9. Statistical Communication: "Lies, Damned Lies, and Statistics"

Readings: None

1

1) Introduction

- Mark Twain: *there are three kinds of lies:* ***lies, damned lies, and statistics***.
- Wikipedia: "The semi-ironic statement refers to the persuasive power of numbers, and succinctly describes how even accurate statistics can be used to bolster an inaccurate argument through such methods as selectively choosing data, ignoring bad results and over-emphasizing good results."
- Even if a study is scientific, we should be sceptical *but not cynical*.

2

2) Problems of Data Collection/Construction
a) Sample Selection Bias

- Is our sample random and representative?
  - "Most young offenders…":
    - Most young offenders are one-time.
    - Most offences are by chronic offenders.
- Solution?

3

b) Excluding Portions of the Data Set

- "Abortion, Depression link?" → which subset of the data is used.
- Dividing up the data set by age, gender, race, etc. → drug tests.
- Related: small sample sizes.
- Solution?

4

c) Poor Data

- Poorly designed questionnaires, poorly answered questions, measurement errors.
- Solution?

5

d) Index Numbers and Weighting Schemes

- Where do the Quarterback weights come from? The Macleans weights? The HDI weights?
- Solution?

6

e) Lurking Variables

- If variables are omitted from the data, the results will usually be biased.
  - E.g. regressing wages on age, but ignoring schooling.
- Solution?

7

3) Problems With Interpreting the Data
a) Treating Samples Like Populations

- Will the results be repeated with another sample?
- Solution?

8

b) Lack of a Proper Baseline

- Control for
  - inflation
  - population
  - trending variables like income
  - present value.

9

"The value of a degree: a million bucks"

- Is this true?
  - In our dataset:
    High school: $702/week = $36,504 per year.
    Bachelors: $973/week = $50,596 per year.
  - Difference: $14,902/yr x 40 years = $563,680.
- BUT, doesn't discount future earnings.

$$PV(\text{earnings}) = Y_0 + \frac{Y_1}{1+r} + \frac{Y_2}{(1+r)^2} + \ldots + \frac{Y_{40}}{(1+r)^{40}}$$

$$\approx \$295{,}000$$

- Versus cost of university education ≈ $27,000 per year?

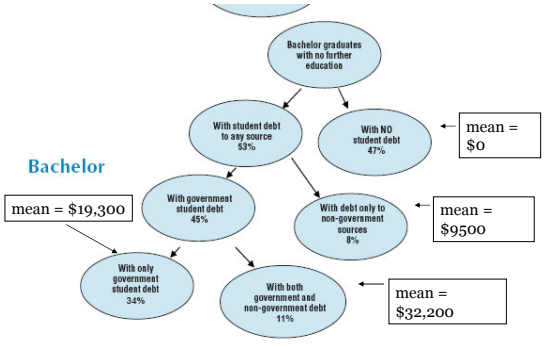
10

c) Focusing on the Mean vs. the Distribution

- Remember our discussions of your summer earnings and student loan data.

11

Student Loans



Source: Statistics Canada, Class of 2000: Profile of postsecondary graduates and student debt

12

d) Correlation is Not Causation

- Stock Markets and:
  - Superbowl winners: "(A) win by a team that played in the old (NFL) would give a lift to stocks."
  - The Democrat effect: "the average gains on the (DJIA) in the 12 months after the Democrats have seized power has been 20 percent. The Republicans … *loss* of 9 percent."
  - The October Effect.
- Solution?

13

e) Statistical and Practical Significance

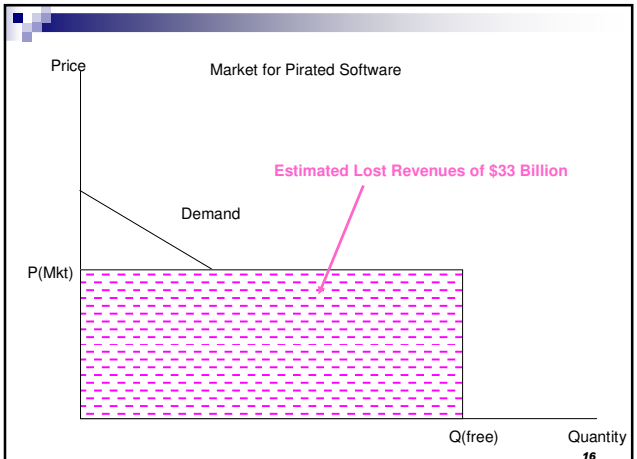
- CJD – is one death statistically or practically significant?
  - If you sample 20 times, at least once (with $\alpha = 0.05$) you will get a meaningless correlation.
    - Meaningless sports stats…
- Note the opposite effect: if sample size too small, may get insignificant result, when it is really significant.
- Even if statistically significant, does it really matter policy-wise?
  - Doubling of 1 death in a million…

14

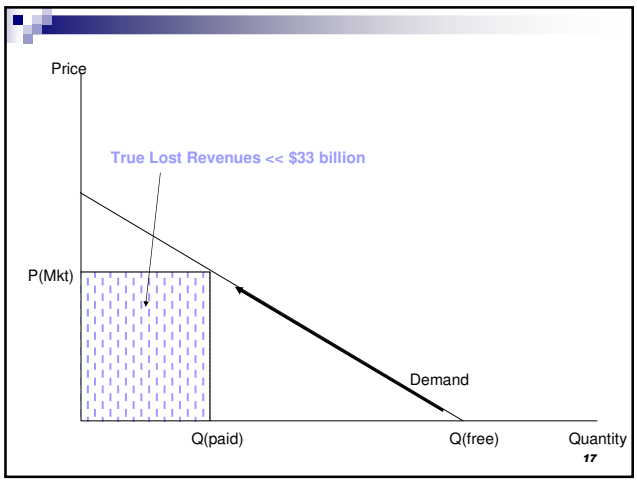
f) Inappropriate Extrapolation

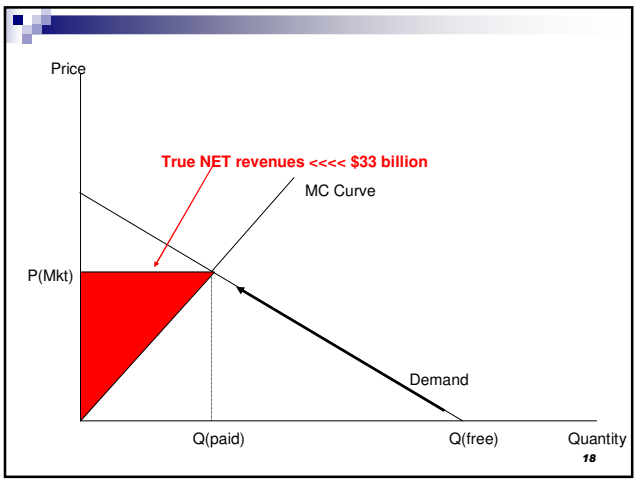
- "BSA or just BS?":
  - Poor sampling.
  - Implicit assumptions about usage.
  - Extrapolating outside the sample ignores the behavioural changes made by rational individuals.

15



16





g) Prior Beliefs Guiding Interpretation

- What do you test?
- What variables do you put in?
- How do you interpret results that go against your hypothesis?
- Nelson: "what often really seems to matter in convincing a male colleague of the existence of sex discrimination is not studies with 10,000 'objective' observations, but rather the particular single direct observation: the experience of his own daughter."

19

h) "Mind Your Language"

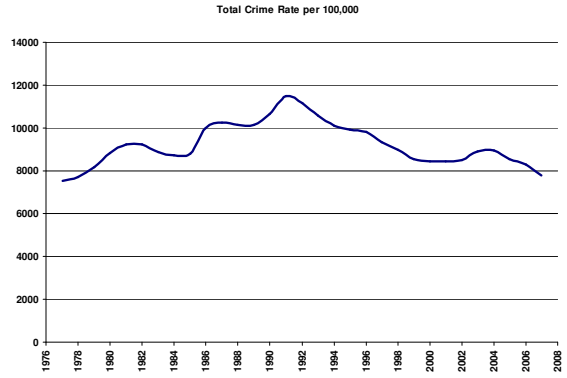
- "Mind your language" → how you word your results can bias how they are perceived by readers/listeners.
- What is wrong with the article "DVD industry prays for holiday upswing"?

20

4) Attitudes/Perceptions Towards Randomness

- There is a huge difference between people's perceptions of randomness and risk and the true outcomes.
  - They interpret the results through the prism of their attitudes.
- Examples:
  - Perceptions of crime vs. true levels.

21



Source: *CANSIM II*

22

Other Examples

- If you have a gun and a swimming pool in your backyard, what is more dangerous to a child in the U.S.?
- Are airplanes or cars more dangerous?
- Are airbags dangerous for kids?
  - Misleading title.
  - Not statistically significant.

23