Integrative Course in Psychology

Pscyhological Assessment

Prepared by:
John Carl Docog, RPm

Topics to discuss:

Psychometric Principles & Properties Research Methods & Statistics	First Meeting
Uses, Benefits, Limitations of Assessment Tools and Benefits Selection of Assessment Methods and Tools	Second Meeting
Test Administration, Scoring, Interpretation and Usage Ethical Principles and Standards of Practices	Third Meeting

Psychological Testing vs Pschological Assessment

Psychological Testing - the process of measuring psychology-related variables by means of devices or procedures designed to obtain a sample behavior.

Psychological Assessment - gathering and integrating of psychology-related data for the purpose of making a psychological evaluation.

	Testing	Assessment
Objective	Quantify behavior	Answer a referral question
Outcome	Test scores, summarized in a psychometric report	Answer to the referral question, reported in a psychological report
Focus	Understand how one person or group compares with each other **Nomothethic Approach:** focuses on the comparison between the client and other samples.	Emphasize the uniqueness of a given individual, group, or situation. **Idiographic Approach:** focuses on what makes the client unique.
Process	Can be done individually or by group	Typically individualized
Duration	Shorter, lasting from few minutes to few hours	Longer, lasting from a few hours to a few days or more.
Sources of Data	Test taker	Multiple sources
Role of Evaluator	Tester is not the key to the process; they can be replaced	Assessor is the key to the process.

Assumptions of Psychological Assessment

1. Psychological traits & states exist
2. Psychological traits & states can be quantified & measured
3. Test-related behavior predicts non-test-related behavior
4. Test and other measurement techniques have strengths and weaknesses
5. Various sources of error are part of the assessment process
6. Testing & assessment can be conducted in a fair and unbiased manner
7. Testing & assessment benefit the society

LAKAS NG LOOB: Lived Experiences of Lesbian Firefighters in a Traditionally Male-Dominated Profession

A Thesis

4. Sources of *'lakas ng loob'*	4.1	Faith	4.1.1 Risk taking 4.1.2 Innate
	4.2	Sense of altruism	4.2.1 Community engagement 4.2.2 Application of professional knowledge 4.2.3 Resiliency 4.2.4 Necessity for work and survival
	4.3	Defying the odds	4.3.1 Fearing probable occurrence 4.3.2 Bowing to fate

February 20 - 26, 2025 / Philippines
(Multiple Responses, up to 12 names allowed)

Base: Total Interviews, 100%*

	Aware	Voting for	Rank
GO, BONG GO (PDPLBN)	100	58.1	1-2
TULFO, ERWIN (LAKAS)	98	56.6	1-2
SOTTO, TITO (NPC)	100	49.0	3-4
BONG REVILLA, RAMON, JR. (LAKAS)	100	46.1	3-6
DELA ROSA, BATO (PDPLBN)	100	44.3	4-7
REVILLAME, WILLIE WIL (IND)	98	42.3	4-9
TULFO, BEN BITAG (IND)	97	40.7	5-11
PACQUIAO, MANNY PACMAN (PFP)	100	39.9	6-12
LAPID, LITO (NPC)	100	39.4	6-13
BINAY, ABBY (NPC)	98	37.6	7-13
CAYETANO, PIA (NP)	100	37.5	7-13
VILLAR, CAMILLE (NP)	96	36.6	8-13
LACSON, PING (IND)	97	35.8	9-13
MARCOS, IMEE R. (NP)	100	30.9	14
AQUINO, BAM (KNP)	95	26.4	15-17
PANGILINAN, KIKO (LP)	99	25.0	15-17
ABALOS, BENHUR (PFP)	89	23.0	15-19
SALVADOR, PHILLIP IPE (PDPLBN)	92	20.7	17-20

Reliability

Reliability refers to **CONSISTENCY** in MEASUREMENT

Reliability Coefficient
- An index of reliability, a proportion that indicates the ratio between the true score variance on a test and the total variance

True Variance – variance from true differences
Error Variance - variance from irrelevant, random sources.

Measurement Error

- All of the factors associated with the process of measuring some variable, other than the variable being measured

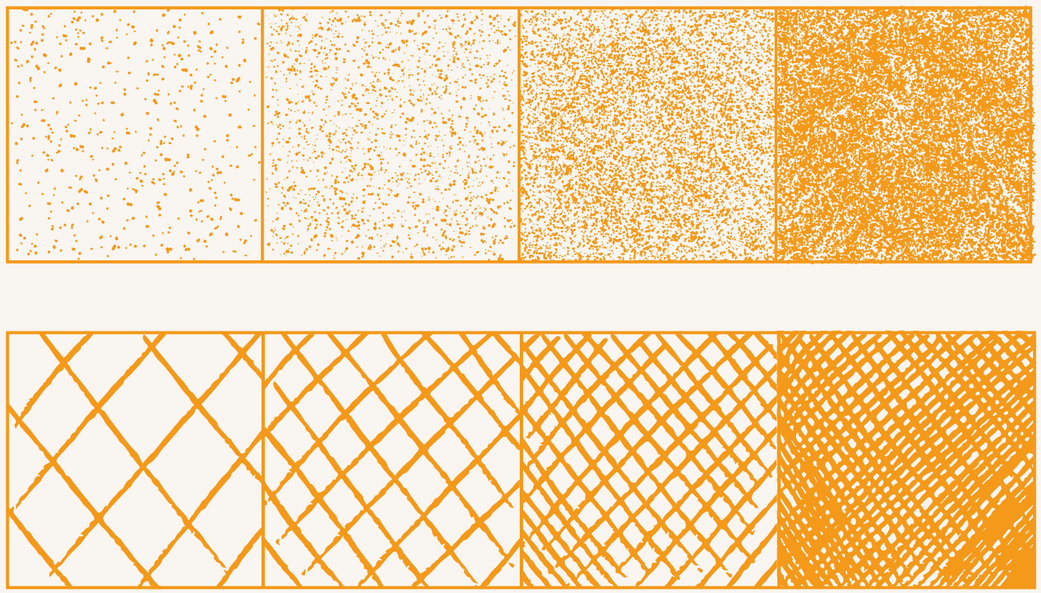
- **Random error/Noise**

o Source of error in measuring a targeted variable caused by unpredictable fluctuations and inconsistencies of other variables in the measurement process

- **Systematic error**

o Source of error in measuring a variable that is typically constant or proportionate to what is presumed to be the true value of the variable being measured

Sources of Error Variance



- Test Construction
- Test Administration

1. Test Environment
2. Test-taker variables
3. Examiner-related variables

- Test Scoring & Interpretation

Reliability Estimates

1. Test-Retest Reliability

- Using the same instrument to measure the same thing as two points in time
- The results of evaluation is called test-retest reliability
- 1 group; 2 different administration
- Measure something that is relatively stable over time such as personality
- Coefficient of Stability - interval between testing is 6 months
- Test-retest is also appropriate in reaction time and perceptual judgment

2. **Parallel Forms and Alternate Forms Reliability**

Parallel Forms

- Each form of the test, the means and the variances of observed test scores are equal

Alternate Forms

- Simply different versions of a test that have been constructed so as to be parallel
- Developing alternate forms of tests can be time consuming and expensive
- 1 group; 2 different administrations

Reliability Estimates

3. **Inter-item/Internal Consistency**

- Single administration; single form
- This relates to the reliability of a test or measurement tool. It evaluates whether the items within a test consistently measure the same construct or concept.
- Assessing test of HOMOGENEITY
- **Homogenous** – items in a scale are unifactorial
- **Heterogeneous** – composed of items that measure more than one trait
- ***MORE HOMOGENOUS; MORE INTER-ITEM CONSISTENCY***

Reliability Estimates

3. **Inter-item/Internal Consistency**

Split-Half Reliability

- Correlating two pairs of scores obtained from equivalent halves of a single test administered once

Kuder-Richardson Formulas

- Developed by G. Frederic Kuder and M.W. Richardson

KR-20

- Homogenous Items
- Dichotomous Items (“right or wrong” items)

KR-21

- It is used for a test where the items are all about the same difficulty

Reliability Estimates

3. **Inter-item/Internal Consistency**

Coefficient Alpha

- Developed by Cronbach
- Homogenous Items
- Non-dichotomous items
- It is the preferred statistic for obtaining an estimate of internal consistency reliability
- Values ranges from (No Similarity 0.00 – Perfectly Identical 1.00)

Reliability Estimates

Sa **UP Questionable** ang **AGE**

No	Coefficient of Cronbach's Alpha	Reliability Level
1	More than 0.90	Excellent
2	0.80-0.89	Good
3	0.70-0.79	Acceptable
4	0.6-.69	Questionable
5	0.5-0.59	Poor
6	Less than 0.59	Unacceptable

Reliability Estimates

4. **Inter-scorer/Interrater Reliability**

- refers to the degree of agreement or consistency between two or more independent scorers, raters, or observers when they evaluate or score the same set of items, behaviors, or responses.
- **Cohenn's Kappa** – 2 raters
- **Fleiss' Kappa** – 3 or more raters

Table 5–1
Summary of Reliability Types

Type of Reliability	Purpose	Typical uses	Number of Testing Sessions	Sources of Error Variance	Statistical Procedures
Test-retest	To evaluate the stability of a measure	When assessing the stability of various personality traits	2	Administration	Pearson *r* or Spearman rho
Alternate-forms	To evaluate the relationship between different forms of a measure	When there is a need for different forms of a test (e.g., makeup tests)	1 or 2	Test construction or administration	Pearson *r* or Spearman rho
Internal consistency	To evaluate the extent to which items on a scale relate to one another	When evaluating the homogeneity of a measure (or, all items are tapping a single construct)	1	Test construction	Pearson *r between equivalent test halves with Spearman Brown correction or Kuder-Richardson for dichotomous items, or coefficient alpha for multipoint items*
Inter-scorer	To evaluate the level of agreement between raters on a measure	Interviews or coding of behavior. Used when researchers need to show that there is consensus in the way that different raters view a particular behavior pattern (and hence no observer bias).	1	Scoring and interpretation	Cohen's kappa, Pearson *r*, or Spearman rho

The True Score Model of Measurement and Alternatives

Classical Test Theory

- focused on understanding the relationship between an individual's true score (the actual measure of their ability or trait), the observed score (their test result), and the error (random or systematic factors that influence the score)

$$X = T + E$$

Where:

- X: Observed Score (what we see in the test result)
- T: True Score (the individual's real ability or trait)
- E: Error Score (random fluctuations affecting the measurement)

The True Score Model of Measurement and Alternatives

Domain Sampling Theory

- Reliability, in the context of Domain Sampling Theory, refers to the degree to which a test's observed scores reflect true scores, regardless of the specific sample of items used.
- The theory supports the idea that longer tests tend to have higher reliability because they include more items and, therefore, better represent the domain as a whole.

The True Score Model of Measurement and Alternatives

Generabilizity Theory

- Generalizability Theory (G-Theory) is a modern and flexible framework for evaluating the reliability of measurements. It expands on Classical Test Theory (CTT) by addressing its limitations, offering a more nuanced understanding of measurement error and reliability.
- based on the idea that a person’s test scores vary from testing to testing because of the variables in the testing situations
- Universe: test situation
- Facets: number of items in the test, amount of review, and the purpose of test administration

The True Score Model of Measurement and Alternatives

Item-Response Theory

- Latent-Trait Theory
- a system of assumption about measurement and the extent to which item measures the trait
- The computer is used to focus on the range of item difficulty that helps assess an individual’s ability level

Theory	Main Idea	Strengths	Limitations
CTT	Test score = true score + error. Focuses on total test score reliability.	Easy to use; widely applied.	Assumes all items are equally important; less precise for items.
DST	Tests are samples from a larger "domain" of possible items.	Highlights how well test items represent the full domain.	Limited by the representativeness of sampled items.
G-Theory	Explores multiple sources of error (e.g., raters, items) and their impact on reliability.	Analyzes reliability in different conditions.	Requires complex statistical methods.
IRT	Examines individual item performance and its relation to underlying traits (e.g., ability).	Provides precise measurement; supports adaptive testing.	Needs large data sets and computational resources.

Standard Error of Measurement

- way to estimate how much a person's observed test score might vary due to random errors.
- A smaller SEM means the test is more reliable, while a larger SEM indicates less reliability.

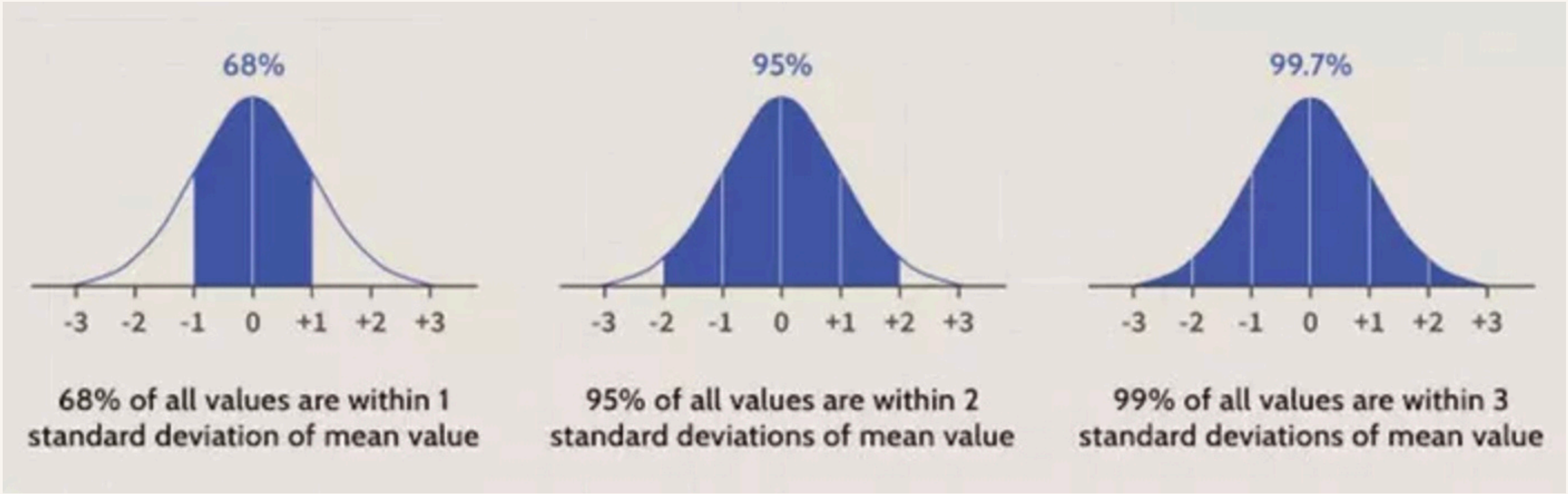
The Formula for SEM:

The formula for SEM is:

$$SEM = SD \times \sqrt{1 - r}$$

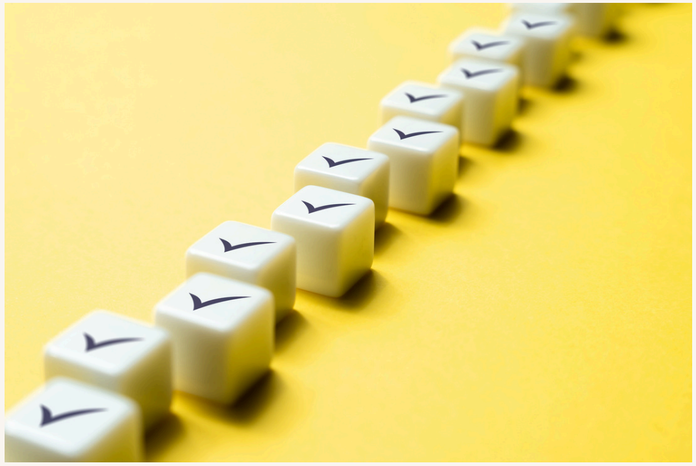
Where:

- **SD** = Standard deviation of test scores.
- **r** = Reliability of the test (a number between 0 and 1).


68%
-3 -2 -1 0 +1 +2 +3
68% of all values are within 1 standard deviation of mean value
95%
-3 -2 -1 0 +1 +2 +3
95% of all values are within 2 standard deviations of mean value
99.7%
-3 -2 -1 0 +1 +2 +3
99% of all values are within 3 standard deviations of mean value

The Concept of Validity

- Estimate of how well a test measures what it purports to measure in a particular context
- NO TEST is universally valid for all time
- The validity of a test must be proven again from time to time
- Determines the accuracy and appropriateness of the inferences drawn from test scores.



Trinitarian View

- Content Validity
- Criterion-related Validity
- Construct Validity

Types of Validity

1. Face Validity: This reflects whether the test appears to measure what it claims to, based on subjective judgment.

- **HIGH FV** – Structured Personality test
- **LOW FV** – Projective Test
- ***Lack of FV may result to. . . . .***
  - Decrease level of cooperation or motivation of testtaker
  - Might not get “buy in” value for test user

Types of Validity

2. Content Validity: This examines whether the test comprehensively covers the domain or content it is supposed to measure.

- More logical than statistical
- Panel of experts can review the test items
- **Construct underrepresentation**: failure to capture important components of a construct
- **Construct-irrelevant variance**: happens when scores are influenced by factors irrelevant to the construct
- Test Blueprint

Types of Validity

3. Criterion-related Validity: a judgement of how adequately a test score can be used to infer an individual's most probable standing on some measure of interest – the measure of interest being criterion

- More statistical than logical
- **Criterion**: standard on which a judgement or decision may be made
- **Criterion Contamination**: occurs when the criterion measure includes aspects of performance that are not part of the job or when the measure is affected by "construct-irrelevant"
- **Predictive Validity**: Assesses the test's ability to predict future performance
- **Concurrent Validity**: Examines the test's relationship with a criterion measured at the same time

Statistical Evidence

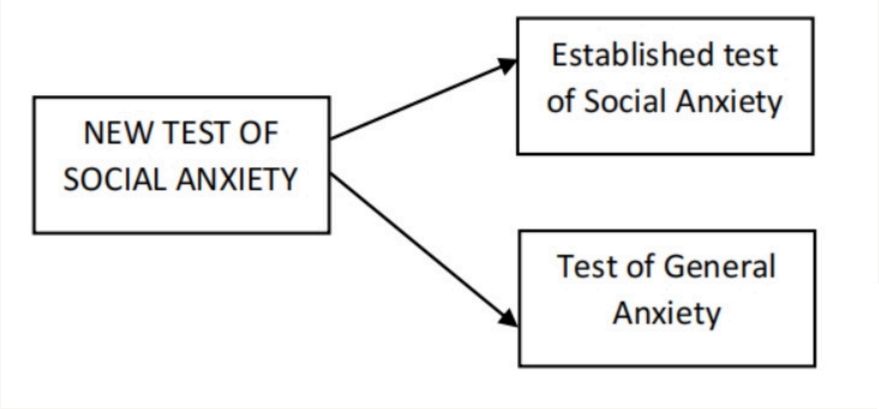
- **Validity Coefficient -** A correlation that provides a measure of the relationship between test scores and scores on the criterion measure
- **Incremental Validity -** The degree to which an additional predictor explains something about the criterion measure that is not explained by predictors already in use

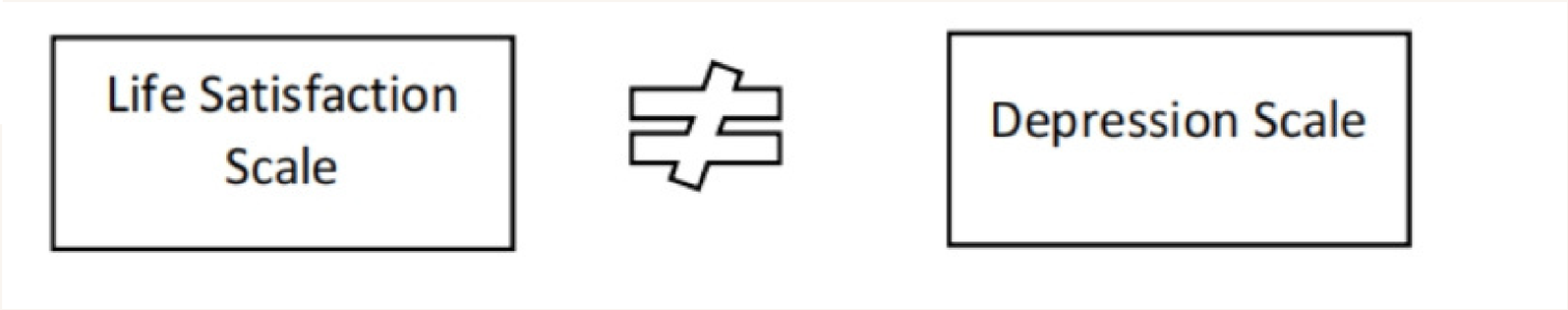
Types of Validity

4. Construct Validity: This addresses whether the test truly measures the theoretical construct or concept it claims to measure.

- covers all types of validity
- Logical and statistical
- Judgement about the appropriateness of inferences drawn from test scores regarding individual standing on variable called construct
- **Construct**: an informed, scientific idea developed or hypothesized to describe or explain behavior; unobservable, presupposed traits that may invoke to describe test behavior or criterion performance

- **Convergent Evidence**: if scores on the test undergoing construct validation tend to highly correlated with another established, validated test that measures the same construct
- **Discriminant Evidence**: a validity coefficient showing little relationship between test scores and/or other variables with which scores on the test being construct-validated should not be correlated





Factor Analysis

- Also helpful in obtaining convergent and discriminant validity
- Class of mathematical procedures designed to identify factors or specific variables that are typically attributes, characteristics, or dimensions on which people may differ
- **Exploratory Factor Analysis**
  - “estimating, or extracting factors; deciding how many factors to retain; and rotating factors to an interpretable orientation”
- **Confirmatory Factor Analysis**
  - Researchers test the degree to which a hypothetical model (which includes factors)fits the actual data
- **Factor Loading**
  - The extent to which the factor determines the test score or scores

Term	Definition	Purpose	Example
Factor Analysis	A statistical method used to identify underlying relationships between variables (factors).	Reduces data complexity and groups variables.	A psychological survey groups questions on mood, energy, and sleep into broader factors like "well-being."
Exploratory Factor Analysis (EFA)	A type of factor analysis used when the structure or number of factors is not predetermined.	Explores potential factor structures.	Researchers analyzing survey data to discover hidden factors without prior assumptions.
Confirmatory Factor Analysis (CFA)	A type of factor analysis used to test if a hypothesized factor structure fits the data.	Confirms the validity of a predefined model.	Testing whether pre-identified factors for emotional intelligence align with survey data.
Factor Loading	A measure of how strongly each variable is associated with a factor.	Shows the strength of the relationship.	A high loading (e.g., 0.8) for a question on "happiness" suggests a strong link to a "positive emotions" factor.

Validity Bias & Fairness

- **Bias** - a factor inherent in a test that systematically prevents accurate, impartial measurement
- **Rating Error**
  - Rating - a numerical or verbal judgment (or both) that places a person or an attribute along a continuum
  - Rating Scale - a scale of numerical or word descriptors
  - Rating Error - a judgment resulting from the intentional orunintentional misuse of a rating scale

- **Leniency Error:** rater is lenient in scoring (Generosity Error)
- **Severity Error:** rater is strict in scoring
- **Central Tendency Error:** rater's rating would tend to cluster in the middle of the rating scale
- **Halo Effect**: tendency to give high score due to failure to discriminate among conceptually distinct and potentially independent aspects of a
- ratee's behavior
- **Fairness** – the extent to which a test is used in an impartial, just, and equitable way
- Attempting to define the validity of the test will be futile if the test is NOT reliable

Validity coefficient value	Interpretation
above .35	very beneficial
.21 - .35	likely to be useful
.11 - .20	depends on circumstances
below .11	unlikely to be useful

- Validity coefficients typically range from -1.0 to +1.0.
- A value closer to +1.0 indicates a strong positive relationship.
- A value closer to 0.0 suggests little to no relationship.
- A value closer to -1.0 indicates a strong negative relationship (less common in validity studies).
- 0.30 to 0.50: Considered moderate validity.
- Above 0.50: Indicates strong validity.
- Higher coefficients mean the test is more reliable for its intended purpose.

Research Methods & Statistics

Research

- **Research** - in psychology is a vital tool for understanding human behavior, thoughts, and emotions.
- Research Methods
  - **Quantitative Research -** Focuses on numerical data and statistical analysis to understand phenomena.
  - **Qualitative Research** - Explores subjective experiences and aims to understand meaning or concepts.
  - **Mixed-Methods Research** - Combines both qualitative and quantitative approaches.

Research Design

1. **Descriptive** - Describes characteristics of a population or phenomenon.
2. **Correlational Design** - Investigates relationships between two or more variables.
3. **Experimental Design** - Tests cause-and-effect relationships by manipulating variables.
4. **Quasi-Experimental Design** - Similar to experimental design but lacks random assignment.
5. **Longitudinal Design** - Observes the same subjects over an extended period.
6. **Cross-Sectional Design** - Examines different groups at one point in time.

Research Design	Key Points	Differences	Examples
Descriptive Design	Describes characteristics or phenomena; answers "what" questions.	Does not explore relationships or causality.	Study describing demographics of university students.
Correlational Design	Examines relationships between variables; determines strength and direction.	Focuses on relationships, not causality.	Investigating if stress levels are associated with academic performance.
Experimental Design	Tests cause-and-effect by manipulating independent variables.	Establishes causation due to controlled manipulation.	Testing effects of a new drug on blood pressure using treatment and placebo groups.

Quasi-Experimental Design	Similar to experimental but lacks random assignment.	Lacks randomization, leading to potential bias.	Studying effects of a teaching method in classrooms assigned by availability.
Longitudinal Design	Observes same subjects over time; tracks changes and developments.	Focuses on changes over time, unlike cross-sectional.	Tracking cognitive development of children over 10 years.
Cross-Sectional Design	Observes different groups at a single point in time; provides a snapshot.	Captures data at one moment, unlike longitudinal studies.	Survey comparing physical activity levels of teenagers and adults at a particular time.

Sampling and it's type

- **Sampling** - the process of selecting the portion of the universe deemed to be representative of the whole population
- **Population** – the complete universe or set of individuals with at least one common observable characteristic
- **Sample** – a portion of the universe deemed to be representative of the whole population

Methods

1. **Stratified Sampling**
    - Reduces bias
    - Members of the sample came from different strata
2. **Random Sampling**
    - Every member of the population has the same chance of being included in the sample
3. **Purposive Sampling**
    - Arbitrarily select some sample because it is believe that it will be the best to represent the population
    - Common in Consumer Psychology
4. **Incidental or Convenience Sampling**
    - Practicality
    - It is done due to budgetary limitations or other constraints

Statistics

Measurement – the act of assigning numbers or symbols to characteristics of things according to rules

Psychological Statistics

Descriptive Statistics	Inferential Statistics
Summarize and describe relatively basic but essential features of a quantitative dataset	Help one make conclusions and predictions based on data.

Continuous Scale

- takes on any value within therange and the possible value within that range is infinite
- used to measure a variable which can theoretically bedivided

Discrete Scale

- can be counted; has distinct, countable values
- used to measure a variable which cannot be theoretically be divided

Scale of Measurement

1. **Nominal**
   - the category labels are not ordered, so it doesn’t matter which number comes first.
2. **Ordinal**
   - in the ordinal scale of data, there is an order. However, the difference between them can not be quantified.
3. **Interval**
   - in the interval scale, we do have an order (just like ordinal data), and we can find the exact difference between the two values
4. **Ratio**
   - the ratio scales has all the features of the interval scale, and in addition there is an absolute or true zero as well.

Measures of Central Tendency

1. **Mean**
   - **Average**; sum of the observations or test scores divided by the number of observations or scores.
   - *Appropriate for: interval and ratio & when the distribution is approximately normal*
2. **Median**
   - **Middle score** in the distribution
   - Appropriate for: ordinal, interval, and ratio
   - *Used when the distribution is skewed and when there are outliers*
3. **Mode**
   - The **most frequently occurring score**
   - Can have more than one mode
   - *The only central tendency that can be used in nominal/categorical data*

Measures of Central Tendency

1. **Mean**
   - **Average**; sum of the observations or test scores divided by the number of observations or scores.
   - *Appropriate for: interval and ratio & when the distribution is approximately normal*
2. **Median**
   - **Middle score** in the distribution
   - Appropriate for: ordinal, interval, and ratio
   - *Used when the distribution is skewed and when there are outliers*
3. **Mode**
   - The **most frequently occurring score**
   - Can have more than one mode
   - *The only central tendency that can be used in nominal/categorical data*

Measures of Variability

- Gives information on the spread or variability of data values
- They give insight into the degree of fluctuation or consistency in the data

1. **Range**:
    - The simplest measure of variability.
    - Calculated as the difference between the highest and lowest values in the dataset.
2. **Variance**:
    - Measures how far the data points are from the mean, on average.
    - It is the average of the squared differences between each data point and the mean.
    - A high variance indicates greater spread in the data.

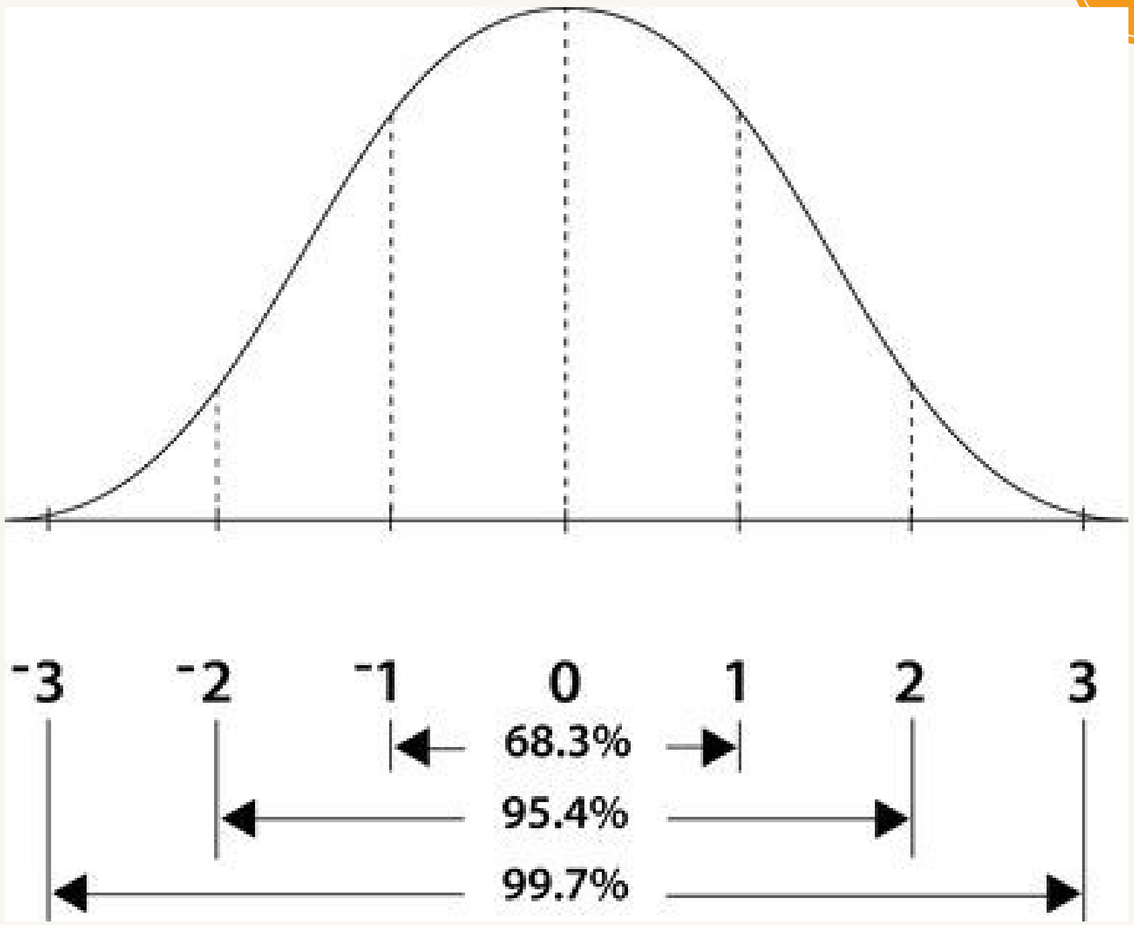
Measures of Variability

3. **Quartile** – dividing points between the four quarters in the distribution
   - Specific point
   - Quarter: refers to an interval
   - **Interquartile Range**: measure of variability equal to the difference between Q3 and Q1
   - **Semi-interquartile Range**: equal to the interquartile range divided by 2

4. **Standard Deviation** – equal to the square root of the average squared deviations about the mean
   - Equal to the square root of the variance
   - Variance: equal to the arithmetic mean of the squares of the differences between the scores in a distribution and their mean
   - Distance from the mean

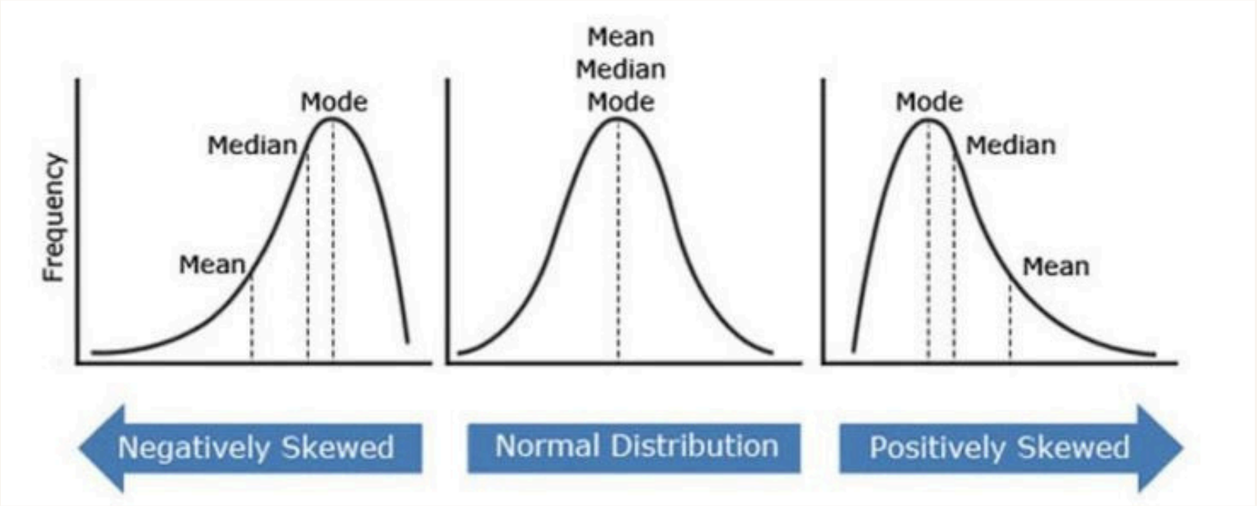
Measures of Normality

- **Symmetrical Distribution** – right side of the graph is mirror image of the left side
- Has only one mode and it is in the center of the distribution
- Mean = median = mode



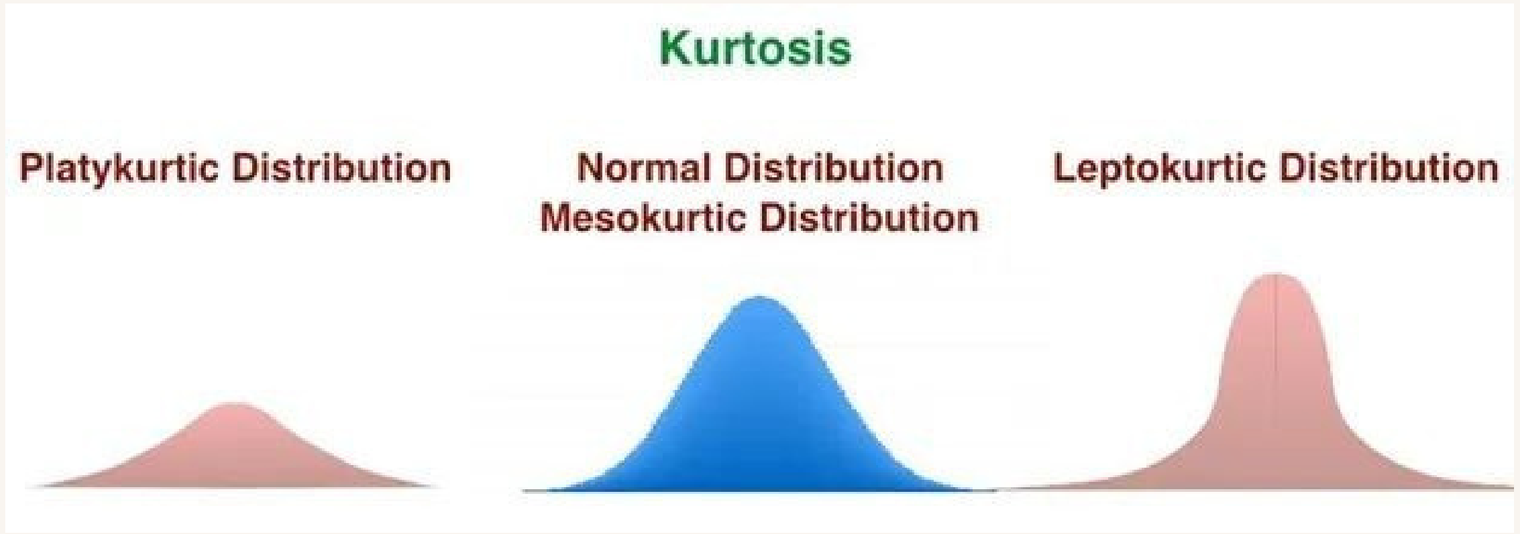
Skewnesss - nature and extent to which symmetry is absent

- **Positive Skewed** - few scores fall the high end of the distribution; Mean > Median > Mode
- **Negative Skewed** - when relatively few of the scores fall at the low end of the distribution; Mean < Median < Mode



Kurtosis - The steepness of a distribution in its center

- **Platykurtic –** A distribution with a flat peak and thinner tails compared to a normal distribution.
- **Leptokurtic –** A distribution with a sharp, high peak and fatter tails compared to a normal distribution.
- **Meso –** This type of distribution has a peak and tails similar to a normal distribution.



Standard Score

Standard Scores

- A raw score that has been converted from one scale to another scale, where the latter scale has some arbitrarily set mean and standard deviation
- More easily INTERPRETABLE than raw scores]

Z – Scores

- **Mean = 0; SD = 1**
- Is equal to the difference between a particular raw score and the mean divided by standard deviation

$$\text{Z Score} = \frac{x - \mu}{\sigma}$$

$$\text{Z Score} = \frac{\text{Raw score - Mean}}{\text{Standard deviation}}$$

- **Z-Scores** – results from the conversion of a raw score into a number indicating how many SD units the raw score is below or above the mean of the distribution
  - Standardize an entire distribution
  - Zero plus or minus one scale
  - Have negative values

- **T-Scores – a scale with a mean set at 50 and a s**tandard deviation set at 10
  - **Fifty plus or minus 10 scale**
  - 5 standard deviations below the mean would be equal to a t-score of 0
  - Raw score that fell in the mean has T of 50
  - Raw score 5 standard deviations about the mean would be equal to a T of 100
  - No negative values
  - Used when the population or variance is unknown
- **Stanine** – a method of scaling test scores on a nine□point standard scale with a mean of five (5) and a standard deviation of two (2
- **STEN** – standard to ten; divides a scale into 10 units

	Mean	SD
Z-Score	0	1
T-Score	50	10
Stanine	5	2
STEN	5.5	2
IQ	100	15
GRE or SAT	500	100

TYPES OF HYPOTHESIS	
NULL HYPOTHESIS	**ALTERNATIVE HYPOTHESIS**
States that the treatment has no effect. Says "there is no change, no effect, no difference – nothing happened, hence the name "null"	The treatment has an effect. There is a change, has effect, has difference – something happened.

ALPHA LEVEL (α)

The probability of obtaining your results due to CHANCE

A small probability that is used to identify the low-probability samples

(a)		C.I.	Common uses
5 %	*0.05*	95 %	Social science research
1 %	*0.01*	99 %	Medical and natural science research
0.01 %	*0.001*	99.99 %	

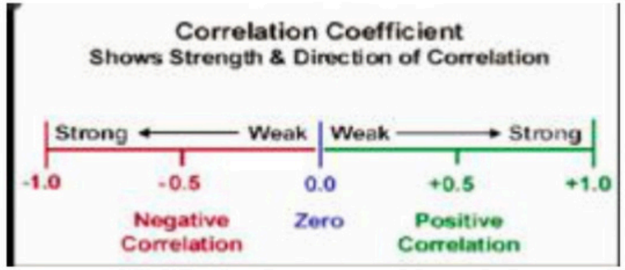
LEVENE'S TEST Tests if variances are equal among groups.			
P-VALUE	**NULL HYPOTHESIS**	**INTERPRETATION**	
Greater than 0.05	Fail to reject	**Equal ang variances** ng bawat grupo	Parametric
Less than 0.05	Reject	**Magkakaiba-iba ang variances** ng bawat grupo	Non-Parametric

Which of these?

- Will I test if there is any significant relationship among variable?
  - √ **If yes, CORRELATION**
- Will I test if a variable significantly predicts another variable?
  - √ **If yes, REGRESSION**
- Will I compare mean/median to different/same group?
  - √ **If yes, TEST OF DIFFERENCE (T test or ANOVA)**

CORRELATION

Measure of the direction (+ or -) and strength (-1 to +1) of a relationship between tow variables. When the r value is closer to +1 or -1, it indicates that there is a stronger linear relationship between the two variables



Positive Correlation	**Negative Correlation**	**No Correlation**
As one variable increases so does the other variable	As one variable increases the other variable decreases.	There is no relationship between the two variables

	Parametric	Non-parametric
Assume distribution	Normal	Any
Assume variance	Homogenous	Homogenous and Heterogenous
Typical data	Ratio or Interval	Ordinal or Nominal
Data set relationships	Independent	Any

*Kapag may pumalya isa, NON-PARAMETRIC TEST na ang gagamitin

PARAMETRIC	NON-PARAMETRIC
Pearson R	SPEARMAN RHO CORRELATION
DEPENDENT T-test	WILCOXON SIGNED RANK TEST
INDEPENDENT T-test	MANN WHITNEY U TEST
REPEATED MEASURES ANOVA	FRIEDMAN TEST
One Way / Two Way ANOVA	KRUSKAL WALLIS TEST

Hypothesis Testing

↳ educated guess

null = H_0 — sig result — H_0 is false — reject H_0

null = H_0 — insig result — H_0 is true — fail to reject H_0

alternative = H_1 — sig result — H_1 is true — fail to reject H_1

alternative = H_1 — insig result — H_1 is false — Reject H_1

$p < 0.05$ significant

$p > 0.05$ insignificant

Required Features

Parametric Test

↳ normal distribution

↳ Homogenous variance

↳ Interval / Ratio

Non-Parametric Test

↳ Normal distribution is not required

↳ Homogenous variance is not required

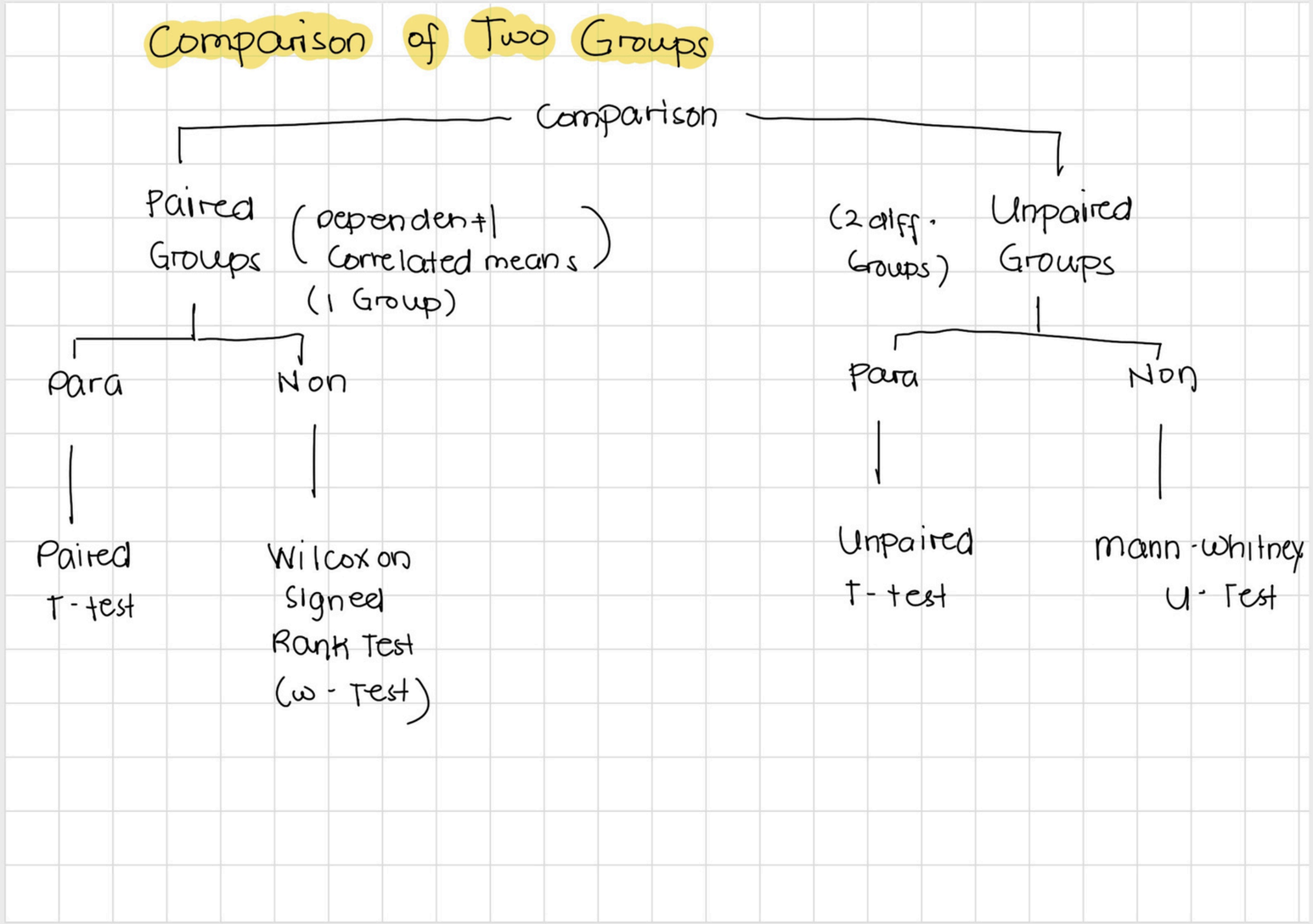
↳ Nominal / ordinal

Measures of Bivariate Correlation

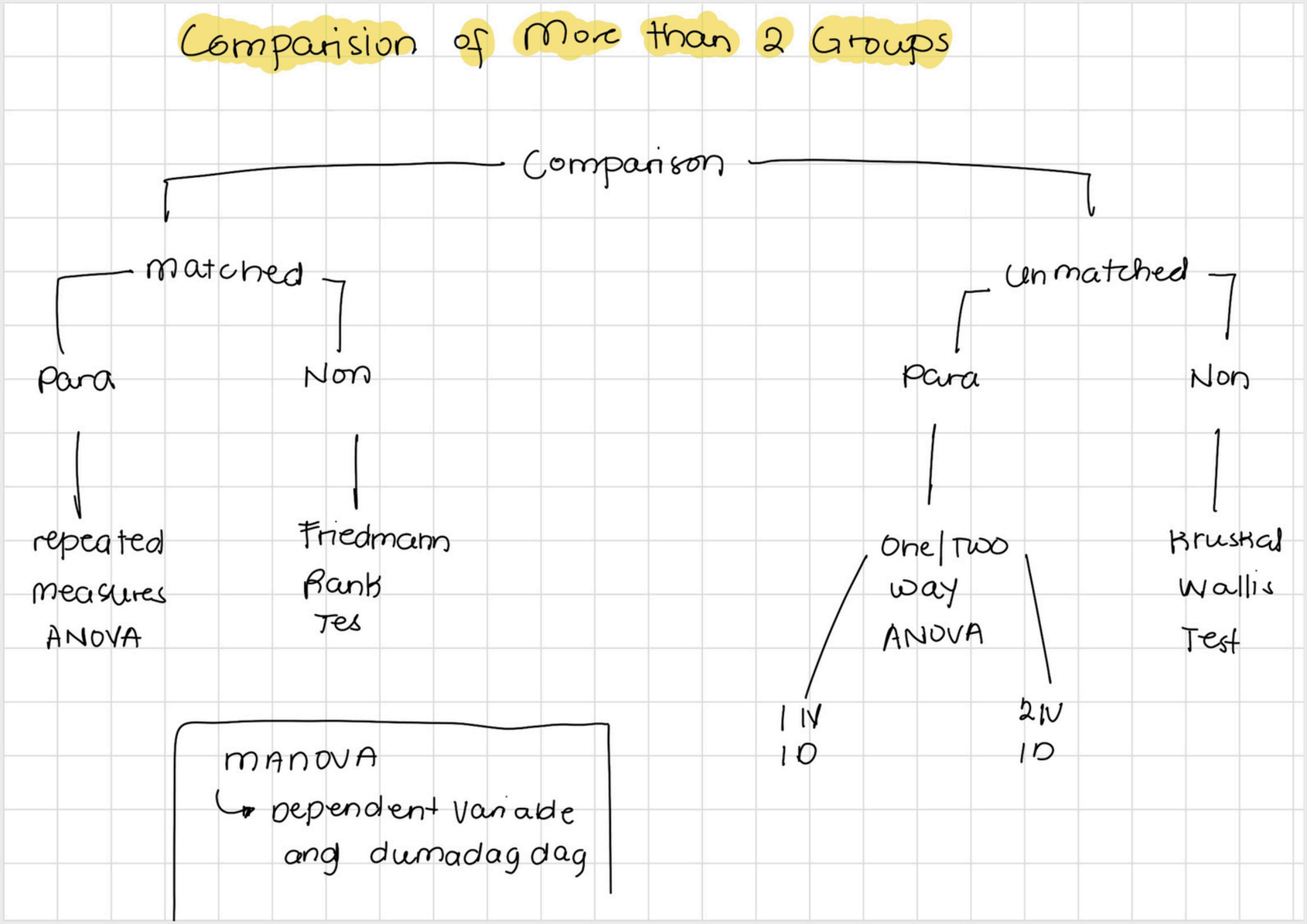
Pearson Correlation → 2 continuous / parametric variables (ratio / interval)

Spearman-Rho Correlation → 2 ordinal data set / nonparametric

- **actual correlation coefficient**
  - 1.00 perfect
  - 0.75 - 0.99 v. strong
  - 0.50 - 0.74 strong
  - 0.25 - 0.49 moderate
  - 0.01 - 0.24 weak
  - 0.00 no relationship
- **algebraic sign**
  - (+) Direct Relationship
  - (-) Inverse Relationship
- **p-value** (significance of the result)
  - 99% confidence interval (0.01)
    - $p < 0.01$ = significant
    - $p > 0.01$ = insignificant
  - 95% confidence interval (0.05)
    - $p < 0.05$ = significant
    - $p > 0.05$ = insignificant



Comparision of More than 2 Groups



Integrative Course in Psychology

The End

Prepared by:
John Carl Docog, RPm

Integrative Course in Psychology

Rationalization

Prepared by:
John Carl Docog, RPm

1. Reliability refers to:

a) The consistency of a test over time

b) The validity of a test across settings

c) The fairness of a test

d) The correlation between variables

2. Which type of validity examines whether a test measures the intended construct?

a) Content validity

b) Criterion validity

c) Construct validity

d) Predictive validity

3. The standard deviation is a measure of:

a) Central tendency

b) Variability

c) Reliability

d) Normality

4. A z-score represents:

a) The mean score of a data set

b) The standard deviation of a score

c) The number of standard deviations a score is from the mean

d) The correlation between two variables

5. Test-retest reliability is best used to measure:

a) Internal consistency

b) Temporal stability

c) Inter-rater agreement

d) Face validity

6. Cronbach's alpha measures inter-item consistency, while Cohen's Kappa measures inter-rater reliability.

a) Both True

b) Neither Both True

c) First True, Latter False

d) First False, Latter True

7. A sample is a subset of a population, designed to represent the key characteristics of that population asaccurately as possible.

a) Both True

b) Neither Both True

c) First True, Latter False

d) First False, Latter True

8. A Likert scale is an example of an ordinal measurement. Same goes with a rating scale ranging 0-10 with equal distances between numbers

a) Both True

b) Neither Both True

c) First True, Latter False

d) First False, Latter True

9. A large p-value indicates strong evidence against the null hypothesis. The p-value indicates the probability of rejecting the null hypothesis when it is true.

a) Both True

b) Neither Both True

c) First True, Latter False

d) First False, Latter True

10. Convergent validity is assessed when unrelated constructs do not correlate. Divergent validity is also known as discriminant validity.

a) Both True

b) Neither Both True

c) First True, Latter False

d) First False, Latter True

11. If a test has high reliability but low validity, what does this mean?

a) The test is inconsistent but measures the correct construct

b) The test is consistent but does not measure the correct construct

c) The test is both consistent and valid

d) The test is neither consistent nor valid

12. Which statistical test is appropriate for examining the relationship between two Continuous variables?

a) ANOVA

b) Chi-square test

c) Pearson correlation

d) Regression analysis

13. A researcher is testing whether a new teaching method improves scores. What is the independent variable?

a) Test scores

b) Teaching method

c) The students

d) The classroom setting

14. Which graph best represents the distribution of a single continuous variable?

a) Bar chart

b) Histogram

c) Scatterplot

d) Pie chart

15. A correlation coefficient of 0 indicates:

a) A strong relationship

b) No relationship

c) A negative relationship

d) A positive relationship

16. The primary goal of exploratory factor analysis (EFA) is to:

a) Confirm existing theoretical models

b) Identify underlying constructs in data

c) Test for normality

d) Predict outcomes

17. Which type of scale includes an absolute zero point?

a) Nominal

b) Ordinal

c) Interval

d) Ratio

18. The null hypothesis in research assumes:

a) No effect or relationship exists

b) A significant effect exists

c) The effect size is meaningful

d) The alternative hypothesis is true

19. The term "standard error" refers to:

a) Variability within a population

b) Variability of a sample statistic

c) Correlation between variables

d) A test's measurement error

20. In a normal distribution, approximately what percentage of data falls within one standard deviation from the mean?

a) 50%

b) 68%

c) 95%

d) 99%

21. Predictive validity is a type of content validity. Face validity is considered logical and statistical.

a) Both True

b) Neither Both True

c) First True, Latter False

d) First False, Latter True

22. The Spearman rank correlation is used for ordinal data, while Pearson's r is used mostly for interval data.

a) Both True

b) Neither Both True

c) First True, Latter False

d) First False, Latter True

23. A skewed distribution has equal mean, median, and mode. The steepness of
data may be categorized as platykurtic, mesokurtic, and leptokurtic.

a) Both True

b) Neither Both True

c) First True, Latter False

d) First False, Latter True

24. A p-value less than 0.05 typically indicates statistical significance. Therefore, the null hypothesis is rejected, and the evidence may suggest using the alternative hypothesis.

a) Both True

b) Neither Both True

c) First True, Latter False

d) First False, Latter True

25. Severity error occurs when the rater is strict in scoring; to prevent rating errors, the rater should use a ranking ratingscale.

a) Both True

b) Neither Both True

c) First True, Latter False

d) First False, Latter True

26. A test has a mean of 50 and a standard deviation of 10. What z-score corresponds to a test score of 70?

a) 1.5

b) 2.0

c) 2.5

d) 3.0

27. Which research design is most appropriate for studying the effect of a training program over time?

a) Cross-sectional study

b) Longitudinal study

c) Case study

d) Meta-analysis

28. A researcher wants to compare the means of three groups. Which statistical test should be used?

a) T-test

b) Chi-square test

c) ANOVA

d) Regression analysis

29. If a dataset has a median of 50 and a mean of 60, the distribution is likely:

a) Symmetrical

b) Positively skewed

c) Negatively skewed

d) Uniform

30. Which measurement scale is used for ranking preferences?

a) Nominal

b) Ordinal

c) Interval

d) Ratio

Integrative Course in Psychology

The End

Prepared by:
John Carl Docog, RPm