

MIT OpenCourseWare  
<http://ocw.mit.edu>

11.220 Quantitative Reasoning & Statistical Methods for Planners I  
Spring 2009

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.

Quantitative Reasoning and  
Statistical Methods for Planners  
Midterm Exam

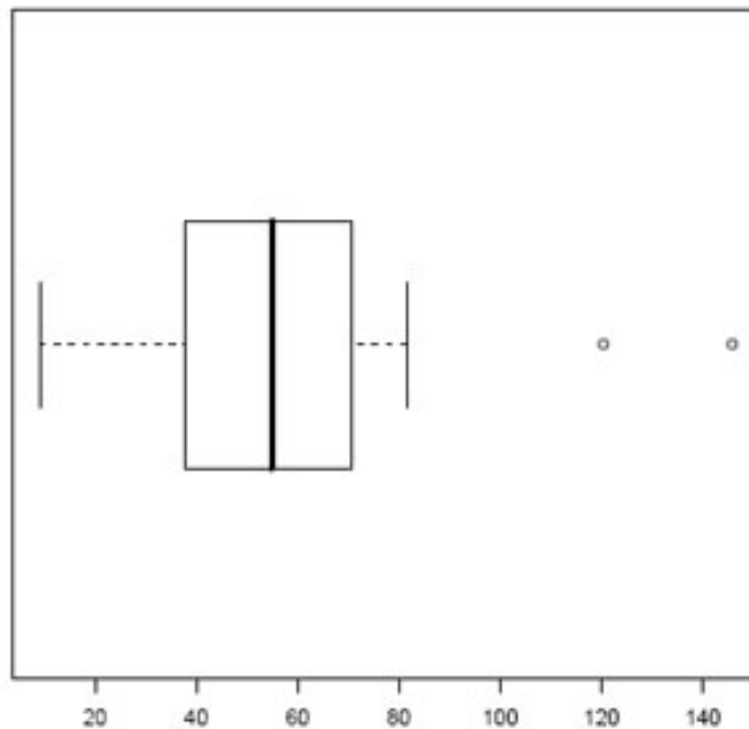
Ezra Glenn

Name: \_\_\_\_\_ email: \_\_\_\_\_

## Problems

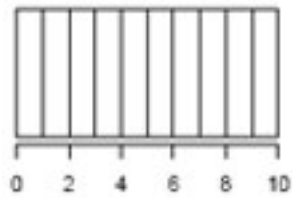
- For all problems involving questions of statistical significance, please use a 95% confidence level (i.e.,  $\alpha = .05$ ).
- For each problem, points (out of 100 total) are given in parenthesis.
- We have included some statistical tables at the end, in case you need them.

1. Please label/describe each of the components of the boxplot shown below. Be sure to explain the significance of each vertical line, as well as the points shown on the right-hand side. (8 points)

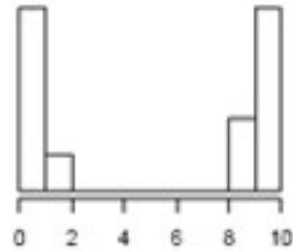


2. For the following questions, please consult the histograms A-D. (5 points each, total of 15 points)

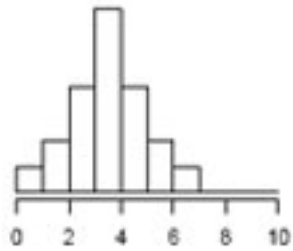
Histogram A



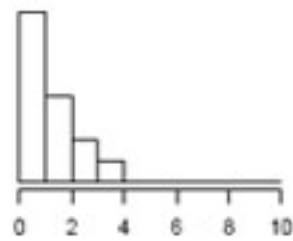
Histogram B



Histogram C



Histogram D



- (a) Rank the distributions, from the one with the lowest mean to the one with the highest mean.
- (b) Rank the distributions, from the one with the lowest median to the one with the highest median.
- (c) Rank the distributions, from the one with the lowest standard deviation to the one with the highest standard deviation.
-

- 
3. The director of community development for the city of Quantopolis is developing guidelines for a program to assist low-income individuals through tuition assistance grants. There is only enough funding for about 500 grants, and the director wants to be sure that demand in the first year of the program is as close to this figure as possible.

Assume that the mean individual income for adults in the City is known to be \$42,000, and that income is normally distributed with a standard deviation of \$10,000. If there are 40,000 adults in the city, and the director is confident that one in two eligible adults will apply to the program, what income threshold should she set in the guidelines to come closest to her target number of applicants? (15 points)

---



6. For the following questions (4 points each, 12 points total), you may assume that:

- 15% of the U.S. population is Latino
- one third of the U.S. Latino population lives in California
- 1 in 8 people in America live in California
- 7 in 10 Latinos in America are Catholic, and that this relationship holds true for all states and regions

What is the chance that a randomly-selected member of the U.S. population will be:

(a) a Catholic Latino living in California?

(b) a non-Catholic Latino who does not live in California?

(c) a non-Latino living in California?

---

7. A researcher interested in the effects of family size on antisocial behavior has conducted a study using records available from the juvenile court system. She has reviewed a random sample of 1,000 court cases in which there was a conviction of a juvenile defendant. Luckily for her research, the records also include information on the size of the offender's family.

She sorts the 1,000 cases based on the number of children in the offender's family and produces the following table to support her hypothesis that children from larger families are more likely to run into trouble with the law:

# of children	# of cases
1	60
2	109
3	147
4	205
5	227
6 or more	252

(You can assume these figures are considered to be “statistically significant” when presented in this way.)

Upon reading her results, a rival researcher argues that her analysis is flawed: families with many children are more likely to have a child who is a juvenile delinquent simply due to the fact *that they have more children*—that is, even if it were random, they would have more chances of “drawing” a delinquent child; this “effect,” he argues, tells us nothing about the relationship between large family size and “moral upbringing” or “antisocial behavior.” He contends that the figures in the second column of the chart should be corrected for this fact, and suggests dividing them by the number of children in each row to reflect a truer picture of how likely children from different size families are to wind up in court. When this is done, he argues, children from larger families are actually slightly *less* likely to be convicted.

Who is correct, and why? How might you modify this analysis to address this issue? (15 points)<sup>1</sup>

---

<sup>1</sup>Adapted from by C. W. Anderson and R. M. Loyne, *The Teaching of Practical Statistics*

---



8. You work for the Mayor in a city with 20,000 registered voters. You have conducted a survey to determine attitudes towards a new zoning proposal. Of the 200 responses, 110 said they were in favor of the proposal. (questions worth 5 points each, 15 points total)
- (a) Assuming this is a good random sample and there were no major flaws with the survey design, estimate the percentage of voters in support of the proposal and provide a 95% confidence interval.
- (b) When you present the results to the Mayor, he questions why you have included a term for the “standard error”—if there was an error with the survey, shouldn’t you do it again? How would you explain this term to him?
- (c) Before agreeing to support the proposal, the Mayor wants more of an assurance that at least 50% of the voters are in favor of it. What advice would you give him concerning this desire?
-

**Extra Credit:** If you still have time, and want extra credit, you can do this problem:

In *Exploratory Data Analysis*, Frederick Hartwig and Brian Dearing describe an approach to learn as much as possible concerning both the “smooth” and the “rough” from a set of data. Using this as a framework—and being *as specific as possible* about the key aspects of EDA, please show how these principles and techniques can be seen in the example of John Snow’s cholera research described in Edward Tufte’s *Visual and Statistical Thinking*. (up to 10 points)

---