9 - The role of Sampling in Inferential Statistics

measures computed for

population = parameters

sample = statistics

Statistics (Sample)	Parameters (Population)
$\bar{X} = M = Sample\ Mean$	$\mu = Population\ Mean$
$S^2 = Sample\ Variance$	$\sigma^2 = Population\ Variance$
$S = Sample\ Standard\ Deviation$	$\sigma = Population\ Standard\ Deviation$
$r = Sample\ Correlation$	$p = population\ correlation$

Central Limit Theorem

shape is approximately normally distributed

the distribution mean (mean of means) is equal to population mean

its standard deviation (standard deviation of means) is equal to Population Standard Deviation/Square Root of N, also called Standard error of mean

N>=30 gives a good enough approximation to a normal distribution

Central Limit...

given a population of values with no specific distribution and sample of size N that is sufficiently large, sampling distribution of means (with replacement) can be described as follows:

N>=30 gives a good enough...

when distributions are known to be nearly normal, 8 or 10 is sufficient, when they are perfectly normal, sample of any size yields a normally distributed sampling distribution

Random Samples

Simple Random Sample

Equal chance to be picked

Independent - pick has no effect on the following pick

Sampling with and w/o replacement

with replacement

with replacem...

probability of first and second... selection stays the same (p of selecting one of 52 is 1/52)

w/o replacement

w/o replacem...

first probability e.g. 1/52, second 1/51, third 1/50 etc.

Infinite Sample

Infinite Sample

for most stat. purposes if population is at least 100 times larger than the sample we treat the population as infinite

Estimators and Bias

sample mean is unbiased estimator of the population mean

sample variance S^2 =SUMA(xi-mean)^2/N is not unbiased estimator of the population variance

sample varian...

the mean of all the S^2 is always smaller than the value of population variance (o^2)

$$\hat{\sigma}^2 = \frac{\sum(X_i - \bar{X})^2}{N-1}$$

unbiased is Variance Estimator

unbiased is Variance Estimator

SPSS calls this equation VARIANCE, if we want to describe the spread of scores in sample and not to estimate variance of the population we have to multiply the outcome by (N-1)/N