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

π — population proportion

p — sample proportion

proportion ranges from 0 to 1

0 → 0%, 1 → 100%

Using a random sample proportion to estimate a population proportion

Statisticians have figured out:

1. Expect sampling error (see the last video)

   $p \neq \pi$ (unless you're really ~~likely~~ lucky)

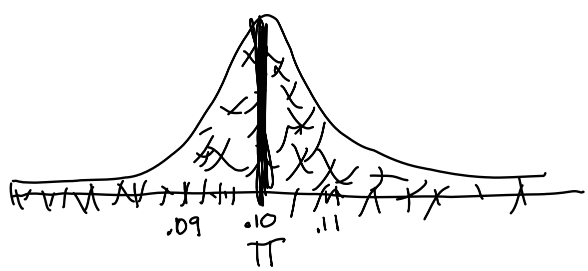
2. A random sample proportion is an UNBIASED estimate of a population proportion.

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

UNBIASED

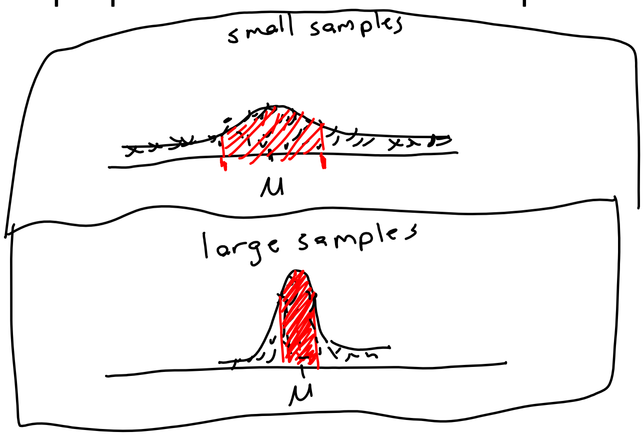
A random sample proportion is just as likely to be above the population proportion as it is to be below the population proportion.

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



CONSISTENT

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



Consistent, part II:

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

$$\sqrt{\frac{\pi(1-\pi)}{n}}$$

π is population proportion

n is number of items in sample

(This formula does not apply to small sample sizes.)

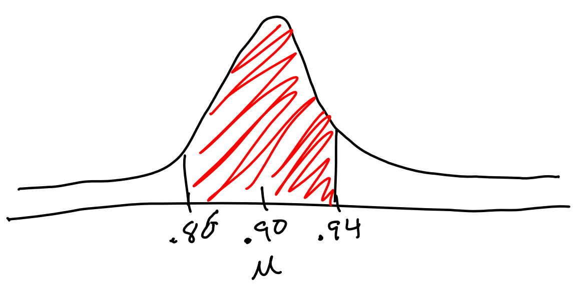
We are mini-god doing a thought experiment.

We know that the proportion of the world's population with brown eyes is .90

We wonder: if we take a random sample of the world's people, how does the sample size influence the accuracy of the sample proportion as an estimate of the true population proportion of .90?

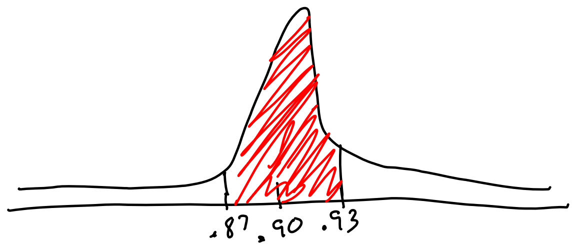
If the sample size is 50, then the standard error of the sample proportion is

$$\sqrt{\frac{\pi(1-\pi)}{n}} = \sqrt{\frac{.90(1-.9)}{50}} = .04$$



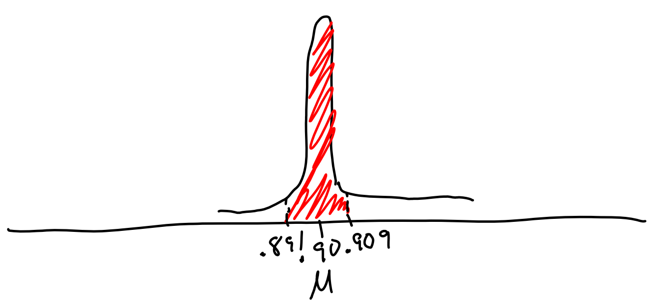
If the sample size is 100, then the standard error of the sample proportions is

$$\sqrt{\frac{\pi(1-\pi)}{n}} = \sqrt{\frac{.90(1-.90)}{100}} = .03$$



If the sample size is 1000, then the standard error of the sample proportions is

$$\sqrt{\frac{\pi(1-\pi)}{n}} = \sqrt{\frac{.90(1-.90)}{1000}} = .009$$



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 proportion by using the sample proportion.

Example: Karl Rove thinks that Sarah Palin has a 50% approval rating among U.S. adults, but he is not sure. He takes a random sample of 1000 U.S. adults; 49% of this sample approves of Sarah Palin.

Should Karl Rove continue to believe that Sarah Palin has a 50% approval rating among U.S. adults?

1.6%

$$\sqrt{\frac{\pi(1-\pi)}{n}} = \sqrt{\frac{.50(1-.50)}{1000}} = .016$$

$$.50 - .49 = \frac{.01}{.016}$$

$$= .632$$

The sample proportion, .49 is only .632 standard errors away from what Rove thinks is the population proportion, .50.

This is not enough standard deviations away for Rove to conclude that the true population proportion is less than .50 (50%).

(Roughly speaking, you need to have an estimate that is more than two standard deviations away for it to provide convincing evidence.)