1. An online learning platform has 80,000 registered students. Fifteen percent of the registered students use the platform's live-tutoring feature in a typical week.

A. Twenty-five registered students are selected at random.

i. Define the random variable of interest for the number of students in the sample who use the live-tutoring feature in a typical week, and state how the random variable is distributed.

ii. Determine the probability that at most 2 of the 25 students use the live-tutoring feature. Show your work.

Part	Model solution	Scoring
A	(i) Let X = the number of the 25 randomly selected students who use live tutoring in a typical week. X has a binomial distribution with $n = 25$ and $p = 0.15$. (Each student uses it or not; n is fixed at 25; p = 0.15 for each student; students are selected at random and 25 ≤ 0.10(80,000), so the trials are approximately independent.) (ii) P($X \le 2$) = P(X = 0) + P(X = 1) + P(X = 2) ≈ **0.254**. (binomcdf(n = 25, p = 0.15, x = 2))	**E** if the response: 1. defines X in context (number of tutoring users out of 25); 2. states binomial with n = 25 and p = 0.15; 3. gives P(X ≤ 2) ≈ 0.254 with the binomial setup shown. **P** if exactly two of the three components are satisfied. **I** otherwise. *Note:* P(X < 2) = P(X ≤ 1) ≈ 0.093 is a boundary error and does not satisfy component 3.

B. Each semester the platform surveys a random sample of 300 registered students. Let $\hat{p}$ be the proportion of the sampled students who use the live-tutoring feature in a typical week.

i. Describe the sampling distribution of $\hat{p}$. Justify your answer.

ii. The platform reports "strong tutoring engagement" if more than 18 percent of the sampled students use the live-tutoring feature. Determine the probability that the platform reports strong tutoring engagement for a particular semester. Show your work.

Part	Model solution	Scoring
B	(i) **Shape:** approximately normal, because $np = 300(0.15) = 45 \ge 10$ and $n(1-p) = 300(0.85) = 255 \ge 10$. **Center:** $\mu_{\hat{p}} = 0.15$. **Variability:** $\sigma_{\hat{p}} = \sqrt{\frac{0.15(0.85)}{300}} \approx 0.0206$; the formula is appropriate because 300 ≤ 0.10(80,000) = 8,000. (ii) $z = \frac{0.18 - 0.15}{0.0206} \approx 1.46$ P($\hat{p} > 0.18$) = P($z > 1.46$) ≈ **0.073**. (normalcdf(lower = 0.18, upper = 1000, μ = 0.15, σ = 0.0206))	**E** if the response: 1. gives center 0.15 and SD ≈ 0.0206, with the 10% condition checked; 2. states approximately normal, checking both Large Counts values; 3. uses the normal model with boundary 0.18 and direction ">", and gives ≈ 0.073. **P** if exactly two of the three components are satisfied. **I** otherwise.

C. On a second learning platform, 20 percent of registered students use live tutoring in a typical week. That platform surveys a random sample of 250 of its registered students, independently of the sample of 300 in part B. Let $\hat{p}_A$ be the sample proportion of tutoring users in the sample of 300 from the first platform, and let $\hat{p}_B$ be the sample proportion in the sample of 250 from the second platform.

i. Calculate the mean and the standard deviation of the sampling distribution of $\hat{p}_B - \hat{p}_A$. Show your work.

ii. Determine the probability that the sample from the first platform has a proportion of tutoring users at least as large as the sample from the second platform. Show your work.

Part	Model solution	Scoring
C	(i) Mean: 0.20 − 0.15 = **0.05**. $\sigma_{\hat{p}_B - \hat{p}_A} = \sqrt{\frac{0.20(0.80)}{250} + \frac{0.15(0.85)}{300}} \approx 0.0326$ (variances add because the samples are independent). (ii) The event is $\hat{p}_B - \hat{p}_A \le 0$. The four counts 45, 255, 50 and 200 are all at least 10, so the sampling distribution is approximately normal. $z = \frac{0 - 0.05}{0.0326} \approx -1.53$; $P(\hat{p}_B - \hat{p}_A \le 0) \approx$ **0.063**. (normalcdf(lower = −1000, upper = 0, μ = 0.05, σ = 0.0326))	**E** if the response: 1. gives the mean 0.05; 2. gives the SD ≈ 0.0326 with both variance terms added; 3. translates the event as $\hat{p}_B - \hat{p}_A \le 0$ and gives ≈ 0.063, with a normal model justified by the four counts. **P** if exactly two of the three components are satisfied. **I** otherwise. A response that uses $\hat{p}_A - \hat{p}_B \ge 0$ with mean −0.05 is equivalent.

Scoring. Each part (A, B, C) is scored essentially correct (E), partially correct (P), or incorrect (I).

4	**Complete response:** three parts essentially correct
3	**Substantial response:** two parts essentially correct and one part partially correct
2	**Developing response:** two parts essentially correct and no parts partially correct; *or* one part essentially correct and one or two parts partially correct; *or* three parts partially correct
1	**Minimal response:** one part essentially correct and no parts partially correct; *or* no parts essentially correct and two parts partially correct
0	**No credit:** no parts essentially correct and fewer than two parts partially correct

2. A company makes rechargeable batteries for wireless earbuds in two factories, and each factory produces thousands of batteries each day. The battery life (time until the charge runs out) of batteries from Factory 1 is approximately normally distributed with mean 20 hours and standard deviation 2.5 hours. The battery life of batteries from Factory 2 has a distribution that is skewed to the left with mean 19 hours and standard deviation 3 hours.

A. Consider the batteries from Factory 1.

i. Determine the probability that a randomly selected battery from Factory 1 lasts more than 24 hours. Show your work.

ii. The company will replace for free any battery whose life is in the shortest 2 percent of all Factory 1 batteries. What is the maximum battery life that qualifies for a free replacement? Show your work.

Part	Model solution	Scoring
A	(i) $z = \dfrac{24 - 20}{2.5} = 1.6$; P($X$ > 24) ≈ **0.055**. (normalcdf(lower = 24, upper = 1000, μ = 20, σ = 2.5)) (ii) The 2nd percentile: $x = 20 + (-2.054)(2.5) \approx 14.9$ hours. (invNorm(area = 0.02, μ = 20, σ = 2.5) ≈ 14.87)	**E** if the response: 1. identifies the normal distribution with mean 20 and SD 2.5 for one battery, with boundary 24 and direction ">"; 2. gives ≈ 0.055; 3. finds the 2nd percentile ≈ 14.9 hours, using z ≈ −2.054 or invNorm with area 0.02. **P** if exactly two of the three components are satisfied. **I** otherwise. *Note:* using the 98th percentile (≈ 25.1 hours) does not satisfy component 3.

B. A quality inspector selects a random sample of 4 batteries from Factory 1 and calculates the sample mean battery life $\bar{x}$.

i. Describe the sampling distribution of $\bar{x}$. Justify your answer.

ii. Determine the probability that the mean battery life of the 4 batteries is more than 22.5 hours. Explain why this probability is smaller than the probability that a single battery lasts more than 22.5 hours.

Part	Model solution	Scoring
B	(i) **Shape:** exactly normal, because the population of Factory 1 battery lives is normal — even though n = 4 is small. **Center:** $\mu_{\bar{x}}$ = 20 hours. **Variability:** $\sigma_{\bar{x}} = \dfrac{2.5}{\sqrt{4}} = 1.25$ hours (4 is less than 10% of the batteries produced). (ii) $z = \dfrac{22.5 - 20}{1.25} = 2$; P($\bar{x}$ > 22.5) ≈ **0.023**. For a single battery, z = (22.5 − 20)/2.5 = 1 and the probability is about 0.159. The mean of 4 batteries varies less than a single battery (SD 1.25 vs 2.5 hours), so a mean as high as 22.5 is less likely.	**E** if the response: 1. gives center 20 and SD 1.25; 2. states normal because the population is normal (not because of the CLT, since n = 4 < 30); 3. gives ≈ 0.023 with the normal model for $\bar{x}$; 4. explains the comparison using the smaller variability of $\bar{x}$. **P** if exactly two or three of the four components are satisfied. **I** otherwise.

C. The inspector also selects a random sample of 40 batteries from Factory 2, independently of the sample of 4 batteries from Factory 1. Let $\bar{x}_1$ be the mean battery life of the 4 batteries from Factory 1 and $\bar{x}_2$ the mean battery life of the 40 batteries from Factory 2.

i. Describe the sampling distribution of $\bar{x}_1 - \bar{x}_2$. Justify your answer.

ii. Determine the probability that the sample of batteries from Factory 2 has a longer mean battery life than the sample from Factory 1. Show your work.

Part	Model solution	Scoring
C	(i) **Shape:** approximately normal: $\bar{x}_1$ is normal because Factory 1 battery lives are normal, and $\bar{x}_2$ is approximately normal by the central limit theorem because n = 40 ≥ 30, even though Factory 2 battery lives are skewed to the left. **Center:** 20 − 19 = **1 hour**. **Variability:** $\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{2.5^2}{4} + \frac{3^2}{40}} = \sqrt{1.5625 + 0.225} \approx 1.34$ hours (independent samples, variances add). (ii) The event is $\bar{x}_1 - \bar{x}_2 < 0$. $z = \frac{0 - 1}{1.34} \approx -0.75$; $P(\bar{x}_1 - \bar{x}_2 < 0) \approx$ **0.227**. (normalcdf(lower = −1000, upper = 0, μ = 1, σ = 1.34))	**E** if the response: 1. gives the mean 1 and SD ≈ 1.34 with both variance terms added; 2. justifies approximate normality for *both* sample means (population normal for Factory 1; CLT with n = 40 for Factory 2); 3. translates the event as $\bar{x}_1 - \bar{x}_2 < 0$ and gives ≈ 0.227. **P** if exactly two of the three components are satisfied. **I** otherwise.

Scoring. Each part (A, B, C) is scored essentially correct (E), partially correct (P), or incorrect (I).

4	**Complete response:** three parts essentially correct
3	**Substantial response:** two parts essentially correct and one part partially correct
2	**Developing response:** two parts essentially correct and no parts partially correct; *or* one part essentially correct and one or two parts partially correct; *or* three parts partially correct
1	**Minimal response:** one part essentially correct and no parts partially correct; *or* no parts essentially correct and two parts partially correct
0	**No credit:** no parts essentially correct and fewer than two parts partially correct

END OF SCORING GUIDELINES