STATISTICS FOR DATA SCIENCE

Confidence Intervals for Large Samples

Prof. Uma D

Department of Computer Science and Engineering

STATISTICS FOR DATA SCIENCE

Confidence Intervals for Large Samples

Prof. Uma D

Topics to be covered...

- Confidence Intervals
- Confidence Intervals for population mean of large samples
- Confidence Levels
- Confidence Co-efficient
- Probability Vs Confidence
- One sided confidence intervals
- Confidence Intervals for Proportions

≥ 30

$n \geq 30$

$n > 30$

$n < 30 \Rightarrow$ Small sample

STATISTICS FOR DATA SCIENCE

Confidence Intervals

1, 2, 3, 4, . . . 10

$N = 10$

$\mu = 5.5$

S_1 1, 3 sample $n = 2$ $\bar{X}_1 = 2$

Single ~~Point~~ value

S_2 1, 5 $\bar{X}_2 = 3$

$\bar{X} = 3 \Rightarrow$ population mean may be 3

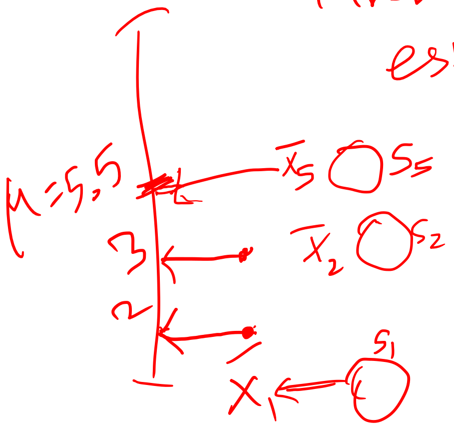
S_3 2, 2 $\bar{X}_3 = 2$

S_4 2, 3 $\bar{X}_4 = 2.5$

S_5 5, 6 $\bar{X}_5 = 5.5$

× $\mu = 3$

Through a point estimate, we will try to hit the population parameter value.



We may or may not hit the population parameter value through point estimate.

However, Interval Estimate may cover population Parameter value

Confidence Interval

Interval contains true parameter value μ.

Confidence Intervals

Population

1, 3, 5, 7

N=4

μ=4

Single value

Point Estimate

Sample → sample mean $\bar{x}$ → Estimate value for μ

Population

μ=?

p=?

$\hat{p} = \frac{x}{n} = \frac{32}{50} = .64$ → p = ? 100 CIs

Statistic s

100 different random samples

Estimate parameter

Confidence Level

90% or 95% 99%

Range of plausible values for μ.

μ=4

(1,7) $\bar{x}=4$

Single

(1,3) $\bar{x}=2$

(1,1) $\bar{x}=1$

(1,1) n=2 $\bar{x}=1$

(1,3) $\bar{x}=2$

(3,3) $\bar{x}=3$

(1,5) $\bar{x}=3$

(1,7) $\bar{x}=4$

Net

covers μ

$\hat{p}_1$ $\hat{p}_2$ $\hat{p}_3$ $\hat{p}_4$ $\hat{p}_5$

p

5% 5 CIs may not cover μ

LB $\bar{x}$ UB

Estimating an Interval

μ=4

3

2

Interval

$\bar{x}=3$

$\bar{x}=2$

100%

STATISTICS FOR DATA SCIENCE

Confidence Intervals

One sample ⇒ 95%
98%
90%

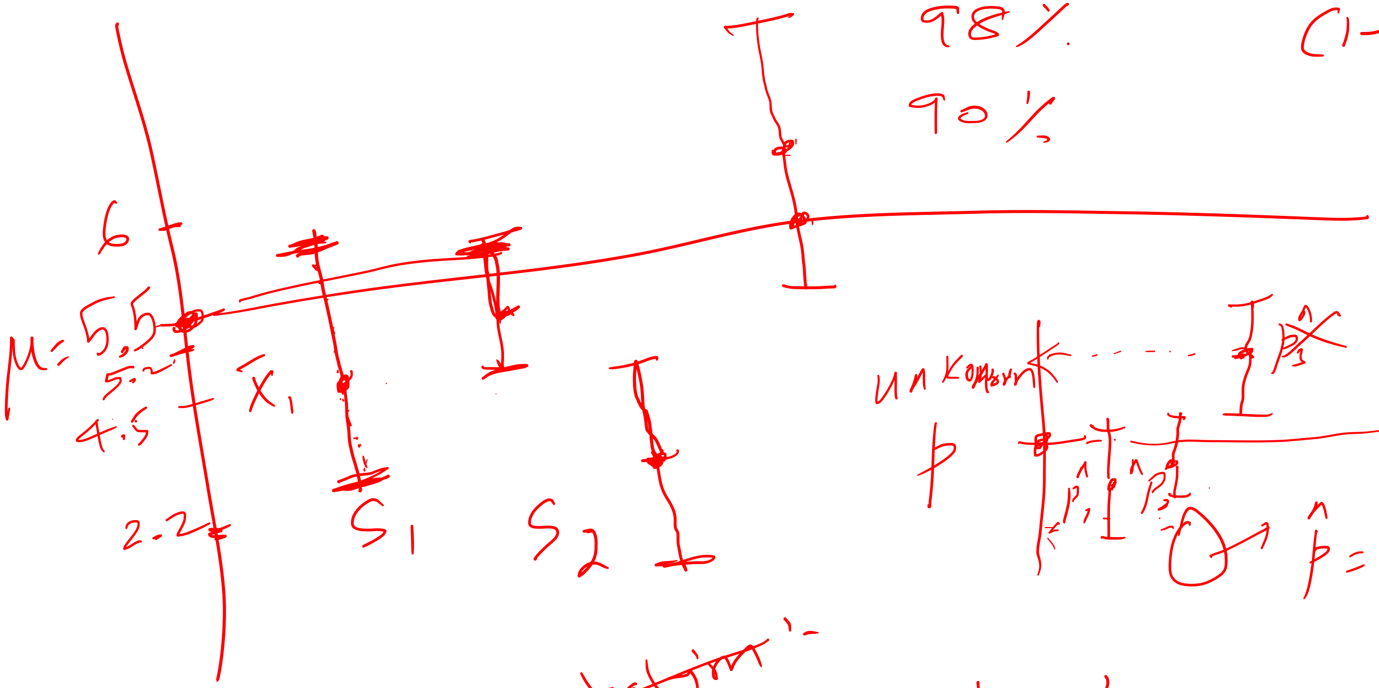
Confidence Level

$(1-\alpha)*100 \Rightarrow 95\% \Rightarrow 0.95$

$98\% \Rightarrow 0.98$

$90\% \Rightarrow 0.90$

Confidence Coefficiency



Significance Level (α)

$\alpha = 5\% \Rightarrow 0.05$

$\alpha = 2\% \Rightarrow 0.02$

$\alpha = 10\% \Rightarrow 0.1$

$1-\alpha \Rightarrow$ confidence coefficient

~~Interpretation~~:

With 95% CL we make certain conclusions.

STATISTICS FOR DATA SCIENCE

Confidence Intervals

100 samples

95 CI covers parameter

5 CIs doesn't cover parameter.

$\bar{X} = \frac{\sum x_i}{n}$ 95%

$\hat{p} = \frac{X}{n}$

μ

1 MoE

$\bar{X}$

LB 2-sided CI UB

$\bar{X} - MoE$ $\bar{X}$ $\bar{X} + MoE$

$(\bar{X} - MoE, +\infty)$

2 am

1-sided CI

LB $+\infty$

Statistic

$\hat{p} = \frac{X}{n}$

σ p p $\hat{p}$ X

1-sided $(-\infty, \bar{X} + MoE)$

$-\infty$ UB

MoE ?

Confidence Intervals

Unknown Parameter (population) | Sample Statistic (sample) | CI

$\mu \leftarrow \bar{X} \pm$ Margin of Error $[\bar{X} - MoE, \bar{X} + MoE]$

Lower Bound, Upper Bound

p | $\hat{p} \pm$ Margin of Error | $[\hat{p} - MoE, \hat{p} + MoE]$

σ | $s \pm$ Margin of Error | $[s - MoE, s + MoE]$

Two Sided Confidence Interval: LB, UB

LB, $+\infty$: $[LB, +\infty)$

$-\infty$, UB: $(-\infty, UB]$

$\bar{X} \Rightarrow \mu \neq \bar{X} \Rightarrow [\bar{X} - MoE, \bar{X} + MoE]$

$\bar{X} - MoE \le \mu \le \bar{X} + MoE$

STATISTICS FOR DATA SCIENCE

Confidence Intervals

Unknown

μ

Sample Statistic

$\bar{X} \pm MoE$

Confidence Interval

$[\bar{X} - MoE, \ \bar{X} + MoE]$

$\bar{X} \pm \boxed{Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}} \rightarrow$ uncertainty of sampling distribution $\rightarrow$ MoE

Two-Sided
Two-Tailed

Confidence Level $= (1-\alpha)*100$

$\mu \Leftarrow \bar{X} \pm \boxed{Z_{\alpha/2} * \frac{s}{\sqrt{n}}}$

$\bar{x} \leftarrow \mu_1$

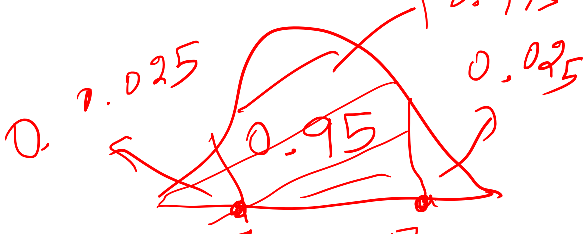
$\alpha = 0.05$
$1-\alpha = 0.95$ $\Rightarrow$ 95% CL

For 95% CL $\Rightarrow$ MoE $= Z_{\alpha/2} * \frac{s}{\sqrt{n}} =$

1.96

90% CL $\Rightarrow$ MoE $=$ For area 0.975 $\Rightarrow$ $Z = 1.96$

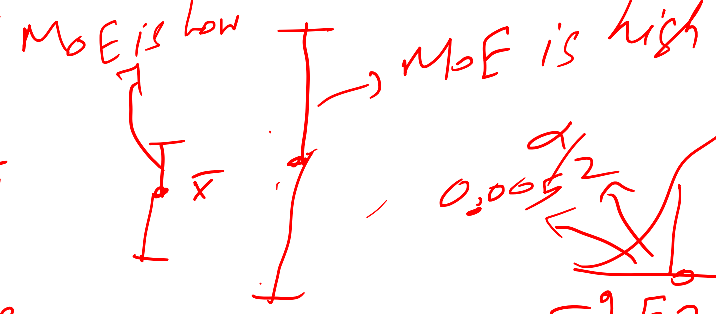
$\alpha = 0.05$



$1 - 0.025 = 0.975$

For 95% CL $\Rightarrow$ $MoE = Z_{\alpha/2} * \frac{s}{\sqrt{n}} = 1.96 * \frac{s}{\sqrt{n}}$

$CI : [\bar{x} - MoE, \; \bar{x} + MoE]$



$\alpha = 0.01$

$\alpha/2 = 0.005$

For 99% CL $\Rightarrow$ $MoE = Z_{\alpha/2} * \frac{s}{\sqrt{n}}$

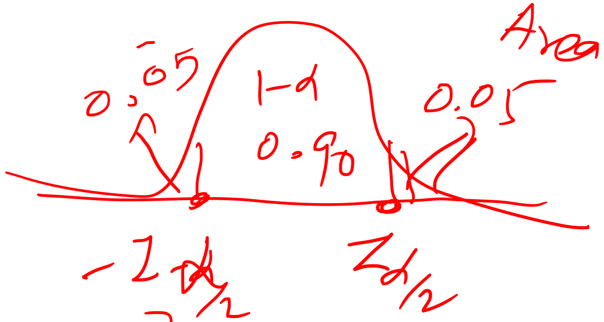
$= 2.57 * \frac{s}{\sqrt{n}}$

Area to the right of Z = 0.005

Area to the left of Z $= 1 - 0.005$
$= 0.995$

For 90% CL $\Rightarrow$ $MoE = Z_{\alpha/2} * \frac{s}{\sqrt{n}}$

$= 1.645 * \frac{s}{\sqrt{n}}$



Confidence Interval for μ $\Rightarrow$ $[\bar{x} - MoE, \; \bar{x} + MoE]$

Two-Sided Confidence Interval for μ (in case of Large Population)

CL	α	$\alpha/2$	$Z_{\alpha/2}$
95% 0.95	5% 0.05	2.5% 0.025	1.96
98% 0.98	2% 0.02	1% 0.01	2.33
99% 0.99	1% 0.01	0.5% 0.005	$Z = -2.57$ or $Z = -2.575$
80% 0.80	20% 0.20	10% 0.10	$Z = -1.28$ or $Z = 1.28$

Confidence Interval

$$\mu \in \bar{x} \pm 1.96\left(\frac{\sigma}{\sqrt{n}}\right)$$

$$\bar{x} \pm 2.33\left(\frac{\sigma}{\sqrt{n}}\right)$$

$$\bar{x} \pm 2.575\left(\frac{\sigma}{\sqrt{n}}\right)$$

$$\bar{x} \pm 1.28\left(\frac{\sigma}{\sqrt{n}}\right)$$

Confidence Intervals

PES UNIVERSITY ONLINE

Large Sample ($n \geq 30$)

Two-Sided CI (Point Estimate ± MoE)

CI for μ: $\bar{X} \pm MoE$ (or) $\bar{X} \pm \boxed{Z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}}$ MoE

Critical value

$(\bar{X} - MoE, \bar{X} + MoE)$

For 95%,

For 98%

For 90%

99%

80%

CI for p: $\hat{p} \pm \underbrace{Z_{\alpha/2} * \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}_{MoE}$

Small Sample ($n < 30$) (student's t-distrn table)

CI for μ: $\bar{X} \pm MoE$ $\quad \bar{X} \pm t_{n-1, \alpha/2} * \frac{s}{\sqrt{n}}$

$n-1$ (df)

One-Sided CI

$\left(-\infty, \bar{X} + Z_{\alpha} * \frac{s}{\sqrt{n}}\right)$ Upper Bound — Lower Interval

$\left(\bar{X} - Z_{\alpha} \cdot \frac{s}{\sqrt{n}}, +\infty\right)$ LB — Upper Interval

$\left(-\infty, \hat{p} + Z_{\alpha} * \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$

$\left(\hat{p} - Z_{\alpha} * \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, +\infty\right)$

$\left(-\infty, \bar{X} + t_{n-1,\alpha} * \frac{s}{\sqrt{n}}\right)$

$\left(\bar{X} - t_{n-1,\alpha} * \frac{s}{\sqrt{n}}, +\infty\right)$

PES UNIVERSITY ONLINE

Two-Sided CI for μ

$$\bar{x} \pm Z_{\alpha/2} * \frac{s}{\sqrt{n}}$$

$\alpha = 10\%$

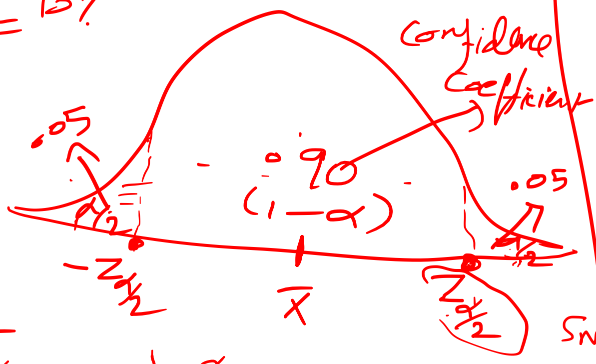
CL $(1-\alpha) * 100\%$ CI

For 90%.

$(1-\alpha) * 100\% \Rightarrow$ CL

$.90 * 100\% \Rightarrow 90\%$ (CL)

Point Estimate: $\bar{x}$



$1-\alpha = .9$

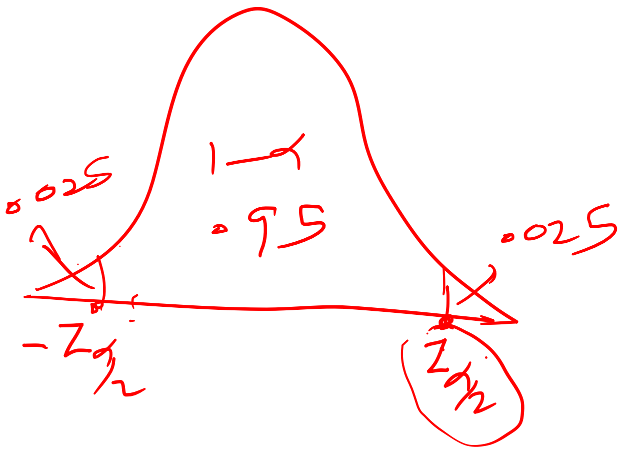
$\alpha = .1 \Rightarrow 10\%$

$\alpha/2 = .05$

For 95%

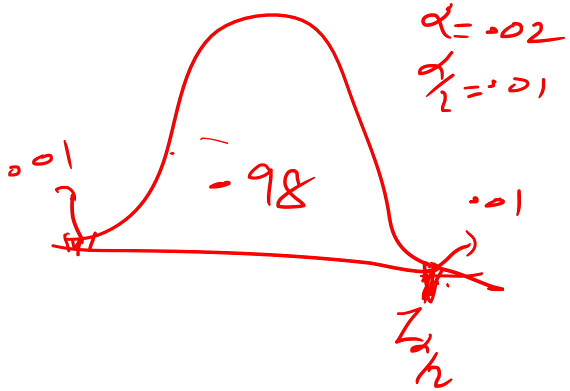
$\alpha = .05$

$\alpha/2 = .025$



$\alpha = .02$

$\alpha/2 = .01$



One-Sided CI for μ

$$\left(\bar{x} - Z_{\alpha} \cdot \frac{s}{\sqrt{n}}, +\infty\right)$$

$$\left(-\infty, \bar{x} + Z_{\alpha} \cdot \frac{s}{\sqrt{n}}\right)$$

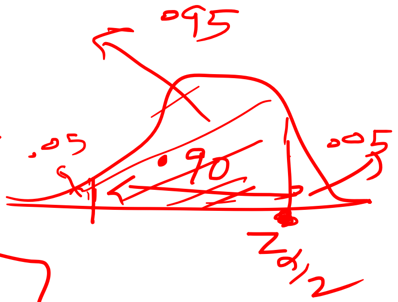
For 90% → LB

→ UB

$\alpha = .10$

Confidence Intervals

CL	$1-\alpha$	α	$\frac{\alpha}{2}$	Left Area	$Z_{\frac{\alpha}{2}}$
For 90%	.90	.1	.05	$\boxed{.95}$	1.645
For 95%	.95	.05	.025	$\boxed{.975}$	$\boxed{1.96}$
For 98%	.98	.02	.01	$\boxed{.99}$	2.33



For 90% .90 .1 Left Area .9 Z_{α} 1.28

Two Sided Confidence Intervals: Example

In a sample of 100 wires the average breaking strength is 50kN, with a standard deviation of 2kN.

$n = 100$
$\bar{x} = 50$
$s = 2$

a) Find **68%** confidence interval for the mean breaking strength of this type of wire.
b) Find **80%** confidence interval for the mean breaking strength of this type of wire.
c) Find **90%** confidence interval for the mean breaking strength of this type of wire.
d) Find 95% confidence interval for the mean breaking strength of this type of wire.
e) Find **99%** confidence interval for the mean breaking strength of this type of wire.

$(1-\alpha) * 100\%$ CI is given by $\bar{x} \pm MoE$ or $\left(\bar{x} - Z_{\alpha/2} * \frac{s}{\sqrt{n}}, \bar{x} + Z_{\alpha/2} * \frac{s}{\sqrt{n}}\right)$

95% CI is given by $\bar{x} \pm Z_{\alpha/2} * \frac{s}{\sqrt{n}}$

$50 \pm 1.96 * \frac{2}{\sqrt{100}}$

$50 \pm .392$

$\therefore$ 95% CI is $(50 - .392, 50 + .392) \Rightarrow (49.608, 50.392)$

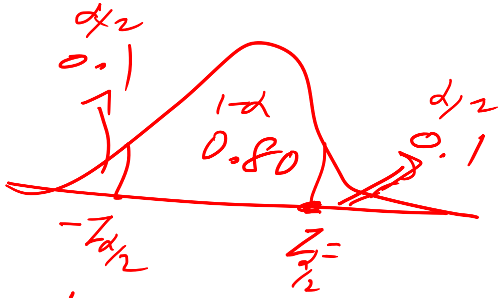
Given $n = 100$, $\bar{x} = 50$, $s = 2$

Construct CI for the given confidence level.

Case 1)

$CL = 80\%$ $\alpha = 20\%$

$\alpha = 0.2$ $\alpha/2 = 0.1$



Area close to 0.1 is 0.1008

corresponding z-score is $Z = -1.28$

$$MoE = Z_{\alpha/2} * \sigma/\sqrt{n} = Z_{\alpha/2} * s/\sqrt{n} = -1.28 * 2/\sqrt{100}$$

$$= 0.256$$

CI is $[\bar{x} - MoE, \bar{x} + MoE] \Rightarrow [50 - 0.256, 50 + 0.256]$

$$= (49.744, 50.256)$$

Confidence Intervals

CL	n	s	$\bar{x}$	$Z_{\alpha/2}$	MoE $\left(Z_{\alpha/2} * \frac{s}{\sqrt{n}}\right)$	C I
68	100	2	50	.99/.1	.2	(49.8, 50.2)
80					.256	(49.744, 50.256)
90					.33	(49.67, 50.33)
95				1.96	.392	(49.608, 50.392)
99					.516	(49.484, 50.516)

CL ⇒ 95%

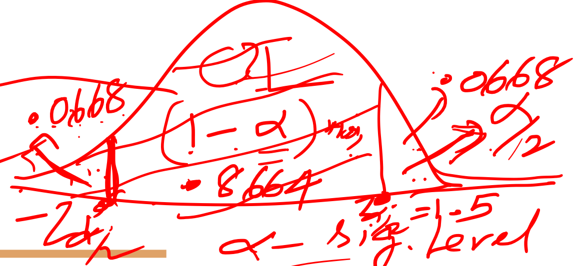
CL ⇒ .95 $(= 1 - \alpha)$

α ⇒ significance level 5%

$(1-\alpha) * 100 \Rightarrow 95\%$

Confidence Intervals

$\alpha/2 = .0668$
$\Rightarrow \alpha = .1336$



PES UNIVERSITY ONLINE

In a sample of 100 wires the average breaking strength is 50kN, with a standard deviation of 2kN.

$\bar{x} = 50$
$n = 100$
$s = 2$

$\boxed{86.64\%}$

$\bar{x} \pm MoE$
$50 \pm .3$

f) An engineer claims that the mean breaking strength is between 49.7kN and 50.3kN. With what level of confidence can this statement be made?

$[49.7, 50.3]$

$-.3 \leftarrow 50 \rightarrow +.3$

$MoE = .3$

$CL = ?$

$(1-\alpha)*100\%$ CI for μ is given by

$$\bar{x} \pm \boxed{Z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}} \; MoE$$

$MoE = 0.3$

$$Z_{\alpha/2} \cdot \frac{s}{\sqrt{n}} = 0.3$$

$$Z_{\alpha/2} \cdot \frac{2}{\sqrt{100}} = 0.3$$

$$Z_{\alpha/2} = \frac{0.3 \times 10}{2} = \frac{3}{2} = 1.5$$

$\alpha = ?$

$(1-\alpha)*100\% = (1-.1336)*100\%$
$= 0.8664*100$
$= 86.64\%$

$\therefore$ Area to the right of $Z_{\alpha/2} = 1.5$
$= 1 - .9332$
$A_2 = .0668 \Rightarrow \alpha = 2(\alpha/2) = .1336$

Two Sided Confidence Intervals

Given $n = 100$, $\bar{X} = 50$, $s = 2$, CI is $(49.7, 50.3)$

$\underbrace{\bar{X}}_{50} - MoE$, $\underbrace{\bar{X}}_{50} + MoE$

Find CL.

From the given CI, we know that MoE is 0.3

$$MoE = Z_{\alpha/2} \cdot \frac{s}{\sqrt{n}} = 0.3$$

$$Z_{\alpha/2} * \frac{2}{\sqrt{100}} = 0.3$$

$$Z_{\alpha/2} = \frac{0.3 \times 10}{2} = \frac{3}{2} = 1.5$$

Area to the right $= 0.0668 = \alpha/2$

$\Rightarrow \alpha = 2(0.0668) = 0.1336$

$\Rightarrow \alpha = 13.36\%$

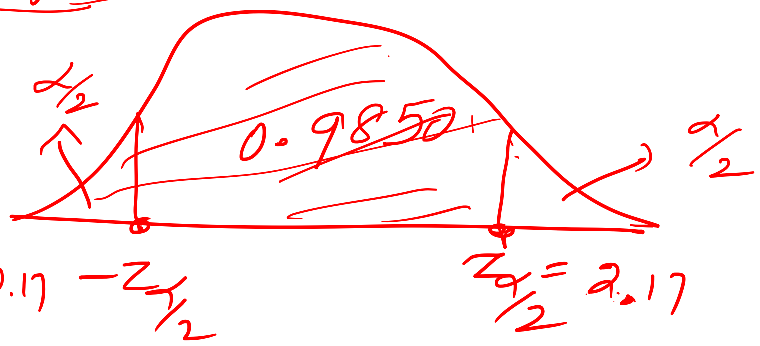
$\therefore$ Confidence Level $= 100 - 13.36\%$
$= 86.64\%$

Two Sided Confidence Intervals

Given: $Z_{\alpha/2} = 2.17$

Find the Confidence Level.

Area to the left of $Z_{\alpha/2}$ (2.17) $= 0.9850$



Area to the right of $Z_{\alpha/2} = 1 - 0.9850 = 0.0150 \; (= \alpha/2)$

$\alpha = 2 * \alpha/2 = 2 * 0.0150 = 0.03$

$\alpha = 0.03 \Rightarrow 3\%$

CL : $(1-\alpha) * 100\% = (1 - 0.03) * 100\% = 97\%$

$\therefore$ CL is 97%

Confidence Intervals

In a sample of 100 wires the average breaking strength is 50kN, with a standard deviation of 2kN.

$n = 100$

g) How many wires must be sampled so that a 95% confidence interval specifies the mean breaking strength to within 0.3 kN?

$n = ?$

For 95% CL

$Z_{\alpha/2} = 1.96$

$MoE = Z_{\alpha/2} \; s/\sqrt{n}$

$0.3 = 1.96 * 2/\sqrt{n}$

$\sqrt{n} = 1.96 * 2/0.3$

$\sqrt{n} = 13.07$

$n = (13.07)^2 = 170.82 \simeq \underline{\underline{171}}$

$\therefore n = \underline{\underline{171}}$

Two Sided Confidence Intervals

Given $s = 2$. $CL = 95\%$, $MoE = 0.3$ Find $n = ?$

$$MoE = Z_{\alpha/2} * \frac{s}{\sqrt{n}}$$

$$0.3 = 1.96 * \frac{2}{\sqrt{n}}$$

$$\sqrt{n} = \frac{1.96 \times 2}{0.3}$$

$$n = 170.650667$$

$$\boxed{n \simeq 171}$$

In a sample of 100 wires the average breaking strength is 50kN, with a standard deviation of 2kN.

h) How many wires must be sampled so that a 99% confidence interval specifies the mean breaking strength to within 0.3 kN?

$$MoE = Z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

For 99% CL $\Rightarrow Z_{\alpha/2} = 2.58$

$$0.3 = 2.58 * \frac{2}{\sqrt{n}}$$

$$\sqrt{n} = \frac{2.58 * 2}{0.3}$$

$$\sqrt{n} = 17.2 \Rightarrow n = (17.2)^2 = 295.84 \simeq 296$$

$$n = \underline{296}$$

Probability vs. Confidence

It is correct to say that there is a **95% chance that the confidence interval** you calculated **contains the true population mean.**

It is **not quite correct to say that** there is a **95% chance that the population mean lies within the interval.** The population mean has one value.

$\mu \in [\quad]$

In contrast, the confidence interval you compute depends on the data you happened to collect.

[Interval]

covers μ

Two -Sided Confidence Intervals

Do It Yourself!!!

A random sample of n = 50 males showed a mean average daily intake of dairy products equal to 756 grams with a standard deviation of 35 grams. Find a 95% confidence interval for the population average μ?

An investigator computes 95% confidence interval for a population mean on the basis of a sample of size 70. If she wishes to compute a 95% confidence interval that is half as wide, how large a sample does she need?

One-Sided Confidence Intervals

One Sided Confidence Intervals

Two-Sided

CI for μ : $[LB, UB]$

$\left[\bar{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$

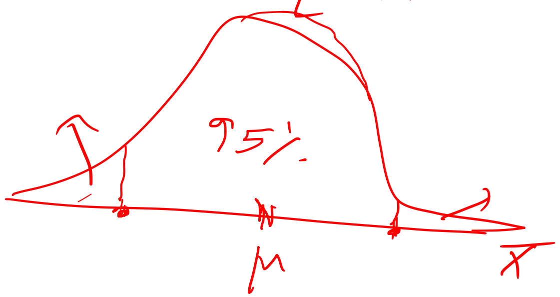
(Or)

$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$ $\left[\bar{X} - Z_{\alpha/2}\sigma_{\bar{X}}, \bar{X} + Z_{\alpha/2}\sigma_{\bar{X}}\right]$

where $\sigma_{\bar{X}}$ is the S.D of $\bar{X}$

(or)

$\left[\bar{X} - MoE, \bar{X} + MoE\right]$



One-Sided CI

$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

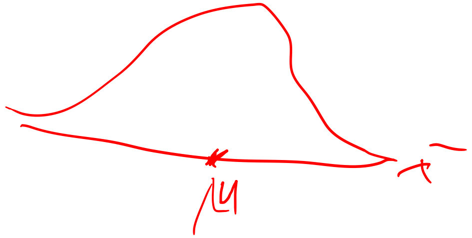
CI for μ :

Lower Interval $\left(-\infty, \bar{X} + \underbrace{Z_\alpha \sigma_{\bar{X}}}_{MoE}\right)$

Upper Bound

CI for μ

Upper Interval $\left(\bar{X} - Z_\alpha \sigma_{\bar{X}}, +\infty\right)$



One Sided Confidence Intervals

$\bar{X} \Rightarrow$ sample mean σ s,

$\bar{X} \sim N(\mu, \sigma^2/n)$

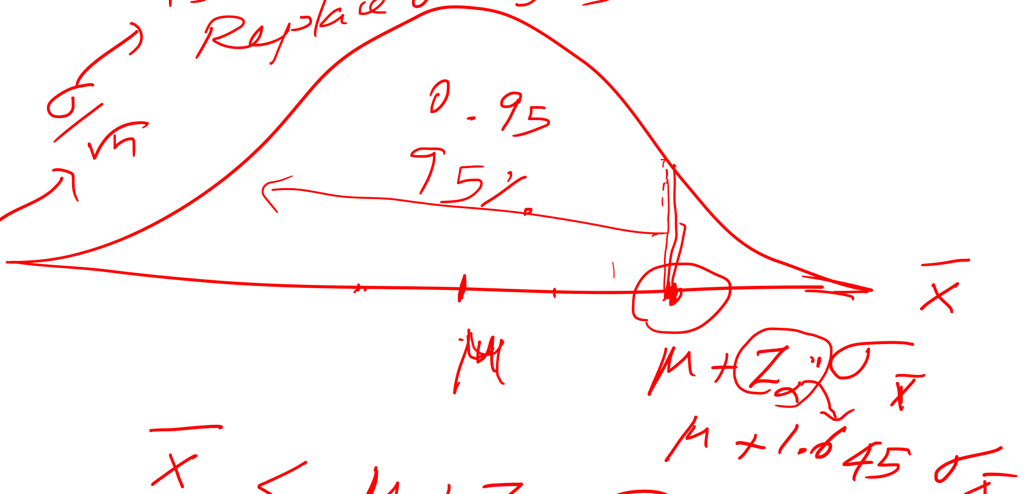
$\sigma/\sqrt{n}$ is unknown, Replace σ by s

Lower Interval : $(-\infty, UB)$

$(-\infty, \bar{x} + Z_\alpha * \sigma_{\bar{x}})$

Upper Interval : $(LB, +\infty)$

$(\bar{X} - Z_\alpha * \sigma_{\bar{x}}, +\infty)$



$\bar{x} < \mu + Z_\alpha \sigma_{\bar{x}}$

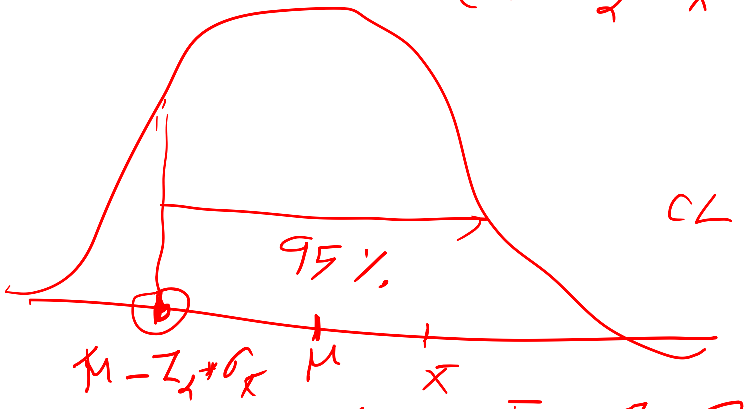
$(\bar{x} - Z_\alpha \sigma_{\bar{x}}, +\infty) \rightarrow$ This interval covers μ

$(\bar{x} - 1.645 * \sigma_{\bar{x}}, +\infty)$

95%

90%

80%



$(-\infty, \bar{X} + Z_\alpha \cdot \sigma_{\bar{x}})$

CL = 95%

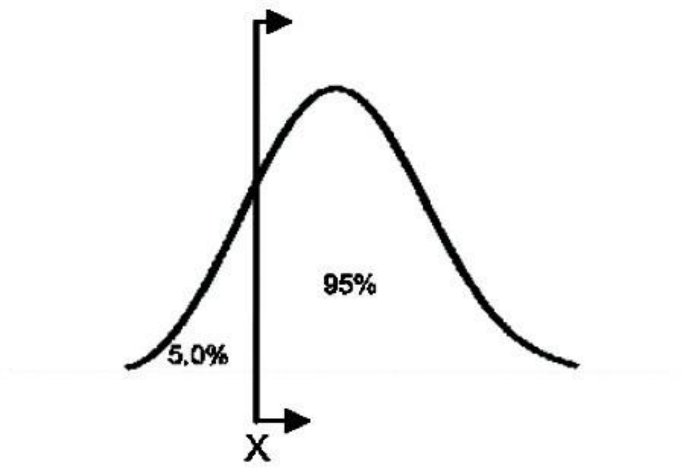
95% of the sample mean values will be greater than $\mu - Z_\alpha \sigma_{\bar{x}}$

$\bar{x} > \mu - Z_\alpha \sigma_{\bar{x}}$

One-Sided Confidence Intervals

1)An upper one-sided bound defines a point that a certain percentage of the population is greater than.

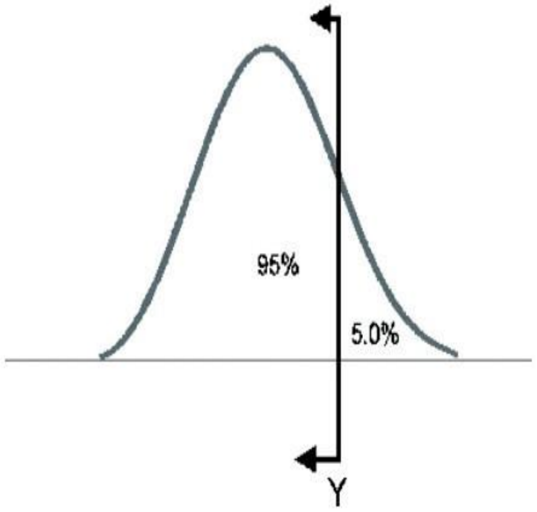
For example, for a 95% upper one-sided bound, this would indicate that 95% of the population parameter is greater than X.



One-Sided Confidence Intervals

2)A lower one-sided bound defines a point that a specified percentage of the population is less than.

For example, For a 95% lower one-sided bound, this would indicate that 95% of the population is less than Y.



Example1

In a sample of 80 ten-penny nails, the average weight was 1.56g and the standard deviation was 0.1g.

a) Find a 90% upper confidence bound for the mean weight.

b) Find a 80% lower confidence bound for the mean weight.

c) Someone says that the mean weight is less than 1.585g. With what level of confidence can this statement be made?

Example1

In a sample of 80 ten-penny nails, the average weight was 1.56g and the standard deviation was 0.1g.

a) Find a 90% upper confidence bound for the mean weight.

$n = 80$ $\quad \bar{x} = 1.56$ $\quad s = 0.1$

Construct Lower CI : $(-\infty, UB)$

$(-\infty, \bar{x} + Z_\alpha * \sigma_{\bar{x}})$

$(-\infty, \bar{x} + Z_\alpha * \frac{\sigma}{\sqrt{n}})$

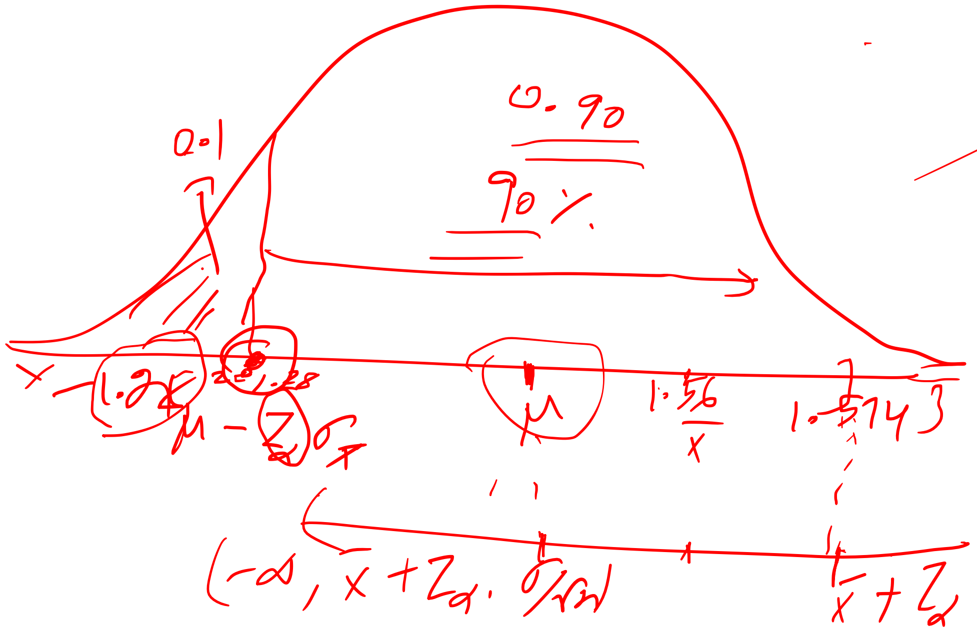
$(-\infty, 1.56 + 1.28 * \frac{0.1}{\sqrt{80}})$

$\bar{x} > \mu - Z_\alpha \sigma_{\bar{x}}$

$(-\infty, \bar{x} + Z_\alpha \frac{\sigma}{\sqrt{n}})$

$(-\infty, 1.56 + 0.0143)$

$(-\infty, \underbrace{1.5743}_{UB})$



Example1

In a sample of 80 ten-penny nails, the average weight was 1.56g and the standard deviation was 0.1g.

b) **Find a 80% lower confidence bound for the mean weight.**

Upper Interval : $(LB, +\infty)$

$(\bar{x} - Z_\alpha \sigma/\sqrt{n}, +\infty)$

$n = 80 \quad \bar{x} = 1.56 \quad s = 0.1 \quad CL = 80\%$

$Z_\alpha = 0.84 \quad 1 - \alpha = 0.8$

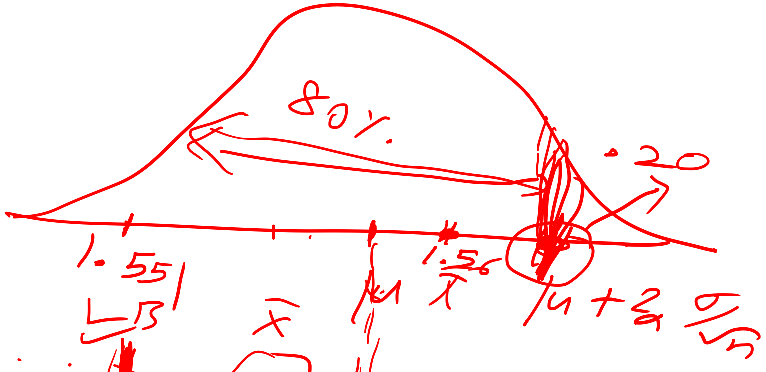
$(\bar{x} - Z_\alpha * s/\sqrt{n}, +\infty)$

$(1.56 - 0.84 * 0.1/\sqrt{80}, +\infty)$ 80%

$(1.56 - \underbrace{0.009}_{MoE}, +\infty)$

$(\underbrace{1.551}_{LB}, +\infty)$

$(\bar{x} - z_\alpha \cdot s/\sqrt{n}, +\infty)$



$(\bar{x} - Z_\alpha * s/\sqrt{n}, \bar{x} +\infty)$

Example1

In a sample of 80 ten-penny nails, the average weight was 1.56g and the standard deviation was 0.1g. $n=80$ $\bar{x}=1.56$ $s=0.1$

c) **Someone says that the mean weight is less than 1.585g. With what level of confidence can this statement be made?**

$$UB = \bar{x} + \boxed{Z_\alpha \cdot \frac{s}{\sqrt{n}}}$$

$$1.585 = 1.56 + Z_\alpha \cdot \frac{0.1}{\sqrt{80}}$$

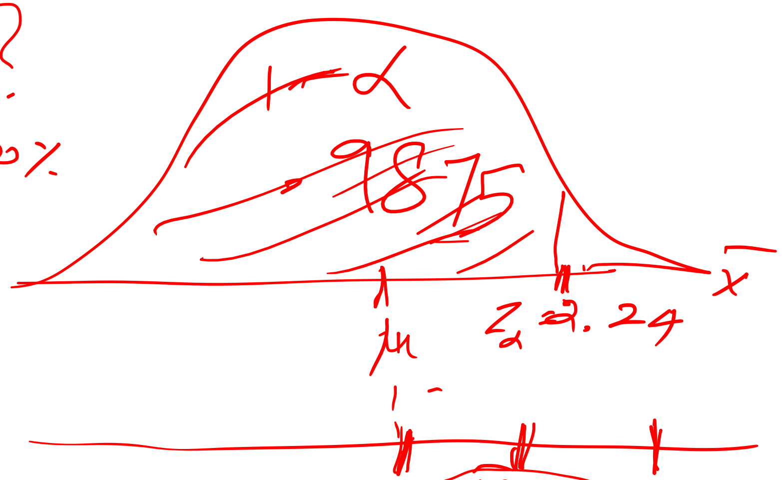
$$\boxed{CL = 98.75\%}$$

$$Z_\alpha \cdot \frac{0.1}{\sqrt{80}} = 1.585 - 1.56 = 0.029$$

$$Z_\alpha = \frac{0.029 * \sqrt{80}}{0.1} = 2.235 \simeq 2.24$$

$CL = ?$

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



$\bar{x}=1.56$ 1.585 $\bar{x}+Z_\alpha \cdot \sigma/\sqrt{n}$

Area to the left of 2.24 = .9875

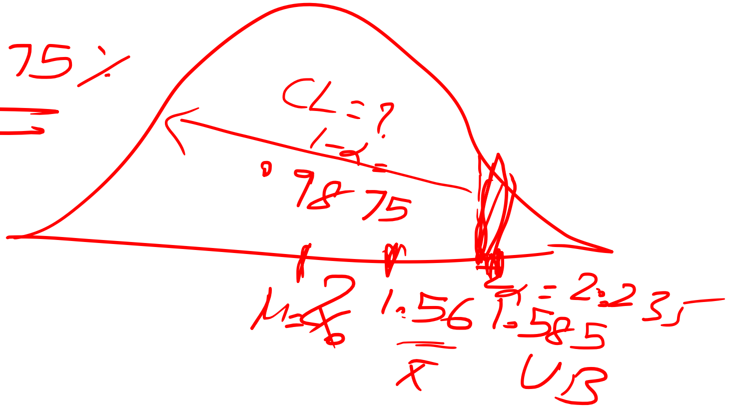
∴ Confidence Level = (.9875)*100 = 98.75%

Example1

In a sample of 80 ten-penny nails, the average weight was 1.56g and the standard deviation was 0.1g.

c) **Someone says that the mean weight is less than 1.585g. With what level of confidence can this statement be made?**

$CI = 98.75\%$



$$\bar{x} + Z_\alpha \cdot \frac{s}{\sqrt{n}} = UB$$

$$1.56 + Z_\alpha \frac{0.1}{\sqrt{80}} = 1.585$$

$$Z_\alpha * \frac{0.1}{\sqrt{80}} = 1.585 - 1.56 = 0.025$$

$$Z_\alpha = \frac{0.025 * \sqrt{80}}{0.1} = 2.235 \simeq 2.24$$

Area to the left of $Z = 2.24$ is .9875

$1-\alpha = .9875$

$(1-\alpha)*100 = .9875*100$

$= 98.75\%$

Solution

a) 90% upper confidence bound for the mean weight.

= 1.5743

b) Find a 80% lower confidence bound for the mean weight.

= 1.551

c) Someone says that the mean weight is less than 1.585g. With what level of confidence can this statement be made?

Hence we can make the statement with 98.75% confidence.

Confidence Interval for population proportion p

Confidence Intervals for Population Proportion p

$X \sim Bin(n, p)$ where p is unknown

Estimate of p is $\hat{p} = \frac{X}{n}$

$$\underbrace{\hat{p} - Z_{\alpha/2} * \sigma_{\hat{p}}}_{LB} \le p \le \underbrace{\hat{p} + Z_{\alpha/2} * \sigma_{\hat{p}}}_{UB}$$

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

$$\hat{p} - Z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} < p < \hat{p} + Z_{\alpha/2} * \sqrt{\frac{p(1-p)}{n}}$$

$$\sqrt{\frac{\hat{p} * (1-\hat{p})}{n}} < p < \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

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

This CI will be shorter in some cases

STATISTICS FOR DATA SCIENCE

Confidence Intervals for Population Proportion p

Little modifications in X and n.

$\tilde{X} = X+2$ and $\tilde{n} = n+4$

W.K.T $\hat{p} = \frac{X}{n} \Rightarrow \tilde{p} = \frac{X+2}{\tilde{n}}$ where $\tilde{n} = n+4$

~~For~~ $100(1-\alpha)\%$ CI for p is given by

$$\left(\underbrace{\tilde{p} - Z_{\alpha/2}\sqrt{\frac{\tilde{p}(1-\tilde{p})}{\tilde{n}}}}_{LB}, \underbrace{\tilde{p} + Z_{\alpha/2}\sqrt{\frac{\tilde{p}(1-\tilde{p})}{\tilde{n}}}}_{UB} \right)$$

Confidence Intervals for Population Proportion p

Two-Sided CI for p

$(1-\alpha)*100\%$ CI is given by

Point Estimate $\pm$ MoE

$$\hat{p} \pm Z_{\alpha/2} * \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

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

UB

LB

(Traditional Approach)

$$\tilde{p} \pm Z_{\alpha/2} * \sqrt{\frac{\tilde{p}(1-\tilde{p})}{\tilde{n}}}$$

shorter CI

$\hat{p}$ & p $\Downarrow$ $\tilde{p}$

n $\Downarrow$ $\tilde{n}$

$$\tilde{p} = \frac{X+2}{\tilde{n}}$$

$$\hat{p} = \frac{X+2}{n+4}$$

$\tilde{p}$

where $\tilde{n} = n+4$

Confidence Intervals for Population Proportion p

The article "Study on the Life Distribution of Microdrills" (Z. Yang, Y. Chen, and Y. Yang,Journal of Engineering Manufacture, 2002:301–305) reports that in a sample of 50 microdrills drilling a low-carbon alloy steel, the average lifetime (expressed as the number of holes drilled before failure) was 12.68 with a standard deviation of 6.83. Find a 95% confidence interval for the mean lifetime of microdrills under these conditions.

μ

Now assume that a specification has been set that a drill should have a minimum lifetime of 10 holes drilled before failure. A sample of 144 microdrills is tested, and 120, or 83.3%, meet this specification. Let p represent the proportion of microdrills in the population that will meet the specification. **We wish to find a 95% confidence interval for p.**

$n = 144 \quad x = 120$ Estimate for p is $\hat{p} = \frac{x}{n} = \frac{120}{144}$

x : No. of drills that meets specification $\quad \hat{p} = 0.833$

Confidence Intervals for Population Proportion p

Now assume that a specification has been set that a drill should have a minimum lifetime of 10 holes drilled before failure. A sample of 144 microdrills is tested, and 120, or 83.3%, meet this specification. Let p represent the proportion of microdrills in the population that will meet the specification. **We wish to find a 95% confidence interval for p.**

$n = 144$ $\quad$ $X = 120$

$\hat{p} = \frac{X}{n} = \frac{120}{144} = 0.833$

$(1-\alpha)*100\%$ CI is given by

$\tilde{p} \pm z_{\alpha/2}\sqrt{\frac{\tilde{p}(1-\tilde{p})}{\tilde{n}}}$

CL = 95%

$(1-\alpha) = .95$

$\Rightarrow \alpha = .05$

$\alpha/2 = .025$

$\tilde{n} = n+4 = 144+4 = 148$

$\tilde{p} = \frac{X+2}{\tilde{n}} = \frac{120+2}{148} = \frac{122}{148} = 0.8243$

$0.8243 \pm 1.96 * \sqrt{\frac{0.8243*(1-0.8243)}{148}}$

$\Rightarrow 0.8243 \pm \cancel{1.96}\ 0.0613$

$(\underbrace{0.8243 - 0.0613}_{LB}, \underbrace{0.8243 + 0.0613}_{UB}) \Rightarrow (0.763, 0.886)$

LB < p < UB

Confidence Intervals for Population Proportion p

Example: $n = 144$ $\quad X = 120$ $\quad \hat{p} = 0.833$

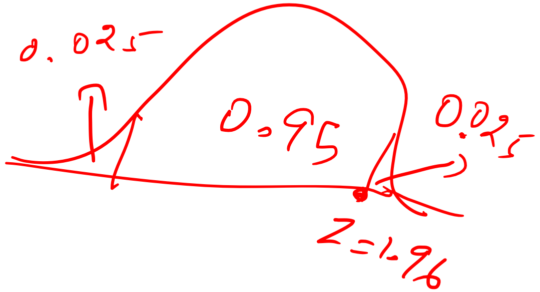
95% CI for p:

$\tilde{n} = n + 4 = 144 + 4 = 148$

$$\tilde{p} = \frac{X+2}{\tilde{n}} = \frac{120+2}{148} = \frac{122}{148} = 0.8243$$

$100(1-\alpha)\%$ CI is given by

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



$$\left(0.8243 - 1.96\sqrt{\frac{0.8243(1-0.8243)}{148}},\ 0.8243 + 1.96\sqrt{\frac{0.8243(1-0.8243)}{148}}\right)$$

$\Rightarrow (0.763,\ 0.886)$ or CI is 0.8243 ± 0.0613

Do It Yourself!!!

One step in the manufacture of a certain metal clamp involves the drilling of four holes. In a sample of 150 clams, the average time needed to complete this step was 72 seconds and the standard deviation was 10 seconds. An efficiency expert says that the mean time is greater than 70 seconds. With what level of confidence can this statement be made?

THANK YOU

D. Uma

Computer Science and Engineering

umaprabha@pes.edu
+91 99 7251 5335