

Stat 431, Spring 2010

Statistical Inference

[Homework assignments](#) [Readings](#) [Schedule of topics](#)

People

		Contact	Office hours
Instructor	Mikhail Traskin	Send e-mail only in case of emergency. See instructor's page for email address. Primarily use webCafé discussion board (see below).	Tuesday/Thursday, 3:15 - 4:30 pm, 466 Jon M. Huntsman Hall
TA	James Piette	jpiette@wharton	Wednesday, 1 - 3 pm, 433 Jon M. Huntsman Hall

Additional help

The Stat Lab provides help with statistics to anybody on campus. It is located in F96 JMHH and is open on days when classes are in session. For staffing and hours, see <http://stat.wharton.upenn.edu/~sivana/statlab.html>. Whoever you find there may or may not be able to help with homework. Most useful for you are the Stat Lab hours staffed by our own TA: James Piette on Wednesday, 3 - 4 pm.

Lectures

Section 001: G55 JMHH, Tuesday/Thursday, 9:00 - 10:30 am.

Section 002: G55 JMHH, Tuesday/Thursday, 10:30 am - 12:00 noon.

Course homepage

Refer to <http://stat.wharton.upenn.edu/~mtraskin/courses/stat431/index.html> (this page) for syllabus.

Statistics 431 is using webCafé. You can gain access by going to

<http://webCafe.wharton.upenn.edu> and following the link to STAT and then to your section. All materials for this course will be distributed and managed via this website, and you will be able to monitor your grade entries throughout the semester.

An important feature of webCafé is the discussion board where everybody can place questions and comments. We will be using it extensively for answering your questions about homeworks, exams and scheduling. You are urged to go there first to see whether your question has already been asked and answered, and, if not, to place your question so it can be answered once for everybody.

Note for non-Wharton students: If you do not have a Wharton computing account, you will need to establish one to access the website. The account also provides access to the computing labs in Wharton and to the intranet. To get an account, on or after the first day of classes, go to <http://accounts.wharton.upenn.edu>. After you have obtained your account, allow up to 12 hours for activation. Wharton students and students who have recently taken a Wharton course have existing accounts.

Course description

This course is about making decisions under uncertainty using statistical methods. The topics include estimation, confidence intervals, hypothesis testing, single and multiple linear regression, one-way and two-way analysis of variance, variable selection, logistic regression and categorical data analysis.

Interpretation of the results and analysis of assumptions is an important part of the course. [Statistical computing package](#) will be extensively used to carry out the computations. However no special emphasis will be made on the details of computations.

Prerequisites

Familiarity with basic probability theory is assumed. Stat 430 or equivalent should provide sufficient background. Otherwise chapters 2, 3, 4 and 5 of the [Devore's book](#) will be a good substitute together with a [short review](#) of fundamental concepts used by the course, written by Prof. David Freedman, University of California, Berkeley.

Statistical computing package

We'll be using JMP version 8, available in the Wharton Computer Labs, F75/F80 Jon M. Huntsman Hall (Wharton account required: see <http://accounts.wharton.upenn.edu>). Individual copies are also available for

purchase at <http://upenn.onthehub.com>. A 3-year license costs \$59.95. You may also purchase six-month (\$29.95) and twelve-month (\$49.95) licenses at <http://estore.e-academy.com>. Please read system requirements carefully before making a purchase to make sure that it will work on your computer. In particular, *JMP 8 is not supported on Windows Vista Home Basic Edition*. When buying, make sure that you are downloading a correct (Windows/Mac) version: there were cases when people bought a wrong version.

You may also use R, an open-source statistical software that is available from [The R Project for Statistical Computing](http://www.r-project.org/).

Text book

J. L. Devore, *Probability and Statistics for Engineering and the Sciences*, 7th ed.

Assignments

Grading

- Homework assignments: 20% (lowest score excluded)
- Midterm: 30%
- Final: 50%

Homework assignments

- There will be about 10 homework assignments.
- Homeworks will be assigned on webCafé and will usually be due a week later. Problems involving computer calculations should be worked using JMP or R. **No extensions to the due date will be given.** However, the lowest homework assignment score will be omitted from the final grade calculation. Unsubmitted work counts as a zero score. Back up your work frequently on a data stick, so you can submit it from a Wharton computer if yours breaks down.
- Homeworks should be written by editing a copy of the MS Word file that contains problem statements. Each solution should be inserted after the respective problem statement following **YOUR SOLUTION:**. Handwriting is not accepted.
- Hand in your homework solutions **always in both of the following ways:**
 - Hand in a stapled paper copy of your solutions in the Statistics Department (JMHH, 4th floor, turn right exiting from the elevator) in the box marked with your section (not in class and not to the

TAs), AND

- Submit your MS Word file electronically via Assignment Submission on webCafé.

For re-grades and missing claims, there must be an electronic copy e-mailed before the deadline.

- Your solutions must have on the cover page your
 - name (as it appears on webCafé),
 - section,
 - school (Wharton/non-Wharton), and
 - year (freshman, sophomore, etc.).
- Homework is designed to teach, and you are encouraged to seek help from the instructor and the TAs if you have questions. You may also work with and help each other. *You must, however, submit your own solutions, with your own write-up and in your own words. Verbatim copying is against the honor code.*
- Graded homeworks will be returned in the same boxes in the Statistics Department.
- Scores for homeworks are finalized one week after the graded copies are handed back. Thereafter there will be no changes and no re-grading. Do not delay checking your graded homeworks to the end of the semester.

Exams

In-class midterm: March 18. No make-up.

Final: Tuesday, May 4, 6:00 – 8:00 pm, location TBA

Both exams are open notes (no textbooks). Calculators may be used but no laptops are allowed.

Readings

- Paul F. Velleman, "Truth, Damn Truth, and Statistics", *Journal of Statistics Education*, Volume 16, Number 2 (2008), <http://www.amstat.org/publications/jse/v16n2/velleman.html>

Schedule of topics

Lectures will closely follow the text which will be occasionally supplemented with handouts on topics going beyond those covered in the book. A review of basic probability theory and common distributions, e.g. chapters 2, 3, 4 and 5 in [Devore](#), might be useful. Topics discussed in sections 3.4 (binomial distribution), 4.3 (normal distribution) and 5.4 (central limit theorem) are of special interest.

Lec#	Date	Topic	Text
01	Thu 14 Jan	Introduction/overview	
02	Tue 19 Jan	Normality; boxplots; QQ plots (probability plots)	1.4, 4.6
03	Thu 21 Jan	Stem and Leaf diagrams; Scatter plots; CLT; Confidence intervals: known variance	1.3, 1.4, 4.3, 4.6 and 7.1
04	Tue 26 Jan	Confidence intervals: known variance; Large-sample CIs; Confidence intervals for population proportion; CIs: unknown variance; Non-normal population distribution	7.1 - 7.4
05	Thu 28 Jan	Confidence intervals for population proportion	7.1 - 7.4
06	Tue 2 Feb	One-sample hypothesis testing	8.1 - 8.5
07	Thu 4 Feb	One-sample hypothesis testing and intervals	8.1 - 8.5
08	Tue 9 Feb	Two-sample inference: testing and intervals	9.1 - 9.5
09	Thu 11 Feb	Two-sample inference: testing and intervals	9.1 - 9.5
10	Tue 16 Feb	Single factor (one-way) ANOVA	10.1 - 10.3
11	Thu 18 Feb	Single factor (one-way) ANOVA	10.1 - 10.3
12	Tue 23 Feb	Two factor (two-way) ANOVA	11.1 - 11.2
13	Thu 25 Feb	Two factor (two-way) ANOVA	11.1 - 11.2
14	Tue 2 Mar	Linear least-squares regression	13.4, 13.5

15	Thu 4 Mar	Simple linear least-squares regression; correlation coefficient	12.1 - 12.5
16	Tue 16 Mar	Review of topics covered so far	
17	Thu 18 Mar	Midterm	
18	Tue 23 Mar	Simple linear least-squares regression; correlation coefficient; regression and ANOVA	12.1 - 12.5
19	Thu 25 Mar	Multiple linear least-squares regression; parameter estimation; hypothesis test for a coefficient of linear regression model; model utility test; diagnostic plots	13.1 - 13.5
20	Tue 30 Mar	Multiple linear least-squares regression; CI for the predicted value; PI for future y value; F-test for a group of predictors	13.1 - 13.5
21	Thu 1 Apr	Multiple linear least-squares regression; F-test for a group of predictors; variable standardization; variable selection: stepwise regression	13.1 - 13.5
22	Tue 6 Apr	Multiple linear least-squares regression; variable selection: stepwise regression	13.1 - 13.5
23	Thu 8 Apr	Maximum Likelihood estimation; logistic regression	13.2
24	Tue 13 Apr	Logistic regression: drop-in-deviance test	
25	Thu 15 Apr	Goodness-of-fit tests	14.1 - 14.2
26	Tue 20 Apr	Categorical data analysis	14.3
27	Thu 22 Apr	Distribution-free procedures	15.1 - 15.2
28	Tue 27 Apr	Course summary	