

Stat 218 - Day 2 Graphical Displays

Two first steps in analyzing data are to:

- construct **graphical displays**
- calculate **numerical summaries**.

Determining which types of displays and summaries depends largely on whether the variable involved is categorical or quantitative.

With categorical variables the most common graphical display is a **bar graph**, and the most common numerical summaries are **proportions**.

Example: Class data

- (a) For one of the three variables collected yesterday, calculate the proportion of students answering in each category and construct a bar graph. Also write a sentence or two revealing what the data reveal.

We will make frequent use of the statistical software package Minitab for analyzing data, particularly with quantitative variables.

Minitab notes:

- Data window
 - Variables in columns, observational units in rows
 - Not a spreadsheet (e.g., no automatic updating)
- Session window
 - Can type commands into Session window or use menus
 - Can create new variables with `let` command at `MTB>` prompt
 - If the `MTB>` prompt does not appear
 - Click on Session window, choose `Editor> Enable commands`
 - Select `Tools> Options> Session window> Submitting commands`, choose “enable”
- Saving work
 - Can open/save worksheet (data only; `mtw`) or project (data and graphs and output; `mpj`)
 - Can copy/paste graphs into Word document (label well; “crop” out unnecessary parts)

Common graphical displays with quantitative data: **dotplots** and **histograms**.

Features to look for include: **shape, center, spread, peaks/gaps, outliers**

Common shapes: **symmetric, skewed** to the left, skewed to the right

Advice: Always relate comments to the **context** of the data!

Example: Day hikes

We will use Minitab to analyze data from the book *Day Hikes in San Luis Obispo County*, saved in the Minitab worksheet `dayhikes.mtw`.

(a) What are the observational units here? What are the variables?

(b) Examine a dotplot (Graph> Dotplot, or MTB> dotp c3) and histogram (Graph> Histogram, or MTB> hist c3) of the distribution of the hike *distances*.

(c) Comment on each of the five features for the distribution of hike distances.

Example: Rowers' weights

The data in the Minitab worksheet `rowers04.mtw` are the weights of the members of the U.S. men's Olympic rowing team in 2004.

(a) Create a dotplot and histogram of the distribution of weights.

(b) Double-click on the axis of the histogram. Select the "binning" tab and click on "number of intervals." Change the number of intervals to 16, and comment on how this changes the histogram. Then change the number of intervals to 4, and comment on how this changes the histogram.

(c) What shape does the distribution of weights have?

(d) Suggest an explanation for the clusters/gaps that are apparent in the distribution of weights.

Example: Exam scores

The following graph displays scores of 62 students who took the first exam in a statistics course:

4	9
5	578
6	344568
7	022444456677799
8	0011122233345577788999
9	000001112223345

This display is called a **stem-and-leaf plot**. The stems appear to the left of the vertical bar; in this case they are the tens digit of the score. The leaves (ones digits here) appear to the right.

(a) How many students received a score of 77 on the exam? Circle their leaves on the plot.

(b) How many and what proportion of students scored 90 or higher?

(c) How many and what proportion of students scored below 70?

(d) Which score appears most often?

(e) There are only two values between 75 and 95 that are not anyone's score. What are they?

(f) Comment on relative advantages and disadvantages of histograms and stem-and-leaf plots.

Example: British rulers' reigns

The following are the lengths of reign (in years) of British rulers since 1066:

William I	21	Edward III	50	Edward VI	6	George I	13
William II	13	Richard II	22	Mary I	5	George II	33
Henry I	35	Henry IV	13	Elizabeth I	44	George III	59
Stephen	19	Henry V	9	James I	22	George IV	10
Henry II	35	Henry VI	39	Charles I	24	William IV	7
Richard I	10	Edward IV	22	Charles II	25	Victoria	63
John	17	Edward V	0	James II	3	Edward VII	9
Henry III	56	Richard III	2	William III	13	George V	25
Edward I	35	Henry VII	24	Mary II	6	Edward VIII	1
Edward II	20	Henry VIII	38	Anne	12	George VI	15

- (a) Who ruled the longest? How long was his/her reign?
- (b) What is the shortest reign? Who ruled it? What do you think this value really means?
- (c) Who is the current monarch? Why do you think her length of reign not included here?
- (d) Create a stem-and-leaf plot of these lengths of reign by filling in the following graph. (I suggest that you take the rulers in order, beginning with William I, putting each leaf on the row with the appropriate stem.)
- | | |
|---|--|
| 0 | |
| 1 | |
| 2 | |
| 3 | |
| 4 | |
| 5 | |
| 6 | |
- (e) Determine a value such that about half of these 40 rulers have reigned longer and half have reigned less long.
- (f) Compare the shape of the distribution of reign lengths with the shape of the distribution of exam scores. What difference is there between these shapes?