Research Data Management:

Store and Analyze

OCTOBER 19, 2017

A. MICHELLE EDWARDS, OAC

CAROL PERRY, LIBRARY

Objectives

Establish sound variable naming and labelling conventions

Identify sound storage options

Recognize the importance of data backups

Research life cycle

Research life cycle

RDM Review…

A sound strategy and best practices used to……

- Organize
- Document
- Store
- Analyze
- Secure
- Preserve/Share/Reuse

……Your data

Create a plan

Elements:

Collecting/creating data

Documenting your work

Managing your files

Storing, securing & backing up files

Preservation

Access, sharing & reuse

Setting the stage

Discussion

What do you need to think about when collecting data?

- File structure
- Comparisons across years
- Changes to variables

Workshop Scenario

The French Blast Research Group received an NSERC grant in 2016-2017 to continue an ongoing feeding trial. The research involves three horses from a number of different farms, where each horse is weighed(kg) at the beginning of the trial, placed on one of three feed regimens (hay, pasture, or silage), their total feed intake(kg) is measured for 2 consecutive weeks, and their weights (kg) are taken at the same time.

Additional information that is gathered includes: the name of the horse's owner, the annual income of the owner, and feed costs of the owner for their horses.

You have been hired by the French Blast Research Group to manage their data collections. They have 5 years of horse feeding trial data that was collected from a number of horse farms across Ontario since 2011. You have been hired to review, manage, and preserve their research trial data according to best practices and requirements of the NSERC grant.

Exercise 1

Work in groups of 2 or 3

Review your job description and the data that has been collected to date for 2017.

Please list any other pieces of data that you would like to collect and explain why.

How would you recommend that the data be collected? Paper, laptop, handheld device, etc…?

Review

Did you encounter any problems?

Commonly Used Statistical Packages

- SAS
- SPSS
- Stata
- R
- Matlab

Variable Name Restrictions and Limits

Length of the Variable Name

SAS: 32 characters long
Stata: 32 characters long
Matlab: 32 characters long
SPSS : 64 bytes long

- 64 characters in English
- 32 characters in Chinese

R: 10,000 characters long

1st Character of the Variable Name

SAS: 1st character MUST be a letter or an underscore
Stata: 1st character MUST be a letter or an underscore
Matlab: 1st character MUST be a letter
SPSS: 1st character MUST be a letter, an underscore, or @, #, $

R: No restrictions found

Variable Name Restrictions and Limits

Special Characters in Variable Names

SAS: NONE
Stata: NONE
Matlab: No restrictions found
SPSS : NONE except
- Period, @

R: NONE except
- Period

Case in Variable Names

SAS: Mixed case – Presentation only
Stata: Mixed case – Presentation only
Matlab: Case sensitive
SPSS: Mixed case – Presentation only

R: Mixed case – Presentation only

NO BLANKS (SPACES) allowed in any of the Statistical Packages
Beware of Function names in all Statistical Packages – these cannot be used as Variable Names

Best Practices for Variable Names

1. Set Maximum length to 32 characters
2. ALWAYS start variable names with a letter
3. Numbers can be used anywhere in the variable name AFTER the first character
4. ONLY use underscores “_” in a variable name
5. Do NOT use blanks or spaces
6. Use lowercase

Exercise 2

Work in groups of 2 or 3

The first 3 observations for 2017 has been collected and provided to you on a slip of paper.

1. Create variable names for ALL the pieces of information that has been collected
2. Create labels/definitions for the variable names – type these out in a new worksheet in Excel
3. Create a data worksheet for your data.
4. Save the file, using the naming conventions you created in the 1st workshop (October 5, 2017)
5. Select someone from your group to write out your variable names on the front board when you are finished.

Review

Did you encounter any problems?

Storage and security

How will you manage data security?

What are your responsibilities?

Are you aware of the university's Guidelines on Categorization and Security of ResearchData and Information?

https://www.uoguelph.ca/ccs/sites/uoguelph.ca.ccs/files/Categorization%20%26%20Security%20of%20Research%20Data_Final.pdf

Storage

How will your data be backed up?

How often will it be backed up?

How much space do you need for your data?

How long will backup copies be retained?

3-2-1 storage backup rule

Keep at least **three** copies of your data

Store the copies on **two** different media

Keep **one** backup copy offsite

Keep a ‘master file’ – original untouched – for emergencies

Storage & Security

Backup files on regular basis

- Keep in separate location

Create ‘master files’ for raw data files & important document files

- Store in physically separate, secure location (preferably on a networked drive)
- If working files are lost – make a copy from the ‘master files’

Synchronize files

Encrypt sensitive data – see CCS encryption services

Security

Banff National Park of Canada

Information Update:

Issued 2016-08-10 (YYYY-MM-DD)

Time 8:59:09

RESTRICTED ACTIVITY: VHF RADIO RECEIVER EQUIPMENT

WHAT:
Pursuant to Section 7(1) of the National Parks General Regulations, of the Canada National Parks Act, the following activity is restricted or prohibited by order of the Superintendent:
Use or possess VHF radio receiver equipment to follow or search for wildlife with radio transmitters

WHERE:
Banff National Park, Kootenay National Park and Yoho National Park

WHY:
To prevent disturbance, harassment and habituation of wildlife that carry radio transmitters for park management purposes

PENALTY:
violators may be charged under the Canada National Parks Act: maximum penalty $25 000.

Conclusion

Contact information:

Michelle Edwards
OAC Statistics Consultant
oacstats@uoguelph.ca

http://www.uoguelph.ca/oac/stats

Carol Perry
Associate Librarian, Research & Scholarship
lib.research@uoguelph.ca

Photo credits

All phots used under CC BY – SA 2.0 licenses

https://www.flickr.com/photos/smerikal/5532818874/in/photostream/ - title image2

https://upload.wikimedia.org/wikipedia/commons/9/93/Cutting_silage_at_Folsetter_1.jpg - title image1

https://www.flickr.com/photos/91035846@N05/9437703460/