People's Democratic Republic of Algeria
University Med Khider of Biskra
Faculty of SNVSTU

Non-parametric tests

Dr. Ben Gherbal Hanane

Biostatistics Level: M1 Biology
Email: hanane.bengherbal@univ-biskra.dz

---

1. Introduction

In biological research, statistical analysis plays a central role in comparing groups, evaluating treatments, and testing scientific hypotheses. Classical (parametric) methods such as the Student t-test or ANOVA rely on specific assumptions, particularly normality of the data and homogeneity of variances.

However, in many biological studies, these assumptions are not satisfied. Data may:

- arise from small samples,
- deviate from normality,
- contain extreme values (outliers),
- or be measured on an ordinal or ranked scale.

In such situations, **non-parametric tests** provide appropriate and robust alternatives. These methods do not require assumptions about the underlying probability distribution of the population and are often based on ranks rather than raw numerical values.

Non-parametric procedures are widely used in biology and medical sciences to compare independent or related samples, to test differences between several groups, and to study associations between variables when parametric conditions are not met.

In this chapter, we introduce the main non-parametric tests commonly used in biological applications.

2. When to Use Non-Parametric Tests

Non-parametric tests are used when:

- The normality assumption is violated.
- The sample size is small.
- The data are ordinal (ranked data).
- There are strong outliers.

These tests are based on **ranks** rather than raw values.

3. Mann–Whitney U Test

In many biological studies, researchers need to compare measurements obtained from two independent samples. When the assumptions required for parametric tests (such as normal distribution) are not satisfied, a non-parametric alternative must be used.

The **Mann–Whitney U test** is a non-parametric method designed to compare two independent populations or two independent measurement methods. Instead of comparing means, this test is based on the ranking of all observed values from both groups combined.

All observations from the two samples are first pooled together and ordered from the smallest to the largest. The ranks corresponding to the individuals of one group are then identified, and the sum of these ranks is calculated. The test statistic is derived from these rank sums and allows us to determine whether a statistically significant difference exists between the two populations.

The Mann–Whitney test is particularly appropriate when:

- Sample sizes are small (less than 30),
- Data are ordinal,
- The distribution of the data is not normal,
- Outliers may influence parametric tests.

This method is widely used in biological and medical research to compare independent groups under non-normal conditions.

3.1 Objective

The Mann–Whitney U test is used to compare **two independent groups** when the data are ordinal or not normally distributed.

Test Hypotheses

Null hypothesis (H_0): The two distributions of the variable in populations (A) and (B) are identical.

Alternative hypothesis (H_1): The two distributions of the variable in populations (A) and (B) are different (two-tailed test).

Alternatively (one-tailed test): the values in population (A) are lower (or higher) than those in population (B).

b. Decision Variable

For practical computational reasons, we assume (possibly after exchanging the roles of the two populations) that the sample size n from population (A) is smaller than the sample size m from population (B).

Let $(x_1, x_2, \ldots, x_n)$ and $(y_1, y_2, \ldots, y_m)$ denote the data from the two samples, arranged in increasing order. We then order all $n + m$ observations together in ascending order.

Example 1.
A biologist wants to compare the effect of two biofertilizers (A and B) on the growth of bean plants.
Independent simple random samples of plants were treated with each fertilizer. After three weeks, the height of each plant (in cm) was measured.
The observed data are:

Fertilizer A: 4 9 11

Fertilizer B: 6 8 10 12

One measurement was lost during the experiment.
Assume that the samples are independent and that normality cannot be assumed.

The decision variable is the random variable denoted by U_x, whose values are calculated as follows:

- For each value x_i, we count the number of values y_j that are strictly less than x_i.
- For each value x_i, we count the number of values y_j that are equal to x_i and multiply this number by $\frac{1}{2}$.
- The statistic U_x is the sum of the two quantities obtained above.

Similarly, we define the random variable U_y.

Example 1:
Here $n = 3$ and $m = 4$.
Let us rank all observed data (highlighting the values from the first sample to identify them). We obtain the following ordered list:

$$4 \ 6 \ 8 \ 9 \ 10 \ 11 \ 12.$$

Let (x_1, x_2, x_3) denote the observations from Fertilizer A. We note that:

- for $x_1 = 4$, no value from Fertilizer B is strictly smaller;
- for $x_2 = 9$, two values from Fertilizer B are strictly smaller: 6 and 8;
- for $x_3 = 11$, three values from Fertilizer B are strictly smaller: 6, 8 and 10.

Therefore,

$$u_x = 0 + 2 + 3 = 5.$$

By doing the same computation for u_y, we obtain:

$$u_y = 1 + 1 + 2 + 3 = 7.$$

- **Properties of U_x and U_y:**

$$U_x + U_y = nm.$$

The statistics U_x and U_y can be expressed in Wilcoxon rank sums W_x and W_y as follows:

$$U_x = W_x - \frac{n(n+1)}{2}, \qquad U_y = W_y - \frac{m(m+1)}{2},$$

where W_x and W_y are the sums of ranks for samples (A) and (B), respectively.

Example 1: We had the following ordering:

$$4 \quad 6 \quad 8 \quad 9 \quad 10 \quad 11 \quad 12$$

with corresponding ranks:

$$1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6 \quad 7.$$

For sample A, the ranks are 1, 4, and 6, hence

$$W_x = 1 + 4 + 6 = 11.$$

The relation is verified:

$$11 - \frac{3 \times 4}{2} = 5 = u_x.$$

Similarly, for sample B:

$$W_y = 2 + 3 + 5 + 7 = 17,$$

and

$$17 - \frac{4 \times 5}{2} = 7 = u_y.$$

Decision Rule

Case of a Two-Tailed Test

The decision variable of the test is defined as:

$$U = \min(U_x, U_y).$$

Rejection Region and Critical Value

Under the null hypothesis H_0, there exists a unique real number a, called the *critical value* of the test, such that $\mathbb{P}(U \leq a) = \alpha$.

The value a is obtained from statistical tables corresponding to the chosen significance level α.

Rejection region of H_0:

$$[0,\, a]$$

Non-rejection region of H_0:

$$(a,\, \frac{nm}{2}]$$

Decision rule:

At significance level α,

$$\begin{cases} \text{If } u_{\text{obs}} \leq a, & \text{reject } H_0, \\ \text{otherwise}, & \text{do not reject } H_0. \end{cases}$$

Here, a denotes the critical value of the test corresponding to the chosen significance level α.

Example

A research team studied the effect of two organic soil treatments (A) and (B) on plant dry biomass (g). The data are:

$$\text{Treatment A: } 2,\ 6,\ 9,\ 10,\ 12,\ 15,\ 18 \qquad (n = 7)$$

$$\text{Treatment B: } 4,\ 7,\ 11,\ 14,\ 16,\ 17,\ 19,\ 20 \qquad (m = 8)$$

Two-tailed test:

$$H_0: \text{ the two distributions are identical} \qquad \text{vs} \qquad H_1: \text{ the two distributions are different.}$$

Step 1: Order all the observations
We order the $n + m = 15$ values in increasing order (values from A are underlined):

$$\underline{2},\ 4,\ \underline{6},\ 7,\ \underline{9},\ \underline{10},\ 11,\ \underline{12},\ 14,\ \underline{15},\ 16,\ 17,\ \underline{18},\ 19,\ 20.$$

Step 2: Compute u_x
Let $(x_1, \dots, x_7)$ be the values from treatment A:

$$(x_1, \dots, x_7) = (2, 6, 9, 10, 12, 15, 18).$$

For each x_i, we count how many values from B are strictly smaller than x_i:

- $x_1 = 2$: none in B is $< 2 \Rightarrow 0$.
- $x_2 = 6$: only $4 < 6 \Rightarrow 1$.
- $x_3 = 9$: $4, 7 < 9 \Rightarrow 2$.
- $x_4 = 10$: $4, 7 < 10 \Rightarrow 2$.
- $x_5 = 12$: $4, 7, 11 < 12 \Rightarrow 3$.
- $x_6 = 15$: $4, 7, 11, 14 < 15 \Rightarrow 4$.
- $x_7 = 18$: $4, 7, 11, 14, 16, 17 < 18 \Rightarrow 6$.

Hence,

$$u_x = 0 + 1 + 2 + 2 + 3 + 4 + 6 = 18.$$

Step 3: Compute u_y
Similarly, for each value y_j from B we count how many values from A are strictly smaller:

- $y_1 = 4$: only $2 < 4 \Rightarrow 1$.
- $y_2 = 7$: $2, 6 < 7 \Rightarrow 2$.
- $y_3 = 11$: $2, 6, 9, 10 < 11 \Rightarrow 4$.
- $y_4 = 14$: $2, 6, 9, 10, 12 < 14 \Rightarrow 5$.
- $y_5 = 16$: $2, 6, 9, 10, 12, 15 < 16 \Rightarrow 6$.
- $y_6 = 17$: $2, 6, 9, 10, 12, 15 < 17 \Rightarrow 6$.
- $y_7 = 19$: all A values are $< 19 \Rightarrow 7$.
- $y_8 = 20$: all A values are $< 20 \Rightarrow 7$.

Thus,

$$u_y = 1 + 2 + 4 + 5 + 6 + 6 + 7 + 7 = 38.$$

Check: Fundamental relation

$$u_x + u_y = 18 + 38 = 56 = nm = 7 \times 8.$$

Step 4: Verify using Wilcoxon rank sums W_x and W_y
The ranks of the ordered values (from 1 to 15) are:

$$2(1),\ 4(2),\ 6(3),\ 7(4),\ 9(5),\ 10(6),\ 11(7),\ 12(8),\ 14(9),\ 15(10),\ 16(11),\ 17(12),\ 18(13),\ 19(14),\ 20(15).$$

Ranks for treatment A: $1, 3, 5, 6, 8, 10, 13$, hence

$$W_x = 1 + 3 + 5 + 6 + 8 + 10 + 13 = 46.$$

Then

$$U_x = W_x - \frac{n(n+1)}{2} = 46 - \frac{7 \times 8}{2} = 46 - 28 = 18,$$

which matches u_x.

Ranks for treatment B: $2, 4, 7, 9, 11, 12, 14, 15$, hence

$$W_y = 2 + 4 + 7 + 9 + 11 + 12 + 14 + 15 = 74,$$

and

$$U_y = W_y - \frac{m(m+1)}{2} = 74 - \frac{8 \times 9}{2} = 74 - 36 = 38,$$

which matches u_y.

Step 5: Decision (two-tailed, $\alpha = 0.05$)

The decision variable is

$$U = \min(U_x, U_y) = \min(18, 38) = 18.$$

From Mann–Whitney tables for $n = 7$ and $m = 8$ at $\alpha = 0.05$ (two-tailed), the critical value is $a = 10$.

$$u_{\text{obs}} = 18 > 10.$$

Conclusion: we **do not reject** H_0 at the 5% level. There is insufficient evidence to conclude that the two treatments lead to different biomass levels.

Case of a One-Tailed Test

Let us consider, for example, the alternative hypothesis

$$H_1 : \text{“The values of the variable in population (A) are smaller than those in population (B).”}$$

The decision variable of the test is U_x. We fix a significance level α.

Rejection region of H_0:

If H_1 is true, the values of U_x tend to be smaller than those of U_y and therefore close to 0. Hence, the rejection region of H_0 is sought in the form $[0, a]$.

Under H_0, there exists a unique real number a, called the *critical value* of the test, such that

$$\mathbb{P}(U_x \leq a) = \alpha.$$

We use the tables to determine an approximate value of a.

Thus, the rejection region of H_0 is the interval $[0, a]$ and the non-rejection region is $(a,\ nm]$.

Decision rule:

At significance level α,

$$\begin{cases} \text{If } u_{\text{obs}} \leq a, & \text{reject} H_0, \\ \text{otherwise,} & \text{do not reject } H_0. \end{cases}$$

Remark: for

$$H_1 : \text{“The values of the variable in population (A) are greater than those in population (B).”}$$

if H_1 is true, the values of U_x tend to be greater than those of U_y and therefore close to nm.

Under H_0, there exists a unique real number a such that

$$\mathbb{P}(U_x \geq a) = \alpha.$$

Moreover,

$$\mathbb{P}(U_x \geq a) = \mathbb{P}(nm - U_x \geq a) = \mathbb{P}(U_x \leq nm - a).$$

Hence, the rejection region of H_0 is $[a, nm]$ and the non-rejection region is $[0, a)$.
Example We test the null hypothesis

$$H_0 : \text{“The scores of product (A) are at the same level as those of product (B).”}$$

against the alternative hypothesis

$$H_1 : \text{“The scores assigned to product (A) are lower than those assigned to product (B).”}$$

Decision variable: U_x.
We choose a significance level $\alpha = 0.01$.

Determination of the rejection region of H_0:
From the Mann–Whitney tables for $m = 8$ and $n = 7$, under H_0 we read:

$$\mathbb{P}(U_x \leq 7) = 0.007, \qquad \mathbb{P}(U_x \leq 8) = 0.010.$$

From the Mann–Whitney tables for $m = 8$ and $n = 7$, we read that the critical value is $a = 7$.
The use of either table leads to the same decision rule.

Decision rule:
If the observed value u_{obs} of U_x is less than or equal to 7, we reject H_0. Otherwise, we do not reject H_0.

Conclusion:
Since $u_{\text{obs}} = 18$, we do not reject H_0 at the 1% significance level.

Example 2 (when some values are repeated)
A biologist studies the effect of two nutrient solutions (A) and (B) on the number of leaves produced by young tomato plants after four weeks.
Independent random samples of plants were treated with each solution.
The observed data (number of leaves per plant) are:

$$\text{Treatment A: } 8,\ 10,\ 10,\ 13,\ 15 \qquad (n = 5)$$

$$\text{Treatment B: } 7,\ 9,\ 10,\ 12,\ 12,\ 16 \qquad (m = 6)$$

Note that some values are repeated (for example, the value 10 appears in both samples and the value 12 appears twice).
Assume that the samples are independent and that the normality assumption cannot be guaranteed. We apply the Mann–Whitney U test.

Step 1: Order all observations

We combine the $n + m = 11$ observations and arrange them in increasing order:

$$7,\ 8,\ 9,\ 10,\ 10,\ 10,\ 12,\ 12,\ 13,\ 15,\ 16$$

When equal values occur (ties), each value receives the **average of the corresponding ranks**.
Step 2: Assign ranks

Value	Rank
7	1
8	2
9	3
10	5
10	5
10	5
12	7.5
12	7.5
13	9
15	10
16	11

(The three values equal to 10 occupy ranks $4, 5, 6$, so each receives the average rank 5.) **Step 3: Compute the rank sum for sample A**

Values of sample A: $8, 10, 10, 13, 15$
Corresponding ranks:

$$2,\ 5,\ 5,\ 9,\ 10$$

Thus

$$W_x = 2 + 5 + 5 + 9 + 10 = 31$$

Then

$$U_x = W_x - \frac{n(n+1)}{2}$$

$$U_x = 31 - \frac{5 \times 6}{2} = 31 - 15 = 16$$

Finally

$$U_y = nm - U_x = 5 \times 6 - 16 = 14$$

4. Paired Wilcoxon Test (Signed-Rank Test)

The paired Wilcoxon test is a **nonparametric** test used to compare two sets of dependent (paired) measurements. It is an alternative to the **paired t-test** when the assumption of normality of the differences is not satisfied.

Hypotheses

$$H_0 : \text{the median of the differences is zero}$$

$$H_1 : \text{the median of the differences is different from zero}$$

Conditions of application

- The observations must be paired.
- The data must be ordinal or quantitative.
- The differences are assumed to be symmetric around their median.

- Zero differences are excluded from the calculation.

Methodology

1. Compute the differences $d_i = Y_i - X_i$ for each pair.
2. Remove the zero differences.
3. Rank the absolute values $|d_i|$ in increasing order.
4. Assign a rank to each value.
5. Attach the corresponding sign to each rank.
6. Compute the sum of positive ranks W^+ and negative ranks W^-.
7. The test statistic is
$$W = \min(W^+, W^-).$$

Example 1

A biologist studies the effect of a fertilizer on the growth of tomato plants. The height of the plants (in cm) is measured before and after the application of the fertilizer.

Plant	Before	After
1	12	15
2	10	13
3	14	14
4	9	11
5	13	17
6	11	12

Step 1: Compute the differences

$$d_i = \text{After} - \text{Before}$$

Plant	d_i
1	3
2	3
3	0
4	2
5	4
6	1

The zero difference (plant 3) is removed.

Step 2: Ranking the absolute values

d_i	$\|d_i\|$	Rank
1	1	1
2	2	2
3	3	3.5
3	3	3.5
4	4	5

Step 3: Sum of the ranks

All signs are positive:

$$W^+ = 1 + 2 + 3.5 + 3.5 + 5 = 15$$

$$W^- = 0$$

Thus,

$$W = \min(W^+, W^-) = 0$$

Decision rule

For a significance level $\alpha = 0.05$ and a sample size $n = 5$ (after removing the zero difference), the critical value from the Wilcoxon signed-rank table is

$$W_{crit} = 0$$

- If $W \leq W_{crit}$, reject H_0.
- If $W > W_{crit}$, do not reject H_0.

Decision for the example

Since

$$W = 0 \leq W_{crit} = 0$$

we reject the null hypothesis H_0.

Conclusion

There is a statistically significant difference between the measurements before and after the treatment. The fertilizer significantly increases the growth of the tomato plants.

Example 2

You have measured the reaction time of a small group of people once in the morning and once in the evening and you want to know if there is a difference.

Case	Morning	Evening	Difference (Evening-Morning)
1	34	45	11
2	33	36	3
3	41	35	-6
4	39	43	4
5	44	42	-2
6	37	42	5
7	39	46	7

Step 1: Hypotheses

$$H_0 : \text{The median difference between evening and morning is 0}$$

$$H_1 : \text{The median difference is different from 0}$$

Step 2: Differences

$$d_i = \text{Evening} - \text{Morning}$$

Case	d_i
1	11
2	3
3	-6
4	4
5	-2
6	5
7	7

There are no zero differences, so the sample size is

$$n = 7$$

Step 3: Ranking the absolute values

d_i	$\lvert d_i \rvert$	Rank	Sign
11	11	7	+
3	3	2	+
-6	6	5	−
4	4	3	+
-2	2	1	−
5	5	4	+
7	7	6	+

Step 4: Sum of the ranks
Positive ranks:

$$W^+ = 7 + 2 + 3 + 4 + 6 = 22$$

Negative ranks:

$$W^- = 5 + 1 = 6$$

Test statistic:

$$W = \min(W^+, W^-) = 6$$

Step 5: Decision rule
For a significance level

$$\alpha = 0.05$$

and a sample size

$$n = 7$$

the critical value from the Wilcoxon table is

$$W_{crit} = 2$$

Decision rule:

- If $W \le W_{crit}$, reject H_0.
- If $W > W_{crit}$, do not reject H_0.

Step 6: Decision

$$W = 6 > 2$$

Therefore we do not reject H_0.

Conclusion
There is no statistically significant difference between the reaction times measured in the morning and in the evening at the 5% significance level.

5. Kruskal–Wallis Test

The Kruskal–Wallis test is a nonparametric alternative to the one-way ANOVA. It is used to compare more than two independent groups when:

- The data are ordinal or not normally distributed,
- The assumption of homogeneity of variances is not satisfied.

Hypotheses
Let k be the number of groups.

- H_0: The k groups come from the same distribution.
- H_1: At least one group differs.

Test Statistic
We combine all observations and rank them from smallest to largest.
Let:

- N : total number of observations,
- n_i : size of group i,
- R_i : sum of ranks of group i.

The Kruskal–Wallis statistic is:

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(N+1)$$

Under H_0, H follows approximately a $\chi^2(k-1)$ distribution.
Example
We study the effect of three treatments on enzyme concentration (U/L).

Group 1 (Drug A)	Group 2 (Drug B)	Group 3 (Drug C)
27	20	34
2	8	31
4	14	3
18	36	23
7	21	30
9	22	6

Step 1: Ranking the Data

We combine all values and rank them:

Value	Rank
2	1
3	2
4	3
6	4
7	5
8	6
9	7
14	8
18	9
20	10
21	11
22	12
23	13
27	14
30	15
31	16
34	17
36	18

Step 2: Sum of Ranks

- Group 1: $R_1 = 14 + 1 + 3 + 9 + 5 + 7 = 39$
- Group 2: $R_2 = 10 + 6 + 8 + 18 + 11 + 12 = 65$
- Group 3: $R_3 = 17 + 16 + 2 + 13 + 15 + 4 = 67$

Step 3: Compute H

Here:

$$N = 18, \quad n_1 = n_2 = n_3 = 6$$

$$H = \frac{12}{18(19)} \left(\frac{39^2}{6} + \frac{65^2}{6} + \frac{67^2}{6} \right) - 3(19) = 2.85$$

Decision Rule
Degrees of freedom:

$$df = k - 1 = 2$$

Critical value at $\alpha = 0.05$:

$$\chi^2_{0.05,2} \approx 5.99$$

Conclusion

$$H = 2.85 < 5.99$$

We **fail to reject** H_0.

Interpretation:
There is no statistically significant difference between the three treatments in terms of enzyme concentration at the 5% level.

The Friedman Test

The Friedman test is a nonparametric alternative to the repeated measures ANOVA. It is used when:

- the same subjects are measured under several conditions or at several times,
- the observations are related (paired or repeated measurements),
- the data are ordinal or the normality assumption is not satisfied.

Biological Context

A biologist wants to study the evolution of the concentration of a protein in blood after applying an experimental treatment. The same 6 laboratory mice are followed for 3 consecutive weeks.

The measured concentrations are given below.

Week 1	**Week 2**	**Week 3**
27	20	34
2	8	31
4	14	3
18	36	23
7	21	30
9	22	6

We want to know whether there is a significant difference between Weeks 1, 2, and 3 at the significance level

$$\alpha = 0.05.$$

Hypotheses

Let the three repeated measurements correspond to the three weeks.

- Null hypothesis:
$$H_0 : \text{there is no difference between the 3 weeks.}$$
- Alternative hypothesis:
$$H_1 : \text{at least one week differs from the others.}$$

Principle of the Test

In the Friedman test, we rank the measurements *within each subject* (within each row).

For each mouse:

- rank 1 is assigned to the smallest value,
- rank 2 to the intermediate value,
- rank 3 to the largest value.

Then we sum the ranks for each week and calculate the test statistic.

Step 1: Ranking Within Each Mouse

Mouse 1

Values: 27, 20, 34

$$20 < 27 < 34$$

So the ranks are:

$$27 \to 2, \qquad 20 \to 1, \qquad 34 \to 3$$

Mouse 2

Values: 2, 8, 31

$$2 < 8 < 31$$

Ranks:

$$2 \to 1, \qquad 8 \to 2, \qquad 31 \to 3$$

Mouse 3

Values: 4, 14, 3

$$3 < 4 < 14$$

Ranks:

$$4 \to 2, \qquad 14 \to 3, \qquad 3 \to 1$$

Mouse 4

Values: 18, 36, 23

$$18 < 23 < 36$$

Ranks:

$$18 \to 1, \qquad 36 \to 3, \qquad 23 \to 2$$

Mouse 5

Values: 7, 21, 30

$$7 < 21 < 30$$

Ranks:

$$7 \to 1, \qquad 21 \to 2, \qquad 30 \to 3$$

Mouse 6

Values: 9, 22, 6

$$6 < 9 < 22$$

Ranks:

$$9 \to 2, \qquad 22 \to 3, \qquad 6 \to 1$$

Rank Table

Mouse	**Week 1**	**Rank**	**Week 2**	**Rank**	**Week 3**	**Rank**
1	27	2	20	1	34	3
2	2	1	8	2	31	3
3	4	2	14	3	3	1
4	18	1	36	3	23	2
5	7	1	21	2	30	3
6	9	2	22	3	6	1

Thus, the sums of ranks are:

$$R_1 = 2 + 1 + 2 + 1 + 1 + 2 = 9$$

$$R_2 = 1 + 2 + 3 + 3 + 2 + 3 = 14$$

$$R_3 = 3 + 3 + 1 + 2 + 3 + 1 = 13$$

Test Statistic
Let:

- $n = 6$ be the number of subjects,
- $k = 3$ be the number of repeated conditions (weeks).

The Friedman statistic is

$$Q = \frac{12}{nk(k+1)} \sum_{j=1}^{k} R_j^2 - 3n(k+1)$$

Here:

$$n = 6, \qquad k = 3, \qquad R_1 = 9, \qquad R_2 = 14, \qquad R_3 = 13$$

So:

$$Q = \frac{12}{6 \times 3 \times 4} \left(9^2 + 14^2 + 13^2\right) - 3 \times 6 \times 4$$

$$Q = \frac{12}{72}(81 + 196 + 169) - 72$$

$$Q \approx 2.33$$

Decision Rule
Under H_0, the statistic Q is approximately distributed as a chi-square with

$$k - 1 = 2$$

degrees of freedom.
At the significance level $\alpha = 0.05$:

$$\chi^2_{0.05,2} \approx 5.99$$

Comparison:

$$Q \approx 2.33 < 5.99$$

Conclusion
Since

$$Q < \chi^2_{0.05,2},$$

we fail to reject the null hypothesis H_0.

Interpretation:
At the 5% significance level, there is no statistically significant difference between Week 1, Week 2, and Week 3 for the protein concentration measured on the same 6 mice.