RESEARCH METHODS

PSY 612

Research Methods

Research methods refer to how the data is collected for a study.

Data collection occurs prior to data analysis.

- However, how the data is collected has important implications for what conclusions can be drawn based on the results of one's analysis.

Measuring Variables

Psychological variables are often **latent constructs**, or constructs that we assume exist but that can't be directly observed.

- e.g., personality traits, emotions, attitudes, beliefs, etc.

Cannot be directly observed because…

- They are tendencies across a variety of situations

- They are a complex set of characteristics that can't be represented by a single thought or behavior

- **Measurement:** the assignment of numerical scores to individuals in a way that represents an individual's level on a construct of interest

Defining Variables

Conceptual Definition: describes the behaviors and internal processes that make up that construct

Example: Extraversion is commonly defined as people's tendency to be talkative, sociable, and assertive across a variety of situations

Defining Variables

Operational Definitions: a definition of a variable in terms of precisely how it is to be measured

- **Self-report measures:** participants report on their own thoughts, feelings, and behaviors

- **Behavioral measures:** some aspect of participants' behavior is observed and recorded

- **Physiological measures:** some aspect of participants' physiological processes are recorded

The Big Five Inventory–2 (BFI-2)

Here are a number of characteristics that may or may not apply to you. For example, do you agree that you are someone who *likes to spend time with others?* Please write a number next to each statement to indicate the extent to which you agree or disagree with that statement.

1 Disagree strongly	2 Disagree a little	3 Neutral; no opinion	4 Agree a little	5 Agree strongly

I am someone who...

1. ___ Is outgoing, sociable.
2. ___ Is compassionate, has a soft heart.
3. ___ Tends to be disorganized.
4. ___ Is relaxed, handles stress well.
5. ___ Has few artistic interests.
6. ___ Has an assertive personality.

31. ___ Is sometimes shy, introverted.
32. ___ Is helpful and unselfish with others.
33. ___ Keeps things neat and tidy.

Operationalization

Operationalization is the process of determining how to measure a variable in a way that accurately represents its conceptual definition.

A psychological construct could be operationalized in multiple different ways.

For example, stress tends to be defined as an adaptive response to a perceived danger or threat that involves physiological, affective, and behavioral components.

- **Operational definition 1:** measuring stress using a self-report questionnaire on which people report on their feelings of stress (e.g., "How often have you felt nervous and stressed?")
- **Operational definition 2:** measuring stress using physiological processes like levels of the stress hormone cortisol

- The researcher desires converging results!

Reliability & Validity

⸺◇⸺

How do researchers know that scores on their measure of a variable represent the psychological construct they are intended to capture?

- When developing a measure, psychologists should empirically examine whether their measure captures the construct it is intended to represent by assessing its **reliability** and **validity**.

Reliability: the consistency of a measure over time, across items, and across researchers

Validity: the extent to which the scores from a measure represent the variable they are intended to

Reliability

Test-Retest Reliability

Test-retest reliability: the extent to which a measure of a construct tends to produce consistent scores over time (for variables that it makes sense to assume temporal consistency)

- **Example:** Since intelligence is expected to be stable over time, people who take an IQ test at two different time points should score approximately the same each time.

To assess test-retest reliability, administer a measure of a construct to the same group of people at one time, then again at a later time, and examine the correlation between the two sets of scores.

- Generally, a test-retest correlation of **+.80 or higher** is considered to indicate good reliability.
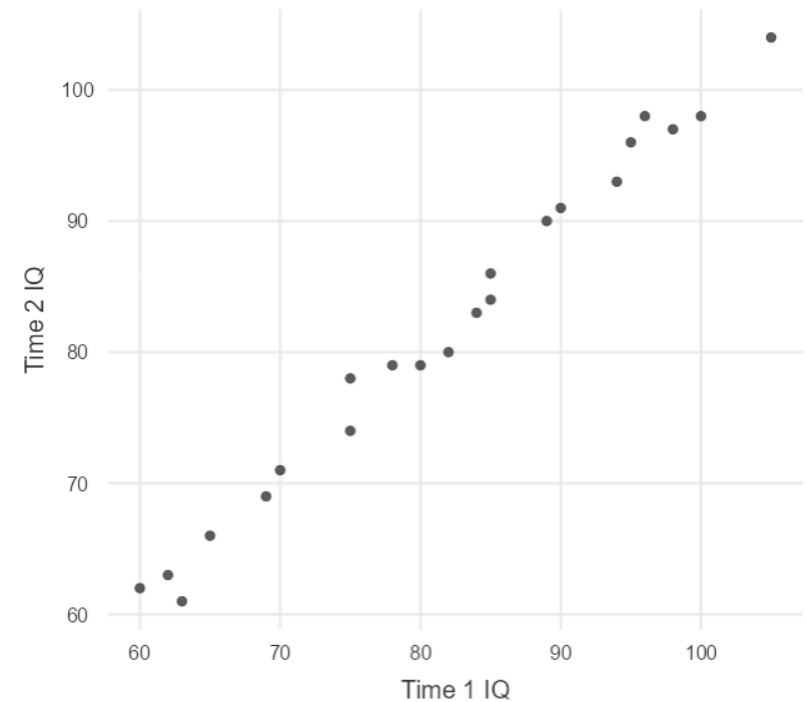
Internal Consistency

On a multiple-item measure of a construct, each of the items is meant to be measuring the same overall concept.

- Thus, the items on the measure should be correlated with each other.

Internal consistency: the consistency of people's responses across the items on a multiple-item measure of a construct

The Big Five Inventory-2

Here are a number of characteristics that may or may not apply to you. On a scale of 1 (Disagree strongly) to 5 (Agree strongly), indicate how much you agree or disagree with each of the following statements.

I am someone who…

1. Is outgoing, sociable
2. Has an assertive personality
3. Is dominant, acts as a leader
4. Is full of energy
5. Is talkative
6. Shows a lot of enthusiasm

Internal Consistency

The most common way of measuring the internal consistency of a measure is by calculating **Cronbach's α.**

- Cronbach's α compares the amount of shared variance, or covariance, among the items making up a multi-item instrument to the amount of overall variance in the items

- To calculate, it requires a multi-item measure with **at least four items**

$$\alpha = \frac{N\bar{c}}{\bar{v} + (N-1)\bar{c}}$$

Cronbach's alpha	Internal consistency
α ≥ 0.9	Excellent
0.9 > α ≥ 0.8	Good
0.8 > α ≥ 0.7	Acceptable
0.7 > α ≥ 0.6	Questionable
0.6 > α ≥ 0.5	Poor
0.5 > α	Unacceptable

- **N** is the total number of items

- c-bar is the mean covariance between pairs of items

- v-bar is the mean variance of all the items

Interrater Reliability

Many behavioral measures involve researchers observing and rating the behaviors performed by participants.

- **Example:** Observers coding how extraverted they perceive participants as being based on their behaviors during a social interaction.

Interrater reliability: the extent to which different observers are consistent in their judgments.

Interrater reliability is also commonly assessed using Cronbach's alpha to measure the level of agreement among coders' ratings.

Validity

Face Validity

Face validity: the extent to which a measurement tool "on its face" appears to measure the construct of interest

- For example, what construct does the following scale appear to you to be a measure of?

- To assess face validity, you can have a large sample of people complete the survey and rate whether it appears to measure the construct it is intended to (but it's often confirmed informally).

Instructions

Below is a list of statements dealing with your general feelings about yourself. Please indicate how strongly you agree or disagree with each statement.

1. On the whole, I am satisfied with myself.

Strongly Agree   Agree   Disagree   Strongly Disagree

2. At times I think I am no good at all.

Strongly Agree   Agree   Disagree   Strongly Disagree

3. I feel that I have a number of good qualities.

Strongly Agree   Agree   Disagree   Strongly Disagree

4. I am able to do things as well as most other people.

Strongly Agree   Agree   Disagree   Strongly Disagree

5. I feel I do not have much to be proud of.

Strongly Agree   Agree   Disagree   Strongly Disagree

Content Validity

Content validity: the extent to which a measure "covers" the construct of interest

- **Example:** Say a researcher defines test anxiety as the experience of bodily reactions and cognitive disruptions due to test taking, as well as general negative attitudes toward test taking. The researcher's measure should capture all three of these aspects of the conceptual definition.

- Multi-dimensional measurement tools can be used to include subscales that capture different components of the psychological phenomenon

This is a difficult type of validity to empirically establish that is often only informally validated.

Test Anxiety Inventory

Read each item below to see if it reflects your experience in test taking. If it does, place a check mark on the line next to the number of the statement.

Bodily Reactions Subscale:

1. I do not enjoy eating before or after an important test.

2. Having to face an important test disturbs my sleep.

3. I cannot relax physically before a test.

Disruptive Thoughts Subscale:

4. During a test, I sometimes find myself having trains of thought that have nothing to do with the test.

5. I mentally freeze up on important tests.

6. I sometimes seem to mentally defeat myself while working on an important test.

Negative Attitudes Toward Test-Taking Subscale:

7. It seems to me that test sessions should not be made the formal, tense situations they are.

8. If exams could be done away with, I think I would actually learn more from my courses.

Criterion Validity

Criterion validity: the extent to which people's scores on a measure are correlated with other variables (i.e., criterions) that one would expect them to be correlated with

- **Concurrent validity:** when people's scores on a measure of interest are correlated with their scores on a criterion that was measured at the same time (Ex: Conscientiousness and GPA)

- **Predictive validity:** when people's scores on a measure of interest are correlated with their scores on a criterion that was measured at a point in the future (Ex: Conscientiousness and early career salaries)

- **Convergent validity:** when people's scores on a measure of interest are correlated with their scores on criterion that measures the *same* construct (Ex: Conscientiousness as measured using the BFI and the TIPI)

- **Discriminant validity:** when people's scores on a measure of interest are *not* highly correlated with their scores on a criterion that should measure a construct that is conceptually distinct (Ex: Conscientiousness and agreeableness)

Choosing a Measure

Using an Existing Measure

Advantages:

- Saves time and trouble of creating your own

- There is already some evidence that the measure is valid if it has been used successfully

- Your results can be more easily compared with and combined with previous results

- Typically, it will probably be expected that you use an existing measure of a construct unless you have a good and clearly stated reason for not doing so

  - PsycTESTS is a great resource for finding already-existing measures: https://psycnet.apa.org/PsycTESTS/testname

Disadvantages:

- An existing measure may not exist for a construct of interest in your study

  - In this case, you may need to design your own.

Tips for Designing Survey Items

- Strive for simplicity

- Clear instructions

- Write items/questions at a reading level that is accessible to a wide audience

- Avoid double-barreled questions (ex: "How satisfied are you with your work environment and pay?")

- Avoid double negatives (ex: "How unwilling are you to not pay extra for a movie ticket?")

- Phrase items in a way that avoids socially desirable responses (ex: "Are you a good person?")

- Use multiple items rather than a single item (at least 4 items per subscale)
  - Multiple items often needed to cover full conceptual definition
  - Multiple items increase measure's reliability
  - Put items on the same response scale so they can be aggregated

- Pre-test the measure with several people for comprehensibility, length, etc.

Scale Options

Likert scales are a very popular scaling option that people are given to respond to surveys items on:

- Likert scales often give 5 or 7 options
  - Provide room for variability in responses
  - True midpoint
  - More built-in response options in Qualtrics

Recommended that you add numerical values to the response options to encourage participants to treat the scale as an interval scale.

How strongly do you agree or disagree with each of the following statements?

	1 Strongly Disagree	2 Disagree	3 Slightly Disagree	4 Neither Agree nor Disagree	5 Slightly Agree	6 Agree	7 Strongly Agree
In most ways my life is close to my ideal.	○	○	○	○	○	○	○
The conditions of my life are excellent.	○	○	○	○	○	○	○
I am satisfied with my life.	○	○	○	○	○	○	○

Handling Survey Data

- **Reverse-coding items**
  - Some measurements will include items that are worded so that *low scores* indicate the presence of the construct being measured
  - These items *must* be reverse-coded prior to aggregating scores for the overall variable

- Items can be reverse coded using the formula: (Max – X) + Min, where
  - Max is the highest possible score
  - X is a participant's actual score
  - Min is the lowest possible score

The Big Five Inventory-2

Here are a number of characteristics that may or may not apply to you. On a scale of 1 (Disagree strongly) to 5 (Agree strongly), indicate how much you agree or disagree with each of the following statements.

<u>**I am someone who...**</u>

1. Is outgoing, sociable
2. Has an assertive personality
3. Rarely feels excited or eager (*reverse*)
4. Tends to be quiet (*reverse*)
5. Is dominant, acts as a leader
6. Is sometimes shy, introverted (*reverse*)

Example: on a 1 to 5 Likert scale
- $(5 - 5) + 1 = 1$
- $(5 - 4) + 1 = 2$
- $(5 - 3) + 1 = 3$
- $(5 - 2) + 1 = 4$
- $(5 - 1) + 1 = 5$

Handling Survey Data

Example in R: Reverse-coding items

Original data

	item1	item2	item3	item4	item5	item6
1	5	4	1	1	5	2
2	1	1	5	5	2	4
3	3	3	3	3	4	3
4	4	5	2	2	2	2
5	2	3	4	4	1	5

```r
data2 <- data %>%
  mutate(across(c(item3, item4, item6), ~5 - . + 1))
```

Reverse-coded data

	item1	item2	item3	item4	item5	item6
1	5	4	5	5	5	4
2	1	1	1	1	2	2
3	3	3	3	3	4	3
4	4	5	4	4	2	4
5	2	3	2	2	1	1

Handling Survey Data

- **Cronbach's alpha**
  - Cronbach's alpha for measuring the internal consistency of the items should be calculated *after* reverse-coding needed items

  - If Cronbach's alpha is unexpectedly low, this could be because some items haven't been appropriately reverse-coded

The Big Five Inventory-2

Here are a number of characteristics that may or may not apply to you. On a scale of 1 (Disagree strongly) to 5 (Agree strongly), indicate how much you agree or disagree with each of the following statements.

<u>**I am someone who...**</u>

1. Is outgoing, sociable

2. Has an assertive personality

3. Rarely feels excited or eager (*reverse*)

4. Tends to be quiet (*reverse*)

5. Is dominant, acts as a leader

6. Is sometimes shy, introverted (*reverse*)

Handling Survey Data

Example in R: Cronbach's alpha

```
{r}
psych::alpha(data2)
```

```
Reliability analysis
Call: psych::alpha(x = data2)

 raw_alpha std.alpha G6(smc) average_r S/N    ase mean  sd median_r
      0.95      0.96    0.96      0.78  21 0.035    3 1.4     0.85

     95% confidence boundaries
            lower alpha upper
Feldt        0.84  0.95  0.99
Duhachek     0.88  0.95  1.02

 Reliability if an item is dropped:
        raw_alpha std.alpha G6(smc) average_r S/N var.r med.r
item1      0.93      0.93    0.97      0.73  14 0.044  0.76
item2      0.96      0.96    0.82      0.82  23 0.020  0.85
item3      0.93      0.93    0.97      0.73  14 0.044  0.76
item4      0.93      0.93    0.97      0.73  14 0.044  0.76
item5      0.97      0.97    0.84      0.88  36 0.010  0.85
item6      0.95      0.95    0.95      0.78  18 0.056  0.85
```

Cronbach's alpha	Internal consistency
$\alpha \geq 0.9$	Excellent
$0.9 > \alpha \geq 0.8$	Good
$0.8 > \alpha \geq 0.7$	Acceptable
$0.7 > \alpha \geq 0.6$	Questionable
$0.6 > \alpha \geq 0.5$	Poor
$0.5 > \alpha$	Unacceptable

Handling Survey Data

- **Aggregating Scores**
  - For multi-item measures of a construct, after reverse-coding items (as needed), a single composite score needs to be calculated to represent how someone scored on that construct overall

- The most common method of aggregating is simply to take the average of the set of items

The Big Five Inventory-2

Here are a number of characteristics that may or may not apply to you. On a scale of 1 (Disagree strongly) to 5 (Agree strongly), indicate how much you agree or disagree with each of the following statements.

I am someone who...

1. Is outgoing, sociable

2. Has an assertive personality

3. Rarely feels excited or eager (*reverse*)

4. Tends to be quiet (*reverse*)

5. Is dominant, acts as a leader

6. Is sometimes shy, introverted (*reverse*)

Handling Survey Data

Example in R: Aggregating Scores

```{r}
data2$extraversion <- data2 %>%
  select(item1:item6) %>%
  rowMeans()
```

	item1	item2	item3	item4	item5	item6	extraversion
1	5	4	5	5	5	4	4.666667
2	1	1	1	1	2	2	1.333333
3	3	3	3	3	4	3	3.166667
4	4	5	4	4	2	4	3.833333
5	2	3	2	2	1	1	1.833333

Factor Analysis

For measures that are used to measure multiple different constructs, **factor analysis** is a statistical analysis tool for examining **how items on the measure tend to be associated, or which items group together**.

Two main types of factor analysis:

- <u>Confirmatory Factor Analysis (CFA)</u>: Used to examine how well the interrelationships between items fit an *a priori*, theoretically motivated model. It's a **confirmatory** technique.

- <u>Exploratory Factor Analysis (EFA)</u>: No *a priori* theoretical model is used to examine how items group together. Items are grouped together based on which tend to have the highest interrelationships. As name suggests, it's an **exploratory** technique.

Example of Factor Analysis

Confirmatory Factor Analysis

- How well does the following model fit the observed relationships between items?

  - Assess fit using various fit statistics for CFA (e.g., χ^2, CFI, TLI, RMSEA)

Construct	Items
Non-norm	**nn1:** I got this clothing item because it was the first thing I saw when I walked into the store or went online to look for clothing.
	nn2: I got this clothing item on impulse without thinking very much about it.
	nn3: I got this clothing item purely by accident (for example, being sent the wrong item).
	nn4: I was unexpectedly given this clothing item as a gift.
Anti-norm	**an1:** This clothing item, or clothing similar to it, is an unconventional thing for me to wear.
	an2: This clothing item, or clothing similar to it, is unusual compared to what other people wear.
	an3: Wearing this clothing item, or clothing similar to it, rebels against expectations about what is acceptable to wear.
	an4: Getting this clothing item, or clothing similar to it, allows me to express my uniqueness.
Personal Rule	**pr1:** My decision to get this clothing item was based only on my personal preferences, meaning it was unrelated to the type of clothing I see people around me wearing.
	pr2: My decision to get this clothing item was based only on my personal preferences, meaning it was unrelated to the expectations people around me have about what clothing is acceptable to wear.
	pr3: This clothing item fits with my personal style.
	pr4: This clothing item fits with personal guidelines I have for myself about what type of clothes I should wear.
	pr5: I decided that getting this clothing item would be a frugal or cost-effective way of spending my money.
Descriptive Norm	**dn1:** This clothing item, or clothing similar to it, is in style right now.
	dn2: This clothing item, or clothing similar to it, is worn by many of my peers.
	dn3: Everybody else seems to frequently get new clothing items.
	dn4: There is no strict expectation from my peers for me to wear clothing similar to the clothing item that I got.
	dn5: Wearing this clothing item will help me successfully achieve one or more of my goals (i.e., career, academic, financial).
	dn6: I got this clothing item for a specific event that I want to make a good impression on others at.
	dn7: Lots of people get clothing similar to the clothing item I got because it is a frugal or cost-effective way of spending money.
Convention	**conv1:** For a long time, people will wear clothing similar to the style of the clothing item that I got.
	conv2: Even if people expect me to wear this type of clothing, no one would judge me if I failed to wear clothing similar to the clothing item that I got.
	conv3: Other people expect me to wear clothing similar to the clothing item that I got.
	conv4: I expect my peers to wear clothing similar to the clothing item that I got.
	conv5: It seems like these days people are expected to frequently get new clothing items.
Social Norm	**sn1:** People would judge me if I failed to wear clothing similar to the clothing item that I got.
	sn2: People would tell me that I am failing to dress appropriately if I failed to wear clothing similar to the clothing item that I got.
	sn3: I got this clothing item for events that I would feel impolite at if I did not wear something similar to the clothing item that I got (e.g., for work, a wedding, a job interview, etc.).
	sn4: I got this clothing item for events that I would feel guilty or shameful at if I did not wear something similar to the clothing item that I got (e.g., for work, a wedding, a job interview, etc.).
	sn5: I find that most of my peers approve of people frequently getting new clothing.
Moral Norm	**mn1:** I think that getting this clothing item or clothing similar to it is good for the economy.
	mn2: I get new clothing items that are recycled (e.g., from a secondhand store, hand-me-downs) because this is good for the environment.
	mn3: Failing to wear this clothing item or clothing similar to it would go against my principles.
	mn4: I feel a moral obligation to wear clothing similar to the clothing item that I got.
	mn5: I got this clothing item to support companies with ethical positions that I agree with.

Example of Factor Analysis

Exploratory Factor Analysis

- How do the items tend to group together without providing an a priori imposed model?
  - Up to the researcher to decide what latent construct is being measure by each group of items

We'll come back to EFA/CFA when we discuss structural equation modeling.

Items	PC1 SN (F)	PC2 DN / Conv	PC3 PR	PC4 SN (I)	PC5 MN	PC6 NN / AN	PC7 Fr
sn4: I got this clothing item for events that I would feel guilty or shameful at if I did not wear something similar to the clothing item that I got (e.g., for work, a wedding, a job interview, etc.).	0.85						
dn6: I got this clothing item for a specific event that I want to make a good impression on others at.	0.79						
sn3: I got this clothing item for events that I would feel impolite at if I did not wear something similar to the clothing item that I got (e.g., for work, a wedding, a job interview, etc.).	0.76						
dn5: Wearing this clothing item will help me successfully achieve one or more of my goals (i.e., career, academic, financial).	0.71						
mn4: I feel a moral obligation to wear clothing similar to the clothing item that I got.	0.36			0.28			
dn2: This clothing item, or clothing similar to it, is worn by many of my peers.		0.76					
dn1: This clothing item, or clothing similar to it, is in style right now.		0.71					
conv1: For a long time, people will wear clothing similar to the style of the clothing item that I got.		0.59					
conv3: Other people expect me to wear clothing similar to the clothing item that I got.		0.52	0.21	0.40			
conv4: I expect my peers to wear clothing similar to the clothing item that I got.		0.50			0.22		
an2: This clothing item, or clothing similar to it, is unusual compared to what other people wear. (negative)		-0.39	0.27		0.37	0.23	
pn2: My decision to get this clothing item was based only on my personal preferences, meaning it was unrelated to the expectations people around me have about what clothing is acceptable to wear.			0.68				
pn1: My decision to get this clothing item was based only on my personal preferences, meaning it was unrelated to the type of clothing I see people around me wearing.			0.64				
pn3: This clothing item fits with my personal style.		0.28	0.59				
pn4: This clothing item fits with personal guidelines I have for myself about what type of clothes I should wear.		0.24	0.57			-0.23	
an4: Getting this clothing item, or clothing similar to it, allows me to express my uniqueness.			0.52			0.27	
conv2: Even if people expect me to wear this type of clothing, no one would judge me if I failed to wear clothing similar to the clothing item that I got. (negative)		0.32		-0.61		0.20	
dn4: There is no strict expectation from my peers for me to wear clothing similar to the clothing item that I got. (negative)			0.25	-0.61			
sn1: People would judge me if I failed to wear clothing similar to the clothing item that I got.	0.34			0.49			
mn3: Failing to wear this clothing item or clothing similar to it would go against my principles.	0.27			0.44	0.22	0.24	
sn2: People would tell me that I am failing to dress appropriately if I failed to wear clothing similar to the clothing item that I got.	0.33			0.43			
mn1: I think that getting this clothing item or clothing similar to it is good for the economy.				-0.22	0.71		
mn5: I got this clothing item to support companies with ethical positions that I agree with.					0.63		
nn4: I was unexpectedly given this clothing item as a gift.			-0.31		0.60		
nn2: I got this clothing item on impulse without thinking very much about it.						0.78	
nn1: I got this clothing item because it was the first thing I saw when I walked into the store or went online to look for clothing.	0.26					0.46	0.24
an1: This clothing item, or clothing similar to it, is an unconventional thing for me to wear.	0.20	-0.25				0.43	
an3: Wearing this clothing item, or clothing similar to it, rebels against expectations about what is acceptable to wear.	0.35	-0.24				0.43	
nn3: I got this clothing item purely by accident (for example, being sent the wrong item).	0.20				0.23	0.29	0.32
pn5: I decided that getting this clothing item would be a frugal or cost-effective way of spending my money.							0.83
dn7: Lots of people get clothing similar to the clothing item I got because it is a frugal or cost-effective way of spending money.							0.83

Research Methods

Review of Research Methods

Research methods refers to how the data is collected for study.

- Although this step comes prior to data analysis, how the data is collected has important implications for what conclusions a researcher can draw based on the results of their analysis.

Three main research designs:

- Experimental Design

- Quasi-Experimental Design

- Correlational Design

Experimental Design

An experimental design is used to test the *causal* effect of one variable on another variable.

- Researchers manipulate which level of an independent variable (IV) participants are assigned to and then measure participants' scores on a dependent variable (DV).

Key Features of an Experimental Design

- The researcher **manipulates** the IV by systematically changing its levels between conditions while keeping all other variables constant.

- The researcher **randomly assigns** participants to a level of the IV.

Experimental Design

Example: A researcher wants to test whether keeping a journal affects people's mental health. The researcher recruits 100 participants. They randomly assign 50 participants to write in a journal every day for two weeks and the other 50 participants to not write in a journal for the next two weeks. At the end of the two-week period, the researcher measures and compares the average stress levels of both groups.
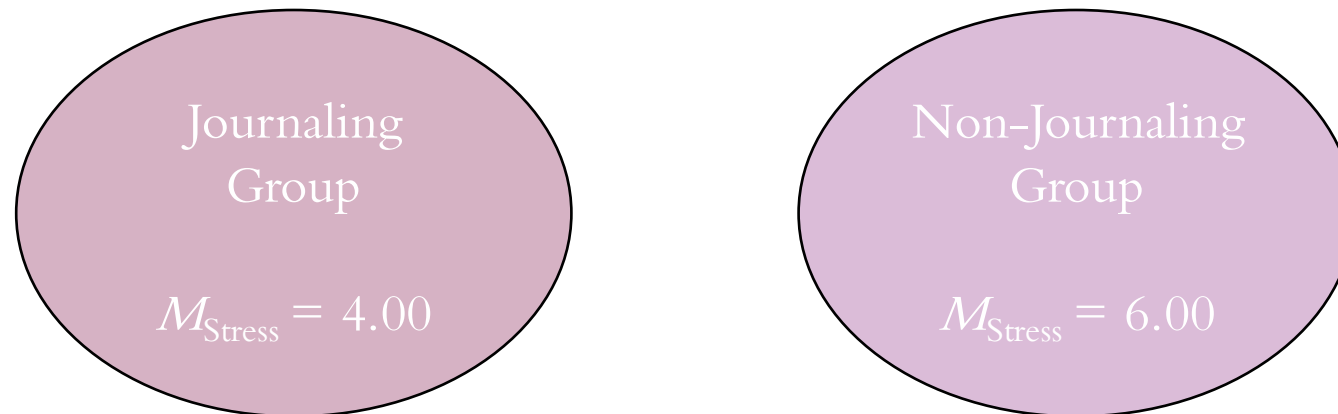
Journaling Group

$M_{Stress} = 4.00$

Non-Journaling Group

$M_{Stress} = 6.00$

Say the researcher found that stress levels in the journaling group were significantly lower than they were in the non-journaling group.

- Why is it that the researcher can additionally make the claim that journaling **caused** stress levels to be lower?

Criteria for Causal Claims

1. Association
   - Demonstrate that changes in the IV are associated with changes in the DV

2. Temporal precedence
   - Demonstrate that the IV occurs temporally prior to the DV

3. Ruling out all possible confounds so they cannot be proposed as alternative explanations
   - **Confounds:** anything that varies between groups other than the independent variable that could have had an effect on the dependent variable (also called third variables).

Quasi-Experimental Design

Similar to an experimental design, but the independent variable cannot be manipulated by the researcher.

- The levels of the IV are chosen by the researcher, but participants' already-existing scores on the IV are measured as well as their scores on the DV.

A quasi-experimental design is often used when the nature of the IV does not lend itself to being manipulated, either for **practical** or **ethical** reasons.

- Example: A researcher wants to test the effect of age group on cognitive abilities.
- Example: A researcher wants to test the effect of smoking on pregnancy outcomes.

Quasi-Experimental Design

Example: A researcher wants to test whether keeping a journal affects people's mental health. The researcher recruits 100 participants and asks them whether they write in a journal regularly or not. Based on their responses, the researcher assigns them to either the journaling group or the non-journaling group. Then, the researcher measures and compares the average stress levels of both groups.
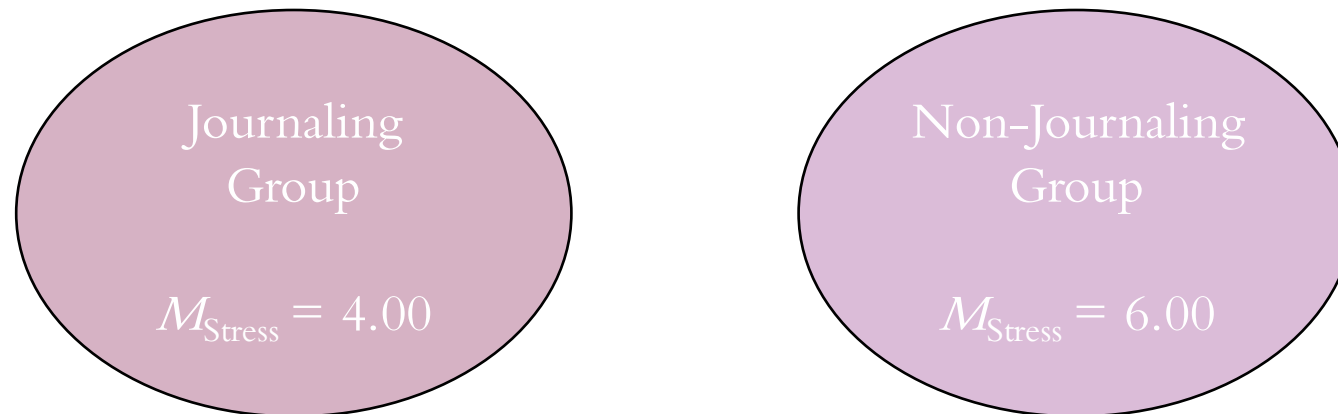
Journaling Group

$M_{Stress} = 4.00$

Non-Journaling Group

$M_{Stress} = 6.00$

Say the researcher found that stress levels in the journaling group were significantly lower than they were in the non-journaling group.

- Can the researcher also make the claim that journaling **caused** stress levels to be lower?

Criteria for Causal Claims

1. Association

2. Temporal precedence

3. Ruling out all possible confounds so they cannot be proposed as alternative explanations

A **quasi-experimental design** can only establish 1) association and 2) temporal precedence.

Correlational Design

A correlation design is used to study whether there is an association between variables.

- The researcher **measures** participants' scores on the predictor variable(s) and outcome variable.

Example: Say a researcher is interested in whether the time students spend studying for an exam is related to their exam scores. The researcher measures these two variables for seven participants and finds a significant, positive relationship between the two.

Why should the researcher **not draw any causal claims** based on these results?