Statistical Inference Theory
Lesson 29
Estimating the Mean μ of a Population

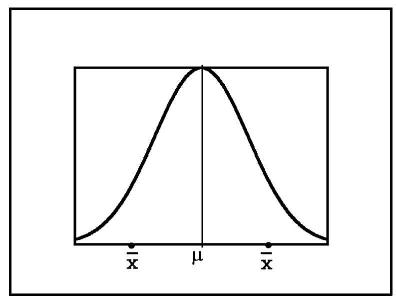
Since μ generally is not known, one of the goals of inference theory is to use $\overline{X}$ as an estimation of μ. There are two types of estimates: point estimate and interval estimate. In either type of estimate, $\overline{X}$ is substituted in place of μ. This substitution creates an error.

29.1 - What is the error created when using a point estimate?

a. The following formula equals the error created when the standard deviation σ of the population is known:

$$\text{Error} = e^* = \pm(\overline{X} - \mu) = \pm z\frac{\sigma}{\sqrt{N}} = \pm z\sigma_{\overline{X}},$$

where σ is the standard deviation of the population and N is the sample size.



b. The following formula is the error created when σ of the population is not known:

$$\text{Error} = e^* = \pm(\overline{X} - \mu) = \pm z\frac{s}{\sqrt{N}}, = \pm z s_{\overline{X}},$$

where s is the standard deviation of the sample of size N.

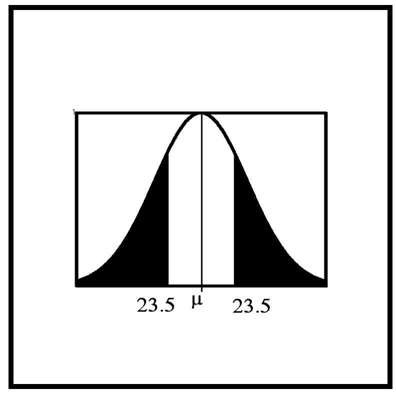
$$Z = \frac{\overline{X} - \mu}{\sigma_{\overline{X}}} = \frac{e^*}{\sigma_{\overline{X}}}$$

29.1 - Example 1: A large university wants to estimate the average age of its students. A random sample of size 100 is taken of the student body. From the sample, the average age is $\overline{X} = 23.5$ years and s = 2.1 years. Assume $\overline{X} = 23.5$ replaces μ.

(a). Find the probability that the error created exceeds 1/2 year.

(b). Find the minimum sample size so that the probability is 0.05 of making an error that exceeds 1/2 year.

1



Solutions:

➤ **(a)**.
It is almost certain that $\overline{X}$ is smaller or larger than μ.
fig. 1

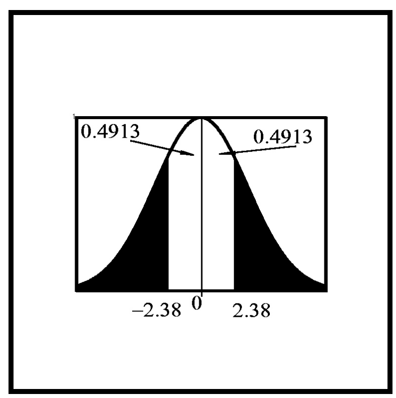
The difference between $\overline{X}$ and μ is the error $e^* = \pm(\overline{X} - \mu)$.

We need to find the probability that the error e* exceeds 1/2 year.

Step 1: Since we only have the standard deviation of the sample, s = 2.1 .

Step 2: The sample size N = 100

2



Step 3: $e^* = \pm(\overline{X} - \mu) = \pm z\frac{s}{\sqrt{N}} = \pm z\frac{2.1}{\sqrt{100}} = \pm z(0.21) = \pm 1/2 = \pm 0.5$

Step 4: Solving for z gives

$z = \pm\frac{0.5}{0.21} = \pm 2.38$.

Step 5: From the normal distribution table:
fig. 2

P{e* > 1/2} = 1 - 0.4913 - 0.4913 = 0.0174

➤ **(b)**.
Step 1: For the probability that the error will exceed 1/2 year is 0.05, we find z for the area 0.5 - 0.05/2 = 0.475: z = 1.96 .

Step 2: Since $e^* = \pm z\frac{2.1}{\sqrt{N}} = \pm(1.96)\frac{2.1}{\sqrt{N}} = \pm(1.96)\frac{2.1}{\sqrt{N}} = \pm\frac{4.116}{\sqrt{N}} = \pm 0.5$

Step 3: Solving for N gives

$\sqrt{N} = 2(4.116) = 8.226$ and

$N \approx 68$, minimum sample size.

Solved Problems

29.1 - Solved Problem 1: A company that manufactures a new gasoline additive is interested in testing the additive to determine the average additional mileage it will give to consumers. It selects 36 different cars and runs each car for 100 miles. The final results showed that the average increase of mileage was 2.1 miles per gallon with a standard deviation of s = 0.5 miles per gallon. Assume $\overline{X} = 2.1$ replaces μ.

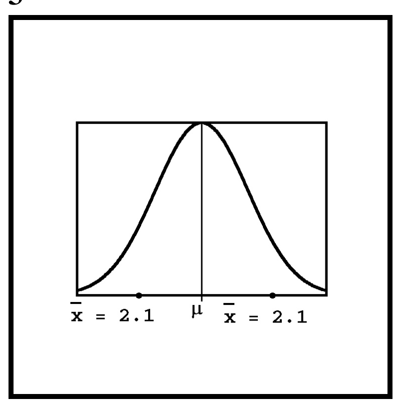
(a). Find the probability that the error created exceeds 0.1 miles per gallon.

(b). Find the minimum sample size so that the probability is 0.02 of making an error that exceeds 0.1 miles per gallon.

Solutions:
➤ **(a)**.
fig. 3

It is almost certain that $\overline{X}$ is smaller or larger than μ. The difference between $\overline{X}$ and μ is the error $e^* = \pm(\overline{X} - \mu)$. We need to find the probability that the error exceeds 0.1 gallons.

3



Step 1: Since we only have the standard deviation of the sample, s = 0.5 .

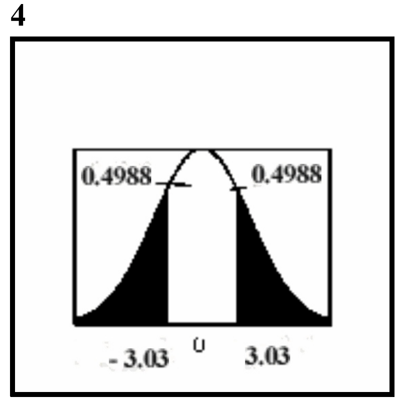
Step 2: The sample size N = 36 .

Step 3: $e^* = \pm(\overline{X} - \mu) = \pm z\frac{s}{\sqrt{N}} = \pm z\frac{0.5}{\sqrt{36}} \approx \pm z(0.08) = \pm 0.1$ gallons

Step 4: Solving for z gives $z = \frac{0.1}{0.08} = 1.25$.

Step 5: From the normal distribution table:
fig. 4

4



$P\{e^* > 0.1\} = 1 - 0.3944 - 0.3944 = 0.2112$.

➤ **(b)**.
Step 1: For the probability that the error will exceed 0.1 gallons is 0.02, we find z for the area 0.5 - 0.02/2 = 0.49: z = 2.33 .

Step 2: $e^* = \pm z\frac{0.5}{\sqrt{N}} = \pm(2.33)\frac{0.5}{\sqrt{N}} = \pm(2.33)\frac{0.5}{\sqrt{N}} = \pm\frac{1.165}{\sqrt{N}} = \pm 0.1$

Step 3: Solving for N gives

$\sqrt{N} = 10(1.165) = 11.65,$

$N \approx 136$, minimum sample size.

Unsolved Problems with Answers

29.1 - Problem 1: A machine fills bottles with orange juice with a standard deviation of $\sigma = 1.5$ ounces. Each hour a sample of 50 filled bottles is taken. Assume the average from this sample is 12.3 ounces.

(a). Find the probability that the error from the true average filled exceeds 0.5 ounces.

(b). Find the minimum sample size so that the probability is 0.01 of making an error that exceeds 0.5 ounces

Answers:

➤ **(a)**. 0.02

➤ **(b)**. $N \approx 60$

⇑ *Refer back to* **29.1 - Example 1 & 29.1 - Solved Problem 1.**

29.2 - What is the error created when using a Confidence Interval estimate ?

The interval estimate of μ is called the confidence interval given by the following formulas:

1. If σ of the population is known:

$$\overline{X} - z\sigma_{\overline{X}} \le \mu \le \overline{X} + z\sigma_{\overline{X}}.$$

2. If σ is not known then use s, the standard deviation of the sample:

$$\overline{X} - z\frac{s}{\sqrt{N}} \le \mu \le \overline{X} + z\frac{s}{\sqrt{N}}$$

$$\overline{X} - zs_{\overline{X}} \le \mu \le \overline{X} + zs_{\overline{X}}.$$

The value z is determined according to the confidence in μ within the given interval.

29.2 - Example 1: A large university wants to estimate the average age of its students. A random sample of size 100 is taken of the student body. From the sample, the average age is $\overline{X}$ = 23.5 years and s = 2.1 years.

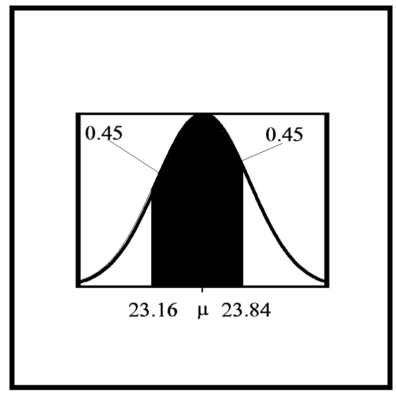
(a). Find a 90% confidence interval for μ.

(b). Find a 95% confidence interval for μ.

Solutions:

➤ **(a)**.
Since the confidence interval is 90%, we use the area 0.90/2 = 0.45 to find z = 1.64 .
fig. 5

5



Step 1: s = 2.1

Step 2: N = 100

Step 3: $s_{\overline{X}} = \pm\frac{s}{\sqrt{N}} = \pm 2.1/10 = \pm 0.21$

Step 4: $\overline{X} - z\frac{s}{\sqrt{N}} \leq \mu \leq \overline{X} + z\frac{s}{\sqrt{N}}$:

$23.5 - 1.64(0.21) \leq \mu \leq 23.5 + 1.64(0.21)$

which gives

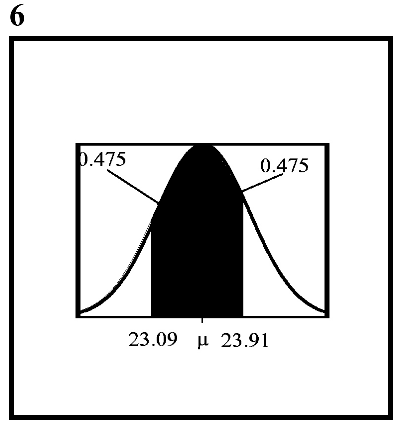
$23.16 \leq \mu \leq 23.84$.

Step 5: The value for μ ranges between 23.16 and 23.84 with 90% probability.

➤ **(b)**.
Since the confidence interval is 95%, we use the area 0.95/2 = 0.475 to find z = 1.96 .

Step 1: s = 2.1
fig. 6

6



Step 2: N = 100

Step 3: $s_{\overline{X}} = \pm\frac{s}{\sqrt{N}} = \pm 2.1/10 = \pm 0.21$

Step 4: $\overline{X} - z\frac{s}{\sqrt{N}} \leq \mu \leq \overline{X} + z\frac{s}{\sqrt{N}}$

$23.5 - 1.96(0.21) \leq \mu \leq 23.5 + 1.96(0.21)$

which gives $23.09 \leq \mu \leq 23.91$.

Step 5: The value for μ ranges between 23.09 and 23.91 with 95% probability.

Solved Problems

29.2 - Problem 1: A company that manufactures a new gasoline additive is interested in testing the additive to determine the average additional mileage it will give to consumers. It selects 36 different cars and runs each car for 100 miles. The final results showed that the average increase of mileage was 2.1 miles per gallon with a standard deviation of s = 0.2 miles per gallon.

(a). Find a 92% confidence interval for μ.

(b). Find a 99% confidence interval for μ.

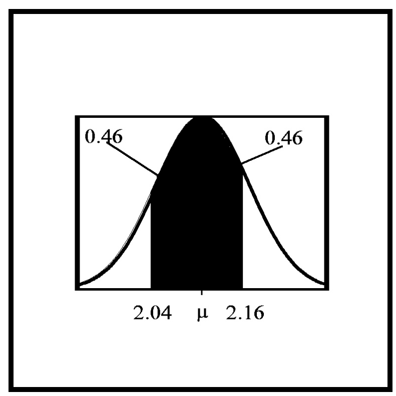
Solutions:

➤**(a)**.
Since the confidence interval is 92%, we use the area $\frac{0.92}{2} = 0.46$ to find z = 1.75 .

Step 1: $\overline{X} = 2.1$ and $s = 0.2$

fig. 7

7



Step 2: N = 36

Step 3: $e^* = \pm\frac{zs}{\sqrt{N}} = \pm 1.75\frac{0.2}{\sqrt{36}} \approx \pm 0.058$

Step 4: $\overline{X} - z\frac{s}{\sqrt{N}} \le \mu \le \overline{X} + z\frac{s}{\sqrt{N}}$

$2.1 - 0.058 \le \mu \le 2.1 + 0.058$

which gives

$2.04 \le \mu \le 2.16$.

Step 5: The value for μ ranges between 2.04 and 2.16 with 92% probability.

➤ **(b)**.
Since the confidence interval is 99%, we use the area 0.99/2 = 0.495 to find z = 2.57 .

Step 1: $\overline{X} = 2.1$ and $s = 0.2$

Step 2: N = 36

Step 3: $e^* = \pm\frac{zs}{\sqrt{N}} = \pm\frac{(2.57)0.2}{6} = \pm 0.033$

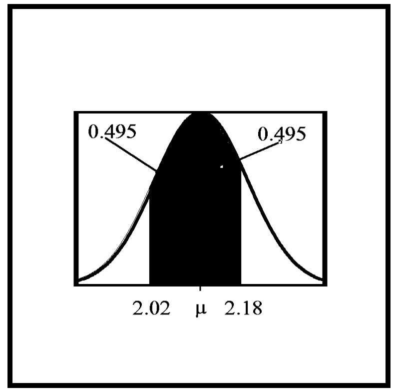
Step 4: $\bar{X} - z\frac{s}{\sqrt{N}} \le \mu \le \bar{X} + z\frac{s}{\sqrt{N}}$

$2.1 - 2.57(0.033) \le \mu \le 2.1 + 2.57(0.033)$

which gives $2.02 \le \mu \le 2.18$.

Step 5: The value for μ ranges between 2.02 and 2.18 with 99% probability.
fig. 8

8



Unsolved Problems with Answers

29.2 - Problem.1: A machine fills bottles with orange juice with a standard deviation of $\sigma = 1.5$ ounces. Each hour a sample of 50 filled bottles is taken. If the average from this sample is 12.3 ounces,

(a). Find a 90% confidence interval for μ.

(b). Find a 98% confidence interval for μ.

Answers:

➤ **(a)**. $11.96 \le \mu \le 12.64$

➤ **(b)**.$11.81 \le \mu \le 12.79$

⇑ *Refer back to* **29.2 - Example 1 & 29.2 - Solved Problem 1.**

29.3 - Determining the Sample Size.

In the first section of this lesson, for each example and problem, we derived an estimate for the minimum sample size needed under the condition that a given error will exceed a given amount for a specifed probability. In this lesson we give below the formula needed to derive the same minimum sample size, within a given confidence interval:

$$N = \frac{z^2\sigma^2}{e^{*2}},$$

where e* is the error, z is determined by the confidence of the estimate, and σ is the standard deviation of the population.

29.3 - Example.1: A machine fills bottles with orange juice with a standard deviation of $\sigma = 1.5$ ounces. Each hour a sample of filled bottles is to be taken. Find the sample size required to assure, with 90% confidence, that the true estimate of μ will not be off by more than

(a). 0.1 ounces.

(b). 0.2 ounces.

Solutions:

➤ **(a)**.
Since we want a confidence of 90%, we look up in the normal distribution table the area 0.45 . This gives $z = 1.64$. From the statement of the problem, $e^* = 0.1$ and $\sigma = 1.5$.
Therefore,

$$N = \frac{z^2\sigma^2}{e^{*2}} = \frac{(1.64)^2(1.5)^2}{(0.1)^2} \approx 605 \text{ bottles per sample.}$$

➤ **(b)**.
Since we want a confidence of 90%, we look up in the normal distribution table the area 0.45 . This gives $z = 1.64$. From the statement of the problem, $e^* = 0.2$ and $\sigma = 1.5$. Therefore,

$$N = \frac{z^2\sigma^2}{e^{*2}} = \frac{(1.64)^2(1.5)^2}{(0.2)^2} \approx 151 \text{ bottles per sample.}$$

Solved Problems

29.3 - Solved Problem 1: A medical journal needs a sample to determine the true average length of time it takes for male patients to recover from heart surgery. They first did a preliminary study and found the standard deviation to be 72 days. For a confidence of 95%, Find the sample size needed to estimate the true mean average within

(a). 2 days.

(b). 10 days.

Solutions:

➤**(a)**.
Since we want a confidence of 95%, we look up in the normal distribution table the area 0.475 . This gives $z = 1.96$. From the statement of the problem, $e^* = 2$ and $\sigma = 72$. Therefore,

$$N = \frac{z^2\sigma^2}{e^{*2}} = \frac{(1.96)^2(72)^2}{(2)^2} \approx 4{,}979 \text{ patients}$$

➤**(b)**.
Since we want a confidence of 95%, we look up in the normal distribution table the area 0.475 . This gives $z = 1.96$. From the statement of the problem, $e^* = 10$ and $\sigma = 72$. Therefore,

$$N = \frac{z^2\sigma^2}{e^{*2}} = \frac{(1.96)^2(72)^2}{(10)^2} \approx 199 \text{ patients}$$

Unsolved Problems with Answers

29.3 - Problem 1: Over the years a large liberal arts college claims that the standard deviation of their students' grade point average is 0.3 . For a confidence of 99%, find the sample size needed to estimate the true mean average within

(a). 0.05 points.

(b). 0.10 .

Answers:

➤**(a)**. $N \approx 239$ students

➤**(b)**. $N \approx 59$ students

⇑ *Refer back to* **29.3 - Example 1 & 29.3 - Solved Problem 1.**

Supplementary Problems

1. A fair coin is tossed 100 times resulting in the value X which is the number of heads that occurs in the sample. Assume X is used as the estimator of the true average number of heads that should appear in the sample.

a. Find the probability that the error created will be 10 heads or more.

b. Find the range of X values that can occur with approximate probability 0.90 .

2. A recent study of medical schools in California showed that 60% of all medical students are female. A random survey of 200 medical students was taken. Let X be the number of female students in the sample. Assume X is used as the estimator of the true average number of females that should result from this sample.

a. Find the probability that X will deviate from μ by 20 females or more.

b. Find the range of X values that can occur with approximate probability 0.95 .

3. The Clear Water Bottling Company has a machine that fills bottles with spring water distributed according to a normal distribution, where $\mu = 16.2$ ounces and $\sigma = 0.15$ ounces. A bottle is said to be under-filled if it contains less than 16 ounces. A random sample of 200 filled bottles is taken and checked for the quantity of spring water in each bottle. Let X be the number of bottles in this sample that contain less than 16 ounces.

a. Find the probability that X will deviate from μ by at least 5 under- filled bottles.

b. Find the range of X values that can occur with probability 0.90 .

4. A Las Vegas casino tested a recently purchased machine that shuffles a card deck containing 52 cards. To test the machine for randomness, it has the machine deal 1000 randomly shuffled 5 card hands. Let X be the number of hands from this sample that contains all black cards.

a. Find μ, the average number of hands that contain all black cards.

b. Find the probability that X will deviate from μ by at least 5 such hands.

c. Find the range of X values that can occur with probability 0.95 .

5. Assume a sample of size 1,000 is taken from a binomial population. Let X be the estimator of μ. If 90% of all X have a range from μ of 12 units or less, find $p > q$, μ and σ of the binomial distribution.

6. From the Central Limit Theorem we have $\mu - z\sigma_{\overline{X}} \leq \overline{X} \leq \mu + z\sigma_{\overline{X}}$.

Show $\overline{X} - z\sigma_{\overline{X}} \leq \mu \leq \overline{X} + z\sigma_{\overline{X}}$.

7. A survey of 100 banks was taken on the interest rates they charge on home mortgages. The following frequency table shows these results:

Interest Rates on Home Mortgages	**Number of Banks**
7.5%	24
8.1%	55
9.0%	11
9.4%	10
Total	100

Find:

a. $\overline{X}$.

b. the standard error of the mean $\sigma_{\overline{X}} = \dfrac{s}{\sqrt{N}}$.

c. the probability that the error $e^* = \overline{X} - \mu$ exceeds 0.1%.

8. The Tammy May Weight Reduction Center recently took a random survey of 36 members that have lost over 10 pounds. The following frequency table shows the amount of weight lost:

Number of pounds lost (rounded to nearest pound)	**Number of members**
23	6
35	10
40	15
45	5
Total	36

Find:

a. $\overline{X}$.

b. the standard error of the mean $\sigma_{\overline{X}} = \frac{s}{\sqrt{N}}$.

c. the probability that the average weight of these 36 members is off by more than 1 pound from the true average of members that lost more than 10 pounds.

9. In a speed reading class, the following table shows the time it took 100 students to each read a given novel::

Time it took to read novel (rounded to nearest minute)	**Number of readers**
128	16
150	34
167	20
205	15
250	15

Find:

a. $\overline{X}$.

b. the standard error of the mean is $\sigma_{\overline{X}} = \frac{s}{\sqrt{N}}$.

c. the probability that the average time of these 100 readers to finish the novel is off by more than 10 minutes from the true average time it should take.

10. The determination of a minimum sample size is given by the formula

$$N = \frac{z^2\sigma^2}{e^{*2}}.$$

where e^* is the desired error.

a. If e* is changed by a factor a >0, show the new sample size is $N_1 = N/a^2$.

b. If N = 100 and e* is decreased by 50%, find N_1.

c. If N = 100 and e* is increased by 10%, find N_1.

11. The error is given by

$$e^* = \frac{z\sigma}{\sqrt{N}}.$$

a. If N is changed by a factor a >0, show the new error is $e_1^* = \frac{e*}{\sqrt{a}}$.

b. If e* = 0.1 and N is decreased by 50%, find e_1*.

c. If e* = 0.01 and N is increased by 100%, find e_1*.