Statistics 1040, Section 006, Midterm 1 (200 Points)

Friday, October 8, 2004

Your Name: _____

from: Stat 1040, Fall 2001, Final Test, December 13, 2001, Question 1
& Stat 1040, Fall 2003, Midterm 1, October 3, 2003, Question 3

Question 1: Controlled Experiment/Observational Study (**60 Points**)

A recent study in Europe looked at a large group of women of childbearing age. The researchers asked each woman how much alcohol they had consumed over the past 12 months. The researchers found that women who drank moderate amount of alcohol were somewhat less likely to have infertillity problems than women who did not drink alcohol at all (November, 2001). The study said it "controlled for age, income, and religion".

1. (**15 Points**) Based on the information above, was this a controlled experiment or an observational study? **Explain briefly.**

(10)

(5) correct explanation
(1) some explanation

Answers from Course Web Page: no intervention was used - nobody was told to drink/not to drink

2. (**15 Points**) Why did they "control for" age, income, and religion?

these factors may be confounding factors

-5 for missing keyword (but otherwise correct explanation)

3. (**15 Points**) Is this convincing evidence that infertility would decrease if women with infertility problems started to drink moderate amounts of alcohol? (Note: we are only asking about infertility. There may be other problems introduced by such behavior, but ignore them for answering this question).

(10) No! - we only know that there is association between drinking and fertility; drinking does not cause fertility
[association is not causation!]

(5) correct explanation
(1) some explanation

4. (**15 Points**) Suggest a possible confounding factor (other than age, income, or religion) and **clearly explain** why you think it might be a confounding factor.

general health (condition):
someone who has some other medical problem may not drink and also be less fertile

(10) for correct confounding factor
(5) for correct explanation
(1) for some explanation

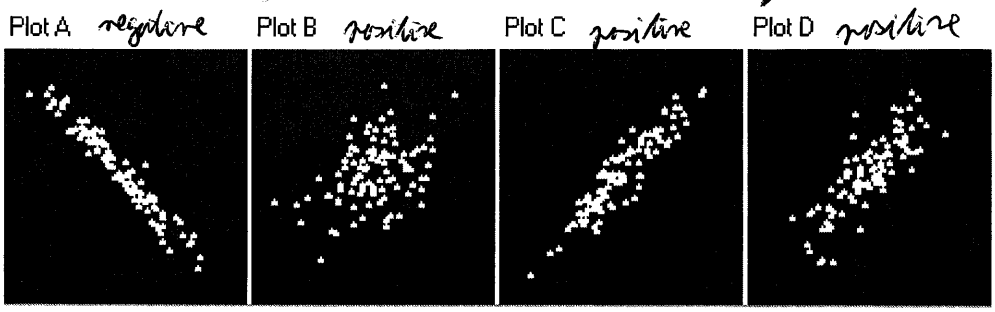
Question 2: Correlation **(30 Points)**

from: Stat 1040, Fall 2002, Midterm 1, October 4, 2002, Question 4 (different plots)
& Stat 1040, Spring 2004, Midterm 1, February 13, 2004, Question 2

1. **(12 Points)** Match each of the scatterplots with **one** of the correlation coefficients below:

straight falling line

-1.00, -0.97, 0.46, 0.82, 0.93, 1.05



(3) • Correlation for Plot A: $r =$ -0.97

(3) • Correlation for Plot B: $r =$ 0.46

(3) • Correlation for Plot C: $r =$ 0.93

(3) • Correlation for Plot D: $r =$ 0.82

Explanation (not required for your answer):
A: strong negative, but not a straight falling line
C, D, B: positive; with C strongest, then D, finally B

There is **no** explanation required for this question.

from: FPP, p. 143-144, Exercise 7, Answers → FPP, p. A57

2. **(18 Points)** Two different investigators are working on a growth study. The first measures the heights of 100 children in inches. The second prefers the metric system, and changes the results to centimeters (multiplying by the conversion factor 2.54 centimeters per inch). A scatter diagram is plotted, showing for each child its height in inches on the horizontal axis, and height in centimeters on the vertical axis.

(a) **(6 Points)** If no mistakes are made in the conversion, what is the correlation?

below -1.0, -1.0, close to -1.0, 0.0, close to 1.0, (1.0,) above 1.0 (6)

(b) **(6 Points)** What happens to r if mistakes are made in the arithmetic?

r goes up, (r goes down,) (6) r stays the same

(c) **(6 Points)** What happens to r if the second investigator goes out and measures the same children again, using metric equipment?

r goes up, (r goes down,) (6) r stays the same

[it is practically impossible to re-measure 100 children without measurement error!]

Circle your answers — There is **no** explanation required for this question.

from: FPP, p.176, Review Exercise 3, (a) & (d)
& Stat 1040, Fall 2002, Midterm 1, October 4, 2002, Question 5, 1., 3. & 4. (different value)

Question 3: Regression **(30 Points)**

Pearson and Lee obtained the following results in a study of about 1,000 families:

-2 for each calculation error

average height of husband $\approx$ 68 inches, SD $\approx$ 2.7 inches
average height of wife $\approx$ 63 inches, SD $\approx$ 2.5 inches,
$r \approx 0.25$

1. **(10 Points)** Predict the height of a wife when the height of her husband is 72 inches. Answer: 64"

Answers from Course Web Page/ Workbook:

x: husband
y: wife

$su_x = \frac{x - avg_x}{SD_x} = \frac{72-68}{2.7} = 1.48$ (3)

$su_y = r \cdot su_x = 0.25 \cdot 1.48 = 0.37$ (3)

$y = su_y \cdot SD_y + avg_y = 0.37 \cdot 2.5 + 63 = 63.925 \approx 64$ inches (4)

2. **(10 Points)** Predict the height of a wife when the height of her husband is unknown. Answer: 63"

63 inches! (8)

(the best guess is the average if nothing else is known) (2)

3. **(10 Points)** Predict the height of a husband when the height of his wife is 65 inches. Answer: 68.54"

x: wife
y: husband

$su_x = \frac{65-63}{2.5} = 0.80$ (3)

$su_y = 0.25 \cdot 0.80 = 0.20$ (3)

$y = 0.20 \cdot 2.7 + 68 = 68.54$ inches (4)

Show your work!

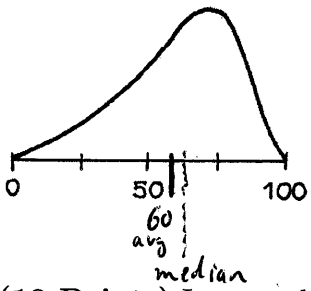
from: FPP, p. 75, Review Exercise 6
& Stat 1040, Fall 2003, Quiz 2, September 12, 2003, Question 1.2

Question 4: Average and Standard Deviation **(40 Points)**

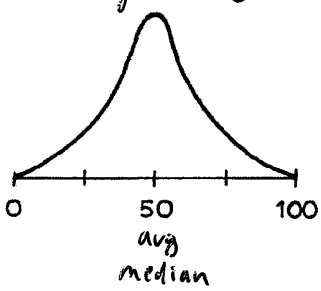
Below are sketches of histograms for three lists.

answers from course Web Page/ Work book:

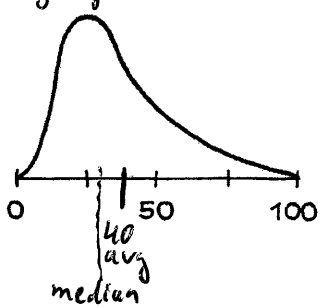
(i) long left tail



(ii) symmetric



long right tail (iii)



1. **(10 Points)** In scrambled order, the averages are 40, 50, 60. Match the histogram with the averages:

   Histogram (i): average = 60 (4)

   Histogram (ii): average = 50 (2)

   Histogram (iii): average = 40 (4)

   Just fill in the correct value.

2. **(10 Points)** Match the histogram with the description (circle your answer):

   - The median is less than the average. Histogram (i), (ii), or (iii). (4) [long right tail]
   - The median is about equal to the average. Histogram (i), (ii), or (iii). (2) [symmetric]
   - The median is bigger than the average. Histogram (i), (ii), or (iii). [long left tail] (4)

   (10)

3. **(10 Points)** Is the SD of histogram (iii) around 5, 15, or 50? Circle your answer.

   [15, because most of the area is within 50 of the average, so 50 is too big; and only a small portion of the area is within 5 of the average, so 5 is too small]

4. **(10 Points)** The SD for histogram (i) is a lot smaller than that for histogram (iii). **True** or **false**? **Circle your answer and explain:** (8)

   The two histograms are almost mirror images and have about the same SD. (2)

Question 5: Normal Distribution **(40 Points)**

The stretched hand spans of women have a distribution that is approximated by a normal curve with a mean of 20 centimeters and a standard deviation equal to 1.5 centimeters.

1. **(15 Points)** For what proportion of women is the stretched hand span greater than 23 centimeters?
   Answer: 2.28%

(3) for graph only

$\frac{23-20}{1.5} = 2.00$ s.u. (5)

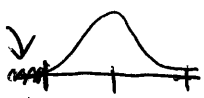
area between −2.00 and 2.00: 95.45% (5)

area above 2.00: $\frac{100\% - 95.45\%}{2} = \frac{4.55\%}{2} = 2.275\% \approx 2.28\%$ (5)

2. **(10 Points)** Would it be unusual for a woman to have a stretched hand span that's less than 14 centimeters? Support your answer with a calculation.

   Answer: **yes, unusual** / **No, not unusual** (4)

   Calculation/Explanation:



(3) for graph only

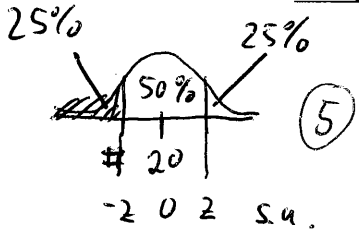
$\frac{14-20}{1.5} = -4.00$ s.u. (2)

area between −4.00 and 4.00: 99.9937% (2)

area below −4.00: $\frac{100\% - 99.9937\%}{2} = \frac{0.0063\%}{2} = 0.00315\% \approx 0\%$ (2)

very unusual!

3. **(15 Points)** What stretched hand span is the 25th percentile of the stretched hand spans of women?
   Answer: about 19.0 cm



(5)

Show your work!

area between −0.65 to 0.65: 48.43% ← closest to 50%

area between −0.70 to 0.70: 51.61%

we need $\#_{s.u.} = -0.65$ (or −0.70) in s.u. (5)

in original units: (5)

$\# = -0.65 \cdot 1.5 + 20 = -0.975 + 20 = 19.025$

(or $\# = -0.70 \cdot 1.5 + 20 = -1.05 + 20 = 18.95$)