

Paired Difference Tests of the Center

EN5423 | Spring 2024

W10_paired_diff_test_center_02.pdf
(Week 10)

Contents

1	THE PAIRED <i>T</i>-TEST	1
1.1	NULL AND ALTERNATIVE HYPOTHESES	1
1.2	COMPUTATION OF THE PAIRED <i>T</i> -TEST.....	1
1.3	PERMUTATION TEST FOR PAIRED DIFFERENCES	3
1.4	THE ASSUMPTION OF NORMALITY FOR THE PAIRED <i>T</i> -TEST.....	7
2	GRAPHICAL PRESENTATION OF RESULTS	8
2.1	BOXPLOTS.....	8
2.2	SCATTERPLOTS WITH A ONE-TO-ONE LINE.....	8
3	ESTIMATING THE MAGNITUDE OF DIFFERENCES BETWEEN TWO GROUPS.....	10
3.1	THE MEDIAN DIFFERENCE (FOLLOWING THE SIGN TEST).....	10
3.2	THE HODGES-LEHMANN ESTIMATOR (FOLLOWING THE SIGNED-RANK TEST).....	10
3.3	MEAN DIFFERENCE (FOLLOWING A <i>T</i> -TEST).....	12

Ask people who are flying on vacation about
their stress level and then ask...

...other people...



...the same people...



...about their stress level after their vacation.

1 The Paired t -test

- The paired t -test evaluates *whether the mean difference*, D , of matched pairs is zero. The test requires the paired differences, D_i , to follow a normal distribution.
- Logarithms may be taken to reduce asymmetry in the differences prior to running a paired t -test, but the results will not indicate whether means in the original scale are similar.
- Transformations change the meaning of a mean. After a log transformation, the t -test evaluates whether the ratio of the two geometric means on the original scale equals 1.
- If the means on the original scale are of interest and the paired differences do not follow a normal distribution, use a permutation test instead of a transformation.

1.1 Null and Alternative Hypotheses

The null hypothesis can be stated as

$H_0: \mu_x = \mu_y$ the means for the x_i and y_i are identical, or

$H_0: \mu_{[D]} = 0$ the mean difference between the x_i and y_i equals 0.

The three possible alternative hypotheses are:

$H_{A1}: \mu_x \neq \mu_y$ the two group means differ, and both possible directions are of interest. Reject H_0 when the two-sided p -value $< \alpha$.

$H_{A2}: \mu_x > \mu_y$ prior to seeing any data, μ_x is expected to be greater than μ_y . Reject H_0 when the one-sided p -value $< \alpha$.

$H_{A3}: \mu_x < \mu_y$ prior to seeing any data, μ_y is expected to be greater than μ_x . Reject H_0 when the one-sided p -value $< \alpha$.

1.2 Computation of the Paired t -test

For two sets of data, x_i and y_i , paired by the attribute $i = 1, 2, \dots, n$, compute the paired differences $D_i = x_i - y_i$ and then D , the sample mean of the differences. Test statistic: The paired t -statistic is

$$t_p = \frac{\bar{D}\sqrt{n}}{s} \quad (1)$$

where s is the standard deviation of the differences.

Example 1. Mayfly nymphs—Paired t -test

As the t -test requires that the paired differences follow a normal distribution, test for normality of the differences (D_i) prior to running the paired t -test. The null hypothesis for the Shapiro-Wilk test for normality is that data follow a normal distribution and a small p -value rejects this hypothesis.

```
Above = np.array([12, 15, 11, 41, 106, 63, 296, 53, 20, 110, 429,
185])
Below = np.array([9, 8, 38, 24, 48, 17, 11, 41, 14, 60, 53, 124])
D_i = Above - Below
nymph_list = pd.DataFrame({'Above': Above, 'Below': Below, 'D_i':
Diff})
statistic, p_value = shapiro(D_i)
print("Test statistic:", statistic)
print("P-value:", p_value)
Test statistic: 0.6833876371383667
P-value: 0.0005857152864336967
```

The very small p -value indicates that the paired differences of the mayfly data do not come from a normal distribution at an α of 0.05.

The t -test is run below for demonstration purposes only, as the assumptions of the test are violated.

```
#the ttest_rel function from the scipy.stats module in Python to
perform a paired t-test with specified alternative hypothesis.
# alternative="greater" means we're testing if the mean of Above is
greater than the mean of Below
t_statistic, p_value = ttest_rel(Above, Below, alternative="greater")
print("Paired t-test")
print("t =", t_statistic)
print("p-value =", p_value)
Paired t-test
t = 2.0823691158441737
p-value = 0.030723019706026842
```

The p -value produced is likely to be pushed upwards by the non-normality. Note that it is an order of magnitude higher than that for the signed-rank test.

In an attempt to obtain a distribution closer to normal, the logarithms of the data are computed. As with the signed-rank test, this implies that a multiplicative rather than an additive relation exists between the two sets of data. The Shapiro-Wilk test for normality of the differences between the logarithms has a p -value of 0.083, higher than the α of 0.05, so normality in these units is not rejected. A paired t -test on the difference in logarithms should work well enough, but it will test whether the geometric means of the x_i and y_i , and not their arithmetic means, are similar.

```
Log_D_i = np.log(Above) - np.log(Below)
statistic, p_value = shapiro(Log_D_i)
print("Test statistic:", statistic)
print("P-value:", p_value)
```

```
Test statistic: 0.8782374858856201
P-value: 0.08320952206850052
```

```
t_statistic, p_value = ttest_rel(np.log(Above), np.log(Below),
alternative="greater")
print("Paired t-test")
print("t =", t_statistic)
print("p-value =", p_value)
Paired t-test
t = 2.4420714059059323
p-value = 0.016351195560573684
```

The one-sided p -value for t_p is 0.016. Therefore, reject that $\mu_{\ln x} = \mu_{\ln y}$ in favor of H_{A2} , the mean of the natural log of the x_i is greater than the mean of natural log of the y_i . Equivalently, reject the null hypothesis that the ratio of geometric means of the groups is 1.

1.3 Permutation Test for Paired Differences

- Although the paired t -test assumes normality of the paired differences, a permutation test on the differences can be used to test whether the mean difference equals zero regardless of the distributional shape of the differences.
- These tests perform better than traditional t -tests in the presence of asymmetry. The use of paired differences in the original scale still assumes that an additive difference is appropriate, and that the mean difference is the best measure of difference between the two groups.
- If the variances of the two groups differ, and if the group with higher variance is also the group with the higher mean, a multiplicative difference is probably a better model.
- That appears to be the case for the mayfly data in **Figure 1**, so we have taken natural logarithms prior to computing the paired differences to more accurately model the variation seen in the data.
- The permutation test equivalent to the paired t -test assumes that the differences are symmetric because a mean is being tested, though a normal distribution is not required. The difference in logs of the mayfly data seems relatively symmetric (see **Figure. 2** below).
- A Python function (`permutation_test_mean_difference`) randomly assigns an algebraic sign to each of the observed differences and computes the test statistic, which is simply the mean difference. A distribution of test statistics representing the null hypothesis (H_0 : the mean difference equals zero) is constructed by repeating this process several thousand times.
- The observed mean difference is compared to the distribution of test statistics, and the proportion of the distribution that is outside (above or below for a one-sided test) the observed mean difference is the permutation p -value of the test.

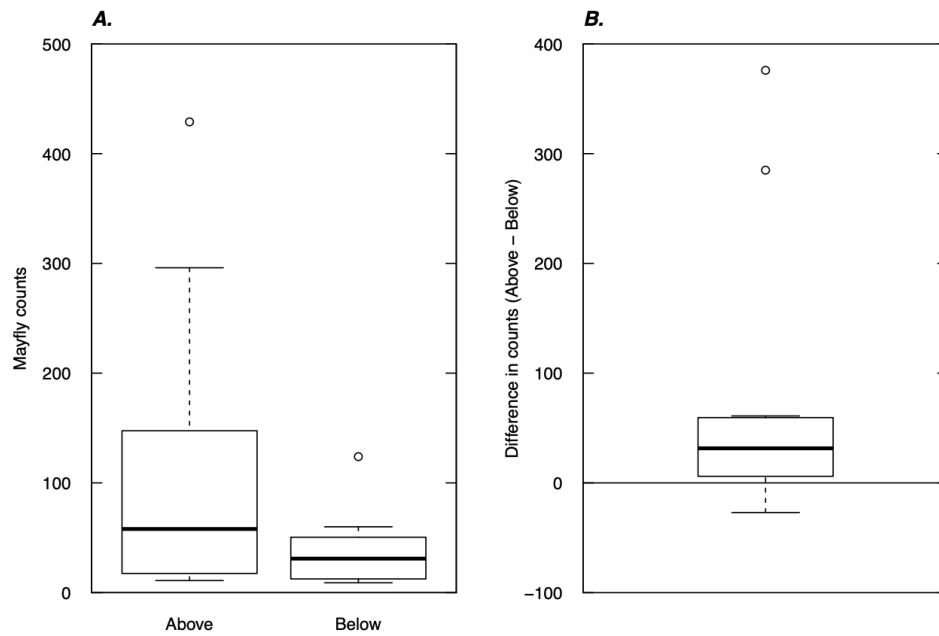
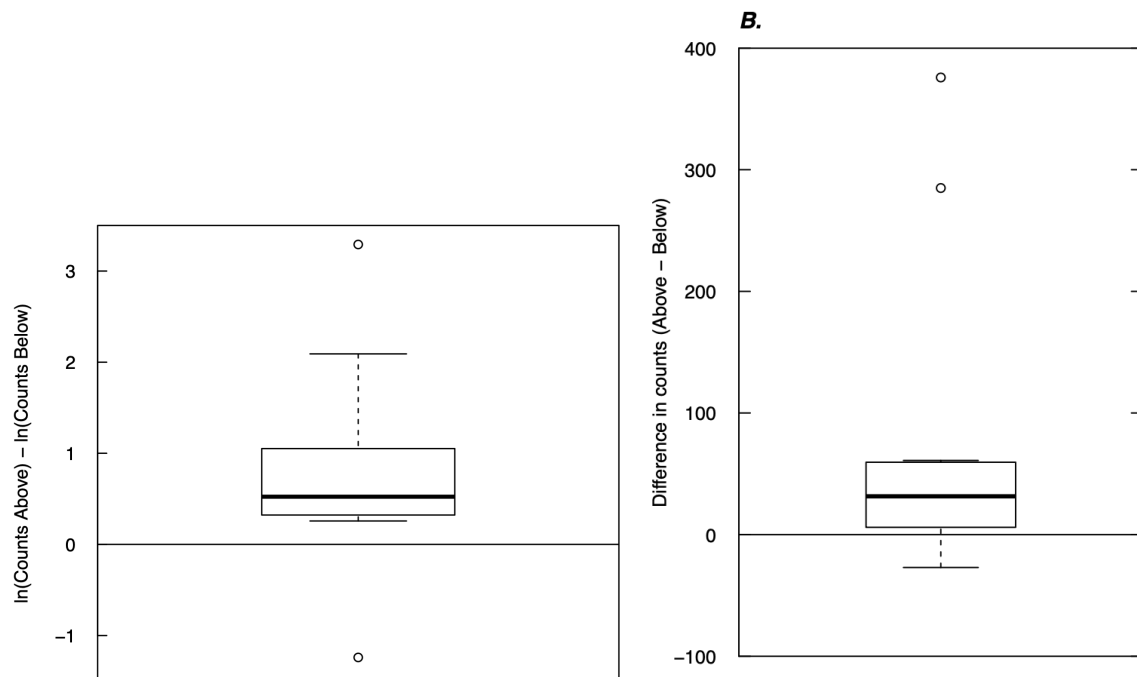


Figure 1. Boxplots of (A) mayfly nymph counts at two different sites, Above and Below, and (B) the differences ($D_i = Above_i - Below_i$)



(Right) Figure 2. Boxplot of the differences of the natural logarithms of the mayfly data from example 1. (left) **Figure 1B**

Example 2. Mayfly nymphs—Permutation test of paired differences.

```
def permutation_test_mean_difference(observed_diff,
n_permutations=10000, alternative='greater'):
    """
    Perform a permutation test of mean differences of matched pairs.

    Parameters:
    observed_diff (array-like): Array of observed differences.
    n_permutations (int): Number of permutations to perform.
    alternative (str): Specifies the alternative hypothesis.
        - 'two-sided': difference is not equal to 0
    (default).
        - 'greater': difference is greater than 0.
        - 'less': difference is less than 0.

    Returns:
    p_value (float): Permutation p-value.
    """

    observed_mean_diff = np.mean(observed_diff)

    # Initialize array to store test statistics from permutations
    permuted_diffs = np.zeros(n_permutations)

    # Perform permutations
    for i in range(n_permutations):
        # Randomly permute signs of differences
        permuted_signs = np.random.choice([-1, 1],
size=len(observed_diff))
        permuted_diff = observed_diff * permuted_signs
        permuted_diffs[i] = np.mean(permuted_diff)

    # Calculate p-value based on the alternative hypothesis
    if alternative == 'two-sided':
        extreme_count = np.sum(np.abs(permuted_diffs) >=
np.abs(observed_mean_diff))
    elif alternative == 'greater':
        extreme_count = np.sum(permuted_diffs >= observed_mean_diff)
    elif alternative == 'less':
        extreme_count = np.sum(permuted_diffs <= observed_mean_diff)
    else:
        raise ValueError("Invalid alternative hypothesis")

    p_value = extreme_count / n_permutations

    return p_value, observed_mean_diff

# Example usage:
```

```
p_value, observed_mean_diff =
permutation_test_mean_difference(Log_D_i)
print("Permutation p-value:", p_value)
print("Mean Difference p-value:", observed_mean_diff)
Permutation p-value: 0.0108
Mean Difference p-value: 0.7764592989644773
```

The permutation test p -value of 0.0108 is slightly less than the normal-theory result of $p = 0.016$, both of which reject the null hypothesis at a 5-percent significance level.

A histogram of the permuted difference in means (**Figure. 3**) shows that the distribution of differences between log counts looks fairly symmetric, though with a broader peak than a normal distribution would have. Even when differences appear to follow a normal distribution the permutation test works well, producing p -values similar to the paired t -test.

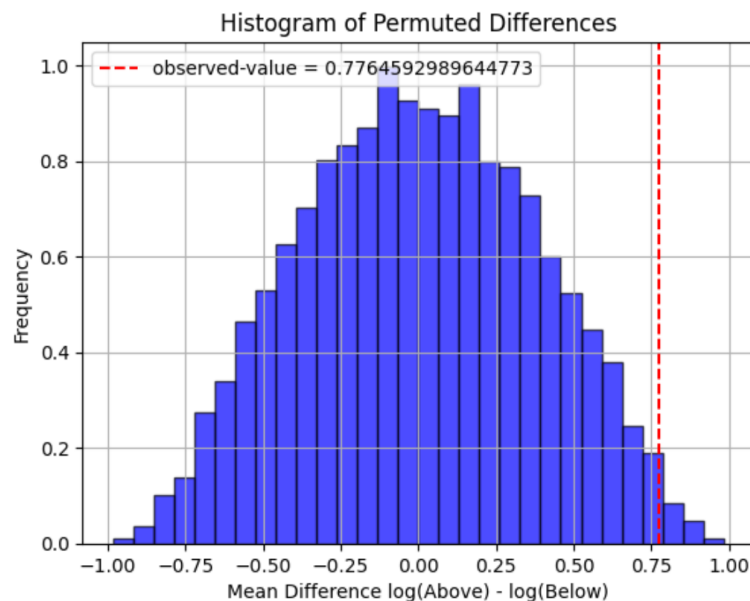


Figure 3. Histogram of permuted differences (representing the null hypothesis) and the observed mean difference from the logs of mayfly data from example 2. The proportion of the entire area at and to the right of the dashed line is the p -value of the test, which equals 0.0108.

Note

The permutation test for mean difference can differ between independent and paired data in terms of how the permutations are performed and what is being tested.

1. Independent Data:

- In the case of independent data, we are typically comparing two separate groups (e.g., treatment vs. control) where each observation in one group is not related to any observation in the other group.
- In a permutation test for independent data, you randomly shuffle the labels (e.g., treatment and control) and compute the test statistic (e.g., difference in means) based on the shuffled labels. This process is repeated many times to obtain the null distribution of the test statistic.

- The null hypothesis usually assumes that there is no difference between the groups, and the permutation test assesses whether the observed difference is statistically significant compared to the null distribution.

2. Paired Data:

- In the case of paired data, each observation in one group is related to a specific observation in the other group (e.g., before-and-after measurements, matched pairs).
- In a permutation test for paired data, you permute the pairs rather than individual observations. This means you shuffle the pairing between the two groups while keeping the individual observations within each group intact.
- The test statistic is typically computed based on the paired differences (e.g., difference between before and after measurements for each pair). The permutation involves randomly permuting the pairing between the two groups and recomputing the test statistic based on the permuted pairs.
- The null hypothesis usually assumes that there is no difference between the paired observations, and the permutation test assesses whether the observed difference is statistically significant compared to the null distribution of permuted differences.

1.4 The Assumption of Normality for the Paired t -test

- The paired t -test is based on the assumption that the differences between paired data (D_i) are **normally distributed** around their mean, with both groups having **the same variance and shape**, differing only in their mean values. If D_i are not normally distributed, particularly if they are asymmetric, the accuracy of p -values derived from the t -test is compromised. Asymmetry also affects the mean's ability to estimate the center of the data, as noted in previous studies.
- Sutton (1993) highlighted that t -tests lose power when the data distribution shows positive skewness, an issue known since the 1920s. This loss of power is more pronounced in one-sided tests than in two-sided tests but occurs in both types. The impact of non-normality leads to incorrect p -values, with traditional t -tests underestimating the extremity of data differences compared to **nonparametric tests like the sign test**.
- An example of the practical implications of choosing the wrong statistical test is shown by computing a t -test on an original scale to determine if one group had more counts than another, ignoring non-normality. The t -test yielded a one-sided p -value of 0.03, significantly higher than the more accurate 0.003 p -value from the sign test. Using a significance level (alpha) of 0.01 would lead to not rejecting the null hypothesis with the t -test, whereas the sign test would reject it.
- It is essential to define the objective of a test clearly to choose the appropriate method. Even if the objective is to compare means, non-normality can confuse t -tests by inflating standard deviation estimates and obscuring differences. The difference in p -values between a normal-theory t -test and a permutation test (which remains effective under normal distributions) underscores the unsuitability of t -tests for data where mean difference is a focus but data are not normally distributed.

- The mean difference of 74.5 counts in the mayfly data was larger than 10 of the 12 pairs, showing that the mean is often not a typical value in skewed data and has limited usefulness for indicating how many more nymphs are found above outfalls than below. When determining whether one group generally has higher values than another, a nonparametric test should be used, with the Hodges-Lehmann median difference estimate providing a more typical measure of group differences.
- Additionally, using transformations before a t -test and then reverting to the original scale can misrepresent the mean difference in the original scale.

2 Graphical Presentation of Results

- Methods for illustrating matched-pair test results are described in week 02 for illustrating a single variable, as the differences between matched pairs are a single variable.
- A probability (Q-Q) plot of the paired differences shows whether or not those differences follow a normal distribution. Here we discuss two plots that illustrate both the test results and the degree of conformity to the test's assumptions.

2.1 Boxplots

- The best method for directly illustrating the results of tests in this chapter is a boxplot of the differences as in **Figure 1**, or as in **Figure 2** for the natural logarithms.
- The number of data above and below zero and the nearness of the median difference to zero are clearly displayed, as is the degree of symmetry of the D_i .
- Although a boxplot is an effective and concise way to illustrate the characteristics of the test for differences, boxplots of the original data for both groups are more intuitive (**Figure 1A**) and might be a good addition for presentations.

2.2 Scatterplots with a One-to-one Line

- Scatterplots effectively demonstrate the relationships between paired data, as seen in **Figure 4**, where each (x_i, y_i) pair is plotted as a point. The line $x = y$ indicates similarity between pairs, showing whether the variable on the x -axis is generally larger or smaller than the variable on the y -axis.
- The scatterplot highlights the general trend that the Above data (x) is greater than the Below data (y), with almost all points falling to the right of a solid line. A line of the form $x = y + d$ is used to visualize the magnitude of differences between x and y , indicating either mean or median differences, depending on the analysis objective.
- In **Figure 4**, the dashed line $x = y - 31.5$ visually represents the median difference of 31.5 counts, suggesting an additive relationship where data points tend to cluster around this line, further illustrating the data's deviation from perfect symmetry.

- It is crucial to understand that a fitted regression line having a slope of 1 does not automatically imply similarity between paired data. Such a line might still exhibit a significant offset from the $x = y$ line, representing an additive difference that reflects disparities between group means or medians.
- The t -test and permutation tests target additive differences between observations, whereas the sign and signed-rank tests assess whether one group consistently shows higher values than the other. The selection of statistical tests should align with the study's objectives, influencing how data should be plotted and interpreted.
- For datasets where a multiplicative relationship is suspected to be more suitable, transformations such as logarithms are recommended. For example, transforming the data and plotting the line $x = y \cdot f^{-1}(d)$ can help visualize relationships in the data more effectively, where $f^{-1}(d)$ for natural logs is $\exp(d)$, as illustrated in **Figure 5** with a transformed line showing the Hodges-Lehmann estimator of group difference in natural log of counts.

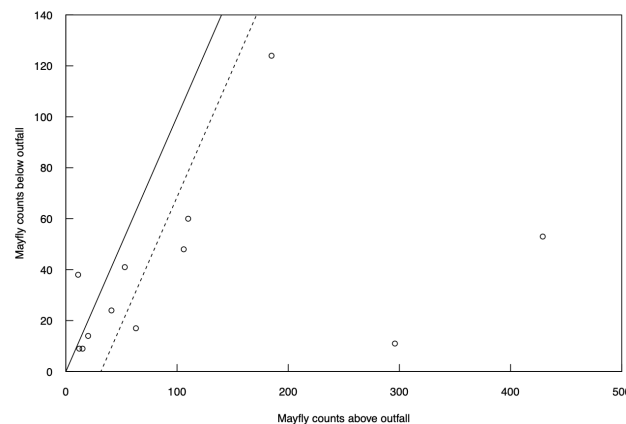


Figure 4. Scatterplot of the mayfly data from Example 1. Solid line is the $x = y$ line. Dashed line is the $y = x - \text{median difference line}$, $y = x - 31.5$. Here x values generally fall to the right of the $x = y$ line, showing that mayfly counts above the outfall are frequently greater than those below the outfall.

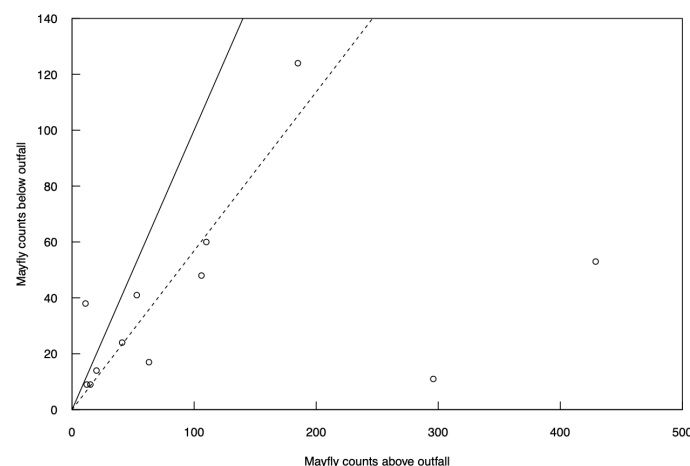


Figure 5. Mayfly data from example 1. The multiplicative relation $y = 0.555 \cdot x$ is shown as the dashed line. For reference, the solid line is the $x = y$ line.

3 Estimating the Magnitude of Differences Between Two Groups

- After testing for differences between matched pairs, it is important to *quantify that difference*. Typically, the mean difference, denoted as \bar{D} , serves as an efficient estimator if the data does not contain outliers and the mean is a reliable central value.
- The mean difference is particularly suitable for situations where the paired t -test or its permutation test equivalent is applied, given the data meets the assumptions of these tests, including normality and symmetry.
- In cases where outliers exist or data exhibits non-normality, the *Hodges-Lehmann estimator*, represented as $\hat{\Delta}$, becomes a more robust choice. This estimator is the median of all possible pairwise averages of the differences.
- The Hodges-Lehmann estimator is the recommended measure of difference when employing the signed-rank test, fitting well with data that may not adhere *strictly to normality assumptions*.
- For data sets where the differences, D_i , are not symmetric, the appropriate measure when using the sign test is the median of the differences, referred to as D_{med} . This estimator directly reflects the central tendency without assuming symmetry or normality.

3.1 The Median Difference (Following the Sign Test)

- For the mayfly data, the median difference (D_{med}) is 31.5, suggesting a significant disparity as the data are asymmetric and do not fit an additive model. This median value is pivotal in understanding the level of difference between the sites above and below the outfalls, indicating that half of the differences are larger and half are smaller than the median.
- Subtraction of the median from the data above the outfalls shows no evidence to reject the null hypothesis (H_0) when tested using the sign test, confirming the median as the most fitting measure for assessing how far from equality the two groups stand on the original scale.
- The confidence interval on the median provides clarity on the precision of the group differences as determined by the sign test. This interval, detailed previously in chapter week04, helps quantify the uncertainty around the median difference, thereby supporting its use in statistical assessments.

3.2 The Hodges-Lehmann Estimator (Following the Signed-rank Test)

- The estimate of difference between groups associated with the signed-rank test is the Hodges-Lehmann difference, $\hat{\Delta}$. *When outliers or non-normality are suspected*, it is a more robust estimator of the difference between groups than is the difference in means.
- Hodges-Lehmann estimators are computed as the median of all possible appropriate combinations of the data; they are associated with many nonparametric test procedures. For the matched-pairs situation, $\hat{\Delta}$ is the median of the $n \cdot \frac{(n+1)}{2}$ possible pairwise averages:

$$\hat{\Delta} = \text{median}[A_{ij}] \quad (2)$$

where $A_{ij} = \lfloor (D_i + D_j)/2 \rfloor$ for all $i \leq j$.

- Note that this version differs from the equation in week08 because this estimator is based on paired differences.
- Week08 presented a Hodges-Lehman estimator for two independent datasets. The estimator is related to the signed-rank test in that subtracting $\hat{\Delta}$ from all paired differences (or equivalently, from the x_i or y_i , whichever is larger) would cause the signed-rank test to have a test statistic W^+ close to 0 and find no evidence of difference between data pairs.
- For the cases of symmetric differences where the signed-rank test is appropriate, the Hodges-Lehmann estimator $\hat{\Delta}$ more efficiently measures the additive difference between two data groups than does the sample median of the differences, D_{med} .
- The provided Python code computes $\hat{\Delta}$, calling it the (psuedo)median. For the mayfly data, $\hat{\Delta}$ of the natural logarithms = 0.5708.

```
# Compute the logarithms of the 'Above' and 'Below' data
log_Above = np.log(Above)
log_Below = np.log(Below)

# Performing the Wilcoxon signed-rank test
result = wilcoxon(log_Above, log_Below, alternative='greater')

# Output the test results
print("Wilcoxon signed rank test")
print(f"V statistic: {result.statistic}, p-value: {result.pvalue}")
n = len(log_Above)
differences = log_Above - log_Below
sorted_diff = np.sort(differences)

# Median estimate (pseudo median)
pseudo_median = np.median(differences)

# Output confidence interval and pseudo median
print("Sample estimates:")
print(f"(pseudo)median: {pseudo_median}")

# Computing exp(-pseudo_median)
exp_neg_median = np.exp(-pseudo_median)
print(f"exp(-pseudo_median): {exp_neg_median}")
Wilcoxon signed rank test
V statistic: 69.0, p-value: 0.008056640625
Sample estimates:
(pseudo)median: 0.5708270199633392
exp(-pseudo_median): 0.5650579312176132
```

- The log of upstream counts minus $\hat{\Delta}$ models the log of the counts below the outfalls. Thus, the counts above the outfalls multiplied by $e^{-0.5708} = 0.5650$ models the counts below the outfalls (the dashed line $y = 0.555 \cdot x$ in **Figure 5**).
- The nonparametric confidence interval around $\hat{\Delta}$ is computed by finding the ranks of the data points representing the ends of the interval. These are a function only of the sample size and α . The pairwise average differences, A_{ij} , are ordered from smallest to largest, and those corresponding to the computed ranks are the ends of the confidence interval.
- For small sample sizes, quantiles for the signed-rank test at the α nearest to $\alpha/2$ and $1 - \alpha/2$ (two-sided interval) or at α (one-sided interval) give the ranks R_u and R_l corresponding to the A_{ij} at the upper and (or) lower confidence limits for $\hat{\Delta}$.
- These limits are the R_l^{th} ranked A_{ij} going from one or both ends of the sorted list of $n \cdot (n + 1) / 2$ differences
- For larger sample sizes where the large sample approximation is used, quantiles of standard normal distribution provide the upper and lower ranks of the pairwise average differences, A_{ij} , corresponding to the ends of the confidence interval. Those ranks are:

$$R_l = \frac{N - z_{\alpha/2} \cdot \sqrt{\frac{n(n+1)(2n+1)}{6}}}{2} \quad (3)$$

$$R_u = \frac{N + z_{\alpha/2} \cdot \sqrt{\frac{n(n+1)(2n+1)}{6}}}{2} + 1 = N - R_l + 1 \quad (4)$$

where $N = \frac{n(n+1)}{2}$

For one-sided intervals, choose the appropriate lower or upper limit using α instead of $\alpha/2$.

3.3 Mean Difference (Following a t -test)

- For the situation where the differences are not only symmetric but also normally distributed and the t -test is used, the most efficient (precise) estimator of the difference between the two groups is the mean difference, \bar{D} .
- However, \bar{D} is in this case only slightly more efficient than is $\hat{\Delta}$, so that when the data depart from normality even slightly the Hodges-Lehmann estimator is just as efficient as \bar{D} . This mirrors the power characteristics of their associated tests, as the signed-rank test is as efficient as the t -test for only slight departures from normality (Lehmann, 1975).
- Therefore, when using field data, which is never exactly normal, \bar{D} has little advantage over $\hat{\Delta}$, whereas $\hat{\Delta}$ is more appropriate in a wider number of situations—for data that are symmetric but not normal.

- A confidence interval on \bar{D} is computed exactly like any confidence interval for a mean. For a two-sided interval.

$$CI = \bar{D} \pm t_{\alpha/2, (n-1)} \frac{s}{\sqrt{n}} \quad (5)$$

where s is the standard deviation of the differences, D_i . A one-sided interval uses α instead of $\alpha/2$.

Example 3. Mayfly nymphs—Estimate of mean difference and two-sided confidence interval.

```
def permutation_test_mean_difference(observed_diff,
n_permutations=10000, alternative='greater'):
    """
    Perform a permutation test of mean differences of matched pairs.

    Parameters:
    observed_diff (array-like): Array of observed differences.
    n_permutations (int): Number of permutations to perform.
    alternative (str): Specifies the alternative hypothesis.
        - 'two-sided': difference is not equal to 0
    (default).
        - 'greater': difference is greater than 0.
        - 'less': difference is less than 0.

    Returns:
    p_value (float): Permutation p-value.
    """

    observed_mean_diff = np.mean(observed_diff)

    # Initialize array to store test statistics from permutations
    permuted_diffs = np.zeros(n_permutations)

    # Perform permutations
    for i in range(n_permutations):
        # Randomly permute signs of differences
        permuted_signs = np.random.choice([-1, 1],
size=len(observed_diff))
        permuted_diff = observed_diff * permuted_signs
        permuted_diffs[i] = np.mean(permuted_diff)

    # Calculate p-value based on the alternative hypothesis
    if alternative == 'two-sided':
        extreme_count = np.sum(np.abs(permuted_diffs) >=
np.abs(observed_mean_diff))
    elif alternative == 'greater':
        extreme_count = np.sum(permuted_diffs >= observed_mean_diff)
```

```
elif alternative == 'less':
    extreme_count = np.sum(permuted_diffs <= observed_mean_diff)
else:
    raise ValueError("Invalid alternative hypothesis")

p_value = extreme_count / n_permutations

return p_value, observed_mean_diff

# Example usage:
p_value, observed_mean_diff =
permutation_test_mean_difference(Log_D_i)
print("Permutation p-value:", p_value)
print("Mean Difference p-value:", observed_mean_diff)
Permutation p-value: 0.0108
Mean Difference p-value: 0.7764592989644773
```

The permutation test p-value of 0.0108 is slightly less than the normal-theory result of $p = 0.016$, both of which reject the null hypothesis at a 5-percent significance level.

A histogram of the permuted difference in means (**Figure. 3**) shows that the distribution of differences between log counts looks fairly symmetric, though with a broader peak than a normal distribution would have. Even when differences appear to follow a normal distribution the permutation test works well, producing p -values similar to the paired t -test.