**BIFX 503: Statistics for Bioinformatics**

**Homework Set #9**

**Due: November 14, 2018**

*Instructions:*

*Use R to complete this assignment. Assignment is to be submitted via Blackboard.*

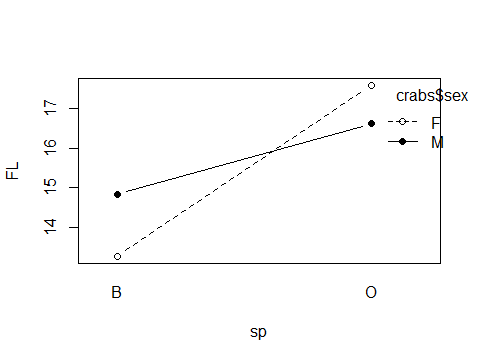
The dataset **crabs** (from the MASS package) contains 5 morphological measurements (in mm) on 200 crabs (*Leptograpsus variegatus*): 50 for each combination of color (blue or orange) and sex (male or female).

Use analysis of variance to determine how each characteristic varies by color and sex.

1. Generate **interaction plots** to visualize how each measurement varies by color and sex. Do the graphs suggest that interaction might be present for any of the variables? What feature of an interaction plot indicates that interaction could be present?

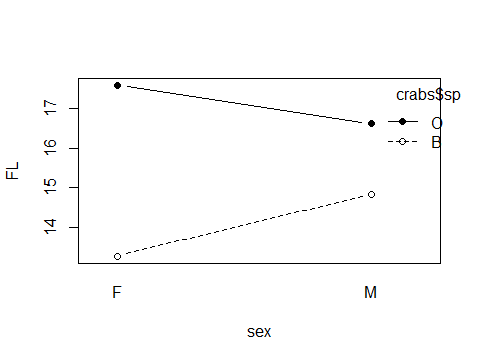
If the two lines cross each other then it means there is a severe interaction between each other. If the lines different from each other or not parallel to each other then it means they have moderate interaction.

interaction.plot(x.factor=crabs$sp, trace.factor=crabs$sex, response=crabs$FL, fun="mean", type="b", ylab="FL", xlab="sp", pch=c(1,19))



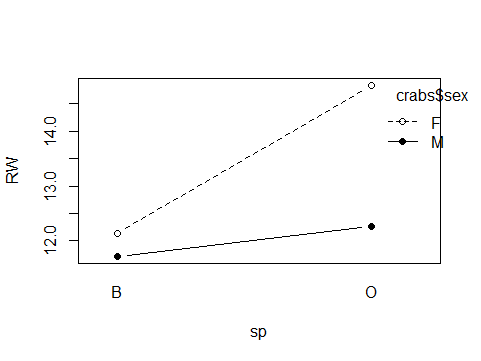
Since two lines cross with each other therefore interaction present between them.

interaction.plot(x.factor=crabs$sex, trace.factor=crabs$sp, response=crabs$FL, fun="mean", type="b", ylab="FL", xlab="sex", pch=c(1,19))



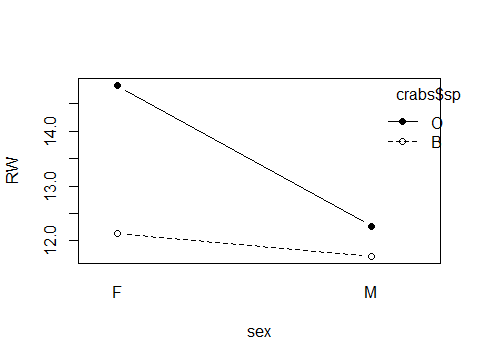
No interaction present between them since the lines are parallel.

interaction.plot(x.factor=crabs$sp, trace.factor=crabs$sex, response=crabs$RW, fun="mean", type="b", ylab="RW", xlab="sp", pch=c(1,19))



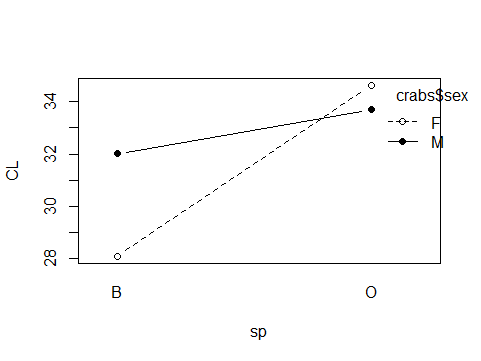
Since the lines are not parallel so moderate interaction present here.

interaction.plot(x.factor=crabs$sex, trace.factor=crabs$sp, response=crabs$RW, fun="mean", type="b", ylab="RW", xlab="sex", pch=c(1,19))



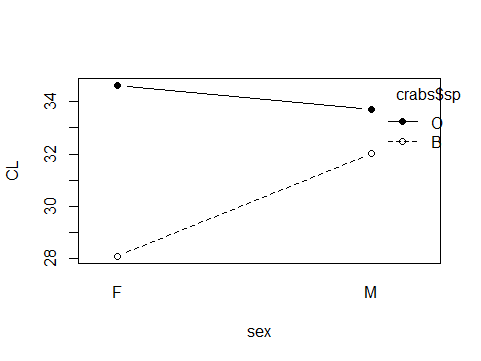
Since the lines are not parallel so moderate interaction present here.

interaction.plot(x.factor=crabs$sp, trace.factor=crabs$sex, response=crabs$CL, fun="mean", type="b", ylab="CL", xlab="sp", pch=c(1,19))



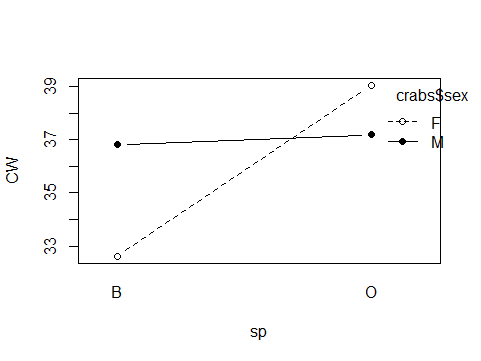
Since the lines cross each other so interaction present in this case.

interaction.plot(x.factor=crabs$sex, trace.factor=crabs$sp, response=crabs$CL, fun="mean", type="b", ylab="CL", xlab="sex", pch=c(1,19))



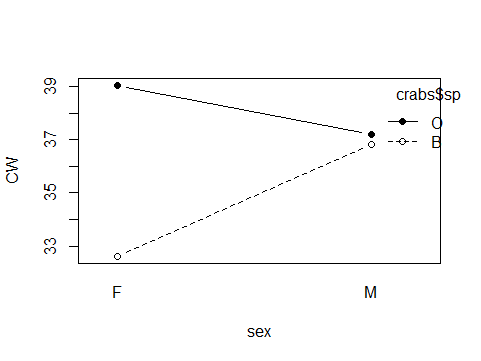
No interaction present between them since the lines are parallel.

interaction.plot(x.factor=crabs$sp, trace.factor=crabs$sex, response=crabs$CW, fun="mean", type="b", ylab="CW", xlab="sp", pch=c(1,19))



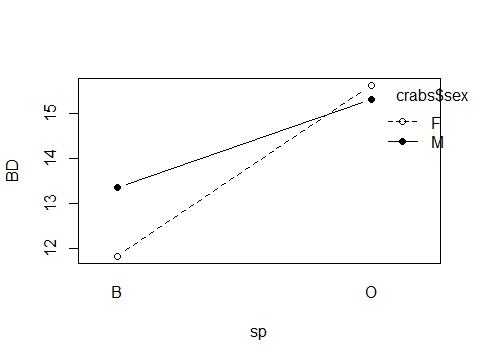
Since the lines cross each other so interaction present in this case.

interaction.plot(x.factor=crabs$sex, trace.factor=crabs$sp, response=crabs$CW, fun="mean", type="b", ylab="CW", xlab="sex", pch=c(1,19))



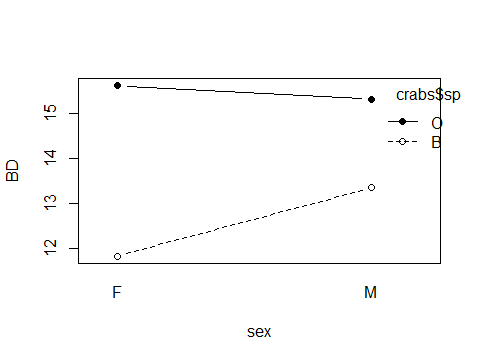
No interaction present between them since the lines are parallel.

interaction.plot(x.factor=crabs$sp, trace.factor=crabs$sex, response=crabs$BD, fun="mean", type="b", ylab="BD", xlab="sp", pch=c(1,19))



Since the lines cross each other so interaction present in this case.

interaction.plot(x.factor=crabs$sex, trace.factor=crabs$sp, response=crabs$BD, fun="mean", type="b", ylab="BD", xlab="sex", pch=c(1,19))



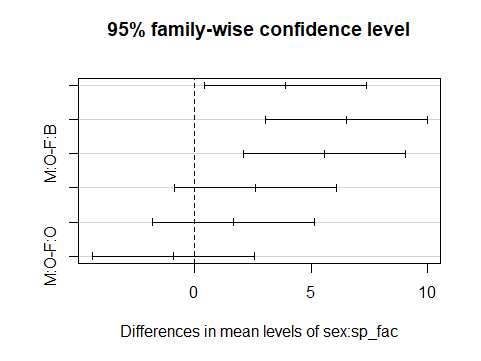
No interaction present between them since the lines are parallel.

1. For each measurement, run an ANOVA that consists of main effects only (color + sex) and describe the results.

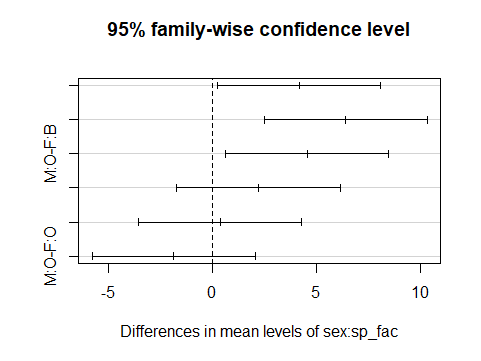
|  |
| --- |
| > summary(aov(FL ~ sp + sex, crabs))  Df Sum Sq Mean Sq F value Pr(>F)  sp 1 466.3 466.3 46.864 **9.43e-11** \*\*\*  sex 1 4.6 4.6 0.458 **0.499**  Residuals 197 1960.3 10.0  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1   * the p-value of sp is **9.43e-11** (significant), which indicates that the species are associated with   significant FL measurement.   * the p-value of sex is 0.499 (not significant), which indicates that the sex are not associated with   FL measurement.  > summary(aov(RW ~ sp + sex, crabs))  Df Sum Sq Mean Sq F value Pr(>F)  sp 1 131.4 131.38 24.09 **1.92e-06 \*\*\***  sex 1 112.1 112.05 20.55 **1.01e-05 \*\*\***  Residuals 197 1074.4 5.45  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  the p-value of sp is **1.92e-06 \*\*\*** (significant), which indicates that the species are associated with  significant RW measurement.  the p-value of sex is **1.01e-05 \*\*\*** ( significant), which indicates that the sex are associated with  RW measurement.  > summary(aov(CL ~ sp + sex, crabs))  Df Sum Sq Mean Sq F value Pr(>F)  sp 1 838 838.5 18.080 **3.27e-05 \*\*\***  sex 1 111 111.2 2.397 **0.123**  Residuals 197 9136 46.4  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  the p-value of sp is **3.27e-05 \*\*\*** (significant), which indicates that the species are associated with  significant CL measurement.  the p-value of sex is 0.123 ( not significant), which indicates that the sex are not associated with  CL measurement.  > summary(aov(CW ~ sp + sex, crabs))  Df Sum Sq Mean Sq F value Pr(>F)  sp 1 576 576.3 9.714 **0.0021 \*\***  sex 1 68 68.3 1.152 **0.2845**  Residuals 197 11687 59.3  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  the p-value of sp is **0.0021 \*\*** (significant), which indicates that the species are associated with  significant CW measurement.  the p-value of sex is 0.2845( not significant), which indicates that the sex are not associated with  CW measurement.  > summary(aov(BD ~ sp + sex, crabs))  Df Sum Sq Mean Sq F value Pr(>F)  sp 1 419.1 419.1 43.535 **3.76e-10 \*\*\***  sex 1 18.8 18.8 1.952 **0.164**  Residuals 197 1896.2 9.6  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  the p-value of sp is **3.76e-10 \*\*\*** (significant), which indicates that the species are associated with  significant BD measurement.  the p-value of sex is 0.164( not significant), which indicates that the sex are not associated with  BD measurement. |
|  |
| |  | | --- | | > | |

1. Run an interaction model for each measurement, and if interaction is present, run a simple main effects analysis.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| |  | | --- | | crabs$sp\_fac <- as.factor(crabs$sp)  > aov.out <- aov(FL ~ sex\*sp\_fac, crabs)  > model.tables(aov.out, type="means", se = TRUE)  Tables of means  Grand mean    15.583  sex  sex  F M  15.432 15.734  sp\_fac  sp\_fac  B O  14.056 17.110  sex:sp\_fac  sp\_fac  sex B O  F 13.270 17.594  M 14.842 16.626  Standard errors for differences of means  sex sp\_fac sex:sp\_fac  0.4380 0.4380 0.6194  replic. 100 100 50  > TukeyHSD(aov.out)  Tukey multiple comparisons of means  95% family-wise confidence level  Fit: aov(formula = FL ~ sex \* sp\_fac, data = crabs)  $`sex`  diff lwr upr p adj  M-F 0.302 -0.5617106 1.165711 0.4912819  $sp\_fac  diff lwr upr p adj  O-B 3.054 2.190289 3.917711 0  $`sex:sp\_fac`  diff lwr upr p adj  M:B-F:B 1.572 -0.03289788 3.1768979 0.0572712  F:O-F:B 4.324 2.71910212 5.9288979 0.0000000  M:O-F:B 3.356 1.75110212 4.9608979 0.0000010  F:O-M:B 2.752 1.14710212 4.3568979 0.0000868  M:O-M:B 1.784 0.17910212 3.3888979 0.0227162  M:O-F:O -0.968 -2.57289788 0.6368979 0.4022824  > plot(TukeyHSD(aov.out)) | |  | | |  | | --- | | > | |     crabs$sp\_fac <- as.factor(crabs$sp)  >  > aov.out <- aov(CL ~ sex\*sp\_fac, crabs)  > model.tables(aov.out, type="means", se = TRUE)  Tables of means  Grand mean    32.1055  sex  sex  F M  31.36 32.85  sp\_fac  sp\_fac  B O  30.06 34.15  sex:sp\_fac  sp\_fac  sex B O  F 28.10 34.62  M 32.01 33.69  Standard errors for differences of means  sex sp\_fac sex:sp\_fac  0.9499 0.9499 1.3434  replic. 100 100 50  > TukeyHSD(aov.out)  Tukey multiple comparisons of means  95% family-wise confidence level  Fit: aov(formula = CL ~ sex \* sp\_fac, data = crabs)  $`sex`  diff lwr upr p adj  M-F 1.491 -0.3823368 3.364337 0.1181121  $sp\_fac  diff lwr upr p adj  O-B 4.095 2.221663 5.968337 2.57e-05  $`sex:sp\_fac`  diff lwr upr p adj  M:B-F:B 3.912 0.431072 7.392928 0.0207290  F:O-F:B 6.516 3.035072 9.996928 0.0000148  M:O-F:B 5.586 2.105072 9.066928 0.0002783  F:O-M:B 2.604 -0.876928 6.084928 0.2154635  M:O-M:B 1.674 -1.806928 5.154928 0.5983099  M:O-F:O -0.930 -4.410928 2.550928 0.8999496  > plot(TukeyHSD(aov.out)) |
|  |
| |  | | --- | | > | |



|  |
| --- |
| crabs$sp\_fac <- as.factor(crabs$sp)  >  > aov.out <- aov(CW ~ sex\*sp\_fac, crabs)  > model.tables(aov.out, type="means", se = TRUE)  Tables of means  Grand mean    36.4145  sex  sex  F M  35.83 37.00  sp\_fac  sp\_fac  B O  34.72 38.11  sex:sp\_fac  sp\_fac  sex B O  F 32.62 39.04  M 36.81 37.19  Standard errors for differences of means  sex sp\_fac sex:sp\_fac  1.071 1.071 1.514  replic. 100 100 50  > TukeyHSD(aov.out)  Tukey multiple comparisons of means  95% family-wise confidence level  Fit: aov(formula = CW ~ sex \* sp\_fac, data = crabs)  $`sex`  diff lwr upr p adj  M-F 1.169 -0.9423008 3.280301 0.2761959  $sp\_fac  diff lwr upr p adj  O-B 3.395 1.283699 5.506301 0.001762  $`sex:sp\_fac`  diff lwr upr p adj  M:B-F:B 4.186 0.2629007 8.109099 0.0314443  F:O-F:B 6.412 2.4889007 10.335099 0.0002045  M:O-F:B 4.564 0.6409007 8.487099 0.0153227  F:O-M:B 2.226 -1.6970993 6.149099 0.4574462  M:O-M:B 0.378 -3.5450993 4.301099 0.9945201  M:O-F:O -1.848 -5.7710993 2.075099 0.6145237  > plot(TukeyHSD(aov.out)) |
|  |
| |  | | --- | | > | |



crabs$sp\_fac <- as.factor(crabs$sp)

>

> aov.out <- aov(BD ~ sex\*sp\_fac, crabs)

> model.tables(aov.out, type="means", se = TRUE)

Tables of means

Grand mean

14.0305

sex

sex

F M

13.724 14.337

sp\_fac

sp\_fac

B O

12.583 15.478

sex:sp\_fac

sp\_fac

sex B O

F 11.816 15.632

M 13.350 15.324

Standard errors for differences of means

sex sp\_fac sex:sp\_fac

0.4349 0.4349 0.6151

replic. 100 100 50

> TukeyHSD(aov.out)

Tukey multiple comparisons of means

95% family-wise confidence level

Fit: aov(formula = BD ~ sex \* sp\_fac, data = crabs)

$`sex`

diff lwr upr p adj

M-F 0.613 -0.2447489 1.470749 0.1602977

$sp\_fac

diff lwr upr p adj

O-B 2.895 2.037251 3.752749 0

$`sex:sp\_fac`

diff lwr upr p adj

M:B-F:B 1.534 -0.05982032 3.12782 0.0639702

F:O-F:B 3.816 2.22217968 5.40982 0.0000000

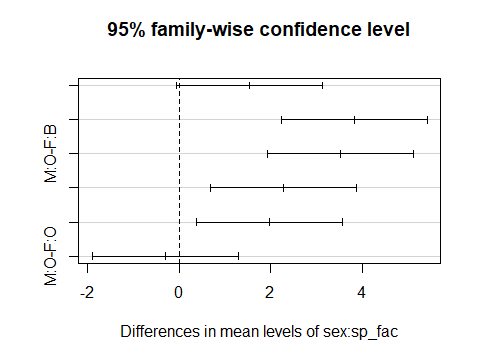
M:O-F:B 3.508 1.91417968 5.10182 0.0000003

F:O-M:B 2.282 0.68817968 3.87582 0.0015279

M:O-M:B 1.974 0.38017968 3.56782 0.0083935

M:O-F:O -0.308 -1.90182032 1.28582 0.9588102

> plot(TukeyHSD(aov.out))



1. What are your overall conclusions regarding how these characteristics vary by color and sex in this crab species?

My overall conclusion is that, all five characteristics have an interaction with species as well as color variable. On the other hand, only FL measurement has an interaction with sex variable and other characteristics have no interaction with sex variable.