

# Study Guide – Math 107, Exam 1

Prof. Adam Loy, Spring 2016

## General Information

Exam 1 will be on Monday 4/18, in class. No books, notes, computers, or cell phones are allowed. A formula sheet will be provided for you.

Below is a rough outline of what material will be covered on Exam 1. To study, I recommend carefully going through class notes, homework problems, and handouts (especially this handout) actively (intermixing reading, thinking, solving problems, and asking questions). After reviewing those materials I recommend solving lots and lots of practice problems (there are many problems in the textbook that may be helpful—remember that the odd problems have solutions in the back of the textbook—I especially encourage you to complete the “Skill Builder” exercises). I will post a practice exam on Moodle.

## Chapter 1 - Collecting Data

- Definitions

Dataset, case/unit, variable, categorical variable, quantitative variable, explanatory variable, response variable, statistical inference, sample, population, sampling bias, simple random sample, bias, association, causation, confounding variable, observational study, experiment, randomized experiment, randomization, treatment

- Know how to answer the two important questions about Data.

- Who are we collecting data on?

- \* Corresponds to subjects or cases/rows in our dataset

- What data are we collecting?

- \* Corresponds to variables/columns in our dataset

- Be able to identify whether a variable is numerical or categorical.
- Be able to identify explanatory and response variable where appropriate.
- Know the relationships between population, parameter, sample and statistic.
- Recognize when it is appropriate to use sample data to infer information about the population.
- Be able to critically examine the way a sample is selected, identifying possible sources of sampling bias
- Be able to identify potential sources of bias that may arise in studies on humans.
- Recognize when associations imply causation.
- Be able to identify potential confounding variables.
- Be able to distinguish between an observational study and a randomized experiment.
- Know when claims of association are appropriate.

## Chapter 2 - Describing Data

- Definitions

parameter, statistic, sample proportion, population proportion, pie chart, bar chart, two-way table, stacked bar chart, mode, symmetry, skew, outliers, dot plot, histogram, box plot, mean, median, resistance, standard deviation, 95% rule, IQR, range, z-score, percentile, five number summary, side-by-side graphs, scatterplot, correlation

- Know how to construct (at least sketch out) and interpret the following tables/plots.

- two-way table, stacked bar chart, scatterplot, bar chart, dot plot, histogram, box plot, pie chart

- Be comfortable with which plots and numerical summaries we use for categorical and quantitative variables, and how to use/calculate them.

- Categorical

- \* Proportion/percentage, bar graph, pie chart

- Quantitative

- \* Dot plot, histogram, density plot, box plot, measure of center (mean, median), measure of spread (range, standard deviation, IQR), measures of position (five number summary – min, Q1, median, Q3, max)

- Be comfortable with describing a distribution (typically with a histogram, but could also be done with dot plots).

- Mode (or peaks)

- \* Unimodal - one mode/peak

- \* Bimodal - two modes/peaks

- \* Multimodal - more than two modes/peaks

- Shape

- \* Symmetric - mirror image of itself

- \* Skewed right - data mounded on left, trails off to the right

- \* Skewed left - data mounded on right, trails off to the left

- Outlier

- \* Extreme values of the variable

- Be comfortable with which measures of center/spread to use depending on the shape of the distribution, and why

- Symmetric distribution - use mean/standard deviation

- Skewed distribution - use median/IQR

- Know how to interpret all of the cells in a two-way table

- Understand the relationship between the shape of data in a scatterplot, association, and the sign and magnitude of the correlation coefficient

- Data shows positive trend → positive association → positive correlation coef

- Data shows negative trend → negative association → negative correlation coef

- Data shows strong linear trend → correlation coef close to -1 or 1

- Data show no relationship/cloud of random points → correlation coef close to 0

## Suggested Textbook Problems

In addition to the skill builder questions at the end of each section, I suggest looking at the following problems in the textbook:

Section	Exercises
1.1	1.16, 1.20
1.2	1.53, 1.58
1.3	1.87–1.91
2.1	2.13, 2.15
2.2	2.57, 2.66
2.5	2.169, 2.172

## Formula Sheet

The below formulas will appear on the formula sheet:

$$\bar{x} = \frac{\sum x}{n}$$

$$s_x = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}}$$

$$z = \frac{x - \bar{x}}{s}$$

$$\text{range} = \text{max} - \text{min}$$

$$\text{IQR} = Q_3 - Q_1$$

$$Q_1 - 1.5 \times \text{IQR}$$

$$Q_3 + 1.5 \times \text{IQR}$$

$$r = \frac{1}{n - 1} \sum \left( \frac{x - \bar{x}}{s_x} \right) \left( \frac{y - \bar{y}}{s_y} \right)$$