UNIVERSITÄT
LEIPZIG

Correlation Tests

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 third practical experience in R. Throughout the following notes, I will introduce you to a couple statistical correlation approaches 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

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("DescTools", # Needed for Contingency Coefficient
                 "ggplot2" # needed for data visualisation
                 )
sapply(package_vec, install.load.package)
```

```
## DescTools   ggplot2
##      TRUE      TRUE
```

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. Nominal Scale - Contingency Coefficient

We can analyse correlations/dependencies of variables of the categorical kind using the contingency coefficient by calling the `ContCoef()` function of base `R`.

Keep in mind that the contingency coefficient is not *really* a measure of correlation but merely of association of variables. A value of $c \approx 0$ indicates independent variables.

2.1 Competition

Are Home Ranges of Passer domesticus related to the stations they are observed at?

Testing this question is as easy as plugging the following into the `ContCoef()` function:

```
table(Data_df$Home.Range, Data_df$Index)
```

```
##
##           AU  BE  FG  FI  LO  MA  NU  RE  SA  SI  UK
##   Large   23  16   0  69   0  22  34  29   0  49  27
##   Medium   0   0   0   0  15  32  30   0   0   0  22
##   Small   65  89 250   0  66  13   0  66 114  17  19
```

```
ContCoef(table(Data_df$Home.Range, Data_df$Index))
```

```
## [1] 0.68
```

As we can see, there seems to be a link between the two. However, we can't be sure if these are linked directly or through another variable.

2.2 Predation

Are colour morphs of Passer domesticus linked to predator presence and/or predator type?

This analysis builds on our findings within our previous exercise (Nominal Tests - Analysing The Sparrow Data Set). Remember that, using the two-sample situation Chi-Squared Test, we found no change in treatment effects (as far as colour polymorphism went) for predator type values but did so regarding the presence of predators. Let's repeat this here:

```
table(Data_df$Colour, Data_df$Predator.Presence)
```

```
##
##          No Yes
##   Black  64 292
##   Brown 211  87
##   Grey   82 331
```

```
ContCoef(table(Data_df$Colour, Data_df$Predator.Presence))
```

```
## [1] 0.44
```

```
table(Data_df$Colour, Data_df$Predator.Type)
```

```
##
##         Avian Non-Avian
##   Black   197        95
##   Brown    60        27
##   Grey    233        98
```

```
ContCoef(table(Data_df$Colour, Data_df$Predator.Type))
```

```
## [1] 0.03
```

Here, we find the same results as when using the Chi-Squared statistic and conclude that colour morphs of the common house sparrow are likely to be driven by predator presence but not the type of predator that is present.

Are nesting sites of Passer domesticus linked to predator presence and/or predator type?

Again, following our two-sample situation Chi-Squared analysis from last exercise, we want to test whether nesting site and predator presence/predator type are linked. The Chi-Squared analyses identified a possible link of nesting site and predator type but nor predator presence.

```
table(Data_df$Nesting.Site, Data_df$Predator.Presence)
```

```
##
##          No Yes
##   Shrub  87 205
##   Tree   94 137
```

```
ContCoef(table(Data_df$Nesting.Site, Data_df$Predator.Presence))
```

```
## [1] 0.11
```

```
table(Data_df$Nesting.Site, Data_df$Predator.Type)
```

```
##
##         Avian Non-Avian
##   Shrub   182        23
##   Tree     49        88
```

```
ContCoef(table(Data_df$Nesting.Site, Data_df$Predator.Type))
```

```
## [1] 0.49
```

Whilst there doesn't seem to be any strong evidence linking nesting site and predator presence, predator type seems to be linked to what kind of nesting site *Passer domesticus* prefers thus supporting our Chi-Squared results.

2.3 Sexual Dimorphism

Are sex ratios of Passer domesticus related to climate types?

Recall that, in our last exercise, we found no discrepancy of proportions of the sexes among the entire data set using a binomial test. What we didn't check yet was whether the sexes are distributed across the sites somewhat homogeneously or whether the sex ratios might be skewed according to climate types. Let's do this:

```
# prepare climate type testing data
Data_df <- Data_df_base
Index <- Data_df$Index
# select all data belonging to the stations at which all parameters except for
# climate type are held constant
Rows <- which(Index == "SI" | Index == "UK" | Index == "RE" | Index == "AU")
Data_df <- Data_df[Rows, ]
Data_df$Climate <- droplevels(Data_df$Climate)

# analysis
table(Data_df$Sex, Data_df$Climate)
```

```
##
##          Coastal Continental
##   Female      91          76
##   Male        72          78
```

```
ContCoef(table(Data_df$Sex, Data_df$Climate))
```

```
## [1] 0.065
```

Quite obviously, they aren't and, if there are any patterns in sex ratios to emerge, these are not likely to stem from climate types. Also keep in mind that we have a plethora of other variables at play whilst the information contained within the climate type variable is somewhat constrained and, in this case, bordering on uninformative (i.e. a coastal site in the Arctic might be more closely resembled by a continental mid-latitude site than by a tropical coastal site).

3. Ordinal Scale - Kendall's Tau

3.1 Climate Warming/Extremes

Do heavier sparrows have heavier/less eggs?

A heavier weight of individual females alludes to a higher pool of resources being allocated by said individuals. There are multiple ways they might make use of it, one of them being investment in reproduction. To test how heavier females of *Passer domesticus* utilise their resources in reproduction, we use a Kendall's Tau approach to finding links between female weight and average egg weight per nest/number of eggs per nest.
Obviously, both weight variables are metric in nature and so we could use other methods as well. On top of that, we first need to convert these into ranks before being able to run a Kendall's Tau analysis as follows:

```r
# overwriting changes in Data_df with base data
Data_df <- Data_df_base
# Establishing Ranks of Egg Weight
RankedEggWeight <- c()
RankedEggWeight[
  order(Data_df$Egg.Weight[which(Data_df$Sex == "Female")])
  ] <- 1:length(which(Data_df$Sex == "Female"))
# Establishing Ranks of Female Weight
RankedWeight_Female <- c()
RankedWeight_Female[
  order(Data_df$Weight[which(Data_df$Sex == "Female")])
  ] <- 1:length(which(Data_df$Sex == "Female"))
# Extracting Numbers of Eggs
RankedEggs <- Data_df$Number.of.Eggs[which(Data_df$Sex == "Female")]
```

Luckily enough, the number of eggs per nest already represent a ranked (ordinal) variable and so we can move straight on to running our analyses:

```r
# Test ranked female weight vs. ranked egg weight
cor.test(x = RankedWeight_Female, y = RankedEggWeight,
         use = "pairwise.complete.obs", method = "kendall")
```

```
##
##  Kendall's rank correlation tau
##
## data:  RankedWeight_Female and RankedEggWeight
## z = 20, p-value <2e-16
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##  tau
## 0.58
```

There is strong evidence to suggest that heavier females tend to lay heavier eggs (tau = 0.58 at p ≈ 0).

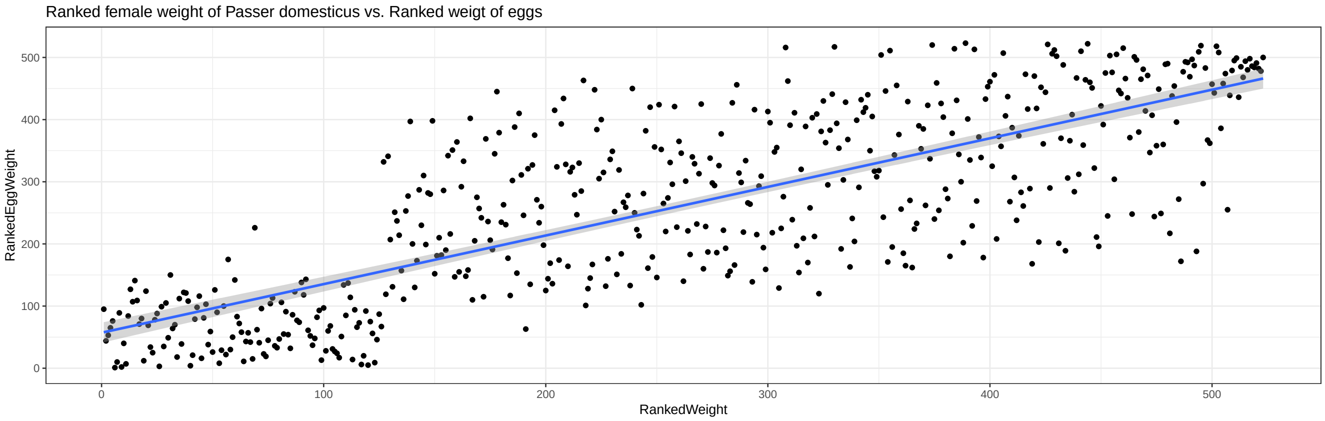
```
# Test ranked female weight vs. number of eggs
cor.test(x = RankedWeight_Female, y = RankedEggs,
         use = "pairwise.complete.obs", method = "kendall")
```

```
##
##  Kendall's rank correlation tau
##
## data:  RankedWeight_Female and RankedEggs
## z = -20, p-value <2e-16
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##   tau
## -0.69
```

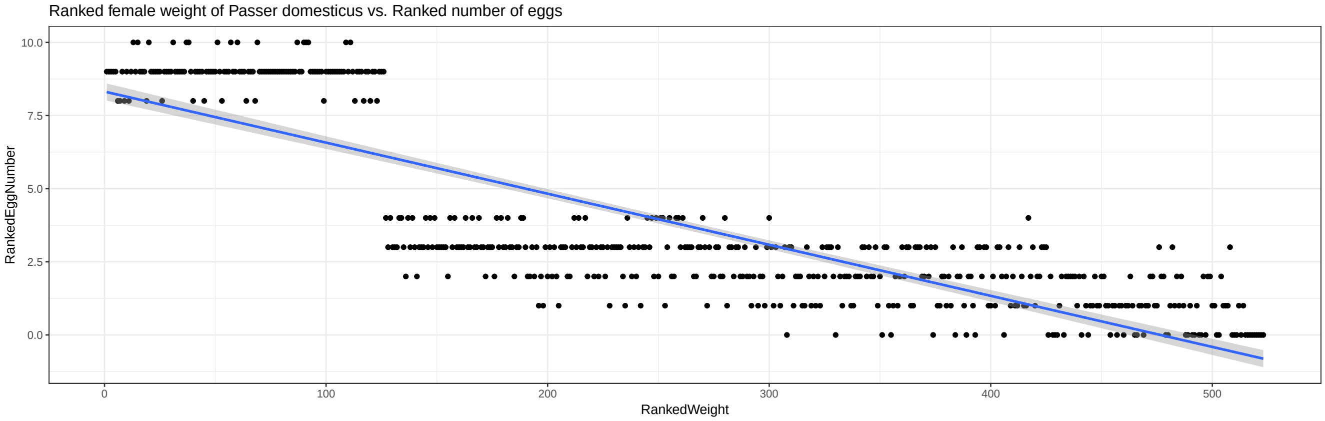
Additionally, the heavier a female of *Passer domesticus*, the less eggs she produces (tau = -0.69 at p ≈ 0).

Now we can visualise the underlying patterns:

```
plot_df <- data.frame(RankedWeight = RankedWeight_Female,
                      RankedEggWeight = RankedEggWeight,
                      RankedEggNumber = RankedEggs)
# plot ranked female weight vs. ranked egg weight
ggplot(data = plot_df, aes(x = RankedWeight, y = RankedEggWeight)) +
  geom_point() + theme_bw() + stat_smooth(method = "lm") +
  labs(title = "Ranked female weight of Passer domesticus vs. Ranked weigt of eggs")
```



```
# plot ranked female weight vs. number of eggs
ggplot(data = plot_df, aes(x = RankedWeight, y = RankedEggNumber)) +
  geom_point() + theme_bw() + stat_smooth(method = "lm") +
  labs(title = "Ranked female weight of Passer domesticus vs. Ranked number of eggs")
```



This highlights an obvious and intuitive trade-off in nature and has us **reject the null hypotheses**.

3.2 Competition

Do heavier sparrows flock together?

Birds of a feather flock together and one may argue that heavier birds may be found in flocks of heavy birds as these flock might share the same beneficial resources:

```
RankedWeight <- c()
RankedWeight[order(Data_df$Weight)] <- 1:length(Data_df$Weight)

cor.test(x = RankedWeight, y = as.numeric(Data_df$Flock),
         use = "pairwise.complete.obs", method = "kendall")
```

```
##
##  Kendall's rank correlation tau
##
## data:  RankedWeight and as.numeric(Data_df$Flock)
## z = -8, p-value = 4e-15
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##   tau
## -0.18
```

ATTENTION! Although this analysis shows a statistically significant negative correlation of ranked bird weight and flock membership this analysis holds **no value at all** since flock assignment at the stations was somewhat random (flock "A" was simply the first flock) and so flock assignment should be treated as intra-site variation and not as inter-site variation as seen here and one could even argue that it isn't an ordinal variable to begin with.

To test what we wanted to test initially (this is done **only to show you how it can be done and not expected of you at this stage**), we'd have to re-run the analysis for every station as follows and give some meaning to the flocks. Therefore, let's order individual sparrows by weight and flocks by their size per site:

```
# establishing an empty data frame and an index vector that doesn't repeat
tau_df <- data.frame(Kendall = as.character(), stringsAsFactors = FALSE)
Indices <- as.character(unique(Data_df$Index))
for(i in 1:length(unique(Data_df$Index))){
  # Ranked weight per station
  RankedWeightStation <- c()
  RankedWeightStation[order(
    Data_df$Weight[which(Data_df$Index == Indices[i])]
    )] <- 1:length(which(Data_df$Index == Indices[i]))
  # ranked flock (by size) per station
  flockranks <- rank(as.numeric(table(Data_df$Flock[which(Data_df$Index == Indices[i])])))
  # flock rank vector and filling
  RankedFlockStation <- rep(NA, length(Data_df$Flock[which(Data_df$Index == Indices[i])]))
    for(k in 1:length(levels(Data_df$Flock))){
      RankedFlockStation[which(
        Data_df$Flock[which(
          Data_df$Index == Indices[i])]
        == levels(Data_df$Flock)[k])] <- flockranks[k]
    }
  # analysis per station
  stationtest <- cor.test(x = RankedWeightStation, y = RankedFlockStation,
                          use = "pairwise.complete.obs", method = "kendall")
  # results to tau_df
  tau_df[1,i] <- as.character(Indices[i])
  tau_df[2,i] <- round(stationtest[["estimate"]][["tau"]],2)
  tau_df[3,i] <- round(stationtest[["p.value"]], 2)
}
```

```
tau_df
```

```
##   Kendall    V2   V3    V4   V5    V6    V7   V8   V9  V10  V11
## 1      SI    UK   AU    RE   NU    MA    LO   BE   FG   SA   FI
## 2   -0.18 -0.08 0.36 -0.11 0.05 -0.36 -0.23 0.09 0.08 0.19 0.31
## 3    0.05  0.41    0  0.14 0.59     0  0.01 0.22 0.07 0.01    0
```

According to these site-wise results, there is no global support for the hypothesis of birds of heavier weights flocking in specific flocks when these are ordered by size. Even the significant site-wise Kendall's Tau results do not tell a holistic story (some show negative correlation, others show a positive correlation). Either we are missing some information here or there simply is no link between sparrow weight and flock membership.

4. Metric and Ordinal Scales

Metric scale correlation analyses call for:

- *Spearman* correlation test (non-parametric)
- *Pearson* correlation test (parametric, requires data to be normal distributed)

4.1 Testing for Normality

Since most of our following analyses are focussing on latitude effects (i.e. "climate warming"), we need to alter our base data. Before we can run the analyses, we need to eliminate the sites that we have set aside for testing climate extreme effects on (Siberia, Unitked Kingdom, Reunion and Australia) from the data set.

```
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 <- Data_df[, -which(colnames(Data_df) == "Population.Status")]  # now redundant
```

In order to know which test we can use with which variable, we need to first identify whether our data is normal distributed using the Shapiro-Test (Seminar 3) as follows:

```
Normal_df <- data.frame(Normality = as.character(), stringsAsFactors = FALSE)
for (i in 1:length(colnames(Data_df))) {
    Normal_df[1, i] <- colnames(Data_df)[i]
    Normal_df[2, i] <- round(shapiro.test(as.numeric(Data_df[, i]))$p.value, 2)
}
colnames(Normal_df) <- c()
rownames(Normal_df) <- c("Variable", "p")
Normal_df
```

```
##
## Variable Index Latitude Longitude Climate Weight Height Wing.Chord Colour Sex
## p            0        0         0       0      0      0          0      0   0
##
## Variable Nesting.Site Nesting.Height Number.of.Eggs Egg.Weight Flock Home.Range
## p                   0              0              0          0     0          0
##
## Variable Predator.Presence Predator.Type
## p                        0             0
```

Unfortunately, none of these variables seems to be normal distributed (this was to be expected for some, to be fair) thus barring us from using Pearson correlation on the entire data set. How about site-wise variable value distributions, though?

```
# establishing an empty data frame and an index vector that doesn't repeat
Normal_df <- data.frame(Normality = as.character(), stringsAsFactors = FALSE)
Indices <- as.character(unique(Data_df$Index))
# remove variables that don't make sense to test for site-wise:
Data_df_Normal <- Data_df[,c("Weight", "Height", "Wing.Chord", "Nesting.Height",
                             "Number.of.Eggs", "Egg.Weight")]
# site-wise shapiro test
for(i in 1:length(colnames(Data_df_Normal))){ # variables
  for(k in 1:length(unique(Data_df$Index))){ # sites
    Normal_df[k,i] <- round(
      shapiro.test(
        as.numeric(
          Data_df_Normal[,i][which(Data_df$Index == Indices[k])])
        )
      $p.value,
      2)
  } # site loop
} # variable loop
colnames(Normal_df) <- colnames(Data_df_Normal)
rownames(Normal_df) <- Indices
Normal_df
```

```
##    Weight Height Wing.Chord Nesting.Height Number.of.Eggs Egg.Weight
## NU   0.57   0.23       0.24           0.02              0       0.05
## MA   0.12   0.03       0.03           0.00              0       0.06
## LO    0.5   0.19       0.17           0.00              0       0.12
## BE   0.38   0.59       0.55           0.00              0       0.71
## FG   0.18   0.88       0.81           0.00              0       0.68
## SA   0.76   0.27       0.26           0.00              0       0.37
## FI   0.43   0.86       0.83           0.00              0       0.29
```

According to these results intra-site correlations can be assessed using Pearson correlation for:

- Weight
- Height
- Wing Chord
- Egg Weight

4.2 Spearman

4.2.1 Climate Warming/Extremes

4.2.1.1 Height/Length

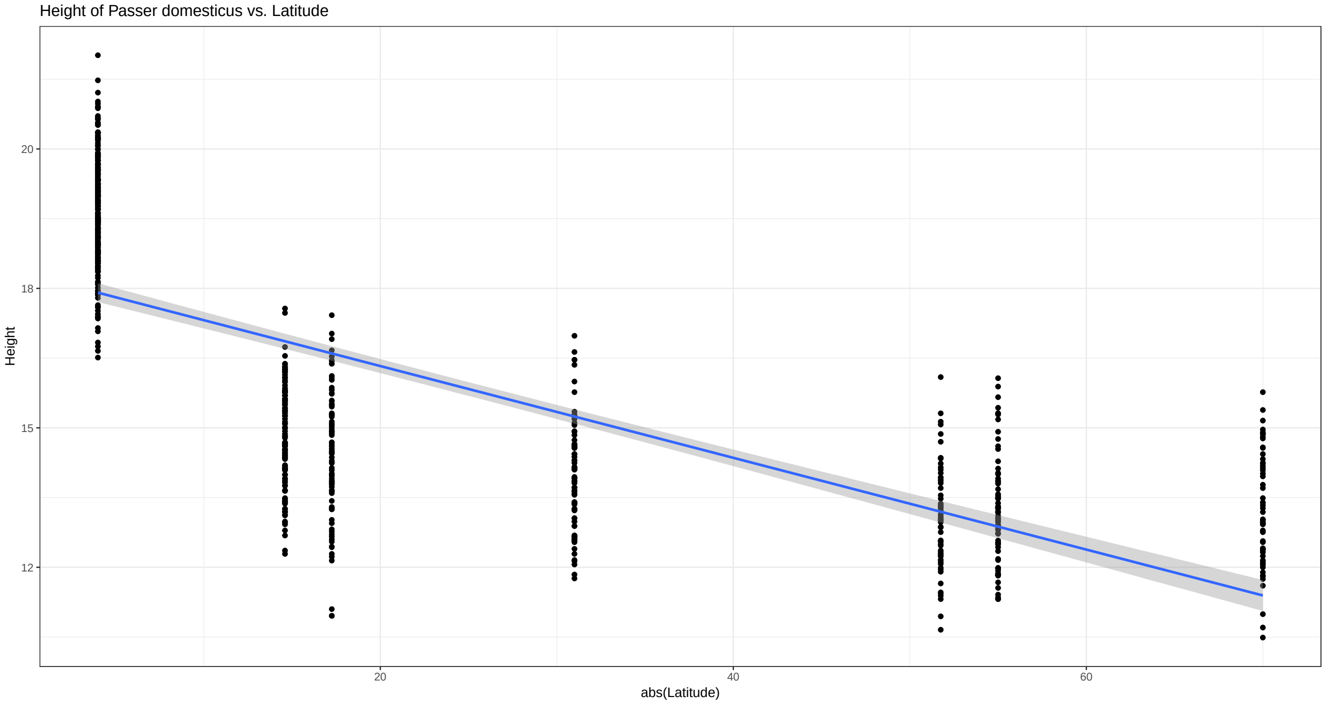
Do height/length records of Passer domesticus and latitude correlate?

Following Bergmann's rule, we expect a positive correlation between sparrow height/length and absolute values of latitude:

```
cor.test(x = abs(Data_df$Latitude), y = Data_df$Height,
         use = "pairwise.complete.obs", method = "spearman")
```

```
##
##  Spearman's rank correlation rho
##
## data:  abs(Data_df$Latitude) and Data_df$Height
## S = 1e+08, p-value <2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##   rho
## -0.82
```

```
ggplot(data = Data_df, aes(x = abs(Latitude), y = Height)) +
  geom_point() + theme_bw() + stat_smooth(method = "lm") +
  labs(title = "Height of Passer domesticus vs. Latitude")
```



Interestingly enough, our analysis yields a negative correlation which would disproof Bermann's rule. This is a good example to show how important biological background knowledge is when doing biostatistics. Whilst a pure statistician might now believe to have just dis-proven a big rule of biology, it should be apparent to any biologist that Bergmann spoke of "bigger" organisms in colder climates (higher latitudes) and not of "taller" individuals. What our sparrows lack in height, they might make up for in circumference. This is an example where we would **reject the null hypothesis** but shouldn't **accept the alternative hypothesis** based on biological understanding.

4.2.1.2 Weight

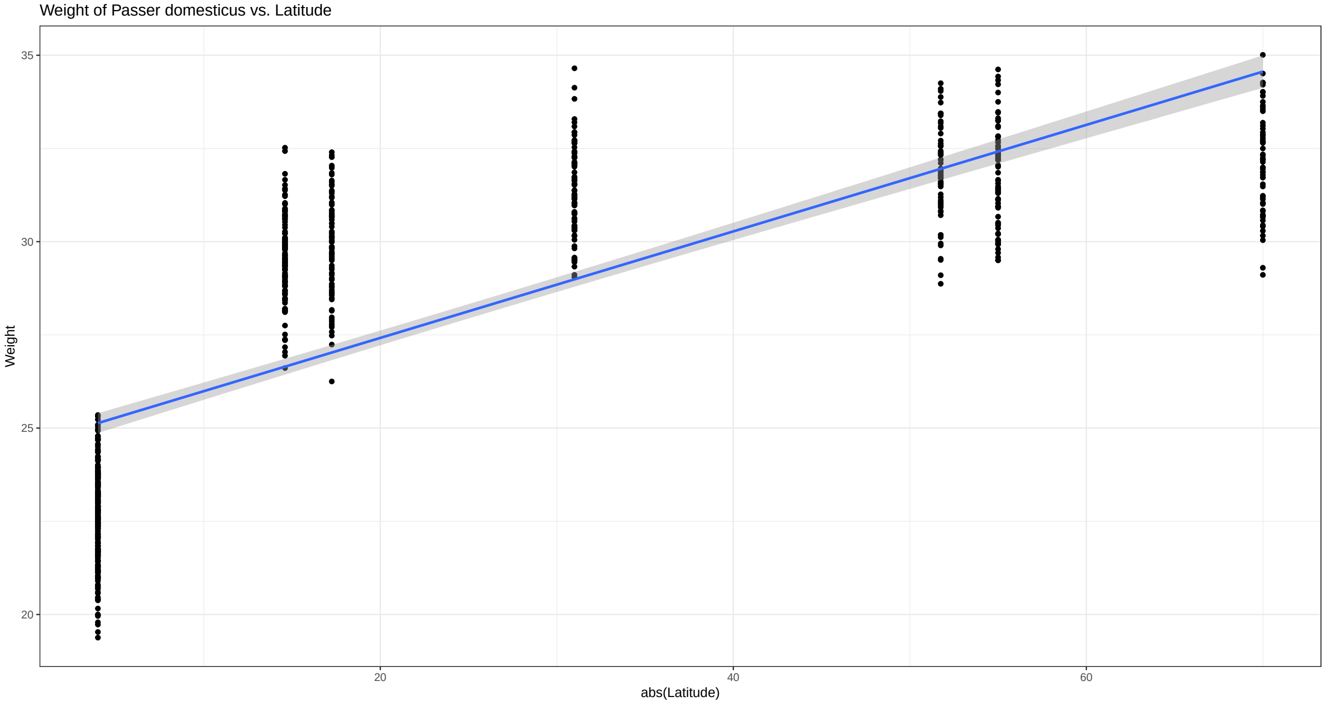
Do weight records of Passer domesticus and latitude correlate?

Again, following Bergmann's rule, we expect a positive correlation between sparrow weight and absolute values of latitude.

```r
cor.test(x = abs(Data_df$Latitude), y = Data_df$Weight,
         use = "pairwise.complete.obs", method = "spearman")
```

```
##
##  Spearman's rank correlation rho
##
## data:  abs(Data_df$Latitude) and Data_df$Weight
## S = 9e+06, p-value <2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##  rho
## 0.87
```

```r
ggplot(data = Data_df, aes(x = abs(Latitude), y = Weight)) +
  geom_point() + theme_bw() + stat_smooth(method = "lm") +
  labs(title = "Weight of Passer domesticus vs. Latitude")
```



Bergmann was obviously right and we **reject the null hypothesis**.

4.2.1.3 Wing Chord

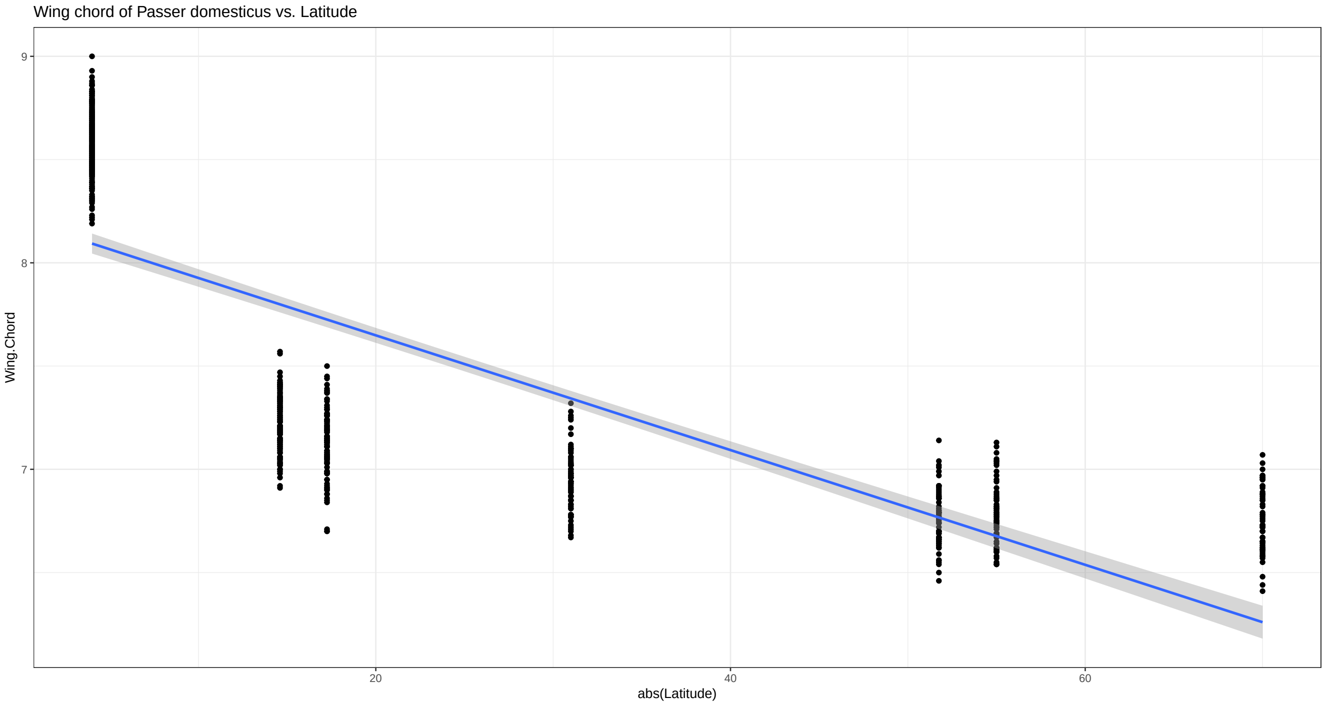
Do wing chord/wing span records of Passer domesticus and latitude correlate?

We would expect sparrows in higher latitudes (e.g. colder climates) to have smaller wings as to radiate less body heat.

```
cor.test(x = abs(Data_df$Latitude), y = Data_df$Wing.Chord,
         use = "pairwise.complete.obs", method = "spearman")
```

```
##
##  Spearman's rank correlation rho
##
## data:  abs(Data_df$Latitude) and Data_df$Wing.Chord
## S = 1e+08, p-value <2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##   rho
## -0.91
```

```
ggplot(data = Data_df, aes(x = abs(Latitude), y = Wing.Chord)) +
  geom_point() + theme_bw() + stat_smooth(method = "lm") +
  labs(title = "Wing chord of Passer domesticus vs. Latitude")
```



And we were right! Sparrows have shorter wingspans in higher latitudes and we **reject the null hypothesis**.

4.2.1.4 Number of Eggs

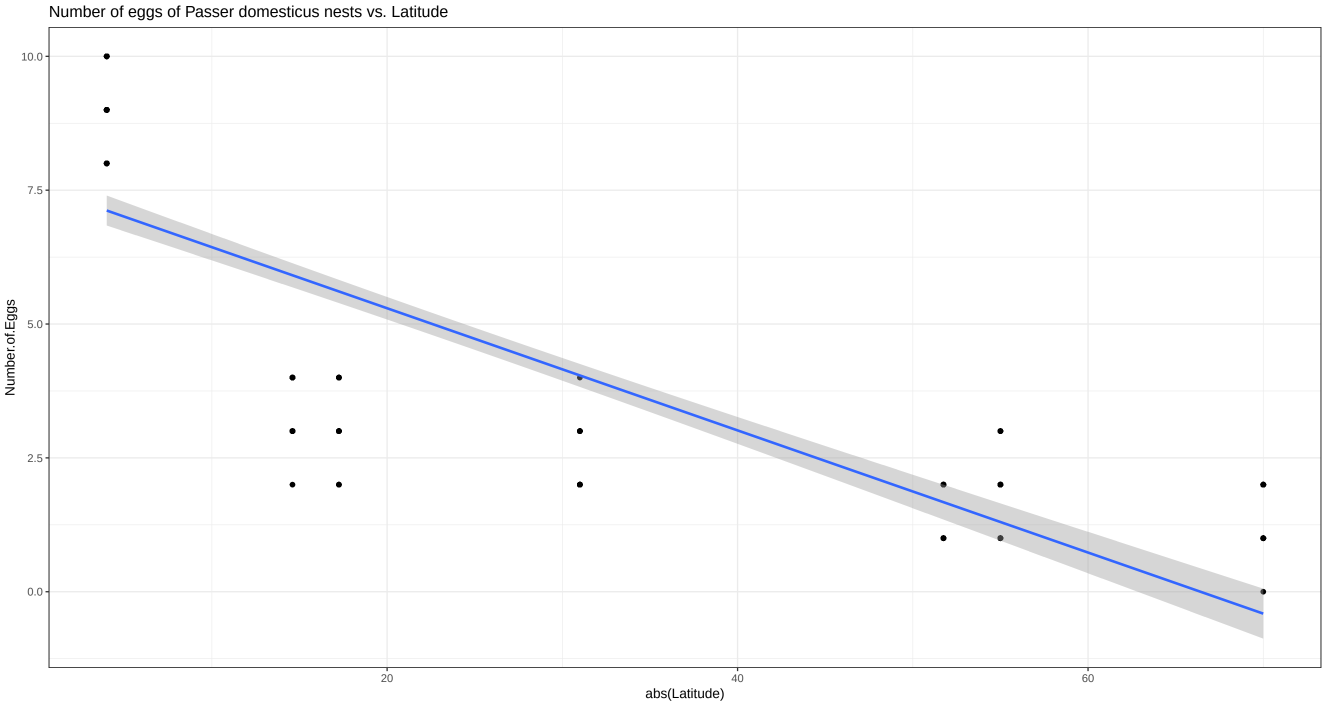
Do numbers of eggs per nest of Passer domesticus and latitude correlate?

Due to resource constraints in colder climates, we expect female *Passer domesticus* individuals to invest in quality over quantity by prioritising caring your fledglings by educing the amount of eggs they produce.

```
cor.test(x = abs(Data_df$Latitude), y = Data_df$Number.of.Eggs,
         use = "pairwise.complete.obs", method = "spearman")
```

```
##
##  Spearman's rank correlation rho
##
## data:  abs(Data_df$Latitude) and Data_df$Number.of.Eggs
## S = 1e+07, p-value <2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##   rho
## -0.93
```

```
ggplot(data = Data_df, aes(x = abs(Latitude), y = Number.of.Eggs)) +
  geom_point() + theme_bw() + stat_smooth(method = "lm") +
  labs(title = "Number of eggs of Passer domesticus nests vs. Latitude")
```



We were right. Female house sparrows produce less eggs per capita in higher latitudes and we **reject the null hypothesis**.

4.2.1.5 Egg Weight

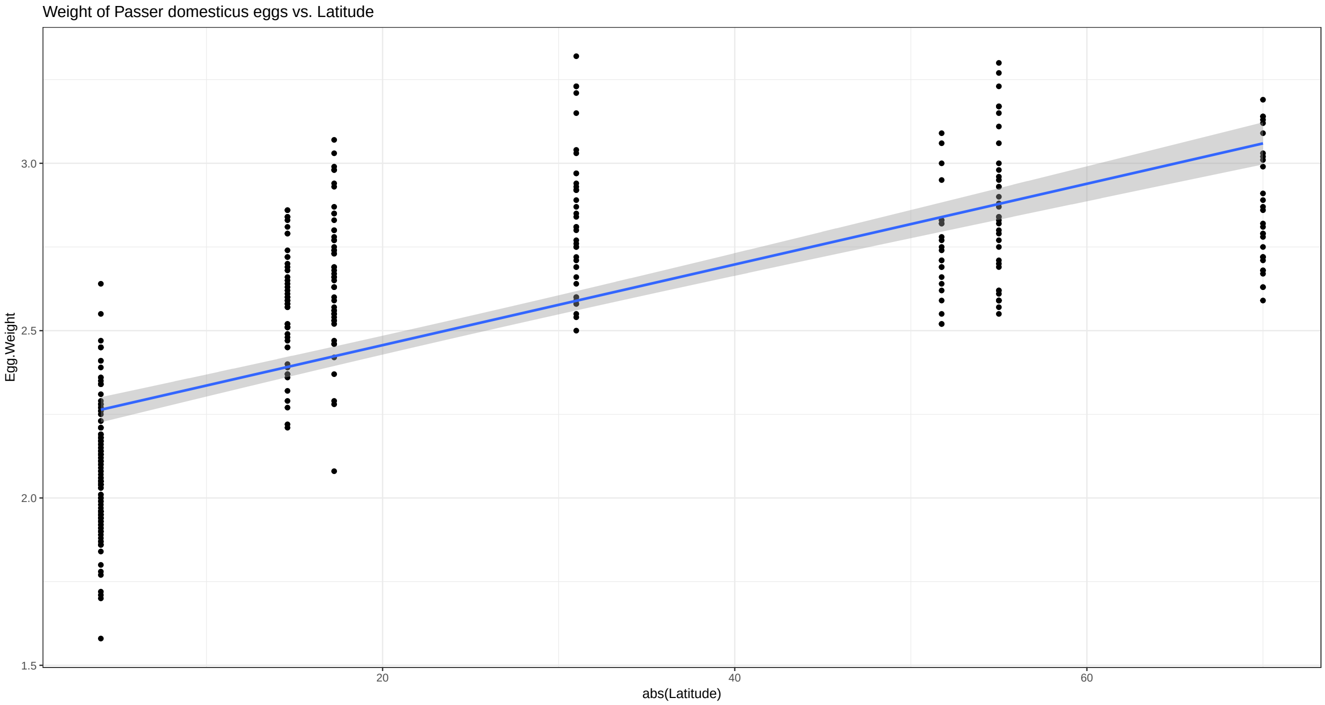
Does average weight of eggs per nest of Passer domesticus and latitude correlate?

Due to the reduced investment in egg numbers that we have just proven, we expect females of *Passer domesticus* to allocate some of their saved resources into heavier eggs which may nurture unhatched offspring for longer and more effectively.

```
cor.test(x = abs(Data_df$Latitude), y = Data_df$Egg.Weight,
         use = "pairwise.complete.obs", method = "spearman")
```

```
##
##  Spearman's rank correlation rho
##
## data:  abs(Data_df$Latitude) and Data_df$Egg.Weight
## S = 1e+06, p-value <2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##  rho
## 0.82
```

```
ggplot(data = Data_df, aes(x = abs(Latitude), y = Egg.Weight)) +
  geom_point() + theme_bw() + stat_smooth(method = "lm") +
  labs(title = "Weight of Passer domesticus eggs vs. Latitude")
```



Indeed, the higher the latitude, the heavier the average egg per nest of *Passer domesticus* and we **reject the null hypothesis**.

4.2.2 Competition

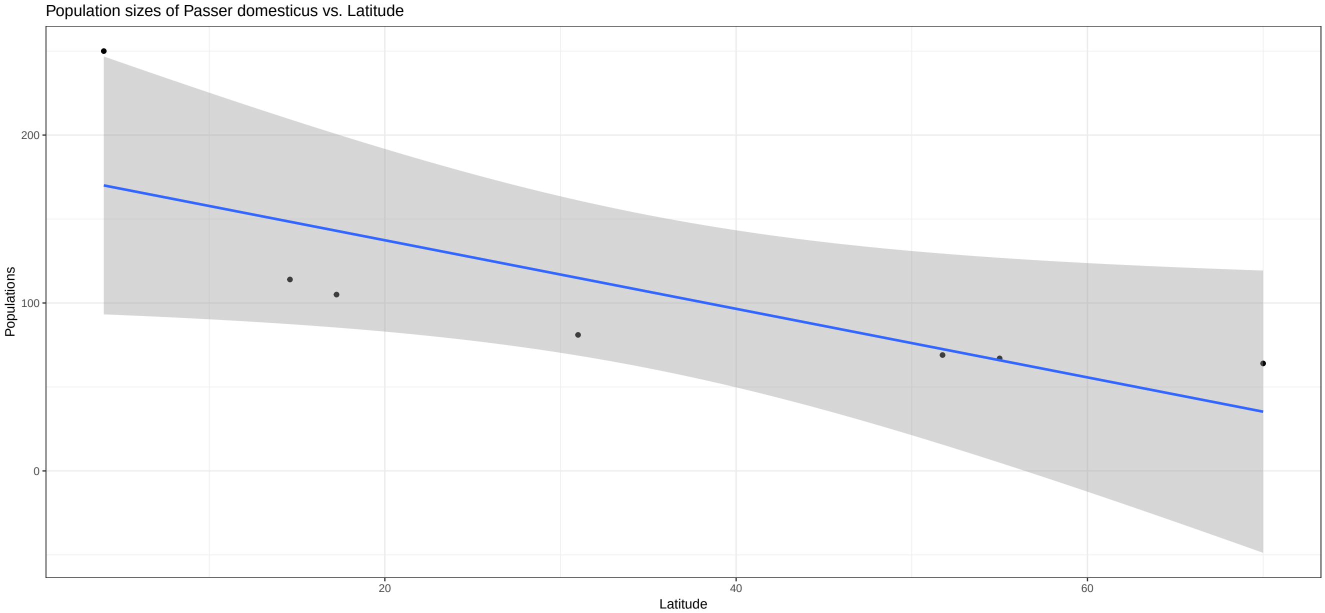
Do population sizes of Passer domesticus and latitude correlate?

Due to increased constraints on avian physiology, we expect habitats of roughly the same size to support less individuals of *Passer domesticus* when moving from warmer to colder climates. To test this, we do the following:

```
# establishing an empty vector and an index vector that doesn't repeat
Pops <- c()
Indices <- as.character(unique(Data_df$Index))
# indexing population sizes
for(i in 1:length(unique(Data_df$Index))){
  Pops[i] <- sum(Data_df$Index == Indices[i])
}
# getting latitudes
Lats <- abs(unique(Data_df$Latitude))
# running test
cor.test(x = Pops, y = Lats,
         use = "pairwise.complete.obs", method = "spearman")
```

```
##
##  Spearman's rank correlation rho
##
## data:  Pops and Lats
## S = 100, p-value = 4e-04
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
##  -1
```

```
# plot
plot_df <- data.frame(Latitude = Lats, Populations = Pops)
ggplot(data = plot_df, aes(x = Latitude, y = Populations)) +
  geom_point() + theme_bw() + stat_smooth(method = "lm") +
  labs(title = "Population sizes of Passer domesticus vs. Latitude")
```



As it turns out, the correlation is not only significant but staggeringly strong and so we **reject the null hypothesis**!

Since we recorded five flocks per site, this trend will translate to average flock size in relation to absolute latitude values and we don't need to run a separate analysis.

4.3 Pearson

In order to show Pearson correlation, we run two simple, site-wise correlation analyses on some of our normal distributed variables.

Do weight and height records of Passer domesticus correlate within each site?

To shed some light on our previous findings, we might want to see whether weight and height of sparrows correlate. Without running the analysis, we can conclude that they do because both correlate with latitude and are thus what we call **collinear**. However, now we are running the analysis on a site level - does the correlation still exist?

Take note that we are now using our entire data set again.

```r
# overwriting altered Data_df
Data_df <- Data_df_base

# establishing an empty data frame and an index vector that doesn't repeat
Pearson_df <- data.frame(Pearson = as.character(), stringsAsFactors = FALSE)
Indices <- as.character(unique(Data_df$Index))

# site-internal correlation tests, weight and height
for(i in 1:length(unique(Data_df$Index))){
  Weights <- Data_df$Weight[which(Data_df$Index == Indices[i])]
  Heights <- Data_df$Height[which(Data_df$Index == Indices[i])]

  Pearson_df[1,i] <- round(cor.test(x = Weights, y = Heights,
                                    use = "pairwise.complete.obs")[["estimate"]][["cor"]], 2)
  Pearson_df[2,i] <- round(cor.test(x = Weights, y = Heights,
                                    use = "pairwise.complete.obs")$p.value, 2)

}
colnames(Pearson_df) <- Indices
rownames(Pearson_df) <- c("r", "p")
Pearson_df
```

```
##     SI   UK   AU   RE   NU   MA   LO   BE   FG   SA   FI
## r 0.76 0.83 0.77 0.75 0.84 0.84 0.79 0.82 0.79 0.76 0.81
## p    0 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
```

Apparently, it does. Heavier birds are taller!

Do nesting height and wing chord records of Passer domesticus correlate within each site?

From a standpoint of flight economics, we might think that birds with shorter wings might prefer to nest closer to the ground to reduce costly flight-time (in terms of fitness):

```r
# establishing an empty data frame and an index vector that doesn't repeat
Pearson_df <- data.frame(Pearson = as.character(), stringsAsFactors = FALSE)
Indices <- as.character(unique(Data_df$Index))
# site-internal correlation tests, nesting.height and wing.chord
for(i in 1:length(unique(Data_df$Index))){
  Wing.Chords <- Data_df$Wing.Chord[which(Data_df$Index == Indices[i])]
  Nesting.Heights <- Data_df$Nesting.Height[which(Data_df$Index == Indices[i])]

  Pearson_df[1,i] <- round(cor.test(x = Wing.Chords, y = Nesting.Heights,
                                    use = "pairwise.complete.obs")[["estimate"]][["cor"]], 2)
  Pearson_df[2,i] <- round(cor.test(x = Wing.Chords, y = Nesting.Heights,
                                    use = "pairwise.complete.obs")$p.value, 2)

}
colnames(Pearson_df) <- Indices
rownames(Pearson_df) <- c("r", "p")
Pearson_df
```

```
##      SI   UK    AU    RE   NU   MA   LO   BE   FG   SA    FI
## r -0.15 0.32 -0.30 -0.20 0.04 0.09 0.27 0.01 0.02 0.04 -0.38
## p   0.4 0.05  0.05  0.15 0.82 0.61 0.10 0.95 0.80 0.78  0.07
```

Although our hypothesis about flight economics was not too far-fetched, we do not see any proof for this here. Keep in mind though that we are looking at intra-population variation here which might not be able to capture all the information that is out there.