

STAT 6200 — Design of Experiments for Research Workers
Lab – Due Thursday, September 2*

In this lab we will get an introduction to the Statlab, Minitab, and how to produce summary graphs and summary statistics in Minitab.

- First, follow the in-class instructions on how to log in to your account and change your password.

Now for the lab:

- Open Minitab by double-clicking the Minitab icon on the desktop.

Minitab allows you to enter data into a worksheet or to bring in data from an external worksheet (e.g., an Excel worksheet, a Minitab worksheet created previously, an ASCII (plain text) file, etc.). We will work with data sets from our text, which are stored on the CD that comes with the text. The Minitab data sets from this CD have also been copied to the 6200 folder of the public Data directory in the Statlab.

At the top of p.25 of our text appears a data set giving the causes of death for 100 children between the ages of 5 and 9 years of age who suffered injury-related deaths. The data are nominal, but they have been coded with numbers 1-5, which represent motor vehicle accident, drowning, home fire, homicide, and other cause, respectively.

- Read these data into Minitab by executing the following commands (using the following menu selections):

File > Open Worksheet >

and then select file "injury deaths" from

K:\6200\Pagano2\ Minitab\chap02\

- To obtain a frequency distribution for cause of death, execute the following commands:

Stat > Tables > Tally

Then select col1 into the list of variables and hit OK. Write down the frequency distribution in the space below.

* This lab should be do-able entirely in class. You will be responsible for doing future labs independently, outside of class.

- Now to obtain a bar chart of Cause of Death, select:

Graph > Chart

Then click on the first cell under “X” and select variable Col1. Under **Annotation > Title** enter a title such as “Causes of injury death for 100 US children, ages 5-9, 1980-1985”. Now hit OK. Changes to the axis labels and tick marks can be made by double-clicking these labels.

- Change the category labels from 1–5 to “Motor Vehicle Accident”, “Drowning”, “Fire”, “Homocide”, and “Other”. Also, change the x and y-axis labels to “Cause of Death” and “Number of Injury Deaths”, respectively.

Now let's produce a histogram. For this, we'll use the low birth weight data discussed in class. Recall these data contain systolic blood pressure, apgar score, and other variables on 100 low birth weight infants.

- Read these data into Minitab by selecting:

File > Open Worksheet

and then select file "lowbwt" from

K:\6200\Pagano2\ Minitab\exercise\

- Now, let's produce a histogram of Apgar score at 5 minutes (just like the one on p.29 of the notes). Select

Graph > Histogram

Then under “X” select variable Apgar5. Under **Annotation > Title** give the histogram a title like “Histogram of Apgar Scores at 5 Minutes”. Under “Options” you can select the y-axis type (frequency, percent, density), the type of intervals or bins (midpoint, cutpoint) and the selection of bins. Specify Frequency, Midpoint, and Automatic, and then hit OK twice.

Now let's compute a frequency polygon of apgar score. Again, select

Graph > Histogram

Under "X" select variable Apgar5. Under **Annotation > Title** give a title like "Frequency Polygon of Apgar Scores at 5 Minutes". Under "Options" select Frequency, Midpoint, and Automatic, and then hit OK. Now under "Display" select "Connect" and hit OK. Alternatively, under "Display" select "Lowess" and under "Attributes" select "Degree of smoothing"=0 and "Number of steps"=1. Hit OK twice. Lowess is a smooth approximation to the histogram, try returning the Lowess Attributes to their default values and see what happens.

- One problem with the frequency polygon we have just created is that it doesn't connect to 0 at the right and left extremes. This can be fixed by using the graph tools. Double-click on the graph and you will see a set of tools. Choose the polygon tool on the bottom-left of the top panel of tools and fix your frequency polygon.

Now let's compute separate frequency polygons of apgar score for each value of germhem. There's a wrong way to do this and a right way:

- The wrong way: From **Graph > Histogram** select apgar5 as the X variable; under "Display" choose Connect; under "For each" choose Group; and select germhem as the grouping variable. Then under "Options" choose "Percent" for the type of histogram; under "Frame" choose "Multiple Graphs" and then select "Overlay graphs on the same page"; and hit OK twice.

- The right way:

1. First create two new variables, one containing the Apgar scores of infants without a germinal hemorrhage, and one containing the Apgar scores of infants with a germinal hemorrhage. This can be done by selecting

Manip > Stack/Unstack > Unstack One Column

Then select variable `apgar5` under “Unstack the data in:”; type in “C7 C8” under “Store data in:”; and select `grmh` under “Using subscripts in:”. Two new variables have been created in columns C7 and C8. Give them names “`apgarh0`” and “`apgarh1`”, respectively.

2. Now create the graph. Select:

Graph > Histogram

and under “X” select `apgarh0` in the first row, and `apgarh1` in the second row. Under “Display” choose “Connect”; under “For Each:” choose “Graph”; under “Annotation” give a title such as “Frequency polygons of apgar scores for infants with and without germinal hemorrhage”; and under “Options” choose “Percent” for the type of histogram. Hit OK twice.

Question: What is the difference between the two graphs we just created containing frequency polygons of apgar scores for infants with and without germinal hemorrhage? Why is the second one better than the first.

Now let's create a cross-tabulation of gender versus toxemia.

- Select

Stat > Tables > Cross Tabulation

Now select variables sex and tox under "Classification variables:". Also, click Counts, Row percents, Column percents, and Total percents, then click OK.

Questions:

1. What percent of girls (sex=0) had mothers with toxemia (tox=1)?
2. What percent of mothers with toxemia gave birth to boys?
3. What percent of the infants were girls with mothers without toxemia?

Finally, let's compute some summary statistics.

- Select

Stat > Basic Statistics > Display Descriptive Statistics

Now select variables sbp and gestage under "Variables:". Then click OK.

Questions:

1. What are the mean and median of gestational age?
2. What's the 25th percentile/first quartile of systolic blood pressure?
3. What's the range (max minus min) of gestational age?