Lesson 05 Multiple LR. Intro to Logistic Regression

Lusine Zilfimian

March 16 (Monday), 2020

Contents

Contents

- Some aspects of Linear Regression
- Quiz (Logistic Regression)

Contents

- Some aspects of Linear Regression
- Quiz (Logistic Regression)
- Intro to Logistic Regression: Materials to read

Contents

- Some aspects of Linear Regression
- Quiz (Logistic Regression)
- Intro to Logistic Regression: Materials to read
- Linear Probability Models

Contents

- Some aspects of Linear Regression
- Quiz (Logistic Regression)
- Intro to Logistic Regression: Materials to read
- Linear Probability Models
- Binominal Logistic Regression

Last Lecture ReCap

- Why we cannot rely on the result of R^2 in Multiple Linear Regression?

Last Lecture ReCap

- Why we cannot rely on the result of R^2 in Multiple Linear Regression?
- Formulate the hypothesis for the significance of the whole model.

Last Lecture ReCap

- Why we cannot rely on the result of R^2 in Multiple Linear Regression?
- Formulate the hypothesis for the significance of the whole model.
- Which are the assumptions of Gauss-Markov theorem?

Last Lecture ReCap

- Why we cannot rely on the result of R^2 in Multiple Linear Regression?
- Formulate the hypothesis for the significance of the whole model.
- Which are the assumptions of Gauss-Markov theorem?
- Interpret the meaning of coefficient of (a) continuous predictor, (b) categorical predictor.

Relaxing the linearity assumption: Non-linearity

1. Non-linear by x

Relaxing the linearity assumption: Non-linearity

1. Non-linear by x

- $y = \beta_0 + \beta_1 x^2 + \varepsilon$

Relaxing the linearity assumption: Non-linearity

1. Non-linear by x

- $y = \beta_0 + \beta_1 x^2 + \varepsilon$
- $y = \beta_0 + \beta_1 x^2 + \beta_2 x^3 + \varepsilon$

Relaxing the linearity assumption: Non-linearity

1. Non-linear by x

- $y = \beta_0 + \beta_1 x^2 + \varepsilon$
- $y = \beta_0 + \beta_1 x^2 + \beta_2 x^3 + \varepsilon$
- $y = \beta_0 + \beta_1 \frac{1}{x} + \varepsilon$

Relaxing the linearity assumption: Non-linearity

1. Non-linear by x
   - $y = \beta_0 + \beta_1 x^2 + \varepsilon$
   - $y = \beta_0 + \beta_1 x^2 + \beta_2 x^3 + \varepsilon$
   - $y = \beta_0 + \beta_1 \frac{1}{x} + \varepsilon$
2. Non-linear by x and β

Relaxing the linearity assumption: Non-linearity

1. Non-linear by x
   - $y = \beta_0 + \beta_1 x^2 + \varepsilon$
   - $y = \beta_0 + \beta_1 x^2 + \beta_2 x^3 + \varepsilon$
   - $y = \beta_0 + \beta_1 \frac{1}{x} + \varepsilon$
2. Non-linear by x and β
   - $y = \beta_0 + x^{\beta_1} + \varepsilon$

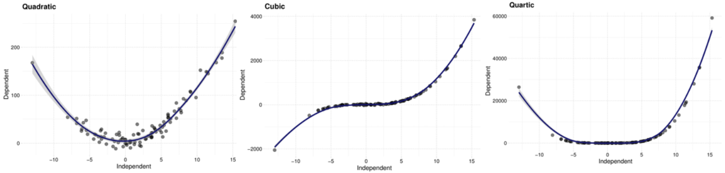
Relaxing the linearity assumption: Non-linearity

1. Non-linear by x
   - $y = \beta_0 + \beta_1 x^2 + \varepsilon$
   - $y = \beta_0 + \beta_1 x^2 + \beta_2 x^3 + \varepsilon$
   - $y = \beta_0 + \beta_1 \frac{1}{x} + \varepsilon$
2. Non-linear by x and β
   - $y = \beta_0 + x^{\beta_1} + \varepsilon$
   - $y = \beta_0 + \beta_1^x + \varepsilon$

Relaxing the linearity assumption: Non-linearity

1. Non-linear by x
   - $y = \beta_0 + \beta_1 x^2 + \varepsilon$
   - $y = \beta_0 + \beta_1 x^2 + \beta_2 x^3 + \varepsilon$
   - $y = \beta_0 + \beta_1 \frac{1}{x} + \varepsilon$
2. Non-linear by x and β
   - $y = \beta_0 + x^{\beta_1} + \varepsilon$
   - $y = \beta_0 + \beta_1^x + \varepsilon$
   - $y = \beta_0 + e^{x\beta_1} + \varepsilon$

Polynomial Regression



- Use non-linear transformations of the predictors, such as $\log x$, $\sqrt{x}, x^2$

Heteroskedasticity

- $var(\varepsilon_i) = \sigma_i^2$

Heteroskedasticity

- $var(\varepsilon_i) = \sigma_i^2$
- Consequences:

Heteroskedasticity

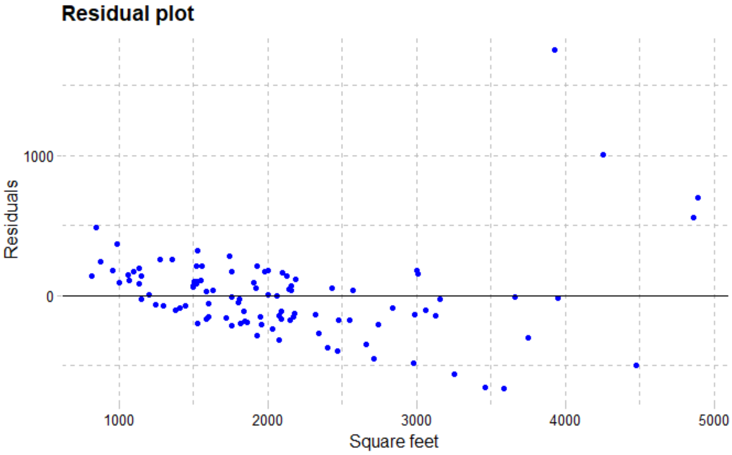
- $var(\varepsilon_i) = \sigma_i^2$
- Consequences:
- Estimated coefficients are unbiased, but not effective.

Heteroskedasticity

- $var(\varepsilon_i) = \sigma_i^2$
- Consequences:
- Estimated coefficients are unbiased, but not effective.
- Standard errors computed for betas are not correct (problem with hypothesis testing, confidence intervals).

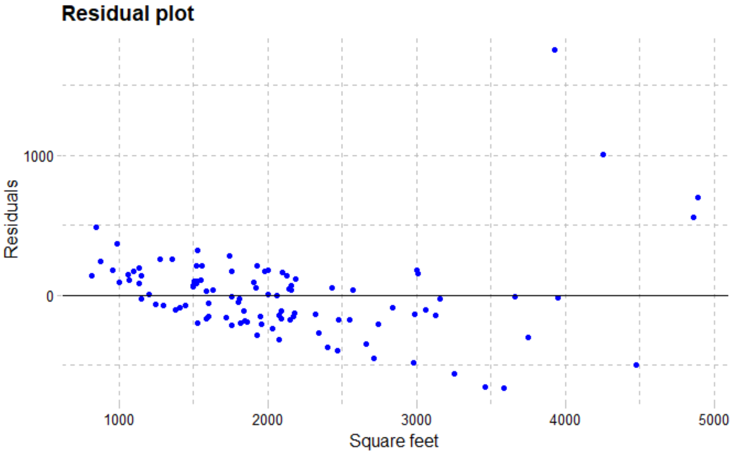
Detecting Heteroskedasticity

- Residual plot



Detecting Heteroskedasticity

- Residual plot



- Tests (Breusch-Pagan, Goldfeld-Quandt, Spearman, etc.)

Model with Heteroskedasticity

- The transformation of the response variable

Autocorrelation

Model with Heteroskedasticity

- The transformation of the response variable
- Weighted least squares

Autocorrelation

Model with Heteroskedasticity

- The transformation of the response variable
- Weighted least squares
- Using heteroscedasticity-consistent (robust) standard errors.

Autocorrelation

Model with Heteroskedasticity

- The transformation of the response variable
- Weighted least squares
- Using heteroscedasticity-consistent (robust) standard errors.

Autocorrelation

- $cov(\varepsilon_i, \varepsilon_j) \neq 0 \ \Rightarrow$

Model with Heteroskedasticity

- The transformation of the response variable
- Weighted least squares
- Using heteroscedasticity-consistent (robust) standard errors.

Autocorrelation

- $cov(\varepsilon_i, \varepsilon_j) \neq 0 \;\Rightarrow$
- The estimated standard errors will tend to underestimate the true standard errors $\Rightarrow$

Model with Heteroskedasticity

- The transformation of the response variable
- Weighted least squares
- Using heteroscedasticity-consistent (robust) standard errors.

Autocorrelation

- $cov(\varepsilon_i, \varepsilon_j) \neq 0 \;\Rightarrow$
- The estimated standard errors will tend to underestimate the true standard errors $\Rightarrow$
- Confidence intervals will be narrower than they should be $\Rightarrow$

Model with Heteroskedasticity

- The transformation of the response variable
- Weighted least squares
- Using heteroscedasticity-consistent (robust) standard errors.

Autocorrelation

- $cov(\varepsilon_i, \varepsilon_j) \neq 0 \;\Rightarrow$
- The estimated standard errors will tend to underestimate the true standard errors $\Rightarrow$
- Confidence intervals will be narrower than they should be $\Rightarrow$
- p-values will be lower than they should be $\Rightarrow$

Model with Heteroskedasticity

- The transformation of the response variable
- Weighted least squares
- Using heteroscedasticity-consistent (robust) standard errors.

Autocorrelation

- $cov(\varepsilon_i, \varepsilon_j) \neq 0 \;\Rightarrow$
- The estimated standard errors will tend to underestimate the true standard errors $\Rightarrow$
- Confidence intervals will be narrower than they should be $\Rightarrow$
- p-values will be lower than they should be $\Rightarrow$
- Erroneously conclude that a parameter is statistically significant

Multicollinearity

It can be difficult to separate out the individual effects of collinear variables on the response variable:

1. Perfect collinearity (Wrong dummy, Different measures, Input mistake)

Multicollinearity

It can be difficult to separate out the individual effects of collinear variables on the response variable:

1. Perfect collinearity (Wrong dummy, Different measures, Input mistake)
    - If $rank(X) < m \Rightarrow (X^TX)^{-1}$ does not exists

Multicollinearity

It can be difficult to separate out the individual effects of collinear variables on the response variable:

1. Perfect collinearity (Wrong dummy, Different measures, Input mistake)
   - If $rank(X) < m \Rightarrow (X^TX)^{-1}$ does not exists
2. Multicollinearity $rank(X) = m$, but there are high correlation

Multicollinearity

It can be difficult to separate out the individual effects of collinear variables on the response variable:

1. Perfect collinearity (Wrong dummy, Different measures, Input mistake)
   - If $rank(X) < m \Rightarrow (X^TX)^{-1}$ does not exists
2. Multicollinearity $rank(X) = m$, but there are high correlation
   - Standard error for $\hat{\beta}_j$ will grow $\Rightarrow\ t = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \downarrow \Rightarrow$

Multicollinearity

It can be difficult to separate out the individual effects of collinear variables on the response variable:

1. Perfect collinearity (Wrong dummy, Different measures, Input mistake)
   - If $rank(X) < m \Rightarrow (X^TX)^{-1}$ does not exists
2. Multicollinearity $rank(X) = m$, but there are high correlation
   - Standard error for $\hat{\beta}_j$ will grow $\Rightarrow \; t = \dfrac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \downarrow \Rightarrow$
   - Fail to reject $H_0 \Rightarrow$ probability of correctly detecting a non-zero coefficient is reduced

Multicollinearity

```
(A <- matrix( c(1, 2, 3, 2, 4, 6, 4, 7, 9, 12, 24, 36),
  nrow = 4, byrow= TRUE))
```

```
##      [,1] [,2] [,3]
## [1,]    1    2    3
## [2,]    2    4    6
## [3,]    4    7    9
## [4,]   12   24   36
```

```
AtA <- t(A) %*% A
qr(A)$rank
```

```
## [1] 2
```

```
qr(A)$rank == qr(AtA)$rank
```

```
## [1] TRUE
```

Multicollinearity

```
solve(AtA)
```

```
## Error in solve.default(AtA): system is computationally sing
```

Detecting collinearity

- Look at the correlation matrix of the predictors.

```
##                       price sqft_living Sqft_with_garden
## price             1.0000000   0.6960595        0.6938874
## sqft_living       0.6960595   1.0000000        0.9972430
## Sqft_with_garden  0.6938874   0.9972430        1.0000000
```

Detecting collinearity

With multicollinearity, small changes in the model or the data can case the erratic change in coefficient estimates and/thus their significance.

Table 1: Multicollinearity

	Dependent variable:	
	price	
	(1)	(2)
Square feet	$\boxed{0.727}$	$\boxed{0.405^{***}}$
	(0.599)	(0.035)
With garder	−0.320	
	(0.594)	
Constant	-247.245^{**}	-275.659^{***}
	(96.426)	(80.419)
Observations	101	101
R^2	0.572	0.570
Adjusted R^2	0.563	0.566
Residual Std. Error	312.706 (df = 98)	311.582 (df = 99)
F Statistic	65.396^{***} (df = 2; 98)	131.445^{***} (df = 1; 99)
Note:		*p<0.1; **p<0.05; ***p<0.01

	Dependent variable:
	sqft_living
Sqft_with_garden	0.990***
	(0.006)
Constant	−80.571***
	(14.005)
Observations	101
R^2	0.996
Adjusted R^2	0.996
Residual Std. Error	52.464 (df = 99)
F Statistic	28,180.940*** (df = 1; 99)
Note:	*p<0.1; **p<0.05; ***p<0.01

Detecting collinearity

- Compute Variance Inflation Factor (VIF)

Detecting collinearity

- Compute Variance Inflation Factor (VIF)
- $VIF(\hat{\beta}_j) = \dfrac{1}{1 - R^2_{X_j|X_{-j}}}$

Detecting collinearity

- Compute Variance Inflation Factor (VIF)
- $VIF(\hat{\beta}_j) = \dfrac{1}{1 - R^2_{X_j|X_{-j}}}$
- $R^2_{X_j|X_{-j}}$ - R^2 from a regression of X_j onto all of the other predictors.

Detecting collinearity

- Compute Variance Inflation Factor (VIF)
- $VIF(\hat{\beta}_j) = \dfrac{1}{1 - R^2_{X_j|X_{-j}}}$
- $R^2_{X_j|X_{-j}}$ - R^2 from a regression of X_j onto all of the other predictors.
- $R^2_{X_j|X_{-j}} \to 1 \Rightarrow$ presence of multicollinearity

Detecting collinearity

- Compute Variance Inflation Factor (VIF)
- $VIF(\hat{\beta}_j) = \dfrac{1}{1 - R^2_{X_j|X_{-j}}}$
- $R^2_{X_j|X_{-j}}$ - R^2 from a regression of X_j onto all of the other predictors.
- $R^2_{X_j|X_{-j}} \to 1 \Rightarrow$ presence of multicollinearity
- $R^2_{X_j|X_{-j}} \to 0 \Rightarrow$ absence of multicollinearity

Solving collinearity

- Drop one of the problematic variables from the regression.

Solving collinearity

- Drop one of the problematic variables from the regression.
- Combine the collinear variables together into a single predictor.

Time for Quiz

Quiz!

Go to **socrative.com** to show your knowledge :)

Logistic Regression

Linear Probability Models

- Simple linear regression: $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \; i = 1, ..., N$

Logistic Regression

Linear Probability Models

- Simple linear regression: $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \; i = 1, ..., N$
- y_i is binary variable, $\mathbb{E}(\varepsilon_i) = 0$

Logistic Regression

Linear Probability Models

- Simple linear regression: $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \; i = 1, ..., N$
- y_i is binary variable, $\mathbb{E}(\varepsilon_i) = 0$
- $\mathbb{E}(y_i) = 1 * \mathbb{P}(y_i = 1) + 0 * (1 - \mathbb{P}(y_i = 1)) = \mathbb{P}(y_i = 1) =$

Logistic Regression

Linear Probability Models

- Simple linear regression: $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \; i = 1, ..., N$
- y_i is binary variable, $\mathbb{E}(\varepsilon_i) = 0$
- $\mathbb{E}(y_i) = 1 * \mathbb{P}(y_i = 1) + 0 * (1 - \mathbb{P}(y_i = 1)) = \mathbb{P}(y_i = 1) =$
- $= \beta_0 + \beta_1 x_i$

Logistic Regression

Linear Probability Models

- Simple linear regression: $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \; i = 1, ..., N$
- y_i is binary variable, $\mathbb{E}(\varepsilon_i) = 0$
- $\mathbb{E}(y_i) = 1 * \mathbb{P}(y_i = 1) + 0 * (1 - \mathbb{P}(y_i = 1)) = \mathbb{P}(y_i = 1) =$
- $= \beta_0 + \beta_1 x_i$
- Linear Probability Model: $\mathbb{P}(y_i = 1) = \beta_0 + \beta_1 x_i$

Why Not Linear Regression and OLS?

1. Problem with alternative coding ($\{0, 1, 2\}, \{1, 5, 3\}$)

$$\varepsilon_i = \begin{cases} 1 - \beta_0 - \beta_1 x_i \\ -\beta_0 - \beta_1 x_i \end{cases}$$

Why Not Linear Regression and OLS?

1. Problem with alternative coding $(\{0,1,2\},\{1,5,3\})$
2. ε_i is not normally distributed and is not continuous:

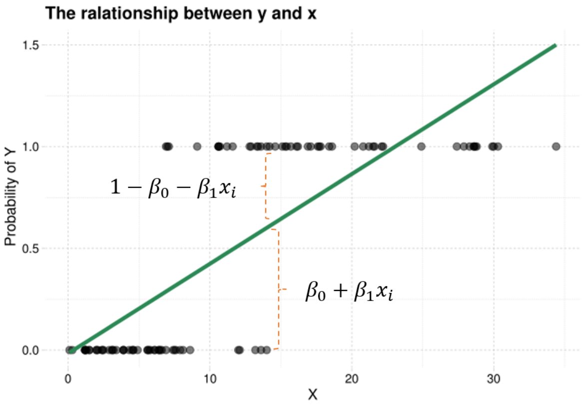
$$\varepsilon_i = \begin{cases} 1 - \beta_0 - \beta_1 x_i \\ -\beta_0 - \beta_1 x_i \end{cases}$$

Why Not Linear Regression and OLS?

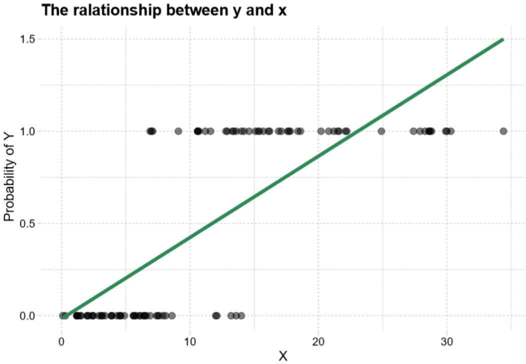
1. Problem with alternative coding $(\{0,1,2\},\{1,5,3\})$
2. ε_i is not normally distributed and is not continuous:

$$\varepsilon_i = \begin{cases} 1-\beta_0-\beta_1 x_i \\ -\beta_0-\beta_1 x_i \end{cases}$$

3. Heteroskedasticity: $var(\varepsilon_i) = (\beta_0+\beta_1 x_i)(1-\beta_0-\beta_1 x_i)$


The ralationship between y and x
Probability of Y
1.5
1.0
0.5
0.0
0
10
20
30
X
$1 - \beta_0 - \beta_1 x_i$
$\beta_0 + \beta_1 x_i$

4. Some of our estimates might be outside the [0,1] interval



Logistic Regression Models

Relationship between **one categorical dependent** variable and one or more **(any) explanatory** variables.

- **Binomial** - Binary dependent variable

Logistic Regression Models

Relationship between **one categorical dependent** variable and one or more **(any) explanatory** variables.

- **Binomial** - Binary dependent variable
- **Multinomial** - Categorical dependent variable with three or more categories

Logistic Regression Models

Relationship between **one categorical dependent** variable and one or more **(any) explanatory** variables.

- **Binomial** - Binary dependent variable
- **Multinomial** - Categorical dependent variable with three or more categories
- Used for prediction (**classification**) and estimation.

Binomial logistic regression

Model description

- Suppose $\exists\ y_i^*$ such that

$$\begin{cases} y_i^* \geq 0, y_i = 1 \\ y_i^* < 0, y_i = 0 \end{cases}$$

Binomial logistic regression

Model description

- Suppose $\exists\ y_i^*$ such that

$$\begin{cases} y_i^* \geq 0, y_i = 1 \\ y_i^* < 0, y_i = 0 \end{cases}$$

- $y_i^* = \beta_0 + \beta_1 x_i + \varepsilon_i$, and $\varepsilon_i \sim F(\cdot)$

Binomial logistic regression

Model description

- Suppose $\exists\ y_i^*$ such that

$$\begin{cases} y_i^* \geq 0, y_i = 1 \\ y_i^* < 0, y_i = 0 \end{cases}$$

- $y_i^* = \beta_0 + \beta_1 x_i + \varepsilon_i$, and $\varepsilon_i \sim F(\cdot)$
- $\mathbb{P}(y_i = 1) = \mathbb{P}(\beta_0 + \beta_1 x_i + \varepsilon_i \geq 0) = \mathbb{P}(\varepsilon_i \leq \beta_0 + \beta_1 x_i) = F(\beta_0 + \beta_1 x_i)$

Binomial logistic regression

Model description

- Suppose $\exists\ y_i^*$ such that

$$\begin{cases} y_i^* \geq 0, y_i = 1 \\ y_i^* < 0, y_i = 0 \end{cases}$$

- $y_i^* = \beta_0 + \beta_1 x_i + \varepsilon_i$, and $\varepsilon_i \sim F(\cdot)$
- $\mathbb{P}(y_i = 1) = \mathbb{P}(\beta_0 + \beta_1 x_i + \varepsilon_i \geq 0) = \mathbb{P}(\varepsilon_i \leq \beta_0 + \beta_1 x_i) = F(\beta_0 + \beta_1 x_i)$
- Often as function F logit distribution function or normal distribution function is used.

Logit distribution function: Properties

- The problem with prediction is solved:

Logit distribution function: Properties

- The problem with prediction is solved:
- $$F(u) = \frac{e^u}{1 + e^u}$$

Logit distribution function: Properties

- The problem with prediction is solved:
- $$F(u) = \frac{e^u}{1 + e^u}$$
- $u \to +\infty, \; F(u) = 1$

Logit distribution function: Properties

- The problem with prediction is solved:
- $$F(u) = \frac{e^u}{1+e^u}$$
- $u \to +\infty,\ F(u) = 1$
- $u \to -\infty,\ F(u) = 0$

Logit distribution function: Properties

- The problem with prediction is solved:
- $$F(u) = \frac{e^u}{1 + e^u}$$
- $u \to +\infty, \ F(u) = 1$
- $u \to -\infty, \ F(u) = 0$
- $\mathbb{P}(y_i = 1) = F(\beta_0 + \beta_1 x_i) = \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}}$

Binary simple case

- Binary response using one predictor

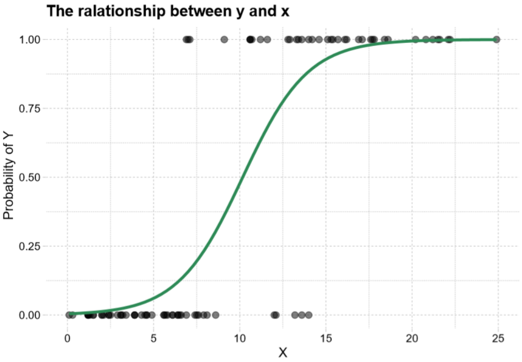
Binary simple case

- Binary response using one predictor
- When $p = 2$, there is only a single linear function to estimate.

Binary simple case

- Binary response using one predictor
- When p $= 2$, there is only a single linear function to estimate.
- The probability: $\mathbb{P}(y_i = 1) = \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}}$

S-shaped curve



Terminology

- Odds: $\frac{\mathbb{P}(y=1)}{1-\mathbb{P}(y=1)} = e^{\beta_0+\beta_1 x} \in [0; +\infty)$

Terminology

- Odds: $\frac{\mathbb{P}(y=1)}{1-\mathbb{P}(y=1)} = e^{\beta_0+\beta_1 x} \in [0; +\infty)$
- Log-odds or logit: $ln\frac{\mathbb{P}(y=1)}{1-\mathbb{P}(y=1)} = \beta_0 + \beta_1 x \in (-\infty; +\infty)$

Terminology

- Odds: $\frac{\mathbb{P}(y=1)}{1-\mathbb{P}(y=1)} = e^{\beta_0+\beta_1 x} \in [0; +\infty)$
- Log-odds or logit: $ln \frac{\mathbb{P}(y=1)}{1-\mathbb{P}(y=1)} = \beta_0 + \beta_1 x \in (-\infty; +\infty)$
- Odds ratio: $\frac{\frac{\mathbb{P}(y=1|x=1)}{1-\mathbb{P}(y=1|x=1)}}{\frac{\mathbb{P}(y=1|x=0)}{1-\mathbb{P}(y=1|x=0)}}$

Estimating the Coefficients

- The most common approach – maximum likelihood: choose that value of parameter, under which it is most likely to get the (your) data.

Estimating the Coefficients

- The most common approach – maximum likelihood: choose that value of parameter, under which it is most likely to get the (your) data.
- In our case, find/choose **estimates** for coefficients such that the **predicted probability** of Y corresponds as closely as possible to Y.

Estimating the Coefficients

- The most common approach – maximum likelihood: choose that value of parameter, under which it is most likely to get the (your) data.
- In our case, find/choose **estimates** for coefficients such that the **predicted probability** of Y corresponds as closely as possible to Y.
- Likelihood Function is the joint PD(M)F, considered as a function of the parameter

Estimating the Coefficients

- The most common approach – maximum likelihood: choose that value of parameter, under which it is most likely to get the (your) data.
- In our case, find/choose **estimates** for coefficients such that the **predicted probability** of Y corresponds as closely as possible to Y.
- Likelihood Function is the joint PD(M)F, considered as a function of the parameter
- Likelihood function:

Estimating the Coefficients

- The most common approach – maximum likelihood: choose that value of parameter, under which it is most likely to get the (your) data.
- In our case, find/choose **estimates** for coefficients such that the **predicted probability** of Y corresponds as closely as possible to Y.
- Likelihood Function is the joint PD(M)F, considered as a function of the parameter
- Likelihood function:
- $$L(\beta_0, \beta_1) = \prod_{y_i=1} F(\beta_0 + \beta_1 x_i) \prod_{y_i=0} (1 - F(\beta_0 + \beta_1 x_i)) \to max$$

Estimating the Coefficients

- The most common approach – maximum likelihood: choose that value of parameter, under which it is most likely to get the (your) data.
- In our case, find/choose **estimates** for coefficients such that the **predicted probability** of Y corresponds as closely as possible to Y.
- Likelihood Function is the joint PD(M)F, considered as a function of the parameter
- Likelihood function:
- $$L(\beta_0, \beta_1) = \prod_{y_i=1} F(\beta_0 + \beta_1 x_i) \prod_{y_i=0} (1 - F(\beta_0 + \beta_1 x_i)) \to max$$
- Likelihood is not a Probability - it can be larger than 1. It is not a PDF either, it is a function of the parameter.

Estimating the Coefficients

- The Maximum Likelihood Method suggests to find a point that makes our Likelihood Maximal:

Estimating the Coefficients

- The Maximum Likelihood Method suggests to find a point that makes our Likelihood Maximal:
- Log-Likelihood function:

Estimating the Coefficients

- The Maximum Likelihood Method suggests to find a point that makes our Likelihood Maximal:
- Log-Likelihood function:
- $l(\beta_0, \beta_1) = \sum_{i=1}^{N}(y_i ln\mathbb{P}(x_i) + (1 - y_i)ln(1 - \mathbb{P}(x_i))$

Estimating the Coefficients

- The Maximum Likelihood Method suggests to find a point that makes our Likelihood Maximal:
- Log-Likelihood function:
- $l(\beta_0, \beta_1) = \sum_{i=1}^{N}(y_i ln\mathbb{P}(x_i) + (1 - y_i)ln(1 - \mathbb{P}(x_i))$
- $\sum_{i=1}^{N}(y_i(\beta_0 + \beta_1 x_i) + log(1 + e^{\beta_0 + \beta_1 x})) \to max$

Estimating the Coefficients

- The Maximum Likelihood Method suggests to find a point that makes our Likelihood Maximal:
- Log-Likelihood function:
- $l(\beta_0, \beta_1) = \sum_{i=1}^{N}(y_i ln\mathbb{P}(x_i) + (1 - y_i)ln(1 - \mathbb{P}(x_i))$
- $\sum_{i=1}^{N}(y_i(\beta_0 + \beta_1 x_i) + log(1 + e^{\beta_0 + \beta_1 x})) \to max$
- The points of maximum of $L(\beta_0, \beta_1)$ and $lnL(\beta_0, \beta_1)$ coincide:

Estimating the Coefficients

- The Maximum Likelihood Method suggests to find a point that makes our Likelihood Maximal:
- Log-Likelihood function:
- $l(\beta_0, \beta_1) = \sum_{i=1}^{N}(y_i ln\mathbb{P}(x_i) + (1 - y_i) ln(1 - \mathbb{P}(x_i))$
- $\sum_{i=1}^{N}(y_i(\beta_0 + \beta_1 x_i) + log(1 + e^{\beta_0 + \beta_1 x})) \to max$
- The points of maximum of $L(\beta_0, \beta_1)$ and $lnL(\beta_0, \beta_1)$ coincide:
- $\underset{\beta}{argmax} L(\beta_0, \beta_1) = \underset{\beta}{argmax} lnL(\beta_0, \beta_1)$

Estimating the Coefficients

- The Maximum Likelihood Method suggests to find a point that makes our Likelihood Maximal:
- Log-Likelihood function:
- $l(\beta_0, \beta_1) = \sum_{i=1}^{N}(y_i ln\mathbb{P}(x_i) + (1 - y_i)ln(1 - \mathbb{P}(x_i))$
- $\sum_{i=1}^{N}(y_i(\beta_0 + \beta_1 x_i) + log(1 + e^{\beta_0 + \beta_1 x})) \rightarrow max$
- The points of maximum of $L(\beta_0, \beta_1)$ and $lnL(\beta_0, \beta_1)$ coincide:
- $\underset{\beta}{argmax} L(\beta_0, \beta_1) = \underset{\beta}{argmax} lnL(\beta_0, \beta_1)$
- The estimates for coefficients are calculated using iterative procedure.