

Biost 517

Applied Biostatistics I

**Final Examination
Tuesday, December 18, 2001**

Instructions: Please provide concise answers to all questions. Rambling answers touching on topics not directly relevant to the question will tend to count against you. Nearly telegraphic writing style is permissible.

The examination is closed book and closed notes. If you come to a problem that you believe cannot be answered without making additional assumptions, clearly state the reasonable assumptions that you make, and proceed.

Problems 1 - 5 refer to a study of the association between findings on magnetic resonance imaging (MRI) of the brain and cognitive function and survival in the elderly. The variables available in this data set include

- **Age** = the patient's age in years
- **Male** = an indicator of the patient's sex (0 = female, 1= male)
- **Marital** = a code indicating the patient's marital status (1= married, 2= widowed, 3= divorced, 4= never married, 5= other)
- **Health** = a code indicating the patient's self-reported health status (1= excellent, 2= very good, 3= good, 4= average, 5= poor)
- **DSST** = score on a test of mental function (possible scores 0 - 150)
- **Infarct** = an indicator of MRI evidence of a brain lesion due to lack of oxygen (0= none, 1= infarct present)
- **Obstime** = time of follow-up in days from start of study until death or the time of data analysis, whichever comes first
- **Death** = indicator that a death was observed

	n	mean	std dev	min	25%-ile	median	75%-ile	maximum
age	58.00	75.34	5.22	67.00	71.25	75.00	78.00	90.00
male	58.00	0.60	0.49	0.00	0.00	1.00	1.00	1.00
marital	58.00	1.38	0.95	1.00	1.00	1.00	1.00	5.00
health	58.00	2.50	1.05	1.00	2.00	2.00	3.00	5.00
DSST	58.00	63.53	29.85	15.00	40.00	57.00	86.50	130.00
infarct	58.00	0.43	0.50	0.00	0.00	0.00	1.00	1.00
obstime	58.00	1539.22	378.29	231.00	1473.25	1610.50	1773.25	2103.00
death	58.00	0.19	0.40	0.00	0.00	0.00	0.00	1.00

1. The above table provides the sample size, sample mean, median, standard deviation, minimum, maximum, 25th and 75th percentiles for the above data. For each of the above variables briefly indicate which of the above descriptive statistics are not of particular scientific interest for this study. Also indicate any other descriptive statistics you might wish to look at.

a. Age

b. Male

c. Marital

d. Health

e. DSST

f. Infarct

g. Obstime

h. Death

2. For each of the following scientific questions, indicate two possible statistical approaches that you might use for statistical inference. Be sure to identify what aspect of the distribution (e.g., a population parameter) that you are comparing, and indicate whether your method of inference can include a confidence interval as well as a test of a hypothesis.

- a. Does the distribution of age differ by sex?

- b. Does the distribution of infarcts on MRI exam differ by sex?

- c. Does the distribution of marital status differ by sex?

- d. Does the distribution of self-reported health status differ by sex?

- e. Does the distribution of DSST scores differ by sex?

- f. Does the survival time differ by sex?

3. Suppose the average DSST is computed separately for males and females in the combined sample and also in strata according to the presence of infarcts. The following table contains the sample mean, standard error, and 95% confidence intervals for those means.

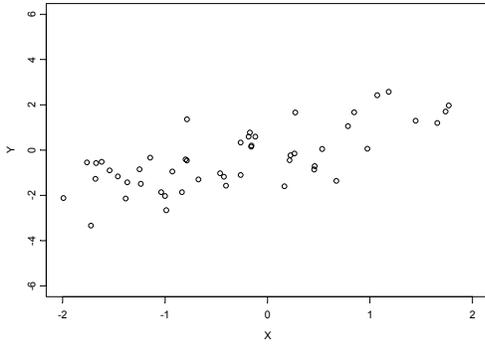
	Mean	Std Err	95% CI Low	95% CI Hi
Females	58.30	6.67	42.26	74.34
Males	66.97	4.79	55.74	78.20
No Infarcts				
Females	85.40	9.08	61.01	109.79
Males	83.48	3.86	74.19	92.77
Infarcts				
Females	37.46	3.65	28.12	46.80
Males	35.33	3.30	26.77	43.89

- a) Based on the results in the above table, how would you characterize the evidence of an association between DSST scores and sex in the combined sample?

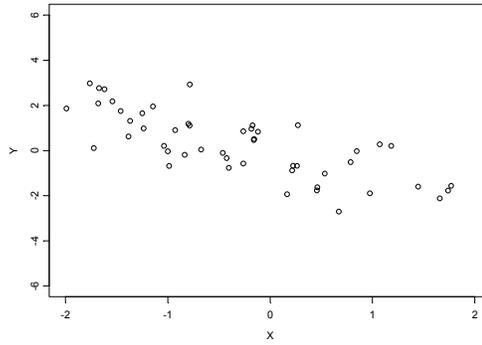
- b) An analysis was performed to compare the prevalence of infarcts between males and females. The difference in the prevalence of infarcts in females minus the prevalence in males was 0.22 with a 95% confidence interval of -0.036 to 0.481. Based on this information, would you conclude that there was sufficient evidence to claim an association between sex and prevalence of infarcts in the population?
- c) Based on the results in the above table, would you regard that there was a statistically significant association between presence of infarcts and average DSST scores in the population? Explain your reasoning.
- d) Based on the results in the above table, is there effect modification by infarct status on any association between DSST scores and sex? Explain your reasoning.
- e) Based on the results in the above table, is there evidence of confounding by infarct status on the association between DSST scores and sex? Explain your reasoning. How does your answer relate to the findings of part b and c above?
- f) What analysis would you regard as the most appropriate way to assess an association between sex and DSST scores in this sample?
4. Suppose that the distribution of age was compared between the sexes using a t test for equal variances (yielding a P value of 0.04) as well as a t test allowing unequal variances (yielding a P value of 0.09). Give two possible interpretations for the seeming discrepancy between these results.

5. Below are 7 scatterplots. List the plots in order according to lowest (most negative) to highest (most positive) correlation. (In all cases, the scale for the x and y axes are the same.)

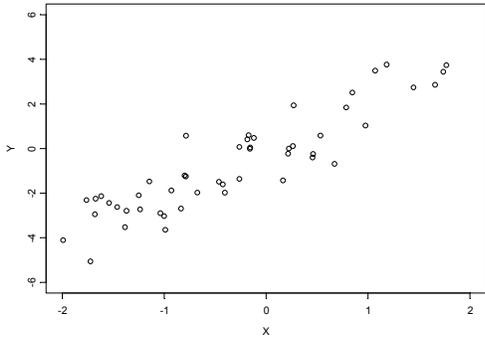
Plot A



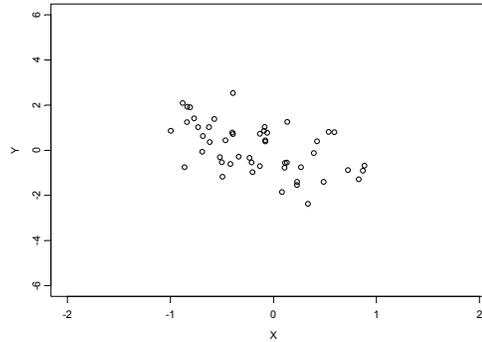
Plot B



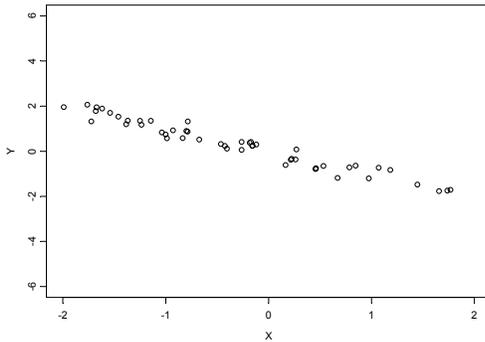
Plot C



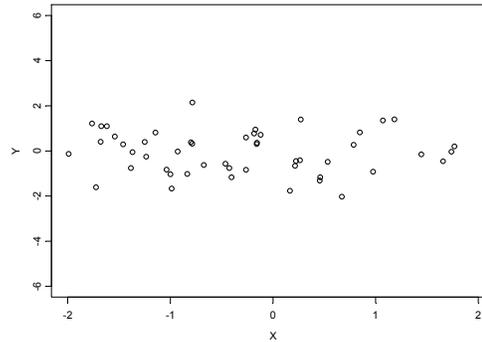
Plot D



Plot E



Plot F



Plot G

