

Logistic Regression

BUS 735: Business Decision Making and Research

- Specific goals:
 - Learn how to conduct regression analysis with a dummy independent variable.
- Learning objectives:
 - LO2: Be able to construct and use multiple regression models (including some limited dependent variable models) to construct and test hypotheses considering complex relationships among multiple variables.
 - LO6: Be able to use standard computer packages such as R to conduct statistical analysis.
 - LO7: Have a sound familiarity of various statistical and quantitative methods in order to be able to approach a business decision problem and be able to select appropriate methods to answer the question.

- Specific goals:
 - Learn how to conduct regression analysis with a dummy independent variable.
- Learning objectives:
 - LO2: Be able to construct and use multiple regression models (including some limited dependent variable models) to construct and test hypotheses considering complex relationships among multiple variables.
 - LO6: Be able to use standard computer packages such as R to conduct statistical analysis.
 - LO7: Have a sound familiarity of various statistical and quantitative methods in order to be able to approach a business decision problem and be able to select appropriate methods to answer the question.

- **Logistic Regression:** method for estimating a regression with a *dummy dependent variable*.
- Will a potential customer purchase a product? (YES=1, NO=0).
 - Might use explanatory variables: age, gender, income, etc.
- Will a potential employee be retained after one year? (YES=1, NO=0).
 - Might use explanatory variables: age, gender, years experience, past income, education dummy (4-year = 1, otherwise = 0).
- Why not run a regression? Which assumption is violated?

- **Logistic Regression:** method for estimating a regression with a *dummy dependent variable*.
- Will a potential customer purchase a product? (YES=1, NO=0).
 - Might use explanatory variables: age, gender, income, etc.
- Will a potential employee be retained after one year? (YES=1, NO=0).
 - Might use explanatory variables: age, gender, years experience, past income, education dummy (4-year = 1, otherwise = 0).
- Why not run a regression? Which assumption is violated?

- **Logistic Regression:** method for estimating a regression with a *dummy dependent variable*.
- Will a potential customer purchase a product? (YES=1, NO=0).
 - Might use explanatory variables: age, gender, income, etc.
- Will a potential employee be retained after one year? (YES=1, NO=0).
 - Might use explanatory variables: age, gender, years experience, past income, education dummy (4-year = 1, otherwise = 0).
- Why not run a regression? Which assumption is violated?

- **Logistic Regression:** method for estimating a regression with a *dummy dependent variable*.
- Will a potential customer purchase a product? (YES=1, NO=0).
 - Might use explanatory variables: age, gender, income, etc.
- Will a potential employee be retained after one year? (YES=1, NO=0).
 - Might use explanatory variables: age, gender, years experience, past income, education dummy (4-year = 1, otherwise = 0).
- Why not run a regression? Which assumption is violated?

- **Logistic Regression:** method for estimating a regression with a *dummy dependent variable*.
- Will a potential customer purchase a product? (YES=1, NO=0).
 - Might use explanatory variables: age, gender, income, etc.
- Will a potential employee be retained after one year? (YES=1, NO=0).
 - Might use explanatory variables: age, gender, years experience, past income, education dummy (4-year = 1, otherwise = 0).
- Why not run a regression? Which assumption is violated?

- **Logistic Regression:** method for estimating a regression with a *dummy dependent variable*.
- Will a potential customer purchase a product? (YES=1, NO=0).
 - Might use explanatory variables: age, gender, income, etc.
- Will a potential employee be retained after one year? (YES=1, NO=0).
 - Might use explanatory variables: age, gender, years experience, past income, education dummy (4-year = 1, otherwise = 0).
- Why not run a regression? Which assumption is violated?
Dependent variable is not interval. As a result, residual will not be normally distributed.

- Normal regression:

$$y_i = b_0 + b_1X_{1,i} + b_2X_{2,i} + \dots + b_{k-1}X_{k-1,i} + e_i$$

- Logistic regression:

$$\log(\text{Odds}) = b_0 + b_1X_{1,i} + b_2X_{2,i} + \dots + b_{k-1}X_{k-1,i} + e_i$$

$$\text{Odds} = \frac{P(y_i = 1)}{1 - P(y_i = 1)} = \frac{P(y_i = 1)}{P(y_i = 0)}$$

- Remember, $y_i \in 0, 1$ indicates YES=1 event did occur, or NO=0 event did not occur.

- Normal regression:

$$y_i = b_0 + b_1X_{1,i} + b_2X_{2,i} + \dots + b_{k-1}X_{k-1,i} + e_i$$

- Logistic regression:

$$\log(\text{Odds}) = b_0 + b_1X_{1,i} + b_2X_{2,i} + \dots + b_{k-1}X_{k-1,i} + e_i$$

$$\text{Odds} = \frac{P(y_i = 1)}{1 - P(y_i = 1)} = \frac{P(y_i = 1)}{P(y_i = 0)}$$

- Remember, $y_i \in 0, 1$ indicates YES=1 event did occur, or NO=0 event did not occur.

- Normal regression:

$$y_i = b_0 + b_1X_{1,i} + b_2X_{2,i} + \dots + b_{k-1}X_{k-1,i} + e_i$$

- Logistic regression:

$$\log(Odds) = b_0 + b_1X_{1,i} + b_2X_{2,i} + \dots + b_{k-1}X_{k-1,i} + e_i$$

$$Odds = \frac{P(y_i = 1)}{1 - P(y_i = 1)} = \frac{P(y_i = 1)}{P(y_i = 0)}$$

- Remember, $y_i \in 0, 1$ indicates YES=1 event did occur, or NO=0 event did not occur.

- Normal regression:

$$y_i = b_0 + b_1 X_{1,i} + b_2 X_{2,i} + \dots + b_{k-1} X_{k-1,i} + e_i$$

- Logistic regression:

$$\log(\text{Odds}) = b_0 + b_1 X_{1,i} + b_2 X_{2,i} + \dots + b_{k-1} X_{k-1,i} + e_i$$

$$\text{Odds} = \frac{P(y_i = 1)}{1 - P(y_i = 1)} = \frac{P(y_i = 1)}{P(y_i = 0)}$$

- Remember, $y_i \in 0, 1$ indicates YES=1 event did occur, or NO=0 event did not occur.

- Normal regression:

$$y_i = b_0 + b_1X_{1,i} + b_2X_{2,i} + \dots + b_{k-1}X_{k-1,i} + e_i$$

- Logistic regression:

$$\log(Odds) = b_0 + b_1X_{1,i} + b_2X_{2,i} + \dots + b_{k-1}X_{k-1,i} + e_i$$

$$Odds = \frac{P(y_i = 1)}{1 - P(y_i = 1)} = \frac{P(y_i = 1)}{P(y_i = 0)}$$

- Remember, $y_i \in 0, 1$ indicates YES=1 event did occur, or NO=0 event did not occur.

- Normal regression:

$$y_i = b_0 + b_1X_{1,i} + b_2X_{2,i} + \dots + b_{k-1}X_{k-1,i} + e_i$$

- Logistic regression:

$$\log(Odds) = b_0 + b_1X_{1,i} + b_2X_{2,i} + \dots + b_{k-1}X_{k-1,i} + e_i$$

$$Odds = \frac{P(y_i = 1)}{1 - P(y_i = 1)} = \frac{P(y_i = 1)}{P(y_i = 0)}$$

- Remember, $y_i \in 0, 1$ indicates YES=1 event did occur, or NO=0 event did not occur.

- Predicted value from a regular regression:

$$\hat{y}_i = b_0 + b_1X_{1,i} + b_2X_{2,i} + \dots + b_{k-1}X_{k-1,i}$$

- For a logistic regression, you can get predicted logist (not too interesting yet):

$$\hat{l}_i = \ln(\widehat{Odds}) = b_0 + b_1X_{1,i} + b_2X_{2,i} + \dots + b_{k-1}X_{k-1,i}$$

- To uncover the **predicted probability of the event occurring**:

$$P(\widehat{y}_i = 1) = \frac{1}{1 + e^{(-\hat{l}_i)}}$$

- Predicted value from a regular regression:

$$\hat{y}_i = b_0 + b_1X_{1,i} + b_2X_{2,i} + \dots + b_{k-1}X_{k-1,i}$$

- For a logistic regression, you can get predicted logist (not too interesting yet):

$$\hat{\eta}_i = \ln(\widehat{Odds}) = b_0 + b_1X_{1,i} + b_2X_{2,i} + \dots + b_{k-1}X_{k-1,i}$$

- To uncover the **predicted probability of the event occurring**:

$$P(\widehat{y}_i = 1) = \frac{1}{1 + e^{(-\hat{\eta}_i)}}$$

- Predicted value from a regular regression:

$$\hat{y}_i = b_0 + b_1X_{1,i} + b_2X_{2,i} + \dots + b_{k-1}X_{k-1,i}$$

- For a logistic regression, you can get predicted logist (not too interesting yet):

$$\hat{l}_i = \ln(\widehat{Odds}) = b_0 + b_1X_{1,i} + b_2X_{2,i} + \dots + b_{k-1}X_{k-1,i}$$

- To uncover the **predicted probability of the event occurring**:

$$P(\widehat{y_i = 1}) = \frac{1}{1 + e^{(-\hat{l}_i)}}$$

- Predicted value from a regular regression:

$$\hat{y}_i = b_0 + b_1X_{1,i} + b_2X_{2,i} + \dots + b_{k-1}X_{k-1,i}$$

- For a logistic regression, you can get predicted logist (not too interesting yet):

$$\hat{l}_i = \ln(\widehat{Odds}) = b_0 + b_1X_{1,i} + b_2X_{2,i} + \dots + b_{k-1}X_{k-1,i}$$

- To uncover the **predicted probability of the event occurring**:

$$P(\widehat{y}_i = 1) = \frac{1}{1 + e^{(-\hat{l}_i)}}$$

- Predicted value from a regular regression:

$$\hat{y}_i = b_0 + b_1X_{1,i} + b_2X_{2,i} + \dots + b_{k-1}X_{k-1,i}$$

- For a logistic regression, you can get predicted logist (not too interesting yet):

$$\hat{l}_i = \ln(\widehat{Odds}) = b_0 + b_1X_{1,i} + b_2X_{2,i} + \dots + b_{k-1}X_{k-1,i}$$

- To uncover the **predicted probability of the event occurring**:

$$P(\widehat{y}_i = 1) = \frac{1}{1 + e^{(-\hat{l}_i)}}$$

- Predicted value from a regular regression:

$$\hat{y}_i = b_0 + b_1X_{1,i} + b_2X_{2,i} + \dots + b_{k-1}X_{k-1,i}$$

- For a logistic regression, you can get predicted logist (not too interesting yet):

$$\hat{l}_i = \ln(\widehat{Odds}) = b_0 + b_1X_{1,i} + b_2X_{2,i} + \dots + b_{k-1}X_{k-1,i}$$

- To uncover the **predicted probability of the event occurring**:

$$P(\widehat{y_i = 1}) = \frac{1}{1 + e^{(-\hat{l}_i)}}$$

- **Marginal effect for regression:** measure of how much y changes when x increases by 1.
 - Example: How much does public expenditure per capita increase (or decrease) when economic ability increases by one unit?
- **Marginal effect for logit:** measure of how much $P(y_i = 1)$ changes when x increases by 1.
 - How much more (or less) likely will an interview candidate be working here after one year if she/he has a four year college education?

- **Marginal effect for regression:** measure of how much y changes when x increases by 1.
 - Example: How much does public expenditure per capita increase (or decrease) when economic ability increases by one unit?
- **Marginal effect for logit:** measure of how much $P(y_i = 1)$ changes when x increases by 1.
 - How much more (or less) likely will an interview candidate be working here after one year if she/he has a four year college education?

- **Marginal effect for regression:** measure of how much y changes when x increases by 1.
 - Example: How much does public expenditure per capita increase (or decrease) when economic ability increases by one unit?
- **Marginal effect for logit:** measure of how much $P(y_i = 1)$ changes when x increases by 1.
 - How much more (or less) likely will an interview candidate be working here after one year if she/he has a four year college education?

- **Marginal effect for regression:** measure of how much y changes when x increases by 1.
 - Example: How much does public expenditure per capita increase (or decrease) when economic ability increases by one unit?
- **Marginal effect for logit:** measure of how much $P(y_i = 1)$ changes when x increases by 1.
 - How much more (or less) likely will an interview candidate be working here after one year if she/he has a four year college education?

- For a regular regression, for coefficient b_2 :
 - The sign (positive/negative) indicates whether x_2 causes y to increase or decrease.
 - The magnitude tells *how much* y increases when increasing x_2 by 1.
 - Coefficient = Marginal Effect.
- For a logistic regression, the coefficient b_2 :
 - The sign (positive/negative) indicates whether x_2 causes y to increase or decrease.
 - The magnitude is pretty meaningless.
 - Need to do more to figure out marginal effect.

- For a regular regression, for coefficient b_2 :
 - The sign (positive/negative) indicates whether x_2 causes y to increase or decrease.
 - The magnitude tells *how much* y increases when increasing x_2 by 1.
 - Coefficient = Marginal Effect.
- For a logistic regression, the coefficient b_2 :
 - The sign (positive/negative) indicates whether x_2 causes y to increase or decrease.
 - The magnitude is pretty meaningless.
 - Need to do more to figure out marginal effect.

- For a regular regression, for coefficient b_2 :
 - The sign (positive/negative) indicates whether x_2 causes y to increase or decrease.
 - The magnitude tells *how much* y increases when increasing x_2 by 1.
 - Coefficient = Marginal Effect.
- For a logistic regression, the coefficient b_2 :
 - The sign (positive/negative) indicates whether x_2 causes y to increase or decrease.
 - The magnitude is pretty meaningless.
 - Need to do more to figure out marginal effect.

- For a regular regression, for coefficient b_2 :
 - The sign (positive/negative) indicates whether x_2 causes y to increase or decrease.
 - The magnitude tells *how much* y increases when increasing x_2 by 1.
 - Coefficient = Marginal Effect.
- For a logistic regression, the coefficient b_2 :
 - The sign (positive/negative) indicates whether x_2 causes y to increase or decrease.
 - The magnitude is pretty meaningless.
 - Need to do more to figure out marginal effect.

- For a regular regression, for coefficient b_2 :
 - The sign (positive/negative) indicates whether x_2 causes y to increase or decrease.
 - The magnitude tells *how much* y increases when increasing x_2 by 1.
 - Coefficient = Marginal Effect.
- For a logistic regression, the coefficient b_2 :
 - The sign (positive/negative) indicates whether x_2 causes y to increase or decrease.
 - The magnitude is pretty meaningless.
 - Need to do more to figure out marginal effect.

- For a regular regression, for coefficient b_2 :
 - The sign (positive/negative) indicates whether x_2 causes y to increase or decrease.
 - The magnitude tells *how much* y increases when increasing x_2 by 1.
 - Coefficient = Marginal Effect.
- For a logistic regression, the coefficient b_2 :
 - The sign (positive/negative) indicates whether x_2 causes y to increase or decrease.
 - The magnitude is pretty meaningless.
 - Need to do more to figure out marginal effect.

- For a regular regression, for coefficient b_2 :
 - The sign (positive/negative) indicates whether x_2 causes y to increase or decrease.
 - The magnitude tells *how much* y increases when increasing x_2 by 1.
 - Coefficient = Marginal Effect.
- For a logistic regression, the coefficient b_2 :
 - The sign (positive/negative) indicates whether x_2 causes y to increase or decrease.
 - The magnitude is pretty meaningless.
 - Need to do more to figure out marginal effect.

- For a regular regression, for coefficient b_2 :
 - The sign (positive/negative) indicates whether x_2 causes y to increase or decrease.
 - The magnitude tells *how much* y increases when increasing x_2 by 1.
 - Coefficient = Marginal Effect.
- For a logistic regression, the coefficient b_2 :
 - The sign (positive/negative) indicates whether x_2 causes y to increase or decrease.
 - The magnitude is pretty meaningless.
 - Need to do more to figure out marginal effect.