More about using a random sample mean to estimate a population mean

Using a random sample mean to estimate a population mean

Statisticians have figured out:

1. Expect sampling error (see the last video)
$\mu \neq \bar{x}$ (unless really lucky)

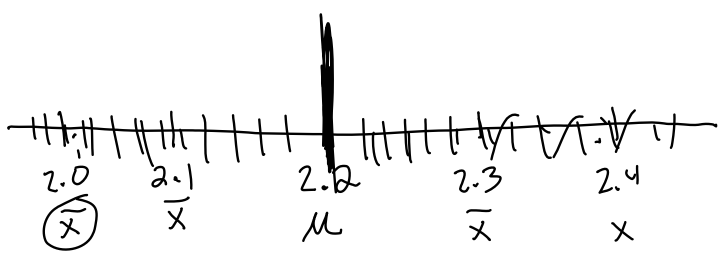
2. A random sample mean is an UNBIASED estimate of a population mean.

3. A random sample mean is a CONSISTENT estimate of a population mean

UNBIASED

A random sample mean is just as likely to be above the population mean as it is to be below the population mean.

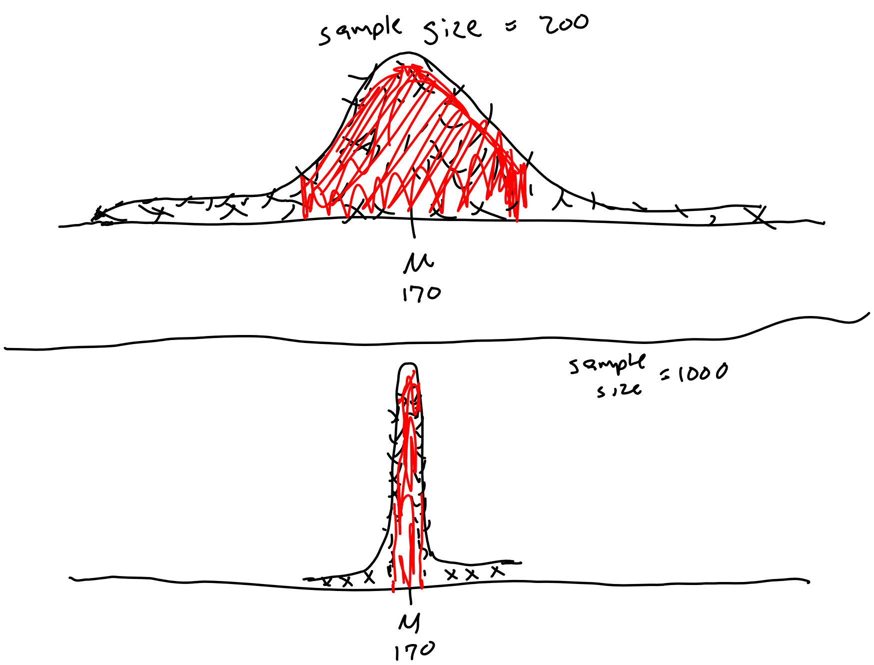
(If somehow the researcher were able to take a whole bunch of random samples, then the average mean of those random samples would equal the population mean.)



CONSISTENT

Larger random sample sizes will usually give you a sample mean that is closer to the population mean than smaller sample sizes.

(Rule of thumb: please always try to have a sample size that is greater than 30 (preferable much greater than 30 if the population if large)).



Consistent, part II:

Suppose somehow that you could take a whole bunch of random samples, each of size "n". The standard deviation of your sample means equals this number:

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

population standard deviation

size of sample

We are mini-god doing a thought experiment.

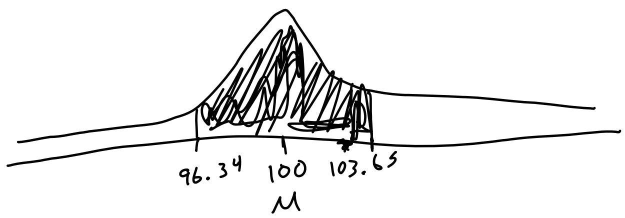
We know that the mean IQ of the world's population is 100 and the standard deviation is 20.

We wonder: if we take a random sample of the world's people, how does the sample size influence the accuracy of the sample mean as an estimate of the true population mean of 100?

If the sample size is 30, then the standard deviation of the sample means is

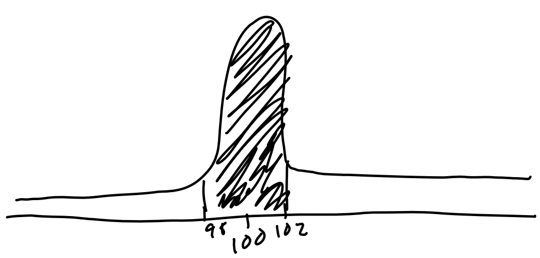
$$\frac{\sigma}{\sqrt{n}} = \frac{20}{\sqrt{30}} = 3.651484$$

random sample size = 30



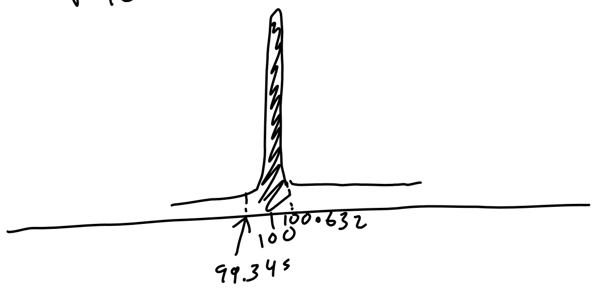
If the sample size is 100, then the standard deviation of the sample means is

$$\frac{\sigma}{\sqrt{n}} = \frac{20}{\sqrt{100}} = 2$$



If the random sample size is 1000, then the standard deviation of the sample means is

$$\frac{\sigma}{\sqrt{n}} = \frac{20}{\sqrt{1000}} = .632$$



Moral of the story (a very vague rule of thumb)

If the researcher can take a true random sample of 1000 from a population then she will usually get a pretty good estimate of the population mean by using the sample mean.

Example: Cereal Inc is sure that the standard deviation of ounces of cereal per box is 2 ounces. It THINKS that the mean ounces of cereal in all of its boxes is 20 ounces but it is not sure. It takes a random sample of 1000 boxes; the sample mean is 19 ounces.

Given the results of the sample, should Cereal Inc stick to its belief that the mean ounces of cereal in all of its boxes is 20 ounces?

$$\frac{\sigma}{\sqrt{n}} = \frac{2 \text{ ounces}}{\sqrt{1000}} = \pm .06 \text{ ounces}$$

$$\frac{20-19}{.06} = 15.8 \text{ standards deviations away!}$$

The sample mean of 19 ounces is 15.8 standard deviations away from 20 ounces. This means that it is virtually impossible that the true population mean is 20 ounces. So cereal inc should stop believing that its machines are filling cereal boxes with of mean of 20 ounces of cereal.