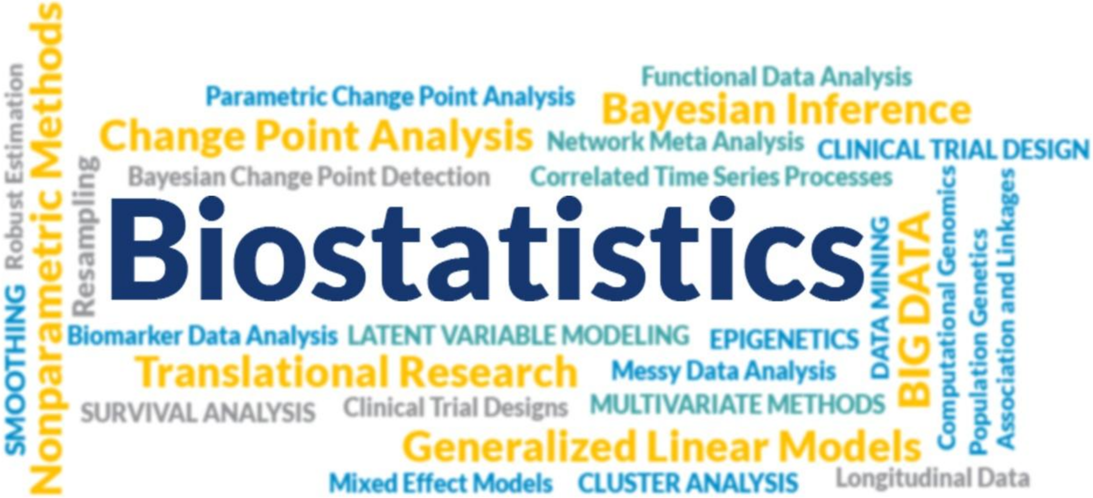


◉ Sheet ○ Slides

Number

10

Done by:

Omayma hassanin

corrected by:

Jamil Nazzal

Doctor

Mahmoud Al hussami

This lecture will demonstrate sample size calculation and determination, and it is an important lecture for the test. Please refer to slides 39-56 from unit 5.

Sample Size Determination:

Sample size calculation/ determination is to determine how many observations or replicates the researcher would include in the sample.

Sample size determination depends on several factors:

1) Extraneous variables: If extraneous variables are present in a population, the researcher should have a large sample size.
2) Homogeneity of the groups in a study: if the groups are homogenous, then the differences between the groups will be very small, so the study should have a large sample size.
   *In inferential statistics, the study should have groups regardless of whether the study depends on nominal, continuous, or ordinal data. Also, if the difference between the groups is very minimal, a small sized sample won't be useful for a study→ a larger sample size is needed when very minimal differences are present, in order to detect differences between the groups (to have more clear differences).
   <u>Example:</u>
   If we want to compare between the four sections, they are considered homogenous in many aspects (high school grades, age, mental abilities, subjects and courses they take, etc.). In this case, we have to have a large sample size to detect the differences.
3) If the phenomena in the population is very rare, a large sample size is needed, we need to make sure that the subjects were exposed to the phenomena and by increasing the size of the sample we increase the chances of having subjects that were exposed to the phenomena.
4) Dropout rate among subjects is expected to be high so we will need a large sample size
5) The Statistical tests that are used require minimum sample sizes.
6) The population must be divided into subgroups, a large sample size is needed

The process by which the sample size is calculated is defined by four methods:

1) Estimation of means
2) Estimation of proportions
3) Power tables (Cohen's table) for effect size
4) Power of a statistical test (G* power)

Note: The above mentioned four methods result in numbers and are used in quantitative studies. The researcher should double the values which result from the four methods in case the phenomena/ case studied is rare. However, if the phenomena/ case normally exists, the values which result from the four methods would stay as it is with no extra changes.

Methods through which a phenomena/ case is studied are two main methods:

1) Qualitative methods: 5 designs→
   a- Phenomenology: an approach to study experiences of patients in a particular issue or the accounts of health professionals about their patients.
      Sample size is between 20 (minimum) and 30 (maximum).
   b- Biography: the experience of one person through a certain case or issue, eg. C.V.
      Sample size is one person.
   c- Case study: a clinical/ medical case.
      Sample size: one person.
   d- Ethnography: the study of cultures.
      Sample size: 15-20
   e- Grounded theory: the study will have participants until the data gets saturated. In other words, when the answers start to get too repetitive, the study should be stopped.

Important note: the above explanation regarding qualitative methods is not that important, the doctor said, but he might ask some questions about it.

2) Quantitative methods:
   a) Observational: in observational studies, the researchers are not manipulating any independent variables. Therefore, no testing of hypothesis occurs in

observational studies, and G* power and Cohen’s table (power table) cannot be used, because both (Cohen’s table and G* power) depend on testing hypothesis.
The researcher can only use mean and proportion in observational studies.

b) Experimental

Note: The best design is the one which answers the research question, so when you choose a design, the best design could be either observational or experimental depending on the type of study your conducting However the stronger the design, the more reliable/ better it is and Experimental designs are stronger than observational designs. usually, we study the phenomena through observational design "descriptive" or through explorational approaches.

Difference between observational (more specifically, descriptive) and explorational studies:

If the phenomena have recently appeared/ is new and no literature explaining it is present, then, the researcher should explore the phenomena (use explorational approaches). Then, after thoroughly studying the phenomena, descriptive studies/ observational are used. →Descriptive studies are the weakest types of studies.

Then, if there are key factors which emerged in the study, correlational studies are used. If the researcher finds out strong correlations between variables and factors which emerged in the study, the following happens:

Steps to study correlations in a phenomenon (phenomena are studied depending on associations)→

There are four approaches (tests) to study correlations: (Very important for the test)

1) If data was continuous→ Independent variable: Continuous

Dependent variable: Continuous

Statistical test used: Pearson Chi Square test or **Pearson’s Product Moment Correlations** (r)

In other words, the variables are intervals or ratios.

2) If data was ordinal→ Independent variable: ordinal
   Dependent variable: ordinal
   Test used: **Spearman test**
3) Dependent: Continuous
   Independent: Ordinal
   Test used: **Point biserial correlation coefficient**

4) Dependent: Continuous
   Independent: Nominal
   Test used: Two types of t-test→ a) If the test is for the same group, **dependent t-test** is used.

   b) If the test is for different groups,

independent t-test is used.

To summarize:

The order by which any phenomena is studied is as follows:

Exploration of the phenomena→ observing the phenomena→ finding correlations between variables in a phenomenon→ then, applying clinical trials if needed.

00:00-21:00

Value of 'r' in Pearson test: (Important for the test)

Value of 'r' is from 0 to 1.

- r=0→ **no correlation**

Eg.: no correlation between height and intelligence→ value of 'r' is 0.

- r≤0.59→ **Weak correlation**
- r value from 0.6 to 0.79→ **Moderate correlation**
- r value from 0.8 to 1→ **strong correlation**

Note: It is very rare 'almost impossible' for a study to have r=1 . Only if we study the same variable, 1 might be a value for the correlation or for r, eg. Comparing two GPAs→ identical variables; from the same type.

Back to Sampling→

To determine a sample is large, small, or moderate in size, we depend on several factors.

The sample is chosen from the reference population and the axis population. The mean of the **sample** (known) chosen from the population is used to determine the mean, μ, of the **population** (unknown).

Estimation for the mean and estimation for the proportions requires a pilot study* to be conducted, and the pilot study which is conducted should not be included in the sample of the original study we are performing. Calculation is as follows:

- $n = (\text{z-score}^2 * \sigma^2) \div \text{Error}^2$

→σ (standard deviation) is taken from the pilot study.

→ 'n' represents sample size.

Example:

Example: Sample Size for Mean

•What sample size is needed to be 90% confident of being correct within ± 5? A pilot study suggested that the standard deviation is 45.

$$n = \frac{Z^2\sigma^2}{\text{Error}^2} = \frac{1.645^2\ 45^2}{5^2} = 219.2 \cong 220$$

Round Up

*Pilot study: a small sample size to detect the standard deviation or the mean from it for the purpose of the original sample (original/ intended study), and also to determine whether the questionnaire is understandable and clear and to modify it if necessary.

A good video explaining pilot studies:

http://study.com/academy/lesson/what-is-a-pilot-study-definition-example.html

Example 2:

Example: Sample Size for Proportion

•What sample size is needed to be within ± 5 with 90% confidence? Out of a population of 1,000, we randomly selected 100 of which 30 were defective.

$$n = \frac{Z^2 p(1-p)}{error^2} = \frac{1.645^2(.30)(.70)}{.05^2} = 227.3$$

$\cong$ **228**

Round Up

Note: The value of the error 5 is converted to 0.05 because the values in the nominator are fractions of one hundred (0.3 and 0.7).

Power Tables for Effect Size (Cohen's table):

Power of the study is defined as the outcome being a result of the independent variable manipulation.

Eg. If the outcome is 90% related to the independent variable, then 90% is the power of the study.

Manipulation of the independent variable usually exists in a study, and the outcome is the dependent variable. There are also extraneous variables which affect the outcome, and there are confounding variables. If the researcher wants to find out the strength of the study and its outcome (in order to determine the sample size), the outcome and strength would be 90 percent affected by the independent variable.

The larger the power, the larger the sample size is. As a result, type 2 error* would be reduced. In other words, 1-β= power; β is type 2 error.

What are the factors which affect Cohen's power table? (Important question for the test)

1- It depends on alpha (type 1 error)
2- It depends on the power of the study
3- Power Tables for Effect Size d: Cohen's d is defined as the **difference between two means divided by a standard deviation** for the data. (Differences between the means; standard deviations of the means; standard error of the mean→ central limit theory*)
4- Power Tables for Effect Size r: is a statistical concept that measures the strength of the relationship between two variables on a numeric scale.

*Type 2 error: a statistical term used within the context of hypothesis testing that describes the error that occurs when one accepts a null hypothesis that is actually false. The error rejects the alternative hypothesis, even though it does not occur due to chance. A type II error fails to reject, or accepts, the null hypothesis, although the alternative hypothesis is the true state of nature.

Good website for understanding the concept: https://onlinecourses.science.psu.edu/stat500/node/40

Central limit theory: a statistical theory that states that given a sufficiently large sample size from a population with a finite level of variance, the mean of all samples from the same population will be approximately equal to the mean of the population.

http://www.investopedia.com/terms/c/central_limit_theorem.asp

Power tables example:

tails	α	z_{crit}
2	0.05	1.96

$$n = \frac{2\left(\left|z_{1-\alpha/2}\right| + \left|z_{\beta}\right|\right)^2}{d^2}$$

	d										
Power	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	1	1.2	1.4
0.25	*	*	*	*	*	*	*	*	*	*	*
0.5	769	193	86	49	31	22	16	13	8	6	4
0.6	980	245	109	62	40	28	20	16	10	7	5
2/3	1144	286	128	72	46	32	24	18	12	8	6
0.7	1235	309	138	78	50	35	26	20	13	9	7
0.75	1389	348	155	87	56	39	29	22	14	10	8
0.8	1570	393	175	99	63	44	33	25	16	11	9
0.85	1796	449	200	113	72	50	37	29	18	13	10
0.9	2102	526	234	132	85	59	43	33	22	15	11
0.95	2599	650	289	163	104	73	54	41	26	19	14
0.99	3675	919	409	230	147	103	75	58	37	26	19

* We are considering only values of beta greater than 0.5
A power value of 0.25 corresponds to a beta equal to 0.75

Notes on the power table (the second one specifically):

- 'd' represents differences→ the lesser the differences (homogenous), the larger the sample size. This can be seen (relatively) on the second table. For instance, when d is 0.1, the corresponding values are large in comparison to the corresponding values when d is 1.4 for example.
  In other words, the larger the power and the smaller the differences (d), the larger the sample size.

- Referring to the second table, if the power is 0.90 and the difference is 0.1, the sample size is 2102. If we increased the power to 0.99, the sample size becomes 3675.
- Another example: if the power is 0.9 and the population is heterogenous and **not** homogenous (more differences; higher 'd'), the sample size would be small as it will correspond to a larger value of d. (The doctor was interrupted while explaining this point and forgot to continue explaining it, so the sentence was completed judging by all the previous explanation.)

To sum up:

To determine the sample size using the power table (Cohen's table), we look at the values for 'd' and the values for the power and find the corresponding value of the sample size from the table.

- 'd' is small
  Power is large
  ➔ Large sample size
- 'd' is large
  Power is small
  ➔ Small sample size

Even though power has an effect on sample size, the main determinant apparently (and (judging from the table) is the value of 'd'.

The value of 'd' represents whether the population is homogenous or heterogenous.

Homogenous population→ small value of 'd'→ large sample size

Heterogenous population→ large value of 'd'→ small sample size

The fourth way to calculate sample size and error is using inferential statistics.

If a design does not use inferential statistics, a hypothesis cannot be tested, and the method to calculate the sample size (n) is invalid.

In other words, if the researcher is not using inferential statistics (whether using parametric or non-parametric techniques), then, the calculation of the sample size won't be valid.

Note: t-test, ANOVA, and correlations are the only tests used to apply Cohen's table.

If you are going to predict a phenomenon, Cohen's table cannot be used to predict a phenomenon, because Cohen's table depends on 'd' and 'r' which are not deduced or calculated in a new phenomenon. Some researchers use correlation to predict a phenomenon. You cannot do a prediction unless a correlation has been made. Correlation is a prerequisite for predictions.

There are factors which influence a prediction, and such factors are called predictands. The researcher should determine the best factor which would result in variance (variability or different factors which would affect the prediction).

If the data was continuous (both dependent and independent are continuous)→ multi-variate statistics, because you are studying more than one factor.

So, if all data is continuous, you use multiple regression which is two parts:

1) Step-wise: if no final studies about the case have been found. We differentiate the factors to two types: a) significant b) insignificant
2) Hierarchal correlations: used if there are studies about the case/ phenomenon. Each factor determined from hierarchal correlations determines the model entered and the value of variance.

If data is nominal→dichotomous, then, logistic regressions.

No type of regression is used for ordinal data to detect the factors

G*POWER:

G*POWER is calculated from the sample, regardless of the type of test/ analysis used.

There should be a research hypothesis for the analysis/ test.

You are going to test either parametric or non-parametric techniques, regardless of the level of the data.

The research has to be familiar with the following:

1) The type of design you are going to use to analyze the data.
2) The power of the study not the power of analyzing the sample size. There is difference between the power of the study and the power of calculating the sample size.
   *The power of calculating the sample size is basically G*POWER.
   *Power of the study which was mentioned before ranges from '1%-99%'
3) Alpha→ when determining alpha, the hypothesis should be taken into consideration (either directional hypothesis [one curve] or non-directional hypothesis [two curves]).

Alpha is type one error

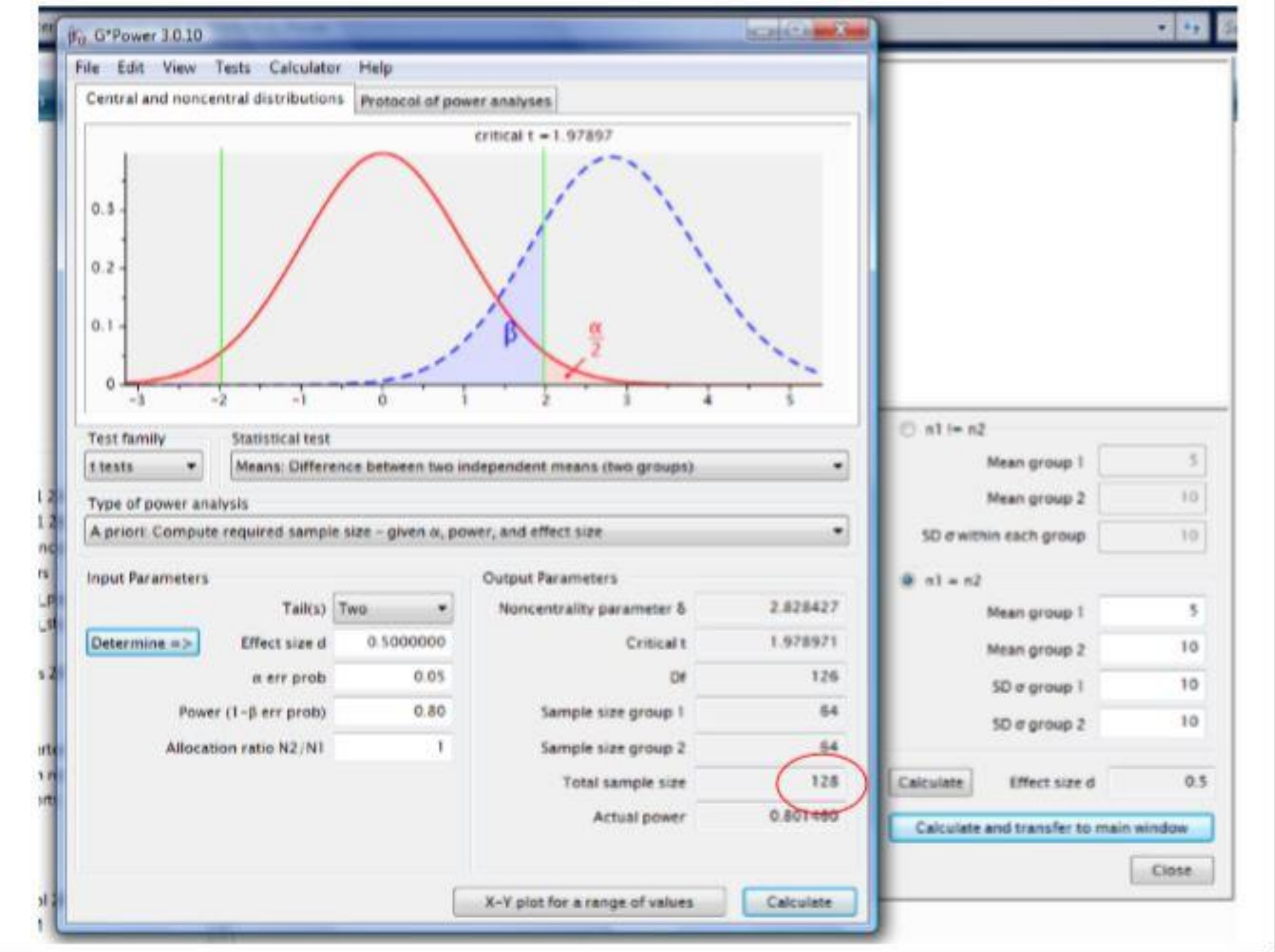
4) Effect size : has three types→
   a) Small
   b) Medium
   c) Large

Effect size is determined from literature. If a literature demonstrates a large study, then the large effect size is used, and the sample size resulting would be small.

A medium effect size results when a study has moderate presence in literature; moderate but not large or strong presence. In this case, the sample size would be medium/ moderate.

If no studies about a case are present in literature, then the effect size would be small and the sample size would be large.

[The above explanation explains the elements present in the interface of the G*POWER program which is illustrated below]:



Explanation about G*POWER from slides:

G*POWER is a FREE program that can make the calculations a lot easier
http://www.psycho.uni-duesseldorf.de/abteilungen/aap/gpower3/

G*Power computes:

• power values for given sample sizes, effect sizes, and alpha levels

• sample sizes for given effect sizes, alpha levels, and power values

• suitable for most fundamental statistical methods

Power is:

– the probability of correctly rejecting a false null hypothesis

– the probability that the study will yield significant results if the research hypothesis is true

– the probability of correctly identifying a true alternative hypothesis

Power of a Statistical Test:

The result of using computer program G* power (Faul et al., 2007) showed that the required sample size was 128 participants. This figure was arrived at by using compromised β=0.80, α = 0.05 (2-tailed) and medium effect size.

To conclude:

Basically, the main purpose of this lecture was to know how the sample size is calculated and determined, and the techniques and programs used to do so.

Best Wishes