Statistical tests

Rachel Warnock

23.02.2022

FAU

Today's adventures

compare the means of 2 groups → z-test

→ t-test

compare 2+ groups → ANOVA

Steps for hypothesis testing (recap)

1. define your research question / hypothesis

2. define your statistical hypothesis, null & alternative distributions

3. find an appropriate test & sampling distribution

4. choose the type I error rate

Steps for hypothesis testing (recap)

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 a useful 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?

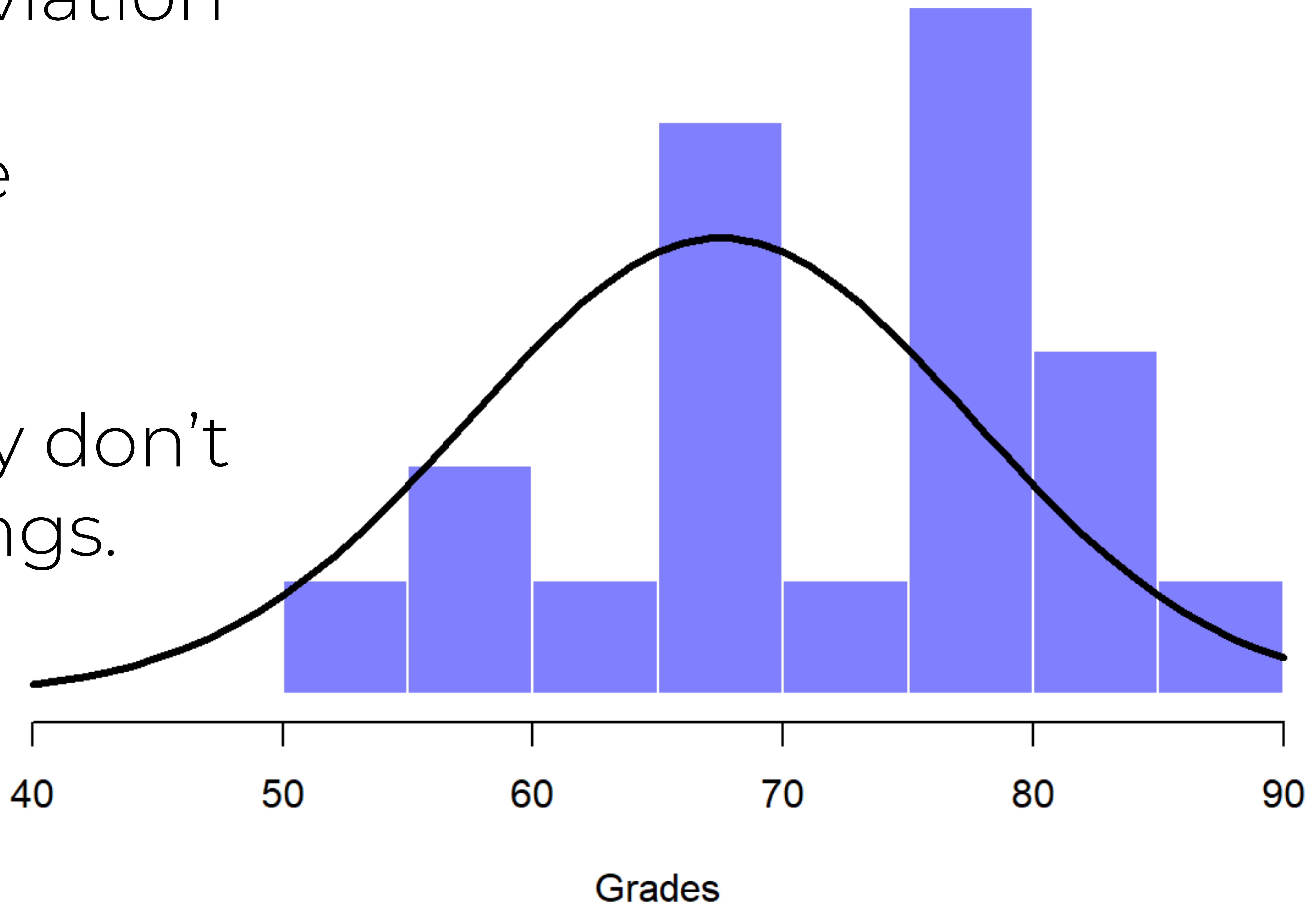
mean average grade = 67.5 and sd = 9.5

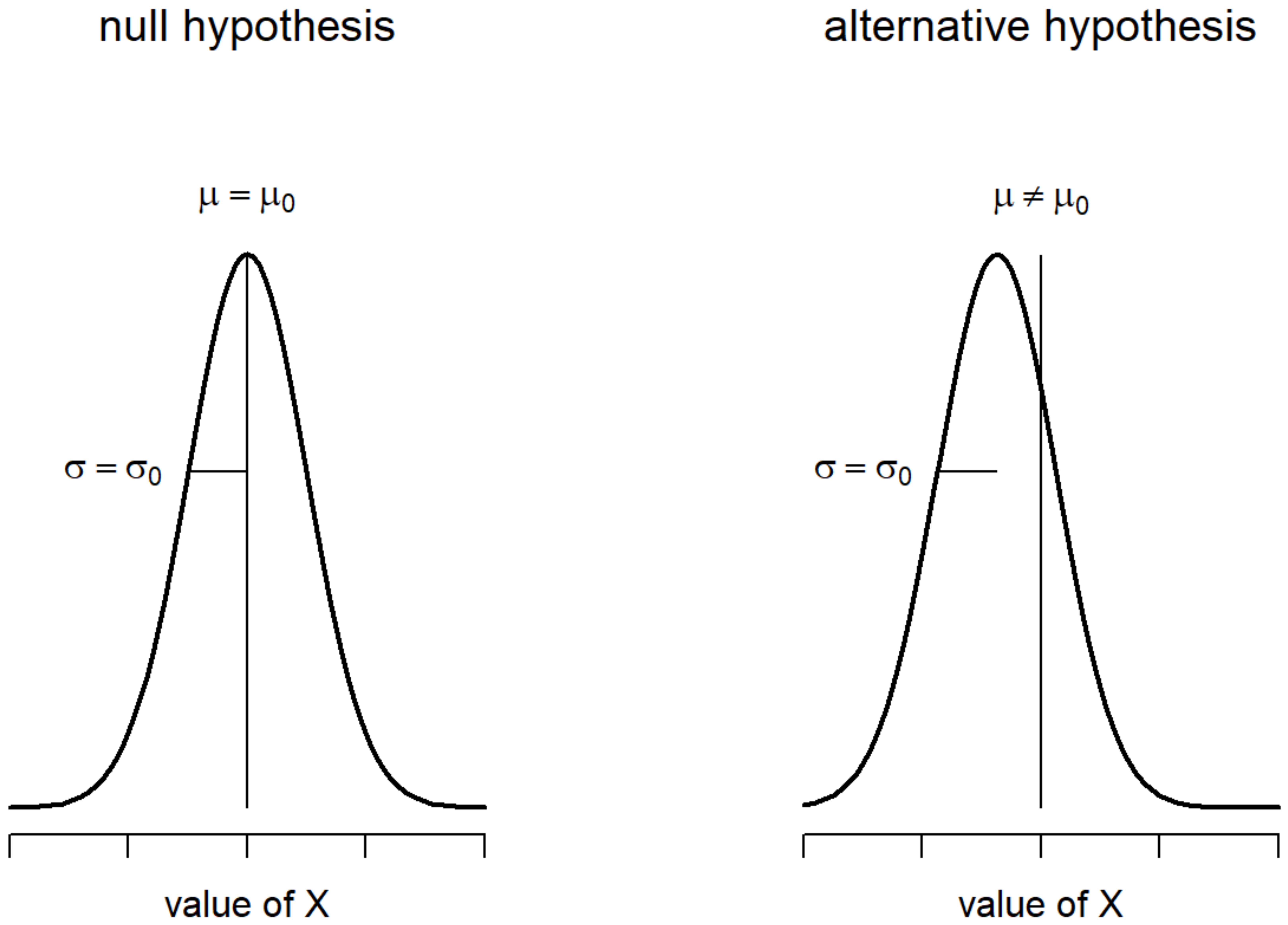
geology students mean = 73.2, N = 20

Assumptions

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

<u>Note:</u> in reality we usually don't know either of these things.



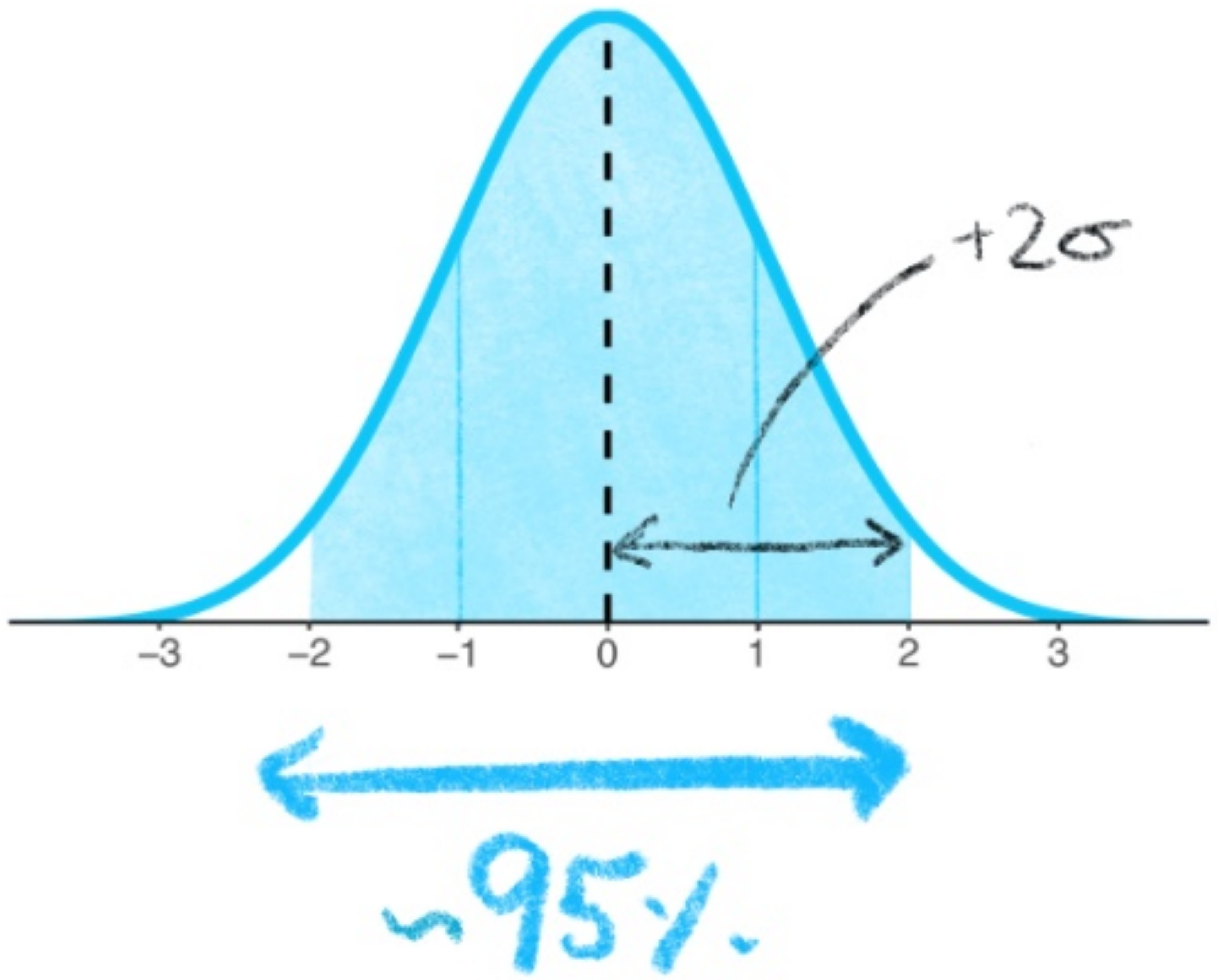

null hypothesis
alternative hypothesis
μ = μ0
μ ≠ μ0
σ = σ0
σ = σ0
value of X
value of X

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

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

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

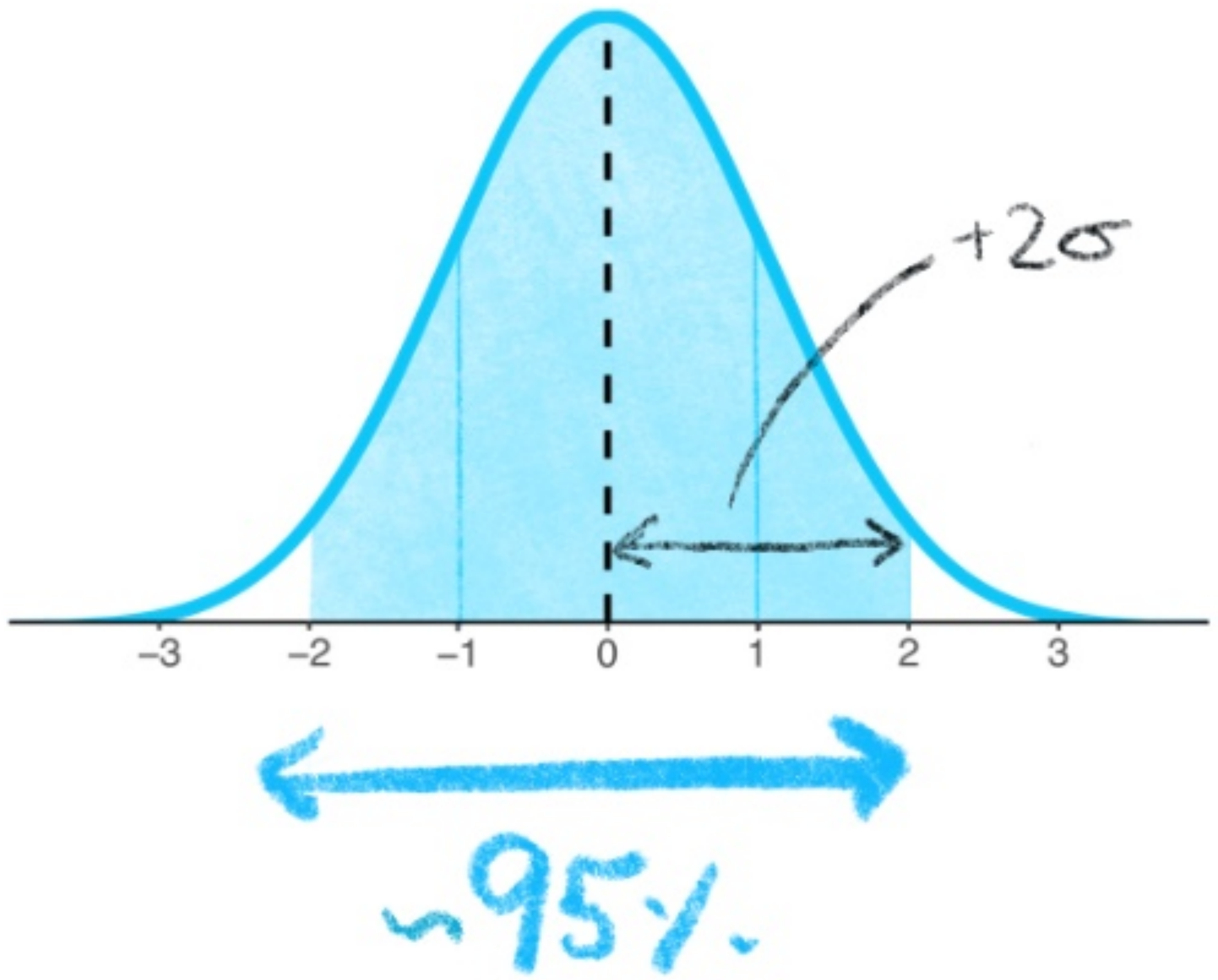
Standard scores, also know as: z-score



Number of standard deviations sample mean is away from the test distribution

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

Calculating the z-score

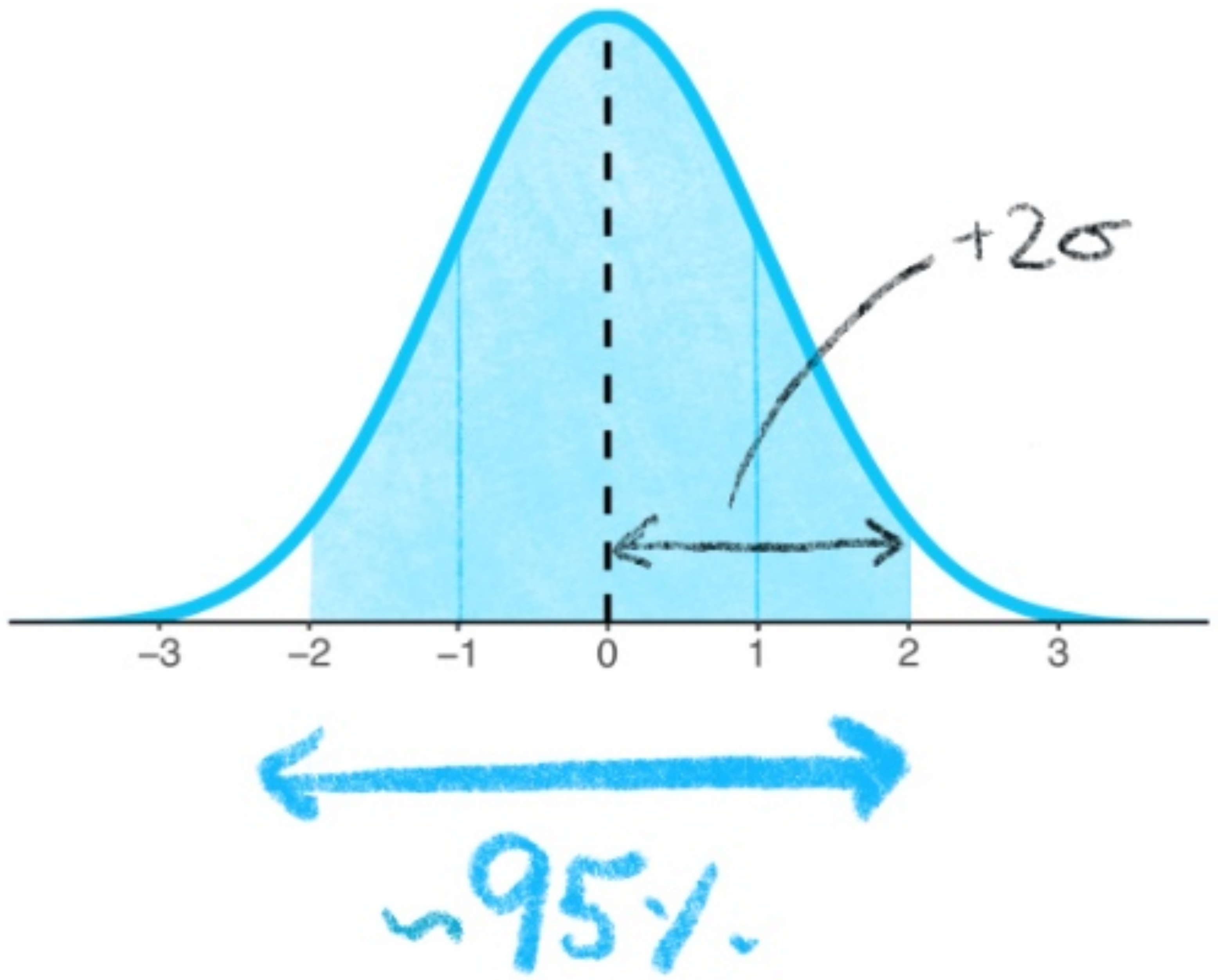


$$z_{\bar{X}} = \frac{\bar{X} - \mu_0}{\mathrm{SE}(\bar{X})}$$

OR

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

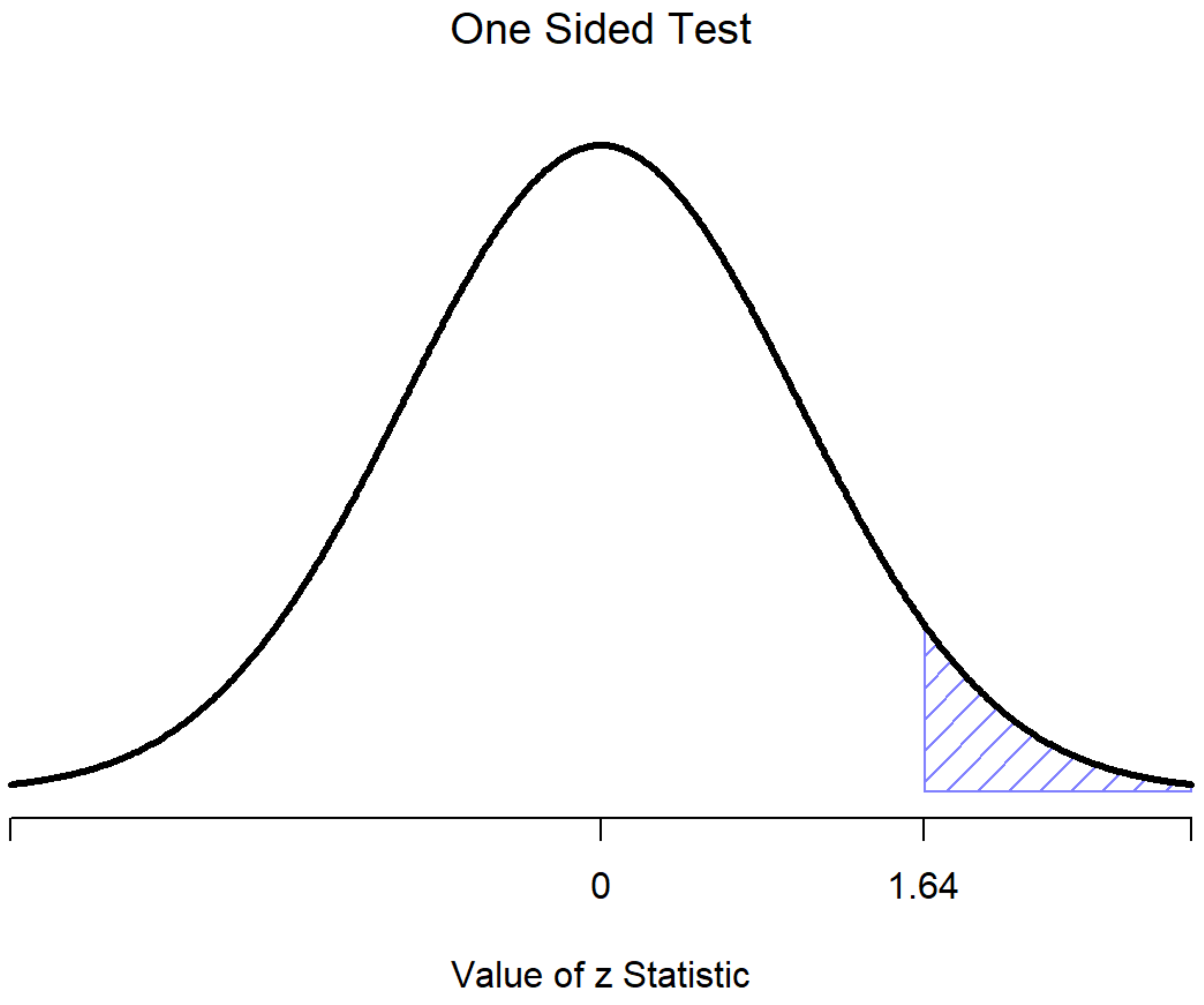
Calculating the z-score

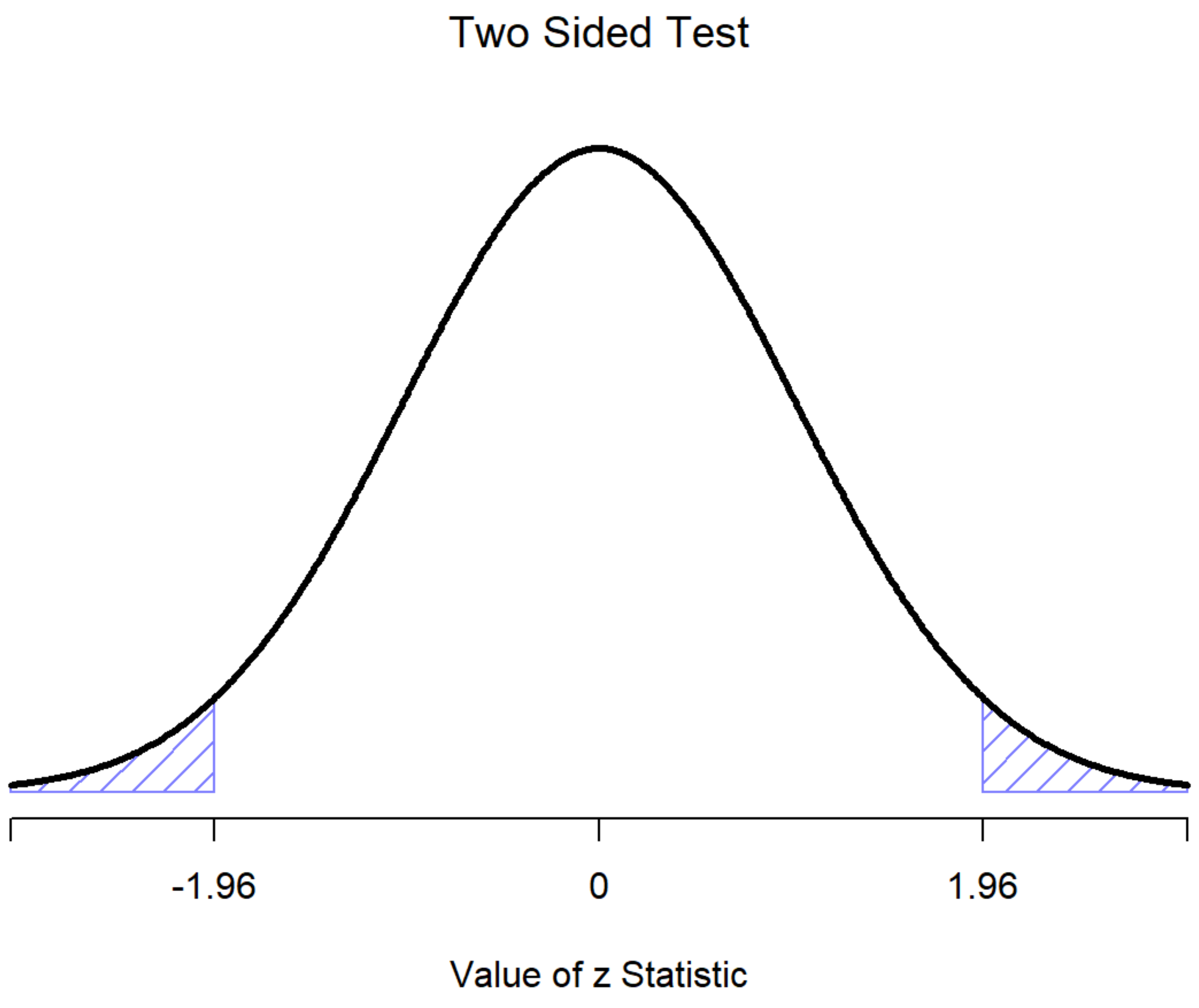


$$z_{\bar{X}} = \frac{\bar{X} - \mu_0}{\mathrm{SE}(\bar{X})}$$

the 5% critical regions for 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





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 do know the <u>estimated</u> standard deviation.

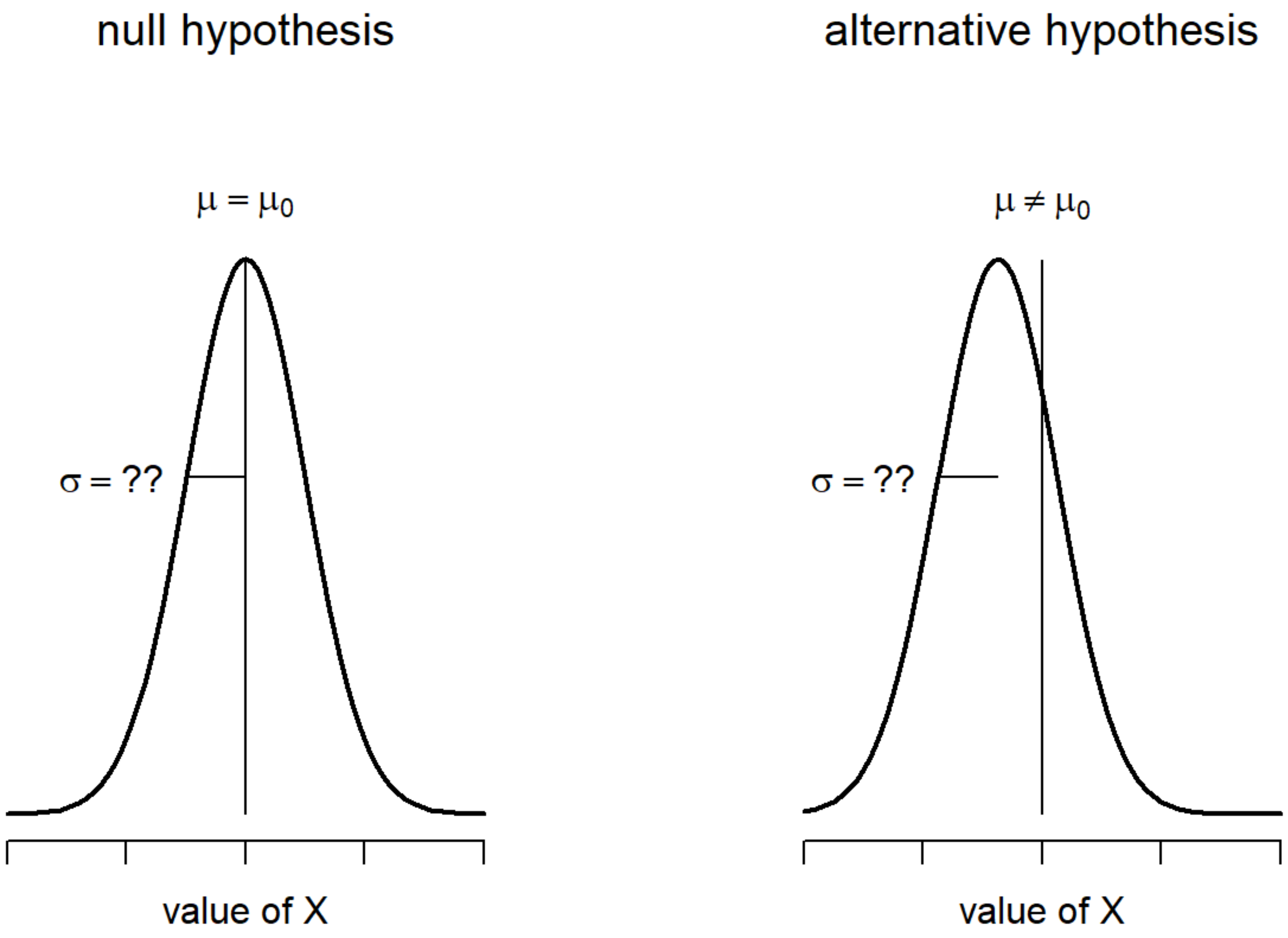
$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n - 1}}$$

From z-test to t-test

We <u>could</u> 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

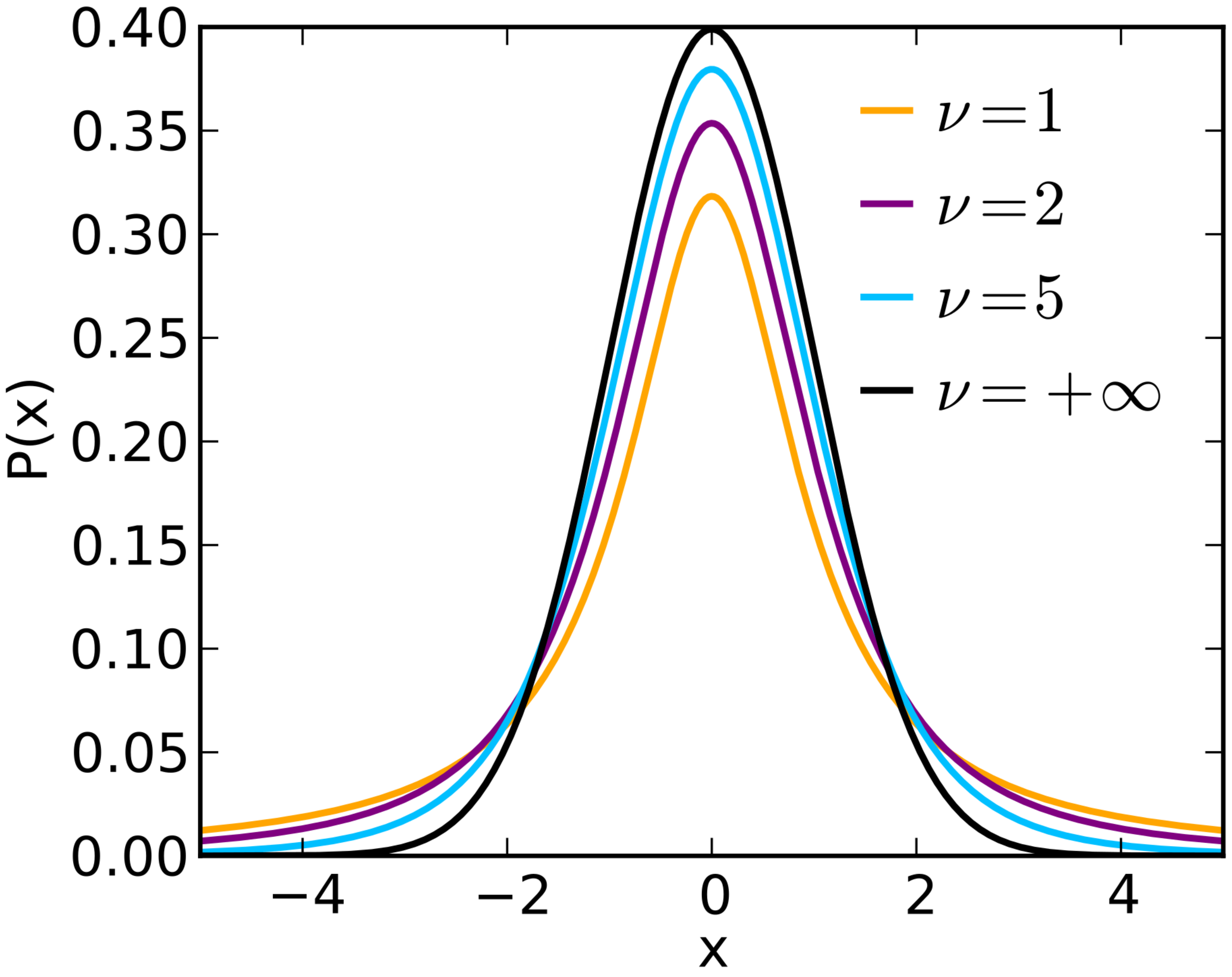
To accommodate the fact we don't know the s.d, we need to subtly change the sampling distribution.

The t-test statistic remains very similar:
The only 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 <u>degrees of freedom</u>.

The one-sample t-test



Assumptions of the one sample t-test

Randomness

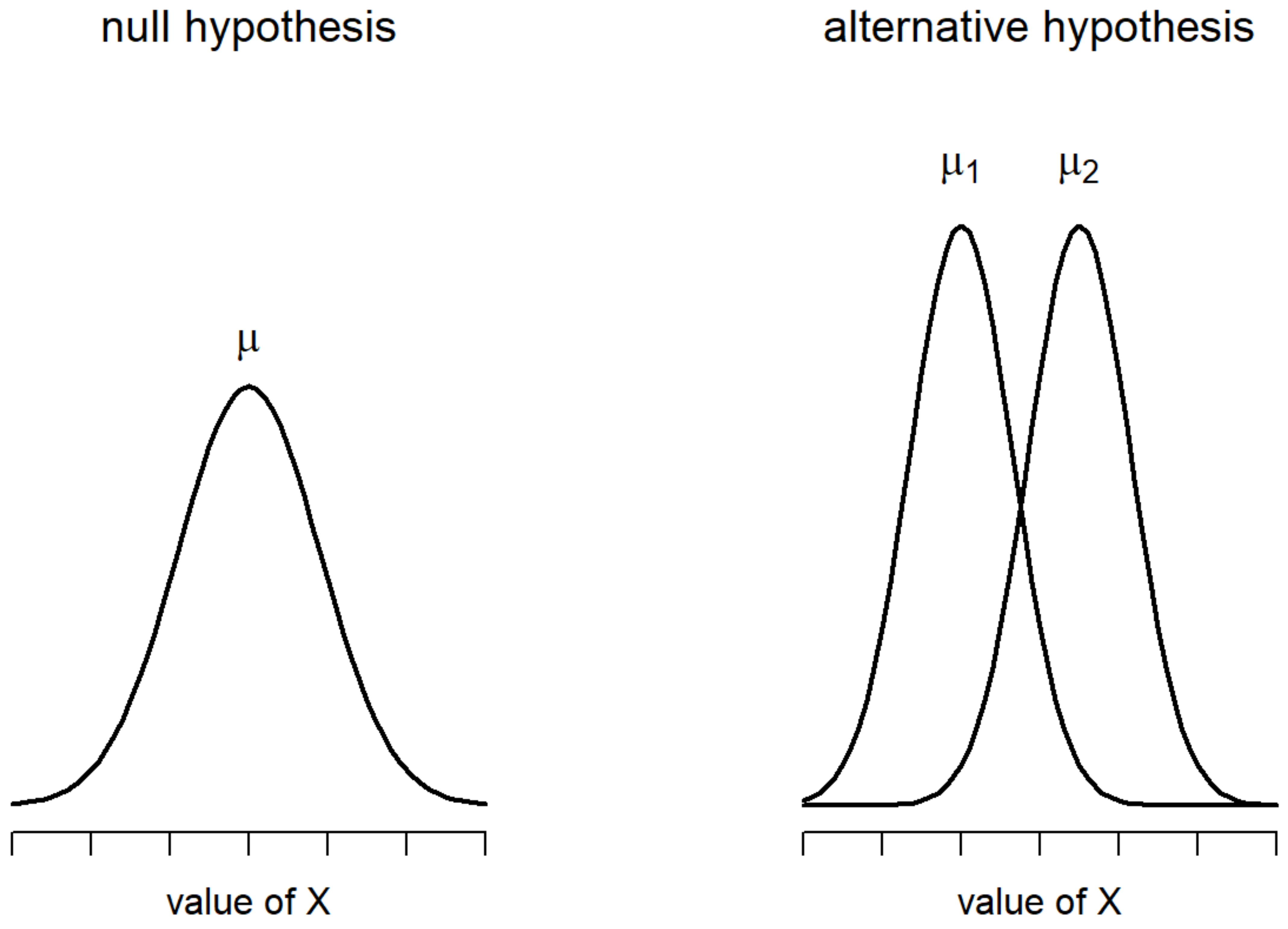
Normality

Independence

Two groups have the same population standard deviation

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

The independent-samples t-test



If the null is true we expect $\bar{X}_1 - \bar{X}_2 = 0$.

But how close?

Test statistics is:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\mathrm{SE}}$$

The independent-samples t-test

To calculate the "pooled" estimate of the variance estimate of the variance we take a weighed average of the variance estimates.

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

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

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} \longrightarrow \hat{\sigma}_p = \sqrt{\frac{w_1 {\hat{\sigma}_1}^2 + w_2 {\hat{\sigma}_2}^2}{w_1 + w_2}}$$

The independent-samples t-test

To calculate the "pooled" estimate of the variance estimate of the variance we take a weighed average of the variance estimates.

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

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

Assumptions of the one sample t-test

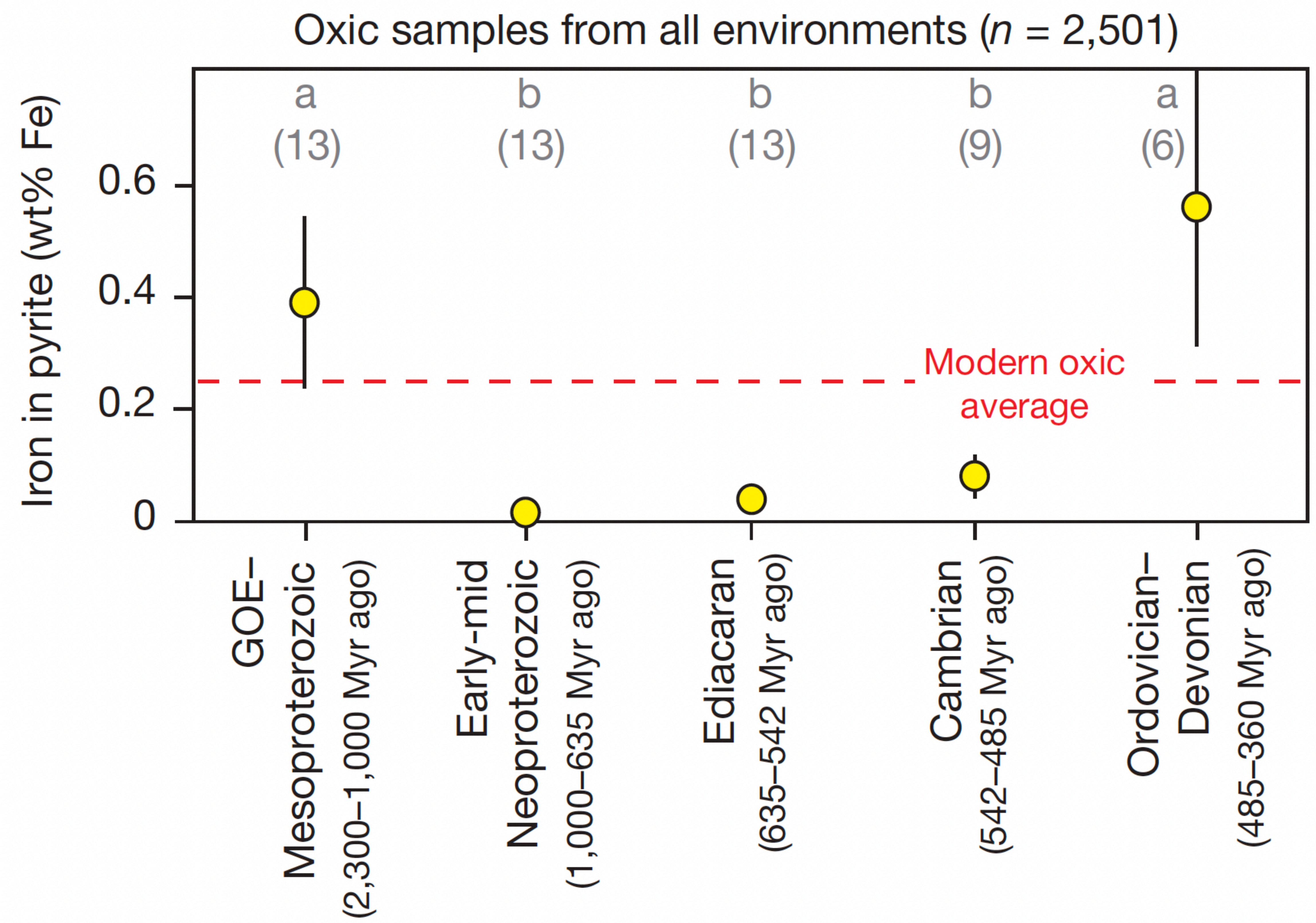
Randomness

Normality

Independence

One-way ANOVA (Analysis of variances)

What if you want to compare more than 2 groups?



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

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

First we calculate the variance:

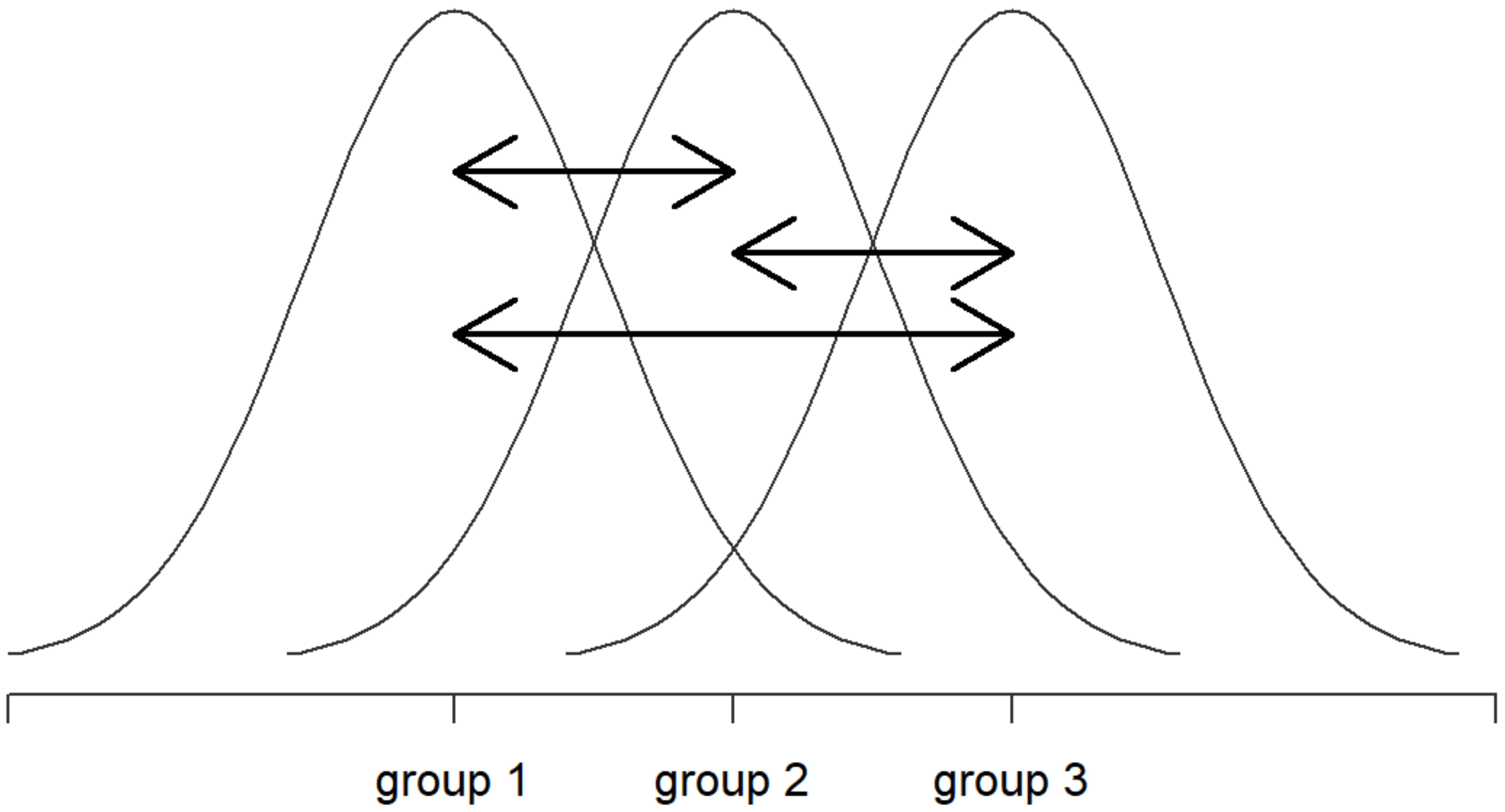
$$\text{Var}(Y) = \frac{1}{N}\sum_{k=1}^{G}\sum_{i=1}^{N_k}\left(Y_{ik} - \bar{Y}\right)^2 \longrightarrow \text{Var}(Y) = \frac{1}{N}\sum_{p=1}^{N}\left(Y_p - \bar{Y}\right)^2$$

Then we calculate the total group (SS tot), within group (SS w) and between group (SS b) **sum of squares**:

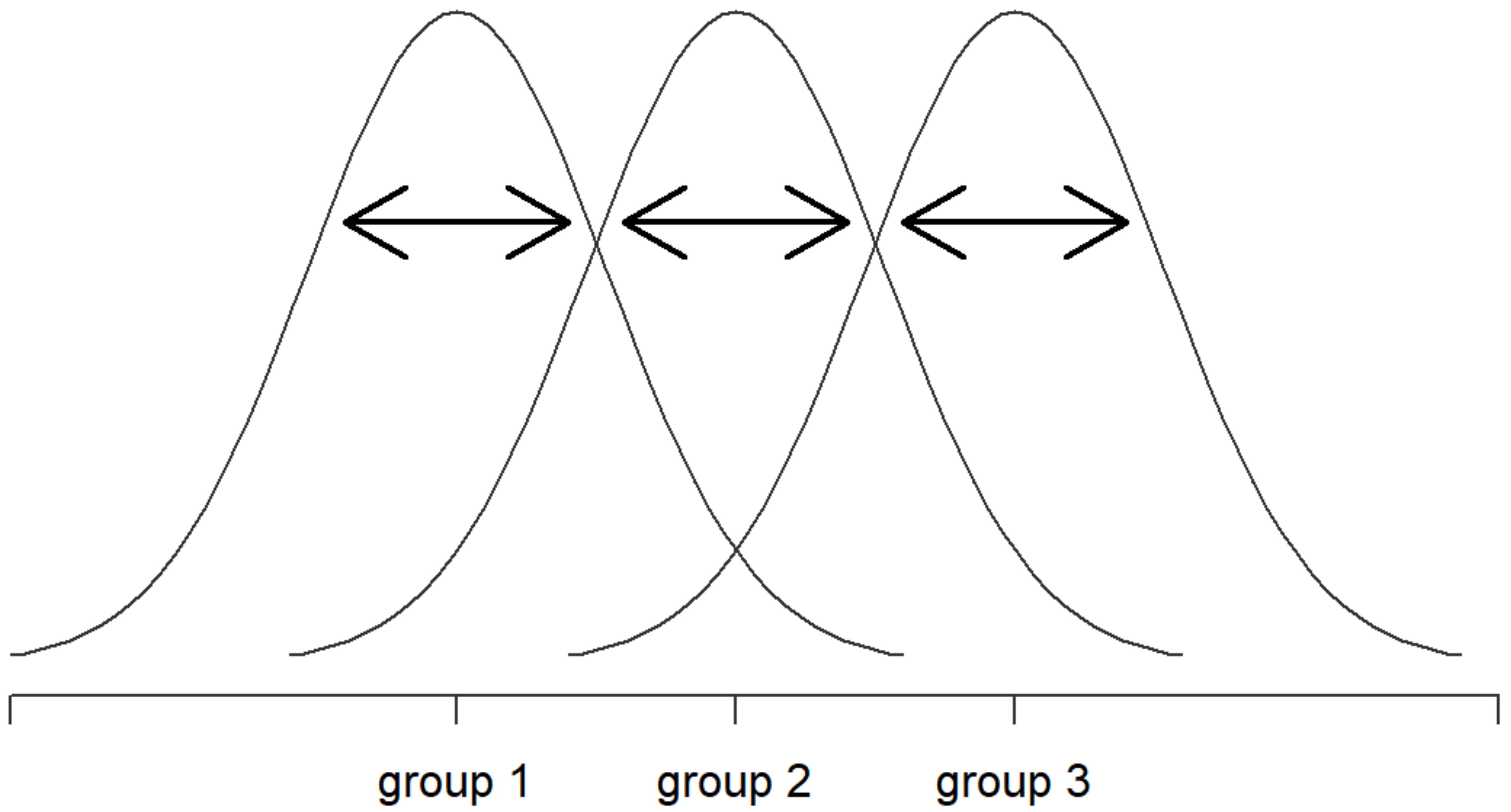
$$\begin{aligned} \text{SS}_b &= \textstyle\sum_{k=1}^{G}\sum_{i=1}^{N_k}\left(\bar{Y}_k - \bar{Y}\right)^2 \\ &= \textstyle\sum_{k=1}^{G} N_k\left(\bar{Y}_k - \bar{Y}\right)^2 \end{aligned}$$

Between group variation (SS b)

i.e. diffs between group means



Within group variation (SS w)



One-way ANOVA (Analysis of variances)

the total variability associated with the outcome variable (SS tot) can be mathematically broken up into 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).

If the null hypothesis is true, then you’d expect all the sample means to be pretty similar to each other, right?

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

We calculate is the degrees of freedom associated with the SS b and SS w values.

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

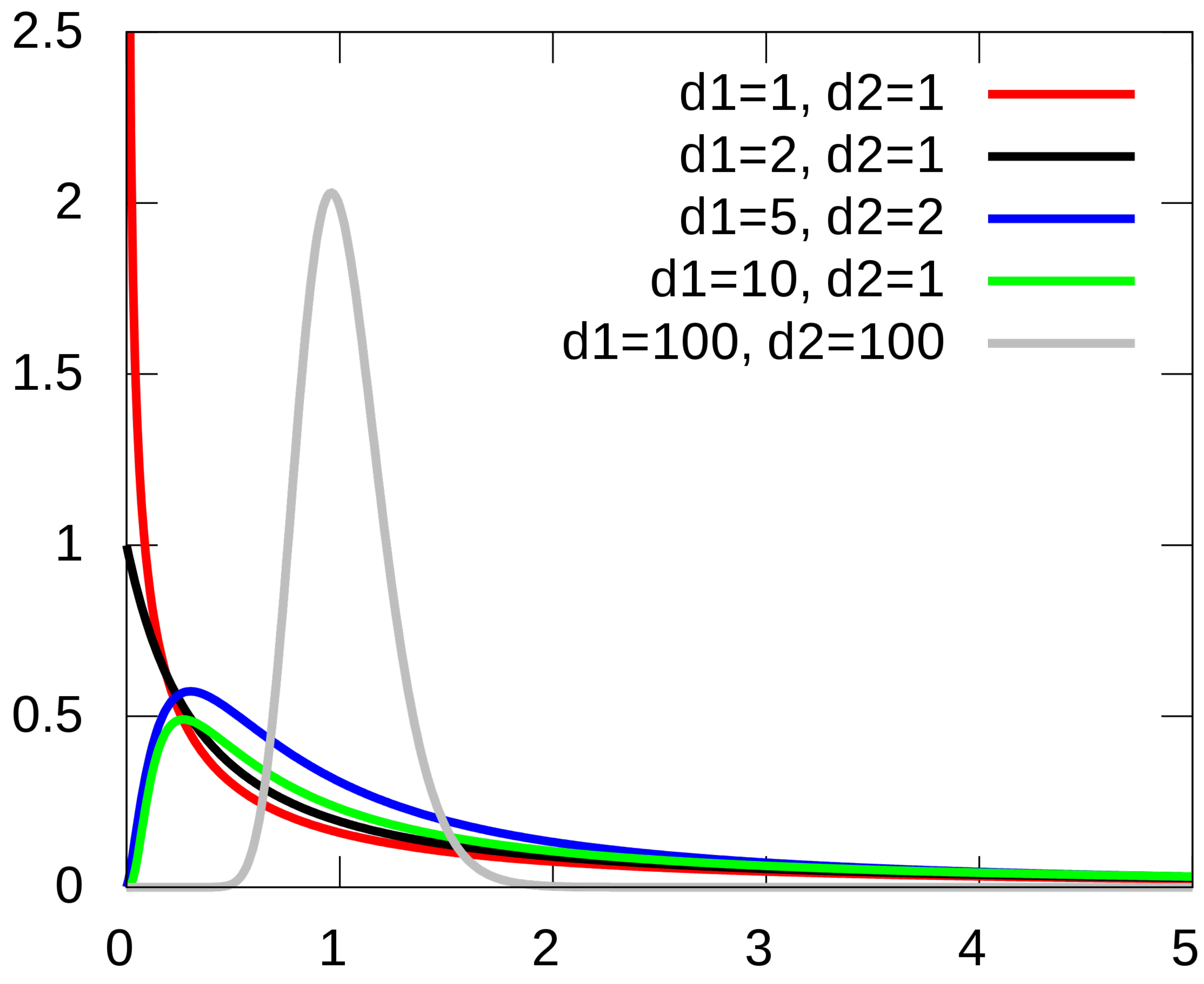
next is convert our summed squares value into a "mean squares" value, by dividing by the degrees of freedom.

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

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.



Today'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