Social Sciences' Research
Methods Centre (SSRMC)

Cambridge Replication Workshop 2013/14

Weekly Assignments

This is the second time the Replication Workhop took place at Cambridge (academic year 2013/14). Over Christmas there was a longer break to start the replication projects. The course had eight two-hour sessions; two weeks had a double session of 4 hours.

Week 1: Picking a paper

In the first session we will pick your paper for replication so that you can get started right away. There are two TAs (and me) helping with the final decision in the practical part of the session.

Either pick a paper from my suggestions (Dropbox folder – check first if data are available!), or find up to three papers yourself. Either way, you have to create a folder with your name and put the pdfs into that folder before class. Please follow the structure for your folders suggested in Dropbox > Student papers.

Please also bring printed versions to class.

Guidelines on how to pick a paper:

- Pick a paper where the data set is available at the Journal's webpage, the author's webpage or a data verse. Do not pick a paper where you cannot locate the data!
- To increase the probability that your paper will eventually be published, pick an article from a top journal.
- The article should be published 2008 or later.
- Your paper should use methods you can manage or learn over the Christmas break.

Please also prepare the readings for week 1 indicated on the syllabus.

Week 2: Start working with your replication paper

In the second session we discuss how a replication is usually structured. This will help you to structure your own project. We also prepare the replication you will do over the Christmas break.

1) Be prepared to discuss the reading above (Gary King).

2) In your dropbox folder:

- Create a word document called "tables" and note all table and figure titles from your project paper (you can also copy screenshots* in there instead)
  Upload this to: Dropbox > Yourname > Analysis > Article > Original Article
  The job for the Christmas break will be to recreate these tables one by one.

- Create a word document named "models" in which you make a note of

  - all Models run in paper, e.g. ARIMA (1,0), Regression
  - which Rcode or package would apply (google beforehand – the TAs will help check if you found the right one), e.g. Package {lme} ...

  Upload this to: Dropbox > Yourname > Analysis > Article > Original Article
  We will help you find the right R Packages and functions to start replicating the first models in class based on that. This will also show you in which cases you have to contact the author before Christmas and ask which models exactly he/she used in case it is not clear.

- Create and upload to dropbox an Rscript called "descriptive.R". In that Rscript, load your dataset into R; and do descriptive statistics (mean, histograms, scatterplots); get familiar with the data
  Upload this to: Dropbox > Yourname > Analysis > R
  This will also show you in which cases you have to contact the author before Christmas to identify the variables.

Screenshots:
*Shortcut macbook: cmd + ctrl + shift + 4 (copies a selected bit of your screen; then paste into word)
Shortcut macbook: cmd + ctrl + shift + 3 (copies a selected bit of your screen on desktop; insert that into word)

Week 3+4 (double session): Presenting first results

In the third session we will discuss the work you have done over Christmas. We will also discuss how to add value to a replication project. The goal for the break is that you are producing a clear, structured replication for new replication partner to cross-check later on.

In your dropbox folder:

1. Create an .R file that contains your replication. Either call it analysis.R and put everything in there, or have separate R.files for each figure, e.g. figure1.R, figure2.R and so on. This file should reproduce all tables and figures, and if your paper does not present the results in such formats, then reproduce all models reported in the text.
   Upload this to: Dropbox > Yourname > Analysis > R
   In the practical part, we will go through this and help with coding.

Guidelines for your R code:

Remember that your replication partner will have to navigate through this file.

Comment everything clearly so that your replication partner can later reproduce it:
load data
transform data
figure 1
...

Do not make any changes in the **original** data set: always load it into R first, then make changes there. This will later be reproducible and transparent. Do not change anything e.g. in the original csv file.

2. Create a word document named "problems.doc" in which you make a note of

   - Problems in R
   - Questions on the models you were running
   - Other issues that came up (e.g. corresponding with the author)
   - make a note of which tables were replicable, and when the results deviated

   Upload this to: Dropbox > Yourname > Analysis > Article
   We will go through this document in class and help to solve the issues in class.

Week 4: Getting your replication ready to hand it over to your replication partner

In your own dropbox folder...

1. Create a folder named replication_partner

2. Upload your original paper, Rscript with your concrete comments, a short report, and questions you have for your replication partner (see guidelines from handout week 3)

3. Send an email to your replication partner giving them info about this folder

Some guidelines:

How to prepare your materials so that your replication partner can work with them:

- Keep it concise. The cross-check should not take more than 3-4 hours.
- Have you provided a list of variable names and explanation?
- Have you provided the updated data set (in which you have already done the recoding)?
- Have you commented your Rcode clearly (see week 2 guidelines)? The TAs will have a look during the practical part
- Have you finished tables and figures?
- Have you written a short results report, 500 words max.?
- You can create a file for open questions if you would like your replication partner to have a look at that.
- Have you uploaded this report including your output figures/tables (screenshot is ok) and Rscript into dropbox?

- Have you sent an email to your replication partner explaining the content of your dropbox (Rcode, Paper, Data)?

Week 5: Cross-check & Analyzing a published replication paper

1. Finish your replication:

If you can't replicate all models yet, then take the time you have to make the code for those you completed readable, and note problems and issues you had:

Go through your file analysis.R (in folder Dropbox > Yourname > Analysis > R) and problems.doc (in folder Dropbox > Yourname > Analysis > Article) and check if this is readable for your replication partner. During the session you will sit with your partner and explain to them what your project is about, and how they can best help you. These two files should be in your dropbox by January 22, 6pm (one day before class!) so I can have a look. If you have last minute questions in R me and the TAs can help before you hand the files over to your partner.

2. Pick a replication partner:

Think about who you would like as your replication partner = who's paper you'd like to replicate (based on similar models, same field etc., if they know R code well and you can learn from them etc.). You can browse the others' projects on dropbox. We will discuss this in class.

3. Read a published replication study:

Each of you will present your thoughts on the question posed below in class. Papers are as pdf in dropbox > Literature > Example Replications.

Aiden	Peterson 1996 - What is the outline of the paper? What did they criticize about the published paper?
Atiyab	Peterson 1996 – How did they add value to the re-analysis (model improvement; theoretical criticism...)?
Audrey	Bell Miller 2013 – What is the outline of the paper? What did they criticize about the published paper?
Brooke	Bell Miller 2013 - How did they add value to the re-analysis (model improvement; theoretical criticism...)?
Mia	Wolferen et al. 2013 - How did they add value to the re-analysis (model improvement; theoretical criticism...)?
Cindy	Wolferen et al. 2013 - What is the outline of the paper? What did they criticize about the published paper?
Ed	Herndon et al. 2013 - What is the outline of the paper? What did they criticize about the published paper?
Rasheed	Herndon et al. 2013 - How did they add value to the re-analysis (model improvement; theoretical criticism...)?

Week 6+7 (double session): Planning your added value part

1. Work on finishing your replication

analysis.R (in folder Dropbox > Yourname > Analysis > R). We have a double session before you upload your final assignment. This is your chance to get help in R with your models.

2. Replicate your partner's work & report to class

In the folder Week6+7 > Students cross-replications > yourname replicated othername:

- upload your Rscript (save their provided Rcode under 'mia_replicates_cindy.R' and make comments within)
- upload max. 500 words report (mia_report_for_cindy.doc) on issues, problems, how you think you could help during this exercise

3. Planning your added value part

Upload a document "value.doc" listing 2-3 ways in which you could potentially add value to your replication and why this is a good idea. You can use the handout from week 5 in the dropbox "how to add value.doc".

Week 8: Wrapping up

By now, you should have finished the final assignment:

- a well commented Rcode of your replication: if you only partially replicated the paper, make sure you mark clearly which tables you replicate
- a word document or pdf with your report on what you replicated, which issues you encountered, how you added value or how future work could add value: this can be a 1-2 page report or a full replication paper
- the data set

Please put them in the dropbox beforehand so I can have a look. Also, read the pdf manual on how to sign up for the data verse and send me your username.