AP Stats
Review for the AP Exam – Chapter 9

Name: Key
Hour: ______

1. What is the difference between a parameter and a statistic?

Parameter = population   statistic - sample

2. What is the notation for the following:

population mean	sample mean	population proportion	sample proportion
μ	$\bar{x}$	p	$\hat{p}$

3. What does it mean if a statistic is unbiased?

The center of the sampling distribution is equal to the center of the population.

4. What is the Central Limit Theorem?

If an SRS of any population is sufficiently large, the sampling distribution of the sampling mean is $\sim N$.

5. What are the two Rules of Thumb that apply to the sampling distributions of sample proportions?

① If $N \geq 10n$, you may use $\sqrt{\frac{p(1-p)}{n}}$

② If $np \geq 10$ and $n(1-p) \geq 10$, you may use normal approximation.

6. Julie generates a sample of 20 random integers between 0 and 9 inclusive. She records the number of 6's in the sample. She repeats this process 99 more times, recording the number of 6's in each sample. What kind of distribution has she simulated?

(A) The sampling distribution of the sample proportion with $n = 20$ and $p = 0.6$.
(B) The sampling distribution of the sample proportion with $n = 100$ and $p = 0.1$.
(C) The binomial distribution with $n = 20$ and $p = 0.1$.
(D) The binomial distribution with $n = 100$ and $p = 0.1$.
(E) The binomial distribution with $n = 20$ and $p = 0.6$.

7. A volunteer for a mayoral candidate's campaign periodically conducts polls to estimate the proportion of people in the city who are planning to vote for this candidate in the upcoming election. Two weeks before the election, the volunteer plans to double the sample size in the polls. The main purpose of this is to...

(A) reduce nonresponse bias
(B) reduce the effects of confounding variables
(C) reduce bias due to the interviewer effect
(D) decrease the variability in the population
(E) decrease the standard deviation of the sampling distribution of the sample proportion

8. A random sample of two observations is taken from a population that is normally distributed with a mean of 100 and a standard deviation of 5. Which of the following is closest to the probability that the sum of the two observations is greater than 221?

(A) 0.0015   (B) 0.0250   (C) 0.0500   (D) 0.4500   (E) 0.9985

$P(\text{sum} > 221) = P\left(z > \frac{221 - (100+100)}{\sqrt{5^2+5^2}}\right) = P(z > 2.970) = .0015$

9. The graphs of the sampling distributions, I and II, of the sample mean of the same random variable for samples of two different sizes are shown above. Which of the following statements must be true about the sample sizes?

(A) The sample size of I is less than the sample size of II.
(B) The sample size of I is greater than the sample size of II.
(C) The sample size of I is equal to the sample size of II.
(D) The sample size does not affect the sampling distribution.
(E) The sample sizes cannot be compared based on these graphs.

10. Suppose that public opinion in a large city is 35 percent against increasing taxes to support the public school system. If a random sample of 500 people from this city are interviewed, what is the approximate probability that more than 200 of these people will be against increasing taxes? Which of the following set-ups would answer the question?

(A) $\binom{500}{200}(0.65)^{200}(0.35)^{300}$

(B) $\binom{500}{200}(0.35)^{200}(0.65)^{300}$

(C) $P\left(z > \dfrac{0.40-0.65}{\sqrt{\dfrac{(0.65)(0.35)}{500}}}\right)$

(D) $P\left(z > \dfrac{0.40-0.35}{\sqrt{\dfrac{(0.40)(0.60)}{500}}}\right)$

(E) $P\left(z > \dfrac{0.40-0.35}{\sqrt{\dfrac{(0.35)(0.65)}{500}}}\right)$

(2005 #2)

11. Let the random variable X represent the number of telephone lines in use by the technical support center of a software manufacturer at noon each day. The probability distribution of X is shown in the table below.

x	0	1	2	3	4	5
$p(x)$	0.35	0.20	0.15	0.15	0.10	0.05

(a) Calculate the expected value (the mean) of X.

(b) Using past records, the staff at the technical support center randomly selected 20 days and found that an average of 1.25 telephone lines were in use at noon on those days. The staff proposes to select another random sample of 1,000 days and compute the average number of telephone lines that were in use at noon on those days. How do you expect the average from this new sample to compare to that of the first sample? Justify your response.

(c) The median of a random variable is defined as any value x such that $P(X \le x) \ge 0.5$ and $P(X \ge x) \ge 0.5$. For the probability distribution shown in the table above, determine the median of X.

(d) In a sentence or two, comment on the relationship between the mean and the median relative to the shape of this distribution.

(a) $E(x) = \mu_x = 0(.35) + 1(.2) + 2(.15) + 3(.15) + 4(.10) + 5(.05)$

$= 1.6$

(b) According to the Central Limit Theorem $\bar{X}$ approaches μ as the sample size increases. So I would expect the average from the new sample of 1000 to be closer to the expected value of 1.6 than the sample size of 20.

(c) The median is 1 since $P(x \le 1) = .55 \ge .5$ and $P(x \ge 1) = .65 > .5$

(d) The distribution is skewed right so the mean is greater than the median.

(1998 #1)

12. Consider the sampling distribution of a sample mean obtained by random sampling from an infinite population. This population has a distribution that is highly skewed toward the larger values.

(a) How is the mean of the sampling distribution related to the mean of the population?

The mean of the sampling distribution and the mean of the population are the same.

(b) How is the standard deviation of the sampling distribution related to the standard deviation of the population?

$$S_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

(c) How is the shape of the sampling distribution affected by sample size?

Small sample sizes will be skewed right. As the sample gets larger, the shape will become more normal.

(2009 #2)

13. A tire manufacturer designed a new tread pattern for its all-weather tires. Repeated tests were conducted on cars of approximately the same weight traveling at 60 miles per hour. The tests showed that the new tread pattern enables the cars to stop completely in an average distance of 125 feet with a standard deviation of 6.5 feet and that the stopping distances are approximately normally distributed.

(a) What is the 70th percentile of the distribution of stopping distances?

$P(z \le z^*) = .7$

invnorm(.7) $z^* = .5244$

$.5244 = \frac{x - 125}{6.5}$

$\boxed{x = 128.409\text{ ft}}$

(b) What is the probability that at least 2 cars out of 5 randomly selected cars in the study will stop in a distance that is greater than the distance calculated in part (a)?

$P(x=2 \cup x=3 \cup x=4 \cup x=5) =$

$\left[\binom{5}{2}.3^2(.7)^3 + \binom{5}{3}.3^3.7^2 + \binom{5}{4}.3^4.7^1 + \binom{5}{5}(.3)^5(.7)^0\right]$

$= \boxed{.472}$

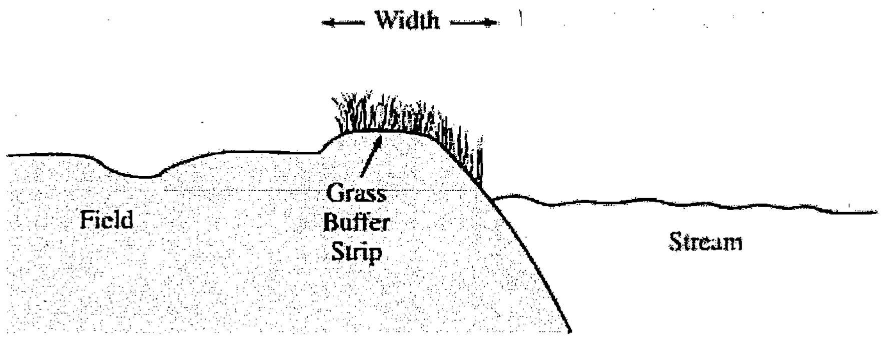
(c) What is the probability that a randomly selected sample of 5 cars in the study will have a mean stopping distance of at least 130 feet?

$P(\bar{x} \ge 130) = P(z \ge 1.72) = \boxed{.0472}$

$$z = \frac{130 - 125}{\frac{6.5}{\sqrt{5}}} = 1.72$$

(2011B #6)

14. Grass buffer strips are grassy areas that are planted between bodies of water and agricultural fields. These strips are designed to filter out sediment, organic material, nutrients, and chemicals carried in runoff water. The figure below shows a cross-sectional view of a grass buffer strip that has been planted along the side of a stream.



A study in Nebraska investigated the use of buffer strips of several widths between 5 feet and 15 feet. The study results indicated a linear relationship between the width of the grass strip (x), in feet, and the amount of nitrogen removed from the runoff water (y), in parts per hundred. The following model was estimated.

$$\hat{y} = 33.8 + 3.6x$$

(a) Interpret the slope of the regression line in the context of this question.

On average, there is an increase of 3.6 parts per hundred of nitrogen removed from the runoff water for each additional foot of width in the grass strip.

(b) Would you be willing to use this model to predict the amount of nitrogen removed for grass buffer strips with widths between 0 feet and 30 feet? Explain why or why not.

No. Do not extrapolate. The model was created using widths between 5 and 15 ft only.

A scientist in California wants to know if there is a similar relationship in her area. To investigate this, she will place a grass buffer strip between a field and a nearby stream at each of eight different locations and measure the amount of nitrogen that the grass buffer strip removes, in parts per hundred, from runoff water at each location. Each of the eight locations can accommodate a buffer strip between 6 feet and 13 feet in width. The scientist wants to investigate which combination of widths will provide the best estimate of the slope of the regression line.

Suppose the scientist decides to use buffer strips of width 6 feet at each of four locations and buffer strips of width 13 feet at each of the other four locations. Assume the model, $\hat{y} = 33.8 + 3.6x$, estimated from the Nebraska study is the true regression line in California and the observations at the different locations are normally distributed with standard deviation of 5 parts per hundred.

(c) Describe the sampling distribution of the sample mean of the observations on the amount of nitrogen removed by the four buffer strips with widths of 6 feet.

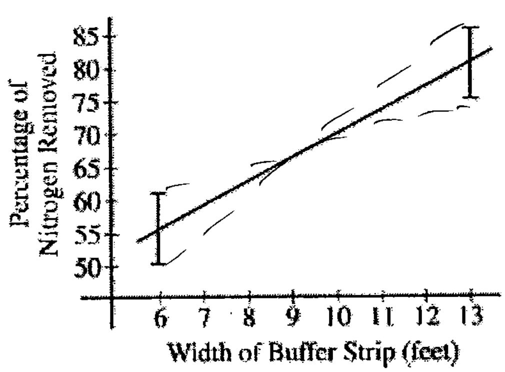
$\hat{y} = 33.8 + 3.6(6) = 55.4 \qquad \sigma_{\bar{x}} = \frac{5}{\sqrt{4}} = 2.5$

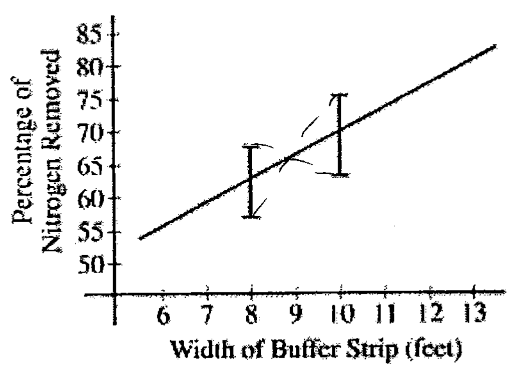
Since the observations are stated as normally distributed, the sampling distribution will also be ~N with a mean of 55.4 parts per hundred and a standard deviation of 2.5 parts per hundred.

(d) Using your result from part (c), show how to construct an interval that has probability 0.95 of containing the sample mean of the observations from four buffer strips with widths of 6 feet.

$55.4 \pm 1.96(2.5) = (50.5, 60.3)$

For the study plan being implemented by the scientist in California, the graph on the left below displays intervals that each have probability 0.95 of containing the sample mean of the four observations for buffer strips of width 6 feet and for buffer strips of width 13 feet. A second possible study plan would use buffer strips of width 8 feet at four of the eight locations and buffer strips of width 10 feet at the other four locations. Intervals that each have probability 0.95 of containing the mean of the four observations for buffer strips of width 8 feet and for buffer strips of width 10 feet, respectively, are shown in the graph on the right below.





If data are collected for the first study plan, a sample mean will be computed for the four observations from buffer strips of width 6 feet and a second sample mean will be computed for the four observations from buffer strips of width 13 feet. The estimated regression line for those eight observations will pass through the two sample means. If data are collected for the second study plan, a similar method will be used.

(e) Use the plots on the previous page to determine which study plan, the first or the second, would provide a better estimator of the slope of the regression line. Explain your reasoning.

The dotted lines connecting low to high and vice versa represent the most extreme cases for the regression line. Since the extreme cases of the first graph are more similar than the second graph, I would recommend the first study plan.

(f) The previous parts of this question used the assumption of a straight-line relationship between the width of the buffer strip and the amount of nitrogen that is removed, in parts per hundred. Although this assumption was motivated by prior experience, it may not be correct. Describe another way of choosing the widths of the buffer strips at eight locations that would enable the researchers to check the assumption of a straight-line relationship.

Rather than clustering the locations with 4 at 6ft and 4 at 13ft, we could space the locations out from 6 to 13. One at each of 6, 7, 8, 9, 10, 11, 12, 13 ft. Then samples could be taken of the amont of nitrogen removed to try to identify a linear relationship.