13/04/23

# Statistics Day-5

## Inferential Statistics

1. Hypothesis Testing → Z test, t test, Chi square test, Annova test (F-Test)
2. P. Value
3. Confidence Interval
4. Significance Value

## Distribution

1. Bernaulli Distribution
2. Binomial Distribution
3. Power law.

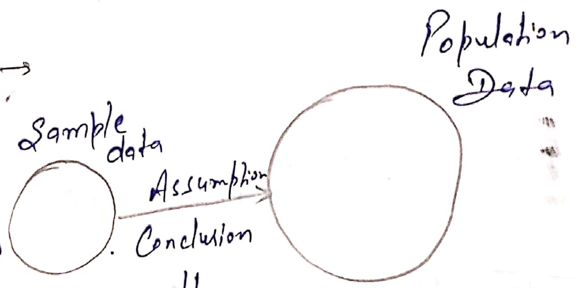
} TRANSFORMATIONS

---

13/04/23

## 1. Hypothesis Testing :→

### Inferential Statistics :→




⇓

Hypothesis Testing.

We have sample data and we make assumptions or conclusions about the population data using sample data. In order to validate the Assumptions/Conclusions we use something called Hypothesis Testing.

### Steps of Hypothesis Testing :→

Eg:- Criminal Case

1. Null Hypothesis :→ (Default statement) i.e the person is not guilty untill it is proven in court. Hence by default that Person is not guilty is our Null Hypothesis.

Person is not a criminal. and by default it's always TRUE.

**Example :→** [Coin is fair or not ?]

How to check ?

$\boxed{P(H) = 0.5 \quad P(T) = 0.5}$ ⇐ iff it satisfies this condition then we can say coin is fair.

Hence

$\boxed{\text{Null Hypothesis : Coin is fair.}}$

② **Alternate Hypothesis :→** It will always be the opp. of Null Hypothesis.

$\boxed{\text{Alternate Hypothesis: Coin is not fair.}}$

$\boxed{\text{Alternate Hypothesis: Person has } \cancel{\text{not}} \text{ committed the crime.}}$

③ **Perform Experiments :→** We perform some experi-ents to proof that whether the coin is fair or not.

or

The person has committed the crime or not?

**What kind of Experiment ?**

Toss the coin → $\boxed{100 \text{ times,}}$

⇓

$\boxed{50 \text{ times Head}}$ ⇒ fair

$\boxed{60 \text{ times Head}}$ ⇒ fair

70 times Head $\xRightarrow{\text{Ask}}$ Domain Expert

⇓ give Range

Confidence Interval

| H | T | |
|---|---|---|
| 60 | 40 | → fair |
| 70 | 30 | → fair |
| 80 | 20 | → not fair |

C.I = [20-80] ← Range

⇓

Coin is fair.

10 20 30 40 50 60 70 80 90

Confidence Interval will be decided by the domain Expert.

C.I → Confidence Interval

(If the value lies b/w this particular interval means the coin is fair)

Q. If we get Head 75 times out of 100 then,

(i) Null Hypothesis :- True ⇒ Accepted

↓

Since 75 lies inside the Confidence Interval.

Q. If we get Head 10 times out of 100 times.

(i) Null Hypothesis :- Null Hypothesis is Rejected.

(ii) Alternate Hypothesis : Accepted.

or

Statement

Conclusion

$\boxed{\text{We fail to Reject the Null Hypothesis.}}$

when it falls $\boxed{\text{inside the Confidence Interval (C.I)}}$

$\boxed{\text{We Reject the Null Hypothesis.}}$

When it falls $\boxed{\text{outside the C.I}}$

(2) Person is criminal or not? { Murder Case }

Steps.

(i) Null Hypothesis: Person is not a criminal.

(ii) Alternate Hypothesis: Person is a criminal.

(iii) Experiment / Proof : DNA, Finger Point, weapon, eye witness, footage.

⇓ Based on Proofs

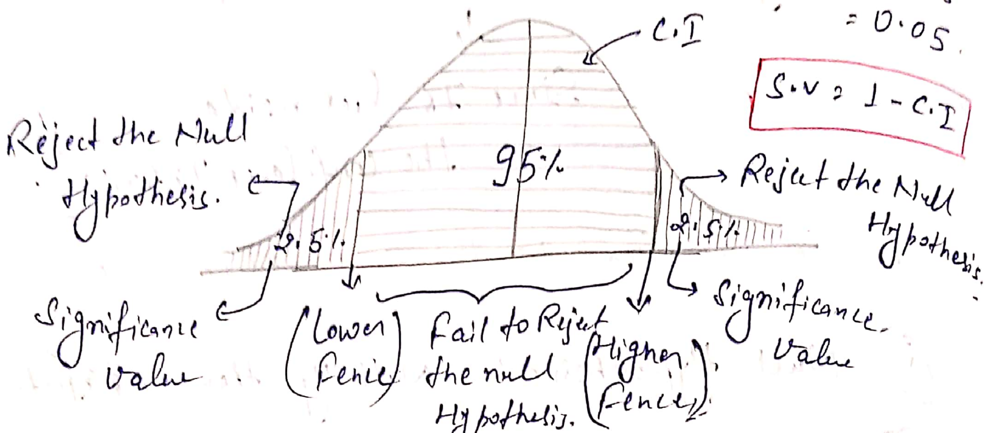
Judge ⇒ Conclusions.

# Confidence Interval (C.I) ← Domain Expert · Significance Value

C.I = 95%

S.V = 1 - 0.95
= 0.05

$\boxed{S.V = 1 - C.I}$




**Scenario :->** In Case Suppose a Doctor is working on Some of the Vaccines. Then the test which is going to ~~In the~~ happen in Vaccine for that the C.I will be very small.

Vaccine ⇒ how is it affecting People
⇓ C.I will be very-very less.
~~Critical~~ Medical — Since they can perform experiments on very few people
⇓
Critical

## # How to find the Lower fence & Higher fence?

⇒ Point Estimate :-> The Value of any statistics that estimates the value of a parameter is called Point Estimate.

Point estimates.
⇑
$\boxed{\bar{x}} \xrightarrow{\text{estimates}} \boxed{\mu}$

Statistic ↓ $\boxed{\bar{x}}$ (↑ Sample mean) $\xrightarrow{\text{estimates}}$ Parameter ↓ $\boxed{\mu}$ (↑ Population mean)

$$\boxed{\text{Point Estimates}} \pm \boxed{\text{Margin of Error}} = \boxed{\text{Parameter i.e Population Mean}}$$

Lower C.I = Point Estimate − Margin of Error

Higher C.I = Point Estimate + Margin of Error.

where, $\boxed{\text{Margin of Error} = Z_{\alpha/2} \boxed{\frac{\sigma}{\sqrt{n}}}}$ ⇒ Standard Error

$\sigma$ → Population SD

$\alpha$ ⇒ Significance Value

Q. Why Sample mean is Called as Point Estimate?

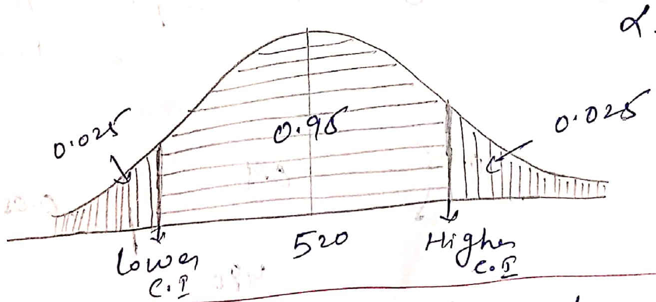
Sol: Since through the Sample mean we estimates the Value of the Population Mean

Q. On the Quant test of CAT Exam, a sample of $\boxed{25 \text{ test}}$ takers has a mean of $\boxed{520}$ with a population Standard deviation of $\boxed{100}$. Construct a $\boxed{95\% \text{ C.I}}$ about the mean?

Sol: $n = 25$, $\bar{x} = 520$, $\sigma = 100$, C.I = 95%

S.V = 1 − 0.95 = 0.05

$\alpha = 0.05$

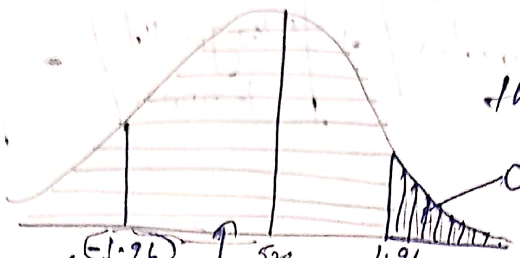



$\boxed{\text{Lower C.I} = \text{Point Estimation} - \text{Margin of Error}}$

$= 520 - Z_{0.05/2}\, \sigma/\sqrt{n}$

$= 520 - Z_{0.025} \frac{100}{\sqrt{25}}$

$= 520 - 1.96 \times \frac{100}{5} = 480.8$




$1 - 0.025 = \boxed{0.975}$ Area under the Curve

-1.96 means the lower C.I is -1.96 away from the mean

Now check in z table where is 0.975.

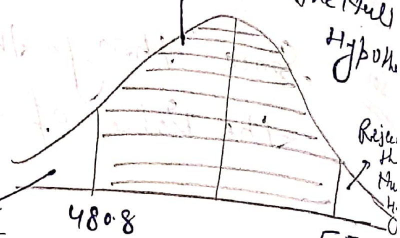
1.96.

Higher C.I $= 520 + Z_{0.05/2} \frac{\sigma}{\sqrt{n}}$

$= 520 + Z_{0.025} \frac{100}{\sqrt{25}}$

$= 520 + 1.96 \times 20$

$= 559.2.$

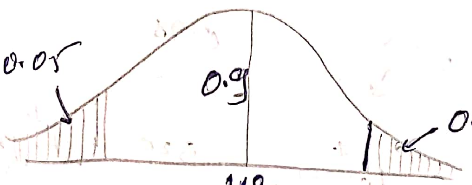



Within this range Whatever Values Comes will be accepted

Q. $\boxed{\bar{x} = 480}$ (Sample mean), $\boxed{\sigma = 85}$ (Population SD), $\boxed{n = 25}$ (Sample size), $\boxed{C.I = 90\%}$

$\alpha = 1 - 0.9 = 0.1.$

Sol$^n$



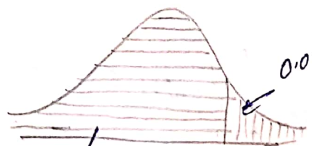

Lower C.I $= 480 - Z_{0.1/2} \frac{\sigma}{\sqrt{n}}$

$= 480 - Z_{0.05} \frac{85^{17}}{5}$

$= 480 - 1.64 \times 17$

$= 480 - 27.8 = \boxed{452.12}$

$\boxed{Z_{0.05} = 1.64}$



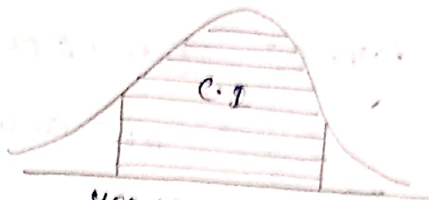

$1 - 0.05 = 0.95$

$\boxed{Z_{0.95} = 1.64}$

Higher C.I $= 480 + 1.64 \times 17$

$= 480 + 27.8$

$= 507.8$




452.12 507.8

$\underbrace{\qquad\qquad}_{\text{Range of C.I.}}$

Hence C.I $= [452.12, 507.8]$

Q. On the Quant test of CAT exam, a sample of $\boxed{25}$ test taken has a mean of $\boxed{520}$, with a sample standard deviation of $\boxed{80}$. Construct $\boxed{95\%}$ C.I about the mean?

Sol^n^ $\boxed{n = 25}$, $\boxed{\bar{x} = 520}$, $\boxed{s = 80}$ → sample SD $\boxed{C.I = 0.95}$

$\alpha = 1 - 0.95 = 0.05$.

Since Population SD not given. In that case we need to use t-table. or If Sample SD is given we use t-table.

Formula

$$\boxed{\bar{x} \pm t_{\alpha/2}\left(s/\sqrt{n}\right)}$$

Whenever we perform t-test, we need to find degree of freedom f

$\boxed{\text{degree of freedom} = n-1}$

Here, degree of freedom $= 25 - 1 = \boxed{24}$

Since last person has no choice all over 2 options

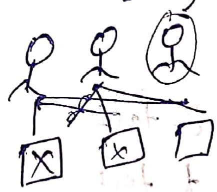
T-test (since n<30 & 's' is given)

Lower C.I $= 520 - t_{0.05/2}\left(\frac{80}{\sqrt{25}}\right)$

$= 520 - t_{0.025}\left(80/5\right)$ (16)

$= 520 - 2.064 \times 16$

$= 486.976.$



1st Person - 3 options

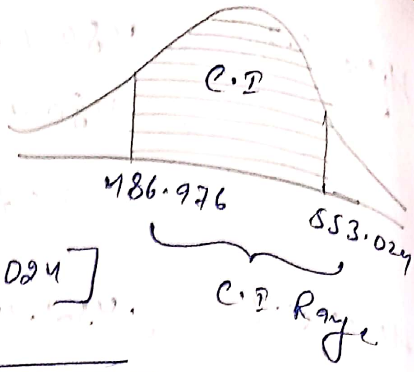
2nd Person - 2 options

3rd Person = 0 options

Total = (n-1) options for degree of freedom

Higher C.I = $520 + 2.064 * 16$

$= 553.024$



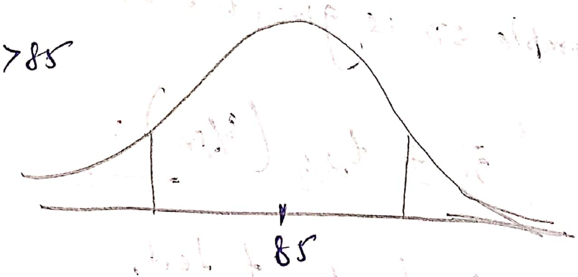

C.I = $[486.976, 553.024]$

# 1 Tail and 2 Tail Test

Q. College in Town A has $\boxed{85\%}$ Placement rate. A new College was recently opened and it was found that a sample of $\boxed{150}$ students had a placement rate $\boxed{88\%}$ with a standard deviation of 4%. Does this college has a different placement rate with 95% CI.

Sol^n

it could be >85 or <85 Hence it is a 2-tail Problem.




~~iff >85%~~

iff greater than 85% then → 1 tail Problem i.e (right side)

iff less than 85% then → 1 tail problem i.e (left side)

1. Z-test
2. t-test

# Hypothesis Testing Problem

Could be > 80ml or < 80ml Hence 2-tail test

Q. A factory has a machine that fills $\boxed{80ml}$ of Baby medicine in a bottle. An employee believe the avg. amt. of baby medicine is $\boxed{\text{not } 80ml}$. Using $\boxed{40 \text{ samples}}$, he measures the avg. amount dispensed by the machine to be $\boxed{78ml}$ with standard deviation of $\boxed{2.5}$

(a) State Null & Alternate Hypothesis.

(b) At 95% CI, is there enough evidence to support machine is working properly or not.

$\boxed{\mu = 80ml, \ n = 40, \ \bar{x} = 78, \ s = 2.5}$

Sol:

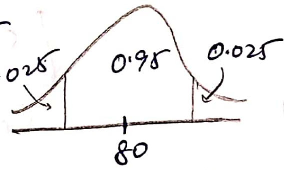
**Step 1**

Null Hypothesis : $\mu = 80$

Alternate Hypothesis : $\mu \neq 80$

**Step 2:**

$C.I = 0.95, \quad S.V(\alpha) = 1 - 0.95 = 0.05$




**Step 3**

any one condition should be satisfied.

Use

(i) $\boxed{n \geq 30 \text{ or Population SD}}$ → Z-Test

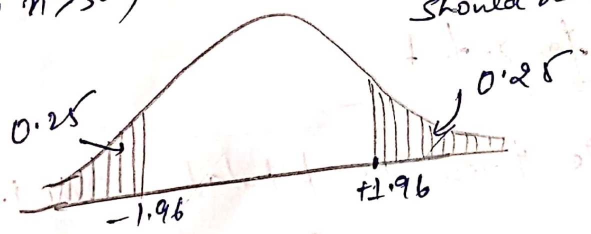
$n = 40$

$s = 2.5$

(ii) $\boxed{n \leq 30 \text{ and Sample SD}}$ → t-test

Both condition should be satisfied

Z-Test (Since $n > 30$)

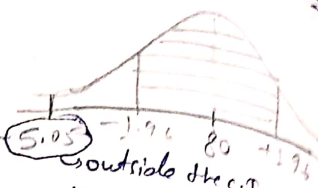



Decision Boundary: $[-1.96, 1.96]$.

## (*) Calculate test statistics (z-test)

$$\boxed{Z\text{-Score} = \frac{\bar{x} - \mu}{(s/\sqrt{n})}} = \frac{78 - 80}{2.5/\sqrt{40}} = \frac{-2 \times 2 \times \sqrt{10}}{2.5} = \boxed{-5.05}$$

$\underbrace{(s/\sqrt{n})}_{\text{Standard Error.}}$




## (*) Conclusions:

Decision Rule: If $z = -5.05$ is less than $-1.96$ & greater than ~~+1.96~~, Reject the Null Hypothesis with 95% C.I.

Reject the Null Hypothesis { There is some fault in the machine.

**Q.** A Complain was registered, the boys in a Government School are underfed. Avg. weight of the boys of age $\boxed{10}$ is $\boxed{32 \text{ kgs}}$ with $\boxed{S.D = 9 \text{ kgs}}$. A sample of $\boxed{25 \text{ boys}}$ were Selected from the government School and the avg. weight was found to be $\boxed{29.5 \text{ kgs}}$ with $\boxed{C.I = 95\%}$ Check it is true or false.

< 32 1-Tail test

**Sol^n**

$$\boxed{n = 25,\quad \mu = 32,\quad \sigma = 9,\quad \bar{x} = 29.5,\quad C.I = 0.95}$$

$\alpha = 1 - 0.95 = 0.05$

### Condition for z-test

(i) We know the population SD i.e $\sigma$

or

(ii) We don't know the Population SD but our Sample is large i.e $n \geq 30$.

Condition for T-test

1. We don't know the population S.D i.e '$\sigma$'
2. Our sample size is small i.e $n < 30$
3. Sample S.D i.e 's' is given

step 1.

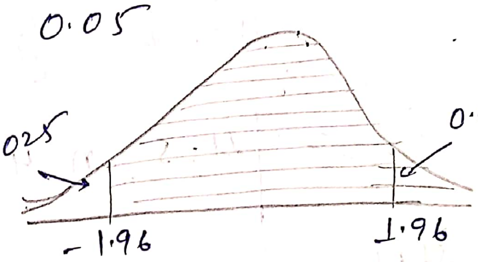
$H_0 : \mu = 32$ kg. (Null Hypothesis)

$H_1 : \mu \neq 32$ (Alternate Hypothesis)

2. C.I = 0.95 $\alpha = 1 - 0.95 = 0.05$

~~1-Tail test~~

3. Z test (since $n > 30$ & '$\sigma$' is given)



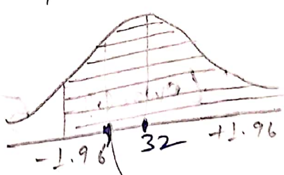

$$\boxed{Z\text{-score} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}}$$

$$Z\text{-score} = \frac{29.5 - 32}{9/\sqrt{25}} = \frac{-2.5 \times 5}{9} = -1.39.$$

Conclusion: $-1.39 > -1.96$ ∴ Accept the Null Hypothesis.

With 95% C.I. We fail to Reject the Null Hypothesis.

The Boys are feel well




(it is inside C.I)

Assignment Q-1

Hypothesis. 1-tail test.

**Q.1** A factory manufactures cars with a warranty of 5 years or more on the engine and transmission. An engineer believes that the engine or transmission will malfunction in less than 5 years. He tests a sample of 40 cars and finds the avg. time to be 4.8 years with a standard deviation of 0.50.

(a) State the Null & alternate Hypothesis.

(b) At a 2% significance level, if there is enough evidence to support the idea that the warranty should be revised?

**Sol^n**

$$\boxed{\mu = 5,\ n = 40,\ s = 0.50,\ \alpha = 0.02,\ \bar{x} = 4.8 \quad C.I = 1 - 0.02 = 0.98}$$

**Step 1**

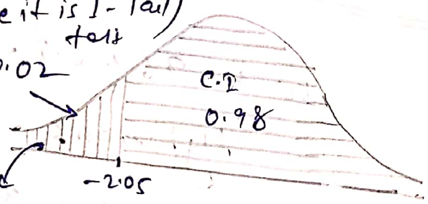
$H_0 : \mu \geq 5$ {Null Hypothesis}

$H_1 : \mu < 5$ {Alternate Hypothesis}

**Step 2**

$\alpha = 0.02$, $C.I = 1 - 0.02 = 0.98$

**Step 3** Z-Test (Since $n > 30$) (Since it is 1-Tail test)




Reject the null Hypothesis

$$\cancel{\text{Lower C.I} = 5 + z_{\alpha}\left(\frac{s}{\sqrt{n}}\right)}$$

$$\cancel{5 + z_{0.02}\left(0.50\right)}$$

**Step 4**

$$\text{Lower C.I} = \bar{x} - z_{\alpha}\left(\frac{s}{\sqrt{n}}\right)$$

$$= 4.8 - z_{0.02}\left(\frac{0.50}{\sqrt{40}}\right)$$

$$= 4.8 - \frac{2.05 \times 0.50}{\sqrt{40}}$$

$$4.8 - 0.1620 = 4.638.$$

Scanned with OKEN Scanner

Steps

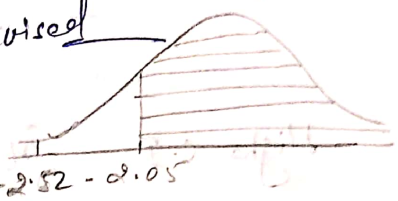
$$Z\text{-score} = \frac{\bar{x}-\mu}{S/\sqrt{n}}$$

$$\frac{4.8-5}{0.50/\sqrt{40}} = \frac{-0.2\times\sqrt{40}}{0.50} = -2.52$$

Conclusion

$-2.52 < -2.05$ Hence Reject the Null Hypothesis.

Hence Warrenty needs to be Revised




lies outside the confidence interval.

Assignment Q-2

Q2 In the population the avg. IQ is $\boxed{100}$ with a standard deviation $\boxed{15}$. A team of scientist wants to test a new medicine to see if it has a +ve or -ve effect, or no effect at all.

A Sample of $\boxed{30}$ Participants who have taken the medicine has a mean of $\boxed{140}$. Did the medicine affect intelligence. $\boxed{C.I = 95\%}$

2-Tail test

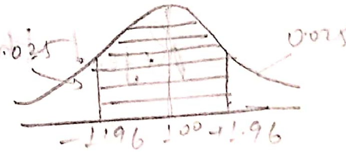
Sol^n

$\mu = 100,\ \sigma = 15,\ n = 30,\ \bar{x} = 140,\ C.I = 0.95$

Step 1 $H_0: \mu = 100$ (null Hypothesis)

$H_1: \mu \neq 100$ (Alternate Hypothesis)

Step 2: $C.I = 0.95$, $S.V(\alpha) = 1-0.95 = 0.05$




Step 3 Z-Test (since $n=30$ & $\sigma$ is given)

Lower C.I $= \bar{x} - Z_{\frac{\alpha}{2}}\left(\sigma/\sqrt{n}\right)$

$= 140 - Z_{\frac{0.05}{2}}\left(15/\sqrt{30}\right)$

$= 140 - Z_{0.025} \times 15/\sqrt{30}$

$$= 140 - 1.96 \times \left(15/\sqrt{30}\right)$$

$$= 140 - 5.367$$

$$= 134.632$$

$$\text{Higher C.I} = \bar{x} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$$= 140 + Z_{\frac{0.05}{2}} \left(\frac{15}{\sqrt{30}}\right)$$

$$= 140 + 1.96 \times 15/\sqrt{30}$$

$$= 145.367$$

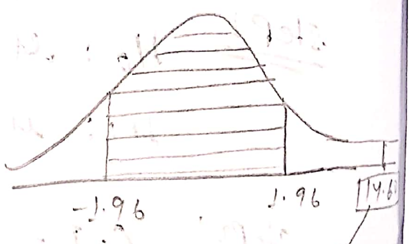
Step 4

$$\boxed{\text{Z-score} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}}$$

$$\text{Z-score} = \frac{140 - 100}{15/\sqrt{30}} = \frac{40 \times \sqrt{30}}{15} = 14.60$$

Step 5

Conclusion:




14.60 > 1.96 Hence we Reject the null Hypothesis.

It lies outside the C.I