การคำนวณขนาดตัวอย่างในผลงานทางวิชาการ

รศ.ดร. วีระศักดิ์ ปัญญาพรวิทยา
คณะสัตวแพทยศาสตร์ มหาวิทยาลัยเชียงใหม่
pveerasak.r@gmail.com

Sample size

- เราต้องการจำนวน sample size ที่เหมาะสม คือ ไม่มากไปจนเกินความจำเป็น และไม่น้อยไปจนสรุปอะไรได้ยาก
- Sample size ที่น้อยๆ อาจจะทำให้ผู้อ่านงานวิจัยไม่เชื่อมั่นในงานวิจัยของเรา
- Sample size ที่มากพอ จะส่งให้การทดสอบทางสถิติมีอำนาจในการทดสอบสูง

- ใช้ magic number = 30 ได้ไหม
- ใช้ sample size ที่มากพอ และพบความแตกต่างทางสถิติ
- ใช้ sample size ที่ไม่มาก แต่พบความแตกต่างทางสถิติ แบบนี้เชื่อถือได้ไหม
- ใช้ sample size ที่ไม่มาก ไม่พบความแตกต่างทางสถิติ
- ใช้ sample size ที่ไม่มาก เกือบพบความแตกต่างทางสถิติ

- เปเปอร์ที่ตีพิมพ์ส่วนใหญ่ ไม่ได้แสดงการคำนวณ sample size ทำไมยังได้ตีพิมพ์
- งานระบาดวิทยา บางครั้งกำหนด sample size ล่วงหน้าไม่ได้ เช่น ไม่ทราบว่าโรคลัมปิสกินจะระบาดกี่ฟาร์ม จะเขียนใน proposal อย่างไร
- หาปัจจัยเสี่ยงต่อการปนเปื้อนเซลล์โซมาติกที่สูงกว่ามาตรฐาน (สูงกว่า / ไม่เกินค่ามาตรฐาน) จากถังนมจากฟาร์ม 40 ฟาร์ม จำนวนฟาร์ม 40 ฟาร์ม เพียงพอไหม ในเมื่อทั้งจังหวัดมีฟาร์มโคนม 40 ฟาร์ม

- ส่วนใหญ่ผู้ประเมินเปเปอร์ ไม่ได้คาดหวังว่าจะเห็นการคำนวณ sample size เนื่องจากอนุมานว่า ผู้วิจัยได้คำนวณแล้ว หรือ ผู้ประเมินบางท่านก็คุ้นเคยว่า ในการวิจัยเรื่องนั้นๆ โดยส่วนใหญ่ใช้ sample size ใกล้เคียงกับเปเปอร์ที่เราส่งให้ประเมิน
- ถ้าเกิดกรณีไม่แน่ใจว่า ขนาดตัวอย่างเพียงพอหรือไม่ แต่ได้ทำวิจัยไปแล้ว ส่งผลงานไปแล้ว ผู้ประเมินถามกลับมาว่าแน่ใจหรือว่า sample size เพียงพอ จะมีวิธีการอย่างไรเพื่อหาว่า sample size เพียงพอหรือไม่เพียงพอ

การระบุขนาดตัวอย่าง (sample size)

- ระบุก่อนทำการศึกษา / ทดลอง
- ระบุใน proposal / concept paper
- แสดง สูตร หรือ วิธีการ ว่า sample size นั้นมีการคำนวณอย่างไร
- ต้องทราบว่า งานเราเป็นงานสำรวจ หรือ งานทดลอง หรือ งานระบาดวิทยา
- ระวังการใช้สูตรของ ยามาเน่ แบบผิดๆ

งานวิจัยเชิงสำรวจ

https://businessanalyticsnida.wordpress.com/2016/10/10/on-yamane-misunderstood/

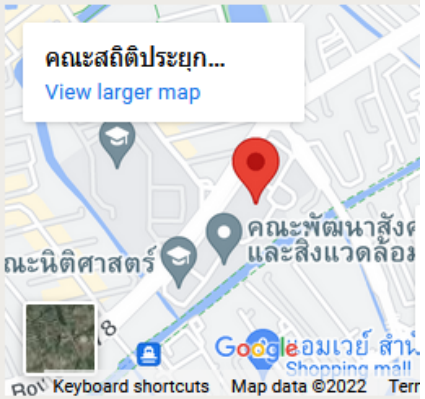
OCTOBER 10, 2016 | SAKWORAWICH | UNCATEGORIZED

ความเข้าใจผิดเรื่องสูตรกำหนดขนาดตัวอย่างของ Yamane ในวงวิชาการไทย

ไม่รู้จะมีสักกี่คนที่จะทราบว่าสูตรในการกำหนดขนาดตัวอย่างของ Taro Yamane ซึ่งมาจากตำรา introductory statistics ของ Yamane เอง ซึ่งไม่ได้ผิดอะไร และ Yamane ก็เขียนอธิบายที่มาที่ไปของสูตรอย่างละเอียดลออ แต่คนไทยกลับไม่อ่าน จำต่อๆ กันมาและเอามาใส่ในงานวิจัย วิทยานิพนธ์ กระทั่งตำราทางสถิติ โดยไม่ได้อ่านว่าสูตรที่ว่านั้นพัฒนามาอย่างไร และควรนำไปใช้เมื่อไหร่

กลายเป็นว่าสูตรกำหนดขนาดตัวอย่าง ของ Yamane เป็น ยาสารพัดนึกในการสุ่มตัวอย่างของไทย

8.00-17.00



คณะสถิติประยุกต์
สถาบันบัณฑิตพัฒนบริหารศาสต
อาคารนวมินทราธิราช ชั้น 12

จะมีสักกี่คนที่รู้ว่า

1. สูตรของ Taro Yamane ใช้สำหรับการวิจัยเชิงสำรวจ (Survey Research) เท่านั้น ดังนั้นพวกที่ทำงานวิจัยเชิงสหสัมพันธ์ (Correlational Research) หรืองานวิจัยเชิงทดลอง (Experimental Research) จึงต้องไม่ใช่สูตรของ Yamane เลย เพราะมาจาก philosophy ที่แตกต่างกัน

2. สูตรของ Taro Yamane ใช้สำหรับงานวิจัยเชิงสำรวจที่เกี่ยวข้องกับค่าสัดส่วนเท่านั้น และสำหรับตัวแปรสุ่มที่เป็น Bernoulli ด้วยเท่านั้น คือต้องตอบแค่ 0,1 เช่น ใช่ ไม่ใช่ หรือ ชอบ กับ ไม่ชอบ หรือ ชาย กับ หญิง ดังนั้นแม้จะเป็นงานวิจัยเชิงสำรวจ แต่ถ้าสนใจจะศึกษาค่าเฉลี่ยหรือความแปรปรวน ก็ไม่น่าจะถูกต้อง แม้จะเป็นตัวแปรสุ่มที่เป็นตัวแปรจัดประเภท แต่ถ้ามีได้หลายคำตอบ ก็จะกลายเป็นการแจกแจงแบบมัลติโนเมียล (Multinomial distribution) หรืออื่นๆ ซึ่งไม่ตรงกับสูตรของ Taro Yamane อยู่ดี

ต้องการทำแบบสอบถามกับเกษตรกรเกี่ยวกับปัจจัยเสี่ยงโรคลัมปิสกิน โดยให้ตอบแบบสอบถาม ซึ่งมีทั้งตอบแบบใช่ -ไม่ใช่ / น้อย - ปานกลาง - มาก สูตรนี้ก็ไม่เหมาะสม

ถ้าจะใช้สูตรนี้เพื่อหาจำนวนคนที่จะทำแบบสอบถาม โดยที่มีคำถามที่ต้องการคำตอบ เป็นแบบ ข้อมูลต่อเนื่อง (จำนวนโคในฟาร์ม) ข้อมูลที่ตอบได้หลายคำตอบ (เห็นด้วยมาก-ปานกลาง-น้อย ให้ข้ามสูตรนี้ไปได้เลย

4. สูตรของ Yamane ใช้กับ Finite population หรือประชากรอันตะ ที่ทราบดีว่ามีประชากรทั้งหมดกี่คน ถ้าไม่ทราบหรือเป็นประชากรอนันต์ ก็ต้องไม่ใช่ Yamane อยู่ดี

5. สูตรของ Taro Yamane ต้องใช้กับงานวิจัยเชิงสำรวจที่มีแผนการเลือกตัวอย่างแบบ Simple random sampling เท่านั้น ถ้าใช้แผนการสุ่มแบบอื่นก็ใช้ไม่ได้

สำรวจความเห็นต่อเกณฑ์ใหม่ ว่าชอบหรือไม่ชอบ ในหลายๆ ตลาด หรือ หลายๆ อำเภอ ที่ไม่ใช่ สุ่มอย่างง่ายก็ใช้สูตรนี้ไม่ได้

สูตรยามาเน่มีตัวแปรที่เราต้องรู้ค่าสองตัวคือ N และ e โดยที่ N คือจำนวนประชากร และ e คือค่าความคลาดเคลื่อน (i.e. error) เวลาแทนค่า e ในสมการต้องเขียนเป็นทศนิยม เช่น e = .05 (error = 5% เป็นค่ามาตรฐานที่เราใช้ในสูตรยามาเน่ สามารถเปลี่ยนได้ตามความเหมาะสม)

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

อาจารย์ยามาเน่ assume ความมั่นใจร้อยละ 95 และความแปรปรวนสูงที่สุด (maximum variance) ที่ค่า p = .5 ทำให้สูตรยามาเน่ใช้งานง่ายขึ้นและมีตัวแปรที่เราต้องแทนค่าแค่สองตัวคือ N และ e ถ้ากลุ่มประชากรที่เราสนใจ N = 5000 และ e = .05 แทนค่าในสมการด้านบนเราต้องเก็บ sample size = 5000/ (1 + 5000*.05^2) = 370 คน ตัวเลขนี้คือ n ขั้นต่ำที่เราต้องการในงานวิจัยนั้นๆ

https://datarockie.com/blog/yamane-sample-size-calculation/

เราใช้สูตรยามาเน่ในกรณีไหน? สูตรนี้จะใช้ได้ถ้าเงื่อนไขต่อไปนี้เป็นจริง

- เรารู้จำนวนประชากรทั้งหมด (finite population)
- ค่าของประชากรที่เราต้องการวัดเป็นค่าร้อยละคิดจากค่าถามแบบ dichotomous (% proportion) ถ้าแบบสอบถามเป็นแบบ likert scale (e.g. 1-5) เราก็จะไม่สามารถใช้สูตรยามาเน่ได้
- การสุ่มตัวอย่างเป็นไปตามทฤษฎีความน่าจะเป็น (random sampling)

Sample Size Table

ถ้าเราใช้ e = .05 ในสูตรยามาเน่และขนาดประชากรใหญ่ขึ้นเรื่อยๆ sample size ที่เรา
คำนวณได้จะ converge สู่ตัวเลข 400 แปลว่าจริงๆเราไม่ต้องเสียเวลาคำนวณ ใช้ n=400
ไปเลย จบๆ 555+ (ลองดูตารางสรุปด้านล่าง)

population	sample size (e=0.05)
1,000	286
10,000	385
100,000	398
1,000,000	400

จากสูตรและตารางนี้ก็นำไปสู่ความเข้าใจผิดๆ ว่า
ไม่รู้อะไรก็ทำแบบสอบถาม 400 คนไปเลย
หรือ
สรุปง่ายๆ ว่า ประชากรมีไม่เกิน 1000 คน
ดังนั้นทำแบบสอบถาม 286 ชุดก็พอ

แล้วโลกนี้มีสูตรที่ดีกว่าของยามาเน่มั้ย? มี! มาดู master formula นี้กัน

เว็บไซต์สร้างแบบสอบถาม/ market research survey ระดับโลกอย่าง Qualtrics และ SurveyMonkey ใช้สูตรการคำนวณ sample size ที่ละเอียดกว่าของยามาเน่ และสามารถประยุกต์ใช้ได้ทั้งแบบรู้หรือไม่รู้จำนวน population ก็ได้ (อ้างอิงจาก Cochran, 1963) กรณีที่เราไม่รู้ N ให้ใช้สูตรนี้

$$n_0 = \frac{z^2 p(1-p)}{e^2}$$

โดยที่ p * (1-p) เรียกว่า variance ถ้าเรากำหนดค่า p = .5; (1-p) = .5 เราจะได้ maximum variance = .5 * .5 = .25 (i.e. ค่า p ที่นิยมใช้ในสูตรนี้คือ .5) ค่า e เป็นตัวเดียวกันกับสูตรยามาเน่ ส่วน z ได้มาจาก Standard Normal Distribution ในตารางด้านล่าง โปรแกรมสถิติส่วนใหญ่กำหนดค่า confidence = .95 (default) แปลว่าตัวเลข z = 1.96 จะถูกหยิบมาใช้บ่อยมาก เป็นตัวเลขที่นักสถิติต้องจำได้ (แอดจำหมดเลยตั้งแต่ 1.65 1.96 2.58)

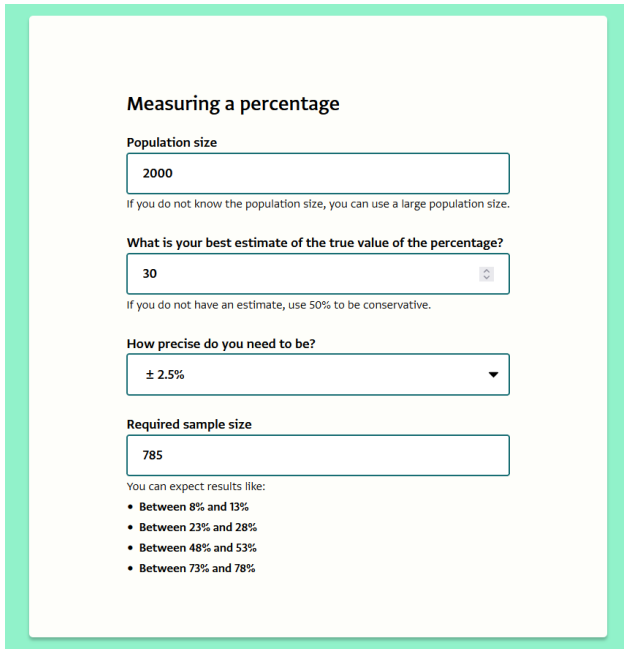
ต้องทราบว่า งานของเรานั้นเป็นงานสำรวจที่สอดคล้องกับการคำนวณนี้จริงๆ หรือ ?

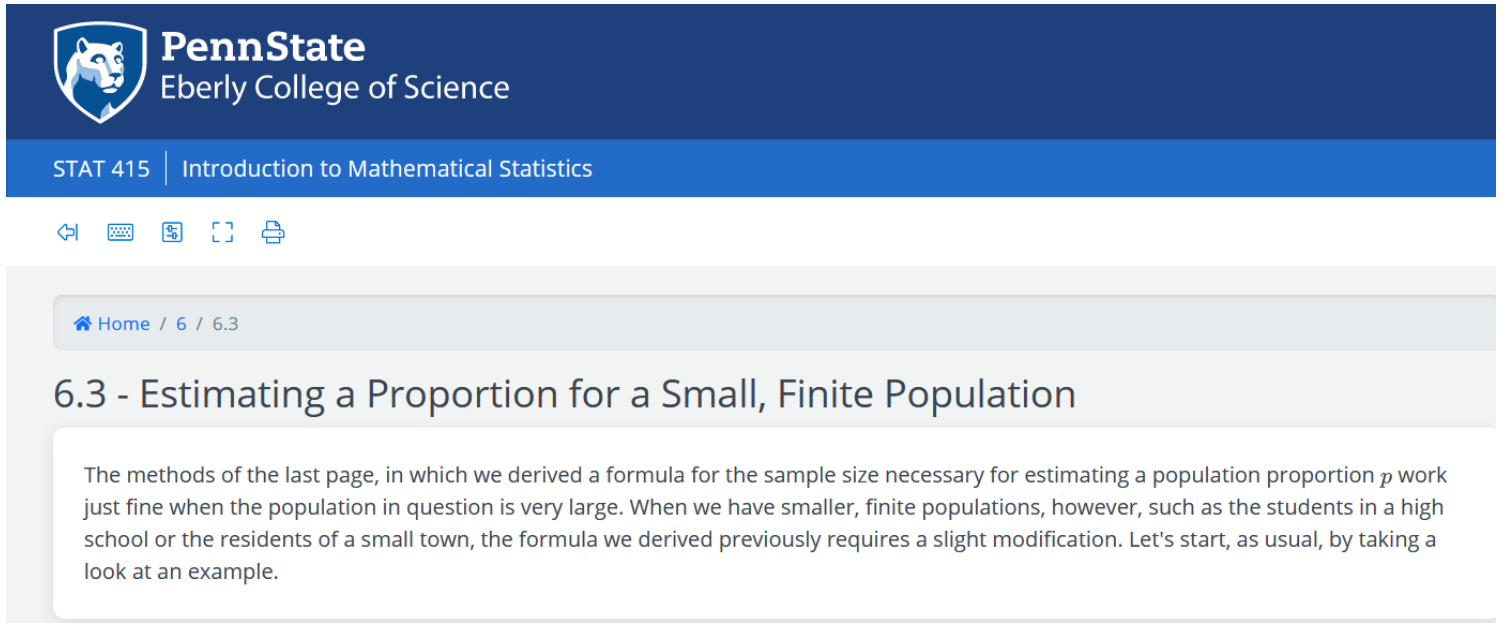
https://www.activityinfo.org/support/docs/sampling/snapshot.html

Support > Documentation > Sampling calculators

Sample size calculator for needs assessments and KAP surveys

The following calculator will help you find an appropriate sample size for a needs assessment survey, a KAP survey, or any other kind of survey intended to provide a "snapshot" of a population at a given point in time. The results of the survey can be used to draw conclusions about the population as a whole, and can provide useful inputs on the design and planning of interventions. To measure change over time in the same population, see our Baseline and endline sample size calculator.





Example 6-3 (continued)

A researcher is studying the population of a small town in India of $N = 2000$ people. She's interested in estimating p for several yes/no questions on a survey.

How many people n does she have to randomly sample (without replacement) to ensure that her estimates $\hat{p}$ are within $\epsilon = 0.04$ of the true proportion p?

Answer

Okay, once and for all, let's calculate this very patient researcher's sample size! Because the researcher has many different questions on the survey, it would behoove her to use a sample proportion of 0.50 in her calculations. If the maximum error ϵ is 0.04, the sample proportion is 0.5, and the researcher doesn't make the finite population correction, then she needs:

$$m = \frac{(1.96^2)(\frac{1}{4})}{0.04^2} = 600.25$$

or 601 people to estimate p with 95% confidence. But, upon making the correction for the small, finite population, we see that the researcher really only needs:

$$n = \frac{m}{1 + \frac{m-1}{N}} = \frac{601}{1 + \frac{601-1}{2000}} = 462.3$$

or 463 people to estimate p with 95% confidence.

Effect of Population Size N

The following table illustrates how the sample size n that is necessary for estimating a population proportion p (with 95% confidence) is affected by the size of the population N. If $\hat{p} = 0.5$, then the sample size n is:

$\hat{p} = 0.5$	ϵ= 0.01	ϵ= 0.03	ϵ= 0.05
N very large	9604	1068	385
N = 10, 000, 000	9595	1068	385
N = 1, 000, 000	9513	1067	385
N = 100, 000	8763	1057	384
N = 10, 000	4900	966	371
N = 1, 000	906	517	279

งานวิจัยที่เป็นการทดลองและงานระบาดวิทยา

งานวิจัยที่มีการวางแผนการทดลอง / มีการ intervention

การสรุปว่าผลการทดลองมีความแตกต่างทางสถิตินั้น
ต้องประกอบด้วย

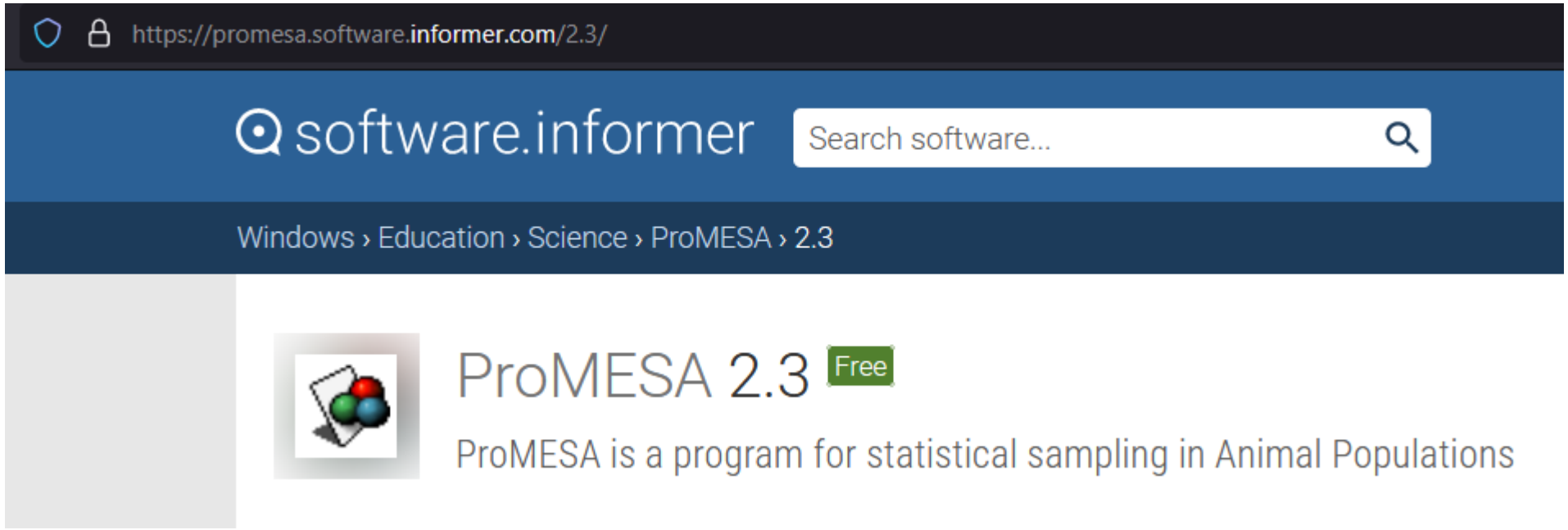
1) มีความแตกต่าง
ระหว่างทรีตเม้นต์เกิดขึ้นจริง

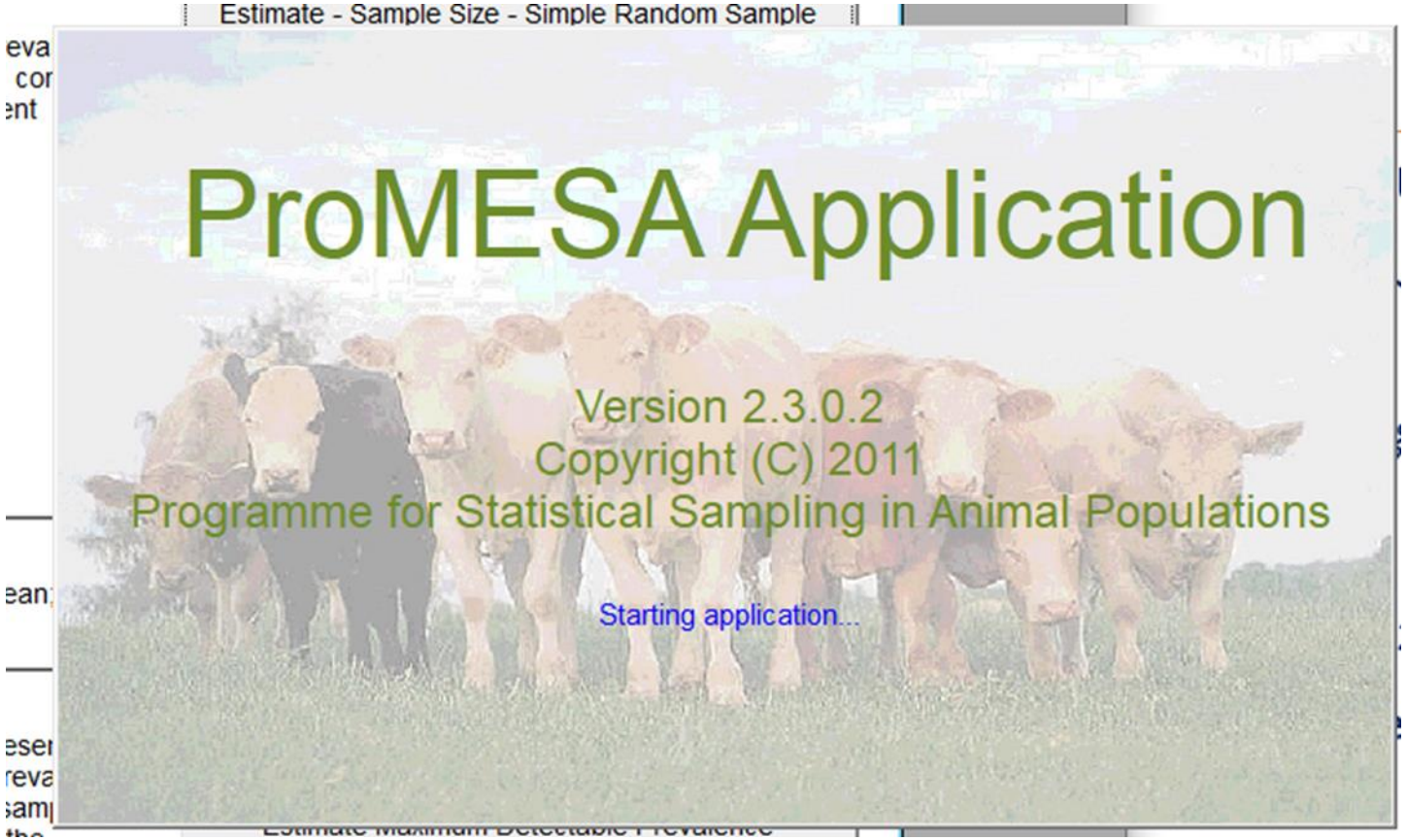
2) มีอำนาจ (power) ที่เพียงพอในการตรวจพบความแตกต่างนั้น

แนะนำโปรแกรมที่ใช้ในการคำนวณ sample size

- งานหาความชุก เช่น ความชุกของโรคเต้านมอักเสบจากเชื้อ S.aureus ในโคนม → PROMESA
- งานหาปัจจัยเสี่ยง เช่น case-control study ตัวอย่างเช่น ทำแบบสอบถามหาปัจจัยเสี่ยงต่อการปนเปื้อน E.coli ในเนื้อสัตว์ที่โรงฆ่า (ตัวแปรตาม คือ ปนเปื้นเกินมาตรฐาน / ปนเปื้อนไม่เกินมาตรฐาน → Gpower
- งานเปรียบเทียบสัดส่วน (proportion test) เช่น เปอร์เซ็นต์การปนเปื้อนของเชื้อ Salmonella ในโรงฆ่าสัตว์ขนาดเล็กเมื่อเทียบกับขนาดกลางว่ามีความแตกต่างกัน หรือไม่ → Gpower
- เปรียบเทียบระดับไตเตอร์จากสัตว์ที่ได้รับวัคซีนลัมปิสกิน ชนิด A กับ ชนิด B ว่ามีความแตกต่างกันหรือไม่ → Gpower

https://promesa.software.informer.com/2.3/

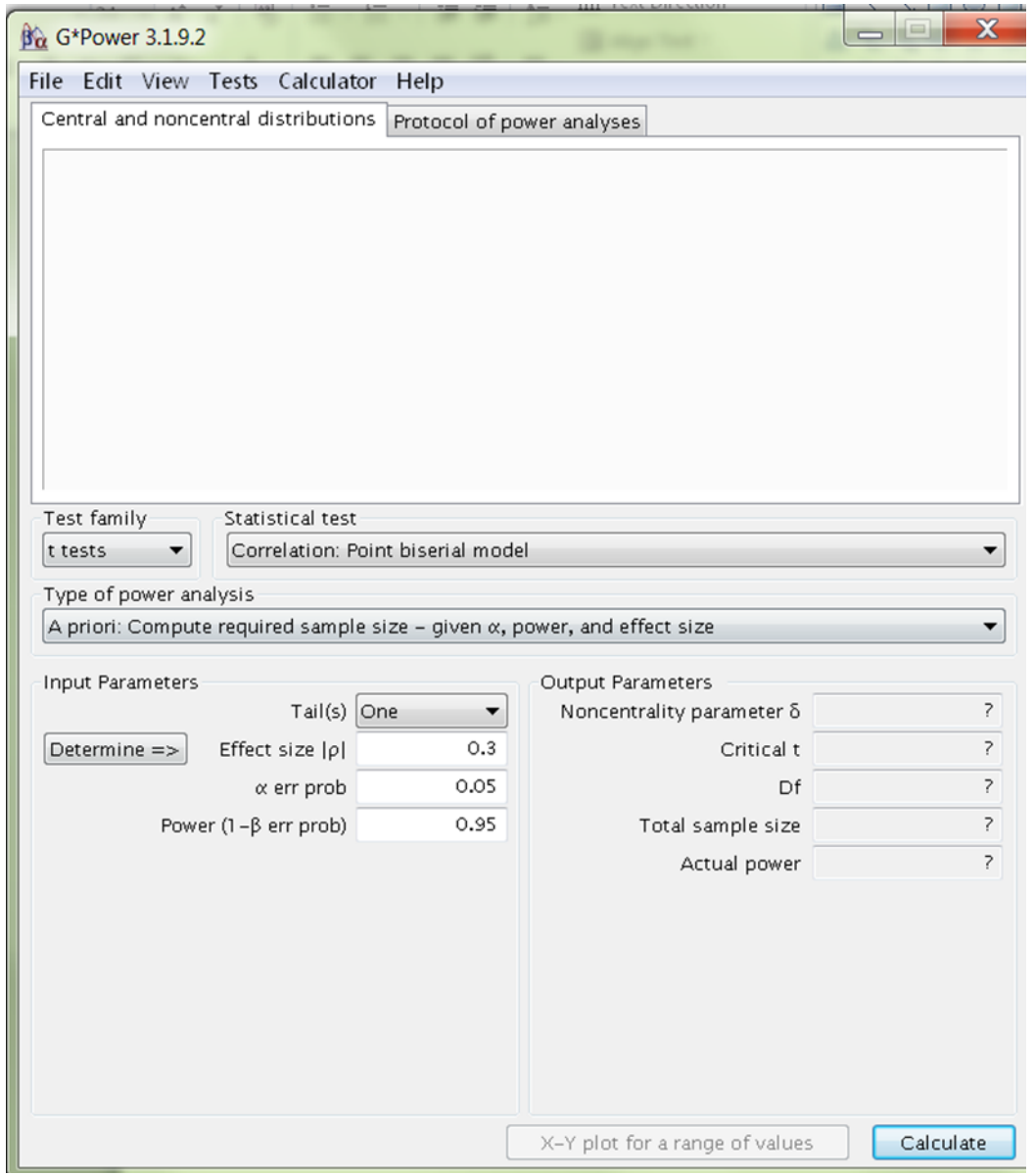



Estimate - Sample Size - Simple Random Sample
ProMESA Application
Version 2.3.0.2
Copyright (C) 2011
Programme for Statistical Sampling in Animal Populations
Starting application...

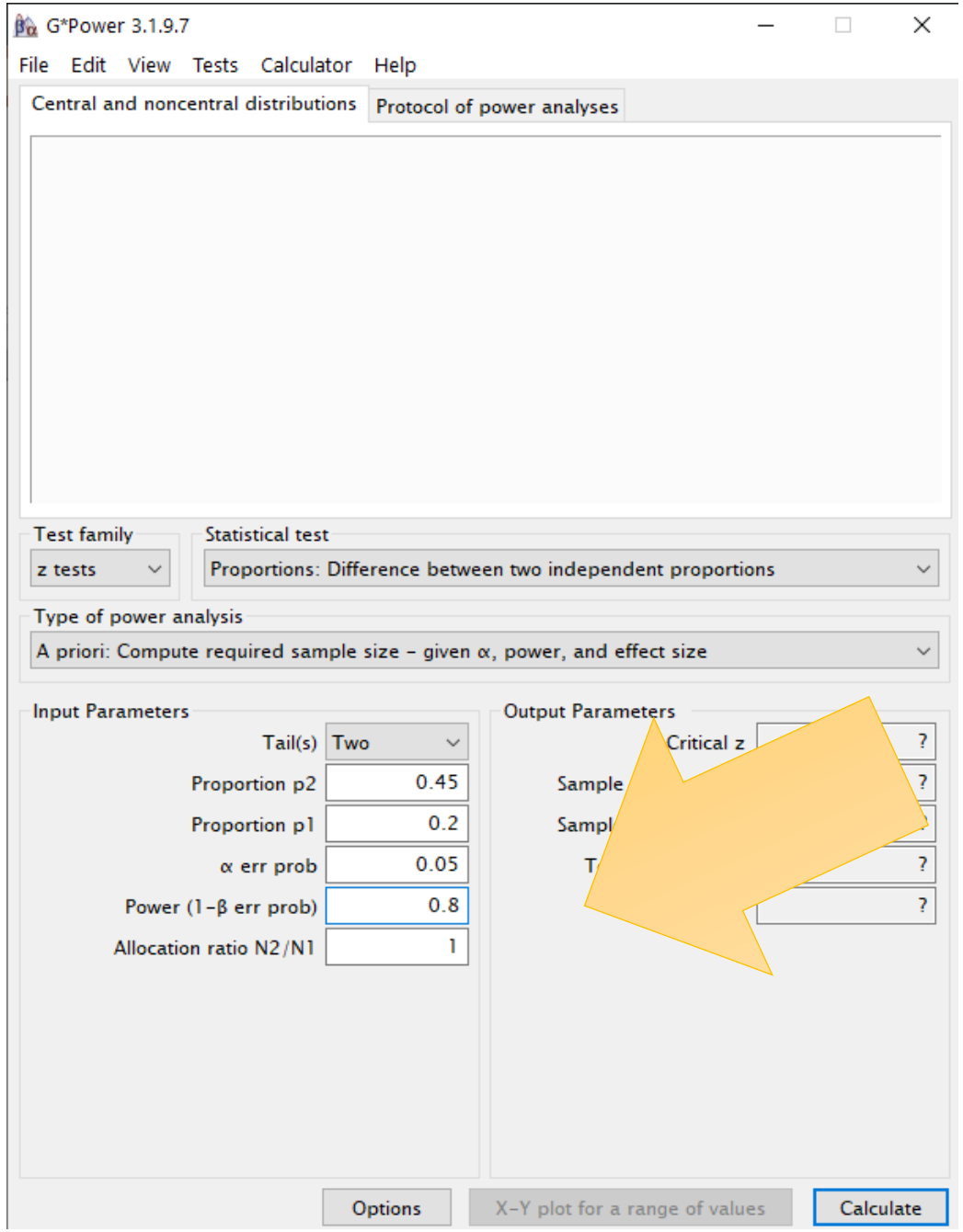
การคำนวณขนาดตัวอย่างโดยใช้โปรแกรม G*power สำหรับงานวิจัยที่มีการใช้สัตว์

รศ. น.สพ. ดร. วิระศักดิ์ ปัญญาพรวิทยา
คณะสัตวแพทยศาสตร์ มหาวิทยาลัยเชียงใหม่
pveerasak.r@gmail.com

free download:



- ต้องการหาว่า เปอร์เซ็นต์การปนเปื้อนของเชื้อ E.coli ที่เกินระดับมาตรฐาน จากตัวอย่างเนื้อไก่ที่เก็บจากตลาดสด กับ ตลาดรับรองมาตรฐานว่าแตกต่างกันหรือไม่ คำนวณหา sample size ได้อย่างไร
- ต้องกำหนดว่า เปอร์เซ็นต์ควรจะแตกต่างกันกี่เปอร์เซ็นต์ ถ้าบอกว่าไม่ทราบได้ไหม ถ้าทราบจะไปทำการศึกษาทำไม ?
- ต้องกำหนด เพราะจะเป็นการทำให้มั่นใจว่า sample size นั้น เพียงพอต่อการวิเคราะห์ทางสถิติ หรือ มี statistical power of test ที่เพียงพอ
- ดังนั้นกำหนดว่าจะแตกต่างกันที่ 25 %



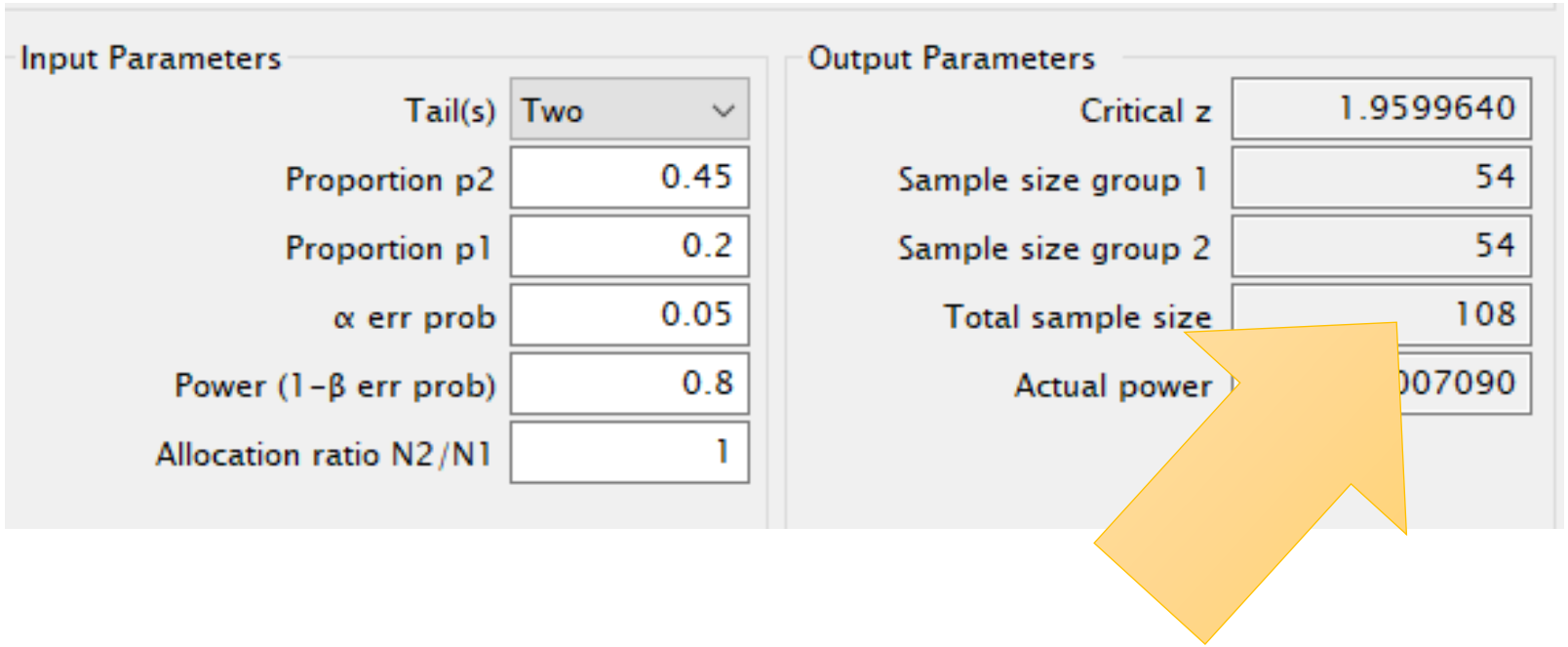
โรงฆ่าขนาดกลางคาดว่าพบการปนเปื้อน 20%

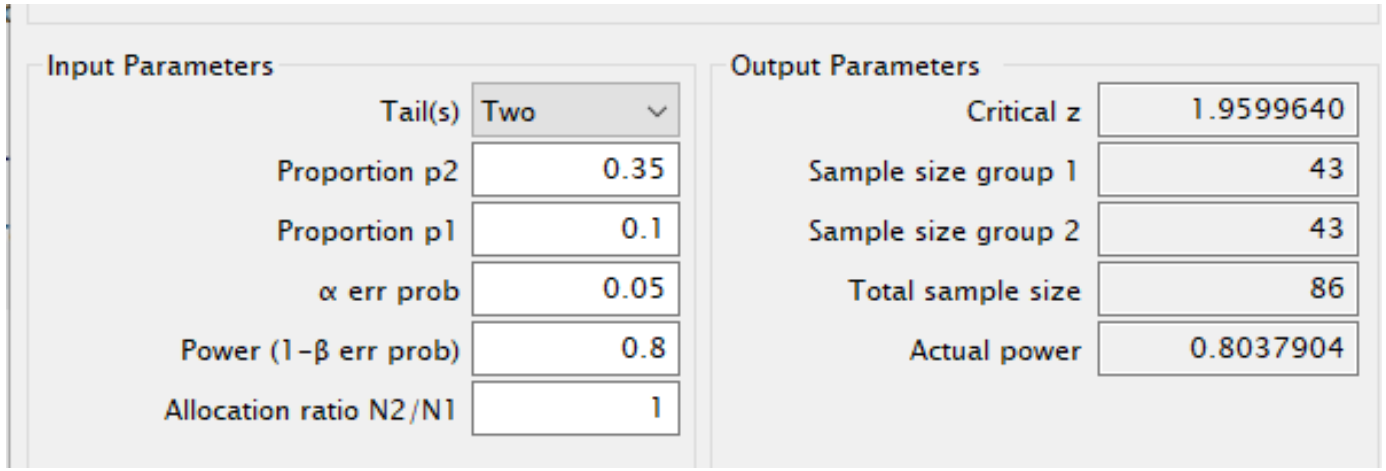
คาดว่าโรงฆ่าขนาดเล็กจะพบการปนเปื้อนสูงกว่าโรงฆ่าขนาดกลาง 25%

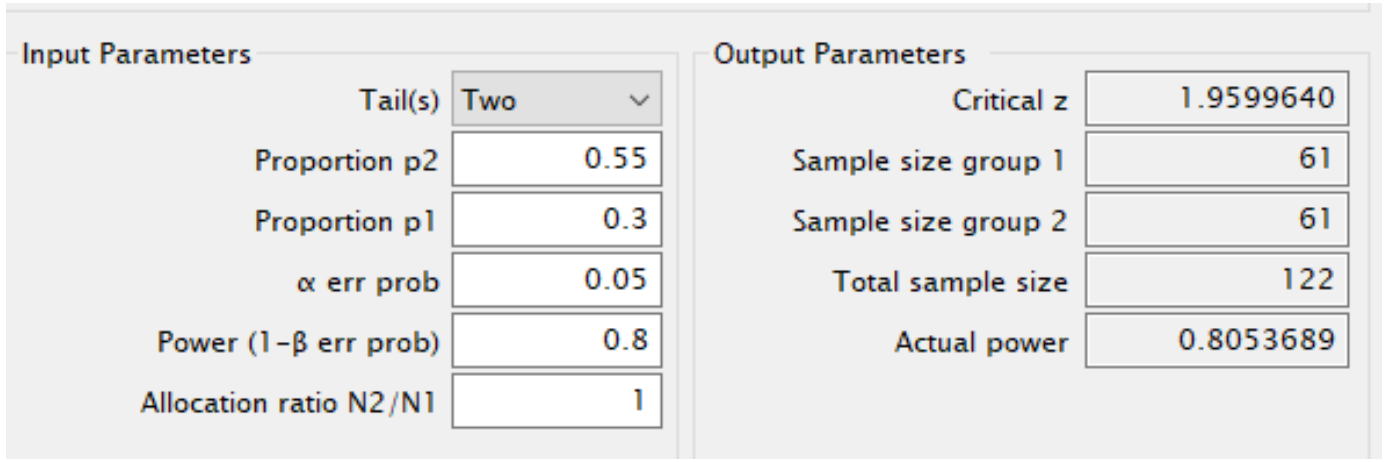
ดังนั้นปนเปื้อน = 20+25 = 45%

ใส่เลข 0.45 และ 0.2

ใส่ค่าแอลฟ่าเป็น 0.05 และค่า Power เป็น 0.8

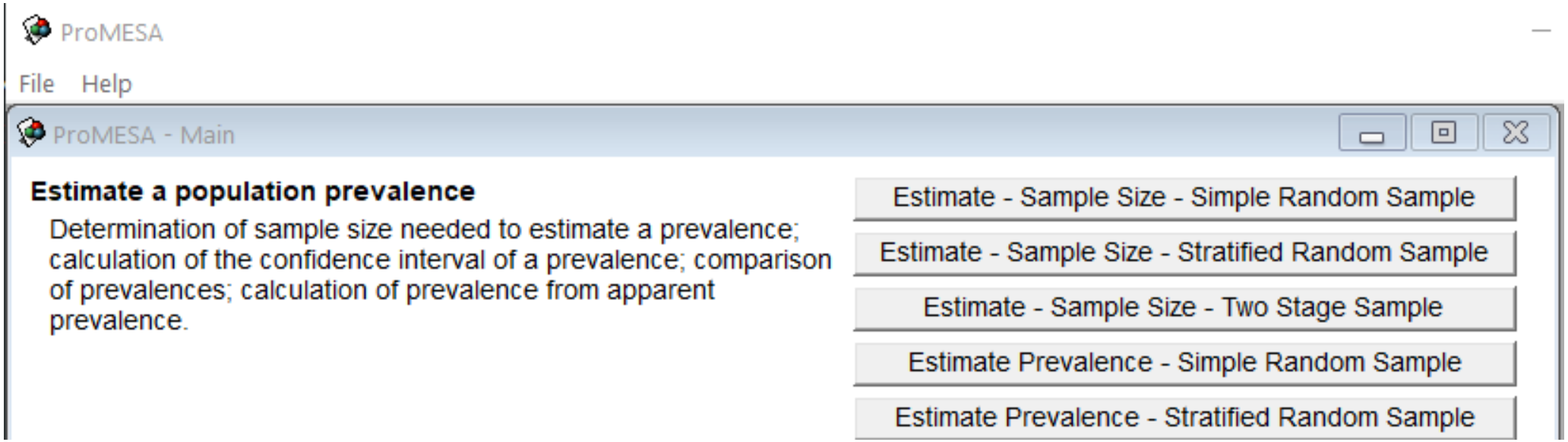

Input Parameters
Tail(s)
Two
Proportion p2
0.45
Proportion p1
0.2
α err prob
0.05
Power (1−β err prob)
0.8
Allocation ratio N2/N1
1
Output Parameters
Critical z
1.9599640
Sample size group 1
54
Sample size group 2
54
Total sample size
108
Actual power
007090



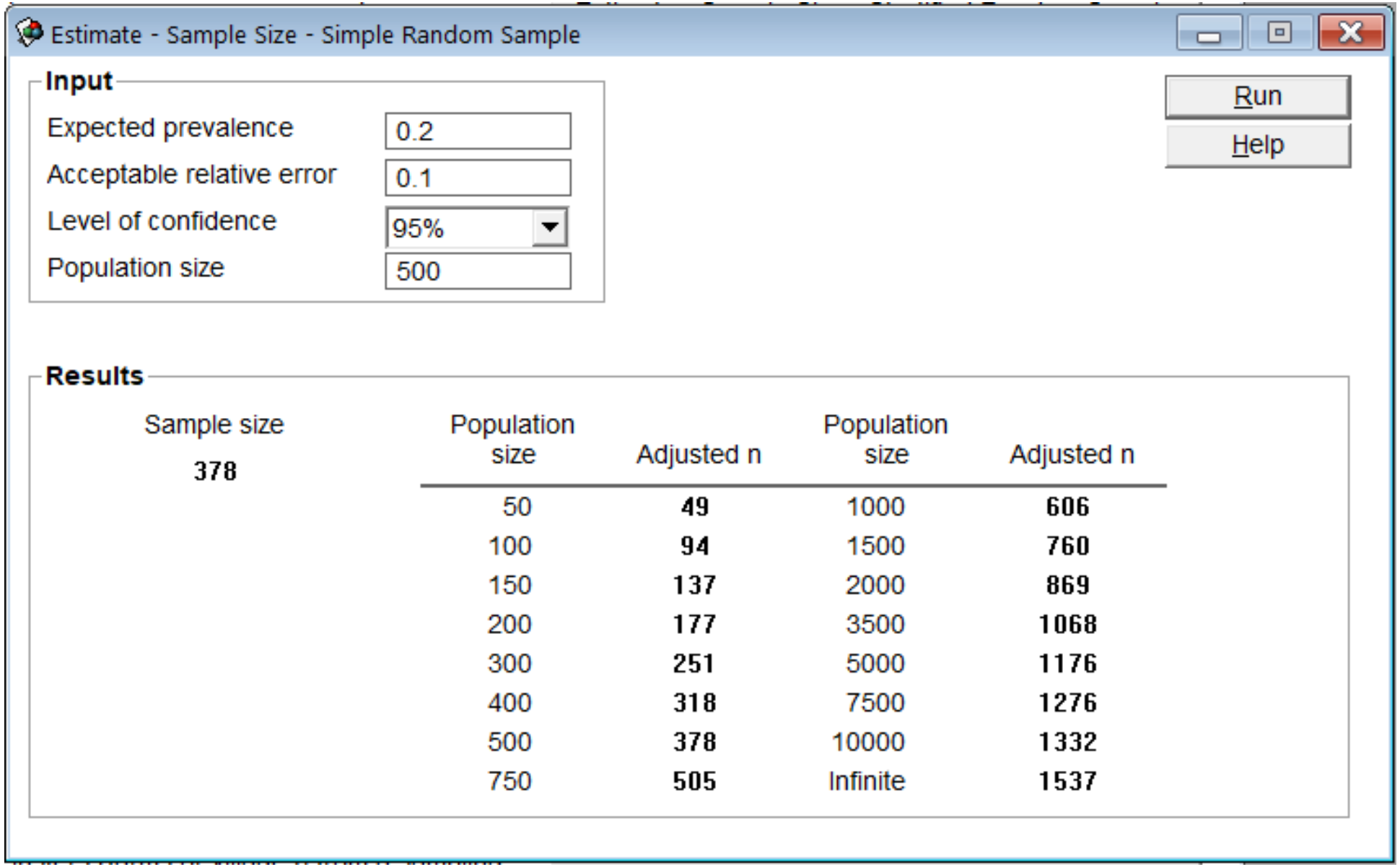


ไม่ใช่ใส่เลขอะไรก็ได้ ขอให้ต่างกัน 25% เพราะ sample size ที่ได้จะแตกต่างกันตาม % positive ที่เราตั้งไว้

- ต้องการหาความชุกของเชื้อแบคทีเรียที่ดื้อยา จากเนื้อหมูที่ขายในร้านขายหมู ควรเก็บตัวอย่างเนื้อหมูจำนวนกี่ตัวอย่าง



ใช้โปรแกรม PROMESA ในการคำนวณ เลือก ประมาณค่า prevalence และในกรณีที่การศึกษาเป็นแบบสุ่มอย่างง่าย ให้เลือก simple random sample


Estimate - Sample Size - Simple Random Sample
Input
Expected prevalence
0.2
Acceptable relative error
0.1
Level of confidence
95%
Population size
500
Run
Help
Results
Sample size
378
Population size
Adjusted n
Population size
Adjusted n
50 49 1000 606
100 94 1500 760
150 137 2000 869
200 177 3500 1068
300 251 5000 1176
400 318 7500 1276
500 378 10000 1332
750 505 Infinite 1537

ple size needed to estimate a prevalence,

Estimate - Sample Size - Simple Random Sample

Input

Expected prevalence	0.6
Acceptable relative error	0.1
Level of confidence	95%
Population size	500

Run

Help

ถ้าความชุกมีเปอร์เซ็นต์สูง
จำนวนตัวอย่างที่เก็บก็จะน้อยลง

Results

Sample size

170

Population size	Adjusted n	Population size	Adjusted n
50	**42**	1000	**204**
100	**72**	1500	**219**
150	**95**	2000	**228**
200	**113**	3500	**239**
300	**139**	5000	**244**
400	**157**	7500	**248**
500	**170**	10000	**250**
750	**191**	Infinite	**257**

Estimate - Sample Size - Simple Random Sample

Input

Expected prevalence	0.2
Acceptable relative error	0.2
Level of confidence	95%
Population size	500

การเพิ่ม acceptable

Relative error จะส่งผลให้

Sample size มีค่าน้อยลง

Results

Sample size

218

Population size	Adjusted n	Population size	Adjusted n
50	**45**	1000	**278**
100	**80**	1500	**306**
150	**108**	2000	**323**
200	**132**	3500	**347**
300	**169**	5000	**357**
400	**196**	7500	**366**
500	**218**	10000	**370**
750	**255**	Infinite	**385**

- ตอบก่อน !!! เราต้องการหาความชุกในระดับใด ระดับฟาร์ม (มีสัตว์อย่างน้อย 1 ตัว ถือว่าเป็น positive ฟาร์ม) หรือ ระดับตัวสัตว์ (มีสัตว์ 10000 ตัว มีกี่ตัวที่ positive) เพราะจำนวน population size ที่ต้องกรอก มีความแตกต่างกัน
- หาความชุกระดับตัวสัตว์ มีสัตว์เป็นหมื่นตัวในพื้นที่ ดังนั้นจำนวนตัวอย่างที่เก็บมักจะต้องมาก (เช่น เก็บตัวอย่างตรวจหาไตเตอร์ต่อโรคแท้งติดต่อ)
- ถ้าหาระดับฟาร์ม จะต้องใช้กี่ฟาร์ม และต้องใช้สัตว์จำนวนกี่ตัว เพราะการเก็บตัวอย่างไม่ใช่ simple random sampling แต่เป็น multistage sampling คือ ต้องหาจำนวนฟาร์มก่อน แล้วหาจำนวนสัตว์ที่เก็บตัวอย่าง

- ตอบก่อน !!! เราต้องการหาความชุกในระดับใด ระดับร้านขาย (มีตัวเนื้ออย่างน้อย 1 ตัวอย่าง ถือว่าเป็นร้านที่มีการขายเนื้อที่ปนเปื้อนเชื้อ) หรือ ระดับเนื้อสัตว์ (มีเนื้อสัตว์จากสัตว์ 10000 ตัว มีกี่ตัวที่ positive) เพราะจำนวน population size ที่ต้องกรอก มีความแตกต่างกัน
- หาความชุกระดับเนื้อสัตว์ มีเนื้อสัตว์เป็นหมื่นตัว ดังนั้นจำนวนตัวอย่างที่เก็บมักจะต้องมาก
- ถ้าหาระดับร้านขาย จะต้องใช้กี่ร้าน และต้องใช้เนื้อสัตว์จำนวนกี่ตัวอย่าง เพราะการเก็บตัวอย่างไม่ใช่ simple random sampling แต่เป็น multistage sampling คือ ต้องหาจำนวนร้านขายก่อน แล้วหาจำนวนเนื้อสัตว์ที่ต้องเก็บตัวอย่าง

การคำนวณขนาดตัวอย่างสำหรับการวิเคราะห์ด้วย logistic regression และ multiple logistic regression

การวิเคราะห์ logistic regression เป็นการหาความสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรต้น โดยที่ตัวแปรตามเป็นข้อมูลแบบ binary outcome คือ เป็นไปได้ 2 แบบ เช่น เป็นโรค/ไม่เป็นโรค, ผลบวก/ผลลบ, ใช่/ไม่ใช่

ตัวอย่างเช่น การเกิดภาวะขาเจ็บในโค (ขาเจ็บ/ขาไม่เจ็บ) ในโคที่เลี้ยงบนพื้นคอนกรีตและโคที่เลี้ยงบนพื้นดิน ทั้งนี้ผู้วิจัยตั้งสมมุติฐานว่าการเลี้ยงบนพื้นคอนกรีตจะพบปัญหาโคเจ็บขาได้มากกว่า ในการบอกถึงโอกาสหรือความเสี่ยงในการเจ็บ ค่าที่นิยมใช้ คือ ค่า odd ratio ซึ่งเป็นค่าที่บอกว่าเมื่อได้รับปัจจัยเสี่ยงแล้วโอกาสเกิดโรคนั้นเป็นกี่เท่าเมื่อเทียบกับการไม่ได้รับปัจจัยเสี่ยง

<u>ตัวอย่างสำหรับ logistic regression</u>

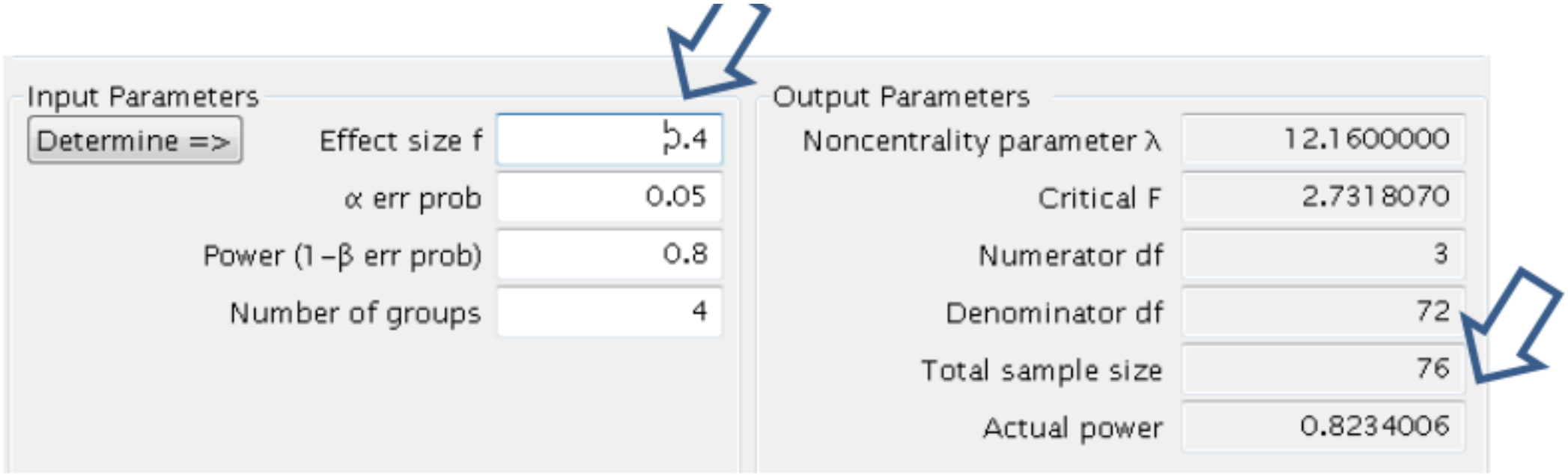
ผู้วิจัยต้องการหาว่า การเลี้ยงโคบนพื้นคอนกรีตจะทำให้เกิดปัญหาขาเจ็บมากน้อยเพียงใดเมื่อเทียบกับการเลี้ยงโคบนพื้นดิน ทั้งนี้อัตราการเกิดโรคขาเจ็บในฟาร์มที่เลี้ยงโคบนพื้นดินเป็น 10% และผู้วิจัยทราบจากรายงานก่อนหน้านี้ หรือ ทราบจากแหล่งข้อมูลว่า หากเลี้ยงโคบนพื้นคอนกรีตน่าจะพบปัญหาขาเจ็บที่ 25% ผู้วิจัยต้องการทำการทดลองเพื่อศึกษาและต้องการทราบว่าต้องใช้โคในการทดลองทั้งหมดกี่ตัว

การคำนวณขนาดตัวอย่างสำหรับการวิเคราะห์ด้วย One-Way ANOVA

การวิเคราะห์ One-way ANOVA จะเป็นการวิเคราะห์สำหรับแผนการทดลองแบบสุ่มสมบูรณ์ ซึ่งผู้วิจัยหน่วยทดลองหรือสัตว์ทดลองที่มีความคล้ายคลึงกัน จากนั้นทำการสุ่มทรีตเมนต์หรือปัจจัยให้กับหน่วยทดลองนั้นๆ

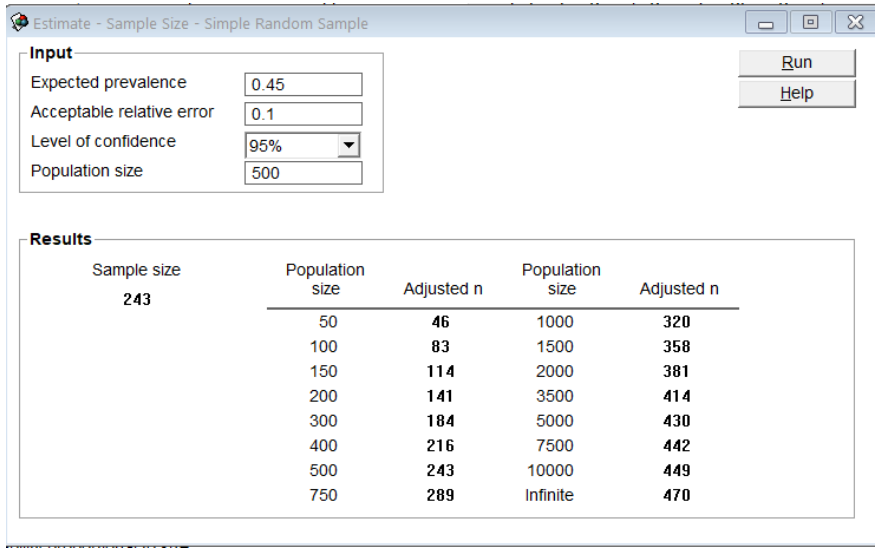
ตัวอย่าง

ผู้วิจัยต้องการทดสอบประสิทธิภาพของยาและสมุนไพร 4 ชนิดในการลดน้ำตาลในกระแสเลือดในหนูทดลอง โดยหนูทดลองมีระดับน้ำตาลในเลือดเริ่มต้นที่ใกล้เคียงกัน ทั้งนี้ผู้วิจัยมีทรีตเมนต์จำนวน 4 ทรีตเมนต์ ประกอบด้วย ทรีตเมนต์ A , B , C และ D ผู้วิจัยต้องการเปรียบเทียบค่าเฉลี่ยระดับน้ำตาลในเลือดของหนูทดลองของแต่ละกลุ่มว่ามีความแตกต่างกันหรือไม่ โดยการวิเคราะห์ One-way ANOVA


Input Parameters
Determine =>
Effect size f
0.4
α err prob
0.05
Power (1 –β err prob)
0.8
Number of groups
4
Output Parameters
Noncentrality parameter λ
12.1600000
Critical F
2.7318070
Numerator df
3
Denominator df
72
Total sample size
76
Actual power
0.8234006

ตัวอย่างการเขียนและนำเสนอการคำนวณหา sample size

- การศึกษานี้เป็นการหาความชุกของการพบการปนเปื้อน โดยวางแผนการเก็บตัวอย่างเป็นแบบสุ่มอย่างง่าย (simple random sampling) ดังนั้นจึงคำนวณหาขนาดตัวอย่างโดยใช้โปรแกรม PROMESA โดยกำหนดค่า Expected prevalence จากข้อมูลที่ได้มีการศึกษามาก่อนหน้านี้ ซึ่งเท่ากับ 45% และกำหนดให้ acceptable relative error เป็น 10% ที่ level of confidence 95% และ population size = 500 ผลการคำนวณเป็นดังนี้



สรุปว่าต้องเก็บตัวอย่างจำนวน 243 ตัวอย่าง

- ต้องการเก็บตัวอย่างภายในเขต 5 เพื่อดูการปนเปื้อนของเชื้อ มีโรงฆ่าจำนวน 3000 โรงฆ่า ต้องเก็บตัวอย่างจากกี่โรงฆ่า และ กี่ตัวอย่างต่อโรงฆ่า