

Course: CVE 593 – Special Topics in Civil Engr

Task: Homework #3

Due Date: Nov 20, 2022 at 11:59 PM

In this homework, you are given data retrieved from the Transportation Secure Data Center website. The data is a 2017 Survey from residential and commercial light-duty fleet owners in California. The objective of the survey was to gather preferences on the use of autonomous and electric vehicles. A sample of these data is presented in the table below.

	self_driving	safety_concern	carshare_level	rideshare_level	pevowner	income	householdvehicles	housing	parkingtype
0	No	Disagree	Not participating	Interested in participating	No	Low income	Three plus	Single family	Personal parking
1	Yes	Disagree	Interested in participating	Interested in participating	No	Low income	Three plus	Single family	Personal parking
2	No	Neutral	Currently participating	Currently participating	No	Low income	Three plus	Single family	Personal parking
3	No	Agree	Not participating	Not participating	No	Low income	Three plus	Apartment	Shared parking
4	Neutral	Neutral	Currently participating	Currently participating	No	Low income	Three plus	Single family	Personal parking

Variable definitions

Variable	Description
self_driving	Consider purchasing a fully self-driving vehicle
safety_concern	Concerned about the safety of self-driving vehicles
Carshare_level	Participation in car-share programs, e.g., Car2Go, Enterprise, Hertz, ZipCar, etc.
Rideshare_level	Participation in ride-sharing and ride-share programs, e.g., Uber, Lyft, etc.
pevowner	Plug-in Electric Vehicle (PEV) ownership
income	The income of the household
Householdvehicles	The number of vehicles in the household
housing	The housing type
Parkingtype	The parking type

Problem 1 (20 Points)

Task: Use Python particularly a *statsmodels* package to calibrate a binary logistic regression that estimates the preference of purchasing a self-driving or autonomous vehicle. To obtain a binary outcome in the data set, create a new column that combines “no” and “neutral” responses in one group. The new column will have “1” and “0” for “yes” and “no” & “neutral”, respectively.

e.g.,

```
1 from statsmodels.discrete.discrete_model import MNLogit
2 from statsmodels.formula.api import probit, logit
3
4 model = .....
```

- a) Fit a binary logit regression between self-driving preference and all other attributes (5 Points)
- b) Identify all the attributes that are significant at a 95% confidence interval (5 Points)
- c) Provide a discussion interpreting the model results of all the variables that are significant at a 95% confidence interval (10 Points)

Problem 2 (30 Points)

Use the same data to calibrate a multinomial regression model that estimates the preference of residents in purchasing self-driving vehicles, that is “no”, “neutral”, and “yes”.

- a) Fit a multinomial regression model with all the attributes, and identify which attributes are significant at a 95% confidence interval (10 Points)
- b) Provide a discussion interpreting the model results of all the variables that are significant at a 95% confidence interval (20 Points)

[Hints: You will have to change the values in data to numerical to fit the regressions. For the discussion of results, use the two literatures provided on BB as example to interpret the results.](#)

Submit your solution together with your code to receive full points.