Estimation & Confidence Interval

Learning Outcomes

Learning Outcomes

At the end of this lesson, the learner should be able to:

1. Compute and interpret a point estimate of population mean by using the sample mean.
2. Calculate the confidence interval for a population mean using the normal distribution and the t-distribution.
3. State the criteria for the t-distribution to be applied in statistical estimation.
4. Solve real-life business problems by applying statistical estimation and confidence interval.

Introduction to Estimation

- As part of your marketing research, you need to estimate the amount spent per student for each meal at the fast food restaurant. **What can you do?**

- You can select a sample of say, 30 students and compute the mean dollar spent per meal. Let's say the amount is $4.50.

- $4.50 is an example of a Point Estimate.

- A **Point Estimate** is a single value estimate to approximate the population mean.

- This case study is an example of a using statistics to make inference about a population parameter.
- In this topic, you will apply concepts of Estimation & Confidence Interval to solve real life problems.

Point Estimation

- In statistics, point estimation involves the use of sample data to calculate a single value (known as a point estimate).
- For example,
  - ✓ Sample mean is a point estimate of the population mean, μ (as illustrated in the previous case study, sample mean of $4.50 is a Point Estimate to approximate the amount spent per student for each meal)
  - ✓ Sample standard deviation, s is a point estimate of the population standard deviation, σ.
- Point estimates are:
  - ✓ Often insufficient. It is either right or wrong.
  - ✓ Unreliable. We cannot be certain of the reliability of the estimate. (i.e. unable to tell how close we are to the true population value)

Confidence Interval

- Instead of using Point Estimate, a better approach is to take into account variability of sample to sample, and construct a Confidence Interval.
- Confidence Interval
  - Gives a **range of values** where the population parameter will likely lie
  - Provides a **degree of confidence** to estimate where the unknown population lies.

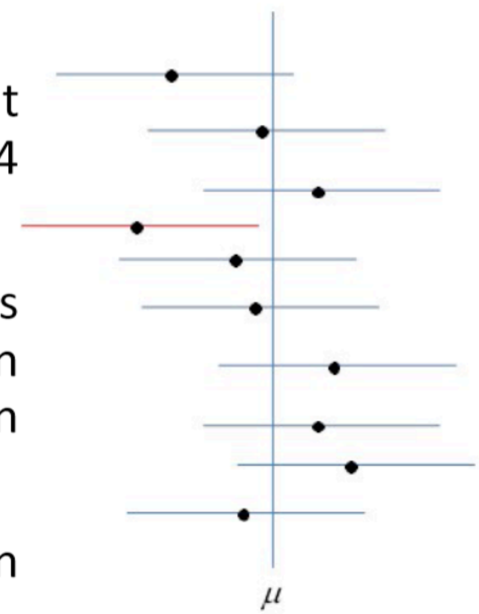
Confidence Interval

A range of values constructed from sample data so that the population parameter is likely to occur within that range at a specified probability. The specified probability is called Level of Confidence.

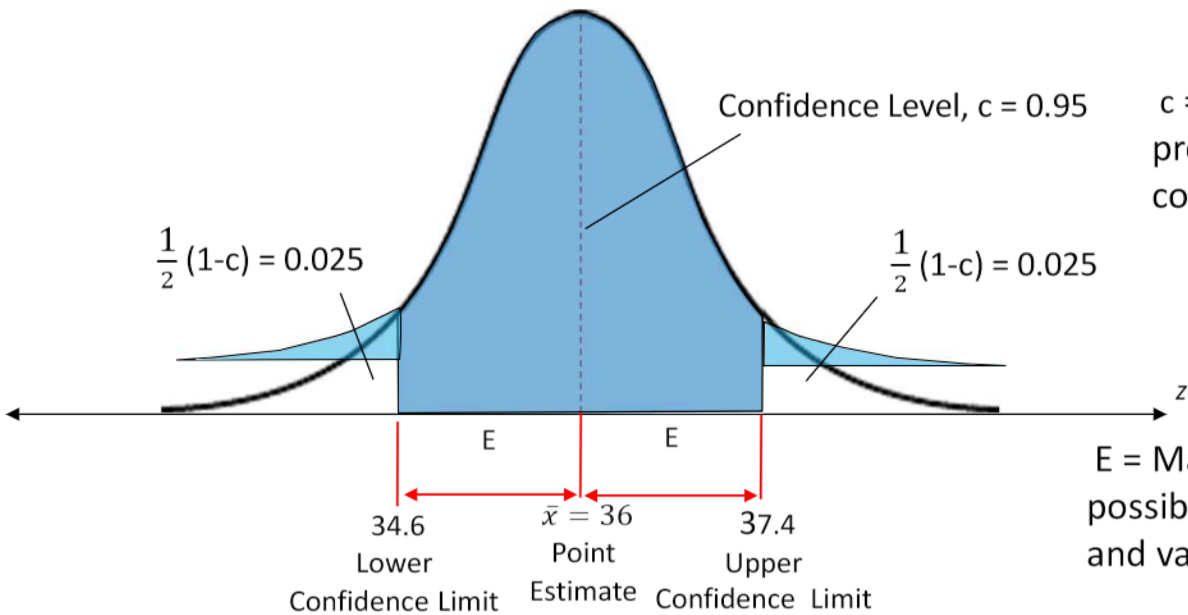
Confidence Interval

Example: The 90% confidence interval for the population mean of battery life is between 34.6 to 37.4 months.

- Based on the above example, we can say "We are 90% confident that the population mean of battery life is between 34.6 to 37.4 months." But what does it mean?
- The diagram on the right shows that there are many intervals obtained from many samples. For a 90% confidence interval, then it is expected that 90% of these intervals contain the population mean, while 10% do NOT contain the population mean.
- It is incorrect to say "There is a 90% probability that the population mean of battery life is between 34.6 to 37.4 months."



Graphical representation of Confidence Interval



c = confidence level, which is the probability that the interval estimate contains the population parameter

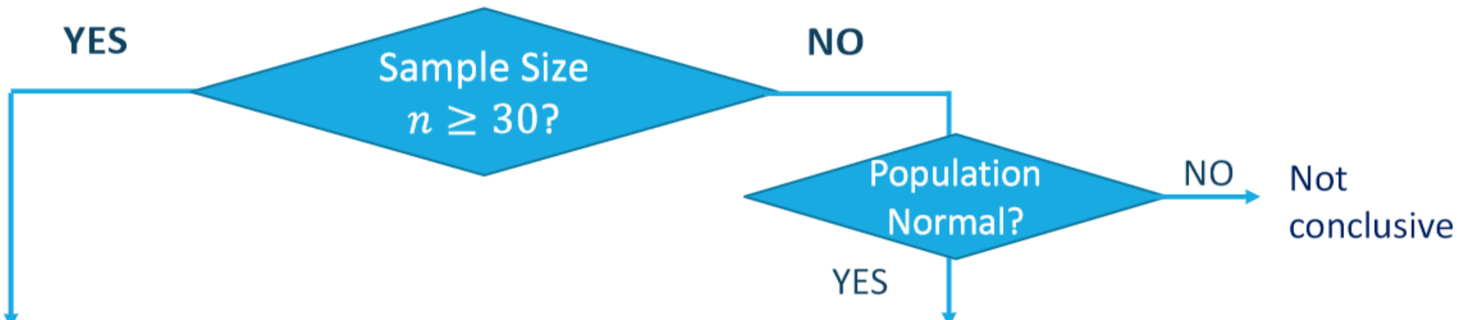
E = Margin of Error, which is the maximum possible distance between point estimate and value of parameter it is estimating

Confidence Interval with confidence level 100c% = Point Estimate $\pm$ Margin of Error

= $\bar{X} \pm$ E

Formulae for Confidence Interval

Confidence Interval, CI = Point Estimate $\pm$ Margin of Error
= $\bar{X} \pm E$ where E can be found using the chart below



σ known	$E = Zc\frac{\sigma}{\sqrt{n}}$ ($\bar{X}$ has normal distribution)	$E = Zc\frac{\sigma}{\sqrt{n}}$ ($\bar{X}$ has normal distribution)
σ unknown	$E = Zc\frac{s}{\sqrt{n}}$ ($\bar{X}$ has normal distribution)	$E = tc\frac{s}{\sqrt{n}}$ ($\bar{X}$ has t distribution with n-1 degree of freedom)

Formulae for Confidence Interval

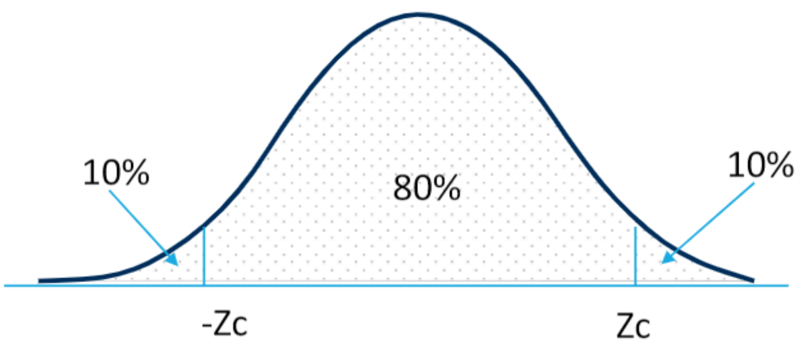
- Typically value of Z_c, also known as the Critical values, based on the Confidence Level are:

Confidence Level, c	Critical value, Z_c
90%	1.645
95%	1.96
99%	2.575

Note: Critical values ($-Z_c$ and Z_c) are values that separate sample statistics that are probable from sample statistics that are improbable, or unusual.

Formulae for Confidence Interval

Example 1: Find the critical values z_c necessary to form a confidence interval at the level of confidence of 80%.



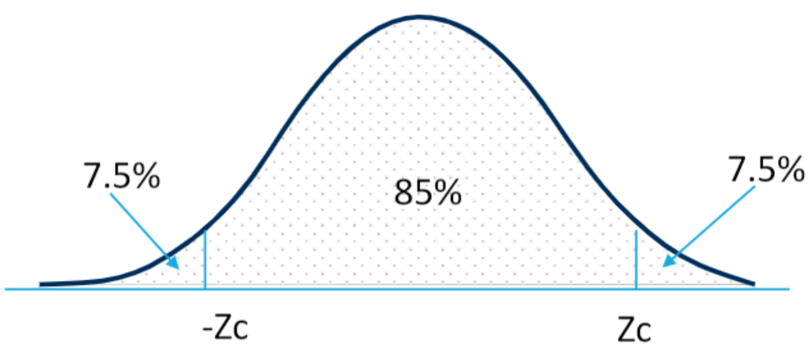
z	0.00	0.01	0.02	0.07	0.08
0.0	0.5000	0.5040	0.5080	0.5279	0.5319
0.1	0.5398	0.5438	0.5478	0.5675	0.5714
0.2	0.5793	0.5832	0.5871	0.6064	0.6103
0.3	0.6179	0.6217	0.6255	0.6443	0.6480
0.4	0.6554	0.6591	0.6628	0.6808	0.6844
0.5	0.6915	0.6950	0.6985	0.7157	0.7190
0.6	0.7257	0.7291	0.7324	0.7486	0.7517
0.7	0.7580	0.7611	0.7642	0.7794	0.7823
0.8	0.7881	0.7910	0.7939	0.8078	0.8106
0.9	0.8159	0.8186	0.8212	0.8340	0.8365
1.0	0.8413	0.8438	0.8461	0.8577	0.8599
1.1	0.8643	0.8665	0.8686	0.8790	0.8810
1.2	0.8849	0.8869	0.8888	0.8980	$\boxed{0.8997}$

$P(Z < Zc) = 90\%$

From table, $Zc = 1.28$

Formulae for Confidence Interval

Example 2: Find the critical values z_c necessary to form a confidence interval at the level of confidence of 85%.



z	0.00	0.01	0.02	0.03	0.04
0.0	0.5000	0.5040	0.5080	0.5120	0.5160
0.1	0.5398	0.5438	0.5478	0.5517	0.5557
0.2	0.5793	0.5832	0.5871	0.5910	0.5948
0.3	0.6179	0.6217	0.6255	0.6293	0.6331
0.4	0.6554	0.6591	0.6628	0.6664	0.6700
0.5	0.6915	0.6950	0.6985	0.7019	0.7054
0.6	0.7257	0.7291	0.7324	0.7357	0.7389
0.7	0.7580	0.7611	0.7642	0.7673	0.7704
0.8	0.7881	0.7910	0.7939	0.7967	0.7995
0.9	0.8159	0.8186	0.8212	0.8238	0.8264
1.0	0.8413	0.8438	0.8461	0.8485	0.8508
1.1	0.8643	0.8665	0.8686	0.8708	0.8729
1.2	0.8849	0.8869	0.8888	0.8907	0.8925
1.3	0.9032	0.9049	0.9066	0.9082	0.9099
1.4	0.9192	0.9207	0.9222	0.9236	$\boxed{0.9251}$

P (Z < Zc) = 92.5%

From table, Zc =1.44

Case 1: Known Population Standard Deviation, $\boldsymbol{\sigma}$

Example 3: The quantity of mineral water dispensed by automated machines into plastic bottles is approximately normally distributed with standard deviation of 24 millilitres. A random sample of 25 such bottles was found to have a mean quantity of 503 millilitres.

a) Find the standard error of the mean.

b) Find a 90 % confidence interval for the mean quantity of mineral water dispensed by the machines.

Case 1: Known Population Standard Deviation, $\boldsymbol{\sigma}$

Given : $\sigma = 24$ n = 25 $\bar{x}$ = 503 Given the population is normal.

X = quantity of mineral water dispensed

a) Standard error of the mean = $\frac{\sigma}{\sqrt{n}} = \frac{24}{\sqrt{25}} = 4.8$

b) For a 90% confidence interval, $Z_c = 1.645$

90% confidence interval = $\bar{X} \pm Z_c\frac{\sigma}{\sqrt{n}}$

$= 503 \pm 1.645\frac{24}{\sqrt{25}}$

$= (495.104 , 510.896)$

The 90% confidence interval for mean quantity of mineral water dispensed by the machines is between 495.104 and 510.896 millilitres

Case 2: Sample size $\geq$ 30 and Population Standard Deviation, $\boldsymbol{\sigma}$ unknown

- If $\bar{X}$ is **normally distributed** with **unknown** population standard deviation σ, but with sample size ≥ 30, then use the sample standard deviation s as an estimate for $\boldsymbol{\sigma}$

- Hence, Confidence Interval $= \bar{\boldsymbol{X}} \pm \mathrm{E}$

$$= \bar{\boldsymbol{X}} \pm \mathrm{Z}c \frac{s}{\sqrt{\boldsymbol{n}}}$$

Case 2: Sample size $\geq$ 30 and Population Standard Deviation, $\boldsymbol{\sigma}$ unknown

Example 4: The heights of a random sample of 40 NYP students yield a mean of 173.8 cm and a standard deviation of 6.8 cm. Assume population is normally distributed.

a) Construct a 95% confidence interval for mean height of all NYP students.

b) With reference to the 95 % confidence interval, what is the maximum possible error of using the sample mean as an estimate of the population mean?

Case 2: Sample size $\geq$ 30 and Population Standard Deviation, $\boldsymbol{\sigma}$ unknown

Given : $s = 6.8$ n = 40 $\bar{x}$ = 173.8

X = height of NYP students

a) For a 95% confidence interval, $Z_c = 1.96$

95% confidence interval = $\bar{X} \pm Z_c \frac{s}{\sqrt{n}}$

$= 173.8 \pm 1.96 \frac{6.8}{\sqrt{40}}$

$= (171.69\ ,\ 175.91)$

The 95% confidence interval for mean height of NYP students is between 171.69 and 175.91 cm

b) Maximum possible error of using the sample mean = E, Margin of error

$= 1.96 \frac{6.8}{\sqrt{40}} = 2.11$

Case 2: Sample size $\geq$ 30 and Population Standard Deviation, $\boldsymbol{\sigma}$ unknown

Example 5 : The average waiting time of customers at a popular restaurant during peak hours is 30 minutes. The service quality manager of the restaurant implemented some time-saving measures and would like to know if the target of reducing customers' average waiting time by at least 10 minutes had been achieved.

35 customers were randomly selected and data on their waiting time at the restaurant during peak hours was recorded. The sample mean waiting time was 23.6 minutes with a standard deviation of 6.4 minutes.

Construct a 90% confidence interval for the population mean waiting time of customers after the time-saving measures were implemented.

Case 2: Sample size $\geq$ 30 and Population Standard Deviation, **σ** unknown

Given : $\mu = 30$ n = 35 $\bar{x}$ = 23.6 $s = 6.4$

X = waiting time for customer

For a 90% confidence interval, Z_c = 1.645

$$90\% \text{ confidence interval} = \bar{X} \pm Z_c \frac{s}{\sqrt{n}}$$

$$= 23.6 \pm 1.645 \frac{6.4}{\sqrt{35}}$$

$$= (21.82, 25.38)$$

The 90% confidence interval for mean waiting time for customer is between 21.82 and 25.38 min

Case 3: Sample size < 30 and Population Standard Deviation, $\boldsymbol{\sigma}$ unknown

- In many real-life situations, the population standard deviation σ is unknown.
- Moreover, due to constraints such as cost and time, it is often not practical to collect samples of size 30 or more.
- For such situations, we can use **t-distribution** provided the **random variable T is normally or approximately normally distributed** .
- Confidence Interval for a t-distribution = $\overline{\boldsymbol{X}} \pm$ E

$$= \overline{\boldsymbol{X}} \pm t_c \frac{s}{\sqrt{\boldsymbol{n}}}$$

where t is the critical value given by the t distribution with n-1 degrees of freedom

Case 3: Sample size < 30 and Population Standard Deviation, $\boldsymbol{\sigma}$ unknown

Example 6: 10 randomly selected people were asked how long they slept at night. The mean time was 7.1 hours and standard deviation 0.78 hours. Find with 95% confidence interval the mean time of sleep. Assume variable is normally distributed.

Case 3: Sample size < 30 and Population Standard Deviation, **σ** unknown

Given : $n = 10, \quad \bar{x} = 7.1, \quad s = 0.78$

X = time slept

For 95% level of confidence and degrees of freedom,n-1= 9, from the t-distribution table, t_c = 2.262

	Level of confidence, c	0.50	0.80	0.90	0.95
	One tail, α	0.25	0.10	0.05	0.025
d.f.	Two tails, α	0.50	0.20	0.10	0.05
1		1.000	3.078	6.314	12.706
2		0.816	1.886	2.920	4.303
3		0.765	1.638	2.353	3.182
4		0.741	1.533	2.132	2.776
5		0.727	1.476	2.015	2.571
6		0.718	1.440	1.943	2.447
7		0.711	1.415	1.895	2.365
8		0.706	1.397	1.860	2.306
9		0.703	1.383	1.833	$\boxed{2.262}$
10		0.700	1.372	1.812	2.228

95% confidence interval = $\bar{X} \pm t_c \frac{s}{\sqrt{n}}$

$= 7.1 \pm 2.262 \left(\frac{0.78}{\sqrt{10}}\right)$

$= 7.1 \pm 0.558$

$= (6.54, 7.66)$

The 95% confidence interval for mean duration of sleep is between 6.54 and 7.66 hours

Case 3: Sample size < 30 and Population Standard Deviation, $\boldsymbol{\sigma}$ unknown

Example 7: A retiree is doing a study on the one-year rate of return of investment funds in Singapore for the past year. A random sample of 15 investment funds was chosen to estimate the mean rate of return of investment funds. The performance of the sample was listed below. Find with 95% confidence interval, the population mean rate of return of investment funds for the past year. Assume the population is normally distributed.

Rate of Return (%) of Investment Funds for the Past Year				
2.33	4.67	-2.58	-8.1	-7.8
7.5	6.3	4.5	-2.9	2.5
10.2	-15.8	4.2	-16.1	6.8

Case 3: Sample size < 30 and Population Standard Deviation, $\boldsymbol{\sigma}$ unknown

Given : $n = 15$, using calculator, : $\bar{x} = -0.2853\%$
$s = 8.3098\%$

For 95% level of confidence and degrees of freedom,n-1= 14, from the t-distribution table, $t_c = 2.145$

At 95% confidence interval = $\bar{X} \pm t_c \frac{s}{\sqrt{n}}$

$= -0.2853 \pm 2.145\,(\frac{8.3098}{\sqrt{15}})$

= (-4.888,4.317)

	Level of confidence, c	0.50	0.80	0.90	0.95
	One tail, α	0.25	0.10	0.05	0.025
d.f.	Two tails, α	0.50	0.20	0.10	0.05
1		1.000	3.078	6.314	12.706
2		0.816	1.886	2.920	4.303
3		0.765	1.638	2.353	3.182
4		0.741	1.533	2.132	2.776
5		0.727	1.476	2.015	2.571
6		0.718	1.440	1.943	2.447
7		0.711	1.415	1.895	2.365
8		0.706	1.397	1.860	2.306
9		0.703	1.383	1.833	2.262
10		0.700	1.372	1.812	2.228
11		0.697	1.363	1.796	2.201
12		0.695	1.356	1.782	2.179
13		0.694	1.350	1.771	2.160
14		0.692	1.345	1.761	$\boxed{2.145}$

The 95% confidence interval for population mean rate of return of investment funds for the past year is between -4.888 %and 4.317 %

Minimum Sample Size to Estimate Population Mean μ

- Sometimes we will need to determine the sample size required before we conduct an experiment.
- We can find the minimum sample size as follows:

Margin of Error, $E = Z_c \frac{\boldsymbol{\sigma}}{\sqrt{\boldsymbol{n}}}$

$$E^2 = (Z_c \frac{\boldsymbol{\sigma}}{\sqrt{\boldsymbol{n}}})^2$$

$$E^2 = \frac{(Z_c \boldsymbol{\sigma})^2}{n}$$

$$n = (\frac{Z_c \sigma}{E})^2$$

Minimum Sample Size to Estimate Population Mean μ

- Example 8: A researcher wants to estimate the mean monthly salary of Admin staff in the private sector. He can tolerate a margin of error of $100 in estimating the mean. He would also prefer to report the interval estimate with a 95% level of confidence. It is reported by the Ministry of Manpower that the standard deviation of $1000. What is the required sample size?

For confidence interval 95%, $Z_c = 1.96$

Given E = 100, σ = 1000.

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

$$n = \left(\frac{1.96\ X\ 1000}{100}\right)^2$$

$$n = 384.16$$

For confidence interval 95%, a sample size of 385 (round up) is required.

Applications of Statistical Estimation

In business, decision makers usually have to make decisions without prefect information of the population mean values. They will have to

- ✓ Conduct business research.
- ✓ Take a random sample of customers to conduct a survey to solicit their feedback.
- ✓ Use the survey results (sample mean and sample standard deviation).
- ✓ Make reliable and precise estimates of the population mean.
- ✓ Make business decisions and policies.

Applications of Statistical Estimation

Example 9: The average waiting time of customers at a popular restaurant during peak hours is 30 minutes. The service quality manager of the restaurant implemented some time-saving measures and would like to know if the target of reducing customers' average waiting time by at least 10 minutes had been achieved.

35 customers were randomly selected and data on their waiting time at the restaurant during peak hours was recorded. The sample mean waiting time was 23.6 minutes with a standard deviation of 6.4 minutes.

Construct a 90% confidence interval for the population mean waiting time of customers after the time-saving measures were implemented.

a) Explain if the target of the service quality manager of the restaurant had been achieved.
b) Suggest how the service quality manager should respond to the results obtained.

CPY2

Applications of Statistical Estimation

a) Based on the solution from Example 5, the 90% confidence interval for mean waiting time for customer is between 21.82 and 25.38 min. Therefore, the target of reducing customers' average waiting time from 30 minutes by at least 10 minutes has **not** been achieved.

b) The Service Quality Manager should review the time-saving measures implemented by relooking at the business processes to see where reduction of waiting can be reduced.

Slide 28

CPY2 CHEONG POH YEE, 3/7/2019

Applications of Statistical Estimation

Example 10: A retiree is doing a study on the one-year rate of return of investment funds in Singapore for the past year. A random sample of 15 investment funds was chosen to estimate the mean rate of return of investment funds. The performance of the sample was listed below. Find with 95% confidence interval, the population mean rate of return of investment funds for the past year.

Rate of Return (%) of Investment Funds for the Past Year				
2.33	4.67	-2.58	-8.1	-7.8
7.5	6.3	4.5	-2.9	2.5
10.2	-15.8	4.2	-16.1	6.8

Explain if the following type of investors should invest in investment funds in Singapore.

a) A risk-averse investor who does not want to incur a loss in his investment.

b) A 30 year-old professional wants to invest his CPF Special Account balance, which currently pays 6% guaranteed interest.

Applications of Statistical Estimation

From Example 7, the 95% confidence interval for population mean rate of return of investment funds for the past year is between -4.888 % and 4.317 %

a) The 95% confidence interval for the population mean rate of return for investment funds in Singapore is between -4.89% and 4.32%. Hence, there is a possibility of a loss of up to -4.89%. A risk-averse investor should not invest in the fund.

b) For the professional, his CPF Special Account is already paying 6% guaranteed interest which is higher than the mean rate of the investment fund (between -4.8876% and 4.3170%). Thus, he should not invest in investment fund in Singapore.