4. An automobile company wants to learn about customer satisfaction among the owners of five specific car models. Large sales volumes have been recorded for three of the models, but the other two models were recently introduced so their sales volumes are smaller. The number of new cars sold in the last six months for each of the models is shown in the table below.

Car Model	A	B	C	D	E	Total
Number of new cars sold in the last six months	112,338	96,174	83,241	3,278	2,323	297,354

The company can obtain a list of all individuals who purchased new cars in the last six months for each of the five models shown in the table. The company wants to sample 2,000 of these owners.

(a) For simple random samples of 2,000 new car owners, what is the expected number of owners of model E and the standard deviation of the number of owners of model E?

(b) When selecting a simple random sample of 2,000 new car owners, how likely is it that fewer than 12 owners of model E would be included in the sample? Justify your answer.

(c) The company is concerned that a simple random sample of 2,000 owners would include fewer than 12 owners of model D or fewer than 12 owners of model E. Briefly describe a sampling method for randomly selecting 2,000 owners that will ensure at least 12 owners will be selected for each of the 5 car models.

---

5. A large pet store buys the identical species of adult tropical fish from two different suppliers—Buy-Rite Pets and Fish Friends. Several of the managers at the pet store suspect that the lengths of the fish from Fish Friends are consistently greater than the lengths of the fish from Buy-Rite Pets. Random samples of 8 adult fish of the species from Buy-Rite Pets and 10 adult fish of the same species from Fish Friends were selected and the lengths of the fish, in inches, were recorded, as shown in the table below.

	Length of Fish	Mean	Standard Deviation
Buy-Rite Pets ($n_B = 8$)	3.4 2.7 3.3 4.1 3.5 3.4 3.0 3.8	3.40	0.434
Fish Friends ($n_F = 10$)	3.3 2.9 4.2 3.1 4.2 4.0 3.4 3.2 3.7 2.6	3.46	0.550

Do the data provide convincing evidence that the mean length of the adult fish of the species from Fish Friends is greater than the mean length of the adult fish of the same species from Buy-Rite Pets?

AP® STATISTICS
2010 SCORING GUIDELINES

Question 4

Intent of Question

The primary goals of this question were to (1) assess students' ability to calculate an expected value and a standard deviation; (2) recognize the applicability of a binomial distribution and perform a relevant binomial probability calculation (or recognize the applicability of a normal approximation and use it to perform a relevant probability calculation); (3) suggest an appropriate sampling method to achieve a given goal.

Solution

Part (a):

Because the population size is so large compared with the sample size ($\frac{297,354}{2,000} \approx 148.7$ times the sample size), far greater than the usual standard of 10 or 20 times larger, we can use the binomial probability distribution even though this is technically sampling without replacement. The parameters of this binomial distribution are the sample size, n, which has a value of $n = 2,000,$ and the proportion of new car buyers who bought model E, p, which has a value of $p = \frac{2,323}{297,354} \approx 0.0078.$ The expected value of the number of model E buyers in a simple random sample of 2,000 is therefore $n \times p = 2,000 \times 0.0078 \approx 15.62.$ The variance is $n \times p \times (1-p) = 2,000 \times 0.0078 \times (1-0.0078) \approx 15.50,$ so the standard deviation is $\sqrt{15.50} \approx 3.94.$

Part (b):

For the reason given in part (a), the binomial distribution with $n = 2,000$ and $p \approx 0.0078$ can be used here. The probability that the sample would contain fewer than 12 owners of model E is calculated from the binomial distribution to be $\sum_{x=0}^{11} \binom{2,000}{x} (0.0078)^x (0.9922)^{2,000-x} \approx 0.147$. This probability is small enough that the result (fewer than 12 owners of model E in the sample) is not likely, but this probability is also not small enough to consider the result very unlikely.

This binomial probability can also be evaluated using a normal approximation. This is reasonable because $n \times p = (2,000) \times (0.0078) = 15.6$ is larger than 10 and $n(1-p) = (2,000) \times (0.9922) = 1,984.4$ is much larger than 10. Using the mean and standard deviation from part (a) gives

$$P(X \le 11) \approx P\left(Z < \frac{12.0 - 15.62}{3.94}\right) = P(Z < -0.92) = 0.179.$$

Part (c):

Stratified random sampling addresses the concern about the number of owners for models D and E. By stratifying on car model and then taking a simple random sample of at least 12 owners from the population of owners for each model, the company can ensure that at least 12 owners are included in the sample for each model while maintaining a total sample size of 2,000. For example, the company could select simple random samples of sizes 755, 647, 560, 22 and 16 for models A, B, C, D and E, respectively, to make the sample size approximately proportional to the size of the owner population for each model.

AP® STATISTICS
2010 SCORING GUIDELINES

Question 4 (continued)

Scoring

Parts (a), (b) and (c) are each scored as essentially correct (E), partially correct (P) or incorrect (I).

Part (a) is scored as follows:

Essentially correct (E) if the response correctly addresses the following two components:

- Calculation of the expected number of owners, showing a proper method for the calculation and providing the correct numerical value
- Calculation of the standard deviation for the number of owners, indicating recognition of the appropriate binomial distribution and providing the calculation and the correct numerical value

Partially correct (P) if the response contains only one of the two components listed above *OR* displays correct formulas for both the expected value and the standard deviation of a binomial distribution but fails to show both of the correct numerical values.

Incorrect (I) if the response provides only numerical values without showing how they were calculated.

Part (b) is scored as follows:

Essentially correct (E) if the student does any of the following:

- Recognizes the applicability of the binomial distribution, identifies the correct parameters, sets up the relevant probability calculation, and completes the calculation correctly
- Uses a normal probability approximation, identifying the relevant mean and standard deviation, and shows a correct calculation of the probability
- Provides an argument based on an appropriate *z*-score, or the number of standard deviations away from the mean, with a reasonable conclusion about likeliness

Partially correct (P) if the student does any of the following:

- Recognizes the applicability of the binomial distribution and identifies the correct parameters *BUT* sets up an incorrect cumulative binomial probability calculation
- Recognizes the applicability of the binomial distribution and shows the calculation correctly *BUT* does not identify the correct parameters in either part (a) or part (b)
- Recognizes the applicability of the normal approximation and identifies the correct parameters *BUT* incorrectly calculates the *z*-score or probability

Incorrect (I) otherwise.

Notes

- If the parameter values were properly identified in part (a), they do not have to be identified in part (b).
- If the response shows a correct calculation of the probability, no comment about likeliness is necessary. But such a comment is necessary if the response contains only a *z*-score without a probability or discusses standard deviations from the mean.
- With the normal calculation, it is acceptable for the response to show the probability that the normal value is below 11 or 11.5 or 12.