

Regression Analysis

Department of Statistics 36-707 Fall 2006

LECTURES: Tuesday and Thursday 10:30 – 11:50
LOCATION: Porter Hall 125B
INSTRUCTOR: Larry Wasserman
OFFICE: Baker Hall 228a
EXTENSION: 8727
EMAIL: larry@stat.cmu.edu
WEB-SITE: www.stat.cmu.edu/~larry/=stat707
TEXT: *Applied Regression, 3rd Edition*, by Sanford Weisberg

OTHER REFERENCES:

1. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, by Trevor Hastie, Robert Tibshirani and Jerome Friedman.
2. *Modern Applied Statistics With S-plus*, by W.N. Venables and B.D. Ripley.
3. *All of Statistics: A Concise Course in Statistical Inference*, by Larry Wasserman.
4. *All of Nonparametric Statistics*, by Larry Wasserman.

WEBSITE

Check the website on a regular basis:

www.stat.cmu.edu/~larry/=stat707

The website has the lecture notes, homeworks, handouts, and announcements.

PREREQUISITES

I will assume you have had a course on statistical inference and that the following topics are familiar:

Expectation, variance, confidence intervals, bias, variance, likelihood function, maximum likelihood estimator, Normal, t , F and χ^2 distributions, basic matrix algebra.

If you want to take this course, you should feel comfortable answering the following questions:

1. Let $X_1, \dots, X_n \sim N(\mu, \sigma^2)$. Find the maximum likelihood estimate of σ^2 . Find the bias and variance of this estimate. Find a $1 - \alpha$ confidence interval for σ .
2. Let $X = (X_1, \dots, X_n)$ where each X_i is $N(0,1)$ and the variables are independent. Let A be an $n \times n$ matrix. Show that the covariance matrix of AX is AA^T .

TOPICS

1. **Linear regression:** simple regression, bias and variance, causation, multiple regression, weighted least-squares, diagnostics, robust regression, logistic regression, generalized linear models.
2. **Model selection:** prediction risk, bias-variance tradeoff, risk estimation, model search, lars and lasso, stepwise regression, boosting.
3. **Smoothing and nonparametric regression:** nonparametric regression, linear smoothers, kernels, local regression, fitting local regressions in R, penalized regression, regularization and splines, wavelets, variance estimation, testing the fit of a linear model, confidence bands, local likelihood and exponential families, multiple regression and additive models.

ASSESSMENT

There will be weekly assignments, two in-class tests and a final exam. The test dates are:

Test 1: Tuesday October 16
Test 2: Tuesday November 13
Final Exam: Set by the university

Don't buy plane tickets until the final exam schedule is out.

Homework	25%
Test 1	25%
Test 2	25%
Exam	25%

COMPUTING

The assignments will involve some computing. We will use R which is available free at <http://www.cran.r-project.org/>

OFFICE HOURS

I will not keep formal office hours. If you wish to meet with me, call or send email and make an appointment. The grader will keep office hours which will be posted on the website.