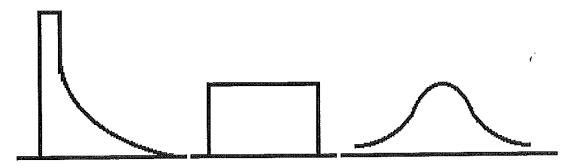
LAB 7 Exploring Data with Summary Statistics

Sometimes the salient features of a histogram can be summarized by a few statistics such as the mean (or average), the median, the standard deviation, or the five-number summary. These quantities allow us to make comparisons of groups and help us to understand the story behind the data.

AT THE COMPUTER

Let's examine how these quantities relate to the histogram that they summarize using data that was gathered in September of 1999 when students responded to a survey in a large statistics class. Use your software to open the data file called 'Au99survey'.

1. Three of the variables in the survey were birthday month (1 for January, 2 for February, etc.), the amount spent on textbooks (in dollars), and the number of sodas consumed over the previous week. Without looking at histograms of the data, guess which of the three rough sketches below will approximately describe the histograms of the three variables. Explain your guesses.



2. Create the histograms for each of these three variables. How do the histograms compare with what you expected?

3. Look at the three histograms for birthday month, textbook costs, and sodas consumed. Guess the value of the median, the mean, and the standard deviation for each variable.

Variable	My guess at the median	My guess at the mean	My guess at the standard deviation		
month					
textbooks					
sodas					

To see how well you did at judging the descriptive statistics in question 3, use your software to compute the actual values.

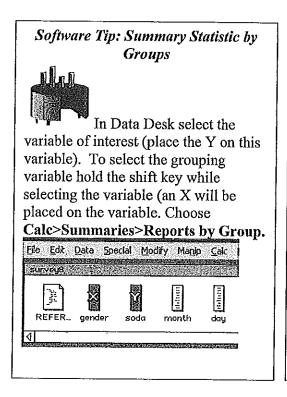
Software	e Tip: Sum	mary Statistics	•
Aschara	In Data De	esk select the	
variable of	interest an	d then choose	
Calc>Sum	maries>R	eports. You m	ay
		ry statistics to	•
produce by		·y	
	_	diama Calcat	
	_	otions>Select	
Summary	_	Julia-Delect	
Summary Summary Options	Statistics 2	<u>8</u>	
Summary Summary Dillions Certes	Statistics	abe 381 ∏General 3	
Summary Summary Cortest.	Statistics	<u>8</u>	
Summary Summary Deltons Certait. C'Hear	Statistics	Gereral XI Group Names P: Conta	
Summary Sunnay Dallots Certer P (Keen P Model CIMUReroe	Statistics Speed P Sun Dev. C Range	Gereral Ger	
Summary Content Conten	Statistics Speed Fr Shin Dev. Frage Volumes	Geretal Group Hornes D' Courts Lida Cates Hon Nynarit	
Summary Cortex Cortex Citeen Magnetic Magn	Statistics Speed 7 Stan Der/ 19 Range 2 Yentroes 10 Metroportie Range 11 Standed Ente	Gerard Group Hornes 7: Courts 1 Indu Cates Non Mynerit Sun Sun of Squark	
Summary Cortex Circles Circ	Statistics Speed F7 Stan Der/ Flange Ventres Standard Eng Pop Ston Der Housel	Gerstal Group Harnes C Conts Tidal Cases Han Mynest Sura G Square	
Summary Summary Cyllons Certex Certex Figure Mod Quarte Percente Figure Percente Figure Percente Figure Percente Figure Percente Figure Figure Figure Percente Figure Figure	Statistics Speed F Sun Der/ Grange Quinces Gipter Durch Brayer Gistarded Era Frank Store Hower Hower King	Gerocal Gerocal Gerocal Gerocal Float Cares Float Cares Float Square	
Summary Content Corten	Statistics Speed Fr Sun Der/ Rhage Ventrus Starterd Enr Prop Ston Der/ Prop Ston Der/ Starterd Enr Prop Ston Der/ Starterd Enr Freschieder Katherine	Greed Group Herrer C Conte Total Cases Hos Namer Sure of Squees OK Cancel	
Summary Summary Cyllons Certex Certex Figure Mod Quarte Percente Figure Percente Figure Percente Figure Percente Figure Percente Figure Figure Figure Percente Figure Figure	Statistics Speed P. Stan Dev. Phays Lystros Lystros Speed Prys Sion Dav Prys Sion Dav Kort Skeed Kort Sin Bayed	Gereral Gereral Gereral V. Courts Clade Cases Hon Nymer Surfal Squar Courts OK	

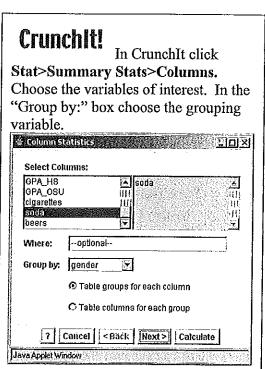
Crunchit! In Crunchit click
Stat>Summary Stats>Columns.
Choose the variables of interest and
click Next. You can choose the
statistics you wish from the list and
then click Calculate.
⊈, Column Statistics
Statistics:
N
Percentiles:optional
☐ Store output in data table
7 Cancel Seack Next > Calculate
Java Applet Wridow

4. What are the actual values for the median, mean, and standard deviation of the three variables?

Variable	Actual median	Actual mean	Actual standard deviation
month			
textbooks			
sodas			

- 5. Compare the actual values with your guesses. Which variable's summary statistics were most difficult to judge? Explain.
- 6. Which measure of location (mean or median) would be the most appropriate in each of the following situations? Explain.
 - a. You want to know how many sodas a typical student in the class drinks per week.
 - b. You want to know how much soda is consumed in a week by the class as a whole.





7. Do males drink more soda than females? One of the big advantages of summary statistics is that they allow us to compare groups such as males and females. Record the following summary statistics for the number of soda's consumed by males and females.

Gender	Mean	Median	1st Quartile	3rd Quartile	Min	Max
Males						
Females						

8. Based on your summary statistics, do you feel males or females drink more soda?

9. Boxplots are useful for making a graphical comparison between groups in a population. Use Boxplots to compare the High School GPA for the men and women in the class. Sketch the comparative Boxplots below.

Software Tip: Comparative Boxplots

In Data Desk select the variable of interest (place the Y on this variable). To select the grouping variable hold the shift key while selecting the variable (an X will be placed on the variable. Choose Plot>Boxplot y by x.

CrunchIt!

In CrunchIt click Graphics>Boxplot. Choose the variables of interest. In the "Group by:" box choose the grouping variable. Select Next and click Use fences to identify outliers.