## *Winter 2021, Empirical Project*

### *Requirements:*

1. *Due on March 8ᵗʰ at the end of the day on Canvas. Both your TA and I will grade your report. The average will be the score for your empirical project. There is no regrading of the empirical project.*
2. *Use Stata to analyze the provided dataset.*
3. *Summarize your results in a report. Describe: your question(s), your data (e.g., show features of the variables of the full sample or across different subsamples using summary statistics and figures), your regression models, your estimates with standard errors (in tables), and the answers to your questions. Remember to interpret your regression results carefully. Pay attention to both statistical significance and economic significance.*
4. *The report shall be no more than 10 pages long (double spaced time new roman font size 12). Be sure not to include unnecessary Stata outputs in your report. Use equations and make your own tables (not Stata output tables) to summarize your regression results.*
5. *Submit your Stata log file together with the report. Make sure that your results could be replicated by the commands in you log file.*
6. *Keep in mind that this is a report for an econometrics class. Asking the right question is important. But to demonstrate that you know what econometrics models to use and how to properly interpret the regression results is important as well.*
7. *Run your report through some online grammar editing website before submission. Although we won't be picky in language and grammar, well-written papers generally make good impressions.*

### *Background of the Stata data file*

In November 1984, the Kentucky Public Service Commission (KPSC) established Administrative Case No. 285 to study the economic feasibility of providing local–measured–service telephone rates in Bowling Green and

Louisville, Kentucky. We are only looking at the data in Louisville in this project.

The experiment took place in the second half of 1986. Before the start date, Louisville consumers were instructed to choose between **unlimited calls** at a cost of $18.70 per month, or a **measured service** option with a monthly fee of $14.02. The measured service tariff included a $5 allowance and distinguished setup, duration, peak periods, and distance. The tariff differentiated among three periods: peak was from 8 a.m. to 5 p.m. on weekdays; shoulder was between 5 p.m. to 11 p.m. on weekdays and Sundays; and off–peak was any other time. For distance band A, measured charges in peak periods were 2 cents for setup and 2 cents per minute. These charges had a 35% discount in shoulder time and 60% discount in off–peak time. For distance band B, setup charges remained equal but duration always had a 100% surcharge relative to the corresponding tariff for band A. The following table summarizes the setup and duration charges for those who choose the measured service option.

|  | Setup Charge per call | Additional charge Per Minute |
|---|---|---|
| Band A, Peak | 2 cents | 2 cents |
| Band A, Shoulder | 2 cents * (1-35%) | 2 cents * (1-35%) |
| Band A, Off-peak | 2 cents * (1-60%) | 2 cents * (1-60%) |
| Band B, Peak | 2 cents | 2 cents * 2 |
| Band B, Shoulder | 2 cents * (1-35%) | 2 cents * (1-35%)* 2 |
| Band B, Off-peak | 2 cents * (1-60%) | 2 cents * (1-60%)* 2 |

If set-up and minute charges do not add up to $5 in a given month, the household pays $14.02 for that month. If set-up and minute charges exceed $5 in a given month, the household pays $14.02+(the total of set-up and minute charges-$5).

The dataset contains randomly sampled households in Louisville. Household demographic information is surveyed before the experiment took place in July 1986. Household detailed telephone usage is based on administrative files recorded by KPSC. In this dataset, we only use household telephone usage information from Oct. to Dec. i.e., 3 months after the beginning of the experiment to make sure that households had all been aware of their new tariff schedules.

**Demographic variables include:**

```
ID      Household ID

MONTH   Month of telephone usage in document

AGE     =1 if 15-14
        =2 if 25-34
        =3 if 35-44
        =4 if 45-54
        =5 if 55-64
        =6 if >65

HHSIZE     # in household

ADULTS     # of adults in household

TEENS      # of kids in household

SENIORS   # of seniors in household

EDUC    education of household head
        =1 if high school drop out
        =2 if high school graduate
        =3 if some college
        =4 if college graduate
        =5 if some graduate school
        =6 if graduate school

MARRIED  marital status of household head
         =1 if married
         =2 if widowed, divorced, separated,
         =3 if never married

RACE        race of household head
     =1 if white
     =2 if black
     =3 if spanish
     =4 if asian
     =5 if other

INCOME     annual household income
     =1 if < $5000
     =2 if $5000-7499
     =3 if $7500-9999
     =4 if $10000-14999
     =5 if $15000-19999
     =6 if $20000-24999
```

```
=7 if $25000-34999
=8 if $35000-49999
=9 if > $50000
```

*BENEFITS =1 if receiving benefits such as social security, food stamps or aid to dependent children, =0 if otherwise.*

*LV =1 if the household resides in Louisville. In this dataset, it's always 1.*

**Telephone usage variables include the tariff variable and a bunch of usage pattern variables.**

```
Tariff  =1 if measured
        =0 if flat
```

Usage pattern variables are named according to the following system of rules:

First character:     M=minutes                C=number of calls
Second character: A=band A (close)        B=band B (far)
Third character:   P= (peak) Weekdays, from 8am to 5pm
                   S= (shoulder) Weekdays and Sundays, from 5pm to 11pm
                   O= (off-peak) Weekdays from 11pm to 8am and from
          Fridays 11pm to Sundays 5pm

*Some Notes (with couple questions for you to start with):*

This data file includes a lot of information. You don't need to use all of it. You could start from pooling all peak calls and minutes, all shoulder calls and minutes, as well as all off-peaks calls and minutes together and ask questions like "Did low income people make more of their calls during off-peak hours?", "Did younger households call more or do elder households call more?", "Did household call more during the holiday season?", "Did households who had chosen the flat rate schedule and those who had chosen the tariff schedule have similar demographic characteristics?", "Did households under measured service rate make less peak hour calls than those under flat rate? Did they make more or less distance band B calls during peak hours? Did they make shorter distance band B calls because of the duration surcharge? Did these households utilize almost all of the $5 allowance?"

If you are willing to program a bit in Stata, you could also recover household's monthly bill from the information in the dataset. Then you ask even more interesting questions related to households' choice between the two telephone plans.


*Final Notes:*

Most real world survey datasets are kind of messy. I have cleaned the dataset before I gave it to you. But if you spot anything that is suspicious, just discuss it in a footnote of your report.