ICMK352
Marketing Intelligence

8

Logistic Regression Analysis

Marketing Intelligence
Assoc. Prof. Dr. Viriya Taecharungroj
Mahidol University International College

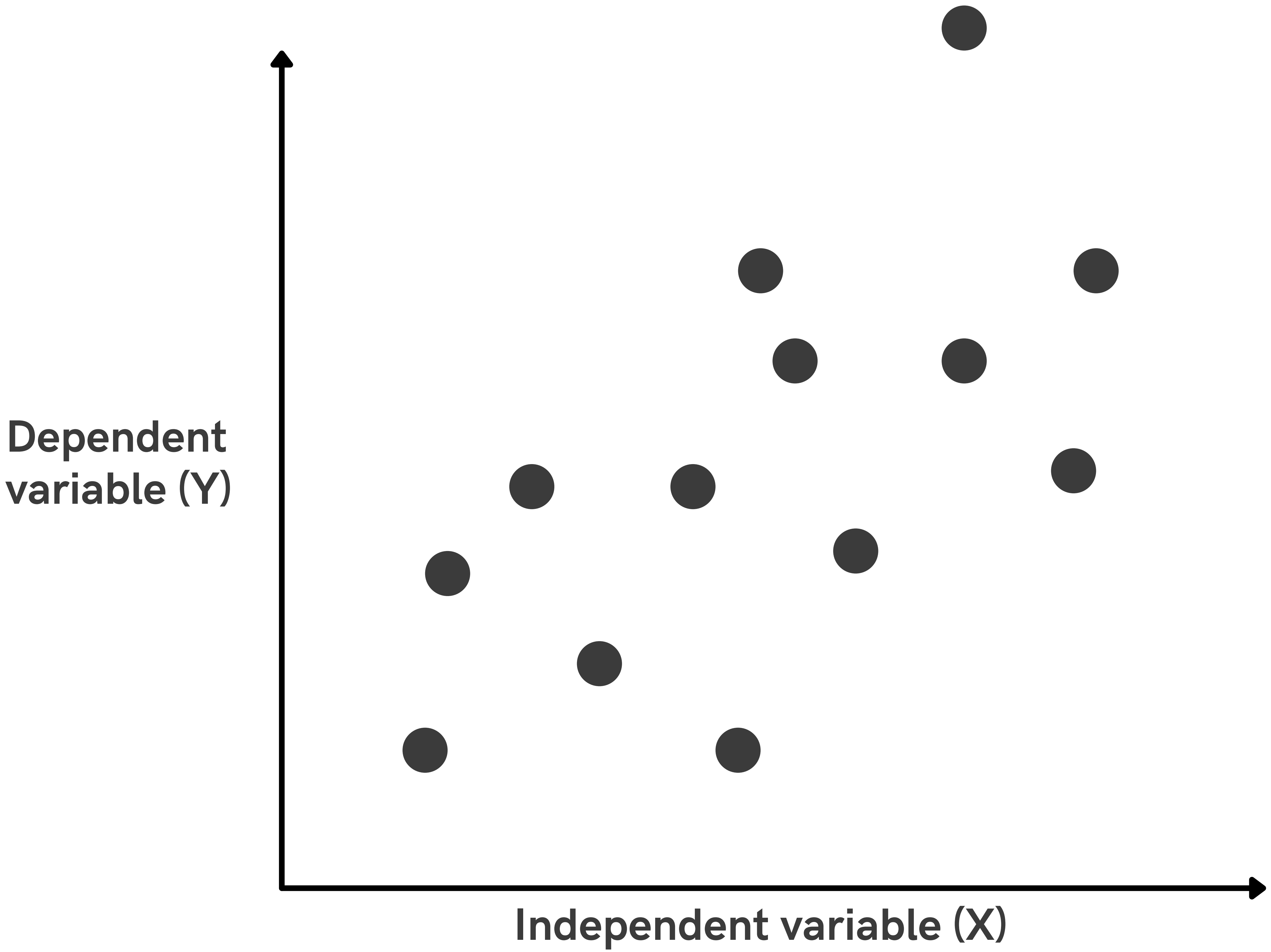
Logistic Regression Analysis

Key topics for discussion

01

Logistic regression model

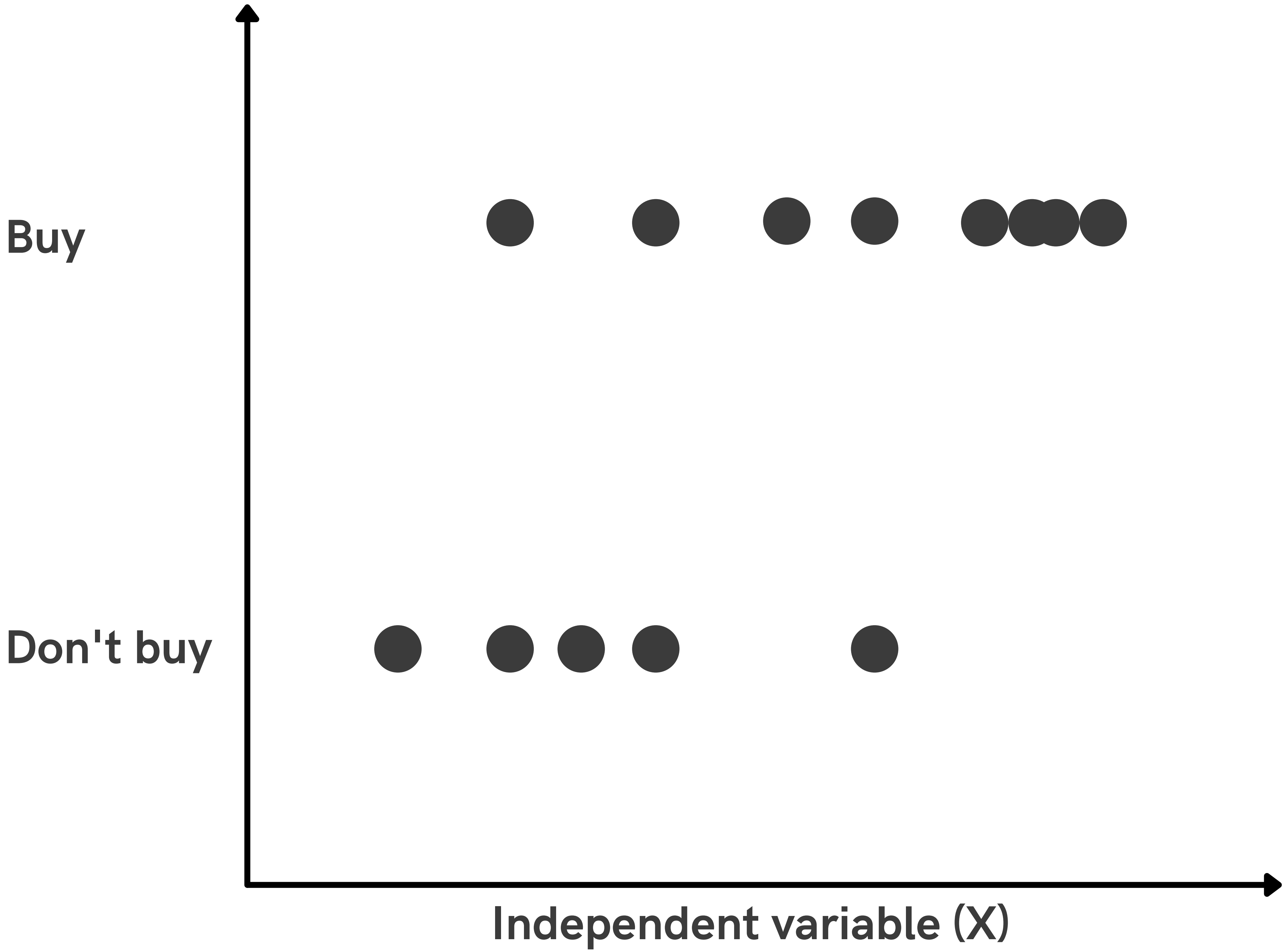
Relationship	Hypothesis examples	Technique
Categorical >>> Numerical	• Does cluster membership affect satisfaction? • Is there a difference in expected design between genders?	ANOVA / T-TEST
Categorical >>> Categorical	• Does income levels affect cluster membership?	Chi-square
Numerical >>> Categorical	• Do perceived design, ease of use, and safety affect purchase decision?	Logistic Regression
Numerical >>> Numerical	• Do perceived design, ease of use, and safety affect satisfaction?	Linear Regression



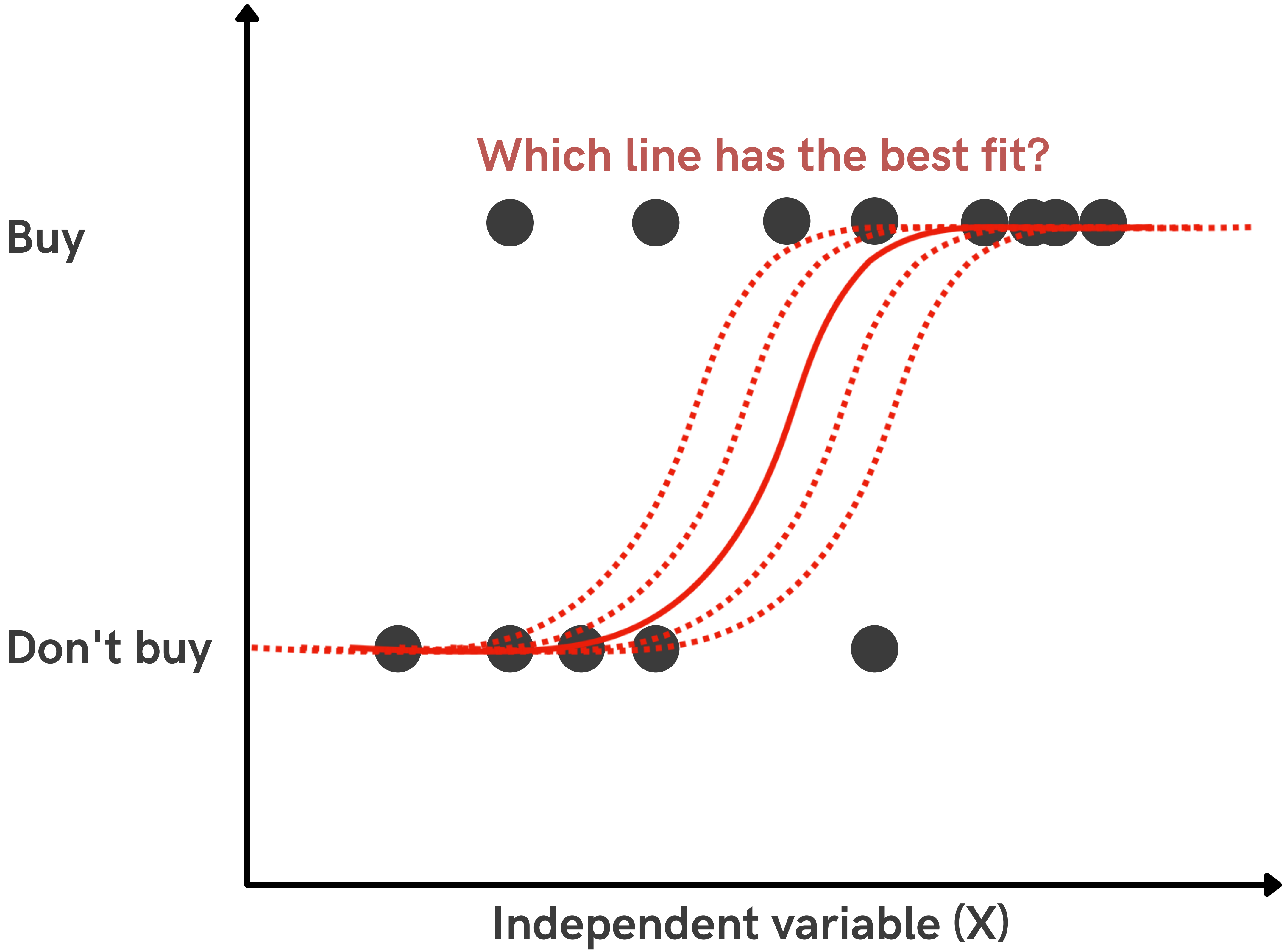
Regression Analysis

- In regression, the independent variable(s) is/are used to predicted dependent variable.
- The least squares criterion guarantees that the “best” straight-line slope and intercept will be calculated

Logistic Regression



- the appropriate regression analysis to conduct when the dependent variable is dichotomous (binary)
- It is used to describe data and to explain the relationship between one dependent binary variable and one or more nominal, ordinal, interval or ratio-level independent variables..

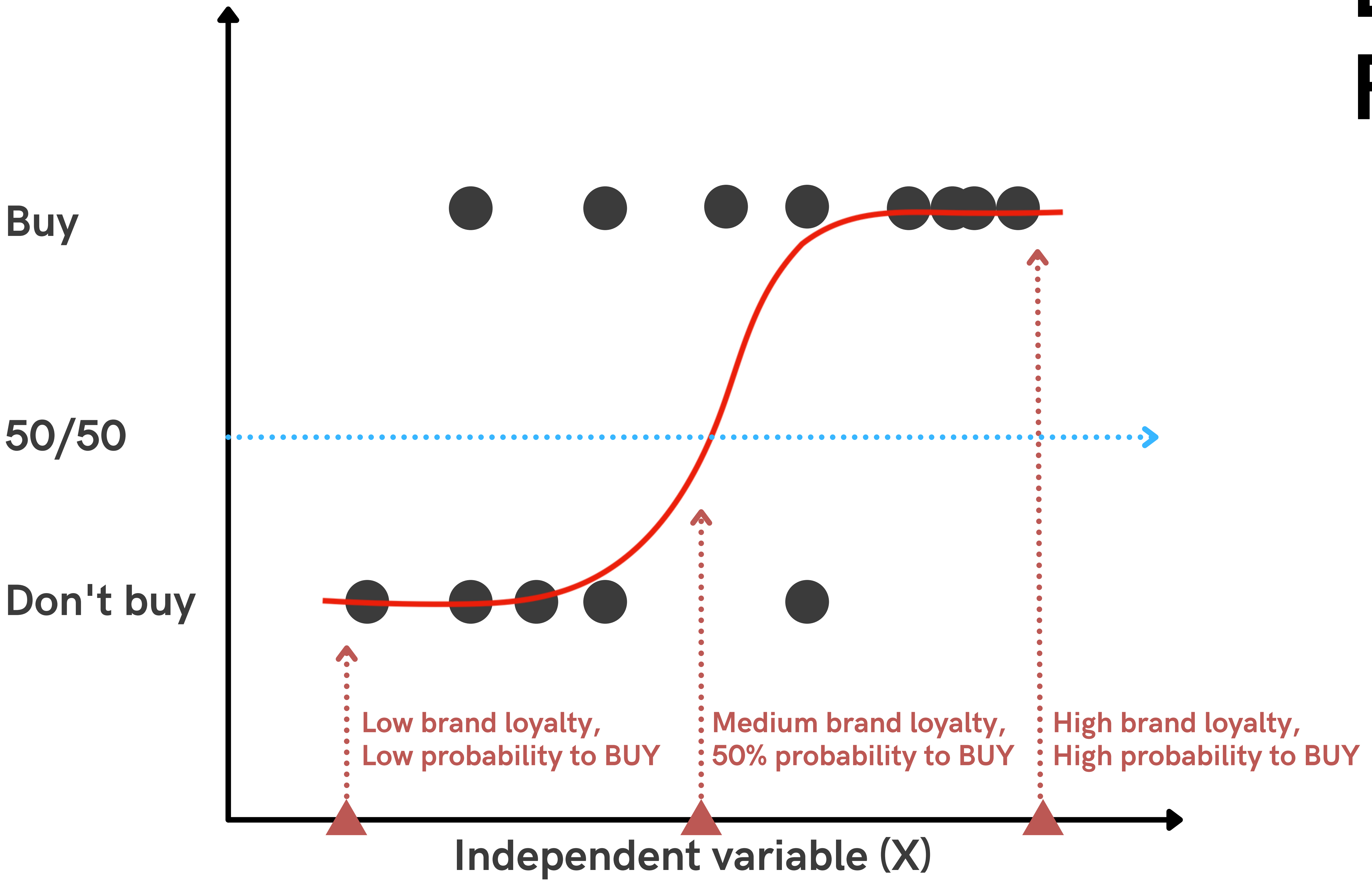


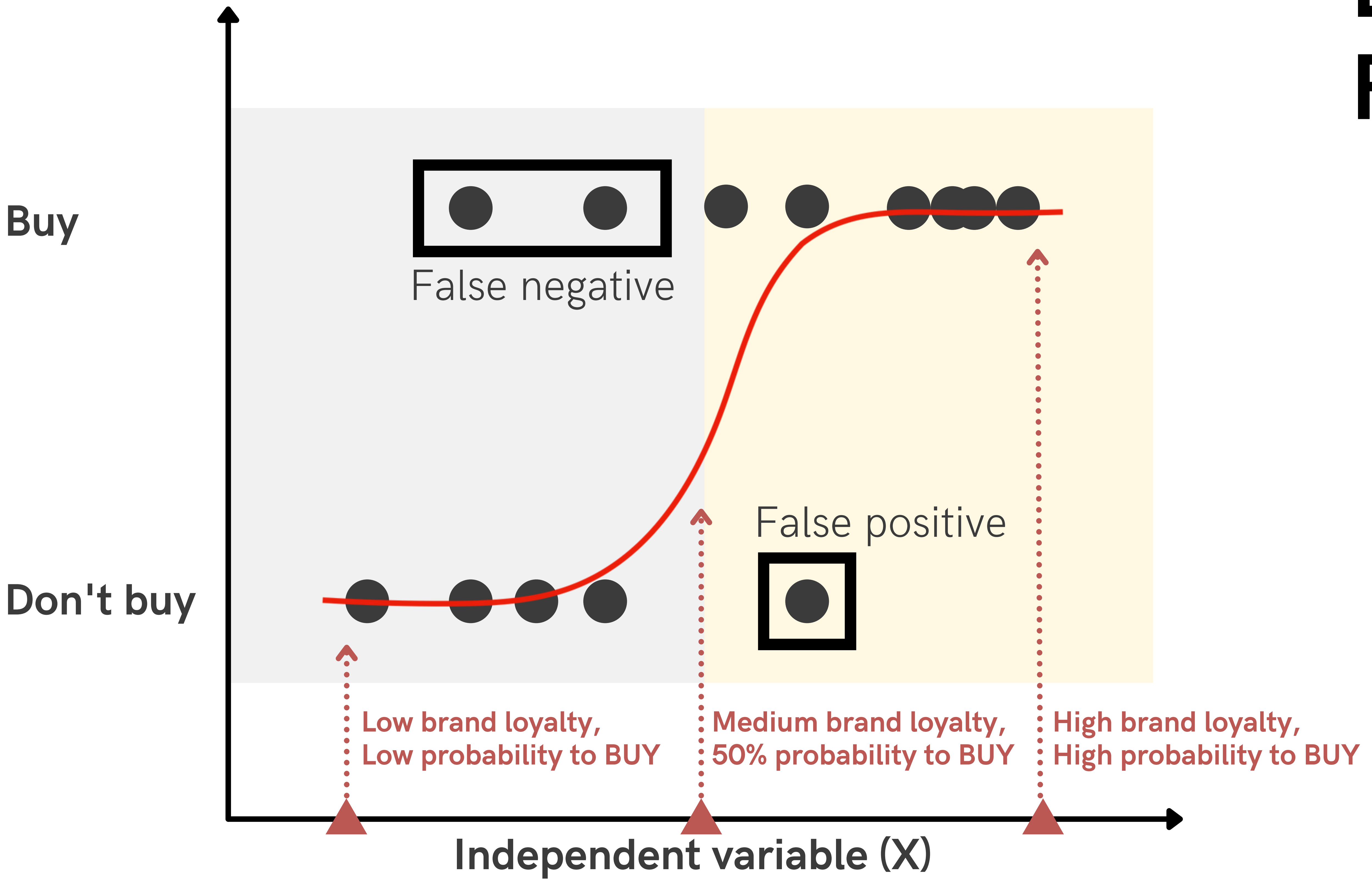
Logistic Regression

- the appropriate regression analysis to conduct when the dependent variable is dichotomous (binary)
- It is used to describe data and to explain the relationship between one dependent binary variable and one or more nominal, ordinal, interval or ratio-level independent variables..

Logistic Regression

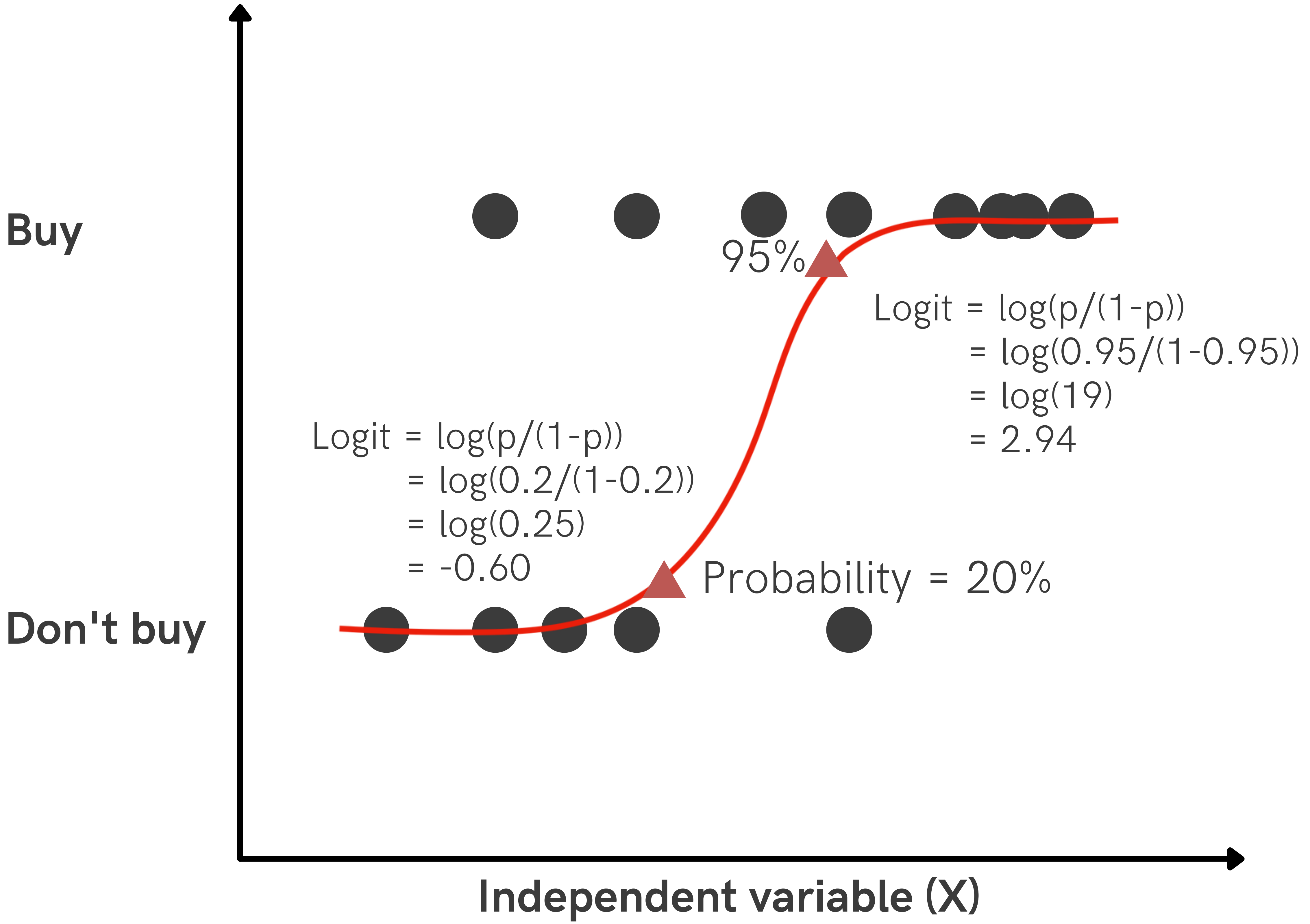
- We can use this line to predict the dependent variable





Logistic Regression

- Logistic regression is usually used for classification. For example, if the probability is greater than 50%, we classify it as such.
- Unlike linear regression which use a "least squares" technique, logistic regression uses "maximum likelihood" technique.
- Probability = p
- Odds = (p/(1-p))
- Logit = log(odds)

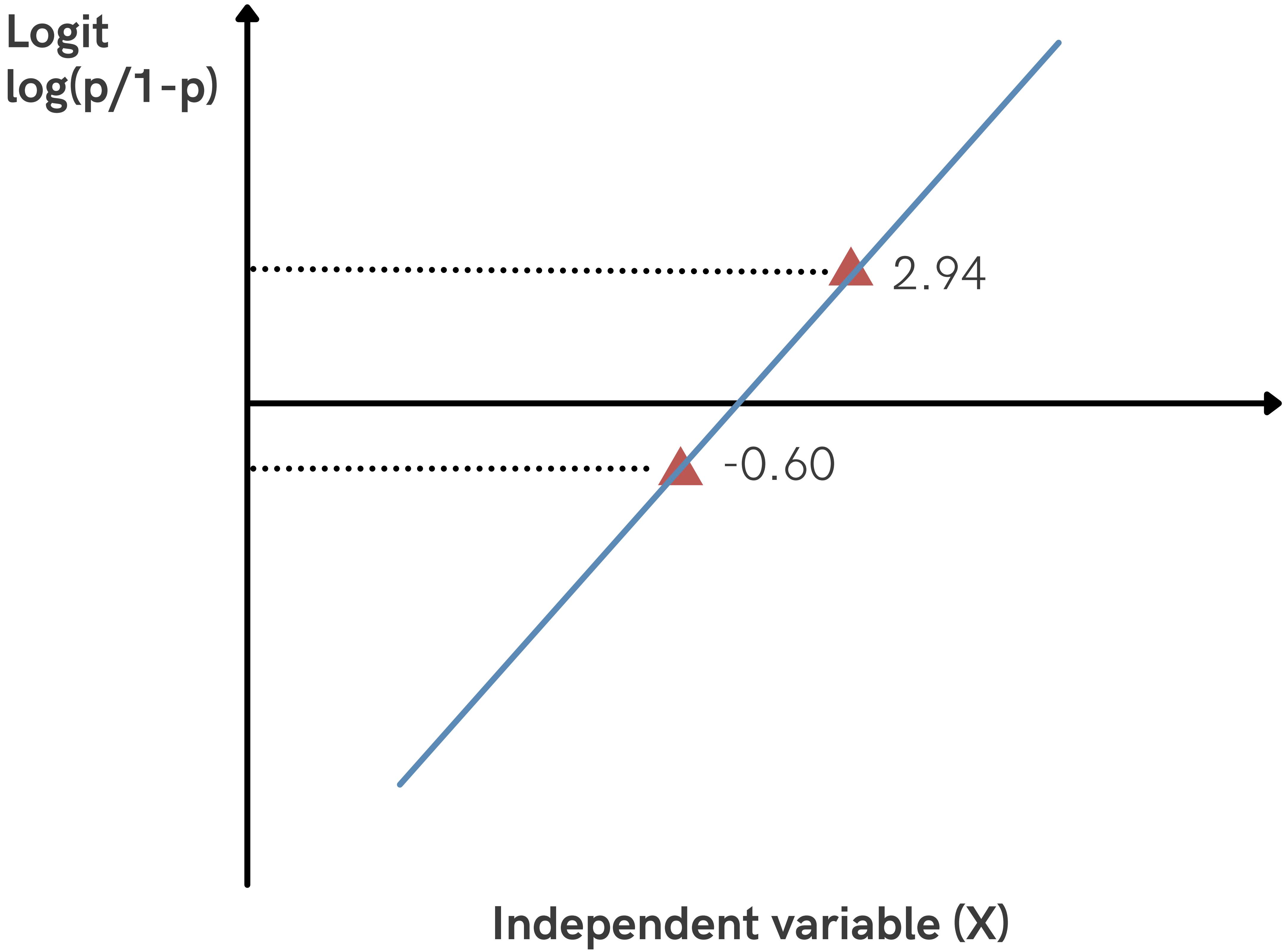


Logistic Regression

- Probability = p
- Odds = (p/1-p)
- Logit = log(odds)

Rationale

- When you transform probability to logit log(odds), you effectively transformed 0.00-1.00 to negative infinity and positive infinity.



Logistic Regression

- Probability = p
- Odds = (p/1-p)
- Logit = log(odds)

Rationale

- When you transform probability to logit log(odds), you effectively transformed 0.00-1.00 to negative infinity and positive infinity.
- By using logit, we can fit a straight line with y-intercept and a slope like a linear regression. However, we cannot interpret those values in the same way.

Logistic Regression

```
Coefficients:
                Estimate Std. Error z value Pr(>|z|)    
(Intercept)     -7.71515    1.87824  -4.108 4.00e-05 ***
iphone_design    0.68128    0.16767   4.063 4.84e-05 ***
iphone_function -0.07184    0.15497  -0.464 0.642934    
iphone_os       -0.30606    0.22977  -1.332 0.182852    
iphone_brand     0.60472    0.16843   3.590 0.000330 ***
iphone_value     0.45547    0.12126   3.756 0.000173 ***
iphone_social    0.31925    0.12855   2.483 0.013013 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Interpreting independent variable’s estimates

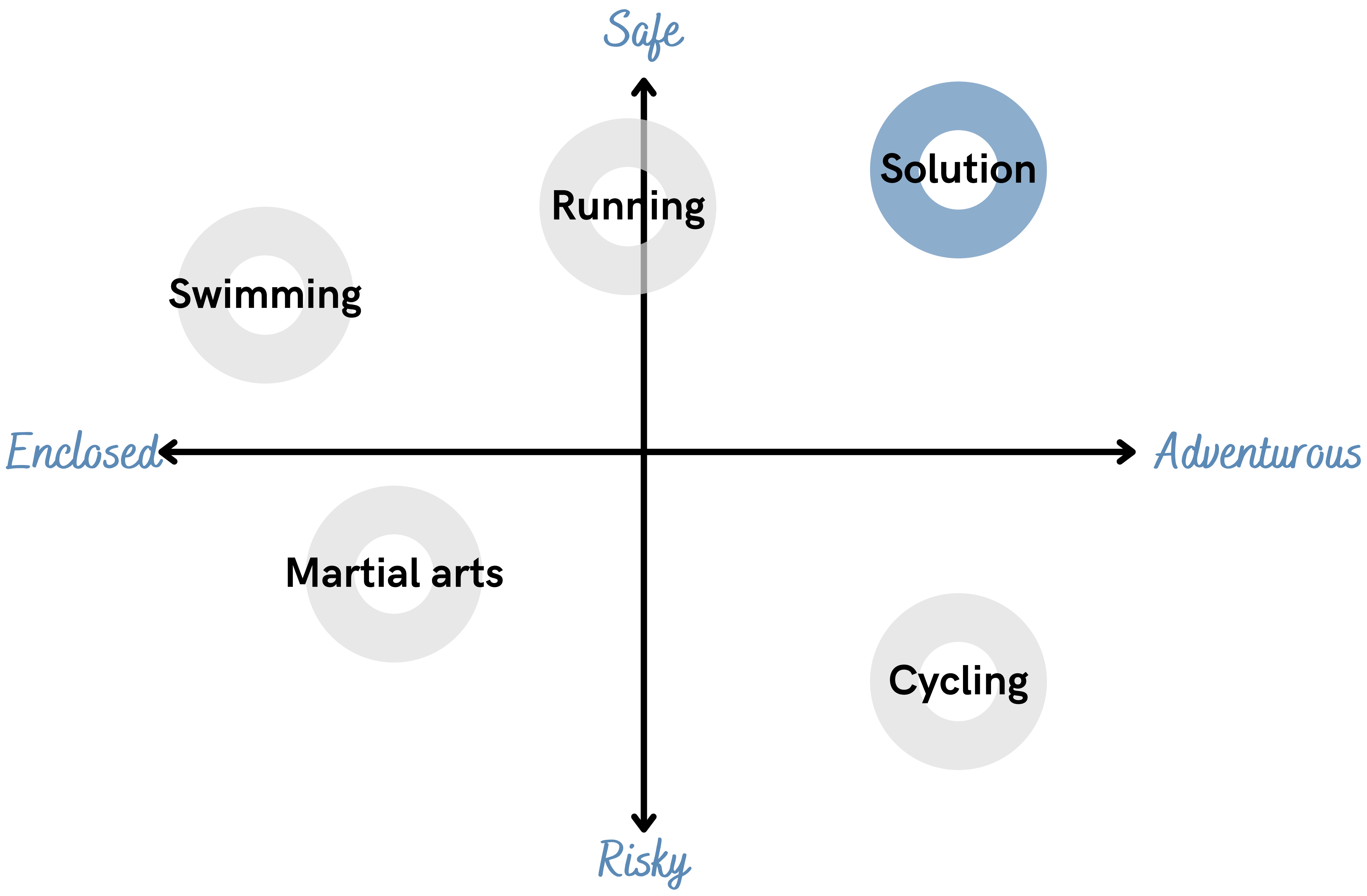
For example, beta (estimate) of design
Logit = 0.68128
Logit = log(odds)
Odds = exp(logit)
= exp(0.68128)
= 1.977473
A one unit increase of a_safe leads to a 98% (1.98-1) increase in odds of purchase intention (Yes)

Interpreting Intercept

- If all independent variables are 0, what is the logit (probability) of a dependent variable being true?

Logit = -7.71515
Logit = log(odds)
Odds = exp(logit)
= exp(-7.71515)
= 0.0004215741
Odds = p/(1-p)
p = <0.001%

Competition



Logistic Regression

- It helps you understand what leads to a particular decision of the customers either desirable or not
- It helps you predict customer's future behaviours and plan interventions

Logistic Regression

- Logistic regression analysis can be used to examine the relationship between a binary dependent variable and one or more independent variables.
- In R, the "glm()" function can be used to perform logistic regression analysis.
- To perform logistic regression analysis between the "iphone_intention" variable and the "iphone_design", "iphone_function", "iphone_os", "iphone_brand", "iphone_value", and "iphone_social" variables in the "iphone_dataframe" dataset, use "glm_iphone_quality <- glm(data = iphone_dataframe, formula = iphone_intention ~ iphone_design + iphone_function + iphone_os + iphone_brand + iphone_value + iphone_social, family = "binomial")".

```
# 7. Logistic regression ----
## 7.1 Logistic regression analysis using glm() to test the perceived quality on purchase intention ----
glm_iphone_quality <- glm(data = iphone_dataframe, formula = iphone_intention ~ iphone_design + iphone_function + iphone_os + iphone_brand + iphone_value + iphone_social,
          family = "binomial")
```

Logistic Regression

- The resulting object "glm_iphone_quality" will contain the logistic regression analysis results.
- The "family = "binomial"" argument specifies the type of model to be used (i.e., binomial logistic regression).
- Other types of logistic regression models include multinomial logistic regression and ordinal logistic regression, which are used when the dependent variable has more than two categories or when the categories have a natural order, respectively.
- To view the results of the logistic regression analysis, use "summary(glm_iphone_quality)". The output will display the coefficients and standard errors for each independent variable, the odds ratios and 95% confidence intervals, and the overall model significance.

```
## Check results
summary(glm_iphone_quality)
```

```
Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)     -4.14240    1.49940  -2.763 0.005733 **
iphone_design   -0.10699    0.14283  -0.749 0.453816
iphone_function  0.23020    0.10757   2.140 0.032354 *
iphone_os        0.10965    0.12148   0.903 0.366756
iphone_brand     0.07688    0.14366   0.535 0.592554
iphone_value     0.39235    0.10281   3.816 0.000136 ***
iphone_social    0.36792    0.11092   3.317 0.000910 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 300.24  on 299  degrees of freedom
Residual deviance: 268.69  on 293  degrees of freedom
AIC: 282.69
```

Logistic Regression

- Another way to view the results is to use the "tidy()" function from the "broom" package. This will display the coefficients, standard errors, and confidence intervals in a more readable format, as well as the log-odds of each predictor variable.
- To view the odds ratios and corresponding 95% confidence intervals for each predictor variable, you can use the "exp()" function to exponentiate the log-odds. The resulting values represent the multiplicative increase in odds associated with a one-unit increase in the predictor variable, holding all other variables constant.
- You can also calculate the percentage increase in odds associated with a one-unit increase in the predictor variable by subtracting one from the exponentiated value and multiplying by 100.

```
## zoom in to the coefficient (log-odds)
tidy(glm_iphone_quality)

## 7.3 Predict the result using glm_iphone_quality model ----
## exp() exponentiate the log-odd to display the odds
tidy(glm_iphone_quality)[,1:2] |> mutate(odds = exp(estimate))
tidy(glm_iphone_quality)[,1:2] |> mutate(odds = exp(estimate),
                              percentage_increase = (exp(estimate) -
1) * 100)
```

A tibble: 7 × 4

	term	estimate	odds	percentage_increase
	<chr>	<dbl>	<dbl>	<dbl>
1	(Intercept)	-4.14	0.0159	-98.4
2	iphone_design	-0.107	0.899	-10.1
3	iphone_function	0.230	1.26	25.9
4	iphone_os	0.110	1.12	11.6
5	iphone_brand	0.0769	1.08	7.99
6	iphone_value	0.392	1.48	48.0
7	iphone_social	0.368	1.44	44.5