**Before beginning this packet – please check your e-mail from Mrs. Cartin for important links to data needed to complete some of these activities and videos to help you if you are having trouble on an activity!!!!**

Also, this packet is scanned to be printed double sided.  There may be blank pages intentionally scanned so that the formatting will be correct.

# AP Statistics Summer Assignment

Welcome to AP Statistics!! The purpose of this assignment is to get a jump on the school year and make you comfortable exploring and analyzing data. It will take several hours to properly complete, so please allow yourself a few days to do it and don't wait until the last minute! I created it with the feedback of previous AP Stats students right after they took the AP exam, so the information in this packet is information in which they felt would best prepare future AP Stats students. Successful completion of this assignment will help us "hit the ground running" and ensure adequate time in 2nd semester for proper review leading up to the AP Exam (and hopefully fewer assignments over Fall/Thanksgiving/Minimester/Spring Breaks).

This Summer Assignment will consist of 3 parts: Videos, Chapter 1 content, and Calculator Practice.

This assignment will be due on the 1st day of class. It consists of 3 parts

1. Videos: Please go to www.apstatsguy.com and watch the videos named "Summer Video One", "Summer Video Two", "Summer Video Three", and "Summer Video Four". You will need to take notes or write a BRIEF summary for these videos. These will be due on the first day of school and will be worth 20% of your summer assignment grade.
2. Chapter 1: These activities are covering Chapter 1 of your textbook and StatsMedic. You do not need your textbook to complete them. However, you will want to go to www.statsmedic.com, create an account, and you will have access to the electronic copies of the worksheets and answer keys for all the classroom activities this year. You can find these under https://www.statsmedic.com/ced-apstats-unit1. Please note that we will be doing the Parkinson's Disease activity on the first day of school, so please, please, please NO SPOILERS!!!! I have shared the data that has been collected in previous years in a google sheet so that you can complete these activities with data from William & Reed students (some of you may remember contributing to this data collection)!!! Please check your email for a link to a google drive folder that contains this data as well as videos by the StatsMedic creators, Luke Wilcox and Lindsay Gallas. These daily activities along with the quizzes in this packet will be due on the first day of school and will be worth the remaining 80% of your summer assignment grade.
3. The third part of the summer assignment is not for a grade but is HIGHLY recommended. You need to become familiar with your calculator. A TI-84 is preferred and is an investment you will not regret. This calculator will be your BFF in AP Statistics. Please take the time to go though the calculator practice in a separate document.

Also, there are a few specific supplies you will need for class on Day 1:
- Graphing calculator (TI-84 is preferred).
- One-inch (or bigger) 3 ring binder
- Colored pens/pencils


I will check emails periodically throughout the summer. If you have questions, please email, but beware that it may take several days for me to get back to you. jcartin@williamandreed.com

Have a great summer!!!

Mrs. Cartin

# How are your favorite classes related? — Unit 1, Day 3

Is your favorite elective class associated with your favorite core class? Collect class data to see if there is a relationship.

1. Which of the following is your favorite elective class? You must choose only one and mark your choice on the board.

| Art | Music | Physical Education | Foreign Language | Technology |
|-----|-------|--------------------|--------------------|------------|
|     |       |                    |                    |            |

2. Identify the individuals and variable?

3. Is the variable categorical or quantitative?

4. Go to stapplet.com to enter the class data. Make a bar graph and a pie chart. Sketch them below.

5. Sometimes it is helpful to investigate more than one variable. Come to the board and put a tally mark where you belong.

**Core Class**

|          |              | Math | English |
|----------|--------------|------|---------|
|          | Art          |      |         |
|          | Music        |      |         |
| Elective | P.E.         |      |         |
|          | Foreign Lang.|      |         |
|          | Tech.        |      |         |

**Find each of the following:**

% of all students who chose P.E.:

% of all students who chose Math and chose Art:

% of the students who prefer math that chose Tech.

**STATS MEDIC**

6. How many variables does the table have? Are the variables categorical or quantitative?

7. Which variable would best explain or predict the other variable?

8. Go to stapplet.com and enter the data. Make a side-by-side bar graph and a segmented bar graph. Sketch them below.

9. How do the bars in the side-by-side-bar graph relate to the bars in the segmented bar graph?

10. Is there an association between favorite core subject and favorite elective? If so, describe it.

11. If there was not an association between favorite core subject and favorite elective, what would the graphs look like? Explain.

Name: _____ Hour: _____ Date: _____

# Analyzing Categorical Data

Important Ideas:

## Check Your Understanding:

1. The following graph was displayed by a national news organization. Explain why the graph may be misleading, and sketch a corrected version of the graph.



**SOUTHWEST BORDER APPREHENSIONS**
OCTOBER - APRIL

195,000
185,000
175,000
165,000
155,000

165,244 (2011)   170,223 (2012)   192,298 (2013)

2. A real estate agent is collecting data on the number of houses built in his town's three neighborhoods during three different decades. The table below gives information.

|  | 1960s | 1970s | 1980s |
|---|---|---|---|
| Shady Lane | 40 | 30 | 10 |
| Oakcrest | 60 | 15 | 5 |
| Pinewood Estates | 0 | 45 | 15 |

a. What proportion of the houses shown were built in Pinewood Estates?

b. Find the distribution of Decade Built for the houses in this town using relative frequencies.

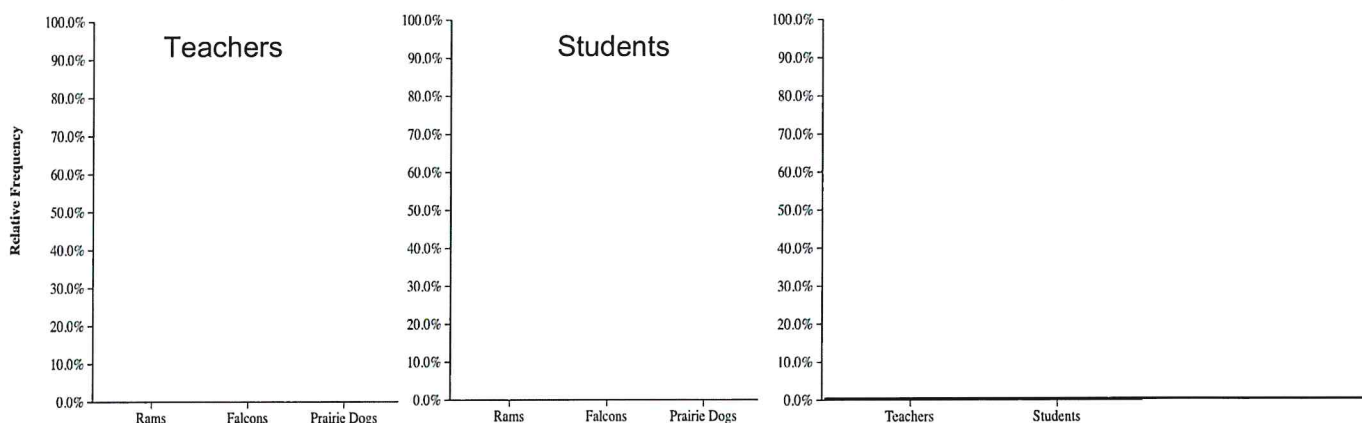c. What percent of homes were built in Oakcrest and in the 1960s?

STATS MEDIC

# What will be the EK mascot? – Unit 1, Day 4

When the high school was built in 1969, the school needed to pick a mascot. The principal decided to have the students and teachers vote between three choices: rams, falcons, or prairie dogs. He took a random sample of students and a random sample of teachers. The results of the surveys are given in the table.

|  | Rams | Falcons | Prairie Dogs |
|---|---|---|---|
| Teachers | 80% | 10% | 10% |
| Students | 30% | 60% | 10% |

1. Create two bar graphs below to display the results. Use three different colors for the bars.

2. Complete the third graph by taking each bar from the teacher sample and stacking them. Use the colors to mark each section. Do the same for the student sample.



3. According to your displays, which mascot appears to have the most support? Explain.

4. Upon hearing the results of the surveys, the students argued that the decision was incorrect because 100 teachers had been surveyed and 500 students had been surveyed. Use this information to fill in the table below with the number of responses.

|  | Rams | Falcons | Prairie Dogs |
|---|---|---|---|
| Teachers |  |  |  |
| Students |  |  |  |

5. How many times more students were sampled than teachers? _____. How can you update the third graph in #1 to take into account the sample size? Adjust your graph.

6. What should they make the EK mascot? Explain.

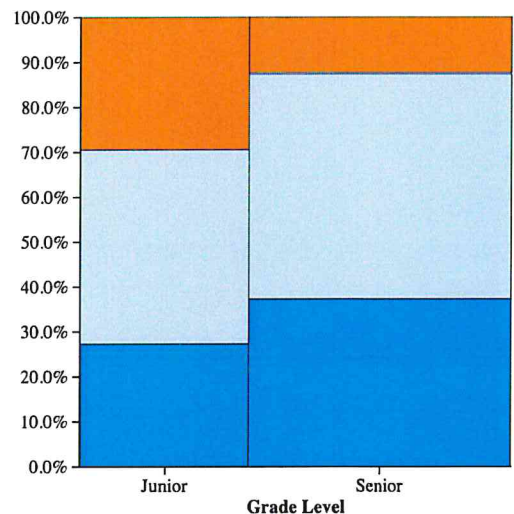# Representing Two Categorical Variables

Important Ideas:

## Check Your Understanding:

The following table gives the result of a random sample of upper level students at Rocky Vista University (the Fighting Prairie Dogs!), along with a mosaic plot.

| Employment Status | Grade Level | |
|---|---|---|
| | Junior | Senior |
| Currently working | 14 | 30 |
| Not working but have had a job | 22 | 40 |
| Never had a job | 15 | 10 |

**Employment status**

- Never had a job
- Not working but did in the past
- Currently working



a. Calculate the proportion of Juniors that are currently working, not working but have had a job, and never had a job.

b. Calculate the proportion of Seniors that are currently working, not working but have had a job, and never had a job.

c. Write a few sentences summarizing what the display in part (a) reveals about the association between grade level and job experience for the students in the sample.

**STATS MEDIC**

Name: _____ Hour: _____ Date: _____

# How many pairs of shoes do you own? – Unit 1, Day 5

1. How many pairs of shoes do you own? Record your answer on the board.

2. Is "Number of pairs of shoes" a categorical or quantitative variable?

3. Enter the data at www.stapplet.com.  Make a dotplot, stemplot, and histogram and sketch each below.

4. List the mean and median of the distribution. Which value do you think is a more appropriate measure of center? Explain.

5. Describe the distribution of the number of pairs of shoes for your class.

    Shape:

    Outliers:

    Center:

    Variability (spread):

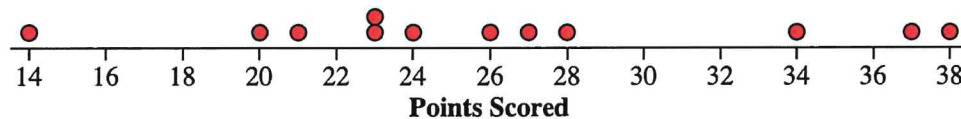6. Which of the three types of display do you prefer?  Why?

STATS MEDIC

# Displaying Quantitative Data

Important Ideas:

## Check Your Understanding:

1. Mr. Wilcox is a huge fan of University of Michigan football. His favorite season was the 199 7season (a perfect season!). The dotplot shows the number of points scored by the U of M team in the 12 games that season.
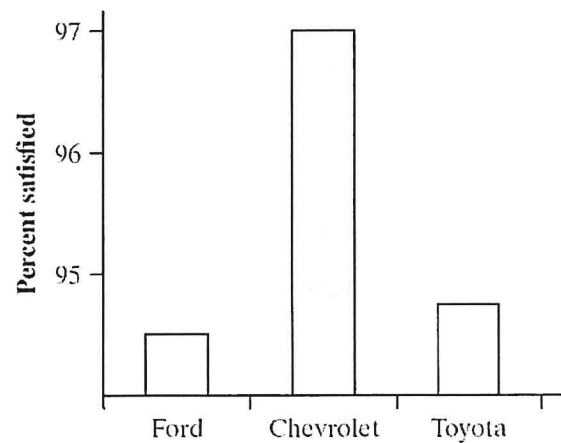


**Points Scored**

(a) Use the dotplot to create a stemplot of the distribution.

(b) Describe the shape of the distribution.

(c) Are there any potential outliers? Why?

(d) What measure of center is most appropriate to describe the distribution? Explain.

STATS MEDIC

1. Mr. Molesky observes a group of monkeys for 24 hours to learn about their behavior. He records how long they slept, how many bananas they ate, gender, age, and the specific breed of monkey.

   (a) What are the individuals in this data set?

   (b) Identify the variables that were recorded, and indicate whether each one is categorical, quantitative (discrete), or quantitative (continuous).

2. The following bar graph gives the percent of owners of three brands of trucks who are satisfied with their truck. What is wrong with the way information is presented in this graph?
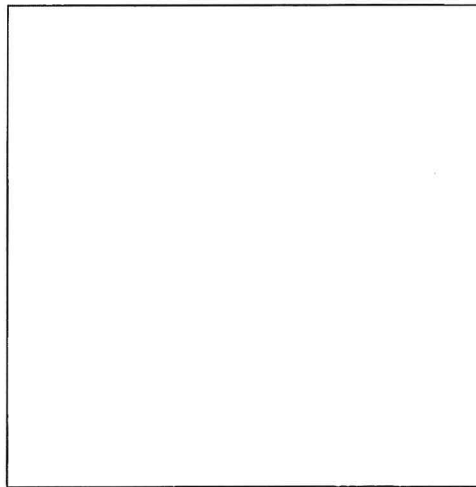
**3.** A research study asked children which of four different emotions they associated with the color red. The response and gender of each child are given in the following table.

|  | Male | Female |
|---|---|---|
| Joy | 28 | 61 |
| Happiness | 20 | 25 |
| Love | 40 | 80 |
| Anger | 18 | 60 |

(a) Find the distribution of emotion for each gender using relative frequencies.

(b) Make a segmented bar graph to compare the distributions in part (a).

(c) Describe what the graph in (b) reveals about the association between gender and emotion for the students in the sample.
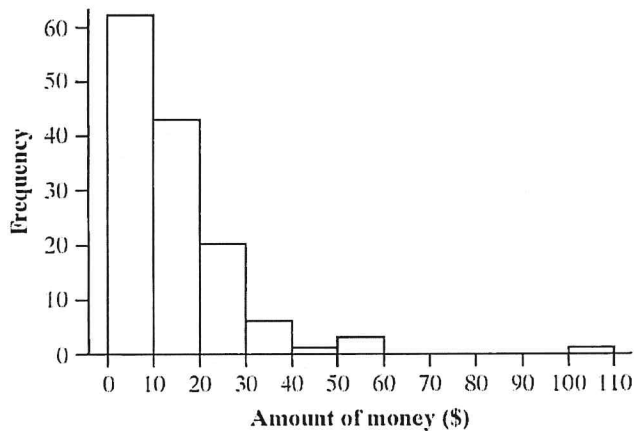
1. A random sample of 20 employees at a large company was selected. Here are the salaries (in thousands of dollars) for these employees during one year.

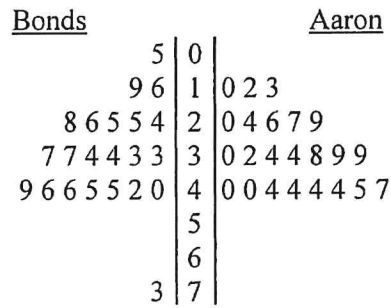| 28 | 31 | 34 | 35 | 37 | 41 | 42 | 42 | 42 | 47 |
| 49 | 51 | 52 | 52 | 60 | 61 | 67 | 72 | 75 | 77 |

(a) Make a dotplot of these data.

(b) What percentage of the sample of employees have a salary of at least $60,000?

2. Professor Windley teaches a statistics class with 136 students. On the first day of class he records how much money each student has in his or her possession (in dollars) during the first class of the semester. The histogram displays the data. Describe the distribution.

**3.** On August 7, 2007 Barry Bonds hit his 756th home run, breaking the all-time career home run record, formerly held by Hank Aaron. Does that make Bonds a better home run hitter than Aaron? Let's compare their annual home run production over their entire careers. A side-by-side stemplot is shown below.

Number of Home Runs Per Year

<u>Bonds</u>                       <u>Aaron</u>

```
         5 | 0 |
        9 6 | 1 | 0 2 3
      8 6 5 5 4 | 2 | 0 4 6 7 9
    7 7 4 4 3 3 | 3 | 0 2 4 4 8 9 9
  9 6 6 5 5 2 0 | 4 | 0 0 4 4 4 4 5 7
              | 5 |
              | 6 |
            3 | 7 |
```

Key: $1|4 = 14$ home runs

Use the plot to write a few sentences discussing the similarities and differences in the distributions of home runs per year for Bonds and Aaron.

Name: _____  Hour: _____  Date: _____

# How many colleges are you applying to? – Unit 1, Day 7

| How many different colleges is your group of 4 applying to? Find the total number of colleges for your whole group. |
|---|

1. Record the data for the class here.


2. Calculate the mean and median for the set of data. Compare them.


3. What is the range of the data?


**Finding Standard Deviation**
4. Finding range is helpful but it does not tell us how spread out the data is between the minimum and maximum. How can we find the *average distance of the values from the mean?*


a. Complete the table.


b. The average you calculated is the average of the **squared distances** from the mean. How do we use this to find the **average distance from the mean**? Find it.

| Value | Distance from mean | (Distance from mean)$^2$ |
|---|---|---|
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  | Total: |  |
| Average (Distance from mean)$^2$: |  |  |

5. Go to stapplet.com. Enter the classroom data and find the summary statistics. Verify our work. How does it compare?



6. We forgot to add one group that applied to 40 colleges! Add this group to the data set. Calculate the new mean, median and standard deviation using the applet. How does it compare to the original measures? Why do you think this is?
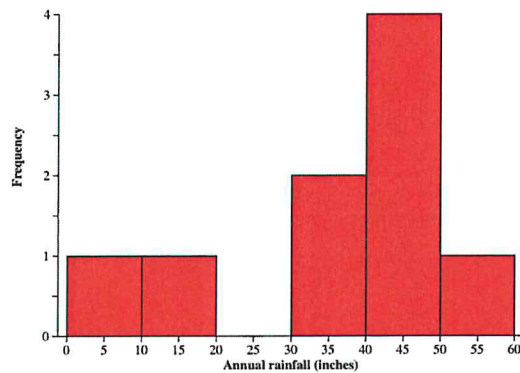
# Describing Quantitative Data

Important Ideas:

## Check Your Understanding:

A researcher is interested in how much annual rainfall is typical in the United States. She takes a random sample of 9 cities in the U.S. and records the annual rainfall, in inches.

| |
|---|
| 8.2 |
| 10.3 |
| 33.5 |
| 39.1 |
| 40.5 |
| 41.9 |
| 42.4 |
| 44.9 |
| 53.7 |



1. Calculate the mean annual rainfall for these cities.

2. Find the median annual rainfall for these cities.

3. Would you use the mean or the median to summarize the typical annual rainfall for a U.S. city? Explain.

4. The standard deviation of the annual rainfall for these 9 cities is 15.52 inches. Interpret this value.
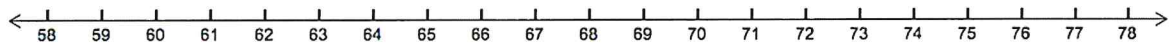
# Where Do I Stand?
## Unit 1, Day 8

How does my height compare with the other AP Stats students in my class? In order to answer this question, Ashmita, a student in 4th hour AP Stats, recorded the heights of everyone in her class. The heights (in inches) were:

68 72 61 62 63 63 64 64 59 62 61 60 65 62 57 77 62 71 65 62 70

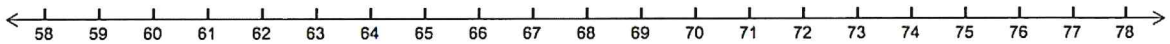1. Create a dotplot to display the class distribution of heights.



2. What is the median height? Describe how you found it.

3. What is $Q_1$ and $Q_3$? Describe how you found them.

4. Record the following values and then use them to make a boxplot.

Minimum:          $Q_1$:          Median:          $Q_3$:          Maximum:



4. The **interquartile range** (or *IQR*) is defined as $Q_3 - Q_1$. Find the *IQR*. Where do you see the *IQR* in the boxplot?

5. An **outlier** is a data value that is way too small or way too big (using the rules below). Are there any outliers? Show your work.

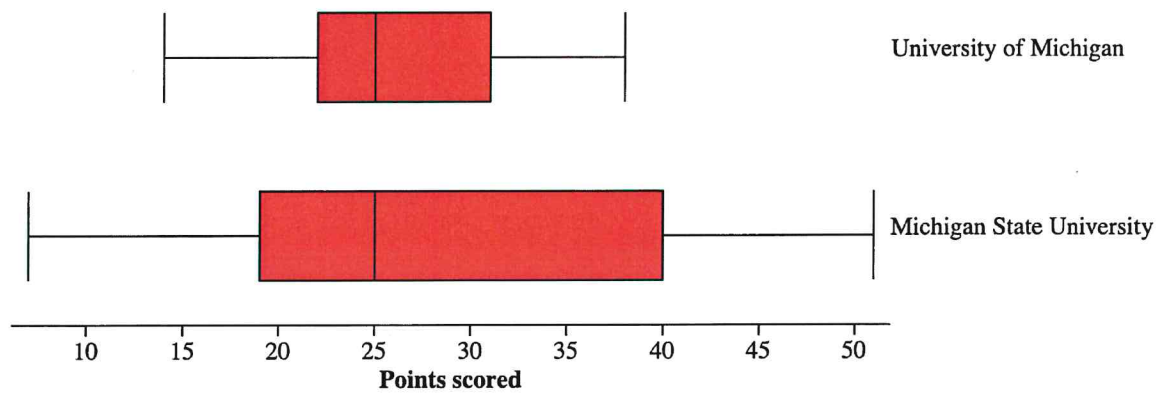Way too small $< Q_1 - 1.5IQR$

Way too big $> Q_3 + 1.5IQR$

6. Ashmita is 63 inches tall. How does her height compare with the other AP Stats students in her class?

**STATS MEDIC**

# Describing Quantitative Data
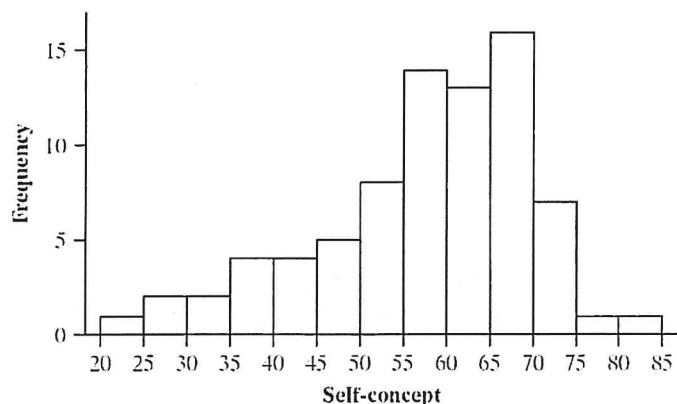
Important Ideas:

## Check Your Understanding:

Mr. Wilcox is a huge fan of University of Michigan football. His favorite season was the 1997 season (a perfect season!). Here parallel boxplots of the points scored by the 1997 University of Michigan football team and the archrival Michigan State University football team. Write a few sentences comparing the distributions.



University of Michigan

Michigan State University

Points scored

1.  A group of 78 third-grade students in a Midwestern elementary school took a "self-concept" test that measured how well they felt about themselves. Higher scores indicate more positive self-concepts. The lowest self-concept scores were 20, 26, 27, 31, and 32.  A histogram and some summary statistics from Minitab for these students' self-concept scores are given below.



| Variable | $n$ | Mean | StDev | Minimum | $Q_1$ | Median | $Q_3$ | Maximum |
|---|---|---|---|---|---|---|---|---|
| SelfConc | 78 | 56.85 | 12.35 | 20 | 50 | 59 | 65 | 84 |

(a) Are there any outliers?  Justify your answer.

(b) Draw a boxplot of this distribution.

(c) One of the students had a self-concept score of 32. If this score had been accidently recorded as 22, what effect would this have had on the value of the mean and the median? Justify your answer.

2. Five students reported the amount of time (in minutes) they spent studying for an AP Statistics test the night before the test. The mean of the reported times is 45 minutes and the standard deviation is 10 minutes.

(a) Interpret the standard deviation in context.

(b) A 6<sup>th</sup> student reported that they studied for 50 minutes. How would the addition of this student to the data set affect the value of the mean and the standard deviation? Explain your answers.

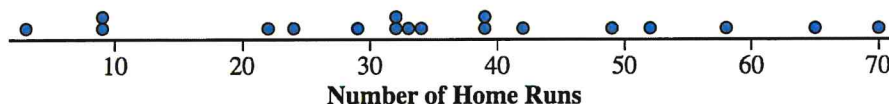# Who is Baseball's Greatest Home Run Hitter?

## VS (unit 1, Day 10)

Barry Bonds broke Mark McGwire's record when he hit 73 home runs in the 2001 season. How does this accomplishment fit with the rest of Bond's career? Here are Bond's home run counts for the years 1986 to 2007.

| 16 | 25 | 24 | 19 | 33 | 25 | 34 | 46 | 37 | 34 | 49 | 73 | 46 | 45 | 45 | 5 | 26 | 28 |
|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|---|----|----|

1. What display did your group get assigned? _____

2. Create the display on the whiteboard and bring it to the front of the room.

3. Describe the distribution.

4. What are the advantages and disadvantages of your type of display?

5. Below is this distribution of the number of home runs per season for Mark McGwire. Compare this distribution to the one for Barry Bonds.



**Number of Home Runs**

| n | mean | SD | min | $Q_1$ | med | $Q_3$ | max |
|----|--------|-------|-----|-----|------|-----|-----|
| 18 | 35.611 | 18.69 | 3 | 24 | 33.5 | 49 | 70 |

STATS MEDIC

# AP Stats: Review of Calculator Commands (TI-84)

| What are we looking for? | Keystrokes | Notes |
|---|---|---|
| Mean<br>Standard Deviation<br>Sample Size<br>5 Number Summary | Stat → Edit → Enter data into $L_1$<br>Stat → Calc → 1-Var Stats | If you put your data in a list other than $L_1$, you must specify the list before you run 1-Var Stats |
| Area under the z-distribution | $2^{nd}$ → VARS → normalcdf (-99, UB)<br>$2^{nd}$ → VARS → normalcdf (LB, 99)<br>$2^{nd}$ → VARS → normalcdf (LB, UB) | LB – Lower Bound<br>UB – Upper Bound<br>LB/UB – use z-scores, leave $\mu = 0$, $\sigma = 1$ |
| Area under the t-distribution | $2^{nd}$ → VARS → tcdf(LB, UB, df) | |
| Area in tail of $\chi^2$–distribution | $2^{nd}$ → VARS → $\chi^2$–cdf(LB, UB, df) | |
| z-score corresponding to area in tail | $2^{nd}$ → VARS → invNorm( | Area must be a percentile |
| t-score corresponding to area in tail | $2^{nd}$ → VARS → invT(area of tail, df) | Area must be a percentile |
| $\chi^2$–score corresponding to area in tail | $2^{nd}$ → VARS → $\chi^2$–cdf | Area must be a percentile |
| Linear Regression<br>Slope & y-intercept<br>Correlation Coefficient<br>Coefficient of Determination | Stat → Edit → Enter x-values into $L_1$<br>Stat → Edit → Enter y-values into $L_2$<br>Stat → Calc → LinReg(a+bx) | Remember to turn diagnostics on for $r$ and $r^2$. If you enter data into lists other than $L_1$ and $L_2$, you must specify |
| Turn Diagnostics ON | $2^{nd}$ → 0 (catalog) → DiagnosticOn | Use the alpha keys for easier navigation of the catalog |
| Graph Scatterplot and Regression Line | x (independent) variable in $L_1$<br>y (dependent) variable in $L_2$<br>LinReg(a+bx)<br>Under StatPlot, select scatterplot<br>Y= → Vars → EQ → RegEQ → Zoom #9 | Make sure StatPlot is turned on |
| Graph Residual Plots | x (independent) variable in $L_1$<br>y (dependent) variable in $L_2$<br>Stat → Calc → LinReg(a+bx)<br>Highlight $L_3$ → $2^{nd}$ → Stat → RESID<br>Under StatPlot, select $L_1$ and $L_3$ (or $2^{nd}$ → Stat → RESID instead of $L_3$) | If graphing by hand, it is much easier to look at $L_1$ and $L_3$ to get ordered pairs |
| Normal Probability Plot | StatPlot → select the last graph | If plot is linear, data is approximately normal |
| Probability – Binomial Distribution | $2^{nd}$ → Vars → binompdf(n, p, x)<br>$2^{nd}$ → Vars → binomcdf (n, p, x) | When you use cdf, you get the sum of the probabilities up to & including x. If you want to find the probability of something greater than x, use 1 – binomcdf |
| Probability – Geometric Distribution | $2^{nd}$ → Vars → geometpdf (p, n)<br>$2^{nd}$ → Vars → geometcdf (p, n) | Use cdf to find the sum of the probabilities |

| | | |
|---|---|---|
| 1-Proportion Confidence Interval | Stat → Tests → 1-PropZInt | Go to pg. 23-24 for conditions |
| 2-Proportion Confidence Interval | Stat → Tests → 2-PropZInt | Go to pg. 23-24 for conditions |
| 1-Proportion Significance Test | Stat → Tests → 1-PropZTest | Go to pg. 23-24 for conditions |
| 2-Proportion Significance Test | Stat → Tests → 2-PropZTest | Go to pg. 23-24 for conditions |
| Confidence Interval for Means | Stat → Tests → TInterval | Go to pg. 23-24 for conditions |
| T-Test for Means | Stat → Tests → T-Test | Go to pg. 23-24 for conditions |
| Chi-Square – Goodness of Fit | Stat → Tests → $X^2$GOF-Test | Go to pg. 23-24 for conditions |
| Chi-Square – Homogeneity | Stat → Tests → $X^2$-Test | Go to pg. 23-24 for conditions |
| Chi-Square – Independence | Stat → Tests → $X^2$-Test | Go to pg. 23-24 for conditions |
| Confidence Intervals for Regression | Independent: L1<br>Dependent: L2<br>Stat → Tests → LinRegTInt | Go to pg. 23-24 for conditions |
| Tests for Regression | Stat → Tests → LinRegTTest | Go to pg. 23-24 for conditions |