The Average and Standard Deviation

Measures of center and spread

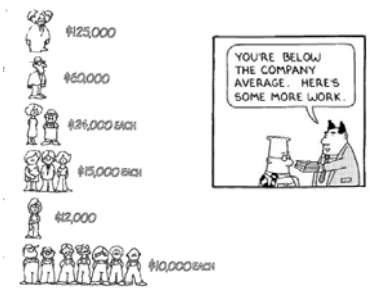
The Average

$$\text{Average} = \text{Mean} = \frac{\text{Sum of Observations}}{\text{Number of Observations}}$$

The Median

The middlemost number in the list

What's the Average Salary? The Median Salary?



Average Salary

125,000
+ 60,000
+ 2 x 24,000
+ 3 x 15,000
+ 12,000
+ 7 x 10,000

= 360,000

$\frac{\$360,000}{15} =$ $24,000

Median Salary

15 people, so 8th is middlemost

$12,000

The mean is the *balance* point.

The median is the *middlemost* point.

Averages depend on the proportions

List: 5, 5, 8: $Average = \frac{5+5+8}{3} = 5\left(\frac{2}{3}\right) + 8\left(\frac{1}{3}\right) = 6$

If there were 20 5's and 10 8's would the average be the same?

Suppose that 20% of the entries in a list are 1, 50% are 2 and 30% are 3. Then the average is

Be Careful What You Average

A class has three sections. On a midterm the first section averaged 68, the second averaged 70 and the third averaged 72. What was the class average?

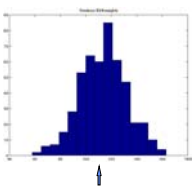
A school has two classes, one with 10 students and one with 100 students.

What is the average class size?

There are 110 students. What is the average size of the class that a student is enrolled in?

Using Averages to Compare: Birthweights

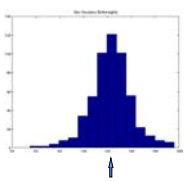
smokers



Average = 114

Median = 115

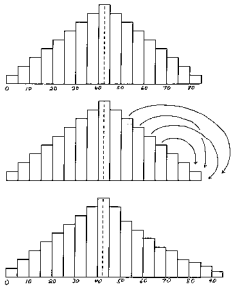
non-smokers



Average = 123

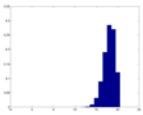
Median = 124

Increasing the mean and leaving the median unchanged



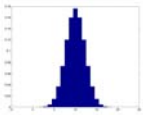
10

Mean, Median, Symmetry and Skewness



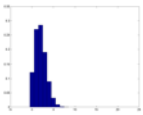
Skewed left

mean < median



Symmetric

mean = median

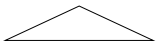


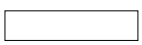
Skewed right

mean > median

11

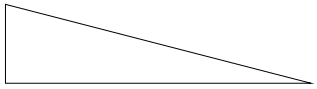
The mean=median for any symmetric histogram





12

Which is greater – the mean or the median?



WEEKLY HOME SALES

On the average

ALAMEDA
TOTAL SALES: 17
LOWEST AMOUNT: $125,000
HIGHEST AMOUNT: $579,000
MEDIAN AMOUNT: $339,000
AVERAGE AMOUNT: $359,000

ALBANY
TOTAL SALES: 8
LOWEST AMOUNT: $150,000
HIGHEST AMOUNT: $500,000
MEDIAN AMOUNT: $220,000
AVERAGE AMOUNT: $261,250

BERKELEY
TOTAL SALES: 20
LOWEST AMOUNT: $185,000
HIGHEST AMOUNT: $980,000
MEDIAN AMOUNT: $353,000
AVERAGE AMOUNT: $433,950
AVERAGE AMOUNT: $173,000

KENSINGTON
TOTAL SALES: 1
LOWEST AMOUNT: $265,000
HIGHEST AMOUNT: $265,000
MEDIAN AMOUNT: $ 0
AVERAGE AMOUNT: $265,000

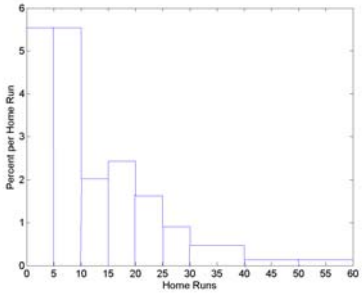
OAKLAND
TOTAL SALES: 90
LOWEST AMOUNT: $ 60,000
HIGHEST AMOUNT: $1,450,000
MEDIAN AMOUNT: $ 200,000
AVERAGE AMOUNT: $ 263,600

PIEDMONT
TOTAL SALES: 1
LOWEST AMOUNT: $510,000
HIGHEST AMOUNT: $510,000
MEDIAN AMOUNT: $ 0
AVERAGE AMOUNT: $510,000

Baseball Salaries

The average salary of players during a recent baseball strike was reported to be $1.2 million. The median salary was $500,000.

What would the histogram look like?



What is the median number of home runs, approximately?

Is the mean greater or less than the median?

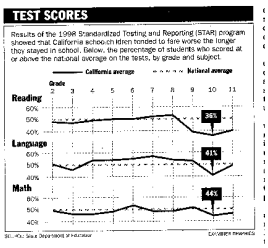
Review Question

In a certain real estate market, the average price of a single family home was $325,000 and the median price was $225,000. Percentiles were computed for this distribution. Is the difference between the 90th and 50th percentile likely to be bigger than, about the same as, or less than the difference between the 50th and 10th percentile? Explain briefly.

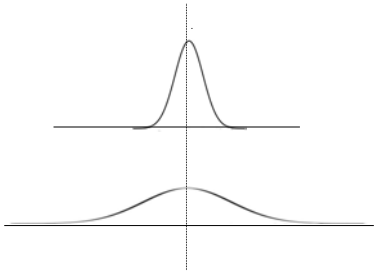
Averages from cross-sectional and longitudinal studies

New test sees weak scores in high school

Statewide trend shows declining results the longer a student has stayed in the system



Quantifying Spread



19

The Standard Deviation (SD)

The SD is a measure of how spread out numbers are around their average. Here is the recipe for calculating it:

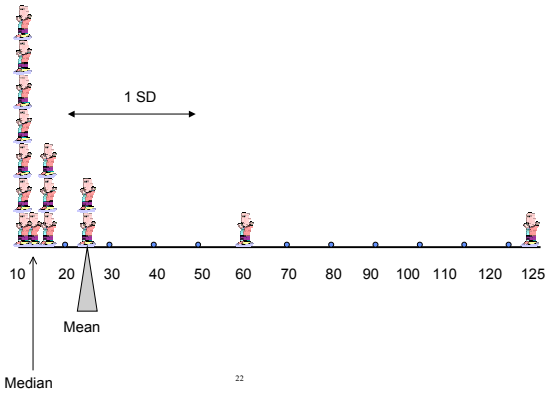
- Subtract mean from each number
- Square the results
- Add them up
- Divide by the length of the list
- Take square root of result

SD is the square root of the average squared deviation from the mean

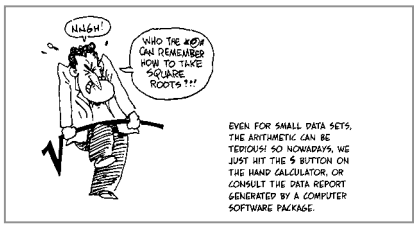
20

$(125000-24000)^2$
$+ (60000-24000)^2$
$+ (24000-24000)^2 + (24000-24000)^2$
$+ (15000-24000)^2 + (15000-24000)^2$
$+ (15000 - 24000)^2$
$+ (12000-24000)^2$
$+ (10000-24000)^2 + (10000-24000)^2$
$+ (10000-24000)^2 + (10000-24000)^2$
$+ (10000-24000)^2 + (10000-24000)^2$
$+ (10000-24000)^2$
$= 13,256,000$
$13,256,000/15 = 883,733$
$SD = \sqrt{883733} = 29,728$

21



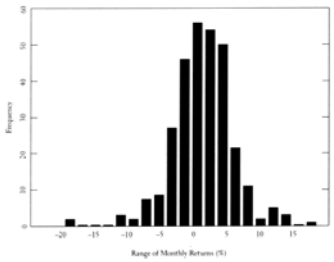
Caveat



On most calculators the **S** button does NOT give the SD as defined in the book, but a slightly larger number. Rather than dividing by the length of the list, they divide by the length of the list minus one.

Statistics and Finance

Distribution of Monthly Returns for a Portfolio Invested in the S&P 500-Stock Index, January 1940–July 1998

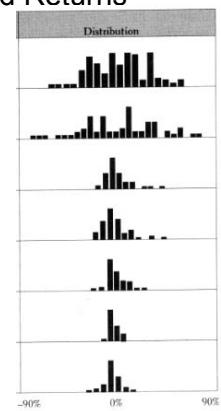


Av = 1%

SD = 4.5%

Risks and Returns

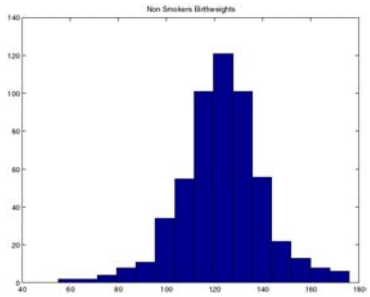
Type	Av	SD
Large Comp Stocks	13.0%	20.3%
Small Comp Stocks	17.7%	33.9%
Long Term Corp Bnds	6.1%	8.7%
Long Term Gov Bnds	5.6%	9.2%
Intermed Trm Gov Bnds	5.4%	5.7%
US Treasury Bills	3.8%	3.2%
Inflation	3.2%	4.5%



General Principle

For virtually any list, most of the entries are in the range “average plus or minus a few SDs”

The SD of Birthweights



One SD

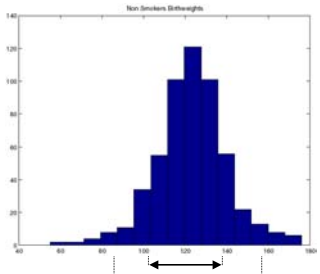
Mean - SD | Mean | Mean + SD

Rule of Thumb for Bell Shaped Histograms

About 68% (two in three) of the entries in a list are within one SD of the average.

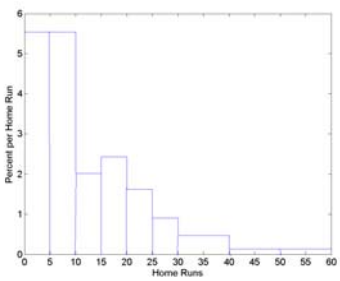
About 95% (19 in 20) are within two SDs of the average.

The SD tells how far numbers in a list are from their average. Most will be around one SD away. Few will be more that two or three SD away



74% within one SD

94% within two SDs



Change of Units

On the previous slide, the mean was about 123 ounces and the SD was 17 ounces. What would the mean and SD be if the birthweights were recorded in pounds?

mean (in pounds) =

SD (in pounds) =

Measuring in SD's

The mean was 123 and the SD was 17.

A child weighs 150 oz. How many SD's larger than the mean is he?

How much does a child who is 2 SDs less than the mean weigh?

Common Notation

$$Mean = \bar{x} = \mu$$

$$SD = s = \sigma$$

Summary

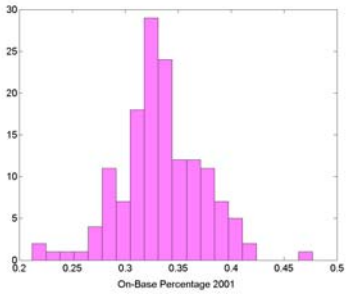
The SD measures spread around the average. It's a sort of average distance of values in the list from their overall average.

Technically, it is the square root of the average squared difference between the numbers and their average.

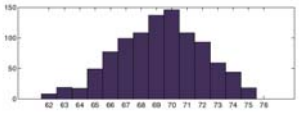
Typically, for bell shaped histograms, about 68% of a set of observations are within 1 SD of their average and about 95% are within two SDs.

American League On-Base Percentages 2001

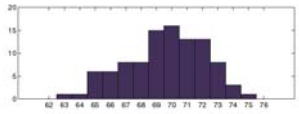
Mean, Median, SD?



8. One investigator takes a sample of 100 men age 18–24 in a certain town. Another takes a sample of 1,000 such men.
   (a) Which investigator will get a bigger average for the heights of the men in his sample? or should the averages be about the same?
   (b) Which investigator will get a bigger SD for the heights of the men in his sample? or should the SDs be about the same?
   (c) Which investigator is likely to get the tallest of the sample men? or are the chances about the same for both investigators?
   (d) Which investigator is likely to get the shortest of the sample men? or are the chances about the same for both investigators?



Sample of 1000
Average = 70.2
SD = 3.1
min= 60.8
max= 78.6



Sample of 100
Average = 69.7
SD = 3.2
min = 62.3
max = 79.6

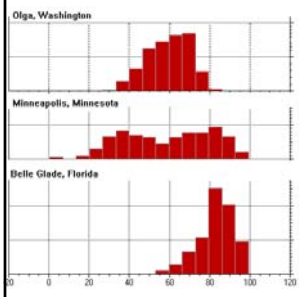
37

Why should the SDs be about the same?

- The range around the average encompassing 68% of the data should be about the same in the two samples.
- The average squared distance from the mean should be about the same in the two samples.

38

Histograms of Daily High Temperatures



1. The average high temperature for Minneapolis is closest to (a) 45 degrees (b) 60 degrees (c) 75 degrees (d) 85 degrees

2. The SD of the high temperatures for Minneapolis is closest to (a) 1 degree (b) 3 degrees (c) 5 degrees (d) 20 degrees

3. The average high temperature for Minneapolis is _____ the average high temperature for Belle Glade. (a) at least ten degrees less than (b) about the same as (c) at least ten degrees higher than

4. The average high temperature for Minneapolis is _____ the average high temperature for Olga. (a) at least ten degrees less than (b) about the same as (c) at least ten degrees higher than

5. The SD of the high temperatures for Minneapolis is _____ the SD of the high temperatures for Belle Glade. (a) about half of (b) about the same as (c) about twice

39