# 1.7 - The Multinomial Distribution

What if the respondents in the survey had three choices:

1. I feel optimistic,
2. I don't feel optimistic,
3. I am not sure?

What would you consider as an appropriate sampling distribution? This is not a binomial distribution anymore. We have more than two possible outcomes. The multinomial distribution, however, is in many ways an extension of the binomial distribution. Let's look at this new distribution...

## Origins

The multinomial distribution arises from an extension of the binomial experiment to situations where each trial has $k \geq 2$ possible outcomes.

Suppose that we have an experiment with

- *n* independent trials, where
- each trial produces exactly one of the events $E_1, E_2, \ldots, E_k$ (i.e. these events are mutually exclusive and collectively exhaustive), and
- on each trial, $E_j$ occurs with probability $\pi_j, j = 1, 2, \ldots, k$.

Notice that $\pi_1 + \pi_2 + \cdots + \pi_k = 1$. The probabilities, regardless of how many possible outcomes, will always sum to 1.

Let's define the random variables:

$X_1$ = number of trials in which $E_1$ occurs,

$X_2$ = number of trials in which $E_2$ occurs,

...

$X_k$ = number of trials in which $E_k$ occurs.

Then $X = (X_1, X_2, \ldots, X_k)$ is said to have a multinomial distribution with index *n* and parameter $\pi = (\pi_1, \pi_2, \ldots, \pi_k)$. In most problems, *n* is regarded as fixed and known.

The individual components of a multinomial random vector are binomial and have a binomial distribution,

$X1 \sim Bin(n, \pi_1)$,

$X2 \sim Bin(n, \pi_2)$,

...

$Xk \sim Bin(n, \pi_k)$.

The trials or each person's responses are independent, however, the components or the groups of these responses are not independent of each other. The sample sizes are different now and known. The number of responses for one can be determined from the others. In other words, even though the individual $X_j$ 's are random, their sum:

$$X_1 + X_2 + \cdots + X_k = n$$

is fixed. Therefore, the $X_j$'s are negatively correlated.

## Notation

If $X = (X_1, X_2, \ldots, X_k)$ is multinomially distributed with index *n* and parameter $\pi = (\pi_1, \pi_2, \ldots, \pi_k)$, then we will write:

$$X \sim Mult(n, \pi)$$

« 1.6.7 - Example of three hypothesis tests
1.7.1 - Distribution function »