Session # Four – Outline for Today

1. Some general announcements
2. Quiz #3 - review of question #16
3. About Z-scores and their use
4. More measures of dispersion including building a box plot
5. Assorted problems, including quiz
6. Discuss "Out of Class Project"
7. Using Excel 2010 for histograms
8. Linear Correlation, ch. 4 begun

1

General Points of Information

Excel #1 due at session 6, 13 Feb

- http://faculty.harford.edu/faculty/dschwank/Stat216/lectures/main_lecture_fall2014.htm
- linked instructions, sample, raw data
- Expectations: HCC faculty webpages

An "opportunity" at every class in many forms: class solo work, InterActMath, group projects

2

3 Measures of Central Tendency

Mean – sum the data values, divide by number of data points

Mode – most frequently occurring

Median – arrange in order, count to the middle

3

Measures of Dispersion in Data

- Range - difference between HI & LO
- Variance - average squared deviation about the mean
- Standard Deviation - square root of variance (for both population & sample)
- Examples (by hand): four point populat.
  - First eight papers from quiz #3
  - #3.2.10, page 151, find σ^2 and σ
  - #3.2.28, Chicago or San Diego, ºF
- Use of Empirical Rule (fig 13, p 149)

4

Measures of Position Definitions

z-Score Definition: the distance the data value is from the mean, expressed in terms of standard deviations

Is a "unitless" measure

For a “standard normal curve”

- Mean of zero
- Standard Deviation of one

5

Measures of Position Definitions

z-score equals [(data value minus mean) divided by standard deviation]

z-score purpose is to provide a way to "compare apples and oranges"

- by converting variables with different centers and/or spreads
- to variables with the same center (0) and spread (1).

6

More Measures of Dispersion

- Percentiles - the percentage of observations that are above and below a certain point
- Quartiles - percentiles at 25%, 50% and 75%, aka Q_1, Q_2, & Q_3
- Example of Super Bowl Scores
  - Mean and Mode
  - Percentiles at 50%, 25% and 75%

7

Numerically summarizing data

- Five number summaries
- Interquartile range ($Q_3 - Q_{1)}$ is resistant to extreme values
- Compute five number summary
- Min value | Q_1 | M | Q_3 | max value
- Summary of formulas on p182-183

8

Numerically summarizing data - Constructing a Box Plot

- Will use the five number summary to create another graph
- Will compute IQR and “fences”
- Will plot the data on horizontal axis
- Quick glance summarizes data

9

More Measures of Dispersion

Upper and lower fences (first find 1.5 times interquartile range)
- Lower fence = Q_1 - 1.5(IQR)
- Upper fence - Q_3 + 1.5(IQR)

Boxplot - shows Q_1, Q_2, Q_3, data between the fences, plus outliers

See pictures: Figure 22, page 177

Example: Super Bowl Score (continued)

10

Building a Box Plot – part 1

1. Calculate interquartile range (IQR)
2. Compute lower & upper fence
   - Lower fence = Q_1 – 1.5 (IQR)
   - Upper fence = Q_3 + 1.5 (IQR)
3. Draw scale then mark Q_1 and Q_3
4. Box in Q_1 to Q_3 then mark M

11

Building a Box Plot – part 2

5. Temporarily mark fences with brackets
6. Draw line from Q_1 to smallest value inside the lower fence and a line from Q_3 to largest value inside the upper fence
7. Put * for all values outside of the fences
8. Erase brackets

12

Box Plot Examples: Super Bowl & Quiz

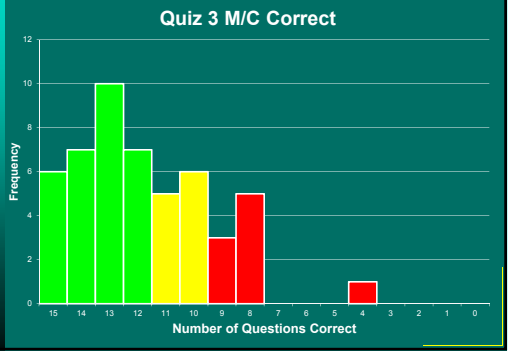
Quiz data from Tuesday, for #1-15:
How many students got this number of questions correct:

15 → 6	7 → 0
14 → 7	6 → 0
13 → 10	5 → 0
12 → 7	4 → 1
11 → 5	3 → 0
10 → 6	2 → 0
9 → 3	1 → 0
8 → 5	0 → 0

Note: there are 50 data points total

13

Quiz #3 Data (frequency histogram)



14

Box Plot Examples: Quiz #3 M/C

Descriptive Data
- Mean = 11.8
- Std Dev (population) = 2.4
- Range = 11.0

Five Number Summary
- Min = 4.0
- Q1= 10.0
- Median= 12.0
- Q3= 13.8
- Max= 15.0

For Box Plot
- IQR= 5.7
- Lower fence 4.3
- Upper fence 19.5

15

Distribution based on Boxplot

- Symmetric
  - median near center of box
  - horizontal lines about same length
- Skewed Right / Positive Skew
  - median towards left of box
  - right line much longer than left line
- Skewed Left / Negative Skew
  - median towards right of box
  - left line much longer than right line

16

Which measure best to report?

- Symmetric distribution
  - Mean
  - Standard Deviation
- Skewed distribution
  - Median
  - Interquartile Range

17

Self Quiz

When can the mean and the median be about equal?

In the 2010 census conducted by the U.S. Census Bureau, two average household incomes were reported: $41,349 and $55,263. One of these averages is the mean and the other is the median. Which is which and why?

18

Self Quiz

The U.S. Department of Housing and Urban Development (HUD) uses the median to report the average price of a home in the United States.

Why do they do that?

19

Self Quiz

A histogram of a set of data indicates that the distribution of the data is skewed right.

Which measure of central tendency will be larger, the mean or the median?

Why?

20

Self Quiz

If a data set contains 10,000 values arranged in increasing order, where is the median located?

Matching: (parameter; statistic)

- _____ is a descriptive measure of a population
- _____ is a descriptive measure of a sample.

21

Self Quiz

A data set will always have exactly one mode. (true or false)

If the number of observations, n, is odd; then the median, M, is the *value* calculated by the formula $M=(n+1)/2$

22

Self Quiz

Find the Sample Mean:
20, 13, 4, 8, 10

Find the Sample Mean:
83, 65, 91, 87, 84

Find the Population Mean:
3, 6, 10, 12, 14

23

Self Quiz

The median for the given list of six data values is 26.5.

7 , 12 , 21 , ___ , 41 , 50

What is the missing value?

24

Self Quiz

The following data represent the monthly cell phone bill for the cell phone for six randomly selected months.

$35.34	$42.09	$39.43
$38.93	$43.39	$49.26

Compute the mean, median, and mode cell phone bill.

25

Self Quiz

Heather and Bill go to the store to purchase nuts, but can not decide among peanuts, cashews, or almonds. They agree to create a mix. They bought 2.5 pounds of peanuts for $1.30 per pound, 4 pounds of cashews for $4.50 per pound, and 2 pounds of almonds for $3.75 per pound. Determine the price per pound of the mix.

26

End Self Quiz, Start Instructor's Quiz

Any questions about chapter three?

Quiz #4 details:

- 15 Multiple Guess questions
- No "long" calculations
- Closed notes, closed book
- Individual effort only
- Calculator may be used (although can do entire quiz without)
- 15 minute time limit enforced

27

Out of Class Project

The purpose of this project is to gather some original data and analyze it, using the methods discussed in the first four chapters of the textbook.

Preliminary summary: covering objectives 1 & 2, due at class #6 (one week) total of 2 - 4 sentences

28

Out of Class Project

Objective is a written report describing:

1. What the question to be answered is
2. The type of sampling used and why
3. A summary of the raw data
4. The statistical analysis of that data
5. A summary of what that analysis actually means
6. Conclusion(s) / answers to the original question.

29

Project Sample Questions (select 1)

Are there really less than 50% peanuts in mixed nuts bags?

Does leg length matter in 40 yard dash times for the HCC baseball team?

Do different branches in the organization have different technical report preparation times?

30

Project Sample Questions (select 1)

- How many hours per week does a full time student spend working a part-time job?
- Are there differences in cell phone minutes used by classmates?

- Show examples of good analysis
- Final report due class #10, 2 Oct

31

To be starting into Chapter 4

- Explanatory variables
- Response variable
- Scatter diagrams
- Linear regression

But first . . .

32

Microsoft Excel 2010 = Spreadsheet

- Available in Library & Math Center
- Four technology assignments (50)
- Problems from text: work both ways
- Excel terms: rows, columns, cells
- Enter text or data or formulas
- Software can do the calculations

33

Looking at Excel Technology

Frequency Tables and Histograms
(needed for Excel assignment #1)
Measures of Central Tendency
(as we have just done in Chapter 3)
Graph / Chart types
(needed for out-of-class project)
Linear Regression
(not required for now)

34

Looking at Excel Technology

Excel as a "spreadsheet"
- Cells in rows and columns
- Words, numbers, or formulas in cells

Example of Greenhouse Gas #2.T.2
- Organize data into a table
- Formulas for sum and divisions
- Make into a pie chart

Example of Super Bowl Margins
- Organize data into a table
- Calculate point margins and sort ↓

35

Still Looking at Excel Technology - 1

Excel / Technology #1 assignment due in one week (next Thursday)
Problem #1 based on Page 96, Section 2.2.31 (a) to (e) & 2.3.15
- Get raw data: book, CD, website
- Create a "bin" for $30,000 and up by $6000 class widths
- Use Excel create frequency table
  - Lower and Upper Class Limits
  - Frequency, Relative Frequency
  - Cumulative Freq, Relative Cum. Freq

36

Still Looking at Excel Technology - 2

Excel / Technology #1 assignment due in one week (next Thursday)
Continuing with Problem #1 based on Page 96, #2.2.31 and #2.3.15
- Insert Excel formulas to sum rows & columns plus do divisions for relative, cumulative, & rel cum freq
- Use Excel to create freq histogram
- Use an open cell or insert a text box to answer part (e)

37

Still Looking at Excel Technology - 3

Excel / Technology #1 assignment due in one week (next Thursday)
Problem #2 based on Page 108, Section 2.3, #21(a) to (e), plus two
- Get raw data: book, CD, website
- Create "bins" for stocks with a class width of 10, beginning at minus 20
- Use Excel create frequency table for both consumer and energy stocks

38

Still Looking at Excel Technology - 4

Excel / Technology #1 assignment due in one week (next Thursday)
Continuing with Problem #2 based on Page 108, Section 2.3, #21
- Use Excel to create frequency "histogram" for both types of stocks
- Copy and paste histograms then change chart type to line chart for frequency polygons and ogives
- Use an open cell or insert a text box to answer part (e)

39

Creating Histograms with Excel

To create a histogram:

- Must first install Data Analysis pack
- All Library, Tutor Center, A-223, and A-258 computers have it already,
- Use Microsoft Office Button/Excel Options/Add-Ins/Analysis ToolPak/OK

Excel calls “lower class limit” a “bin”

Enter lower class limits manually

40

Creating Histograms with Excel

Raw data into Excel
Sort to get idea for class limits
Use data/data analysis/histogram
Enter data array and bin
Select options for output
Make histogram look presentable
Complete frequency table
Example of Chocolate Chips #2.R.7

41

Still Looking at Excel Technology - 5

Problem #3 is 3.R.1 on Page 183

- Use Excel for descriptive statistics

Problem #4 based on Page 242, Section 4.4, #7 (b), (d), and (e)

- Use Excel to create relative frequency marginal distribution table
- Use Excel to create a conditional distribution table
- Use Excel to insert a side-by-side column graph of the conditional distribution table

42

Excel Projects in Statistics

- CD icon in text means raw data in an Excel format is available
- Four Excel "technology" assignments throughout semester
- First submission preferred by email with Excel file attached (by paper)
- Future assignments will have mandatory electronic submission
- Grading increasingly stringent to "business quality" standards

43

Definitions (starting into Chapter 4)

explanatory variables = factors = variable whose value can not be explained = independent variable = predictor variable = X-axis number

response variable = variable of interest = variable whose value can be explained = dependent variable = Y-axis number

44

Build a Scatter Diagram

Use data on page 201-2, problem #27
Height versus Head Circumference

1. 27.75 // 17.5
2. 24.5 // 17.1
3. 25.5 // 17.1
4. 26 // 17.3
5. 25 //16.9
6. 27.75 // 17.6
7. 26.5 // 17.3
8. 27.0 // 17.5
9. 26.75 // 17.3
10. 26.75 // 17.5
11. 27.5 // 17.5

45

Linear Correlation

- Measure of the strength of linear relations between two quantitative variables
- Represented by Greek letter “rho” *r*

46

Linear Correlation

- Equals sum for all *i* of
- [(x sub *i* minus x bar divided by sample standard deviation of x)
- Times (y sub *i* minus y bar divided by sample standard deviation of y)]
- All divided by (number of individuals in the sample minus 1)

47

Properties of Linear Cor Coefficient

- Always between -1 and +1
- The closer to +1 the stronger the positive linear relationship
- The closer to -1 the stronger the negative linear relationship
- Close to zero means little linear relation between the two variables
- Is a "unitless" measure

48

Linear Correlation

Sample problem to work by hand (1)

x	2	3	5	6	6
y	5.7	5.2	2.8	1.9	2.2

Step 1: create table with five columns

Step 2: for both x and y, calculate mean and standard deviation

49

Linear Correlation

Sample problem to work by hand (2)

Step 3: compute $(x_i\text{-mean}_x / s_x)$

Step 4: compute $(y_i\text{-mean}_y / s_y)$

Step 5: (step 3) times (step 4)

50

Linear Correlation

Sample problem to work by hand (3)

Step 6: (step 5) divided by (n-1)

Step 7: examime (step 6) to determine degree of linear correlation

51

Finding a Linear Equation

Recall: you already know a method to find linear equations

Point-slope method

- use two points to find slope
- then one point to find y intercept

Example

52

Notes of the day

We have finished the first three chapters, so be sure to have all these suggested problems worked

Use weekend to think of a project

Start on Excel Project #1

- Use Math Center to help T/R/S
- Use group statistics sessions M/W/F in Aberdeen Hall Computer Labs

53

Other Notes of the Day

1.
2,.
3.
4.
5.
6.

54