

Biost 517
Applied Biostatistics I

Midterm Examination
November 4, 2011

Name: _____ Disc Sect: M Tu W F

Instructions: Please provide concise answers to all questions. The exam is worth a total of 147 points.

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. You may use calculators, but you may not use any special programs written for programmable calculators.

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.

Please adhere to and sign the following pledge. Should you be unable to truthfully sign the pledge for any reason, turn in your paper unsigned and discuss the circumstances with the instructor on Monday.

PLEDGE:

On my honor, I have neither given nor received unauthorized aid on this examination:

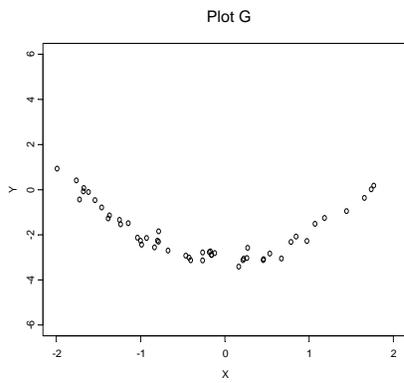
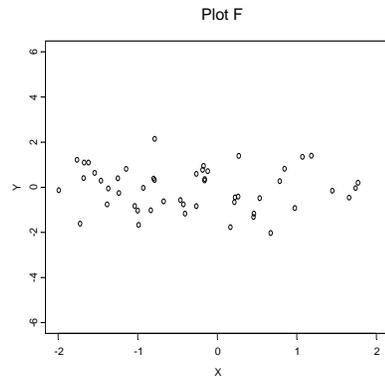
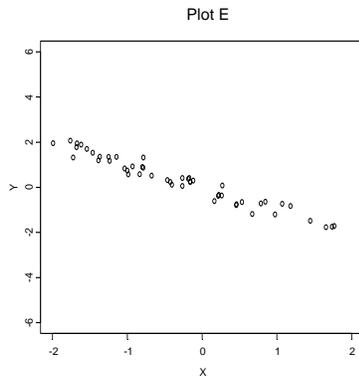
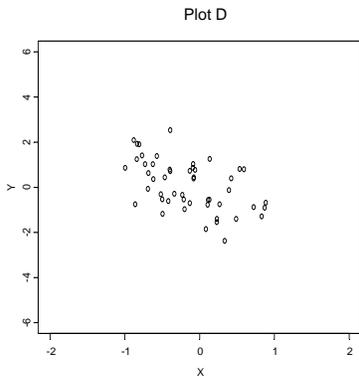
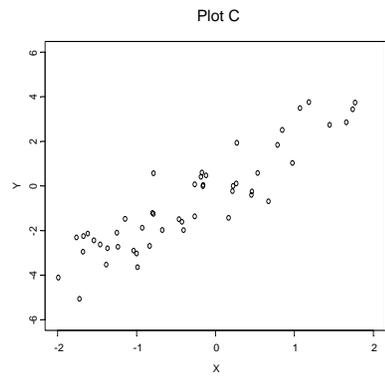
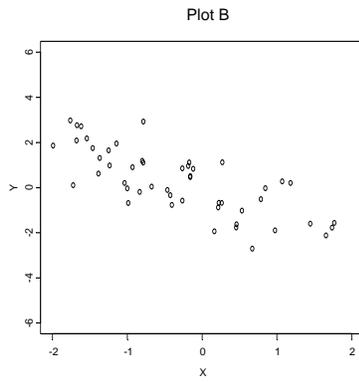
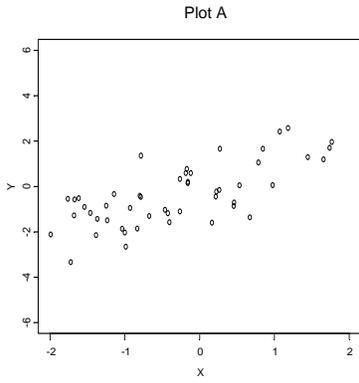
Signed: _____

1. (30 points total) Suppose we are interested in studying whether levels of the chemical lactate in the cerebrospinal fluid (CSF - the fluid surrounding the brain and spinal cord) can be used to diagnose bacterial meningitis (an infection of the membranes covering the brain and spinal cord).
 - Suppose we sample 120 subjects with culture proven bacterial meningitis and we measure the CSF lactate. Among these subjects we find that 96 subjects have a CSF lactate greater than 3 mmol/L.
 - Suppose we also sample 250 otherwise healthy subjects who were undergoing radiologic examinations due to low back pain. CSF samples from these subjects were measured for lactate levels, and 10 were found to have CSF lactate levels greater than 3 mmol/L.
- a. (5 points) Can the above data be used to estimate the probability of bacterial meningitis at the hospital where the study was performed? If so, provide the estimate. If not, briefly explain why not.

- b. (5 points) Can the above data be used to estimate the sensitivity of a high CSF lactate (greater than 3 mmol/L) in diagnosing bacterial meningitis? If so, provide the estimate. If not, briefly explain why not.
- c. (5 points) Can the above data be used to estimate the specificity of a high CSF lactate (greater than 3 mmol/L) in diagnosing bacterial meningitis? If so, provide the estimate. If not, briefly explain why not.
- d. (5 points) Can the above data be used to estimate the predictive value of the positive of a high CSF lactate (greater than 3 mmol/L) in diagnosing bacterial meningitis? If so, provide the estimate. If not, briefly explain why not.
- e. (5 points) Can the above data be used to estimate the predictive value of the negative of a high CSF lactate (greater than 3 mmol/L) in diagnosing bacterial meningitis? If so, provide the estimate. If not, briefly explain why not.
- f. (5 points) Suppose at another hospital the probability of bacterial meningitis is 10% among all patients for whom a CSF sample is obtained. Can the above data be used to estimate the positive predictive value of high CSF lactate (greater than 3 mmol/L) at that hospital? If so, provide the estimate and explain how you derived your answer.

2. (15 points) Below are 7 scatterplots labeled A - G. 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.)

Most.....Most
Neg Pos



Problems 3 - 4 relate to a longitudinal observational study examining relationships between selected biomarkers of inflammation and cardiovascular disease in 5,000 elderly, generally healthy subjects. The following variables are available:

Name	Description
Site	Indication of geographic site (1, 2, 3, or 4)
Age	Subject age at study accrual (years)
Smoker	Indicator of subject's current smoking status at study accrual (0= no, 1= yes)
BMI	Subject's body mass index (weight / height ²) at time of study accrual (kg / m ²)
SBP	Subject's systolic blood pressure at time of study accrual (mm Hg)
Cholesterol	Subject's serum cholesterol at time of study accrual (mg/dl).
CRP	Subject's blood level of C reactive protein—a biomarker of inflammation (mg/l)
Ttodth	Observation time from randomization to death or data analysis, whichever came first (months)
Death	Indicator that subject was observed to die while on study

The following table contains descriptive statistics on the sample.

	N	Mean	SD	Min	q25	Median	q75	Max
Females								
Site	2904	2.45	1.13	1	1	2	3	4
Age	2904	72.6	5.52	65	68	71	76	100
Smoker	2901	0.126	0.332	0	0	0	0	1
BMI	2895	26.9	5.31	14.7	23.2	26.1	29.6	58.8
SBP	2897	137	22.3	77	122	135	151	235
Cholesterol	2870	221	38.9	88	195	219	245	430
CRP	2861	3.63	5.49	0	1	2	4	86
Ttodth	2904	79.9	20.2	0.8	80.1	88.3	92.3	96.8
Death	2904	0.170	0.376	0	0	0	0	1
Males								
Site	2096	2.50	1.13	1	1	2	4	4
Age	2096	73.2	5.69	65	69	72	77	95
Smoker	2093	0.114	0.318	0	0	0	0	1
BMI	2092	26.4	3.78	15.6	23.9	26.1	28.5	46.2
SBP	2093	136	21.3	79	121	133	148	219
Cholesterol	2083	198	35.7	73	174	197	221	407
CRP	2072	3.59	6.97	0	1	2	3	108
Ttodth	2096	75.2	24.6	0.2	54.7	87.6	92.3	96.7
Death	2096	0.299	0.458	0	0	0	1	1

3. (3 points each part) For each of the following variables, circle the descriptive statistics that you **WOULD NOT CONSIDER USING** to provide a scientifically meaningful description of a possible association between sex and the relevant scientific variable. Very briefly explain your reasons (just a few words should suffice to justify your entire answer).
- Consider **site**. Circle the descriptive statistics that WOULD NOT be useful to describe associations between that variable and sex. Briefly explain why.

Mean Std Dev Minimum 25th Pctile Median 75th Pctile Maximum

- b. Consider **age**. Circle the descriptive statistics that WOULD NOT be useful to describe associations between that variable and sex. Briefly explain why.

Mean Std Dev Minimum 25th Pctile Median 75th Pctile Maximum

- c. Consider **smoker**. Circle the descriptive statistics that WOULD NOT be useful to describe associations between that variable and sex. Briefly explain why.

Mean Std Dev Minimum 25th Pctile Median 75th Pctile Maximum

- d. Consider **BMI**. Circle the descriptive statistics that WOULD NOT be useful to describe associations between that variable and sex. Briefly explain why.

Mean Std Dev Minimum 25th Pctile Median 75th Pctile Maximum

- e. Consider **SBP**. Circle the descriptive statistics that WOULD NOT be useful to describe associations between that variable and sex. Briefly explain why.

Mean Std Dev Minimum 25th Pctile Median 75th Pctile Maximum

- f. Consider **cholesterol**. Circle the descriptive statistics that WOULD NOT be useful to describe associations between that variable and sex. Briefly explain why.

Mean Std Dev Minimum 25th Pctile Median 75th Pctile Maximum

- g. Consider **CRP**. Circle the descriptive statistics that WOULD NOT be useful to describe associations between that variable and sex. Briefly explain why.

Mean Std Dev Minimum 25th Pctile Median 75th Pctile Maximum

- h. Consider **ttodth**. Circle Circle the descriptive statistics that WOULD NOT be useful to describe associations between that variable and sex. Briefly explain why.

Mean Std Dev Minimum 25th Pctile Median 75th Pctile Maximum

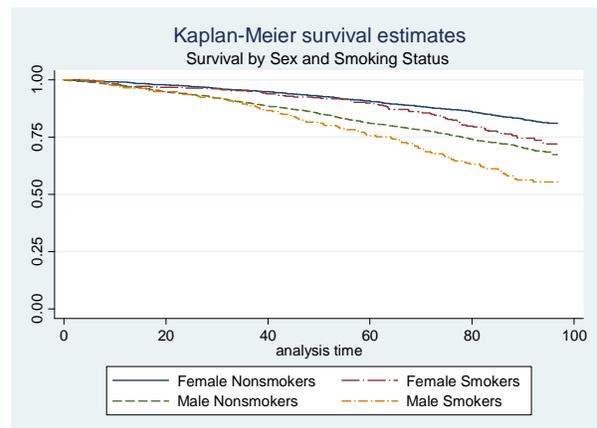
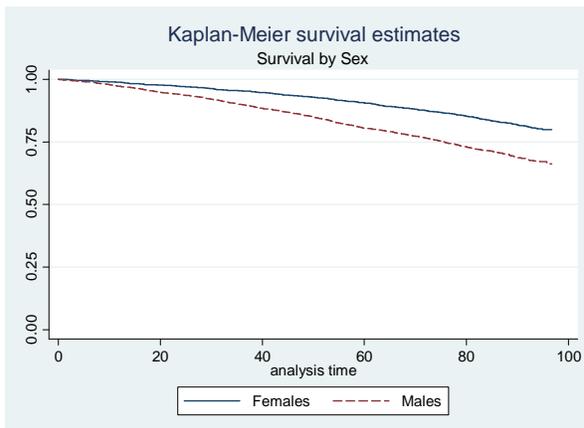
- i. Consider **death**. Circle the descriptive statistics that WOULD NOT be useful to describe associations between that variable and sex. Briefly explain why.

Mean Std Dev Minimum 25th Pctile Median 75th Pctile Maximum

4. (10 points) Do the above descriptive statistics provide evidence for associations with sex for any of the variables? If so, list the variable and provide the descriptive estimate that you used to make your decision. (I want to know what you looked at.)

5. (10 points) For which of the variables is it of scientific interest to assess skewness caused by outlying values? Briefly list all the possible evidence for any such variable that you believe is affected by outliers.

6. (35 points) The following are results from Kaplan-Meier analyses of the time to death within strata defined by sex, smoking status, and all combinations of those two variables.



	Females			Males			Both Sexes		
	n	Survival Probability		n	Survival Probability		n	Survival Probability	
		60 mos	84 mos		60 mos	84 mos		60 mos	84 mos
Nonsmokers	2536	0.907	0.847	1854	0.811	0.728	4390	0.866	0.796
Smokers	365	0.899	0.778	239	0.757	0.612	604	0.843	0.713
All Patients	2901	0.906	0.838	2093	0.805	0.716	4994	0.864	0.787

- a. (5 points) Under what conditions will the Kaplan-Meier estimates provide unbiased estimation of the distribution of times to death?
- b. (10 points) Based on the above statistics, would you conclude that there is overall an association between sex and the probability of survival? Provide statistics to quantify your answer.
- c. (10 points) Based on the above statistics, would you conclude that smoking behavior confounds any association between sex and survival? Provide statistics to quantify your answer.

- d. (10 points) Based on the above statistics, would you conclude smoking behavior modifies any association between sex and survival? Provide statistics to quantify your answer.

Problem 7 pertains to a study of dementia in 735 elderly adults. Available data include

- **male** = Patient sex (0= female, 1= male)
- **fev** = Forced expiratory volume (l/sec).
- **dsst** = Cognitive function as measured by the Digit Symbol Substitution Test (DSST) on a scale from 0 (worst) to 100 (best)

