Assignment 2

Sanjiv Kumar

11/7/2021

Part 1. One-way ANOVA

Introduction

Lime disease in human is caused by the bacterial spirochete Borrelia (*Borrelia burgdorferi*) which is transmitted by the bite of hard tick from the genus *Ixodes*. The present dataset (Ref. 1) is collected wherein ticks were sampled from three different islands (Askö (ASK), Torö (TOR) and Öja(OJA)) in Sweden. Prevalence of the *Borrelia* infected ticks was quantified at each sampling site.

Method

Since, the prevalence of *Borrelia* infection is presented in the form of percentages (%), the first step is to transform the data using Arcsine transformation. Prevalence of *Borrelia* infection was examined at different island. One-way ANOVA (`lm`) models were generated to see the difference of various factors in the prevalence of infection. The data was normalized with `sqrt` transformation and results were interpreted.

```
# Arcsine data transformation
borrelia$borrelia.prev.arc <- asin(sqrt(borrelia$Borrelia_prevalence/100))
```

Result

Prevalence of *Borrelia* infection

Following results shows the prevalence of *Borrelia* infection was examined at different island.

Prevalence of Borrelia infection in ticks at different Island
Measured central tendencies

Island	variable	n	mean	sd
ASK	borrelia.prev.arc	30	0.810	0.252
OJA	borrelia.prev.arc	30	0.312	0.119
TOR	borrelia.prev.arc	30	0.676	0.218

The data was visualized using boxplot.

Prevalence of Borrelia infection in ticks at different Islands
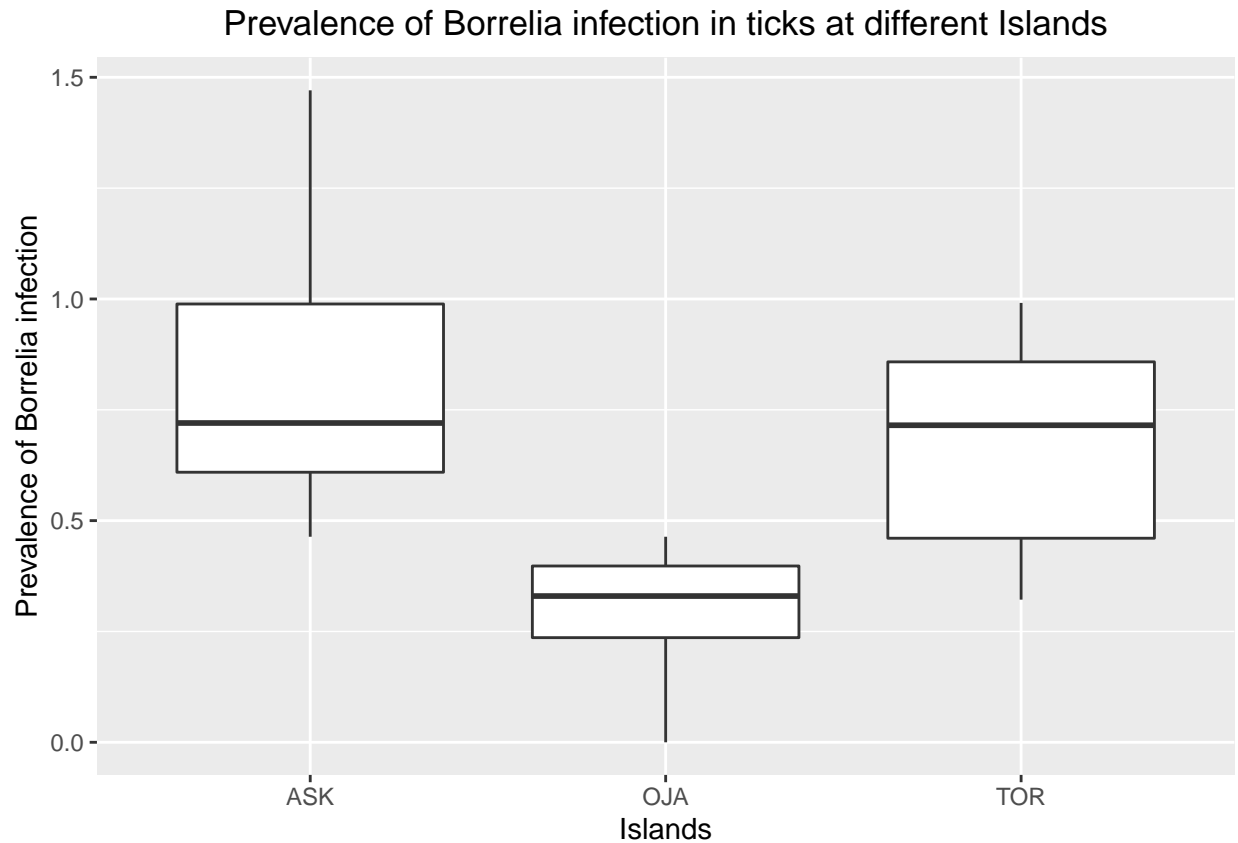
One-way ANOVA

Though, from the `Q-Q plot`, the residuals appear to be on straight line, but `Residuals vs Fitted` graphs shows spread in the largest group more than three times the spread in the smallest group, the data needs to be transformed.

```
##
## Call:
## lm(formula = borrelia.prev.arc ~ Island, data = borrelia)
##
## Residuals:
##      Min      1Q  Median      3Q     Max
## -0.35446 -0.14036  0.00965  0.12892  0.66098
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.80965    0.03730  21.706  < 2e-16 ***
## IslandOJA   -0.49755    0.05275  -9.432 5.86e-15 ***
## IslandTOR   -0.13344    0.05275  -2.530   0.0132 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2043 on 87 degrees of freedom
## Multiple R-squared:  0.5229, Adjusted R-squared:  0.5119
## F-statistic: 47.67 on 2 and 87 DF,  p-value: 1.051e-14
```

Residuals vs Fitted

Normal Q–Q

Scale–Location

Constant Leverage:
Residuals vs Factor Levels
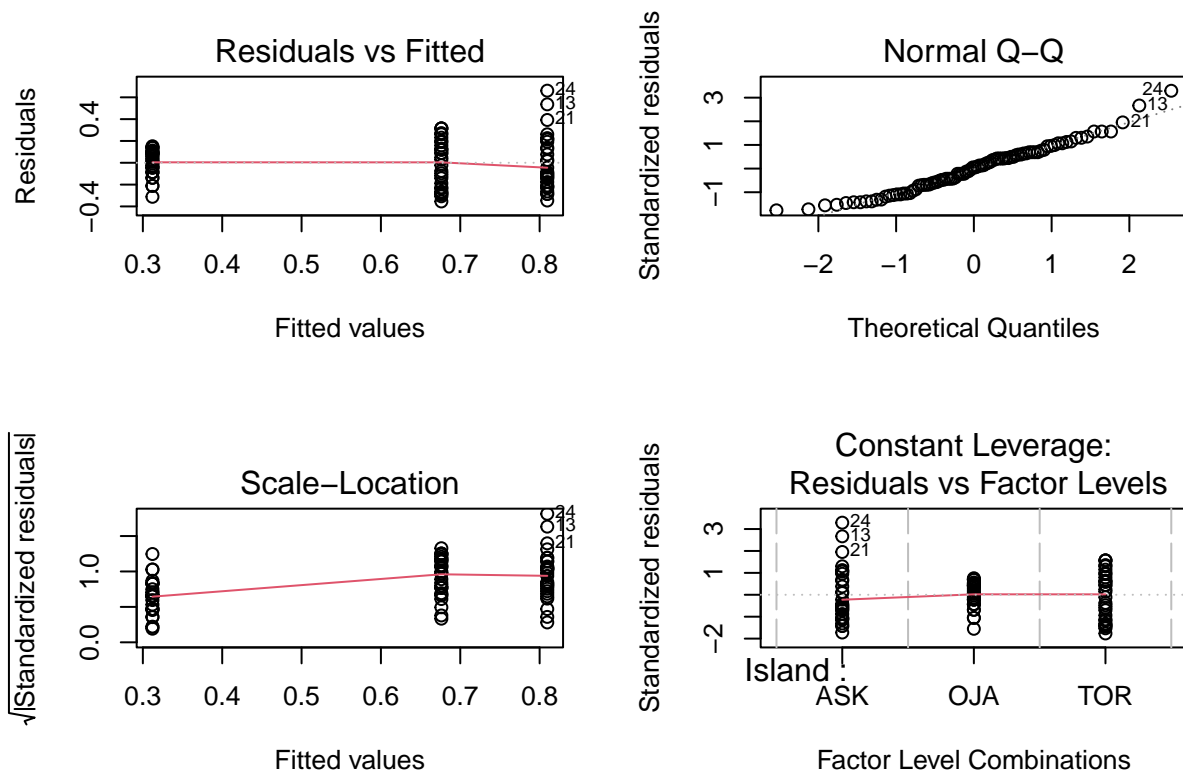
```
## Anova Table (Type II tests)
##
## Response: borrelia.prev.arc
##           Sum Sq Df F value   Pr(>F)
## Island    3.9794  2  47.666 1.051e-14 ***
## Residuals 3.6316 87
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```
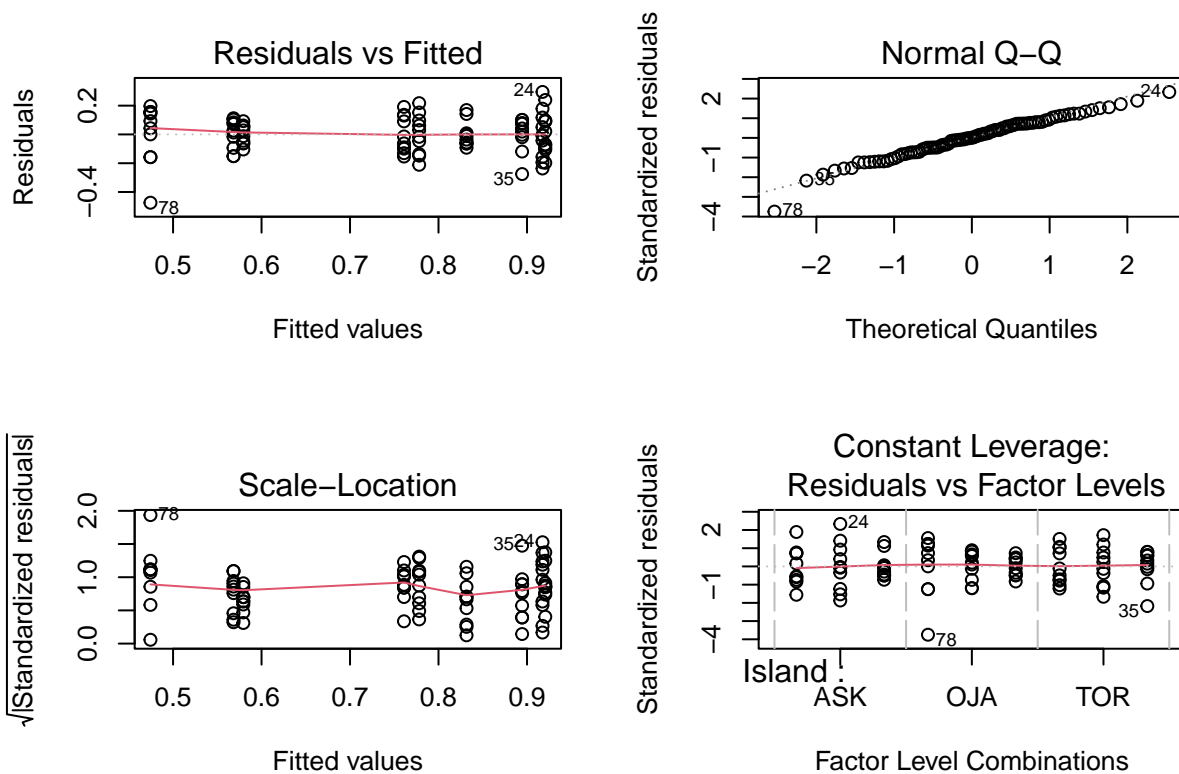
Stats Report

- Test: One-way ANOVA
- F-value: 47.6664057
- df: 87
- p-value: $1.0506727 \times 10^{-14}$

Sqrt transformation Because of prevalence of no infection (0, zero), the data was transformed using the `sqrt` transformation, which normalizes the `Residuals vs Fitted` graph.

```
##
## Call:
## lm(formula = sqrt(borrelia.prev.arc) ~ Island * Habitat, data = borrelia)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.47447 -0.08580  0.00119  0.09199  0.29515
##
## Coefficients:
```

```
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)               0.920650   0.042194  21.819  < 2e-16 ***
## IslandOJA                -0.446182   0.059672  -7.477 7.97e-11 ***
## IslandTOR                -0.159651   0.059672  -2.675  0.00903 **
## HabitatForest            -0.003109   0.059672  -0.052  0.95858
## HabitatMeadow            -0.088997   0.059672  -1.491  0.13973
## IslandOJA:HabitatForest   0.096985   0.084389   1.149  0.25383
## IslandTOR:HabitatForest   0.020248   0.084389   0.240  0.81098
## IslandOJA:HabitatMeadow   0.193950   0.084389   2.298  0.02412 *
## IslandTOR:HabitatMeadow   0.222462   0.084389   2.636  0.01005 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1334 on 81 degrees of freedom
## Multiple R-squared:  0.6079, Adjusted R-squared:  0.5692
## F-statistic:  15.7 on 8 and 81 DF,  p-value: 1.025e-13
```



```
## Anova Table (Type II tests)
##
## Response: sqrt(borrelia.prev.arc)
##               Sum Sq Df F value    Pr(>F)
## Island        2.01291  2 56.5306 4.287e-16 ***
## Habitat       0.03966  2  1.1138   0.33328
## Island:Habitat 0.18333  4  2.5743   0.04373 *
## Residuals     1.44210 81
## ---
```
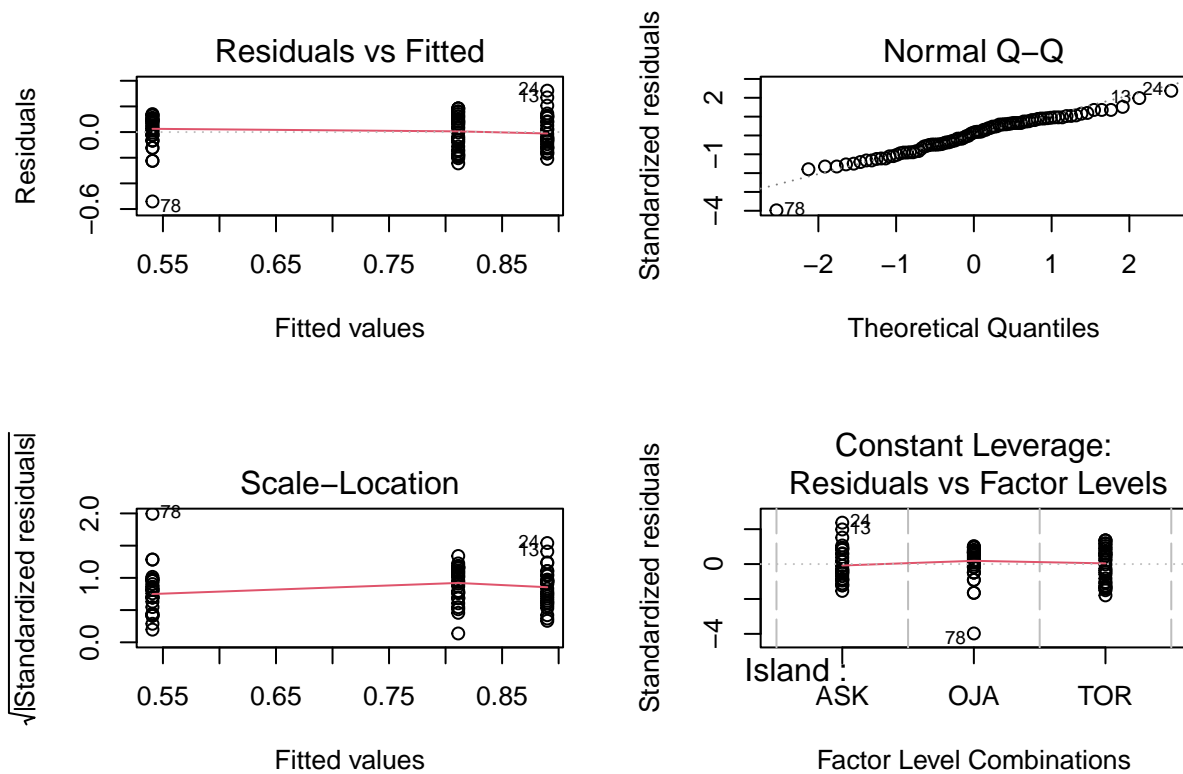
4

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Stats Report

- Test: One-way ANOVA
- F-value: 56.5305913
- df: 2
- p-value: $4.2867385 \times 10^{-16}$

Interpretation of results   Anova table shows that the there was a significant different in `Borrelia` infections at various Islands. Therefore, `lm` model was generated for the `Island` using `sqrt` transformation. Both `Q-Q plot` and `Residuals vs Fitted` looks good.

```
##
## Call:
## lm(formula = sqrt(borrelia.prev.arc) ~ Island, data = borrelia)
##
## Residuals:
##      Min      1Q   Median      3Q      Max
## -0.54074 -0.09066  0.02134  0.10075  0.32275
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.88995    0.02526  35.234  < 2e-16 ***
## IslandOJA   -0.34920    0.03572  -9.776 1.16e-15 ***
## IslandTOR   -0.07875    0.03572  -2.205   0.0301 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1383 on 87 degrees of freedom
## Multiple R-squared:  0.5473, Adjusted R-squared:  0.5369
## F-statistic: 52.59 on 2 and 87 DF,  p-value: 1.068e-15
```

Residuals vs Fitted

Normal Q–Q

Scale–Location

Constant Leverage:
Residuals vs Factor Levels

Stats Report

- Test: One-way ANOVA
- F-value: 52.5867633
- df: 87
- p-value: $1.0676223 \times 10^{-15}$

Since we used One-way ANOVA, pairwise comparison was made using post hoc test, `glth` from `multicomp` package.

```
## Anova Table (Type II tests)
##
## Response: sqrt(borrelia.prev.arc)
##           Sum Sq Df F value    Pr(>F)
## Island    2.0129  2  52.587 1.068e-15 ***
## Residuals 1.6651 87
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## Analysis of Variance Table
##
## Response: borrelia.prev.arc
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Island     2 3.9794 1.98970  47.666 1.051e-14 ***
## Residuals 87 3.6316 0.04174
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##
```

```
##    Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lm(formula = sqrt(borrelia.prev.arc) ~ Island, data = borrelia)
##
## Linear Hypotheses:
##                Estimate Std. Error t value Pr(>|t|)
## OJA - ASK == 0 -0.34920    0.03572  -9.776   <1e-04 ***
## TOR - ASK == 0 -0.07875    0.03572  -2.205    0.076 .
## TOR - OJA == 0  0.27046    0.03572   7.572   <1e-04 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

The results shows that there is significant difference between prevalence of *Borrelia* infection between Askö (ASK) - Öja(OJA) and Torö (TOR) - Öja(OJA) but the difference was not significant between Askö (ASK) - Torö (TOR). The table above show the significant differences among Island groups.

Plot of Means
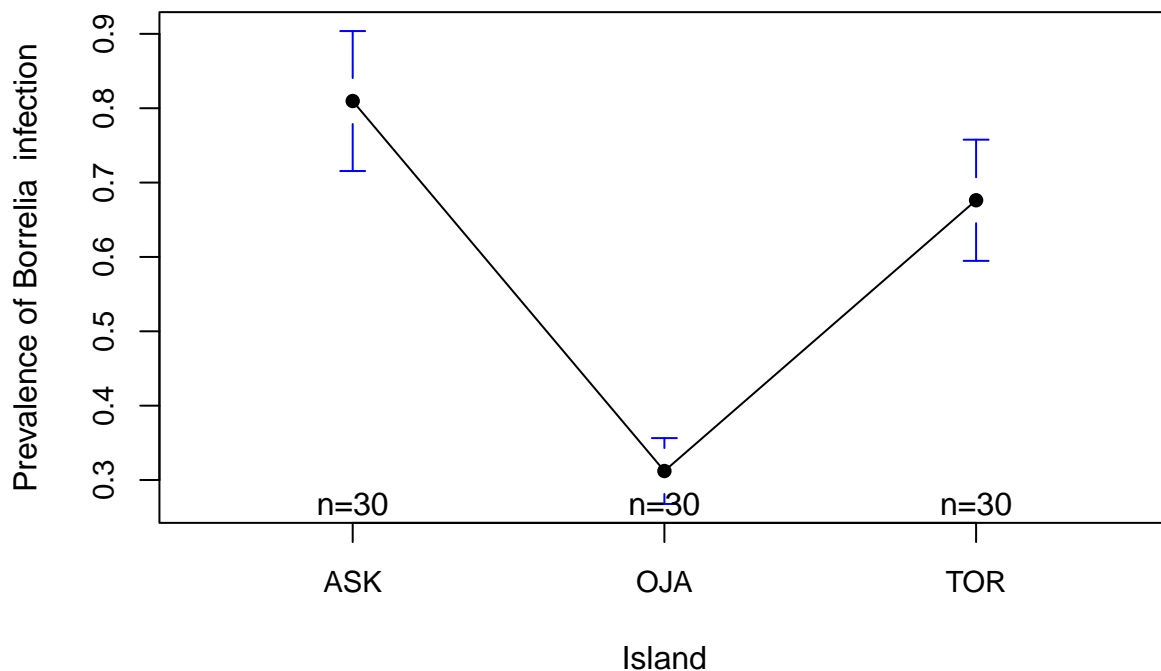
Discussion

Analysis of the data shows that the prevalence of *Borrelia* infection in ticks is significantly influenced by the location (i.e. Island) but not significantly affected by the habitat. It was also concluded that the difference between OJA - ASK and TOR - OJA were significant with p-value < 0.001, as compare to the TOR - ASK, which was non-significant.

References

1. borrelia.txt, data provided with the assignment.

Part 2. Two-way ANOVA with interaction

Introduction

Additional parameters i.e. soil pH and temperature was also included in the dataset at each site. Here, we are analyzing how Temperature and location influences the prevalence of infections in ticks. Here we convert Temperature data into factors to check the influence of low, medium, high and very high temperature on prevalence of *Borrelia* infection in comparision to the location.

Method

Since, the prevalence of *Borrelia* infection is presented in the form of percentages (%), the first step is to transform the data using Arcsine transformation. Both pH and Temperature data is converted into three equal factors (i.e. Low, Medium and High).

```
# head(borrelia)
# pH_fact = cut(borrelia$PH, 4, labels=c('pH_Low', 'pH_Medium', 'pH_High', 'ph_VHigh'))
# borrelia$pH_fact = pH_fact

# head(borrelia)
temp_fact = cut(borrelia$Temp, 3, labels=c('temp_Low', 'temp_Medium', 'temp_High'))
borrelia$temp_fact = temp_fact
# head(borrelia)
```

Following results shows the influence of pH and temperature on the prevalence of *Borrelia* infection.

Prevalence of Borrelia (effect of Temperature and Island)
Measured central tendencies

Island	temp_fact	variable	n	mean	sd
ASK	temp_Low	borrelia.prev.arc	11	0.720	0.165
OJA	temp_Low	borrelia.prev.arc	10	0.318	0.134
TOR	temp_Low	borrelia.prev.arc	16	0.673	0.236
ASK	temp_Medium	borrelia.prev.arc	5	1.014	0.383
OJA	temp_Medium	borrelia.prev.arc	5	0.319	0.111
TOR	temp_Medium	borrelia.prev.arc	3	0.684	0.100
ASK	temp_High	borrelia.prev.arc	14	0.807	0.230
OJA	temp_High	borrelia.prev.arc	15	0.306	0.118
TOR	temp_High	borrelia.prev.arc	11	0.679	0.229

The data was visualized using boxplot.

## Prevalence of Borrelia (effect of Temperature and Island)
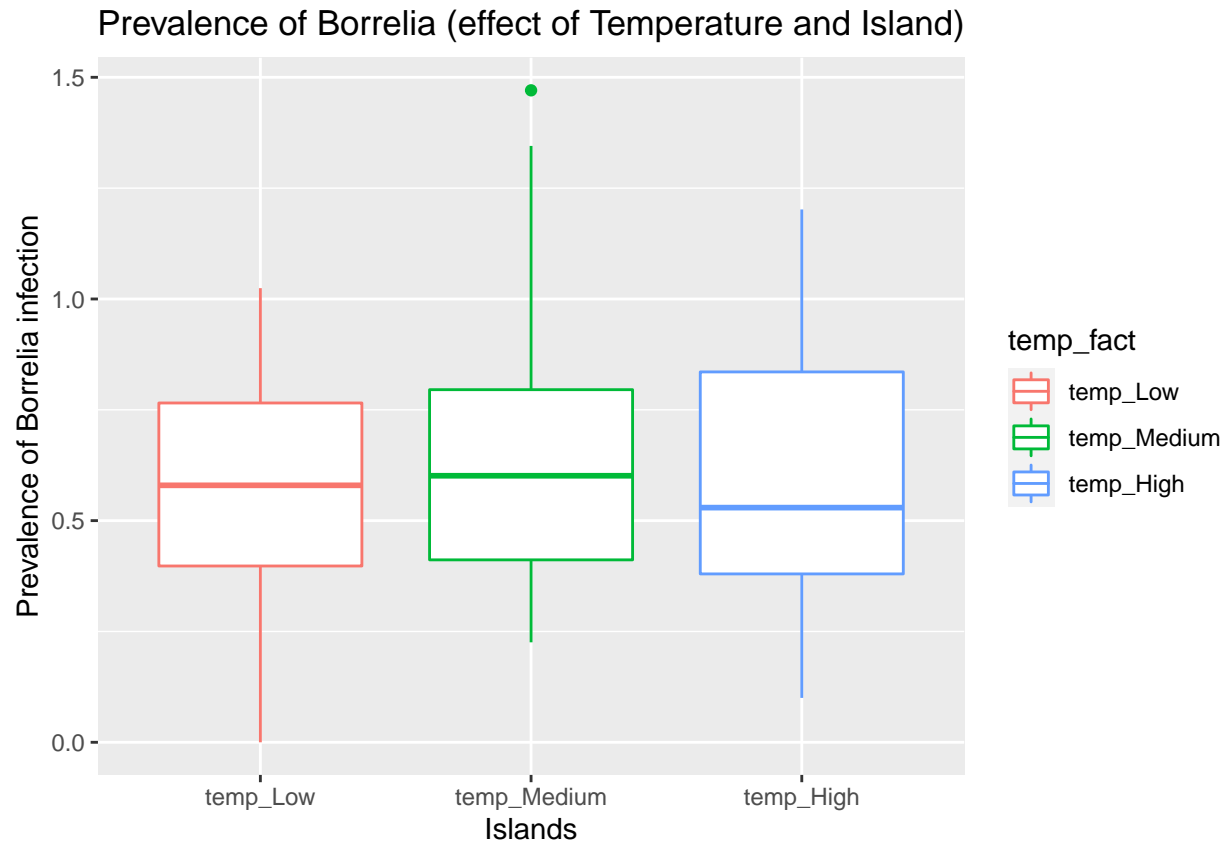
Two-way ANOVA

Two-way ANOVA was performed for prevalence of *Borrelia* infection against temperature and location (i.e. Island). `lm` model was generated for the `temp_fact` and `Island`. Both `Q-Q plot` and `Residuals vs Fitted` looks good.

```
##
## Call:
## lm(formula = borrelia.prev.arc ~ temp_fact * Island, data = borrelia)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.42338 -0.13234  0.00332  0.12974  0.45675
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    0.72008    0.06116  11.773  < 2e-16 ***
## temp_facttemp_Medium           0.29380    0.10941   2.685  0.00879 **
## temp_facttemp_High             0.08700    0.08173   1.065  0.29026
## IslandOJA                     -0.40165    0.08863  -4.532    2e-05 ***
## IslandTOR                     -0.04692    0.07945  -0.591  0.55648
## temp_facttemp_Medium:IslandOJA -0.29344   0.15593  -1.882  0.06345 .
## temp_facttemp_High:IslandOJA  -0.09979    0.11635  -0.858  0.39363
## temp_facttemp_Medium:IslandTOR -0.28345   0.16810  -1.686  0.09562 .
## temp_facttemp_High:IslandTOR  -0.08151    0.11398  -0.715  0.47658
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2029 on 81 degrees of freedom
## Multiple R-squared:  0.5621, Adjusted R-squared:  0.5188
## F-statistic:    13 on 8 and 81 DF,  p-value: 7.204e-12
```
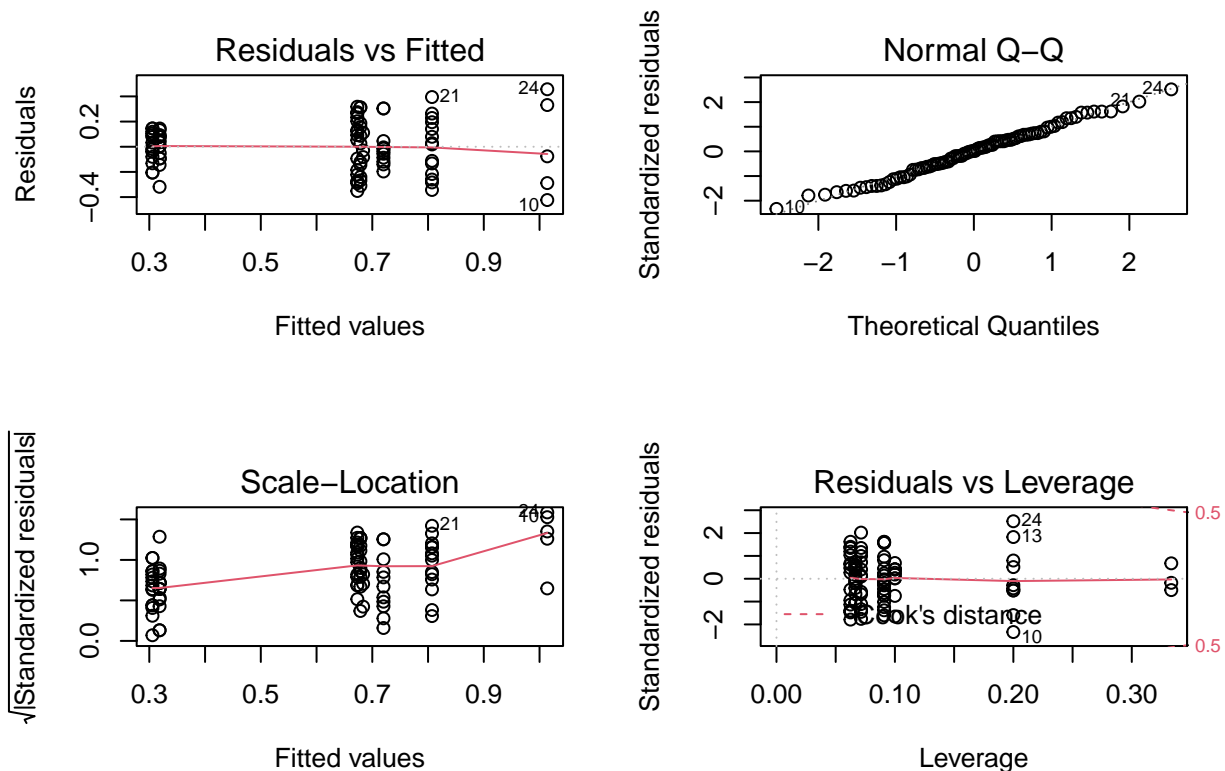

Stats Report

- Test: Two-way ANOVA
- F-value: 0.9434006
- df: 2
- p-value: 0.3935368

Pairwise comparison was made using post hoc test, `glth` from `multicomp` package.

```
## Anova Table (Type II tests)
##
## Response: borrelia.prev.arc
##                 Sum Sq Df F value    Pr(>F)
## temp_fact        0.1170  2  1.4214    0.2473
## Island           4.0187  2 48.8318 1.219e-14 ***
## temp_fact:Island 0.1815  4  1.1029    0.3609
## Residuals        3.3330 81
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## Analysis of Variance Table
##
```

```
## Model 1: borrelia.prev.arc ~ Island
## Model 2: borrelia.prev.arc ~ temp_fact * Island
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     87 3.6316
## 2     81 3.3330  6   0.29852 1.2091 0.3102
```

post-hoc

```
summary(glht(lm_borrelia_temp_Island, linfct = mcp(Island = "Tukey")))
```

```
## Warning in mcp2matrix(model, linfct = linfct): covariate interactions found --
## default contrast might be inappropriate
```

```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lm(formula = borrelia.prev.arc ~ temp_fact * Island, data = borrelia)
##
## Linear Hypotheses:
##               Estimate Std. Error t value Pr(>|t|)
## OJA - ASK == 0 -0.40165    0.08863  -4.532  < 1e-04 ***
## TOR - ASK == 0 -0.04692    0.07945  -0.591 0.825197
## TOR - OJA == 0  0.35473    0.08177   4.338 0.000121 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

There difference was only found among various Islands, the effect of temperature was not significant.

for Temperature

```
# for Temperature
testInteractions(lm_borrelia_temp_Island, pairwise="temp_fact", adjustment="holm")
```

```
## F Test:
## P-value adjustment method: holm
##                          Value Df Sum of Sq      F Pr(>F)
##   temp_Low-temp_Medium -0.101505  1    0.0940 2.2838 0.4039
##      temp_Low-temp_High -0.026571  1    0.0132 0.3201 0.5731
## temp_Medium-temp_High  0.074934  1    0.0525 1.2762 0.5239
## Residuals                        81    3.3330
```
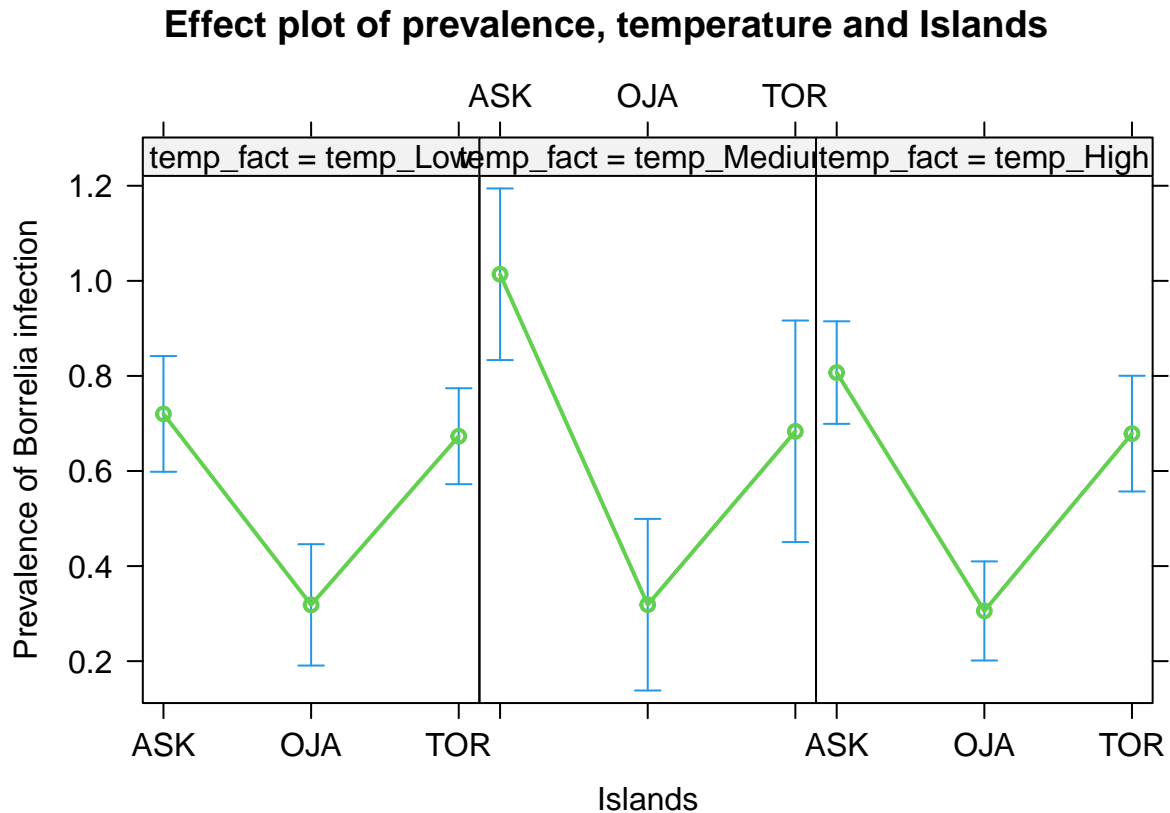
for Island

```
testInteractions(lm_borrelia_temp_Island, pairwise="Island", adjustment="holm")
```

```
## F Test:
## P-value adjustment method: holm
##            Value Df Sum of Sq      F     Pr(>F)
## ASK-OJA   0.53273  1    3.5036 85.146 8.465e-14 ***
## ASK-TOR   0.16857  1    0.3012  7.320  0.008312 **
## OJA-TOR  -0.36415  1    1.3985 33.986 2.177e-07 ***
```

```
## Residuals          81    3.3330
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Effect plot

Effect plot of prevalence, temperature and Islands

Discussion

Analysis of the data shows that the prevalence of *Borrelia* infection in ticks is significantly influenced by the location (i.e. Island) but not significantly affected by the temperature.

References

1. borrelia.txt, data provided with the assignment.