Empirical problem set
BUS456 Fall 2022

Group 12
Ekaterina Poruchikova - s222650
Jhih-Jiuan Cheng - s222648
Shuta Yamada - s222422

2022-10-03

Question 1

Table 1 shows the frequency of the claimcolour: Green claim is 5674, Red is 1492, Yellow is 13190. Table 2 presents the the insurance type, where auto is 7562, life is 2305, other is 35, property is 5865, travel is 4589.

Table 1: Frequency of the claim colour

Green	Red	Yellow
5677	1492	13191

Table 2: Frequency of the insurance type

	auto	life	other	property	travel
2	7563	2305	35	5865	4590

```
## [1] "Insurance claim with empty insurance type: 17857"
## [2] "Insurance claim with empty insurance type: 18515"
```

Question 2

Table 3 presents the fraction of claims by colour, while Table 4 provides the breakdown by claim colour and insurance type. In the green claim color other has the highest share, while the second largest category is property insurance. In the yellow category most prominent are life and travel insurance policies. At the same time red claim color is most common in the travel category. However, the travel has the lowest green claim color, other and property have two of the lowest values in yellow category, while auto has the lowest value in red. Therefore it can be concluded that travel is the least likely category to be suitable for automatic evaluation.

Table 3: Percentage of claim color

green	yellow	red
0.2787602	0.6479517	0.0732881

Table 4: Percentage of claim color by insurance type

insurancetype	green	yellow	red
auto	0.3411345	0.6402221	0.0186434
life	0.2125813	0.7422993	0.0451193
other	0.8857143	0.0857143	0.0285714
property	0.3636829	0.5459506	0.0903666
travel	0.0960784	0.7479303	0.1559913

Question 3

Put in here

Table 5: Balance table

treatmentgroup	0		1		2		3		4		
Variable	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean	Test
claiminternalnumber	4101	10172.012	4133	10158.391	4142	10254.186	3962	10248.514	4020	10064.951	F= 0.7
claimcolour	4101		4133		4142		3962		4020		X2= 7.923
... Green	1155	28.2%	1138	27.5%	1117	27%	1159	29.3%	1106	27.5%	
... Red	300	7.3%	307	7.4%	304	7.3%	300	7.6%	281	7%	
... Yellow	2646	64.5%	2688	65%	2721	65.7%	2503	63.2%	2633	65.5%	
insurancetype	4101		4133		4142		3962		4020		X2= 4.86
...	0	0%	0	0%	0	0%	0	0%	0	0%	
... auto	1531	37.3%	1523	36.8%	1524	36.8%	1467	37%	1518	37.8%	
... life	463	11.3%	496	12%	457	11%	447	11.3%	442	11%	
... other	7	0.2%	7	0.2%	6	0.1%	8	0.2%	7	0.2%	
... property	1167	28.5%	1192	28.8%	1218	29.4%	1147	29%	1141	28.4%	
... travel	933	22.8%	915	22.1%	937	22.6%	893	22.5%	912	22.7%	
privatefinanced	4101		4133		4142		3962		4020		X2= 14.422***
... 0	676	16.5%	661	16%	662	16%	713	18%	601	15%	
... 1	3425	83.5%	3472	84%	3480	84%	3249	82%	3419	85%	

Statistical significance markers: * p<0.1; ** p<0.05; *** p<0.01

Question 4

Table 6 presents the regression output. In the first model the constant term is positive and statistically significant (at 1% level). From this result, we can estimate that without treatments, customers have a general tendency to accept automated claim procedures. If we think about the fact that over 80% of customers have either green or yellow claim color, this result makes intuitive sense. All treatment variables have positive coefficients. However, among four treatment variables, only social norm treatment and combined treatment are statistically significant. The P-value of social norm treatment is below 5% significance level, and that of combined treatment is below 1% level. In terms of the result of social norm treatment, it aligns with the theory of reciprocity (conditional cooperation). Although customers cannot observe how other customers behave in this experiment, the information about others' pro-social behaviors (accepting automated procedures) might have affected their choices. As to combined treatment, it has a larger coefficient than the social norm treatment. In addition to the effects of social norm treatment, simplification might reduce the cognitive efforts of customers and contribute to increase the rate of acceptance.

Question 5

In the second regression in the Table 6, the coefficient of the variable "privatefinanced" is negative and statistically significant (1% level). Thus, we can assume that there is a negative correlation between paying expenses on insurance from a personal budget and accepting automated procedures. This can be backed up by Prospect Theory. Since the value function is steeper in the loss domain than in the gain domain, it is difficult for customers paying insurance fees by themselves (loss), to be driven by the utility from being "good" (gain from pro-social behavior, accepting automation). The change in the social norm group coefficient and its significance can most likely be attributed to a random chance as the sign of the coefficient remains the same and the value of the coefficient decreased only slightly and p-value changed from 3% to 5%. Therefore, the interpretation would not change drastically.

Question 6

In the Table 7 separate regressions by claim color are presented. Combined treatment had the largest effect on the group with the "Yellow" claim color. Also, the coefficient is only statistically significant for this group (1% level). However, this might not be the very result the company wanted, because claims labeled "Yellow"

Table 6: Regression output

	Dependent variable:	
	accept_automatic	
	(1)	(2)
simplification	0.007	0.008
	(0.007)	(0.007)
personalization	0.004	0.006
	(0.007)	(0.007)
social.norm	0.015**	0.014*
	(0.007)	(0.007)
combined	0.026***	0.028***
	(0.007)	(0.007)
insurancetypelife		0.078***
		(0.008)
insurancetypeother		−0.023
		(0.055)
insurancetypeproperty		−0.058***
		(0.006)
insurancetypetravel		0.011*
		(0.007)
red		0.004
		(0.009)
green		0.052***
		(0.005)
privatefinanced1		−0.059***
		(0.007)
Constant	0.867***	0.907***
	(0.005)	(0.009)
Observations	20,358	20,358
R^2	0.001	0.020
Adjusted R^2	0.001	0.020
Residual Std. Error	0.328 (df = 20353)	0.324 (df = 20346)
F Statistic	4.090*** (df = 4; 20353)	38.036*** (df = 11; 20346)
Note:		*p<0.1; **p<0.05; ***p<0.01

require some elements of personal evaluation. Although they can expect an increase in operational/financial efficiency from automation, the effect would not be as big as they expected. They will still need humans to process these claims and the time spent on procedures might not change so significantly. This may damage customer experiences because it kills one of the good features of automated claim procedures for customers. Since automation leads to a reduction of time for procedures, it is natural for customers who accept it to expect getting answers for their claims in a shorter time. This would shift their reference point to evaluate time spent on the procedures: for example, from setting 2-3 days as the origin (reference point) to setting 1 day or several hours as the origin. However, it would take longer than 1 day or several hours if their claims are not very suitable for automation, of which the claim color is "Yellow." This will not be a problem for customers with the former reference point, but for those who with the latter reference point, it is interpreted as a loss. This will degrade customer experiences and ultimately, may lead to the unwanted result, the loss of customers.

Table 7: Regressions by claim color

	Dependent variable:		
	accept_automatic		
	Green	Yellow	Red
	(1)	(2)	(3)
combined	0.013	0.036***	−0.005
	(0.013)	(0.009)	(0.028)
Constant	0.894***	0.854***	0.873***
	(0.009)	(0.006)	(0.019)
Observations	2,261	5,279	581
R^2	0.001	0.003	0.0001
Adjusted R^2	0.0001	0.003	−0.002
Residual Std. Error	0.299 (df = 2259)	0.333 (df = 5277)	0.336 (df = 579)
F Statistic	1.137 (df = 1; 2259)	15.175*** (df = 1; 5277)	0.032 (df = 1; 579)
Note:			*p<0.1; **p<0.05; ***p<0.01

Question 7

As an additional analysis we ran a logistic regression to predict the automatic acceptance. The results of the regression are presented in the Table 8. The same variables as in the linear regression in Table 6 were found significant. However, due to larger presentce of automatic asseptances in the sample, as demonstrated by Table 9, the predictions made by the model were exclusively automatic acceptances. Therefore the model accuracy suffered, as demonstrated by the ROC figure.

```
## Setting levels: control = 0, case = 1
```

```
## Setting direction: controls < cases
```

Table 8: Logistic regression output

	Dependent variable:
	accept_automatic
simplification	0.068
	(0.066)
personalization	0.055
	(0.066)
social.norm	0.132*
	(0.068)
combined	0.271***
	(0.069)
insurancetypelife	0.993***
	(0.100)
insurancetypeother	−0.235
	(0.534)
insurancetypeproperty	−0.478***
	(0.052)
insurancetypetravel	0.127**
	(0.065)
red	0.044
	(0.082)
green	0.496***
	(0.053)
privatefinanced1	−0.568***
	(0.069)
Constant	2.278***
	(0.087)
Observations	20,358
Log Likelihood	−7,349.621
Akaike Inf. Crit.	14,723.240
Note:	*p<0.1; **p<0.05; ***p<0.01

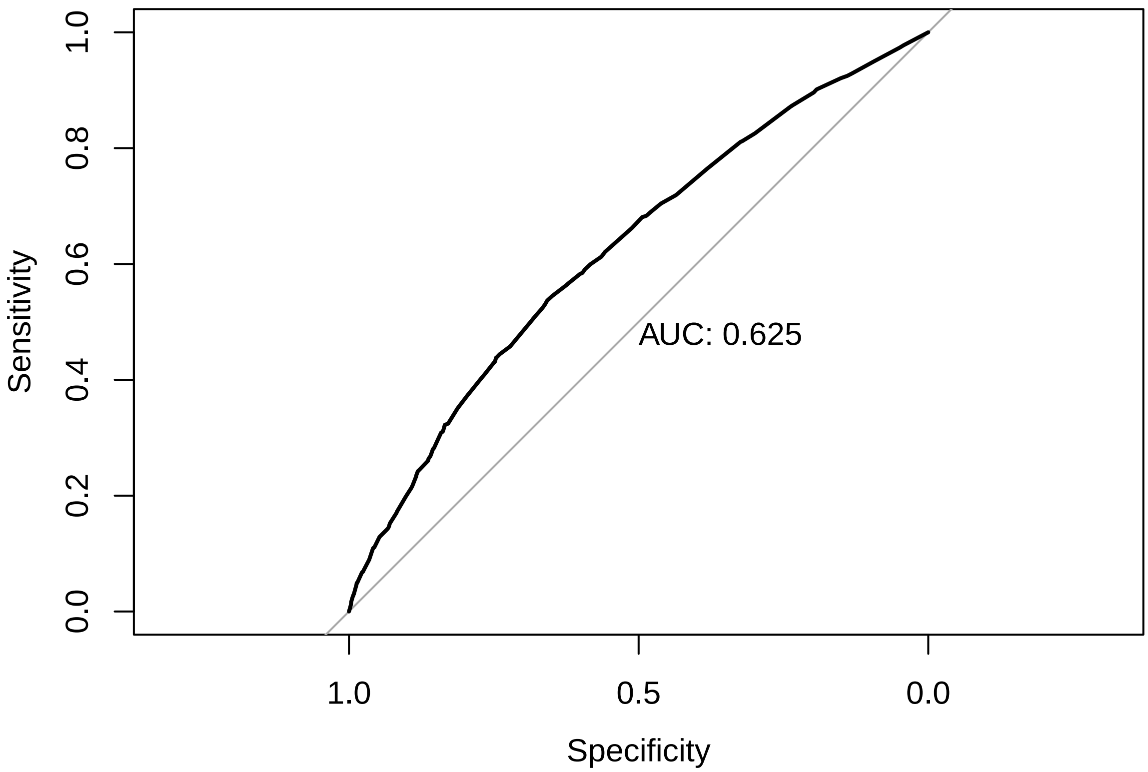

Table 9: Summary statistics of the logit predictions

Var1	Freq
0	2490
1	17868

[1] "Lowest probability prediction is 0.77"