STT 200 – Lecture 1, section 2,4 Recitation 7 (10/16/2012)

TA: Zhen (Alan) Zhang

zhangz19@stt.msu.edu
Office hour: (C500 WH) 1:45 – 2:45PM Tuesday
(office tel.: 432-3342)
Help-room: (A102 WH) 11:20AM-12:30PM, Monday, Friday

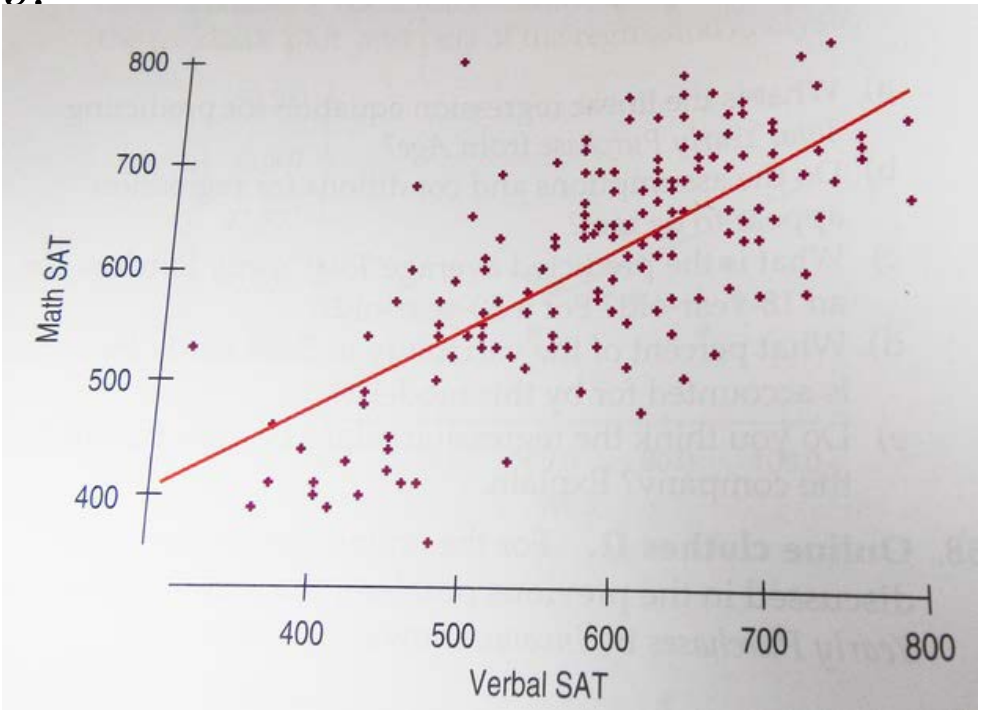
Class meet on Tuesday:
3:00 – 3:50PM A122 WH, Section ***02***
12:40 – 1:30PM A322 WH, Section ***04***

Overview

- We will discuss following problems:
  - Chapter 8 "*Liner Regression*" (Page 220): #39, 40
  - Chapter 14 "*From randomness to Probability*"(Page 379) #1,2,3,4,11,12

- All recitation PowerPoint slides available at here

- Chapter 8 (Page 220): #39:



- Describe the relationship;
- Any students that do not fits the overall pattern?
- Interpret r = 0.685;
- Verbal: mean 596.3, Stdev.=99.5, Math: mean 612.2, Stdev.= 96.1, write the equation of regression.
- Interpret the slope
- Predict for verbal = 500;
- What is the residual for student with total 1600?

- Chapter 8 (Page 220): #39 (continued):
  - Strategies:

1. Identify the explanatory variable (x) and response variable (y) in a regression analysis; We are interested in predicting y based on x, or explain the information (variation) of y using x.
2. find their respective information such as *mean* and *standard deviation*, (i.e. $\bar{x}, S_x, \bar{y}, S_y$), then you can calculate the slope and intercept using

$$b_1 = r\frac{S_y}{S_x}; \quad b_0 = \bar{y} - b_1\bar{x}$$

We can also calculate **R-square** statistic by taking square of the correlation r for model assessment. It represents *the proportion of total variation of y that has been explained by the regression model*.

So R-square closer to 1 indicates better model, while closer to 0 indicates a poor model fit (x might be not so useful in predicting y!).

- Chapter 8 (Page 220): #39 (continued):
  - Describe the relationship;

Moderately strong, fairly straight, positive. Possible outliers.

  - Any students that do not fits the overall pattern?
  - Interpret r = 0.685;

Positive, fairly strong linear relationship. 46.9% variation in math scores is explained by verbal scores.

  - Verbal: mean 596.3, Stdev.=99.5, Math: mean 612.2, Stdev.= 96.1, write the equation of regression.

Math = 217.7 + 0.662 Verbal based on $b_1 = r\frac{s_y}{s_x};\ b_0 = \bar{y} - b_1\bar{x}$

  - Interpret the slope

Every point of verbal score adds an average 0.662 points to the predicted math score.

  - Predict for verbal = 500;

Ans = 548.5 points

  - What is the residual for student with total 1600?

Ans = 52.7 points

- Chapter 8 (Page 220): #40:

  SAT score: mean 1833, Stden. = 123;

  GPA: mean 2.66, Stden. = 0.56;

  Scatterplot is reasonably linear, correlation = 0.47.

  - write the equation of regression.
  - Interpret the intercept;
  - Predict GPA with SAT = 2100;
  - How effective is SAT predicting GPA?
  - Would you rather have a positive or negative residual?

- Chapter 8 (Page 220): #40 (continued):

  SAT score: mean 1833, Stden. = 123;

  GPA: mean 2.66, Stden. = 0.56;

  Scatterplot is reasonably linear, correlation = 0.47.

  - write the equation of regression.

  GPA = -1.262 + 0.00214 SAT

  - Interpret the intercept;

  0 SAT score would have -1.262 GPA. *Impossible*. Adjust the height of the line and is meaningless itself.

  - Predict GPA with SAT = 2100; (3.23)

  - How effective is SAT predicting GPA? (R square = 0.47 ^2=0.221), somewhat useful. Might be affected by other factors.

  - Would you rather have a positive or negative residual?

  (positive, since this indicates the actual GPA is higher than expected GPA based on the overall performance. Recall residual = observed y – predicted y)

- Chapter 14 (Page 379): #1:

  Find sample space and whether you think the events are equally likely.

  - Toss a coin; record the order of heads and tails.
  - A family has 3 children; record the number of boys.
  - Flip a coin until you get a head or 3 consecutive tails.
  - Roll two dice; record the larger number.

  Tips: Sample space is the collection of all possible outcome events.

- Chapter 14 (Page 379): #1 (continued):

  Find sample space and whether you think the events are equally likely.

  - Toss two coins; record the order of heads and tails.

  (S={TT,TH,HH,HT}. Equally likely, each assumes a probability of 0.5*0.5=0.25.)

  - A family has 3 children; record the number of boys.

  (S={0,1,2,3}. Unequally likely)

  *Tips: There are <u>four</u> events in total and if they are equally likely, each should assume probability of 0.25 to ensure the sum is 1. Now check the event that "none of them are boys", the probability is 0.5*0.5*0.5 = 0.125 which does not equal 0.25, so they can not be equally likely.*

  - Flip a coin until you get a head or 3 consecutive tails.

  (S={H,TH,TTH,TTT}. Unequally likely)

  The probability of event = observe H is 0.5, which does not equal 0.25.

  - Roll two dice; record the larger number.

  (S={1,2,3,4,5,6}. Unequally likely)

  The probability of event = observe 1, i.e., observe the pair (1,1) is $\frac{1}{6} \times \frac{1}{6} = \frac{1}{36}$

  which does not equal $\frac{1}{6}$.

- Chapter 14 (Page 379): #1 (continued):

Summary:

1. **Toss a coin**: outcome is either head or tail. Without particular specification, we assume the coin is fair, that is, the chances of getting a head and a tail are equal; each assumes 0.5.

2. **Gender of a child**: outcome is either boy or girl. Without particular specification, we assume the chances of observing a boy and a girl are equal; each assumes 0.5.

3. **Toss a dice**: outcome is among {1,2,3,4,5,6}. Without particular specification, we assume equal chances of observing a number from 1 to 6; each assumes $\frac{1}{6}$.

- Chapter 14 (Page 379): #2:

  Find sample space and whether you think the events are equally likely.

  - Roll two dice; record the sum of the numbers;
  - A family has 3 children; record each child's sex in order of birth;
  - Toss four coins; record the number of tails.
  - Toss a coin 10 times; record the longest run of heads.

- Chapter 14 (Page 379): #2 (continued):

  Find sample space and whether you think the events are equally likely.

  - Roll two dice; record the sum of the numbers;

  (S={2,3,4,…,11,12}. Unequally likely)

  - A family has 3 children; record each child's sex in order of birth;

  (S={BBB,BBG,BGB,BGG,GGG,GBG,GBB,GGB}. Equally likely. *Tips: the total number of events should be 2*2*2 = 8, help check.*)

  - Toss four coins; record the number of tails.

  (S={0,1,2,3,4}. Unequally likely)

  - Toss a coin 10 times; record the longest run of heads.

  (S={0,1,2,3,…,9,10}. Unequally likely)

- Chapter 14 (Page 379): #3:

  A casino claims that its roulette wheel is truly random.

  What should that claim mean?

  (Every number is equally likely to occur)

- Chapter 14 (Page 379): #4:

  The weather reporter on TV makes prediction such as 25% chance of rain. What do you think is the meaning of such a phrase?

  (In circumstances "like this", rain occurs 25% of the time.)

- Chapter 14 (Page 380): #11:

Which of following probability assignments are possible?

	red	yellow	green	blue	Valid?
a)	0.25	0.25	0.25	0.25	?
b)	0.10	0.20	0.30	0.40	?
c)	0.20	0.30	0.40	0.50	?
d)	0	0	1.00	0	?
e)	0.10	0.20	1.20	-1.50	?

- Chapter 14 (Page 380): #11 (continued):

  Which of following probability assignments are possible?

  Strategy:

  1. Check each cell has a number between 0 and 1 (including 0 and 1).

  2. Check the sum is 1.

- Chapter 14 (Page 380): #11 (continued):

Which of following probability assignments are possible?

	red	yellow	green	blue	Valid?
a)	0.25	0.25	0.25	0.25	Yes
b)	0.10	0.20	0.30	0.40	Yes
c)	0.20	0.30	0.40	0.50	No
d)	0	0	1.00	0	Yes
e)	0.10	0.20	1.20	-1.50	No

- Chapter 14 (Page 379): #12:

Which of following probability assignments are possible?

	10% off	20% off	30% off	50% off	Valid?
a)	0.20	0.20	0.20	0.20	?
b)	0.50	0.30	0.20	0.10	?
c)	0.80	0.10	0.05	0.05	?
d)	0.75	0.25	0.25	-0.25	?
e)	1.00	0	0	0	?

- Chapter 14 (Page 379): #12 (continued):

Which of following probability assignments are possible?

	10% off	20% off	30% off	50% off	Valid?
a)	0.20	0.20	0.20	0.20	No
b)	0.50	0.30	0.20	0.10	No
c)	0.80	0.10	0.05	0.05	Yes
d)	0.75	0.25	0.25	-0.25	No
e)	1.00	0	0	0	Yes

Thank you.