

## Logistic Regression

Odds ratios are below, cigs variable is significant.

```
library(faraway)
data(wcgs, package="faraway")
lmod <- glm(chd ~ height + cigs, family = binomial, wcgs)
summary(lmod)

##
## Call:
## glm(formula = chd ~ height + cigs, family = binomial, data = wcgs)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.0041  -0.4425  -0.3630  -0.3499   2.4357
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -4.50161     1.84186  -2.444  0.0145 *
## height       0.02521     0.02633   0.957  0.3383
## cigs         0.02313     0.00404   5.724 1.04e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 1781.2  on 3153  degrees of freedom
## Residual deviance: 1749.0  on 3151  degrees of freedom
## AIC: 1755
##
## Number of Fisher Scoring iterations: 5
beta <- coef(lmod)
```

We can say that the odds of heart disease increase by 2.6% with each additional inch in height and by 2.3% with each additional cigarette smoked per day. Note that  $\exp x \approx 1+x$  for small values of  $x$ . We observe that so the  $\beta_2 = 2.3\%$  increase in odds due to smoking a cigarette a day could have been quickly estimated from the original model output without further computation. This approximation is useful for quick intuitions. It is more natural to compute the effect of a pack a day (20 cigarettes):

```
exp(beta[3]*20)
```

```
##      cigs
## 1.588115
```

So we have 59% increase in the odds of heart disease due to smoking.

The model is significant based on the anova table above. Can we drop height?

```
anova(lmod, test="Chi")
```

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: chd
##
## Terms added sequentially (first to last)
##
##          Df Deviance Resid. Df Resid. Dev Pr(>Chi)
## NULL                3153      1781.2
## height  1      1.125      3152      1780.1  0.2888
## cigs    1    31.070      3151      1749.0 2.49e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

lmodc <- glm(chd ~ cigs, family = binomial, wcsv)
anova(lmodc, lmod, test="Chi")
```

```
## Analysis of Deviance Table
##
## Model 1: chd ~ cigs
## Model 2: chd ~ height + cigs
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1      3152      1750
## 2      3151      1749  1  0.92025  0.3374
```

Yes height can be dropped from the model. 95% confidence intervals for the coefficients

```
confint(lmod)
```

```
## Waiting for profiling to be done...
##              2.5 %      97.5 %
## (Intercept) -8.13475465 -0.91297018
## height      -0.02619902  0.07702835
## cigs         0.01514949  0.03100534
```

Doing Prediction:

```
test<-data.frame(cbind(height=72,cigs=0))
predict(lmod,test,type="response",se.fit=TRUE)
```

```
## $fit
##      1
## 0.06376553
##
## $se.fit
##      1
## 0.006434148
##
## $residual.scale
## [1] 1
```

Goodness of Fit Test:

Ho: The model is a good fit Ha: The model is not a good fit.

```
library(ResourceSelection)

## ResourceSelection 0.3-5    2019-07-22
hoslem.test(lmod$y,fitted(lmod),g=5)

##
## Hosmer and Lemeshow goodness of fit (GOF) test
##
## data:  lmod$y, fitted(lmod)
## X-squared = 2.8341, df = 3, p-value = 0.4179
Accept Ho. Fit is good.
```