Understanding and Applying the Logit Model

Prepared & presented by: Prof. Guechari Yasmina

1. Introduction to Qualitative Variables

1) What are Qualitative Variables?

- Variables that represent categories rather than numbers.
- Examples:
  - ✓ **Binary variables**: Yes/No, Male/Female, Success/Failure, approved/not approved (values: 0 or 1).
  - ✓ **Categorical variables**: Color (Red, Blue, Green), Education Level (Primary, Secondary….).

2) Why Analyze Qualitative Variables?

- To model and predict outcomes like purchase decisions, loans approval, or voting preferences.

2. The Logit Model

- A statistical model used to analyze relationships between one or more independent variables (predictors) and a binary or categorical dependent variable is called Logistic model or logit.

1) Definition

A logit model is commonly used for binary outcomes, such as predicting whether a loan application is approved (1) or rejected (0). The model estimates the probability of an event occurring, given certain independent variables

2) Formula:

$$P(Y = 1/X) = \frac{1}{1+e^{-(\beta_0+\beta_1 X_1+\beta_2 X_2 \cdots\cdots+\beta_k X_k)}}$$

P(Y=1): Probability that the dependent variable Y equals 1

β_0: Intercept.

$\beta_1,\beta_2,\ldots,\beta_k$: Coefficients for the independent variables ($X_1,X_2,\ldots,X_k$): Independent variables

e: Euler's number (approximately 2.718).

2. The Logit Model

- **Log-Odds (Logit) Form**
- The logit regression is the natural logarithm of the odds (logit transformation), the model becomes:
- $\ln(\frac{P(Y=1)}{1-(P(Y=1)}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots\cdots + \beta_k X_k$ **(prove it)**
- Where: $\frac{P(Y=1)}{1-(P(Y=1)}$ is the odds ratio
- Odds is the probability of something occurring dividing by the probability of not occurring.
- $odds\ ratio = \frac{P(Occurring)}{P(not\ occurring)} = \frac{P}{1-P}$

2. The Logit Model

- In logistic regresssion we are estimating an unknown ***P*** for any given linear combination of independent variables.

- In logistic regression we do not know ***P***. so, the goal of logistic regression is to estimate *P* for a linear regression combination of independent variables the estimates is $\hat{P}$

- $\hat{P} = \frac{e^{\beta_0+\beta_1 X_1}}{1+e^{\beta_0+\beta_1 X_1}}$ (estimated probability or estimated regression equation)

2. The Logit Model

3) Applications of the Logit Model

- **Example 1:**
- **Objective**: Predict whether a customer will purchase a product (Y=1) or not (Y=0).
- **Independent Variables**:

✓Income Age

✓Advertisement exposure (number of ads seen)

✓Product price

- **Application**: Optimizing marketing campaigns by identifying factors that influence purchasing decisions.
- $P(Y = 1) = \frac{1}{1+e^{-(\beta_0+\beta_1 Inc+\beta_2 Age+\beta_3 Adv+\beta_4 Price)}}$

2. The Logit Model

3) Applications of the Logit Model

- **Example 2:**
- **Objective**: Predict whether a student will pass (Y=1) or fail (Y=0) a course.
- **Independent Variables**:

✓Attendance rate

✓Hours of study per week

✓GPA from previous semesters

- **Application**: Identifying at-risk students and designing interventions to improve performance.
- $P(Y = 1) \frac{1}{1+e^{-(\beta_0+\beta_1 Att+\beta_2 Hour+\beta_3\ GPA)}}$

2. The Logit Model

3) Applications of the Logit Model

- **Example 3:**
- **Objective**: Predict whether a customer will default on their credit card payment (Y=1) or not (Y=0).
- **Independent Variables**:

✓Monthly income

✓Credit utilization ratio

✓Previous payment history

✓Outstanding debt

- **Application**: Assessing credit risk and setting credit limits for customers.

2. The Logit Model

3) Applications of the Logit Model

- **Example 4:**
- **Objective**: Predict the mode of transportation chosen (Y=1 for public transport, Y=0 for private car).
- **Independent Variables**:

✓Income

✓Distance to travel

✓Cost of public transport

✓Availability of parking

- **Application**: Urban planning and improving public transport systems.

2. The Logit Model

3) Applications of the Logit Model

- **Example 5:**
- **Objective:** Predict whether an employee will leave (Y=1) or stay (Y=0) in company.
- **Independent Variables**:
  - Job satisfaction
  - Salary
  - Work-life balance
  - Length of employment
- **Application**: Reducing employee turnover by addressing key factors.

3. Example

- **Numerical example 1:** A company wants to predict whether a customer will purchase a product (Yes = 1, No = 0) based on:
- **Income (X1):** Monthly income of the customer (in $).
- **Advertisement Exposure (X2):** Number of ads the customer has seen.
- **Data:**

Income	Ads Seen	Purchase
2000	3	0
3000	5	0
4000	7	1
2500	4	1
4500	8	1

3. Example

- **Estimating a Logit Model in EViews**

1. **Prepare the Data**:
    1. Ensure your dataset is loaded into EViews.
    2. Include the dependent variable (binary, coded as 0 and 1) and independent variables.
2. **Specify the Equation**:
    1. Go to **Quick > Estimate Equation** or open the Equation Estimation window.
    2. In the Equation Estimation dialog, specify your model in the format:

3. Example

- Dependent Variable C Independent Variables
- Purchase c Income Ads_Seen
- In the **Estimation Method**, select Binary Logit.
- Click **OK** to estimate the model.
- EViews will display the results, including coefficients, standard errors, and statistical significance.

2) Eviews results:

- EViews estimates:
- Interceptβ_0=−4. 38; β_1= - 0.0001 (income) β_2= 1.013 (Ads seen).

3. Example

- Interpretation: The estimated logistic model is:
- $\text{logit(Purchase)} = \beta_0 + \beta_1 \cdot \text{Income} + \beta_2 \cdot \text{Ads Seen}$
- Substituting the coefficients:
- $\text{logit(Purchase)} = -4.38 - 0.0001 \cdot \text{Income} + 1.013 \cdot \text{Ads Seen}$.
- Assume that the p-value associated with coefficient is less than 5%
- Here, the dependent variable (Purchase) is binary (0 or 1), and the independent variables are Income and Ads Seen.

3. Example

- Since **β_1 = -0.0001** is very close to zero, the effect of income on purchase behavior is **minimal**.
- To translate this into odds, we exponentiate the coefficient:$e^{-0.0001} \approx 0.9999$, This means that for every **one-unit increase in income**, the odds of making a purchase **decrease by approximately 0.01%** (0.9999**-** 1= -0.00011 or **0.01% decrease** in odds).
- A one-dollar increase in income has a negligible effect on purchasing decision.

3. Example

- exposure to advertisements has a strong positive association with the likelihood of purchasing.
- To translate this into odds, we exponentiate the coefficient: $e^{1.013} \approx 2.75$.
- This means that for every one additional ad seen, the odds of making a purchase increase by approximately 175% ($2.75 - 1 = 1.75$, or a 175% increase in odds).This suggests that advertising has a substantial impact on purchase behavior.

3. Example

- The estimated logit regression is:
- $ln\ (\frac{p}{1-p})$=$\beta_0 + \beta_1 income + \beta_2 add\ seen$
- $ln\ (\frac{p}{1-p})$=$-4.38 - 0.0001 Incom + 1.013 Add\ seen$
- The estimated probability that the customer with Income=3000 and Ads Seen=5 purchase the product is
- $\boldsymbol{\widehat{P}} = \frac{\boldsymbol{e^{(-4.38-0.0001*\ 3000+1.013*\ 5)}}}{\boldsymbol{1+e^{(-4.38-0.0001*3000+1.013*5)}}}$ =$\frac{e^{0.385}}{1+e^{0.385}}$ = 0.595
- The predicted probability of purchase is **59.5%** for this customer.

3. Example

- **Insights**
- **Ads Seen** is the most important factor influencing purchase decision, with a strong positive relationship.
- **Income** has a very weak and slightly negative impact on purchase likelihood.
- Additional variables or interaction effects might improve the model's explanatory power.

3. Example

- **Numerical example 2: Example: Predicting Loan Approval**

Applicant	Income (X_1)	Credit Score (X_2)	Loan Approval (Y)
1	45,000	680	1
2	42,000	700	0
3	50,000	720	1
4	48,000	650	0
5	55,000	730	1
6	30,000	600	0
7	35,000	670	1
8	28,000	590	0
9	60,000	740	1
10	25,000	560	0

3. Example

- **Model Specification**: The logit regression for loan approval could be written as:
- $\text{Logit}(P(Y=1))=\beta_0+\beta_1\cdot\text{Income}+\beta_2\cdot\text{Credit Score}$
- $P(Y=1)$: Probability that the loan is approved.
- Independent variables: X_1: Income. X_2: Credit Score
- **Interpretation of Coefficients**:
- Intercept (β_0): odds of loan approval when all independent variables are 0.
- Slope Coefficients (β_1,β_2): A positive β_1 indicates higher income increases the probability of loan approval.
- A positive β_2 indicates a higher credit score increases the probability of loan approval.

3. Example

1) Eviews steps

- Save the data in a CSV file or enter it directly into EViews.
- **Import into EViews**:
- Open the file in EViews using **File → Open → Foreign Data as Workfile**.
- **Estimate the Model**:
- Go to **Quick → Estimate Equation** and enter the model:
- Loan_Approval c Income Credit_Score
- Select **Binary → Logit** as the estimation method.

3. Example

Dependent Variable: LOAN_APPROVAL
Method: ML - Binary Logit (Newton-Raphson / Marquardt steps)
Date: 01/11/25 Time: 16:44
Sample: 1 10
Included observations: 10
Convergence achieved after 5 iterations
Coefficient covariance computed using observed Hessian

Variable	Coefficient	Std. Error	z-Statistic	Prob.
INCOME	-0.034876	0.175205	-0.199061	0.8422
CREDIT_SCORE	0.053162	0.040060	1.327070	0.1845
C	-34.24072	22.91808	-1.494048	0.1352

McFadden R-squared	0.514762	Mean dependent var	0.500000
S.D. dependent var	0.527046	S.E. of regression	0.398892
Akaike info criterion	1.272682	Sum squared resid	1.113805
Schwarz criterion	1.363458	Log likelihood	-3.363412
Hannan-Quinn criter.	1.173102	Deviance	6.726824
Restr. deviance	13.86294	Restr. log likelihood	-6.931472
LR statistic	7.136119	Avg. log likelihood	-0.336341
Prob(LR statistic)	0.028211		
Obs with Dep=0	5	Total obs	10
Obs with Dep=1	5		

3. Example

2) Eviews results:

- EViews estimates:
- Interceptβ_0=−34.24; β_1= - 0.034 (income) β_2= 0.053 (Credit Score).

Example (interpretation)

- The logit regression shows that the credit score has a coefficient of 0.053, indicating a weak positive relationship with the likelihood of loan approval.
- A one-unit increase in credit score increases the odds of loan approval by approximately 5.4%.
- Since the **p-value is 0.18 (> 0.05)**, we **cannot conclude that this relationship is statistically significant**. This means that while a positive trend is observed, it **may not be reliable** for decision-making.

Example (interpretation)

- The coefficient of **-0.0348** suggests a **negative relationship** between **income and loan approval probability**. This means that, **as income increases, the odds of loan approval slightly decrease**.
- Since **-0.0348** is a small value, the effect of income on loan approval is **minimal**.
- Converting this to odds: $e^{-0.0348} \approx 0.966$
- This means that for **each one-unit increase in income, the odds of loan approval decrease by about 3.4%**(0.966 - 1 = - 0.034 or **3.4%**).
- However, the very high p-value (0.8422) indicates that this effect is not statistically significant, meaning there is no strong evidence that income influences loan approval in this model."

- The probability that applicant with credit score 700 and income 45, get his request approved is
- $\widehat{\boldsymbol{P}} = \frac{e^{\beta_0+\beta_1 income+\beta_2 credit\ score}}{1+e^{\beta_0+\beta_1 income+\beta_2 credit\ score}}$
- Logit(P(Y=1))=-34,24 – 0,034·Income+0,053·Credit Score
- Logit(P(Y=1))=-34,24 – 0,034·(48)+0,053·(700)=1,19
- $\widehat{\boldsymbol{P}} = \frac{e^{1,19}}{1+e^{1,19}}$=0.768
-

Applicant	Credit Score	Income	Loan Amoun	Default
1	720	5,5	15	0
2	680	4,2	20	1
3	650	3,8	18	0
4	710	6	12	1
5	600	2,9	25	1
6	580	2,5	30	0
7	750	6,8	10	1
8	690	4,5	16	0
9	630	3,2	22	1
10	770	7,5	8	0