SAMPLE SIZE DETERMINATION

10/25/2024 By Digafe Tsegaye

SAMPLE SIZE IN QUANTITATIVE STUDIES

- How many completed questionnaires do we need to have a **representative sample**?
- Generally the larger the better, but that takes more time and money.
- **Answer depends on:**
  - How different or dispersed the population is.
  - Desired level of confidence.
  - Desired degree of accuracy.
  - Desired margin of error

CONT'D

In general sample size determination depends on

- **Objectives of the study**
- **Design of the study**
- **Plan for statistical analysis**
- **Degree of precision**
- **Degree of confidence**
- **Accuracy of the measurements**

SAMPLE SIZE …

- **Which variables should be included in sample size calculation?**
  - It should relate to the study's **primary outcome variable.**
  - If the study have secondary outcome variables which are considered important, the sample size should also be sufficient for the analysis of these variables.

SAMPLE SIZE …

- **Sample size may be determined by using:**
  - **Subjective methods** (***less sophisticated methods***)
    - **The rule of thumb approach: e.g. 5% of population**
    - **Conventional approach: e.g. Average of sample sizes of similar other studies;**
    - **Cost basis approach: The number that can be studied with the available funds;**
    - **Depending on the number of independent variables**
  - **Statistical formulae** (**more sophisticated methods**)
    - **Confidence interval approach (CIA).**

THE CONFIDENCE INTERVAL APPROACH:

- **To determine sample sizes using statistical formulae, researchers use the CI approach based on the following factors:**
  - ✓ *Desired level of data precision or accuracy;*
  - ✓ *Amount of variability in the population (**homogeneity**);*
  - ✓ *Level of confidence required in the estimates of population values.*
- Availability of resources such as money, manpower and time may prompt the researcher to modify the computed sample size.

SAMPLE SIZE DETERMINATION

- Given confidence interval

$$mean(proportion) \pm z_{\frac{\alpha}{2}} s.e$$

- Hence the **absolute precision** denoted by *d is given as*

$$d = z_{\frac{\alpha}{2}} s.e$$

- *Where* ***s.e*** *is the* ***standard error*** *of the estimator of the parameter of interest.*

ESTIMATING A SINGLE POPULATION MEAN

$$d = Z_{\frac{\alpha}{2}} SE$$

$$= Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

$$\Rightarrow \sqrt{n} = Z_{\frac{\alpha}{2}} \frac{\sigma}{d}$$

$$\Rightarrow \boxed{n = Z_{\frac{\alpha}{2}}^{2} \frac{\sigma^{2}}{d^{2}}}$$

SAMPLE SIZE

- **POPULATION VARIABILITY:** This refers to the target population's diversity.
  - If the target population exhibits large variability in the behaviors and attitudes of interest being researched, a large sample is needed.
- **LEVEL OF CONFIDENCE:** The level of confidence indicates the probability that the true value of the estimate in fact will fall within the boundaries of the confidence interval.
- **MARGIN OF ERROR:** Also known as **sampling error**, indicates the ***desired level of precision*** of the estimate.
- **POPULATION SIZE:**

SAMPLE SIZE FOR SINGLE POPULATION PROPORTION

- If the study aims to be conducted on single population, then we need the following
  - What is the **probability of the event** occurring?
  - How much **error is tolerable** (confidence interval)?
  - How much **precision** do we need?
  - How **confident** do we need to be that the true population value falls within the CI?

SINGLE POPULATION PROPORTION

- Let p denotes proportion of success, then

$$d = Z_{\frac{\alpha}{2}} SE$$

$$= Z_{\frac{\alpha}{2}} \sqrt{\frac{P(1-P)}{n}}$$

$$\Rightarrow \sqrt{n} = Z_{\frac{\alpha}{2}} \frac{\sqrt{P(1-P)}}{d}$$

$$\Rightarrow \boxed{n = Z_{\frac{\alpha}{2}}^{2} \frac{P(1-P)}{d^{2}}}$$

SAMPLE SIZE FOR SINGLE…

Where:

- **n**-is minimum sample size
- **p**- is estimate of the prevalence rate for the population
- **d**- is the margin of sampling error tolerated

- **Zα/2 -** is the standard normal variable at (**1-α**)100% confidence level and **α** is mostly 5%

POINT TO BE CONSIDERED

- **Use correction formula, if study population(N) is <10,000**

1. If sampling is from a finite population of size N, then

$$n = \frac{n_0}{\left(1 + \frac{n_0}{N}\right)}$$

where n_0 is the sample from an infinite population

2. The initial sample size approached in the study may need to be increased in accordance with the expected response rate, loss to follow up, lack of compliance, and any other predicted reasons for loss of subjects
3. Design effect for complex cluster sampling

Common values multiply n by 2, 3, ...5.

EXAMPLE

- A hospital administrator wishes to know what proportions of discharged patients are unhappy with the care received during hospitalization.

- If 95% Confidence interval is desired to estimate the proportion within 5%, how large a sample should be drawn?

- $\mathbf{n = Z^2\, p(1\text{-}p)/d^2} = (1.96)^2\,(.5\times.5)/(.05)^2 = 384.2$

  $\approx$ **384 patients**

SAMPLE SIZE FOR TWO POPULATION

- Equal Sample size for the Difference of proportions
- Unequal Sample size for the Difference of proportions

EQUAL SAMPLE SIZE FOR THE DIFFERENCE OF PROPORTIONS

- Here the objective of the study is to check whether there is significant difference between two proportions coming from two different population.
- The sample size to be taken from each group is assumed to be equal.

- Let us denote:
  - $\hat{p}_1$ =current estimate of population proportion P_1 (Non-exposed or control group)
  - $\hat{q}_1 = 1 - \hat{p}_1$
  - $\hat{p}_2$ = current estimate of population proportion P_2 (exposed or treated group)
  - $\hat{q}_2 = 1 - \hat{p}_2$
  - $\bar{P}$ =estimated avarage of P_1 and P_2

- $\bar{q}$ =estimated average of Q_1 and Q_2
- Z_α=the Z value corresponding to the alpha error. This is the value at two tailed.
- Z_β =the Z value corresponding to the beta error. The Z value for beta is always based on one tailed test. So if beta is 0.05, 0.1, 0.2, or 0.3, then the corresponding z values are 1.65, 1.28, 0.85 and 0.52 respectively.

- Then the formula for simple size from each group is defined as:

$$n = \left[\frac{Z_\alpha \sqrt{2\bar{p}\bar{q}} + Z_\beta \sqrt{\hat{p}_1\hat{q}_1 + \hat{p}_2\hat{q}_2}}{\hat{p}_1 - \hat{p}_2} \right]^2$$

COMPARISON OF TWO PROPORTIONS..

- The number of samples required in each group to compare the proportion is given by:

  n (in each region) = $(p_1q_1 + p_2q_2)$ $(f(\alpha, \beta))$ / $((p_1 - p_2)^2$

- ✓ α = type I error (level of significance)
- ✓ β = type II error (1-β = power of the study)

- ✓ power = the probability of getting a significant result
- ✓ f (α, β) =10.5, when the power = 90% and the level of significance = 5%

SAMPLE SIZE USING STATISTICAL SOFTWARE

- As an alternative method, we can use EPI INFO statistical software to calculate the sample size required for the study.
- Let us assume the population that we want to conduct the study has target population of size N=100,000.
- **The proportion of the variable of interest is not known which means there is no previous study done and hence we decided to use 50% as an estimate of the prevalence for that variable.**
- Then the steps that we need to follow to get the required sample size using EPI INFO statistical software are given below:

STEPS TO COMPUTE SAMPLE SIZE

- First make sure you install the software EPI INFO.
- If your computer has the software, then go to the start menu and open it.
- This is the window that you are going to get when you open the software:

Start page

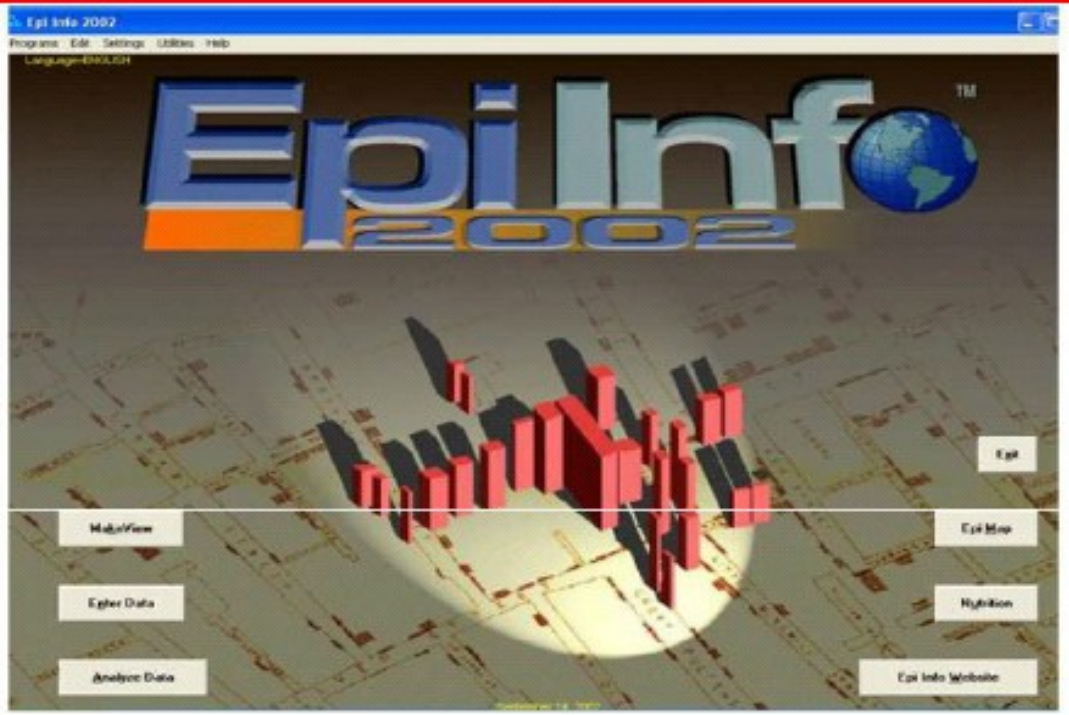


Figure 3.8: EPI INFO starting page

- Then go to *Utilities* and select **StatCalc** as shown below:
- When you click **StatCalc**, then the following option will display for you. Select *Sample size and power* using the direction arrow of your computer as:
- At this step you are going to get options of the study design you planned to use for the study: select "population survey" as:

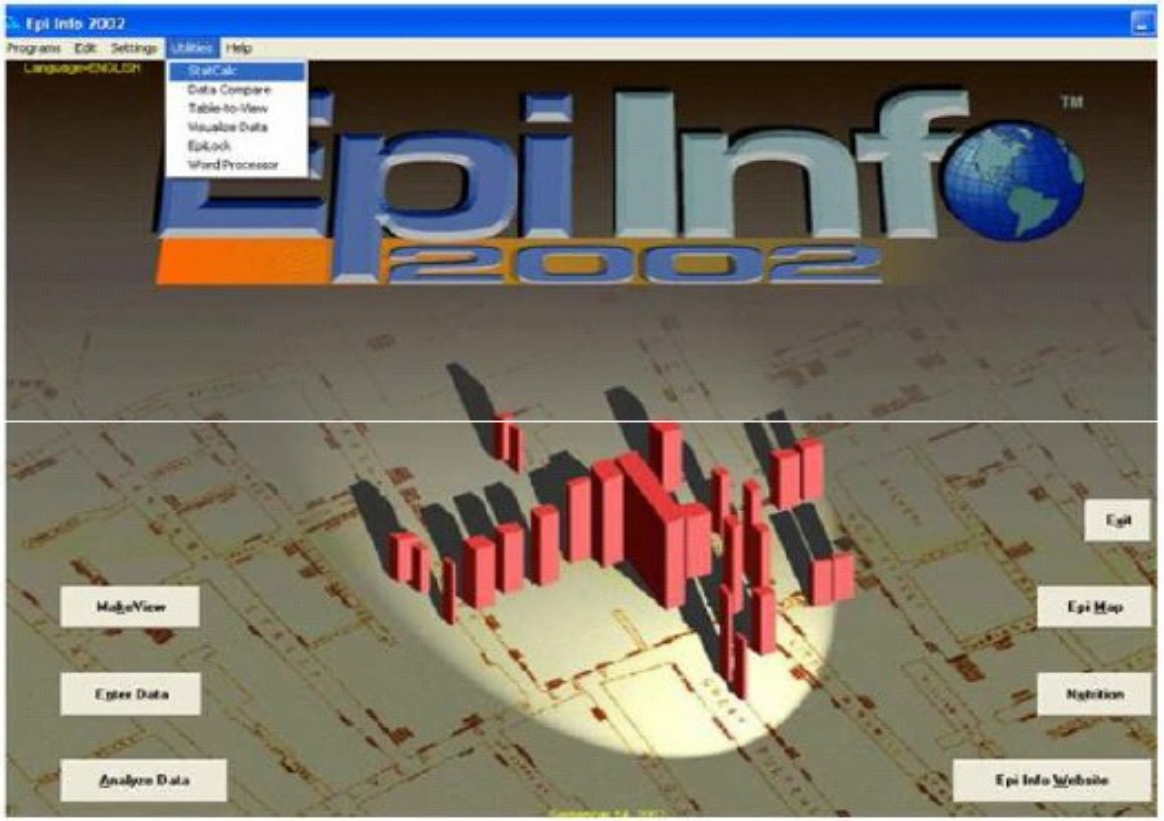


Figure 3.9: The StatCalc option of EPI INFO

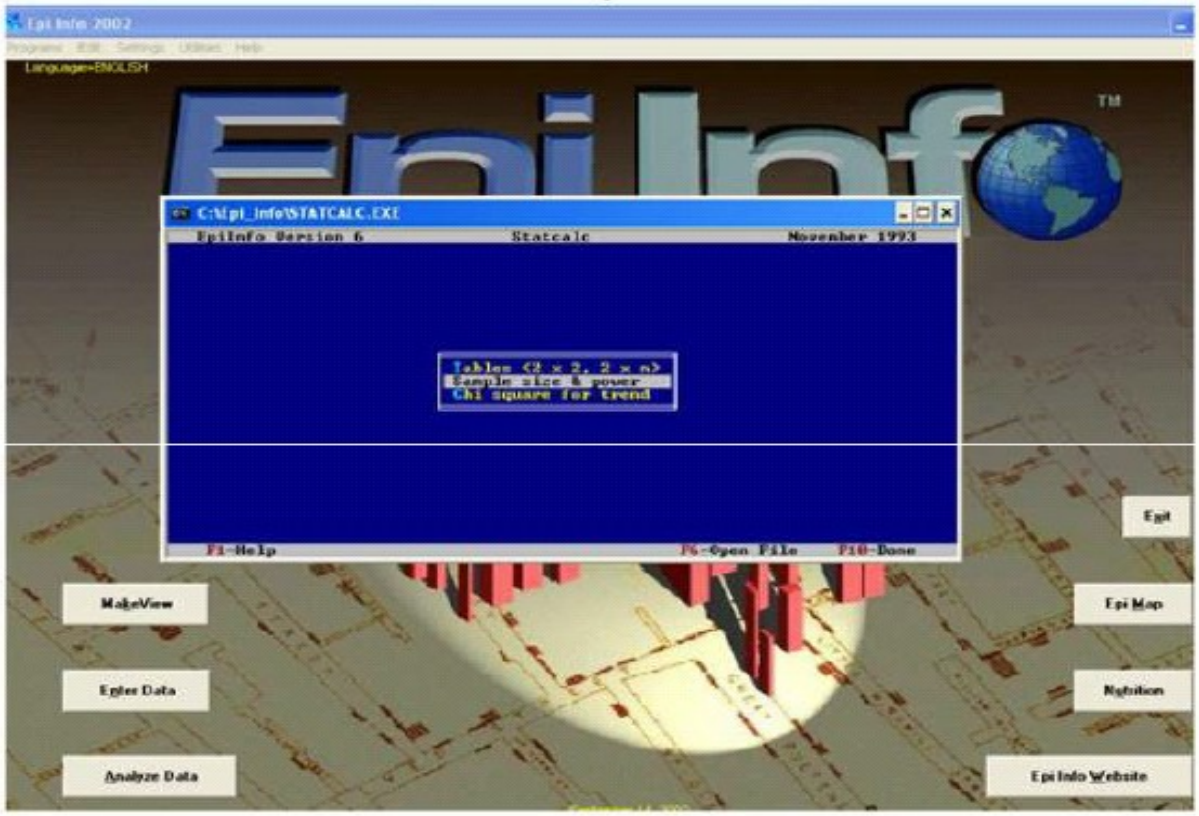


Figure 3.10: Options given when you select statcalc

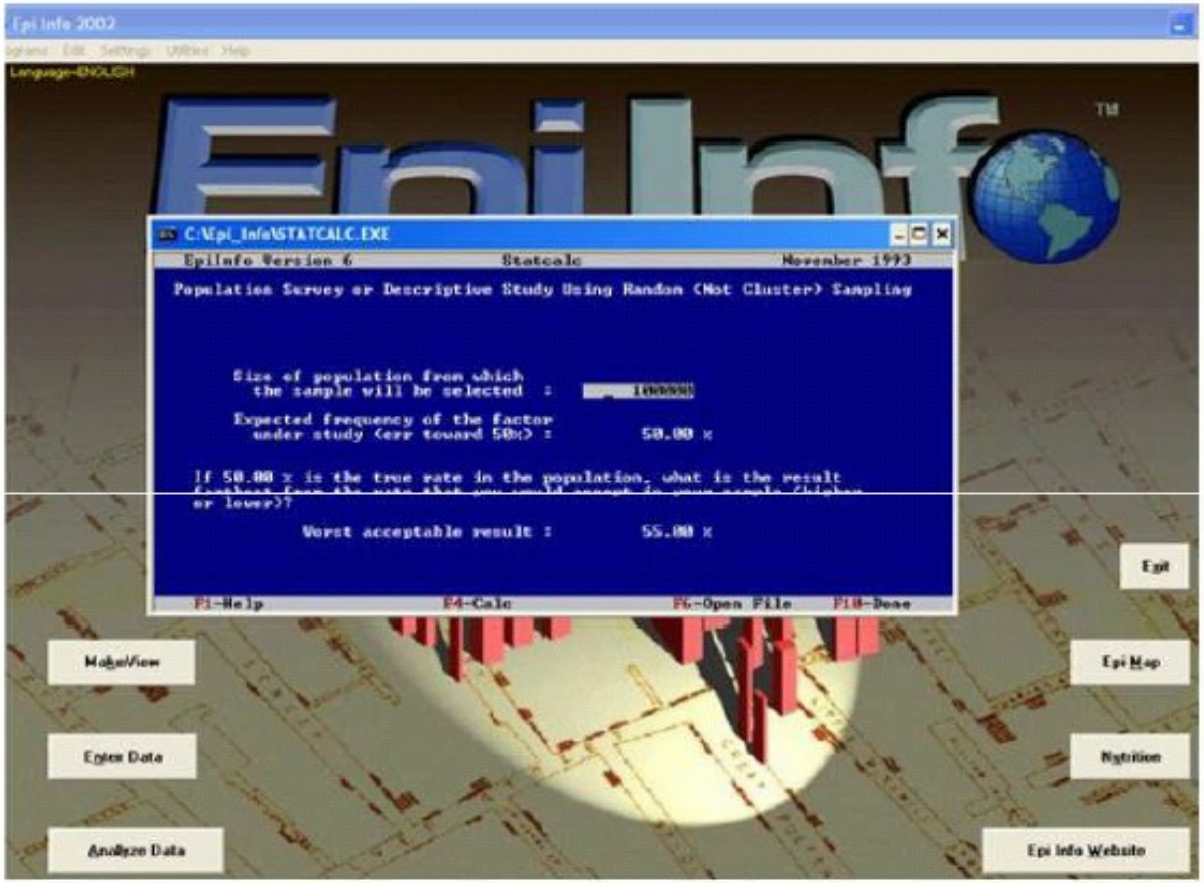


Figure 3.12: The required information to compute the sample

SAMPLE SIZE FOR NON-PROBABILITY SAMPLING

- When using non-probability sampling, sample size is **unrelated to accuracy**, so **cost benefit** consideration must be used.