Sample Size Estimation using Yamane and Cochran and Krejcie and Morgan and Green Formulas and Cohen Statistical Power Analysis by G*Power and Comparisons

Chanuan Uakarn[1]
Kajohnsak Chaokromthong[2]
Nittaya Sintao[3]
Kasem Bundit University[1]
Rajamangala University of Technology Suvarnabhumi[2]
Stamford International University[3]
E-mail: chanuan.uak@kbu.ac.th[1]
E-mail: kajohnsak7@gmail.com[2]
E-mail: nitsin07@gmail.com[3]

Received: December 1, 2021; Revised: December 14, 2021; Accepted: December 20, 2021

ABSTRACT

Sample Size Determination of Quantitative Research; population size as research fields for the researchers to collect data for analyzing Descriptive Statistics and Inferential Statistics. Therefore sample size refers to a representative of the population. Research Purposes are: to study sample size determination of academic officer for references to determine the sample size of the research appropriate for population size, and to calculate sample size by G*Power. Derivation of the sample size of this research is Documentary Research by calculating from various formulas of the academic officer such as Taro Yamane, Krejcie & Morgan, Cochran, Green, S. B., and Cohen J. which including calculation formulas, table, comparison and computer program of G*Power - depending on variable types, variable measurement scales, Type **I** error (Alpha, α=0.05), Type **II** error (Beta, β), Power of Test: $1-\beta$ =0.80 and Effect Size=0.30.

KEYWORDS: Sample size**,** Population, Significance, Size of a test, Effect size, Power of test.

Introduction

In Quantitative Research; it is tough for researchers to access a large population size; therefore, the researchers need to reduce the population size into correct, adequate and appropriate sample size for collecting data from research fields by processing reference number analyzed from sample called Statistics back to Parameters called population. Population has classified into 3 types as follows 1) Classified by group into 2 groups such as 1.1 General population 1.2 Hypothetical population 2) Classified by scope of population into 2 types such as 2.1 Finite population is every unit of study can specify scope or completely count all the numbers 2.2 Infinite Population is every unit of study but this cannot specify scope or completely count all the numbers and 3) Classified by population characteristics into 2 types (Sedlack & Stanley, 1992) such as 3.1 Homogeneity is population in every unit has similar structural features 3.2 Heterogeneity is population in each unit has different structural features.

Sample is a representative as part of the population that the researchers are interested in. A good sample is the sample with complete important characteristics that are similar to the population, and is also a good representative of the population.

Good sample characteristics

A Good and reliable sample as a representative of the population in every research shall have characteristics as follows

1. Sample must have the same characteristics as population. The more of similarity, numbers of research will not be required. On the other hand, the fewer of similarity, many more numbers of research will not be required.

2. Sample should derive by probability sampling in accordance with preliminary agreement or conditions in statistic that is used for data analysis, especially Inferential Statistics that is used for referencing research results to population; there will be most preliminary agreement or conditions that sample should derive by probability sampling.

Theory

Sample Size Determination; the researchers should consider various factors as follows;

1) Expense, time, labor and data collection tools.

2) Population size.

3) Similarity; the more of similar population, the usage of samples size will be small. If the population is very different, there will be plenty of variances, therefore, the usage of sample size will be large.

4) Accuracy

5) Sampling error

6) Reliability

Sample size determination:

1. Sample size determination by using criteria

- Hundreds of populations; use 15-30% of sample size
- Thousands of populations; use 10-15% of sample size
- Tens Thousands of populations; use 5-10% of sample size
- Hundreds Thousands of population; use 1-5% of sample size

2. Sample size determination by calculation formulas.

3. Sample size determination by using tables.

4. Sample size determination by calculating computer programs.

Purposes

1. To study sample size determination of academic officer for references.

2. To determine sample size of the research appropriate for population size.

3. To calculate sample size by G*Power.

Benefit of Research

1. Researchers can determine the correct sample size.

2. Guidelines of sample size selection or determination of different methods.

3. Facilities for researchers by using computer tools.

Research Process

A researchers conduct a research by Documentary Research.

1. Population and sample determination; researchers conduct a research by documentary research in order to search for the truth of the phenomenon of sample size determination that appropriates for the real representatives of population size in Quantitative Research (Chanuan, 2020).

2. Creation of data collection tools which types of source document; researchers collect data from both national and international researches as published international journal.

2.1 Primary Document: Data that is the most closely matches the study.

2.2 Secondary Document: Data that has some inaccuracies from primary document.

3. Data Collection: All research data, researchers set criteria of data collection including 1) Authenticity 2) Credibility 3) Representativeness and 4) Meaning.

4. Data Analysis: Researchers analyze data from primary document and secondary document by both Content Analysis and Statistical Analysis depending on variable types, variable measurement scales, Type I error (Alpha, α), Type II error (Beta, β), Power of Test: 1- β and Effect Size by G*Power Ver.3.1.9.7.

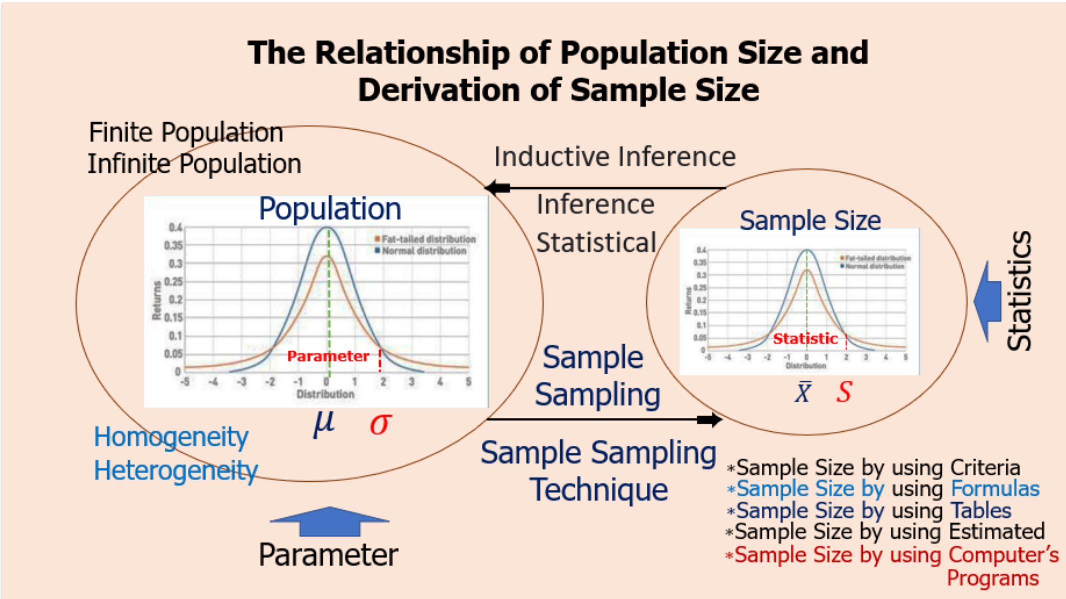

Figure 1 The relationship of population size and derivation of sample size (Chanuan, 2020).

Research Methods and Research Results

Sample Size Determination by calculation formula of Yamane compares with Cochran Table. All 5 cases Cohen compares with sample size by calculating computer programs and Green.

1.Taro Yamane Formula (Yamane, 1973)

$$n = \frac{\text{N}}{1 + \text{N}e^2}$$

$$n = \frac{37{,}581}{1 + 37{,}581\text{x}0.05^2} = 395.78$$

$$\approx 396$$

Where n = sample size

N = population size = 37,581

e = error (0.05) reliability level 95%

or; *e* = level of precision always set the value of 0.05

Yamane (1973) adjusted calculation formula to be more accurate; by increasing of π = population variance from Dichotomous Variable equal to 0.50 and z = z score at significance level α (where z = 2 at α = 0.05 and z = 3 at α = 0.01) as the following formula

$$n = \frac{(z)^2(\pi)(1-\pi)(N)}{(z)^2(\pi)(1-\pi) + (N)(\text{e})^2}$$

$$n = \frac{(2)^2(0.50)(1-0.50)(37{,}581)}{(2)^2(0.50)(1-0.50) + (37{,}581)(0.05)^2}$$

$$= 395.79 \approx 396$$

From calculation formula of given 37,581 population; sample size equals to 396 compares with Taro Yamane Table at reliability level 95% ($e = \pm 5\%$) and various errors; sample size equal to 394–397 which have a similar value.

2. Krejcie & Morgan Formula (Krejcie & Morgan, 1970) If the population size is known

$$n = \frac{\chi^2 Np(1-p)}{e^2(N-1) + \chi^2 p(1-p)}$$

$$n = \frac{3.841 \times 37{,}581 \times 0.5(1-0.5)}{0.05^2(37{,}581-1) + 3.841 \times 0.5(1-0.5)}$$

$$n = \frac{36{,}087.16}{93.95 + 0.96} = 380.225 \approx 380$$

n = sample size

N = population size $= 37{,}581$

e = acceptable error of sample size

χ^2= Chi-square $d_f = 1$ and reliability level 95% ($X^2 = 3.841$)

p = the population proportions (Assumed to be 0.5)

Compare calculation formula with Krejcie & Morgan Table; sample size is 380 which equals to the population proportion and assumed to be 0.5; acceptable error 5% and reliability level 95% but the advantage by using table can calculate small and 10 or more sample size and population.

3. Cochran Formula (Cochran, 1977)

3.1 If the population size is unknown but a lot, the population proportion is known

$$n = \frac{p(1-p)z^2}{e^2}$$

$$n = \frac{(0.1)(1-0.1)(2.58)^2}{(0.05)^2} = 240$$

n = sample size

p = the population proportion ($p = 0.1$)

e = acceptable sampling error ($e = 0.05$)

z = z value at reliability level or significance level.

- Reliability level 95% or significance level 0.05; z = 1.96
- Reliability level 99% or significance level 0.01; z = 2.58

3.2 If the population size is unknown, the population proportion is unknown.

$$n = \frac{z^2}{4e^2}$$

$$n = \frac{(1.96)^2}{4(0.05)^2} = 384.16$$

n = sample size

p = the population proportions

e = acceptable sampling error ($e = 0.05$)

z = z value at reliability level or significance level.

- Reliability level 95% or significance level 0.05; z = 1.96
- Reliability level 99% or significance level 0.01; z = 2.58

3.3 If the population size is unknown and estimated population mean

$$n = \frac{\sigma^2 z^2}{e^2}$$

$$n = \frac{15^2 x\ 1.96^2}{5^2} = 35\ unit$$

n = sample size (unit)

σ= standard deviation of the sample ($\sigma = 15$)

e = acceptable sampling error $\approx (\pm 5\%)$

[If σ is unknown, defined e as % of σ such as 8% of σ ($e = 0.08\sigma$) or 10% of σ ($e = 0.10\sigma$)]

z = z value at reliability level or significance level.

- Reliability level 95% or significance level 0.05; z = 1.96
- Reliability level 99% or significance level 0.01; z = 2.58

3.4 If the population size is known and estimated the population proportion.

$$n = \frac{p(1-p)}{\frac{e^2}{z^2} + \frac{p(1-p)}{N}}$$

$$n = \frac{0.5(1-0.5)}{\frac{0.05^2}{1.96^2} + \frac{0.5(1-0.5)}{2,000}} = 322$$

n = sample size
N = population size
e = acceptable sampling error
p = the population proportions
z = z value at reliability level or significance level.

- Reliability level 95% or significance level 0.05; z = 1.96
- Reliability level 99% or significance level 0.01; z = 2.58

3.5 If the population size is known and estimated population mean

$$n = \frac{NZ^2\sigma^2}{(N-1)e^2 + Z\sigma^2}$$

$$n = \frac{400x(1.96)^2x\,(15)^2}{(400-1)(5)^2 + 1.96x(15)^2} = 33\ unit$$

n = sample size (unit)
N = population size (unit)
σ= standard deviation of the sample
e = acceptable sampling error

[If σ is unknown, defined e as % of σ such as 8% of σ $(e = 0.08\sigma)$ or 10% of σ $(e = 0.10\sigma)$]

z = z value at reliability level or significance level.

- Reliability level 95% or significance level 0.05; z = 1.96
- Reliability level 99% or significance level 0.01; z = 2.58

4, Samuel B. Green Formula (Green, 1991)

Regression Analysis is calculation of sample size from Green Formula which is the representative of population; it suitable for research by Sample Survey in data collection in order to test hypothesis by Regression Analysis. The dominant characteristic of Green Formula which does not take into account in population size; number of independent variables are only known. Green had developed equation (Harris, 1975) as n ≥ 50 + m and (Howell, 2002) noted that Green Formula is appropriate for Regression Analysis more than other formulas but not suitable for few predictors. For 5 predictors, the power of a hypothesis test is 0.80

$$n \geq 50 + 8(m)$$

where n is sample size
m is predictor or independent variables If m = 13, sample size estimated

$$n \geq 50 + 8(13) = 154$$

5. Jacob Cohen Formula

5.1 Calculation of sample size by group as follow:

$$n = \frac{N_{0.05}}{400f^2} + 1$$

$$n = \frac{1,096}{400(0.70)^2} + 1 = 7$$

Where $N_{0.05}$ = derived from table (Cohen, 1977: p.384); effect size from $f = \frac{\sigma_m}{\sigma} = 0.70$ One-way Analysis of Variance test) and Power of test as defined at significance level (α) = 0.05 or 0.01 such as significance level 0.05 and Power of test 80% and u = 3 (Group 1) = 1096 unit; therefore, calculation of sample size is 7 unit per group.

5.2 Calculation of sample size for Parametric Statistics by Regression Analysis (Chua, 2006) as follow;

Calculation of sample size of Cohen; independent variables k can be maximum of 10 (u = 10), Power of Test = 0.80, Effect Size : $f^2 = 0.15$ where Type I error (Alpha, α) = 0.05

$$N = \frac{\lambda}{f^2}$$

where f^2 : (Effect Size) from equation (Cohen, J. 1992: pg.115)

$$f^2 = \frac{R^2}{1 - R^2}$$

substitute f^2 to the main equation, so Calculation Formula of sample size of Cohen,

$$N = \frac{\lambda\,(1 - R^2)}{R^2}$$

For a trail value of v = 120, λ = 17.4 (from table 9.4.2, Cohen, 1988). Substitute λ to sample size equation ($N = \frac{\lambda}{f^2}$), therefore N = 17.4/0.15 = 116, and (v = N-u-1) = 116-10-1 = 105

However, v value at v = 60 and v = 120; N is the best value so

$$\lambda = \lambda_L - \frac{\frac{1}{V_L} - \frac{1}{V}}{\frac{1}{V_L} - \frac{1}{V_U}}\,(\lambda_L - \lambda_U)$$

λ_L = where v = 60
λ_U = where v = 60
V_L = v low
V_U = v high
where v = 60, λ = 18.7, and v = 120, λ = 17.4. substitute λ to equation

$$\lambda = \left[18.7 - \frac{\frac{1}{60} - \frac{1}{105}}{\frac{1}{60} - \frac{1}{120}}(18.7 - 17.4)\right]$$
$$= 17.58$$

therefore

$$N = \frac{\lambda}{f^2} = \frac{17.8}{0.15} = 117$$

Results from calculation many times; it found that λ value at v = 60 and v = 120 will little decrease N, therefore the appropriate value from calculation of sample size is N = 116.

5.3 Calculation of sample size by G*Power Program

It can calculate sample size or power test of many analytical statistics such as exact statistics, t - tests, F - tests, Chi - square, Z tests, ANOVA, Correlation and Regression.

Calculation of sample size by G*Power Program in case of One-way ANOVA; data input into computer program, analytical statistics and other values including types of variables, variables scale, Type I error (Alpha, α), Type II error (Beta, β), Power of Test: 1- β and Effect Size as follow (Cohen, 1988)

Estimated Sample Size for One-way ANOVA (Faul et al., 2009) as shown in the figure 2, 3.

F- test for Group Effect
H_0: delta = 0 versus H_A: delta! = 0
Study Parameters:
Alpha = 0.0500
Power = 0.8000
delta = 0.7014
N_g = 4
m1 = 68.0000
m2 = 72.0000
m3 = 77.0000
m4 = 80.0000
Var_m = 21.1875
Var_e = 43.0650
Estimated sample sizes:
N = 28
N per Group = 7

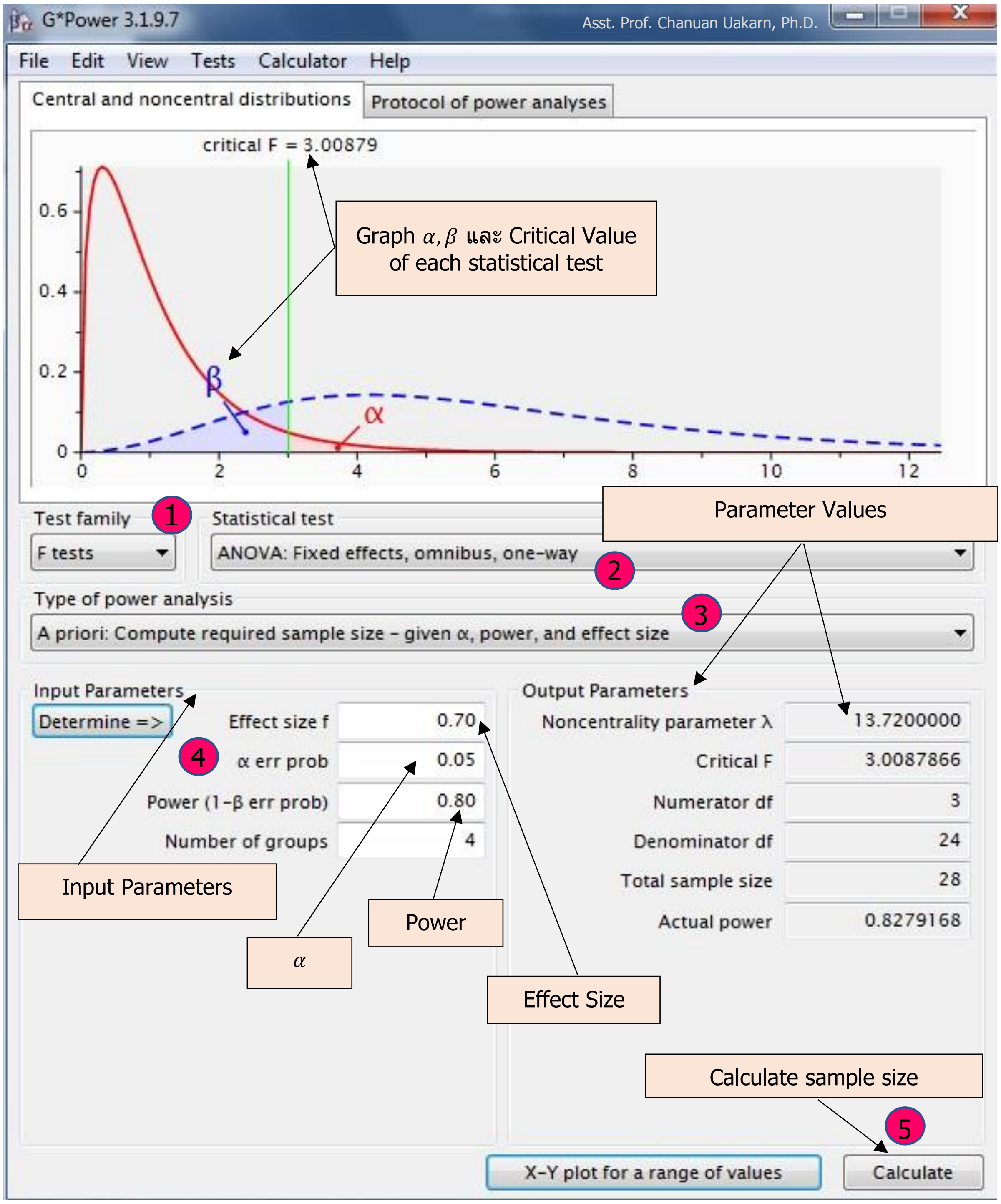

Figure 2 Shows input parameters of various statistical test by G*Power Program in case of One-way ANOVA.

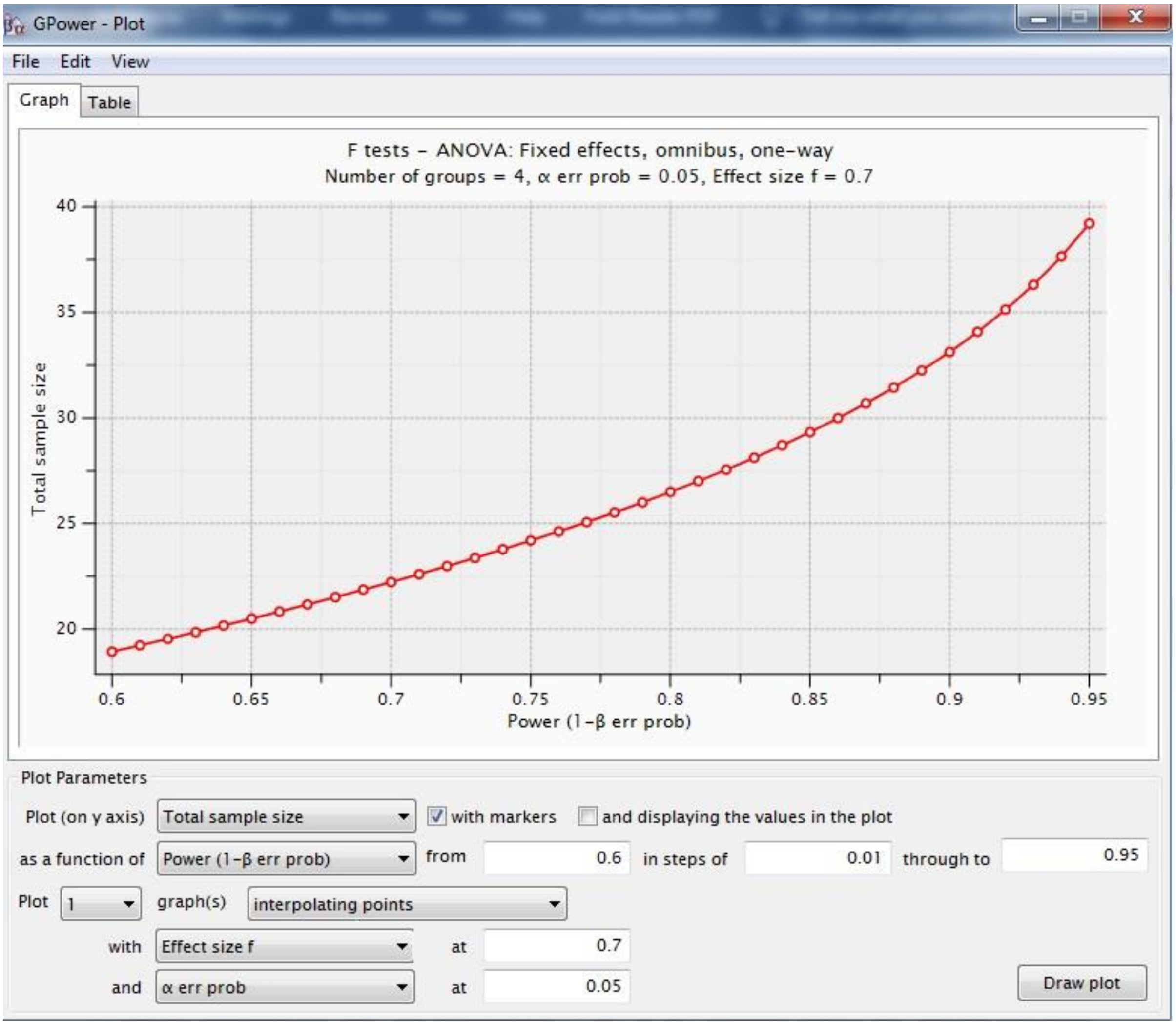

Figure 3 The relationship of total sample size and power of test: Power $(1-\beta)$.

Table 1 Input procedure as statistical values required for calculation of sample size by G*Power (3.1.9.7). (Faul et al., 2009) as shown in the figure 4, 5 and 6.

Statistical Values $\alpha = 0.05$, power (1-β) = 0.80 Effect size = 0.30 and μ^2=0.088						
Procedure			Data Input	Calculation of Sample Size		
Type 1	Type 2	Type 3		Type 1	Type 2	Type 3
1	1	1	Test family	F-Test	F-Test	F-Test
2	2	2	Statistics test	ANOVA: Fixed effect, omnibus, one-way	ANOVA: Fixed effect, omnibus, one-way	ANOVA: Fixed effect, Special, main effects and interactions
3	3	3	Type of power analysis	A priori compute required sample size-given α,	A priori compute required sample size-given α,	A priori compute required sample size-given α,

				power, and effect size	power, and effect size	power, and effect size
–	4	4	Input Parameter	–	Determine=≫	Determine=≫
–	5	–	Select Procedure	–	Effect size from means	–
4	6	5	Number of groups	3	3	3
–	7	–	SD σ within each group	–	18.25	–
–	8	–	Mean	–	45.7, 43.5, 32.6	–
–	9	–	Equal n	–	50	–
–	10	6	Size	–	50	Direct
–	11	7	Total sample size	–	150	Partial μ^2=0.088
–	12	8	Calculate	–	Effect size f=0.3138471	Effect size f=0.31063
5	13	9	Calculate and transfer to main window	Effect size f=0.30 (Cohen)	Effect size f=0.3138471	Effect size f=0.31063
6	14	10	Calculate	Total sample size 111	Total sample size 102	Total sample size 178
			Critical F	3.08039	3.08824	1.88578

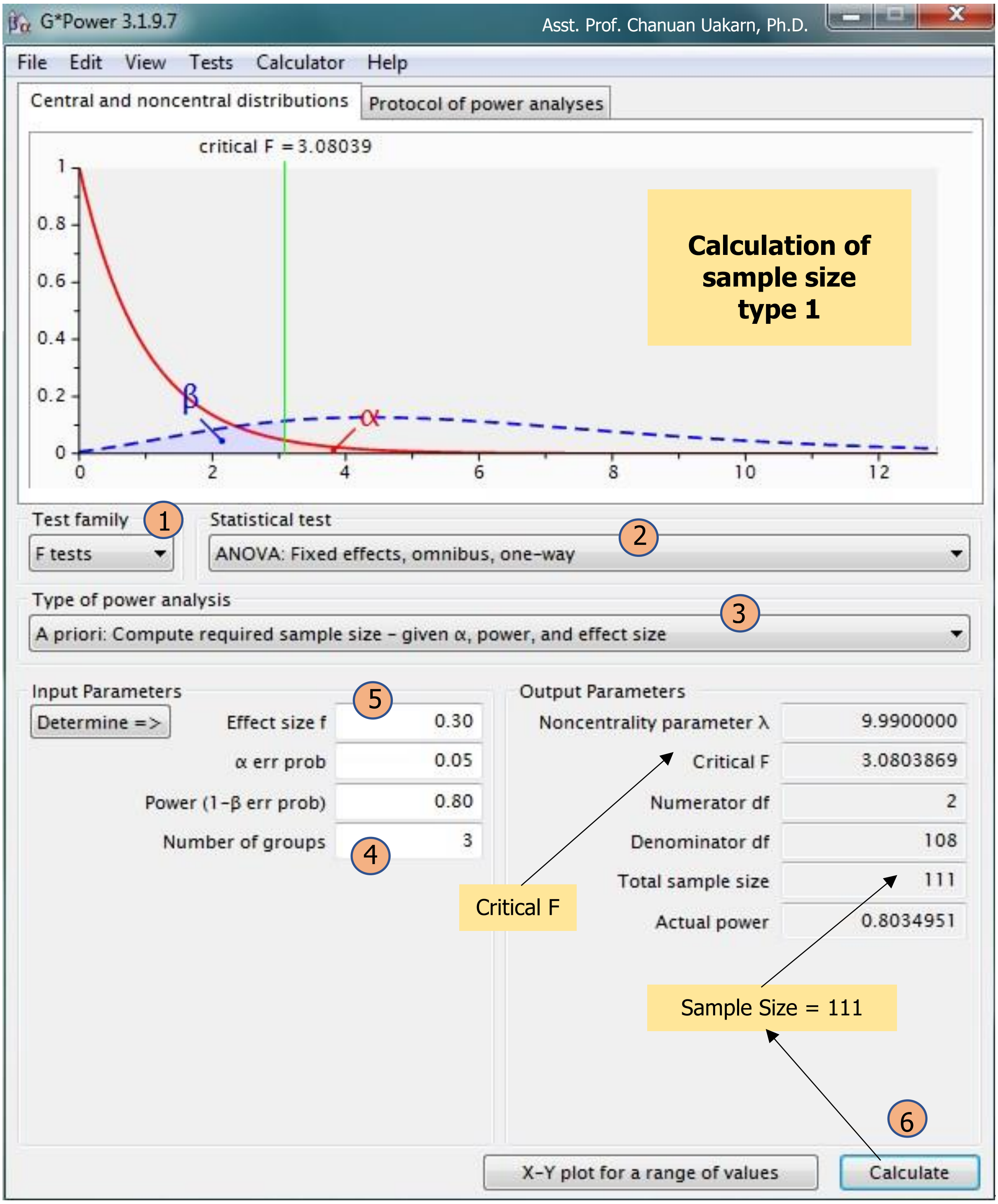

Figure 4 Shows calculation of sample size type 1.

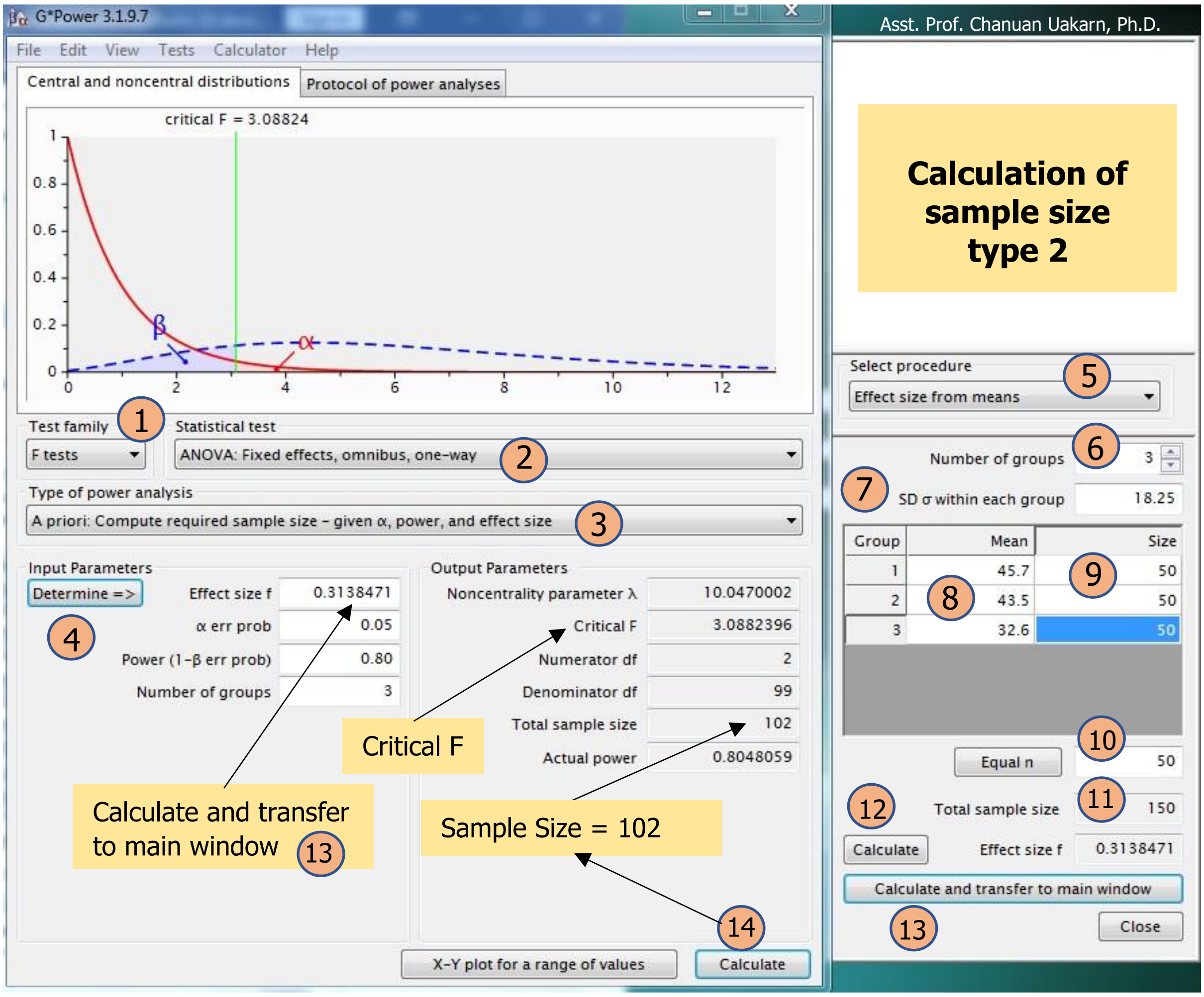

Figure 5 Shows calculation of sample size type 2.

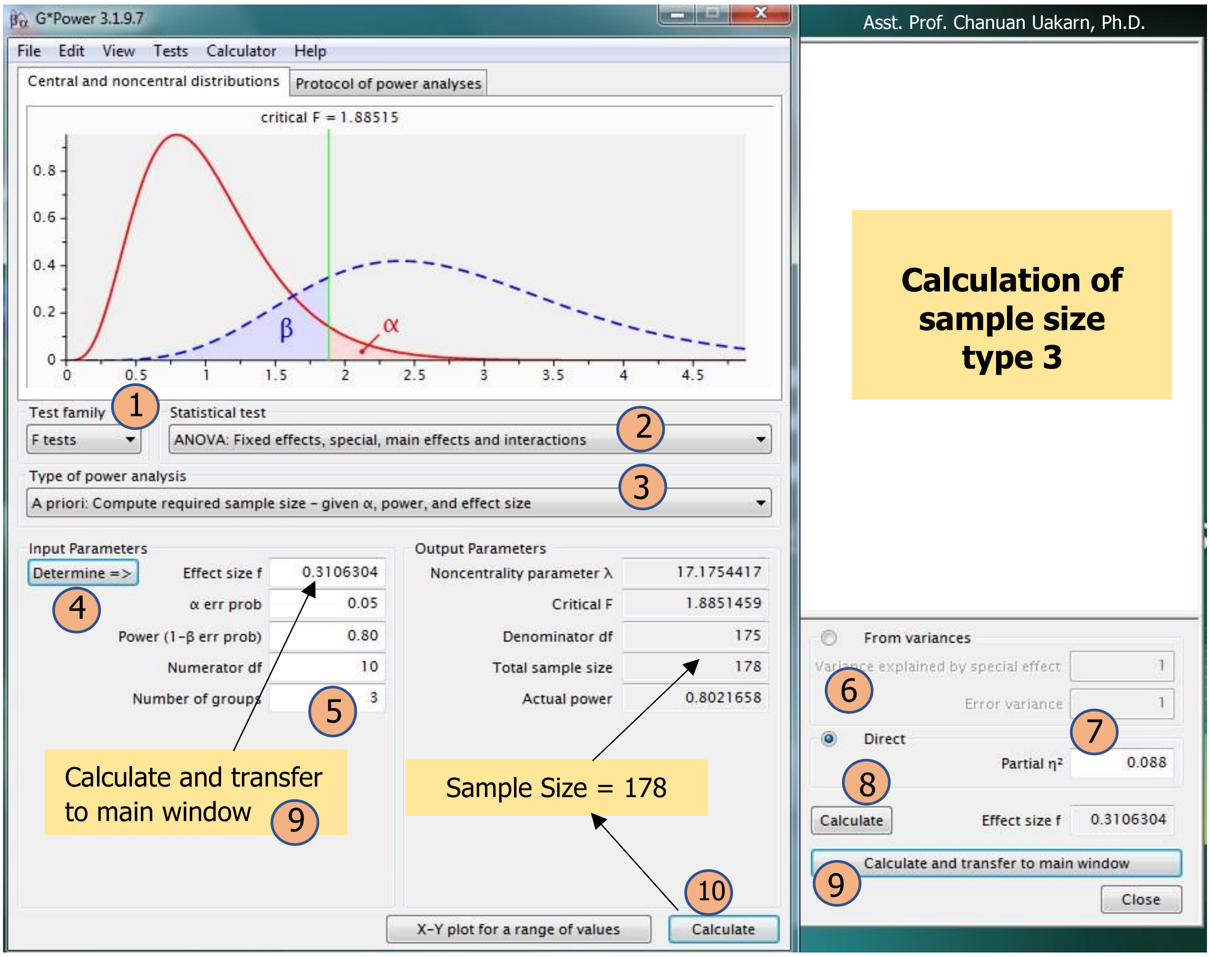

Figure 6 Shows calculation of sample size type 3.

Discussion and recommendation

According to Calculation of Sample Size by formulas and tables of academic officer together with input various statistical values, these are in order to receive the real value of representatives of population and correct sample size. Results from similar experiments such as Taro Yamane and Krejcie & Morgan However, the sample size is still high. In terms of Green and Cohen, the sample size is smaller, or not relatively not close to Cochran these are depending on size or number of inputs, statistical values, acceptable error, power of test, effect size, population size, variable size. The most important is the selection of statistical methods for analyzing Inferential Statistics with parameters such as t-Test, F-Test, z-Test, ANOVA, Regression Analysis or Inferential Statistics without parameters for instance correlation and Chi-Square.

Conclusion and recommendation

Sample Size Determination of Yamane and Krejcie & Morgan are suitable for Survey Research and Finite Population; but not suitable for Experiment Research. In case the population size is unknown but the number is high, population proportion is known or the population size is unknown, population proportion is unknown or the population size is unknown and estimated population mean or the population size is known and estimated the population proportion Cochran Formula is suitable for all these. Green Formula is easy and suitable for Regression Analysis; number of independent variables are only known. Cohen Formula is another choice; the calculation depends on (α), Power of Test, Effect Size, Type I or II error and correspond with calculation by G*Power.

Reference

Chua Lee Chuan. (2006). Sample Size Estimation Using Krejcie and Morgan and Cohen Statistical Power Analysis: A Comparison. *Jurnal Penyelidikan IPBL,* Jilid 7.

Cochran, W. G. (1977). *Sampling techniques.* 3rd Ed. New York: John Wiley & Sons.

Cohen, J. (1977). *Statistical Power Analysis for Behavioral Sciences***.** 1st Ed. New York: Academic Press. StataCorp. 2003. STATA Reference Manual Release 10. Texas: Stata Press.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences***,** 2nd Ed. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Cohen, J. (1992). Quantitative Methods in Psychology: Power Premier. *Psychological Bulletin;* July1992; Vol.112, No.1; pp.155-159

Chanuan Uakarn. (2020). *Advanced Quantitative Research Methods in Public Policy*. Kasem Bundit University: Public Policy and management, Lecture document 2-2020.

Faul, F., Erdfelder, E., Lang, A.-G., and Buchner, A. (2007). *G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior Research Methods,* 39, 175-191.

Faul, F., Erdfelder, E., Buchner, A., and Lang, A.-G. (2009). *Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. Behavior Research Methods,* 41, 1149-1160.

Krejcie, R.V., and Morgan, D.W. (1970) *Determining Sample Size for Research Activities. Educational and Psychological Measurement.* 30, 607 – 610.

Samuel B. Green, (1991). *How Many Subjects Does It Take To Do A Regression Analysis?* Jul 1;26(3), 499-510.

Sedlack R. G., and Stanley J. (1992). *Social Research: Theory and Methods.* 1st Ed., Indiana University: Allyn and Bacon.

Yamane, Taro. (1973), *Statistics: An Introductory Analysis.* London: John Weather Hill, Inc.