Sampling Generalities

Dr. Mansour Rezaei

Prof. of Biostatistics

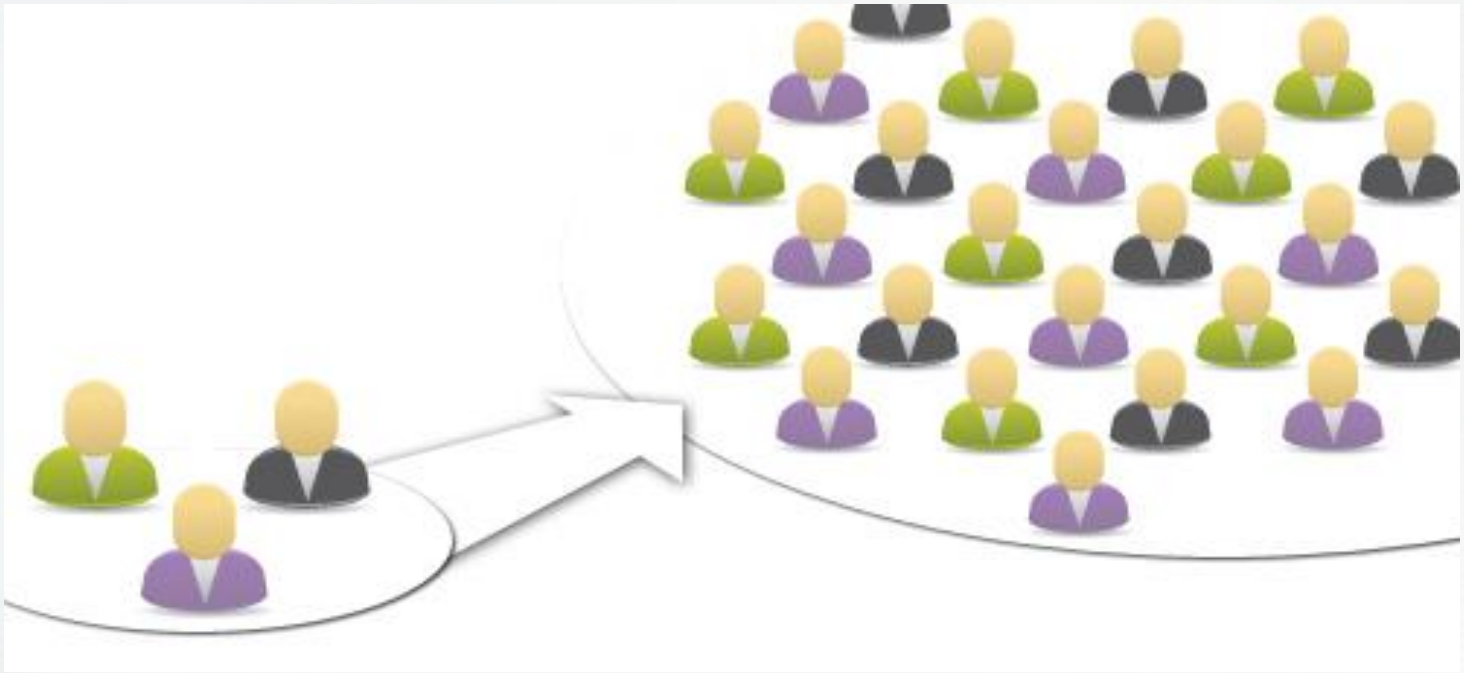
Kermanshah University of Medical Sciences

(KUMS)

Introduction

- **You want to determine a parameter.**
- **Should you take the measurement from every individual in the population?**
- **Is it feasible to approach each individual in the area to get the information?**

More practical to do sampling



Results from the sample will represent the data of the population of interest.

“ How do I determine the sample size that I need???”

Why must life be so complicated?

- **“Why can’t I just take 6 of my best friends to be the sample?”**

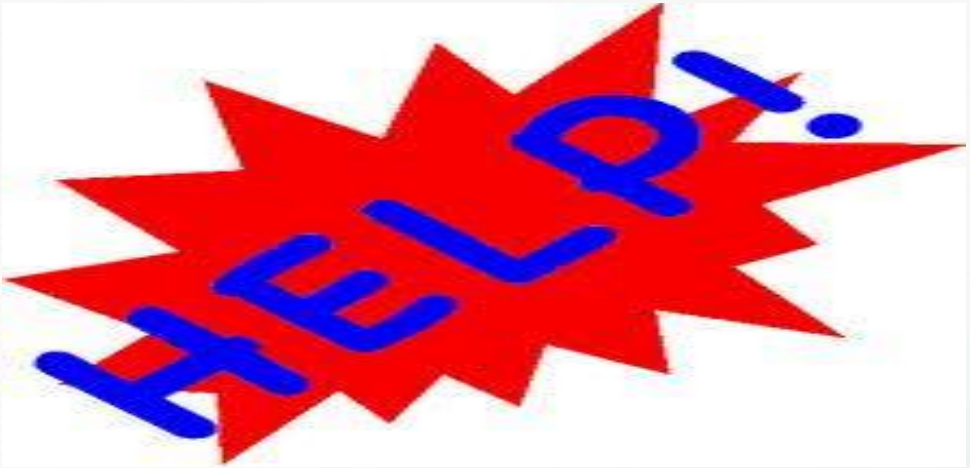
1- easier,

2- cheaper,

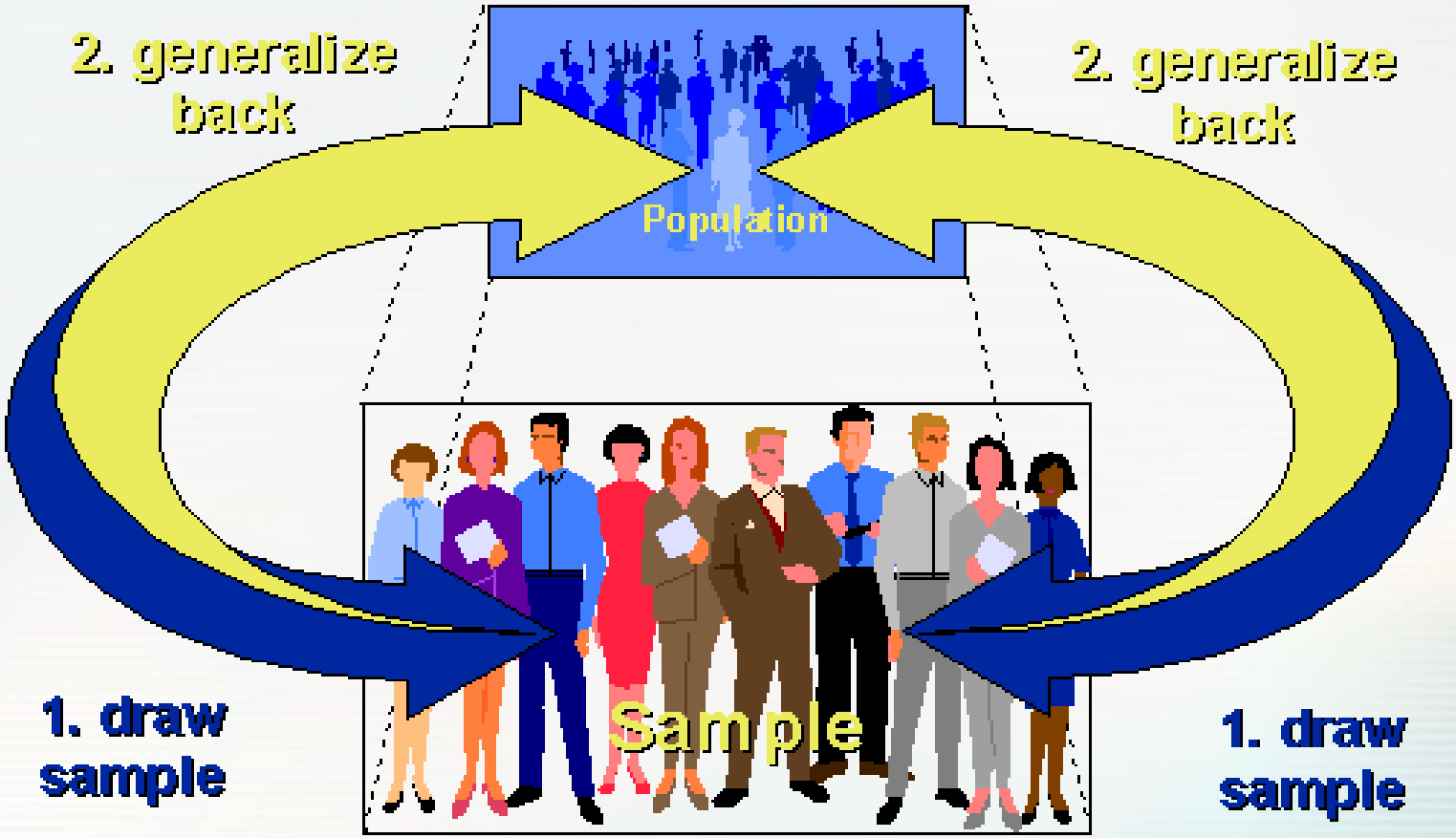
3- good bonding time,

4-more co-operative ...

- **“Other studies using very few subjects have been published, why can’t I do the same?”**



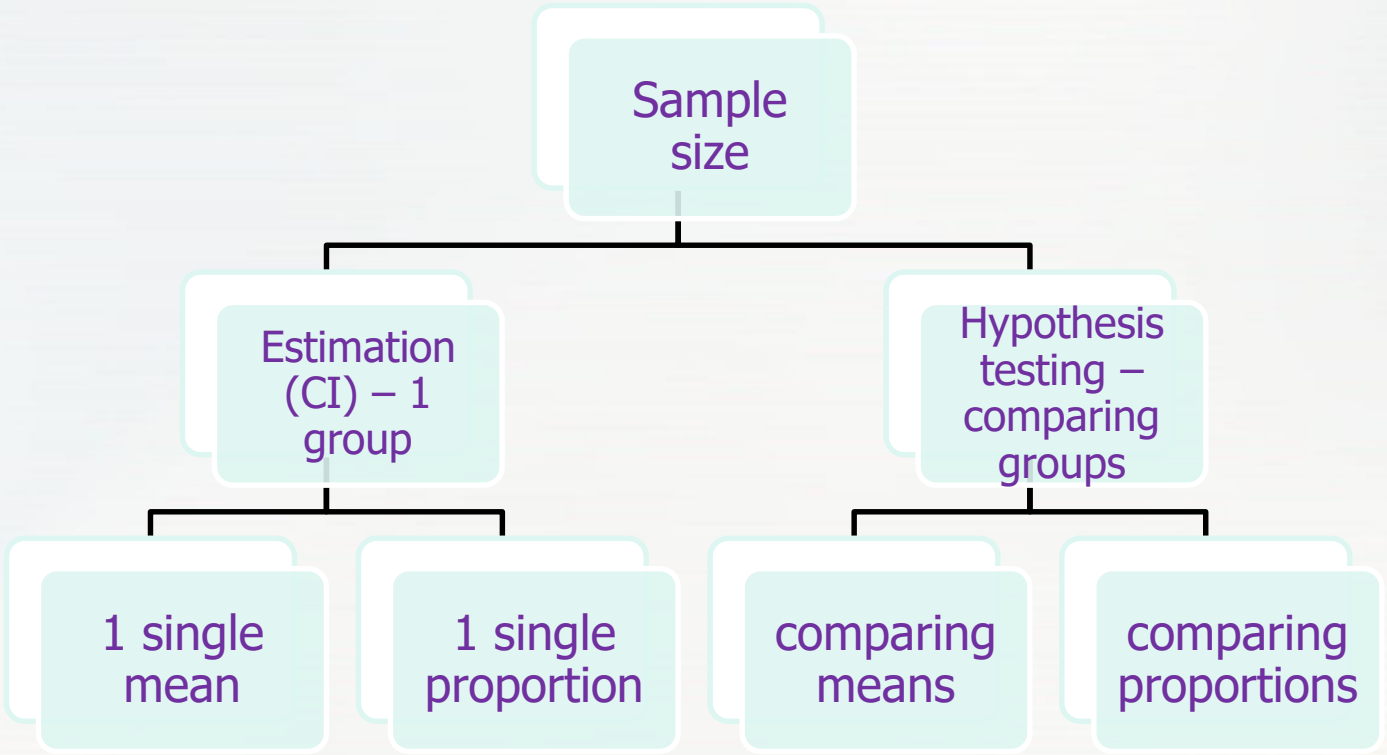
We need to have an adequate sample size to ensure the study results can represent our study population.



- **Sample size calculation must correspond to the study objectives.**
- **Different study objectives may require different calculations.**
- **Calculate the sample size required for all the main objectives & select the largest affordable size.**

Must also consider
available resources

We will cover:



Sample size calculation based on estimation (confidence interval)

- **used when we want to estimate the population parameter (e.g. mean BMI)**

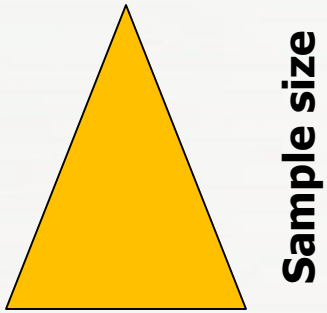
- **Estimate by giving lower and upper limits (i.e. 2 values - “confidence interval”).**

- **Confidence interval (CI): an interval within the true population parameter is likely to fall.**

- **However, how confident are we that the interval really contains the true value of the population parameter?**

- **Impossible to be 100% confident.**

- **How about being:**
  - **90% confident?**
  - **95% confident?**
  - **99% confident?**



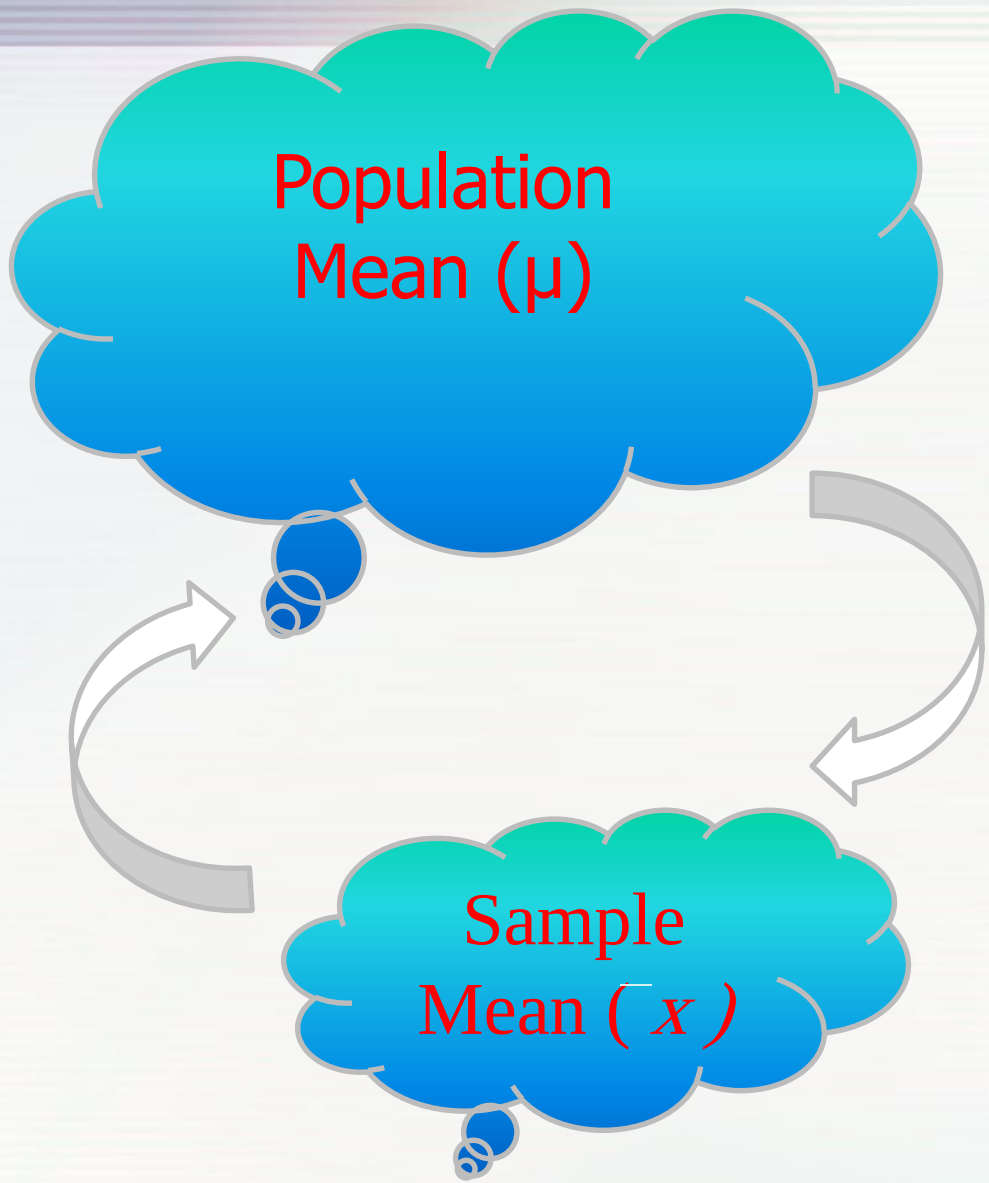
Normally researchers choose 95% confidence level

Sample size calculation for single mean (numerical variable)

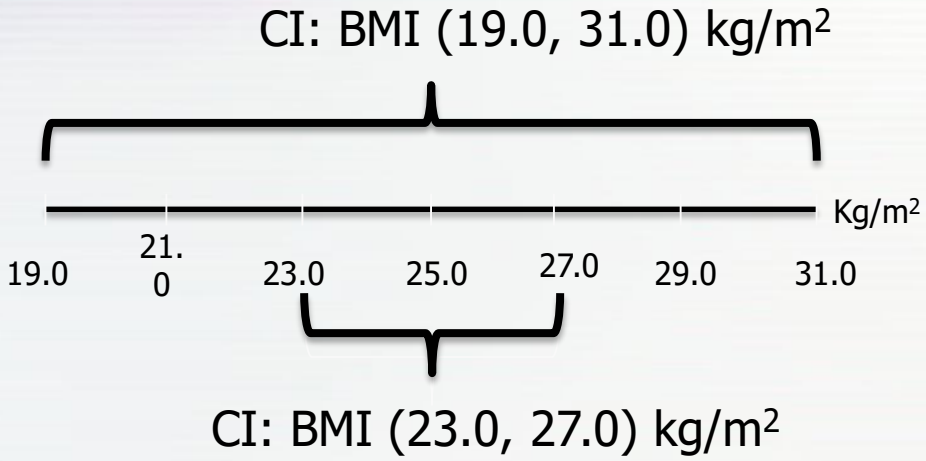
Estimation for single mean (numerical variable)

- **Objective 1: To determine the mean BMI (kg/m^2) of females aged 20-30 years living in Kermanshah.**

- **We want to be confident that the estimation (CI) is as close as possible to the mean BMI from all females aged 20-30 years living in Kermanshah (i.e. the population of our interest).**


Population
Mean (μ)
Sample
Mean ($\bar{x}$)

Narrow vs wide CI:



- **E.g.:**
- **which will give a more accurate estimation??**
- **Narrower CI : estimation of the population mean is more accurate, but we'll need a larger sample!**
- **Small sample size will give a wider CI: estimation is less accurate**

$$n = \left(\frac{Z * \sigma}{\Delta} \right)^2$$

Where:

Z = critical value for 95% confidence level is 1.96 (fixed)

σ = standard deviation: from previous literature, or pilot study

Δ : (*2 x Δ = CI width)* you decide based on practicality.

SAMPLE SIZE CALCULATOR FOR ESTIMATING MEAN

* USING RANDOM (NOT CLUSTER) SAMPLING

Level of Confidence = 95%

Standard Deviation = 6.4

Precision = 4

Population Size (*N*) =

From previous literature / pilot study

Step = 1

You decide

Precision (*d*)	Sample Size (*n*)		Suggestion for FPC application
	No FPC	With FPC	
± 4.00	10		Infinite population is assumed.
± 3.00	18		Infinite population is assumed.
± 2.00	40		Infinite population is assumed.
± 1.00	158		Infinite population is assumed.

List of suggested sample sizes, select the one most practical

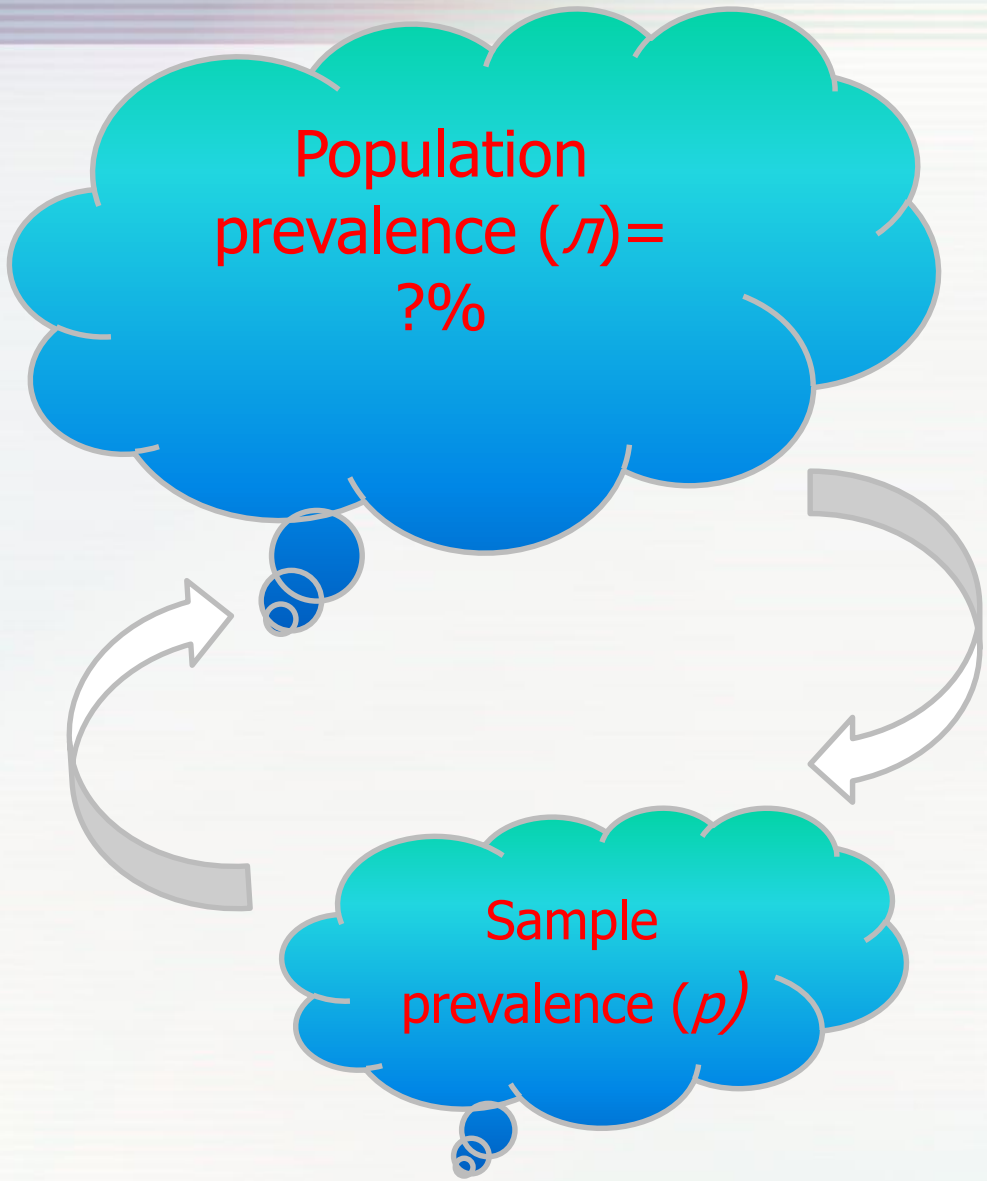
How to report the sample size determination in the 'methods' section?

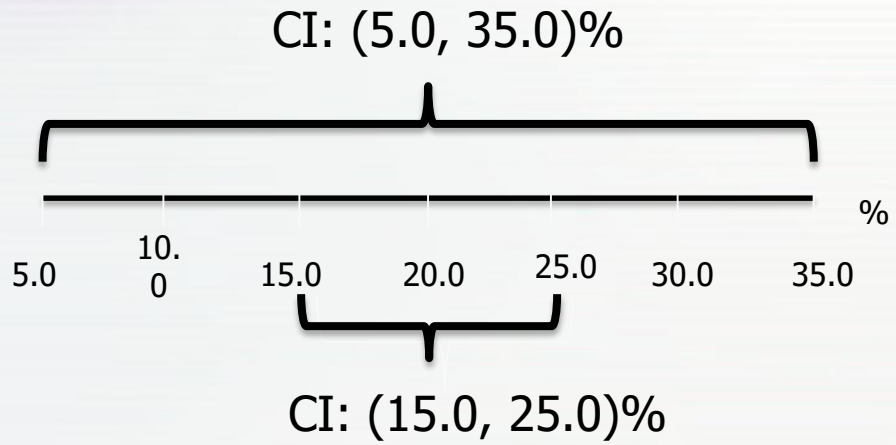
For objective 1 (to determine the mean BMI (kg/m^2) of females aged 20-30 years living in Kermanshah), at an estimated SD of 6.4 kg/m^2 *(reference?)* and precision of 2.0 kg/m^2, we need 40 subjects (determined using Sample Size Calculator for Prevalence Studies, Naing *et. al., 2006).* With the anticipation of 20% attrition rate, we will recruit 48 subjects for this study.

Attrition: non-response, drop-outs (cohort), missing data

Sample size calculation for single proportion (categorical variable)

- **Objective 2: To determine the proportion of females aged 20-30 years living in Kermanshah who are overweight (BMI ≥ 25 kg/m^2).**


Population prevalence (π)= ?%
Sample prevalence (p)



- **E.g.:**
- **Same concept as before,**
  - **large sample size will give narrower CI (more accurate estimation)**
  - **Small sample size will give wider CI (less accurate estimation)**

Sample size formula for single proportion:

$$n = \left(\frac{Z}{\Delta}\right)^2 * P(1-P)$$

where:

n = sample size

P = expected proportion (from previous literature / pilot study)

Δ = precision (remember, 2 x Δ = CI width)

Z value = 1.96 for 95% CI (fixed)

SAMPLE SIZE CALCULATOR FOR PREVALENCE STUDIES

* USING RANDOM (NOT CLUSTER) SAMPLING

Level of Confidence = 95%

Expected *P* = 0.2

Population Size (*N*) =

* Suggested precision (d) is 0.05.

0.2 = 20%
From previous literature / pilot study

Sample Size Table

Precision (*d*)	Sample Size (*n*)		Suggestion for FPC application	Assumption (Normality)
	No FPC	With FPC		
± 0.01	6147		◄ Infinite population is assumed.	OK
± 0.02	1537		◄ Infinite population is assumed.	OK
± 0.03	683		◄ Infinite population is assumed.	OK
± 0.04	385		◄ Infinite population is assumed.	OK
± 0.05	246		◄ Infinite population is assumed.	OK
± 0.06	171		◄ Infinite population is assumed.	OK
± 0.07	126		◄ Infinite population is assumed.	OK
± 0.08	97		◄ Infinite population is assumed.	OK
± 0.09	76		◄ Infinite population is assumed.	OK
± 0.10	62		◄ Infinite population is assumed.	OK
± 0.11	51		◄ Infinite population is assumed.	OK
± 0.12	43		◄ Infinite population is assumed.	OK
± 0.13	37		◄ Infinite population is assumed.	OK
± 0.14	32		◄ Infinite population is assumed.	OK

List of suggested sample sizes

How to report the sample size determination in the 'methods' section?

- **For objective 2 [to determine the proportion of females aged 20-30 years living in Kermanshah who are overweight (BMI ≥ 25 kg/m^2)], at an estimated proportion of 20% (*reference?)* and precision of 5%, we will need 246 subjects (determined using Sample Size Calculator for Prevalence Studies, Naing *et. al., 2006)*. Anticipating 20% attrition rate, we will recruit 300 subjects for this study.**

Sample size calculation for comparing groups – hypothesis testing

Some sample size formulae (don't memorise!)

Comparing 2 independent means (independent *t* test)

$$n = \frac{2\sigma^2}{\Delta^2}\left(z_{\alpha/2} + z_{\beta}\right)^2$$

Comparing 2 dependent means (paired *t* test)

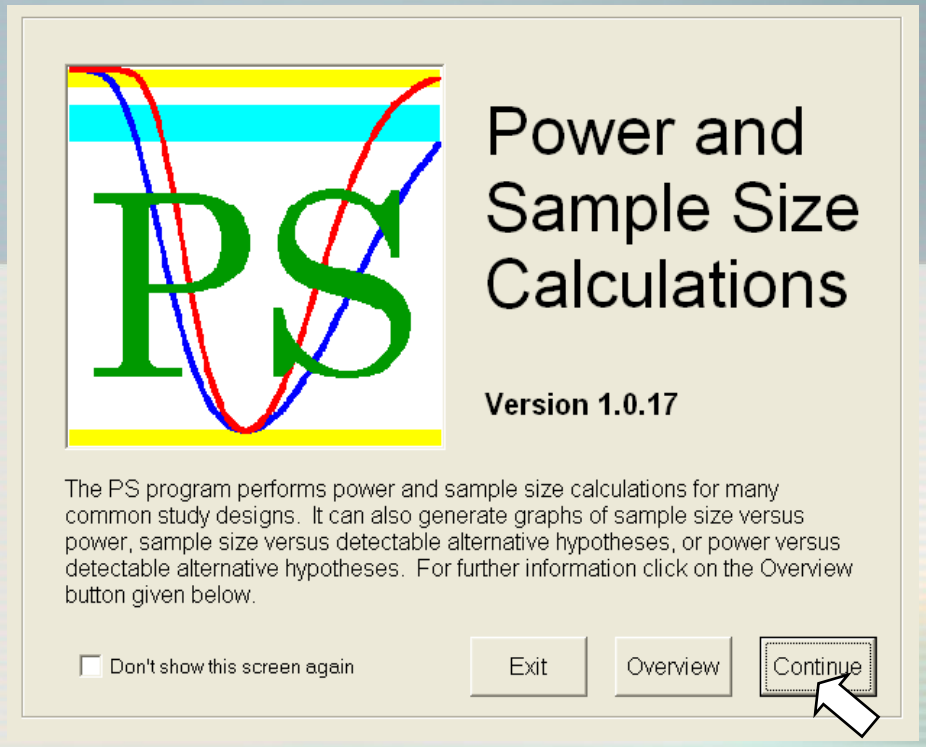
$$n = \left(\frac{\left(z_{\alpha} + z_{\beta}\right)\sigma_d}{\Delta}\right)^2$$

Comparing 2 proportions (Chi-square test)

$$n = \frac{p_1(1-p_1) + p_2(1-p_2)}{(p_1 - p_2)^2}(z_{\alpha} + z_{\beta})^2$$

Using PS software ….

- Comparing means
- Comparing proportions



http://biostat.mc.vanderbilt.edu/twiki/bin/view/Main/PowerSampleSize

E.g. we want to compare between 2 means:

is there a significant difference between:

the mean BMI of females aged 20-30 years living in Kermanshah

and

the mean BMI of females aged 20-30 years living in Kurdestan

(Mean BMI from 2 independent/separate groups, - researchers will use independent *t*-test analysis later)

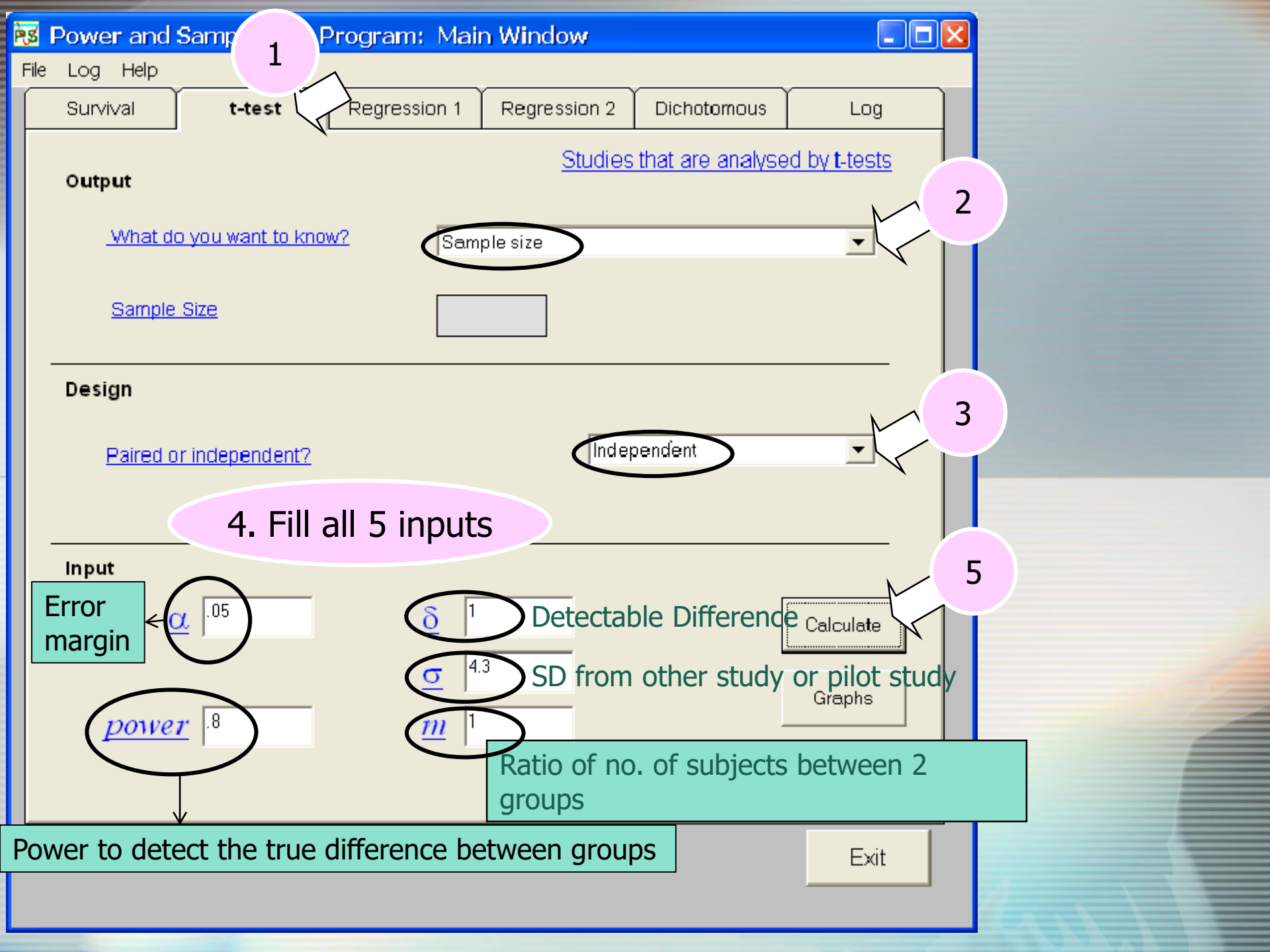
- **Information needed by PS sample size calculator for comparing means:**

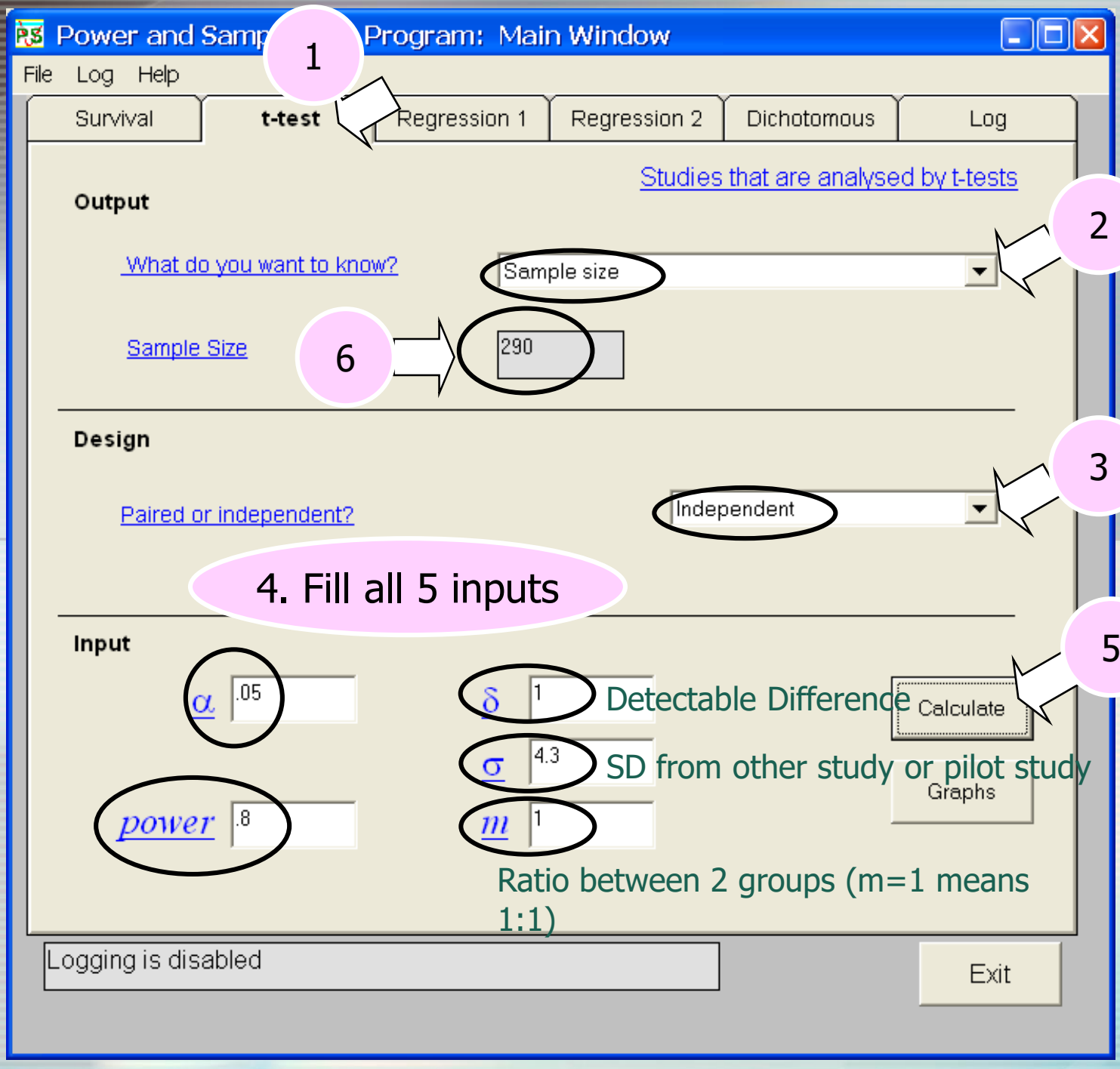
  i. α (alpha): margin of error. Normally we allow 5% (0.05).

  ii. power of the study: power to detect the true difference between the groups. Normally set at 80% (0.8)

  iv. δ (detectable difference) : the smallest difference that would be statistically significant when comparing different groups

  v. m (ratio): ratio of no. subjects between groups. Ideally '1' to indicate 1:1


Power and Samp Program: Main Window
File Log Help
Survival
t-test
Regression 1
Regression 2
Dichotomous
Log
1
Studies that are analysed by t-tests
Output
What do you want to know?
Sample size
2
Sample Size
Design
Paired or independent?
Independent
3
4. Fill all 5 inputs
Input
Error margin
α .05
δ 1
Detectable Difference
Calculate
5
σ 4.3
SD from other study or pilot study
Graphs
power .8
m 1
Ratio of no. of subjects between 2 groups
Power to detect the true difference between groups
Exit


Power and Samp Program: Main Window
File Log Help
Survival
t-test
Regression 1
Regression 2
Dichotomous
Log
1
Studies that are analysed by t-tests
Output
What do you want to know?
Sample size
2
Sample Size
6
290
Design
Paired or independent?
Independent
3
4. Fill all 5 inputs
Input
α .05
δ 1
Detectable Difference
Calculate
5
σ 4.3
SD from other study or pilot study
Graphs
power .8
m 1
Ratio between 2 groups (m=1 means 1:1)
Logging is disabled
Exit

How to report?

We used PS software (Dupont & Plummer, 1997) to calculate the sample size based on comparing 2 means.

To detect the difference of 1 unit (BMI kg/m^2) with 80% power and alpha 0.05, we need 291 subjects in each study group (SD was estimated as 4.3 kg/m^2, *source?*). Anticipating a 20% attrition rate, we will recruit 350 subjects in each group.

Reference:

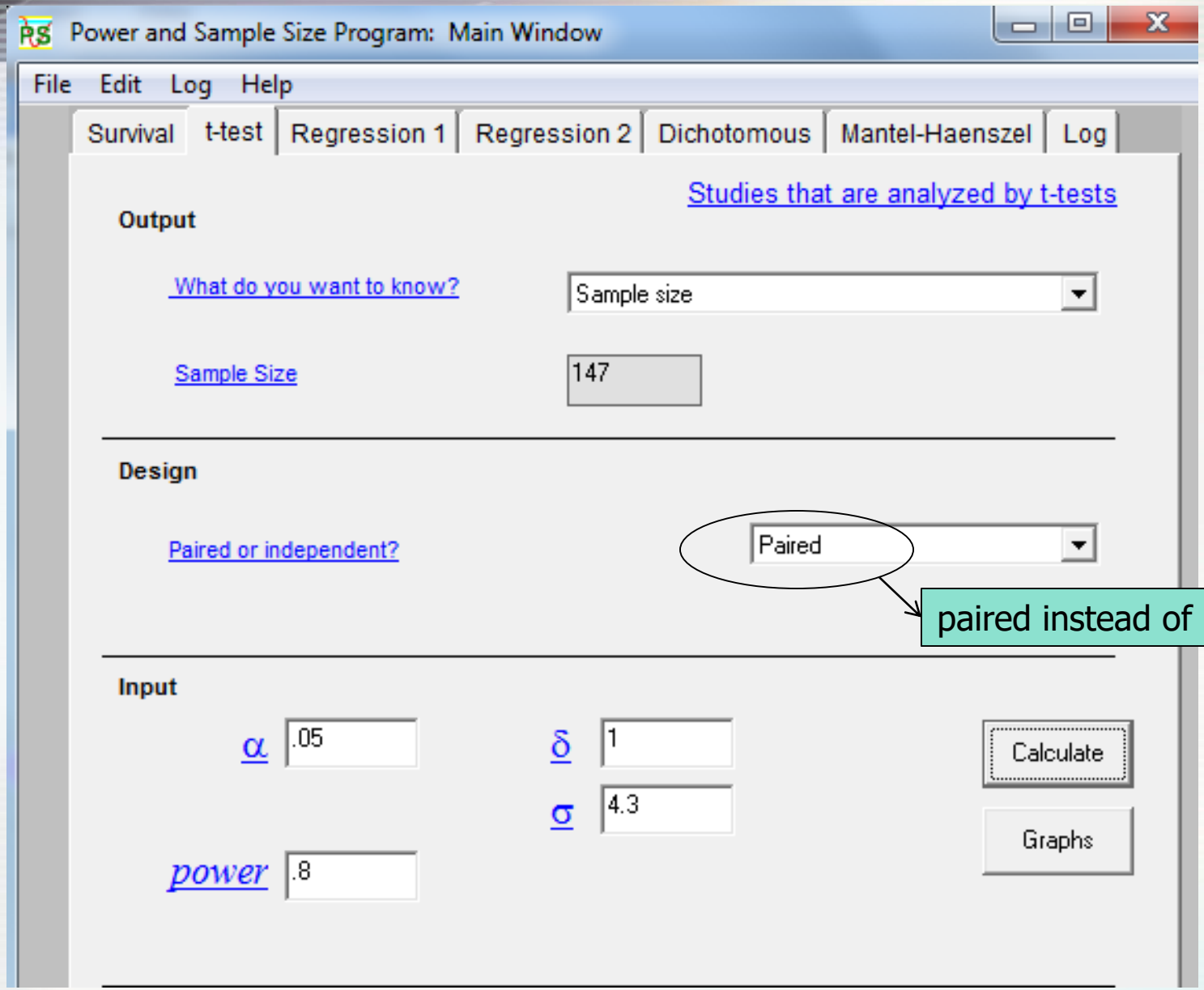
Dupont WD and Plummer WD: PS power and sample size program available for free on the Internet. Controlled Clin Trials,1997;18:274.
http://www.mc.vanderbilt.edu/prevmed/ps.htm

Using PS software ….

- Comparing 2 means- paired groups

Researchers want to compare BMI (kg/m^2) of females aged 20-30 years living in Kermanshah before and after health intervention (i.e. 2 paired/ same groups).

This means that researchers need to use paired *t*-test.


Power and Sample Size Program: Main Window
File Edit Log Help
Survival t-test Regression 1 Regression 2 Dichotomous Mantel-Haenszel Log
Studies that are analyzed by t-tests
Output
What do you want to know?
Sample size
Sample Size
147
Design
Paired or independent?
Paired
paired instead of independent
Input
α .05
δ 1
σ 4.3
power .8
Calculate
Graphs

How to report?

We used PS software (Dupont & Plummer, 1997) to calculate the sample size based on comparing 2 means. To detect the difference of 1 unit (BMI kg/m^2) with 80% power and alpha 0.05, we need 147 subjects (SD was estimated as 4.3 kg/m^2, *source?*). Anticipating a 20% attrition rate, we will recruit 180 subjects in each group.

Reference:

Dupont WD and Plummer WD: PS power and sample size program available for free on the Internet. Controlled Clin Trials,1997;18:274.
http://www.mc.vanderbilt.edu/prevmed/ps.htm

Using PS software ....

- Comparing 2 proportions

  Researchers want to determine if there is any difference between:

  the prevalence of females aged 20-30 years who are overweight (BMI ≥ 25 kg/m^2) living in Kermanshah

  and

  the prevalence of females aged 20-30 years who are overweight (BMI ≥ 25 kg/m^2) living in Kurdestan

- **Information needed by PS sample size calculator for comparing proportions:**

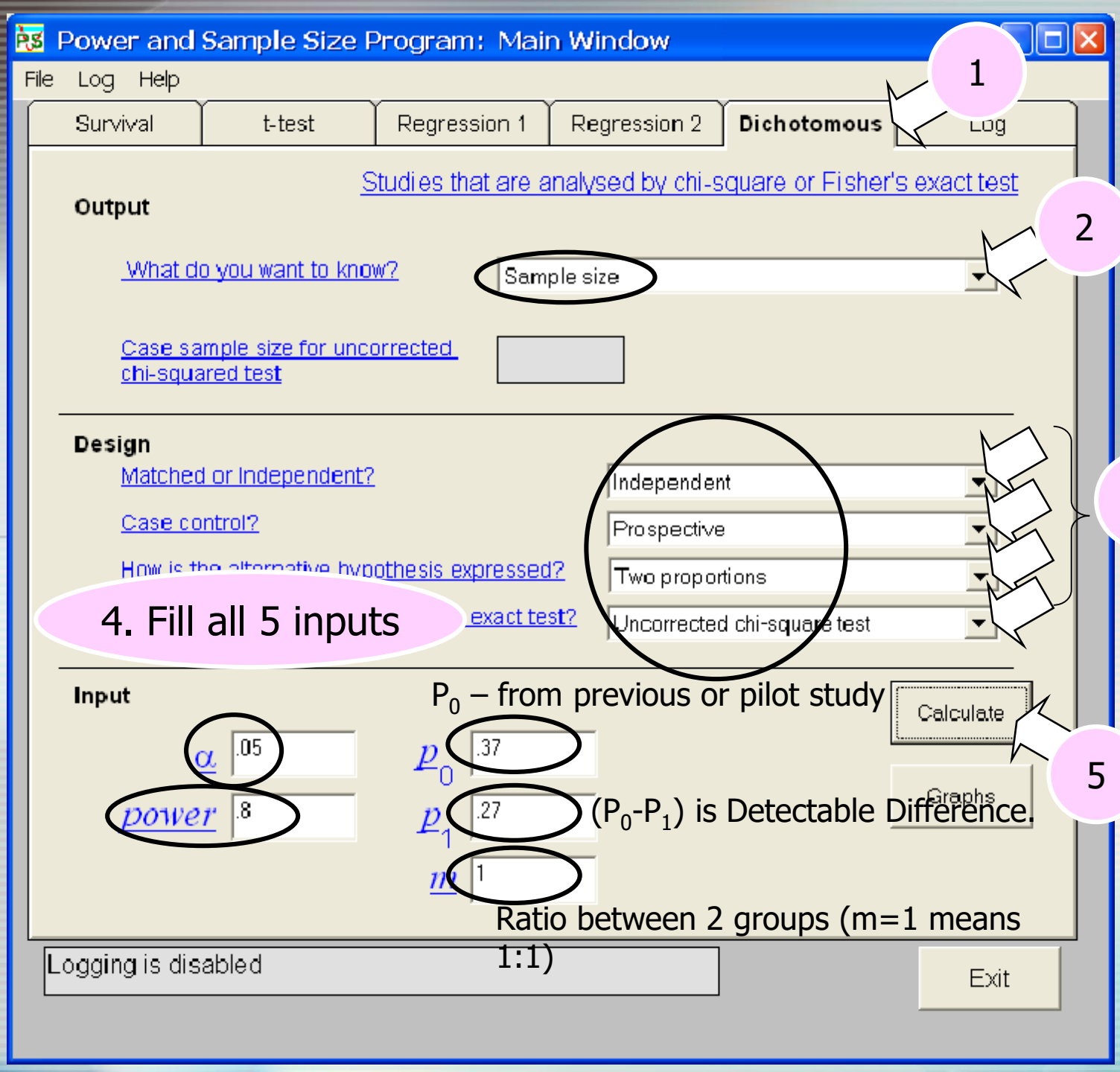
  i. α (alpha): margin of error. Normally we allow 5% (0.05).

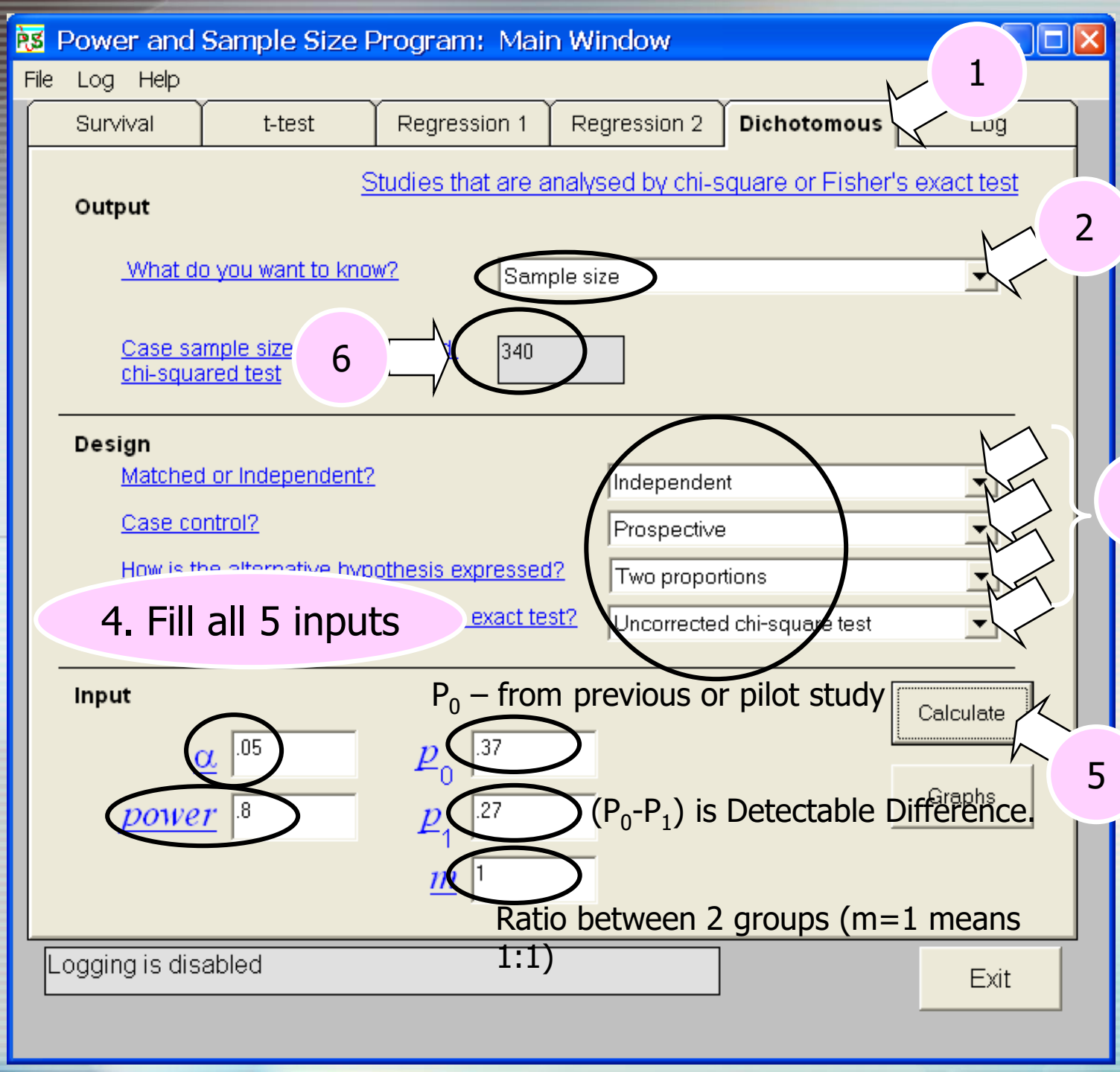
  ii. power of the study: power to detect the true difference between the groups. Normally set at 80% (0.8)

  iv. p_0 : estimated proportion in population 1– estimated from the results from previous literature / pilot study

  v. p_1 : estimated proportion in population 2 – you decide. Small gap between p_0 & p_1 requires large sample size. Must consider available resources and adjust accordingly.

  vi. m (ratio): ratio of no. subjects between groups. Ideally '1' to indicate 1:1


Power and Sample Size Program: Main Window
File Log Help
Survival t-test Regression 1 Regression 2 Dichotomous Log
1
Studies that are analysed by chi-square or Fisher's exact test
Output
What do you want to know?
Sample size
2
Case sample size for uncorrected chi-squared test
Design
Matched or Independent?
Independent
Case control?
Prospective
How is the alternative hypothesis expressed?
Two proportions
exact test?
Uncorrected chi-square test
3
4. Fill all 5 inputs
Input
P_0 – from previous or pilot study
Calculate
α .05
p_0 .37
power .8
p_1 .27
(P_0-P_1) is Detectable Difference.
Graphs
5
m 1
Ratio between 2 groups (m=1 means 1:1)
Logging is disabled
Exit


Power and Sample Size Program: Main Window
File Log Help
Survival t-test Regression 1 Regression 2 Dichotomous Log
1
Studies that are analysed by chi-square or Fisher's exact test
Output
What do you want to know?
Sample size
2
Case sample size chi-squared test
6
340
Design
Matched or Independent?
Independent
Case control?
Prospective
How is the alternative hypothesis expressed?
Two proportions
exact test?
Uncorrected chi-square test
3
4. Fill all 5 inputs
Input
α .05
power .8
p_0 .37
p_1 .27
m 1
P0 – from previous or pilot study
(P0-P1) is Detectable Difference.
Ratio between 2 groups (m=1 means 1:1)
Calculate
Graphs
5
Logging is disabled
Exit

How to report?

We used PS software (Dupont & Plummer, 1997) to calculate the sample size based on comparing 2 proportions.
To detect the difference of 10% in proportion (P_1-P_0) with 80% power and alpha 0.05, we need 340 subjects in each study group. (P_o- the prevalence of females aged 20-30 years who are overweight in Bertam Indah was estimated at 37%, source?). Anticipating a 20% attrition rate, we will recruit 410 subjects in each group.

Reference:
Dupont WD and Plummer WD: PS power and sample size program available for free on the Internet. Controlled Clin Trials,1997;18:274.
http://www.mc.vanderbilt.edu/prevmed/ps.htm


i'M READY!

Acknowledgement

- **Some of the materials in this lecture were adapted from the slides of Assoc. Prof. Dr Lin Naing @ Mohd Ayub Saddiq (with permission).**
- **Thanks for my dear daughter Khansa**