Exam 1 Study Guide for Math 12

How to study for the test:
- Study the lecture notes
- Make sure you are comfortable with the homework problems
- Make sure you know each item under “What You Need to Know” below
- Make sure you understand the formulas given on the attached formula sheet as you will get one just like it for the exam. You need to know how and when to use these formulas.
- Look through the old exam that is posted. This is mainly posted to give you an idea how an exam of mine might look, but also for you to “test” yourself on some of the material. Take the exam and then check with the posted key.

Calculator: You will be able to use your calculator for the exam. Make sure you know how to use it, as I will not answer any calculator questions during the exam. Realize that you might be asked to find certain things by hand using formulas but you can use your calculator to double check your answers.

What to bring to exam: pencil, eraser, homework, and your TI-83 or TI-84 calculator.

What You Need to Know

Terms and concepts that you need to be familiar with:

Chapter 1
- population/target population
- sample
- survey/poll
- census
- sample survey
- data, data set
- elements/subjects/individuals
- variable
- qualitative vs. quantitative variable
- continuous vs. discrete variable
- representative sample
- simple random sampling
- convenience sampling
- systematic sampling
- stratified sampling
- cluster sampling
- voluntary response sampling
- Grouped vs. Ungrouped data
- randomized experiment
- observational study
- double-blind experiment
- placebo
- placebo effect
- treatment group
- control group
- confounder
- biased vs. unbiased study
- voluntary bias
- social acceptability bias
- self interest bias
- leading question bias
- nonresponse bias
- sampling bias
- raw data
• Be able to construct and interpret a **frequency distribution** table, for qualitative and **ungrouped** quantitative data (grouped data in ch 2).

• Be able to find and interpret **relative frequency** and **percentage** distributions.

• Be able to find and interpret **cumulative** frequency and relative cumulative frequency distributions.

• Be able to construct and interpret **bar graphs**.

• Be able to construct and interpret **pie charts**.

• Understand how **graphs can be misleading** with improper positioning of the vertical scale or with the violation of the area principle.

• **statistics vs. parameter**

**Chapter 2**

• Be able to construct and interpret a **frequency distribution** table for grouped data.
  ✓ Be able to divide data into reasonable number of **classes** of equal width (except possibly some open ended class(es) at the beginning or end of data set).
  ✓ Make sure classes don’t overlap and that each data value belongs to exactly one class.

• Outliers (extreme values)

• Clusters

• Class midpoint

• Be able to construct and interpret **histograms**. Recognize different shapes of histograms/polylines: skewed left/right, symmetric, bell shaped, uniform.

• Be able to construct and interpret **polygons**.

• Be able to construct and interpret **line graphs** and **time-series plot**.

• Be able to construct and interpret a **stem-and-leaf display**.

• Be able to construct and interpret **dotplots** and **stacked dotplots**.

• Be able to find **mean, median, and mode** for ungrouped data and approximate the mean for grouped data (you need to use midpoints for each class).
  ✓ Know the characteristics of the above measurements, such as which are sensitive to outliers, which can be used for qualitative data also, etc.
  ✓ Understand the relationship between them. Ex. in a distribution that is skewed right, approximately where the mean, median, and mode are located, and where in relationship to each other?
  ✓ Make sure to use the same units as of the original data for these measurements.
  ✓ Recognize if a distribution is bimodal, multimodal, or has no mode.

• Be able to find the **range, variance and standard deviation** for ungrouped and data, and approximate the variance and standard deviation for grouped data (using the formulas).
✓ Have an understanding of the **meaning of a standard deviation**: Looking at some histograms, can you tell which distributions have smaller/bigger standard deviation? Could you make a rough estimation of what the standard deviation would be?

✓ Know the relationship between variance and standard deviation (variance is standard deviation squared).

✓ Make sure to use the same units as of the original data for the standard deviation. I will not ask you to include units for the variance.

✓ Make sure you use the correct formula for the variance based on whether the data is from a sample or population and from grouped or ungrouped data.

- Know what **outlier sensitive** means and be able to determine which parameters are more outlier sensitive.

- Understand the meaning and be able to use the symbols \( \bar{x}, \mu, s, \sigma, n, N \)

- Be able to use and interpret **Chebyshev’s Theorem**.

- Be able to use and interpret the **Empirical Rule**, in both directions. That is, given an interval, find about what percentage of the data values lie in this interval, or given a percentage of the data values, find what interval this corresponds to. Remember that this rule only can be applied to bell-shaped distributions.

- Be able to find and interpret **quartiles** and the **interquartile range (IQR)**. Be able to find upper and lower outlier boundaries.

- Be able to find and interpret **percentiles**, both directions. That is, be able to find the approximate value that correspond to a percentile, and be able to find which percentile rank a data value has.

- Be able to construct and interpret a **boxplot** (also called box-and-whisker plot). Remember, you are expected to be able to construct the kind of boxplot that shows outliers, as we did in class, and not the kind that the textbook shows.
  ✓ You need to be able to recognize outliers.
  ✓ You should be able to recognize whether a distribution is skewed left, skewed right, or symmetric.
  ✓ Based on above properties, be able to match up a histogram with its corresponding boxplot.

- Be able to find and interpret **z-scores**, both directions.
Formula Sheet

To find the approximate value of the $k$th percentile:

$$i = \frac{k}{100}(n + 1)$$

If $i$ is a whole number, the $k$th percentile is the number in the $k$th position in the ordered set of data. If $i$ is NOT a whole number, the $k$th percentile is the average of the closest two whole numbers to the $i$th position.

Percentile of a specific value $y$ in a data set

$$= \frac{x + 0.5y}{n} \cdot (100)$$

where $x =$ number of data values less than the number you want to find the percentile for

where $y =$ number of data values equal to the data value for which you want to find the percentile

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

Lower outlier boundary $= Q_1 - 1.5 \times IQR$

Upper outlier boundary $= Q_3 + 1.5 \times IQR$

**Empirical Rule**

- about 68% of all values fall within 1 standard deviation of the mean
- about 95% of all values fall within 2 standard deviations of the mean
- almost all values (about 99.7%) fall within 3 standard deviations of the mean

**Chebyshev’s Theorem:** At least \(1 - \frac{1}{k^2}\) of the data values lie within $k$ standard deviations of the mean

$$z = \frac{x - \mu}{\sigma}$$

**Ungrouped data**

Population mean: \(\mu = \frac{\sum x}{N}\)

Sample mean: \(\bar{x} = \frac{\sum x}{n}\)

Population Variance: \(\sigma^2 = \frac{\sum (x - \mu)^2}{N}\)

Sample Variance: \(s^2 = \frac{\sum (x - \bar{x})^2}{n - 1}\)

**Grouped data**

Population mean: \(\mu = \frac{\sum fm}{N}\)

Sample mean: \(\bar{x} = \frac{\sum fm}{n}\)

Population Variance: \(\sigma^2 = \frac{\sum (m - \mu)^2 \cdot f}{N}\)

Sample Variance: \(s^2 = \frac{\sum (m - \bar{x})^2 \cdot f}{n - 1}\)

Standard Deviation = \(\sigma = \sqrt{\sigma^2}\) or \(s = \sqrt{s^2}\)