

**AKM 1  
5EN306**

*Martin Janíčko*

Logit a probit

Funkční zápis

Vhodnost modelu

Otázky, nejasnosti

# **AKM 1 5EN306**

## **Logit a probit**

*Martin Janíčko*

30/11/2017

# Struktura prezentace

AKM 1  
5EN306

*Martin Janíčko*

Logit a probit

Funkční zápis

Vhodnost modelu

Otázky, nejasnosti

**1** Logit a probit

**2** Funkční zápis

**3** Vhodnost modelu

**4** Otázky, nejasnosti

# Logit a probit I

AKM 1  
5EN306

Martin Janíčko

Logit a probit

Funkční zápis

Vhodnost modelu

Otázky, nejasnosti

- Logistická regrese se využívá v některých specifických situacích v modelování či předpovědích, kde ostatní metody nesplňují intuitivní či jiné požadavky na výsledky.
- Logistickou regresi používáme, nás zajímá binární výsledek (*dummy*)
- Takovým výsledkem může být například
  - 1 ano/ne;
  - 2 úspěch/neúspěch;
  - 3 bankrot/přežití;
  - 4 zaměstnaný/nezaměstnaný.
- Zároveň ale můžeme vystavět ordered či unordered logit i s více proměnnými, jako např.:

# Logit a probit II

AKM 1  
5EN306

*Martin Janíčko*

Logit a probit

Funkční zápis

Vhodnost modelu

Otázky, nejasnosti

- 1 ano/možná/ne;
  - 2 pravice/umírněná pravice/střed/umírněná levice/levice;
  - 3 žena/muž/dítě;
- Logistická regrese i probit jsou si velice podobné a v podstatě substituovatelné.
  - Rozdíl spočívá zejména v rozdělení závisle proměnné: v probitu se jedná čistě o normální rozdělení, v logitu o logistické.
  - Ř., že logit se blíží k ose o něco pomaleji než probit a zároveň má o něco lepší interpretaci.

# Logit a probit III

AKM 1  
5EN306

Martin Janíčko

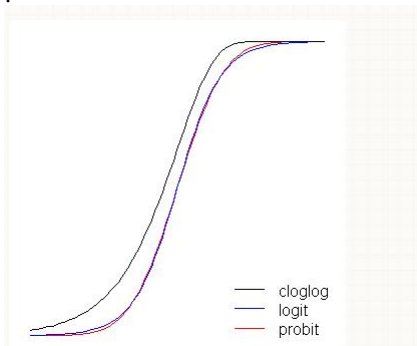
Logit a probit

Funkční zápis

Vhodnost modelu

Otázky, nejasnosti

- Obě metody ale vedou k velice podobným, jakkoli ne zcela identickým, predikcím, a jejich použití je v podstatě indiferentní.



# Logit a probit IV

AKM 1  
5EN306

Martin Janíčko

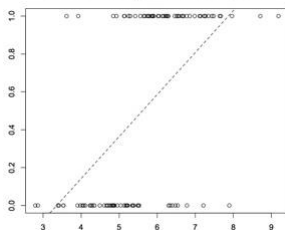
Logit a probit

Funkční zápis

Vhodnost modelu

Otázky, nejasnosti

- Za předpokladu, že používáme logit/probit, můžeme psát, že  $y$  v  $y = \alpha + \beta x + \epsilon$  je sklon k určitému výsledku, například k bankrotu, úspěchu či "ano" na svatbě.
- Použití standardní regrese (např. MNČ) je



nehodné.

- Základním problémem je to, že regresní přímka může uvádět predikce mimo interval  $\langle 0, 1 \rangle$ .

# Logit a probit V

AKM 1  
5EN306

Martin Janíčko

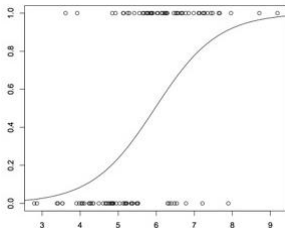
Logit a probit

Funkční zápis

Vhodnost modelu

Otázky, nejasnosti

- Druhým zásadním problémem je očekávání stejného mezního efektu jednotlivých nezávisle proměnných.
- Třetím problémem je výskyt heteroskedasticity, tj. nestejný rozptyl náhodné veličiny v čase.
  - Toto souvisí s nemožností obdržet normálně distribuované chyby odhadu, jelikož výsledek má čistě binární charakter.
- Logit nebo probit tyto problémy řeší.



# Logit a probit VI

AKM 1  
5EN306

Martin Janíčko

Logit a probit

Funkční zápis

Vhodnost modelu

Otázky, nejasnosti

- Meze, v níž může být závisle proměnná definována, jsou jasné dány. Nelineární vztah mezi závisle proměnnou a nezávisle proměnnými zaručuje nekonstantní derivace. Správná specifikace nezávisle proměnných zaručuje homoskedasticitu.
- Zatímco standardní lineární regrese může mít zápis  $y_t = \alpha + \beta x_t + \epsilon_t$ , u logitu a probitu píšeme, že  $y_t = f(\alpha + \beta x_t + \epsilon_t)$
- U logitu používáme tzv. cumulative distribution function logistického rozdělení, zatímco u probitu je to cdf normálního rozdělení.



# Logit a probit VII

AKM 1  
5EN306

Martin Janíčko

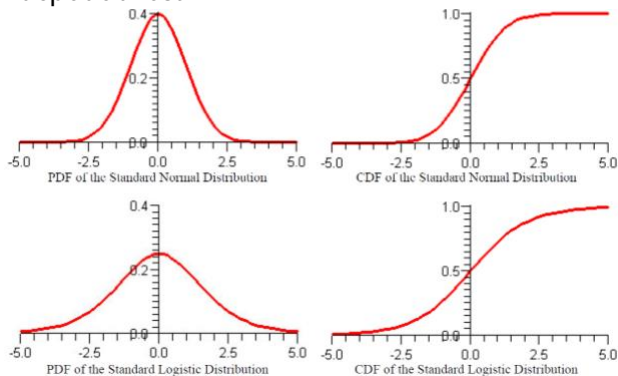
Logit a probit

Funkční zápis

Vhodnost modelu

Otázky, nejasnosti

- U obou fcí tak obdržíme předpokládanou pravděpodobnost.



# Logit a probit VIII

AKM 1  
5EN306

Martin Janíčko

Logit a probit

Funkční zápis

Vhodnost modelu

Otázky, nejasnosti

- Výsledky interpretujeme buď pomocí logaritmované pravděpodobnosti, např.  $\log \text{pst}$  je o 0,5 vyšší u  $y$  vůči  $z$ .
- Nebo naopak, že  $y$  je 1,65krát pravděpodobnější než  $z$ .
- U probitu k interpretaci používáme tzv.  $z$  – score, které má identickou interpretaci, nicméně na různých intervalech různé hodnoty.
- Doprovodné testy, které používáme, jsou věrohodnostní poměr (likelihood ratio) podovnávajících dva různé modely, pseudo R-sq., Waldův test, AIC a BIC.

# Logit a probit IX

AKM 1  
5EN306

*Martin Janíčko*

Logit a probit

Funkční zápis

Vhodnost modelu

Otázky, nejasnosti

- Všechny tyto testy mohou být použity jak pro hodnocení daného modelu vůči jiným specifikacím, tak i pro hodnocení např. logitu vůči jinému modelu.

# Funkční zápis I

AKM 1  
5EN306

Martin Janíčko

Logit a probit

Funkční zápis

Vhodnost modelu

Otázky, nejasnosti

- Máme binární specifikaci a specifikujeme, že  $\pi$  je poměr jedniček (ano, přežití, bankrot, výhra) v množině všech možných stavů světa.
- Můžeme tedy psát, že  $\pi = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$
- Což odpovídá log fci  $\ln\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x$
- Množství nezávisle proměnných v regresi není ničím omezeno a závisí na specifikaci modelu.
- Pokud bychom ale ponechali  $\alpha\beta = 0$ , pak bude pravděpodobnost výskytu daného jevu přesně  $p = 0,5$ .
- A to proto, že  $p = \frac{1}{1 + e^{\beta_0 + \beta_1 x}}$

# Funkční zápis II

AKM 1  
5EN306

Martin Janíčko

Logit a probit

Funkční zápis

Vhodnost modelu

Otázky, nejasnosti

- Pokud jsou naopak bety vysoké, pak se  $p$  blíží 1; pokud jsou nízké, pak nule.
- Definujme si poměr pravděpodobnosti  $odds = \frac{p}{1-p}$ .
- Tento poměr nabývá hodnot v intervalu  $\langle 0, \infty \rangle$
- Transformujeme-li zápis do logaritmické podoby, tedy  $\ln(odds) = \ln\left(\frac{p}{1-p}\right) = \ln(p) - \ln(1-p)$  Tato transformace je užitečná, jelikož se nám proměnná změní do intervalu  $\langle -\infty, +\infty \rangle$

# Vhodnost modelu a možné problémy I

AKM 1  
5EN306

Martin Janíčko

Logit a probit

Funkční zápis

Vhodnost modelu

Otázky, nejasnosti

- Vhodnost modelu určujeme pomocí několika standardních testů, zejm. pak Chi-Square, procento správných předpovědí a pseudo R-sq.
- Chi-Square nám ukáže, zda je celkový model statisticky signifikantní na zvolené hladině významnosti a při zvolených stupních volnosti.
- Procento správných předpovědí nám ukáže, zda model dokáže predikovat s procento správných výsledků z více než poloviny.
- Pseudo R-sq. je McFaddenovo R-sq. a využívá se podobně jako běžné R-sq. s tím rozdílem, že ukazuje rozdíly mezi jednotlivými modely.

# Vhodnost modelu a možné problémy II

AKM 1  
5EN306

Martin Janíčko

Logit a probit

Funkční zápis

Vhodnost modelu

Otázky, nejasnosti

- 1 Omitted variable bias.
- 2 Irrelevant variable bias
- 3 Funkční forma
- 4 Multikolinearita
- 5 Strukturální zlomy

- Omitted variable bias zjišťujeme pomocí likelihood ratio chi-square testu. Pokud je hodnota nižší než kritická a stupně volnosti jsou neporušeny, pak změněný model není zlepšením od jeho původní verze.
- Irrelevant variable bias je podobná situace, která může vyústit ve špatný fit daného modelu. Zde můžeme použít Waldův test nebo opět likelihood ratio chi-square test.

# Vhodnost modelu a možné problémy III

AKM 1  
5EN306

Martin Janíčko

Logit a probit

Funkční zápis

Vhodnost modelu

Otázky, nejasnosti

- Špatná funkční forma může mít za následek nepřesnost odhadu a nesprávné koeficienty. V takovém případě volíme transformaci proměnných, např. jejich logaritmizaci či zavádění poměrů.
- I zde můžeme použít Waldovu statistiku nebo likelihood ratio chi-square test.
- Multikolinearita sice nepovede k nepřesně odhadnutým koeficientům, ale zvýší směrodatnou chybu.
- To povede k nižší statistické signifikanci některých proměnných. V tk případě je vhodné z modelu vyloučit tu proměnnou, o níž se domníváme, že je teoreticky nejméně důležitá.



# Vhodnost modelu a možné problémy IV

AKM 1  
5EN306

*Martin Janíčko*

Logit a probit

Funkční zápis

Vhodnost modelu

Otázky, nejasnosti

- V případě, že data vykazují možné strukturální zlomy, je dobré do specifikace zahrnout dummy proměnné.

# Děkuji za pozornost.

AKM 1  
5EN306

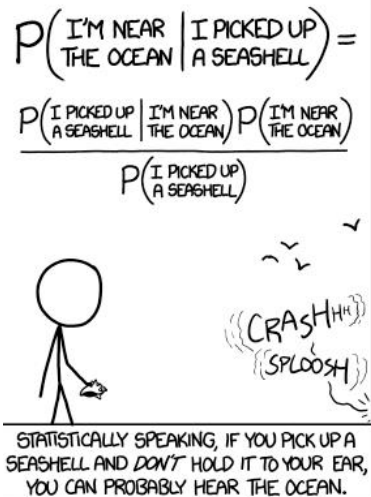
Martin Janíčko

Logit a probit

Funkční zápis

Vhodnost modelu

Otázky, nejasnosti





EVROPSKÁ UNIE  
Evropské strukturální a investiční fondy  
Operační program Výzkum, vývoj a vzdělávání



Toto dílo podléhá licenci Creative Commons  
*Uveďte původ – Zachovejte licenci 4.0 Mezinárodní.*

