

Midterm Bonus Question

Ravin Lathigra

2019-03-18

Model 1: No Interaction Terms

$$low = lwt + smoke$$

The first model has 3 coefficients:

- Intercept = 0.62199682
- lwt = -0.01332433
- smoke = 0.67667325

It is immediately obvious that as **lwt** increases, the odds of having a birth weight, **low**, less than 2.5kg decreases. This is shown by the negative coefficient. Smoking on the otherhand increases the odds.

Logistic regression models the log odds therefore, to gain more insight into the interpretation of the coefficients, we exponentiate the coefficients to gather the change in odds holding other features constant. Table 1 shows that holding smoking status constant, a unit increase in lwt decreases 1%. The presence of smoking on the other hand increases the odds by 97%.

$$\frac{p}{(1-p)} = e^{\beta_0} e^{\beta_1(lwt)} e^{\beta_2(smoke)}$$

Table 1: Change in Odds | Model 1

Change in Odds	
(Intercept)	0.86
lwt	-0.01
smoke	0.97

Model 2: Interaction Model - 1

$$low = \beta_0 + (\beta_1 + \beta_2(smoke))(lwt)$$

The second model has 3 coefficients:

- Intercept = 1.00969013
- lwt = -0.01664904
- lwt:smokeyes = 0.00592139

When an interaction term is introduced, interpretation of the coefficients becomes more complicated. We now have 2 cases we need to consider i.e. smoking is present and smoking is not present. Table 2 shows the change in odds considering these two cases for a unit change in **lwt**. The odds of having a birth that is less than 2.5kg decreases 1% faster when smoking is not present for a unit change in **lwt**. Since **smokeyes** is a binary feature it makes sense to see how odds change with **lwt** under both scenarios. **Lwt** on the other hand is a continuous feature. Assessing how odds change with smoking status conditioned on **lwt** cannot be separated into cases. Instead, we could make statements such as: *the odds of having a low weight birth change by _____ when smoking status equals yes and **lwt** is ____.*

Table 2: Change in Odds | Model 2

	Smoking Present	Non-Smoking
lwt	-0.01	-0.02

Model 3: Interaction Model - 2

Model 3:

$$low = \beta_0 + \beta_1(lwt) + \beta_2(smoke) + \beta_3(lwt)(smoke)$$

We can rearrange it to be as follows:

$$low = \beta_0 + (\beta_1 + \beta_3(smoke))(lwt) + \beta_2(smoke)$$

The third model has 3 coefficients:

- Intercept = 1.93233640
- lwt = -0.02388637
- smoke = -1.51088524
- lwt:smokeyes = 0.01756803

When an interaction term is introduced, interpretation of the coefficients becomes more complicated. We now have 2 cases we need to consider i.e. smoking is present and smoking is not present. Table 3 shows the change in odds considering these two cases for a unit change in `lwt`. The odds of having a birth that is less than 2.5kg decreases 1% faster when smoking is not present for a unit change in `lwt`. We could if we wanted to see the change in odds when smoking is present by setting `lwt` equal to 0, however 0 is not in the possible domain for `lwt` so similar to model two we can make statements about odds if we assume a level for `lwt` but that is it.

Table 3: Change in Odds | Model 3

	Smoking Present	Non-Smoking
lwt	-0.01	-0.02

```
library(corrplot)      #Visualize Correlation between variables
library(kableExtra)    #Style tables
library(tidyverse)     #contains ggplot2,dplyr,tidyr, readr, purr, tibble, stringr, forcats
library(formatR)       #Improve readability of code
library(e1071)         #Functions for latent class analysis, Fourier transform ect.
library(VIM)           #Knn
library(ggfortify)     #Add on to ggplot2 to allow for more plot types
library(Rtsne)         #Dimension reduction classification
library(caret)         #streamlined model development
library(RColorBrewer)  #Control colours of visualizations
library(GGally)        #Contains ggpairs plots
library(lmtest)        #Test for linear assumptions
library(MASS)
library(faraway)
library(cowplot)
library(pROC)
```

```

library(ggpmisc)
library(plotROC)

data(birthwt)
fit_glm_1 <- glm(low ~ lwt + smoke, family=binomial, data=birthwt)
glm_1_coef <- coefficients(fit_glm_1)

chg_odds_1 <- data.frame(round(exp(glm_1_coef)-1,2))
colnames(chg_odds_1) = "Change in Odds"

kable(chg_odds_1,
      booktabs = T,
      caption = "Change in Odds | Model 1",
      align = rep("c",ncol(chg_odds_1))) %>%
  kable_styling(position = "center", latex_options = "hold_position")

fit_glm_2 <- glm(low ~ lwt + lwt:smoke, family=binomial, data=birthwt)
glm_2_coef <- coefficients(fit_glm_2)

chg_odds_2 <- data.frame(round(exp(glm_2_coef[2]+glm_2_coef[3])-1,2),round(exp(glm_2_coef[2])-1,2))
colnames(chg_odds_2) = c("Smoking Present","Non-Smoking")

kable(chg_odds_2,
      booktabs = T,
      caption = "Change in Odds | Model 2",
      align = rep("c",ncol(chg_odds_2))) %>%
  kable_styling(position = "center", latex_options = "hold_position")

fit_glm_3 <- glm(low ~ lwt + smoke + lwt:smoke, family=binomial, data=birthwt)
glm_3_coef <- coefficients(fit_glm_3)

chg_odds_3 <- data.frame(round(exp(glm_3_coef[2]+glm_3_coef[4])-1,2),round(exp(glm_3_coef[2])-1,2))
colnames(chg_odds_3) = c("Smoking Present","Non-Smoking")

kable(chg_odds_3,
      booktabs = T,
      caption = "Change in Odds | Model 3",
      align = rep("c",ncol(chg_odds_3))) %>%
  kable_styling(position = "center", latex_options = "hold_position")

```