Statistics

Part 2

Central Limit Theorem (CLT): The distribution of sample statistics is nearly normal, centered at the population mean, and with a standard deviation equal to the population standard deviation divided by square root of the sample size.

$$\bar{x} \sim N\left(mean = \mu, SE = \frac{s\,\cancel{\sigma}}{\sqrt{n}}\right)$$

shape center spread

- If we report a point estimate , we probably won’t hit the exact population parameter.
- If we report a range of plausible values we have a good shot at capturing the parameter.

- Fishing with Spear   Fishing with Net

Confidence Interval for a mean

Confidence Interval is defined as a plausible range of values for the population parameter

Using only a sample statistic for estimation of parameter is unreliable. If we report a point estimate, we most probably wont hit the exact population parameter, but if we report a range of plausible values, we have a good shot at capturing the parameter

Sample mean **X** is our best guess for the unknown population mean. Hence any interval we construct is to be around that **X** that we know, to be our best guess.

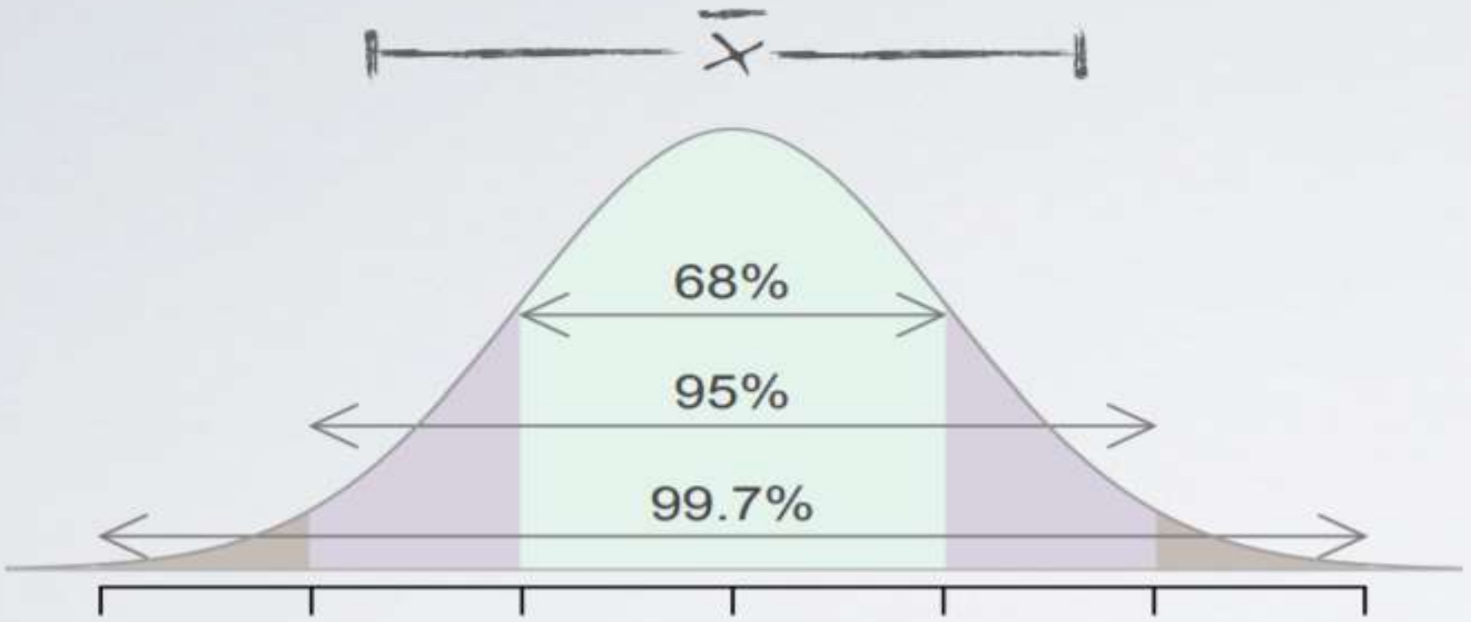
From CLT we know **X** is distributed nearly normally and the centre of the distribution is at the unknown population mean

Confidence Interval for a mean

- Considering the nearly normally distribution rule, we can state that roughly 95% of random samples will have sample means that are within 2 standard errors of the population mean.
- Clearly ,then for the 95%of the random samples , the unknown true population mean is going to be within the standard errors of that sample's mean.
- So the 95% Confidence Interval can be constructed approximately as sample mean ± 2 SE i.e. Approx. 95% CI : X ± 2 SE
- ± 2 SE is called Margin of Error (ME)

Central Limit Theorem (CLT):

$$\bar{x} \sim N\left(mean = \mu, SE = \frac{\sigma}{\sqrt{n}}\right)$$

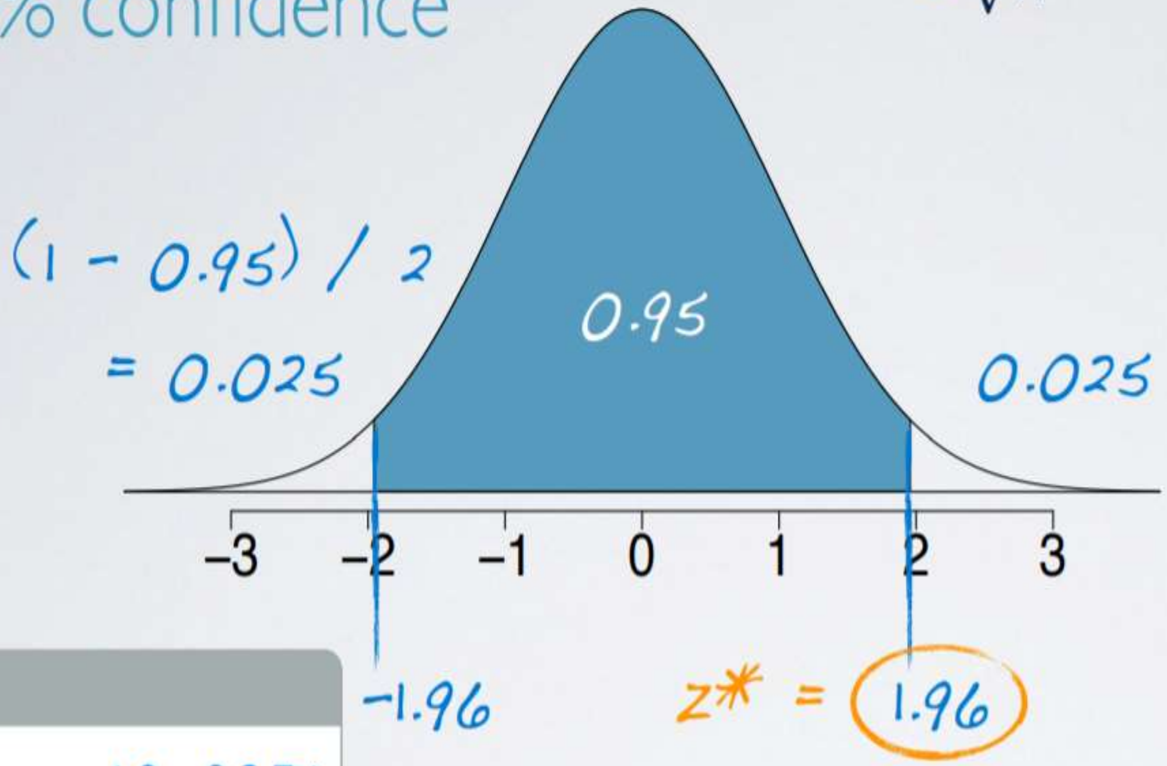


approximate 95% CI: $\bar{x} \pm \boxed{2SE}$

margin of error (ME)

finding the critical value
95% confidence

$$\bar{x} \pm z^{\star} \frac{s}{\sqrt{n}}$$



R

```
> qnorm(0.025)
[1] -1.96
```

Second decimal place					
0.07	0.06	0.05	0.04	0.00	Z
0.0003	0.0003	0.0003	0.0003	0.0003	−3.4
0.0004	0.0004	0.0004	0.0004	0.0005	−3.3
0.0005	0.0006	0.0006	0.0006	0.0007	−3.2
0.0008	0.0008	0.0008	0.0008	0.0010	−3.1
0.0011	0.0011	0.0011	0.0012	0.0013	−3.0
0.0015	0.0015	0.0016	0.0016	0.0019	−2.9
0.0021	0.0021	0.0022	0.0023	0.0026	−2.8
0.0028	0.0029	0.0030	0.0031	0.0035	−2.7
0.0038	0.0039	0.0040	0.0041	0.0047	−2.6
0.0051	0.0052	0.0054	0.0055	0.0062	−2.5
0.0068	0.0069	0.0071	0.0073	0.0082	−2.4
0.0089	0.0091	0.0094	0.0096	0.0107	−2.3
0.0116	0.0119	0.0122	0.0125	0.0139	−2.2
0.0150	0.0154	0.0158	0.0162	0.0179	−2.1
0.0192	0.0197	0.0202	0.0207	0.0228	−2.0
0.0244	0.0250	0.0256	0.0262	0.0287	−1.9
0.0307	0.0314	0.0322	0.0329	0.0359	−1.8

Confidence Interval of a Population Mean

```
> qnorm(0.025)
[1] -1.96
```

What is the critical value for the 98% Confidence Interval?

a. Z = 2.05
b. Z = -1.96
c. Z = 2.33
d. Z = -2.33
e. Z = 1.96

Ans. ‘c’

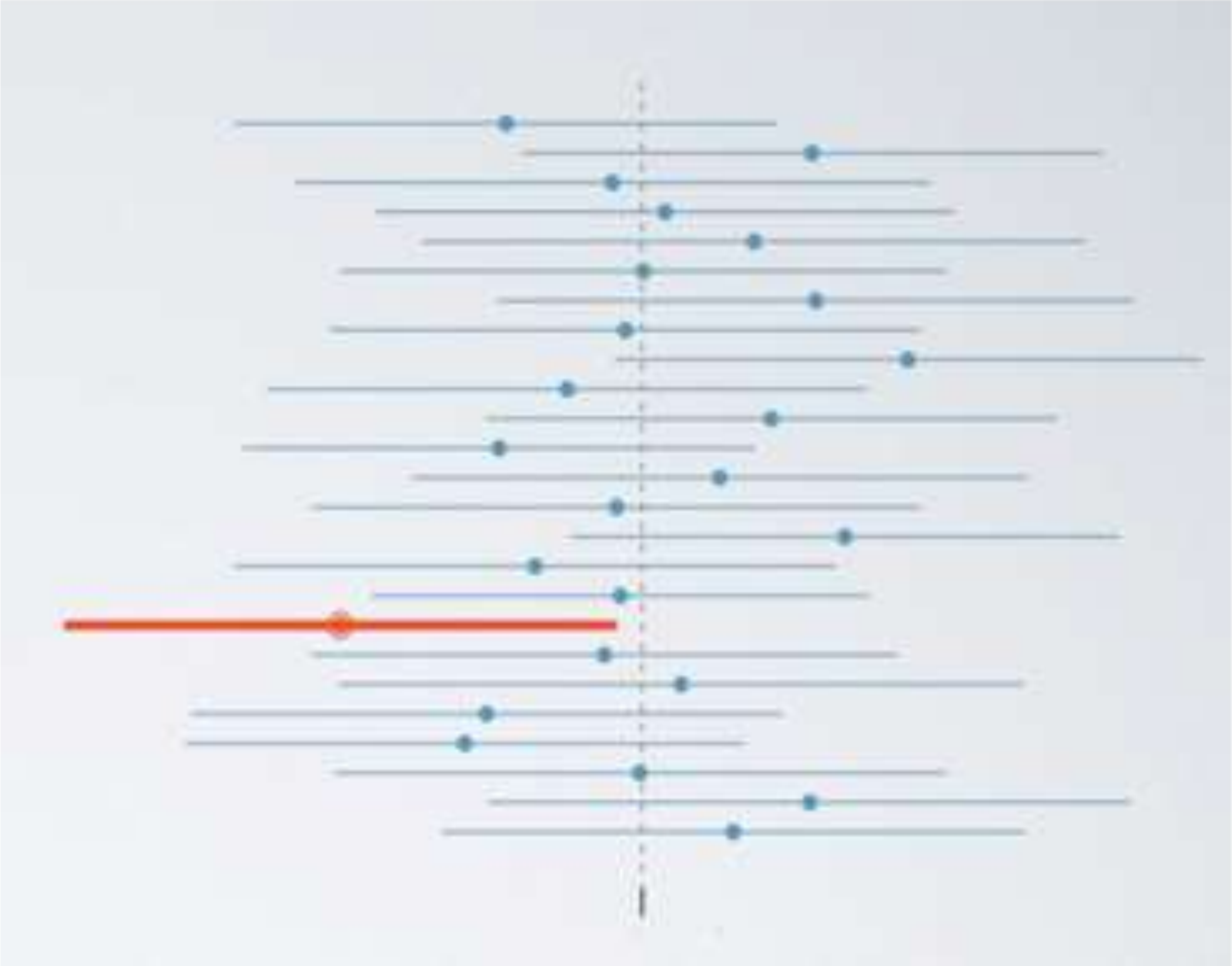
Accuracy V/s Precision

Suppose we took many samples and built a confidence interval for each sample using the equation : point estimate ± 1.96 SE, then about 95% of these intervals would contain the true population parameter μ.

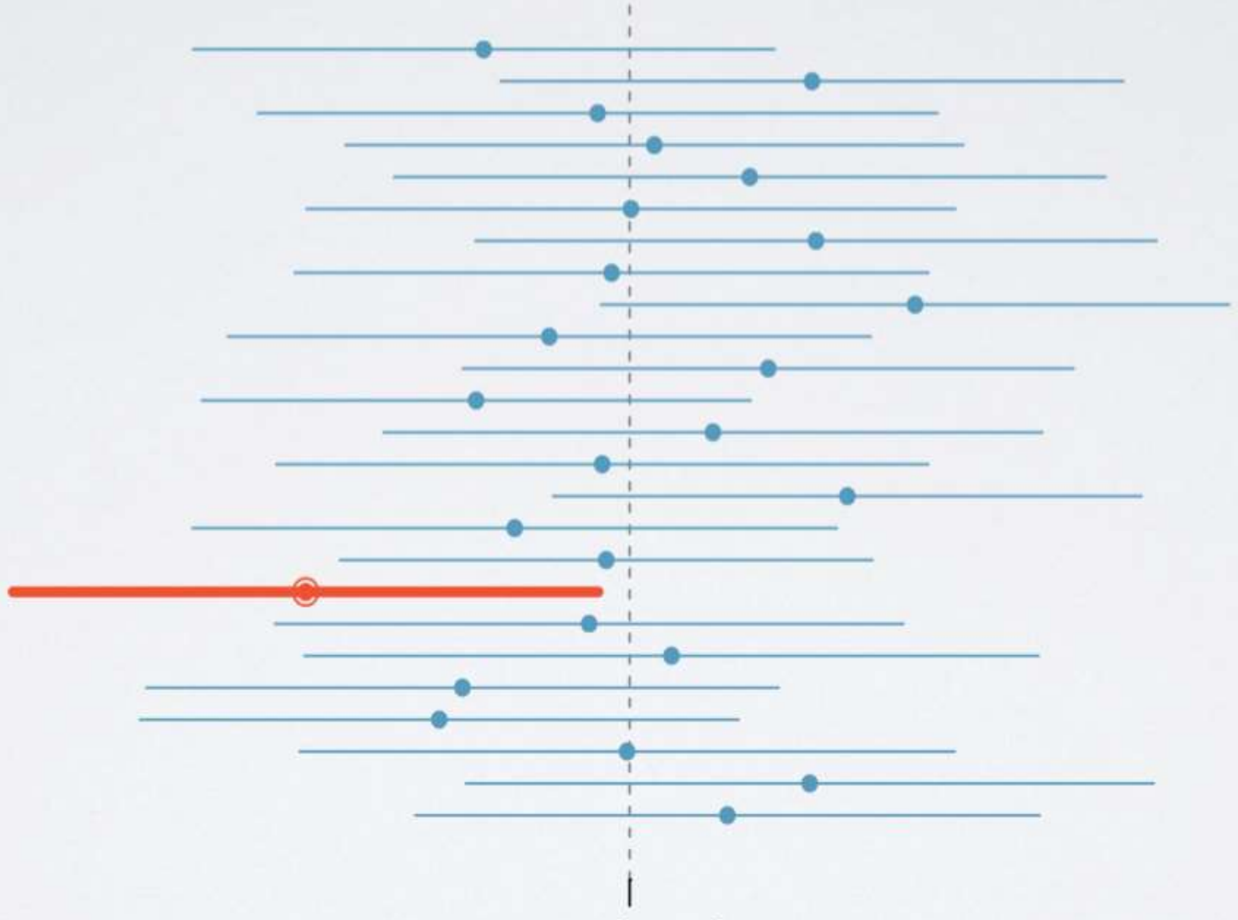
Hence the Confidence level for these intervals would be 95%

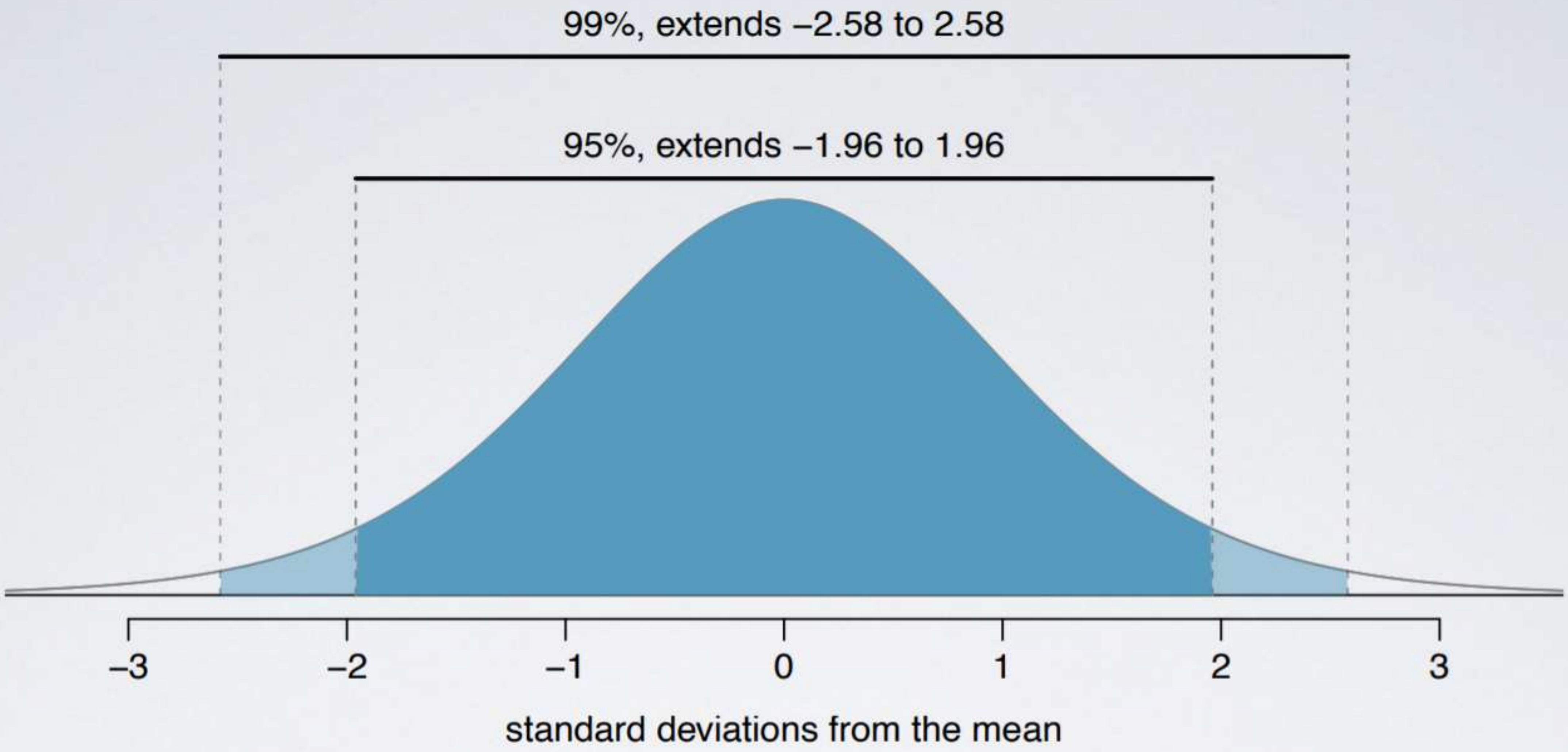
Confidence levels commonly used are 90%, 95%, 98% and 99%

Confidence Level



If we want to be very certain that we capture the population parameter, should we use a wider interval or a narrower interval?




99%, extends −2.58 to 2.58
95%, extends −1.96 to 1.96
−3
−2
−1
0
1
2
3
standard deviations from the mean

CL ↑ width ↑

Accuracy V/s Precision

- As the confidence level increases, the larger the critical value, hence the larger ME and hence the width of the confidence interval also increases

- Weather forecast….. Next day maximum temperature would be between 5 deg to 45 deg.
- Such weather forecast is not precise. It is not informative. It doesn't help me decide whether to wear a sweater or light cotton attire.
- As the confidence level increases, the width of CI increases, which increases accuracy. However precision goes down.

Accuracy V/s Precision

- In order to get higher precision as well as higher accuracy, increase sample size. It reduces SE and hence ME. Therefore we can remain at a high confidence level while not needing to increase the Confidence Interval