STAT 231

October 7, 2016

Road map

- 5 min recap of model selection
- PPDAC approach to STAT PROBLEMS (Ch 3).

$\{y_1, \ldots y_n\} \rightarrow$ SAMPLE

↓

Model. $\quad Y_i \sim f(y_i; \theta)$

↓

Test the model (Does the data agree?)

Yes ↓ ESTIMATION

NO :· LOOK FOR ANOTHER MODEL

Graphical → pdf (theoretical and empirical)

cdf (")

Q-Q plot.

Numerical: Compare observed frequencies to expected frequencies

Suggested Model

$$Y_i \sim Poi(\gamma)$$

Data is arranged

# of hits	Obs. Freq. (f_i)
0	2
1	6
2	22
3	40
4	20
5	10
≥ 6	0

Total # of hits

$$\sum_j f_j = 300$$

Av. # of hits

$$\hat{\gamma} = 300/100 = 3$$

Y_i = # of hits on a website in a 15 min interval.

For n days $n = 100$, we collect a sample

$$\{y_1, \ldots y_{100}\}$$

hits

y_i = # of ~~accidents~~ in day i

Objective: To estimate the average number of hits

What is the MLE for μ?

(a) 20

(b) 3 ——————→ 62%

(c) 16.67

(d) 2.5

(e) ~~Cannot be determined.~~
none of the above.

Expected frequency of r hits $= \dfrac{e^{-\hat{\mu}}\hat{\mu}^{r}}{r!} \times n.$

$\therefore$

PPDAC

Step by step algorithm to approach a statistical problem:

P → Problem

P → Plan

D → Data

A: → Analysis

C → Conclusion

Examples

- To figure out what Canadians think of cops (approval rating).

- A random sample from the phone book of K-W. and conduct interviews.

Example 2: ~~A Flue~~ To figure out whether a new flu shot that came out is effecting in reducing flu symptoms amongst Canadia

PROBLEM

- What is your target population?
- units → a member of the Target population
- variate
- attributes → characteristic of the unit

1st problem → Variate : Whether or not they approve of cops

Attribute → function of the variate

Randomly select people from the UW.

Control group → placebo

& Test " → actual flu shot.

Example 3: Suppose we want to find the relationship between Canadian parents smoking and smoking among children.

& A random sample is chosen from K-W and the test conducted.

PLAN:

Question:

(a) What is the study population?

(b) How is the data collected?

(SAMPLING PROTOCOL)

(c) What are the possible studies!

A study population is the population from which your sample is drawn.

In Example 1

Target pop → set of all Canadian

Study pop → people in K-W
with a phone line

Study population need not
be a subset of the target population

but in most cases they are.

Observational
Cofs, Smoking

Experimental
Flu

Estimation

Hypothesis Testing

Prediction:

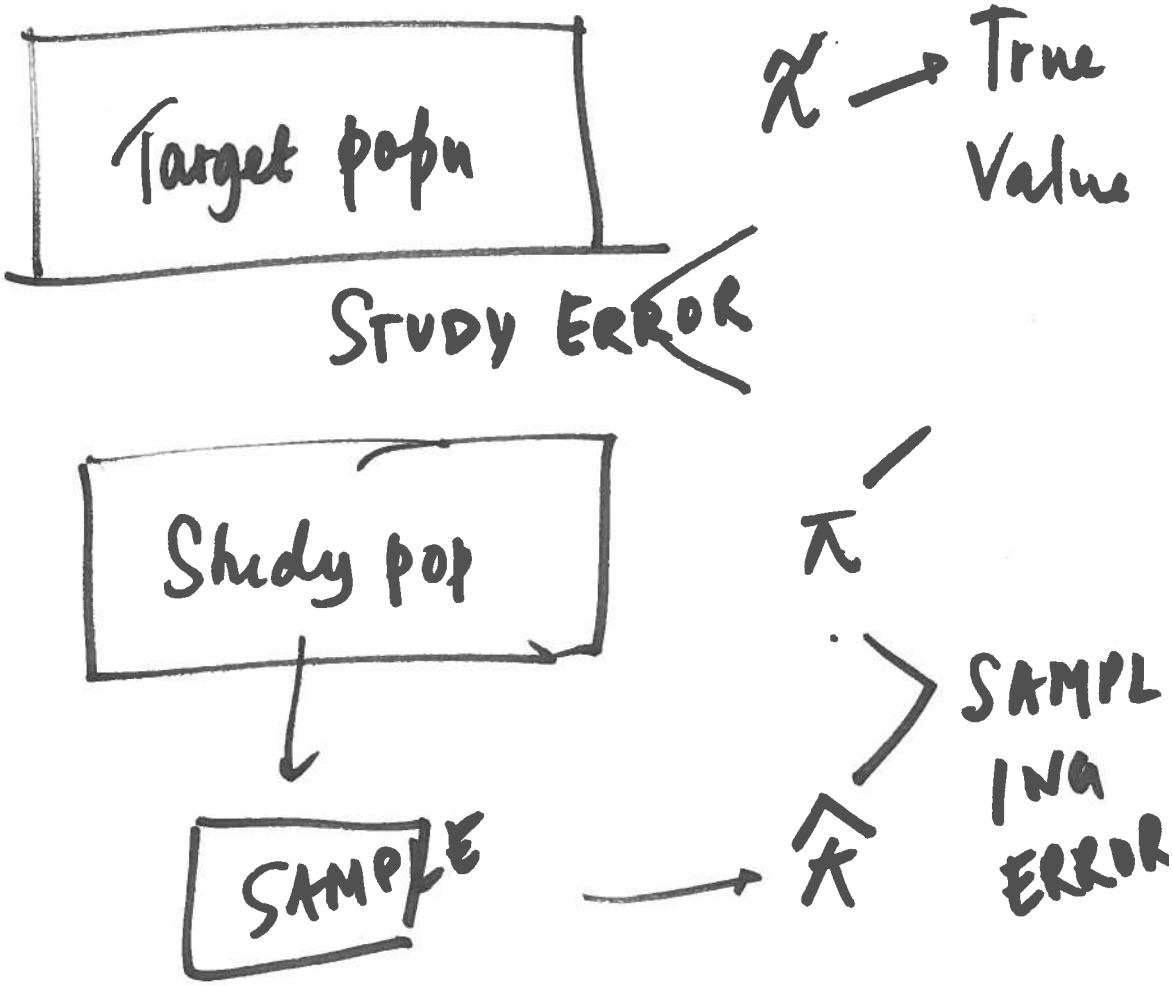
Types of Errors.

(i) Study Error

(ii) Sampling Error

(iii) Measurement Error.

We have to know what these errors mean, and what could be the source of it?



The difference in the value of the attribute between

Target and Study = STUDY ERROR

Study and Sample = SAMPLING ERROR

Measurement Errors are made when we tabulate data.