STA 371G Outline
Fall 2019

Instructor: Mingyuan Zhou, Ph.D., Associate Professor of Statistics
Office: CBA 6.458
Phone: 512-232-6763
Email: mingyuan.zhou@mccombs.utexas.edu
Website: http://mingyuanzhou.github.io/
Office Hours: Monday & Wednesday 3:30-4:30 PM. You are welcome to come by my office at other times.

Wednesday, August 28

Topics:

- Introduction
- Probability
- Random variables
- Probability distributions

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

Wednesday, September 4

Topics:

- Mean, variance and standard deviation of a random variable
- Add a constant to a random variable
- Multiply a random variable by a constant
- Independent random variables, sum of independent random variables

Reading Assignments:
If you are not familiar with the topics discussed in class, you are recommended to read:
pp. 140-142, 740-741 Business Analytics: Business Analytics: Data analysis and decision making, 6th edition
or
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

You are also recommended to read:
pp. 1-14 of "1TopicSummary_ProbabilityConceptsAndNormalDistributions.pdf" (available in Canvas/files)

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

Monday, September 9

- Measure uncertainty with probability
- Frequency probability and subjective probability
- Probability, lotteries and betting odds
- Payoff tables
- Payoffs and Losses
- Nonprobabilistic criteria for decision making: maximin, maximax, and minimax loss

Wednesday, September 11

- Probabilistic criteria for decision making: expected payoff, expected loss
- Utility functions
- Conditional probability and conditional bets or conditional reference contracts
- Conditional, joint and marginal probabilities
- Bayes' theorem

Reading Assignments:
Chapter 6 of Business Analytics: Data analysis and decision making, 6th edition
or
Chapter 6 of Data analysis and decision making, 4th edition
or
Chapter 7 of Data analysis and decision making, 3rd edition

Monday, September 16

- Decision trees, risk profile, sensitivity analysis
- Risk profile, sensitivity analysis

Reading Assignments:

Chapter 6 of Business Analytics: Data analysis and decision making, 6th edition
or
Chapter 6 of Data analysis and decision making, 4th edition
or
Chapter 7 of Data analysis and decision making, 3rd edition

Wednesday, September 18

- The value of information
- Expected value of perfect information (EVPI)
- Expected value of sample information (EVSI)
- Expected value of sample information (EVSI)
- Case study: Freemark Abbey Winery

Monday, September 23

- 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. 167-171, 174-182 of Business Analytics: Data analysis and decision making, 6th edition
or
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 are also recommended to read:
pp. 15-30 of "1TopicSummary_ProbabilityConceptsAndNormalDistributions.pdf" (available in Canvas/files)

You may further read:
Chapters 3.1.1, 3.1.2, 3.1.4 and 3.1.5 of OpenIntro Statistics, 3rd edition
To learn the binomial distribution and its normal approximation, please read:
pp. 233-239 of Data analysis and decision making, 4th edition
or

pp. 268-273 of Data analysis and decision making, 3rd edition

You may further read:
Chapters 3.4.1 and 3.4.2 of OpenIntro Statistics, 2nd edition

Wednesday, September 25

- If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $P(X < x) = P(\frac{X-u}{\sigma} < \frac{x-u}{\sigma}) = P(Z < \frac{x-u}{\sigma})$.
- 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).
- 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.
- Plot a normal distribution in Excel and R

Monday, September 30

- Example: Testing at ZTel, we will make an Excel spreadsheet for calculations
- Case study, Texas BBA Salary Statistics
- Expectation of a continuous random variable
- Population mean, variance, standard deviation
- Sample mean, sample variance, standard error of the sample mean
- Sampling distribution of the sample mean
- Binomial distribution and its normal approximation
- 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 $X \sim \mathcal{N}\left(np, np(1-p)\right)$
- Important concepts: Population and Sample
- Sampling distribution of a sample proportion

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

Reading Assignments:

To learn more about estimation and sampling distribution, please read:

pp. 280-281, 292-297, 299, 312-318 of Business Analytics: Data analysis and decision making, 6th edition

or

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

You are also recommended to read:

"2TopicSummary_EstimationAndSamplingDistributions.pdf" (available in Canvas/files)

For this topic, you may further read:

Chapters 4.1, 4.2, 4.4 and 5.3 of OpenIntro Statistics, 3rd edition

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

Wednesday, October 2

- Please install R and Rstudio on your laptop and bring it to class
- Simulation using Excel and R
- Simulate random numbers from a discrete distribution
- Find the sample mean and variance, compare them with the true mean and variance
- Simulate the sampling distribution of the sample mean
- Uniform random numbers, flip a coin, toss a die, flip two coins, toss two dice, law of large numbers
- Estimate π with Monte Carlo simulation
- Simulate normal random numbers $X \sim \mathcal{N}(\mu, \sigma^2)$.
- Find $P(X < x)$ and $P(X <?) = p$ using simulation
- Demonstrate Central Limit Theorem using simulation
- Simulation of weekly demand

Reading Assignments:

Chapters 4.4, pp 740-741 of Business Analytics: Data analysis and decision making, 6th edition

or

Chapters 4.4–4.8 of Data analysis and decision making, 4th edition

or

Chapters 5.4–5.8 of Data analysis and decision making, 3rd edition

Monday, October 7

- Sampling distribution of the sample mean
- Confidence interval
- Simple linear regression
- Linear prediction: $Y = b_0 + b_1 X$

Reading Assignments:

Chapters 7.1 and 7.2 of OpenIntro Statistics, 3rd edition

pp. 418-441 of Business Analytics: Data analysis and decision making, 6th edition

or

pp. 531-551 of Data analysis and decision making, 4th edition

or

pp. 562-584 of Data analysis and decision making, 3rd edition

Wednesday, October 9

- 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

Reading Assignments:

PDF "Simple Linear Regression" posted in Canvas/files

Monday, October 14

- Linear relationship between X and Y
- $b_0 = \bar{y} - b_1\bar{x}$, $b_1 = r_{xy} \times \frac{s_y}{s_x}$
- 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: $\beta_0 + \beta_1 X \pm 2\sigma$

Reading Assignments:

"3TopicSummary_RegressionModelAndEstimation.pdf" (available in Canvas/files)
"5TopicSummary_CorrelationAndCovariance.pdf" (available in Canvas/files)
"6TopicSummary_ComputingAndInterpretingRSquare.pdf" (available in Canvas/files)
"7TopicSummary_InterpretingAndEstimatingVarianceOfEpsilon.pdf" (available in Canvas/files)

Wednesday, October 16

- 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 $b_0 + b_1 X$
- Case study: A stock's beta coefficient

Monday, October 21

- Midterm Exam, 6:45-9:45 pm

Wednesday, October 23

- Case study: Milk and Money
- 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: $s = \sqrt{SSE/(n-2)}$
- Sampling distributions of regression parameters
- Confidence intervals of regression parameters

Monday, October 28

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

Wednesday, October 30

- Multiple regression
- T-test
- Examples: Auto MPG, Baseball

Monday, November 4

- Example: Supervisor performance data
- F-test
- Understanding multiple regression
- Correlation and causation

Wednesday, November 6

- Multicollinearity
- Example: Number of beer and weight & height

Monday, November 11

- Dummy variables and interactions
- Example: Gender Discrimination in Salary at Fifth National Bank
- Case study: Orion Bus Industries—Contract Bidding Strategy

Reading Assignments:

Chapters 10, 11.1-11.5, and 11.7-11.10 of Business Analytics: Data analysis and decision making, 6th edition

or

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

"4TopicSummary_NonlinearRelationships.pdf " (available in Canvas/files)

Wednesday, November 13

- Example: MidCity House Price
- Diagnostics
- Polynomial regression

Reading Assignments:

"9TopicSummary_MeasuringTheQualityOfTheEstimateOfBeta.pdf" (available in Canvas/files)

"10TopicSummary_HypothesisTestingInRegression.pdf" (available in Canvas/files)

Monday, November 18

- Slides 3.3 pages 25-46
- Variable interaction
- Log transformation
- Case Study, Oakland A's (A)
- Case Study, Oakland A's (B)
- Slides 4.1, Slides 4.2 pages 1-15
- Time series: fitting a trend
- Autocorrelation
- Time series regression, Hotel Occupancy Case

Reading Assignments:

Chapter 12 of Business Analytics: Data analysis and decision making, 6th edition

or

Chapter 12 of Data analysis and decision making, 4th edition

or

Chapter 13 of Data analysis and decision making, 3rd edition

Wednesday, November 20

- Slides 4.2 pages 16-end, Slides 4.3 pages 1-15
- Random walk models
- Autoregressive models
- Example: Monthly stock closing prices
- Example: Daily/Monthly temperature
- Example: Monthly Boston Armed Robberies Jan.1966-Oct.1975
- Seasonal models
- Example: Fisher river daily temperatures

Reading Assignments:
"8TopicSummary_ForecastingModelForSPSSSales.pdf" (available in Canvas/files)

Monday, November 25

- Slides 4.3 pages 16-end
- Example: Monthly airline passengers
- Case study: Northern Napa Valley Winery, Inc.
- Example: Monthly liquor sales

Wednesday, November 27

- Thanksgiving holidays

Monday, December 2

- Slides 5, Slides 6.1 pages 1-3
- Outliers
- 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, December 4

- Simulation and decision
- Multivariate distributions, covariance and correlation
- Sum of correlated random variables
- Simulate portfolio return
- Sample from a finite population (with/without replacement)
- Simulate binomial random variables
- Simulate student t random variables
- Simulate a random walk model
- Simulate an AR+Trend model
- Simulate prediction intervals for an AR model

Monday, December 9 (Last class of the semester)

- Practice questions for the final exam