## Class 1 Outline

1. Sample Size Based on Precision (estimation)
2. Sample Size Based on Desired Precision: Continuous Outcome
3. Sample Size Based on Desired Precision: Binary Outcome
4. Sample Size Based on Hypothesis Testing
5. Sample Size : Single Sample
   - Continuous Outcome or Binary Outcome

1

## 1. Sample Size for Purpose of Precision (Estimation)

- An aim of sample size calculation for a single sample is to have a large enough sample with which to estimate a population mean (or difference in means) or proportion (or difference in proportions) within a narrow interval with high reliability.

- Goal: Concerned with the precision of the estimate ("narrowness of the CI")

estimate ± d units

(or consider the width of the interval as w = 2d)

2

## 2.1 Size of Single Sample: Continuous Outcome

Aim : Estimate $\mu$

Want: $\overline{X}$ ± d units where d = Z•SEM

(95% CI of width=2d)

Steps:

1. Specify d (or w = 2d)
2. Use known $\sigma^2$ or estimate using $s^2$
3. $$d = z_{\alpha/2}\sqrt{\sigma^2/n}$$
   $$d^2 = (z_{\alpha/2})^2 \cdot \sigma^2/n$$
   $$n = \frac{(z_{\alpha/2})^2 \cdot \sigma^2}{d^2} = \frac{4(z_{\alpha/2})^2 \cdot \sigma^2}{w^2}$$

3

## 2.2a Example: Sample Size for Estimating a Mean

- Suppose that for a certain group of cancer patients we are interested in estimating the mean age of diagnosis. We would like a 95% confidence interval that is 5 years wide. If the population standard deviation is 12 years, how large should our sample be?

$$n = \frac{z^2 \cdot \sigma^2}{d^2} = \frac{(1.96)^2(144)}{(2.5)^2} = 88.5 \approx 89$$

4

## 2.2b Example: Sample Size for Estimating a Mean

- Suppose that we would like d=1 (half of the width of the CI = 1)
- Then the necessary sample size increases

$$n = \frac{z^2 \cdot \sigma^2}{d^2} = \frac{(1.96)^2 (144)}{(1)^2} = 553.2 \approx 554$$

5

## 2.3 Example: Sample Size for Estimating a Difference in Means

Aim: Estimate $\mu_1$-$\mu_2$

Want $\bar{X}_1 - \bar{X}_2$ within ± d units where d = Z•SE

(95% CI of width=2d)

If equal sample sizes in both groups, then

$$d = z \cdot \sqrt{\frac{\sigma_1^{\ 2}}{n} + \frac{\sigma_2^{\ 2}}{n}}$$

$$n = \frac{z^2 \cdot (\sigma_1^{\ 2} + \sigma_2^{\ 2})}{d^2}$$

Use $\sigma_1^2$, $\sigma_2^2$ or estimate using $s_1^2$and $s_2^2$

6

## 3.1 Size of Single Sample: Binary Outcome

Aim: Estimate p

Want: $\hat{p}$ ± d units where d = Z•SE

(95% CI of width=2d)

Steps:

1. Specify d (or w = 2d)
2. Use estimated p (use p=0.5 if no information)
3. Solve for n

$$d = z_{\alpha/2}\sqrt{\frac{pq}{n}}$$

$$d^2 = (z_{\alpha/2})^2 \cdot \frac{pq}{n}$$

$$n = \frac{(z_{\alpha/2})^2 \cdot pq}{d^2} = \frac{4(z_{\alpha/2})^2 \cdot pq}{w^2},$$

## 3.2a Example: Sample Size for Estimating a Proportion

- Suppose that an epidemiologist would like to determine the proportion of infants in a large rural area in Bangladesh who continue to be breastfed by their mothers beyond 18 months of age. Suppose that in a similar area, the proportion of breastfed infants is reported to be 0.20. What sample size is required to estimate the true proportion to within ± 3 percentage points with 95% confidence? Assume p=0.2, d=0.03, $\alpha$ =0.05

$$n = \frac{z^2 \cdot pq}{d^2} = \frac{(1.96)^2(0.2)(0.8)}{(0.03)^2} = 683$$

8

## 3.2b Example: Sample Size for Estimating a Proportion

- Suppose there is no prior information about the proportion who breastfeed
- Assume p=q=0.5 (most conservative)
- Then the necessary sample size increases

$$n = \frac{z^2 \cdot pq}{d^2} = \frac{(1.96)^2(0.5)(0.5)}{(0.03)^2} = 1{,}068$$

9

## 3.3 Example: Sample Size for Estimating a Difference in Proportions

Aim: Estimate $p_1$-$p_2$

Want $\hat{p}_1 - \hat{p}_2$ within ± d units where d = Z•SE

(95% CI of width = w = 2d)

If equal sample sizes in both groups, then

$$d = z \cdot \sqrt{\frac{p_1 q_1}{n} + \frac{p_2 q_2}{n}}$$

$$n = \frac{z^2 \cdot (p_1 q_1 + p_2 q_2)}{d^2}$$

Use estimates of $p_1$, $p_2$ or (or $p_1$=$p_2$ =0.5 if unknown)

10

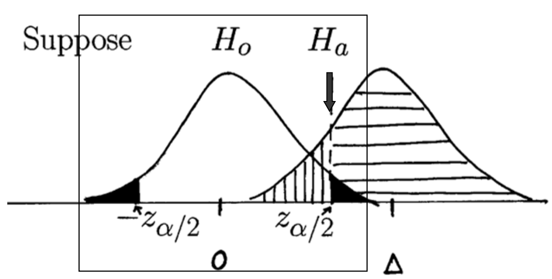
## 4. Sample Size Based on Hypothesis Testing

- Aim: Have large enough samples to detect a difference in population means (or in population proportions)
- Given a difference of interest $\Delta$ which is defined as either:
  - $\Delta = \mu_a - \mu_0$ or $\Delta = p_a - p_0$ for one sample
  - $\Delta = \mu_1 - \mu_2$ or $\Delta = p_1 - p_2$ for two samples
- We would like to maintain low probability of a Type I error ($\alpha$) and low probability of a Type II error ($\beta$) [high power = $1 - \beta$].

11

## 4.1a Review of Hypothesis Testing

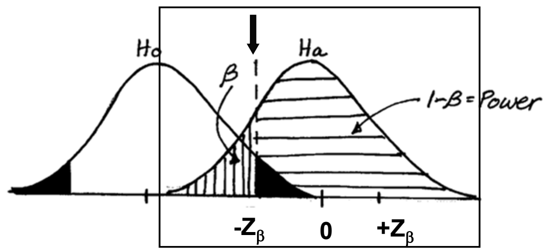
- Define Ho and Ha
- Set $\alpha$
- Reject Ho when $Z_{obs} > Z_{\alpha/2}$ or $Z_{obs} < -Z_{\alpha/2}$ <u>under the assumption that Ho is true</u>




12

## 4.1b Review of Hypothesis Testing

- For a particular Ha, we use the critical value as the cut point for determining β (and power) <u>under the assumption that Ha is true</u>




13

## 5. Sample Size for One Group for Hypothesis Testing

| Population Value | Estimator | Sample Size |
|---|---|---|
| μ | $\overline{X}$ | $n = \dfrac{(z_{\alpha/2} + z_{\beta})^2 \sigma^2}{\Delta^2}$ |
| p | $\hat{p}$ | $n = \left[ \dfrac{z_{\alpha/2}\sqrt{p_0 q_0} + z_{\beta}\sqrt{p_a q_a}}{\Delta} \right]^2$ |

14

## 5.1a Example: Single Sample Size for Test of Population Mean

Suppose we are interested in investigating whether there is an association between familial cardiovascular disease and cholesterol level. Suppose we know that the average cholesterol level in children is 175 mg/dl and $\sigma$ = 50 mg/dl. A study was performed to measure the cholesterol levels of children of men with heart disease.

We are interested in testing:

$H_o$: $\mu \leq 175$

$H_a$: $\mu > 175$

What sample size is needed when $\alpha$ = 0.05 and $1-\beta$= 0.80?

15

## 5.1b Example: Single Sample Size for Test of Population Mean

- We can only calculate statistical power using a specific alternative hypothesized value, e.g. 190.
- Test $H_o$: $\mu \leq 175$ vs. $H_a$: $\mu = 190$ such that $\Delta = 15$
- Then we can calculate that we would need approximately 69 children

$$n = \frac{(z_\alpha + z_\beta)^2 \sigma^2}{\Delta^2} = \frac{(1.645 + 0.84)^2 50^2}{15^2} = 68.6$$

16

## 5.2 Example: Single Sample Sizes for Test of Population Mean: Using Stata

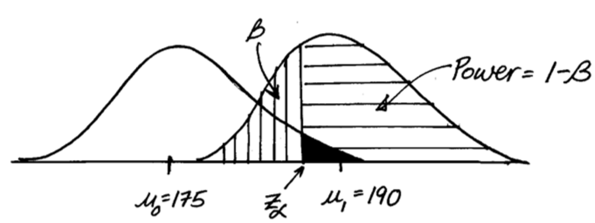
```
. sampsi 175 190, sd(50) onesam onesid a(0.05) p(.8)

Estimated sample size for one-sample comparison of mean
  to hypothesized value

Test Ho: m =    175, where m is the mean in the population

Assumptions:

         alpha =   0.0500  (one-sided)
         power =   0.8000
 alternative m =      190
            sd =       50

Estimated required sample size:

             n =       69
```

17

## 5.3a Derivation of Sample Size Formula

- How is this sample size formula derived?

$H_o$ $H_a$




18

## 5.3b Derivation of Sample Size Formula

- Recall under the assumption that $H_o$ is true

We reject $H_o$ if $z_{obs} \geq z_\alpha$

or if $\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} \geq z_\alpha$

- Calculate power under the assumption that $H_a$ is true ($H_o$ is false)

19

## 5.3c Calculation of Power

$$\begin{aligned}
\text{Power} &= 1-\beta \\
&= 1 - P(\text{accept } H_o | H_o \text{ false}) \\
&= P\ (\text{reject } H_o | H_o \text{ false}) \\
&= P(z_{obs} \geq z_\alpha | \mu = \mu_1) \\
&= P\left(\frac{\overline{X}-\mu_o}{\sigma/\sqrt{n}} \geq z_\alpha | \mu = \mu_1\right) \\
&= P\left(\overline{X} \geq \mu_0 + z_\alpha \cdot \sigma/\sqrt{n} | \mu = \mu_1\right) \\
&= P\left(\frac{\overline{X}-\mu_1}{\sigma/\sqrt{n}} \geq \frac{\mu_0}{\sigma/\sqrt{n}} + z_\alpha - \frac{\mu_1}{\sigma/\sqrt{n}}\right) \\
&= P\left(z \geq z_\alpha + \frac{\mu_0-\mu_1}{\sigma/\sqrt{n}}\right)
\end{aligned}$$

20

## 5.3d Solving for Sample Size

$$-Z_{\beta} \geq Z_{\alpha} + \frac{(\mu_0 - \mu_1)}{\frac{\sigma}{\sqrt{n}}}$$

$$\frac{-(\mu_0 - \mu_1)}{\frac{\sigma}{\sqrt{n}}} \geq Z_{\alpha} + Z_{\beta}$$

$$\frac{(\mu_0 - \mu_1)^2}{\frac{\sigma^2}{n}} \geq (Z_{\alpha} + Z_{\beta})^2$$

$$n \geq \frac{(Z_{\alpha} + Z_{\beta})^2 \sigma^2}{(\mu_0 - \mu_1)^2}$$

21

## 5.4 Calculating Power for a Fixed Sample Size n=30

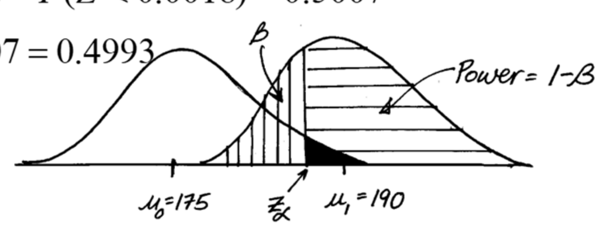
$$n = \frac{(z_{\alpha} + z_{\beta})^2 \sigma^2}{\Delta^2} \Rightarrow$$

$$z_{\beta} = \sqrt{\frac{n\Delta^2}{\sigma^2}} - z_{\alpha} = \sqrt{\frac{30(15)^2}{50^2}} - 1.645$$

$$z_{\beta} = 1.6432 - 1.645 = -0.0018$$

$$\beta = P(Z < -z_{\beta}) = P(Z < 0.0018) = 0.5007$$

$$1 - \beta = 1 - 0.5007 = 0.4993$$




22

# 5.5 Calculating Power for a Fixed Sample Size n=30: Using Stata

```
. sampsi 175 190, sd(50) onesam onesid a(0.05) n(30)

Estimated power for one-sample comparison of mean
  to hypothesized value

Test Ho: m =    175, where m is the mean in the population

Assumptions:
        alpha =   0.0500  (one-sided)
 alternative m =      190
           sd =       50
 sample size n =       30

Estimated power:
        power =   0.4993
```

23

# 5.6a Example: Single Sample Size for Test of Population Proportion

Suppose we are interested in investigating the incidence of disease in a certain rural area. Suppose we know that the incidence of disease in a similar area was 0.15. A study was performed to measure disease incidence and we are interested in knowing whether the incidence in our study area is the same as that of the similar area.

We are interested in testing:

$H_o$: p = 0.15

$H_a$: $p \neq 0.15$

What sample size is needed when $\alpha = 0.05$ and $1-\beta = 0.80$?

24

## 5.6b Example: Single Sample Size for Test of Population Proportion

- We can only calculate statistical power using a specific alternative hypothesized value, e.g. 0.25
- Test $H_o$: p=0.15 vs. $H_a$: p=0.25 such that $\Delta$ = 0.10
- Then we can calculate that we would need approximately 114 individuals

$$n = \left[ \frac{z_{\alpha/2}\sqrt{p_0 q_0} + z_\beta \sqrt{p_a q_a}}{\Delta} \right]^2$$

$$= \left[ \frac{(1.96\sqrt{(0.15)(0.85)} + 0.84\sqrt{(0.25(0.75)}}{(0.10)} \right]^2 = 113.1$$

25

## 5.7 Example: Single Sample Sizes for Test of Proportion: Using Stata

```
. sampsi 0.15 0.25, onesam a(0.05) p(.8)

Estimated sample size for one-sample comparison of proportion
  to hypothesized value

Test Ho: p = 0.1500, where p is the proportion in the population

Assumptions:

         alpha =   0.0500  (two-sided)
         power =   0.8000
 alternative p =   0.2500

Estimated required sample size:

             n =     114
```

26

# 6. Summary

- Sample size can be calculated for one or two groups for purposes of:
  - Precision (requires desired width of CI and specification of $\alpha$)
  - Hypothesis testing (requires null hypothesis and specific alternative hypothesis [to calculate $\Delta$] and specification of assumed $\alpha,\beta$, variance)

27