

DATA SCIENCE WITH R

HYPOTHESIS TESTING

Introduction to Hypothesis Testing

Basic Framework of a Hypothesis Test

Distance Measures



Central Limit Theorem

Types of Hypothesis Tests



Central Limit Theorem

Applied in hypothesis testing to aid in calculating probability or chance



Central Limit Theorem

Applied in hypothesis testing to aid in calculating probability or chance

Example:

There is a population of 10,000 respondents to a survey, and we pick a sample of 500. We review data on one of the attributes, income.



Central Limit Theorem

Applied in hypothesis testing to aid in calculating probability or chance

Example:

There is a population of 10,000 respondents to a survey, and we pick a sample of 500. We review data on one of the attributes, income.

1. How many samples are possible?



Central Limit Theorem

Applied in hypothesis testing to aid in calculating probability or chance

Example:

There is a population of 10,000 respondents to a survey, and we pick a sample of 500. We review data on one of the attributes, income.

1. How many samples are possible?
2. What will be the mean income of the different samples?



Central Limit Theorem

Applied in hypothesis testing to aid in calculating probability or chance

Example:

There is a population of 10,000 respondents to a survey, and we pick a sample of 500. We review data on one of the attributes, income.

1. How many samples are possible?
2. What will be the mean income of the different samples?
3. What if we plot a frequency distribution of the sample means in Excel?



Central Limit Theorem

Central Limit Theorem:

As sample size grows sufficiently large, the sampling distribution of the means will tend towards a normal distribution (even if the underlying population is not normal)



Central Limit Theorem

Central Limit Theorem:

As sample size grows sufficiently large, the sampling distribution of the means will tend towards a normal distribution (even if the underlying population is not normal)

Mathematically:

When we select simple random samples of size n , the distribution of these sample means can be modeled with a probability model that is



Central Limit Theorem

Central Limit Theorem:

As sample size grows sufficiently large, the sampling distribution of the means will tend towards a normal distribution (even if the underlying population is not normal)

Mathematically:

When we select simple random samples of size n , the distribution of these sample means can be modeled with a probability model that is

$$N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$



Central Limit Theorem

Central Limit Theorem:

As sample size grows sufficiently large, the sampling distribution of the means will tend towards a normal distribution (even if the underlying population is not normal)

Mathematically:

When we select simple random samples of size n , the distribution of these sample means can be modeled with a probability model that is

$$N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Implications

If sample size is sufficiently large (>30), you can always use a normal distribution as your test distribution without worrying about true population distribution



Recap

➤ Central Limit Theorem



THANK YOU

