Confidence Intervals

The reason behind taking a SRS from some population is to get an accurate estimate about the center (average) of that population. It is more accurate to express the estimate (or guess) for that average as an interval rather than a single #.

Think about picking a number between 1 & 10. You could guess a single number, but you would only have a 10% chance of being right. If you guessed that the number would be between 2 and 8 (inclusively), you would have a 70% chance of being right.

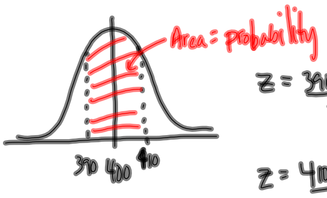
In statistics, we take a sample and find the mean to make a guess about the mean for the population. It is better to try to "catch" the population mean in an interval rather than guessing just a single value.

<u>Example</u>:

The recent scores of Gr9 students in a literacy test had the following results: $N(400, 36)$.

If we choose <u>one</u> student at random, what is the probability that their mark is between 390 + 410?

Since the population is normal, we can use z-scores!



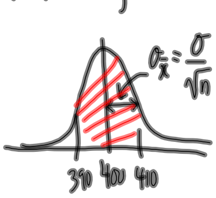
$$z = \frac{390-400}{36} = -0.28 \quad (0.3897)$$

$$z = \frac{410-400}{36} = +0.28 \quad (0.6103)$$

$$61.03\% - 38.97\% = 22.06\%$$

There is a 22.06% probability that the student's score is between 390 and 410.

What if we take a SRS of 25 from these scores, what will be the probability of the mean of this sample ($\bar{x}$) being between 390 and 410?



$$\sigma_{\bar{x}} = \frac{36}{\sqrt{25}} = 7.2$$

$$z = \frac{390-400}{7.2} = -1.39 \quad (0.0823)$$

$$z = \frac{410-400}{7.2} = +1.39 \quad (0.9177)$$

$$91.77\% - 8.23\% = 83.54\%$$

There is an 83.54% chance of the mean of the sample being between 390 and 410.

There is a much greater chance that a sample mean will be close to the population mean rather than a single piece of data. We can get a better idea about the population from a sample than we can get from an individual data piece.

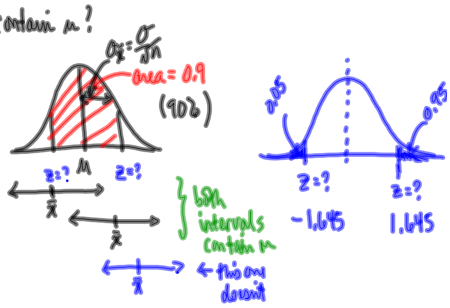
What we want to do is to take a SRS from a population and using the mean ($\bar{x}$), create an interval that we feel will contain the population mean (σ).

Let's say we take a SRS ($n=25$) and get $\bar{x} = 407$.
How can we create an interval that will contain the population mean ($\mu = 400$)?

If my $\bar{x}$ was different, I may have to +/- more than 1 st. dev. in order to contain μ.
(recall $\sigma_{\bar{x}} = 7.2$)

The problem is, we don't know the value of μ when we take our SRS. From the CLT (Central Limit Theorem) we know that any sampling distribution is normal if $n \geq 30$.

How many std. dev do I have to +/- if I want to be 90% certain that the interval will contain μ?



If your z score is ±1.645 then you will have a 90% probability that the interval will contain μ.
the number of st. dev units.

Recall: $\sigma_{\bar{x}} = 7.2$

interval: $\bar{x} \pm z\sigma_{\bar{x}}$

$407 \pm 1.645(7.2)$

407 ± 11.844

$395.156 - 418.844$

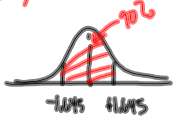
There is a 90% chance of the pop mean falling between 395.156 and 418.844.

(90% confidence interval)

90% probability: $z = \pm 1.645$

95% probability: $z = \pm 1.96$

99% probability: $z = \pm 2.576$



Confidence Interval: $\bar{x} \pm z\sigma_{\bar{x}}$

Practice - Confidence Intervals

1. $n = 300$ large sample

$\bar{x} = 509$

Sample st. dev $S_x = 130$ use S_x as a reasonable estimate for σ

90% confidence interval?

$(z = 1.645)$

$\bar{x} \pm z\sigma_{\bar{x}}$

$509 \pm 1.645(7.51)$

509 ± 12.35395

496.65 to 521.35

90% confidence interval

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

$\sigma_{\bar{x}} = \frac{130}{\sqrt{300}}$

$\sigma_{\bar{x}} = 7.51$

There is a 90% chance that the interval will contain the population mean.