 52.
The regression equation is
FWSVAL = 4732 + 749 WKSWORK

Predictor       Coef       Stdev    t-ratio        p
Constant        4732        4932       0.96    0.339
WKSWORK        749.2       107.3       6.99    0.000
s = 23021       R-sq = 24.5%     R-sq(adj) = 24.0%

Analysis of Variance
SOURCE       DF          SS          MS         F        p
Regression    1 25857605632 25857605632     48.79    0.000
Error       150 79495389184   529969248
Total       151 1.05353E+11

     Fit  Stdev.Fit         95% C.I.             95% P.I.
    4732       4932   (   -5015,   14479)  (  -41798,   51262) X
   24211       2577   (   19118,   29304)  (  -21571,   69993)
   43690       2124   (   39492,   47888)  (   -2001,   89380)
X  denotes a row with X values away from the center
```

USING MINITAB GRAPHS FOR EXPLORATORY DATA ANALYSIS

Command: **PLOT Cvert * Choriz**
Function: Generate an XY plot of bivariate data
Goal: To see if data is linearly related and if not, then why not.
a. Is it non-linear in the center, (curvilinear)?
b. Is it jointed in the center (two separate parts)?
c. Is it linear but running into edges: floors, ceilings or walls?

```
MTB > PLOT 'FWSVAL' * 'WKSWORK'
   150000+
         -
 FWSVAL  -                                                      *
         -                                              *       *
         -                                                      *
   100000+                                                      *
         -                                                      *
         -                                                 *    3
         -                                                      +
         -                                    *    *            +
    50000+                                                 * * * +
         - *             *                        *           2  +
         - 2                                              3   **  +
         - *             *            *           *          *   * +
         - *            *              *    *     *     2         7
        0+ + *     *    *  * *   *          *                    *
           +---------+---------+---------+---------+---------+------WKSWORK
           0        10        20        30        40        50
MTB > # Wall:  subjects who make money with no time (WKSWORK = 0)
MTB > # Wall:  subjects who work every week (WKSWORK = 52)
MTB > # Floor: subjects who make no money yet still work (FWSVAL=0)
```

Command **LPLOT Cvert Choriz using Csymbol**
Function: To display different symbols depending on value of 3rd variable for each subject/row.
Limits: Values of third variable must be integers (0=> Z, 1=> A, 2 => B, etc.)
Goal: To have visual display of the "effect" of a third variable on correlation between first two.
Limits: If multiple subjects at same coordinates, number is shown instead of letters
Action: May separate data if symbols form clusters or groups

```
MTB > LPLOT 'FWSVAL'*'WKSWORK' indicate 'SEX'
   150000+
         -
 FWSVAL  -                                                      A
         -                                              A       A
         -                                                      A
   100000+                                                      A
         -                                                      A
         -                                                 A    3
         -                                                      +
         -                                    A   A             +
    50000+                                                 A A A +
         - Z             Z                        A           2  +
         - 2                                              3   AA  +
         - A             A            A           Z          A   Z +
         - A            Z              Z   Z     Z     2         7
        0+ + Z     Z    Z  Z Z   Z          Z                    A
           +---------+---------+---------+---------+---------+------WKSWORK
           0        10        20        30        40        50
MTB > # Letter Z indicates SEX = 0 (Female)
MTB > # Letter A indicates SEX = 1 (Male)
MTB > # Numbers 1-9 indicate number of multiple subjects at same place
MTB > # Plus sign (+) indicates more than 9 subjects at the same place
```

Command: **DOTPLOT C1**
Function: Frequency plot for a single variable (or for each variable if multiple variables listed)
Variable: Best for quantitative continuous data
Not good for qualitative data (binomial, multinominal or ordinal data)
Not real good for discrete data with just a few values.
Problem: When * indicates multiple points, it indicates UP TO the maximum number of points.
Goal #1: Look for multiple peaks. Multiple peaks indicate different groups.
Action: If multiple peaks, you may want to separate the data into different groups
Example: Say two peaks in Wages: -- first at zero and second at $35,000
Break out data into two groups: No income versus some income.
Goal #2: Look for shape (see if it might be from a population which is Normal).
Action: If it looks mound-shaped, we have better reason to use TTEST for small samples.

```
MTB > DOTPLOT 'FWSVAL'
          .
          :                 .
          :           :     :
          :          ::  .  :
          : .  .:  ::: .:. ::         :
          :.: :::::::::.:::.::.. :.: ..:
          ::: :::::::::::::::::::::.. : .  .   ..      .. .
          +---------+---------+---------+---------+---------+-------FWSVAL
          0     25000     50000     75000    100000    125000
MTB > # Notice at least two modes -- first at zero.
```

Command **BOXPLOT**
Function: To summarize three key percentiles of a distribution and display outliers
Output: Box formed from 25th percentile to 75th percentile. + in center is 50th percentile.
Outliers are points more than 1.5 IQR beyond hinges (the ends of the box).

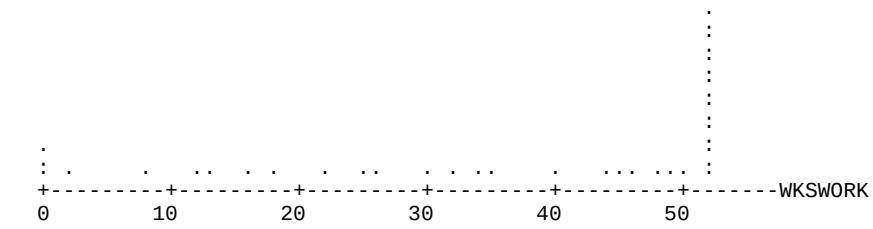
```
MTB > BOXPLOT 'FWSVAL'

                ---------------
          -------I      +      I---------------   **      ** *
                ---------------
          +---------+---------+---------+---------+---------+------FWSVAL
          0     25000     50000     75000    100000    125000

MTB > BOXPLOT 'WKSWORK'

                                                 -------------
          O O     *   **  * *   ------------------I           +
                                                 -------------
          +---------+---------+---------+---------+---------+------WKSWORK
          0        10        20        30        40        50

MTB > # Note that + sign (median) is at right end of box -- highly asymmetric.
```

Command **HISTOGRAM**
Function: Space-efficient summary of data into a small number of groups.
Benefit: Space efficient summary. Can be more meaningful than a DOTPLOT of the same data.
Output: Box formed from 25th percentile to 75th percentile. + in center is 50th percentile.
Outliers are points more than 1.5 IQR beyond hinges (the ends of the box).
Problem: Command automatically calculates center and width of each interval.
Shape of display depends on selection of center and spread for each interval.
Each symbol contains UP TO the maximum number of observations indicated.

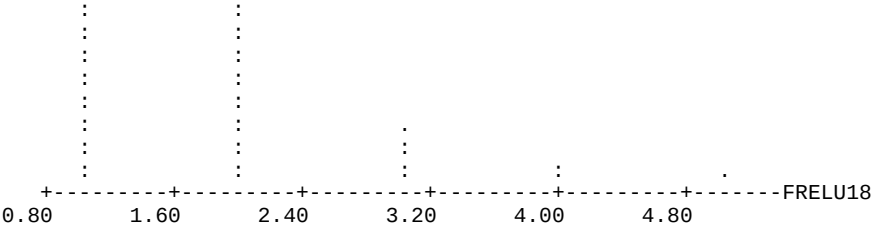
```
MTB > Histogram 'GRADE'.
Histogram of GRADE   N = 152
Each * represents 2 obs.
Midpoint   Count
       0       2  *
       2       0
       4       0
       6       0
       8       1  *
      10       5  ***
      12      59  ******************************
      14      30  ***************
      16      34  *****************
      18      21  ***********
```

```
MTB > Histogram 'WKSWORK'.           # Tricky -- each * represents UP TO 5 observations.
Histogram of WKSWORK   N = 152
Each * represents 5 obs.
Midpoint   Count
       0      16  ****
       5       0
      10       2  *
      15       4  *
      20       2  *
      25       2  *
      30       3  *
      35       4  *
      40       6  **
      45       4  *
      50     109  **********************
```

```
MTB > DOTPLOT 'WKSWORK'          # Tricky -- each dot represents UP TO 7 points.
Each dot represents 7 points
```




```
MTB > DOTPLOT 'FRELU18'              # This is not a good display (use a Histogram)
Each dot represents 4 points
```



MINITAB COMMANDS TO MANIPULATE DATA:

Command: **CODE**
Function: To map values from one column to another based on range-based rules.
Limits: Rule cannot relate to any condition outside the one column involved.
Rule involves transforming values within a range into a single value.
Uses: To generate ordinal data from quantitative data
E.g., Generate indicators for short (1), medium (2) and tall (3) from heights.
CODE (0:62) to 1, (62:70) to 2, (70-99) to 3 from C1 to C11.
To generate binomial data (indicator variables) from quantitative or ordinal data.
E.g., Generate indicators for not-tall (0) and tall (1) from height
CODE (0:66) to 0, (66:99) to 1 from C1 to C12

Command **LET**
Function: To assign values to a column (or constant)
Power: Can assign values based on formulas using multiple columns (variables)
E.g., To compute z values of position relative to the mean:
LET C9 = (C1-Mean(C1))/STDEV(C1)
Can assign values of 1 or 0 depending on whether a logical condition is true or false.
E.g., To identify both low and high values with an indicator of 1.
LET C9 = (C1 < 2) OR (C1 > 90)
E.g., To identify records satisfying complex criteria.
LET C9 = (C1<2) AND ((C2>9) OR (C3 < 1))

Command: **COPY**
Function: To select a subset of the data based on USE and OMIT selection criteria.
Example: MTB > COPY from C1-C9 to C11 to C19;
SUBC> USE C2 = 1;
SUBC> OMIT C3 = 0.
Limit: Limited to a maximum of one USE and one OMIT subcommand per COPY.
USE and OMIT subcommands are limited to
* a single column variable
* a relationship of equals only (=).
Power Can copy multiple columns in a single pass.
COPY commands can be stages so output from one is the input for the next.

Command: **INDICATOR**
Use: Advanced
Function: To create separate binomial indicator variables for ordinal data.
E.g., Breakout C1 into its 9 values:
INDICATOR C1 into C11 - C19
Use: Allow use of ordinal data in a regression.

METHODS TO ACHIEVE CERTAIN GOALS:

Goal: Pick a variable that will have a significant model.
Problem: Some variables are unable to generate a significant regression model.
Methods: 1. Trial and error
2. From the correlation matrix,
   a. Identify the cells having the highest absolute |correlations|.
   b. Pick variables related as columns or rows to such cells.

==========

Goal: Pick a variable whose model will tell us something new or interesting.
Problem: Some correlations are high but they don't tell us anything new.
Reason: Some correlations are high because of the relationship involved.
Highly-correlated variables can be uninformative because they are

1. related as original and transformed or mapped. Examples include
   * Height (inches) and height coded ordinally (tall, medium, short)
   * Income in dollars and income as a percentile (50th %)

   High correlation is expected between near likenesses (twins)
2. related as parts of a formula
   Additive examples involving part and whole include:
   * Costs of goods and revenues; Inventory and Total Assets.
   * Wages and total income; Cost of land and cost of house.
   * Score at midterm and score in entire course.

   Multiplicative examples involving part and whole include:
   * Quantity of a part and the total value of that part.
   * Ave income and total income for a given population.

   Other examples involving input and result include:
   * Number of children and official Poverty level income
3. related systematically but accidentally. Examples include
   * Season of year and closeness of earth to the sun
4. related due to a known common cause. Examples include:
   * Rainfall and season of the year.

Action: Summarize correlations in order by decreasing r-squared or |r|.
Place correlations in two groups: informative and uninformative
Select variables with highest correlations which you find informative.
Work from high (r^2) to low in selecting variables for analysis.

==========

Goal: To identify whether a non-linear correlation exists between two variables
Problem: A high non-linear correlation may yield a moderate or low linear correlation.
Action: Plot the two variables (PLOT C1 C2)
   a. Look for non-linear patterns (single curves, double s-curves, etc.)
   b. Look for the data dividing into two parts
   c. Look for the data running into maximums or minimums

==========

Goal: To minimize the output from the REGRESS command:
Action: Precede REGRESS with BRIEF 1.
Value: This eliminates the “Unusual Observations” part of the REGRESS output.
Warning: This command stays in effect until changed.

METHODS (continued)

Goal: To transform a non-linear correlation into a linear correlation
Problem: Regression builds models based only on linear correlations.
Actions: a. Transform data.
1. To model single curves, square data: Let C11 = C1*C1
2. To model exponential change, log data: Let C11 = LOGE(C1)
3. To model double s-curves, cube data: Let C11 = C1*C1*C1
b. Separate data by creating a new indicator to identify the separation.
Suppose one group has low values of C1 and C2. Other has high values.
c. Create a new indicator: Let C11 = (C1<0.5) OR (C1>9) #Edges at 0 and 10
Copy omitting the data at the edges.

==========

Goal: To use wording that is relevant and meaningful.
Problem: Sometimes the wording is inappropriate
Example: Age is not a factor in determining .....
Better: In this linear model, age was not a statistically significant factor in determining .....

==========

Goal: Coding can be used for two reasons
1. To group data for analytical purposes
2. To group data for ease in using and omitting data in COPY.

==========

Goal: To use language that is precise and meaningful.
Vague: Some language is extremely vague
Reasons: #1: Failure to specify the exact nature of a relationship
Using vague words like “relates”, “changes”, and “affects”.
Example: Smoking affects lung capacity.
This statement doesn't assert 'increase' or 'decrease'
#2: Speaking potentially rather than actually.
Using words like “can”, “may”, and “might”
Example: Smoking can/may/might decrease capacity. John may save Jim
This doesn't omit idea that smoking can increase capacity.
This doesn't omit the possibility that John may not save Jim.
Reason: Avoids taking a stand.
Advice: Use words that are more specific in describing behavior or modality.
#1. Specify direction of relation: increase or decrease
Specify any causality in relation: causes, generates, produces, etc.
#2 Specify how likely something is: “is probable”, “is highly likely”, “almost certain”

==========

Goal: To communicate effectively during your project
Problem: Reader doesn’t know what variables in new columns are.
Advice: After creating new variables (COPY) and naming them (NAME),
make a list (INFORMATION) of the complete set of variables.

==========

ERROR MESSAGES OR ERROR CONDITIONS

Problem: "ALL VALUES IDENTICAL". No output generated.
Command: CORRELATE, STEPWISE or REGRESS
Why: None of these apply to columns where all values are identical.
For example, if all people are female in a sample, one cannot compute a correlation or perform a regression involving that column.
Without some variability, one can't measure closeness.
Choices:
1. Omit that column from such procedures.
2. Don't build that column in the first place

==========

Problem: "COLUMN LENGTHS NOT EQUAL"
Command: CORRELATE, STEPWISE or REGRESS
Why: Correlation between two variables (columns) requires that all subjects (rows) have both properties (values). There is no way to plot a point on an XY plot if that point has no value for one of the two coordinates.
Causes: User errors in using Minitab.

```
1. MTB > CORR C1 C  2   C3     # Wrong syntax
      Error: Column lengths not equal
*  2. MTB > COPY from C1 to C11
   MTB > COPY from C2 to C12;
   SUBC> OMIT when C3 = 1.
   MTB > CORR C11 C12
      Error: Column lengths not equal
   MTB > INFO
         C1  10
         C2  10
         C11 10
         C12  7
```

Fixes:
1. Fix syntax errors
   `MTB > CORR C1 C2 C11 C12`
2. Erase bad data
   MTB > ERASE C11 C12
3. Redo COPY properly.
   ```
   MTB > COPY C1 and C2  to C11 and C12;
   SUBC > OMIT when C3 = 1.
   ```

==========

Problem: "NO VARIABLES ENTERED OR REMOVED". No model built.
Command: STEPWISE
Explain: When no correlation is statistically significant, stepwise will not create a regression model.
Choices:
1. Choose a new variable to model
2. See if an expected relationship is really non-linear (PLOT)
3. See if outliers are affecting an expected relationship (PLOT)
4. Force the model to accept variables regardless of size:
   ```
   MTB > STEPWISE C1 C2-C6;
   SUBC > FENTER = 0;
   SUBC > FREMOVE = 0.
   ```
5. Change (CODE) zeroes to missing values. This may result in creating a model. However the action may not be justified.
   If the zero data does not apply, then it is better to use
   a. COPY and OMIT such records when the variable equals zero.
      If the zero data does apply, then changing it to zero is like falsifying the data.

ERROR (continued)

Problem: "UNUSUAL OBSERVATIONS"
Command: REGRESS
Cause: These are outliers in some sense
Fix: There are several alternatives -- but each must be justified:

1. Leave alone (outliers may be representative and thus good)
2. Isolate them: Create a new indicator (C90) which is 1 for those values.
   All other records will have a value of 0.
   Use this new variable in the regression (STEPWISE, etc.)
   Example: Suppose C1 has a single record with a value of 23,456 (far above 5,000)
   One way is to use the LET command along with a logical condition:
   MTB > LET C90 = C1 > 23000
   A second way is to use the CODE command:
   MTB > CODE (*:9999) as 0, (9999:99999) as 1 from C33 to C90
   A third way is to manually set C90(33) equal to 1.
   MTB > LET K1 = Count(C33)
   MTB > SET c90;
   DATA> K1(0);
   DATA> END.
   MTB > LET C90(33) = 1
   Example: Suppose we want to identify records having multiple conditions:
   Use the LET command with a logical condition involving AND or OR.
   MTB > LET C90 = (C1<1000) OR (C1 > 23000)
3. Adjust them: set them to average value for that variable.
4. Eliminate them (change value to asterisk or delete record)

==========

Problem: Visual problem: output has too many rows or columns (more than 4 or 5) to read.
Command: TABLE or TALLY
Cause: One variable has too many values (probably quantitative continuous)
Fix: Group values into smaller groups (bins) using CODE command.

==========

Problem: Minitab will not generate the table
Command: TABLE or TALLY
Cause: Input has values that are non-integers, that are negative numbers or are greater than 32,000.
Fix: Don't use on quantitative continuous variable. Use HISTOGRAM or DOTPLOT

==========

Problem: Column contains alphabetic (non-numeric data)
Cause: Who knows!
Fix: If created by a copy command, then erase column and recreate data using Copy command.
If original data and non-numeric character is improper, edit using Data Manager.

==========

Problem: Column contains data from an old COPY command that needs to be removed.
Fix: Erase column (ERASE C99)

FILES IN MINITAB:

Minitab uses three kinds of files:

1. Data files which normally have an extension of *.MTW, *.DAT or *.ASC.
   When you view (TYPE), edit (EDIT) or print (PRINT) a Minitab data file, you get weird stuff.
2. Session output files which normally have an extension of *.LIS or *.TXT
   When you view (TYPE), edit (EDIT) or print (PRINT) a Minitab session output file, you see the same stuff you saw on your screen.
3. Command files which normally have an extension of *.MTB or *.MTJ.
   When you view (TYPE), edit (EDIT) or print (PRINT) a Minitab command file, you see nothing but Minitab commands. A journal file can be executed by typing EXEC 'filename.mtb'.

The kind of file is often indicated by the extension, but choosing an extension does not force a Minitab file to be a particular kind. The kind of file is always determined by the command that created it:

1. A data file is created in Minitab by the SAVE command.
2. A session output file is created by the OUTFILE command.
3. A session history file is created by the JOURNAL command

A. If you accessing stored data, you normally don't use the SAVE command
   If you made changes in your data and want to save them, then use SAVE.
B. If you are printing directly onto a printer, you might not use OUTFILE.
   If you want to edit your session output, you should use OUTFILE.
C. If you don't want to repeat you work, you might not use JOURNAL.
   If you want to repeat your work (with some changes), you should use JOURNAL.

Normal procedure for storing session output (OUTFILE)

```
MTB > # Retrieve data set
MTB > RETRIEVE 'C:\Minitab\data\pulse.mtw'
MTB > OUTFILE 'a:\pulse1a.lis'
MTB > # Enter commands and view work until done.
MTB > NOOUTFILE
MTB > # Do not "SAVE" your data (unless you changed it).
MTB > STOP
```

==========

MANIPULATING OUTFILE OUTSIDE MINITAB

----While in DOS using EDIT (MSDOS 5.xx or later)

C:> EDIT a:\pulse1.lis

[From File Menu, open the Minitab OUTFILE: a:\pulse1.lis.]

[Print after editing and saving the OUTFILE data]

--- While in Windows using NOTEPAD

[Open Notepad in Accessories Group]

[File Open OUTFILE: pulse1.lis

[Print data after editing and saving the session OUTFILE data]

--- While in Windows using WORD

[Open File and select Text as kind of data]

Note: You may need to change Files-List of Types to *.* (all files)

[Select OUTFILE: pulse1.lis]

Select entire document.

Format using a non-proportional font (such as New Courier or Courier)

[Print data after editing and saving the session OUTFILE data]