DA503 - APPLIED STATISTICS

from a Practitioner's Point of View

Instructor:
Eren Ocakverdi

As a business practitioner myself, I have been using and benefiting from statistical modeling techniques for over a decade in the field of Turkish economy and its financial markets. Statistics and its cousin Econometrics provide us with indispensable tools to turn our intuition into insight. Regardless of the subject you study and the size of the data you have, these techniques will prevent you from making invalid inferences, at the very least.

As H. G. Wells once said: "*Statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write.*"

And I think, we are already there.

Self-test:
Just give the following questions some thought:
Suppose you are trying to predict the amount of credit given to customers using various economic and financial measures. You observe that adding new variables (predictors) increases your R-squared (R2) value and it has already reached to 0.8:

- What does 0.8% actually mean? How do you interpret this number?
- How do you choose between the alternative predictors?
- Do you think you should continue adding new predictors? Why?
- How do you know that the model is not incorrectly specified?

Course description:
Content of this course relies respectfully on the main pillars of statistical analysis, but does not follow any particular convention. In other words, any available kind of Statistics or Applied Statistics textbook is a good reference, but no single book is enough to cover these subjects nor the particular path taken in this course.

Participants of the program (fortunately) come from various backgrounds and with different job descriptions. Course is designed so as to embrace the blessing in such diversity and to avoid the "one size fits all" approach (or at least intended to do so).

Main focus of the lectures will be on learning the practical aspects of statistical analyses, but related theoretical background will also be provided to some extent in order to help you understand the gist of approaches/methods that we will study.

Software implementation and real-life applications are more than instrumental and of crucial importance in this respect. We will be using EViews software and related R packages along with various genuine data sets. This, however, should not be considered as a binding constraint or as an additional burden. What matters most here is the accurate and easy implementation of the techniques. You can use any other software package or programming language as long as you understand and master the methods covered in the class.

Grade points will be earned through the assignments (see outline below). You are encouraged, but not required, to form a team of two.

Suggested references

1) Introductory Statistics, Sheldon M. Ross, Academic Press, 3rd ed. 2010

2) Statistics for Business and Economics, Paul Newbold, William Carlson, Betty Thome, Pearson, 8th ed., 2012

Outline:

1. **Refresher and Warm-up (0.5 week)**
   - Random variable
   - Distribution function
   - Properties of data
   - Levels of measurement
   - Software literacy
   - Descriptive statistics
   - Lab exercise

     Assignment-1 (5 pts)

2. **Hypothesis testing and Inference (1.5 week)**
   - Inferential statistics
   - Validity and reliability
   - Central limit theorem
   - Sampling
   - Confidence interval
   - Sample size
   - Hypothesis testing
   - Power analysis

     Assignment-2 (5 pts)
   - One-sample t-test
   - Paired sample t-test
   - Independent (two) samples t-test

     Assignment-3 (5 pts)
   - One-way ANOVA
   - Two-way ANOVA

     Assignment-4 (5 pts)
   - Factor analysis
   - Lab exercise

     Assignment-5 (10 pts)

3. **Basic estimation and Causality (1.5 week)**
   - Regression analysis
   - Least squares estimation
   - Model selection
   - Diagnostic tools
   - Lab exercise
   - Maximum likelihood estimation
   - Lab exercise

     Assignment-6 (10 pts)
   - ANOVA vs. Regression
   - Time series analysis
   - Lab exercise

     Assignment-7 (10 pts)

4. **Bootstrap and Simulation (1 week)**
    - Resampling
    - Simple bootstrap
    - Regression bootstrap
        Assignment-8 (10 pts)
    - Random number generation
    - Simulation
    - Distribution fitting
    - Simulating a regression
    - Simulating critical values
    - Lab exercise
        Assignment-9 (15 pts)

5. **Forecasting and Prediction (1 week)**
    - Prediction vs. Forecast
    - Measurement error
    - Bias-variance trade off
    - Prediction evaluation
    - Forecast evaluation
    - Cross validation
    - Static vs. Dynamic forecasts
    - Stepwise regression
    - Lab exercise
        Assignment-10 (25 pts)

6. **Advanced Models (1 week)**
    - Bayesian linear model
    - Markov Chain Monte Carlo
    - Gibbs sampling
    - Nonlinear regression
    - Binary response model
    - Ordered choice model
    - Count model
    - Censored regression model
    - Truncated regression model
    - Heckman selection model
    - Variance model