

# Statistics 1040, Section 003, Midterm 1 (200 Points)

Friday, October 3, 2008

Your Name: \_\_\_\_\_

from: Stat 1040, Spring 2008, Final Test, May 1, 2008, Question 1 (Solutions: → Course Web Page)

## Question 1: Controlled Experiments / Observational Studies (35 Points)

**Does Aspartame Cause Cancer?** Aspartame is an artificial sweetener found in thousands of products — sodas, chewing gum, dairy products and even many medicines. Some research has suggested that aspartame can cause lymphoma or leukemia in rats. A recent study by the National Cancer Institute involved 340,045 men and 226,945 women, ages 50 to 69. From surveys they filled out in 1995 and 1996 detailing food and beverage consumption, researchers calculated how much aspartame they consumed. Over the next five years, 2,106 developed cancers such as lymphoma or leukemia. No association was found between aspartame consumption and occurrence of these cancers.

1. (10 Points) Was the study a controlled experiment or an observational study? (6)  
Why? Circle your answer and explain.

There was no intervention. The subjects decided themselves what (4)  
to eat and to drink.

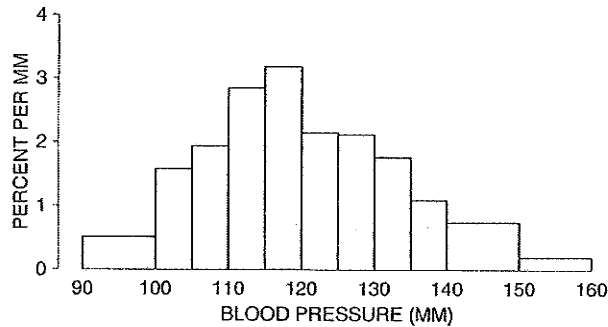
2. (15 Points) Suggest a possible confounding factor for this study and explain why your confounding factor might make you doubt their results. (10) for 1 valid factor  
(5) for explanation  
There might be various confounding factors, e.g.  
- race: some types of cancer are more frequent in one race than in other races  
- environment & work place: subjects may have been exposed to other chemicals or pollutants at home or at their work place and these chemicals or pollutants may have contributed to their cancer  
- genetics: certain genetic diseases may be associated with the type of cancer
3. (10 Points) "It's very reassuring ... It's a large study with a lot of power," said Richard Adamson, a senior science consultant to the American Beverage Association, the leading industry group. Does the large sample size prove that aspartame does not cause cancers such as lymphoma or leukemia? YES or NO (6) — Circle your answer and explain.

An invalid study remains invalid no matter what the sample size is. (4)  
Here, they have ignored several <sup>1</sup> confounding factors at a large scale.

from: FPP, Chapter 3, Review Exercise 4, pp. 51-52  
 Question 2: Histograms (35 Points)

(Solutions: → Workbook)

The figure below is a histogram showing the distribution of blood pressure for all 14,148 women in the Drug Study from Chapter 3, Section 5. Use the histogram to answer the following questions. Circle your answer.



1. (5 Points) Is the percentage of women with blood pressures above 130 mm around 25%, 50%, or 75%?

5

2. (5 Points) Is the percentage of women with blood pressures between 90 mm and 160 mm around 1%, 50%, or 99%?

5

3. (5 Points) In which interval are there more women: 135-140 mm or 140-150 mm?

5

4. (5 Points) Which interval is more crowded: 135-140 mm or 140-150 mm?

5

5. (5 Points) On the interval 125-130 mm, the height of the histogram is about 2.1% per mm. What percentage of the women had blood pressures in this class interval? About 10.5%, 12.6%, or 21%?

5  
 $[ 5 \times 2.1\% = 10.5\% ]$

6. (5 Points) Which interval has more women: 97-98 mm or 102-103 mm?

5

7. (5 Points) Which is the most crowded millimeter of all? Somewhere between 90 and 100 mm, somewhere between 115 and 120 mm, or somewhere between 140 and 150 mm?

5

New Question

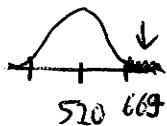
(No solutions available)

Question 3: Normal Curve (48 Points)

The Graduate Record Examination (GRE) is a test taken by college students who intend to pursue a graduate degree in the United States. For around 210,000 female US citizens who took the General GRE Test in 2005-06, the mean for the quantitative ability portion of the exam was about 520 and the standard deviation was about 135 ([http://www.ets.org/Media/Tests/GRE/pdf/05-06\\_factors\\_final.%20pdf.pdf](http://www.ets.org/Media/Tests/GRE/pdf/05-06_factors_final.%20pdf.pdf)). We can assume that the histogram follows a normal curve. Show your work!

- (15 Points) The percentage of female US citizens who scored more than 669 on the GRE test is roughly 13.57 %.

-2 for each calculation error  
+3 for correct graph (but nothing else)



① s.u.:  $\frac{669-520}{135} = 1.10$  s.u. (5)

-1.10 0 1.10 s.u. ② area from -1.10 to 1.10: 72.87% (5)

③ area above 1.10:  $\frac{100\% - 72.87\%}{2} = 13.57\%$  (5)

- (18 Points) The percentage of female US citizens who scored between 351 and 574 is about 54.98 %.



① s.u.:  $\frac{351-520}{135} = -1.25$  s.u. (3)

s.u.:  $\frac{574-520}{135} = 0.40$  s.u. (3)

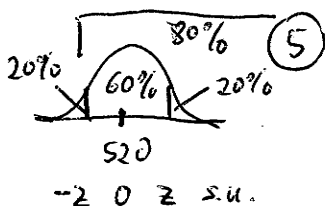
-1.25 0 0.40 s.u.

② area from -1.25 to 1.25: 78.87% (3)

area from -0.40 to 0.40: 31.08% (3)

③ area from -1.25 to 0.40:  $\frac{78.87\%}{2} + \frac{31.08\%}{2} = 54.98\%$  (6)

- (15 Points) In order to be among the top 80% of all female US citizens, a student must have obtained a minimum GRE score of about 405.



① area from  $-0.85$  to  $0.85$ : 60.47% (Closest to 60%) (5)

② original units:  $-0.85 \cdot 135 + 520 = 405.25 \approx 405$  (5)

from: FPP, Chapter 10, Review Exercise 4, pp. 176-177  
 & Stat 1040, Spring 2006, Midterm 1, Question 5  
 Question 4: Regression (50 Points) [Part 2 is new!]

(Solutions → Workbook)  
 ( → Course Web Page)

In one study, the correlation between the educational level of husbands and wives in a certain town was about 0.50; both averaged 12 years of schooling completed, with an SD of 3 years.

Show your work.

1. (15 Points) Predict the educational level of a woman whose husband has completed 18 years of schooling.

The answer is: 15 years

$$s_{y|x} = \frac{18-12}{3} = 2 \quad (5)$$

$$s_{ay} = 2 \cdot 0.5 = 1 \quad (5)$$

$$y = 1 \cdot 3 + 12 = \underline{15} \quad (5)$$

-2 for each calculation error

(in 1., 2., & 3.)

2. (10 Points) Calculate the corresponding r.m.s. error.

The answer is: 2.6 years

$$r.m.s. error = \sqrt{1 - 0.5^2} \cdot 3$$

$$= \sqrt{0.75} \cdot 3 = 0.866 \cdot 3 = \underline{2.6}$$

-4 for each major mistake,

e.g.  $\sqrt{\quad}$  of everything,  
 $\tau$  instead of  $\tau^2$ , etc.

3. (15 Points) Predict the educational level of a man whose wife has completed 15 years of schooling.

The answer is: 13.5 years

$$s_{u|x} = \frac{15-12}{3} = 1 \quad (5)$$

$$s_{ay} = 1 \cdot 0.5 = 0.5 \quad (5)$$

$$y = 0.5 \cdot 3 + 12 = \underline{13.5} \quad (5)$$

4. (10 Points) Apparently, well-educated men marry women who are less well-educated than themselves. But the women marry men with even less education. How is this possible?

Workbook: "Appearances are deceiving; all that is going on here is the regression effect."

(10) for correct keyword

(5) for reasonable explanation

(without keyword)

Web Page: "Nothing unexpected - this is just the regression effect! The explanation given is an example of the regression fallacy."

(1) for some explanation

New Question

(No solutions available)

Question 5: Guessing the Correlation Coefficient (32 Points)

Match the four scatterplots with their correlations from the list:

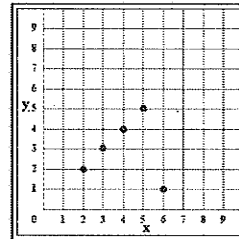
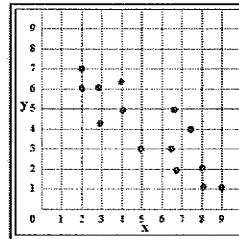
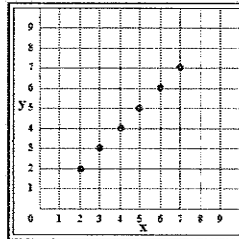
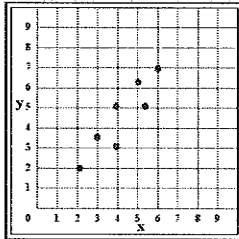
-1.03, -1.00, -0.90, -0.85, -0.30, 0.00, 0.30, 0.85, 0.90, 1.00, 1.03

A

B

C

D



Correlation for Plot A:  $r = 0.90$  (8) [narrow football; increasing]

Correlation for Plot B:  $r = 1.00$  (8) [all points on increasing line]

Correlation for Plot C:  $r = -0.85$  (8) [somewhat wider than A; decreasing]

Correlation for Plot D:  $r = 0.00$  (8) [almost perfect, but outlier; see p. 76 in wed 9/24/08 lecture & recall applet from web page]

Formulas:

$$\text{avg} = \frac{\text{sum of all numbers}}{\text{how many numbers}}$$

$$\text{SD} = \sqrt{\text{average of } [( \text{deviations from avg} )^2]}$$

$$\text{r.m.s. error} = \sqrt{1 - r^2} \times \text{SD}_y$$