



UNIVERSITÄT  
LEIPZIG

# Metric Tests For Two-Sample Situations

## ANALYSING THE SPARROW DATA SET

-

### BASIC STATISTICS FOR BIOLOGISTS

**Erik Kusch**

*Research Assistant*

University of Leipzig

Faculty of Life Sciences

Institute of Biology

Behavioral Ecology Research Group

Talstrasse 33

D-04103 Leipzig

Germany

email: erik.kusch@uni-leipzig.de

# Summary:

Welcome to our fourth practical experience in R. Throughout the following notes, I will introduce you to a couple statistical approaches for metric or ordinal data when wanting to compare two samples/populations that might be useful to you and are, to varying degrees, often used in biology.

To do so, I will enlist the sparrow data set we handled in our first exercise.

# Contents

<b>1</b>	<b>Preparing Our Procedure</b>	<b>2</b>
1.1	Packages . . . . .	2
1.2	Loading Data . . . . .	2
<b>2</b>	<b>Mann-Whitney U Test</b>	<b>3</b>
2.1	Climate Warming/Extremes . . . . .	3
2.1.1	Sparrow Weight . . . . .	3
2.1.2	Sparrow Height . . . . .	4
2.1.3	Sparrow Wing Chord . . . . .	4
2.2	Predation . . . . .	5
2.2.1	Predator Presence . . . . .	5
2.2.2	Predator Type . . . . .	5
2.3	Competition . . . . .	6
2.4	Sexual Dimorphism . . . . .	7
2.4.1	Sparrow Weight . . . . .	7
2.4.2	Sparrow Height . . . . .	7
2.4.3	Sparrow Wing Chord . . . . .	8
<b>3</b>	<b>Wilcoxon Signed Rank Test</b>	<b>9</b>
3.1	Preparing Data . . . . .	9
3.2	Climate Warming/Extremes . . . . .	9
3.2.1	Sparrow Weight . . . . .	9
3.2.2	Sparrow Height . . . . .	10
3.2.3	Sparrow Wing Chord . . . . .	10
3.3	Predation . . . . .	11
3.4	Competition . . . . .	12

# 1. Preparing Our Procedure

To ensure others can reproduce our analysis we run the following three lines of code at the beginning of our R coding file.

```
rm(list = ls()) # clearing environment
Dir.Base <- getwd() # soft-coding our working directory
Dir.Data <- paste(Dir.Base, "Data", sep = "/") # soft-coding our data directory
```

## 1.1 Packages

Using the following, user-defined function, we install/load all the necessary packages into our current R session.

```
# function to load packages and install them if they haven't been installed yet
install.load.package <- function(x) {
  if (!require(x, character.only = TRUE))
    install.packages(x)
  require(x, character.only = TRUE)
}
package_vec <- c()
sapply(package_vec, install.load.package)
```

```
## list()
```

As you can see, we don't need any packages for our analyses in this practical.

## 1.2 Loading Data

During our first exercise (Data Mining and Data Handling - Fixing The Sparrow Data Set) we saved our clean data set as an RDS file. To load this, we use the `readRDS()` command that comes with base R.

```
Data_df_base <- readRDS(file = paste(Dir.Data, "/1 - Sparrow_Data_READY.rds", sep = ""))
Data_df <- Data_df_base # duplicate and save initial data on a new object
```

## 2. Mann-Whitney U Test

We can analyse the significance of two population/sample medians of metric variables which are independent of one another using the `wilcox.test()` function in base R whilst specifying `paired = FALSE`.

### 2.1 Climate Warming/Extremes

*Does morphology of *Passer domesticus* depend on climate?*

Logically, we'd expect morphological aspects of *Passer domesticus* to change given different frequencies and severities of climate extremes. Don't forget, however, that our statistical procedures are usually built on the null hypothesis of no differences or correlations being present and so is the Mann-Whitney U Test.

Our data set recorded three aspects of sparrow morphology and three climate levels (Coastal, Continental, Semi-Coastal). Remember, however, that we set aside four stations (Siberia, United Kingdom, Reunion and Australia) to test climate effects on which are strictly limited to continental and coastal climate types. We need to exclude all other sites records from our data:

```
# prepare climate type testing data
Data_df <- Data_df_base
Index <- Data_df$Index
Rows <- which(Index == "SI" | Index == "UK" | Index == "RE" | Index == "AU")
Data_df <- Data_df[Rows, ]
Data_df$Climate <- droplevels(Data_df$Climate)
```

#### 2.1.1 Sparrow Weight

Let's start with the weight of *Passer domesticus* individuals as grouped by the climate type present at the site weights have been recorded at:

```
# Analysis
with(Data_df,
      wilcox.test(x = Weight[Climate == "Continental"],
                  y = Weight[Climate != "Continental"],
                  paired = FALSE)
)

##
## Wilcoxon rank sum test with continuity correction
##
## data: Weight[Climate == "Continental"] and Weight[Climate != "Continental"]
## W = 20000, p-value <2e-16
## alternative hypothesis: true location shift is not equal to 0

# Direction of effect
median(Data_df$Weight[which(Data_df$Climate == "Continental")], na.rm = TRUE) >
  median(Data_df$Weight[which(Data_df$Climate != "Continental")], na.rm = TRUE)

## [1] TRUE
```

Quite obviously, the weight of sparrows seems to be dependent on the type of climate the individuals are suspected to ( $p = 1.72 \times 10^{-32}$ ) and **reject the null hypothesis**. Weights of sparrows in continental climates are, on average, heavier than respective weights of their peers in coastal climates.

### 2.1.2 Sparrow Height

```
# Analysis
with(Data_df,
      wilcox.test(x = Height[Climate == "Continental"],
                  y = Height[Climate != "Continental"],
                  paired = FALSE)
)

##
## Wilcoxon rank sum test with continuity correction
##
## data: Height[Climate == "Continental"] and Height[Climate != "Continental"]
## W = 10000, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0

# Direction of effect
median(Data_df$Height[which(Data_df$Climate == "Continental")], na.rm = TRUE) >
  median(Data_df$Height[which(Data_df$Climate != "Continental")], na.rm = TRUE)

## [1] FALSE
```

We conclude that the height of sparrows seems to not be dependent on the type of climate the individuals are suspected to ( $p = 0.2$ ) and **accept the null hypothesis**. Sparrows in continental climates are, on average, smaller than their peers in coastal climates but not to a statistically significant degree.

### 2.1.3 Sparrow Wing Chord

```
# Analysis
with(Data_df,
      wilcox.test(x = Wing.Chord[Climate == "Continental"],
                  y = Wing.Chord[Climate != "Continental"],
                  paired = FALSE)
)

##
## Wilcoxon rank sum test with continuity correction
##
## data: Wing.Chord[Climate == "Continental"] and Wing.Chord[Climate != "Continental"]
## W = 10000, p-value = 0.01
## alternative hypothesis: true location shift is not equal to 0

# Direction of effect
median(Data_df$Wing.Chord[which(Data_df$Climate == "Continental")], na.rm = TRUE) >
  median(Data_df$Wing.Chord[which(Data_df$Climate != "Continental")], na.rm = TRUE)

## [1] FALSE
```

Apparently, the wing chord of sparrows seems to be dependent on the type of climate the individuals are suspected to ( $p = 0.01$ ) and **reject the null hypothesis**. Sparrows in continental climates have, generally speaking, shorter wings than their peers in coastal climates.

## 2.2 Predation

*Does nesting height of nest sites of *Passer domesticus* depend on predator characteristics?*

In our second practical (Nominal Tests), we used a Chi-Squared approach in a two-sample situation to identify whether predator presence and type had any influence over the nesting sites that individuals of *Passer domesticus* preferred. Our findings showed that they did and so we should expect similar results here when using **Nesting Site** as our response variable instead of **Nesting Height** as these two are highly related to each other.

Additionally, to save some space in these notes, I am not showing how to identify the direction of the effect via code any more for now. We may wish to use the entirety of our data set again for this purpose:

```
Data_df <- Data_df_base
```

### 2.2.1 Predator Presence

First, we start with a possible link to predator presence and nesting height chosen by common house sparrows:

```
with(Data_df,
      wilcox.test(x = Nesting.Height[Predator.Presence == "Yes"],
                  y = Nesting.Height[Predator.Presence != "Yes"],
                  paired = FALSE)
)

##
## Wilcoxon rank sum test with continuity correction
##
## data: Nesting.Height[Predator.Presence == "Yes"] and Nesting.Height[Predator.Presence != "Yes"]
## W = 30000, p-value = 8e-04
## alternative hypothesis: true location shift is not equal to 0
## [1] FALSE
```

The nesting height of sparrows depends on whether a predator is present or not ( $p = 7.68 \times 10^{-4}$ ) thus **rejecting the null hypothesis**. Sparrows tend to go for nesting sites in more elevated positions when no predator is present.

### 2.2.2 Predator Type

Again, we might want to check whether the position of a given nest might be related to what kind of predator is present:

```
with(Data_df,
      wilcox.test(x = Nesting.Height[Predator.Type == "Avian"],
                  y = Nesting.Height[Predator.Type != "Avian"],
                  paired = FALSE)
)

##
## Wilcoxon rank sum test with continuity correction
##
## data: Nesting.Height[Predator.Type == "Avian"] and Nesting.Height[Predator.Type != "Avian"]
## W = 5000, p-value <2e-16
## alternative hypothesis: true location shift is not equal to 0
## [1] FALSE
```

We conclude that the nesting height of sparrows depends on what kind of predator is present ( $p = 7.44 \times 10^{-19}$ ) conclusively **rejecting the null hypothesis**. Therefore, we are confident in stating that sparrows tend to go for nesting sites in more elevated positions when non-avian predators are present.

## 2.3 Competition

*Do home ranges of *Passer domesticus* depend on climate?*

We might expect different behaviour of *Passer domesticus* given different climate types. Since **Home Range** is on an ordinal scale (Large, Medium, Small) we can run a Mann-Whitney U Test on these whilst taking **Climate** into account as our predictor variable.

Take note that we need to limit our analysis to our climate type testing sites again as follows:

```
# prepare climate type testing data
Data_df <- Data_df_base
Index <- Data_df$Index
Rows <- which(Index == "SI" | Index == "UK" | Index == "RE" | Index == "AU")
Data_df <- Data_df[Rows, ]
Data_df$Climate <- droplevels(Data_df$Climate)
```

To be able to use the `wilcox.test()` function on **Home Range**, we need to transform its elements into a numeric type. Luckily, this is as easy as using the `as.numeric()` function on the data since it will assign every factor level a number corresponding to its position in `levels()` as follows:

```
levels(Data_df$Home.Range)

## [1] "Large" "Medium" "Small"
HR_vec <- as.numeric(Data_df$Home.Range)
```

As you can see, the levels of **Home.Range** are ordered alphabetically. The `as.numeric()` command will thus transform every record of "Large" into 1, every record of "Medium" into 2 and every record of "Small" into 3.

We are ready to run the analysis:

```
# Analysis
with(Data_df,
  wilcox.test(x = HR_vec[Climate == "Continental"],
              y = HR_vec[Climate != "Continental"],
              paired = FALSE)
)

##
## Wilcoxon rank sum test with continuity correction
##
## data: HR_vec[Climate == "Continental"] and HR_vec[Climate != "Continental"]
## W = 10000, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0

# Direction of effect
median(HR_vec[which(Data_df$Climate == "Continental")], na.rm = TRUE) >
median(HR_vec[which(Data_df$Climate != "Continental")], na.rm = TRUE)

## [1] FALSE
```

According to the output of our analysis, the home ranges of *Passer domesticus* do not depend on the climate characteristics of their respective habitats ( $p = 0.28$ ). Thus, we **accept the null hypothesis**.

As you can see, the median of numeric home ranges is smaller in continental climates (just not statistically significant). Remember that small numeric ranges mean large actual ranges in this set-up and so we can conclude that climates force common house sparrows to adapt to bigger home ranges.

## 2.4 Sexual Dimorphism

*Does morphology of *Passer domesticus* depend on sex?*

If we assume a sexual dimorphism to have manifested itself in *Passer domesticus* over evolutionary time, we'd expect different morphological features of males and females. In our second practical (Nominal Tests) we already assessed this using a Chi-Squared approach in a two-sample situation on colour morphs of the common house sparrows. At the time, we concluded that colouring is not equal for the sexes. So what about characteristics of our sparrows we can put into a meaningful order?

We may wish to use the entirety of our data set again for this purpose:

```
Data_df <- Data_df_base
```

### 2.4.1 Sparrow Weight

We start by assessing the median weight of sparrows again, as driven by their sexes:

```
# Analysis
with(Data_df,
  wilcox.test(x = Weight[Sex == "Male"],
             y = Weight[Sex != "Male"],
             paired = FALSE)
)

##
## Wilcoxon rank sum test with continuity correction
##
## data: Weight[Sex == "Male"] and Weight[Sex != "Male"]
## W = 2e+05, p-value <2e-16
## alternative hypothesis: true location shift is not equal to 0

# Direction of effect
median(Data_df$Weight[which(Data_df$Sex == "Male")], na.rm = TRUE) >
  median(Data_df$Weight[which(Data_df$Sex != "Male")], na.rm = TRUE)

## [1] TRUE
```

We already identified *Climate* to be a major driver of median sparrow weight within this practical. Now, we need to add *Sex* to the list of drivers of sparrow weight ( $p = 8.26 \times 10^{-20}$ ) and **reject the null hypothesis** that sparrow weights do not differ according to *Sex*. Males tend to be heavier than females.

### 2.4.2 Sparrow Height

Let's move on and see if the height/length of our observed sparrows are dependent on their sexes:

```
with(Data_df,
  wilcox.test(x = Height[Sex == "Male"],
             y = Height[Sex != "Male"],
             paired = FALSE)
)

##
## Wilcoxon rank sum test with continuity correction
##
## data: Height[Sex == "Male"] and Height[Sex != "Male"]
## W = 1e+05, p-value = 1
## alternative hypothesis: true location shift is not equal to 0
```



We already identified `Climate` to be a major driver of median sparrow height within this practical. Sparrow `Sex` does not seem to be an informative characteristic when trying to understand sparrow heights ( $p = 0.95$ ). So we **accept the null hypothesis** and don't identify any direction of effects since there is no effect.

### 2.4.3 Sparrow Wing Chord

```
with(Data_df,
      wilcox.test(x = Wing.Chord[Sex == "Male"],
                  y = Wing.Chord[Sex != "Male"],
                  paired = FALSE)
)

##
## Wilcoxon rank sum test with continuity correction
##
## data: Wing.Chord[Sex == "Male"] and Wing.Chord[Sex != "Male"]
## W = 1e+05, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
```

According to our previous analysis within this practical, `Climate` has been determined to be a major driver wing chords of common house sparrows. With our current analysis in mind, we can conclude that the `Sex` of any given *Passer domesticus* individual does not influence the wing chord of said individual ( $p = 0.9$ ). Therefore we **accept the null hypothesis** and don't identify any direction of effects since there is no effect.

## 3. Wilcoxon Signed Rank Test

We can analyse the significance of two population/sample medians of metric variables which are dependent of one another using the `wilcox.test()` function in base R whilst specifying `paired = TRUE`.

### 3.1 Preparing Data

Obviously, none of our data records are paired as such. Whilst one may want to make the argument that many characteristics of individuals that group together might be dependant on the expressions of themselves found throughout said group, we will not concentrate on this possibility within these practicals.

Conclusively, we need an **additional data set with truly paired records** of sparrows. Within our study set-up, think of a **resettling experiment**, were you take *Passer domesticus* individuals from one site, transfer them to another and check back with them after some time has passed to see whether some of their characteristics have changed in their expression.

To this end, presume we have taken the entire *Passer domesticus* population found at our **Siberian** research station and moved them to the **United Kingdom**. Whilst this keeps the latitude stable, the sparrows *now experience a coastal climate instead of a continental one*. After some time (let's say: a year), we have come back and recorded all the characteristics for the same individuals again.

You will find the corresponding *new data* in 2 - Sparrow\_Resettled\_READY.rds. Take note that this set only contains records for the transferred individuals in the **same order** as in the old data set.

```
Data_df_Resettled <- readRDS(file = paste(Dir.Data, "/2 - Sparrow_Resettled_READY.rds",
  sep = ""))
```

With these data records, we can now re-evaluate how the characteristics of sparrows can change when subjected to different conditions than previously thus shedding light on their **plasticity**.

### 3.2 Climate Warming/Extremes

*Does morphology of Passer domesticus depend on climate?*

We have already concluded that the overall morphological aspects of populations of *Passer domesticus* are shaped by climate, but what happens if we take birds from one climate and resettle them to another climate?

#### 3.2.1 Sparrow Weight

First, let's see how the average weight of our individual sparrows changed a year after they were relocated from Siberia to the UK:

```
with(Data_df[which(Data_df$Index == "SI"),],
  wilcox.test(x = Weight,
    y = Data_df_Resettled$Weight,
    paired = TRUE))

##
## Wilcoxon signed rank test with continuity correction
##
## data: Weight and Data_df_Resettled$Weight
## V = 2000, p-value = 2e-09
## alternative hypothesis: true location shift is not equal to 0
```

Apparently, the weight of the individual sparrows have significantly changed following their relocation ( $p = 2.07 \times 10^{-9}$ ) and we **reject the null hypothesis**.

Earlier, we identified sparrows to be heavier in continental climates when compared to coastal climates - does this sentiment hold true with relocated birds?

```
# Direction of effect
median(Data_df$Weight[which(Data_df$Index == "SI")], na.rm = TRUE) >
  median(Data_df_Resettled$Weight, na.rm = TRUE)
```

```
## [1] TRUE
```

Yes, it does. The resettled birds have reduced their median weight (probably not a conscious decision on behalf of the sparrows).

### 3.2.2 Sparrow Height

Secondly, have our relocated sparrows become taller or shorter?

```
with(Data_df[which(Data_df$Index == "SI"),],
      wilcox.test(x = Height,
                  y = Data_df_Resettled$Height,
                  paired = TRUE))
```

```
##
## Wilcoxon signed rank test with continuity correction
##
## data: Height and Data_df_Resettled$Height
## V = 0, p-value = NA
## alternative hypothesis: true location shift is not equal to 0
```

Interestingly enough, we do not receive either meaningful W (V) statistic nor an informative p-value (NA).

This could only be due to one reason:

```
## [1] TRUE
```

Our sparrows have not become any shorter or taller! In fact, no height/length record has changed for any of the sparrows we relocated. This may usually be indicative of a data handling error but, in this case, makes a lot of sense when considering how difficult it may be for mature individuals to change in size.

### 3.2.3 Sparrow Wing Chord

Third, let's check whether wing chords have changed across the board. We can expect them to behave just like height records did:

```
with(Data_df[which(Data_df$Index == "SI"),],
      wilcox.test(x = Wing.Chord,
                  y = Data_df_Resettled$Wing.Chord,
                  paired = TRUE))
```

```
##
## Wilcoxon signed rank test with continuity correction
##
## data: Wing.Chord and Data_df_Resettled$Wing.Chord
## V = 0, p-value = NA
## alternative hypothesis: true location shift is not equal to 0
unique(Data_df$Wing.Chord[which(Data_df$Index == "SI")] == Data_df_Resettled$Wing.Chord)
```

```
## [1] TRUE
```

Indeed, none of the wing chord records have changed.

### 3.3 Predation

*Does nesting height of nest sites of *Passer domesticus* depend on predator characteristics?*

We have already identified predator characteristics at our sites to be influential in the overall nesting site and height of *Passer domesticus*. Does this trend hold true when considering a relocation experiment?

Firstly, we will test whether nesting heights have changed after the relocation. Before we do so, we should first check whether we'd expect a change based on whether **predator presence** is different between Siberia and the UK:

```
PP_Sib <- unique(Data_df$Predator.Presence[which(Data_df$Index == "SI")])
PP_UK <- unique(Data_df_Resettled$Predator.Presence)

PP_Sib == PP_UK
```

```
## [1] TRUE
```

Apparently, predators are present at both of these sites and so we would not expect a significant change in nesting height. Let's check this:

```
with(Data_df[which(Data_df$Index == "SI"),],
      wilcox.test(x = Nesting.Height,
                  y = Data_df_Resettled$Nesting.Height,
                  paired = TRUE))
```

```
##
## Wilcoxon signed rank test
##
## data: Nesting.Height and Data_df_Resettled$Nesting.Height
## V = 60, p-value = 7e-05
## alternative hypothesis: true location shift is not equal to 0
```

There actually is an effect after the resettling! Therefore, we have to **reject the null hypothesis** ( $p = 7.4 \times 10^{-5}$ )! How can this be?

Maybe it has to do with the kind of predator at each site:

```
unique(Data_df$Predator.Type[which(Data_df$Index == "SI")])
```

```
## [1] Avian
## Levels: Avian Non-Avian
```

```
unique(Data_df_Resettled$Predator.Type)
```

```
## [1] Non-Avian
## Levels: Avian Non-Avian
```

As you can see, Siberian sparrows are subject to avian predation whilst the sparrow populations that we monitored in the UK are experiencing non-avian predator presence. A **causal link between nesting height and predator type** seems to be logical!

Which direction is the effect headed? Earlier within this practical, we hypothesized that avian predation forces lower nesting heights in *Passer domesticus* - does this hold true?

```
NH_Sib <- mean(Data_df$Nesting.Height[which(Data_df$Index == "SI")], na.rm = TRUE)
NH_UK <- mean(Data_df_Resettled$Nesting.Height, na.rm = TRUE)

NH_Sib < NH_UK
```

```
## [1] TRUE
```

Yes, it does!

### 3.4 Competition

*Do home ranges of *Passer domesticus* depend on climate?*

Earlier in this practical, we have shown that home ranges of flocks of *Passer domesticus* are affected by the climate conditions they are experiencing. Let's see if our relocated sparrows have altered their behaviour:

```
with(Data_df[which(Data_df$Index == "SI"),],
      wilcox.test(x = as.numeric(Home.Range),
                  y = as.numeric(Data_df_Resettled$Home.Range),
                  paired = TRUE))
```

```
##
## Wilcoxon signed rank test with continuity correction
##
## data: as.numeric(Home.Range) and as.numeric(Data_df_Resettled$Home.Range)
## V = 600, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
```

However, they haven't! Given the p-value of 0.94, we **accept the null hypothesis** and conclude that home ranges of our flocks of sparrows have not changed significantly after the relocation to the UK.

Due to our earlier analysis, we would expect smaller home ranges of sparrows in the UK when compared to their previous home ranges in Siberia. Before testing this, remember that, when converted to numeric records, low values indicate larger home ranges:

```
HR_Sib <- as.numeric(Data_df$Home.Range[which(Data_df$Index == "SI")])
HR_UK <- as.numeric(Data_df_Resettled$Home.Range)

median(HR_Sib, na.rm = TRUE) < median(HR_UK, na.rm = TRUE)
```

```
## [1] TRUE
```

We were right! The assignment of home ranges did shift to accommodate smaller home ranges in the coastal climate of the UK it is just not intense enough for statistical significance - this will be further evaluated in our next seminar.