Customer Analytics

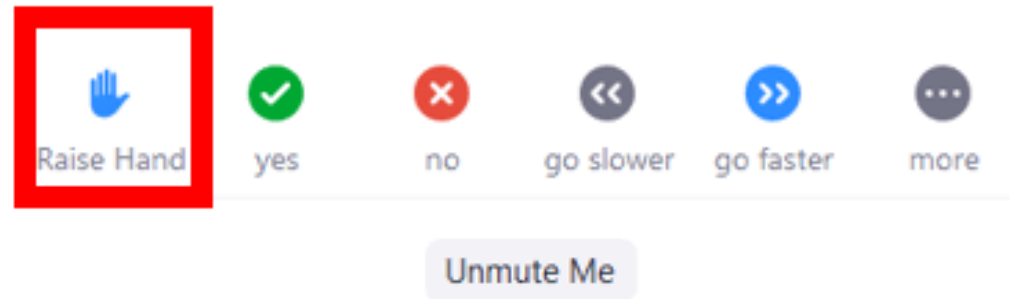Computer Lab Session – Uncertainty
October 28, 2021

Agenda

1. Kick-Off
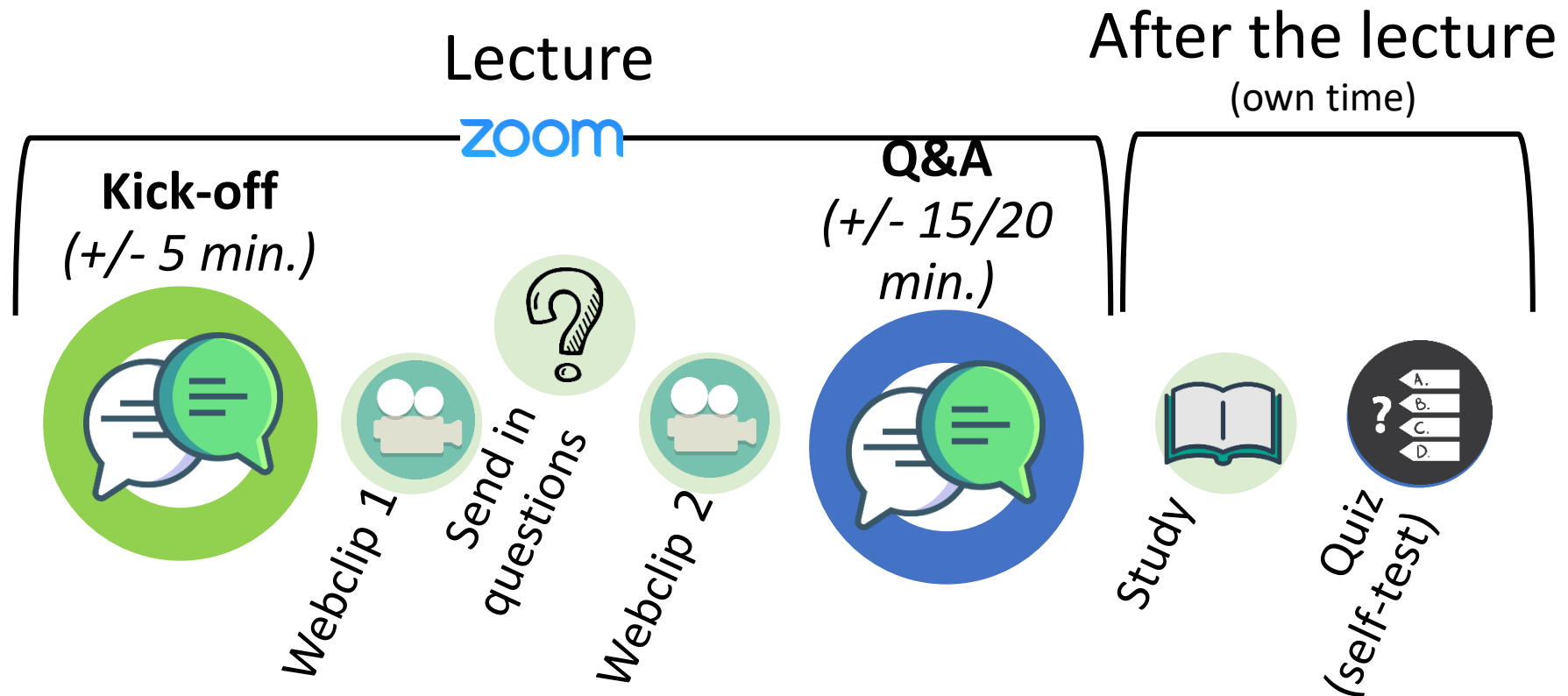2. Webclips & Questions
3. Next week

1. Kick-Off

House Rules

- Keep your microphone muted when you don't speak

- Keep your camera on for this lecture

- If you would like to speak, raise your hand with the Raise Hand function and wait to be called on

- Please use the chat to ask questions throughout the class

Structure lectures

Lecture

After the lecture
(own time)

zoom

Kick-off
(+/- 5 min.)

?

Q&A
(+/- 15/20 min.)

Webclip 1

Send in questions

Webclip 2

Study

Quiz (self-test)

Individual: watch the webclips and send in questions via email
(+/- 60/70 mins)

Data sets & software

- The course is organized around <u>several data sets</u> that illustrate an important concept.
  - All these examples will be "<u>hands-on</u>" and have an emphasis on real-time problem solving.

- We're using **R** (Version 4.1.1 – "Kick Things")
  - Advantages: widely used & lots of contributed software, free
  - Disadvantages: programming language, unpredictability of packages, updates

- R notebooks in the computer lab

Today

1. We assume that you have all watched the Webclips of Module 1

2. Testing and Uncertainty: Why test? Quantifying uncertainty; how large should the test be?

3. Go ahead and watch the **Webclips '1.1 – 1.5'** on Canvas. You get the most out of this if you actively work along with the webclip! So make sure you do so.

4. If you have any questions while watching or after, you can send these to us via e-mail customeranalytics@tilburguniversity.edu

5. We will cover these questions during next week's Q&A session. From next week onwards we will cover them in the same session.

Since George is out of office this week, we will **not** discuss any (theoretical) questions today. These will be covered during next week's Q&A session. However, I will stick around for a bit to help with any practical questions you might have (regarding R, course material etc.)!

Questions?

Good luck with the lab session!
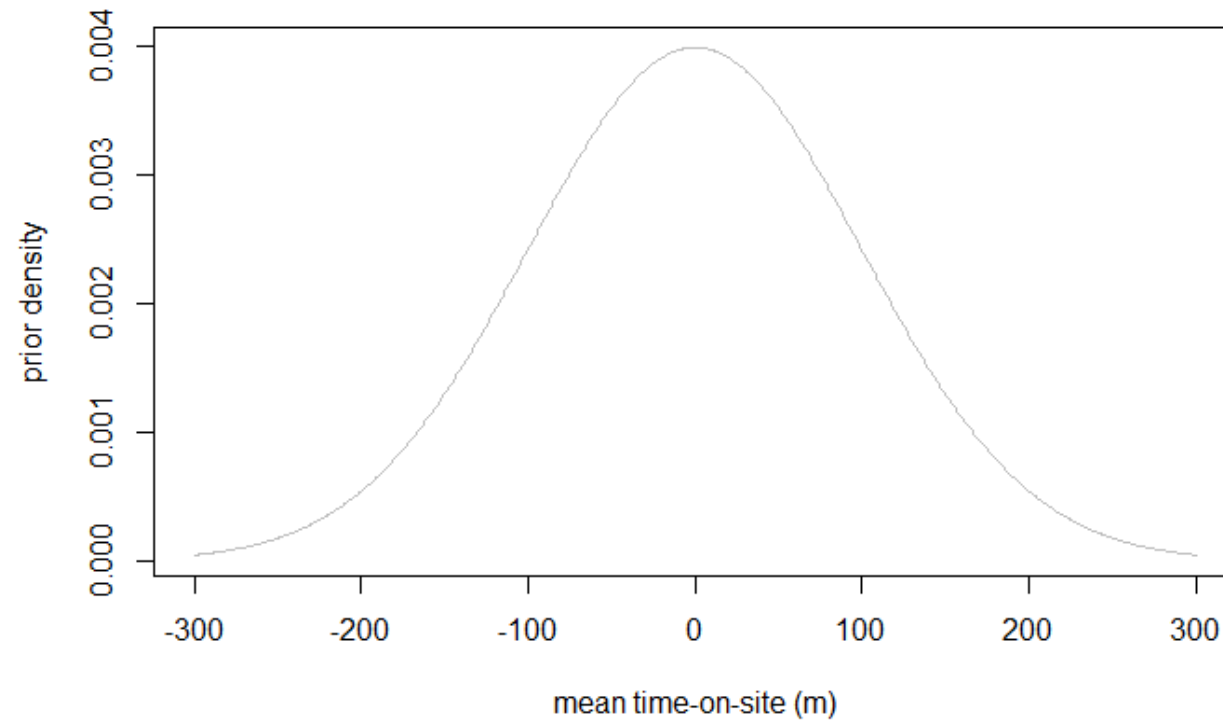
3. Comprehension Check

1. Do you see how respmail has a lot of NA's? What's going on there? Does it make sense?

- *Yes, it makes sense. The variable indicates whether a customer has responded to a mail (=1) or not (=0), but out of the total customer set (9964) only 4952 were actually mailed (test set). The other 5012 did not receive a mail, so were also not able to respond or not. Which means their data is NA.*

2. How can you change the distribution to be a flat, noninformative prior?

Flat or diffuse prior

$$(a = 1, b = 1)$$

prior density

response rate
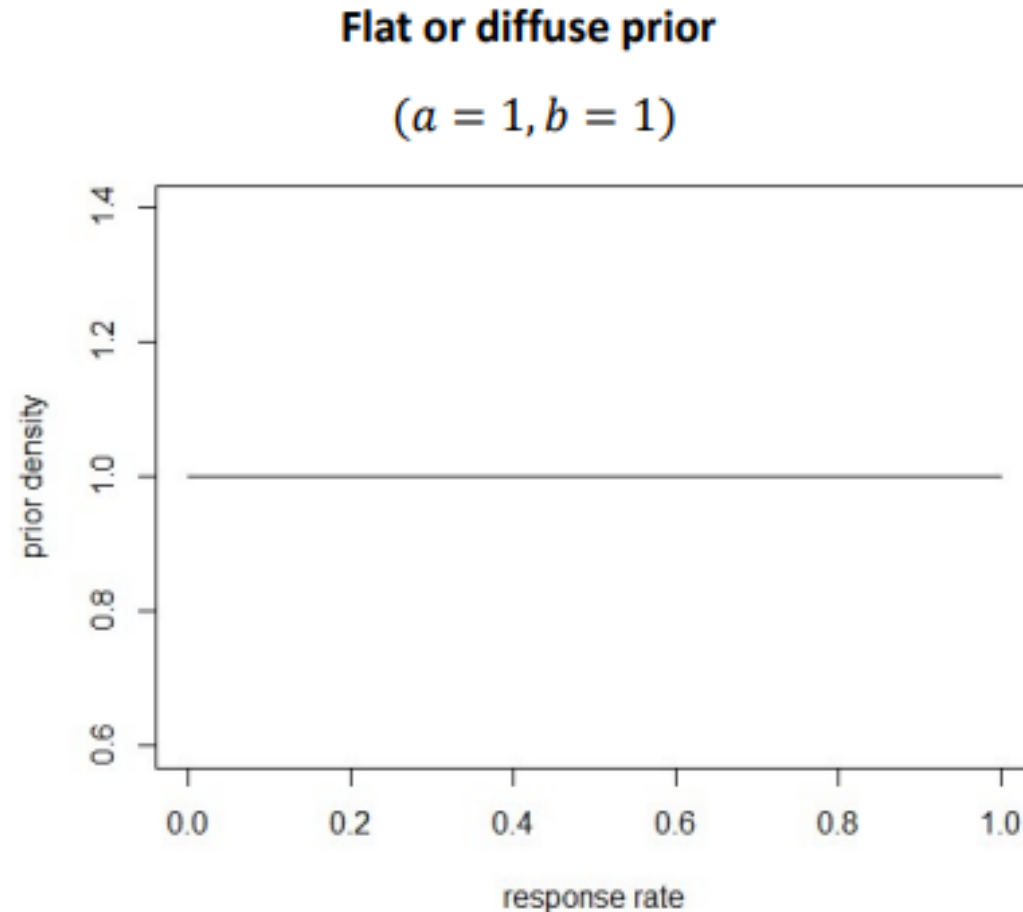
3. How could you change the distribution so that you were pretty certain that the average time on site is 2 and 8 minutes?

Q&A Session Next week