Module 10 notebook

Neil Yetz & Gemma Wallace

Clear environment

```r
rm(list = ls())
```

Demo activity

Load Libraries

```r
library(tidyverse)
```

```
## -- Attaching packages ---------------------------------------------------------------------------

## v ggplot2 3.3.0     v purrr   0.3.3
## v tibble  3.0.0     v dplyr   0.8.5
## v tidyr   1.0.2     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0

## Warning: package 'ggplot2' was built under R version 3.6.3

## Warning: package 'tibble' was built under R version 3.6.3

## Warning: package 'tidyr' was built under R version 3.6.3

## Warning: package 'dplyr' was built under R version 3.6.3

## Warning: package 'forcats' was built under R version 3.6.3

## -- Conflicts ------------------------------------------------------------------------------------
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
library(psych)
```

```
##
## Attaching package: 'psych'

## The following objects are masked from 'package:ggplot2':
##
##     %+%, alpha
```

```r
library(olsrr)
```

```
## Warning: package 'olsrr' was built under R version 3.6.3

##
## Attaching package: 'olsrr'

## The following object is masked from 'package:datasets':
##
```

```
##     rivers
```

Read in data

```r
lr <- read_csv("Logistic2.csv")
```

```
## Parsed with column specification:
## cols(
##   Y = col_double(),
##   X1 = col_double(),
##   X2 = col_double(),
##   X3 = col_double(),
##   X4 = col_double()
## )
```

```r
#lr <- mutate(lr, X2 = X2*10)
```

```r
#write_csv(lr, "Logistic2.csv", na = "")
```

Describe variables

```r
describe(lr)
```

```
##    vars   n mean   sd median trimmed  mad min max range  skew kurtosis   se
## Y     1 164 0.69 0.46    1.0    0.73 0.00 0.0 1.0   1.0 -0.81    -1.35 0.04
## X1    2 164 0.46 0.50    0.0    0.45 0.00 0.0 1.0   1.0  0.15    -1.99 0.04
## X2    3 164 6.15 1.76    6.7    6.25 1.63 1.1 8.9   7.8 -0.58     0.08 0.14
## X3    4 164 1.42 1.37    1.0    1.26 1.48 0.0 5.0   5.0  0.64    -0.56 0.11
## X4    5 164 3.17 1.01    3.0    3.33 1.48 0.0 4.0   4.0 -1.12     0.53 0.08
```

OLS Regression

Model 1

```r
ols_mod1 <- lm(Y ~ X1, data = lr)
ols_regress(ols_mod1)
```

```
##                          Model Summary
## --------------------------------------------------------------
## R                        0.439       RMSE               0.418
## R-Squared                0.193       Coef. Var         60.719
## Adj. R-Squared           0.188       MSE                0.175
## Pred R-Squared           0.174       MAE                0.346
## --------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                 ANOVA
## --------------------------------------------------------------------
##                Sum of
##               Squares         DF    Mean Square      F         Sig.
## --------------------------------------------------------------------
## Regression      6.785          1          6.785    38.764    0.0000
## Residual       28.355        162          0.175
```

```
## Total           35.140        163
## --------------------------------------------------------------------
##
##                                 Parameter Estimates
## -------------------------------------------------------------------------------------
##       model      Beta    Std. Error    Std. Beta      t        Sig      lower    upper
## -------------------------------------------------------------------------------------
## (Intercept)    0.500         0.045                 11.211    0.000    0.412    0.588
##          X1    0.408         0.066         0.439    6.226    0.000    0.279    0.537
## -------------------------------------------------------------------------------------
```

Intercept: When X1 is zero, the expected Y is .500. X1: For every one-unit increase X1, there is an expected .408 increase in Y.

This model explains 19.31% of the variance in Y.

Model 2

```r
ols_mod2 <- lm(Y ~ X1 + X2, data = lr)
ols_regress(ols_mod2)
```

```
##                          Model Summary
## --------------------------------------------------------------
## R                        0.442       RMSE                0.419
## R-Squared                0.196       Coef. Var          60.809
## Adj. R-Squared           0.186       MSE                 0.176
## Pred R-Squared           0.166       MAE                 0.345
## --------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## --------------------------------------------------------------------
##                Sum of
##               Squares         DF    Mean Square      F         Sig.
## --------------------------------------------------------------------
## Regression      6.876          2          3.438    19.585    0.0000
## Residual       28.264        161          0.176
## Total          35.140        163
## --------------------------------------------------------------------
##
##                                  Parameter Estimates
## ---------------------------------------------------------------------------------------
##       model       Beta    Std. Error    Std. Beta      t         Sig      lower    upper
## ---------------------------------------------------------------------------------------
## (Intercept)      0.584         0.125                  4.676    0.000     0.337    0.831
##          X1      0.405         0.066        0.436     6.157    0.000     0.275    0.535
##          X2     -0.013         0.019       -0.051    -0.721    0.472    -0.050    0.023
## ---------------------------------------------------------------------------------------
```

Intercept: When all predictors are zero, the expected Y is .584. X1: Holding all other variables constant; For every one-unit increase X1, there is an expected .405 increase in Y. X2: Holding all other variables constant; For every one-unit increase X2, there is an expected -.013 increase in Y.

This model explains 19.58% of the variance in Y. This is not much difference in R^2 as compared rto model 1.

Model 3

```r
ols_mod3 <- lm(Y ~ X1 + X2 + X3, data = lr)
ols_regress(ols_mod3)
```

```
##                         Model Summary
## ---------------------------------------------------------------
## R                       0.465       RMSE                0.415
## R-Squared               0.216       Coef. Var          60.210
## Adj. R-Squared          0.202       MSE                 0.172
## Pred R-Squared          0.176       MAE                 0.336
## ---------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## ---------------------------------------------------------------------
##                Sum of
##               Squares         DF    Mean Square      F         Sig.
## ---------------------------------------------------------------------
## Regression      7.603          3          2.534    14.724    0.0000
## Residual       27.538        160          0.172
## Total          35.140        163
## ---------------------------------------------------------------------
##
##                                   Parameter Estimates
## ----------------------------------------------------------------------------------------
##       model      Beta    Std. Error    Std. Beta       t        Sig      lower     upper
## ----------------------------------------------------------------------------------------
## (Intercept)     0.680         0.132                  5.143    0.000      0.419     0.941
##          X1     0.345         0.071        0.372     4.838    0.000      0.204     0.486
##          X2    -0.012         0.018       -0.046    -0.661    0.509     -0.049     0.024
##          X3    -0.053         0.026       -0.158    -2.054    0.042     -0.105    -0.002
## ----------------------------------------------------------------------------------------
```

Intercept: When all predictors are zero, the expected Y is .680. X1: Holding all other variables constant; For every one-unit increase X1, there is an expected .345 increase in Y. X2: Holding all other variables constant; For every one-unit increase X2, there is an expected -.0.12 increase in Y. X3: Holding all other variables constant; For every one-unit increase X3, there is an expected -.053 increase in Y.

This model explains 21.64% of the variance in Y. This is not much difference in R^2 as compared to model 1.

Model 4

```r
ols_mod4 <- lm(Y ~ X1 + X2 + X3 + X4, data = lr)
ols_regress(ols_mod4)
```

```
##                         Model Summary
## ---------------------------------------------------------------
## R                       0.518       RMSE                0.402
## R-Squared               0.268       Coef. Var          58.360
## Adj. R-Squared          0.250       MSE                 0.162
## Pred R-Squared          0.222       MAE                 0.314
## ---------------------------------------------------------------
```

```
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                               ANOVA
## --------------------------------------------------------------------
##               Sum of
##              Squares         DF    Mean Square      F         Sig.
## --------------------------------------------------------------------
## Regression      9.431          4          2.358    14.581    0.0000
## Residual       25.710        159          0.162
## Total          35.140        163
## --------------------------------------------------------------------
##
##                                   Parameter Estimates
## ---------------------------------------------------------------------------------------
##       model      Beta    Std. Error    Std. Beta      t        Sig      lower    upper
## ---------------------------------------------------------------------------------------
## (Intercept)     0.308         0.169                  1.823    0.070    -0.026    0.643
##          X1     0.286         0.071        0.308     4.012    0.000     0.145    0.427
##          X2    -0.013         0.018       -0.050    -0.734    0.464    -0.049    0.022
##          X3    -0.029         0.026       -0.086    -1.114    0.267    -0.081    0.023
##          X4     0.117         0.035        0.255     3.362    0.001     0.048    0.185
## ---------------------------------------------------------------------------------------
```

Intercept: When all predictors are zero, the expected Y is .316. X1: Holding all other variables constant; For every one-unit increase X1, there is an expected .287 increase in Y. X2: Holding all other variables constant; For every one-unit increase X2, there is an expected -.013 increase in Y. X3: Holding all other variables constant; For every one-unit increase X3, there is an expected -.028 increase in Y. X4: Holding all other variables constant; For every one-unit increase X4, there is an expected .115 increase in Y.

This model explains 26.87% of the variance in Y. This is not much difference in R^2 as compared rto model 1.

Hierarchical comparison

```r
anova(ols_mod1,
      ols_mod2,
      ols_mod3,
      ols_mod4)
```

```
## Analysis of Variance Table
##
## Model 1: Y ~ X1
## Model 2: Y ~ X1 + X2
## Model 3: Y ~ X1 + X2 + X3
## Model 4: Y ~ X1 + X2 + X3 + X4
##   Res.Df    RSS Df Sum of Sq       F    Pr(>F)
## 1    162 28.355
## 2    161 28.264  1   0.09126  0.5644 0.4536063
## 3    160 27.538  1   0.72643  4.4926 0.0355952 *
## 4    159 25.709  1   1.82804 11.3055 0.0009681 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Logistic regression

```
log_mod <- glm(Y ~ X1 + X2 + X3 + X4, family = binomial, data = lr)
summary(log_mod)
```

```
##
## Call:
## glm(formula = Y ~ X1 + X2 + X3 + X4, family = binomial, data = lr)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.3956  -0.7618   0.3744   0.7864   1.6046
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.86309    0.98886  -0.873 0.382766
## X1           1.77209    0.48405   3.661 0.000251 ***
## X2          -0.08569    0.11189  -0.766 0.443785
## X3          -0.15597    0.15370  -1.015 0.310210
## X4           0.59549    0.20668   2.881 0.003962 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 203.32  on 163  degrees of freedom
## Residual deviance: 155.59  on 159  degrees of freedom
## AIC: 165.59
##
## Number of Fisher Scoring iterations: 5
```

The model above displays the log odds of each predictor variable (While controlling for all other predictors in the model) on the outcome of Y. We can see that X1 and X4 are statistically significant. However, to have a better interpretation of each with odds ratios, we need to exponentiate the coefficients.

Get ORs & 95% confidence intervals

```
exp(coefficients(log_mod))
```

```
## (Intercept)          X1          X2          X3          X4
##   0.4218572   5.8831439   0.9178798   0.8555857   1.8139261
```

```
exp(confint(log_mod))
```

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

```
##                  2.5 %    97.5 %
## (Intercept) 0.05830423  2.898238
## X1          2.37624313 16.216290
## X2          0.73114151  1.138133
## X3          0.63116883  1.157472
## X4          1.22321689  2.762664
```

Intercept: When all of the X variables are zero, the odds are .421 of developing the outcome of Y (Or we can take the inverse and state the they are 2.38 times as likely NOT to develop the outcome of Y). This is not statistically significant. X1 (Binary Variable): After controlling for all variables in the model, Those coded as

1 are 5.88 times as likely to develop the outcome of Y as compared to those coded 0. This is statistically significant. X2 (Continuous): After controlling for all variables in the model, For every one unit increase in X2, there is an expected increase of 0.918 times of developing Y (Or we can take the inverse and state that for every one unit increase in X2, there is a 1.09 increase in the odds of NOT developing the outcome of Y). This is not statistically significant. X3: (Continuous): After controlling for all variables in the model, For every one unit increase in X3, there is an expected increase of 0.856 times in the odds of developing Y (Or we can take the inverse and state that for every one unit increase in X2, there is a 1.17 increase in the odds of NOT developing the outcome of Y). This is not statistically significant. X4: (Continuous): After controlling for all variables in the model, For every one unit increase in X4, there is an expected increase of 1.81 times in the odds of developing Y. This is statistically significant.

Deviancy test

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

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: Y
##
## Terms added sequentially (first to last)
##
##
##      Df Deviance Resid. Df Resid. Dev Pr(>Chi)
## NULL                   163     203.32
## X1    1   34.604       162     168.72 4.04e-09 ***
## X2    1    0.484       161     168.23 0.486443
## X3    1    3.694       160     164.54 0.054620 .
## X4    1    8.946       159     155.59 0.002781 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The difference between the deviance for each model and the null is one measure of model fit. These comparisons tell us whether adding information to a null model will lead to better prediction. Each row in the deviance table compares that model to the null model.

X1 = model with just X1 X2 = model wth X1 + X2 X3 = model with X1 + X2 + X3 X4 = model with X1 + X2 + X3 + X4

In this case, Only adding in X4 adds to the predictive power.

Calculate Mcfadden R^2

```
m1_mcfadden <- 1 - (168.72/203.32)
m2_mcfadden <- 1 - (168.23/203.32)
m3_mcfadden <- 1 - (164.54/203.32)
m4_mcfadden <- 1 - (155.59/203.32)

m1_mcfadden
```

```
## [1] 0.1701751
```

```
m2_mcfadden
```

```
## [1] 0.1725851
```

```r
m3_mcfadden
```

```
## [1] 0.1907338
```

```r
m4_mcfadden
```

```
## [1] 0.2347531
```

Compare conclusions from the OLS vs. logistic regression analyses

We get similar conclusions between the 2 analysis aproaches.

Try it yourself

Data prep

```r
#obs <- read_csv("bac_obs.csv")
# Create dichotomized versions of bac (>.08 vs. <= .08) and typ_drks (> average of 2 per day vs. <= ave
#obs <- mutate(obs, bac_over = ifelse(bac > .08, 1, 0))
#table(obs$bac_over)
#describe(obs$weight)
#obs <- mutate(obs, weight_low = ifelse(weight <= 58.54, 1, 0))
#write_csv(obs, "bac_module10.csv")
```

Import data

```r
bac <- read_csv("bac_module10.csv")
```

Use OLS regression to predict bac_over with alcexp, pmood, weight_low, and typ_drks

Build up models step by step

```r
ols_m1 <-lm(bac_over ~ alcexp, data = bac)
ols_regress(ols_m1)
```

```
##                         Model Summary
## ---------------------------------------------------------------
## R                       0.457       RMSE                 0.446
## R-Squared               0.209       Coef. Var           83.352
## Adj. R-Squared          0.205       MSE                  0.199
## Pred R-Squared          0.195       MAE                  0.397
## ---------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## ---------------------------------------------------------------------
##                Sum of
##               Squares         DF    Mean Square      F         Sig.
## ---------------------------------------------------------------------
## Regression     10.381          1         10.381     52.203     0.0000
```

```
## Residual       39.374        198          0.199
## Total          49.755        199
## -------------------------------------------------------------------
##
##                                   Parameter Estimates
## ----------------------------------------------------------------------------------------
##       model       Beta     Std. Error     Std. Beta       t        Sig       lower      upper
## ----------------------------------------------------------------------------------------
## (Intercept)     -0.662          0.169                  -3.925    0.000     -0.994     -0.329
##      alcexp      0.293          0.041         0.457     7.225    0.000      0.213      0.373
## ----------------------------------------------------------------------------------------
```

```
ols_m2 <-lm(bac_over ~ alcexp + pmood, data = bac)
ols_regress(ols_m2)
```

```
##                         Model Summary
## -------------------------------------------------------------
## R                         0.477       RMSE                 0.442
## R-Squared                 0.228       Coef. Var           82.562
## Adj. R-Squared            0.220       MSE                  0.195
## Pred R-Squared            0.207       MAE                  0.390
## -------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## -------------------------------------------------------------------
##                 Sum of
##                Squares         DF    Mean Square      F         Sig.
## -------------------------------------------------------------------
## Regression     11.319          2          5.660    29.009    0.0000
## Residual       38.436        197          0.195
## Total          49.755        199
## -------------------------------------------------------------------
##
##                                   Parameter Estimates
## ----------------------------------------------------------------------------------------
##       model       Beta     Std. Error     Std. Beta       t        Sig       lower      upper
## ----------------------------------------------------------------------------------------
## (Intercept)     -0.854          0.189                  -4.528    0.000     -1.227     -0.482
##      alcexp      0.278          0.041         0.433     6.810    0.000      0.197      0.358
##       pmood      0.050          0.023         0.139     2.193    0.029      0.005      0.095
## ----------------------------------------------------------------------------------------
```

```
ols_m3 <-lm(bac_over ~ alcexp + pmood + weight_low, data = bac)
ols_regress(ols_m3)
```

```
##                         Model Summary
## -------------------------------------------------------------
## R                         0.494       RMSE                 0.438
## R-Squared                 0.244       Coef. Var           81.896
## Adj. R-Squared            0.232       MSE                  0.192
## Pred R-Squared            0.215       MAE                  0.384
## -------------------------------------------------------------
```

```
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## --------------------------------------------------------------------
##                Sum of
##               Squares          DF    Mean Square      F         Sig.
## --------------------------------------------------------------------
## Regression     12.129           3          4.043     21.06    0.0000
## Residual       37.626         196          0.192
## Total          49.755         199
## --------------------------------------------------------------------
##
##                                   Parameter Estimates
## ----------------------------------------------------------------------------------------
##       model      Beta    Std. Error    Std. Beta      t        Sig      lower      upper
## ----------------------------------------------------------------------------------------
## (Intercept)    -0.834         0.187                 -4.450    0.000    -1.204     -0.464
##      alcexp     0.269         0.041        0.420     6.619    0.000     0.189      0.349
##       pmood     0.048         0.023        0.133     2.106    0.036     0.003      0.092
##  weight_low     0.182         0.089        0.128     2.053    0.041     0.007      0.357
## ----------------------------------------------------------------------------------------
```

```r
ols_m4 <-lm(bac_over ~ alcexp + pmood + weight_low + typ_drks, data = bac)
ols_regress(ols_m4)
```

```
##                          Model Summary
## --------------------------------------------------------------
## R                        0.609       RMSE                0.401
## R-Squared                0.371       Coef. Var          74.880
## Adj. R-Squared           0.358       MSE                 0.160
## Pred R-Squared           0.340       MAE                 0.336
## --------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## --------------------------------------------------------------------
##                Sum of
##               Squares          DF    Mean Square      F         Sig.
## --------------------------------------------------------------------
## Regression     18.460           4          4.615    28.756    0.0000
## Residual       31.295         195          0.160
## Total          49.755         199
## --------------------------------------------------------------------
##
##                                   Parameter Estimates
## ----------------------------------------------------------------------------------------
##       model      Beta    Std. Error    Std. Beta      t        Sig      lower      upper
## ----------------------------------------------------------------------------------------
## (Intercept)    -0.891         0.172                 -5.194    0.000    -1.230     -0.553
##      alcexp     0.103         0.046        0.160     2.251    0.025     0.013      0.193
##       pmood     0.054         0.021        0.151     2.606    0.010     0.013      0.095
```

```
##  weight_low     0.207         0.081        0.146     2.551    0.012     0.047     0.367
##    typ_drks     0.015         0.002        0.439     6.281    0.000     0.010     0.020
## ---------------------------------------------------------------------------------------
```

Compare the OLS models by comparing R^2 values

Model 1 vs Model 2: R^2 increases from 0.209 to 0.228, indicating that adding pmood when controlling for alc_exp does not add much explained variance (~2%) to the model.

Model 2 vs. Model 3: R^2 increases from 0.228 to 0.244, indicating that adding weight_low when controlling for alcexp and pmood does not add much explained variance (~2%) to the model

Mode 3 vs. Model 4: R^2 increases from 0.244 to 0.371, indicating that about 13% additional varince in bac_over is explained when you add typ_drks to the model while controlling for alc_exp, pmood, and weight_low.

Compare the OLS models via significance testing with hierarchical regression

```r
anova(ols_m1,
      ols_m2,
      ols_m3,
      ols_m4)
```

```
## Analysis of Variance Table
##
## Model 1: bac_over ~ alcexp
## Model 2: bac_over ~ alcexp + pmood
## Model 3: bac_over ~ alcexp + pmood + weight_low
## Model 4: bac_over ~ alcexp + pmood + weight_low + typ_drks
##   Res.Df    RSS Df Sum of Sq       F    Pr(>F)
## 1    198 39.374
## 2    197 38.436  1    0.9385  5.8476   0.01652 *
## 3    196 37.626  1    0.8091  5.0415   0.02587 *
## 4    195 31.295  1    6.3312 39.4498 2.146e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The significance tesing indicates that each subsequwnt model explains a significantly larger percent of the variance in Y than the previous model. Keep in mind that, even though the increase in R^2 is significant, 2% is not a huge increase.

Use Logistic Regression to predict bac_over with alcexp, pmood, weight_low, and typ_drks

```r
logreg_mod <- glm(bac_over ~ alcexp + pmood + weight_low + typ_drks, data = bac)
summary(logreg_mod)
```

```
##
## Call:
## glm(formula = bac_over ~ alcexp + pmood + weight_low + typ_drks,
##     data = bac)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -1.06738  -0.32135   0.05886   0.31073   0.89553
##
```

```
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)    
## (Intercept) -0.891359   0.171627  -5.194 5.17e-07 ***
## alcexp       0.102739   0.045641   2.251  0.02550 *  
## pmood        0.053904   0.020684   2.606  0.00987 ** 
## weight_low   0.207061   0.081163   2.551  0.01150 *  
## typ_drks     0.014957   0.002381   6.281 2.15e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## (Dispersion parameter for gaussian family taken to be 0.1604883)
## 
##     Null deviance: 49.755  on 199  degrees of freedom
## Residual deviance: 31.295  on 195  degrees of freedom
## AIC: 208.6
## 
## Number of Fisher Scoring iterations: 2
```

The model above displays the log odds of each predictor variable (While controlling for all other predictors in the model) on the outcome of bac_over. We can see that all of the predictors statistically significant. However, to have a better interpretation of each with odds ratios, we need to exponentiate the coefficients.

Exponentiate coefficients to get odds ratios and confidence intervals

```r
exp(coefficients(logreg_mod))
```

```
## (Intercept)      alcexp       pmood  weight_low    typ_drks 
##   0.4100982   1.1082027   1.0553834   1.2300581   1.0150698
```

```r
exp(confint(logreg_mod))
```

```
##                 2.5 %    97.5 %
## (Intercept) 0.2929537 0.5740856
## alcexp      1.0133733 1.2119060
## pmood       1.0134537 1.0990480
## weight_low  1.0491552 1.4421537
## typ_drks    1.0103431 1.0198187
```

Intercept: When all of the X variables are zero, the odds of having a bac greater than 0.08 are 0.41 times as likely. Or we can take the inverse and state the they are 2.43 times as likely NOT to develop the outcome of bac over 0.08 (calculated by dividing 1/0.41). The confidence interval does not include 1, indicating that this is statistically significant.

alcexp (continuous): for every one-unit increase in alcexp, the odds of having a bac > 0.08 increased by 1.12. In other words, for every one-unit increase in alcexp, participants are 1.12 times more likely to have a bac > 0.08. The confidence interval does not include 1, indicating that this is statistically significant.

pmood (continuous): when controlling for alcexp, for every one-unit increase in pmood, the odds of hacing a bac > 0.08 increased by 1.05. In other words, for every one-unit increase in pmood, participants are 1.05 times more likely to have bac > 0.08. The confidence interval does not include 1, indicating that this is statistically significant.

weight_low (binary): when controlling for alcexp and pmood, participants with lower body weight (coded as 1) wre 1.23 times as likely to have a bac > 0.08 than participants who did not have lower body weight. The confidence interval does not include 1, indicating that this is statistically significant.

typ_drks (continuous): when controlling for alcexp, pmood, and weight_low, for a one-unit increase in typ_drks the odds of having a bac > 0.08 increase by 1.015. In other wors, for every one-unit increase in typ_drks, participants are 1.015 times as likely to have a bac > 0.08. The confidence interval does not include 1, indicating that this is statistically significant.

Deviance testing

```r
anova(logreg_mod,test="Chisq")
```

```
## Analysis of Deviance Table
##
## Model: gaussian, link: identity
##
## Response: bac_over
##
## Terms added sequentially (first to last)
##
##
##            Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
## NULL                         199     49.755
## alcexp      1  10.3810       198     39.374 8.794e-16 ***
## pmood       1   0.9385       197     38.436   0.01560 *
## weight_low  1   0.8091       196     37.626   0.02475 *
## typ_drks    1   6.3312       195     31.295 3.366e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The difference between the deviance for each model and the null is one measure of model fit. These comparisons tell us whether adding information to a null model will lead to better prediction. Each row in the deviance table compares that model to the null model. alcexp = model with just alcecp pmood = model wth alcexp + pmood weight_low = model with alcexp + pmood + weight_low typ_drks = model with alcexp + pmood + weight_low + typ_drks

In this case, adding each variable adds to the predictive power of the model (i.e., reducing model deviance).

Calculate McFadden's R^2 from deviance table

This serves as an effect size! McFadden R2 = 1-(Deviance model/Deviance Null)

```r
# get the null deviance from the original model output
# get the deviance for each model from the residual deviance column of the devaince table

m1_mcfaddens <- 1-(39.374/49.755)
m1_mcfaddens
```

```
## [1] 0.2086423
```

```r
m2_mcfaddens  <- 1-(38.436/49.755)
m2_mcfaddens
```

```
## [1] 0.2274947
```

```r
m3_mcfaddens <- 1-(37.626/49.755)
m3_mcfaddens
```

```
## [1] 0.2437745
```

```r
m4_mcfadens <- 1-(31.295/49.755)
m4_mcfadens
```

```
## [1] 0.371018
```

Now we can compare the McFadden's R^2 to answer the same questions we asked about with hierarchical regression the the OLS models.

Model 1 vs Model 2: McFaddens R^2 increases from 0.209 to 0.228, indicating that adding pmood when controlling for alc_exp does not add much explained variance (~2%) to the model.

Model 2 vs. Model 3: R^2 increases from 0.228 to 0.244, indicating that adding weight_low when controlling for alcexp and pmood does not add much explained variance (~2%) to the model

Mode 3 vs. Model 4: R^2 increases frorm 0.244 to 0.371, indicating that about 13% additional varince in bac_over is explained when you add typ_drks to the model while controlling for alc_exp, pmood, and weight_low.

Note, these values are essentially identical to what we got from the R^2 vaues in the OLS models above!

Compare conclusions from the OLS vs. logistic regression analyses

We get similar conclusions between the 2 analysis aproaches.