**USC School of Cinematic Arts, Fall 2015**

Big Data and Statistics for Game Design: Understanding Gamers for Fun and Profit

**Instructor:** Dennis Wixon,

**Course number:** GESM 160g 35486D

**Course title: Statistical Analysis for Games: understanding the principles and tools of classical statistics and data science for fun and profit**

**Student Assistant:** TBD,

**Unit Value**: 4 Units

**Class Locations:** SCA B148, SCI 207

**Class meetings:** Monday, Friday 11:00 – 12:20

(Note the class located in the basement of SCA; take the elevator and turn to the right, the room as at the end of the hall)

**Instructor Contact Information:**

**Email**: dwixon@cinema.usc.edu

**Phone/Text 425-210-6831**

**Office hours**: Monday, Friday 1:00 – 2:00 and by appointment

**Location:** SCI 201K

**Course Description:**

The course is divided into two complementary phases.

1. In the first phase the concepts and tools of statistics will be presented and explained. Exercises with SPSS and Excel will embody the concepts and provide practice in data analysis and interpretation. Mastery will be judged by a series of assignments and quizzes and exercises.
2. The second phase contains all the stages of doing a data science project. Specifically it includes developing a proposal and presentation, review of same, collecting and scrubbing data, performing analysis and writing up conclusions.

|  |  |
| --- | --- |
| **Learning Objectives:**  Upon completion, students should be able to:   1. Understand explain and apply statistical concepts in proposing and managing a data science project 2. . Complete a project by:    1. Finding and downloading the appropriate data set,    2. Planning an analysis for the set    3. Proposing the project a review committee    4. Conducting and reviewing the analysis    5. Presenting the results to a review committee   Note: the course does NOT rely on mathematical proofs or derivations, no computations are done by hand, instead use of tools SPSS and Excel is taught with examples |  |

**Required Readings and Supplementary Materials**

**Required Readings**

* Hays: *Statistics* – selected pages see below, will be on blackboard
* Williams and Monge *Reasoning with Statistics* –selected chapters see below, will be on blackboard
* Provost and Fawcett *Data Science for Business* –selected chapters see below, available at the bookstore

Specific readings from these books are listed in table 3 below

**Required software**

* Excel
* SPSS (see note below)

**Prerequisite:**

It’s recommended that students have completed an introductory course in any of the following areas: experimental design, statistics, or research methods. If you are uncertain if your experience is sufficient, please contact the instructor.

**Submitting assignments:**

Slides, assignments, take-home quizzes, and supplementary materials will be posted on Blackboard.

Assignments will be collected on Blackboard. Most assignments should be uploaded by Wednesday Noon unless otherwise noted

**Evaluation Breakdown**

|  |  |
| --- | --- |
| **Deliverable** | **Points** |
| Short assignments   * Quizzes * Homework | 40 Total   * 8 assignments * 5 points each |
| Presentation on a Research Topic | 10 Total   * 5 (class) * 5 (instructor) |
| Research Proposal   * Presentation * Write up | 15 Total   * 10 presentation   + 7 (instructor)   + 3 (class) * 5 write up |
| Final research project   * Draft write up * Presentation * Write up | 25 Total   * 10 presentation   + 7 (instructor)   + 3 (class) * 15 write up |
| Class contribution   * Instructor rating | 10 total |
| **Total** | **100 points** |

**Due dates**

All deliverables should be submitted on the due date specified (see Course Content).   
***Late assignments will receive a 50% reduction in points for each week late.***   
After 2 weeks (or end of term) late assignments cannot be evaluated

Exceptions to the due date or substitute assignments should be discussed with the instructor and confirmed via email, i.e. you must receive an email from the instructor confirming any extension or alternate arrangement.

**Resources for Assignments**

Many resources are provided for assignments and your research projects. Most of these are data bases or collections of data bases that you can use for projects.

Many are provided in Blackboard. Some are available on the Web.

Most of these data bases are available in a format (e.g. CSV) that can be downloaded into Excel. Only one requires programming skills

**Data Sources**

|  |  |  |  |
| --- | --- | --- | --- |
| **Data Name** | **Description** | **Location** | **Internet source** |
| Survey | data from class | Blackboard | none |
| Mushroom |  | Web | http://archive.ics.uci.edu/ml/datasets/mushroom |
| Iris 1 | Subset of original data | Web | http://archive.ics.uci.edu/ml/datasets/iris |
| Iris 2 | Subset of original data | Web | http://archive.ics.uci.edu/ml/datasets/iris |
| Breast Cancer | 569 cases | Blackboard | http://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29 |
| Titanic | 891 | Blackboard | none |
| Various | A number of data bases on interesting topics | Web | <https://github.com/fivethirtyeight/data> |
| Valve (games data base) | A number of data based | Web | [https://developer.valvesoftware.com/wiki/Steam\_Web\_API](https://cinemail.usc.edu/owa/UrlBlockedError.aspx)  [http://valvesoftware.com/](https://cinemail.usc.edu/owa/redir.aspx?C=fz_cl4jaoU2ga4SUxhsLnT-yLwkUkNFIyolY5XUHifmyo-g4uLuJ1bwYrP-hb7hSzndx-Oy9i0k.&URL=http%3a%2f%2fvalvesoftware.com%2f) |

**Course content: dates, topics, deliverables**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Wednesday Friday**  **Dates** | **Topics/Daily Activities** | **Readings** will be available on blackboard unless otherwise indicated | **Deliverables  Due Dates \_** | **Points** |
| **Week 1** | | | | |
| 24-Aug    26-Aug | * Introduction: course expectations, * Syllabus review * Course goals * Introduction to computer lab * Intro to SPSS (Excel) * Chi Square * Hypothesis testing * Mushroom data | * Data Science for Business (DSB) chapter 1 * How do people really use text editors? Whiteside * NY times: Age of Big Data * Box plot explained   http://www.cbsolution.net/ techniques/on target/ box\_plots\_clearly\_explained.html     * DSB chapter 2 | 1. Complete Survey and upload results to blackboard   **Due Thurs 25-Aug (noon)**   1. Pick a variable from the survey use SPSS to analyze it , compute descriptive statistics and do a boxplot Upload results to blackboard   **Due: Monday 29 –Aug (EOD)**  **Data: Survey** . | None  5 |
| **Week 2** | | | | |
| 31-Aug    2-Sept | * In Class review of box plot. * Decision trees      * In class review of CHI square * Logic of CHAID | * DSB Chapter 3 | 1. Pick a variable other than the one we did and compute a CHI square using SPSS and EXCEL write a paragraph about your conclusions upload all files to blackboard   **Due Thursday 1-Sept (noon)**  **Data: Mushroom**   1. Compute decision tree , interpret the results upload the output and paragraph to blackboard   **Due: Monday 5-Sept (EOD)**  **Data: Breast Cancer** | 5  5 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Wednesday Friday**  **Dates** | **Topics/Daily Activities** | **Readings  will be available on blackboard unless otherwise indicated** | **Deliverables  Due Dates** | **Points** |
| Week 3 | | | | |
| 7-Sept 9-Sept• | In Class review of decision tree  * How to skim a research paper      * Correlation * Regression * Partailing of variance * SPSS example regression, correlation: sepal length, petal width | * Reasoning with statistics Chapters 11,12, p 127-152 | 1. Skim research papers and summarize   **Due 8-Sept (Noon)**   1. Compute a correlation and regression in SPSS and Excel compare the results and interpret them; use variables sepal-width and petal length Upload all the analyses and results to blackboard   **Due Monday 12-Sept (EOD)  Data: IRIS** | 5  5 |
| **Week 4** | | | | |
| 14-Sept     16-Sept | * In Class review of Correlation and Regression * Logistic Regression * In Class review of Logistic Regression * Parametric statistics * Distributions: normal curve, t-distribution, CHI square, F-distribution * Transforming distributions | * DSB chapter 4 * Statistical modeling: the two cultures | 1. Download and compute a logistic regression for to predict the survival of Titanic passengers   **Due Monday 19-sept Data Titanic** | 5 |
| **Week 5** | | | | |
| 21-Sept 23-Sept | * **Presentation on Paper** | * Reading of your choice | Upload presentations to Blackboard **Due: Sunday 18-Sept (EOD)** | 10 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Wednesday Friday**  **Dates** | **Topics/Daily Activities** | **Readings  will be available on blackboard unless otherwise indicated** | **Deliverables  Due Dates** | **Points** |
| Week 6 | | | | |
| 28-Sept  **30-Sept** | * The Central Limit Theorem * Myths about sample size      * Guest Lecture: JJ Cadiz, Microsoft | * Hayes Statistics Chapter 7-9   **None – work on proposal presentation** | 1. Prove the central limit theorem   **Due Monday 3-Oct (noon)  Data: Titanic (again)** | 5 |
| **Week 7** | | | | |
| 5-Oct  7-Oct | * Presentation on proposed research | * Power point presentation | Develop proposal for data set to analyze and question to answer  pick data set  **Due: Presentation Friday 7-Oct (EOD)** | 10 |
| **Week 8** | | | | |
| 10-Oct    14-Oct | * Bayesian statistics * cost benefit * models * over-fitting * **Guest Lecture: Justin Shacklette  Activision** | * DSB chapter 5 * <http://www.r-bloggers.com/evaluating-model-performance-a-practical-example-of-the-effects-of-overfitting-and-data-size-on-prediction> | Turn your presentation into a one-page proposal for discussion following week, proposal should be in Word and submitted to blackboard  **Due:Proposal  Friday 14-Oct(EOD)** | 5 |
| **Week 9** | | | | |
| 19-Oct  21-Oct | * One on one meetings review write up and presentation of results | None revise presentation and prepare for data analysis |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Wednesday Friday**  **Dates** | **Topics/Daily Activities** | **Readings  will be available on blackboard unless otherwise indicated** | **Deliverables  Due Dates** | **Points** |
| Week 10 | | | | |
| 26-Oct    28-Oct | * Comparisons of models for data fitting, * Entropy * ROC curves * Guest Lecture: Dmitri Williams. Ninja Metrics |  | * Collect/analyze data Start write-up/presentation |  |
| **Week 11** | | | | |
| 2--Nov    4-Nov | * Parametric and non-parametric statistics * ANOVA – planned comparison * **<TBD>** |  | * Collect/analyze data Start write-up/presentation |  |
| **Week 12** | | | | |
| 9--Nov 11-Nov | * **<TBD>** * **Guest Lecture: Nathan Blau RIOT games** |  | * Collect/analyze data Continue write up/presentation |  |
| **Week 13** | | | | |
| 16, 18-Nov | * Present project to class |  | 1. Write up and presentation   **Due Friday -Nov 18 (EOD**) | 25 |
| **Week 14** | | | | |
| 23-Nov 25-Nov | **Thanksgiving** | | | |
| **Week 15** | | | | |
| 30-Nov  2--Dec | * Feedback on paper * Course evals * Notification of “final” grade |  | Optional revision of analysis and write-up  **Due 5-Dec EOD** | Can add points to research write up |
| **Week 16** | | | | |
| 3-Dec  6-Dec | * “study days “ |  | Final grades assigned, notification |  |
| **Week 17** | | | | |
|  | * No final |  |  |  |

**Statement for Students with Disabilities**

Any student requesting academic accommodations based on a disability is required to register with Disability Services and Programs (DSP) each semester. A letter of verification for approved accommodations can be obtained from DSP. Please be sure the letter is delivered to me (or to TA) as early in the semester as possible. DSP is located in STU 301 and is open 8:30 a.m.–5:00 p.m., Monday through Friday. Website and contact information for DSP: http://sait.usc.edu/academicsupport/centerprograms/dsp/home\_index.html, (213) 740-0776 (Phone), (213) 740-6948 (TDD only), (213) 740-8216 (FAX) [ability@usc.edu](mailto:ability@usc.edu).

**Statement on Academic Integrity**

USC seeks to maintain an optimal learning environment. General principles of academic honesty include the concept of respect for the intellectual property of others, the expectation that individual work will be submitted unless otherwise allowed by an instructor, and the obligations both to protect one’s own academic work from misuse by others as well as to avoid using another’s work as one’s own. All students are expected to understand and abide by these principles. *SCampus*, the Student Guidebook, ([www.usc.edu/scampus](http://www.usc.edu/scampus) or <http://scampus.usc.edu>) contains the University Student Conduct Code (see University Governance, Section 11.00), while the recommended sanctions are located in Appendix A.

Students will be referred to the Office of Student Judicial Affairs and Community Standards for further review, should there be any suspicion of academic dishonesty. The Review process can be found at: <http://www.usc.edu/student-affairs/SJACS/>. Information on intellectual property at USC is available at: <http://usc.edu/academe/acsen/issues/ipr/index.html>.

**Emergency Preparedness/Course Continuity in a Crisis**

In case of a declared emergency if travel to campus is not feasible, USC executive leadership will announce an electronic way for instructors to teach students in their residence halls or homes using a combination of Blackboard, teleconferencing, and other technologies.

**Instructor Bio:**

Dennis has worked in user research since 1981. He was a usability manager at Digital Equipment Corporation, where methods such as Usability Engineering, Contextual Inquiry, and data logging were developed. For the past 14 years he managed research teams at Microsoft, which have covered a wide spectrum of products, including: the Games User Research team, which developed RITE (Rapid Iterative Testing and Evaluation) and TRUE (Tracking Real-time User Experience) methods. TRUE has been successfully applied to many products, including the Halo franchise. Dennis has also been an active member of the CHI community for many years, serving as in a number of roles including Conference Co-chair and was elected Vice President for Conferences in 2004. He has co-authored over 50 articles, book chapters and talks on HCI with many valued colleagues. Dennis has also coauthored of two books: Field Methods Casebook for Software Design (with Dr. Judy Ramey) and Brave NUI World (with Dr. Daniel Wigdor).He is adjunct full professor at University of Washington in the department of Human Centered Design and Engineering Design. He holds a PhD. in experimental social psychology from Clark University

**SA Bio:**