



ETC513 Assignment 3: Comparison of Energy and Pollution by Country

Shaohu Chen

Master of Business Analytics(In Progress)

Qian Duan

Master of Business Analytics(In Progress)

Tina Tsou

Master of Business Analytics(In Progress)

Report for
Australian Government

**Our consultancy
add names &
add names**

📞 (03) 9905 2478
✉️ questions@company.com

ABN: 12 377 614 630

24 May 2021

Introduction

Step 1

Step 2

Step 3

```
drivedum <- dummy_cols(drive, select_columns = c('Booking Type', 'Exam Result',
                                                    'Product Type Name'),
                         remove_selected_columns = TRUE)

train <- drivedum[1:252813,]

test <- drivedum[252814:337084,]

logmodel <- glm(`Exam Result_PASS` ~ `Booking Type_Private` + `Number of Examinations`, family=binomial(link="logit"))

summary(logmodel)

## Call:
## glm(formula = 'Exam Result_PASS' ~ 'Booking Type_Private' + 'Number of Examinations',
##      family = binomial(link = "logit"), data = train)
##
## Deviance Residuals:
##       Min        1Q     Median        3Q       Max
## -1.6008  -1.3915   0.9544   0.9764   0.9796
##
## Coefficients:
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)          0.5417475  0.0062682 86.428 < 2e-16 ***
## 'Booking Type_Private' -0.0595670  0.0082481 -7.222 5.13e-13 ***
## 'Number of Examinations' 0.0027617  0.0003107  8.889 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##  
## (Dispersion parameter for binomial family taken to be 1)  
##  
## Null deviance: 333535 on 252812 degrees of freedom  
## Residual deviance: 333398 on 252810 degrees of freedom  
## AIC: 333404  
##  
## Number of Fisher Scoring iterations: 4
```

Looking at the p value, we see that both variables have p-value close to 0 which means they are both statistically significant when it comes to the regression. From the table above, we see that *Booking Type_Private* affects the passing of an exam negatively compared to *Booking Type_Driving School*. The *Number of Examinations* have a very small but positive affect on the passing of an exam. So from the logistic regression we can conclude that by booking Privately it will negatively affect the passing of the exam in relations to booking by Driving School and that the number of examinations taken prior will have a very small but positive affect on the outcome.

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

```
## Analysis of Deviance Table  
##  
## Model: binomial, link: logit  
##  
## Response: Exam Result_PASS  
##  
## Terms added sequentially (first to last)  
##  
##  
##  
## Df Deviance Resid. Df Resid. Dev Pr(>Chi)  
## NULL 252812 333535  
## 'Booking Type_Private' 1 55.472 252811 333479 9.478e-14 ***  
## 'Number of Examinations' 1 81.045 252810 333398 < 2.2e-16 ***  
## ---  
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
fitted.results <- predict(logmodel,newdata=test,type='response')
fitted.results <- ifelse(fitted.results > 0.5,1,0)
misClasificError <- mean(fitted.results != test$`Exam Result_PASS`)
print(paste('Accuracy',1-misClasificError))

## [1] "Accuracy 0.626253396779438"
```