Spot_variability

In this document we will assess the hypotheses tested in the manuscript concerning the differences in canalization between native and invasive populations. Our predictions are:

1. **Extreme temperatures will increase the variation in each geographic population, as they will stress regular developmental pathways.**
2. **Invasive populations might show different variation patterns, as a product of their adaptation to new environments.**

Our data consists in a data.frame consisting in an individual label, the genetic lineage (population) to which it belongs, the temperature at which it has been raised, its wing size, its spot size and the spot size/wing size ratio:

```
spot.raw <- read.table(paste('/Users/ceferino_vg/Documents/Drosophila_suzukii/',
                              'Spot_evolution/data/results_plasticity.txt',
                              sep=''),
                        header=T, sep=',')

spot.data <- aggregate(Wing ~ IND*TEMP+POP, FUN = mean, data = spot.raw)

spot.data$Spot <- aggregate(Spot ~ IND*TEMP+POP, FUN = mean, data = spot.raw)$Spot

spot.data$Ratio <- aggregate(Ratio ~ IND*TEMP+POP, FUN = mean, data = spot.raw)$Ratio

spot.data$TEMP <- factor(spot.data$TEMP)

spot.ind <- aov(Spot ~ Wing, data=spot.data)

spot.data$Res <- spot.ind$residuals

head(spot.data)
```

```
##   IND TEMP   POP    Wing    Spot      Ratio         Res
## 1 L04   16 PARIS 1986144 92220.0 0.04644350    196.1911
## 2 L06   16 PARIS 1946261 84597.0 0.04346038  -5451.1637
## 3 L07   16 PARIS 2026222 78718.5 0.03884989 -15290.6384
## 4 L09   16 PARIS 1891875 74713.0 0.03949119 -12641.0682
## 5 L11   16 PARIS 2006696 94066.5 0.04691930   1024.6204
## 6 L13   16 PARIS 1942214 90485.5 0.04658440    637.8099
```

For future analyses, we import the libraries we are going to use and we write the function to estimate Cohen's d as an effect size measure:

```
library(ggplot2)
library(ggsignif)
library(wesanderson)
library(cvequality)
library(car)

cohens.d <- function(pop1, pop2) {
  d <- (mean(pop1)-mean(pop2))/
    sqrt(((length(pop1)-1)*var(pop1)+(length(pop2)-1)*var(pop2))/(length(pop1)+length(pop2)-2))
```

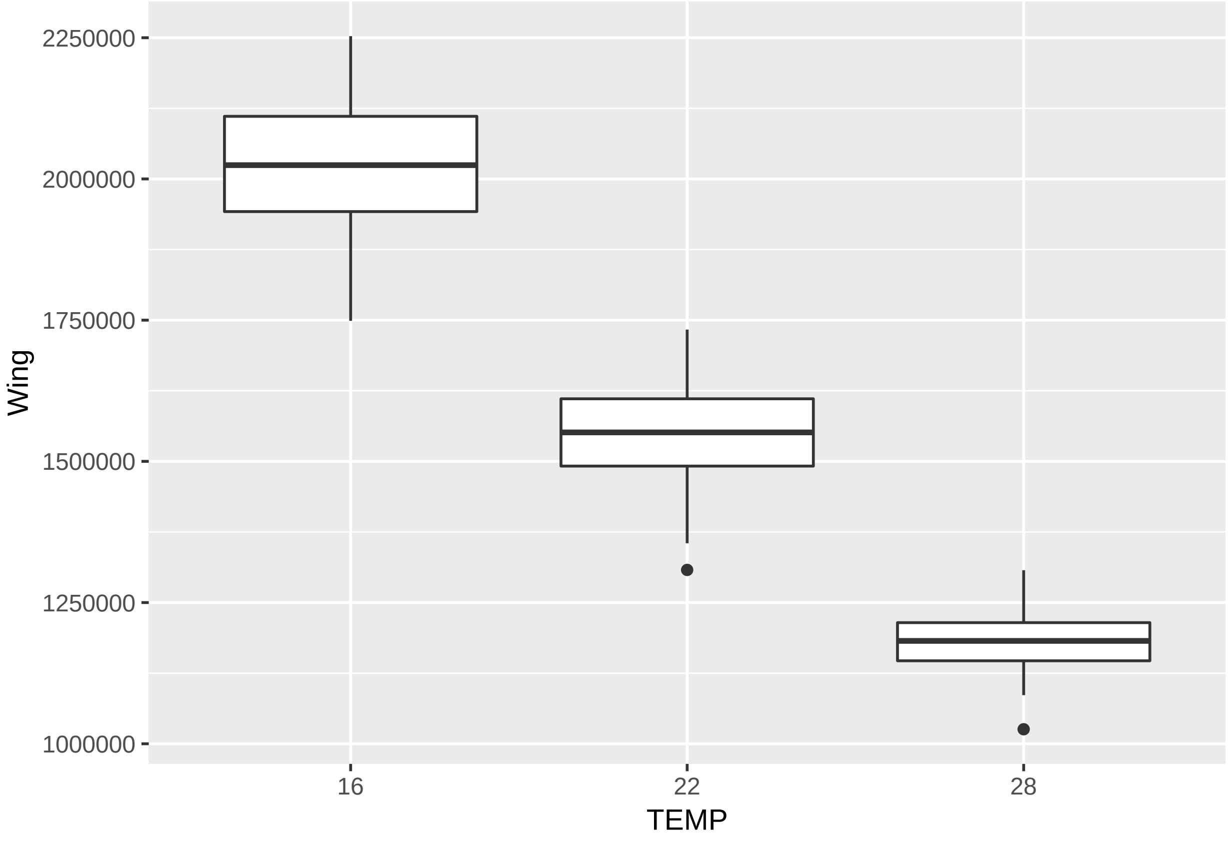
```r
  return(d)
}
```

We take different indexes to study the different phenotypes within and among temperatures and populations.

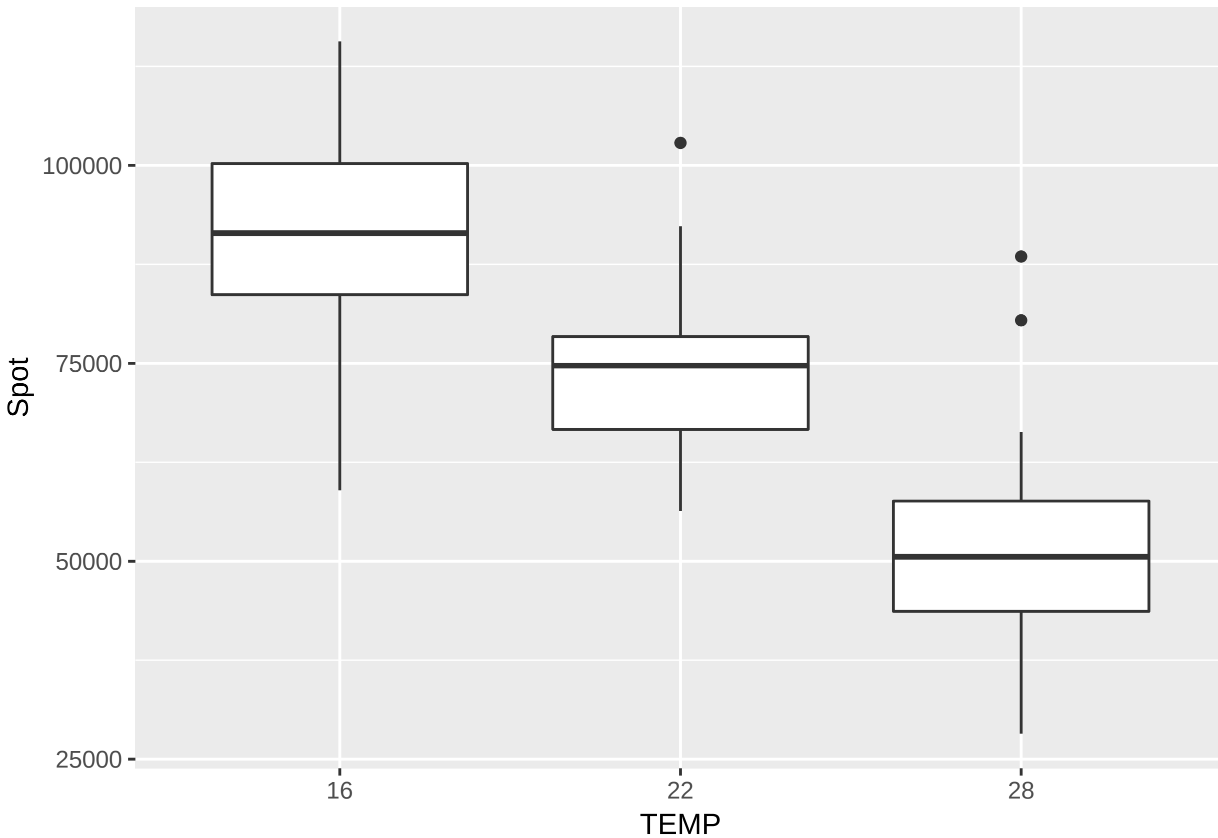
```r
Paris_idx <- which(spot.data$POP=='PARIS')
Sappo_idx <- which(spot.data$POP=='SAPPORO')
Sokol_idx <- which(spot.data$POP=='SOKOL')
temp16_idx <- which(spot.data$TEMP==16)
temp22_idx <- which(spot.data$TEMP==22)
temp28_idx <- which(spot.data$TEMP==28)
```

We start by exploring visualliy differences in variance among temperatures for each trait (with populations pooled):

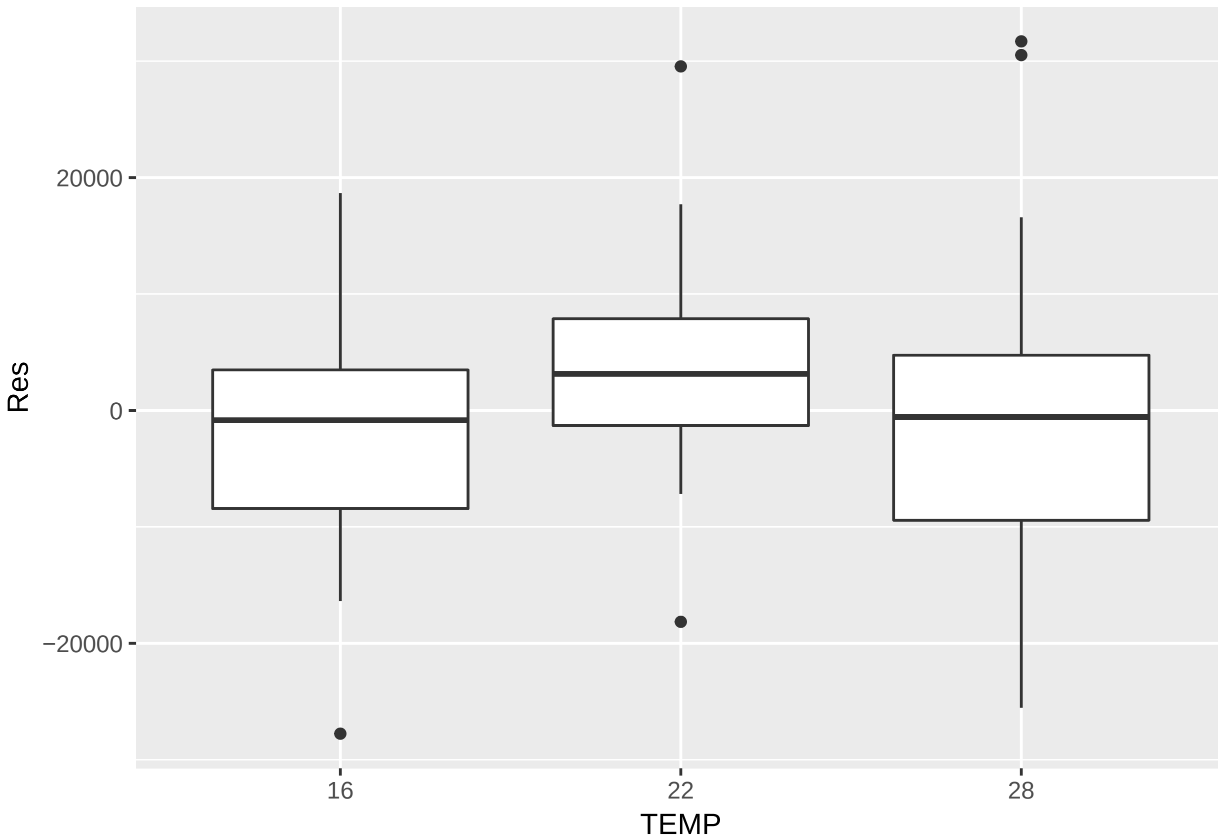
```r
ggplot(data = spot.data,
aes(x = TEMP, y = Wing)) +
geom_boxplot()
```



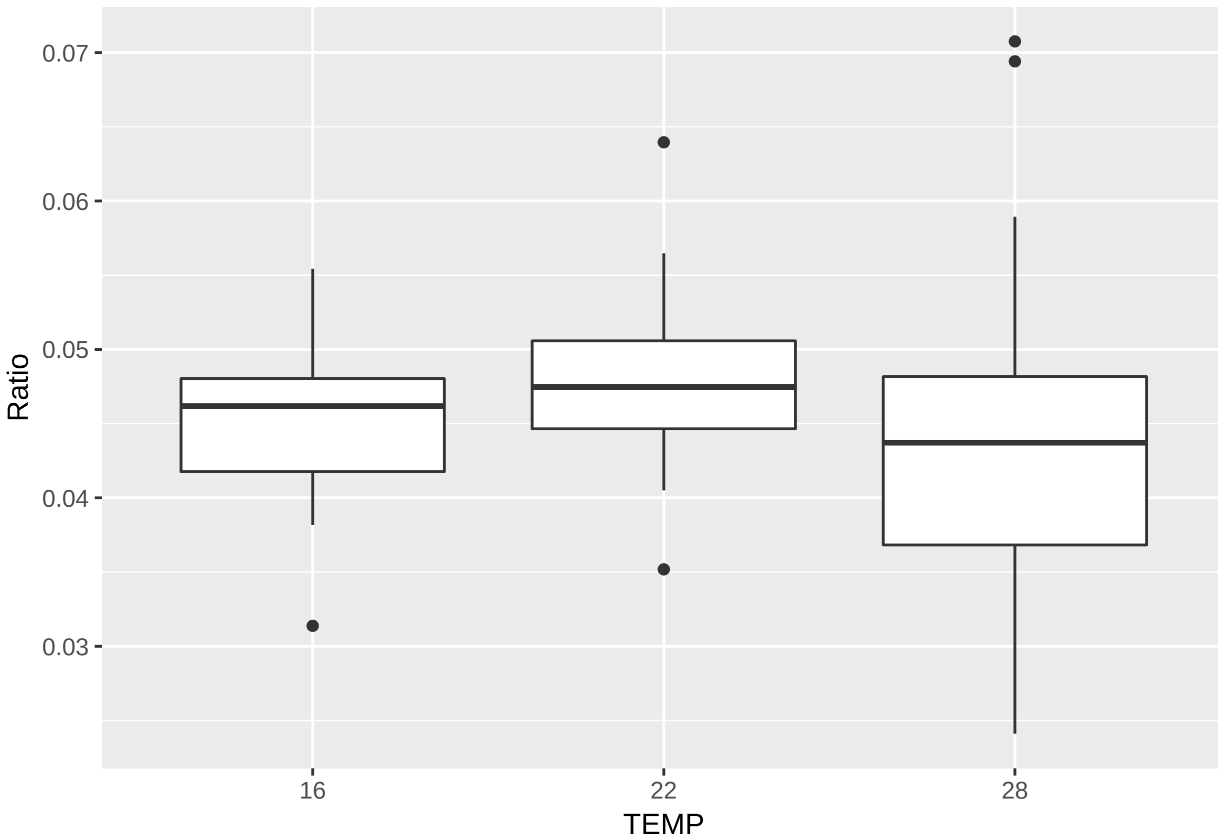

```r
ggplot(data = spot.data,
aes(x = TEMP, y = Spot)) +
geom_boxplot()
```




```r
ggplot(data = spot.data,
aes(x = TEMP, y = Res)) +
geom_boxplot()
```




```
ggplot(data = spot.data,
aes(x = TEMP, y = Ratio)) +
geom_boxplot()
```



Although if we get a sense of what's going on, it wouldn't be fair to compare variation among temperatures as such since we know temperature substiantially decrease size. Compare variation for such different sizes isn't ideal. For that, we are going to compare coefficients of variation, standardized by the mean.

So we are going to compare the coefficients of variation among temperatures (within population) for each phenotypic trait. Sorry if things get a bit messy.

So, for each geographic population we run a general MSLR test to check for differences in variation among temperatures for a trait. If the test comes up significant, we run three pairwise-tests to check between which temperatures there's a difference. We repeat this process for each phenotypic trait and then for each geographic population.

We start by looking at Sapporo population, where we didn't find variation differences among temperatures for any of the traits:

```
mslr_test(nr = 1000, x = spot.data$Wing[Sappo_idx],
          y = as.character(spot.data$TEMP[Sappo_idx]), seed = 2344)
```

```
## $MSLRT
## [1] 0.8279009
##
## $p_value
## [1] 0.6610337
```

```
sd(spot.data$Wing[intersect(temp16_idx,Sappo_idx)])/
  mean(spot.data$Wing[intersect(temp16_idx,Sappo_idx)])
```

```
## [1] 0.06429659
```

```
sd(spot.data$Wing[intersect(temp22_idx,Sappo_idx)])/
  mean(spot.data$Wing[intersect(temp22_idx,Sappo_idx)])
```

```
## [1] 0.05072137
```

```r
sd(spot.data$Wing[intersect(temp28_idx,Sappo_idx)])/
  mean(spot.data$Wing[intersect(temp28_idx,Sappo_idx)])
```

```
## [1] 0.06326157
```

```r
mslr_test(nr = 1000,x= spot.data$Spot[Sappo_idx],
          y= as.character(spot.data$TEMP[Sappo_idx]), seed = 982)
```

```
## $MSLRT
## [1] 4.248949
##
## $p_value
## [1] 0.1194957
```

```r
sd(spot.data$Spot[intersect(temp16_idx,Sappo_idx)])/
  mean(spot.data$Spot[intersect(temp16_idx,Sappo_idx)])
```

```
## [1] 0.1854069
```

```r
sd(spot.data$Spot[intersect(temp22_idx,Sappo_idx)])/
  mean(spot.data$Spot[intersect(temp22_idx,Sappo_idx)])
```

```
## [1] 0.1147414
```

```r
sd(spot.data$Spot[intersect(temp28_idx,Sappo_idx)])/
  mean(spot.data$Spot[intersect(temp28_idx,Sappo_idx)])
```

```
## [1] 0.2033109
```

```r
leveneTest(spot.data$Res[Sappo_idx], group = spot.data$TEMP[Sappo_idx])
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  2  1.0234 0.3686
##       40
```

```r
sd(spot.data$Res[intersect(temp16_idx,Sappo_idx)])
```

```
## [1] 12116.12
```

```r
sd(spot.data$Res[intersect(temp22_idx,Sappo_idx)])
```

```
## [1] 8376.84
```

```r
sd(spot.data$Res[intersect(temp28_idx,Sappo_idx)])
```

```
## [1] 9874.78
```

```r
mslr_test(nr = 1000, x = spot.data$Ratio[Sappo_idx],
          y = as.character(spot.data$TEMP[Sappo_idx]), seed = 3542)
```

```
## $MSLRT
## [1] 4.598332
##
## $p_value
## [1] 0.1003425
```

```r
sd(spot.data$Ratio[intersect(temp16_idx,Sappo_idx)])/
  mean(spot.data$Ratio[intersect(temp16_idx,Sappo_idx)])
```

```
## [1] 0.1436217
```

```r
sd(spot.data$Ratio[intersect(temp22_idx,Sappo_idx)])/
  mean(spot.data$Ratio[intersect(temp22_idx,Sappo_idx)])
```

```
## [1] 0.1079327
```

```r
sd(spot.data$Ratio[intersect(temp28_idx,Sappo_idx)])/
  mean(spot.data$Ratio[intersect(temp28_idx,Sappo_idx)])
```

```
## [1] 0.1953703
```

Now we can look at one invasive population, Sokol population:

```r
mslr_test(nr = 1000, x = spot.data$Wing[Sokol_idx],
          y = as.character(spot.data$TEMP[Sokol_idx]), seed = 4567)
```

```
## $MSLRT
## [1] 5.832362
##
## $p_value
## [1] 0.05414005
```

```r
mslr_test(nr = 1000, x = spot.data$Spot[Sokol_idx],
          y = as.character(spot.data$TEMP[Sokol_idx]), seed = 13)
```

```
## $MSLRT
## [1] 18.4557
##
## $p_value
## [1] 9.826435e-05
```

```r
leveneTest(spot.data$Res[Sokol_idx], group = spot.data$TEMP[Sokol_idx])
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value  Pr(>F)
## group  2  3.1658 0.05266 .
##       41
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
mslr_test(nr = 1000, x = spot.data$Ratio[Sokol_idx],
          y = as.character(spot.data$TEMP[Sokol_idx]), seed = 876)
```

```
## $MSLRT
## [1] 26.06361
##
## $p_value
## [1] 2.189573e-06
```

All traits but the residuals showed differences among temperatures. Let's start by looking at the wing:

```r
mslr_test(nr = 1000, x = spot.data$Wing[c(intersect(temp22_idx,Sokol_idx),
                                          intersect(temp16_idx,Sokol_idx))],
          y = as.character(spot.data$TEMP[c(intersect(temp22_idx,Sokol_idx),
                                            intersect(temp16_idx,Sokol_idx))]), seed = 98)
```

```
## $MSLRT
## [1] 1.835844
##
## $p_value
## [1] 0.1754389
```

```
mslr_test(nr = 1000, x = spot.data$Wing[c(intersect(temp28_idx,Sokol_idx),
                                          intersect(temp16_idx,Sokol_idx))],
          y = as.character(spot.data$TEMP[c(intersect(temp28_idx,Sokol_idx),
                                            intersect(temp16_idx,Sokol_idx))]), seed = 463)
```

```
## $MSLRT
## [1] 0.7375501
##
## $p_value
## [1] 0.3904468
```

```
mslr_test(nr = 1000, x = spot.data$Wing[c(intersect(temp28_idx,Sokol_idx),
                                          intersect(temp22_idx,Sokol_idx))],
          y = as.character(spot.data$TEMP[c(intersect(temp28_idx,Sokol_idx),
                                            intersect(temp22_idx,Sokol_idx))]), seed = 8789)
```

```
## $MSLRT
## [1] 5.307605
##
## $p_value
## [1] 0.02123252
```

```
sd(spot.data$Wing[intersect(temp16_idx,Sokol_idx)])/
  mean(spot.data$Wing[intersect(temp16_idx,Sokol_idx)])
```

```
## [1] 0.04892553
```

```
sd(spot.data$Wing[intersect(temp22_idx,Sokol_idx)])/
  mean(spot.data$Wing[intersect(temp22_idx,Sokol_idx)])
```

```
## [1] 0.07236896
```

```
sd(spot.data$Wing[intersect(temp28_idx,Sokol_idx)])/
  mean(spot.data$Wing[intersect(temp28_idx,Sokol_idx)])
```

```
## [1] 0.03810683
```

For the wing, extreme temperatures reduce variation but especially high temperature.

Now we look at the absolute spot size:

```
mslr_test(nr = 1000, x = spot.data$Spot[c(intersect(temp22_idx,Sokol_idx),
                                          intersect(temp16_idx,Sokol_idx))],
          y = as.character(spot.data$TEMP[c(intersect(temp22_idx,Sokol_idx),
                                            intersect(temp16_idx,Sokol_idx))]), seed = 982)
```

```
## $MSLRT
## [1] 0.05974309
##
## $p_value
## [1] 0.8069025
```

```
mslr_test(nr = 1000, x = spot.data$Spot[c(intersect(temp28_idx,Sokol_idx),
                                          intersect(temp16_idx,Sokol_idx))],
```

```r
         y = as.character(spot.data$TEMP[c(intersect(temp28_idx,Sokol_idx),
                                          intersect(temp16_idx,Sokol_idx))]), seed = 654)
```

```
## $MSLRT
## [1] 10.76825
##
## $p_value
## [1] 0.00103256
```

```r
mslr_test(nr = 1000, x = spot.data$Spot[c(intersect(temp28_idx,Sokol_idx),
                                          intersect(temp22_idx,Sokol_idx))],
         y = as.character(spot.data$TEMP[c(intersect(temp28_idx,Sokol_idx),
                                          intersect(temp22_idx,Sokol_idx))]), seed = 1248)
```

```
## $MSLRT
## [1] 12.01211
##
## $p_value
## [1] 0.0005285612
```

```r
sd(spot.data$Spot[intersect(temp16_idx,Sokol_idx)])/
  mean(spot.data$Spot[intersect(temp16_idx,Sokol_idx)])
```

```
## [1] 0.1022187
```

```r
sd(spot.data$Spot[intersect(temp22_idx,Sokol_idx)])/
  mean(spot.data$Spot[intersect(temp22_idx,Sokol_idx)])
```

```
## [1] 0.1038831
```

```r
sd(spot.data$Spot[intersect(temp28_idx,Sokol_idx)])/
  mean(spot.data$Spot[intersect(temp28_idx,Sokol_idx)])
```

```
## [1] 0.2858967
```

We find that the absolute spot size is much more variable at high temperature.

Now we look at the residual spot size:

```r
leveneTest(spot.data$Res[c(intersect(temp22_idx,Sokol_idx),
                                          intersect(temp16_idx,Sokol_idx))],
          group = spot.data$TEMP[c(intersect(temp22_idx,Sokol_idx),
                                          intersect(temp16_idx,Sokol_idx))])
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  1  0.6745 0.4192
##       25
```

```r
leveneTest(spot.data$Res[c(intersect(temp28_idx,Sokol_idx),
                                          intersect(temp16_idx,Sokol_idx))],
          group = spot.data$TEMP[c(intersect(temp28_idx,Sokol_idx),
                                          intersect(temp16_idx,Sokol_idx))])
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  1  2.0037 0.1683
##       27
```

```r
leveneTest(spot.data$Res[c(intersect(temp22_idx,Sokol_idx),
                                          intersect(temp28_idx,Sokol_idx))],
          group = spot.data$TEMP[c(intersect(temp22_idx,Sokol_idx),
                                            intersect(temp28_idx,Sokol_idx))])
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value  Pr(>F)
## group  1  5.5451 0.02527 *
##       30
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
sd(spot.data$Res[intersect(temp16_idx,Sokol_idx)])
```

```
## [1] 8809.549
```

```r
sd(spot.data$Res[intersect(temp22_idx,Sokol_idx)])
```

```
## [1] 5344.59
```

```r
sd(spot.data$Res[intersect(temp28_idx,Sokol_idx)])
```

```
## [1] 13819.08
```

And finally, the ratio:

```r
mslr_test(nr = 1000, x = spot.data$Ratio[c(intersect(temp22_idx,Sokol_idx),
                                          intersect(temp16_idx,Sokol_idx))],
          y = as.character(spot.data$TEMP[c(intersect(temp22_idx,Sokol_idx),
                                            intersect(temp16_idx,Sokol_idx))]), seed = 674)
```

```
## $MSLRT
## [1] 0.3043865
##
## $p_value
## [1] 0.5811454
```

```r
mslr_test(nr = 1000, x = spot.data$Ratio[c(intersect(temp28_idx,Sokol_idx),
                                          intersect(temp16_idx,Sokol_idx))],
          y = as.character(spot.data$TEMP[c(intersect(temp28_idx,Sokol_idx),
                                            intersect(temp16_idx,Sokol_idx))]), seed = 74)
```

```
## $MSLRT
## [1] 10.16051
##
## $p_value
## [1] 0.001434809
```

```r
mslr_test(nr = 1000, x = spot.data$Ratio[c(intersect(temp28_idx,Sokol_idx),
                                          intersect(temp22_idx,Sokol_idx))],
          y = as.character(spot.data$TEMP[c(intersect(temp28_idx,Sokol_idx),
                                            intersect(temp22_idx,Sokol_idx))]), seed = 2874)
```

```
## $MSLRT
## [1] 18.56054
##
## $p_value
## [1] 1.645929e-05
```

```r
sd(spot.data$Ratio[intersect(temp16_idx,Sokol_idx)])/
  mean(spot.data$Ratio[intersect(temp16_idx,Sokol_idx)])
```

```
## [1] 0.08923806
```

```r
sd(spot.data$Ratio[intersect(temp22_idx,Sokol_idx)])/
  mean(spot.data$Ratio[intersect(temp22_idx,Sokol_idx)])
```

```
## [1] 0.07553107
```

```r
sd(spot.data$Ratio[intersect(temp28_idx,Sokol_idx)])/
  mean(spot.data$Ratio[intersect(temp28_idx,Sokol_idx)])
```

```
## [1] 0.2721465
```

For the relative spot size, we find the same pattern than for the absolute spot size: at high temperature variation is much larger.

Finally we check for the Paris population, where we find differences for all traits but the wing:

```r
mslr_test(nr = 1000, x = spot.data$Wing[Paris_idx],
          y = as.character(spot.data$TEMP[Paris_idx]), seed = 3)
```

```
## $MSLRT
## [1] 3.102327
##
## $p_value
## [1] 0.2120012
```

```r
mslr_test(nr = 1000, x = spot.data$Spot[Paris_idx],
          y= as.character(spot.data$TEMP[Paris_idx]), seed = 357)
```

```
## $MSLRT
## [1] 7.063888
##
## $p_value
## [1] 0.029248
```

```r
leveneTest(spot.data$Res[Paris_idx], group = spot.data$TEMP[Paris_idx])
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  2  1.2449 0.3004
##       35
```

```r
sd(spot.data$Res[intersect(temp28_idx,Paris_idx)])
```

```
## [1] 12935.25
```

```r
sd(spot.data$Res[intersect(temp22_idx,Paris_idx)])
```

```
## [1] 8990.495
```

```r
sd(spot.data$Res[intersect(temp16_idx,Paris_idx)])
```

```
## [1] 7207.257
```

```r
mslr_test(nr = 1000, x = spot.data$Ratio[Paris_idx],
          y = as.character(spot.data$TEMP[Paris_idx]), seed = 982)
```

```
## $MSLRT
```

```
## [1] 12.19957
##
## $p_value
## [1] 0.002243351
```

We start by looking at the wing variation:

```
sd(spot.data$Wing[intersect(temp28_idx,Paris_idx)])/
  mean(spot.data$Wing[intersect(temp28_idx,Paris_idx)])
```

```
## [1] 0.04384594
```

```
sd(spot.data$Wing[intersect(temp22_idx,Paris_idx)])/
  mean(spot.data$Wing[intersect(temp22_idx,Paris_idx)])
```

```
## [1] 0.06443954
```

```
sd(spot.data$Wing[intersect(temp16_idx,Paris_idx)])/
  mean(spot.data$Wing[intersect(temp16_idx,Paris_idx)])
```

```
## [1] 0.04056821
```

Here again, as for Sokol, the wing is less variable at extreme temperatures.

For the absolute spot size:

```
mslr_test(nr = 1000, x = spot.data$Spot[c(intersect(temp22_idx,Paris_idx),
                                          intersect(temp16_idx,Paris_idx))],
          y = as.character(spot.data$TEMP[c(intersect(temp22_idx,Paris_idx),
                                            intersect(temp16_idx,Paris_idx))]), seed = 7605)
```

```
## $MSLRT
## [1] 1.008613
##
## $p_value
## [1] 0.3152352
```

```
mslr_test(nr = 1000, x = spot.data$Spot[c(intersect(temp28_idx,Paris_idx),
                                          intersect(temp16_idx,Paris_idx))],
          y = as.character(spot.data$TEMP[c(intersect(temp28_idx,Paris_idx),
                                            intersect(temp16_idx,Paris_idx))]), seed = 234)
```

```
## $MSLRT
## [1] 5.126877
##
## $p_value
## [1] 0.02355807
```

```
mslr_test(nr = 1000, x = spot.data$Spot[c(intersect(temp28_idx,Paris_idx),
                                          intersect(temp22_idx,Paris_idx))],
          y = as.character(spot.data$TEMP[c(intersect(temp28_idx,Paris_idx),
                                            intersect(temp22_idx,Paris_idx))]), seed = 4562)
```

```
## $MSLRT
## [1] 3.424738
##
## $p_value
## [1] 0.06422642
```

```r
sd(spot.data$Spot[intersect(temp28_idx,Paris_idx)])/
  mean(spot.data$Spot[intersect(temp28_idx,Paris_idx)])
```

```
## [1] 0.247059
```

```r
sd(spot.data$Spot[intersect(temp22_idx,Paris_idx)])/
  mean(spot.data$Spot[intersect(temp22_idx,Paris_idx)])
```

```
## [1] 0.1426551
```

```r
sd(spot.data$Spot[intersect(temp16_idx,Paris_idx)])/
  mean(spot.data$Spot[intersect(temp16_idx,Paris_idx)])
```

```
## [1] 0.1035649
```

Here also as for Sokol, the absolute spot size of the Paris population at 28º is much more variable.

Finally, for the relative spot size:

```r
mslr_test(nr = 1000, x = spot.data$Ratio[c(intersect(temp22_idx,Paris_idx),
                                           intersect(temp16_idx,Paris_idx))],
          y = as.character(spot.data$TEMP[c(intersect(temp22_idx,Paris_idx),
                                             intersect(temp16_idx,Paris_idx))]), seed = 1204)
```

```
## $MSLRT
## [1] 1.03837
##
## $p_value
## [1] 0.3082008
```

```r
mslr_test(nr = 1000, x = spot.data$Ratio[c(intersect(temp28_idx,Paris_idx),
                                           intersect(temp16_idx,Paris_idx))],
          y = as.character(spot.data$TEMP[c(intersect(temp28_idx,Paris_idx),
                                             intersect(temp16_idx,Paris_idx))]), seed = 4501)
```

```
## $MSLRT
## [1] 8.699608
##
## $p_value
## [1] 0.003182784
```

```r
mslr_test(nr = 1000, x = spot.data$Ratio[c(intersect(temp28_idx,Paris_idx),
                                           intersect(temp22_idx,Paris_idx))],
          y = as.character(spot.data$TEMP[c(intersect(temp28_idx,Paris_idx),
                                             intersect(temp22_idx,Paris_idx))]), seed = 652)
```

```
## $MSLRT
## [1] 7.719397
##
## $p_value
## [1] 0.005463063
```

```r
sd(spot.data$Ratio[intersect(temp28_idx,Paris_idx)])/
  mean(spot.data$Ratio[intersect(temp28_idx,Paris_idx)])
```

```
## [1] 0.249608
```

```r
sd(spot.data$Ratio[intersect(temp22_idx,Paris_idx)])/
  mean(spot.data$Ratio[intersect(temp22_idx,Paris_idx)])
```

```
## [1] 0.1151799
```

```r
sd(spot.data$Ratio[intersect(temp16_idx,Paris_idx)])/
  mean(spot.data$Ratio[intersect(temp16_idx,Paris_idx)])
```

```
## [1] 0.08280693
```

Finally, here also the Paris population is alike to Sokol population and the results are similar to the absolute spot size: the relative spot size is much larger at 28?.

In summary:

Table 1: Sapporo population

	16º	22º	28º
Spot	ND	ND	ND
Wing	ND	ND	ND
Resid	ND	ND	ND
Ratio	ND	ND	ND

Table 2: Sokol population

	16º	22º	28º
Spot	-	-	+
Wing	-	+	-
Resid	-	-	+
Ratio	-	-	+

Table 3: Paris population

	16º	22º	28º
Spot	-	-	+
Wing	-	+	-
Resid	ND	ND	ND
Ratio	-	-	+

While invasive populations only show larger variation in their spots at high temperature, the native population also showed increased variation at low temperature.