# Concordia University

## Department of Economics

**Econ 324**                                                                **Ivan Tchinkov**
**Economic Data Analysis**                                                            **Fall 2023**

# Problem Set # 1

(due online on Sunday, Oct. 15, 11:59 p.m.)

### Instructions:

- Please, write your answers in a separate file (i.e. Word), save it as a PDF document and name it with your last name and your student ID number linked by an underscore (i.e. Yourname_12345678.pdf).

- Answer **all** questions. Clearly indicate which question each answer belongs to.

- Include only the **relevant** parts of the SAS output, summarize the results in a separate table, if needed.

- You need to upload 2 files:

    1. The PDF file described above with your answers and relevant SAS output.
    2. Your SAS code file, with comments, named the same way as the PDF file, with an extension *.sas (i.e. Yourname_12345678.sas).

## I. True/False/Uncertain - Briefly explain. No credit without an explanation (7 marks each).

1. Omitted variable bias (OVB) occurs if the excluded variable is correlated with any of the included variables.

2. The problem with over-fitting is imprecision.

3. With 10 Xs, in the first step of both the forward and backward automatic search procedure, there are 10 regressions.

4. If the confidence interval (CI) for the beta coefficient is (0.98, 1.17), one should fail to reject the null hypothesis $H_0 : e^\beta = 1$.

5. A problem with the Linear Probability Model (LPM) is that $\hat{\beta} \notin [0, 1]$.

## II. Problem - Use SAS for your computations. You have to show your work. No credit without an explanation (8 marks each).

1. We are interested in the factors explaining whether consumers make online purchases or not. A questionnaire was administered with a sample of 435 people. The data are in the file **ps1.sas7bdat** in the 324_data folder. Here is the variables description:

    - Sex: 0=male, 1=female;
    - Age - in years;
    - Purchase: 1=online purchase made last year, 0=no online purchase made last year;

    Variables X1 to X34 are measured on a scale 1 to 7 where 1=strongly disagree and 7=strongly agree and are the answers to the question: "Indicate to what extent you agree with the following statements":

| 1 | I always purchase the types of products I want from the Internet |
|---|---|
| 2 | There is a high risk for purchasing online |
| 3 | Internet retailers encourage me to make suggestions |
| 4 | The Internet retailers' websites provide in-depth information to answer my questions |
| 5 | I can buy the products that are not available in retail shops through the Internet |
| 6 | The website designs of the Internet retailers are aesthetically attractive |
| 7 | Online shopping is not as secure as traditional retail shopping |
| 8 | I do not feel secure about providing my bank card details to a payment platform |
| 9 | I have regular access to a computer |
| 10 | Internet shopping offers a wide variety of products |
| 11 | Online shopping offers better value for my money compared to traditional retail shopping |
| 12 | Family/friends encourage me to make purchases through the Internet |
| 13 | Online shopping allows me to bay the same, or similar products, at cheaper prices than traditional retailing stores |
| 14 | It is easy to receive a personalized customer service from an Internet retailer |
| 15 | Online shopping allows me to save money as I do not need to pay transportation costs |
| 16 | It is quick and easy for me to complete a transaction through the website |
| 17 | It takes only a little time and effort to make a purchase through the Internet |
| 18 | I have knowledge about how to make purchases through the Internet |
| 19 | I have regular access to the Internet |
| 20 | Internet retailers offer good after sales service |
| 21 | Internet retailers understand my needs |
| 22 | I am very skilled at using the Internet |
| 23 | I am not confident that the information I provide to an Internet retailer is not used for other purposes |
| 24 | Internet retailers honour their product guarantees |
| 25 | I am not confident that my personal information is protected by an Internet retailer |
| 26 | The products I ordered are delivered to me within the time promised by the Internet retailers |
| 27 | The quantity and quality of the products I receive from Internet retailers are exactly the same as I order |
| 28 | Internet shopping saves me time, so I can do other activities |
| 29 | The links within the website allow me to move back and forth easily between pages of the website |
| 30 | It is more convenient to shop through the Internet when compared to traditional retail shopping |
| 31 | Internet retailers' websites are easy to navigate |
| 32 | I think the Internet offers lower prices compared to retail stores |
| 33 | Marketing communication influenced my decision to make purchases through the Internet |
| 34 | The media influenced my decision to make purchases through the Internet |

(a) Provide descriptive statistics for the variables "Sex", "Age" and "Purchase".

(b) Present in a table the mean and standard deviation of each of the 34 variables X1-X34 ordered from the highest to the lowest value of the mean. Which of them seem the most important?

(c) Present the means of the 34 Xs for women and men separately. Are there any differences?

(d) Do a Linear Probability Model (LPM) with the variable "Purchase" as a dependent variable and 36 independent variables - 34 Xs, "Sex" and "Age". Test them for statistical significance at 1% and interpret the ones that are.

(e) Do a Logistic regression model (Logit) with the same variables as in (d). Test for statistical significance at 1% and interpret the coefficients that are. Are they different that the ones with the LPM?

(f) Now do a Logit with only X1-X17 in addition to "Age" and "Sex". Which of the two models is preferable - logit with all or with half the Xs in addition to "Age" and "Sex"?

(g) The data file **ps1_g.xlsx** contains one line - the profile of a potential respondent. Using the unrestricted Logit model (with all 36 explanatory variables) estimate the probability and the 95% confidence interval that that person made a purchase last year.

(h) Do the same with the restricted Logit model (with "Age", "Sex" and X1-X17) - which of the two predictions is more precise and why?