Základy ekonometrie

VIII. Modely kvalitativních a omezených vysvětlovaných proměnných

Obsah tématu

Úvod

- Umělé vysvětlující proměnné.
- Umělé vysvětlované proměnné.
- Kategoriální vysvětlované proměnné.
- Omezené vysvětlované proměnné.
- Vysvětlované proměnné vyjadřující počet.

Obsah tématu

Motivace

- Lineární regresní model:

$$Y_i = \alpha + \beta X_i + \epsilon_i.$$

- $Y = 1$ rodina vlastní dům (0 jinak); X příjem rodiny
- Podmíněná pravděpodobnost: $E(Y_i|X_i)$

$$E(Y_i|X_i) = \alpha + \beta X_i.$$

- P_i = pravděpodobnost vlastnictví domu $(Y_i = 1)$; $(1 - P_i)$ = pravděpodobnost $Y_i = 0$.

$$E(Y_i) = 0(1 - P_i) + 1(P_i) = P_i,$$

$$E(Y_i|X_i) = \alpha + \beta X_i = P_i.$$

Problémy

- Nenormalita rozdělení ϵ_i (Bernoulliho rozdělení) $\rightarrow$ není problém.

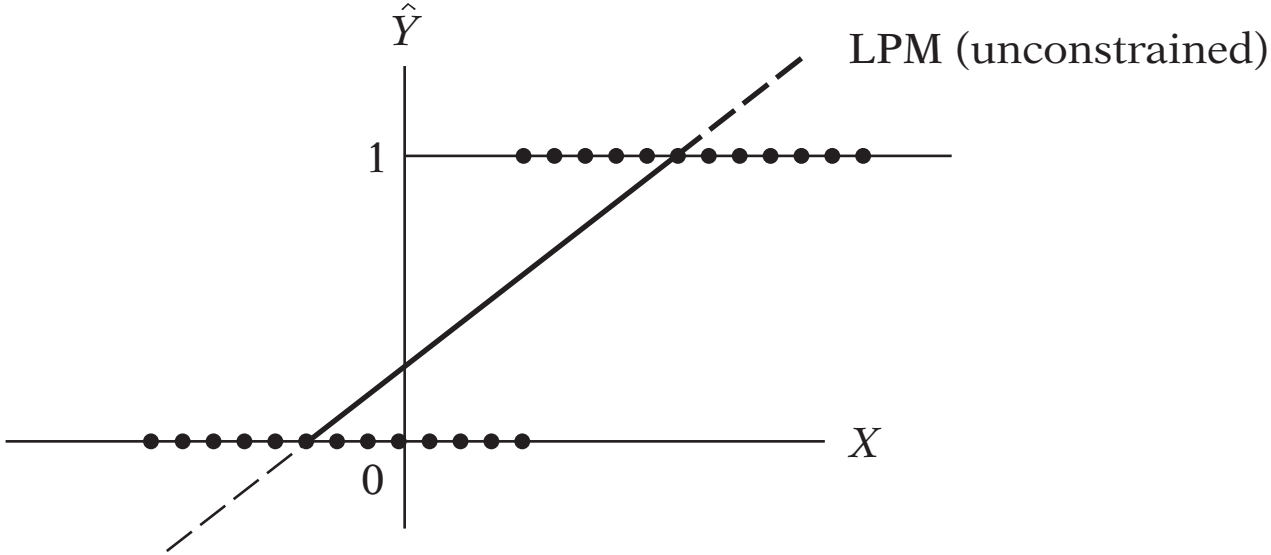
$$\begin{array}{lcc} & \epsilon_i & \text{prav.} \\ Y_i = 1 & 1 - \alpha - \beta X_i & P_i \\ Y_i = 0 & -\alpha - \beta X_i & (1 - P_i) \end{array}$$

- Heteroskedasticita rozptylu náhodných složek: $var(\epsilon_i) = P_i(1 - P_i)$ a $P_i = E(Y_i|X_i) = \alpha + \beta X_i \rightarrow$ WLS s transformací dělením

$$\sqrt{E(Y_i|X_i)[1 - E(Y_i|X_i)]} = \sqrt{P_i(1 - P_i)} = \sqrt{w_i}.$$

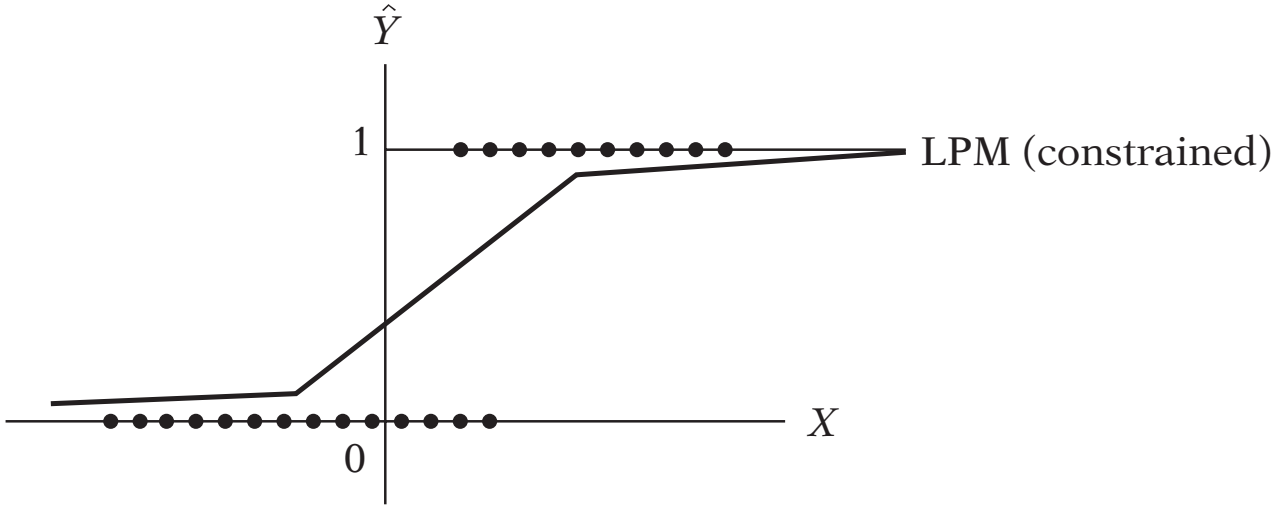
  (OLS regrese + $\widehat{Y}_i$ jako odhad $E(Y_i|X_i) \rightarrow \widehat{w}_i = \widehat{Y}_i(1 - \widehat{Y}_i)$)
- Nesplnění $0 \leq E(Y_i|X) \leq 1$.
- Zpochybnění R^2 jako měřítko kvality vyrovnání (obvykle velmi nízké).

Neomezený LPM



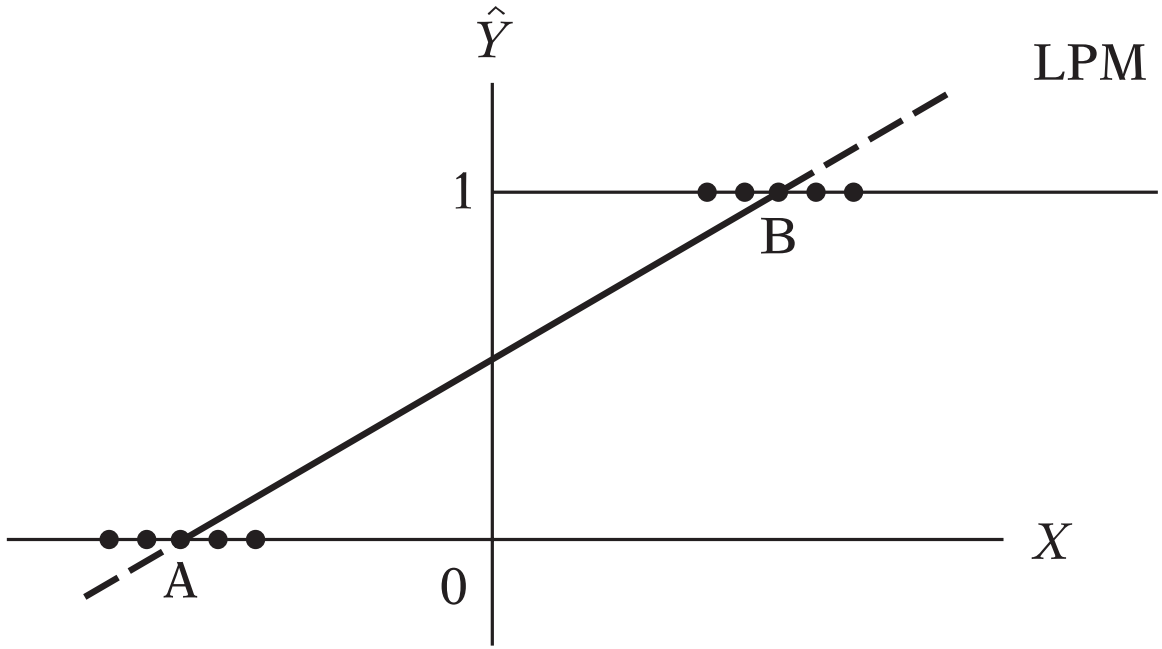
Zdroj: Gujarati, Porter (2009) – Basic econometrics.

Omezený LPM



Zdroj: Gujarati, Porter (2009) – Basic econometrics.

LPM s vyšším R^2

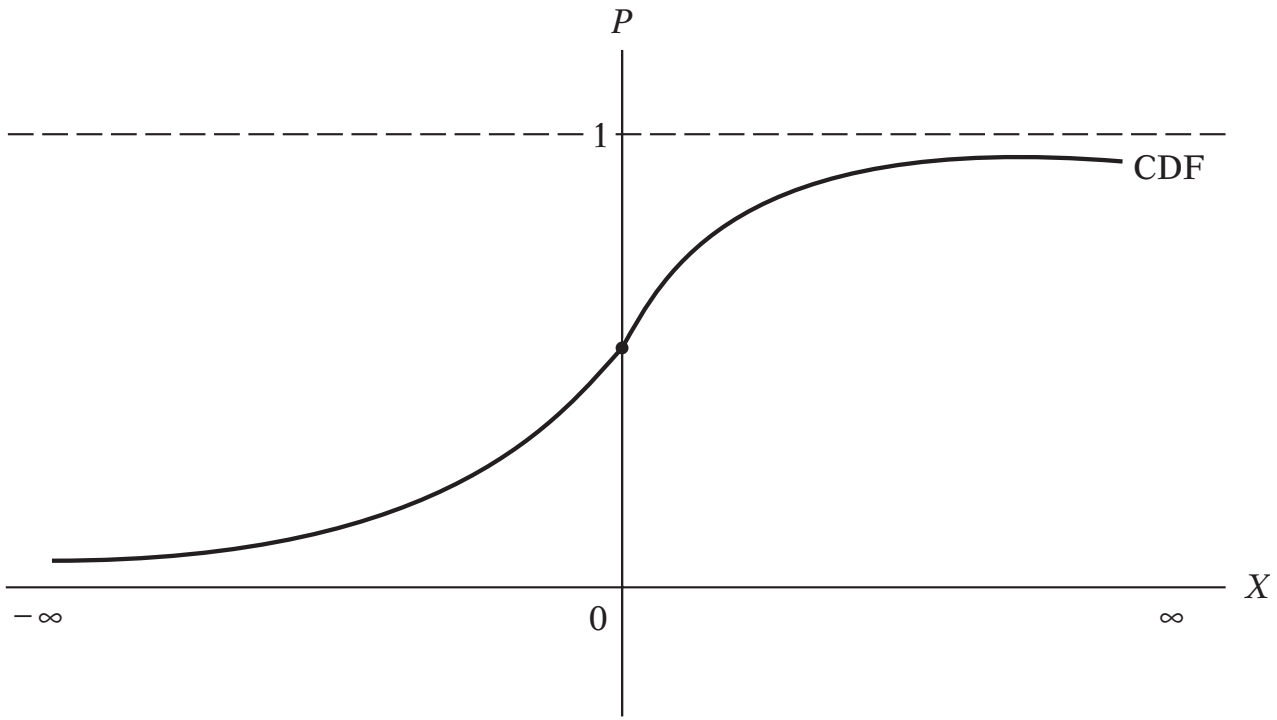


Zdroj: Gujarati, Porter (2009) – Basic econometrics.

Interpretace a zásadní problém

- Parametry: marginální vliv závisle proměnné na pravděpodobnost vysvětlované proměnné.
- Neatraktivní vlastnost: $P_i = E(Y = 1|X)$ roste lineárně s X!
- Příklad vlastnictví domů: $\widehat{\beta} = 0.10 \Rightarrow$ s růstem X o jednotku (1000\$) roste pravděpodobnost o 10 %.
- Rozumné pro důchod: 8000\$, 10000\$, 18000\$, 22000\$?
- Raději P_i nelineárně vztažené k X_i:
  1. S růstem X_i růst $P_i = E(Y = 1|X) \times$ v hranicích $0 - 1$.
  2. Nelineární vztah P_i a X_i (zpomalený pokles k nule pro klesající X_i a zpomalený růst k jedničce pro rostoucí X_i).

Kumulativní distribuční funkce



Zdroj: Gujarati, Porter (2009) – Basic econometrics.

Obsah tématu

Motivace

- Příklad vlastnictví domu:

$$P_i = E(Y = 1|X_i) = \frac{1}{1 + e^{-(\alpha+\beta X_i)}}$$

- (Kumulativní) logistická funkce:

$$P_i = \frac{1}{1 + e^{-Z_i}} = \frac{e^{Z_i}}{1 + e^{Z_i}},$$

kde $Z_i = \alpha + \beta X_i$.

- Splňuje naše požadavky! $\times$ nelze OLS (místo toho ML odhad – logistické rozdělení ϵ_i).

$$1 - P_i = \frac{1}{1 + e^{Z_i}}.$$

Podíl šancí

- Podíl šancí:

$$\frac{P_i}{1-P_i} = \frac{1+e^{Z_i}}{1+e^{-Z_i}} = e^{Z_i}$$

- Interpretace pro příklad vlastnictví domů?
- Přirozený logaritmus:

$$L_i = \ln\left(\frac{P_i}{1-P_i}\right) = Z_i = \alpha + \beta X_i.$$

- $L = \text{logit} \Rightarrow$ logit model.

Vlastnosti logitu

1. P mezi 0 a 1 $\times$ logit neomezen.
2. L lineární v X $\times$ pravděpodobnosti ne!
3. Počet vysvětlujících proměnných dle libosti.
4. Logit zvyšující se hodnotou (záporný) pro klesající podíl šancí z 1 do 0 a růst do nekonečna (kladný) pro růst podílu šancí z 1 do nekonečna.
5. Interpretace β: změna L (logaritmu podílu šancí) pro jednotkovou změnu X.
6. Pro danou úroveň příjmu, X^*, možný výpočet pravděpodobnosti vlastnictví domu (nejen podíl šancí).
7. Oproti LPM: logaritmus podílu šancí lineárně vztažený k X_i.

Logit model – interpretace výsledků

- Značení Koop (jednoduchá regrese).
- Mezní vliv X na pravděpodobnost volby 1 (na základě derivace):

$$\frac{\exp(\beta X_i)}{1+\exp(\beta X_i)}\frac{1}{1+\exp(\beta X_i)}\beta.$$

- Podíl šancí:

$$\frac{\Pr(Y_i=1)}{\Pr(Y_i=0)} = \frac{\frac{\exp(\beta X_i)}{1+\exp(\beta X_i)}}{\frac{1}{1+\exp(\beta X_i)}} = \exp(\beta X_i).$$

- Mezní vliv X na logaritmus podílu šancí: β.
- Vliv jednotkové změny X na podíl šancí: $\exp(\beta)$.

Maximálně věrohodný odhad

- Značení Koop (jednoduchá regrese).

$$L(\beta) = p(Y_1, \ldots, Y_N) = \prod_{i=1}^{N} p\left(Y_i\right).$$

- Logit:

$$L(\beta) = \prod_{i=1}^{N} \left(\frac{\exp(\beta X_i)}{1+\exp(\beta X_i)}\right)^{Y_i} \left(\frac{1}{1+\exp(\beta X_i)}\right)^{1-Y_i}.$$

- Robustní odhad rozptylů (možný problém heteroskedasticity).

Příklad – mimomanželské poměry

- Fair (1978), datový soubor `affair.gdt`:
  - *AFFAIR* = 1 pokud měl jednotlivec tento druh poměru (= 0 jinak);
  - *MALE* = 1 pokud je jednotlivec mužem (= 0 jinak);
  - *YEARS* je počet let manželství daného jednotlivce;
  - *KIDS* = 1 pokud má jednotlivec děti z manželství (= 0 jinak);
  - *RELIG* = 1 pokud se jednotlivec pokládá za nábožensky založeného;
  - *EDUC* je počet ukončených let vzdělání;
  - *HAPPY* = 1 pokud se jednotlivec cítí v manželství šťastný (= 0 jinak).
- Užitečná funkce v *gretlu*: `$coeff` (matice koeficientů odhadu).

Logit – mimomanželské poměry

	Logit			*Podíl šancí*	*Logit (robust)*	
Proměnná	Koef.	*p*-hodn. $\beta_j = 0$	95% int. spol.	Koef.	Koef.	*p*-hodn. $\beta_j = 0$
Konstanta	-1.29	0.07	[-2.71;0.13]	—	-1.29	0.09
MALE	0.25	0.26	[-0.18;0.67]	1.28	0.25	0.27
YEARS	0.05	0.03	[0.01;0.09]	1.05	0.05	0.03
KIDS	0.44	0.12	[-0.12;1.00]	1.55	0.44	0.13
RELIG	-0.89	0.00	[-1.32;-0.47]	0.41	-0.89	0.00
EDUC	0.01	0.75	[-0.07;0.10]	1.01	0.01	0.75
HAPPY	-0.87	0.00	[-1.28;-0.46]	0.42	-0.87	0.09

Skóringové modely

- $\Pr(Y_i = 1)$: pravděpodobnost splacení úvěru.
- $\Pr(Y_i = 0)$: pravděpodobnost nesplacení úvěru.
- Logit: charakteristiky žadatelů.
- Odhad $\rightarrow$ predikční schopnosti modelu.
- Rozdělení na dobré a špatné klienty $\rightarrow$ „cutoff“ hranice (C) + odpovídající skóre .
- Chyba prvního (α) a druhého druhu $(\beta) \rightarrow$ sensitivita a specificita modelu.
- Diskriminační síla modelu – ROC křivka, Cumulative Accuracy Profile (CAP) křivka, Giniho koeficient, Pietra koeficient, Brier skóre, Kolmogorov-Smirnov test apod.

ROC křivka

- Receiver Operating Characteristic
- Poměr úspěšnosti (hit rate):

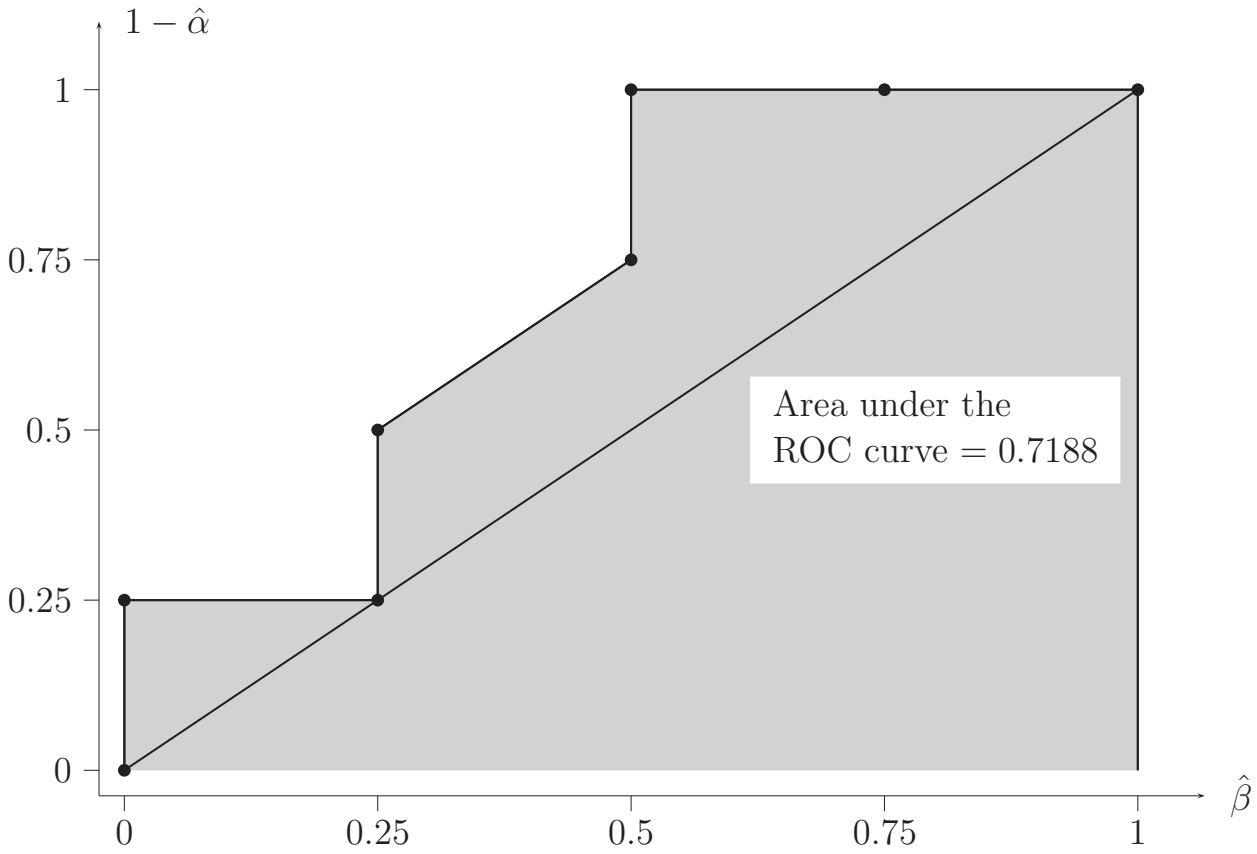
$$HR(C) = \frac{H(C)}{N_D}.$$

- $H(C)$ počet špatných klientů se skóre menším než C; N_D celkový počet špatných klientů $\Rightarrow (1 - \alpha)$.
- Poměr falešného varování (false alarm rate):

$$FAR(C) = \frac{F(C)}{N_{ND}}.$$

- $F(C)$ počet dobrých klientů se skóre menším než C; N_{ND} celkový počet dobrých klientů $\Rightarrow (\beta)$.
- ROC křivka: zobrazení $FAR(C)$ vzhledem k $HR(C)$ pro různá C.
- Obsah plochy pod ROC křivkou = pravděpodobnost, že špatní klienti mají nižší skóre než dobří klienti.

Receiver Operating Characteristic křivka pro umělá data



Zdroj: Winkelmann, Boes (2006) - Analysis of Microdata.

Číselné charakteristiky diskriminační síly modelu

- Z obrázku: A obsah pod ROC křivkou.
- Giniho koeficient – poměr plochy mezi ROC křivkou a diagonálou jednotkového čtverce:

$$GC = 2A - 1$$

- $GC \in (0, 1) \rightarrow$ v praxi uspokojivé 0.60
- Pietra index – obsah plochy největšího trojúhelníku vepsaného mezi ROC křivku a diagonálu jednotkového čtverce:

$$PI = \frac{\sqrt{2}}{4} \max_C |HR(C) - FAR(C)|.$$

Kolmogorovův-Smirnovův test distribučních funkcí HR a FAR.

Obsah tématu

Motivace

- Příklad vlastnictví domů: rozhodnutí o vlastnictví závisí na nepozorovaném rozdílů užitků (latentní proměnná):

$$
\begin{aligned}
Y_i^* &= U_{1i} - U_{0i}; \\
Y_i^* &= \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + \epsilon_i; \\
Y_i^* &= \beta X_i + \epsilon_i.
\end{aligned}
$$

- Pozorujeme rozhodnutí:

$$
\begin{aligned}
Y_i = 1 \quad &\text{pokud} \quad Y_i^* \geq 0, \\
Y_i = 0 \quad &\text{pokud} \quad Y_i^* < 0.
\end{aligned}
$$

Probit funkce

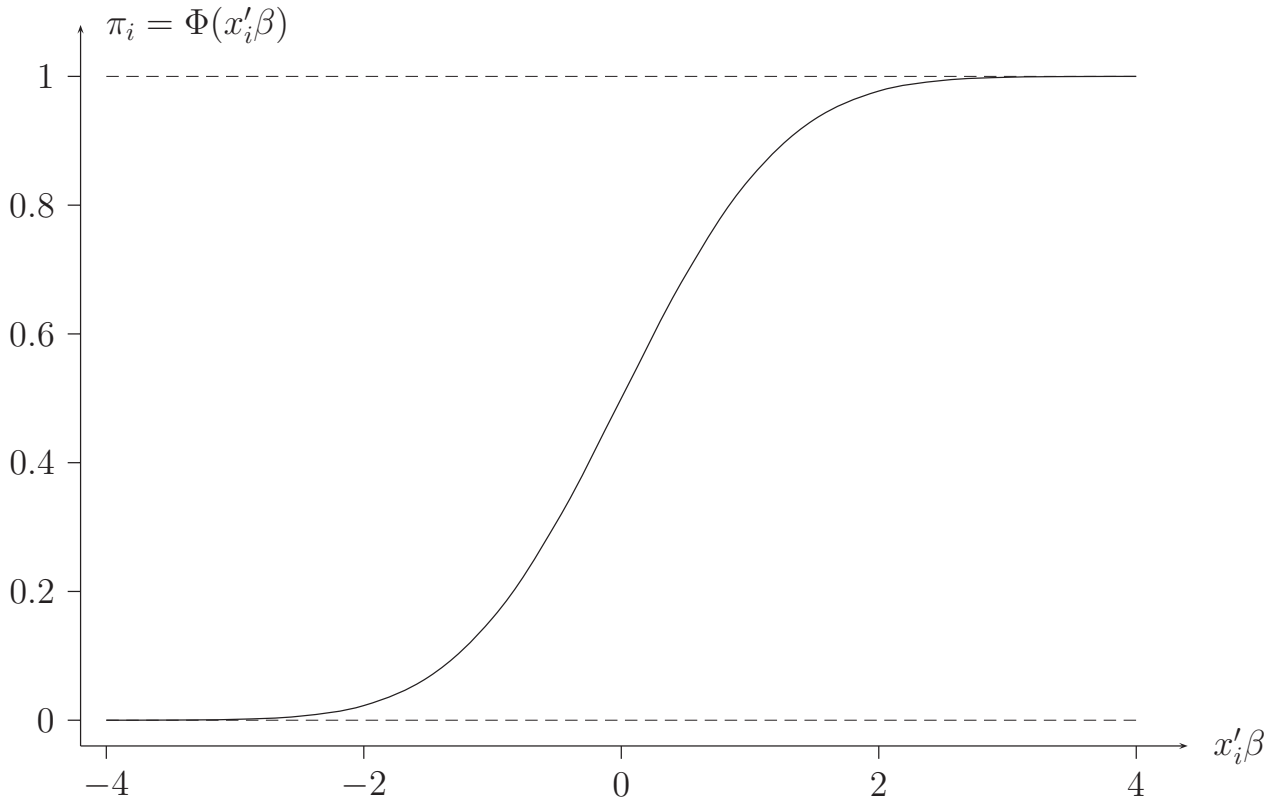
$$\Pr(Y_i = 1) = \Pr(Y_i^* \geq 0) = \Pr(\beta X_i + \epsilon_i \geq 0) = \Pr(\epsilon_i \geq -\beta X_i).$$

- Normální rozdělení náhodné složky.
- Kumulativní distribuční funkce: $\Pr(Z \leq z)$.
- Z standardizovaná normální náhodná veličina (tzn. $N(0,1)$): $\Phi(z)$.
- Probit model:

$$\Pr(Y_i = 1) = \Pr(\epsilon_i \geq -\beta X_i) = 1 - \Phi(-\beta X_i) = \Phi(\beta X_i).$$

- $\Pr(Y_i = 0) = 1 - \Pr(Y_i = 1) \Rightarrow \Pr(Y_i = 0) = \Phi(-\beta X_i)$.
- Probit funkce = inverzní funkce k distribuční funkci: $\Phi^{-1}(p_i)$.

Funkce pravděpodobnosti v probit modelu



Zdroj: Winkelmann, Boes (2006) - Analysis of Microdata.

Mezní vlivy

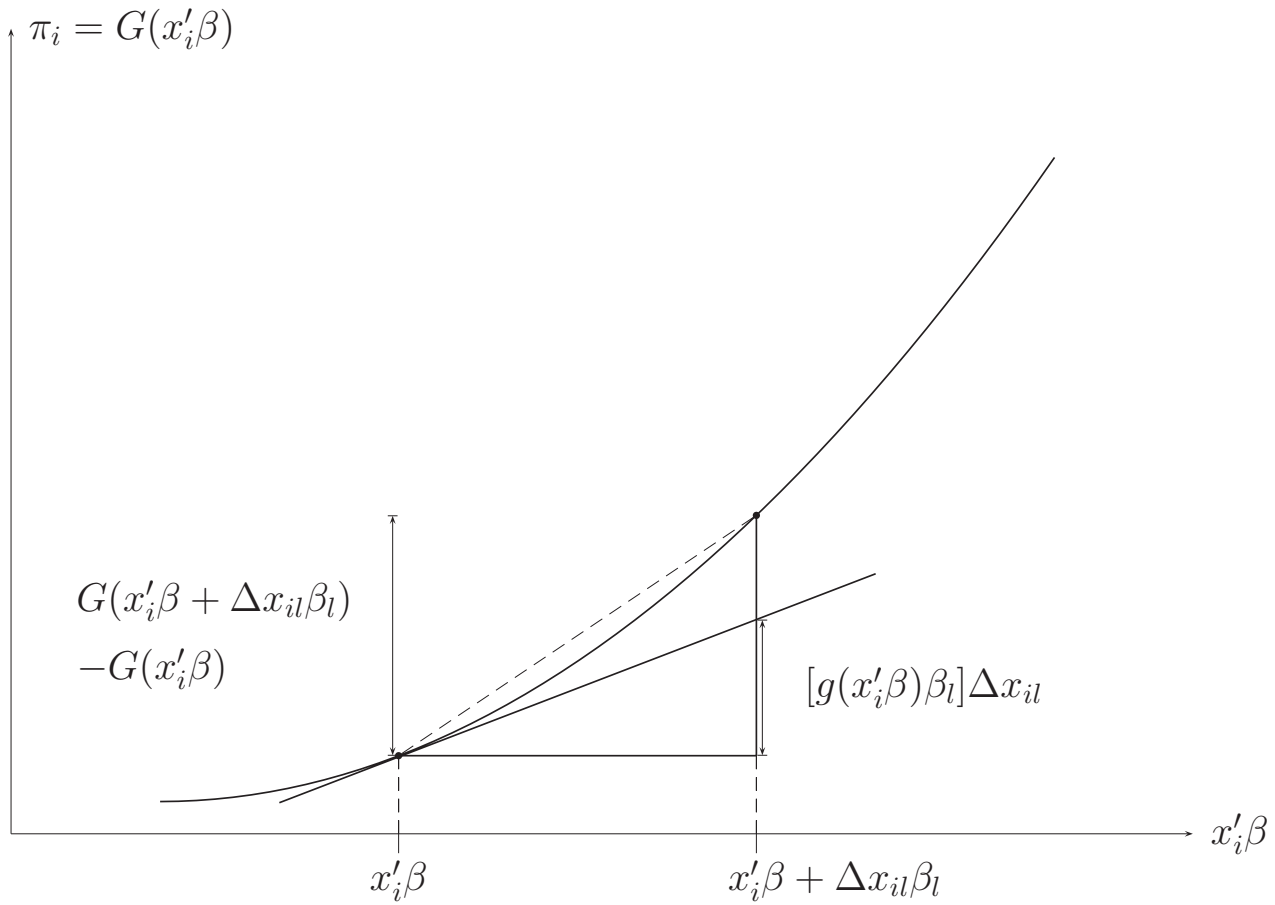
- „Jak se změní *pravděpodobnost volby* 1, pokud změníme X?“.
- Mezní vliv X na pravděpodobnost volby 1 $\rightarrow$ derivace $\Phi(\beta X)$:

$$\phi(\beta X)\beta$$

  $\phi(\cdot)$ = funkce hustoty pravděpodobnosti normálního rozdělení
- Zobecnění pro více regresorů.
- Mezní vlivy pro průměrné hodnoty vysvětlujících proměnných:

$$\phi\left(\widehat{\alpha} + \widehat{\beta}_1\overline{X}_1 + \ldots + \widehat{\beta}_k\overline{X}_k\right)\widehat{\beta}_j.$$

Diskrétní a mezní změny v nelineárních modelech



Zdroj: Winkelmann, Boes (2006) - Analysis of Microdata.

ML odhad

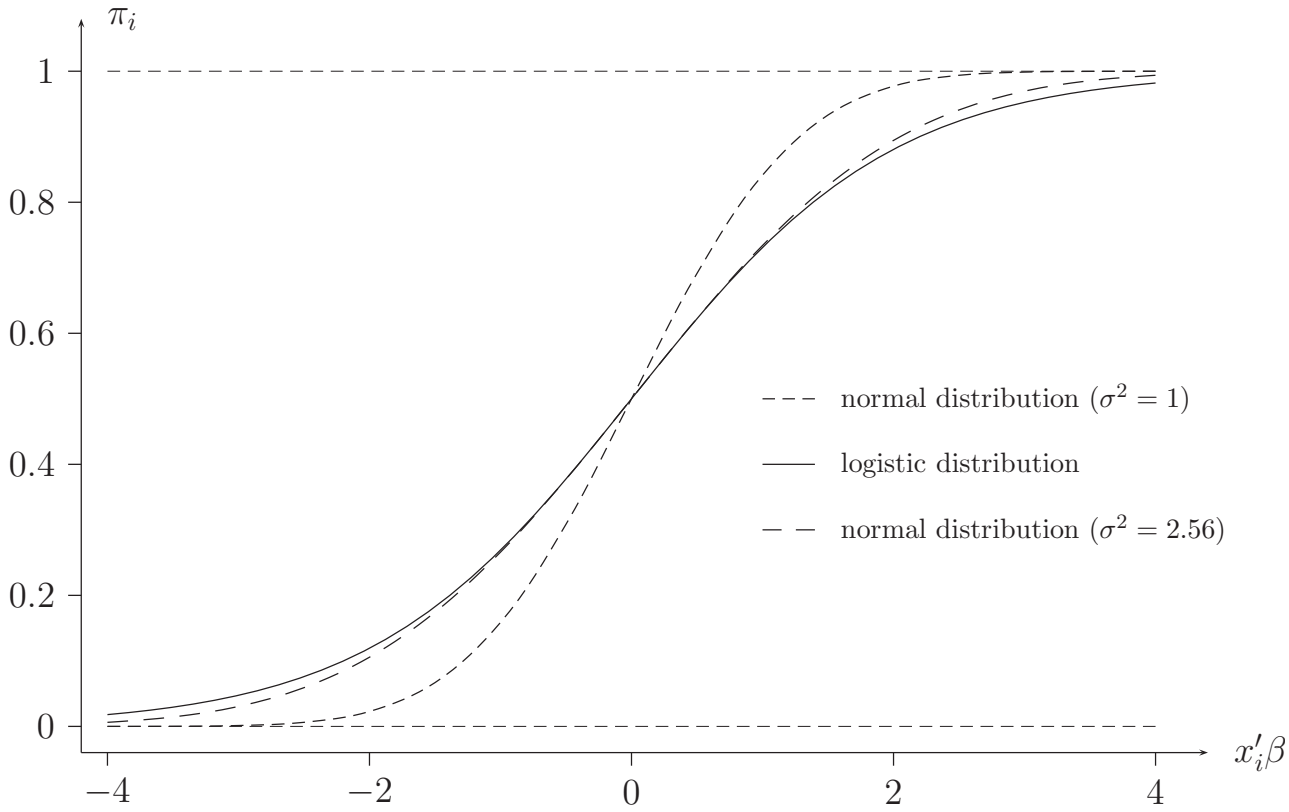
- Pro jediný parametr (snadné zobecnění):

$$L(\beta) = p(Y_1, \ldots, Y_N) = \prod_{i=1}^{N} p\left(Y_i\right).$$

- Probit:

$$L(\beta) = \prod_{i=1}^{N} p(Y_i) = \prod_{i=1}^{N} \Phi(\beta X_i)^{Y_i} \Phi(-\beta X_i)^{1-Y_i}.$$

Probit a logit



Zdroj: Winkelmann, Boes (2006) - Analysis of Microdata.

Probit – mimomanželské poměry

	Probit			*Mezní ef.*	*Probit (robust)*	
Proměnná	Koef.	*p*-hodn. $\beta_j = 0$	95% int. spol.	Koef.	Koef.	*p*-hodn. $\beta_j = 0$
Konstanta	-0.74	0.08	[-1.56;0.09]	—	-0.74	0.11
MALE	0.15	0.23	[-0.10;0.40]	0.05	0.15	0.24
YEARS	0.03	0.03	[0.00;0.05]	0.01	0.03	0.02
KIDS	0.25	0.12	[-0.07;0.57]	0.07	0.25	0.13
RELIG	-0.51	0.00	[-0.75;-0.27]	-0.15	-0.51	0.00
EDUC	0.01	0.81	[-0.04;0.06]	0.00	0.01	0.81
HAPPY	-0.51	0.00	[-0.76;-0.27]	-0.17	-0.51	0.09

Obsah tématu

Motivace

- Y_i hodnoty $0, 1, \ldots, J$.
- Volba alternativy s nejvyšším užitkem.
- Základní alternativa, $j = 0$ (benchmark)

$$Y_{ji}^* = U_{ji} - U_{0i}.$$

- Nepozorovaná diference užitků $\times$ pozorovaná volba.

$$Y_{ji}^* = \alpha_j + \beta_{j1} X_{1i} + \beta_{j2} X_{2i} + \ldots + \beta_{jk} X_{ki} + \epsilon_{ji}.$$

- Vysvětlující proměnné bez indexu j! (variabilita jen mezi jednotlivci + lze „obejít“)

Obsah tématu

Motivace

- Náhodné složky: normální rozdělení.
- Problém: ϵ_{ji} vzájemně korelované.
- Potřeba odhadů všech možných korelací.
- Pokud více alternativ $\rightarrow$ problém s přesností odhadu.

Obsah tématu

Motivace

- Vhodný i pro více alternativ.
- Pravděpodobnost i-tého jednotlivce pro volbu j:

$$\Pr(Y_i = j) = \frac{\exp(\beta_j X_i)}{1 + \sum_{s=1}^{J} \exp(\beta_s X_i)}.$$

- Zobecnění pro vícenásobnou regresi (v rámci exponentu).
- Odhad j regresních rovnic.
- Mezní vliv na základě derivace.

Nezávislost irelevantních alternativ

- Předpoklad použití!
- Podíly šancí se s přidáním alternativy nemění.
- Dopravní příklad: auto ($Y = 0$), věřejná doprava ($Y = 1$), kolo ($Y = 2$).
- Porušení: auto ($Y = 0$), červený autobus ($Y = 1$), modrý autobus ($Y = 2$).
- Řešení skrze vnořený (nested) logit model: nejdříve auto $\times$ hromadná doprava $\rightarrow$ po volbě hromadné dopravy logit pro červený $\times$ modrý autobus.

Příklad – poptávka po crackerech

- Paap a Franses (2000), datový soubor `cracker.gdt`:
- $N = 136$, domácností, čtyři druhy crackerů.
- Nezáleží na volbě základní alternativy!
- Užitečná funkce v *gretlu*: tvorba matice z vysvětlujících proměnných a maticové násobení $\rightarrow$ tvorba proměnné z vektoru: v konzoli `series promenna=vektor`.
- Pro výpočty jednotlivých pravděpodobností volby a popisných statistik.

Multinomiální logit – crackery

	Stř. hodnota	p-hodnota pro $\beta_j = 0$	95% int. spol.
Sunshine			
α_1	-10.06	0.15	[-23.59;3.46]
β_{11}	-7.98	0.01	[0.77;24.02]
β_{12}	12.39	0.04	[0.77;24.02]
β_{13}	0.37	0.91	[-5.83;6.57]
β_{14}	4.83	0.36	[-5.54;15.20]
Keebler			
α_2	-2.53	0.73	[-16.90;11.85]
β_{21}	-3.10	0.30	[-9.01;2.81]
β_{22}	-0.60	0.92	[-12.99;2.81]
β_{23}	1.15	0.70	[-4.67;6.97]
β_{24}	5.33	0.25	[-3.66;14.32]
Nabisco			
α_3	-7.01	0.09	[-15.09;1.07]
β_{31}	-1.38	0.48	[-5.23;2.48]
β_{32}	5.57	0.12	[-1.37;12.50]
β_{33}	0.86	0.65	[-2.84;4.56]
β_{34}	4.72	0.06	[-0.23;9.67]

Multinomiální logit – predikované pravděpodobnosti

Pravděpodobnost nákupu	Stř. hodnota	Sm. odch.	Min.	Max.
Sunshine	0.08	0.11	0.01	0.64
Keebler	0.07	0.03	0.02	0.16
Nabisco	0.60	0.10	0.31	0.80
Private label	0.25	0.11	0.02	0.49

Obsah tématu

Motivace

- Model:

$$Y_{ji}^{*} = \alpha_j + \beta_{j1} X_{1i} + \beta_{j2} X_{2i} + \ldots + \beta_{jk} X_{ki} + \epsilon_{ji}.$$

- Multinomiální logit a pravděpodobnost:

$$\Pr(Y_i = j) = \frac{\exp(\beta_j X_i)}{1 + \sum_{s=1}^{J} \exp(\beta_s X_i)}.$$

- Podmíněný logit a pravděpodobnost:

$$\Pr(Y_i = j) = \frac{\exp(\beta X_{ji})}{1 + \sum_{s=1}^{J} \exp(\beta X_{si})}.$$

- Mezní vlivy(efekty):

$$\frac{\partial \Pr(Y_i = j)}{\partial X_{ji}}.$$

- Přechod na multinomiální logit: $Z_i \times D_{ji}$.

Obsah tématu

Motivace

- Krátce: dotazníková šetření, vysvětlované proměnné kvalitativní (dobrý, průměrný, slabý), ale uspořádatelné.
- Klíčový vztah mezi (vektory) y^* a y (y_i má hodnoty $j = 1, \ldots, J$; J je počet uspořádaných alternativ):

$$y_i = j \quad \text{pokud} \quad \gamma_{j-1} < y_i^* \leq \gamma_j,$$

- $\gamma = (\gamma_0, \gamma_1, \ldots, \gamma_J)'$ je vektor parametrů, kde $\gamma_0 \leq \ldots \leq \gamma_J$.
- Normalita regresního modelu pro latentní data:

$$\begin{aligned} \Pr(y_i = j|\beta, \gamma) &= \Pr(\gamma_{j-1} < y_i^* \leq \gamma_j|\beta, \gamma) \\ &= \Pr(\gamma_{j-1} < x_i'\beta + \epsilon_i \leq \gamma_j|\beta, \gamma) \\ &= \Pr(\gamma_{j-1} - x_i'\beta < \epsilon_i \leq \gamma_j - x_i'\beta|\beta, \gamma). \end{aligned}$$

- ϵ_i z $N(0,1)$: $\Pr(y_i = j|\beta, \gamma) = \Phi(\gamma_j - x_i'\beta) - \Phi(\gamma_{j-1} - x_i'\beta)$
- Obvyklé řešení problému identifikace: $\gamma_0 = -\infty$, $\gamma_1 = 0$ a $\gamma_J = \infty$.

Obsah tématu

Motivace

- Závisle proměnná cenzorovaná na hodnotě nula:

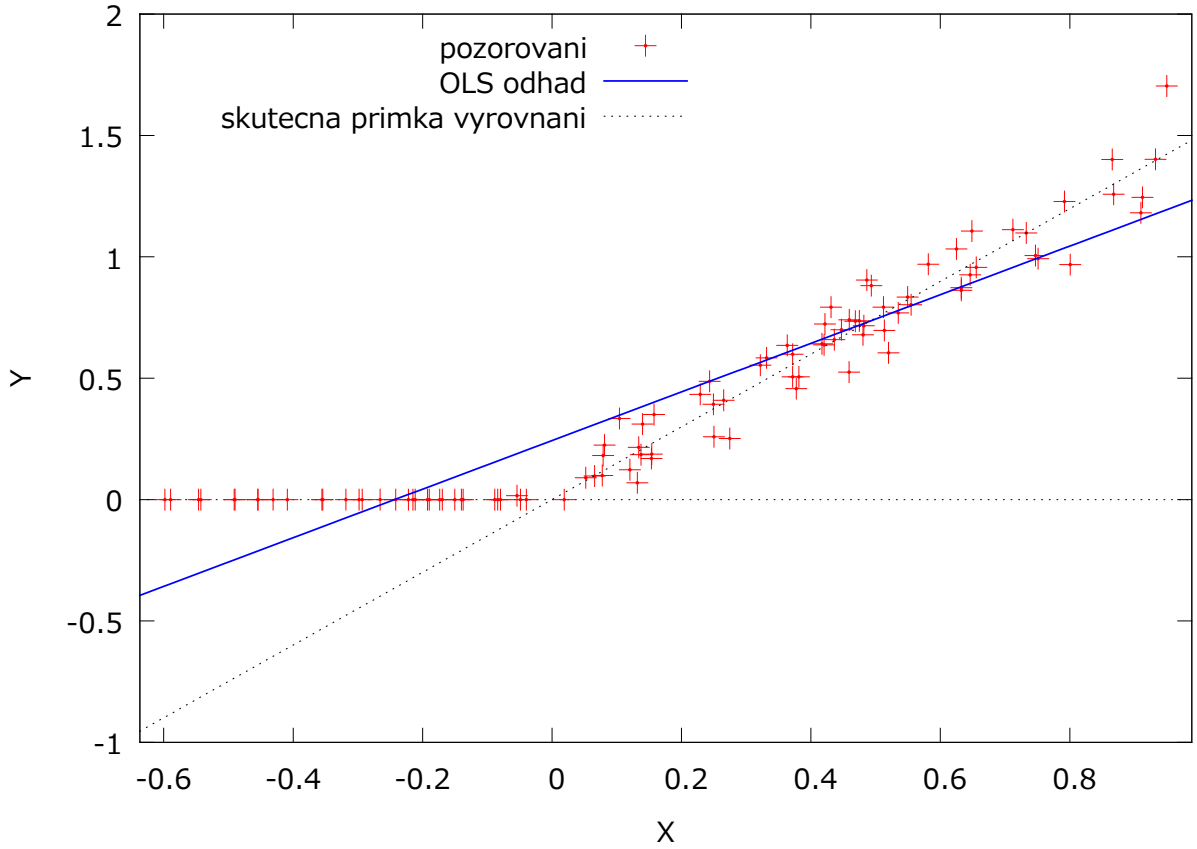
$$Y_i^* = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + \epsilon_i.$$

- Pozorujeme Y_i:

$$\begin{aligned} Y_i &= Y_i^* \quad \text{pokud} \quad Y_i^* > 0, \\ Y_i &= 0 \quad \text{pokud} \quad Y_i^* \leq 0. \end{aligned}$$

- Příklad: závislost požadovaných investic na charakteristikách firmy.
- Obvyklá interpretace výsledků.

Odhady – OLS a tobit



Umělý datový soubor `tobit.gdt`.

Obsah tématu

Motivace

- Práce s daty vyjadřujícími počet.
- Nenormalita rozdělení (není asymptoticky problém) $\Rightarrow$ LRM a OLS $\times$ lepší modely.
- Poissonův regresní model: klasické předpoklady s Poissonovým rozdělením vysvětlované proměnné.

$$E(Y_i) = \lambda_i;$$
$$E(Y_i) = \lambda_i = \beta X_i;$$
$$E(Y_i) = \lambda_i = \exp(\beta X_i).$$

Mezní vlivy

- Vícenásobná regrese:

$$E(Y_i) = \lambda_i = \exp(\alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki}).$$

- Mezní vliv:

$$\frac{\mathrm{d}E(Y_i)}{\mathrm{d}X_{ji}} = \beta_j \exp(\alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki}).$$

- Podíl relativních incidencí:

$$\frac{\exp(\alpha + \beta_1 X_1 + \ldots + \beta_j (X_j + 1) + \ldots + \beta_k X_k)}{\exp(\alpha + \beta_1 X_1 + \ldots + \beta_j X_j + \ldots + \beta_k X_k)} = \exp(\beta_j).$$

- Logaritmus vysvětlující proměnné $\Rightarrow$ mezní vliv je $\exp(\beta)$:

$$E(Y_i) = \lambda_i = \exp(\beta \ln(X_i)) = X_i \exp(\beta).$$

Testování přerozptýlenosti

- Pro vhodnost Poissonova modelu $\times$ jinak např. negativní binomiální regresní model.
- Poissonův regresní model: $E(Y_i) = \lambda_i$; $var(Y_i) = \lambda_i$.
- $H_0 : E(Y_i) = var(Y_i) \rightarrow$ Cameronův-Trivediho test:
  1. Odhad Poissonova modelu; vyrovnané hodnoty $\widehat{\lambda}_i$.
  2. Nová proměnná:
$$Z_i = \frac{\left(Y_i - \widehat{\lambda}_i\right)^2 - Y_i}{\widehat{\lambda}_i \sqrt{2}}.$$
  3. Při platnosti H_0 má Z_i nulovou střední hodnotu.
  4. Regrese Z na úrovňovou konstantu + t-test.

Příklad – poptávka po zdravotní péči

- Vysvětlení faktorů ovlivňujících poptávku po zdravotní péči mezi seniory.
- Deb a Trivedi (1987); data o $N = 4406$ Američanů ve věku 66 a více let; `deb_trivedi.gdt`.
  - *DRVISIT* = počet návštěv u lékaře v minulém roce;
  - *FAMINC* = rodinný příjem (v desítkách tisíc dolarů);
  - *MALE* = 1 pokud je jednotlivec muž (= 0 jinak);
  - *EXCHLTH* = 1 pokud osoba cítí, že má výborné zdraví (= 0 jinak);
  - *POORHLTH* = 1 pokud osoba cítí, že má chatrné zdraví (= 0 jinak);
  - *AGE* věk respondenta (v letech dělený stovkou);
  - *MARRIED* = 1 pokud je osoba ženatá nebo vdaná (= 0 jinak);
  - *PRIVINS* = 1 pokud má osoba soukromé zdravotní pojištění (= 0 jinak).

Poissonův model – poptávka po zdravotní péči.

Proměnná	Koef.	p-hodnota pro $\beta_j = 0$	95% int. spol.	IRR*
Konstanta	1.78	0.00	[1.62;1.94]	—
FAMINC	0.004	0.08	[-0.001;0.008]	1.004
MALE	-0.09	0.00	[-0.11;-0.06]	0.92
EXCHLTH	-0.49	0.00	[-0.54;-0.43]	0.62
POORHLTH	0.53	0.00	[0.49;0.56]	1.69
AGE	-0.03	0.00	[-0.05;-0.01]	0.97
MARRIED	-0.06	0.00	[-0.03;-0.09]	0.94
PRIVINS	0.29	0.00	[0.26;0.32]	1.33

* *Incidence rate ratio* – podíl relativních incidencí.

Obsah tématu

Motivace

- *Duration models* – data vyjadřující množství času, který uběhne před tím, než nastane nějaká událost (např. doba než nezaměstnaný nalezne práci popř. samotná doba nezaměstnanosti, čas mezi dvěma nákupy jednoho výrobku, délka stávky).
- V řadě případů data omezena zprava – v době měření událost ještě nenastala (např. pozorovaná osoba je ještě nezaměstnaná, spotřebitel ještě produkt podruhé nekoupiů, pozorovaná stávka ještě neskončila) $\rightarrow$ potřeba zakomponovat v rámci odhadové metody.
- Často data omezena jen na pozorování, kdy událost nastala před dobou měření – potřeba zohlednění tohoto omezení.

Hazard rate

- Otázka doba trvání, pokud ještě událost nenanastala $\rightarrow$ *riziková funkce* měří šanci (pravděpodobnost), že trvání bude ukončeno nyní, za podmínky, že nebylo ukončeno v minulosti (např. šance nalezení práce, zakoupení produktu, ukončení stávky).
- Modely trvání vyjádřeny skrze „hazard rate“ $\rightarrow$ ekonometrická otázka odhadu této rizikové funkce z pozorovaných dat trvání.
- Data n trvání: $y_1, \ldots, y_n$; předpoklad, že pocházejí z náhodného výběru z populace s funkcí hustoty f a s odpovídající kumulativní distribuční funkcí F.
- *Funkce přežití (survival function)* $S(t)$ a riziková funkce $\lambda(t)$:

$$S(t) = P[y_i > t] = 1 - F(t)$$

$$\lambda(t) = \lim_{\delta \downarrow 0} \frac{P[t < y_i \leq t + \delta | y_i > t]}{\delta}.$$

Hazard rate (pokračování)

- Odhad λ místo f.

$$\lambda(t) = \frac{f(t)}{S(t)} = -\frac{\mathrm{d}\log(S(t))}{\mathrm{d}t},$$

- a lze odvodit

$$f(t) = \lambda(t)S(t), \quad S(t) = e^{-\int_0^t \lambda(s)\mathrm{d}s}.$$

- Modely rizikové funkce: dle požadavku na konstantnost, růst nebo pokles pravděpodobnosti realizace události v čase.
- *Exponential hazard model*: konstantní riziková funkce (pro všechna t)

$$\lambda(t) = \gamma$$

odpovídá funkci hustoty $f(t) = \gamma e^{-\gamma t}$, tedy exp. rozdělení.

- *Weibull hazard model* s Weibullovým rozdělením $f(t) = \alpha\gamma t^{\alpha-1}e^{-\gamma t^{\alpha}}$

$$\lambda(t) = \alpha\gamma t^{\alpha-1}.$$

Růst pro $\alpha > 1$, pokles pro $\alpha < 1$ a konstantnost pro $\alpha = 1$.

Hazard rate (dokončení)

- *Log-normální rozdělení*, kde logaritmus trvání $\log(y_i)$ má normální rozdělení se střední hodnotou μ a rozptylem σ^2

$$\lambda(t) = \frac{\phi\left(\frac{\log(t)-\mu}{\sigma}\right)}{\sigma t\left(1-\Phi\left(\frac{\log(t)-\mu}{\sigma}\right)\right)}$$

Riziková funkce nejdříve roste a následně klesá s bodem obratu daným řešením rovnice $t\sigma\lambda(t) = \sigma + (\log(t) - \mu)/\sigma$.

Proporční hazard model

- Riziková funkce různá pro jednotlivce.
- Předpokládá individuální rizikové funkce jako $\lambda_i(t) = g_i\lambda(t)$, kde faktor $g_i > 0$ odpovídá individuálně specifickým vlivům.
- Pro $g_i = e^{x_i'\beta}$, kde x_i jsou proměnné ovlivňující rizikovou funkci

$$\lambda_i(t) = e^{x_i'\beta}\lambda(t).$$

- Základní riziková funkce $\lambda(t)$ obvykle obsahuje škálovací parametr, tudíž potřeba mít x_i bez úrovňové konstanty.
- Lineární závislost logaritmu rizikové funkce:

$$\log(\lambda_i(t)) = x_i'\beta + \log(\lambda(t)).$$

- Podobné LRM, ale logaritmus základní rizikové funkce je nepozorovatelný; parametr β měří mezní relativní vliv vysvětlující proměnné na rizikovou funkci:

$$\beta = \frac{\partial \log(\lambda_i(t))}{\partial x_i} = \frac{1}{\lambda_i(t)}\frac{\partial \lambda_i(t)}{\partial x_i}.$$

Rozšíření

- Kombinace s modely panelových dat $\rightarrow$ probit model náhodných vlivů.
- Varianty obecných logit a probit modelů.
- *Treatment effects models (modely efektů léčby)*: z medicíny $\times$ i v ekonomii (např. efekt programu rekvalifikací či jiných politik).