Sample Size

Prof Dr. Adinegara bin Lutfi

II) Methodology

- a) Describe study design
- b) Identify study population
- c) Sample size calculation
- d) Case –diagnostic criteria, inclusion and exclusion criteria
- Non-cases-describe
- e) Sampling method - SRS
- - Systematic RS
- - Cluster sampling
- - Stratified sampling
- f) Data collection
- g) Data processing
- h) Data analysis / Interpretation

Sample size

-the required size of the respondents (optimum) that is to be included in the study

Sample size

- Depends on:
- - degree of precision required
- - type of methodology used

Table 1
Comparison of characteristics via quantitative analysis amongst case (PIH) and control (NPIH)

Independent Variables	Case Mean (SD)	Control Mean (SD)	Student t test		P value
Demographic / Socio-economic variables					
Age (years)	33.50 (5.93)	31.63 (6.33)	1.18		0.243
Income (Ringgit Malaysia)	790.74 (464.51)	1354.16 (1532.46)	-1.82		0.075
Monthly grocery bills (Ringgit Malaysia)	361.07 (165.62)	328.00 (139.88)	0.78		0.439
Physiological findings					
Weight (in kg)	71.09 (16.70)	60.13 (16.32)	2.57		0.013 **
Height (in cm)	1.54 (0.05)	1.53 (0.06)	1.00		0.784
Body mass index (kg/meter^2)	29.86 (6.68)	25.34 (6.50)	2.65		0.010**
Haemoglobin level (g/dL)	10.99 (1.43)	11.38 (1.05)	-1.18		0.241
Lifestyle					
Exercise sessions per week	0.71 (1.62)	0.80 (1.34)	-0.22		0.830
Number of cigarettes smoked by husband per day	6.34 (7.51)	5.60 (8.26)	0.35		0.726

P value

- Probability that the observed data is the result of chance

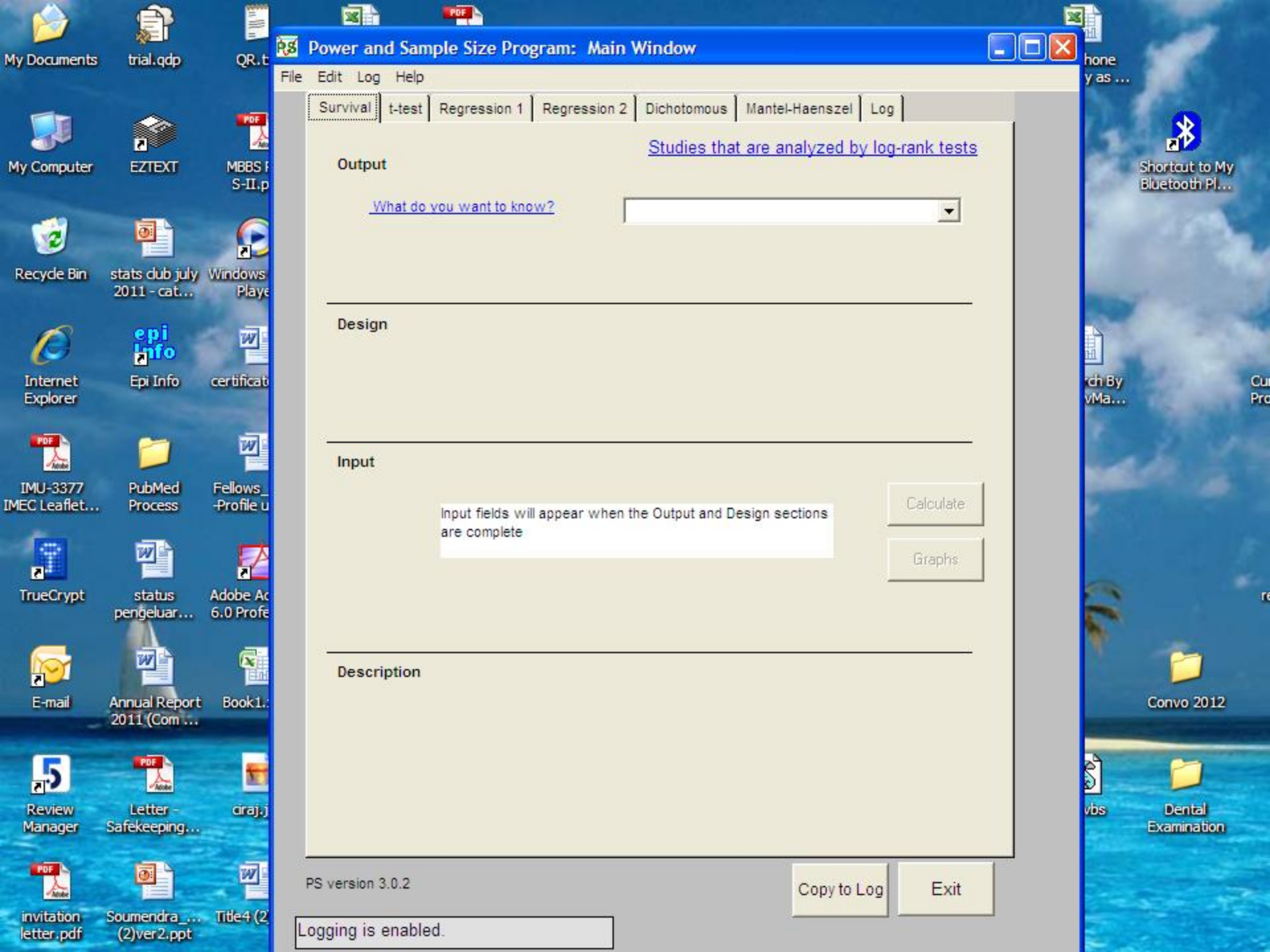
- Observed data of variable
- =chance + true difference
- Observed data of variable (100%)
- =chance (3%) + true difference (97%)

- The smaller chance is, the greater is the true difference of data – statistically significant
- P value < 0.05 (or 5%)– statistically significant

Test result	Actual condition: Disease	Actual condition: Non- Disease
Test shows Positive result	**Correct**	**False positive**
Test shows Negative result	**False negative**	**Correct**

Test result	Actual condition: Difference	Actual condition: No Difference
Test shows Positive result $P \leq 0.05$	Power	**False positive**
Test shows Negative result $P > 0.05$	**False negative**	Confidence level

	Null is false	Null is true
Reject Null	**Correct**	**Type I error (alpha)**
Accept Null	**Type II error (beta)**	**Correct**


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 log-rank tests
Output
What do you want to know?
Design
Input
Input fields will appear when the Output and Design sections are complete
Calculate
Graphs
Description
PS version 3.0.2
Copy to Log
Exit
Logging is enabled.

Using PS Software to compare 2 proportions.....

Analysis for Project shall involve P value

Parameters needed

- α = Type 1 error
- Power = 1- β (Type 2 error)
- m (ratio) = 1
- For case control:
- Proportion 0 (P0) = Proportion of exposure in NOT ILL group
- Proportion 1 (P1) = Proportion of exposure in ILL group

Parameters needed

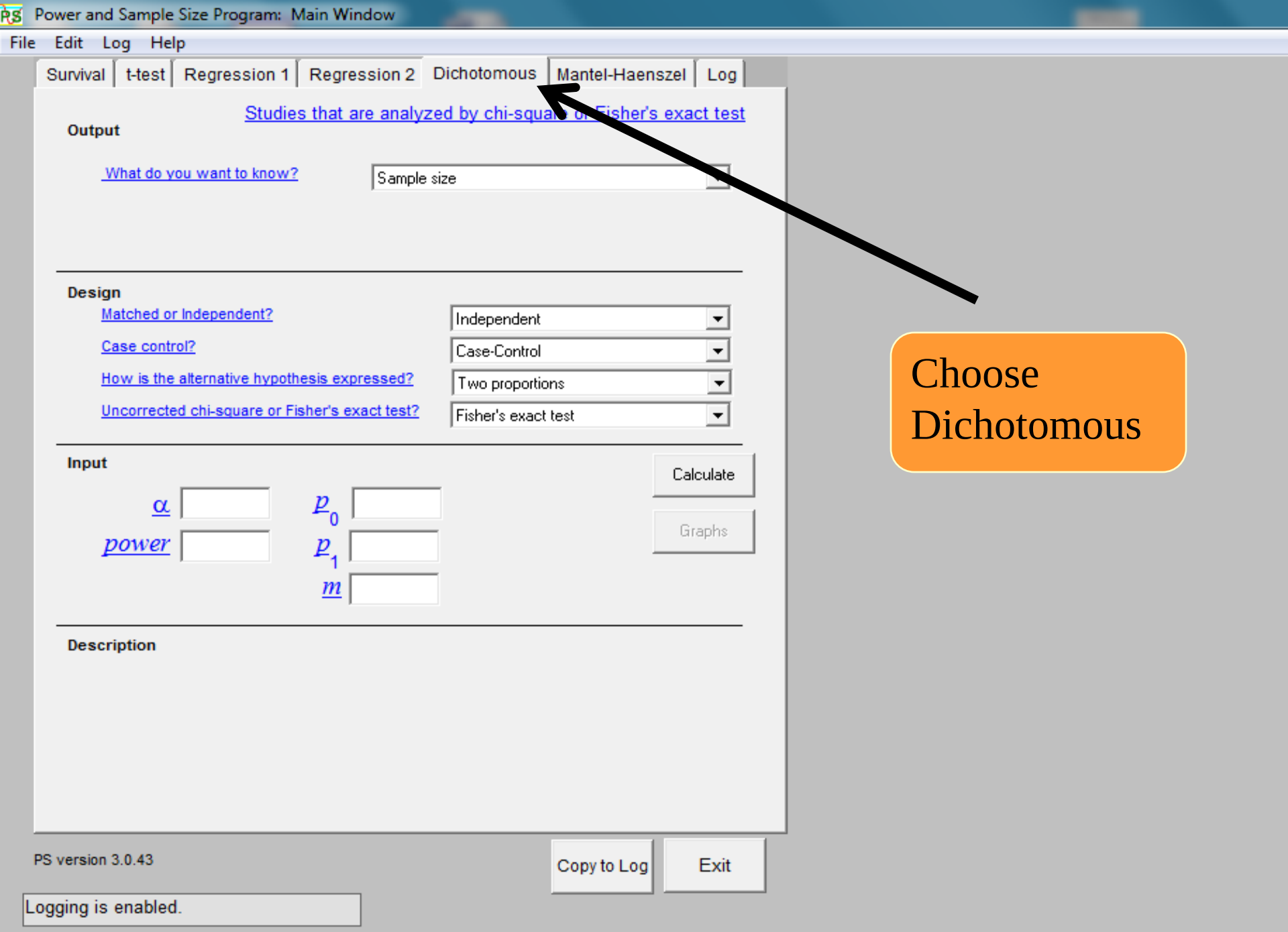
- α = Type 1 error
- Power = 1- β (Type 2 error)
- m (ratio) = 1
- For cohort / cross-sectional:
- Proportion 0 (P0) = Proportion of disease in UNEXPOSED group
- Proportion 1 (P1) = Proportion of disease in EXPOSED group

Data analysis: Creating a contigency table

- Ex. In a case control study, it was noted that out of 78 cases of lung cancer, 55 of these cases has had history of smoking before developing the disease.
- The control group consisted of 82 non-lung cancers in which only 5 of them has had history of smoking.

	Cancer of Lung	Non-Ca	Total
Smokers	55	25	80
Non-Smokers	45	77	122
Total	100	102	202

Row X Column Table



Choose Dichotomous

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 chi-square or Fisher's exact test

Output

What do you want to know? Sample size

Design

Matched or Independent? Independent

Case control? Case-Control

How is the alternative hypothesis expressed? Two proportions

Uncorrected chi-square or Fisher's exact test? Fisher's exact test

Input

α p_0

power p_1

m

Calculate

Graphs

Description

PS version 3.0.43

Copy to Log Exit

Logging is enabled.

Choose Sample Size

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 chi-square or Fisher's exact test

Output

What do you want to know? — Sample size

Design

Matched or Independent? — Independent

Case control? — Case-Control

How is the alternative hypothesis expressed? — Two proportions

Uncorrected chi-square or Fisher's exact test? — Fisher's exact test

Input

α ____ p_0 ____

power ____ p_1 ____

m ____

Calculate

Graphs

Description

PS version 3.0.43

Copy to Log

Exit

Logging is enabled.

Follow accordingly

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 chi-square or Fisher's exact test

Output

What do you want to know? Sample size

Design

Matched or Independent? Independent

Case control? Case-Control

How is the alternative hypothesis expressed? Two proportions

Uncorrected chi-square or Fisher's exact test? Fisher's exact test

Input

α 0.05

power 0.8

p_0 0.245

p_1 0.55

m 1

Calculate

Graphs

Description

PS version 3.0.43

Copy to Log

Exit

Logging is enabled.

- We are planning a study of independent cases and controls with 1 control(s) per case. Prior data indicate that the probability of exposure among controls is 0.245. If the true probability of exposure among cases is 0.55, we will need to study 45 case patients and 45 control patients to be able to reject the null hypothesis that the exposure rates for case and controls are equal with probability (power) 0.8. The Type I error probability associated with this test of this null hypothesis is 0.05. We will use a continuity-corrected chi-squared statistic or Fisher's exact test to evaluate this null hypothesis.

How to write

Edit View Favorites Tools Help

Back Search Favorites

E:\from office\SemVII\Epid Study Design\Preterm labour\previous preterm delivery -- Hay et al_ 308 (6924) 295 -- BMJ.htm

Go

births were excluded because they occurred in only three women. Table III shows the results for the logistic analysis of the 467 women recruited before weeks' gestation including those who had a late miscarriage. No woman miscarried before 16 weeks' gestation. In this analysis ethnic origin was not nificant (P>0.25) in the independent logistic analysis.

TABLE II - Outcome of pregnancy in women recruited before 24 weeks' gestation who delivered from 24 weeks' gestation

Variable	No (%) who had preterm delivery	Total	Odds ratio (95% confidence interval)	P value
Grade of flora:				
I	17 (2.9)	584	1	0.04
II	1 (3.1)	32	2.8 (1.1 to 7.4)	
III	8 (9.6)	83		
Obstetric history:				
No previous pregnancy	9 (3.9)	231	1	
Previous preterm labour:				
No	9 (2.0)	450	1	<0.001
Yes	9 (37.5)	24	24.8 (8.8 to 69.9)	
Ethnic origin:				
White	11 (2.6)	427	1	0.192
Afro-Caribbean	3 (7.3)	41	3.1 (0.8 to 12.9)	
Asian	9 (4.4)	206	2.0 (0.8 to 5.2)	
Other	2 (8.3)	24	4.3 (0.8 to 24.0)	

ABLE III - Outcome of pregnancy in women recruited before 16 weeks' gestation who delivere
6 weeks' gestation

	No (%) who had late miscarriage	No (%) who had preterm delivery	Total	Odds rati (95% confidence

one

How to obtain data from literature

Using PS Software to compare 2 means.....

Analysis for Project shall involve P value

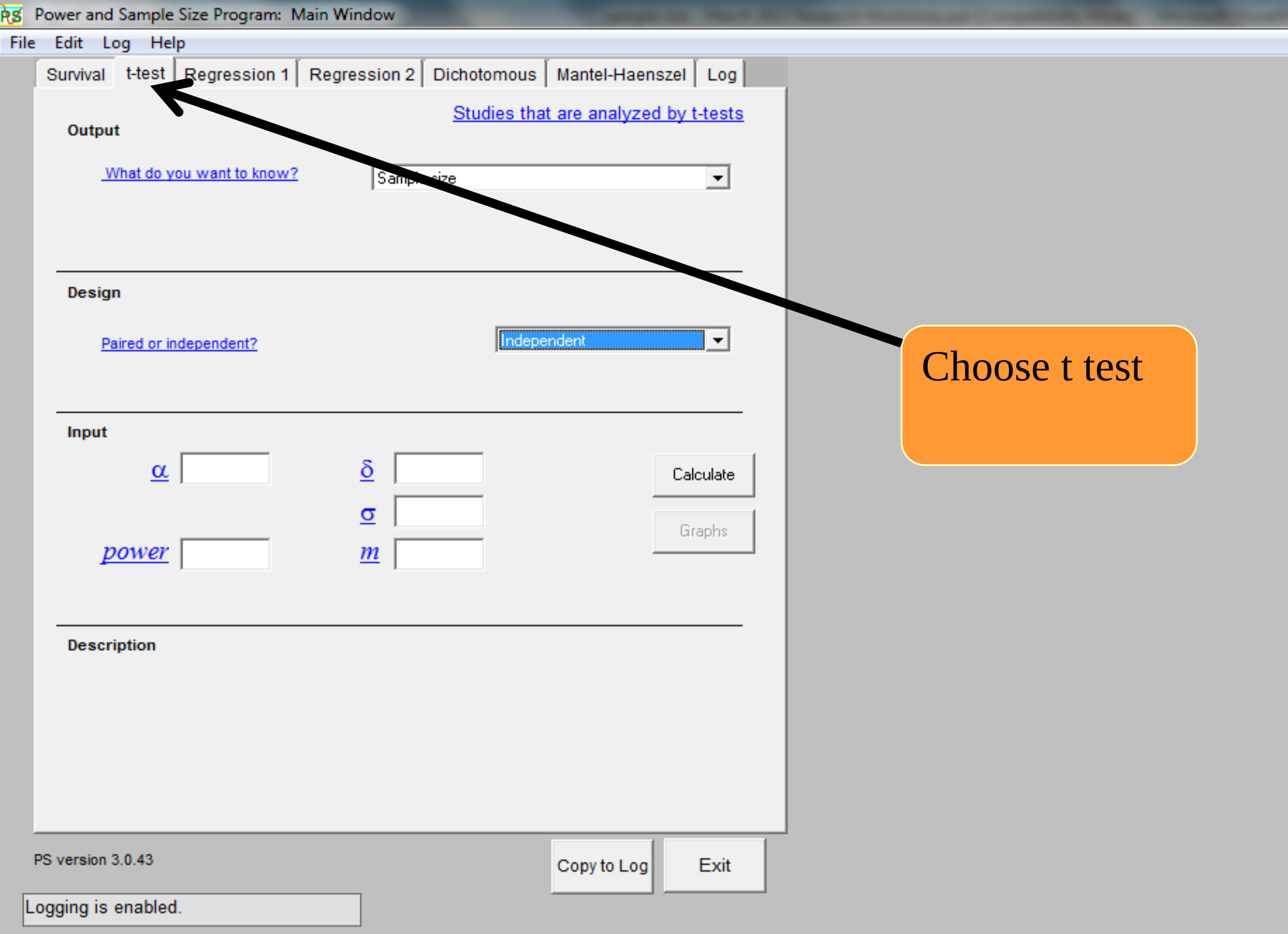
Parameters needed

- α = Type 1 error
- Power = 1- β (Type 2 error)
- m (ratio) = 1
- δ = detectable difference (mean 1 minus mean 2)
- σ = standard deviation

Continuous outcome

The primary outcome of interest is the mean difference in an outcome variable between two treatment groups.

For example, it is postulated that a good clinical response difference between the active and placebo groups is 0.2 units with an SD of 0.5 units, how many subjects will be required to obtain a statistical significance for this clinical difference?


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
Design
Paired or independent?
Independent
Input
α
δ
σ
power
m
Calculate
Graphs
Description
PS version 3.0.43
Copy to Log
Exit
Logging is enabled.
Choose 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

Design

Paired or independent? Independent

Input

α 0.05

δ 0.2

σ 0.5

power 0.8

m 1

Calculate

Graphs

Description

PS version 3.0.43

Copy to Log

Exit

Logging is enabled.

- We are planning a study of a continuous response variable from independent control and experimental subjects with 1 control(s) per experimental subject. In a previous study the response within each subject group was normally distributed with standard deviation 0.5. If the true difference in the experimental and control means is 0.2, we will need to study 99 experimental subjects and 99 control subjects to be able to reject the null hypothesis that the population means of the experimental and control groups are equal with probability (power) 0.8. The Type I error probability associated with this test of this null hypothesis is 0.05.

How to write

Using EPI 7 to look into prevalence.....

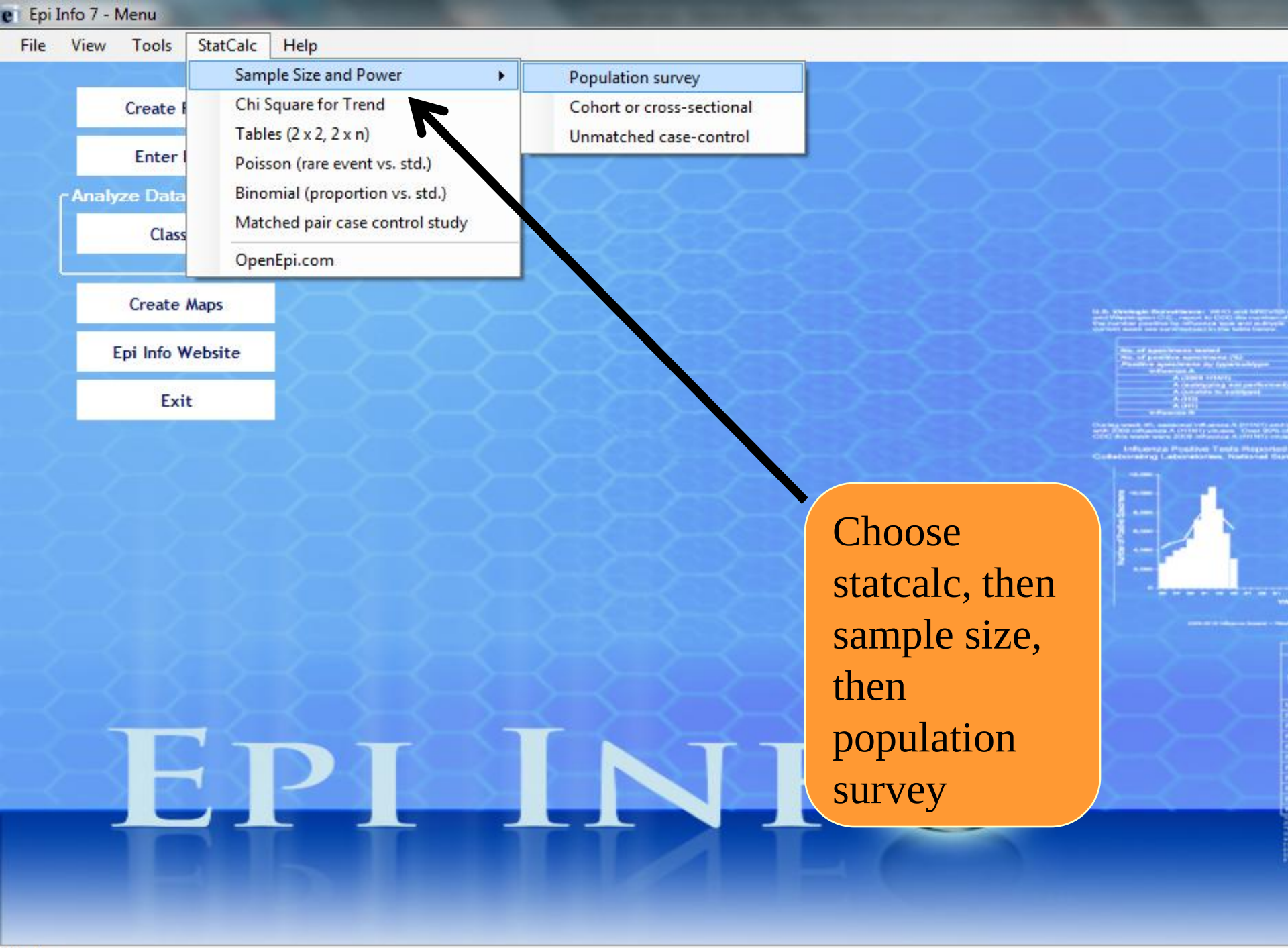
Analysis for Project shall involve only percentages, graphs – NO P value

Epi Info 7 - Menu
File
View
Tools
StatCalc
Help
Create Forms
Enter Data
Analyze Data
Classic
Visual Dashboard
Create Maps
Epi Info Website
Exit
EPI INFO™
Ready

Launch EPI 7

Parameters needed

- Population size = population to be faced during data collection
- Expected frequency = Frequency of outcome / disease
- Confidence limits = Margin of error – amount of error in % you can tolerate. The lower the margin, the larger the sample size - usually we go for 5% or 6%
- Cluster = usually 1 unless we do cluster sampling – then need to count the clusters



Choose statcalc, then sample size, then population survey

Population survey or descriptive study
For simple random sampling, leave design effect and clusters equal to 1.

Population size:	500
Expected frequency:	35 %
Confidence limits:	6 %
Design effect:	1.0
Clusters:	1

Confidence Level	Cluster Size	Total Sample
80%	86	86
90%	127	127
95%	163	163
97%	187	187
99%	228	228
99.9%	289	289
99.99%	328	328

Population = 500
Obesity = 35%
Margin of error = 6%
Confidence level =95%

Thank-you