

Population

Populasi

Every research is about a group of object, i.e., people, events, companies, things, etc., of interest to the researcher. This group of object is called the **population**. It is actually the scope of the study, about which the researcher wants to make conclusions or inferences.

*Setiap penelitian, selalu mengenai kumpulan object, i.e., orang, perusahaan, benda, dll., yang menjadi pokok perhatian peneliti. Kumpulan objek ini disebut **populasi**. Bisa dikatakan, ini adalah merupakan ruang lingkup dari penelitian itu sendiri dimana tentang populasi inilah kesimpulan hendak ditarik oleh peneliti.*

EXAMPLE #1:

TELADAN:

1. A researcher wants to know performance of small scale and medium enterprises (SME) in Indonesia who are practicing strategic planning. The population in this case is the set of all small scale enterprises which are practicing strategic planning.

Peneliti ingin mengetahui kinerja dari Usaha kecil menengah di Indonesia yang mempraktekkan perencanaan strategis. Untuk kasus ini, populasinya adalah semua usaha kecil menengah yang mempraktekkan perencanaan strategis.

2. Customers of a particular hospital are being studied in relation to their perception on the hospital. The population in this case is the set of customers of the hospital. Not the hospital.

Pelanggan suatu rumah sakit sedang dipelajari berkaitan dengan persepsi mereka terhadap rumahsakit tersebut. Populasi dalam hal ini adalah semua pelanggan dari rumah sakit tersebut. Bukan rumah sakit tersebut.

Because a research is about a population, or is to make inferences about a population, hence all *elements* (note : in general, members of a population are called elements) of the population must be measured (or interviewed).

Karena penelitian selalu mengenai populasi, atau untuk menarik suatu kesimpulan mengenai populasi, maka semua elemen (catatan: pada umumnya anggota dari populasi disebut elemen) dari populasi itu harus diukur (atau diinterview).

EXAMPLE #2:

TELADAN:

Suppose that the population of small scale and medium enterprises (SME) in Indonesia who are practicing strategic planning consists of only 10 firms. A researcher wants to test the claim: "that if a company exercises strategic planning their profit will increase with average rate of increase above 10 billion IDR". The hypothesis for such problem would be :

Misalkan bahwa populasi dari usaha kecil menengah (UKM) di Indonesia yang mempraktekkan perencanaan strategis hanya terdiri dari 10 perusahaan saja. Peneliti ingin menguji klaim : “bahwa apabila perusahaan mempraktekkan perencanaan strategis maka keuntungan perusahaan akan meningkat dengan rata-rata peningkatan diatas 10 Milliyar rupiah”. Secara statistic hipotesis tersebut dirumuskan sebagai berikut:

$$H_0 : \beta = 10$$

$$H_1 : \beta > 10$$

Where β is the population average increase in profit (*note: population parameters always uses Greece letter*). Suppose X is the random variable representing increases in profit. Because the population size is only ten, and the research is about making conclusion on the population, then the researcher decided to collect all x_i for all member of the population, i.e., $x_1 = 9.8$; $x_2 = 10.5$; $x_3 = 9.6$; $x_4 = 11$; $x_5 = 12$; $x_6 = 13$; $x_7 = 9$; $x_8 = 12$; $x_9 = 11$; $x_{10} = 10$. He then calculate the average increases:

Di mana β adalah rata-rata peningkatan keuntungan (catatan: parameter populasi dalam statistika selalu menggunakan huruf Yunani). Misalkan X adalah peubah acak yang mewakili peningkatan keuntungan. Karena besarnya populasi hanya 10 perusahaan, dan penelitian sejatinya adalah mengenai populasi, maka peneliti memutuskan untuk mengumpulkan semua data, x_i , dari populasi tersebut, yaitu semua perusahaan, i.e., $x_1 = 9.8$; $x_2 = 10.5$; $x_3 = 9.6$; $x_4 = 11$; $x_5 = 12$; $x_6 = 13$; $x_7 = 9$; $x_8 = 12$; $x_9 = 11$; $x_{10} = 10$. Kemudian peneliti lalu menghitung rata-rata peningkatan keuntungan sebagai berikut:

$$\beta = \frac{\sum_{i=1}^{i=10} x_i}{10} = \frac{9.8+10.5+9.6+11+12+13+9+12+11+10}{10} = 10.79$$

(*note: the population average is β not $\hat{\beta}$ which is a symbol used for sample average*)

Since $\beta = 10.79$ is greater than 10, then we reject H_0 or accept H_1 , meaning that the claim is substantiated.

Karena $\beta = 10.79$ lebih besar dari 10, maka hipotesis H_0 ditolak dan hipotesis H_1 diterima, yang artinya bahwa klaim, yaitu keuntungan meningkat dengan rata-rata peningkata diatas 10 M adalah didukung penelitian ini.

Sample and Sample size

Contoh dan Besarnya Contoh

There are many cases where to collect data about a population is prohibitively expensive if not impossible at all. A case at hand for example is, suppose we want to check a claim by a tooling company producing bottling machine capable of producing 1.5 liters of coke in each bottle. The population for this study of course would be all bottles of coke produced now and in the future.

Since future productions have not been produced, hence there is no way of knowing the exact population in this case.

Ada banyak kasus dimana mengumpulkan data dari populasi sangat mahal, bahkan bisa saja tidak mungkin sama sekali. Sebagai contoh, misalkan kita ingin mengecek apakah klaim oleh perusahaan peralatan-peralatan yang memproduksi mesin pengisian botol yang mampu mengisi 1.5 liter tiap botol coca cola. Populasi untuk studi seperti ini tentu semua botol coke yang diproduksi sekarang dan dimasa yang akan datang. Oleh karena produksi coke di masa datang belum dihasilkan, maka tidak ada cara untuk mengetahui populasi yang pasti untuk kasus ini.

In the absence of ability to measure all elements of the population, we can still make conclusion or inferences about the population, i.e., by taking only part of the population, called **sample**, a sub set of the population. The statistics generated from samples are called **estimates**, i.e., estimators of the population parameters.

*Manakala kita tidak mampu mengukur semua elemen dari populasi, kita masih bisa membuat kesimpulan tentang populasi, i.e., dengan mengambil dan mengukur hanya sebagian dari populasi tersebut, yang disebut **sample**, yaitu sub-set dari populasi. Statistik yang dihasilkan dari sample disebut **penduga**, yaitu **penduga** dari parameter populasi tersebut.*

EXAMPLE #3:

TELADAN #3:

In EXAMPLE #2 the population parameter is the population average increase in profit. Suppose we take a sample of 4 out of 10 SMEs (*exercise: try to find out how many samples of 4 can be generated from the population of 10 SMEs*). From one possible sample, you can calculate the sample average, usually we give notation $\hat{\beta}$ (beta hat) to represent estimates. Can we still make conclusion about the population from this sample of 4? The answer is yes; of course with some degree of confidence, or **confidence level** $(1 - \alpha)$, where α is the significance level, that is, you can only trust, or rely upon the result of the hypothesis testing by 95 % for instance. Significance level α is the allowed error committed by rejecting a true Null Hypothesis (H_0). This is dependent on you, the researcher. For instance, if you are willing to take significance level of 5 %, i.e., allowable error of 5 %, then the confidence level in this case would be $(1 - 5\%) = 95\%$.

*Pada TELADAN #2, parameter populasi adalah rata-rata kenaikan untung dari populasi. Misalkan kita mengambil sample (contoh) sebanyak 4 dari 10 elemen populasi SME tersebut (Latihan: coba hitung berapa banyak contoh berukuran 4 yang bisa dibentuk dari 10 elemen populasi tersebut?). Dari salah satu contoh, anda dapat menghitung rata-rata contoh, biasanya diberi notasi $\hat{\beta}$ (beta topi) untuk menyatakan penduga. Apakah masih bisa kita hasilkan kesimpulan tentang populasi dari contoh yang hanya 4 unit ini? Jawabannya adalah Ya; sudah barang tentu dengan tingkat kesalahan tertentu, atau dengan **tingkat kepercayaan** sebesar $(1 - \alpha)$, dimana α adalah **tingkat nyata**, yaitu, bahwa kita hanya dapat percaya, atau hanya dapat mengandalkan hasil uji hipotesis tersebut sebesar 95% misalnya. Tingkat nyata sebesar 5 % adalah besarnya error yang diizinkan sebagai akibatberbuat kesalahan, yaitu, menolak hopotesis nol (H_0) yang benar. Besarnya α*

ini tergantung pada anda, peneliti. Sebagai contoh, bila anda bersedia menetapkan tingkat nyata sebesar 5 %, i.e., error yang diizinkan sebesar 5 %, maka tingkat kepercayaan terhadap hasil uji nantinya adalah sebesar (1 – 5 %), yaitu 95 %.

From EXAMPLE #3, It can be seen that there is a risk in term of error committed by making conclusion from only part of the population – a sample. The size of the error, sampling error that is, is dependent of the sample size (n). The larger the sample size, the smaller would be the error. In this module though, we will turn this around by first deciding on the size of error we are willing to take, then, based on that we decide on the size of the sample. The second thing that we need to decide in determining the sample size is the level of precision of the estimates. In sampling theory, this error is called **margin of error (e)**, which is basically the precision level of our estimate, i.e., how far would be the interval $\hat{\beta} \pm e$ where we want our population parameter β to fall upon. Therefore, the second thing we need to decide to determine the sample size is the size of the margin of error (**e**) (**note: e is not the same as α , or significant level**).

*Dari TELADAN #3, dapat dilihat bahwa ada resiko dalam bentuk error sebagai akibat membuat kesimpulan dari hanya sebagian dari populasi – atau contoh. Besarnya error ini, yaitu sampling error, tergantung pada besarnya contoh (n). Makin besar ukuran contoh, maka konsekuensi errornya mekin kecil. Dalam modul ini logikanya akan kita balik dengan lebih dulu menentukan besarnya error yang bersedia kita terima, lalu berdasarkan besar error yang bersedia kita terima tersebut, lalu kita tentukan besarnya contoh. Hal kedua yang perlu kita putuskan sebelum menentukan besarnya contoh adalah tingkat ketelitian yang kita mau dari penduga. Dalam teori penarikan contoh, ini disebut **margin of error (e)**, yang pada prinsipnya merupakan ukuran tingkat ketelitian dari penduga kita, i.e., besarnya interval $\hat{\beta} \pm e$ di mana kita menghendaki nilai parameter β dari populasi (yang kita tidak tahu) berada di dalamnya. Oleh karena itu hal kedua yang perlu kita tentukan dalam menentukan besarnya contoh adalah tingkat ketelitian yang kita inginkan yaitu besarnya margin of error (e)nya (**catatan: e tidak sama dengan α , atau tingkat nyata**).*

Back to the first error that has to be considered, the allowable error, which has to do with the confidence level, it is basically the **probability** that the population parameter β is within the range or interval of $\hat{\beta} \pm e$. In other words, confidence level can be written as follows:

*Kembali kepada error yang pertama yang harus dipertimbangkan dalam menentukan besar contoh, yaitu error yang diizinkan, yang berkaitan dengan tingkat kepercayaan, ia pada dasarnya merupakan **peluang** bahwa parameter populasi β berada pada selang $\hat{\beta} \pm e$. Dengan kata lain tingkat kepercayaan dapat ditulis sebagai berikut :*

$$\text{Probability } (\hat{\beta} - e < \beta < \hat{\beta} + e) = 1 - \alpha \dots\dots\dots (1)$$

If X is the random variable having β as its population mean, and follows a Normal Distribution with standard deviation σ , i.e., statistically written as follows,

Bila X adalah peubah acak dengan nilai tengah (rata-rata) populasi β , dan mengikuti distribusi normal dengan simpangan baku σ , i.e., yang dalam notasi statistika sebagai berikut,

$$X \sim N(\beta, \sigma^2)$$

then, following The Central Limit Theorem, $\hat{\beta}$, i.e., the sample average will follow a Normal Distribution with mean β and standard deviation $\frac{\sigma}{\sqrt{n}}$ in statistical notation it is written as follows:

$$\hat{\beta} \sim N(\beta, \frac{\sigma^2}{n})$$

Given the above information about $\hat{\beta}$, then the following statistic,

$$z = \frac{\hat{\beta} - \beta}{\frac{\sigma}{\sqrt{n}}} \dots\dots\dots (1c)$$

follows standard normal distribution.

Using equation (1c), then equation (1) can be re-written as follows

maka persamaan (1) dapat ditulis kembali sebagai berikut,

$$\text{Probability} \left(\hat{\beta} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \beta < \hat{\beta} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right) = 1 - \alpha \dots\dots\dots (2)$$

Where $Z_{\frac{\alpha}{2}}$ is the standard normal variate value of the standard normal distribution such that $\text{Prob} (z > Z_{\frac{\alpha}{2}}) = \alpha/2$.

Dimana $Z_{\frac{\alpha}{2}}$ adalah nilai peubah normal baku dari distribusi normal baku, dimana Peluang $(z > Z_{\frac{\alpha}{2}}) = \alpha/2$.

From equation (1) and (2) then,

Dari persamaan (1) dan (2) kemudian diturunkan persamaan berikut,

$$e = Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \dots\dots\dots (3)$$

By simplifying equation (3) we get the sample size,

Degan menyederhanakan persamaan (3) selanjutnya kita dapat rumus besarnya contoh sebagai berikut,

$$n = \frac{Z_{\frac{\alpha}{2}}^2 \sigma^2}{e^2} \dots\dots\dots (4)$$

In conclusion, to determine the sample size, n, you need to determine (1) the significant level (α) or confidence level ($1 - \alpha$), (2) the error margin or the degree of precision, **e**, and (3) the

variance of the random variable, σ^2 . Notice, that the sample size is not dependent on the population size.

Sebagai kesimpulan, untuk menentukan besarnya contoh, n, perlu ditentukan terlebih dahulu (1) tingkat nyata (α) atau tingkat kepercayaan ($1 - \alpha$), (2) error margin atau tingkat ketelitian, e, dan (3) ragam dari peubah acak X, yaitu σ^2 . Perhatikan, besarnya contoh tidak tergantung pada besarnya populasi.

EXAMPLE # 4: Determining sample size for estimation of the Mean

TELADAN # 4: Penentuan besarnya contoh untuk menduga Rata-rata

We would like to estimate the mean teacher's salary in the East Jakarta High school, with 95% confidence level, to an accuracy within Rp. 100.000. In this case we have literally no idea what σ would be, because we haven't even do the survey to estimate the value of σ . In such case, we have to deduce the best possible values which could be from previous study or other similar setting, and suppose was given to be Rp. 1000.000. So in the absence of anything better, let's use that as our guess for σ .

Kita hendak menduga gaji rata-rata guru Sekolah Menengah Atas di Jakarta Timur dengan tingkat kepercayaan 95 %, dan dengan tingkat ketelitian plus-minus Rp 100.000,-. Dalam hal ini kita tidak mengetahui sama sekali besarnya simpangan baku (σ), karena sama sekali survey belum dilakukan untuk menduga nilai dari σ . Dalam keadaan seperti ini kita terpaksa harus menduga nilai σ yang terbaik, bisa saja dari studi terdahulu atau dugaan dari situasi yang mirip di tempat lain, dan katakanlah sebesar Rp. 1000.000,-. Oleh karena itu, dalam situasi dimana tidak ada informasi yang lebih baik lagi, mari kita gunakan itu sebagai pengganti dari σ .

In this case the value 95 % percent confidence level translates to a $z = 1.96$ (see Table Z). Therefore, equation (4) become:

Dari nilai tingkat kepercayaan 95 %, kita bisa dapatkan nilai $z = 1.96$ (lihat Tabel Z). Oleh karena itu, persamaan (4) menjadi :

$$n = \left(\frac{(1.96)(1000000)}{100000} \right)^2 = 384.16$$

or 384 teachers, rounding to the nearest whole number. (**Note:** formula (4) actually should use Student-t distribution – and hence t-Table; however, for a large sample size, $n > 30$, student-t can be closely approximated by standard normal – z distribution).

atau sebesar 384 guru, dengan membulatkan ke atas. (Catatan: persamaan (4) sebenarnya harus menggunakan distribusi student-t – oleh karenanya menggunakan Tabel t; akan tetapi, untuk contoh, n. yang cukup besar, yaitu > 30 , student-t dapat didekati oleh sebaran normal baku – sebaran Z).

EXERCISE: try to change the precision level in EXAMPLE #4 from Rp. 100.000 to Rp. 500.000, find out what would be the sample size? Comment on the result.

Coba ubah tingkat ketelitian pada TELADAN #4 dari Rp. 100.000 menjadi Rp 500.000. Hitunglah besarnya contoh. Apa komentar anda dari hasil tersebut?

Often times, the value that we want to estimate is not the population mean, but the proportion (p), e.g., the proportion of Binusian who endorse gay lifestyle for example. In such case, formula (4) would still be the same except that the variance (σ^2) should be replaced by the variance of proportion, which is $p(1-p)$. Again determining the value for proportion would still be a problem. As usual, we might resort to using result of previous study for this. Notice however, that proportion variance would be maximum when $p = 0.5$ (50 %). Therefore calculating sample size using $p = 0.5$ will always generate sample size that is bigger than the required ones; so it always safe.

Sering terjadi dimana nilai dari peubah yang ingin kita duga bukan rata-rata populasi, tapi proporsi dari populasi (p), e.g., proporsi dari Binusian yang mendukung lifestyle gay misalnya. Dalam kasus seperti ini, rumus (4) tetap sama kecuali ragam (σ^2) harus diganti dengan ragam dari proporsi, yaitu $p(1-p)$. Lagi-lagi, dalam penentuan nilai dari proporsi, p , tetap menjadi masalah. Seperti biasanya, kita sering sekali jadinya menggunakan nilai proporsi dari penelitian sebelumnya atau yang mirip. Perhatikan tapinya, ragam dari proporsi akan maksimum pada saat proporsi $p = 0.5$ (50 %). Oleh karena itu perhitungan besarnya contoh dengan menggunakan $p = 0.5$ selalu memberikan besar contoh yang lebih besar dari yang dibutuhkan; oleh sebab itu, menggunakan nilai ini akan selalu aman.

EXAMPLE #5: Sample size to estimate a proportion

TELADAN #5: Besar contoh untuk menduga proporsi

A professor in BINUS's Sociology department is trying to determine the proportion of students who support gay lifestyle. How large a sample size should be needed?

Seorang professor di Departemen sosiologi di Universitas Binus hendak menduga proporsi mahasiswa yang mendukung gaya hidup gay di Binus. Berapa besar contoh yang dibutuhkan?

As usual, she needs to determine first the margin of error, say 0.05 (5 %). Secondly, she also need to decide on the degree of confidence, say it is 95 % (or significance level of 5 %), so that $z = 1.96$. Also she decided to use the maximum variance of the proportion, i.e., with $p = 0.5$. Now, using a modified formula, having variance as $p(1-p)$, based on equation (4), we have:

Seperti biasa, pertama dia perlu mentukan margin of error, katakanlah 0.05 (5 %). Kedua, dia juga perlu menentukan tingkat kepercayaan dari dugaan itu nantinya, katakanlah 95 % (atau tingkat nyatanya adalah 5 %), sehingga $z = 1.96$. Kemudian dia juga memutuskan menggunakan keragaman dari proporsi yang maksimum, yaitu $p = 0.5$. Sekarang dengan menggunakan formula yang telah dimodifikasi, dengan mengubah ragam menjadi $p(1-p)$, maka berdasarkan rumus (4) kita dapatkan :

$$n = \frac{Z_{\alpha/2}^2(p(1-p))}{e^2} = \frac{1.96^2(0.5(1-0.5))}{0.05^2} = 384.16$$

or, the sample size is 384 students, i.e., by rounding down to the nearest whole number. *atau, besarnya contoh adalah 384, yaitu dengan membulatkan ke bilangan bulat terdekat.*

- EXERCISES:** (1) try to change the precision level or margin of error from 0.05 (5 %) to 0.025 (2.5 %). Find out what would be the sample size? Comment on the result.
- (2) try to change **p**, from 0.5 to 0.2, with the same precision level of 0.05 (5%). What is the new sample size. Comment on the result.
- (1) *Coba ubah tingkat ketelitian atau margin error dari 0.05 (5 %) ke 0.025 (2.5 %). Hitunglah kembali berapa besarnya contoh? Berikan komentar tentang hasil tersebut.*
- (2) *Coba ubah proporsi p, dari 0.5 menjadi 0.2, dengan tingkat ketelitian yang sama pada tingkat 0.05 (5%). Berapa besar contoh yang baru? Berikan juga komentar terhadap hasil ini.*

From the two examples discussed above it can be seen that if the variable of interest is measuring proportion, then using **p** = 0.5 will always give a conservative sample size, i.e., is always safe and at the same time not too large in excess of the sample size calculated in the presence of estimates of the proportion. However, if the variable of interest is the mean (not proportion), even if we can have estimates of σ^2 from previous studies, we might still have problems, because in most studies, you will most likely have many variables of interest and also they would most likely have different variances. However in the presence of many variables or for that matter multi variates situation, you could kind of think of having one of them as proportion, or there you would always be able to create proportion variables among these variables even if it is basically not needed. Given that the use of formula to derive sample size for proportion assuming the maximum level of variability (**p** = 0.5) is generally produce a more conservative sample size, including as compare to calculating sample size for the mean, than it is suggested to use this formula even if we are working on estimating the mean.

Dari dua teladan yang didiskusikan di atas dapat dilihat bahwa bila variable yang menjadi pokok perhatian kita adalah mengukur proporsi, maka dengan memilih p = 0.5 akan selalu memberikan besar contoh yang konservatif, yaitu, akan selalu memberikan besar contoh yang aman dan juga tidak akan memberikan besar contoh yang jauh melebihi besar contoh yang dihitung bila proporsi populasi diketahui. Akan tetapi, bila variable yang menjadi pokok perhatian kita adalah rata-rata populasi (bukan proporsi), sekalipun penduga dari ragam (σ^2) dari penelitian terdahulu diketahui, kita masih menghadapi masalah, karena pada setiap studi, kita akan selalu berhadapan dengan banyak variable yang menjadi pokok perhatian kita, dan hampir pasti ragam dari masing-masing variable ini pun akan berbeda satu sama lain. Akan tetapi dalam situasi dimana banyak variable, kita bisa mengasumsikan seolah-olah ada variable proporsi diantara variable tersebut, sekalipun pada prinsipnya tidak dibutuhkan. Bahwa dengan menggunakan rumus besarnya contoh untuk kasus proporsi dengan menggunakan p = 0.5 akan selalu memberikan besar contoh yang selalu aman atau konservatif, termasuk dibandingkan

dengan besar contoh untuk kasus yang bukan proporsi (rata-rata populasi), maka dianjurkan untuk selalu menggunakan rumus proporsi dengan $p = 0.5$. Pembuktian bahwa inilah yang paling konservatif diuraikan di bawah ini.

Other things to consider is first, in deriving the above formula it is assumed or based on a simple random sampling design (simple random sampling technique) of selecting the samples. Special attention need to be made in the case of using other sampling techniques. Secondly, that the formula is only to ensure representativeness of the sample given a certain level of acceptable error and confidence level. A particular data analysis technique often requires certain minimum number of samples, or more accurately, certain minimum number of observations. As a general rule, would be to use the largest number amongst the two. For example, if the sample size based on the formula is 350, and the requirement of the data analysis is 200, then we should go for 350. On the other hand, if the data analysis requirement is 300 and the calculated sample size is 250, then we should use 300 as sample size.

Hal lain yang perlu mendapat perhatian, pertama, bahwa dalam menurunkan rumus di atas, diasumsikan atau didasarkan atas pengambilan contoh secara acak sederhana (teknik penarikan contoh acak sederhana) dalam memilih contohnya. Perhatian khusus perlu apabila teknik penarikan contoh dilakukan selain dengan cara acak sederhana. Kedua, bahwa formula tersebut bertujuan hanya untuk memastikan bahwa besarnya contoh tersebut cukup untuk **mewakili** populasi dengan tingkat margin error dan tingkat kepercayaan yang telah ditetapkan. **Data analisis tertentu** kadang-kadang membutuhkan sejumlah minimum contoh (ulangan atau pengamatan) tertentu. Sebagai aturan umum, kita menggunakan yang terbesar dari kedua persyaratan tersebut. Sebagai contoh, bila besarnya contoh berdasarkan rumus yaitu dalam memenuhi persyaratan perwakilan adalah 350, dan kebutuhan observasi sesuai dengan kebutuhan teknik data analisis memerlukan 200, maka kita harus memilih banyaknya contoh sebesar 350. Sebaliknya, bila teknik data analisis membutuhkan banyaknya observasi sebesar 300 sedangkan besar contoh sesuai rumus (4) adalah 250, maka kita harus menggunakan angka 300 sebagai besarnya contoh.

A Note on Slovin's Formula

Sebuah Catatan Berkaitan dengan Formula Slovin

Many researchers use the so called Slovin's formula in determining sample size, including in writing thesis and/or skripsi in Binus. Some of them indicate a wrong understanding of the formula and the uses of the formula. For this reasons, it is important here to elaborate the situation where the formula can be used and understanding of components of the formula including how the formula is derived. The Slovin's formula itself is as follows:

Banyak penelitian yang menggunakan rumus Slovin dalam penentuan besarnya contoh dalam penelitian mereka, termasuk thesis atau skripsi di Binus. Sebagian diantaranya menunjukkan pemahaman yang sedikit keliru dengan penggunaan dan pengertian rumus tersebut. Oleh karena itu sangat penting di sini diuraikan dalam kasus atau kondisi seperti apa rumus tersebut bisa digunakan dan pengertian tentang apa yang dimaksud dengan komponen-komponen dari rumus tersebut. Slovin's formula adalah sebagai berikut:

$$n = \frac{N}{1 + Ne^2} \dots\dots\dots (5)$$

Where n = sample size; N = population size; and e is the margin of error.

Dimana n = besarnya contoh; N = besarnya populasi; dan e adalah margin of error.

Actually whether Slovin is the name of a person or a aliases is not clear; in fact the name can not be found in literature as a person who derive the formula. The formula itself is presented and derived by Yamane (Yamane, 1967), eventhough he never label the formula as Slovin's formula. Notice, formula (5) is not dependent on the variance. For this reason, it is a bit difficult to accept the same sample size for differen population with different variances but only having the same population size (N). This is one of the flaws in using this formula without clear understanding of it.

Sebenarnya kata Slovin apa berkaitan dengan nama orang atau nama samaran tidak jelas; bahkan literature di mana Slovin sebagai penulisnya tidak pernah ditemukan di pustaka. Rumus Slovin sendiri pernah diturunkan di literature yaitu dalam Yamane (1967) sekalipun tidak diberi label "Slovin". Perhatikan, rumus (5) tidak tergantung sama sekali dengan ragam (variance). Oleh karena itu adalah sangat aneh kalau kita menentukan besarnya contoh yang sama untuk beberapa populasi dengan besar populasi yang sama (N) namun dengan keragaman yang berbeda. Inilah salah satu kesalahan besar yang akan terjadi bila menggunakan rumus ini.

The second mistake that often occurs in using the formula is that researchers often do not distinguish the differen between margin of error (e) and significant level (α). For example, user of this formula often times would say : "with margin of error of 5 % which is the same as saying that the confidence level of 95 %". Eventhough you could have confidence level of 95 % but the margin of error does not necessarily 5 %, it could be 1 % for example, because both are two different concepts (see explanation of these concepts above).

Kesalahan kedua yang sering terjadi bila menggunakan rumus ini adalah tidak membedakan antara margin of error (e) dan tingkat nyata (α). Misalnya pengguna rumus ini akan sering mengatakan demikian: " dengan margin of error 5 % atau sama dengan tingkat kepercayaan 95 %.....". Pada hal bisa saja dengan tingkat kepercayaan 95 % margin of errornya 1 % misalnya, karena dua hal tersebut merupakan dua konsep yang berbeda (lihat penjelasan sebelumnya).

The third, if a researcher use confidence level different than 95 %, for example 90 %, and stated that margin of error is 10 %, this case also is wrong, because the Slovin's formula is only working, or derived given confidence level only for 95 %. Margin of error could be any figure. This is explained in the following derivation of the formula.

Ketiga, bila seorang peneliti misalnya menggunakan tingkat kepercayaan selain 95 %, misalnya 90 %, dan mengatakan bahwa margin of errornya adalah 10 %, inipun adalah salah, karena rumus Slovin ini hanya berlaku untuk tingkat kepercayaan 95 %. Margin of error bisa berapa saja. Hal ini akan diuraikan melalui penurunan rumus Slovin berikut.

Cochran (1977) derived the formula of sample size for the case of proportion variable, given finite population size as follows:

Cochran (1977) menurunkan rumus besarnya contoh untuk kasus variable proporsi (P) dalam keadaan populasi yang terbatas (finite population) adalah sebagai berikut:

$$n = \frac{n_o}{1 + \frac{n_o}{N}} \dots\dots\dots (6)$$

Where n = sample size; N = population size; and

Dimana n = besarnya contoh; N = besarnya populasi; dan

$$n_o = \frac{Z_{\alpha}^2(p(1-p))}{e^2} \dots\dots\dots (7)$$

with proportion p = 0.5, that is the proportion that gives the maximum proportion variance; and with confidence level of 95 %, so that z = 1.96, rounded up to 2; then formula (7) become

dengan p = 0.5, yaitu proporsi yang memberikan ragam proporsi terbesar; dan dengan tingkat kepercayaan 95 %, sehingga z = 1.96, dibulatkan menjadi 2, maka rumus (7) menjadi

$$n_o = \frac{2^2(0.5(1-0.5))}{e^2} = \frac{1}{e^2} \dots\dots\dots (8)$$

Substituting equation (8) into equation (6) we then have

Dengan mensubstitusikan persamaan (8) ke persamaan (6) kita dapatkan persamaan berikut:

$$n = \frac{1/e^2}{1 + \frac{1/e^2}{N}} = \frac{N}{1 + Ne^2} \dots\dots\dots (9)$$

Which is basically the Slovin’s formula. Therefore, if we use Slovin’s formula, the assumptions are (1) confidence level is 95 %; (2) the proportion is assumed to be 5 % (0.05). The margin of error can be any value.

Yang merupakan rumus Slovin. Jadi kalau menggunakan rumus Slovin, asumsinya adalah (1) tingkat kepercayaan adalah 95 %; (2) proporsi diasumsikan 5 % (0.05). Margin error bisa berapa saja.

Sampling Technique

Cara Penarikan Contoh

Having decided the sample size, the number of elements of the sample to be drawn from the population – *element of the sample is called subject*, the next thing to be done in sampling is how the subjects are chosen or the sampling technique. There are two ways to choose the subjects, number (1) is the probabilistic technique, and number (2) is the non-probabilistic

technique. The probabilistic technique is a method whereby each sample chosen is ensured to have equal probability. The non-probabilistic technique is a method whereby there is no known probability of a sample being chosen let alone having equal probability. If in the data analysis, statistical inferences are going to be used, then the method of choosing the sample should be probabilistic. When the situation is as such that almost impossible to use probabilistic sampling, then we have to explain and elaborate as much as possible the justification that the non-probabilistic technique chosen follows follow probabilistic sampling technique closely.

*Setelah mendiskusikan bagaimana menentukan besarnya contoh, yaitu penentuan banyaknya **subjek** dari contoh yang akan diambil dari populasi – element dari contoh disebut **subjek**, hal berikutnya yang harus dilakukan dalam penarikan contoh adalah bagaimana subjek nya dipilih atau disebut teknik penarikan contohnya. Ada dua cara memilih subjek, pertama (1) adalah dengan cara probabilistik, dan yang kedua (2) adalah dengan cara non-probabilistik. Teknik probabilistic adalah metode dimana setiap contoh dipilih dengan memastikan bahwa peluang terpilih harus sama. Cara non-probabilistik adalah metode dimana tidak ada cara untuk mengetahui peluang terpilihnya contoh tertentu, apalagi untuk memastikan bahwa semua contoh memiliki peluang yang sama. Apabila pada data analisisnya akan menggunakan statistika inferensial, maka metode penarikan contohnya harus dengan probabilistic. Pada situasi dimana hampir tidak mungkin menggunakan penarikan contoh probabilistic, maka kita harus menjelaskan dan menguraikan cara yang kita tempuh dan menunjukkan sedemikian rupa bahwa cara non-probabilistik yang kita lakukan itu sedapat mungkin telah mendekati metode probabilistic, yaitu, bahwa peluang terpilihnya semua contoh mendekati sama.*

The most common probabilistic sampling technique used is the **simple random sampling** technique. Out of N population members, we choose n subjects randomly. First we number all elements of the population from 1 to N . Then we generate random number between 1 to N , in this case usually using Excel function `RANDBETWEEN(1, N)`. If the sample size is n , we simply call the `RANDBETWEEN()` function n times, or we just copy the formula of function in n cells of the Excel spreadsheet. In the case where the same number appears, then the function can be repeated until different numbers are drawn.

*Cara penarikan contoh probabilistic yang paling umum adalah dengan **teknik penarikan contoh acak sederhana**. Dari sebanyak N anggota populasi, kita memilih sebanyak n subjek secara acak. Pertama kita menomeri semua elemen dari populasi dari 1 sampai N . Lalu kita generate bilangan acak antara 1 sampai N , biasanya dengan menggunakan fungsi EXCEL : `RANDBETWEEN(1, N)`. Apabila besarnya contoh adalah n , maka kita tinggal memanggil fungsi ini sebanyak n kali, atau dengan meng-copy fungsi itu ke sebanyak n cell yang berbeda. Manakala ada pengulangan angka yang sama, maka fungsi tersebut dipanggil kembali sampai menghasilkan angka yang berbeda dari angka-angka sebelumnya.*

EXAMPLE: Simple random sampling technique

TELADAN: Penarikan contoh dengan cara acak sederhana

Recall EXAMPLE # 3. Suppose we assign number for all the SMEs from 1 to 10. In sampling terminology, we call this the **sampling frame** or just **the frame**. To generate a sample having members (or subjects) of 4, we just call `RANDBETWEEN(1,10)` placed the

function in 4 cells in a column. The four SMEs to be chosen would be the corresponding number listed in the four cell.

*Coba kita lihat lagi TELADAN #3. Misalkan kita memberikan nomor bagi semua SME dari 1 sampai 10. Dalam istilah teori sampling daftar populasi yang telah dinomori ini disebut **sampling frame** atau **frame** saja. Menghasilkan contoh sebanyak 4 subjek, kita bisa memanggil fungsi **RANDBETWEEN(1, 10)** sebanyak empat kali dan taro di 4 cell pada kolom tertentu. SME yang terpilih adalah empat SME dengan nomer yang terpilih seperti yang tertera pada keempat cell tersebut .*

Other probabilistic sampling techniques are: (1) stratified random sampling technique – this technique is used when the population can be grouped having more homogeneity within group as opposed to between group; in addition this method can be utilized when we are interested in comparing groups, (2) cluster random sampling – this is used when the frame is not available; so we create clusters that have list (frame). We choose using simple random sampling from the list of clusters, and then do complete enumeration within a cluster, (3) systematic sampling - Every n^{th} element in the population is sampled, starting from a randomly chosen element, (4) area - It is actually a cluster sampling confined to a particular area sampling, (5) double sampling - Collect preliminary data from a sample, and choose a sub-sample of that sample for more detailed investigation.

Teknik penarikan contoh probabilistic yang lainnya antara lain : (1) stratified random sampling technique – tehnik ini digunakan manakala populasi dapat dikelompokkan sedemikian rupa dimana masing-masing kelompok lebih homogeny relative disbanding dengan kelompok lainnya; disamping itu, tehnik ini dapat digunakan apabila kita tertarik juga untuk membandingkan antar kelompok. (2) Cluster random sampling – tekni ini digunakan manakala framenya tidak tahu atau tidak tersedia; oleh karena itu kita menciptakan clusters yang ada daftarnya atau framenya. Selanjutnya kita memilih cluster dengan acak sederhana dari daftar cluster ini, dan selanjutnya melakukan sensus pada cluster terpilih, atau memilih lagi sample dari cluster terpilih. (3) Systematic sampling – pada tehnik ini setiap elemen ke n^{th} dalam populasi dipilh, dimulai dengan elemen yang sebelumnya dipilih secara acak. (4) area sampling - Ini sebenarnya sama dengan cluster sampling yang dibatasi hanya pada area tertentu. (5) double sampling – Pertama kita kumpulkan data awal dari contoh tertentu, lalu kemudian kita pilih sub-sample dari contoh tersebut untuk penyelidikan yang lebih mendalam.

Non-probabilistic sampling are (1) **Convenience Sampling** (collection of information from members of the population who are conveniently provide it)

- Survey whoever is easily available,
- Used for quick diagnosis situations,
- Simplest and cheapest,
- Weak representativeness (least reliable),

Non-probabilistic sampling antara lain (1) Convenience sampling (pengumpulan data atau informasi dari anggota populasi yang secara lebih convenient menyediakan informasi tersebut)

- *Kita survey objek yang gampang dihubungi dan tersedia,*
- *Digunakan untuk diagnose secara cepat,*
- *Sangat sederhana dan murah,*
- *Keterwakilan populasi sangat lemah (kurang terandalkan).*

(2) **Purposive Sampling** (confined to specific types of people who can provide the desired information) consist of:

- **Judgment sampling:** experts' opinion could be sought ,e.g., doctor surveyed for cancer causes.
- **Quota sampling:** establish quotas for numbers or proportion of people to be sampled, e.g., survey for research on dual career families: 50% working men and 50% working women surveyed.

For detail information and explanation on these refer to Sekaran (2010).