2. An administrator at a large university wants to conduct a survey to estimate the proportion of students who are satisfied with the appearance of the university buildings and grounds. The administrator is considering three methods of obtaining a sample of 500 students from the 70,000 students at the university.

   (a) Because of financial constraints, the first method the administrator is considering consists of taking a convenience sample to keep the expenses low. A very large number of students will attend the first football game of the season, and the first 500 students who enter the football stadium could be used as a sample. Why might such a sampling method be biased in producing an estimate of the proportion of students who are satisfied with the appearance of the buildings and grounds?

   (b) Because of the large number of students at the university, the second method the administrator is considering consists of using a computer with a random number generator to select a simple random sample of 500 students from a list of 70,000 student names. Describe how to implement such a method.

   (c) Because stratification can often provide a more precise estimate than a simple random sample, the third method the administrator is considering consists of selecting a stratified random sample of 500 students. The university has two campuses with male and female students at each campus. Under what circumstance(s) would stratification by campus provide a more precise estimate of the proportion of students who are satisfied with the appearance of the university buildings and grounds than stratification by gender?

---

3. Each full carton of Grade A eggs consists of 1 randomly selected empty cardboard container and 12 randomly selected eggs. The weights of such full cartons are approximately normally distributed with a mean of 840 grams and a standard deviation of 7.9 grams.

   (a) What is the probability that a randomly selected full carton of Grade A eggs will weigh more than 850 grams?

   (b) The weights of the empty cardboard containers have a mean of 20 grams and a standard deviation of 1.7 grams. It is reasonable to assume independence between the weights of the empty cardboard containers and the weights of the eggs. It is also reasonable to assume independence among the weights of the 12 eggs that are randomly selected for a full carton.

   Let the random variable X be the weight of a single randomly selected Grade A egg.

   i) What is the mean of X?

   ii) What is the standard deviation of X?

GO ON TO THE NEXT PAGE.

4. The Behavioral Risk Factor Surveillance System is an ongoing health survey system that tracks health conditions and risk behaviors in the United States. In one of their studies, a random sample of 8,866 adults answered the question “Do you consume five or more servings of fruits and vegetables per day?” The data are summarized by response and by age-group in the frequency table below.

Age-Group (years)	Yes	No	Total
18–34	231	741	972
35–54	669	2,242	2,911
55 or older	1,291	3,692	4,983
Total	2,191	6,675	8,866

Do the data provide convincing statistical evidence that there is an association between age-group and whether or not a person consumes five or more servings of fruits and vegetables per day for adults in the United States?

GO ON TO THE NEXT PAGE.

Question 3

Intent of Question

The primary goals of this question were to assess a student's ability to (1) calculate a probability from a normal distribution and (2) apply properties of means and variances of functions of random variables.

Solution

Part (a):

Let W denote the weight of a randomly selected full carton of eggs. W has a normal distribution with mean 840 grams and standard deviation 7.9 grams.

The z-score for a weight of 850 grams is $z = \frac{850 - 840}{7.9} \approx 1.27.$

The standard normal probability table reveals that
$P(W > 850) = P(Z > 1.27) \approx 1 - 0.8980 = 0.1020.$

Part (b):

(i) Let W represent the weight of a randomly selected full carton of eggs, P the weight of the packaging, and X_i the weight of the *ith* egg, for $i = 1, 2, \ldots, 12.$

Note that $W = P + X_1 + X_2 + \ldots + X_{12}.$
Properties of expected values establish that $\mathrm{E}(W) = \mathrm{E}(P) + \mathrm{E}(X_1) + \ldots + \mathrm{E}(X_{12}).$
Because all 12 eggs have the same mean weight, this becomes $\mathrm{E}(W) = \mathrm{E}(P) + 12 \times \mathrm{E}(X_i).$
We were told that $\mathrm{E}(W) = 840$ and $\mathrm{E}(P) = 20$, so we can solve
$840 = 20 + 12 \times \mathrm{E}(X_i)$ to find $\mathrm{E}(X_i) = \frac{840 - 20}{12} \approx 68.33$ grams.

(ii) Because of independence, properties of variance establish that
$$\mathrm{Var}(W) = \mathrm{Var}(P) + \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \ldots + \mathrm{Var}(X_{12}).$$

Because all 12 eggs have the same variance of their weights, this becomes
$$\mathrm{Var}(W) = \mathrm{Var}(P) + 12 \times \mathrm{Var}(X_i).$$

We were told that $\mathrm{SD}(W) = 7.9$ and $\mathrm{SD}(P) = 1.7$. Therefore, $\mathrm{Var}(W) = (7.9)^2 = 62.41$ and $\mathrm{Var}(P) = (1.7)^2 = 2.89.$

We can solve $62.41 = 2.89 + 12 \times \mathrm{Var}(X_i)$ to find $\mathrm{Var}(X_i) = \frac{62.41 - 2.89}{12} = 4.96.$ Thus, $\mathrm{SD}(X_i) = \sqrt{(4.96)} \approx 2.23$ grams.

Question 3 (continued)

Scoring

Parts (a), (b-i), and (b-ii) were scored as essentially correct (E), partially correct (P), or incorrect (I). (Minor arithmetic errors in any part were not penalized).

Part (a) is scored as follows:

Essentially correct (E) if the response correctly includes the following three components:

1. Indicates use of a normal distribution and clearly identifies the correct parameter values (using a z-score is sufficient);
2. Uses the correct boundary value;
3. Reports the correct normal probability consistent with components 1 and 2.

Partially correct (P) if the response correctly includes two of the three components listed above.

Incorrect (I) if the response does not satisfy the criteria for an E or a P.

Notes:

1. An error in statistical notation in the response lowers the score one level (that is, from E to P or from P to I).
2. Responses that calculate a probability for a sample mean with n not equal to 1 should be scored an I. For example, using $z = \frac{x - \mu}{\sigma/\sqrt{n}}$, even if the parameters were correctly identified.
3. In component 1, a sketch of a normal curve with the mean labeled is sufficient for indicating use of a normal distribution and identifying the mean.
4. The following were examples of clearly identified parameters for component 1:
   - Writes "$\mu = 840, \sigma = 7.9$."
   - Explicitly labels the mean and standard deviation in a normalcdf calculator statement.
   - Sketches a normal curve, labels 840 as the mean, and labels two additional consecutive values separated by 7.9.
5. For component 3, acceptable correct values were all in the interval from 0.1020 to 0.1038.

Part (b-i) is scored as follows:

Essentially correct (E) if the response correctly uses properties of expected values to set up the correct equation to be solved *AND* correctly solves the equation for the desired expected value
OR
If the response follows a correct numerical procedure to find the correct expected value for one egg.

Partially correct (P) if the response indicates a correct procedure but makes an error in applying properties of expected values.
OR
If the response provides poor communication of the procedure.

Question 3 (continued)

Incorrect (I) if the response does not satisfy the criteria for an E or P.

Note: $\frac{820}{12} = 68.33$ is an example of poor communication, because the two arithmetic steps of subtraction $(840 - 20)$ and division $\left(\frac{820}{12}\right)$ were not documented.

Part (b-ii) is scored as follows:

Essentially correct (E) if the response combines variances and correctly includes the following three components:

1. Subtracts variances
2. Correctly uses the "12" in the calculations
3. Reports the correct standard deviation, consistent with components (1) and (2)

Partially correct (P) if the response combines variances and correctly includes two of the three components listed above.

Incorrect (I) if the response does not satisfy the criteria for an E or P.

Notes:

1. Examples of incorrect calculations with variances that should be scored P (one component incorrect):

$$\sqrt{\frac{7.9^2 + 1.7^2}{12}} = 2.33\text{g}$$

$$\sqrt{7.9^2 - 1.7^2} = 7.71\text{g}$$

$$\frac{\sqrt{7.9^2 - 1.7^2}}{12} = 0.643\text{g}$$

$$\frac{7.9^2 - 1.7^2}{12} = 4.96\text{g}$$

Examples of incorrect calculations with variances that should be scored I (more than one component incorrect):

$$\sqrt{7.9^2 + 1.7^2} = 8.08\text{g}$$

$$\frac{\sqrt{7.9^2 + 1.7^2}}{12} = 0.673\text{g}$$

2. Example of a response that does not combine variances and should be scored I:

$$\sqrt{\frac{7.9^2}{12}} = 2.28$$