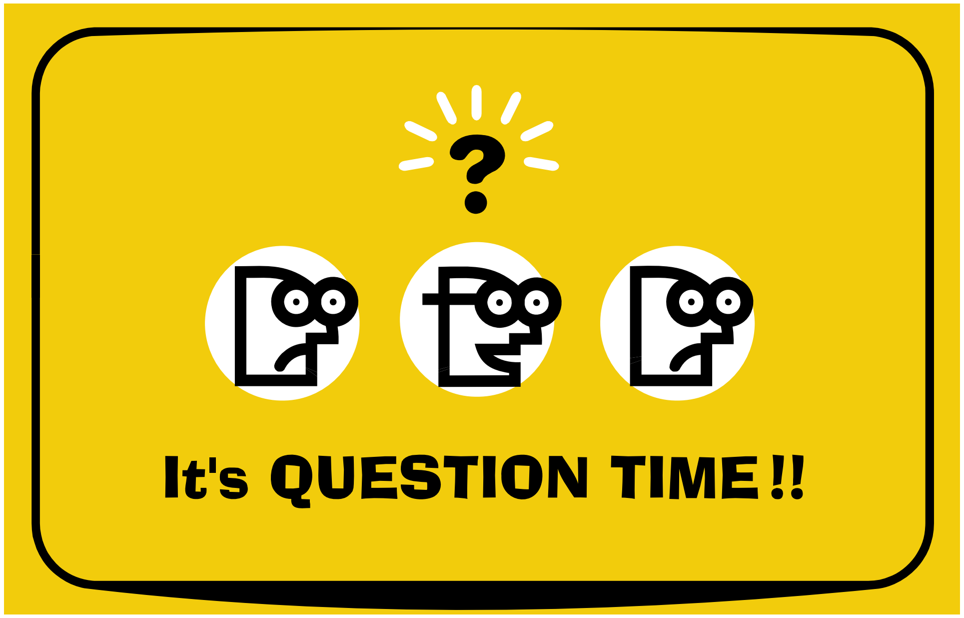
# ESTIMATES



The best single estimate for a population parameter is the corresponding sample statistic.  Such an estimate is called a point estimate.

$\overline{x}$ is the best point estimate for $\mu$

$s$ is the best point estimate for $\sigma$

$\hat{p}$ is the best point estimate for $p$

Often, though, we will want to find an interval that we are confident that our population parameter lies within. Such an estimate is called an interval estimate, and the resulting interval is called a confidence interval.

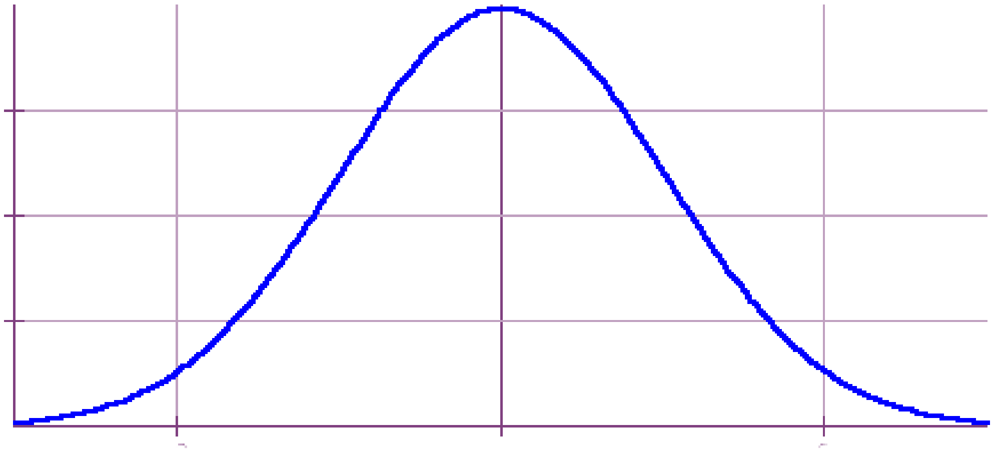
$$69.362 < \mu < 78.638$$

Let's suppose that we take a sample of size *n* > 30 and mean *x-bar*, and that we want to find a 95% confidence interval for the population mean, *mu*.

$$69.362 < \mu < 78.638$$

95% confidence interval

Recall that even if the population, itself, isn't normally distributed, the distribution of sample means of size *n > 30* for this population will be approximately normally distributed.

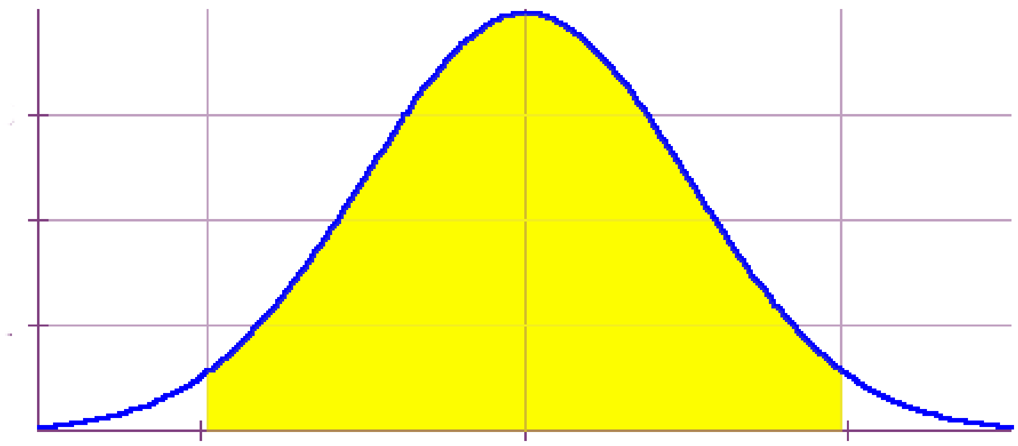


$$\mu_{\overline{x}} = \mu$$

$$\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}}$$

Furthermore, there is a 95% chance that our sample mean will lie within 1.96 standard deviations of the population mean.

$$z_{.05/2} = 1.96$$




95% chance that $\mu - 1.96 \cdot \frac{\sigma}{\sqrt{n}} < \overline{x} < \mu + 1.96 \cdot \frac{\sigma}{\sqrt{n}}$

We now just do a little algebra.

$$\mu - 1.96 \cdot \frac{\sigma}{\sqrt{n}} < \bar{x} < \mu + 1.96 \cdot \frac{\sigma}{\sqrt{n}}$$

$$-1.96 \cdot \frac{\sigma}{\sqrt{n}} < \bar{x} - \mu < 1.96 \cdot \frac{\sigma}{\sqrt{n}}$$

$$-\bar{x} - 1.96 \cdot \frac{\sigma}{\sqrt{n}} < -\mu < -\bar{x} + 1.96 \cdot \frac{\sigma}{\sqrt{n}}$$

$$\bar{x} + 1.96 \cdot \frac{\sigma}{\sqrt{n}} > \mu > \bar{x} - 1.96 \cdot \frac{\sigma}{\sqrt{n}}$$

$$\bar{x} - 1.96 \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1.96 \cdot \frac{\sigma}{\sqrt{n}}$$

The last inequality is called the *95% confidence interval for the mean.* There is a 95% chance that the interval we have constructed contains the true population mean.

$$\bar{x} - 1.96 \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1.96 \cdot \frac{\sigma}{\sqrt{n}}$$

In general, for a sample of size *n,* the *(1-alpha)% confidence interval for the mean* is given by:

$$\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

EXAMPLE: If a population has a standard deviation of 14 and if a sample of size 35 has a mean of 74, find the *95% confidence interval for the mean.*

$$\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

$$74 - 1.96 \cdot \frac{14}{\sqrt{35}} < \mu < 74 + 1.96 \cdot \frac{14}{\sqrt{35}}$$

$$69.362 < \mu < 78.638$$

We also say that this estimate has the following margin of error.

$$E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = 1.96 \cdot \frac{14}{\sqrt{35}} \approx 4.638$$

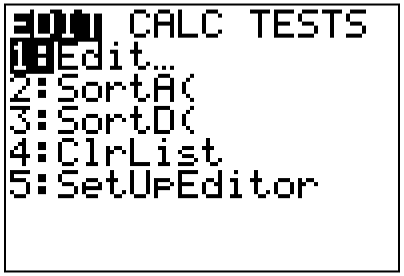
We can also do this on our calculator.

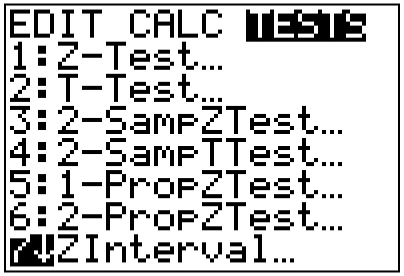
$\sigma = 14$

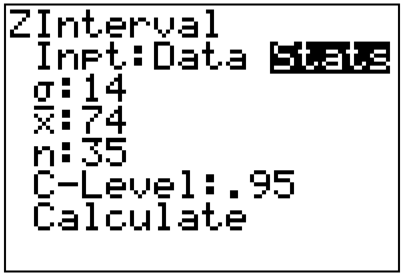
$\bar{x} = 74$

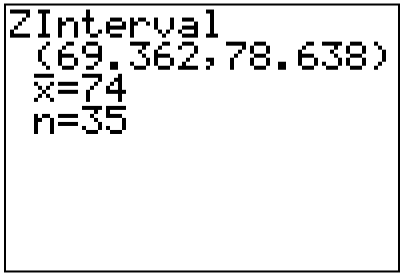
$n = 35$

$C - level = .95$



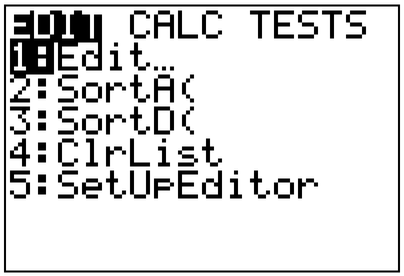



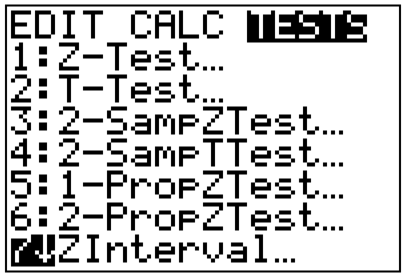



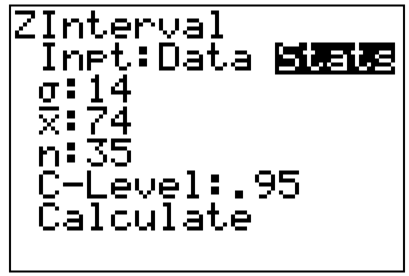



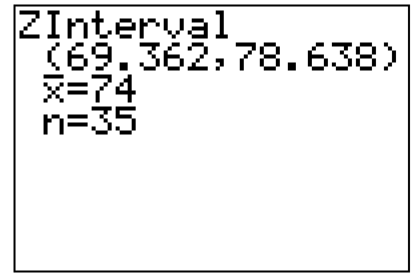

$$69.362 < \mu < 78.638$$

We can additionally determine our sample mean and maximum error from this result.



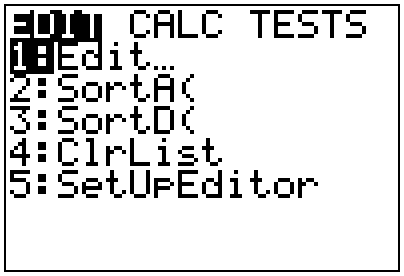



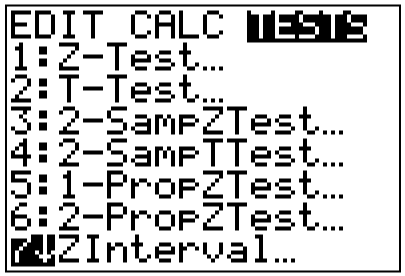



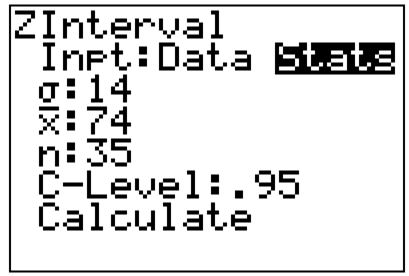



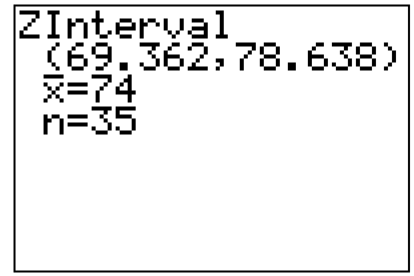

$$69.362 < \mu < 78.638$$

We can additionally determine our sample mean and maximum error from this result.



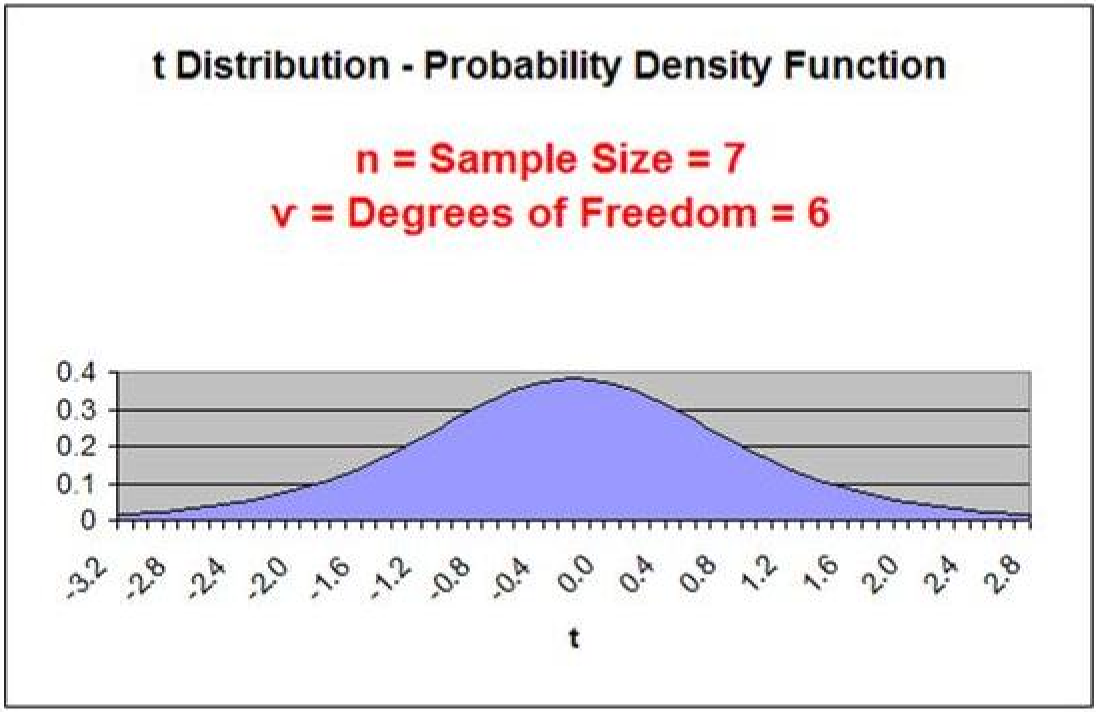




$$69.362 < \mu < 78.638$$

$$\bar{x} = \frac{78.638 + 69.362}{2} = 74$$

$$E = \frac{78.638 - 69.362}{2} = 4.638$$

Typically, we don't know either the true population mean or the true population standard deviation.  However, if our population is either normally distributed or our sample size *n* is greater than 30, then we may find a *(1-alpha)% confidence interval for the mean* using the *t-distribution.*



The *t-distribution* has the following properties.

1. Bell shaped
2. Symmetrical
3. Thicker at the tails than the normal distribution
4. There is a unique *t-distribution* for each sample size *n*
5. A *t-distribution* for sample size *n* has *n-1* degrees of freedom
6. The *t-distribution* approaches the normal distribution as *n* increases

Here's the formula for finding a (1-*alpha)%* confidence interval for the mean using the *t-distribution.*

$$\bar{x} - t_{\alpha/2} \cdot \frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

$$\text{degrees of freedom} \;=\; n - 1$$

$$\text{margin of error} = E = t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

EXAMPLE: If a population is normally distributed, *n=35, x-bar=1.97,* and *s=1.44*, find the 95% confidence interval for the *mean*.

$$\text{degrees of freedom } = 35 - 1 = 34$$

$$\bar{x} - t_{.05/2} \cdot \frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{.05/2} \cdot \frac{s}{\sqrt{n}}$$

$$1.97 - 2.03 \cdot \frac{1.44}{\sqrt{35}} < \mu < 1.97 + 2.03 \cdot \frac{1.44}{\sqrt{35}}$$

$$1.4753 < \mu < 2.4647$$

$$E = t_{.05/2} \cdot \frac{s}{\sqrt{n}} = 2.03224 \cdot \frac{1.44}{\sqrt{35}} = 0.4947$$

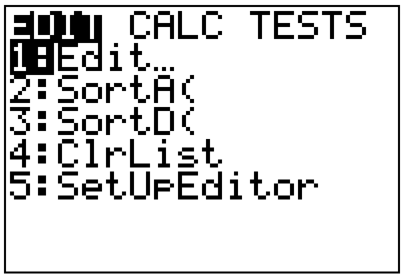
We can do this one, too, on our calculator.

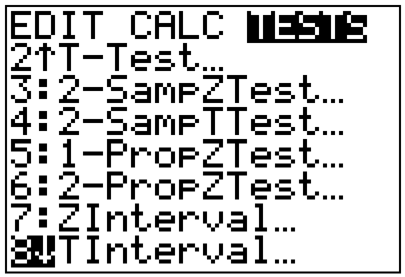
$\bar{x} = 1.97$

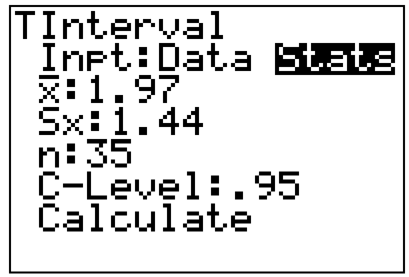
$s = 1.44$

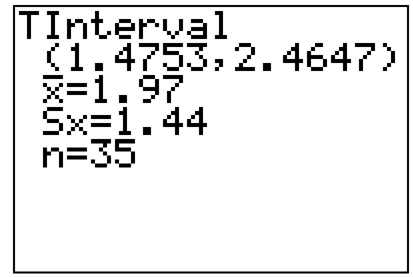
$n = 35$

$C - level = .95$



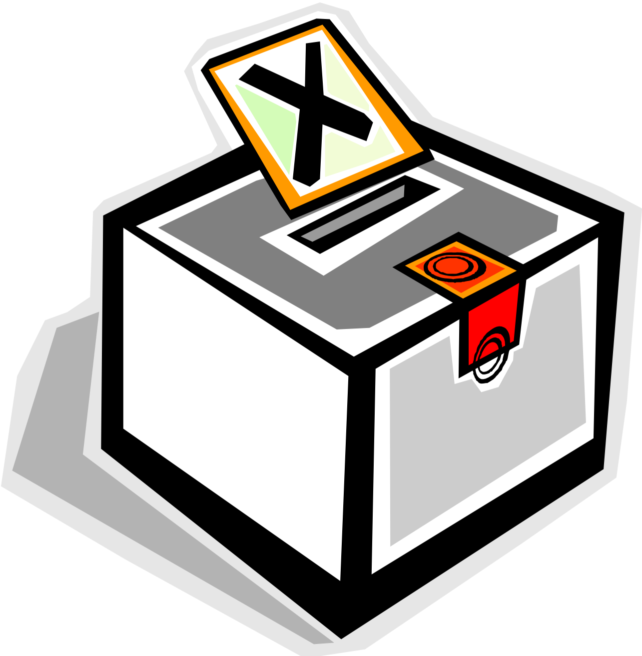




$$1.4753 < \mu < 2.4647$$

$$E = \frac{2.4647 - 1.4753}{2} = 0.4947$$

Now let's consider another situation. Suppose a sample of 100 people votes on a proposition called *proposition 1* to require everyone to take a course in statistics, and when the votes are tallied, 60% are in favor of the proposition and 40% are against.



$$n = 100$$

$$x = 60$$

$$\hat{p} = .6 = 60\%$$

We could, of course, consider this just one of many samples of size 100 that we could take from our population, and this results in a *binomial distribution*.

$$n = 100$$

$$x = \text{number of "yes" votes}$$

$$\hat{p} = x/n$$

If our sample size is sufficiently large, then we can approximate our binomial distribution by a corresponding normal distribution. In this case, our normal distribution will have, in general, the following mean and standard deviation.

$$\mu = np$$

$$\sigma = \sqrt{np(1-p)} = \sqrt{npq}$$

Hence, *z*-scores can be computed in this distribution by the following formula.

$$\mu = np$$

$$\sigma = \sqrt{np(1-p)} = \sqrt{npq}$$

$$z = \frac{x-\mu}{\sigma} = \frac{x-np}{\sqrt{npq}} = \frac{\frac{x-np}{n}}{\frac{\sqrt{npq}}{n}} = \frac{\frac{x}{n}-p}{\sqrt{\frac{npq}{n^2}}} = \frac{\hat{p}-p}{\sqrt{\frac{pq}{n}}}$$

Furthermore, if we don’t know the true value of *p* for our population, then we generally use *p-hat* and *q-hat* in our formulas as estimates.

$$\mu = np$$

$$\sigma = \sqrt{np(1-p)} = \sqrt{npq}$$

$$z = \frac{\hat{p} - p}{\sqrt{\frac{\hat{p}\hat{q}}{n}}}$$

We're now ready to find a $1\text{-}\alpha$ confidence interval. Thus, suppose our proportion, when converted to a $z$-score, lies between $-z_{\alpha/2}$ and $z_{\alpha/2}$. Then here's what happens.

$$-z_{\alpha/2} < z < z_{\alpha/2} \Rightarrow -z_{\alpha/2} < \frac{\hat{p} - p}{\sqrt{\frac{\hat{p}\hat{q}}{n}}} < z_{\alpha/2}$$

$$\Rightarrow -z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} < \hat{p} - p < z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

We're now ready to find a *1-α* confidence interval. Thus, suppose our proportion, when converted to a *z*-score, lies between $-z_{\alpha/2}$ and $z_{\alpha/2}$. Then here's what happens.

$$\Rightarrow -\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} < -p < -\hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

$$\Rightarrow \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} > p > \hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

$$\Rightarrow \hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

We're now good to go! Suppose in our example that we only had a sample of 100 voters with *p-hat* = *.6* and *q-hat* = *.4*. We can now set up a 95% confidence interval for the proportion of voters in favor of the proposition.

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

$$.6 - 1.96\sqrt{\frac{(.6)(.4)}{100}} < p < .6 + 1.96\sqrt{\frac{(.6)(.4)}{100}}$$

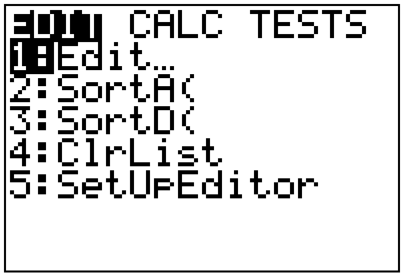
$$.50398 < p < .69602$$

$$50.398\% < p \cdot 100\% < 69.602\%$$

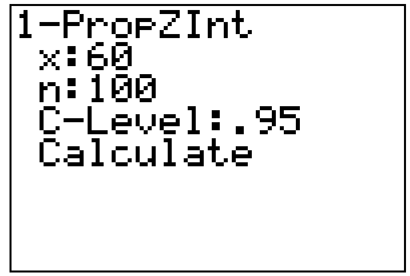
The margin of error in this case is:

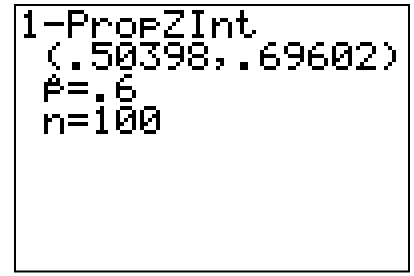
$$E = z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} = 1.96\sqrt{\frac{(.6)(.4)}{100}} = .09602 = 9.602\%$$

And this, too, can be done on the calculator.








$$E = \frac{.69602 - .50398}{2} = .09602 = 9.602\%$$

Notice that we can solve our margin of error formula for *n.* We can use this to determine the best sample size for a desired margin of error.

$$E = z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

$$\Rightarrow E^2 = \left[z_{\alpha/2}\right]^2 \cdot \frac{\hat{p}\hat{q}}{n}$$

$$\Rightarrow n = \left[z_{\alpha/2}\right]^2 \cdot \frac{\hat{p}\hat{q}}{E^2}$$

Also, we normally don't know *p-hat* or *q-hat* before taking a sample, so we'll just use 0.5 as the estimate for each. And now if we want a 95% confidence interval with a margin of error of 2%, then this is what our estimated sample size should be:

$$n = \left[ z_{\alpha/2} \right]^2 \cdot \frac{\hat{p}\hat{q}}{E^2} = 1.96^2 \cdot \frac{(.5)(.5)}{(.02)^2} = 2401$$

If our result had contained a fractional part, then we would always round up to the next whole number

$$n = \left[ z_{\alpha/2} \right]^2 \cdot \frac{\hat{p}\hat{q}}{E^2} = 1.96^2 \cdot \frac{(.5)(.5)}{(.02)^2} = 2401$$