

Random Sampling

1. each possible sample of a given size has equal chance of being selected
2. each individual member of the population has equal chance of being selected into the sample

Why do we take random samples?

- to generalize from a sample to a population, we must be able to apply the laws of probability
- to generalize from a sample to a population, the sample must be representative of the population

Sampling with or without replacement [DEFINE]

So why sample with replacement? Generally, we don't sample with replacement. It happens that the mathematical techniques used to compute inferential statistics are simpler in the "with replacement" case. When the sample size is small compared to the population size, the differences are negligible.

Reading Study Example

You want to know if students in your school are faring well nationally, you might compare the average reading scores of your students with the national norms. Say the scores nationally have a mean of 75 and a standard deviation of 16. You randomly select 100 students and obtain a sample mean of 72. What do you conclude?

You need to know something about the sampling distribution of the mean.

Sampling Distribution of the Mean

Draw the population (μ , σ)

Take all possible samples of a given size N

Calculate the mean of each sample

Calculate the probability of each mean

The sampling distribution of the mean:

1. is a distribution of scores where each score is a sample mean
2. a mean equal to the mean of the raw-score population: $\mu_{\bar{x}} = \mu$
3. a standard deviation equal to: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}}$
4. has a normal shape, given the shape of the raw-score population and sample size

Note: As the number of scores in each sample increases, each sample mean becomes a more accurate estimate of the population mean. This will mean that the sample means will vary less.

There is a different sampling distribution for each different sample size.

The Central Limit Theorem suggests that regardless of the shape of the population of raw scores, the sampling distribution of the mean approaches a normal distribution as sample size N increases.

With most sample sizes ($N \geq 30$), the sampling distribution of the mean will be normal. For highly skewed distributions $N \geq 300$ is required to achieve normality.

Reading Example

From the sample of 100, we can determine the sampling distribution

$$\mu_{\bar{X}} = \mu = 75, \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{N}} = \frac{16}{\sqrt{100}} = 1.6$$

Now we can compute z_{obt} for our sample mean and compare it to the probability of obtaining a sample mean equal or less than 72, if chance is responsible (“if the sample is a random sample from the null-hypothesis population”)

$$z_{obt} = \frac{\bar{X}_{obt} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} = \frac{72 - 75}{1.6} = -1.88$$

Is the obtained mean a random sample from a population where $\mu = 75$

$p(\bar{X}_{obt} \leq 72) = 0.0301$, thus the assumption is unreasonable, reject H_0