Confidence Intervals

by Sophia

WHAT'S COVERED

This tutorial will cover the basics of confidence intervals, focusing on how to identify the z-critical value needed for a given confidence interval. Our discussion breaks down as follows:

1. Confidence Intervals

Before we begin, it's important to note that sampling error is the inherent variability in the process of sampling. In a random sample, it occurs when you use a statistic, like a sample mean, to estimate the parameter, like a population mean. You won't always get exact accuracy with the sample mean, but you can use it to estimate the population mean. The idea is that you can be close.

When you take a larger sample, you're going to be, on average, closer. The sampling error, which is the amount by which the sample statistic is off from the population parameter, decreases. You get more consistently close values to the parameter when you take samples. When you calculate a margin of error in a study, you are approximating the sampling error.

When you take a sample, you try to obtain values that accurately represent what's going on in the population.

> EXAMPLE For example, suppose you took a simple random sample of 500 people getting ready for an upcoming election in a town of 10,000, and found that 285 of those 500 plan to vote for a particular candidate. Your best guess, for the true proportion, in the population of the town that will vote for candidate y is the proportion that you got in your sample--285 out of 500, which is 57% of the town. That's your best guess, but you might be off by a little bit.

You don't know if the true proportion of people who will vote for that candidate is 57%, and that's why you report a margin of error in your poll.

From the margin of error, you can create what is called a **confidence interval**. A confidence interval is an interval that contains the likely values for a parameter. We base the confidence interval on our point estimate, and the width of the interval is affected by the confidence level and sample size.

The confidence interval can be found using the following formula:

FORMULA TO KNOW

Confidence Interval

$$CI = Point\ Estimate \pm Margin\ of\ Error$$

The confidence interval is your point estimate, which is your best guess from your simple random sample, plus or minus the margin of error. You believe you are within a certain amount of the right answer with your point estimate.

TERM TO KNOW

Confidence Interval

An interval that contains likely values for a parameter. We base our confidence interval on our point estimate, and the width of the interval is affected by confidence level and sample size.

---

2. Margin of Error and the Affect of Confidence Level and Sample Size

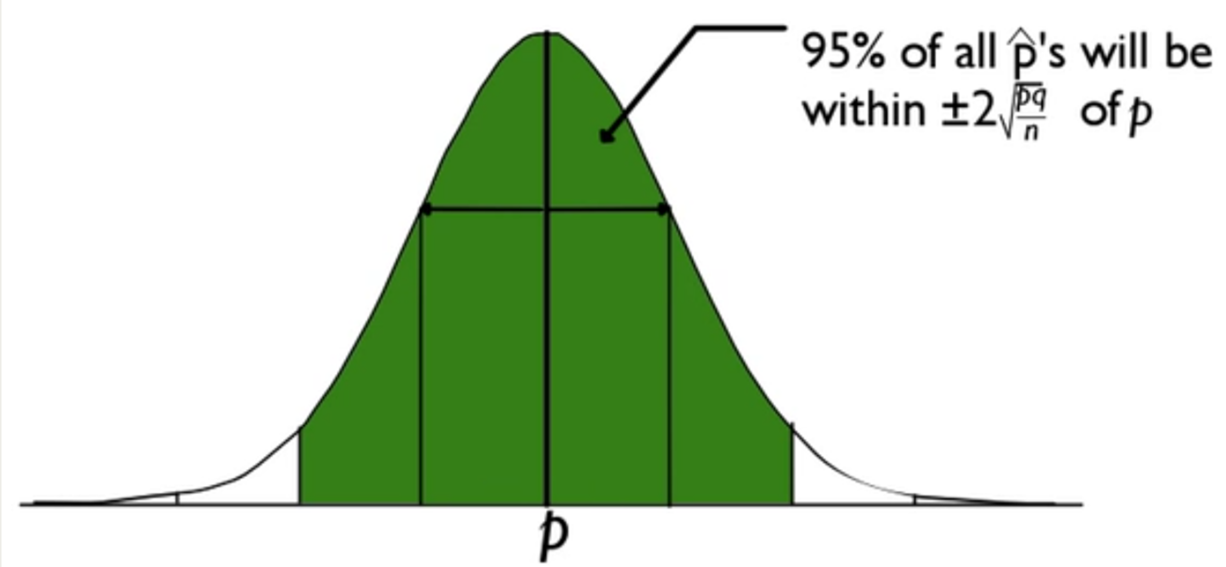
The margin of error depends on two things:

- Sample size: You knew this from before when you said that a larger sample size results in less sampling error, and therefore a lower margin of error.
- Confidence level: You're going to learn more about this, but a higher confidence level results in a larger margin of error.

EXAMPLE If you want to be very confident that you're going to accurately describe what percent of people are going to vote for that particular candidate, you have to go out a little bit further on each side. Maybe you have to go out plus or minus 5%, as opposed to plus or minus 3%.

IN CONTEXT

95% Confidence

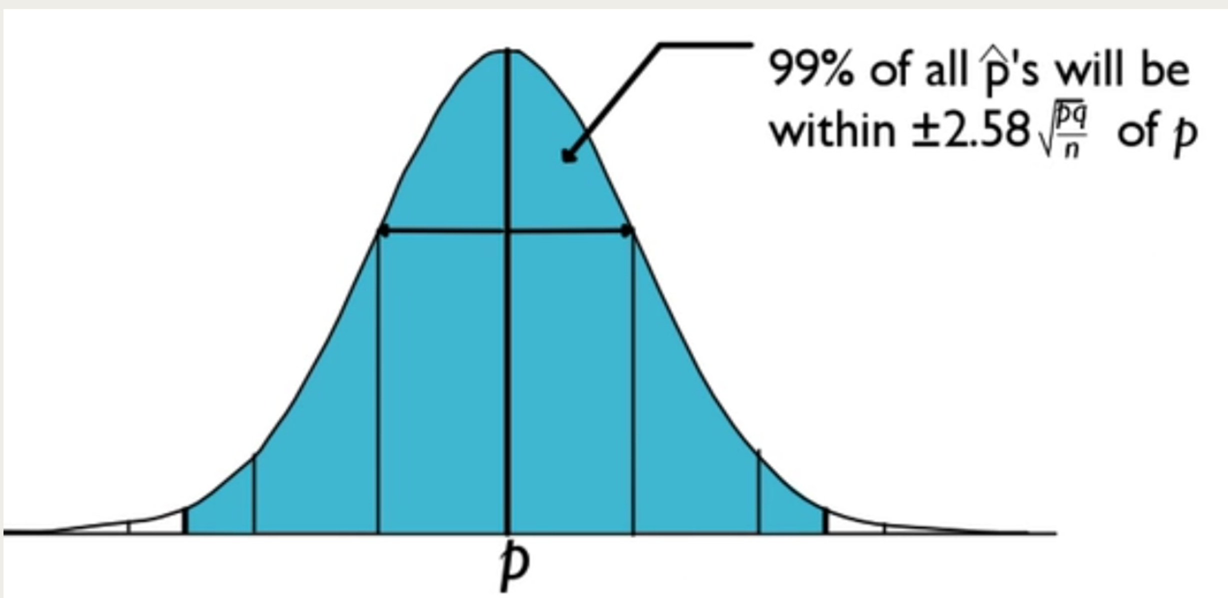


If the sampling distribution of p-hat is approximately normal, it will be centered at p, the population parameter. 95% of all sample proportions will be within two standard deviations of p.

So p plus or minus two standard deviations will contain 95% percent of all p-hat. This is called 95% confidence. Approximately 19 out of every 20 samples, in the long term, that you take will be within two standard deviations of the right answer. 95% percent of all p-hats are within two standard deviations of p.

If you want to be more confident, you can go out even further.

99% Confidence



For instance, 99% of all p-hats will be within 2.58 standard deviations of p. This means that when you take a sample proportion, 99% of sample proportions will be within 2.58 standard deviations of the right answer, the value of p.

> Take your p-hat value, and plus or minus 2.58 standard deviations, and you're 99% likely to capture the value of p.

---

3. Confidence Interval Formulas

When stating the confidence interval, we will use the following phrase:

> *In **C%** of samples, the **parameter** will be within **z* standard errors** of the **sample statistic**.*

Use this language to describe your confidence interval. The bold words will be replaced with your numbers.

3a. For Sample Means

What does this look like if you're using means? If you are using means, it looks like that Mu (μ), the parameter, will be contained in the interval statistic, which is x-bar, plus or minus z*, times the standard error of the statistic. In other words,

> *For C% of the time, the parameter μ will be contained in the interval* $\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}$.

FORMULA TO KNOW

Confidence Interval of Means

$$CI = \bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}$$

3b. For Sample Proportions

If you're using proportions, that means that the sample proportion, p-hat, plus or minus z* times the standard error, will contain the value of the parameter, p, some percent of the time, such as 95% or 99% of the time.

> *For C% of the time, the parameter p will be contained in the interval* $\hat{p} \pm z^* \sqrt{\frac{pq}{n}}$.

FORMULA TO KNOW

Confidence Interval of Proportions

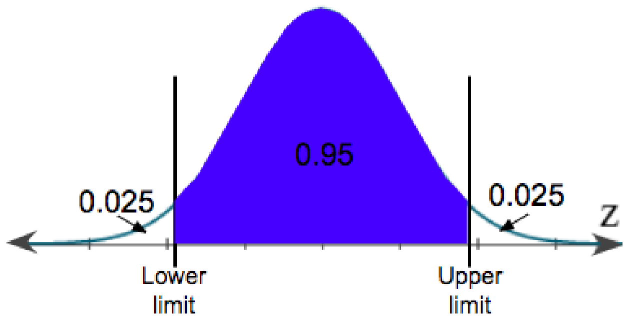
$$CI = \hat{p} \pm z^* \sqrt{\frac{pq}{n}}$$

---

4. Finding Z*

The confidence level determines the value of z*. Depending on what you choose for your confidence level, z* will be affected that way. To find the z* critical value, we can use a z-table. For a confidence interval, we can follow the same steps as a two-sided test.

EXAMPLE If we have a 95% confidence interval, this is actually the same as a 5% significance level. However, this is split between two tails, the lower and upper part of the distribution. Each tail will have 2.5%, or 0.025.



We can use the upper limit to find the critical z-score. Remember, a distribution is 100%, so to find the upper limit, we can subtract 0.025 from 1, which gives us 0.975. Now, we can use a z-table.

Standard Normal Distribution
Z-Table

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441

1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	**0.9750**	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890

In a z-table, the value 0.975 corresponds with a 1.9 in the left column and 0.06 in the top row. This tells us that the z-score is 1.96.

Another way is to use a t-table, which you will learn more about in a later tutorial. We don't use the t-distribution for proportions; however, we can use the last row in this table to find the confidence levels.

t-Distribution Critical Values												
Tail Probability, p												
One-tail	0.25	0.20	0.15	0.10	0.05	0.025	0.02	0.01	0.005	0.0025	0.001	0.0005
Two-tail	0.50	0.40	0.30	0.20	0.10	0.05	0.04	0.02	0.01	0.005	0.002	0.001
df												
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	0.816	1.080	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073

16	0.690	0.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.610	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485	3.767
24	0.685	0.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756	3.038	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
1000	0.675	0.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	3.300
>1000	0.674	0.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291
	Confidence Interval between -t and t											
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%

Z confidence level, critical values, are found in the last row of this t-table, under the infinity value, or ">1000". Essentially the normal distribution is the t distribution with infinite degrees of freedom. We're going to look in this row to find the z critical value that we should use, which is the same as the 1.96 we previously got.

SUMMARY

When you take a sample, you obtain a sample statistic that is a point estimate of the population parameters. You can create a confidence interval where you can be a certain percent confident that the parameter lies within the interval. This means that the percent of sample statistics in the sample distribution are within the margin of error of the parameter. Perhaps you'll say 95% of all the x-bars in the sampling distribution of x-bar will be within the margin of error of the true parameter Mu. That percent of confidence intervals will contain the parameters. If you did samples over and over again, and took confidence intervals each time, 90% or 95% of confidence intervals would contain the answer of Mu or p, or whatever parameters you're trying to estimate.

Good luck!

Source: THIS TUTORIAL WAS AUTHORED BY JONATHAN OSTERS FOR SOPHIA LEARNING. PLEASE SEE OUR **TERMS OF USE**.

TERMS TO KNOW

Confidence Interval

An interval that contains likely values for a parameter. We base our confidence interval on our point estimate, and the width of the interval is affected by confidence level and sample size.

FORMULAS TO KNOW

Confidence Interval

$$CI = \text{Point Estimate} \pm \text{Margin of Error}$$

Confidence Interval of Means

$$CI = \bar{x} \pm z^* \sigma / \sqrt{n}$$

Confidence Interval of Proportions

$$CI = \hat{p} \pm z^* \sqrt{\frac{pq}{n}}$$

