

STAT 8020 R Lab 15: Logistic Regression

Whitney

October 22, 2020

Contents

Odds Ratio	1
Heart Attack Example	1
Gender vs. Handness Example	1
Logistic Regression	1
Diagnostic	4
Model selection	5

Odds Ratio

Heart Attack Example

```
n11 = 189; n12 = 10845; n21 = 104; n22 = 10933
## odds ratio
theta <- (n11 * n22) / (n12 * n21)
SE_log.theta <- sqrt(1 / n11 + 1 / n12 + 1 / n21 + 1 / n22)
z <- qnorm(0.975)
CI_log.theta <- c(log(theta) - z * SE_log.theta, log(theta) + z * SE_log.theta)
exp(CI_log.theta)

## [1] 1.440042 2.330780
```

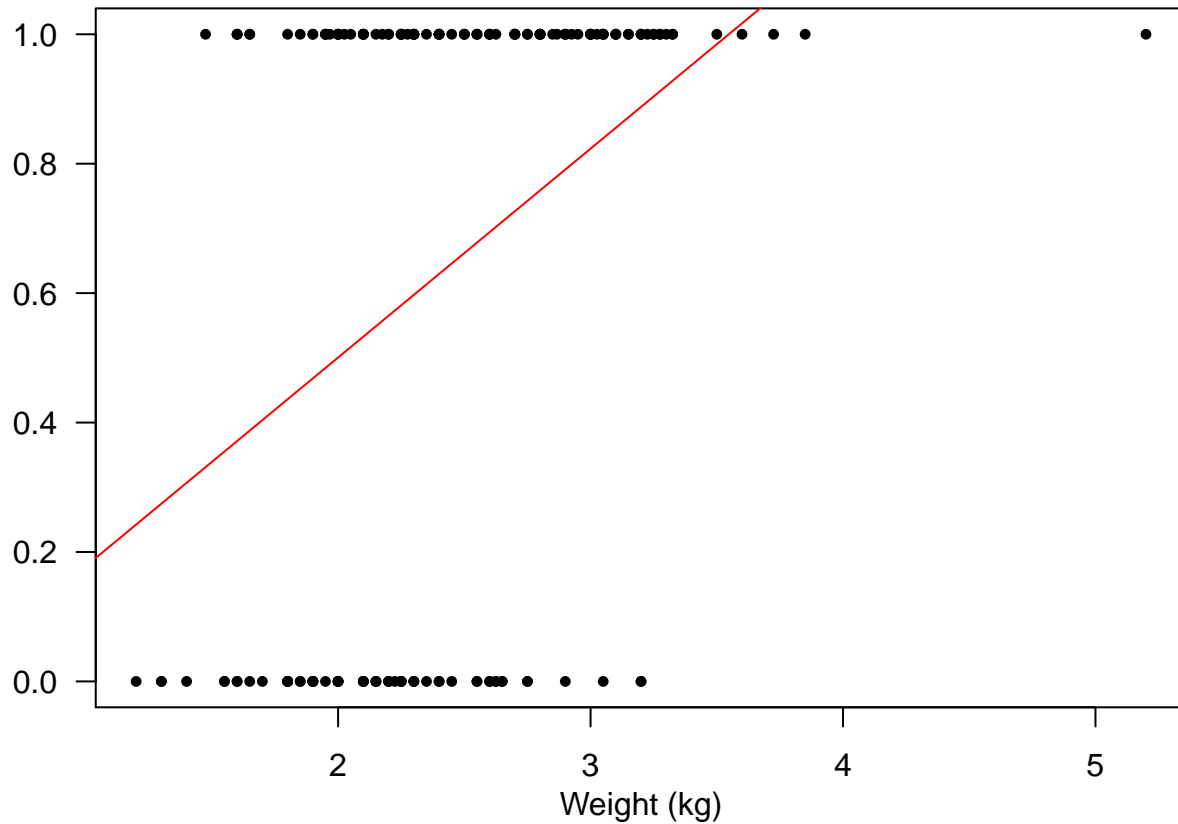
Gender vs. Handness Example

```
n11 = 9; n12 = 43; n21 = 4; n22 = 44
theta <- (n11 * n22) / (n12 * n21)
SE_log.theta <- sqrt(1 / n11 + 1 / n12 + 1 / n21 + 1 / n22)
z <- qnorm(0.975)
CI_log.theta <- c(log(theta) - z * SE_log.theta, log(theta) + z * SE_log.theta)
exp(CI_log.theta)

## [1] 0.6592751 8.0401992
```

Logistic Regression

```
crab <- read.table("http://users.stat.ufl.edu/~aa/cda/data/Crabs.dat", header = T)
# Linear Regression
lmFit <- lm(crab$y ~ crab$weight)
par(mar = c(3.5, 3.5, 0.8, 0.6))
plot(crab$weight, crab$y, pch = 16, cex = 0.75, las = 1, xlab = "",
     ylab = "")
mtext("Weight (kg)", side = 1, line = 2)
abline(lmFit, col = "red")
```



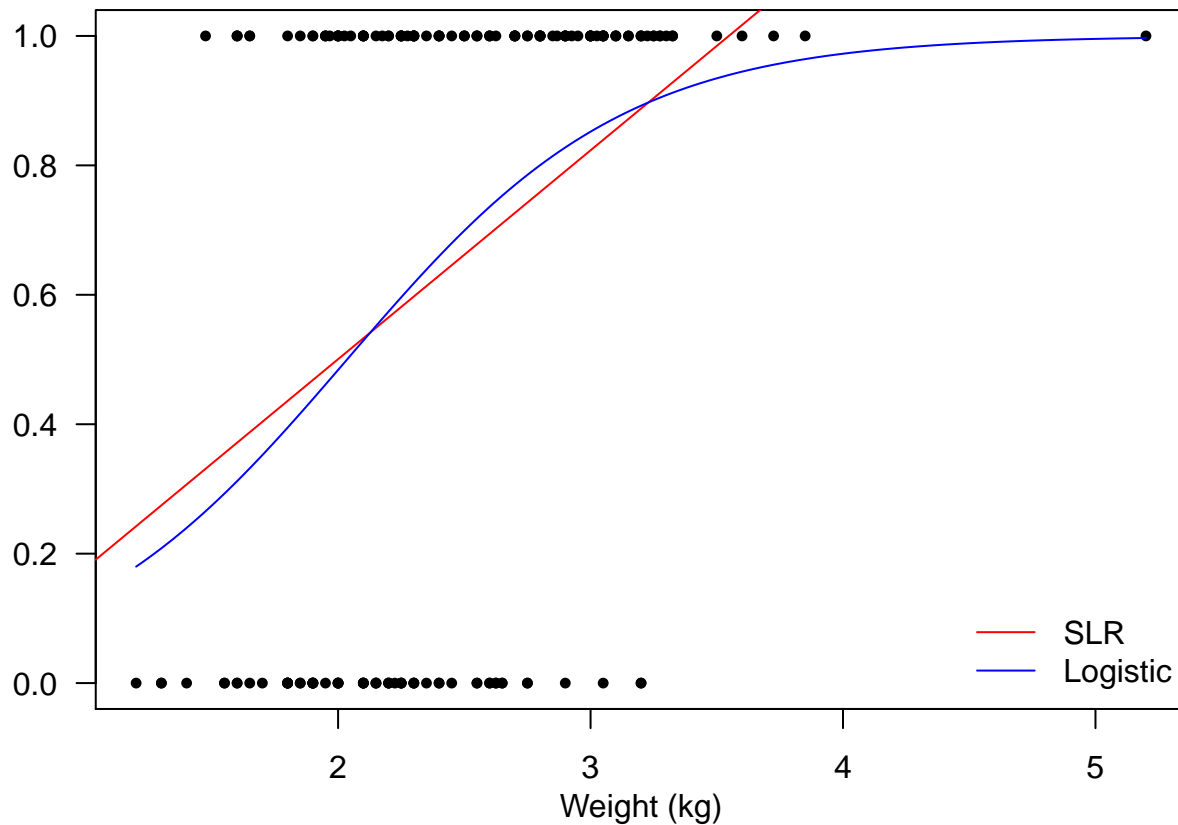
```
# Logistic Regression
logitFit <- glm(y ~ weight, data = crab, family = "binomial")
summary(logitFit)
```

```
##
## Call:
## glm(formula = y ~ weight, family = "binomial", data = crab)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.1108  -1.0749   0.5426   0.9122   1.6285
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -3.6947     0.8802  -4.198 2.70e-05 ***
## weight         1.8151     0.3767   4.819 1.45e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 225.76  on 172  degrees of freedom
## Residual deviance: 195.74  on 171  degrees of freedom
## AIC: 199.74
##
## Number of Fisher Scoring iterations: 4
```

```

# Plot the fits
rg <- range(crab$weight)
xg <- seq(rg[1], rg[2], 0.01)
pred <- predict(logitFit, newdata = data.frame(weight = xg), type = "response")
par(mar = c(3.5, 3.5, 0.8, 0.6))
plot(crab$weight, crab$y, pch = 16, cex = 0.75, las = 1, xlab = "",
      ylab = "")
mtext("Weight (kg)", side = 1, line = 2)
abline(lmFit, col = "red")
lines(xg, pred, col = "blue")
legend("bottomright", legend = c("SLR", "Logistic"),
      col = c("red", "blue"), lty = 1, bty = "n")

```



```

# Another CI
library(MASS)
confint(logitFit)

```

```
## Waiting for profiling to be done...
```

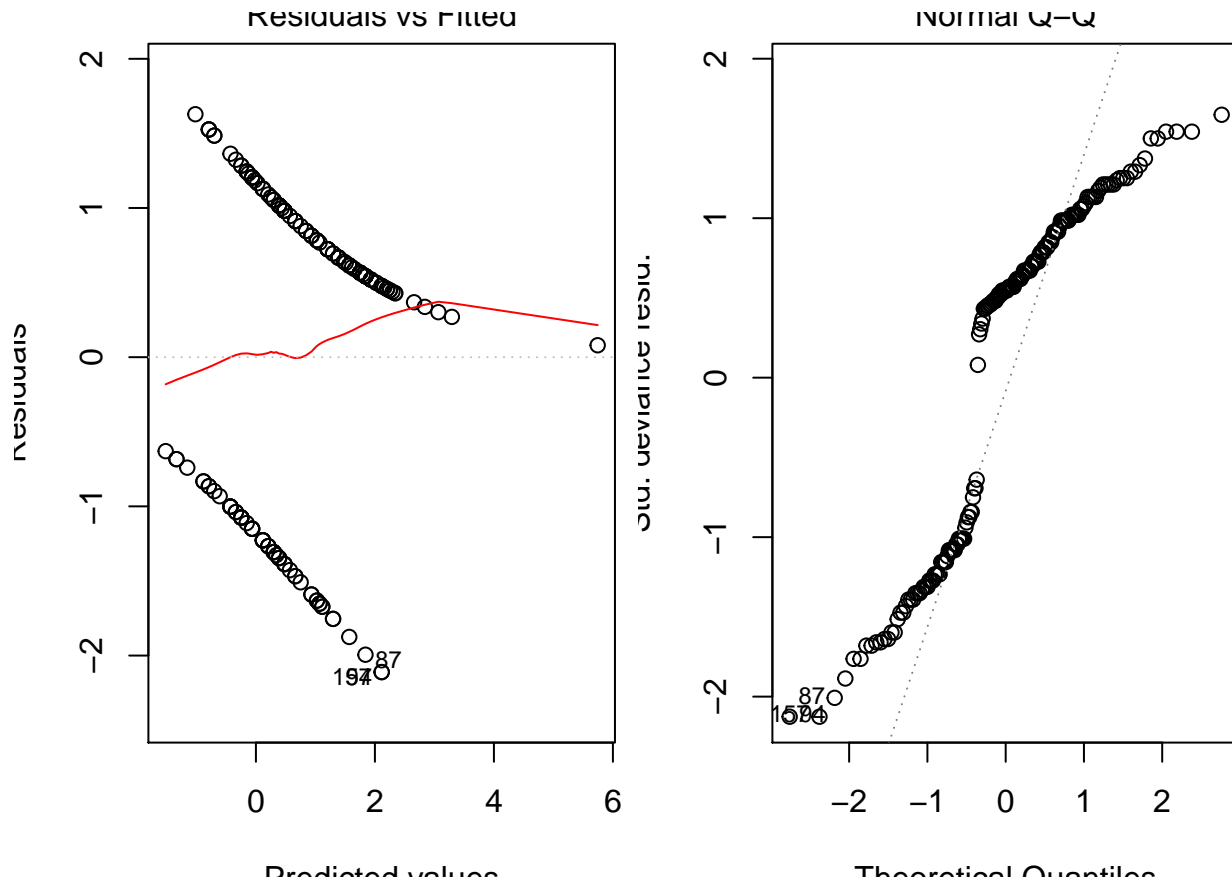
```
##           2.5 %    97.5 %
## (Intercept) -5.505932 -2.039701
## weight      1.113790  2.597305

```

```

# Residual plot
par(mfrow = c(1, 2))
plot(logitFit, which = 1:2)

```



Diagnostic

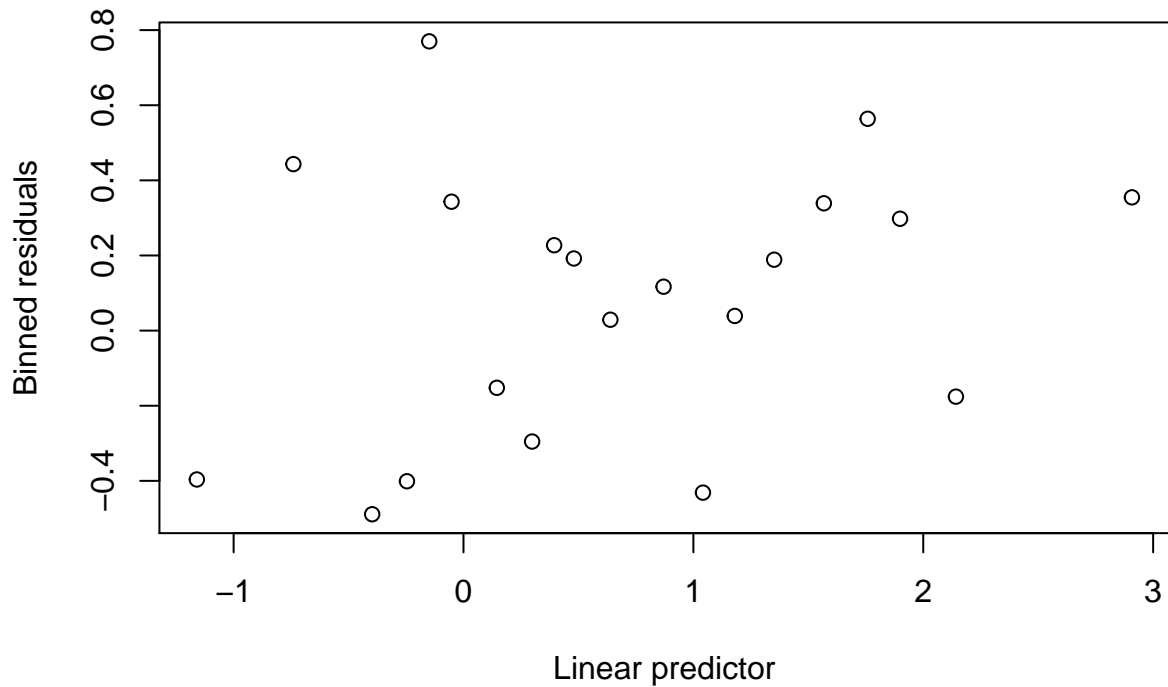
```
wt_bin <- findInterval(crab$weight, unique(quantile(crab$weight, seq(0, 1, length.out = 20 + 1))), right = FALSE)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
##
## The following object is masked from 'package:MASS':
##
##   select
##
## The following objects are masked from 'package:stats':
##
##   filter, lag
##
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

crab.res <- mutate(crab, residuals = residuals(logitFit), Linpred = predict(logitFit), bin = wt_bin)

res_bin <- tapply(crab.res$residuals, crab.res$bin, mean)
Lpred_bin <- tapply(crab.res$Linpred, crab.res$bin, mean)

plot(res_bin ~ Lpred_bin, xlab = "Linear predictor",
      ylab = "Binned residuals")
```



Model selection

```
logitFit2 <- glm(y ~ weight + width, data = crab, family = "binomial")
step(logitFit2)
```

```
## Start: AIC=198.89
## y ~ weight + width
##
##           Df Deviance   AIC
## - weight  1   194.45 198.45
## <none>           192.89 198.89
## - width   1   195.74 199.74
##
## Step: AIC=198.45
## y ~ width
##
##           Df Deviance   AIC
## <none>           194.45 198.45
## - width  1   225.76 227.76
##
## Call: glm(formula = y ~ width, family = "binomial", data = crab)
##
## Coefficients:
## (Intercept)      width
##   -12.3508     0.4972
##
## Degrees of Freedom: 172 Total (i.e. Null); 171 Residual
## Null Deviance:      225.8
## Residual Deviance: 194.5   AIC: 198.5
```