DSS
Decision Sciences & Systems

Technische Universität München

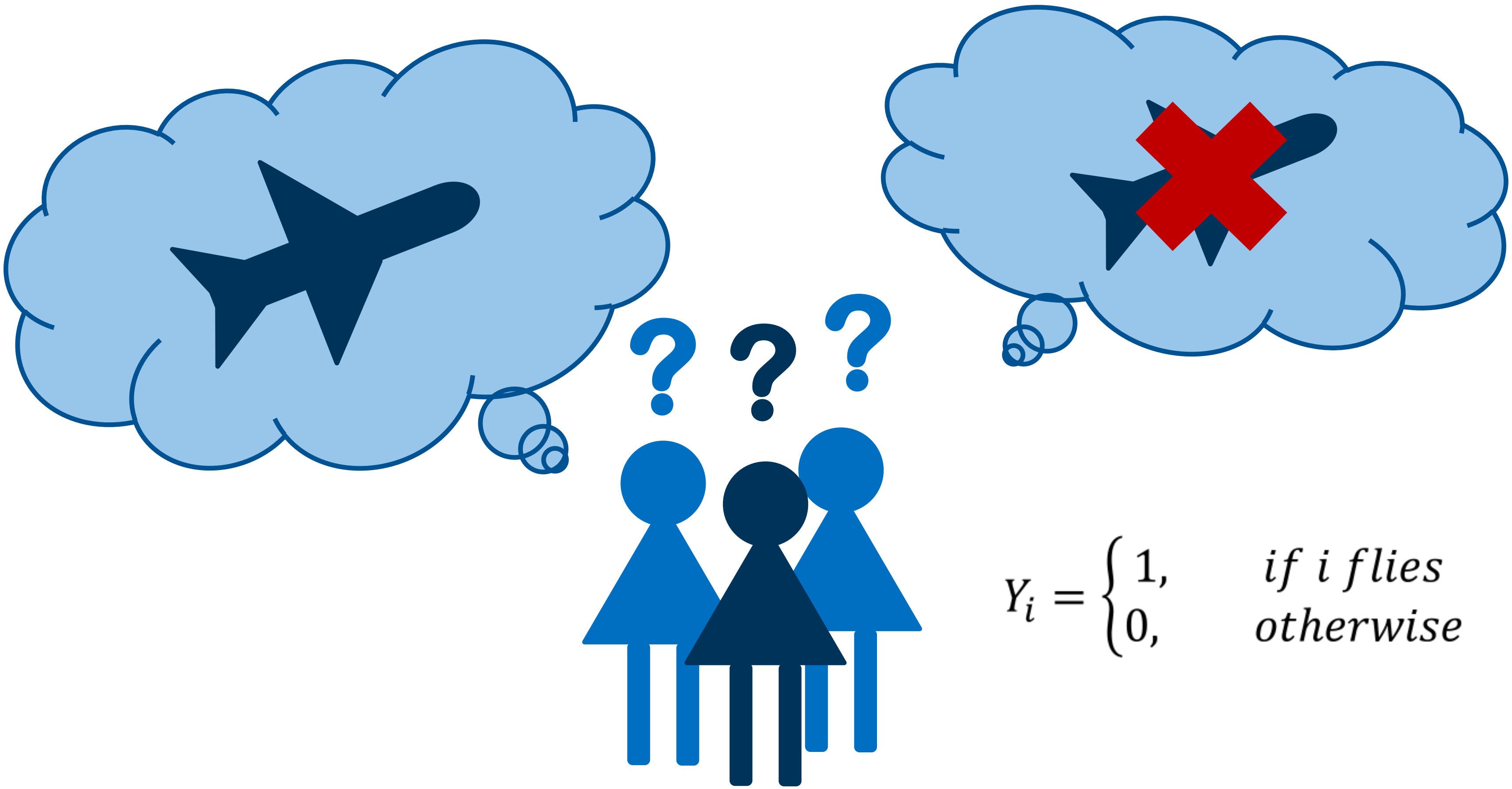
Tutorial Business Analytics

Tutorial 4: Generalized Linear Models

Decision Sciences & Systems (DSS)
Department of Informatics
TU München

Tutorial Business Analytics

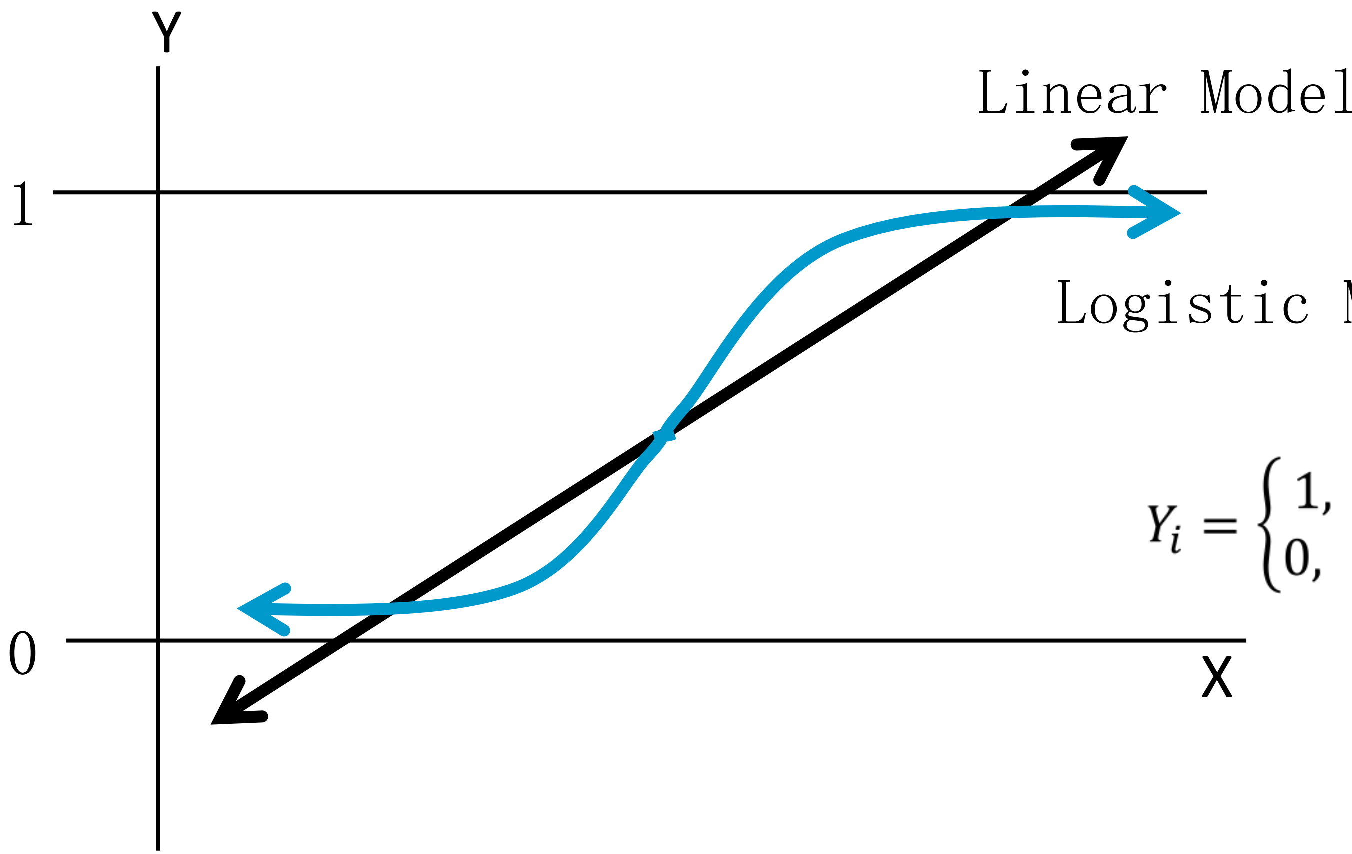
Generalized Linear Models – Motivation



$$Y_i = \begin{cases} 1, & if\ i\ flies \\ 0, & otherwise \end{cases}$$

Tutorial Business Analytics

Generalized Linear Models – Motivation



$$Y_i = \begin{cases} 1, & \text{if } i \text{ flies} \\ 0, & \text{otherwise} \end{cases}$$

Tutorial Business Analytics

Generalized Linear Models

- GLMs are a general class of linear models
- Consist of three components:

- **Random:** Identifies dependent variable μ and probability distribution
- **Systematic:** Identifies the set of explanatory variables $(X_1, \dots, X_k)$
- **Link function:** Identifies function of μ that is linear

$$g(\mu) = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k$$

Example: Linear regression uses identity link ($g(\mu) = \mu$)

Question: Which link function could be useful for a binary dependent variable?

Tutorial Business Analytics

From Logistic Function to Logit

Logistic Function: $p(x_i) = \frac{e^{x_i{}'\beta}}{1+e^{x_i{}'\beta}}$

transform …

Logit: $\ln(\frac{p(x_i)}{1-p(x_i)}) = x_i{}'\beta$

$\Leftrightarrow$ $\frac{p(x_i)}{1-p(x_i)} = e^{x_i{}'\beta}$ **odds**

Logistic Regression: $\ln(\frac{p(x_i)}{1-p(x_i)}) = x_i{}'\beta + \varepsilon_i$

Tutorial Business Analytics

Interpreting the coefficient of logistic regression

$x_{ij} \in x_i$: $\ln(\frac{p(x_i)}{1-p(x_i)}) = x_i{}'\beta$

$(x_{ij}+1) \in \tilde{x}_i$: $\ln(\frac{p(\tilde{x}_i)}{1-p(\tilde{x}_i)}) = \tilde{x}_i{}'\beta$

$$\ln(\frac{p(\tilde{x}_i)}{1-p(\tilde{x}_i)}) - \ln(\frac{p(x_i)}{1-p(x_i)}) = \tilde{x}_i{}'\beta - x_i{}'\beta = \beta_j$$

$$\Leftrightarrow \qquad \beta_j = \ln(\frac{\frac{p(\tilde{x}_i)}{1-p(\tilde{x}_i)}}{\frac{p(x)}{1-p(x)}})$$

$$\Leftrightarrow \qquad e^{\beta_j} = \frac{\frac{p(\tilde{x}_i)}{1-p(\tilde{x}_i)}}{\frac{p(x_i)}{1-p(x_i)}}$$

odds ratio

Tutorial Business Analytics

Summary: Interpreting the coefficient of logistic regression

Effect of change in x_{ij}: on log-odds (A), odds (B) and probability (C)

$$\Delta x_{ij} = 1 > 0$$

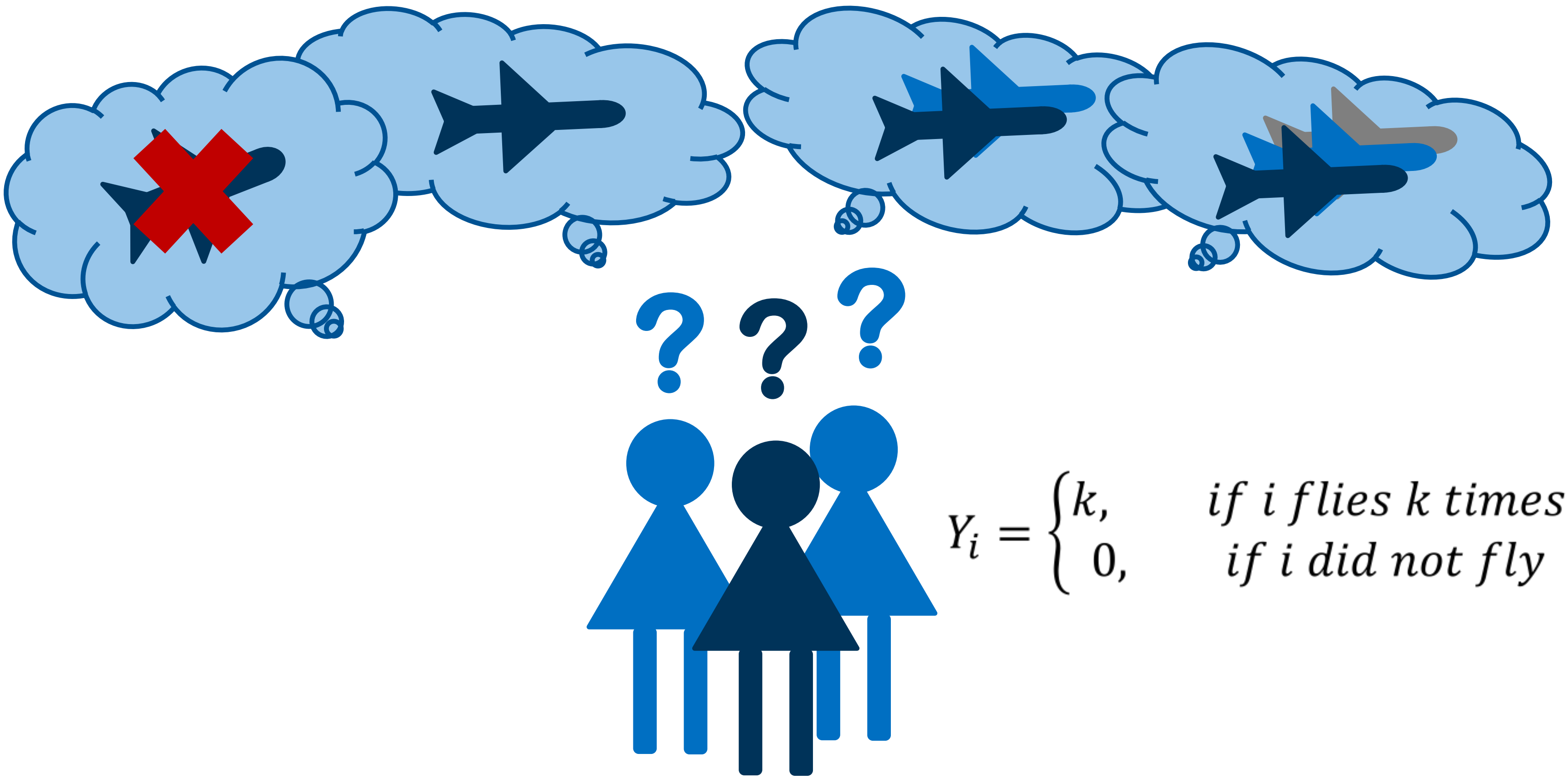
$$\Rightarrow \quad \Delta\, ln(\frac{p(x_i)}{1-p(x_i)}) = \ln(\frac{p(\tilde{x}_i)}{1-p(\tilde{x}_i)}) - \ln(\frac{p(x_i)}{1-p(x_i)}) = \beta_j \quad \text{(A)}$$

$$\Leftrightarrow \quad e^{\beta_j} = \frac{\frac{p(\tilde{x}_i)}{1-p(\tilde{x}_i)}}{\frac{p(x_i)}{1-p(x_i)}} \quad \text{(B), (C)}$$

$\boldsymbol{\beta_j}$	$\boldsymbol{ln(\frac{p}{1-p})}$ (A)	$\boldsymbol{\frac{p}{1-p}}$ (B)	$\boldsymbol{p}$ (C)
$\beta_j > 0$	increases by β_j	increases by a factor of e^{β_j}	Magnitude of increase unknown
$\beta_j < 0$	decreases by β_j	decreases by a factor of e^{β_j}	Magnitude of decrease unknown

Tutorial Business Analytics

Poisson Regression – Motivation



$$Y_i = \begin{cases} k, & \text{if } i \text{ flies } k \text{ times} \\ 0, & \text{if } i \text{ did not fly} \end{cases}$$

Tutorial Business Analytics

From Incidence Rate to Link Function

Incidence Rate: $\mu(x) = e^{x_i{}'\beta}$

transform ...

Link Function: $\ln\big(\mu(x)\big) = x_i{}'\beta$

Poisson Regression: $\ln\big(\mu(x)\big) = x_i{}'\beta + \varepsilon_i$

Tutorial Business Analytics

Interpreting the coefficient of poisson regression

$x_{ij} \in x_i$: $\qquad \ln(\mu(x_i)) = x_i{}'\beta$

$(x_{ij} + 1) \in \tilde{x}_i$: $\qquad \ln(\mu(\tilde{x}_i)) = \tilde{x}_i{}'\beta$

$$\ln(\mu(\tilde{x}_i)) - \ln(\mu(x_i)) = \tilde{x}_i{}'\beta - x_i{}'\beta = \beta_j$$

$$\Leftrightarrow \qquad \beta_j = \ln\left(\frac{\mu(\tilde{x}_i)}{\mu(x_i)}\right)$$

$$\Leftrightarrow \qquad e^{\beta_j} = \frac{\mu(\tilde{x}_i)}{\mu(x_i)}$$

incidence rate ratio

Tutorial Business Analytics

Summary: Interpreting the coefficient of poisson regression

Effect of change in x_{ij}: on log-incidence rate (A), incidence rate (B)

$$\Delta x_{ij} = 1 > 0$$

$$\Rightarrow \quad \Delta \ln\big(\mu(x_i)\big) = \ln\big(\mu(\tilde{x}_i)\big) - \ln\big(\mu(x_i)\big) = \beta_j \qquad \text{(A)}$$

$$\Leftrightarrow \quad e^{\beta_j} = \frac{\mu(\tilde{x}_i)}{\mu(x_i)} \qquad \text{(B)}$$

$\boldsymbol{\beta_j}$	$\boldsymbol{ln\big(\mu(x_i)\big)}$ **(A)**	$\boldsymbol{\mu(x_i)}$ **(B)**
$\beta_j > 0$	increases by β_j	**increases by a factor of** e^{β_j}
$\beta_j < 0$	decreases by β_j	**decreases by a factor of** e^{β_j}