

Empirical problem set
BUS456 Fall 2022

Group 12
Ekaterina Poruchikova - s222650
Add
Add
Add

2022-09-23

Question 1

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"
```

Comment on question 1

Question 2

Table 3: Percentage of claim color

green	yellow	red
0.278831	0.647888	0.0732809

Table 4: Percentage of claim color by insurance type

insurancetype	green	yellow	red
	1.0000000	0.0000000	0.0000000
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

```
auto %>%  
  group_by(treatmentgroup) %>%  
  summarise(  
    green = mean(green),  
    yellow = mean(yellow),  
    red = mean(red),
```

```

auto = mean(insurancetype == 'auto'),
life = mean(insurancetype == 'life'),
property = mean(insurancetype == 'property'),
travel = mean(insurancetype == 'travel'),
other = mean(insurancetype == 'other'),
privatefinanced = mean(as.numeric(as.character(privatefinanced)))
) %>%
t(.) %>%
knitr::kable(caption = 'Title')

```

Table 5: Title

treatmentgroup	0	1	2	3	4
green	0.2816386	0.2753448	0.2696765	0.2927075	0.2753047
yellow	0.6452085	0.6503750	0.6569290	0.6315922	0.6548122
red	0.07315289	0.07428018	0.07339450	0.07570023	0.06988311
auto	0.3733236	0.3684975	0.3679382	0.3701741	0.3775180
life	0.1128993	0.1200097	0.1103332	0.1127933	0.1099229
property	0.2845647	0.2884104	0.2940608	0.2894272	0.2837603
travel	0.2275055	0.2213888	0.2262192	0.2253343	0.2268093
other	0.001706901	0.001693685	0.001448576	0.002018673	0.001740860
privatefinanced	0.8351622	0.8400677	0.8401738	0.8200858	0.8505347

Create a so-called “balance table” that shows variable means of the variables (green, red, yellow, auto, life, property, travel, other, and privatefinanced) per treatment group and p-values from a test of joint significance (testing whether the mean of a variable is statistically different across treatment groups). The variable privatefinanced is a dummy variable equal to 1 if the individual paid for the insurance out of own pocket (0 if paid via employer/union/creditcard). What is the main takeaway from the table?