

CHAPTER 2

SAMPLING DISTRIBUTION & CONFIDENCE INTERVAL

Expected Outcomes

- ✓ Able to identify the sampling distribution for sample means and sample proportions.
- ✓ Able to find the confidence interval for the population mean.
- ✓ Able to find the confidence interval for the difference between two populations mean.
- ✓ Able to find the confidence interval for the paired data.
- ✓ Able to find the confidence interval for a population proportion.
- ✓ Able to find the confidence interval for the difference between two population proportions.
- ✓ Able to find the confidence interval for population variance and standard deviation
- ✓ Able to find the confidence interval for two population variances and standard deviations

CONTENT

- 2.1 Sampling Distributions**
- 2.2 Estimate, Estimation, Estimator**
- 2.3 Confidence Interval for the Population Mean**
- 2.4 Confidence Interval for the Difference between Two Population Means**
- 2.5 Confidence Intervals with Paired Data**
- 2.6 Confidence Interval for the Population Proportion**
- 2.7 Confidence Interval for the Difference between Two Population Proportion**
- 2.8 Confidence Interval for Population Variance and Population Standard Deviation**
- 2.9 Confidence Interval on the ratio of Two Population Variances and Standard Deviations**

2.1 SAMPLING DISTRIBUTION

- ✓ A sampling distribution is the *probability distribution*, under **repeated sampling of the population, of a given statistic** (a numerical quantity calculated from the *data* values in a *sample*).
- ✓ The formula for the sampling distribution depends on
 - i. the *distribution* of the population,
 - ii. the statistic being considered, and
 - iii. the sample size used.
- ✓ The experiment consists of choosing a sample of size n from the population and measuring the statistic S . *The sampling distribution is the resulting probability distribution.*

Sampling Distribution of Means

- ✓ Imagine carrying out the following procedure:
 - ❑ Take a random sample of n independent observations from a population.
 - ❑ Calculate the mean of these n sample values. (Mean of samples)
 - ❑ Repeat the procedure until you have taken k samples of size n , calculate the sample mean of each k .
 - ❑ Therefore, there will be $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_k$ which will form a distribution of all the sample means.

The distribution is called sampling distribution of means.

Sampling Distribution of Means

- ✓ The mean of the sampling distribution of mean is equal to the population mean, $\mu_{\bar{X}} = \mu$.

- ✓ The standard deviation of the sampling distribution of means is

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \quad \text{for infinite population}$$

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{(N-n)}{(N-1)}} \quad \text{for finite population}$$

- ✓ If the population is normally distributed, the sampling distribution is normal regardless of sample size.

- ✓ By using Central Limit Theorem, $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

- ✓ If the population distribution is not necessarily normal and has mean μ , and standard deviation σ , then for sufficiently large n the sampling distribution of \bar{X} is approximately normal, with mean $\mu_{\bar{X}} = \mu$ and standard deviation

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}.$$

Example 1

In a study, it is reported that the amount of lead found in the printed sections of paper chocolate wrappers is normally distributed with mean 26 mg and standard deviation 10 mg. A sample of 500 paper chocolate wrappers is randomly selected. Find a sampling distribution for mean amount of lead found in the printed sections of paper chocolate wrappers.

Solution:

X : the amount of lead found in the printed sections of paper chocolate

$$\mu_{\bar{X}} = \mu = 26mg$$

$$\sigma_{\bar{X}}^2 = \frac{\sigma^2}{n} = \frac{100}{500} = 0.2$$

$$\bar{X} \sim N(26, 0.2)$$

Sampling Distribution for the Sum and Difference Between Two Means

Let $\bar{X}_1 \sim N\left(\mu_1, \frac{\sigma_1^2}{n_1}\right)$ and $\bar{X}_2 \sim N\left(\mu_2, \frac{\sigma_2^2}{n_2}\right)$; \bar{X}_1 and \bar{X}_2 is independent



The sampling distribution for the sum between two means:

$$\bar{X}_1 + \bar{X}_2 \sim N\left(\mu_1 + \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$



The sampling distribution for the difference between two means:

$$\bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

Example 2

Starting salaries for computer graduates per year at two universities are normally distributed with the following means and standard deviations. Sample from each university are taken and showed below.

	<i>University A</i>	<i>University B</i>
<i>Mean</i>	<i>RM 30 000</i>	<i>RM 26 400</i>
<i>Standard Deviation</i>	<i>RM 400</i>	<i>RM 200</i>
<i>Sample size</i>	<i>55</i>	<i>65</i>

Find a sampling distribution for the sum and difference between two means of the starting salaries for computer graduates.

Example 2: solution

X_A : Starting salaries for computer graduates per year at University A

X_B : Starting salaries for computer graduates per year at University B

$$\bar{X}_A \sim N\left(30000, \frac{400^2}{55}\right) \text{ and } \bar{X}_B \sim N\left(26400, \frac{200^2}{65}\right)$$

Sampling distribution for the sum between two means

$$\bar{X}_A + \bar{X}_B \sim N(56400, 3524.4755)$$

Sampling distribution for the difference between two means

$$\bar{X}_A - \bar{X}_B \sim N(3600, 3524.4755)$$

Sampling Distribution of Proportion

- ✓ π → Proportion, Probability and Percent for population
- ✓ $p = \frac{x}{n}$ → Sample proportion of x successes in a sample of size n
- ✓ $q = 1 - p$ → Sample proportion of failures in a sample of size n
- ✓ x is the binomial random variable created by counting the number of successes picked by drawing n times from the population.

$$X \sim \text{Bin}(n, \pi)$$

- ✓ The shape of the binomial distribution looks fairly Normal as long as n is large and/or p is not too extreme (not close to 0 or 1).

$$np(1-p) > 5, \text{ then } p \sim N\left(\pi, \frac{\pi(1-\pi)}{n}\right)$$

Sampling Distribution of Proportion

- ✓ The sampling distribution for proportions is a distribution of the proportions of all possible n samples that could be taken in a given situation.
- ✓ That is, the sample proportion (percent of successes in a sample), is approximately Normally distributed with

- Mean, $\mu_p = \pi$

- Standard Deviation, $\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}}$

Example 3

Adam of the Gambang Football Club had a 90% kick trials in the previous training. That is, on average, Adam made 9 of every 10 kicks. Assume 100 kicks were randomly selected, find the sampling distribution for proportion of Adam's kick.

Solution:

X : number of success kick

$$\mu_P = 0.90 \quad \sigma_P^2 = \frac{0.9(1-0.9)}{100} = 0.0009$$

$$\therefore P \sim N(0.90, 0.0009)$$

Sampling Distribution for the Sum and Difference Between Two Proportions

Let $P_1 \sim N\left(\pi_1, \frac{\pi_1(1-\pi_1)}{n_1}\right)$ and $P_2 \sim N\left(\pi_2, \frac{\pi_2(1-\pi_2)}{n_2}\right)$; P_1 and P_2 is independent



The sampling distribution for the sum between two proportions:

$$P_1 + P_2 \sim N\left(\pi_1 + \pi_2, \frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}\right)$$



The sampling distribution for the difference between two proportions:

$$P_1 - P_2 \sim N\left(\pi_1 - \pi_2, \frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}\right)$$

Example 4

A Ministry of Higher Education Learning survey shows about 89% of male students and 95% of female students rate their university experience as “good” or “excellent”. Assume this result is true for the current population of all students. Let the sample proportion of students in a random sample of 900 male and 1500 female who hold this view. Find its sampling distribution.

Solution:

X_{male} : No. of male students that rate their university experience as good/excellent

X_{female} : No. of female students that rate their university experience as good/excellent

$$P_{male} \sim N\left(0.89, \frac{0.89(1-0.89)}{900}\right) \text{ and } P_{female} \sim N\left(0.95, \frac{0.95(1-0.95)}{1500}\right)$$

$$P_{male} \sim N(0.89, 0.00011) \quad \text{and} \quad P_{female} \sim N(0.95, 0.00003)$$

Sampling Distribution for the sum between two proportions

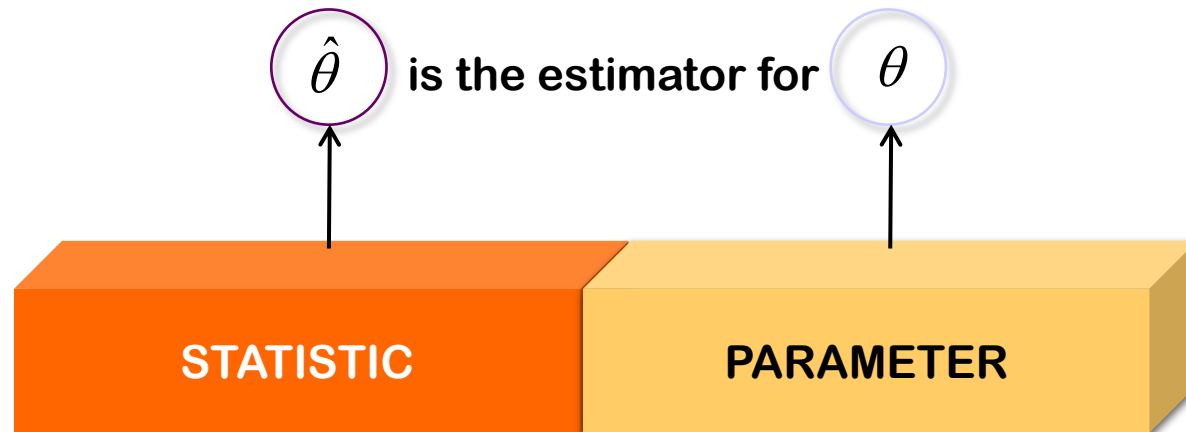
$$\rightarrow P_{male} + P_{female} \sim N(1.8400, 0.00014)$$

Sampling Distribution for the difference between two proportions

$$\rightarrow P_{male} - P_{female} \sim N(-0.06, 0.00014)$$

2.2 ESTIMATE, ESTIMATION, ESTIMATOR

- *Probability function* are actually families of models in the sense that each include one or more parameters.
- **Example:** *Poisson, Binomial, Normal*
- Any function of a random sample whose objective is to approximate a parameter is called a *statistic or an estimator*.



Estimations & Estimate

- ✓ **Estimation** – Is the entire process of using an estimator to produce an estimate of the parameter
- ✓ 2 types of estimation

1. *Point Estimate*

is a specific numerical value calculated from a sample data representing the unknown population parameter.

Parameter	Point Estimator
μ	\bar{X}
σ^2	S^2
π	p

2. *Interval Estimate*

A pair numbers calculated from the sample data, creating an interval within which the parameter is estimated to lie, $a < \theta < b$.

Properties of Good Point Estimator

i. Unbiased

As estimator $\hat{\theta}$ is said to be unbiased if $E(\hat{\theta}) = \theta$ for all θ .

ii. Consistency

An estimator $\hat{\theta}$ is said to be consistent for θ if it converges in probability to θ , that is for all $\varepsilon > 0$ where ε is random error.

ii. Efficiency

Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two unbiased estimators for a parameter, θ . If $Var(\hat{\theta}_1) < Var(\hat{\theta}_2)$, we can say that $\hat{\theta}_1$ is more efficient than $\hat{\theta}_2$.

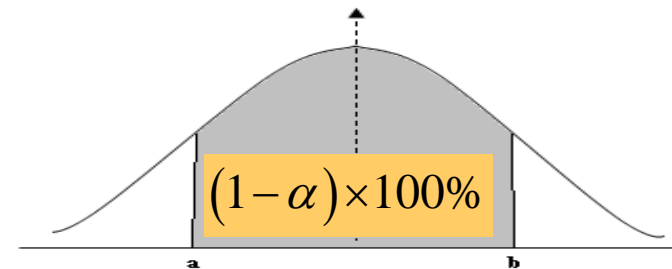
v. Sufficient

The estimator $\hat{\theta}$ is said to be sufficient for θ if for all θ and all possible sample points, the conditional probability distribution functions X_1, X_2, \dots, X_n of given $\hat{\theta}$ do not depend on θ .

Confidence Interval

✓ **CONFIDENCE INTERVAL**

- ❑ Range of numbers that have a high probability of containing the unknown parameter as an interior point.
- ❑ Are usually calculated so that this percentage is 95%, but 90%, 99%, 99.9% (or whatever) may also be applied.
- ❑ By looking at the width of a confidence interval, we can get a good sense of the estimator precision.
- ❑ Width = $b - a$



✓ **CONFIDENCE COEFFICIENT, $1 - \alpha$**

- ❑ The probability of correctly including the population parameter being estimated in the interval that is produced

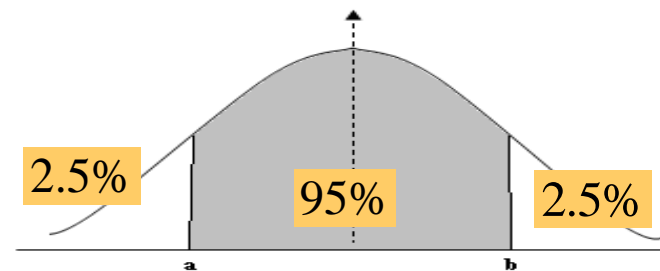
Level Of Confidence

- The confidence coefficient expressed as a percent.

$$(1 - \alpha) \times 100\%$$

EXAMPLE: 95% Confidence Interval for θ

- ✓ 95% confidence interval constructed will contain the unknown θ
- ✓ 5% will lie either entirel
- ✓ $\alpha = 0.05 = 5\%$
- ✓ $P(a < \theta < b) = 0.95$



Confidence Interval for θ

In General

$(1-\alpha)100\%$ confidence interval for θ is given by,

$$P\left(\underbrace{\hat{\theta} - \left(\text{distribution for } \hat{\theta}\right)\left(s.d \text{ for } \hat{\theta}\right)}_{\text{Estimation Error, } E} < \theta < \hat{\theta} + \underbrace{\left(\text{distribution for } \hat{\theta}\right)\left(s.d \text{ for } \hat{\theta}\right)}_{\text{Estimation Error, } E}\right) = 1 - \alpha$$

OR

$(1-\alpha)100\%$ confidence interval for θ

$$\left(\hat{\theta} \pm \left(\text{distribution for } \hat{\theta}\right)\left(s.d. \text{ for } \theta\right)\right)$$

- **For normal distribution,**

- (i) Sample mean, \bar{X} is the best estimator for the population mean, μ .
- (ii) $\bar{X} \sim N\left(\mu, \sigma^2/n\right)$.

Based on the **Central Limit Theorem**, the confidence interval can be constructed as follows:

$$\begin{aligned}
 1 - \alpha &= P\left(-Z_{\alpha/2} < Z < Z_{\alpha/2}\right) \\
 &= P\left(-Z_{\alpha/2} < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < Z_{\alpha/2}\right) \\
 &= P\left(-\bar{X} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < -\mu < -\bar{X} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right) \\
 &= P\left(\bar{X} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right)
 \end{aligned}$$

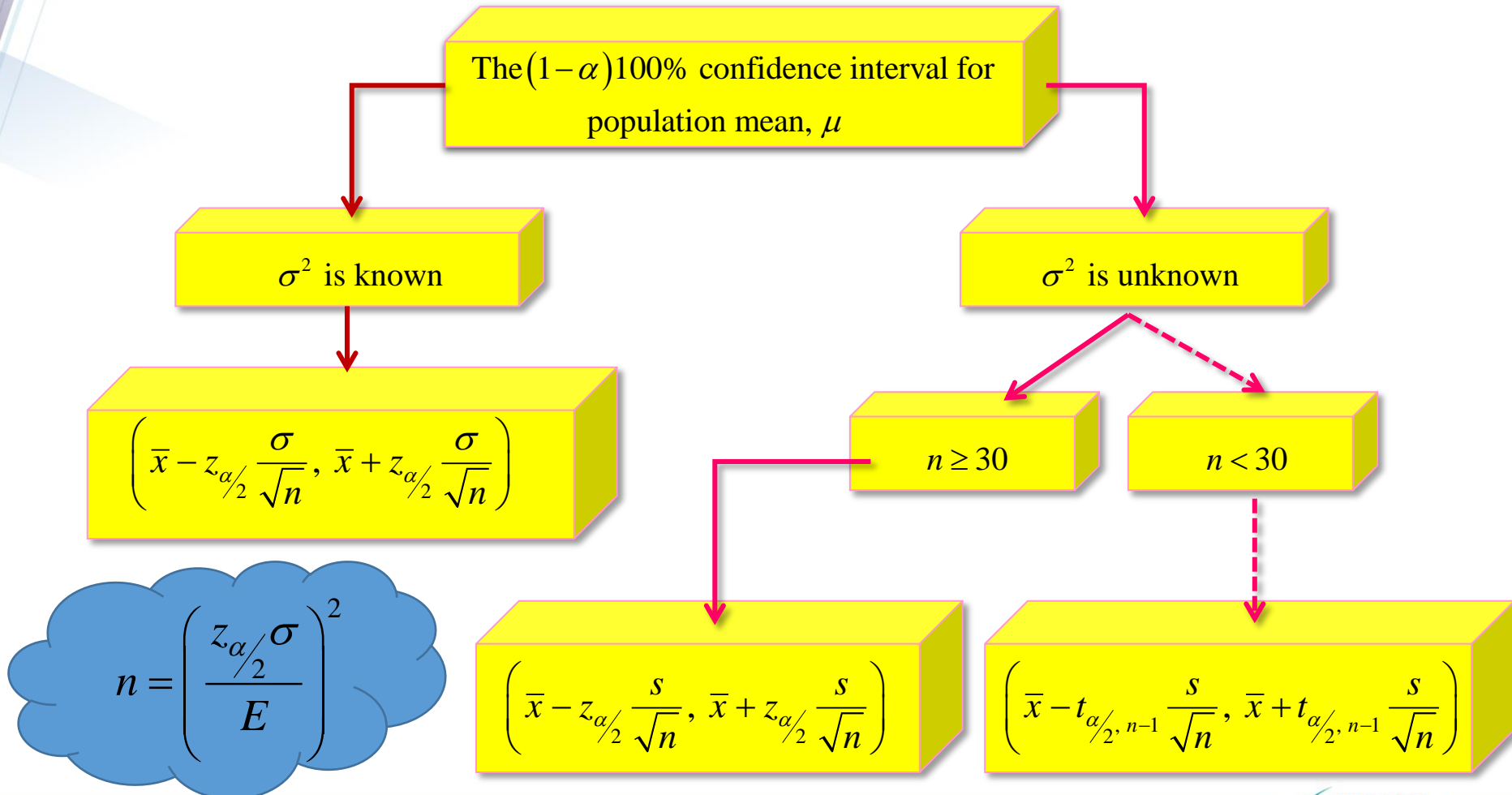
$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

Thus, the $(1 - \alpha)100\%$ confidence interval for μ is given by

$$\left(\bar{X} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{X} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right) \quad \text{or} \quad \left(\bar{X} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right)$$

if population variance, σ^2 is known.

2.3 CONFIDENCE INTERVAL FOR POPULATION MEAN



Example 5

A plant produces steel sheets whose weights are known to be normally distributed with a standard deviation of 2.4 kg. A random sample of 26 sheets had a mean weight of 3.14 kg. Find the 99% confidence interval for the population mean weight and give the interpretation of the parameter estimate.

Solution: X : the weight of steel sheets

$$n = 26, \quad \bar{x} = 3.14 \text{ kg}, \quad \sigma = 2.4 \text{ kg}, \quad z_{0.01/2} = z_{0.005} = 2.5758$$

A 99% confidence interval for population mean, μ

$$\begin{aligned} &= \left(\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right) \\ &= \left(3.14 - 2.5758 \left[\frac{2.4}{\sqrt{26}} \right], 3.14 + 2.5758 \left[\frac{2.4}{\sqrt{26}} \right] \right) \\ &= (1.9276, 4.3524) \text{ kg} \end{aligned}$$

Interpretation: We are 99% confident that the true mean weight of steel sheets lies within 1.9276 and 4.3524 kg.

Example 6

A random sample of 25 bulbs was selected and the specified brightness was evaluated for each bulb by measuring the amount of electric current required. The sample mean of the current required is 280.3 micro amps and the sample standard deviation is 10.3 micro amps. Construct a 90% confidence interval of mean for the **population** and explain your answers.

Solution:

X : the amount of electric current required for bulbs

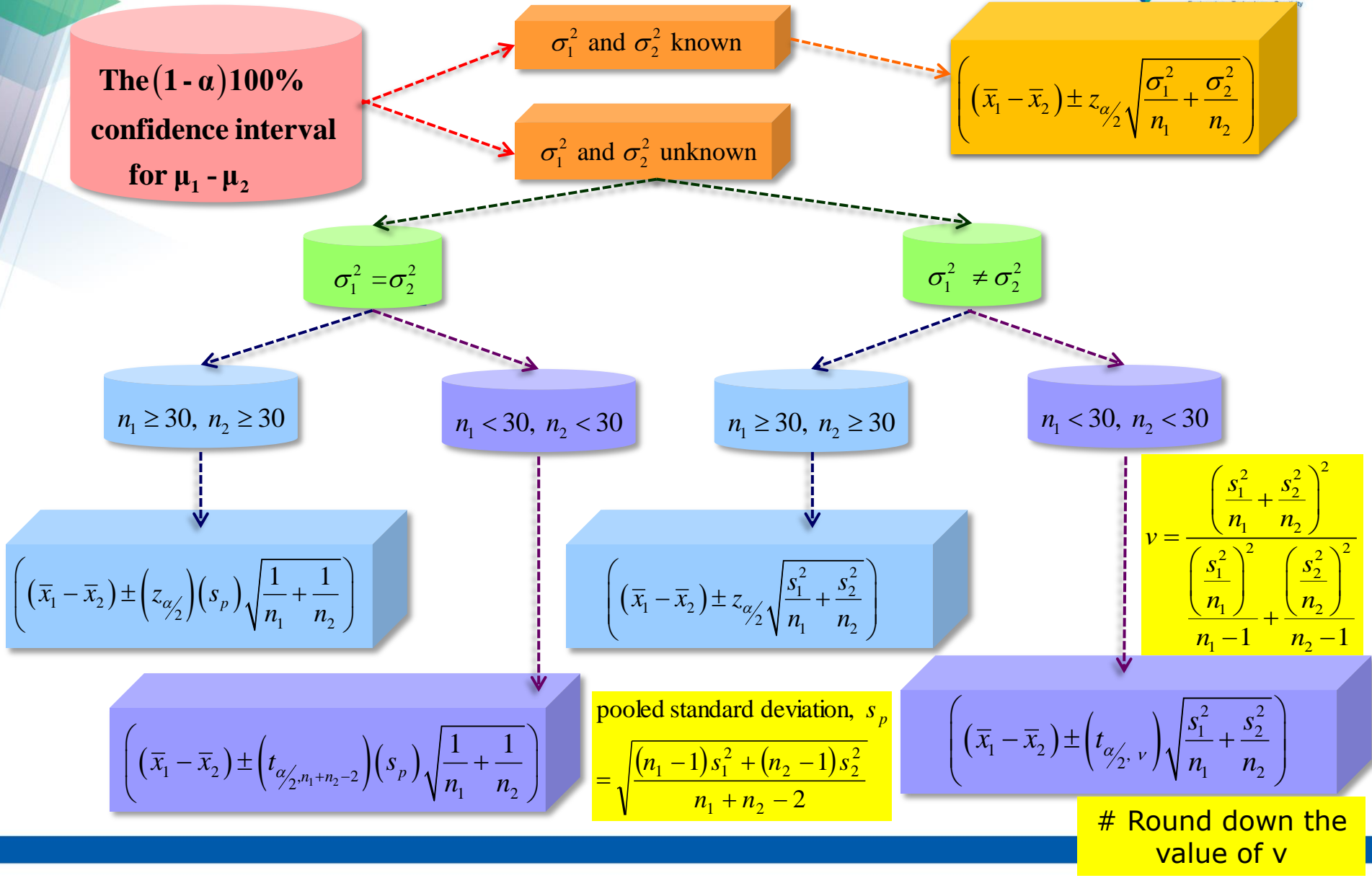
$$n = 25, \quad \bar{x} = 280.3 \text{ micro amps}, \quad s = 10.3 \text{ micro amps}, \quad t_{0.1/2, 25-1} = t_{0.05, 24} = 1.7109$$

A 90% confidence interval for population mean, μ

$$\begin{aligned} &= \left(\bar{x} - t_{\alpha/2, n-1} \frac{s}{\sqrt{n}}, \bar{x} + t_{\alpha/2, n-1} \frac{s}{\sqrt{n}} \right) \\ &= \left(280.3 - 1.7109 \frac{10.3}{\sqrt{25}}, 280.3 + 1.7109 \frac{10.3}{\sqrt{25}} \right) \\ &= (276.7755, 283.8245) \text{ micro amps} \end{aligned}$$

Interpretation: We are 90% confident that the true mean of current required for bulb lies within 276.7755 and 283.8245 micro amps.

2.4 CONFIDENCE INTERVAL FOR THE DIFFERENCE BETWEEN TWO POPULATION MEANS



Example 7

Find the 95% confidence interval for the **difference mean** of children's sleep time and adults sleep time if given that the variances for children's sleep time is 0.81 hours while for adults is 0.25 hours. The mean sample sleep time for 30 children's are 10 hours while for 40 adults are 7 hours and give the comment on parameter estimate.

Solution:

X_C : the children's sleep time
 X_A : the adult's sleep time

The populations should be independent with each other.

Example 7: solution

<i>Children (C)</i>	<i>Adults (A)</i>	$z_{0.05/2} = z_{0.025} = 1.9600$
$n_C = 30$	$n_A = 40$	
$\bar{x}_C = 10$	$\bar{x}_A = 7$	
$\sigma_C^2 = 0.81$	$\sigma_A^2 = 0.25$	

A 95% confidence interval for difference between two population means, $\mu_C - \mu_A$

$$\begin{aligned}
 &= \left((\bar{x}_C - \bar{x}_A) \pm z_{\alpha/2} \sqrt{\frac{\sigma_C^2}{n_C} + \frac{\sigma_A^2}{n_A}} \right) \\
 &= \left((10 - 7) \pm (1.9600) \sqrt{\frac{0.81}{30} + \frac{0.25}{40}} \right) \\
 &= (2.6426, 3.3574) \text{ hours}
 \end{aligned}$$

Interpretation: We are 95% confident that the difference between the true means of children's sleep time and adults sleep time lies within 2.6426 and 3.3574 hours.

Example 8

A biotechnology company produces a therapeutic drug whose concentration is normally distributed. Two new methods of producing this drug have been proposed, and the method that produces more drugs will be chosen by management. The researchers chose a sample of size 15 from each method and obtained the following data in grams per liter.

Type of Method	Mean	Standard deviation
Method X	16.22	0.05
Method Y	17.41	1.00

By assuming that the both population variances are equal, construct a 99% confidence interval on the difference between mean of the concentration by method Y and method X. Interpret the answer resulted.

Example 8: solution

X_X : the concentration of therapeutic drug by using Method X

X_Y : the concentration of therapeutic drug by using Method Y

$$\begin{aligned}\text{pooled standard deviation, } s_p &= \sqrt{\frac{(n_Y - 1) s_Y^2 + (n_X - 1) s_X^2}{n_Y + n_X - 2}} \\ &= \sqrt{\frac{(15 - 1) 1.00^2 + (15 - 1) 0.05^2}{15 + 15 - 2}} \\ &= 0.7080\end{aligned}$$

$$\begin{aligned}t_{\alpha/2, n_Y + n_X - 2} \\ = t_{0.01/2, 15 + 15 - 2} = 2.7633\end{aligned}$$

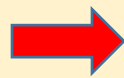
A 99% confidence interval for difference between two population means, $\mu_Y - \mu_X$

$$\begin{aligned}&= \left((\bar{x}_Y - \bar{x}_X) \pm t_{\alpha/2, n_Y + n_X - 2} s_p \sqrt{\frac{1}{n_Y} + \frac{1}{n_X}} \right) \\ &= \left((17.41 - 16.22) \pm (2.7633)(0.7080) \sqrt{\frac{1}{15} + \frac{1}{15}} \right) \\ &= (0.4756, 1.9044) \text{ grams per litre}\end{aligned}$$

Interpretation: We are 99% confident that the difference between the true means of the concentration produced by using method Y and method X lies within 0.4756 and 1.9044 grams per litre.

How to interpret negative confidence interval for μ ?

Two chemical companies, Company A and Company B supply a raw material where the most important element in this material is the concentration. The standard deviations of concentration produced by both companies are 5.81 and 4.70, respectively. The sample mean of concentration in a random sample of 10 batches produced by Company A is 90.25 grams per liter, while for Company B, a random sample 15 batches yields 87.54 grams per liter. Construct a 96% confidence interval for the difference mean for the concentration produced by both companies

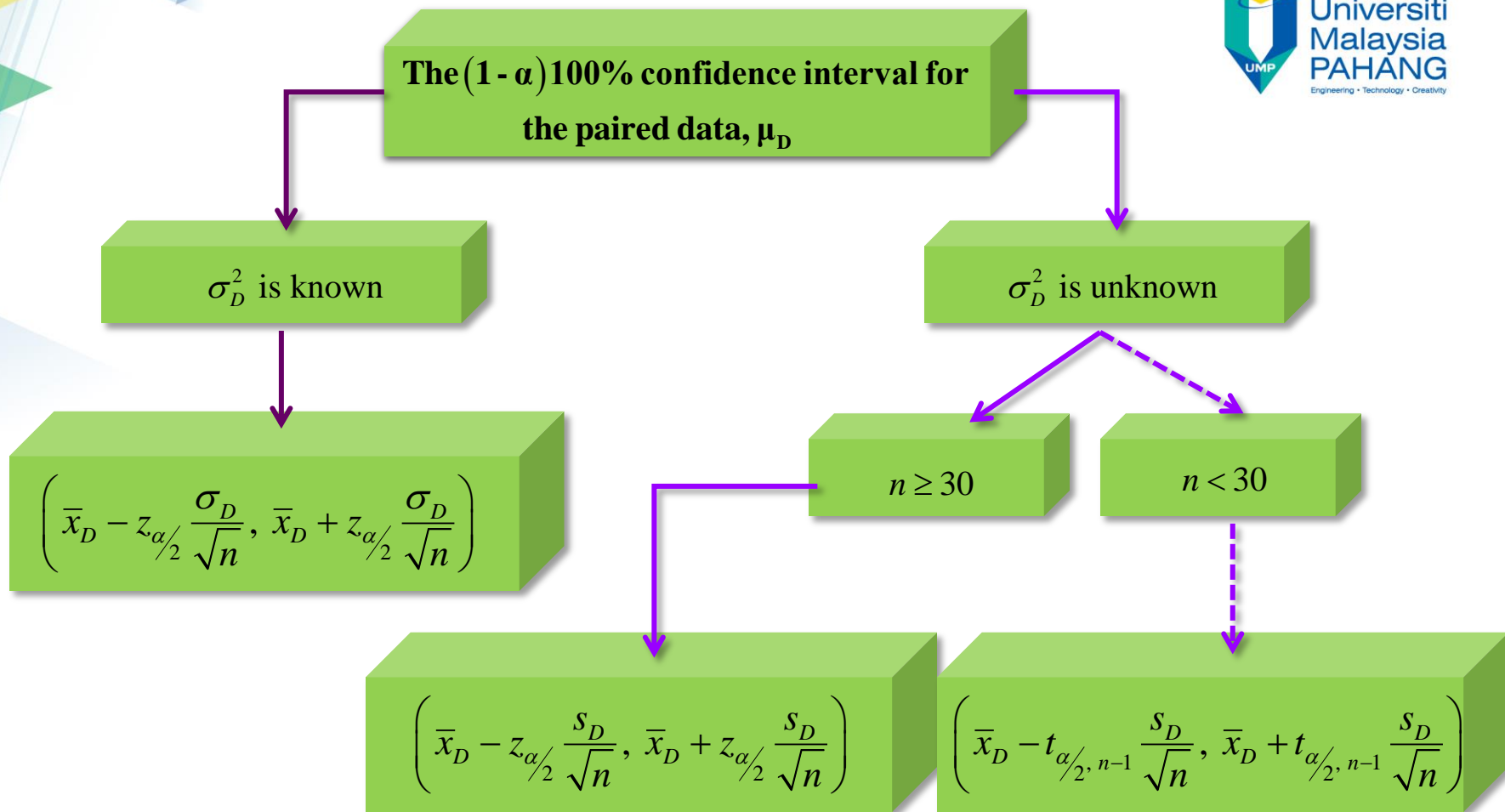
 $(-1.8120, 7.2320)$ grams per liter

We are 96% confident that the population mean difference for the concentration produced by both companies is between 1.8120 grams per litre more by Company B to 7.2320 grams per litre more by Company A, compared to each other, respectively.

2.5 CONFIDENCE INTERVAL FOR THE PAIRED DATA

Paired Data?

- ✓ **Paired data is the difference of two dependent samples.**
- ✓ In paired data, each data point in one sample is **matched** to a unique data point in the second sample (matched couplings).
- ✓ **Example** of a paired data is a **pre-test/post-test** study design in which a factor is measured **before and after** an intervention/treatment.



where

$D = X_1 - X_2$ is the difference between two data sets,
 μ_D is a mean difference of the population,
 \bar{x}_D is a sample mean of the differences, and
 s_D is a sample standard deviation of the differences.

Example 9

A tyre manufacturer wishes to compare the made of thread wear of tyres between the new and conventional materials. One tyre of each type is placed on each front wheel of each of 10 front wheel drive vehicles. The choice as to which type of tyre goes on the right wheel and which goes on the left is made with the flip of a coin. Each car is driven for 40 000 miles, then the tyres are removed, and the depth of the thread on each is measured. The results were as follows:

<i>Car</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	<i>8</i>	<i>9</i>	<i>10</i>
<i>New material</i>	4.35	5.00	4.21	5.03	5.71	4.61	4.70	6.03	3.80	4.70
<i>Old material</i>	4.19	4.62	4.04	4.72	5.52	4.26	4.27	6.24	3.46	4.50
<i>Difference, D</i>	0.16	0.38	0.17	0.31	0.19	0.35	0.43	- 0.21	0.34	0.20

Calculate a 90% confidence interval for the **mean difference in thread wear between new and old materials** in a way that takes advantage of the reduces variability produced by the paired design. Hence give the comment on your answer.

Example 9: solution

X_N : the depth of the thread wears of tyres made by using new materials

X_O : the depth of the thread wears of tyres made by using old materials

$$n = 10, \quad \bar{x}_D = 0.2320, \quad s_D = 0.1826, \quad t_{0.10/2, 10-1} = t_{0.05, 9} = 1.8331$$

A 90% confidence interval for μ_D

$$\begin{aligned} &= \left(\bar{x} - t_{\alpha/2, n-1} \frac{s_D}{\sqrt{n}}, \bar{x}_D + t_{\alpha/2, n-1} \frac{s_D}{\sqrt{n}} \right) \\ &= \left(0.2320 - [1.8331] \frac{0.1826}{\sqrt{10}}, 0.2320 + [1.8331] \frac{0.1826}{\sqrt{10}} \right) \\ &= (0.1262, 0.3378) \end{aligned}$$

Interpretation: We are 90% confident that the true mean difference in thread wear between new and old materials lie within 0.1262 and 0.3378.

2.6 CONFIDENCE INTERVAL FOR THE POPULATION PROPORTION

The $(1 - \alpha)$ 100% confidence interval for a population proportion, π

$$\left(p - z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}}, p + z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}} \right)$$

Sample proportion $\leftarrow p = \frac{x}{n}$ and $q = 1 - p$ \rightarrow *Number of successive*
 \rightarrow *Sample size*

where $np \geq 5$ and $nq \geq 5$

$$n = \left(\frac{z_{\alpha/2}}{E} \right)^2 p(1-p)$$

Note that if the sample proportion is not given, consider equal value of sample proportion that is $p = q = 0.5$.

Example 10

The fraction of defective integrated circuits produced in a photolithography process is being studied. A random sample of 200 circuits is tested, revealing 13 defectives.

- a) Calculate a 95% confidence interval on the fraction of defective circuits produced by this process and give a comment on the resulted confidence interval.*

- b) How large the sample if we wish to be at least 95% confident that the error in estimating p is less than 0.02?*

Example 10: solution

a) X : The defective integrated circuits produced in a photolithography process

$$p = \frac{x}{n} = \frac{13}{200} = 0.0650, \quad z_{\alpha/2} = z_{0.05/2} = 1.9600$$

A 95% confidence interval for population proportion π

$$\begin{aligned} &= \left(p - z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}}, p + z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}} \right) \\ &= \left((0.0650) - (1.9600) \sqrt{\frac{0.0650(1-0.0650)}{200}}, (0.0650) + (1.9600) \sqrt{\frac{0.0650(1-0.0650)}{200}} \right) \\ &= (0.0308, 0.0992) \end{aligned}$$

Interpretation: We are 95% confident that the fraction of the defective integrated circuits produced in a photolithography process is between 3.08% and 9.92%.

b)

$$n = \left(\frac{z_{\alpha/2}}{E} \right)^2 p(1-p) = \left(\frac{1.9600}{0.02} \right)^2 (0.0650)(1-0.0650) = 583.6831 \approx 584 \text{ circuits}$$

2.7 CONFIDENCE INTERVAL FOR THE DIFFERENCE BETWEEN TWO POPULATION PROPORTIONS

The $(1 - \alpha)100\%$ confidence interval for the difference between two population proportions, $\pi_1 - \pi_2$

$$\left((p_1 - p_2) \pm z_{\alpha/2} \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \right)$$

Example 11

Two different types of injection-molding machines are used to form plastic parts. A part is considered defective if it has excessive shrinkage or is discoloured. Two random samples, each of size 300, are selected, and 15 defective parts are found in the sample from machine 1 while 8 defective parts are found in the sample from machine 2. Construct a 98% confidence interval on the difference in the two population proportions that are defective. Give an interpretation for your answer.

Solution:

X_1 : the number of defective plastics part produced by Machine 1

X_2 : the number of defective plastics part produced by Machine 2

<i>Machine 1</i>	<i>Machine 2</i>	$z_{\alpha/2} = z_{0.02/2} = 2.3263$
$n_1 = 300$	$n_2 = 300$	
$x_1 = 15$	$x_2 = 8$	
$p_1 = \frac{x_1}{n_1} = 0.0500$	$p_2 = \frac{x_2}{n_2} = 0.0267$	

Example 11: solution

<i>Machine 1</i>	<i>Machine 2</i>	$z_{\alpha/2} = z_{0.02/2} = 2.3263$
$n_1 = 300$	$n_2 = 300$	
$x_1 = 15$	$x_2 = 8$	
$p_1 = \frac{x_1}{n_1} = 0.0500$	$p_2 = \frac{x_2}{n_2} = 0.0267$	

A 98% confidence interval for difference between two population proportions $\pi_1 - \pi_2$

$$\begin{aligned}
 &= \left((p_1 - p_2) \pm z_{\alpha/2} \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \right) \\
 &= \left((0.0500 - 0.0267) \pm 2.3263 \sqrt{\frac{0.0500(0.9500)}{300} + \frac{0.0267(0.9733)}{300}} \right) \\
 &= (-0.0131, 0.0597)
 \end{aligned}$$

Interpretation: We are 98% confident that the difference in the two true proportions for the defective of plastics part produced by using Machine 1 and Machine 2 is between 1.31% more by Machine 2 to 5.97% more by Machine 1, compared to each other respectively.

2.8 CONFIDENCE INTERVAL FOR POPULATION VARIANCE AND POPULATION STANDARD DEVIATION

The $(1 - \alpha)$ 100% confidence interval for a population variance, σ^2

$$\left(\frac{(n-1)s^2}{\chi_{\alpha/2, n-1}^2}, \frac{(n-1)s^2}{\chi_{1-\alpha/2, n-1}^2} \right)$$

The $(1 - \alpha)$ 100% confidence interval for a population standard deviation, σ

$$\left(\sqrt{\frac{(n-1)s^2}{\chi_{\alpha/2, n-1}^2}}, \sqrt{\frac{(n-1)s^2}{\chi_{1-\alpha/2, n-1}^2}} \right)$$

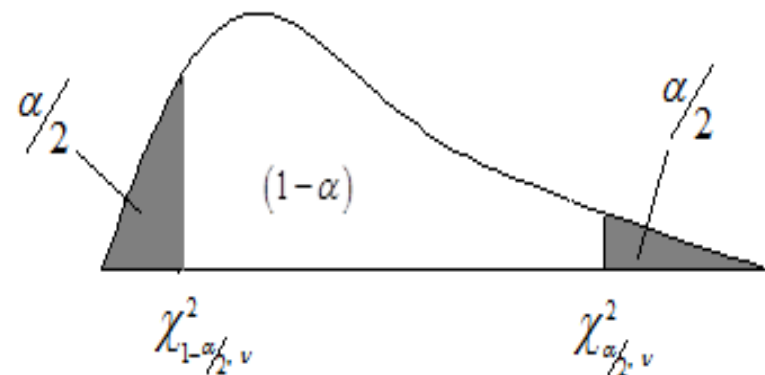
where

$$v = n - 1$$

and

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi_{n-1}^2$$

(Chi-square distribution).



Example 12

A random sample of 10 rulers produced by a machine gives a set of data below (in cm).

<i>100.131</i>	<i>100.072</i>	<i>100.023</i>	<i>99.994</i>	<i>99.885</i>
<i>100.146</i>	<i>100.037</i>	<i>100.108</i>	<i>99.929</i>	<i>100.210</i>

Find the 95% confidence interval for the variance and standard deviation of all the rulers produced by the machine. Give the comment on the 95% confidence interval for the true variance.

Example 12: solution

X : the length of rulers produced by a machine

A 95% confidence interval for population variance

$$\begin{aligned}
 &= \left(\frac{(n-1)s^2}{\chi_{\alpha/2, n-1}^2}, \frac{(n-1)s^2}{\chi_{1-\alpha/2, n-1}^2} \right) \\
 &= \left(\frac{9(0.0101)}{19.0228}, \frac{9(0.0101)}{2.7004} \right) \\
 &= (0.0048, 0.0337)
 \end{aligned}$$

$$n = 10, \quad s = 0.1006, \quad s^2 = 0.0101$$

$$\chi_{\alpha/2, n-1}^2 = \chi_{0.05/2, 10-1}^2 = 19.0228$$

$$\chi_{1-\alpha/2, n-1}^2 = \chi_{1-0.05/2, 10-1}^2 = 2.7004$$

Interpretation: We are 95% confident that the true variance of the length for the rulers produced by a machine lie within 0.0048 and 0.0337.

A 95% confidence interval for population standard deviation

$$= (\sqrt{0.0048}, \sqrt{0.0337}) = (0.0693, 0.1836) \text{ cm}$$

2.9 CONFIDENCE INTERVAL ON THE RATIO OF TWO POPULATION VARIANCES

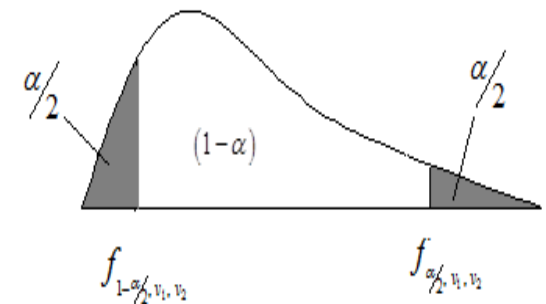
The $(1 - \alpha)$ 100% confidence interval for $\frac{\sigma_1^2}{\sigma_2^2}$

$$\frac{s_1^2 / \sigma_1^2}{s_2^2 / \sigma_2^2} \sim F_{v_1, v_2}$$

(F distribution)

$$\left(\frac{s_1^2}{s_2^2} f_{1-\alpha/2, v_2, v_1}, \frac{s_1^2}{s_2^2} f_{\alpha/2, v_2, v_1} \right)$$

where $v_1 = n_1 - 1$, $v_2 = n_2 - 1$ and $f_{1-\frac{\alpha}{2}, v_2, v_1} = \frac{1}{f_{\frac{\alpha}{2}, v_1, v_2}}$



Example 13

The machine in Example 12 is serviced. A random sample of 12 rulers produced by the machine after the service gives a set of data below.

100.030	100.011	100.022	100.043	99.904	99.965
100.046	100.067	100.088	99.989	100.110	100.051

Find the 95% confidence interval for the ratio of variances for all rulers produced by the machine before and after the service. Hence, interpret your answer.

Solution:

X_B : the length of rulers produced by a machine before the service

X_A : the length of rulers produced by a machine after the service

Example 13: solution

Before	After	
$n_B = 10$	$n_A = 12$	$f_{\alpha/2, n_A-1, n_B-1} = f_{0.05/2, 10-1, 12-1} = 3.5879$
$s_B = 0.1006$	$s_A = 0.0557$	$f_{\alpha/2, n_B-1, n_A-1} = f_{0.05/2, 12-1, 10-1} = 3.9121$
$s_B^2 = 0.0101$	$s_A^2 = 0.0031$	

A 95% confidence interval on the ratio of two population variances $\frac{\sigma_B^2}{\sigma_A^2}$

$$\begin{aligned}
 &= \left(\frac{s_B^2}{s_A^2} \frac{1}{f_{\alpha/2, n_B-1, n_A-1}}, \frac{s_B^2}{s_A^2} f_{\alpha/2, n_A-1, n_B-1} \right) \\
 &= \left(\frac{0.0101}{0.0031} \frac{1}{3.5879}, \frac{0.0101}{0.0031} (3.9121) \right) \\
 &= (0.9081, 12.7459)
 \end{aligned}$$

Interpretation: We are 95% confident that the ratio of two true variances for the length of rules produced by a machine before and after service lie within 0.9081 and 12.7459.

REFERENCES

1. Montgomery D. C. & Runger G. C. 2011. *Applied Statistics and Probability for Engineers*. 5th Edition. New York: John Wiley & Sons, Inc.
2. Walpole R.E., Myers R.H., Myers S.L. & Ye K. 2011. *Probability and Statistics for Engineers and Scientists*. 9th Edition. New Jersey: Prentice Hall.
3. Navidi W. 2011. *Statistics for Engineers and Scientists*. 3rd Edition. New York: McGraw-Hill.
4. Bluman A.G. 2009. *Elementary Statistics: A Step by Step Approach*. 7th Edition. New York: McGraw–Hill.
5. Triola, M.F. 2006. *Elementary Statistics*. 10th Edition. UK: Pearson Education.
6. Satari S. Z. et al. Applied Statistics Module New Version. 2015. Penerbit UMP. Internal used.

Thank You

NEXT: CHAPTER 3 HYPOTHESIS TESTING