

S12 - 0.0 - Definitions

Population - (Parameters) The complete collection of all individuals. (Data)

Sample - (Statistic) A subcollection of members selected from a population. (Data)

Data Types

Quantitative - Measurable (Height/Weight)

Categorical - Not measurable (Colour)

Discrete (Penny Weight)

Continuous (Human Weight)

Nominal - Names, Labels, Categories. ie. Yes/No

Ordinal - Can be arranged in order. ie. Grades A,B,C,D

Interval - Like Ordinal but difference has meaning. ie. Temp

Ratio - Interval with zero as starting point. ie. Distance

Samples (Voluntary Response?)

Each member of Population :

- Random - is as likely (Simple : Sample)
- Probability - is known (not the same*)

Systematic - Select on every 50th* person

Convenience - Results that are easy to get

Stratified - At least two subgroups. ie. Male/Female

Cluster - Divide into sections

Data Characteristics

Centre - Average

Variation - Measure of vary

Distribution - Shape of spread

Outliers - Lie far from

Time - Changing over time

Biases (Source/Context)

-Selection (Non-representative sample)

-Response (Non-response)

-Voluntary

-Leading Question

-Loss Aversion

-Framing Bias

-Anchoring Bias

...

Correlation (Outcomes related) vs

Causation (One Causes other)

Experimental - Altered

Observational - Unaltered

Statistical/Practical

Significance/Conclusion

Mean:	Average of Data (Information)	
Median:	Middle # or Mean of Middle two #'s ($n = 5, 3rd \#, n = 6, average \ 3rd \ \& \ 4th \ #'s$)	
Mode:	Value that occurs most often. Can have no/multiple mode/s. <u>Bi/Multi-modal</u>	
Range (UL/LL):	Highest Data minus Lowest Data (Quartiles Q1, Q2, Q3)	1,2,2,3,3,4
Frequency (f):	Number of Data Points	1,2,2,3,3,4,4,5
Standard deviation:	Measure of Spread (Pop vs. Sample)	
Variance:	σ^2 is the square of the standard deviation	

Subset: All data within another Set (A group of things that belong together.)

Probability: Measure of the likelihood of a random event taking place.

Outliers : An abnormal observation

μ : mean

n : number of observations

x_i : data

Σ : Sum

σ : standard deviation

σ^2 : Variance

s : sample standard deviation

z : the number of standard deviations that x is away from the mean μ

\bar{p} : sample proportion

X^2 : Chi Squared

\bar{x} : Sample (n)

μ : population (N)

Rule of Thumb -

About 95% of

the area is 2

standard

deviations from

the mean

Estimators

Unbiased

(Target Population

Parameter)

\bar{x}, s^2, \bar{p}

Biased

Median, Range, s

$S_n = n\mu$; The sum of n numbers with a mean of μ

S12 - 0.0 - Definitions

Number of Trials : Number of times experiment is repeated.

Outcomes : Different possible results.

Event : A set of outcomes ie. Coin Toss

Sample Space "S": The set of all possible outcomes in an event. ie. Heads/Tails {H,T}

$P(E)$: Probability of event "E"

Frequency : Number of times a particular outcome observed.

Relative Frequency : Frequency divided by number of trials = Experimental Probability

Dependent : Dependent on previous Event.

Independent : Independent on previous Event.

Mutually Exclusive : Can't happen at the same time (aka Disjointed).

Compliment : Events Not in "A". $P(\bar{E}) = 1 - P(E)$ \bar{E} : Not E

Compound Event : Two or more simple events.

5% Rule*

$p(x) \leq 0.05$ is Rare.

Point Estimate : a single value used to approximate population parameter.

Confidence interval : a range, or interval, of values used to estimate a true value of a population parameter.

Confidence Level : The probability $(1 - \alpha)$ that the confidence interval actually contains the population parameter.

Critical Value : the number on the borderline separating sample statistics that are likely to occur from those that are unlikely to occur.

Degrees of Freedom : For a collection of sample data, is the number of sample values that can vary after certain restrictions have been imposed on all values.

Random Variable : a variable that has a single numerical value.

Probability Distribution : a description that gives the probability for each value of the random variable.

(Graph/Table/Formula) $\sum = 1.0$

Discrete random variable : has either a finite number of values or a countable number of values. (ie. # of Books)

Continuous random variable: has infinite many values, and those values can be associated with measurements on a continuous scale without gaps or interruptions. (ie. Distance)

Binomial Probability Distribution: results from a procedure that has a fixed number of independent trials of success or failure at a constant probability.

Percentiles of $x = \frac{\text{number of values less than } x}{\text{total number of values}}$

$L = \frac{k}{100} n$
L Whole ; $\frac{x_L + x_{L+1}}{2}$
Or Round up

Frequency
Distribution
If x is a range
use midpoint

Interquartile Range (IQR) = $Q_3 - Q_1$

Semi-interquartile Range (IQR) = $\frac{Q_3 - Q_1}{2}$

Midquartile = $\frac{Q_3 + Q_1}{2}$