

What is Bonferroni Correction?

Authored by
stats writer

December 9, 2025

RECOMMENDED CITATION

stats writer (2025). *What is Bonferroni Correction?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=106803>

The **Bonferroni Correction** is a fundamental technique in statistics designed to manage the complexities that arise when researchers perform multiple simultaneous hypothesis tests. It functions primarily by adjusting the required significance level, often denoted as the alpha level (α), for each individual test. This adjustment is crucial for mitigating the inflation of the global error rate. Specifically, it drastically reduces the probability of committing a **Type I error**, which occurs when one falsely rejects the **null hypothesis** even though it is true.

In essence, the method controls the family-wise error rate (FWER) across a set of comparisons. The core mechanism is straightforward: the original chosen alpha (typically 0.05) is divided by the total number of comparisons (n) being performed. This calculation results in a new, much smaller and more stringent threshold for determining statistical significance, demanding stronger evidence (a lower **p-value**) before the researcher can reject the null hypothesis.

The Problem of Multiple Comparisons

When statistical analysis involves only a single test, the risk of a **Type I error** is directly controlled by the chosen **alpha level**. For instance, if $\alpha = 0.05$, there is a 5% chance of incorrectly concluding that an effect exists when it does not. However, in modern scientific research, particularly in fields like genomics, pharmacology, and psychology, it is common practice to perform dozens or even hundreds of simultaneous statistical comparisons. This scenario is known as the multiple comparison problem, and it fundamentally alters the probability landscape.

If we maintain the standard $\alpha = 0.05$ for every individual test in a family of tests, the probability that at least one of these tests yields a significant result purely by chance (a false positive) increases dramatically. If a researcher conducts 20 independent tests, each with an individual 5% error rate, the probability of observing at least one spurious significant result can climb far above 5%. Without correction, this inflation of the error rate can lead to false discoveries, wasted resources, and non-replicable scientific findings, undermining the integrity of the research.

The **Bonferroni Correction** provides a highly conservative and reliable solution to this pervasive issue. It directly addresses the need to keep the overall probability of making any false discovery within acceptable limits, often returning the effective error rate of the entire "family" of tests back toward the original specified alpha (e.g., 0.05).

Understanding the Family-Wise Error Rate (FWER)

To appreciate why correction methods like Bonferroni are necessary, one must grasp the concept of the family-wise error rate (FWER). The FWER is defined as the probability of making at least one **Type I error** (falsely rejecting a true **null hypothesis**) within an entire set or "family" of hypothesis tests. If the individual tests are independent, the relationship between the individual

error rate (α) and the FWER is easy to visualize, although the actual calculation becomes complex when tests are dependent.

Consider a family of n tests where the **null hypothesis** is true for all of them. The probability of correctly accepting the null hypothesis in a single test is $(1 - \alpha)$. The probability of correctly accepting the null hypothesis in all n tests, assuming independence, is $(1 - \alpha)^n$. Therefore, the FWER, which is the probability of rejecting at least one true null hypothesis, is calculated as $1 - (1 - \alpha)^n$. Even for a small number of tests, say $n=10$, the FWER jumps to approximately 0.40, meaning there is a 40% chance of a false positive lurking among the results if no correction is applied.

The goal of the **Bonferroni Correction** is to ensure that the FWER remains controlled at the desired level, typically 0.05. By drastically lowering the individual test threshold, the correction ensures that the accumulated probability of error across the entire study remains manageable and trustworthy. This protection is achieved by using the simple but powerful formula derived from the union bound inequality.

The Bonferroni Formula and Calculation

The mathematical basis for the **Bonferroni Correction** is rooted in the union bound, which states that the probability of the union of events is less than or equal to the sum of their individual probabilities. In simpler terms, to ensure the FWER ($P(\text{at least one Type I error})$) is less than or equal to the desired overall **alpha level** (α_{overall}), we must set the individual alpha level (α_{adjusted}) according to the number of tests (n).

The formula for the adjusted **alpha level** (α_{adjusted}) is elegantly simple:

$$\alpha_{\text{adjusted}} = \alpha/n$$

Where:

α : Represents the initial significance level desired for the entire family of tests (e.g., 0.05).

n : Represents the total number of comparisons or statistical tests being performed simultaneously.

For example, if a researcher plans to conduct four independent tests ($n=4$) and desires an overall FWER of 0.05, the adjusted alpha for each individual test would be $0.05 / 4 = 0.0125$. This means that for any of the four comparisons to be deemed statistically significant, its corresponding **p-value** must be less than 0.0125, which is a much stricter requirement than the standard 0.05 threshold.

Step-by-Step Implementation of the Correction

Applying the **Bonferroni Correction** involves a clear, sequential process that researchers must

follow after designing their experiment but before interpreting their results.

Determine the Family of Tests: First, the researcher must clearly define what constitutes the "family" of comparisons. These are the tests whose results are interpreted together and which are subject to the same overall FWER control.

Specify the Overall Alpha (α): Select the desired **alpha level** for the family-wise error rate (usually 0.05 or 0.01).

Count Comparisons (n): Determine the exact number of statistical tests (n) within that family. For pairwise comparisons among k groups, n is often calculated as $k(k-1)/2$.

Calculate Adjusted Alpha (α_{adjusted}): Use the Bonferroni formula (α/n) to find the new, lower threshold.

Execute and Compare: Perform all n statistical tests and calculate their respective **p-values**. Compare each test's **p-value** against the calculated α_{adjusted} . Only tests where $P < \alpha_{\text{adjusted}}$ are considered statistically significant and lead to the rejection of the **null hypothesis**.

This systematic approach ensures that the researcher maintains control over false discoveries, even when the analysis involves a large number of concurrent hypotheses. The stringent nature of the correction guarantees a high degree of confidence in any reported significant finding.

Using the Bonferroni Calculator

To assist in determining the necessary threshold quickly, the following tool applies the **Bonferroni Correction** formula. By inputting the original desired significance level (α) and the total number of comparisons (n), the calculator provides the required adjusted **alpha level** (α_{adjusted}) needed to maintain the FWER control.

```
@import url('https://fonts.googleapis.com/css?family=Droid+Serif|Raleway');
```

```
.axis--y .domain {  
display: none;  
}
```

```
h1 {  
color: black;  
text-align: center;  
margin-top: 15px;  
margin-bottom: 0px;  
font-family: 'Raleway', sans-serif;  
}
```

```
h2 {
```

```
color: black;
font-size: 20px;
text-align: center;
margin-bottom: 15px;
margin-top: 15px;
font-family: 'Raleway', sans-serif;
}
```

```
p {
color: black;
text-align: center;
margin-bottom: 15px;
margin-top: 15px;
font-family: 'Raleway', sans-serif;
}
```

```
#words_intro {
color: black;
font-family: Raleway;
max-width: 550px;
margin: 25px auto;
line-height: 1.75;
}
```

```
#words_intro_center {
text-align: center;
color: black;
font-family: Raleway;
max-width: 550px;
margin: 25px auto;
line-height: 1.75;
}
```

```
#words_outro {
color: black;
font-family: Raleway;
max-width: 550px;
margin: 25px auto;
line-height: 1.75;
}
```

```
#words {
color: black;
font-family: Raleway;
max-width: 550px;
margin: 25px auto;
line-height: 1.75;
padding-left: 100px;
}

#calcTitle {
text-align: center;
font-size: 20px;
margin-bottom: 0px;
font-family: 'Raleway', serif;
}

#hr_top {
width: 30%;
margin-bottom: 0px;
margin-top: 10px;
border: none;
height: 2px;
color: black;
background-color: black;
}

#hr_bottom {
width: 30%;
margin-top: 15px;
border: none;
height: 2px;
color: black;
background-color: black;
}

.input_label_calc {
display: inline-block;
vertical-align: baseline;
width: 350px;
}
```

```
#button_calc {
border: 1px solid;
border-radius: 10px;
margin-top: 20px;
padding: 10px 10px;
cursor: pointer;
outline: none;
background-color: white;
color: black;
font-family: 'Work Sans', sans-serif;
border: 1px solid grey;
/* Green */
}

#button_calc:hover {
background-color: #f6f6f6;
border: 1px solid black;
}

.label_radio {
text-align: center;
}
```

When conducting a single statistical test, researchers usually compare the **p-value** to a predetermined alpha level (α), such as 0.05.

If the calculated **p-value** is less than 0.05, the researcher rejects the **null hypothesis**, concluding a statistically significant difference exists between the group means.

However, when running multiple comparisons simultaneously, the likelihood of incorrectly rejecting a true null hypothesis (committing a **Type I error**) increases exponentially.

To control this cumulative error, you must perform a **Bonferroni Correction**, adjusting the alpha level according to the following formula:

$$\alpha_{\text{adjusted}} = \alpha/n$$

where:

α : Represents the initial, unadjusted alpha level (e.g., 0.05)

n: Represents the total number of comparisons or individual tests being evaluated

Subsequently, you only reject the **null hypothesis** for an individual test if its p-value is less than the calculated α_{adjusted} .

To utilize the **Bonferroni Correction** and swiftly determine the adjusted α level, input your values

below and click the "Calculate" button.

Original Alpha Level (e.g., 0.05)

Total Number of Comparisons (n)

Calculated Adjusted α : 0.01250

Interpretation Example: If you plan to conduct 4 comparisons, you must only reject the null hypothesis of each comparison if it has a p-value less than 0.01250 to maintain an overall family-wise error rate of 0.05.

```
function calc() {  
  //get input values  
  var a = document.getElementById('a').value*1;  
  var n = document.getElementById('n').value*1;  
  
  //find number of bins  
  var adj = a/n;  
  
  //output  
  document.getElementById('adj').innerHTML = adj.toFixed(5);  
  document.getElementById('n_out').innerHTML = n;  
  document.getElementById('adj_out').innerHTML = adj.toFixed(5);  
}
```

Advantages and Criticisms of the Bonferroni Method

The **Bonferroni Correction** is widely used due to its simplicity and its powerful ability to strictly control the family-wise error rate (FWER). Its major advantage lies in its mathematical guarantee that the probability of making one or more false positive conclusions across the entire set of tests will not exceed the chosen overall **alpha level** (α). It is also highly versatile, as it does not require assumptions about the independence or dependence structure among the various individual tests.

However, the primary criticism leveled against the Bonferroni method is its extreme conservativeness. By dividing the alpha level by the number of tests (n), especially when n is large, the resulting α_{adjusted} becomes so small that it significantly increases the risk of committing a **Type II error** (falsely failing to reject a false **null hypothesis**). This means the researcher is highly likely to miss genuine effects, leading to a loss of statistical power.

This trade-off between strict Type I error control and loss of power is a critical consideration. Researchers must weigh the cost of a false positive against the cost of a false negative. If

controlling the FWER is paramount (e.g., in clinical trials where a false positive could lead to harmful treatment), Bonferroni remains an excellent choice. If the study is exploratory and minimizing Type II errors is more important, alternative methods might be preferred.

Alternatives to the Bonferroni Correction

Because of Bonferroni's inherent power reduction, several less conservative adjustments have been developed to handle the multiple comparison problem while maintaining reasonable statistical rigor. These alternatives often provide a better balance between Type I and Type II error control.

One popular alternative is the **Holm-Bonferroni Method** (or Holm's Sequential Bonferroni Procedure). This step-down procedure is uniformly more powerful than the standard Bonferroni correction. It involves ordering the individual **p-values** from smallest to largest and then testing them sequentially against adjusted alpha levels that increase as the rank of the p-value increases. This sequential approach retains the strict control over FWER but provides a substantial increase in statistical power compared to the standard, fixed α adjusted of the original method.

Another important approach is controlling the **False Discovery Rate (FDR)**, most famously exemplified by the Benjamini-Hochberg procedure. Instead of focusing on the FWER (the probability of making *any* false discovery), FDR methods control the expected proportion of false discoveries among all rejected **null hypotheses**. FDR methods are significantly less conservative than Bonferroni and are frequently used in high-throughput data analysis, such as gene expression studies, where thousands of comparisons are made and some false positives can be tolerated if the overall proportion is low.

Choosing the Right Correction Method

The choice of correction method should always be guided by the nature of the research questions and the associated risk profile. When performing a small number of planned comparisons (e.g., pairwise tests following an ANOVA), the **Bonferroni Correction** is simple to calculate and provides the strongest guarantee against false positives, making it a reliable choice for dedicated hypothesis testing.

If the number of comparisons (n) is large or if the individual tests are highly dependent (i.e., not independent), the standard Bonferroni method may become prohibitively conservative, forcing the researcher toward the Holm-Bonferroni method or other FWER-controlling procedures that retain more power. For large-scale exploratory studies where the primary goal is hypothesis generation and strict FWER control is too restrictive, FDR-controlling methods offer a viable, power-sensitive alternative.

Ultimately, understanding the potential inflation of the error rate in multiple testing scenarios is

paramount. The **Bonferroni Correction** serves as the foundational approach, highlighting the necessary rigor required to maintain scientific integrity when dealing with multiple inferences. It remains a crucial tool in the statistical toolkit for ensuring that reported findings are genuinely significant and replicable.

ARABPSYCHOLOGY.COM