

What is a Conditional Distribution in Statistics?

Authored by
stats writer

April 25, 2024

RECOMMENDED CITATION

stats writer (2024). *What is a Conditional Distribution in Statistics?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=139180>

A conditional distribution in statistics is a probability distribution that shows the likelihood of an event occurring based on the occurrence or non-occurrence of another event or condition. It is used to understand the relationship between two variables and how their values affect the probability of a certain outcome. This type of distribution can provide valuable insights into the conditional dependencies and interactions between variables, and is often used in data analysis and decision-making processes.

What is a Conditional Distribution in Statistics?

If X and Y are two jointly distributed, then the **conditional distribution** of Y given X is the probability distribution of Y when X is known to be a certain value.

For example, the following two-way table shows the results of a survey that asked 100 people which sport they liked best: baseball, basketball, or football.

	Baseball	Basketball	Football	Total
Male	13	15	20	48
Female	23	16	13	52
Total	36	31	33	100

If we want to know the probability that a person prefers a certain sport *given* that they are male, then this is an example of a conditional distribution.

The value of one random variable is known (the person is male), but the value of the other random variable is unknown (we don't know their favorite sport).

To find the conditional distribution of sports preference among males, we would simply look at the values in the row for **Male** in the table:

	Baseball	Basketball	Football	Total
Male	13	15	20	48
Female	23	16	13	52
Total	36	31	33	100

The conditional distribution would be calculated as:

Males who prefer baseball: $13/48 = .2708$

Males who prefer basketball: $15/48 = .3125$

Males who prefer football: $20/48 = .4167$

Notice that the sum of the probabilities adds up to 1: $13/48 + 15/48 + 20/48 = 48/48 = 1$.

We can use this conditional distribution to answer questions like: *Given that an individual is male, what is the probability that baseball is their favorite sport?*

From the conditional distribution we calculated earlier, we can see that the probability is **.2708**.

In technical terms, when we calculate a conditional distribution we say that we're interested in a particular **subpopulation** of the overall population. The subpopulation in the previous example was males:

	Baseball	Basketball	Football	Total
Male	13	15	20	48
Female	23	16	13	52
Total	36	31	33	100

Subpopulation

And when we want to calculate a probability related to this subpopulation, we say that we're interested in a particular **character of interest**. The character of interest in the previous example was baseball:

Character of Interest

	Baseball	Basketball	Football	Total
Male	13	15	20	48
Female	23	16	13	52
Total	36	31	33	100

Subpopulation

To find the probability that the character of interest occurs in the subpopulation, we simply divide

the value of the character of interest (e.g. 13) by the total values in the subpopulation (e.g. 48) to get $13/48 = .2708$.

Conditional Distributions & Independence

We can say that random variables X and Y are independent if and only if the conditional distribution of Y given X is, for all possible realizations of X , equal to the unconditional distribution of Y .

For example, in the previous table can we see that the events "prefers baseball" and "male" are independent?

To answer this, let's calculate the following probabilities:

$P(\text{prefers baseball})$

$P(\text{prefers baseball} \mid \text{male})$ "prefers baseball, given that they are male"

The probability that a given individual prefers baseball is:

$P(\text{prefers baseball}) = 36/100 = .36$.

	Baseball	Basketball	Football	Total
Male	13	15	20	48
Female	23	16	13	52
Total	36	31	33	100

The probability that a given individual prefers baseball, given that they are male is

$P(\text{prefers baseball} \mid \text{male}) = 13/48 = .2708$.

	Baseball	Basketball	Football	Total
Male	13	15	20	48
Female	23	16	13	52
Total	36	31	33	100

Since $P(\text{prefers baseball})$ is not equal to $P(\text{prefers baseball} \mid \text{male})$, the random variables of Sports Preference and Gender are *not* independent.

Why Use Conditional Distributions?

Conditional probability distributions are useful because we often collect data for two variables (like Gender and Sports Preference) but we're interested in answering questions about probability when we happen to *know* the value of one of the variables.

In the previous example, we considered the scenario where we knew that a given individual was male and we simply wanted to know the probability that the individual preferred baseball.

There are many instances in real life where we happen to know the value of one variable and we can use a conditional distribution to find the probability of another variable taking on a certain value.