

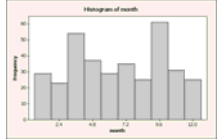
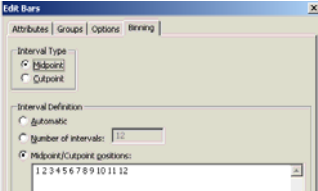
**140,000-acre blaze continues to grow; \$21 million has been spent battling flames**  
By Ari B. Bloomekatz  
12:22 p.m.

**Chapter 4 and 5 – Quantitative Data II**

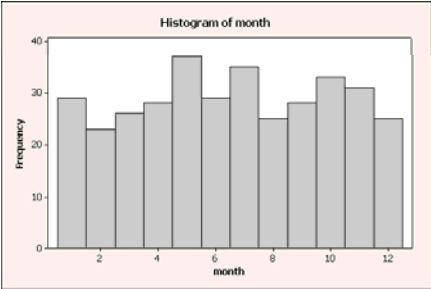
2 WORKSHEETS

### Quick Interlude: Changing Bin Width

- Features of a distribution can appear to change as you adjust the bin size
- <http://www.canyons.edu/faculty/morrowa/140/datasets>
  - Math 140 Survey Results – Month
  - Plot a histogram of month
- Graph the Histogram
  - Double click on the bars
  - Click on the Binning tab
  - Change Midpoints
    - 1 2 3 4 5 6 7 8 9 10 11 12
    - To put a new bin, centered at each month

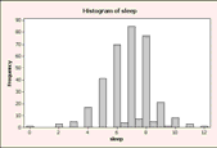



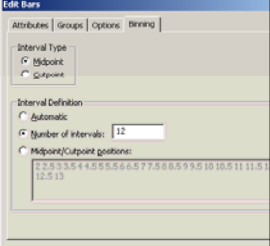
### Just Checking



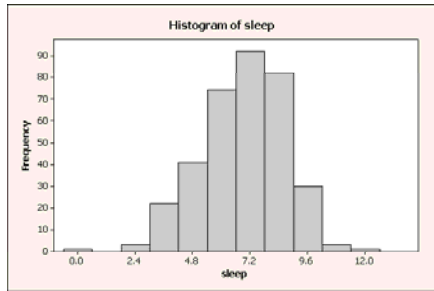
- Now we can see a trend in months of birth...
- We can use this feature to 'fix' the bins Minitab assigns.

### Changing Bin Width: Smoothing

- Why is the histogram for sleep so jagged?
 
- Smooth it out with fewer bins.
- Plot the histogram
  - Double click the bars
  - Click on the Binning tab
  - Change the Number of Intervals
    - Cut it in half to start



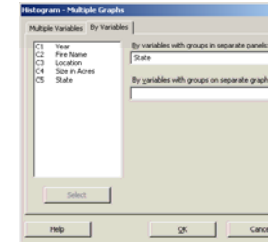
### Just Checking...



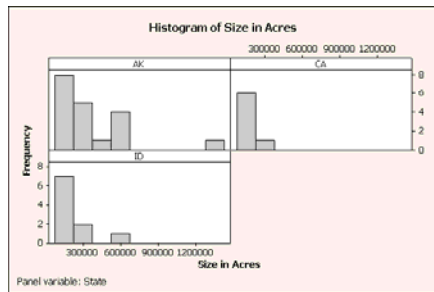
- Preserve the main features (modes/gaps), smooth the ones that are "too jagged".
- Be careful of oversmoothing...

### Comparing Histograms Graph Side-by-Side Histograms

- Talking Points
  - Shape
  - Center
  - Spread
- Minitab > Graph > Histogram > Simple
  - Select Graph Variables
  - Settings under "Multiple Graphs"
    - "By Variables" tab
    - By Variables with groups in separate panels
- <http://www.canyons.edu/faculty/morrowa/140/datasets/>
  - With States: 1997-2008 Large Fires (100,000+ fires) in AK, CA, ID
  - Graph the histogram of Size in Acres, by State



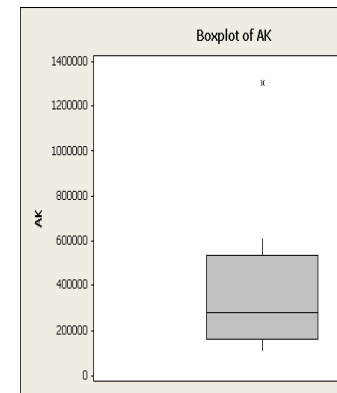
### Just Checking...



- Describe shape of each
- Compare and Contrast

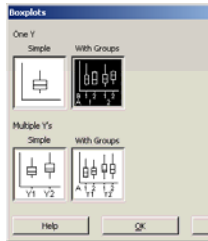
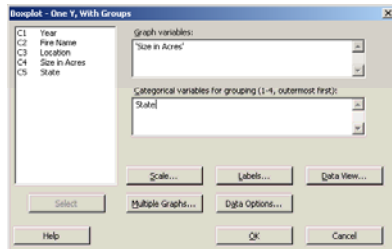
### More on Boxplots

- Upper bar at \_\_\_\_\_
- Middle bar at \_\_\_\_\_
- Lower bar at \_\_\_\_\_
- Length of 'Whiskers'
  - On the top drawn to smaller of maximum value or upper fence
    - Upper Fence =  $Q3 + 1.5 IQR$
  - On the bottom, drawn to larger of minimum value of lower fence
    - Lower Fence =  $Q1 - 1.5 IQR$
    - Note: Typo in Text pg 90
  - Values beyond whisker? Mark with asterisk (outlier)



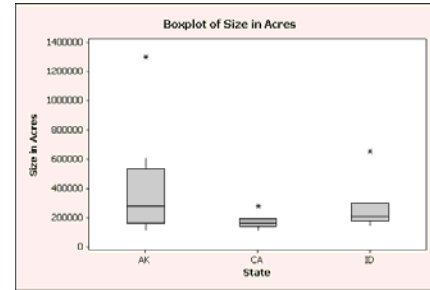
### Compare with Boxplots

- Minitab > Graph > Boxplots
  - One Y > With Groups
  - Select Variable to Graph
  - Select Categorical Variable for Grouping



- Plot Size in Acres by State

### Just Checking...



1. Compare shape, center, spread.
2. Which state had the largest fire?
3. True/False: Approximately 50% of large fires in AK are bigger than all of the large fires in CA.

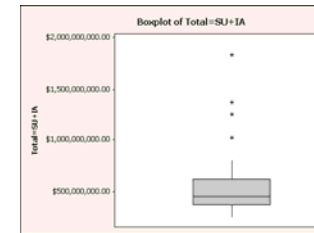
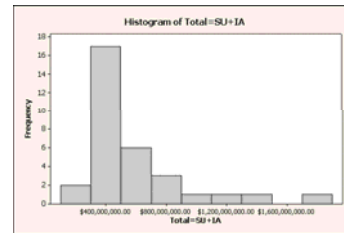
### Brief Interlude: Problems with Data

- My hunt for \$\$ spent on California wildfires
- Data Sets > California Fire Dollar Damage
- What's the issue? (*Hint: Think early years vs late ones*)

|    | C1   | C2            | C3          | C4           | C5 | C6 |
|----|------|---------------|-------------|--------------|----|----|
|    | YEAR | NUMBEROFFIRES | ACRESBURNED | DOLLARDAMAGE |    |    |
| 1  | 1933 | 1994          | 129210      | 318636       |    |    |
| 2  | 1934 | 2338          | 363052      | 563710       |    |    |
| 3  | 1935 | 1447          | 127262      | 165543       |    |    |
| 4  | 1936 | 3605          | 756696      | 1877147      |    |    |
| 5  | 1937 | 2907          | 71312       | 151584       |    |    |
| 6  | 1938 | 4150          | 221061      | 404225       |    |    |
| 7  | 1939 | 2491          | 513620      | 847579       |    |    |
| 8  | 1940 | 4497          | 156015      | 272178       |    |    |
| 9  | 1941 | 5460          | 278599      | 515737       |    |    |
| 10 | 1942 | 5236          | 573597      | 1,484,964    |    |    |

### Solution: Costs Normalized to 2002

- Data Sets > Total Fire Costs
- Histogram vs. Boxplot? How can we tell what one looks like based on the other?



WORKSHEET

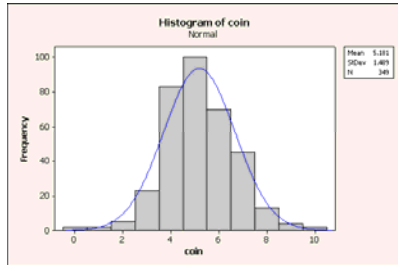
### Measure of Center: The Mean

- The **mean** is the usual average – sum all values and divide by the total number of values.

$$\bar{y} = \frac{\sum y}{n}$$

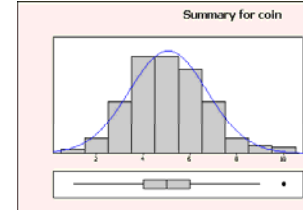
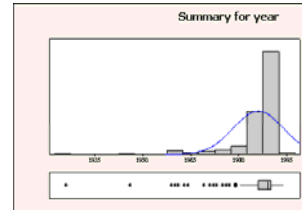
- Minitab > Stat > Basic > Display Descriptive
- Minitab > Graph > Hist > With Fit

- The mean acts like a point of balance



### Center of Symmetric Data: Mean

- The mean works best as a measure of center on symmetric data. For skew data, the median is best.



### Spread of Symmetric Data: SD

- The **standard deviation** (SD) is approximately the average distance a typical value is from the mean

$$s = \sqrt{\frac{\sum (y - \bar{y})^2}{n - 1}}$$

- Variance =  $s^2$
- The standard deviation is **ONLY** an appropriate measure of spread for symmetric data.
  - IQR is better for skew data.

WORKSHEET

### Class Work

- To get credit, it is your responsibility to get checked off.
  - Chapter 4 and 5 Handout II
    - Rules for checking answers: No Pens in the Front!!!

### Homework

- Textbook/Routine Homework
  - Due Next Week (25% chance of collection)
  - 1. Read Chapters 4 and 5
  - 2. Pg 78-86 #15, 17, 19, 21, 35, 53
  - 3. Pg 109-120 #7, 9, 19, 25, 29, 43
- NOTE: PLAN ON WORKING ON THE NEXT HW IN TLC!!!**
  - This is one of the times when using Minitab 15 is worth it.
- Project/Exploration Homework
  - None this time (coming next week!)