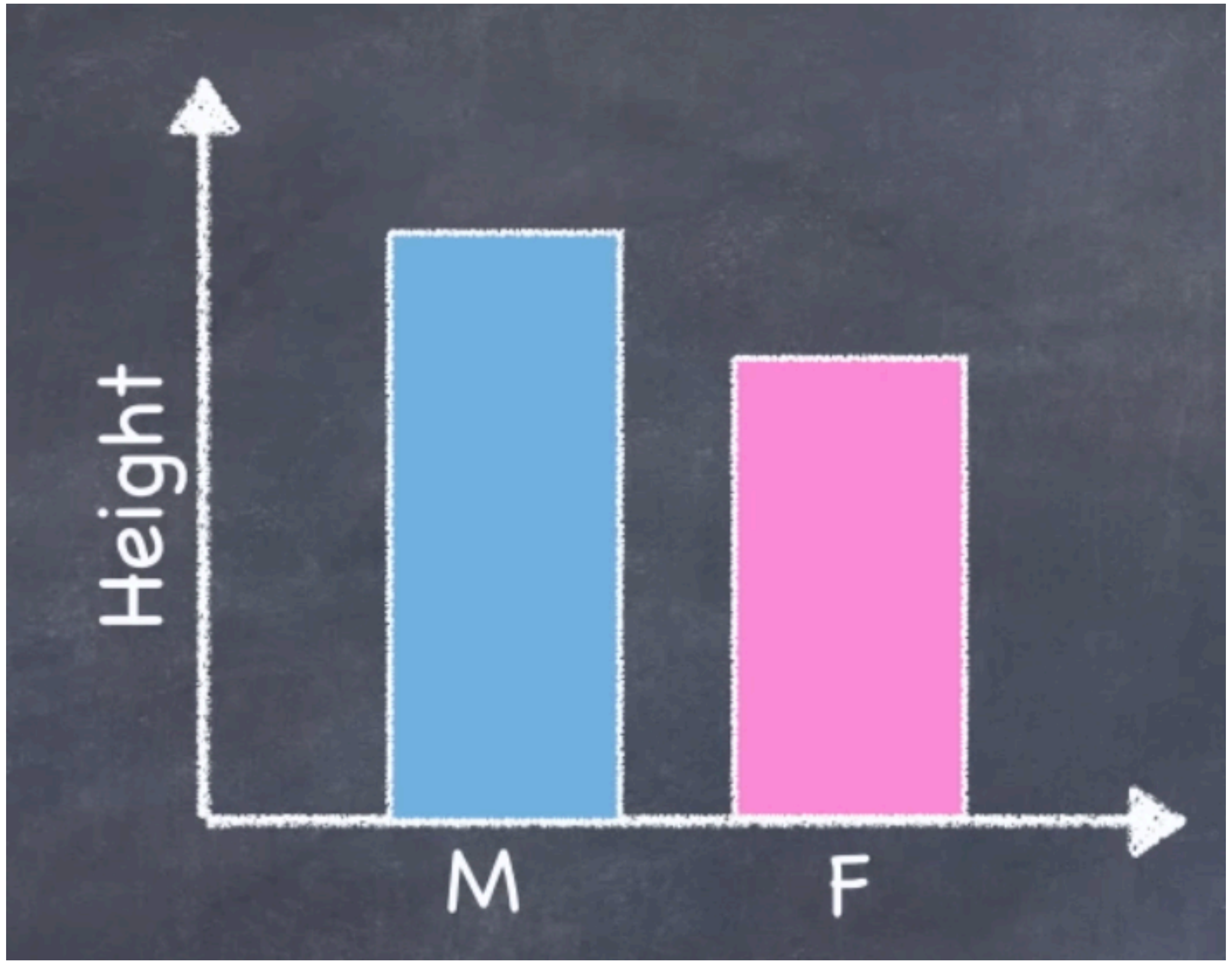
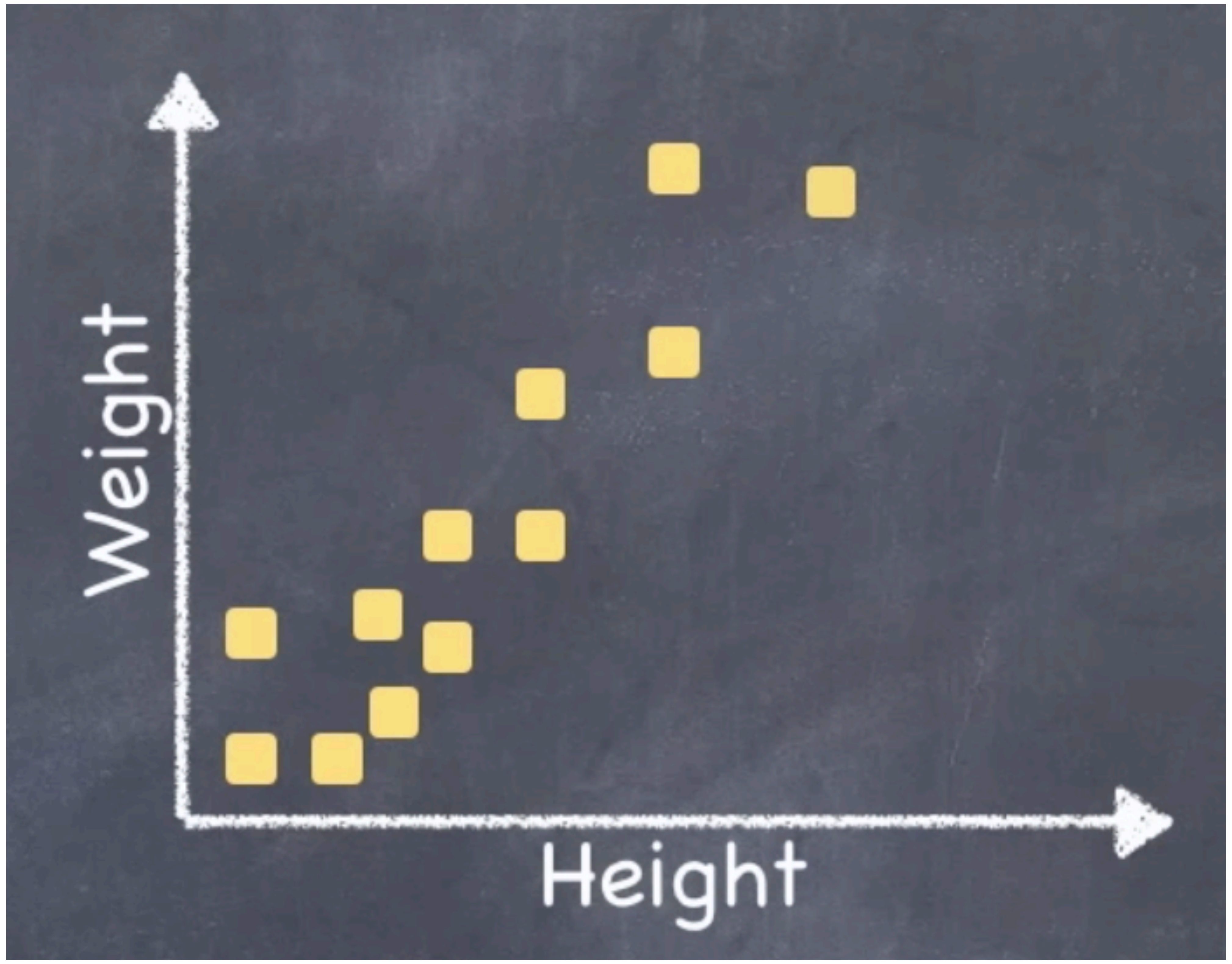
Statistical tests I

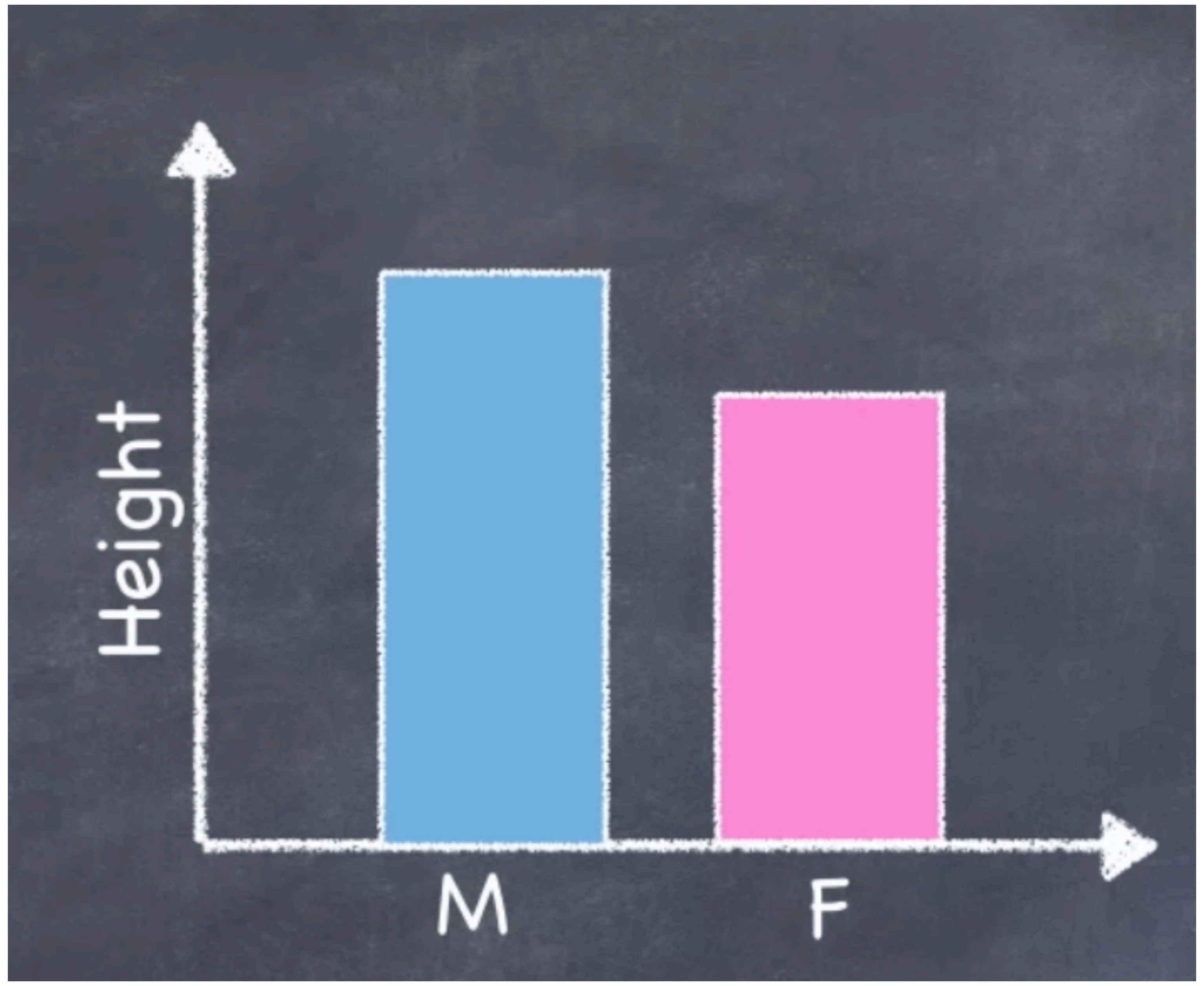
Rachel Warnock

10.04.2024


Height
M
F


Weight
Height

Today's adventures



compare the means of 2 groups → z-test

→ t-test

compare 2+ groups → ANOVA

You do not need to remember all the details of these statistical tests, you **only need to get the gist**.

The most important thing: what test is most appropriate for my data?

Steps for hypothesis testing

1. Define your research question / hypothesis

2. Define your statistical hypothesis (null & alternative)

3. *Find an appropriate test* & sampling distribution

4. Choose the type I error rate

Steps for hypothesis testing

5. Collect the data

6. Calculate test statistics

7. State the **statistical** conclusion

8. Interpret your results

Comparing 2 means — *z*-test and *t*-test

We often want to know if the average value of some variable is different (or higher or lower) than some out value or group.

The *z*-test — the most useless of all statistical tests

Almost never applied in real life (because it relies on a known standard deviation).

It's useful as a stepping stone to understanding the *t*-test.

A simple z-test

You might ask, are students from geology scoring higher than the average student in statistics?

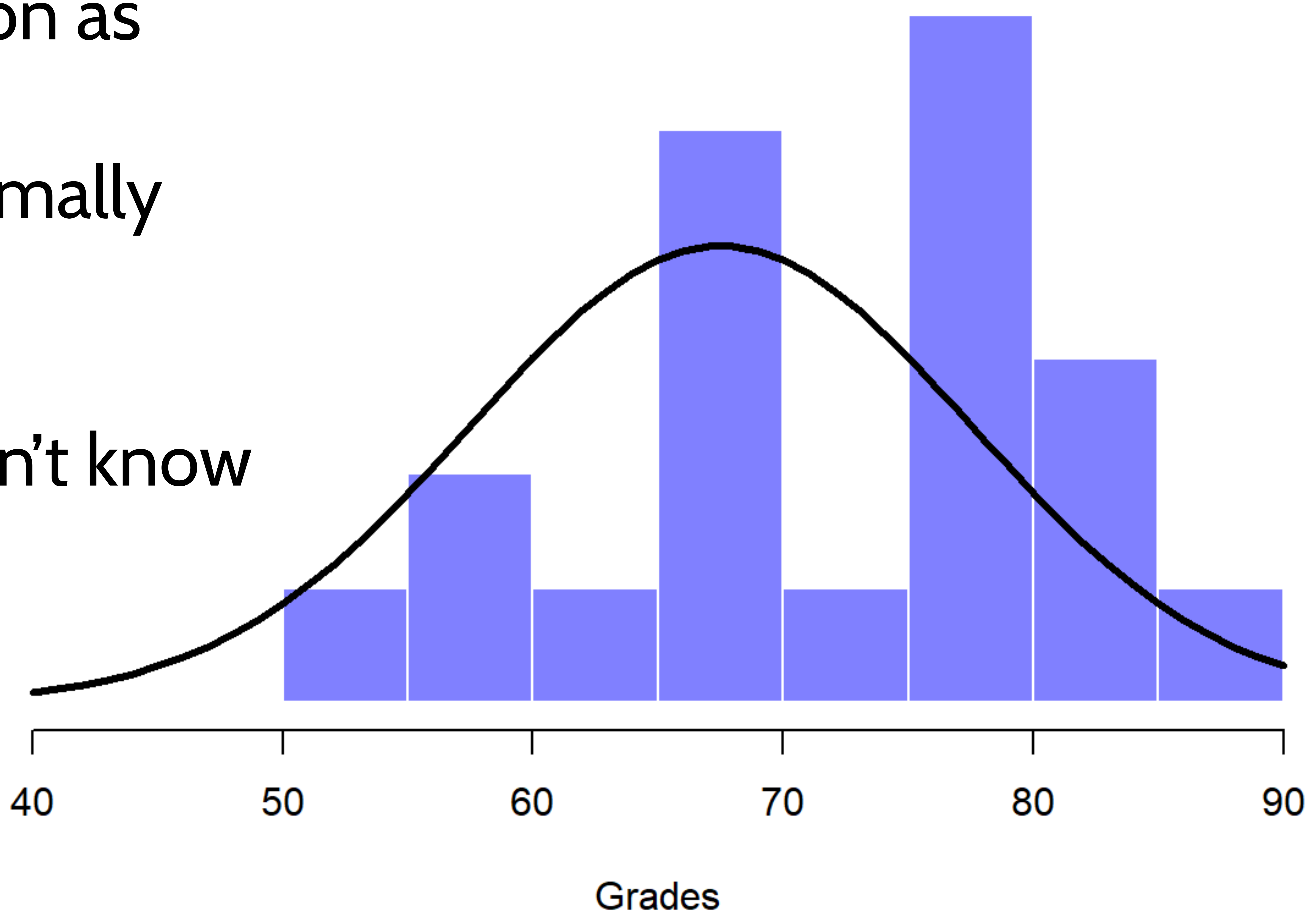
mean average grade = 67.5 and s.d. = 9.5

geology students mean = 73.2, N = 20

Assumptions

- the same standard deviation as the rest of the class
- the student grades are normally distributed

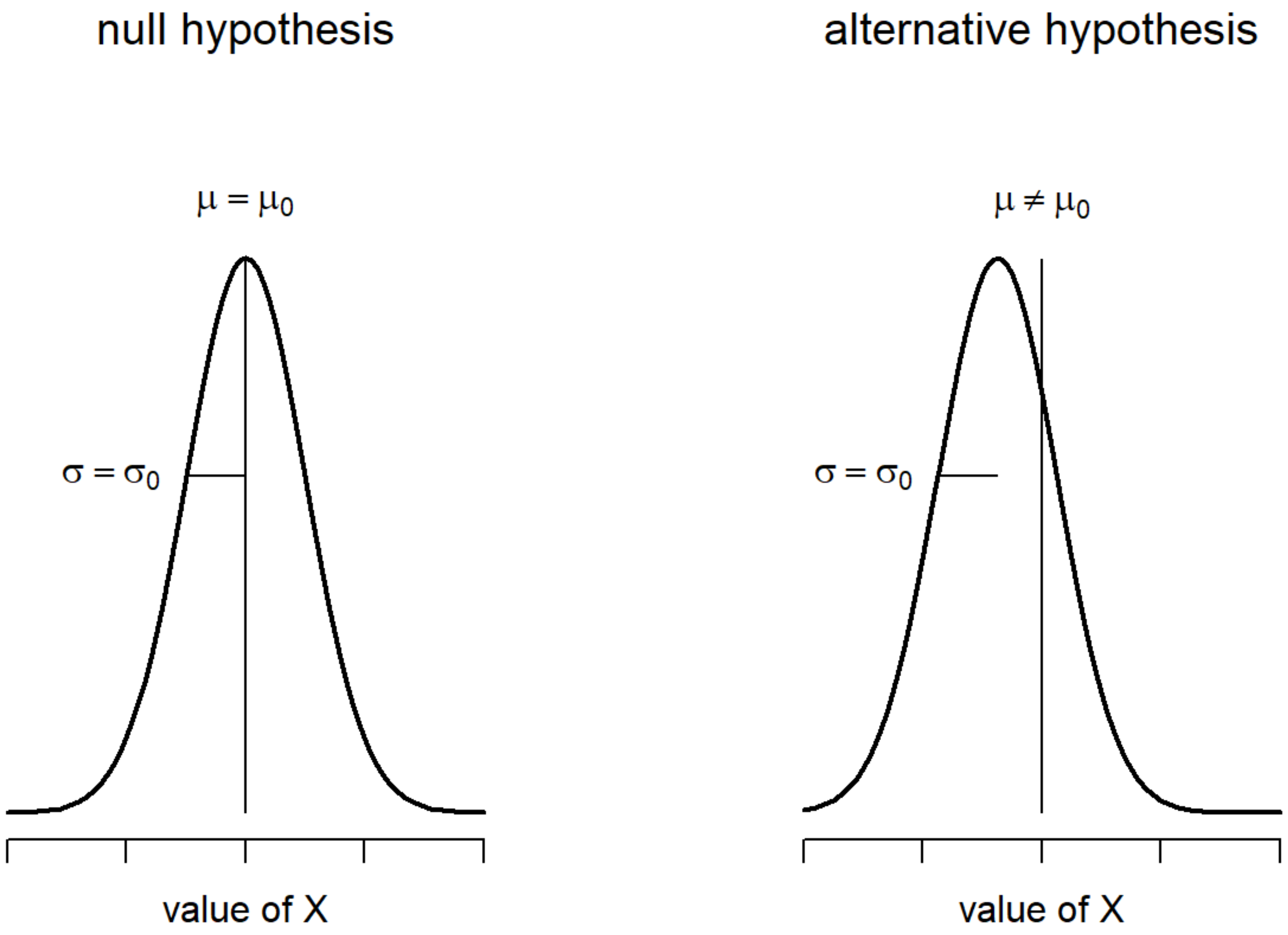
Note: in reality we usually don't know either of these things.



What are our hypotheses? 🤔

- research
- statistical (null and alternative)

The hypotheses for a one-sided or two-sided *z*-test



If the null hypothesis is true then the **sampling distribution** of the mean can be written as:

$$\bar{X} \sim \mathrm{Normal}(\mu_0, \mathrm{SE}(\bar{X}))$$

i.e., **comes from a distribution with the same mean & SD**

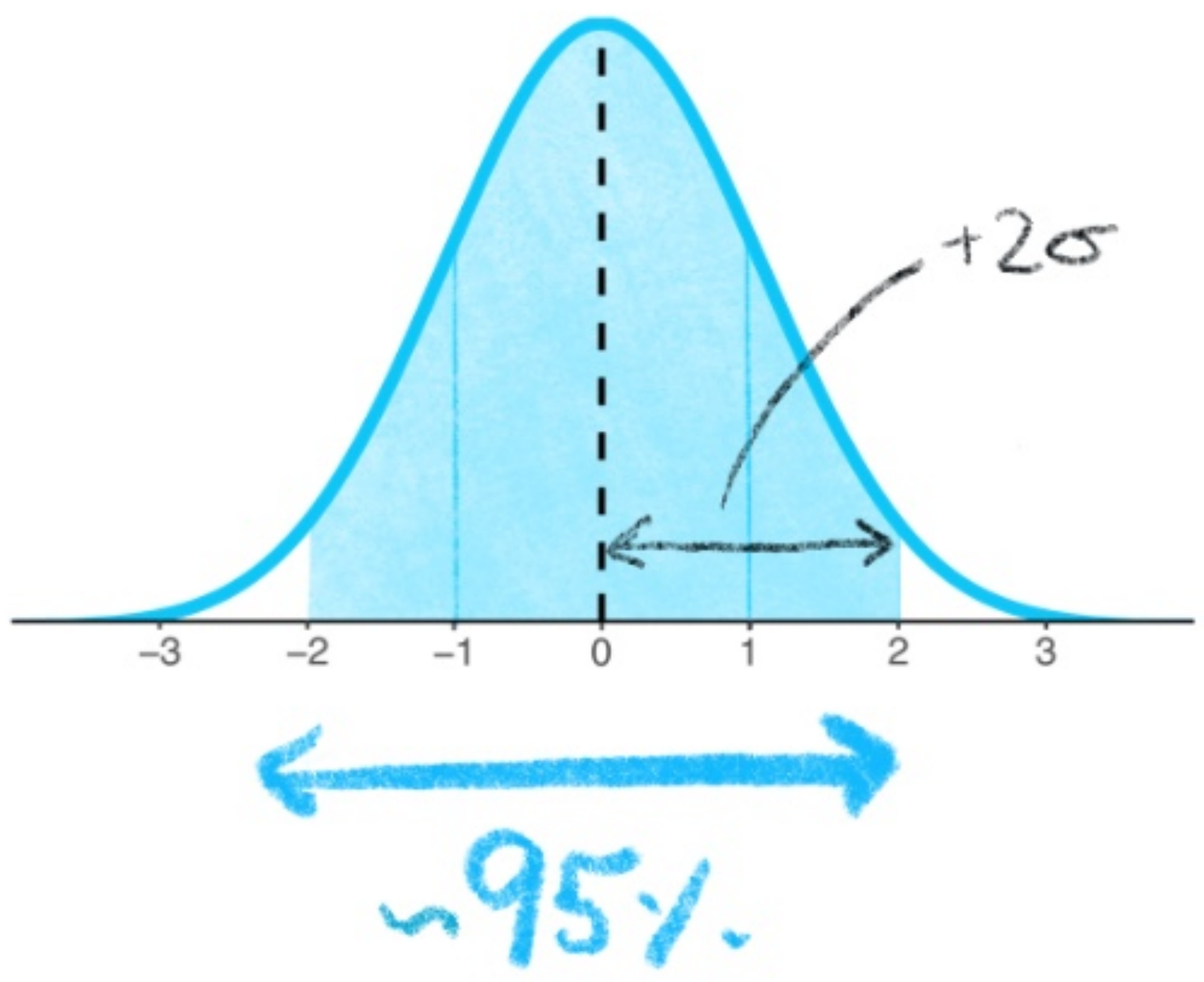
We focus on the difference between the means.

If $\mu - \mu_0$ equals or is very close to zero, this indicates support for the null hypothesis.

If $\mu - \mu_0$ is a long way away from zero, then it's looking less likely that we'd accept the null hypothesis.

How far from zero should it be for us to reject H_0?

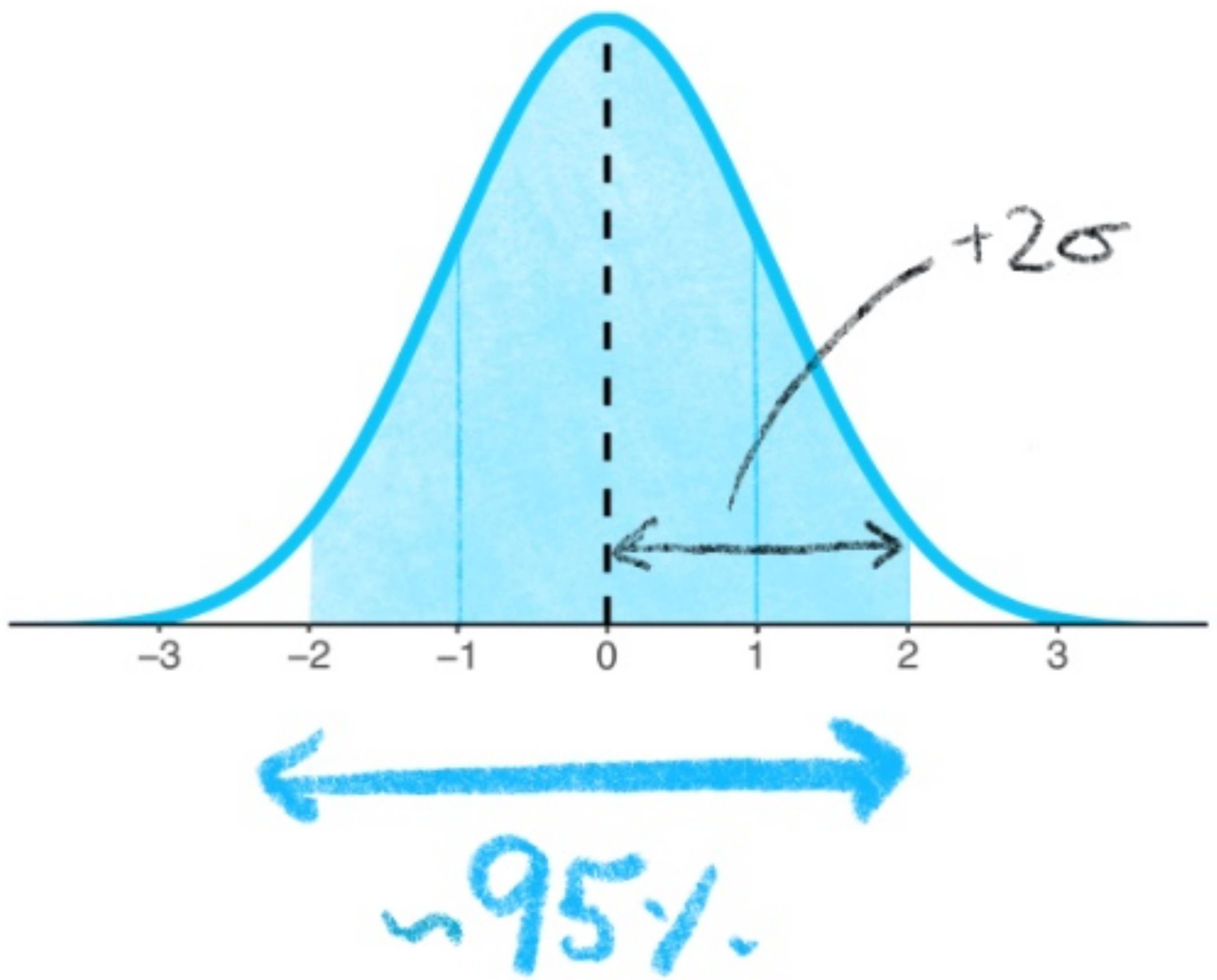
Standard scores, also know as: z-scores



= number of standard deviations the sample mean is away in the test distribution

$$\text{standard score} = \frac{\text{raw score} - \text{mean}}{\text{standard deviation}}$$

Calculating the z-score



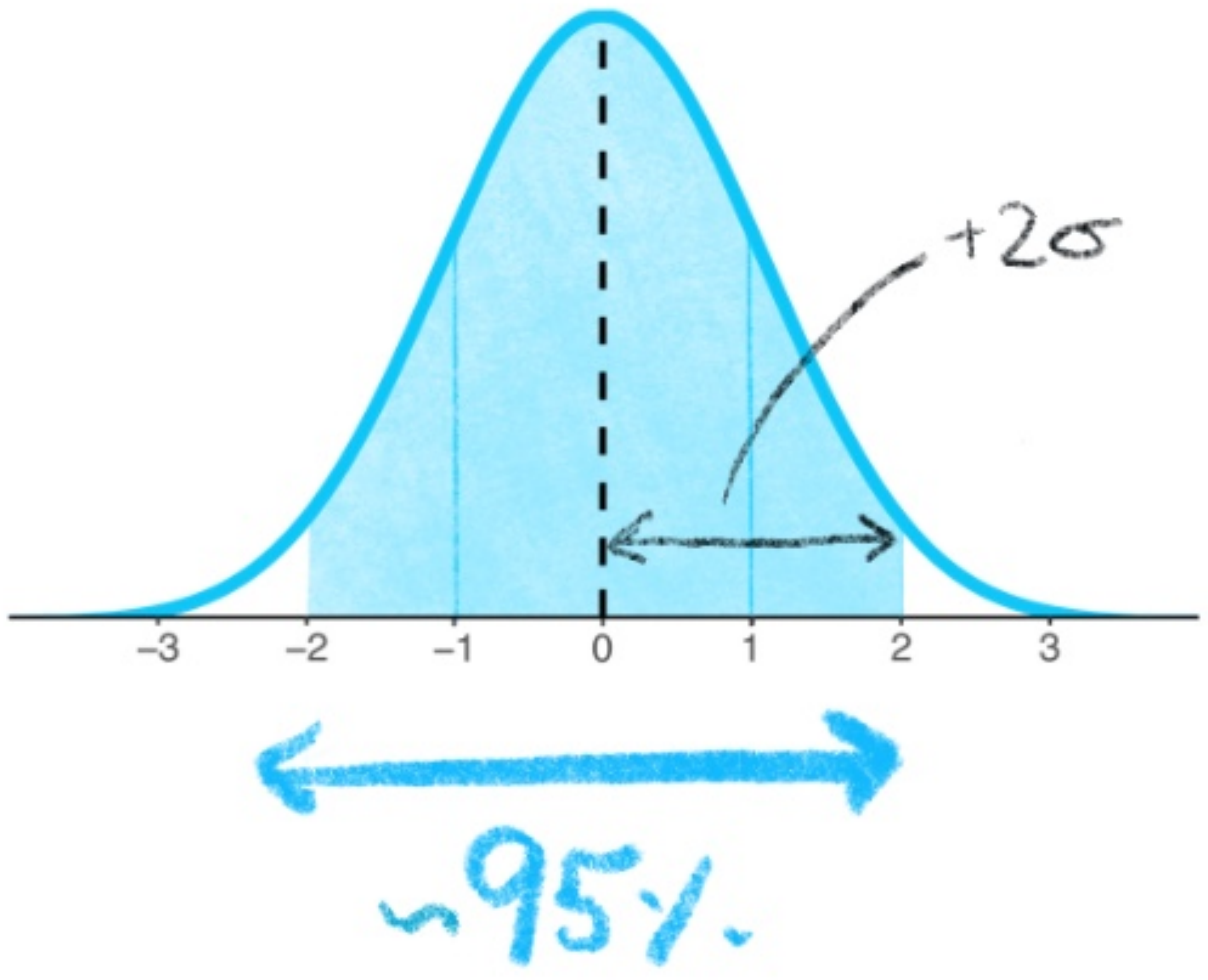
$$z_{\bar{X}} = \frac{\bar{X} - \mu_0}{\mathrm{SE}(\bar{X})}$$

OR

$$z_{\bar{X}} = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{N}}$$

You don't need to know these formulas!
Only here for reference

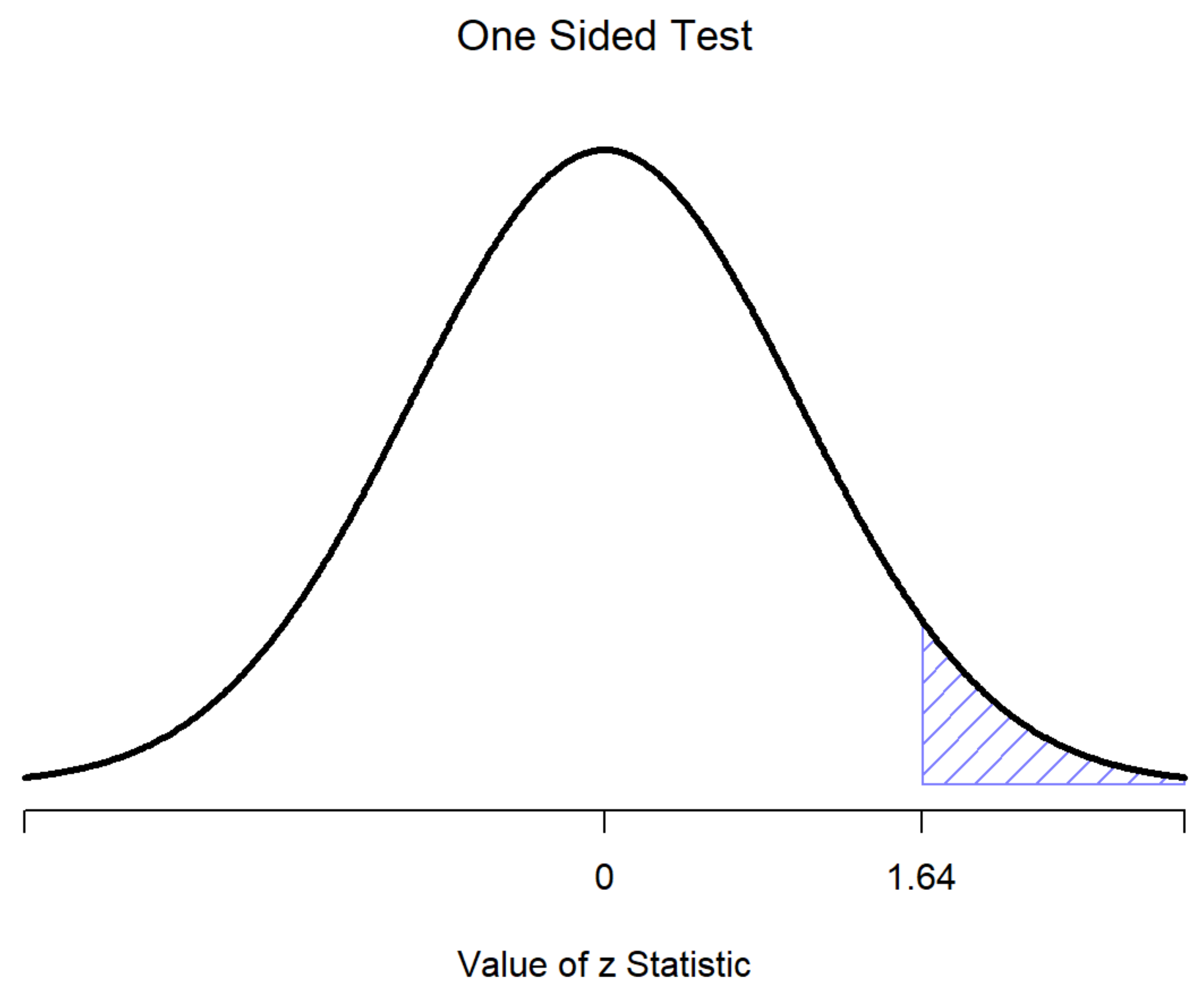
Calculating the z-score

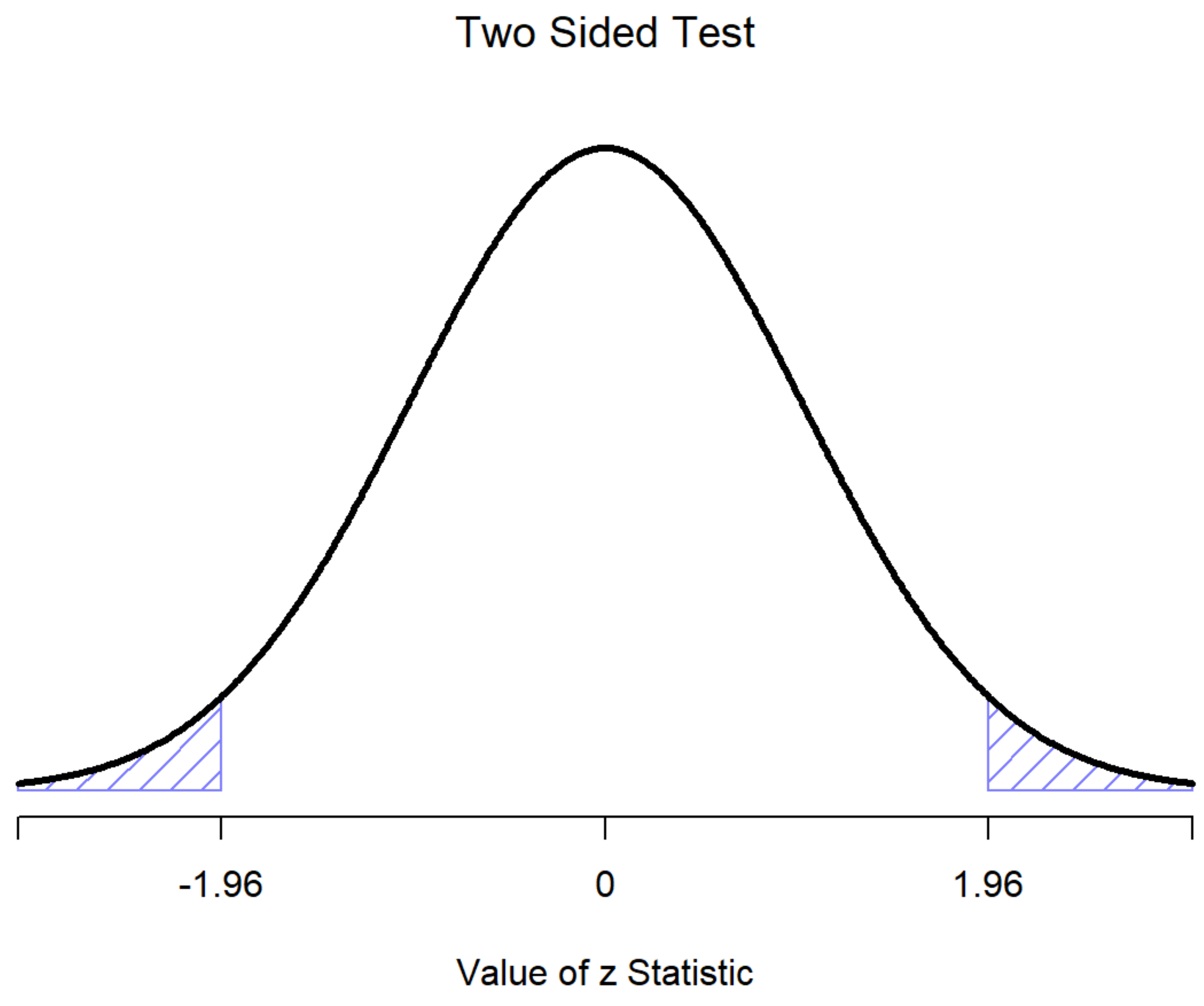


The *z*-statistic is equal to the number of standard deviation that separates the observed sample mean μ from the population mean μ_0 predicted by the null hypothesis.

The 5% critical regions for the *z*-test are always the same.

desired α level	two-sided test	one-sided test
.1	1.644854	1.281552
.05	1.959964	1.644854
.01	2.575829	2.326348
.001	3.290527	3.090232





A simple z-test

You might ask, are students from geology scoring higher than the average student in statistics?

mean average grade = 67.5 and s.d. = 9.5

geology students mean = 73.2, N = 20

desired α level	two-sided test	one-sided test
.1	1.644854	1.281552
.05	1.959964	1.644854
.01	2.575829	2.326348
.001	3.290527	3.090232

With a mean grade of 73.2 in the sample of geology students, and assuming a true population standard deviation of 9.5, we get $z = 2.26$.

What is p?

Assumptions of the *z*-test

Randomness

Normality

Independence

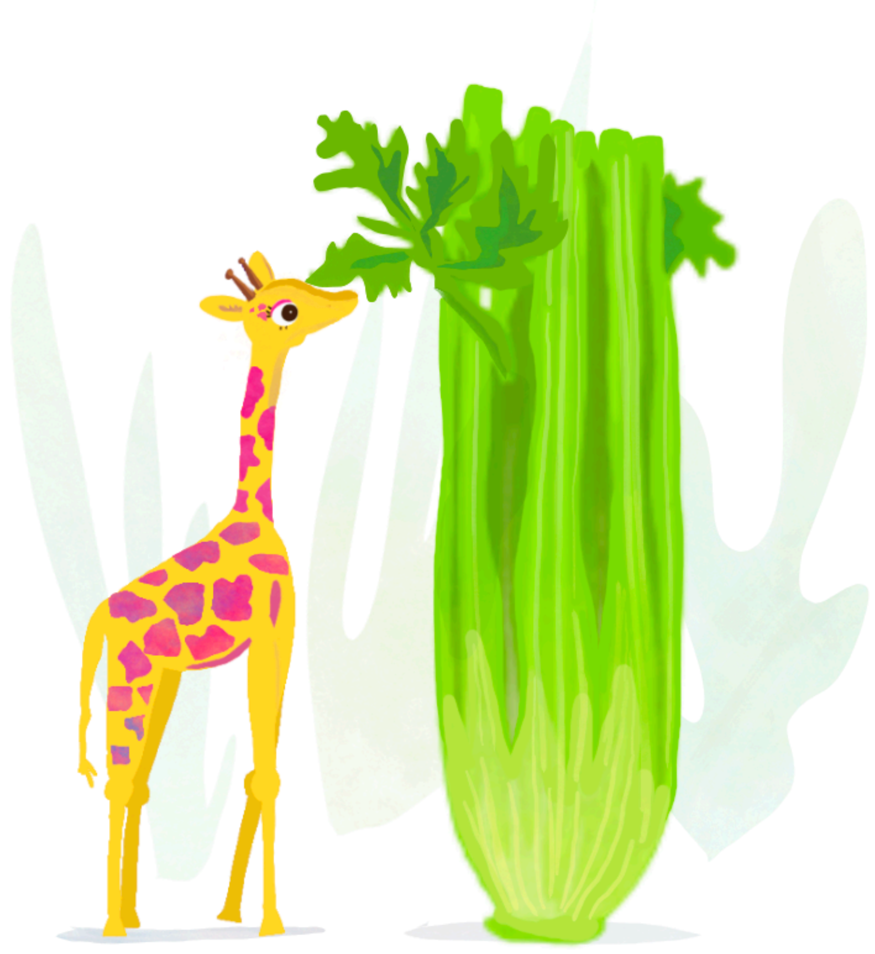
Known standard deviation

From *z*-test to *t*-test

Suppose we don't know the true standard deviation?

We only know the **estimated** standard deviation.

See Teacup Giraffes for more on the normal distribution

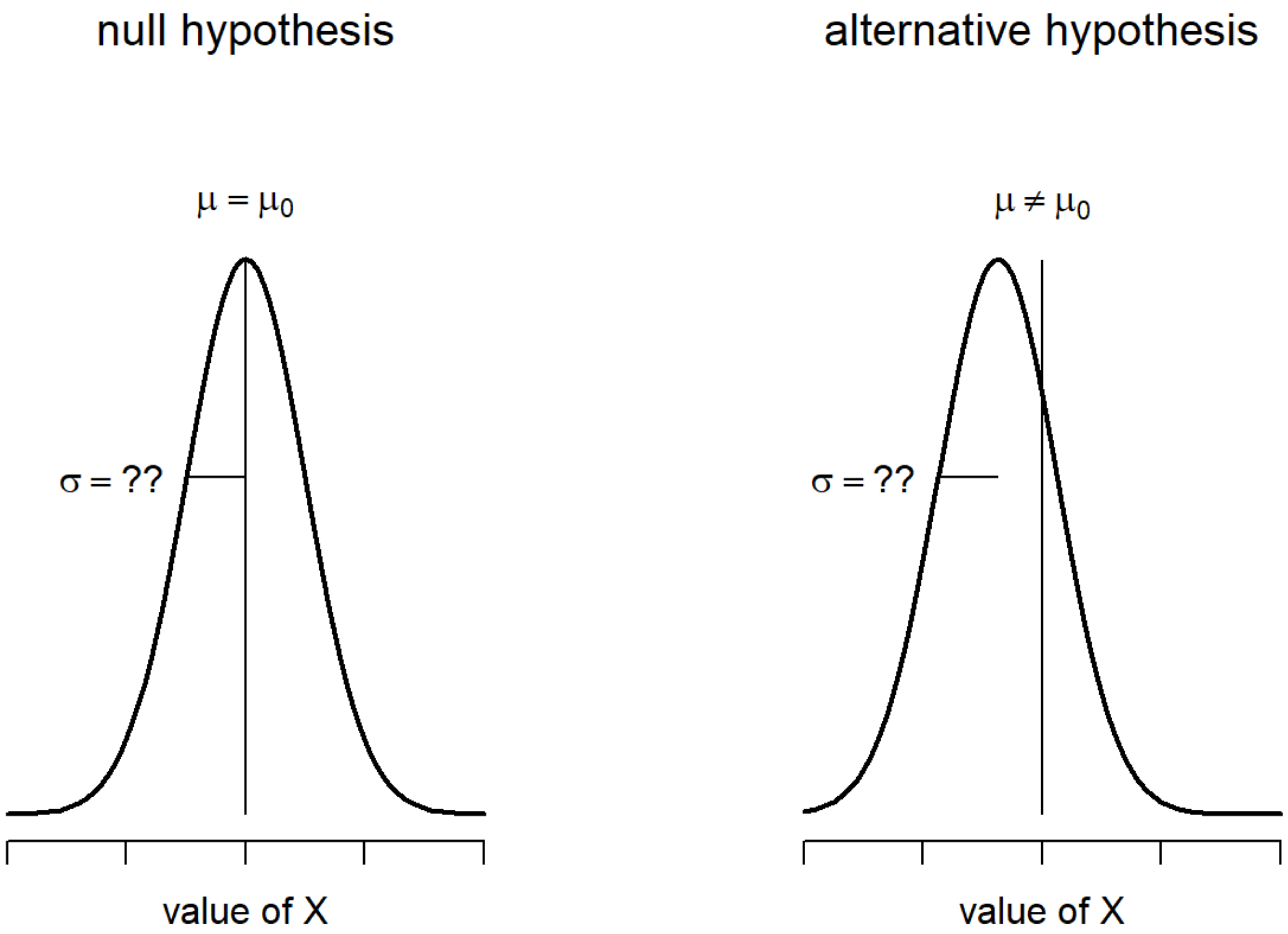


From *z*-test to *t*-test

We *could* use the estimated s.d. but this wouldn't be strictly appropriate.

What if the estimate is incorrect?

In the previous example, if the s.d. changes from 9 to 11, the results become non-significant.



null hypothesis
$\mu = \mu_0$
$\sigma = ??$
value of X
alternative hypothesis
$\mu \neq \mu_0$
$\sigma = ??$
value of X

The one-sample *t*-test

To accommodate the fact we don't know the s.d, we need to subtly change the sampling distribution.

The t-test statistic is very similar:

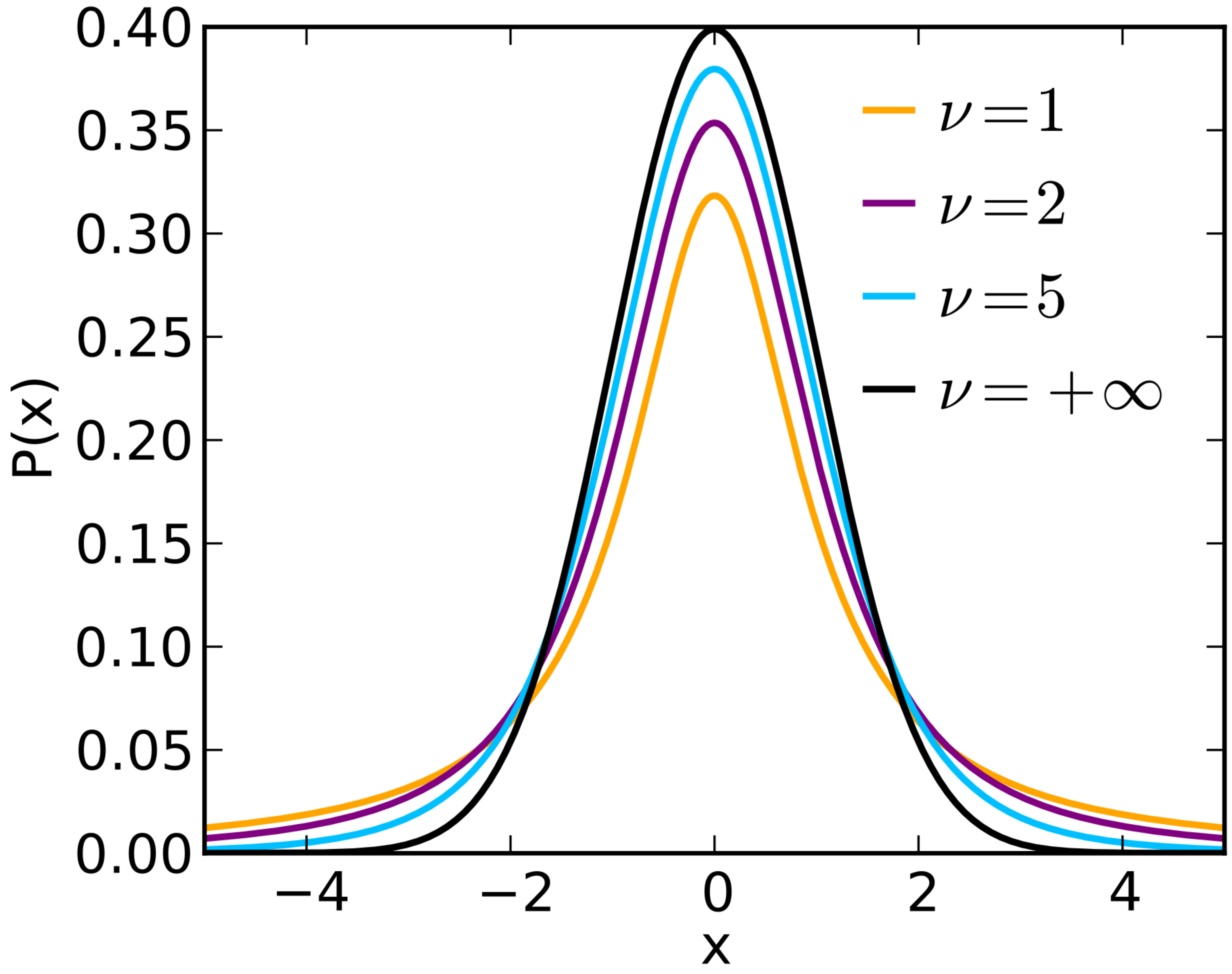
The first diff. is that we use the est. s.d.

$$t = \frac{\bar{X} - \mu}{\hat{\sigma}/\sqrt{N}}$$

If the estimate has been constructed from *N* observations, then the sampling distribution becomes a ***t*-distribution** with *N* - 1 degrees of freedom.

From here we can obtain a *p*-value for the *t*-distribution.

The one-sample *t*-test



Assumptions of the one sample *t*-test

Randomness

Normality

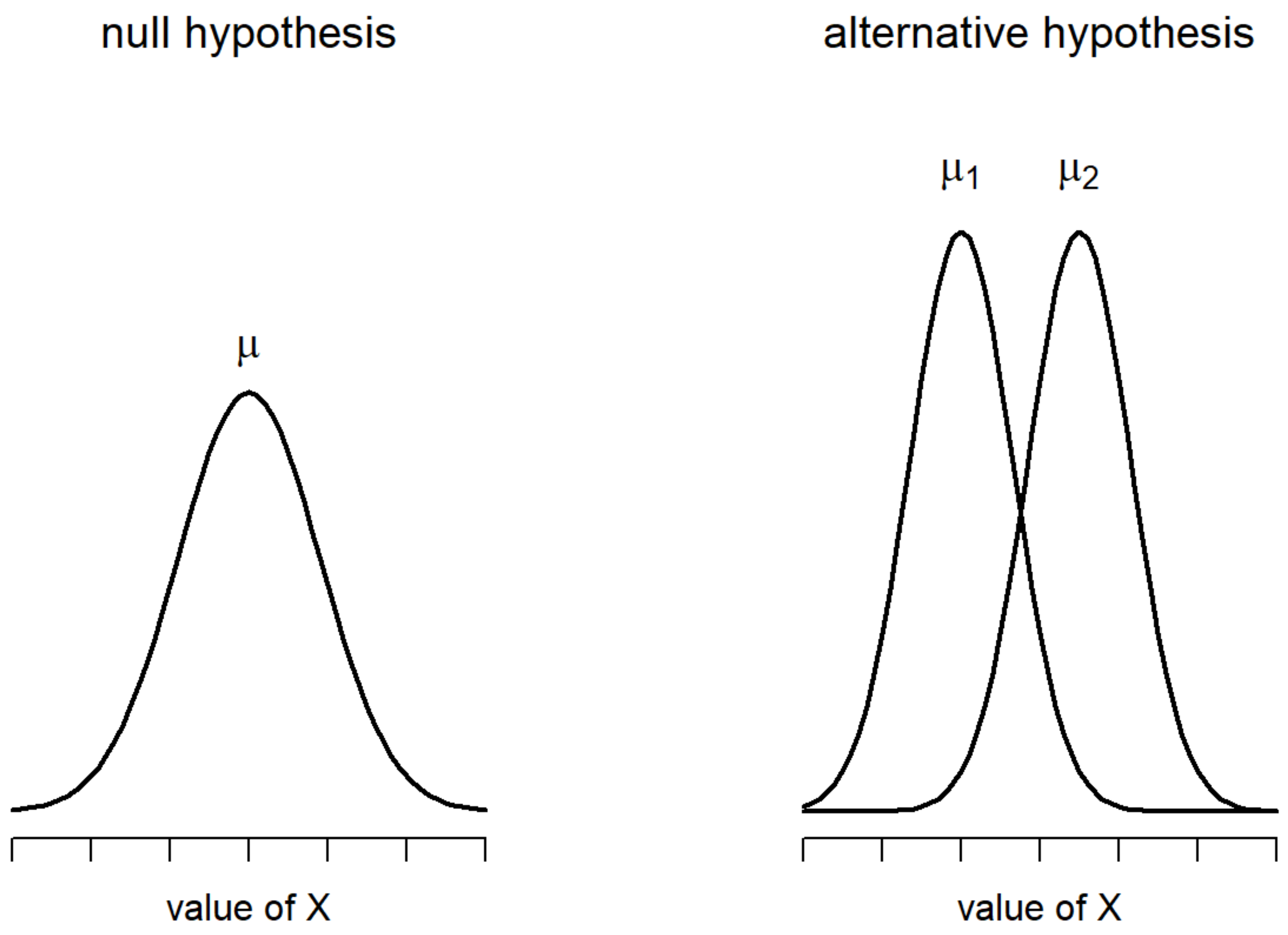
Independence

There are many varieties of *t*-test that get round some of these assumptions.

The one-sample *t*-test compares the mean of a single sample to some pre-existing value to test whether the sample mean is significantly greater or less than that value.

The independent samples *t*-test compares the mean of one distinct group to the mean of another group.

The independent-samples *t*-test



If the null is true we expect the difference between the means = 0.

But how close?

Test statistic is:

$t = (\mu - \mu_0) / SD$

The independent-samples *t*-test

For the variance, we calculate a "pooled" estimate by taking a weighted average of the variance estimates for each sample.

The weight assigned to each sample is equal to the number of observations in that sample, minus 1.

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} \longrightarrow$$

$$\begin{aligned} w_1 &= N_1 - 1 \\ w_2 &= N_2 - 1 \end{aligned}$$

$$\hat{\sigma}_p = \sqrt{\frac{w_1 \hat{\sigma}_1{}^2 + w_2 \hat{\sigma}_2{}^2}{w_1 + w_2}}$$

The independent-samples *t*-test

Compared to the one sample *t*-test:

$$t = \frac{\bar{X} - \mu}{\hat{\sigma}/\sqrt{N}} \longrightarrow t = \frac{\bar{X}_1 - \bar{X}_2}{\mathrm{SE}(\bar{X}_1 - \bar{X}_2)}$$

Again, from here we can obtain a *p*-value for the *t*-distribution.

Assumptions of the independent samples *t*-test

Randomness

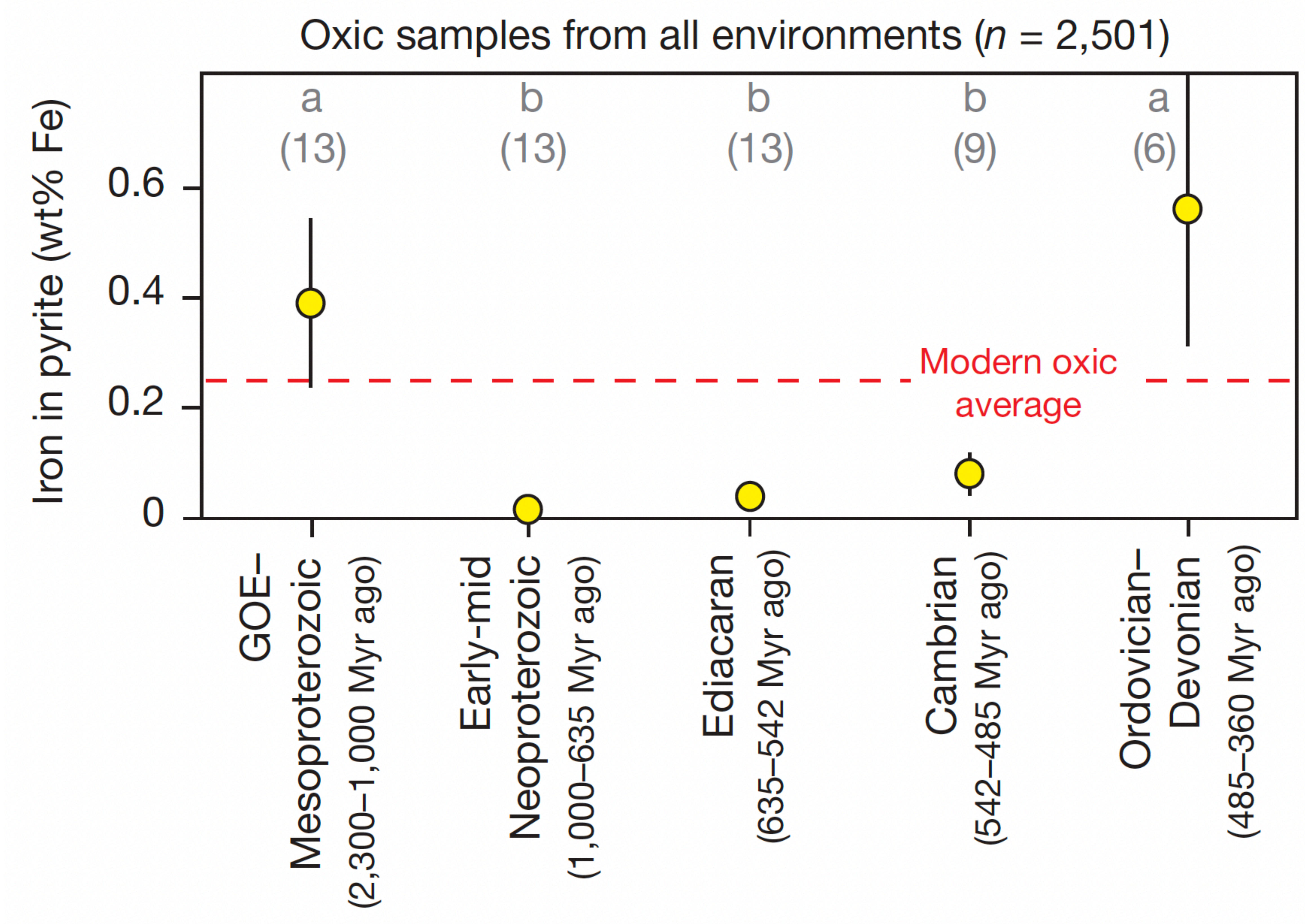
Normality

Independence (between groups)

Homogeneity of variance (the s. d. is the same between both groups)

One-way ANOVA (Analysis of variances)

What if you want to compare more than 2 groups?

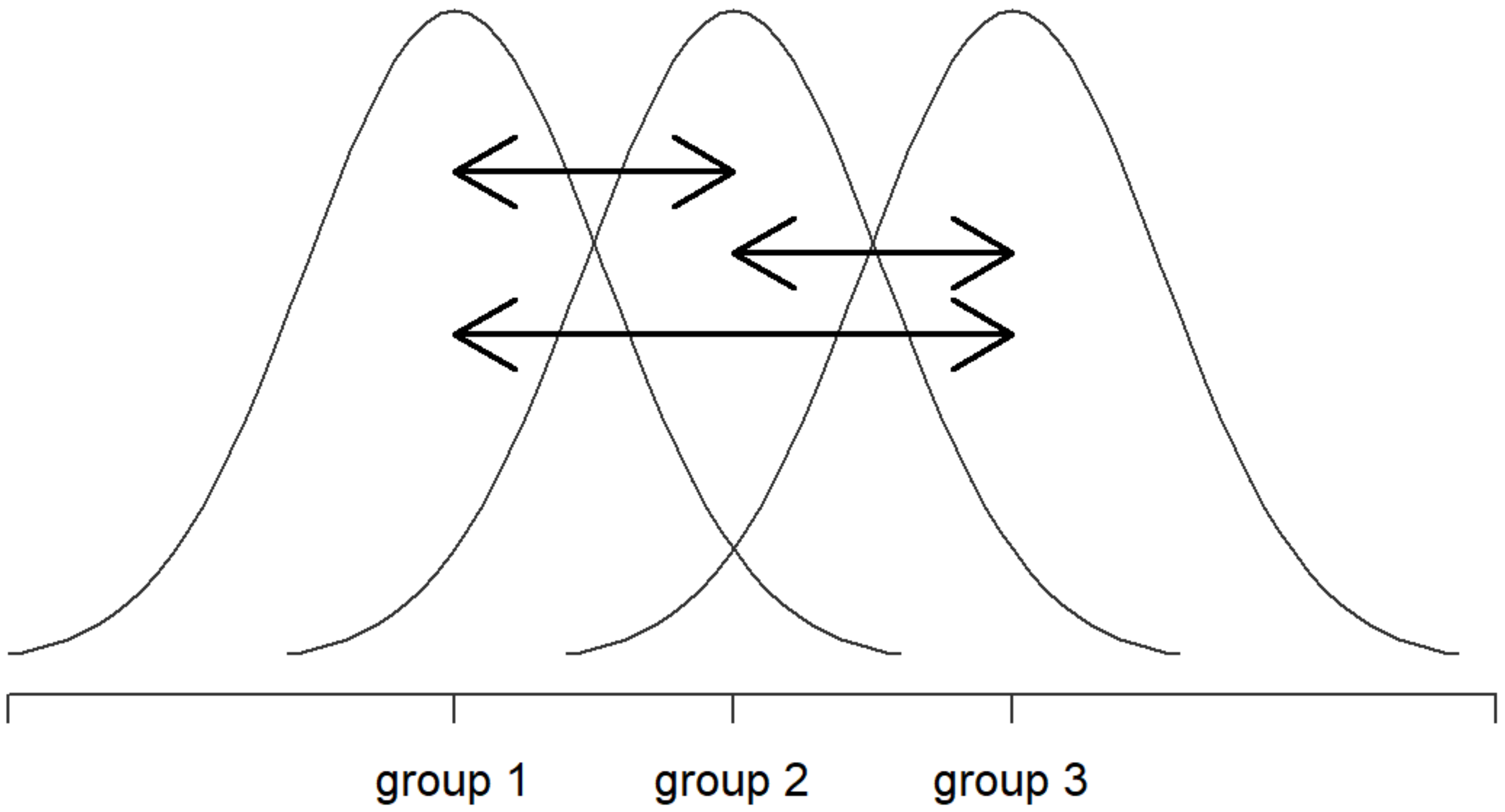


$$H_0 \quad : \quad \text{it is true that } \mu_P = \mu_A = \mu_J$$

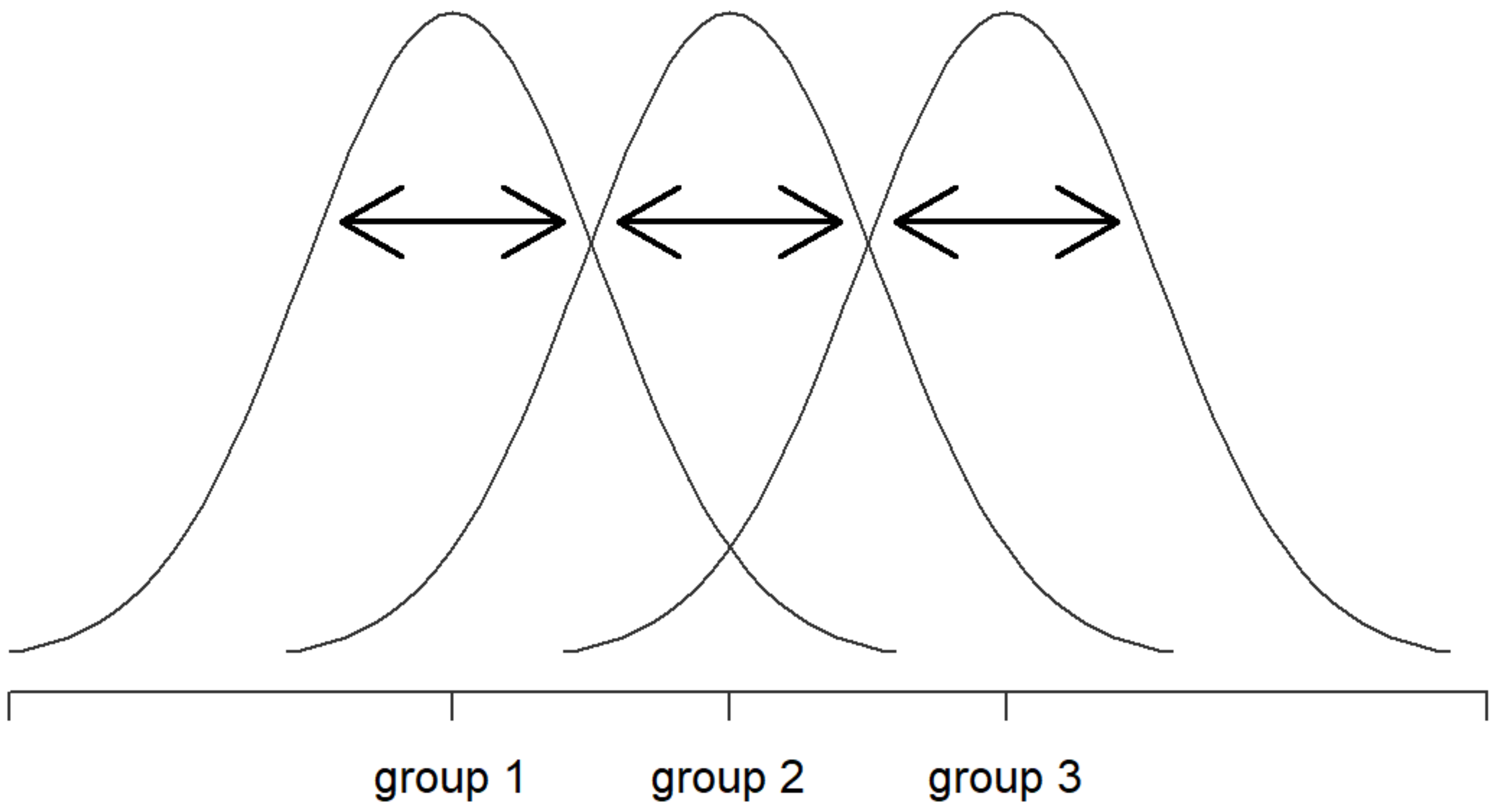
$$H_1 \quad : \quad \text{it is *not* true that } \mu_P = \mu_A = \mu_J$$

Between group variation (SS_b)

i.e., the differences between group means



Within group variation (SS_w)



Total group variation (SS_{tot})

SS_{tot} is the sum of "the variation due to the differences in the sample means for the different groups" (SS_b) plus "all the rest of the variation" (SS_w).

$SS_{tot} = SS_b + SS_w$

If the null hypothesis is true, **we expect all the sample means to be similar** and SS_b to be small, with $SS_{tot} = SS_w$.

First we calculate the variance:

$$\mathrm{Var}(Y) = \frac{1}{N}\sum_{k=1}^{G}\sum_{i=1}^{N_k}\left(Y_{ik} - \bar{Y}\right)^2 \longrightarrow \mathrm{Var}(Y) = \frac{1}{N}\sum_{p=1}^{N}\left(Y_p - \bar{Y}\right)^2$$

Then we calculate the total group (SS_{tot}), within group (SS_w) and between group (SS_b) **sum of squares**:

$$\begin{aligned} SS_b &= \textstyle\sum_{k=1}^{G}\sum_{i=1}^{N_k}\left(\bar{Y}_k - \bar{Y}\right)^2 \\ &= \textstyle\sum_{k=1}^{G} N_k\left(\bar{Y}_k - \bar{Y}\right)^2 \end{aligned}$$

From sums of squares to the *F*-test via the *F* ratio

df1 = no. of groups - 1
df2 = no. of obs. - groups

We calculate is the degrees of freedom associated with the SS_b and SS_w values.

$$\begin{aligned} \mathrm{df}_b &= G - 1 \\ \mathrm{df}_w &= N - G \end{aligned}$$

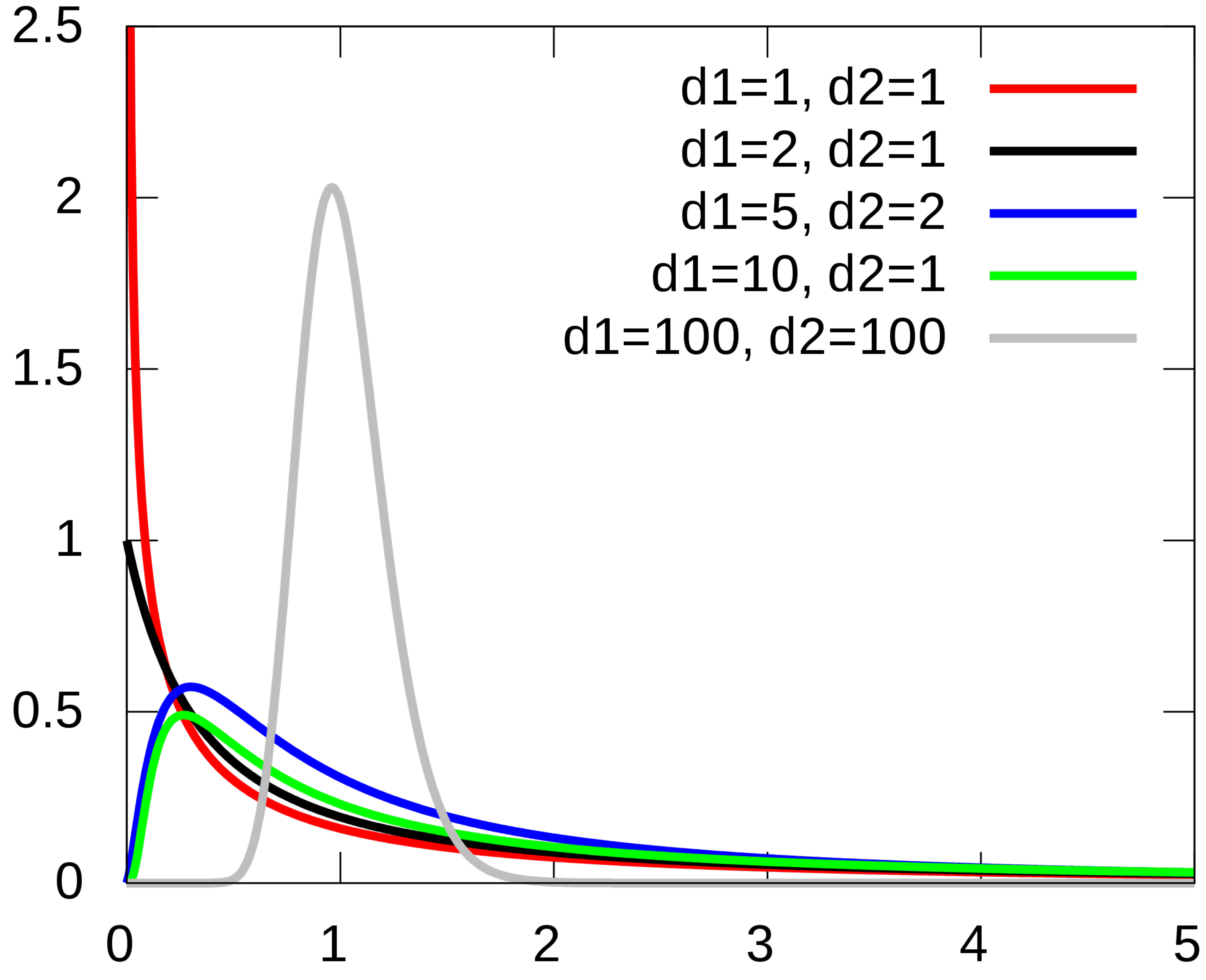
Next is convert our summed squares value into a "mean squares" value, by dividing by the degrees of freedom.

$$\begin{aligned} \mathrm{MS}_b &= \frac{\mathrm{SS}_b}{\mathrm{df}_b} \\ \mathrm{MS}_w &= \frac{\mathrm{SS}_w}{\mathrm{df}_w} \end{aligned} \longrightarrow F = \frac{\mathrm{MS}_b}{\mathrm{MS}_w}$$

The gist behind the *F* statistic

The intuition behind the *F* statistic is straightforward: bigger values of *F* means that the between-groups variation is large, relative to the within-groups variation.

The larger the value of *F*, the more evidence we have against the null hypothesis.



Tomorrow's adventures

post hoc ANOVA test → Tukey-Kramer

Steel Dwass

equivalent to a one-sample t-test, non-parametric → Wilcoxon test

equivalent to ANOVA, non-parametric → Kruskal-Wallis

Questions?

teacups-giraffes-and-statistics

Further info

Video: [ANOVA (Analysis of variance) simply explained]
Video: [ANOVA: One-way analysis of variance]

ANOVA in R - A Complete Step-by-Step Guide

Two-way ANOVA - Examples and when to use it