Data Science

Midterm Review

Before we start...

Before we start...

1. Our mod of the day.

Before we start...

1. Our mod of the day.
2. Project 2

Before we start...

1. Our mod of the day.
2. Project 2
3. Equation(s) you were promised

Before we start...

1. Our mod of the day.
2. Project 2
3. Equation(s) you were promised
4. How this lecture is going to work.

Our moderator

Our moderator

1. Laura!

Project 2

Project 2

1. Well done, submission rate was much higher than Project 1.

An equation you were promised:

This was written wrong:

$$(1 + \frac{\lambda}{N})^{-2}$$

An equation you were promised:

Should have been:

$$\sqrt{1 + \frac{\lambda}{N}}$$

An equation you were promised:

Let's say you had a variable where half the data was missing ($\lambda = 0.5$) and you used $N = 5$ for the number of generated data sets:

$$\sqrt{1 + \frac{0.5}{5}} = 1.049$$

An equation you were promised:

How much better would it be if you used an 'infinite' number of generated data sets?

$$\sqrt{1 + \frac{0.5}{5}} = 1.0$$

To a notebook!

Another equation you were promised:

Pooled Slope Estimate: One way to ‘average’ a pooled analysis (multiple imputation)

Another equation you were promised:

Pooled Slope Estimate: One way to 'average' a pooled analysis (multiple imputation)

1. Calculate the linear regression for each imputed data-set

Another equation you were promised:

Pooled Slope Estimate: One way to 'average' a pooled analysis (multiple imputation)

1. Calculate the linear regression for each imputed data-set
2. Each linear regression has its own slope: β_{1i}

Another equation you were promised:

Pooled Slope Estimate: One way to 'average' a pooled analysis (multiple imputation)

1. Calculate the linear regression for each imputed data-set
2. Each linear regression has its own slope: β_{1i}
3. Just average them!

Another equation you were promised:

Pooled Slope Estimate: One way to 'average' a pooled analysis (multiple imputation)

1. Calculate the linear regression for each imputed data-set
2. Each linear regression has its own slope: β_{1i}
3. Just average them!

$$\beta_{1p} = \frac{\beta_{11} + \cdots + \beta_{1n}}{n}$$

Another equation you were promised:

That was pretty naïve, let's try something more sophisticated:

Another equation you were promised:

That was pretty naïve, let's try something more sophisticated:

1. Calculate the linear regression for each imputed data-set

Another equation you were promised:

That was pretty naïve, let's try something more sophisticated:

1. Calculate the linear regression for each imputed data-set
2. Each linear regression has its own slope: β_{1i}

Another equation you were promised:

That was pretty naïve, let's try something more sophisticated:

1. Calculate the linear regression for each imputed data-set
2. Each linear regression has its own slope: β_{1i}
3. Each linear regression has its own standard error: Z_i

Another equation you were promised:

That was pretty naïve, let's try something more sophisticated:

1. Calculate the linear regression for each imputed data-set
2. Each linear regression has its own slope: β_{1i}
3. Each linear regression has its own standard error: Z_i
4. Do the following fancy weighted average:

Another equation you were promised:

That was pretty naïve, let's try something more sophisticated:

1. Calculate the linear regression for each imputed data-set
2. Each linear regression has its own slope: β_{1i}
3. Each linear regression has its own standard error: Z_i
4. Do the following fancy weighted average:

$$V_s = \frac{\sum Z_i}{n} + (1 + n^{-1}) * \frac{1}{n-1} * \sum (\beta_{1i} - \beta_{1p})^2$$

How this lecture is going to work

How this lecture is going to work

1. I will be clear about what will not be on the exam

How this lecture is going to work

1. I will be clear about what will not be on the exam
2. I will explain how the midterm will be administered.

How this lecture is going to work

1. I will be clear about what will not be on the exam
2. I will explain how the midterm will be administered.
3. We will discuss expectations

Part I: What's not on the exam

NLTK

NLTK

1. We didn't cover NLTK in the lectures, so don't worry about it.

NLTK

1. We didn't cover NLTK in the lectures, so don't worry about it.
2. Everything else (lectures and readings and projects) is fair game.

NLTK

1. We didn't cover NLTK in the lectures, so don't worry about it.
2. Everything else (lectures and readings and projects) is fair game.
3. Aside: It has been brought to my attention that NLTK is not on the course docker images: I will fix this.

git

`git`

1. You should still learn it.

docker

docker

1. I'm not evil.

Part II: Administering the Exam

Taking the exam

Taking the exam

1. At 00:00 EDT on March 31st the exam will go live on ELMS

Taking the exam

1. At 00:00 EDT on March 31st the exam will go live on ELMS
2. At 23:59 EDT on March 31st the exam will close on ELMS

Taking the exam

1. At 00:00 EDT on March 31st the exam will go live on ELMS
2. At 23:59 EDT on March 31st the exam will close on ELMS
3. We will only answer clarifying questions that are asked in private

Taking the exam

1. At 00:00 EDT on March 31st the exam will go live on ELMS
2. At 23:59 EDT on March 31st the exam will close on ELMS
3. We will only answer clarifying questions that are asked in private
4. This is the only thing this semester where public discussion is forbidden.

Taking the exam

1. At 00:00 EDT on March 31st the exam will go live on ELMS
2. At 23:59 EDT on March 31st the exam will close on ELMS
3. We will only answer clarifying questions that are asked in private
4. This is the only thing this semester where public discussion is forbidden.
5. Please respect these rules.

Part III: Great Expectations

What to expect

What to expect

1. Exam is designed for approx. 90 minutes

What to expect

1. Exam is designed for approx. 90 minutes
2. Feel free to typeset, our eyes will thank you

What to expect

1. Exam is designed for approx. 90 minutes
2. Feel free to typeset, our eyes will thank you
3. Explain. your. process.

What to expect

1. Exam is designed for approx. 90 minutes
2. Feel free to typeset, our eyes will thank you
3. Explain. your. process.
4. Nothing will be accepted after 23:59 EDT on March 31$^{\text{st}}$, 2021 CE (AD).

Thanks for your time!

:)