Module: Descriptive and Inferential Statistics in Cybercrime Investigation

March 5, 2014

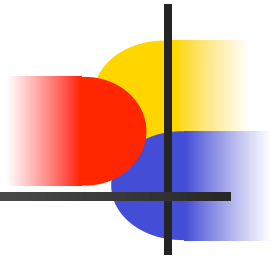
Wingyan Chung

http://cbia.stetson.edu/cic

STETSON UNIVERSITY
School of Business Administration
Center for Business Intelligence and Analytics

Work supported by the NSF under Grants Nos. DUE-1141209 and by Stetson University. Findings, conclusions, or recommendations expressed are those of the authors and do not necessarily reflect NSF's views.

Introductory Statistics



- A two-course series at business school
  - Descriptive Statistics (STAT201, half unit)
  - Inferential Statistics (STAT301, 1 unit)
- Data analytics initiative at Stetson
- Potential to combine into one course, and offer another intermediate statistics course to connect to advanced domain-specific analytics courses

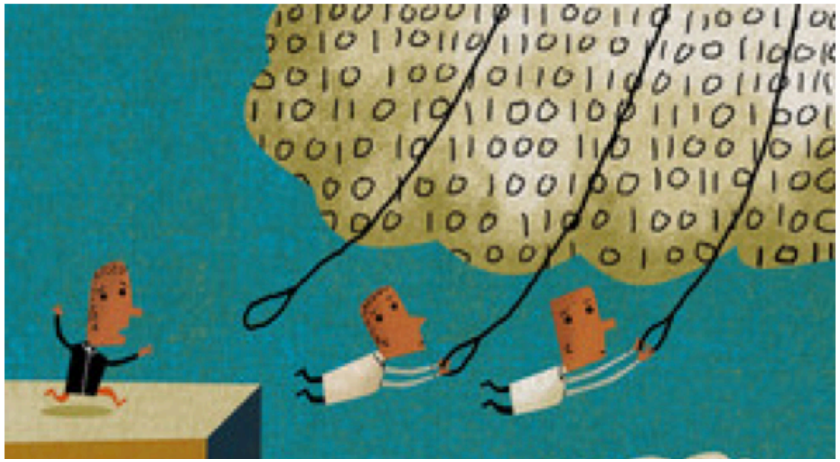
FEBRUARY 4, 2014, 12:40 AM ET

BY JAMES WILLHITE

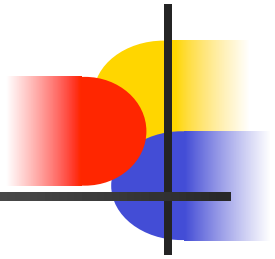
Getting Started in 'Big Data'

Wanted: Ph.D.-level statistician with the technical skill to use data-visualization software and a deep understanding of the _____ industry.

Fill in the blank with almost any business: consumer products, entertainment, health care, semiconductors or fast food. The list reflects the growing range of companies trying to mine mountains of data in hopes of improving product design, supply chains, customer service or other operations.

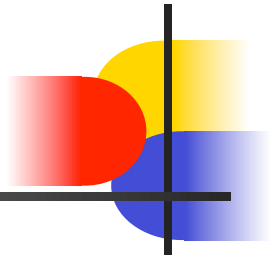


Module Development



- Cybercrime investigation can produce a large amount of data, requiring analyses to support understanding and decision making
- Measures taken to reduce cyber crime include user education, website certification, case reporting, etc.
- Crime Prevention and Prediction
  - Number of cases vs. intervention measures

Module Case

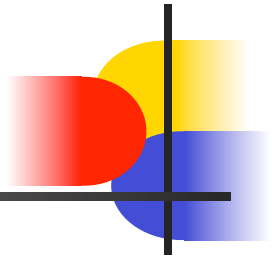


- Marvo Inc. is a growing company in the entertainment and hotel industry
- Operates 3 theme parks and 7 resorts and hotels near theme parks, and sells souvenirs, books, and multimedia
- Faces challenges of cyber attacks as it expands its online stores and online guest services
- Needs to better prevent and predict cyber crime through the use of statistics

Module Design

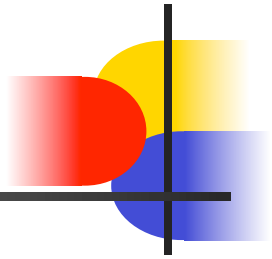
	Topic	Module
1	Understanding data types (ordinal, interval, etc.)	Classify types of data used in cyber crime investigation
2	Measures of location and dispersion (mean, median, standard deviation, etc.)	Understand how to compute central tendency and variation of number of cyber crime in different places and different times
3	Measures of shape and association (skewness, correlation, etc.)	Understand how to identify shape and association of pairs of data variables used in cyber crime investigation
4	Probability Distribution	Describe likelihood of cyber crime occurrence using different probability distribution
5	Linear Regression: Technique and Application	Predict the numbers of cyber crime cases based on intervention measures taken
6	Linear Regression: Testing Significance of Coefficients	Determine significance of the coefficients of a linear regression equation
7	Analysis of Categorical Data	Determine effectiveness of intervention measures in different types of cyber crime

Sample Module (1)



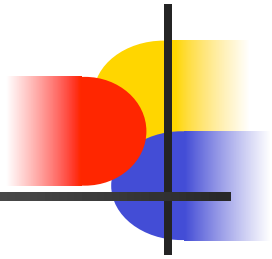
- Classifying types of cyber crime data
  - Video: Identity theft cases
  - Reading on different data types
  - Teamwork: Identify different data types in a spreadsheet file of cyber crime data
  - BOK: data type and data representation
  - Ex: MyStatLab - Assignment 1

Sample Module (2)



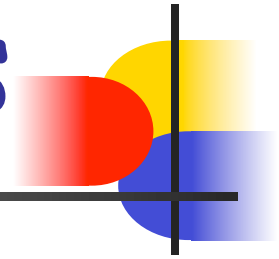
- Probability Distribution
  - Pair work: Plotting frequency distribution of crime cases with Excel
  - Estimating cumulative probability of cyber crime occurrences
  - Applying different distribution to describe likelihood of crime occurrence
    - Poisson, Normal, Uniform, etc.
  - BOK: probability distribution functions
  - Ex: MyStatLab – Assignment 4

Evaluation of Learning



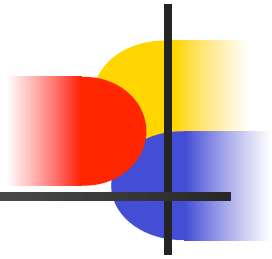
- Control vs. Treatment
  - Sections × Context (random assignment)
- Performance Measure
  - Assignment Score via MyStatLab Portal
    - http://www.mystatlab.com/
    - A commercial website associated with the textbook
    - Automated grading of student submission with automatic Q/A variation
  - Examination Score

Implementation of Modules



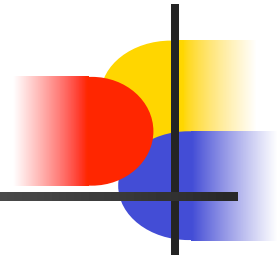
- STAT301 Business Statistics, Fall 2014
  - The second introductory statistics course
  - Mostly business majors, with others
- Cyber crime investigation support is a theme used in the module development
- Video, group activities, textbook tutorials, and exercises were used
- Evaluation of learning done in MyStatLab portal

Module Evaluation Plan



- Student learning outcomes
  - Assignment scores
- Student feedback
  - Helpfulness of modules
  - Connection with BOK
  - Satisfaction
- Instructor feedback
- Community feedback

Ongoing and Future Works



- Module Packaging
- Organizing and disseminating modules
- Evaluating modules
- Community adoption