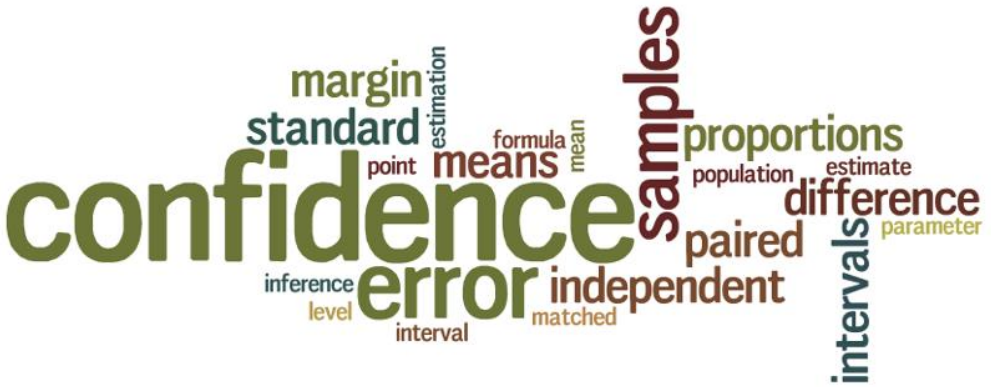


Confidence Interval, t-Distribution, Hypothesis Testing, t-Tests

Review

- Gaussian Distribution
  - Areas under the curve
  - Reading the normal distribution table
  - Approximating B(n,p) with Normal distribution
  - Continuity correction

- Central Limit Theorem

Activity - R

According to the Indian Bureau of the Census, about 75% of the commuters in the India drive to work alone. Suppose 150 Indian commuters are randomly sampled.

- What is the probability that fewer than 105 commuters drive to work alone?
- What is the probability that between 110 and 120 (inclusive) commuters drive to work alone?
- What is the probability that more than 95 commuters drive to work alone?

Activity - R

- Expected Mean = 0.75*150=112.5
- Variance = n*p*q= 150*0.75*0.25

1. Area under the curve from –Infinity to 104.5 (continuity correction)

```
> pnorm(104.5,112.5,sqrt(150*0.75*0.25))
[1] 0.06571401
```

2. Area under the curve between 120.5 and 109.5

```
> pnorm(120.5,150*0.75,sqrt(150*0.75*0.25)) -  pnorm(109.5,150*0.75,sqrt(150*0.75*0.25))
[1] 0.6484822
```

3. Area under the curve above 95.5

```
> 1-pnorm(95.5,112.5,sqrt(150*0.75*0.25))
[1] 0.999326
```


Activity - R

According to National Center for Health Statistics of the US, the distribution of serum cholesterol levels for 20-74 year old males has a mean of 211mg/dl with a standard deviation of 46mg/dl.

- What is the probability that the serum cholesterol level of a male is >230mg/dl?

- What is the probability that the average serum cholesterol level of a random sample of 36 males will be >230mg/dl?

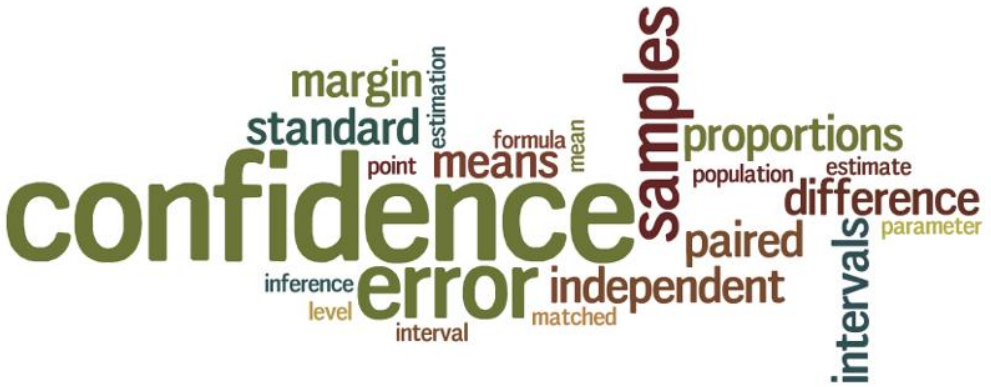
```
> 1-pnorm(230,mean=211,sd=46)
[1] 0.3397874
```

2) Expected value for the Average of the group = 211 SD of the sampling distribution for the group of 36

= 46/sqrt(36)

```
> 1-pnorm(230,mean=211,sd=46/sqrt(36))
[1] 0.006601229
```


" I got the instructions from my Statistics Professor. He was 80% confident that the true location of the restaurant was in this neighborhood."



Confidence Interval

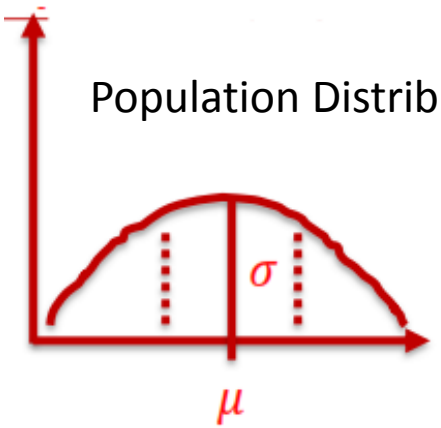
CI

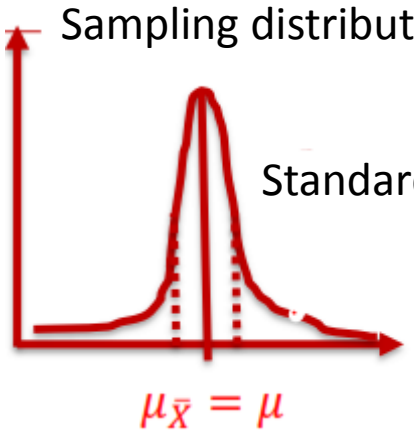
When we use samples to provide population estimates, we cannot be CERTAIN that they will be accurate. There is an amount of uncertainty, which needs to be calculated.

Publish Date	Source	Polling Organisation	NDA	UPA	Other
12 May 2014	[177]	CNN-IBN – CSDS – Lokniti	**276 (±6)**	97 (±5)	148 (±23)
	[177][178]	India Today – Cicero	**272 (±11)**	115 (±5)	156 (±6)
	[177][179]	News 24 – Chanakya	**340 (±14)**	70 (±9)	133 (±11)
	[177]	Times Now – ORG	**249**	148	146
	[177][180]	ABP News – Nielsen	**274**	97	165
	[177]	India TV – CVoter	**289**	101	148
14 May 2014	[181][182]	NDTV – Hansa Research	**279**	103	161
12 May 2014	[177]	Poll of Polls	**283**	105	149
16 May 2014	Actual Results[2]		**336**	58	**149**

Polling Organisation	NDA	UPA	Other
CNN-IBN – CSDS – Lokniti	**276 (±6)**	97 (±5)	148 (±23)
India Today – Cicero	**272 (±11)**	115 (±5)	156 (±6)
News 24 – Chanakya	**340 (±14)**	70 (±9)	133 (±11)

Incorrect way to present data as it gives the feeling that the population parameter will lie within these ranges.

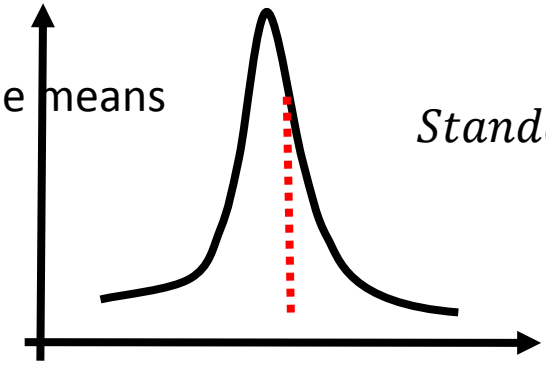




Standard Error (SE) is the same as Standard Deviation of the sampling distribution and a sample with 1 SE may or may not include the population parameter.

CI

Sampling distribution of sample means



$$Standard\ Error\ of\ the\ mean = \frac{\sigma}{\sqrt{n}}$$

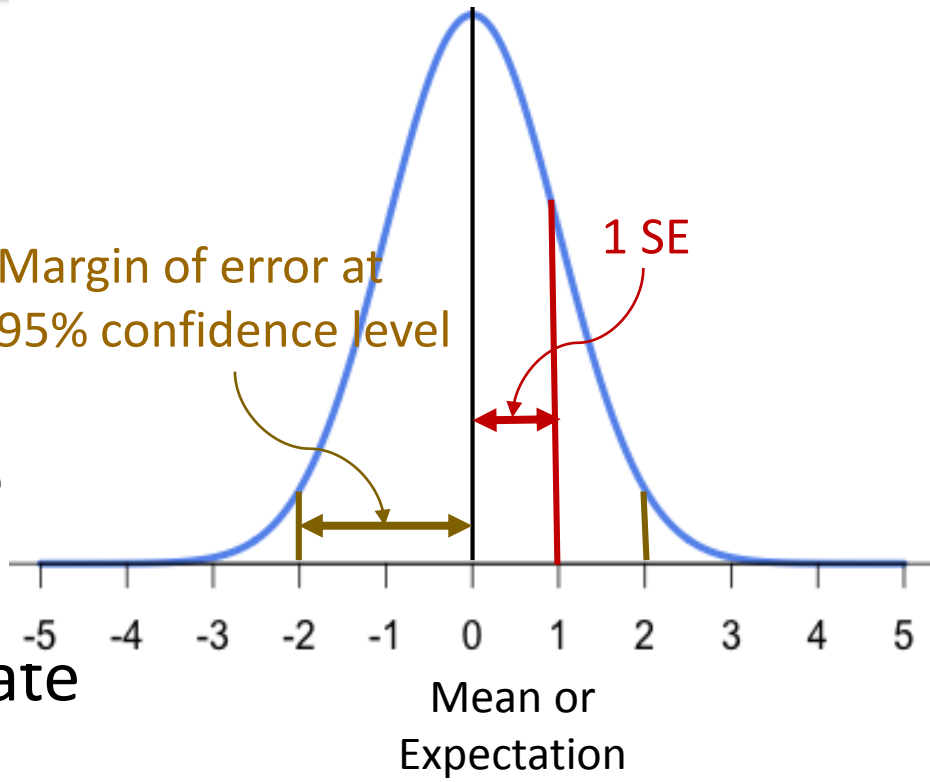
- We have seen that ~ 95% of the samples will have a mean value within the interval +/- 2 SE of the population mean *(recall the Empirical Rule for Normal Distribution).*

- Alternatively, 95% of such intervals include the population mean. Here, 95% is the Confidence Level and the interval is called the Confidence Interval.

SE, Margin of Error, Confidence Interval and Sample Size

$$SE = \frac{\sigma}{\sqrt{n}}$$

$$Margin\ of\ Error = z * SE$$



Margin of error is the **maximum expected difference between the** true population parameter and a sample estimate of that parameter.

Margin of error is meaningful only when stated in conjunction with a probability (confidence level).

SE, Margin of Error, Confidence Interval and Sample Size

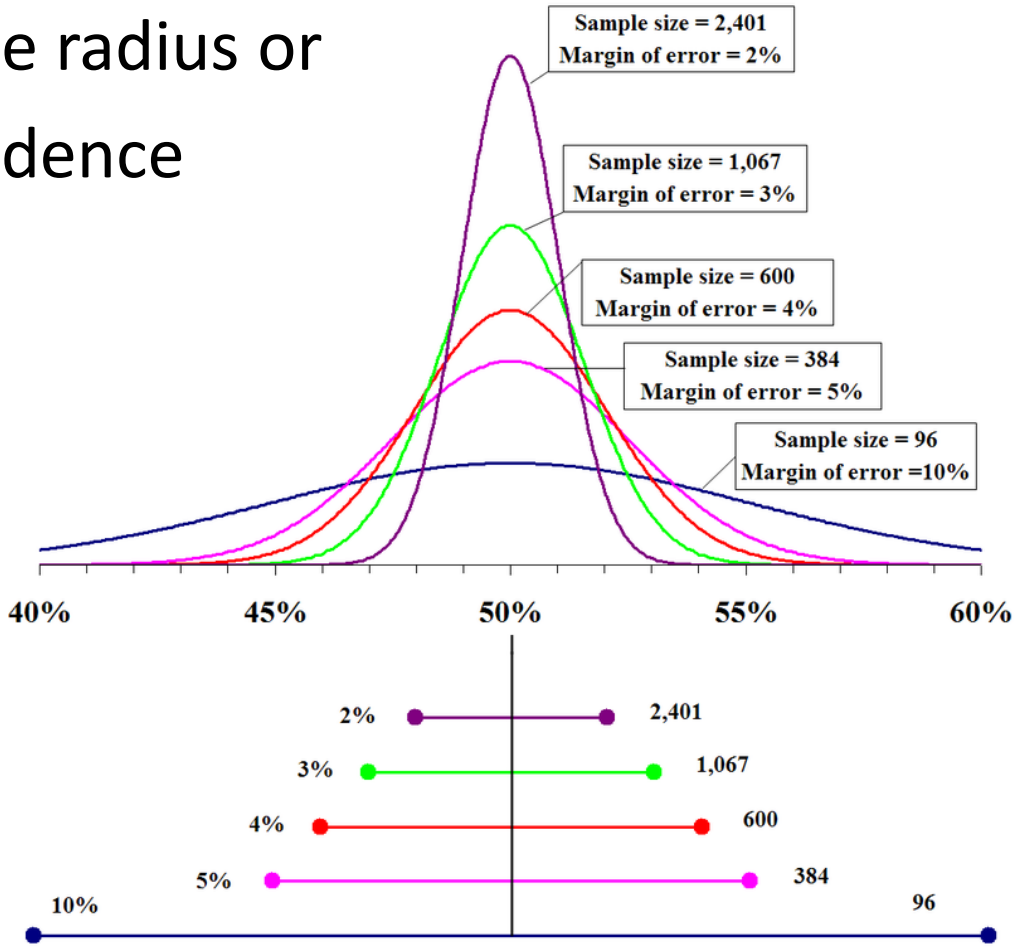
Just like Mean, Proportion is another common parameter of interest in many problems.

Expectation of a sample proportion = p

SE of a sample proportion = $\sqrt{\frac{pq}{n}}$

SE, Margin of Error, Confidence Interval and Sample Size

Margin of error is the radius or half-width of a confidence Interval.



SE, Margin of Error, Confidence Interval and Sample Size

In a poll by FOX News conducted between Apr 23 – 25, 2017, a survey of 1009 randomly sampled voters showed that the approval rating for President Trump down to 45%.

What is the margin of error at 95% confidence level (z = 1.96)? *Check qnorm(0.975, 0, 1). Why 0.975?*

SE, Margin of Error, Confidence Interval and Sample Size

Margin of error =

$$1.96 * \sqrt{\frac{0.45 * 0.55}{1009}} \approx 3.07\ \%$$

SE, Margin of Error, Confidence Interval and Sample Size

If the desired margin of error at 95% confidence level is 1%, what should be the sample size?

$$0.01 = 1.96 * \sqrt{\frac{0.45 * 0.55}{n}}$$

$$\Rightarrow n = \left(\frac{1.96}{0.01} * \sqrt{0.45 * 0.55}\right)^2 = 9508$$

Confidence Intervals

A survey was taken of US companies that do business with firms in India. One of the survey questions was: Approximately how many years has your company been trading with firms in India?

A random sample of 44 responses to this question yielded a mean of 10.455 years. Suppose the population standard deviation for this question is 7.7 years.

Using this information, construct a 90% confidence interval for the mean number of years that a company has been trading in India for the population of US companies trading with firms in India.

Confidence Intervals

$n = 44$,
$\bar{x} = 10.455$,
$\sigma = 7.7$

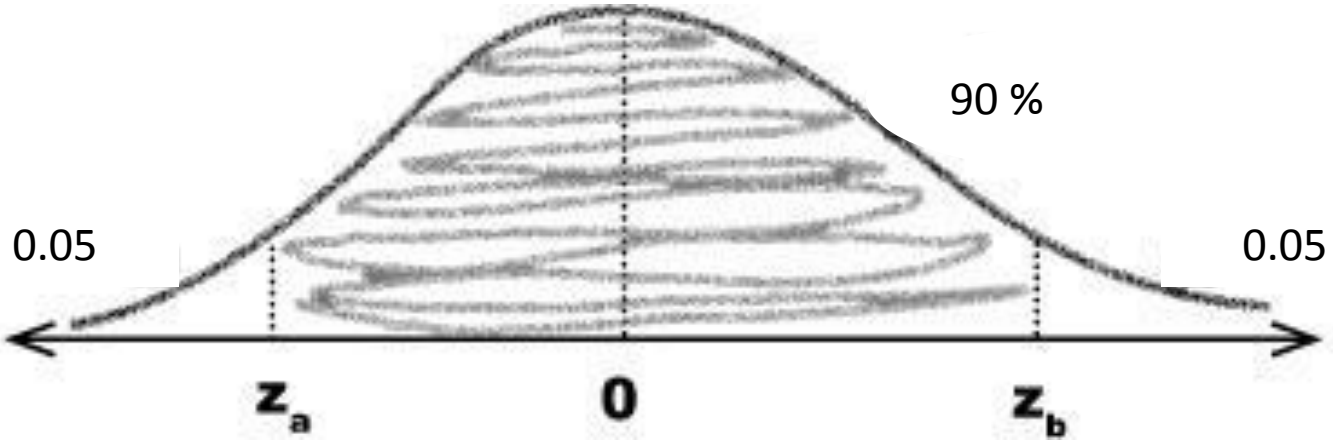
$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$ or Margin of error = $z * \frac{\sigma}{\sqrt{n}}$

Confidence Interval for the population Mean is

Sample Mean $\pm$ Margin of Error

Confidence Intervals

Find Z_a and Z_b where $P(Z_a < Z < Z_b) = 0.90$



$P(Z < Z_a) = 0.05$ and $P(Z > Z_b) = 0.05$

Confidence Intervals

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

From probability tables using interpolation, we get Z_a= -1.645 and Z_b= 1.645.

Check qnorm(0.05, 0,1) and qnorm(0.95, 0, 1) in R.

Confidence Interval

$Margin\ of\ error\ at$ 90% $Confidence\ Level$ = 1.645 $*\frac{7.7}{\sqrt{44}}$ = 1.91

Recall Confidence Interval for the Population Mean is Sample Mean ± Margin of Error

$\bar{X}$ - 1.91 < μ < $\bar{X}$ + 1.91

Since the sample mean is 10.455 years, we get the confidence interval for 90%

as 8.545 < μ < 12.365.

The analyst is 90% confident that if a census of all US companies trading with firms in India were taken at the time of the survey, the actual population mean number of trading years of such firms would be between 8.545 and 12.365 years.

Shortcuts for Calculating Confidence Intervals

Population Parameter	Population Distribution	Conditions	Confidence Interval
μ	Normal	You know σ^2 n is large or small $\bar{X}$ is the sample mean	$\left(\bar{X} - z\frac{\sigma}{\sqrt{n}}, \quad \bar{X} + z\frac{\sigma}{\sqrt{n}}\right)$
μ	Non-Normal	You know σ^2 n is large (>30) $\bar{X}$ is the sample mean	$\left(\bar{X} - z\frac{\sigma}{\sqrt{n}}, \quad \bar{X} + z\frac{\sigma}{\sqrt{n}}\right)$
μ	Normal or Non-Normal	You don't know σ^2 n is large (>30) $\bar{X}$ is the sample mean s^2 is the sample variance	$\left(\bar{X} - z\frac{s}{\sqrt{n}}, \quad \bar{X} + z\frac{s}{\sqrt{n}}\right)$
p	Binomial	n is large p_s is the sample proportion q_s is 1 - p_s	$\left(\bar{X} - z\frac{s}{\sqrt{n}}, \quad \bar{X} + z\frac{s}{\sqrt{n}}\right)$

Shortcuts for Calculating Confidence Interval

Level of Confidence	Value of z
90 %	1.64
95 %	1.96
99 %	2.58

You took a sample of 50 Gems and found that in the sample, the proportion of red Gems is 0.25. Construct a 99% confidence interval for the proportion of red Gems in the population.

$$0.25 - 2.58 * \sqrt{\frac{0.25 * 0.75}{50}} < p < 0.25 + 2.58 * \sqrt{\frac{(0.25 * 0.75)}{50}}$$

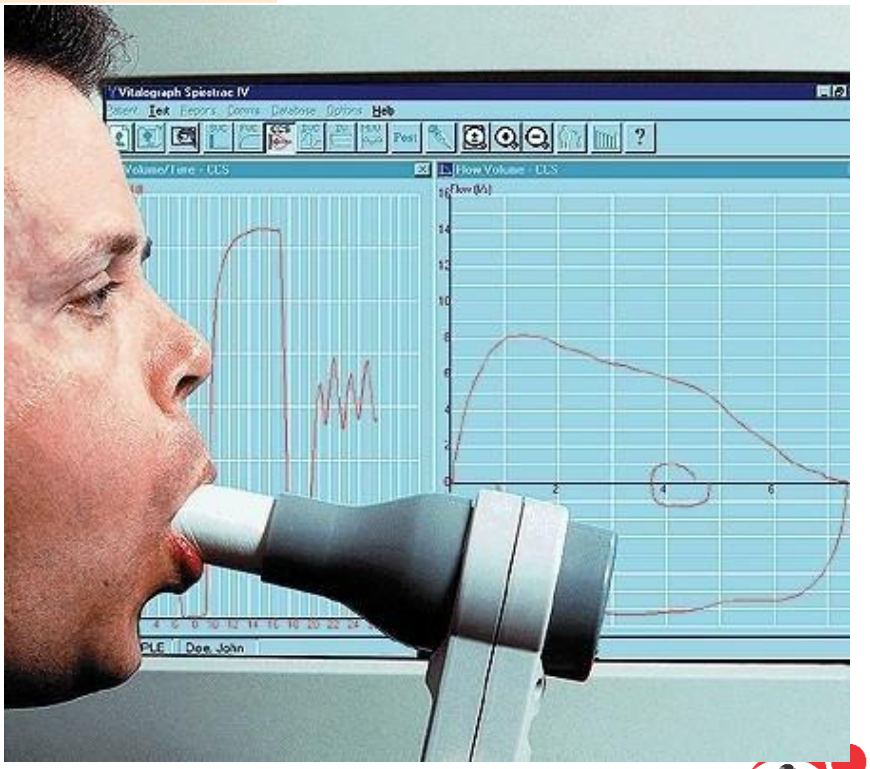
.

$$0.09 < p < 0.41$$

Shortcuts for Calculating Confidence Interval

Level of Confidence	Value of z
90 %	1.64
95 %	1.96
99 %	2.58

The lung function in 57 people is tested using FEV1 (Forced Expiratory Volume in 1 Second) measurements. The mean FEV1 value for this sample is 4.062 liters and standard deviation, *s* is 0.67 liters. Construct the 95 % confidence Interval.



FEV1 values of 57 male medical students

Level of Confidence	Value of Z
90 %	1.64
95 %	1.96
99 %	2.58

2.85	2.85	2.98	3.04	3.10	3.19	3.20	3.30	3.39
3.42	3.48	3.50	3.54	3.57	3.60	3.60	3.69	3.70
3.70	3.75	3.78	3.83	3.90	3.96	4.05	4.08	4.10
4.14	4.16	4.20	4.20	4.30	4.30	4.32	4.44	4.47
4.47	4.50	4.50	4.56	4.68	4.70	4.71	4.80	4.90
5.00	5.10	5.10	5.30	5.43				

$$95\%\ CI: (4.062\ - 1.96\ * \frac{0.67}{\sqrt{50}}\ , 4.062 + 1.96\ * \frac{0.67}{\sqrt{57}}$$

$$= (3.89, 4.23)$$

Attention Check

What happens to confidence interval as confidence level changes?

As confidence level increases, the confidence interval becomes wider and *vice-versa.*

What happens to the confidence interval as sample size changes ?

As sample size increases, the confidence interval become narrower.

Remember $\left(\bar{X} - z\frac{\sigma}{\sqrt{n}}, \bar{X} + z\frac{\sigma}{\sqrt{n}}\right)$.

Confidence Intervals for a Sample Median

Confidence limits are given by actual values in the sample using the following formulae:

Lower 95 % Confidence Limit: $\frac{n}{2} - 1.96 * \frac{\sqrt{n}}{2}$ ranked value.

Upper 95 % Confidence Limit: $1 + \frac{n}{2} + 1.96 * \frac{\sqrt{n}}{2}$ ranked values.

Confidence Intervals for a Sample Median

2.85	2.85	2.98	3.04	3.10	3.10	3.19	3.20	3.30	3.39
3.42	3.48	3.50	3.54	3.54	3.57	3.60	3.60	3.69	3.70
3.70	3.75	3.78	3.83	3.90	3.96	4.05	4.08	4.10	4.14
4.14	4.16	4.20	4.20	4.30	4.30	4.32	4.44	4.47	4.47
4.47	4.50	4.50	4.56	4.68	4.70	4.71	4.78	4.80	4.80
4.90	5.00	5.10	510	5.20	5.30	5.43			

Lower 95 % CI

Median

Upper 95 % CI

Lower 95 % confidence Limit: $\frac{57}{2} - 1.96 * \frac{\sqrt{57}}{2} = 21.10$ ranked value. 21st ranked value in 3.70

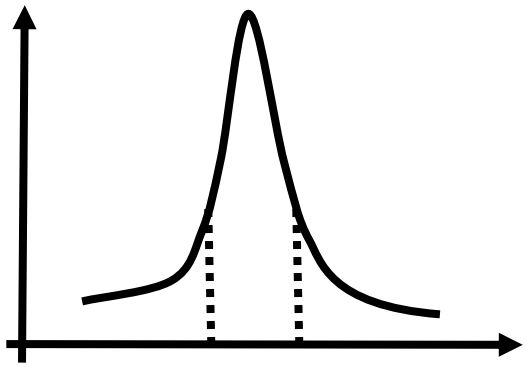
Upper 95 % confidence Limit: $1 + \frac{57}{2} + 1.96 * \frac{\sqrt{57}}{2} = 36.90$ ranked values. 37th ranked value is 4.32.

95 %CI: (3.70, 4.32) *Recall 95% CI using Mean: (3.89, 4.23)*

Confidence Intervals for a Sample Median

- Lack of distributional assumptions makes it difficult to obtain an exact CI for the median.
- CI are not necessarily symmetric around the sample estimate.

The Summary of CI



Confidence Interval = Sample statistics $\pm$ Margin of Error

$$\left(\bar{X} - z\frac{\sigma}{\sqrt{n}}, \bar{X} + z\frac{\sigma}{\sqrt{n}}\right)$$

Margin of error = z * Standard Error *(Recall the standardization formula)*

Depends on the Confidence level

Probability density.
Area under the curve between the limits.
Probability that a certain % of samples will contain the population mean within this interval.

$$\frac{\sigma}{\sqrt{n}}$$

Standard deviation of the population: Measure of deviation from the mean

A Short detour – Variance Formula Differences

Population Parameter

$$\mu = \frac{\sum x}{N}$$

$$\text{Variance } \sigma^2 = \frac{\sum (x-\mu)^2}{N}$$

Sample Statistic

$$\bar{x} = \frac{\sum x}{N}$$

$$\text{Variance } S^2 = \frac{\sum (x-\bar{x})^2}{n-1}$$