11.220 Quantitative Reasoning and Statistical Methods for Planning I

Department of Urban Studies and Planning, MIT, Spring 2015, Credits: 4-2-6
Lectures: MW 11:00-12:30; 9-354
Recitations: *W2* (1-242, 1-273) or *R EVE (5 PM)* (9-450A)
Stellar website: https://stellar.mit.edu/S/course/11/sp15/11.220/

Instructor: Jinhua Zhao jinhua@mit.edu
Office 9-523, Office Hour: R 2-3:30pm; Assistant: Phil Sunde, 9-316, psunde@mit.edu

TAs:

- Kelly Heber k_heber1@mit.edu; Office: 9-319; W3-4pm or by appointment
- Lyndsey Rolheiser lyndseyr@mit.edu; Office: 9-324; W4-5pm or by appointment
- Faizan Siddiqi fjawed@mit.edu; Office: 9-266; R6-7pm or by appointment

Human Beings are not Statistically Intuitive

Many, if not most, planners work with quantitative data. Some summarize, analyze, and present data they have collected themselves or have obtained from secondary sources; others must review quantitative analyses and assess the validity of arguments made therein. This course is designed to prepare you to critically review analyses prepared by others, as well as to conduct basic statistical analysis of data yourself.

People use *narratives* as the primary thinking and communication instruments. More true is this for planners! We empathize dearly with students who find statistics difficult--and this is precisely the reason for a class dedicated to it. We present six planning cases in the course:

1. India Slum and Informality
2. Infrastructure Financing (optimism bias and strategic misrepresentation)
3. African Education (program evaluation, policy design)
4. Climate Change and Cities (resilience)
5. What's in a House Price? (Tax Capitalization)
6. Natural Resources Protection

which start with the planning questions, explain in these particular contexts QR can be the most fitting approach, and naturally package the statistical concepts within. Using these examples of quantitative analyses related to the planning profession, you will become familiar with a variety of tools for describing and comparing sets of data, as well as those used to generate estimates and test hypotheses. The course also emphasizes the development of sound arguments and research design, such that students appreciate both the power and limits of quantitative analysis in argumentation. Unlike many other statistics classes, 11.220 pays particular attention to developing the skill of expressing statistical ideas in clear, simple language. We view these skills as essential for effective planning practice.

Figure 1: Mapping between QR Cases and Planning Topics

Cases	**Planning Topics**
India Slum and Informality	IDG/HCED/CDD: Informal Settlement, Affordable House
Infrastructure Financing (optimism bias)	UIS/CDD/Transportation: Reference Class Forecasting; Benefit/Cost Analysis
African Education (which policy works)	IDG: Program evaluation, Policy Design, Randomized Control Trial
Climate Change and Cities (resilience)	EPP: Climate Change, Resilience
What's in a House Price? (Tax Capitalization)	CRE/HCED/CDD: Urban Economics, Public Policy, Property Tax
Natural Resources Protection	EPP: Resource Management

Figure 2: Mapping between QR Cases and Statistical Concepts

Cases	**Statistical Concepts**
India Slum and Informality	Descriptive Statistics and Intro to Probability
Infrastructure Financing (optimism bias)	Probability Distribution
African Education (which policy works)	Sampling, Confidence Interval, Test of Significance
Climate Change and Cities (resilience)	Hypothesis Testing, Differences in Means
What's in a House Price? (Tax Capitalization)	Scatterplot, Correlation, Multivariate Linear Regression
Natural Resources Protection	Nominal/Ordinal Data, Logistic regression

In this course students will learn to use Stata for quantitative and statistical analyses and may also use Access, Excel, and ArcGIS when analyzing a dataset of their choice for the final paper. In addition to lecture, there is a required 1-hour session each week that will alternate between a laboratory session for hands-on practice with the software and a classroom recitation consisting of problem solving using statistical methods and equations.

Student expectations

Learning Norms

1. Reading before the class: **to establish the norm of reading**
2. No laptops in class
3. In-class idea notes
4. Learning from each other

Grading

Items	%
Reading and class participation (including in-class idea notes)	10%
Five Problem Sets (10%, 6%, 6%, 6%, 12%)	40%
Quizzes (10%, 10%)	20%
Term project (Idea 1%; Proposal 4%; Interim report 5%; Full report 12%; Revised report 8%)	30%

Readings

1. Case Materials

2. Required textbook: Meier, K., J. Brudney and J. Bohte (MBB). 2011. *Applied Statistics for Public and Nonprofit Administration*. 8th edition. Belmont, CA: Wadsworth. (Available at the Coop and on reserve in Rotch Library.)

3. Required Articles

- Flyvbjerg, Bent. "Curbing optimism bias and strategic misrepresentation in planning: Reference class forecasting in practice." *European Planning Studies* 16.1 (2008): 3-21.
- Plantinga, Andrew J., and Stephanie Bernell. "The Association between Urban Sprawl and Obesity: Is It a Two-Way Street?." Journal of Regional Science 47.5 (2007): 857-879.
- Duranton, Gilles, and Matthew A. Turner. "The fundamental law of road congestion: Evidence from US cities." *The American Economic Review* (2011): 2616-2652.

In-class idea notes

At the end of the each class, we'll reserve 5 minutes for you to write an in-class idea note. You may use it to

- What does the lecture **inspire** you to think? Either as a practitioner, as a researcher, or as a citizen
- How does the lecture **connect** to any of your personal experience, public events, planning debates, other courses, or your future career?
- Do you learn any new **data source** that might of use for your future work?
- Any puzzles, confusions or misunderstandings of the **statistical concepts** discussed today
- Anything you've learned on research design? Any ideas for your **term project** topics?

It is NOT graded. Everyone gets the full score as long as you submit it. We will collect and use them to gauge the class pace, target our recitations and improve the next lectures. You may consider this as a mini class evaluation and constant feedback.

Problem Sets

1. Encourage group work but individual submission required; please write down collaborators' names.
2. Please submit a hard copy to your TA at the beginning of lecture on the day that it is due. It is your responsibility to ensure the printed copy makes it to the TA in time.

3. Late policy:
   a. For problem sets 1, 2, 4, late assignments will receive a **20% deduction per day up to two days**. After 2 days, late assignments will not be accepted and a 0 will be given.
   b. For problem sets 3, 5, **NO late assignment will be accepted and a 0 will be given.** This is because problem sets 3 and 5 are due before the two quizzes and we need to distribute solutions as a study aid for the quizzes.

Term Project

Each student will work on a term project with the topic selected by the student. We've introduced five milestones throughout the term: Idea; Proposal; Interim report; Full report; Revised report. This design serves two purposes: 1) to avoid procrastination by breaking it into smaller and more manageable parts; 2) for us to provide you with timely feedback. We'll distribute the Term Paper Guide with detailed instruction and clear expectations. Please see Appendix I for the term project titles from the QR2014 class.

Recitation

Students are strongly encouraged to join the weekly recitation.

- Regular sessions: Wed 2-3pm 1-242; Wed 2-3pm 1-273; Thu 5-6pm 9-450A
- Special arrangement for Thu, Feb 12: no recitation that day. Please come to the Wed sessions on Feb 11 instead.

Please complete this survey to indicate your preferred time slot. You will do our best to accommodate your preference but cannot guarantee it. http://goo.gl/forms/uDhlXeZAlq

Statistical computing:

This course uses the statistical software program Stata. Prior experience in Stata is not required. Teaching assistants will teach basic Stata programming during recitations and labs. The following is a list of resources about how to use Stata:

- Kohler, U., & Kreuter, F. (2009). Data analysis using Stata (2nd ed.). College Station, TX: Stata Press. Table of contents: http://www.stata.com/bookstore/daus2.html
- Official Resources and support for Stata: http://www.stata.com/support/
- Stata Resources (UCLA): http://www.ats.ucla.edu/stat/stata/default.htm
- Introduction to Stata (UNC): http://www.cpc.unc.edu/research/tools/data_analysis/statatutorial/index.html
- Stata Tutorial (Princeton Univ.): http://data.princeton.edu/stata/default.html

STATA, R and RStudio are available for both Windows and Macintosh. Further, they are all available under both Macintosh and Windows on our CRON cluster computers.

- SATA: <http://ist.mit.edu/stata/all>
- R: <http://cran.r-project.org/bin/macosx/>
- R Studio: <http://www.rstudio.com/products/rstudio/>

R and RStudio are available to students FREE to install on their own computers.

STATA requires the purchase of a license ($30-$100 depending on sophistication of package). You can also run STATA WITHOUT PURCHASING THE SW, albeit, with a tiny bit of inconvenience. <http://cronlasso.mit.edu/cron/p.lasso?t=8:5:10>

Class Schedule

	Cases and Topics	Reading	Out	In	Paper
L1	02/04 QR in a Nutshell				
Case 1: Slums in India					
L2	02/09 Slum in India (Faizan); Measurement and Data	Case1 M2, 3	1		
L3	02/11 Descriptive Statistics; Brief Intro to Probability	M4-6			
R1	02/11 Research Design *Special: No Thu recitation				
	Feb 16 Presidents Day -- Holiday				
Case 2: Infrastructure Finance					
L4	02/17 Reference Class Forecasting	Case 2	2		
L5	02/18 Probability	M7			
R2	Lab 02/18, 02/19 Intro to Stata				02/20 Idea
L6	02/23 The Normal and Binomial Distributions	M8,9		1	
Case 3: African Education					
L7	02/25 African Education (Miguel & Lyndsey); RCT	Case 3	3		
R3	02/25, 02/25 Probability				
L8	03/02 Inferential Statistics	M11			
L9	03/04 Confidence Intervals; Tests of Significance	M11		2	
R4	03/04, 03/5 Sampling, Estimation, Confidence Interval				
Case 4: Climate Change and Cities					
L10	03/09 Climate Change and Cities	Case 4	4		
L11	03/11 Hypothesis Testing	M12		3	
R5	03/11, 03/12 Hypothesis Testing				
L12	03/16 T-test, Z-test, Differences in Means	M13,14			
L13	03/18 Quiz 1 (in class)				
R6	03/18, 03/19 Difference of Means and Proportions				
	Mar 23~27 Spring Break				
L14	03/30 Guest Lecture (Fadi Masoud)			4	
Case 5: What's in the housing price?					
L15	04/01 What's in the housing price?	Case 5	5		

	Introduction to Causal Inference				
R7	Lab 04/01, 04/02 Hypothesis Testing etc in Stata				
L16	04/06 Scatterplots and Correlation; Katherine McNeill (tbc)	M18			
L17	04/08 Linear Regression in a Nutshell	M18			
R8	04/08, 04/09 Regression Analysis				
L18	04/13 Multiple Regression	M21			Prpsl
L19	04/15 Practical Matters	M23			
R9	Lab 04/15, 04/16 Regression Analysis				
	Apr 20 Patriots Day -- Vacation.				
L20	04/22 Art of Model Building	M24			
R10	Lab 04/22, 04/23 Linear Regression				
L21	04/27 Assumptions and Endogeneity	M19		5	
L22	04/29 Quiz 2 (in class)				
Case 6: Natural Resources Management					
L23	05/04 Natural Resources; Nominal and Ordinal Data	Case 6, M15			Interim
L24	05/06 Logistic Regression	M16			
R11	Lab 05/06, 05/07 Logistic Regression; Term Paper				
L25	05/11 Evidencing Causality: Advanaed Topics				
L26	05/13 Ten Ideas from QR: Course summary				Full
	05/18 Revised Paper Due				Rvsd

Appendix I Term Project Titles from 2014 QR Class

- The Relationship between Crime Cases and Living Patterns in Neighborhoods of Boston
- The Impact of LIHTC Construction on Neighborhood Crime in San Francisco
- Calculating the Impact of Residential Energy Efficiency Upgrades in Somerville, Massachusetts
- Is Home Ownership the Answer to Canada's First Nations Housing Crisis?
- Shored Up? Investigating if Public Beach Nourishment Funds Advantage Communities in New York and New Jersey with Higher Socioeconomic Status
- Displacement after the Loma Prieta Earthquake
- An Initial Analysis of the Potential for Energy Efficiency-Focused Workforce Development in Eastern Kentucky
- The Knowledge-Driven Economy and Housing Affordability Evaluating the Relationship between High-Skill Jobs and Metropolitan Rents
- Using Regression to Explore City Parks and Crime Rates: State level analysis of hate crimes in the US
- Vehicle Ownership In Cambridge, Massachusetts
- What influences recycling rates? Exploring the relationship between income, age and recycling diversion rates in New York City's community districts
- Statistical analysis of historical districts in Boston
- Job Creation and the Shale Gas Industry A Case Study of Pennsylvania
- Leaders actions and employee activity
- The Mutual Influence of Park Funding and Exercise An examination of the relationship between total park expenditures and physical activity in US cities
- Affordable for All? Examining the Correlation between Rental Burden and Race in Cook County, Illinois
- Transit Ridership in the City of Chicago Impacts of Race, Income, Automobile Access and Distance to Transit
- Water Pollution & Agriculture The Relationship between Crop Area, Population, and Nitrate-Based Water Quality Violations In Pennsylvania
- Is the Conflict Between the United States & Iran Associated With the Change in the Number of Iranian Students in the U.S. Universities?
- Pawnshops and Poverty: The Relationship between Alternative Financial Institutions and the Communities They Serve
- The Suburban Advantage: Student Academic Achievement From City Center to Suburban Sprawl
- Decentralization And Access To Development By The Most Poor
- Budgeting for Homelessness: What Factors Should Impact Hud Funding Allocations
- Opening The Black Box: The Role Of Institutions In Development

- Is Ethnic Diversity Associated With Human Development In Africa?
- Concentrated Poverty and Depression: An Analysis of New York City Neighborhoods
- Does Transportation Mode Have Relationship with Public Health? A Regression Analysis of Obesity Rates and Public Transit Mode Share & Walking/Biking Mode Share
- More parks or not? The Relationship between Crime and Parks in Community of Chicago

Appendix II Supplementary Learning Materials

Free Online Courses

- EdX
  - Introduction to Statistics
  - Introduction to Statistics: Inference (Stat 2.3x)
- Coursera
  - Duke University: Data Analysis and Statistical Inference (coursera)
  - Statistics One by Andrew Conway, Princeton: https://www.coursera.org/course/stats1
  - https://class.coursera.org/dataanalysis-001/class/index
  - Data Analysis by Jeff Leek: https://www.coursera.org/course/dataanalysis
  - Visualizing data: https://itunes.apple.com/us/course/visualizing-data/id510187558
  - Passion Driven Statistics: https://www.coursera.org/course/pdstatistics
  - Statistics: Making Sense of Data by Alison Gibbs: https://www.coursera.org/course/introstats
  - Think Again: How to Reason and Argue
- Udacity: Introduction to Statistics
- Berkeley SticiGUI: http://www.stat.berkeley.edu/~stark/SticiGui/index.htm
- On Endogeneity (and Instrument Variables):
  - https://www.youtube.com/watch?v=CCiIfjm8qjw
  - https://www.youtube.com/watch?v=lLI-0pK9MD8
  - https://www.youtube.com/watch?v=OWHCbEP56ms
  - https://www.youtube.com/watch?v=cX5q_dKt6iU
- Omitted Variable Bias
  - https://www.youtube.com/watch?v=6I1tUM0RB6I
  - https://www.youtube.com/watch?v=_Ka_PAvdDjk
  - https://www.youtube.com/watch?v=CndHm9WDVIE

Data Analytics Examples

- TED Ed
  - http://ed.ted.com/lessons/hans-rosling-shows-the-best-stats-you-ve-ever-seen
  - http://ed.ted.com/lessons/peter-donnelly-shows-how-stats-fool-juries
  - http://ed.ted.com/lessons/david-mccandless-the-beauty-of-data-visualization
- Khan Academy:
  - https://www.khanacademy.org/math/probability/descriptive-statistics
  - https://www.khanacademy.org/math/probability/statistics-inferential

- Youtube Jeff Leek Statistics Channel
- Simply Statistics: http://simplystatistics.org/
- Economist: http://www.economist.com/blogs/graphicdetail
- New York Times:
    - Data http://data.nytimes.com
    - http://www.nytimes.com/interactive/2009/04/29/business/2009-wide-housing-graphic.html
    - http://www.nytimes.com/2009/05/09/business/09charts.html
- **Simulation Tools:**
    - Sampling Distributions http://onlinestatbook.com/stat_sim/sampling_dist/index.html
    - Impact of Outliers on a Regression Fit http://www.stat.sc.edu/~west/javahtml/Regression.html
    - Effect of Sampling on a Regression Fit http://lstat.kuleuven.be/java/version2.0/Applet003.html
    - Minimizing the Sum of Squared Residuals http://hspm.sph.sc.edu/courses/J716/demos/LeastSquares/LeastSquaresDemo.html

Additional Reading

- Kennedy, P. (2008). *A Guide to Econometrics* (6 edition.). Malden, MA: Wiley-Blackwell. (5th or higher edition is fine)
- Daisley, P. (1979). Statistical Thinking Rather Than Statistical Methods. Journal of the Royal Statistical Society, Series D, 28, 4, pp. 231–239.
- Naked Statistics
- Deborah Stone: Interperative Numbers
- Glaeser, E. 2005. "Review of Richard Florida's the Rise of the Creative Class." Regional Science and Urban Economics 35, no. 5: 593-96.
- Horwitz, L., and L. Ferleger. 1980. Chapter 1: Statistics and logic. In *Statistics for Social Change*. Boston: South End Press.
- Hodge, G. 1963. The use and mis-use of measurement scales in city planning. *Journal of the American Institute of Planners* 29, no. 2: 112-121.
- Moore, D. and W. I. Notz. 2006. Chapters 1 – 4 in *Statistics: Concepts and Controversies*. 6th edition. New York: W.H. Freeman.
- Moore, David S. 2004. Scatterplots and correlation. Chapter 4 in *The Basic Practice of Statistics*. 3rd edition. New York: W.H. Freeman.

Acknowledgement

I would like to thank Kelly, Lyndsey, Faizan for their amazing effort of creating the cases; Chris Zegras, Albert Saiz and Ezra Glenn for their concrete advices on the case based approach and overall pedagogy, and Dennis Frenchman, Eran Ben-Joseph, and Sandra Wellford for their generous institutional, financial and attitudinal support!