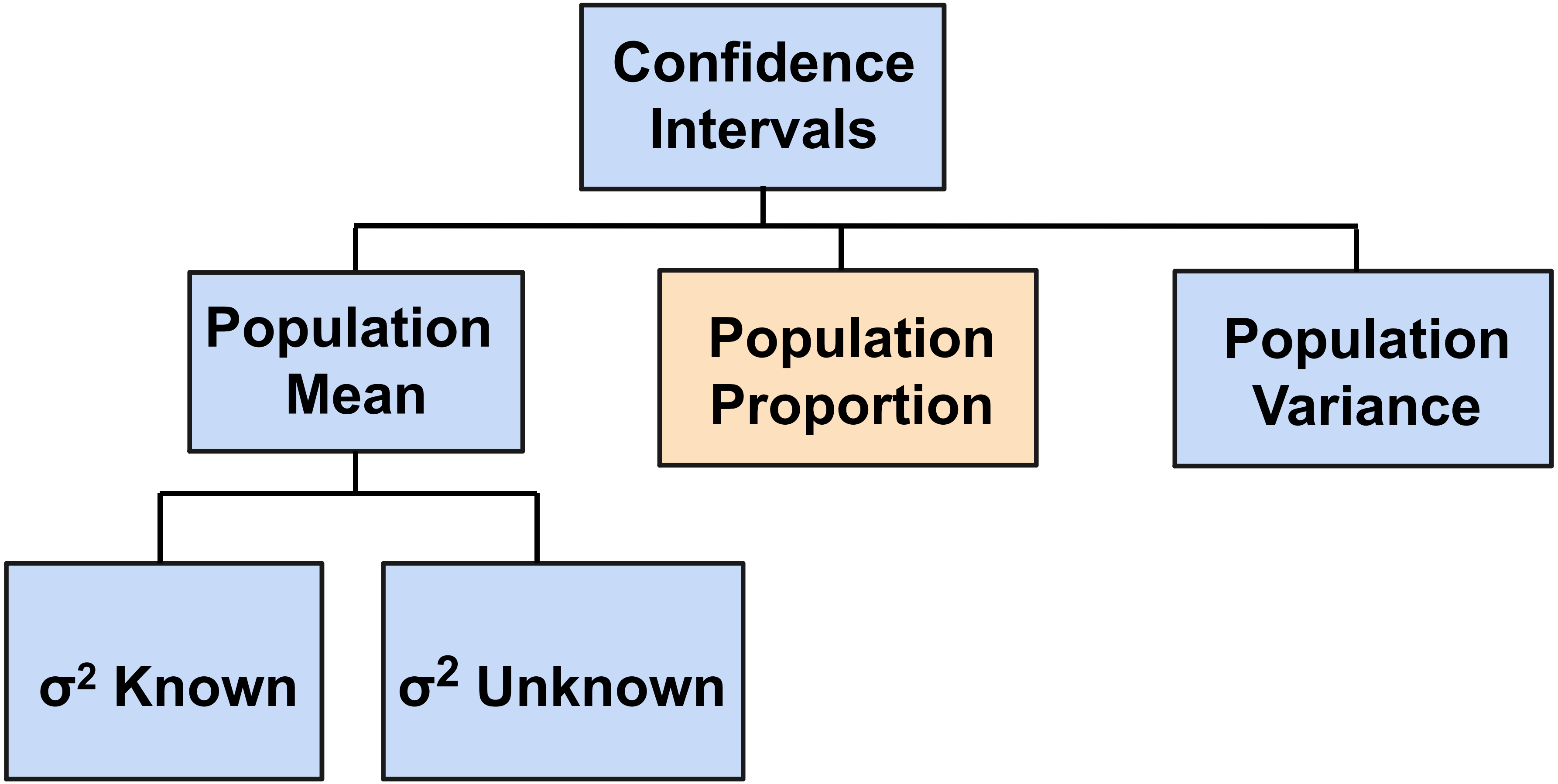
7.4

Confidence Interval Estimation for Population Proportion



Confidence Interval Estimation for Population Proportion

- An interval estimate for the population proportion (P) can be calculated by adding an allowance for uncertainty to the sample proportion ($\hat{p}$)

Confidence Intervals for the Population Proportion

- Recall that the distribution of the sample proportion is approximately normal if the sample size is large, with standard deviation

$$\sigma_P = \sqrt{\frac{P(1-P)}{n}}$$

- We will estimate this with sample data:

$$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Confidence Interval Endpoints

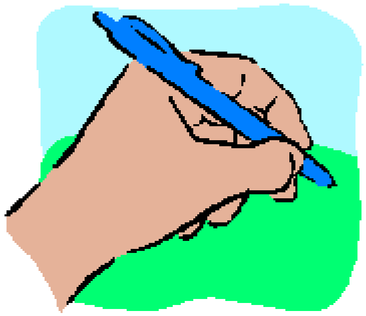
- The confidence interval for the population proportion is given by

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

- where
  - $z_{\alpha/2}$ is the standard normal value for the level of confidence desired
  - $\hat{p}$ is the sample proportion
  - n is the sample size
  - $nP(1-P) > 5$

Example

- A random sample of 100 people shows that 25 are left-handed.
- Form a 95% confidence interval for the true proportion of left-handers



Example

(continued)

- A random sample of 100 people shows that 25 are left-handed. Form a 95% confidence interval for the true proportion of left-handers.

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$\frac{25}{100} \pm 1.96\sqrt{\frac{.25(.75)}{100}}$$

$$0.1651 < P < 0.3349$$

Interpretation

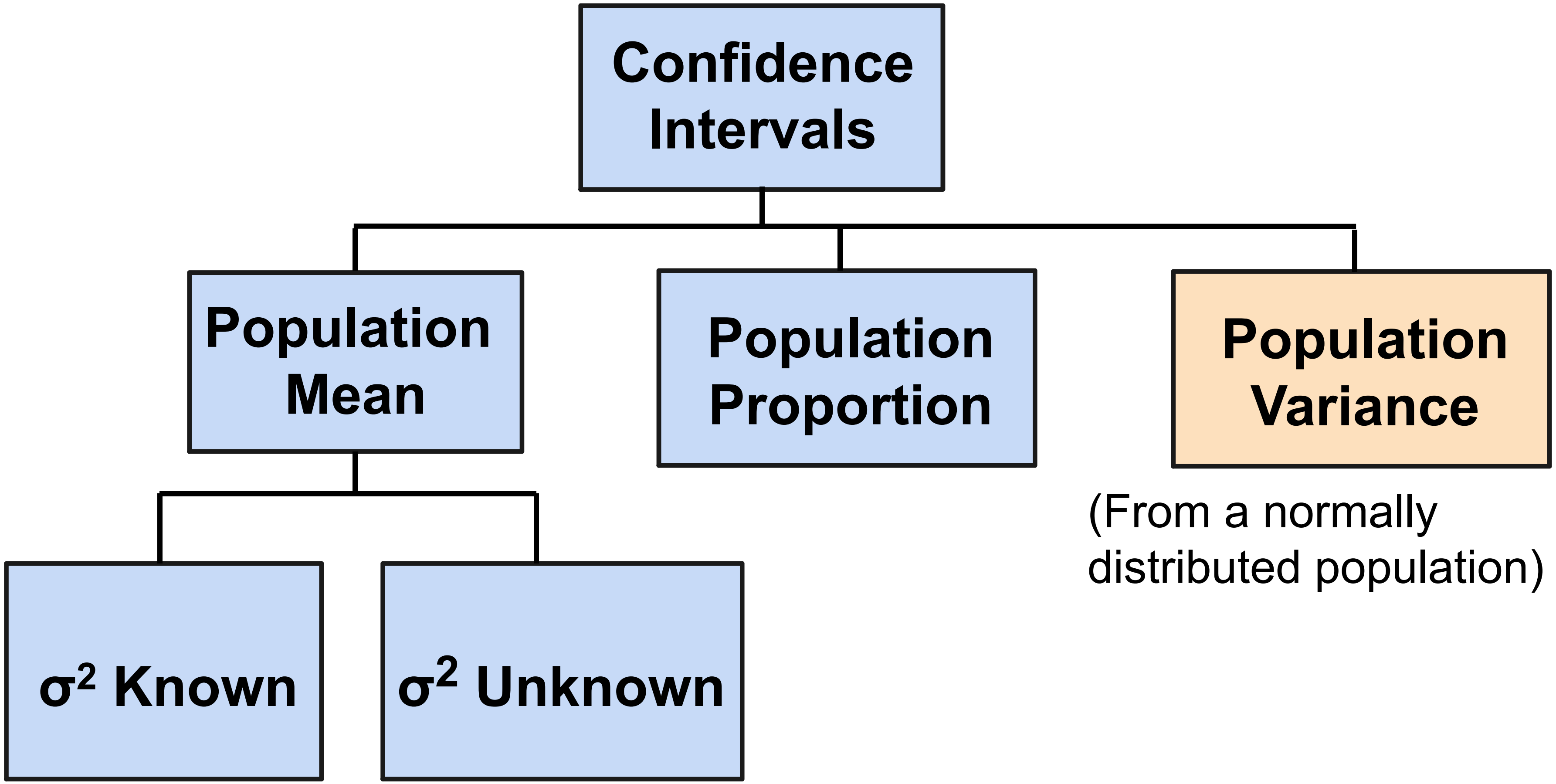
- We are 95% confident that the true proportion of left-handers in the population is between

  16.51% and 33.49%.

- Although the interval from 0.1651 to 0.3349 may or may not contain the true proportion, 95% of intervals formed from samples of size 100 in this manner will contain the true proportion.

7.5

Confidence Interval Estimation for the Variance



Confidence Intervals for the Population Variance

- Goal: Form a confidence interval for the population variance, σ^2

  - The confidence interval is based on the sample variance, s^2

  - Assumed: the population is normally distributed

Confidence Intervals for the Population Variance

(continued)

The random variable

$$\chi^2_{n-1} = \frac{(n-1)s^2}{\sigma^2}$$

follows a chi-square distribution with (n – 1) degrees of freedom

Where the chi-square value $\chi^2_{n-1,\alpha}$ denotes the number for which

$$P(\chi^2_{n-1} > \chi^2_{n-1,\alpha}) = \alpha$$

Confidence Intervals for the Population Variance

(continued)

The 100(1 - α)% confidence interval for the population variance is given by

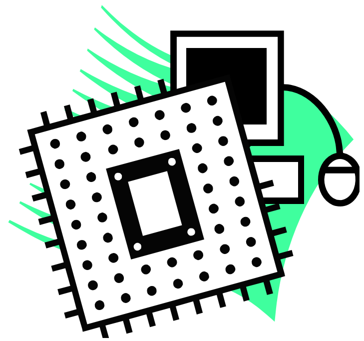
$$\text{LCL} = \frac{(n-1)s^2}{\chi^2_{n-1,\,\alpha/2}}$$

$$\text{UCL} = \frac{(n-1)s^2}{\chi^2_{n-1,\,1-\alpha/2}}$$

Example

You are testing the speed of a batch of computer processors. You collect the following data (in Mhz):

Sample size	**17**
Sample mean	**3004**
Sample std dev	**74**



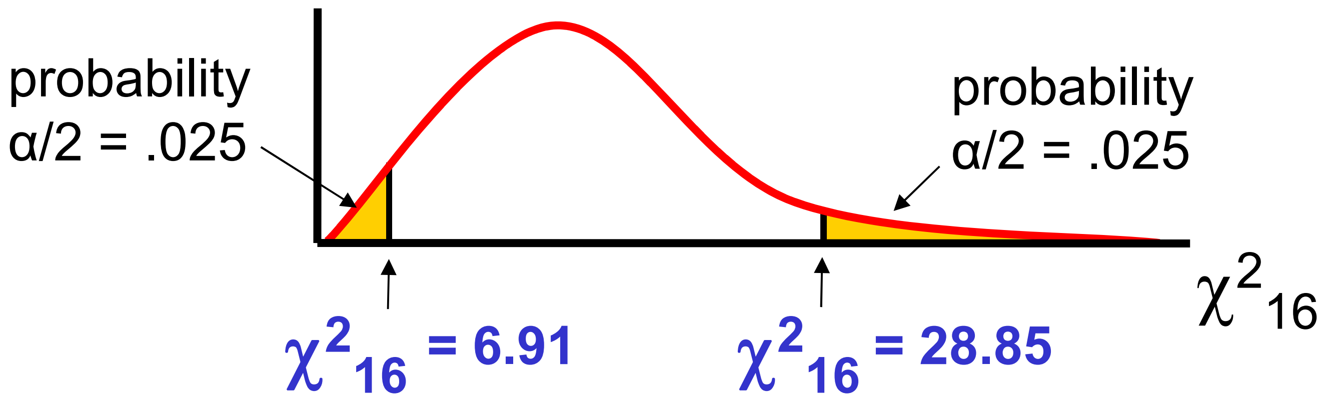
Assume the population is normal.
Determine the 95% confidence interval for σ_x^2

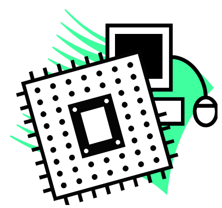
Finding the Chi-square Values

- n = 17 so the chi-square distribution has (n – 1) = 16 degrees of freedom
- α = 0.05, so use the the chi-square values with area 0.025 in each tail:

$$\chi^2_{n-1,\,\alpha/2} = \chi^2_{16\,,\,0.025} = 28.85$$

$$\chi^2_{n-1,\,1-\alpha/2} = \chi^2_{16\,,\,0.975} = 6.91$$





Calculating the Confidence Limits

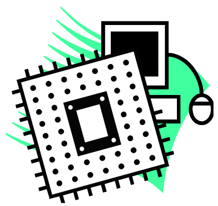
- The 95% confidence interval is

$$\frac{(n-1)s^2}{\chi^2_{n-1,\,\alpha/2}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{n-1,\,1-\alpha/2}}$$

$$\frac{(17-1)(74)^2}{28.85} < \sigma^2 < \frac{(17-1)(74)^2}{6.91}$$

$$3037 < \sigma^2 < 12680$$

Converting to standard deviation, we are 95% confident that the population standard deviation of CPU speed is between 55.1 and 112.6 Mhz



7.6

Confidence Interval Estimation: Finite Populations

- If the sample size is more than 5% of the population size (and sampling is without replacement) then a finite population correction factor must be used when calculating the standard error

Finite Population Correction Factor

- Suppose sampling is without replacement and the sample size is large relative to the population size
- Assume the population size is large enough to apply the central limit theorem
- Apply the finite population correction factor when estimating the population variance

$$\text{finite population correction factor} = \frac{N-n}{N-1}$$

Estimating the Population Mean

- Let a simple random sample of size n be taken from a population of N members with mean μ
- The sample mean is an unbiased estimator of the population mean μ
- The point estimate is:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Finite Populations: Mean

- If the sample size is more than 5% of the population size, an unbiased estimator for the variance of the sample mean is

$$\hat{\sigma}^2_{\bar{x}} = \frac{s^2}{n}\left(\frac{N-n}{N-1}\right)$$

- So the 100(1-α)% confidence interval for the population mean is

$$\bar{x} \pm t_{n-1,\alpha/2}\,\hat{\sigma}_{\bar{x}}$$

Estimating the Population Total

- Consider a simple random sample of size n from a population of size N
- The quantity to be estimated is the population total $N\mu$
- An unbiased estimation procedure for the population total $N\mu$ yields the point estimate $N\bar{x}$

Estimating the Population Total

- An unbiased estimator of the variance of the population total is

$$N^2\hat{\sigma}_{\bar{x}}^2 = N^2 \frac{s^2}{n}\left(\frac{N-n}{N-1}\right)$$

- A 100(1 - α)% confidence interval for the population total is

$$N\bar{x} \pm t_{n-1,\alpha/2} N\hat{\sigma}_{\bar{x}}$$

Confidence Interval for Population Total: Example

A firm has a population of 1000 accounts and wishes to estimate the value of the total population balance

A sample of 80 accounts is selected with average balance of $87.60 and standard deviation of $22.30

Find the 95% confidence interval estimate of the total balance

Example Solution

$$N = 1000, \quad n = 80, \quad \bar{x} = 87.6, \quad s = 22.3$$

$$N^2 \hat{\sigma}_{\bar{x}}^2 = N^2 \frac{s^2}{n} \frac{(N-n)}{N-1} = (1000)^2 \frac{(22.3)^2}{80} \frac{920}{999} = 5724559.6$$

$$N\hat{\sigma}_{\bar{x}} = \sqrt{5724559.6} = 2392.6$$

$$N\bar{x} \pm t_{79,0.025} N\hat{\sigma}_{\bar{x}} = (1000)(87.6) \pm (1.9905)(2392.6)$$

$$82837.53 < N\mu < 92362.47$$

The 95% confidence interval for the population total balance is \$82,837.53 to \$92,362.47

Estimating the Population Proportion: Finite Population

- Let the true population proportion be P
- Let $\hat{p}$ be the sample proportion from n observations from a simple random sample
- The sample proportion, $\hat{p}$, is an unbiased estimator of the population proportion, P

Confidence Intervals for Population Proportion: Finite Population

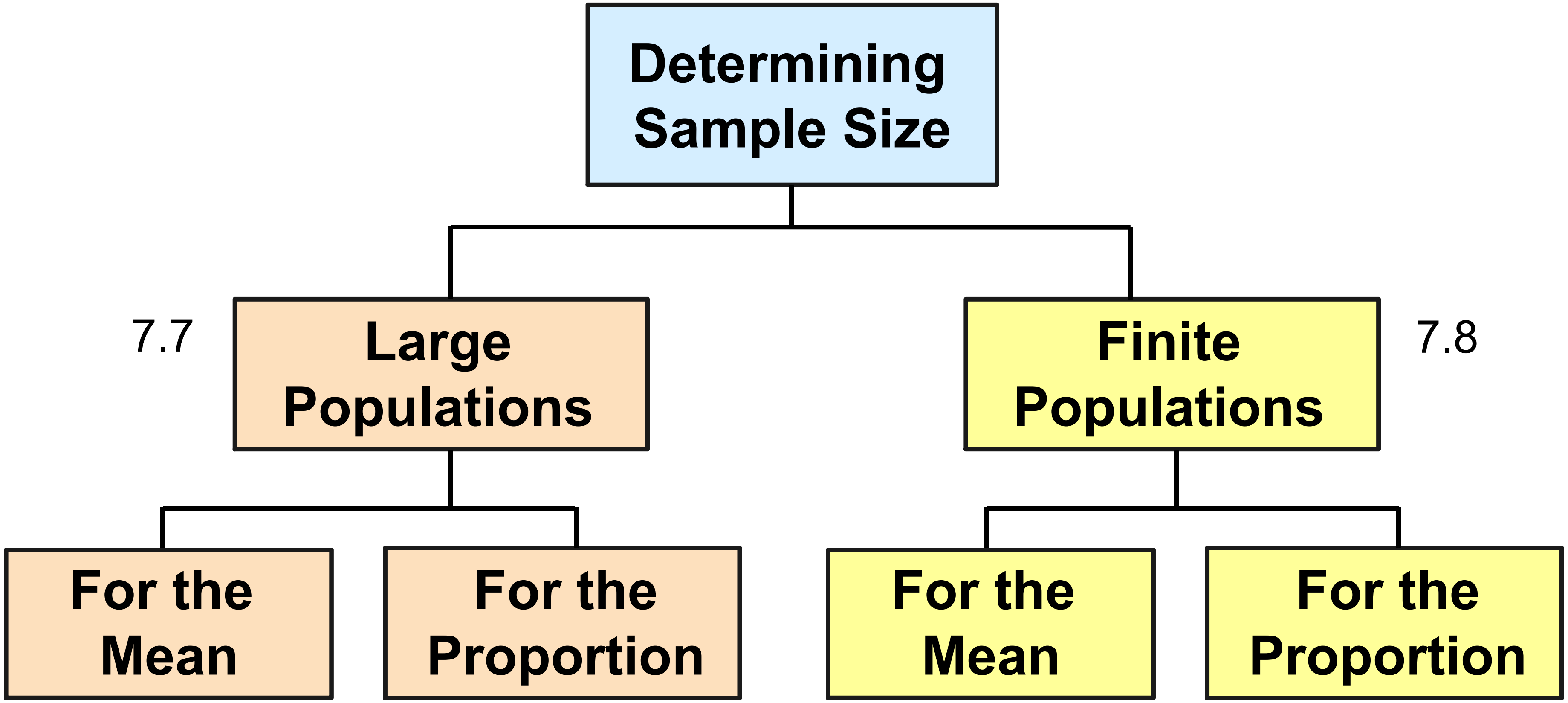
- If the sample size is more than 5% of the population size, an unbiased estimator for the variance of the population proportion is

$$\hat{\sigma}^2_{\hat{p}} = \frac{\hat{p}(1-\hat{p})}{n}\left(\frac{N-n}{N-1}\right)$$

- So the 100(1-α)% confidence interval for the population proportion is

$$\hat{p} \pm z_{\alpha/2}\hat{\sigma}_{\hat{p}}$$

Sample-Size Determination



Sample-Size Determination: Large Populations

Large Populations

For the Mean

(Known population variance)

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Margin of Error (sampling error)

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

Sample-Size Determination: Large Populations

(continued)

Large Populations

For the Mean

(Known population variance)

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

→ Now solve for n to get →

$$n = \frac{z_{\alpha/2}^2 \sigma^2}{ME^2}$$

Sample-Size Determination

(continued)

- To determine the required sample size for the mean, you must know:

  - The desired level of confidence (1 - α), which determines the $z_{\alpha/2}$ value
  - The acceptable margin of error (sampling error), ME
  - The population standard deviation, σ

Required Sample Size Example

If σ = 45, what sample size is needed to estimate the mean within ± 5 with 90% confidence?

$$n = \frac{z_{\alpha/2}^2 \, \sigma^2}{ME^2} = \frac{(1.645)^2 (45)^2}{5^2} = 219.19$$

So the required sample size is **n = 220**

(Always round up)

Sample Size Determination: Population Proportion

Large Populations

For the Proportion

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$ME = z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Margin of Error (sampling error)

Sample Size Determination: Population Proportion

(continued)

Large Populations

For the Proportion

$$ME = z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$\hat{p}(1-\hat{p})$ cannot be larger than 0.25, when $\hat{p}$ = 0.5

→ Substitute 0.25 for $\hat{p}(1-\hat{p})$ and solve for n to get

→ $$n = \frac{0.25\, z_{\alpha/2}^2}{ME^2}$$

Sample Size Determination: Population Proportion

(continued)

- The sample and population proportions, $\hat{p}$ and P, are generally not known (since no sample has been taken yet)
- P(1 – P) = 0.25 generates the largest possible margin of error (so guarantees that the resulting sample size will meet the desired level of confidence)
- To determine the required sample size for the proportion, you must know:
  - The desired level of confidence (1 - α), which determines the critical $z_{\alpha/2}$ value
  - The acceptable sampling error (margin of error), ME
  - Estimate P(1 – P) = 0.25

Required Sample Size Example: Population Proportion

How large a sample would be necessary to estimate the true proportion defective in a large population within ±3%, with 95% confidence?

Required Sample Size Example

(continued)

Solution:

For 95% confidence, use $z_{0.025} = 1.96$

ME = 0.03

Estimate P(1 – P) = 0.25

$$n = \frac{0.25\, z_{\alpha/2}^2}{ME^2} = \frac{(0.25)(1.96)^2}{(0.03)^2} = \boxed{1067.11}$$

So use n = 1068

Sample-Size Determination: Finite Populations

7.8

Finite Populations

For the Mean

A finite population correction factor is added:

$$Var(\overline{X}) = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$$

1. Calculate the required sample size n_0 using the prior formula:

$$n_0 = \frac{z_{\alpha/2}^2 \sigma^2}{ME^2}$$

2. Then adjust for the finite population:

$$n = \frac{n_0 N}{n_0 + (N-1)}$$

Sample-Size Determination: Finite Populations

Finite Populations

For the Proportion

A finite population correction factor is added:

$$\text{Var}(\hat{p}) = \frac{P(1-P)}{n}\left(\frac{N-n}{N-1}\right)$$

1. Solve for n:

$$n = \frac{NP(1-P)}{(N-1)\sigma_{\hat{p}}^2 + P(1-P)}$$

2. The largest possible value for this expression (if P = 0.25) is:

$$n = \frac{0.25(1-P)}{(N-1)\sigma_{\hat{p}}^2 + 0.25}$$

3. A 95% confidence interval will extend $\pm 1.96\,\sigma_{\hat{p}}$ from the sample proportion

Example: Sample Size to Estimate Population Proportion

How large a sample would be necessary to estimate within ±5% the true proportion of college graduates in a population of 850 people with 95% confidence?

Required Sample Size Example

(continued)

Solution:

- For 95% confidence, use $z_{0.025} = 1.96$
- ME = 0.05

$$1.96\,\sigma_{\hat{p}} = 0.05 \quad \Rightarrow \quad \sigma_{\hat{p}} = 0.02551$$

$$n_{max} = \frac{0.25N}{(N-1)\sigma_{\hat{p}}^2 + 0.25} = \frac{(0.25)(850)}{(849)(0.02551)^2 + 0.25} = 264.8$$

So use n = 265