

# sign,rank sum, signed rank

BWB

## 目录

<b>1</b>	<b>符号检验的基本原理</b>	<b>3</b>
1.1	假设 . . . . .	3
1.2	检验思想 . . . . .	3
<b>2</b>	<b>符号检验的步骤</b>	<b>3</b>
<b>3</b>	<b>示例：新疗法对血压的影响</b>	<b>3</b>
3.1	问题描述 . . . . .	3
3.2	数据 . . . . .	4
3.3	按照 HATPC 框架进行分析 . . . . .	4
3.3.1	假设 (Hypotheses) . . . . .	4
3.3.2	假设条件 (Assumptions) . . . . .	4
3.3.3	检验统计量 (Test Statistic) . . . . .	4
3.3.4	P 值计算 (P-value) . . . . .	5
3.3.5	结论 (Conclusion) . . . . .	5
3.4	使用 R 语言进行符号检验 . . . . .	5
<b>4</b>	<b>检验背景和适用场景</b>	<b>6</b>
4.1	1. 何时使用符号秩检验 . . . . .	6
4.2	2. 为什么选择符号秩检验 . . . . .	6
<b>5</b>	<b>检验假设</b>	<b>6</b>
<b>6</b>	<b>检验原理</b>	<b>7</b>
6.1	1. 基本思想 . . . . .	7
6.2	2. 数学推导 . . . . .	7
<b>7</b>	<b>检验步骤</b>	<b>7</b>
7.1	步骤 1: 计算差值 . . . . .	7
7.2	步骤 2: 剔除零差值 . . . . .	7
7.3	步骤 3: 计算绝对差值 . . . . .	7

7.4	步骤 4: 对绝对差值进行排序并赋予秩次	8
7.5	步骤 5: 赋予符号	8
7.6	步骤 6: 计算正负秩次和	8
7.7	步骤 7: 计算检验统计量	8
7.8	步骤 8: 确定检验的显著性	8
7.8.1	计算期望值 $E(W)$ 和方差 $Var(W)$	8
7.8.2	计算标准化的 $Z$ 值	9
7.8.3	计算 $p$ 值	9
7.8.4	做出决策	9
<b>8</b>	<b>实例解析</b>	<b>9</b>
8.1	假设情境	9
8.2	数据	9
8.3	手动计算步骤	9
8.3.1	步骤 1: 计算差值	9
8.3.2	步骤 2: 剔除零差值	9
8.3.3	步骤 3: 计算绝对差值	10
8.3.4	步骤 4: 排序并赋予秩次	10
8.3.5	步骤 5: 赋予符号	10
8.3.6	步骤 6: 计算正负秩次和	10
8.3.7	步骤 7: 计算检验统计量	11
8.3.8	步骤 8: 确定检验的显著性	11
8.4	使用 R 语言进行 Wilcoxon 符号秩检验	12
8.4.1	在 R 中输入数据	12
8.4.2	使用 <code>wilcox.test</code> 进行符号秩检验	12
8.4.3	输出结果	12
8.4.4	结果解释	13
8.4.5	参数说明	13
8.4.6	单尾检验	13
<b>9</b>	<b>结论与解释</b>	<b>13</b>
9.1	统计结论	13
9.2	实际意义	13
<b>10</b>	<b>注意事项</b>	<b>13</b>
10.1	1. 数据要求	13
10.2	2. 平秩处理	14
10.3	3. 样本量	14
10.4	4. 单尾检验与双尾检验	14

# 引言

当数据不满足正态性假设，或者我们无法确定数据的分布形态时，符号检验是一种有效的非参数统计方法。本文将介绍符号检验的基本原理，并通过一个具体的示例，按照 HATPC（假设、假设条件、检验统计量、P 值、结论）框架进行分析。

## 1 符号检验的基本原理

假设我们有一组样本数据  $X_1, X_2, \dots, X_n$ ，希望检验总体中位数是否等于某个特定值  $\mu_0$ 。

### 1.1 假设

- 零假设  $H_0: \mu = \mu_0$
- 备择假设  $H_1: \mu \neq \mu_0$

### 1.2 检验思想

在  $H_0$  下，假设数据分布关于  $\mu_0$  对称。因此，差值  $D_i = X_i - \mu_0$  应该均匀地分布在零的两侧。也就是说，每个  $D_i$  为正或负的概率都是 0.5。

因此，正符号的数量  $S$  服从参数为  $n$  和  $p = 0.5$  的二项分布，即  $S \sim B(n, 0.5)$ 。

## 2 符号检验的步骤

1. 计算差值并确定符号：

$$D_i = X_i - \mu_0$$

记录每个  $D_i$  的符号，忽略  $D_i = 0$  的情况。

2. 统计正符号的数量  $S$ ，样本量为  $n$ （不包括  $D_i = 0$  的数量）。

3. 计算 P 值：

$$P\text{-value} = 2 \times P(S \leq \min(S, n - S))$$

或根据具体的备择假设（单侧或双侧）计算 P 值。

4. 作出统计决策：如果 P 值小于显著性水平  $\alpha$ ，则拒绝  $H_0$ ，认为中位数与  $\mu_0$  不等。

## 3 示例：新疗法对血压的影响

### 3.1 问题描述

研究人员想要检验一种新疗法是否能够降低患者的血压。随机选取了 12 名患者，测量了他们接受治疗前后的血压。

## 3.2 数据

表 1: 治疗前后血压数据 (mmHg)

患者编号	治疗前血压	治疗后血压	差值
1	150	145	+5
2	160	155	+5
3	148	150	-2
4	155	149	+6
5	162	158	+4
6	149	149	0
7	151	147	+4
8	157	152	+5
9	153	154	-1
10	158	153	+5
11	150	151	-1
12	152	148	+4

## 3.3 按照 HATPC 框架进行分析

### 3.3.1 假设 (Hypotheses)

- 零假设  $H_0$ : 新疗法对血压没有影响,  $p_+ = 0.5$
- 备择假设  $H_1$ : 新疗法降低了血压,  $p_+ > 0.5$

其中,  $p_+$  表示差值  $D_i$  为正的的概率。

### 3.3.2 假设条件 (Assumptions)

- 样本是独立随机抽取的。
- 差值  $D_i$  来自连续、对称分布。

### 3.3.3 检验统计量 (Test Statistic)

计算正符号的数量  $S$ 。

从数据中, 我们有:

- 正符号数量  $S = 8$
- 负符号数量 = 3
- 忽略  $D_i = 0$  的一项, 样本量  $n = 11$

### 3.3.4 P 值计算 (P-value)

在零假设  $H_0$  下,  $S \sim B(n, 0.5)$ 。

计算单侧检验的 P 值:

$$P\text{-value} = 1 - P(S \leq 7) = P(S \geq 8)$$

使用二项分布的累积分布函数:

$$P(S \leq k) = \sum_{i=0}^k \binom{n}{i} p^i (1-p)^{n-i}$$

其中  $n = 11, p = 0.5$ 。

计算  $P(S \leq 7)$ :

$$\begin{aligned} P(S \leq 7) &= \sum_{i=0}^7 \binom{11}{i} (0.5)^{11} \\ &= 0.8867 \text{ (使用二项分布计算器或统计软件得出)} \end{aligned}$$

因此,

$$P\text{-value} = 1 - P(S \leq 7) = 1 - 0.8867 = 0.1133$$

P 值为 0.1133。

### 3.3.5 结论 (Conclusion)

由于 P 值 (0.1133) 大于显著性水平  $\alpha = 0.05$ , 我们无法拒绝零假设  $H_0$ 。因此, 没有足够的证据表明新疗法显著降低了血压。

## 3.4 使用 R 语言进行符号检验

我们可以使用 R 语言的 `binom.test` 函数来进行符号检验。以下是具体的 R 代码:

Listing 1: 使用 `binom.test` 进行符号检验

```
# 正符号数量
S <- 8
# 样本量 (不包括差值为零的项)
n <- 11

# 执行单侧符号检验
result <- binom.test(x = S, n = n, p = 0.5, alternative = "greater")

# 输出结果
print(result)
```

运行上述代码，得到以下结果：

```
Exact binomial test
```

```
data: S and n
```

```
number of successes = 8, number of trials = 11, p-value = 0.1133
```

```
alternative hypothesis: true probability of success is greater than 0.5
```

P 值为 0.1133，与手工计算结果一致。

## Wilcoxon 符号秩检验 (Signed-Rank Test) 详解

### 4 检验背景和适用场景

#### 4.1 1. 何时使用符号秩检验

- **配对样本**：同一组对象在不同时间点或不同条件下的测量值。
  - 例如：评估新药对患者血压的影响，测量患者服药前后的血压值。
- **数据类型**：连续型或有序数据，不要求满足正态分布。
- **目的**：检验两个相关样本的中位数差值是否为零，即是否存在显著差异。

#### 4.2 2. 为什么选择符号秩检验

- **非参数方法**：不要求数据符合特定分布，适用于分布未知或非正态分布的数据。
- **对异常值不敏感**：相比于配对 t 检验，符号秩检验对异常值的影响较小。
- **适用于小样本量**：在样本量较小的情况下，符号秩检验仍然具有较好的统计效率。

### 5 检验假设

Wilcoxon 符号秩检验基于以下假设：

- **配对数据**：样本由配对观测值组成。
- **独立性**：每对观测值之间相互独立。
- **连续性**：差值 ( $D_i = X_i - Y_i$ ) 来自连续分布。
- **对称性**：差值的分布关于 0 对称。
- **可比性**：差值可以被排序。

这些假设对于检验的有效性至关重要。

## 6 检验原理

### 6.1 1. 基本思想

符号秩检验基于配对差值的符号和大小，通过比较正负差值的秩次和，来判断两个相关样本之间是否存在显著差异。

### 6.2 2. 数学推导

- 原假设 ( $H_0$ ): 差值的中位数为零，即两个样本的中位数相等。
- 备择假设 ( $H_1$ ): 差值的中位数不为零，即两个样本的中位数不相等。
- 检验统计量:

– 理论检验统计量:

\* 单侧检验:  $W^+ = \sum_{i:D_i>0} R_i$

\* 双侧检验:  $W = \min(W^+, W^-)$

– 观察到的检验统计量:

\* 单侧检验:  $w^+$

\* 双侧检验:  $w = \min(w^+, w^-)$

## 7 检验步骤

### 7.1 步骤 1: 计算差值

对于每一对样本，计算差值:

$$d_i = x_i - y_i$$

其中,  $x_i$  为第  $i$  个对象在条件 1 下的测量值,  $y_i$  为条件 2 下的测量值。

### 7.2 步骤 2: 剔除零差值

如果某个差值  $d_i = 0$ , 则从分析中剔除, 以免影响秩次的计算。

### 7.3 步骤 3: 计算绝对差值

对每个非零差值, 计算其绝对值:

$$|d_i|$$

## 7.4 步骤 4：对绝对差值进行排序并赋予秩次

- 将所有  $|d_i|$  按升序排列。
- 为每个  $|d_i|$  赋予秩次  $R_i$ 。
- **处理相同值（平秩处理）**：如果有相同的绝对差值，计算这些差值的平均秩次，赋予它们。

## 7.5 步骤 5：赋予符号

根据原始差值  $d_i$  的正负，为对应的秩次赋予符号：

- 如果  $d_i > 0$ ，则秩次为正  $(+R_i)$ 。
- 如果  $d_i < 0$ ，则秩次为负  $(-R_i)$ 。

## 7.6 步骤 6：计算正负秩次和

- 正秩和  $W^+$ ：

$$W^+ = \sum_{d_i > 0} R_i$$

- 负秩和  $W^-$ ：

$$W^- = \sum_{d_i < 0} R_i$$

## 7.7 步骤 7：计算检验统计量

取较小的秩和作为检验统计量  $W$ ：

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

## 7.8 步骤 8：确定检验的显著性

### 7.8.1 计算期望值 $E(W)$ 和方差 $\text{Var}(W)$

- 有效样本量  $n$ （剔除零差值后的样本数）。
- 期望值  $E(W)$ ：

$$E(W) = \frac{n(n+1)}{4}$$

- 方差  $\text{Var}(W)$ ：

$$\text{Var}(W) = \frac{n(n+1)(2n+1)}{24}$$



### 7.8.2 计算标准化的 $Z$ 值

$$Z = \frac{W - E(W)}{\sqrt{\text{Var}(W)}}$$

### 7.8.3 计算 $p$ 值

$p$  值的计算取决于备择假设的形式：

- 对于  $H_1: \mu > \mu_0$  (右侧检验):  $P(W^+ \geq w^+)$
- 对于  $H_1: \mu < \mu_0$  (左侧检验):  $P(W^+ \leq w^+)$
- 对于  $H_1: \mu \neq \mu_0$  (双侧检验):  $2P(W^+ \leq w)$

在选择适当的  $p$  值计算方法时，需要注意两点：

- 是单侧检验还是双侧检验
- 备择假设是大于、小于还是不等于

### 7.8.4 做出决策

如果计算得到的  $p$  值小于预先设定的显著性水平  $\alpha$  (通常为 0.05)，则拒绝原假设，认为两个样本之间存在显著差异。

## 8 实例解析

### 8.1 假设情境

研究者想评估一种新药对患者血压的影响，测量了 10 位患者在服药前后的血压。

### 8.2 数据

### 8.3 手动计算步骤

#### 8.3.1 步骤 1：计算差值

已在上表中计算。

#### 8.3.2 步骤 2：剔除零差值

无零差值，继续下一步。

患者编号	服药前血压 ( $x_i$ )	服药后血压 ( $y_i$ )	差值 $d_i$
1	150	140	10
2	160	155	5
3	145	148	-3
4	155	150	5
5	170	160	10
6	165	162	3
7	158	155	3
8	152	150	2
9	149	147	2
10	160	158	2

表 2: 患者血压数据及差值

患者编号	差值 $d_i$	绝对差值 $ d_i $
1	10	10
2	5	5
3	-3	3
4	5	5
5	10	10
6	3	3
7	3	3
8	2	2
9	2	2
10	2	2

表 3: 绝对差值计算

### 8.3.3 步骤 3: 计算绝对差值

### 8.3.4 步骤 4: 排序并赋予秩次

### 8.3.5 步骤 5: 赋予符号

根据  $d_i$  的正负, 赋予符号:

### 8.3.6 步骤 6: 计算正负秩次和

- 正秩和  $W^+$ :

$$W^+ = 2.0 + 2.0 + 2.0 + 5.0 + 5.0 + 7.5 + 7.5 + 9.5 + 9.5 = 50.0$$

- 负秩和  $W^-$ :

$$W^- = 5.0$$

排序	患者编号	$ d_i $	秩次 $R_i$
1	8	2	2.0
2	9	2	2.0
3	10	2	2.0
4	3	3	5.0
5	6	3	5.0
6	7	3	5.0
7	2	5	7.5
8	4	5	7.5
9	1	10	9.5
10	5	10	9.5

表 4: 排序并赋予秩次

患者编号	$d_i$	秩次 $R_i$	带符号的秩次
8	2	2.0	+2.0
9	2	2.0	+2.0
10	2	2.0	+2.0
3	-3	5.0	-5.0
6	3	5.0	+5.0
7	3	5.0	+5.0
2	5	7.5	+7.5
4	5	7.5	+7.5
1	10	9.5	+9.5
5	10	9.5	+9.5

表 5: 赋予符号的秩次

### 8.3.7 步骤 7: 计算检验统计量

$$W = \min(W^+, W^-) = \min(50.0, 5.0) = 5.0$$

### 8.3.8 步骤 8: 确定检验的显著性

- 有效样本量  $n = 10$ 。
- 期望值  $E(W)$ :

$$E(W) = \frac{10 \times 11}{4} = 27.5$$

- 方差  $\text{Var}(W)$ :

$$\text{Var}(W) = \frac{10 \times 11 \times 21}{24} = 96.25$$

- 计算  $Z$  值:

$$Z = \frac{W - E(W)}{\sqrt{\text{Var}(W)}} = \frac{5.0 - 27.5}{\sqrt{96.25}} \approx -2.293$$

- 计算  $p$  值:

$$p = 2 \times \Phi(Z) = 2 \times \Phi(-2.293) \approx 0.0218$$

- 由于  $p < 0.05$ , 拒绝原假设, 认为服药前后的血压存在显著差异。

(此处保留之前的手动计算步骤, 内容与之前相同。)

## 8.4 使用 R 语言进行 Wilcoxon 符号秩检验

在实际应用中, 可以使用 R 语言的 ‘wilcox.test()’ 函数来进行符号秩检验, 下面是具体的实现步骤。

### 8.4.1 在 R 中输入数据

```
# 服药前血压数据
x <- c(150, 160, 145, 155, 170, 165, 158, 152, 149, 160)

# 服药后血压数据
y <- c(140, 155, 148, 150, 160, 162, 155, 150, 147, 158)
```

### 8.4.2 使用 wilcox.test 进行符号秩检验

```
# 进行符号秩检验
result <- wilcox.test(x, y, paired = TRUE, exact = FALSE, correct = FALSE)

# 显示结果
print(result)
```

### 8.4.3 输出结果

运行上述代码, 将得到以下输出:

```
Wilcoxon signed rank test

data:  x and y
V = 5, p-value = 0.02194
alternative hypothesis: true location shift is not equal to 0
```

#### 8.4.4 结果解释

- 检验统计量  $V = 5$ ，与手动计算的  $W = 5$  一致。
- $p$  值约为 0.02194，小于显著性水平  $\alpha = 0.05$ 。
- 结论：拒绝原假设，认为服药前后的血压存在显著差异。

#### 8.4.5 参数说明

- $x, y$ ：两个配对样本的数据。
- `paired = TRUE`：表示进行配对检验。
- `exact = FALSE`：对于较大的样本量，设置为 `FALSE`，使用正态近似。
- `correct = FALSE`：不进行连续性校正。

#### 8.4.6 单尾检验

如果需要进行单尾检验（例如，检验服药后血压是否降低），可以设置参数 `alternative`：

```
# 单尾检验：检验服药后血压是否降低
result_one_tailed <- wilcox.test(x, y, paired = TRUE, alternative = "greater", exact
    = FALSE, correct = FALSE)

# 显示结果
print(result_one_tailed)
```

## 9 结论与解释

### 9.1 统计结论

- 拒绝原假设：根据检验结果，服药前后的血压存在显著差异。

### 9.2 实际意义

- 新药有效：新药对降低患者的血压具有显著效果，具有实际应用价值。

## 10 注意事项

### 10.1 1. 数据要求

- 配对数据：样本必须由配对观测值组成。

- **独立性**：每对观测值之间应相互独立。
- **连续性**：差值应来自连续分布。
- **对称性**：虽然不要求正态分布，但差值应关于 0 近似对称分布。
- **可排序性**：差值必须可以被排序。

## 10.2 2. 平秩处理

- **必要性**：当存在相同的绝对差值时，需要进行平秩处理，以保证秩次的正确分配。

## 10.3 3. 样本量

- **小样本量**：对于  $n \leq 25$  的样本，可以使用精确的临界值表进行检验。
- **大样本量**：当  $n > 25$  时，可以使用正态近似，计算  $Z$  值和  $p$  值。

## 10.4 4. 单尾检验与双尾检验

- **双尾检验**：检验差值的中位数是否不等于零。
- **单尾检验**：检验差值的中位数是否大于或小于零，需要调整  $p$  值的计算。

# Wilcoxon 秩和检验 (Rank Sum Test) 详解

在统计分析中,当我们需要比较两个独立样本的平均数但无法满足正态性假设时,**Wilcoxon 秩和检验**是一种常用的非参数检验方法。它放宽了对数据分布的要求,不需要假设数据服从正态分布,也不需要对称性假设。

## 1. 概述

**Wilcoxon 秩和检验**用于比较两个独立样本的平均数,适用于数据不满足正态分布或存在异常值的情况。通过对数据进行秩转换,比较两个样本的秩和,以判断它们是否来自具有相同平均数的总体。

## 2. 与 Wilcoxon 符号秩检验的区别

- **Wilcoxon 秩和检验 (Rank Sum Test)**:
  - **应用场景**：比较两个**独立样本**的平均数。
  - **数据要求**：两个样本之间互相独立。
  - **适用检验**：类似于**独立样本 t 检验** (Two-sample t-test)。

- **Wilcoxon 符号秩检验 (Signed Rank Test):**
  - **应用场景:** 比较两个相关样本、配对样本或同一样本的两两测量。
  - **数据要求:** 数据成对出现, 每对数据具有相关性。
  - **适用检验:** 类似于配对样本 t 检验 (Paired t-test)。

### 3. 检验假设

- **原假设 ( $H_0$ ):** 两个总体的平均数相等,  $\mu_x = \mu_y$ 。
- **备择假设 ( $H_1$ ):**
  - 双尾检验:  $\mu_x \neq \mu_y$ 。
  - 单尾检验:  $\mu_x > \mu_y$  或  $\mu_x < \mu_y$ 。

**假设条件:**

- 样本来自连续分布的总体。
- 两个样本独立, 观测值之间相互独立。
- 总体的分布形状相同, 但平均数可能不同。

### 4. 检验步骤与公式

**步骤 1: 数据合并与排序**

- 将样本  $X$  (大小为  $n_x$ ) 和样本  $Y$  (大小为  $n_y$ ) 的数据合并, 形成总样本量  $N = n_x + n_y$ 。
- 对所有数据按从小到大的顺序进行排序, 赋予秩值  $R_i$ 。
- 若有相同的值 (平级), 赋予它们的平均秩。

**步骤 2: 计算样本  $X$  的秩和  $W$**

$$W = \sum_{i=1}^{n_x} R_i$$

**步骤 3: 计算检验统计量的期望值  $E(W)$  和方差  $\text{Var}(W)$**

- **期望值:**

$$E(W) = \frac{n_x(N+1)}{2}$$

- 方差：

$$\text{Var}(W) = \frac{N(N-1)}{n_x n_y} \left( \sum_{i=1}^N R_i^2 - \frac{N(N+1)^2}{4} \right)$$

**步骤 4：计算标准化检验统计量  $Z$**

$$Z = \frac{W - E(W)}{\sqrt{\text{Var}(W)}}$$

**步骤 5：确定  $p$  值并作出决策**

- 根据  $Z$  值，查找标准正态分布表，得到相应的  $p$  值。
- **决策规则：**
  - 如果  $p$  值小于显著性水平  $\alpha$ （如 0.05），则拒绝原假设  $H_0$ 。
  - 如果  $p$  值大于或等于  $\alpha$ ，则不拒绝  $H_0$ 。

## 5. 手动计算示例

**假设情境：**

某研究人员想比较两种教学方法对学生考试成绩的影响。样本数据如下：

- **教学方法 A（样本  $X$ ， $n_x = 8$ ）：**

85, 78, 90, 88, 76, 95, 89, 84

- **教学方法 B（样本  $Y$ ， $n_y = 7$ ）：**

80, 82, 75, 79, 77, 81, 83

研究人员想知道这两种教学方法是否导致学生**平均成绩**有显著差异。

**步骤 1：数据合并与排序**

将两个样本的数据合并并排序，赋予秩值：



数据值	来自样本	秩值 $R_i$
75	Y	1
76	X	2
77	Y	3
78	X	4
79	Y	5
80	Y	6
81	Y	7
82	Y	8
83	Y	9
84	X	10
85	X	11
88	X	12
89	X	13
90	X	14
95	X	15

**步骤 2：计算样本  $X$  的秩和  $W$**

提取样本  $X$  的秩值并求和：

$$W = 2 + 4 + 10 + 11 + 12 + 13 + 14 + 15 = 81$$

**步骤 3：计算期望值  $E(W)$  和方差  $\text{Var}(W)$**

- 总样本量  $N = n_x + n_y = 8 + 7 = 15$
- 期望值：

$$E(W) = \frac{n_x(N+1)}{2} = \frac{8 \times 16}{2} = 64$$

- 计算总秩值的平方和：

$$\sum_{i=1}^N R_i^2 = 1^2 + 2^2 + 3^2 + \dots + 15^2 = 1240$$

- 方差：

$$\text{Var}(W) = \frac{N(N-1)}{n_x n_y} \left( \sum_{i=1}^N R_i^2 - \frac{N(N+1)^2}{4} \right)$$

计算各项：

$$\begin{aligned}
N(N-1) &= 15 \times 14 = 210 \\
n_x n_y &= 8 \times 7 = 56 \\
\sum_{i=1}^N R_i^2 &= 1240 \\
\frac{N(N+1)^2}{4} &= \frac{15 \times 16^2}{4} = 960 \\
\text{Var}(W) &= \frac{210}{56} \times (1240 - 960) = \frac{210}{56} \times 280 = 1050
\end{aligned}$$

**步骤 4：计算标准化检验统计量  $Z$**

$$Z = \frac{W - E(W)}{\sqrt{\text{Var}(W)}} = \frac{81 - 64}{\sqrt{1050}} = \frac{17}{32.4037} \approx 0.5247$$

**步骤 5：确定  $p$  值并作出决策**

- 查找标准正态分布表，对于  $Z = 0.5247$ ，单尾  $p$  值约为 0.3001。
- **双尾检验：**  
 $p$  值为  $2 \times 0.3001 = 0.6002$ 。
- **决策：**  
由于  $p$  值 (0.6002) 大于显著性水平 0.05，不拒绝原假设  $H_0$ ，认为两种教学方法的平均成绩没有显著差异。

## 6. 使用 R 语言的 `wilcox.test` 进行检验

**步骤 1：输入数据**

```
# 样本 X
x <- c(85, 78, 90, 88, 76, 95, 89, 84)

# 样本 Y
y <- c(80, 82, 75, 79, 77, 81, 83)
```

**步骤 2：执行 Wilcoxon 秩和检验**

```
# 进行 Wilcoxon 秩和检验，不进行连续性校正
result <- wilcox.test(x, y, alternative = "two.sided", exact = FALSE, correct =
  FALSE)

# 查看结果
print(result)
```

### 输出结果：

```
Wilcoxon rank sum test

data:  x and y
W = 81, p-value = 0.6002
alternative hypothesis: true location shift is not equal to 0
```

### 解释：

- **W 值：**输出的  $W$  值为 81，与手动计算结果一致。
- **$p$  值：** $p\text{-value} = 0.6002$ ，符合手动计算的双尾  $p$  值。
- **结论：**由于  $p$  值大于 0.05，不拒绝原假设，认为两种教学方法的平均成绩没有显著差异。

## 7. 总结

**Wilcoxon 秩和检验**是一种强大的非参数统计方法，适用于比较两个独立样本的**平均数**，尤其在数据不满足正态分布或存在异常值的情况下。

- **优点：**
  - 不需要假设数据服从正态分布。
  - 对异常值不敏感。
  - 可以处理不同样本量的情况。
- **限制：**
  - 假设两个总体的分布形状相同，仅平均数可能不同。
  - 当存在大量重复值时，检验的准确性可能受到影响。

在本示例中，通过手动计算和 R 语言的验证，我们得出了相同的结论：两种教学方法的平均成绩没有显著差异。