

22S:105
Statistical Methods and Computing

Sampling Distributions

Lecture 11
 Mar. 3, 2014

Kate Cowles
 374 SH, 335-0727
 kate-cowles@uiowa.edu

To answer this question, we would like to

- Take a large number of samples of size 10 from the same population
- Calculate the sample mean \bar{x} for each sample
- Make a histogram of the values of \bar{x}
- Examine the distribution displayed in the histogram for shape, center, spread, and outliers

Sampling distributions

Suppose our sample size was 10.

What can we say about \bar{x} from a sample of size 10 as an estimate of μ ?

We can get an idea of how good an estimator \bar{x} is likely to be by asking "What would happen if we took many samples of 10 subjects from this population?"

Simulating the sampling distribution of \bar{x}

In practice, it is too difficult and expensive to draw many samples from a large population such as all adult Chinese males. But we can imitate random sampling by using a computer to do *simulation*.

In this way we can study the distribution of sample means.

- Imagine that we knew that upper arm skin-fold thicknesses in Chinese males followed a normal distribution with mean $\mu = 10$ mm and standard deviation $\sigma = 3$ mm.
- Then we could use the computer to generate many random samples of size 10 from this distribution and calculate \bar{X} from each of these samples

The **sampling distribution** of a statistic is the distribution of values taken by the statistic in *all possible* samples of the same size from the same population.

Our simulated distribution is only an approximation to the true sampling distribution, but it gives us an idea of what it would look like.

The Law of Large Numbers

Assume that observations are drawn at random from any population with finite mean μ .

As the number of observations increases, the sample mean \bar{x} of the observed values gets closer and closer to the true mean μ of the population.

The mean and standard deviation of \bar{x}

Let \bar{x} be the sample mean of a simple random sample of size n drawn from a large population with mean μ and standard deviation σ .

- Then the *mean* of the sampling distribution of \bar{x} is μ .
- The *standard deviation* is of the sampling distribution of $\frac{\sigma}{\sqrt{n}}$.
 - Averages are less variable than individual observations.
 - The results of large samples are less variable than the results of small samples

The central limit theorem

We have described the center and spread of the sampling distribution of \bar{x} . What about its shape?

The shape depends on the shape of the population distribution.

Special case:

- If a population has a *normal* distribution with mean μ and standard deviation σ , then the sample mean \bar{x} of a random sample of n observations has a *normal* distribution with mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$

The **Central Limit Theorem** says that

- Provided that n is large enough, the shape of the sampling distribution of \bar{x} is approximately normal.

Example: birthweights

We have been told that the birthweights of individual infants in the US follow a normal distribution with mean 120 oz. and standard deviation 15 oz.

What is the sampling distribution of the mean in samples of size 25 from this population?

Normal with mean 120 oz and standard deviation $\frac{15}{5} = 3$ oz.

Deciding whether \bar{x} in a particular sample is unusual

Suppose that the sample mean $\bar{x} = 115$ oz for a sample of 25 infants. Is this sample unusual?

We will use our standard approach of finding the probability of an event as *extreme* as this *or more so*, that is

$$P(\bar{X} \leq 115)$$

We can use z-scores to do this.

$$\begin{aligned} z &= \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \\ &= \frac{115 - 120}{15/5} \\ &= \frac{-5}{3} = -1.67 \end{aligned}$$

We then use Table A to find

$$P(Z \leq -1.67) = .0465$$

What are the upper and lower limits that enclose 95% of the means of samples of size 25 drawn from this population?

We begin by finding the value z such that 2.5% of the area under the standard normal curve lies below z .

- This is $z = -1.96$.
- Then 2.5% of the area under the standard normal curve lies *above* 1.96.

In other words

$$P(-1.96 \leq Z \leq 1.96) = 0.95$$

Now we have to *unstandardize*

$$\begin{aligned} -1.96 &\leq \frac{\bar{x} - 120}{3} \leq 1.96 \\ -1.96(3) &\leq \bar{x} - 120 \leq 1.96(3) \\ 120 - 1.96(3) &\leq \bar{x} \leq 120 + 1.96(3) \\ 114.12 &\leq \bar{x} \leq 125.88 \end{aligned}$$

In other words, 95% of all the possible samples of size 25 drawn from this population would have sample means between 114.12 and 125.88 oz.