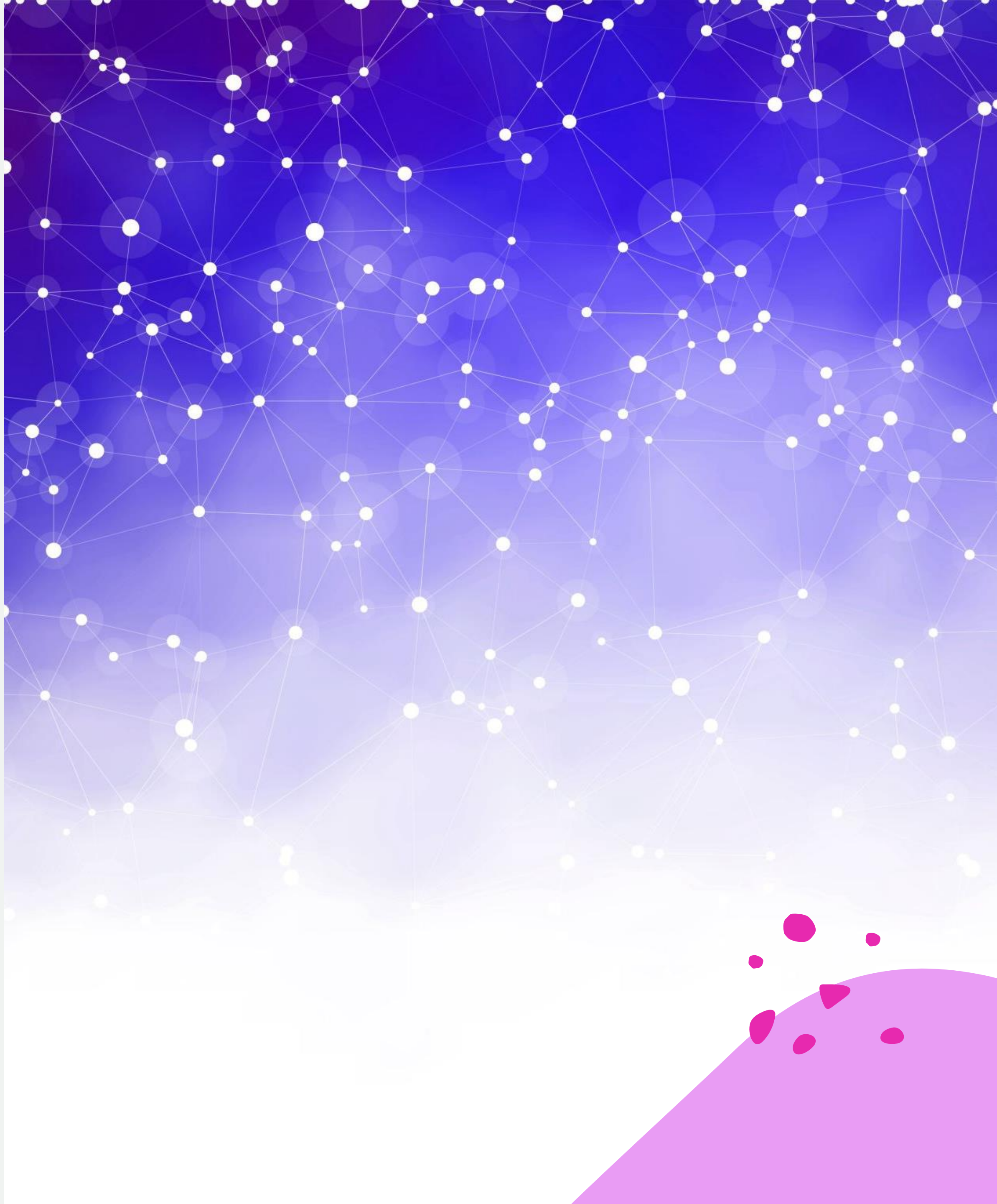
The Logic of Hypothesis Testing

PSY 611: Fall 2023

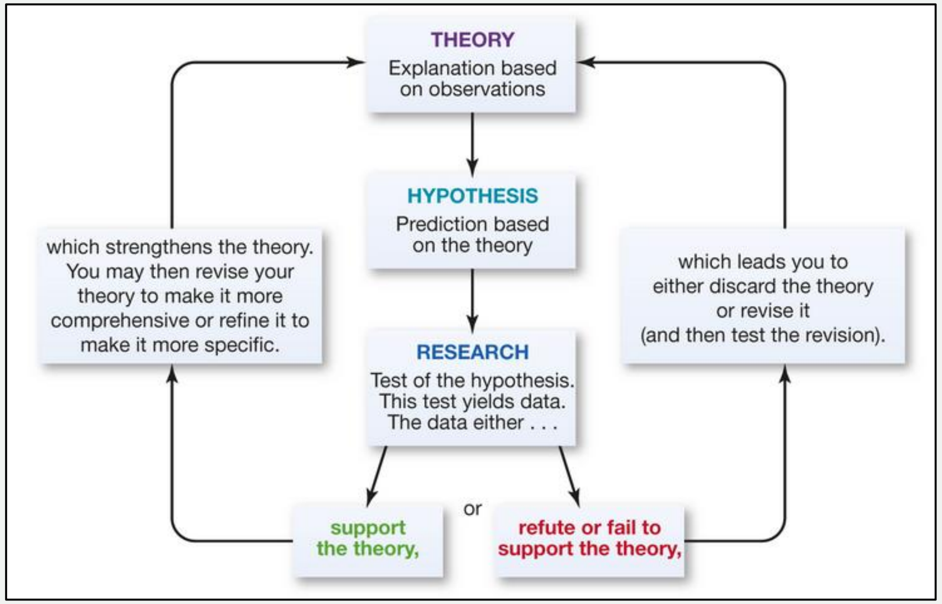


Theory Testing in Psychology

- The field of psychology is concerned with developing theories that explain the human mind and behaviors.

- In pursuit of this goal, researchers read the literature to get up-to-date with what has already been discovered in their field of study

- Based on the knowledge that has been gathered so far, psychology researchers:
    - Identify gaps in current knowledge, develop informed hypotheses, collect empirical observations, and use the steps of hypothesis testing to conclude whether their observations support, or refute, their hypotheses



The Steps of Hypothesis Testing

- ❑ Develop a **theory** by performing a literature review in your area of study
- ❑ State your **hypotheses** (a specific, testable prediction about the expected results)
  - Hypotheses are typically about the predicted <u>effect</u> of one variable on another or about the predicted <u>relationship</u> between two or more variables.
  - Typically, we specify two hypotheses: a null and an alternative hypothesis
- ❑ Specify the regression **models** corresponding to each of your hypotheses
- ❑ State your **decision rule** (i.e., Based on what criteria will you decide that the observations support, or refute, your hypothesis?)
- ❑ **Fit** both models to the data and then **evaluate** which model fits the real-world observations better
- ❑ **Interpret** the results & consider the **effect size**

State your Hypotheses

Example: Say a researcher develops a theory that working from home versus working in an office affects how satisfied people are with their jobs such that working from home makes people feel more satisfied. Let's say for this example that we know from prior research than on a work satisfaction scale (1 = *very unsatisfied*, 10 = *very satisfied*), people who work from an office score, on average, 5.00. For the current study, the researcher collected work satisfaction scores from 50 people working from home.

People working from an office

Average work satisfaction = 5.00

People working from home

Average work satisfaction = ??

State your Hypotheses

People working from an office

Average work satisfaction = 5.00

People working from home

Average work satisfaction = ??

- How should we state our hypothesis?
  - The intuitive way that we might want to state our hypothesis is that people who work from home will have higher work satisfaction than people who work from an office.
  - Hypothesis: Average work satisfaction of people working from home > Average work satisfaction of people working from an office

- There are a couple of issues with stating the hypothesis this way
  - This hypothesis does not make a **specific, testable** prediction. There are *many potential realities* that fit with this prediction, so we cannot very easily test whether our real-world observations support, or refute, this hypothesis.
  - The second issue is the **problem of induction**.

State your Hypotheses

People working from an office

Average work satisfaction = 5.00

People working from home

Average work satisfaction = ??

- Induction: the process of inferring universal rules given only particular observations
  - One school of thought within the philosophy of science (Karl Popper, 1902-1994) argues that we cannot argue for a universal theory based on repeated, similar observations.

Sam the swan is white;
Georgina the swan is white;
Fred the swan is white;
. . .
Emma the swan is white

Conclusion: All swans are wh

vho works from home is more satisfied
er job;
vho works from home is more satisfied
ieir job;
vho works from home is more
ed with their job

usion: All people who work from
are more satisfied with their jobs (?)

State your Hypotheses

- Karl Popper argued that while observations cannot be used to justify that a universal theory is *true*, they **can** be used to justify the claim that a universal theory is *false* (or… unlikely to be true)
  - If we have a theory that "all swans are white," then the first black swan we observe refutes this theory!

- This means we want to state our hypothesis not as the expected result that we hope our observations will *support*, but as the expected result that we hope our observations will *refute*
  - Null hypothesis

State your Hypotheses

People working from an office

Average work satisfaction = 5.00

People working from home

Average work satisfaction = ??

- For our scenario, let's state the null hypothesis in terms of the results we would expect to see if working environment ***has no effect*** on work satisfaction.

Null Hypothesis:

$$H_0: \mu_{WorkFromHome} = 5.00$$

Alternative Hypothesis:

$$H_1: \mu_{WorkFromHome} \neq 5.00$$

- The hypotheses are statements about *populations.*
- The two hypotheses must be *all-inclusive* and *mutually exclusive.*

State your Hypotheses

Null Hypothesis:

$$H_0: \mu_{WorkFromHome} = 5.00$$

Alternative Hypothesis:

$$H_1: \mu_{WorkFromHome} \neq 5.00$$

- The hypothesis that we *actually end up testing* is the **null hypothesis**.
- The frequentist approach to hypothesis testing examines ***the probability of obtaining the data we observed if the null hypothesis is true***.
  - $p < .05$ → Reject the null hypothesis
  - $p > .05$ → Fail to reject the null hypothesis

Review of Week 1 Concepts

Describing Data

- Measures of Central Tendency (mean, median, mode)
- Measures of Variability (min, max, range, IQR, variance, standard deviation, mean absolute deviation)
- Tables & Graphs (frequency table, bar graphs, boxplots, histograms, scatterplots)

Cyclical Nature of Theory Testing in Science

- Theory → Hypothesis → Empirical Research to Test Hypothesis → Findings *support* or *refute* theory → Original theory is *strengthened* or *revised*

The Steps of Hypothesis Testing

- The problem of induction: While observations *cannot* be used to justify a claim that a universal theory is *true,* observations *can* be used to justify a claim that a universal theory is *false* (e.g., "all swans are white")

  Null hypothesis: A statement of the predicted outcome if there is *no* effect, or *no* relationship between, two variables

Hypotheses

Null Hypothesis:

$$H_0: \mu_{WorkFromHome} = 5.00$$

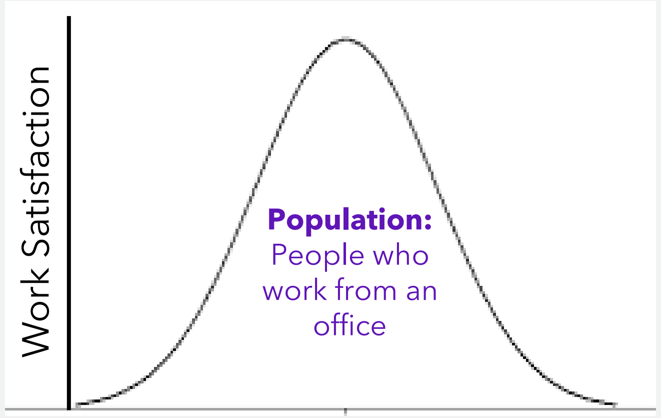
Alternative Hypothesis:

$$H_1: \mu_{WorkFromHome} \neq 5.00$$

- The hypotheses are statements about *populations*.
- The two hypotheses must be *all-inclusive* and *mutually exclusive*.
- The hypothesis that we *actually end up testing* is the **null hypothesis**.

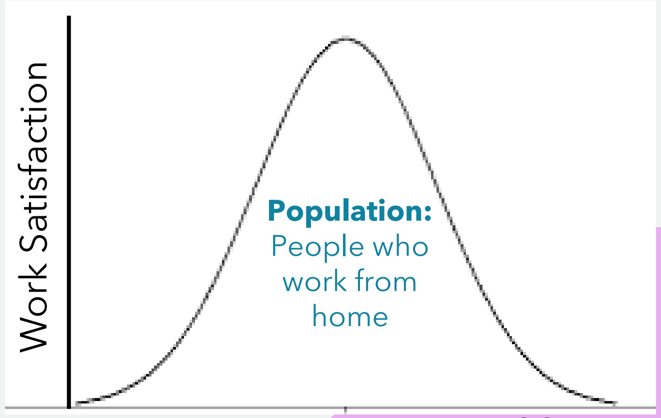
Samples vs Populations

Example: Say a researcher develops a theory that working from home versus working in an office affects how satisfied people are with their jobs such that working from home makes people feel more satisfied. Let's say for this example that we know from prior research than on a work satisfaction scale (1 = *very unsatisfied*, 10 = *very satisfied*), people who work from an office score, on average, 5.00. For the current study, the researcher collected work satisfaction scores from 50 people working from home.



$\mu_{WorkFromOffice} = 5.00$

Null Hypothesis (H_0): $\mu_{WorkFromHome} = 5.00$



$\mu_{WorkFromHome} = ??$

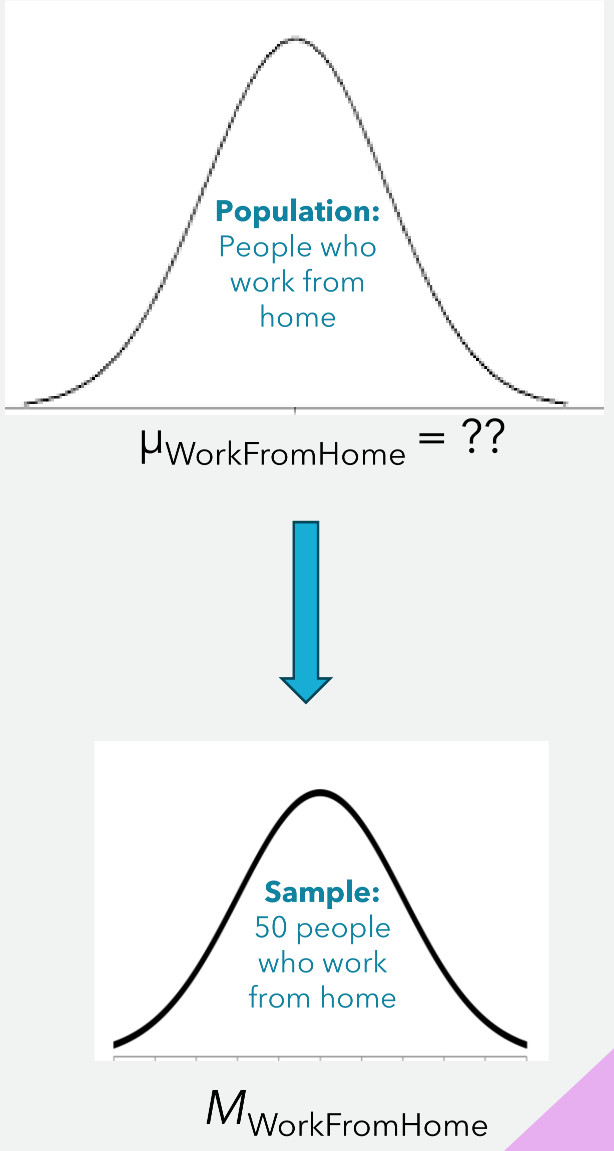
Q: If we're interested in how the average work satisfaction of these two populations compares, why doesn't the researcher just collect population-level data for the second population by surveying everyone who works from home and measuring their work satisfaction?

Samples vs Populations

- **Population:** the larger group of individuals that a researcher wants to draw conclusions about
- **Sample:** a smaller subset of participants drawn from the larger population

Due to *practical limitations*, researchers often collect data from samples instead of populations. Collecting data requires:

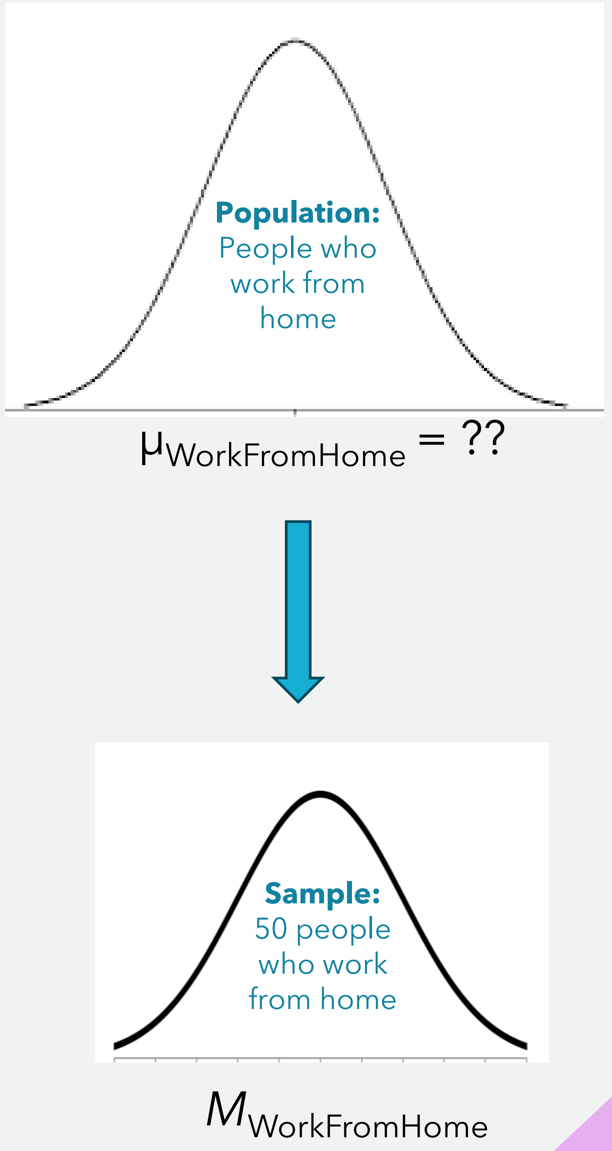
- Time, resources, & willing participants



Sampling

Let’s say this researcher only has enough time and resources to collect a **sample** of 50 participants (n = 50) from the Work from Home population.

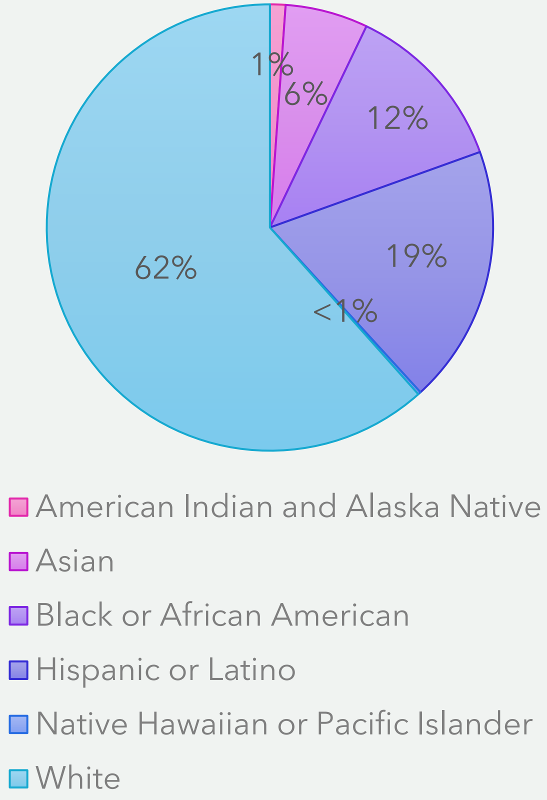
- The researcher is trained in good methods and follows all the recommendations for obtaining a **representative sample**.

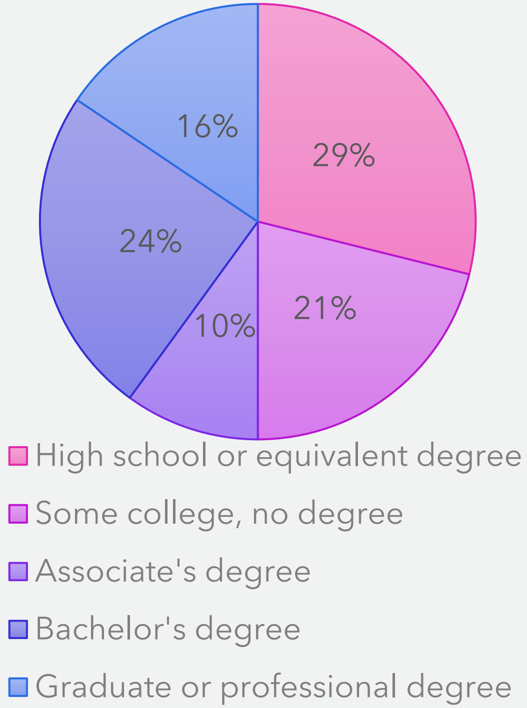


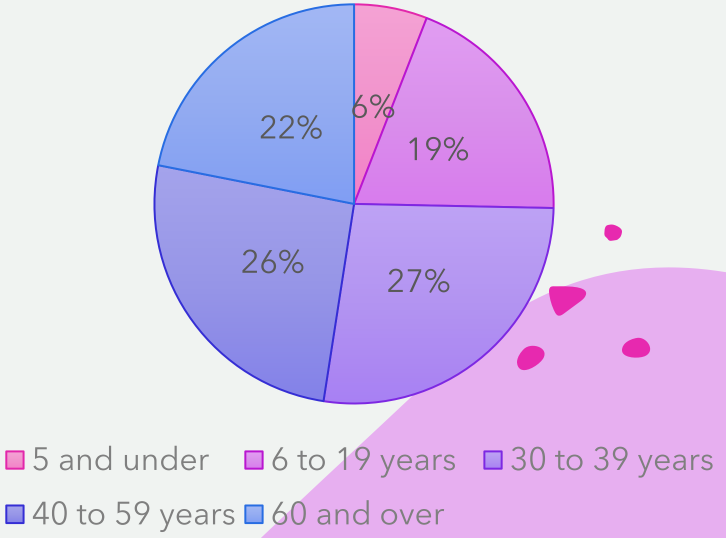
A Representative Sample

Representative Sample: the participants have the same attributes as those that exist in the population and in approximately the same proportions, making *generalization* possible

- Example: US population demographics (based on 2020 Census data)







Obtaining a Representative Sample

Methods for obtaining a representative sample:

- **Random Sample**: recruiting participants from the population completely randomly; all cases have an equal chance of being selected for the sample
- **Large sample size:** a larger sample is more likely to represent the population

Even when good methods are used to collect one's data (e.g., random sample, adequate sample size) there are <u>still</u> issues that arise from the sampling process.

Practical Issues with Sampling

- **Self-selection bias:** occurs when not everyone who is asked to participate in a study agrees to do so
  - The sample will be non-representative because only the characteristics of individuals who self-select into the study will be represented (Ex: a study on cheating behaviors).

- **Sampling Error:** the discrepancy, due to random factors, between a sample statistic and the population parameter the sample is attempting to estimate.
  - Ex: The difference between a sample mean and the true mean of the population the sample came from.

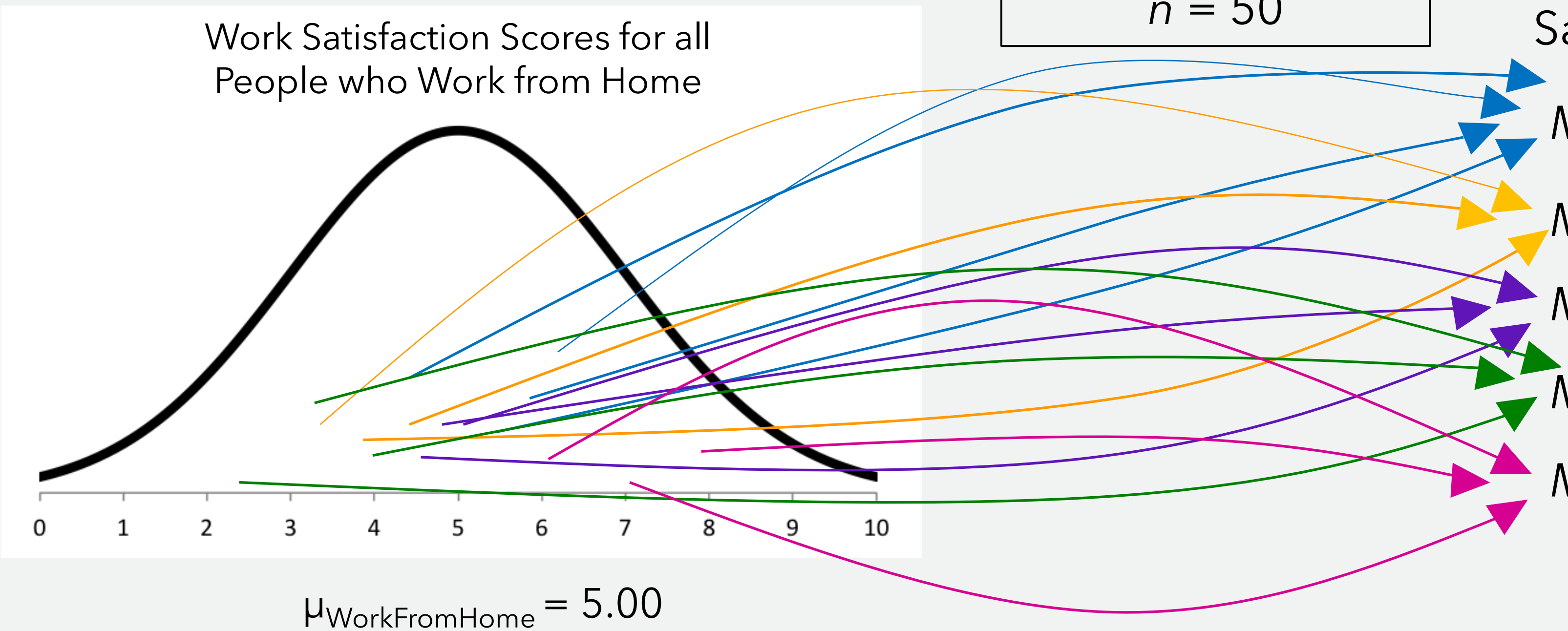
Sampling Error

Thought experiment:
Let's propose that the true population mean on work satisfaction for all people who work from home *is* 5.00.

- (This is just for this exercise. We don't actually know the true population mean – if we did, we wouldn't need to do any further steps!)

Randomly Sample
n = 50

Sample Means



M_1 = 6.00

M_2 = 4.00

M_3 = 5.25

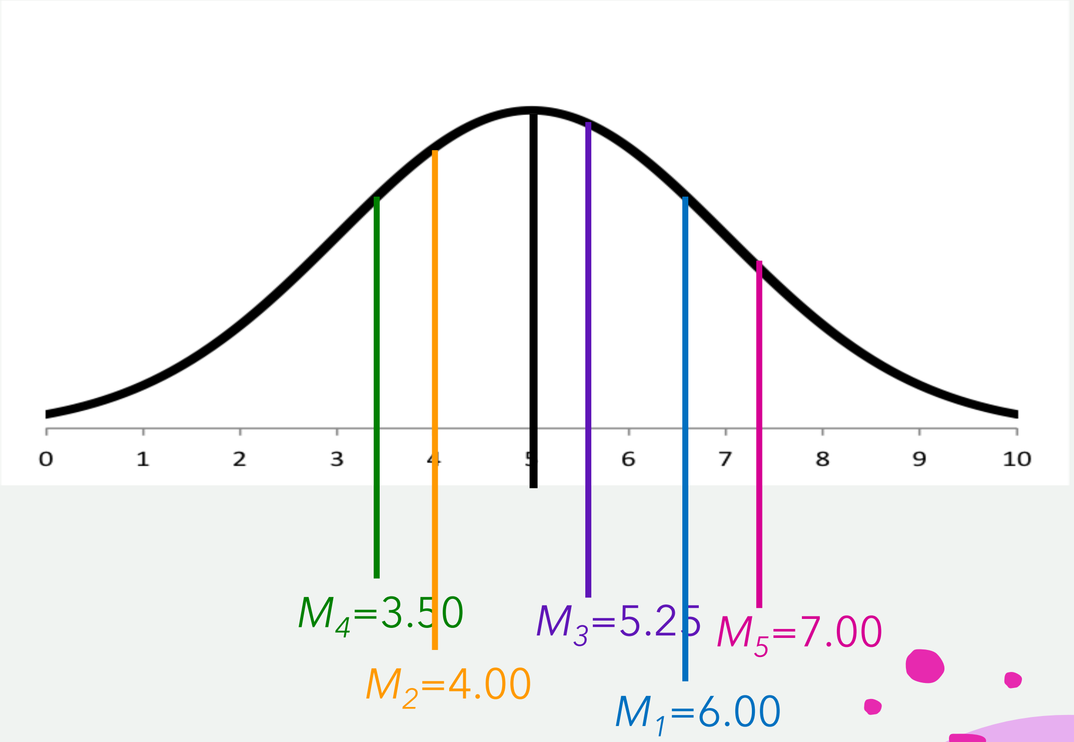
M_4 = 3.50

M_5 = 7.00

$\mu_{WorkFromHome}$ = 5.00

Sampling Error

- Sampling error is due to completely random factors (i.e., chance)
  - Who happened to end up in your sample when you randomly sampled from the population?
- You are equally likely to obtain a sample mean above or below the actual population mean.
  - This means that the sample mean is an *unbiased* estimator of the true mean of the population.



Sampling Distributions

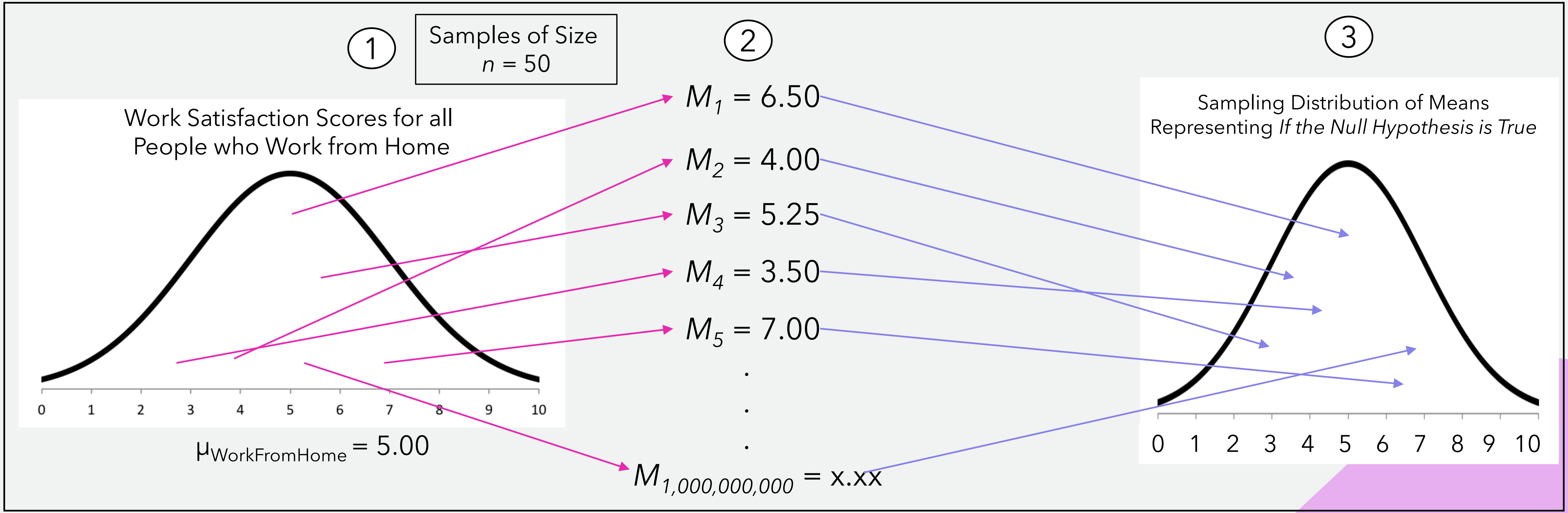
A tool for representing all the possible results *if the null hypothesis is true*.

Sampling Distributions

- A **sampling distribution** is a frequency distribution that's created by calculating all of the *possible* sample means (or another type of statistic) that could be obtained by randomly sampling from a given population.

To generate a sampling distribution:

1. Take repeated, random samples of a specified size from a population.
2. Calculate the mean (or another statistic) for each sample.
3. Plot all of the sample means (or the other statistic you calculated) in a frequency distribution.



Central Limit Theorem

Instead of having to construct a sampling distribution of means from scratch…

- **Central Limit Theorem:** describes the characteristics of a sampling distribution of means when the sample size is large and every possible sample is obtained

<u>If n is large ($n \geq 30$), then…</u>

(1) The mean of the sampling distribution of means is equal to the population mean

$$\mu_M = \mu$$

(2) The standard deviation of the sampling distribution of means, aka the **standard error of the mean**, is equal to the standard deviation of the population (or the estimate of the population standard deviation) divided by the square root of the sample size

$$\sigma_M = \frac{\sigma}{\sqrt{n}} \qquad or \qquad s_M = \frac{s}{\sqrt{n}}$$

(3) The sampling distribution of the mean will be normally distributed, no matter what the shape of the population is.

Standard Error

- The term **"standard error"** refers to *the standard deviation of a sampling distribution*.

- The standard error of a *sampling distribution of means* is the standard deviation of the population of individual scores divided by the square root of the sample size:

$$\sigma_M = \frac{\sigma}{\sqrt{n}} \qquad \textit{or} \qquad s_M = \frac{s}{\sqrt{n}}$$

- Why? See https://www.youtube.com/watch?v=eVeeJDgkjrQ&ab_channel=MathsandStats for a proof.

Decision Rules

Decision Errors

- *Thought Exercise*: Say the researcher calculates the average work satisfaction of the 50 participants they gathered data from who work from home and finds that the mean of the sample is 6.00 (M = 6.00). This is higher than the average work satisfaction of people who work from an office (μ = 5.00).

What are **two possible explanations** for this result?

1. People who work from home are actually, on average, more satisfied with work than people who work from an office.

OR

2. The means of the two populations are equal, and *we just happened to obtain this sample mean due to* ***sampling error*** (i.e., due to chance)**.**

Decision Errors: Type I and Type II

- **Type I Error:** the researcher concluding that there is an effect, or a relationship, between variables when, in reality, there is not (i.e., the null hypothesis is actually true)

- **Type II Error:** the researcher concluding that there is *not* an effect, or a relationship, between variables when, in reality, there is (i.e., the null hypothesis is actually false)

Type I and Type II Decision Errors

	In reality, H_0 is False	In reality, H_0 is True
Researcher decides to Reject H_0	Correct Decision	Type I Error
Researcher decides to Fail to Reject H_0	Type II Error	Correct Decision

Setting a Decision Rule

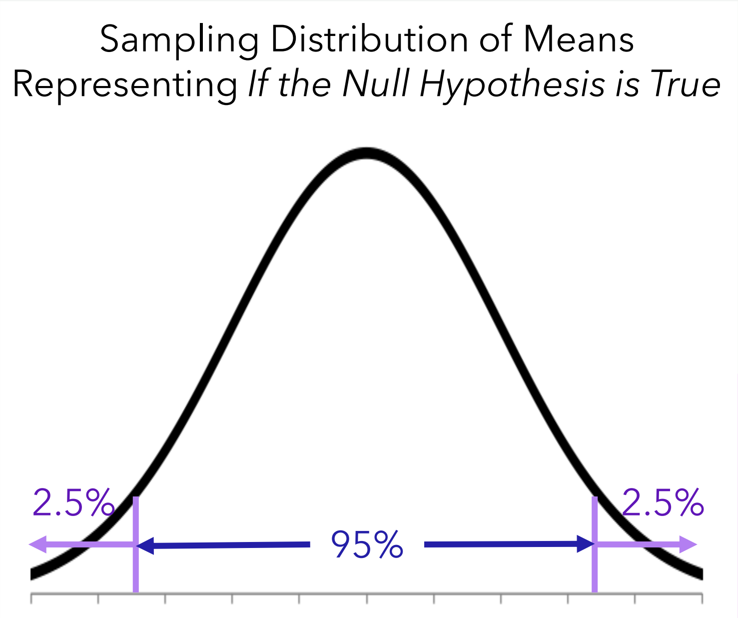
- Deciding on the **decision rule** means deciding *how willing you are to make a Type I error.*
  - *Q:* Say *you* are the researcher conducting this study to see whether or not people's work satisfaction is better when people work at home compared to when they work in an office. How willing are you to make a Type I error (to conclude there is a real effect when, in fact, there is none?)

- α (alpha) is a researcher's willingness to make a Type I error
  - The convention in psychology research is to set alpha at 5% ($\alpha = .05$)

Setting a Decision Rule

- Remember that we can construct a sampling distribu *ll the sample means we could obtain when the null hypothesis is true.*
  - Setting a decision rule means deciding *where* on your s you need to land to reject, or fail to reject, the null hypothesis

Where is there a less than 5% chance of landing on this sampling distribution?



- Setting α = .05 means we will:
  - **Reject the null hypothesis** when the probability of obtaining our results is less than 5% when the null hypothesis is true.
  - **Fail to reject the null hypothesis** when the probability of obtaining our results is greater than 5% when the null hypothesis is true.

Where did your results land on the sampling distribution?

- Calculate a test statistic:

$$z = \frac{M - \mu}{\sigma_M} \qquad \text{or} \qquad t = \frac{M - \mu}{s_M}$$

- The *z*- and *t*-statistics describe where your sample mean lies compared to the mean of the sampling distribution in standard error units.

Making a Decision about the Null Hypothesis

What is the probability (p) of obtaining *your sample mean* when the null hypothesis is true?

- You can obtain a p-value for your results using a statistical software (like R)

If $p < .05$

- Decision: Reject the null hypothesis
- The results are statistically significant

If $p > .05$

- Decision: Fail to reject the null hypothesis
- The results are non-significant

Effect Size

- When we say an effect is "significant," that means the results we obtained are unlikely simply due to chance
  - More likely due to a true effect (or true relationship)

- However, statistical significance does not tell you the *size* of your study's results

- One commonly used measure of effect size is Cohen's *d*:

$$d = \frac{|M - \mu|}{\sigma} \quad \text{or} \quad d = \frac{|M - \mu|}{s}$$

The size of the difference between the two means being compared in standard deviation units.

Comparing a Sample Mean to a Population Mean

The Traditional Approach

Theory

Example: A researcher believes that practicing yoga can reduce people's anxiety levels. In general, people in the US population score a 6.00 on general anxiety ($\mu = 6.00$). The researcher randomly samples 8 people from the US population, has them practice yoga for six months, and then measures their anxiety levels. At the end of the six month period, the individuals' anxiety scores were:

4, 6, 3, 4, 8, 7, 6, 2

State the Hypotheses

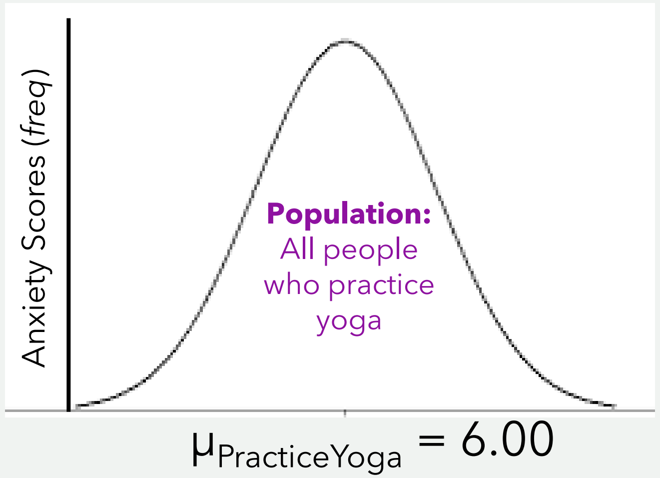
Example: A researcher believes that practicing yoga can reduce people's anxiety levels. In general, people in the US population score a 6.00 on general anxiety ($\mu = 6.00$). The researcher randomly samples 8 people from the US population, has them practice yoga for six months, and then measures their anxiety levels. At the end of the six month period, the individuals' anxiety scores were: 4, 6, 3, 4, 8, 7, 6, 2

Null hypothesis:

- **H_0:** $\mu_{PractiveYoga} = 6.00$

Alternative hypothesis:

- **H_1:** $\mu_{PractiveYoga} \neq 6.00$



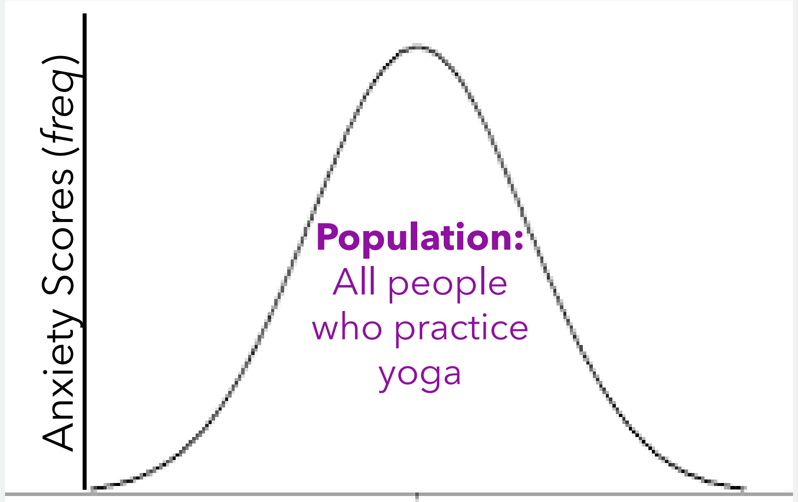
Set the Decision Rule

- Unless you have reason to alter it, typically we use the default alpha level that is the convention in psychology research
  - $\boldsymbol{\alpha = .05}$

- I am 5% willing to make a Type I error

Construct a Sampling Distribution Representing all the Possible Results <u>if the null hypothesis is true</u>

Null hypothesis: H_0: $\mu_{PractiveYoga} = 6.00$



$\mu_{PracticeYoga} = 6.00$

***s* is calculated from the data:**

$$s = \sqrt{\frac{\Sigma(X - M)^2}{n-1}}$$

Mean of the SDoM:

- $\mu_M = \mu = 6.00$

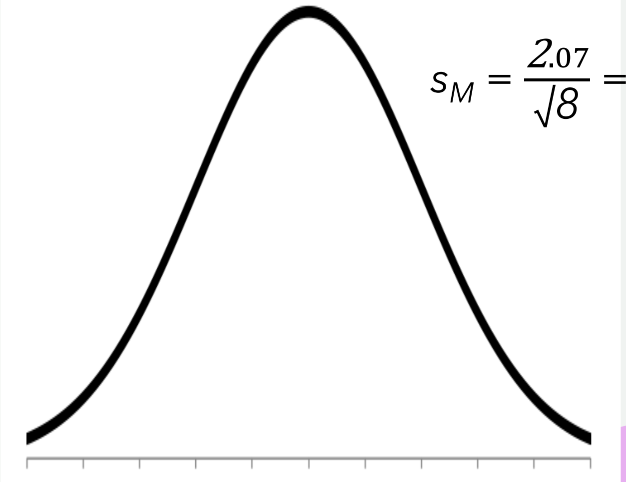
Standard Error of the SDoM:

- $s_M = \frac{s}{\sqrt{n}} = \frac{2.07}{\sqrt{8}} = 0.7319$

Shape

- If the population SD (σ) is known
  - Normally distributed
- If the population SD (σ) is unknown
  - ***t*-distribution**

Sampling Distribution of Means (SDoM) Representing *If the Null Hypothesis is True*



$s_M = \frac{2.07}{\sqrt{8}} = 0.7319$

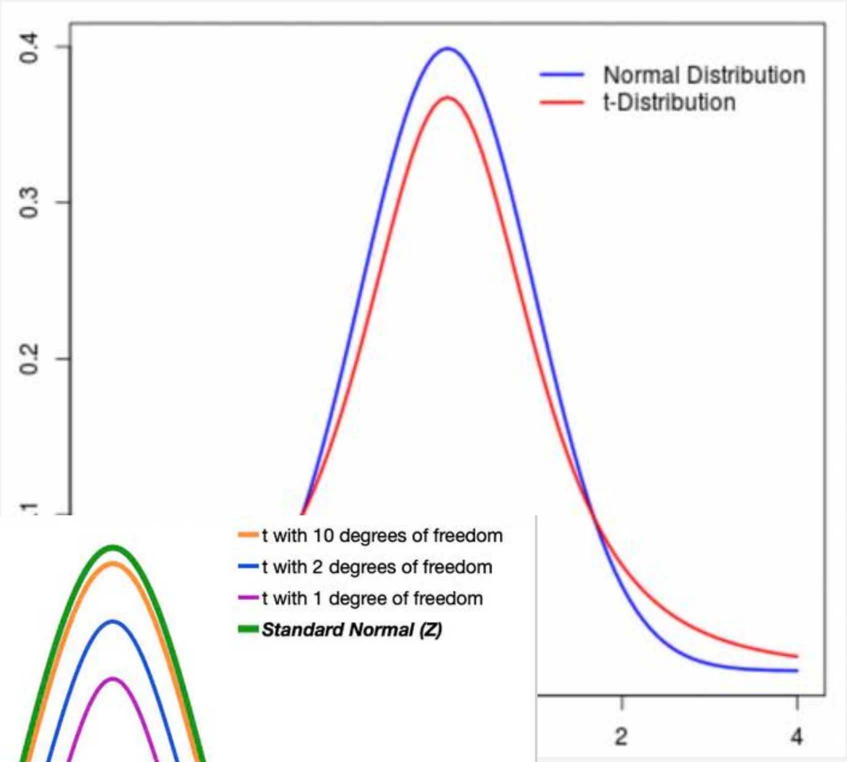
$\mu_M = 6.00$

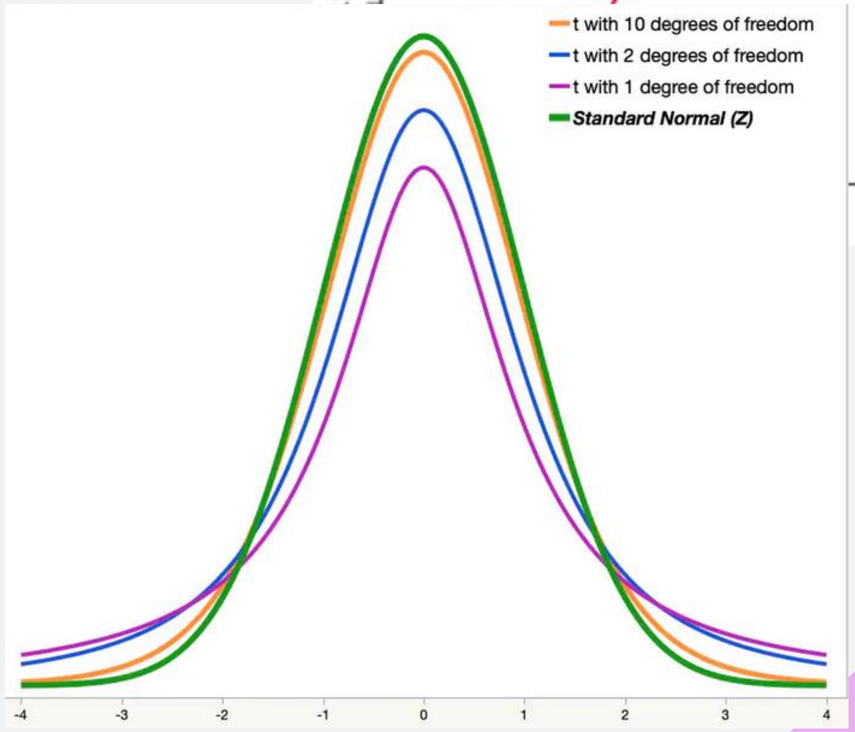
The t-distribution

- Similar to the normal distribution, but has thicker tails

- The shape of the *t*-distribution changes depending on the degrees of freedom
  - $df = n - 1$

- As *df* increase, the shape of the *t*-distribution becomes more and more similar to a normal distribution





Where did your sample mean land on the sampling distribution?

- Raw scores of our sample

```r
anxiety_yoga <- c(4, 6, 3, 4, 8, 7, 6, 2)
```

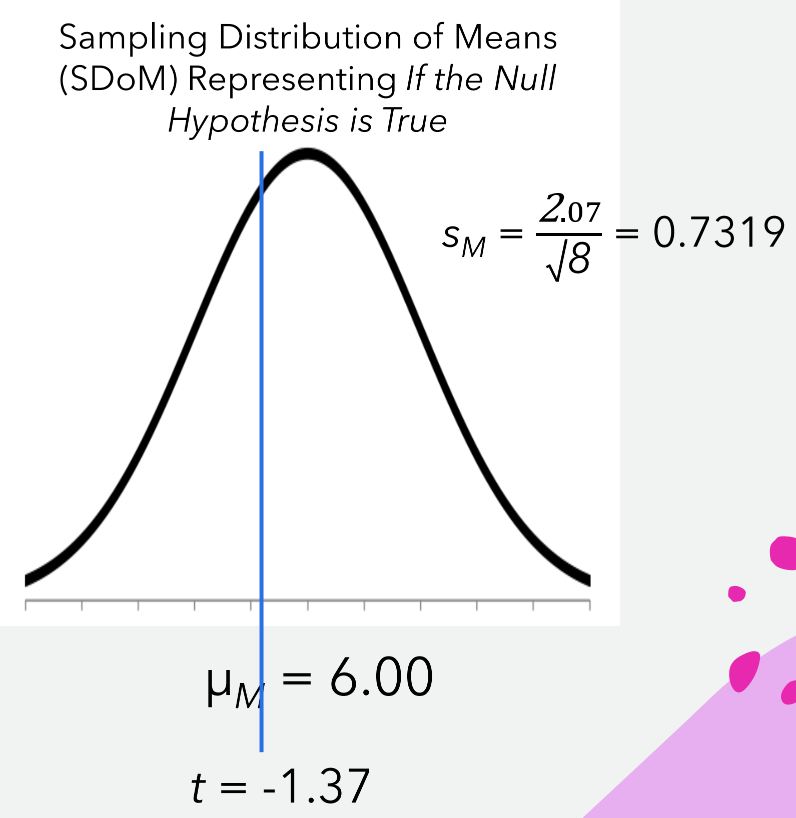
- Sample mean:

```r
> mean(anxiety_yoga)
[1] 5
```

- Calculate the test statistic to see where this sample mean lands on the SDoM:

$$t = \frac{M - \mu}{s_M} = \frac{5 - 6}{0.7319} = -1.37$$

  - Our sample mean is 1.37 standard errors below the mean



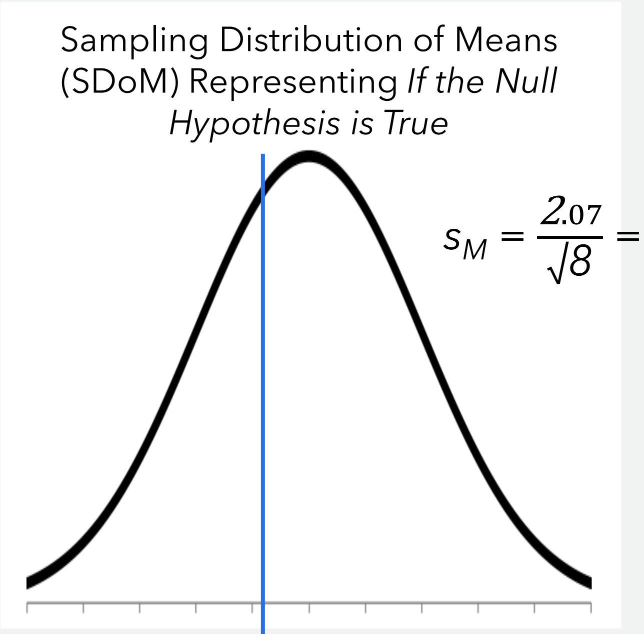
What is the probability of obtaining this sample mean if the null hypothesis is true?

- You can obtain a *p*-value for your results using a statistical software (like R)

```
> t.test(anxiety_yoga, mu = 6)

        One Sample t-test

data:  anxiety_yoga
t = -1.3663, df = 7, p-value = 0.2141
alternative hypothesis: true mean is not equal to 6
95 percent confidence interval:
 3.269272 6.730728
sample estimates:
mean of x
        5
```



$s_M = \frac{2.07}{\sqrt{8}} = 0.7319$

$\mu_M = 6.00$

$t = -1.37$

$\boxed{p = .214}$

Making a Decision about the Null Hypothesis

If $p < .05$

- Decision: Reject the null hypothesis
- The results are statistically significant

If $p > .05$

- Decision: Fail to reject the null hypothesis
- The results are non-significant

Our Decision: Fail to reject the null hypothesis because $p > .05$

- There is no significant difference in the anxiety levels of people who practice yoga compared to the general US population

Summary:
Comparing a Sample Mean to a Population Mean using "The Traditional Approach"

Research Question: Does practicing yoga reduce people's anxiety? The average anxiety in the general US population is currently $\mu = 6.00$.

State the hypotheses:

- $\mathbf{H_0}$: $\mu_{PractiveYoga} = 6.00$
- $\mathbf{H_1}$: $\mu_{PractiveYoga} \neq 6.00$

Set the decision rule:

- $\alpha = .05$, two-tailed

Construct a sampling distribution of means representing <u>if the null hypothesis is true</u>.

Calculate the test statistic (where did your sample mean land?):

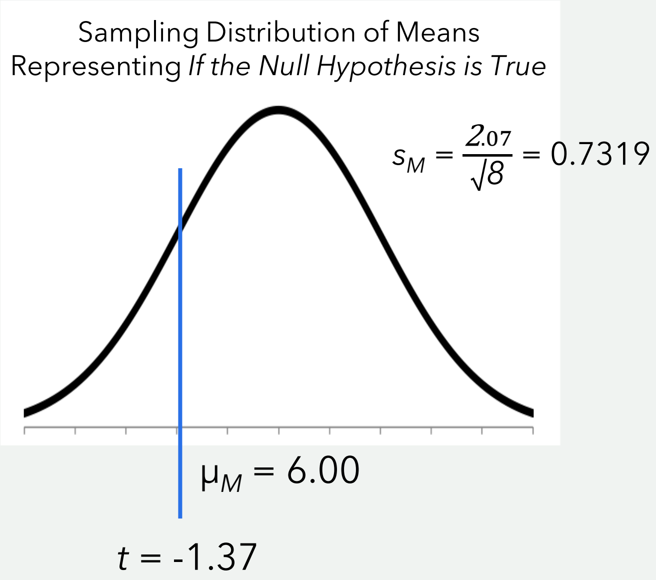
- $t = \dfrac{M - \mu}{s_M} = \dfrac{5 - 6}{0.7319} = -1.37$

Obtain the *p*-value using R:

```
p-value = 0.2141
```

Make a decision about the null hypothesis:

- Fail to reject H_0 because $p > .05$
- No significant difference in anxiety



What's missing?

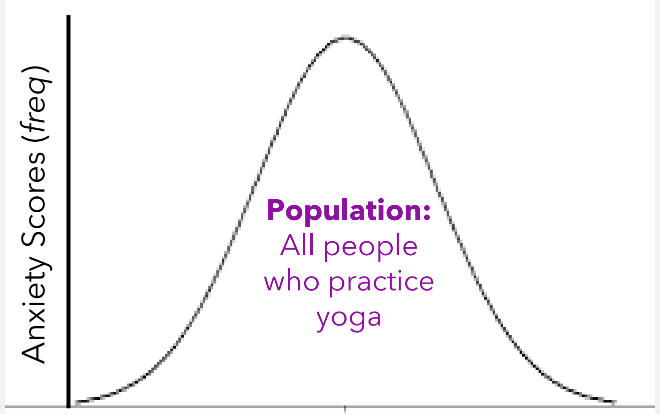
The effect size!

Effect Size

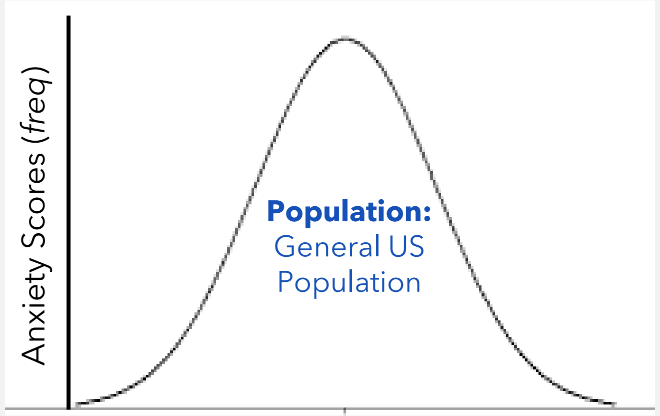
- $d = \frac{|M - \mu|}{s} = \frac{|5 - 6|}{2.07} = 0.48$

```
> cohensD(anxiety_yoga, mu = 6) # lsr
[1] 0.4830459
```

The size of the difference between the two means being compared in standard deviation units.



$\mu_{PracticeYoga} = 5.00$



$\mu_{US\_Population} = 6.00$

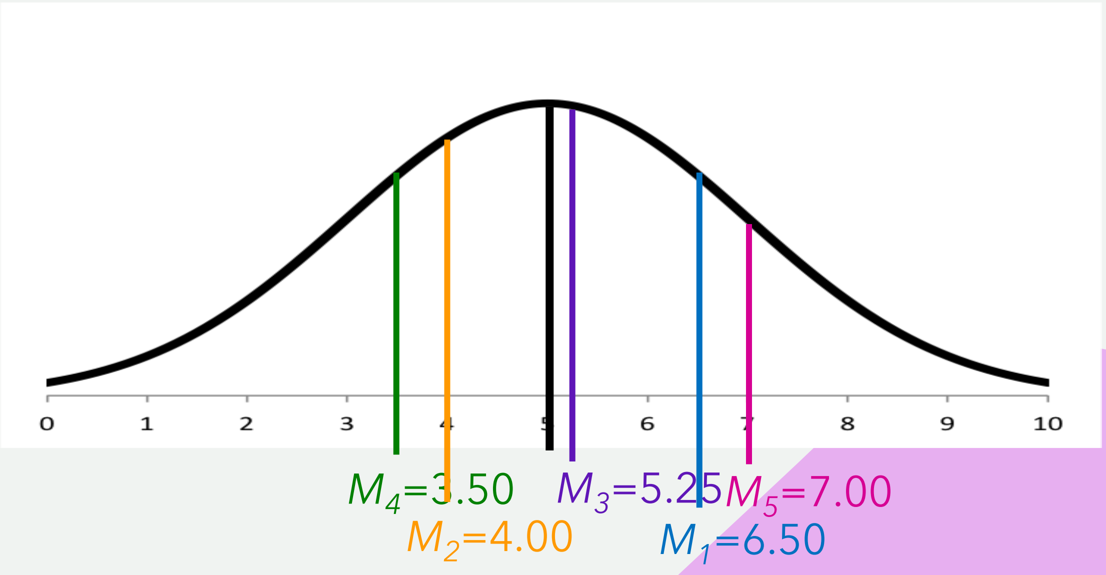
0.48 standard deviations

Confidence Intervals

Point Estimates

- **Point Estimate:** a single value that is used to estimate a population value
  - **Example:** A sample mean (*M*) is used to estimate the population mean (μ)

- Due to sampling error, a point estimate is unlikely to be exactly equal to the population parameter it's estimating.



Interval Estimates

- **Interval Estimate:** a range of values around a point estimate within which a population parameter is likely to exist
  - More likely to contain the true population parameter than a point estimate

- A commonly reported interval estimate in psychological research is the **95% confidence interval**

95% Confidence Interval

95% Confidence Interval around the Sample Mean

$$95\%CI = M \pm (t_{cv} \text{ x } s_M)$$

- M = sample mean
- t_{cv} = t-critical values corresponding to an $\alpha = .05$, $df = n - 1$, and two-tailed test
- s_M = the estimated standard error of the sampling distribution of means

95% Confidence Interval around the Sample Mean

$$95\%\text{CI} = M \pm (t_{cv} \text{ x } s_M)$$

For our example:

- $M = 5.00$
- t_{cv} (α = .05, df = 8-1 = 7, two-tailed) = 2.365
- $s_M = \frac{2.07}{\sqrt{8}} = 0.7319$

- Lower Bound: 5 – (2.365*0.7319) = 3.27
- Upper Bound: 5 + (2.365*0.7319) = 6.73

- Reporting in APA Style:
  95%CI[3.27, 6.73]

df	α = .05, one-tailed -or- α = .10, two-tailed	α = .025, one-tailed -or- α = .05, two-tailed	α = .01, one-tailed -or- α = .02, two-tailed	α = .005, one-tailed -or- α =.01, two-tailed
1	6.314	12.706	31.821	63.657
2	2.920	4.303	6.965	9.925
3	2.353	3.182	4.541	5.841
4	2.132	2.776	3.747	4.604
5	2.015	2.571	3.365	4.032
6	1.943	2.447	3.143	3.707
7	1.895	2.365	2.998	3.499
8	1.860	2.306	2.896	3.355
9	1.833	2.262	2.821	3.250
10	1.812	2.228	2.764	3.169

```
> t.test(anxiety_yoga, mu = 6)

        One Sample t-test

data:  anxiety_yoga
t = -1.3663, df = 7, p-value = 0.2141
alternative hypothesis: true mean is not equal to 6
95 percent confidence interval:
 3.269272 6.730728
sample estimates:
mean of x
        5
```

95% Confidence Interval around the Sample Mean

$$95\%CI = M \pm (t_{cv} \text{ x } s_M)$$

- When constructing a 95%CI around a sample mean:
  - If the 95%CI around the sample mean **contains the population mean (μ) proposed by the null hypothesis**

    → The findings are *not significant*

  - If the 95%CI around the sample mean **does *not* contain the population mean (μ) proposed by the null hypothesis**

    → The findings are *significant*

What does a 95%CI around a sample mean… mean?

- *Officially*, a 95% confidence interval tells you that if you were to repeat the process of sampling and calculating 95% CI's for each sample mean, 95% of the time, the CI will contain the true population mean

- If you repeatedly sampled and calculated 95%CI's:
  - 95% of the time, the CI captures the true population mean, μ
  - 5% of the time, the CI does not capture the true population mean, μ

- Simulation of 95%CIs: https://shiny.rit.albany.edu/stat/confidence/

95% Confidence Interval

95% Confidence Interval around the Difference Between Means

$$95\%\text{CI}\mu_{\text{Diff}} = (M - \mu) \pm (t_{cv} \text{ x } s_M)$$

- M = sample mean
- μ = population mean for the known population
- t_{cv} = *t*-critical values corresponding to an α = .05, *df* = *n* – 1, and two-tailed test
- s_M = the estimated standard error of the sampling distribution of means

95% Confidence Interval around the Difference Between Means

$$95\%CI\mu_{Diff} = (M - \mu) \pm (t_{cv} \times s_M)$$

For our example:

- $M = 5.00$
- $\mu = 6.00$
- t_{cv} (α = .05, df = 8-1 = 7, two-tailed) = 2.365
- $s_M = \frac{2.07}{\sqrt{8}} = 0.7319$

df	α = .05, one-tailed -or- α = .10, two-tailed	α = .025, one-tailed -or- α = .05, two-tailed	α = .01, one-tailed -or- α = .02, two-tailed	α = .005, one-tailed -or- α =.01, two-tailed
1	6.314	12.706	31.821	63.657
2	2.920	4.303	6.965	9.925
3	2.353	3.182	4.541	5.841
4	2.132	2.776	3.747	4.604
5	2.015	2.571	3.365	4.032
6	1.943	2.447	3.143	3.707
7	1.895	2.365	2.998	3.499
8	1.860	2.306	2.896	3.355
9	1.833	2.262	2.821	3.250
10	1.812	2.228	2.764	3.169

- Lower Bound: (5-6) – (2.365*0.7319) = -2.73
- Upper Bound: (5-6) + (2.365*0.7319) = 0.73

- Reporting in APA Style:
  $95\%CI\mu_{Diff}[-2.73, 0.73]$

95% Confidence Interval around the Difference Between Means

$$95\%\mathrm{CI}\mu_{\mathrm{Diff}} = (M - \mu) \pm (t_{cv} \times s_M)$$

- When constructing a 95%CI around the difference between means:
  - If the 95%CI **contains zero**, the two means are *not significantly different*
  - If the 95%CI **does not contain zero**, the two means are *significantly different*

What is a p-value?

p-values

- Earlier, I mentioned that we can get the *p*-value corresponding to our test statistic using a statistical software, like R
  - How was this *p*-value calculated?

```
> t.test(anxiety_yoga, mu = 6)

        One Sample t-test

data:  anxiety_yoga
t = -1.3663, df = 7, p-value = 0.2141
alternative hypothesis: true mean is not equal to 6
95 percent confidence interval:
 3.269272 6.730728
sample estimates:
mean of x
        5
```

p-values

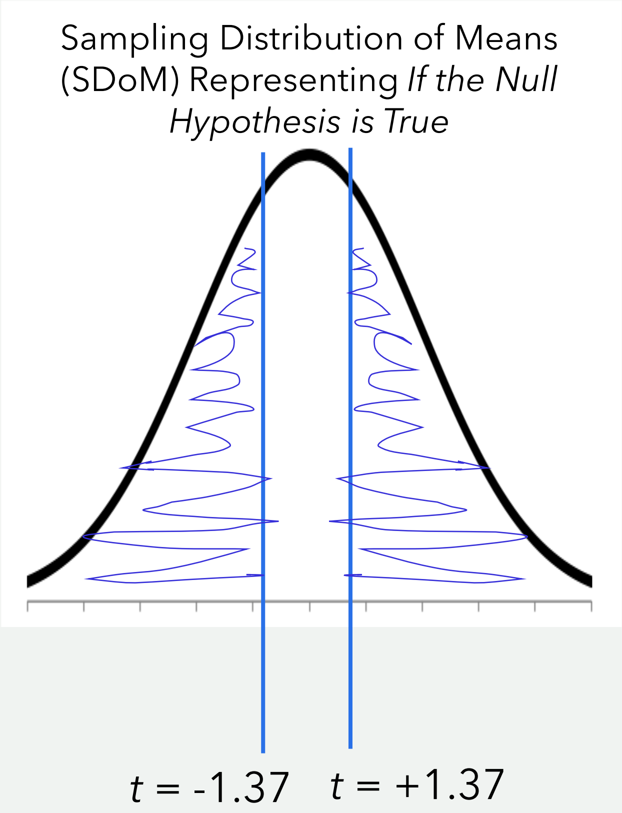
- Remember: When we have to estimate the population standard deviation, our sampling distribution of means is a ***t*-distribution**

- The *p*-value is the probability of obtaining a test statistic as extreme, or more extreme, than our test statistic when the null hypothesis is true
  - *In either direction* because we conducted a two-tailed test

- The pt() function in R gives us the probability of landing *below a particular t-statistic* for a *t-distribution* with a specified *df*.

```
> pt(-1.3663, df = 7)
[1] 0.1070566
```

```
> pt(-1.3663, df = 7)*2
[1] 0.2141131
```



Testing our knowledge of p-values

- Suppose we run an experiment testing the effectiveness of a new drug for treating anxiety. We compare a treatment group and a control group and find that patients in the treatment group improve significantly more than those in the control group (p = .001). Based on this finding, which of the following statements is valid and which invalid?

Which of the following statements are valid?

Write down which of the following statements are true or false.

1. The probability that the null hypothesis is true is .001.
2. The probability of mistakenly rejecting the null hypothesis is .001.
3. The difference between the two groups is large.
4. The probability of obtaining a difference as extreme or more extreme than ours is .001.
5. The probability of rejecting the null hypothesis is .999.
6. The probability of obtaining a difference as extreme or more extreme than ours, if the null hypothesis is true, is .001.
7. The probability that our findings are due to chance is .001.