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.

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.

Question 4 (continued)

Part (c) is scored as follows:

Essentially correct (E) if the response describes an appropriate sampling method (e.g., stratified random sampling) that ensures all of the following:

- Total sample size of 2,000
- At least 12 owners for each of the five car models
- Random selection of owners

Partially correct (P) if the response mentions stratified random sampling but gives a weak description, or no description, of how to implement the procedure *OR* describes another appropriate sampling method but includes only two of the three components listed above.

Incorrect (I) otherwise.

4 Complete Response

All three parts essentially correct

3 Substantial Response

Two parts essentially correct and one part partially correct

2 Developing Response

Two parts essentially correct and one part incorrect

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 two parts incorrect

OR

Two parts partially correct and one part incorrect

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.

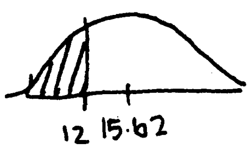
(a) For simple random samples of 2,000 new car owners, what is the expected number of owners of model E and the standard deviation of the number of owners of model E?

$P(E) = \frac{2323}{297,354} = 0.00781$

$E(x) = 2000 \times 0.00781 = \boxed{15.62}$

$\sigma_X = \sqrt{np(1-p)} = \sqrt{15.62(1-0.00781)} = \boxed{3.94}$

(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.



$z = \frac{12-15.62}{3.94} = -0.919$

$P(z < -0.919) = \boxed{0.179}$

(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.

Use a stratified random sample.

$E(A) = 2000 \times 0.378 = 756$

$E(B) = 2000 \times 0.323 = 646$

$E(C) = 2000 \times 0.2799 = 560$

$E(D) = 2000 \times 0.011 = 22$

$E(E) = 2000 \times 0.00781 = 16$

Randomly pick 756 from A, 646 from B, 560 from C, 22 from D and 16 from E. This way, the proportion would be represented as well as more than 12 people will be selected from D and E.

GO ON TO THE NEXT PAGE.

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?

$p_E = \frac{x}{n} = \frac{2323}{297354} = .0078$

expected number of owners in a sample of 2000 new car owners is $p_E n$ = .0078 x 2000 = 15.624 owners of E

standard deviation = $\sqrt{\frac{pq}{n}} = \sqrt{\frac{(.0078)(.9922)}{2000}} = .0020$ owners of 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.

This is a binomial setting, there are two categories that are mutually exclusive and exaustive, and there is a fixed sample size (2000).

binomcdf(n=2000, p=.0078, k=11) gives the probability of getting fewer than 12 owners (0 through 11) in a sample of size 2000 with the probability of success .0078

probability = .1971

(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.

If you seperated the list of new car owners by model, and then you randomly chose 400 owners from each of the groups of the different models, then you would be garunteed 400 owners of each of the different models

GO ON TO THE NEXT PAGE.

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 <u>and</u> the standard deviation of the number of owners of model E?

For simple random samples of 2,000 new car owners, the expected number of owners of model E is approximately 15.62. I obtained this answer by multiplying the proportion of car E owners ($2323/297354 \approx .0078$) by 2,000.

(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.

When using the function binomcdf on the calculator, n represents the number of trials (2000), p equals the proportion of successes (.0078), and X equals the upper limit of the successes. In this case, 12 is not included so X must be 11. Using these values in the function gives a probability of approximately .1471 that fewer than 12 car owners own model E

(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.

To ensure that at least 12 owners will be selected for each model, the ~~the~~ sample size multiplied by the proportion of car owners that own each model must be greater than 12.

GO ON TO THE NEXT PAGE.

Question 4

Overview

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.

Sample: 4A
Score: 4

This is an essentially complete response that clearly demonstrates an understanding of the calculation of an expected value and a standard deviation, the use of z-scores and the calculation of a related probability, and the concept of stratified random sampling. The response to part (a) provides correct formulas and correct numerical values for the expected value and the standard deviation of the number of owners of car model E who would be included in a simple random sample of 2,000 owners from the population of 297,354 new car owners. In part (b) the normal approximation to the binomial distribution is used to obtain the probability of observing fewer than 12 owners of car model E in a simple random sample of 2,000 new car owners. A correct formula is given for the z-score, and it is correctly evaluated. This response uses both a sketch and standard probability notation to clearly indicate that the probability of fewer than 12 owners should be calculated, and the probability is correctly evaluated. The response to part (c) adequately describes an appropriate stratified random sampling procedure. It indicates that a separate sample will be randomly selected from the population of owners for each of the five car models, and it specifies a set of sample sizes that all exceed 12 and yield a total sample size of 2,000. The communication in this part of the response could have been improved by clearly indentifying the strata and indicating what should be sampled from each stratum — for example, randomly selecting 756 car owners from the population of new car owners who purchased car model A in the past six months. A statement justifying the use of the normal approximation to the binomial distribution would have strengthened the communication in part (b). Although communication weaknesses were a concern, all three parts were scored as essentially correct, and the response earned a score of 4.

Sample: 4B
Score: 3

This is a substantial response that is very well written. A correct procedure for computing the expected value is described in part (a), and the correct numerical value is calculated. This response presents the formula for a standard deviation of a sample proportion instead of the standard deviation of the number of owners of car model E who would be included in a simple random sample of 2,000 owners. This was a common mistake. The response to part (a) was scored as partially correct. Using the binomial distribution with parameters $n = 2{,}000$ and $p = 0.0078$, the probability of obtaining fewer than 12 owners of car model E in a random sample of 2,000 owners is correctly calculated in part (b). This response shows additional understanding by providing some justification for using the binomial distribution, although it does not address the issue of sampling without replacement. The response to part (c) describes an appropriate stratified random sampling procedure. Car owners are grouped with respect to the car model they own, and a random sample of 400 owners is selected from the car owners in each group. This ensures that the sample will include exactly 2,000 new car owners with 400 owners for each of the five car models. Parts (b) and (c) were both scored as essentially correct. With one part partially correct and two parts essentially correct, the response earned a score of 3.

Question 4 (continued)

Sample: 4C
Score: 2

This is an example of a developing response that clearly demonstrates understanding of the calculation of an expected value and the calculation of a related probability. The response to part (a) provides a good description of a method for calculating the expected number of owners of car model E who would be included in a simple random sample of 2,000 owners from the population of 297,354 new car owners, and it also provides the correct numerical value. Because it fails to address the standard deviation, it was scored as partially correct. The response to part (b) is very well expressed. It identifies the sample size and the relevant population proportion and clearly describes the use of the binomial distribution to compute the probability of obtaining fewer than 12 owners of car model E in a random sample of 2,000 owners. The correct value of the probability is obtained, and this part was scored as essentially correct. The response to part (c) was scored as incorrect. It says something about multiplying the total sample size "by the proportion of car owners that own each model," but it does not link those values to sample sizes used in taking random samples from the populations of owners for the five car models. This response neither adequately describes a sampling procedure for obtaining at least 12 owners for each car model, nor does it address random selection. Because one part was scored as essentially correct, one part was scored as partially correct and one part was scored as incorrect, this response earned a score of 2.