Hypothesis testing

Introduction - part 2

Rachel Warnock

13.02.2026

Recap
Introduction to hypothesis testing

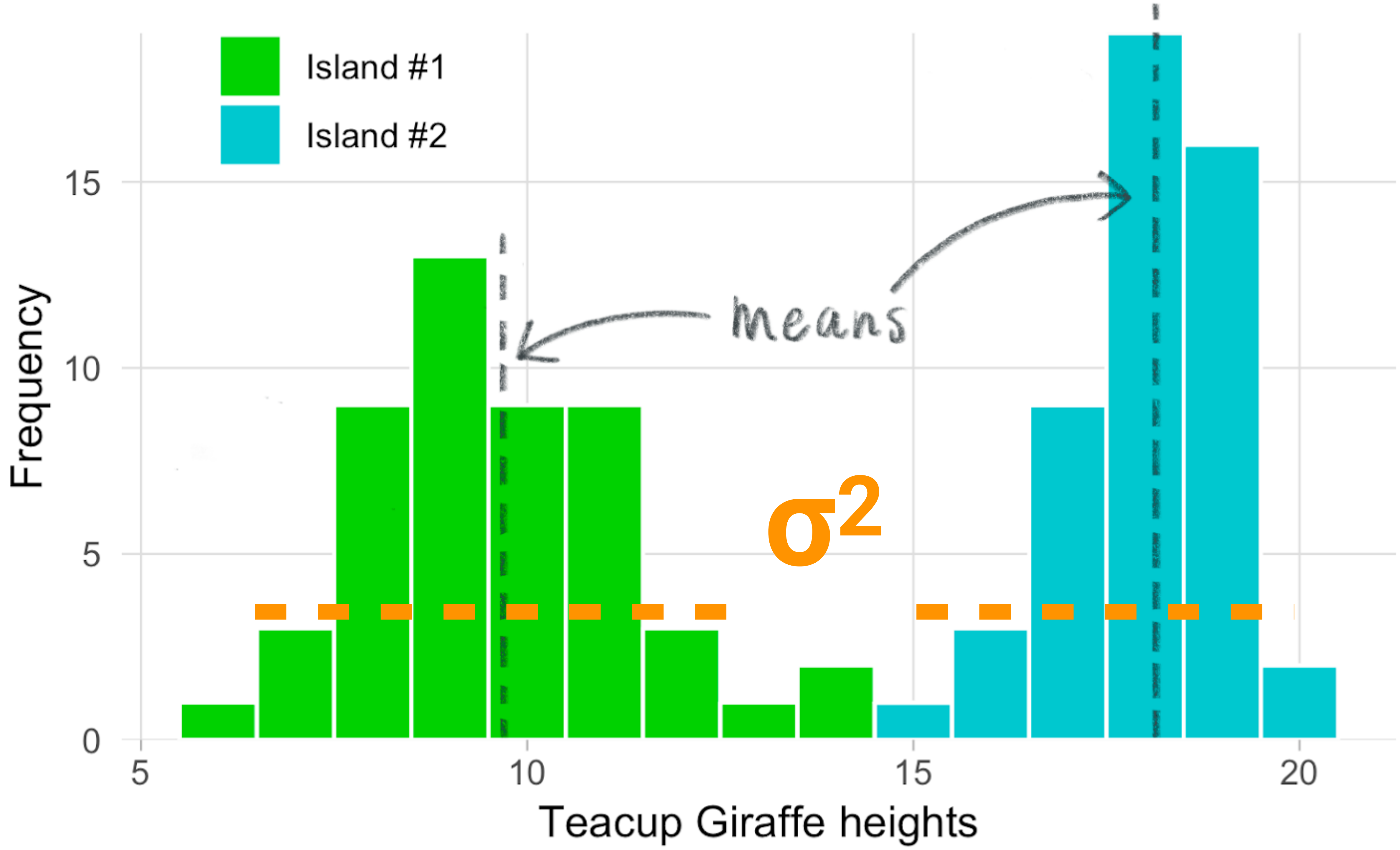
Learn more about the tiny giraffes @ tinystats.github.io

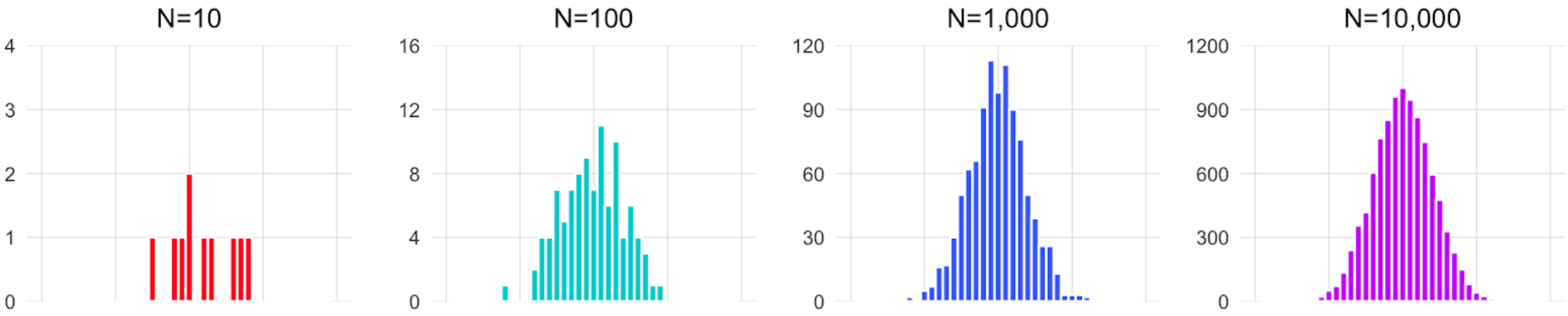
Parameters of the normal distribution

μ - the mean or expectation

σ - the standard deviation

or **σ2** - the variance



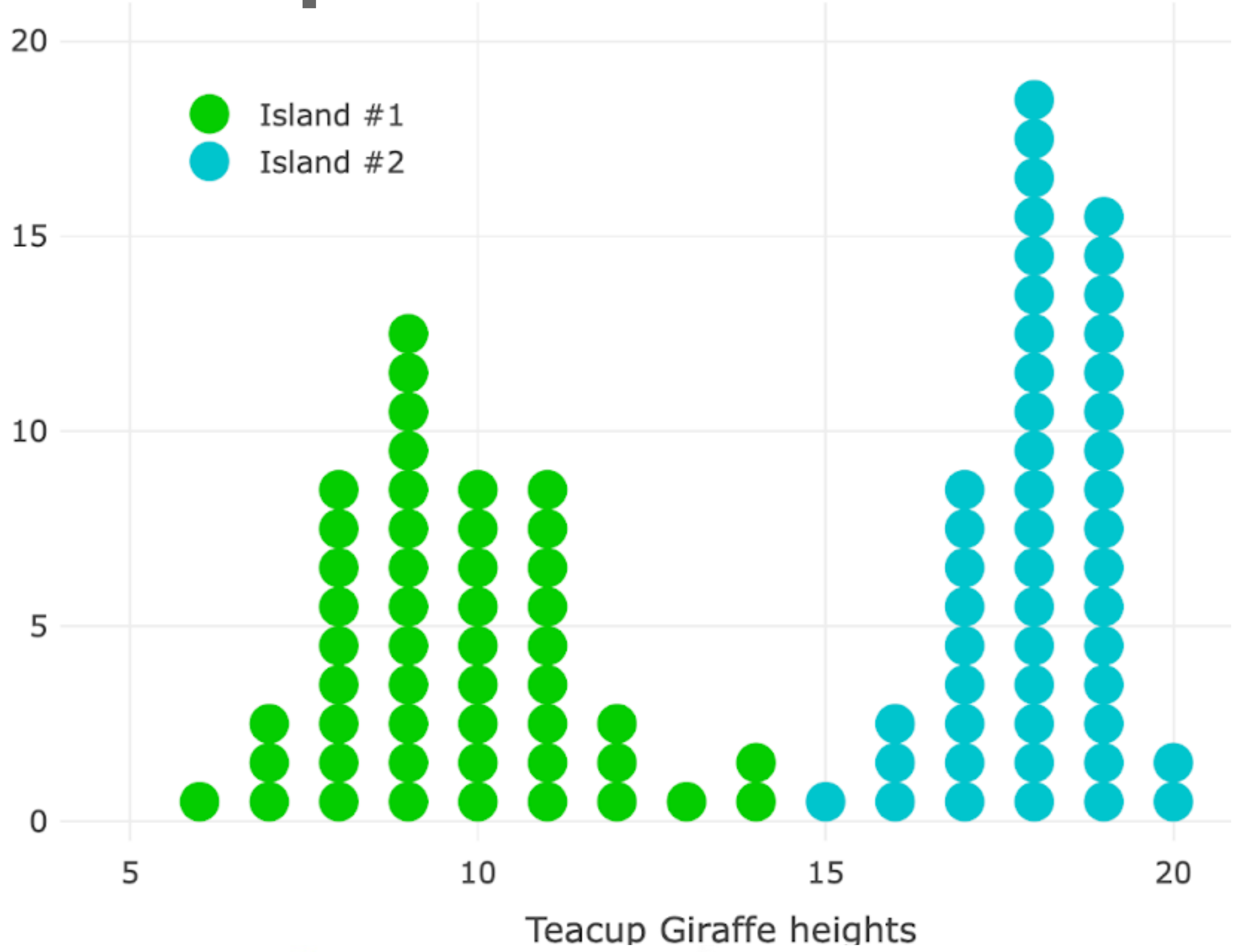


N=10
N=100
N=1,000
N=10,000

Differences between samples

Independent samples *t*-test

Is there a significant difference between the heights of giraffes on Island 1 and Island 2?



Defining hypothesis

Comparing two sample means

We have two sample means for Island 1 and Island 2

The **null hypothesis** H_0 is there is no difference between the means

The **alternative hypothesis** H_1 is that the means are different

→ If the H_0 is true the difference between the mean ~0, taking into account sample size / variance

The *t*-statistic

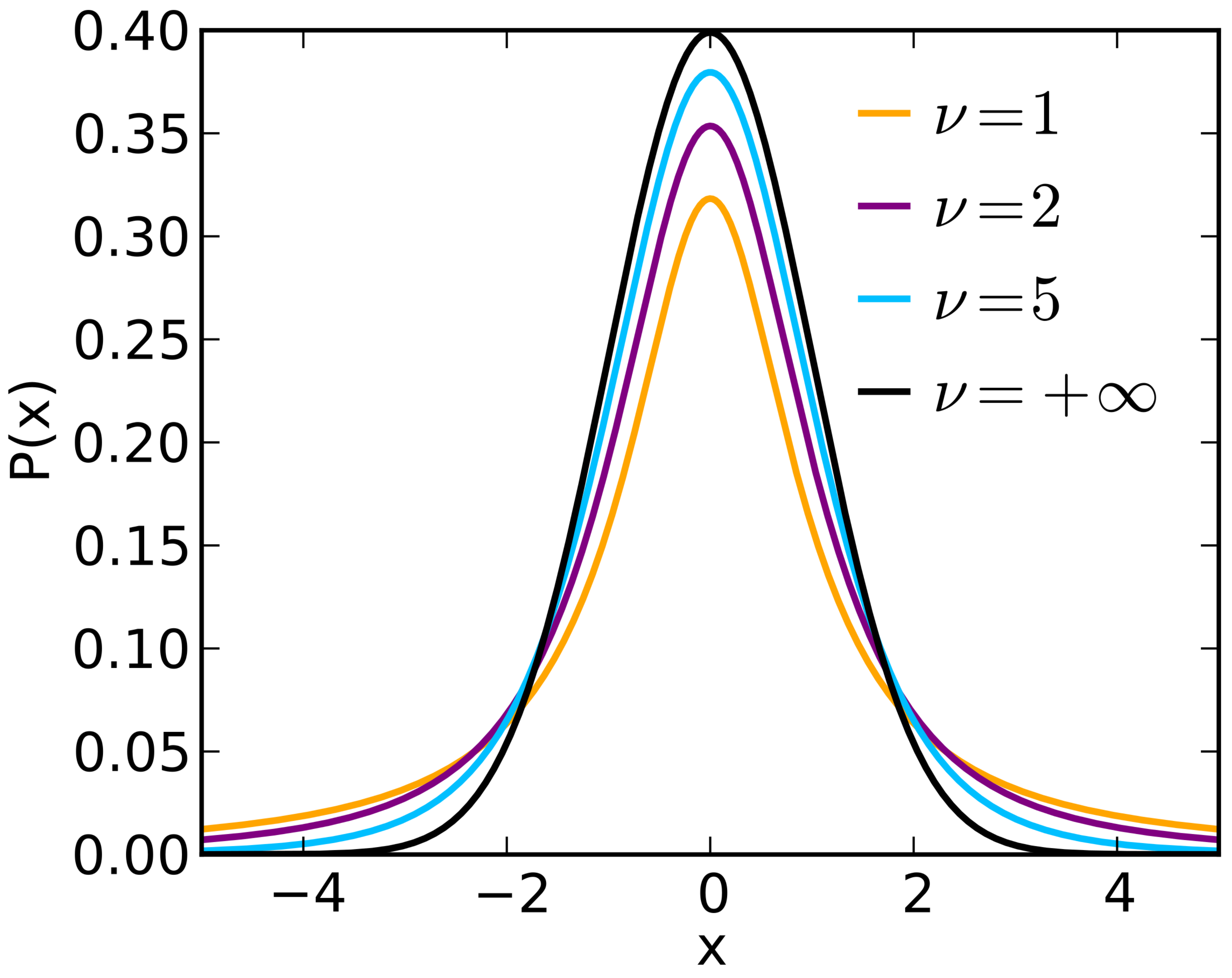
Independent samples *t*-test

Difference between the sample means

$$t = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}}$$

Standard deviation

If our hypothesis is true the mean of *t* will be close to zero



The parameters of the distribution are **degrees of freedom (n_1 + n_2 - 2)**, accounts for uncertainty in *s*

If H_0 is true, how extreme is our sample?

Intuitively, we are more likely to reject H_0 if:

- sample difference is greater
- number of observations is greater

Type I and II errors

A **type I error** occurs when we reject a null hypothesis that is *true*
Prob = α (also known as the alpha value)

A **type II error** occurs when we **accept** (don't reject) a null hypothesis that is *false*
Prob = β - beta (also known as the the beta value or the **power** of the test)

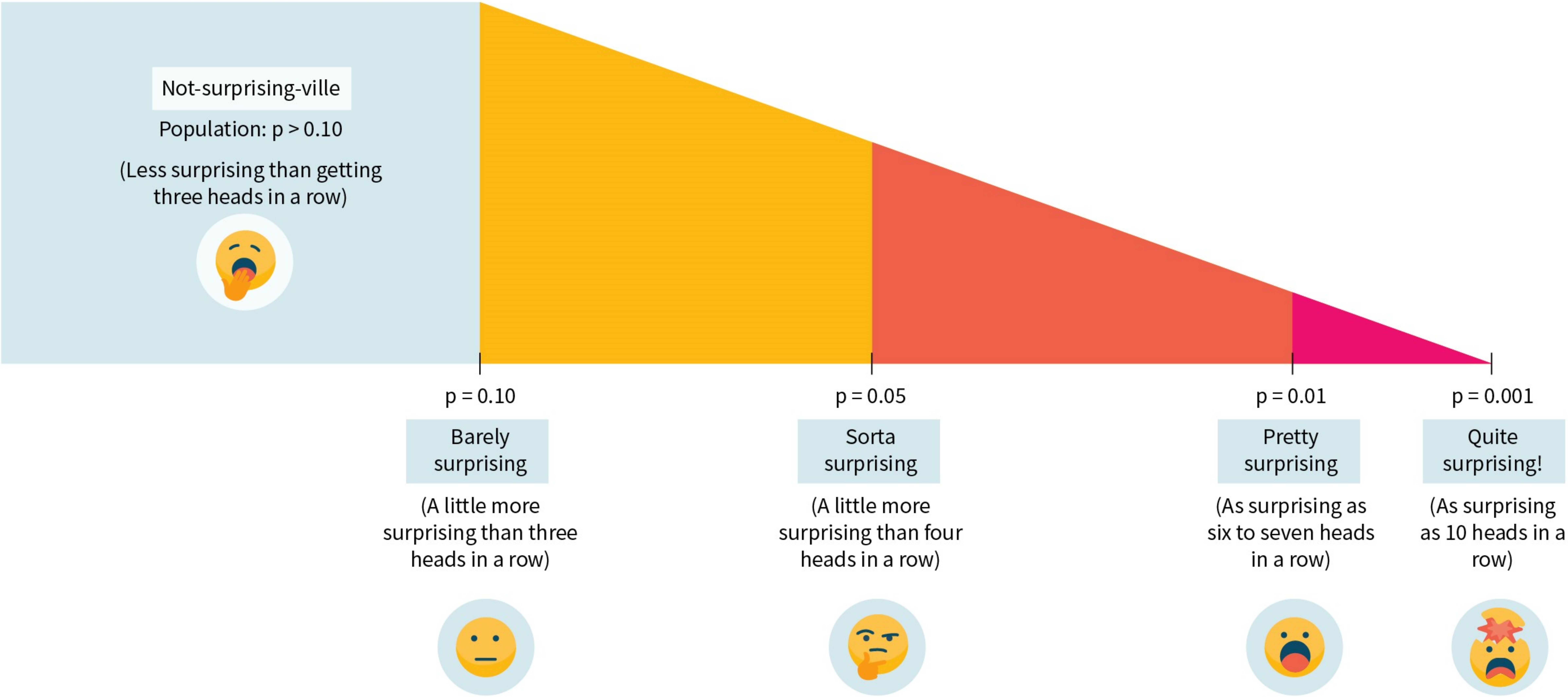
Type I and type II errors are *always* possible

p-values

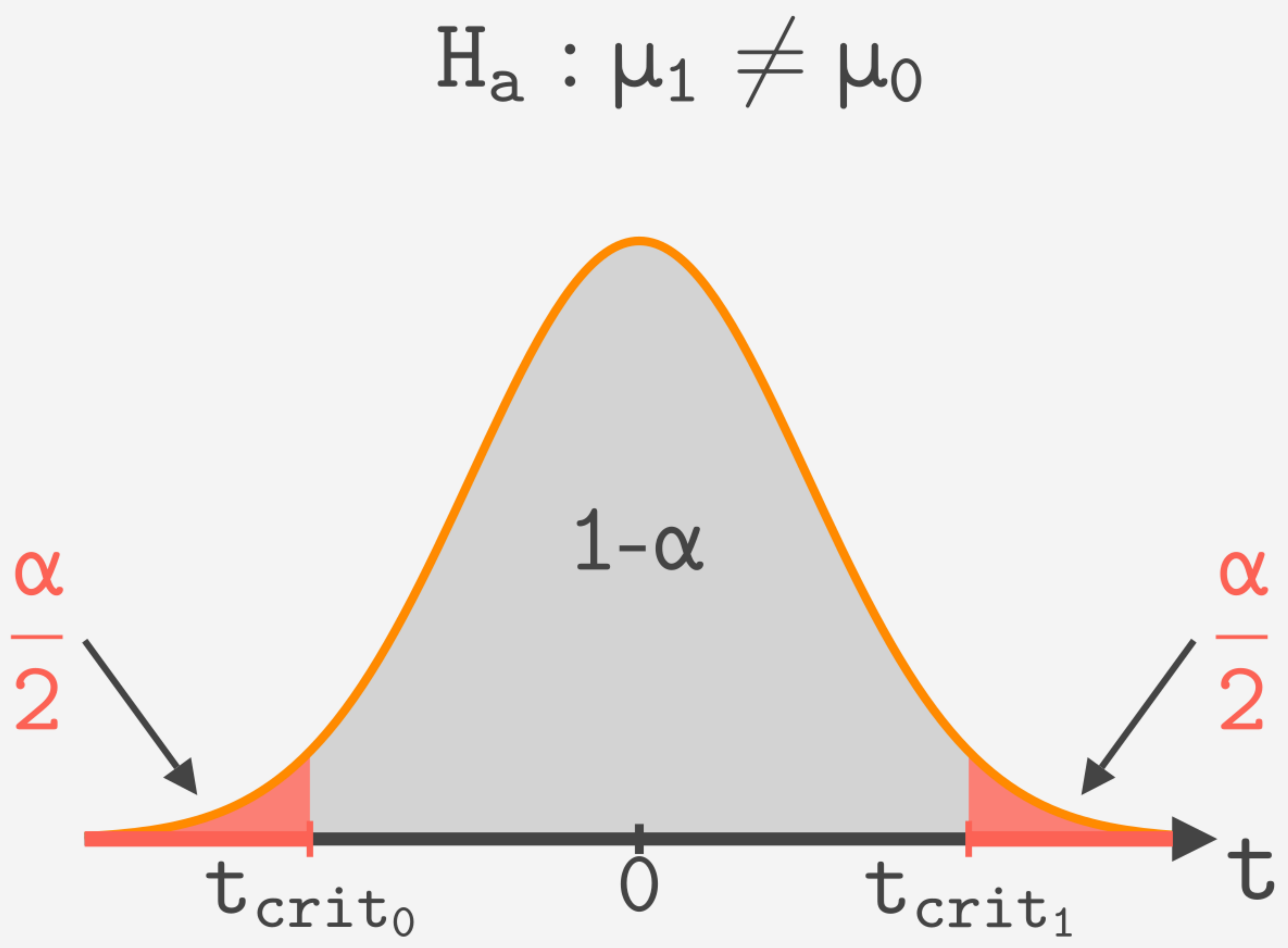
α is a predetermined threshold

The *p*-value is the probability of observing the results if the null hypothesis is true. This calculated from the data

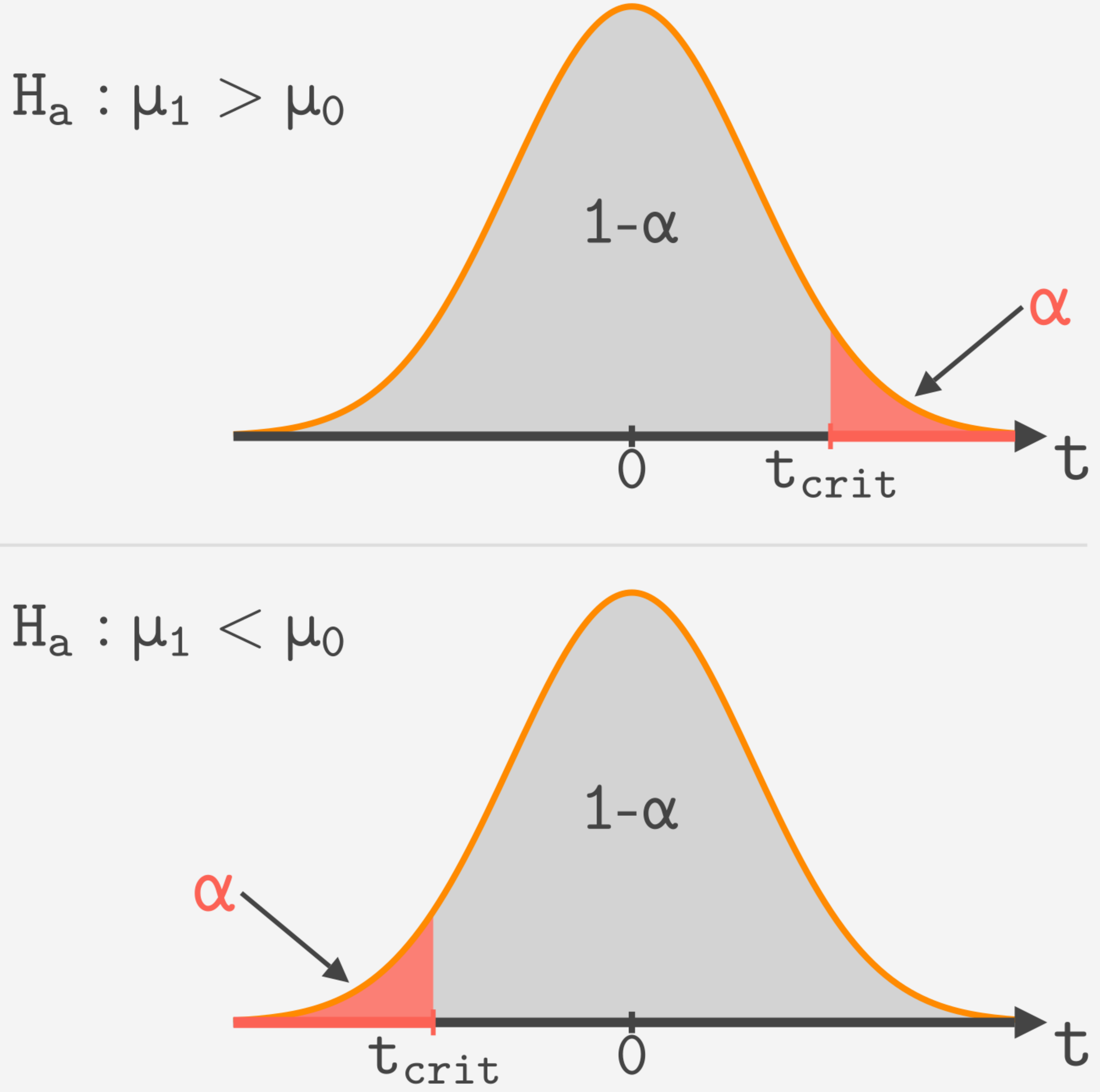
If the *p*-value is **less than or equal to** alpha, the result is considered **statistically significant**



Not-surprising-ville
Population: p > 0.10
(Less surprising than getting three heads in a row)
p = 0.10
Barely surprising
(A little more surprising than three heads in a row)
p = 0.05
Sorta surprising
(A little more surprising than four heads in a row)
p = 0.01
Pretty surprising
(As surprising as six to seven heads in a row)
p = 0.001
Quite surprising!
(As surprising as 10 heads in a row)



Two — Tailed Test



One — Tailed Test

Steps in hypothesis testing

1. Define your research question and hypothesis
2. Define your statistical hypothesis (null & alternative)
3. Find an appropriate test & sampling distribution
4. Choose the type I error rate, i.e., your alpha value

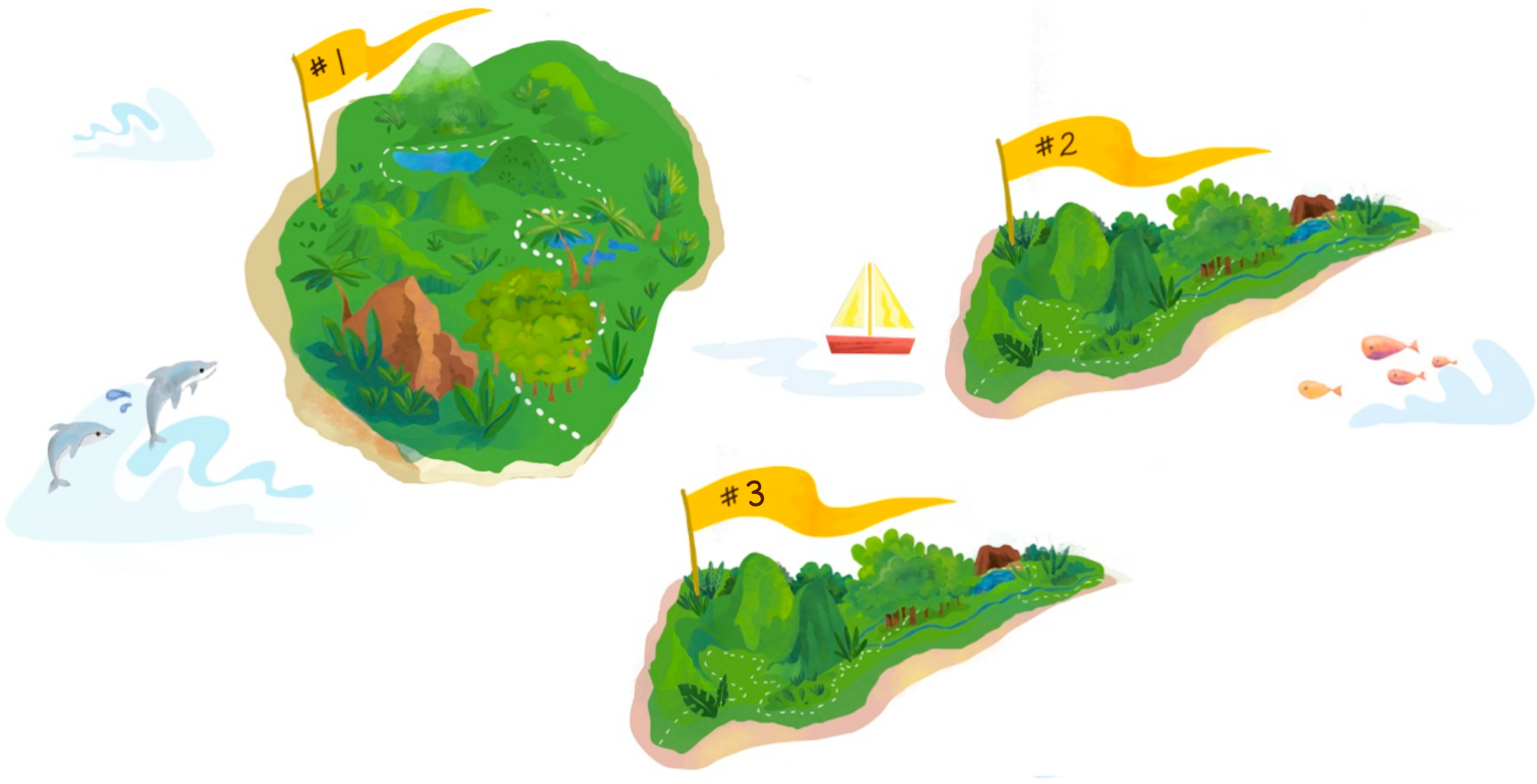
Steps in hypothesis testing

5. Collect the data
6. Calculate test statistics
7. State the **statistical** conclusion
8. Interpret your results

ANOVA

Analysis of variance

>2 groups



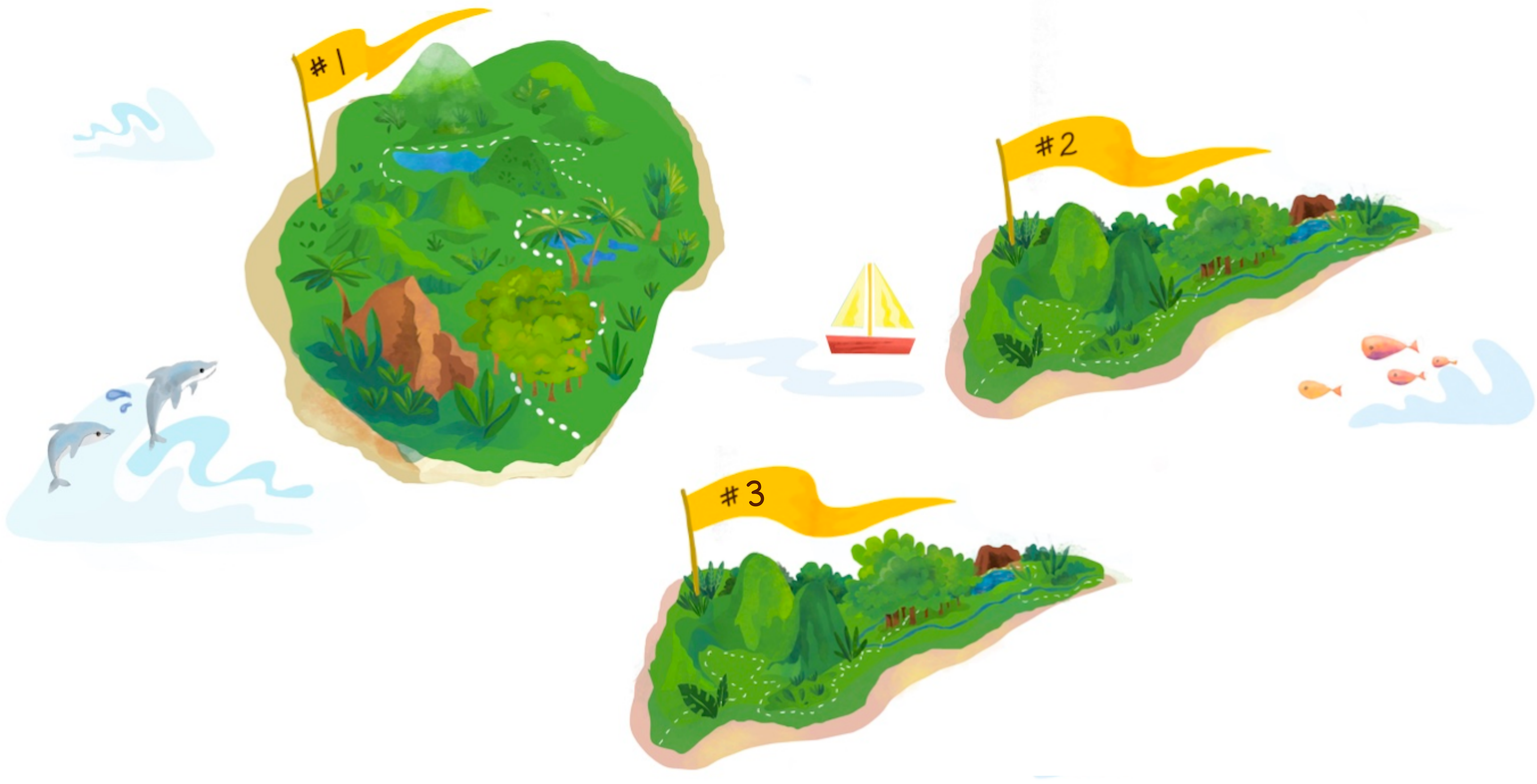
Imagine we've collected data for **three** populations that live on **three** different islands

What is our null hypothesis?

The null hypothesis is there is **no difference between the means**

The alternative hypothesis H_1 is that the means are different

>2 groups

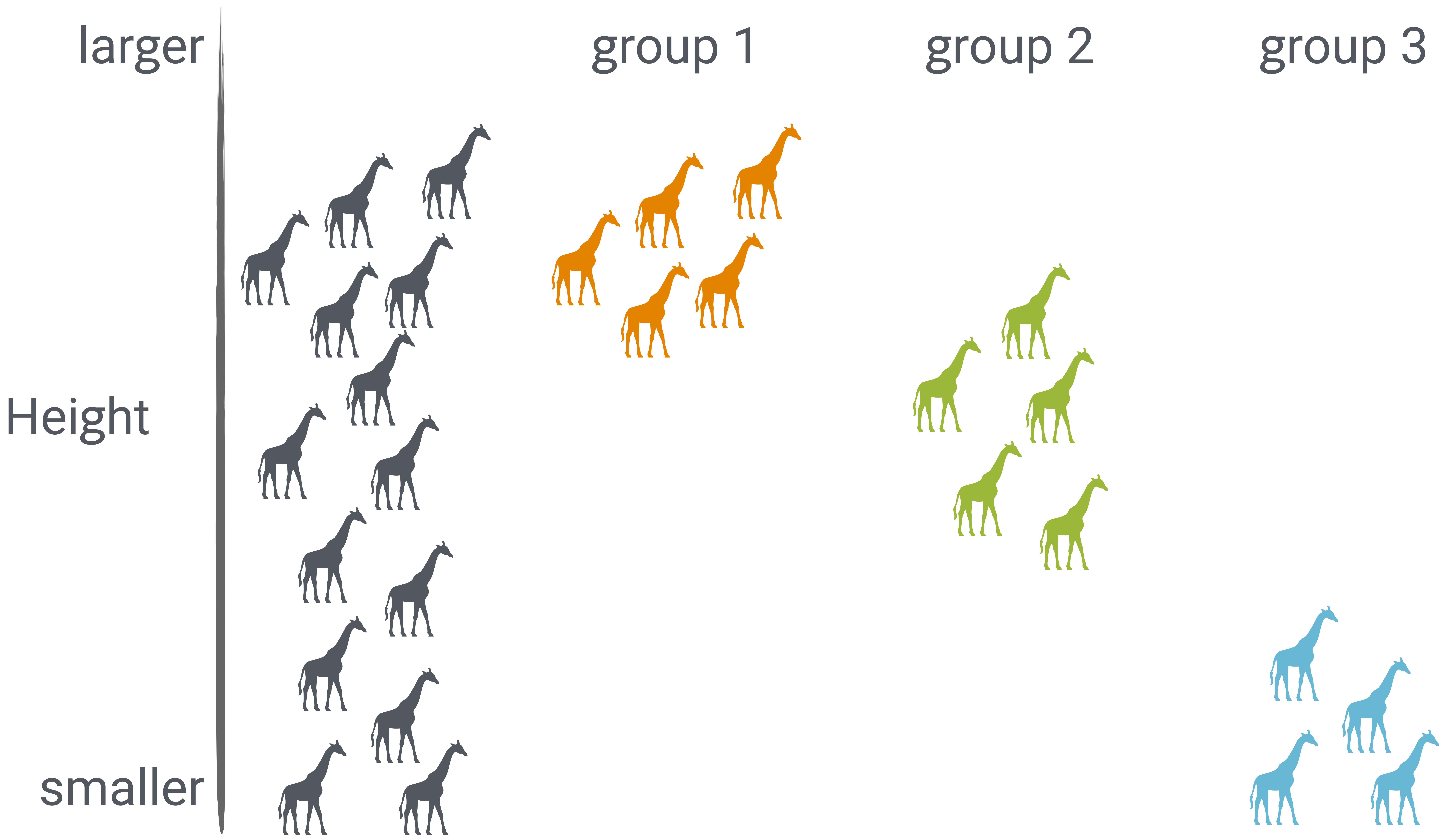


Imagine we've collected data for **three** populations that live on **three** different islands

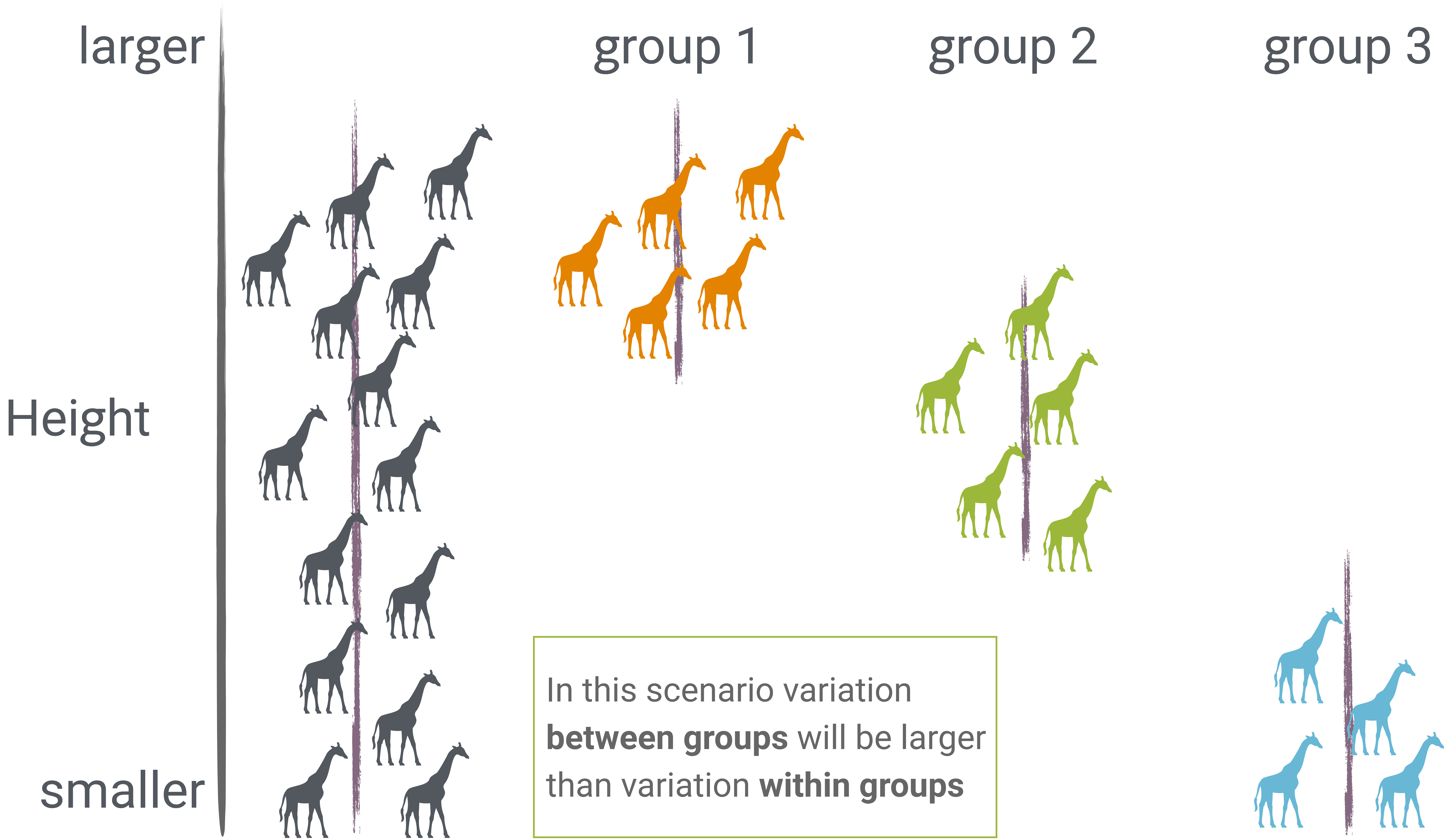
What is our null hypothesis?

H_0: $\mu_1 = \mu_2 = \mu_3$

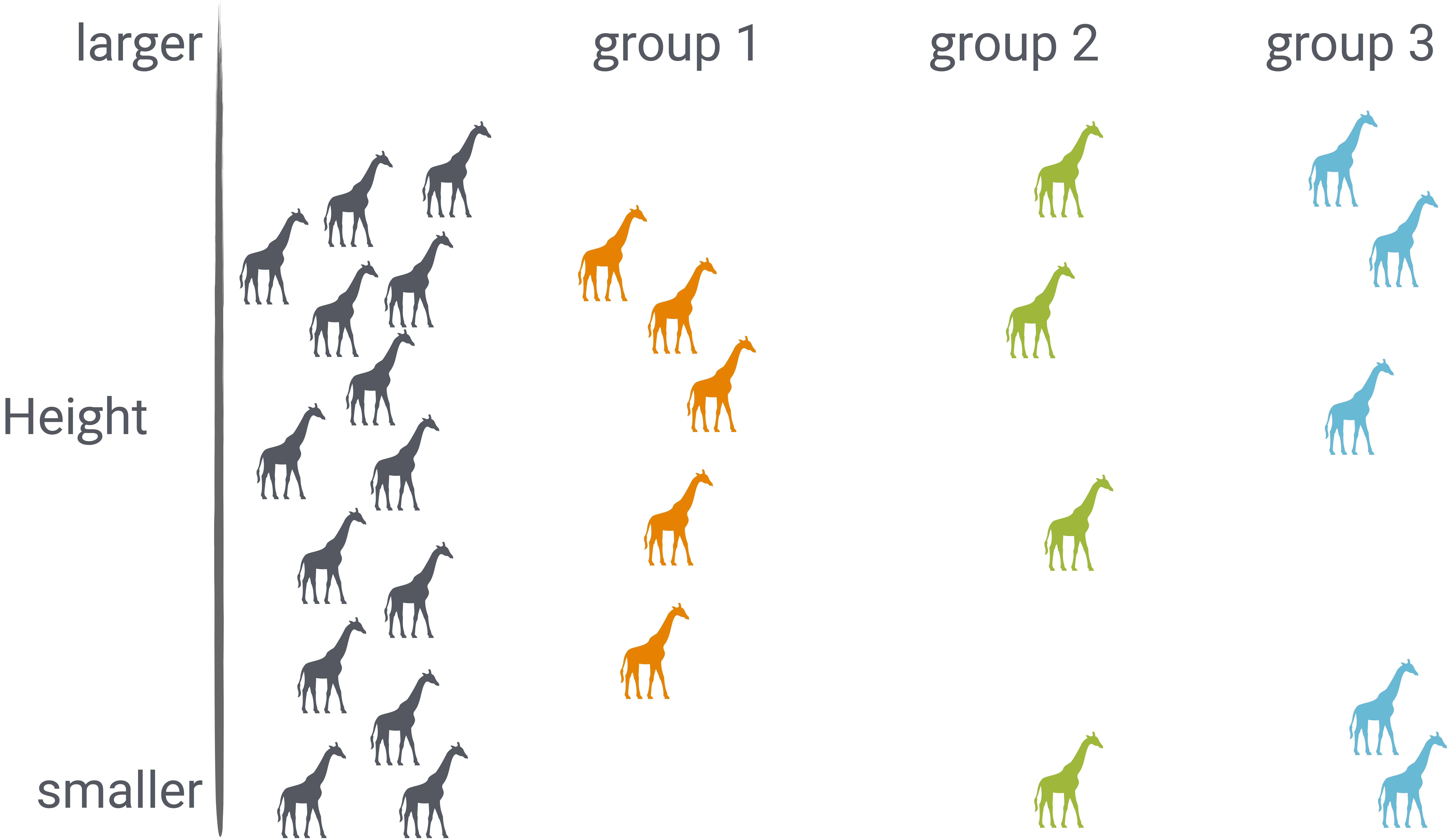
H_1: at least one group mean differs

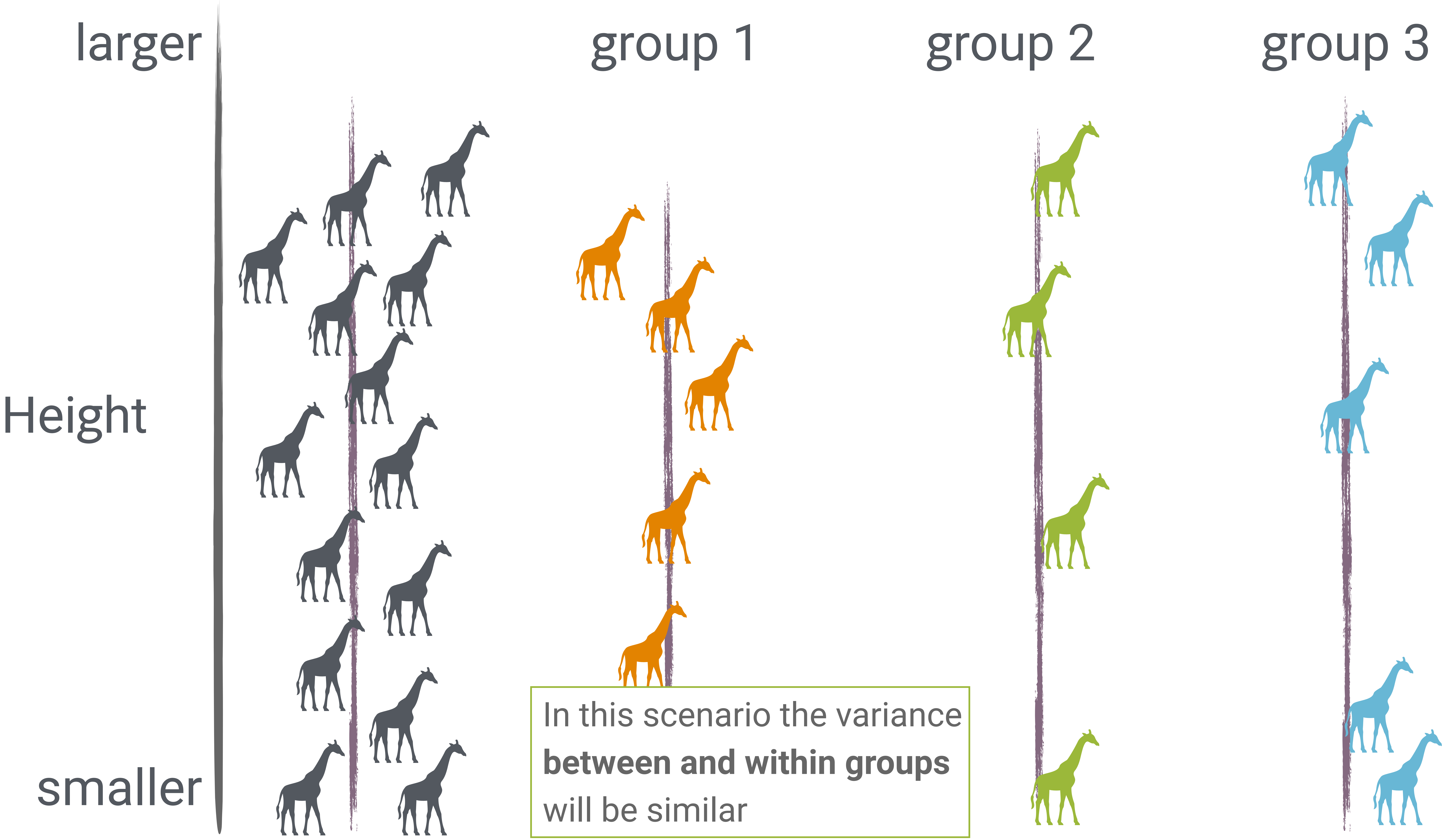


larger
Height
smaller
group 1
group 2
group 3



In this scenario variation **between groups** will be larger than variation **within groups**


larger
group 1
group 2
group 3
Height
smaller



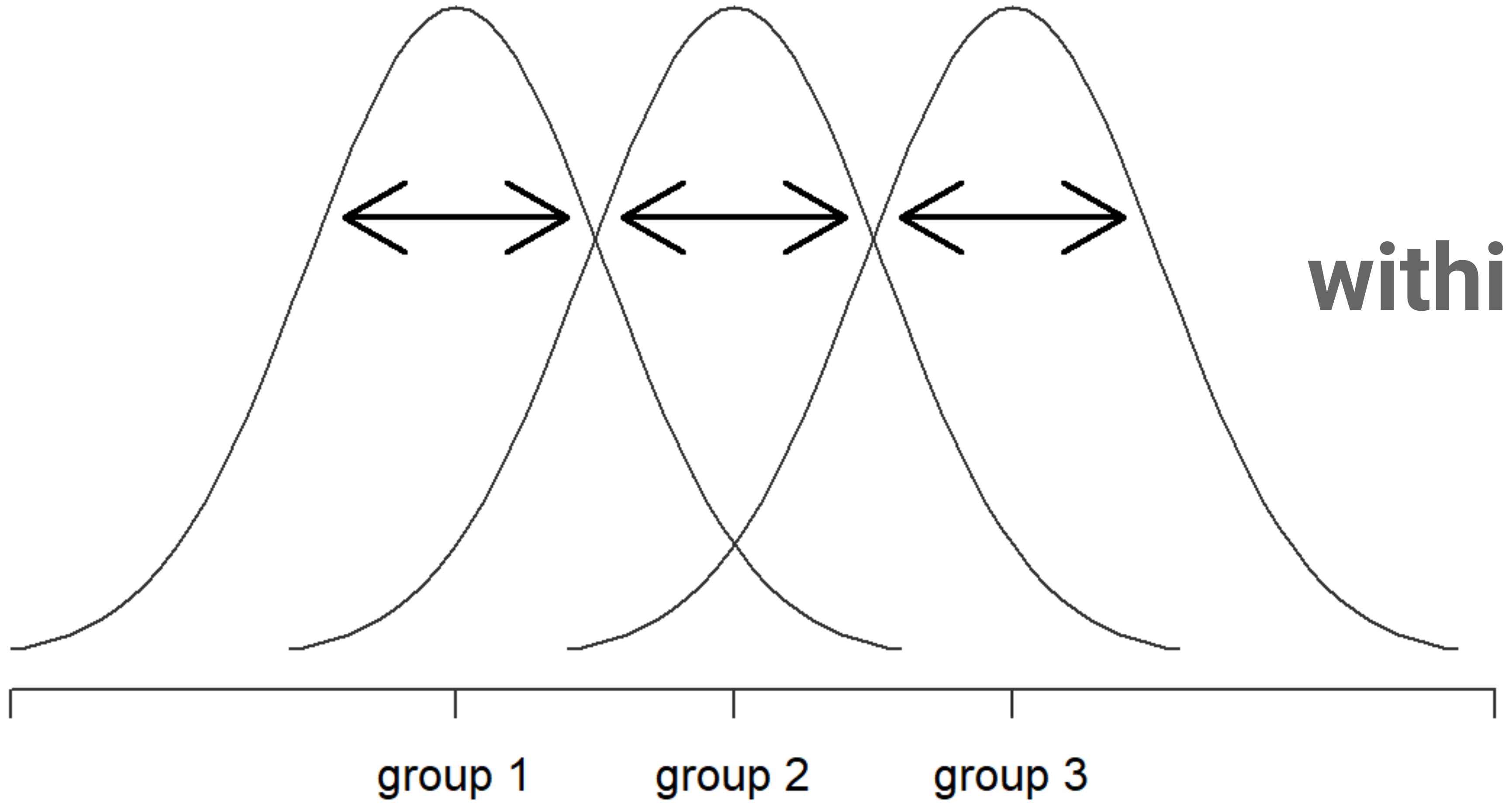
In this scenario the variance **between and within groups** will be similar

ANOVA

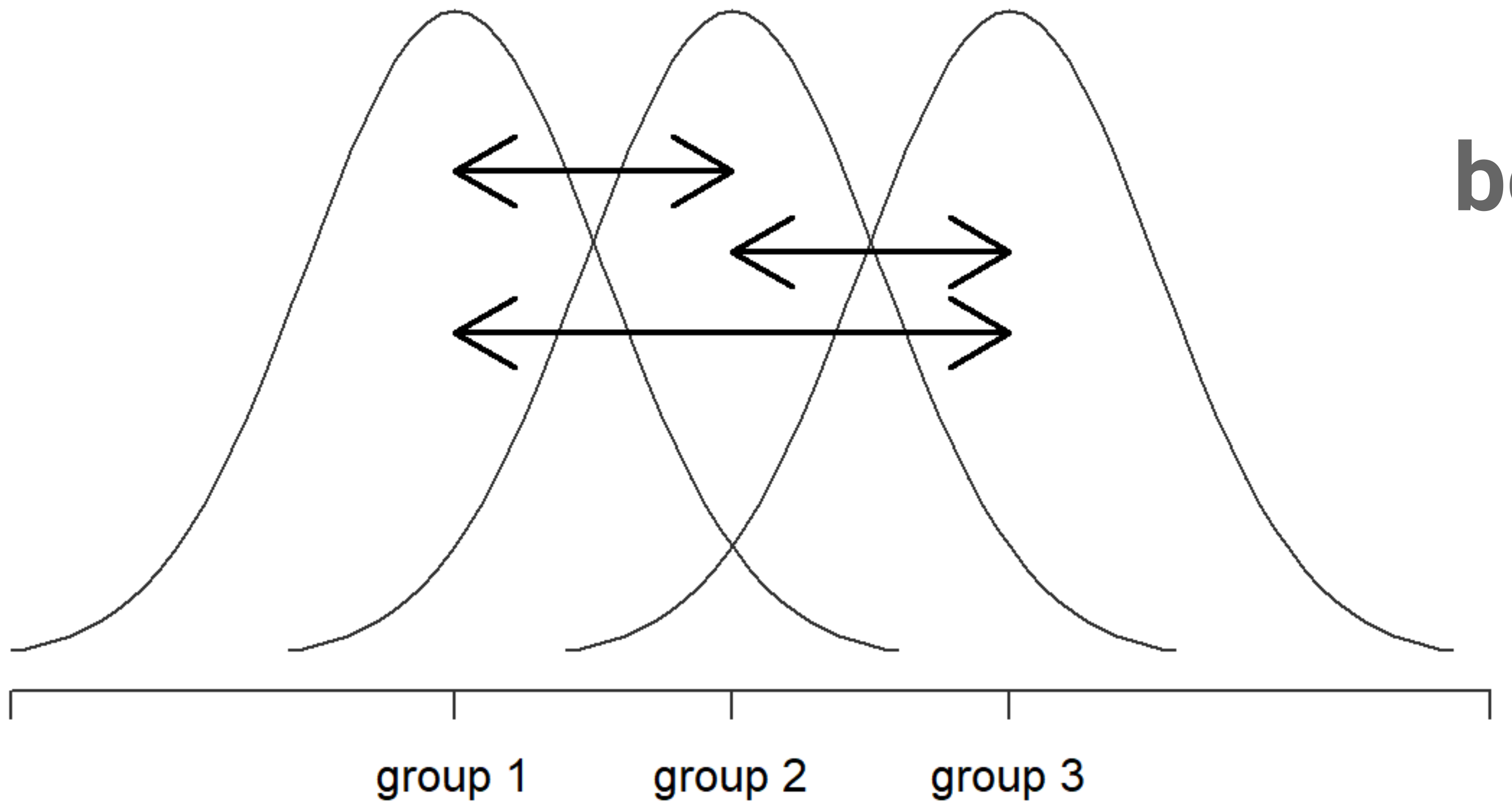
There is more than one way that the means could differ - we can't use a *t*-test

Instead we look at the variance

If H_0 is true, variation between the means is no larger than expected from within group variability



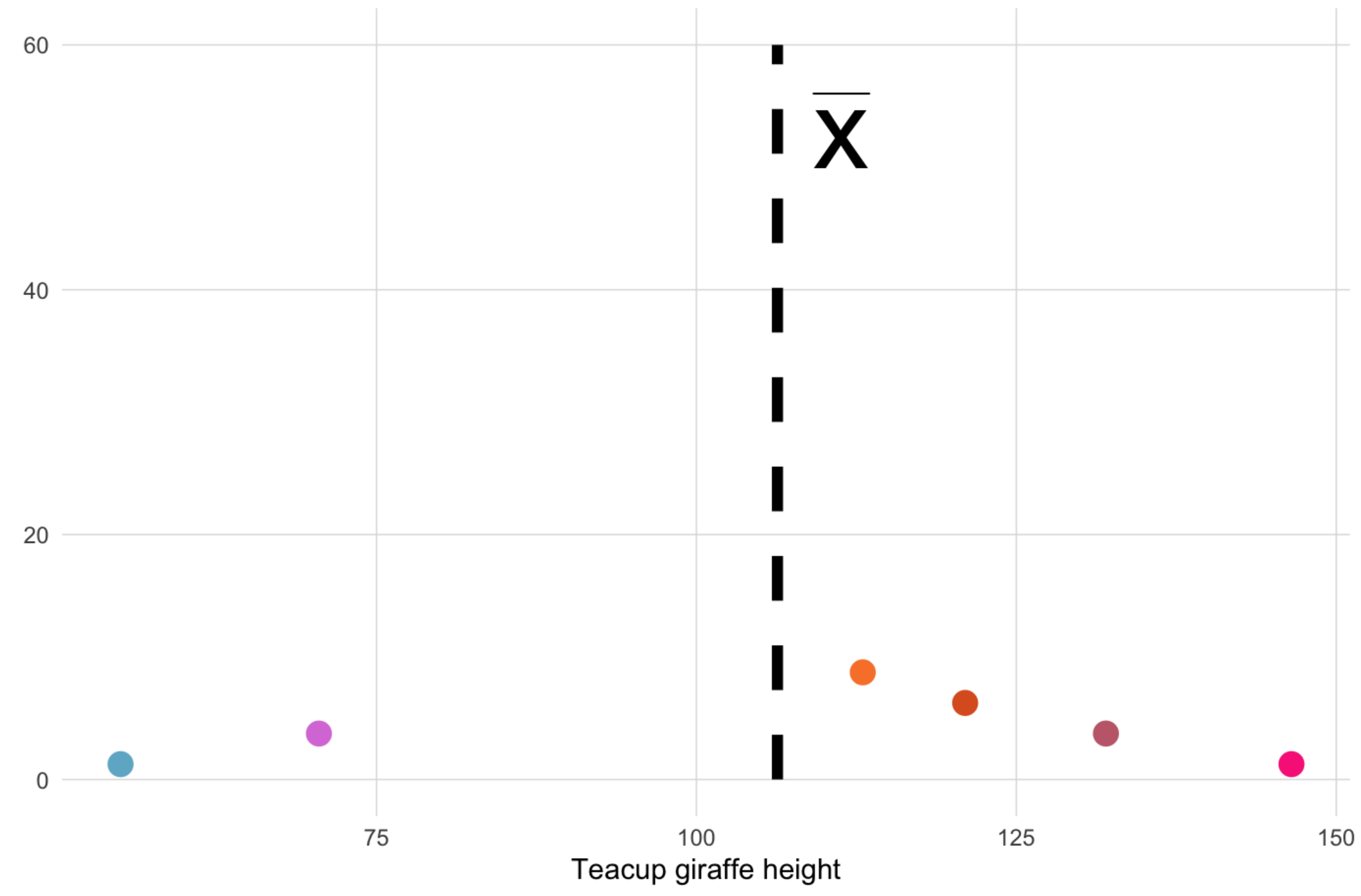
within group variance



between group variance

ANOVA

Steps



Calculate the sum of squares for:

Sum of squares between groups (SSB)

Sum of squares within groups (SSW)

Note: Sum of squares total (SST)

And SST = SSB + SSW, under H_0 is SSW accounts for almost all SST

ANOVA

Steps

Degrees of freedom:

Between groups: df = k - 1

Within groups: df = N - k

Note for the total sample: df = N - 1

Compute mean squares

$$MS_{\text{between}} = \frac{SS_{\text{between}}}{k-1}$$

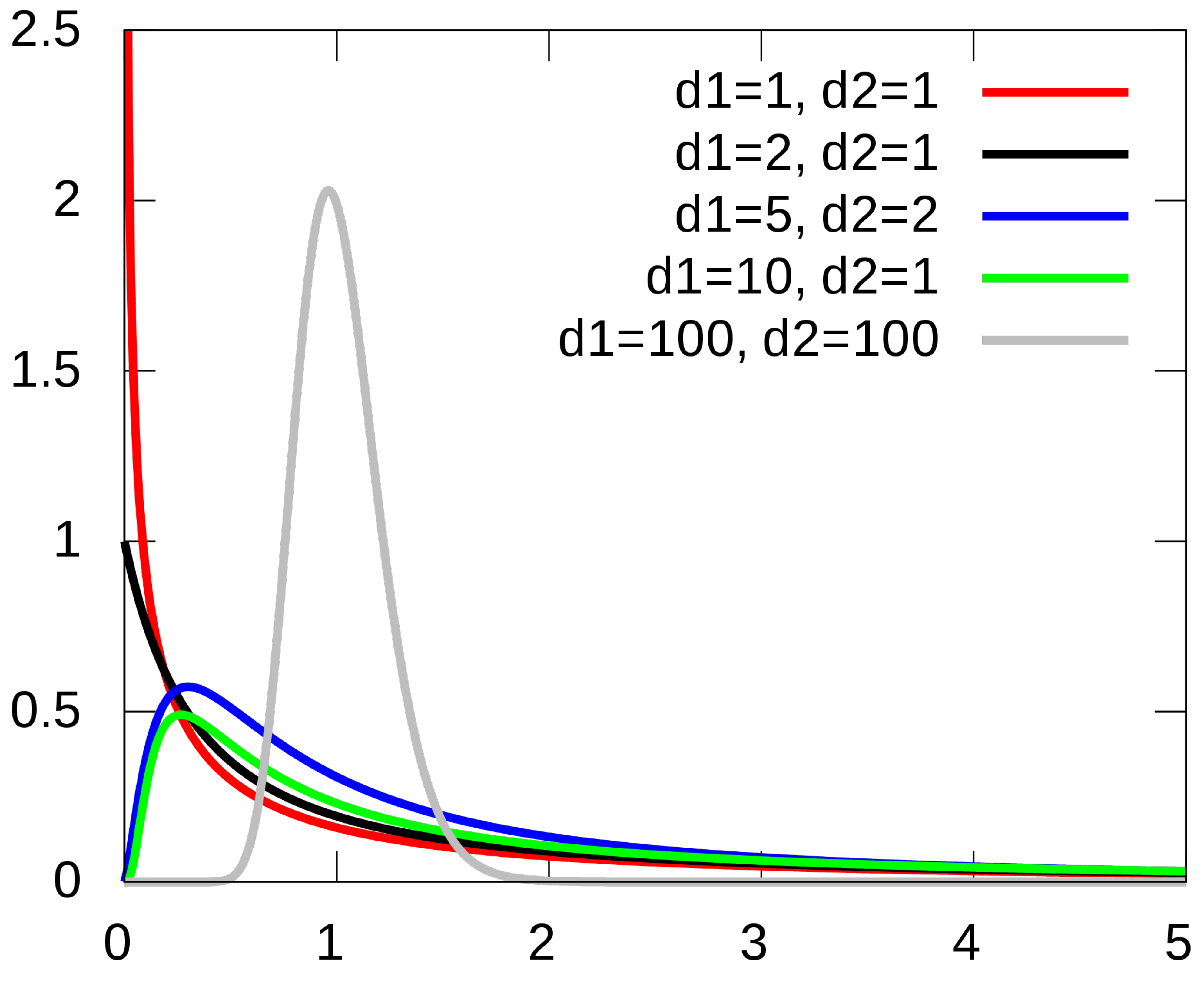
$$MS_{\text{within}} = \frac{SS_{\text{within}}}{N-k}$$

The *F*-statistic

Calculate the *F*-statistic

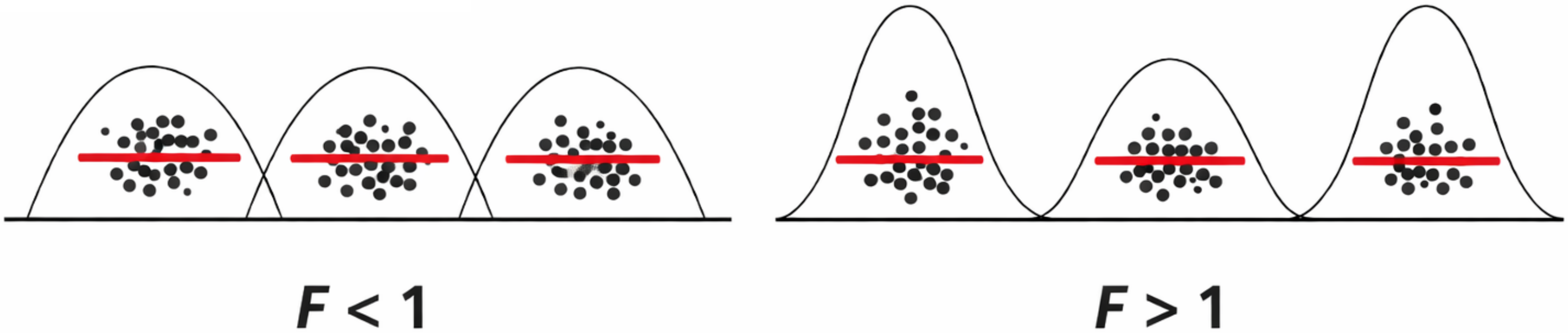
$$F = \frac{MS_{\text{between}}}{MS_{\text{within}}}$$

Compare this to the *F*-distribution, with $df_1 = k - 1$ and $df_2 = N - k$, calculate *p*-value



If MSB and MSW are similar, *F* ~ 1, if F >> 1 H_0 is rejected

$$F = \frac{MS_{\text{between}}}{MS_{\text{within}}}$$



MSW > MSB
More variation
within groups than
between groups

MSB > MSW
More variation
between groups
than **within** groups

www.desmos.com/calculator/1ha4o1c2sq

Kruskal-Wallis test

If the data are not normally distributed

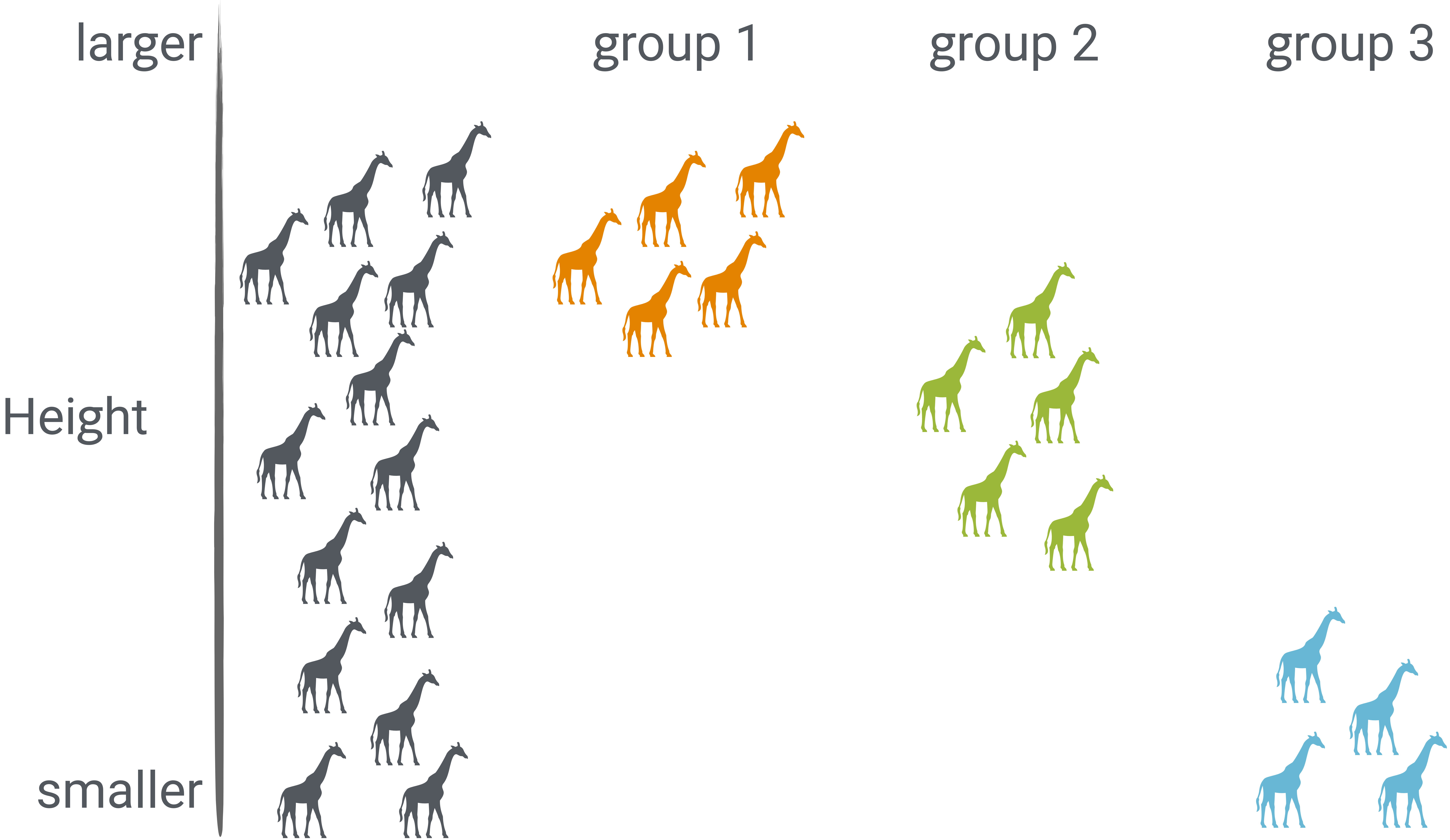
Parametric vs. non-parametric statistics

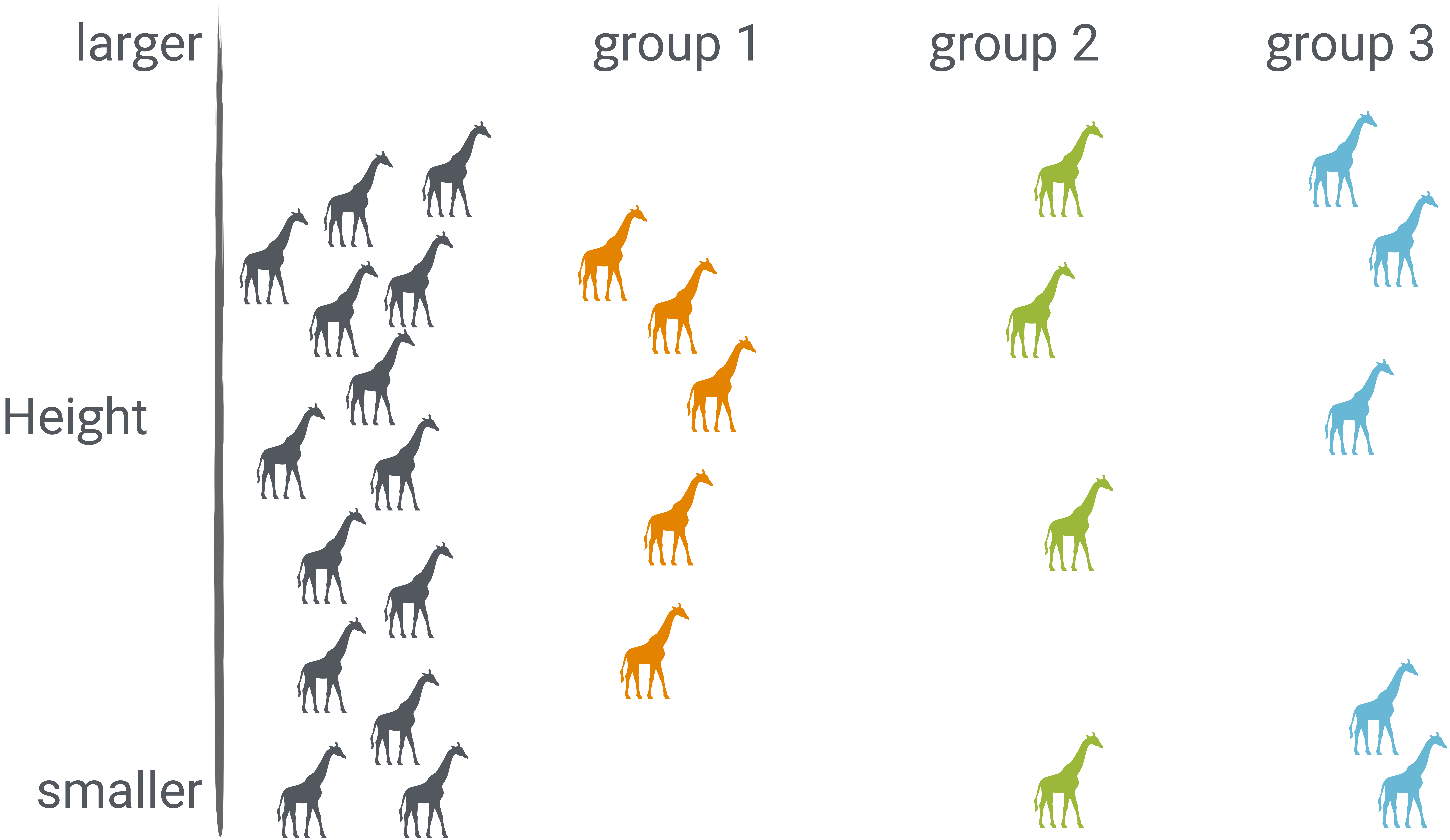
Parametric statistics are based on assumptions about the distribution of population

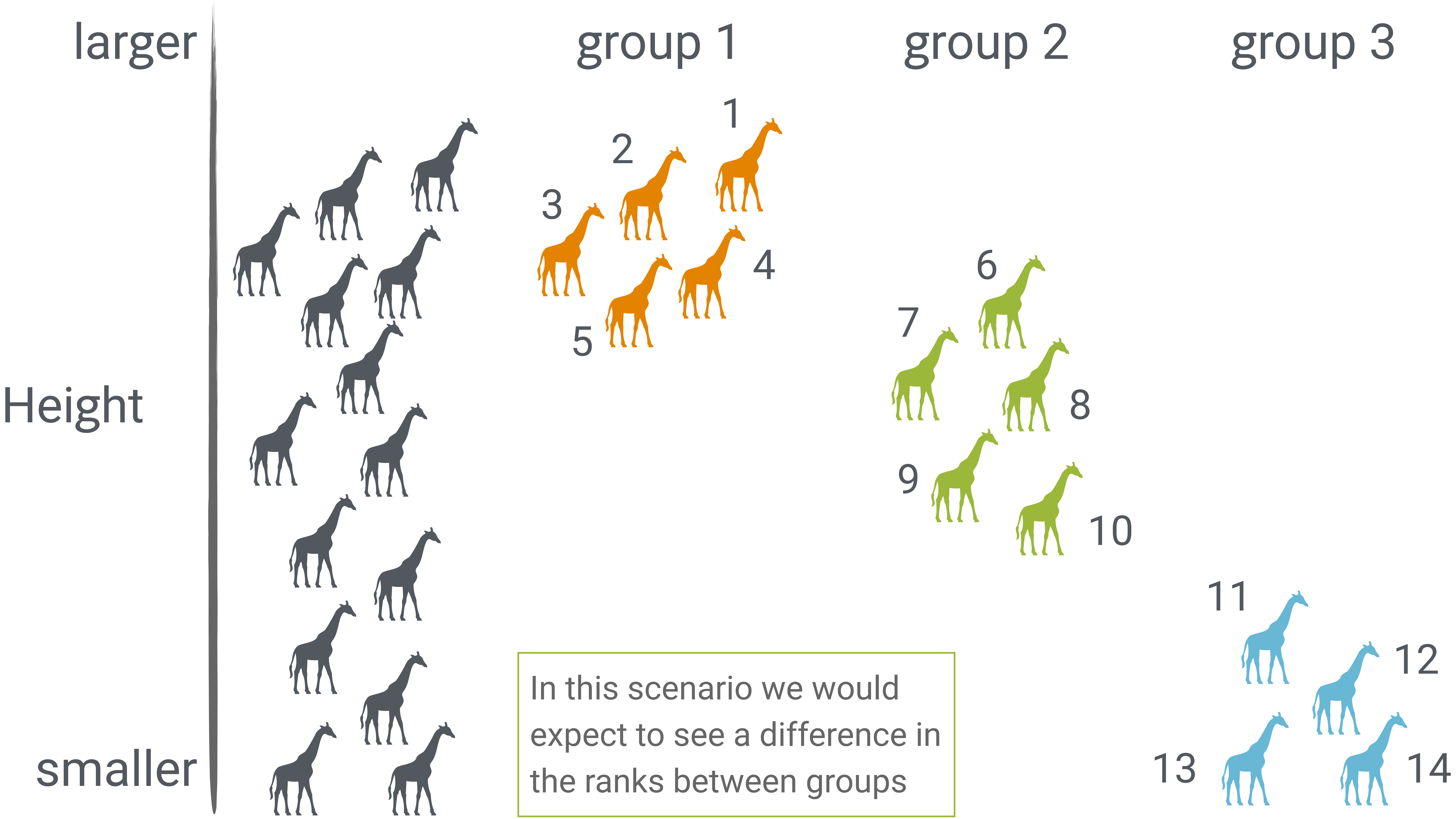
Non-parametric statistics do not make this assumption — the data can be collected from a sample that does not follow a specific distribution

If your data are **normally distributed** you can use a **parametric test**

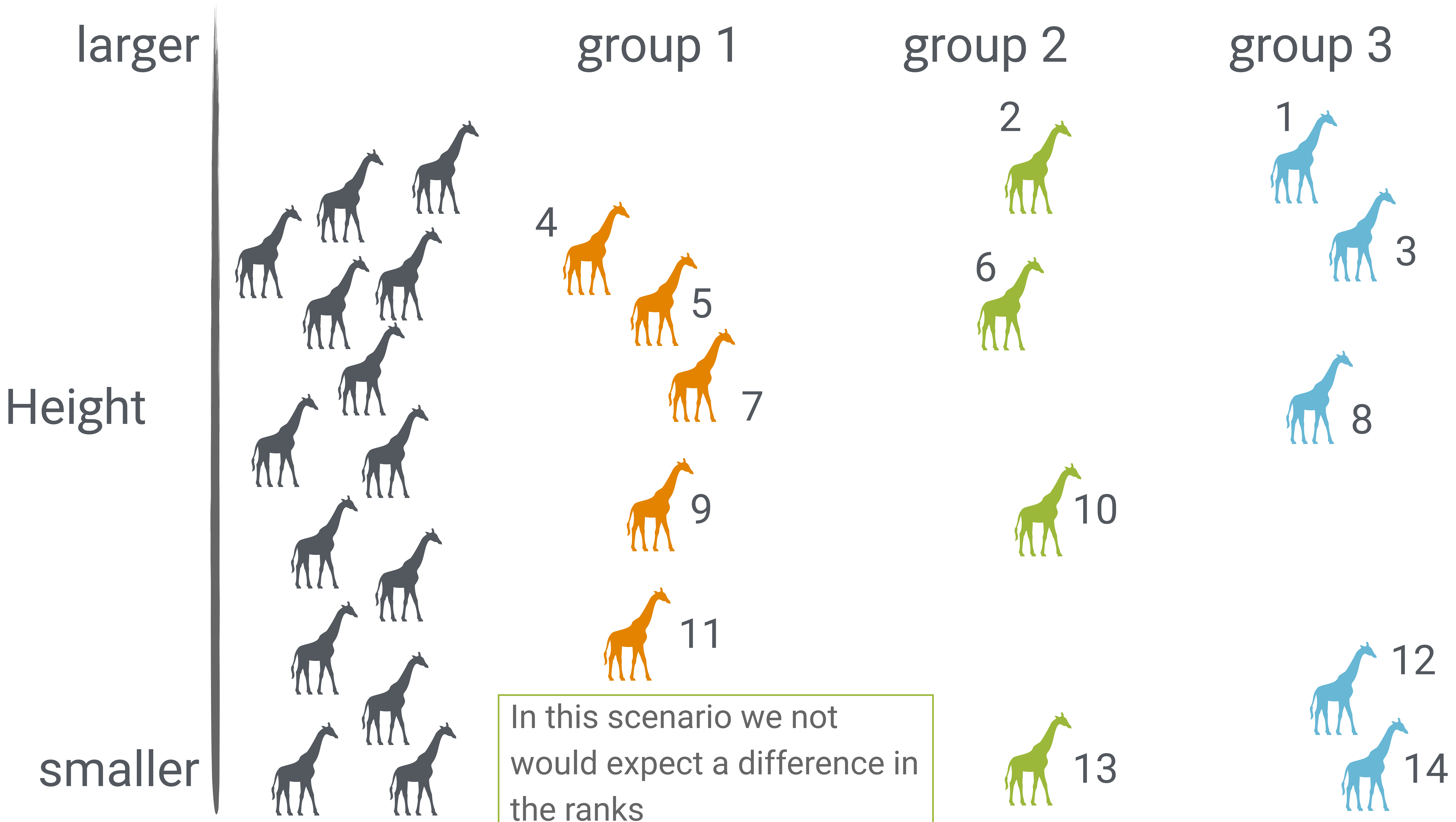
If your data are **not normally distributed** you use a **non-parametric test**


larger
group 1
group 2
group 3
Height
smaller


larger
group 1
group 2
group 3
Height
smaller



In this scenario we would expect to see a difference in the ranks between groups



In this scenario we not would expect a difference in the ranks

Steps involved in Kruskal-Wallis test

Define your hypothesis:

H_0 = all groups come from the same distribution

H_1 = at least one group tends to have larger or smaller values than at least one other group

Combine all data points from all groups and assign ranks

Put the data back into groups and calculate the mean

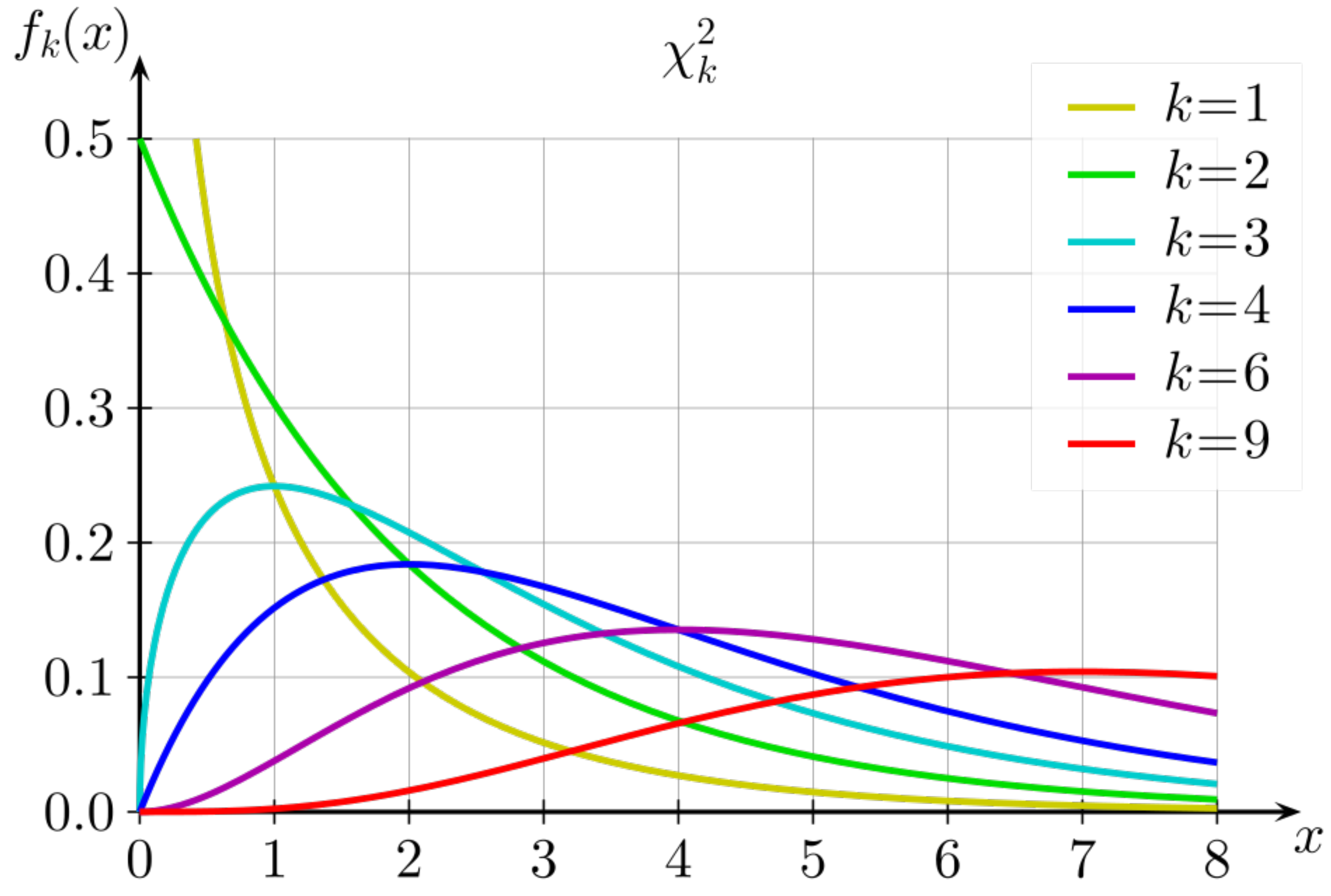
Steps involved in Kruskal-Wallis test

Calculate the Kruskal-Wallis statistic, *H*

Small *H* → ranks are well mixed

Large *H* → ranks are unevenly distributed

Compare this to the X^2-distribution, with df = *k* - 1, calculate the *p*-value



The Kruskal-Wallis statistic, *H*

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(N+1)$$

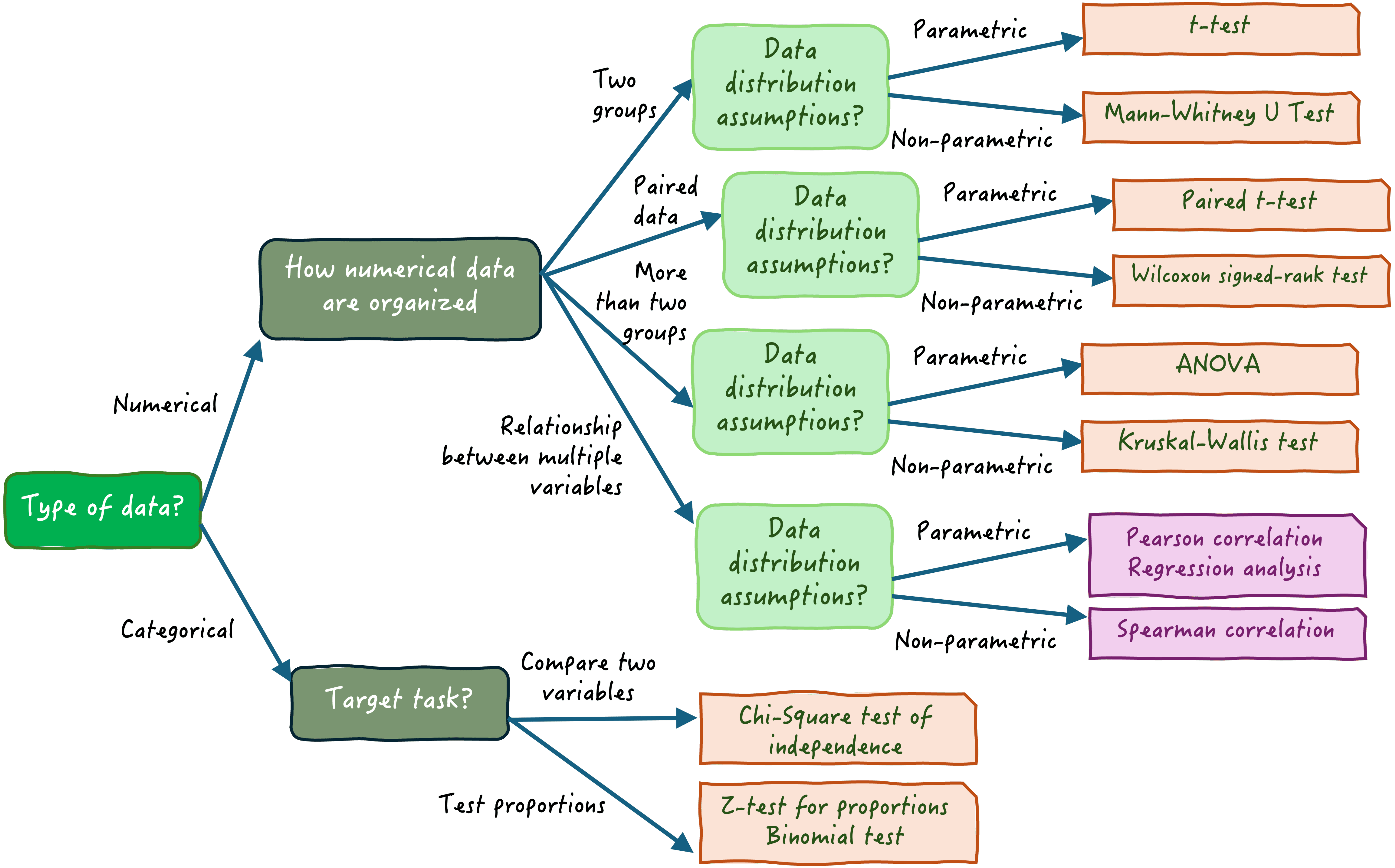
k = number of groups

N = total number of observations

R = sum of ranks in group

	Parametric Tests	Nonparametric Tests
One Sample	Simple t-Test	Wilcoxon test for one sample
Two dependent samples	Paired Sample t-Test	Wilcoxon Test
Two inpendent samples	Unpired Sample t-Test	Mann-Whitney U Test
More than two independent samples	One factorial ANOVA	Kruskal-Wallis Test
More than two dependent samples	Repeated Measures ANOVA	Friedman Test
Correlation between two variables	Pearson-Korrelation	Spearman-Korrelation

How to choose a statistical test


Type of data?
Numerical
How numerical data are organized
Two groups
Data distribution assumptions?
Parametric
t-test
Non-parametric
Mann-Whitney U Test
Paired data
Data distribution assumptions?
Parametric
Paired t-test
Non-parametric
Wilcoxon signed-rank test
More than two groups
Data distribution assumptions?
Parametric
ANOVA
Non-parametric
Kruskal-Wallis test
Relationship between multiple variables
Data distribution assumptions?
Parametric
Pearson correlation
Regression analysis
Non-parametric
Spearman correlation
Categorical
Target task?
Compare two variables
Chi-Square test of independence
Test proportions
Z-test for proportions
Binomial test

You do **not** need to remember all this!!

3 important concepts you need to navigate this chart

Types of data

Paired vs. **unpaired** (are your data points across samples independent)

Parametric vs. **non-parametric**

www.graphpad.com/support/faqid/1790/

Can you propose the statistical hypothesis?

Question: *Does temperature influence species richness across elevation gradients?*

Hypothesis: *Species richness decreases with increasing elevation.*

Question: *Is sedimentation rate associated with fossil preservation quality?*

Hypothesis: *Higher sedimentation rates lead to better fossil preservation.*

Reading group

Group exercise

In your groups, discuss the following:

- What type of data is in your paper?
- What - if any - statistical tests are used?
- How are *p*-values interpreted?

Make a note 📝