Chapter 6 Group Practice

EN.553.211 Fall 2026

Each original question is followed by its worked solution. Try the question before reading the next page. **Notation.** $N(\mu, \sigma)$ uses mean μ and standard deviation σ. Each group should show the random variable, model, probability statement, and reasoning. Each member should be ready to explain the solution.

All numbers are illustrative. For these exercises, assume the CLT approximation is adequate. Sample size alone does not guarantee its accuracy.

z	0	1	1.2816	1.6449	2
$\Phi(z)$	0.5000	0.8413	0.9000	0.9500	0.9772

Table values are rounded. Use $\Phi(-z) = 1 - \Phi(z)$, $z_{0.10} \approx 1.2816$, and $z_{0.05} \approx 1.6449$.

Problem 1 Cell fluorescence

Fluorescence measurements for individual cells in a stable population are strongly right-skewed, with mean 40 units and standard deviation 20 units. A laboratory reports the average fluorescence of 100 independently sampled cells. Let $\overline{X}$ be this sample average and $T = 100\overline{X}$ be the total fluorescence.

(a) Find the mean, standard deviation, and approximate distribution of both T and $\overline{X}$. Which statements are exact, and which are approximate?

(b) Approximately what proportion of repeated samples have average fluorescence between 38 and 44 units? Calculate this probability in **two ways**: using the average $\overline{X}$ and using the total T. For each route, state the event, standardize both endpoints, and use the table. Explain why the answers agree. Can this calculation determine the proportion of *individual cells* between 38 and 44 units?

(c) Find the approximate 90th percentile of $\overline{X}$, starting with a probability equation. Interpret it in terms of repeated samples of 100 cells.

Problem 1 Worked Solution

Problem 1 Cell fluorescence

(a) Independence gives the exact moments

$$\begin{aligned} \mathrm{E}[T] &= 100(40) = 4000, \qquad \mathrm{Var}(T) = 100(20^2) = 40000, \qquad \mathrm{SD}(T) = 200, \\ \mathrm{E}[\overline{X}] &= 4000/100 = 40, \qquad \mathrm{Var}(\overline{X}) = 40000/100^2 = 4, \qquad \mathrm{SD}(\overline{X}) = 2. \end{aligned}$$

The CLT gives $T \overset{\text{approx}}{\sim} N(4000, 200)$ and $\overline{X} \overset{\text{approx}}{\sim} N(40, 2)$. Their normal shapes are approximate; the means and standard deviations are exact.

(b) The average and total describe the same event:

$$\begin{aligned} \Pr(38 < \overline{X} < 44) &\approx \Phi\left(\frac{44-40}{2}\right) - \Phi\left(\frac{38-40}{2}\right) = \Phi(2) - \Phi(-1), \\ \Pr(3800 < T < 4400) &\approx \Phi\left(\frac{4400-4000}{200}\right) - \Phi\left(\frac{3800-4000}{200}\right) = \Phi(2) - \Phi(-1). \end{aligned}$$

Both routes give $0.9772 - (1 - 0.8413) = \boxed{0.8185}$ using the table (unrounded CDFs give 0.8186 to four decimals). About 81.85% of repeated samples have an average in this interval. This does *not* determine the proportion of individual cells in the interval: their distribution is skewed, their SD is 20, and the first two moments alone do not specify that probability.

(c) Start with $\Pr(\overline{X} \le c) \approx 0.90$. Then

$$\Phi\left(\frac{c-40}{2}\right) = 0.90 \quad \Longrightarrow \quad c \approx 40 + 2z_{0.10} \approx 40 + 2(1.2816) \approx \boxed{42.56 \text{ units}}.$$

About 90% of repeated samples of 100 cells have an average at or below this cutoff.

Problem 2 Question

Use the normal table on page 1. Assume the CLT approximation is adequate, as in Problem 1.

Problem 2 Comparing two fertilizers

A greenhouse compares two-week biomass gains under two fertilizers. For fertilizer A, the population mean is 14 g and the standard deviation is 8 g. For fertilizer B, the population mean is 12 g and the standard deviation is 6 g. Both distributions may be skewed.

Initially, 100 seedlings are independently sampled for each fertilizer. Observations within each group are identically distributed, and the two groups are independent. Define $D = \overline{X}_A - \overline{X}_B$.

(a) Find the mean, standard deviation, and approximate normal distribution of D. Show the variance calculation.

(b) Despite fertilizer A's greater population mean, what is the approximate probability that its observed sample mean is below fertilizer B's? Express the event in terms of D first.

(c) Now use n seedlings in *each* group. Find the smallest integer n for which the normal approximation gives at least a 95% probability that A's observed mean exceeds B's by more than 1 g. Show how n enters the standard deviation and probability.

Problem 2 Worked Solution

Problem 2 Comparing two fertilizers

(a) For $D = \overline{X}_A - \overline{X}_B$, independence makes the variances *add*:

$$\mathrm{E}[D] = 14 - 12 = 2, \qquad \mathrm{Var}(D) = \frac{8^2}{100} + \frac{6^2}{100} = 1, \qquad \mathrm{SD}(D) = 1.$$

The moments are exact; the approximate distribution is $\boxed{D \overset{\text{approx}}{\sim} N(2,1)}$ (in grams).

(b) A's observed mean is below B's exactly when $D < 0$. Thus

$$\Pr(\overline{X}_A < \overline{X}_B) = \Pr(D < 0) \approx \Phi\left(\frac{0-2}{1}\right) = \Phi(-2) = 1 - \Phi(2) \approx \boxed{0.0228}.$$

The ordering reverses in about 2.28% of repeated pairs of samples.

(c) With n seedlings in each group, $\mathrm{E}[D] = 2$, $\mathrm{Var}(D) = 100/n$, and $\mathrm{SD}(D) = 10/\sqrt{n}$. The requested event is $D > 1$, so

$$\Pr(D > 1) \approx 1 - \Phi\left(\frac{1-2}{10/\sqrt{n}}\right) = \Phi\left(\frac{\sqrt{n}}{10}\right) \geq 0.95 \quad \Longleftrightarrow \quad n \geq (10 z_{0.05})^2 \approx 270.554.$$

Round up: $\boxed{n = 271 \text{ seedlings per group}}$, or 542 total. Unrounded normal CDF values give 0.949826 at $n = 270$ and 0.950140 at $n = 271$. This minimum is under the normal approximation, not an exact guarantee for arbitrary population distributions.