Cathy Sun
October 28, 2013
Period 4

+ 2

Hans Rosling: The Magic Washing Machine

His mother and grandmother were so excited and mesmerized to have electricity do the laundry. They thought this machine was a miracle and was doing wonders by taking away so much of the hard and menial labor. At the time of Rosling's talk, there were 7 billion people on Earth. There are two billion people who live on less than $2 a day, while 1 billion people use $80 a day. That means that four billion people are considered "middle class." The washing machine, although thought of to have permeated all of society, only has 2 billion people using it, those who use more than $40 a day. However, the rest of the 5 million people still wash clothes by hand, mostly by poor women. 1/7 of world population uses 6 units of energy out of 12, half of the energy consumption. 2/7 of the world population uses only 1 unit of energy consumption. While giving a talk to hard-core environmentalists and people who "go green," none of them wash their clothes by hand. Population growth and economic growth will increase energy usage. However, population growth will not have as much of an impact as the economy will. As the economy grows, people will generate more income and the newer, more modernist people will double their energy use. Even still, the richest people use the most energy and lecturing them will not have that much of an impact. To change this, we must look towards greener and more sustainable energy sources and products for all so that all of society benefits and decreases energy usage.

Graph?

MEDIA CLIPS solutions

"Weirdness Factor" answers

1. The formula is for standard deviation.

2. No. In some ways the formula suggests population standard deviation (e.g., the denominator being N rather than $n - 1$), but in other ways the formula suggests sample standard deviation (e.g., the use of s instead of σ and the use of $\bar{x}$ instead of μ).

3. Yes. The standard deviation can be interpreted roughly as an average or typical amount of deviation from the norm (in this case, the mean). For example, the first panel could be referring to how people's behavior in a high school can deviate greatly from an outside-of-high-school human behavior norm, with greatest deviation being interpreted as most weird.

4. {0, 0, 0, 0, 0, 10, 10, 10, 10, 10}. This data set yields a sample standard deviation

$$s = \left(\frac{5}{3}\right)\sqrt{10} \approx 5.27$$

or a population standard deviation

$$\sigma = 5.$$

More generally, if the data set contains an even number of elements, the standard deviation of that data set will be maximized when half the elements are of the minimum allowable value and the other half are of the maximum allowable value. (If the data set contains an odd number of elements, the "extra" element can be either the minimum or the maximum value.) To explore this concept, students can start with a data set that can be readily modified in interactive technology such as Excel, a TI-84 calculator, or an applet (e.g., www.stat.tamu.edu/~west/ph/stddev.html; illuminations.nctm.org/LessonDetail.aspx?ID=L449; or illuminations.nctm.org/ActivityDetail.aspx?ID=78).

Another approach is to look at the formula for standard deviation. We can see that standard deviation will be maximized if the squared deviations from the mean are all as great as possible. Without loss of generality, assume a data set with b elements, where b is even. Assume that the greatest standard deviation occurs for some combination of b elements when each element has an extreme value (maximum or minimum). Without loss of generality, assume that the extreme values are zero and a and that we have x elements of value a and $b - x$ elements of value zero. The mean of the b values would be ax/b. The sum of the squares of the deviations from the mean would be

$$x \cdot (a - (ax/b))^2 + (b - x) \cdot (0 - (ax/b))^2 = a^2x - a^2x^2/b.$$

Using calculus or algebra to find the downward-facing parabola's vertex, we see that the sum of the squares of the deviations is maximized when $x = b/2$. Thus, when b is even, half the elements have one extreme value and half have the other.

5. (*a*) Yes. Similar arguments to the previous answer can apply here, but another type of informal reasoning can be used to justify the answer to question 5. The variance is simply the square of the standard deviation, and the square function $f(x) = x^2$ is a monotonically increasing function, so it preserves order (i.e., if $a < b$, then $f(a) < f(b)$). Therefore, whatever data set of values that maximizes the standard deviation will also maximize any monotone function of the standard deviation.

   (*b*) Range is the difference between the greatest and the smallest values of a data set. Therefore, the range of this data set will be maximized if its greatest value is maximized and its smallest value is minimized. Because each value can be any number between 0 and 10, inclusive, we must make sure that the data set contains at least one 0 and at least one 10. The data set in question 4 certainly meets this condition.

   (*c*) Interquartile range (IQR) is the difference between the first and third quartiles of a data set. Therefore, the IQR of this data set will be maximized if its third quartile is maximized and its first quartile is minimized. Because each value can be any number between 0 and 10, inclusive, we must make sure that the data set contains enough 0s and enough 10s for the first and third quartiles to equal 0 and 10, respectively. As students can verify by using the TI-84 calculator (or by hand), the data set {0, 0, 0, 0, 0, 10, 10, 10, 10, 10} meets this condition, but so would {0, 0, 0, 0, a, b, 10, 10, 10, 10} or even {0, 0, 0, a, b, c, d, 10, 10, 10} where $0 \le a \le b \le c \le d \le 10$.

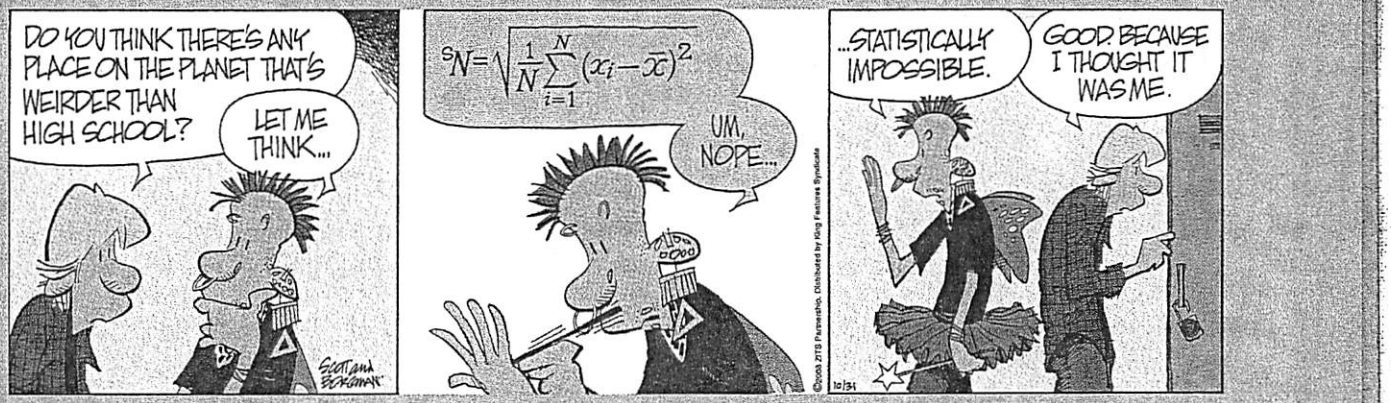
6. The standard deviation has the same units as the original variable x.

7. No. Knowing context is necessary to know what would be a large standard deviation in a particular situation. We could, however, look at a modified measure, such as the coefficient of variation, a unitless measure consisting of the standard deviation divided by the mean.

MEDIA clips

Lawrence M. Lesser // Aaron J. Warner

The High School Weirdness Factor



Zits © Zits Partnership, King Features Syndicate

"Media Clips" aims to offer readers contemporary, authentic applications of quantitative reasoning that are based on print or electronic media. The source material may contain mathematical errors or inaccuracies. We encourage submissions that include questions as well as clips. Submissions must include a complete citation of the source, along with the full text of the cited material. If possible, submissions should also include a copy of the actual clip. Please send submissions to "Media Clips" editor.

Edited by ***Louis Lim***
louis.lim1@gmail.com
Richmond Hill High School
Richmond Hill, Ontario L4S 1A2, Canada

1. What measure of variability does the formula in the second panel convey?
standard deviation
2. Is the notation in this formula consistent?
yes
3. Can this formula be connected to the theme of the first panel? yes, calculating how different school is from normal
4. What set of 10 numbers, with each number being between 0 and 10, inclusive, would *maximize* the value of the measure of variability in question 1?
0,0,0,0,0,10,10,10,10,10
5. Would the answer to question 4 also yield the greatest possible value for—
(*a*) the variance? ✓
(*b*) the range? ✓
(*c*) the interquartile range? ✓

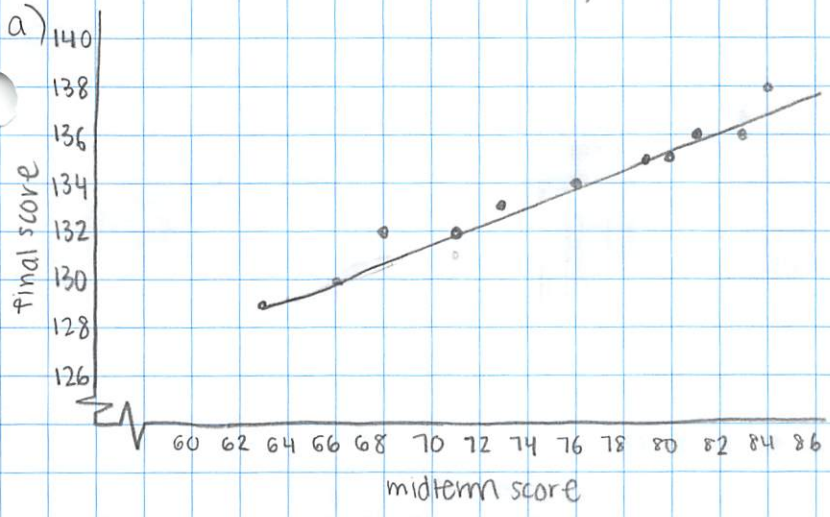
What characteristics of the data set yield the greatest possible value for (*a*), (*b*), and (*c*)?

6. What would be the units of the formula shown in the second panel?
7. Is it possible to specify a large value for a standard deviation?
yes
8. Form a set of 10 numbers (each of which could be any number between 0 and 10, inclusive) so as to *minimize* the value of the measure of variability in question 1.
5,5,5,5,5,5,5,5,5,5
9. Considering the context of the comic strip, determine a possible variable that is represented by *x*.
10. Is there a use of the word *variance* in a nonstatistical context that still involves some aspect of *deviation*?
yes, how strange or weird something is from normal

#13,14 Handout

Cathy Sun
11/13/13

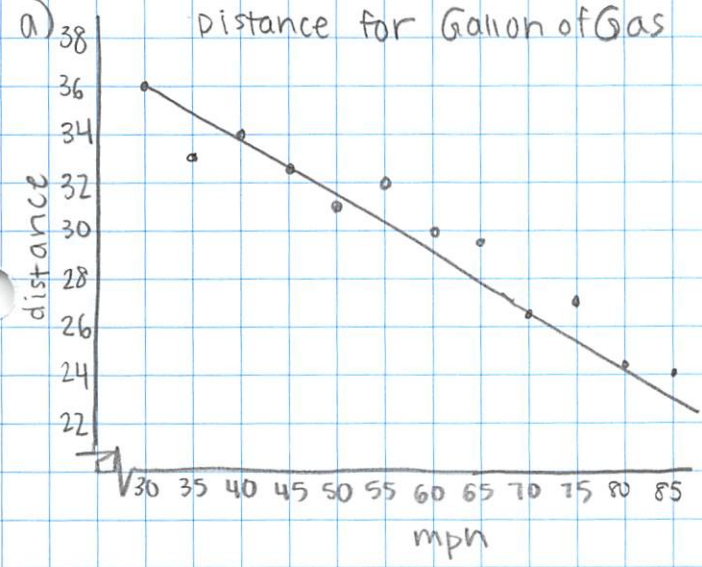
13. a) Scores of Sociology Students



Ana
- positive association
- as midterm scores increased, final exam scores increased as well

c) high, the dots are all pretty close to the line of best fit

14. a) Distance for Gallon of Gas



Ana
- negative association
- as mph increases, the distance traveled decreases

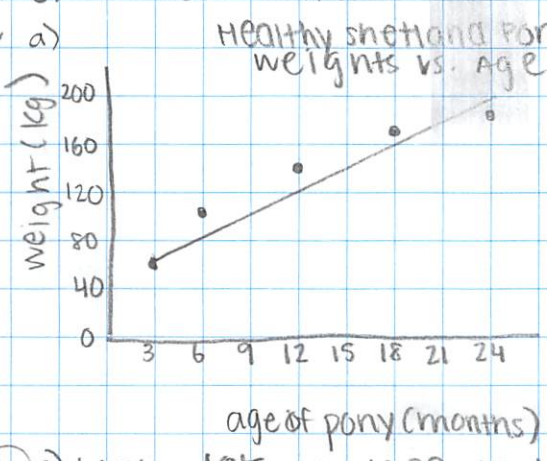
c) high, the values are very close to the line of best fit

Cathy Sun
11/13/13

4.1 #4-7, 10

/10

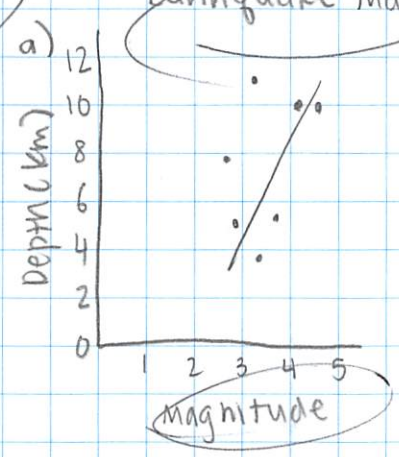
4. b) moderate or low linear correlation
5. a) high linear correlation
6. c) no linear correlation
7. a)



Ans
- positive association
- as age of pony increases, we see an increase in the weight as well

c) high, dots are close to best fit line

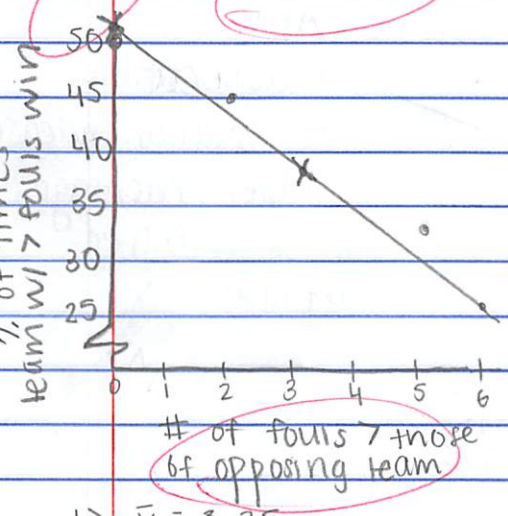
10.



Ans
- + association
- as magnitude increases, we see an increase in depth as well

c) low correlation

4a) Bball Fouls



Ana
- - association
- as a team gets more fouls than opposing team, the % of times it wins decreases

b) $\bar{x} = 3.25$

$\bar{y} = 38.5$

$b = -3.93$

$\hat{y} = 51.29 - 3.93x$

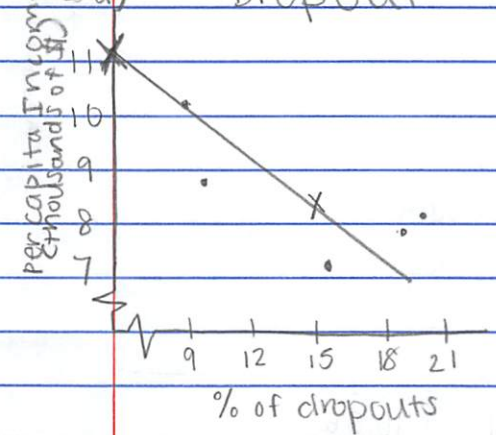
c)

X	Y
0	51.29
3.25	38.5

d) $y(4) = 51.29 - (3.93)(4)$

$= 35.57$ % of times will win

6a) dropouts



Ana
- - association
- as percentages of dropouts increase, per capita income decreases

b) $\bar{x} = 15.02$

$\bar{y} = 8.46$

$b = -0.1759$

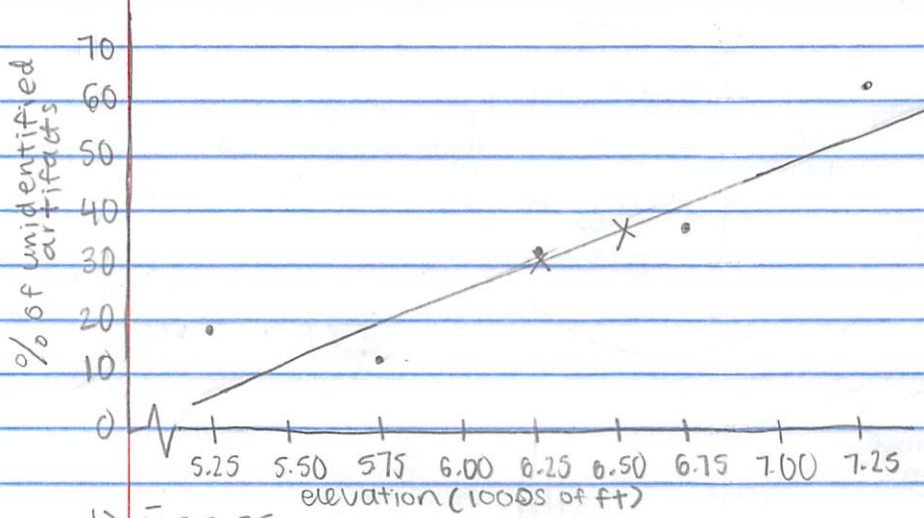
$\hat{y} = 11.1 - 0.1759x$

c)

X	Y
0	11.1
15.02	8.46

d) $y(17) = 11.1 - (0.1759 \times 17) =$ 8.11 → $8,110 per capita income

7a) Artifacts



Ana
- + association
- as elevation increases so does percentage of unidentified artifacts

b) $\bar{x} = 6.25$

$\bar{y} = 32.8$

$b = 22$

$\hat{y} = -104.7 + 22x$

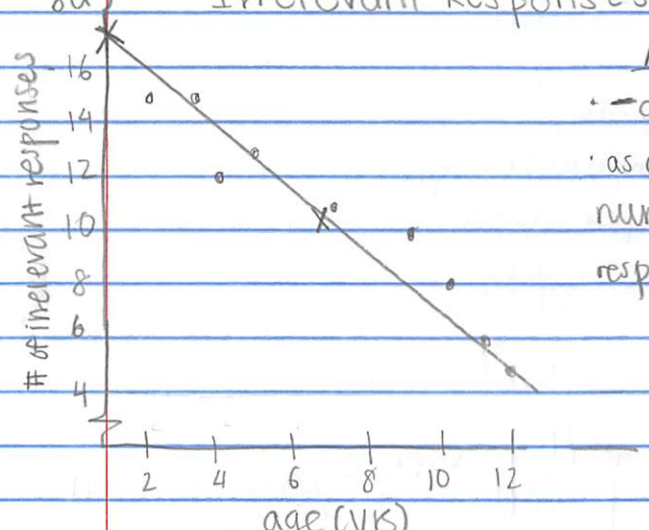
c)

x	y
6.25	32.8
6.5	38.3

d) $y(6.5) = -104.7 + 22(6.5)$

$= 38.3$ % of unidentified artifacts

8a) Irrelevant Responses



Ana
- − association
- as age increases, number of irrelevant responses decreases

c)

x	y
7	10.56
0	17.30

b) $\bar{x} = 7$ $b = -0.96$

$\bar{y} = 10.56$ $\hat{y} = 17.30 - 0.96x$

d) $y(9.5) = 17.30 - (0.96 \times 9.5)$

$= 8.18$ irrelevant responses

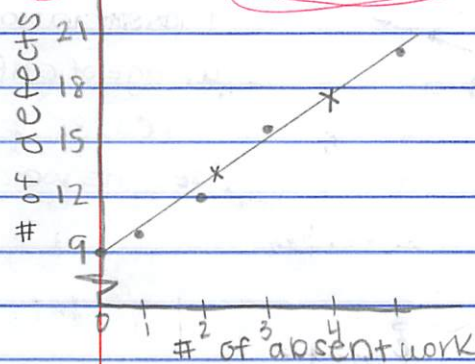
Cathy Sun
11/17/13
P4 - Stats

10
Good

4.2 (1-4, 6-8)

1.

a) Assembly Line Defects



Ana
- positive association
- as number of absent workers increase, number of defects increase as well

b)

x	y	xy	x^2
3	16	48	9
5	20	100	25
0	9	0	0
2	12	24	4
1	10	10	1

$\Sigma x = 11$ $\Sigma y = 67$ $\Sigma xy = 182$ $\Sigma x^2 = 39$

$\bar{x} = 2.2$ $\bar{y} = 13.4$

$SS_{xy} = 182 - \frac{(11)(67)}{5} = 34.6$

$SS_x = 39 - \frac{(11)^2}{5} = 14.8$

$b = \frac{34.6}{14.8} = 2.34$

$a = 13.4 - (2.34)(2.2) = 8.25$

$\hat{y} = 8.25 + 2.34x$

c)

x	y
4	17.61
2.2	13.4

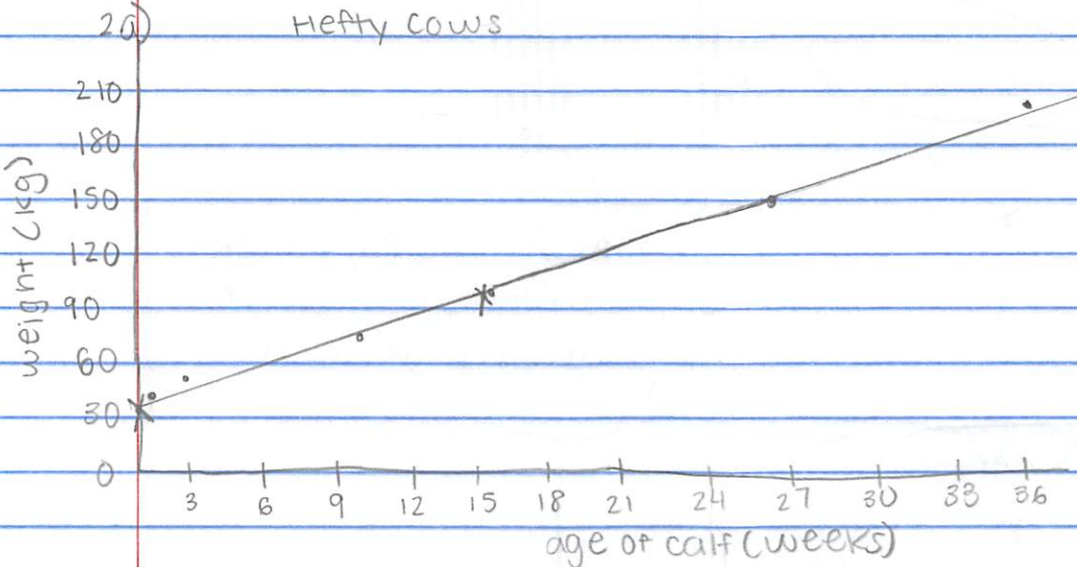
see graph above

d) $y = 8.25 + 2.34x$

$y(4) = 8.25 + 2.34(4)$

$= 17.61$ defects

2a) Hefty Cows



Ana
+ association
· as age of calf increases, so does its weight

b) $\bar{x} = 15.33$

$\bar{y} = 102.83$

$b = 4.51$

$\hat{y} = 33.70 + 4.51x$

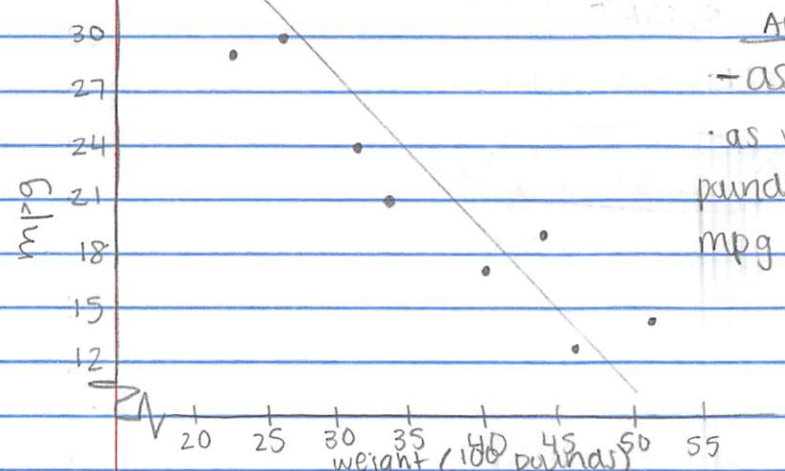
c)

X	Y
0	33.70
15.33	102.83

d) $y(12) = 33.70 + 4.51(12)$

$= 87.82$ kg

3a) Heavy Cars, Gas Guzzlers



Ana
- association
· as weight (hundred pounds) of car increases, mpg decreases

b) $\bar{x} = 37.38$ $b = -0.60$

$\bar{y} = 20.88$ $\hat{y} = 43.33 - 0.60x$

c)

X	Y
20.88	37.38
0	43.33

d) $y(38) = 43.33 - 0.60(38)$

$= 20.53$ mpg

cathy sun
11/19/13
P4-stats

4.3 (1-6, 8)

1a) no

b) more people

2a) no

b) with more money, people might be more inclined to relax and spend more money on liquor

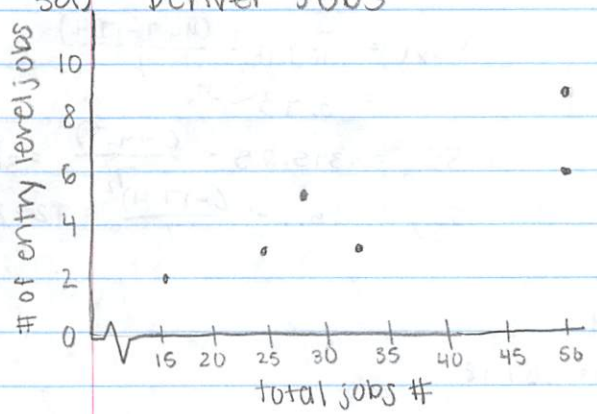
3a) no

b) better medical practices

4a) no

b) increase in population

5a) Denver Jobs



Ana

- + association
- correlation: $r = 0.8602$, high
- $(50,6)$ $(50,9)$ are influential observation or outlier
- trend: as total # of jobs increase, so do # of entry level jobs
- COD = 0.7399

b) closer to 1

c)

x	y	x^2	y^2	xy
16	2	256	4	32
33	3	1089	9	99
50	6	2500	36	300
28	5	784	25	140
50	9	2500	81	450
25	3	625	9	75

$\Sigma x = 202$ $\Sigma y = 28$ $\Sigma x^2 = 7754$ $\Sigma y^2 = 164$ $\Sigma xy = 1096$

$SS_{xy} = \Sigma xy - \frac{\Sigma x \Sigma y}{n}$

$= 1096 - \frac{(202)(28)}{6}$

$= 153.33$

$SS_x = \Sigma x^2 - \frac{(\Sigma x)^2}{n}$

$= 7754 - \frac{(202)^2}{6}$

$= 953.33$

$SS_y = \Sigma y^2 - \frac{(\Sigma y)^2}{n}$

$= 164 - \frac{(28)^2}{6}$

$= 33.33$

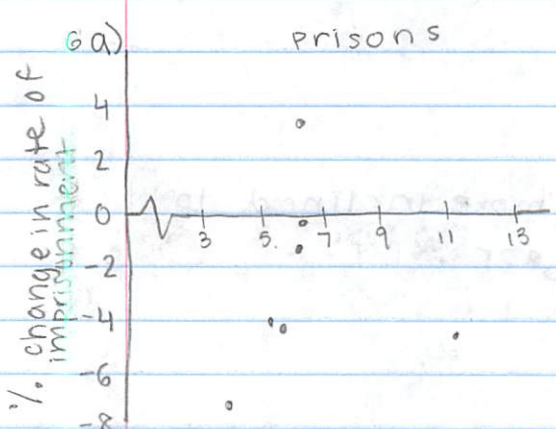
$r = \frac{153.33}{\sqrt{953.33 \times 33.33}}$

$\boxed{r = 0.8602}$ $\boxed{r^2 = .7399}$

- 75.99% of variation in y (# of entry level) jobs is explained by LSRL and variation in x (# of total jobs)

26.01% is unexplained, perhaps due to the economy, kinds of businesses in the area, overall level of education in area

6a) prisons



b) closer to 0

c)

x	y	x^2	y^2	xy
6.1	-1.4	37.21	1.96	-8.54
5.7	-4.1	32.49	16.81	-23.37
3.9	-7.0	15.21	49	-27.3
5.2	-4.0	27.04	16	-20.8
6.2	3.6	38.44	12.96	22.32
6.5	-0.1	42.25	0.01	-0.65
11.1	-4.4	123.21	19.36	-48.84

$\Sigma x = 44.7$ $\Sigma y = -17.4$ $\Sigma x^2 = 315.85$ $\Sigma y^2 = 116.1$ $\Sigma xy = -107.18$

Ana

- + association
- correlation $r = 0.0835$, low
- (11.1, -4.4) is outlier
- as percent change in rate of violent crime increases, so does percent change in imprisonment rate
- COD - $r^2 = 0.0070$

$SS_{xy} = -107.18 - \frac{(44.7)(-17.4)}{7}$

$= 3.93$

$SS_x = 315.85 - \frac{(44.7)^2}{7} = 30.41$

$SS_y = 116.1 - \frac{(-17.4)^2}{7} = 72.85$

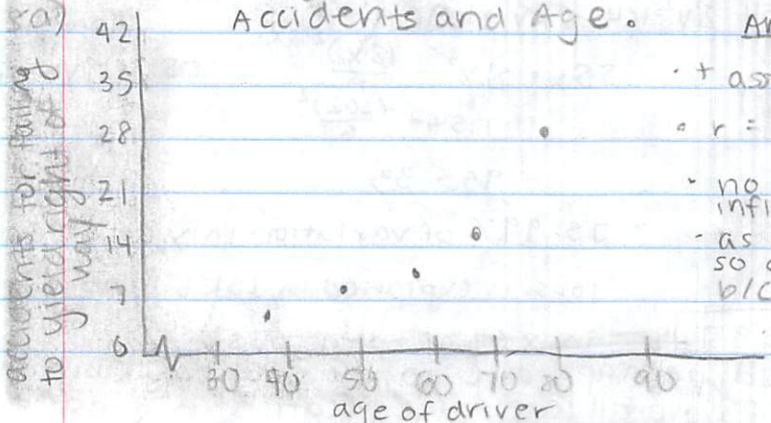
$r = \frac{3.93}{\sqrt{30.41 \times 72.85}} = 0.0835$

$r^2 = 0.0070$

7% of variation in y (% change in rate of imprisonment) can be attributed to variation in x (% change in rate of violent crime) and LSRL.

93% of data is unexplained, perhaps b/c of different personalities, parole options, not thinking of prison as a real consequence for violent crime

8a) Accidents and Age.



Ana

- + association
- r = 0.9428, high
- no outliers or influential observations
- as age of driver increases, so do % of fatal accidents b/c of right of way
- COD - 0.8888

→ 88% of variation in y (% of fatal accidents b/c of failing to have right of way) can be explained by variation in x (age of driver) and LSRL

12% is unexplained maybe because of road conditions, high speed, congestion

Take Worthwhile Notes

SUN

Taking worthwhile notes is one of your critical high school survival skills. It is easy to take good notes.

To begin with, take notes with a purpose other than to just fill up a page with words. No one gives you credit for the number of words in your notebook. You only get credit for a few right words - on exams

The purpose of note taking is to get key points from textbooks and lectures. A key point is an answer to an exam question. Every class day, your teacher is giving you about five to ten key points in the reading assignment and lecture. (Which also means missing class, for whatever reason, is like throwing away a page of valuable notes.)

* Take notes with a purpose

* Take short notes. use key points.

* don't miss class.

Once you have these key points, use them to get exam answers by making up your own set of exam questions. This is exactly what the teacher does in making up an exam, giving us most of the same questions - and the answers, too.

Students at an eastern high school who made up possible test questions later found 75% of the same or similar questions on actual exams. These same students scored almost 10% higher than students who studied without this method.

* Use notes to make up test questions

When studying from textbooks, most of us mark everything we think is important.
A more effective method may be to mark only the parts you don't know. Later on, go over those parts. When you finally know them check them off. There is no need to go over and over information that you already know.

* Make notes of what you don't know in the text.

As with lecture notes, a good way to check how well you understand something is to write down potential test questions about the material just read. Start by briefly summarizing to yourself the meaning of the chapter title and subtitles. Practice for the exam while studying the material.

* Use text to make up test questions.

An easy way to take notes is to draw a line down the center of a sheet. Use the left side for key points from the textbook, which you will read before class. Use the right side to add any important key points the teacher makes if you don't already have them on the left side. This text-lecture process will make class material much easier to understand and add to your retention.

After class, find a quiet place to review your notes. If you don't understand something, ask the teacher right away, or ask someone else in class to give you an explanation. Don't let unanswered questions pile up.

* Review notes immediately after class.

* If you don't understand, ask.

13. The state criminology laboratory must sometimes estimate a person's height from partial skeleton remains. How strong a correlation is there between body height and bone size? A random sample of eight adult male x-rays gave the following information, where x = length of femur (thigh bone) and y = body height.

x (*in.*)	17.5	20	21	19	15.5	18.5	16	18
y (*in.*)	50	80	78	73	63	71	64	71

Complete parts a through c for these data, and comment on the meaning of r and r^2 in the context of this problem.

14. Seven children in the third grade were given a verbal test to study the relationship between the number of words used in a response and the silence interval before the response. A psychologist asked each child a number of questions, and then the total number of words used in answering and the time before the child began answering were recorded. The following data were obtained, where x = total silence time in seconds and y = total number of words in response.

x	25	16	19	27	38	42	31
y	60	55	50	64	73	70	65

Complete parts a through c for these data, and comment on the meaning of r and r^2 in the context of this problem.

15. The following data are taken from the *Statistical Abstract of the United States*. In this data, x = *M1* (readily available cash) in tens of billions of dollars and y = passenger car sales (in millions of cars).

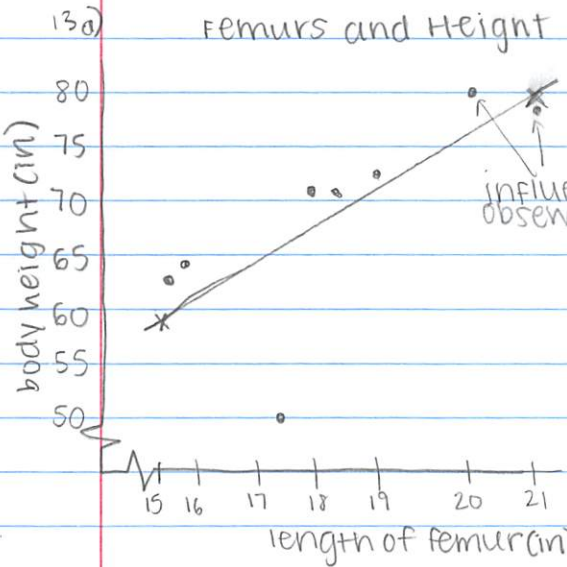
x	4.1	4.4	4.8	5.2	5.5	6.2	7.3	7.5
y	6.4	6.3	5.0	6.7	7.6	8.0	7.5	7.1

Complete parts a through c for these data, and comment on the meaning of r and r^2 in the context of this problem.

cathy sun, Wendy Xiao

11/22/13
P4-Stats

Skeletal Remains



Ana
- + association
- correlation $r = 0.7061$, high
- outliers (17.5, 50)
- as length of femur increases (inches), the body height goes up (3.632 inches)
- COD = 0.4985

b) $\hat{y} = a + bx = 2.690 + (3.632)x$

$r = 0.70608$

$r^2 = 0.4985$

x	y
18.19	68.8
15.5	58.99
21.0	79.0

c) approximately 50% of the variation in y (body height) can be explained by LSRL and variation in x (femur length)

Cathy Sun
11/22/13
P4-Stats

WHAT WAS THE MOST POPULAR SONG OF 2011?

10

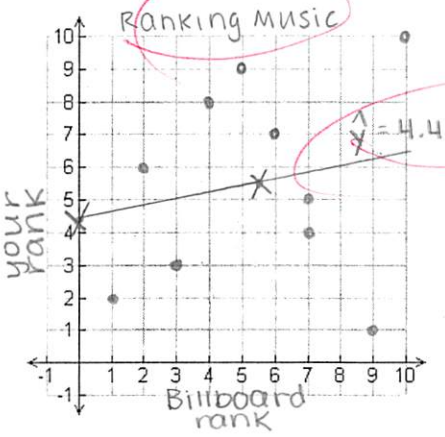
We use an alphabetical list of songs that were top songs on Billboard's weekly charts to explore this question.

A. E.T.
B. Firework
C. Forget You
D. Give Me Everything
E. Grenade
F. Just Can't Get Enough
G. Moves Like Jagger
H. Party Rock Anthem
I. Rolling In the Deep
J. Super Bass

Rank the songs from 1 for first place to 10 for tenth place. When you have finished your rankings, Mr. Micek will give you the Billboard rankings.

Song	Billboard Rank	Your Rank
A.	4	8
B.	3	3
C.	7	4
D.	5	9
E.	6	7
F.	10	10
G.	9	1
H.	2	6
I.	1	2
J.	8	5

Plot your results on the grid. (x-axis is billboard rank, y-axis is your rank)



X	Y
0	4.4
5.5	5.5

Ana

- + association
- 0.2, low
- (10,10), (9,1), (5,9), (4,8) are outliers
- as Billboard rank increases, my rank increases as well
- COD = r^2 = 0.04

Use your calculator to determine and explain:

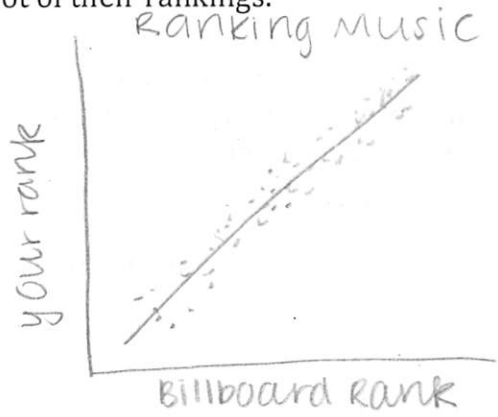
R= 0.2

R^2= 4 %

4% of the Ⓥ in y (my rank) can be explained by LSRL and Ⓥ in x (Billboard rank).
96% unexplained, perhaps because of personal music taste or radio choice

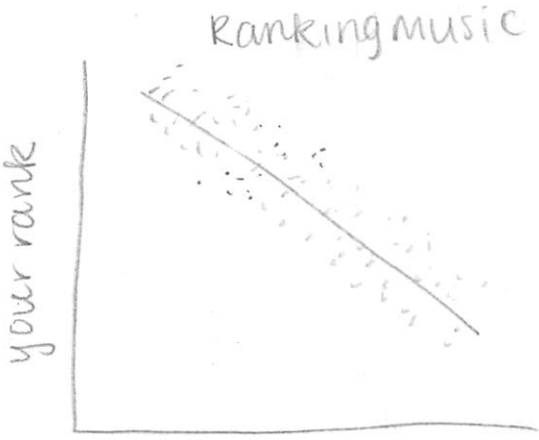
Extension: You know that Aileen had a measure of 0.86 while Mei had -0.76 and Juanita had 0.24, sketch a plot of their rankings.

Aileen



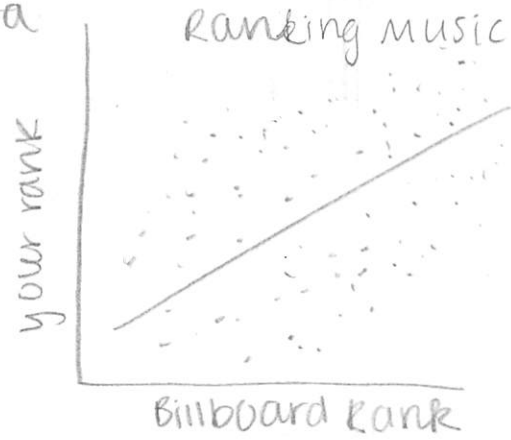
$r = 0.86$

Mei



$r = -0.76$

Juanita



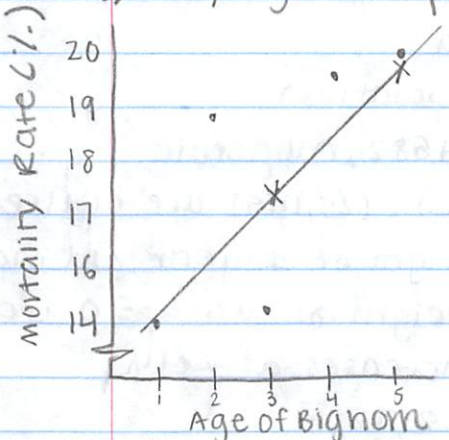
$r = 0.24$

Cathy Sun
11/21/13
P4-Stats

10

Ch 4 RVW 1-3, 6

1a) Dying sheep



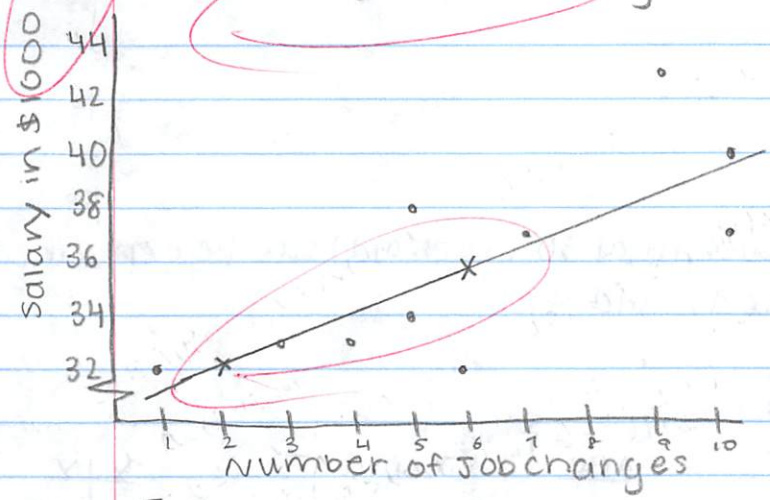
Ana
- + association
- r = 0.6847 moderate
- (2, 18.9), (3, 14.4) outliers
- as age of bighorn sheep increases, so does mortality rate in percent
- COD = 0.4688 ↓ 46.88% of variation in y (mortality rate) can be explained by LSRL and variation in x (age of bighorn)

x	y
3	17.38
5	19.92

b) $\hat{y} = 13.57 + 1.27x$

c) $r = 0.6847$
$r^2 = 0.4688$

2a) Job changes and salary



Ana
- + association
- r = 0.7609, high
- ins (10, 40) out (9, 43), (10, 37)
- as number of job changes increases, so does salary in $1000
- $r^2 = 0.5789$, 57.89% of variation in y (salary) can be explained by LSRL and ⓥ in x (number of job changes)

x	y
6	35.9
2	32.15

b) $\bar{x} = 6$
$\bar{y} = 35.9$
$b = 0.9390$

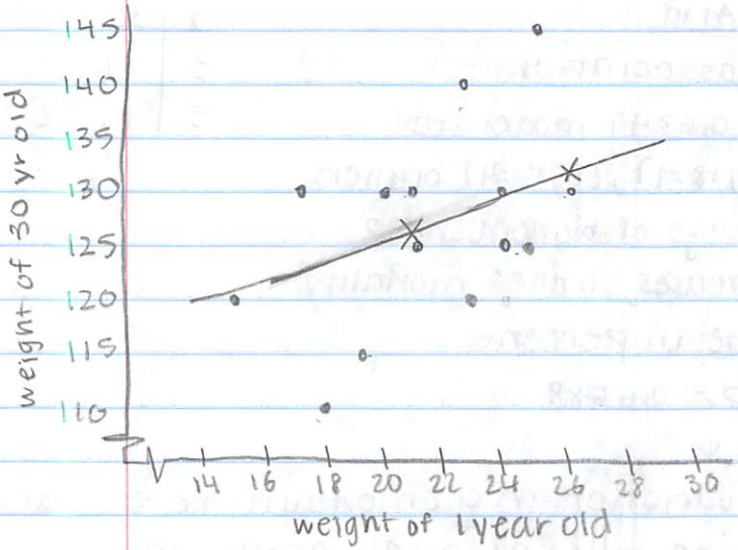
c) $\hat{y} = 30.27 + 0.9390x$

d) $\hat{y}(2) = 30.27 + 0.9390(2) = \$32150$ salary

e) positive

f) $r = 0.7609$
$r^2 = 0.5789$

3a) Fat Babies → Adults



Ana
- + association
- $r = 0.4682$, moderate
- (18,30), (25,145) are outliers
- As weight of 1 year old increases, the weight of same 30 year old increases as well
- $r^2 = 0.2192$

b) $\bar{x} = 21.43$

$\bar{y} = 126.79$

$b = 1.28$

x	y
21.43	126.79
26	132.53

c) $\hat{y} = 99.25 + 1.28x$

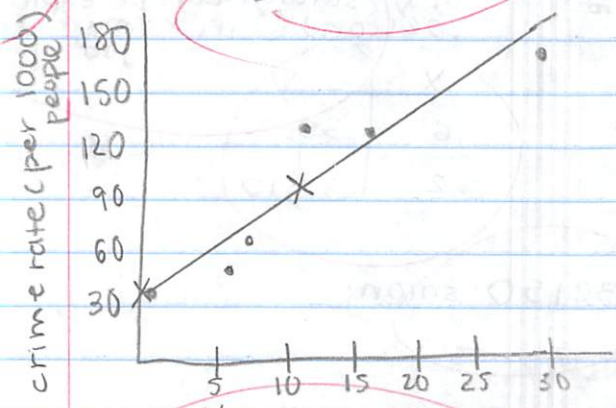
d) 124.85 lbs

e) positive

f) $r = 0.4682$, $r^2 = 0.2192$

21.92% of variation in y (weight at 30 years old) can be explained by (N) in x (weight of 1 year old) and LSRL

6a) Changing pop and crime



Ana
- + association
- $r = 0.9267$
- (29, 173) : influential observation
- As percent change in population increases, so does crime rate (per 1000 population)
- $COD = 0.8588$

x	y
0	36.88
12	98.17

b) $\hat{y} = 36.88 + 5.11x$

c) 98.2 crime rate (per 1000 people)

d) $r = 0.9267$

e) Doesn't guarantee cause and effect because there will always be some part left unexplained since r^2 almost always < 1

Zainab Pi
Cathy Su

1, 15-20

1. E
15. $r^2 = 0.399$
C
16. ~~$r^2 = 96.28$~~ A
C
17. $\hat{y} = 7386.87 - 3.63x$
E
18. B
19. E
20. I, II, III