3. Natural experiment design

3.1 Data collection

The goal of this work is to discover the impact of Green IT usage. Thus, we planned to choose CO2 emissions as dependent variable to show the drivers’ driving behavior changes and the environmental impact of the usage of our app, a product of Green IT.

*（下面这我用了过去时和现在完成时）*

The sample of vehicles monitored in the experiment has been selected from 63 different taxi drivers in xxx(公司). We obtained their driving data from July 2019 to October 2020 using OBD systems, which have been incorporated into the computers on-board new vehicles to monitor vehicle components and driving behaviors in recent years. Meanwhile, we invited these 63 taxi drivers to use our software application (护驾宝就是“Hujiabao”吗？), which can send alert to them when it detects risky driving behaviors and can provide a driving behavior ranking at the end of the day. We used the drivers' check-ins in the app to identify whether they had used the software that day.

*(这里没有介绍数据特征，好像是要放在results里？——去除所有碳排放量为0，即缺失的记录项之后，得到11189 observations，21个变量。其中，为了保证正态性，将xxxx取logged data,更名为xxxxx。这些变量的缺失值用每个driver的每日数据均值替换。这里可以有个变量的表？但是21个变量确实似乎要缩减一点。).*

3.2 Analytic strategy *(这个模型我们好像没完全确定下来，所以这里有些细节我没有补充全)*

We planned to construct a linear regression model to investigate how the app usage and driving behaviors affect the CO2 emissions, as given next:

(+公式)

(这里解释公式)As described earlier, CO2 emissions *(或者这里写log\_co2\_1)*, being the dependent variable, is get directly through the ODB systems of a car per day. (+Xxxxxxx) is the independent variable. xxxx and xxxxxx are the moderate variables. We use the driver's personal features (gender, age, driving experience) as control variables *(这个应该是控制了好像，没有就后面控制了啥就写啥)*, and xxx is the idiosyncratic error term. Thus, we use Equation (1) to fit our data. The effect of aggressive driving behaviors and the app usage can be found by the sign and significance level of xxxx . The moderating effects are shown in the coefficients xxxxx. Further, we adopt xxxxx to verify the stability of the results.

Significance of the equation, each independent variable（*还是说contributing factor）* and coefficient was assessed using z tests and measured the effects of the presence of each independent variable on the strength of predictive relationships in the model. An α level of 0.05*（我们的模型好像是这样,0.05）* was adopted for our z test on each contributing factor. Analyses were conducted in 2021.