MS Module 1 Background and data visualization (overview)

(The attached PDF file has better formatting.)

Readings: §4.3 the normal distribution; §4.6 probability plots, §7.1 point estimation general concepts and criteria (first three sections and fifth subsection), §7.2 methods of point estimation (first two subsections)

Chapters 1-7 are background (covered in Course P). The textbook uses these concepts; final exam problems assume you are familiar with them and tests the material in §4.3, §4.6, §7.1, and §7.2. This course assumes you know the material in

§1.2 Pictorial and tabular methods (especially dotplots and histograms)
§1.3 Measures of location (especially mean, median, and quartiles)
§1.4 Measures of variability (especially σ, σ^2, and the computing formulas; boxplots; outliers)

§2.4 Conditional probability (especially the multiplication rule and Bayes' theorem)
§2.5 Independence

§3.1 Random variables
§3.2 Probability distributions (especially parameters; the cumulative distribution function)
§3.3 Expected values (means and variances)
§3.4 Binomial distributions (distribution function; means and variances)
§3.3 Poisson distributions (distribution function; means and variances)

§5.2 Covariance and correlation
§5.3 Conditional distributions (especially regression to the mean)
§5.4 Transformations of random variables

§6.1 Statistics and their distributions
§6.2 Distribution of the sample mean (especially the central limit theorem and the law of large numbers)
§6.3 Distribution of a linear combinations (especially of differences between random variables)

Read §4.3, "the normal distribution," excluding the last subsection (the normal moment generating function). You are not responsible for the proof that the normal curve satisfies the requirement that its integral is one. This course uses the percentiles of the standard normal distribution, *z* values, *z* critical values, standardized residuals, and normal approximations to other distributions.

Read §4.6, "Probability plots," subsections on "sample percentiles" and "a probability plot," and the two supplementary readings on the discussion forum (see below).

Read §6.4 distributions based on a normal random sample (general concepts, not the mathematics of the χ^2 distribution, the *t* distribution, and *F* distribution). We use these distributions throughout the course (in addition to the normal distribution). The final exam does not test the mathematics in this section.

Read §7.1, "Point estimation - general concepts and criteria," subsections on "mean squared error," "unbiased estimators," "estimators with minimum variance," and "reporting a point estimate: the standard error." The final exam tests the bias, variance, standard error, and mean squared error of estimators. You are not responsible for the subsections on "more complications" and "the bootstrap."

Read §7.2 "Methods of point estimation," subsections on "method of moments" and "maximum likelihood estimation." You are not responsible for the subsections "some properties of MLEs," "large-sample behavior of the MLE," and "some complications." Estimation is the core of mathematical statistics, and these methods are tested on the final exam.

Probability plots are used throughout the textbook and they are tested on the final exam. The textbook section of probability plots is brief. Two readings on the discussion forum (and tested on the final exam) are required:

- The section "Normal Quantile Plot" from a reading on *Univariate Analysis*. Know the shapes of the normal probability plots for the four distributions in Figure 3.8. You are not responsible for the Lilliefors confidence bounds in the first graphic.
- Yunsi Wang, Tyler Steele, and EVA Zhang on the "QQ Plot," pages 1-7 only (until the section titled "QQ plot application").

The attachments to this posting have the supplementary readings.

The final exam tests four types of distributions:

- heavy tailed, such as *t* distributions with few degrees of freedom
- light tailed, such the uniform distribution
- right skewed (positively skewed), such as lognormal, Poisson, gamma, and exponential distributions
- left skewed (negatively skewed), such as capped distributions

Know the attributes of the normal distribution, the χ^2 distribution, the *t* distribution, and the *F* distribution. This course does not test the mathematical attributes of these distributions, but you should know their properties to understand the statistical concepts.

Question: Which parts of Chapters 1-7 must we review for this course?

Answer: Each candidate differs. Some sections are necessary for understanding the subsequent chapters, such as the discussions of measures of location (mean and median) and measures of variability, such as variance and standard deviation – but most candidates taking this course already know these topics. Some sections are new to some candidates, such as the moment generating function, but the textbook sections on the syllabus for this course do not use moment generating functions. Skim chapters 1-7 to know what they cover, so that if a module uses a probability measure you are not familiar with, you can review that section.