STATISTICS 101 SECTION 1 – FALL 2011

SYLLABUS AND OTHER INFORMATION

Lecturer: Dr W J Ewens

wewens@sas.upenn.edu
215 898 7109
Office: 324 Leidy Labs (corner 38th Street and Hamilton Walk)

Teaching Assistant*: Alex Goldstein alexlg@wharton.upenn.edu

* Please contact the TA with questions about JMP: specifically, how to install it and how to use it to solve statistical problems. (See more on JMP below.) You should also contact him about accessing Canvas.

The lectures and the material discussed in this section (Section 1) are very similar to that given in Sections 2/3. However there are some differences between Section 1 and Sections 2/3 in the order in which the material will be presented. Thus THE HOMEWORKS AND THE EXAMS IN THIS SECTION WILL BE DIFFERENT FROM THOSE IN SECTIONS 2 AND 3.

Homeworks will be handed out every Monday, starting on Monday September 12. Homework answers will be due in on the Monday following the day on which any homework is handed out. More details will be given in class.

There will be one mid-term exam, to be given on Thursday 27 October, 6 - 8 pm. More details as to where the exam will be held will be given later. The final exam is on Friday, December 16, 6 - 8 pm. The timing of this final exam is set by the university and is outside the control of the lecturers in the course.

The assessment in this course is by homeworks (20%), the mid-term exam (30%) and the final exam (50%).

The textbook for this course (and also for some subsequent Statistics courses) is Stine and Foster (SF), "Statistics for Business", Addison Wesley, 2011.

SYLLABUS

The content of this course falls into three broad categories, namely probability theory, descriptive statistics and inferential statistics. The reason why we even discuss probability theory, and why we discuss it first, will be given in the first lecture. A more detailed list of the specific topics covered within these three broad categories is given below, together with references to relevant pages in Stine and Foster (SF).

The course will be given in association with use of the statistical package JMP. You should either install this package on your computer soon or get used to using the Wharton computers that have it installed. Either way, it is a good idea to buy a copy of the JMP manual "JMP Start Statistics", SAS. Initially we won't use JMP much, but we will use it more and more as the course progresses.

1. Probability theory

1.1 The relation between probability theory and Statistics.
1.2 Events and their unions, intersections and complements. (SF 156).
1.3 Probability calculations involving events. Independence of events. (SF 156–166).
1.4 Conditional probabilities. (SF 174–195).
1.5 Discrete random variables and their probability distributions.
1.6 The fair die example in more detail.
1.7 The (identical) concepts of the mean μ and the expected value of a discrete random variable.
1.8 The variance and standard deviation of a discrete random variable. The expected value of a function of a random variable. Examples in the "fair die" case.
1.9 Various rules concerning means, variances and standard deviations. (SF 201–207).
1.10 *Into the unknown.* The thumbtack example. The concept of a parameter and the binomial parameter θ.
1.11 The binomial distribution and the parameter θ. The mean, variance and standard deviation of a binomial random variable. (SF 245–257).
1.12 The proportion of successes in the binomial context. The mean, variance and standard deviation of this proportion.
1.13 Many *iid* random variables whose distribution is known. (SF 230–236).
1.14 The mean and variance of a sum and an average of many random variables. Examples

in the "fair die" case. The mean and the variance of the difference between two independent random variables. (SF 326–330).
1.15 *Into the unknown again.* Many *iid* random variables whose distribution is unknown. Examples from the " unfair die" case.
1.16 Continuous random variables and their density functions. The mean and variance of a continuous random variable.
1.17 The normal distribution. The "Z" transform and the standard normal distribution. The use of normal distribution charts. The "one- and two- standard deviation rule". The central limit theorem and its uses. (SF 261–272).
1.18 Two dependent random variables. The concepts of the covariance (σ_{XY}) and the correlation (ρ_{XY}) between random variables X and Y. (SF 226–229).

2. Descriptive statistics

2.1 Measures of central tendency: the sample mean, the sample median and the sample mode.
2.2 Measures of variation in a data set: the sample variance s^2, sample percentiles and sample quartiles.
2.3 Association, regression and correlation. (SF 79-93, 109-119).
2.4 Aspects of curvilinear regression. Transformations. (SF 488-503).
2.5 The Sharpe ratio. (SF 208-210).
2.6 Quality control. (SF 325-332).
2.7 Time series. (SF 134-141).

3. Inferential statistics

3.1 Estimation

3.1.1 Estimation of the binomial parameter θ and the precision of this estimate. (SF 353–357).
3.1.2 The "unfair die" case again.
3.1.3 The average ($\bar{y}$) as an estimate of the mean μ.
3.1.4 s^2 as an estimate of the variance of each Y_j. (SF 203).
3.1.5 s^2/n as an estimate of the variance of $\bar{Y}$.
3.1.6 An approximate 95% confidence interval for the mean.
3.1.7 More on regression. (SF 464–473).

3.2 Hypothesis testing

3.2.1 General principles of hypothesis testing. (SF 378–381).
3.2.2 The concept of the p-value.

3.2.3 Testing hypotheses about the binomial parameter θ. (SF 384–386).
3.2.4 2×2 table testing for association (= testing for the equality of two binomial parameters). (SF–456–464).
3.2.5 The one-sample t test. (SF 388–391).
3.2.6 Further examples of hypothesis testing.