STATISTICAL METHODS

Lisbon School of Economics & Management
Universidade de Lisboa

Master in Industrial Management,

Operations and Sustainability (MIMOS)

2nd year/1st Semester

2025/2026

CONTACT

Professor: Elisabete Fernandes
E-mail: efernandes@iseg.ulisboa.pt

https://doity.com.br/estatistica-aplicada-a-nutricao

https://basiccode.com.br/produto/informatica-basica/

PROGRAM

1. Fundamental Concepts of Statistics
2. Descriptive Data Analysis
3. Introduction to Inferential Analysis
4. Parametric Hypothesis Testing
5. Non-Parametric Hypothesis Testing
6. Linear Regression Analysis

LECTURE 8 HOMEWORK: QUESTIONS AND SOLUTIONS

EXERCISE 6.8

6.8 Given a population with mean $\mu = 400$ and variance $\sigma^2 = 1,600$, the central limit theorem applies when the sample size is $n \geq 25$. A random sample of size $n = 35$ is obtained.

a. What are the mean and variance of the sampling distribution for the sample means?
b. What is the probability that $\bar{x} > 412$?
c. What is the probability that $393 \leq \bar{x} \leq 407$?
d. What is the probability that $\bar{x} \leq 389$?

Newbold et al (2013)

EXERCISE 6.8 A): SOLUTION

✅ Answer:

Given: population mean μ = 400, population variance $\sigma^2 = 1600 \rightarrow \sigma = 40$, sample size n = 35.

By the Central Limit Theorem:

$$\bar{X} \sim N\left(400,\ \frac{1600}{35}\right)$$

Standard error:

Central Limit Theorem: If $n \geq 25$ then $\overline{X} \sim \text{Normal}(\mu, \sigma^2/n)$

$$\sigma_{\bar{X}} = \frac{40}{\sqrt{35}} = 6.76$$

$\mu_{\bar{X}} = \mu$ and $\sigma_{\bar{X}} = \dfrac{\sigma}{\sqrt{n}}$

(a) Mean and variance of the sampling distribution

$$E(\bar{X}) = 400, \quad Var(\bar{X}) = \frac{1600}{35} = 45.71$$

EXERCISE 6.8 B): SOLUTION

Answer:

Given: population mean μ = 400, population variance $\sigma^2 = 1600 \rightarrow \sigma = 40$, sample size n = 35.

By the Central Limit Theorem:

$$\bar{X} \sim N\left(400,\ \frac{1600}{35}\right)$$

Standard error:

$$\sigma_{\bar{X}} = \frac{40}{\sqrt{35}} = 6.76$$

(b) Probability that $\bar{X} > 412$

$$z = \frac{412 - 400}{6.76} = 1.77$$

Standard Normal Distribution Table

$$P(\bar{X} > 412) = 1 - \Phi(1.77) = 1 - 0.9616 = 0.0384$$

$\bar{X} \sim$ **Normal(μ, σ²/n)** → $Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim$ **Normal(0, 1)**

EXERCISE 6.8 C): SOLUTION

✅ Answer:

Given: population mean $\mu = 400$, population variance $\sigma^2 = 1600 \rightarrow \sigma = 40$, sample size $n = 35$.

By the Central Limit Theorem:

$$\bar{X} \sim N\left(400,\ \frac{1600}{35}\right)$$

Standard error:

$$\sigma_{\bar{X}} = \frac{40}{\sqrt{35}} = 6.76$$

(c) Probability that $393 \leq \bar{X} \leq 407$

$$z_1 = \frac{393 - 400}{6.76} = -1.04, \quad z_2 = \frac{407 - 400}{6.76} = 1.04$$

$$P = \Phi(1.04) - \Phi(-1.04) = 2 \times \Phi(1.04) - 1 = 2 \times 0.8508 - 1 = 0.7016$$

Standard Normal Distribution Table

EXERCISE 6.8 D): SOLUTION

Answer:

Given: population mean μ = 400, population variance $\sigma^2 = 1600 \rightarrow \sigma = 40$, sample size n = 35.

By the Central Limit Theorem:

$$\bar{X} \sim N\left(400,\ \frac{1600}{35}\right)$$

Standard error:

$$\sigma_{\bar{X}} = \frac{40}{\sqrt{35}} = 6.76$$

(d) Probability that $\bar{X} \leq 389$

$$z = \frac{389 - 400}{6.76} = -1.63$$

$$P(\bar{X} \leq 389) = \Phi(-1.63) = 1 - \Phi(1.63) = 0.052$$

Standard Normal Distribution Table

EXERCISE 6.17

6.17 The times spent studying by students in the week before final exams follows a normal distribution with standard deviation 8 hours. A random sample of four students was taken in order to estimate the mean study time for the population of all students.

a. What is the probability that the sample mean exceeds the population mean by more than 2 hours?
b. What is the probability that the sample mean is more than 3 hours below the population mean?
c. What is the probability that the sample mean differs from the population mean by more than 4 hours?
d. Suppose that a second (independent) random sample of 10 students was taken. Without doing the calculations, state whether the probabilities in parts (a), (b), and (c) would be higher, lower, or the same for the second sample.

Newbold et al (2013)

EXERCISE 6.17 A): SOLUTION

Answer:

Given: population standard deviation $\sigma = 8$ hours, first sample size $n = 4$.

Standard error of the sample mean:

$$SE = \frac{\sigma}{\sqrt{n}} = \frac{8}{\sqrt{4}} = 4$$

$\bar{X} \sim N(\mu, 4^2)$

$\bar{X}$ ~ Normal(μ, σ^2/n)

(a) Probability that the sample mean exceeds μ by more than 2 hours

$$P(\bar{X} - \mu > 2) = P(Z > \frac{2}{4}) = P(Z > 0.5)$$

$$P = 1 - \Phi(0.5) \approx 1 - 0.6915 = 0.3085$$

Standard Normal Distribution Table

EXERCISE 6.17 B): SOLUTION

Answer:

Given: population standard deviation $\sigma = 8$ hours, first sample size $n = 4$.

Standard error of the sample mean:

$$SE = \frac{\sigma}{\sqrt{n}} = \frac{8}{\sqrt{4}} = 4$$

$\bar{X} \sim N(\mu, 4^2)$

$\bar{X}$ ~ Normal(μ, σ^2/n)

(b) Probability that the sample mean is more than 3 hours below μ

$$P(\mu - \bar{X} > 3) = P(\bar{X} - \mu < -3) = P(Z < \frac{-3}{4}) = P(Z < -0.75)$$

$$P \approx 0.2266$$

Standard Normal Distribution Table

EXERCISE 6.17 C): SOLUTION

Answer:

Given: population standard deviation $\sigma = 8$ hours, first sample size $n = 4$.

Standard error of the sample mean:

$$SE = \frac{\sigma}{\sqrt{n}} = \frac{8}{\sqrt{4}} = 4$$

$\bar{X} \sim N(\mu, 4^2)$

$\bar{X}$ ~ Normal(μ, σ^2/n)

(c) Probability that the sample mean differs from μ by more than 4 hours

Two-sided:

$$P(|\bar{X} - \mu| > 4) = 1\text{-}P(-4/4 < (\bar{X} - \mu)/4 < 4/4)$$
$$= 1 - P(-1 < Z < 1)$$

$$P = 2(1 - \Phi(1)) = 2(1 - 0.8413) = 0.3174$$

Standard Normal Distribution Table

EXERCISE 6.17 D): SOLUTION

Answer:

Given: population standard deviation $\sigma = 8$ hours, first sample size $n = 4$.

Standard error of the sample mean:

$$SE = \frac{\sigma}{\sqrt{n}} = \frac{8}{\sqrt{4}} = 4$$

$\bar{X} \sim N(\mu, 4^2)$

$\bar{X}$ ~ Normal(μ, σ^2/n)

(d) Second sample of $\boxed{n = 10}$

The standard error decreases:

$$SE = \frac{8}{\sqrt{10}} \approx 2.53$$

- For (a) probability of >2 hours: **increases** in z (2/2.53 = 0.791) → **lower probability**? Wait, check carefully:

  Actually, smaller SE → larger z for the same deviation → probability of exceeding 2 hours **decreases**. ✅

- For (b) probability of >3 hours below: same logic → **lower probability**.
- For (c) probability of >4 hours deviation: smaller SE → probability **decreases**.

So for the second sample (n=10): **all probabilities are lower**.

(a) Probability that the sample mean exceeds μ by more than 2 hours

$$P(\bar{X} - \mu > 2) = P(Z > \frac{2}{4}) = P(Z > 0.5)$$

(b) Probability that the sample mean is more than 3 hours below μ

$$P(\mu - \bar{X} > 3) = P(\bar{X} - \mu < -3) = P(Z < \frac{-3}{4}) = P(Z < -0.75)$$

(c) Probability that the sample mean differs from μ by more than 4 hours

Two-sided:

$$P(|\bar{X} - \mu| > 4) = P(Z > \frac{4}{4}) = P(Z > 1)$$

$$P = 2(1 - \Phi(1)) = 2(1 - 0.8413) = 0.3174$$

EXERCISE 6.19

6.19 The price-earnings ratios for all companies whose shares are traded on the New York Stock Exchange follow a normal distribution with a standard deviation of 3.8. A random sample of these companies is selected in order to estimate the population mean price-earnings ratio.

a. How large a sample is necessary in order to ensure that the probability that the sample mean differs from the population mean by more than 1.0 is less than 0.10?
b. Without doing the calculations, state whether a larger or smaller sample size compared to the sample size in part (a) would be required to guarantee that the probability of the sample mean differing from the population mean by more than 1.0 is less than 0.05.
c. Without doing the calculations, state whether a larger or smaller sample size compared to the sample size in part a would be required to guarantee that the probability of the sample mean differing from the population mean by more than 1.5 hours is less than 0.10.

Newbold et al (2013)

EXERCISE 6.19 A): SOLUTION

Answer:

Given: population standard deviation $\sigma = 3.8$. We want

$$P(|\bar{X} - \mu| > E) < \alpha,$$

for $E = 1.0$ and $\alpha = 0.10$. For a two-sided bound $2(1 - \Phi(z)) = \alpha \Rightarrow \Phi(z) = 1 - \alpha/2$. For $\alpha = 0.10$ we have $z = \Phi^{-1}(0.95) = 1.645$.

(a) How large must n be?

We solve $\frac{E\sqrt{n}}{\sigma} = z \Rightarrow \boxed{n = \left(\frac{z\sigma}{E}\right)^2}.$

$$n = \left(\frac{1.645 \times 3.8}{1.0}\right)^2 = (6.251)^2 \approx 39.075.$$

Round up to guarantee the requirement: $n = 40$.

Determine n such as P(| $\bar{X}$ - μ|> E) < α

$$n = \left(\frac{z\sigma}{E}\right)^2$$

Alternative Solution:

$1 - P\left(\frac{-1}{\frac{3.8}{\sqrt{n}}} < \frac{\bar{X}-\mu}{\frac{3.8}{\sqrt{n}}} < \frac{1}{\frac{3.8}{\sqrt{n}}}\right) = 0.1 \Leftrightarrow \Phi\left(1/\frac{3.8}{\sqrt{n}}\right) - \Phi\left(-1/\frac{3.8}{\sqrt{n}}\right) = 0.9 \Leftrightarrow$

$2\Phi\left(1/\frac{3.8}{\sqrt{n}}\right) - 1 = 0.9 \Leftrightarrow \Phi\left(1/\frac{3.8}{\sqrt{n}}\right) = 0.95 \Leftrightarrow$

$1/\frac{3.8}{\sqrt{n}} = z_{0.95} = 1.645 \Leftrightarrow n = (3.8 \times 1.645)^2 = 39.075$

Round up to guarantee the requirement: n = 40

EXERCISE 6.19 B): SOLUTION

Answer:

Given: population standard deviation $\sigma = 3.8$. We want

$$P\big(|\bar{X} - \mu| > E\big) < \alpha,$$

for $E = 1.0$ and $\alpha = 0.10$. For a two-sided bound $2(1 - \Phi(z)) = \alpha \Rightarrow \Phi(z) = 1 - \alpha/2$. For $\alpha = 0.10$ we have $\boxed{z = \Phi^{-1}(0.95) = 1.645.}$

(b) To make the probability < 0.05 (stricter), we need a larger critical z $\boxed{\text{(e.g. } z_{0.975} = 1.96)}$ while E stays the same, so $\left(\frac{z\sigma}{E}\right)^2$ is larger.

Answer: A larger sample than in (a) is required.

Determine n such as P(| $\bar{X}$ - μ|> E) < α → $n = \left(\frac{z\sigma}{E}\right)^2$

If z increases, then n increases

EXERCISE 6.19 C): SOLUTION

Answer:

Given: population standard deviation $\sigma = 3.8$. We want

$$P\big(|\bar{X} - \mu| > E\big) < \alpha,$$

for $E = 1.0$ and $\alpha = 0.10$. For a two-sided bound $2(1 - \Phi(z)) = \alpha \Rightarrow \Phi(z) = 1 - \alpha/2$. For $\alpha = 0.10$ we have $z = \Phi^{-1}(0.95) = 1.645$.

(c) If the allowable deviation E is increased to 1.5 (with the same target probability 0.10), n scales like $1/E^2$, so increasing E decreases the required n.

Answer: A smaller sample than in (a) is required.

Determine n such as P(| $\bar{X}$ - μ|> E) < α → $n = \left(\frac{z\sigma}{E}\right)^2$

If E increases, then n decreases

LECTURE 9: SAMPLING AND SAMPLING DISTRIBUTIONS

SAMPLING DISTRIBUTION OF THE SAMPLE PROPORTION

P = the proportion of the population having some characteristic

- Sample proportion $(\hat{p})$ provides an estimate of *P*:

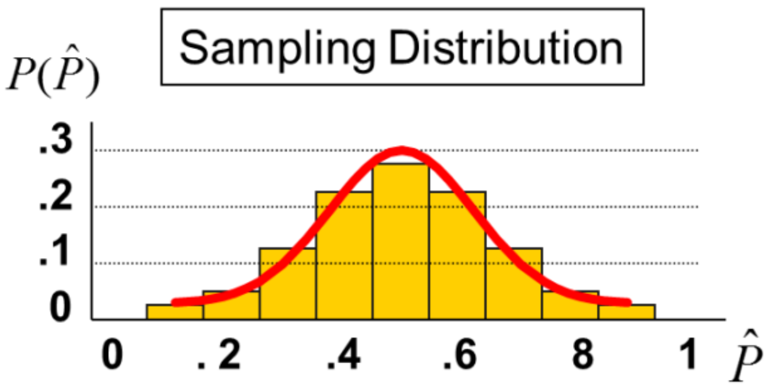
$$\boxed{\hat{p} = \frac{X}{n} = \frac{\text{number of items in the sample having the characteristic of interest}}{\text{sample size}}}$$

- $0 \leq \hat{p} \leq 1$
- $\hat{p}$ has a binomial distribution, but can be approximated by a normal distribution when $nP(1-P) > 5$

Newbold et al (2013)

SAMPLING DISTRIBUTION OF P HAT

- Normal approximation:



Properties: $E(\hat{p}) = P$ and $\sigma_{\hat{p}} = \sqrt{\dfrac{P(1-P)}{n}}$

(where *P* = population proportion)

Newbold et al (2013)

Z-VALUE FOR PROPORTIONS

Standardize $\hat{p}$ to a Z value with the formula:

$$Z = \frac{\hat{p} - P}{\sigma_{\hat{p}}} = \frac{\hat{p} - P}{\sqrt{\frac{P(1-P)}{n}}}$$

Where the distribution of Z is a good approximation to the standard normal distribution if $nP(1-P) > 5$

Newbold et al (2013)

Central Limit Theorem: If n ≥ 25 then $Z = \frac{\hat{p} - P}{\sqrt{\frac{P(1-P)}{n}}}$ ~ Normal(0, 1)

SAMPLE PROPORTION: EXAMPLE

- If the true proportion of voters who support Proposition *A* is $P = 0.4$, what is the probability that a sample of size 200 yields a sample proportion between 0.40 and 0.45?

- i.e.: **if $P = 0.4$ and $n = 200$, what is**

  $$P(0.40 \le \hat{p} \le 0.45)?$$

Newbold et al (2013)

SAMPLE PROPORTION: EXAMPLE

- **if P = 0.4 and n = 200, what is**

$$P\left(0.40 \le \hat{p} \le 0.45\right)?$$

$$Z = \frac{\hat{p} - P}{\sqrt{\frac{P(1-P)}{n}}} \sim \text{Normal}(0, 1)$$

Find $\sigma_{\hat{p}}$: $\sigma_{\hat{p}} = \sqrt{\frac{P(1-P)}{n}} = \sqrt{\frac{.4(1-.4)}{200}} = .03464$

Convert to standard normal:

$$\boxed{\begin{aligned} P(.40 \le \hat{p} \le .45) &= P\left(\frac{.40-.40}{.03464} \le Z \le \frac{.45-.40}{.03464}\right) \\ &= P(0 \le Z \le 1.44) \end{aligned}}$$

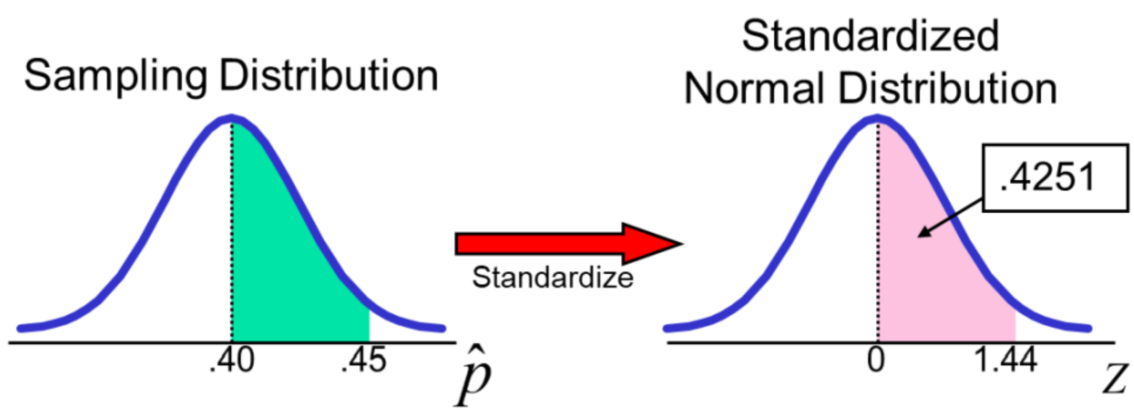
Newbold et al (2013)

SAMPLE PROPORTION: EXAMPLE

- if P = 0.4 and n = 200, what is

$$P\left(0.40 \le \hat{p} \le 0.45\right)?$$

Use standard normal table: $P(0 \le Z \le 1.44) = \Phi(1.44) - \Phi(0) =$ 0.9251 - 0.5 = 0.4251



Newbold et al (2013)

EXERCISE 6.31

6.31 According to the Internal Revenue Service, 75% of all tax returns lead to a refund. A random sample of 100 tax returns is taken.

a. What is the mean of the distribution of the sample proportion of returns leading to refunds?
b. What is the variance of the sample proportion?
c. What is the standard error of the sample proportion?
d. What is the probability that the sample proportion exceeds 0.8?

Newbold et al (2013)

EXERCISE 6.31 A): SOLUTION

Answer:

Given: population proportion $p = 0.75$, sample size $n = 100$.

For a sample proportion $\hat{p}$, the sampling distribution is approximately normal (by CLT) because n is large:

$$\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right)$$

Standard error: $SE = \sqrt{\frac{p(1-p)}{n}}$.

$$Z = \frac{\hat{p} - P}{\sqrt{\frac{P(1-P)}{n}}} \sim \text{Normal}(0, 1)$$

(a) Mean of the sampling distribution

$$E(\hat{p}) = p = 0.75$$

EXERCISE 6.31 B): SOLUTION

Answer:

Given: population proportion $p = 0.75$, sample size $n = 100$.

For a sample proportion $\hat{p}$, the sampling distribution is approximately normal (by CLT) because n is large:

$$\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right)$$

Standard error: $SE = \sqrt{\frac{p(1-p)}{n}}$.

$$Z = \frac{\hat{p} - P}{\sqrt{\frac{P(1-P)}{n}}} \sim \text{Normal}(0, 1)$$

(b) Variance of the sampling distribution

$$Var(\hat{p}) = \frac{p(1-p)}{n} = \frac{0.75 \cdot 0.25}{100} = \frac{0.1875}{100} = 0.001875$$

EXERCISE 6.31 C): SOLUTION

Answer:

Given: population proportion $p = 0.75$, sample size $n = 100$.

For a sample proportion $\hat{p}$, the sampling distribution is approximately normal (by CLT) because n is large:

$$\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right)$$

Standard error: $SE = \sqrt{\frac{p(1-p)}{n}}$.

$$\boldsymbol{Z = \frac{\hat{p} - P}{\sqrt{\frac{P(1-P)}{n}}} \sim \textbf{Normal}(0, 1)}$$

(c) Standard error of the sample proportion

$$SE = \sqrt{0.001875} \approx 0.0433$$

EXERCISE 6.31 D): SOLUTION

Answer:

Given: population proportion $p = 0.75$, sample size $n = 100$.

For a sample proportion $\hat{p}$, the sampling distribution is approximately normal (by CLT) because n is large:

$$\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right)$$

Standard error: $SE = \sqrt{\frac{p(1-p)}{n}}$.

$$Z = \frac{\hat{p} - P}{\sqrt{\frac{P(1-P)}{n}}} \sim \text{Normal}(0, 1)$$

(d) Probability that $\hat{p} > 0.8$

Compute z-score:

$$z = \frac{0.8 - 0.75}{0.0433} = \frac{0.05}{0.0433} \approx 1.154$$

$$P(\hat{p} > 0.8) = 1 - \Phi(1.154) \approx 1 - 0.875 = 0.125$$

Standard Normal Distribution Table

SAMPLING DISTRIBUTION OF THE SAMPLE VARIANCE

- Let $x_1, x_2, \ldots, x_n$ be a random sample from a population. The sample variance is

$$\boxed{s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

- the square root of the sample variance is called the sample standard deviation
- the sample variance is different for different random samples from the same population

Newbold et al (2013)

MEAN AND VARIANCE OF THE SAMPLE VARIANCE

- The sampling distribution of s^2 has mean σ^2

$$\boxed{E(s^2) = \sigma^2}$$

- If the population distribution is normal, then

$$\boxed{Var(s^2) = \frac{2\sigma^4}{n-1}}$$

Newbold et al (2013)

SAMPLE VARIANCE AND CHI-SQUARE DISTRIBUTION

1. **Sample variance for a normal population:**

$$\boxed{S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

2. **Relationship with the chi-square distribution:**

   If $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$, then:

$$\boxed{\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}}$$

Q = $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2{}_{(n-1)}$

3. **Implications:**

- The sample variance is **proportional to a chi-square variable**.
- Allows the construction of **confidence intervals for** σ^2.
- Degrees of freedom: $n - 1$.

CHI-SQUARE DISTRIBUTION OF SAMPLE AND POPULATION VARIANCES

- If the population distribution is normal then

$$\boxed{Q = \frac{(n-1)s^2}{\sigma^2}}$$

has a chi-square (χ^2) distribution with $n - 1$ degrees of freedom

Newbold et al (2013)

EXERCISE 6.49

6.49 A random sample of size $n = 18$ is obtained from a normally distributed population with a population mean of $\mu = 46$ and a variance of $\sigma^2 = 50$.

a. What is the probability that the sample mean is greater than 50?
b. What is the value of the sample variance such that 5% of the sample variances would be less than this value?
c. What is the value of the sample variance such that 5% of the sample variances would be greater than this value?

Newbold et al (2013)

EXERCISE 6.49 A): SOLUTION

Answer:

Given: population mean $\mu = 46$, population variance $\sigma^2 = 50$ (hence $\sigma = \sqrt{50}$), sample size $n = 18$.

(a) Probability that $\bar{X} > 50$

Standard error:

$$SE = \sqrt{\frac{\sigma^2}{n}} = \sqrt{\frac{50}{18}} = \sqrt{2.777\ldots} = 1.6667.$$

Z-score:

$$z = \frac{50 - 46}{SE} = \frac{4}{1.6667} = 2.40.$$

Probability:

$$P(\bar{X} > 50) = 1 - \Phi(2.40) \approx 0.00820.$$

Standard Normal Distribution Table

EXERCISE 6.49 B): SOLUTION

✅ Answer:

Given: population mean $\mu = 46$, population variance $\sigma^2 = 50$ (hence $\sigma = \sqrt{50}$), sample size $n = 18$.

(b) Value s_L^2 such that 5% of the sample variances are less than s_L^2

For a normal population,

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}.$$

$$\mathbf{Q} = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}$$

We want $P(S^2 < s_L^2) = 0.05$. Let $\chi^2_{0.05,\,17}$ be the 5th percentile of χ^2 with 17 df (here $n-1 = 17$). Then

$$s_L^2 = \frac{\sigma^2}{n-1}\,\chi^2_{0.05,17}.$$

Numerically $\chi^2_{0.05,17} \approx 8.6717602$, so

$$s_L^2 = \frac{50}{17} \times 8.6717602 \approx 25.51.$$

Alternative Solution:

$P(S^2 < a) = 0.05 \Leftrightarrow P\left(\frac{(n-1)S^2}{\sigma^2} < \frac{(n-1)a}{\sigma^2}\right) = 0.05 \Leftrightarrow$

$P\left(Q < \frac{17 \times a}{50}\right) = 0.05$

Then $\frac{17 \times a}{50} = \chi^2_{0.05;17} = \mathbf{8.672}$

$\Leftrightarrow$ **a = 50 × 8.672/ 17 = 25.51**

Chi-Square Distribution Table

EXERCISE 6.49 C): SOLUTION

Answer:

Given: population mean $\mu = 46$, population variance $\sigma^2 = 50$ (hence $\sigma = \sqrt{50}$), sample size $n = 18$.

$$Q = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}$$

(c) Value s_U^2 such that 5% of the sample variances are greater than s_U^2

Now we want $P(S^2 > s_U^2) = 0.05$ so s_U^2 is the 95th percentile. Use $\chi^2_{0.95,17} \approx 27.5871116$:

$$s_U^2 = \frac{50}{17} \times 27.5871116 \approx 81.14.$$

Chi-Square Distribution Table

Alternative Solution:

$P(S^2 > a) = 0.05 \Leftrightarrow P\left(\frac{(n-1)S^2}{\sigma^2} > \frac{(n-1)a}{\sigma^2}\right) = 0.05 \Leftrightarrow$

$P\left(Q < \frac{17 \times a}{50}\right) = 0.95$

Then $\frac{17 \times a}{50} = \chi^2_{0.95;17}$ **= 27.587**

⇔ a = 50 × 27.587/ 17 = 81.14

LECTURE 9: ESTIMATORS AND CONFIDENCE INTERVALS

ESTIMATOR VS. ESTIMATE

- An estimator of a population parameter is
  - a random variable that depends on sample information . . .
  - whose value provides an approximation to this unknown parameter
- A specific value of that random variable is called an estimate

Newbold et al (2013)

Example:

Let $\bar{X}$ be the **estimator** of the population mean weight μ.

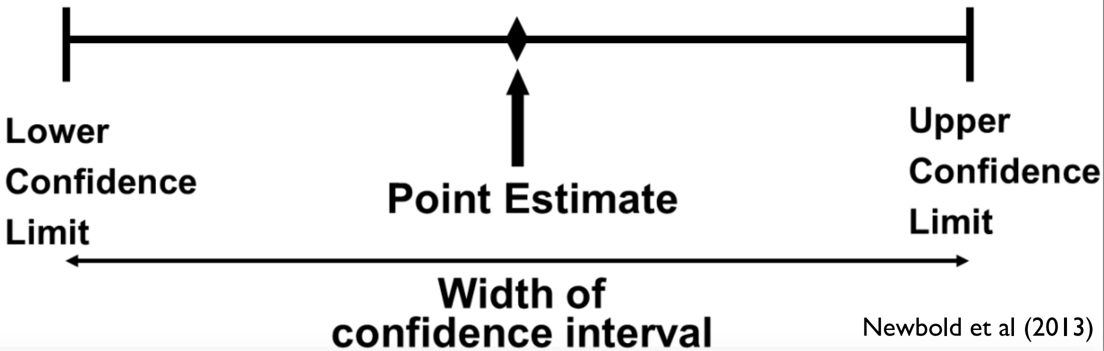
If the sample mean is $\bar{x} = 70\ \mathrm{kg}$, then **70 kg is the estimate** of the average weight.

Summary:

- An **estimator** is a **random variable (or statistic)** used to estimate an unknown population parameter.
- An **estimate** is the **numerical value** that the estimator takes for a given sample.

POINT ESTIMATE VS. CONFIDENCE INTERVAL

- A point estimate is a single number,
- a confidence interval provides additional information about variability



Newbold et al (2013)

Example: Point Estimate and Confidence Interval

We took a random sample of 100 adults and found an **average weight of 70 kg**.

- **Point estimate:** 70 kg
- **95% confidence interval: CI = (68.4, 71.6)**

➡️ We are 95% confident that the true average weight of all adults lies **between 68.4 kg and 71.6 kg**.

POINT ESTIMATES: EXAMPLE

We can estimate a Population Parameter …		with a Sample Statistic (a Point Estimate)
Mean	μ	$\bar{x}$
Proportion	P	$\hat{p}$

Population Mean

Sample Mean

Population Proportion

Sample Proportion

Newbold et al (2013)

PROPERTIES OF ESTIMATORS: UNBIASED ESTIMATOR

- A point estimator $\hat{\theta}$ is said to be an unbiased estimator of the parameter θ if its expected value is equal to that parameter:

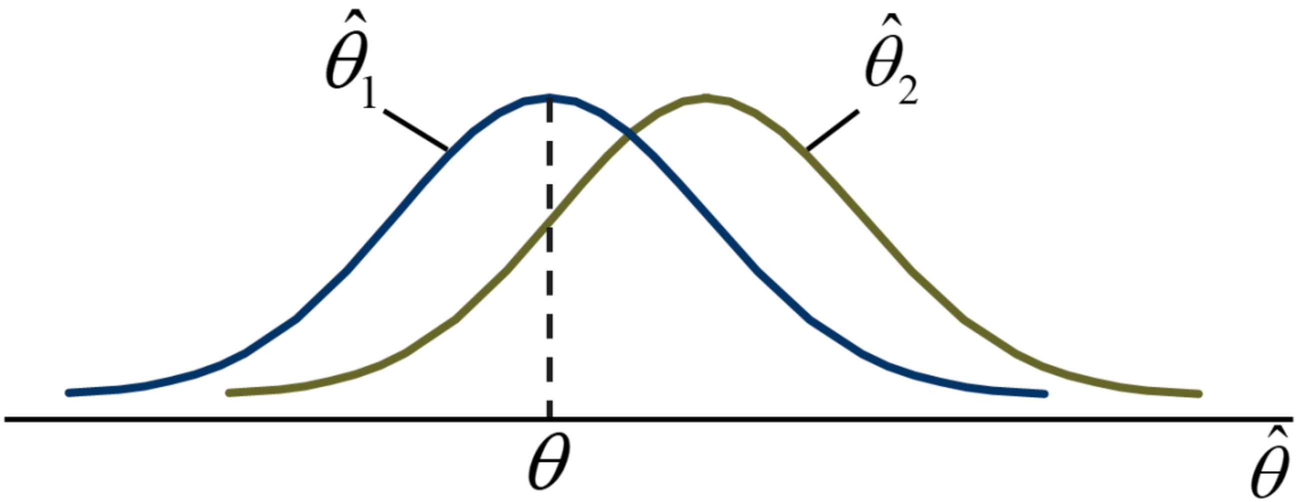
$$\boxed{E(\hat{\theta}) = \theta}$$

- Examples:
  - The sample mean $\bar{x}$ is an unbiased estimator of μ
  - The sample variance s^2 is an unbiased estimator σ^2
  - The sample proportion $\hat{p}$ is an unbiased estimator of P

Newbold et al (2013)

PROPERTIES OF ESTIMATORS: UNBIASED ESTIMATOR

- $\hat{\theta}_1$ is an unbiased estimator, $\hat{\theta}_2$ is biased:



Newbold et al (2013)

PROPERTIES OF ESTIMATORS: BIAS

- Let $\hat{\theta}$ be an estimator of $\boldsymbol{\theta}$
- The bias in $\hat{\theta}$ is defined as the difference between its mean and $\boldsymbol{\theta}$

$$\boxed{\text{Bias}(\hat{\theta}) = E(\hat{\theta}) - \theta}$$

- The bias of an unbiased estimator is 0

Newbold et al (2013)

PROPERTIES OF ESTIMATORS: MOST EFFICIENT ESTIMATOR

- Suppose there are several unbiased estimators of $\boldsymbol{\theta}$
- The most efficient estimator or the minimum variance unbiased estimator of $\boldsymbol{\theta}$ is the unbiased estimator with the smallest variance
- Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two unbiased estimators $\boldsymbol{\theta}$, based on the same number of sample observations. Then,
  - $\boxed{\hat{\theta}_1 \text{ is said to be more efficient than } \hat{\theta}_2 \text{ if } Var(\hat{\theta}_1) < Var(\hat{\theta}_2)}$
  - The relative efficiency of $\hat{\theta}_1$ with respect to $\hat{\theta}_2$ is the ratio of their variances:

$$\boxed{\text{Relative Efficiency} = \frac{Var(\hat{\theta}_2)}{Var(\hat{\theta}_1)}}$$

Newbold et al (2013)

CONFIDENCE INTERVAL ESTIMATES

- How much uncertainty is associated with a point estimate of a population parameter?
- An interval estimate provides more information about a population characteristic than does a point estimate
- Such interval estimates are called confidence interval estimates

Note:
- A **confidence interval** is the concept, while a **confidence interval estimate** is the specific interval calculated from a sample.

Example: For a sample of 100 adults with mean weight 70 kg, the 95% confidence interval estimate is **CI = (68.4, 71.6)**.

CONFIDENCE INTERVAL ESTIMATE

- An interval gives a range of values:
  - Takes into consideration variation in sample statistics from sample to sample
  - Based on observation from one sample
  - Gives information about closeness to unknown population parameters
  - Stated in terms of level of confidence
    - Can never be 100% confident

Newbold et al (2013)

CONFIDENCE INTERVAL AND CONFIDENCE LEVEL

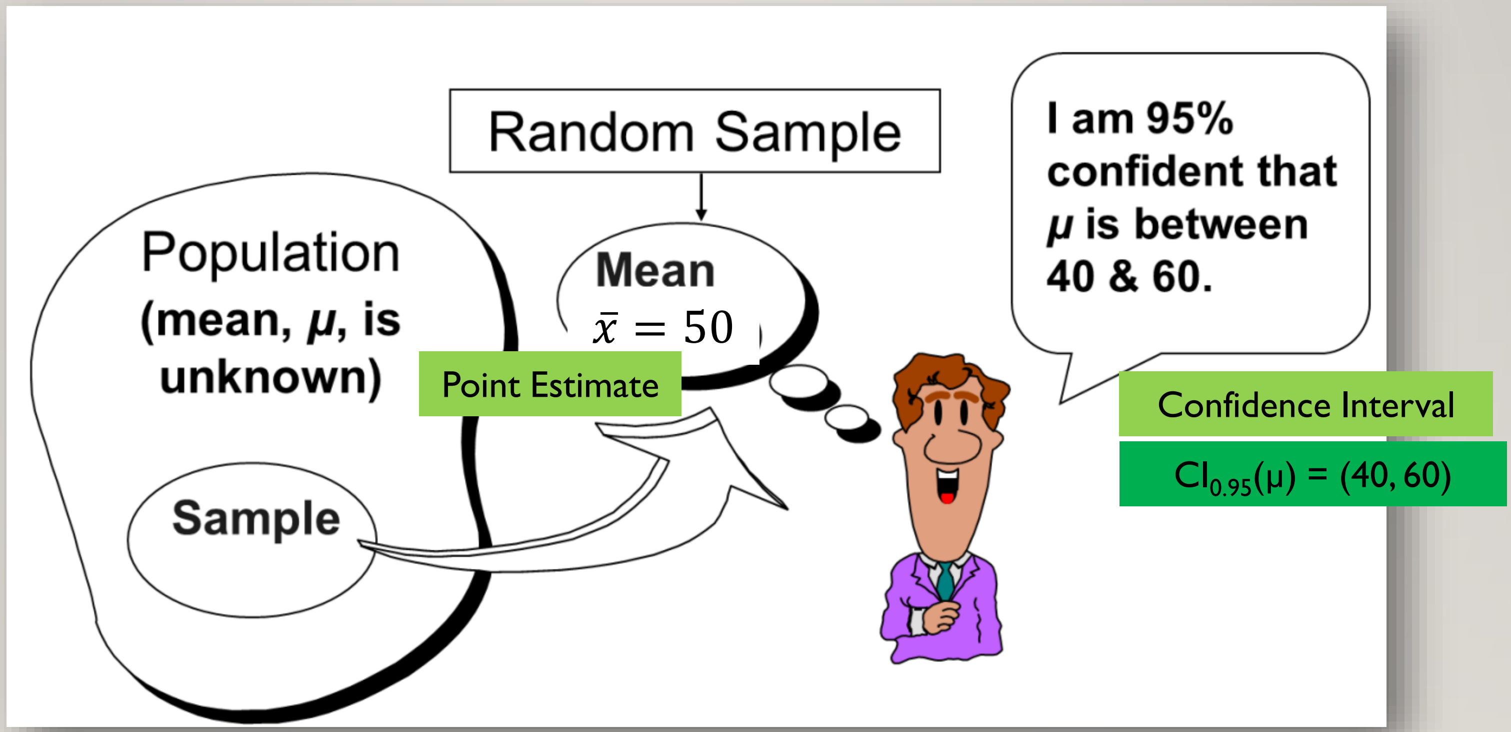
- If $P(a < \theta < b) = 1 - \alpha$ then the interval from *a* to *b* is called a $100(1-\alpha)\%$ confidence interval of θ.
- The quantity $100(1-\alpha)\%$ is called the confidence level of the interval
  - α is between 0 and 1
  - In repeated samples of the population, the true value of the parameter θ would be contained in $100(1-\alpha)\%$ of intervals calculated this way.
  - The confidence interval calculated in this manner is written as $a < \theta < b$ with $100(1-\alpha)\%$ confidence

Newbold et al (2013)

Note:

- The **confidence level** of an interval indicates how certain we are that the interval contains the true population parameter. It is equal to $1-\alpha$ (e.g., 95% confidence level corresponds to $\alpha = 0.05$).
- The **significance level** (α) is the probability of making a **Type I error**, or the probability that the interval does **not** contain the true parameter.

ESTIMATION PROCESS



Newbold et al (2013)

ESTIMATION PROCESS

- Suppose confidence level = 95%
- Also written $(1-\alpha) = 0.95$
- A relative frequency interpretation:
  - From repeated samples, 95% of all the confidence intervals that can be constructed of size *n* will contain the unknown true parameter
- A specific interval either will contain or will not contain the true parameter

Newbold et al (2013)

GENERAL FORMULA FOR CONFIDENCE INTERVAL

- The general form for all confidence intervals is:

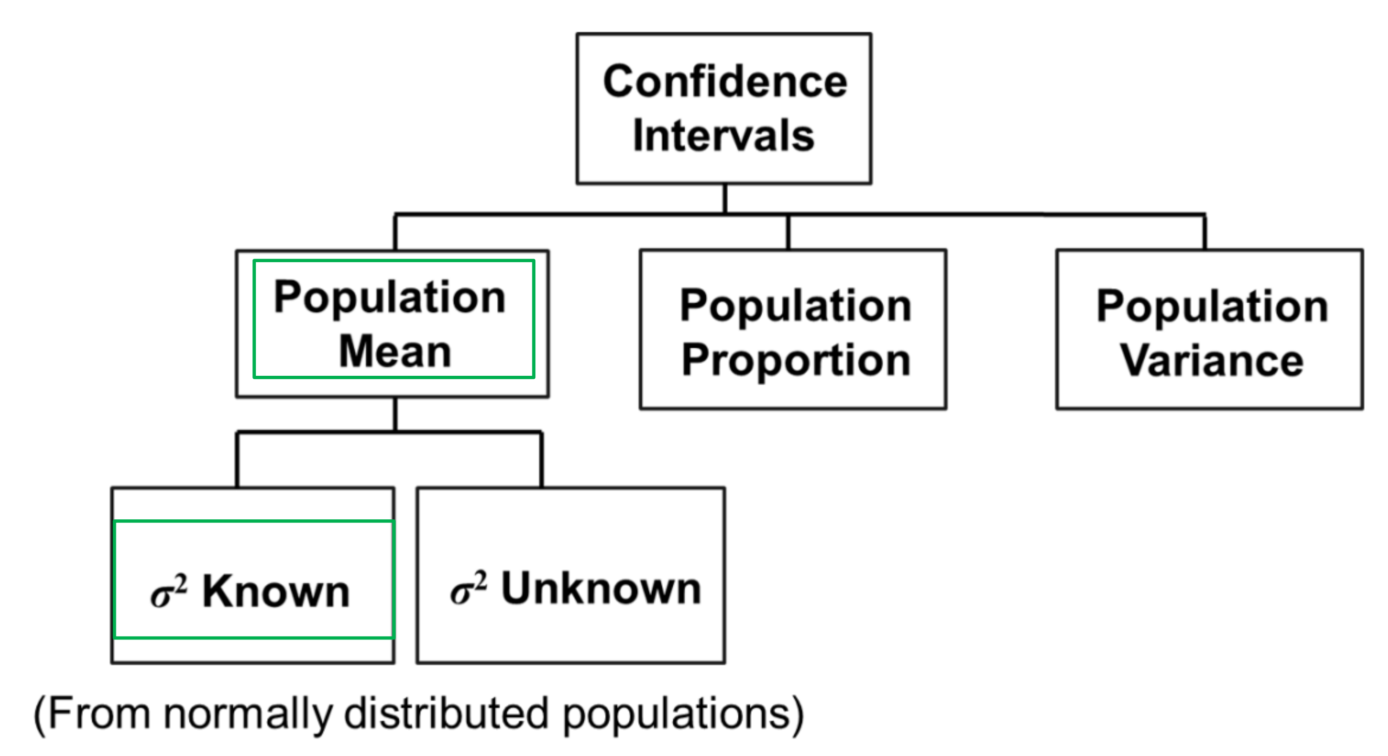
$$\boxed{\hat{\theta} \pm ME}$$

Point Estimate ± Margin of Error

- The value of the margin of error depends on the desired level of confidence

Newbold et al (2013)

CONFIDENCE INTERVALS WE WILL CONSIDER



Newbold et al (2013)

CONFIDENCE INTERVAL ESTIMATE FOR THE MEAN (σ^2 KNOWN)

- Assumptions
  - Population variance σ^2 is known
  - Population is normally distributed
  - If population is not normal, use large sample
- Confidence interval estimate:

n ≥ 25

$$\boxed{\bar{x} \pm z_{1-\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}}}$$

($z_{1-\frac{\alpha}{2}}$ is the normal distribution value for a probability of $1-\frac{\alpha}{2}$ in each tail)

Quantile of the standard normal distribution corresponding to probability 1-α/2

Newbold et al (2013)

CI ESTIMATE FOR THE MEAN (σ^2 KNOWN): CONFIDENCE LIMITS

- The confidence interval is

$$\bar{x} \pm z_{1-\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}}$$

- The endpoints of the interval are

$$\text{UCL} = \bar{x} + z_{1-\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}}$$ Upper confidence limit

$$\text{LCL} = \bar{x} - z_{1-\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}}$$ Lower confidence limit

Newbold et al (2013)

CI ESTIMATE FOR THE MEAN (σ^2 KNOWN): MARGIN OF ERROR

- The confidence interval,

$$\bar{x} \pm z_{1-\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}}$$

- Can also be written as $\bar{x} \pm ME$
  where *ME* is called the margin of error

$$\boxed{ME = z_{1-\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}}}$$

- The interval width, *w*, is equal to twice the margin of error

Note:
The **margin of error** is half the width of the confidence interval. Equivalently, the **width of the interval** is twice the margin of error.

- Margin of error = (interval width) / 2
- Interval width = 2 × (margin of error)

Newbold et al (2013)

Width of Confidence Interval for the Mean (σ known)

$$\boxed{\text{Width} = 2 \cdot z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}}$$

Where:

- $z_{1-\alpha/2}$ = critical value (quantile) from the standard normal distribution
- σ = population standard deviation
- n = sample size

CI ESTIMATE FOR THE MEAN (σ^2 KNOWN): REDUCING THE MARGIN OF ERROR

$$ME = z_{1-\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}}$$

The margin of error can be reduced if

- the population standard deviation can be reduced $(\sigma \downarrow)$
- The sample size is increased $(n \uparrow)$
- The confidence level is decreased, $(1-\alpha)\downarrow$

Note:
The **margin of error** decreases when the **standard deviation** is smaller or the **sample size** is larger, and increases when the **critical value** (confidence level) is higher.

Newbold et al (2013)

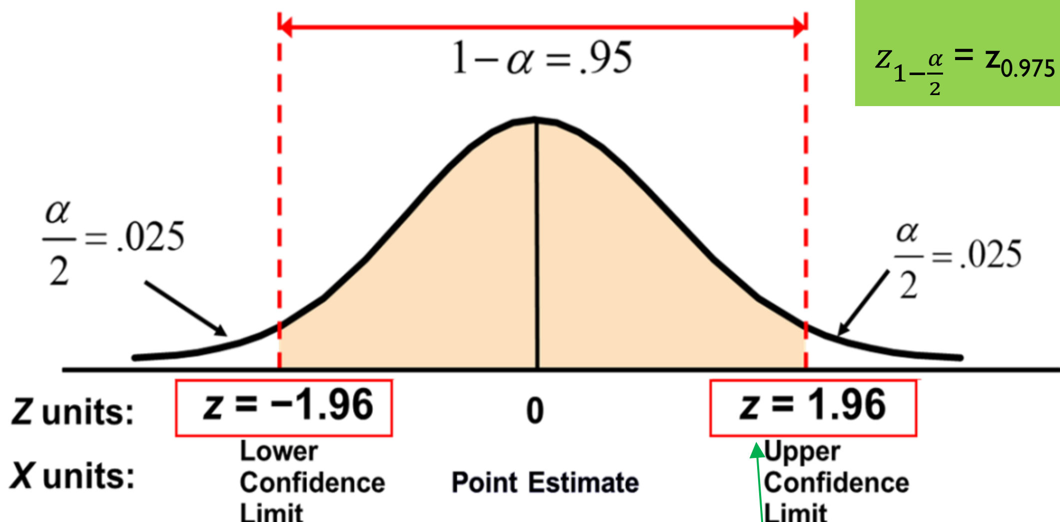
Z-QUANTILES FOR A 95% CONFIDENCE INTERVAL

$1-\alpha = 0.95 \Leftrightarrow \alpha = 0.05$

$1- \alpha/2 = 0.975$

$z_{1-\frac{\alpha}{2}} = z_{0.975} = 1.96$

- Consider a 95% confidence interval:



- Find $Z_{0.975} = 1.96$ from the standard normal distribution table

Newbold et al (2013)

$z_{0.975} = 1.96$ is the quantile of the standard normal distribution corresponding to a cumulative probability of 0.975.

COMMON LEVELS OF CONFIDENCE

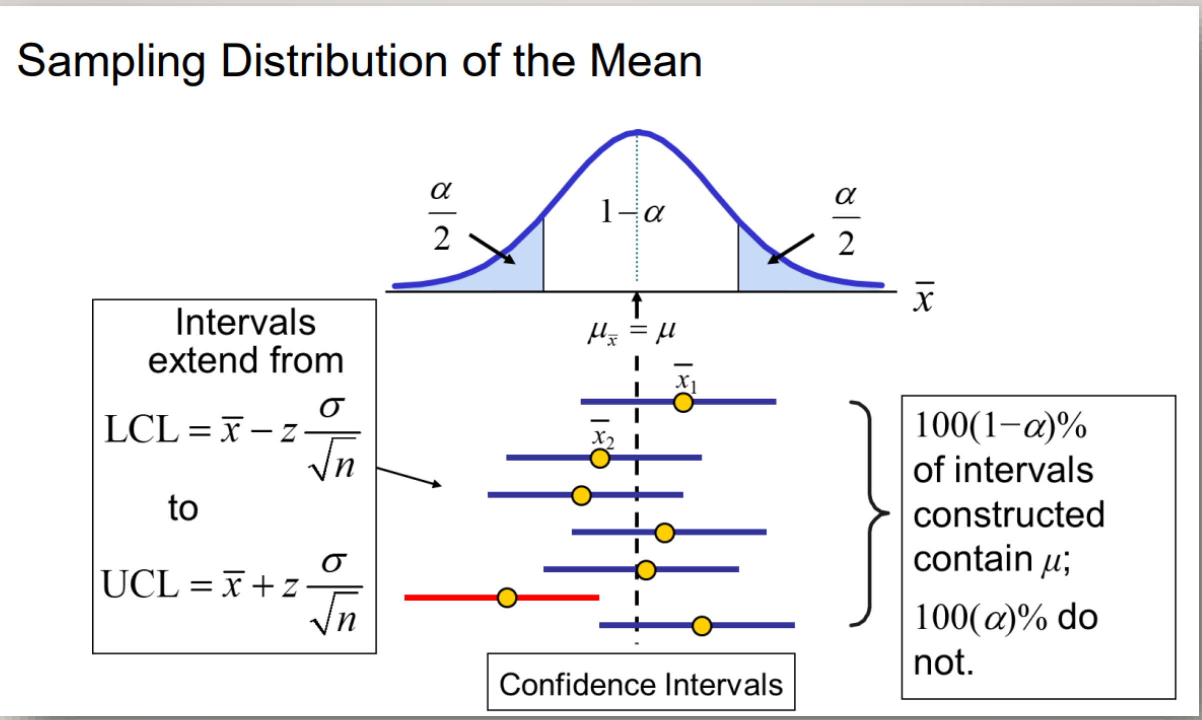
- Commonly used confidence levels are 90%, 95%, 98%, and 99%

Confidence Level	Confidence Coefficient, $1-\alpha$	Quantiles $z_{1-\frac{\alpha}{2}}$
80%	.80	1.28
90%	.90	1.645
95%	.95	1.96
98%	.98	2.33
99%	.99	2.58
99.8%	.998	3.08
99.9%	.999	3.27

Newbold et al (2013)

Note:
•The most commonly used confidence levels are **90%, 95%, and 99%**.

INTERVALS AND LEVEL OF CONFIDENCE



Newbold et al (2013)

Note:
•The **confidence level** is the probability that the interval, not the parameter, **contains the true population value**.

CI ESTIMATE FOR THE μ (σ^2 KNOWN): EXAMPLE

- A sample of 11 circuits from a large normal population has a mean resistance of 2.20 ohms. We know from past testing that the population standard deviation is 0.35 ohms.
- Determine a 95% confidence interval for the true mean resistance of the population.

Newbold et al (2013)

CI ESTIMATE FOR THE μ (σ^2 KNOWN): EXAMPLE

- A sample of 11 circuits from a large normal population has a mean resistance of 2.20 ohms. We know from past testing that the population standard deviation is .35 ohms.
- Solution:

$$\bar{x} \pm z_{1-\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}}$$

$$= 2.20 \pm 1.96 \left(\frac{.35}{\sqrt{11}} \right)$$

$$= 2.20 \pm .2068$$

$$1.9932 < \mu < 2.4068$$

n = 11 (sample size)
$\bar{x} = 2.20$ (sample mean)
σ = 0.35 (population standard deviation)
1 – α = 0.95 (confidence level), then α = 0.05 and $z_{1-\alpha/2} = z_{0.975} = 1.96$ (see previous slide)

Newbold et al (2013)

$CI_{0.95}(\mu) = (1.9932; 2.4068)$

- We are 95% confident that the true mean resistance is between 1.9932 and 2.4068 ohms
- Although the true mean may or may not be in this interval, 95% of intervals formed in this manner will contain the true mean

EXERCISE 7.14

7.14 It is known that the standard deviation in the volumes of 20-ounce (591-millliliter) bottles of natural spring water bottled by a particular company is 5 millliliters. One hundred bottles are randomly sampled and measured.

a. Calculate the standard error of the mean.
b. Find the margin of error of a 90% confidence interval estimate for the population mean volume.
c. Calculate the width for a 98% confidence interval for the population mean volume.

Newbold et al (2013)

EXERCISE 7.14 A): SOLUTION

Answer:

Given: $\sigma = 5$ ml, $n = 100$.

So the standard error is $\mathrm{SE} = \sigma/\sqrt{n} = 5/\sqrt{100} = 5/10 = 0.5$ ml.

(a) Standard error of the mean

$$\boxed{\mathrm{SE} = 0.5\ \mathrm{ml}}$$

EXERCISE 7.14 B): SOLUTION

Answer:

Given: $\sigma = 5$ ml, $n = 100$.

So the standard error is $\mathrm{SE} = \sigma/\sqrt{n} = 5/\sqrt{100} = 5/10 = 0.5$ ml.

$1\text{-}\alpha = 0.90 \Leftrightarrow \alpha = 0.10$

$1\text{-}\ \alpha/2 = 0.95$

$z_{1-\frac{\alpha}{2}} = z_{0.95} = 1.645$

(b) Margin of error for a 90% confidence interval

For 90% two-sided CI, $Z_{0.95} = 1.645$

$$\mathrm{ME} = Z_{0.95} \cdot \mathrm{SE} = 1.645 \times 0.5 = 0.8225 \text{ ml}$$

Rounded sensibly:

$$\boxed{\mathrm{ME}_{90\%} \approx 0.823 \text{ ml}}$$

EXERCISE 7.14 C): SOLUTION

Answer:

Given: $\sigma = 5$ ml, $n = 100$.

So the standard error is $\mathrm{SE} = \sigma/\sqrt{n} = 5/\sqrt{100} = 5/10 = 0.5$ ml.

(c) Width of a 98% confidence interval

For 98% two-sided CI, $Z_{0.99} = 2.3263$

1-α = 0.98 ⇔ α = 0.02

1- α/2 = 0.99

$z_{1-\frac{\alpha}{2}} = z_{0.98} = 2.3263$

Margin of error:

$$\mathrm{ME}_{98\%} = 2.3263 \times 0.5 = 1.16315 \text{ ml}$$

Width of the interval (two-sided) = $2\times$ ME:

$$\mathrm{Width}_{98\%} = 2 \times 1.16315 = 2.3263 \text{ ml}$$

Rounded:

$$\boxed{\mathrm{Width}_{98\%} \approx 2.326 \text{ ml}}$$

EXERCISE 7.15

7.15 A college admissions officer for an MBA program has determined that historically applicants have undergraduate grade point averages that are normally distributed with standard deviation 0.45. From a random sample of 25 applications from the current year, the sample mean grade point average is 2.90.

a. Find a 95% confidence interval for the population mean.
b. Based on these sample results, a statistician computes for the population mean a confidence interval extending from 2.81 to 2.99. Find the confidence level associated with this interval.

Newbold et al (2013)

EXERCISE 7.15 A): SOLUTION

Answer:

Problem data:

$\sigma = 0.45,\ n = 25,\ \bar{x} = 2.90.$

Because the population standard deviation σ is known and the underlying distribution is normal, use the normal (z) distribution.

n = 25 (sample size)
$\bar{x} = 2.90$ (sample mean)
σ = 0.45 (population standard deviation)
$1 - \alpha = 0.95$ (confidence level), then
$z_{0.975} = 1.96$ (see previous slide)

(a) 95% confidence interval for μ

Standard error:

$$SE = \frac{\sigma}{\sqrt{n}} = \frac{0.45}{\sqrt{25}} = \frac{0.45}{5} = 0.09.$$

Critical z for 95% (two-sided): $Z_{0.975} = 1.96$

Margin of error:

$$ME = Z_{0.975} \times SE = 1.96 \times 0.09 = 0.1764.$$

Confidence interval:

$$\bar{x} \pm ME = 2.90 \pm 0.1764 \Rightarrow$$ $CI_{95\%}(\mu) = (2.7236, 3.0764)$

EXERCISE 7.15 B): SOLUTION

Answer:

Problem data:

$\sigma = 0.45,\ n = 25,\ \bar{x} = 2.90.$

Because the population standard deviation σ is known and the underlying distribution is normal, use the normal (z) distribution.

(b) Confidence level for the interval

$CI_{1-\alpha}(\mu) = (2.81, 2.99)$

$$ME = z_{1-\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}} \Leftrightarrow$$

$$(2.99 - 2.81)/2 = z_{1-\frac{\alpha}{2}} \times 0.45/5 \Leftrightarrow$$

$$z_{1-\frac{\alpha}{2}} = 1, \text{ then } 1 - \alpha/2 = 0.8413 \Leftrightarrow \alpha = 0.3174$$

$$1 - \alpha = 0.6826$$

Standard Normal Distribution Table

n = 25 (sample size)
$\bar{x} = 2.90$ (sample mean)
$\sigma = 0.45$ (population standard deviation)
$1 - \alpha = 0.95$ (confidence level), then
$z_{0.975} = 1.96$ (see previous slide)

Thus, the confidence level is about 68.27%

Note:

- The **margin of error** is **half the width** of the confidence interval.

HOMEWORK OF LECTURE 9: QUESTIONS

EXERCISE 6.32

6.32 A record store owner finds that 20% of customers entering her store make a purchase. One morning 180 people, who can be regarded as a random sample of all customers, enter the store.

a. What is the mean of the distribution of the sample proportion of customers making a purchase?
b. What is the variance of the sample proportion?
c. What is the standard error of the sample proportion?
d. What is the probability that the sample proportion is less than 0.15?

Newbold et al (2013)

EXERCISE 6.48

6.48 A random sample of size $n = 25$ is obtained from a normally distributed population with a population mean of $\mu = 198$ and a variance of $\sigma^2 = 100$.

a. What is the probability that the sample mean is greater than 200?

b. What is the value of the sample variance such that 5% of the sample variances would be less than this value?

c. What is the value of the sample variance such that 5% of the sample variances would be greater than this value?

Newbold et al (2013)

EXERCISE 7.13

7.13 A personnel manager has found that historically the scores on aptitude tests given to applicants for entry-level positions follow a normal distribution with a standard deviation of 32.4 points. A random sample of nine test scores from the current group of applicants had a mean score of 187.9 points.

a. Find an 80% confidence interval for the population mean score of the current group of applicants.
b. Based on these sample results, a statistician found for the population mean a confidence interval extending from 165.8 to 210.0 points. Find the confidence level of this interval.

Newbold et al (2013)

THANKS!

Questions?