Stats for Clinical Trials, Math 150
Jo Hardin
Logistic Regression example # 1

1. Consider the HERS data described in your book (age 30); variable description also given on the book website `http://www.epibiostat.ucsf.edu/biostat/vgsm/data/hersdata.codebook.txt`

   For now, we will try to predict whether the individuals had a medical condition, `medcond`. We will use the variables `age`, `weight`, `diabetes` and `drinkany`

```
> HERS <- read.table("HERS.txt", sep="\t", header=T, na.strings=".")
> attach(HERS)


> summary(glm(medcond ~ age, family="binomial"))$coef
               Estimate  Std. Error   z value     Pr(>|z|)
(Intercept) -1.60404454 0.400644718 -4.003658 6.237044e-05
age          0.01619155 0.005965348  2.714267 6.642259e-03

> summary(glm(medcond ~ age + weight, family="binomial"))$coef
                Estimate  Std. Error   z value     Pr(>|z|)
(Intercept) -2.169846602 0.496466231 -4.370582 1.239155e-05
age          0.018926204 0.006132171  3.086379 2.026105e-03
weight       0.005279148 0.002742218  1.925138 5.421212e-02

> summary(glm(medcond ~ age+diabetes, family="binomial"))$coef
               Estimate  Std. Error   z value     Pr(>|z|)
(Intercept) -1.89085342 0.407555871 -4.639495 3.492616e-06
age          0.01848156 0.006027953  3.065977 2.169602e-03
diabetes     0.48714064 0.088177630  5.524538 3.303543e-08

> summary(glm(medcond ~ age*diabetes, family="binomial"))$coef
                Estimate  Std. Error   z value     Pr(>|z|)
(Intercept)  -2.51762428 0.478275141 -5.263966 1.409802e-07
age           0.02780399 0.007072117  3.931495 8.441917e-05
diabetes      2.83494074 0.913870625  3.102125 1.921369e-03
age:diabetes -0.03540210 0.013718549 -2.580601 9.862861e-03

> summary(glm(medcond ~ age*drinkany, family="binomial"))$coef
                 Estimate  Std. Error   z value   Pr(>|z|)
(Intercept)  -0.990678407 0.511047367 -1.938526 0.05255913
age           0.008847923 0.007588988  1.165890 0.24365904
drinkany     -1.439698282 0.831289196 -1.731886 0.08329383
age:drinkany  0.016793128 0.012391965  1.355163 0.17536577
```

2. Consider the following made up (partly) data set. For 26 people, I measured the following variables: amount of change in their pocket; number of coins in their pocket; and number of coins in their pocket that weren't quarters (so called "low" coins). Then I also noted whether they had enough money to buy a candy bar for $1.25 using just their change.

```
> amount
 [1] 1.37 1.01 1.50 0.56 0.61 3.06 5.42 1.75 5.40 0.56
[11] 0.34 2.33 3.34 1.14 1.02 1.72 0.45 0.85 3.46 5.58
[21] 2.16 5.01 0.32 0.39 2.41 3.64

> num.coins
 [1]  9 10  3  5 10 37 28  9 11  4  6 17 15 14 10  7  7 14 38 29 11
[22] 14  5  9 20 17

> num.lowcoins
 [1]  4  8  0  4  9 34  9  3  2  2  5 12 11  4  8  0  7 10 35 13  3
[22]  3  3  6 14 11
> candy
 [1] 1 0 1 0 0 1 1 1 1 0 0 1 1 0 0 1 0 0 1 1 1 1 0 0 1 1

> summary(glm(candy~num.coins,family="binomial"))$coef
             Estimate Std. Error   z value   Pr(>|z|)
(Intercept) -2.1775981  1.1895702 -1.830576 0.06716392
num.coins    0.2138296  0.1066116  2.005688 0.04488959

> summary(glm(candy~num.lowcoins,family="binomial"))$coef
               Estimate Std. Error    z value  Pr(>|z|)
(Intercept)  -0.30521199 0.62884461 -0.4853536 0.6274255
num.lowcoins  0.08017993 0.06975297  1.1494841 0.2503564

> summary(glm(candy~num.coins + num.lowcoins,family="binomial"))$coef
              Estimate Std. Error   z value   Pr(>|z|)
(Intercept)  -3.3288980  1.3715694 -2.427072 0.01522123
num.coins     0.6291259  0.2518994  2.497528 0.01250626
num.lowcoins -0.5153716  0.2520362 -2.044831 0.04087151
```

Notice that the directionality of the low coins changes when it is included in the model that already contains the number of coins total. Moral of the story: be very very very careful interpreting coefficients when you have multiple explanatory variables.