PSYCHOLOGY 10: STATISTICAL METHODS (M-F 9am, 420-040)
General Objectives and Syllabus

This course encourages you to acquire an understanding of how numerical methods can be applied productively in social science research. We will consider (a) a few aspects of research design (how to generate data that can be used to answer a specific question), (b) how to summarize a large data set by means of descriptive statistics (e.g., the mean), (c) what assumptions and what calculations are needed for us to make inferences from the data, and (d) how statistics can help us to "explain" our data. Many of the ideas discussed in the course are already part of your everyday language (e.g., probability, distribution, error, mean and relationship) and, although the precise definitions of, and the computations based on, these ideas will be novel, you should be mindful of this familiarity. You will be encouraged to write your statistics exercises with due respect for the rules of grammar (noting that the symbols, $<$, $>$, $\leq$, Σ, etc., stand for words or phrases), and for the ideal of arithmetic accuracy. **The daily *Lecture Notes* and weekly *Class Handouts* (also called *Quiz HW's*) will be posted on Coursework.**

Student Output

Weekly Quizzes. Starting on **Tuesday, 1/15/13**, there will be a **CLOSED-BOOK Quiz** in **420-040** and in **420-041** every Tuesday from **8:50 to 9:50 am**. There will be a quiz even on the Tuesdays following the MLK (1/19) and Presidents' Day (2/16) holidays. There will be no "final" exam; the last quiz will be on Tuesday, 3/10/15, and there will be no lectures after that date. You will be allowed to bring to the quiz up to **one sheet** (with formulae and notes that **you** prepare) **for every week of lectures** – 1 sheet for Quiz #1, 2 sheets for Quiz #2, …, 9 sheets for Quiz #9. The preparation of **your own** 'quiz-notes' facilitates learning, and this is why you may bring **only your own notes** to the quizzes. Students with last names starting with **A-O** should take the quiz in **420-040**; those with last names starting with **P-Z** should take the quiz in the **420-041**.

Your grade in the course will be based on your performance on these 9 quizzes. To earn a "CR" (i.e., C- or better), a student needs to get 60% or more; and to earn an "A-" or better, a student needs to get 90% or more. **Missing scores are set to 0**. Students do **not** simply get to drop their lowest score, and **missing a quiz will hurt your grade**. Students are allowed to "replace" their lowest score from among the first **6** quizzes by their average score on **all** the quizzes 1 to 6; students' scores on the last **3** quizzes enter into the final grade without any modification. (Frankly, this is my attempt to discourage you from starting your vacation a week or two ahead of schedule!) For helpful tips on taking quizzes and preparing for them, please read the memos, "**Quiz.Guidelines.pdf**" and "**Quiz.Preparation.pdf**", in the *Read.me.now* folder.

Quiz regrades. Each week we provide solutions to the quiz problems, and you are encouraged to use these solutions to improve your understanding of the material. If you think you have been unfairly graded, you may consider submitting your blue-book for a regrade to one of the co-Head TAs, Molly or Kody, or to the TA who graded the blue-book (identified by the initials on the cover). Please submit your regrade requests **within 2 weeks of the quiz**, and try to limit the number to **two requests** for the quarter. We keep track of these requests, and **I hope that you will not abuse this facility** because regrade requests are time-consuming and often have little impact on your total score (e.g., because the 're-grader' notices both positive and negative deviations from 'fair' grading in a given blue-book!).

"*Quiz Homeworks*". I feel that practice is a necessary prerequisite for the understanding of statistics. Some of this practice will start in class, but the bulk of it will have to be done as "homework" and, in this 5-unit class, I hope that you will spend about 15 hours a week on the course material. Obviously, practice is facilitated if the assigned material is read and if good notes are taken in class. Please read all my handouts, especially the weekly **Quiz HW-n** handouts, which give helpful summaries of relevant information from my lecture notes, the course textbook, and other sources. However, the proof of understanding an idea is the ability to work through the appropriate exercises. There will be homework exercises that you will be expected to do; these assignments represent the minimum you should do in order to learn a reasonable amount of statistics. The homework will *not* be graded, but I hope that you will contact the TAs or myself as soon as you encounter difficulties. Our Office Hours and Section Hours are given in a separate handout.

Sections (optional) and Office Hours

Sections are held at all hours on Monday, and office hours are distributed across Wednesday, Thursday and Friday. Students need not attend sections if they are mastering the material easily. Our hope, however, is that you will attend as many sections as you need to in order to do well. Note that the new material for a quiz is presented mainly in the **WThF** lectures. Therefore, attendance at lectures and office hours on these days can be rewarding.

Recommended textbook: *Basic Statistics for the Behavioral Sciences*, by Gary W. Heiman (5^{th} **or other ed**).

Some other books that you may want to look at are: an **online text**, *SurfStat.australia*, available at http://surfstat.anu.edu.au/surfstat-home/ ; *Statistics* by Freedman, Pisani, Purves and Adhikari; and *Statistics*, by Spiegel (Schaum's Outline Series). **Please note**, however, that some students have felt that they did not use the recommended text enough to justify the purchase of the book; these students felt that they need not have bought the book. I accept this evidence. For other students, it is likely that the recommended text would be helpful because, e.g., the text might be easier to read than my handouts, or the text might have more examples and explanations than my handouts. Also, the recommended text could come in handy when you attend graduate or professional school!

On-line Applets: We often need to calculate the **probability** that a variable (e.g., X = the number of successes in an experiment, or X = the chi-square index) lies in **a certain range**: e.g., P($X > 3.84$), P($X < 5.2$) or P($-1.2 < X < 2.1$). Statistical Tables provide these probabilities. For example, a hard-copy of the Table for X = the chi-square index is given in Quiz HW-1. Very conveniently, these Tables are provided in applets on the web, e.g., the SurfStat Tables at http://surfstat.anu.edu.au/surfstat-home/tables/normal.php. **Please familiarize yourself with them immediately**.

Introduction to R: In this introductory course, I prefer to emphasize **'by hand'** calculations so as to facilitate the understanding of statistical practice. Because of this, and because our enrollment is more than 200, I've chosen to not include computing packages as a formal part of our syllabus. However, in the later weeks of the course, I will refer to output from R, a powerful package that we use in our graduate statistics courses in the Department. In the **R Intro** folder in Coursework, there is a self-paced online tutorial in R for interested students in Psych 10/Stat 60 who expect to be doing quantitative research in the near future. I hope that some of you will find this tutorial helpful, and that you will consult the TAs or me if you need quick tips on using R.

COURSE OBJECTIVES (Weeks 1 and 2)

Readings and Exercises

1) The handout "Quiz HW-1".
2) *Heiman*, Chaps. 1 & 2. Please read these in the first two days of the week.
3) Chap 3. See Quiz HW-1 for `rf at mode', `range', and `semi-interquartile range' as measures of *dispersion*.
4) Chap 15 on chi-square, plus Exercises at the end of the chap. Also, the various Problem Sets.
5) I will give you a method for calculating the `necessary' sample size in a planned study.
6) Handout and **Exercises** on Probability Theory. *Heiman*, A-3, on "Probability".

Objectives

1) Distinguish between (a) sample and population, (b) variable and *value* of a variable, (c) description and explanation.
2) Identify qualitative (or categorical or nominal) variables, versus quantitative (e.g., ordinal, interval, ratio and absolute) variables. Note that discrete quantitative variables are sometimes treated as continuous, and vice versa.
3) Distinguish between (a) random and non-random variables, and (b) dependent and independent variables.

4) Construct frequency and relative frequency distributions for nominal data. Describe this distribution by stating its `location' (the *mode*) and `dispersion' (the *relative frequency at the mode*).
5) Construct a bivariate frequency distribution. Describe this distribution by (a) describing the two *univariate* (or marginal) distributions (as in (4) above), and (b) using the cell frequencies to state whether or not there appears to be a *relationship* between the two variables.
6) Calculation of expected frequencies: $E_i = Np_i$; $E_{ij} = (R_i C_j)/N$. (Compare the formulae in *Heiman*.)
7) Calculate the χ^2 (chi-square) measure of (a) goodness of fit between expected and observed frequencies, (b) the association between two variables.
8) Use the Chi-Square Table for statistical inference (i.e., for the `goodness-of-fit test' and for the `test of contingency' between two variables). (*Note*: An abridged chi-square Table is given in Quiz HW-1; please hold on to it for much future use!)
9) Use the Chi-Square Table to calculate *how large a sample is needed* so that one can reliably detect small differences.
10) Simple laws of probability; mutually exclusive events, and the "addition rule"; an exhaustive set of events; joint probability; the definition of independent events, and the "multiplication rule", *P(A and B) = P(A) P(B)*; conditional probability and Bayes' Rule; base rates; statistical independence between random variables.

Syllabus for Weeks 3-9

So far we have had some experience with the testing of statistical hypotheses that belong to the two most frequently encountered types of hypotheses. The goodness of fit test answers the question: Is the *univariate* distribution, from which a single sample of size *N* has been drawn, that which is specified in the null hypothesis, H_0? The χ^2 test provides a satisfactory answer provided *N* is large (for which reason it is called a *large sample test*), and provided the expected frequencies are not too small. The contingency test answers the question: Are the two variables that have been observed in the sample significantly related to each other?

Part of the logic of the contingency test, especially the meaning of *independence* between variables, comes from the Simple Laws of Probability. For example, the formula, $E_{ij} = (R_i C_j)/N$, is derived from the "product rule" of Probability. In addition, we will consider questions, such as: What is meant by the concept, `*probability distribution* of a test statistic'? Why does the test statistic, $\Sigma[(E_i - O_i)^2/E_i]$ , have the $\chi^2$ distribution shown in QHW-1? In what way does a `large' value of χ^2 depend on the *df* and on the significance level? How should one construct test statistics in other situations, such as, when (i) the sample size is small, (ii) the data are not frequencies but scores (e.g., 55 inches), or (iii) the situation is novel?

In order to improve our understanding of testing procedures, we will study a few exercises in Discrete Probability (flipping coins, etc.). These exercises help us to understand some of the mathematical models that can be used to derive the expected frequencies in the χ^2 formula. Then we will consider *empirical* distributions for interval scale variables, how to summarise them, how to pool them, and what happens when you linearly transform them. Next we will consider these same issues with *theoretical* distributions, the distributions of discrete and continuous *random variables*. Next we consider the big one, the *Normal* distribution, and derive from it the *Z*-test, *t*-test, χ^2 test, and *F*-test. Then we consider *Correlation and Regression*; and *1- and 2-way Analysis of Variance.* If time permits we can chat about special topics such as *multiple regression*, and *decision theory*.

Weeks 3-4

The *Readings and Exercises* include a Class Handout on Empirical Distributions, and *Heiman*, Chap. 6, and A-1, A-2. We will depend almost exclusively on Class Handouts for the results on pooling estimates of mean and variance across samples; *expected value* of a random variable, *X* (μ or $\mu_x \equiv E(X)$) and variance (σ^2 or $\sigma_x^2 = E[(X - \mu)^2]$) of *X*; the use of μ and σ^2 in describing gambles and in rationalising people's preferences among gambles; the linearity of the operator, *E(.)*; the expected value and variance of $\overline{X}$; the expected value of the sum of squares, *E(SS)*; proof that the mean of *Z*-scores is 0, and the variance of *Z*-scores is 1 [$Z = (X- \mu)/\sigma$].

Week 5

Chap. 6. Here again we will depend heavily on Class Handouts for material on probability *density* functions (for continuous random variables); simple continuous distributions (e.g., the uniform or rectangular distribution.); the Central Limit Theorem and the Normal distribution; calculations involving the Normal distribution; sampling distributions; 1- and 2-tailed Z-tests of hypotheses about μ; the Binomial distribution and the Normal approximation to the Binomial; the sign test.

Week 6

Chaps. 10, 11, 12. Confidence intervals for μ and p; t-tests for 1-sample and 2-sample problems.

Week 7

Chaps. 7 and 8. Correlation and Regression; amount of explained variance; correlation and causation; multiple regression.

Weeks 8 and 9

Chaps. 13, 14, and A-4, A-5. Analysis of Variance.