

Statistics 1040, Section 004, Midterm 1 (200 Points)

Friday, February 13, 2004

Your Name: _____

From: Stat 1040, Summer 2000, Test 1, June 26, 2000, Question 1
Question 1: Observational Studies and Experiments (50 Points)

Forty men have agreed to be subjects in an experiment on the effectiveness of a new throat spray that is supposed to reduce snoring. These people were randomly divided into a treatment group and a control group.

1. (10 Points) Is the study described in this article an **observational study** or a **controlled experiment**? Circle your answer and explain!

People were assigned to a treatment (the new throat spray) or a placebo, ^{(2) explanation}
i.e., there is intervention.

2. (15 Points) When the groups were chosen, each of 40 men was given a spray bottle and told to use it every night for a week. Why are *all* the men given spray bottles? Are all the contents the same? Explain! ^{(5) explanation}

They are all given spray bottles to make the experiment blind, i.e., to minimize the effects of people reacting to the idea of treatment rather than the treatment itself.

The placebo group (control group) gets a placebo, i.e., a spray that lacks the active ingredients ⁽⁵⁾

3. (15 Points) To find out whether the spray works, at the end of the study the men were asked the following question:

Do you think that with the spray you snore more than before, less than before, or about the same as before?

It turns out that on average the men in *both groups* thought that they snored less than before. One explanation why the treatment group might answer this way is that the spray works. But why did the men in the control group answer this way? Provide two plausible reasons.

- They thought they were getting the treatment so they only thought they snored less.
- The placebo had some effect (e.g., moisturizing the throat)
- They actually snored less because of psychosomatic effects or other confounding factors (such as sleeping on back/side, length of sleep, etc.) ^{- 1 if only 1 reason}

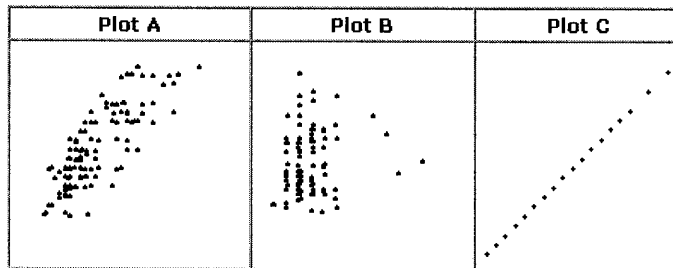
4. (10 Points) Suggest a better way to evaluate the effectiveness of the spray. Record the snoring or have an observer measure it. Self-reporting is quite unreliable here! Obviously, they should record (observe) before and after the experiment and compare the outcomes.

Question 2: Correlation (30 Points)

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

1. (18 Points) The following three plots are based on the heights and weights of 126 college students. Match each plot with the description:

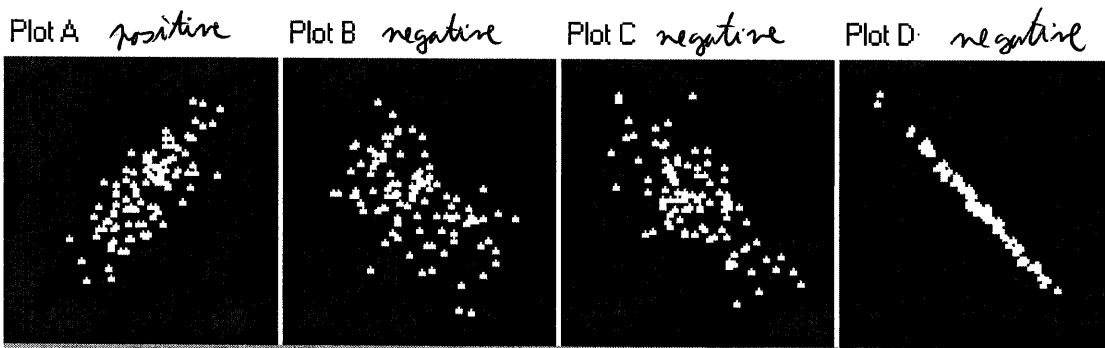
- (a) X = Weight in pounds and Y = Height in inches for the 126 people;
- (b) X = Age in years and Y = Weight in pounds;
- (c) X = Weight in pounds and Y = Weight in kilograms for the 126 people.



- ⑥ • Plot A: (a) [medium (strong) association between height & weight]
- ⑥ • Plot B: (b) [weak (none) association between age (here 18-30) & weight]
- ⑥ • Plot C: (c) [perfect correlation (change of scale!).]

2. (12 Points) For the following plots, match each plot with one of the following correlation coefficients:

~~-1.03~~, -0.99, -0.70, -0.50, 0, 0.20, 0.76, 0.99



- ③ • Correlation for Plot A: $r = \underline{0.76}$
- ③ • Correlation for Plot B: $r = \underline{-0.50}$
- ③ • Correlation for Plot C: $r = \underline{-0.70}$
- ③ • Correlation for Plot D: $r = \underline{-0.99}$

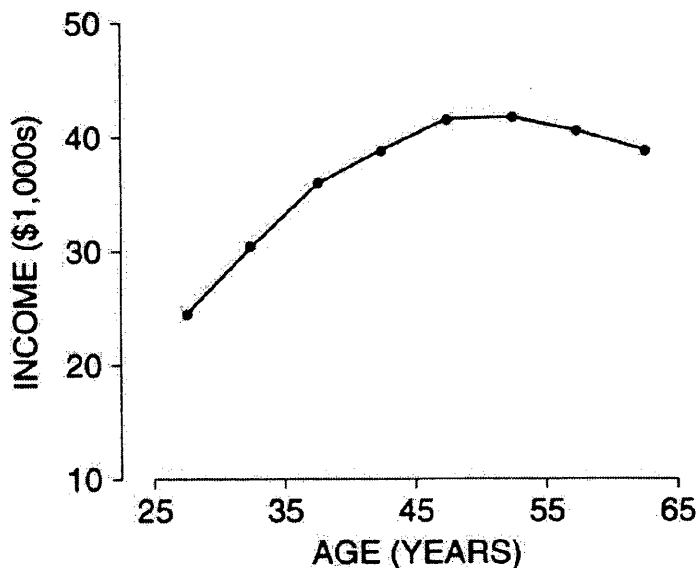
Explanation (not required):
 -1.03 impossible (r takes values between -1 and 1)
 Plot A shows positive association, all other plots show negative association.
 Plot D: strongest negative association $\rightarrow -0.99$
 Plot C: 2nd strongest negative association $\rightarrow -0.70$
 Plot B: 3rd strongest negative association $\rightarrow -0.50$
 Plot A: somewhat strong positive association than C's 0.76

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

also: Stat 1040, Midterm 1, October 4, 2002, Question 1

Question 3: Representative Sample (35 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.



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

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?

15

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).

10

explanation:

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 careers.

Also, the overall educational level has increased over the last 50 years, resulting in better-paid jobs for younger people.

Perhaps people with higher income might also retire a bit earlier than people with less income - therefore resulting in a lower average towards age 65.

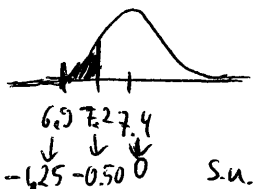
But, people close to age 65 on average will not earn less than 5 or 10 years ago!

From: Stat 1040, Midterm 1, October 4, 2002, Question 2 (different values)

Question 4: Normal Distribution (45 Points)

According to the U.S. Department of Agriculture, a 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. Find the percentage of servings that contain:

1. (15 Points) Between 6.9 and 7.2 grams of fat: 20.29 % of servings



$$\frac{6.9-7.4}{0.4} = -1.25 \text{ s.u. } \textcircled{3}$$

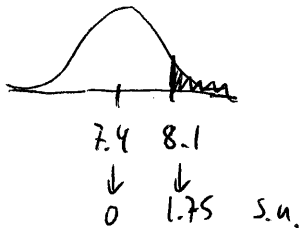
$$\frac{7.2-7.4}{0.4} = -0.50 \text{ s.u. } \textcircled{3}$$

area between -1.25 and 1.25: 78.87% $\textcircled{3}$ area between -0.50 and 0.50: 38.29% $\textcircled{3}$

$$\text{area between } -1.25 \text{ and } -0.50: \frac{78.87\% - 38.29\%}{2} = \frac{40.58\%}{2} = \underline{\underline{20.29\%}} \textcircled{3}$$

$\textcircled{3}$ for graph only
(in each part)

2. (15 Points) More than 8.1 grams of fat: 4.0 % of servings

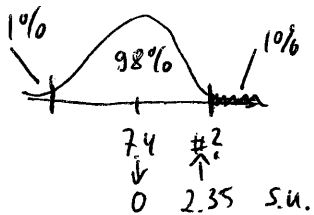


$$\frac{8.1-7.4}{0.4} = 1.75 \text{ s.u. } \textcircled{5}$$

area between -1.75 and 1.75: 91.99% \approx 92% $\textcircled{5}$

area above 1.75: $\frac{100\% - 92\%}{2} = \frac{8\%}{2} = \underline{\underline{4\%}} \textcircled{5}$

3. (15 Points) Only one percent of servings will contain more than 8.34 grams of fat.



area between -2.30 and 2.30: 97.86%

area between -2.35 and 2.35: 98.12% \leftarrow closest to 98% $\textcircled{5}$

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

in original units:

$$\# = 2.35 \cdot 0.4 + 7.4 = 0.94 + 7.4 = \underline{\underline{8.34}} \textcircled{5}$$

(or $\# = 2.30 \cdot 0.4 + 7.4 = 0.92 + 7.4 = 8.32$)

Show your work!

2.) From: Freedman, Pisani, Purves, page 155, Exercise 8

Question 5: Regression (35 Points)

For women age 25 and over in the U.S. in 1993, the relationship between age and educational level (years of schooling completed) can be summarized as follows:

- \bar{x} average age ≈ 48.7 years, $SD_x \approx 16.8$ years
 \bar{y} average ed. level ≈ 12.5 years, $SD_y \approx 3.1$ years,
 $r \approx -0.28$

1. (20 Points) The predicted educational level (number of years of schooling) for a woman who is 45 years old is 12.7 years. Show your work!

$$s.u._x = \frac{x - \text{avg}_x}{SD_x} = \frac{45 - 48.7}{16.8} = -0.22 \quad (6) \quad -2 \text{ for each calculation error}$$

$$s.u._y = r \cdot s.u._x = -0.28 \cdot (-0.22) = 0.0616 \quad (6)$$

$$\hat{y} = s.u._y \cdot SD_y + \text{avg}_y = 0.0616 \cdot 3.1 + 12.5 = 12.69 \approx \underline{12.7} \quad (8)$$

2. (15 Points) True or false, and explain: as you get older, you become less educated. If this statement is false, what accounts for the negative correlation?

(10) False - these are cross-sectional data and not longitudinal data. (5) explanation
Younger people were born more recently and education levels have been going up over time.
[If you have 12 years of schooling for example, you will always have 12 years (or more) throughout your entire life - but never less than 12 years.]

Question 6: Valentine's Day Question (5 Points)

How many **free points** do you want to get for taking the Midterm Test just before the Valentine's Day (circle one)?

- 0 points
- 1 point
- 3 points

0/0
• 5 points