

Quiz 1 - Reproducible Research

Jean-Luc BELLIER

19 juillet 2016

Question 1

Suppose I conduct a study and publish my findings. Which of the following is an example of replication of my study ?

- An investigator at another institution conducts a study addressing a different scientific question and publishes her findings.
- I give my data to an independent investigator at another institution, she analyzes the data and gets the same results as I originally obtained
- I take my own data, analyze it again, and publish new findings
- An investigator at another institution conducts a study addressing the same question, collects her own data, analyses it separately from me, and publishes her own findings.

Answer : An investigator at another institution conducts a study addressing the same question, collects her own data, analyses it separately from me, and publishes her own findings.

Question 2

Which of the following is a requirement for a published data analysis to be reproducible ?

- The full computer code for doing the data analysis is made publicly available.
- The data analysis is conducted using R.
- The investigator's final publication is made available free of charge.
- The data analysis is conducted on a variant of the Unix operating system.

Answer : The full computer code for doing the data analysis is made publicly available.

Question 3

Which of the following is an example of reproducible study ?

- The study's analytic data and computer code are not publicly available, but the study is simple enough to be repeated by an independent investigator.
- The study's original authors re-run their computer code on their analytic data and confirm publicly that the findings match those of the published results.
- The study's analytic data and computer code for the data analysis are publicly available. When the code is run on the analytic data, the findings are identical to the published results.
- The study's analytic data are publicly available, but the computer code is not.

Answer : The study's analytic data and computer code for the data analysis are publicly available. When the code is run on the analytic data, the findings are identical to the published results.

Question 4

Which of the following is a reason that a study might NOT be fully *replicated* ?

- The original study was conducted by a well-known investigator.
- The original study was opportunistic in its timing and it would be difficult to find a similar context in which to repeat it.
- The original study was published in a high impact journal and is considered authoritative.
- The original study has null findings.

Answer : The original study was opportunistic in its timing and it would be difficult to find a similar context in which to repeat it.

Question 5

Which of the following is a reason why publishing *reproducible research* is increasingly important ?

- New technologies are increasing the rate of data collection, creating datasets that are more complex and extremely high dimensional.
- The statistical methods for most studies can be accurately described using plain language.
- Most studies today are small-scale and easily replicated.
- Computing power is limited today, making it difficult to apply.

Answer : New technologies are increasing the rate of data collection, creating datasets that are more complex and extremely high dimensional.

Question 6

What is the role of *processing code* in research pipeline ?

- It transforms the measured data into analytic data.
- It conducts the statistical analysis of the primary outcome.
- It transforms the computational results into figures and tables.
- It transforms the analytic data into computational results.

Answer : It transforms the measured data into analytic data.

Question 7

Which is a goal of literate statistical programming ?

- Separate figures and tables from other data analytic summaries.
- Ensure that data analysis documents are always exported in PDF format.
- Require that data analysis summaries are always written in LaTeX.
- Combine explanatory text and data analysis in a single document.

Answer : Combine explanatory text and data analysis in a single document.

Question 8

What does it mean to *weave* a literate statistical program ?

- Transform a literate program from R to Python.
- Transform the literate program into a human readable document.
- Compress the literate program so that it takes up less space.
- Transform the literate program into a machine readable code file.

Answer : Transform the literate program into a human readable document.

Question 9

Which of the following is required to implement a literate programming system ?

- A Unix-based computer system.
- A programming language like R.
- A web server for publishing documents.
- A program that views PDF files.

Answer : A programming language like R.

Question 10

What is one way in which the knitr system differs from Sweave ?

- knitr lacks features like caching of code chunks.
- knitr allows for the use of markdown instead of LaTeX.
- knitr was developped by Friedrich Leisch.
- knitr is written on Python instead of R.

Answer : knitr allows for the use of markdown instead of LaTeX.