Basic Data

Stephen T

Business Administrati

Lecture 1: Ap

The Times They

Come gather 'round people, where
And admit that the waters around
And accept it that soon, you'll be a
If your time to you is worth saving
Then you'd better start swimming

We Need to Chan

- Everybody in Japan hopes for
  monozukuri. *Shakō* teaches ab
- Those are still value-added are
  they probably won't lead to gr

Everything I need t
Basic Data

- This is one of the most import
  MBA-MPP program.
  - The only material that is m

I'm glad y

- Modern organizations generate
- According to psychologists, hu items of data directly.
- Some processing goes on subco

Descriptive statistic

- In the U.S., the Senate (legisla
representatives from each state
California. However, we don't
anything else about states, to
Since there are two Senators p

Descriptive statistic
elect

- Most modern democracies insi
representation of citizens. For
"small electoral districts" whe

I'm really gla

- Many fields of business science
related disciplines. For exampl
- Even with modern precision en
production process may not be

I'm still glad

- But the *customer* is likely to d
  claims.
- A failure is very costly; the cus
- A *Bayesian analysis* can help y

"Hens

- The Japanese educational syst
  ranked score on a 20-80 scale v
  distribution of scores.
  - Schools use the hensachi of
  - Prospective students use th

Hensachi,

- Should hensachi of students be
  hensachi of universities?
- Can the hensachi of a student
  university?

Financial in

- The *market efficiency hypothes*
about a security affects its ma

- We say that information alread
discounted by the market. We
use it to make unusually profit

Financial instr

- For example, *martingale theor*
  strategy "buy and hold until p
  can be successful, *i.e.*, on avera
- The answer for a simple rando

Don't Over

Use several different methods, whic
to increase both breadth and depth

- In the automobile industry, lar
surveyed using questionnaires.

An Example fro
Econ

- In telecommunications, there i
  which is designed to bring in la
  selling permission to use certai
  wireless internet) in a given re

Brief course

Goal Understanding of the basic i
statistics, including the underl
(probability and linear algebra
statistics including factor anal
regression analysis will be intr

Prerequisites

Prerequisites Although not abso
students should have taken col
algebra courses.

Language of Instruction I plan
course materials will generally

Manual C

- Calculation by hand will be a
  N.B. "By hand" includes use o
  I can't permit that on examina
- Intended to improve your unde
  computations

Computation

- Computational exercises will a
- Intended to familiarize you wit
  output (*e.g.*, organization of da
  common statistics.

Reso

- Just about anything you need
  the class home page, `http://t`
  `Teach/DataAnalysis/`. If it's
  didn't know (about the assign
  acceptable excuse.

Recommende

- *Statistics*, by David Freedman,
  (*i.e.*, expensive but worth it).
  many examples showing how t

- *Principles and Practice of Stru*
  Kline. An advanced topic, but

What can we learn

Earlier, we mentioned some fields w
finance, education, and so on. Wha
There are three basic kinds of ques

- What happened? or What are

What ha

- We would like to measure quan
- Most "happenings" (we call th
  outcomes at different times, or
  least we can count *frequency* o
  histogram to display frequencie

Types of

- The *type* of a variable determi
observations may be compared
- Variables may be *qualitative* (
equality), *ordinal* (values can b
(values can be combined using

Working wi

- *Cardinal* variables may be trea
  ordinal values ("forgetting" th
  vice versa. That is, it is reason
  cardinal variable by size, but i
  list of values for an ordinal var

Explain

- We use *models* to explain why
  models expressed as equations
  handle other kinds of model.
- There are two aspects to any
  model. Statistics does not tell

Domain mo

- In the domain of economics we
  a *demand function*.
- We sometimes assume that the
  specific form (such as linear or
  statistics easier to calculate. T

Statistical mo

- In a food supply model the un
  contributes positively, sometim
  supply. Statistics can *measure*
  some degree, and the uncertain
- Probability theory shows that

Variable type, mo

- Each variable type has particu
  for its analysis. The statisticia
  appropriate to her problem.
  - Especially watch out for co
    3 to represent the Minshute

Reliability

- If the data are random, can we
  After all, anything could happ
- Yes! We can derive *moments* (
  are quite predictable even if in

Designing Stat

The following notes correspond rou
Pisani, & Purves.

- Yes, Virginia, statistical analys
- We often have a choice of *what*
- the *type* of each variable

Controlled e

- When
  - we have substantial control
    observations and their *quar*
  - exploit that control to achi
    results in different circumst

Observatio

- Though we have less control in
  preferred for
  - *ethical* reasons: experiment
    consent is generally frowned
  - *financial* reasons: it's often

The Salk vacc

- Case study in the ethics of exp
- Background: *polio* is a disease
  killing some, and paralyzing m
  unknown, because of the succe
  the 1950s

The Di

- *Planning* and *execution* of a pr
  and coordination with related
  effectiveness of the treatment i
  - But we didn't know accura
  - Do more testing to find out

Effectiveness o

- Experiments can give more acc
- Baseline: just give vaccine to e
  incidence (rate of infection) fa
  - Problem: incidence varies a

Various difficu

- **Principle:** relationships amon
  be controlled. Ideally, the only
  and control groups is the treat
- *Imbalanced sample: counts* wil
  results toward large group

Addressing th

- *Imbalanced sample:* computing
- *Self-selection:* imposing treatm
- *Confounding:* assign treatment
- *Placebo effect:* use a *placebo* or

Salk vacci

Two studies of effectiveness of the

NFIP The *National Foundation f*
and conducted an experiment
age groups (Grades 1–3) were
treatment group) if in Grade 2

Salk vaccine stu

Incidence of polio: rate per 10

Experts group	
Size	Rate

Salk vaccine stud

- "Refused permission" has a ve
"control" in both studies. This
children in high-income familie
vaccination, and more likely to

Comparison: porta

Number of studies
Design
No controls
Controlled, not randomized

Comparison: rando

Number of studies

Therapy	Ra
	Positiv
Coronary bypass surgery	1

The bias in favor o

- In studies of new medical treat
  biases in favor of the effectiven
- Especially for *surgical* treatme
  than ineligible patients; other
  should be better

Homework 1: due

Submit homework to `data-hw@tu`
email. Note the due date is April
time of receipt by the server.

For this homework, please submit
PDF attachments). In other words

Problems

1. In student evaluations of a cou
   asked is "was the pace of lectu
   (3) too fast?" The students' an
   for a variable named **pace**. Is
   cardinal variable? If cardinal,

3. In the mass media (*e.g.*, newpa
pages, but *not* a statistical tex
example of each type of variab

(a) qualitative

(b) ordinal

(c) discrete cardinal

(d) continuous cardinal

4. In the mass media, find an exa
each of

(a) a controlled experiment

(b) an observational study

Also give the URL or bibliogra
found the example. You may u
or different sources, as conveni

Measurement Proje

18, 11

Submit homework to `data-hw@tu`
email. Note the due date is April
time of receipt by the server.

Tasks

1. Find "something" to observe. career relevance to you, as you throughout this class. It needs complexity, because you will b Describe your topic briefly, and in it

- continuous cardinal

Describe the variables you hav
each one, its unit of observatio
and origin (if relevant to that

Example: For the traffic exa
gender and mode of transporta
motorcycle), and weather cond

types (qualitative, ordinal, disc
use variables defined in part **??**

Example: In the traffic exam
person passing the point betwe
observation. Some variables m
observations, for example the
weather condition will be the s

date will be perfectly correlate

Keep in mind that once your obser
you will be asked to collect a *data*
measuring all the variables you hav
observation. You will also be asked
your data, *i.e.*, the relations you ex

You may augment your data set w