Sample calculation for medical research

Monthly Research Meeting - 31 January 2019



Dr. Dileepa Ediriweera MBBS, MSc(Biostat), MSc(Biomedical Informatics), DCSD

Population vs Sample

- Population value vs Sample value
 - Parameter vs statistic
- Notations
 - Greek vs English











Centre for Health Informatics, Biostatisics and Epidemiology, Faculty of Medicine, University of Kelaniya

Confidence interval

 Thus a 95% confidence interval for the <u>Population mean</u> is

$$\overline{x} \pm 1.96 \cdot \frac{\sigma}{\sqrt{n}}$$

This is in the form

Point estimate ± margin of error

The margin of error here is $1.96 \cdot \sigma / \sqrt{n}$



Sampling Distribution of \hat{p} (p=0.70)





Rentre for Health Informatics, Biostatisics and Epidemiology, Faculty of Medicine, University of Kelaniya

Sample size calculation

- 1. Estimate a prevalence (percentage/proportion)
 - Eg. Prevalence of obesity among school children
- 2. Estimate a mean
 - Eg. Mean Hb level among post-op ca patients
- 3. Estimate a percentage difference
 - Eg. Smoking prevalence difference in lung caner and normal people
- 4. Estimate a mean difference
 - Eg. FBS difference in group A and B



Calculating sample size for a crosssectional and longitudinal studies

- For proportions (i.e. binary exposure)
- For means (i.e. continuous exposure)



Sample size formula for percentages (nominal or ordinal) $\frac{Z^2 * ([P * Q])}{e^2}$

Where

n = required sample size

Z = the Z value for your desired confidence level

P = estimation of the population %

Q = (100 - P)

e = desired accuracy range



Exercise

- I want to estimate the prevalence of overweight in my clinic patients.
- Method:
 - Draw a sample of clinical records and estimate the prevalence
 - I want an estimate with a margin of error of 5%, with 95% confidence, and assume that the national population overweight prevalence of adult (i.e. 25%) also applies to my patients
 - How large a sample do I need to draw?
- Answer: 288



Exercises

- **1**. P = 20%, e = 5
- **2**. P = 40%, e = 5
- **3**. P = 50%, e = 5
- **4**. P = 60%, e = 5
- 5. P = 80%, e = 5
- 6. P = 1%, e = 5?



Sample size formula for means (interval data)

 $Z^2 * \sigma^2$

 $n = e^2$

where

n = required sample size

Z = the Z value for your desired confidence level $\sigma =$ estimated standard deviation for the population mean

e = desired accuracy range



Exercise

- I want to estimate the average weight of my clinic patients
- Method
 - Draw a sample of clinical records and estimate the average weight
 - I want an estimate with a margin of error of 5 kg, with 95% confidence, and assume that the national population standard deviation of adult weight of 25 also applies to my patients..
 - How large a sample do I need to draw?
- Answer: 96



Exercises

SD = 20, e = 5
SD = 40, e = 5
SD = 50, e = 5
SD = 60, e = 5
SD = 80, e = 5
SD = 1, e = 5?



Calculating sample size for case-control studies

- Use difference in proportions (i.e. binary exposure)
- Use difference in means (i.e. continuous exposure)



Difference in proportions





Example

- How many cases and controls do you need to sample assuming
 - 80% power
 - An equal number of cases and controls (r=1)
 - The proportion exposed in the control group is 20%
 - The proportion exposed in the case group is 33%



Example

$$n = (\frac{r+1}{r}) \frac{(\bar{p})(1-\bar{p})(Z_{\beta} + Z_{\alpha/2})^2}{(p_1 - p_2)^2}$$

- For 80% power, Z_{β} =.84
- For 0.05 significance level, $Z_{\alpha} = 1.96$
- r=1 (equal number of cases and controls)
- The proportion exposed in the control group is 20%
- The proportion exposed in the case group is 33%
- Average proportion exposed = (.33+.20)/2=.265
- Answer: n=362 (181 cases, 181 controls)



Difference in means





Centre for Health Informatics, Biostatisics and Epidemiology, Faculty of Medicine, University of Kelaniya

Example

- How many cases and controls do you need to sample assuming
 - 80% power
 - The standard deviation of the characteristic you are comparing is 15.0
 - You want to detect a difference in your characteristic of 7.0 (one half standard deviation)
 - An equal number of cases and controls (r=1)



Example

$$n = (\frac{r+1}{r}) \frac{\sigma^2 (Z_\beta + Z_{\alpha/2})^2}{(difference)^2}$$

- For 80% power, Z_{β} =.84
- For 0.05 significance level, Z_{α} =1.96
- r=1 (equal number of cases and controls)
- σ=15.0
- Difference = 7.0
- Answer: n=144 (72 cases, 72 controls)







Centre for Health Informatics, Biostatisics and Epidemiology, Faculty of Medicine, University of Kelaniya

Feedback

- 1. what did you like about this session?
- 2. what didn't you like about this session?
- 3. what did you learn from this session?



Thank you



Centre for Health Informatics, Biostatisics and Epidemiology, Faculty of Medicine, University of Kelaniya