

1/5

MD = modern drive

Lec 37: 2019/4/26

Set up
projector

Open course webpage to Lec 37 so students can see Table 8.6 + Fig 6.4 + Table 6.2

in MD Table 8.6

Recap of sampling scenarios, we've covered

- ① p bowl of balls
- ② M pennies
- ③ $p_1 - p_2$ resumes ~~submitted~~ @ Bank

Today: ⑤ & ⑥. Regression intercept & slope.

Say
verbally

Read over Fig 6.4, describe verbally relationship between teaching score & "beauty" score.
Then tell them "You've previously seen the corresponding regression table"

Recall estimate column gives you equation of regression line in MD Fig 6.4

General $\hat{y} = b_0 + b_1 x$

Here $\text{score} = b_0 + b_{\text{bty avg}} \times \text{bty avg}$

$$\text{score} = 3,880 + 0.067 \times \text{bty avg}$$

↑
⊕ve relationship

2/5

Q: What about other columns?

A: Relates to sampling / statistical inference.

Recall Sampling framework/terminology from
MD 8.3

(study) Population: Who do we view these $n=463$
as a sample of? All UT instructors, not
just Austin? All UT Austin ~~instructors~~ instructors
in 2019? Who do we want to generalize
our study to? Open-ended questions.

Population parameter $\beta_0 = \text{pop intercept}$
 $\beta_1 = \text{" slope}$

where $y = \beta_0 + \beta_1 x$

Unknown, like ρ & μ .

Sample: Sample of $n=463$ instructors

Point estimate/sample statistic:

Fitted intercept $b_0 = 3.880 \rightarrow$ Estimates β_0
" slope $b_1 = 0.067 \rightarrow$ " β_1

The same way "sample proportion" \hat{p} estimates ρ
" mean \bar{x} " μ

even.

Say this,
wait for
someone to
say standard

Ok folks? What is the precise name of the statistical ~~quantity~~ term that quantifies variation in point estimates due to sampling variation?

37 ~~22~~ Column in Table 6.2: std-error
standard error. Just like to did in Lab 08
where you virtually simulated sampling from
bowl

Important As n goes up,
SE goes down!

4th ~~3rd~~ column statistic

This is the "test statistic" used in hypothesis testing in regression.

Ex: Diff in promotion rates @ bank from

$$\hat{V}_m - \hat{V}_A$$

more next week.

Lec 32

5th

~~Handwritten scribble~~

column

p-value

of the following HT
bottom row.

$$H_0: \beta_0 = 0$$

$\Rightarrow H_1: \beta_0 \neq 0$

4

is interest

0 or ∞

$$H_0: \beta_1 = 0$$

$\therefore H_1: \beta_1 \neq 0$

4

is slope

○ ~ wt?

4/5

both p-values are very small,
reject both null hypotheses &
declare both intercept & slope are
≠ 0.

5th & 6th column \Rightarrow [Duh!!! Look @ Fig 6.4. You don't
need no PhD in statistics to know
the intercept & slope aren't 0.]

In particular, we say slope ~~is~~ isn't 0,
it's positive, i.e. there is a statistically
significant relationship between
teaching score & beauty score.

6th & 7th column 95% confidence interval.

In particular 95% CI for population
slope β_1

is $[0.035, 0.099]$

↑

Does not contain 0!!!

Suggesting there is a meaningful \oplus ve
relationship between teaching & ~~teach~~ beauty score.
beauty

5/5

Q: How does R compute these p-values & conf ints? Does it do resampling like we've done in class.

A: No! There are actually two ways to generate CIs & p-values.

Using Simulation/Resampling & Using Formula

Say this
verbally

We've been doing this
to focus on conceptual
understanding.

talk about Galton?

We'll do
this next
week in
practice
midterm III

Also, we'll see on Monday

- Checking conditions for inference via residual analysis.