Chapter 14 Confidence Intervals: The Basic

**Statistical Inference**

**Situation:** We are interested in estimating some parameter (population mean, \( \mu \)) that is unknown. We take a random sample from this population.

**Goal:** Draw inference about the population parameter from the sample data.

**Confidence Intervals**

Example: Beetle cars

Suppose EPA wants to estimate \( \mu \), the average CO\(_2\) emitted by all Beetle cars. 49 Beetle cars are randomly selected at the VW plant in Detroit and tested for CO\(_2\) emissions. The test results show these 49 cars have an average of 1.5 grams CO\(_2\) emission per mile.

What is our best guess about the unknown parameter \( \mu \)?

But we need to be careful in making a conclusion about \( \mu \) because

- We know any other random sample of 49 cars would give a different value of \( \bar{x} \).
- We can’t say how much confident we are with our estimate.

So instead of just giving one number (the value of \( \bar{x} \)) as our estimate of \( \mu \), it seems more desirable to give an __________ of values that may contain \( \mu \) with some degree of __________.

Let’s try to find an interval of values within which we can say that average CO\(_2\) emission (\( \mu \)) of Beetle cars lies with some high degree of confidence, say 95%.

For this, let’s recall from Chapter 11 how \( \bar{x} \) behaves in repeated sampling.
In the “Beetle cars” example, suppose we repeatedly sample 49 cars and for each sample note their mean CO₂ emission. We know that the distribution of sample mean, \( \bar{x} \) is then

Assume we know that the standard deviation (\( \sigma \)) of CO₂ emissions of Beetle cars is 0.84 (it is unrealistic to assume \( \sigma \) to be known, we would get rid of this assumption in Chapter 18!).

Then,

Using the 68-95-99.7 rule, we know that \( \bar{x} \) will fall within 2 standard deviations

\[
2 \left( \frac{\sigma}{\sqrt{n}} \right) = \mu
\]

of the population mean \( \mu \) with probability

This means that in 95% of all samples, the observed mean score \( \bar{x} \) will be within _____ points of the population mean \( \mu \). (Note: \( \mu \) is unknown and fixed - it doesn’t vary from sample to sample).

To say that \( \bar{x} \) is within 0.24 points of \( \mu \) is equivalent to saying that \( \mu \) is within 0.24 points of observed \( \bar{x} \). This happens in 95% of all samples.
Combining these facts we can say, “In 95% of all samples (of size \( n = 49 \)) the true but unknown mean \( \mu \) lies in the interval”. We can rewrite this interval as

Recall that mean of our sample was 1.5 grams. So we say that “We are 95% confident that

This interval we just calculated is a 95% confidence interval for the unknown average CO\(_2\) emission (\( \mu \)) of all Beetle cars.

In general, confidence intervals for any parameter consists of two parts:

1) An interval calculated from the data of the form

The margin of error conveys how accurate we believe our guess of the true parameter value is, based on the variability of the estimate.

2) A confidence level, \( C \), gives the probability that the random interval captures the true parameter value in repeated samples. (Note that it is NOT the probability that any one specific interval calculated from a random sample captures the true parameter.)

Many types of confidence intervals exist for various kinds of parameters.
Ch. 14: Confidence Intervals for mean \( \mu \) (s.d. \( \sigma \) is known).
Ch. 18: Confidence Intervals for mean \( \mu \) (s.d. \( \sigma \) is unknown).
Ch. 19: Confidence Intervals for difference between two population means.
Ch. 20: Confidence Intervals for population proportion.
Ch. 21: Confidence Intervals for comparing two proportions.

We will concentrate on confidence intervals for the mean \( \mu \) of a population.
Confidence Intervals for a Population Mean (standard deviation is known)

Confidence Intervals for a Population Mean

Choose an SRS of size \( n \) from a population having unknown mean \( \mu \) and known standard deviation \( \sigma \).

A level \( C \) confidence interval for \( \mu \) is

Here, \( z^* \) is the value on the standard normal curve with area \( C \) between \(-z^*\) and \( z^* \). The interval is exact when the original population distribution is normal and is approximately correct for large \( n \) otherwise.

The most commonly used confidence levels are

<table>
<thead>
<tr>
<th>( C )</th>
<th>90%</th>
<th>95%</th>
<th>99%</th>
</tr>
</thead>
<tbody>
<tr>
<td>( z^* )</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\( z^* \) for other confidence levels can be found similarly from Table A or more conveniently from Table C.
Ex: Beetle Cars
Recall that the CO$_2$ emissions of Beetle cars have a mean $\mu$ and standard deviation $\sigma = 0.84$. Our sample mean was $\bar{x} = 1.5$ and $n = 49$. We had calculated the 95% confidence interval for the mean CO$_2$ emission as

a. Find a 90% confidence interval for the mean CO$_2$ emission.

b. Find a 99% confidence interval for the mean CO$_2$ emission.

c. Find an 80% confidence interval for the mean CO$_2$ emission.

Note: We don’t know if any of the above confidence intervals contain $\mu$ or not! Then what do we mean by confidence?
The meaning of “Confidence”: When we say “95% confident”, we mean that

Our confidence is in the procedure, not in any one specific interval.

So it is **completely wrong to say**:

(because any one interval either contains the parameter or not, there is no randomness in it!)

Remember that probability (chance) is associated only with a random phenomenon. After you have constructed a CI from a random sample, there is no randomness left it. Hence it doesn’t make sense to attach any probability statement to a specific (numerical) CI.
Behavior of Confidence Intervals

What happens to the margin of error if we increase the confidence level $C$? Does it increase, decrease or stay the same? (Hint: what happens to the value of $z^*$?)

How does this affect the width of the resulting confidence interval?

**Note the tradeoff:** We would like to have a smaller margin of error (narrower interval) as well as high confidence but...

Q: What could we do to get a narrower interval (smaller margin of error) without lowering confidence?

E.g. in the “Beetle cars” example the 99% confidence interval for $\mu$ was found to be ________________ with a sample of size 49. If instead we had taken a sample of size 100 cars and suppose their mean CO$_2$ emission was 1.5 grams, then

How does the size of $\sigma$ affect the margin of error?

Thus we have 3 ways of reducing the width of the confidence interval:

1) 

2) 

3)
Choosing the Sample Size

We saw that we can have high degree of confidence as well as small margin of error by

Usually researchers will have a desired confidence level and margin of error they want to attain. So one aspect of designing any study is to decide the number of observations needed.

Let \( m \) represent the desired margin of error. Recall the formula of margin of error:

\[
\text{Margin of Error} = \frac{m \cdot \sigma}{\sqrt{n}}
\]

Solving for \( n \) we get:

\[
\frac{m \cdot \sigma}{\text{Margin of Error}} \leq n
\]

*****Always round your answer up!!*****

Ex: Suppose PGSA (Poor Graduate Students Association) at the Texas state wants to estimate the mean monthly income of SMU graduate students within $100 with 95% confidence. How many students should PGSA sample? Assume that the standard deviation of incomes of SMU graduate students is $421.