Jerome Wong INDE546 HW4

Exercise 1. Estimate a binary logit model using a dependent variable that you specify. Include three independent variables (IV), with at least 2 being significant at p<0.10.

As part of this exercise, (a) create a bar plot of the dependent variable and (b) prepare the model results in a journal ready table that can be submitted for publication. As needed, re-label the variable names so they are meaningful to an outside reader.

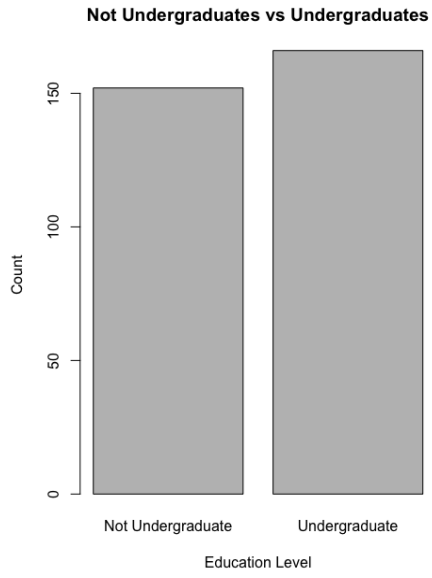
Dependent Variable:

Qn2. Undergraduate students

Independent Variable:

Q30. Age 26 and above, Qn3. Credits taken in Autumn 2021, Qn5 Class missed in Autumn 2021

(a)

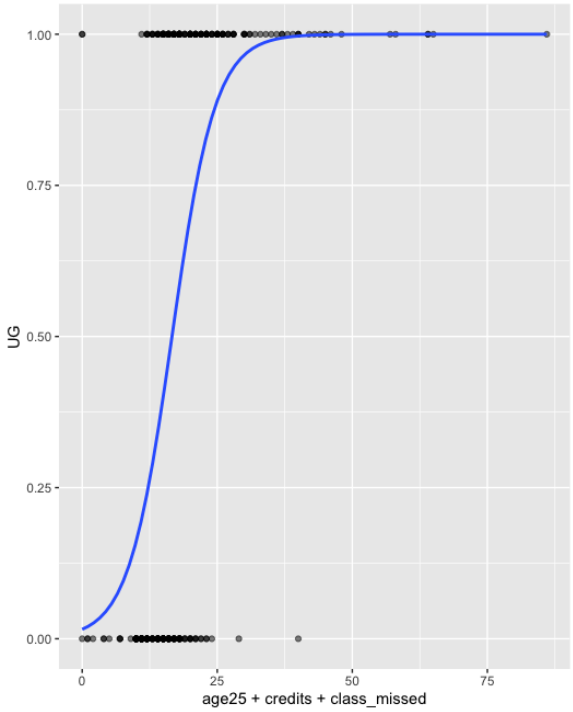


(b)

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-6.0218	0.8667	-6.95	<0.001
age25	-3.4640	0.8441	-4.10	<0.001
credits	0.4409	0.0643	6.86	<0.001
class_missed	0.1770	0.0378	4.68	<0.001

Exercise 2. Interpret your results (including model fit). Include data visualizations as appropriate to support your interpretation.

Undergraduates are shown to be inversely proportional to an older age as seen by the negative estimate. This makes sense as undergraduates tend to be younger than graduates in a university. Undergraduates are also shown to be directly proportional to credits taken. Undergraduates tend to take more credits compared to graduates as the subjects taken are of a lower workload. Finally, undergraduates are shown to be directly proportional to the class periods missed. Undergraduates' class sizes are often larger than graduates and because of that, it is easier for the undergraduates to miss a class without being noticed.



Predicted undergraduate education level by age over 25, number of credits taken and number of class periods missed

```
    Null deviance: 421.87  on 304  degrees of freedom
Residual deviance: 241.66  on 301  degrees of freedom
  (14 observations deleted due to missingness)
AIC: 249.66
```

With the inclusion of the variables age over 25, number of credits taken and number of class periods missed, devicen decreased from 421.87 to 241.66 with a loss of 3 degrees of freedom. Finding the chi square value by taking the subtracting the residual deviance from the null deviance, and taking a degree of freedom of 3, we can calculate a p-value that is <0.001 which is less than 0.05. We can conclude the model is useful for predicting that a given individual is an undergraduate or not.

Exercise 3. Create a 2nd binary logit model that includes the IVs used in Exercise 1, and at least one interaction terms (Compare the two model in terms of the AIC. Which model is better and why.

A 2nd binary logit model is created with the interaction term that combines undergraduates and those aged over 25.

```
    Null deviance: 421.87  on 304  degrees of freedom
Residual deviance: 228.42  on 300  degrees of freedom
  (14 observations deleted due to missingness)
AIC: 238.42
```

The inclusion of the interaction term led to a reduction in the AIC score from 249.66 in the 1st logit model to 238.42 in the 2nd logit model. Since the AIC score is lower in the 2nd binary logit model, the 2nd binary logit model is better.