UNIVERZITA PARDUBICE

FAKULTA EKONOMICKO-SPRÁVNÍ

BAKALÁŘSKÁ PRÁCE

2021 Robin Kováč

Univerzita Pardubice

Fakulta ekonomicko-správní

Logistický růstový model a jeho aplikace v ekonometrické analýze

Bakalářská práce

2021 Robin Kováč

Univerzita Pardubice
Fakulta ekonomicko-správní
Akademický rok: 2020/2021

ZADÁNÍ BAKALÁŘSKÉ PRÁCE
(projektu, uměleckého díla, uměleckého výkonu)

Jméno a příjmení:	**Robin Kováč**
Osobní číslo:	**E17321**
Studijní program:	**B6208 Ekonomika a management**
Studijní obor:	**Ekonomika a provoz podniku**
Téma práce:	**Logistický růstový model a jeho aplikace v ekonometrické analýze**
Zadávající katedra:	**Ústav matematiky a kvantitativních metod**

Zásady pro vypracování

Cíl práce: popis reálných úloh managementu podniku, které vedou na logistický růstový model. Dále bude tato problematika teoreticky popsána a v závěru bude řešeno několik příkladů.

Osnova:
- Regresní analýza a její aplikace.
- Logistická regrese.
- Logistický růstový model, jeho využití a interpretace.
- Reálné praktické příklady použití logistické regrese.

Rozsah pracovní zprávy: **35**
Rozsah grafických prací:
Forma zpracování bakalářské práce: **tištěná/elektronická**

Seznam doporučené literatury:

HEBÁK, Petr. Vícerozměrné statistické metody. [3]. Praha: Informatorium, 2005. ISBN 80-7333-039-3
HENDL, Jan. Přehled statistických metod: analýza a metaanalýza dat. 3., přeprac. vyd. Praha: Portál, 2009. ISBN 978-80-7367-482-3.
HOSMER, David W. a Stanley LEMESHOW. Applied logistic regression. 2nd ed. New York: John Wiley, 2000. Wiley series in probability and statistics. ISBN 0-471-35632-8.
HUŠEK, Roman a Jan PELIKÁN. Aplikovaná ekonometrie: teorie a praxe. Praha: Professional Publishing, 2003. ISBN 80-86419-29-0.
CHATTERJEE, Samprit, Ali S. HADI a Bertram PRICE. Regression analysis by example. 3rd ed. New York: John Wiley, 2000. Wiley series in probability and statistics. ISBN 0-471-31946-5.
MELOUN, Milan a Jiří MILITKÝ. Kompendium statistického zpracování dat: metody a řešené úlohy. Vyd. 2., přeprac. a rozš. Praha: Academia, 2006. ISBN 80-200-1396-2.
MELOUN, Milan a Jiří MILITKÝ. Statistical data analysis: a practical guide : complete with 1250 exercises and answer key on CD. New Delhi: Woodhead Publishing India, 2011. Woodhead Publishing India in materials. ISBN 978-0-85709-109-3
MELOUN, Milan, Jiří MILITKÝ a Martin HILL. Počítačová analýza vícerozměrných dat v příkladech. Praha: Academia, 2005. ISBN 80-200-1335-0 PECÁKOVÁ, Iva. Statistika v terénních průzkumech. Praha: Professional Publishing, 2008. ISBN 978-80-86946-74-0.
STANKOVIČOVÁ, Iveta a Mária VOJTKOVÁ. Viacrozmerné štatistické metódy s aplikáciami. Bratislava: Iura Edition, 2007. Ekonómia. ISBN 978-80-8078-152-1.

Vedoucí bakalářské práce: **Mgr. Jana Heckenbergerová, Ph.D.**
Ústav matematiky a kvantitativních metod

Datum zadání bakalářské práce: **1. září 2020**
Termín odevzdání bakalářské práce: **30. dubna 2021**

L.S.

prof. Ing. Jan Stejskal, Ph.D.
děkan

doc. Ing. Marcela Kožená, Ph.D.
vedoucí ústavu

Prohlašuji:

Práci s názvem Logistický růstový model a jeho aplikace v ekonometrické analýze jsem vypracoval samostatně. Veškeré literární prameny a informace, které jsem v práci využil, jsou uvedeny v seznamu použité literatury.

Byl jsem seznámen s tím, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon), ve znění pozdějších předpisů, zejména se skutečností, že Univerzita Pardubice má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle § 60 odst. 1 autorského zákona, a s tím, že pokud dojde k užití této práce mnou nebo bude poskytnuta licence o užití jinému subjektu, je Univerzita Pardubice oprávněna ode mne požadovat přiměřený příspěvek na úhradu nákladů, které na vytvoření díla vynaložila, a to podle okolností až do jejich skutečné výše.

Beru na vědomí, že v souladu s § 47b zákona č. 111/1998 Sb., o vysokých školách a o změně a doplnění dalších zákonů (zákon o vysokých školách), ve znění pozdějších předpisů, a směrnicí Univerzity Pardubice č. 7/2019 Pravidla pro odevzdávání, zveřejňování a formální úpravu závěrečných prací, ve znění pozdějších dodatků, bude práce zveřejněna prostřednictvím Digitální knihovny Univerzity Pardubice.

V Pardubicích dne 1. 04. 2021

Robin Kováč v.r.

PODĚKOVÁNÍ

Velice bych chtěl poděkovat své vedoucí, Mgr. Jana Heckenbergerová, Ph.D., za veškerou pomoc, čas a rady, které mi poskytla během zpracování bakalářské práce.

ANOTACE

Cílem práce je vysvětlit regresní analýzy za využití regresní funkce a její aplikace na lineární modely a také na nelineární modely, ze kterých se popíše převážně funkce logistické regrese. Dále se popíší 3 metody pro odhad parametrů a jejich využití za pomocí transformace na potřebný tvar. Jednotlivé modely a příklady regresní funkce a logistické regrese budou zpracovány pomocí programů MS Excel a Statistica 12.

KLÍČOVÁ SLOVA

Regrese, regresní funkce, logistická regrese, nelineární modely, lineární modely, odhad parametrů

TITLE

Logistic growth model and its application in econometric analysis

ANNOTATION

The aim of this work is to explain the regression analysis using the regression function and its application to linear models and also to nonlinear models, from which the logistic regression function is described. Next, 3 methods for estimating parameters and their uses of transformations into the required shape will be described. Individual models and examples of regression functions and logistic regressions will be processed using MS Excel and Statistica 12.

KEYWORDS

Regression, regression function, logistic regression, nonlinear models, linear models, parameter estimation

Obsah

SEZNAM TABULEK

SEZNAM OBRÁZKŮ

SEZNAM ZKRATEK A ZNAČEK

E (Y | X)	podmíněné očekávání
ROC	Receiver Operating Characteristic (prahová operační charakteristika)
BMI	index tělesné hmotnosti

ÚVOD

Logistická regrese se zabývá problematikou odhadu pravděpodobnosti, jestli daný jev nastal nebo nenastal. Nejčastější využití má v medicíně, nebo v sociálních vědách. Například to může být zjišťování toho, jaká je pravděpodobnost, že člověk na určitou nemoc onemocní nebo ne. V práci se prvně vysvětlí lineární regrese, která tvoří první kapitolu a věnuje se teorii, kde je popsána jak lineární regrese, tak i modely, které jsou linearizovatelné a nelinearizovatelné. Dále ukáže modelování lineární regrese, její odhad parametrů. Druhá kapitola se zabývá samostatnou logistickou regresí, která se zabývá teorií a kde pomocí ROC křivky ukáže kvalitu vyhodnocení logistické regrese. Třetí kapitola bude o aplikaci logistické regrese a v následující kapitola je zaměřená na praktické využití logistické regrese, která se zabývá odhadem parametrů Gompertzovy regrese, logistické regrese a praktického příkladu na získání logistického modelu pomocí programu STATISTICA 12, kde budou detailněji vysvětleny kroky k dosažení potřebných výsledků.

1. Regresní analýza a její aplikace

V kapitole bylo čerpáno z následujících literatur (KUBANOVÁ 2003) a (KUBANOVÁ 2008).

Regresní analýza se zabývá metodami, které umožňují zkoumat vztahy, které vznikají mezi dvěma proměnnými. V této analýze nám půjde o to, abychom přesně popsali tvar mezi vztahem proměnných X a Y a charakterizovat vhodnost toho vztahu pro předpovědění hodnot závislé proměnné za pomoci hodnot nezávislé proměnné.

V regresní analýze se obecně analyzuje vztah mezi proměnnou, která se nazývá cílová nebo také závislá proměnná, a ostatními, které se nazývají nezávislé nebo také ovlivňující proměnné.

Cílová proměnná (závislá), nebo také regresand, se označuje písmenem Y a ovlivňující proměnnou (nezávislá), nebo také regresor, se označuje písmenem X. Vztah mezi těmito proměnnými se reprezentuje pomocí matematického modelu, což je rovnice, která svazuje regresand s regresorem a také svazuje pravděpodobnostní předpoklady, které by měl vztah splňovat.

Stochasticky závislé veličiny jsou takové veličiny, kdy změna hodnoty jedné náhodné veličiny vyvolá změnu rozdělení pravděpodobností druhé náhodné veličiny. Pro takovéto závislé veličiny je charakteristické, že:

- změny závislé proměnné jsou vysvětlovány ne všemi, ale jen některými činiteli těchto změn
- bereme v úvahu působení náhodných vlivů
- při zjišťování údajů připouštíme možnost chyb

1.1. Regresní funkce

Podmíněnou střední hodnotu E (Y | X), která je považována za funkci proměnné x, budeme nazývat regresní funkcí náhodné veličiny Y vzhledem k X. Regresní funkce vyjadřuje změny podmíněné střední hodnoty jedné náhodné veličiny při změně hodnot druhé náhodné veličiny. Grafem je pak regresní křivka.

Regresní funkce nám napomáhá k tomu, že pomocí ní můžeme předpovídat, jaké hodnoty bude mít jedna náhodná veličina, když známe hodnoty náhodné veličiny druhé. Proto náhodné veličině Y nemusí pokaždé při určité hodnotě x náhodné veličiny X nabýt hodnoty E (Y | X), ale bude nabývat hodnoty, které budou rozptýlené okolo ní.

Náhodná veličina Y má pro určitou hodnotu x $(x_1, x_2, \dots, x_m)$ a parametry $(\beta_0, \beta_1, \beta_2, \dots, \beta_k)$ podmíněnou střední hodnotu E (Y | X) = g (x, $\beta_0, \beta_1, \beta_2, \dots, \beta_k$). Regresní funkcí se tu nazývá funkce g proměnné x a parametry $(\beta_0, \beta_1, \beta_2, \dots, \beta_k)$ se nazývají regresní koeficienty

Podle tvaru regresní funkce se rozlišují různé typy regresních modelů:

a) Lineární modely – ty mají vzhledem k parametrům regresní tvar ve tvaru

$$g\ (x, \beta_0, \beta_1, \beta_2, \dots, \beta_k) = \sum_{i=0}^{k} \beta_i\, g_i(x) \tag{1. 1}$$

kde g_i jsou funkce nezávisle proměnných $x = (x_1,\ x_2,\ \dots,\ x_m)$

b) Nelineární modely – ty je možné transformací upravit na lineární tvar:

regresní mocninná funkce:

$$g\ (x, \beta_0, \beta_1) = \beta_0 * x^{\beta_1} \tag{1. 2}$$

regresní exponenciální funkce:

$$g\ (x, \beta_0, \beta_1) = \beta_0 * \beta_1^{x} \tag{1. 3}$$

c) Nelineární modely, které se nedají jednoduše transformovat na lineární tvar, jako například:

(1. 4)

$$g\ (x, \beta_0, \beta_1, \beta_2) = \ \beta_0 * \beta_1^x + \beta_2$$

Jednoduchý model pro lineární regresi je takový, kdy grafem regresní funkce je přímka. Model lineární regrese se bude nazývat model:

(1. 5)

$$Y_i = \alpha + \beta x_i + E,$$

kde pro parametry β_0 a β_1 jsme zvolili značení α a β a E nám značí nezávislé náhodné veličiny a nazývá se jako náhodná složka v lineárním modelu. Náhodná složka zahrnuje působení veličin nebo náhodných vlivů, které však nejsou zařazeny do modelu.

Přímka $y = \alpha + \beta x$ se nazývá regresní přímka, kde α a β jsou neznámé parametry, které musíme odhadnout, a β je její směrnice. Pomocí metody nejmenších čtverců získáme jednotlivé bodové odhady parametrů α a β.

1.1.1. Metoda nejmenších čtverců

Řekněme, že máme dvojici naměřených hodnot (x_1, y_1), (x_2, y_2), …, (x_m, y_m) a k těmto bodům hledáme funkci $\hat{y} = a + bx$, aby k nim co nejvíce tzv. „přiléhala“, kde to „přiléhání“ budeme měřit jako součet rozdílů hodnot $\hat{y}_i - y_i$, který nazýváme jako tzv. rezidui). Jelikož se může stát, že se kladné a záporné rozdíly mezi odchylkami $\hat{y}_i$ a y_i navzájem odečtou, nebudeme brát součet reziduí jako míru přiléhání, ale spíše vezmeme součet jejich čtverců, který zapíšeme jako

(1. 6)

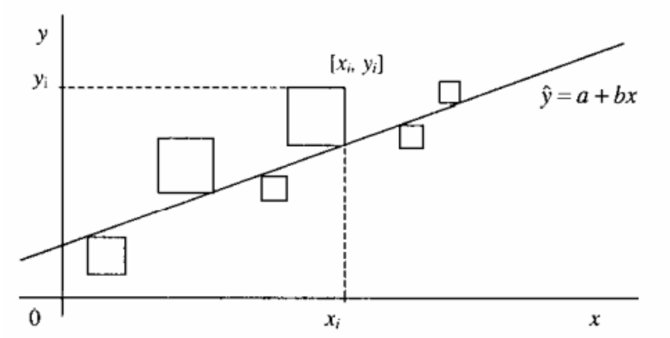
$$\sum_{i=1}^{m} (y_i - \hat{y}_i)^2.$$

Funkce $\hat{y}$ bude mnohem lépe přiléhat k bodům, které naměříme, čím menší bude ten součet. Proto se budeme snažit najít takové odhady parametrů a, b, aby mohlo platit

(1. 7)

$$\sum_{i=1}^{m}(y_i - \hat{y}_i)^2 = min.$$

Grafické znázornění myšlenky v následujícím obrázku:



Obrázek č. 1: Metoda nejmenších čtverců

Zdroj: KUBANOVÁ 2003

Abychom mohli spočítat vzdálenost jednotlivých čtverců, tak budeme muset najít jejich minimum a to pomocí funkce

(1. 8)

$$S(a,b) = \sum_{i=1}^{m}(y_i - \hat{y})^2 = \sum_{i=1}^{m}(y_i - a - bx_i)^2.$$

Poté spočítáme parciální derivaci prvního řádu pro jednotlivé parametry a, b

(1. 9)

$$\frac{\partial S}{\partial a} = -2\sum_{i=1}^{m}(y_i - a - bx_i),$$

(1. 10)

$$\frac{\partial S}{\partial b} = -2\sum_{i=1}^{m}(y_i - a - bx_i) * x_i.$$

Aby mohl extrém funkce o dvou proměnných existovat, je potřebná nutná podmínka a to taková, že položíme obě parciální derivace prvního řádu rovny nule

$$\frac{\partial S}{\partial a} = -2\sum_{i=1}^{m}(y_i - a - bx_i) = 0, \tag{1. 11}$$

$$\frac{\partial S}{\partial b} = -2\sum_{i=1}^{m}(y_i - a - bx_i) * x_i = 0. \tag{1. 12}$$

Po následovné úpravě dostaneme soustavu rovnic o dvou neznámých

$$a\sum_{i=1}^{m} x_i + b\sum_{i=1}^{m} x_i^2 = \sum_{i=1}^{m} x_i\, y_i\,, \tag{1. 13}$$

ze kterých poté vypočítáme odhady pro jednotlivé parametry a, b

$$a = \frac{\sum_{i=1}^{m} x_i^2 \sum_{i=1}^{m} y_i - \sum_{i=1}^{m} x_i \sum_{i=1}^{m} x_i y_i}{m\sum_{i=1}^{m} x_i^2 - (\sum_{i=1}^{m} x_i)^2}, \tag{1. 14}$$

$$b = \frac{m\sum_{i=1}^{m} x_i y_i - \sum_{i=1}^{m} x_i \sum_{i=1}^{m} y_i}{m\sum_{i=1}^{m} x_i^2 - (\sum_{i=1}^{m} x_i)^2}. \tag{1. 15}$$

Pro ověření, že funkce S nabývá ve stacionárním bodě minimum, se musí určit parciální derivace druhého řádu:

$$\frac{\partial^2 S}{\partial a^2} = 2m, \tag{1. 16}$$

$$\frac{\partial^2 S}{\partial a \partial b} = 2\sum_{i=1}^{m} x_i, \tag{1. 17}$$

$$\frac{\partial^2 S}{\partial b^2} = 2\sum_{i=1}^{m} x_i^2. \tag{1. 18}$$

Regresní přímka, která byla získaná pomocí metody nejmenších čtverců, má následující tvar

(1. 19)

$$\hat{y} = a + bx.$$

Rovnice lze poté upravit na tvar:

(1. 20)

$$\hat{y} = \bar{y} + b(x - \bar{x}).$$

PŘÍKLAD Č. 1: Odhad parametrů regresní funkce

Na příkladu si ukážeme, jak odhadnout parametry regresní funkce $\hat{y}(x) = \alpha + \beta x$ na základě dat, které máme uvedeny v následující [Tabulka 1], kde máme naměřené hodnoty:

Tabulka 1 Hodnoty pro odhad parametrů regresní funkce

BODY	x	y
1	2	1
2	4	2
3	3	3
4	5	8
5	8	4
6	6	7
7	9	9
8	10	5
9	7	6
10	11	11

Zdroj: vlastní zpracování

Abychom získali odhady a, b parametrů α, β, použijeme metodu nejmenších čtverců.

Chceme tedy najít minimum a to za použití funkce:

(1. 21)

$$\sum_{i=1}^{10} (y_i - a - bx_i)^2,$$

kde provedeme parciální derivaci prvního řádu pro a, a pro b už po úpravě rovnic:

$$\frac{\partial S}{\partial a} = \sum_{i=1}^{10}(y_i x_i - a x_i - b x_i^2) = 0, \tag{1. 22}$$

$$\frac{\partial S}{\partial b} = \sum_{i=1}^{10}(y_i - a - b x_i) = 0. \tag{1. 23}$$

Po dosazení a úpravách rovnice vypadají následovně:

$$\frac{\partial S}{\partial a} = 427 - a65 - b505, \tag{1. 24}$$

$$\frac{\partial S}{\partial b} = 56 - a10 - b65. \tag{1. 25}$$

Nyní vyřešíme soustavu dvou rovnic o dvou neznámých pro *a*, a pro *b*, kde:

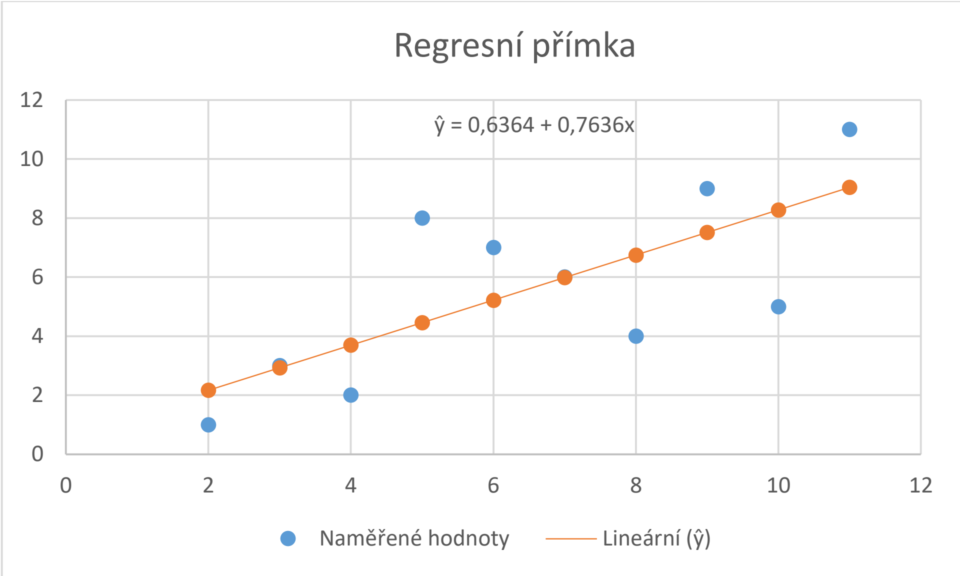
$$a = 0{,}6363637,$$

$$b = 0{,}763636.$$

Poté co známe jednotlivé hodnoty *a*, *b*, má regresní přímka tvar:

$$\hat{y} = 0{,}6363637 + 0{,}763636x \tag{1. 26}$$

a graf této přímky vypadá následovně:



Obrázek č 2: Graf regresní přímky

Zdroj: vlastní zpracování

1.1.2. Linearizovatelné modely

1.1.2.1. Mocninná funkce

V nelineárním tvaru má mocninná funkce tvar

$$y = \alpha x^{\beta} \tag{1. 27}$$

a po zlogaritmování této mocninné funkce vznikne rovnice v lineárním tvaru:

$$\ln y = \ln \alpha + \beta \ln x. \tag{1. 28}$$

Neznáme parametry a, b se vypočítají náhodným výběrem metodou nejmenších čtverců.

Získáváme potom tedy následující soustavu rovnic:

$$m \ln a + b \sum_{i=1}^{m} \ln x_i = \sum_{i=1}^{m} \ln y_i \tag{1. 29}$$

(1. 30)

$$\ln a \sum_{i=1}^{m} \ln x_i + b \sum_{i=1}^{m} \ln^2 x_i = \sum_{i=1}^{m} \ln x_i \ln y_i$$

Poté jednotlivé parametry upravíme:

(1. 31)

$$b = \frac{m \sum_{i=1}^{m} \ln x_i \ln y_i - \sum_{i=1}^{m} \ln x_i \sum_{i=1}^{m} \ln y_i}{m \sum_{i=1}^{m} \ln^2 x_i - \left(\sum_{i=1}^{m} \ln x_i\right)^2},$$

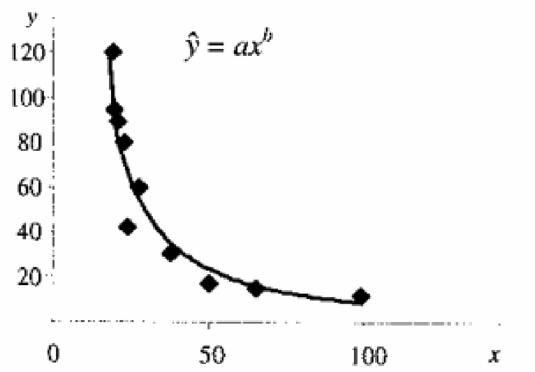
(1. 32)

$$\ln a = \frac{\sum_{i=1}^{m} \ln y_i - b \sum_{i=1}^{m} \ln x_i}{m}.$$

Grafické znázornění nelineární regresní funkce, která má tvar:

(1. 33)

$$\hat{y} = a x^b$$



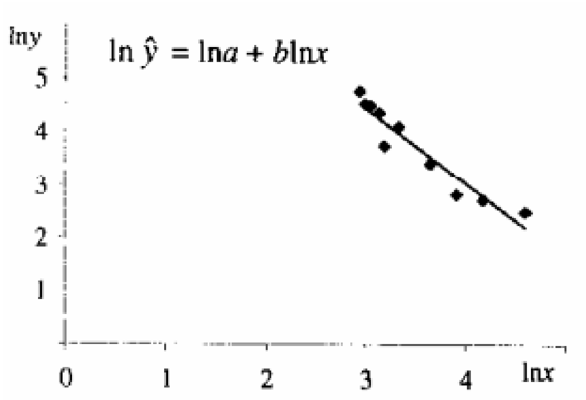
Obrázek č 3: Graf mocninné funkce před trensformací

Zdroj: KUBANOVÁ 2003

Transformací nelineární regresní funkcí vzniká regresní funkce, která má tvar:

(1. 34)

$$\ln \bar{y} = \ln a + b \ln x.$$



Obrázek č. 4: Graf mocninné funkce po užití transformace

Zdroj: KUBANOVÁ 2003

1.1.2.2. Exponenciální funkce

Jako další typ funkce, při kterém se využívá linearizace, je exponenciální funkce, která má tvar:

$$y = \alpha \beta^x, \tag{1.35}$$

kde při zlogaritmování funkce dostaneme rovnici v lineárním tvaru:

$$\ln y = \ln \alpha + x \ln \beta. \tag{1.36}$$

Po použití metody nejmenších čtverců získáme následující soustavu rovnic:

$$m \ln a + \ln b \sum_{i=1}^{m} x_i = \sum_{i=1}^{m} \ln y_i \tag{1.37}$$

$$\ln a \sum_{i=1}^{m} x_i + b \sum_{i=1}^{m} x_i^2 = \sum_{i=1}^{m} x_i \ln y_i \tag{1.38}$$

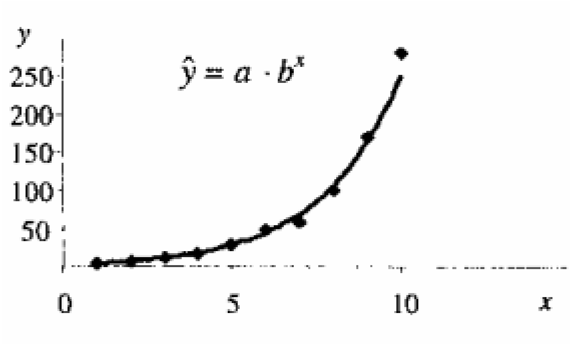
Jednotlivé parametry a, b poté vypočítáme podle vztahů:

$$\ln b = \frac{m \sum_{i=1}^{m} x_i \ln y_i - \sum_{i=1}^{m} x_i \sum_{i=1}^{m} \ln y_i}{m \sum_{i=1}^{m} x_i^2 - (\sum_{i=1}^{m} x_i)^2}, \tag{1.39}$$

$$\ln a = \frac{\sum_{i=1}^{m} \ln y_i - \ln b \sum_{i=1}^{m} x_i}{m}. \tag{1. 40}$$

Grafické znázornění nelineární regresní funkce má tvar:

$$\hat{y} = ab^x \tag{1. 41}$$

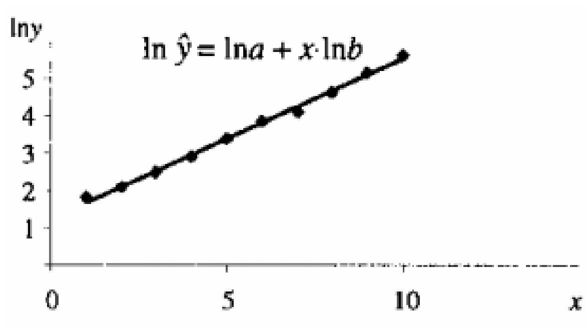


Obrázek č. 5: Graf exponenciální funkce před transformací

Zdroj: KUBANOVÁ 2003

Transformací nelineární regresní funkcí vzniká regresní funkce, která má tvar:

$$\ln \bar{y} = \ln a + x \ln b. \tag{1. 42}$$



Obrázek č. 6: Graf exponenciální funkce po užití transformace

Zdroj: KUBANOVÁ 2003

1.1.3. Nelineární regrese

V nelineární regresi jsou některé typy modelů, které však nelze jednoduše převést transformací na lineární tvar. Pro odhady parametrů se potom nedá použít metoda nejmenších čtverců, jak tomu bylo při lineární regresi.

Nelineárních regresních funkcí je spousta. My se však podíváme jen na ty, se kterými se setkáváme nejčastěji, jsou to:

- Zobecněná exponenciální funkce
- Gompertzova křivka
- Logistická regresní funkce

1.1.3.2. Zobecněná exponenciální funkce

Zobecněná exponenciální funkce se liší od exponenciální funkce, která má v základu tvar funkce:

$$y = \alpha\beta^x, \tag{1. 43}$$

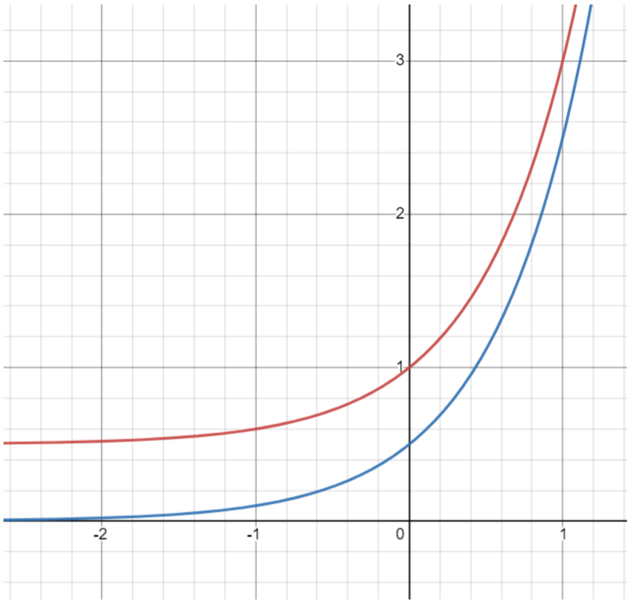
tak zobecněná funkce má navíc přidanou konstantu k. Vzorec pro zobecněnou exponenciální funkci vypadá následovně:

$$y = k + \alpha\beta^x. \tag{1. 44}$$

Na následujícím grafu si můžeme všimnout rozdílu, mezi — exponenciální funkcí a — zobecněnou exponenciální funkcí.

Jednotlivé body na grafu mají následující hodnoty:

$\alpha = 0{,}5$ $\beta = 5$ $k = 0{,}5$,

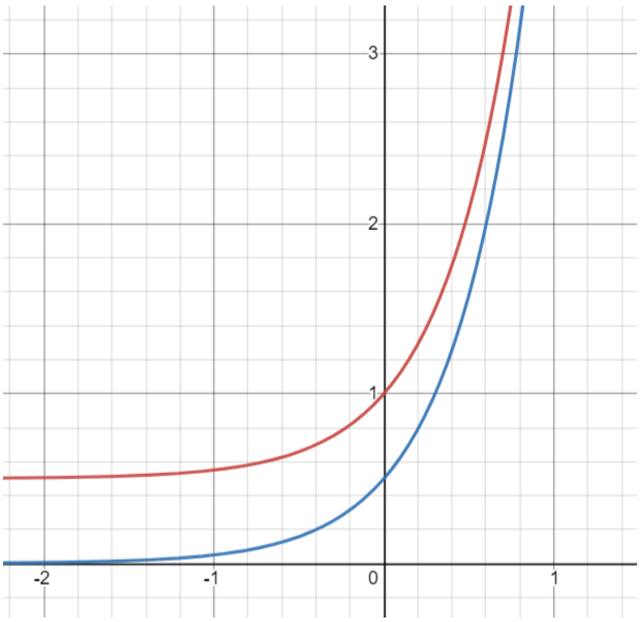


Obrázek č. 7: Grafické znázornění rozdílu mezi exponenciální funkcí a zobecněnou exponenciální funkcí

Zdroj: vlastní zpracování

kde hodnota *k* udává, od kterého minima na ose x zobecněná exponenciální funkce začíná růst, hodnota *β* udává rychlost růstu a hodnota *α* udává její sklon.

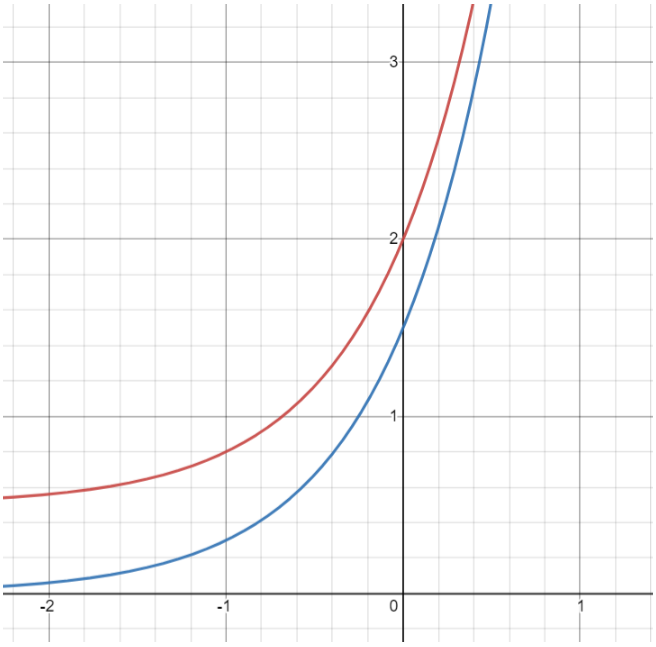
Pro ukázku změníme hodnotu z původního obrázku pro $\beta = 10$



Obrázek č. 8: Grafické znázornění změný hodnoty β

Zdroj: vlastní zpracování

Pro další ukázku změníme hodnotu z původního obrázku pro α = 1,5



Obrázek č. 9: Grafické znázornění změny hodnoty α

Zdroj: vlastní zpracování

Zobecněná exponenciální funkce má své využití jak u Gompertzovy křivky, tak u Logistické regresní funkce, protože aby bylo možné využít metody pro odhad parametrů právě u těchto dvou křivek, které si později popíšeme, musíme je obě převést pomocí transformace na tvar, který bude odpovídat zobecněné exponenciální funkci.

U Gompertzovy křivky se to provede logaritmováním:

$$y = k\alpha^{\beta^x} \tag{1.45}$$

$$\ln y = \ln k\, \alpha^{\beta^x} \tag{1.46}$$

$$\ln y = \ln k + \ln \alpha^{\beta^x} \tag{1.47}$$

$$\ln y = \ln k + \beta^x \ln \alpha \tag{1.48}$$

Názorný příklad, jak pomocí exponenciální regrese dostaneme potřebné parametry, si ukážeme v kapitole 4.1.

V případě logistické regresní funkce se to provede pomocí převrácené hodny:

$$y = \frac{k}{1 + \alpha\beta^x} \tag{1.49}$$

$$\frac{1}{y} = \frac{1 + \alpha\beta^x}{k} \tag{1.50}$$

$$\frac{1}{y} = \frac{1}{k} + \frac{\alpha}{k}\beta^x \tag{1.51}$$

Názorný příklad, jak pomocí exponenciální regrese dostaneme potřebné parametry, si ukážeme v kapitole 4.2.

Metody pro odhady parametrů nelineárních modelů máme následující:

- Metoda částečných součtů
- Metoda dílčích průměrů
- Metoda vybraných bodů

1.1.3.2.1. Metoda částečných součtů

V této metodě postupujeme tak, že si prvně rozdělíme soubor *n* pozorování na tři stejně velké části o délce m. Pokud však máme takový počet prvků v souboru, který není dělitelný třemi, tak vynecháme až tři nejstarší pozorování, podle toho, kolik musíme vynechat, aby byl takový počet prvků, který už je možné dělit třemi. Poté co máme soubor rozdělen, sečteme hodnoty v jednotlivých skupinách na S_1, S_2 a S_3. Podle následujících vztahů se vypočítají odhady parametrů:

$$k = \frac{1}{m}\left(S_1 - ab\frac{b^m - 1}{b - 1}\right) \tag{1. 45}$$

$$a = \frac{(b-1)(S_2 - S_1)}{b(b^m - 1)^2} \tag{1. 46}$$

$$b = \sqrt[m]{\frac{S_3 - S_2}{S_2 - S_1}} \tag{1. 47}$$

PŘÍKLAD Č. 2: Odhad parametrů pomocí metody částečných součtů

Tabulka 2 Hodnoty pro odhad parametrů užití metody částečných součtů

x	y	
1	60	
2	100	
3	180	
4	200	
5	260	
6	350	
7	410	S_1 =1500
8	480	
9	530	
10	600	
11	640	
12	700	
13	770	S_2 =3720
14	820	
15	870	
16	900	
17	930	
18	960	
19	980	S_3 =5460

Zdroj: vlastní zpracování

$$b = \sqrt[6]{\frac{5460 - 3720}{3720 - 1500}} = 0{,}9602 \tag{1. 48}$$

$$a = \frac{(0{,}9602 - 1)(3720 - 1500)}{0{,}9602(0{,}9602^6 - 1)^2} = -1967{,}84 \tag{1. 49}$$

$$k = \frac{1}{6}\left[1500 - (-1967{,}84)0{,}9602\frac{0{,}9602^6 - 1}{0{,}9602 - 1}\right] = 1961{,}2 \tag{1. 50}$$

1.1.3.2.2. Metoda dílčích průměrů

V další metodě si opět rozdělíme celý soubor na tři části s tím, že oproti předchozí metodě dílčích součtů se nemusí rozdělit na 3 stejně velké části o stejném počtu pozorovaných. Zde se v každé části vypočítá hodnota aritmetického průměru a průměry, které vyjdou, si označíme jako S_D, S_P, S_H. U této metody se hodnota parametru pro b odhadne podle vzorce:

$$b = \sqrt[\frac{n-5}{2}]{\frac{S_H - S_P}{S_P - S_D}} \tag{1. 51}$$

Abychom odhadli ostatní parametry, tak použijeme metodu nejmenších čtverců.

PŘÍKLAD Č. 3: Odhad parametrů pomocí metody dílčích průměrů

Tabulka 3 Hodnoty pro odhad parametrů užití metody dílčích průměrů

x	y	
1	60	
2	100	
3	180	
4	200	
5	260	
6	350	S_D =191,67
7	410	
8	480	
9	530	
10	600	
11	640	
12	700	S_P =560
13	770	
14	820	
15	870	
16	900	
17	930	
18	960	
19	980	S_H = 890

Zdroj: vlastní zpracování

(1. 52)

$$b = \sqrt[\frac{19-5}{2}]{\frac{890-560}{560-191{,}67}} = 0{,}9844$$

A pro odhady ostatních parametrů se použije metoda nejmenších čtverců.

$$Y = k + \alpha \cdot \beta^X$$

1.1.3.2.3. Metoda vybraných bodů

U poslední metody, kterou si popíšeme, si z hodnot *y* vybereme tři hodnoty y_x, y_{x+m}, a y_{x+2m}. Vybírají se tak, aby hodnota pro y_x odpovídala některé z prvních hodnot *y*, hodnota y_{x+m}odpovídala některé ze středních hodnot a hodnota y_{x+2m} některé z posledních hodnot *y*, kde *m* je přirozené číslo. Odhad parametrů se následovně vypočítá podle vzorců:

(1. 53)

$$k = y_x - ab^x$$

$$a = \frac{y_{x+m} - y_x}{b^x(b^m - 1)} \tag{1. 54}$$

$$b = \sqrt[m]{\frac{y_{x+2m} - y_{x+m}}{y_{x+m} - y_x}} \tag{1. 55}$$

PŘÍKLAD Č. 4: Odhad parametrů pomocí metody vybraných bodů

Tabulka 4 Hodnoty pro odhad parametrů užití metody vybraných bodů

x	y	
1	60	y_1
2	100	
3	180	
4	200	
5	260	
6	350	
7	410	
8	480	
9	530	
10	600	y_{10}
11	640	
12	700	
13	770	
14	820	
15	870	
16	900	
17	930	
18	960	
19	980	y_{19}

Zdroj: vlastní zpracování

$$b = \sqrt[9]{\frac{980 - 600}{600 - 60}} = 0{,}9617 \tag{1. 56}$$

$$a = \frac{600 - 60}{0{,}9617^1(0{,}9617^9 - 1)} = -1894{,}7367 \tag{1. 57}$$

$$k = 60 - (-1894{,}7367)0{,}9617^1 = 1882{,}1683 \tag{1. 58}$$

1.1.3.1. Gompertzova křivka

Gompertzova křivka, nebo také Gompertzova funkce, je typ matematického modelu časové řady. Je to funkce mající sigmoidální tvar popisující růst, který je nejpomalejší na začátku a postupně roste do konečného bodu časového období.

Gompertzova funkce se většinou aplikuje v medicíně, nebo například při nárůstu populace.

Gompertzova funkce má následující tvar:

$$y = k\alpha^{\beta^x}, \tag{1. 59}$$

kdy α je asymptotou, β nastavuje posun podél osy x a kde x nastavuje rychlost růstu. Maximální hranice Y – ové souřadnice u Gompertzovy křivky se nazývá *Saturační mez*.

Na grafu[Obrázek č. 14] máme zaznamenané hodnoty se změnou hodnot pro:

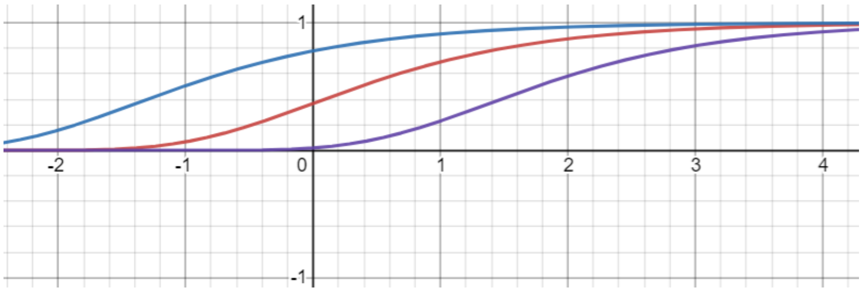
-k = 2 -k = 1 -k = -1/2,

Obrázek č. 10: Grafické znázornění změn hodnoty k u Gompertzovy křivky

Zdroj: vlastní zpracování

Na grafu[Obrázek č. 15] máme zaznamenané hodnoty se změnou hodnot pro:

-α = 1/2 -α = 1 -α = 2

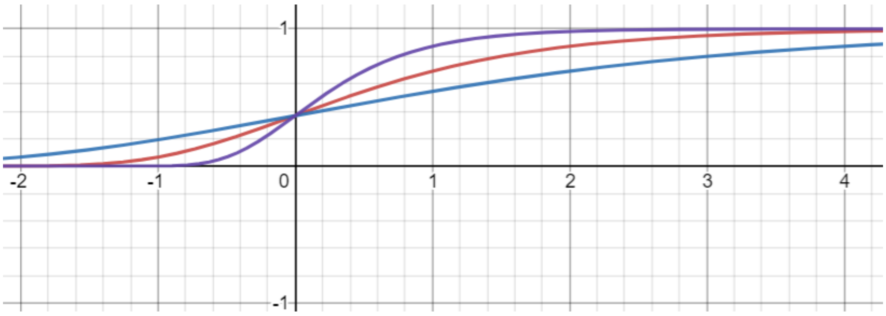


Obrázek č. 11: Grafické znázornění změn hodnoty α u Gompertzovy křivky

Zdroj: vlastní zpracování

Na grafu[Obrázek č. 16] máme zaznamenané hodnoty se změnou hodnot pro:

- β = 1/4 - β = 1 - β = 4



Obrázek č. 12: Grafické znázornění změn hodnoty β u Gompertzovy křivky

Zdroj: vlastní zpracování

2. Logistický regresní model

V této kapitole, bylo čerpáno z následujících zdrojů (HEBÁK 2005), (HENDL 2009), (HOSMER, LEMESHOW 2000), (HUŠEK, PELIKÁN 2003), (CHATTERJEE, HADI, PRICE 2000), (MELOUN, MILITKÝ 2006) a (MELOUN, MILITKÝ 2005).

Logistická regrese je metoda, která se používá v matematické statistice a popisuje případ, kdy vysvětlovaná (závisle) proměnná je binární. Dříve se aplikace logistické regrese objevovala v oblastech medicíny, kdy závisle proměnná představuje přítomnost nebo nepřítomnost choroby.

Rozdíl mezi lineární regresí a logistickou regresí je ten, že sledujeme určitý jev, který buď nastane $P(y=1)=\pi$ nebo nenastane $P(y=0)=1-\pi$. Jelikož hodnota pravděpodobnosti leží mezi čísly 0 a 1, tak hodnota *šance* $\frac{P}{1-P}$ jevu, který zkoumáme, patří do intervalu (0, ∞) a logaritmus *šance* $\ln(\frac{P}{1-P})$ patří do intervalu (-∞, ∞).

Šance, jestli daný jev nastal nebo nenastal, má následující tvar:

$$\frac{P(y=1)}{P(y=0)}=\frac{\pi}{1-\pi}, \tag{2. 1}$$

kde za pomocí logitové transformace vznikne zlogaritmovaná *šance*.

Logit (logitová transformace) má tvar funkce:

$$\ln\frac{\pi}{1-\pi}. \tag{2. 2}$$

Regresní model, který využil logitovou transformaci má následující tvar funkce:

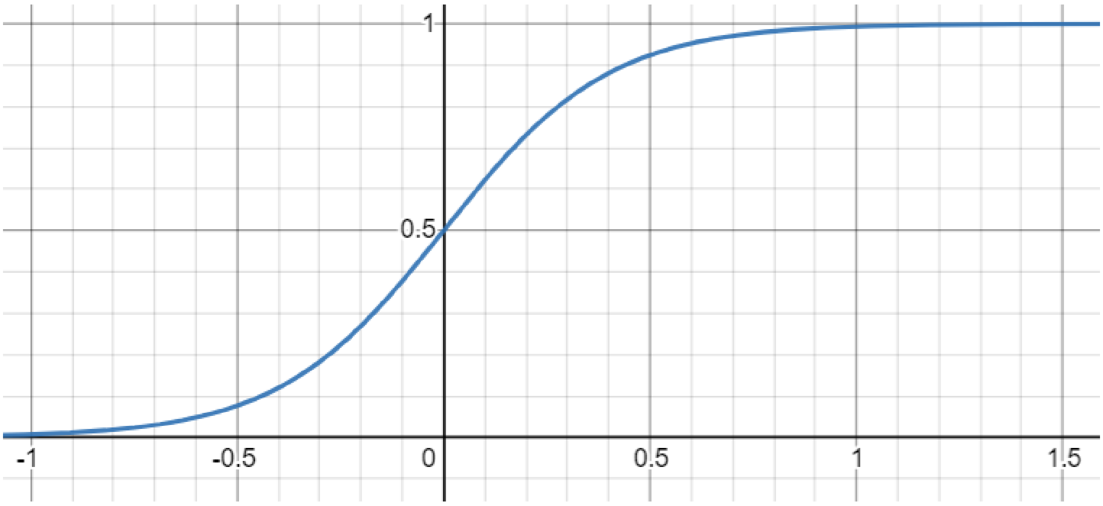
$$\ln\frac{\pi(x)}{1-\pi(x)}=\alpha+\beta x. \tag{2. 3}$$

Pravděpodobnost, že náhodná veličina x bude rovna 1, tedy že určitý jev nastane, má tvar rovnice:

(2. 4)

$$\pi(x) = \frac{e^{(\alpha+\beta x)}}{1 + e^{(\alpha+\beta x)}}.$$

Graf logistické regresní funkce má tvar s-křivky:



Obrázek č. 13: Grafické znázornění logistické regresní funkce

Zdroj: vlastní zpracování

Po uplatnění logitové transformaci má regresní model tvar funkce:

(2. 5)

$$\ln \frac{\pi(x)}{1 - \pi(x)} = \alpha + \beta x.$$

Hodnoty α a β jsou regresní koeficienty. Abychom mohli odhadnout a, b, je za potřebí použití metody nejmenších čtverců, pomocí které získáme maximálně věrohodné odhady pro hodnoty α a β.

Jelikož v logistické regresi porovnáváme, zda se pravděpodobná událost stala $P(y = 1) = \pi$ nebo nestala $P(y = 0) = 1 - \pi$, tak po užití logitové transformaci, kdy porovnáváme událost, která se stala s událostí, která se nestala využitím pravděpodobnostního poměru $\frac{\pi}{1-\pi}$ (také zvaný jako *poměr šancí*), vzniká logistická funkce, která má tvar:

(2. 6)

$$\pi = \frac{1}{1 + e^{C-Z}}.$$

Při modelování více proměnných $x = x_1, x_2, \dots, x_m$, kde odhadované koeficienty $\beta_0, \beta_1, \beta_2, \dots, \beta_m$ jsou míry změny poměru pravděpodobností $\frac{\pi}{1-\pi}$, tak je pravděpodobnostní poměr vyjádřen tvarem:

(2. 7)

$$\frac{\pi}{1-\pi} = e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_m x_m}$$

Poměr šancí je lineární funkcí diskriminační funkce o *m* nezávisle proměnných

(2. 8)

$$Z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_m x_m,$$

kde po zlogaritmování funkce a úpravě vyjde:

(2. 9)

$$C - Z = \ln \frac{1-\pi}{\pi},$$

kde *C* je absolutním členem β_0.

2.1. Volba proměnných

Při volbě proměnných máme většinou k dispozici zadaný soubor nezávisle proměnných, abychom mohli zařadit objet do třídy. Častěji se ale stává, že uživatel předem neví nic o nezávisle proměnných, které by do analýzy měly být zařazeny. Do analýzy se moc nedoporučuje zařazovat velké počty nezávisle proměnných, jsou proměnné nejdřív vyšetřovány na to, která z nich je nejvíce spjata s dichotomní závisle proměnnou.

Samotné vyšetření se pak provádí pomocí obecného χ^2 – testu dobré shody nebo také pomoci Studentova t-testu významnosti jednotlivých parametrů. Abychom zjistili, jestli určitá proměnná zlepší predikci modelu, použijeme krokovou logistickou regresní analýzu. Tyto testy jsou většinou v dopředné krokové analýze postaveny na χ^2-statistice. Když je velká hodnota χ^2 nebo malá vypočítaná hladina významnosti P, tak to znamená, že by nezávisle proměnná měla být přidána do výběru proměnných. Tento test by se ale měl brát spíš jako doporučení, než jako pravý test významnosti.

2.2. Kvalita vyhodnocení logistickou regresí

Pokud máme regresní model, který nám predikuje zařazení objektu do jedné ze dvou tříd, tak je pro nás důležité, jak moc této predikci můžeme věřit. Avšak není nemožné, abychom nemohli získat statisticky významné výsledky. Ty však nemusí jednotlivé objekty správně přiřadit.

Pro to, abychom mohli jednotlivé objekty roztřídit do jednotlivých tříd, tak musíme prvně nalézt tzv., *prahový bod pravděpodobnosti*, který budeme značit jako P_C a jako vizuální informaci o zařazení do jednotlivých tříd, nám poskytne histogram nebo rozptylový diagram. Většinou bývají histogramy dva, kde první slouží pro objekty v „události“ a druhý slouží pro objekty v „neudálosti“.

Po nalezení P_C budeme objekt zařazovat jako v „události“, jejíž pravděpodobnost bude větší nebo rovna té hodnotě P_C. My se zaměříme na graf prahové operační charakteristiky, ve zkratce jen ROC (z anglického Receiver Operating Characteristic).

Pokud chceme znázornit schopnosti regresního modelu klasifikovat správně jednotky podle hodnoty Y, tak lze použít čtyřpolní klasifikační tabulku, na které jsou vyobrazena správná a chybná zařazení jednotek.

Klasifikační tabulka

Tabulka 5: Klasifikační tabulka

		SKUTEČNÁ HODNOTA	
PREDIKOVANÁ HODNOTA		1	0
	1	SP	FP
	0	FN	SN

Zdroj: vlastní zpracování

Následující terminologie, které si za chvíli popíšeme, nám pomůžou v hodnocení kvality modelu:

Skutečně Pozitivní = SPj

Skutečně Negativní = SN

Falešně Pozitivní = FP

Falešně Negativní = FN

Senzitivita = $\frac{SP}{SP+FN}$

Specificita = $\frac{SN}{FP+SN}$

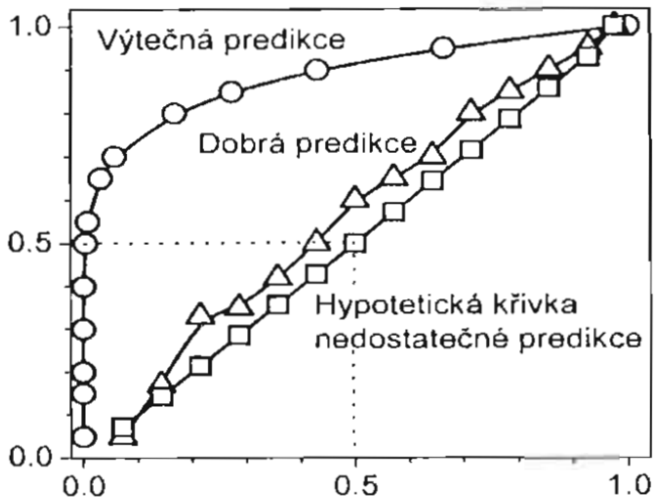
Míra Falešně Pozitivní = $\frac{FP}{FP+SN}$

Prevalence = $\frac{SP+FN}{SP+FN+FP+SN}$

Prediktivní hodnota pozitivního testu = $\frac{SP}{SP+FP}$

Prediktivní hodnota negativního testu = $\frac{SN}{FN+SN}$

Ukázku, jak může vypadat výsledná ROC křivka máme na následujícím obrázku.



Obrázek č. 14: Grafické znázornění ROC křivky

Zdroj: MELOUN, MILITKÝ 2005

3. Logistický růstový model, jeho využití a interpretace

V kapitole bylo čerpáno z následujících zdrojů (HUŠEK, PELIKÁN 2003), (CHATTERJEE, HADI, PRICE 2000), (MELOUN, MILITKÝ 2006), (MELOUN, MILITKÝ 2005),

3.1. Aplikace logistické regrese

Použití modelu vícenásobné logistické regrese má své využití nejčastěji k odhadnutí pravděpodobnosti určitě události, která nastala. Daná událost, která nastala, může být například výskyt viru, infarktu, smrt na určitou nemoc atd. Jestliže některá z těchto událostí nastanou, musí být vyznačeny v časovém období, ve kterém se odehráli. Například sledování nemoci v průběhu několika let.

Prvně abychom mohli takový regresní model určit, potřebujeme mít určitý výběr dat, ve kterém byly objekty sledovány za určitě časové období a hodnoty závažných proměnných byly zaznamenávány už od začátku sledování. Výběr můžeme uskutečnit a použít na to dva způsoby:

1. Výběr se získal náhodným způsobem a pozorování bylo provedeno za určité časové období. Tento výběr se potom nazývá jako *výběr cross validation*. Prvním pod výběr je ten, ve kterém je nejvíce zkušeností co se týče události, kde se vyčíslí logistický regresní model, který se potom může aplikovat na člen druhého pod výběru, a druhým pod výběr obsahuje veškeré údaje, které zbyly. Aby aplikace mohla fungovat správně, tak v původním výběru se nesmí objevovat podstatné změny, aby se nezměnil vztah mezi nezávisle proměnnými a výskytem události.
2. Za použití druhého způsobu se výběr získá, když dostaneme dva náhodné výběry. Jeden z výběrů je ten, ve kterém se daná událost objeví a ve druhém se událost neobjeví. Tomuto způsobu se říká *případ řídícího výběru*. Hodnoty předpovídaných proměnných se získají pomocí retrospektivního způsobu, což znamená např. ze vzpomínek nebo minulých záznamů. Data, která se získají tak poslouží k tomu, aby se odhadnul logistický regresní model. Tato metoda má oproti předchozí metodě výhodu v tom, že dokáže specifikovat počet objektů, jedinců v „události“ a „neudálosti“. Pokud chceme aplikovat model, tak regresní koeficienty $b_1, b_2, \dots, b_m$ jsou platné a konstanta a musí vyjadřovat

skutečný poměr objektu v dané události. Aby to mohlo nastat, musí se vyhodnotit pravděpodobnost π události v tzv. *cílovém souboru*. Nastavení pro konstantu a^* má provedení podle vzorce, který je:

$$a^* = a + \ln\left[\frac{\pi n_0}{(1-\pi)n_1}\right]. \tag{3. 1}$$

V tomto vzorci je pro nás a konstantou, π je odhad pravděpodobnosti π události v konečném souboru, n_0 označuje počet objektů, které jsou ve výběru, ale událost pro ně nenastala a n_1 označuje počet objektů, které jsou ve výběru a pro něž se daná událost v konečném souboru stala.

4. Praktické příklady použití logistické regrese

V této kapitole si ukážeme odhad parametrů Gompertzovy regrese a logistické regrese, kde za použití tří metod a úpravě vzorců Gompertzovy regrese a logistické regrese na patřičný exponenciální tvar získáme jednotlivé parametry.

4.1. Odhad parametrů Gompertzovy regrese

Jak jsem se již zmínil v úvodu, budeme používat tří metod pro odhad parametrů. Abychom však získali správné parametry, musíme funkci upravit na požadovaný tvar, a to zlogaritmováním její původní rovnice do nové, kde jednotlivé kroky a postup jsme si vysvětlili a popsali v podkapitole 1.1.3.2.

Jako první krok provedeme zlogaritmování hodnot y_i, kde na následujícím obrázku můžeme vidět výsledné hodnoty $\ln y_i$

Tabulka 6 Hodnoty pro odhad parametrů Gompertzovy regrese

xi	1	2	3	4	5	6	7	8	9	10	11	12
yi	4	6	8	12	17	21	23	26	30	35	40	48
ln yi	1,3863	1,7918	2,0794	2,4849	2,8332	3,0445	3,1355	3,2581	3,4012	3,5553	3,6889	3,8712
xi	13	14	15	16	17	18	19	20	21	22	23	
yi	51	56	59	62	67	72	75	77	80	86	90	
ln yi	3,9318	4,0254	4,0775	4,1271	4,2047	4,2767	4,3175	4,3438	4,3820	4,4543	4,4998	

Zdroj: vlastní zpracování

Metoda částečných součtů

Další kroky, které provedeme, budou výpočty jednotlivých parametrů β, α a k. Kde po dosazení do vzorců získáme následující hodnoty koeficientů.

$b = \sqrt[7]{\frac{30{,}4788-27{,}2773}{27{,}2773-20{,}2369}} = 0{,}8935,$

$\ln a = \frac{(0{,}8935-1)(27{,}2773-20{,}2369)}{0{,}8935(0{,}8935^7-1)^2} = -2{,}8217,$ kde po úpravě získáme koeficient o hodnotě $a = 0{,}0595,$

$\ln k = \frac{1}{7}\left(20{,}2369 - 2{,}8217 * 0{,}8935\frac{0{,}8935^7-1}{0{,}8935-1}\right) = 4{,}7356$ a po úpravě máme hodnotu $k = 113{,}9269.$

Funkce, kterou jsme získali pomocí této metody, má tvar:

$$y = 113{,}9269 * 0{,}0595^{0{,}8935^x}$$

Metoda dílčích průměrů

V této metodě bude menší rozdíl ve vypočítávání dvou koeficientů a to a a k. Právě pro tyto dva koeficienty budeme muset použít metodu nejmenších čtverců. Prvně ale začneme výpočtem koeficientu b, kde dosadíme do vzorce a získáme

$b = \sqrt[\frac{23-5}{2}]{\frac{4{,}3257-3{,}7262}{3{,}7262-2{,}3937}} = 0{,}9151.$ Jak již bylo zmíněno, pro zbylé dva koeficienty budeme muset použít metodu nejmenších čtverců, kdy podle upravené rovnice $\ln y = \ln k + \beta^x \ln \alpha$ se za y dosadí hodnoty $\ln y$ a za x se dosadí hodnoty β^x.Za pomocí regrese, kterou dostaneme tak, že v programu MS Excelu si v kartě Data ►Analýza dat ►Regrese, dostaneme požadované hodnoty koeficientů α, β naší upravené funkce a podle vztahů $\alpha = \ln k$ a $\beta = \ln a$ získáme a = 0,0231 a k = 158,3181.

Funkce, kterou jsme pomocí této metody získali, má tvar:

$$y = 158{,}3181 * 0{,}0231^{0{,}9151^x}$$

Metoda vybraných bodů

Poslední metoda se liší v tom, že si musíme zvolit hodnoty pro x a pro m. Hodnota pro x nejčastěji bývá rovna 1 a m je přirozené číslo. My máme hodnoty zvolené následovně: x =1, m = 11. Nyní se můžeme pustit do výpočtu koeficientů:

$$b = \sqrt[11]{\frac{4{,}4998-3{,}8712}{3{,}8712-1{,}3863}} = 0{,}8825,$$

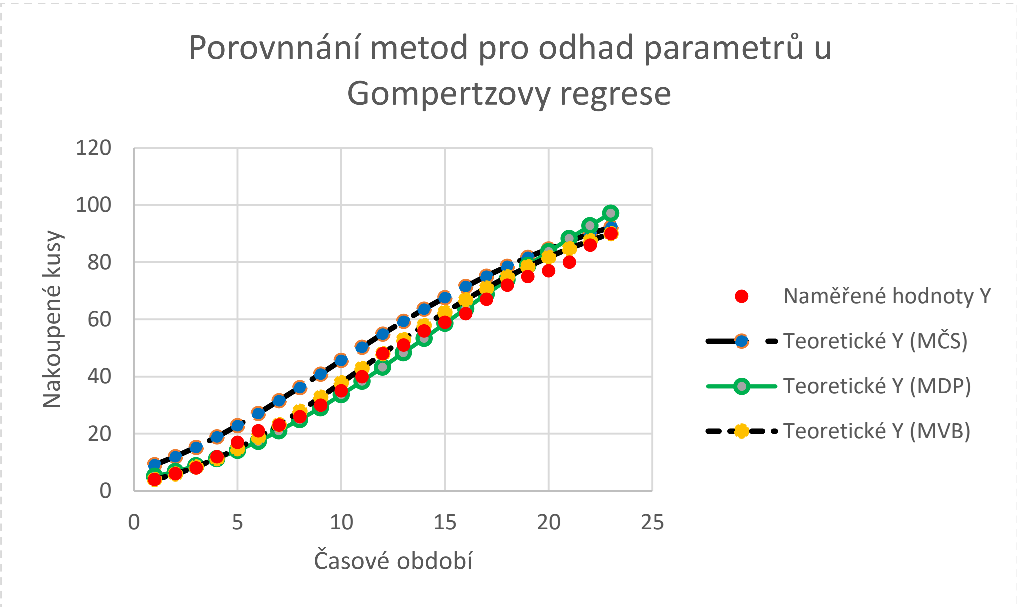
$\ln a = \frac{3{,}8712-1{,}3863}{0{,}8825^1(0{,}8825^{11}-1)} = -3{,}7691$, po úpravě dostaneme a = 0,02307

$\ln k = 1{,}3863 - (-3{,}7691) * 0{,}8825^1 = 4{,}7127$, po úpravě dostaneme k = 111,35

Funkce, kterou jsme pomocí této metody získali, má tvar:

$$y = 111{,}35 * 0{,}02307^{0{,}8825^x}.$$

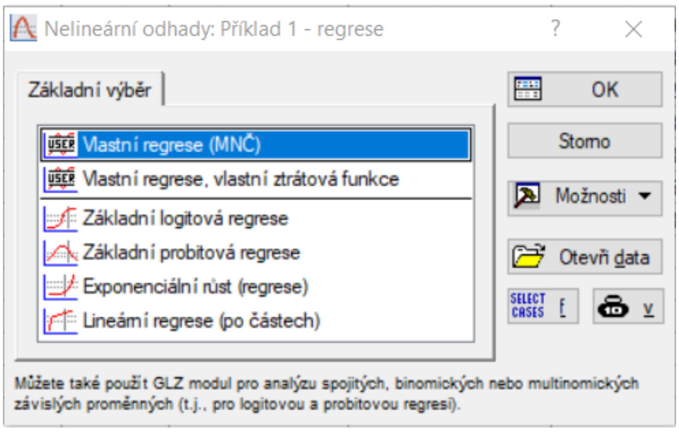
Na grafu si můžeme ukázat a porovnat rozdíly mezi funkcemi jednotlivých metod.



Obrázek č. 15 Graf Porovnání metod pro odhad parametrů u Gompertovy regrese

Zdroj: vlastní zpracování

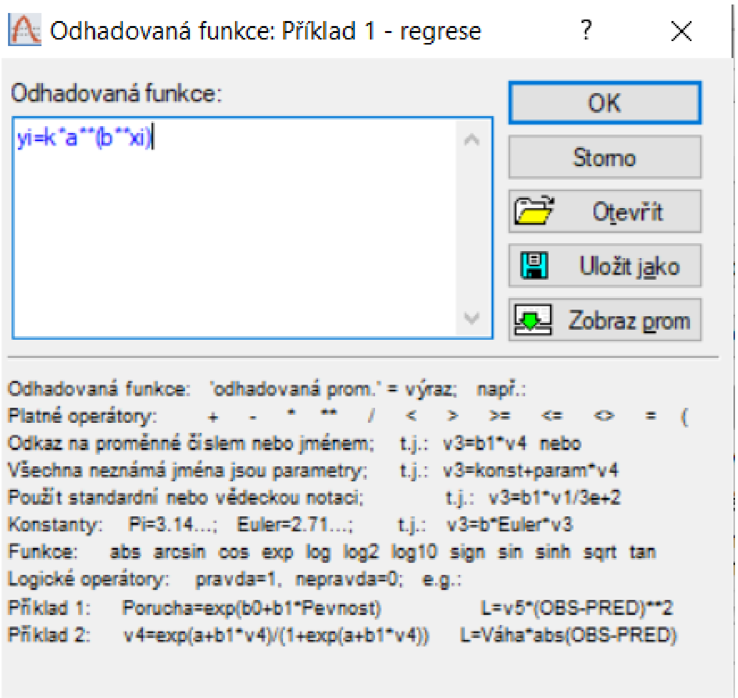
Pokud bychom chtěli upřesnit odhad parametrů na daná data, která máme zadaná je použití programu STATISTICA 12. Nyní si ukážeme jednotlivé kroky po tom, co máme připravená data v potřebném programu. Přes kartu Statistika ►Pokročilé lineární/nelineární modely ►Nelineární odhady zvolíme Vlastní regrese (MNČ) a klikneme na tlačítko OK.



Obrázek č. 16 STATISTICA 12 Nelineární odhady

Zdroj: vlastní zpracování

Dále musíme ručně zadat patřičný vzorec pro odhad parametrů tak, jak jsme si zvolili zápis (pojmenování) jednotlivých proměnných. Poté zvolíme OK.



Obrázek č. 17 STATISTICA 12 Odhadovaná funkce Gompertzovy regrese

Zdroj: vlastní zpracování

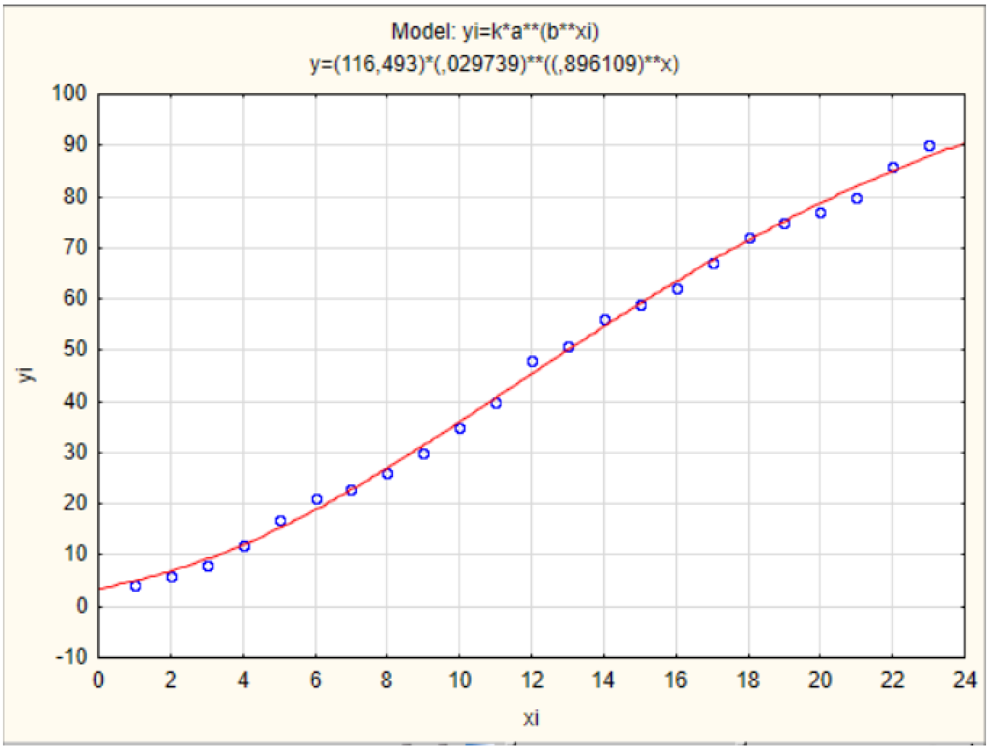
Jako další krok bude dobré si zvolit počáteční body pro parametry a = 0,02, b = 0,8 a k = 113.

	Odhad	Standard chyba	t-hodn. sv = 20	p-hodn.	Dol. sp. Mez	Hor. sp. Mez
k	116,4928	4,235766	27,5022	0,000000	107,6571	125,3284
a	0,0297	0,002776	10,7134	0,000000	0,0239	0,0355
b	0,8961	0,004775	187,6668	0,000000	0,8861	0,9061

Obrázek č. 18 STATISTICA 12 Odhad parametrů Gompertzovy regrese

Zdroj: vlastní zpracování

Graf funkce má následující tvar.



Obrázek č. 19 STATISTICA 12 Model grafu Gompertzovy regrese

Zdroj: vlastní zpracování

4.2. Odhad parametrů logistické regrese

Pro odhad parametrů logistické regrese budeme také vycházet z podkapitoly 1.1.3.2, ale funkci upravíme do lomeného tvaru, tedy na tvar:

$$\frac{1}{y} = \frac{1}{k} + \frac{\alpha}{k}\beta^x.$$

Prvně si musím hodnoty pro y upravit do potřebného tvaru, tedy na $\frac{1}{y}$.

Tabulka 7 Hodnoty pro odhad parametrů logistické regrese

xi	1	2	3	4	5	6	7	8	9	10	11	12
yi	2,564	5,541	9,521	16,482	24,621	29,965	34,632	40,496	50,513	64,263	72,528	83,651
1/y1	0,3900	0,1805	0,1050	0,0607	0,0406	0,0334	0,0289	0,0247	0,0198	0,0156	0,0138	0,0120

Zdroj: vlastní zpracování

Jednotlivé postupy budou stejné, jako bylo u Gompertzovy regrese s tím rozdílem, že máme jinak upravenou funkci.

Metoda částečných součtů

Opět si dosadíme do vzorců, abychom získali odhad pro jednotlivé parametry *β*, *α* a *k*, kde $b = \sqrt[4]{\frac{0{,}0611-0{,}1276}{0{,}1276-0{,}7362}} = 0{,}5748$, koeficient *α* získáme dosazením do následujícího vzorce $\frac{a}{k} = \frac{(0{,}5748-1)(0{,}1276-0{,}7362)}{0{,}5748(0{,}5748^4-1)^2} = 0{,}5673$, po vynásobení koeficientem *k*, dostaneme hodnotu *a* = 42,8503. Po dosazení pro získání koeficientu *k*, získáme $\frac{1}{k} = \frac{1}{4}\left(0{,}7362 - 0{,}5673 * 0{,}5748\,\frac{0{,}5748^4-1}{0{,}5748-1}\right) = 0{,}0132$ a po převrácení hodnoty získáme *k* = 75,5368.

Funkce získaná pomocí této metody má tvar:

$$y = \frac{75{,}537}{1 + 42{,}850 * 0{,}437^x}$$

Metoda dílčích průměrů

Postup je stejný jako u Gompertzovy regrese, tedy dosadíme do vzorce a získáme $b = \sqrt[\frac{12-5}{2}]{\frac{0{,}0153-0{,}0319}{0{,}0319-0{,}1840}} = 0{,}5311$. Pro získání zbylých dvou koeficientů se musí opět provést metoda nejmenších čtverců, kdy v tomto případě za *y* dosadíme hodnoty $\frac{1}{y}$ a za *x* se dosadí hodnoty β^x. Získání regrese má stejný postup, tedy že v programu MS Excelu si v kartě Data ►Analýza dat ►Regrese, dostaneme požadované hodnoty koeficientů *α*, *β* naší upravené funkce a podle vztahů $\alpha = \frac{1}{k}$ a $\beta = \frac{a}{k}$ získáme *a* = 52,4703 a *k* = 77,2083.

Funkce získaná pomocí této metody má tvar:

$$y = \frac{77{,}2083}{1 + 52{,}4703 * 0{,}5311^x}$$

Metoda vybraných bodů

Hodnota pro *x* zůstává stejná, tedy *x* = 1, ale hodnota pro *m* je v pro tento příklad jiná a to *m* = 5. Postup pro získání odhad parametrů zůstává stejný:

$$b = \sqrt[5]{\frac{0{,}0138-0{,}0334}{0{,}0334-0{,}3900}} = 0{,}5597,$$

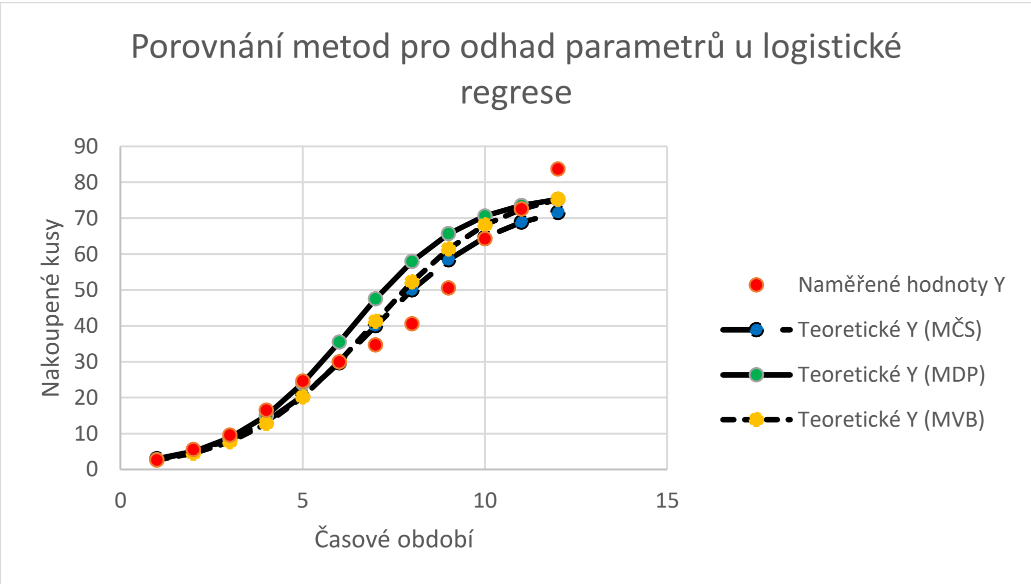
$\frac{a}{k} = \frac{0{,}0334-0{,}3900}{0{,}5597^1(0{,}5597^5-1)} = 0{,}6743$, po úpravě dostaneme *a* = 53,3018

$\frac{1}{k} = 0{,}3900 - 0{,}6743 * 0{,}5597^1 = 0{,}0127$, po úpravě dostaneme *k* = 79,0524

Funkce získaná pomocí této metody má tvar:

$$y = \frac{79{,}0524}{1 + 53{,}3018 * 0{,}5597^x}$$

Na grafu si můžeme ukázat a porovnat rozdíly mezi funkcemi jednotlivých metod.

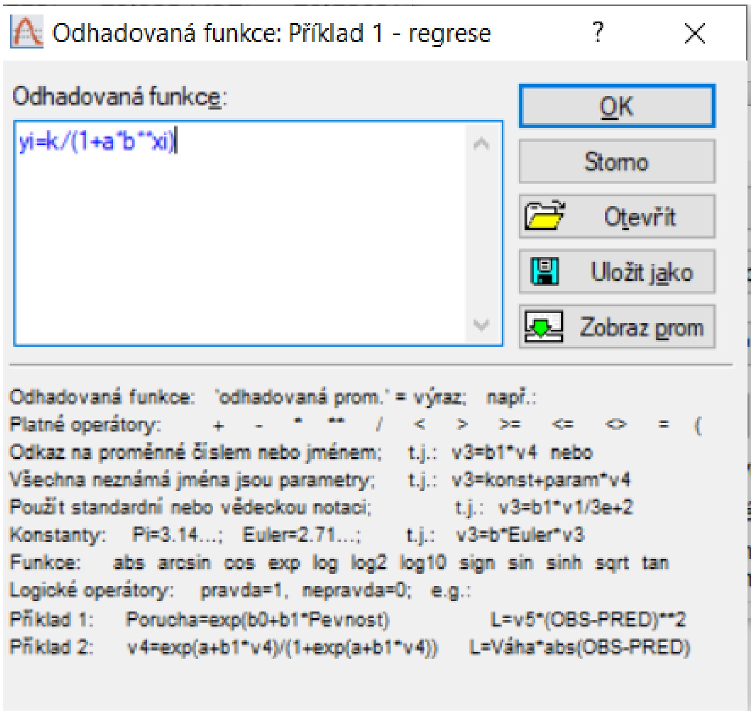


Obrázek č. 20 Graf Porovnání metod pro odhad parametrů u logistické regrese

Zdroj: vlastní zpracování

Pokud bychom chtěli upřesnit odhad parametrů na daná data, která máme zadaná je použití programu STATISTICA 12. Postup bude stejný, jako bylo u Gompertzovy regrese, tedy přes kartu Statistika ►Pokročilé lineární/nelineární modely ►Nelineární odhady zvolíme Vlastní regrese (MNČ) a klikneme na tlačítko OK.

Ručně se zadá vzorec a zvolíme OK.



Obrázek č. 21 STATISTICA 12 Odhadované funkce logistické regrese

Zdroj: vlastní zpracování

Jako další krok bude, že si zvolíme počáteční body pro parametry a = 42, b = 0,5 a k = 75.

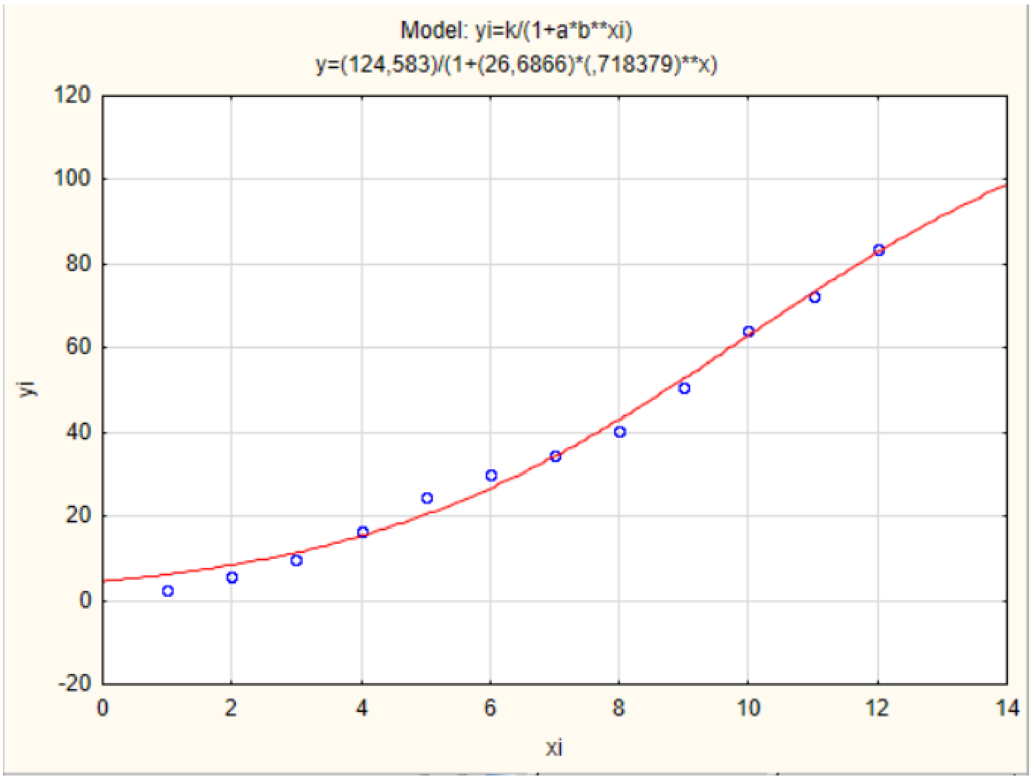
Parametry pro náš odhad parametrů logistické regrese vypadají následovně.

	Odhad	Standard chyba	t-hodn. sv = 9	p-hodn.	Dol. sp. Mez	Hor. sp. Mez
k	124,5826	19,63629	6,34451	0,000134	80,16220	169,0029
a	26,6866	3,51096	7,60096	0,000033	18,74431	34,6290
b	0,7184	0,02665	26,96081	0,000000	0,65810	0,7787

Obrázek č. 22 STATISTICA 12 Odhad parametrů logistické regrese

Zdroj: vlastní zpracování

Graf funkce má následující tvar



Obrázek č. 23 STATISTICA 12 Model grafu logistické regrese

Zdroj: vlastní zpracování

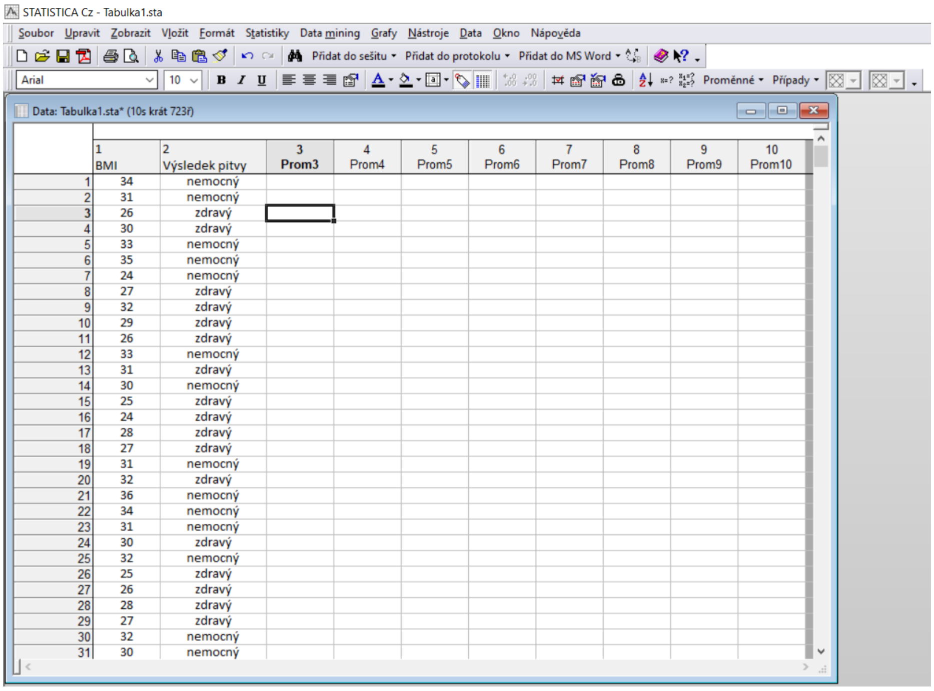
4.3. Příklad využití logistického regresního modelu

Jak už bylo zmíněno ve druhé kapitole, logistická regrese se nejčastěji objevovala v oblastech medicíny. Veškeré příklady, které si ukážeme, budou prováděný v softwaru STATISTICA 12 a ukážeme si jednotlivé postupy, jak se dopracovat ke grafickému znázornění logistické křivky.

Jako příklad praktického využití logistické regrese jsem si zvolil pitvu těl, která byla podle hodnot BMI buď nemocná, nebo zdravá. Pro tento příklad máme 723 předpovězených hodnot.

V softwaru STATISTICA 12 máme připravená naměřená data v Tabulce.

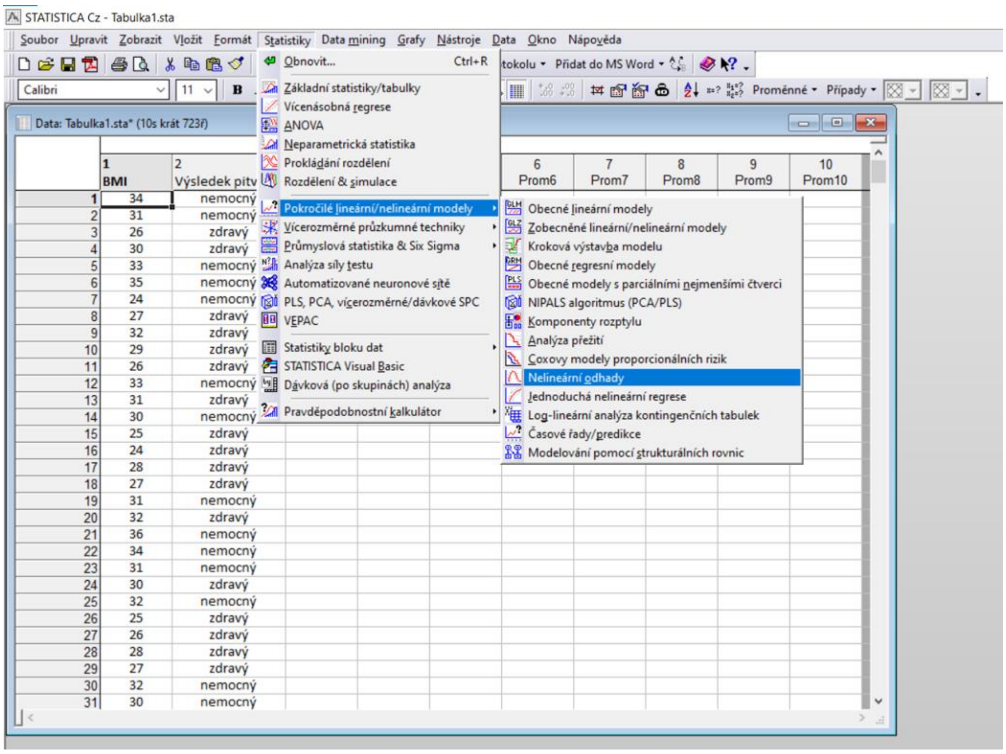
Tabulka 8 STATISTICA 12 Data s hodnotami



	1 BMI	2 Výsledek pitvy
1	34	nemocný
2	31	nemocný
3	26	zdravý
4	30	zdravý
5	33	nemocný
6	35	nemocný
7	24	nemocný
8	27	zdravý
9	32	zdravý
10	29	zdravý
11	26	zdravý
12	33	nemocný
13	31	zdravý
14	30	nemocný
15	25	zdravý
16	24	zdravý
17	28	zdravý
18	27	zdravý
19	31	nemocný
20	32	zdravý
21	36	nemocný
22	34	nemocný
23	31	nemocný
24	30	zdravý
25	32	nemocný
26	25	zdravý
27	26	zdravý
28	28	zdravý
29	27	zdravý
30	32	nemocný
31	30	nemocný

Zdroj: vlastní zpracování

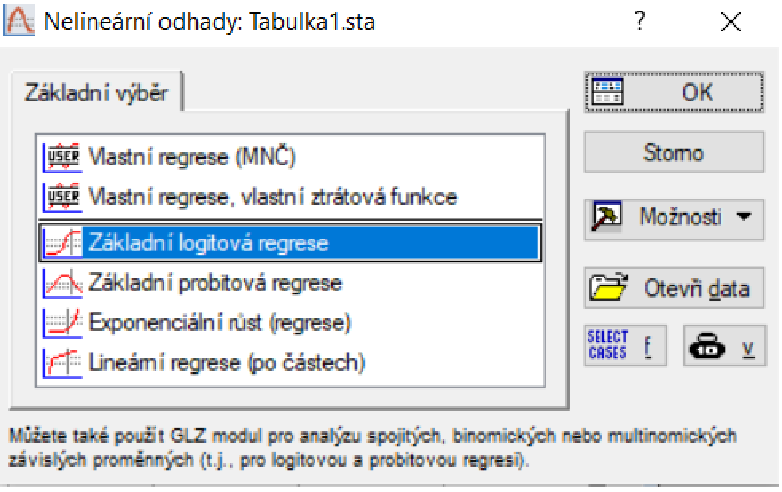
Abychom mohli z dat získat potřebou logistickou funkci, zvolíme si v menu Statistiky ►Pokročilé lineární/nelineární modely ►Nelineární odhady.



Obrázek č. 24 STATISTICA 12 Zvolení správného odhadu

Zdroj: vlastní zpracování

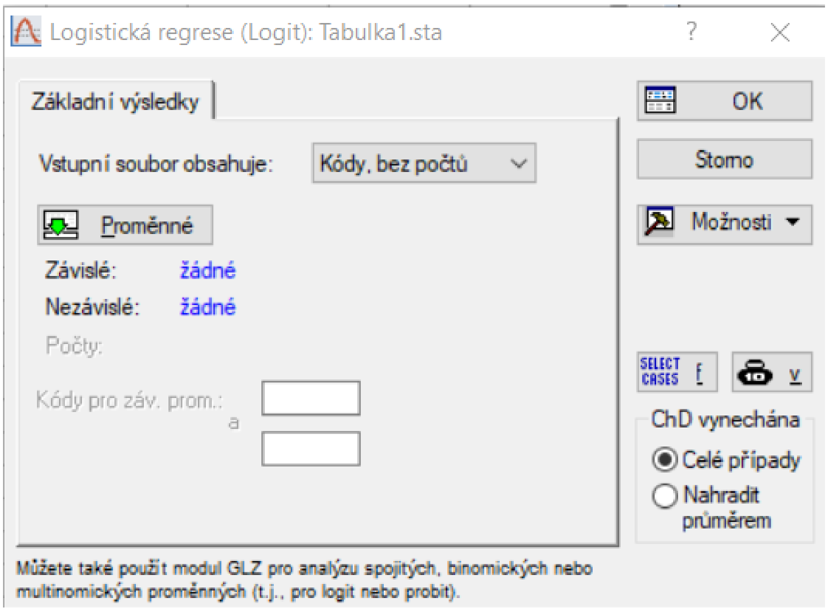
Poté se objeví okénko, kde zvolíme Základní logitovou regresi a klikneme na tlačítko OK.



Obrázek č. 25 STATISTICA 12 Nelineární odhady – Základní logitová regrese

Zdroj: vlastní zpracování

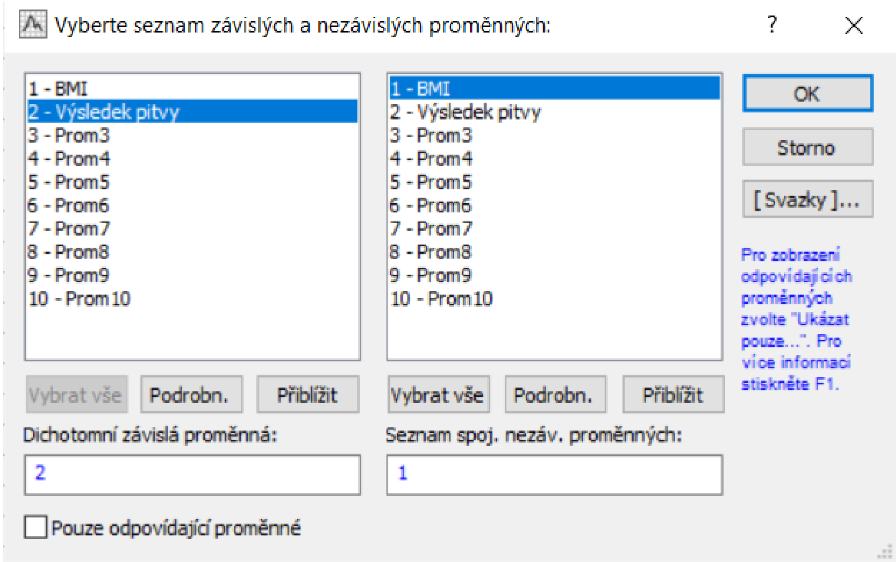
Po zvolení základní logitové regrese se ukáže tabulka, kde si zvolíme Proměnné,



Obrázek č. 26 STATISTICA 12 Logistická regrese (logit)

Zdroj: vlastní zpracování

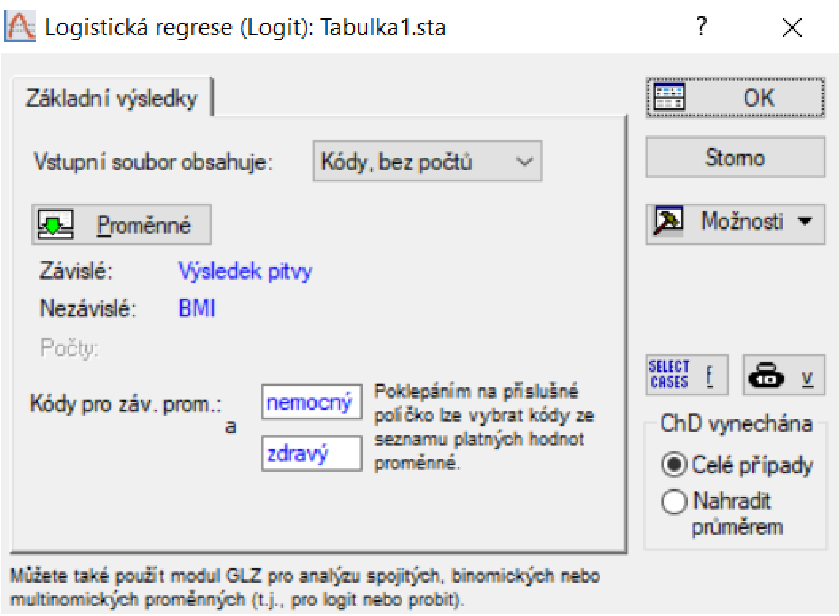
kde závislá proměnná je Výsledek pitvy a nezávisle proměnné jsou hodnoty BMI. Po zvolení proměnných opět klikneme na tlačítko OK.



Obrázek č. 27 STATISTICA 12 Volba proměnných

Zdroj: vlastní zpracování

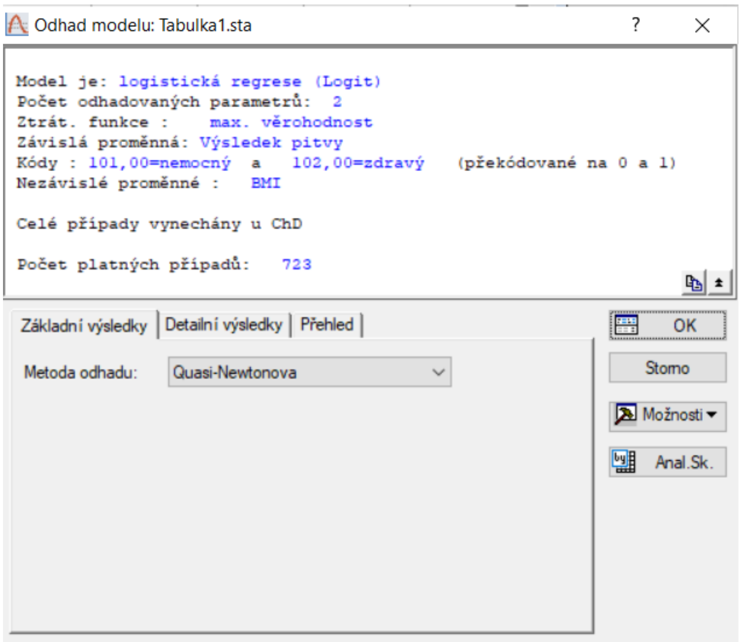
Jelikož máme navolené proměnné, můžeme opět kliknout na tlačítko OK,



Obrázek č. 28 STATISTICA 12 Logistická regrese (logit) – potvrzení proměnných

Zdroj: vlastní zpracování

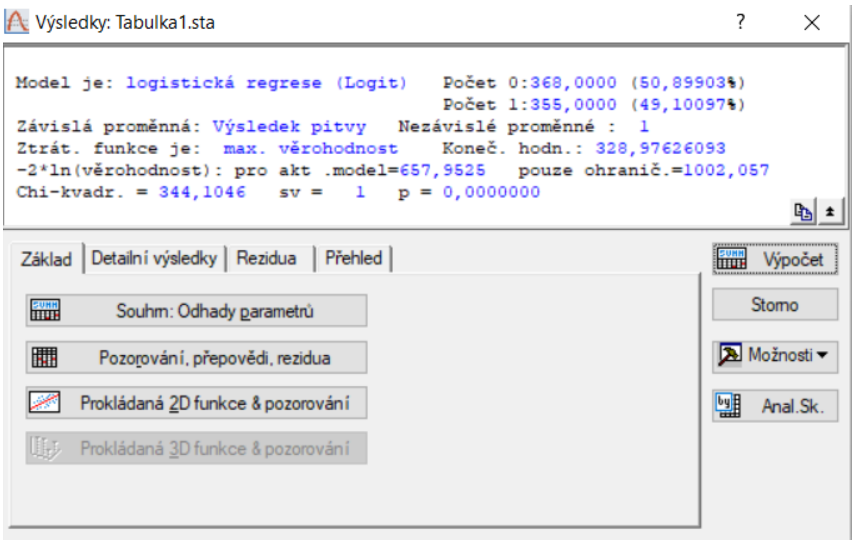
kde se nám ukáže tabulka Odhadu modelu



Obrázek č. 29 STATISTICA 12 Odhad modelu logistické regrese

Zdroj: vlastní zpracování

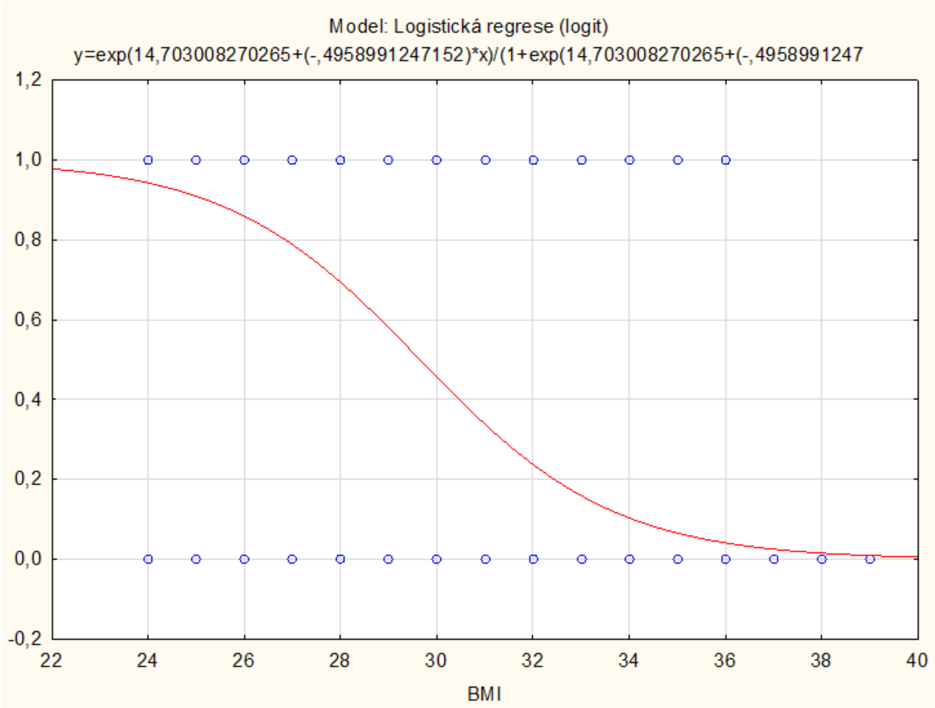
a pro detailnější údaje klikneme na tlačítko OK, které nás přesunou do výsledků, kde se můžeme podívat na jednotlivé výsledky.



Obrázek č. 30 STSTISTICA 12 Výsledky logistické regrese

Zdroj: vlastní zpracování

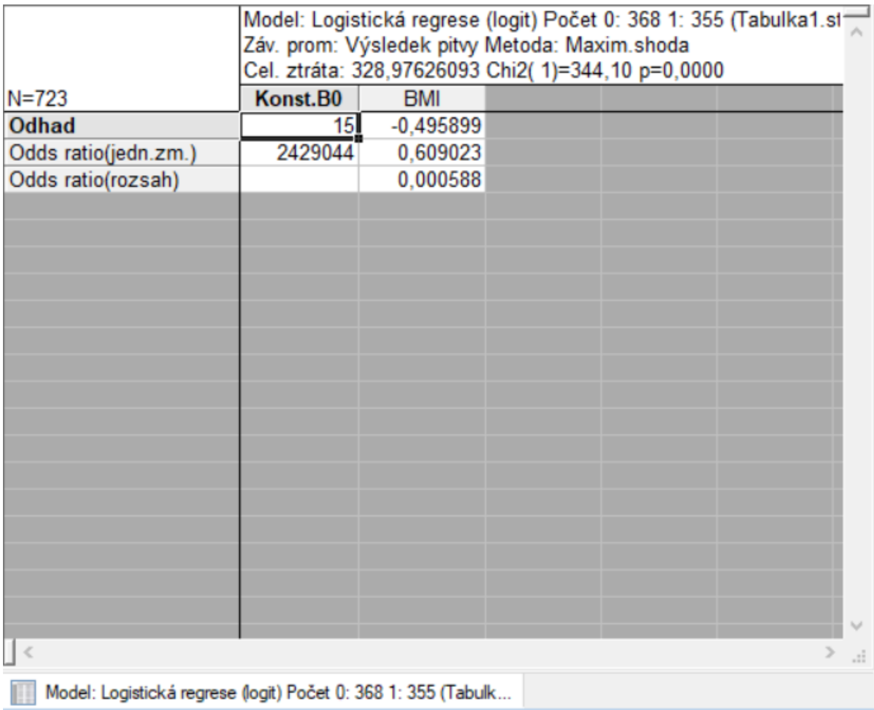
Nás bude převážně zajímat grafické znázornění logistického modelu pro daná data.



Obrázek č. 31 STSTISTICA 12 Grafické znázornění logistického modelu

Zdroj: vlastní zpracování

Dále co můžeme získat, je odhad parametrů logistické regrese.



Model: Logistická regrese (logit) Počet 0: 368 1: 355 (Tabulka1.st
Záv. prom: Výsledek pitvy Metoda: Maxim.shoda
Cel. ztráta: 328,97626093 Chi2(1)=344,10 p=0,0000

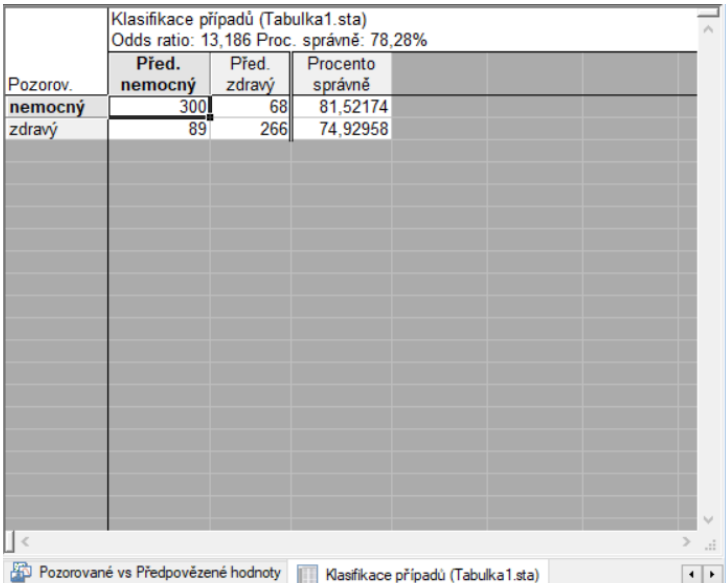
N=723	Konst.B0	BMI
Odhad	15	-0,495899
Odds ratio(jedn.zm.)	2429044	0,609023
Odds ratio(rozsah)		0,000588

Model: Logistická regrese (logit) Počet 0: 368 1: 355 (Tabulk...

Obrázek č. 32 STATISTICA 12 Odhad parametrů logistiké regrese

Zdroj: vlastní zpracování

Co také můžeme zjistit, tak je Klasifikace případů, jinak řečeno, že z pozorovaných hodnot, jako například pitva, u které vyšlo, že bylo tělo nemocné, můžeme na následujícím obrázku vidět, že z celkem naměřených 368 nemocných bylo správně zařazených 300 a nesprávně pouze 68. Kde u zdravých z celkem naměřených 355 bylo správně zařazených 266 a 89 špatně.



Klasifikace případů (Tabulka1.sta)
Odds ratio: 13,186 Proc. správně: 78,28%

Pozorov.	Před. nemocný	Před. zdravý	Procento správně
nemocný	300	68	81,52174
zdravý	89	266	74,92958

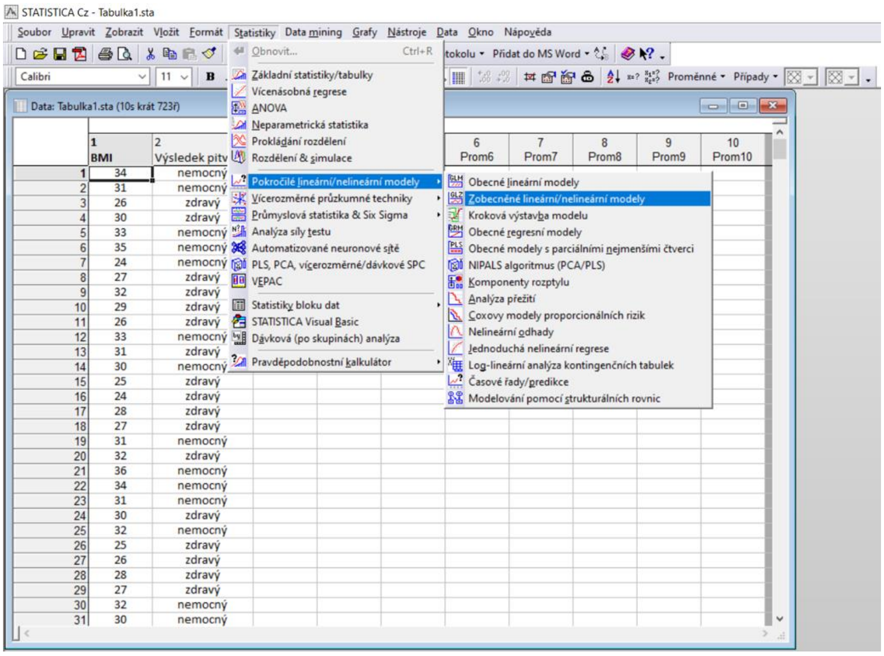
Pozorované vs Předpovězené hodnoty
Klasifikace případů (Tabulka1.sta)

Obrázek č. 33 STSTISTICA 12 Klasifikace případů

Zdroj: vlastní zpracování

Software STATISTICA je také schopný vygenerovat na základě našich dat ROC křivku. Na následujících krocích si ukážeme, jak se dostat k výpočtu a grafu ROC křivky.

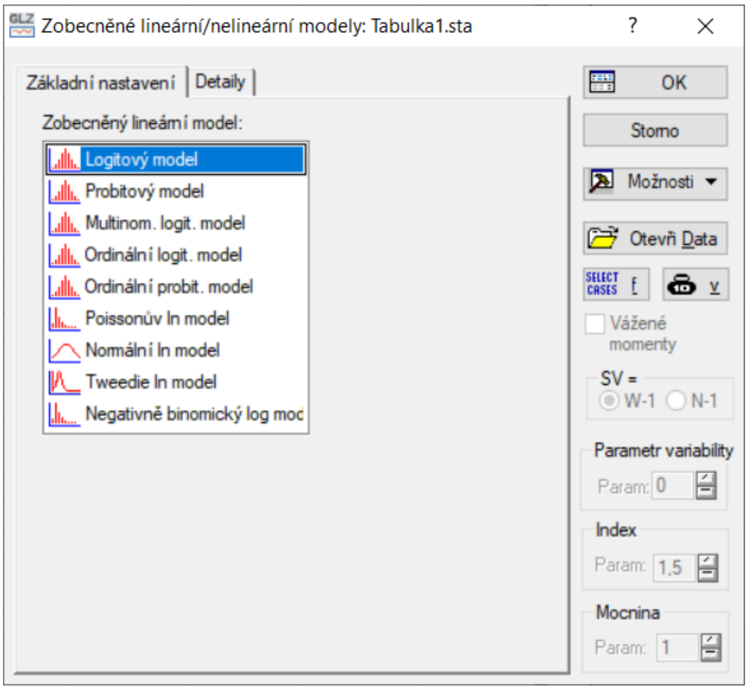
Na horním menu si zvolíme Statistiky ►Pokročilé lineární/nelineární modely ►Zobecněné lineární/nelineární modely.



Obrázek č. 34 STSTISTICA 12 Zvolení správného modelu pro ROC křivku¨

Zdroj: vlastní zpracování

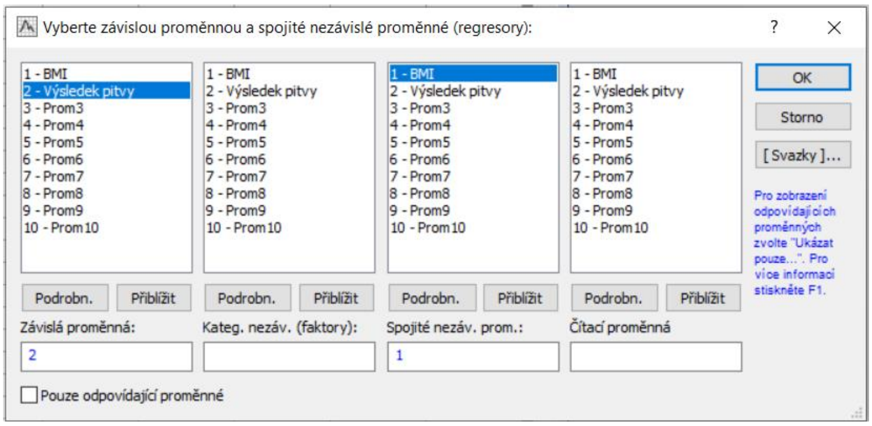
Poté se objeví tabulka, kde se vybere Logitový model a potvrdí se tlačítkem OK.



Obrázek č. 35 STSTISTICA 12 Volba logitového modelu

Zdroj: vlastní zpracování

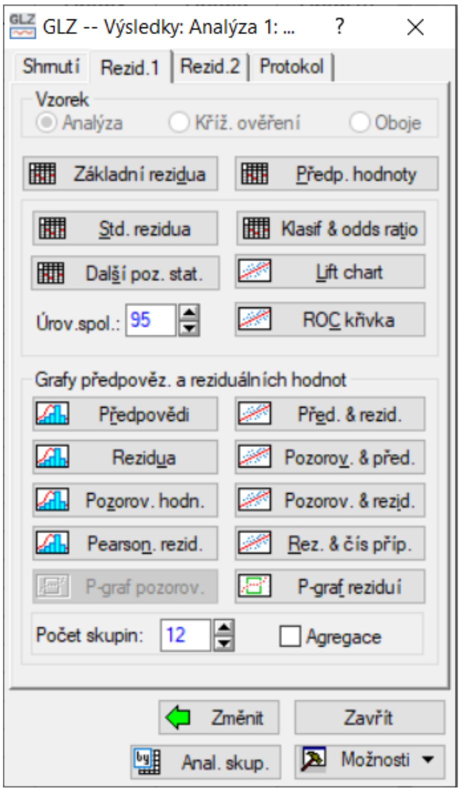
Na následující tabulce se vyberou proměnné, kde závisle proměnná je Výsledek pitvy a Spojité nezávisle proměnné je hodnota BMI.



Obrázek č. 36 STSTISTICA 12 Volba proměnných pro ROC křivku

Zdroj: vlastní zpracování

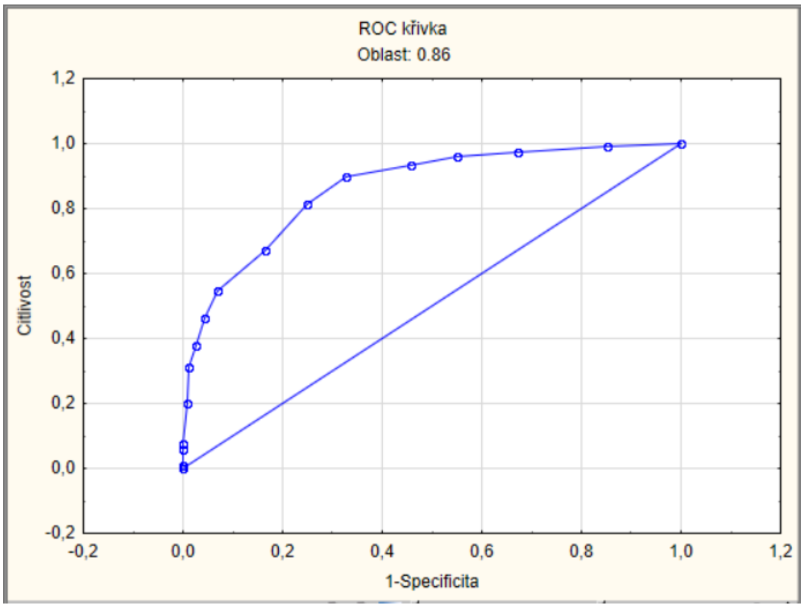
Po potvrzení se objeví tabulka, kde se vybere Rezid.1



Obrázek č. 37 STSTISTICA 12 Výsledky analýzy

Zdroj: vlastní zpracování

a zvolí se ROC křivka, kde se ukáže výsledný graf.



Obrázek č. 38 STSTISTICA 12 Grafické znázornění ROC křivky

Zdroj: vlastní zpracování

ZÁVĚR

Cílem bakalářské práce bylo popsat reálné úkoly, které vedou na logistický růstový model. Prvně se v práci popsala potřebná teorie a ukázkové příklady. V dalších kapitolách byla popsána logistická regrese, její využití a aplikace. Ve čtvrté kapitole bylo ukázáno, jak přes zobecněnou exponenciální rovnici můžeme pomocí užití správné transformace Gompertzovy regrese a logistické regrese získat odhad parametrů. Poté použitím programu STATISTICA 12 získat přesnější odhad parametrů pro obě regrese.

Ve využití logistického regresního modelu se zkoumala pravděpodobnost zdraví člověka na základě jeho hodnoty BMI. Dále bylo na příkladu pitvy těl znázorněno, jak moc velký vliv má hodnota BMI na zdraví člověka. Zjistilo se, že čím větší je hodnota BMI, tím větší je šance, že bude člověk nemocný a obráceně, čím menší je hodnota BMI, tím větší je šance, že bude člověk zdraví. Podle grafického znázornění logistického regresního modelu je vidět, že střední hodnotou BMI 31. Poté bylo pomocí klasifikační tabulky upřesněno, kolik bylo správně a špatně zaznačených hodnot.

POUŽITÁ LITERATURA

HEBÁK, Petr. Vícerozměrné statistické metody. [3]. Praha: Informatorium, 2005. ISBN 80-7333-039-3.

HENDL, Jan. Přehled statistických metod: analýza a metaanalýza dat. 3., přeprac. vyd. Praha: Portál, 2009. ISBN 978-80-7367-482-3.

HOSMER, David W. a Stanley LEMESHOW. Applied logistic regression. 2nd ed. New York: John Wiley, 2000. Wiley series in probability and statistics. ISBN 0-471-35632-8.

HUŠEK, Roman a Jan PELIKÁN. Aplikovaná ekonometrie: teorie a praxe. Praha: Professional Publishing, 2003. ISBN 80-86419-29-0.

CHATTERJEE, Samprit, Ali S. HADI a Bertram PRICE. Regression analysis by example. 3rd ed. New York: John Wiley, c2000. Wiley series in probability and statistics. ISBN 0-471-31946-5.

KUBANOVÁ, Jana. Statistické metody pro ekonomickou a technickou praxi. Bratislava: Statis, 2003. ISBN 80-85659-31-X.

KUBANOVÁ, Jana. Statistické metody pro ekonomickou a technickou praxi. Vyd. 3., dopl. Bratislava: Statis, 2008, 247 s. ISBN 978-80-85659-47-4.

MELOUN, Milan a Jiří MILITKÝ. Kompendium statistického zpracování dat: metody a řešené úlohy. Vyd. 2., přeprac. a rozš. Praha: Academia, 2006. ISBN 80-200-1396-2.

MELOUN, Milan, Jiří MILITKÝ a Martin HILL. Počítačová analýza vícerozměrných dat v příkladech. Praha: Academia, 2005. ISBN 80-200-1335-0.

PECÁKOVÁ, Iva. Statistika v terénních průzkumech. Praha: Professional Publishing, 2008. ISBN 978-80-86946-74-0.