1/9

1. EDA via Data Wrangling

Done in Practice Midterm I

2. EDA via Visualizations

Done in Practice Midterm I

3. Regression model using priming number

a) Mean # of countries guessed for baseline group: those primed with "14 countries" (as opposed to "94 countries")

b) Difference in mean # of countries guessed for 94 group relative to 14 group.

ie. on avg they guessed + 34.7 countries more

c) intercept
priming 14 countries
priming 94 countries

2/3

d) XXX = 29.5
YYY = 29.5 + 34.7 = 64.2

e) XXX = 29.5 YYY = 36 - 29.5 = 6.5
AAA = 64.2 BBB = 120 - 64.2 = 55.8

IMPORTANT FOR FINAL PROJECT

f) No wrong answer, this is your opinion.
However, look @ 95% confidence interval for priming 94 countries in regression table two pages back (page 6). It is

[20.2, 49.2] countries

&

DOES NOT CONTAIN

0 countries.

ie. Suggestive of meaningful difference!

3/9

4. Regression model using height

** Done in Practice Midterm I.

IMPORTANT FOR FINAL PROJECT

but look @ f):

Is slope for height = -1.86 "significantly different" than 0?

Look @ regression table two pages back (page 9)

The 95% CI for the slope for height is

$$[-4.57, \quad 0.854]$$

includes 0!

If slope = 0, suggestive of NO Relationship between height and # of cantrips guessed.

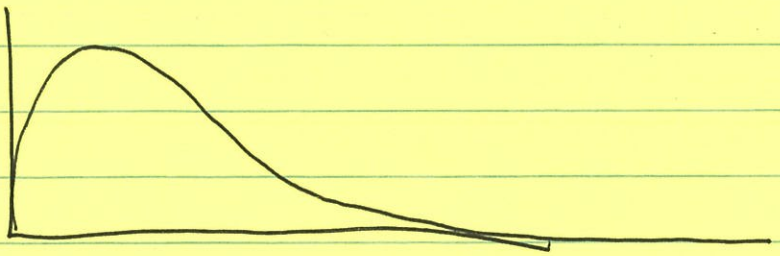
DUH!!! Why would these two variables be related?

4/9

Much harder than you can expect

1. Seattle House Prices

a) b/c these variables were right-skewed



b) $\log_{10}(\text{price}) = 3.33 + 0.69 \times \log_{10}(\text{size})$

c) $\log_{10}(\text{price}) = 3.33 + (-0.883) + (0.69 + 0.31)\log_{10}(\text{size})$

$= 2.447 + 1 \times \log_{10}(\text{size})$

d) $2.447 + \log_{10}(1000) = 2.447 + 3 = 5.447 = \log_{10}(\text{price})$

thus $\widehat{\text{price}} = 10^{5.447} = \$279,898$

5/9

Parallel slopes model

e) intercept
log10_size ← single slope
condition 2
" 3
" 4
" 5

2. Sampling scenarios

a) See next page

b) Yes, blc it is random sampling, thus the sample is representative of bowl, so estimate is good!

c) While the sample size is large, are these 1000 pennies REPRESENTATIVE of all US pennies in circulation? Maybe? Are they newer?

d) Again, are these 38 220 students REPRESENTATIVE of ALL Smithies?

6/6

Scenario	1	2	3
Population	$N = ?$ 2400	$N = ?$ Who knows?	$N = ?$ About 2400 Smithies
Population parameter name	? population proportion	? population mean	Population slope
Population parameter mathematical notation	? p	? μ	β_1
Sample size	$n = ?$ 50	$n = ?$ 1000	$n = ?$ 38
Point estimate name	? sample proportion	? sample mean	Fitted slope
Point estimate mathematical notation	? $\hat{p}$	? $\bar{x}$	b_1
Point estimate numerical value	? $\frac{20}{50}$	? 2013.56	? 34.7

7/9

2. d) Continued. Note this involves asking yourself about BOTH random assignment AND random sampling.

- Random sampling: Is sample of 38 220 students representative of population of all Smithies.

- Random assignment: If we randomly assign treatment & control, can we make a causal statement?

3. Short Answer

a) - Clinical Trial
- Randomized Controlled trial
- Randomized experiment that uses blocking.

- Note A/B test is used in the context of internet testing.

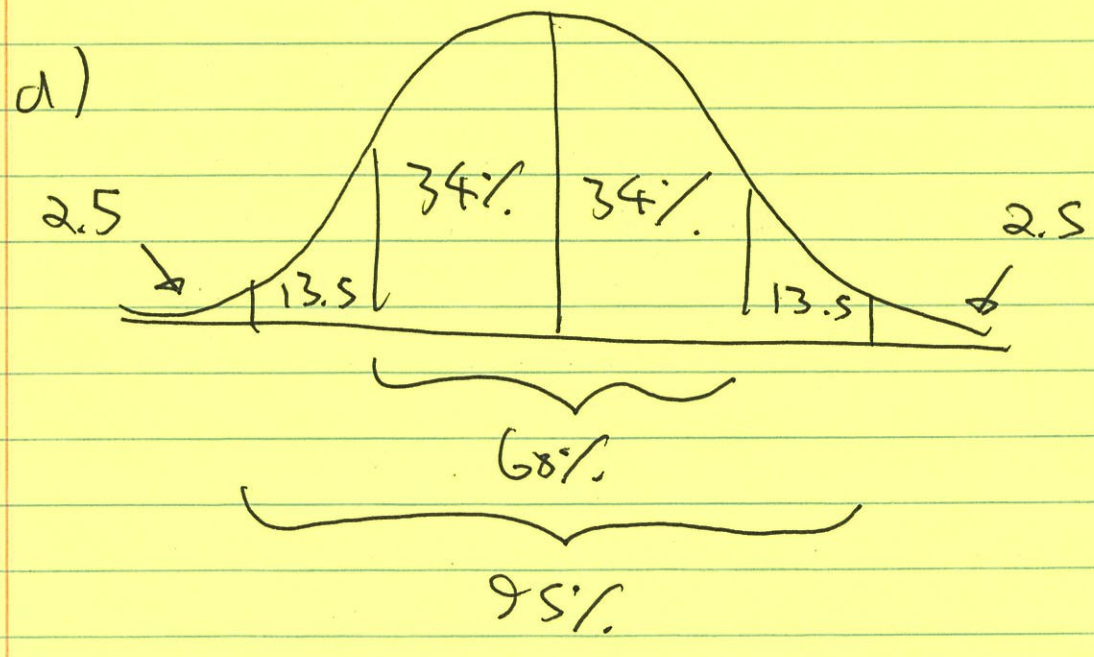
b) While Kinsey has a large sample size n

8/9

3.b) continued.
his sample is probably NOT representative of all US males. Thus his conclusion probably does NOT generalize to all US males.

c) b/c researchers did not explicitly assign "treatment" vs "control" variable. ie. who slept with shoes on ~~was~~ vs off.

d)



9/9

4. Sampling distn

a) SKIP

b) Zt will get narrower
i.e. SE will go down

c) Standard Error