Case Study #1 - The Travel Cost Model

PUT YOUR NAME HERE

9/19/2022

Case Study #1 - The Travel Cost Model (100 points)

This repo provides instructions, code, and data to estimate a travel cost model from Seller, Stoll, and Chavas (1985).

Install R and Rstudio (25 points)

As mentioned above, this case study requires the use of the *R* programming language. *R* is free and available for download here. While you can use *R* on it's own, downloading *RStudio* will provide you with a welcoming environment (an integrated development environment, or IDE) that is useful for replication. It is free and available for download here.

Optional: Install GitHub (0 points)

Github is free and available for download here. *Github* is used to house this repository and by installing and using it to clone the repository one will simplify the replication procedure. However, you could also simply download a zipped file version of this repository, unzip in the desired location on your machine, and follow the replication procedures outlined below.

Cloning the Repository (25 points)

Begin by either cloning this repository or downloading the zipped file. Cloning can be done from the main page of this repo by clicking on the green "code" button in the top right corner of the page and following those instructions (this will also provide you with an option to download a .zip file of the repository), or by navigating in the terminal to the desired location of the clone and then typing:

```
git clone https://github.com/bryanparthum/environmental_economics
```

If you chose to download a .zip file of the repository, simply unzip it to wherever you would like to have it and proceed with the following steps in that directory.

The Travel Cost Model

To get started with the replication, navigate in your file explorer (or equivalent) to `case_studies/case_study_1` and double click on (open) the markdown file `travel_cost.Rmd`. This will prompt your machine to open the file in *RStudio*. This markdown document includes all the instructions and code to estimate a travel cost model from Seller, Stoll, and Chavas (1985).

You can run individual code sections or lines by navigating in the section of code marked with `{r}` and hit the `Run` button at the top of the code editor (other options in its drop down menu), or use a keyboard shortcut and type `Ctrl + Enter` (on a Mac, type `Cmd + Return`). The Keyboard shortcuts are really handy for running the code line-by-line. Whatever line the cursor is on, regardless of where the cursor is on that line, can be run by simply typing `Ctrl + Enter` (on a Mac, type `Cmd + Return`).

Running individual lines is useful when testing the commands, but the entire markdown document (compete with code chunks) is ran by hitting the `Knit` button at the top of the page (this is called "knitting"). When you `knit` a document, it will run all the text and all the code and populate a output file (in this case, a .pdf). **This output file is what you will submit for your case study.**

Next, load the data set:

```
## load data
data(RecreationDemand)
```

Explore what this data set is by typing:

```
## look at the help file
? RecreationDemand
```

```
## starting httpd help server ... done
```

The question mark `?` is really useful in *R* whenever we want to access the help file for something. For example, typing `? library` will automatically bring you to the help file for loading packages. When help command (the `?`) are included in an Rmarkdown document it opens your browser to the relevant help file.

Continue by answering the following questions and adding your responses to this markdown document.

1. (5 points) What are the data that you just loaded? Copy the `Description` from the help file.

- Description:

Export the table of summary statistics from the data using the `stargazer` package.

```
## export summary statistics
stargazer(RecreationDemand,
          title  = 'Summary Statistics',
          digits = 1,
          dep.var.labels.include = FALSE,
          type   = 'latex',
          header = FALSE)
```

Table 1: Summary Statistics

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
trips	659	2.2	6.3	0	0	2	88
quality	659	1.4	1.8	0	0	3	5
income	659	3.9	1.9	1	3	5	9
costC	659	55.4	46.7	4.3	28.2	69.7	493.8
costS	659	59.9	46.4	4.8	33.3	72.6	491.5
costH	659	56.0	46.1	5.7	29.0	68.6	491.0

2. (5 points) What was the total number of respondents (observations)?
    - Answer:
3. (5 points) What was the average number of trips taken?
    - Answer:
4. (5 points) What was the average cost of a trip to Sommerville Lake? (this is the `costS` variable) And the other two substitute lakes? (Conroe Lake = `costC` and Houston Lake = `costH`)
    - Answer:

Next, find out what the generalized linear model is.

```
## what is a generalized linear model?
? glm
```

5. (5 points) What is a glm model? Copy and paste the description of the model from the help file.
    - Description:

You will estimate two other models as well. Find out what the negative binomial model is and what the zero-inflated model is.

```
## what is a negative binomial model?
? glm.nb

## what is a zero-inflated model?
? zeroinfl
```

6. (5 points) What is a glm.nb model? Copy and paste the description of the model from the help file.

    - Desription:

7. (5 points) What is a zeroinfl model? Copy and paste the description of the model from the help file.

    - Desription:

Next, run the generalized linear model (`glm`), the negative binomial model (`glm.nb`), and the zero-inflated model (`zeroinfl`). Look at the regression output summaries.

```
## run the model generalized linear model
glm <- glm(trips ~ costS + costC + costH + quality + ski + income + userfee,
           data   = RecreationDemand,
           family = poisson)

## look at the summary output
summary(glm)
```

```
##
## Call:
## glm(formula = trips ~ costS + costC + costH + quality + ski +
##     income + userfee, family = poisson, data = RecreationDemand)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -11.8465   -1.1411   -0.8896   -0.4780   18.6071
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.264993   0.093722   2.827  0.00469 **
## costS       -0.042536   0.001670 -25.467  < 2e-16 ***
## costC       -0.003430   0.003118  -1.100  0.27131
## costH        0.036134   0.002710  13.335  < 2e-16 ***
## quality      0.471726   0.017091  27.602  < 2e-16 ***
## skiyes       0.418214   0.057190   7.313 2.62e-13 ***
## income      -0.111323   0.019588  -5.683 1.32e-08 ***
## userfeeyes   0.898165   0.078985  11.371  < 2e-16 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 4849.7  on 658  degrees of freedom
## Residual deviance: 2305.8  on 651  degrees of freedom
## AIC: 3074.9
##
## Number of Fisher Scoring iterations: 7
```

```
## run the negative binomial model
nb <- glm.nb(trips ~ costS + costC + costH + quality + ski + income + userfee,
             data = RecreationDemand)

## look at the summary output
summary(nb)
```

```
##
## Call:
## glm.nb(formula = trips ~ costS + costC + costH + quality + ski +
##     income + userfee, data = RecreationDemand, init.theta = 0.7292568331,
##     link = log)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.9727  -0.6256  -0.4619  -0.2897   5.0494
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.121936   0.214303  -5.235 1.65e-07 ***
## costS       -0.092691   0.006653 -13.931  < 2e-16 ***
## costC        0.048009   0.009185   5.227 1.72e-07 ***
## costH        0.038836   0.007751   5.011 5.42e-07 ***
## quality      0.721999   0.040117  17.998  < 2e-16 ***
## skiyes       0.612139   0.150303   4.073 4.65e-05 ***
## income      -0.026059   0.042453  -0.614    0.539
## userfeeyes   0.669168   0.353021   1.896    0.058 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(0.7293) family taken to be 1)
##
##     Null deviance: 1244.61  on 658  degrees of freedom
## Residual deviance:  425.42  on 651  degrees of freedom
## AIC: 1669.1
##
## Number of Fisher Scoring iterations: 1
```

```
##
##
##               Theta:  0.7293
##           Std. Err.:  0.0747
##
##  2 x log-likelihood:  -1651.1150
```

```
## run the zero-inflated poisson
zi <- zeroinfl(trips ~  costS + costC + costH + quality + ski + income + userfee | quality + in
               data = RecreationDemand)

## look at the summary output
summary(zi)
```

```
##
## Call:
## zeroinfl(formula = trips ~ costS + costC + costH + quality + ski + income +
##     userfee | quality + income, data = RecreationDemand)
##
## Pearson residuals:
##     Min      1Q  Median      3Q     Max
## -6.3255 -0.2714 -0.1809 -0.1646 13.3126
##
## Count model coefficients (poisson with log link):
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  2.099162   0.111397  18.844  < 2e-16 ***
## costS       -0.037600   0.002038 -18.454  < 2e-16 ***
## costC        0.002369   0.003818   0.620    0.535
## costH        0.025234   0.003355   7.522 5.40e-14 ***
## quality      0.033833   0.023914   1.415    0.157
## skiyes       0.471691   0.058187   8.106 5.21e-16 ***
## income      -0.099780   0.020779  -4.802 1.57e-06 ***
## userfeeyes   0.610488   0.079435   7.685 1.53e-14 ***
##
## Zero-inflation model coefficients (binomial with logit link):
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  3.29191    0.51608   6.379 1.79e-10 ***
## quality     -1.91407    0.20619  -9.283  < 2e-16 ***
## income      -0.04502    0.10797  -0.417    0.677
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Number of iterations in BFGS optimization: 18
## Log-likelihood: -1181 on 11 Df
```

Export the regression outputs to a table using the `stargazer` package.

```
## export summary statistics
stargazer(glm,
          nb,
          zi,
          title  = 'Travel Cost Regression Results',
          digits = 2,
          type   = 'latex',
          header = FALSE,
          label  = "tab:results")
```

Using Table 2, answer the questions below.

9. (5 points) What are the coefficients interpreted as? Recall, they are the partial derivatives of the utility function.

- Answer:

10. (5 points) Are the coefficients on the cost variables as expected? Recall, `costS` is the own price (the price of visiting Sommerville Lake), and the other two cost variable `costC` and `costH` are the price of visiting the "substitute" lakes.

- Answer:

11. (5 points) What is the marginal willingness to pay for lake quality at Sommerville Lake? Recall, this is the ratio of two coefficients (preference parameters).

- Answer:

Table 2: Travel Cost Regression Results

	Dependent variable:		
	trips		
	Poisson	*negative binomial*	*zero-inflated count data*
	(1)	(2)	(3)
costS	-0.04^{***}	-0.09^{***}	-0.04^{***}
	(0.002)	(0.01)	(0.002)
costC	-0.003	0.05^{***}	0.002
	(0.003)	(0.01)	(0.004)
costH	0.04^{***}	0.04^{***}	0.03^{***}
	(0.003)	(0.01)	(0.003)
quality	0.47^{***}	0.72^{***}	0.03
	(0.02)	(0.04)	(0.02)
skiyes	0.42^{***}	0.61^{***}	0.47^{***}
	(0.06)	(0.15)	(0.06)
income	-0.11^{***}	-0.03	-0.10^{***}
	(0.02)	(0.04)	(0.02)
userfeeyes	0.90^{***}	0.67^{*}	0.61^{***}
	(0.08)	(0.35)	(0.08)
Constant	0.26^{***}	-1.12^{***}	2.10^{***}
	(0.09)	(0.21)	(0.11)
Observations	659	659	659
Log Likelihood	$-1{,}529.43$	-826.56	$-1{,}180.80$
θ		0.73^{***} (0.07)	
Akaike Inf. Crit.	3,074.86	1,669.12	
Note:		*p<0.1; **p<0.05; ***p<0.01	