STA 110B Spring 1998

Midterm Exam 2

April 2, 1998

Name:	**Section:**

I understand and agree to abide by the Duke honor code,

Signed:

Instructions

This is a closed-book exam, however, one 8.5 by 11 inch "crib sheet" is permitted. You may use a calculator if you find it useful. Show your work in the space provided, but be concise. Correct but unsubstantiated answers will receive no credit.

Point assignments for each of the 3 problems are given in parentheses in the table below. You have 1 hour and 15 minutes total; plan accordingly. You must hand the exam in at 12:10pm, no extra time will be given. Good luck!

		Page 1	**Page 2**	**Page 3**
1.	(40)			
2.	(30)			
3.	(30)			
Total	(100)			

1) The following is the joint distribution of high school grade point average (center of grade category) and year of matriculation for college freshmen in the U.S. (reconstructed from Statistical Abstract of the United States, 1996, tables 296 and 284).

	Year of Matriculation (**Y**)		
HS GPA (**G**)	**Y**=1970	**Y**=1980	**Y**=1990
G=1.5	0.08	0.07	0.06
G=2.5	0.17	0.22	0.19
G=3.5	0.05	0.08	0.08

a) (**4 points**) Write the conditional distribution of **G** given **Y**=1970.

b) (**5 points**) Calculate $E(\mathbf{G}|\mathbf{Y})$ for each value of **Y**.

c) (**4 points**) Plot (scatter plot) the conditional means you calculated in part **b)** (vertical axis) against **Y** (horizontal axis).

1) Continued. The joint distribution of high school grade point average and year of matriculation for college freshmen in the U.S. is:

	Year of Matriculation (**Y**)		
HS GPA (**G**)	**Y**=1970	**Y**=1980	**Y**=1990
G=1.5	0.08	0.07	0.06
G=2.5	0.17	0.22	0.19
G=3.5	0.05	0.08	0.08

d) (**5 points**) are **G** and **Y** independent random variables? Why/why not?

e) (**4 points**) Calculate the mean and variance of **G**.

f) (**4 points**) Calculate the mean and variance of **Y**.

1) Continued. The joint distribution of high school grade point average and year of matriculation for college freshmen in the U.S. is:

	Year of Matriculation (**Y**)		
HS GPA (**G**)	**Y**=1970	**Y**=1980	**Y**=1990
G=1.5	0.08	0.07	0.06
G=2.5	0.17	0.22	0.19
G=3.5	0.05	0.08	0.08

g) (**5 points**) The expected value of **G** times **Y**, $E(\mathbf{GY})$, is 4951.25. Calculate the covariance of **G** and **Y**.

h) (**4 points**) What is the correlation of **G** and **Y**?

i) (**5 points**) Briefly, what can you conclude about the relationship between **G** and **Y**?

2) In 1990, the average time workers in the United States (those who did not work at home) spent traveling to work was 21 minutes, the variance of travel time was 166 (derived from a 1990 Census Bureau table).

a) (**5 points**) What is the mean, variance and approximate distribution of the sample mean of a sample of size $n = 4$ drawn from this population?

b) (**5 points**) What is the mean, variance and approximate distribution of the sample mean of a sample of size $n = 400$ drawn from this population?

c) (**10 points**) What is the probability that the sample mean of a sample of size $n = 400$ drawn from this population would be between 20 and 22 minutes?

2) Travel times, continued.

d) (10 points) The proportion of workers who worked at home in 1990 was 3% (same source as before). What is the probability that the number of workers in a sample of size 400 that worked at home would be greater than or equal to 8 **and** less than or equal to 16? Justify any approximation you might use.

3) ACORN (Association of Community Organizations for Reform Now) presented the following data to a Joint Congressional Hearing on discrimination in lending in October of 1991. The data are loan refusal rates at 15 major banks in the U.S. for high income white and high income minority applicants. The difference between the two rates (minority minus white) is also included.

Loan Refusal Rates

Bank	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Minority	41.1	41.3	21.4	24.2	5.8	36.6	38.3	39.1	29.5	33.3	21.7	28.6	17.3	38.0	32.9
White	26.8	25.1	2.2	14.1	4.2	15.3	15.0	15.8	7.3	10.3	7.4	10.1	5.5	7.6	9.2
Difference	14.3	16.2	19.2	10.1	1.6	21.3	23.3	23.3	22.2	23.0	14.3	18.5	11.8	30.4	23.7

A table of sample statistics for the data follows:

	n	**sample mean**	**sample variance**
Minority	15	29.94	103.93
White	15	11.73	50.31
Difference	15	18.21	49.81

a) (**10 points**) Give a 95% confidence interval for the mean difference in refusal rates between low income minority and low income white applicants. Show your work.

3) Continued. These questions continue our analysis of the ACORN data, whose summary data is repeated here.

Loan Refusal Rates

	n	**sample mean**	**sample variance**
Minority	15	29.94	103.93
White	15	11.73	50.31
Difference	15	18.21	49.81

b) (**15 points**) It is of interest to test the (null) hypothesis that refusal rates are equal for high income minority and high income white applicants (against the alternative that they are different). Calculate the p-value associated with this hypothesis.

c) (**5 points**) Can the null hypothesis be rejected at the 5% level? At the 1% level? Why/why not?