Data Management

Concerns

1. Data used to reach conclusions must be of high quality

2. Data is expensive. Archive data so that it can be retrieved and interpreted in the future.

3. Well managed data makes processing and analysis efficient.

Steps in Data Management

1. Plan data management, considering
   - objectives and outputs
   - resources and skills

2. Design field data sheets
   - Form layout should match computer file
   - Include date and collectors name and signature
   - Lab notebook should have numbered pages and be permanently bound (data sheets can be glued in)
   - For inventions, each page must be witnessed

3. Collect data

4. Check data

5. Data entry and organization. Common errors
   - data entry errors
   - incorrect conversion or combination of numbers
   - confusing variables and data sets

6. Back up data

7. Process data

8. Check processed data
   - keep a log

9. Archive data
   - document
   - describe

Data Management Problems

Technical
- not being able to use software
- not being able to set up data checking procedures
- organizing data in ways not compatible with some uses

Organizational
- multiple copies of files
- no one responsible for checking data
- no feedback on data quality
- no policy on archiving and making data available

Conceptual
- multiple entry of the same data or hand pre-processing of data
- missing links between numbers and information on source of numbers
- no audit trail

Data Checks

1. Consistency checks, eg length of pregnancy can not be 12 months unless you are an elephant

2. Frequency tables for categorical items

3. For real numbers produce summary statistics such as mean, maximum, minimum, inter-quartile range, etc.

4. Graphical summaries
   - scatter plot of one item against another shows cases that do not follow the general pattern
   - box-plots are good at showing outliers

Study Questions

1. Is the study clearly identified?

2. Is there information to help find related studies?

3. Are the objectives well-stated?

4. Is the location recorded?

5. Is the principal investigator clearly shown?

Measurements

1. Are the measurements clearly described?

2. Are the names clear?

3. Are the measurement units clear?

4. Is the object that was measured clear (eg plot vs sample)?

Field Layout

1. Is the field plan discernible from the data set?

2. Is the type of design stated?

3. Are the treatments clearly defined?

4. Are the treatment levels clear?

Data

1. Are these original or summary data? If summary, where are the originals?
2. Is it clear what measurement each value represents?
3. Can you tell the design, field layout and/or treatment level for each value?
4. When were the values recorded?
5. How were missing values recorded?
6. For discrete measurements, is each level described clearly?

Graphs

1. Title identifies what it is?
2. Axes are labeled, including units?
3. Categorical values are shown discretely, eg using bars?
4. Fitted lines are clearly identified and equations shown?
5. Number of data for each point is shown?
6. Standard error bars are included?
7. Tests of significance are shown?