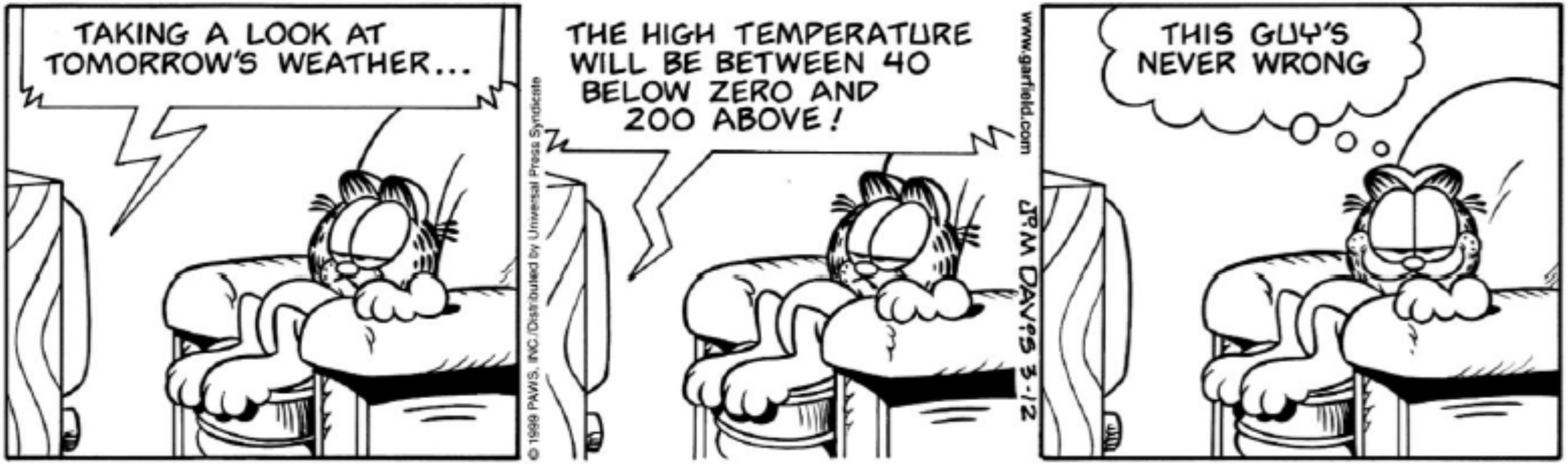
Lecture 06

10.12.21



Refresher Quiz

The heights of men in a certain population follow a normal distribution with mean 69.7 inches and standard deviation 2.8 inches.

a) If a man is chosen at random from the population, find the probability that he will be more than 72 inches tall.

b) If two men were chosen at random from the population, find the probability that (i) both of them will be more than 72 inches tall; (ii) their mean height will be more than 72 inches.

The heights of men in a certain population follow a normal distribution with mean 69.7 inches and standard deviation 2.8 inches.

a) If a man is chosen at random from the population, find the probability that he will be more than 72 inches tall.

```
> pnorm(72, 69.7, 2.8, lower.tail = F)
```

= 20.5%

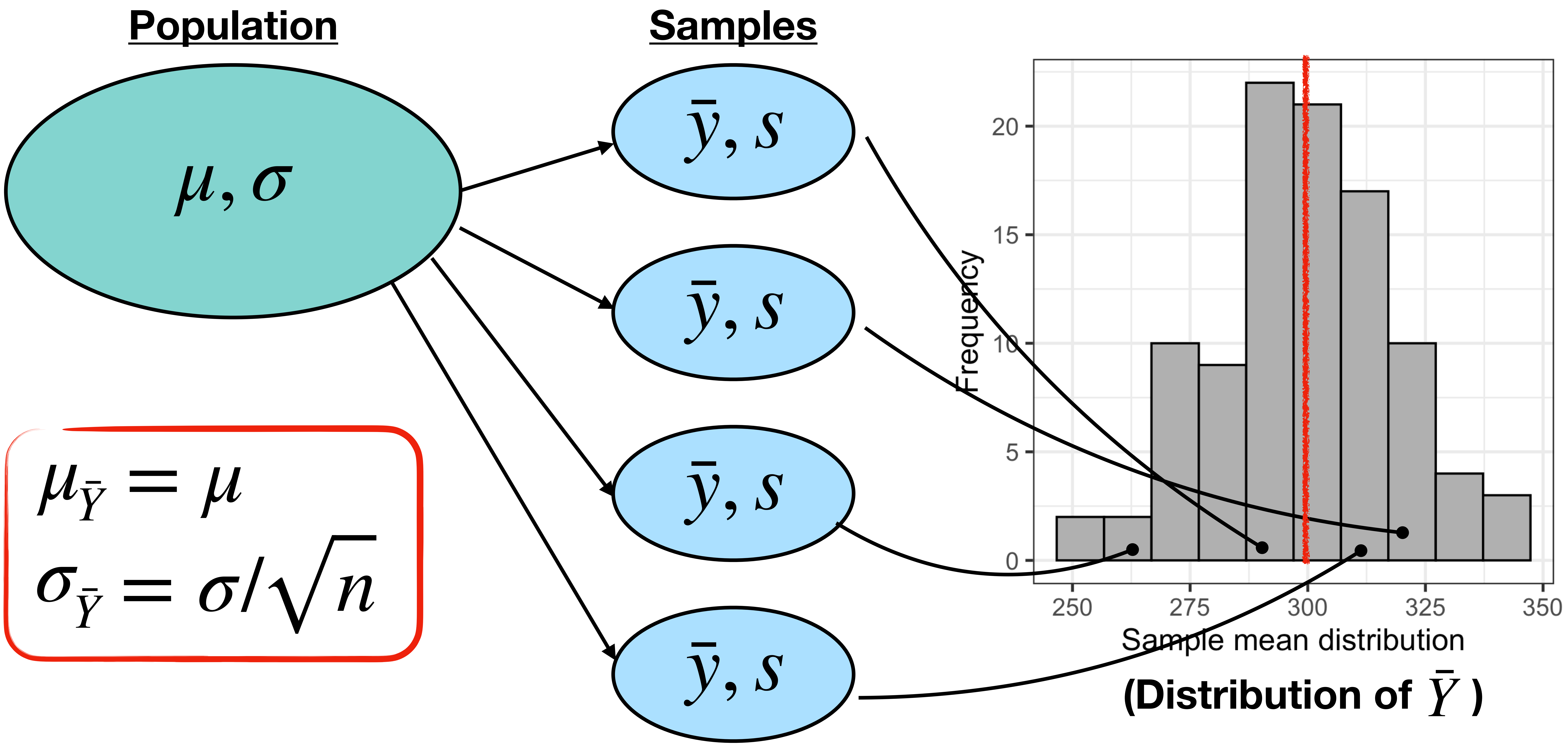
b) If two men were chosen at random from the population, find the probability that (i) both of them will be more than 72 inches tall; (ii) their mean height will be more than 72 inches.

P(both > 72 in tall) = P(> 72) * P(> 72) = 0.205 * 0.205 = 4.20%

```
> pnorm(72, 69.7, 2.8/sqrt(2), lower.tail = F)
```

= 12.26%

Summary: sampling distribution



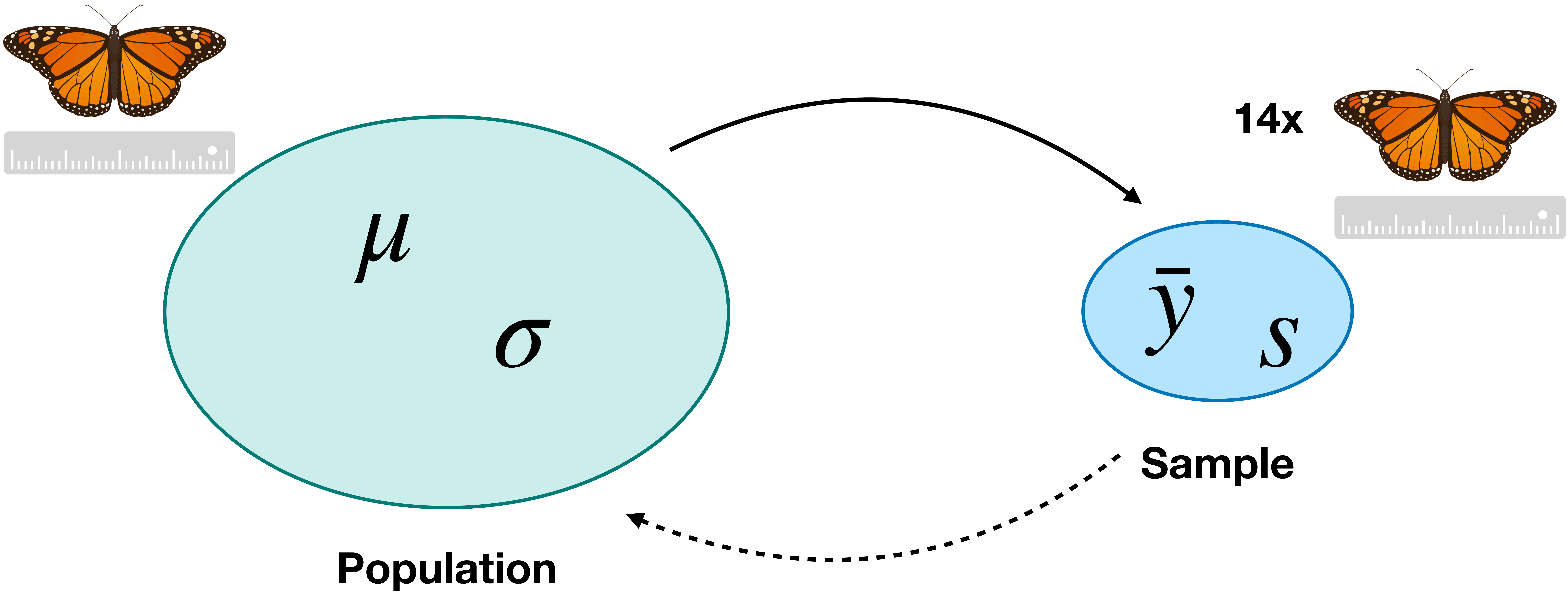
$$\boxed{\begin{aligned}\mu_{\bar{Y}} &= \mu \\ \sigma_{\bar{Y}} &= \sigma/\sqrt{n}\end{aligned}}$$

Statistical estimation

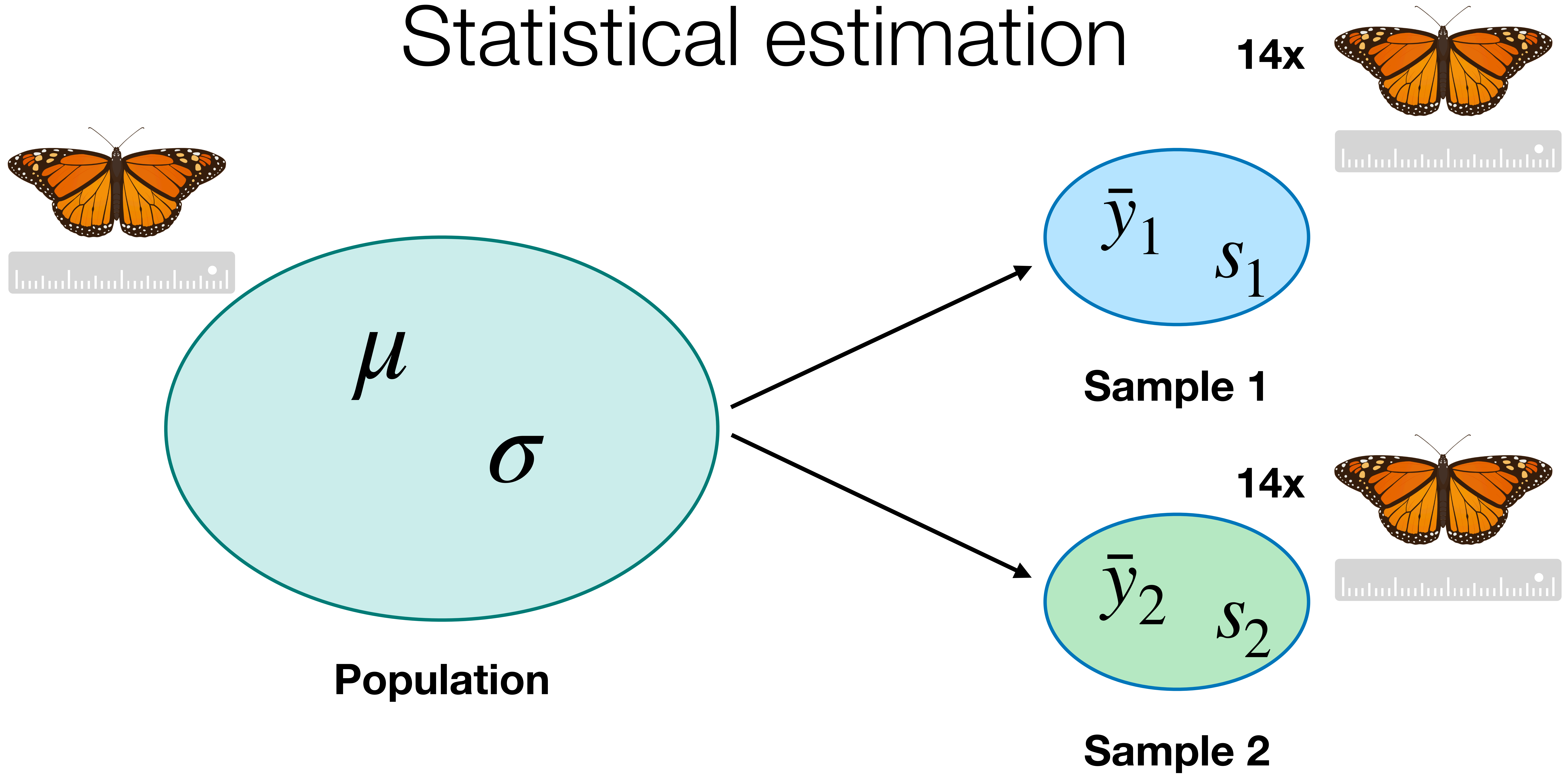
- We view our data as a **random sample from a population** and use the information about <u>our data</u> to *infer* facts about the <u>population</u>
- Goals:
    - (1) Determine estimate of some feature of the population (i.e. mean)
    - (2) Assess the precision of the estimate

Statistical estimation

Statistical estimation



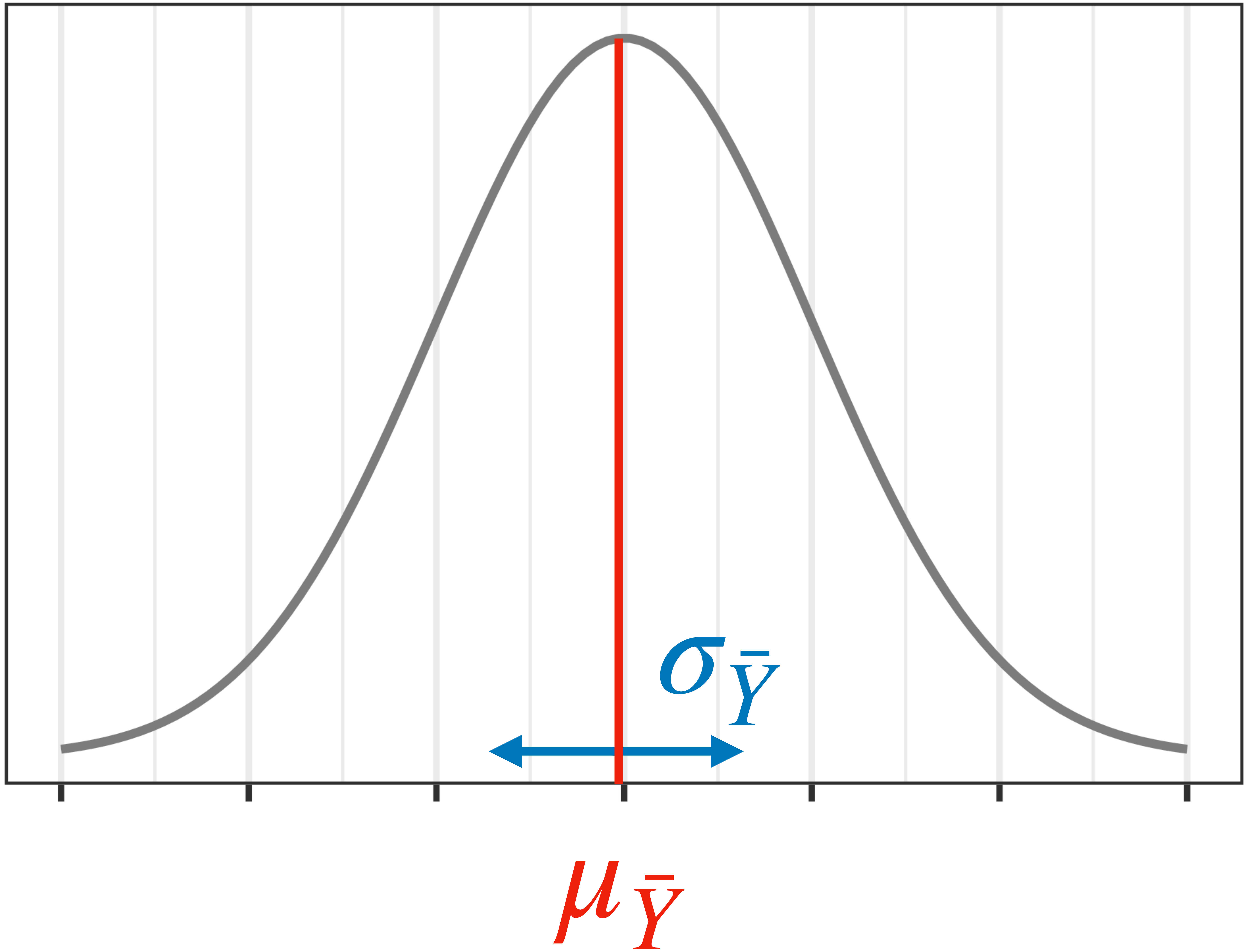
Statistical estimation



Standard error of the mean is estimated from the sampling distribution

Standard deviation of sampling distribution

$$\sigma_{\bar{Y}} = \frac{\sigma}{\sqrt{n}}$$



Standard error of the mean

$$SE_{\bar{Y}} = \frac{s}{\sqrt{n}}$$

Standard error of the mean is estimated from the sampling distribution

- **Standard error of the mean (SE)** is a measure of <u>reliability</u> or <u>precision</u> of the sample mean as an estimate of the population mean
- SE incorporates the two factors that influence reliability:
    - Variability of observations (s)
    - Sample size (*n*)

Standard error of the mean

$$SE_{\bar{Y}} = \frac{s}{\sqrt{n}}$$

Standard error of the mean is estimated from the sampling distribution

n = 14

$\bar{y} = 32.81cm^2$

$s = 2.48cm^2$

$$SE_{\bar{Y}} = \frac{s}{\sqrt{n}}$$

Standard error of the mean is estimated from the sampling distribution

***n* = 14**

$$\bar{y} = 32.81 cm^2$$

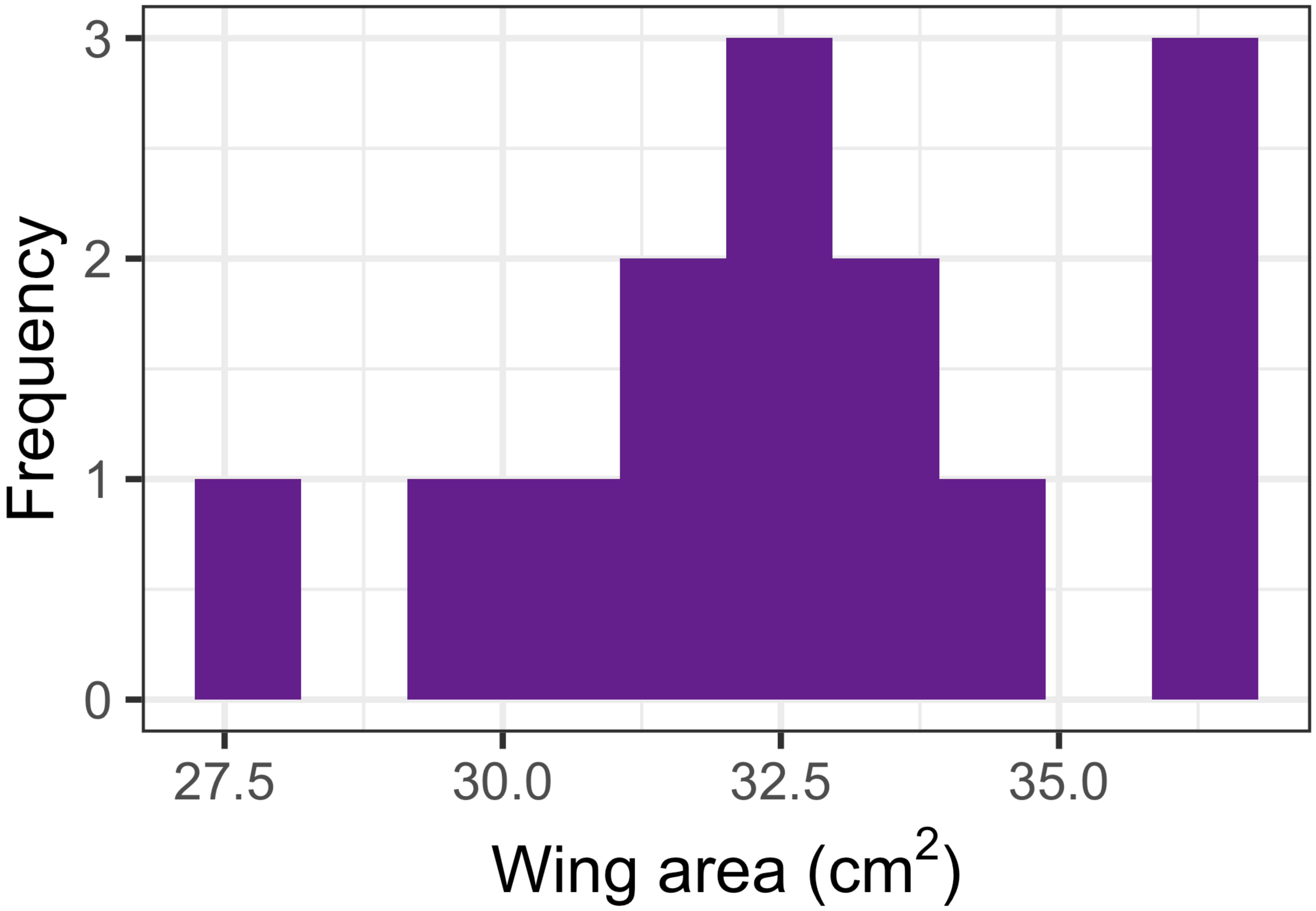
$$s = 2.48 cm^2$$

$$SE_{\bar{Y}} = \frac{2.48}{\sqrt{14}}$$

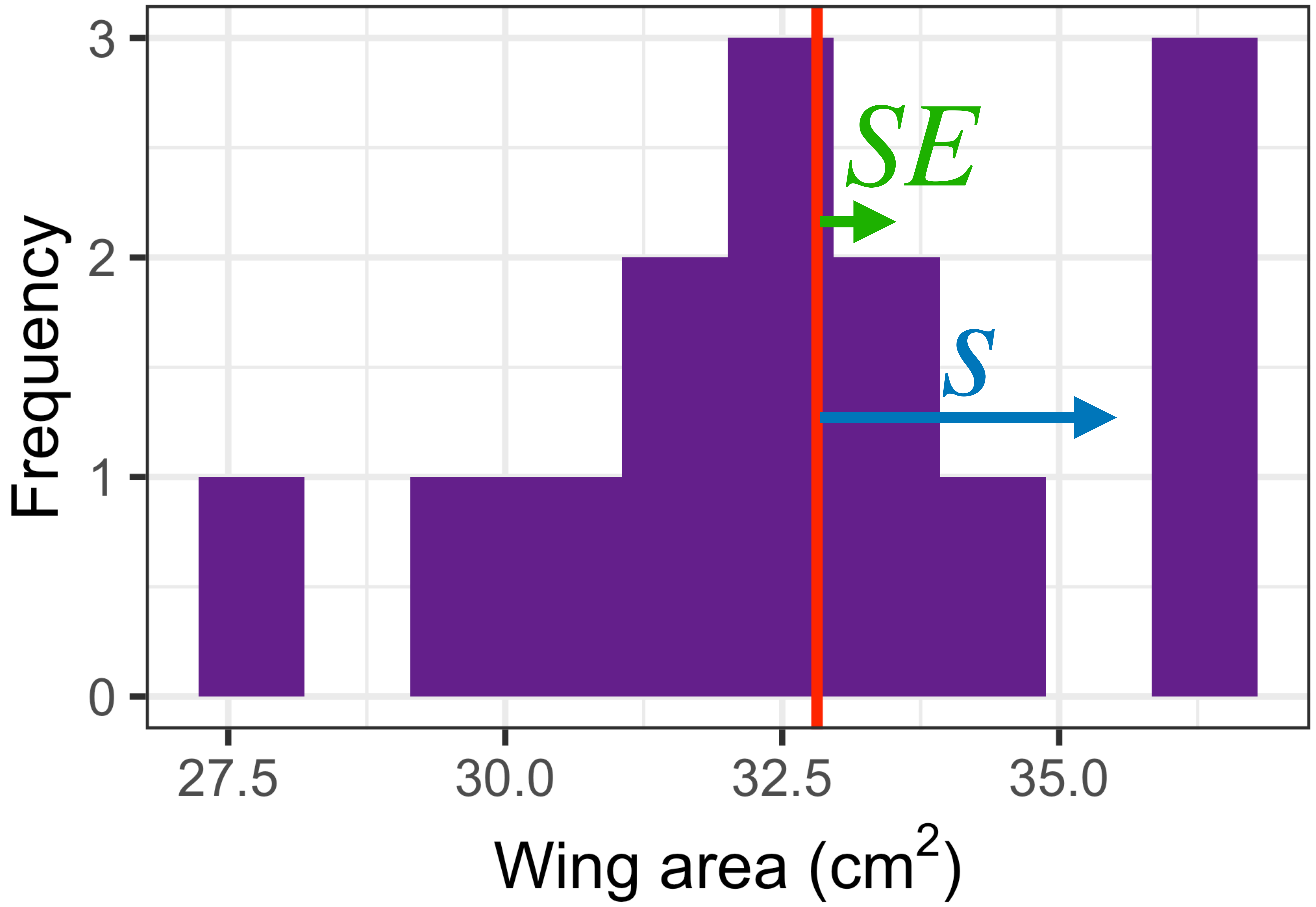
$$= 0.66 cm^2$$

What does the SE mean?

Standard error of the mean is estimated from the sampling distribution

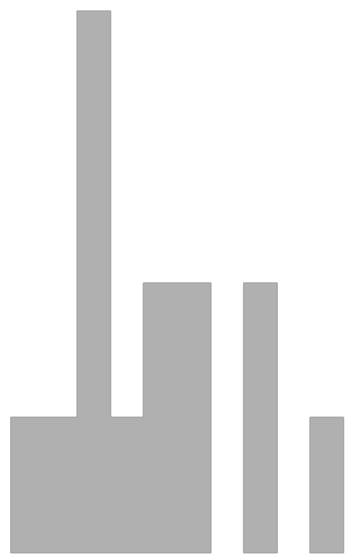


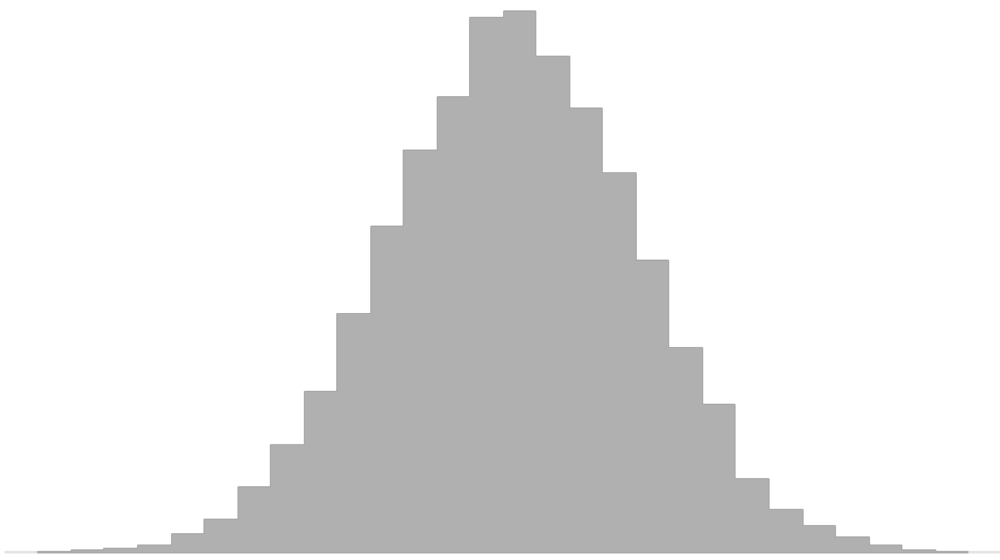
Standard error of the mean is estimated from the sampling distribution



Standard error (SE) versus **standard deviation (SD)**

	$n = 14$	$n = 140$	$n = 1400$	$n \to \infty$
$\bar{y}$	32.69	32.12	31.97	$\bar{y} \to \mu$

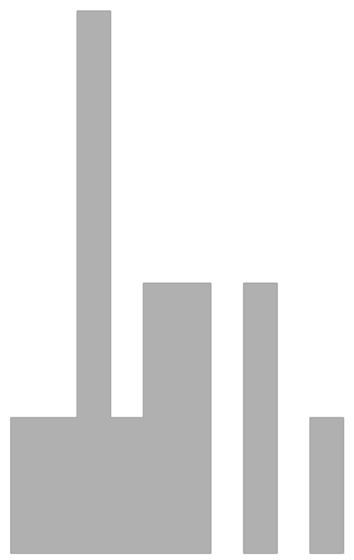


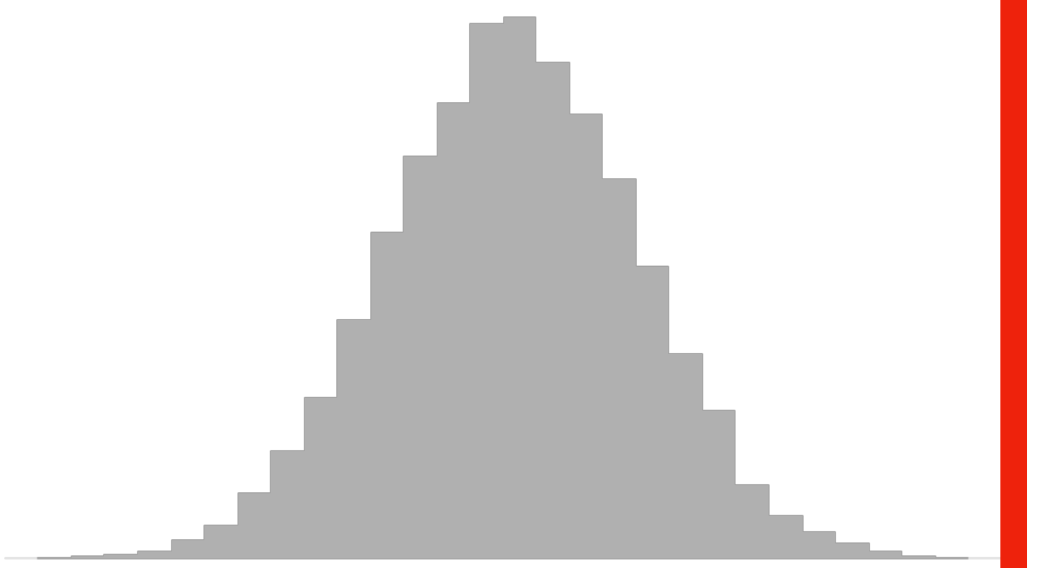



$\bar{y} = 32.81cm^2$ $\quad$ $s = 2.48cm^2$

Standard error (SE) versus **standard deviation (SD)**

	$n = 14$	$n = 140$	$n = 1400$	$n \to \infty$
$\bar{y}$	32.69	32.12	31.97	$\bar{y} \to \mu$
s	1.68	2.47	2.52	$s \to \sigma$

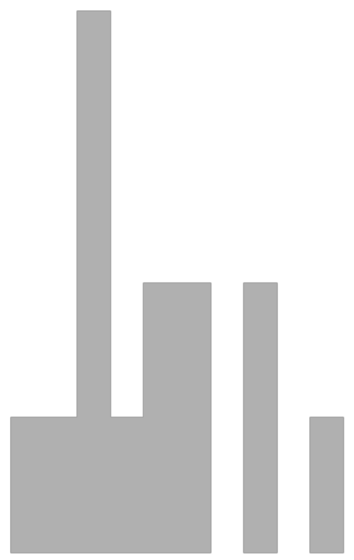


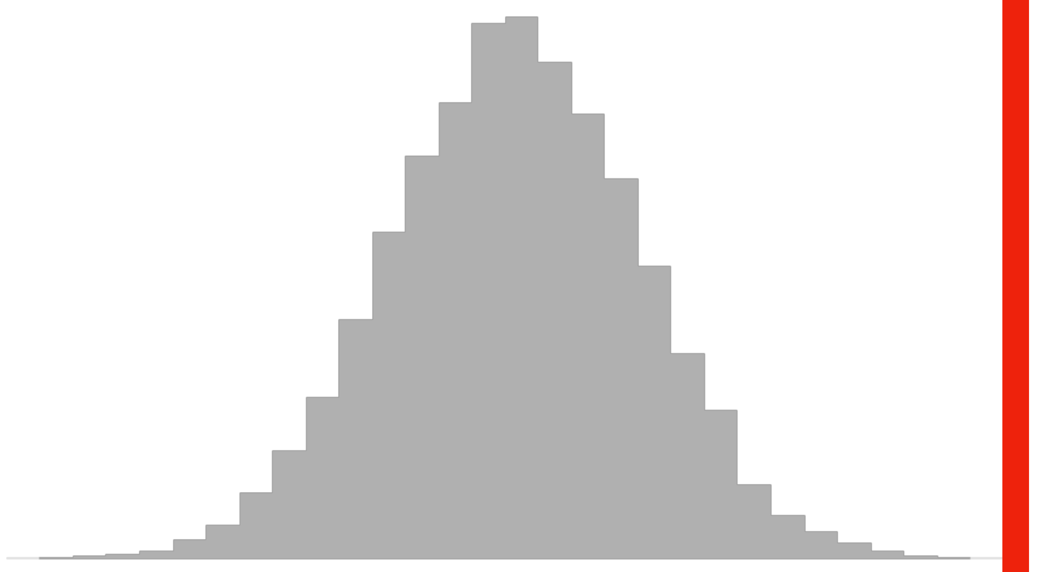



$\bar{y} = 32.81cm^2$ $s = 2.48cm^2$

Standard error (SE) versus **standard deviation (SD)**

	$n = 14$	$n = 140$	$n = 1400$	$n \to \infty$
$\bar{y}$	32.69	32.12	31.97	$\bar{y} \to \mu$
s	1.68	2.47	2.52	$s \to \sigma$
SE	0.451	0.209	0.067	$SE \to 0$

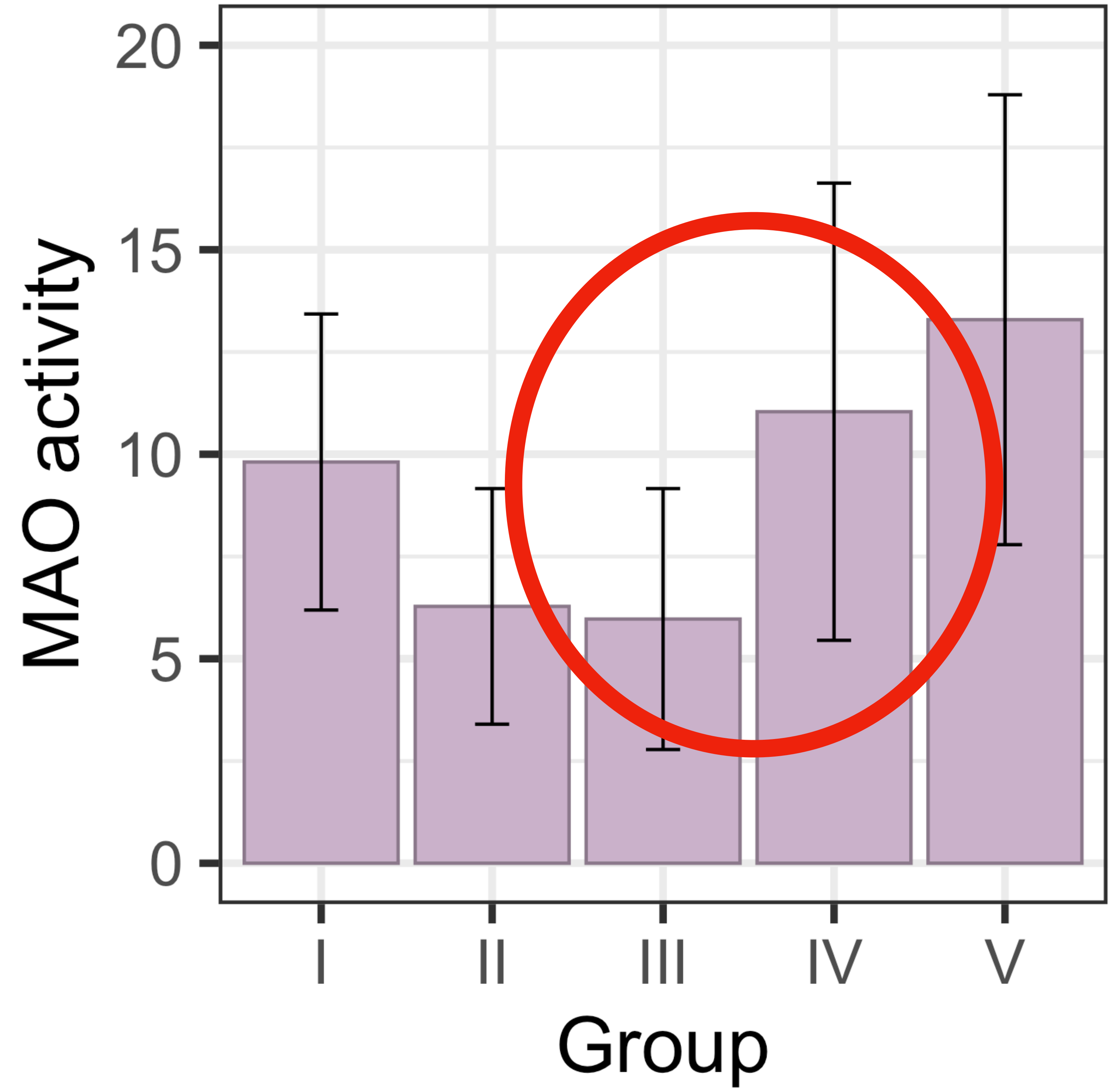


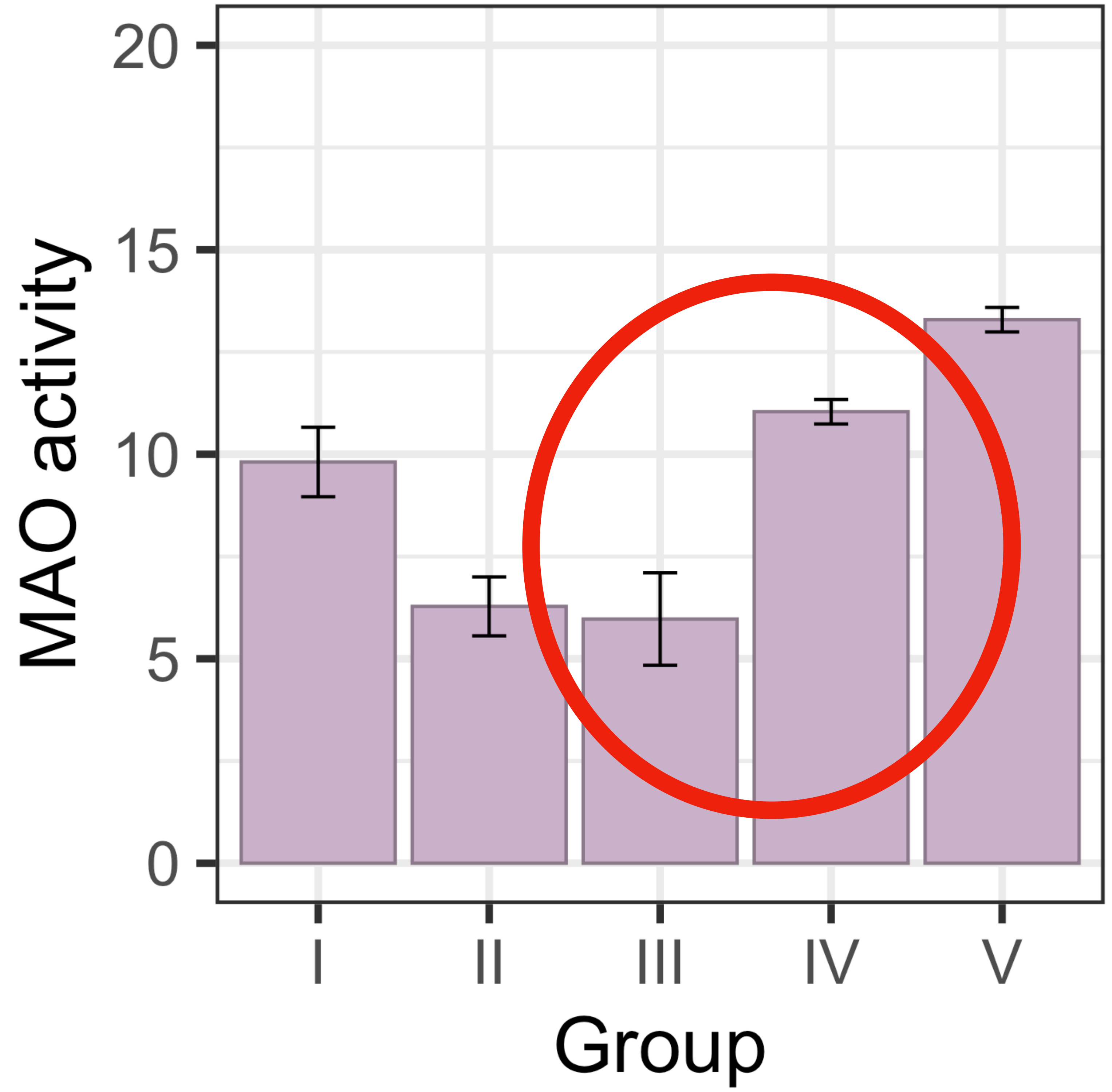
$\bar{y} = 32.81 cm^2$ $s = 2.48 cm^2$

Standard error (SE)
versus **standard deviation (SD)**

SD = dispersion of data



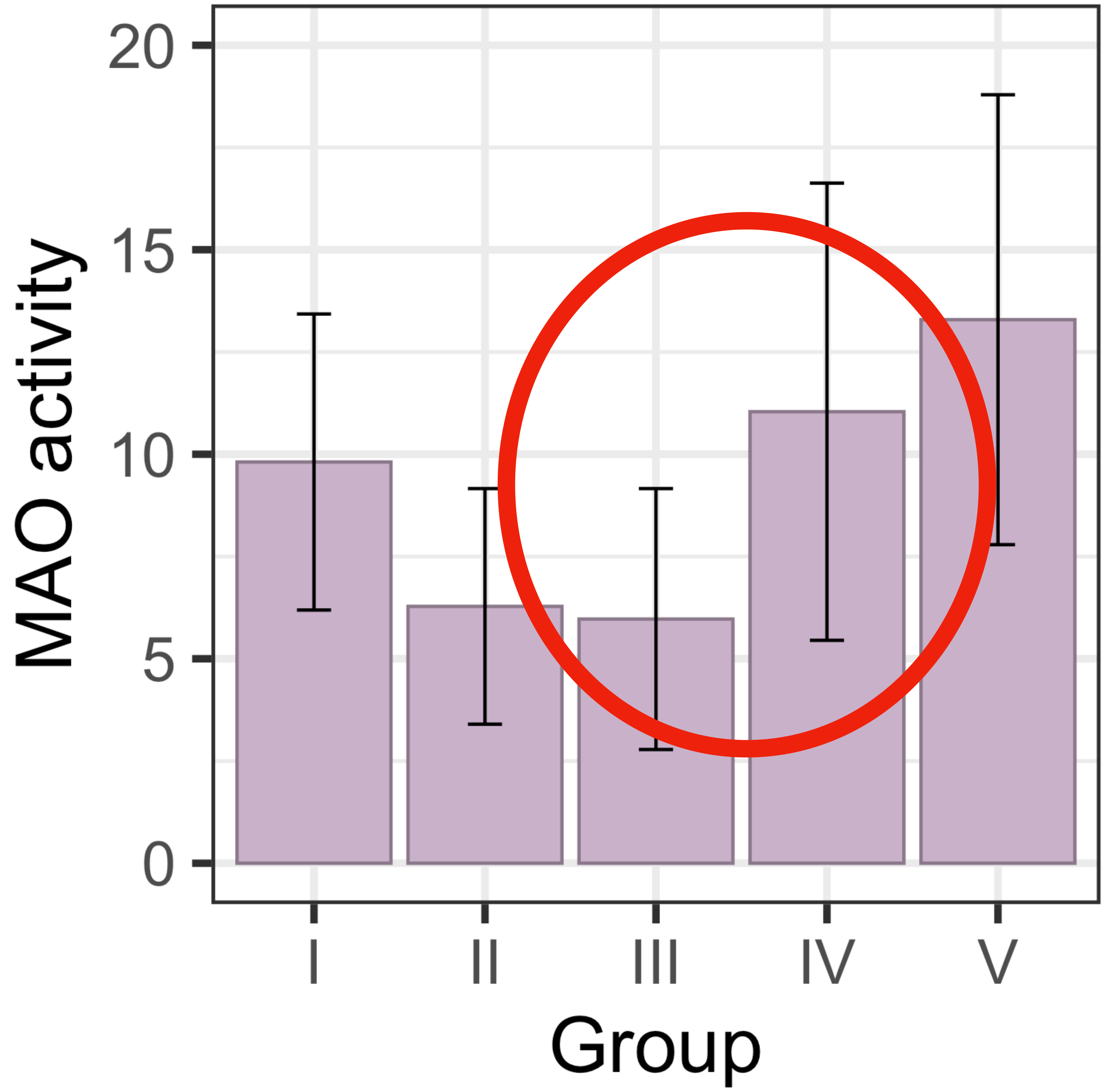
SE = unreliability in the estimate of the population mean



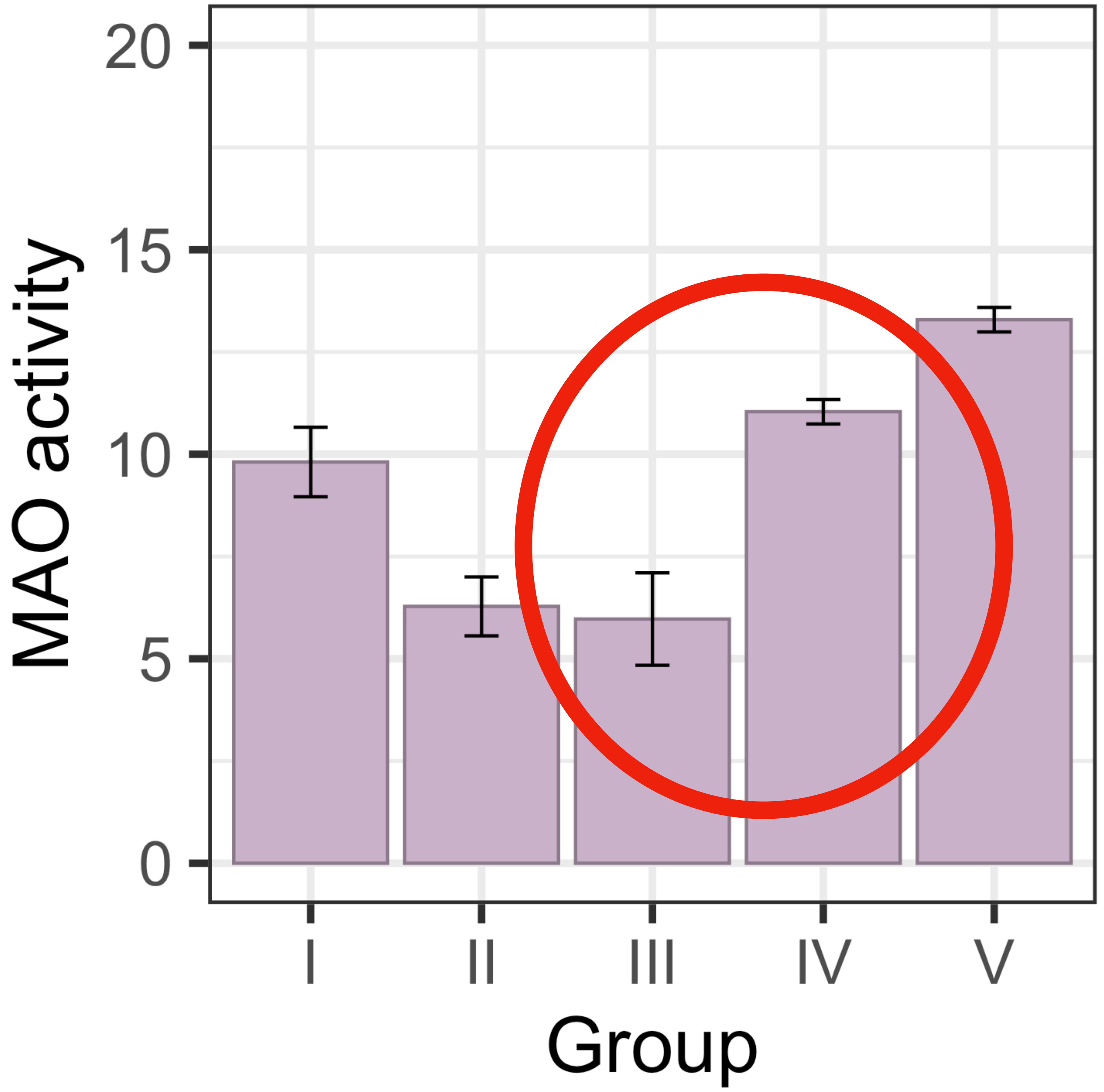
When would we choose to plot SE v. SD?

Do you want to compare means or summarize data variability?

SD = dispersion of data



SE = unreliability in the estimate of the population mean



Example

A pharmacologist measured the concentration of dopamine in the brains of eight rats. The mean concentration was 1,269 ng/gm and the standard deviation was 145 ng/gm. What was the standard error of the mean?

$$SE = \frac{s}{\sqrt{n}}$$

$$SE = \frac{145}{\sqrt{8}} = 51.2$$

Practice

This quantity is a measure of the accuracy of the sample mean as an estimate of the population mean

Standard Error (SE) | Standard Deviation (SD)

This quantity tends to stay the same as the sample size goes up

Standard Error (SE) | Standard Deviation (SD)

This quantity tends to go down as the sample size goes up

Standard Error (SE) | Standard Deviation (SD)

Practice

This quantity is a measure of the accuracy of the sample mean as an estimate of the population mean

Standard Error (SE)

Standard Deviation (SD)

This quantity tends to stay the same as the sample size goes up

Standard Error (SE)

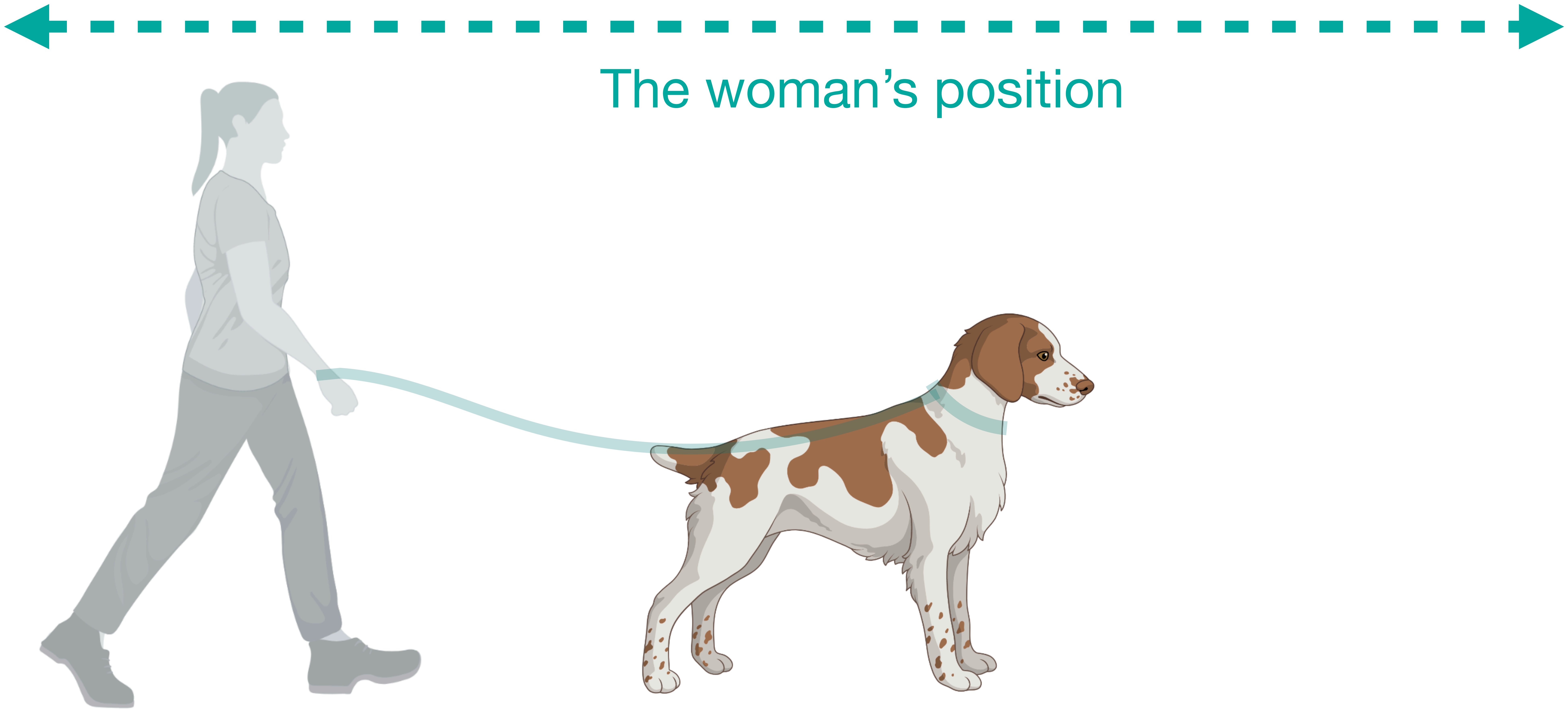
Standard Deviation (SD)

This quantity tends to go down as the sample size goes up

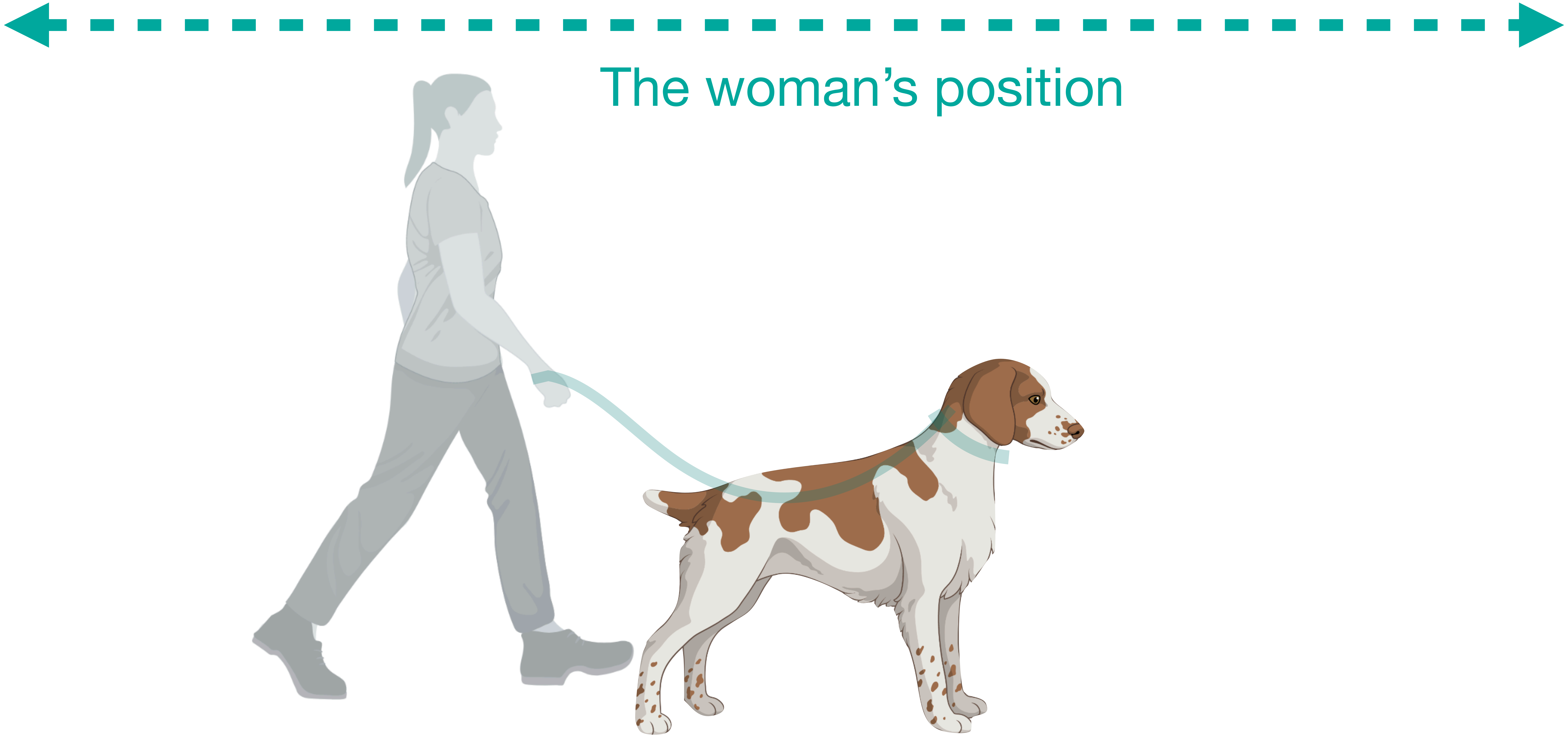
Standard Error (SE)

Standard Deviation (SD)

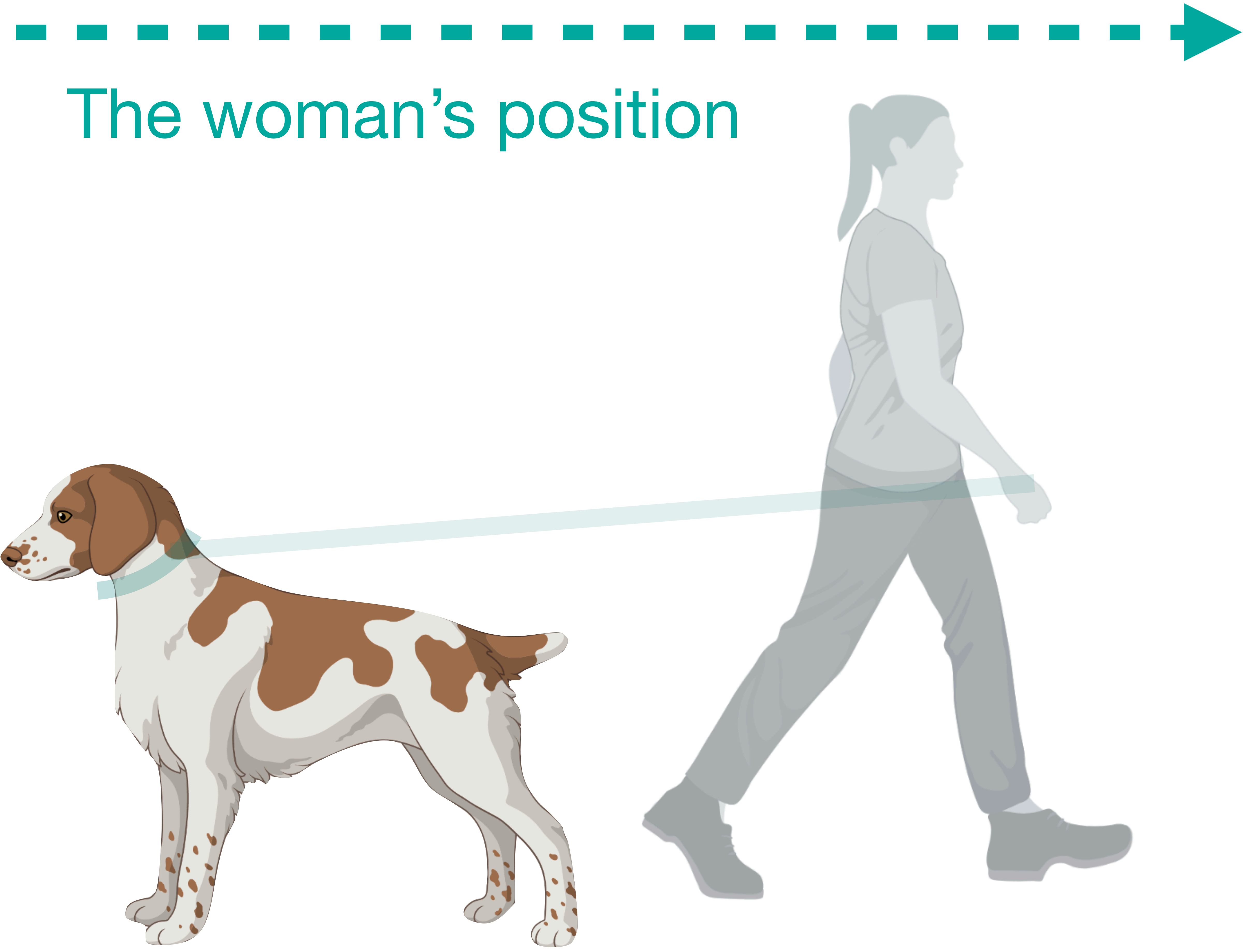
The confidence interval



The confidence interval



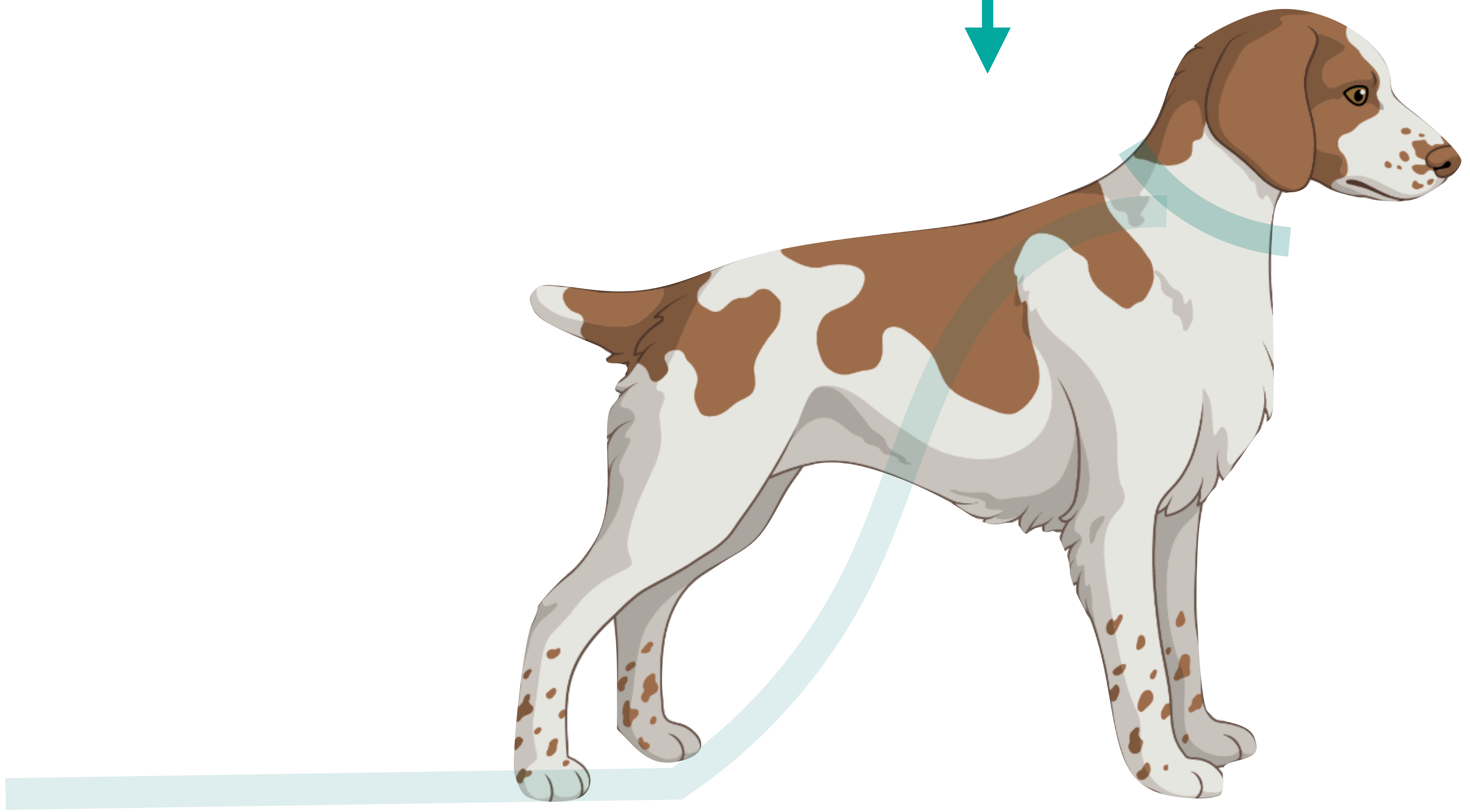
The confidence interval



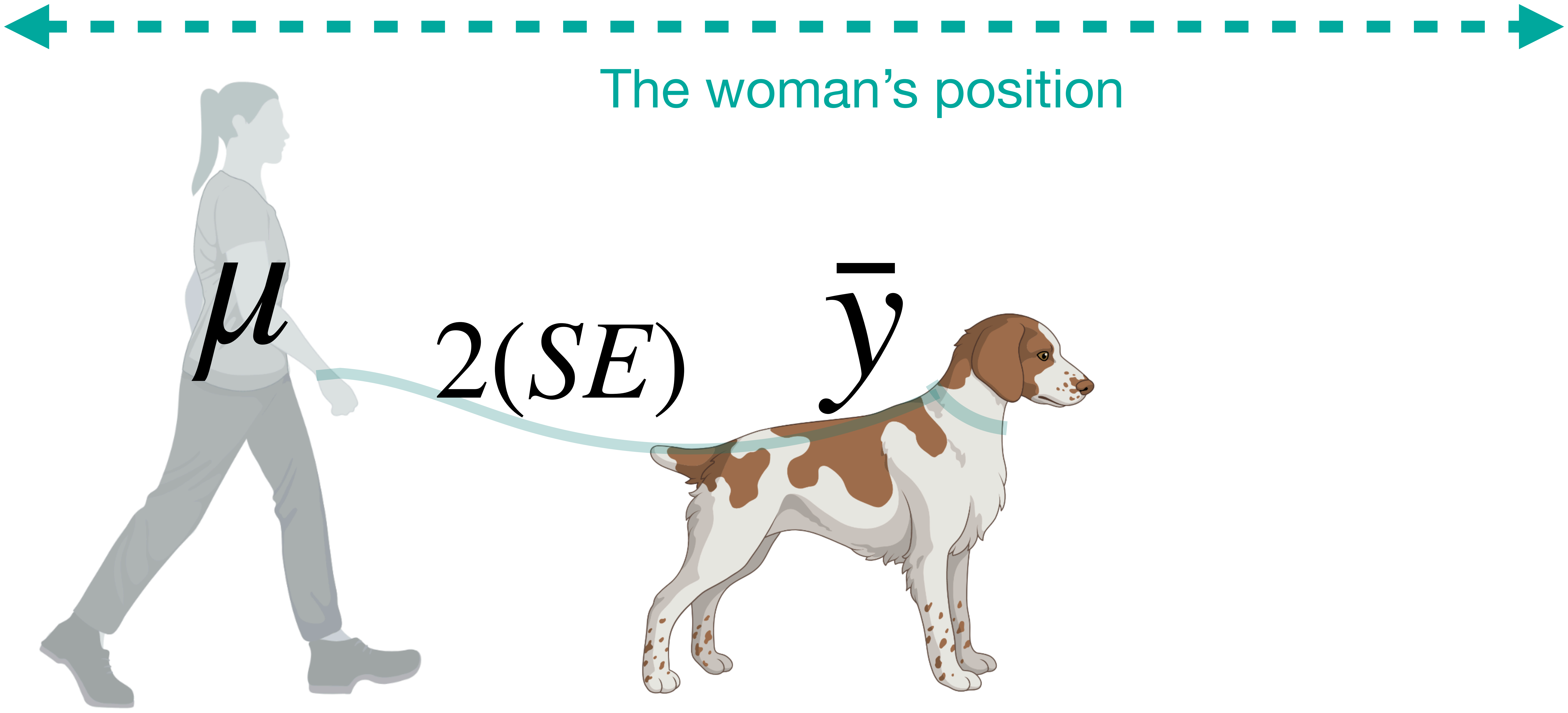
The confidence interval

The woman’s position

Not likely, but possible

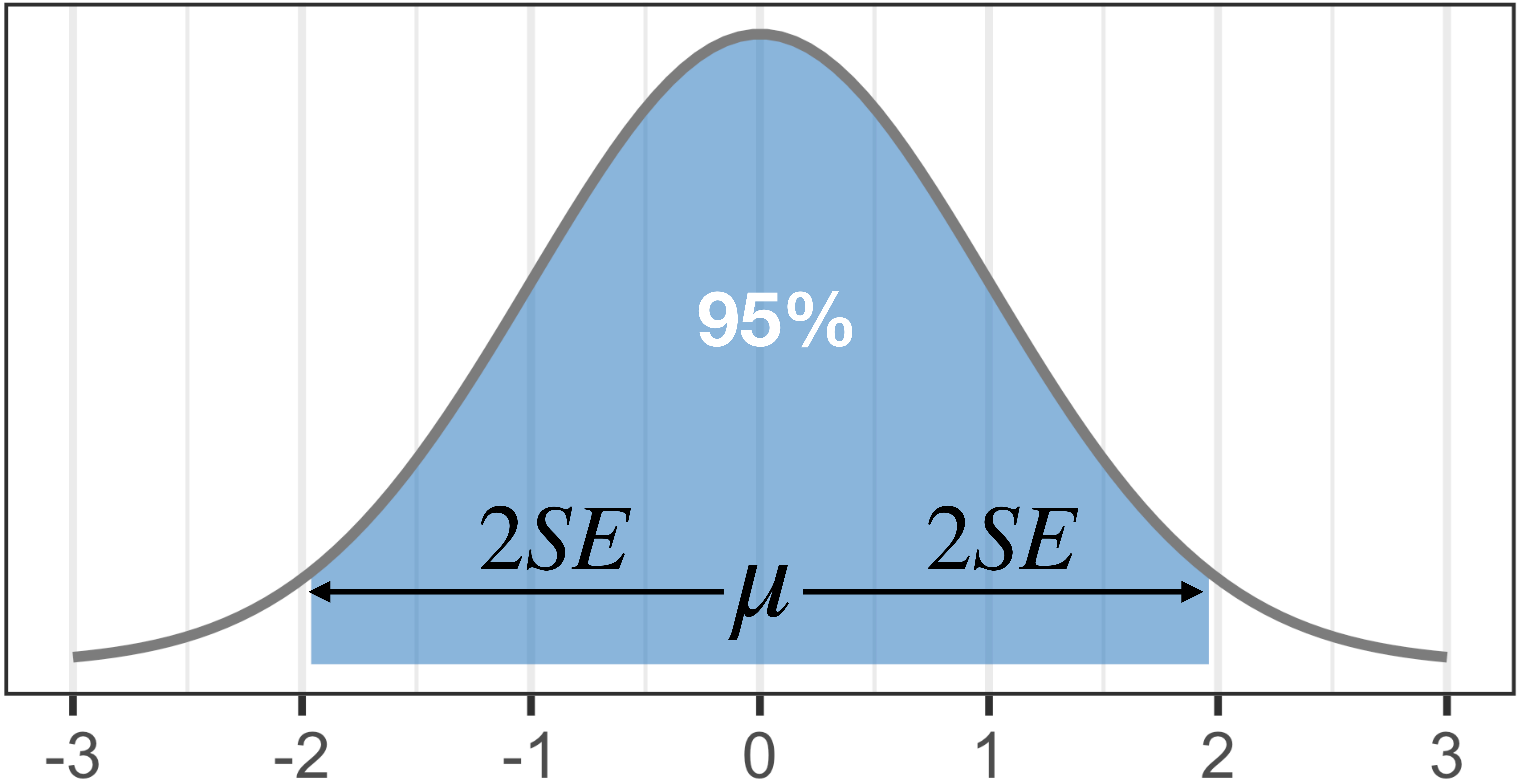


The confidence interval



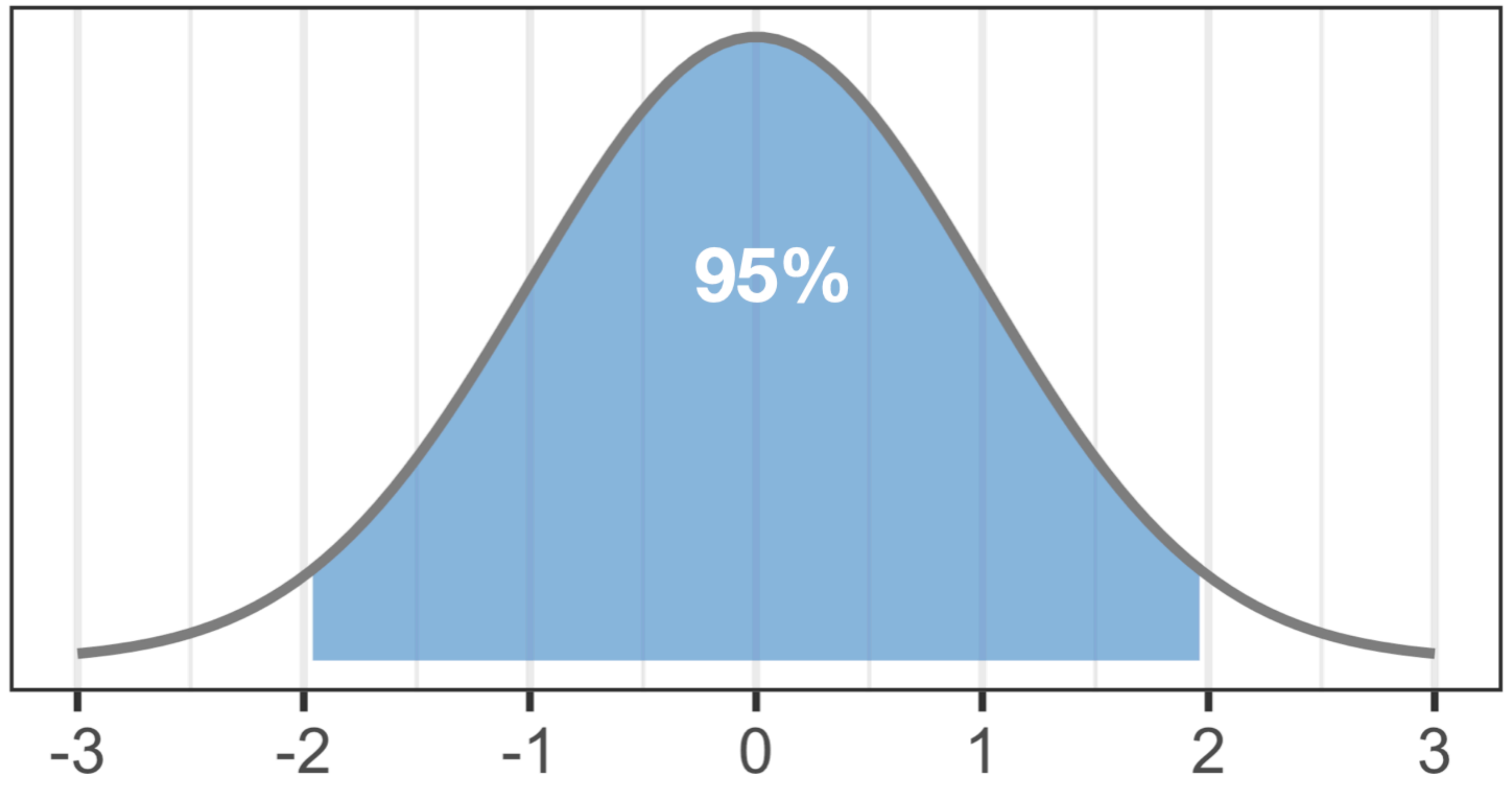
95% confidence interval of woman's position = position of the dog +/- 2 x SE

The confidence interval



Sampling distribution of Y - random sample from normal distribution

The confidence interval



$$Pr[-1.96 < Z < 1.96] = 0.95$$

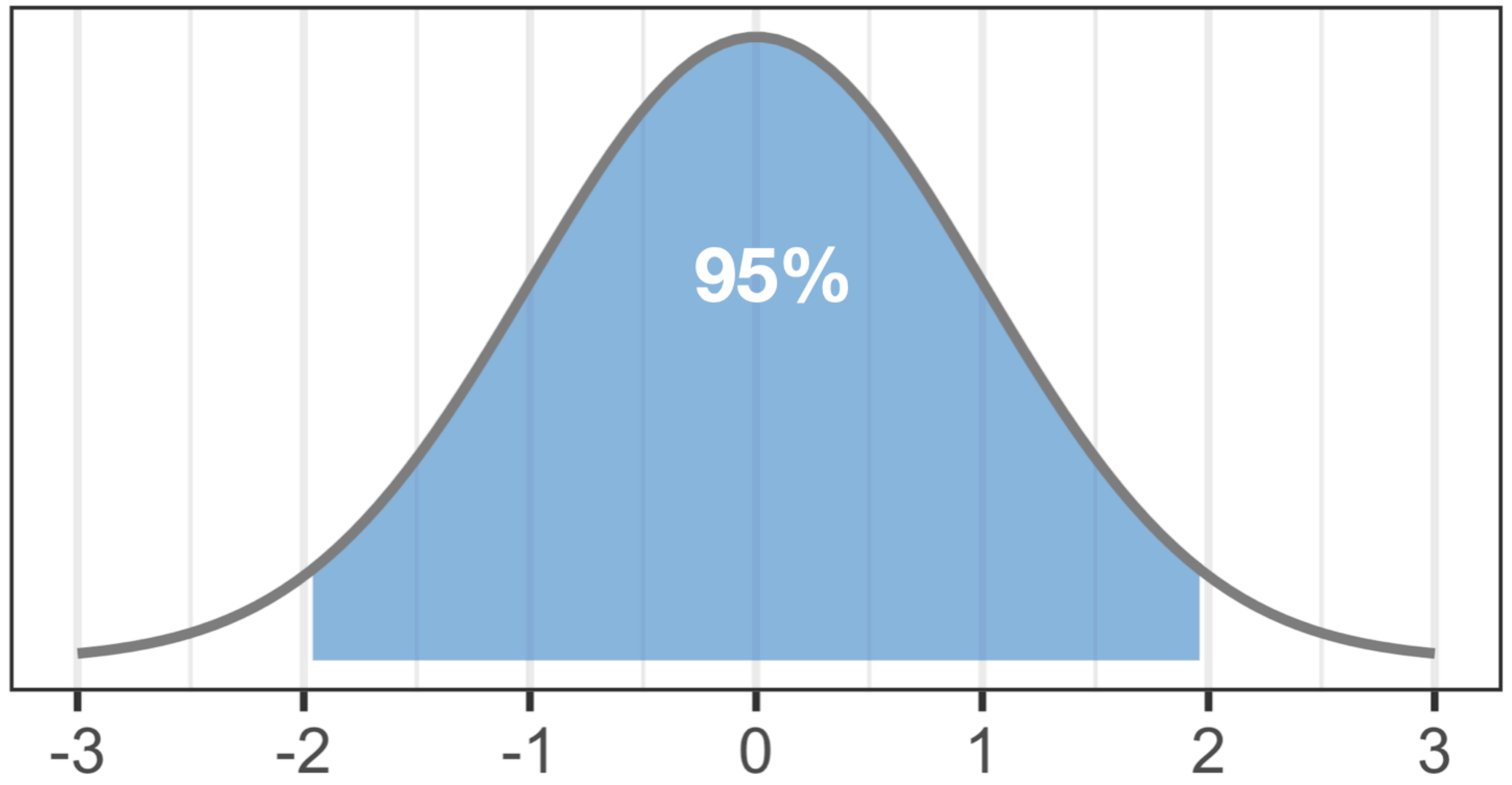
$$\uparrow$$

$$\frac{\bar{Y} - \mu}{\sigma/\sqrt{n}}$$

But… we don't know σ

Will contain μ for 95% of all samples $\longrightarrow$ $\boxed{\bar{Y} \pm 1.96 \frac{\sigma}{\sqrt{n}}}$

Student's t distribution for confidence intervals



$$Pr[-1.96 < Z < 1.96 \;\; = 0.95$$

$$\uparrow$$

$$\frac{\bar{Y}-\mu}{\sigma/\sqrt{n}}$$

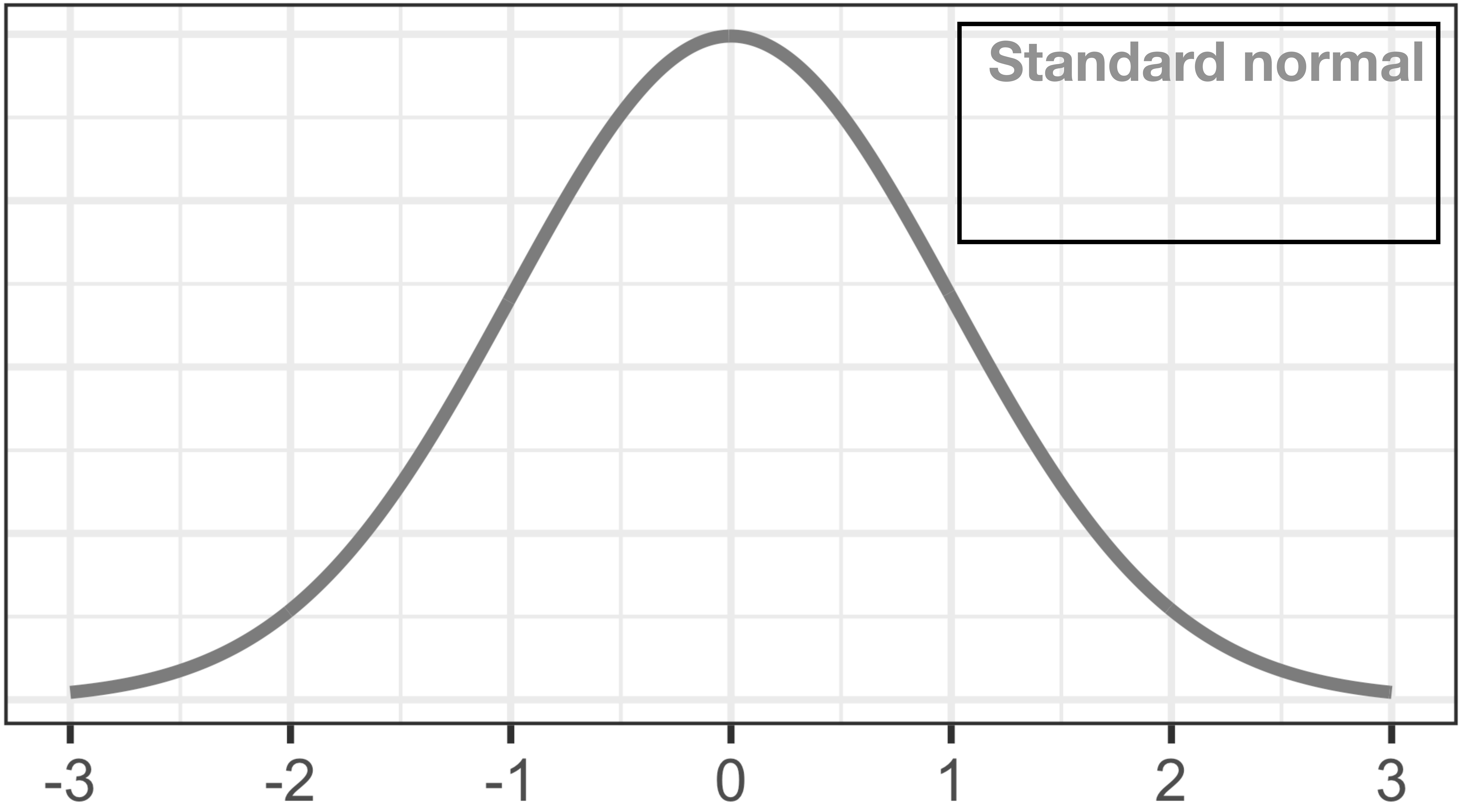
But… we don't know σ

"Critical value"

Will contain μ for 95% of all samples

$$\boxed{\bar{y}\ \cancel{\bar{Y}} \pm \cancel{1.96}\ t_{0.025}\ \frac{\cancel{\sigma}\ s}{\sqrt{n}}}$$

Student's *t* distribution for confidence intervals

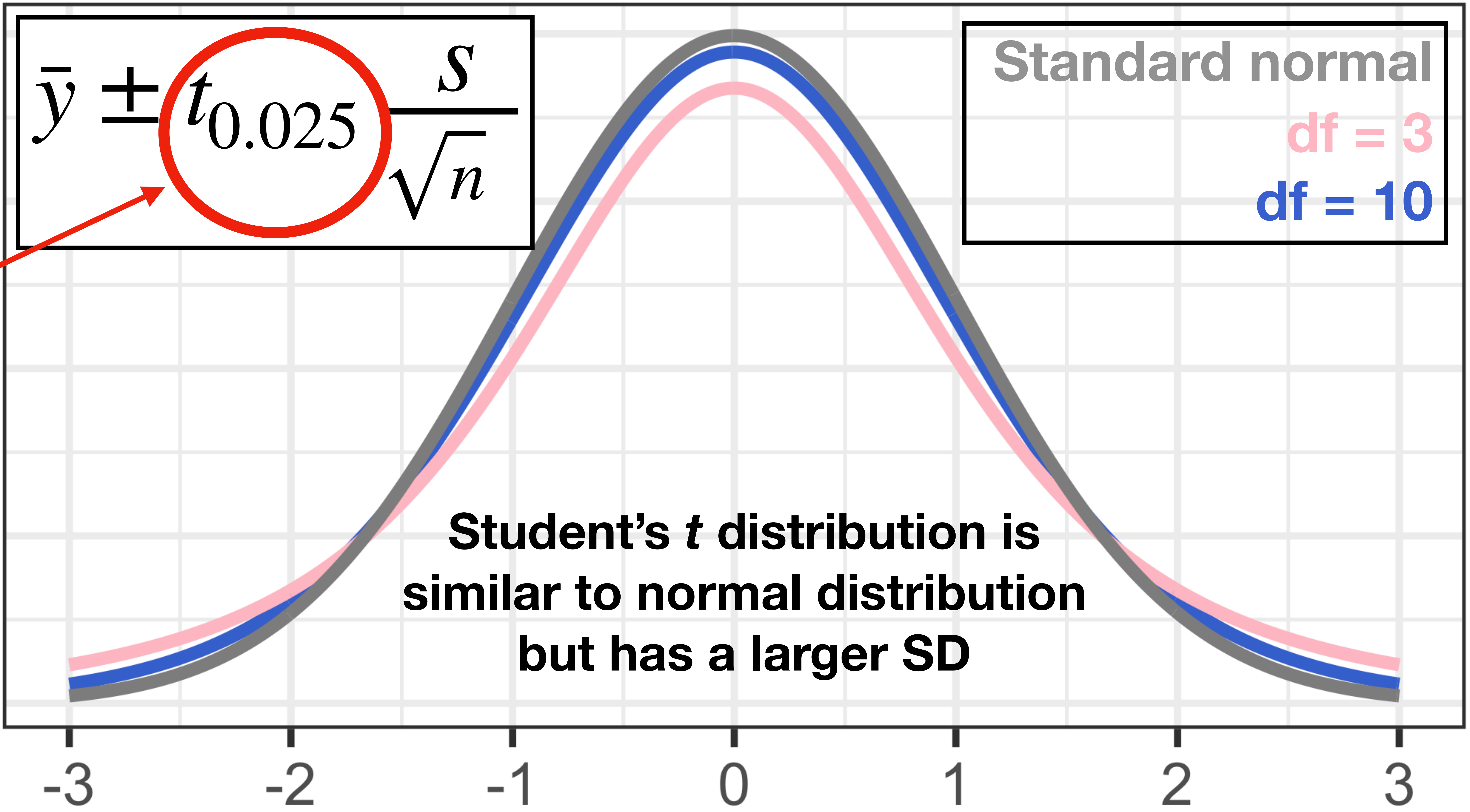


Student's *t* distribution for confidence intervals

Shape of distribution depends on ***degrees of freedom*** → ***(df = n - 1)***

"Critical value"

"Two-tailed 5% critical value"

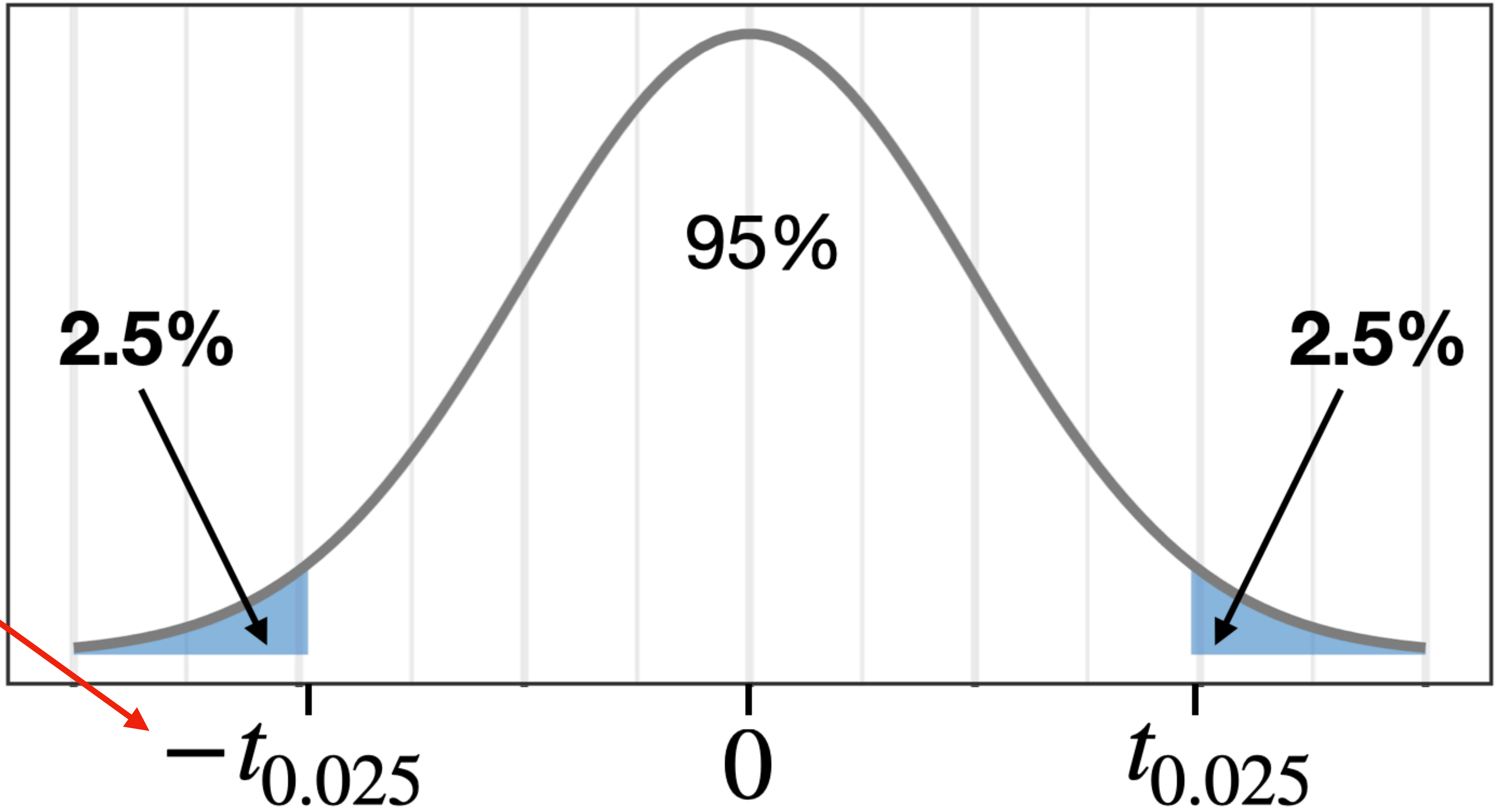


Critical value and Student's *t* distribution

"Two-tailed 5% critical value" = Combined area above t and below -t = 5%

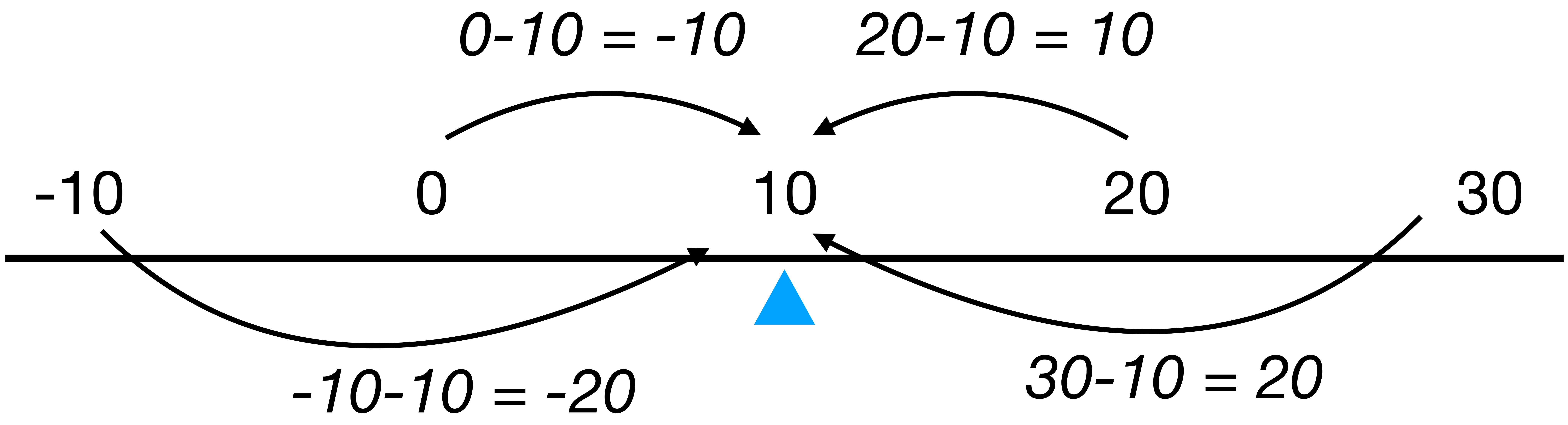
"Two-tailed 5% critical value" = Area between two tails = 95%

"Two-tailed 5% critical value"



	P						
one-tail	0.1	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	0.2	0.1	0.05	0.02	0.01	0.002	0.001
DF							
1	3.078	6.314	12.706	31.821	63.656	318.289	636.578
2	1.886	2.92	4.303	6.965	9.925	22.328	31.6
3	1.638	2.353	3.182	4.541	5.841	10.214	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.61
5	1.476	2.015	2.571	3.365	4.032	5.894	6.869
6	1.44	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.86	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.25	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.93	4.318
13	1.35	1.771	2.16	2.65	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.14
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
120	1.289	1.658	1.98	2.358	2.617	3.16	3.373
1000	1.282	1.646	1.962	2.33	2.581	3.098	3.3
Inf	1.282	1.645	1.96	2.326	2.576	3.091	3.291

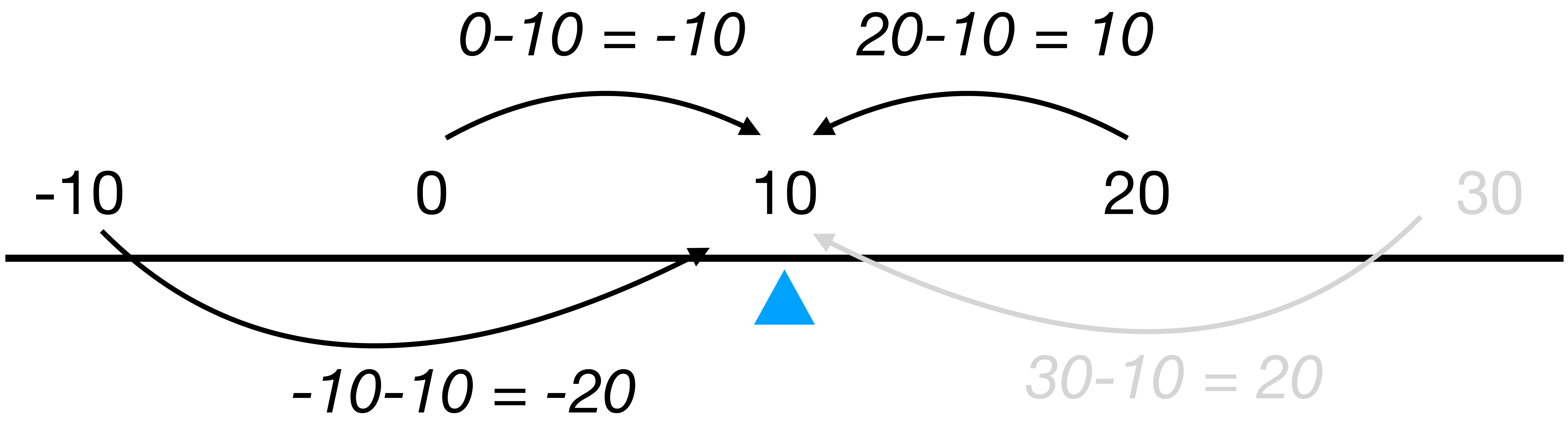
N-1: **degrees of freedom** explained



(-20) + (-10) + (10) + (20) = 0

Sum of deviations is always zero!

N-1: **degrees of freedom** explained



Has to be +20 from mean…

$$(-20) + (-10) + (10) + X = 0$$

Sum of deviations is always zero!

Calculating the confidence interval: butterflies

$$\xrightarrow[\text{df} = 13]{n = 14}$$

$$\bar{y} = 32.81cm^2$$

$$s = 2.48cm^2$$

$$\bar{y} \pm t_{0.025} \frac{s}{\sqrt{n}}$$

	P						
one-tail	0.1	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	0.2	0.1	0.05	0.02	0.01	0.002	0.001
DF							
1	3.078	6.314	12.706	31.821	63.656	318.289	636.578
2	1.886	2.92	4.303	6.965	9.925	22.328	31.6
3	1.638	2.353	3.182	4.541	5.841	10.214	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.61
5	1.476	2.015	2.571	3.365	4.032	5.894	6.869
6	1.44	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.86	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.25	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.93	4.318
13	1.35	1.771	2.16	2.65	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.14
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
120	1.289	1.658	1.98	2.358	2.617	3.16	3.373
1000	1.282	1.646	1.962	2.33	2.581	3.098	3.3
Inf	1.282	1.645	1.96	2.326	2.576	3.091	3.291

Calculating the confidence interval: butterflies

$n = 14$

df = 13

$\bar{y} = 32.81cm^2$

$s = 2.48cm^2$

Critical value

```
qt(p, df)

qt(0.975, 13)
```

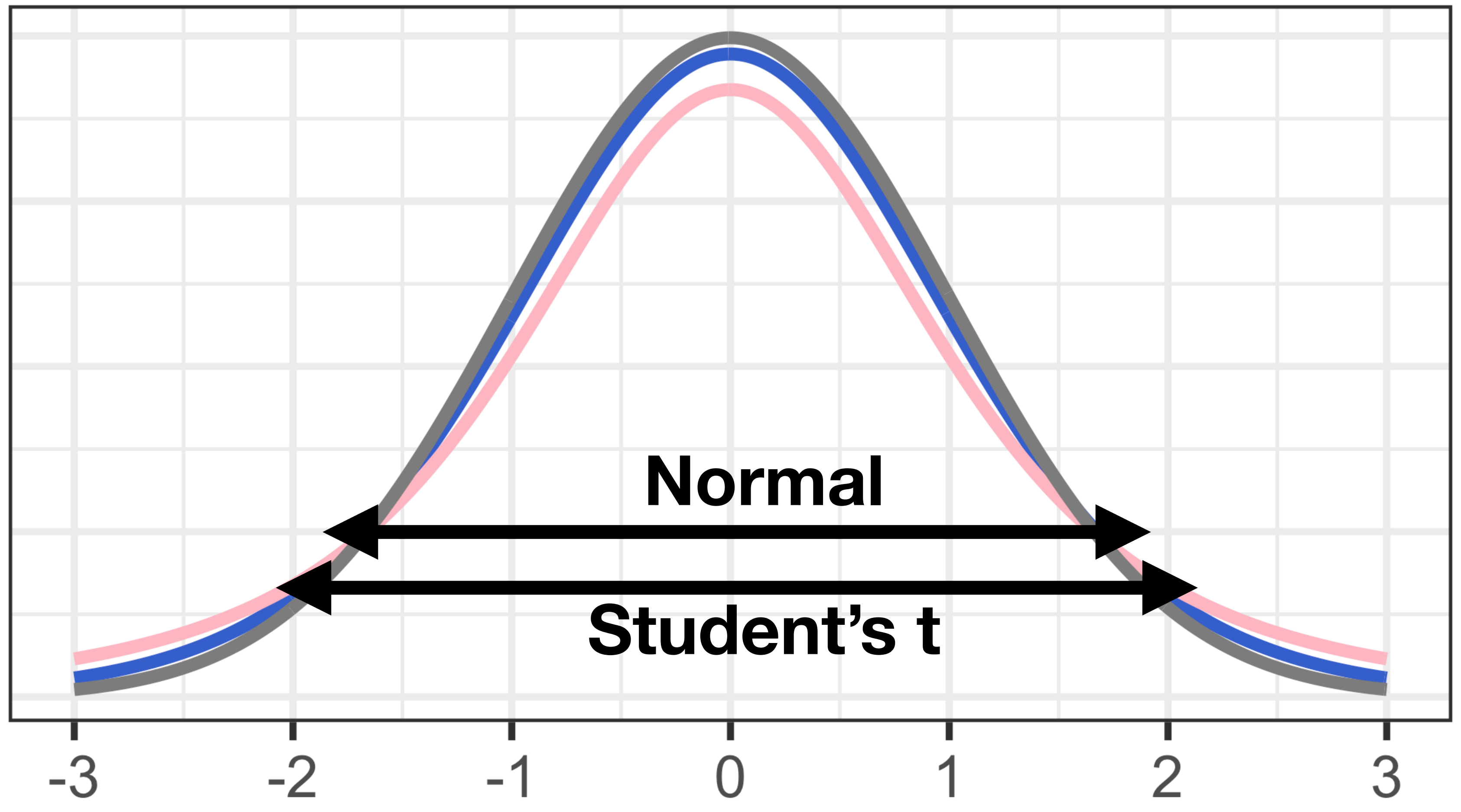
95%

$$\bar{y} \pm t_{0.025} \frac{s}{\sqrt{n}}$$

$$32.81 \pm 2.16 \frac{2.48}{\sqrt{14}}$$

$$32.81 \pm 1.43 \qquad (31.4, 34.2)$$

Note: why use Student's *t* distribution?



***n* = 14**

df = 13

$\bar{y} = 32.81cm^2$

$s = 2.48cm^2$

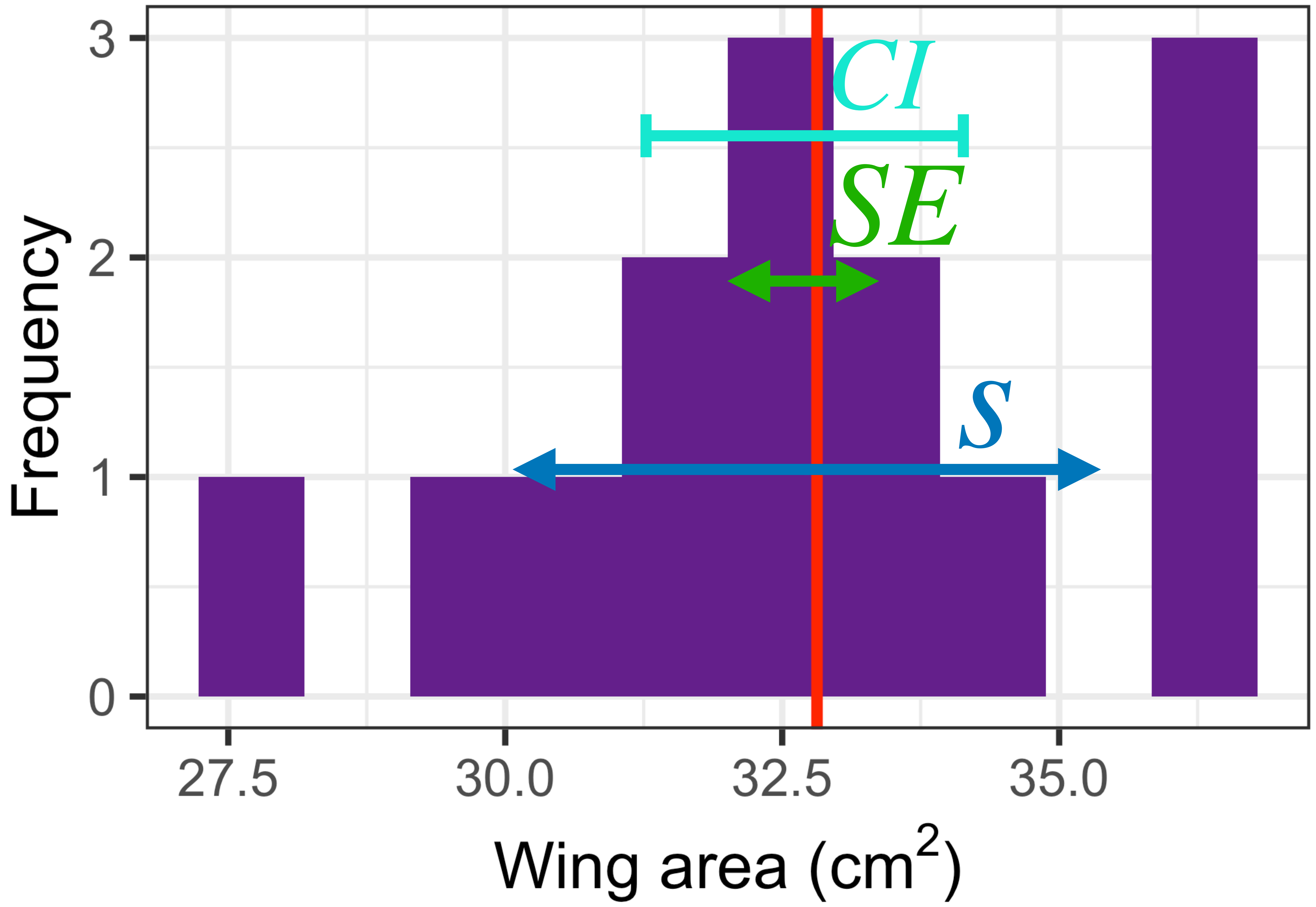
95%

$$\bar{y} \pm z_{0.025} \frac{s}{\sqrt{n}}$$

$$32.81 \pm 1.96 \frac{2.48}{\sqrt{14}}$$

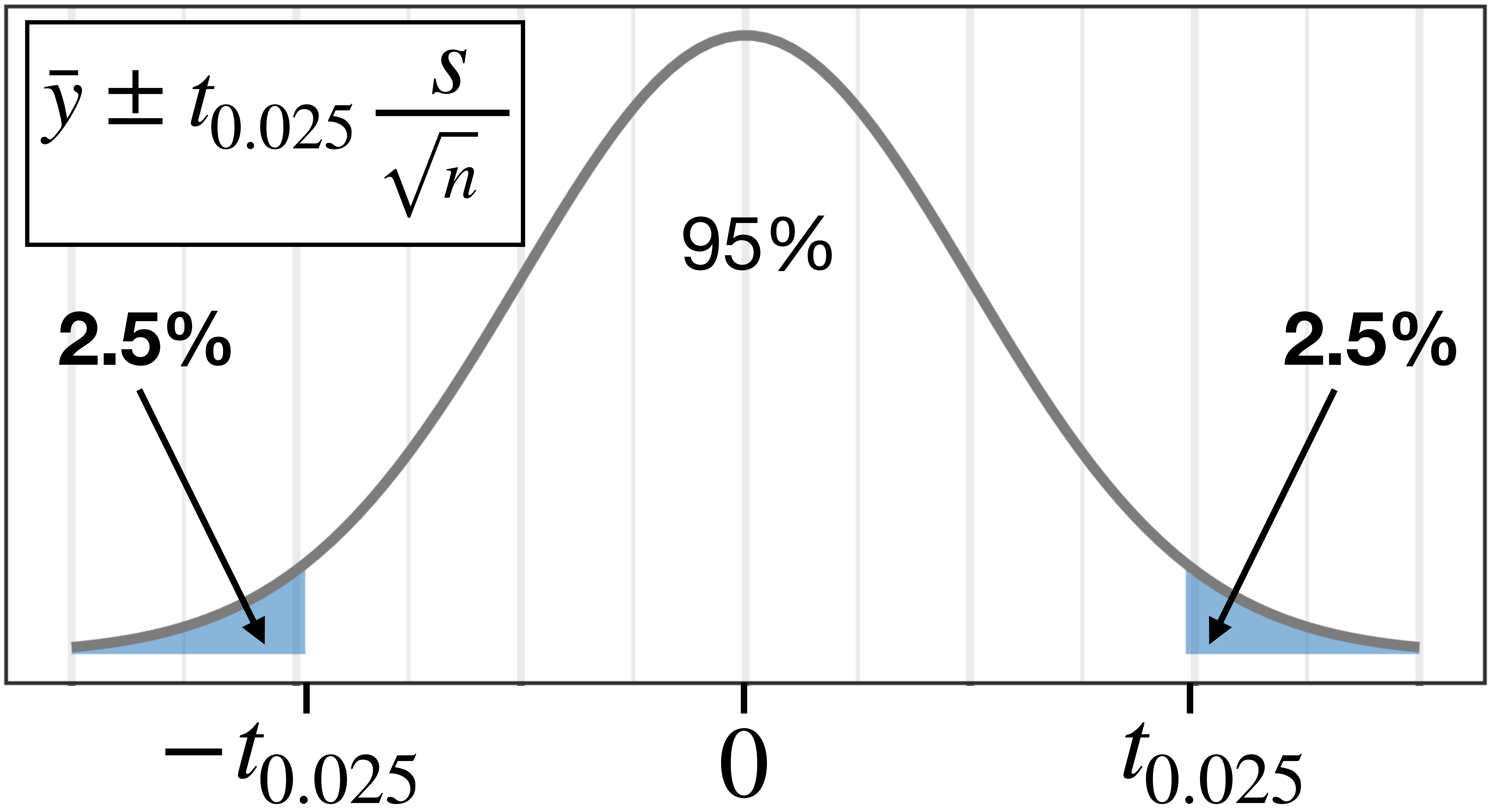
$$32.81 \pm 1.29 \qquad (31.51, 34.11)$$

Calculating the confidence interval: butterflies

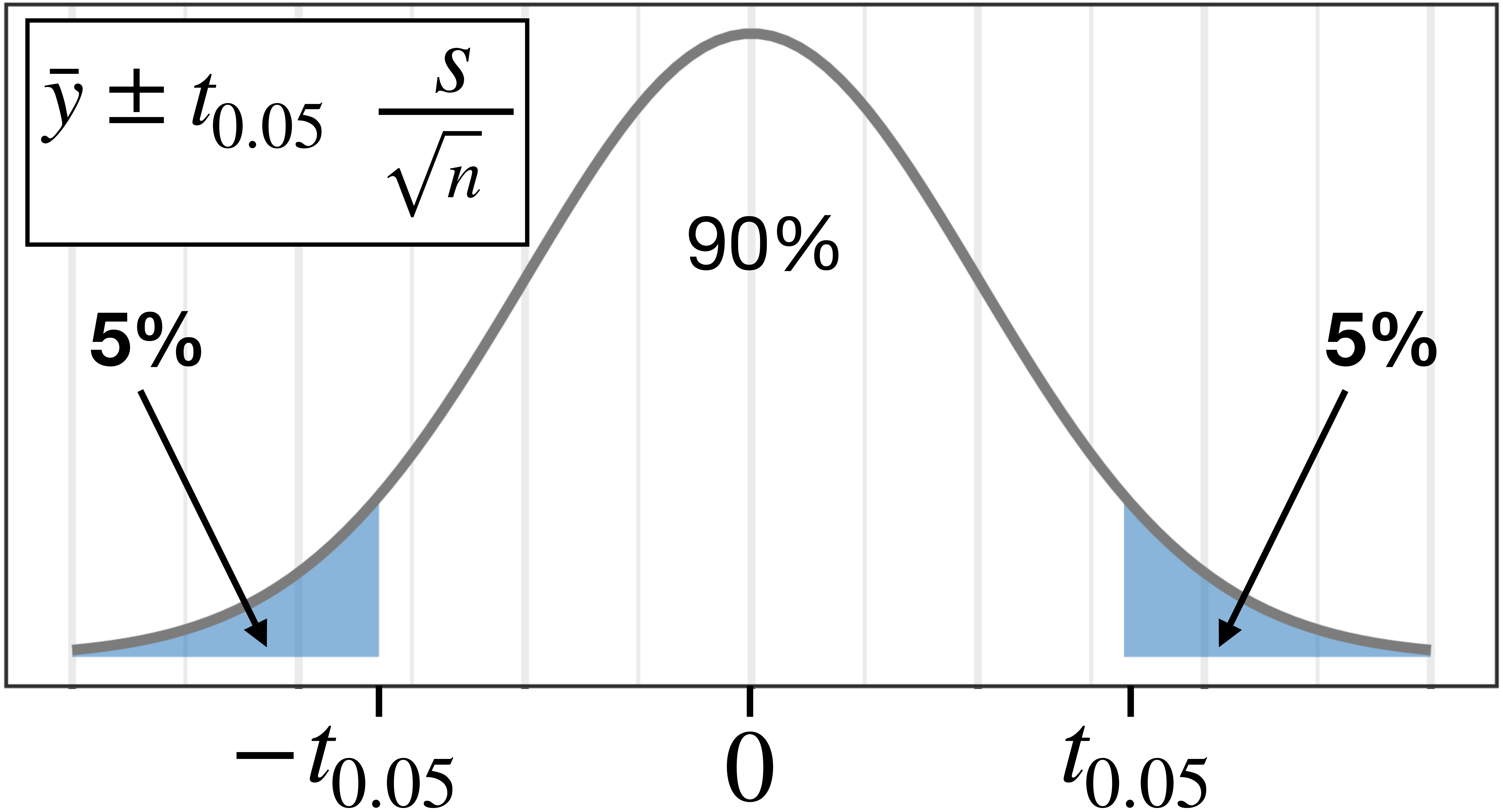


32.81 ± 1.43

Critical value and Student's *t* distribution



Critical value and Student's *t* distribution



Calculating the confidence interval: butterflies

n **= 14**

$\longrightarrow$

df = 13

$\bar{y} = 32.81cm^2$

$s = 2.48cm^2$

90%

$$\bar{y} \pm t_{0.05} \frac{s}{\sqrt{n}}$$

	P						
one-tail	0.1	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	0.2	0.1	0.05	0.02	0.01	0.002	0.001
DF							
1	3.078	6.314	12.706	31.821	63.656	318.289	636.578
2	1.886	2.92	4.303	6.965	9.925	22.328	31.6
3	1.638	2.353	3.182	4.541	5.841	10.214	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.61
5	1.476	2.015	2.571	3.365	4.032	5.894	6.869
6	1.44	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.86	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.25	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.93	4.318
13	1.35	1.771	2.16	2.65	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.14
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
120	1.289	1.658	1.98	2.358	2.617	3.16	3.373
1000	1.282	1.646	1.962	2.33	2.581	3.098	3.3
Inf	1.282	1.645	1.96	2.326	2.576	3.091	3.291

Calculating the confidence interval: butterflies

$$\xrightarrow[\text{df} = 13]{n = 14}$$

$\bar{y} = 32.81cm^2$

$s = 2.48cm^2$

90%

$$\bar{y} \pm t_{0.05} \frac{s}{\sqrt{n}}$$

$$32.81 \pm 1.77 \frac{2.48}{\sqrt{14}}$$

32.81 ± 1.17 $(31.6,34.0)$

The higher the confidence level, the wider the confidence interval

Example

A pharmacologist measured the concentration of dopamine in the brains of eight rats. The mean concentration was 1,269 ng/gm and the standard deviation was 145 ng/gm. Construct a 95% confidence interval for the population mean.

$$SE = \frac{s}{\sqrt{n}}$$

$$SE = \frac{145}{\sqrt{8}} = 51.2$$

$$\bar{y} \pm t_{0.025}\frac{s}{\sqrt{n}}$$

(df = n - 1 = 7)

$$1269 \pm (t_{0.025})(51.2)$$

	P						
one-tail	0.1	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	0.2	0.1	0.05	0.02	0.01	0.002	0.001
DF							
1	3.078	6.314	12.706	31.821	63.656	318.289	636.578
2	1.886	2.92	4.303	6.965	9.925	22.328	31.6
3	1.638	2.353	3.182	4.541	5.841	10.214	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.61
5	1.476	2.015	2.571	3.365	4.032	5.894	6.869
6	1.44	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.86	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.25	4.297	4.781
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	3.93	4.318
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.14
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
120	1.289	1.658	1.98	2.358	2.617	3.16	3.373
1000	1.282	1.646	1.962	2.33	2.581	3.098	3.3
Inf	1.282	1.645	1.96	2.326	2.576	3.091	3.291

```
qt(p = 0.025, df = 7, lower.tail = F)
```

```
qt(p = 0.975, df = 7, lower.tail = T)
```

Example

A pharmacologist measured the concentration of dopamine in the brains of eight rats. The mean concentration was 1,269 ng/gm and the standard deviation was 145 ng/gm. Construct a 95% confidence interval for the population mean.

$$SE = \frac{s}{\sqrt{n}}$$

$$SE = \frac{145}{\sqrt{8}} = 51.2$$

$$\bar{y} \pm t_{0.025} \frac{s}{\sqrt{n}}$$

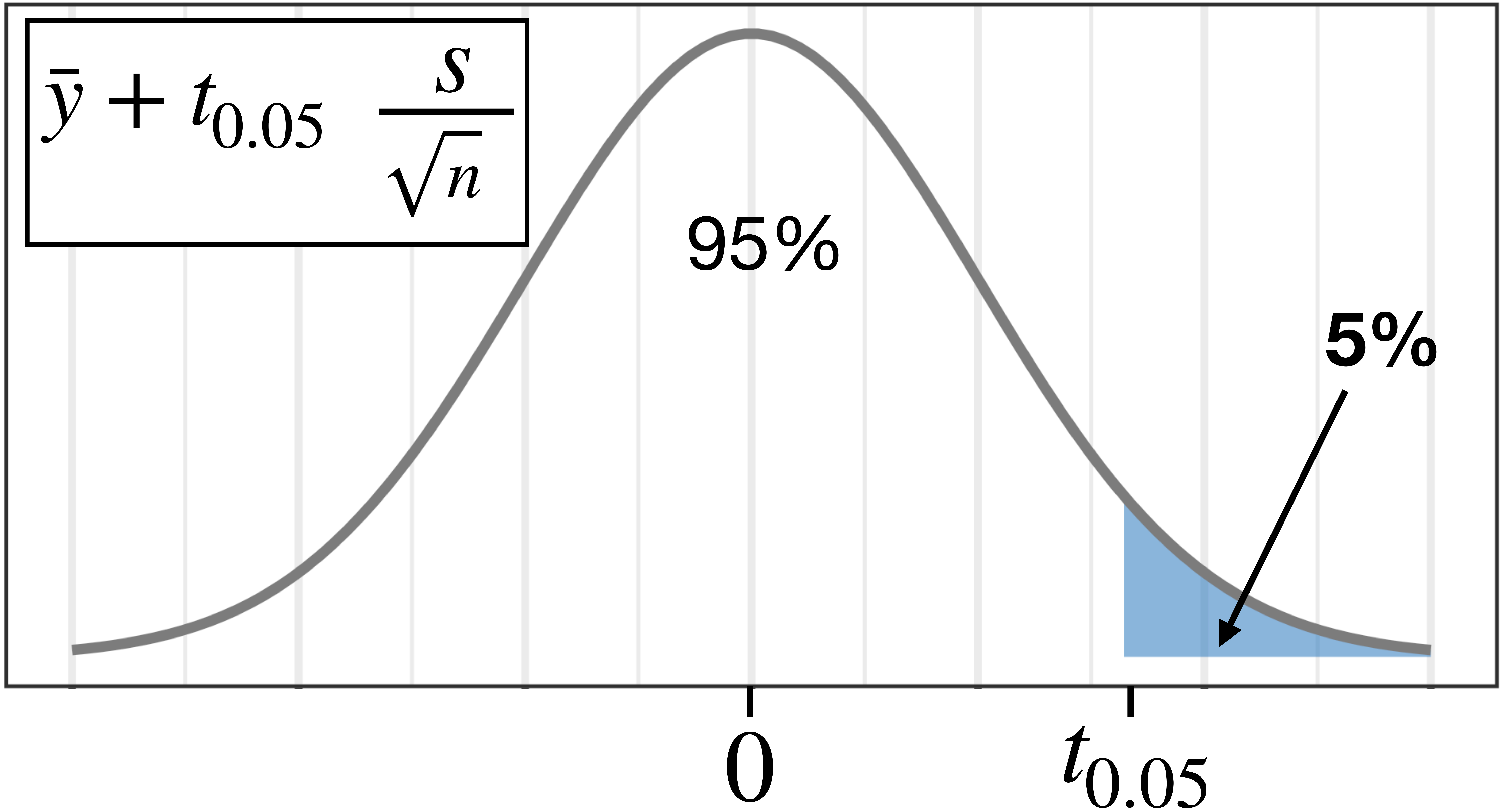
(df = n - 1 = 7)

$$1269 \pm (2.365)(51.2)$$

$$1269 \pm 121.1$$

$$(1147.9, 1390.1)$$

One-sided confidence intervals



One-sided confidence intervals

A pharmacologist measured the concentration of dopamine in the brains of eight rats. The mean concentration was 1,269 ng/gm and the standard deviation was 145 ng/gm. Construct a one-sided, upper-bound 95% confidence interval for the population mean.

$$\bar{y} + t_{0.05} \frac{s}{\sqrt{n}}$$

(df = n - 1 = 7)

	P						
one-tail	0.1	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	0.2	0.1	0.05	0.02	0.01	0.002	0.001
DF							
1	3.078	6.314	12.706	31.821	63.656	318.289	636.578
2	1.886	2.92	4.303	6.965	9.925	22.328	31.6
3	1.638	2.353	3.182	4.541	5.841	10.214	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.61
5	1.476	2.015	2.571	3.365	4.032	5.894	6.869
6	1.44	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.86	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.25	4.297	4.781
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	3.93	4.318
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.14
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
120	1.289	1.658	1.98	2.358	2.617	3.16	3.373
1000	1.282	1.646	1.962	2.33	2.581	3.098	3.3
Inf	1.282	1.645	1.96	2.326	2.576	3.091	3.291

```
qt(p = 0.05, df = 7, lower.tail = F)
qt(p = 0.95, df = 7, lower.tail = T)
```

One-sided confidence intervals

A pharmacologist measured the concentration of dopamine in the brains of eight rats. The mean concentration was 1,269 ng/gm and the standard deviation was 145 ng/gm. Construct a one-sided, upper-bound 95% confidence interval for the population mean.

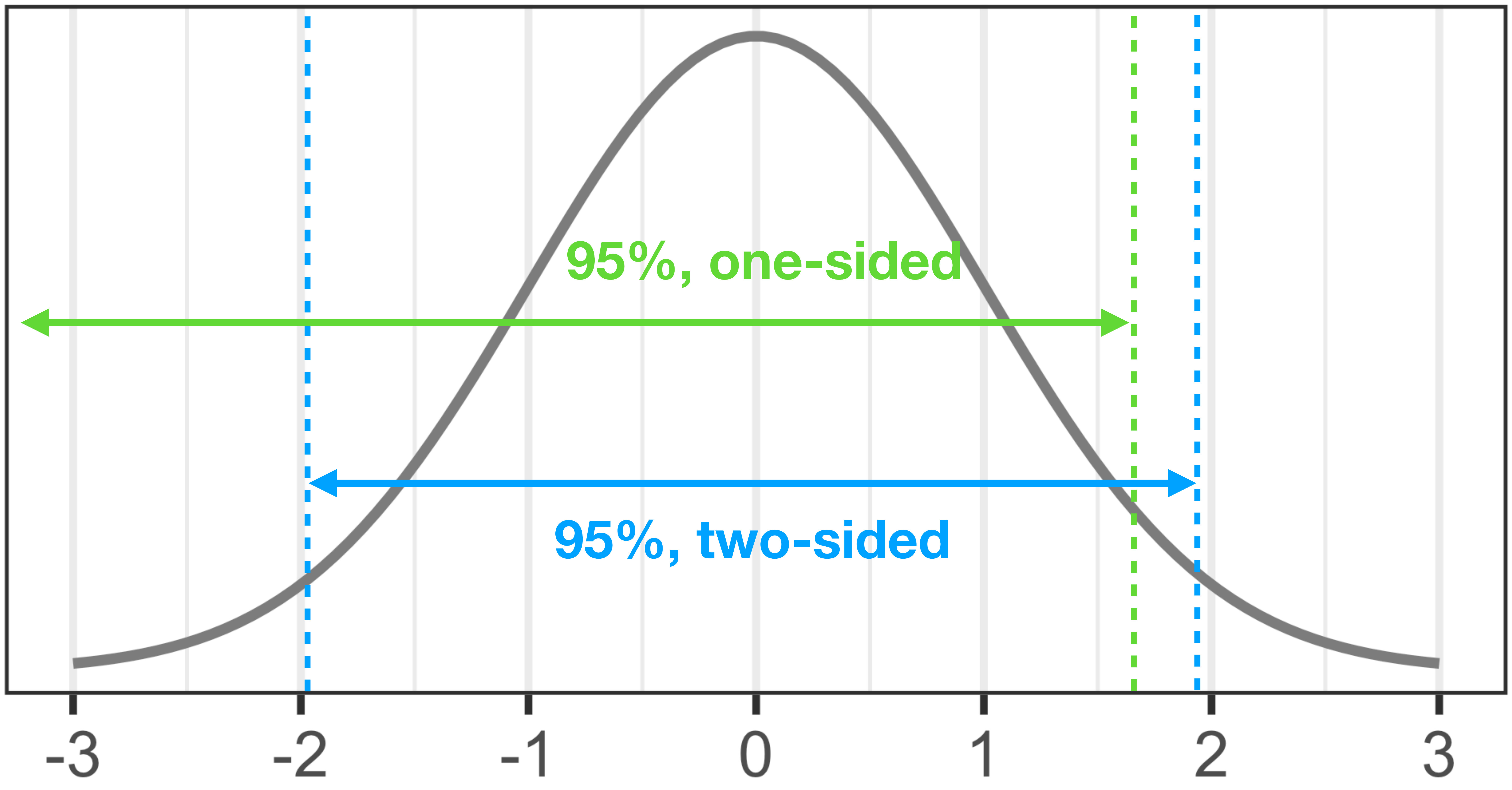
$$\bar{y} + t_{0.05} \frac{s}{\sqrt{n}}$$

(df = n - 1 = 7)

$$1269 + (1.895)(51.2)$$

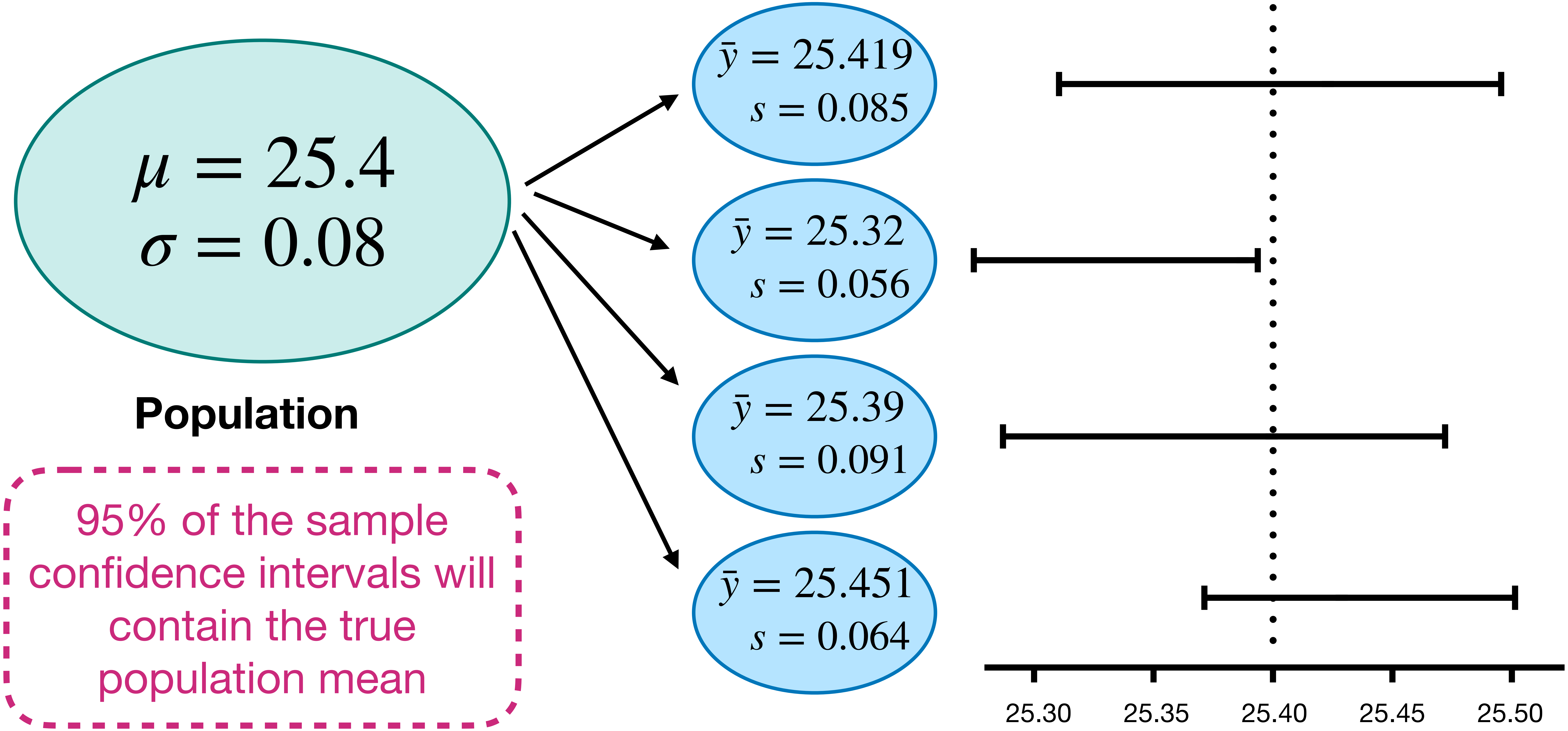
$$(\infty, 1366.0)$$

One-sided confidence intervals

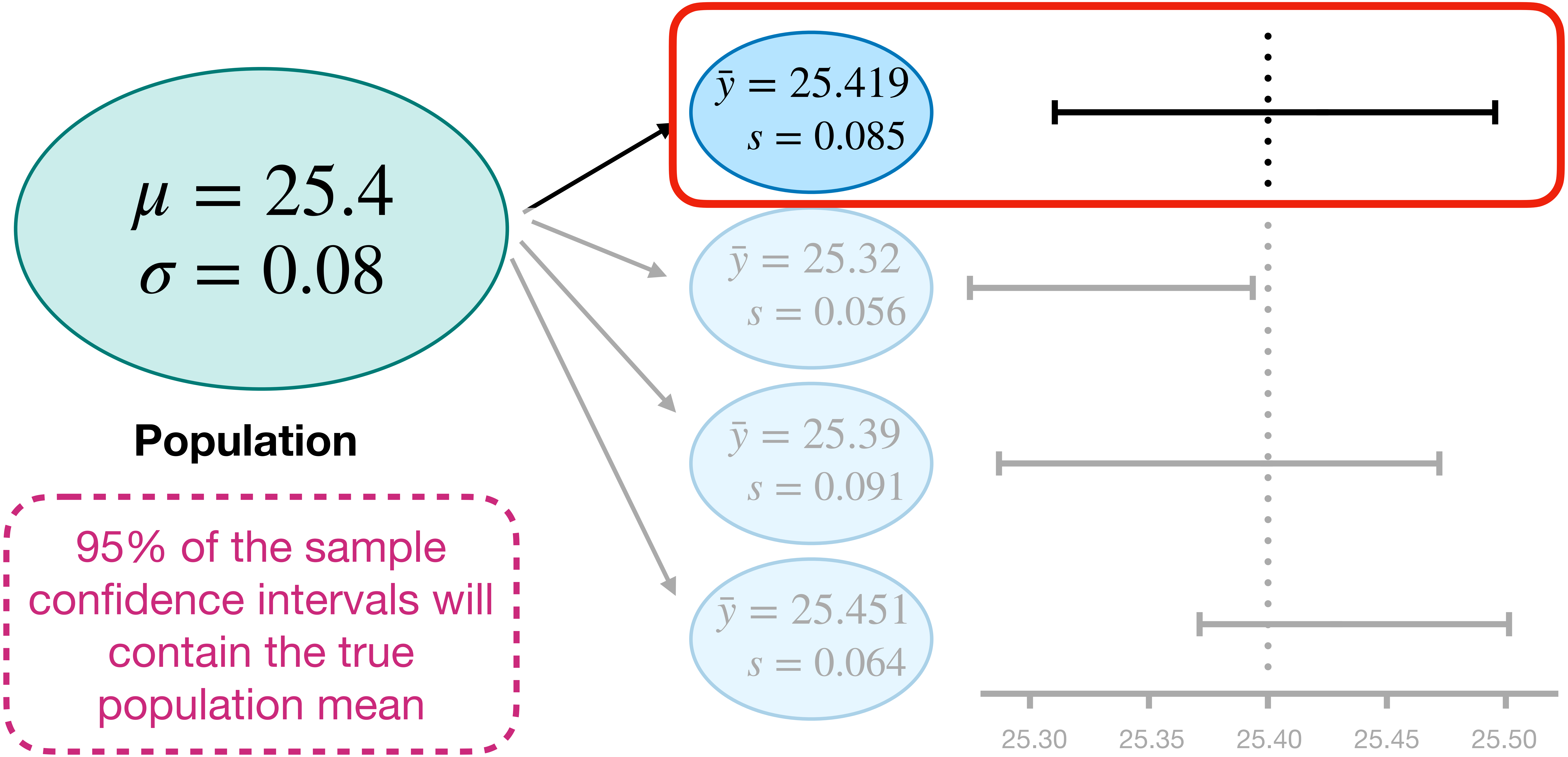


$(-\infty, 1366.0)$ $(1147.9, 1390.1)$

Confidence intervals and randomness



Confidence intervals and randomness

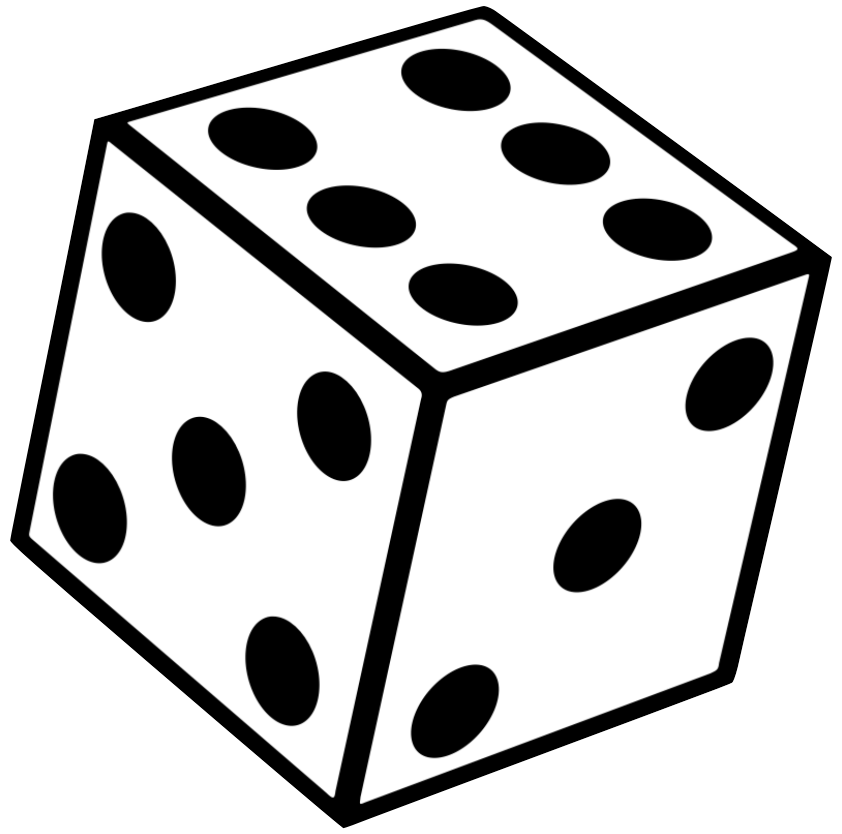


95% of the sample confidence intervals will contain the true population mean

Confidence intervals and randomness

- **Larger samples** produce **narrower** confidence intervals
    - *Because SE is smaller (divide by square root of n)*
- A **confidence interval** can be interpreted as a probability… with caution!

Confidence intervals and randomness



$$Pr[Y = 2] = \frac{1}{6}$$

Pr{a sample will give us a CI that contains the true mean} = 0.95

$$Y = 5$$

$$Pr[5 = 2] \neq \frac{1}{6}$$

Pr{the true mean is within our CI} = 0.95

Pr{31 < μ < 34} $\neq$ 0.95

Confidence intervals and randomness

- **Larger samples** produce **narrower** confidence intervals
    - *Because SE is smaller (divide by square root of n)*
- A **confidence interval** can be interpreted as a probability… with caution!
    - Pr{a sample will give us a CI that contains the true mean} = 0.95
    - Pr{the true mean is within our CI} = 0.95
    - *An individual statement can be TRUE or FALSE, but if you create numerous statements, one statement will be TRUE 95% of the time*
- **"We are 95% confident that the true mean is between X and X"**

A pharmacologist measured the concentration of dopamine in the brains of eight rats. The mean concentration was 1,269 ng/gm and the standard deviation was 145 ng/gm. Construct a 95% confidence interval for the population mean.

$$SE = \frac{s}{\sqrt{n}}$$

$$SE = \frac{145}{\sqrt{8}} = 51.2$$

$$\bar{y} \pm t_{0.025}\frac{s}{\sqrt{n}}$$

(df = n - 1 = 7)

$$1269 \pm (2.365)(51.2)$$

$$1269 \pm 121.1$$

$$(1147.9, 1390.1)$$

We are 95% confident that the mean concentration of dopamine of all rats is between 1,147.9 and 1,390.1 ng/gm

Planning a study with sufficient precision

There is no substitute for bad data.

- **(1) Population variability of observed variable**
    - Sometimes you cannot (or should not) reduce this, you want a random sample of the <u>entire</u> population
    - However, reducing environmental variation and fixing the variables you <u>can</u> will lead to the cleanest data

Planning a study with sufficient precision

There is no substitute for bad data.

- **(1) Population variability of observed variable**
  - Sometimes you cannot (or should not) reduce this, you want a random sample of the entire population
  - However, reducing environmental variation and fixing the variables you can will lead to the cleanest data
- **(2) Sample size**
  - You can try to guess-calculate how large of a sample you need to prove/disprove your hypothesis
  - $SE = s/\sqrt{n}$

Planning a study with sufficient precision

$$\xrightarrow{\boldsymbol{n} = \mathbf{14}}$$

$$\bar{y} = 32.81cm^2$$

$$s = 2.48cm^2$$

$$SE = \frac{2.48}{\sqrt{n}} \leq 0.4$$

$$n \geq 38.4$$

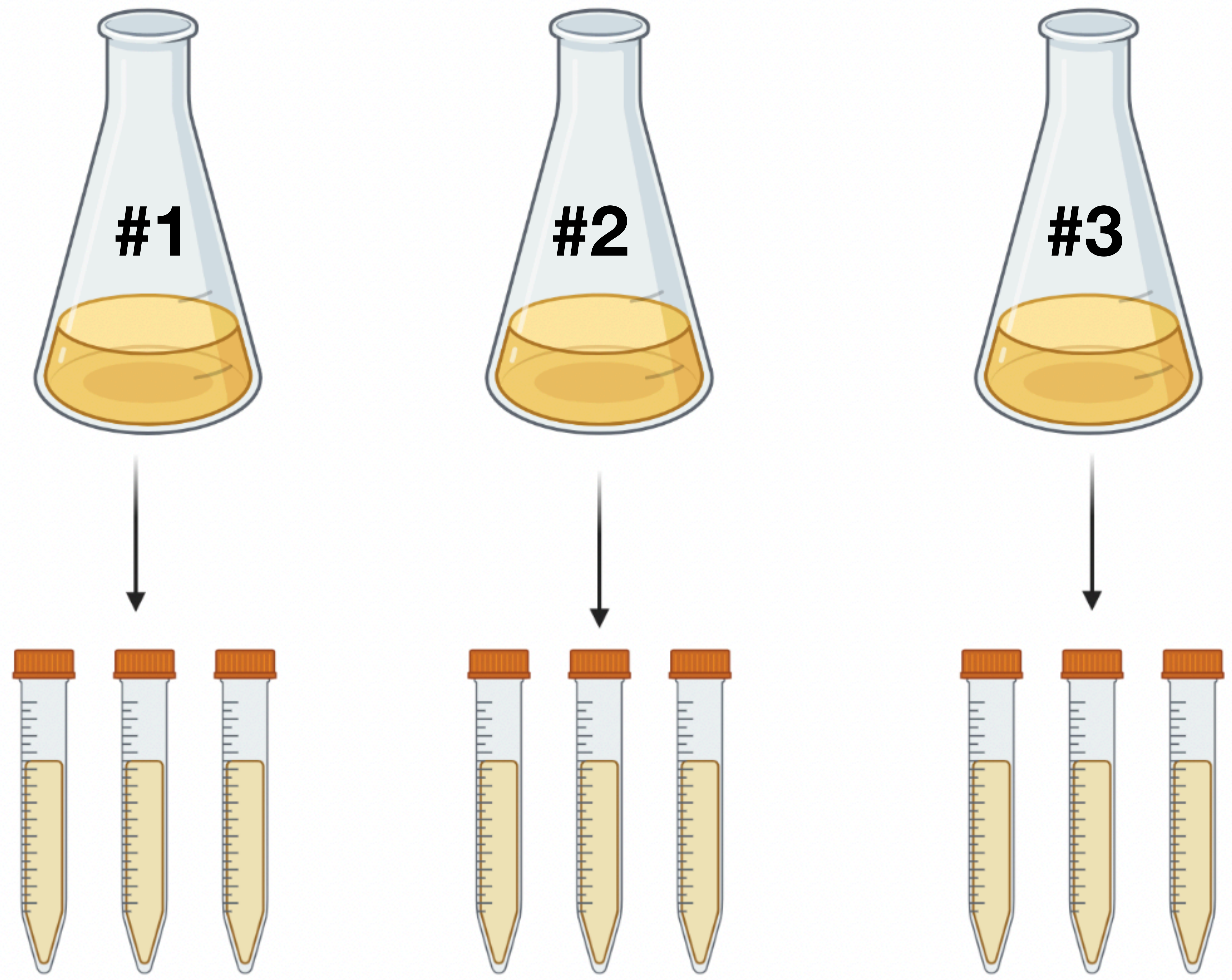
Assumptions for estimating confidence intervals

- **Conditions on the design of the study:**
  - (1) Data is a <u>random sample</u> from a <u>large population</u>
  - (2) Observations in the sample must be <u>independent</u> of each other

What does it mean for samples to be independent of each other?

Still a good experimental design! Means > individual sample

***n* = 3?**



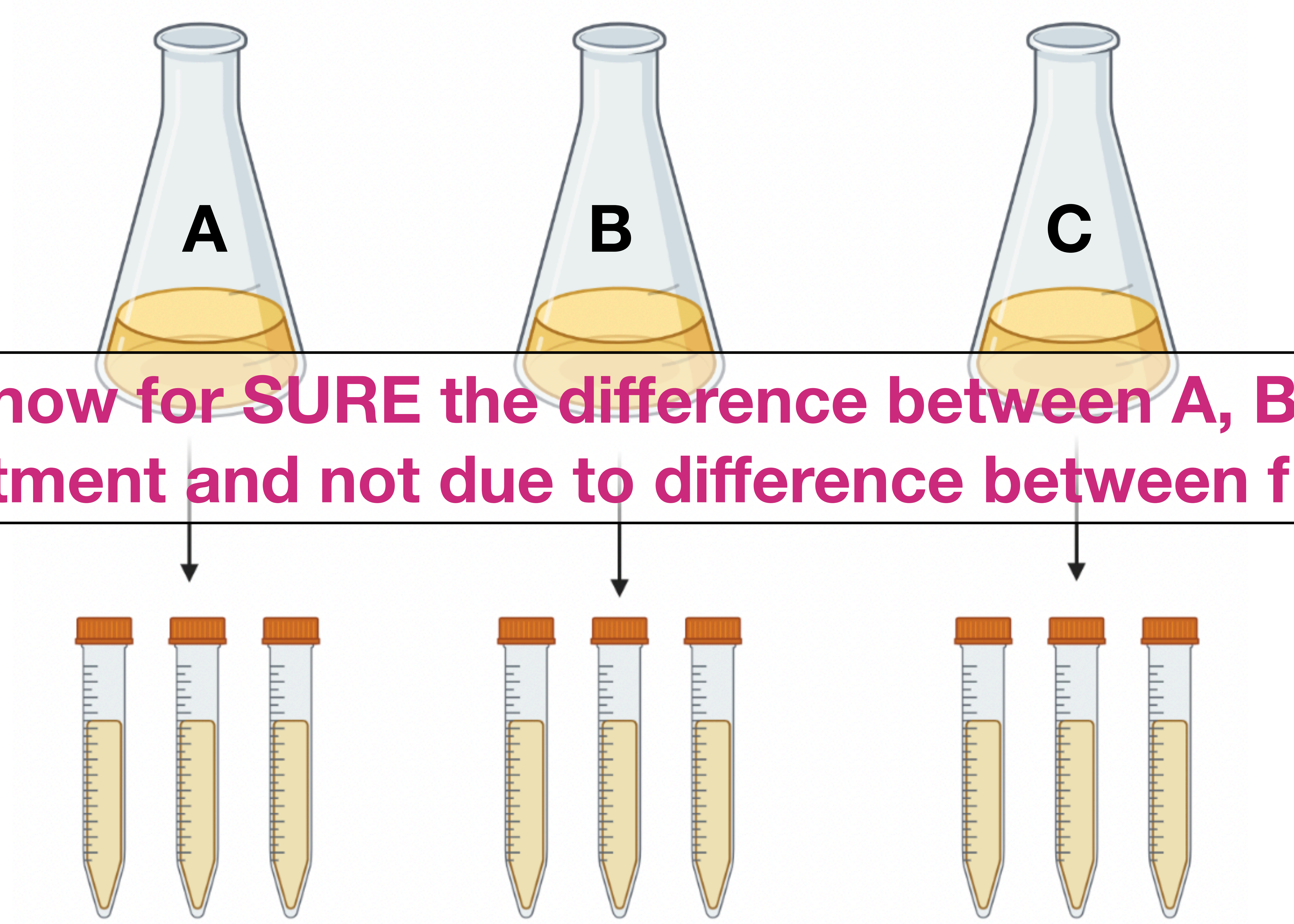
"Biological replicates"

***n* = 9?**

"Technical replicates"

"Hierarchical data structures"

What does it mean for samples to be independent of each other?



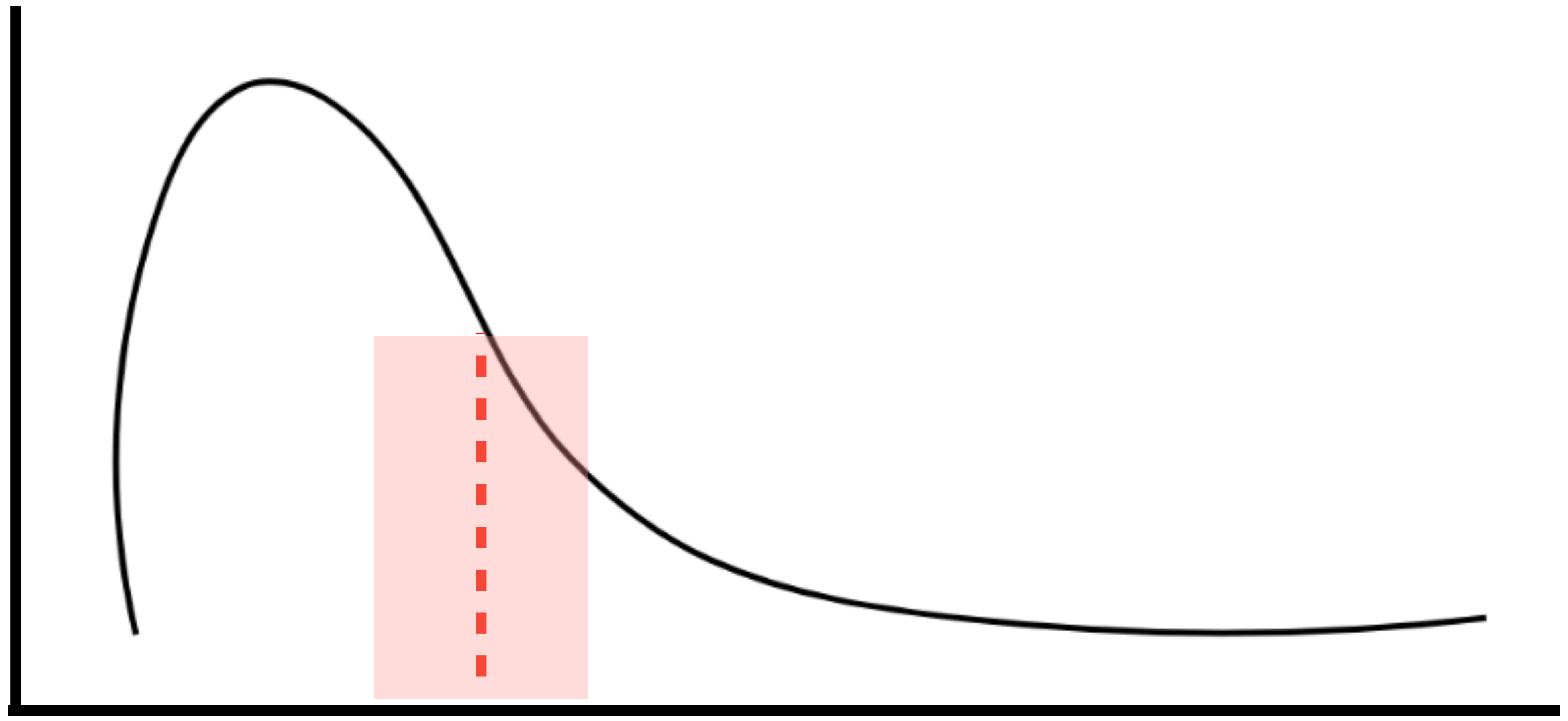
How can we know for SURE the difference between A, B, and C is due to treatment and not due to difference between flasks?

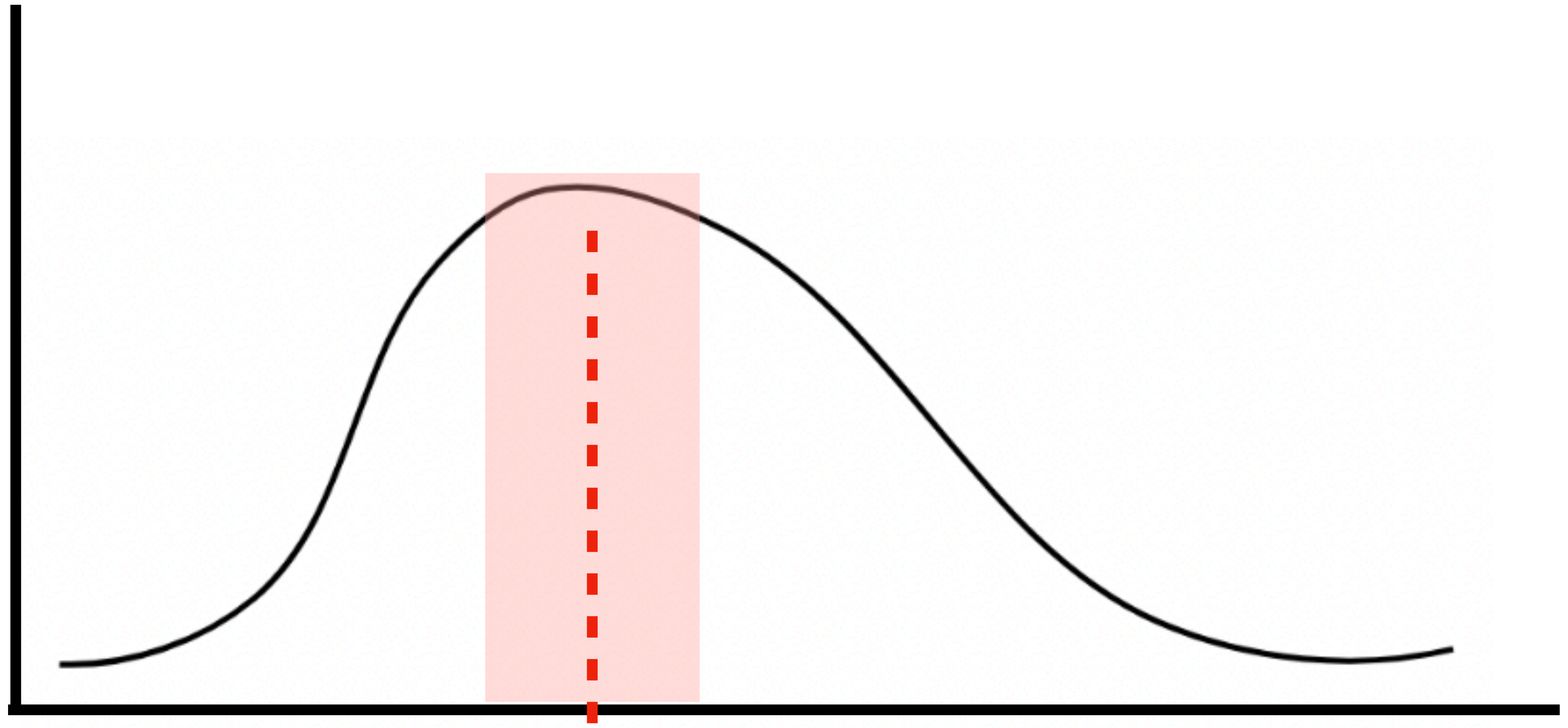
“Hierarchical data structures”

Assumptions for estimating confidence intervals

- **Conditions on the design of the study:**
    - (1) Data is a random sample from a large population
    - (2) Observations in the sample must be independent of each other
- **Conditions on the form of the population distribution**
    - (3) If n is small, the population distribution must be ~normal
    - (4) If n is large, the population distribution doesn't have to be normal

Why does it matter if our population is normally distributed?





Is the mean even a meaningful measure for this population?

If *n* is large enough, the sample distribution will be normal regardless of the shape of the population

How can we tell if a population is normally distributed?

- **Plot distribution (i.e. histogram)**
  - *Every analysis should begin with an inspection of the data and the points that lie far from the center*
- **Quantile plot**
- **Shapiro-wilks test for non-normality**
- **If not normal distribution, try a data transformation**

Announcements

- Extra practice problems from the textbook posted to GitHub and Canvas
    - Note: no solutions, but use your classmates/TAs for help!
- If you need help keeping track of the stats R functions we have been using, check out the `stats_R_cheatsheet.md` on GitHub!