/

Confidence Intervals

As part of a study on plant growth, a plant physiologist grew 13 individually potted soybean seedlings of the "Wells II" type. The seedlings were raised in a greenhouse under identical environmental conditions (i.e. light, temperature, soil, water, etc.). The total stem length in centimeters (cm) of each seedling after 16 days of growth was recorded.

Total Stem Length (in cm)

20.2	22.9	23.3	20.0	19.4	22.0	22.1
22.0	21.9	21.5	19.7	21.5	20.9	

The (conceptual) population consists of the billions and billions of individually potted soybean seedlings of the "Wells II" type that could have been grown under the exact same conditions under which the 13 of this experiment were grown. The random variable Y is the total stem length (in cm) after 16 days of growth of *one* seedling randomly selected from this population.

$$E(Y) = \mu_Y \quad \text{and} \quad V(Y) = \sigma_Y^2$$

We envision the 13 seedlings of this experiment as a random sample of size 13 taken from this population *with* replacement.

Seedling	1	2	...	13
Total stem length (in cm) after 16 days of growth	Y_1	Y_2	...	Y_{13}

Since the population size is large relative to the sample size, it doesn't matter whether the sampling was conducted with or without replacement.

Our objective is to use $Y_1 \;\; Y_2 \;\; \ldots \;\; Y_{13}$ to construct confidence intervals for μ_Y.

(For the 13 stem lengths, $\overline{Y} = 21.3384615$, and $S_Y{}^2 = 1.4858974$).

Confidence Intervals for μ_Y

Y is any random variable (discrete or continuous) with $E(Y) = \mu_Y$ = the population average value for Y and $V(Y) = \sigma_Y^2$ = the population average squared deviation of Y from μ_Y. Randomly sample n items from the population *with* replacement. Measure Y on each.

Trial number	1	2	3	...	n
Y values	Y_1	Y_2	Y_3	...	Y_n

If the sample size is *small* relative to the population size, it doesn't matter whether the sampling is done *with* or *without* replacement.

Recall that the Central Limit Theorem (CLT) says

$$\overline{Y} = \sum_{j=1}^{n} Y_j \div n \quad \text{is approximately } N(\ \mu = \mu_Y\ ,\ \sigma^2 = \sigma_Y^2 / n\ )$$

if n is "large enough".

However, if Y itself is normally distributed, then

$$\overline{Y} = \sum_{j=1}^{n} Y_j \div n \quad \text{is } N(\ \mu = \mu_Y\ ,\ \sigma^2 = \sigma_Y^2 / n\ )$$

for *all* values of n ; i.e. n = 1, 2, 3, 4, 5, 6, 7, ... , etc. .

To calculate confidence intervals (CI's) for μ_Y when σ_Y^2 is *known*, we shall require $\overline{Y}$ to be (at least approximately) normally distributed. Thus, if n is large, $\overline{Y}$ will be approximately normally distributed by the CLT, regardless of the pdf for Y (or probability function for discrete Y). For small to moderate n, however, Y itself must be normally distributed to guarantee that $\overline{Y}$ is normally distributed.

Therefore,

$(\overline{Y} - \mu_Y) \div (\sigma_Y / \sqrt{n})$ is N[0, 1] (or ≈ N[0, 1] if n is "large enough").

Since P(−1.96 < N[0, 1] < 1.96) = .95 , P(N[0, 1] < 1.645) = .95 , and P(−1.645 < N[0, 1]) = .95 ,

$$P[\ -1.96 < (\overline{Y} - \mu_Y) \div (\sigma_Y / \sqrt{n}) < 1.96\ ] = .95 \text{ (or } \approx .95 \text{)},$$

$$P[\ (\overline{Y} - \mu_Y) \div (\sigma_Y / \sqrt{n}) < 1.645\ ] = .95 \text{ (or } \approx .95 \text{), and}$$

$$P[\ -1.645 < (\overline{Y} - \mu_Y) \div (\sigma_Y / \sqrt{n})\ ] = .95 \text{ (or } \approx .95 \text{)}.$$

$P[\ \overline{Y} - 1.96 (\sigma_Y / \sqrt{n}) < \mu_Y < \overline{Y} + 1.96 (\sigma_Y / \sqrt{n}\ ] = .95$ (or ≈ .95), making $\overline{Y} \pm 1.96 (\sigma_Y / \sqrt{n})$ a **two sided** (perhaps approximate) 95% confidence interval for μ_Y when σ_Y^2 is *known*. The length of this interval is $2 \times 1.96 (\sigma_Y / \sqrt{n})$.

$P[\ \overline{Y} - 1.645 (\sigma_Y / \sqrt{n}) < \mu_Y\ ] = .95$ (or ≈ .95), making $[\ \overline{Y} - 1.645 (\sigma_Y / \sqrt{n})\ ,\ +\infty\ )$ an **upper** (perhaps approximate) 95% confidence interval for μ_Y when σ_Y^2 is *known*.

$P[\ \mu_Y < \overline{Y} + 1.645 (\sigma_Y / \sqrt{n})\ ] = .95$ (or ≈ .95), making $(\ -\infty\ ,\ \overline{Y} + 1.645 (\sigma_Y / \sqrt{n})\ ]$ a **lower** (perhaps approximate) 95% confidence interval for μ_Y when σ_Y^2 is *known*.