STAT151A Quiz 5

Please write your full name and email address here:

Also, please put your intials on each page in case the pages get separated.

You have 30 minutes for this quiz.

There are two questions, each weighted equally..

There are extra pages at the end if you need more space for solutions.

Question 1

Suppose we have a randomized controlled trial in which 100 patients are randomly chosen to either receive medication or not. After a period of time their self-reported well-being is recorded, as well as some additional covariates.

For this question, assume that we have an `R` dataframe `df` with the following columns:

- `df$wellbeing`: Self-reported feeling of well-being after taking the medication
- `df$medicine`: A binary (one–hot) encoding of whether medication was administered
- `df$age`: The patient's age in years
- `df$health`: A continuous variable summarizing the patient's pre–treatment health status

Suppose we run the regression

```
reg <- lm(wellbeing ~ medicine + age + health, df)
ci_medicine <- confint(reg, "medicineTRUE", level=0.95)
betahat_medicine <- coefficients(reg)["medicineTRUE"]
```

Each question peforms a calculation in `R` and draws an *invalid conclusion*. For each question, state in a single sentence why the conculsion is invalid using concepts from class. You do not need to mathematically prove your statement; you may refer to proofs from lecture.

Note that *each question is separate*, so a stated observation in one question does not necessarily apply in the other question.

Here is an example of the expected level of detail in the response:

(Example): Suppose we find that `betahat_medicine` is equal to `-0.1873705`.

Invalid conclusion: We conclude that the medicine improves well–being by -0.1873705. **Why is the conclusion invalid?**

Answer: Even if the regression is correctly specified, the regression estimate has some uncertainty, and is not necessarily equal to the the true expected effect of medicine.

(a) Suppose we find that `ci_medicine` takes the value `c(-0.618182 0.243441)`.

Invalid conclusion: Because `ci_medicine` contains zero, we conclude that the medicine has no effect. **Why is the conclusion invalid?**

(b) Suppose we plot the fitted residuals and they are highly right–skewed.

Invalid conclusion: Because the residuals are non–normal, `ci_medicine` cannot be used for valid hypothesis testing. **Why is the conclusion invalid?**

(c) Suppose we plot the fitted residuals and find that they are approximately normally distributed. However, their variance is much higher for low valuse of `health` than for high values.

Invalid conclusion: Because the residuals are approximately normal, `ci_medicine` can be used for valid hypothesis testing. **Why is the conclusion invalid?**

Question 2

Let $s \sim \chi^2_K$ and $z \sim \mathcal{N}(0,1)$, independently of one another, so that $t = \frac{z}{\sqrt{s/K}} \sim \text{StudentT}_K$.

(a) Use the LLN to argue that, for large K, it is approximately true that $t \sim \mathcal{N}(0,1)$.

(b) Suppose that $x \sim \mathcal{N}(0,4)$ independently of s. Show that

$$\frac{x/2}{\sqrt{s/K}} \sim \text{StudentT}_K.$$

(c) Suppose that $r_k \overset{\text{IID}}{\sim} \mathcal{N}(0,9)$, independently of z, for $k = 1, \ldots, K$. Show that

$$\frac{z}{\sqrt{\sum_{k=1}^{K} r_k^2/(9K)}} \sim \text{StudentT}_K.$$

Extra space for answers (indicate clearly which problem you are working on)

Extra space for answers (indicate clearly which problem you are working on)