



MAT116 Project 3 Chapter 12

1

Histograms

- As we have already hinted at in 298B, histograms can be used to group data together.
- Given a large pile of numeric data, histograms can provide a visual representations of how the data is distributed.
- We do a short review here.

2

Definitions

- An **observation** is a single number. It could be a measurement, monetary amount, etc. It can also be considered to be one particular outcome to some random trial.
- Raw data** is a collection of observations
- Frequency** is how many observations are in each bin.

3

Definitions

- **Relative frequency** is the percentage observations in each bin.
- A **frequency distribution** is a chart which shows the bins and frequencies.

4

Example

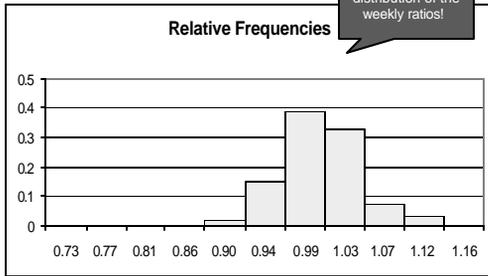
- Let's build a histogram of the "Ratio of Closes" for our Disney stock.
- One new feature we will graph on the histogram is what is called the *relative frequency*, which is just the percent of observations in each bin.

5

Notes

6

The Result



7

The Mean

- In statistics, one of the main measurements is known as the mean, or the average.
- Question for Thought: How are the ideas of mean and expected value related? When are the two equal to each other?

8

Formal Definition

- Given n observations, the mean of these observations is given by:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

What does this actually mean? Why so complicated?

9

Example

- Write the sum of the first 8 odd numbers in summation notation.
- Find the sum.

10

Central Tendency

- The mean of a set of data is one way to measure what is called the central tendency of that data.
- Central tendency tells you about what is going on at the “center” or in the middle of your data.

11

Central Tendency

- Another measure of central tendency is the median, which is the point at which half of your observations lie above the median and half lie below.
- =MEDIAN(range) can be used in Excel to compute the median.

12

Other Statistical Measures

- The mean or median are fine, but they don't tell you anything about how data is spread out.
- For example, are all the data points closely clustered together or are thinly spread out?

13

Range

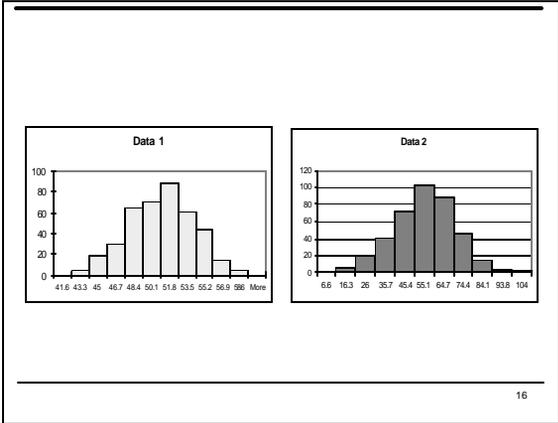
- The first measure of spread is the *range*. We've already been exposed to this in our histogram exercises. The range is simply the MAX-MIN.
- The major disadvantage to the range as a measure of spread is that it only takes into account TWO observations of data...the others are completely ignored.

14

Example

- For example, consider the two sets of data in the sheets for Ch12.xls called Data 1 and Data 2.
- What is the range of each set of data? The mean? Median?
- What about the histograms? (We won't worry about the midpoints or relative frequencies for now.)

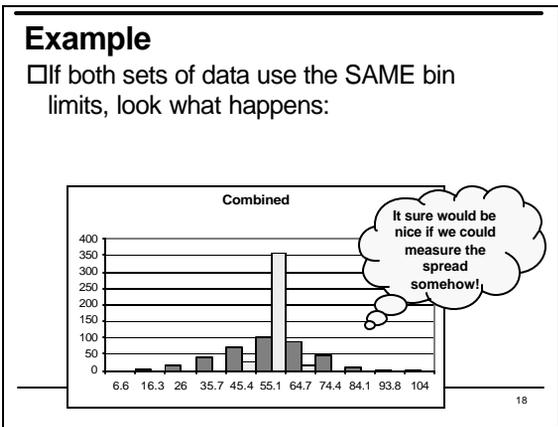
15



Example

- How do they compare?
- What can we say about the spread?

17



Measuring Spread or Variation

- The range only takes into account two observations of data. If we want to take into account all our data points, one way to do it would be to look at how far each data point lies from the mean, \bar{x} .

The symbol for the mean is read "x-bar"

19

Measuring Spread

- Hence, for each value of x_i , we can compute the value of $(x_i - \bar{x})$ and then add all those up!
- Aahh...if only life were that simple... Watch what happens...

20

Example

- Let's say we have the following data:
 - 5,6,9,8,5,6,3,2,10
- The mean of this data is $\bar{x}=6$
- Now we subtract

X	$X - \bar{X} =$	
5	5-6	-1
6	6-6	0
9	9-6	3
8	8-6	2
5	5-6	-1
6	6-6	0
3	3-6	-3
2	2-6	-4
10	10-6	4
	Total	0

21

Harumph!

- The sum of the $(x_i - \bar{x})$ values will always be zero so this method of measuring spread is useless.
- Notice that some of the $(x_i - \bar{x})$ values are positive and some are negative...if they were all positive, we could avoid the sum being zero.
- So...after finding $(x_i - \bar{x})$, we will square that and compute $(x_i - \bar{x})^2$.

22

New Method

- Now when we square them and add, we get 56.
- Finally, we divide by $(n-1)$, where n is the number of data points in the set. This gives us what we call the variance of this data set.

x	$x - \bar{x}$	$(x - \bar{x})^2$
5	5-6	1
6	6-6	0
9	9-6	9
8	8-6	4
5	5-6	1
6	6-6	0
3	3-6	9
2	2-6	16
10	10-6	16
	Total	56

23

Variance

$$\frac{56}{9-1} = 7$$

Hello, my name is VARIANCE

24

Formulas

□ Variance = s^2

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

□ Note that we divide by $n-1$, not n .

□ Standard Deviation = s

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

25

Standard Deviation

□ The *standard deviation* of a set of observations is the square root of the variance.

□ We take square root essentially to undo the part where we squared the $(x - \bar{x})$ values. Hence...

26

27

Example

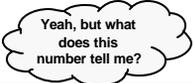
□ Find the standard deviation of the following set of data.

■ 5, 6, 9, 8, 5, 6, 3, 2, 10

□ This is the same data as before.

□ We know the variance = 7, so we know the standard deviation.

■ $\sqrt{7} \approx 2.65$



28

29

Example

□ Find the standard deviation of the set of data found in the STDEV sheet of the Ch. 12 Excel file.

30

Using Excel to find s

□ To find the standard deviation of a set of data, use the following command:

□ =STDEV(*range*)

□ To find the variance of a set of data, use the following command:

□ =VAR(*range*)

31

Example

□ Find the variance and standard deviation of the weekly ratios for the Disney stock.

□ What is the meaning of the standard deviation in this context?

32

Adjusting Data

□ Until now, we have been using the weekly ratios for our stock, which essentially violates one of our assumptions. Namely, we cannot use historical data to predict the value of a stock in the future.

33

Adjusting Data

- We need to come up with a way to do this, however, to find a fair price for our stock option.
- To do so, we will fall back on another assumption: our investments will grow at the risk-free rate and we can somewhat reliably predict volatility from past volatility.

34

Adjusting Data

- To proceed in this way, we need to adjust EACH of our weekly ratios so that our overall behavior lines up with that of our risk-free rate of 4%.
- However, we need to be sure that whatever adjustments we make do not affect the fundamental characteristics of our data...specifically, we want to make sure our standard deviation remains preserved.

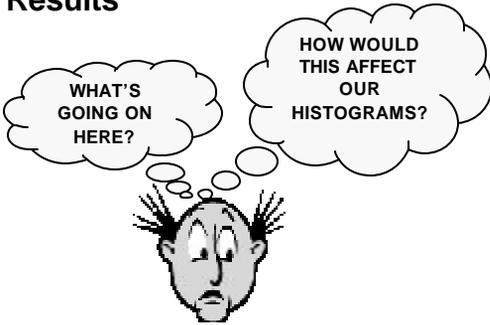
35

Adjusting

- What happens when we add or subtract a constant to all of our weekly ratios?
 - What happens to the mean?
 - What happens to the variance?
 - What happens to the standard deviation?
- We'll use our actual data to explore this.

36

Results



37

Sampling: Definitions

- A **population** is everyone or everything that is being studied.
- A **parameter** is a number which describes a population.
- A **sample** is a portion of the population.
- A **statistic** is a number which describes a sample.

38

Using Samples

- When and how do we use samples?
- What do we mean when we say we have a random sample?
- What is meant by the term "sample size?"

39

Example

- In the Random sheet of Chapter 12.xls, there is a set of data that will represent a population. (Most populations are much larger.) Let's say each cell gives the age of a person.
- Let's generate a random sample of size 10.

40

Example

- What is the *sample mean*? That is, what is the average age of the 10 randomly chosen persons?
- How does the sample mean compare to the population mean?

41

Law of Large Numbers

- The sample mean can be used to estimate the mean (expected value) of a random variable. In general, the larger the sample, the more accurate the estimate.

42

Law of Large Numbers

Another way of saying this is that if we were to choose a random sample of size 10 “many many” times and compute each sample mean, then the average of these means will be very close to the population mean. (We will call this μ later.)

43

Project Note

- Here's where we are headed.
- Each team has several hundred weekly ratios in their Excel file.
- Think of this set as a population of weekly ratios.

44

Project Note

- For our Class Project, we had a 20-week option, so we will randomly choose 20 of these many ratios and compute a number associated with each one. (The number to be computed is yet to be specified.)
- We will do this “many many times;” in our case several thousand times.
- We will use the results of this to estimate our stock's closing price at the end of the option period.

45

Volatility

- Volatility is a measurement of how much the value of a stock fluctuates.
- One of our assumptions is that we can generally use past volatility to predict future volatility.
- Hence, we need a way to define this term.

46

Definition

- There are a few different ways to measure volatility.
- The one we will use is as follows:
- Volatility is the standard deviation of the (weekly) ratios

47

Example

- What is the volatility of the Disney stock based on our data?

- Note that this is based on data collected weekly, as opposed to daily, monthly, or yearly.

48

Adjusting the Volatility

- Because we have weekly ratios, we need to *annualize* the volatility. To do this, we multiply by $\sqrt{52}$.
- To annualize monthly ratios, multiply by $\sqrt{12}$.
- To annualize weekly ratios, multiply by $\sqrt{52}$.
- To annualize daily ratios, multiply by $\sqrt{252}$ (252 business days per year)

49

Example

- What is our annualized volatility for the Disney stock?

50

Focus on the Project

- We can now take our team data and compute all the weekly ratios.
- We can also summarize this data by finding
 - the minimum and maximum ratio
 - the mean ratio and median ratio
 - the variance of the ratios
 - the standard deviation of the ratios

51

Focus on the Project

In addition, we can compute the annualized volatility of our stock and talk about what that means in our particular cases.

52

Focus on the Project

We are also now in the position to adjust our weekly ratios so that we don't violate the assumption that we can use past price data to predict future price data. However, we want to do so in a way that does not affect the volatility (standard deviation of ratios) of our stock. (Why?)

53

Focus on the Project

So, how should we adjust our weekly ratios?

Whatever method we use should result in our average weekly ratio being the same as the weekly risk-free ratio!!! (Important: Why?)

54

Focus on the Project

- Let's review what we know:
- Our mean weekly ratio for the class project is approximately 1.001894.
 - Let's call this R_m for the "mean of the Ratios"
- Our computed weekly risk-free ratio is approximately 1.000795.
 - Let's call this R_f for the risk free Rate
- Notice that R_m is much too large.

55

Focus on the Project

- This means that on average, each of our weekly ratios is too large.
- In particular, each one is $(R_m - R_f)$ too big! (This needs to make sense before you move on.)
- Let's call $(R_m - R_f)$ the *ratio excess*

56

Focus on the Project

- If we took every weekly ratio and reduced it by the quantity $(R_m - R_f)$, then the new set of data we would get would have an average equal to the weekly risk-free rate. This would be good! (Why?)

57

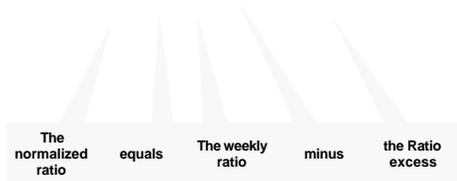
Focus on the Project

- Therefore, for each ratio, R_i , in our list, we will adjust it by removing or subtracting the excess that is represented by $(R_m - R_{rf})$.
- We call this *normalizing* each ratio.

58

Focus on the Project

$$R_{norm} = R_i - (R_m - R_{rf})$$



59

Focus on the Project

- Let's do this to our Option Data file. AND..
- Make sure the new mean of the R_{norms} matches the weekly risk-free rate.

60
