Sampling Distributions

Quick Reminders
1. Know the difference between a parameter and a statistic. A parameter is a population
value (i.e. a population mean or proportion) and a statistic estimates a population
parameter based upon a sample.
2. A sampling distribution of a statistic is the distribution of the values of that statistic
obtained from all possible samples of a given size from a given population.
3. Know what is meant by an unbiased statistic. A statistic is unbiased if it’s sampling
distribution is centered at the population parameter.
4. When sampling without replacement, the formulas for $\sigma_\hat{p}$ and $\sigma_x$ should be used only
when the population is at least 10 times as large as the sample.
5. The CLT is a statement about shape! It says that the sampling distribution of sample
means becomes more normally distributed as the sample size increases.
6. Sampling distributions are an extension of probability, so many free response questions
that include questions on sampling distributions will also include parts that relate to
material discussed and reviewed in the earlier prep session on probability.

Sampling Distribution of the sample mean...
Let $\bar{x}$ be the mean of a simple random sample of size $n$ drawn from a population whose
mean is $\mu_x$ and whose standard deviation is $\sigma_x$. The sampling distribution of $\bar{x}$ will have
- a mean equal to the population mean, $\mu_\bar{x} = \mu_x$
- and a standard deviation equal to the population standard deviation divided by the square
  root of the sample size, $\sigma_\bar{x} = \frac{\sigma_x}{\sqrt{n}}$

The shape of this distribution is approximately normal provided that
- The population from which the sample was drawn is approximately normal
  OR
- The sample is sufficiently large, $n \geq 30$ is what most texts agree on.

This is stated on the formula sheet as...
If $\bar{x}$ is the mean of a random sample of size $n$ from an infinite population with mean $\mu$ and
standard deviation $\sigma$, then:

$\mu_\bar{x} = \mu$ and $\sigma_\bar{x} = \frac{\sigma}{\sqrt{n}}$
Sampling Distribution of the sample proportion

Let \( \hat{p} \) be the proportion of “successes” in a simple random sample of size \( n \) drawn from a population in which the proportion of “successes” is \( p \). If \( x \) is the number of successes in the sample then \( \hat{p} = \frac{x}{n} \). The sampling distribution of \( \hat{p} \) will have...

- a mean equal to the population proportion \( \mu_p = p \)
- and a standard deviation \( \sigma_p = \sqrt{\frac{p(1-p)}{n}} \)

The shape of this distribution is approximately normal provided that...

- \( np \geq 10 \) and \( n(1-p) \geq 10 \).
- AND
- The sample size \( n \) is not more than 10% of the population. This allows us to consider the outcome of each of the \( n \) trials to be independent.

This is stated on the formula sheet under the notes on the binomial distribution. This is because the distribution of the sample proportion is just a linear transformation of the count of successes. \( \hat{p} = \frac{1}{n} \cdot x \)
Multiple Choice Questions
Questions 1 and 2 refer to the following information:

Every Tuesday, Jim and John’s Video has “lucky-chance” day. A customer may choose to flip a coin and roll a fair die and rent a second movie for an amount (in cents) equal to ten or twenty times the number uppermost on the die, ten if the coin is heads and twenty if the coin is tails. For example, if the customer rolls a two and flips heads, a second movie may be rented for $0.20. If a two and a tail turn up, a second movie may be rented for $0.40. Let $X$ represent the amount paid for a second movie on lucky-chance day. The expected value of $X$ is $0.525 and the standard deviation of $X$ is $0.322.

1. If a customer rolls the dice and rents a second movie every Tuesday for 52 consecutive weeks, what is the total amount that the customer would expect to pay for these second movies?

A. $0.32
B. $0.53
C. $0.85
D. $16.74
E. $27.30

2. If a customer rents a second movie every Tuesday for 35 consecutive weeks, what is the approximate probability that the total amount paid for these second movies will exceed $20.00?

A. 0
B. 0.20
C. 0.44
D. 0.56
E. 0.80
3. The population \{5, 8, 9, 12, 14\} has mean \(\mu = 9.6\) and standard deviation \(\sigma = 3.14\). When sampling with replacement, there are 32 different possible ordered samples of size 2 that can be selected from this population. The mean of each of these 32 samples is computed. For example, 1 of the 32 samples is (5, 9), which has a mean of 7. The distribution of the 32 sample means has its own mean \(\mu_x\) and its own standard deviation \(\sigma_x\). Which of the following statements is true?

A. \(\mu_x = 9.6\) and \(\sigma_x = 3.14\)
B. \(\mu_x = 9.6\) and \(\sigma_x > 3.14\)
C. \(\mu_x = 9.6\) and \(\sigma_x < 3.14\)
D. \(\mu_x > 9.6\)
E. \(\mu_x < 9.6\)

4. Suppose the distribution of weights of adult dogs of a particular breed is strongly skewed right with a mean of 15 pounds and a standard deviation of 4 pounds. Describe the sampling distribution of sample means for a random sample of 40 dogs from the population.

A. The sampling distribution will be strongly skewed right with a mean of 15 pounds and a standard deviation of 4 pounds.
B. The sampling distribution will be strongly skewed right with a mean of 15 pounds and a standard deviation of 0.632 pounds.
C. The sampling distribution will be approximately normally distributed with a mean of 15 pounds and standard deviation of 4 pounds.
D. The sampling distribution will be approximately normally distributed with a mean of 15 pounds and standard deviation of 0.1 pounds.
E. The sampling distribution will be approximately normally distributed with a mean of 15 pounds and standard deviation of 0.632 pounds.
5. Forty-five percent of households in a large metropolitan area have at least one pet. What is the probability of obtaining a random sample of 100 households in the area, in which 35 or less have at least one pet?

A. 0.011  
B. 0.018  
C. 0.022  
D. 0.978  
E. 0.982

6. Suppose that the cans of soda that are filled by a particular bottling company are filled to an average volume of 12 ounces with a standard deviation 0.2 ounces. What is the probability that a random sample of 4 cans has an average volume of at least 12.2 ounces?

A. 0  
B. 0.023  
C. 0.046  
D. 0.159  
E. 0.318
7. The diameter of a particular variety of oranges is normally distributed with a mean of 5 cm and a standard deviation of 0.5 cm. Suppose an orchard sells bags of 16 oranges, assuming that the bags are filled at random, 25 percent of have a mean diameter greater than which of the following?

A. 4.663 ounces  
B. 4.916 ounces  
C. 5.021 ounces  
D. 5.084 ounces  
E. 5.337 ounces

8. Suppose that public opinion in a large city is 72 percent in favor of allowing a large recycling plant to be constructed using taxpayer funds and 28 percent against such a build. If a random sample of 300 people from this city is interviewed, what is the approximate probability that more than 100 of these people will be against increasing taxes?

A. \( \binom{300}{100} (0.72)^{100} (0.28)^{200} \)  
B. \( \binom{300}{100} (0.28)^{100} (0.72)^{200} \)  
C. \( P \left( z > \frac{0.33 - 0.28}{\sqrt{\frac{(0.33)(0.67)}{300}}} \right) \)  
D. \( P \left( z > \frac{0.33 - 0.28}{\sqrt{\frac{(0.33)(0.67)}{300}}} \right) \)  
E. \( P \left( z > \frac{0.33 - 0.28}{\sqrt{\frac{(0.28)(0.72)}{300}}} \right) \)
Sampling Distributions

We have intentionally included more material than can be covered in most Student Study Sessions to account for groups that are able to answer the questions at a faster rate. Use your own judgment, based on the group of students, to determine the order and selection of questions to work in the session. Be sure to include a variety of types of questions (multiple choice and free response) in the time allotted.

Multiple Choice
1. E (Similar to 1997 Q19)
   \[ E(52X) = 52E(X) = 52(0.525) = 27.30 \]
2. B (Similar to 1997 Q20)
   Let \( T \) be the total amount spent on second movies for the 35 weeks.
   \[ \sigma^2(T) = \sigma^2(X_1) + \sigma^2(X_2) + \ldots + \sigma^2(X_{35}) = \sqrt{35(0.322)^2} = 1.905 \]
   \[ E(T) = 35(0.525) = 18.375 \]
   Since there are 35 independent trials it is reasonable to assume that \( T \sim N(18.375, 1.905) \) so
   \[ P(T > 20) = P\left( z > \frac{20 - 18.375}{1.905} \right) \approx 0.20 \]
3. C (Similar to 2002 Q30)
   Since the mean of the sampling distribution of sample means is equal to the population mean
   \[ \mu_x = 9.6 \]
   and the standard deviation of the sampling distribution is \( \sigma_x = \frac{\sigma}{\sqrt{n}} \) so it must be less than 3.14.
4. E (AP Style Question)
   Since the sample size is greater than thirty it is reasonable (by the CLT) to assume the sampling distribution will be normally distributed. The mean is equal to the population mean and the standard deviation will be \( \sigma_x = \frac{\sigma}{\sqrt{n}} = \frac{4}{\sqrt{40}} = 0.632 \)
5. C (AP Style Question)
   \[ P(\hat{p} < .35) = P\left( z < \frac{.35 - .45}{\sqrt{(0.45)(0.55)\over 100}} \right) = P(z < -2.01) = 0.022 \]
6. B (AP Style Question)
   \[ P(\bar{x} > 12.2) = P\left( z > \frac{12.2 - 12}{0.2} \right) = P(z > 2) = 0.023 \]
7. D (Similar to 2007 Q3)

\[ P(z > z^*) = 0.25 \ \text{therefore} \ z^* = 0.6745; \ \text{Solve} \ \frac{x - 5}{0.5} = \frac{\bar{x} - 5}{0.5} = 5.084 \ \text{ounces} \]

8. E (Similar to 2002 Q8)

\[ P\left( \hat{p} > \frac{100}{300} \right) = P\left( z > \frac{0.33 - 0.28}{\sqrt{0.28 \cdot 0.72 / 300}} \right) \]