BIOS14 - Processing and Analysis of Biological Data

Per-Erik Isberg, Dept of Statistics

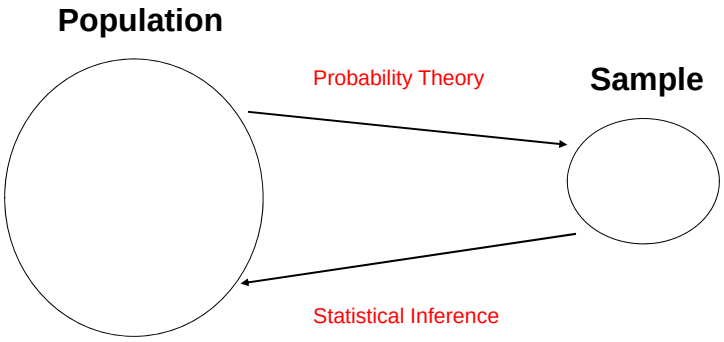
Introduction

Litterature

Quinn & Keough (2002) – Experimental Design and Data Analysis for Biologists, Cambridge University Press (ISBN 0521009766)

$\Rightarrow \quad y = f(x)$

Introduction



Probability

Probability ⇒ Probability distributions

- **Discrete distributions**
  - Binomial
  - Poisson
- **Continuous distributions**
  - Exponential
  - Normal
  - t, χ^2, F

Estimation

Example:

We sample n=20 individuals from a population

Calculate the mean in the sample $\Rightarrow \ \bar{x}$

Repeat the whole procedure $\Rightarrow \ \bar{x}_1, \ \bar{x}_2, \ \bar{x}_3, \ \ldots$

Q: What distribution?

A: Next thursday!

Estimation

Keywords:

- Confidence intervals
- Standard errors
- Central Limit Theorem

Methods:

- Maximum Likelihood (ML)
- Ordinary Least Squares (OLS)
- Bootstrap
- Jackknife
- Bayesian

Hypothesis testing

Statistical tests

H_0: Null hypothesis (assumed true)

H_1: Alternative hypothesis (one-tailed/two-tailed)

α = significance level of the test (5%)

Hypothesis testing

The statistical test

The truth	**H_0 not rejected**	**H_0 rejected**
H_0 true	Correct	type I error Pr( ) = α = often 5 %
H_0 false	type II error Pr( ) = β	Correct Pr( ) = 1 – β = power

Correlation and regression

$$y = f(x_1, x_2, \ldots)$$

Numerical variable

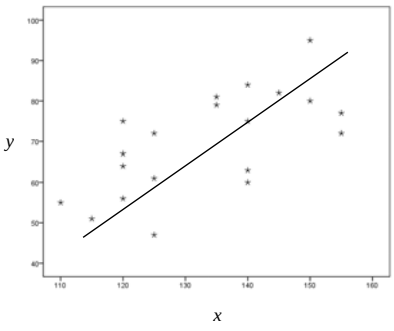
If the x's are numerical ⇒ Regression analysis

If the x's are categorical ⇒ Analysis of Variance (ANOVA)

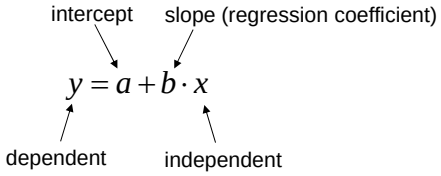
If we have both types ⇒ Analysis of Covariance (ANCOVA)

Correlation and regression

Of special interest ⇒ Linear models (OLS)



Correlation and regression



Correlation – measures the strength of the linear relation

Correlation and regression

Multiple linear regression

$$y = a + b_1 \cdot x_1 + b_2 \cdot x_2 + \ldots + b_p \cdot x_p$$

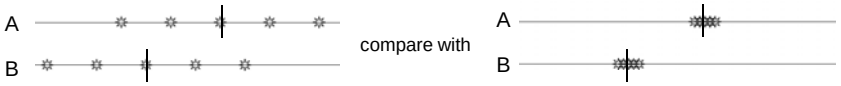
Tests, confidence intervals, predictions, …

$\Rightarrow$ Non-linear regression

Analysis of Variance

$$y = f(x)$$

Numerical Categorical



Keywords: between group variation – within group variation

Analysis of Variance

⇒ ANOVA table

	Sum of Squares	df	Mean Square
Between Groups			
Within Groups			
Total			

⇒ One-way analysis of variance

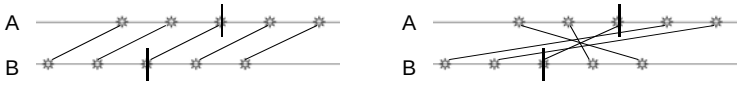
⇒ if just two groups = t-test (for independent samples)

Analysis of Variance

Sometimes the groups consists of the same indivuals measured at several occasions

⇒ Repeated measurements

⇒ if measured twice = t-test (paired)



Keywords: within individual variation

Analysis of Variance

Model I - Fix effects

Conclusions are only valid for the observed levels of the factor ⇒ Treatment A, B, C

Model II - Random effects

The levels used are a sample from a population of levels ⇒ Operator 1, 2, 3, 4

Mixed models include both fix and random effects

Analysis of Variance

Generalizations

- Factorial designs
- Nested designs
- Repeated measurements
- General Linear Models (GLM)

Analysis of Variance

Assumptions

- Independence
- Equal variances
- Normality

⇒ **Non-parametric methods**

- Mann-Whitney
- Kruskal-Wallis
- Kolmogorov-Smirnov

Analysis of frequencies

Goodness of fit

- χ^2-test
- G-test
- Fisher's exact test

Contingency tables (two-way)

Count

		Disease		
		Yes	No	Total
Exposed to riskfactor	Yes			
	No			
Total				

Analysis of frequencies

Higher order tables (Three-way and more)

⇒ Log-linear models

Keywords: Dependence - Independence

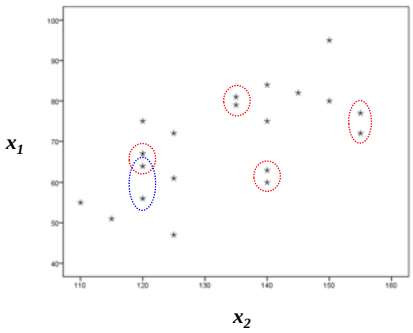
⇒ Logistic regression

$$y = f(x)$$

Probability (0/1)

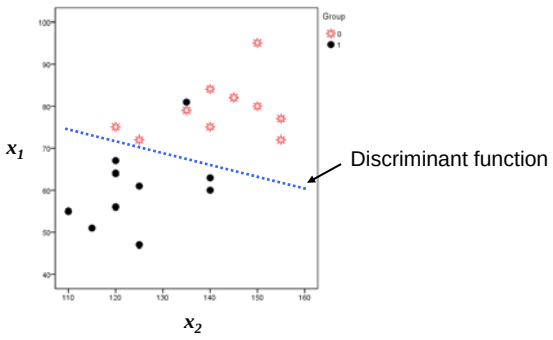
Multivariate analysis

Cluster analysis – Similarity among individuals



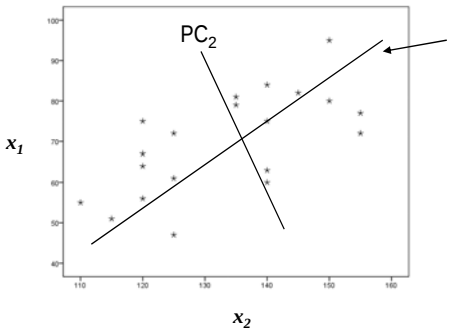
Multivariate analysis

Discriminant analysis – Prediction of group membership



Multivariate analysis

Principal component – Summarization of information



Survival analysis

We have a group of individuals that are ”alive” and want to study **when** they move into state ”dead”.

Of special interest:
T = Time for death, $T \geq 0$

Survival analysis

Special feature:

Censored observations – We can't observe the true death time. The individual is still alive at the end of the study.

Example: The time for completing a PhD!