SCL Mini Manual 3
Minitab in the SCL

September 24, 1999

Contents

1 Introduction **3**

2 Getting Started **5**

3 Data Manipulation **7**
3.1 Data Entry . . . 7
3.1.1 Enter Data from the Keyboard . . . 7
3.1.2 Read Data from a File . . . 8
3.1.3 Retrieving Data from a File . . . 8
3.1.4 Naming Columns . . . 8
3.2 Data Output . . . 9
3.2.1 View Your Data . . . 9
3.2.2 Save Worksheet . . . 9
3.3 Record Your Session in a File . . . 9
3.3.1 The UNIX Script Command . . . 9
3.3.2 Minitab Script Command . . . 10
3.4 Entering Patterned Data . . . 10
3.5 Editing Data . . . 11
3.6 Arithmetic . . . 11
3.7 Summary Statistics . . . 12

4 Statistical Analysis **13**
4.1 Computing Descriptive Statistics . . . 13
4.2 The One Sample Problem . . . 14
4.2.1 t-test . . . 14
4.2.2 Confidence Interval . . . 15
4.2.3 Sign Test . . . 15
4.2.4 Sign Confidence Interval . . . 15
4.2.5 Wilcoxon Signed-Rank Test . . . 16
4.2.6 Wilcoxon Signed-Rank Confidence Interval . . . 16
4.3 The Two Sample Problem . . . 17
4.3.1 t-test and Confidence Interval . . . 17
4.3.2 Mann-Whitney-Wilcoxon(MWW) rank sum test and confidence interval 18

4.3.3 Mood median test 19

5 Graphics **21**

5.1 Histogram 21

5.2 Boxplots 22

5.3 Scatter Plot 23

6 Distributions and Random Data **25**

6.1 Random Sample Generation 25

6.2 Distribution 26

6.2.1 Density Function 26

6.2.2 Cumulative Distribution Functions 27

6.2.3 Quantiles 27

Chapter 1

Introduction

Minitab is a powerful statistical software package that provides a wide range of basic and advanced data analysis capabilities. It was developed at Penn State University in the early 70's by Tom Ryan, Brian Joiner, and Barbara Ryan. It is now a corporation housed in State College, Pa.

Essentially MINITAB is a collection of commands which operate on columns of data that are entered via the READ and SET commands or are generated internally. In this mini manual, we will briefly introduction how to use MINITAB in SCL.

Chapter 2

Getting Started

To start Minitab on SCL Sparcstation, simply type `mtb` and press `<RETURN>`. The screen displays information about Minitab, and the prompt `MTB >`. This prompt is Minitab's way of telling you that it is waiting for you to enter a command. The follow messages are displayed in the screen when you type `mtb`:

```
MINITAB Statistical Software, Enhanced Version
Release 9.1 for SUN
(C) Copyright 1992 Minitab Inc. - All Rights Reserved

NOV. 18, 1995 - Western Michigan University

Worksheet size: 100000 cells

  For information on:          Type:
  --------------------------   ---------------
  How to use Minitab           HELP
  Customer service             HELP OVERVIEW 14
  Documentation                HELP OVERVIEW 15
  What's new in this release   NEWS

MTB >
```

To exit Minitab, type `STOP` or `QUIT` after the `MTB>` prompt.

```
MTB > QUIT
Worksheet changes have not been saved.  Stop anyway (Y/N)? y
*** Minitab Release 9.1 *** Minitab Inc. ***
Worksheet size: 100000 cells
```

Chapter 3

Data Manipulation

Minitab places the data that you either enter or retrieve into a temporary storage area called the worksheet. The worksheet is arranged into columns (variables) and rows (individual cases). A worksheet can also contain stored constants and matrices. When you end a Minitab session, your current worksheet and its data disappear since they only existed in temporary memory. If you want to use this data later you should save it into a file.

3.1 Data Entry

Minitab is a collection of commands that manipulate columns of data. The columns are stored in a worksheet. They are called c1, c2, etc. But as described below, they can be named.

3.1.1 Enter Data from the Keyboard

For example, you want to read in the two sets of data $\{2, 8, 12\}$ and $\{5.6, 7.8, 16.7\}$ simply type

```
MTB > READ c1 c2
 DATA> 2 5.6
 DATA> 8 7.8
 DATA> 12 16.7
 DATA> END
       3 rows read.
```

or equivalently

```
 MTB > SET c1
 DATA> 2 8 12
 DATA> END
 MTB > SET c2
```

```
DATA> 5.6 7.8 16.7
DATA> END
```

Now the data sets are stored in the two vectors called `c1` and `c2`.
Note: Minitab is not case-sensitive.

3.1.2 Read Data from a File

Suppose we have a data file, say *example.dat* which contains two columns of data (separated by spaces). To enter this data into columns `c1` and `c2` in Minitab, use the Minitab command:

```
MTB > READ 'example.dat' c1 c2
```

3.1.3 Retrieving Data from a File

This file should be the Minitab saved worksheet which has suffix `.MTW`.
Suppose you have Minitab saved worksheet `example.mtw`, you can use the following command to retrieve data.

```
MTB > RETRIEVE 'example'
Retrieving worksheet from file: example.MTW
Worksheet was saved on  3/19/1996
MTB >
```

3.1.4 Naming Columns

Use the `NAME` command and always enclose the name in single quotation marks.

```
MTB > NAME c1 'Var1' c2 'Var2'
```

Now print your data again and you can see:

```
MTB > PRINT c1 c2


 ROW   Var1   Var2

   1      2    5.6
   2      8    7.8
   3     12   16.7
```

3.2 Data Output

3.2.1 View Your Data

When you are working with a data set, you may want to see a listing of all the variables in the current worksheet. To display a summary that lists column numbers, column names, and the number of rows in each column, use the `INFO` command.

```
MTB > INFO

Column   Name          Count
C1       Var1              3
C2       Var2              3
```

To look at the data you just entered by using the `PRINT` command.

```
MTB >  PRINT c1 c2


 ROW   Var1   Var2

   1      2    5.6
   2      8    7.8
   3     12   16.7
```

3.2.2 Save Worksheet

It is a good idea to save your worksheet frequently. This is a good time to do that since you have just entered new data. Minitab automatically adds the extension `.MTW` to a file you save with the `SAVE` command to indicate that it is a binary saved worksheet file.

```
MTB > SAVE 'example'
Saving worksheet in file: example.MTW
```

3.3 Record Your Session in a File

There are two ways to record a Minitab session.

3.3.1 The UNIX Script Command

This method is UNIX command. Before you login Minitab, at UNIX prompt, type:

`script` *filename*

Then screen will display:

```
Script started, file is filename
```

After then, you can use Minitab, the whole your typing and screen display will be saved in file *filename*. After you quit Minitab and return to UNIX prompt, then type `exit`, the system will stop scripting. The Minitab session will be in *filename*.

3.3.2 Minitab Script Command

Using the `OUTFILE` '*filename*' in Minitab, then the system will record your session and save them in file *filename*. When you want to stop, just type `NOOUTFILE`.

Example:

```
/home/student/zhao : mtb

 MINITAB Statistical Software, Enhanced Version
 Release 9.1 for SUN
 (C) Copyright 1992 Minitab Inc. - All Rights Reserved

 NOV. 20, 1995 - Western Michigan University

 Worksheet size: 100000 cells

   For information on:            Type:
   --------------------------     ---------------
   How to use Minitab             HELP
   Customer service               HELP OVERVIEW 14
   Documentation                  HELP OVERVIEW 15
   What's new in this release     NEWS

 MTB > OUTFILE 'file.lis'
 Collecting Minitab session in file: file.lis
 MTB >
 MTB > NOOUTFILE
```

3.4 Entering Patterned Data

You can always enter data by typing it in. However, there is an easier way if your data follow a pattern. If you want to read data set $\{1, 1, 1, 1, 1, 2, 2, 2, 2, 3, 3, 3\}$, type

```
 MTB > SET c3
```

```
DATA> 5(1) 4(2) 3(3)
DATA> ENd
```

SET command provides abbreviations that allow you to enter patterned data in a convenient way. For example.

$1:4$	expands to $1, 2, 3, 4$
$4:1$	expands to $4, 3, 2, 1$
$1:3/.5$	expands to $1, 1.5, 2, 2.5, 3$
$4(1)$	expands to $1, 1, 1, 1$
$2(1:3)$	expands to $1, 2, 3, 1, 2, 3$
$(1:3)2$	expands to $1, 1, 2, 2, 3, 3$
$2(1:3)2$	expands to $1, 1, 2, 2, 3, 3, 1, 1, 2, 2, 3, 3$

3.5 Editing Data

`LET c(i)=k`	Corrects a number in the worksheet, i.e. $c_i = k$
`DELETE i j c1 c2`	Deletes rows i and j of vectors **c1** and **c2** and moves the remaining rows up to close the gap.
`COPY c1 c2 c3 c4;`	Copy from **c1** to **c3**, **c2** to **c4**,
`USE c1=k;`	Copy selected rows ($c1_i = k$)
`OMIT i j.`	Omit rows i and j.

3.6 Arithmetic

Using LET command to do some simple arithmetic.
For example:

`LET c3=c1*5`	$c3_i = 5c1_i, i = 1, \ldots, n$
`LET c4=c1+c2`	$c4_i = c1_i + c2_i, i = 1, \ldots, n$
`LET c5=c1/c2`	$c5_i = c1_i/c2_i, i = 1, \ldots, n$
`LET c6=c1**2`	$c6_i = c1_i^2, i = 1, \ldots, n$
`LET c7=sqrt(c1)`	$c7_i = \sqrt{c1_i}, i = 1, \ldots, n$

3.7 Summary Statistics

`SUM c1 k`	Computes the sum of **c1**, $k = \sum c1_i$
`MAXIMUM c1 k`	Computes the maximum value of **c1**, $k = \max\{c1_i\}$
`MINIMUM c1 k`	Computes the minimum value of **c1**, $k = \min\{c1_i\}$
`MEAN c1 k`	Computes the mean of **c1**, $k = \frac{1}{n} \sum c1_i$
`MEDIAN c1 k`	Computes the median of **c1**, $k = \text{median}\{c1_1, \ldots, c1_n\}$
`STDEV c1 k`	Computes the standard deviation of **c1**, $k = \sqrt{\sum (c1_i - \overline{\mathbf{c1}})^2/(n-1)}$

Chapter 4

Statistical Analysis

Minitab offers a wide array of basic statistics to help you to analyze your data. For example, you can do t-tests, correlations, and more.

4.1 Computing Descriptive Statistics

For an example data set, let's use some data collected on 59 professional baseball players. This data set consists 6 columns. The first column is "Height", the second column is "Weight", the third column is "Bats" (a "1" is coded for right-handed, "2" for left-handed, and "3" for switch-hitters), the fourth column is "Throws" ("0" for right, and "1" for left), the fifth column is "Pitch" ("0" if a pitcher, "1" if not), and the sixth column is "Average" (ERA if a pitcher, and batting average if not a pitcher).
Use the `DESCRIBE` command to produce a summary table describing the three variables Height, Weight, and Average. The `BY` subcommand with `DESCRIBE` tells Minitab to produce separate summary statistics for each Pitch.

```
MTB > READ 'baseball.dat' c1-c6
Entering data from file: baseball.dat
     59 rows read.
MTB > NAME c1 'Height' c2 'Weight' c3 'Bats'
MTB > NAME c4 'Throws' c5 'Pitch' c6 'Average'
MTB > DESCRIBE 'Height' 'Weight' 'Average';
SUBC> BY 'Pitch'.
```

```
          Pitch       N     MEAN   MEDIAN   TRMEAN    STDEV   SEMEAN
Height        0      26   75.192   75.500   75.208    1.960    0.384
              1      33   72.667   73.000   72.655    2.217    0.386
Weight        0      26   201.00   200.00   201.42    17.48     3.43
              1      33   185.42   185.00   185.10    17.27     3.01
Average       0      26    3.722    3.560    3.706    0.595    0.117
              1      33  0.26185  0.27000  0.26234  0.02352  0.00410

          Pitch     MIN      MAX       Q1       Q3
Height        0   71.000   79.000   73.000   76.000
              1   68.000   77.000   71.000   74.000
Weight        0   160.00   232.00   188.75   218.25
              1   155.00   230.00   175.00   195.00
Average       0    2.840    4.990    3.287    3.973
              1  0.21200  0.30300  0.24250  0.28200
```

4.2 The One Sample Problem

Here we look at the one sample problem, including t-test, R-test, and confidence interval.

4.2.1 t-test

Minitab command: TTEST $[\mu = \mu_0]$ c1 c2

Performs a separate t-test on the data in each column. If μ_0 is not specified, $\mu = 0$ is used. TTEST does a two-sided test unless you use the subcommand ALTERNATIVE=k to do a one-sided test. $k = 1$ gives $H_1 : \mu > \mu_0$ and $k = -1$ gives $H_1 : \mu < \mu_0$

TTEST calculates the test statistics:

$$t = \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}}$$

Suppose we wish to test the hypothesis for variable "Height"
$H_0 : \mu = 70$ vs $H_1 : \mu > 70$

```
MTB > TTEST 70 'Height';
SUBC> ALTERNATIVE=1.

TEST OF MU = 70.000 VS MU G.T. 70.000

              N     MEAN    STDEV  SE MEAN        T   P VALUE
Height       59   73.780    2.443    0.318    11.88    0.0000
```

4.2.2 Confidence Interval

We can construct confidence intervals based on the t distribution, with a relatively simple Minitab command:

```
TINTERVAL k c1 ... cn
```

Calculates a sparate $k\%$ t-confidence interval for each column. The default value of k is .95.

Example: Creates 90% CI for variables "Height" and "Weight".

```
MTB > TINTERVAL .90 c1 c2

              N      MEAN    STDEV  SE MEAN   90.0 PERCENT C.I.
Height       59    73.780    2.443    0.318  (  73.248,  74.311)
Weight       59    192.29    18.90     2.46  (  188.18,  196.40)
```

4.2.3 Sign Test

```
STEST [median=k1] c1 ... cn;
ALTERNATIVE=k2
```

Performs a separate sign test of the null hypothesis H_0 : median $= k1$ for each column. If $k1$ is not specified, $k1 = 0$ is used. Unless the subcommand ALTERNATIVE is used, Minitab tests the alternative hypothesis: H_1 : median $\neq k1$. $k2 = 1$ gives H_1 : median $> k1$ and $k2 = -1$ gives H_1 : median $< k1$

Example: Does sign test H_0 : median(Height) $= 70$, vs H_1 : median(Height) $\neq 70$

```
MTB >  STEST 70 'Height'

SIGN TEST OF MEDIAN = 70.00 VERSUS  N.E.  70.00

                N  BELOW  EQUAL  ABOVE   P-VALUE     MEDIAN
Height         59      3      2     54    0.0000      74.00
```

4.2.4 Sign Confidence Interval

```
SINTERVAL k c1 ... cn
```

Calculates three separate sign confidence intervals for each variable. The first interval gives the $k\%$ confidence just below k, and the third the $k\%$ confidence just above. Only rarely can you achieve exact confidence $k\%$ using the standard procedure. The middle confidence interval is found by a nonlinear interpolation procedure, and gives an approximate interval with confidence $k\%$. If k is not specified, 95% confidence is used.

Example: Computes 95% sign confidence intervals for variables "Weight" and "Height".

```
MTB > SINTERVAL 'Height' 'Weight'

SIGN CONFIDENCE INTERVAL FOR MEDIAN

                                  ACHIEVED
                   N    MEDIAN   CONFIDENCE    CONFIDENCE INTERVAL   POSITION
Height            59     74.00     0.9316      (    73.00,   75.00)        23
                                   0.9500      (    73.00,   75.00)       NLI
                                   0.9628      (    73.00,   75.00)        22

Weight            59     190.0     0.9316      (    185.0,   195.0)        23
                                   0.9500      (    185.0,   195.0)       NLI
                                   0.9628      (    185.0,   195.0)        22
```

4.2.5 Wilcoxon Signed-Rank Test

`WTEST [k] c1 ... cn;`

Does a one-sample Wilcoxon signed-rank test of the null hypothesis H_0 : median $= k$ vs H_1 : median $\neq k$ If k is not given, $k = 0$ is used. The subcommand `ALTERNATIVE` works the same way as it does with `STEST`.

Example: Does Wilcoxon signed-rank test H_0 : median(Height) $= 70$, vs H_1 : median(Height) $\neq$ 70

```
MTB > WTEST 70 'Height'

TEST OF MEDIAN = 70.00 VERSUS MEDIAN N.E. 70.00

                    N FOR   WILCOXON             ESTIMATED
                N    TEST  STATISTIC  P-VALUE      MEDIAN
Height         59      57     1632.5    0.000       74.00
```

4.2.6 Wilcoxon Signed-Rank Confidence Interval

`WINTERVAL k c1 ... cn`

Calculates separate confidence intervals for the median of each variable, using a procedure that corresponds to the Wilcoxon signed-rank test. The default value of k is 95.

Example: Computes 95% Wilcoxon signed-rank confidence intervals for variables "Weight" and "Height".

```
MTB > WINTERVAL 95 'Height' 'Weight'

                   ESTIMATED    ACHIEVED
               N      MEDIAN  CONFIDENCE  CONFIDENCE INTERVAL
Height        59       74.00        95.0  (    73.00,    74.50)
Weight        59       192.5        95.0  (    187.5,    197.5)
```

4.3 The Two Sample Problem

Here we look at the two sample problem, including two-sample t-test, Mann-Whitney-Wilcoxon(MWW) rank sum test and confidence interval, and Mood's median test.

4.3.1 t-test and Confidence Interval

```
TWOSAMPLE [k] c1 c2
```

Does a two (independent) sample t-test and confidence interval. Minitab tests the null hypothesis $H_0 : \mu_1 = \mu_2$ vs $H_1 : \mu_1 \neq \mu_2$ and calculates a confidence interval for $\mu_1 - \mu_2$. If k is omitted, a 95% confidence interval is calculated.

Example: Compares the weight between pitcher group and hitter group.
We first use command `UNSTACK` to devide variable "Weight" into two group. One group(`c10`) is for pitcher, the other group(`c11`) is for hitter.

```
MTB > UNSTACK 'Weight' c10 c11;
SUBC> SUBSCRIPTS 'Pitch'.
MTB > NAME c10 'Pitcher' c11 'Hitter'
MTB > TWOSAMPLE 'Pitcher' 'Hitter'

TWOSAMPLE T FOR Pitcher VS Hitter
          N      MEAN     STDEV   SE MEAN
Pitcher  26     201.0      17.5       3.4
Hitter   33     185.4      17.3       3.0

95 PCT CI FOR MU Pitcher - MU Hitter: ( 6.4,  24.7)

TTEST MU Pitcher = MU Hitter (VS NE): T= 3.42  P=0.0012  DF=  53
```

4.3.2 Mann-Whitney-Wilcoxon(MWW) rank sum test and confidence interval

`MANN [k] 'y' 'x'`

Performs a two-sample rank test (also called Mann-Whitney-Wilcoxon test) for the difference between the two population medians. It also calculates the corresponding point and confidence interval estimates.

By default, a two-sided test is performed. To do one-sided tests, use the ALTERNATIVE subcommand. This is same as `STEST`.

The test is carried out as follows.

1. the two samples are ranked together;

2. computes the sum of the ranks of the first sample, $W = \sum_{j=1}^{n_2} R(y_j)$

3. computes $S_R^+ = W - \frac{n_2(n_2+1)}{2}$, where n_2 is sample size of $\mathbf{y}$

4. does hypothesis $H_0 : \Delta = 0$, vs $H_A : \Delta \neq 0$, the test statistics is:

$$Z = \frac{S_R^+ - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n+1)}{12}}}$$

5. computes $k\%$ confidence interval

Example: Compares the weight between pitcher group and hitter group.

```
MTB > MANN 'Pitcher' 'Hitter'

Mann-Whitney Confidence Interval and Test

Pitcher    N =  26     Median =      200.00
Hitter     N =  33     Median =      185.00
Point estimate for ETA1-ETA2 is       15.00
95.0 pct c.i. for ETA1-ETA2 is (5.00,24.99)
W = 986.0
Test of ETA1 = ETA2  vs.  ETA1 n.e. ETA2 is significant at 0.0017
The test is significant at 0.0016 (adjusted for ties)
```

4.3.3 Mood median test

`MOOD c1 c2`

Performs Mood's median test on sample data contained in the `c1` with subscript (group) identifiers in `c2`.

Example: Compares the weight between pitcher group and hitter group.

```
MTB > MOOD 'Weight' 'Pitch'

Mood median test of Weight

Chisquare = 8.84   df = 1   p = 0.003

                                        Individual 95.0% CI's
   Pitch   N<=     N>   Median    Q3-Q1 --+---------+---------+---------+----
       0     8     18    200.0     29.5               (-------+------------)
       1    23     10    185.0     20.0 (--+------)
                                        --+---------+---------+---------+----
                                      184.0     192.0     200.0     208.0
Overall median = 190.0

A 95.0% C.I. for median(0) - median(1): (5.0,20.9)
```

Chapter 5

Graphics

Minitab supplies two kinds graphics mode. One is Standard Graphics mode, the other one is Professional Graphics mode. In SCL, when you login Minitab, the default mode is Professional Graphics mode. You can enter the Minitab command `GSTD` to switch into Standard Graphics mode. Enter the command `GPRO` to switch back to Professional Graphics mode.

Standard Graphics commands allow you to do simple character-style and high-resolution graphs. Professional Graphics commands allow you to produce a much wider range of high-resolution graphs. In this manual, we will use Standard Graohics mode. For Professional Graphics mode, you can reference MINITAB Graphics Manual.

5.1 Histogram

A histogram for a data list is generated with the `HISTOGRAM` command.

```
HISTOGRAM c1 ... cn
```

Produces a histogram of the data in specified variable(s). The range of the data is divided into intervals, and the count of observations in each interval is displayed both numerically and graphically.

Character style histograms use a horizontal line of '*'s to represent the bars. If there are more than 50 observations in a class, each * represents two or more observations. In this case, the number of observations represented by each * is displayed. Observations falling on a boundary are put in the interval with the larger midpoint. Up to 100 levels (or classes) can be displayed in a single histogram.

Example: Does histogram of variable "Weight".

```
MTB > HISTOGRAM 'Weight'
```

```
Histogram of Weight   N = 59

Midpoint   Count
     160       6  ******
     170       2  **
     180       6  ******
     190      17  *****************
     200      12  ************
     210       8  ********
     220       4  ****
     230       4  ****
```

5.2 Boxplots

Boxplots display the main features of a batch of data, and permit simple comparisons of several batches. The middle half of each variable is represented by a box and the median is marked with a "+". The extent of the data and the location of possible extraordinary values are indicated on either side of the box with special symbols. The command is:

```
BOXPLOT c1
```

Example: Making boxplot of variable "Weight".

```
MTB > BOXPLOT 'Weight'


                          ---------------
     --------------------I    +         I------------------
                          ---------------
  +---------+---------+---------+---------+---------+------Weight
150       165       180       195       210       225
```

We can use subcommand BY to display boxplot for different group. For example, we want to compare the distribution of weight between pitcher and hitter groups, then we can use two group's boxplot to do that.

```
MTB > BOXPLOT 'Weight';
SUBC> BY 'Pitch'.

Pitch

                                     -------------------
0                ---------------------I     +           I----------
                                     -------------------

                           --------------
1            --------------I     +      I-----------           *
                           --------------
          +---------+---------+---------+---------+---------+------Weight
        150       165       180       195       210       225
```

5.3 Scatter Plot

A Scatterplot is a useful tool for determining the relationship between two variables. The command is:

```
PLOT c1 c2
```

Produces a scatter plot of two variables; the first(`c1`) on the vertical axis and the second(`c2`) on the horizontal axis. You can add titles, footnotes, and axis labels to your plot with the subcommands `TITLE, FOOTNOTE, XLABEL, YLABEL`, change the plotting symbol with `SYMBOL`, and change Minitab's default scaling with `XINCREMENT, XSTART, YINCREMENT, YSTART`

Example: If we want to look at the relationship between height and weight, the one method is through graphics display, using scatterplot.

```
MTB > PLOT 'Weight' 'Height';
SUBC> TITLE 'The relationship between weight and height';
SUBC> XLAB 'Height';
SUBC> YLAB 'Weight'.
```

```
                  The relationship between weight and height
         -
         -                                              *         *
      225+                                                   2
         -                          *    *              *    *
         -
W        -                          2              2
e        -                          *         2    *
i     200+                          *         3    *
g        -                *         2    2    *    *
h        -                     2    *         *    2    *
t        -                2    *    3    2    *    *
         -                     *                   *
      175+ *         *         *    *
         -                               *
         -                *
         -     2         *         *
         -          *    *
           +---------+---------+---------+---------+---------+------
        68.0      70.0      72.0      74.0      76.0      78.0
                                    Height
```

Chapter 6

Distributions and Random Data

6.1 Random Sample Generation

We can use Minitab to generate random sample which follows specified distribution. The distributions available in Minitab are the bernoulli, binomial, possion, uniform, normal, t, F, cauchy, laplace, logistic, lognormal, χ^2, exponential, Γ, weibull and β.

The Minitab command is:
`RANDOM n c1 ... cn;`
`BERNOULLI` p =k
`BINOMIAL` n =k p =k
`POISSON` μ =k
`NORMAL` μ =k σ =k
`UNIFORM` a =k b =k
`t` v =k
`F` v_1 =k v_2 =k
`CAUCHY` a =k
`LAPLACE` a =k
`LOGNORMAL` μ =k σ =k
`CHISQUARE` v =k
`EXPONENTIAL` θ =k
`GAMMA` α =k β =k
`WEIBULL` α =k β =k
`BETA` α =k β =k

Generates a separate random sample of n observations into each column. The subcommand specifies which distribution you want. If no subcommand is given, data are simulated from a standard normal distribution.

Example: Generates 5 random samples, each containing 10 observations, from a binomial distribution with $n = 5$ and $p = 0.3$.

```
MTB > RANDOM 10 c1-c5;
SUBC> BINOMIAL 5 0.3.
MTB > PRINT c1-c5


 ROW   C1   C2   C3   C4   C5

   1    2    1    0    3    1
   2    1    2    0    2    0
   3    1    1    1    1    1
   4    3    1    0    1    1
   5    0    0    1    2    2
   6    2    2    0    3    1
   7    1    1    1    1    1
   8    3    1    3    1    3
   9    0    3    2    2    2
  10    1    3    3    1    0
```

6.2 Distribution

6.2.1 Density Function

Densities are obtained via the command:

```
PDF k
```

Calculates probability density function of the distribution specified in the subcommand. All the subcommands work the same way as they do with `RANDOM`. For a discrete distribution, `PDF` calculates probabilities for the specified values. For a continuous distribution, `PDF` calculates the(continuous) probability density function.

Example of `PDF` with Binomial Data

```
MTB > SET c1
DATA> 2 3 4
DATA> END
MTB > PDF c1;
SUBC> BINOMIAL 5 0.3.
        K          P( X = K)
     2.00             0.3087
```

```
      3.00             0.1323
      4.00             0.0284
```

6.2.2 Cumulative Distribution Functions

Calculating probabilities can be done via command:

```
PDF k
```

The `CDF` for any value x is the probability that a random variable with the specified distribution has a value less than or equal to x. That is:

$$\text{CDF}(x) = \Pr(X \leq x)$$

Example of `cdf` with Binomial Data

```
MTB > CDF c1;
SUBC> BINOMIAL 5 0.3.
        K  P( X LESS OR = K)
     2.00             0.8369
     3.00             0.9692
     4.00             0.9976
```

6.2.3 Quantiles

Critical values(quantiles) are obtained via the command:

```
INVCDF k
```

`INVCDF` returns the inverse of the CDF. This means that for a given probability, p, `INVCDF p` returns the value $F^{-1}(p)$. By definition, `INVCDF` requires $0 \leq p \leq 1$.

Example

```
MTB > INVCDF .95;
SUBC> CHISQUARE 1.
    0.9500    3.8415
```