

If experimentation is the Queen of the Sciences, surely statistics must be regarded as the Guardian of the Royal Virtue.
— Myron Tribus

Sound principles for simple statistics

Check the data	
Frequencies	Summary statistics
Missing data	Sample sizes
Reasonableness	Get feedback
Everyone involved	
Early	Often
Keep documentation	
Data	Analyses
Results	
Check the results	Cross-tabulations
Reasonable	Ask a statistician

Brochures

- Database ownership (1 of 3).
- Avoiding pitfalls that result in bad data (2 of 3).
- Guidelines for detecting bad data (3 of 3).
- How Quantitative Health Sciences can satisfy your research needs.
- Working with spreadsheets.

OHS Section

Pippa M. Simpson, PhD
Director

Raymond G. Hoffmann, PhD
Associate Director

Shun-Hwa Li, PhD
Senior Biostatistician

Ke Yan, PhD
Senior Biostatistician

Mahua Dasgupta, MS
Biostatistician

Melodee Nugent, MA
Biostatistician

Chris Cronk, ScD
Senior Epidemiologist

JoAnn Gray-Murray, PHD
Qualitative Researcher

Database Support

Kathy Divine, MS
Database Administrator

Haydee Zimmerman, BA
Database Analyst II

Kim Gajewski, BA
Database Analyst II

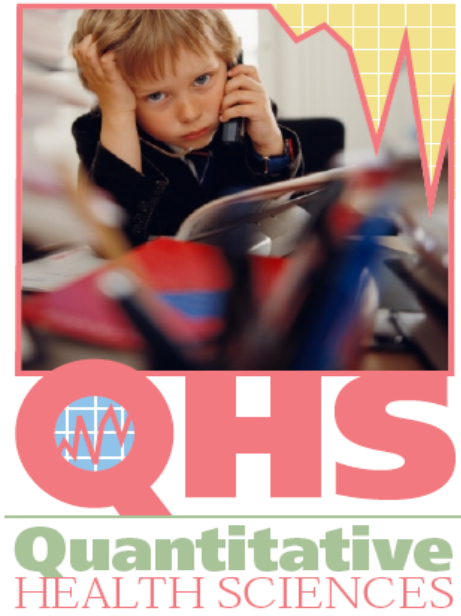
Robert Thielke, PhD
Manager IS II



Phone: (414) 955-7675 • Fax: (414) 955-6331

www.chw.org/qhs

© 2010 Children's Hospital and Health System. All rights reserved.



Sound principles
for simple statistics

Quantitative Health Sciences was established to provide help in the design and analysis of research studies.

Sound principles for simple statistics

The purpose of this document is for you to obtain a better preliminary analysis of your data and to be aware of issues that could jeopardize the quality of your analysis. These suggestions apply to all sizes of statistical analyses.

Before starting analyses, check the data for errors. Validating frequencies, minimum and maximum values will help pinpoint incorrect values.

Summary statistics such as means, cross-tabulations and plots will help in checking the data and analysis.

Backup

Backup data regularly. This can be done on a CD or on a server. Regardless of how you choose to backup your data, it always should be located on something external to your computer.

Check

Many things can go wrong in analysis. Regularly checking the data can facilitate early detection of error.

Consistency, such as:

- Ages (150 year olds).
- Weights (5000 lb. people).
- Pregnant males.

Sample size

- Cases may inadvertently be deleted.

Recodes

- Check cross-tabulations using the original variables. This will ensure that recodes were executed correctly.

Software

- A program may not be doing what you expected.
- Check that the results make sense.
- Enter in data for which you know the result.
- Discuss the results with your statistician.

For more complex analyses, get an expert involved. *Assumptions are built into many analyses, and you may be using tools incorrectly.*

Document

There often are many people involved in data management and analysis. Good documentation can help everyone stay informed. In addition to database documentation, also document the analysis process.

- Give value labels, especially when you recode variables (for example, 1=Yes, 0=No).
- Keep the syntax of the analysis.
- Save all versions of data files. This allows you to have an audit trail of changes and will provide protection in case of accidental deletion.
- Use meaningful labels for variables.
- Encourage feedback.

Obtain feedback from all project members about the interpretation of data. Ask for feedback early and often.

Know the source of the data

Knowing the basics about your data may be helpful in identifying obvious errors. Some of these basics may be:

- The extent and pattern of missing data.
- The number of observations in the subpopulations of interest.

No one should be nonexpendable

Good documentation will help transition new employees after the departure of old group members, especially when replacing key personnel. All members of the research group should be familiar with the analysis.

Protocol

- The analysis should relate to the protocol.
- The hypotheses and the tests performed should have the same aims.
- Consider possible confounders and effect modifiers.

Reports

When cleaning up output for “publication,” be especially careful about changing labels since this kind of change can be difficult to check.

Security

This always should be a concern. When data is collected, personal identifiers should be removed from the data to satisfy HIPAA requirements. Examples of possible identifiers are:

- Name.
- Address.
- Telephone number.
- Social Security number.

All such identifiers should be removed from the data set and stored in a safe place. Each identifier should be replaced with another unique identifier to match records back to the analysis data set, if needed.