Lecture 22

Today we start on the road towards formal statistical inference. As mentioned previously the purpose of statistical inference is too draw conclusions about something based upon data (eg draw conclusions about the population parameter based on sample data)

eg Suppose we give one group of patients a placebo and another group an active pill (ie a pill containing the drug). Assume that patients are assigned to groups randomly. For group one (placebo) assume the measurements of some pertaurbant variable has mean $\bar{x} = 15$ and standard deviation $s = 0.5$ and for group ~~two~~ (drug) the variable has mean $\bar{y} = 18$ with std dev = 1.5. Do we have evidence that the drug works? This is the sort of problem statistical inference is designed to address.

Statistical inference is based on the idea that
- data is from random sample (or randomized experiment)
- based on what would happen if we used the inference technique many times. (ie based on sampling distributions)
- two main types
  - Confidence intervals - for estimating value of population parameter → today
  - Hypothesis tests - for assessing evidence for a claim. → next time

Confidence interval

Consists of two parts

1. An interval of form

$$\text{estimate} \pm \text{margin of error}$$

this is computed based the sample data

Sometimes written in form (estimate − margin, estimate + margin)

2. A confidence level (this is the probability that the method gives an interval that contains true value of parameter)

Formal Definition

A level C confidence interval for a parameter is an interval computed from sample data by a method that has probability C of producing an interval containing the true value of the parameter. (see 3A for a little bit more)

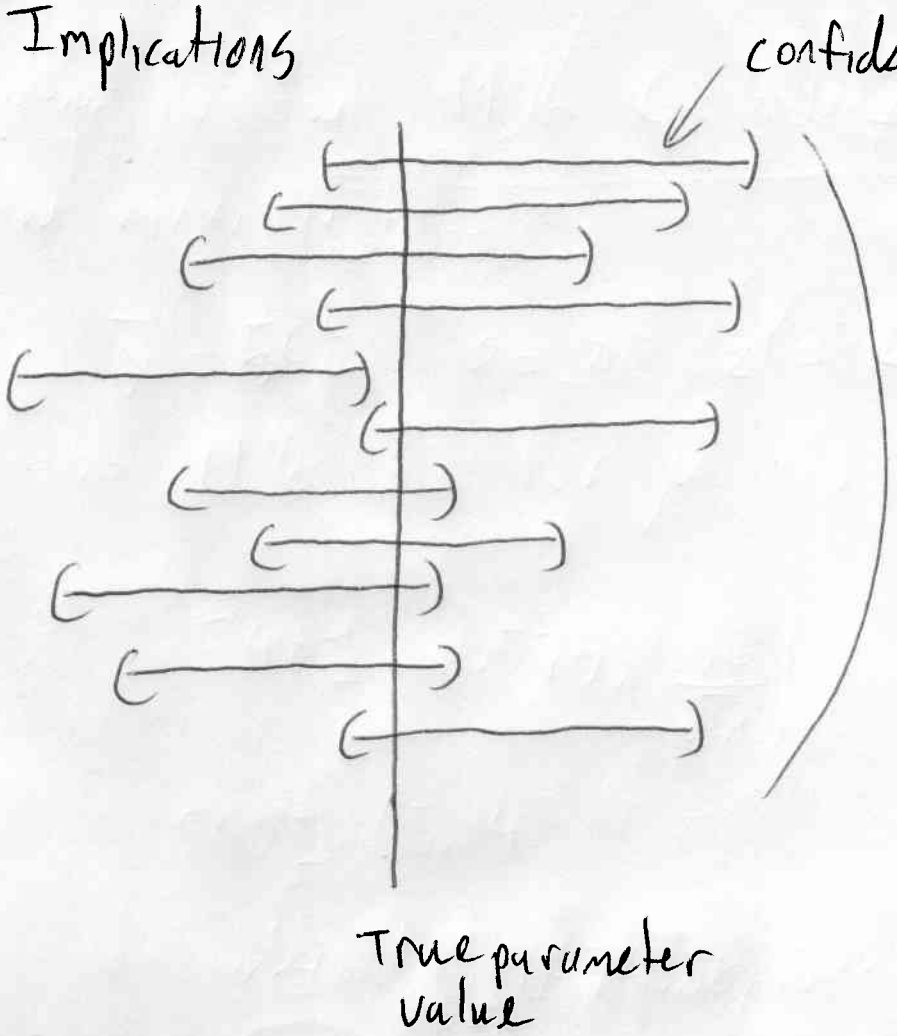
Confidence interval for population mean

(here we assume that σ is known)

Recall that the sampling distribution of the sample mean is $N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$ (exact when the population is normal, approximate by the CLT when n is large. This tells us that there is probability C that $\bar{x}$ lies in

(skip to 4)

Implications



Note We can not say whether the specific interval we compute contains the true value of the parameter or not. Only that the procedure will give an interval containing the true value C% of the time.

(cont from 3)

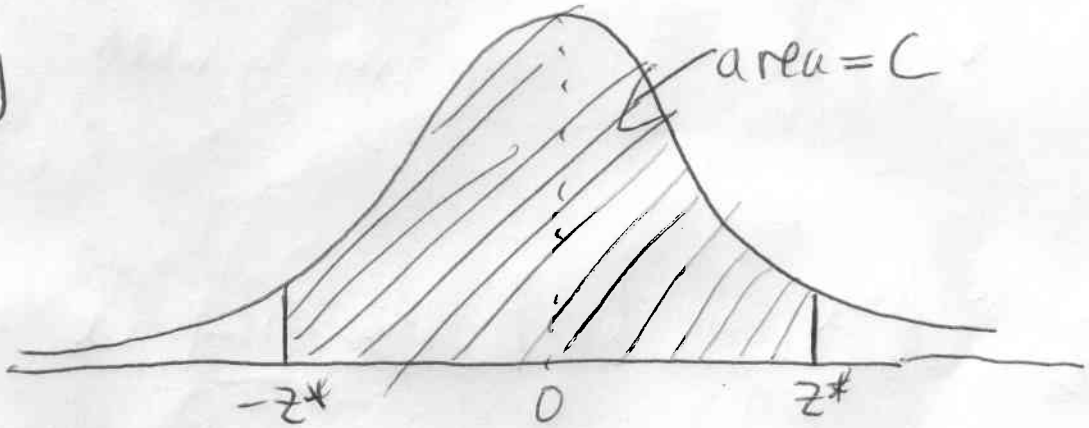
The interval

$$\left(\mu - z^* \frac{\sigma}{\sqrt{n}} \;,\; \mu + z^* \frac{\sigma}{\sqrt{n}}\right)$$

where z^* is the number such that

$$P(-z^* < z < z^*) = C$$

eg



However in real life μ is unknown so we use the equivalent interval

$$\left(\bar{x} - z^* \frac{\sigma}{\sqrt{n}} \;,\; \bar{x} + z^* \frac{\sigma}{\sqrt{n}}\right)$$

as an interval for μ

More formally

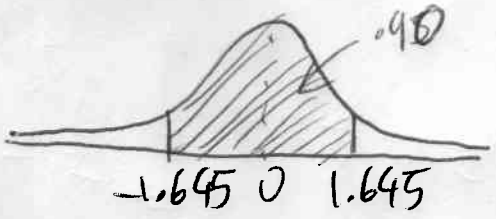
Given a SRS of size n from a population with unknown mean μ and known s.d. σ a level C confidence interval for μ is

$$\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}$$

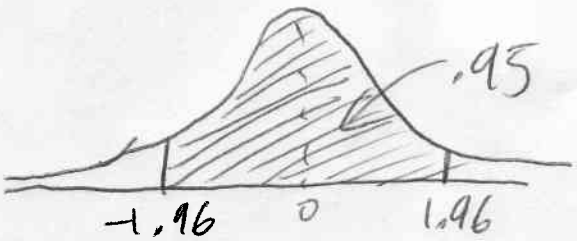
where $P(-z^* < z < z^*) = C$.

Typical values for z^*

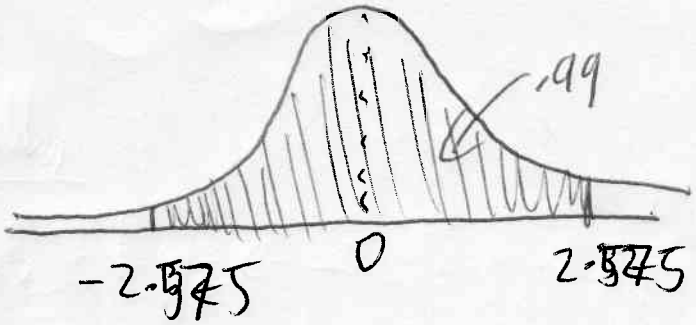
C	Probability	z^*
90%	$P(-1.645 < z < 1.645) = .90$	1.645



95%	$P(-1.96 < z < 1.96) = .95$	1.96



99%



2.575

$P(-2.575 < Z < 2.575) = .99$

Example

Suppose you are interested in renting an apartment. A simple random sample of 10 apartments advertised in the newspaper has mean \$540 and std dev \$80 is known about the population. What is a 95% confidence interval for the mean rent in this community?

$\bar{x} = 540 \quad \sigma = 80 \quad z^* = 1.96 \quad n = 10$

So 95% CI for μ (the mean rent in community) is

$$540 \pm (1.96)\left(\frac{80}{\sqrt{10}}\right)$$

$$\Rightarrow 540 \pm 49.58$$

$$\text{i.e. } (490.41, 589.58)$$

which means 95% of the time we gather a sample of size 10 this interval will contain true mean

what is the 99% Confidence interval for the mean rent?

$\bar{X} = 540$, $\sigma = 80$ $z^* = 2.575$ $n = 10$

so 99% CI for μ is

$$540 \pm (2.575)\left(\frac{80}{\sqrt{10}}\right)$$

$$540 \pm 65.14$$

ie interval is $(474.86, 605.14)$

In general the "margin of error" decreases as

- the sample size n increases
- the population sd σ decreases
- the confidence level C decreases

Note your book covers how to get the sample size for a confidence interval with a certain length. I will not cover this topic in this class.