Upgrading business statistics curriculum to meet the needs of knowledge workers

2018 Stata User Group Meeting, Vancouver

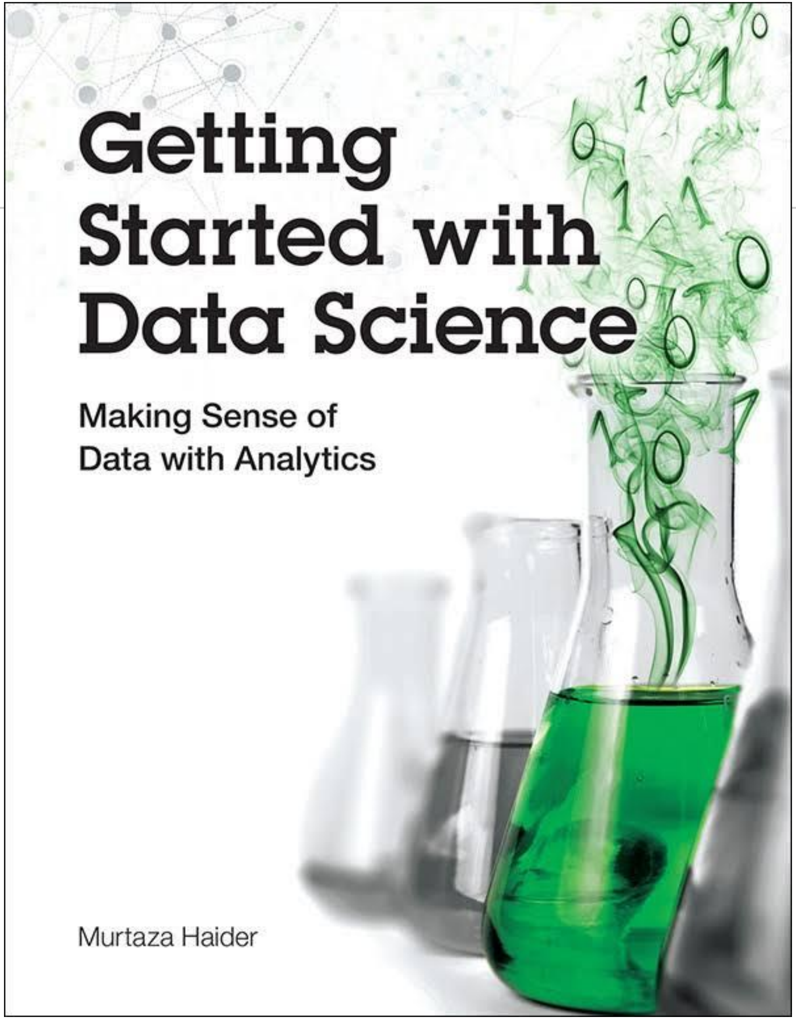
Murtaza Haider
Ted Rogers School of Management
Ryerson University, Canada

Outline

- A word about myself
- Questions:
  - Why are we teaching t-tests today?
  - Why business students are being taught the same curriculum as stats majors?
  - What needs to be taught: business statistics or data science?
  - What we teach, what has changed, what must be taught in Business Statistics

Murtaza Haider

- Academic
  - Teaching number crunching to non-statisticians
- Author
- Syndicated columnist with the Financial Post


Getting Started with Data Science
Making Sense of Data with Analytics
Murtaza Haider

1

Teaching Statistics

To non-statisticians

B Schools

- Business and management faculties are one of the largest in most schools
- The Ted Rogers School of Management enrollment stands at over 10,000 FTE
- Each student takes at least two courses in business statistics

300,000

Degrees conferred by North American business schools (2013/14)

1,100,000

Students enrolled in Business Faculties

Two

Business stats courses taken by undergraduate students

What is being taught?

First Course

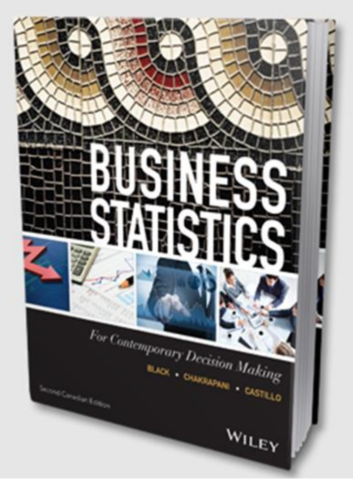
- Descriptive statistics
- Probability
- CLT
- Probability distributions
  - Normal
  - Binomial
- Hypothesis testing
  - T-tests
  - Correlation tests
  - ANOVA

Second Course

- Use of statistical software
  - Mostly SPSS or SAS
  - Rarely R or Stata
- Use of non textbook data sets
- Data collection and sampling
- Regression
  - OLS/ Simple Regression
  - Multivariate Regression
- May be Time series forecasting/GLM

The distribution of effort

- Focus remains on statistical theory and not data
- Calculator not software
- A mountain of topics before Regression

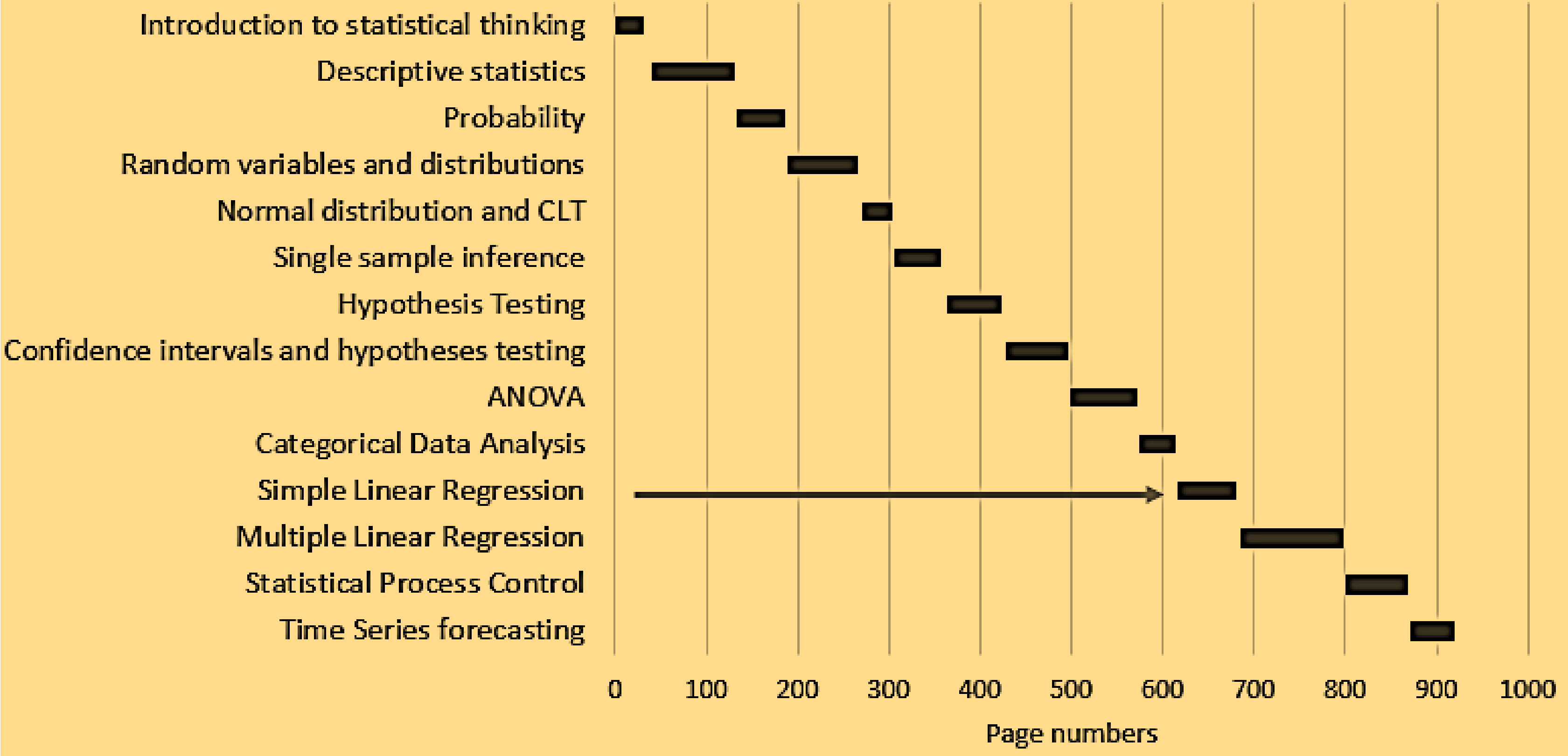

BUSINESS STATISTICS
For Contemporary Decision Making
BLACK • CHAKRAPANI • CASTILLO
Second Canadian Edition
WILEY

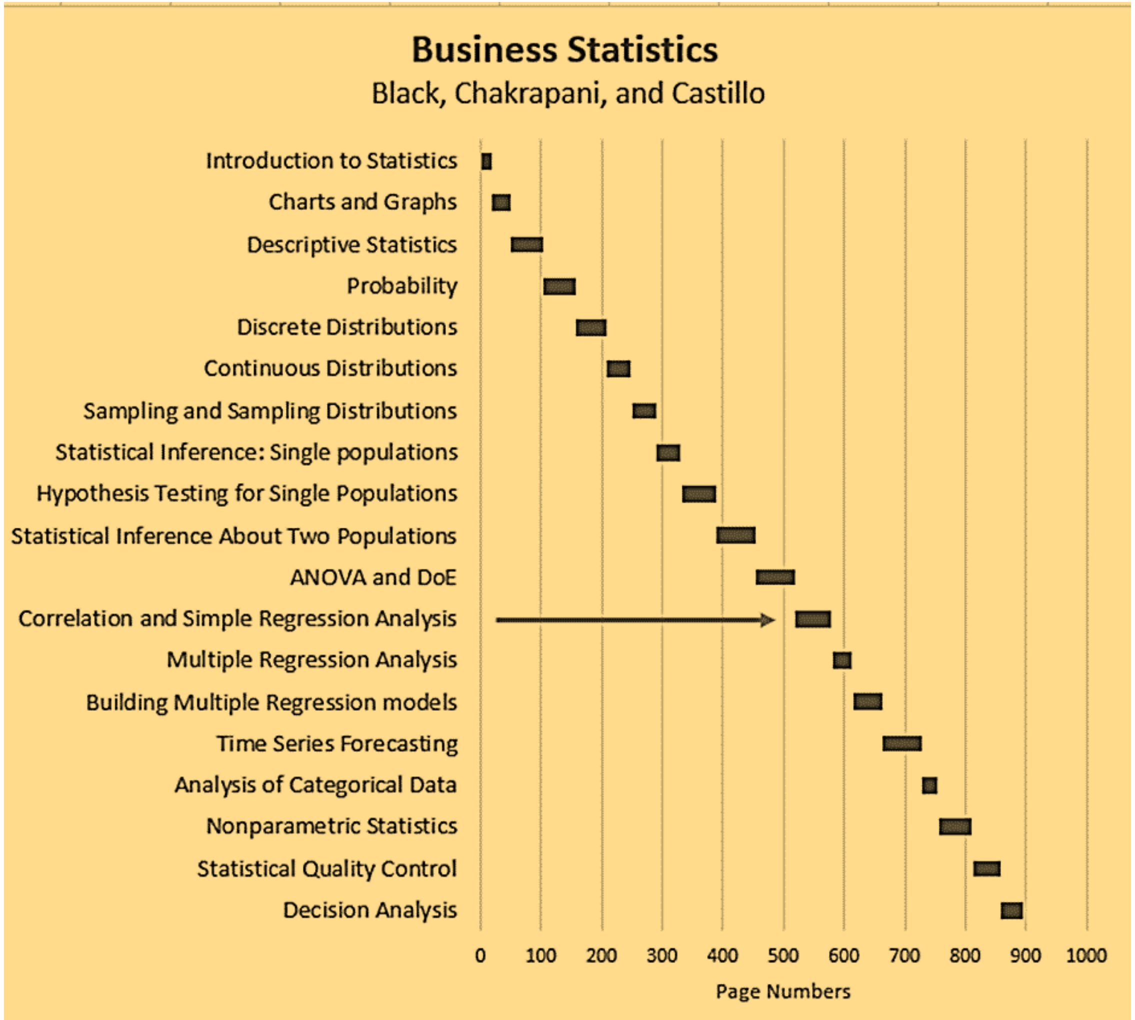
The 800 lbs. guerilla!

The road to Regression is paved with redundant statistical tools

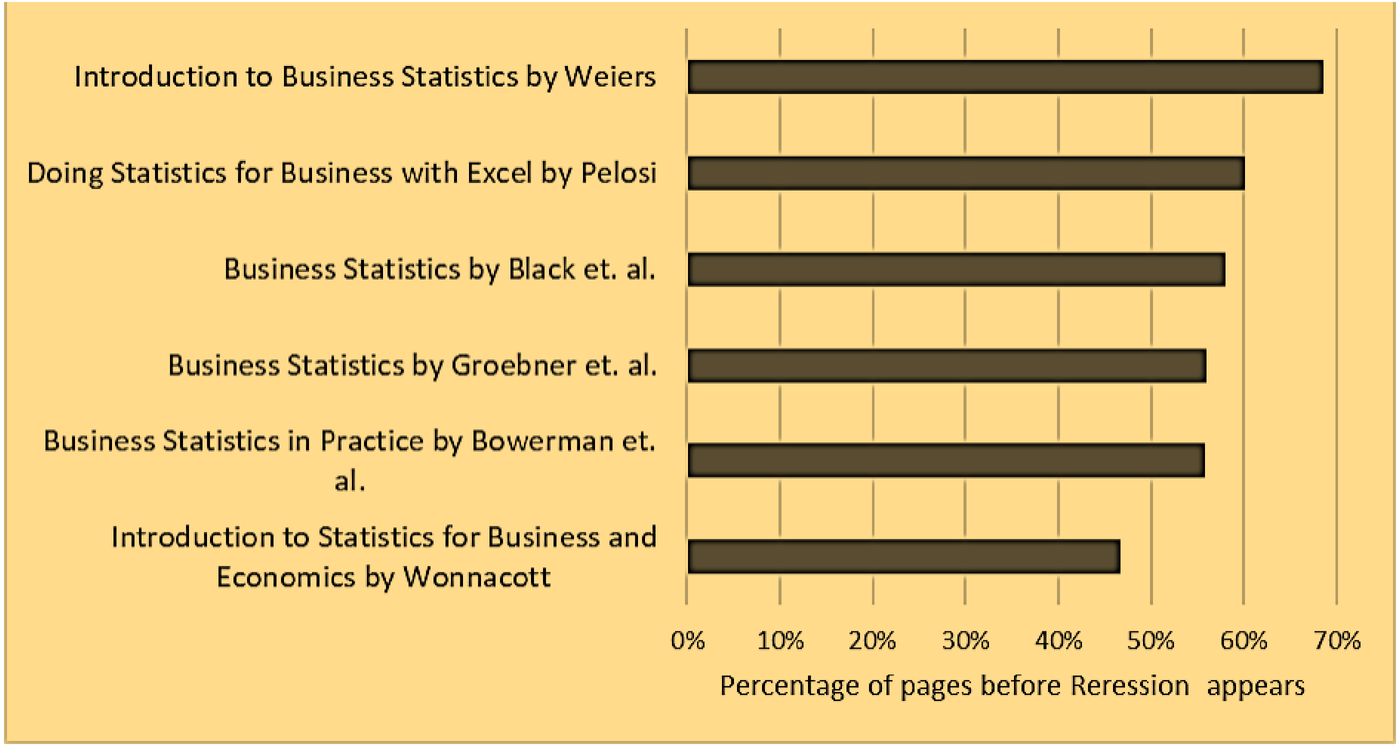
Statistics for Business and Economics

by McClave and Benson




Business Statistics
Black, Chakrapani, and Castillo
Introduction to Statistics
Charts and Graphs
Descriptive Statistics
Probability
Discrete Distributions
Continuous Distributions
Sampling and Sampling Distributions
Statistical Inference: Single populations
Hypothesis Testing for Single Populations
Statistical Inference About Two Populations
ANOVA and DoE
Correlation and Simple Regression Analysis
Multiple Regression Analysis
Building Multiple Regression models
Time Series Forecasting
Analysis of Categorical Data
Nonparametric Statistics
Statistical Quality Control
Decision Analysis
0
100
200
300
400
500
600
700
800
900
1000
Page Numbers

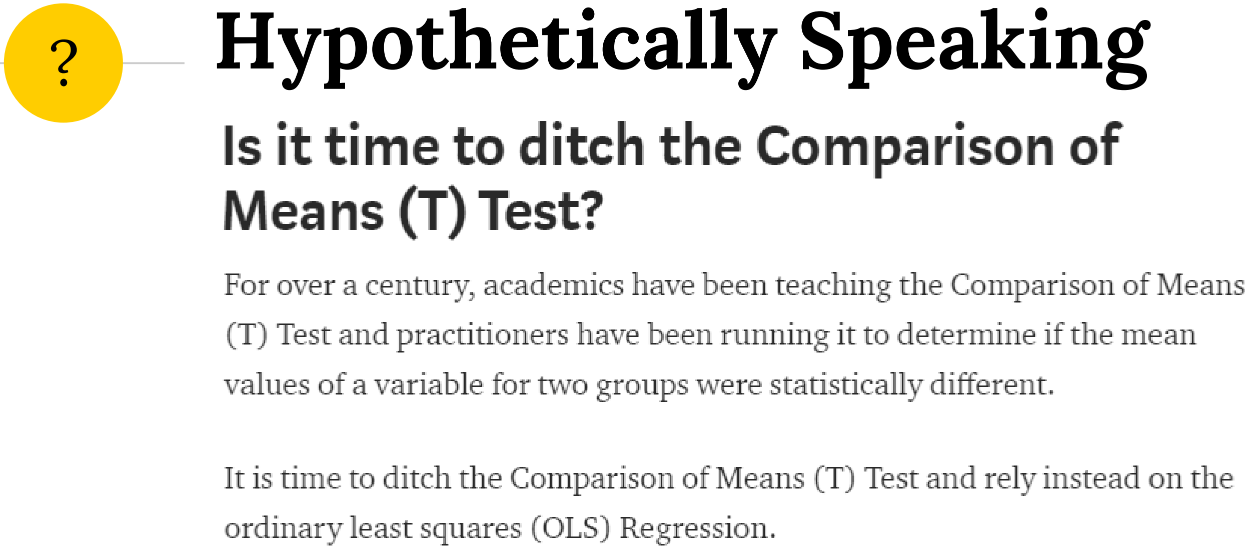
The Road to Regression



What’s up with Simple Linear Regression When All Else is Supposed to be Equal

“

https://medium.com/@regionomics/is-it-time-to-ditch-the-comparison-of-means-t-test-73571ccd8dd2

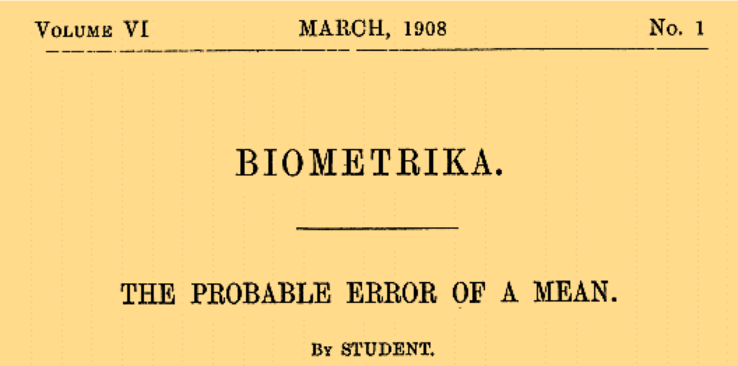


Hypothetically Speaking

Is it time to ditch the Comparison of Means (T) Test?

For over a century, academics have been teaching the Comparison of Means (T) Test and practitioners have been running it to determine if the mean values of a variable for two groups were statistically different.

It is time to ditch the Comparison of Means (T) Test and rely instead on the ordinary least squares (OLS) Regression.



Volume VI MARCH, 1908 No. 1

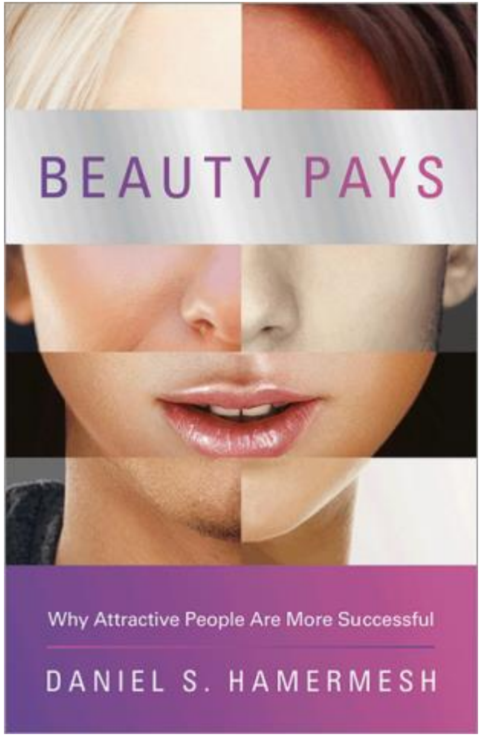
BIOMETRIKA.

THE PROBABLE ERROR OF A MEAN.

By STUDENT.

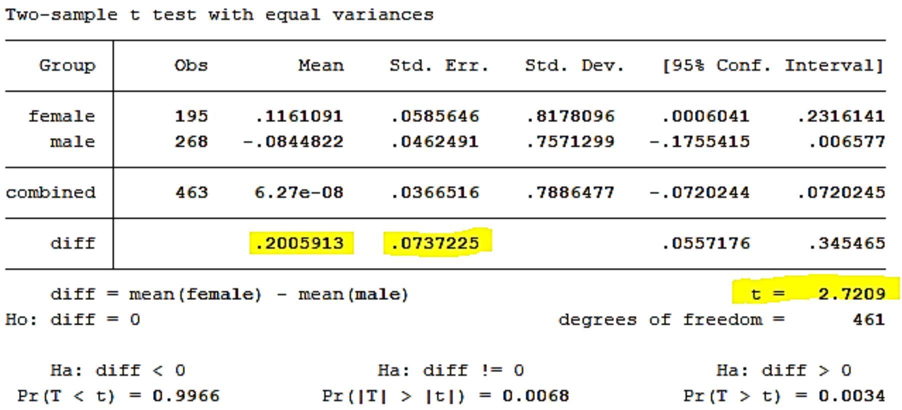
OLS with a continuous dependent variable and a categorical explanatory variable is the same as a T-test for comparison of means

The ultimate beauty test



With Equal Variances

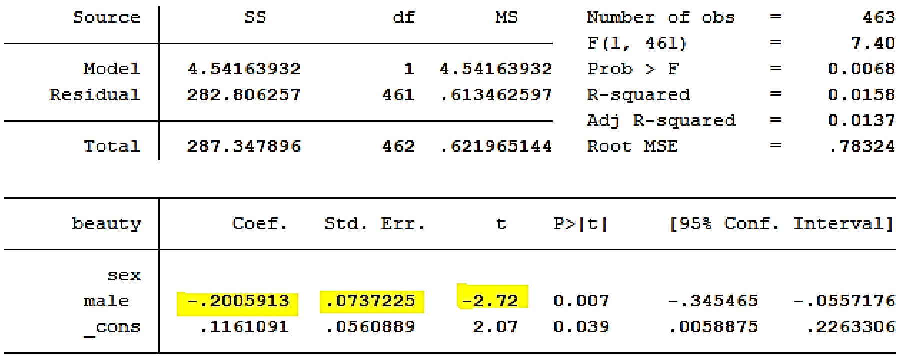
T Test



```
Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
  female |     195    .1161091    .0585646    .8178096    .0006041    .2316141
    male |     268   -.0844822    .0462491    .7571299   -.1755415     .006577
---------+--------------------------------------------------------------------
combined |     463    6.27e-08    .0366516    .7886477   -.0720244    .0720245
---------+--------------------------------------------------------------------
    diff |            .2005913    .0737225                .0557176     .345465
------------------------------------------------------------------------------
    diff = mean(female) - mean(male)                              t =   2.7209
Ho: diff = 0                                     degrees of freedom =      461

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.9966         Pr(|T| > |t|) = 0.0068          Pr(T > t) = 0.0034
```

OLS Regression

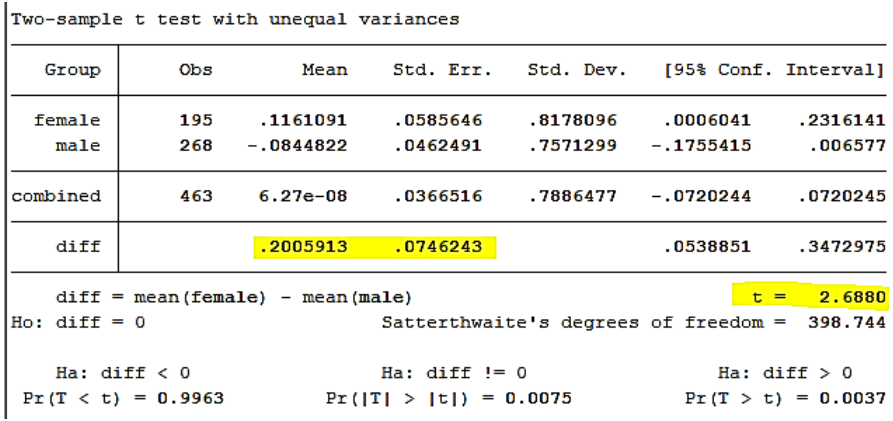


```
      Source |       SS           df       MS      Number of obs   =       463
-------------+----------------------------------   F(1, 461)       =      7.40
       Model |  4.54163932         1  4.54163932   Prob > F        =    0.0068
    Residual |  282.806257       461  .613462597   R-squared       =    0.0158
-------------+----------------------------------   Adj R-squared   =    0.0137
       Total |  287.347896       462  .621965144   Root MSE        =    .78324

------------------------------------------------------------------------------
      beauty |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         sex |
        male |  -.2005913   .0737225    -2.72   0.007     -.345465   -.0557176
       _cons |   .1161091   .0560889     2.07   0.039     .0058875    .2263306
------------------------------------------------------------------------------
```

With Unequal Variances

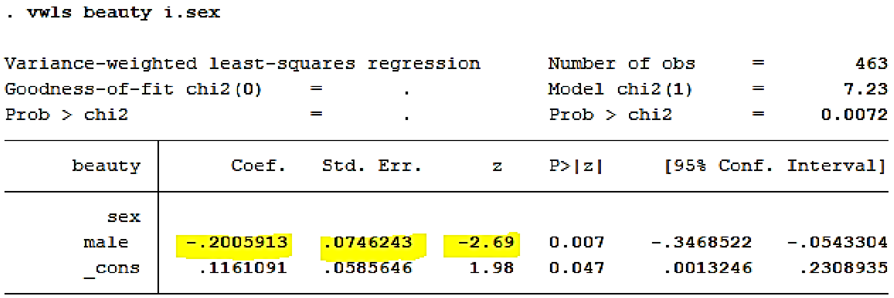
T Test



```
Two-sample t test with unequal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
  female |     195    .1161091    .0585646    .8178096    .0006041    .2316141
    male |     268   -.0844822    .0462491    .7571299   -.1755415     .006577
---------+--------------------------------------------------------------------
combined |     463    6.27e-08    .0366516    .7886477   -.0720244    .0720245
---------+--------------------------------------------------------------------
    diff |            .2005913    .0746243                .0538851    .3472975
------------------------------------------------------------------------------
    diff = mean(female) - mean(male)                              t =   2.6880
Ho: diff = 0                     Satterthwaite's degrees of freedom =  398.744

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.9963         Pr(|T| > |t|) = 0.0075          Pr(T > t) = 0.0037
```

OLS Regression



```
. vwls beauty i.sex

Variance-weighted least-squares regression           Number of obs   =        463
Goodness-of-fit chi2(0)  =        .                  Model chi2(1)   =       7.23
Prob > chi2              =        .                  Prob > chi2     =     0.0072

------------------------------------------------------------------------------
      beauty |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         sex |
        male |  -.2005913   .0746243    -2.69   0.007    -.3468522   -.0543304
       _cons |   .1161091   .0585646     1.98   0.047     .0013246    .2308935
------------------------------------------------------------------------------
```

The same goes for ANOVA *and* Correlation

Ditch what you can

Think Data Science, *not* Statistics

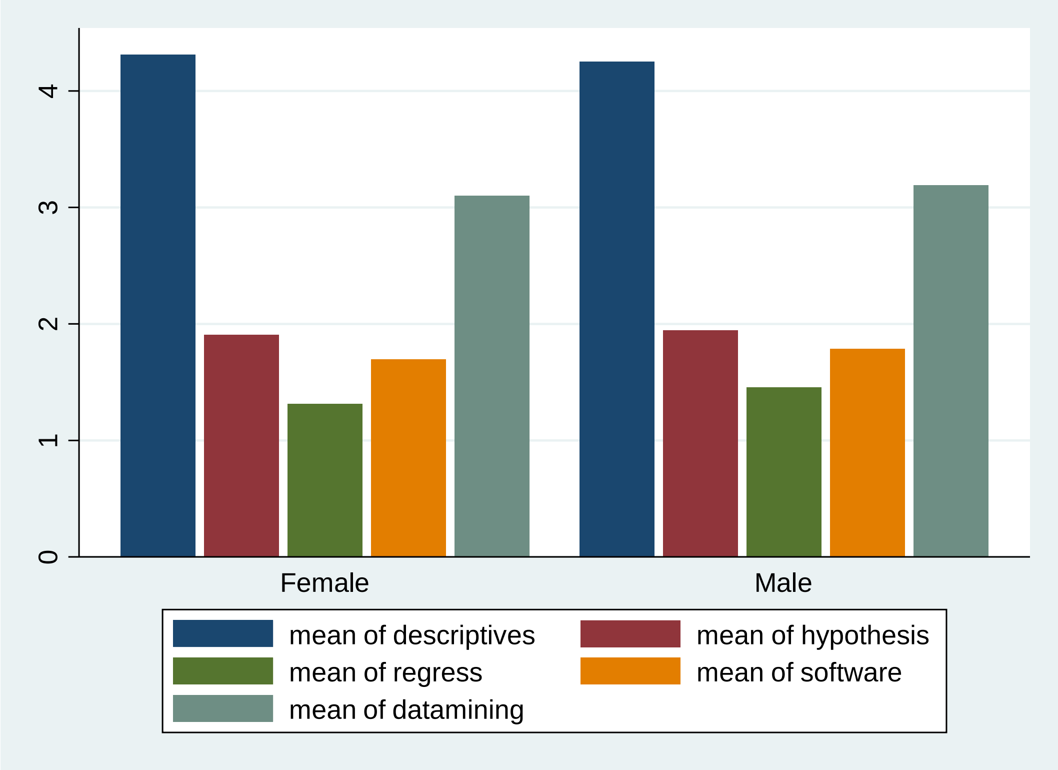
What are students learning?

A Case-Control Experiment

- 1700 students taking the second course in Business Statistics in the second semester at a certain school
  - The course contents are typical of a second course in business statistics
  - Working with a collaborator
- Divided in two groups:
  - Treated: Blended learning with online videos
  - Control: Same old same old
- Surveyed in the second half of course
- Some findings

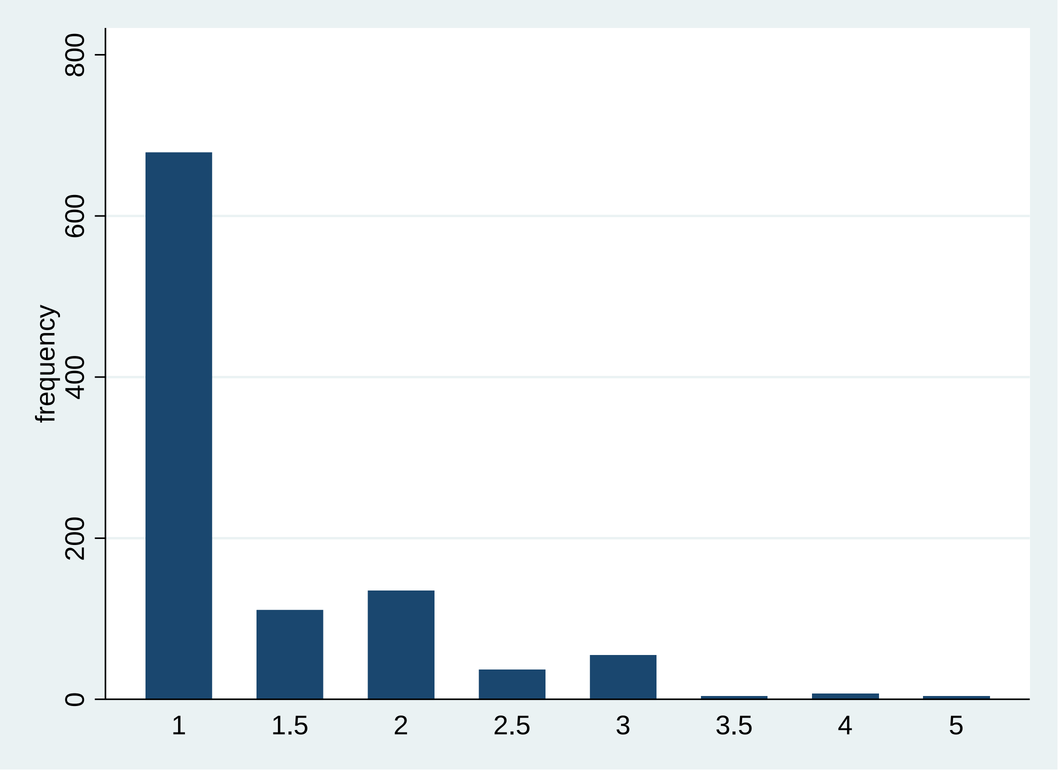
Competencies

2nd course in Biz Stats in Semester II

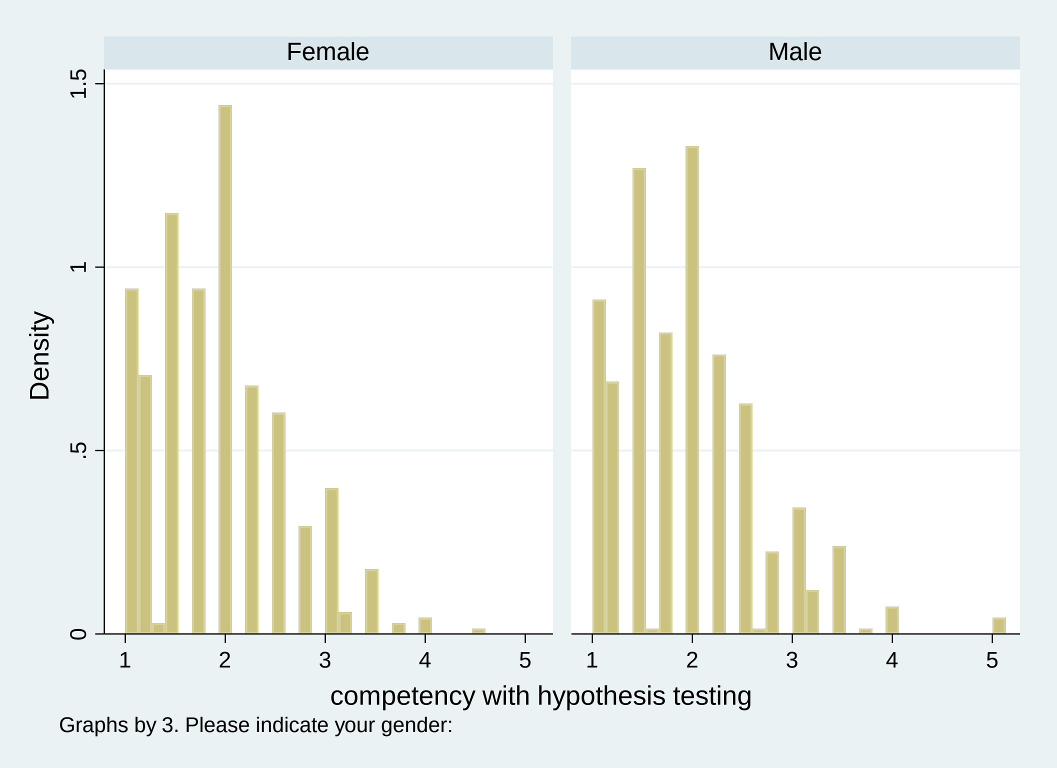


Regressing in Regression

2nd course in Biz Stats in Semester II

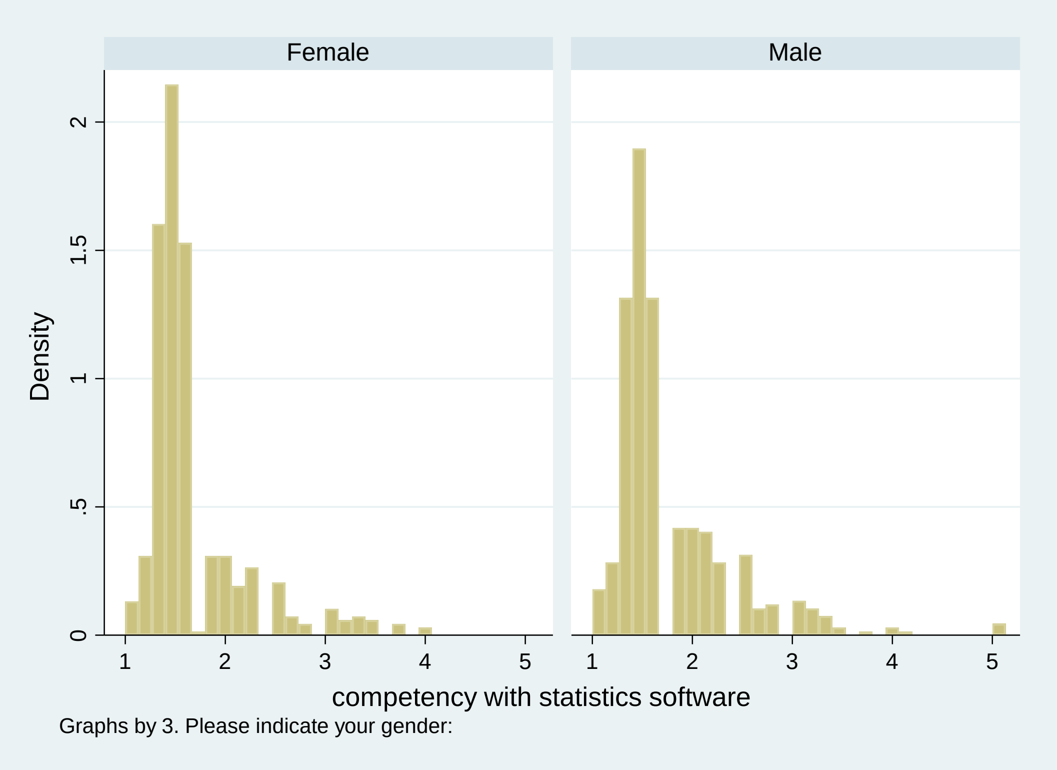


Hypothetically different



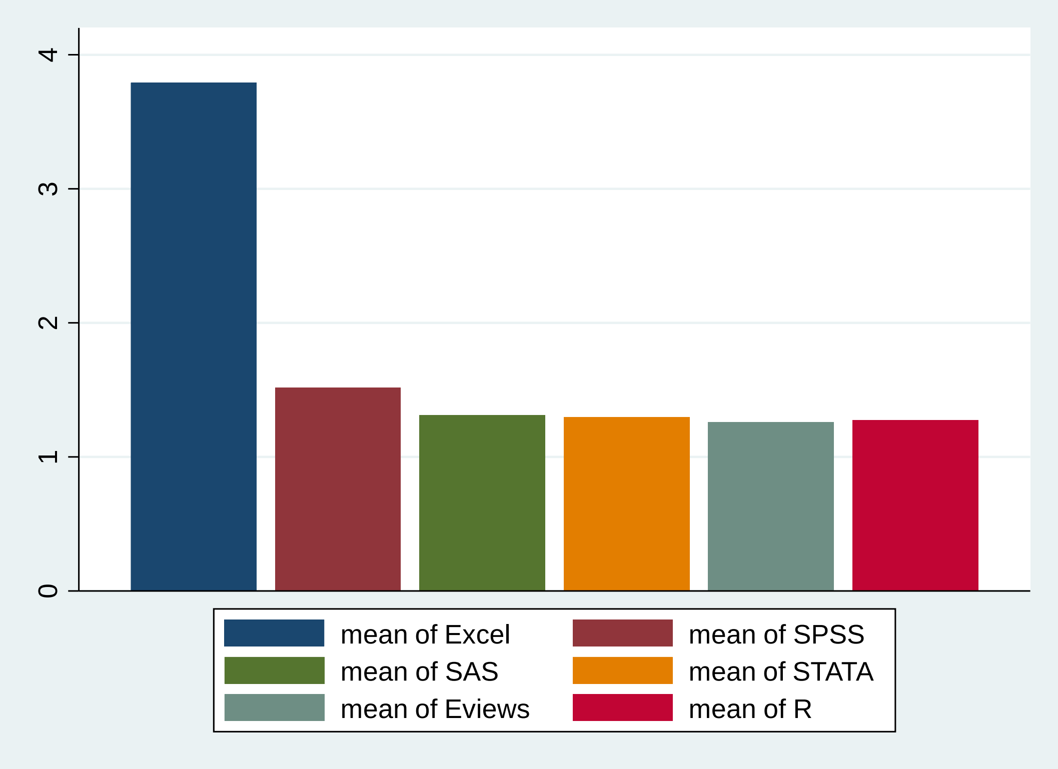
Soft skills

2nd course in Biz Stats in Semester II



Excelling in Excel

2nd course in Biz Stats in Semester II



Say hello to Big Data Science

What has changed?

- Lots of data ... CIT
  - Open data of all types
  - Machine generated
  - Survey data … Census, PEW, others
  - Consumption data
  - Web engagement data
- Open source software
  - R, Hadoop, etc.
- SAAS
- Cloud computing

Changing the computation engine from Mathematics to Computing in Statistics

The death of statistical inference
From Sample to Big Population Data

What should be taught

- Data comes first
- Start with a Puzzle
  - Curriculum should match the needs of the industry
  - Life as a biz analyst is about data-driven questions
- Data wrangling
- Data visualization
- Tabulations, X Tabulations
- Regression
- Machine Learning

We must get unstuck

Needless dependence on mathematics has made our thinking sticky

Teaching of Regression Methods, even if inference is postponed until late, nevertheless belongs to the mainstream

“

George Cobb, *The American Statistician*, 2015

Thanks!

Questions / Comments?

You can find me at

- @regionomics
- murtaza.haider@ryerson.ca
- +1-416-318-1365