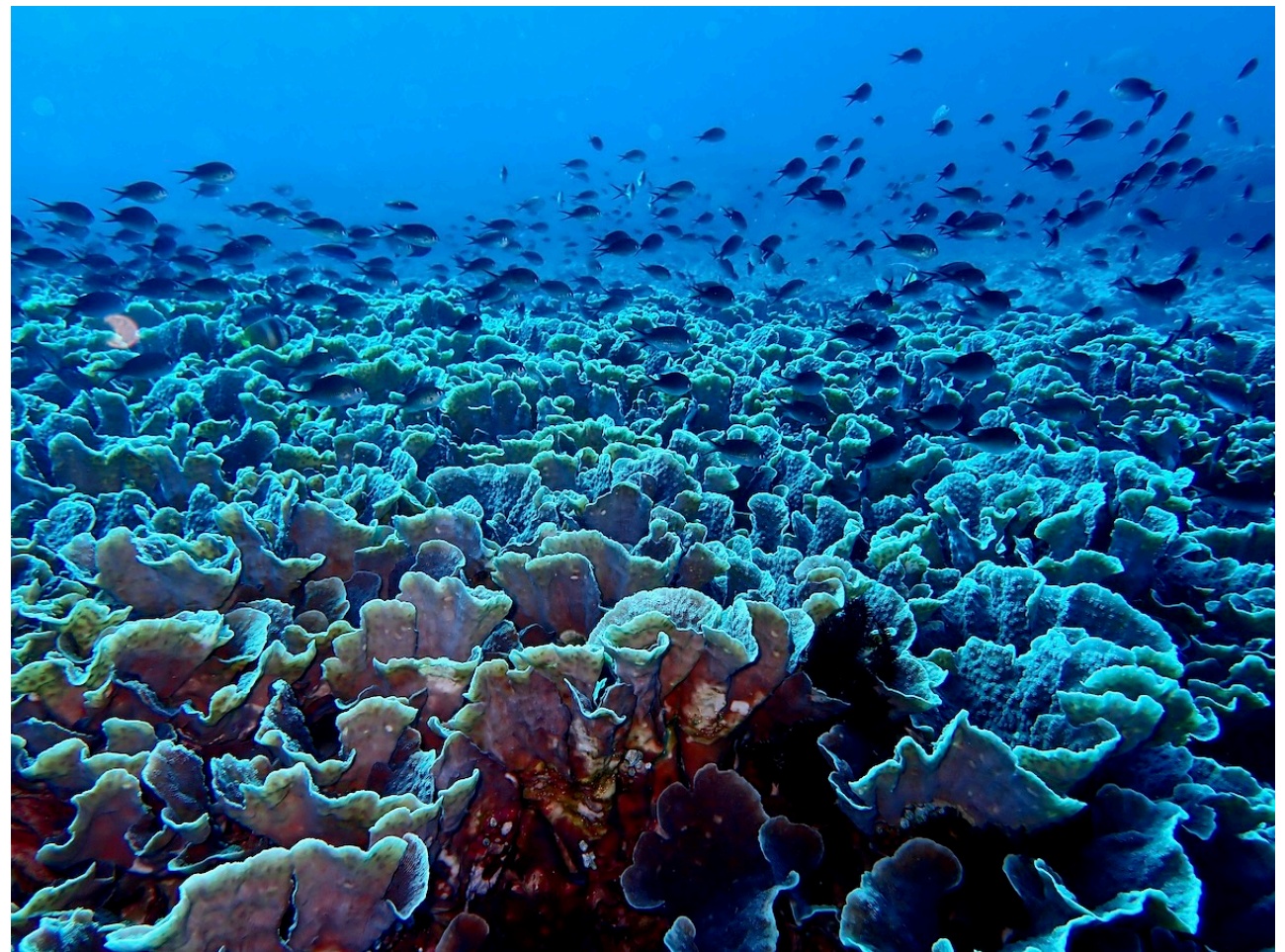


Exercises in Marine Ecological Genetics

01. Introduction

- General info and course outline
- Installing required software
- Downloading course materials
- Test for HWE using R

Martin Helmkamp



General course info

- Suggestion: we start at 14:00 and finish at 15:30
- Course language will be English, but questions can always be asked in German
- There will be no tests or grades
- Slides will be provided, but please do not post them online
- Contact: martin.helmkamp@uni-oldenburg.de or martin.helmkamp@leibniz-zmt.de

Objectives

- Apply theory and concepts of ecological genetics in practice
- Learn to analyze, visualize and interpret real world data
- Learn how to work on the command line / a scripting environment (R)
- Become familiar with the most common data types and file formats

Course outline

May be subject to change

Class	Date	Topics	Script
01	Apr 14	Introduction, software installation	01_intro.R
02	Apr 21	Within population genetics	02_within.R
03	Apr 28	Genetic drift and effective population size	03_Ne.R
04	May 05	Population structure and gene flow	04_structure.R
05	May 12	Isolation by distance (lecture online, exercises in person)	05_ibd.R
–	May 19	Himmelfahrt break	–
06	May 26	Whole-genome sequencing and genome assembly	06_genseq.sh
07	Jun 02	Genotyping, SNPs and population genomics	07_snps.R
08	Jun 09	Recombination and linkage disequilibrium	08_linkage.R
–	Jun 16	Student presentations	–
09	Jun 23	Selection and mutation	09_selection.R
10	Jun 30	DNA barcoding	10_barcode.sh
11	Jul 07	Metabarcoding	11_meta.sh
–	Jul 14	To be determined	–

Required software

- R
- RStudio
- git
- bash

Install the latest version of R from <https://cran.r-project.org>
(for Windows, choose the base version)

Select and install the latest RStudio Desktop version from
<https://www.rstudio.com/products/rstudio/download>

git and bash come already installed on computers running macOS.
For Windows, install git including bash by following the instructions at
<https://www.computerhope.com/issues/ch001927.htm>

Working with the command line / bash

Windows

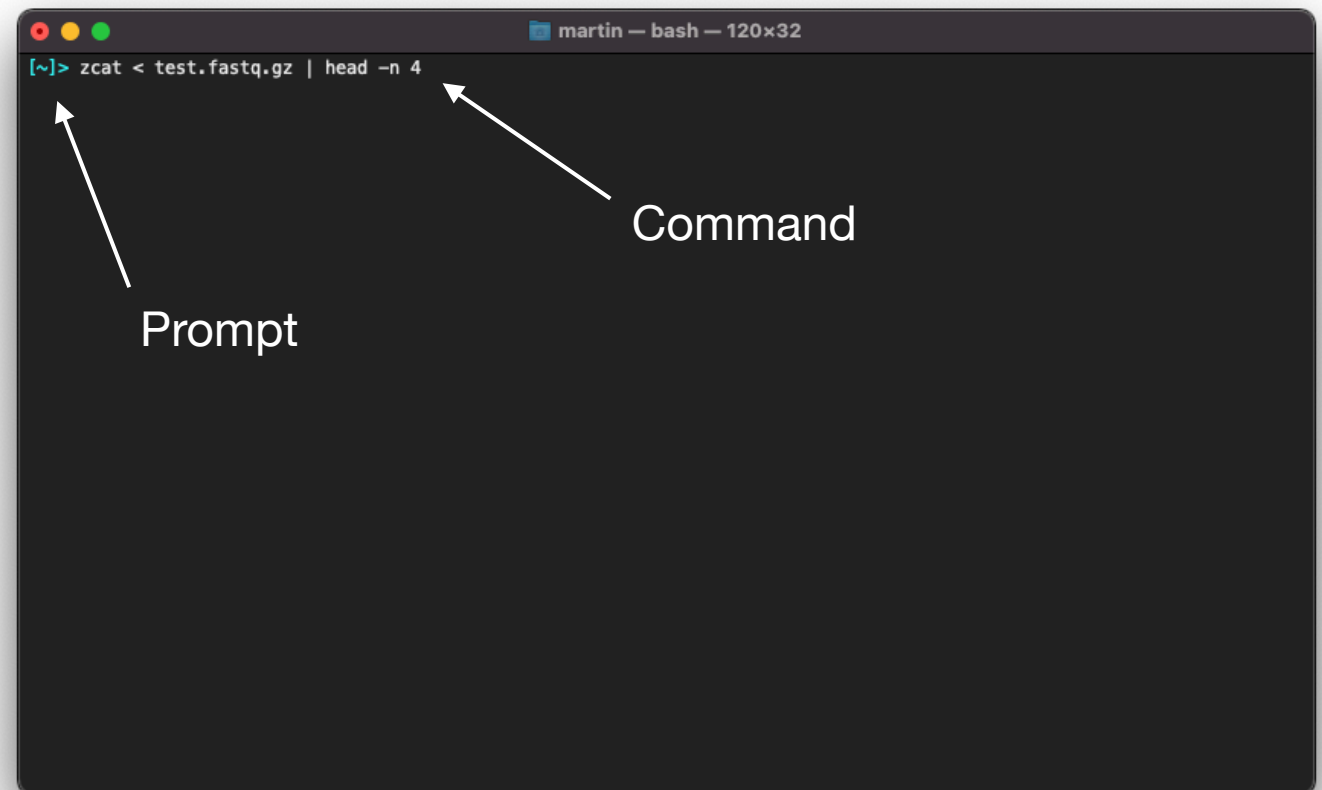
- launch git bash from Start menu
- alternatively, install Windows Subsystem for Linux (WSL) on Windows 10 or above (see <https://learn.microsoft.com/en-us/windows/wsl/install>)

macOS

- open Terminal app
in /Applications/Utilities,
type and execute “bash”

Typical usage

`command [-options] [file]`



Advantages of command line / scripting tools

- Highly flexible
- Can be automated and combined into complex workflows
- Reproducible, easy to document
- Can run on high performance computers

Download course materials using git

Create course directory (e.g. “Documents/meg23_exercises”)

```
cd Documents  
mkdir meg23_exercises
```

Create subdirectory “local”

```
cd meg23_exercises  
mkdir local
```



Download course repository from GitHub

```
git clone https://github.com/mhelmkamp/meg23\_repo.git
```


Download course materials using git

List directory contents

```
cd meg23_repo
ls -l                # -l: long format
```

Update course repository

```
git pull
```



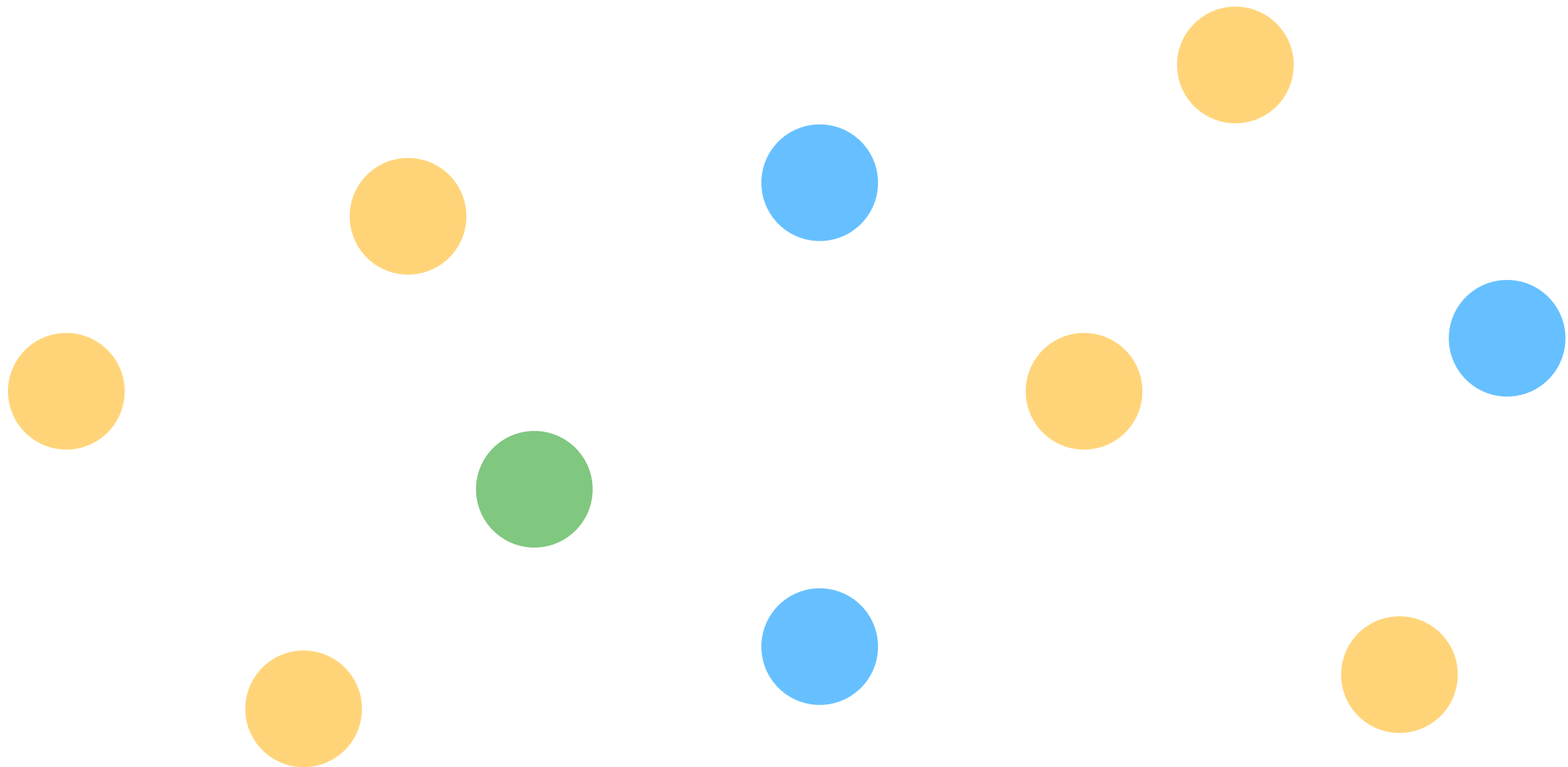
Please do not save over files in the course repository. Instead, save your own scripts to the local subdirectory (including copies of course scripts you would like to edit)

Hardy-Weinberg Equilibrium

Exercise 1

Is this population in HWE?

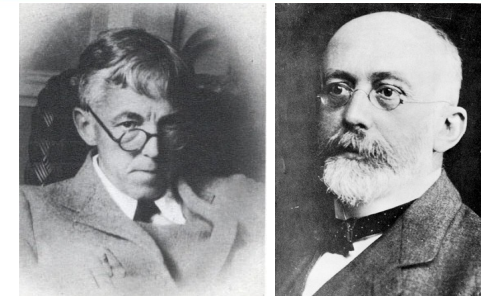
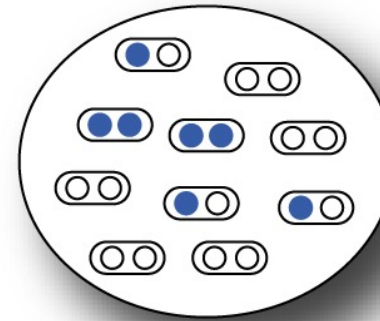
Diploid, 1 locus, 2 co-dominant alleles (yellow, blue)



HARDY-WEINBERG (1908)

Godfrey H. Hardy (1877-1947)

Wilhelm Weinberg (1862-1937)

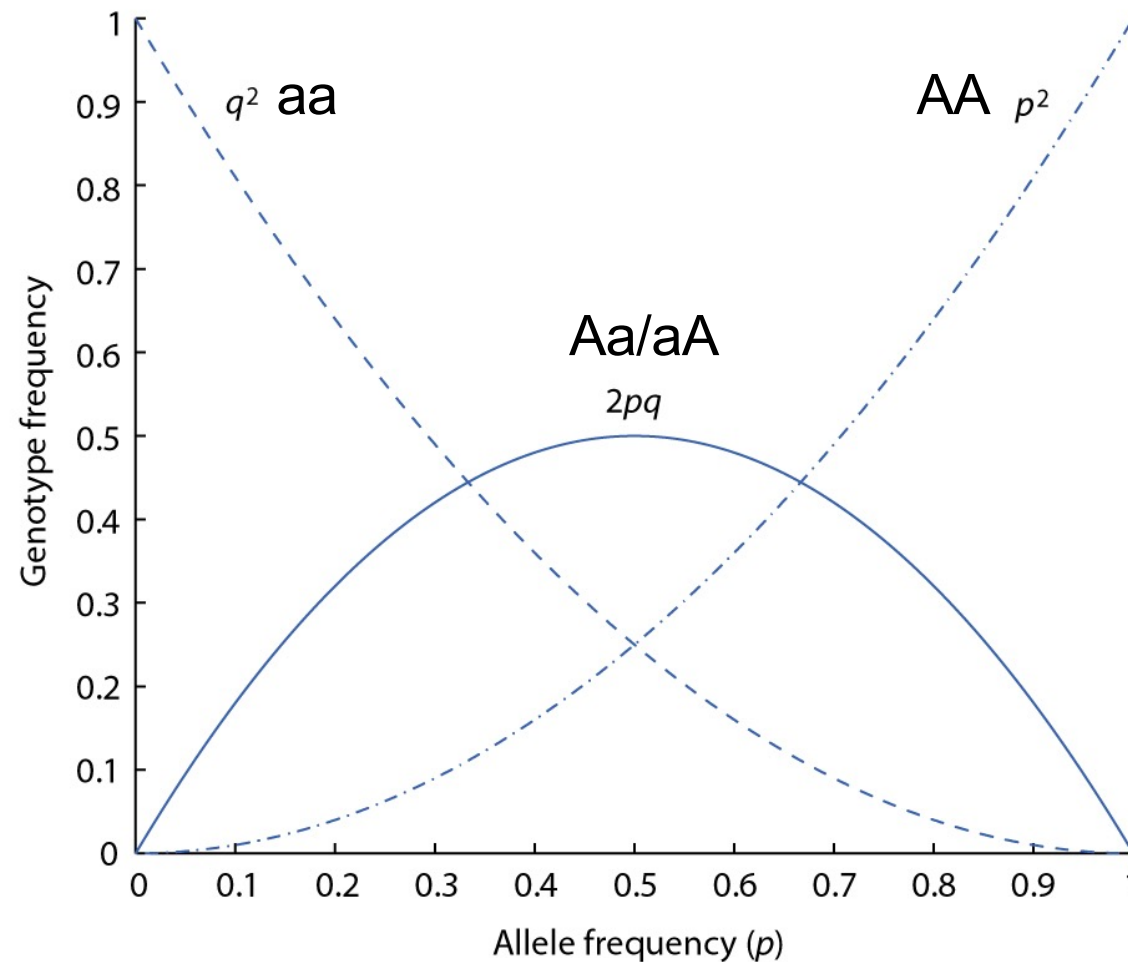


Establish the relationship between allele frequencies and genotype frequencies in a population

$$p^2 + 2pq + q^2 = 1$$

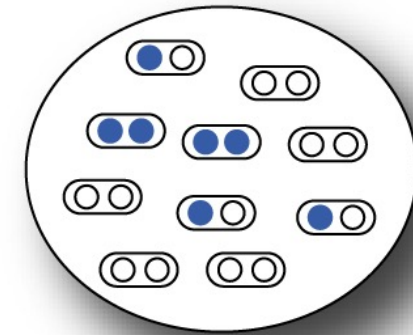
AA Aa/aA aa

p and q : allele frequencies for a locus with two alleles (A and a)
($p + q = 1$)



HETEROZYGOSITY

In one population



H_o = proportion of heterozygote individuals, observed heterozygosity

$H_e = 2pq = 1 - p^2 - q^2$, expected heterozygosity (assuming HW equilibrium)

$$F = \frac{H_e - H_o}{H_e}$$

Fixation index: *proportion by which heterozygosity is reduced or increased relative to the heterozygosity of a population at HW equilibrium with the same allele frequencies.*

Divided by $H_e \rightarrow$ *proportion* (of expected heterozygosity)

Varies between -1 and 1

$F < 0$: heterozygote excess

$F > 0$ heterozygote deficit (homozygote excess)

May be averaged over several loci \rightarrow reduces bias

May be extended to k alleles

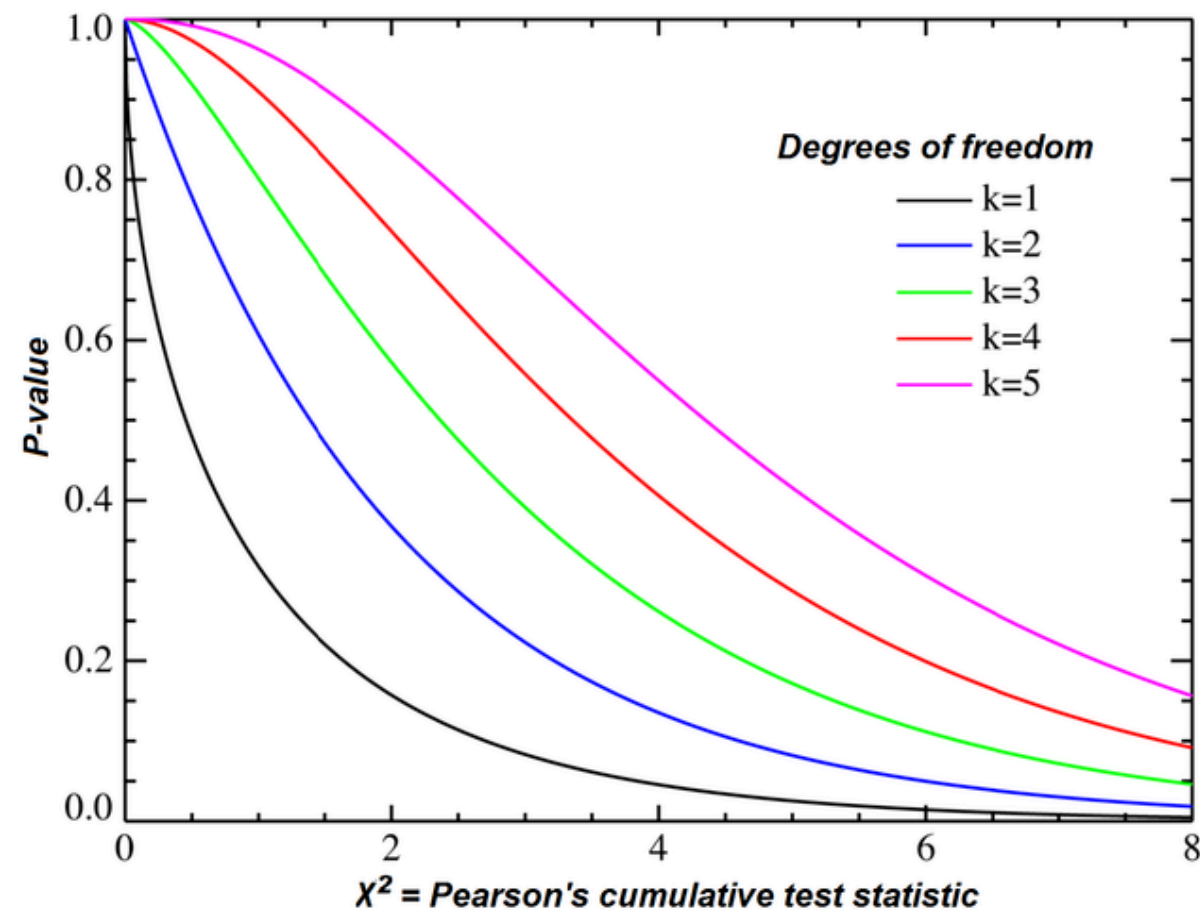
Pearson's chi-squared test

Exercise 1

Chi-square statistic:

$$\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}$$

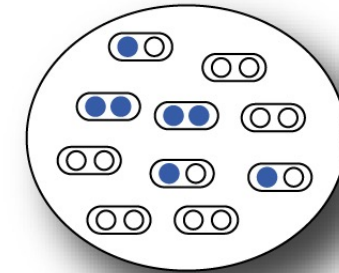
Chi-square distribution:



A single generation of reproduction will result in a population that meets the expected Hardy-Weinberg frequencies, i.e. is at Hardy-Weinberg (HW) equilibrium

Assuming an “ideal” population, i.e. :

- Diploid organisms
- Sexual reproduction (as opposed to clonal)
- Random mating (as opposed to e.g. assortative) with respect to genotype
- Random union of gametes
- Discrete, non-overlapping generations
- Very large (infinite) population
- No migration
- No population structure
- No natural selection
- Two alleles
- Identical allele frequencies in both sexes



-> Departures from HW equilibrium may indicate:

- Inbreeding
- Assortative mating
- Self-fertilization
- Natural selection
- Population structure
- ...