**01:960:463:01 Index 29880 – Regression Methods (Spring/2020)**

**Time: Tuesday 7, 8 (6:40-9:30 pm)**

**WebEx will be started at 6:20 pm**

**Instructor:** Jack Mardekian, PhD

Recently retired from Pfizer Inc

**Contact information:** 908.304.8400 (cell)

848.445.7647 (office) (my cell number is a better option, I am sheltering at home)

Rutgers e-mail: [mardekia@stat.rutgers.edu](mailto:mardekia@stat.rutgers.edu)

**Virtual Office Hours:** By WebEx : Monday 1-2 pm, Tuesday 2-3 pm

Adhoc WebEx is available, I will respond to e-mails quickly

**Course Description in Catalog**:

Prerequisite: Level II Statistics.

Multiple and nonlinear correlation and regression techniques for analysis of events in time and space: analysis of variance and covariance (ANOVA), related multivariate techniques, response surface approaches.

**Course objective**: Students will gain an understanding of regression methods including simple linear regression, least squares estimation, multiple linear regression, checking model adequacy, diagnostics for leverage and influence, variable selection and model building, multicollinearity, weighted least squares, and logistic regression. Students will be able to perform analyses using regression methods. Students will be able to execute regression procedures in R software to analyze data. We will review SAS output but programming in SAS is out of scope for the course.

**Prerequisites:** Basic working knowledge of statistical methods typically covered in an introductory statistics course. Previously covered concepts will be reviewed as they are needed. Pre-calculus level mathematical skills including an introduction to matrices. Basic computer skills and programming basics are assumed.

**Textbook:** *A Second Course in Statistics: Regression Analysis*, William Mendenhall and Terry Sincich, 8th edition, Pearson Education, 2020

**Textbook site:** [**https://www.pearson.com/us/higher-education/program/Mendenhall-A-Second-Course-in-Statistics-Regression-Analysis-8th-Edition/PGM2492408.html**](https://www.pearson.com/us/higher-education/program/Mendenhall-A-Second-Course-in-Statistics-Regression-Analysis-8th-Edition/PGM2492408.html)

**Canvas class site:** <https://rutgers.instructure.com/courses/39544>

**Class notes, assignments, announcements will be posted on this site.**

**Rutgers Honor Pledge is in effect at all times:** *“On my honor, I have neither received nor given any unauthorized assistance on this examination.”*

**Grading (weights are tentative):**

In-class Midterm Exam (40%) completed on March 10

Final Exam will be based on material since Midterm Exam and

is tentatively scheduled as take-home (40%)

HW assignments (20%) (submitted electronically using Canvas). Please pay attention to due dates, only HW submitted by the “available until date” will be graded.

Points for attendance added to Midterm only. No quizzes planned during remainder of course.

**Considerations resulting from the cross-over to Online Teaching:**

1. All meeting and office hour times including assignment due times are scheduled using Eastern Standard Times. Please adjust these times to your geographic location.

2. All lectures will be recorded in MP4 format (audio + desktop) and posted in Canvas. All material continues to be posted in Canvas.

3. You can collaborate (following the guidelines for social distance) on HW assignments.

4. Rutgers Honor Pledge will be in effect for the tentative take-home Final Exam.

**Relevant dates:**

|  |  |
| --- | --- |
| Regular Classes End | Monday, May 4 (Our last class is April 28) |
| Final Exam Period | Thursday, May 7 - Wednesday, May 13 |

**Timely completion of HW assignments in Canvas is an important factor for passing this course.**

Message from Susan Lawrence, Vice Dean for Undergraduate Education:

* If your faculty or your students experience any technology problems as you begin the roll out of remote instruction, please refer them to the Teaching and Learning with Technology [Help Desk](https://tlt.rutgers.edu/tlt-help-desk-247365) web page.
* Students with concerns about access to appropriate technology should referred to this [Technology Resources for Students](https://coronavirus.rutgers.edu/technology-resources-for-students/) page.
* Students without access to computers can be referred to Dean of Students [deanofstudents@echo.rutgers.edu](mailto:deanofstudents@echo.rutgers.edu).
* The University source for information is [Rutgers University Teaching and Learning with Technology: Emergency Preparedness](https://tlt.rutgers.edu/node/489)
* We will continue to update and post answers to FAQs at [Keep Teaching](https://sasoue.rutgers.edu/teaching-learning/keep-teaching).

**Rutgers Office of Information Technology website: https://oit-nb.rutgers.edu/**

**The R Project for Statistical Computing:** [**http://www.r-project.org/**](http://www.r-project.org/)

**For R material on the Web – See** [**www.rseek.org**](http://www.rseek.org)

**also Google “An Introduction to R”**

**All web sites below were successfully accessed on January 16, 2020**

**Resources to help you learn and use R**

[**http://www.ats.ucla.edu/stat/r/**](http://www.ats.ucla.edu/stat/r/)

**Useful Statistics web sites:**

[**http://www.claviusweb.net/statistics.shtml**](http://www.claviusweb.net/statistics.shtml)

[**http://stattrek.com/**](http://stattrek.com/)

|  |  |
| --- | --- |
| **Topics from textbook covered, to be covered (there are 14 class meetings + Final Exam night):**  **March 24 is the 9th class meeting. Remaining classes: 3/31, 4/7, 4/14, 4/21, 4/28** | |
| **3.  Simple Linear Regression** | **6.  Variable Screening Methods** |
| 3.1 Introduction | 6.1 Introduction: Why Use a Variable-Screening Method? |
| 3.2 The Straight-Line Probabilistic Model | 6.2 Stepwise Regression |
| 3.3 Fitting the Model:  The Method of Least Squares | 6.3 All-Possible-Regressions-Selection Procedure |
| 3.4 Model Assumptions | 6.4 Caveats |
| 3.5 An Estimator of σ2 |  |
| 3.6 Assessing the Utility of the Model:  Making Inferences About the Slope β1 | **7.  Some Regression Pitfalls** |
| 3.7 The Coefficient of Correlation | 7.3 Parameter Estimability and Interpretation  Standardized coefficients |
| 3.8 The Coefficient of Determination | 7.4 Multicollinearity |
| 3.9 Using the Model for Estimation and Prediction |  |
| 3.10 A Complete Example | **8.  Residual Analysis** |
|  | 8.1 Introduction |
| **4.  Multiple Regression Models** | 8.2 Plotting Residuals and Detecting Lack of Fit |
| We will also cover Appendix B: The Mechanics of a Multiple Regression Analysis | 8.3 Detecting Unequal Variances |
| 4.1 General Form of a Multiple Regression Model | 8.4 Checking the Normality Assumption |
| 4.2 Model Assumptions | 8.5 Detecting Outliers and Identifying Influential Observations |
| 4.3 A First-Order Model with Quantitative Predictors | 8.6 Detection of Residual Correlation:  The Durbin-Watson Test |
| 4.4 Fitting the Model:  The Method of Least Squares |  |
| 4.5 Estimation of σ2, the Variance of ε | **9. Special Topics in Regression** |
| 4.6 Testing Overall Model Utility:  The Analysis of Variance F-Test | 9.1 Introduction |
| 4.7 Inferences About the Individual β Parameters | 9.4 Weighted Least Squares |
| 4.8 Multiple Coefficients of Determination: R2 and R2adj | 9.6 Logistic Regression |
| 4.9 Using the Model for Estimation and Prediction | 9.7 Poisson Regression |
| 4.10 An Interaction Model with Quantitative Predictors | 9.8 Ridge and Lasso Regression |
| 4.11 A Quadratic (Second-Order) Model with a Quantitative Predictor | 9.9 Robust Regression |
| 4.14 A Complete Example |  |
|  | **10.  Introduction to Time Series Modeling and Forecasting** |
| **5.  Principles of Model-Building** | 10.1 What Is a Time Series? |
| 5.1 Introduction: Why Model Building is Important | 10.2 Time Series Components |
| 5.2 The Two Types of Independent Variables: Quantitative and Qualitative | 10.3 Forecasting Using Smoothing Techniques (Optional) |
| 5.3 Models with a Single Quantitative Independent Variable | 10.4 Forecasting: The Regression Approach |
| 5.4 First-Order Models with Two or More Quantitative Independent Variables | 10.5 Autocorrelation and Autoregressive Error |
| 5.5 Second-Order Models with Two Quantitative Independent Variables | 10.7 Constructing Time Series Models |
|  | 10.8 Fitting Time Series Models with Autoregressive Errors |
|  | 10.9 Forecasting with Time Series Autoregressive Models |

**The Comprehensive R Archive Network (CRAN)**

**What are R and CRAN?**

R is ‘GNU S’, a freely available language and environment for statistical computing and graphics which provides a wide variety of statistical and graphical techniques: linear and nonlinear modelling, statistical tests, time series analysis, classification, clustering, etc. Please consult the [R project homepage](https://www.r-project.org/) for further information.

CRAN is a network of ftp and web servers around the world that store identical, up-to-date, versions of code and documentation for R. Please use the CRAN [mirror](https://cran.r-project.org/mirrors.html) nearest to you to minimize network load.

|  |  |  |  |
| --- | --- | --- | --- |
| |  | | --- | | Download and Install R Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:   * [Download R for Linux](https://cran.r-project.org/bin/linux/) * [Download R for (Mac) OS X](https://cran.r-project.org/bin/macosx/) * [Download R for Windows](https://cran.r-project.org/bin/windows/)   R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above. | | Source Code for all Platforms Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!   * The latest release (2019-12-12, Dark and Stormy Night) [R-3.6.2.tar.gz](https://cran.r-project.org/src/base/R-3/R-3.6.2.tar.gz), read [what's new](https://cran.r-project.org/doc/manuals/r-release/NEWS.html) in the latest version. * Sources of [R alpha and beta releases](https://cran.r-project.org/src/base-prerelease/) (daily snapshots, created only in time periods before a planned release). * Daily snapshots of current patched and development versions are [available here](https://stat.ethz.ch/R/daily). Please read about [new features and bug fixes](https://cran.r-project.org/doc/manuals/r-devel/NEWS.html) before filing corresponding feature requests or bug reports. * Source code of older versions of R is [available here](https://cran.r-project.org/src/base/). * Contributed extension [packages](https://cran.r-project.org/web/packages/index.html) | | Questions About R  * If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](https://cran.r-project.org/faqs.html) before you send an email. | |

**NOTE:** Rutgers Computer Labs software includes R.3.51 (MAC) and 3.5.2 (PC) and Rstudio 1.1.463