Regression Models on Survey Data

Live Session

8th February, 2021

Marju Kaps

Agenda for Today

- Announcements / Reminders
- Revision Question (Poll)
- Go through Survey GLM Tutorial
- Discussion Question: Coding data and interpretation
- Open floor for questions

Where we are

- This week: GLMs on survey data
- 15th Feb: No class – Reading week
- 22nd Feb: No class – Workshop week
  - Social Research Dissertation Workshop 22nd Feb, 6pm
- 1st March: Project Deadline (12 midday)
- 1st & 8th March: Multilevel models
- 15th March: Revising for the exam

Note: The exam timetable will be published in March

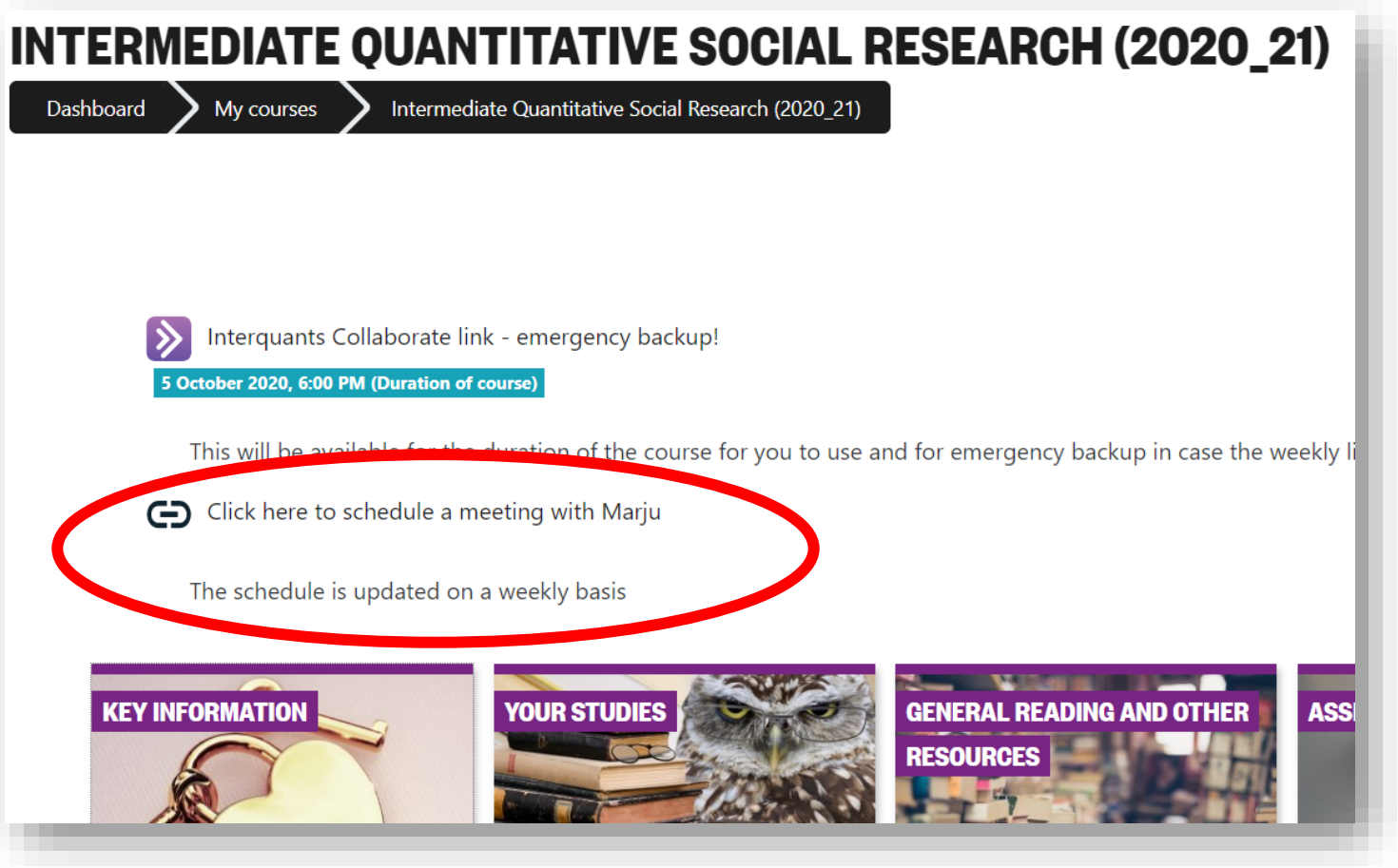
Meetings

I will continue to hold meetings through Reading Week and Workshop Week.

Sign-up link on Moodle

We can discuss:

- Projects
- Class content



A question from last time

Are MSc and MRes degrees at Birkbeck classified based on marks?

YES

Classification system (from the Registry pages):

- Distinction: 70% or above
- Merit: 60% - 69.99%
- Pass: 50% - 59.99%

For a Distinction, Masters students must typically also achieve a distinction mark in the dissertation.

POLL

Treatment of ordinal variables

What is an ordinal variable?

Ordered levels of a factor

Not equidistant

No meaningful 0 necessary

Examples?

```
Age groups
School years
Responses on a Likert scale: do you like math?
NO 1 2 3 4 5 YES
```

Treatment of ordinal variables

Independent Variables

No “perfect” treatment of ordinal ***predictor*** *variables* in R regression models. Coded either as

1. Categorical (factor) or
2. Continuous (numeric)

Dependent Variables

If the ***outcome*** *variable* is ordinal, you can run an “ordinal logistic regression”

- See Additional Resources on Moodle (Not on the exam)

Discussion question

Which of the two “importance of religion” variables from CES would you use in an analysis?

Factor better: can see all the levels individually, easier to talk about the levels

No real linear relationship

Continuous better: Variables that "feel" more continuous (political affiliation, scores on test)

If there are many variables with many levels, makes summary tables easier to track