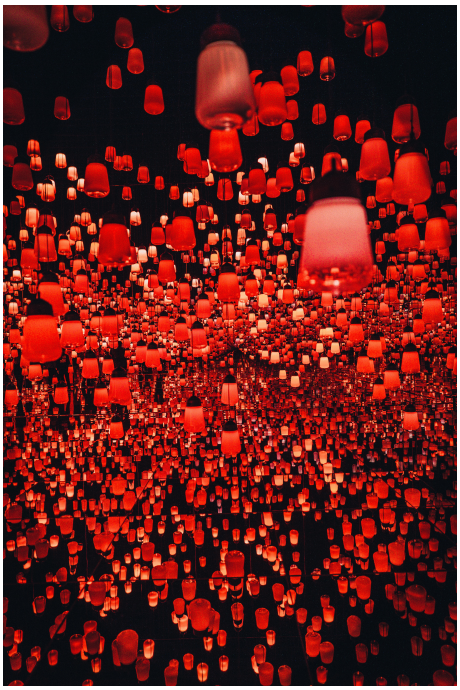

ETC513 Assignment 3: Comparison of Energy and Pollution by Country

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 COVID19

29 May 2021

Our consultancy
add names &
add names

(03) 9905 2478
questions@company.com

ABN: 12 377 614 630

```
knitr::opts_chunk$set(echo = FALSE,
                      message = FALSE,
                      warning = FALSE,
                      fig.align = 'center',
                      cache = TRUE)
```

Introduction

Step 1

Step 2

Step 3

In the following section, we will be analyzing the relationship between *Booking Type* and *Exam Result*.

The frequency plot, Figure 1, between *Booking Type* and *Exam Result* shows that the percentages of people who passed the exam are similar for both driving school and private.

Since the response variable and predictor variable are categorical variables, they will have to be converted into dummy variables(0 & 1). Then, using logistic regression to analyze their relationship, we get the following equation:

$Y\ B(p), log(\frac{p}{1-p}) = \beta_0 + \beta_1 X + \epsilon$

- β_0 is the intercept.
- β_1 is the coefficient of *Booking Type_Private*
- X is *Booking Type_Private* taking values 0 or 1

\begin{table}[!htbp] \caption{Regression with *Booking Type_Private*}

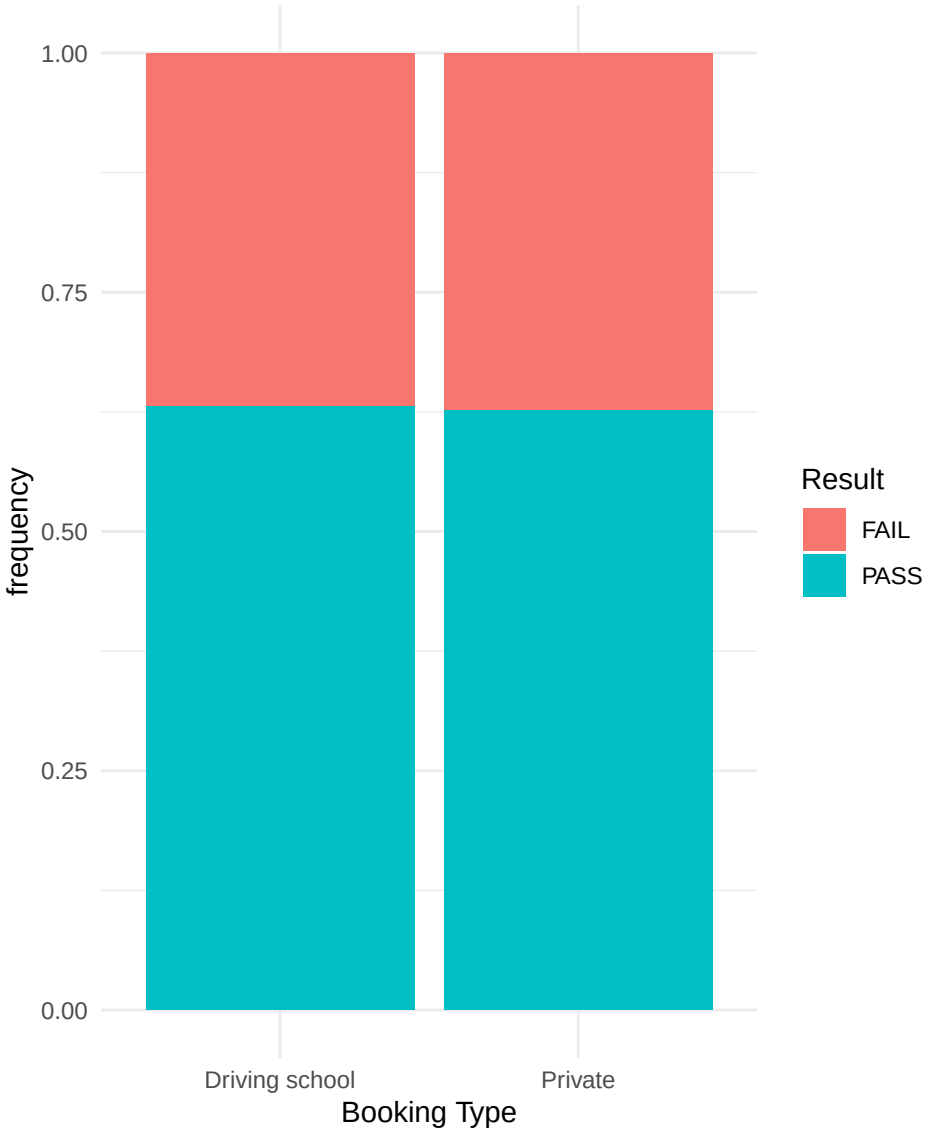

Figure 1: *Frequency Plot between Booking Type and Exam Result*

	Dependent variable:
	‘Exam Result_PASS‘
‘Booking Type_Private‘	−0.061***
	(0.008)
Constant	0.559***
	(0.006)
Observations	252,813
Log Likelihood	-166,739.600
Akaike Inf. Crit.	333,483.300

	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)
NULL	NA	NA	252812	333534.7	NA
‘Booking Type_Private‘	1	55.47238	252811	333479.3	0

\end{table} Table **??** shows the regression summary. *Booking Type_Private* has p-value close to 0 which means it is statistically significant. Due to the variable being a dummy variable relative to booking type driving school, the coefficient indicates that *Booking Type_Private* affects the passing of an exam negatively compared to *Booking Type_Driving School*. Private booking reduces the log odds by 0.061.

A quick run of ANOVA test on the *logmodel* to analyze the table of deviance shows how well the x variable is doing in comparison to the null model. Here we can see that the drop in deviance is quite small despite having low p-value. We try to improve the model by adding more variables to the function:

$$Y\ B(p), log(\frac{p}{1-p}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$$

X_2 is *Number of Examinations* taken by each examinee.

The regression with *Number of Examinations* has AIC of 333404. It is slightly lower than the AIC of the previous regression which was 333483. In comparison, having this one extra variable improved the function significantly (statistically).

Table 1: *Regression Results*

	Dependent variable:	
	‘Exam Result_PASS‘	
	(1)	(2)
‘Booking Type_Private‘	−0.061***	−0.060***
	(0.008)	(0.008)
‘Number of Examinations‘		0.003***
		(0.0003)
Constant	0.559***	0.542***
	(0.006)	(0.006)
Observations	252,813	252,813
Log Likelihood	-166,739.600	-166,699.100
Akaike Inf. Crit.	333,483.300	333,404.200
Note:	*p<0.1; **p<0.05; ***p<0.01	

x

\begin{table}[!htbp] \centering

\caption{Regression Results}

\label{}

\begin{tabular}{@{\extracolsep{5pt}}lD{.}{.}{-3} D{.}{.}{-3} }

\\[-1.8ex]\hline

\hline \\[-1.8ex]

& \multicolumn{2}{c}{\textit{Dependent variable:}} \\

\cline{2-3}

\\[-1.8ex] & \multicolumn{2}{c}{‘Exam Result_PASS‘} \\

\\[-1.8ex] & \multicolumn{1}{c}{(1)} & \multicolumn{1}{c}{(2)}\\

\hline \\[-1.8ex]

‘Booking Type_Private‘ & -0.061^{***} & -0.060^{***} \\

& (0.008) & (0.008) \\

& & \\

‘Number of Examinations‘ & & 0.003^{***} \\

& & (0.0003) \\

& & \\

Constant & 0.559^{***} & 0.542^{***} \\

& (0.006) & (0.006) \\

& & \\

\hline \\[-1.8ex]

Observations & \multicolumn{1}{c}{252,813} & \multicolumn{1}{c}{252,813} \\

Log Likelihood & \multicolumn{1}{c}{-166,739.600} & \multicolumn{1}{c}{-166,699.100} \\

Akaike Inf. Crit. & \multicolumn{1}{c}{333,483.300} & \multicolumn{1}{c}{333,404.200} \\

\hline

\hline \\[-1.8ex]

\textit{Note:} & \multicolumn{2}{r}{*p$<$0.1; **p$<$0.05; ***p$<$0.01} \\

\end{tabular}

\end{table}

// end of logistic regression model testing

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  drive$'Booking Type' and drive$'Exam Result'
## X-squared = 6.2967, df = 1, p-value = 0.0121
```

However, when we run a chi-square test, the p-value is 0.0121 so we have statistical evidence that there is a relationship between *Booking Type* and *Exam Result*.