Statistical Inference Theory
Lesson 38
The Distribution of Differences of Sample Means

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

38.1 - Problem 1:

➤(a).

From the table, $\overline{X}_d = \overline{X}_1 - \overline{X}_2$ = 79.45 - 70.13 = 9.32 degrees.

➤(b).

From the table, $\mu_d = \mu_1 - \mu_2$ = 76.5 - 72.12 = 4.38 degrees

➤(c).

$$\sigma_d = \sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}} = \sqrt{\frac{12.13^2}{36} + \frac{5.9^2}{36}} \approx 2.25 \text{ degrees}$$

➤(d).

$E = \pm(\overline{X}_d - \mu_d)$ = ± (9.32 - 4.38) = ± 4.94

➤(e).

$$z = \frac{\overline{X}_d - \mu_d}{\sigma_d} = \frac{9.32 - 4.38}{2.25} \approx 2.2$$

38.2 - Problem 2:

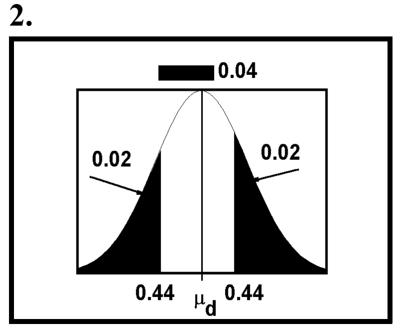
$e^* = \pm(\overline{X}_d - \mu_d) = \pm z\sigma_d = \pm 0.50.$

$$\sigma_d = \sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}} = \sqrt{\frac{1.2^2}{100} + \frac{2.12^2}{100}} \approx 0.24$$

(0.24)z = ±0.50

Solving for z ≈ ±2.08.

2.



From the normal distribution table, the area for z = 2.08 is 0.4812.

Therefore, $P\{(\overline{X}_d - \mu_d) \geq 0.5\} = 1 - 0.4812 - 0.4812 \approx 0.04$.

38.2 - Statistical Decision Theory

38.2 - Problem 1:

➤(a).

Since the administrator claims that there is a difference between the two majors:

H_o: $\mu_d = 0$
H_a: $\mu_d \neq 0$

➤(b).

D.R. If $-e* \leq \overline{X}_d \leq e*$ *then reject* H_a*; otherwise, reject* H_o *and accept* H_a*.*

$$\sigma_d = \sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}} = \sqrt{\frac{0.3^2}{400} + \frac{0.5^2}{400}} \approx 0.03$$

$e* = z\sigma_d = 0.03z$

Since $\alpha = 0.02$, from the normal distribution table, for the area 0.5 - 0.02/2 = 0.49: z = 2.33 .

Therefore, $e* \approx 0.07$.

The decision rule now reads:

If $-0.07 \leq \overline{X}_d \leq 0.07$ *then reject* H_a*; otherwise reject* H_0 *and accept* H_a .

➤(c).

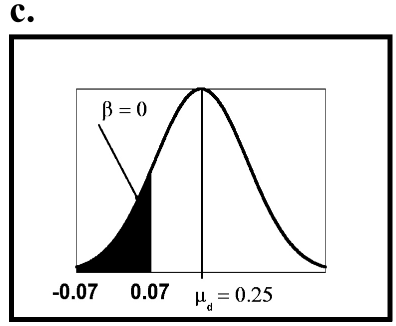
$$z = \frac{\overline{X}_d - \mu_d}{\sigma_d}$$

$$z = \frac{0.07 - 0.25}{0.03} \approx -6$$

$$z = \frac{-0.07 - 0.25}{0.03} \approx -10.67$$

Therefore, $\beta = 0$

c.



➤(d).

$\overline{X}_d = \overline{X}_1 - \overline{X}_2 = 3.1 = 2.9 = 0.2.$

Since $\overline{X}_d = 0.2 \geq 0.07$, we reject H_0 and accept H_a.

➤(e).
Since we reject H_o and accept H_a, we conclude there is a significant difference between the grade point average of Physics and Math majors.

38.2 - Problem 2:
➤(a).
μ_1 : the average price for foreign lap top computers.
μ_2: the average price for domestic lap top computers.
$\mu_d = \mu_1 - \mu_2$

We test the null hypothesis $\mu_d = 0$ against the alternative hypothesis $\mu_d < 0$:

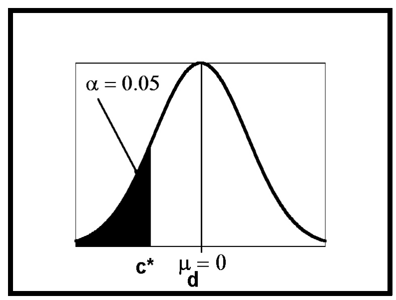
$H_o : \mu_d = 0$
$H_a : \mu_d < 0$

➤(b).

Since we are testing on one side, we state the decision rule as: If $\overline{X}_d < c^* < 0$ then reject H_0 and accept H_a; otherwise reject H_a.

$$\sigma_d = \sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}} = \sqrt{\frac{100.75^2}{400} + \frac{120.76^2}{400}} \approx 7.86$$

b.



$c^* = \mu_d + z\sigma_d = 0 + 7.86z$

For $\alpha = 0.05$, from the normal distribution table, for the area $0.5 - 0.05 = 0.45$, $z = -1.64$.

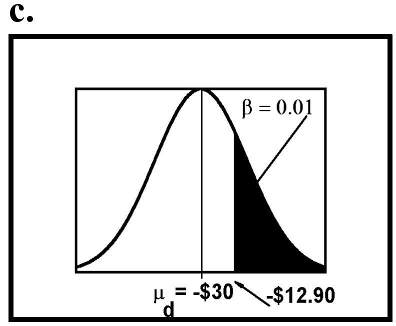
$c^* = 7.86(-1.64) \approx -12.90$

Therefore, the decision rule is:

If $\overline{X}_d < -\$12.90$ then reject H_0 and accept H_a; otherwise reject H_a.

➤(c).
For $\mu_d = -\$20$, we need to find $\beta = P\{\overline{X} \geq -\$12.90\}$.

c.



$$z = \frac{\overline{X}_d - \mu_d}{\sigma_d} = \frac{-12.90 - -30}{7.86} \approx 2.18$$

From the distribution table, $P\{z \geq 2.18\} = 0.5 - 0.4854 = 0.0146$.

Therefore, $\beta \approx 0.01$.

➤(d).

$\overline{X}_d = \overline{X}_1 - \overline{X}_2 = \$1,989 - \$2,110 = -\$121.$

Since -\$121 < -\$12.90, we reject H_o and accept H_a.

➤(e).

The foreign price is significantly lower than the domestic prices for lap top computers.

Supplementary Problems

1.

$\overline{X}_1$: The average weight of the fuses produced from machine A.

$\overline{X}_2$: The average weight of the fuses produced from machine B.

$100\overline{X}_1$: The total weight of the fuses produced from machine A.

$100\overline{X}_2$: The total weight of the fuses produced from machine B.

$100\overline{X}_2 - 100\overline{X}_1 > 16$ ounces.

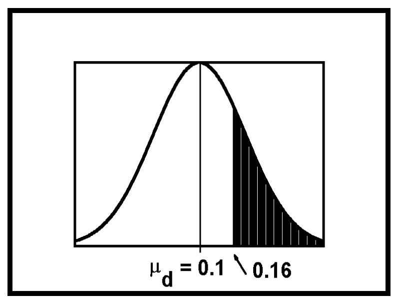
$\overline{X}_2 - \overline{X}_1 > 16/100 = 0.16$

$\overline{X}_d = X_2 - X_1 > 0.16$

We need to find the $P\{\overline{X}_d > 0.16\}$.

$\mu_d = \mu_2 - \mu_1 = 0.6 - 0.5 = 0.10$

1.



$$\sigma_d = \sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}} = \sqrt{\frac{0.04^2}{100} + \frac{0.05^2}{100}} \approx 0.006$$

$$z = \frac{\overline{X}_d - \mu_d}{\sigma_d} = \frac{0.16 - 0.10}{0.006} = 10$$

Since z = 10 and off the normal distribution table, $P\{\overline{X}_d > 0.16\} = P\{Z > 10\} = 0$.

2.

$\overline{S}_1$: The total number of heads tossed by Bill.

$\overline{S}_2$: The total number of heads tossed by Jim.

$$\overline{S}_d = \overline{S}_1 - \overline{S}_2$$

Since tossing of the fair coin 100 times, we have a binomial distribution. Therefore,

$\sigma^2 = Npq = 100(0.5)(0.5) = 25.$

First, we find $P\{\overline{S}_d > 0\}$, the probability that Bill wins \$100.

$$\sigma_d = \sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}} = \sqrt{\frac{25}{100} + \frac{25}{100}} \approx 0.71$$

$$Z = \frac{S_d - \mu_d}{\sigma_d} = \frac{1 - 0}{0.71} \approx 1.41$$

$P\{\overline{S}_d > 0\} = P\{Z \geq 1.41\} = 0.5 - 0.4207 = 0.0793$, the probability that Bill will win \$100.

Therefore, Bill's expected winning on for this game is \$100(0.0793) - \$10(0.9207) ≈ -\$1.28 .

3.

$$\overline{X}_d - z\sigma_d \leq \mu_d \leq \overline{X}_d + z\sigma_d$$

$\overline{X}_d = 35{,}000 - 33{,}500 = 1{,}500$

$$\sigma_d = \sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}} = \sqrt{\frac{3100^2}{1000} + \frac{2990^2}{1000}} \approx 136.20$$

For a confidence of 95%, we find the z value for the area 0.95/2 = 0.475: z = 1.96.

$1{,}500 - 1.96(136.20) \leq \mu_d \leq 1{,}500 + 1.96(136.20)$

$1233 \leq \mu_d \leq 1767$

4.

➤a.

The claim is that there is no difference in price-earnings ratio. The counter-claim is that there is a difference.

H_o: $\mu_d = 0$

H_a: $\mu_d \neq 0$

➤b.

$\overline{X}_d = 16.089 - 14.666 = 1.423$

$$\sigma_d = \sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}} = \sqrt{\frac{7.374^2}{44} + \frac{7.295^2}{44}} \approx 1.56$$

$$z = \frac{1.423 - 0}{1.56} \approx 0.91$$

For the area 0.5 - 0.01/2 =0.5 - 0.005 = 0.495, from the normal distribution table: z = 2.56 .

Since z = 0.91 < 1.56, there is no significant difference between the two companies.

➤c.

$$\overline{X}_d - z\sigma_d \le \mu_d \le \overline{X}_d + z\sigma_d$$

$\overline{X}_d = 1.423$

For a confidence of 95%, we find the z value for the area 0.95/2 = 0.475: z = 1.96.

$1.423 - 1.96(1.56) \le \mu_d \le 1.423 + 1.96(1.56)$

$-1.63 \le \mu_d \le 4.48$

5.

➤a.

Since the counter-claim is that the grade point average of female students is higher than male students:

H_o: $\mu_d = 0$, (there is no difference in the grade point averages)

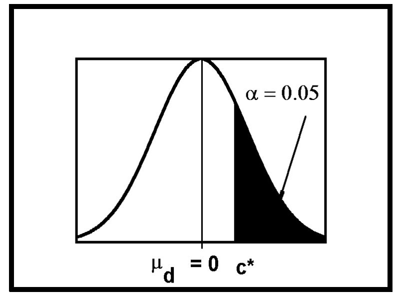
H_a: $\mu_d > 0$

➤b.

Decision Rule: Take a sample of size N. If $\overline{X} \geq c^*$ then accept Ms. Floss's claim; otherwise reject her claim and assume there is no difference in grade points or reserve judgement.

$$c^* = z\,\sigma_d = z\sqrt{\frac{0.38^2}{N} + \frac{0.35^2}{N}} \approx z\frac{0.52}{\sqrt{N}}$$

5.



For $\alpha = 0.05$, the area is 0.5 - 0.05 = 0.45 and from the normal distribution table: z = 1.64.

$$c^* = \frac{(1.64)(0.52)}{\sqrt{N}} \approx \frac{0.85}{\sqrt{N}}\text{, equation 1.}$$

For $\beta = 0.02$ and $\mu_d = 0.1$

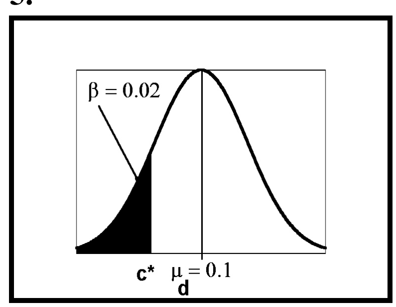
$$c^* = \mu_d + z\sigma_d = 0.1 + z\sigma_d = z\sqrt{\frac{0.38^2}{N} + \frac{0.35^2}{N}} \approx 0.1 + z\frac{0.52}{\sqrt{N}}\text{, equation 2 .}$$

For $\beta = 0.02$, the area is 0.5 - 0.02 = 0.48 and from the normal distribution table: z = -2.05.

$$c^* = \approx 0.1 - (2.05)\frac{0.52}{\sqrt{N}} = 0.1 - \frac{1.066}{\sqrt{N}}\text{, equation 2.}$$

Setting equation 1 and equation 2 equal,

5.



$$0.1 - \frac{1.066}{\sqrt{N}} = \frac{0.85}{\sqrt{N}}.$$

$$0.1 = \frac{1.916}{\sqrt{N}}$$

$N \approx 367$

$$c^* = \frac{0.85}{19.16} \approx 0.04$$

Decision Rule: Take a sample of size N = 367. If $\overline{X} \geq 0.04$ *then accept Ms. Floss's claim; otherwise reject her claim and assume there is no difference in grade points or reserve judgement.*

6.

➤a.

Let $\mu_d = \mu_w - \mu_m$, where μ_w is the average age for deceased women insured by the company and μ_m is the average age for deceased men insured by the company.

Therefore,

H_o: $\mu_d = 4$
H_a: $\mu_d < 4$

➤b.

$$\sigma_d = \sqrt{\frac{7.21^2}{1000} + \frac{10.59^2}{1000}} \approx 0.41$$

$\overline{X}_d = 81.45 - 78.79 = 2.66$

$$z = \frac{\overline{X}_d - \mu_d}{\sigma_d} = \frac{2.66 - 4}{0.41} = -3.21$$

For $\alpha = 0.05$, we have for the area 0.5 - 0.05 = 0.45: z = -1.64.

Since -3.21 < -1.64, reject H_o and accept H_a. We would conclude that the deceased insured women live less than 4 years longer than the deceased insured men.

➤c.

$$\overline{X}_d - z\sigma_d \le \mu_d \le \overline{X}_d + z\sigma_d$$

$$2.66 - z(0.41) \le \mu_d \le 2.66 + z(0.41)$$

For a 95% confidence interval, we find z for the area 0.95/2 = 0.475: z = 1.96.

$$2.66 - 1.96(0.41) \le \mu_d \le 2.66 + 1.96(0.41)$$

$$1.86 \le \mu_d \le 3.46$$

7.

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, $\sigma^2(\overline{X}_d) = \sigma^2(\overline{X}_1 - \overline{X}_2) =$
$\sigma^2(\overline{X}_1) + \sigma^2(\overline{X}_2)$