

Cheat Sheet: Introduction to Sampling Distributions

Essential Concepts

- One of the goals of statistical inference is to draw a conclusion about a population on the basis of a random sample from the population. Obviously, random samples vary, so we need to understand how much they vary and how they relate to the population. Our ultimate goal is to create a probability model that describes the long-run behavior of sample measurements. We use this model to make inferences about the population.
- In statistics, when we want to describe the characteristics of a sample, we call the values statistics. However, when we want to describe the characteristics of a population, we call those values parameters.
- We can use mathematical theory to derive expressions for the mean and standard deviation of the sampling distribution of the sample proportion. When taking many random samples of size n from a population distribution with population proportion p :
 - The mean of the distribution of sample proportions is p .
 - The standard deviation of the distribution of sample proportions is $\sqrt{\frac{p(1-p)}{n}}$.
- In order to get a sense of the pattern of variation in sample proportions, we need to generate more than five samples. The distribution showing how sample proportions vary from sample to sample is called a sampling distribution of the sample proportion.
- The Central Limit Theorem states that, as the sample size gets larger, the distribution of the sample proportion will become closer to a normal distribution.
- When taking many random samples of size n from a population distribution with population proportion p :
 - The mean of the distribution of sample proportions is p .
 - The standard deviation of the distribution of sample proportions is $\sqrt{\frac{p(1-p)}{n}}$.

- If $np \geq 10$ and $n(1 - p) \geq 10$, then the Central Limit Theorem states that the distribution of the sample proportions follows an approximate normal distribution with mean p and standard deviation $\sqrt{\frac{p(1-p)}{n}}$
 - In practice, we do not know the population proportion, nor do we have the luxury of taking thousands of random samples. Instead, we observe a single random sample. In this case, we need to estimate the mean and standard deviation of the sample proportion:
 - The estimated mean of the distribution of sample proportions is \hat{p} .
 - To distinguish it from the true standard deviation of sample proportions, we call the estimated standard deviation of sample proportions the standard error of \hat{p} :
-

Key Equations

sample size

$$n = \frac{p(1-p)}{(\sigma_p)^2}$$

standard deviation

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

standard error

$$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Glossary

Central Limit Theorem

If $np \geq 10$ and $n(1 - p) \geq 10$, then the Central Limit Theorem states that the distribution of the sample proportions follows an approximate normal distribution with mean p and standard deviation $\sqrt{\frac{p(1-p)}{n}}$

parameters

numbers that describe a population

population

the population is the entire collection of individuals or objects that you want to learn about

sample

a sample is a part of the population that is selected for study

sampling distribution of a statistic

a probability distribution that describes the long-term behavior of the statistic.

sampling distribution of a sample proportion

a probability distribution that describes how sample proportions vary from sample to sample

standard deviation

a measure that describes the variability of a population

standard error

an estimate of the variability across the samples of a population

statistics

numbers that are calculated from a sample