STA 371G Outline
Spring 2015

Instructor: Mingyuan Zhou, Ph.D., Assistant Professor of Statistics
    Office: CBA 6.462
    Phone: 512-232-6763
    Email: mingyuan.zhou@mccombs.utexas.edu
    Office Hours: Monday Wednesday 5:00-6:30 PM. You are welcome to come by my office at other times.

Wednesday, January 21

Topics:

- Introduction

- Probability

- Random variables

- Probability distributions

Monday, January 26

Topics:

- Mean, variance and standard deviation of a random variable

- Add a constant to a random variable

- Multiply a random variable by a constant

- Conditional, joint and marginal probabilities

Reading Assignments:
    If you are not familiar with the topics discussed in class, you are recommended to read:
    pp. 156-168, 189-195, of Data analysis and decision making, 4th edition
    or
    pp. 196-206, 225-231 of Data analysis and decision making, 3rd edition

    To learn more about these topics, you may further read:
    Chapters 2.1, 2.2, 2.4, and 2.5 of OpenIntro Statistics, 2nd edition

Wednesday, January 28

- Conditional, joint and marginal probabilities

- Independent random variables, sum of independent random variables

- Continuous random variables

- Probability density function: area under the curve represents probability

- Standard normal distribution $Z \sim \mathcal{N}(0, 1)$

Monday, February 2

- Standard normal calculations in Excel: NORMSDIST,
  or in R: pnorm (type "?pnorm" in R for help).

- Normal distribution $X \sim \mathcal{N}(\mu, \sigma^2)$

- Understand the meaning of the standard deviation σ in a normal distribution: $P(\mu - \sigma < X < \mu + \sigma) = ?$ and $P(\mu - 2\sigma < X < \mu + 2\sigma) = ?$

- Normal calculations in Excel:
  NORMSDIST, NORMDIST
  NORMSINV, NORMINV
  or in R:
  pnorm, qnorm (type "?pnorm" and "?qnorm" in R for help).

- If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $P(X < x) = P(\frac{X-u}{\sigma} < \frac{x-u}{\sigma}) = P(Z < \frac{x-u}{\sigma})$.

- Standardizing a normal random variable $Z = \frac{X-\mu}{\sigma} \sim \mathcal{N}(0, 1)$
  Interpretation: the value of Z is the number of standard deviations that X deviates towards the left (if $Z < 0$) or the right (if $Z > 0$) of the mean.

  Reading Assignments:

  To get familiar with the normal distribution, you are recommended to read:

  pp. 211-215, 217-225 of Data analysis and decision making, 4th edition

  or

  pp. 247-250, 253-260 of Data analysis and decision making, 3rd edition

  You may further read:

  Chapters 3.1.1, 3.1.2, 3.1.4 and 3.1.5 of OpenIntro Statistics, 2nd edition

Wednesday, February 4

- Plot a normal distribution in Excel and R

- Example: Testing at ZTel, we will make an Excel spreadsheet for calculations

- Case study, Texas BBA Salary Statistics

- Expectation of a random variable

- Binomial distribution and its normal approximation

Reading Assignments:
Lecture notes 3 and 4 posted on the course website.

To learn more about the binomial distribution, its normal approximation, and the sampling distribution of a sample proportion, please read:
pp. 233-239, 403-404 of Data analysis and decision making, 4th edition
or
pp. 268-273, 438-439 of Data analysis and decision making, 3rd edition

For this topic, you may further read:
Chapters 3.4.1, 3.4.2 and 6.1 of OpenIntro Statistics, 2nd edition

Monday, February 9

- Case study: Texas BBA Demographics

- Binomial distribution $X \sim \text{Binomial}(n, p)$. Examples: the number of "Heads" in 100 coin flips, the number of votes for Republican in 1000 voters

- The normal approximation to the binomial $X \sim \mathcal{N}(np, np(1-p))$

- Important concepts: Population and Sample

- Sampling distribution of a sample proportion

- Case study: A national poll of 803 adults by Anzalone Liszt Grove Research

- Population mean, variance, standard deviation

- Sample mean, sample variance, standard error of the sample mean

- Sampling distribution of the sample mean

Reading Assignments:
To learn more about estimation and sampling distribution, please read:
pp. 352-353, 366-371, 374, 388-395 of Data analysis and decision making, 4th edition
or
pp. 378-379, 393-398, 400-401, 422-430 of Data analysis and decision making, 3rd edition

For this topic, you may further read:
Chapters 4.1, 4.2, 4.4 and 5.3 of OpenIntro Statistics, 2nd edition

Wednesday, February 11

- Sampling distribution of the sample mean

- Central limit theorem

- t distribution (optional)

- Confidence interval

- Introduction to linear regression

Monday, February 16

- Simple linear regression

- Linear prediction: $Y = b_0 + b_1 X$

- Least squares estimation of b_0 and b_1

- Examples: predict house price, baseball runs per game

- Using Excel and R to do the calculation

- Excel add-in: if you are using Mac, please install StatPlus:mac LE (available at http://www.analystsoft.com/en/products/statplusmacle/); if you are using windows, please install Analysis ToolPak or Decision Tools Standard 6.1

  (available at http://www.mccombs.utexas.edu/tech/computer-services/coe#DecisionTools).

- Sample mean, variance, and standard deviation

- Sample covariance, sample correlation

- Linear relationship between X and Y

- $b_0 = \bar{y} - b_1 \bar{x}$, $b_1 = r_{xy} \times \frac{s_y}{s_x}$

Reading Assignments:
Chapters 7.1 and 7.2 of OpenIntro Statistics, 2nd edition
pp. 531-551 of Data analysis and decision making, 4th edition
or
pp. 562-584 of Data analysis and decision making, 3rd edition

Wednesday, February 18

- mean(e)=0, Corr(e, X)=0, Corr($e, \hat{Y}$)=0, Corr($\hat{Y}, X$)=1

- SST, SSR, SSE

- Coefficient of determination: $R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$

- $R^2 = r_{xy}^2$ measures the proportion of variation in Y explained by X.

- Statistical model for simple linear regression

- Statistical model for simple linear regression:
  $Y = \beta_0 + \beta_1 X + \epsilon, \ \epsilon \sim \mathcal{N}(0, \sigma^2)$
  $Y \sim \mathcal{N}(\beta_0 + \beta_1 X, \sigma^2)$

- Conditional distribution of Y given X

- 95% prediction interval of Y given X: $\beta_0 + \beta_1 X \pm 2\sigma$

- Conditional and marginal distributions of Y

- Interpretation of ϵ and σ

- The error terms ϵ_i are independently and identically distributed

- Least squares estimation and Gaussian maximum likelihood (optional)

- True line $\beta_0 + \beta_1 X$ and least squares line $b_0 + b_1 X$

- Degrees of freedom

- In SLR, σ^2 is estimated with $s^2 = \frac{\sum_{i=1}^{n} e_i^2}{n-2} = \frac{SSE}{n-2}$.

- SLR regression standard error: $s = \sqrt{SSE/(n-2)}$

Reading Assignments:
PDF "Simple Linear Regression" posted in Canvas/files

Monday, February 23

- Sampling distributions of regression parameters

- Confidence intervals of regression parameters

- Case study: Waite First Securities, Milk and Money

Wednesday, February 25

- Topic summary for Midterm #1

- Discuss Practice Exam #1

- Common problems in homework assignments

- Hypothesis testing in SLR: t-statistic and p-value

Monday, March 2

- Midterm Exam #1

Wednesday, March 4

- Forecasting with linear regression models

- Multiple regression

- Example: Auto MPG data

Monday, March 9

- Multiple regression

- T-test and F-test

- Example: Supervisor performance data

- Understanding multiple regression

- Correlation and causation

- Example: Number of beer and weight & height

- Examples: Auto MPG, Baseball

Wednesday, March 11

- Multicollinearity

- Dummy variables and interactions

- Example: Gender Discrimination in Salary at Fifth National Bank

- Example: MidCity House Price

Monday, March 23

- Case study: Orion Bus Industries—Contract Bidding Strategy

- Diagnostics

- Polynomial regression

Wednesday, March 25

- Variable interaction

- Log transformation

- Outliers

- Discuss Homework Assignment 6

- Case Study, Oakland A's (A)

- Case Study, Oakland A's (B)

Reading Assignments:
Chapters 10, 11.1-11.5, and 11.8-11.11 of Data analysis and decision making, 4th edition
or
Chapters 11, 12.1-12.5, and 12.8-12.11 of Data analysis and decision making, 3rd edition

Monday, March 30

- Time series: fitting a trend

- Autocorrelation

- Time series regression, Hotel Occupancy Case

- Random walk models

- Autoregressive models

- Example: Monthly stock closing prices

- Example: Daily/Monthly temperature

- Example: Monthly Boston Armed Robberies Jan.1966-Oct.1975

Wednesday, April 1

- Seasonal models

- Example: Fisher river daily temperatures

- Example: Monthly airline passengers

- Example: Monthly liquor sales

- Case study: Northern Napa Valley Winery, Inc.

Reading Assignments:
Chapter 12 of Data analysis and decision making, 4th edition
or
Chapter 13 of Data analysis and decision making, 3rd edition

Monday, April 6

- Moving averages, exponential smoothing and ARMA

- Hypothesis testing: Type I Error, Type II Error, significant level, and power

- Understanding prediction errors in linear regression

- Model selection

Wednesday, April 8

- Review for Midterm Exam #2

- Model selection

Monday, April 13

- Midterm Exam #2