Exercise series N°2

Exercise1: Choose the correct answer for the following questions

1. **What is a qualitative variable in statistical modeling?**
   a) A variable that represents quantities and can take any numerical value.
   b) A variable that represents categories or characteristics, often non-numerical.
   c) A variable that is always continuous and measurable.
   d) A variable that only includes time-related data.

2. **Which of the following is an example of a qualitative variable?**
   a) Age of an individual (in years).
   b) Monthly income (in dollars).
   c) Employment status (Employed, Unemployed).
   d) Temperature (in Celsius).

3. **The Logit Model is used to:**
   a) Model a continuous dependent variable based on multiple predictors.
   b) Predict probabilities of outcomes for a binary dependent variable.
   c) Determine the linear relationship between independent and dependent variables.
   d) Identify time-series patterns in data.

4. **The Logit Model formula can be expressed as:**
   a) $P(Y=1) = \beta_0 + \beta_1 X$.
   b) $P(Y=1) = 1/(1 + e^{-(\beta_0 + \beta_1 X)})$
   c) $Y = \ln(\beta_0 + \beta_1 X)$
   d) $Y = e^{\beta_0 + \beta_1 X}$

5. **The Logit Model is best suited for:**
   a) Predicting stock prices.
   b) Modeling binary outcomes such as "success vs. failure."
   c) Determining relationships between continuous variables.
   d) Identifying seasonal trends in data.

6. **In a Logit Model, the coefficients (β) represent:**
   a) The change in the log-odds of the dependent variable for a one-unit change in the predictor.
   b) The probability of success.
   c) The likelihood of multicollinearity.
   d) The mean value of the independent variable.

7. **Which of the following is an application of the Logit Model?**
   a) Predicting the likelihood of a customer purchasing a product (Yes/No).
   b) Estimating the annual revenue of a company.
   c) Modeling temperature changes over time.

d) Determining the variance of residuals in linear regression.

8. The dependent variable in a Logit Model:
a) Can take on any real number.
b) Must be a binary variable (e.g., 0 or 1).
c) Must be a categorical variable with more than two levels.
d) Represents the residual error.

9. If a logistic regression model produces a coefficient of 0.7 **for a variable, what does the corresponding** odds ratio **indicate?**
a) The odds increase by **70%** for a one-unit increase in the independent variable
b) The odds increase by **100%** for a one-unit increase in the independent variable
c) The odds increase by $e^{0.7} \approx$ **2.01 or 101%** for a one-unit increase in the independent variable
d) The probability of the event occurring is exactly 0.7

10. In a dataset for a Logit Model, which of the following is most appropriate for the dependent variable?
a) The height of individuals in centimeters.
b) The income of individuals in dollars.
c) Whether a student passes or fails an exam (Pass = 1, Fail = 0).
d) The number of hours worked per week.

11. Which of the following metrics is commonly used to assess the goodness-of-fit of a logistic regression model?
a) Adjusted R^2
b) Pseudo R^2 (e.g., McFadden's R^2)
c) Mean Squared Error (MSE)
d) Variance Inflation Factor (VIF)

Exercise 2

The goal of this study is to **predict whether a customer will purchase a product (Y=1) or not (Y=0)** based on four key independent variables: **Income** – The financial capability of the customer, which may affect purchasing power. **Age** – The age of the customer, which could influence spending behavior. **Advertisement Exposure** – The number of ads seen by the customer, which may impact product awareness and decision-making. **Product Price** – The cost of the product, which plays a critical role in affordability and purchase likelihood.

Customer_ID	Income	Age	Ad Exposure	Product Price	Purchase Decision
1	40	24	5	150	0
2	60	28	10	180	1
3	55	35	8	190	0
4	70	45	12	250	1
5	30	30	4	120	0
6	50	26	6	160	1
7	45	50	9	200	1
8	65	40	11	220	1
9	80	38	15	270	1
10	20	29	3	100	0
11	55	27	7	160	0
12	50	33	13	210	1
13	40	25	5	130	0
14	60	31	10	190	1
15	45	34	6	150	0
16	35	42	8	140	0
17	55	37	14	240	1
18	70	29	9	210	1
19	50	31	12	200	1
20	30	32	4	110	0
21	65	28	11	230	1
22	40	26	7	150	0
23	55	39	13	220	1
24	45	33	6	160	0
25	50	30	8	170	1
26	60	35	14	240	1
27	45	31	5	140	0
28	80	40	12	260	1
29	30	25	4	130	0
30	50	29	10	180	1
31	55	50	6	160	0
32	65	27	14	250	1
33	40	42	7	150	0
34	75	38	10	210	1
35	30	32	5	120	0
36	50	30	8	170	1
37	60	39	13	220	1
38	45	31	6	150	0
39	80	28	18	280	1
40	50	33	12	190	1

1. Write the logit model, and the logit regression equation

2. Using the given data:Perform logistic regression (using EViews). Obtain the estimated values of β_0, β_1, β_2 and β_3.
3. Explain the effect of independent variables on the likelihood of purchase based on the sign and magnitude of β_1 and β_2, β_3.
4. Predict the probability of purchase for a customer with income 70, age 40, Ads seen 7 and price of 67
5. Which factor has the **strongest influence** on customer purchasing decisions?