**P-values**

In statistical hypothesis testing, the ***p*-value** or **probability value** is, for a given statistical model, the probability that, when the null hypothesis is true, the statistical summary (such as the absolute value of the sample mean difference between two compared groups) would be greater than or equal to the actual observed results.

1. **Hypothesis Testing**
2. Shape

   Description automatically generated
3. [**Hypothesis testing**](https://www.statisticshowto.datasciencecentral.com/probability-and-statistics/hypothesis-testing/#HTMean) is used to test the validity of a claim *(****null hypothesis****)* that is made about a population using sample data.



1. The ***alternative hypothesis***is the one you would believe if the null hypothesis is concluded to be untrue.







we’ll make a claim *(null hypothesis)* and use a sample data to check if the claim is valid.

If the claim isn’t valid, then we’ll choose our *alternative hypothesis* instead

To know if a claim is valid or not, we’ll use **a p-value to weigh the strength of the evidence to see if it’s statistically significant.** If the evidence supports the *alternative hypothesis*, then we’ll reject the *null hypothesis* and accept the *alternative hypothesis*

🍕 🍕 [**Example**](https://www.dummies.com/education/math/statistics/what-a-p-value-tells-you-about-statistical-data/)**:** Suppose a pizza place claims their delivery times are 30 minutes or less on average but you think it’s more than that. So you conduct a hypothesis test and randomly sample some delivery times to test the claim

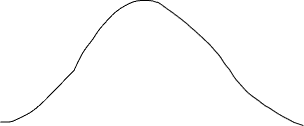
* **Null hypothesis** — The mean delivery time is 30 minutes or less
* **Alternative hypothesis** — The mean delivery time is greater than 30 minutes

**Ho**: delivery <=30

**Ha**: delivery > 30



The goal here is to determine which claim — the null or alternative — is better supported by the evidence found from our sample data.



What we want to test here is to see if there is a chance that the mean delivery time is greater than 30 minutes. In other words, we want to see if the pizza place lied to us somehow.😂

One of the common ways to do the hypothesis testing is to use [**Z-test**](https://www.analyticsvidhya.com/blog/2015/09/hypothesis-testing-explained/).

Chart, histogram

Description automatically generated

The normal distribution has two parameters — the **mean (μ)** and **standard deviation, also called sigma (σ).**

The **mean** is the central tendency of the distribution. It defines the location of the peak for normal distributions.

The **standard deviation** is a measure of variability. It determines how far away from the mean the values tend to fall.

The normal distribution is commonly associated with the [**68-95-99.7 rule**](https://en.wikipedia.org/wiki/68%E2%80%9395%E2%80%9399.7_rule)(image above).

* 68% of the data is within 1 standard deviation (σ) of the mean (μ)
* 95% of the data is within 2 standard deviations (σ) of the mean (μ)
* 99.7% of the data is within 3 standard deviations (σ) of the mean (μ)

you may be wondering, “**How does the normal distribution apply to our previous hypothesis testing?”**

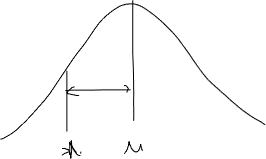
Since we used **Z-test to conduct our hypothesis testing,**

we need to calculate [**Z-scores**](https://www.statisticshowto.datasciencecentral.com/probability-and-statistics/z-score/) (to be used in our [**test-statistic**](https://www.statisticshowto.datasciencecentral.com/test-statistic/)) which is the number of standard deviations from the mean a data point is.

In our case, **each data point is the pizza delivery time that we collected.**

Graphical user interface, text

Description automatically generated with medium confidence



Notice that when we’ve calculated all the Z-scores for each pizza delivery time and plotted a **standard normal distribution** curve as below, the unit on the X-axis will change from minutes to standard deviation unit since we’ve **standardized the variable by subtracting the mean and dividing by its standard deviation** (see the formula above).

Chart, histogram

Description automatically generated

A Z-score can tell us where the overall data lies compared to the average population.

**The higher or lower the Z-score, the more unlikely the result is to happen by chance and the more likely the result is meaningful.**

Recall that we randomly sampled some pizza delivery times and the goal is to check if the mean delivery time is greater than 30 minutes.

If the **final evidence supports** the claim by the pizza place (mean delivery time is 30 minutes or less), then we **will not reject the null hypothesis.**

Otherwise, we’ll reject the null hypothesis.

The job of p-value therefore here is to answer this question:

**If I’m living in a world where the pizza delivery time is 30 minutes or less (null hypothesis is true), how surprising is my evidence in real life?**

P-value answers this question with a number — **probability**.

**The lower the p-value, the more surprising the evidence is, the more ridiculous our null hypothesis looks.**

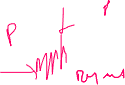
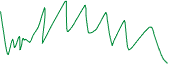
And what do we do when we feel ridiculous with our null hypothesis?

**We reject that** and choose our **alternative hypothesis instead**.

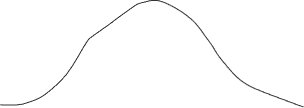
If the p-value is lower than a predetermined **significance level** (people call it *alpha*, I call it *the threshold of being ridiculous —* don’t ask my why, I just find it easier for me to understand), then we reject the null hypothesis.



Now we understand what p-value means. Let’s apply that in our case.



1. **Normal Distribution**







1. **What is P-value?**
2. **Statistical Significance**

Now we understand what p-value means. Let’s apply that in our case.

## 🍕 P-value in pizza delivery times 🍕

Now that we’ve collected some sampled delivery times, we perform the calculation and find that the **mean delivery time is longer by 10 minutes with a p-value of 0.03**.

What this means is that **in a world** where the pizza delivery time is 30 minutes or less (**null hypothesis is true**), there’s a **3% chance** we would see the mean delivery time is **at least 10 minutes longer due to random noise**.

The lower the p-value, the more meaningful the result because it is less likely to be caused by noise.

There’s a common misinterpretation of p-value for most people in our case:

The p-value 0.03 means that there’s 3% (probability in percentage) that the result is due to chance — **which is not true**.

People often want to have a definite answer (including me), and this is how I got myself confused for a long time to interpret p-values.

A p-value doesn’t \*prove\* anything. It’s simply a way to use surprise as a basis for making a reasonable decision.

**4. Statistical Significance**

Finally, this is the final stage where we put everything together and test if the result is ***statistically significant.***

Having just the p-value is not enough, we need to set a threshold (aka **significance level — alpha**). The alpha should always be set before an experiment to avoid bias. If the observed p-value is lower than alpha, then we conclude that the result is ***statistically significant.***

The rule of thumb is to set alpha to be either 0.05 or 0.01 (again, the value depends on your problems at hand).

As mentioned before, assume that we set the alpha to be 0.05 before we began the experiment, the result obtained is statistically significant since the p-value of 0.03 is lower than the alpha.

**For reference purposes, below are the** [**basic steps for the whole experiment**](http://G)**:**

1. State the null hypothesis
2. State the alternative hypothesis
3. Determine the value of alpha to be used
4. Find the Z-score associated with your alpha level
5. Find the test statistic using this formula
6. If the value of test statistic is less than the Z-score of alpha level (or p-value is less than alpha value), reject the null hypothesis. Otherwise, don’t reject the null hypothesis.

Text

Description automatically generated with low confidence

Formula to calculate test statistic for Step 5

If you want to know more about statistical significance, feel free to check out this article — [Statistical Significance Explained](https://towardsdatascience.com/statistical-significance-hypothesis-testing-the-normal-curve-and-p-values-93274fa32687) written by

[Will Koehrsen](https://medium.com/u/e2f299e30cb9?source=post_page-----f40a746cfc8--------------------------------)

<https://towardsdatascience.com/p-values-explained-by-data-scientist-f40a746cfc8>