STA 371G Outline
Spring 2016

Instructor: Mingyuan Zhou, Ph.D., Assistant Professor of Statistics
Office: CBA 6.462
Phone: 512-232-6763
Email: `mingyuan.zhou@mccombs.utexas.edu`
Website: http://mingyuanzhou.github.io/
Office Hours: Tuesday & Thursday 5:00-6:30 PM. You are welcome to come by my office at other times.

Tuesday, January 19

Topics:

- Introduction
- Probability
- Random variables

Reading Assignments:
You are recommended to read:
Chapter 1 of OpenIntro Statistics, 3rd edition

Thursday, January 21

Topics:

- Probability distributions
- Mean, variance and standard deviation of a random variable
- Add a constant to a random variable
- Multiply a random variable by a constant

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, 3rd edition

Tuesday, January 26

- Conditional, joint and marginal probabilities
- Independent random variables, sum of independent random variables

Thursday, January 28

- Continuous random variables
- Probability density function: area under the curve represents probability
- Standard normal distribution $Z \sim \mathcal{N}(0, 1)$
- Normal distribution $X \sim \mathcal{N}(\mu, \sigma^2)$

  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, 3rd edition

Tuesday, February 2

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

Thursday, February 4

- Plot a normal distribution in Excel and R
- Normal calculations in Excel:
  NORMSDIST, NORMDIST
  NORMSINV, NORMINV
  or in R:
  pnorm, qnorm (type "?pnorm" and "?qnorm" in R for help).
- Example: Testing at ZTel, we will make an Excel spreadsheet for calculations
- Case study, Texas BBA Salary Statistics
- Expectation of a continuous 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, 3rd edition

Tuesday, February 9

- Binomial distribution and its normal approximation
- 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

Thursday, February 11

- Population mean, variance, standard deviation
- Sample mean, sample variance, standard error of the sample mean
- Sampling distribution of the sample mean
- Sampling distribution of the sample mean
- Central limit theorem
- t distribution (optional)
- Confidence interval
- Introduction to linear regression

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, 3rd edition

Tuesday, 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: Palisade Decision Tools (including StatTools) for Windows, StatPlus:mac LE for Mac.
- 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, 3rd edition
pp. 531-551 of Data analysis and decision making, 4th edition
or
pp. 562-584 of Data analysis and decision making, 3rd edition

Thursday, 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 independent, 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