Statistics Applied to Bioinformatics

Introduction

Jacques van Helden

Jacques.van-Helden@univ-amu.fr
Aix-Marseille Université (AMU), France
Technological Advances for Genomics and Clinics
(TAGC, INSERM Unit U1090)
http://jacques.van-helden.perso.luminy.univmed.fr/

Scope or the course

- Why statistics ?
  - Almost every problem in bioinformatics involves large data sets.
  - Background in statistics is very different depending on the scientific trainings
- Scope of the course
  - A general introduction to statistics
  - Concepts and methods selected for their importance in bioinformatics
  - Problem-driven approach based on a few concrete examples
    - Microarray analysis
    - Detection of over-represented patterns in non-coding sequences
    - Analysis of ORF lengths in different genomes
    - ...

Exercises

- At the end of each course, I will give some exercises, which aim at ensuring that the concepts are well understood.
- These exercises do not require calculation or computer, they consist in selecting the appropriate test or statistics and identifying the parameter values.
- At the beginning of each course, we will check the solutions and discuss questions

Practical training

- During the course: demonstration of a statistical program (language): R
- Practical introduction to R

Evaluation

- The examination will consist in exercises such as those proposed at the end of each course

Quick look at the dictionary

- Probabilité (Probability)
  - *Grandeur par laquelle on mesure le caractère aléatoire (possible et non certain) d'un événement, d'un phénomène par l'évaluation du nombre de chances d'en obtenir la réalisation. [Robert, 1982]*
  - *Value used to measure the stochastic (possible but not certain) character of an event or phenomenon, by evaluating its chances of realizations. [Robert, 1982]*
  - We will see an operational definition in the course

Quick look at the dictionary

- Statistique (Statistics)
  - *Ensemble des données numériques concernant une catégorie de faits (et utilisable selon ces méthodes d'interprétation) [Robert, 1982]*
  - *Set of numerical data about a category of facts (and which can be used according to these interpretation methods) [Robert, 1982]*

Quick look at the dictionary

- Inférence (Inference)
  - *Opération logique par laquelle on admet une proposition en vertu de sa liaison avec d'autres propositions déjà tenues pour vraies. [Robert, 1982]*
  - *Logical operation by which one admits a proposition on the basis of its relationship with other propositions considered as true.*
  - We will use a more precise definition in the course.

Quick look at the dictionary

- Déduction
  - Procédé de la pensée par lequel on conclut d'une ou de plusieurs propositions données (prémisses) à une proposition qui en résulte, en vertu de règles logiques. [Robert, 1982]