**Statistics 213**
**Introduction to MINITAB**

**Objective:** To become familiar with MINITAB:

i) windows format – how to use the menus and dialog boxes
ii) inputting data – how to input your own data
– how to retrieve particular files
iii) finding descriptive statistics
iv) some graphical routines

Click on **Start** in the lower left-hand corner of the screen. Click on **MathStats Packages** and then **Minitab 13 for Windows**. Finally, click on **Minitab.**
Spend a few moments finding your way around the MINITAB windows. Our main interest will be in the DATA window, which serves as the worksheet for entering data, and the MENU bar that allows the selection of commands to perform various tasks. Results are presented in the SESSION window.

**Example 1:** Entering data
The values below represent the grade point averages of 30 randomly selected U. of C. Students:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 2.0 | 3.1 | 1.9 | 2.5 | 1.9 | 2.3 | 2.6 | 3.1 | 2.5 | 2.1 |
| 2.9 | 3.0 | 2.7 | 2.5 | 2.4 | 2.7 | 2.5 | 2.4 | 3.0 | 3.4 |
| 2.6 | 2.8 | 2.5 | 2.7 | 2.9 | 2.7 | 2.8 | 2.2 | 2.7 | 2.1 |

i) Enter this data into the first column, c1, of the worksheet in the DATA window. Note that it will be necessary to use the "down arrow" key after typing each value to move down the column.
Enter GPA in the blank cell under the c1. This assigns a name for the "variable" in that column.
Before proceeding, be sure that the highlighted cell is not one of your data values – click anywhere else on the worksheet.

ii) From the MENU BAR select GRAPH > HISTOGRAM... a dialog box appears.
Click the mouse on c1 GPA in the rectangle on the left of the box: then press SELECT. This enters the name of your variable as one to be graphed. {Or type c1 in the GRAPH VARIABLES cell.}
Move to the table headed DATA DISPLAY and click the arrow in the DISPLAY cell. Select BAR from the choices given. Be sure the FOR EACH column specifies GRAPH. Press OK and note the results.
Return to the HISTOGRAM dialog box and investigate the variations that are possible with the buttons and menus available in that box.

iii) Again, select GRAPH from the MENU BAR. Choose CHARATCER GRAPHS > STEM-AND-LEAF.
Highlight c1 GPA as before and press SELECT. Select INCREMENT and type 1. Then click on OK.
Have a look at the dotplot for this data by using CHARACTER GRAPHS > DOTPLOTS.

(iv) To find the sum, mean, standard deviation etc. Click on Calc>Column Statistics. Click on Input variable: Click on CI. Click on any of the statistics (eg. Sum, mean) and then click on Okay.

**Example 2:** Using MINITAB files
Students in an introductory statistics course participated in a simple experiment. Each student recorded his or her height, weight, gender, smoking preference, usual activity level, and resting pulse. Then they each flipped a coin and those who's coin came up heads ran in place for one minute. The entire class then recorded their pulse rates again. The data is saved in a worksheet called PULSE.MTW

To retrieve the worksheet: Choose FILE>OPEN WORKSHEET ... a dialog box appears. Click on the icon Data. Type PULSE.MTW in the FILE NAME box – or – scroll through the files listed below it until you find the pulse file and click on it.
Press OK. The results of this experiment appear in the worksheet in the DATA window.

| Variable | Description | Variable | Description |
|---|---|---|---|
| C1 'PULSE1' | Resting pulse rate | C5 'SEX' | 1=male 2=female |
| C2 'PULSE2' | Second pulse rate | C6 'HEIGHT' | Height in inches |
| C3 'RAN' | 1= ran in place;<br>2=did not run in place | C7 'WEIGHT' | Weight in pounds |
| C4 'SMOKES' | 1=smokes regularly<br>2=does not smoke regularly | C8 'ACTIVITY' | Usual level of physical activity;<br>1=slight;<br>2=moderate;3= a lot |

i) Compare the resting pulse rates of the students according to their level of activity by examining the three dotplots: GRAPH>CHARACTER>DOTPLOTS
Select PULSE1 as the variable to be plotted {highlight and SELECT}
Click the BY VARIABLE check box: click in the text box following it; double click on c8 ACTIVITY in the box listing the variables {or click and SELECT}
Click OK and the dotplots are given one above the other.
To compare the actual values: STAT>Basic Statistics> Display Descriptive Stats. Click the Variable Box , double click on PULSE1. Click the By Variable check box. Double click on Activity. Click on OK or hit enter.

ii) Compare the descriptive statistics of the resting pulse rate according to activity level in a similar fashion.

iii) Draw boxplots of the resting pulse rate according to activity level:
GRAPH>BOXPLOT. Designate PULSE1 as the Y (measurement) and ACTIVITY as the X (category). Click OK.

iv) Use MINITAB to help you answer the following question:

1. Compare the resting pulse rate of males and females by discussing the relative position of their central measures; the spread and shape of the respective distributions.
2. What is the range of resting pulse rate for those classified as regular smokers?
3. What is the difference in the mean 'second pulse' rate for those who ran in place and the mean of those who did not?

**Additional Questions:**

**1.** Suppose you select 20 students from the class (Stat 213) and collect data for these 3 variables

X: # of courses enrolled in
Y: total cost of textbooks and supplies for courses enrolled in
Z: method of payment used to by textbooks and supplies for courses enrolled in

(a) Describe the population.
(b) Describe the sample.
(c) Classify the responses of the above variables as qualitative or quantitative. If quantitative, identify as discrete or continuous.

**2.** A consumer group has made a study of the amount of music on classical compact discs. The discs have a capacity of 70 minutes of music but many discs are released with much less. The results of their study of a sample of 318 classical discs are shown below:

| Playing Time | Relative Frequency |
|---|---|
| 0 and under 30 | 0.025 |
| 30 and under 35 | 0.186 |
| 35 and under 40 | 0.308 |
| 40 and under 45 | 0.211 |
| 45 and under 50 | 0.135 |
| 50 and under 55 | 0.094 |
| 55 and under 60 | 0.006 |
| 60 and under 65 | 0.025 |
| 65 and under 70 | |

(a) Complete the table by filling in the missing relative frequency. (0.01)
(b) Draw the first three classes of the relative frequency histogram for this data.
(c) What **percentage** of the sample had playing times of at least 50 minutes? (13.5%)
(d) **How many** CD's in the sample had playing times of at least 30 minutes but less than 60 minutes? 298.92

**3.** The following summary statistics were obtained from a data set with 1000 values:

Mean = 60.5 Median = 64 Standard deviation = 10.5 $Q_1 = 55.5$ $Q_3 = 76.0$

Indicate approximately **how many** observation are in each of the intervals below. Explain your answer including a statement of any assumptions that you made and any observations about the validity of those assumptions.

(a) The interval from 55.5 to 76.0 (500)
(b) The interval from 39.5 to 81.5 (950 Emperical, at least 750 Chebyshev's Rule)

**4.** The campus branch of the Royal Bank randomly selected a sample of student accounts and checked the corresponding balance (rounded to the nearest dollar amount). The resulting data follows :

458, 623, 449, 721, 441, 498, 609, 557, 587, 435, 567, 760, 505, 876, 535, 684

a) Find the following: mean, median, mode, Q1, Q3, standard deviation, and the interquartile range. ( $\bar{x}$ = 581.5625, median = 562, no mode, Q1 = 478, Q3 = 653.5, s = 126.7217, IQR = 175.5
b) Using the Chebyshev's inequality, within what range would you expect to find at least 75% of the account balances? (328.1205, 835.0045)
c) Find the $80^{th}$ percentile and interpret its meaning. (684)

**5.** A sample of 100 values has a mean of 41 and a standard deviation of 6. The shape of the distribution is unknown. How many values would you expect to be contained in the interval 20 to 62? (91.84)

**6.** For a list of 50 sample values, $\Sigma X_i = 50$ and $\Sigma X^2_i = 646$. In what interval would you expect to find at least 8/9 of the data? (-9.463, 11.462)