

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

Friday, October 7, 2005

Your Name: _____

Question 1: Controlled Experiment/Observational Study (25 Points)

The New Zealand Herald reported on September 26, 2005:

"Left-handed women are more than twice as likely as right-handers to suffer from breast cancer before reaching menopause, Dutch scientists said on Monday. [...] Cuno Uiterwaal, an assistant professor [...] and his colleagues studied 12,000 healthy, middle-aged women born between 1932-1941 who were part of a breast screening programme. The scientists determined their hand preference and followed up their medical history to see which women developed breast cancer. [...] Left-handed women are more than twice as likely to develop pre-menopausal breast cancer as non-left handed women. [...] Other risk factors such as family history of breast cancer, number of pregnancies, smoking habits, and social and economic status were considered. [...]"

1. (5 Points) Is this an example of a **controlled experiment** or an **observational study**? Circle your answer and explain.

② No intervention took place - the women were just asked about their hand preference and medical history. ③

2. (10 Points) Why does the article emphasize "Other risk factors such as family history of breast cancer, number of pregnancies, smoking habits, and social and economic status were considered"? Explain clearly.

These may be confounding factors that could have an effect on obtaining breast cancer. ⑩

- 3 if correct description, but keywords missing

3. (10 Points) Does this result necessarily imply that if all left-handed women immediately switch over to right-handedness, they will reduce, on average, their risk to obtain breast cancer to the same rate as it is reported for right-handed women? Explain clearly.

⑤ Clearly no. By switching from left hand to right hand, a woman will not reduce her risk of obtaining breast cancer. Association (between left-handedness and higher risk of breast cancer) is not causation. ⑤

- 3 if correct description, but keywords missing

From: Stat 1040, Fall 2002, Midterm 1, October 4, 2002, Question 2

Question 2: Normal Distribution (45 Points)

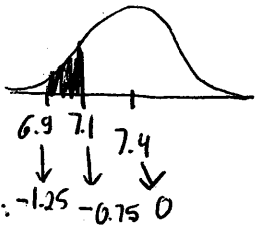
According to the U.S. Department of Agriculture, one 3-ounce serving of trimmed sirloin beef contains, on average, 7.4 grams of fat. Assume that the amount of fat for such servings closely follows the normal curve, with a standard deviation of 0.4 gram.

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%.

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

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



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)

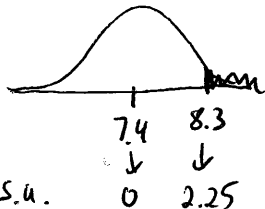
area between -1.25 and 0: $78.87\% / 2 = 39.435\%$ (3)

area between -1.25 and -0.75: $39.435\% - 27.335\% = 12.1\%$ (3)

3 for graph only (in each part)

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

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

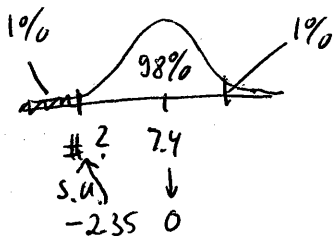


from Table: area between -2.25 and 2.25: 97.56% (5)

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 # = -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, Spring 2003, Midterm 1, February 14, 2003, Question 5

Question 3: Change of Scale (40 Points)

From the subjects in a health survey, the following data were collected:

average height = 68 inches SD = 2.5 inches
average blood pressure = 120 mm SD = 15 mm
correlation $r = -0.2$

You want to provide a summary of these results to a friend in Europe and report heights in centimeters (instead of inches) to make it easier for your friend to interpret the results. Recall that 1 inch = 2.54 centimeters.

1. (10 Points) The average height (in centimeters) is: $2.54 \cdot 68 = 172.72$ centimeters
2. (10 Points) The SD [for the height] (in centimeters) is: $2.54 \cdot 2.5 = 6.35$ centimeters
3. (10 Points) The correlation r is: -0.2 (this does not change when we change units!)

4. (10 Points) Circle the correct answer: r is measured in

- centimeters
- inches
- mm
- inches · mm
- centimeters · mm
- inches · centimeters
- none of these – r is a unitless number

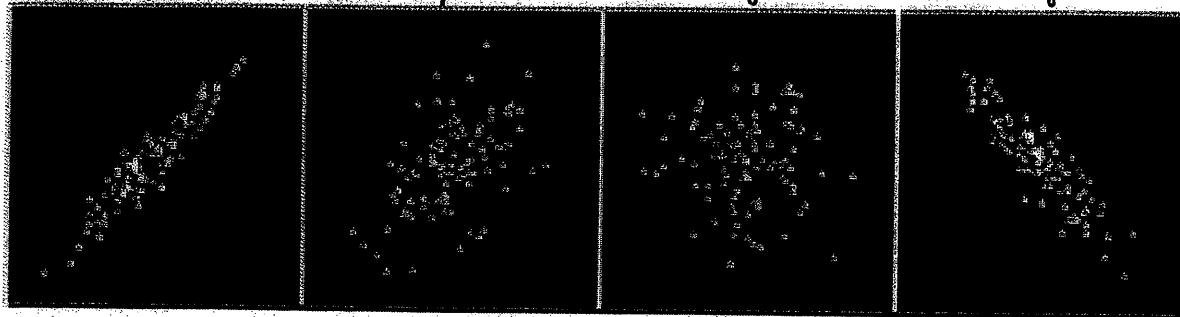
From: Stat 1040, Spring 2002, Final Test, April 30, 2002, Question 6 (different plots)
 & Stat 1040, Fall 2002, Midterm 1, October 4, 2002, Question 4 (same plots)

Question 4: Guessing the Correlation Coefficient (30 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



⑦ Correlation for Plot A: $r = \underline{0.93}$

⑧ Correlation for Plot B: $r = \underline{0.50}$

⑧ Correlation for Plot C: $r = \underline{-0.05}$

⑦ Correlation for Plot D: $r = \underline{-0.89}$

for each plot:

4 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: \rightsquigarrow 0.93

Plot D shows an association that is slightly weaker than in A: \rightsquigarrow -0.89

Plot B shows a weak positive association: \rightsquigarrow 0.50

Plot C shows almost no association: \rightsquigarrow -0.05

