11.220: Quantitative Reasoning and Statistical Methods for Planning I

Test-out exam: January 2005

Name: _____ E-mail address: _____

Academic advisor's name: _____

Did you attend the brush-up? _____ All _____ Some _____ None

Instructions:

1. Relax. Breathe. We recommend that you flip through the exam and start with those problems you feel most comfortable answering.
2. You are allowed to use up to two textbooks of your choice and your notes during the exam. Please note that you may not share these materials with other students.
3. You may use a calculator during the exam; however, you may not use any statistical functions on the calculator. Please read and sign the statement at the bottom of this page regarding these restrictions.
4. For each problem, show your work in the space provided. We encourage you to draw pictures where appropriate. Partial credit will be given if we have evidence that you framed the problem correctly and/or were headed in the right direction to obtain the correct answer. Points will be deducted if you are asked to show your work for a problem and fail to do so. You may use the back of any page to complete your work. In all cases, indicate clearly what your (single) final answer is for each question.
5. Remember to keep your writing clear, tight, and to the point. Think before you write. You do not need more than a few well-written sentences to answer any of the questions on the exam. (This will save you time!)
6. Point values for each question are included in [brackets] next to each question on the exam; there are 100 total possible points.
7. During the exam, please let us know if you find unclear printing, typing, or errors on the exam. We will not give any hints about how to answer a question. If you think a question is unclear, clearly state your assumptions on the exam and complete the question to the best of your ability.
8. The exams will be scored and returned to your mailboxes in the student common room. You will be notified *via* email whether you have placed out of 11.220.

Please read this statement and sign your name below.

I certify that I have neither given to nor received assistance from another person on this exam, and that I have used only simple arithmetic functions on my calculator.

Signature: _____

Univariate data analysis & Probability ☐

Consider the following set of ratings for college scholarship applicants produced by a selection committee of two persons. Each evaluator was asked to rank each application on a scale of 1 (excellent) to 5 (poor).

	Evaluator #1	Evaluator #2
J. Savola	1.5	2.4
K. Sparks	2.5	3.0
S. Steelman	5.0	2.7
C. Cunningham	3.5	2.9
A. Clark	4.5	2.7

[8] 1. (a) Compute standardized scores for C. Cunningham by both evaluators (*show your work*).
(b) In 1-2 clear sentences, interpret each of the two scores in simple language.
(c) Which evaluator ranked K. Sparks' application more favorably?

[6] 2. In Question 1 above, what is the purpose of computing standardized scores to compare the evaluators' rankings? Explain in 2 or 3 clear sentences why we wouldn't simply look at the average scores for each applicant to decide which received the most favorable rating.

3. In many countries, low-income households are entitled to subsidized public services such as water supply and electricity. Two of the greatest challenges in administering such programs are (a) to ensure that as many qualified families receive the assistance as possible, and (b) to prevent unqualified families from claiming resources to which they are not entitled.
Suppose you were asked to evaluate the success of the Lifeline Services Program of the country Hilpan in terms of these two objectives. Data on the eligibility of the country's families, as well as on whether they received Lifeline Services or not, are provided below.

	Eligible for Lifeline services	Not eligible for Lifeline services	
	Low income	Middle income	High income
Received Lifeline services	1,489	602	114
Did not receive Lifeline services	794	1,342	589

[3] a. What is the probability that a randomly selected low income family actually received Lifeline services? [*Show your work.*]

[3] b. What is the probability that a randomly selected family received Lifeline services, given that it was not eligible for the Lifeline program? [*Show your work.*]

[5] c. Based on your results to parts (a) and (b) above, what is your impression about the success of the Hilpan Lifeline Services Program in tackling the two challenges of targeting described above? [*2 or 3 clear sentences is plenty.*]

Probability distributions ☐

[7] 4. According to the 2001 *Kids Count Data Book*, an annual survey of quality-of-life indicators for children in the US, 80% of children under the age of 2 in the US have been fully immunized. Suppose a day care center in Massachusetts wanted to know whether immunization rates among its children were better than the national average. Staff of the center draw a random sample of 15 children at the center and check their immunization records. What is the probability that more than 80% of those children will have been immunized? [*Show your work.*]

[4] 5. In comparing their immunization rates for under-2 children to that of the nation, the day care center is assuming that the two populations (all under-2 children in the US and all under-2 children in their care) are also comparable. Name <u>one</u> way in which the two populations could differ systematically, <u>and</u> how this difference might bias the results of the center's analysis.

[8] 6. Suppose that the *Kids Count Data Book* also found that the average age of a US child when s/he completes his immunizations is normally distributed with a mean of 20 months and a standard deviation of 4.2 months. Suppose further that the Centers for Disease Control (CDC) wanted to launch a public health campaign to target parents of children who are more than six months overdue on receiving their two-year-old vaccinations. What percentage of all children will not have been immunized by the time they are 30 months old? [*Show your work.*]

[6] 7. Suppose that a survey of undergraduate college seniors was undertaken to explore the relationship between alcohol use and place of residence. The collected data appear in the table below, as does the test statistic that was obtained. Using the appropriate statistical table, test the null hypothesis that there is no association between alcohol use and on-campus *versus* off-campus residency. Use a .05 level of significance. To answer this question you must (a) state clearly whether you reject the null hypothesis, and (b) summarize your findings in one clear sentence that someone with no statistics/QR training could understand.

	No alcohol use	Light alcohol use	Heavy alcohol use
Lives on campus	35	21	47
Lives off campus	49	35	38

$X^2 = 5.22$

Hypothesis testing and confidence intervals

[10] 8. There are other statistical methods that can be used to examine the relationship between alcohol use and place of residence using these data. Suppose that you were interested only in the 'heavy alcohol use' category of these data. Conduct a test at the .05 level of significance of the null hypothesis that the proportion of students living off campus who are heavy alcohol users is equal to that of students living on campus. To answer this question you must (a) show your work, (b) indicate clearly the value of test statistic you compute, (c) indicate clearly the p-value associated with this test statistic, and (d) state clearly whether you reject the null hypothesis or not and what your conclusion about the hypothesis is in simple language.

	No alcohol use	Light alcohol use	Heavy alcohol use
Lives on campus	35	21	47
Lives off campus	49	35	38

Proportion of on-campus students with 'heavy' alcohol use: 0.46
Proportion of off-campus students with 'heavy' alcohol use: 0.31

[8] 9. Suppose that a school administrator obtains a copy of your report on alcohol use among students and asks you whether it's possible that "more than half of students who live on-campus report 'heavy alcohol use.'" Compute a 90% confidence interval for the proportion of on-campus residents who report 'heavy use." Show your work, and clearly indicate the values in your confidence interval. On the basis of that answer, what would your response to the administrator be and why? [*Hint: Some of the computation you need to complete for this question should have already been done above!*]

[7] 10. If you repeated the analysis in Question 9 above, which of the following changes would result in a wider confidence interval? [*Check all that apply*.]

_____ The sample size were larger.

_____ The confidence level were raised to 95%.

_____ The standard error of *p* were larger.

_____ All of the above.

Regression analysis [ ]

Consider the following regression analysis results from a recent study on poverty rates among children of different racial and ethnic backgrounds. The dependent variable for the model is "time in poverty," expressed as a proportion of the child's total months of life. (For example, a 10-month-old child who had lived in poverty for 4 months would have a value of 40 for the dependent variable).

Table 2 CDA 01-04

Regression Analysis of Percent Time in Poverty

Effects of Independent Variables

	Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t value	Significance
(constant)***	7.123	1.202		5.926	0.000
Percent Time in Single–Parent Family***	0.208	0.009	0.227	22.159	0.000
Percent Time on AFDC***	0.686	0.014	0.511	48.625	0.000
Mother's Math and Verbal Skill Level***	-0.157	0.011	-0.144	-14.224	0.000
Number of Children Ever Born***	3.068	0.249	0.115	12.333	0.000
Number of Years at First Birth Above 15***	-0.380	0.068	-0.056	-5.610	0.000
Percent Time Living in Northeast***	0.027	0.008	0.032	3.458	0.001
Percent Time Living in South***	0.056	0.007	0.081	8.206	0.000
Percent Time Living in West	0.009	0.008	0.011	1.148	0.251
Lives in Rural Area***	2.322	0.595	0.033	3.905	0.000
Race = Black	1.213	0.849	0.014	1.428	0.153
Race = Other	2.167	1.440	0.013	1.505	0.132

Analysis of Variance

	Sum of Squares	Degrees of Freedom	Mean Square	F value	Significance
Regression	13,287,489,597	11	1,207,953,600	896.556	0.000
Residual	7,399,516,125	5492	1,347,326		
Total	20,687,005,722	5503			

Explanatory Power of Model	R	R Square	Std. Error of the Estimate
	0.801	0.642	1160.744

Note: * Significant at more than a 90 percent level, ** Significant at more than a 95 percent level,
*** Significant at more than a 99.9 percent level.
Source: Center for Data Analysis WLS calculations of data from the National Longitudinal Survey of Youth, 1979-96

Notes: Reference category for Region is North Central; Reference category for Race is White

[6] 11. Explain in clear, simple language what "R Square 0.642" in the results summary means. [*Don't use jargon, and focus your discussion on this model rather than on regression analysis generally.*]

[8] 12. All t-statistics (or 't value,' as written in column 4 of the table) in regression analysis refer to the same general null hypothesis. State in clear, simple language (a) what this hypothesis is and (b) what it means to reject the null hypothesis in regression analysis.

[6] 13. In clear, simple language, interpret the value of the regression coefficient for the variable 'Race-other.' Use the unstandardized coefficient in your analysis. You do not need to comment on the significance value for this variable in your response. *[1 or 2 sentences is all you need. Don't use jargon; talk about this specific variable in this example.]*

[5] 14. Which of these two characteristics is expected to have a greater impact on a child's time in poverty? [*Show your work and indicate your answer clearly.*]

A) Having lived 75% of his/her life in a single-parent family
B) Being one of four children ever born to his/her mother

End of the exam.