

BM06BIM Research Methods I – Assignment 1

Linear Regressions with Cross-Sectional and Panel Data

Tuesday, September 20, 2022

INSTRUCTIONS

- Read this entire assignment before beginning. Follow the directions and answer all questions. Provide graphs/figures where appropriate and refer to them in the text.
- The points you achieve in each exercise will reflect the informativeness and conciseness of your results and quality of presentation. All group members will receive the same grade.
- If you want to cite references in your assignment, which you do not have to do, please use the [APA Style](#).
- Submit your answers digitally via Canvas as **one PDF file**. The file name should follow the convention: `assignment_1_group_xx`, where `xx` corresponds to the group number.
- You submit your assignment with your names, student IDs, and group number contained.
- Use R Studio for the assignment.
- On Canvas, you will find the datasets `twitch.csv` and `whr.csv` that you will need to complete this assignment.
- Please keep the answers to the questions crisp, concise, and to the point. Please adhere to Calibri or Times New Roman 12 pt font with 1.5 line spacing.

Deadline: 4 October, 2022 (23:59, late submissions not allowed)

Exercise I: What makes streamers contribute? Linear regression analysis with cross-sectional data

The data that you will be using in this exercise stems from the live streaming platform twitch.tv. It is the most popular platform for live streaming content and its most popular content is gaming.

Download the `twitch.csv` from Canvas. The dataset contains the following variables.

Variable	Description
<code>user</code>	A unique ID of a streamer
<code>length</code>	How many minutes a streamer streamed on the observation day
<code>t</code>	The observation day
<code>nstreams</code>	how many streams the streamer has started on the observation day
<code>viewgain</code>	The total viewers a streamer gained on the observation day
<code>numgames</code>	The number of games a streamer played on the observation day
<code>maxviewers</code>	The maximum number of viewers viewing the stream at any given point on the observation day
<code>avgviewers</code>	The average number of viewers viewing the stream at any given point on the observation day
<code>followergain</code>	How many followers the streamer gained on the observation day
<code>channelurl</code>	The account name of the streamer
<code>stream_start_hour</code>	The hour the stream started
<code>played_star_game</code>	Whether the streamer played Fortnite (=1) or not (=0)
<code>topgame</code>	The preferred game of the streamer
<code>played_topgame</code>	Whether the streamer played their topgame (=1) or not (=0)
<code>mature</code>	Whether the content is rated mature (=1) or not (=0)
<code>content</code>	Content category
<code>topgame_genre</code>	The genre of the streamer's topgame

- Load the data into R Studio and call it `twitch_data`. Use `stargazer(twitch_data)` to obtain the descriptive statistics of the dataset. Present it in a table (N, mean, standard deviation, min, and max) and briefly discuss the mean values of variables. [max. 1 page]
- Use `length` as the dependent variable and choose three additional independent variables. Use `ggplot` and choose adequate plots to depict the relationship between the `length` and each of the three independent variables. Briefly describe the main insights from each of the three plots. [max. 1 page]
- Run four different linear regression models with `length` as the dependent variable and four different independent variables. For each model, increase the number of independent variables stepwise. E.g., use one independent variable for the first model,

then add a second variable, then add a third variable, and so forth. Also motivate why the chose the variables you chose [max. 1 page]

1. Write down the regression equations of your four models.
 2. Use `stargazer` to present a results table. Describe your resulting coefficients in a concise way (the direction, magnitude, and statistical significance).
 3. Which observation do you make regarding the R^2 when you increase the number of variables?
- d. Focus on the statistically significant variables in your previous analysis and assume the relationships you found are causal. How much would each of the variables have to increase (or decrease) to make the streamer stream 6 hours on the observation day? [max. 0.25 page]
- e. Please elaborate on two theoretical reasons for why the regression coefficients cannot be interpreted as a causal effect in this case. [max. 0.25 page]

Exercise II: What makes countries happy? Linear regression analysis with panel data

The data that you will be using in this exercise stems from the [World Happiness Report](#). It measures a country's aggregated happiness and provides several other variables that might affect it.

Download the `whr.csv` from Canvas. The variables contained in the dataset are explained [here](#).

- a. Load the data into R Studio and call it `whr_data`. Use `stargazer(whr_data)` to obtain the descriptive statistics of the dataset. Present it in a table (N, mean, standard deviation, min, and max) and briefly discuss the mean values of variables. [max. 1 page]
- b. What are the five countries with the highest happiness on average? Is happiness normally distributed? [max. 0.5 page]
- c. Run a regression model with happiness (`Life.Ladder`) as the dependent variable and include at least 5 independent variables of your choice. [max. 2 pages]
 1. Write down the regression equations of your two models.
 2. Run a fixed effects model and a random effects model. Use `stargazer` to present a results table.
 3. Describe your resulting coefficients in a concise way (the direction, magnitude, and statistical significance). When interpreting the coefficients, pay attention to whether variables are logarithms or not.
- d. Please use an appropriate test to decide whether or not you should prefer the random or fixed effects model for your regression specification. Furthermore, elaborate on two theoretical reasons why fixed effects should be preferred. [max. 0.25 page]