Psychology 200 Statistics: An Activity-Based Approach

Section 003  Tu 7:00-9:50 p.m. Allbritton 204

Faculty Instructor: Emil Coman  ecoman@wesleyan.edu  ph. #2990
Office Hours Tu 5:00-6:30 pm; Allbritton 204 (104A alternative)
Mailbox in the 104 Judd Hall mailroom.

Course Description

In this course, you will learn how to interpret, organize, describe, and make inferences from psychological data. Specific goals of this course are for you: (1) to develop an understanding of the basic principles of statistical reasoning from a causal perspective (2) to learn how to interpret and apply statistical inference tests commonly used in psychology, and (3) to learn how to use Excel, AMOS and other statistical software towards achieving these goals. The course is geared towards psychology students who need to understand and critically evaluate published psychological data, and those who will eventually be conducting their own studies and/or analyzing data. However, the course is also relevant to anyone who wishes to better understand and evaluate statistical data and research encountered in everyday life (such as the results of studies reported in the media). The course is intensive and requires the devotion of considerable time to reading materials and doing problem sets and working to grasp complex concepts.

The course will be organized around work sessions aimed at together-discovering by examples and applied exercises key concepts and statistical techniques. Lectures are not going to be formally delivered, nor are we expected to rely heavily on Powerpoints; PPTs will be rarely used, and when used, it will be primarily to keep us on track. We will instead approach each topic as a hands-on mutual-discovery investigation motivated by current concrete examples related to psychological research: media coverage of research, claims heard/seen about numbers, as well as published research from psychological peer-reviewed literature. Therefore the use of Internet searches in class are one such mechanism of discovery; getting fast answers from reliable sources to your psych research questions is a skill that will stay with you; these will be however guided together-searches, i.e. focused on specific tasks. I will lead you in this discovery using a visual approach, with formulas and the math behind the graphical depictions to be derived from them. This is then expected to assist you when reading the chapter after the class; we will recap every week some issues from the prior week.

The ordering of topics to be discovered varies from the textbook one: we will use one that is most relevant for statistical practice, following Humberto, & Howland (2006) below. So: regression 1st, then some 1 variable statistics, then measurement, multiple regression, comparing groups on nominal, categorical, and continuous variables, etc. We will be able to measure how much we advance every week, by taking a quick self-report survey of the grasp of topics we aim to understand; this will provide us with real data on how things start making sense/not.

Class Website

Check the Moodle website for this class, even though most of our communication will be done in class. However, I will use it to post class materials, all handouts, and grades, links to online stuff, etc. You can reach the website from Moodle within your portfolio.
RESOURCES: Textbooks, Software, Online materials

The book should be available at the Broad Street Bookstore and is also on reserve at Olin Library. The book is required, but its utility is somewhat secondary: we want to make sure we cover all the topics listed in there. Also essential is a PC/laptop/computing device (brain chip?) that can handle worksheets and calculations [Excel is the 1st aim, Amos 2nd, SPSS 3rd.


Free online

AMOS 5 Guide AMOS 16 Guide AMOS 20 guide

Not required


Additional PDF and Excel materials

Software

1. MS Excel (OpenOffice or Google docs )

2. AMOS will be used in class for illustration more often than SPSS (because it can do anything SPSS can, for the purpose of this class). There is a free perpetual Amos 5 Student version: http://amosdevelopment.com/download EXE file

3. SPSS statistical software will also be used. A full version of SPSS is available on all PCs on campus. Please leave time in your schedule to work in the computer lab.

Online materials

E.G. Statnotes: Topics in Multivariate Analysis, by G. David Garson faculty.chass.ncsu.edu/garson/PA765/statnote.htm Became now www.statisticalassociates.com

Grading policy

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Points</th>
<th>Cumulative points</th>
</tr>
</thead>
<tbody>
<tr>
<td>14 Weekly reactions</td>
<td>12 * 3</td>
<td>36</td>
</tr>
<tr>
<td>3 problem sets</td>
<td>3 * 10</td>
<td>30</td>
</tr>
<tr>
<td>3 quizzes</td>
<td>3 * 10</td>
<td>30</td>
</tr>
<tr>
<td>1 group project</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>1 final exam</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>TOTAL</td>
<td></td>
<td>161</td>
</tr>
</tbody>
</table>

The final class grade will be the percentage of points received out of 161 points. These points will come from the 12 weekly reactions assignment test grades (3 pts x 12 = 42), the 3 problem set grades (10 pts x 3 = 30), the 3 quizzes grades (3 pts x 10 = 30), the group project (25 pts), and the final exam grade (40 pts). While the point system is fixed, I reserve the right to adjust individual final grades to reflect great improvement or serious decline in performance over time. Note that A+’s are reserved for near-perfect scores.

Course grade cutoffs: A+ 98, A 93.35, A- 90; B+ 86.65, B 83.35, B- 80; C+ 76.65, C 73.35, C- 70 (etc.)
<table>
<thead>
<tr>
<th>Week</th>
<th>Date</th>
<th>Topic</th>
<th>Chapters***</th>
<th>Excel materials</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>1/29</td>
<td>Variability and applied modeling - Simple linear regression</td>
<td>5: Regression</td>
<td>Reg.xls + Reading 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4: Scatterplots and Correlation</td>
<td></td>
</tr>
<tr>
<td>02</td>
<td>2/4</td>
<td>One variable modeling</td>
<td>1: Picturing Distributions with Graphs</td>
<td>MonteCarlo.xls BoxModel.xls</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2: Describing Distributions with Numbers</td>
<td>PresidentialHeights.xls</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>8: Producing Data: Sampling</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>11: Sampling Distributions</td>
<td></td>
</tr>
<tr>
<td>03</td>
<td>2/12</td>
<td>Summarizing data, inference to population</td>
<td>18: Inference about a Population Mean</td>
<td>Histogram.xla</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1st encounter with unobserved variables</td>
<td>14: Confidence Intervals: The Basics</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>15: Tests of Significance: The Basics</td>
<td></td>
</tr>
<tr>
<td>04</td>
<td>2/19</td>
<td>Variable types: nominal, categorical, continuous</td>
<td>3: The Normal Distributions</td>
<td>MonteCarlo.xls BoxModel.xls</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>13: Binomial Distributions</td>
<td>PresidentialHeights.xls</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>20: Inference about a Population Proportion</td>
<td></td>
</tr>
<tr>
<td>05</td>
<td>2/26</td>
<td>Groups and populations</td>
<td>19: Two-Sample Problems</td>
<td>TwoBoxModel.xls CPS90Workers.xls</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2nd encounter with unobserved variables</td>
<td>21: Comparing Two Proportions + Supplemental material: AMOS</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>guide: 35 Mixture Modeling without Training Data</td>
<td></td>
</tr>
<tr>
<td>06</td>
<td>3/5</td>
<td>Comparing groups from different populations: nominal and categorical variables</td>
<td>6: Two-Way Tables</td>
<td>Crosstabs.xls ChiSquareDist.xls</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>23: Two Categorical Variables: The Chi-Square Test</td>
<td></td>
</tr>
<tr>
<td>07-8</td>
<td>-</td>
<td>3/12&amp;19 Spring break</td>
<td>- enjoy</td>
<td></td>
</tr>
<tr>
<td>09</td>
<td>3/26</td>
<td>Comparing groups from different populations: continuous variables</td>
<td>25: One-Way Analysis of Variance + Supplemental material: AMOS</td>
<td>AMOS model</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>guide: 12 Simultaneous Factor Analysis for Several Groups</td>
<td>+ Reading 3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TwoBoxModel.xls CPS90Workers.xls</td>
</tr>
<tr>
<td>10</td>
<td>4/2</td>
<td>Measurement in psychological research</td>
<td>9: Producing Data: Experiments + Supplemental material: 11</td>
<td>Measure.xls</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3rd encounter with unobserved variables</td>
<td>The Measurement Box Model</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>4/9</td>
<td>Correlation and its Dangers</td>
<td>4: Scatterplots and Correlation</td>
<td>Correlation.xls + Reading 2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>But: covariances and their utility</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>4/16</td>
<td>Multiple linear regression</td>
<td>16: Inference in Practice</td>
<td>Galileo.xls</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Nonlinear functional forms</td>
<td>24: Inference for Regression</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>4/23</td>
<td>Group projects presentations + Reality and imaginary in psychological research</td>
<td>Your slideshows!</td>
<td>Powerpoints</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+ Group projects presentations + Reality and imaginary in psychological research</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>4/30</td>
<td>Extensions: Latent variables and simultaneous regressions OR Putting all in the larger perspective</td>
<td>Materials will be made available</td>
<td>AMOS models</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+ Reality and imaginary in psychological research</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>5/7</td>
<td>Analyzing ourselves: how did we do?</td>
<td>Weekly progress data</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Review for final exam</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>5/14 Final exam time TBA</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*** Chapter readings will be recommended for limited sections announced in class.***
**Suggested readings.** A number of articles are proposed (we can change them if you have other suggestions); we will make sense of them together in class after reading individually before class. Focus only on the methods and results for this class.

<table>
<thead>
<tr>
<th>READINGs from applied literature</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>+ Wainer, H. (2007). The Most Dangerous Equation. American Scientist, 95(3), 249-256.</td>
<td>Variability $\sigma(\overline{X}) = \sigma/\sqrt{n}$</td>
</tr>
</tbody>
</table>

**Weekly reactions.** These will be 1-2 paragraph reactions/responses/commentaries to a question posed by the instructor; email them/send though Moodle.  
 **Problem sets.** Three problem sets will be handed and then be due the following week; these will be applications of Excel/Amos demonstrations like those done in class.  
 **Group project.** The group research project & group presentation: groups of 4-5 students will choose a topic of their own interest for an applied psychological research feasible during one semester. A main research question or hypothesis will be formulated, an appropriate research design will be identified, then data will be either obtained (secondary data sources) or collected (if needed), and an analysis will be performed (with the instructor’s assistance) to answer the research question. The final product will be a class presentation as a group, based on a Powerpoint slideshow.  
 **Quizzes.** Will be made up of 10 multiple-choice questions.  
 **Final exam.** Will consist of ~20 multiple-choice questions and 3 open-ended questions.  
 **Class participation.** I will take attendance, and I expect to be told if you will miss class for a serious reason. The design of the course presumes attendance (for class demonstrations, learning Excel, Amos, SPSS, taking quizzes, etc.), and so I will expect (and look forward) to see you all regularly.  
 **Exceptional Circumstances.** The assignment and grading policy is designed to take into account exceptional and/or unforeseen circumstances such as personal illness, family emergencies, having multiple tests on same day, oversleeping, desiring to leave campus early at final exam time, sports schedules, and so forth. You should see the flexibility already built into the policies as being precisely for these circumstances, and should not request any other special arrangements except in the case of an incapacitating and extensive medical emergency extending over many weeks. The latter will be given further consideration only if accompanied by a doctor’s letter indicating that you were unable to attend class or to perform class-related work during a stated time period.  
 **Disclaimers**  
 **Syllabus.** I reserve the right to change any information in the syllabus, when necessary, with notice given to you during class. It is your responsibility to keep up with any changes.  
 **Weather.** If I am unable to make it to campus or must leave early because of weather conditions, I will send you an email (as it is often not possible for me to get to the classroom to leave a note on the door).  
 **Comments.** Your comments and suggestions are welcome, during and after class; you can come to my office hours. No appointment is needed for coming to my office hours. I am happy to respond to questions about homework or other types of statistical questions by email at any time, and this is usually easier than setting up an in-person meeting; I just do not make guarantees about responding immediately, especially during non-business hours.
Chapter 1: Picturing Distributions with Graphs
- Individuals and variables
- Categorical variables: pie charts and bar graphs
- Quantitative variables: histograms
- Interpreting histograms
- Quantitative variables: stemplots
- Time plots

Chapter 2: Describing Distributions with Numbers
- Measuring center: the mean
- Measuring center: the median
- Comparing the mean and the median
- Measuring spread: the quartiles
- The five-number summary and boxplots
- Spotting suspected outliers
- Measuring spread: the standard deviation
- Choosing measures of center and spread
- Using technology
- Organizing a statistical problem

Chapter 3: The Normal Distributions
- Density curves
- Describing density curves
- Normal distributions
- The 68-95-99.7 rule
- The standard Normal distribution
- Finding Normal proportions
- Using the standard Normal table
- Finding a value given a proportion

Chapter 4: Scatterplots and Correlation
- Explanatory and response variables
- Displaying relationships: scatterplots
- Interpreting scatterplots
- Adding categorical variables to scatterplots
- Measuring linear association: correlation
- Facts about correlation

Chapter 5: Regression
- Regression lines
- The least-squares regression line
- Using technology
- Facts about least-squares regression
- Residuals
- Influential observations
- Cautions about correlation and regression
- Association does not imply causation

Chapter 6: Two-Way Tables
- Marginal distributions
- Conditional distributions
- Simpson's paradox

Chapter 7: Exploring Data: Part I Review
- Part I summary
- Test yourself
- Supplementary exercises

PART II: FROM EXPLORATION TO INFERENCE
Chapter 8: Producing Data: Sampling
- Population versus sample
- How to sample badly
- Simple random samples
- Inference about the population
- Other sampling designs
- Cautions about sample surveys
- The impact of technology

Chapter 9: Producing Data: Experiments
- Observation versus experiment
- Subjects, factors, treatments
- How to experiment badly
- Randomized comparative experiments
- The logic of randomized comparative experiments
- Cautions about experimentation
- Matched pairs and other block designs

Commentary: Data Ethics
- Institutional review boards
- Informed consent
- Confidentiality
- Clinical trials
- Behavioral and social science experiments

Chapter 10: Introducing Probability
- The idea of probability
- The search for randomness
- Probability models
- Probability rules
- Discrete probability models
- Continuous probability models
- Random variables
- Personal probability

Chapter 11: Sampling Distributions
- Parameters and statistics
- Statistical estimation and the law of large numbers
- Sampling distributions
- The sampling distribution of x
- The central limit theorem

Chapter 12: General Rules of Probability
- Independence and the multiplication rule
- The general addition rule
- Conditional probability
- The general multiplication rule
Independence again
Tree diagrams

Chapter 13: Binomial Distributions
The binomial setting and binomial distributions
Binomial distributions in statistical sampling
Binomial probabilities
Using technology
Binomial mean and standard deviation
The Normal approximation to binomial distributions

Chapter 14: Confidence Intervals: The Basics
The reasoning of statistical estimation
Margin of error and confidence level
Confidence intervals for a population mean
How confidence intervals behave

Chapter 15: Tests of Significance: The Basics
The reasoning of tests of significance
Stating hypotheses
P-value and statistical significance
Tests for a population mean
Significance from a table*

Chapter 16: Inference in Practice
Conditions for inference in practice
Cautions about confidence intervals
Cautions about significance tests
Planning studies: sample size for confidence intervals
Planning studies: the power of a statistical test

Chapter 17: From Exploration to Inference: Part II Review
Part II summary
Review exercises
Test yourself
Supplementary exercises

PART III: INFERENCE ABOUT VARIABLES
Chapter 18: Inference about a Population Mean
Conditions for inference about a mean
The t distributions
The one-sample t confidence interval
The one-sample t test
Using technology
Matched pairs t procedures
Robustness of t procedures

Chapter 19: Two-Sample Problems
Two-sample problems
Comparing two population means
Two-sample t procedures
Using technology
Robustness again
Details of the t approximation
Avoid the pooled two-sample t procedures
Avoid inference about standard deviations

Chapter 20: Inference about a Population Proportion
The sample proportion p
Large-sample confidence intervals for a proportion
Accurate confidence intervals for a proportion
Choosing the sample size
Significance tests for a proportion

Chapter 21: Comparing Two Proportions
Two-sample problems: proportions
The sampling distribution of a difference between proportions
Large-sample confidence intervals for comparing proportions
Using technology
Accurate confidence intervals for comparing proportions
Significance tests for comparing proportions

Chapter 22: Inference about Variables: Part III Review
Part III summary
Test yourself
Supplementary exercises

PART IV: INFERENCE ABOUT RELATIONSHIPS
Chapter 23: Two Categorical Variables: The Chi-Square Test
Two-way tables
The problem of multiple comparisons
Expected counts in two-way tables
The chi-square test statistic
Cell counts required for the chi-square test
Using technology
Uses of the chi-square test
The chi-square distributions
The chi-square test for goodness of fit

Chapter 24: Inference for Regression
Conditions for regression inference
Estimating the parameters
Using technology
Testing the hypothesis of no linear relationship
Testing lack of correlation
Confidence intervals for the regression slope
Inference about prediction
Checking the conditions for inference

Chapter 25: One-Way Analysis of Variance
Comparing Several Means
Comparing several means
The analysis of variance F test
Using technology
The idea of analysis of variance
Conditions for ANOVA
F distributions and degrees of freedom
Some details of ANOVA

PART V: OPTIONAL COMPANION CHAPTERS
(available on the BPS CD and online)
Chapter 26: Nonparametric Tests
Comparing two samples: the Wilcoxon rank sum test
The Normal approximation for W
Using technology
What hypotheses does Wilcoxon test?
Dealing with ties in rank tests
Matched pairs: the Wilcoxon signed rank test
The Normal approximation for W+
Dealing with ties in the signed rank test
Comparing several samples: the Kruskal-Wallis test
Hypotheses and conditions for the Kruskal-Wallis test
The Kruskal-Wallis test statistic

Chapter 27: Statistical Process Control
Processes
Describing processes
The idea of statistical process control
x charts for process monitoring
s charts for process monitoring
Using control charts
Setting up control charts
Comments on statistical control
Don't confuse control with capability!
Control charts for sample proportions
Control limits for p charts

Chapter 28: Multiple Regression
Parallel regression lines
Estimating parameters
Using technology
Inference for multiple regression
Interaction
The multiple linear regression model
The woes of regression coefficients
A case study for multiple regression
Inference for regression parameters
Checking the conditions for inference

Chapter 29: More about Analysis of Variance
Beyond one-way ANOVA
Followup analysis: Tukey pairwise multiple comparisons
Followup analysis: contrasts
Two-way ANOVA: conditions, main effects, and interaction
Inference for two-way ANOVA
Some details of two-way ANOVA