Data Management for Your Research – From the Simple to the Complex

- Jon A. Detterich, MD – Division of Cardiology

- Robinder Khemani, MD – Division of Critical Care Medicine

- Mark James, BS – Information Technology

- Paul Vee

- Stephan G. Erberich, PhD – Chief Data Officer & Director of Bioinformatics

Objectives

1. To understand various methods of data collection – from simple tables to complex time varying data

    1. Excel, Access, Redcap, Acknowledge (Biopac)

2. To understand how to integrate data collection with data Analysis – Integrating the collection methods with analysis using Matlab, R, ….

    1. SAS/JMP, Matlab, R

3. To learn some of the pitfalls of data collection with integrated analysis and how to avoid them

    1. Data integrity – automated entry –
    2. Pulling from larger data pools

Study Design

Data collection methods are only as good as your study design!

Data Collection

- Critical first step of data analysis

- Entry directly into a data table
  - Advantages and Pitfalls

- Entry into an electronic form
  - Free Text
  - Drop down

Go to excel and Access examples

Excel – the simple data table

- What type of data are you working with?
  - Categorical
  - Continuous
  - Repeated Measures
  - Time Varying or Time Series

Gender	E
Male	H
M	H
Male	W
Male	H
Male	H
female	H
Female	W
Male	A
Male	A
Fem	H
Male	H
Female	W
Male	H
Female	H
M.	H
Female	W
Female	H
Male	W
Female	H
Female	H
Female	W
Female	H
Male	H
Female	A
Female	W

Gender	E
0	H
0	H
0	W
0	H
0	H
1	H
1	W
0	A
0	A
1	H
0	H
1	W
0	H
1	H
0	H
1	W
1	H
0	W
1	H
1	H
1	W
1	H
0	H
1	A
1	W

Go to excel table

- What is JMP?
  - JMP is SAS but in a graphical user interface (GUI) format
- What type of data are you working with?
  - Categorical
  - Continuous
  - Repeated Measures
  - Time Varying or Time Series

```
PROC PRINT DATA=idre.sales;
WHERE Country='AU' AND Salary<30000;
RUN;


PROC PRINT data=idre.sales;
WHERE Country='AU' & Salary<30000;
RUN;
```

Obs	Employee_ID	First_Name	Last_Name	Gender	Salary	Job_Title	Country	Birth_Date	Hire_Date
3	120121	Irenie	Elvish	F	26600	Sales Rep. II	AU	-4169	6575
4	120122	Christina	Ngan	F	27475	Sales Rep. II	AU	-523	8217
5	120123	Kimiko	Hotstone	F	26190	Sales Rep. I	AU	3193	10866
6	120124	Lucian	Daymond	M	26480	Sales Rep. I	AU	1228	8460
8	120126	Satyakam	Denny	M	26780	Sales Rep. II	AU	11951	18475
9	120127	Sharryn	Clarkson	F	28100	Sales Rep. II	AU	8404	15645
12	120130	Kevin	Lyon	M	26955	Sales Rep. I	AU	10575	18383
13	120131	Marinus	Surawski	M	26910	Sales Rep. I	AU	8668	17167
14	120132	Fancine	Kaiser	F	28525	Sales Rep. III	AU	-2462	8309

Go to JMP

Time Series - Collection

- How can you collect time series data
  - Download from pre-existing data - positives? Negatives?
  - Acquire in real time – how many signals are you recording?

- When time series data is collected, the method will set you up for your analysis
  - Acquire in real time – **time stamp (critical)**
  - Sample rate – sampling frequency must be established prior to collection and then not adjusted
  - Sample rate – **Nyquist rate (limit)**

Acknowledge (Biopac) Example

Matlab GUI Example