Lecture 9

Last time

Difference between population distribution, data distribution and sampling distribution.

Sampling distribution of $\bar{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$ when n $\geq$ 30 by the Central Limit Theorem.

Confidence

Statistical Inference

We have finally reached the point towards which we have been working all semester. We will introduce some specific methods for drawing conclusions about a population based on a sample taken from that population. The term **statistical inference** refers to these methods.

Here are some characteristics of statistical inference:
- it uses probability to say how trustworthy the conclusions are (since a different sample will produce different numbers)
- it is based on sampling distributions, so it only provides probabilities based on the longrun behavior of the statistic
- it can only be used when the data come from a random sample or randomized experiment

Example:

In a national park in 1993, 196 female grizzly bears were tracked and the number of young produced per bear was recorded. The U.S. Fish and Wildlife Service uses this data to monitor the population growth. The mean for the sample is $\bar{X} = 1.239$. Given this information, what can we say about the number of cubs produced per bear for all 1,100 bears in the park? This is the question we will attempt to answer in this section.

How does $\bar{X}$ relate to μ?

Estimators

There are two types of estimators we are using to make inferences about the population parameter: **point estimators** and **interval estimators**.

Point estimates are the most common form of inference reported by the mass media. For example, every month the Gallup organization reports survey results on the public approval of President Bush's performance in office. In April 2005, 45% of Americans approved while 32% of Canadians, 30% of British, 19% of French, 19% in Spain and 17% in Germany approved. These were point estimates rather than parameters, since they used a sample of about 1000 people in each country rather than the entire population.

An interval estimate indicates precision by giving an interval of numbers around the point estimate. The interval is made up of numbers that are the most possible values for the unknown parameters, based on the data observed. For example, a survey of recent college graduates predicts that the mean salary of all graduates working fulltime falls somewhere between $35,000 and $39,000, that is, within a **margin of error** of $2000 of the point estimator $37,000. An interval estimate is designed to contain the parameter with some chosen probability, usually 0.95. Because interval estimates contain the parameter with a certain degree of confidence (probability), they are referred to as **confidence intervals**.

Confidence Intervals for Mean μ

Example:
Refer back to the bear example. Imagine that we know that σ=.57 for the entire population. In reality we wouldn't know this. Then,

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

We base all of our inference on sampling distributions.

Remember the Empirical Rule (68-95-99.7): It says that in 95% of all samples, the sample mean will be within (approximately) 2 standard deviations of the population mean μ.

This interval of numbers is called the confidence interval and has the general form

$$POINT\ ESTIMATE \pm MARGIN\ OF\ ERROR \tag{1}$$

where the margin of error (m) tells us how accurate the (point) estimate is likely to be in estimating a parameter. It is a multiple of the standard deviation of the sampling distribution of the estimate, such as 1.96 $\times$ (standard deviation) when the sampling distribution is normal.

The confidence level (C) of the confidence interval determines the probability that the interval contains the parameter in repeated sampling.

Example:
In the long run (many, many samples) a 95% confidence interval for $\bar{X}$ means that only 5% of the confidence intervals produced from the samples will not contain the population mean μ.

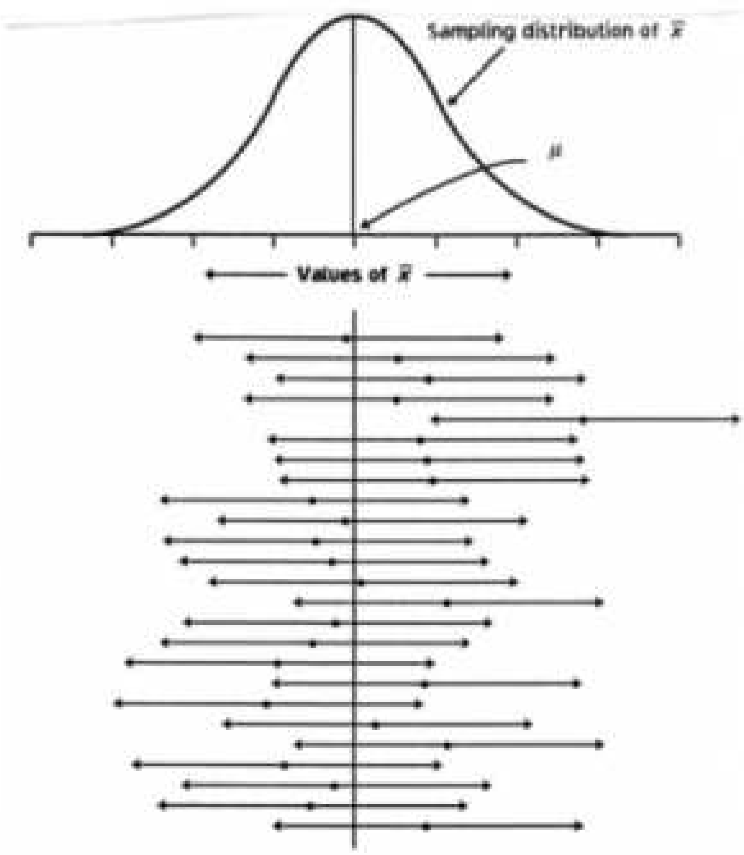


Figure 6.4 *Twenty-five samples from the same population gave these 95% confidence intervals. In the long run, 95% of all samples give an interval that contains the population mean μ.*

When we have a large enough sample size, $\bar{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$. We will use a **one-sample Z statistic (Z-score)** to calculate our confidence interval (specifically our margin of error):

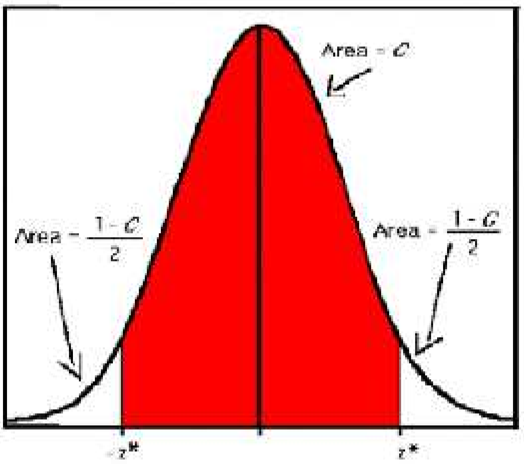
$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \tag{2}$$

This is the standardized form of $\bar{X}$ where $Z \sim N(0, 1)$

Z tells us how many standard deviations away $\bar{X}$ is from μ.

When Z is used to calculate a margin of error it is referred to as the critical value and sometimes written as Z^*

The picture below demonstrates how the critical value Z^* and C are related to each other with the normal distribution. Z^* is the number such that the probability C is captured between $-Z^*$ and Z^*.



Example:
Use the standard normal table to find Z^* for C=.85. Draw a diagram labeling Z^*, the area in the tails and the confidence level.

If we start with a sample mean $\bar{X}$ and go out Z^* standard deviations, we have an interval that contains μ in C of all the samples.

A level C confidence interval for μ:

$$\bar{X} \pm Z^*(\frac{\sigma}{\sqrt{n}}) \tag{3}$$

Example:
At a stockyard in Missouri 1000 two-year old steers are delivered to be slaughtered. The manager takes an SRS of n=50 steers to weigh to estimate the mean weight of all 1000 steers. The weight of this two-year old variety of beef cattle is known to have standard deviation of σ=35.2 lbs. Find a 90% CI for μ if $\bar{X} = 1653$ lbs. The first thing to do when working these problems is to list all the important information given in the problem.

$\bar{X}$= σ= n=

Z^*=

$\bar{X} \pm Z^*(\frac{\sigma}{\sqrt{n}})$=

The 90% CI for μ is

The width of a CI is the upper bound minus the lower bound or twice the margin of error. The smaller the width of a CI, the more accurate it is.
What is the width of the 90% CI in this example?

How Confidence Intervals Behave

Confidence Level

- The user chooses the confidence level, preferably before the study or experiment is completed

- High confidence means that the parameter is almost always in the interval

- Fields like physics and chemistry use Cs of .99 and .98 because they deal with very precise laboratory equipment. Psychology and sociology researchers more often use Cs of .90 since there is so much more variability in their human subjects.

Margin of Error

- We want the margin of error, $Z^*(\frac{\sigma}{\sqrt{n}})$, to be small since that means the width of the CI will be smaller and the parameter will be bounded more compactly.

- Z^* and σ are in the numerator of the margin or error and n is in the denominator. So, the margin of error gets smaller when:

1. Z^* is smaller. The smaller Z^* is, the lower the confidence level is.
2. σ is smaller, which implies lower variability in the population.
3. n gets larger. Increasing the sample size reduces the margin of error for any fixed C.

— #1 and #2 can't always be controlled by the researcher. You may have a fixed C that you're willing to accept, and σ can't always be changed. The best way to reduce the margin of error is to increase the sample size.

Example:

(a) Find a 95% CI for the mean steer weight in the previous example. Compare the margins of error and widths for this CI and the 90% CI that we already calculated.

$\bar{X}$= σ= n=

Z^*=

$\bar{X} \pm Z^*(\frac{\sigma}{\sqrt{n}})$=

The 90% CI for μ is

	90%	95%
Margin of Error		
Width		

(b) If the sample size is increased to n=100, find a 90% CI.

$\bar{X} \pm Z^*(\frac{\sigma}{\sqrt{n}})$=

The 90% CI for μ is

Is this margin of error smaller or larger than the margin of error for a 90% CI of n=50?