

What is the Assumption of Independence in Statistics?

Authored by
stats writer

April 27, 2024

RECOMMENDED CITATION

stats writer (2024). *What is the Assumption of Independence in Statistics?*.

PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=139929>

The Assumption of Independence in Statistics is a fundamental concept that assumes each data point or observation in a dataset is independent from one another. This means that the value of one data point does not affect or influence the value of another data point. This assumption allows for the use of statistical methods and models to analyze and make inferences about a population based on a sample. It is a crucial assumption in many statistical tests and models, as violations of independence can lead to inaccurate results and conclusions. Therefore, it is important to carefully consider and assess the independence of data points before conducting any statistical analysis.

What is the Assumption of Independence in Statistics?

Many statistical tests make the assumption that are independent. This means that no two observations in a dataset are related to each other or affect each other in any way.

For example, suppose we want to test whether or not there is a difference in mean weight between two species of cats. If we measure the weight of 10 cats from species A and 10 cats from species B, we would violate the assumption of independence if each of the groups of cats came from the same litter.

It's possible that the mother cat of species A simply had all low-weight kittens while the mother cat of species B had heavy kittens. In this regard, the observations in each sample are not independent of each other.

There are three common types of statistical tests that

make this assumption of independence:

1.

2.

3.

In the following sections, we explain *why* this assumption is made for each type of test along with how to determine whether or not this assumption is met.

Assumption of Independence in t-tests

A two sample t-test is used to test whether or not the means of two populations are equal.

Assumption: This type of test assumes that the observations *within* each sample are independent of each other and that the observations *between* samples are also independent of each other.

Test this Assumption: The easiest way to check this assumption is to verify that each observation only appears in each sample once and that the observations in each sample were collected using random sampling.

Assumption of Independence in ANOVA

An ANOVA is used to determine whether or not there is a significant difference between the means of three or more independent groups.

Assumption: An ANOVA assumes that the observations in each group are independent of each other and the observations within groups were obtained by a random sample.

Test this Assumption: Similar to a t-test, the easiest way to check this assumption is to verify that each observation only appears in each sample once and that the observations in each sample were collected using random sampling.

Assumption of Independence in Regression

Assumption: Linear regression assumes that the in the fitted model are independent.

Test this Assumption: The easiest way to check this assumption is to look at a residual time series plot, which is a plot of residuals vs. time. Ideally, most of the residual autocorrelations should fall within the 95%

confidence bands around zero, which are located at about ± 2 -over the square root of n , where n is the sample size. You can also formally test if this assumption is met using the .

Common Sources of Non-Independence

There are three common sources of non-independence in datasets:

1. Observations are close together in time.

For example, a researcher may be collecting data on the average speed of cars on a certain road. If he chooses to track the speeds during the evening hours, he may find that the average speed is much higher than he expected simply because each driver is rushing home from work.

This data violates the assumption that each observation is independent. Since each observation was observed during the same time of day, the speed of each car is likely to be similar to each other.

2. Observations are close together in space.

For example, a researcher may collect data on the annual income of individuals who happen to all live in the same high-income neighborhood because it's convenient to do so.

In this regard, all of the individuals included in the sample data are likely to have similar incomes since they all live near each other. This violates the assumption that each observation is independent.

3. Observations appear multiple times in the same dataset.

For example, a researcher may need to collect data for 50 individuals but instead decides to collect data on 25 individuals twice because it's much easier to do so.

This violates the assumption of independence because each observation in the dataset will be related to itself.

How to Avoid Violating the Assumption of Independence

The easiest way to avoid violating the assumption of independence is to simply use when obtaining a sample from a population.

Using this method, every individual in the of interest has an equal chance of being included in the sample.

For example, if our population of interest contains 10,000 individuals then we may randomly assign a number to every individual in the population and then use a random number generator to select 40 random numbers. The individuals who match up with these numbers would then be included in the sample.

Using this method, we minimize the chances that we select two individuals who may be in close proximity to each other or who may be related in some way.

This is in direct contrast to other sampling methods such as:

Convenience sampling: Including individuals in a sample who are simply convenient to reach. Voluntary sampling: Including individuals in a sample who to be included.

By using a random sampling method, we can minimize the chances that we violate the assumption of independence.