

MONT 106N – Identifying Patterns
Information and Sample Questions for Midterm 1

General Information

The first midterm exam will be given in class on Monday, October 19. It will cover the material from Chapters 1 - 7 of Freedman, Pisani, Purves. This means in particular:

1. Issues about the design of studies – controlled experiments versus observational studies, questions like the ones from Problem Set 1 about whether conclusions are justified, possible confounding factors, and so forth. A question here might ask for a paragraph-length written response rather than just a calculation.
2. Histograms – be able to construct histograms for “small” data sets, explain the conceptual meaning of the histogram, etc.
3. Averages, medians, and standard deviations – be able to compute these for “small” data sets by hand. Also be prepared for conceptual questions about what these measures tell us about data.
4. Normal approximation to data – using the normal curve and the normal curve area table (questions similar to Problem Set 3), percentiles, etc.
5. Equations of lines in the plane.

No use of cell phones, pagers, I-pods, or any other electronic devices beyond a calculator will be allowed during the exam – turn them off and stow them in your backpack.

Important Note: I will be out of town starting from about noon on Wednesday, October 14, until the middle of the day on Sunday, October 18. I will have office hours on Wednesday morning and I will be in email contact, though. So if you have questions as you prepare, I will try to respond as soon as possible.

Sample Exam Questions

Be sure to provide explanations for your answers if that is asked for.

I.

- A) What is a *confounding factor*? What are the techniques used in well-designed controlled experiments to try to minimize their effects?
- B) Here is a passage from Dr. Dean Edell’s column in the San Francisco Chronicle of August 1, 1990:

“The more experienced a doctor is, the better. As obvious as that sounds, there are still too many people out there who never ask surgeons for a history of their work. The importance of knowing this is illustrated by this study. Peter Starek, a surgeon at the University of North Carolina, reviewed 460 heart valve replacement operations and found that only 4% of the patients of the three most senior surgeons died. But one junior surgeon lost almost 33% of his patients. Since that surgeon was technically

the best in the group, says Starek, something was obviously lacking – perhaps the kind of good judgment that grows out of experience...”

1. Was Starek’s study a controlled experiment or an observational study?
2. Is Starek’s claim justified by this evidence? Or are there possible confounding factors he is overlooking?

II.

- A) Find the equation of the line passing through the points $P = (2, 6)$ and $Q = (-5, 2)$ in the plane.
- B) Find the equation of the line perpendicular to the line in part A and passing through the origin $O = (0, 0)$.

III. The following data on people without health insurance by family income level for 2006 is from the Current Population Survey.

Family Income	No Insurance $N = 44,815$	All People $N = 293,834$
Less than \$25,000	18,590	70,478
\$25,000 to \$49,999	13,620	72,963
\$50,000 to \$74,999	6,445	55,258
\$75,000 to \$200,000	6,160	95,136

Notice $N =$ at the top of each column is the total number of people in the four class intervals for the column. Also, the number of people is in units of 1000.

1. For each of the four income class intervals, compute the percentage of No Insurance people in the class interval and the percentage of all people in the interval.
2. Use your percentages from 1. to construct two histograms, one for people without insurance and one for all people using the class intervals in the table. (Use the graph paper.)
3. By comparing your histograms, what can you say about the distribution of people without health insurance by family income in comparison to the distribution of all people by family income?
4. Is the average income or the median income larger for the families with no insurance? Explain how you can tell.

IV. The grades in a large university statistics class are normally distributed with a mean of 60 (out of a possible 100 points) and an SD of 12.

1. Draw a smooth histogram for the data.
2. If 200 students took the test, estimate how many scored between 70 and 80.

V. (Hypothetical) A big town near Worcester administers a spelling test to all students in fourth grade. The test consists of 100 words. The average for the number of words correctly spelled was 50 with a standard deviation of 10. Suppose a student was in the 70th percentile. What would you estimate for the number of words the student spelled correctly? (Be sure to show your calculations and draw normal approximations as needed.)