Inference in SLR

11/11/24

Housekeeping

- Last set of homework problems are released today!
- Office hours changed this week:
  - Today: 2-3pm only
  - Wednesday 4-5pm
  - Friday: cancelled, moved to next week before midterm

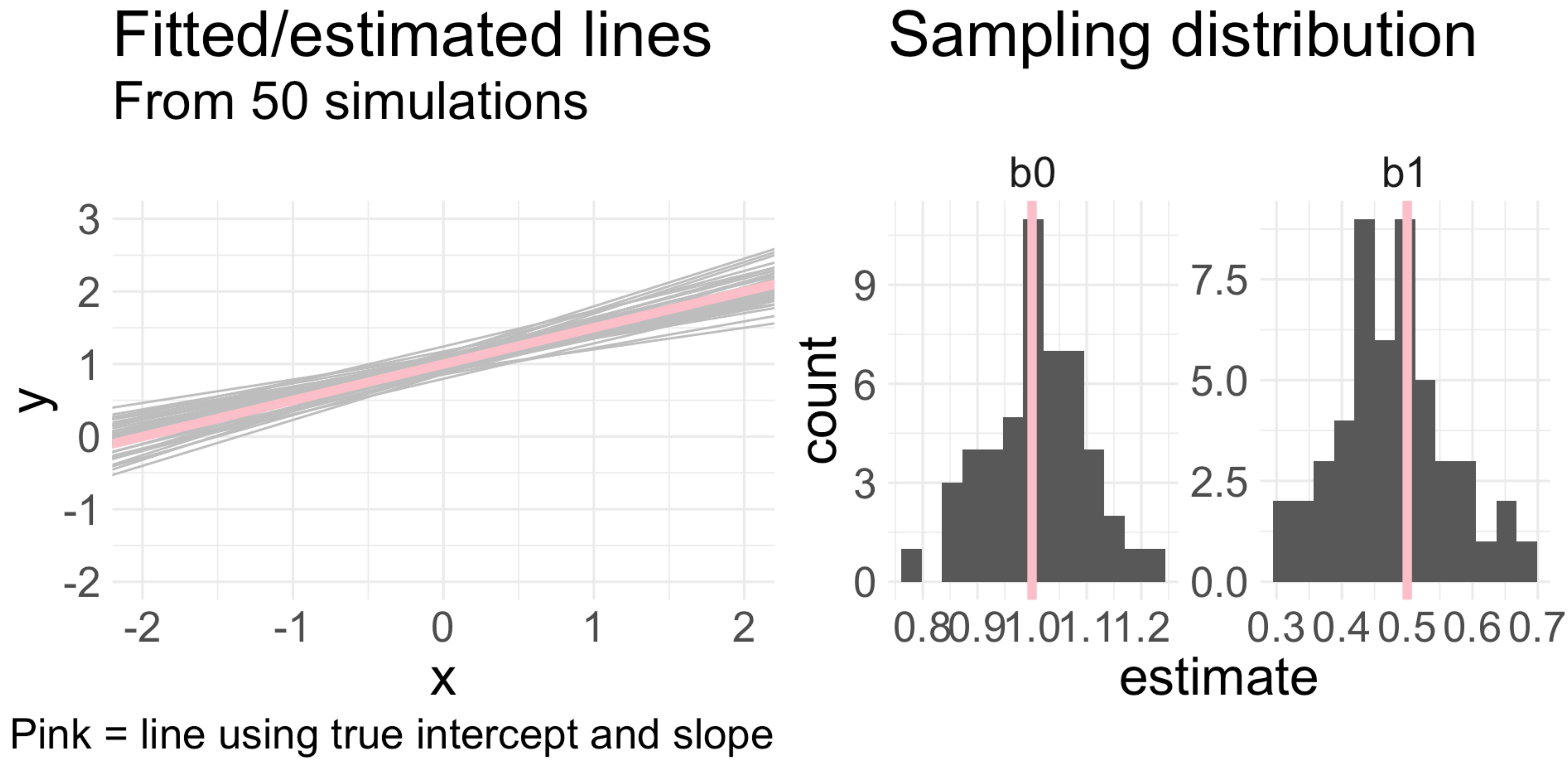
Recap

- Learned how to interpret slope and intercept of fitted model
  - b_0 is expected value of response when $x = 0$
  - b_1 is expected change in y for a one unit increase in x
- When explanatory x is categorical, we have a slightly more nuanced interpretation
- Coefficient of determination R^2 assesses strength of linear model fit

Variability of coefficient estimates

- Remember, a linear regression is fit using a sample of data
- Different samples from the same population will yield *different* point estimates of (b_0, b_1)
  - I will generate 30 data points under the following model: $y = 1 + 0.5x + \epsilon$
    - How? Randomly generate some x and ϵ values and then plug into model to get corresponding y
  - Fit SLR to these (x, y) data, and obtain estimates (b_0, b_1)
  - Repeat this 50 times

Variability of coefficient estimates



What are we interested in?

Remember: we fit SLR to understand how x is (linearly) related to y:

$$y = \beta_0 + \beta_1 x + \epsilon$$

- What would a value of $\beta_1 = 0$ mean?
  - If $\beta_1 = 0$, then the effect of x disappears and there is in fact no linear relationship between x and y
- We don't know β_1, so let's look at estimate b_1:
  - Is an estimate of $b_1 = 0$ a convincing evidence of no relationship? What if $b_1 = 0.1$?
  - It depends! We saw that b_1 varies by sample! So let's perform inference for β_1

Inference for SLR

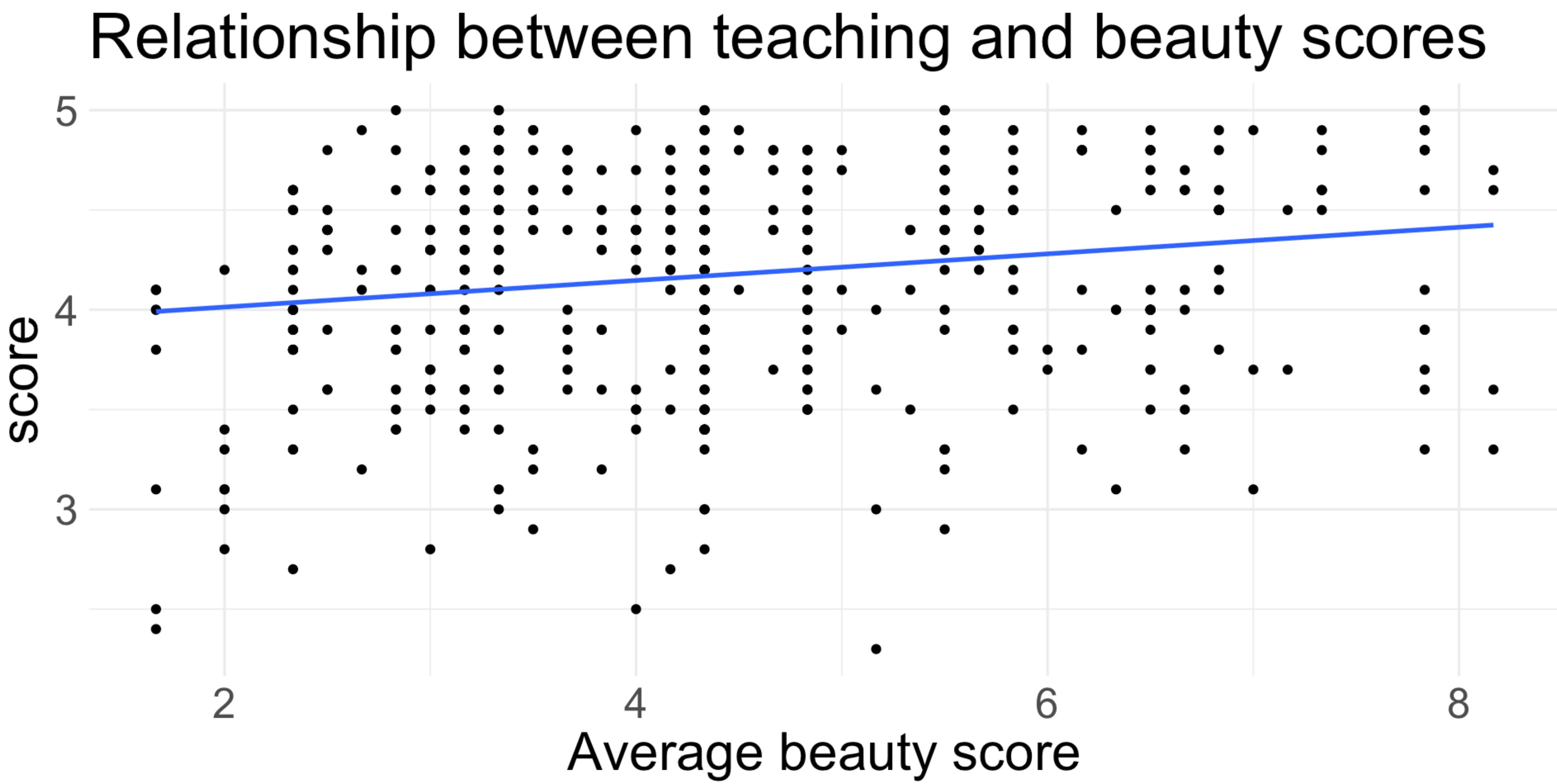
Using mathematical model, not simulation!

Running example: `evals` data

Data on 463 courses at UT Austin were obtained to answer the question: "What factors explain differences in instructor teaching evaluation scores?"

- One hypothesis was that more attractive instructors receive better teaching evaluations
- We will look at the variables:
  - `score`: course instructor's average teaching score, where average is calculated from all students in that course. Scores ranged from 1-5, with 1 being lowest.
  - `bty_avg`: course instructor's average "beauty" score, where average is calculated from six student evaluations of "beauty". Scores ranged from 1-10, with 1 being lowest.

Teaching evaluations data



Does this line really have a non-zero slope?

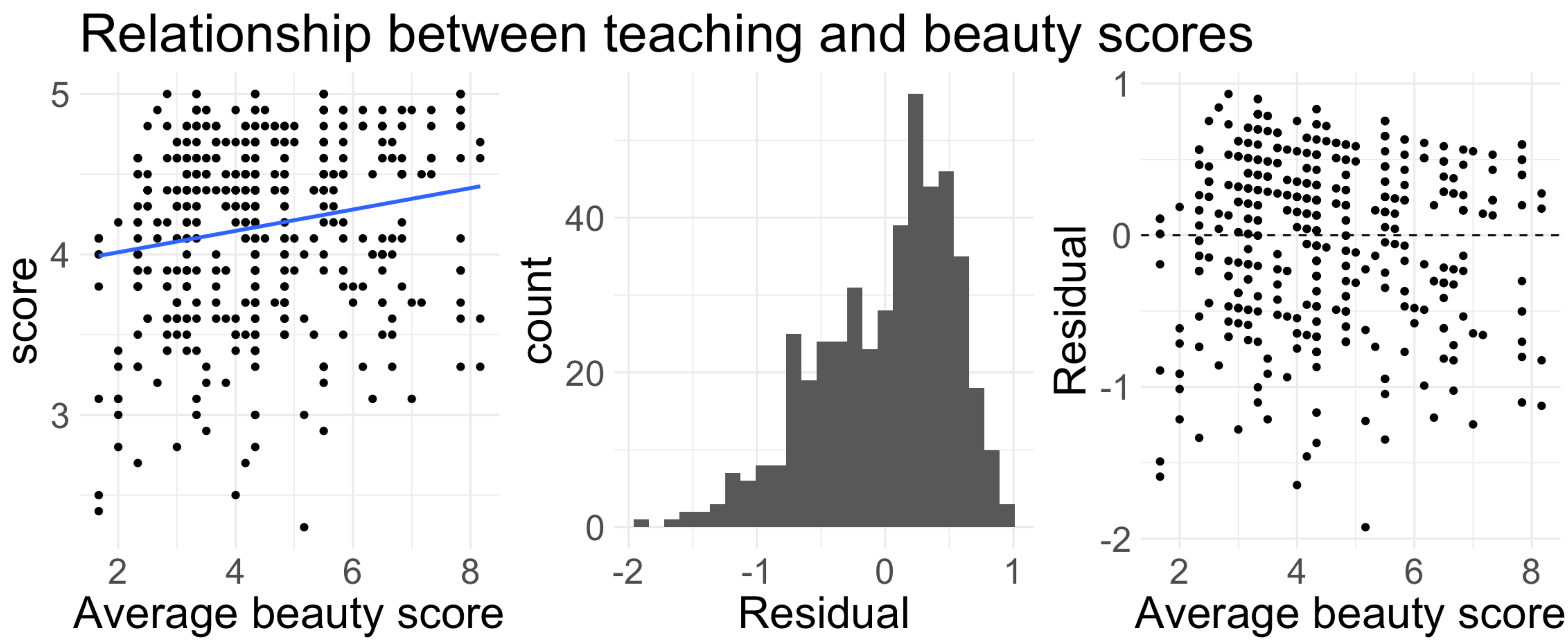
Hypothesis test for slope

- We have the following hypotheses:
  - $H_0 : \beta_1 = 0$: the true linear model has slope zero
  - $H_A : \beta_1 \neq 0$: the true linear model has a non-zero slope. An instructor's average beauty score is predictive of their average teaching evaluation score.
- To assess, we do what we usually do:
  - Check if methods are appropriate
  - If so: obtain an estimate, identify/estimate standard error of the estimate, find an appropriate test statistic, and calculate p-value
- The output from `lm()` actually does all of this for us, but we will see how the test statistic and p-value are calculated!

Teaching evaluations: model assessment

We fit the model in R, and obtain the following plots.

Are all conditions of LINE met?



Looking at `lm()` output

```r
library(broom)
eval_mod <- lm(score ~ bty_avg, data = evals)
eval_mod |>
  tidy()
```

term	estimate	std.error	statistic	p.value
(Intercept)	3.880338	0.0761430	50.961213	0.00e+00
bty_avg	0.066637	0.0162912	4.090382	5.08e-05

Assuming the linear model is appropriate, interpret the coefficients!

- Intercept: an instructor with an average beauty score of 0 would be expected to have an average evaluation score of 3.88
- Slope: for every one point increase in average beauty score an instructor receives, their evaluation score is expected to incrase by 0.067 points

Looking at `lm()` output

term	estimate	std.error	statistic	p.value
(Intercept)	3.880338	0.0761430	50.961213	0.00e+00
bty_avg	0.066637	0.0162912	4.090382	5.08e-05

- `estimate`: the observed point estimate (b_0 or b_1)
- `std.error`: the estimated standard error of the estimate
- `statistic`: the value of the test statistic
- `p.value`: p-value associated with the *two-sided alternative* $H_A : \beta_1 \neq 0$

- Let's confirm the test statistic calculation:

$$t = \frac{\text{observed} - \text{null}}{\text{SE}_0} = \frac{b_{1,obs} - \beta_{1,0}}{\widehat{\text{SE}}_0} = \frac{0.066637 - 0}{0.0162912} = 4.0903823 \sim t_{df}$$

where $df = n - 2$

p-value and conclusion

term	estimate	std.error	statistic	p.value
(Intercept)	3.880338	0.0761430	50.961213	0.00e+00
bty_avg	0.066637	0.0162912	4.090382	5.08e-05

Let's confirm the p-value calculation:

$$\text{p-value} = \Pr(T \geq 4.09) + \Pr(T \leq -4.09)$$

where $T \sim t_{461}$

- So our p-value is: `2 * (1 - pt(4.09, 461))` = 5.0827307^{-5}
- **Assuming the LINE conditions are met**: since our p-value 5.0827307^{-5} is extremely small, we would reject H_0 at any reasonable significant level. Thus, the data provide convincing evidence that there is a linear relationship between instructor's beauty score and evaluation score.

Different H_A

term	estimate	std.error	statistic	p.value
(Intercept)	3.880338	0.0761430	50.961213	0.00e+00
bty_avg	0.066637	0.0162912	4.090382	5.08e-05

What would your p-value be if your alternative was $H_A : \beta_1 > 0$? What would your conclusion be?

- $\Pr(T \geq 4.09)$ = `(1-pt(4.09, 461)` = 2.5413653^{-5}
- The data provide convincing evidence that there is a *positive* relationship between instructor's beauty score and evaluation score.

What would your p-value be if your alternative was $H_A : \beta_1 < 0$? What would your conclusion be?

- $\Pr(T \leq 4.09)$ = `pt(4.09, 461)` = 0.9999745
- The data do *not* provide convincing evidence that there is a *negative* relationship between instructor's beauty score and evaluation score.

Confidence intervals

term	estimate	std.error	statistic	p.value
(Intercept)	3.880338	0.0761430	50.961213	0.00e+00
bty_avg	0.066637	0.0162912	4.090382	5.08e-05

We can also construct confidence intervals using the output from `lm()`! Remember:

$$\text{CI} = \text{point est.} \pm \text{critical value} \times \text{SE}$$

- Critical value also comes from t_{n-2} distribution
- Suppose we want a 95% confidence intervals for β_1:
  - What code would you use to obtain critical value?
  - `qt(0.975, 461)` = 1.97
- So our 95% CI for β_1 is: $0.067 \pm 1.97 \times 0.016 = (0.035, 0.099)$

Remarks

- Note: for β_1, the null hypothesis is **always** of the form $H_0 : \beta_1 = 0$
- LINE conditions must be met for underlying mathematical and probability theory to hold here! If not met, simulation-based methods would be a better choice
- Here, the Independence conditions did not seem to be met
  - Take STAT 412 or other course to learn how to incorporate dependencies between observations!
- **So what can we say**?
  - The results suggested by our inference should be viewed as preliminary, and not conclusive
  - Further investigation is certainly warranted!
  - Checking LINE can be very subjective, but that's how real-world analysis will be!