

Topics - A tale that grew with the telling.

1) Designed experiments observational studies

double blind, crossover, controlled, placebo
known^v sampling, population
probabilities

[confounding
correlation / other dependence
attributing

Simpson's paradox - Berkeley example
- death penalties example
confounding factor

2) Summary statistics

mean, median mode | transformations
SD, range IQR
visuals: histograms etc.
boxplots w/did not

3) Tchebyshev + Typical range good for reading graphs.

4) probabilities density / histogram

$N(\mu, \sigma) \leftrightarrow N(\mu, \sigma)$
 $N(\mu, \sigma)$ example. ~~X~~ = $\mu + Z\sigma$
Forwards & backwards, normal
tables.

5) Regression : str. line

regression to mean
relation to correlation / coeff of det. r^2
bias - not a str. line?
estimate a, b, σ - z

6) correlation - in $[-1, 1]$
- not causation
- read from scatter plots
- linear

ecological correlation (groups)
bivariate data

7) More regression

- cond pop.
- residual analysis - resid plot $\hat{\epsilon}_i$
- prediction
- extrapolation
linearity
const σ
outliers
other patterns
 $i=1 \dots n$

8) Probability - limit definition.
- Bayes rule / beliefs.
- K's axioms
- prob as area
- counting points

conditional prob

9) Conditional Prob

3

- definition, partial info.
- $\cap$, $\cup$, complement, disjoint
- indep. events / mult. rule.
- addition rule.
- w/ & w/out replacement
(counting)
use 1 rule
- permutations / combinations.

10) Binomial (prob)

- r, n, p
- meaning: # successes (no order)
indep. trials

11) Using Bayes rule - false +, false - w/ counting rules

12) contingency tables

- Mosaic plots + how they represent probs & cond probs.
- relative risk
- odds ratio.

13) random sampling

Simple Random sampling (IID) 4
parameter, population, data,
RV, outcome, statistic/estimator
estimate.

$$SE \text{ vs } SD : \begin{matrix} SE = \frac{SD}{\sqrt{n}} \\ \uparrow \\ \text{estimator} \end{matrix} \quad \text{pop.}$$

response bias, sampling bias,
other sources of bias

14) Variance: types of sampling
cluster, quota,
for average vs. sum
"box models"

15) Deep Theorems: Law of Averages $\bar{X} \rightarrow \mu$
Expected value, SE, | gambling,
lottery tickets
CLT !!!

Use of these theorems to
approx. probs. $\rightarrow$ leads to C.I.'s

For averages & sums.

pop. dist. (not normal) $\rightarrow$ Normal
estimators

$\rightarrow$ CLT for binomial $SD = \sqrt{p(1-p)}$

?? FPCF

16) CI'S $L, U = \bar{X} \pm SE Z_{\text{confidence}}$

- in tups - right + wrong.
- use w/ APCF ???
- means + proportions, difference
of means
proportions
- estimate SE.
- use of CLT.

17) Testing

H_0, H_A — what you want to show.

test stat

p-value / OLS / sig. Prob.

Chinese menu of tests.

write out hypothesis in (CI's)
notation + in words

State conclusion in words.
data snooping is bad.

18) T-test — σ unknown

- t dist, deg. of freedom
(shape)
- goes normal

19) Differ of means, proportions (t & normal)

20) Goodness of fit / Indep: χ^2
types of hypotheses for ~~both~~ ^{both} cases.

→ O_i 's, E_i 's, deg. of freedom (2 cases)
shape of χ^2 — problems w/ using it.
both cases for O_i, E_i .

21) paired df tests - effect on SE,
deg. of freedom
- why better to pair

6

22) Bayes - Bayes rule (discrete)
- how to find post data dts
in pple
- prior, conditional density,
posterior
- beliefs & updating them
by the data.
- Bayes testing. vs Freq testing
estimation estimation
- pred. probs.

23) mult. regression.