

Math 218 - Minitab

Homework #3

1) This first problem involves simulating a random situation in order to estimate an unknown probability, and you need to supply exactly the following four pages. The first page is a printout of your session window for this problem, which begins with the comment symbol # and gives your name, ID, and lab time. The next two pages are labeled graphs, and the last page is your handwritten answers to the questions that follow. Please do not hand in printouts of your data windows.

After entering the information requested above, using **Calc/Set Base**, enter the last four digits of your student ID number as a base for the random number generator. With each person using a different base, different people will generate different data, and you or the TA can replicate your particular experiment using your base.

Experiment 1: Use $n=100$. Experiment 2: Use $n=2500$.

Using **Calc/Random Data/Normal**, put n standard normal random values in column $c1$, and put n normals with mean 0, standard deviation 9 in $c2$. Make a plot of $c1$ versus $c2$ using the **Graph/Plot** window menu. Using the **Graph/Plot** option, put column $c1$ as the x variable, and $c2$ as y . Using the **Annotations/Title** submenu inside the **Graph/Plot** routine, label your graph with

"YOUR NAME: Normal $\sigma=9$ versus $\sigma=1$ ".

If there is too much clutter on your plot made by points overlapping, use the **Edit Attributes** submenu inside the **Graph/Plot** routine to change the symbol of the plot, or its size so that the individual points can be seen clearly. When all the options mentioned above have been entered, have the routine now plot the graph, and obtain a printout.

Now count how many times the absolute value of the normal with $\sigma=1$ in column $c1$ exceeds the absolute value of the normal with $\sigma=9$ in column in $c2$. To do this, type in the session window the command **let c3=Abs(c1) > Abs(c2)**. This will put a 1 in column $c3$ whenever the absolute value of the entry in $c1$ is larger than the

absolute value of the entry in c2; a zero gets placed in c3 otherwise. (In order for this command to work, the command language needs to be enabled. If it is not already enabled, select **Editor/Enable Command Language**.) Taking the sum of c3 now gives the number of times the absolute value of the normal with standard deviation 1 exceeded the absolute value of the normal with standard deviation 9. We can use **Count/Column Statistic** to count how many ones there are in c3 that is, how many times we obtained such excellencies. You will use this sum to estimate an unknown probability.

Questions for you to answer:

The sum of c3 is the number of times the absolute value of a normal $\sigma=1$ (lets call it X) exceeds the absolute value of a normal $\sigma=9$ (called Y). You will use the information you generated by your experiments to estimate the unknown probability p that $|X|$ exceeds $|Y|$, that is, the value of $p=P(|X|>|Y|)$.

a) From the data on each of your two experiments, give a 95% confidence interval for the unknown p. Do this by hand using a calculator and the normal tables in the text following the formulas given in our text. Do not use the confidence interval procedure in Minitab. To get full credit, please show your calculations.

b) Before looking at the data, you can predict roughly how much shorter (i.e. by what factor) the confidence interval for experiment 2 will be when compared with the interval from experiment one. What is this factor: 1,2,5,10,25, or 625? The widths may not be exactly in this ratio in your data because the two experiments will likely result in different values of the estimate of p. Calculate the ratio of the widths of the intervals obtained from the two experiments, to five decimal places accuracy, and compare the observed ratio to your predicted ratio. Are the observed ratio of widths larger or smaller than those you predicted?

c) On the graph with $n=100$ points find and circle in red all those points which got counted by c3, that is, those points that fell in the region where the values in c1 exceeded the corresponding values in c2. Also draw in red, with a straightedge, the two lines which form the boundary of this region. You must pay attention to the scale printed for your graph: the x-axis and y-axis will get different scales, to fit the spread of c1 and c2 respectively.

d) With your calculator set in **radians**, approximate to 5 decimal places $p=\text{Arctan}(1/9)/(\pi/2)$, where $\pi=3.14159\dots$. For each of your two experiments, does your confidence interval cover this particular value?

e) Assume the value $\text{Arctan}(1/9)/(\pi/2)$ is the true value of p (make sure that on your calculator you use radians and not degrees when calculate this value). If 60 students turn in their homework, each doing two experiments, then how many of the 120 combined experiments would you expect to fail to cover the true value of p ?

2) In Minitab, open the grades.mtw (Note: to open any of the Minitab Data sets, from the File menu choose Open Worksheet, and making sure that file types is set to mtw, locate the desired data set in the Data subdirectory of the Minitab directory). The file consists of three columns, the first two of which give the SAT Verbal and Math scores for 200 college bound students, and the last column their GPA.

a) Find the 95% confidence interval (**Stat>Basic Statistics>1-sample t**) for the population mean Verbal SAT score.

b) What is the 95% confidence interval of the GPA scores for those students whose Math SAT score was higher than their Verbal SAT score?

c) What is the 95% confidence interval of the GPA scores for those students whose Math SAT score was lower than their Verbal SAT score?

To answer parts b) and c), you might want to use the command line editor as you did in the first problem, to enter, for part b) say, a command like **let c4=(c2>c1)*c3**. The value of $(c2>c1)$ is 1 if $c2$ is bigger than $c1$ and 0 otherwise, so the result of this computation is to enter the GPA in column 4 if the math SAT is higher than the verbal SAT, and 0 otherwise.

However, the zeros which get created in this process should not be entered in the computations, since they correspond to students not in the population we are interested in. In order to get rid of these zeros, first sort $c4$ in descending order (using the **Manip>Sort** menu command, sort $c4$ and put the result in $c5$, sort by column $c4$). Now use the mouse to select all entries at the bottom of $c5$ which are 0 and delete them. Now you are ready to do, say, a **Stat>Basic Statistics>1-sample t** to get a 95% confidence interval for the column.
