

# OCN 750 - HW5

*Maia Kapur*

*September 23, 2015*

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

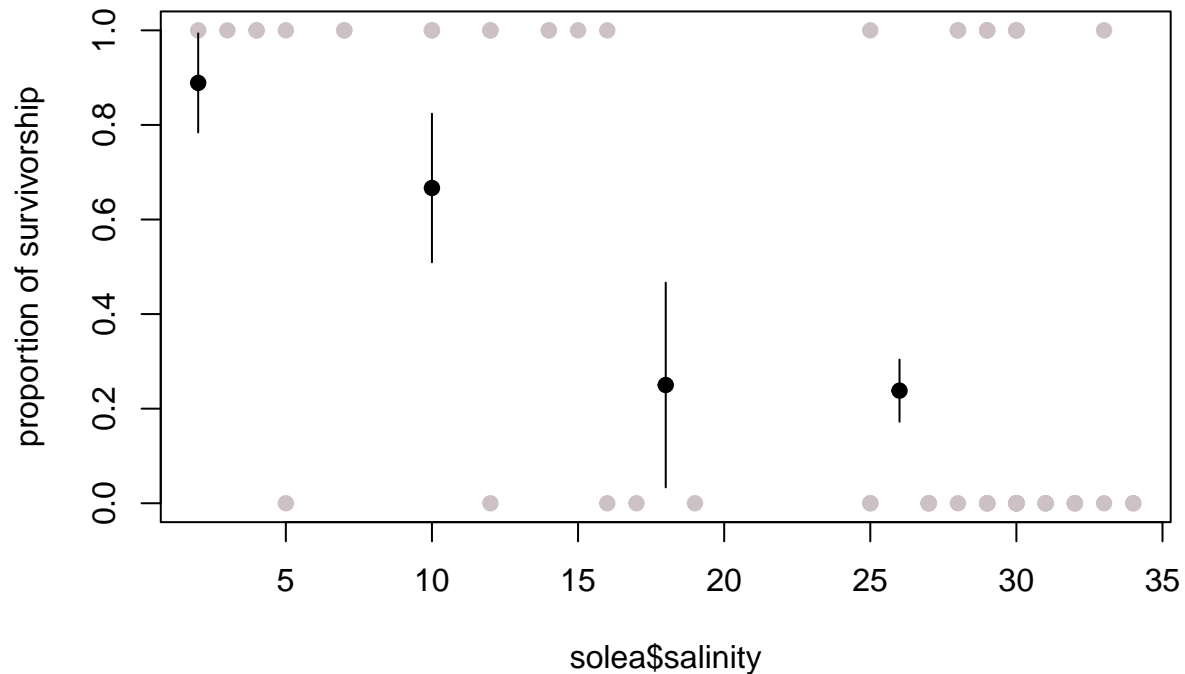
When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
setwd("~/Dropbox/2015 Fall/OCN 750")
solea = read.table("Solea.txt", header = TRUE)

#define 8 bins, using a sequence of 4 breaks
breaks = with(solea, seq(min(salinity), max(salinity), length = 5))
#make a binning factor from logPar
cut.sal = cut(solea$salinity, breaks = breaks)
#calculate the proportion of presence by bin
means = with(solea, tapply(Solea_solea, cut.sal, mean))
#function for the standard error of a binary variable
binomial.SE = function(x) sqrt((mean(x)*(1-mean(x)))/length(x))
#calculate the standard error for each bin
ses = with(solea, tapply(Solea_solea, cut.sal, binomial.SE))
#plot the raw data
plot(solea$Solea_solea ~ solea$salinity,
     main = "proportional survivorship vs. salinity, solea",
     ylab = "proportion of survivorship",
     col = "lavenderblush3", pch = 19, col.main = "hotpink4")
#plot the bin means
points(means ~ breaks[1:4], pch = 19, ylim = c(0,1), xlim =
range(solea$salinity))
#plot the bin SEs
segments(breaks[1:4], means+ses, breaks[1:4], means-ses)
```

Plot the raw data of sole presence/absence vs. salinity. Break the salinity data into four equally sized bins, and within each bin calculate the proportion of presences. Plot the binned proportions vs. the midpoints of the bins (e.g. if the bin is from 3 to 7, then 5 is the midpoint). Also calculate the standard error of the proportion for each bin, and plot that with error bars.

## proportional survivorship vs. salinity, solea



#####Fit a binomial glm to test whether presence/absence of sole is driven by salinity. #####Use curve() to plot the fitted logistic relationship on the same plot as the raw data.

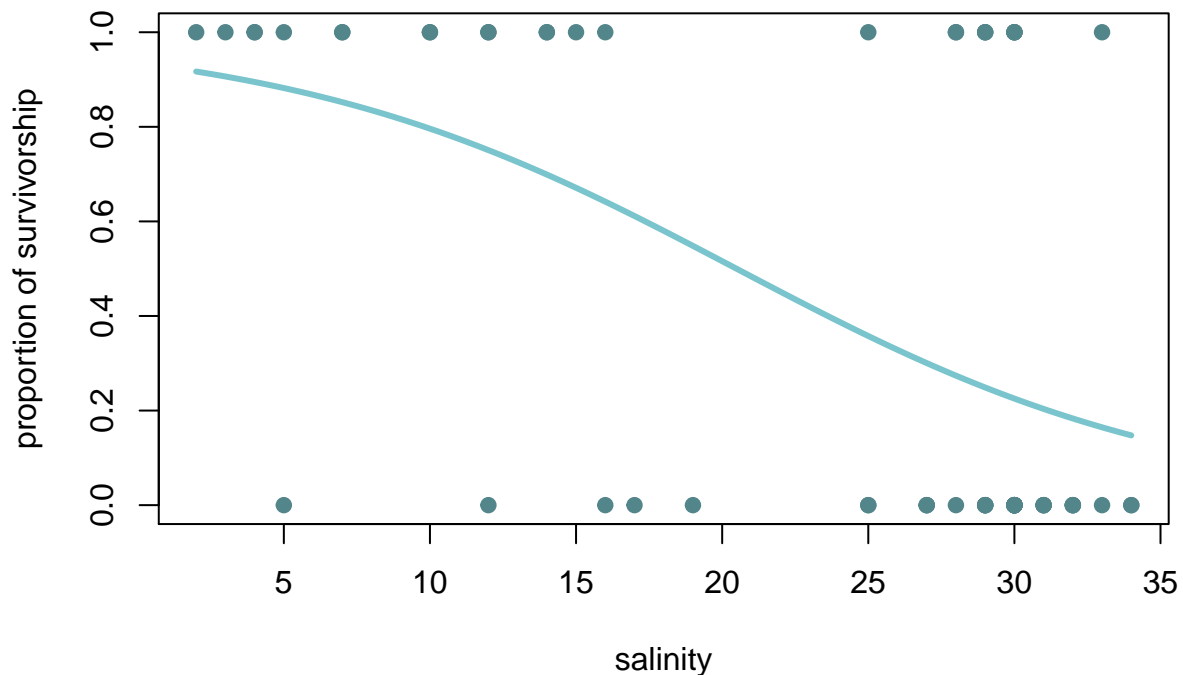
```
##create binomial model
mod = glm(Solea_solea ~ salinity, data = solea, family = binomial)
summary(mod)

##
## Call:
## glm(formula = Solea_solea ~ salinity, family = binomial, data = solea)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.0674  -0.7146  -0.6362   0.7573   1.8997
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  2.66071    0.90167   2.951 0.003169 **
## salinity     -0.12985    0.03494  -3.716 0.000202 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 87.492  on 64  degrees of freedom
## Residual deviance: 68.560  on 63  degrees of freedom
## AIC: 72.56
##
## Number of Fisher Scoring iterations: 4
```

```
##define logistic function
logistic = function(x) exp(x)/(1+exp(x))

##plot the raw data and add on the fittedcurves
plot(Solea_solea ~ salinity, data = solea,
     main = "proportional survivorship vs. salinity, solea",
     ylab = "proportion of survivorship", col = "cadetblue4",
     pch = 19, col.main = "cadetblue3")
curve(logistic(coef(mod)[1]+coef(mod)[2]*x), add = T, col = 'cadetblue3', lwd = 3)
```

### proportional survivorship vs. salinity, solea



#####What is the effect of salinity on sole presence/absence? Would you consider it a strong effect or a weak effect? Why?

```
##It looks to be a gentle yet definite effect.
##The model summary indicates a high AIC for the salinity variable.
```

```
require(car)
Anova(mod) ##it's significant
```

Do a likelihood ratio test to see whether the effect of salinity is significant.

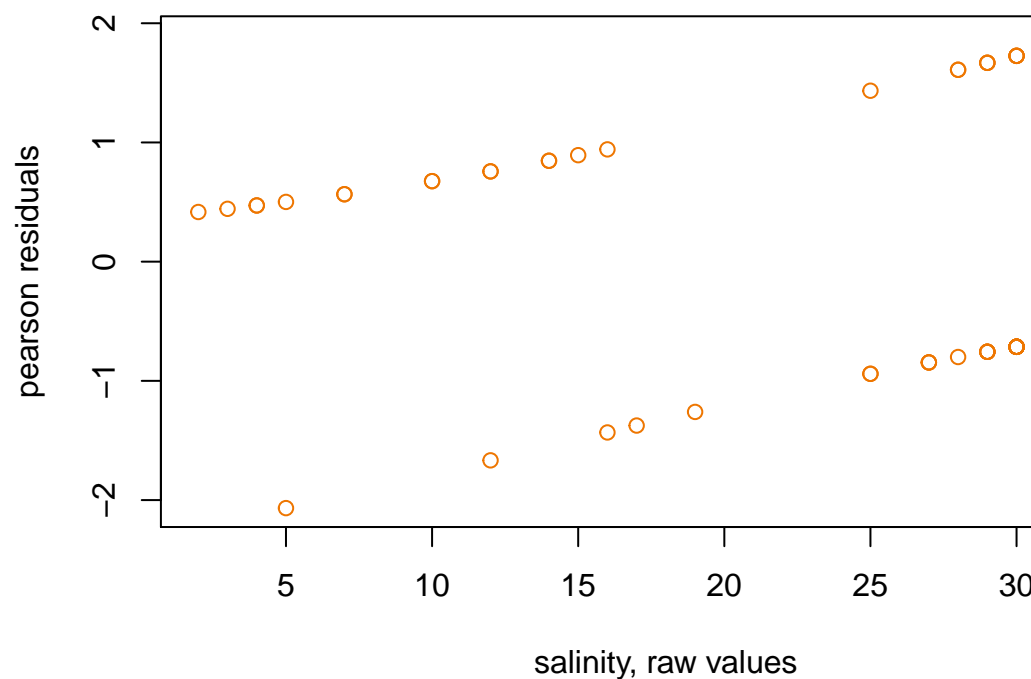
```
## Analysis of Deviance Table (Type II tests)
##
## Response: Solea_solea
##           LR Chisq Df Pr(>Chisq)
## salinity  18.931  1  1.355e-05 ***
```

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

```
plot(residuals(mod, family = "pearson") ~ solea$salinity,
     xlab = "salinity, raw values",
     ylab = "pearson residuals", main = "Deviance Residuals vs. Predicted Values",
     col = "darkorange2", col.main = "darkorange4")
```

Examine a plot of residuals (deviance residuals or pearson residuals) vs. predicted values. How

## Deviance Residuals vs. Predicted Values



would you interpret this plot?

```
##This plot depicts a quasi-linear trend in residual distribution, which is expected for log predictors.
```

```
hs = read.csv("~/Dropbox/2015 Fall/OCN 750/heat_shock_subset.csv")

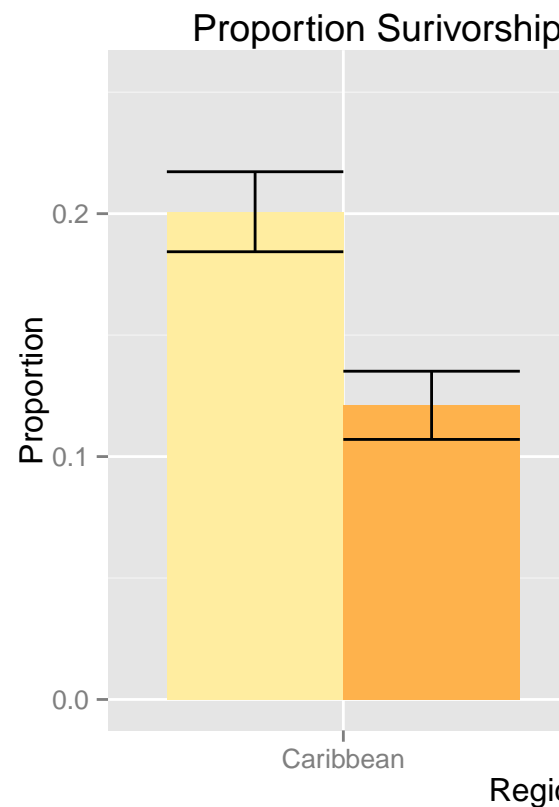
##aggregate data to find means
means = aggregate(formula = proportion ~ sex + region, data = hs, FUN = mean)

##calculate standard error for sex and region
se = aggregate(formula = proportion ~ sex + region, data = hs, FUN = function(x)
               sqrt(var(x)/length(x)) )

##create a list that contains the values from the above functions
mu = c(means$proportion)
stderrs = c(se$proportion)
```

```
##plot it
require(ggplot2)

ggplot(means, aes(x = region, proportion, fill = sex)) +
  geom_bar(stat = "identity", position = "dodge") +
  geom_errorbar(aes(ymax = mu + stderrs, ymin = mu - stderrs), position = "dodge") +
  ggtitle("Proportion Survivorship by Sex and Region") +
  xlab("Region") +
  ylab("Proportion") +
  scale_fill_brewer(palette = "YlOrRd")
```



Make an exploratory plot of proportion of survivors vs. sex\*region.

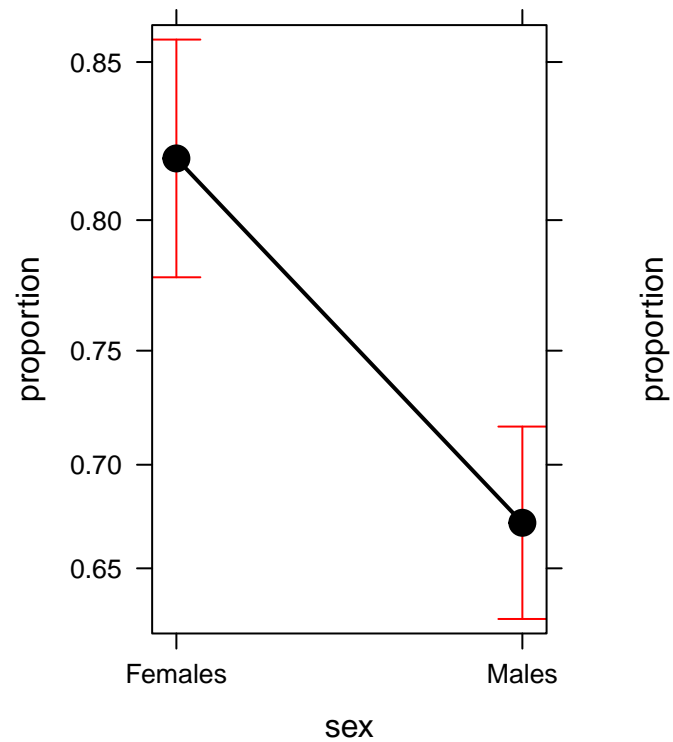
```
##for the binomial glm, you must have integer 0 or 1 values (1 = survived)
##replace all non-zero values with 1
hs$proportion[hs$proportion != 0] <- 1

##fit binomial model for non-interaction
hsmod.noint = glm(proportion ~ sex + region, data = hs, family = "binomial")
hsmod.int = glm(proportion ~ sex * region, data = hs, family = "binomial")
```

Fit a binomial glm where you test for effects of sex and region on the probability of surviving, as well as an interaction between sex and region. Remember there are two different ways to fit a binomial glm where  $n > 1$ . They are equivalent.

```
require(effects)
plot(allEffects(hsmod.noint), main = "Effect of Sex and Region on Proportion Survivorship")
```

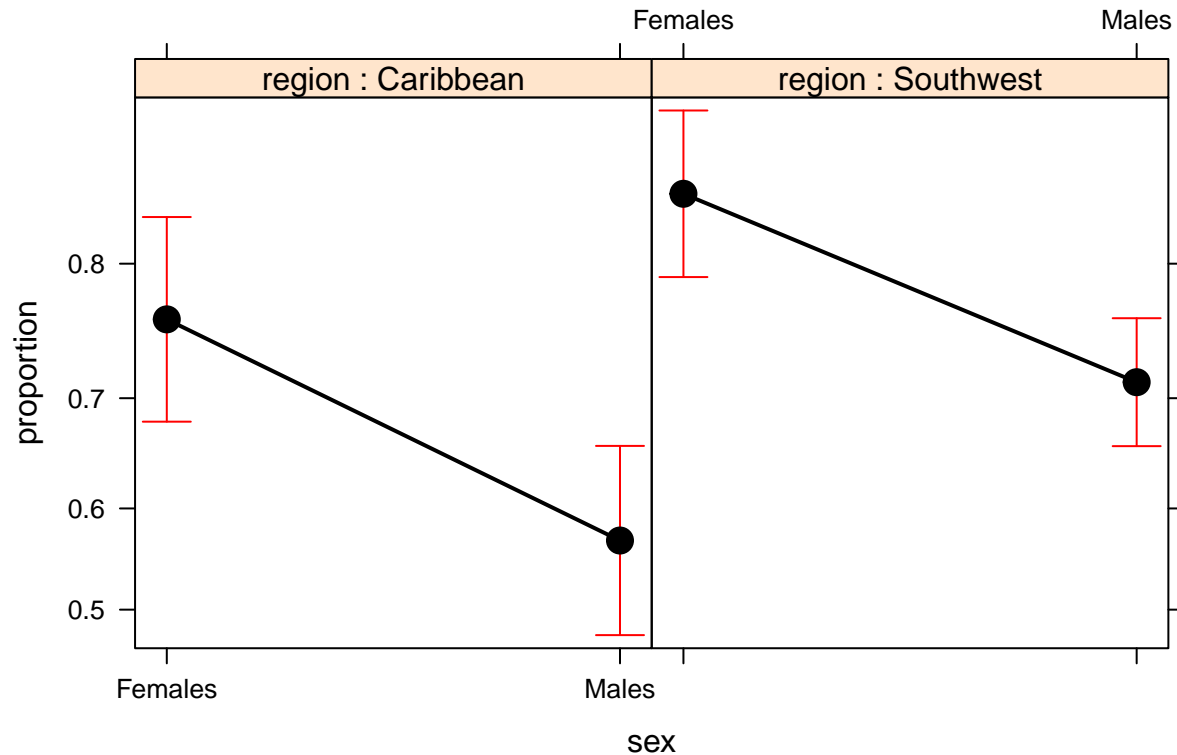
## Sex and Region on Proportion Survivorship



Plot the fitted effects. How do you interpret the results?

```
plot(allEffects(hsmod.int), main = "Interaction of Sex and Region on Proportion Survivorship")
```

## Interaction of Sex and Region on Proportion Survivorship



##There don't seem to be very different effects between the two groups; both have a lower prop. of male

```
require(car)
Anova(hsmod.int)
```

Do a likelihood ratio test for the interaction between sex and region.

```
## Analysis of Deviance Table (Type II tests)
##
## Response: proportion
##          LR Chisq Df Pr(>Chisq)
## sex          22.6834 1  1.91e-06 ***
## region        10.5796 1  0.001143 **
## sex:region     0.1355 1  0.712824
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
require(car)
##fit quasi-binomial lm
hsqbin = glm(proportion ~ sex * region, data = hs, family = "quasibinomial")
```

```
##define overdispersion parameter
overdis = function(model) {
  sum(residuals(model, type = "pearson")^2)/(length(model$y) - length(model$coefficients))
}
overdis(hsqbin) ##it's almost perfectly 1 (e.g. almost 0 overdisp.)
```

Oops, we didn't consider overdispersion. Fit the same model using the quasibinomial method. How big is the dispersion parameter?

```
## [1] 1.005161
```

“ “####Test the interaction with the quasibinomial model, using the F-test that accounts for the dispersion parameter. How do you interpret this result, compared to the result from the first model?

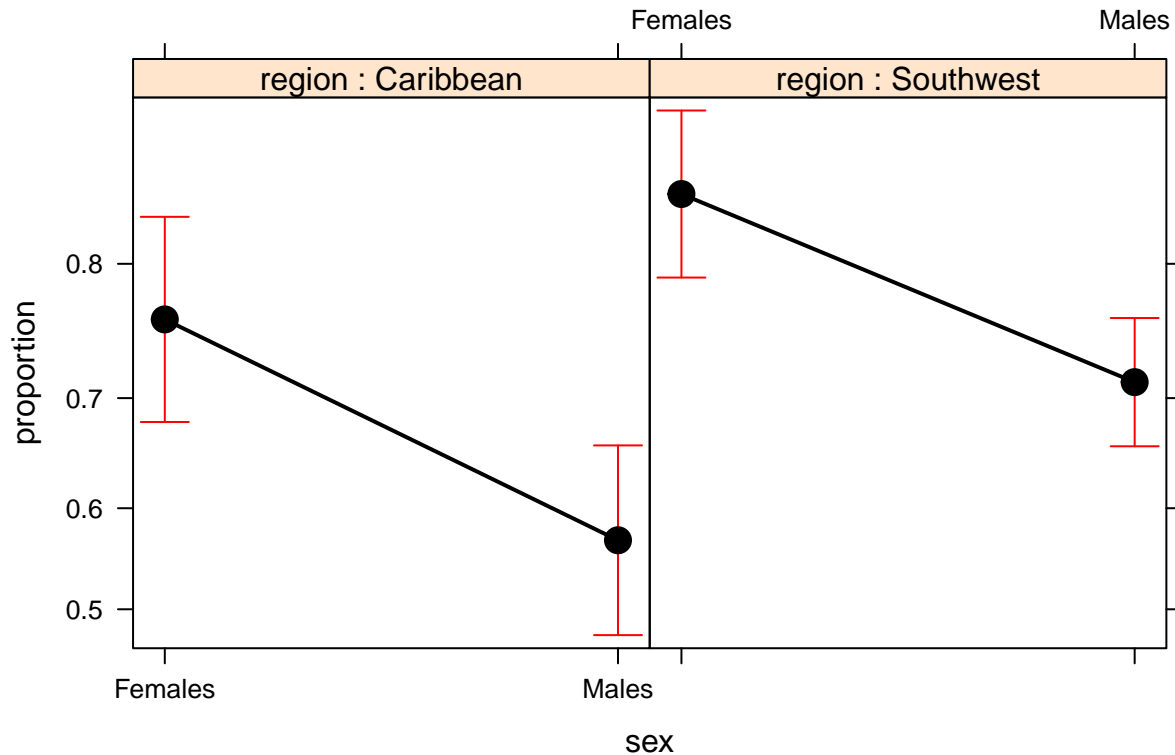
```
Anova(hsqbin, test = 'F')
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: proportion
##          SS Df      F    Pr(>F)
## sex      22.68  1 22.5669 2.418e-06 ***
## region   10.58  1 10.5252 0.001228 **
## sex:region  0.14  1  0.1348 0.713630
## Residuals 779.00 775
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
plot(allEffects(hsqbin), main = "Effects plot for Q-Binom distribution, Sex x Region")
```



## Effects plot for Q-Binom distribution, Sex x Region



##this doesn't look very different from before -- likely cause overdispersion was so low

##For the interaction, Anova() compares the full model to a restricted model that has the interaction removed. `Anova(hsqbin, test = 'F')`

Test the main effects of the model (sex and region), using F-tests where the sex\*region interaction is no longer in the 'full' model. What's your final take on these results? Are the results consistent with the researchers' hypothesis of local adaptation?

```
## Analysis of Deviance Table (Type II tests)
##
## Response: proportion
##          SS Df      F    Pr(>F)
## sex      22.68  1 22.5669 2.418e-06 ***
## region   10.58  1 10.5252  0.001228 **
## sex:region  0.14  1  0.1348  0.713630
## Residuals 779.00 775
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

##the Anova() reveals that indeed, the interaction is non-significant. Sex is the most significant driver