W271 Assignment 3

Emanuel Mejía

Contents


```r
library(tidyverse)
library(stargazer)
library(car)
```

1 Customer churn study: Part-3 (100 Points)

In the last two homework assignments, you initiated modeling a binary variable and used logistic regression to study the churn tendencies of customers.

Now, in Part-3, we're going to explore different interactions, transformations, and categorical explanatory variables to create a more comprehensive model.

```r
telcom_churn <- read.csv("./data/Telco_Customer_Churn.csv", header=T,na.strings=c("","NA"))
```

For the remainder of this section, pay particular attention to all variables.

1.1 Data Preprocessing (5 Points)

In this section, Convert variables as needed, and manage any missing values.

We'll convert the following variables and treat categorical explanatory variables as factors.

```r
telcom_churn["Churn"][telcom_churn["Churn"] == "No"] <- 0
telcom_churn["Churn"][telcom_churn["Churn"] == "Yes"] <- 1
telcom_churn["SeniorCitizen"][telcom_churn["SeniorCitizen"] == 0] <- "No"
telcom_churn["SeniorCitizen"][telcom_churn["SeniorCitizen"] == 1] <- "Yes"
```

```r
telcom_churn$Churn <- as.integer(telcom_churn$Churn)
telcom_churn$SeniorCitizen <- as.factor(telcom_churn$SeniorCitizen)
telcom_churn$gender <- as.factor(telcom_churn$gender)
```

Finally, we'll omit all the NA values since we have enought data, even without them.

```r
telcom_churn <- na.omit(telcom_churn)
```

1.2 Estimate a logistic regression (10 Points)

Estimate the following binary logistic regressions and report the results in a table using stargazer package.

$$
\begin{aligned}
Churn &= \beta_0 + \beta_1 tenure + \beta_2 MonthlyCharges + \beta_3 TotalCharges + \beta_4 SeniorCitizen + \beta_5 gender + e && \text{(Model 1)} \\
Churn &= \beta_0 + \beta_1 tenure + \beta_2 MonthlyCharges + \beta_3 TotalCharges + \beta_4 SeniorCitizen + \beta_5 gender && \text{(Model 2)} \\
&\quad + \beta_6 tenure^2 + \beta_7 MonthlyCharges^2 + \beta_8 TotalCharges^2 + e \\
Churn &= \beta_0 + \beta_1 tenure + \beta_2 MonthlyCharges + \beta_3 TotalCharges + \beta_4 SeniorCitizen + \beta_5 gender && \text{(Model 3)} \\
&\quad + \beta_6 tenure^2 + \beta_7 MonthlyCharges^2 + \beta_8 TotalCharges^2 \\
&\quad + \beta_9 SeniorCitizen \times tenure + \beta_{10} SeniorCitizen \times MonthlyCharges \\
&\quad + \beta_{11} SeniorCitizen \times TotalCharges + \beta_{12} gender \times tenure \\
&\quad + \beta_{13} gender \times MonthlyCharges + \beta_{14} gender \times TotalCharges + e
\end{aligned}
$$

- where $SeniorCitizen \times MonthlyCharges$ denotes the interaction between `SeniorCitizen` and `MonthlyCharges` variables.

So we have the following 3 models:

```r
mod.fit1 <- glm(
  formula = Churn ~ tenure + MonthlyCharges + TotalCharges + SeniorCitizen + gender,
  family = binomial(link = logit),
  data = telcom_churn
)

mod.fit2 <- glm(
  formula = Churn ~ tenure + MonthlyCharges + TotalCharges + SeniorCitizen + gender +
    I(tenure^2) + I(MonthlyCharges^2) + I(TotalCharges^2),
  family = binomial(link = logit),
  data = telcom_churn
)

mod.fit3 <- glm(
  formula = Churn ~ tenure + MonthlyCharges + TotalCharges + SeniorCitizen + gender +
    I(tenure^2) + I(MonthlyCharges^2) + I(TotalCharges^2) + SeniorCitizen:tenure + SeniorCitizen:Monthly
    SeniorCitizen:TotalCharges + gender:tenure + gender:MonthlyCharges + gender:TotalCharges,
  family = binomial(link = logit),
  data = telcom_churn
)
```

Table 1: Results

	Dependent variable:		
	Churn		
	(1)	(2)	(3)
Intercept	-1.581^{***}	-1.241^{***}	-1.358^{***}
	(0.122)	(0.201)	(0.236)
Tenure	-0.068^{***}	-0.125^{***}	-0.123^{***}
	(0.005)	(0.013)	(0.014)
MonthlyCharges	0.028^{***}	0.023^{***}	0.024^{***}
	(0.002)	(0.007)	(0.007)
TotalCharges	0.0002^{**}	0.001^{***}	0.001^{***}
	(0.0001)	(0.0002)	(0.0002)
SeniorCitizen(Yes)	0.630^{***}	0.634^{***}	1.477^{***}
	(0.079)	(0.080)	(0.399)
Gender(Male)	-0.004	-0.007	0.247
	(0.062)	(0.063)	(0.235)
Tenure**2		0.001^{***}	0.001^{***}
		(0.0001)	(0.0001)
MonthlyCharges**2		0.00003	0.0001
		(0.0001)	(0.0001)
TotalCharges**2		-0.00000^{***}	-0.00000^{***}
		(0.00000)	(0.00000)
Inter: Tenure | SeniorCitizen(Yes)			0.013
			(0.013)
Inter: MonthlyCharges | SeniorCitizen(Yes)			-0.013^{**}
			(0.005)
Inter: TotalCharges | SeniorCitizen(Yes)			-0.0001
			(0.0002)
Inter: Tenure | Gender(Male)			-0.010
			(0.010)
Inter: MonthlyCharges | Gender(Male)			-0.006^{*}
			(0.003)
Inter: TotalCharges | Gender(Male)			0.0002^{*}
			(0.0001)
Observations	7,032	7,032	7,032
Log Likelihood	-3,156.802	-3,138.899	-3,126.703
Akaike Inf. Crit.	6,325.604	6,295.799	6,283.406
Note:			*p<0.1; **p<0.05; ***p<0.01

1.3 Test a hypothesis: linear effects (15 Points)

Using `Model 1`, test the hypothesis of linear effects of variables on customer churn using a likelihood ratio test.

So we have the following:

$$Model_1 : Churn = \beta_0 + \beta_1 tenure + \beta_2 MonthlyCharges + \beta_3 TotalCharges + \beta_4 SeniorCitizen + \beta_5 gender + e$$

And we'll perform an Anova Test to test each variable's effect on customer churn using LRT:

```
Anova.tests <- Anova(mod.fit1, test = "LR")
LR.chisq <- Anova.tests[["LR Chisq"]]
p.val.chisq <- Anova.tests[["Pr(>Chisq)"]]
Anova.tests
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: Churn
##                LR Chisq Df Pr(>Chisq)
## tenure          192.288  1  < 2.2e-16 ***
## MonthlyCharges  289.800  1  < 2.2e-16 ***
## TotalCharges      6.021  1    0.01414 *
## SeniorCitizen    62.612  1  2.517e-15 ***
## gender            0.004  1    0.94700
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The results are as follows:

- For the test of `tenure` with $H_0 : \beta_1 = 0$ VS $H_a : \beta_1 \neq 0$ we obtain $-2log(\Delta) = 192.2879$ and a p-value of $\mathbb{P}(A > 192.2879) = 1.0068245 \times 10^{-43}$ meaning that we have a strong evidence that `tenure` is important to include in the model (given that the other variables are in the model as well).
- For the test of `MonthlyCharges` with $H_0 : \beta_2 = 0$ VS $H_a : \beta_2 \neq 0$ we obtain $-2log(\Delta) = 289.7996$ and a p-value of $\mathbb{P}(A > 289.7996) = 5.4981731 \times 10^{-65}$ meaning that we have strong evidence that `MonthlyCharges` is important to include in the model (given that the other variables are in the model as well).
- For the test of `TotalCharges` with $H_0 : \beta_3 = 0$ VS $H_a : \beta_3 \neq 0$ we obtain $-2log(\Delta) = 6.0211$ and a p-value of $\mathbb{P}(A > 6.0211) = 0.0141358$ meaning that we have marginal evidence that `TotalCharges` is important to include in the model (given that the other variables are in the model as well). So, depending on the confidence level we select we might include this variable or not (with $\alpha = 0.05$ we would include it, and we wouldn't include it if using $\alpha = 0.01$).
- For the test of `SeniorCitizen` with $H_0 : \beta_4 = 0$ VS $H_a : \beta_4 \neq 0$ we obtain $-2log(\Delta) = 62.6124$ and a p-value of $\mathbb{P}(A > 62.6124) = 2.5165905 \times 10^{-15}$ meaning that we have strong evidence that `Senior` is important to include in the model (given that the other variables are in the model as well).
- For the test of `Gender` with $H_0 : \beta_5 = 0$ VS $H_a : \beta_5 \neq 0$ we obtain $-2log(\Delta) = 0.0044$ and a p-value of $\mathbb{P}(A > 0.0044) = 0.9470004$ meaning that we have NO evidence that `Gender` is important to include in the model (given that the other variables are in the model as well). Meaning that we would fail to reject $H_0 : \beta_5 = 0$.

1.4 Test a hypothesis: Non linear effect (15 Points)

Perform a likelihood ratio test to assess the hypothesis that $\beta_6 = 0$, $\beta_7 = 0$, and $\beta_8 = 0$ within the context of `Model 2`. Interpret the implications of this test result in the context of the estimated `Model 2`.

Then, test the same hypothesis in `Model 3` using a likelihood ratio test. Interpret what this test result means in the context of a model like what you have estimated in `Model 3`.

So we have the following hypotheses for both tests:

$H_0 : \beta_6 = \beta_7 = \beta_8 = 0$ VS $H_a : \beta_6 \neq 0$ and/or $\beta_7 \neq 0$ and/or $\beta_8 \neq 0$

We can run the first one by comparing `Model 2` with `Model 1` (which is the same as having a model 2 under H_0) as follows:

```
mod2.quad.anova <- anova(mod.fit1, mod.fit2, test = "Chisq")
LR.chisq.m2quad <- mod2.quad.anova[["Deviance"]]
p.val.m2quad <- mod2.quad.anova[["Pr(>Chi)"]]
mod2.quad.anova
```

```
## Analysis of Deviance Table
##
## Model 1: Churn ~ tenure + MonthlyCharges + TotalCharges + SeniorCitizen +
##     gender
## Model 2: Churn ~ tenure + MonthlyCharges + TotalCharges + SeniorCitizen +
##     gender + I(tenure^2) + I(MonthlyCharges^2) + I(TotalCharges^2)
##   Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
## 1      7026     6313.6
## 2      7023     6277.8  3   35.806 8.232e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We obtain $-2log(\Delta) = 35.8055$ and a p-value of $\mathbb{P}(A > 35.8055) = 8.2318333 \times 10^{-8}$ which means that (under the context of `Model 2`) there is a strong evidence that including the Quadratic Terms is important.

On the other hand to run the second test we need to compare model 3 against a model without the Quadratic terms, as follows:

```
mod.fit3.H0.quad <- glm(
  formula = Churn ~ tenure + MonthlyCharges + TotalCharges + SeniorCitizen + gender +
    SeniorCitizen:tenure + SeniorCitizen:MonthlyCharges + SeniorCitizen:TotalCharges +
    gender:tenure + gender:MonthlyCharges + gender:TotalCharges,
  family = binomial(link = logit),
  data = telcom_churn
)

mod3.quad.anova <- anova(mod.fit3.H0.quad, mod.fit3, test = "Chisq")
LR.chisq.m3quad <- mod3.quad.anova[["Deviance"]]
p.val.m3quad <- mod3.quad.anova[["Pr(>Chi)"]]
mod3.quad.anova
```

```
## Analysis of Deviance Table
##
## Model 1: Churn ~ tenure + MonthlyCharges + TotalCharges + SeniorCitizen +
##     gender + SeniorCitizen:tenure + SeniorCitizen:MonthlyCharges +
##     SeniorCitizen:TotalCharges + gender:tenure + gender:MonthlyCharges +
##     gender:TotalCharges
## Model 2: Churn ~ tenure + MonthlyCharges + TotalCharges + SeniorCitizen +
##     gender + I(tenure^2) + I(MonthlyCharges^2) + I(TotalCharges^2) +
##     SeniorCitizen:tenure + SeniorCitizen:MonthlyCharges + SeniorCitizen:TotalCharges +
```

```
##     gender:tenure + gender:MonthlyCharges + gender:TotalCharges
##   Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
## 1      7020     6285.5
## 2      7017     6253.4  3   32.111 4.958e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

And now we obtain $-2log(\Delta) = 32.1114$ and a p-value of $\mathbb{P}(A > 32.1114) = 4.9581056 \times 10^{-7}$ which means that (under the context of `Model 3`) there is a strong evidence that including the Quadratic Terms is also important.

1.5 Test a hypothesis: Total effect of gender (15 Points)

Test the hypothesis that `gender` has no effect on the likelihood of churn, in `Model 3`, using a likelihood ratio test.

We have the following model:

$$\begin{aligned} Churn = \beta_0 &+ \beta_1 tenure + \beta_2 MonthlyCharges + \beta_3 TotalCharges + \beta_4 SeniorCitizen + \beta_5 gender \\ &+ \beta_6 tenure^2 + \beta_7 MonthlyCharges^2 + \beta_8 TotalCharges^2 \\ &+ \beta_9 SeniorCitizen \times tenure + \beta_{10} SeniorCitizen \times MonthlyCharges \\ &+ \beta_{11} SeniorCitizen \times TotalCharges + \beta_{12} gender \times tenure \\ &+ \beta_{13} gender \times MonthlyCharges + \beta_{14} gender \times TotalCharges + e \end{aligned}$$

To test this hypothesis under `Model 3` we could just use the `gender` term by itself to test its specific impact in the model, and have the following test: $H_0 : \beta_5 = 0$ VS $H_a : \beta_5 \neq 0$

```r
mod.fit3.H0.gen <- glm(
  formula = Churn ~ tenure + MonthlyCharges + TotalCharges + SeniorCitizen +
    I(tenure^2) + I(MonthlyCharges^2) + I(TotalCharges^2) + SeniorCitizen:tenure +
    SeniorCitizen:MonthlyCharges + SeniorCitizen:TotalCharges + gender:tenure +
    gender:MonthlyCharges + gender:TotalCharges,
  family = binomial(link = logit),
  data = telcom_churn
)

mod3.gen.anova <- anova(mod.fit3.H0.gen, mod.fit3, test = "Chisq")
LR.chisq.m3gen <- mod3.gen.anova[["Deviance"]]
p.val.m3gen <- mod3.gen.anova[["Pr(>Chi)"]]
mod3.gen.anova
```

```
## Analysis of Deviance Table
##
## Model 1: Churn ~ tenure + MonthlyCharges + TotalCharges + SeniorCitizen +
##     I(tenure^2) + I(MonthlyCharges^2) + I(TotalCharges^2) + SeniorCitizen:tenure +
##     SeniorCitizen:MonthlyCharges + SeniorCitizen:TotalCharges +
##     gender:tenure + gender:MonthlyCharges + gender:TotalCharges
## Model 2: Churn ~ tenure + MonthlyCharges + TotalCharges + SeniorCitizen +
##     gender + I(tenure^2) + I(MonthlyCharges^2) + I(TotalCharges^2) +
##     SeniorCitizen:tenure + SeniorCitizen:MonthlyCharges + SeniorCitizen:TotalCharges +
##     gender:tenure + gender:MonthlyCharges + gender:TotalCharges
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
```

```
## 1      7018     6254.5
## 2      7017     6253.4  1    1.1091   0.2923
```

And so we obtain $-2log(\Delta) = 1.1091$ and a p-value of $\mathbb{P}(A > 1.1091) = 0.2922712$ which means that (under the context of `Model 3`) there is NO evidence that including the gender term is important. We couldn't reject $H_0 : \beta_5 = 0$.

Although, gender is also included in many interaction terms that might be (or not be) important to the model. So if we want to test the FULL impact of `gender` in our model we would have a test: $H_0 : \beta_5 = \beta_{12} = \beta_{13} = \beta_{14} = 0$ VS $H_a : \beta_5 \neq 0$ and/or $\beta_{12} \neq 0$ and/or $\beta_{13} \neq 0$ and/or $\beta_{14} \neq 0$

```
mod.fit3.H0.nogen <- glm(
  formula = Churn ~ tenure + MonthlyCharges + TotalCharges + SeniorCitizen +
    I(tenure^2) + I(MonthlyCharges^2) + I(TotalCharges^2) + SeniorCitizen:tenure +
    SeniorCitizen:MonthlyCharges + SeniorCitizen:TotalCharges,
  family = binomial(link = logit),
  data = telcom_churn
)

mod3.nogen.anova <- anova(mod.fit3.H0.nogen, mod.fit3, test = "Chisq")
LR.chisq.m3nogen <- mod3.nogen.anova[["Deviance"]]
p.val.m3nogen <- mod3.nogen.anova[["Pr(>Chi)"]]
mod3.nogen.anova
```

```
## Analysis of Deviance Table
##
## Model 1: Churn ~ tenure + MonthlyCharges + TotalCharges + SeniorCitizen +
##     I(tenure^2) + I(MonthlyCharges^2) + I(TotalCharges^2) + SeniorCitizen:tenure +
##     SeniorCitizen:MonthlyCharges + SeniorCitizen:TotalCharges
## Model 2: Churn ~ tenure + MonthlyCharges + TotalCharges + SeniorCitizen +
##     gender + I(tenure^2) + I(MonthlyCharges^2) + I(TotalCharges^2) +
##     SeniorCitizen:tenure + SeniorCitizen:MonthlyCharges + SeniorCitizen:TotalCharges +
##     gender:tenure + gender:MonthlyCharges + gender:TotalCharges
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1      7021     6262.9
## 2      7017     6253.4  4    9.5332  0.04907 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

And so we have $-2log(\Delta) = 9.5332$ and a p-value of $\mathbb{P}(A > 9.5332) = 0.0490696$ which means that (under the context of `Model 3`) there is marginal evidence that including all of these terms might be important (depending on the chosen α),

1.6 Senior V.S. non-senior customers (20 Points)

Estimate a new model, `Model 4`, by excluding all insignificant variables from `Model 3`. Then, predict how the likelihood of churn changes for senior customers compared to non-senior customers, while keeping `tenure, MonthlyCharges,` and `TotalCharges` at their average values.

To do this we can perform a LRT as follows.

```r
Anova(mod.fit3)
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: Churn
##                               LR Chisq Df Pr(>Chisq)
## tenure                         104.850  1  < 2.2e-16 ***
## MonthlyCharges                  11.405  1  0.0007324 ***
## TotalCharges                    14.810  1  0.0001189 ***
## SeniorCitizen                   63.458  1  1.638e-15 ***
## gender                           0.026  1  0.8727620
## I(tenure^2)                     31.383  1  2.118e-08 ***
## I(MonthlyCharges^2)              1.435  1  0.2309311
## I(TotalCharges^2)               16.528  1  4.795e-05 ***
## tenure:SeniorCitizen             0.961  1  0.3270439
## MonthlyCharges:SeniorCitizen     5.645  1  0.0175095 *
## TotalCharges:SeniorCitizen       0.261  1  0.6093226
## tenure:gender                    0.829  1  0.3626033
## MonthlyCharges:gender            3.407  1  0.0649308 .
## TotalCharges:gender              3.193  1  0.0739758 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

These results are consistent with Table 1. So, taking any test and $\alpha = 0.05$ we could use the following model:

$$\begin{aligned} Model_4 : Churn = \beta_0 + \beta_1 tenure + \beta_2 MonthlyCharges + \beta_3 TotalCharges + \beta_4 SeniorCitizen \\ + \beta_5 tenure^2 + \beta_6 TotalCharges^2 + \beta_7 SeniorCitizen \times MonthlyCharges + e \end{aligned}$$

```r
mod.fit4 <- glm(
  formula = Churn ~ tenure + MonthlyCharges + TotalCharges + SeniorCitizen +
    I(tenure^2) + I(TotalCharges^2) + SeniorCitizen:MonthlyCharges,
  family = binomial(link = logit),
  data = telcom_churn
)
```

Then, we can compute the averages and build two dataframes (one for Seniors, another for Non-Seniors) as follows:

```r
avg_ten <- mean(telcom_churn$tenure)
avg_Mon <- mean(telcom_churn$MonthlyCharges)
avg_Tot <- mean(telcom_churn$TotalCharges)

predict.data.sen <- data.frame(
  tenure = avg_ten,
  MonthlyCharges = avg_Mon,
  TotalCharges = avg_Tot,
  SeniorCitizen = "Yes"
)

predict.data.nonsen <- data.frame(
  tenure = avg_ten,
```

```r
  MonthlyCharges = avg_Mon,
  TotalCharges = avg_Tot,
  SeniorCitizen = "No"
)
```

Finally, we can compute the probability of churning for both groups:

```r
predict.sen <- predict(
  object = mod.fit4,
  newdata = predict.data.sen,
  type = "link",
  se = TRUE
)
pi.hat_sen <- exp(predict.sen$fit) / (1 + exp(predict.sen$fit))
pi.hat_sen
```

```
##         1
## 0.2794051
```

```r
predict.nonsen <- predict(
  object = mod.fit4,
  newdata = predict.data.nonsen,
  type = "link",
  se = TRUE
)
pi.hat_nonsen <- exp(predict.nonsen$fit) / (1 + exp(predict.nonsen$fit))
pi.hat_nonsen
```

```
##         1
## 0.1476657
```

So we have the probability of churning:

- For Seniors: $\hat{\pi}_{Senior} = 27.94\%$.
- For Non-Seniors: $\hat{\pi}_{NonSenior} = 14.77\%$

When keeping `tenure`, `MonthlyCharges`, and `TotalCharges` at their average values.

1.7 Construct a confidence interval (20 Points)

Use `Model 4` and construct the 95% wald confidence interval for the churn probability for the customers with the following profile:

- $tenure = 55.00$;
- $MonthlyCharges = 89.86$;
- $TotalCharges = 3794.7$;
- $SeniorCitizen = "No"$;

```r
predict.data1 <- data.frame(
  tenure = 55.00,
  MonthlyCharges = 89.86,
  TotalCharges = 3794.7,
  SeniorCitizen = "No"
)
linear.pred1 <- predict(object = mod.fit4, newdata = predict.data1, type = "link", se = TRUE)
pi.hat1 <- exp(linear.pred1$fit) / (1 + exp(linear.pred1$fit))
CI.lin.pred1 <- linear.pred1$fit + qnorm(p = c(0.05/2, 1 - 0.05/2)) * linear.pred1$se
CI.pi1 <- exp(CI.lin.pred1) / (1+exp(CI.lin.pred1))
round(data.frame(pi.hat1, lower = CI.pi1[1], upper = CI.pi1[2]),4)
```

```
##   pi.hat1  lower  upper
## 1  0.1245 0.1056 0.1463
```

Meaning, that for this profile we have a $\hat{\pi} = 12.45\%$ and a CI of $0.1056 < \hat{\pi} < 0.1463$

and

- $tenure = 29.00$;
- $MonthlyCharges = 18.25$;
- $TotalCharges = 401.4$;
- $SeniorCitizen = "Yes"$

```r
predict.data2 <- data.frame(
  tenure = 29.00,
  MonthlyCharges = 18.25,
  TotalCharges = 401.4,
  SeniorCitizen = "Yes"
)
linear.pred2 <- predict(object = mod.fit4, newdata = predict.data2, type = "link", se = TRUE)
pi.hat2 <- exp(linear.pred2$fit) / (1 + exp(linear.pred2$fit))
CI.lin.pred2 <- linear.pred2$fit + qnorm(p = c(0.05/2, 1 - 0.05/2)) * linear.pred2$se
CI.pi2 <- exp(CI.lin.pred2) / (1+exp(CI.lin.pred2))
round(data.frame(pi.hat2, lower = CI.pi2[1], upper = CI.pi2[2]),4)
```

```
##   pi.hat2  lower  upper
## 1  0.0895 0.0563 0.1393
```

And for this profile we have a $\hat{\pi} = 8.95\%$ and a CI of $0.0563 < \hat{\pi} < 0.1393$