End of 6-2 and 6-3

Determining Sample Size Required to Estimate

p and μ

Determining Sample Size to Estimate *p*

$$E = z_{\alpha/2}\sqrt{\frac{\hat{p}\,\hat{q}}{n}}$$

(solve for *n* by algebra)

$$n = \frac{(z_{\alpha/2})^2\,\hat{p}\,\hat{q}}{E^2}$$

Sample Size for Estimating Proportion *p*

When an estimate of $\hat{p}$ is known:

$$n = \frac{(z_{\alpha/2})^2\,\hat{p}\,\hat{q}}{E^2}$$ Formula 6-2

When no estimate of $\hat{p}$ is known:

$$n = \frac{(z_{\alpha/2})^2\,0.25}{E^2}$$ Formula 6-3

$\hat{p}$	$\hat{q}$	$\hat{p}\hat{q}$
0.1	0.9	0.09
0.2	0.8	0.16
0.3	0.7	0.21
0.4	0.6	0.24
0.5	0.5	0.25
0.6	0.4	0.24
0.7	0.3	0.21
0.8	0.2	0.16
0.9	0.1	0.09

Two Formulas for Estimating Proportion *p*

When an estimate of $\hat{p}$ is known:

$$n = \frac{(z_{\alpha/2})^2 \, \hat{p}\,\hat{q}}{E^2}$$

When no estimate of $\hat{p}$ is known:

$$n = \frac{(z_{\alpha/2})^2 \; 0.25}{E^2}$$

Round-Off Rule for Sample Size *n*

In order to ensure that the required sample size is at least as large as it should be, if the computed sample size is not a whole number, round it up to the next higher whole number.

Example: We want to determine, with a margin of error of four percentage points, the current percentage of U.S. households using e-mail. Assuming that we want 90% confidence in our results, how many households must we survey? A 1997 study indicates 16.9% of U.S. households used e-mail.

$$n = \frac{[z_{\alpha/2}]^2 \hat{p}\hat{q}}{E^2}$$

$$= \frac{[1.645]^2 (0.169)(0.831)}{0.04^2}$$

$= 237.51965$

$= 238$ households

To be 90% confident that our sample percentage is within four percentage points of the true percentage for all households, we should randomly select and survey 238 households.

Example: We want to determine, with a margin of error of four percentage points, the current percentage of U.S. households using e-mail. Assuming that we want 90% confidence in our results, how many households must we survey? There is no prior information suggesting a possible value for the sample percentage.

$$n = \frac{[z_{\alpha/2}]^2 (0.25)}{E^2}$$

$$= \frac{(1.645)^2 (0.25)}{0.04^2}$$

$= 422.81641$

$= 423$ households

With no prior information, we need a larger sample to achieve the same results with 90% confidence and an error of no more than 4%.

Sample Size for Estimating Mean μ

$$E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

(solve for *n* by algebra)

$$n = \left[\frac{z_{\alpha/2}\ \sigma}{E}\right]^2$$

Formula 6-5

$z_{\alpha/2}$ = critical *z* score based on the desired degree of confidence

E = desired margin of error

σ = population standard deviation

Round-Off Rule for Sample Size *n*

In order to ensure that the required sample size is at least as large as it should be, if the computed sample size is not a whole number, round it up to the next higher whole number.

n = 216.09 = 217 (rounded up)

Example: **If we want to estimate the mean weight of plastic discarded by households in one week, how many households must be randomly selected to be 99% confident that the sample mean is within 0.25 lb of the true population mean? (A previous study indicates the standard deviation is 1.065 lb.)**

$\alpha = 0.01$

$z_{\alpha/2} = 2.575$

$E = 0.25$

$s = 1.065$

$$n = \left[\frac{z_{\alpha/2}\,\sigma}{E}\right]^2 = \left[\frac{(2.575)(1.065)}{0.25}\right]^2$$

$$= 120.3 = 121 \text{ households}$$

We would need to randomly select 121 households and obtain the average weight of plastic discarded in one week. We would be 99% confident that this mean is within 1/4 lb of the population mean.

What if σ is Not Known ?

1. Use the range rule of thumb to estimate the standard deviation as follows: $\sigma \approx \frac{\text{range}}{4}$

2. Conduct a pilot study by starting the sampling process. Based on the first collection of at least 31 randomly selected sample values, calculate the sample standard deviation *s* and use it in place of σ. That value can be improved as more sample data are obtained.

3. Estimate the value of σ by using the results of some other study that was done earlier.

What happens if you settle for less accurate results; that is, you increase your margin of error ?

What happens when E is doubled ?

E = 1 : $$n = \left[\frac{z_{\alpha/2}\sigma}{1}\right]^2 = \frac{(z_{\alpha/2}\sigma)^2}{1}$$

E = 2 : $$n = \left[\frac{z_{\alpha/2}\sigma}{2}\right]^2 = \frac{(z_{\alpha/2}\sigma)^2}{4}$$

- **Sample size *n* is decreased to 1/4 of its original value if E is doubled.**
- **Larger errors allow smaller samples.**
- **Smaller errors require larger samples.**