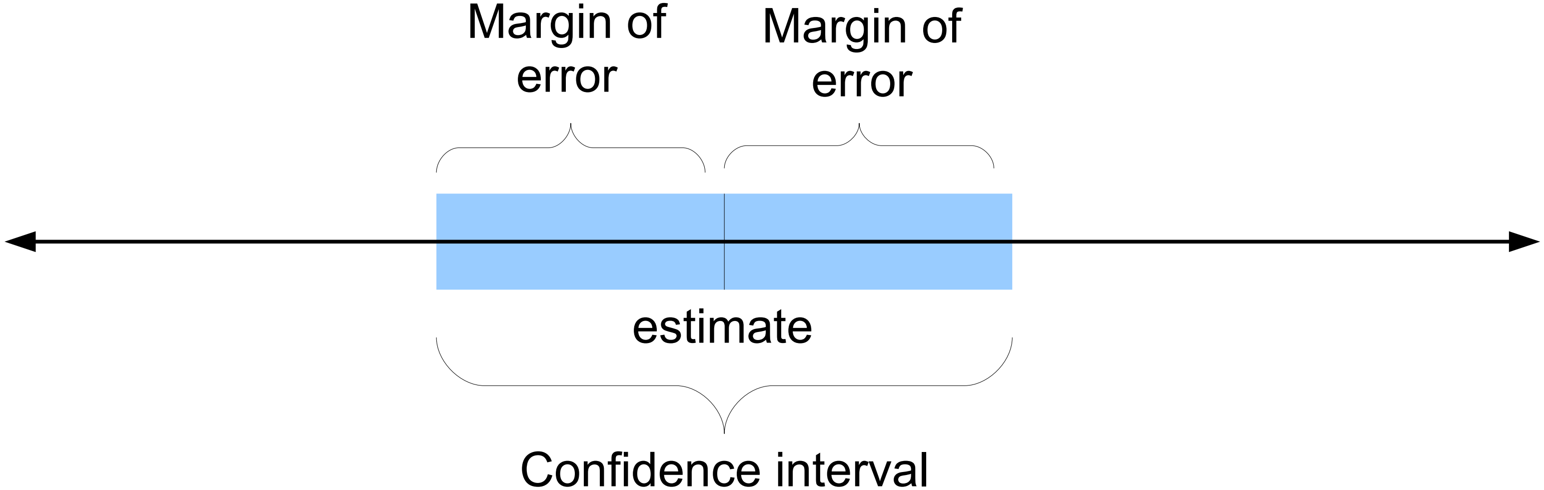
Confidence Intervals

- If we know that each member of the population has probability *p* of having a certain characteristic, we can use the CLT theorem to study the distribution of a sample mean.
- What if we don't know p, all we have is our data from the sample. We want to make an estimate of p, and give some margin of error. This is essentially what a <u>confidence interval</u> is.
- For a prescribed level of confidence (less than 100%), we want to determine **a range** for which we are THAT confident the true population probability "p" is within the range.

Confidence Intervals, cont.

- Usually we want a fairly high confidence level: 75%, 95% or 99% are common, but really any percentage less than 100 is possible. The larger the confidence, the wider the interval.
- The more sure we are of the confidence interval, the less precise it is.



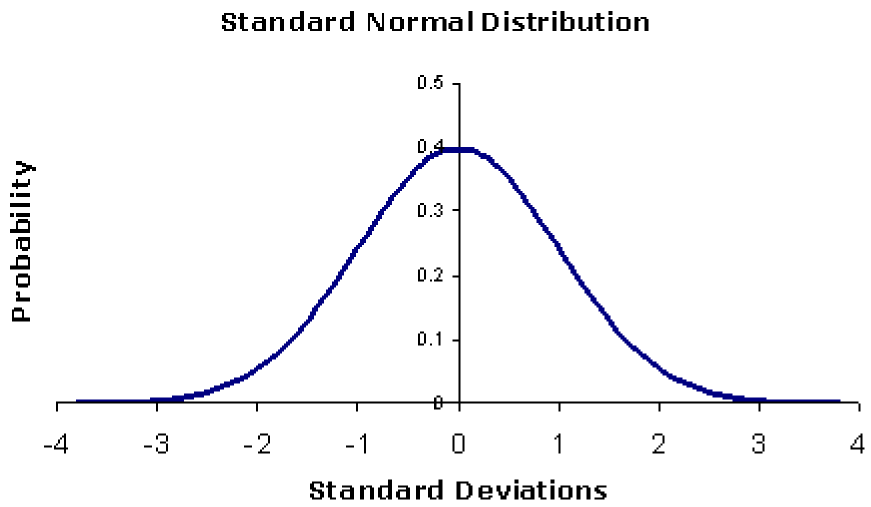
Confidence Interval for Proportion

- p is the population proportion (of a certain characteristic)
- To find a C% confidence interval, we need to know the z-score of the central C% in a standard-normal distribution. Call this 'z'
- Our confidence interval is $\hat{p} \pm z * SE(\hat{p})$
- $\hat{p}$ is the sample proportion
- $SE(p) = \sqrt{(\hat{p}(1-\hat{p})/n}$

Z values for some CIs

- For your reference, these could be useful:

Confidence %	# standard deviations (z)
50%	0.67449
75%	1.15035
90%	1.64485
95%	1.95996
97%	2.17009
99%	2.57583
99.9%	3.29053



To calculate, use
invNorm(CI + (1-CI)/2)

e.g. for 75% confidence,
invNorm(.75 + (1-.75)/2)
=invNorm(.75+ .25/2)
=invNorm(.875)

Example: Bad Apples

You want to give a 95% confidence interval of how many apples in a given orchard are bad this year. Of all harvested apples, you randomly test 1000 apples and find 35 of them are bad.

- p estimate is $\hat{p}$=.035, so $\hat{q}$=.965
- SD($\hat{p}$)=√(.035*.965/1000)=.0058
- The middle 95% is within 1.96 sds
- Our confidence interval is .035±1.96*.0058, i.e. **between and .0236 and .0464**
- **We are 95% confident that in this orchard between 2.36% and 4.64% of apples are bad.**

Margin of Error

- Based on a certain % confidence interval, the amount we add/subtract from our estimate is the **margin of error**.
- In the previous example, the margin of error was 1.96*.0058=.011368 which is 1.1368%
- For C% confidence, $ME_C = z_C\sqrt{(pq/n)}$

Example: Margin of Error

A poll of 1654 adults asked whether they have ever bobbed for apples. 76% said “Yes.”

For 93% confidence, what is the margin of error?

- To find the z-score for the central 93%, remember that 7% is in the tails, 3.5% in the upper tail and 3.5% in the lower tail. So invNorm(.965)=1.812 is our z
- $ME_{93\%} = z\sqrt{(pq/n)} = 1.812 * \sqrt{(.76 * .24/1654)}$ $= .01903$, or 1.903%

Example: Margin of Error

A poll of 1654 adults asked whether they have ever bobbed for apples. 76% said “Yes.”

What is the margin of error for 99% confidence?

- Similarly, the z value for central 99% is invNorm(.995)=2.576
- $ME_{99\%}$=2.576*.010501=.02705 or 2.705%
- **As confidence level of the interval increases, so does the margin of error!**

Example: Determine Sample Size for given Confidence & ME

It is estimated that 43% of adults 25-35 sing in the shower. We want to see if this is true for adults 35 and older. How many do we need to sample to have a margin of error of 5% at a 90% confidence level.

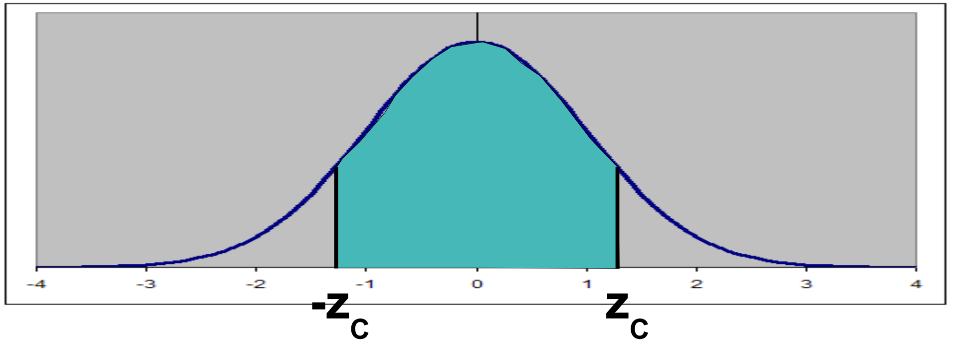
- $ME=z\sqrt{(pq/n)}$, z for 90% confidence is 1.64485
- $.05=1.64485\sqrt{(.43*.57)}/\sqrt{n}$
- So $\sqrt{n}=1.64485\sqrt{(.43*.57)}/.05 = 16.2865$
- So $n=16.2865^2=265.25$, so we need 266 people sampled (round up to the next whole person)

Decreasing Margin of Error by increasing n

- For C% confidence, $ME_c = z_c\sqrt{(pq/n)}$
- If we increase the sample size, the margin of error goes down, but at a rate of the square root of the change in “n”.
- To **halve** ME, we need to **quadruple (x4)** the sample size
- To get **$1/10^{th}$** the ME, we need to increase sample size to be **100 times** as large

Determine CI from Margin of Error

- You can use the formula $ME_c = z_c\sqrt{(pq/n)}$ to give you the confidence level, because you can determine z_c, and from that figure out the confidence level.
- Divide both sides by $\sqrt{(pq/n)}$ to give you: $z_c = ME_c/\sqrt{(pq/n)}$
- Then Confidence level is found: normalcdf($-z_c$, z_c)



Example: Approval Rating

The results of a poll of 656 random citizens give the mayor's approval rating at 67% with a 4% margin of error. How confident are we that the city-wide approval is between 63% and 71%?

- $z = .04/\sqrt{(.67 * .33/656)} = 2.1788$
- normalcdf(-2.1788, 2.1788) = .97065
- So the poll used a 97.065% confidence level.