

Introduction to statistics

Computational biology week - London, April 2022

Jesús Urtasun Elizari, PhD



Imperial College
London

Outline

① Probability in a nutshell

- Discrete and continuous probability
- Mean and variance of a distribution
- Exercises in R

② Linear models

- What is a linear model
- Fitting a linear model
- Prediction vs inference
- Exercises in R

③ Hypothesis testing

- Probability distributions
- Statistical tests
- Real genetics problem
- Exercises in R

Chapter I

Probability in a nutshell

Probability in a nutshell

Discrete and continuous probability

- Discrete output \rightarrow probability well defined, compute by counting
- Continuous output \rightarrow need for probability distributions
- Definition of probability in both scenarios
- Computing probabilities in both scenarios

Probability in a nutshell

Discrete probability

Compute probability of getting 4 heads in a row

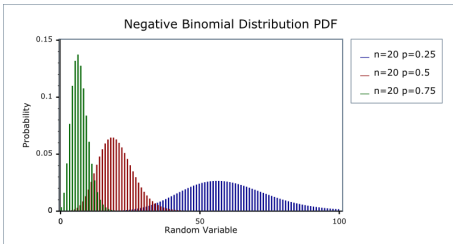
Probability in a nutshell

Discrete probability

Compute probability of getting a specific genotype

Probability in a nutshell

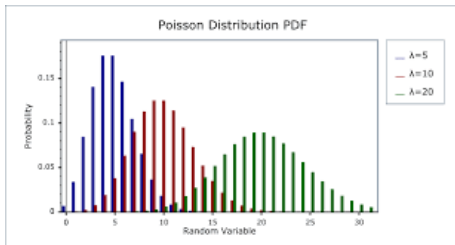
Binomial distribution



- Binomial distribution

Probability in a nutshell

Poisson distribution



- Poisson distribution

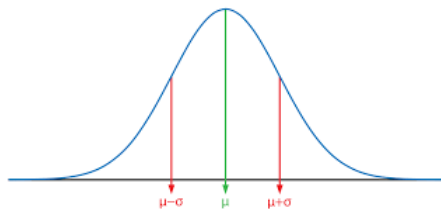
Probability in a nutshell

Continuous probability

Compute probability of measuring a particular height
Used when comparing means

Probability in a nutshell

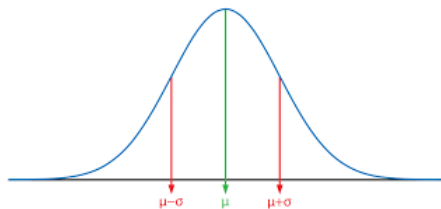
Gaussian distribution



- Gaussian distribution

Probability in a nutshell

Mean and variance



- Momenta of a distribution
- Mean, variance, skewness, kurtosis

Probability in a nutshell

Exercises in R

https://lmsbioinformatics.github.io/LMS_StatisticsInR/course/CBW_StatisticsInR_course

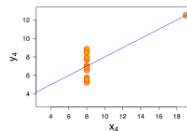
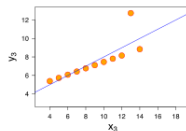
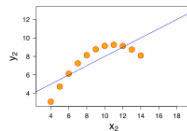
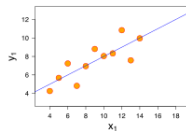
Chapter II

Linear models

Linear models

What is a linear model

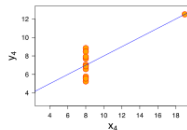
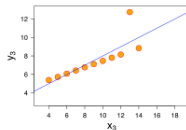
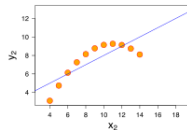
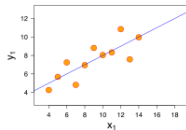
- 1 Find a function that describes a set of observations



Linear models

Fitting a linear model

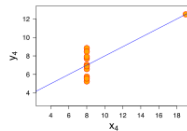
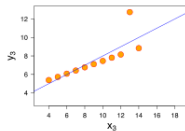
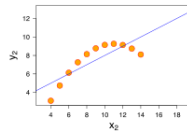
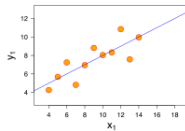
- 1 Fit a linear model
- 2 Measuring differences - residuals
- 3 Evaluate fit - correlation coefficient



Linear models

Advanced topics

1 (...)



Linear models

Exercises in R

https://lmsbioinformatics.github.io/LMS_StatisticsInR/course/CBW_StatisticsInR_course

Chapter III

Hypothesis testing

Hypothesis testing

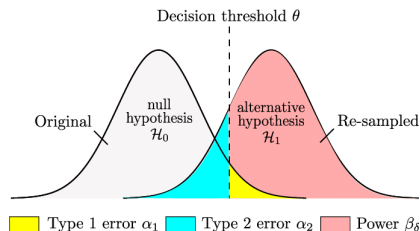
Introduction

- Null hypothesis and alternative hypothesis
- Statistic tests and p-values
- χ^2 -test, t -test, Wald test
- Exercises in R

Hypothesis testing

Compare different models

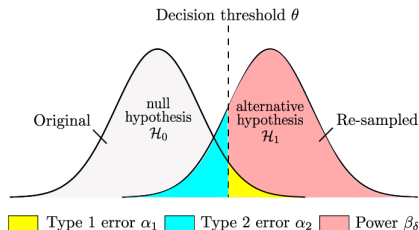
1 Null and alternative hypothesis



Hypothesis testing

Statistic test

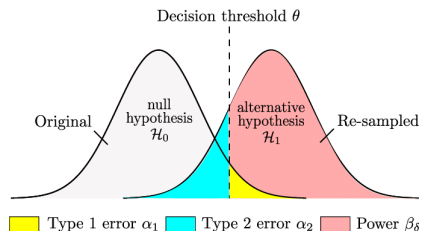
- 1 Quantify the significance of an observation
- 2 Certainty when accepting / rejecting a hypothesis



Hypothesis testing

p-values

1 p-values



Hypothesis testing

Exercises in R

https://lmsbioinformatics.github.io/LMS_StatisticsInR/course/CBW_StatisticsInR_course