STATISTICS II

Bachelor's degrees in Economics, Finance and Management

2nd year/2nd Semester
2025/2026

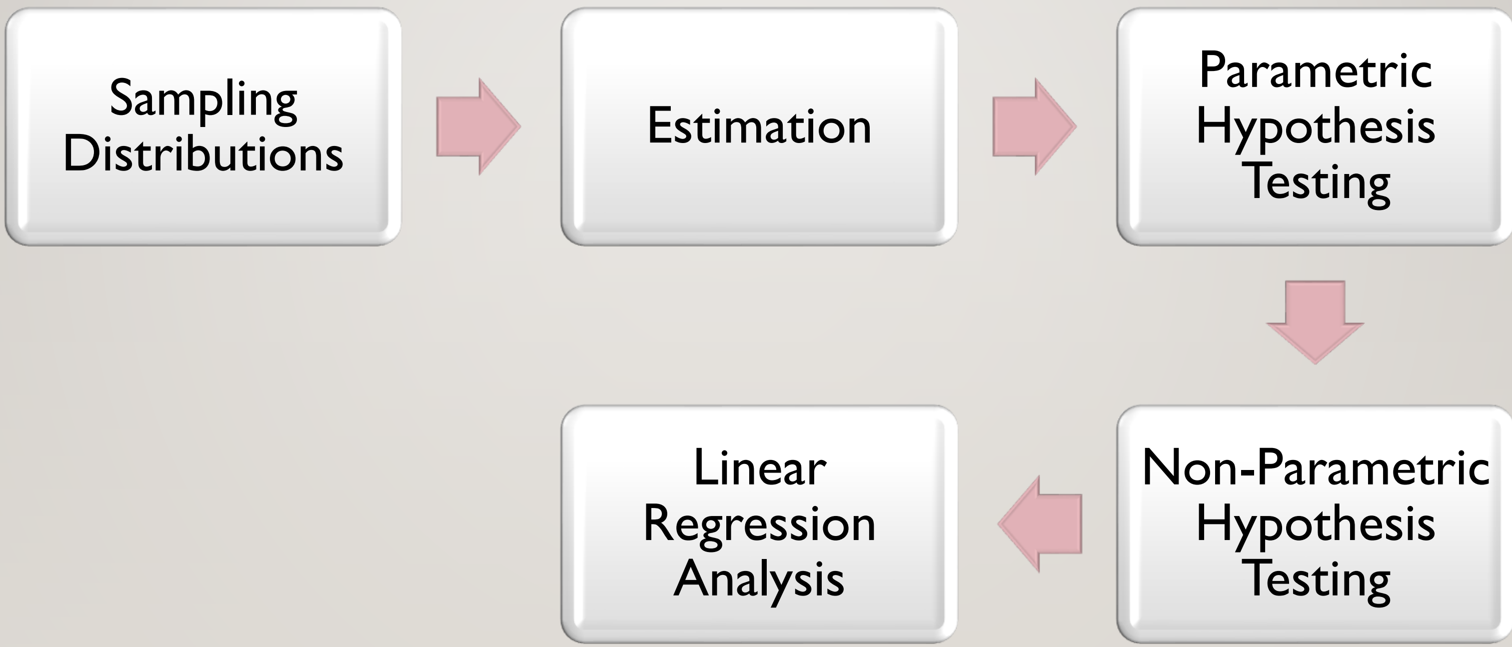
CONTACT

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


Sampling Distributions
Estimation
Parametric Hypothesis Testing
Non-Parametric Hypothesis Testing
Linear Regression Analysis

LECTURE 2 HOMEWORK: QUESTIONS AND SOLUTIONS

EXERCISE 6.18

6.18 An industrial process produces batches of a chemical whose impurity levels follow a normal distribution with standard deviation 1.6 grams per 100 grams of chemical. A random sample of 100 batches is selected in order to estimate the population mean impurity level.

a. The probability is 0.05 that the sample mean impurity level exceeds the population mean by how much?
b. The probability is 0.10 that the sample mean impurity level is below the population mean by how much?
c. The probability is 0.15 that the sample mean impurity level differs from the population mean by how much?

Newbold et al (2013)

EXERCISE 6.18 A): SOLUTION

Answer:

Given: population standard deviation σ = 1.6 g per 100 g, sample size n = 100.

Standard error:

$$\sigma_{\bar{X}} = \frac{1.6}{\sqrt{100}} = \frac{1.6}{10} = 0.16$$

X ~ Normal(μ, 0.16²)

(a) Probability is 0.05 that $\bar{X}$ exceeds μ by how much?

We need $P(\bar{X} - \mu > E) = 0.05.$

$$P\left(\frac{\bar{X}-\mu}{0.16} > \frac{E}{0.16}\right) = 0.05 \Leftrightarrow P(Z > E/0.16) = 0.05$$

Then, $E/0.16 = z_{0.95} = 1.645 \Leftrightarrow E = 0.16 \times 1.645 = 0.2632$

EXERCISE 6.18 B): SOLUTION

Answer:

Given: population standard deviation σ = 1.6 g per 100 g, sample size n = 100.

Standard error:

$$\sigma_{\bar{X}} = \frac{1.6}{\sqrt{100}} = \frac{1.6}{10} = 0.16$$

***X* ~ Normal(μ, 0.16²)**

(b) Probability is 0.10 that $\bar{X}$ is below μ by how much?

We need $P(\bar{X} - \mu < E) = 0.10$

$$P\left(\frac{\bar{X}-\mu}{0.16} < \frac{E}{0.16}\right) = 0.10 \Leftrightarrow P(Z < E/0.16) = 0.10$$

Then, $E/0.16 = z_{0.10} = -1.2816 \Leftrightarrow E = 0.16 \times -1.2816 = -0.2051$

EXERCISE 6.18 C): SOLUTION

Answer:

Given: population standard deviation σ = 1.6 g per 100 g, sample size n = 100.

Standard error:

$$\sigma_{\bar{X}} = \frac{1.6}{\sqrt{100}} = \frac{1.6}{10} = 0.16$$

***X* ~ Normal(μ, 0.16²)**

(c) Probability is 0.15 that $\bar{X}$ differs from μ by how much?

"Differs" means two-sided:

$$P(|\bar{X} - \mu| > E) = 0.15$$

$P\left(\frac{-E}{0.16} < \frac{\bar{X}-\mu}{0.16} < \frac{E}{0.16}\right) = 1-0.15 \Leftrightarrow P\left(\frac{-E}{0.16} < Z < \frac{E}{0.16}\right) = 0.85 \Leftrightarrow$

$\Phi\,(E/0.16) - \Phi\,(-E/0.16) = 0.85 \Leftrightarrow 2\Phi\,(E/0.16) - 1 = 0.85 \Leftrightarrow$

$\Phi\,(E/0.16) = 0.925 \Leftrightarrow E/0.16 = z_{0.925} = 1.44 \Leftrightarrow 0.16 \times 1.44 = 0.2304$

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, σ²/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:

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

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

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 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:

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

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

LECTURE 3: SAMPLING DISTRIBUTIONS OF SAMPLE PROPORTIONS

SAMPLING DISTRIBUTION OF THE SAMPLE PROPORTION

P = the proportion of the population having some characteristic

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

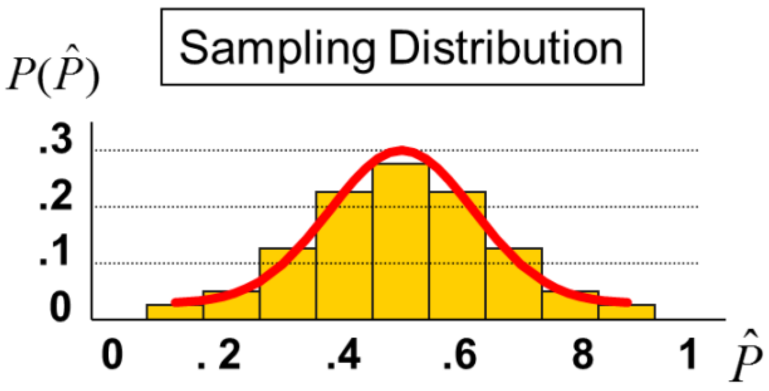
$$\hat{p} = \frac{X}{n} = \frac{\text{number of items in the sample having the characteristic of interest}}{\text{sample size}}$$

- $0 \leq \hat{p} \leq 1$
- $\hat{p}$ 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 $\sigma_{\hat{p}} = \sqrt{\frac{P(1-P)}{n}}$

(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 $nP\left(1-P\right) > 5$

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**

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

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:

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

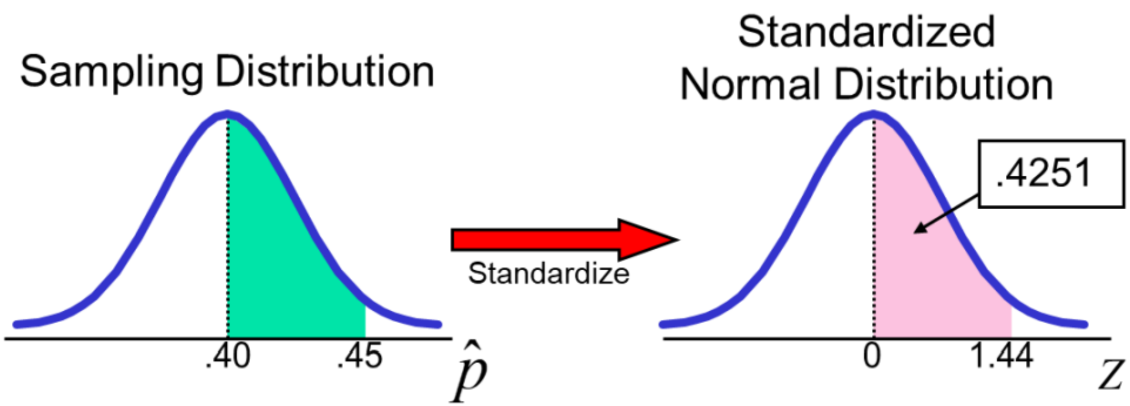
Newbold et al (2013)

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:

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

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

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

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

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

(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

LECTURE 4: SAMPLING DISTRIBUTIONS OF SAMPLE VARIANCES

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

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

- 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

$$E(s^2) = \sigma^2$$

- If the population distribution is normal, then

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

Newbold et al (2013)

SAMPLE VARIANCE AND CHI-SQUARE DISTRIBUTION

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

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

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

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

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

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

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

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

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,

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

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 $\frac{17 \times a}{50} = \chi^2_{0.05;17}$ **= 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 4: 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.

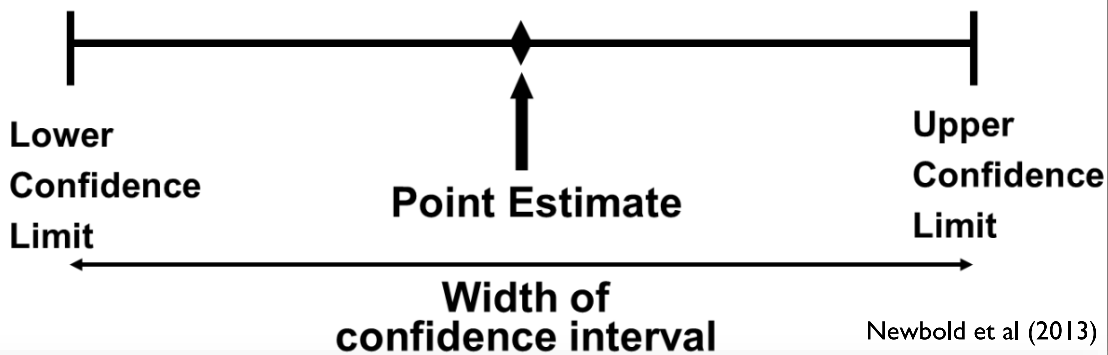
If the sample mean is $\bar{x} = 70 \text{ kg}$, then **70 kg is the estimate** of the average weight.

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 $\boldsymbol{\theta}$ if its expected value is equal to that parameter:

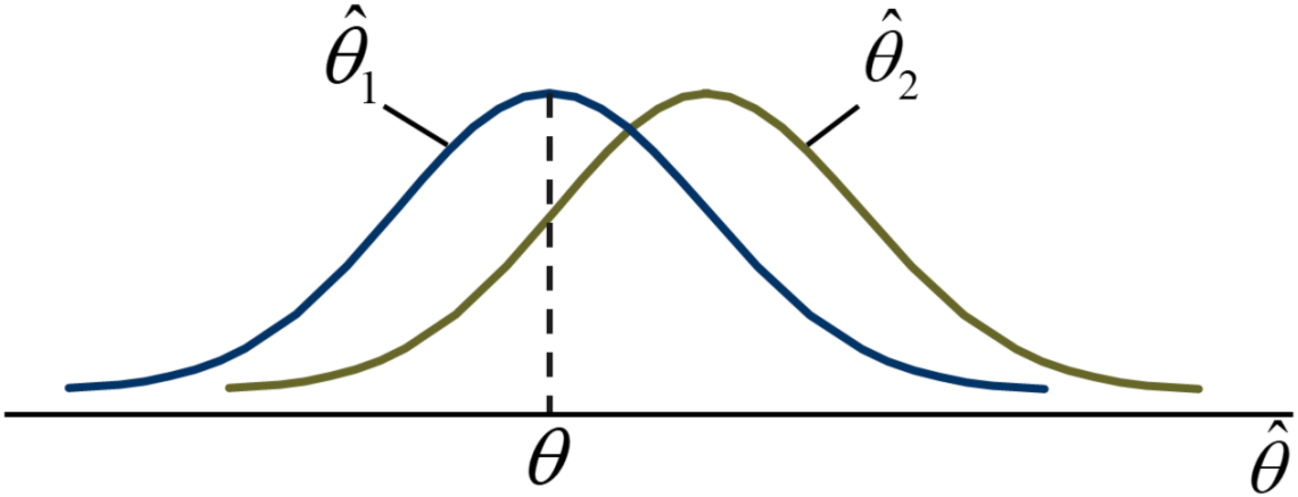
$$E(\hat{\theta}) = \theta$$

- Examples:
  - The sample mean $\overline{x}$ 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 $\boldsymbol{\theta}$

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

- 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,
  - $\hat{\theta}_1$ is said to be more efficient than $\hat{\theta}_2$ 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:

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

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

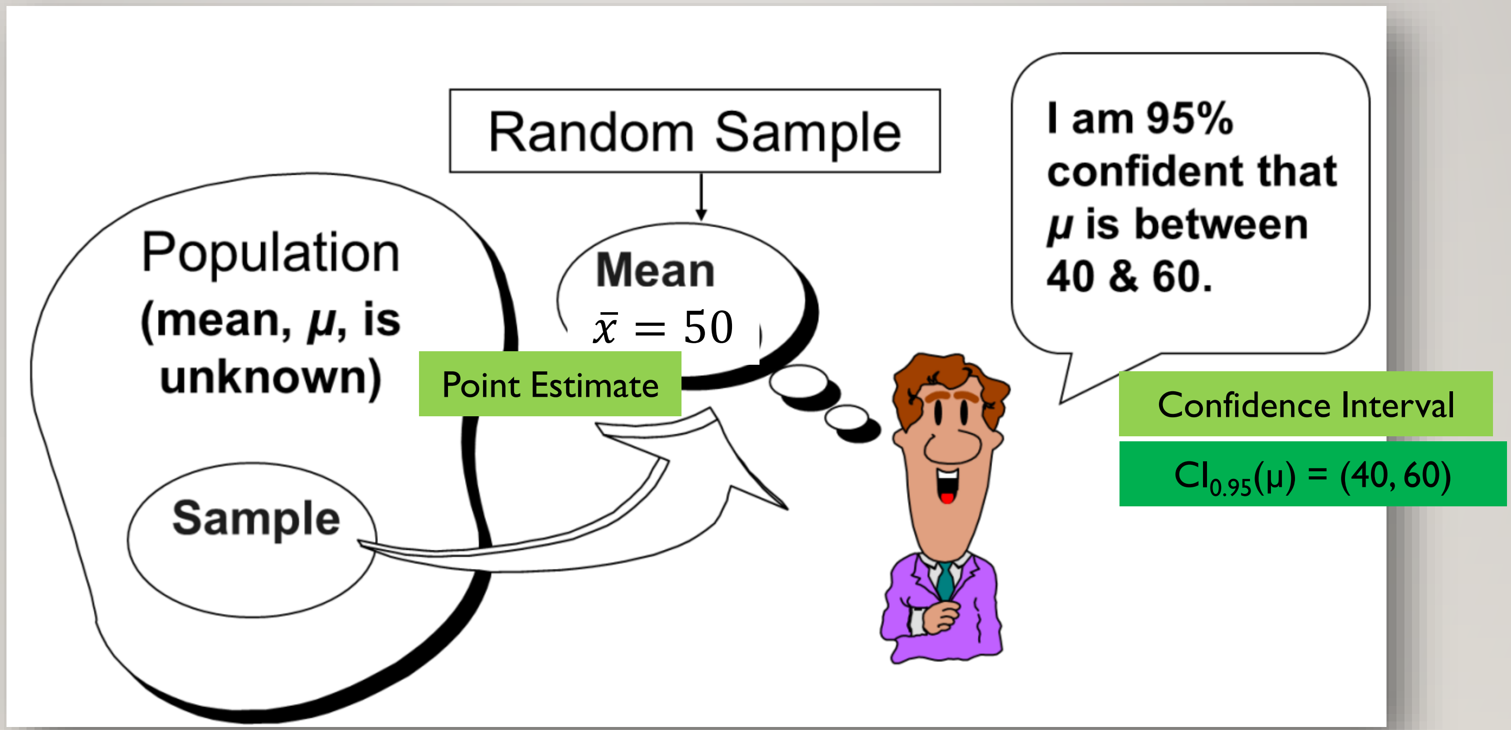
- If $P(a < \theta < b) = 1 - \alpha$ then the interval from *a* to *b* is called a $100(1-\alpha)\%$ confidence interval of θ.
- The quantity $100(1-\alpha)\%$ is called the confidence level of the interval
  - α is between 0 and 1
  - In repeated samples of the population, the true value of the parameter 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−α (e.g., 95% confidence level corresponds to α = 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:

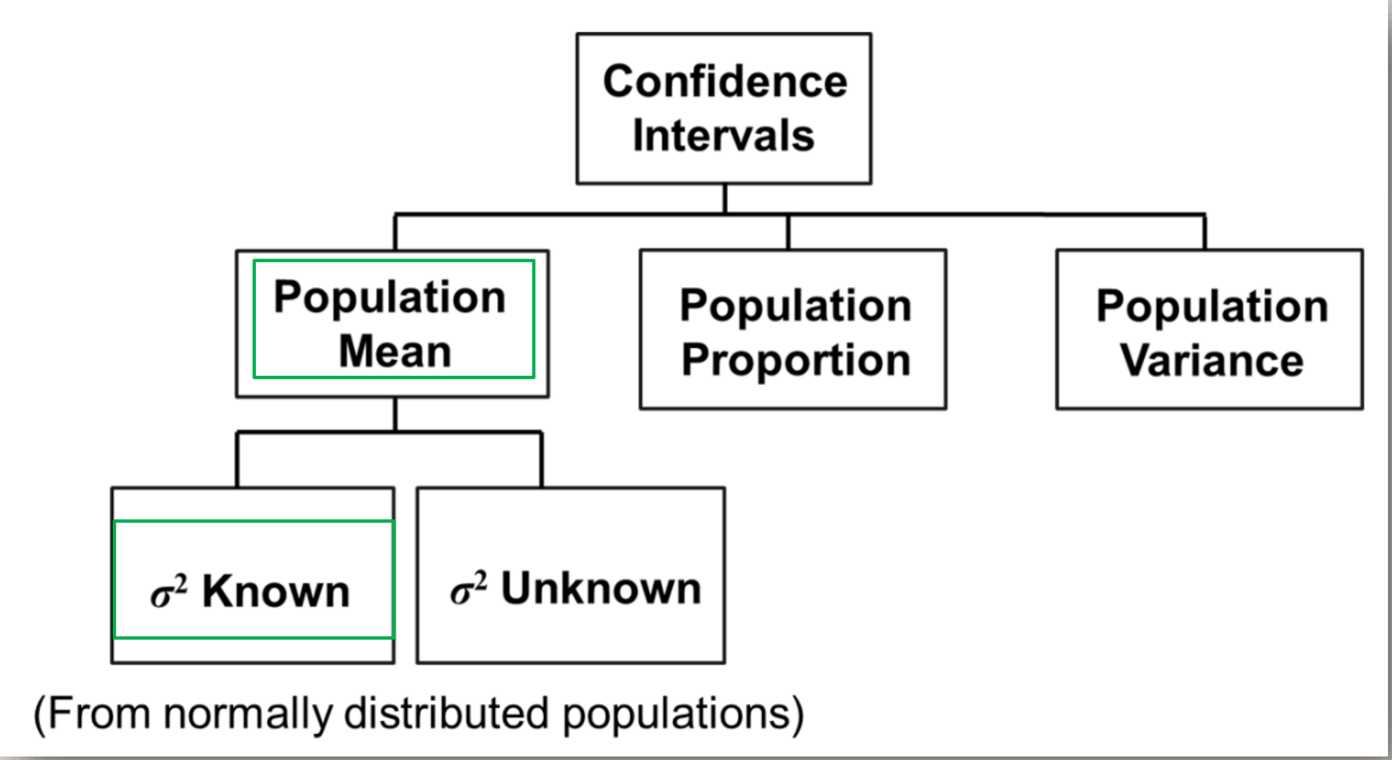
$$\hat{\theta} \pm ME$$

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)

HOMEWORK OF LECTURE 4: QUESTIONS

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)

THANKS!

Questions?