Statistical Inference Theory
Lesson 39
The Distribution of Differences of Sample Proportions

39.1-What is the Central Limit Theorem for $\overline{P}_d = \overline{P}_1 - \overline{P}_2$?

39.1 - Problem 1:

➤(a).

$$\overline{P}_d = \overline{P}_1 - \overline{P}_2 = 0.65 - 0.41 = 0.24$$

➤(b).

$$p_d = p_1 - p_2 = 0.5 - 0.5 = 0$$

➤(c).

$$\sigma_{\overline{P}_d} = \sqrt{\frac{(0.65)(0.35)}{100} + \frac{(0.41)(0.59)}{100}} \approx 0.07$$

➤(d).

$$E = \pm(\overline{P}_d - p_d) = \pm 0.24$$

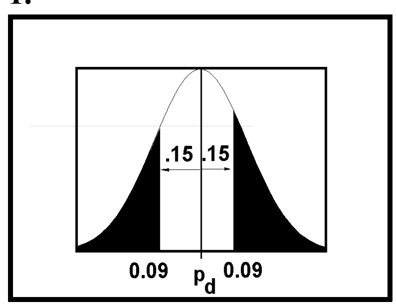
➤(e).

$$z = \frac{\overline{P}_d - p_d}{\sigma_{\overline{P}_d}} = \frac{0.24}{0.07} \approx 3.43$$

39.1 - Problem 2:

$$e* = \pm(\overline{P}_d - p_d) = \pm(0.09 - p_d) = \pm z\,\sigma_{\overline{P}_d} = \pm 0.15$$

1.



$$\sigma_{\overline{P}_d} = \sqrt{\frac{\overline{P}_1(1 - \overline{P}_1)}{N_1} + \frac{\overline{P}_2(1 - \overline{P}_2)}{N_2}} = \sqrt{\frac{(0.72)(1 - 0.72)}{49} + \frac{(0.81)(1 - 0.81)}{49}} \approx 0.085$$

$$z = \pm\frac{0.15}{0.085} \approx 1.76$$

From the normal distribution table, for z = 1.76, we find 0.4608.

Therefore, the probability that $\overline{P}_d$ = 0.09 exceeds p_d by more than 0.15 is

2(0.5 - 0.4608) ≈ 0.08.

39.2 - Statistical Decision Theory

39.2 - Problem 1:

➤(a).

Since the challenge is that there is a difference,

H_o: $p_d = 0$

H_a: $p_d \neq 0$.

➤(b).

From the null and alternative hypothesis we state the decision rule as:

If $-c^* \leq \overline{P}_d \leq c^*$ then reject H_a ; otherwise reject H_o and accept H_a .

$$\sigma_{\overline{P}_d} = \sqrt{\frac{\overline{P}_1(1 - \overline{P}_1)}{N_1} + \frac{\overline{P}_2(1 - \overline{P}_2)}{N_2}} = \sqrt{\frac{(0.07)(1 - 0.93)}{400} + \frac{(0.04)(1 - 0.96)}{400}} \approx 0.02$$

$c^* = z\sigma_d = 0.02z$

From the normal distribution table, z = 2.57 for the area 0.5 - (0.01)/2 = 0.495. Therefore,

$c^* = 0.02(2.57) \approx 0.05$.

Therefore, the decision rule is

If $-0.05 \leq \overline{P}_d \leq 0.05$ then reject H_a; otherwise reject H_0 and accept H_a .

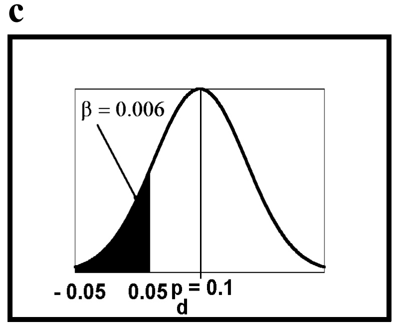
➤(c).

Since $p_d = 0.10$,

$$z = \frac{-0.05 - 0.10}{0.02} = -7.5$$

$$z = \frac{0.05 - 0.10}{0.02} = -2.5$$

c



Therefore, $P\{-0.05 \leq \overline{P}_d \leq 0.05\} = P\{-7.5 \leq Z \leq -2.5\} = 0.5 - 0.4938 \approx 0.006$.

➤(d).

$$\overline{P}_d = \overline{P}_1 - \overline{P}_2 = 0.07 - 0.04 = 0.03$$

Since $-0.05 \leq 0.03 \leq 0.05$, we reject H_a .

➤(e).

The study shows that there is no significant difference in the percentage of students that major in English and Economics as far as having a GPA of 3.5 or more.

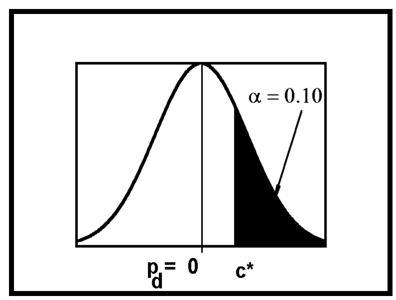
39.2 - Problem 2:

➤(a).

Since the claim is that a higher percentage of adult Catholics believe in birth control than Protestants, we have

H_o: $p_d = 0$
H_a: $p_d > 0$.

b



➤(b).

$c^* \leq \overline{P}_d$ than reject H_0 and accept H_a; otherwise reject H_a.

From the normal distribution, z = 1.28 for the area 0.5 - 0.10 = 0.4.

$$c^* = z\sigma_{\overline{P}_d}$$

$$\sigma_{\overline{P}_d} = \sqrt{\frac{\overline{P}_1(1 - \overline{P}_1)}{N_1} + \frac{\overline{P}_2(1 - \overline{P}_2)}{N_2}} = \sqrt{\frac{(0.71)(1 - 0.71)}{500} + \frac{(0.63)(1 - 0.63)}{200}} \approx 0.04$$

$$c^* = z\sigma_{\overline{P}_d} = 1.28(0.04) \approx 0.05$$

Therefore, the decision rule is

$0.05 \leq \overline{P}_d$ than reject H_o and accept H_a; otherwise reject H_a .

➤(c).

Since, $p_d = 0.05$ and decision rule, we have $\beta = P\{\overline{P}_d \geq 0.05\} = 0.5$.

➤(d).

$$\overline{P}_d = \overline{P}_1 - \overline{P}_2 = 0.071 - 0.63 = 0.08$$

Since $0.08 \geq 0.05$, we reject H_0 and accept H_a.

➤(e).

A significant higher percentage of Catholics support birth control as compared to Protestants in this country.

Supplementary Problems

1.

The confidence interval formula is

$$\overline{P}_d - z\sigma_{\overline{P}_d} \leq \mu_d \leq \overline{P}_d + z\sigma_{\overline{P}_d} \text{ where}$$

$$\sigma_{\overline{P}_d} = \sqrt{\frac{\overline{P}_1(1 - \overline{P}_1)}{N_1} + \frac{\overline{P}_2(1 - \overline{P}_2)}{N_2}} = \sqrt{\frac{(0.05)(1 - 0.05)}{100} + \frac{(0.03)(1 - 0.03)}{100}} \approx 0.03$$

and

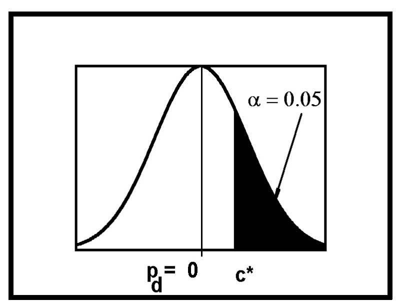
$$\overline{P}_d = 0.05 - 0.03 = 0.02 .$$

For a 95% confidence interval, we have $z = 1.96$.

$0.02 - 1.96(0.03) \leq p_d \leq 0.02 + 1.96(0.03)$

Therefore, the 95% confidence interval is $-0.04 \leq p_d \leq 0.08$.

b



2.

➤a.

Since we wish to see if the medication from company B lowers the blood pressure for a significantly number of people, we have $p_d = p_2 - p_1$ and

H_o: $p_d = 0$

H_a: $p_d > 0$

➤b.

D.R. If $\overline{P}_d = \overline{P}_2 - \overline{P}_1 \geq c^*$ *than reject* H_0 *and accept* H_a*; otherwise reject* H_a.

$$\sigma_{\overline{P}_d} = \sqrt{\frac{\overline{P}_1(1 - \overline{P}_1)}{N_1} + \frac{\overline{P}_2(1 - \overline{P}_2)}{N_2}} = \sqrt{\frac{(0.15)(1 - 0.15)}{250} + \frac{(0.29)(1 - 0.29)}{250}} \approx 0.04$$

$c^* = 0 + z0.04 = 1.64(0.04) \approx 0.07$.

Therefore, the decision rule is read:

D.R. If $\overline{P}_d = \overline{P}_2 - \overline{P}_1 \geq 0.07$ *than reject* H_0 *and accept* H_a*; otherwise reject* H_a..

➤c.

$\overline{P}_d = \overline{P}_2 - \overline{P}_1 = 0.29 - 0.15 = 0.14 > 0.07$, reject H_o and conclude that company's B medication

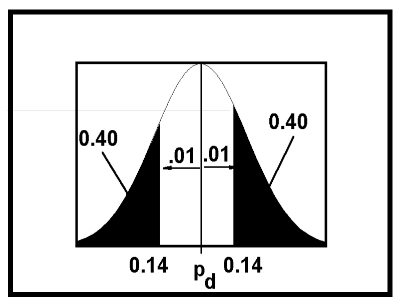
is effective for a significant number of people suffering from high blood pressure .

➤d. *If conclude that* $p_d = 0.14$*, what is the probability that the error is greater than 1%?*

$$e^* = \pm z\sigma_{\overline{P}_d} = \pm z(0.04) = 0.01$$

$z = \pm 0.25$.

d



From the normal distribution table, the area is 0.5 - 0.0987 = 0.4013.

Since the error can occur on two side,s, the probability that the error is greater than 1% is approximately 0.8.

➤e.

$$\sigma_{\overline{P}_d} = \sqrt{\frac{\overline{P}_1(1 - \overline{P}_1)}{N} + \frac{\overline{P}_2(1 - \overline{P}_2)}{N}} = \sqrt{\frac{(0.15)(1 - 0.15)}{N} + \frac{(0.29)(1 - 0.29)}{N}} \approx \frac{0.6}{\sqrt{N}}$$

$$e^* = \pm \frac{0.6}{\sqrt{N}} z = \pm 0.01$$

Since we want the probability of such an error to be 0.05/2 + 0.05/2, z = 1.96.

$$e^* = \pm \frac{0.6}{\sqrt{N}}(1.96) \approx \pm \frac{1.176}{\sqrt{N}} = \pm 0.01$$

Solving for N we have N = 13,830.

➤f.
We use the above decision rule:

D.R. If $\overline{P}_d = \overline{P}_2 - \overline{P}_1 \geq 0.07$ *than reject* H_0 *and accept* H_a*; otherwise reject* H_a.

$$\sigma_{\overline{P}_d} = \sqrt{\frac{P_1(1 - P_1)}{N} + \frac{P_2(1 - P_2)}{250}} = \sqrt{\frac{(0.20)(1 - 0.20)}{250} + \frac{(0.35)(1 - 0.35)}{250}} \approx 0.04$$

$\beta = P\{\overline{P}_d < 0.07\}$

$$z = \frac{0.07 - 0.15}{0.04} = -2$$

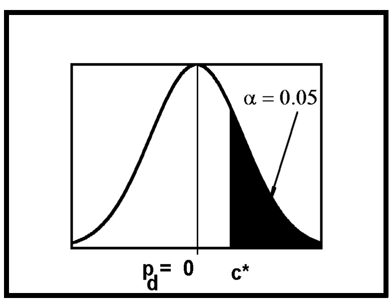
$\beta = P\{\overline{P}_d < 0.07\} = P\{Z < -2\} = 0.5 - 0.4772 = 0.0228.$

3.
➤a. State H_o and H_a
$p_d = p_2 - p_1$

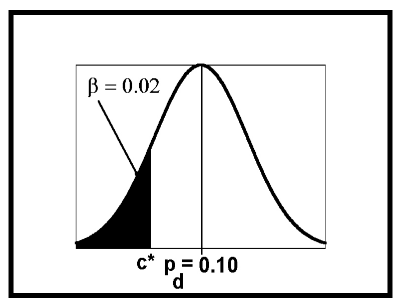
H_o: $p_d = 0$

H_a: $p_d > 0.$

b



b



➤b.
D.R. Take a sample of size N. If $c^* \leq \overline{P}_d$ *than reject* H_0 *and accept* H_a *; otherwise reject* H_a .

$$c^* = z\sigma_{\overline{P}_d}$$

$$\sigma_{\overline{P}_d} = \sqrt{\frac{P_1(1-P_1)}{N} + \frac{P_2(1-P_2)}{N}} = \sqrt{\frac{(0.07)(1-0.07)}{N} + \frac{(0.07)(1-0.07)}{N}} \approx \frac{0.36}{\sqrt{N}}$$

For $\alpha = 0.05$, $z = 1.64$,

$$c^* = z\sigma_{\overline{P}_d} = (1.64)\frac{0.36}{\sqrt{N}} \approx \frac{0.59}{\sqrt{N}}, \text{ equation 1.}$$

For $\beta = 0.02$ when $p_d = 0.10$ we have $P\{\overline{P}_d \le c^*\} = 0.02$.

For the area 0.5 - 0.02 = 0.48, $z = -2.05$.

$$\sigma_{\overline{P}_d} = \sqrt{\frac{P_1(1-P_1)}{N} + \frac{P_2(1-P_2)}{N}} = \sqrt{\frac{(0.15)(1-0.15)}{N} + \frac{(0.05)(1-0.05)}{N}} \approx \frac{0.42}{\sqrt{N}}$$

$$c^* = 0.10 + z\sigma_{\overline{P}_d} = 0.10 - (2.05)\frac{0.42}{\sqrt{N}} \approx 0.10 - \frac{0.86}{\sqrt{N}}, \text{ equation 2.}$$

Setting equation 1 and equation 2 equal, we need to solve N.

$$\frac{0.59}{\sqrt{N}} \approx 0.10 - \frac{0.86}{\sqrt{N}}$$

$$\frac{0.59}{\sqrt{N}} + \frac{0.86}{\sqrt{N}} = 0.10$$

$$\sqrt{N} = \frac{1.45}{0.10} = 14.5$$

$N \approx 210$

From equation 1, $c^* = \frac{0.59}{\sqrt{N}} = \frac{0.59}{14.5} \approx 0.04$.

Therefore, the decision rule becomes:

D.R. Take a sample of size N = 210. If 0.04 ≤ $\overline{P}_d$ *than reject* H_0 *and accept* H_a *; otherwise reject* H_a

4.

We can write $\overline{X}_d = \overline{X}_1 - \overline{X}_2 = \overline{X}_1 + -\overline{X}_2$

$\sigma^2(-X) = E[(-X)^2] - [E(-X)]^2 = E[(X)^2] - [E(X)]^2 = \sigma^2(X)$

From Lesson 16, since $\overline{X}_1$, $\overline{X}_2$ are independent,

$$\sigma^2(\overline{P}_d) = \sigma^2(\overline{P}_1 - \overline{P}_2) = \sigma^2(\overline{P}_1) + \sigma^2(\overline{P}_2) = p_1(1 - p_1)/N + p_2(1 - p_2)/N .$$

5.

Jack: $p_1 = 0.50$

Jill: $p_2 = 0.55$

$p_d = p_1 - p_2 = 0.50 - 0.55 = -0.05$

$$\sigma(\overline{P}_d) = \sqrt{\frac{(0.55)(0.45)}{100} + \frac{(0.5)(0.5)}{100}} \approx 0.07$$

Jack will win if $\overline{P}_d = \overline{P}_1 - \overline{P}_2 > 0$.

$$z = \frac{0 - 0.05}{0.07} = -0.71$$

$P\{0 \le z \le 0.71\} = 0.2612$

$P(\overline{P}_d \ge 0) \approx 0.5 - 0.262 = 0.238$