AP Stats: Lunchtime Review #5 and #6
Probability

Binomial Distribution

Geometric Distribution

Probability Distribution

Transforming Random Variables

What does probability look like?

Combining Random Variables

Sampling Distribution of...

Chapter 6 Multiple Choice Practice

Directions. *Identify the choice that best completes the statement or answers the question. Check your answers and note your performance when you are finished.*

1. A marketing survey compiled data on the number of cars in households. If X = the number of cars in a randomly selected household, and we omit the rare cases of more than 5 cars, then X has the following probability distribution:

X	0	1	2	3	4	5
$P(X)$	0.24	0.37	0.20	0.11	0.05	0.03

What is the probability that a randomly chosen household has at least two cars?

A. 0.19
B. 0.20
C. 0.29
D. 0.39
E. 0.61

2. What is the expected value of the number of cars in a randomly selected household?

A. 2.5
B. 0.1667
C. 1.45
D. 1
E. Can not be determined

3. A dealer in Las Vegas selects 10 cards from a standard deck of 52 cards. Let Y be the number of diamonds in the 10 cards selected. Which of the following best describes this setting?

A. Y has a binomial distribution with n = 10 observations and probability of success p = 0.25.
B. Y has a binomial distribution with n = 10 observations and probability of success p = 0.25, provided the deck is shuffled well.
C. Y has a binomial distribution with n = 10 observations and probability of success p = 0.25, provided that after selecting a card it is replaced in the deck and the deck is shuffled well before the next card is selected.
D. Y has a geometric distribution with n = 10 observations and probability of success p = 0.25.
E. Y has a geometric distribution with n = 52 observations and probability of success p = 0.25.

4. In the town of Lakeville, the number of cell phones in a household is a random variable W with the following probability distribution:

Value w_i'	0	1	2	3	4	5
Probability p_i	0.1	0.1	0.25	0.3	0.2	0.05

The standard deviation of the number of cell phones in a randomly selected house is

A. 1.7475
B. 1.87
C. 2.5
D. 0.09
E. 2.9575

5. A random variable Y has the following probability distribution:

Y	-1	0	1	2
$P(Y)$	$4C$	$2C$	0.07	0.03

The value of the constant C is:

A. 0.10.
B. 0.15.
C. 0.20.
D. 0.25.
E. 0.75.

6. The variance of the sum of two random variables X and Y is

A. $\sigma_X + \sigma_Y$.
B. $(\sigma_X)^2 + (\sigma_Y)^2$.
C. $\sigma_X + \sigma_Y$, but only if X and Y are independent.
D. $(\sigma_X)^2 + (\sigma_Y)^2$, but only if X and Y are independent.
E. None of these.

7. It is known that about 90% of the widgets made by Buckley Industries meet specifications. Every hour a sample of 18 widgets is selected at random for testing and the number of widgets that meet specifications is recorded. What is the approximate mean and standard deviation of the number of widgets meeting specifications?

A. $\mu = 1.62$; $\sigma = 1.414$
B. $\mu = 1.62$; $\sigma = 1.265$
C. $\mu = 16.2$; $\sigma = 1.62$
D. $\mu = 16.2$; $\sigma = 1.273$
E. $\mu = 16.2$; $\sigma = 4.025$

8. A raffle sells tickets for $10 and offers a prize of $500, $1000, or $2000. Let C be a random variable that represents the prize in the raffle drawing. The probability distribution of C is given below.

Value c_i	$0	$500	$1000	$2000
Probability p_i	0.60	0.05	0.13	0.22

The expected profit when playing the raffle is

A. $145.
B. $585.
C. $865.
D. $635.
E. $485.

9. Let the random variable X represent the amount of money Carl makes tutoring statistics students in the summer. Assume that X is Normal with mean $240 and standard deviation $60. The probability is approximately 0.6 that, in a randomly selected summer, Carl will make less than about

A. $144
B. $216
C. $255
D. $30
E. $360

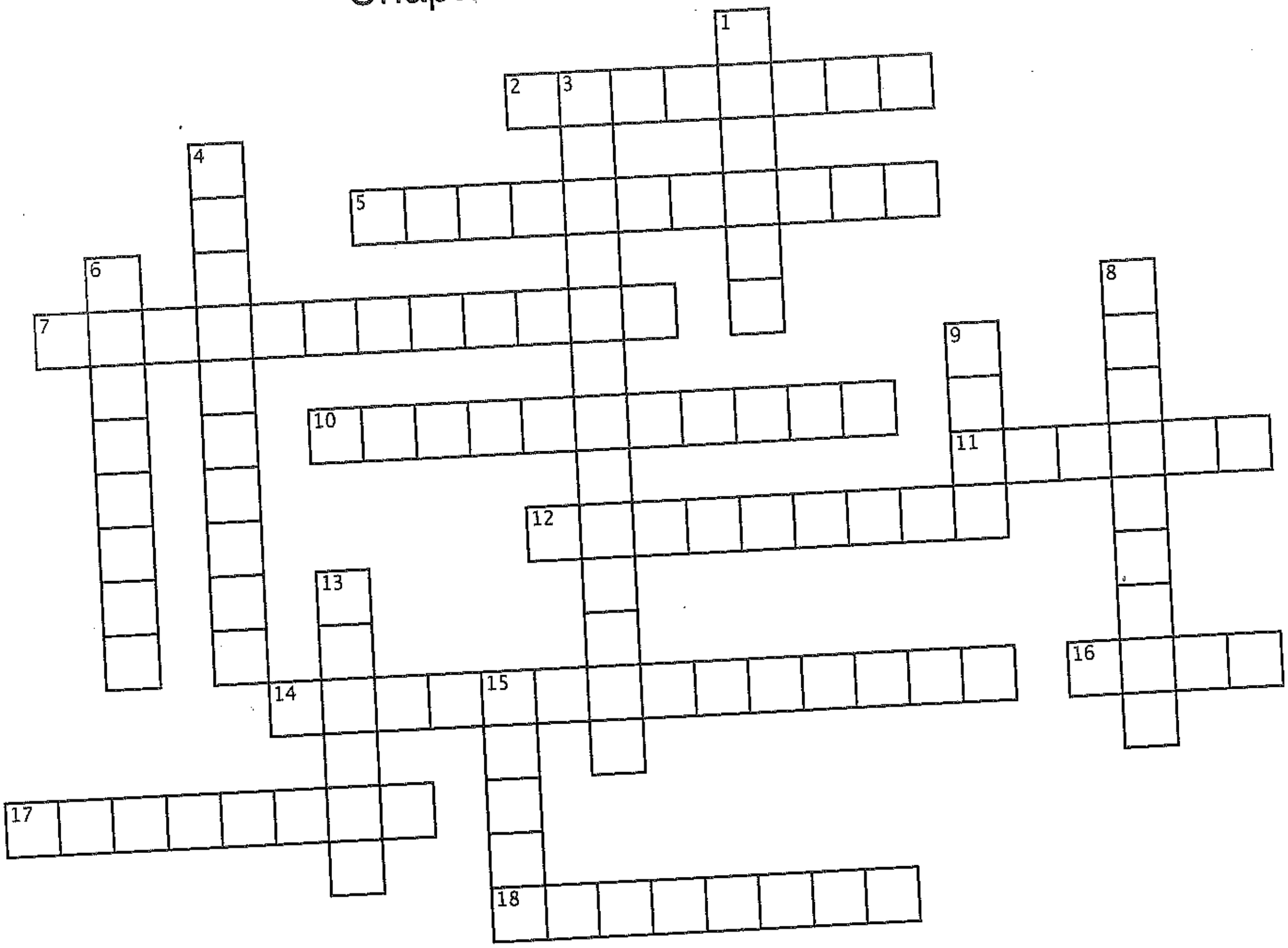
10. Which of the following random variables is geometric?

A. The number of phone calls received in a one-hour period
B. The number of times I have to roll a six-sided die to get two 5s.
C. The number of digits I will read beginning at a randomly selected starting point in a table of random digits until I find a 7.
D. The number of 7s in a row of 40 random digits.
E. All four of the above are geometric random variables.

Multiple Choice Answers

Problem	Answer	Concept	Right	Wrong	Simple Mistake?	Need to Study More
1	D	Discrete Random Variable				
2	C	Expected Value of Discrete Random Variables				
3	C	Binomial Settings				
4	A	Standard Deviation of Discrete Random Variables				
5	B	Probability Distribution				
6	D	Combining Random Variables				
7	D	Binomial Approximations				
8	B	Expected Value				
9	C	Normal Approximations				
10	C	Geometric Random Variables				

Chapter 6: Random Variables



Across

2. The average of the squared deviations of the values of a variable from its mean.
5. Random variables are ______ if knowing whether an event in X has occurred tells us nothing about the occurrence of an event involving Y.
7. The probability ______ of a random variable gives its possible values and their probabilities.
10. The number of ways of arranging k successes among n observations is the binomial _____.
11. The sum or difference of independent Normal random variables follows a ______ distribution.
12. When you combine independent random variable, you always add these.
14. A linear ______ occurs when we add/subtract and multiply/divide by a constant.
16. An easy way to remember the requirements for a geometric setting.
17. This setting arises when we perform several independent trials of a chance process and record the number of times an outcome occurs.
18. The mean of a discrete random variable is also called the ______ value.

Down

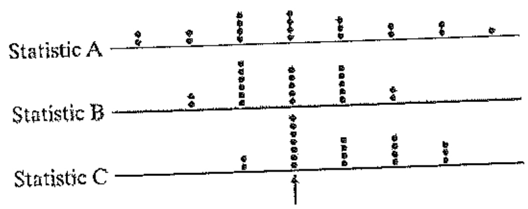
1. A _____ variable takes numerical values that describe the outcomes of some chance process.
3. When n is large, we can use a Normal ______ to determine probabilities for binomial settings.
4. A random variable that takes on all values in an interval of numbers.
6. A random variable that takes a fixed set of possible values with gaps between.
8. A _____ setting arises when we perform independed trials of the same chance process and record the number of trials until a particular outcome occurs.
9. An easy way to remember the requirements for a binmial setting.
13. Adding a constant to each value of a random variable has no effect on the shape or _____ of the distribution.
15. Multiplying each value of a random variable by a constant has no effect on the _____ of the distribution.

Chapter 7 Multiple Choice Practice

Directions. *Identify the choice that best completes the statement or answers the question. Check your answers and note your performance when you are finished.*

1. The variability of a statistic is described by
A. the spread of its sampling distribution.
B. the amount of bias present.
C. the vagueness in the wording of the question used to collect the sample data.
D. probability calculations.
E. the stability of the population it describes.

2. Below are dot plots of the values taken by three different statistics in 30 samples from the same population. The true value of the population parameter is marked with an arrow.



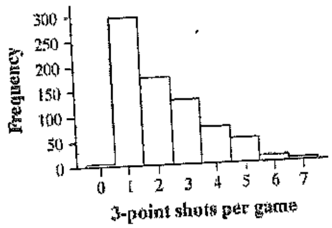
The statistic that has the largest *bias* among these three is
A. statistic A.
B. statistic B.
C. statistic C.
D. A and B have similar bias, and it is larger than the bias of C.
E. B and C have similar bias, and it is larger than the bias of A.

3. According to a recent poll, 27% of Americans prefer to read their news in a physical newspaper instead of online. Let's assume this is the parameter value for the population. If you take a simple random sample of 25 Americans and let $\hat{p}$ = the proportion in the sample who prefer a newspaper, is the shape of the sampling distribution of $\hat{p}$ approximately Normal?
A. No, because $p < 0.50$
B. No, because $np = 6.75$
C. Yes, because we can reasonably assume there are more than 250 individuals in the population.
D. Yes, because we took a simple random sample.
E. Yes, because $n(1-p) = 18.25$

4. The time it takes students to complete a statistics quiz has a mean of 20.5 minutes and a standard deviation of 15.4 minutes. What is the probability that a random sample of 40 students will have a mean completion time greater than 25 minutes?
A. 0.9678
B. 0.0322
C. 0.0344
D. 0.3851
E. 0.6149

5. A fair coin (one for which both the probability of heads and the probability of tails are 0.5) is tossed 60 times. The probability that more than 1/3 of the tosses are heads is closest to
A. 0.9951.
B. 0.33.
C. 0.109.
D. 0.09.
E. 0.0049.

6. The histogram below was obtained from data on 750 high school basketball games in a regional athletic conference. It represents the number of three-point baskets made in each game.



What is the range of sample sizes a researcher could take from this population without violating conditions required for performing Normal calculations with the sampling distribution of $\bar{x}$?

A. $0 \leq n \leq 30$
B. $30 \leq n \leq 50$
C. $30 \leq n \leq 75$
D. $30 \leq n \leq 750$
E. $75 \leq n \leq 750$

7. The incomes in a certain large population of college teachers have a normal distribution with mean \$60,000 and standard deviation \$5000. Four teachers are selected at random from this population to serve on a salary review committee. What is the probability that their average salary exceeds \$65,000?

A. 0.0228
B. 0.1587
C. 0.8413
D. 0.9772
E. essentially 0

8. A random sample of size 25 is to be taken from a population that is Normally distributed with mean 60 and standard deviation 10. The mean $\bar{x}$ of the observations in our sample is to be computed. The sampling distribution of $\bar{x}$

A. is Normal with mean 60 and standard deviation 10.
B. is Normal with mean 60 and standard deviation 2.
C. is approximately Normal with mean 60 and standard deviation 2.
D. has an unknown shape with mean 60 and standard deviation 10.
E. has an unknown shape with mean 60 and standard deviation 2.

9. The scores of individual students on a college entrance examination have a left-skewed distribution with mean 18.6 and standard deviation 6.0. At Milllard North High School, 36 seniors take the test. The sampling distribution of mean scores for random samples of 36 students is

A. approximately Normal.
B. symmetric and mound-shaped, but non-Normal.
C. skewed right.
D. neither Normal nor non-normal. It depends on the particular 36 students selected.
E. exactly Normal.

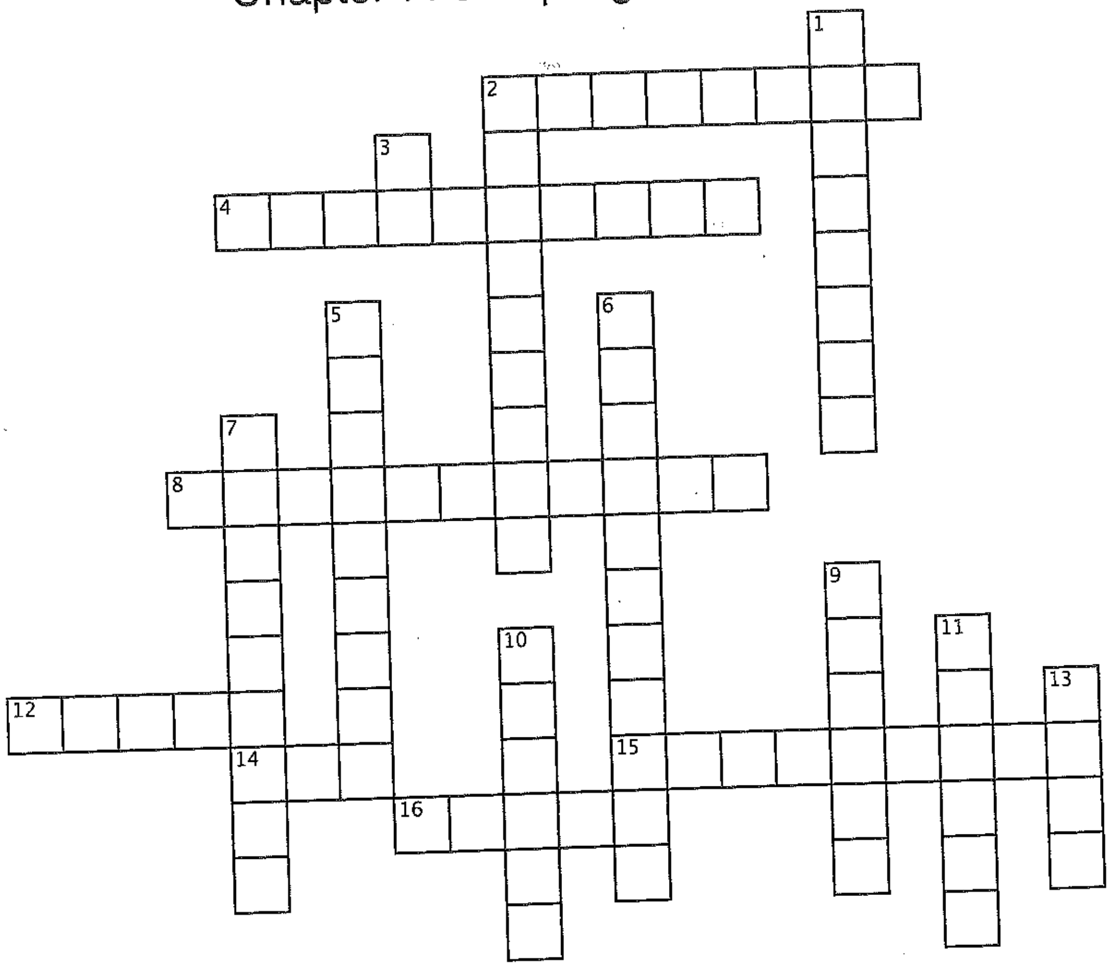
10. The distribution of prices for home sales in Minnesota is skewed to the right with a mean of \$290,000 and a standard deviation of \$145,000. Suppose you take a simple random sample of 100 home sales from this (very large) population. What is the probability that the mean of the sample is above \$325,000?

A. 0.0015
B. 0.0027
C. 0.0079
D. 0.4046
E. 0.4921

Multiple Choice Answers

Problem	Answer	Concept	Right	Wrong	Simple Mistake?	Need to Study More
1	A	Sampling Variability				
2	C	Bias and Variability				
3	B	Normality Condition				
4	B	Normal Probability Calculation				
5	A	Normal Probability Calculation				
6	C	10% Condition and CLT				
7	A	Normal Probability Calculation				
8	B	Sampling Distribution for Means				
9	A	Sampling Distribution for Means				
10	C	Normal Probability Calculation				

Chapter 7: Sampling Distributions



Across

2. _____ distribution: the distribution of values taken by the statistic in all possible samples of the same size from the population
4. ______ distribution: the distribution of all values of a variable in the population
8. _____ of a statistic is described by the spread of the sampling distribution
12. Greek letter used for the population standard deviation
14. the Normal approximation for the sampling distribution of a sample proportion can be used when both the number of successes and failures are greater than _____
15. sampling distributions and sampling variability provide the foundation for performing ______
16. central _____ theorem tells us if the sample size is large, the sampling distribution of the sample mean is approximately Normal, regardless of the shape of the population

Down

1. a statistic is an _____ estimator if the mean of the sampling distribution is equal to the true value of the parameter being estimated.
2. a number, computed from sample data, that estimates a parameter
3. Greek letter used for the population mean
5. standard ______ : measure of spread of a sampling distribution
6. sampling ______ notes the value of a statistic may be different from sample to sample
7. a number that describes a population
9. the rule of thumb for using the central limit theorem - the sample size should be greater than ______
10. when the sample size is large, the sampling distribution of a sample proportion is approximately ______
11. to draw a conclusion about a population parameter, we can look at information from a _____ sample
13. center of a sampling distribution

2. When a law firm represents a group of people in a class action lawsuit and wins that lawsuit, the firm receives a percentage of the group's monetary settlement. That settlement amount is based on the total number of people in the group—the larger the group and the larger the settlement, the more money the firm will receive.

   A law firm is trying to decide whether to represent car owners in a class action lawsuit against the manufacturer of a certain make and model for a particular defect. If 5 percent or less of the cars of this make and model have the defect, the firm will not recover its expenses. Therefore, the firm will handle the lawsuit only if it is convinced that more than 5 percent of cars of this make and model have the defect. The firm plans to take a random sample of 1,000 people who bought this car and ask them if they experienced this defect in their cars.

   (a) Define the parameter of interest and state the null and alternative hypotheses that the law firm should test.

   (b) In the context of this situation, describe Type I and Type II errors and describe the consequences of each of these for the law firm.

3. Men's shirt sizes are determined by their neck sizes. Suppose that men's neck sizes are approximately normally distributed with mean 15.7 inches and standard deviation 0.7 inch. A retailer sells men's shirts in sizes S, M, L, XL, where the shirt sizes are defined in the table below.

Shirt size	Neck size
S	$14 \leq$ neck size < 15
M	$15 \leq$ neck size < 16
L	$16 \leq$ neck size < 17
XL	$17 \leq$ neck size < 18

   (a) Because the retailer only stocks the sizes listed above, what proportion of customers will find that the retailer does not carry any shirts in their sizes? Show your work.

   (b) Using a sketch of a normal curve, illustrate the proportion of men whose shirt size is M. Calculate this proportion.

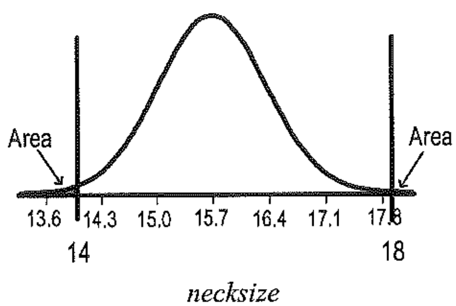
   (c) Of 12 randomly selected customers, what is the probability that exactly 4 will request size M ? Show your work.

Question 3

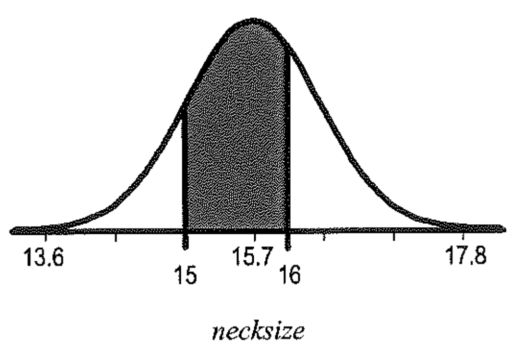
Solution

Part (a):

$P(necksize < 14 \;\; or \;\; necksize \geq 18)$
$= P(necksize < 14) + P(necksize \geq 18)$
$= P\left(z < \frac{14-15.7}{0.7}\right) + P\left(z \geq \frac{18-15.7}{0.7}\right)$
$= P(z < -2.429) + P(z \geq 3.286)$
$= 0.00758 + 0.00051$
$= 0.00809$



Part (b):



$P(15 \leq necksize < 16)$
$= P\left(\frac{15-15.7}{0.7} \leq z < \frac{16-15.7}{0.7}\right)$
$= P(-1.000 \leq z < 0.429)$
$= 0.50723$

Part (c):

X = number of customers who request size M
X is binomial with $n = 12$ customers and $p = 0.5072$
$P(X = 4) = {}_{12}C_4(0.5072)^4(0.4928)^8 = 495(0.06618)(0.00348) = 0.1139$

AP® STATISTICS
2003 SCORING GUIDELINES

Question 3 (cont'd)

Scoring

Each part is scored as either essentially correct (E), partially correct (P), or incorrect (I).

Part (a) is essentially correct (E) if the response
1. recognizes the need to look at neck sizes below 14 and above 18
2. correctly computes the two tail probabilities (except for minor arithmetic or transcription errors) and adds those probabilities

Part (a) is partially correct (P) if the response

considers only neck sizes below 14 (or above 18) but computes that corresponding tail area correctly
OR
recognizes the need to look at neck sizes below 14 and above 18 but does not compute both tail probabilities correctly
OR
recognizes the need to look at neck sizes below 14 and above 18 but approximates tail probabilities using the Empirical Rule
OR
computes the proportion of customers that will find the store carries their size
(i.e., 1 – correct answer)
OR
States the correct answer (0.0081) without supporting work

NOTE: A normal curve with correct regions shaded showing both correct end points (14 and 18) and the mean and the standard deviation may be used for element 1.

Part (b) is essentially correct (E) if
1. the appropriate probability is illustrated using a normal curve in which the end points are identified and the mean and standard deviation are implied
2. the required probability is correctly computed (except for minor arithmetic errors)

Part (b) is partially correct (P) if only one of the above elements is correct.

NOTES:

(1) If part (a) was not essentially correct because the student interchanged the mean and standard deviation, and the same values for mean and standard deviation are used in part (b), then part (b) can be considered essentially correct if the probability calculated is correct for the mean and standard deviation used.

(2) A reasonable approximation using the Empirical Rule in part (b) is only acceptable if the computation in part (a) is done correctly (i.e., without using the Empirical Rule).

(1)

Question 3 (cont'd)

Part (c) is essentially correct (E) if

1. the student recognized the setting as binomial
2. the probability calculated in part (b) is used for p
3. work is shown — that is, the correct values for n and x are given and the desired probability calculated, or the binomial formula is correctly evaluated.

Part (c) is partially correct (P) if

the student recognizes the situation as binomial and identifies p from part (b) but does not compute the desired probability

OR

the student computes the probability as either $(0.5072)^4\,(0.4928)^8$ or $\binom{12}{4}(0.5072)^4$

OR

the student gives the correct probability of 0.1139 but work is not shown

NOTE: Rounding the probability in part (b) for use in part (c) is acceptable.

4 **Complete Response (3E)**

All three parts essentially correct

3 **Substantial Response (2E 1P)**

Two parts essentially correct and one part partially correct

2 **Developing Response (2E 0P or 1E 2P or 3P)**

Two parts essentially correct and no parts partially correct
OR
One part essentially correct and two parts partially correct
OR
Three parts partially correct

1 **Minimal Response (1E 1P or 1E 0P or 0E 2P)**

One part essentially correct and either zero or one parts partially correct
OR
No parts essentially correct and two parts partially correct

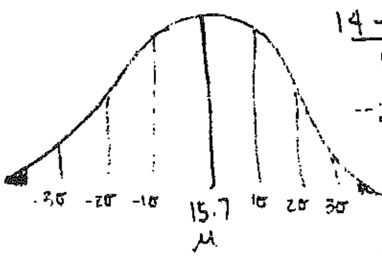
1

3. Men's shirt sizes are determined by their neck sizes. Suppose that men's neck sizes are approximately normally distributed with mean 15.7 inches and standard deviation 0.7 inch. A retailer sells men's shirts in sizes S, M, L, XL, where the shirt sizes are defined in the table below.

A

Shirt size	Neck size
S	14 ≤ neck size < 15
M	15 ≤ neck size < 16
L	16 ≤ neck size < 17
XL	17 ≤ neck size < 18

(a) Because the retailer only stocks the sizes listed above, what proportion of customers will find that the retailer does not carry any shirts in their sizes? Show your work.



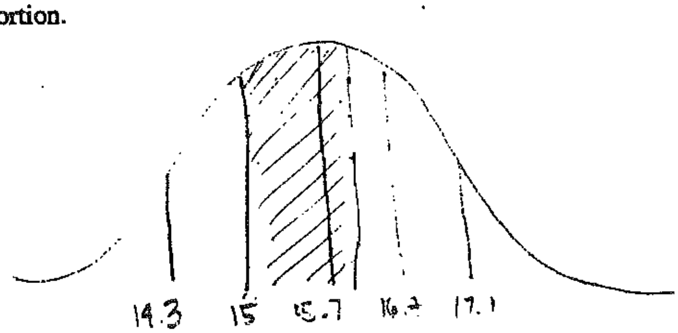
$\frac{14-15.7}{0.7} < z < \frac{18-15.7}{0.7}$

$-2.43 < z < 3.29$

P = .9995 - .0075 = .992

1 - P = shirts not carried = .008

(b) Using a sketch of a normal curve, illustrate the proportion of men whose shirt size is M. Calculate this proportion.



$\frac{15-15.7}{0.7} < z < \frac{16-15.7}{0.7}$

$-1 < z < .4286$

P = .6628 - .1587

P = .5041

(c) Of 12 randomly selected customers, what is the probability that exactly 4 will request size M ? Show your work.

$P(X=4) = \binom{12}{4}(.5041)^4(.4959)^8$

= 495(.06457)(.003657)

= .1169

P = .12

Binomial setting

1. all observations are independent
2. either success or failure
3. all observations have probability, p, of success
4. fixed number for n

GO ON TO THE NEXT PAGE.

(1)

B

3. Men's shirt sizes are determined by their neck sizes. Suppose that men's neck sizes are approximately normally distributed with mean 15.7 inches and standard deviation 0.7 inch. A retailer sells men's shirts in sizes S, M, L, XL, where the shirt sizes are defined in the table below.

Shirt size	Neck size
S	14 ≤ neck size < 15
M	15 ≤ neck size < 16
L	16 ≤ neck size < 17
XL	17 ≤ neck size < 18

(a) Because the retailer only stocks the sizes listed above, what proportion of customers will find that the retailer does not carry any shirts in their sizes? Show your work.

1 − normalcdf(14, 18, 15.7, 0.7) ≅ [0.0081] = 0.81%

I found the proportion of people within the necksize 14 to 18 because that is what retailers carry. Subtracting from one to obtain the proportion that retailers do not carry, I found that .0081 of population of customers are out of the range of shirt sizes that retailer carries.

(b) Using a sketch of a normal curve, illustrate the proportion of men whose shirt size is M. Calculate this proportion.

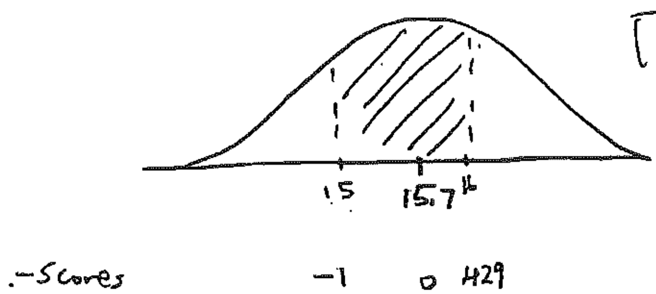
normalcdf(15, 16, 15.7, 0.7) ≅ [.5072]

[or]

$P(x<16) = P(z< \frac{16-15.7}{.7}) = P(z<.429)$
$= .6664$

$P(x \le 15) = P(z< \frac{15-15.7}{.7}) = P(z<-1)$
$= .1587$

$P(x<16) - P(x<15) = P(15 \le x < 16)$
$.6664 - .1587 = [.5077]$



(c) Of 12 randomly selected customers, what is the probability that exactly 4 will request size M? Show your work.

$_{12}C_4 (.5072)^4 (.4928)^8 \cong$ [.11394] = 11.394%

This is a binomial distribution with success probability of .5072, 12 trials, and expect exactly 4 successes.

binompdf(12, .5072, 4) = .11394

GO ON TO THE NEXT PAGE.

2006 AP® STATISTICS FREE-RESPONSE QUESTIONS

3. The depth from the surface of Earth to a refracting layer beneath the surface can be estimated using methods developed by seismologists. One method is based on the time required for vibrations to travel from a distant explosion to a receiving point. The depth measurement (M) is the sum of the true depth (D) and the random measurement error (E). That is, $M = D + E$. The measurement error (E) is assumed to be normally distributed with mean 0 feet and standard deviation 1.5 feet.

   (a) If the true depth at a certain point is 2 feet, what is the probability that the depth measurement will be negative?

   (b) Suppose three independent depth measurements are taken at the point where the true depth is 2 feet. What is the probability that <u>at least one</u> of these measurements will be negative?

   (c) What is the probability that the <u>mean</u> of the three independent depth measurements taken at the point where the true depth is 2 feet will be negative?

4. Patients with heart-attack symptoms arrive at an emergency room either by ambulance or self-transportation provided by themselves, family, or friends. When a patient arrives at the emergency room, the time of arrival is recorded. The time when the patient's diagnostic treatment begins is also recorded.

   An administrator of a large hospital wanted to determine whether the mean wait time (time between arrival and diagnostic treatment) for patients with heart-attack symptoms differs according to the mode of transportation. A random sample of 150 patients with heart-attack symptoms who had reported to the emergency room was selected. For each patient, the mode of transportation and wait time were recorded. Summary statistics for each mode of transportation are shown in the table below.

Mode of Transportation	Sample Size	Mean Wait Time (in minutes)	Standard Deviation of Wait Times (in minutes)
Ambulance	77	6.04	4.30
Self	73	8.30	5.16

   (a) Use a 99 percent confidence interval to estimate the difference between the mean wait times for ambulance-transported patients and self-transported patients at this emergency room.

   (b) Based only on this confidence interval, do you think the difference in the mean wait times is statistically significant? Justify your answer.

GO ON TO THE NEXT PAGE.

2

Question 3

Intent of Question

The primary goals of this question are to assess a student's ability to: (1) recognize the random variable of interest, identify its probability distribution, and calculate a probability for a linear combination of a normal random variable and a constant; (2) use basic probability rules to find a different probability; and (3) use the sampling distribution of the sample mean to find a probability about the mean of three observations.

Solution

Part (a):

Since $M = D + E$ (a normal random variable plus a constant is a normal random variable), we know that M is normally distributed with a mean of 2 feet and a standard deviation of 1.5 feet. Thus,

$$P(M < 0) = P\left(Z < \frac{0-2}{1.5}\right) < P(Z < -1.33) = 0.0918\text{, where } Z = \frac{M-\mu}{\sigma}.$$

Part (b):

$$\begin{aligned} P(\text{at least one measurement} < 0) &= 1 - P(\text{all three measurements} \geq 0) \\ &= 1-(1-0.0918)^3 \\ &= 1-(0.9082)^3 \\ &= 1-0.7491 \\ &= 0.2509 \end{aligned}$$

Part (c):

Let $\overline{X}$ denote the mean of three independent depth measurements taken at a point where the true depth is 2 feet. Since each measurement comes from a normal distribution, the distribution of $\overline{X}$ is normal with a mean of 2 feet and a standard deviation of $\frac{1.5}{\sqrt{3}} = 0.8660$ feet. Thus,

$$P(\overline{X} < 0) = P\left(Z < \frac{0-2}{\frac{1.5}{\sqrt{3}}}\right) < P(Z < -2.31) = 0.0104\text{, where } Z = \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}.$$

AP® STATISTICS
2006 SCORING GUIDELINES

Question 3 (continued)

Scoring

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

Part (a) is essentially correct (E) if the student clearly does ALL three of the following:

- identifies the distribution as normal;
- specifies BOTH μ and σ; AND
- calculates the correct probability.

Part (a) is partially correct (P) if the student:

- calculates the correct probability but fails to identify the distribution as normal with BOTH μ and σ specified;

  OR
- correctly identifies the distribution as normal with BOTH μ and σ specified but fails to calculate the correct probability.

Part (a) is incorrect (I) if any of the following occur:

- the student indicates the probability is 0.5 because the random error is symmetric about zero;

  OR
- the student uses a mean of zero and a standard deviation of 1;

  OR
- the student conducts a hypothesis test.

Notes:

- The student may use the distribution of the error, E, to solve the problem. That is, finding the area below -2 for a normal distribution with mean 0 and standard deviation 1.5 should be scored essentially correct (E).

  Thus $P(E < -2) = P\left(Z < \frac{-2-0}{1.5}\right) < P(Z < -1.33) = 0.0918$, where $Z = \frac{E-\mu}{\sigma}$.
- If only the calculator command normalcdf $(-\infty, 0, 2, 1.5)$ is provided along with the probability 0.0912, then the response should be scored as partially correct (P).

Part (b) is essentially correct (E) if the student calculates the correct probability AND:

- correctly applies complement and probability rules using the value obtained in part (a);

  OR
- clearly identifies the distribution as binomial AND specifies BOTH n and p using the value obtained in part (a).

Part (b) is partially correct (P) if the student:

- clearly identifies the distribution as binomial AND specifies BOTH n and p, using the value obtained in part (a), but does not calculate the correct probability;

  OR

AP® STATISTICS
2006 SCORING GUIDELINES

Question 3 (continued)

- clearly identifies the distribution as binomial AND specifies BOTH n and p using a value of p that is unrelated to the value obtained in part (a) and calculates the correct probability based on their value of p;

  OR
- calculates the correct probability using the value obtained in part (a) but fails to correctly identify the distribution as binomial with BOTH n and p specified;

  OR
- recognizes the solution as the sum of the product of the probabilities of successes and failures, using the answer from part (a), but omits only the binomial coefficients.

Part (b) is incorrect (I) if the student calculates $P(\text{at least one measurement} < 0) = 1 - p^3$, where p is the solution to part (a).

Notes:

- The solution using the binomial distribution with $p = 0.0918$ is:

$$\begin{aligned} P(\text{at least one measurement} < 0) &= P(B=1) + P(B=2) + P(B=3) \\ &= \binom{3}{1} 0.0918^1 (1-0.0918)^2 + \\ &\quad \binom{3}{2} 0.0918^2 (1-0.0918)^1 + \binom{3}{3} 0.0918^3 \\ &= 0.2272 + 0.0230 + 0.0008 \\ &= 0.2510 \end{aligned}$$

- If only the calculator command $1 - \text{binomcdf}(3, 0.0918, 0)$ is provided along with the probability 0.2509, then the response should be scored as partially correct (P).

Part (c) is essentially correct (E) if the student clearly does ALL three of the following:

- identifies the distribution of the sample mean as normal;
- specifies BOTH $\mu_{\bar{x}}$ and $\sigma_{\bar{x}}$; AND
- calculates the correct probability.

Part (c) is partially correct (P) if the student:

- calculates the correct probability, but fails to identify the distribution of the sample mean as normal with BOTH $\mu_{\bar{x}}$ and $\sigma_{\bar{x}}$ specified;

  OR
- correctly identifies the distribution of the sample mean as normal with BOTH $\mu_{\bar{x}}$ and $\sigma_{\bar{x}}$ specified, but fails to calculate the correct probability.

Part (c) is incorrect (I) if any of the following occur:

- the student uses the same calculation as in part (a);

  OR

AP® STATISTICS
2006 SCORING GUIDELINES

Question 3 (continued)

- the student uses an incorrect standard deviation (e.g., $\frac{1.5}{\sqrt{2}}$ or $\sqrt{3(1.5)}$;

  OR
- the student conducts a hypothesis test.

Notes:

- An alternate solution using the sum instead of the mean is:
  Let $\bar{X}$ denote the mean of three independent depth measurements taken at a point where the true depth is 2 feet. Since each measurement comes from a normal distribution, the distribution of the sum of the three measurements, $S = (X_1 + X_2 + X_3)$, is normal with a mean $\mu_S = 6$ feet and a standard deviation $\sigma_S = 2.598$ feet $\left(\sigma_S = 3\left(\frac{1.5}{\sqrt{3}}\right)\right.$, often calculated as $\left.\sqrt{(1.5)^2 + (1.5)^2 + (1.5)^2}\right)$.

  Thus $P(S < 0) = P\left(Z < \frac{0-6}{2.598}\right) = P(Z < -2.31) = 0.0104$, where $Z = \frac{S - \mu_S}{\sigma_S}$.

- If only the calculator command normalcdf $(-\infty, 0, 2, 0.866)$ is provided along with the probability 0.01046, then the response should be scored as partially correct (P).
- If the student does not consistently specify a correct μ and σ from the same distribution, i.e., for the mean or the sum, the response should be scored at most partially correct (P).

Question 3 (continued)

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 no parts partially correct

OR

One part essentially correct and two parts partially correct

OR

Three parts partially correct

1 **Minimal Response**

One part essentially correct and one part partially correct

OR

One part essentially correct and no parts partially correct

OR

No parts essentially correct and two parts partially correct

③

2. As dogs age, diminished joint and hip health may lead to joint pain and thus reduce a dog's activity level. Such a reduction in activity can lead to other health concerns such as weight gain and lethargy due to lack of exercise. A study is to be conducted to see which of two dietary supplements, glucosamine or chondroitin, is more effective in promoting joint and hip health and reducing the onset of canine osteoarthritis. Researchers will randomly select a total of 300 dogs from ten different large veterinary practices around the country. All of the dogs are more than 6 years old, and their owners have given consent to participate in the study. Changes in joint and hip health will be evaluated after 6 months of treatment.

   (a) What would be an advantage to adding a control group in the design of this study?

   (b) Assuming a control group is added to the other two groups in the study, explain how you would assign the 300 dogs to these three groups for a completely randomized design.

   (c) Rather than using a completely randomized design, one group of researchers proposes blocking on clinics, and another group of researchers proposes blocking on breed of dog. How would you decide which <u>one</u> of these two variables to use as a blocking variable?

3. Big Town Fisheries recently stocked a new lake in a city park with 2,000 fish of various sizes. The distribution of the lengths of these fish is approximately normal.

   (a) Big Town Fisheries claims that the mean length of the fish is 8 inches. If the claim is true, which of the following would be more likely?

   - A random sample of 15 fish having a mean length that is greater than 10 inches

   or

   - A random sample of 50 fish having a mean length that is greater than 10 inches

   Justify your answer.

   (b) Suppose the standard deviation of the sampling distribution of the sample mean for random samples of size 50 is 0.3 inch. If the mean length of the fish is 8 inches, use the normal distribution to compute the probability that a random sample of 50 fish will have a mean length less than 7.5 inches.

   (c) Suppose the distribution of fish lengths in this lake was nonnormal but had the same mean and standard deviation. Would it still be appropriate to use the normal distribution to compute the probability in part (b) ? Justify your answer.

GO ON TO THE NEXT PAGE.

AP® STATISTICS
2007 SCORING GUIDELINES

Question 3

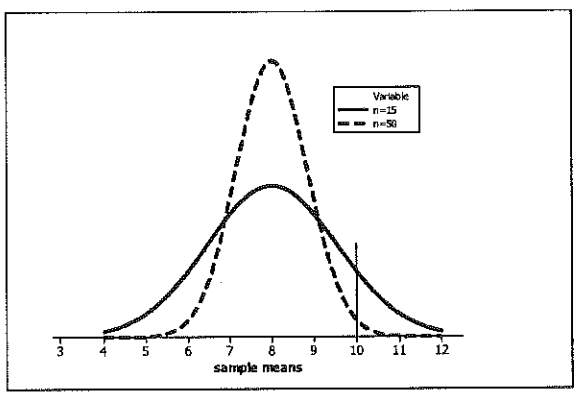
Intent of Question

This question was developed to assess a student's understanding of the sampling distribution of the sample mean: in particular, a student's ability to: (1) compare probabilities concerning sample means from different sample sizes; (2) compute an appropriate probability; and (3) recognize that an application of the Central Limit Theorem is being evaluated.

Solution

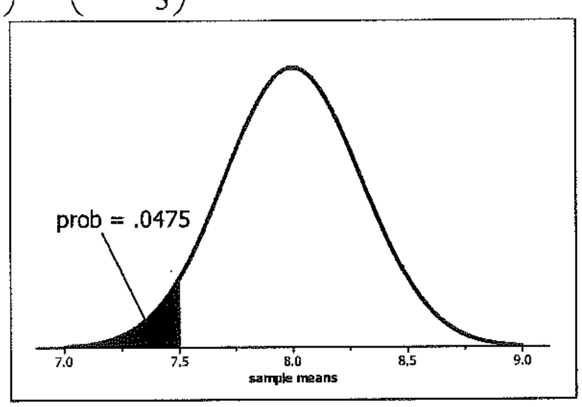
Part (a):

The random sample of $n = 15$ fish is more likely to have a sample mean length greater than 10 inches. The sampling distribution of the sample mean $\bar{x}$ is normal with mean $\mu = 8$ and standard deviation $\sigma/\sqrt{n}$. Thus, both sampling distributions will be centered at 8 inches, but the sampling distribution of the sample mean when $n = 15$ will have more variability than the sampling distribution of the sample mean when $n = 50$. The tail area ($\bar{x} > 10$) will be larger for the distribution that is less concentrated about the mean of 8 inches when the sample size is $n = 15$, as shown in the following graph.



Part (b):

$$P(\bar{x} < 7.5) = P\left(z < \frac{7.5-8}{0.3}\right) = P\left(z < -\frac{5}{3}\right) = P(z < -1.67) = 0.0475$$



AP® STATISTICS
2007 SCORING GUIDELINES

Question 3 (continued)

Part (c):

Yes. The Central Limit Theorem says that the sampling distribution of the sample mean will become approximately normal as the sample size n increases. Since the sample size is reasonably large ($n = 50$), the calculation in part (b) will provide a good approximation to the probability of interest even though the population is nonnormal.

Scoring

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

Part (a) is scored as essentially correct (E) if the student says that the sample of 15 fish is more likely to have a mean length that is greater than 10, AND the justification is based on *variability* in the *sampling* distributions.

Part (a) is scored as partially correct (P) if:

the student makes correct statements about the sampling distribution of the sample mean or the probabilities but does not specifically refer to the variability in these two sampling distributions;

OR

the student remarks that the sample mean approaches the population mean as the sample size increases (an argument based on the Law of Large Numbers).

Some examples of partially correct statements are:

- With the smaller sample size we will be more likely to get an extreme value for the sample mean.
- Variability in the smaller sample is larger.
- Variability in the larger sample is smaller.
- The sample mean approaches the population mean as the sample size increases.

Part (a) is scored as incorrect (I) if an answer is provided with no justification or incorrect justification.

Note: If a student chooses a particular value for a standard deviation and goes through the correct calculations, or shows the result algebraically based on a generic standard deviation, then the response should be scored essentially correct.

Part (b) is scored essentially correct (E) if the probability is calculated correctly and a reasonable sketch or evidence of calculation is shown.

Part (b) is scored partially correct (P) if:

an incorrect but plausible calculation is shown. Examples include using an incorrect standard deviation (such as $0.3/\sqrt{50}$) to obtain the probability;

OR

the student switches the sample mean and the population mean to obtain a standardized z value of 1.67.

AP® STATISTICS
2007 SCORING GUIDELINES

Question 3 (continued)

Part (b) is scored incorrect (I) if an answer is provided with no justification or incorrect justification.

Note: Normalcdf (…) with no additional work is at best partially correct. If an appropriate sketch accompanies the calculator command, *OR* if the components of the calculator command are clearly identified/labeled, then the solution should be scored essentially correct.

Part (c) is scored as essentially correct (E) if the student says that the probability is a reasonable approximation because of the Central Limit Theorem and also refers to the large sample size in this case.

Part (c) is scored partially correct (P) if the student indicates that the response in part (b) would not change but provides a weak justification. Examples of a weak justification include mentioning CLT without reference to sample size, and mentioning sample size without reference to CLT.

Part (c) is scored incorrect (I) if an answer is provided with no justification or incorrect justification.

Note: An E counts for 2 points in part (a), and an E counts for 1 point in each of parts (b) and (c). Similarly, a P counts for 1 point in part (a), and a P counts for ½ point in parts (b) and (c). When the total number of points earned is not an integer, the final score earned will be rounded down to the integer value.

4 Complete Response

4 points earned

3 Substantial Response

3 or 3½ points earned

2 Developing Response

2 or 2½ points earned

1 Minimal Response

1 or 1½ points earned

2014 AP® STATISTICS FREE-RESPONSE QUESTIONS

2. Nine sales representatives, 6 men and 3 women, at a small company wanted to attend a national convention. There were only enough travel funds to send 3 people. The manager selected 3 people to attend and stated that the people were selected at random. The 3 people selected were women. There were concerns that no men were selected to attend the convention.

   (a) Calculate the probability that randomly selecting 3 people from a group of 6 men and 3 women will result in selecting 3 women.

   (b) Based on your answer to part (a), is there reason to doubt the manager's claim that the 3 people were selected at random? Explain.

   (c) An alternative to calculating the exact probability is to conduct a simulation to estimate the probability. A proposed simulation process is described below.

   > Each trial in the simulation consists of rolling three fair, six-sided dice, one die for each of the convention attendees. For each die, rolling a 1, 2, 3, or 4 represents selecting a man; rolling a 5 or 6 represents selecting a woman. After 1,000 trials, the number of times the dice indicate selecting 3 women is recorded.

   Does the proposed process correctly simulate the random selection of 3 women from a group of 9 people consisting of 6 men and 3 women? Explain why or why not.

---

3. Schools in a certain state receive funding based on the number of students who attend the school. To determine the number of students who attend a school, one school day is selected at random and the number of students in attendance that day is counted and used for funding purposes. The daily number of absences at High School A in the state is approximately normally distributed with mean of 120 students and standard deviation of 10.5 students.

   (a) If more than 140 students are absent on the day the attendance count is taken for funding purposes, the school will lose some of its state funding in the subsequent year. Approximately what is the probability that High School A will lose some state funding?

   (b) The principals' association in the state suggests that instead of choosing one day at random, the state should choose 3 days at random. With the suggested plan, High School A would lose some of its state funding in the subsequent year if the mean number of students absent for the 3 days is greater than 140. Would High School A be more likely, less likely, or equally likely to lose funding using the suggested plan compared to the plan described in part (a)? Justify your choice.

   (c) A typical school week consists of the days Monday, Tuesday, Wednesday, Thursday, and Friday. The principal at High School A believes that the number of absences tends to be greater on Mondays and Fridays, and there is concern that the school will lose state funding if the attendance count occurs on a Monday or Friday. If one school day is chosen at random from each of 3 typical school weeks, what is the probability that none of the 3 days chosen is a Tuesday, Wednesday, or Thursday?

AP® STATISTICS
2014 SCORING GUIDELINES

Question 2

Intent of Question

The primary goals of this question were to assess a student's ability to (1) calculate a probability; (2) assess whether a claim about randomness is questionable in light of a calculated probability; and (3) determine whether a description of a simulation method achieves a correct simulation of a random process.

Solution

Part (a):

The probability that all 3 people selected are women can be calculated using the multiplication rule, as follows:

P(all three selected are women)

$= P(\text{first is a woman}) \times P(\text{second is a woman}|\text{first is a woman}) \times P(\text{third is a woman}|\text{first two are women})$

$= \frac{3}{9} \times \frac{2}{8} \times \frac{1}{7} \approx 0.012$

Part (b):

The probability calculated in part (a) does provide a reason to doubt the manager's claim that the selections were made at random. The calculation shows that there is only about a 1.2% chance that random selection would have resulted in three women being selected. The probability is small enough that it may cast doubt on the manager's claim that the selections were made at random.

Part (c):

No, the process does not correctly simulate the random selection of three women from a group of nine people of whom six are men and three are women. The random selection of three people among nine is done *without* replacement. However, in the simulation with the dice, the three dice rolls in any given trial are independent of one another, indicating a selection process that is done *with* replacement.

AP® STATISTICS
2014 SCORING GUIDELINES

Question 2 (continued)

Scoring

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

Part (a) is scored as follows:

Essentially correct (E) if the response correctly computes the probability of selecting the three women, and shows how the probability was computed.

Partially correct (P) if the response shows only one of the following:

Gives the correct probability of $\frac{1}{84}$ (0.012 or 0.011 is acceptable) but does not show how it was computed;

OR

Correctly shows how the probability should be computed, but does not carry the computation through correctly;

OR

Correctly computes (showing work) only the numerator, or only the denominator of the correct answer. (For example, $\frac{1}{9}\times\frac{1}{8}\times\frac{1}{7}\approx 0.002$, or $\frac{3}{9}\times\frac{2}{9}\times\frac{1}{9}\approx 0.008$, or $\frac{3}{9}\times\frac{3}{8}\times\frac{3}{7}\approx 0.054$);

OR

Mistakenly assumes independence and calculates (showing work) the binomial probability $\frac{1}{3}\times\frac{1}{3}\times\frac{1}{3}=\frac{1}{27}\approx 0.037$.

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

Part (b) is scored as follows:

Essentially correct (E) if the response states that the probability from part (a) is small (or insufficiently small), makes an appropriate decision consistent with the probability being small (or insufficiently small), and does so in the context of this situation.

Partially correct (P) if the response shows only one of the following:

Otherwise satisfies the criteria for an E but does so without any context;

OR

States a significance level and makes a decision in context that is appropriate to the given probability in part (a) and the stated significance level, but does not explicitly compare the probability and the significance level;

OR

Otherwise satisfies the criteria for an E but does not explicitly make a decision about whether there is reason to doubt the manager's claim. (For example: "The probability of selecting the three women from among the nine employees is very small so it is unlikely to occur by chance.")

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

AP® STATISTICS
2014 SCORING GUIDELINES

Question 2 (continued)

Notes:

- Each of the following situations is one in which a response that otherwise would be scored as E should be scored as P, and a response that otherwise would be scored as P should be scored as I:
  - The response includes a statement that the small probability *proves* that the manager did not make the selection at random (or any equivalent wording).
  - The response includes a statement that clearly interprets the probability from part (a) to be the probability that the manager selected the three people at random.
- Each of the following situations is one in which the response is scored as I:
  - The decision is inconsistent with the justification (*e.g.*, "The probability is very small, so there is no reason to doubt the manager's claim").
  - The response states or implies that because the selection of three women was *not impossible,* there is *no reason to doubt* the manager's claim.

Part (c) is scored as follows:

Essentially correct (E) if the response answers no *AND* states that the dice outcomes in the proposed simulation are independent *AND* states that the genders of the selected convention attendees are dependent. The table below shows statements that should be considered equivalent to the required statements of independence and dependence.

Independence of dice outcomes	Dependence of genders
• The three dice outcomes are independent. • The probability of rolling a 5 or a 6 is the same on all three dice. • The dice simulation actually simulates sampling with replacement.	• The genders of the three people are dependent (or not independent). • The probability of selecting a woman changes after each selection. • The people are sampled without replacement.

OR

Essentially correct if the response answers no *AND* computes the correct probability that a trial of the simulation will indicate the selection of three women $\left(\frac{1}{3} \times \frac{1}{3} \times \frac{1}{3} \approx 0.037\right)$ *AND* states that the probability is different from the probability found in part (a).

Partially correct (P) if the response correctly answers no and either:

States only that the dice outcomes are independent or states only that the genders of the selected convention attendees are dependent, but not both;

OR

Otherwise meets the criteria for E but has poor communication. An example of poor communication is: "No, because it selects with replacement. It isn't possible for the same person to be selected twice." (There is an apparent shift between the two sentences from describing the simulation to describing the actual selection of people, but that is not made clear.)

AP® STATISTICS
2014 SCORING GUIDELINES

Question 2 (continued)

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

Note: Pointing out that a sample of three people is more than 10% of the population of nine people should be considered equivalent to stating that the selection of a woman is not independent among the three people selected to attend the convention.

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

AP® STATISTICS
2014 SCORING GUIDELINES

Question 3

Intent of Question

The primary goals of this question were to assess a student's ability to (1) perform a probability calculation from a normal distribution; (2) explain an implication of examining the distribution of a sample mean rather than the distribution of a single measurement; and (3) perform a probability calculation involving independent events using the multiplication rule.

Solution

Part (a):

Because the distribution of the daily number of absences is approximately normal with mean 120 students and standard deviation 10.5 students, the z-score for an absence total of 140 students is $z = \frac{140-120}{10.5} \approx 1.90$. The table of standard normal probabilities or a calculator reveals that the probability that 140 or fewer students are absent is 0.9713. So the probability that more than 140 students are absent (and that the school will lose some state funding) is $1 - 0.9713 = 0.0287$.

Part (b):

High School A would be *less* likely to lose state funding. With a random sample of 3 days, the distribution of the sample mean number of students absent would have less variability than that of a single day. With less variability, the distribution of the sample mean would concentrate more narrowly around the mean of 120 students, resulting in a smaller probability that the mean number of students absent would exceed 140.

In particular, the standard deviation of the sample mean number of absences, $\bar{x}$, is $\frac{\sigma}{\sqrt{n}} = \frac{10.5}{\sqrt{3}} = 6.062$. So the z-score for a sample mean of 140 is $\frac{140-120}{6.062} \approx 3.30$. The probability that High School A loses funding using the suggested plan would be $1 - 0.9995 = 0.0005$, as determined from the table of standard normal probabilities or from a calculator, which is less than a probability of 0.0287 obtained for the plan described in part (a).

Part (c):

For any one typical school week, the probability is $\frac{2}{5} = 0.4$ that the day selected is not Tuesday, not Wednesday, or not Thursday. Therefore, because the days are selected independently across the three weeks, the probability that none of the three days selected would be a Tuesday or Wednesday or Thursday is $(0.4)^3 = 0.064$.

AP® STATISTICS
2014 SCORING GUIDELINES

Question 3 (continued)

Scoring

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

Part (a) is scored as follows:

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

1. Indicates use of a normal distribution and clearly identifies the correct parameter values (showing correct components of a z-score calculation is sufficient).
2. Uses the correct boundary value (140, 140.5, or 141 is acceptable).
3. Reports the correct normal probability consistent with components 1 and 2

OR

if the response reports a probability of 0.025 with justification based on the empirical rule for an acceptable boundary value (140, 140.5, or 141 is acceptable).

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

OR

if the response provides an incorrect probability of 0.05 with justification based on the empirical rule for an acceptable boundary value (140, 140.5, or 141 is acceptable).

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

Note: An inconsistency in calculations lowers the score for part (a) by one level (that is, from E to P or from P to I).

Part (b) is scored as follows:

Essentially correct (E) if the response provides the correct answer of less likely *AND* the following three components:

1. Clearly references the distribution of the sample <u>mean</u>.
2. Indicates that the variability of the distribution is smaller.
3. Indicates that the distribution is centered at 120.

OR

if the response provides the correct answer of less likely *AND* the following two components:

1. Correctly calculates the probability that the sample mean would exceed 140 (arithmetic errors are not penalized).
2. Correctly compares this probability to the probability in part (a).

Partially correct (P) if the response provides the correct answer of less likely *AND* only two of the following three components:

1. Clearly references the distribution of the sample <u>mean</u>.
2. Indicates that the variability of the distribution is smaller.
3. Indicates that the distribution is centered at 120.

OR

if the response provides the correct answer of less likely *AND* correctly calculates the probability that the sample mean would exceed 140 (arithmetic errors are not penalized) *BUT* does <u>not</u> correctly compare this probability with the probability in part (a).

AP® STATISTICS
2014 SCORING GUIDELINES

Question 3 (continued)

Incorrect (I) if the response does not meet the criteria for E or P, including if the response provides the incorrect answer or provides the correct answer of less likely with no explanation or an incorrect explanation.

Note: An equivalent approach is to use the total number of absences for 3 days. The sampling distribution of the total number of absences for the 3 days is approximately normal, with mean $3(120) = 360$ absences and standard deviation $3(6.026) \approx 18.187$ absences. The z-score for a total of $3(140) = 420$ absences is: $\frac{420 - 360}{18.187} \approx 3.30$. Such a response is scored E if the response provides the correct answer of less likely and references the distribution of the sample total, and includes the correct mean and standard deviation.

Part (c) is scored as follows:

Essentially correct (E) if the response correctly calculates the probability *AND* shows sufficient work.

Partially correct (P) if the response reports the correct probability but shows no work or does not show sufficient work;
OR
if the response uses the multiplication rule involving three events but does so incorrectly and/or with an incorrect probability of not selecting a Tuesday, Wednesday, or Thursday.

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

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