Homework 4

STA 138 | Camden Possinger | Winter 2022

1. For the model $\log(\frac{\pi}{1-\pi}) = \alpha + \beta_1 x_1 + \beta_2 x_2$

```
angina <- read.csv("angina.csv")
angina %<>% select(y,myofam,age)

angina$myofam %>% as.factor
```

```
##   [1] no  no  yes no  yes no  no  no  yes yes no  no  no  yes no  no  yes yes
##  [19] no  no  yes no  no  no  no  no  no  no  no  no  no  yes yes yes yes no
##  [37] no  no  no  yes no  yes yes yes yes no  yes no  no  no  no  no  yes yes
##  [55] yes no  yes yes no  no  no  yes no  no  yes no  yes yes yes no  no  no
##  [73] no  no  yes no  no  yes no  yes yes no  no  yes yes yes no  no  no  no
##  [91] yes no  no  yes yes yes yes no  no  no  no  no  no  no  no  yes no  yes
## [109] no  no  no  no  no  no  no  no  no  yes no  no  yes no  yes yes no  no
## [127] no  no  no  no  no  yes no  no  no  no  yes no  yes no  no  no  yes yes
## [145] no  yes no  no  no  yes no  no  yes no  yes no  no  yes yes no  no  no
## [163] yes no  no  yes yes no  yes no  no  no  no  no  no  yes no  no  no  no
## [181] no  no  no  no  no  no  no  no  no  no  no  no  yes yes no  no  no  no
## [199] yes no
## Levels: no yes
```

(a) What are the estimated parameters for this model?

```
no_interact <- glm(y ~ myofam + age, data = angina, family = "binomial")
est_params <- data.frame("Parameter" = c("Alpha","Beta 1","Beta 2"),
                         "Estimated_Values" = no_interact$coefficients)

est_params %>% kbl
```

	Parameter	Estimated_Values
(Intercept)	Alpha	-5.2783645
myofamyes	Beta 1	2.1302028
age	Beta 2	0.0887068

(b) Interpret the parameters. From this model, what would you estimate to be the probability of angina for a patient with age 50 and with a family history of myocardial infarction?

$\hat{\alpha}$ is the estimated log-odds ratio when a patient does not have a family history of myocardial infarction and has 0 age

$\hat{\beta}_1$ is the estimated log-odds ratio of angina between when a patient does have a family history of myocardial infarction and when a patient does not keeping age constant.

$\hat{\beta}_2$ is the estimated log-odds ratio of angina based on a 1 year change in age keeping everything else constant.

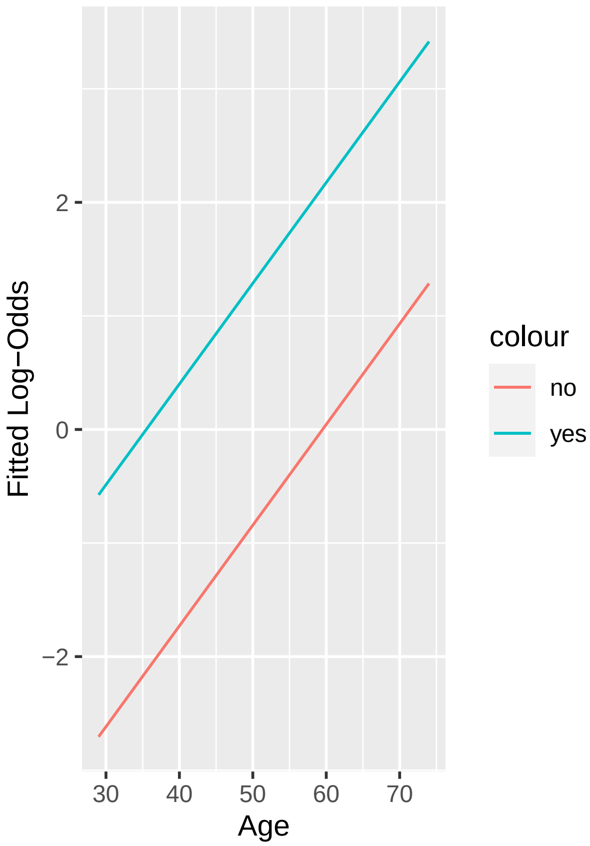
```r
est_odds <- no_interact %>% predict(newdata = list("myofam" = "yes", "age" = 50)) %>% unname

est_prob <- est_odds %>% plogis

print(est_prob)
```

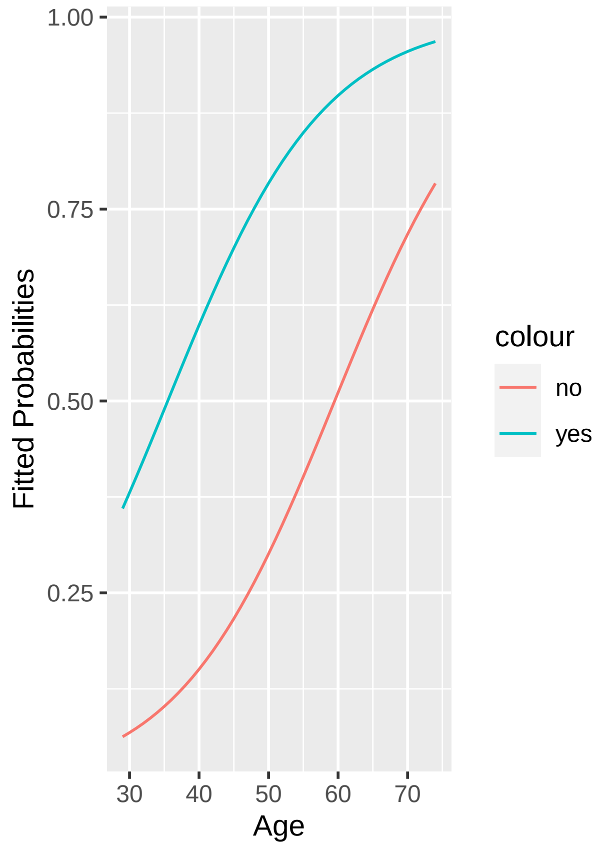
```
## [1] 0.7836694
```

Here the estimated probability of this individual who has a family history of myocardial infarction and is aged 50 is about 0.78, which is fairly high.

(c) Plot both the fitted log-odds and fitted probability of angina for patients as a function of age and myofam status.

```r
max_age <- angina$age %>% max
min_age <- angina$age %>% min

age <- min_age:max_age


fitted_odds_1 <- no_interact %>% predict(newdata = data.frame("age" = age, "myofam" = rep("yes",age %>%
fitted_odds_0 <- no_interact %>% predict(newdata = data.frame("age" = age, "myofam" = rep("no",age %>% l

fitted_probs_1 <- fitted_odds_1 %>% plogis
fitted_probs_0 <- fitted_odds_0 %>% plogis

plot_data <- data.frame(age,fitted_odds_1,fitted_odds_0,fitted_probs_1,fitted_odds_0)

fitted_odds_plot <- ggplot()+geom_line(data = plot_data, aes(x = age, y = fitted_odds_1,color = "yes"))+
  geom_line(data = plot_data, aes(x = age, y = fitted_odds_0,color = "no"))+
  xlab("Age")+
  ylab("Fitted Log-Odds")

fitted_probs_plot <-  ggplot()+geom_line(data = plot_data, aes(x = age, y = fitted_probs_1,color = "yes
  geom_line(data = plot_data, aes(x = age, y = fitted_probs_0,color = "no"))+
  xlab("Age")+
  ylab("Fitted Probabilities")


ggarrange(fitted_odds_plot, fitted_probs_plot, nrow = 1)
```





(d) Under this model, estimate the odds ratio of angina for a patient if they did have a family history of myocardial infarction vs. the same patient if they didn't have such a history. Does it depend on age?

```
est_odds_ratios <- data.frame( "Parameter" = c("Alpha","Beta 1","Beta 2"),
                              "Estimated_Odds_Ratios" = no_interact$coefficients %>% exp)

est_odds_ratios %>% kbl
```

	Parameter	Estimated_Odds_Ratios
(Intercept)	Alpha	0.0051008
myofamyes	Beta 1	8.4165739
age	Beta 2	1.0927602

In this case the estimated log odds ratios for angina for a patient shows that as the patient ages the odds of angina increases by about 9% while if a patient has a family history of myocardial infarction the odds of angina increases by about 742%. Angina does depend on age, but mostly depends on a family history of myocardial infarction. Here the odds ratio for family history of myocardial infarction does not depend on age.

$\frac{exp(-3.1481617+(0.0887068)x_2)}{exp(-5.2783645+(0.0887068)x_2)}$ does not depend on x_2

2. For the model $\log(\frac{\pi}{1-\pi}) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 \cdot x_2$

(a) What are the estimated parameters for this model?

```
interact <- glm(y ~ myofam + age + myofam:age, data = angina, family = "binomial")
est_params <- data.frame("Parameter" = c("Alpha","Beta 1","Beta 2","Beta 3"),
                         "Estimated_Values" = interact$coefficients)

est_params %>% kbl
```

	Parameter	Estimated_Values
(Intercept)	Alpha	-4.0470314
myofamyes	Beta 1	-3.1408954
age	Beta 2	0.0655033
myofamyes:age	Beta 3	0.1060486

(b) Interpret the parameters. From this model, what would you estimate to be the probability of angina for a patient with age 50 and with a family history of myocardial infarction?

$\hat{\alpha}$ is the estimated log odds-ratio of angina when both age $= 0$ and a patient does not have a family history of myocardial infarction.

$\hat{\beta}_1$ is the estimated log odds-ratio of angina between if a patient has a family history of myocardial infarction or not when age $= 0$.

$\hat{\beta}_2$ is the estimated log odds-ratio of angina between a one year difference in a patient's age when a patient does not have a family history of myocardial infarction.

$\hat{\beta}_3$ quantifies the interaction effect between age and family history of myocardial infarction on the presence of angina.

```
est_odds <- interact %>% predict(newdata = list("myofam" = "yes", "age" = 50, "myofamyes:age" = 50)) %>%

est_prob <- est_odds %>% plogis

print(est_prob)
```

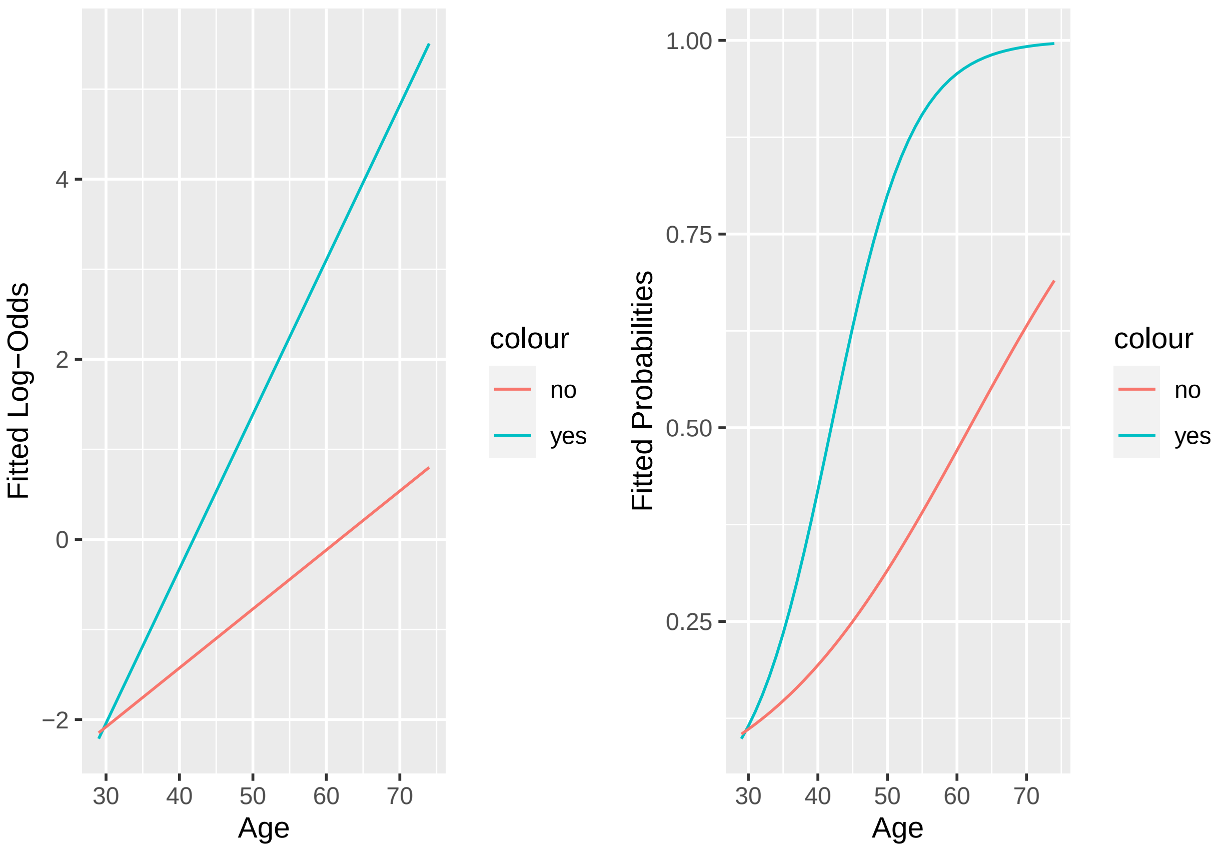
```
## [1] 0.8005392
```

Here the estimated probability of this individual who has a family history of myocardial infarction and is aged 50 is about 0.80, which is still high and doesn't change much from the the model in question 1

(c) Plot both the fitted log-odds and fitted probability of angina for patients as a function of age and myofam status.

```
fitted_odds_1 <- interact %>% predict(newdata = data.frame("age" = age, "myofam" = rep("yes",age %>% ler
fitted_odds_0 <- interact %>% predict(newdata = data.frame("age" = age, "myofam" = rep("no",age %>% leng

fitted_probs_1 <- fitted_odds_1 %>% plogis
fitted_probs_0 <- fitted_odds_0 %>% plogis

plot_data <- data.frame(age,fitted_odds_1,fitted_odds_0,fitted_probs_1,fitted_odds_0)

fitted_odds_plot <- ggplot()+geom_line(data = plot_data, aes(x = age, y = fitted_odds_1,color = "yes"))+
  geom_line(data = plot_data, aes(x = age, y = fitted_odds_0,color = "no"))+
  xlab("Age")+
  ylab("Fitted Log-Odds")
```

```
fitted_probs_plot <-  ggplot()+geom_line(data = plot_data, aes(x = age, y = fitted_probs_1,color = "yes
  geom_line(data = plot_data, aes(x = age, y = fitted_probs_0,color = "no"))+
  xlab("Age")+
  ylab("Fitted Probabilities")


ggarrange(fitted_odds_plot, fitted_probs_plot, nrow = 1)
```



(d) Under this model, estimate the odds ratio of angina for a patient if they did have a family history of myocardial infarction vs. the same patient if they didn't have such a history. Does it depend on age?

```
est_odds_ratios <- data.frame("Parameter" = c("Alpha","Beta 1","Beta 2","Beta 3"),
                              "Estimated_Odds_Ratios" = interact$coefficients %>% exp)

est_odds_ratios %>% kbl
```

	Parameter	Estimated_Odds_Ratios
(Intercept)	Alpha	0.0174742
myofamyes	Beta 1	0.0432441
age	Beta 2	1.0676962
myofamyes:age	Beta 3	1.1118759

```
interact_sum <- interact %>% summary
p_values <- data.frame("Parameter" = c("Alpha","Beta 1","Beta 2","Beta 3"),
                              "p-values" =  interact_sum$coefficients[,4] )
```

```
p_values %>% kbl
```

	Parameter	p.values
(Intercept)	Alpha	0.0004475
myofamyes	Beta 1	0.2340170
age	Beta 2	0.0023232
myofamyes:age	Beta 3	0.0483167

Here the odds-ratio depends on age since the interaction term is included.

$\frac{exp(-7.1879268+(0.1715519)x_2)}{exp(-4.0470314+(0.0655033)x_2)}$ depends on x_2