

# “Case-based Social Statistics II”

Spring 2002

Professor: Mark S. Handcock

## Syllabus

**MOS**  $\equiv$  *Mind on Statistics* text.

**CHS**  $\equiv$  *A Casebook for a First Course in Statistics* text.

Lecture Date	Topics	Readings (see website)
12	Introduction to Statistical Inference. Random sampling and sampling distributions: sampling distribution of the sample mean $\bar{X}$ .	<b>MOS</b> 8.1 – 8.3 $\hookrightarrow$ The Central Limit Theorem for census data (p. 48 - 50)
13	The standard error of the mean. The Central limit theorem: theory and applications. The standard error of a binomial proportion.	<b>MOS</b> 8.4 $\hookrightarrow$ The Central Limit Theorem for census data
14	Inference about the population mean: point and interval estimation of a population mean; Confidence intervals based on the $t$ -distribution: assumptions and interpretation.	<b>MOS</b> 9.1 – 9.3 $\hookrightarrow$ The return on stocks in the Over the Counter market
15	Prediction intervals for a future observation. Inferences about proportions: point and interval estimates; confidence intervals for a binomial proportion.	<b>MOS</b> 9.1, 9.5 $\hookrightarrow$ Volume and weight from a vineyard harvest (p. 95-99)
16	Testing statistical hypotheses: concept of a null hypothesis; testing using confidence intervals. Formal framework of hypothesis testing: the alternative hypothesis; Type I and Type II errors.	<b>MOS</b> 6.1, 6.2, 6.4
17	The practice of hypothesis testing. Statistical significance, Test levels and the $p$ -value of a hypothesis test. Qualms about $p$ -values and alternatives	<b>MOS</b> 6.2
18	One-sided testing. Testing of hypotheses about proportions. Inference about the population median.	<b>MOS</b> 6.2
19	Hypothesis testing – comparisons between groups. Comparisons when the samples are paired.	<b>MOS</b> 7.1, 7.2, 7.4 $\hookrightarrow$ Baseball free agency: do teams get what they pay for?
20	Two-sample $t$ -test and non-parametric alternatives.	<b>MOS</b> 7.3 – 7.5 $\hookrightarrow$ Reporting of sexual partners by men and women
21	The analysis of qualitative data: nominal and ordinal discrete data. Contingency tables, $\chi^2$ tests of independence.	<b>MOS</b> 8.1 – 8.3 $\hookrightarrow$ Air bags and types of automobiles

22	Introduction to simple linear regression: Modeling the relationship between variables. Linear relationships: graphical (scatter plot) and theoretical (model building) analysis.	<b>MOS</b> 9.1 ↔ Productivity versus quality in the assembly plant
23	The principle of least-squares. The assumptions of linear least squares regression. Determining the regression line.	<b>MOS</b> 9.2 – 9.4 ↔ Emergency calls to the New York Auto Club
24	Measuring uncertainty in regression: $\sigma$ , the standard deviation of $Y$ about the regression line. The distribution of the regression estimates; standard errors of estimates.	<b>MOS</b> 9.5 ↔ PCB contamination of U.S. bays and estuaries
25	Confidence intervals and testing: confidence intervals for the regression estimates. An extended example: Maintenance costs.	<b>MOS</b> 9.5
26	What regression explains about the sources of variation: $R^2$ : the coefficient of determination. Prediction intervals for a single new value of $Y$ at a given value of $x$ .	<b>MOS</b> 9.6 ↔ Predicting international adoption visas from earlier years
27	Checking assumptions in regression: examination of residuals, diagnostics and residual plots. The dangers of extrapolation.	<b>MOS</b> 9.6 ↔ Long and short term performance of stock mutual funds
28	Multiple linear regression. Nature of the model: least-squares fitting of a plane & the geometric picture.	<b>MOS</b> 10.1 – 10.3, 11.1 ↔ Estimating a demand function
29	Interpretation and estimation for the regression coefficients: confidence intervals and hypothesis testing for the regression coefficients.	<b>MOS</b> 11.1 ↔ The possibility of voting fraud in an election
30	Developing a regression model: Predicting the response at given values of the explanatory variables; deciding if an explanatory variable significantly improves the model. Testing the overall fit of the regression line.	<b>MOS</b> 10.4, 11.2, 11.3 ↔ Prediction of the time interval between “Old Faithful” eruptions
31	The Practice of Regression: Representation of qualitative variables in regression: “Indicator” variables. Transformations of variables. Choosing a model from all possible models.	<b>MOS</b> 11.4, 11.5, 14.1, 14.2 ↔ Purchasing power parity and high inflation countries
	Summary and Review class (how far have we come)	