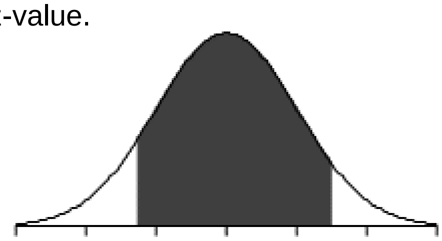
Estimation

- Remember that we use **statistics** (information about a sample) to estimate **parameters** (information about a population.
- There are 2 types of estimates:
    - **point estimate –** the statistic itself
        - parameter is almost never exactly equal to point estimate
    - **interval estimate** – a range of scores
        - given as a range of scores (from bottom to top) or as number ± a margin of error
        - the actual parameter is almost always somewhere in the interval
        - how big the interval is depends on how large our sample is and how confident we want to be of our prediction
        - bigger samples → smaller interval
        - more confidence → bigger interval

Confidence

- how certain you are that the actual mean will be in the interval you predict.
- usually expressed as a percentage (80% confidence, 95% confidence, etc.)
- essentially this is the chance your answer is correct
- literally it is the percentage of the normal curve (sampling distribution) that is in the middle, out to a given z-value.



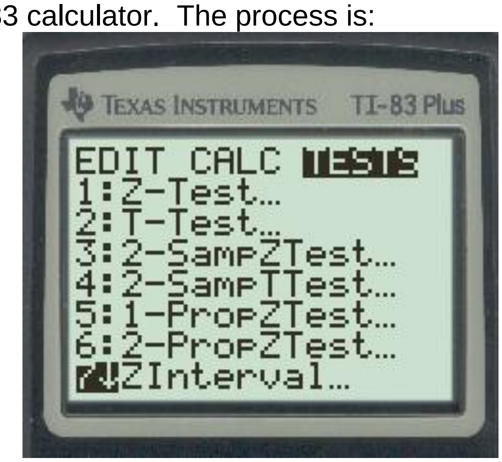
- as you increase your confidence, you have to go out farther (a higher z-value) to include more of the normal curve.
- variable for confidence is "c" … usually given as a decimal

TYPICAL PROBLEM:
You want to know the average number of students in a course at Iowa Lakes. You take a sample of 36 courses and find that the average is 14.7 students with a standard deviation of 4.9 students. Find a point estimate and an interval estimate, with 95% confidence.

- POINT ESTIMATE = 14.7 students (This is just the average from the sample given in the problem.)

You can do the interval estimate very easily with a TI-83 calculator. The process is:

- Go to STAT mode.
- Select TESTS.
- Select "**ZInterval**" from the "TESTS" menu.



- Select “**Stats**” for “Inpt”
- Enter the standard deviation (says “σ” instead of “s”), mean($\bar{x}$), number (n), and confidence level.
- Highlight “**Calculate**”, and hit ENTER.

The answer screen looks like this:

You may be asked to find the margin of error. To do this, just take

UPPER BOUND – MEAN

16.301 – 14.7 = 1.601

If you didn’t have a TI-83, you’d have to use this formula (which is built into your calculator):

- $$E = \frac{z_c \bullet s}{\sqrt{n}}$$
  - E = margin of error
  - n = # in sample
  - s = sample S.D.
  - z_c = z-score associated with a given level of confidence

The z-scores associated with standard confidence levels are given at the bottom of the page, near the “big” z-table in your book.

There are several types of interval estimates, each of which is used in different situations.

- We will look at three of these.

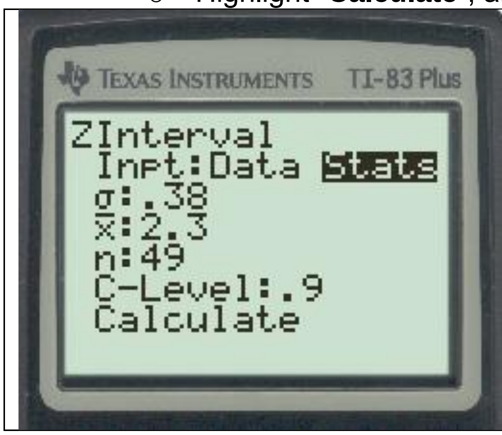
Z-INERVAL

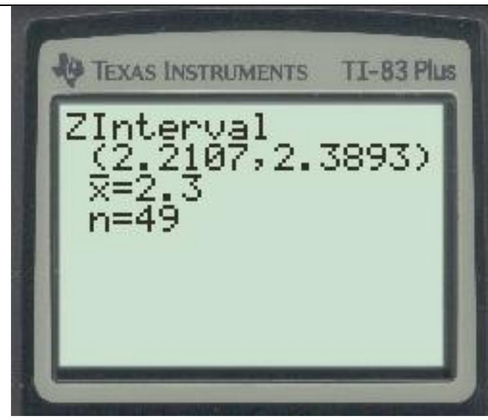
- Used to estimate the μ, the mean (average) of the population.
- Usually used when you have a large sample or when you know a lot about the population.

Example:
You want to find the average number of children for families in Kossuth County. You take a sample of 49 families and find that the average is 2.3 children. The standard deviation is .38 children. Find a point estimate and an interval estimate, with 90% confidence.

NOTE: From a statistical point of view, 49 is a large enough sample that we can use a Z-Interval.

- ○ STAT → TESTS → Zinterval.
- ○ Enter the standard deviation, mean, number, and confidence level.
- ○ Highlight "**Calculate**", and hit ENTER.





So there would be between 2.2107 and 2.3893 children.

Margin of Error = 2.3893 – 2.3 = .0893

Example:
You want to know the average number of unpopped kernels per bag in a brand of popcorn. You take a sample of 40 bags and find an average of 23 unpopped kernels, with a standard deviation of 5.8. Find an interval estimate, with 99% confidence.

z-intervals generally work well with **large** samples.

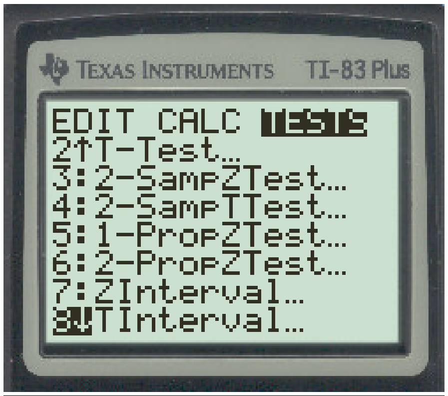
- "Large" in statistics generally means n $\geq$ 30
- This is because at 30, $s \approx \sigma$.
- If you have a smaller sample, you must know the standard deviation of the population (like for IQ) in order to use a z-interval.

We can also se an alternative statistic, called "t", that is better adapted to small samples.

t-intervals

- On a TI-83, just go to "TInterval" in the TESTS menu.
    - The order is different, but otherwise things work exactly like a ZInterval.

EXAMPLE:
A sample of 8 people shows that the average height is 67 inches, with a standard deviation of 1.5 inches. Find a point estimate and an interval estimate for the average height, with 99% confidence.



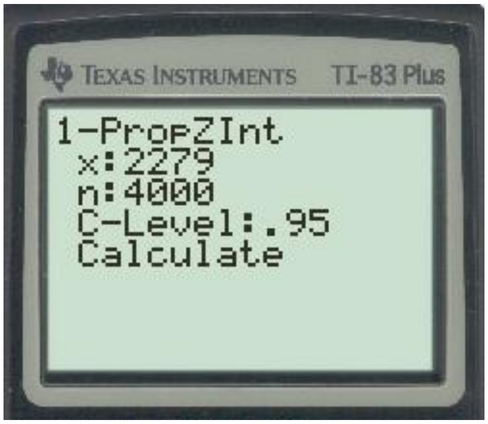
EXAMPLE:
A sample of 6 twenty-ounce bottles of pop showed that 259 was the average number of calories. The standard deviation was 34.7 calories. Find a point estimate and an interval estimate for the average calories of all twenty- ounce bottles of pop, with 95% confidence.

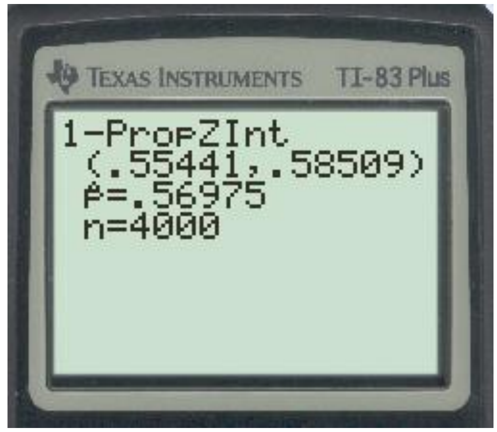
EXAMPLE:
Joe wanted to know the average number of runs his favorite baseball team scored per game. He watched 12 games and found they scored an average of 6.3 runs with a standard deviation of 2.4 runs. Use this to find an interval estimate with 85% confidence.

Binomial Estimates
(on a TI-83 this is 1-Prop Z-Int)

- This means estimating a **PERCENTAGE** (proportion) with a given characteristic, rather than an average.
- It's the sort of estimate used in election forecasts.

EXAMPLE:
NBS takes a poll of 4000 people to see whether they like what the President is doing in office. They find that 2279 people support the President. Give a point and an interval estimate of what the President's support, with 95% confidence. Also find the margin of error.





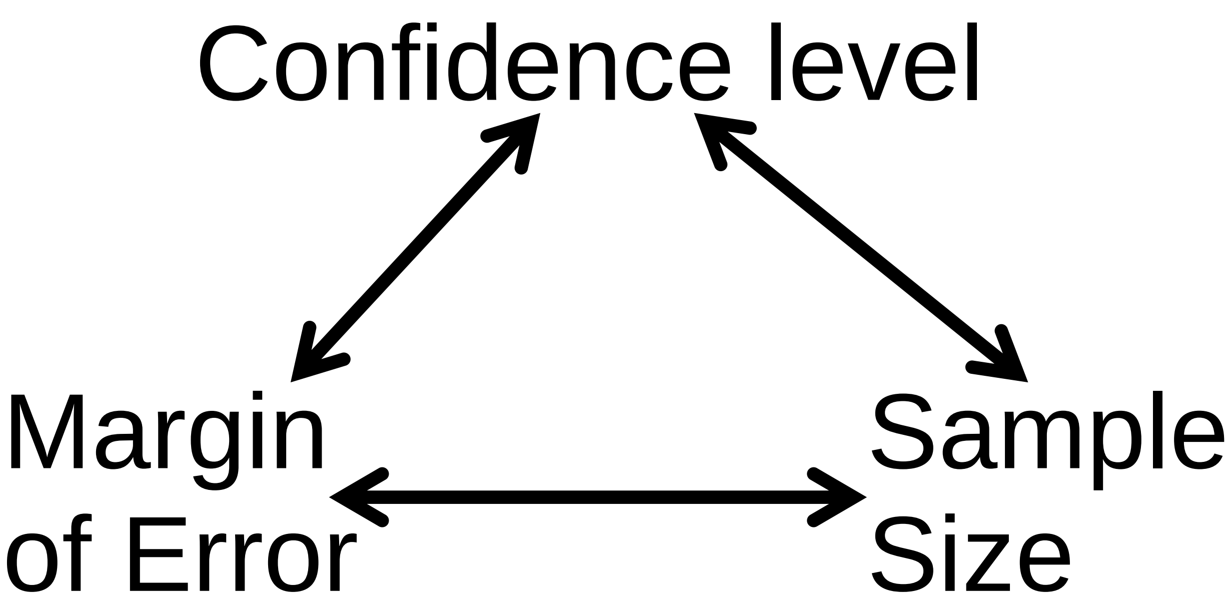
- Point Estimate
  - The actual percentage of the sample with that characteristic
  - This is given on your calculator with the variable p-hat.
  - You could easily find this yourself with the formula $\hat{p} = \frac{x}{n}$
- Interval Estimate
  - Just like before, you just copy the lower and upper bound from your calculator.

- Margin of Error
  - Just like before, subtract Upper Bound – Middle
  - .58509 – .56975 = .01534
  - On TV or in a newspaper they'd say "57% $\pm$ 1½ %".

If you had to do an interval estimate by hand, you'd use the formula $E = (z_c)\sqrt{\frac{\hat{p} \bullet \hat{q}}{n}}$

- Here $\hat{q}$ is the percent that DON'T have the characteristic

<u>SAMPLE SIZE PROBLEMS</u>
There are three things that play against each other in estimation problems:



Changing any one of these three things will automatically affect the others.

- We generally want the confidence level to be as high as possible and the margin of error to be as low as possible.
- These are often set by outside forces.

The only factor you can control is the sample size.

- If you make your sample large enough, you can guarantee the confidence and margin of error your want.

SAMPLE SIZE WITH AVERAGES
FORMULA:

$$n = \left(\frac{z_c \bullet s}{E} \right)^2$$

- s = standard deviation
- E = margin of error (giveaway in problem is "within")
- Z_c = z-score for given confidence (found in inset table)
- This is just the z-error formula, solved for "n".
- You **<u>can't</u>** do sample size problems in the TI-83 stats mode. (You **<u>must</u>** use the formula.)

<u>EXAMPLE</u>:
A battery manufacturer wants to know the average time their AAA batteries will last when used in calculators. They test 16 batteries in conditions similar to calculator use and find they last an average of 13.7 hours, with a standard deviation of 1.9 hours.
How many batteries will they have to test to estimate the actual average within 0.5 hours, with 99% confidence?

- z_c = 2.58 (for 99%)
- $$n = \left(\frac{2.58 \bullet 1.9}{.5} \right)^2$$
  $$n = 96.118416$$
- We **ALWAYS ROUND UP** on sample size problems, so they should include 97 in the sample.

BINOMIAL (PERCENT) SAMPLE SIZE PROBLEMS

- If you have an estimate of "p" ahead of time:

$$n = \hat{p} \bullet \hat{q} \left(\frac{z_c}{E} \right)^2$$

  - p-hat is the percent that do have the characteristic.(as a decimal)
  - q-hat is the percent that don't have the characteristic (as a decimal)
  - E is the margin of error (as a decimal)
  - Z_c isz-score for given confidence (found in inset table)

- If you don't have an estimate of "p" ahead of time, you assume it's 50-50 (the worst possible case):
- So p-hat and q-hat are both .5

$$n = .5 \bullet .5 \left(\frac{z_c}{E} \right)^2$$

EXAMPLE:
In the first games of the season, the players on a basketball team made 8 of 11 free throws they attempted. A sports analyst wants to predict the team's free throw percentage for the whole season. How many attempts will have to be analyzed to estimate the season free throw percentage within 10% either way, with 90% confidence.

First find p-hat and q-hat

8 out of 11 is .7272…

So p-hat is appx. .73
and q-hat is appx. .27

(Answers may vary slightly due to rounding.)

- Since it's 90% confidence, $z_c = 1.645$
- $$n = .7272 \bullet .2728 \left(\frac{1.645}{.10} \right)^2$$
  $$n = 53.68216725$$
- about 54 free throws

EXAMPLE:
Mr. X and Mrs. Y are running for governor. A preliminary poll shows that 57% of the voters favor Mr. X. How large of a sample should the Associated Press take to estimate the winner within 2% either way, and with 95% confidence?

EXAMPLE:
The Des Moines *Register* is trying to estimate the percentage of Iowans who support Governor Branstad. They don't have a previous estimate of what that rating should be.

How many people should they sample to be within 4% of the actual approval rating, with 95% confidence?

- Since it's 95% confidence,
  $z_c = 1.96$

$$n = .5 \bullet .5 \left(\frac{1.96}{.04} \right)^2$$

$$n = 600.25$$

They should sample 601 people.

EXAMPLE:
You have no clue what percent of Americans shop at Wal-Mart. How large of a sample should you select to estimate this percentage within 3% either way, with 99% confidence?

For 99%, $z_c = 2.58$

```
.5*.5(2.58/.03)^2
```

1849 (Since it comes out even you don't have to round up.)