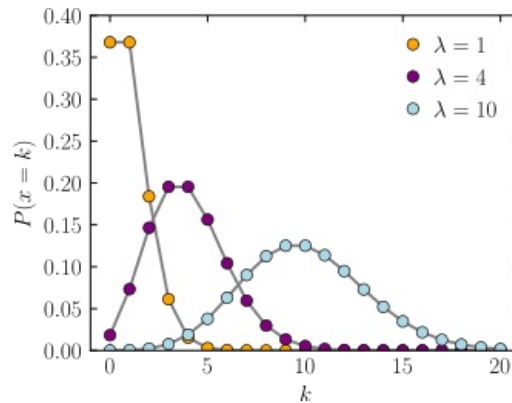# Poisson distribution

A **Poisson distribution'** (distribution) is a statistical distribution of a random variable determining the **number of occurrences of events over time**, where these events occur randomly and independently of each other. It is typical of the Poisson distribution that it is governed by the frequencies of events, which are relatively rare − therefore, they have a very ""small probability of occurrence"" (for example, it can be the number of bleeding events in the monitored persons per 100 person-years and .). It is therefore sometimes possible to come across a less typical designation of this particular distribution, namely the ``distribution of sparse phenomena*.

The random variable itself has a distribution: $Po(\lambda)$.

Prerequisites

For a random variable to follow the Poisson distribution, the following conditions must be met:

1. a certain event can occur at **any moment in time**;
2. the number of events over time depends only on the time interval − but its beginning or end does not matter;
3. λ is considered to be the **mean value of the number of occurrences of phenomena per time unit'**.

The random variable λ_i depends on k variables labeled x_i, where $i = 1, 2, \ldots, k$. So the final notation would be $x_1, x_2, \ldots, x_k$, which are quantities we call **covariates**. ***The job of the Poisson distribution is to find** relationships between them''*. To be able to find these relations, we define a function:

$$\lambda_i = \lambda(x_i) = \lambda(x_1, \ldots, x_k).$$

Formula

$$P(X = x) = \frac{\lambda^x}{x!} e^{-\lambda}$$

e represents **Euler number**.

Example

Incidence of gallbladder cancer

Let's consider that five cases of gallbladder cancer occur in one region in the male population, aged between 35 and 40 years. We consider the variable X to be the number of cases of this disease during the year. Since we know that there are five cases of gallbladder cancer per year, the variable X will follow a Poisson distribution of the form $\lambda = 5(1) = 1$. Our goal is to find the probabilities that cancer will 1) not occur, 2) occur once, 3) occur twice in a given year:

1. $P(X = 0) = \dfrac{e^{-5}}{0!} (5)^0 = e^{-5} = 0,00674$

2. $P(X = 1) = \dfrac{e^{-5}}{1!} (5)^1 = 5e^{-5} = 0,03369$

3. $P(X = 2) = \dfrac{e^{-5}}{2!} (5)^2 = 12,5e^{-5} = 0,08422$

From these results, it follows that the probability of the absence of the disease in a given year is less than one percent, the probability of a single occurrence of gallbladder cancer is around three percent, and the occurrence of three cases is 8.4%.

Incidence of leukemia

With the second example, we will show how we must proceed to calculate the expected cases of events if we want to reflect the size of the monitored population and the time interval.

Assume that the annual incidence of leukemia is 11.2 cases per 100,000 persons. If we follow these 100,000 people for one year, what is the probability that we will not see any new cases of this disease? And what will this probability be if it is applied to only 1000 persons?

Both groups must be adequately divided. For 100,000 people, the incidence of leukemia is 11.2, that is:

$$P(X=0) = \frac{e^{-11,2}}{0!}(11,2)^0 = 0.0000137$$

For a group of 1000 people, the incidence of new cases is equal to 0.112 ((11.2/100,000)x1000). Therefore, the formula will look like this:

$$P(X=0) = \frac{e^{-0.112}}{0!}(0.112)^0 = 0.89$$

The probability that leukemia will not occur in our hypothetical population of 100,00 individuals is well below one percent. Conversely, for a population of 1000 individuals, this probability is 89%.

Links

Related Articles

- Attributable risk
- Absolute Risk
- Relative risk
- Incidence
- Normal distribution
- Meta-analysis
- Confidential Interval

External links

Skripta o Poissonově distribuci s ukázkovými grafy, v anglickém jazyce (https://www.itl.nist.gov/div898/handbook/eda/section3/eda366j.htm)

Used literature

- WOOLSON, Robert F. – CLARKE, William. *Statistical Methods for the Analysis of Biomedical Data.* 2. edition. John Wiley & Sons. Inc., 2002. 368 pp. ISBN 9780471394051.