Chapter .1 Day 1

Confidence Intervals for the Mean (σ known)

Recall: Due to time, money, etc. we usually do not have access to *all* measurements of an *entire* population so instead we rely on a ___Sample___.

Example: An economics researcher is collecting data about the number of hours worked by grocery store employees in the Indianapolis area:

He collects data from a random sample of 30 employees:
26, 25, 32, 31, 28, 28, 28, 22, 28, 25, 21, 40, 32, 22, 25, 22, 26, 24, 46, 20, 35, 22, 32, 48, 32, 36, 38, 32, 22, 19

The mean of this data is: $\bar{x} = 28.9$

Do you think that the population mean μ is also 28.9 ?

New Terminology:

A ___point estimate___ of a population parameter is an estimate of the parameter using a single number.

We will use ___$\bar{X}$___ to represent the point estimate for the population mean μ. The sample mean $\bar{x}$ is the most unbiased point estimate.

Goal:
1. Find an **estimate for the population** mean from a sample.
2. Recognize the estimate is not perfect and will be "off" by some **margin of error.**
   Calculate the amount of error.
4. **Create an interval estimate** for the population mean.

Example: Sample mean $\bar{x} = 28.9$ *hours*
Margin of error is 2.5 hours

28.9 – 2.5 = 26.4

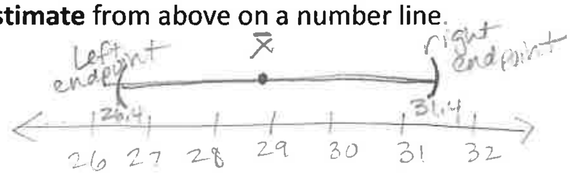
28.9 + 2.5 = 31.4

We estimate that the true mean is in the interval

___26.4___ to ___31.4___.

___26.4___ < μ < ___31.4___

Sketch the **interval estimate** from above on a number line.



If you were given the sketch first and asked to get the margin of error and sample mean, what would we need to do?

the Sample mean, $\bar{X}$ is the middle of the interval.

So $\bar{X} = \frac{26.4 + 31.4}{2} = \frac{57.8}{2} = 28.9$

margin of error is
$\bar{X}$ – left endpoint
or
right endpoint – $\bar{X}$

28.9 – 26.4 = 2.5

Assumptions about the random variable x:
The following must be true for us to proceed with the mathematics we want to do here...

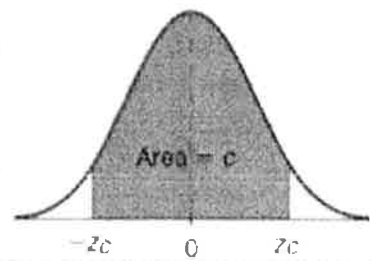
1. We have a **simple random sample of size "n"** drawn from a population of x values.
2. The value of the standard deviation of x, **σ is known.**
3. One of two things must be true:
   a. The x **distribution is normal** and our methods work for any sample size
   b. The x distribution is unknown BUT we have a **sample size of n = 30 or more.** Significantly skewed or not mound shape may require higher sample sizes.

An estimate is not very valuable unless we know how "good" it is.

Reliability of an estimate: **CONFIDENCE LEVEL** ("*c*")
You can choose *c* to be any value between 0 and 1.
(usually use **0.90, 0.95, or 0.99**)

The values of $-z_c$ and z_c are called the **CRITICAL VALUES**.

Confidence Level c and Corresponding Critical Value z_c Shown on the Standard Normal Curve



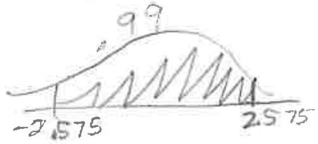
SOME LEVELS OF CONFIDENCE AND THEIR CORRESPONDING CRITICAL VALUES:

Level of Confidence c	Critical Value z_c
0.70, or 70%	1.04
0.75, or 75%	1.15
0.80, or 80%	1.28
0.85, or 85%	1.44
0.90, or 90%	1.645
0.95, or 95%	1.96
0.98, or 98%	2.33
0.99, or 99%	2.575

Example: What is the critical value z_c necessary to construct a confidence interval at the given level of confidence c?

a) c = 0.99

.99



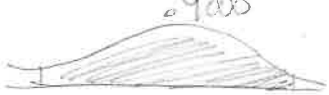
$\frac{1-.99}{2} = \frac{.01}{2} = .0050$

Look up on normal curve

$z_c = 2.575$

b) c = 0.90

.9000



$\frac{1-.9}{2} = \frac{.1}{2} = .05$

$z_c = -1.645$

Know how to find them

The difference between the point estimate and the actual parameter value is called the Sampling error.
When μ is estimated, the **sampling error** is the difference $\bar{x}-\mu$. In most cases, of course μ is unknown, and $\bar{x}$ varies from sample to sample. If $\bar{x}=8.2$ and $\mu=9.1$, the **sampling error** is $8.2-9.1 = -.9$.

However, you can calculate a maximum value for the error when you know the level of confidence and the sampling distribution.

MARGIN OF ERROR

...ample: Calculate the margin of error for a confidence interval with ...e following values:

$Z_c = 1.44$

a) c = 0.85, σ = 8.5, n = 40

$\frac{1-.85}{2} = \frac{.15}{2} = .075$

$E = 1.44 \cdot \frac{8.5}{\sqrt{40}} = 1.935$

What does this MEAN?

We are __85%__ confident the true mean is within __1.935__ of our estimate.

It DOES NOT mean that __85%__ of the time the mean is within __1.935__ of our estimate.

Maximal Margin of Error

- Since μ is unknown, the margin of error $|\bar{x} - \mu|$ is unknown
- Using confidence level c, we can say that $\bar{x}$ differs from μ by at most

$$E = z_c \frac{\sigma}{\sqrt{n}}$$

We can use the point estimate: $\bar{x}$ and the margin of error: **E** to construct a "Confidence Interval" for μ.

CONFIDENCE INTERVAL FOR μ:
An interval such that c is the PERCENTAGE of all intervals generated by the same process that contain μ.

$$\bar{x} - E < \mu < \bar{x} + E$$

95%

...ample: c = 0.95, $\bar{x} = 5.7$, σ = 1.3, n = 45
...onstruct a confidence interval.

$\frac{1-c}{2} = \frac{1-.95}{2} = \frac{.05}{2} = .025$

$Z_c = 1.96$

$E = 1.96\left(\frac{1.3}{\sqrt{45}}\right) = .4$

Rounding follow #5 in the problem

$5.7 - .4 = 5.3$
$5.7 + .4 = 6.1$

$5.3 < \mu < 6.1$
$(5.3, 6.1)$

I am 95% confident that the interval from 5.3 to 6.1 contains the true mean.

Step 1: Check your assumptions
- ☐ simple random sample
- ☐ σ is known
- ☐ normal distribution OR $n \geq 30$

Step 2: Identify σ, n, z

Step 3: Calculate the margin of error, E using formula

Step 4: Identify $\bar{x} \pm E$ and set up your confidence interval

Step 5: Interpret your interval in the context of the problem

"I am C% confident that the interval _____ to _____ contains the true mean."

Example: Suppose that the standard deviation of all high school seniors' SAT scores in a certain year was σ = 150. A random sample of 100 scores yielded a sample mean $\bar{x} = 1010$. Let μ be the mean of all SAT scores in that year. Find a 0.99 confidence interval for μ. Round your answers to integers.

$\sigma = 150 \quad n = 100 \quad \bar{x} = 1010$

$\frac{1-.99}{2} = \frac{.01}{2} = .005$

$Z_c = 2.575$

$E = 2.575\left(\frac{150}{\sqrt{100}}\right) \approx 39.$

$1010 - 39 < \mu < 1010 + 39$

$971 < \mu \; 1049$

I am 99% confident that the true mean is in the interval 971 to 1049.

Example: Use the indicate interval to find the margin of error and the sample mean.

Confidence interval: (20.5, 40.5)

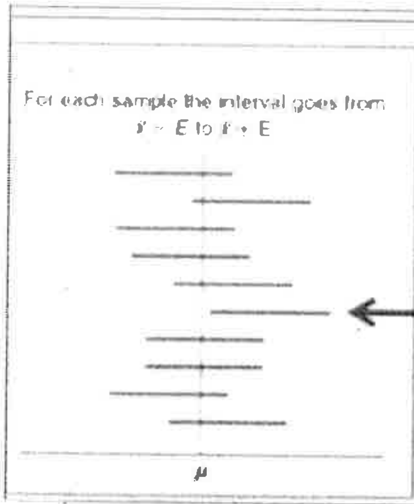
$$\bar{X} = \frac{20.5+40.5}{2} = \frac{61}{2} = 30.5$$

margin of error $= 30.5 - 20.5 = 10$

INTERPRETATION OF CONFIDENCE INTERVALS

Once we have a specific confidence interval for μ, such as 3 < μ < 5, all we can say is that:

We are c% confident that the mean of the population μ will be in that interval.



At the 0.90 confidence level, 1 in 10 samples, on average, will fail to enclose the true mean μ within the confidence interval.

MINIMUM SAMPLE SIZE NEEDED TO ESTIMATE A MEAN

more confident but less accurate

As the confidence level increases, the interval increases. WIDER BAR -> MORE CONFIDENT

As the interval increases, the precision of the estimate decreases. WIDER BAR -> LESS ACCURATE

A more accurate answer can be found by <u>increasing</u> the sample size, n. But what's the minimum sample size needed?

> Given a c-confidence level and a margin of error E, the minimum sample size n needed to estimate the population mean is
>
> $$n = \left(\frac{z_c \sigma}{E}\right)^2$$
>
> If σ is unknown, it can be estimated with s provided you have a preliminary estimate with at least 30 members.

Example: Find the minimum sample size needed to be:

need z_c

a) 95% confident a sample is within 2 units of the population mean if σ is known to be 3.6.

E

$$\frac{1-.95}{2} = \frac{.05}{2} = .025$$

$$z_c = 1.96$$

Always round up

$$n = \left(\frac{1.96 \cdot 3.6}{2}\right)^2 = 12.44\ldots$$

$$\boxed{n = 13}$$

b) 90% confident a sample is within 0.5 units of the population mean if σ is known to be 4.7.

E

$$\frac{1-.9}{2} = \frac{.1}{2} = .05$$

$$z_c = 1.645$$

$$n = \left(\frac{1.645 \cdot 4.7}{.5}\right)^2 = 239.1\ldots$$

$$\approx \boxed{240}$$