Statistical Inference Theory
Lesson 36
Estimating the Proportion p of a Population

36.1- What is the error created when using a point estimate ?

36.1 - Problem 1:

➤(a).

Here $\overline{P}$ = 0.15 and N = 49.

The difference between $\overline{P}$ and p is the error $(\overline{P} - p)$. We need to find the probability that the error exceeds 10%.

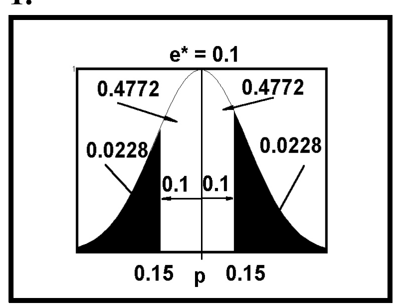
Step 1: Since $\overline{P}$ = 0.15 and N = 49, we compute the standard deviation of the sample

$$\sqrt{\frac{\overline{P}(1-\overline{P})}{N}} = \sqrt{\frac{(0.15)(0.85)}{49}} \approx 0.05$$

Step 2: $e^* = \overline{P} - p = \pm z\sqrt{\frac{\overline{P}(1-\overline{P})}{N}} = \pm z(0.05) = \pm 0.10.$

Step 3:Solving for z gives $z = \pm\frac{0.10}{0.05} = \pm 2$

1.



Step 4: From the normal distribution table for z = 2:

P{e* > 0.10} = 1 -0.4772 - 0.4772 ≈ 0.05

➤(b).

Step 1: $e^* = \overline{P} - p = \pm z\sqrt{\frac{0.15(0.85)}{N}} = 0.10$

Step 2: Since the probability of this error is to be 0.01, we have 0.01/2 = 0.005 .

From the normal distribution table for 0.5 - 0.005 = 0.495, z = 2.57 .

Step 3: $2.57\sqrt{\dfrac{0.15(0.85)}{N}} = 0.10$

$$\sqrt{\frac{0.15(0.85)}{N}} = \frac{0.10}{2.57}$$

$$\sqrt{\frac{N}{0.1275}} = \frac{2.57}{0.10}$$

$N \approx 84$

36.2 - What is the error created when using a Confidence Interval estimate ?

36.2 - Problem 1:

➤(a).

Step 1: Here $\overline{P} = 0.15$

Step 2: $N = 49$

Step 3: $e* = \pm z\sqrt{\dfrac{\overline{p}(1 - \overline{p})}{N}} = \pm z\sqrt{\dfrac{(0.15)(0.85)}{49}} \approx \pm z0.05$

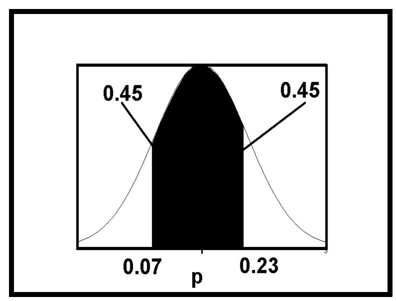
Step 4: $0.15 - z(0.05) \leq p \leq 0.15 + z(0.05)$

Step 5: Since the confidence interval is 0.9, z = 1.64 and

$$0.15 - 1.64(0.05) \leq p \leq 0.15 + 1.64(0.05)$$

Step 6: $0.07 \leq p \leq 0.23$

a

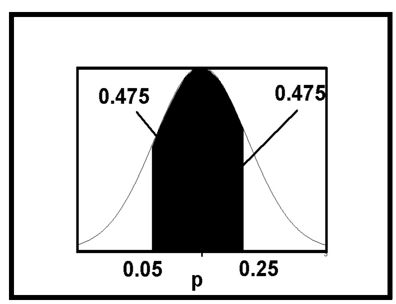


➤(b).

Step 1: Here $\overline{P} = 0.15$

Step 2: $N = 49$

b



Step 3: $e* = \pm z\sqrt{\dfrac{\overline{p}(1 - \overline{p})}{N}} = \pm z\sqrt{\dfrac{(0.15)(0.85)}{49}} \approx \pm z0.05$

Step 4: $0.15 - z(0.05) \le p \le 0.15 + z(0.05)$

Step 5:Since the confidence interval is 0.95/2 = 0.475, z = 1.96

and $0.15 - 1.96(0.05) \le p \le 0.15 + 1.96(0.05)$.

Step 6: $0.05 \le p \le 0.25$

36.6 - Determining the Sample Size.

36.6 - Problem 1:

➤(a).

Step 1: p = 0.30

Step 2: e* = 0.05

Step 3: Since we want a confidence of 99%, z = 2.58.

Step 4: $N = \dfrac{z^2 p(1-p)}{e*^2} = \dfrac{(2.57)^2(0.30)(0.70)}{(0.05)^2} \approx 555.$

➤(b).

If there is no estimate for p then let p = 0.5.

Therefore, $N = \dfrac{z^2 p(1-p)}{e*^2} = \dfrac{(2.57)^2(0.50)(0.50)}{(0.05)^2} \approx 661$

Supplementary Problems

1.

➤a.

$$\overline{P} = \frac{N_1\overline{P}_1 + N_2\overline{P}_2 + N_3\overline{P}_3 + N_4\overline{P}_4 + N_5\overline{P}_5}{N_1 + N_2 + N_3 + +N_4 + N_5} =$$

$$\frac{500(0.35) + 150(0.42) + 250(0.39) + 400(0.34) + 200(0.25)}{500 + 150 + 250 + 400 + 200} \approx 0.35$$

➤b.

$$\sigma_{\overline{P}} = \frac{\sqrt{N_1\overline{P}_1(1-\overline{P}_1) + N_2\overline{P}_2(1-\overline{P}_2) + \ldots + N_n\overline{P}_n(1-\overline{P}_n)}}{N_1 + N_2 + \ldots + N_n} =$$

$$\frac{\sqrt{500[0.35(0.65)]+150[0.42(0.58)]+250[0.39(0.61)]+400[0.34(0.66)]+200[0.25(0.75)]}}{500+150+250+400+200} \approx 0.01$$

➤c.

Step 1:Here $\overline{P} = 0.35$ and $\sigma_{\overline{P}} = 0.01$

The difference between $\overline{P}$ and p is the error $(\overline{P} - p)$. We need to find the probability that the error exceeds 2%.

Step 2: $e^* = \overline{P} - p = \pm z\sigma_{\overline{P}} = \pm z(0.01) = \pm 0.02$.

Step 3: Solving for z gives $z = \pm\frac{0.02}{0.01} \approx \pm 2$

Step 4: From the normal distribution table for z =2: P{e* > 0.02} = 1 - 0.4772 - 0.4772 = 0.0456

➤d.

The z value from the normal distribution table for 095/2 = 0.475 is z = 1.96

For $\sigma_{\overline{P}}$ =0.01 and $\overline{P}$ = 0.35, $0.35 - 1.96(0.01) \leq p \leq 0.35 + 1.96(0.01)$

$0.33 \leq p \leq 0.37$

2.

➤a.

$$\overline{P} = \frac{N_1\overline{P}_1 + N_2\overline{P}_2 + \ldots + N_n\overline{P}_n}{N_1 + N_2 + \ldots + N_n} = \frac{N\overline{P}_1 + N\overline{P}_2 + \ldots + N\overline{P}_n}{N + N + \ldots + N} =$$

$$= \frac{N(\overline{P}_1 + \overline{P}_2 + \ldots + \overline{P}_n)}{nN} = \frac{\overline{P}_1 + \overline{P}_2 + \ldots + \overline{P}_n}{n}$$

$$\sigma_{\overline{P}} = \frac{\sqrt{N_1\overline{P}_1(1 - \overline{P}_1) + N_2\overline{P}_2(1 - \overline{P}_2) + \ldots + N_n\overline{P}_n(1 - \overline{P}_n)}}{N_1 + N_2 + \ldots + N_n} =$$

$$\frac{\sqrt{N\overline{P}_1(1 - \overline{P}_1) + N\overline{P}_2(1 - \overline{P}_2) + \ldots + N\overline{P}_n(1 - \overline{P}_n)}}{nN} =$$

$$\frac{\sqrt{\overline{P}_1(1 - \overline{P}_1) + \overline{P}_2(1 - \overline{P}_2) + \ldots + \overline{P}_n(1 - \overline{P}_n)}}{n\sqrt{N}}$$

➤b.

$$\sigma_{\overline{P}} = \frac{\sqrt{\overline{P}_1(1 - \overline{P}_1) + \overline{P}_2(1 - \overline{P}_2) + ...+\overline{P}_n(1 - \overline{P}_n)}}{n\sqrt{N}}$$

$$e^* = \pm z\sigma_{\overline{P}} = \pm z\frac{\sqrt{\overline{P}_1(1 - \overline{P}_1) + \overline{P}_2(1 - \overline{P}_2) + ...+\overline{P}_n(1 - \overline{P}_n)}}{n\sqrt{N}}$$

Solving for N, we get $N = z^2[\frac{\overline{P}_1(1 - \overline{P}_1) + \overline{P}_2(1 - \overline{P}_2) + ...+\overline{P}_n(1 - \overline{P}_n)}{e*^2n^2}]$

3.

➤a.

We use the formula $\sigma_{\overline{P}} = \frac{\sqrt{\overline{P}_1(1 - \overline{P}_1) + \overline{P}_2(1 - \overline{P}_2) + ...+\overline{P}_n(1 - \overline{P}_n)}}{n\sqrt{N}}$

where n = 9

N = 20

MONTH	$\overline{P}_k$	**1 -** $\overline{P}_k$	**(1 -** $\overline{P}_k$**)**
October	0.57	0.43	0.2451
November	0.59	0.41	0.2419
December	0.61	0.39	0.2379
January	0.62	0.38	0.2356
February	0.52	0.48	0.2496
March	0.51	0.49	0.2499
April	0.49	0.51	0.2499
May	0.55	0.45	0.2475
June	0.60	0.4	0.24
	Total = 5.06		**Total** ≈ 2.2

$\overline{P} = 5.06/9 \approx 0.56$

$$\sigma_{\overline{P}} = \frac{\sqrt{\overline{P}_1(1 - \overline{P}_1) + \overline{P}_2(1 - \overline{P}_2) + ...+\overline{P}_n(1 - \overline{P}_n)}}{n\sqrt{N}} = \frac{\sqrt{2.2}}{(9)\sqrt{20}} \approx 0.04$$

For 95% confidence, z = 1.96.

$0.56 - 1.96(0.04) \le p \le 0.56 + 1.96(0.04)$

$0.48 \le p \le 0.64$

➤b.

From above, $N = z^2[\frac{\overline{P}_1(1-\overline{P}_1) + \overline{P}_2(1-\overline{P}_2) + ...+\overline{P}_n(1-\overline{P}_n)}{e*^2n^2}]$

For 0.90/2 = 0.45, z = 1.64.

e* = 0.02
n = 9

From the above table,

$$N = z^2[\frac{\overline{P}_1(1-\overline{P}_1) + \overline{P}_2(1-\overline{P}_2) + ...+\overline{P}_n(1-\overline{P}_n)}{e*^2n^2}] = \frac{(1.64)^2(2.2)}{(0.02)^2(9^2)} \approx 183$$

➤c.
The following table is a partial list of the amount he can lose for any given month:

Number of Games Lost (k)	Amount Lost (X)
10 11 12	-\$100 -\$310 -\$520
...	...

Let X be the amount he loses for a given month. Therefore, we need to find P{X ≤ -\$100}, where p = 0.44 (the chance he loses a single game) . Since there are 20 games per month, we have P{X ≤ -\$100} = P{$\overline{P}$ ≥ 10/20} = P{$\overline{P}$ ≥ 0.5}.

$$\sigma_{\overline{P}} = \sqrt{\frac{(0.44)(0.56)}{20}} \approx 0.11$$

$$z = \frac{0.5 - 0.44}{0.11} \approx 0.545$$

Therefore, P{X ≤ -\$100} = P{$\overline{P}$ ≥ 0.5} = P{z ≥ 0.545} 0.5 - 0.2088 = 0.2912.

➤d.
From c, we know that the probability that he will lose at the end of any given month is p = 0.2912.

Since this is a binomial problem, we need to use the formula:

$$P\{X = 0\} = \binom{N}{k} p^k q^{N-k},$$

where N = 9, k = 0. and X equals the number of months that he loses money.

Therefore, we have

$$P\{X = 0\} = \binom{9}{0}(0.2912)^0(0.0.7088)^9 \approx 0.05$$

4.

Step 1: $N = \dfrac{z^2 p(1-p)}{e*^2} = \dfrac{z^2(p - p^2)}{e^{*2}}$

Step 2: $-(e^{*2}/z^2)N = p^2 - p$

Step 3: Completing the square: $-(e^{*2}/z^2)N = (p - 0.5)^2 - 0.25$

Step 4: $N = -(z^2/e^{*2})(p - 0.5)^2 + (z^2/e^{*2})0.25$

Step 5: Since $-(z^2/e^{*2})(p - 0.5)^2 < 0$, N is maximum when p = 0.5.