Introduction to R
Session 5 – Advanced Graphics

Statistical Consulting Centre

20 July, 2017

1. Logistic regression

1. Suppose the only information available about mercury concentration were that the lakes were either clean (0) or contaminated (1). Create a factor called `cont`, such that $cont = 1$ if `mercury > 0.53` and $cont = 0$ if `mercury ≤ 0.53`.

```
cont <- ifelse(joined.long.df$Mercury > 0.53, 1, 0)
cont <- as.factor(cont)
```

2. Check the one-way frequency table of `cont`.

```
table(cont)
```

```
## cont
##  0  1
## 91 68
```

3. Add `cont` to `joined.long.df`.

```
joined.long.df$cont <- cont
```

4. (a) Perform a logistic regression to assess how `pH` influences the probability of contamination.

```
myglm <- glm(cont~pH, family = binomial, data = joined.long.df)
```

(b) What is the overall significance of `pH`.

```
anova(myglm, test = "Chisq")
```

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: cont
##
## Terms added sequentially (first to last)
##
##
##      Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
## NULL                   158     217.08
## pH    1   50.445       157     166.64 1.226e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

(c) How does a 1 unit increase in `pH` change the log odds of contamination?

```
summary(myglm)
```

```
##
## Call:
```

```
## glm(formula = cont ~ pH, family = binomial, data = joined.long.df)
##
## Deviance Residuals:
##     Min       1Q    Median       3Q      Max
## -1.8899  -0.8550  -0.4251   0.9729   1.9320
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   6.8833     1.2482   5.515  3.5e-08 ***
## pH           -1.1002     0.1897  -5.798  6.7e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 217.08  on 158  degrees of freedom
## Residual deviance: 166.64  on 157  degrees of freedom
## AIC: 170.64
##
## Number of Fisher Scoring iterations: 4
```

A 1 unit increase in pH decreases the log odds of contamination by 1.03