```
---
title: "Taking a step towards reproducibility in research"
author: "A Michelle Edwards, Ph.D., MLIS"
date: "September 9, 2019"
output:
  word_document:
    toc: true

---
```

```{r echo=FALSE}
library(knitr)
library(SASmarkdown)

```

What is Reproducible Research?

Definitions

Reproducible:

* Able to be shown, done, or made again
(https://dictionary.cambridge.org/dictionary/english/reproducible)

* Able to be reproduced or copied
(https://www.lexico.com/en/definition/reproducible)

* Capable of being reproduced
(https://www.vocabulary.com/dictionary/reproducible)

##

Using these definitions - how would you define "Reproducible
Research"?

##

##

##

"Reproducible Crisis"

A study has claimed that up to 2/3 of researchers have tried and failed
to reproduce another scientist's experiments
(https://www.bbc.com/news/science-environment-39054778).

Experiments are meant to be replicable or reproducible. Materials and
Methods section should be similar to a recipe - with step-by-step type of

guidelines to follow. Data collection and statistical analysis methods
form part of the Materials and Methods section of papers and theses.

##

Exercise No. 1

Read the Materials and Methods section of the paper that was handed out.
Working with a partner, can you identify the steps that were used for
data collection and statistical analysis? Is anything missing? Can you
identify all the data (variables) that were collected? Did the author(s)
create any new variables for their analysis? Did they delete any
observations? Were there any transformations? What analysis did they
conduct? Did they check the assumptions of the analysis they used?

##

What can YOU do to help make your research REPRODUCIBLE?

  1) _

  2) _

  3) _

  4) _

  5) _

##

Documenting your R or SAS syntax

R syntax

Here is a sample of R syntax:

```{r}
#  Using the base aov() function in base R this will conduct an ordinary
least squares ANOVA              #
model <- aov(nitrogen ~ block + trmt, data=rcbd_data)
model

summary(model)


#  Let's fix our block and trmt effects
#
rcbd_data$block_fac = as.factor(rcbd_data$block)
rcbd_data$trmt_fac = as.factor(rcbd_data$trmt)

#  Running the aov() with block and trmt as factors
#
```

```
model2 <- aov(nitrogen ~ block_fac + trmt_fac, data=rcbd_data)
model2

#  Run summary on your new updated model
summary(model2)

#  What if we wanted to see Tukey means comparisons?
model3 <- aov(nitrogen ~ block_fac + factor(trmt_fac), data=rcbd_data)
TukeyHSD(model3,which='factor(trmt_fac)')
summary(model3)

```

SAS syntax

Here is a sample of SAS code:

```{sas }
*  Partioning of the Variation for an RCBD trial;
 Proc glimmix data=rcbd;
  class block trmt;
  model Nitrogen = trmt/s;
  random block;
  title "Proc GLIMMIX Results";
  lsmeans trmt / pdiff adjust=tukey ilink lines;
  output out=second predicted=pred residual=resid residual(noblup)=mresid
student=studentresid
 student(noblup)=smresid;
Run;

```

Exercise No. 2

Working with a partner, can you identify the steps that were used in the
statistical analysis? Do you understand what the analysis is doing? Are
these pieces of syntax helpful to have in a thesis? or accessible to a
published paper?

##

Adding Comments to your code

For R syntax - you can add information embedded in your code by using the
"#" at the beginning of the line of code. In the sample code above - you
will notice that there are a number of comments included in the syntax.
You can make these are long or as short as you wish. These will help YOU
remember what you did and why you did it.

For SAS syntax - you have two options to add comments in your program.
If you only have one line - then you would add an "*" at the beginning of
the line and a ";" at the end of the line. If you have several lines to
add as a comment you can add "/ *" at the beginning and a "* /" at the
end of your comment.

Exercise No. 3

Take a few moments to review the R and the SAS code in the previous
section to see how comments have been used. Would you add more comments
in the code? Could you add more comments? Why?

##

Code and results ALL together

Whether you use R or SAS, your syntax or code is in one file, while your
output is usually in a second file. We tend to create the code, run the
code, and save our output. We may then print the output and use it to
writeup our results. What happens when you look at the output, say 6
months down the road and you notice a problem? Maybe you decide that you
would like to rerun the code. Will you be able to find the exact code
you ran? Chances are you've moved on from that code and modified it.

Maybe the answer is to create a document where you see your code and your
output together?

MARKDOWN

Check out these examples - the first one for R and the second for SAS.

   1. Example using R Markdown:

##

```{r}
# Analysis conducted 20190913
# Reading data from the Excel file that contains the RCBD trial sample
data using the READXL package
library(readxl)

rcbd <- read_excel("~/Workshops/Common_Files/RCBD_excel.xlsx",
col_names=T)
rcbd


# Running the Mixed Model ANOVA for the RCBD trial data using the LME4
package
library(lme4)

model1 <- lmer(Nitrogen ~ ((1|block) + trmt), data=rcbd)
model1
```

##

   2. Example using SAS Markdown

##

````
```{sas ex-sas, echo=TRUE}
* Analysis conducted 20190913;
* Reading data that contains the RCBD trial sample data;

Data rcbd;
  input block trmt Nitrogen Weed Bin_weed;
  datalines;
1 1 34.98 81 0.81
2 1 41.22 87 0.87
3 1 36.94 89 0.89
4 1 39.97 79 0.79
1 2 40.89 88 0.88
2 2 46.69 85 0.85
3 2 46.65 99 0.99
4 2 41.90 86 0.86
1 3 42.07 54 0.54
2 3 49.42 23 0.23
3 3 52.68 11 0.11
4 3 42.91 16 0.16
1 4 37.18 10 0.10
2 4 45.85 23 0.23
3 4 40.23 10 0.10
4 4 39.20 69 0.69
1 5 37.99 61 0.61
2 5 41.99 06 0.06
3 5 37.61 02 0.02
4 5 40.45 75 0.75
1 6 34.89 21 0.21
2 6 50.15   08 0.08
3 6 44.57   27 0.27
4 6 43.29   03 0.03
;
Run;


* Running the Mixed Model ANOVA for the RCBD trial data;

Proc glimmix data=rcbd;
  class block trmt;
  model Nitrogen = trmt/s;
  random block;
  title "Proc GLIMMIX Results";
  lsmeans trmt / pdiff adjust=tukey ilink lines;
  output out=second predicted=pred residual=resid residual(noblup)=mresid
student=studentresid student(noblup)=smresid;
Run;


```
````

What can you do with R markdown?

##

   1. Keep code and results together - in one file!

   2. A way to archive your code

   3. A way to submit your code to journals

##

How do you create R markdown

   1. Create PDF, Word, HTML documents

   2. R markdown documents are made up of R (SAS) code and notes - like you would in a Word or text editor.

   3. Each R or SAS code section creates the output.

R Markdown example

One way to learn R markdown is to review an example. This document was created using R markdown. Please [click here](C:/Users/edwardsm/Documents/Workshops/Tutorials/Reproducible/Reproducible_tutorial.rmd).

Concluding remarks and thoughts

   * We should all try to ensure that our research is transparent and reproducible

   * Including comments in your syntax or code is a great place to start. It can help you remember what you did when you've been away from your analysis for a while. It can also encourage you to be more efficient with your coding, explaining what you did every step. YES! It will take time, and it can be difficult to slow down and add all the information into your code as you work.

   * Creating a document that contains both your syntax/code along with the results is a great way to keep your syntax/code together. Requires time to learn the R markdown and it needs you to spend more time documenting. But it will be worth the extra work in the end.