

Primary Scales:

Nominal Scales: Numbers serve only as labels (frequency counts i.e. Modes, percentages)

Ordinal Scales: Ranking. No difference between figures. (centile based stats. i.e. percentile, quartile, etc.)

Interval Scale: Equal distances represent equal values. $y = a + bx$. (nom & ord techniques + mean, std dev, etc)

Ratio Scales: Like interval but possesses absolute 0 point. $y = bx$. (all statistical techniques)

Scaling Techniques (Comparative & Non-Comparative)

Comparative - direct comparisons between objects. (Only ordinal or ran order properties)

Paired Comparison - compare one object to another based off criterion (ex. Which one do you prefer?)

Rank Order - rank each object

Constant Sum - (ex. Allocate 100 points to each object)

Non-Comparative - each object is scaled independently (interval or ratio scaled)

Continuous Rating Scale - rate by placing mark at appropriate position on a line that runs from one extreme to another.

Itemized Rating Scale - Respondents are provided with a scale that has a number or brief description with each category.

Semantic Differential - 7 point rating scale with bipolar labels (1-7 or -3 - 3)

Stapel Scale - (-5 to 5) vertical scale. 0 is neutral point.

Likert Scale - indicate level of agreement or disagreement from 5 levels (ex. 1-5 PNC has poor customer service)

$$X_o \text{ (Observed Score)} = X_t \text{ (True Score)} + X_s \text{ (Systematic Error)} + X_e \text{ (Random Error)}$$

Reliability - extent to which scale produces consistent results if repeated. Related to (X_e)

Validity - Extent to which differences in observed scale scores reflect true differences among objects. No systematic or random error.

Questionnaire Design - structured technique for data collection consisting of a series of questions, written or verbal.

Open Ended Q's - Adv: easy & quick. Disadv: Costly, interview bias, various responses

Multiple Choice - Needs to be mutually exclusive & collectively exhaustive (can't find answer) Adv: Easy data coding, less bias
Disadv: Position Bias (tendency to choose option based on position)

Dichotomous Questions - T/F, Y/N, etc.

Pre-testing - to identify and eliminate potential problems

Sampling Error - when sample is not reflective of population

Non-Sampling Error - all other errors besides sampling (measurement, analysis, response bias, etc.)

Sampling Frame - representation of the elements of the target population (ex. Phone directory, listserv)

Probability Sampling

SRS - each element in population has equal chance of selection

Systematic - start at random point and pick every i^{th} element

Stratified - population is split into sub populations. Take samples then compare the populations

Cluster Sampling - samples are selected from geographic locations.

Non-Probability Sampling

Convenience Sampling - taking a sample from those most convenient

Judgmental Sampling - experiences individual selects sample based off judgement. Ex. Test Markets

Quota Sampling - Ex. We know student population is 40% male, 60% female. We choose a sample where 40% are male.

Snowball Sampling - respondents are asked to refer researcher to other potential respondents

Data Analysis Methods

$$\text{Standard Dev} = \text{SQRT}[\text{sum of } (X - \bar{X}) / n - 1]$$

$$Z = (p - \pi) / \sigma \quad \text{where: } \sigma = \text{SQRT}[p(1-p) / n]$$

Type I Error - Rejecting Null when Null is True Type II Error - Failing to reject Null when Null is False

Power - $(1 - \beta)$ Probability of Rejecting the Null when it is False (Rejecting when you should)

Cross Tabulation - describes two or more variables simultaneously. (Chi-Square χ^2)

Null Hypothesis: "No association between variables"

$$df = (r - 1) (c - 1)$$

$$\text{Expected Frequency} = (\# \text{ in row}) (\# \text{ in column}) / (\text{total sample size})$$

$$\chi^2 = \sum (\text{observed} - \text{expected})^2 / \text{Total}$$

t Distribution - assumes normal distribution and mean is unknown.

Null Hypothesis: $\mu \leq \text{something}$

$$t = (\bar{X} - \mu) / s \quad df = n - 1 \quad s_x = S / \text{sqrt} [n]$$

Paired Sample t Test -

$$H_0: \mu \text{ of Differences} = 0 \quad H_a: \mu \text{ of Differences} \neq 0$$

$$df = n - 1 \quad t_{n-1} = (\bar{D} - \mu_{\text{subD}}) / (s_{\text{subD}} / \text{sqrt} [n])$$

One Way ANOVA - used to test means for two or more populations. Must have a dependent variable that is metric (interval or ratio) and independent (nominal or ordinal) called *factors*

$$H_0: \text{All means of each of the populations are equal.} \quad H_a: \text{At least one mean is different} \quad df \text{ within} = (c - 1) \quad \text{and} \quad df \text{ error} = (N - c)$$

$$[SS_y = SS \text{ between} + SS \text{ within} \quad \text{or} \quad \text{Total Variation} = SS_x + SS \text{ error}] \quad [F = (SS \text{ between} / c - 1) / SS \text{ within} \quad \text{or} \quad MS_x / MS \text{ error}]$$

$$\text{Eta-Square} = \eta^2 = SS_x / SS_y$$

Product Moment Correlation - summarizes strength of association between two *metric* variables (interval or ratio). Measures strength and direction.

$$H_0: \text{Two variables are not related} \quad H_a: \text{Two variables are related}$$

Regression Analysis -

$$Y = \beta_0 + \beta_1 X \quad \text{where} \quad Y = \text{dependent variable} \quad X = \text{independent variable} \quad \beta_0 = \text{intercept} \quad \beta_1 = \text{slope}$$

$$\text{For each individual observation:} \quad Y_i = \beta_0 + \beta_1 X_i + e_i \quad \text{where} \quad 'e' = \text{error} \quad i = i^{\text{th}} \text{ observation}$$

$$H_0: \beta_1 = 0 \quad H_a: \beta_1 \neq 0 \quad t = (b / SE_b)$$