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

October 4, 2002

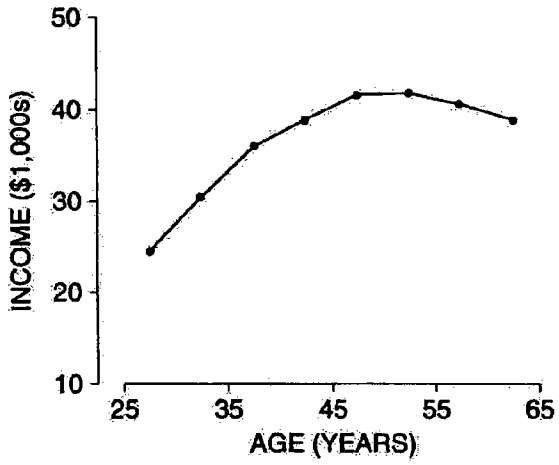
Your Name: _____

From: Freedman, Pisani, Purves, "Special Review Exercises", page 265, question 6

Question 1: Representative Sample (**25 Points**)

The graph below summarizes the results from a study based on a representative sample of men age 25–64 in 1993, who were working full time that year; the graph shows average income for each age group.

True or false, and **explain**: the data show that on average, if a man keeps working, his income will increase until age 50 or so, then start decreasing. If false, how do you account for the pattern in the data?



Source: March 1993 Current Population Survey; CD-ROM supplied by the Bureau of the Census through the U.C. Survey Research Center.

(*) Also, the overall education level has increased over the last 50 years, resulting in better-paid jobs for younger people. People with higher income might also retire a bit earlier than people with less income – therefore resulting in a lower average towards age 65 (but people getting close to 65 will not earn less!).

(10) (10)

False – this is a cross-sectional study. To find out whether men have a decrease in income after an age of 50 years, we need a longitudinal study, e.g., follow a group of men for several years and obtain their income and see if those above 50 have a decline in income. Alternatively, we might ask men of age 50 and above about their income during the last 15 years and see whether this did decline (although this might be very unreliable – who keeps IRS records for 15 years or still remembers his income from 15 years ago).

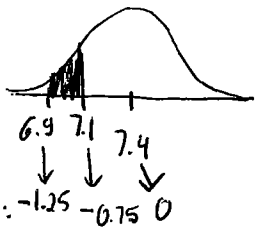
(5) The pattern in the plot might be explained by increasing starting salaries: Young men today might earn as much as men that started their worklife 40 years ago had earned 10 or 20 years into their career. (*)

Question 2: Normal Distribution (**45 Points**)

According to the U.S. Department of Agriculture, one 3–ounce serving of trimmed sirloin beef contains, on average, 7.4 grams of fat. Assume that the amount of fat for such servings closely follows the normal curve, with a standard deviation of 0.4 gram.

Fill the blanks in the statements below and **show all the work needed to obtain the answer:**

1. (**15 Points**) The proportion of servings that contain between 6.9 and 7.1 grams of fat is 12.1%.



3 for graph only (in each part)

$\frac{7.1-7.4}{0.4} = -0.75$ s.u. (3)

$\frac{6.9-7.4}{0.4} = -1.25$ s.u (3)

from Table: area between −0.75 and 0.75: 54.67%

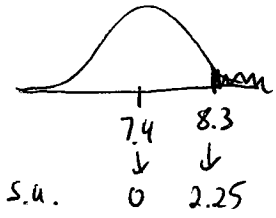
area between −1.25 and 1.25: 78.87%

area between −0.75 and 0: 54.67% / 2 = 27.335% (3)

area between −1.25 and 0: 78.87% / 2 = 39.435% (3)

area between −1.25 and −0.75: 39.435% − 27.335% = 12.1% (3)

2. (**15 Points**) The proportion of servings that contain more than 8.3 grams of fat is 1.22%.



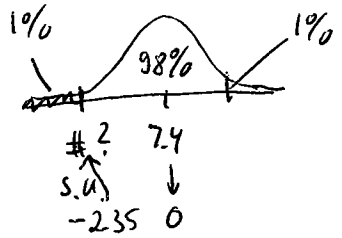
$\frac{8.3-7.4}{0.4} = 2.25$ s.u. (5)

from Table: area between −2.25 and 2.25: 97.56% (5)

area above 2.25: $\frac{100\% - 97.56\%}{2} = 1.22\%$ (5)

−3 if not divided by 2

3. (**15 Points**) Only 1% of the servings will contain less than 6.46 grams of fat.



from Table:

area between −2.30 and 2.30: 97.86% (5)

area between −2.35 and 2.35: 98.12%

we need $\#_{s.u.} = -2.35$ (or −2.30) in s.u. (5)

in original units:

$\# = -2.35 \cdot 0.4 + 7.4 = 6.46$ (5)

(or $\# = -2.30 \cdot 0.4 + 7.4 = 6.48$)

From: Stat 1040, Fall 1999, Final Test, December 17, 1999. Question 1 a,b

Question 3: Controlled Experiment/Observational Study (**30 Points**)

As part of a study on exercise and health, a group of 1,000 people was followed for 5 years. At the beginning of the study, the researchers asked each person whether they exercised regularly or not. At the end of the study, the researchers measured several health–related variables, and in doing so, they noticed that the death rate for the exercise group was much lower than for the no–exercise group.

(7)

1. (**10 Points**) Is this an example of a controlled experiment or an observational study? **Explain.**

(3) correct explanation
(1) none 11

Worklook:

They were asked what they did - they weren't told what to do.

2. (**20 Points**) Does this result necessarily imply that if people who do not exercise start to exercise regularly, they will live longer, on average, than if they do not? **Explain clearly.**

(10) correct explanation
(5)(1) some 4

Worklook:

(10)

No - this result merely shows that the type of people who choose to exercise differ from the type who do not. This is an observational study and there could be many confounding factors. For example, people who exercise might smoke less and it might be smoking that's harmful not exercise that's beneficial. Taking a smoker & making them exercise might have no impact

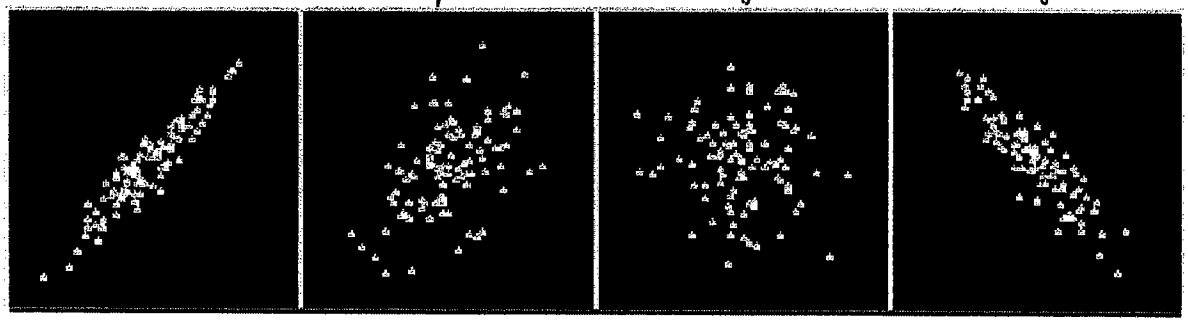
From: Stat 1040, Spring 2002, Final Test, April 30, 2002. Question 6 (different plots)

Question 4: Guessing the Correlation Coefficient (**40 Points**)

Match the four scatterplots with their correlations from the list:

~~-1.03~~, -0.99, -0.89, -0.50, -0.05, 0.50, 0.93, ~~1.03~~

Plot A positive Plot B positive Plot C negative Plot D negative



(10) Correlation for Plot A: r = 0.93

(10) Correlation for Plot B: r = 0.50

(10) Correlation for Plot C: r = −0.05

(10) Correlation for Plot D: r = −0.89

for each plot:
5 for 1 level off, i.e.,
A: 0.50
B: −0.05, 0.93
C: −0.50, 0.50
D: −0.50, −0.99
2 for correct direction but 2 or more levels off
0 for −1.03 or 1.03

Explanation (not required for your answer):

−1.03 and 1.03 are absolutely impossible as values for the correlation coefficient r; only values between −1 and 1 are possible

Plots A and B show a positive association; Plots C and D a negative association.

Plot A shows the strongest association overall: ⇝ 0.93

Plot D shows an association that is slightly weaker than in A: ⇝ −0.89

Plot B shows a weak positive association: ⇝ 0.50

Plot C shows almost no association: ⇝ −0.05

From: Freedman, Pisani, Purves, Chapter 10, page 176, question 3 a, c, d

Question 5: Regression **(60 Points)**

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

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.

Show all the work needed to obtain the answer.

1. **(15 Points)** Predict the height of a wife when the height of her husband is 72 inches.

X: husband
Y: wife

$$s.u._X = \frac{X - avg_X}{SD_X} = \frac{72-68}{2.7} = 1.48$$

$$s.u._Y = r \cdot s.u._X = 0.25 \cdot 1.48 = 0.37$$

$$Y = avg_Y + s.u._Y \cdot SD_Y = 63 + 0.37 \cdot 2.5 = 63.925 \approx 64 \text{ inches}$$

15 for correct result + work
5 for each correct step
1 for incorrect result (and no work)

2. **(15 Points)** Predict the height of a wife when the height of her husband is 68 inches.

nothing to calculate - the regression line goes through the point of averages, i.e., for an average husband, it predicts an average wife, i.e., 63 inches

as above in 1.
or with explanation

3. **(15 Points)** Predict the height of a wife when the height of her husband is unknown.

63 inches!

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

10 correct result
5 explanation

4. **(15 Points)** Predict the height of a husband when the height of his wife is 68 inches.

X: wife
Y: husband

$$s.u._X = \frac{68-63}{2.5} = 2$$

$$s.u._Y = 0.25 \cdot 2 = 0.5$$

$$Y = 68 + 0.5 \cdot 2.7 = 69.35 \text{ inches}$$

as above in 1.