GLOSSARY

Prevalence: is the proportion of persons in a population who have a particular disease or attribute at a specified point in time OR over a specified period of time. Usually a %

$$\text{Prevalence} = \frac{\text{Number of cases in a defined population}}{\text{Number of persons in a defined population}}$$

Usually refers to the present moment, or the past.

Lifetime prevalence : the proportion of a population who, at some point in life has **ever** had the characteristic. Lifetime IPV prevalence among adult women = the % that who have experienced IPV at some point in their life .

Period prevalence : the proportion of a population who has the characteristic at **any** point during a given time period **Past year IPV prevalence among women** = % of women who experienced IPV in the 12 months before measurement.

Incidence : occurrence of **new** cases of an event of interest in a specified population (disease, violence) over a specified time period (or the collective time you spend observing them).

Incidence(risk)

$$\text{Incidence} = \frac{\text{Number of new cases in a given time period}}{\text{Number of people at risk at the start of the time period}}$$

In violence research, this is often a prediction about a steady state. "One in four college women experience sexual assault"

Incidence rate: A measure of the frequency with which an event–eg a new case of illness, occurs in a population over a period of time

Hypothesis: a statement of expectation or prediction that will be tested

Null hypothesis: mean (group A) = mean (group B)

Alternative hypothesis: mean (group A) ≠ mean (group B)

p-value (probability value): the probability of getting a result that is either the same or more extreme than the actual observations. The P-value is known as the level of marginal significance within the hypothesis testing that represents the probability of occurrence of the given event. Tells you **how likely it is that your data could have occurred under the null hypothesis.** The P-value is used as an alternative to the rejection point to provide the least significance at which the null hypothesis would be rejected. If the P-value is small, then there is stronger evidence in favour of the alternative hypothesis.

Eg In testing the association between men's IPV perpetration and their experience of childhood trauma, the null hypothesis would be: There is no association between IPV perpetration and childhood trauma experience.

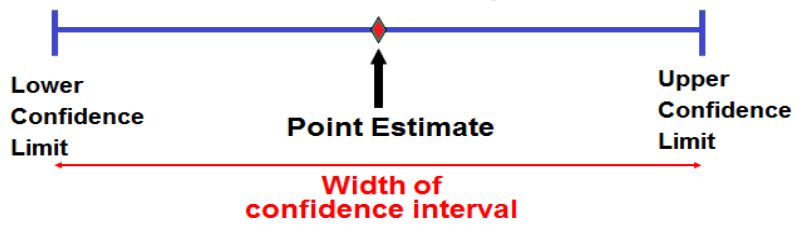
The alternative hypothesis would be: There is association between IPV perpetration and childhood trauma experience.

Small p-value (<0.05) would be in favour of the alternative hypothesis (i.e the data provided is in favour of the alternative).

Point Estimate: A best guess about the true value of something. “30% of women experience IPV” “Women without advanced education are twice as likely to experience IPV”

Confidence Interval: the range of values that **could** be true you repeated to your study in the same way OR how “confident” you are about your point edtimate

Confidence level: Usually 95%. the percentage of times you would expect to get a point estimate inside your confidence interval if you repeat the study over and over .

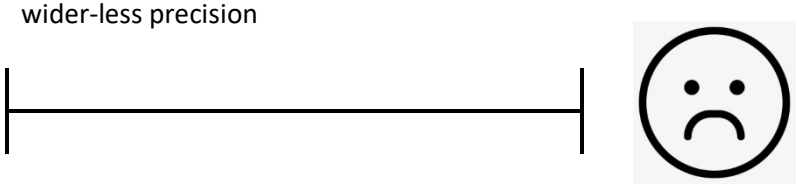


Odds: Probability of event occurring divided by the probability of the event not occurring. A one in 6 chance of getting a 6 on a 6 sided die.

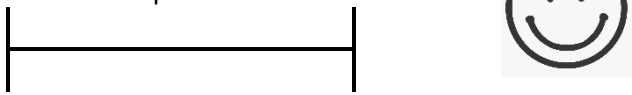
Odds ratio: The probability in Group A vs probability in Group B. Ex Odds ratio of rolling a 6 on one of these dice

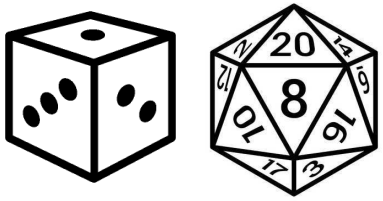
Confidence Interval

wider-less precision



narrow-more precision





1 in 6 chance

1 in 20 chance

Risk Ratio or Relative Risk: measure of the **risk** of a certain event (eg IPV experience) happening in one group compared to the risk of the same event happening in another group.

20% of women under 25 will experience IPV next year
16% of women over 25 will experience IPV next year

RR = 1.25

Interpreting risk and odds ratios:

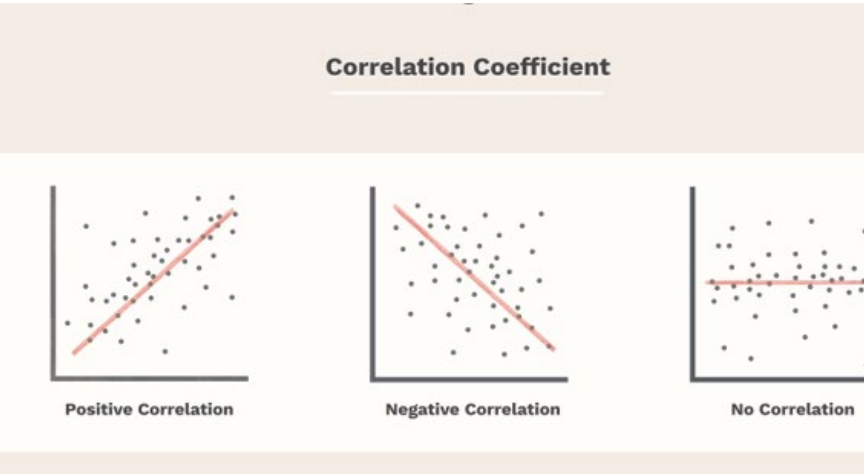
1= NO difference >1 = more risk <1 less risk

Coefficients (beta or ß): in public health use, a measure of how much something goes up or down in relation to something else

Interpreting coefficients:

0= NO difference >0 = more risk <0 less risk

Correlation coefficient: a numeric measure of the relationship between two variables (the degree to which changes to the value of one variable predicts change to the value of another)

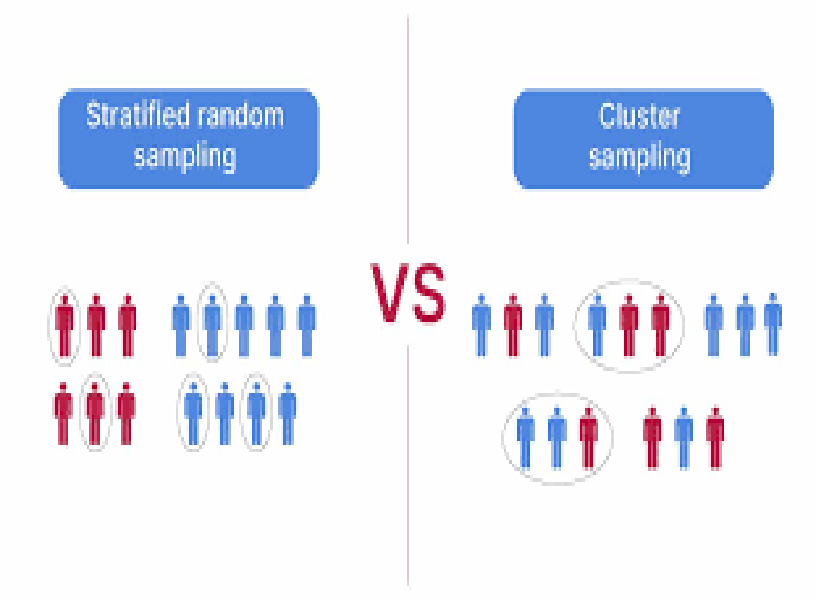


Population: Who you WANT to talk about

Study population: Who you have AVAILABLE to look at

Sample: Who you ACTUALLY get data from

Generalizability: How well the sample matches the population. Who got left out? Why? What does it mean for your conclusion?



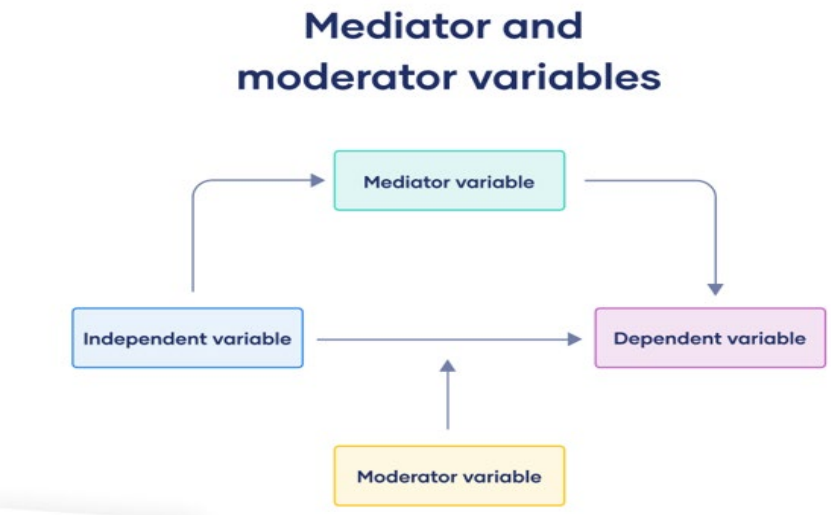
Stratified samples: Divides population into groups, then includes some members from each groups. Example: Take some people from each census area.

Cluster samples: Divides population into groups, then picks random groups and includes everyone. Examples: Choosing random classrooms in a school

Stratified Cluster Sample: You can stratify and then cluster sample! Example: Divide population into villages (stratify). From each village, select random households (cluster) and interview everyone in the household.

Outcome/dependent/response variables: The thing you are trying to understand and talk about

Predictor/independent/explanatory variable(s): The thing(s) you are measuring & using to try to explain your outcome



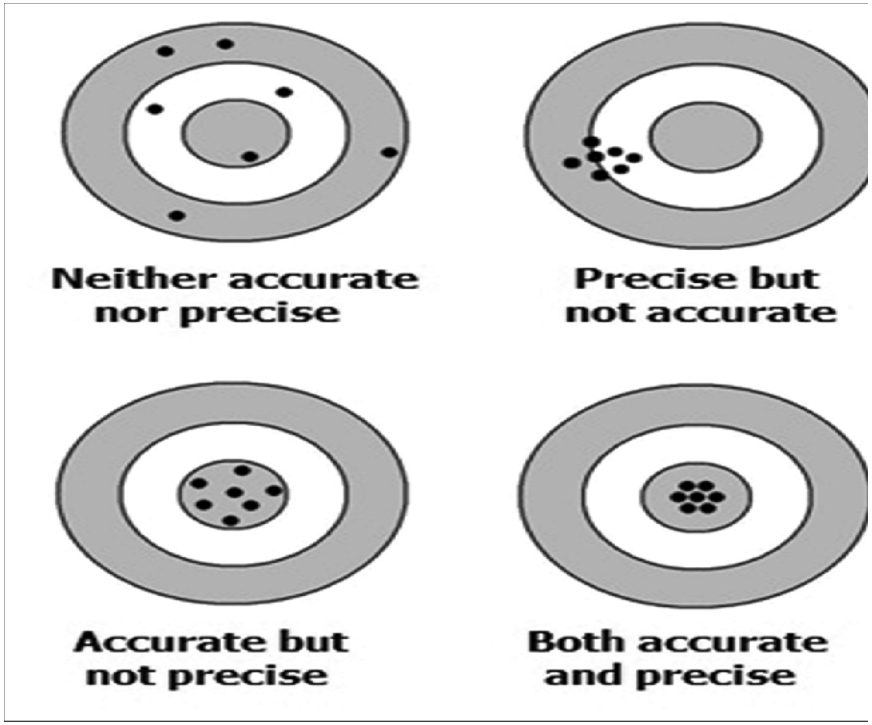
Mediator variable: a variable that explains the process through which two variables are related

Moderator variable: a variable that affects the strength and direction of a relationship

Measure of observational error

Accuracy: refers to how close a measurement is to the true or accepted value.

Precision: refers to how close measurements of the same item are to each other.



Logistic regression: Used for yes/no outcomes like violence or HIV infection. Most common in public health. Findings usually expressed as risk or odds ratios.

Linear regression: Used for continuous outcomes, like scores on a psychometrics test for depression or self-confidence. Findings usually expressed as coefficients.

Hierarchical regression: Used with stratified or cluster samples to express the impact of variables measured at different levels. Example: Characteristics of a classroom are expressed at one level, characteristics of students at another level.

Structural equation models or path models: Draws a picture showing possible lines/paths of causes and effects across different variables. Each path gets a coefficient.