STUDY SET 3

Sampling Distributions and Estimation

THE NEXT THREE QUESTIONS ARE BASED ON THE FOLLOWING INFORMATION:
Suppose you flip a coin four times. For every head, you receive one point, and for every tail, you lose one point.

1. Develop the sampling distribution for the mean number of points you receive from flipping four coins.

   Answer:

$\bar{x}$	-1	-0.5	0	0.5	1
Fav Outcomes	1	4	6	4	1
$P(\bar{x})$	1/16 = 0.0625	4/16 = 0.25	6/16 = 0.375	4/16 = 0.25	1/16 = 0.0625

2. What is the probability that the mean number of points you receive on four flips is 0.5?
   Answer:
   0.25
3. What is the probability that the mean number of points you receive on four flips is 0?

   Answer:
   0.375

THE NEXT TWO QUESTIONS ARE BASED ON THE FOLLOWING INFORMATION:
You have recently joined a country club. Suppose that the number of times you expect to play golf in a month is represented by a normally distributed random variable with a mean of 10 and a standard deviation of 2.4.

4. Over the course of the next year, what is the probability that you average more than 11 games a month?

   Answer:
   0.0749

5. Over the course of the next year, the probability is 85% that you average less than how many games per month?

   Answer:
   10.721

6. The number of orders that come into a mail-order sales office each month is normally distributed with a mean of 298 and a standard deviation of 15.4. For a particular sample size, the probability that the sample mean exceeds 300 is 0.2. How big must the sample be?

   Answer: 42

THE NEXT TWO QUESTIONS ARE BASED ON THE FOLLOWING INFORMATION:
The length of time it takes to fill an order at a local sandwich shop is normally distributed with a mean of 4.1 minutes and a standard deviation of 1.3 minutes.

7. What is the probability that the average waiting time for a random sample of ten customers is between 4.0 and 4.2 minutes?

   Answer:
   0.1896

8. The probability is 95% that the average waiting time for a random sample of ten customers is greater than how many minutes?

   Answer:
   3.42

THE NEXT TWO QUESTIONS ARE BASED ON THE FOLLOWING INFORMATION:
In a recent survey of high school students, it was found that the average amount of money spent on entertainment each week was normally distributed with a mean of $52.30 and a standard deviation of $18.23. Assume that these values are representative of all high school students.

9. What is the probability that for a sample of 25, the average amount spent exceeds $60?

   Answer:
   0.0174

10. The probability is 65% that the average spending of a sample of 25 randomly-selected students will spend at least how much?

    Answer:
    50.88

THE NEXT THREE QUESTIONS ARE BASED ON THE FOLLOWING INFORMATION:
Let the random variable X follow a normal distribution with a mean of μ and a standard deviation of σ. Let $\bar{X}_1$ be the mean of a sample of 16 observations randomly chosen from this population, and $\bar{X}_2$ be the mean of a sample of 25 observations randomly chosen from the same population.

11. Evaluate the statement $P(\mu - 0.2\sigma < \bar{X}_1 < \mu + 0.2\sigma) < P(\mu - 0.2\sigma < \bar{X}_2 < \mu + 0.2\sigma)$ as to whether it is true or false. Please state if you are unable to determine whether the statement is true or false. Support your answer.

    Answer:
    Since $\bar{X}_1$ is less concentrated around the mean (larger standard error), the statement $P(\mu - 0.2\sigma < \bar{X}_1 < \mu + 0.2\sigma) < P(\mu - 0.2\sigma < \bar{X}_2 < \mu + 0.2\sigma)$ is true.

12. Evaluate the statement $P(\bar{X}_1 < \mu + \frac{\sigma}{4}) < P(\bar{X}_2 < \mu + \frac{\sigma}{5})$ as to whether it is true or false. Please state if you are unable to determine whether the statement is true or false. Support your answer.

    Answer:
    Both of these expressions are asking the same question: What is the probability of observing a sample mean below one standard deviation above its expected value? Therefore, these probabilities would be equal. Hence, $P(\bar{X}_1 < \mu + \frac{\sigma}{4}) < P(\bar{X}_2 < \mu + \frac{\sigma}{5})$ is a false statement.

13. Evaluate the statement $P(\bar{X}_1 < \mu + \frac{\sigma}{16}) < P(\bar{X}_2 < \mu + \frac{\sigma}{25})$ as to whether it is true or false. Please state if you are unable to determine whether the statement is true or false. Support your answer.

    Answer:
    In $P(\bar{X}_1 < \mu + \frac{\sigma}{16})$, we are going 1/4 of a standard error above the mean. In $P(\bar{X}_2 < \mu + \frac{\sigma}{25})$, we are only going 1/5 of a standard error above the mean. Therefore, the expression on the left hand side has to be larger than the probability on the right hand side. Hence, $P(\bar{X}_1 < \mu + \frac{\sigma}{16}) < P(\bar{X}_2 < \mu + \frac{\sigma}{25})$ is a false statement.

Suppose that 15% of all invoices are for amounts greater than $1,000. A random sample of 60 invoices is taken. What is the mean and standard error of the sample proportion of invoices with amounts in excess of $1,000? What is the probability that the proportion of invoices in the sample is greater than 18%?

Answer:
0.2578

THE NEXT FOUR QUESTIONS ARE BASED ON THE FOLLOWING INFORMATION:
It has been found that 62.1% of all unsolicited third class mail delivered to households goes unread. Over the course of a month, a household receives 100 pieces of unsolicited mail.

14. What is the mean of the sample proportion of pieces of unread mail?

 Answer:
 0.621

15. What is the variance of the proportion?

 Answer:
 0.00235

16. What is the standard error of the sample proportion?

 Answer:
 0.0485

17. What is the probability that the sample proportion is greater than 0.60?

 Answer:
 0.6664

THE NEXT TWO QUESTIONS ARE BASED ON THE FOLLOWING INFORMATION:
The filling machine at a bottling plant is operating correctly when the variance of the fill amount is equal to 0.3 ounces. Assume that the fill amounts follow a normal distribution.

18. What is the probability that for a sample of 30 bottles, the sample variance is greater than 0.5?

 Answer:
 Using Excel, *p*-value = CHIDIST(48.33,29) = 0.0136.

19. The probability is 0.10 that for a sample of 30 bottles, the sample variance is less than what number?

 Answer: 0.2045

THE NEXT TWO QUESTIONS ARE BASED ON THE FOLLOWING INFORMATION:
A sample of 25 bottles is taken from the production line at a local bottling plant. Assume that the fill amounts follow a normal distribution.

20. What is the probability that the sample standard deviation is more than 70% of the population standard deviation?

    Answer:
    0.982
    (Obtained using Excel)
    Using table, *p*-value lies between 0.975 and 0.990.

The probability is 90% that the sample variance is less than what percent of the population variance?

    Answer:
    k= 1.383 or 13.83%

21. The length of time it takes a stock analyst to complete an evaluation of a company's earnings forecast is normally distributed with a mean of 6.7 hours. From looking at the time spent by an analyst evaluating 36 randomly selected stocks you find that the sample mean was 6.4 hours. If you are told that the probability of getting a sample mean this small or smaller is 0.10, what must the population standard deviation have been?
    Answer:
    1.406

THE NEXT TWO QUESTIONS ARE BASED ON THE FOLLOWING INFORMATION:
It is estimated that 1.3% of all items produced on an assembly line have some form of defect. You randomly select 60 items for inspection.

22. What is the probability that more than 1.5% of the 60 items have some form of defect?
    Answer:
    0.4443
23. The probability is 86% that the proportion of items with a defect is less than what number?
    Answer:
    $\Rightarrow K = 0.0288$
24. An advertisement claims that four out of five doctors recommend a particular product. A consumer group wants to test that claim, and takes a random sample of 30 doctors. The consumer group finds that of this group of doctors, only 0.75 would recommend the product. What is the probability that the proportion of doctors in this sample who recommend the product is 0.75 or less? Do you have reason to doubt the manufacturer's claim? Explain.
    Answer:

    $P(Z < -0.68) = 0.2483$

It would not be unusual to have 75% or fewer doctors recommending the product out of a group of 30. We cannot really doubt the manufacturer's claim.

THE NEXT TWO QUESTIONS ARE BASED ON THE FOLLOWING INFORMATION:
The average checking account balance at a local bank is $1,742 with a standard deviation of $132. A random sample of 70 accounts is selected.

25. What is the probability that the average balance for these accounts is less than $1,700?
Answer:
0.0039
26. The probability is 0.23 that the average balance for these accounts is greater than what amount?

Answer:
$1,753.67.

THE NEXT TWO QUESTIONS ARE BASED ON THE FOLLOWING INFORMATION:
The number of nails in a five-pound box of nails is normally distributed with a mean of 563.3 and a standard deviation of 33. Suppose you select a random sample of 30 boxes.

27. What is the probability that the average number of nails per box for this sample is between 560 and 580 nails?

Answer:
0.706

28. The probability is 0.22 that the sample mean will be more than what value?
Answer:
567.94

29. The length of time it takes for a stock analyst to complete an evaluation of a company's earnings forecast is normally distributed with a standard deviation of 1.7 hours. From looking at the time spent by an analyst evaluating 15 randomly-selected stocks you find that the sample mean was 6.4 hours. If you know that the probability of getting a sample mean this large or larger is 0.36, what is the mean of the population?
Answer:
6.242

30. In examining the invoices issued by a company, an auditor finds that the dollar amount of invoices have a mean of $1,732 and standard deviation of $298. Which pair of symmetric numbers around the mean make the statement $P(a < \bar{X} < b)=0.853$ correct for a random sample of 55 invoices?
Answer:
a =1673.74 and b =1790.26.

31. It has been estimated that 43% of all college students change their major at least once during the course of their college career. If we take a random sample of 55 college students, what is the probability that more than 40% will change their major?
Answer:
0.6736

93. It has been estimated that 53% of all college students change their major at least once during the course of their college career. Suppose you are told that the sample proportion for a random sample was 0.48. Furthermore, you are told that the probability of getting a sample proportion of this size or smaller is 14%. What must have been the sample size?

Answer:
117

94. The number of students using the ATM on campus daily is normally distributed with a mean of 237.6 and a standard deviation of 26.3. You take a random sample of 30 days. Find two values "*a*" and "*b*" that are symmetric around the mean such that the probability is 0.95 that the sample mean is greater than "*a*" and less than "*b*".

Answer:
a =228.189 and b =247.01.

95. The results of a recent survey indicated that 17.7% of all U.S. adults had taken a commercial airplane flight during the last year. If we take a random sample of 200 adults, what is the probability that more than 20% flew during the past year?

Answer:
0.1977

THE NEXT FOUR QUESTIONS ARE BASED ON THE FOLLOWING INFORMATION:

Given a population with mean μ=120 and variance σ^2= 81, the central limit applies when the sample size is large-enough. A random sample of size n = 36 is obtained.

96. What are the mean and standard deviation of the sampling distribution for the sample means?
Answer:
1.50

97. What is the probability that the sample mean is greater than 123?
Answer:
0.0228

98. What is the probability that the sample mean is between 118 and 121?
Answer:
0.6568

99. What is the probability that the sample mean is at most 121.5?
Answer:
0.8413

THE NEXT FIVE QUESTIONS ARE BASED ON THE FOLLOWING INFORMATION:
The mean selling price of new homes in a city over a year was $120,000.The population standard deviation was $28,000. A random sample of 100 new home sales from this city was taken.

100. What is the probability that the sample mean selling price was more than $116,000?
Answer:
0.9236

101. What is the probability that the sample mean selling price was between $118,000 and $122,000?
Answer:
0.5222

102. What is the probability that the sample mean selling price was between $119,000 and $121,000?
Answer:
0.2812

103. Without doing the calculations, state in which of the following ranges the sample mean selling price is most likely to lie:
$118,000 to $120,000 $119,000 to $121,000
$120,000 to $122,000 $121,000 to $123,000
Answer:
$119,000 to $121,000

104. Suppose that, after you had done these calculations, a friend asserted that the population distribution of selling prices of new homes in this city was almost certainly not normal. How would you respond?

Answer:
Even with non-normal populations, the sampling distribution of the sample means will be normal for sufficient sample *n*. Since *n* is 100, the sampling distribution of the sample means can be assumed to be a normal distribution.

THE NEXT THREE QUESTIONS ARE BASED ON THE FOLLOWING INFORMATION:
Candidates for employment at a city fire department are required to take a written aptitude test. Scores on this test are normally distributed with mean 260 and standard deviation 51. A random sample of nine test scores was taken.

105. What are the mean and standard error of the sampling distribution for the mean score?
Answer:
17.0

106. What is the probability that the sample mean score is less than 250?
Answer:
0.2776

107. What is the probability that the sample mean score is more than 230?
Answer:
0.9608

Chapter 6

THE NEXT FIVE QUESTIONS ARE BASED ON THE FOLLOWING INFORMATION:
Times spent studying by students in the week before final exams follow a normal distribution with standard deviation 9 hours. A random sample of 5 students was taken in order to estimate the mean study time for the population of all students.

108. What is the standard error of the sampling distribution for the mean score?
Answer:
4.025

109. What is the probability that the sample mean exceeds the population mean by more than 2.1 hours?
Answer:
0.3015

110. What is the probability that the sample mean is more than 3.2 hours below the population mean?
Answer:
0.2119

111. What is the probability that the sample mean differs from the population mean by ± 4.1 hours?
Answer:
0.6922

112. Suppose that a second (independent) random sample of 10 students was taken. Without doing the calculations, state whether the probabilities in the second, third and fourth questions would be higher, lower, or the same for the second sample.
Answer:
The probabilities in the second, third, and fourth questions would be lower for the second sample than that for the first sample.

THE NEXT THREE QUESTIONS ARE BASED ON THE FOLLOWING INFORMATION:
The number of hours spent studying by students on Michigan State campus in the week before final exams follows a normal distribution with standard deviation 9.5 hours. A random sample of these students is taken to estimate the population mean number of hours studying.

113. How large a sample is needed to ensure that the probability that the sample mean differs from the population mean by more than 2.1 hours is less than 0.05?
Answer:
79

114. Without doing the calculations, state whether a larger or smaller sample than that in the first question would be required to guarantee that the probability that the sample mean differs from the population mean by more than 2.1 hours is less than 0.10.
Answer:
Smaller sample

115. Without doing the calculations, state whether a larger or smaller sample than that in the first question would be required to guaranteed that the probability that the sample mean differs from the population mean by more than 1.8 hours is less than 0.05.

Answer:
Larger sample

<u>THE NEXT THREE QUESTIONS ARE BASED ON THE FOLLOWING INFORMATION:</u>

In taking a sample of *n* observations from a population of *N* members, the variance of the sampling distribution of the sample means is $\sigma_{\bar{X}}^2 = \frac{\sigma^2}{n} \cdot \frac{N-n}{N-1}$. The quantity $\frac{(N-n)}{(N-1)}$ is called the finite population correction factor.

116. To get some feeling for possible magnitudes of the finite population correction factor, calculate it for samples of $n = 15$ observations from populations of $N = 15$, 30, 60, 100, 500, 1,000, 5,000, and 10,000 members.

Answer:

Population size N	15	30	60	100	500	1,000	5,000	10,000
FPCF (N – n) / (N -1)	0	0.517	0.763	0.859	0.972	0.986	0.997	0.999

117. Explain why the result for $n = 15$, found in the first question, is precisely what one should expect on intuitive grounds.

Answer:

When the population size (*N*) equals the sample size (*n*), then there is no variation away from the population mean and the standard error will be zero. As the sample size becomes relatively small compared to the population size, the correction factor tends toward 1 and becomes less significant in the calculation of the standard error.

118. Given the results in the first question, discuss the practical significance of using the finite population correction factor for samples of 20 observations from populations of different sizes.

Answer:
The correction factor tends toward a value of 1 and becomes progressively less important as a modifying factor when the sample size decreases relative to the population size.

THE NEXT FOUR QUESTIONS ARE BASED ON THE FOLLOWING INFORMATION:
According to the Internal Revenue Service, 80% of all tax returns lead to a refund. A random sample of 100 tax returns is taken.

119. What is the mean of the sampling distribution of the sample proportion of returns leading to refunds?

Answer:
0.80

120. What is the standard error of the sample proportion?
Answer:
0.04

121. What is the probability that the sample proportion exceeds 0.85?
Answer:
0.1056

122. What is the probability that the sample proportion is between 0.78 and 0.84?
Answer:
0.5328

THE NEXT THREE QUESTIONS ARE BASED ON THE FOLLOWING INFORMATION:
A store has determined that 28% of all refrigerators purchasers will also purchase a service agreement. In one month, 250 refrigerators are sold to customers who can be regarded as a random sample of all purchasers.

123. What are the mean and standard error of the sampling distribution for the proportion of those who will purchase a service agreement?
Answer:
0.0284

124. What is the probability that the sample proportion will be less than 0.30?
Answer:
0.7580

125. Without doing the calculations, state in which of the following ranges the sample proportion is most likely to be: 0.27 to 0.29, 0.28 to 0.30, 0.29 to 0.31, and 0.30 to 0.32.
Answer:
The sample proportion is most likely to be in the range 0.27 to 0.29.

THE NEXT FOUR QUESTIONS ARE BASED ON THE FOLLOWING INFORMATION:

A random sample of size n =25 is obtained from a normally distributed population with population mean μ =200 and variance σ^2 =100.

126. What are the mean and standard deviation of the sampling distribution for the sample means?
Answer:
2.0

127. What is the probability that the sample mean is greater than 203?
Answer:
0.0668

128. What is the value of the sample variance such that 5% of the sample variances would be less than this value?
Answer:
0.05
Hence, 0.24k = 13.85 and k = 57.7. Therefore, if the sample variance is 57.7, then 5% of the sample variances would be less than this value.

129. What is the value of the sample variance such that 5% of the sample variances would be greater than this value?
Answer:
Hence, 0.24k = 36.42 and k = 151.75. Therefore, if the sample variance is 151.75, then 5% of the sample variances would be greater than this value.

THE NEXT TWO QUESTIONS ARE BASED ON THE FOLLOWING INFORMATION:

Monthly rates of return on the shares of a particular common stock are independent of one another and normally distributed with a standard deviation of 1.8. A sample of 15 months is taken.

130. Find the probability that the sample standard deviation is less than 2.6.
Answer:
0.99

131. Find the probability that the sample standard deviation is more than 1.2.
Answer:
$\approx$ Between 0.95 and 0.975.

THE NEXT THREE QUESTIONS ARE BASED ON THE FOLLOWING INFORMATION:

A random sample of 10 stock market mutual funds was taken. Suppose that rates of returns on the population of stock market mutual funds follow a normal distribution.

132. The probability is 0.10 that sample variance is bigger than what percentage of the population variance?
Answer:
1.6316 or 163.16%

133. Find any pair of numbers, a and b, to complete the following sentence: The probability is 0.95 that the sample variance is between a% and b% of the population variance.
Answer:
a = 0.30 and b = 2.1137, and the probability is 0.95 that the sample variance is between 30% and 211.37% of the population variance.

134. Suppose that a sample of 20 mutual funds had been taken. Without doing the calculations, indicate how this would change your answer to the second question.
Answer:
The 95% interval in the second question would be smaller.

THE NEXT THREE QUESTIONS ARE BASED ON THE FOLLOWING INFORMATION:

Each member of a random sample of 20 business economists was asked to predict the rate of inflation for the coming year. Assume that the predictions for the whole population of business economists follow a normal distribution with standard deviation 2%.

135. The probability is 0.01 that the sample standard deviation is larger than what value?
Answer:
Hence k = 2.76. Therefore, the probability is 0.01 that the sample standard deviation is larger than 2.76%.

136. The probability is 0.025 that the sample standard deviation is smaller than what value?
Answer:
Hence k = 1.369. Therefore, the probability is 0.025 that the sample standard deviation is smaller than 1.369%.

137. Find any pair of values such that the probability that the sample standard deviation lies between these values is 0.90.
Answer:
The probability that the sample standard deviation lies between 1.459% and 2.519% is 0.90.

138. Find the sampling distribution of sample means if all possible samples of size 2 are drawn with replacement from the following population:

X	-2	0	2
$p(x)$	0.2	0.6	0.2

Answer:

$\bar{x}$	-2	-1	0	1	2
$p(\bar{x})$	0.04	0.24	0.44	0.24	0.04

139. Why does the sample size play such an important role in reducing the standard error of the mean? What are the implications of increasing the sample size?
Answer:
The standard error of the mean is the standard deviation of the population you are sampling from divided by the square root of the sample size. So, mathematically as the sample size increases, the standard error naturally decreases. But there is more to this, because the standard error is the standard deviation of the sample means. So, as the sample size increases, the sample means are deviating less and less from the true population mean. Hence, as we sample more, we get statistics which are closer to the true parameters and our inference methods will improve. This is true for sampling distributions of mean, proportions, and variances.

THE NEXT FIVE QUESTIONS ARE BASED ON THE FOLLOWING INFORMATION:
The following data give the number of pets owned for a population of 4 families.

Family	A	B	C	D
Number of Pets Owned	2	1	4	3

140. Find the population mean and population standard deviation.
Answer:
μ =2.5 and σ =1.118

141. Samples of size 2 will be drawn at random from the population. Use the formulas $E(\bar{X}) = \mu$ and $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$ to calculate the mean and the standard deviation of the sampling distribution of sample means.
Answer:
$E(\bar{X}) = \mu$ =2.5, and $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$ = 0.6455

142. List all possible samples of 2 families that can be selected without replacement from this population, and compute the sample mean for each sample.
Answer:

Sample	AB	AC	AD	BC	BD	CD
$\bar{x}$	1.5	3.0	2.5	2.5	2.0	3.5

143. Use your answer to the previous question to determine the sampling distribution of sample means.
Answer:

$\bar{x}$	1.5	2.0	2.5	3.0	3.5
$p(\bar{x})$	1/6	1/6	1/3	1/6	1/6

THE NEXT FOUR QUESTIONS ARE BASED ON THE FOLLOWING INFORMATION:
The chairman of the statistics department in a certain college believes that 70% of the department's graduate assistantships are given to international students. A random sample of 50 graduate assistants is taken.

144. Assume that the chairman is correct and P = 0.70. What is the sampling distribution of the sample proportion $\hat{p}$?
Answer:
The sampling distribution of the sample proportion is approximately normal, since $nP(1 - P) > 5$.

145. Find the expected value and the standard error of the sampling distribution of $\hat{p}$.
Answer:
$E(\hat{p})$ = 0.70, and $\sigma_{\hat{p}}$ = 0.0648

146. What is the probability that the sample proportion $\hat{p}$ will be between 0.65 and 0.73?
Answer:
0.4566

147. Why is the central limit theorem important to statistical analysis?

ANSWER:
Although the normal distribution occurs frequently and describes many populations, it is not the only probability distribution in existence. The value of the central limit theorem is found in its conclusion that regardless of the shape of the population distribution (i.e. samples can come from any type of distribution), the sampling distribution of the mean will form a normal distribution. The central limit theorem provides the basis for considerable work in applied statistical analysis. Many random variables can be modeled as sums or means of independent random variables, and the normal distribution provides a good approximation of the true distribution. It can be applied to both discrete and continuous random variables.

148. What is an acceptance interval?

Answer:
An acceptance interval is an interval within which a sample mean has a higher probability of occurring, given that the population mean and variance is known. If the sample mean is within that interval, then the conclusion can be accepted that the random sample came from the population with the known population mean and variance. Thus acceptance intervals provide an operating rule for process-monitoring applications. The probability that the sample mean is within a particular interval can be computed if the sample means have a distribution that is close to normal. Acceptance intervals can also be computed for nonnormal probability distributions. Acceptance intervals find wide application for monitoring manufacturing processes to determine if product standards continue to be achieved.