

Az általánosított lineáris modell (GLM)

Ferenci Tamás
tamas.ferenci@medstat.hu

Utoljára frissítve: 2023. május 12.

Tartalom

1 Az általánosított lineáris modell (GLM)

Tartalom

1 Az általánosított lineáris modell (GLM)

A lineáris és a logisztikus regresszió közös keretben

- Vegyük észre a hasonlóságokat!

- 1 Van valamilyen eredményváltozó-eloszlás

- Lineárisnál normális, logisztikusnál Bernoulli

- 2 A feltételes várhatóérték valamilyen transzformáltját modellezzük:

$$g \left[\mathbb{E} (Y|X) \right] = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k$$

- Lineárisnál az identitás, logisztikusnál a logit (vagy logitálisan fordított logit) link-funkció

- 3 Elvileg valamit mondani kellhet a varianciáról is

- Lineárisnál az állandó, logisztikusnál a variancia függ a várható értéktől (vagy a valószínűségektől, az eloszlásunktól)

A lineáris és a logisztikus regresszió közös keretben

- Vegyük észre a hasonlóságokat!
 - ① Van valamilyen eredményváltozó-eloszlás
 - Lineárisnál normális, logisztikusnál Bernoulli
 - ② A feltételes várhatóérték valamilyen transzformáltját modellezzük:

$$g \left[\mathbb{E} (Y|X) \right] = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k$$

Lineáris regresszió: az eredményváltozó eloszlása a normális eloszlás, a transzformáció az identitás

- ③ Elvileg valamit mondani kellhet a varianciáról is

Lineáris regresszió: az eredményváltozó eloszlása a normális eloszlás, a transzformáció az identitás, a varianciája konstans (a GLM keretében ez az a varianciák)

A lineáris és a logisztikus regresszió közös keretben

- Vegyük észre a hasonlóságokat!
 - ① Van valamilyen eredményváltozó-eloszlás
 - Lineárisnál normális, logisztikusnál Bernoulli
 - ② A feltételes várhatóérték valamilyen transzformáltját modellezzük:

$$g \left[\mathbb{E} (Y|X) \right] = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k$$

- Lineárisnál az identitás, logisztikusnál a korábban látott f (pontosabban szöveve annak az inverze)
- ③ Elvileg valamit mondani kellhet a varianciáról is
 - Lineárisnál az eredményváltozó normális eloszlású, logisztikusnál Bernoulli eloszlású, azaz a varianciát mindig meg kell adnunk a modellben. (Egyébként nem lehet az eloszlásunk)

A lineáris és a logisztikus regresszió közös keretben

- Vegyük észre a hasonlóságokat!
 - 1 Van valamilyen eredményváltozó-eloszlás
 - Lineárisnál normális, logisztikusnál Bernoulli
 - 2 A feltételes várhatóérték valamilyen transzformáltját modellezzük:
$$g \left[\mathbb{E} (Y|X) \right] = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k$$
 - Lineárisnál az identitás, logisztikusnál a korábban látott f (pontosabban szólva annak az inverze)
 - 3 Elvileg valamit mondani kellhet a varianciáról is

A lineáris és a logisztikus regresszió közös keretben

- Vegyük észre a hasonlóságokat!
 - ① Van valamilyen eredményváltozó-eloszlás
 - Lineárisnál normális, logisztikusnál Bernoulli
 - ② A feltételes várhatóérték valamilyen transzformáltját modellezzük:

$$g \left[\mathbb{E} (Y|X) \right] = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k$$
 - Lineárisnál az identitás, logisztikusnál a korábban látott f (pontosabban szólva annak az inverze)
 - ③ Elvileg valamit mondani kellhet a varianciáról is

Lineárisnál azt, hogy $\sigma^2 = \sigma_0^2$, logisztikusnál megspóroltuk ezt, mert a várható érték meghatározta a szórás is (egy paramétere volt az eloszlásnak)

A lineáris és a logisztikus regresszió közös keretben

- Vegyük észre a hasonlóságokat!
 - ① Van valamilyen eredményváltozó-eloszlás
 - Lineárisnál normális, logisztikusnál Bernoulli
 - ② A feltételes várhatóérték valamilyen transzformáltját modellezzük:

$$g \left[\mathbb{E} (Y|X) \right] = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k$$
 - Lineárisnál az identitás, logisztikusnál a korábban látott f (pontosabban szólva annak az inverze)
 - ③ Elvileg valamit mondani kellhet a varianciáról is
 - Lineárisnál azt, hogy $\sigma_i^2 = \sigma_0^2$, logisztikusnál megspóroltuk ezt, mert a várható értéke meghatározta a szórást is (egy paramétere volt az eloszlásnak)

A lineáris és a logisztikus regresszió közös keretben

- Vegyük észre a hasonlóságokat!
 - ① Van valamilyen eredményváltozó-eloszlás
 - Lineárisnál normális, logisztikusnál Bernoulli
 - ② A feltételes várhatóérték valamilyen transzformáltját modellezzük:

$$g \left[\mathbb{E} (Y|X) \right] = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k$$
 - Lineárisnál az identitás, logisztikusnál a korábban látott f (pontosabban szólva annak az inverze)
 - ③ Elvileg valamit mondani kellhet a varianciáról is
 - Lineárisnál azt, hogy $\sigma_i^2 = \sigma_0^2$, logisztikusnál megspóroltuk ezt, mert a várható értéke meghatározta a szórás is (egy paramétere volt az eloszlásnak)

Az általánosított lineáris modell (GLM)

- A fenti komponensek határozzák meg az ún. általánosított lineáris modellt (generalized linear modell, GLM)
- Az eredményváltozó eloszlása legyen exponenciális eloszláscsaládból származó
- A g függvény neve: link függvény
- Becslés maximum likelihood-dal
- A lineáris és logisztikus regresszió mind speciális esete ennek (alkalmasan választott eredményváltozó eloszlással, link függvénnyel és szórás-függvénnyel)
- Sok minden más is ide tartozik, lássunk még egy példát

Az általánosított lineáris modell (GLM)

- A fenti komponensek határozzák meg az ún. általánosított lineáris modellt (generalized linear modell, GLM)
- Az eredményváltozó eloszlása legyen exponenciális eloszláscsaládból származó
 - A g függvény neve: link függvény
 - Becslés maximum likelihood-dal
 - A lineáris és logisztikus regresszió mind speciális esete ennek (alkalmasan választott eredményváltozó eloszlással, link függvénnyel és szórás-függvénnyel)
 - Sok minden más is ide tartozik, lássunk még egy példát

Az általánosított lineáris modell (GLM)

- A fenti komponensek határozzák meg az ún. általánosított lineáris modellt (generalized linear modell, GLM)
- Az eredményváltozó eloszlása legyen exponenciális eloszláscsaládból származó
- A g függvény neve: link függvény
 - Becslés maximum likelihood-dal
 - A lineáris és logisztikus regresszió mind speciális esete ennek (alkalmasan választott eredményváltozó eloszlással, link függvénnyel és szórás-függvénnyel)
 - Sok minden más is ide tartozik, lássunk még egy példát

Az általánosított lineáris modell (GLM)

- A fenti komponensek határozzák meg az ún. általánosított lineáris modellt (generalized linear modell, GLM)
- Az eredményváltozó eloszlása legyen exponenciális eloszláscsaládból származó
- A g függvény neve: link függvény
- Becslés maximum likelihood-dal
- A lineáris és logisztikus regresszió mind speciális esete ennek (alkalmasan választott eredményváltozó eloszlással, link függvénnyel és szórás-függvénnyel)
- Sok minden más is ide tartozik, lássunk még egy példát

Az általánosított lineáris modell (GLM)

- A fenti komponensek határozzák meg az ún. általánosított lineáris modellt (generalized linear modell, GLM)
- Az eredményváltozó eloszlása legyen exponenciális eloszláscsaládból származó
- A g függvény neve: link függvény
- Becslés maximum likelihood-dal
- A lineáris és logisztikus regresszió mind speciális esete ennek (alkalmasan választott eredményváltozó eloszlással, link függvénnyel és szórás-függvénnyel)
- Sok minden más is ide tartozik, lássunk még egy példát

Az általánosított lineáris modell (GLM)

- A fenti komponensek határozzák meg az ún. általánosított lineáris modellt (generalized linear modell, GLM)
- Az eredményváltozó eloszlása legyen exponenciális eloszláscsaládból származó
- A g függvény neve: link függvény
- Becslés maximum likelihood-dal
- A lineáris és logisztikus regresszió mind speciális esete ennek (alkalmasan választott eredményváltozó eloszlással, link függvénnyel és szórás-függvénnyel)
- Sok minden más is ide tartozik, lássunk még egy példát

Poisson regresszió

- Mi van, ha az eredményváltozó valamilyen darabszám, események száma jellegű változó (count data)?
- Ilyenekre tipikusan feltételezett eloszlás első közelítésben: Poisson-eloszlás
- Ez exponenciális családbeli
- Várható értéke itt is épp a paramétere
- Tipikus link függvény választás: a log
- Összerakva mindezeket a modellünk:

$$Y \sim \text{Poi}(\lambda)$$
$$\log \left[\mathbb{E}(Y|\underline{X}) \right] = \log \lambda = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k$$

Poisson regresszió

- Mi van, ha az eredményváltozó valamilyen darabszám, események száma jellegű változó (count data)?
- Ilyenekre tipikusan feltételezett eloszlás első közelítésben: Poisson-eloszlás
- Ez exponenciális családbeli
- Várható értéke itt is épp a paramétere
- Tipikus link függvény választás: a log
- Összerakva mindezeket a modellünk:

$$Y \sim \text{Poi}(\lambda)$$
$$\log \left[\mathbb{E}(Y|\underline{X}) \right] = \log \lambda = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k$$

Poisson regresszió

- Mi van, ha az eredményváltozó valamilyen darabszám, események száma jellegű változó (count data)?
- Ilyenekre tipikusan feltételezett eloszlás első közelítésben: Poisson-eloszlás
- Ez exponenciális családbeli
 - Várható értéke itt is épp a paramétere
 - Tipikus link függvény választás: a log
 - Összerakva mindezeket a modellünk:

$$Y \sim \text{Poi}(\lambda)$$
$$\log \left[\mathbb{E}(Y|\underline{X}) \right] = \log \lambda = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k$$

Poisson regresszió

- Mi van, ha az eredményváltozó valamilyen darabszám, események száma jellegű változó (count data)?
- Ilyenekre tipikusan feltételezett eloszlás első közelítésben: Poisson-eloszlás
- Ez exponenciális családbeli
- Várható értéke itt is épp a paramétere
- Tipikus link függvény választás: a log
- Összerakva mindezeket a modellünk:

$$Y \sim \text{Poi}(\lambda)$$
$$\log \left[\mathbb{E}(Y|\underline{X}) \right] = \log \lambda = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k$$

Poisson regresszió

- Mi van, ha az eredményváltozó valamilyen darabszám, események száma jellegű változó (count data)?
- Ilyenekre tipikusan feltételezett eloszlás első közelítésben: Poisson-eloszlás
- Ez exponenciális családbeli
- Várható értéke itt is épp a paramétere
- Tipikus link függvény választás: a log
- Összerakva mindezeket a modellünk:

$$Y \sim \text{Poi}(\lambda)$$
$$\log \left[\mathbb{E}(Y|\underline{X}) \right] = \log \lambda = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k$$

Poisson regresszió

- Mi van, ha az eredményváltozó valamilyen darabszám, események száma jellegű változó (count data)?
- Ilyenekre tipikusan feltételezett eloszlás első közelítésben: Poisson-eloszlás
- Ez exponenciális családbeli
- Várható értéke itt is épp a paramétere
- Tipikus link függvény választás: a log
- Összerakva mindezeket a modellünk:

$$Y \sim \text{Poi}(\lambda)$$
$$\log \left[\mathbb{E}(Y|\underline{X}) \right] = \log \lambda = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k$$