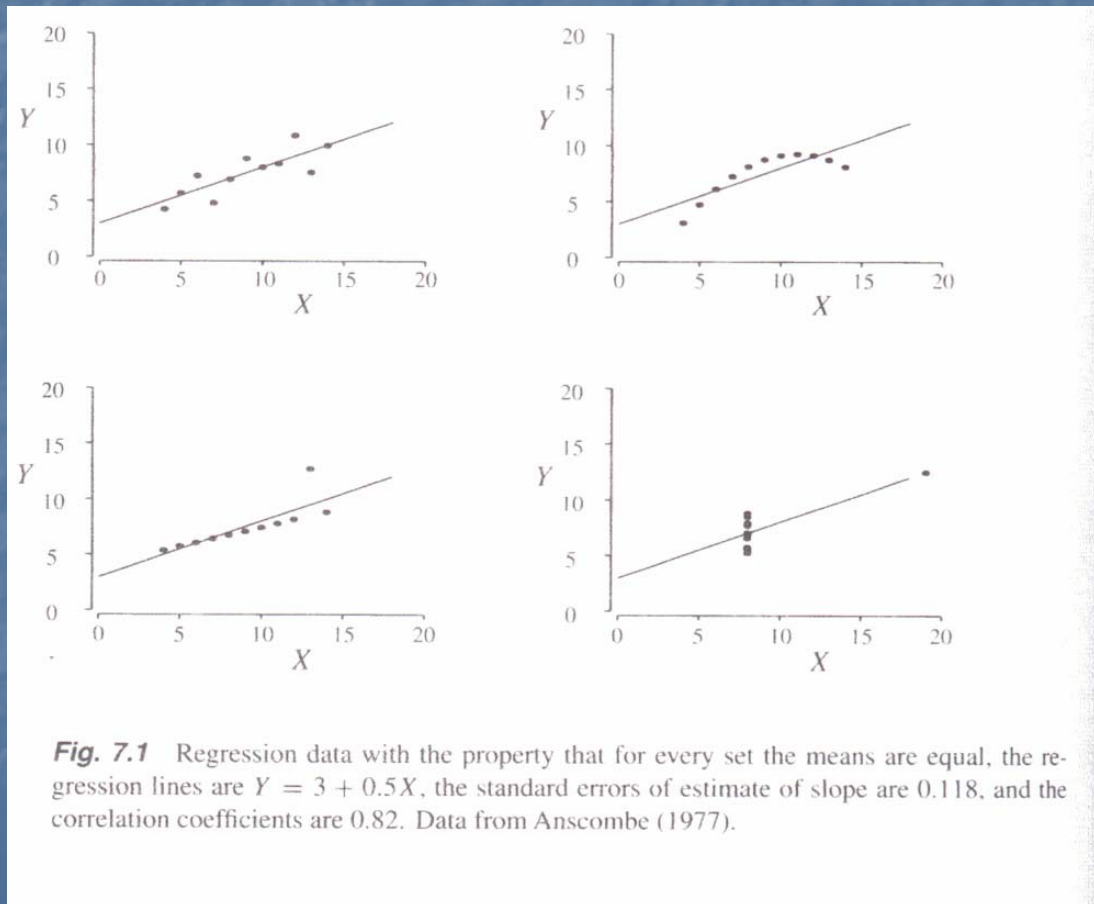


# Introductory Statistics with R



# The R environment

- Web site at <http://www.r-project.org/>
- A quick look at pull down menus and command lines
  - ls(), rm()
- ->R

# CRAN and Libraries

- Libraries are provided by users and extend the functionality of R
- They are loaded in a session with `library`
  - `library(qtl)` for example loads the qtl library
- They are unloaded with `detach`
  - `detach(qtl)` for example
- WHY? Because R is memory resident and you may need RAM for your data...

# CRAN

## Comprehensive R Archive Network

### CRAN Mirrors

The Comprehensive R Archive Network is available at the following URLs, please choose a location close to you:

#### Australia

<http://cran.au.r-project.org/>

PlanetMirror, Brisbane

<http://cran.ms.unimelb.edu.au/>

University of Melbourne

#### Austria

<http://cran.at.r-project.org/>

Wirtschaftsuniversitaet Wien

#### Brazil

<http://cran.br.r-project.org/>

Universidade Federal do Parana

<http://www.insecta.ufv.br/CRAN/>

Federal University of Vicosa

<http://cran.fiocruz.br/>

Oswaldo Cruz Foundation, Rio de Janeiro

<http://lmq.esalq.usp.br/CRAN/>

University of Sao Paulo, Piracicaba

<http://www.vps.fmvz.usp.br/CRAN/>

University of Sao Paulo, Sao Paulo

#### Canada

# CRAN

- Pick a download site and then:
- Software: Packages
  - This will open up a list of hundreds of possible libraries
- Task for next week:
  - explore the library list on one of the CRAN sites
  - Install qtl, maanova, and ISwR
  - Find a library that YOU find interesting and explore it

# CRAN

- Some packages are better loaded from a script (they need many related libraries).
- For next week: Go to <http://www.bioconductor.org/docs/install-howto.html> and load the bioconductor libraries into R.
- R libraries will open up the full power of R and the code is open source, so explore.

# Demos, tutorials, data sets

- Manual pages load with the libraries
- Tours
  - For example: C:\Program Files\R\R-2.3.0\library\qtl\docs should have a qtl tour, C:\Program Files\R\R-2.3.0\library\maanova\demo has several maanova tours: take these for next week.
  - Explore the file structure of the qtl library
- NOTE: CRAN provides a template for library submission, but the quality of code and documentation varies widely.

# Reproducible Research

- Always use scripts for your R code.
- One of the more attractive features of R for implementing reproducible research is that each R session provides the ability to store all of the commands used as an R script file and a copy of all the objects in the R session at any time. This allows for the complete analysis to be reproduced by another researcher or for the continuation of a study from a known stable configuration.

# Data frames

- Attach and detach allows for access to data within a data frame without the \$ notation:
- $\rightarrow R$

# Mouse Phenome Database

- <http://aretha.jax.org/pub-cgi/phenome/mpdcgi?rtn=docs/home>
- We will be using data sets:
  - 177 (pulse, blood pressure)
  - And the High Fat diet studies...)
    - 99 (HDL, Fat, Chol)
    - 143 (glucose, insulin, leptin, body wt)
- For next week, download these and import them into R (I'll send you scripts with one method)

# L2: Descriptive statistics and graphics

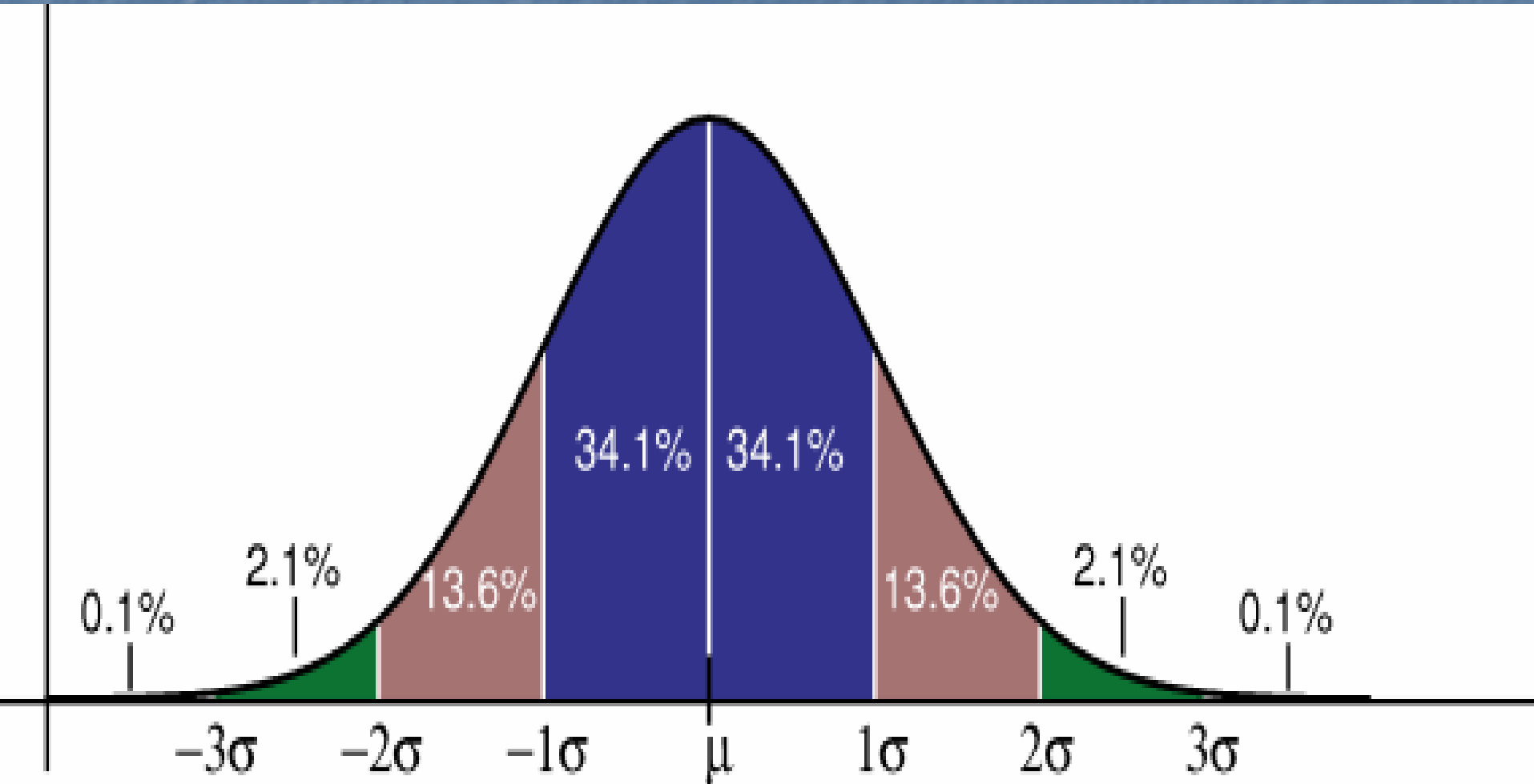
- For a set of data with components  $x_i$  with  $n$  entries

$$\bar{x} = \sum_1^n x$$

$$\text{var} = \frac{1}{n} \sum_1^n (x_i - \bar{x})^2$$

$$sd = \text{var}^{1/2}$$

# Normal distribution



# Not all data will be normally distributed

- Biological data is often
  - Lognormal distribution
  - Binomial distribution
  - Poisson distribution
- The important thing to know is what type of distribution your data has and to take this into account in the analysis
- >R