Teaching Statistics to students with limited language skill using MINITAB

Eunmee Koh
Department of Mathematical Science
Teikyo Loretto Heights University
Denver, Colorado 80236
303)937-4286(office)
303)937-4243(Fax)
e-mail:kohe@essex.hsc.colorado.edu

Many statistical software made student easy in obtaining a statistic they want. For example, if they want to get the sample standard deviation of the data set, then they can simply type "stdev c1" after the prompt for the case using MINITAB statistical software assuming the data are stored in c1(column 1). Whenever we, instructors, introduce a new statistic we give a definition of the statistic and explanation of the meaning of the statistic but many times we find that sudents do not grasp the meaning of that statistic and they just get what computer gives without full understanding of what they are getting. I found that when students are to calculate the sample variance they sometimes answer in negative values, which shows that they do not know what they are getting.   This is even more popular case when they have limited language skill such as foreign students.

 From my experience of teaching classes of mostly Japanese students they are found to have difficulties in understanding what is explained in lecture and also they tend not to listen the explanation. Thye'd rather choose to look at the formula and plug in numbers to get the statistic. To cope with that situation I tried different teaching method when we are to use computer software for the course. When we teach them how to get the statistic we want using Minitab it was found that letting them use first "let" command only and get the statistic step by step, not by one simple command was very helpful. MINITAB statistical software not only gives the statistic we want by typing simple command but also allows us to do calculation column by column . For example if we want to get the deviation from the sample mean then type "let c2=c1-mean(c1)". It will create another column, c2, which contains an array of deviation. Thus at each step they are supposed to print the colum they newly generated and get possible plots to have a feel for what they

obtain so that they can see what the computer does. At the end of step they are to compare their answer to the one by a simple command. After they practiced few times they seem to understand what they are getting. Then they can obtain a statistic by a simple command afterward. Exercising both ways helped students to understand what they are getting step by step. Then later it was much easier for the students to get the meaning of the result they get using one command. We can apply this to when we obtain other statistics such as sample mean, sample variance, and those for the grouped data too.

And also when we explain about The Central Limit Theorem, they can simulate the data from sevral different distributions such as Chi-square, F, Uniform. They can generate 100 sets of different sample size such as n=2,5,10,25, and 100. They can get the histogram of the original data for the comparison. And for each set of data they calculate the sample means and draw the histograms of the sample means. They can easily see what the Central Limit Theorem says- if random samples of n observations are drawn from a nonnormal population with finite mean μ and standard deviation σ , then when n is large, the sampling distribution of the sample mean x is approximately normally distributed, with mean and standard deviation

$$\mu_{\bar{x}} = \mu \qquad\qquad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

and the approximation will become more and more accurate as n becomes large.

Here is an example of how to get a statistic - sample standard deviation in sevral different ways.

*By defining formula:

$$s = \sqrt{\frac{\Sigma(X_i - \bar{X})^2}{n-1}}$$

*By a Minitab command

MTB> stdev c1 (let's say c1 stores our data)

*By using "let" command only:

MTB>let c2=c1-mean(c1) : to get $x_t - \overline{x}$

MTB>let c3=c2**2   : to get $(x_t - \overline{x})^2$

MTB>let k1=sum(c3)   : to get $\Sigma(x_t - \overline{x})^2$

MTB>let k2=k1/(n(c1)-1) : to get $\Sigma\dfrac{(x_t - \overline{x})^2}{n-1}$

MTB>let k3=sqrt(k2)   : to get $\sqrt{\dfrac{\Sigma(x_t - \overline{x})^2}{n-1}}$

MTB>print k3   : to print standard deviation we calculated

or they can do

MTB>let k1=sqrt(sum((c1-mean(c1))**2)/(n(c1)-1))

MTB>print k1

Example

A sample of 25 residential properties from the community and has recorded the following mortgage rate. We would like to get the standard deviation of the data.

 9.2 10.0 11.5 12.0 14.0
14.8 14.5 13.2 11.4 10.5
12.0 13.0 16.5 15.3 11.0
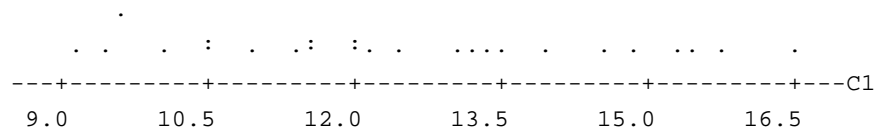15.7 12.5   9.5   13.3 13.5
10.5 11.5 10.5 12.1 15.5

We store the data in C1(column 1).

** Step 1) First let student see the dotplot of the data
MTB > dotplot c1
Character Dotplot

```
                    .
            . .    .   :   .   .:   :. .      ....    .    . .   .. .         .
         ---+---------+---------+---------+---------+---------+---------+---C1
           9.0       10.5      12.0      13.5      15.0      16.5
```
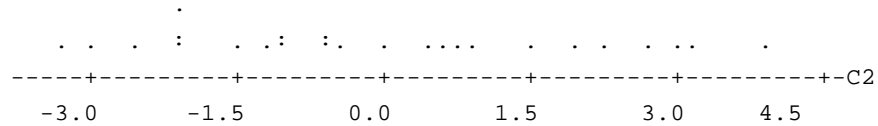** Step 2) Generate column 2(C2) which is $(x_t - \overline{x})$.

MTB > let c2=c1-mean(c1)

Let them get the print of C1 and C2. It is shown below. Then let them have another dotplot of C2.
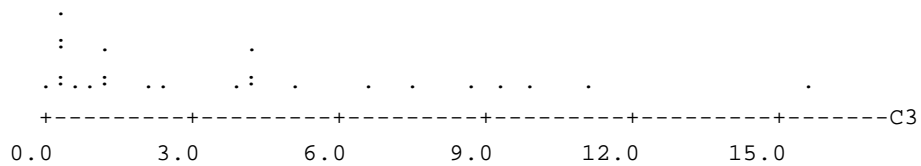
MTB > dotplot c2

Character Dotplot

```
                       .
          . .   .  :    . .:  :.  .  ....   .  .   .  . ..      .
       -----+---------+---------+---------+---------+---------+-C2
          -3.0      -1.5       0.0       1.5       3.0       4.5
```

**Step 3) Generate another column $(X_t - \overline{X})^2$

MTB > let c3=c2**2

Then get the dotplot of C3 and printout of C1-C3.

MTB > dotplot c3

Character Dotplot

```
          .
          :   .           .
       .:..:   ..        .:   .    .   .   . . .    .               .
       +---------+---------+---------+---------+---------+-------C3
        0.0       3.0       6.0       9.0      12.0      15.0
```

MTB > print c1-c3

Row	C1	C2	C3
	X_t	$(X_t - \overline{X})$	$(X_t - \overline{X})^2$
1	9.2	-3.34	11.1556
2	14.8	2.26	5.1076
3	12.0	-0.54	0.2916
4	15.7	3.16	9.9856
5	10.5	-2.04	4.1616
6	10.0	-2.54	6.4516
7	14.5	1.96	3.8416
8	13.0	0.46	0.2116
9	12.5	-0.04	0.0016

10	11.5	-1.04	1.0816
11	11.5	-1.04	1.0816
12	13.2	0.66	0.4356
13	16.5	3.96	15.6816
14	9.5	-3.04	9.2416
15	10.5	-2.04	4.1616
16	12.0	-0.54	0.2916
17	11.4	-1.14	1.2996
18	15.3	2.76	7.6176
19	13.3	0.76	0.5776
20	12.1	-0.44	0.1936
21	14.0	1.46	2.1316
22	10.5	-2.04	4.1616
23	11.0	-1.54	2.3716
24	13.5	0.96	0.9216
25	15.5	2.96	8.7616

**Step 4) Get $\dfrac{\Sigma(X_i - \overline{X})^2}{n-1}$

MTB > let k1=sum(c3)/(n(c1)-1)
MTB > print k1

Data Display

K1   4.21750

** Step 5)   Get $\sqrt{\dfrac{\Sigma(X_i - \overline{X})^2}{n-1}}$

MTB > let k2=sqrt(k1)
MTB > print k2

Data Display

K2   2.05366

**Step 6) We can compare this result with the one we can get by a simple command as follows:

MTB > stdev c1

Column Standard Deviation

Standard deviation of C1   =   2.0537

References

Mendenhall, Reinmuth, and Beaver,"Statistics for Management and Economics"
        7th edition, Duxbury Press
Minitab, Minitab Inc, State College, Pennsylvania