t-tests (comparing means)

Ling250/450: Data Science for Linguistics
C.M. Downey
Spring 2025

Quick review

Quick review

- We test new ideas/claims with **statistical hypothesis testing**

Quick review

- We test new ideas/claims with **statistical hypothesis testing**
- We always assume the **negation** of our hypothesis, called the **Null hypothesis**, until we provide evidence otherwise

Quick review

- We test new ideas/claims with **statistical hypothesis testing**
- We always assume the **negation** of our hypothesis, called the **Null hypothesis**, until we provide evidence otherwise
    - **Refuting** the Null is how we provide **evidence for our new claim**

Quick review

- We test new ideas/claims with **statistical hypothesis testing**
- We always assume the **negation** of our hypothesis, called the **Null hypothesis**, until we provide evidence otherwise
  - **Refuting** the Null is how we provide **evidence for our new claim**
  - We refute the Null by **minimizing** a test's **significance value** (α)

Quick review

- We test new ideas/claims with **statistical hypothesis testing**
- We always assume the **negation** of our hypothesis, called the **Null hypothesis**, until we provide evidence otherwise
  - **Refuting** the Null is how we provide **evidence for our new claim**
  - We refute the Null by **minimizing** a test's **significance value** (α)
- One measure of significance is called the **p-value**

Quick review

- We test new ideas/claims with **statistical hypothesis testing**
- We always assume the **negation** of our hypothesis, called the **Null hypothesis**, until we provide evidence otherwise
  - **Refuting** the Null is how we provide **evidence for our new claim**
  - We refute the Null by **minimizing** a test's **significance value** (α)
- One measure of significance is called the **p-value**
  - Interpretable as the **probability** that we are **incorrectly refuting** the Null

Quick review

- We test new ideas/claims with **statistical hypothesis testing**
- We always assume the **negation** of our hypothesis, called the **Null hypothesis**, until we provide evidence otherwise
  - **Refuting** the Null is how we provide **evidence for our new claim**
  - We refute the Null by **minimizing** a test's **significance value** (α)
- One measure of significance is called the **p-value**
  - Interpretable as the **probability** that we are **incorrectly refuting** the Null
  - OR as the **probability** of seeing **data like ours**, when **assuming the Null**

Quick review

- We test new ideas/claims with **statistical hypothesis testing**
- We always assume the **negation** of our hypothesis, called the **Null hypothesis**, until we provide evidence otherwise
  - **Refuting** the Null is how we provide **evidence for our new claim**
  - We refute the Null by **minimizing** a test's **significance value** (α)
- One measure of significance is called the **p-value**
  - Interpretable as the **probability** that we are **incorrectly refuting** the Null
  - OR as the **probability** of seeing **data like ours**, when **assuming the Null**
  - $p < 0.05$ means "less than 5% chance the Null would give us the data we observe"

Purpose of t-test

Purpose of t-test

- A type of **statistical test** to check if a sample has the **same mean** as something else (either a **known mean** or a **second sample**)

Purpose of t-test

- A type of **statistical test** to check if a sample has the **same mean** as something else (either a **known mean** or a **second sample**)
- We'll look at **three variants**, each with its own purpose

Purpose of t-test

- A type of **statistical test** to check if a sample has the **same mean** as something else (either a **known mean** or a **second sample**)
- We'll look at **three variants**, each with its own purpose
  - **One-sample**: checks if sample comes from a population with a certain mean

Purpose of t-test

- A type of **statistical test** to check if a sample has the **same mean** as something else (either a **known mean** or a **second sample**)
- We'll look at **three variants**, each with its own purpose
  - **One-sample**: checks if sample comes from a population with a certain mean
  - **Independent samples (Student's)**: checks if two samples have the same population mean (assumes populations have same standard deviation)

Purpose of t-test

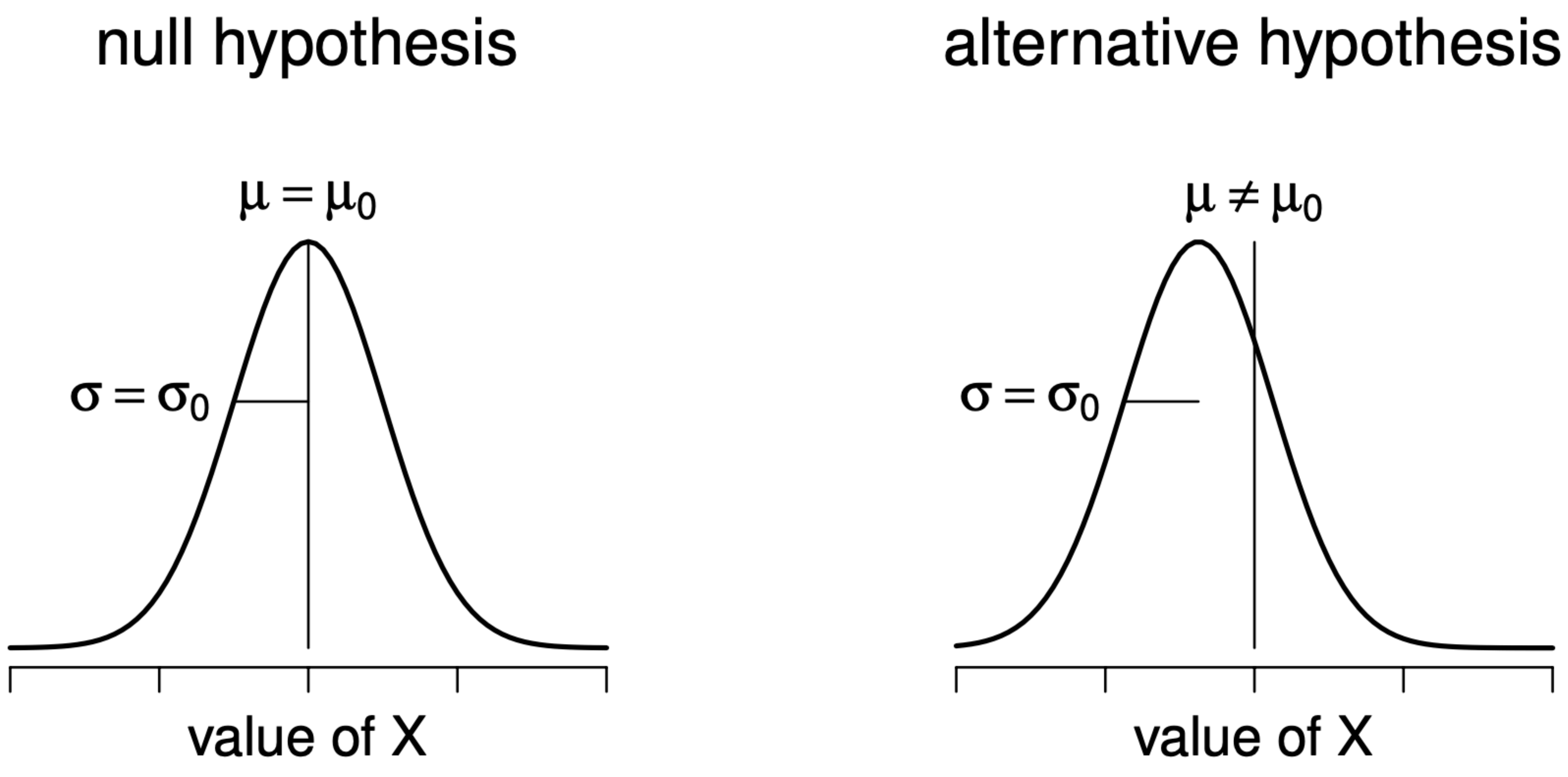
- A type of **statistical test** to check if a sample has the **same mean** as something else (either a **known mean** or a **second sample**)
- We'll look at **three variants**, each with its own purpose
  - **One-sample**: checks if sample comes from a population with a certain mean
  - **Independent samples (Student's)**: checks if two samples have the same population mean (assumes populations have same standard deviation)
  - **Independent samples (Welch)**: same as Student's, but **does not** assume both populations have same standard deviation

Purpose of t-test

- A type of **statistical test** to check if a sample has the **same mean** as something else (either a **known mean** or a **second sample**)
- We'll look at **three variants**, each with its own purpose
  - **One-sample**: checks if sample comes from a population with a certain mean
  - **Independent samples (Student's)**: checks if two samples have the same population mean (assumes populations have same standard deviation)
  - **Independent samples (Welch)**: same as Student's, but **does not** assume both populations have same standard deviation
- We'll only look at the math of the one-sample test in detail

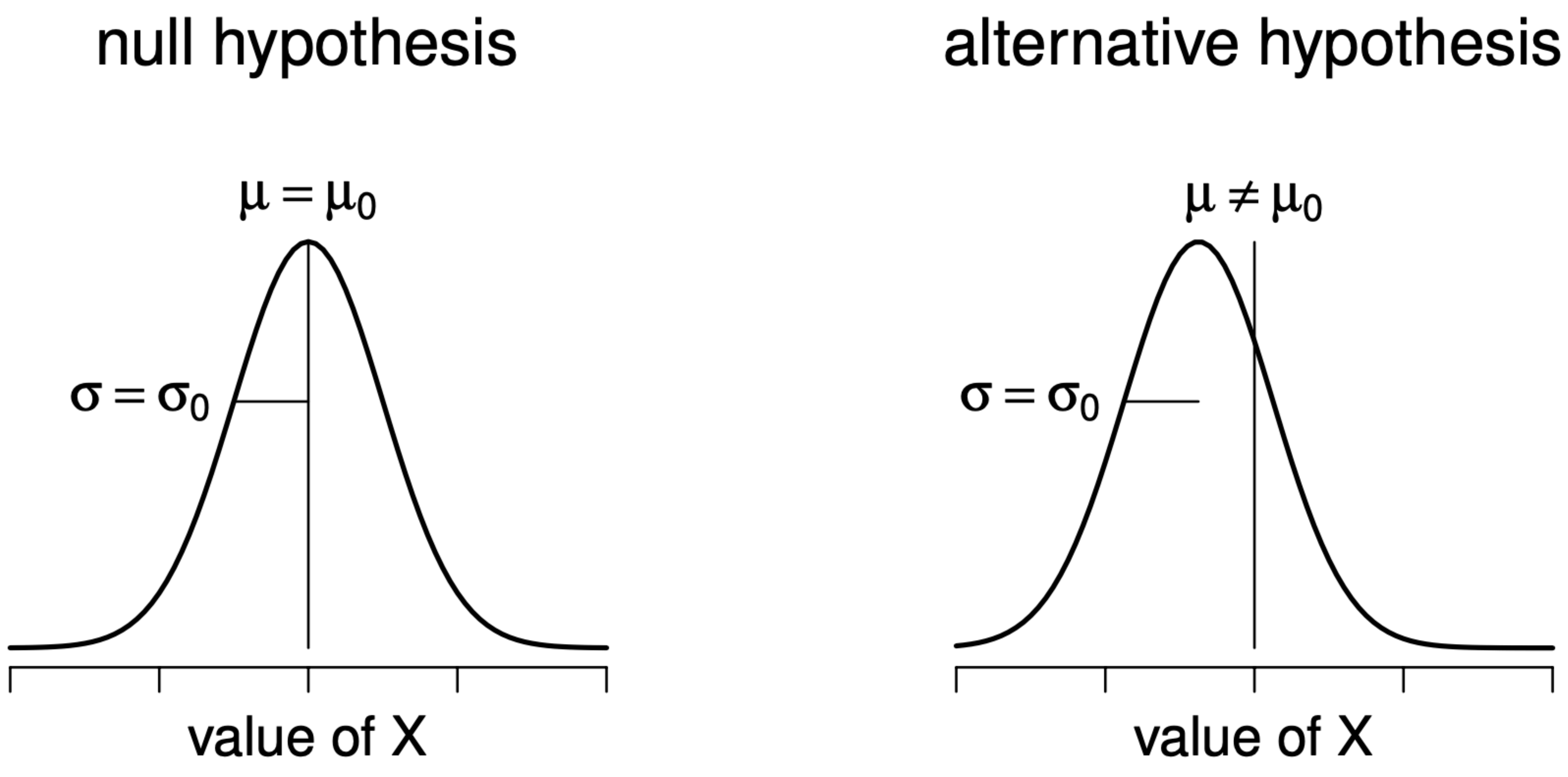
One-sample t-test

First, the z-test



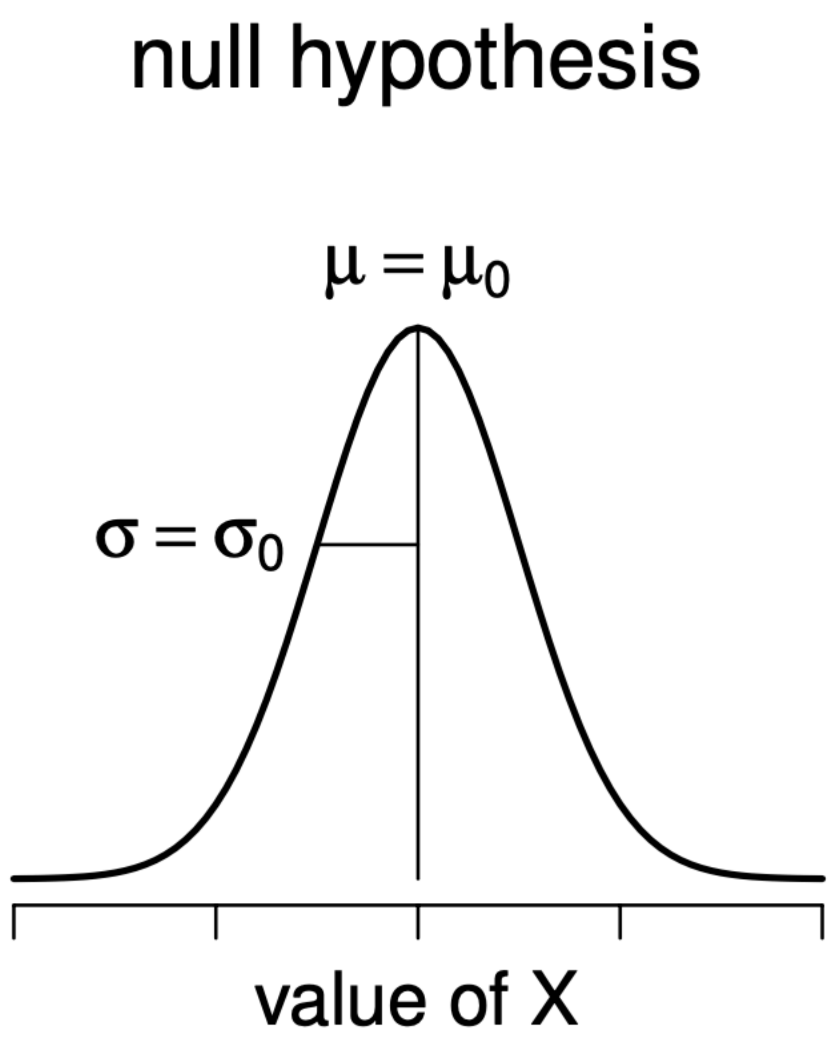
First, the z-test

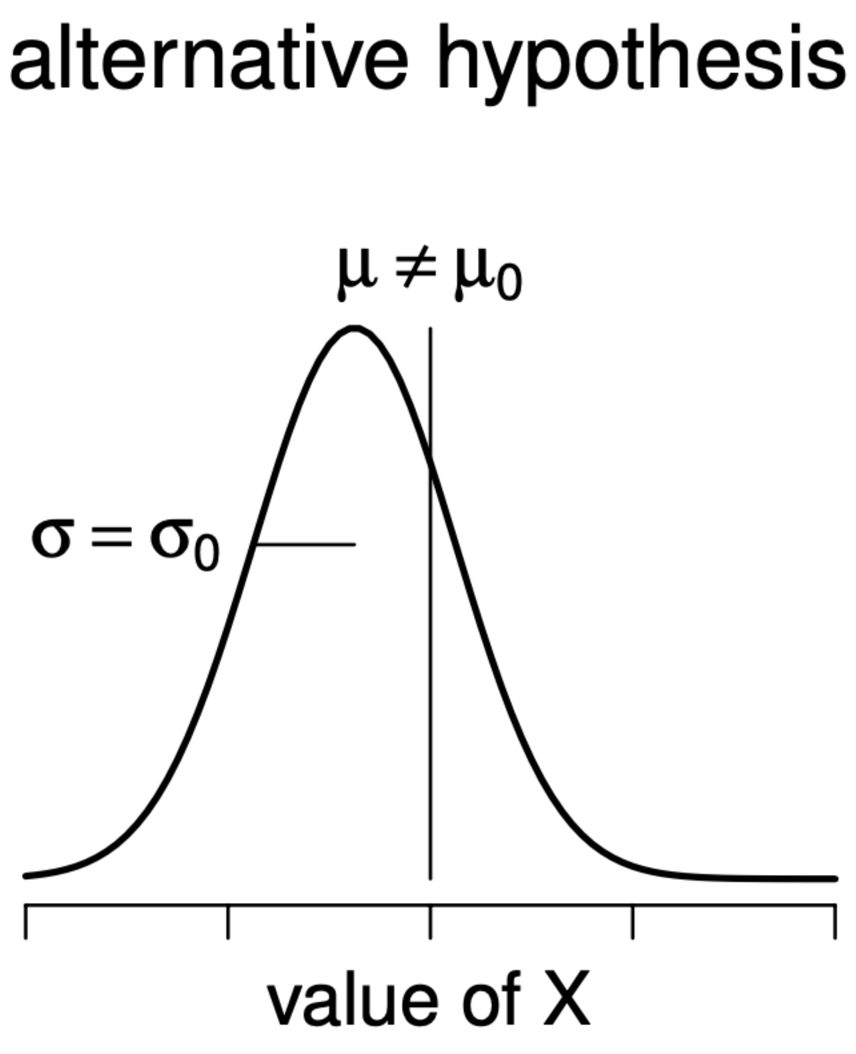
- As a **"stepping-stone"** to learning the t-test, we'll describe the **z-test**



First, the z-test

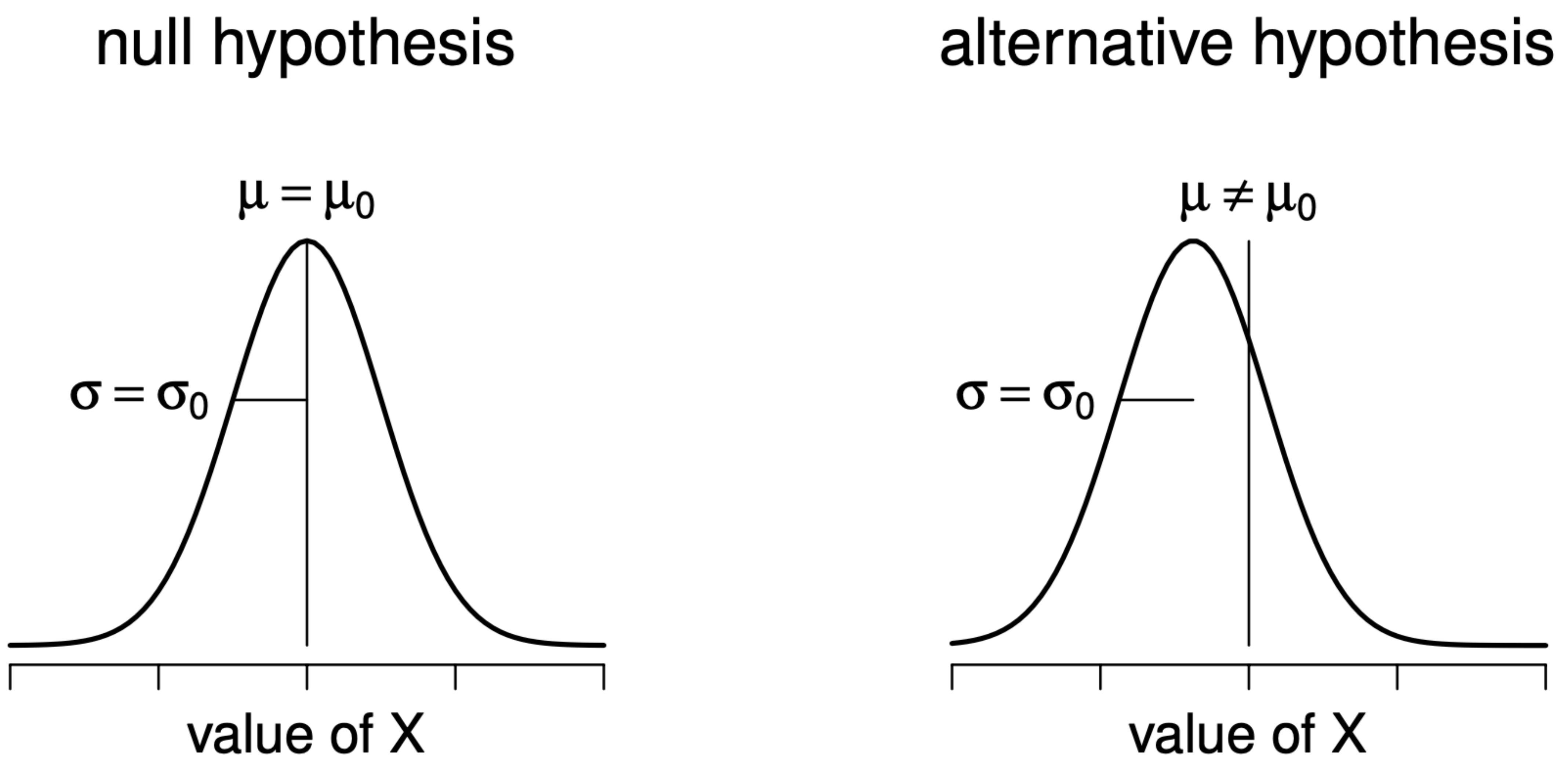
- As a **"stepping-stone"** to learning the t-test, we'll describe the **z-test**
- Slightly **simpler** than the t-test, but **never used in practice**





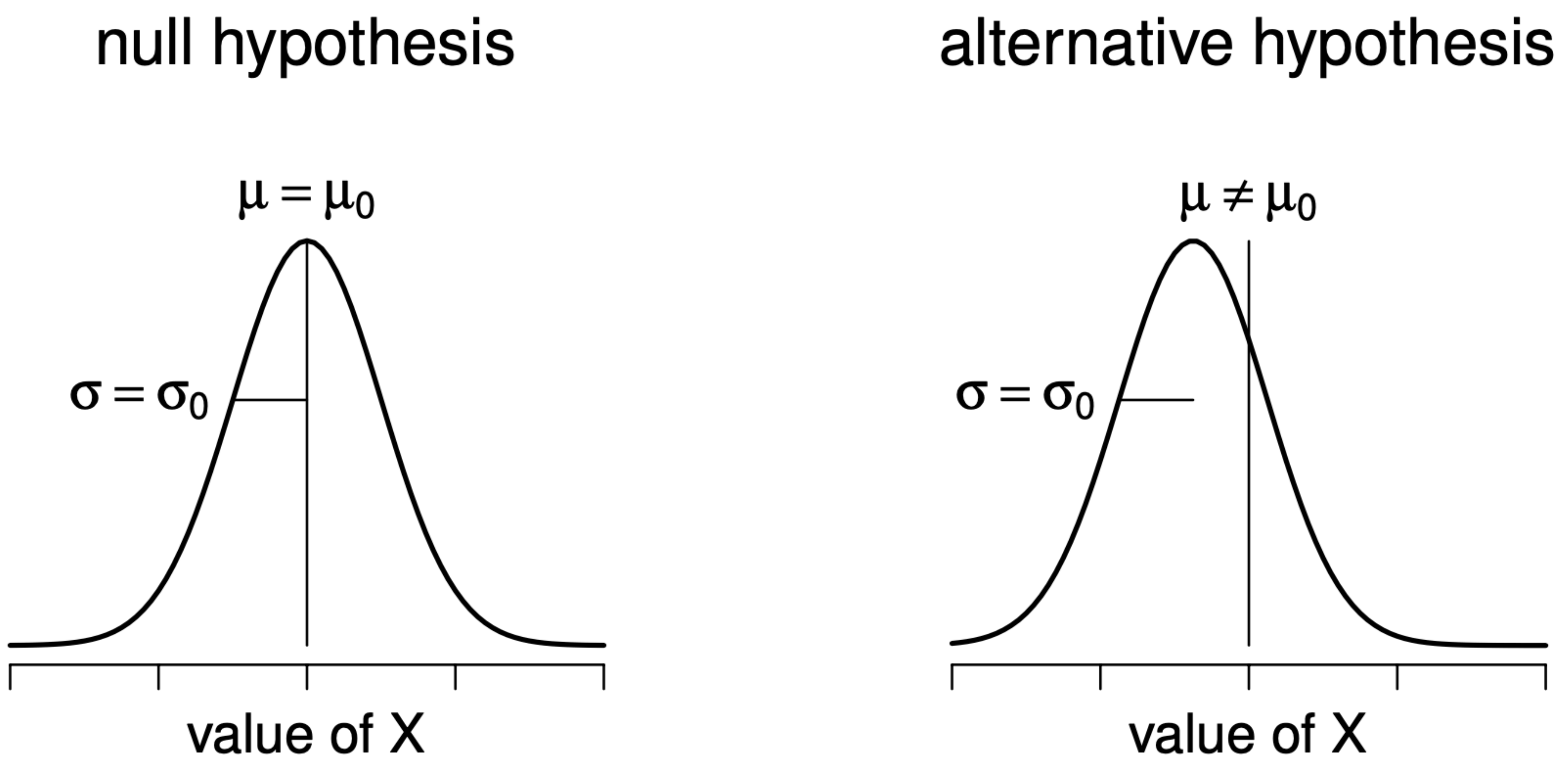
First, the z-test

- As a **"stepping-stone"** to learning the t-test, we'll describe the **z-test**
- Slightly **simpler** than the t-test, but **never used in practice**
- This is a **one-sample** test



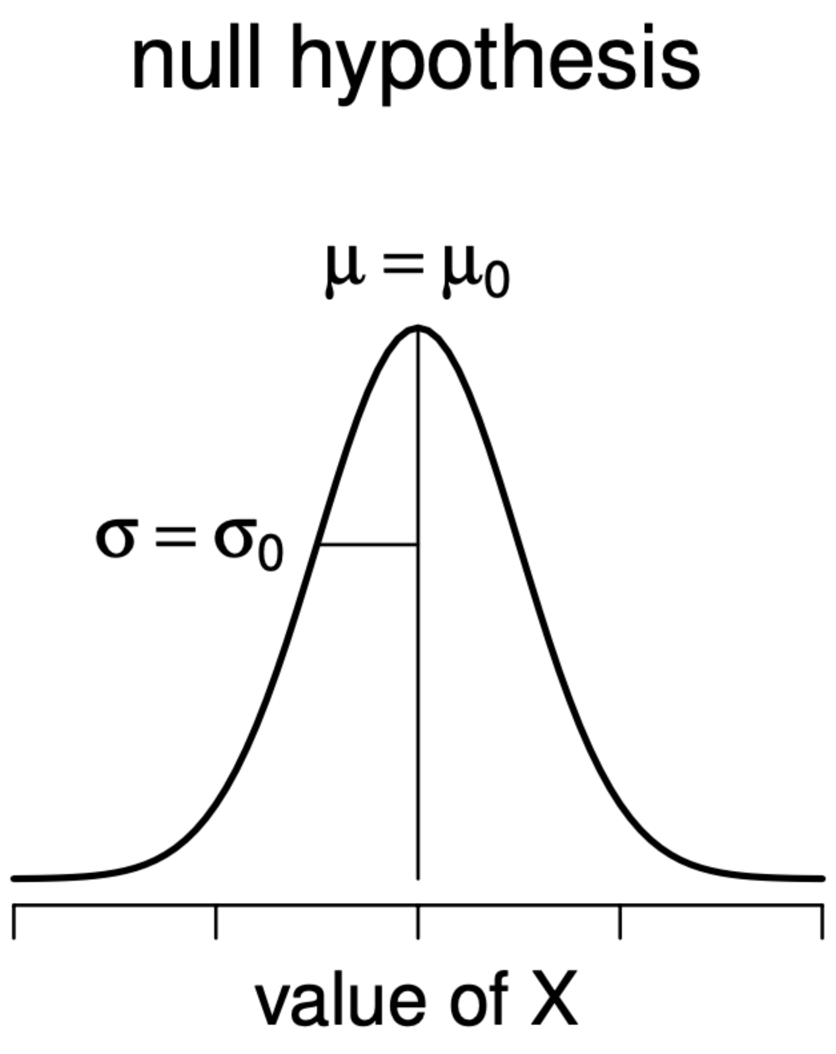
First, the z-test

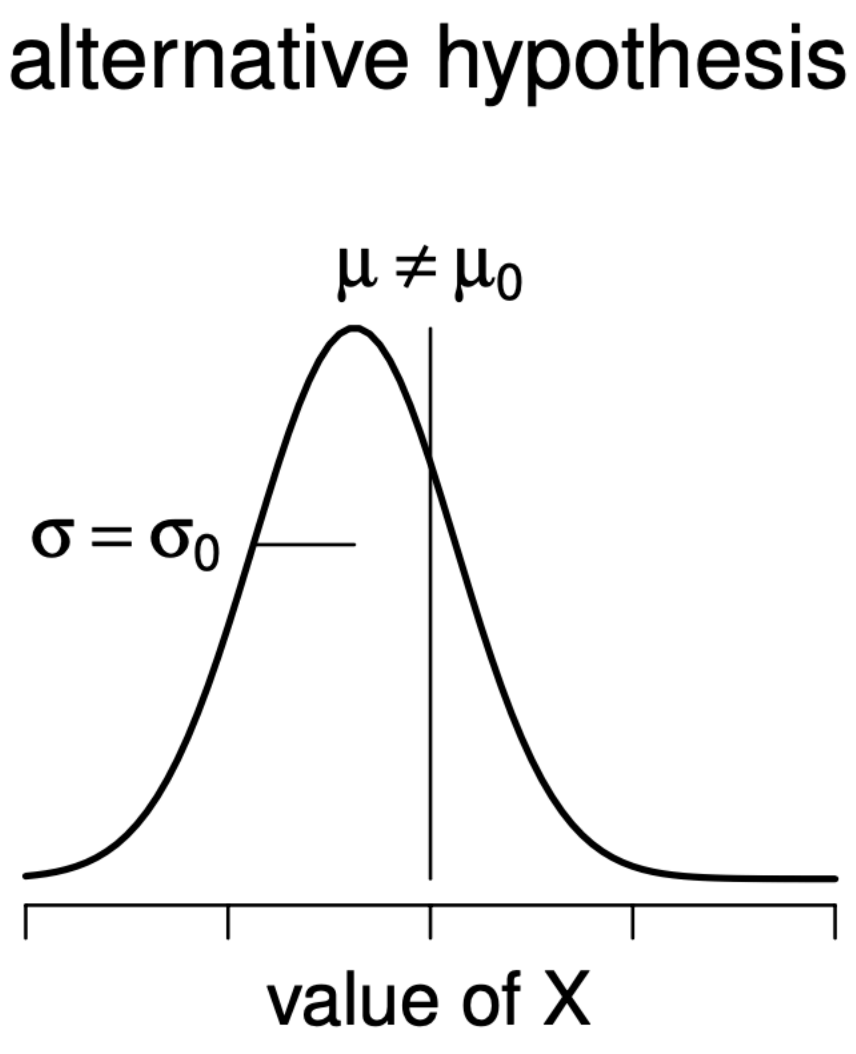
- As a **"stepping-stone"** to learning the t-test, we'll describe the **z-test**
- Slightly **simpler** than the t-test, but **never used in practice**
- This is a **one-sample** test
  - We want to know if a **sample** comes from a **population** with a **certain mean**



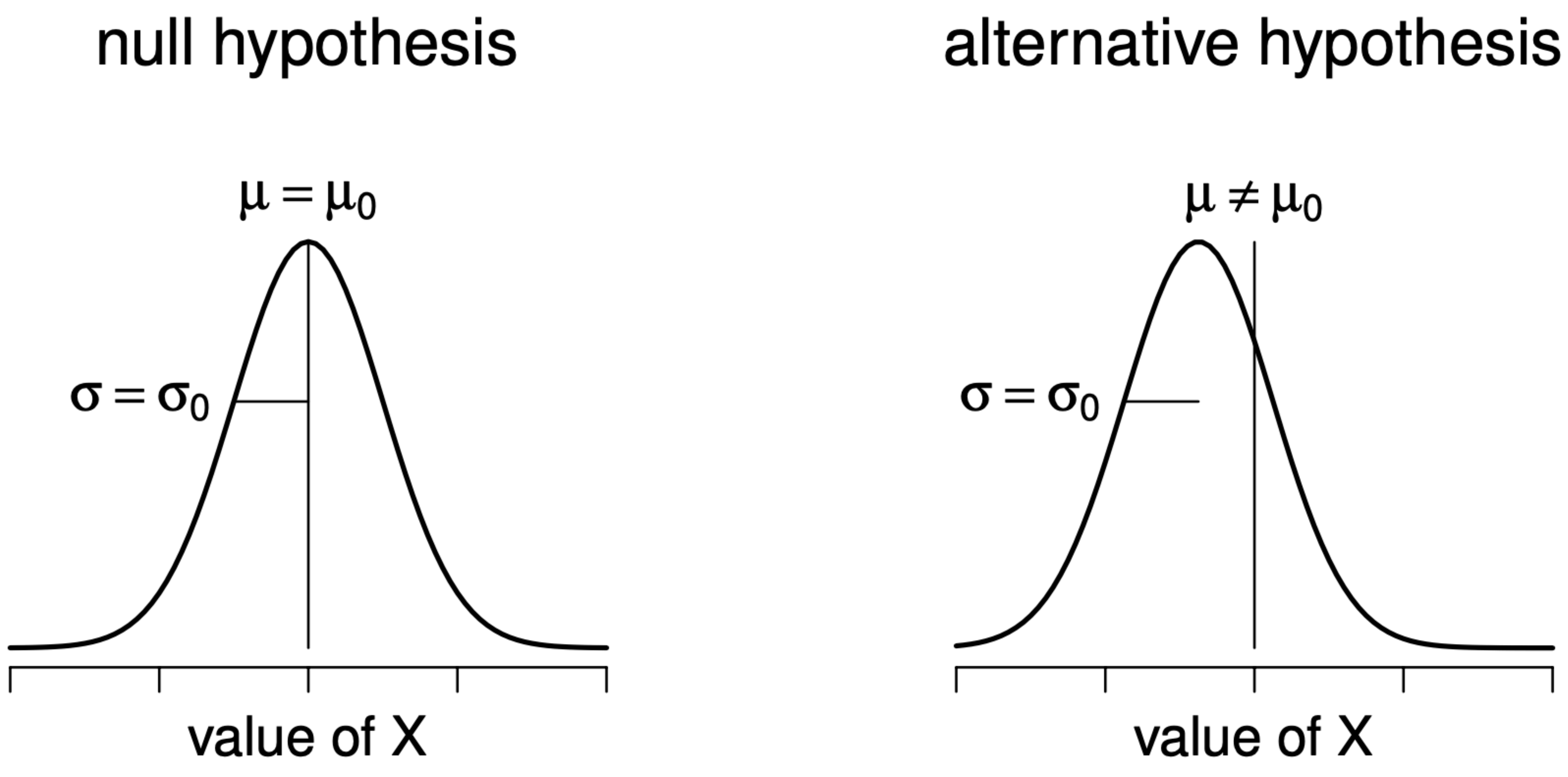
First, the z-test

- As a **"stepping-stone"** to learning the t-test, we'll describe the **z-test**
- Slightly **simpler** than the t-test, but **never used in practice**
- This is a **one-sample** test
  - We want to know if a **sample** comes from a **population** with a **certain mean**
  - Pretend example: do linguistics students have the **same average grade** as the entire class taking a foreign language?



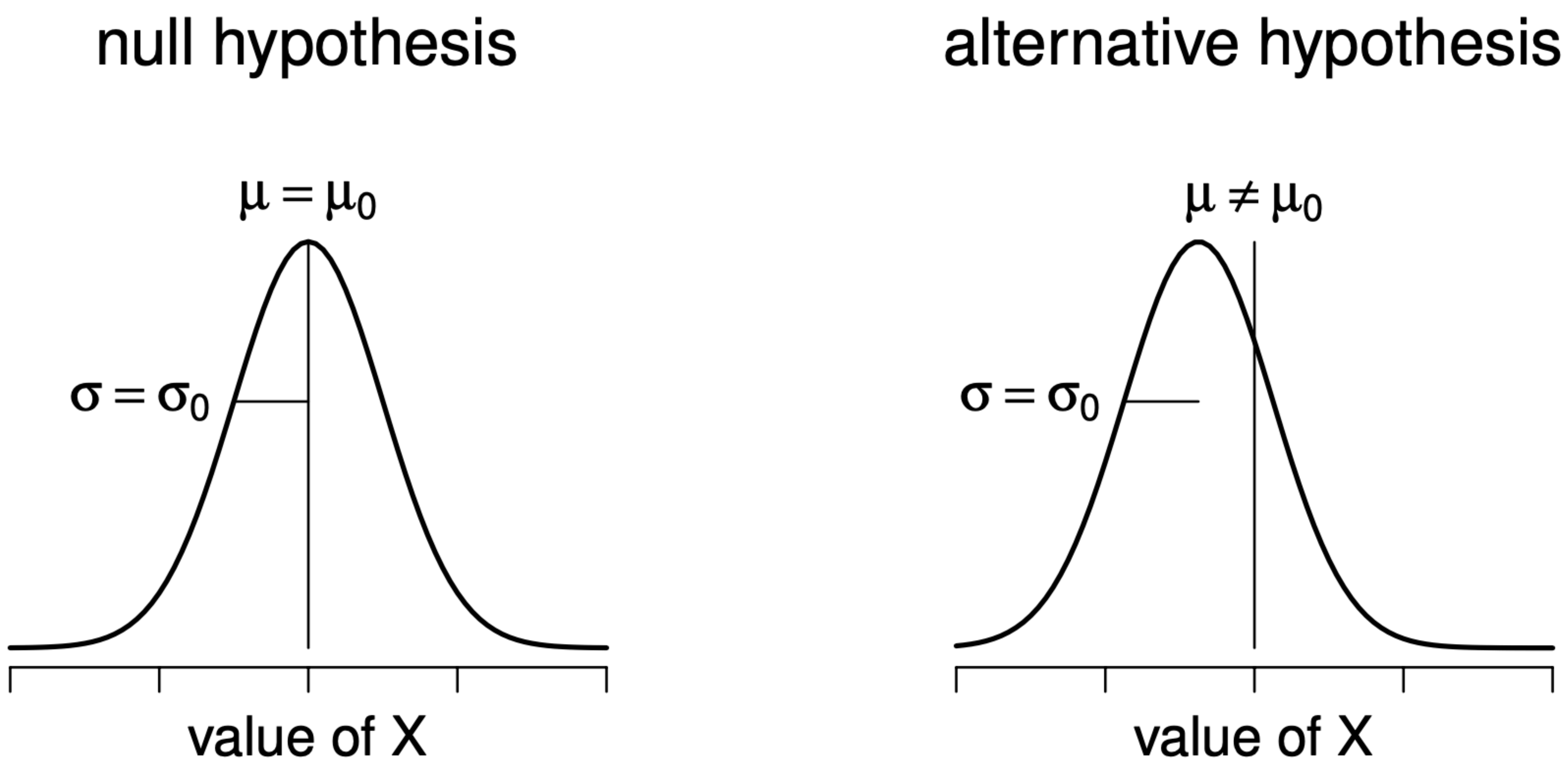


z-test hypothesis



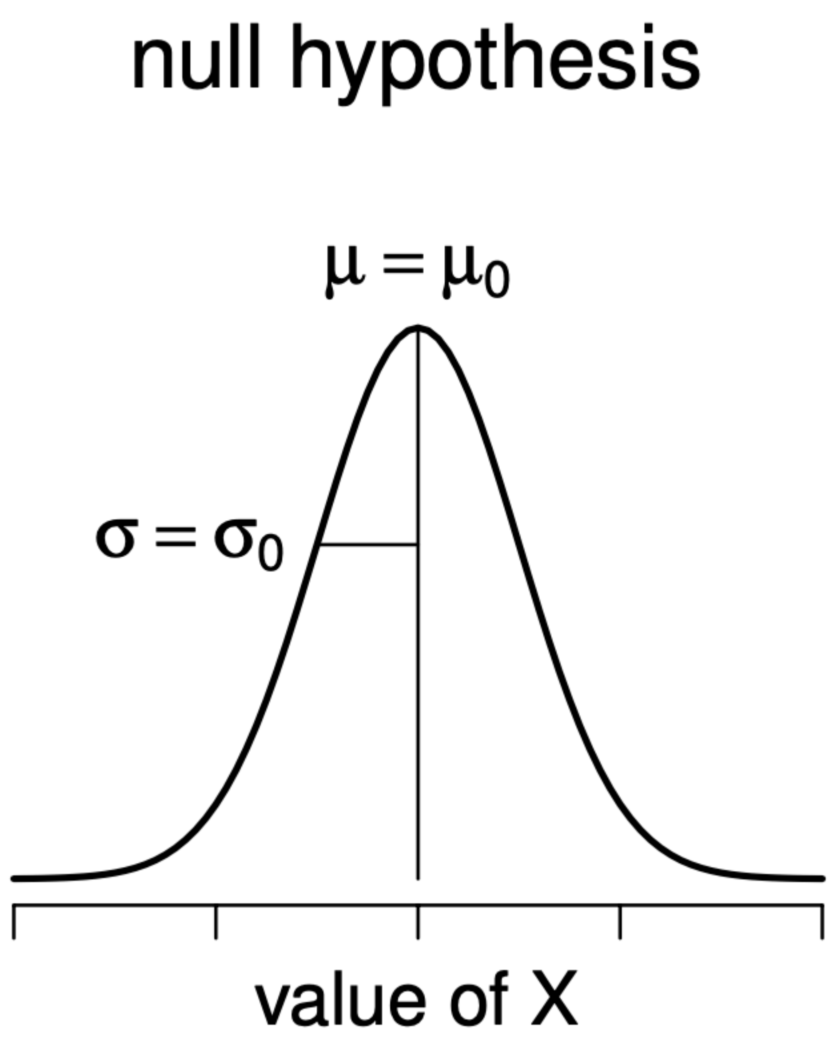
z-test hypothesis

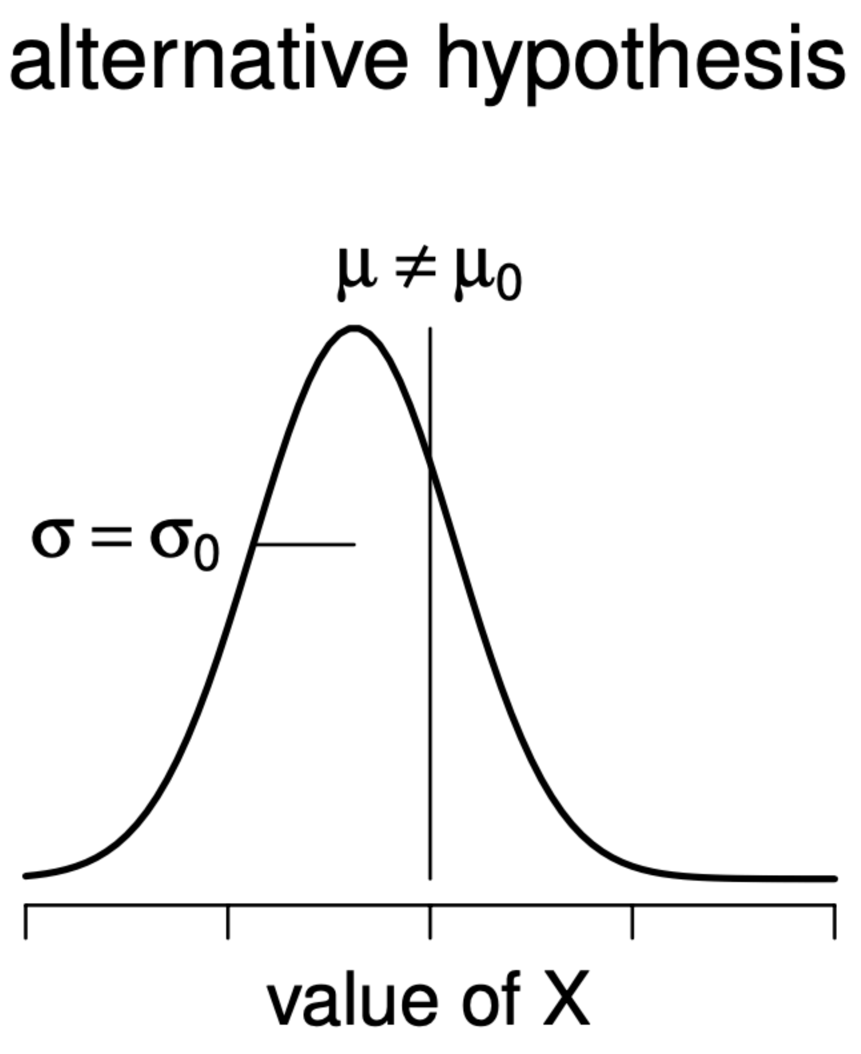
- Our **Null hypothesis** is that the Ling students have the **same mean grade** as the course overall
  - μ_0 is the mean course grade (67.5)
  - $H_0 : \mu = \mu_0 = 67.5$



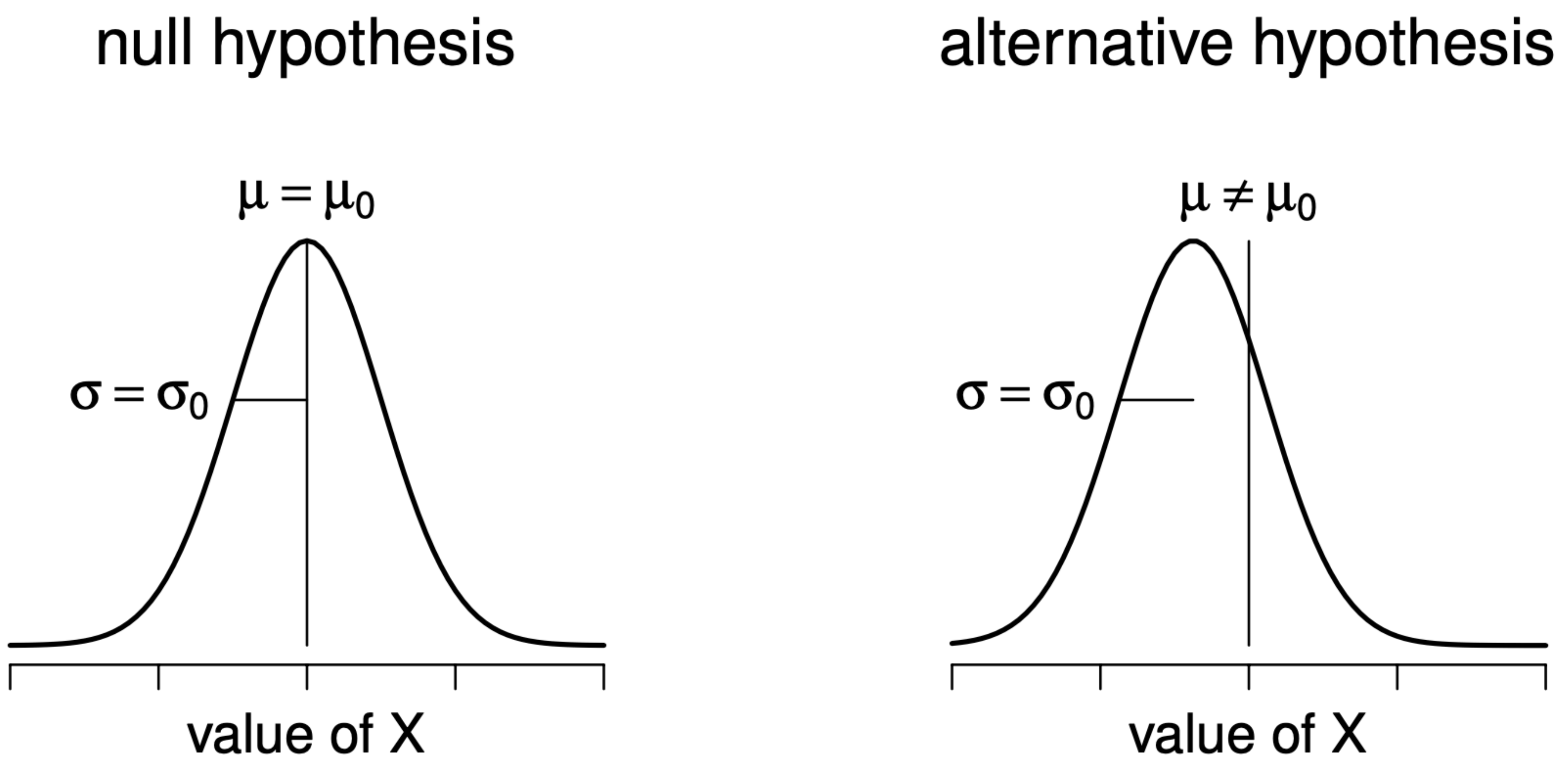
z-test hypothesis

- Our **Null hypothesis** is that the Ling students have the **same mean grade** as the course overall
  - μ_0 is the mean course grade (67.5)
  - $H_0 : \mu = \mu_0 = 67.5$
- The **alternative hypothesis** is the Ling students have a **different mean grade**
  - $H_1 : \mu \neq \mu_0 = 67.5$



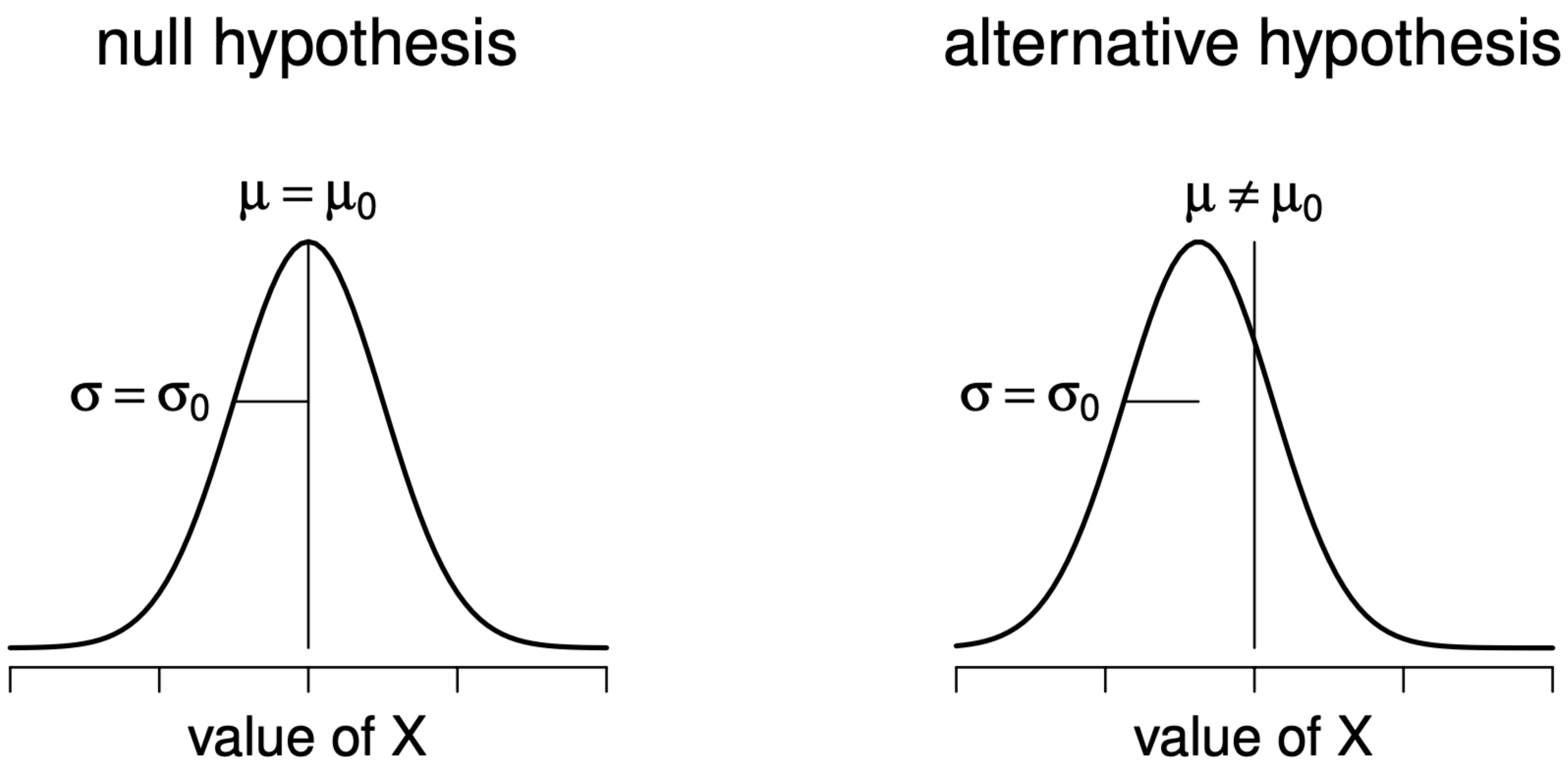


z-test assumptions



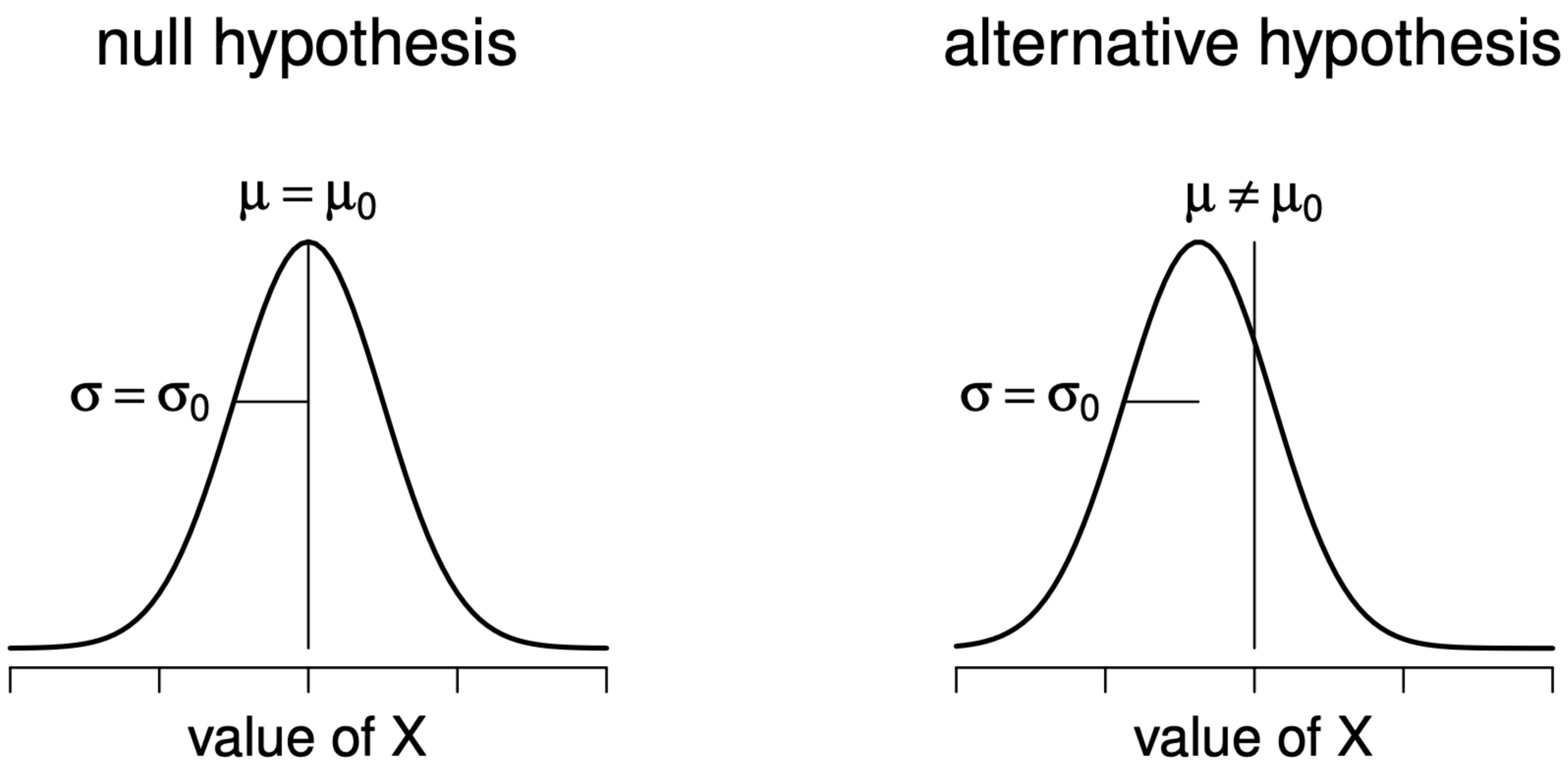
z-test assumptions

- **Big assumption**: we assume we know the **standard deviation** of the **population we compare against** (e.g. the whole class; σ_0)
  - Also that our **sample population** (Ling students) has the same stdev
  - These are **bad assumptions**, which will lead us to the t-test



z-test assumptions

- **Big assumption**: we assume we know the **standard deviation** of the **population we compare against** (e.g. the whole class; σ_0)
  - Also that our **sample population** (Ling students) has the same stdev
  - These are **bad assumptions**, which will lead us to the t-test
- We also assume the **sampling distribution** of the mean is **Normal**
  - This is usually a **safe-ish** assumption (see lecture on probability distributions)



z-test outline

z-test outline

- We want to make an **inference** about the **population** of Ling students in general (their average grade μ) based on a **sample** (X; those currently taking the class)

z-test outline

- We want to make an **inference** about the **population** of Ling students in general (their average grade μ) based on a **sample** (X; those currently taking the class)
  - $\overline{X}$ is the **sample mean**; N is the **sample size**

z-test outline

- We want to make an **inference** about the **population** of Ling students in general (their average grade μ) based on a **sample** (X; those currently taking the class)
  - $\overline{X}$ is the **sample mean**; N is the **sample size**
- **Null hypothesis**: Ling students have the **same average grade** as other students ($\mu = \mu_0 = 67.5$)

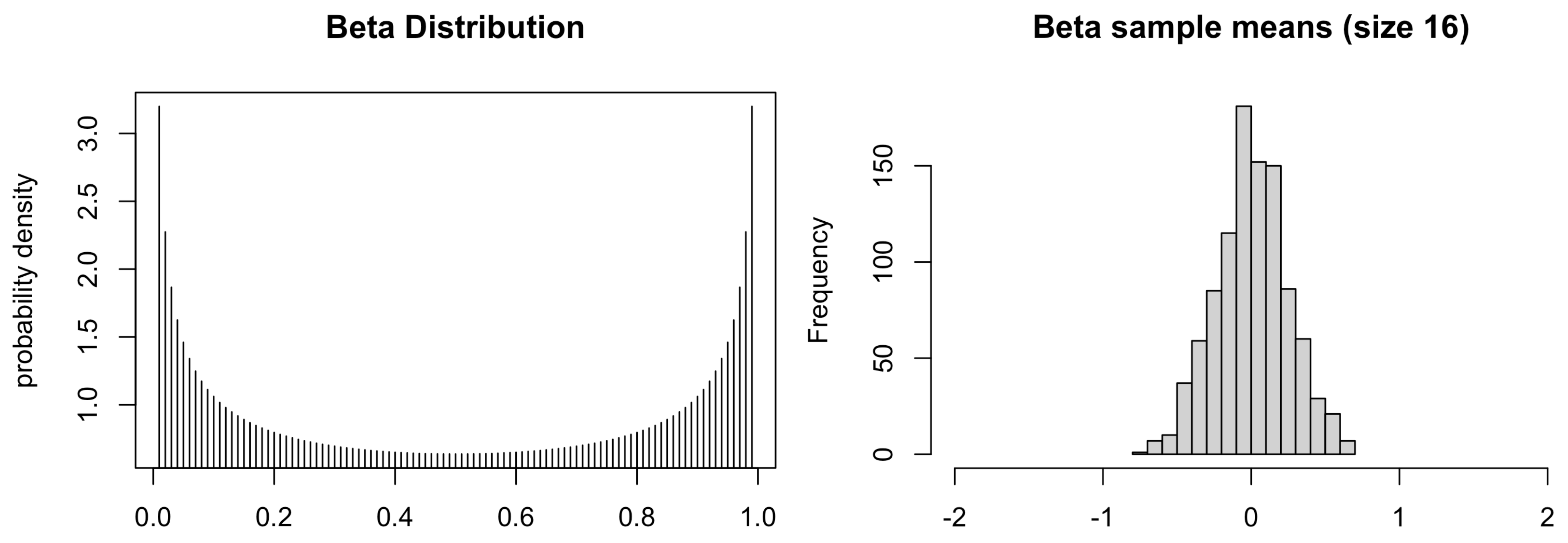
z-test outline

- We want to make an **inference** about the **population** of Ling students in general (their average grade μ) based on a **sample** (X; those currently taking the class)
  - $\overline{X}$ is the **sample mean**; N is the **sample size**
- **Null hypothesis**: Ling students have the **same average grade** as other students ($\mu = \mu_0 = 67.5$)
  - If $\overline{X}$ is **close** to μ_0, this is evidence **for** the Null

z-test outline

- We want to make an **inference** about the **population** of Ling students in general (their average grade μ) based on a **sample** (X; those currently taking the class)
  - $\bar{X}$ is the **sample mean**; N is the **sample size**
- **Null hypothesis**: Ling students have the **same average grade** as other students ($\mu = \mu_0 = 67.5$)
  - If $\bar{X}$ is **close** to μ_0, this is evidence **for** the Null
  - If $\bar{X}$ is **far** from μ_0, this is evidence **against** the Null

Recall: Sampling distributions

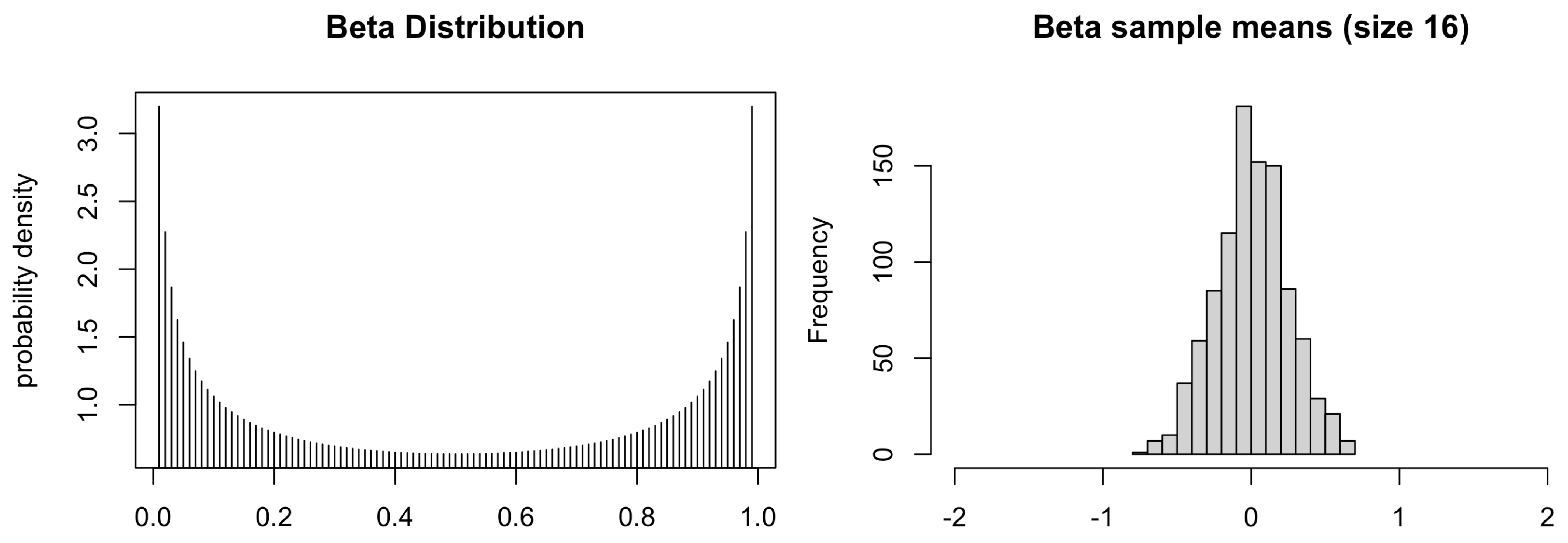


$$SEM(\bar{X}) = \frac{\sigma}{\sqrt{N}}$$

$$\bar{X} \sim \text{Normal}(\mu, SEM(\bar{X}))$$

Recall: Sampling distributions

- Remember we can model the **distribution of sampling means**

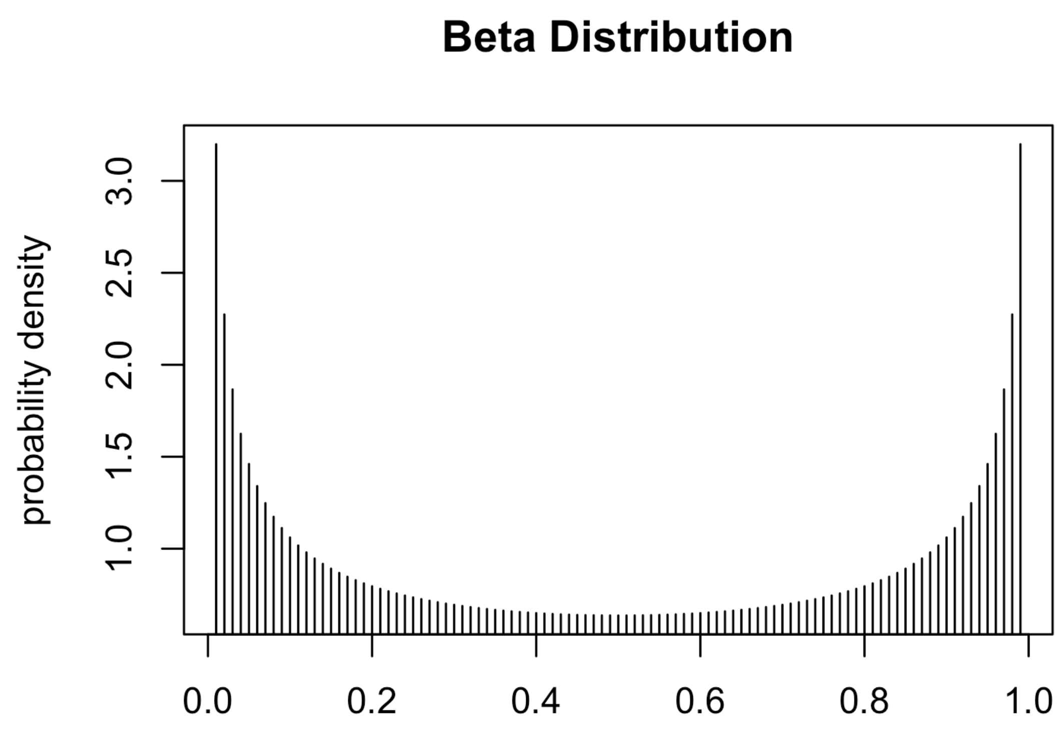


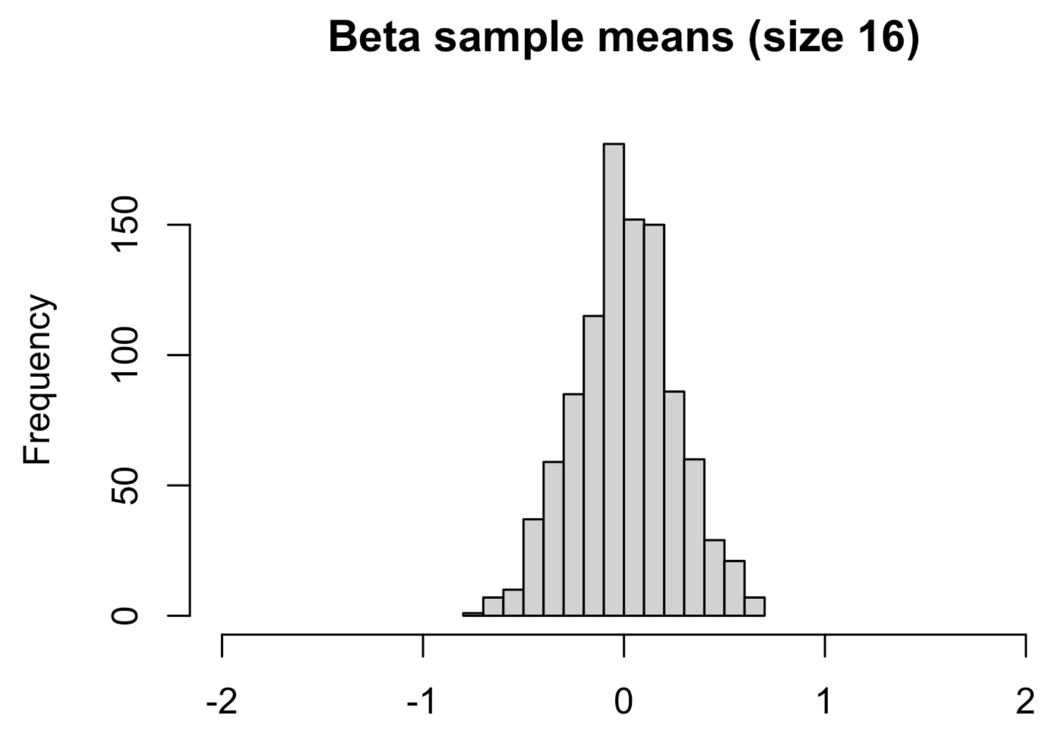
$$SEM(\bar{X}) = \frac{\sigma}{\sqrt{N}}$$

$$\bar{X} \sim \text{Normal}(\mu, SEM(\bar{X}))$$

Recall: Sampling distributions

- Remember we can model the **distribution of sampling means**
  - i.e. for **any sample** X of size N from a **certain population**, what is the **expected value** of $\overline{X}$?



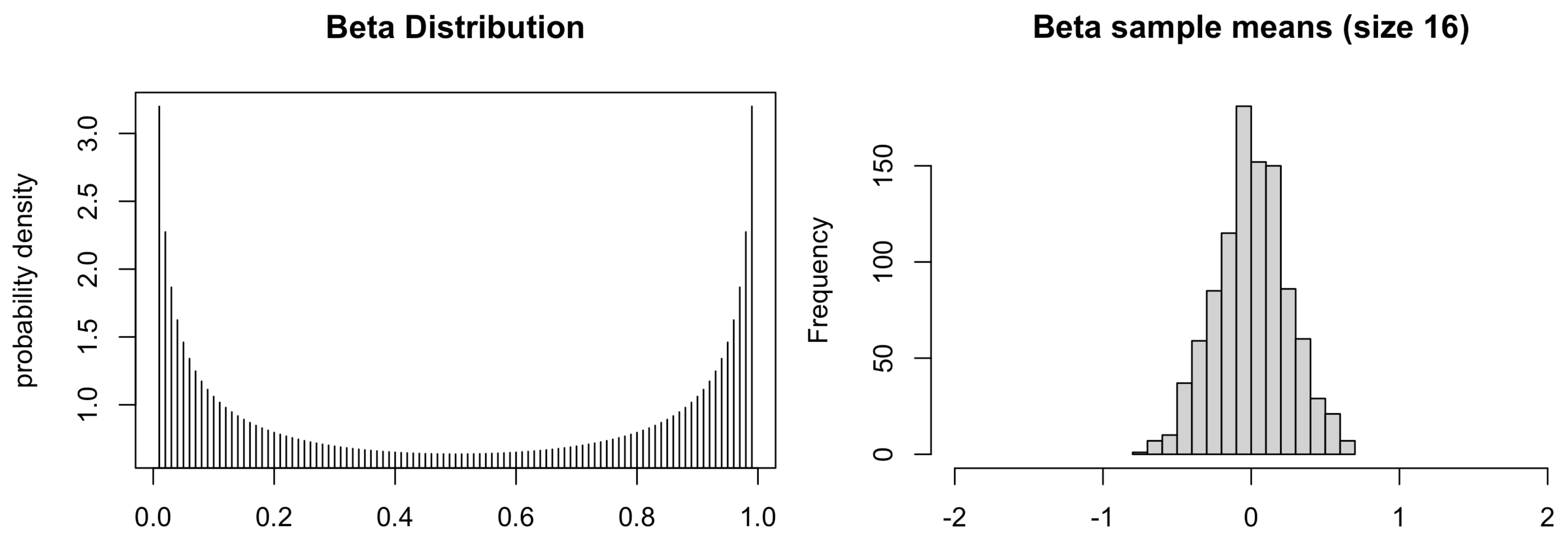


$$SEM(\overline{X}) = \frac{\sigma}{\sqrt{N}}$$

$$\overline{X} \sim \text{Normal}(\mu, SEM(\overline{X}))$$

Recall: Sampling distributions

- Remember we can model the **distribution of sampling means**
  - i.e. for **any sample** X of size N from a **certain population**, what is the **expected value** of $\bar{X}$?
  - This distribution is **(almost) always Normal**, and **centered** on the **population mean** (μ)

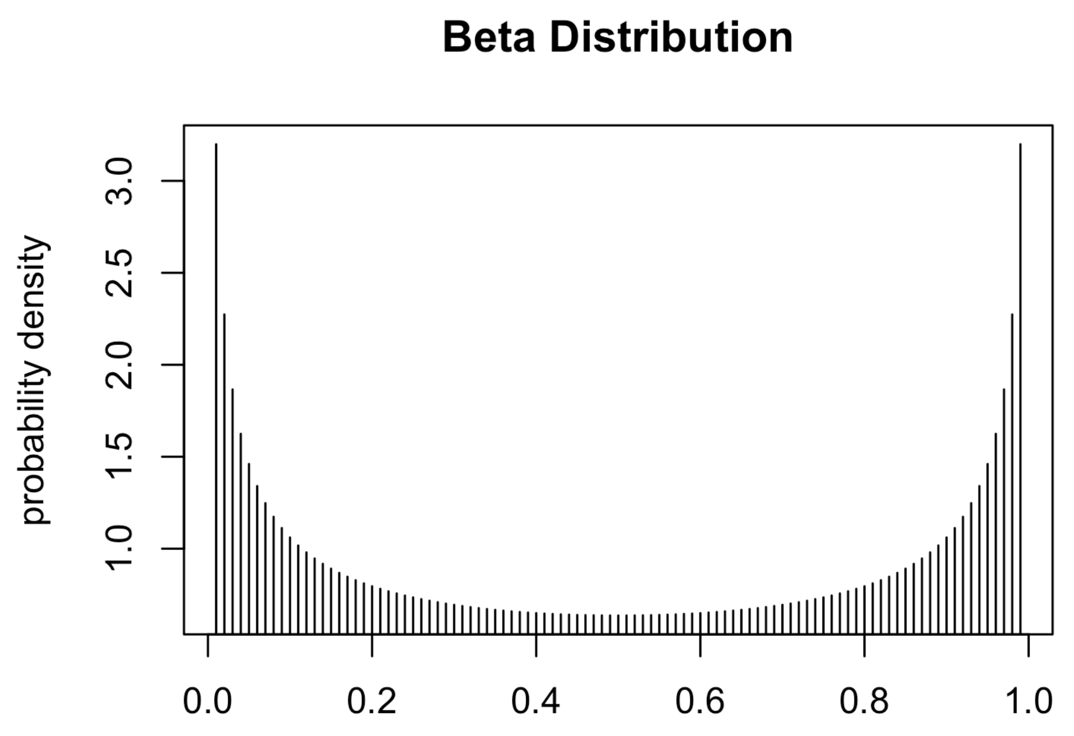


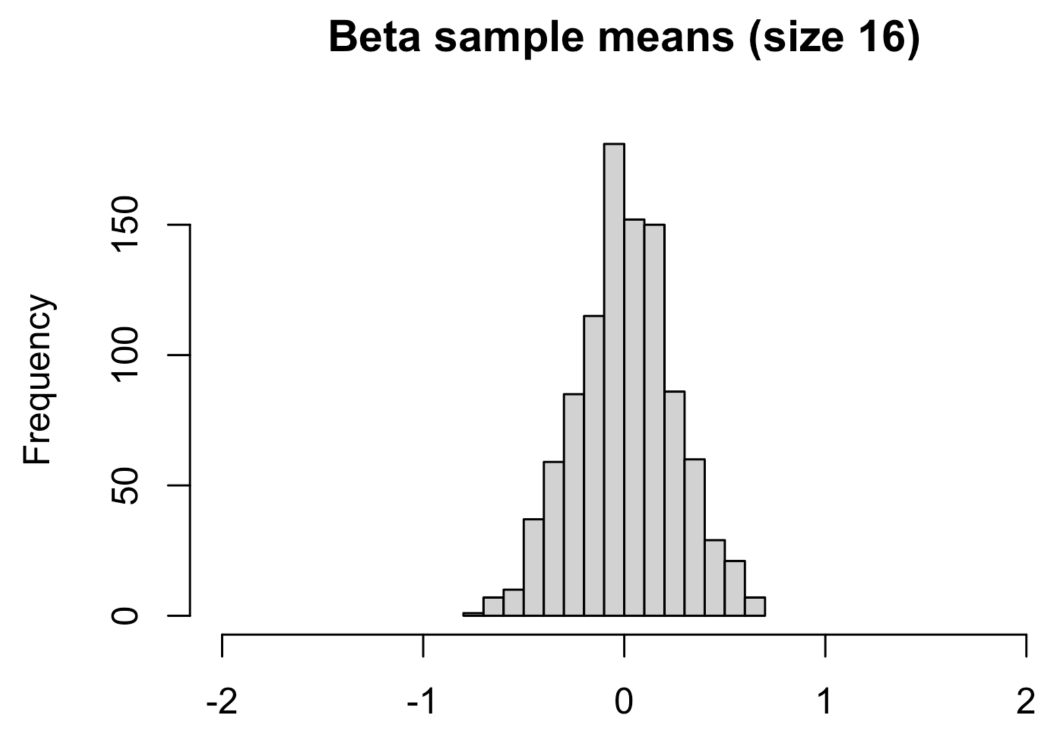
$$SEM(\bar{X}) = \frac{\sigma}{\sqrt{N}}$$

$$\bar{X} \sim \text{Normal}(\mu, SEM(\bar{X}))$$

Recall: Sampling distributions

- Remember we can model the **distribution of sampling means**
  - i.e. for **any sample** X of size N from a **certain population**, what is the **expected value** of $\overline{X}$?
  - This distribution is **(almost) always Normal**, and **centered** on the **population mean** (μ)
- Lets us calculate the **Standard Error of the Mean (SEM)**



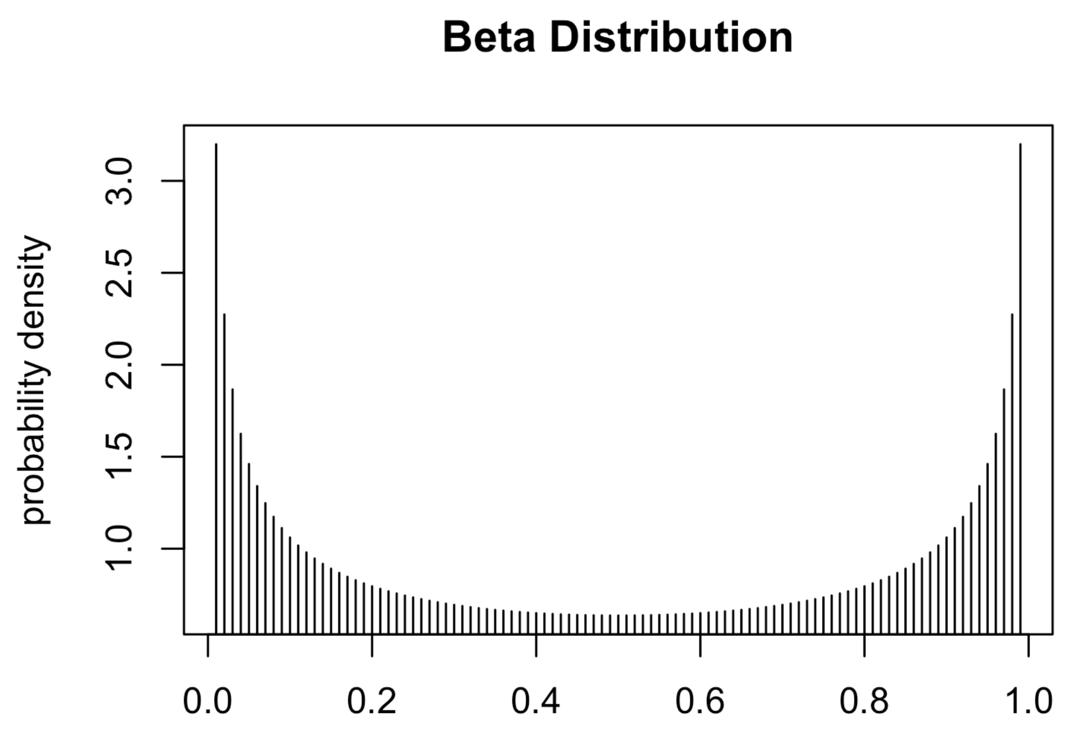


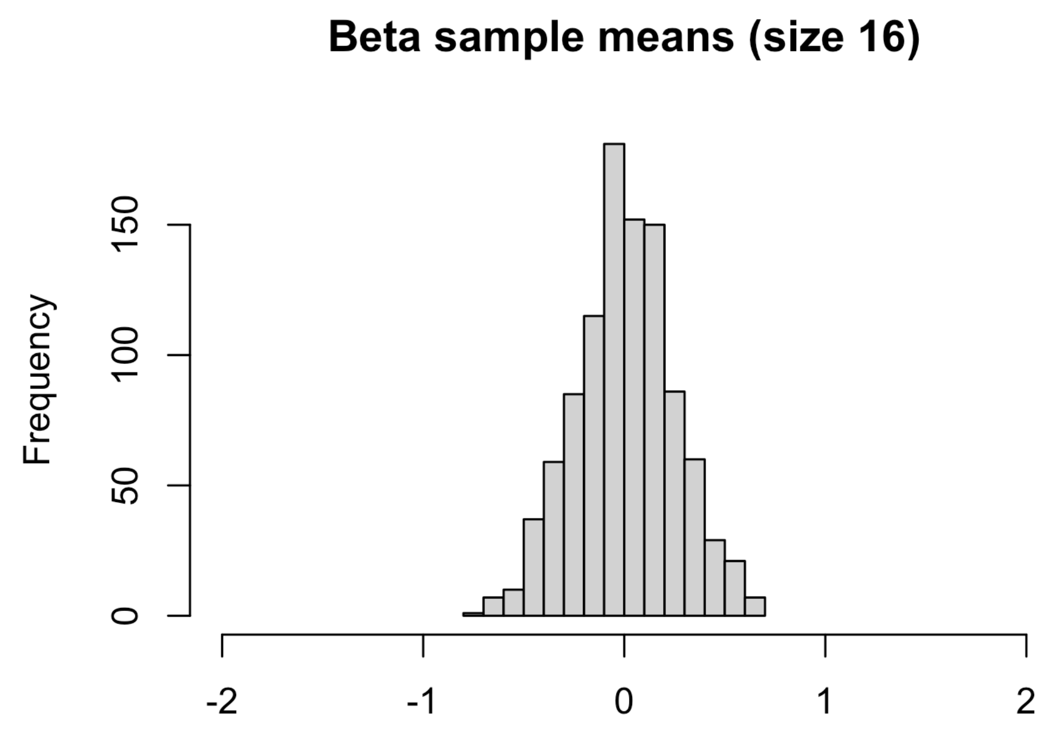
$$SEM(\overline{X}) = \frac{\sigma}{\sqrt{N}}$$

$$\overline{X} \sim \text{Normal}(\mu, SEM(\overline{X}))$$

Recall: Sampling distributions

- Remember we can model the **distribution of sampling means**
  - i.e. for **any sample** X of size N from a **certain population**, what is the **expected value** of $\overline{X}$?
  - This distribution is **(almost) always Normal**, and **centered** on the **population mean** (μ)
- Lets us calculate the **Standard Error of the Mean (SEM)**
  - How much to we expect $\overline{X}$ to differ from μ?





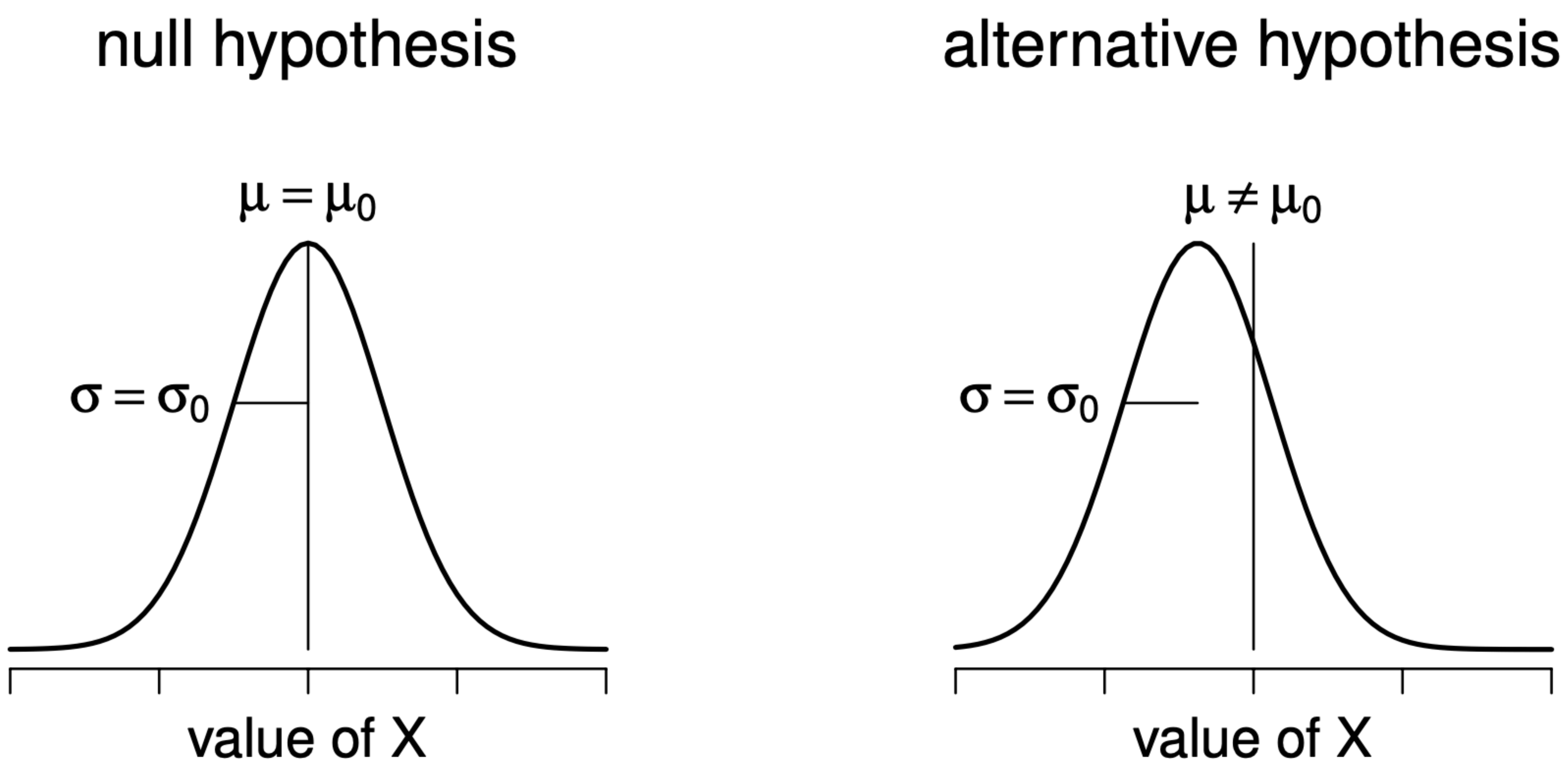
$$SEM(\overline{X}) = \frac{\sigma}{\sqrt{N}}$$

$$\overline{X} \sim \text{Normal}(\mu, SEM(\overline{X}))$$

Using the Standard Error

$$\bar{X} \sim \text{Normal}(\mu_0, SEM(\bar{X}))$$

$$\bar{X} \sim \text{Normal}(\mu_0, \frac{\sigma_0}{\sqrt{N}})$$

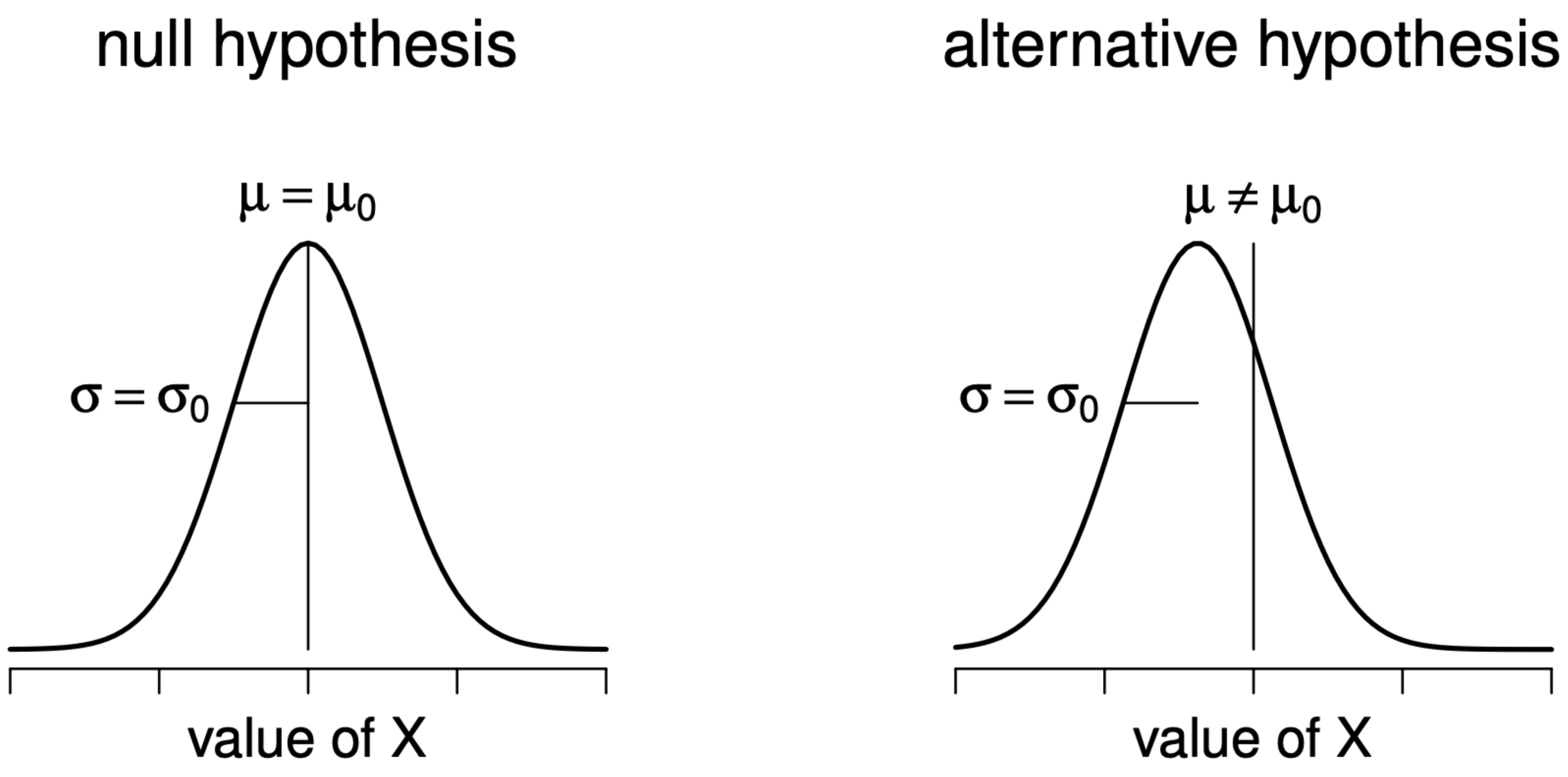


Using the Standard Error

- We can put things together to tell us **how the sample mean should be distributed assuming the Null hypothesis**

$$\overline{X} \sim \text{Normal}(\mu_0, SEM(\overline{X}))$$

$$\overline{X} \sim \text{Normal}(\mu_0, \frac{\sigma_0}{\sqrt{N}})$$

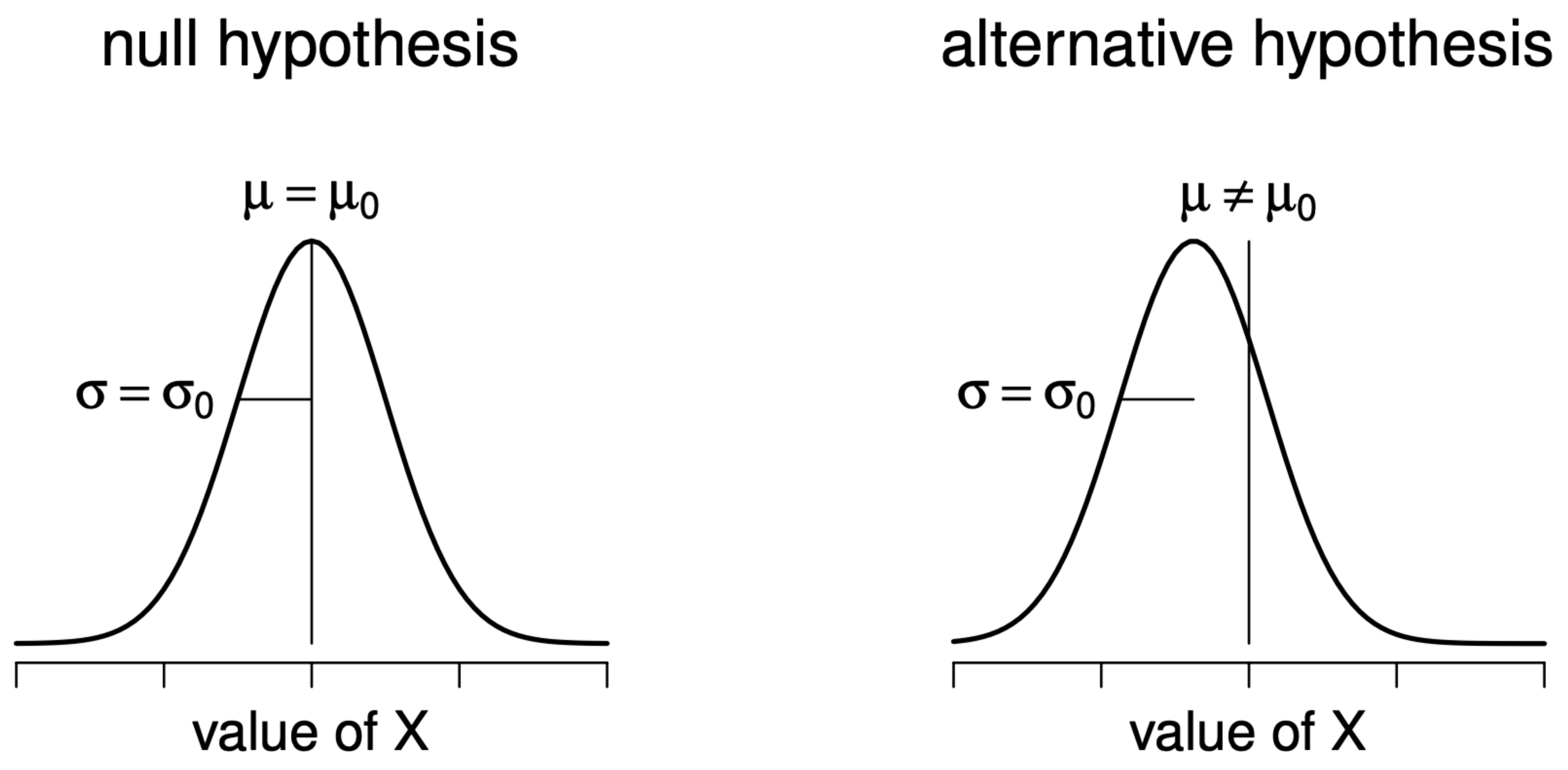


Using the Standard Error

- We can put things together to tell us **how the sample mean should be distributed assuming the Null hypothesis**
  - μ_0 : the **population mean** under the Null (average grade of all students)

$$\bar{X} \sim \text{Normal}(\mu_0, SEM(\bar{X}))$$

$$\bar{X} \sim \text{Normal}(\mu_0, \frac{\sigma_0}{\sqrt{N}})$$

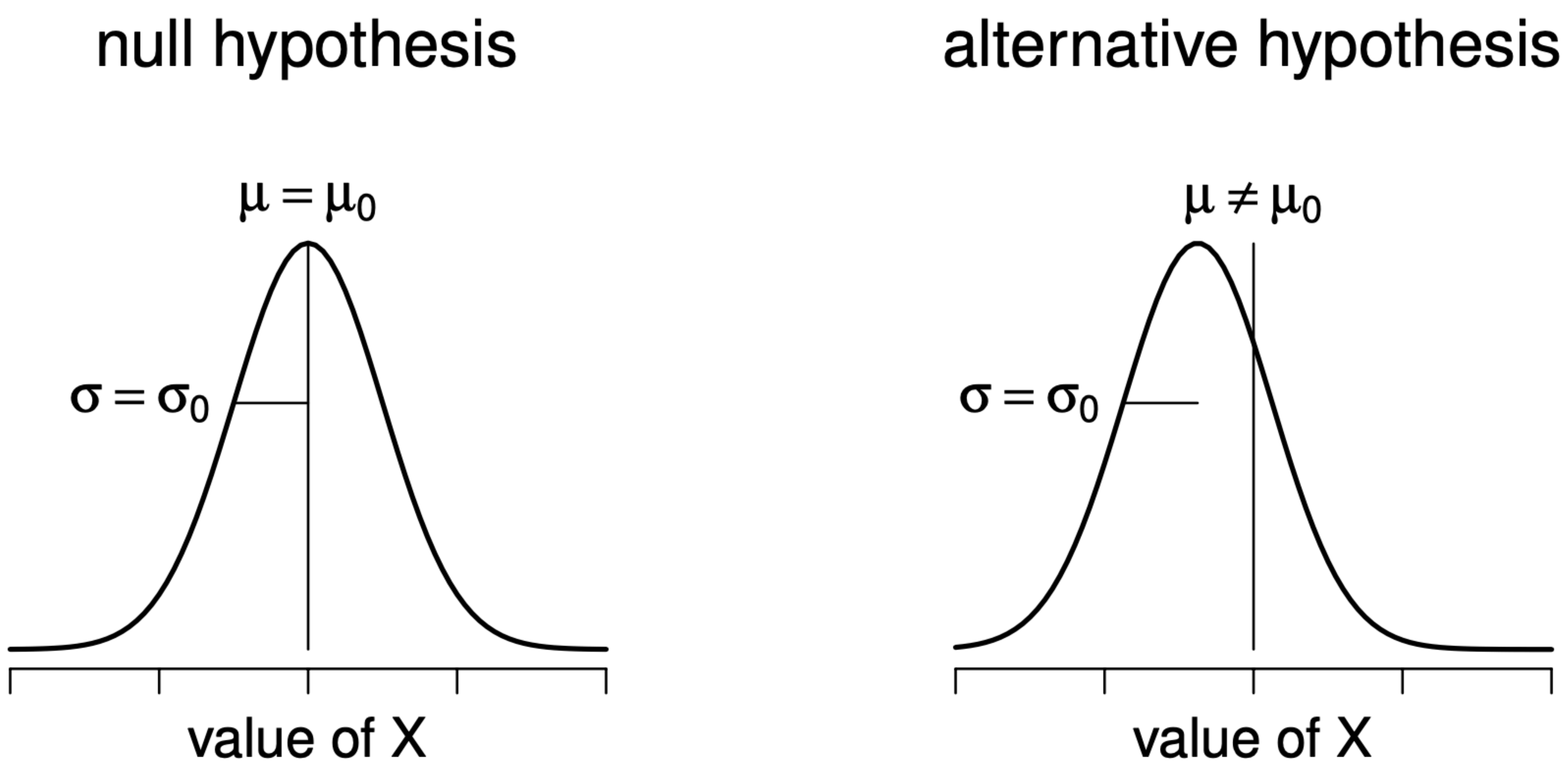


Using the Standard Error

- We can put things together to tell us **how the sample mean should be distributed assuming the Null hypothesis**
  - μ_0 : the **population mean** under the Null (average grade of all students)
  - $SEM(\overline{X})$: the **Standard Error of the Mean** (calculated based on σ_0 and N)

$$\overline{X} \sim \text{Normal}(\mu_0, SEM(\overline{X}))$$

$$\overline{X} \sim \text{Normal}(\mu_0, \frac{\sigma_0}{\sqrt{N}})$$

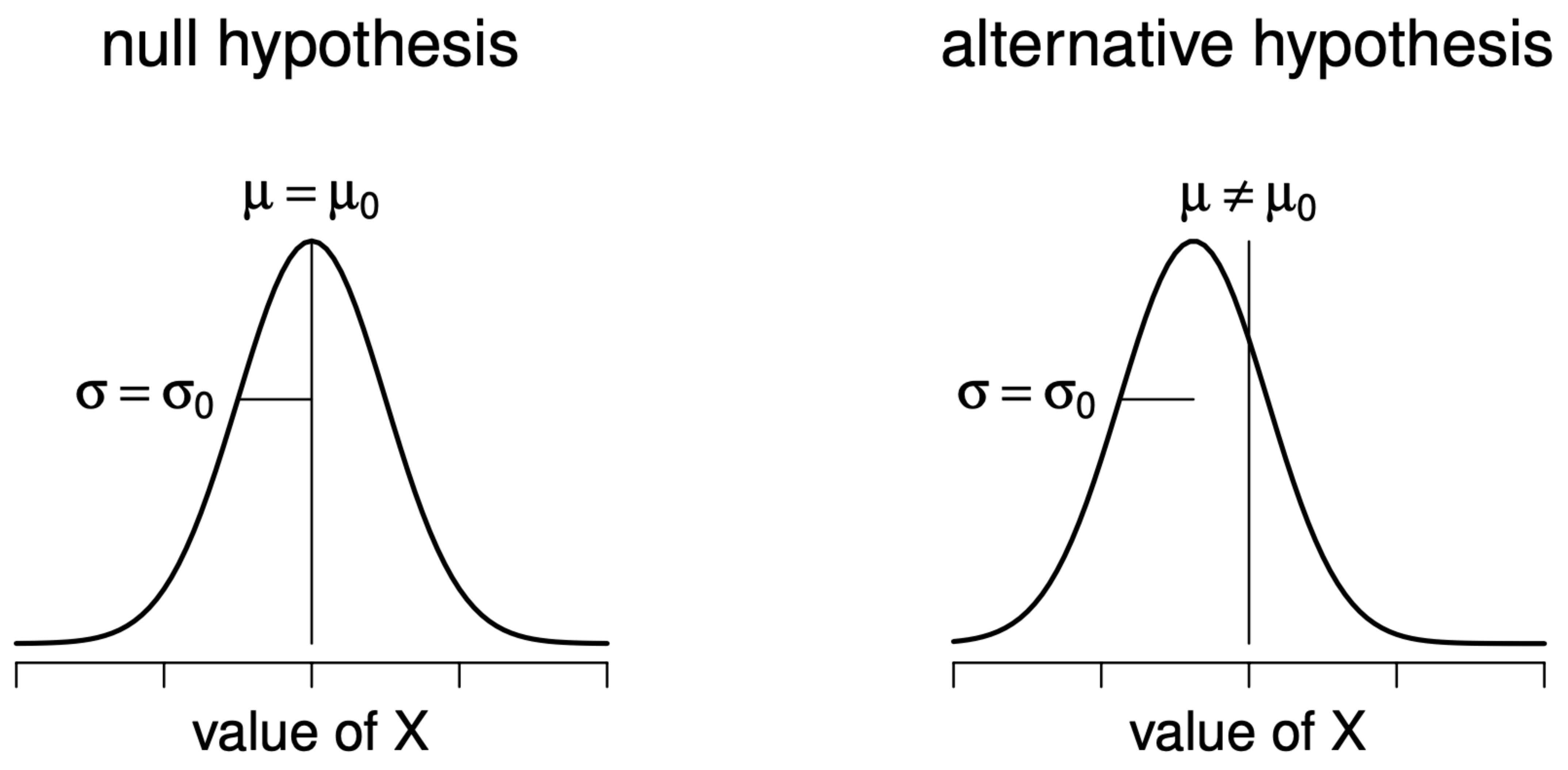


Using the Standard Error

- We can put things together to tell us **how the sample mean should be distributed assuming the Null hypothesis**
  - μ_0 : the **population mean** under the Null (average grade of all students)
  - $SEM(\overline{X})$: the **Standard Error of the Mean** (calculated based on σ_0 and N)
  - σ_0 : the **population standard deviation** under the Null

$$\overline{X} \sim \text{Normal}(\mu_0, SEM(\overline{X}))$$

$$\overline{X} \sim \text{Normal}(\mu_0, \frac{\sigma_0}{\sqrt{N}})$$

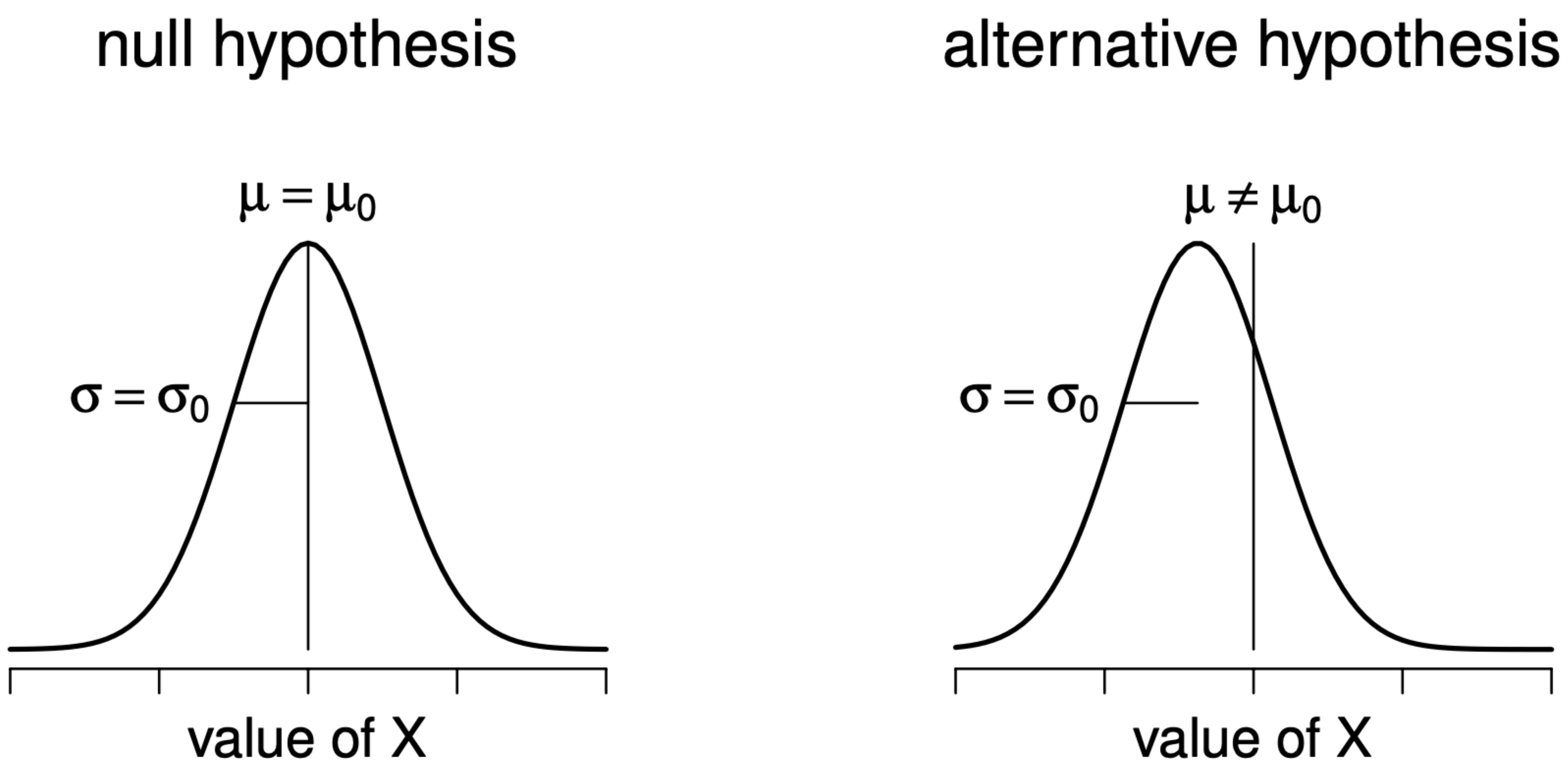


Using the Standard Error

- We can put things together to tell us **how the sample mean should be distributed assuming the Null hypothesis**
  - μ_0 : the **population mean** under the Null (average grade of all students)
  - $SEM(\overline{X})$: the **Standard Error of the Mean** (calculated based on σ_0 and N)
  - σ_0 : the **population standard deviation** under the Null
  - N : the **sample size** (number of Ling students taking the class)

$$\overline{X} \sim \text{Normal}(\mu_0, SEM(\overline{X}))$$

$$\overline{X} \sim \text{Normal}(\mu_0, \frac{\sigma_0}{\sqrt{N}})$$

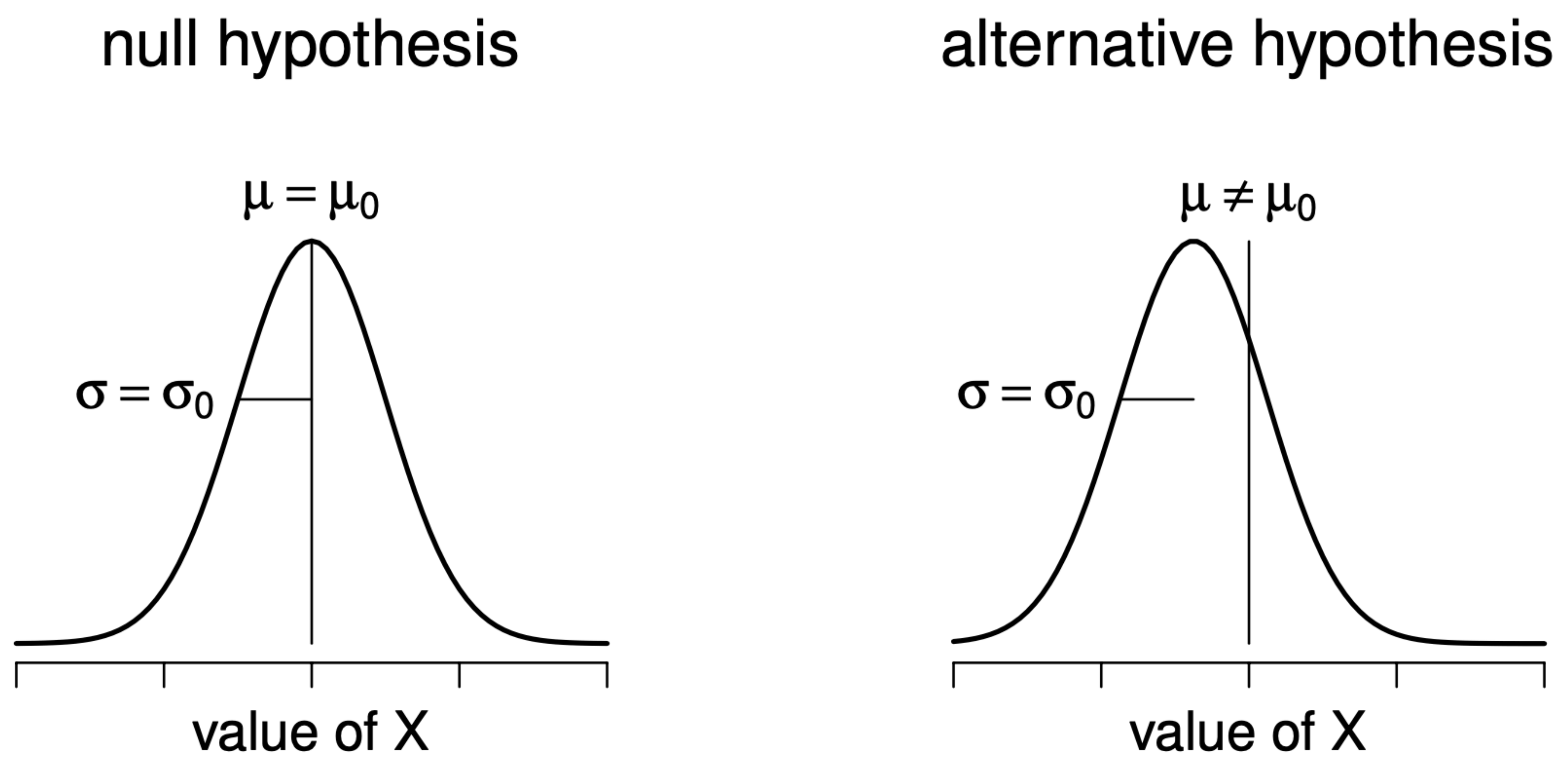


Using the Standard Error

- We can put things together to tell us **how the sample mean should be distributed assuming the Null hypothesis**
  - μ_0 : the **population mean** under the Null (average grade of all students)
  - $SEM(\overline{X})$: the **Standard Error of the Mean** (calculated based on σ_0 and N)
  - σ_0 : the **population standard deviation** under the Null
  - N : the **sample size** (number of Ling students taking the class)
- This tells us **which values of $\overline{X}$ are more probable**, assuming the Null!

$$\overline{X} \sim \text{Normal}(\mu_0, SEM(\overline{X}))$$

$$\overline{X} \sim \text{Normal}(\mu_0, \frac{\sigma_0}{\sqrt{N}})$$



Using the z-score

$$z_{\overline{X}} = \frac{\overline{X} - \mu_0}{SEM(\overline{X})}$$

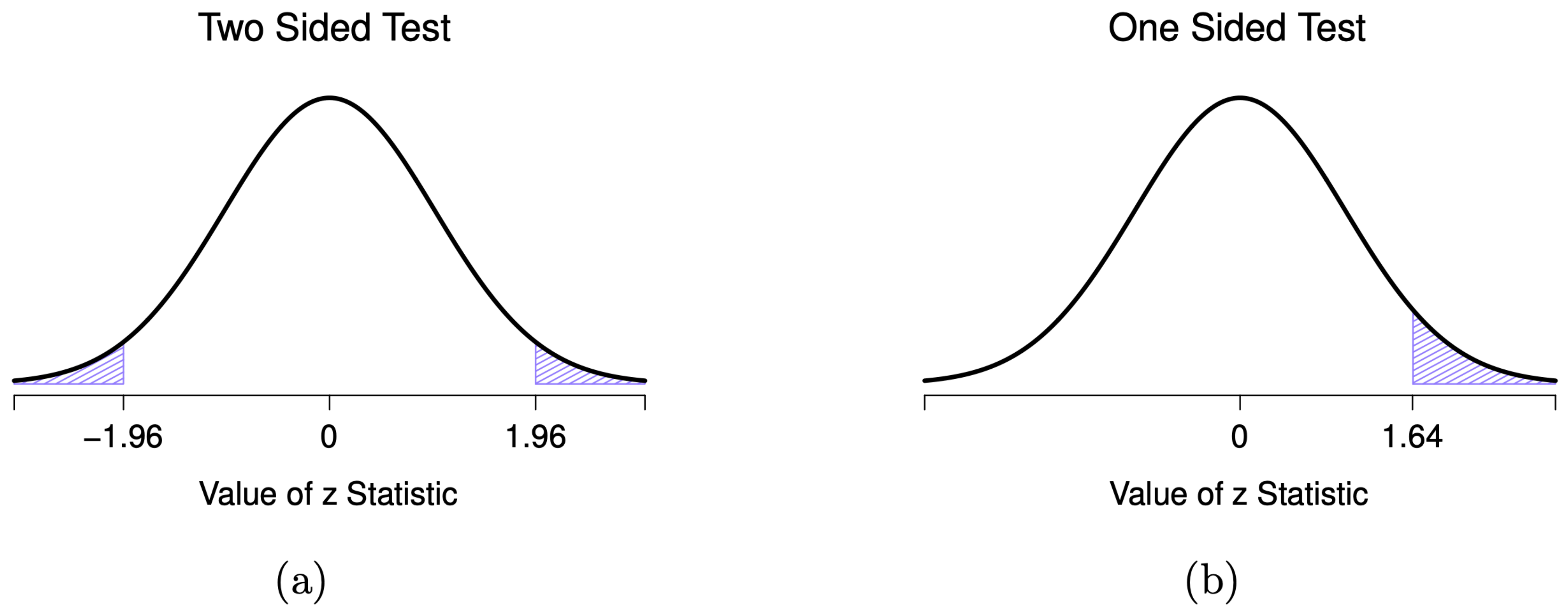


Figure 13.3: Rejection regions for the two-sided z-test (panel a) and the one-sided z-test (panel b).

Using the z-score

- The z-score is **how many standard deviations away** a certain value is from the mean, in a **Normal distribution**

$$z_{\overline{X}} = \frac{\overline{X} - \mu_0}{SEM(\overline{X})}$$

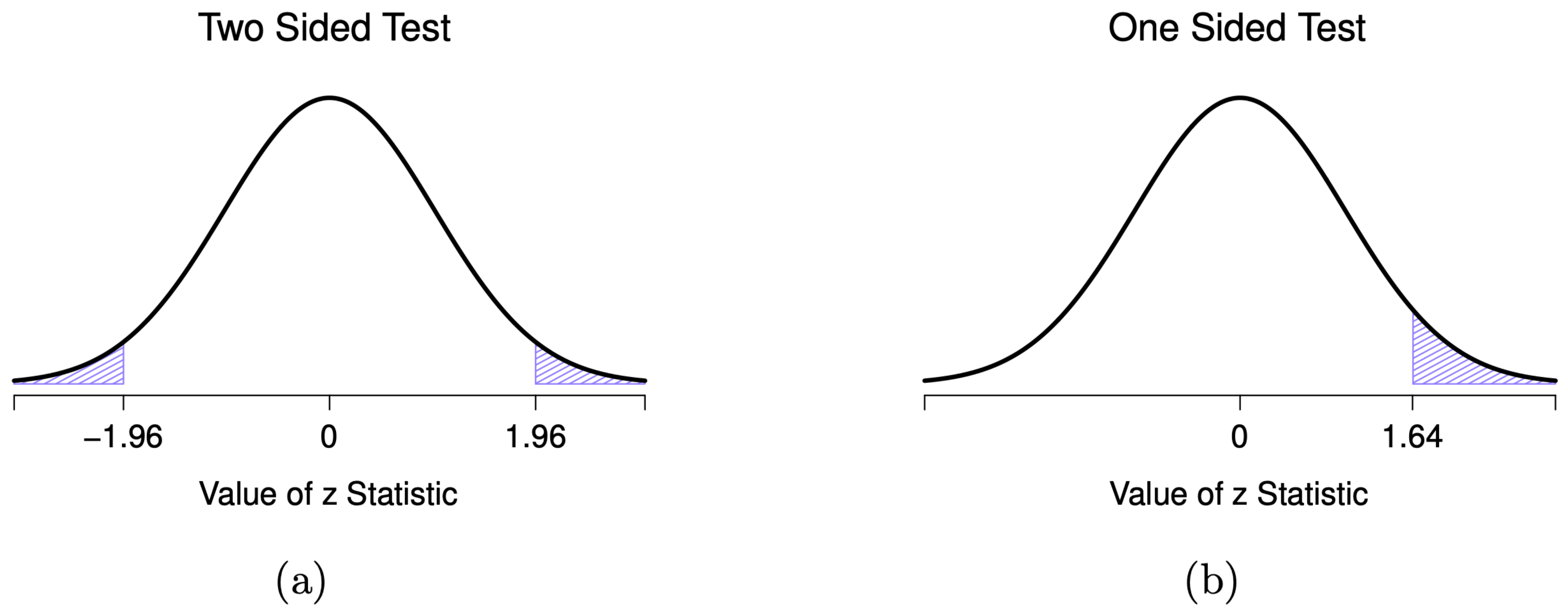


Figure 13.3: Rejection regions for the two-sided z-test (panel a) and the one-sided z-test (panel b).

Using the z-score

- The z-score is **how many standard deviations away** a certain value is from the mean, in a **Normal distribution**
  - The **area under the curve** is easily calculable from z-scores!

$$z_{\overline{X}} = \frac{\overline{X} - \mu_0}{SEM(\overline{X})}$$

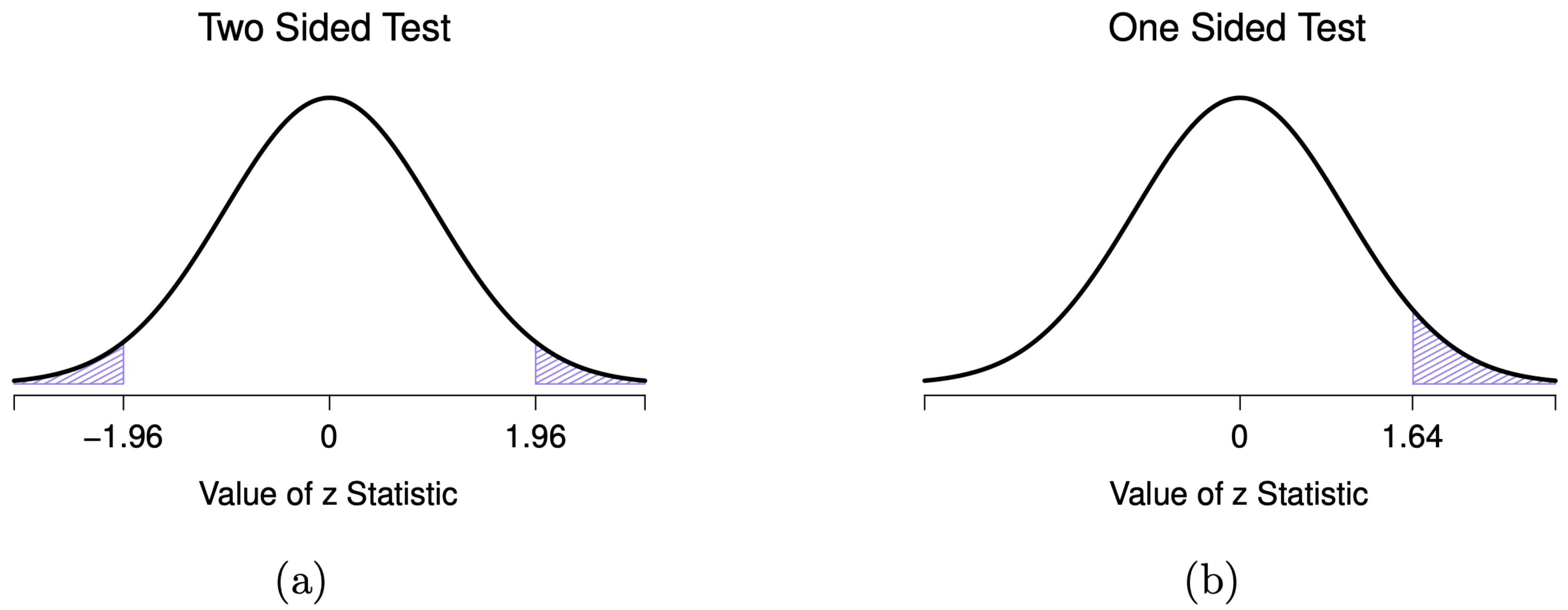


Figure 13.3: Rejection regions for the two-sided z-test (panel a) and the one-sided z-test (panel b).

Using the z-score

- The z-score is **how many standard deviations away** a certain value is from the mean, in a **Normal distribution**
  - The **area under the curve** is easily calculable from z-scores!
- This lets us define a **critical region** for a significant result!

$$z_{\overline{X}} = \frac{\overline{X} - \mu_0}{SEM(\overline{X})}$$

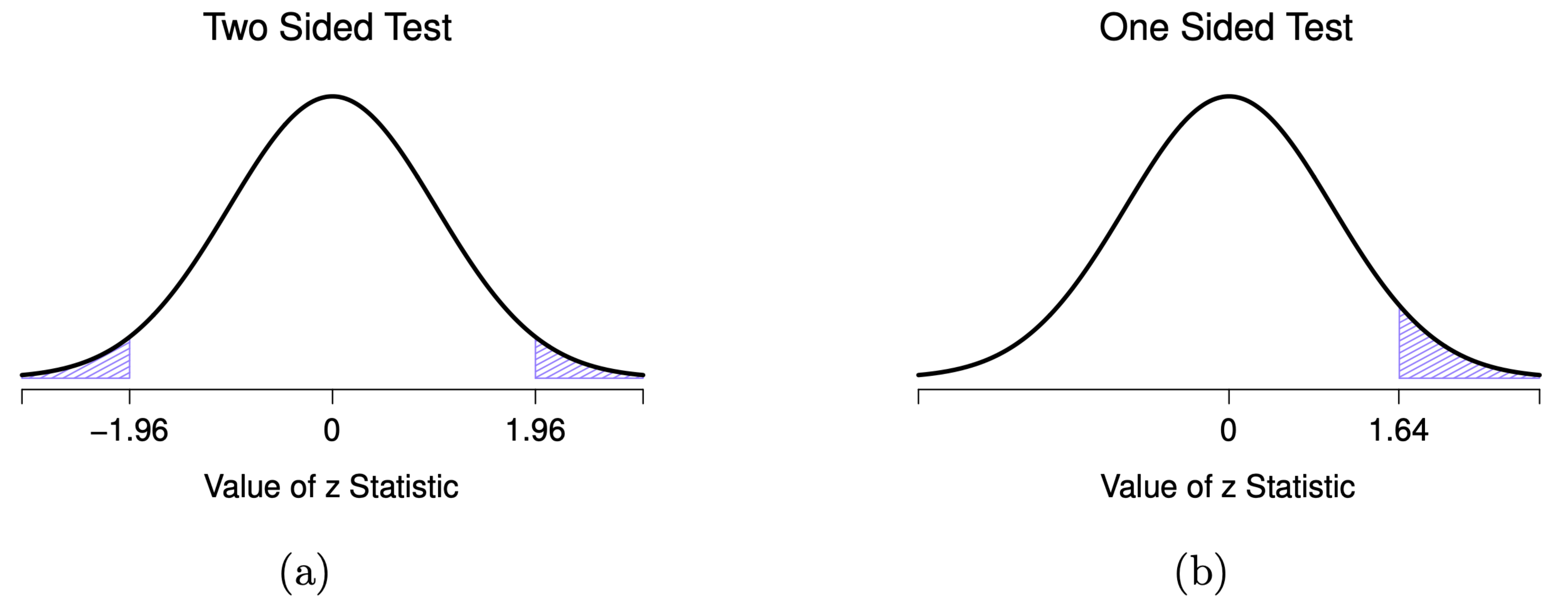


Figure 13.3: Rejection regions for the two-sided z-test (panel a) and the one-sided z-test (panel b).

Using the z-score

- The z-score is **how many standard deviations away** a certain value is from the mean, in a **Normal distribution**
  - The **area under the curve** is easily calculable from z-scores!
- This lets us define a **critical region** for a significant result!
  - i.e. if $\overline{X}$ is a **certain z-score away** from μ_0, we can say $\mu \neq \mu_0$ with **statistical significance!**

$$z_{\overline{X}} = \frac{\overline{X} - \mu_0}{SEM(\overline{X})}$$

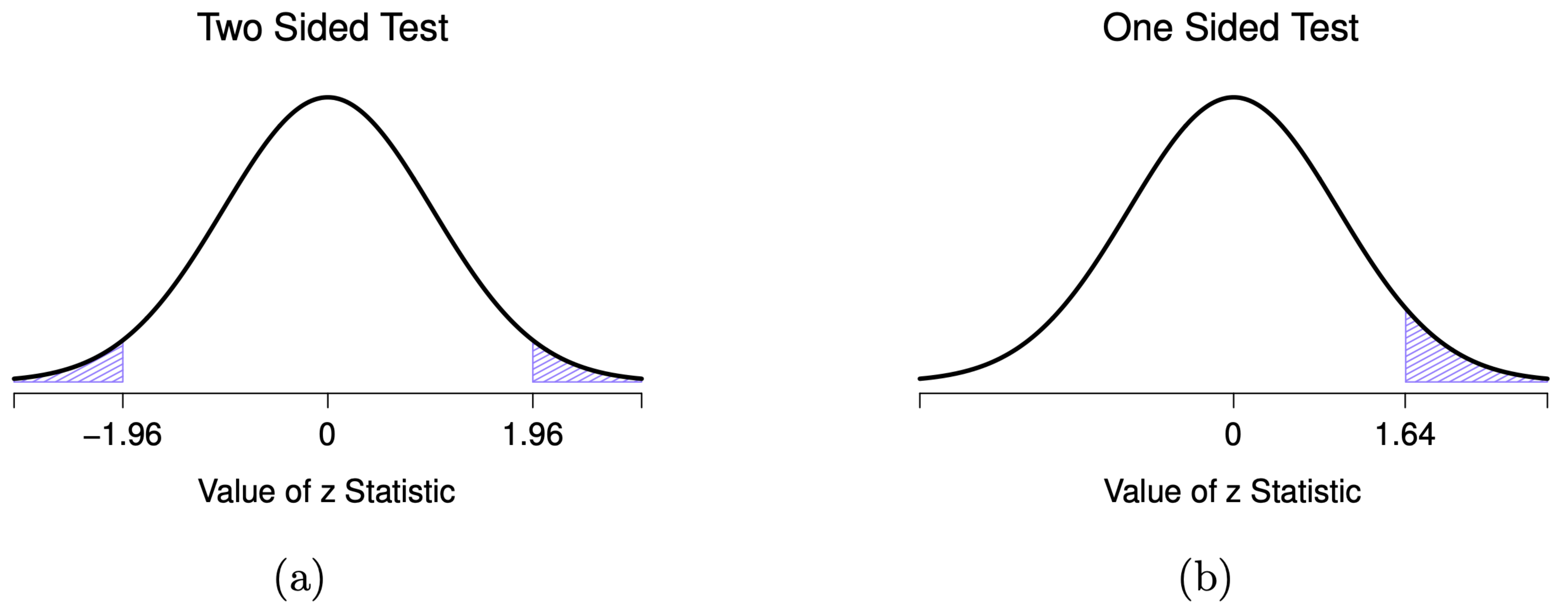


Figure 13.3: Rejection regions for the two-sided z-test (panel a) and the one-sided z-test (panel b).

Calculating the p-value

```
> null_mean = 67.5
> null_stdev = 9.5
> N = 20
> SEM = null_stdev / sqrt(N)
> SEM
[1] 2.124265
>
> ling_mean = 72.3
> z_score = (ling_mean - null_mean) / SEM
> z_score
[1] 2.259606
>
> upper_area = pnorm(z_score, lower.tail=FALSE)
> upper_area
[1] 0.01192287
> # this is our p-value!
```

Calculating the p-value

- These are all the ingredients we need to **calculate the significance**

```
> null_mean = 67.5
> null_stdev = 9.5
> N = 20
> SEM = null_stdev / sqrt(N)
> SEM
[1] 2.124265
>
> ling_mean = 72.3
> z_score = (ling_mean - null_mean) / SEM
> z_score
[1] 2.259606
>
> upper_area = pnorm(z_score, lower.tail=FALSE)
> upper_area
[1] 0.01192287
> # this is our p-value!
```

Calculating the p-value

- These are all the ingredients we need to **calculate the significance**
- Let's plug in some values
  - **Sample size** (number of Ling students): $N = 20$
  - **Sample mean** (their average grade): $\overline{X} = 72.3$

```
> null_mean = 67.5
> null_stdev = 9.5
> N = 20
> SEM = null_stdev / sqrt(N)
> SEM
[1] 2.124265
>
> ling_mean = 72.3
> z_score = (ling_mean - null_mean) / SEM
> z_score
[1] 2.259606
>
> upper_area = pnorm(z_score, lower.tail=FALSE)
> upper_area
[1] 0.01192287
> # this is our p-value!
```

Calculating the p-value

- These are all the ingredients we need to **calculate the significance**
- Let's plug in some values
  - **Sample size** (number of Ling students): $N = 20$
  - **Sample mean** (their average grade): $\overline{X} = 72.3$
- Calculation on the right $\rightarrow$

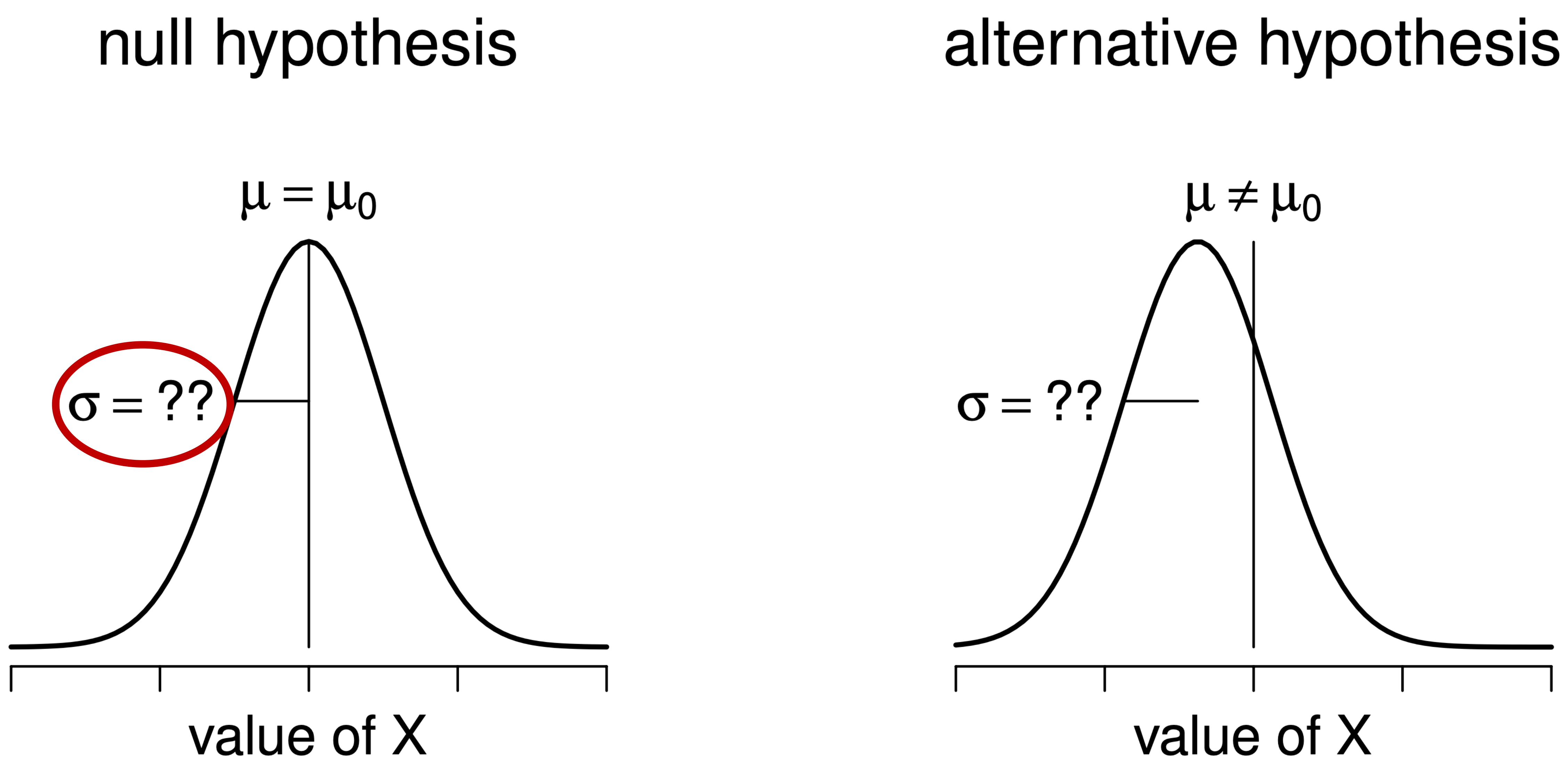
```
> null_mean = 67.5
> null_stdev = 9.5
> N = 20
> SEM = null_stdev / sqrt(N)
> SEM
[1] 2.124265
>
> ling_mean = 72.3
> z_score = (ling_mean - null_mean) / SEM
> z_score
[1] 2.259606
>
> upper_area = pnorm(z_score, lower.tail=FALSE)
> upper_area
[1] 0.01192287
> # this is our p-value!
```

Calculating the p-value

- These are all the ingredients we need to **calculate the significance**
- Let's plug in some values
  - **Sample size** (number of Ling students): $N = 20$
  - **Sample mean** (their average grade): $\overline{X} = 72.3$
- Calculation on the right →
- p-value is **0.012**
  - We'd have a **1.2% chance** of seeing a sample mean as high as that if the Null were true
  - (i.e. a fairly significant result)

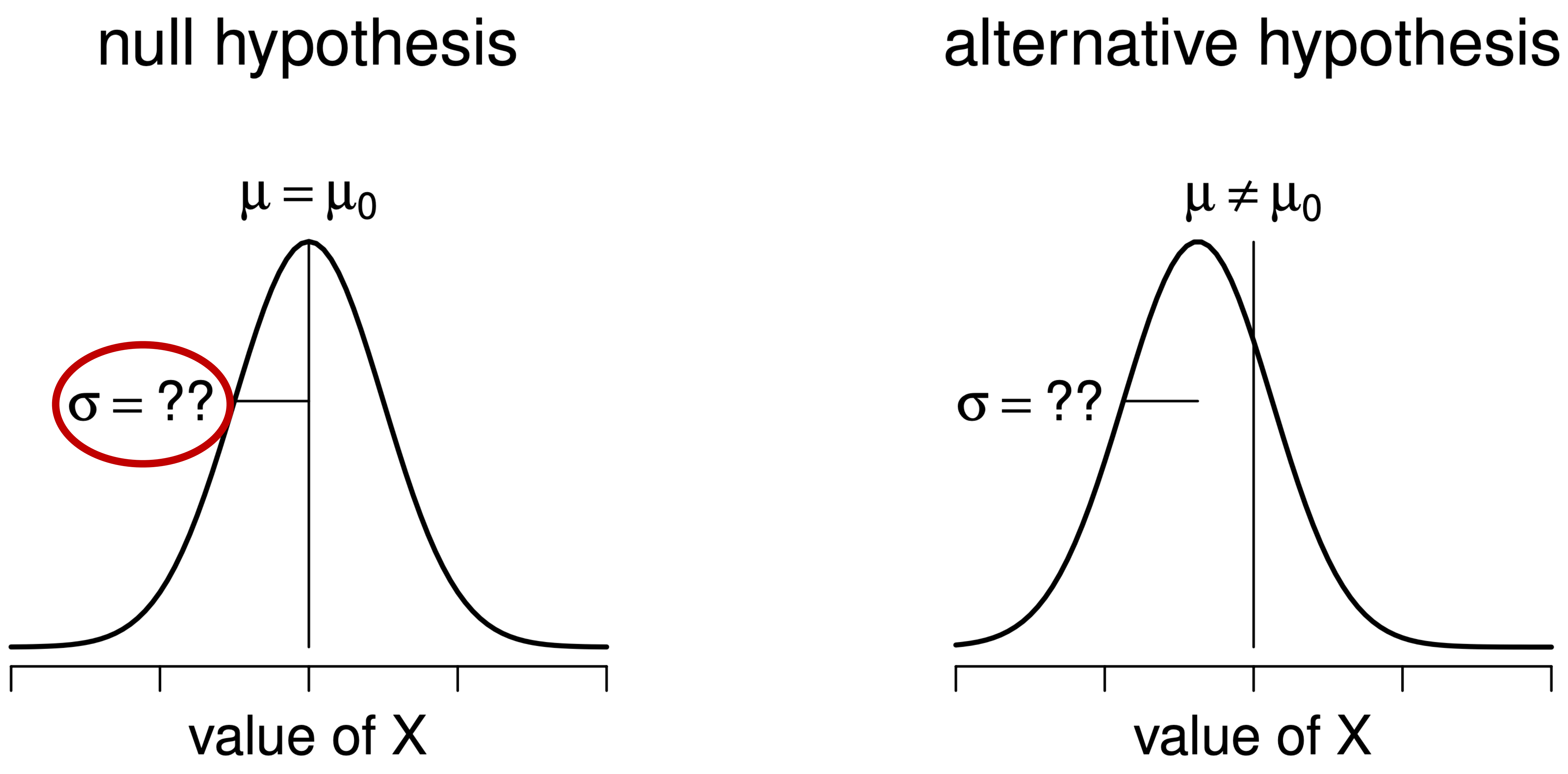
```
> null_mean = 67.5
> null_stdev = 9.5
> N = 20
> SEM = null_stdev / sqrt(N)
> SEM
[1] 2.124265
>
> ling_mean = 72.3
> z_score = (ling_mean - null_mean) / SEM
> z_score
[1] 2.259606
>
> upper_area = pnorm(z_score, lower.tail=FALSE)
> upper_area
[1] 0.01192287
> # this is our p-value!
```

Why we need a t-test



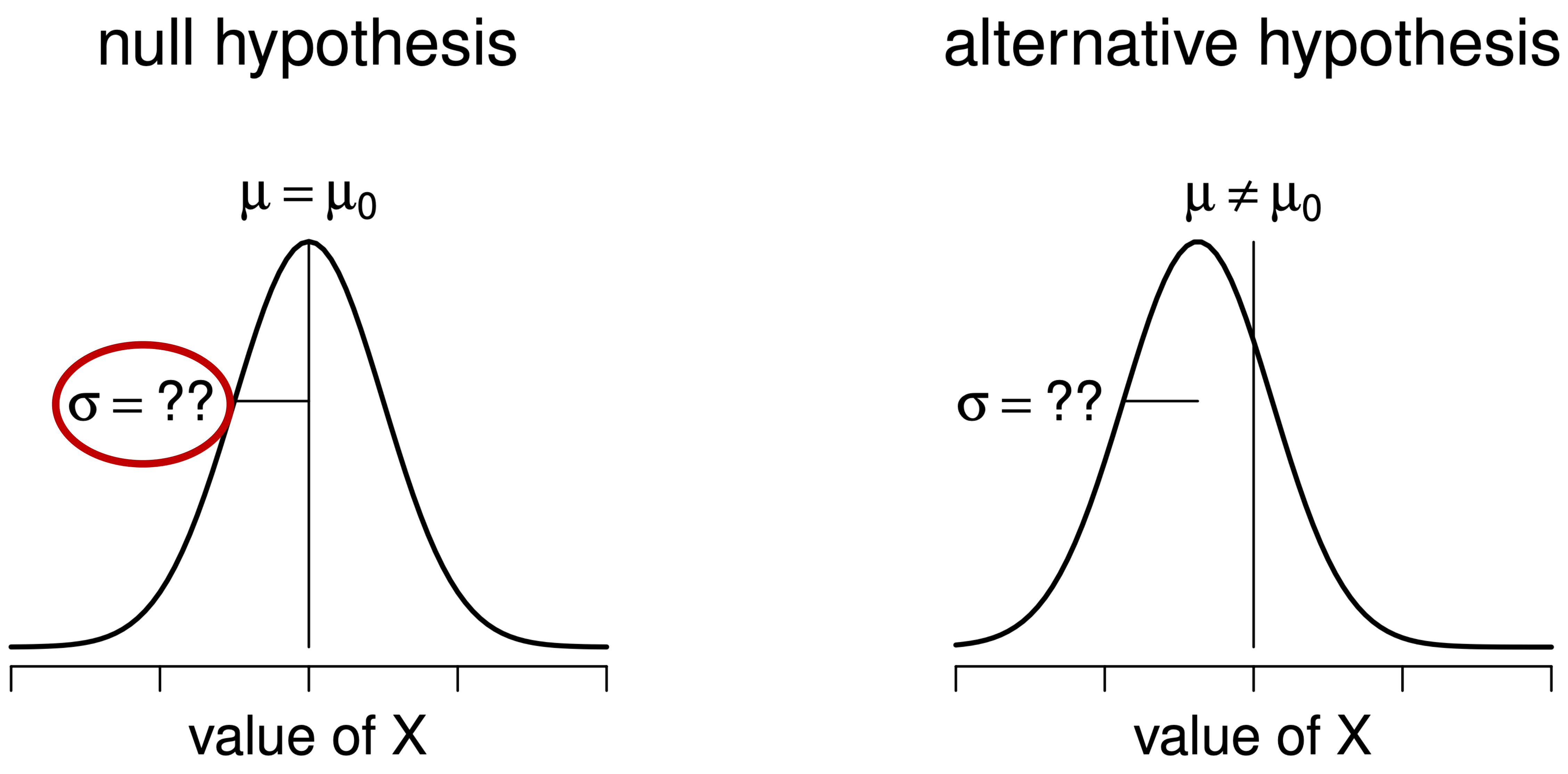
Why we need a t-test

- Remember we're **assuming weird things** about the standard deviation



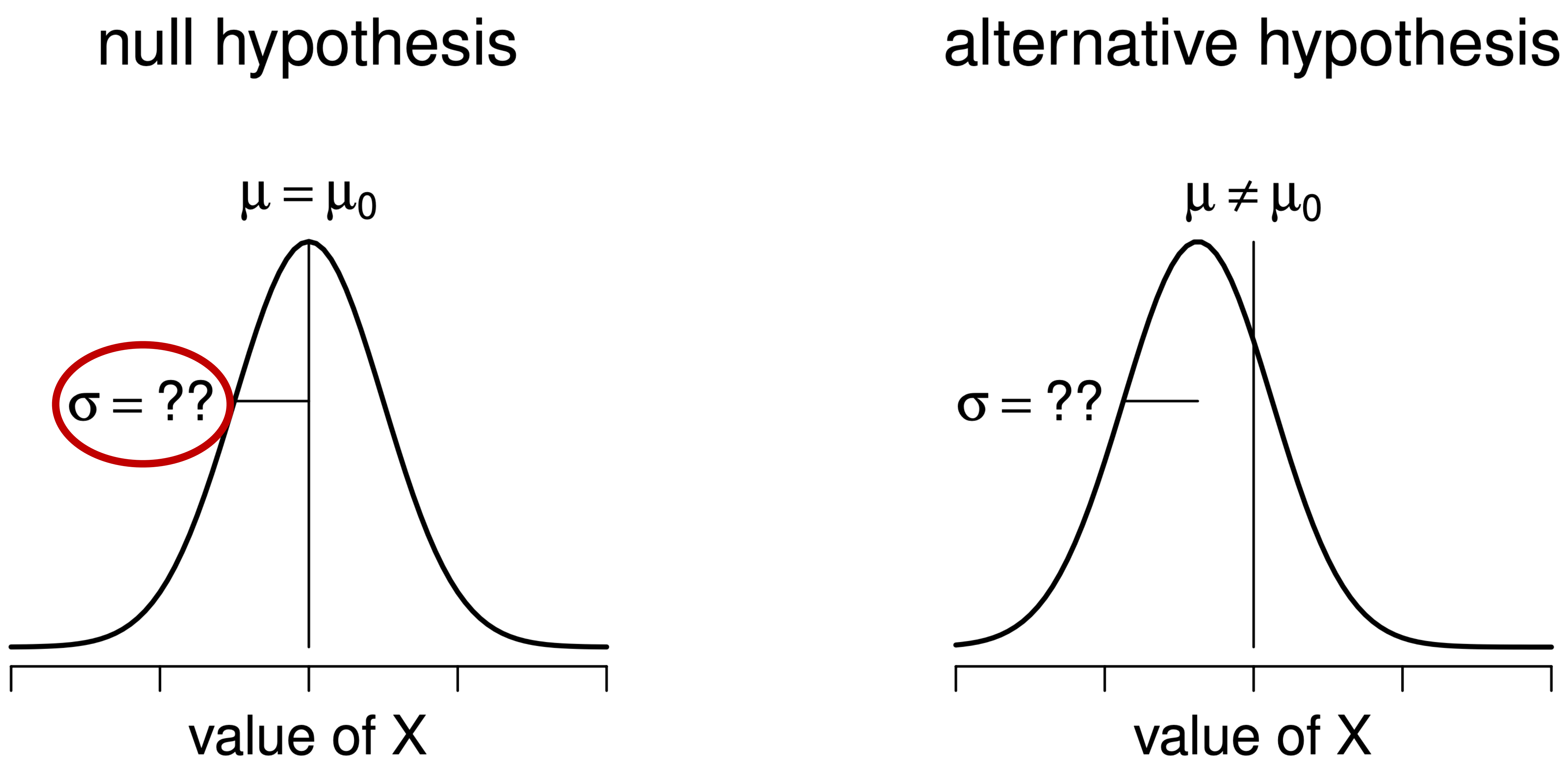
Why we need a t-test

- Remember we're **assuming weird things** about the standard deviation
  - We assume we **know the overall population stdev** (σ_0, the standard deviation of all students who take the class)



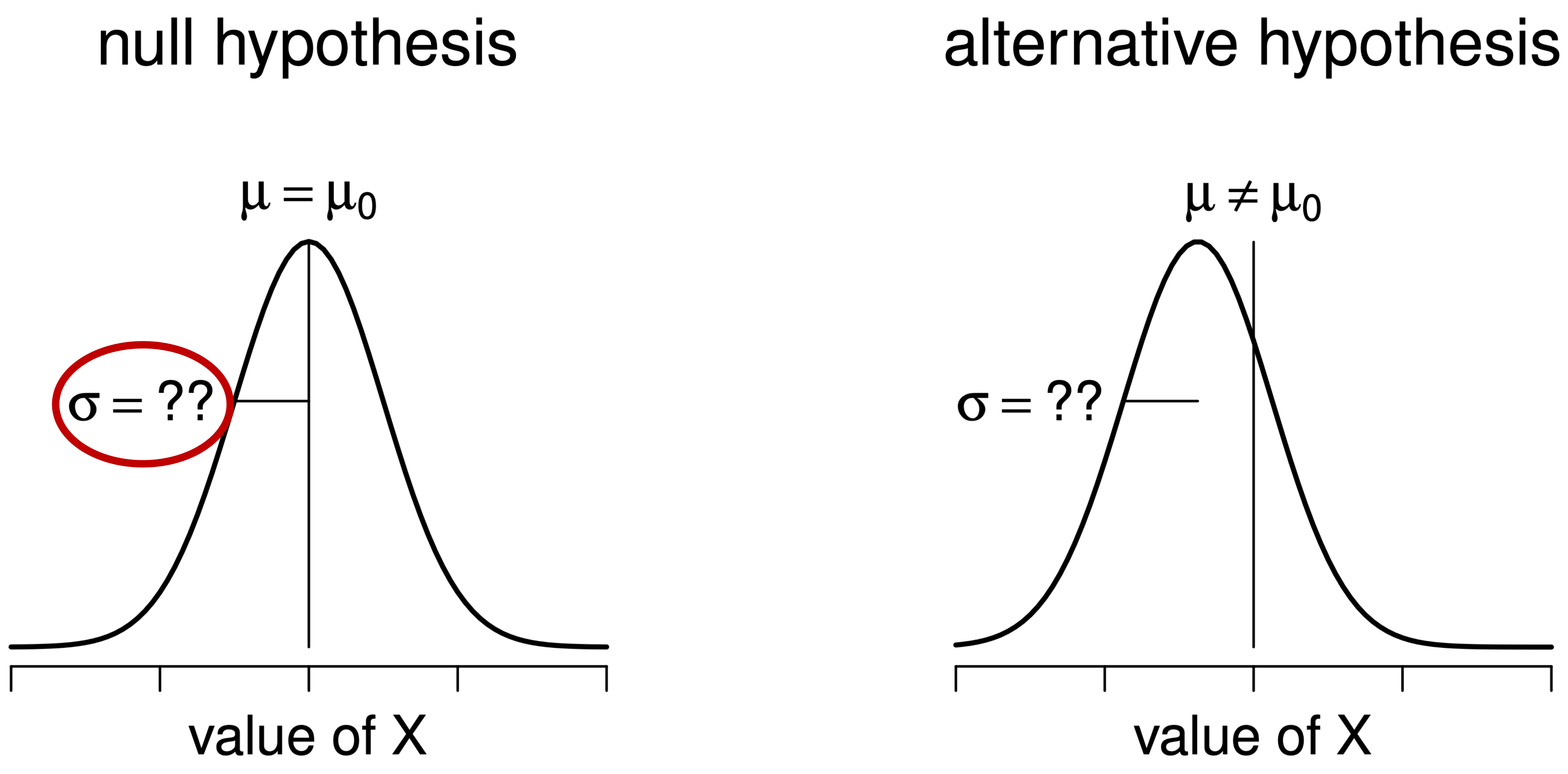
Why we need a t-test

- Remember we're **assuming weird things** about the standard deviation
  - We assume we **know the overall population stdev** (σ_0, the standard deviation of all students who take the class)
  - We assume that the Ling student stdev is **the same** as the overall stdev ($\sigma = \sigma_0$)



Why we need a t-test

- Remember we're **assuming weird things** about the standard deviation
  - We assume we **know the overall population stdev** (σ_0, the standard deviation of all students who take the class)
  - We assume that the Ling student stdev is **the same** as the overall stdev ($\sigma = \sigma_0$)
- The t-test **eliminates these assumptions**



t-test changes

t-test changes

- The t-test is **almost the same** as the z-test (that's why we did it first)

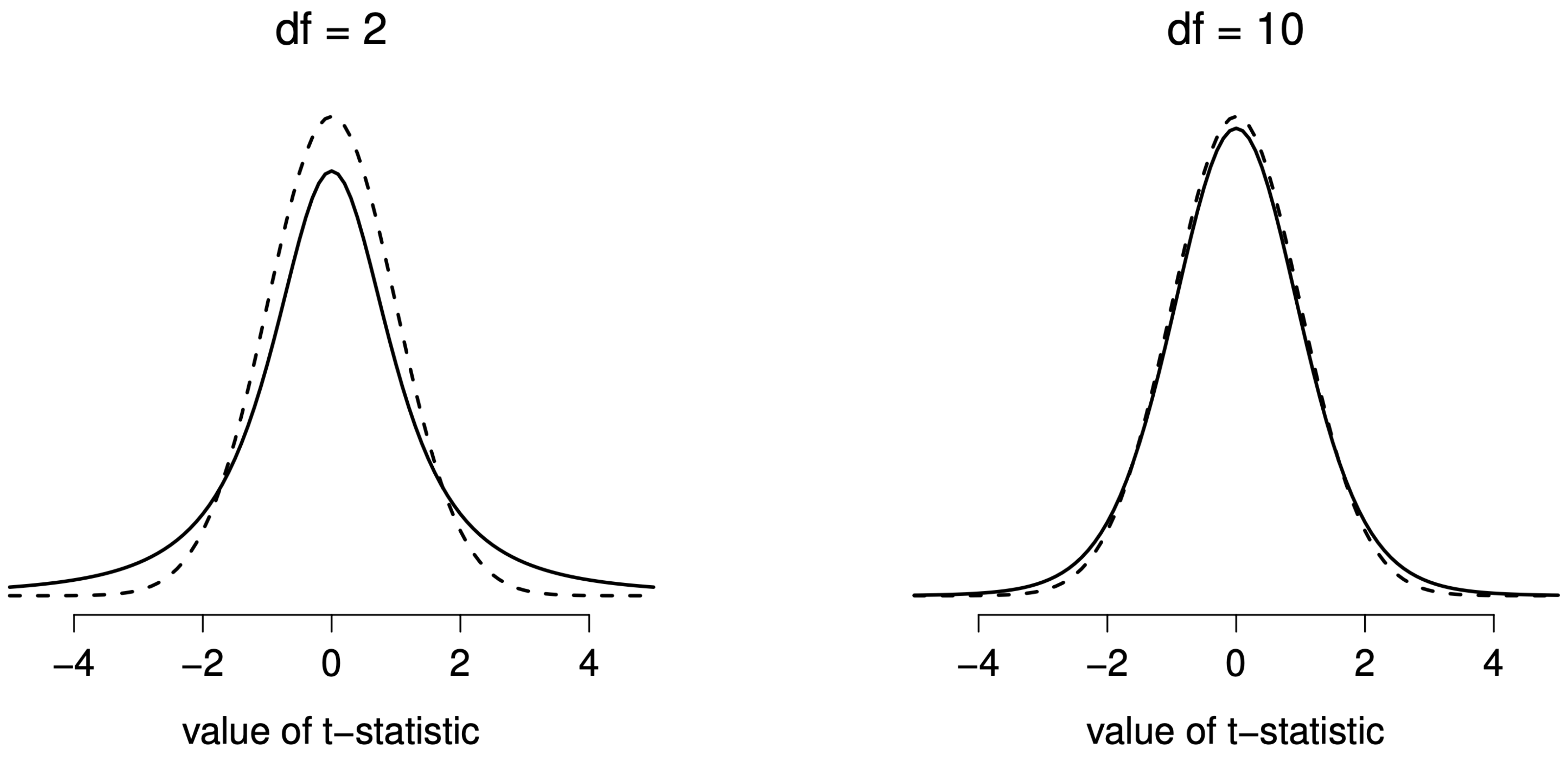
t-test changes

- The t-test is **almost the same** as the z-test (that's why we did it first)
- Change 1:
  - We **don't have the true population stdev** (σ)
  - We instead **estimate it from our sample** ($\hat{\sigma} = \sigma(X)$)

t-test changes

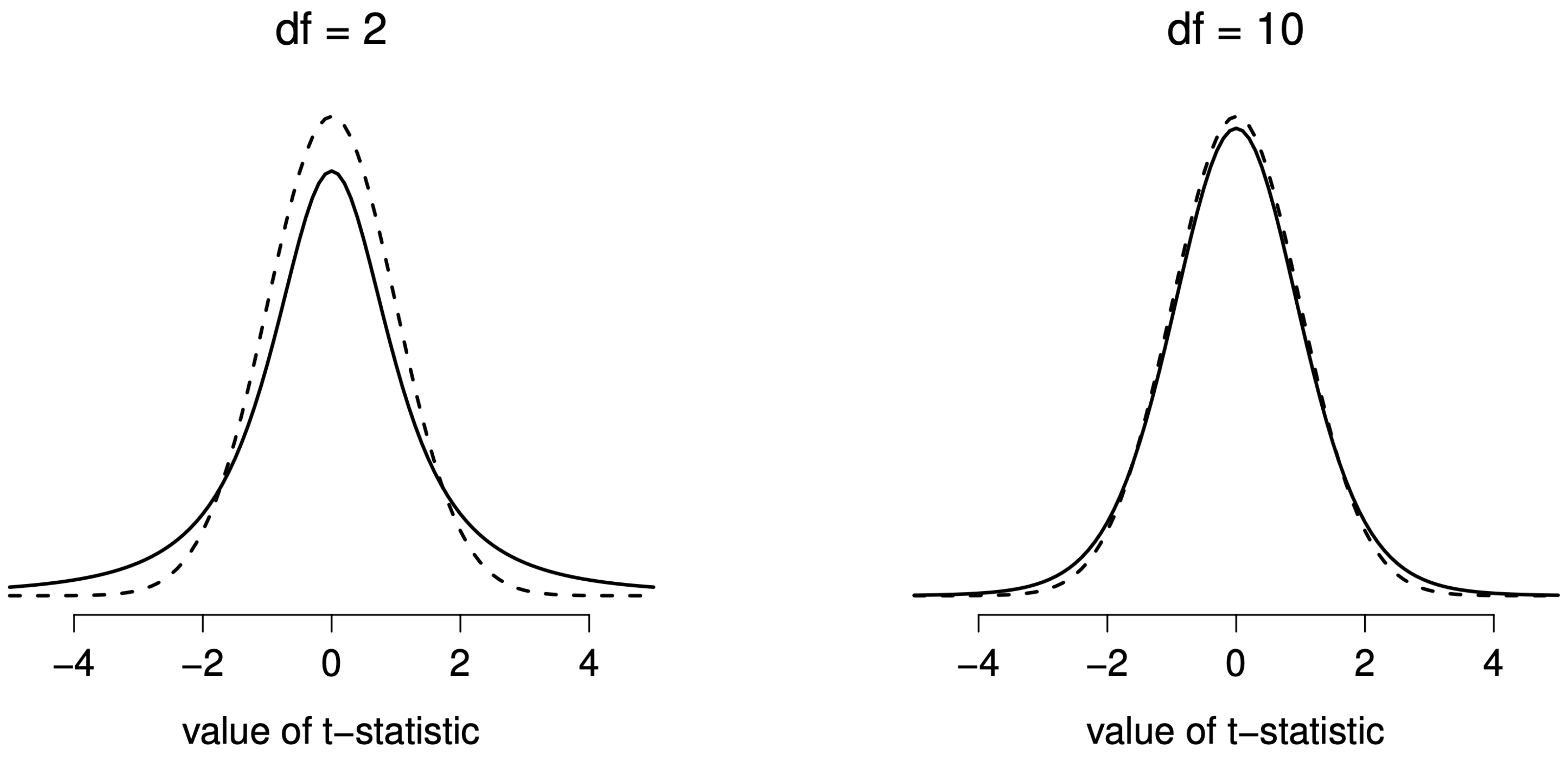
- The t-test is **almost the same** as the z-test (that's why we did it first)
- Change 1:
  - We **don't have the true population stdev** (σ)
  - We instead **estimate it from our sample** ($\hat{\sigma} = \sigma(X)$)
- Change 2:
  - We use a **t-distribution** instead of a Normal distribution
  - What's a t-distribution?

t-distribution



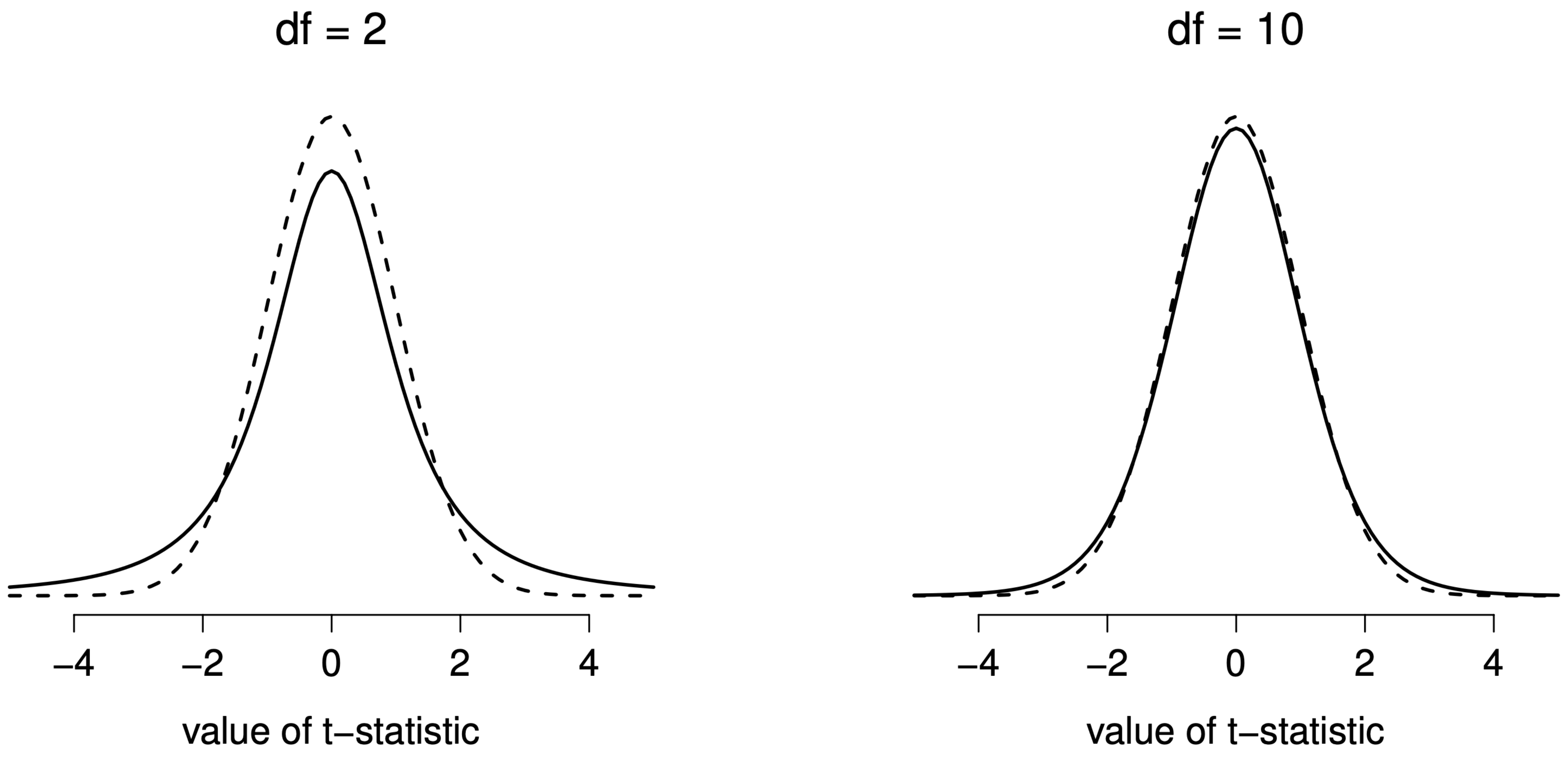
t-distribution

- The t-distribution **fairly similar** to the Normal distribution, except:



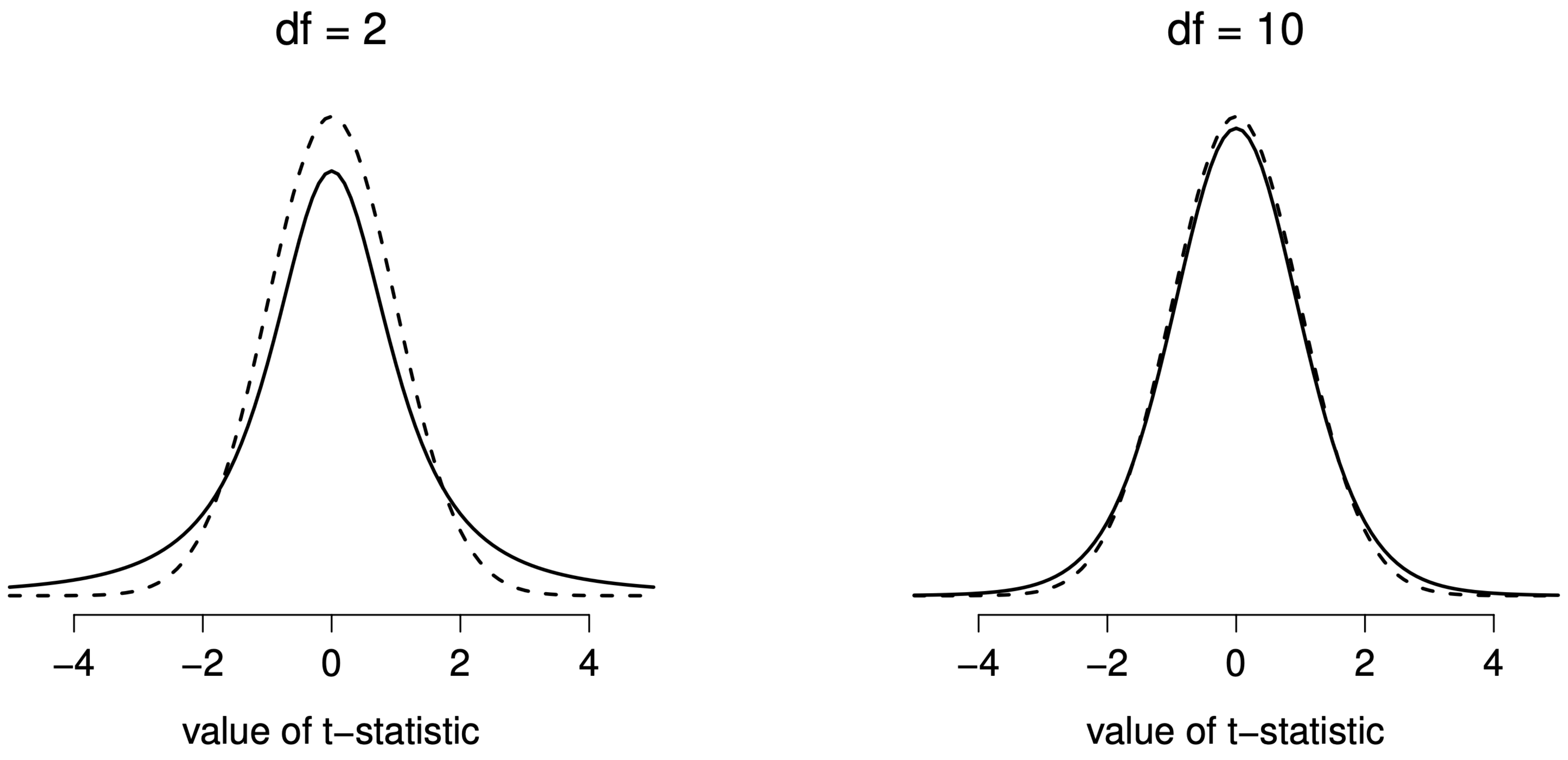
t-distribution

- The t-distribution **fairly similar** to the Normal distribution, except:
  - It has **fatter tails**



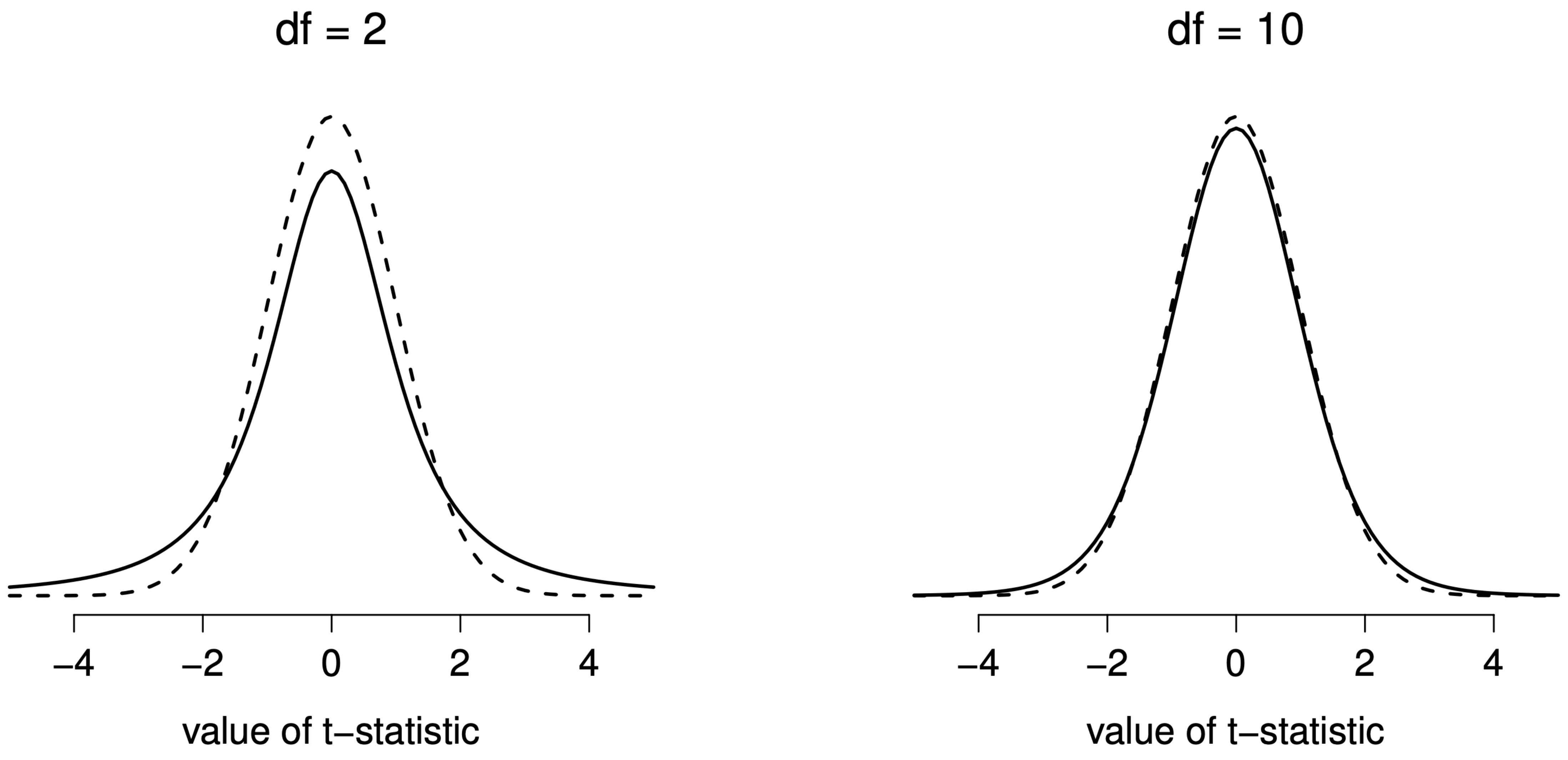
t-distribution

- The t-distribution **fairly similar** to the Normal distribution, except:
  - It has **fatter tails**
  - The fatness of the tails **depends on the sample size** (called "degrees of freedom")



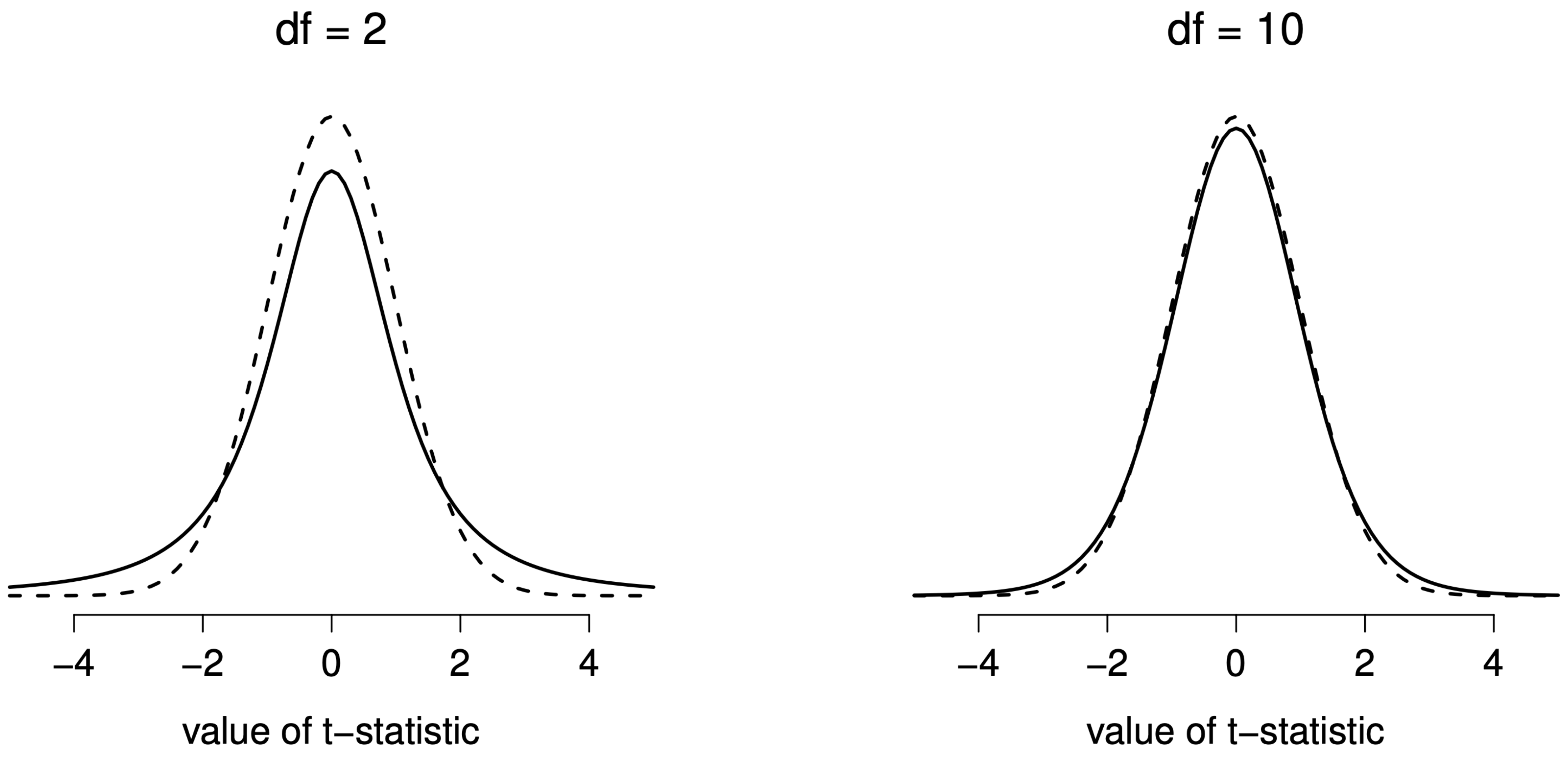
t-distribution

- The t-distribution **fairly similar** to the Normal distribution, except:
  - It has **fatter tails**
  - The fatness of the tails **depends on the sample size** (called "degrees of freedom")
- Lower df → fatter tails



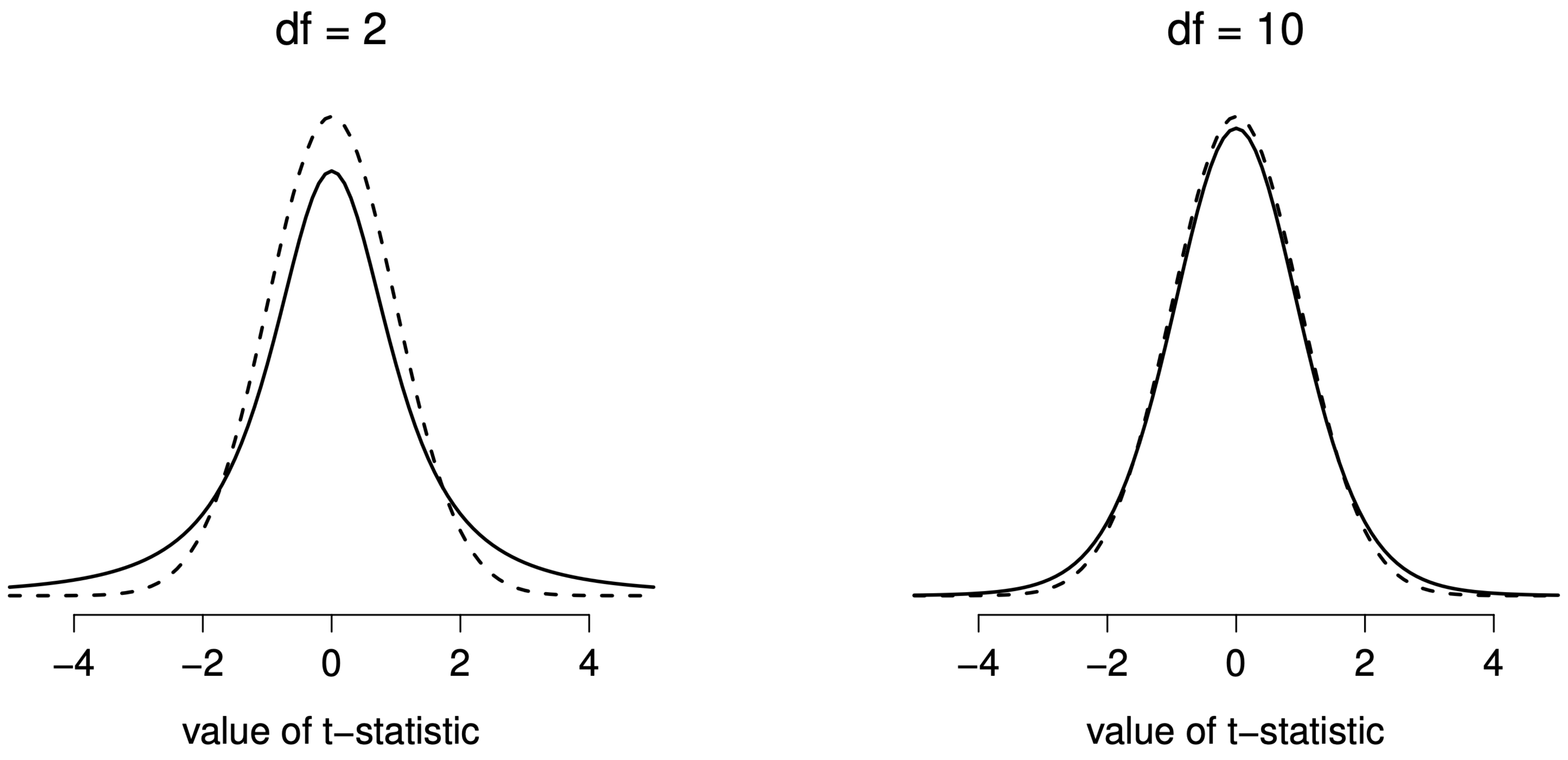
t-distribution

- The t-distribution **fairly similar** to the Normal distribution, except:
  - It has **fatter tails**
  - The fatness of the tails **depends on the sample size** (called "degrees of freedom")
- Lower df → fatter tails
  - Accounts for **uncertainty** in small samples



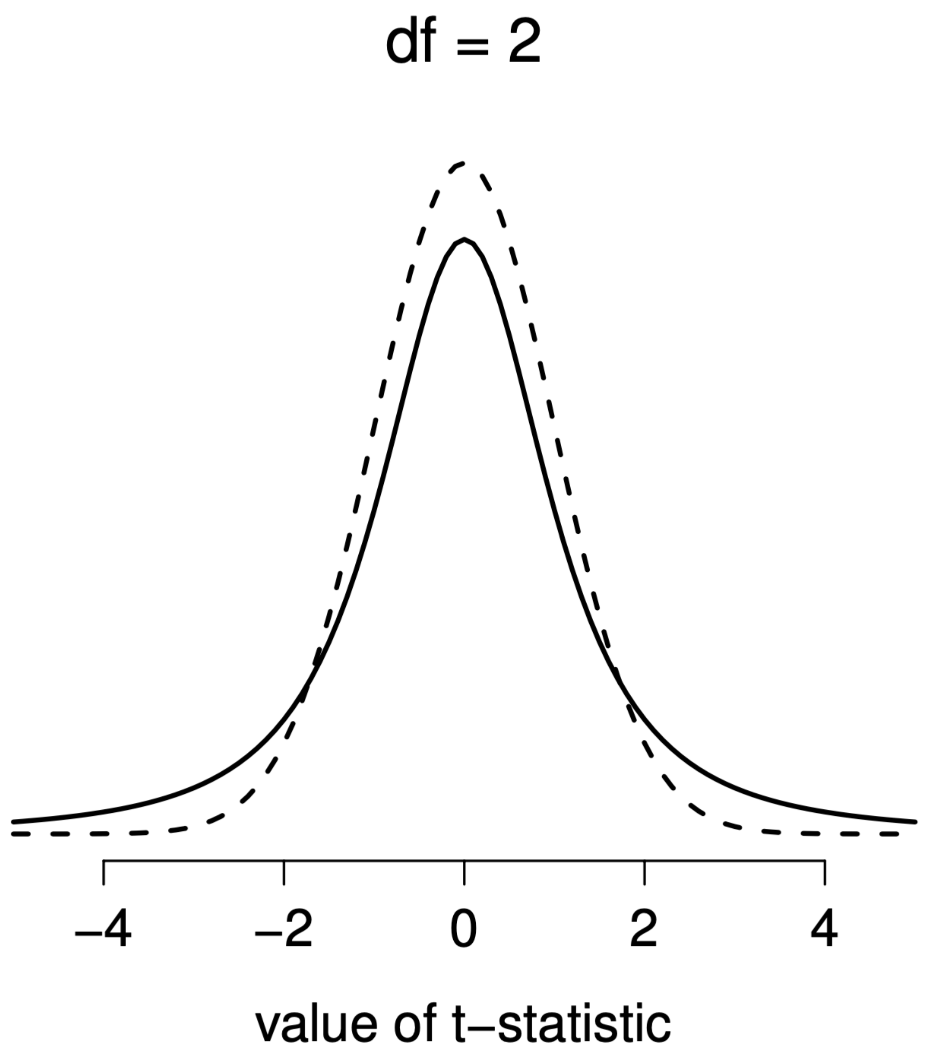
t-distribution

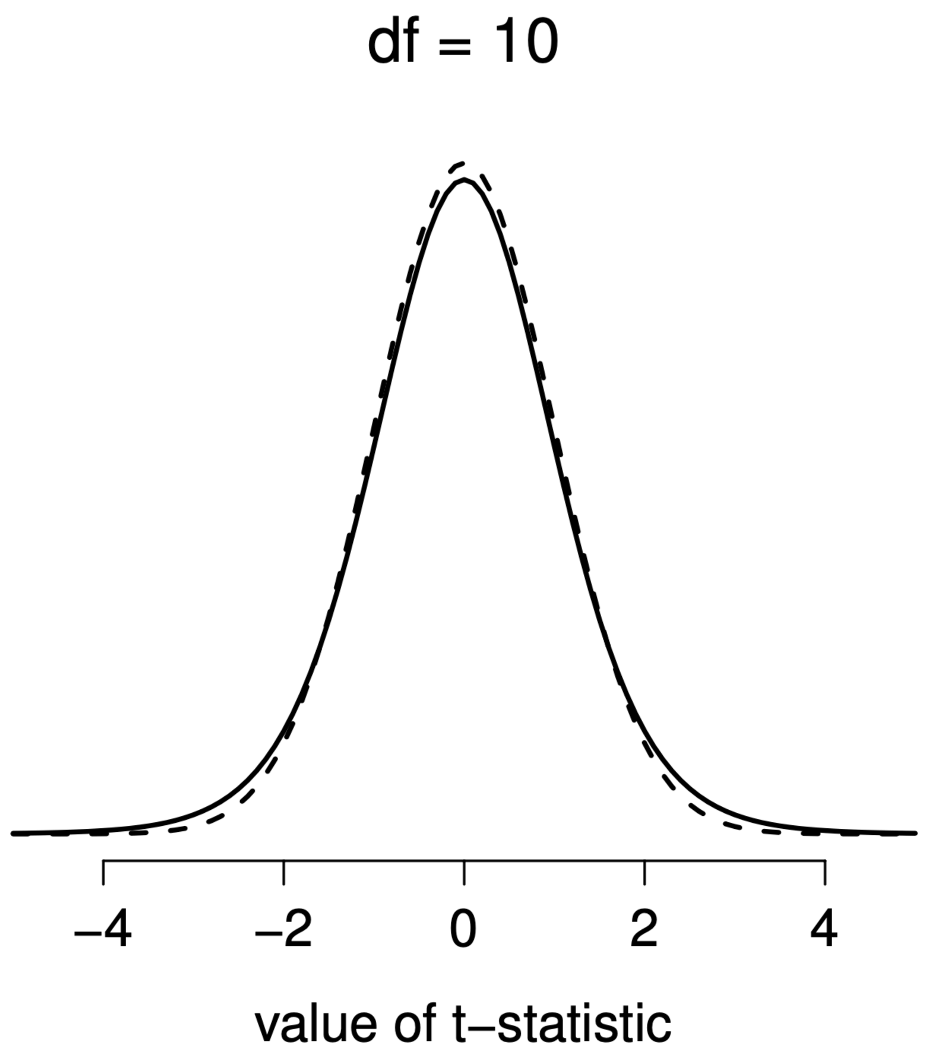
- The t-distribution **fairly similar** to the Normal distribution, except:
  - It has **fatter tails**
  - The fatness of the tails **depends on the sample size** (called "degrees of freedom")
- Lower df → fatter tails
  - Accounts for **uncertainty** in small samples
  - For sample size N, df = $N - 1$



t-distribution

- The t-distribution **fairly similar** to the Normal distribution, except:
  - It has **fatter tails**
  - The fatness of the tails **depends on the sample size** (called "degrees of freedom")
- Lower df → fatter tails
  - Accounts for **uncertainty** in small samples
  - For sample size N, df = $N - 1$
- Higher df → **closer to Normal**

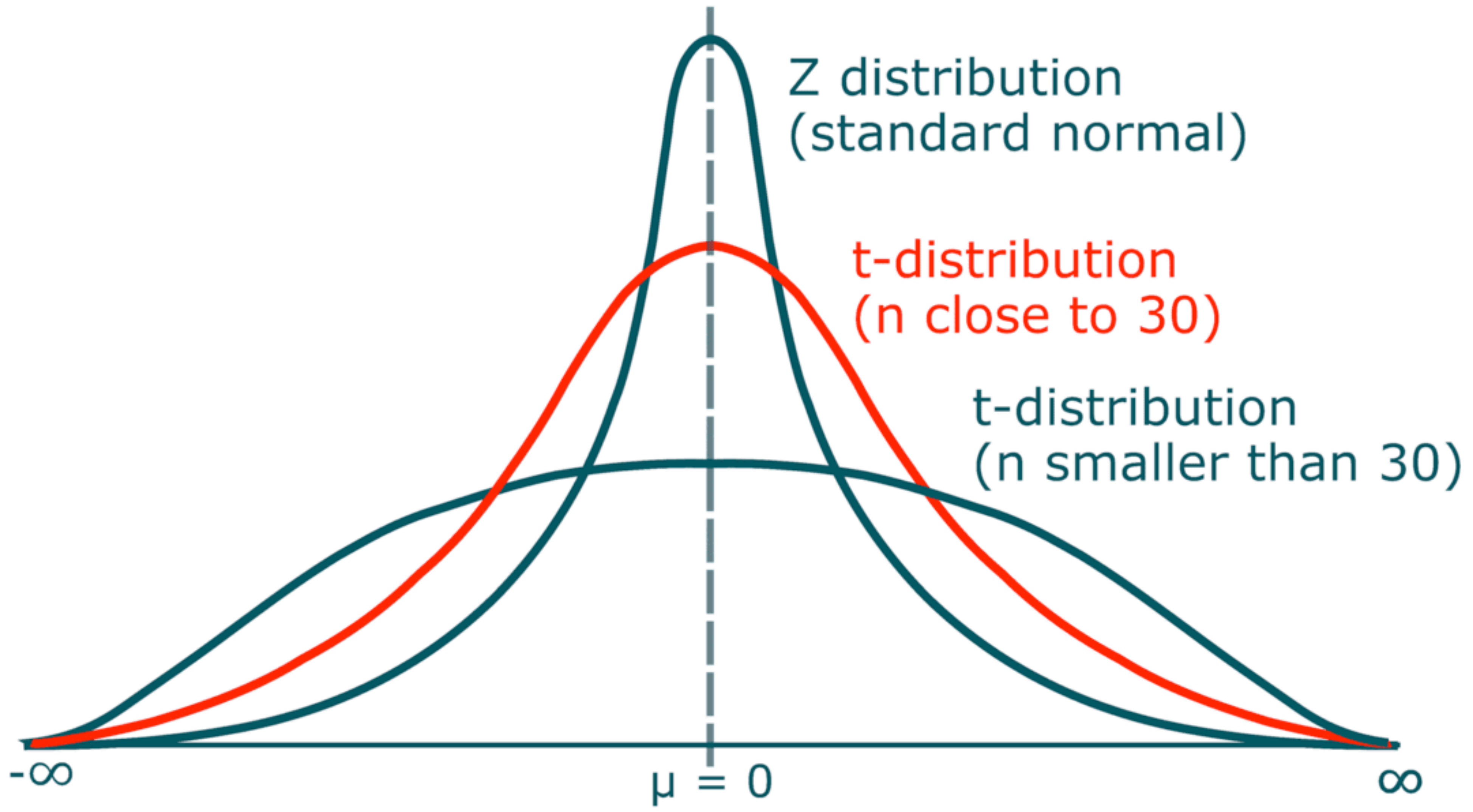




t-score

- Instead of the z-score, we calculate the **t-score**
  - **Identical** to z-score, except for the **estimated stdev**
- **Significance** found by plugging the t-score into the t-distribution
- That's it!

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



t-test in R (from scratch)

```
> null_mean = 67.5
> ling_mean = 72.3
> ling_stdev = 10.1
> N = 20
> # SEM using the estimated stdev from the sample
> SEM = ling_stdev / sqrt(N)
> SEM
[1] 2.258429
>
> t_score = (ling_mean - null_mean) / SEM
> t_score
[1] 2.125372
>
> upper_area = pt(t_score, df=N-1, lower.tail=FALSE)
> upper_area
[1] 0.02344606
> # This is our p-value using the t-test
```

t-test in R (the correct way)

t-test in R (the correct way)

- You can conduct a t-test with R's `t.test()` command
  - `x`: the sample values
  - `mu`: the Null hypothesis mean
  - `alternative="greater"` makes it a **one-sided** test

```
> t.test(x=ling_grades, mu=67.5, alternative="greater")

        One Sample t-test

data:  ling_grades
t = 2.1254, df = 19, p-value = 0.02345
alternative hypothesis: true mean is greater than 67.5
95 percent confidence interval:
 68.39488      Inf
sample estimates:
mean of x
     72.3
```

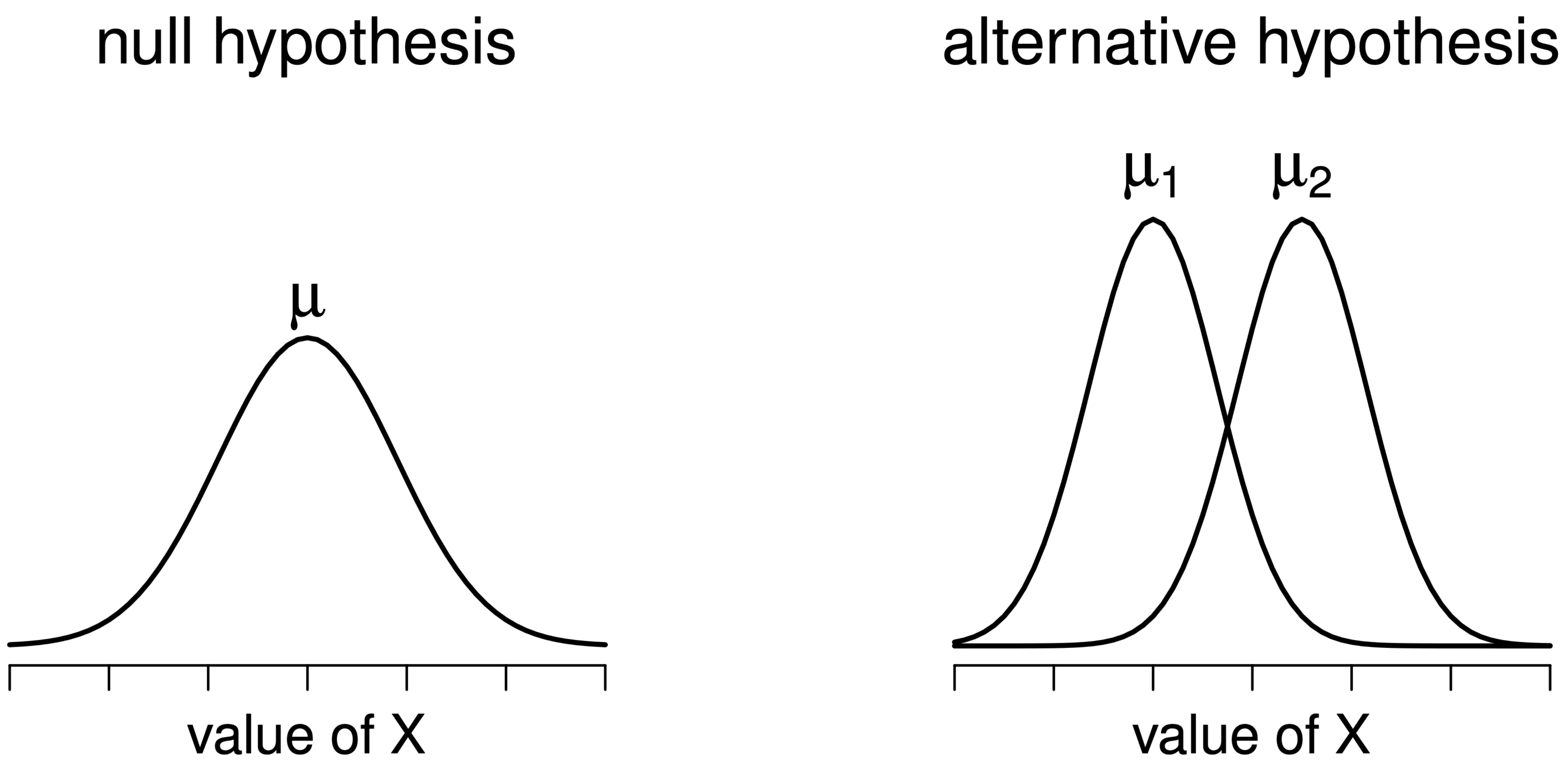
t-test in R (the correct way)

- You can conduct a t-test with R's `t.test()` command
  - `x`: the sample values
  - `mu`: the Null hypothesis mean
  - `alternative="greater"` makes it a **one-sided** test
- Reminder: this is called a **One-Sample t-test**
  - We'll cover other types in less detail

```
> t.test(x=ling_grades, mu=67.5, alternative="greater")

	One Sample t-test

data:  ling_grades
t = 2.1254, df = 19, p-value = 0.02345
alternative hypothesis: true mean is greater than 67.5
95 percent confidence interval:
 68.39488      Inf
sample estimates:
mean of x 
     72.3 
```

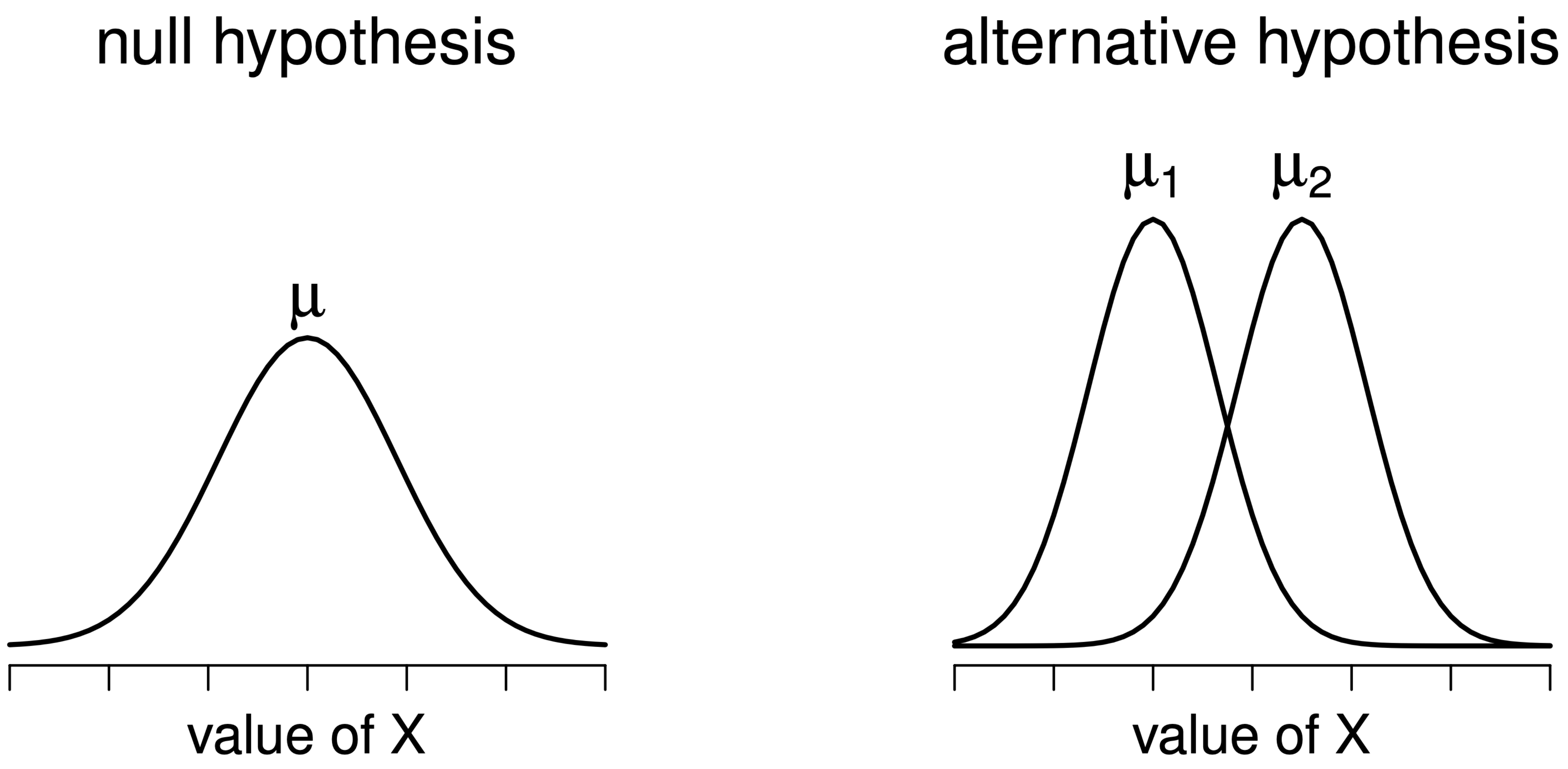
Other t-test varieties

Student's Independent Samples t-test



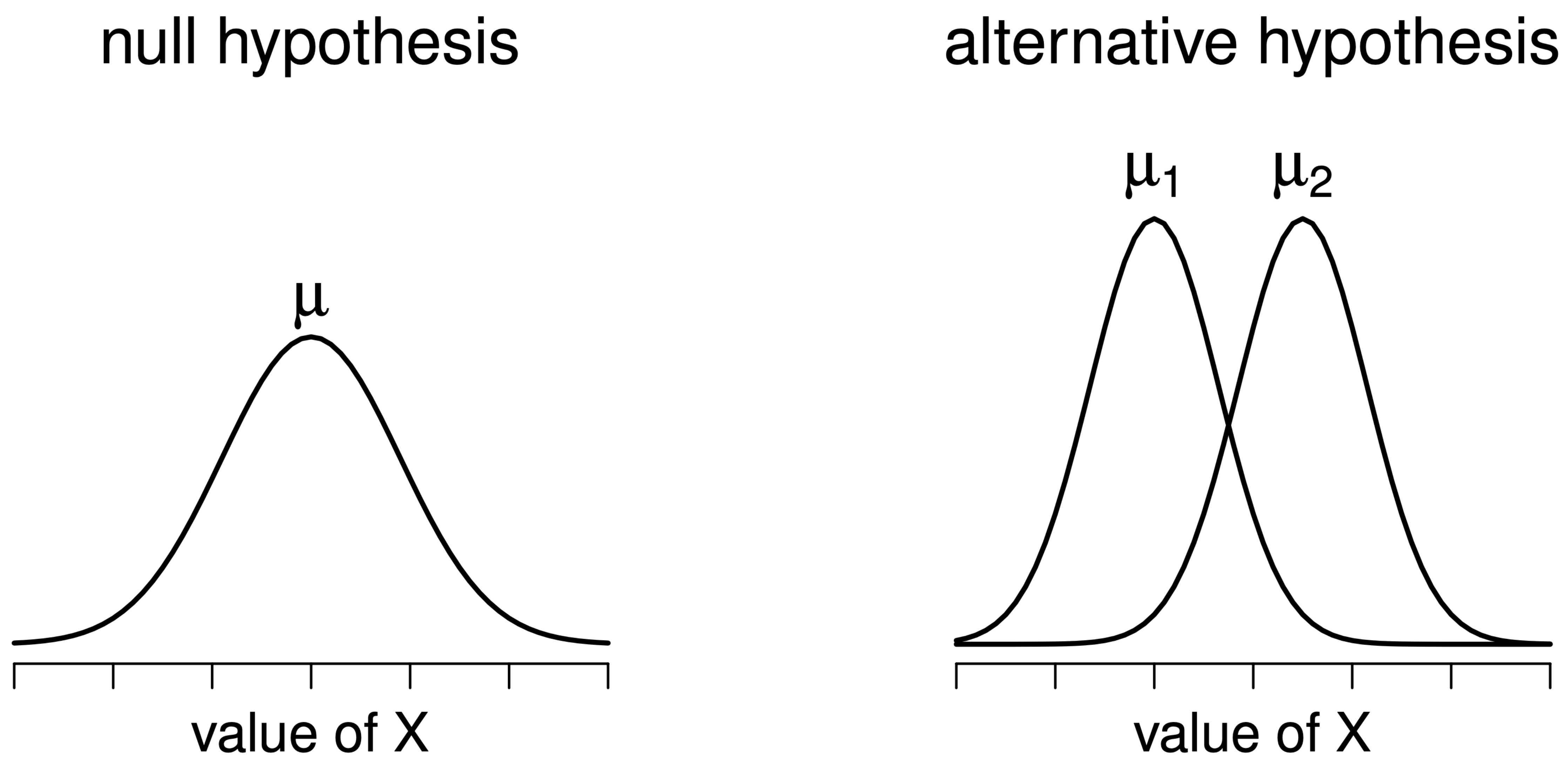
Student's Independent Samples t-test

- There are **two varieties** of t-test for **comparing independent samples**



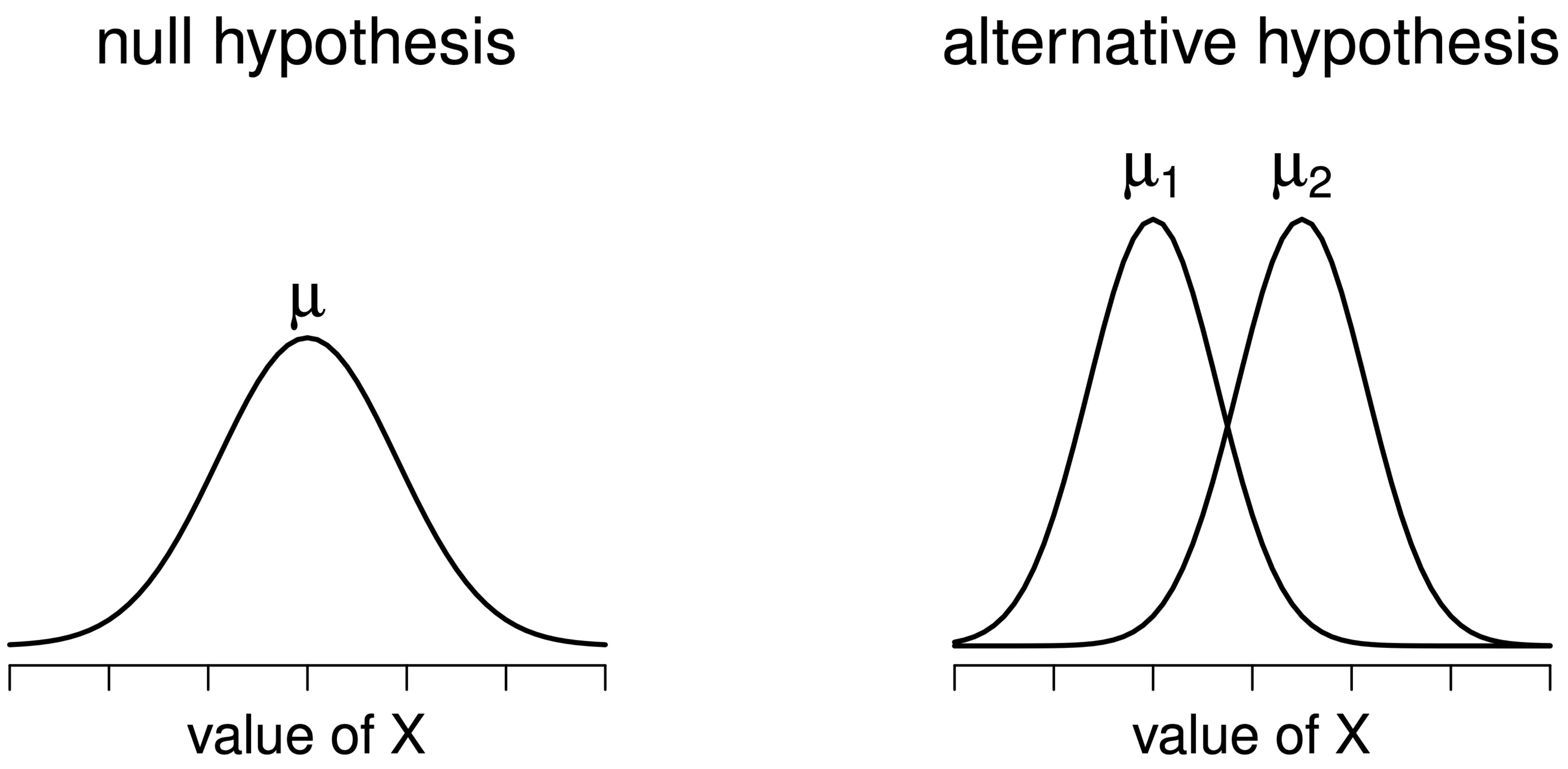
Student's Independent Samples t-test

- There are **two varieties** of t-test for **comparing independent samples**
  - The first is called **"Student's"** (this was the pen-name of the inventor)



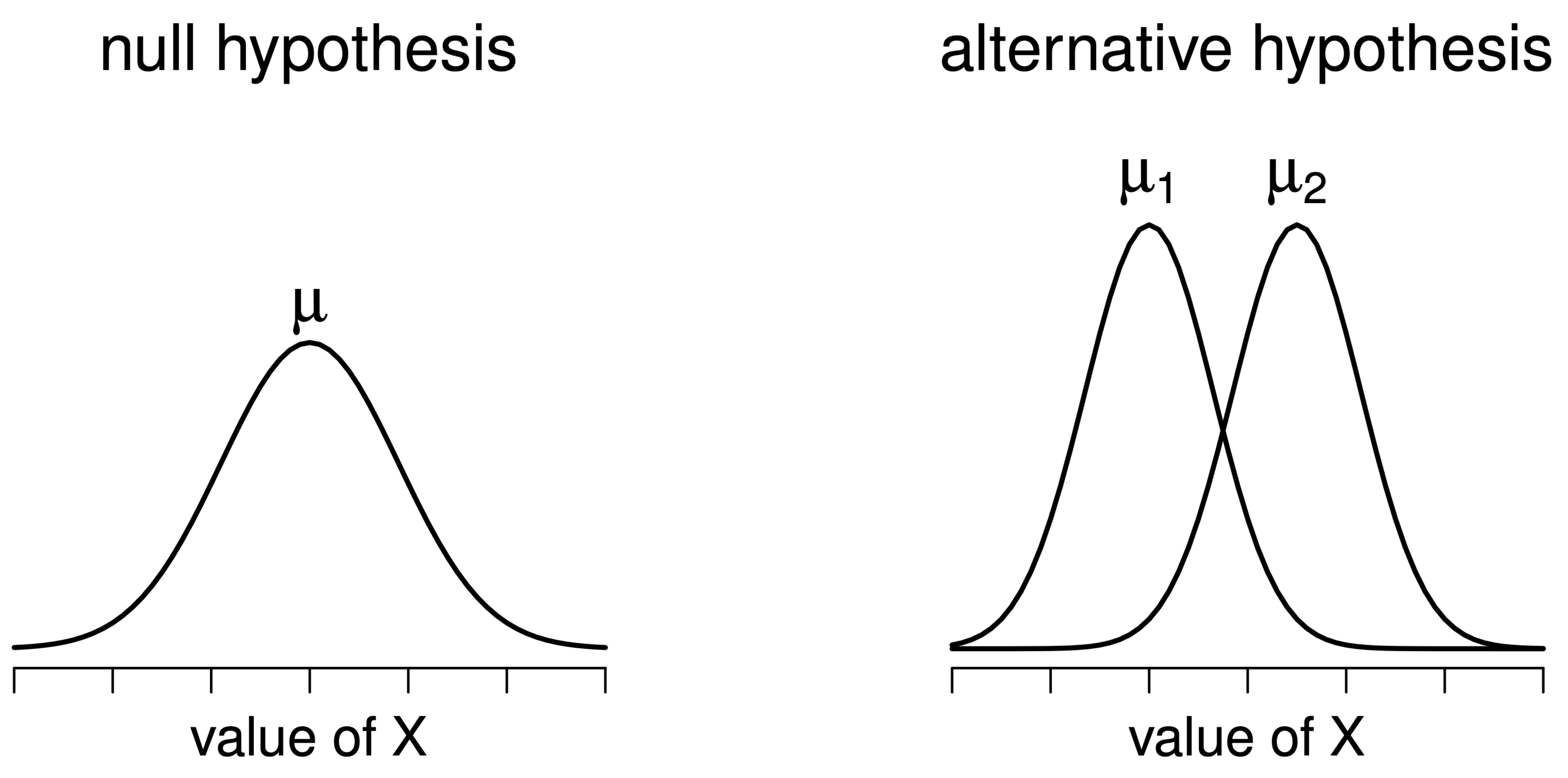
Student's Independent Samples t-test

- There are **two varieties** of t-test for **comparing independent samples**
  - The first is called **"Student's"** (this was the pen-name of the inventor)
- Idea: see if **two samples** have the **same mean**



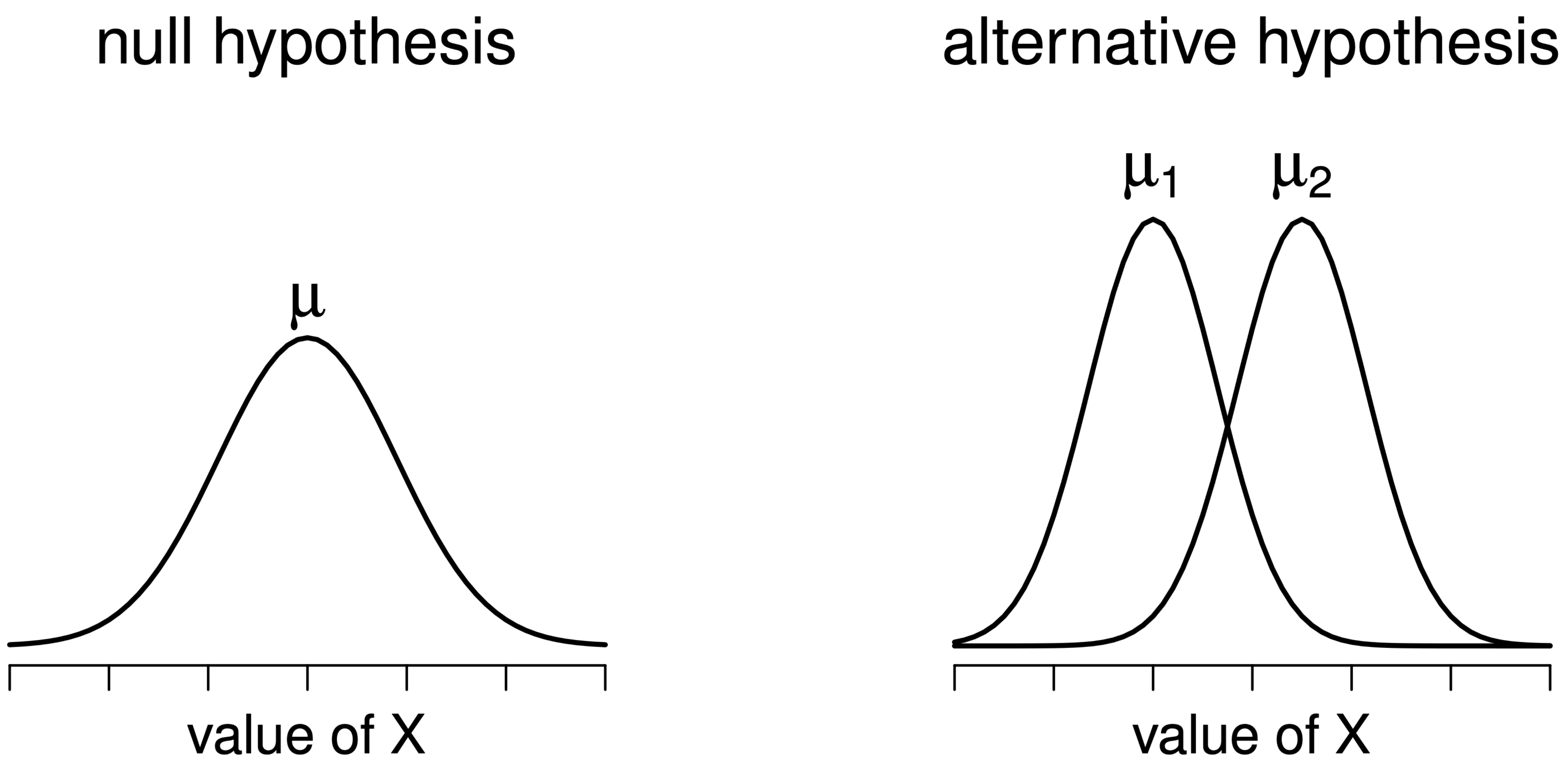
Student's Independent Samples t-test

- There are **two varieties** of t-test for **comparing independent samples**
  - The first is called **"Student's"** (this was the pen-name of the inventor)
- Idea: see if **two samples** have the **same mean**
  - As opposed to one-sample tests, where we compare a sample to a population



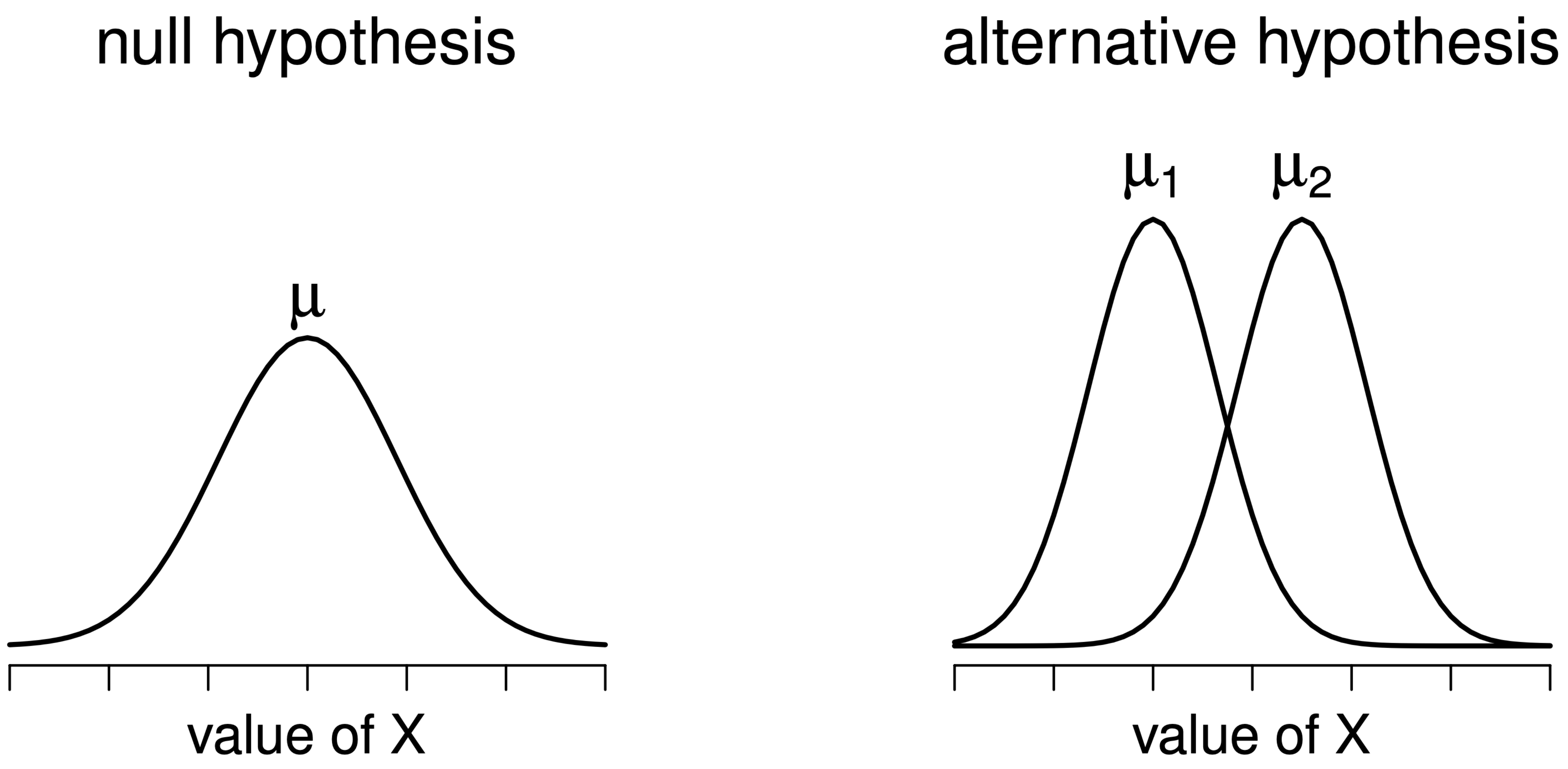
Student's Independent Samples t-test

- There are **two varieties** of t-test for **comparing independent samples**
  - The first is called **"Student's"** (this was the pen-name of the inventor)
- Idea: see if **two samples** have the **same mean**
  - As opposed to one-sample tests, where we compare a sample to a population
  - $H_0 : \mu_1 = \mu_2$



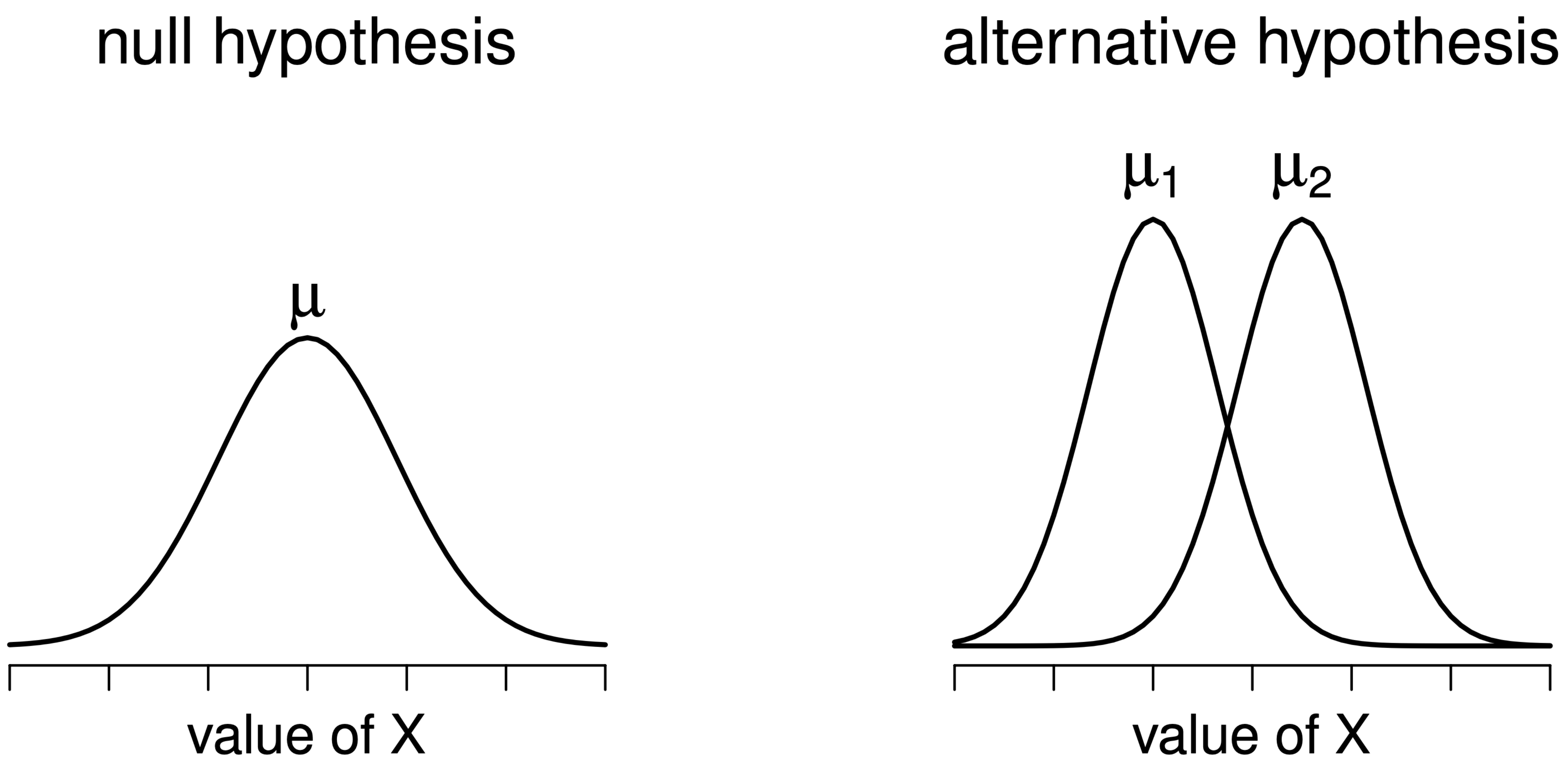
Student's Independent Samples t-test

- There are **two varieties** of t-test for **comparing independent samples**
  - The first is called **"Student's"** (this was the pen-name of the inventor)
- Idea: see if **two samples** have the **same mean**
  - As opposed to one-sample tests, where we compare a sample to a population
  - $H_0 : \mu_1 = \mu_2$
  - $H_1 : \mu_1 \neq \mu_2$



Student's Independent Samples t-test

- There are **two varieties** of t-test for **comparing independent samples**
  - The first is called **"Student's"** (this was the pen-name of the inventor)
- Idea: see if **two samples** have the **same mean**
  - As opposed to one-sample tests, where we compare a sample to a population
  - $H_0 : \mu_1 = \mu_2$
  - $H_1 : \mu_1 \neq \mu_2$
- Example: do **two tutors** get their students **significantly different grades?**



Student's t-test details

Student's t-test details

- Still gets a **t-score** in the end, but has to **pool** the standard deviations of **both samples** together to calculate
  - Don't worry about how this is done. See the book if interested.

Student's t-test details

- Still gets a **t-score** in the end, but has to **pool** the standard deviations of **both samples** together to calculate
  - Don't worry about how this is done. See the book if interested.
- Student's test **assumes** the two populations have the **same standard deviation**
  - This is the key difference from **Welch's t-test**

Student's t-test details

- Still gets a **t-score** in the end, but has to **pool** the standard deviations of **both samples** together to calculate
    - Don't worry about how this is done. See the book if interested.
- Student's test **assumes** the two populations have the **same standard deviation**
    - This is the key difference from **Welch's t-test**
- The two samples are **assumed** to be **independent** from each other
    - i.e. having the **same participant** in both samples would break this assumption

Student's t-test in R

```
> head(tutor_grades)
     grade     tutor
1 93.60010 Anastasia
2 84.59462 Anastasia
3 79.06581 Anastasia
4 61.45601 Anastasia
5 73.03190 Anastasia
6 73.46387 Anastasia
> t.test(formula = grade ~ tutor, data=tutor_grades, var.equal=TRUE)

	Two Sample t-test

data:  grade by tutor
t = 1.1026, df = 31, p-value = 0.2787
alternative hypothesis: true difference in means between group Anasta
sia and group Bernadette is not equal to 0
95 percent confidence interval:
 -3.053475 10.240356
sample estimates:
 mean in group Anastasia mean in group Bernadette
                72.34231                 68.74887
```

Student's t-test in R

- We still use `t.test()`, but the **syntax is different** from one-sample tests

```
> head(tutor_grades)
     grade     tutor
1 93.60010 Anastasia
2 84.59462 Anastasia
3 79.06581 Anastasia
4 61.45601 Anastasia
5 73.03190 Anastasia
6 73.46387 Anastasia
> t.test(formula = grade ~ tutor, data=tutor_grades, var.equal=TRUE)

	Two Sample t-test

data:  grade by tutor
t = 1.1026, df = 31, p-value = 0.2787
alternative hypothesis: true difference in means between group Anasta
sia and group Bernadette is not equal to 0
95 percent confidence interval:
 -3.053475 10.240356
sample estimates:
 mean in group Anastasia mean in group Bernadette
                72.34231                 68.74887
```

Student's t-test in R

- We still use `t.test()`, but the **syntax is different** from one-sample tests
- `formula = grade ~ tutor`
  - Indicates our samples are **by tutor**
  - Grades are our **test statistic**
  - This makes it a **multiple-sample test**

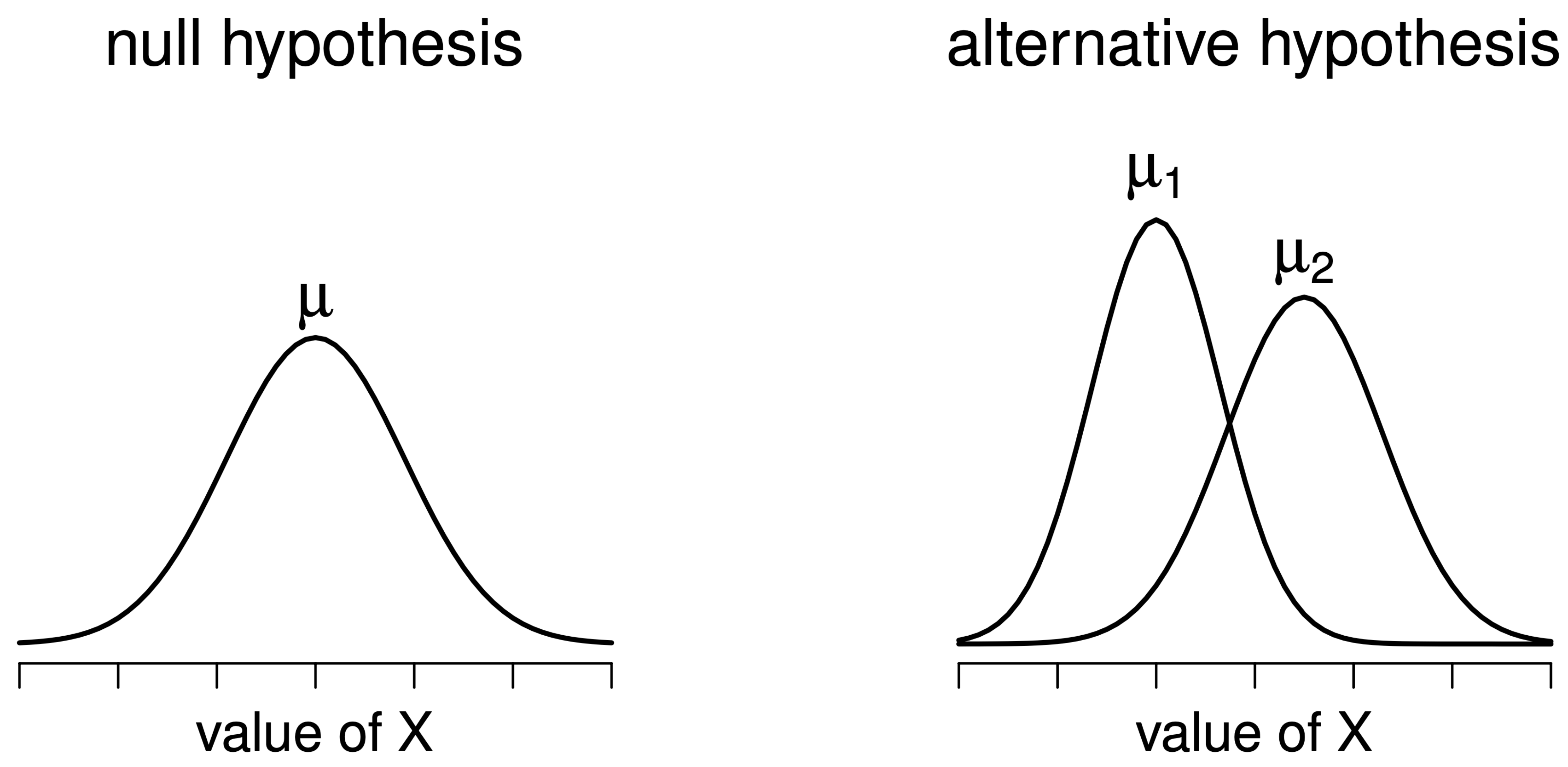
```
> head(tutor_grades)
     grade     tutor
1 93.60010 Anastasia
2 84.59462 Anastasia
3 79.06581 Anastasia
4 61.45601 Anastasia
5 73.03190 Anastasia
6 73.46387 Anastasia
> t.test(formula = grade ~ tutor, data=tutor_grades, var.equal=TRUE)

	Two Sample t-test

data:  grade by tutor
t = 1.1026, df = 31, p-value = 0.2787
alternative hypothesis: true difference in means between group Anasta
sia and group Bernadette is not equal to 0
95 percent confidence interval:
 -3.053475 10.240356
sample estimates:
 mean in group Anastasia mean in group Bernadette
                72.34231                 68.74887
```

Student's t-test in R

- We still use `t.test()`, but the **syntax is different** from one-sample tests
- `formula = grade ~ tutor`
  - Indicates our samples are **by tutor**
  - Grades are our **test statistic**
  - This makes it a **multiple-sample test**
- `var.equal = TRUE` gives us the **Student's t-test**
  - If this is left out or set to FALSE we get **Welch Independent-samples t-test**

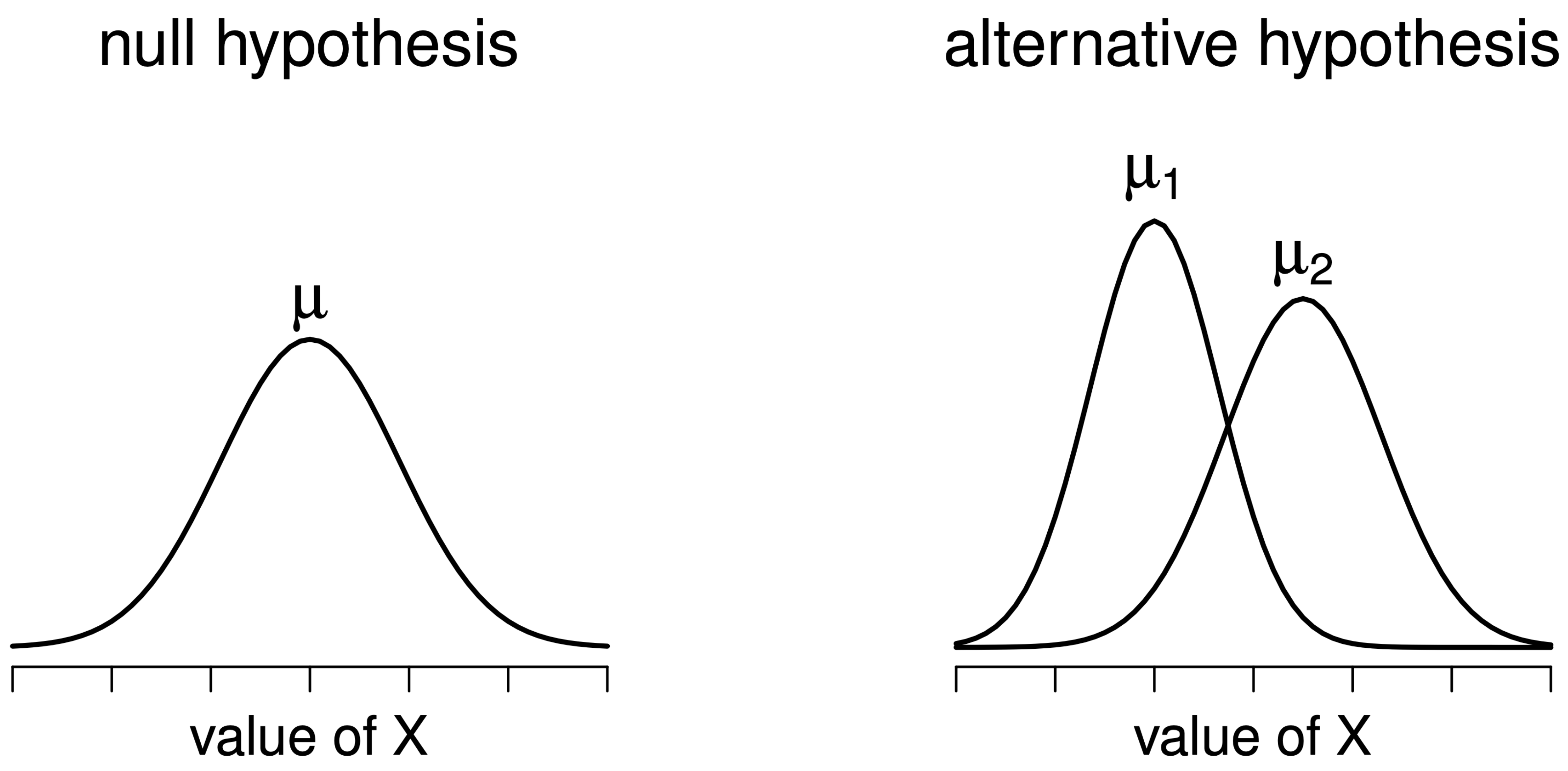
```
> head(tutor_grades)
     grade     tutor
1 93.60010 Anastasia
2 84.59462 Anastasia
3 79.06581 Anastasia
4 61.45601 Anastasia
5 73.03190 Anastasia
6 73.46387 Anastasia
> t.test(formula = grade ~ tutor, data=tutor_grades, var.equal=TRUE)

	Two Sample t-test

data:  grade by tutor
t = 1.1026, df = 31, p-value = 0.2787
alternative hypothesis: true difference in means between group Anasta
sia and group Bernadette is not equal to 0
95 percent confidence interval:
 -3.053475 10.240356
sample estimates:
 mean in group Anastasia mean in group Bernadette 
                72.34231                 68.74887 
```

Welch Independent Samples t-test



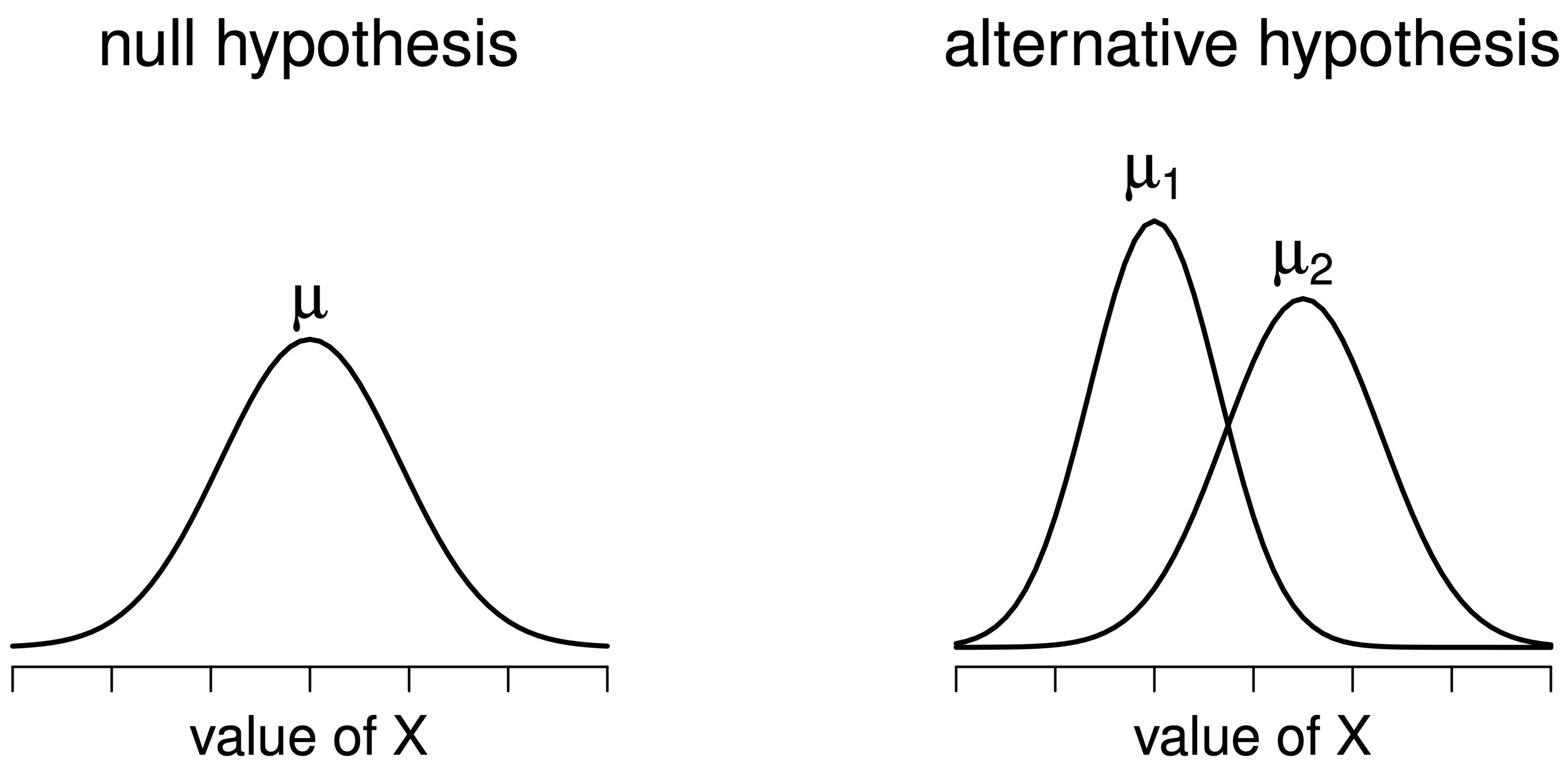
Welch Independent Samples t-test

- Welch's test is **just like Student's**, except we **don't assume the same stdev** for each population
  - Again, there's some tricky math you shouldn't worry about for this course



Welch Independent Samples t-test

- Welch's test is **just like Student's**, except we **don't assume the same stdev** for each population
  - Again, there's some tricky math you shouldn't worry about for this course
- Setting `var.equal = FALSE` in `t.test()` gets you the Welch test
  - This is the **default value**, so you can also leave the argument out



Reporting t-test results

Reporting t-test results

- Fields of science have various **conventions** for reporting the **results of statistical tests**, which might look something like this:
  - "With a mean grade of 72.3, the Linguistics students scored slightly higher grades than the class average of 67.5 ($t(19) = 2.25$, $p < 0.05$); the 95% confidence interval is [67.8, 76.8]"

Reporting t-test results

- Fields of science have various **conventions** for reporting the **results of statistical tests**, which might look something like this:
  - "With a mean grade of 72.3, the Linguistics students scored slightly higher grades than the class average of 67.5 ($t(19) = 2.25$, $p < 0.05$); the 95% confidence interval is [67.8, 76.8]"
- $t(19) = 2.25$: the t-score with 19 degrees of freedom was 2.25

Reporting t-test results

- Fields of science have various **conventions** for reporting the **results of statistical tests**, which might look something like this:
    - "With a mean grade of 72.3, the Linguistics students scored slightly higher grades than the class average of 67.5 ($t(19) = 2.25$, $p < 0.05$); the 95% confidence interval is [67.8, 76.8]"
- $t(19) = 2.25$: the t-score with 19 degrees of freedom was 2.25
- $p < 0.05$: the p-value was less than a 0.05 significance threshold

Reporting t-test results

- Fields of science have various **conventions** for reporting the **results of statistical tests**, which might look something like this:
  - "With a mean grade of 72.3, the Linguistics students scored slightly higher grades than the class average of 67.5 ($t(19) = 2.25$, $p < 0.05$); the 95% confidence interval is [67.8, 76.8]"
- $t(19) = 2.25$: the t-score with 19 degrees of freedom was 2.25
- $p < 0.05$: the p-value was less than a 0.05 significance threshold
- 95% confidence interval : the range the model thinks that the Linguistics student population mean falls into (sometimes abbreviated CI_{95})