General Biostatistics

Part 8

Common Statistical Methods

Sample Size and Power

Outline

- Sample size for one group
  - precision
  - hypothesis testing
- Sample size for two independent groups
  - hypothesis testing
- Other methods for sample size calculation

Sample Size (SS) for One Group

- Based on precision of estimation: desire confidence interval of a certain width

- Based on hypothesis testing

One Group SS (Estimation)

- Based on *precision* of estimation: desire confidence interval of a certain width
  - Requires:
    - α (significance level)
    - width of interval
    - estimate of variance or proportion

One Group SS- Continuous

- For a continuous variable, the 95% CI is

$$\bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}} = \bar{x} \pm d \qquad n = \frac{1.96^2 \sigma^2}{d^2}$$

- Need to specify d (1/2 the width of the CI)
- Need to estimate σ^2

Example: One Group SS- Continuous

- We are interested in estimating the mean age at cancer diagnosis for a certain group of patients.
  - Suppose we would like to estimate the mean age within ± 2.5 years (95% CI of width 5 years).
  - Suppose that we estimate the population standard deviation as 12 years.

Example: One Group SS- Continuous

- We would need a sample size of 89 patients in order to estimate the mean age at diagnosis to within ± 2.5 years

$$n = \frac{1.96^2 \sigma^2}{d^2} = \frac{1.96^2 12^2}{2.5^2} = 89 \text{ patients}$$

One Group SS - Dichotomous

- For a dichotomous variable, the 95% CI is

$$\hat{p} \pm 1.96 \sqrt{\frac{\hat{p}\hat{q}}{n}} = \hat{p} \pm d \qquad n = \frac{1.96^2 \hat{p}\hat{q}}{d^2}$$

- Need to specify d (1/2 the width of the CI)
- Need to estimate p

Example: One Group SS- Dichotomous

- We are interested in estimating the proportion of patients with sleep disturbance who are symptom-free 18 months after treatment.
  - Suppose we would like to estimate the proportion symptom-free within ± 3 % (95% CI of width 6%).
  - Suppose that we estimate p as 0.20 (based on a previous report in the literature).

Example: One Group SS- Dichotomous

- We would need a sample size of 683 patients in order to estimate the proportion symptom-free to within ± 3%

$$n = \frac{1.96^2 \, pq}{d^2} = \frac{1.96^2 (0.2)(0.8)}{0.03^2} = 683 \text{ patients}$$

Example: One Group SS- Dichotomous

- What sample size is needed to estimate the proportion symptom-free to within ± 5%?

$$n = \frac{1.96^2 \, pq}{d^2} = \frac{1.96^2 (0.2)(0.8)}{0.05^2} = 246 \text{ patients}$$

- What sample size is needed to estimate the proportion symptom-free if p is unknown?

$$n = \frac{1.96^2 \, pq}{d^2} = \frac{1.96^2 (0.5)(0.5)}{0.05^2} = 384 \text{ patients}$$

One Group SS (Hypothesis Testing)

- Sample size for one group based on *hypothesis testing*
  - Requires:
    - α (significance level)
    - 1- β (statistical power)
    - clinically meaningful difference Ho and Ha
    - estimate of variance in the group

Review of Errors

- Type I error = reject a true Ho
- α = the probability of a Type I error
- Type II error = "accept" a false Ho
- β = the probability of a Type II error
- Power = 1 - β = the probability of correctly rejecting Ho

One Group SS: Continuous

- Ho: $\mu = \mu_1$
- Ha: $\mu = \mu_2$
- $\Delta = \mu_1 - \mu_2$

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

One Group SS: Dichotomous

- Ho: p= p_1
- Ha: p= p_2
- $\Delta = p_1 - p_2$

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

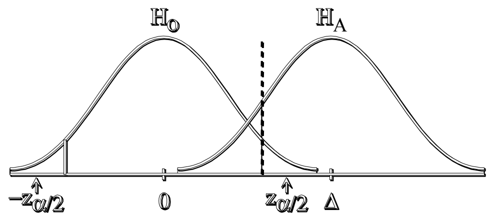
Sample Size for Two Groups

- Sample size for two group hypothesis testing
  - Requires:
    - α (significance level)
    - 1- β (statistical power)
    - clinically meaningful difference between groups
    - estimate of variance in each group

Sample Size for Two Groups

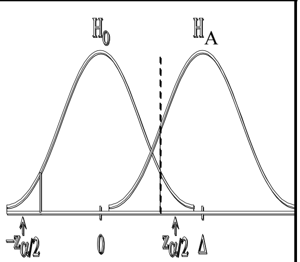
- Continuous outcome:
  - Ho: $\mu_1 - \mu_2 = 0$
  - Ha: $\mu_1 - \mu_2 = \Delta$
- Dichotomous outcome:
  - Ho: $p_1 - p_2 = 0$
  - Ha: $p_1 - p_2 = \Delta$

α (significance level)



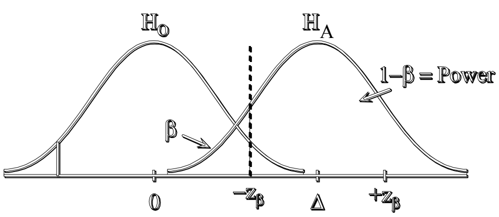
Choices of α

- Z_α may be one-sided or two sided
- Usually α =0.05 and Z_α =1.645, $Z_{\alpha/2}$ = 1.96



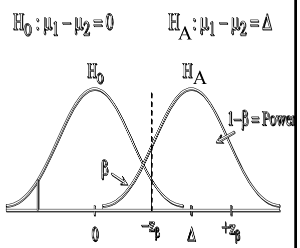
β and Power

$H_0 : \mu_1 - \mu_2 = 0$ $\quad$ $H_A : \mu_1 - \mu_2 = \Delta$



Choices of β

- Z_β is always one-sided
- Usually β=0.20 and Z_β =0.84 or β=0.10 and Z_β =1.28



Sample Size for Two Groups

- Equal sample sizes for 2 groups: continuous outcome

$$n = \frac{(z_{\alpha/2} + z_\beta)^2(\sigma_1^2 + \sigma_2^2)}{\Delta^2}$$

- Equal sample sizes for 2 groups: dichotomous outcome

$$n = \frac{\left(z_{\alpha/2}\sqrt{\overline{p}\,\overline{q}} + z_\beta\sqrt{p_1q_1 + p_2q_2}\right)^2}{\Delta^2}$$

What influences sample size?

- Continuous outcome

$$n = \frac{(z_{\alpha/2} + z_\beta)^2(\sigma_1^2 + \sigma_2^2)}{\Delta^2}$$

  - Estimates of population variances
  - Estimated difference in average outcome
  - α and β

What influences sample size?

- Dichotomous outcome

$$n = \frac{\left(z_{\alpha/2}\sqrt{\bar{p}\bar{q}} + z_{\beta}\sqrt{p_1 q_1 + p_2 q_2}\right)^2}{\Delta^2}$$

  - Estimates of p_1 or p_2
  - Estimated difference in outcome
  - α and β

Example: Two Group SS- Continuous

- We would like to determine the sample sizes required to detect a difference of 5 mm in average blood pressure between individuals receiving placebo versus drug.
  - Assume a significance level of 0.05 and power of 0.80
  - Assume $\sigma_1 = \sigma_2 = 15$ mm
  - Assume equal sample sizes in both groups

Example: Two Group SS- Continuous

- We would need a sample size of 142 patients in each group in order to detect a 5 mm difference in average blood pressure

$$n = \frac{(z_{\alpha/2} + z_{\beta})^2 (\sigma_1^2 + \sigma_2^2)}{\Delta^2}$$

$$n = \frac{(1.96 + 0.84)^2 (15^2 + 15^2)}{5^2} = 142$$

Stata Output

```
. sampsi 0 5, sd1(15) sd2(15) p(0.8)

Estimated sample size for two-sample comparison of means

Test Ho: m1 = m2, where m1 is the mean in population 1
                    and m2 is the mean in population 2
Assumptions:

         alpha =   0.0500  (two-sided)
         power =   0.8000
            m1 =        0
            m2 =        5
           sd1 =       15
           sd2 =       15
         n2/n1 =     1.00

Estimated required sample sizes:

            n1 =      142
            n2 =      142
```


Sample Size Table

(α=0.05, β=0.20)

	Difference of Interest between Groups (Δ)		
Assumed SD	5	10	15
15	142	36	16
20	566	142	63

Example: Two Group SS- Dichotomous

- We would like to determine the sample sizes required to detect a difference in response rate of 10% between individuals receiving new versus standard treatment.
  - Assume a significance level of 0.05 and power of 0.80
  - Assume p_1 =0.25 in the standard treatment
  - Assume equal sample sizes in

Example: Two Group SS- Dichotomous

- We would need a sample size of 207 patients in each group in order to detect a 10% increase with the new treatment.

$$n = \frac{(z_{\alpha/2}\sqrt{\bar{p}\bar{q}} + z_{\beta}\sqrt{p_1 q_1 + p_2 q_2})^2}{\Delta^2}$$

$$n = \frac{(1.96\sqrt{(0.3)(0.7)} + 0.84\sqrt{(0.25)(0.75) + (0.35)(0.65)})^2}{0.10^2}$$

$$n = 207$$

31

Sample Size Considerations

- The best estimated sample size is not just a single calculation
- What sample size is required when the assumptions are varied?
- The estimated sample size should allow for fluctuations in assumptions and our lack of knowledge

32

Sample Size Considerations

- Sample size may need to be inflated for drop-out, treatment cross-over
- Statistical versus practical considerations
  - Cost
  - Feasibility

33

Other Methods for SS Calculation

- Time to event analysis
- Rates and person-time
- Repeated measures over time; correlated observations

Summary

- Sample size calculation for a single group may be based on precision or on hypothesis testing.
- Sample size calculation for two groups is often based on hypothesis testing
  - Ho of no difference between groups
  - The aim is to reject the Ho and conclude that there appears to be a difference