

Math 1040, Spring 2007
Instructor: Alla Borisyuk
Review and practice problems for Midterm 2
(March 14, 2007; 8:35 to 9:25 am)

Please, keep in mind that the test may include any of the material that we encountered in class/homeworks, even if it does not directly appear in the practice problems.

During the test you will be provided with a copy of Table B, formulae for computing mean, standard deviation, correlation and the regression line.

Primary topics covered

- Chapter 11: **Histograms and stemplots.** How to build them, how to describe them (symmetric and skewed distributions, center, spread, outliers)
- Chapter 12: **Numerical summaries of the distributions.** Five-number summaries, mean and standard deviation. Computing each of these, deciding when to use one or the other. Boxplots
- Chapter 13: **Density curves and normal density curves.** Normal computations. Mean and standard deviation of normal curves. Finding a fraction of individuals that belong to a certain range of values of the variable (using 68-95-99.7 rule or normal scores). Reversed calculation: Find a range of variable values that contain a certain fraction of individuals.
- Chapter 14: **Scatterplots and correlation.** Plot and describe scatterplots (direction, form, strength of the relationship, outliers, positive and negative association). Computing correlation. Interpreting correlation (positive, negative, near -1,0,1). Properties of correlation (between -1 and 1, computed for quantitative variables, does not depend on units of measurement or distinction between response and explanatory variables, affected by outliers)
- Chapter 15-: **Regression, Causation.** Calculating the least-squares line. Plotting the least-squares line. Using regression line for prediction. Correlation does not mean causation, lurking variables.

Practice problems

1. Stemplot, boxplot (Ch. 11, 12):

Here are the blood concentrations of a certain chemical (call it NN) in 13 healthy patients:

13 15 16 23 27 28 29 31 34 42 48 51 60

- Make a stemplot of these data
- Describe its shape
- Find the 5-number summary of these data.
- Why 5-number summary may be better suited for these data than the mean-standard deviation summary?
- Make a boxplot of these data.
- Here is the 5-number summary for a group of 13 patients with a liver disease:
21 32 40 55 58. Put a boxplot for these patients on the same graph as the healthy ones.
- Compare the distributions of the chemical concentration in healthy and diseased patients, listing as many points of comparison as you can
- Do you think the comparison of these distributions supports the claim that the liver disease is accompanied by the elevated blood level of NN?

2. Correlation, regression (Ch. 14,15):

Here are the data of average time spent reading per day per person and the rates of Alzheimer's disease for 5 unnamed countries.

reading time (minutes)	26	30	34	27	15
Alzh. rate (# deaths per 100,000 people)	1.9	1.7	1.5	1.7	2.3

We would like to know whether rate of death from Alzheimer's can be predicted from the average reading time.

- Find the mean reading time and the mean Alzheimer's rates in this data set.
- Find the standard deviation of both reading times and death rates
- Find the correlation (r) between reading times and death rates
- Does the r value that you obtained, indicates positive association, negative association or no linear relationship between these variables?
- Make a scatterplot (Hint: think about which variable to put on the horizontal axis)
- Find the equation for the regression line
- Plot the regression line on the same graph as the scatterplot
- Looking at the graph and the r value, do you think the following statement is justified: "Lack of reading causes high rates of death from Alzheimer's?"
- Based on your results, what Alzheimer's death rate would you predict for a country with average reading time 20 minutes per day per person?

3. Normal computations (Ch. 13):

Newborn babies have mean weight $\mu = 3200$ g and the standard deviation $\sigma = 500$ g

- What is the fraction of babies who have weight below 3100 g?
- What is the fraction of babies who have weight between 3100 and 3250 g?
- What is the weight level such that 50% of all newborns weigh below that level?
- What is the weight level such that approximately 35% of all newborns weigh above that level?