CONFIDENCE LEVELS AND CONFIDENCE INTERVALS

Exit polls see photo finish

Today's Chanakya predicts landslide victory for BJP, allies

DC CORRESPONDENT
NEW DELHI, NOV. 5

The Bihar Assembly elections appear to be living up to its billing as something of a cliffhanger, with a series of exit polls offering varied verdicts. At the end of the fifth and final round of voting on Thursday, the pollsters released their predictions, with four of them favouring the Grand Secular Alliance and two betting on the NDA.

Today's Chanakya, that hit the bull's eye in predicting the spectacular maiden electoral entry of Arvind Kejriwal's Aam Aadmi Party in 2013, and later followed it up by coming bang on in the 2014 Lok Sabha polls, forecast a near two-thirds majority for the BJP-led NDA, saying it could win 155 seats in an Assembly of 243. The pollster also claimed the Grand Alliance could win 83 seats.

In contrast, C-Voter, Neilsen and NewsX gave clear wins for the Grand Alliance led by Chief Minister Nitish Kumar. However, C-Voter also noted that the BJP would be the single largest party in the Assembly, winning 91 seats. The pollster's verdict is a close call, with the Grand Alliance pegged to win about 122 seats, against 111 for the NDA.

NewsX predicted 130 seats for the Grand Alliance against 90 for the NDA. The India Today-Cicero poll predicted 113 to 127 seats for the NDA and 111 to 123 for the JD(U)-led alliance. India TV forecast 112 to 132 seats for the JDU-led alliance and 101 to 121 for the NDA.

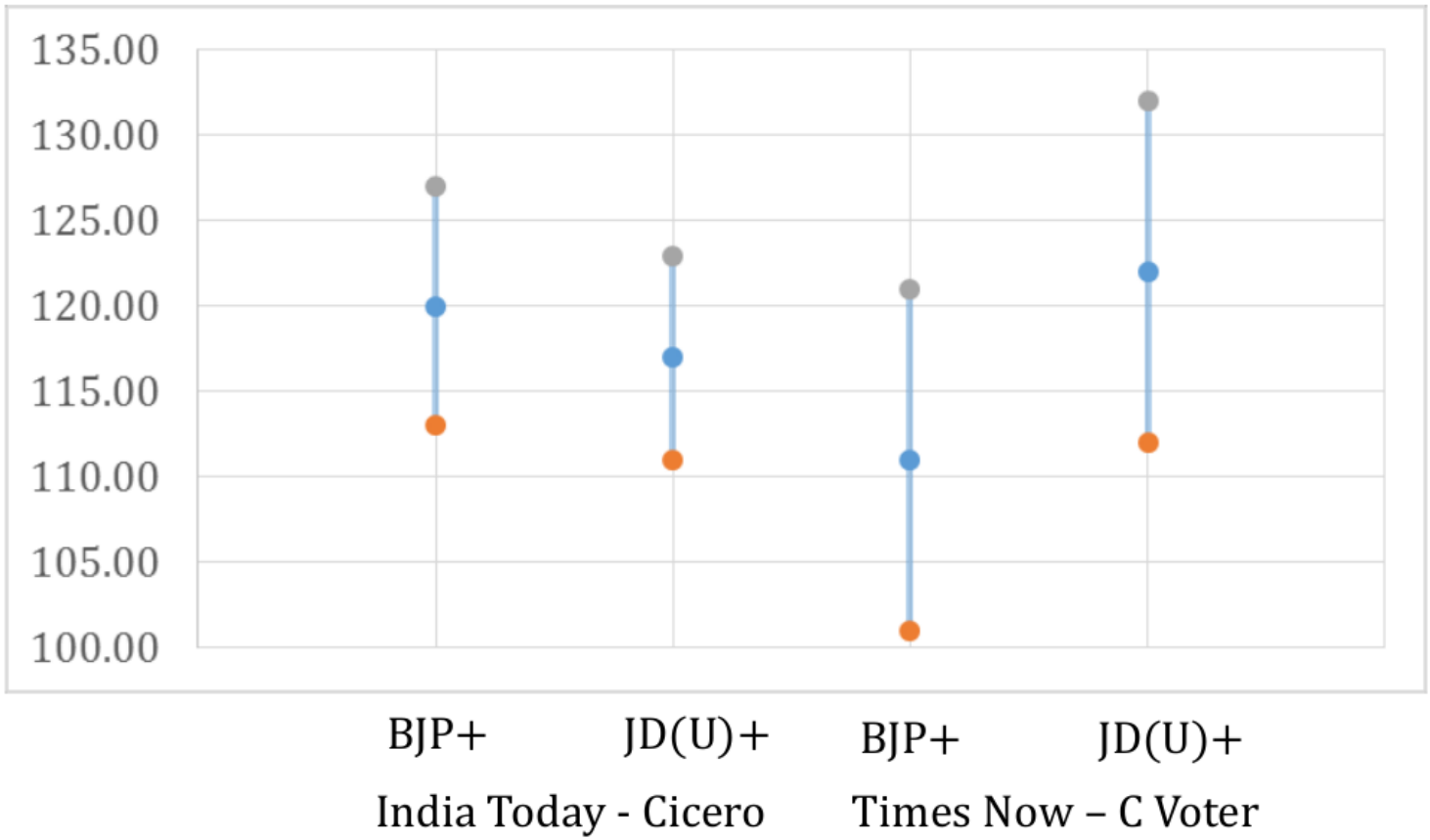
In the 2010 Bihar elections, the NDA, comprising the JD(U) and BJP, had swept the polls, winning 115 and 91 seats respectively, with the RJD getting 22 seats and the Congress just four.

DEAD HEAT

	India Today-Cicero	News X-CNX	ABP-Nielsen	Today's Chanakya	Times Now-C Voter	Average
BJP+	120 ±7	95	108	155	111 ±10	118
JD(U)+	117 ±6	130	130	83	122 ±10	118
Others	6	13	5	5	10	7

VOTING FOR BIHAR ASSEMBLY ELECTIONS

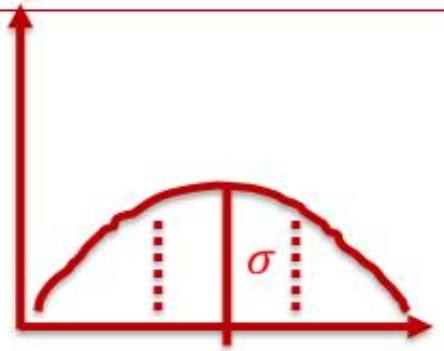
Phase	Date	No.of districts	No. of constituencies (Total-243)	Turn out	NO. OF SEATS CONTESTED			
					NDA	Grand Secular Alliance	Others	Total
1	Oct. 12	10	49	57%	21	28	0	49
2	Oct. 16	6	32	55%	12	19	1	32
3	Oct. 28	6	50	53.32%	28	22	0	50
4	Nov. 1	7	55	57.59%	30	22	3	55
5	Nov. 5	9	57	60%	32	24	1	57


135.00
130.00
125.00
120.00
115.00
110.00
105.00
100.00
BJP+
JD(U)+
BJP+
JD(U)+
India Today - Cicero
Times Now – C Voter

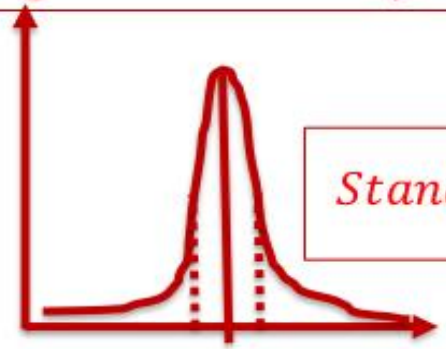
Polling Organisation	NDA	UPA	Other
CNN-IBN – CSDS – Lokniti	**276 (±6)**	97 (±5)	148 (±23)
India Today – Cicero	**272 (±11)**	115 (±5)	156 (±6)
News 24 – Chanakya	**340 (±14)**	70 (±9)	133 (±11)

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

Population distribution

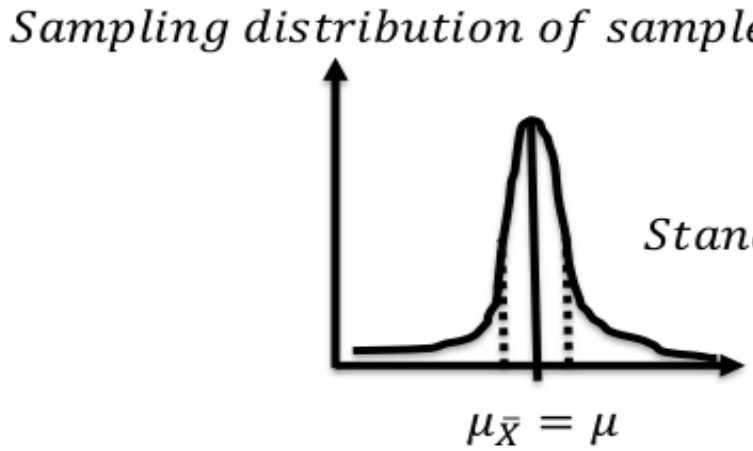


Sampling distribution of sample means



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

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.



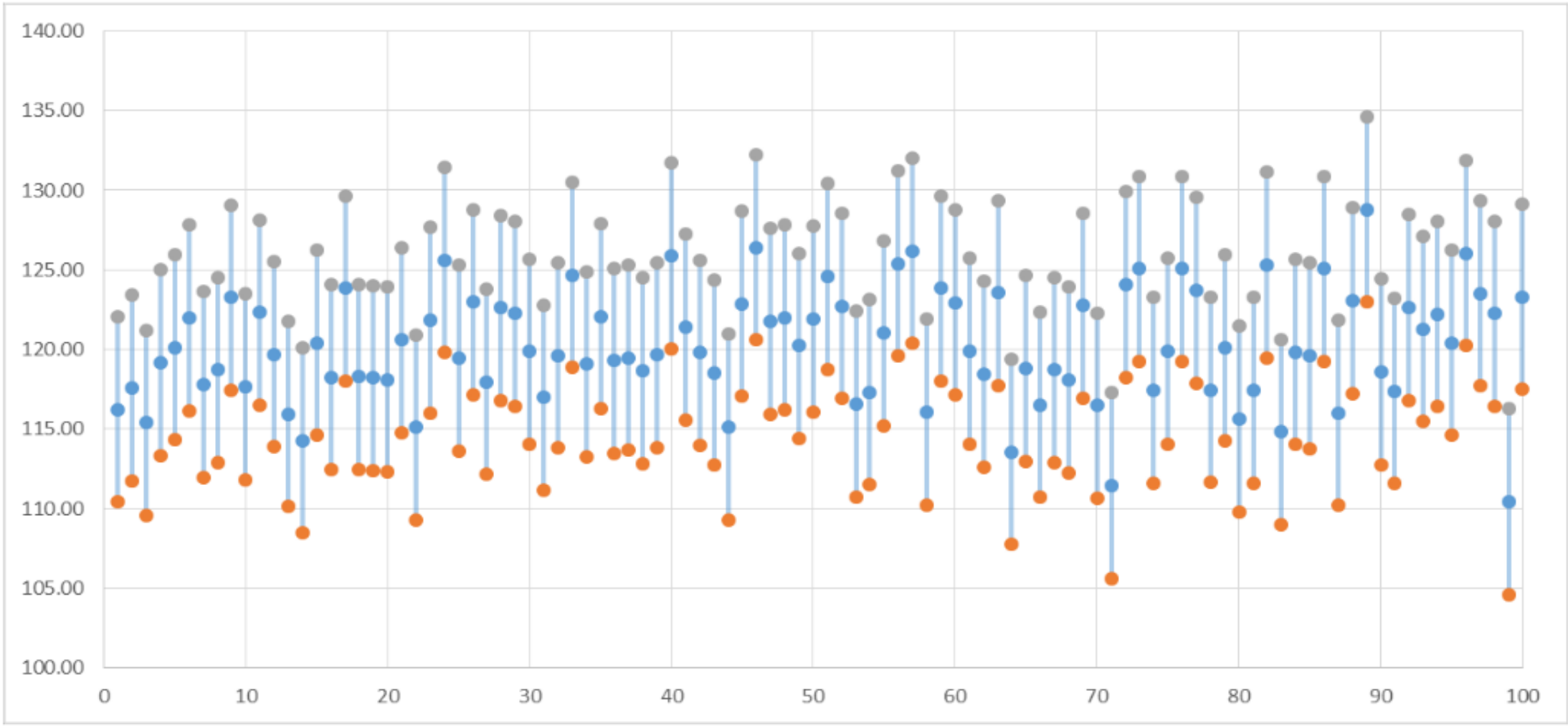
$$Standard\ Error\ of\ the\ Mean = \frac{\sigma}{\sqrt{n}}$$

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.

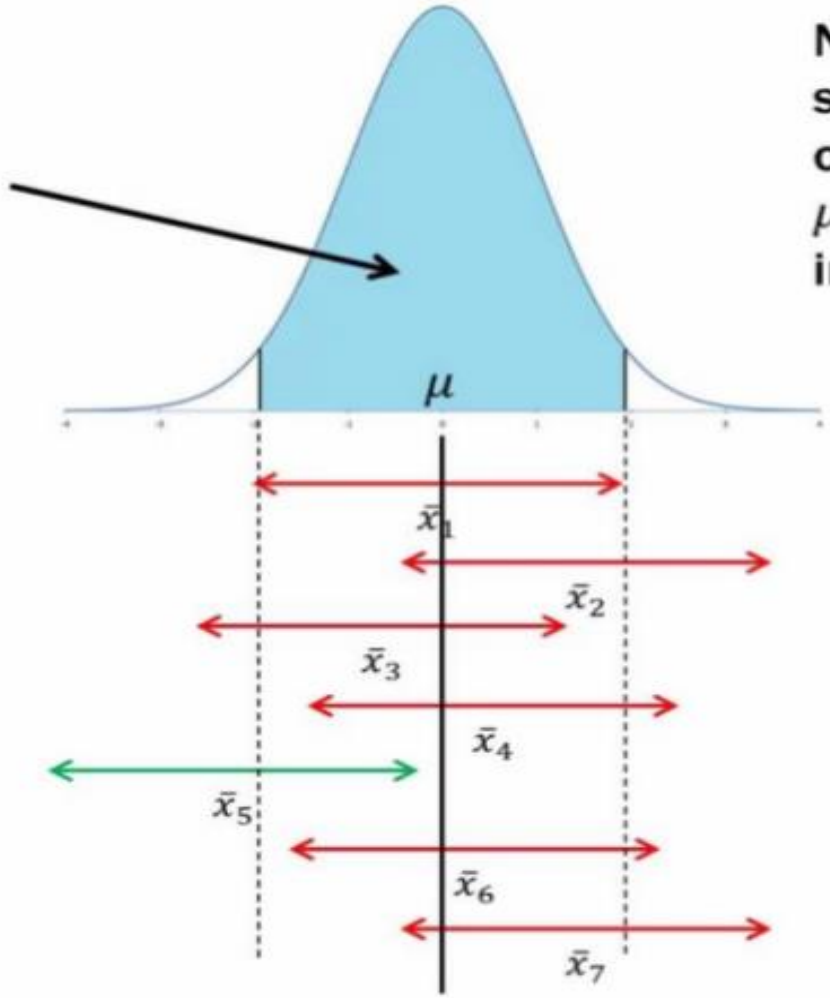
Confidence Level and Interval - Excel



94 of the 100 intervals contain the population mean.

Confidence Level and Interval

95% of all sample means $(\bar{x})$ are in here. THEN THIS.



Notice that as soon as a sample mean steps outside the dotted line, μ is no longer in its interval.

Many samples of the same size. THESE COME FIRST.

Samples of the same size have the same standard error $\sigma_{\bar{x}}$. So the 95% "width" is the same for all samples of that size.

HUFFPOST POLLSTER

FIND A CHART | Try 'New York,' 'Obama' or 'health care'

AdChoices

← PREVIOUS POLL | NEXT POLL →

POLL UPDATE

2016 National Republican Primary - Trump 40%, Cruz 17% (Ipsos/Reuters (Web) 2/13-2/17)

Population	1,473 Adults
Margin of Error	±2.9 percentage points
Polling Method	Internet
Source	Ipsos/Reuters [PDF]

This poll asked respondents 2 questions tracked by HuffPost Pollster. Read our FAQ.

Margin of Error is the range of expected variation for a given survey result or, more specifically, to how confident we can be that, if repeated using the same methodology, the results of a survey would fall within that range of variation.

1) 2016 National Republican Primary

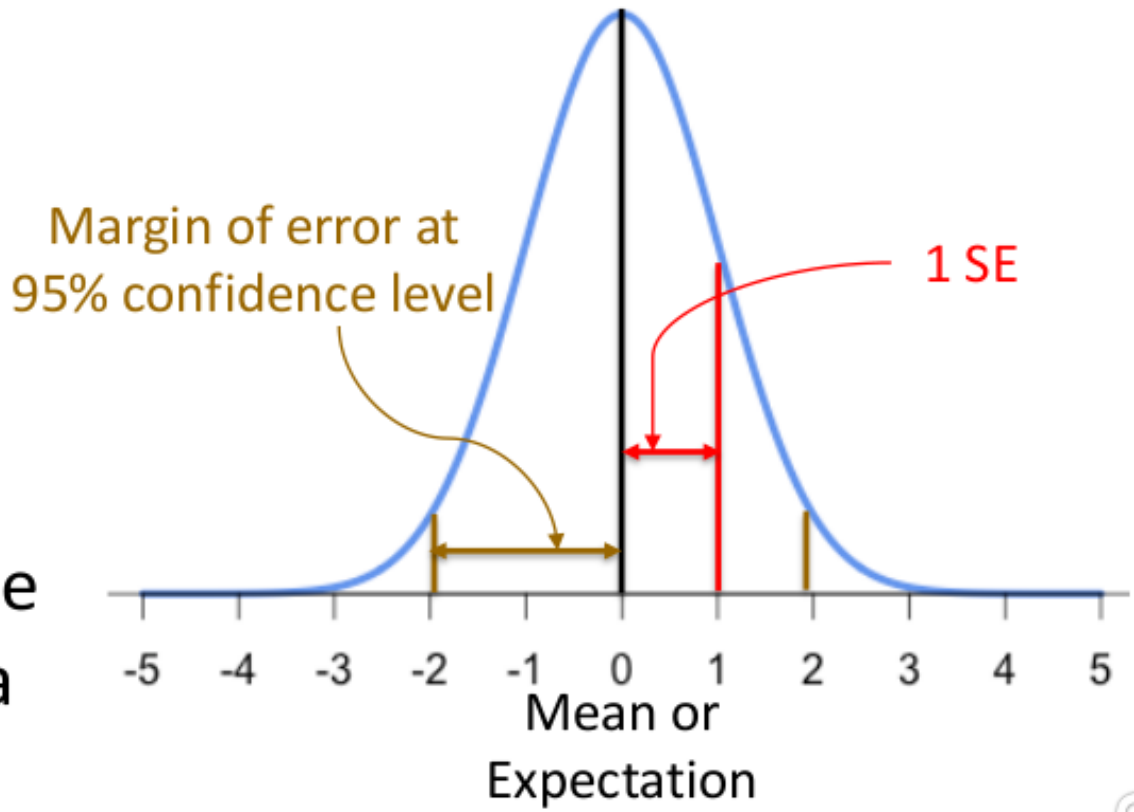
Asked of 476 Republican registered voters

Jeb Bush (R)	9%
Ben Carson (R)	10%
Ted Cruz (R)	17%
Jim Gilmore (R)	1%
John Kasich (R)	9%
Marco Rubio (R)	11%
Donald Trump (R)	40%
Wouldn't vote	4%

SE, Margin of Error, Confidence Interval and Sample Size

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

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

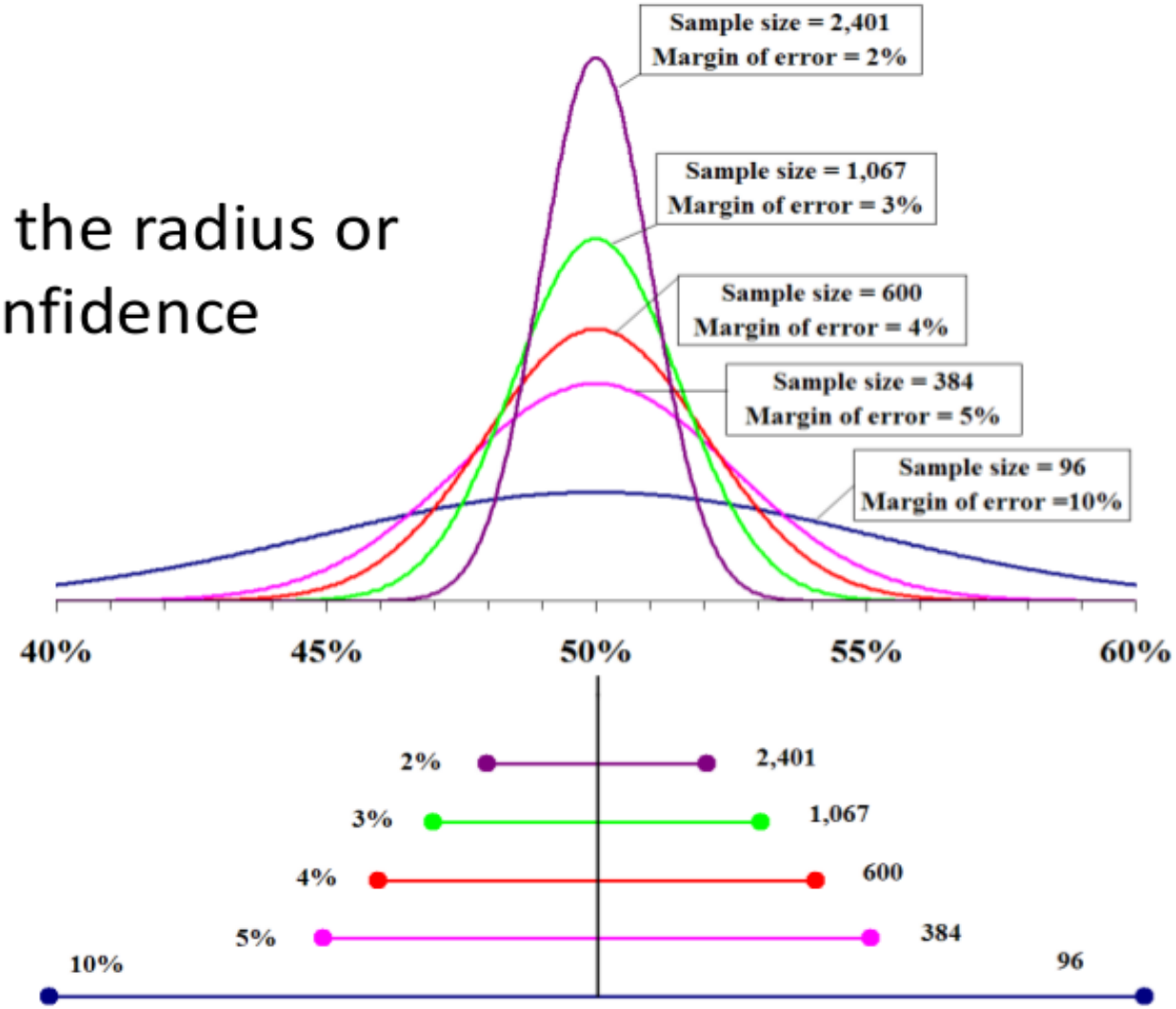


Margin of error is the **maximum expected difference** between the true population parameter and a sample estimate of that parameter.

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

SE, Margin of Error, Confidence Interval and Sample Size

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



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

In a poll by CNN/ORC conducted between November 27 – December 1, 2015, a survey of 930 randomly sampled registered voters predicted that 49% would vote for Hillary Clinton.

What is the margin of error at 95% confidence level (z = 1.96)?

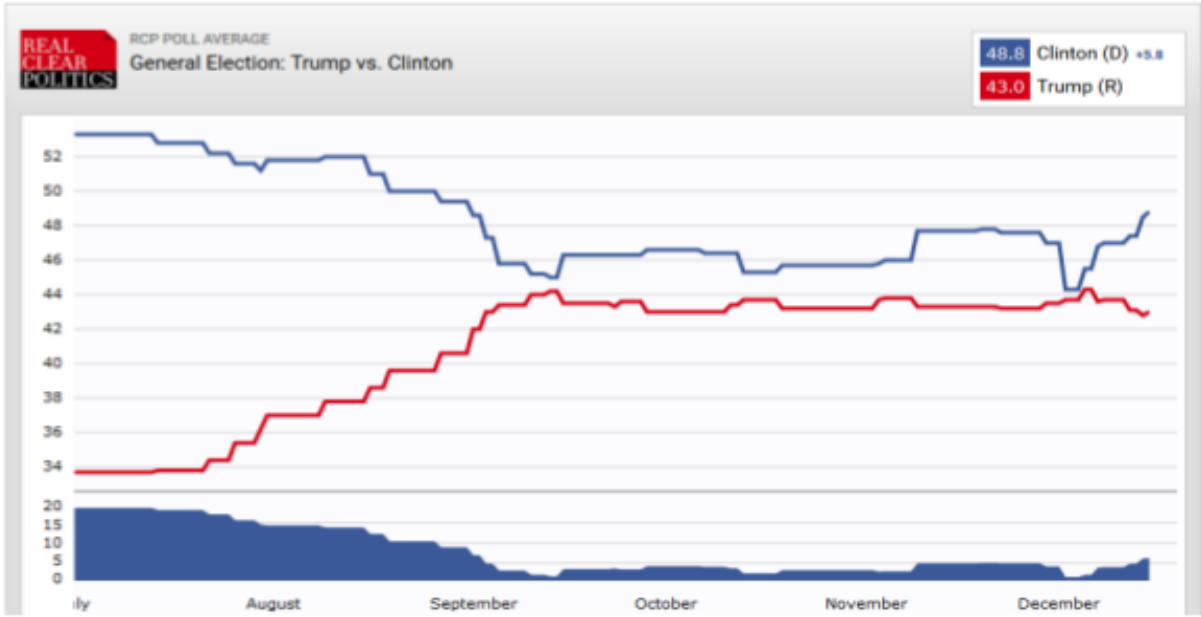
Check qnorm(0.975, 0, 1). Why 0.975?

RealClear Politics — Polls | Election 2016 | Video | Wri

Polling Data

Poll	Date	Sample	MoE	Clinton (D)	Trump (R)	Spread
RCP Average	11/23 - 12/13	--	--	48.8	43.0	Clinton +5.8
ABC News/Wash Post	12/10 - 12/13	851 RV	4.0	50	44	Clinton +6
NBC/WSJ	12/6 - 12/9	849 RV	3.4	50	40	Clinton +10
USA Today/Suffolk	12/2 - 12/6	1000 LV	3.0	48	44	Clinton +4
CNN/ORC	11/27 - 12/1	930 RV	3.0	49	46	Clinton +3
Quinnipiac	11/23 - 11/30	1473 RV	2.6	47	41	Clinton +6

All General Election: Trump vs. Clinton Polling Data



$$\text{Margin of error} = 1.96 * \sqrt{\frac{0.49 * 0.51}{930}} \approx 3.2\%$$

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?

$$0.01 = 1.96 * \sqrt{\frac{0.49 * 0.51}{n}}$$

$$\therefore n = \left(\frac{1.96}{0.01} * \sqrt{0.49 * 0.51}\right)^2 = 9600$$

Other ways of estimating the data size

- The rule of thumb
  - Count the total number of levels (assuming 10 levels for numeric)
  - Multiply with number classes
  - Multiply with 75-150

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

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

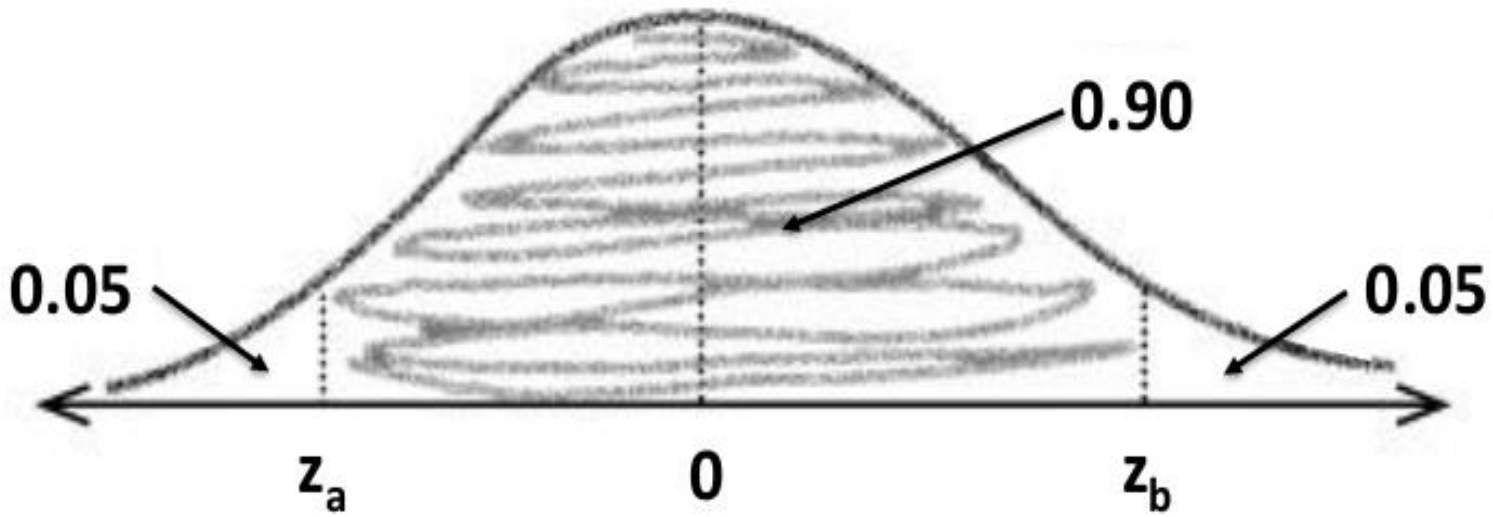
$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$ or Margin of error = $z * \frac{\sigma}{\sqrt{n}}$

$\therefore$ 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

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 Intervals

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

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

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

Since the sample mean is 10.455 years, we get the confidence interval for 90% as $8.545 < \mu < 12.365$.

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.

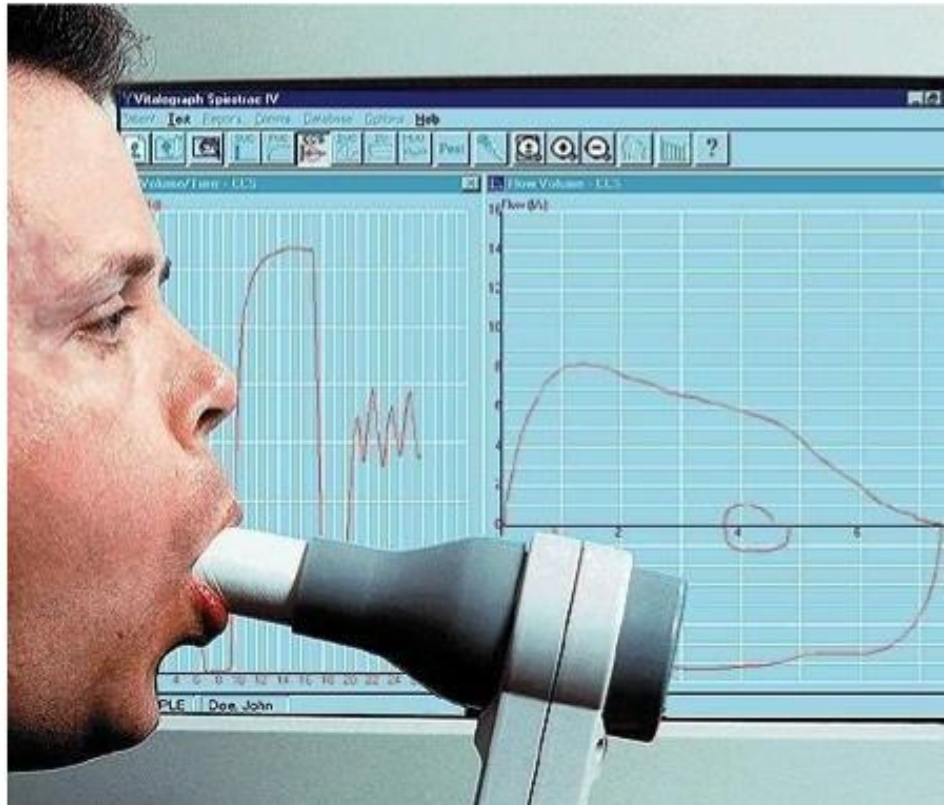
Shortcuts for Calculating Confidence Intervals

Population Parameter	Population Distribution	Conditions	Confidence Interval
μ	Normal	You know σ^2 n is large or small $\bar{X}$ is the sample mean	$(\bar{X} - z\frac{\sigma}{\sqrt{n}}, \bar{X} + z\frac{\sigma}{\sqrt{n}})$
μ	Non-normal	You know σ^2 n is large (> 30) $\bar{X}$ is the sample mean	$(\bar{X} - z\frac{\sigma}{\sqrt{n}}, \bar{X} + z\frac{\sigma}{\sqrt{n}})$
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	$(\bar{X} - z\frac{s}{\sqrt{n}}, \bar{X} + z\frac{s}{\sqrt{n}})$
p	Binomial	n is large p_s is the sample proportion q_s is $1 - p_s$	$(p_s - z\sqrt{\frac{p_s q_s}{n}}, p_s + z\sqrt{\frac{p_s q_s}{n}})$

Shortcuts for Calculating Confidence Intervals

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 litres and standard deviation, *s* is 0.67 litres. Construct the 95% Confidence Interval.



FEV1 values of 57 male medical students

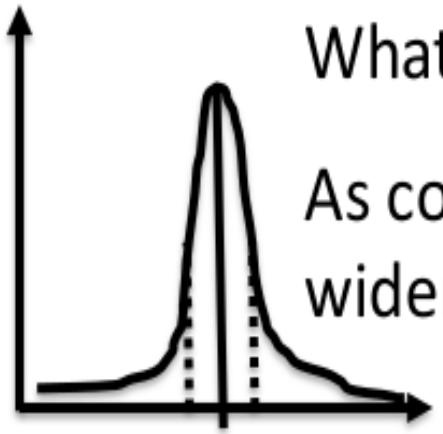
Level of confidence	Value of z
90%	1.64
95%	1.96
99%	2.58

2.85	2.85	2.98	3.04	3.10	3.10	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	5.10	5.20	5.30	5.43			

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

$$= (3.89, 4.23)$$

Attention Check



What happens to confidence interval as confidence level changes?

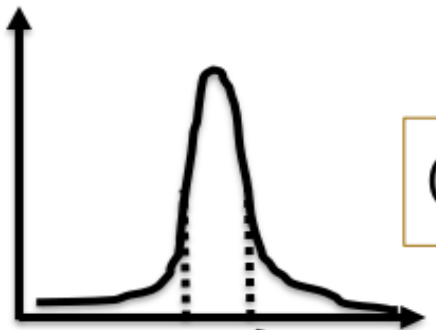
As confidence level increases, the confidence interval becomes wider and *vice-versa*.

What happens to the confidence interval as sample size changes?

As sample size increases, the confidence interval becomes narrower.

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

The Connection



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

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

$$Max(\bar{x} - \mu)$$

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

Depends on the confidence level

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

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

Standard deviation of the population: A sort of "average" deviation from the mean.