

Kurz - 1vf343 -  
GRETl ÚVOD A PRAKTICKÁ  
UKÁZKA LOGIT REGRESE



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



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY

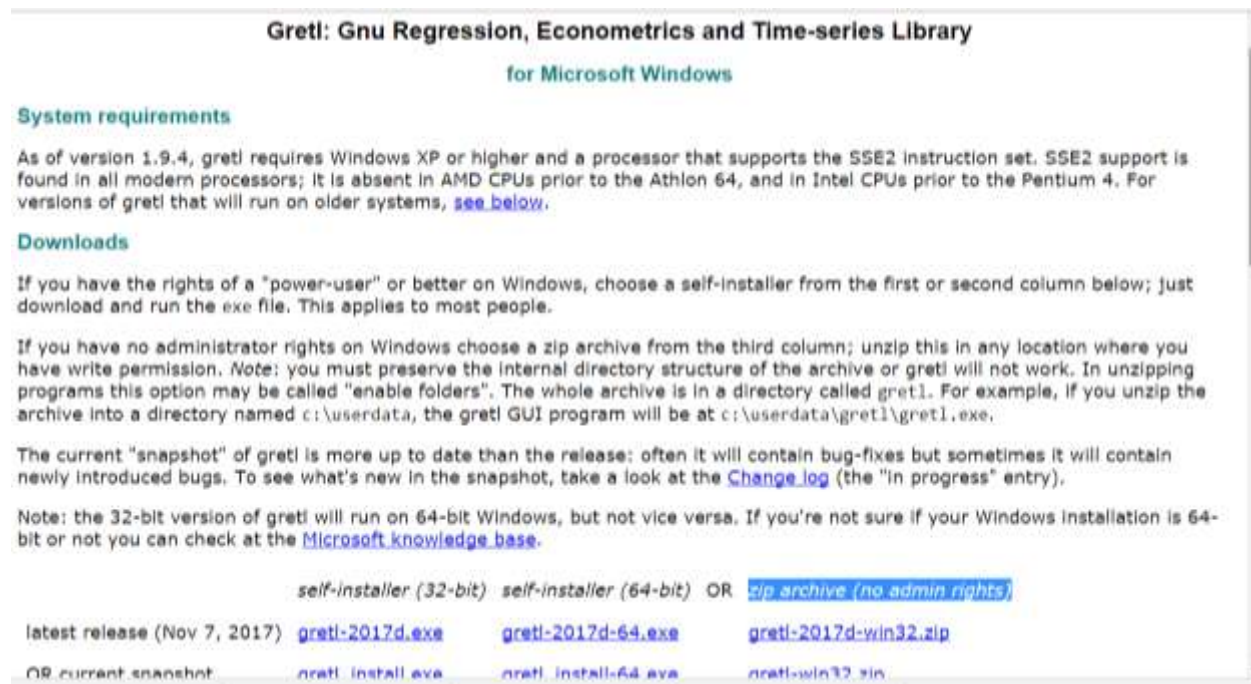
- sada pptx slidů Sada slide pro 1VF343 Ekonomická data a jejich analýza seznamující studenty se základním nastavením systému GRETL. 5,00
- +++ inspirace ve STATa

# Gretl pro Windows - představení

- Názorný (menu) / i v liště (pro pokročilejší)
- Všestranný (umí vše potřebné)
- Zdarma
- Solidní import dat
- Solidní instruktážní podpora (videa, vzorová data)
- Mobilní (portable)
- Více méně "Stata pro nemajetné a pro méně zkušené"

# Gretl pro Windows - instalace

- Ke stažení - <http://gretl.sourceforge.net/win32/>
- Volba - *zip archive (no admin rights)* – vhodné i pro práce na VŠE síti



**Gretl: Gnu Regression, Econometrics and Time-series Library**  
for Microsoft Windows

**System requirements**

As of version 1.9.4, gretl requires Windows XP or higher and a processor that supports the SSE2 instruction set. SSE2 support is found in all modern processors; it is absent in AMD CPUs prior to the Athlon 64, and in Intel CPUs prior to the Pentium 4. For versions of gretl that will run on older systems, [see below](#).

**Downloads**

If you have the rights of a "power-user" or better on Windows, choose a self-installer from the first or second column below; just download and run the exe file. This applies to most people.

If you have no administrator rights on Windows choose a zip archive from the third column; unzip this in any location where you have write permission. Note: you must preserve the internal directory structure of the archive or gretl will not work. In unzipping programs this option may be called "enable folders". The whole archive is in a directory called gretl. For example, if you unzip the archive into a directory named c:\userdata, the gretl GUI program will be at c:\userdata\gretl\gretl.exe.

The current "snapshot" of gretl is more up to date than the release: often it will contain bug-fixes but sometimes it will contain newly introduced bugs. To see what's new in the snapshot, take a look at the [Change log](#) (the "in progress" entry).

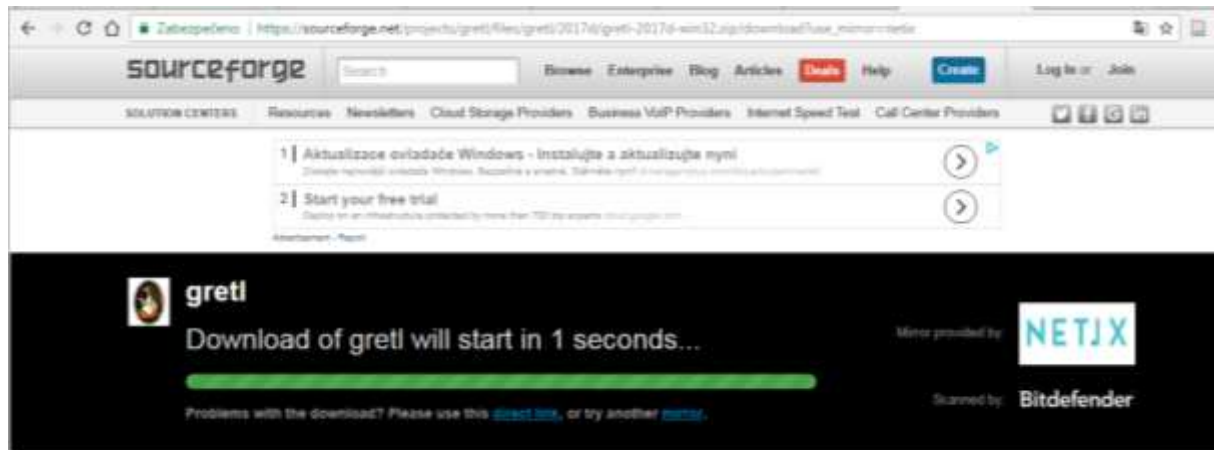
Note: the 32-bit version of gretl will run on 64-bit Windows, but not vice versa. If you're not sure if your Windows installation is 64-bit or not you can check at the [Microsoft knowledge base](#).

	<a href="#">self-installer (32-bit)</a>	<a href="#">self-installer (64-bit)</a>	OR	<a href="#">zip archive (no admin rights)</a>
latest release (Nov 7, 2017)	<a href="#">gretl-2017d.exe</a>	<a href="#">gretl-2017d-64.exe</a>		<a href="#">gretl-2017d-win32.zip</a>
OR current snapshot	<a href="#">gretl-install.exe</a>	<a href="#">gretl-install-64.exe</a>		<a href="#">gretl-win32.zip</a>

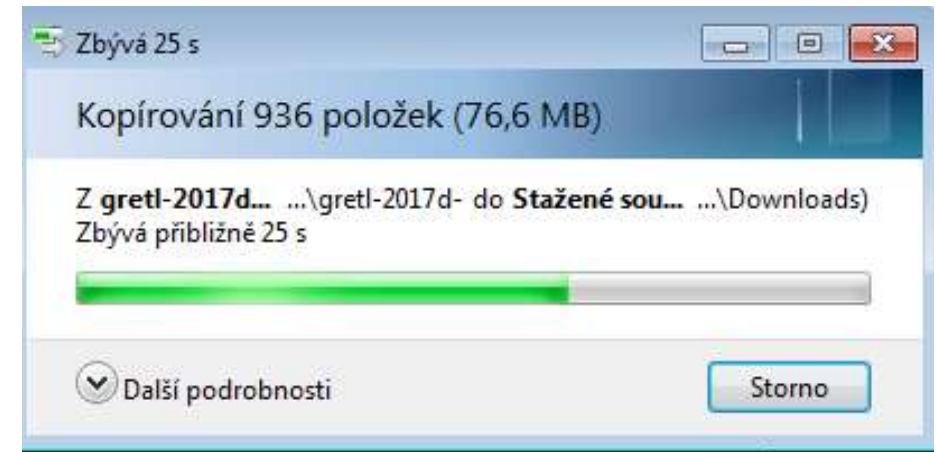
Zdroj: <http://gretl.sourceforge.net/win32/>

# Gretl pro Windows - instalace

- Ke stažení - <http://gretl.sourceforge.net/win32/>
- Volba - *zip archive (no admin rights)* – vhodné i pro práce na VŠE síti

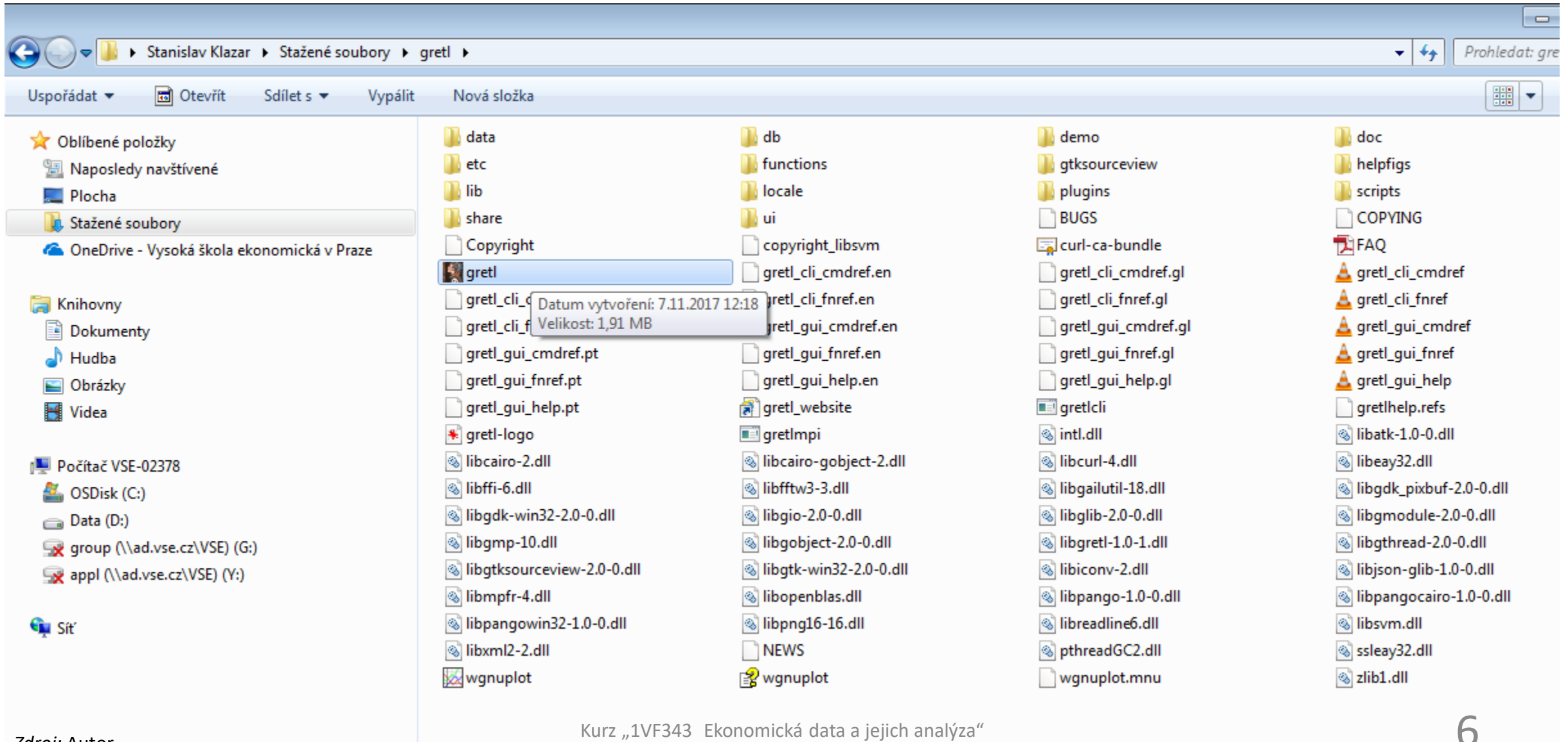


Zdroj: Autor

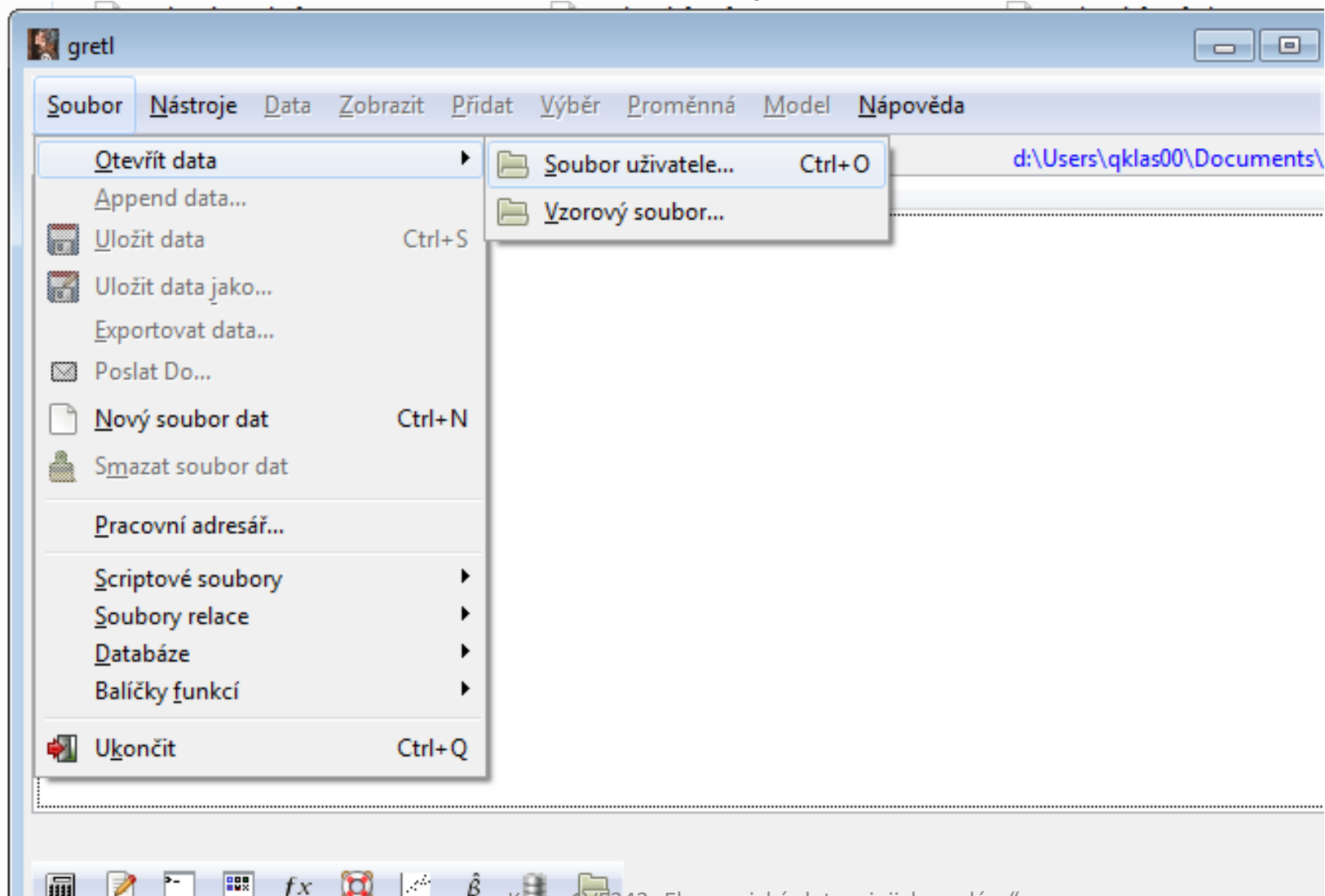


Zdroj: Autor

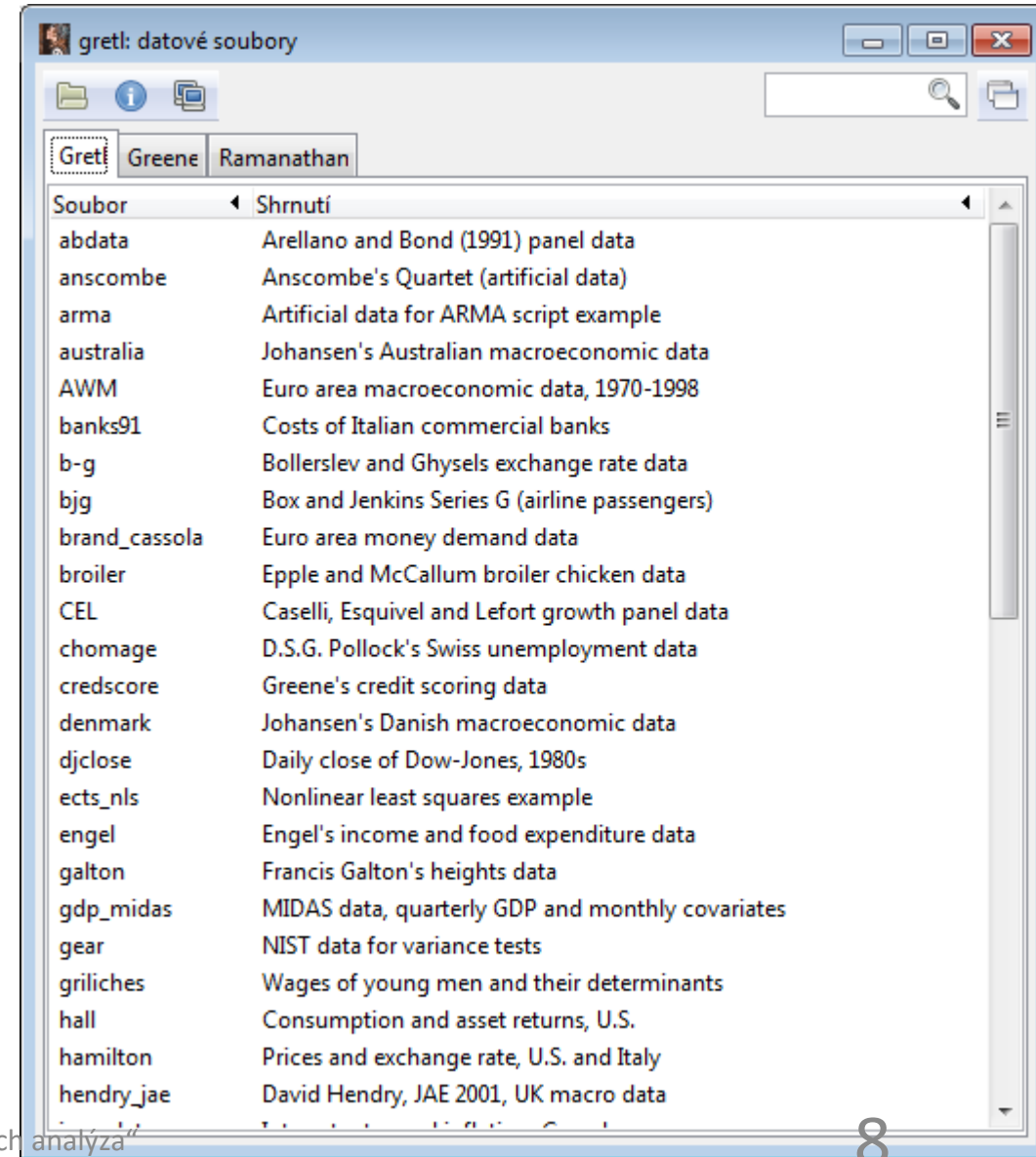
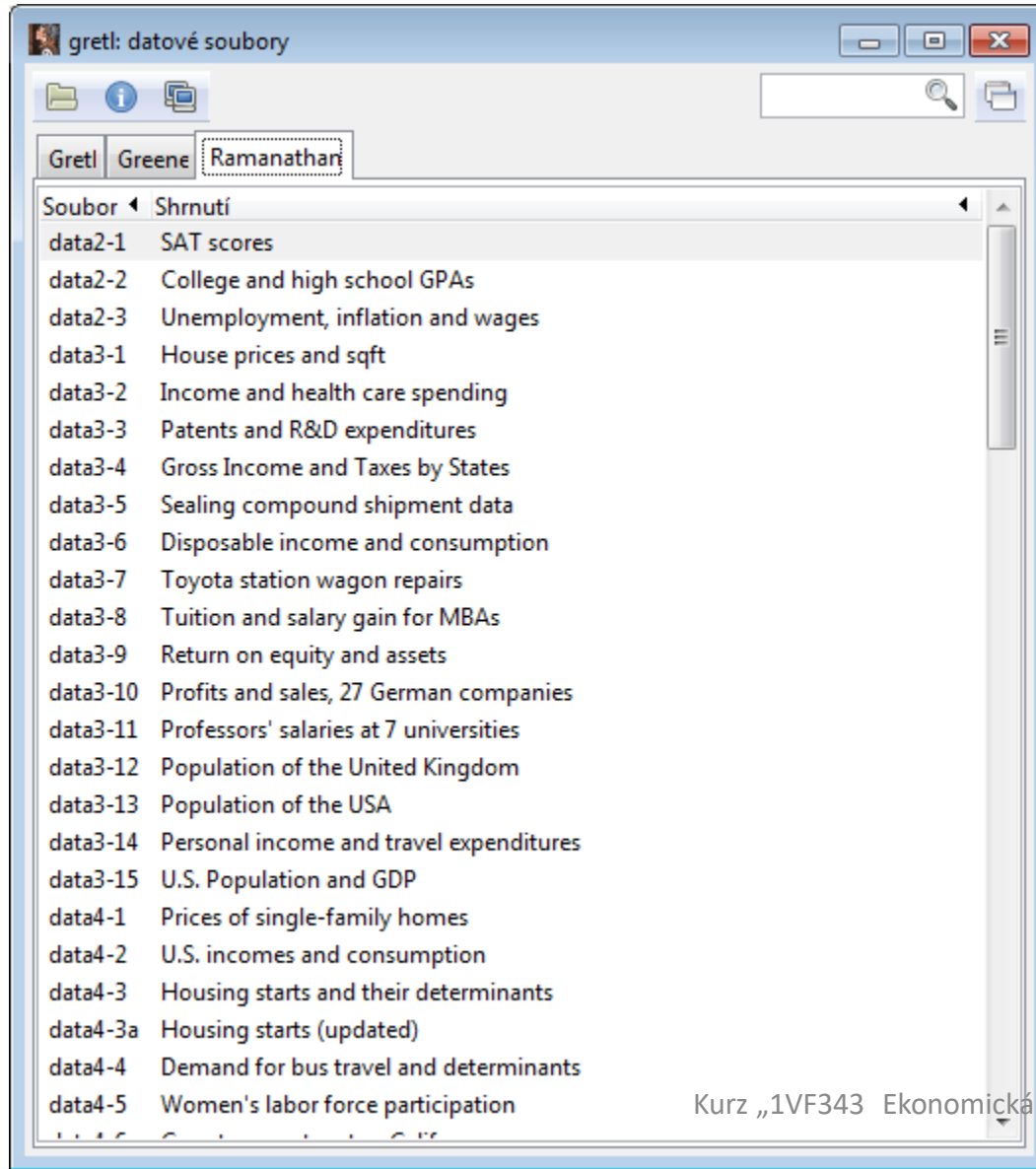
# První spuštění



# Úvodní obrazovka a import dat (vzorová)



# Příklady vzorových dat





# Logit v Gretlu - menu

The screenshot shows the Gretl software interface. The 'Model' menu is open, and the 'Logit' option is selected, which has opened a sub-menu with 'Binary' selected. The main window displays a list of variables for the file 'pension.gdt'.

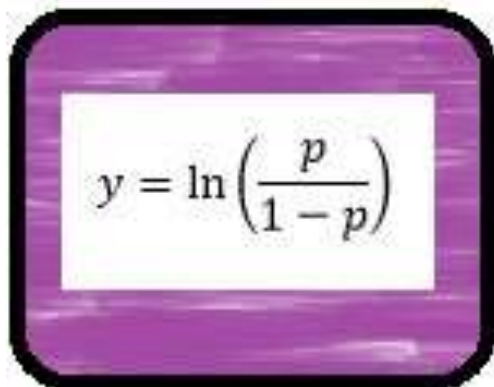
ID #	Jméno proměnné	Popisek
0	const	
1	id	family identifier
2	pyears	years in pension plan
3	prftshr	=1 if profit sharing plan
4	choice	=1 if can choose method invest
5	female	=1 if female
6	married	=1 if married
7	age	age in years
8	educ	highest grade completed
9	finc25	\$15,000 < faminc92 <= \$25,000
10	finc35	\$25,000 < faminc92 <= \$35,000
11	finc50	\$35,000 < faminc92 <= \$50,000
12	finc75	\$50,000 < faminc92 <= \$75,000
13	finc100	\$75,000 < faminc92 <= \$100,000
14	finc101	\$100,000 < faminc92
15	wealth89	net worth, 1989, \$1000
16	black	=1 if black
17	stckin89	=1 if owned stock in 1989
18	irain89	=1 if had IRA in 1989
19	retretck	0=methode 50=mixed 100=metretck

# Logit model vs „klasická regrese“ - k čemu slouží?

- Doplnění klasické regrese pro případy, kdy vysvětlovaná proměnná je tzv. binární (0, 1), například stavy
  - „vyhrál“ vs „nevyhrál“
  - „absovoval“ vs „neabsolvoval“
  - „zaměstnán“ vs „nezaměstnán“
  - „splácí úvěr“ vs „nesplácí“ (bankovníctví)
  - „černý pasažér“ vs „řádné placení daní a poplatků“ (veřejné finance)
- Odpoví na otázku, jak moc se změní pravděpodobnost úspěchu (vysvětlovaná proměnná = 1), při změně proměnných vysvětlujících ( $x_1, x_2, \dots$ ).

# Logit model - jak funguje (určení Y)

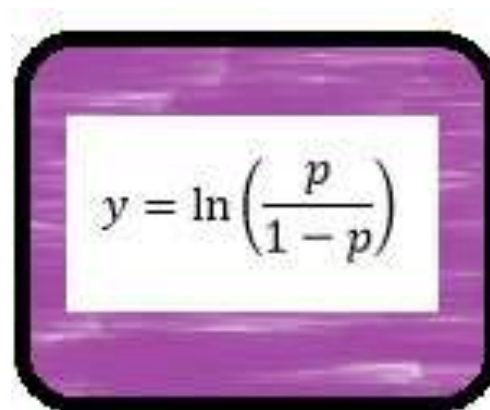
- Je nutný mezistupeň při výpočtu. Nejdříve standardně určíme vztah
- $Y = a_0 + a_1X_1 + a_2X_2 \dots$  kde:
  - Y je binární (0, 1)
  - $X_1, X_2$  jsou standardní vysvětlující proměnné
- Ve druhém kroku je proveden přepočítání Y na příslušnou pravděpodobnost. Využívá se vztahu:


$$y = \ln\left(\frac{p}{1-p}\right)$$

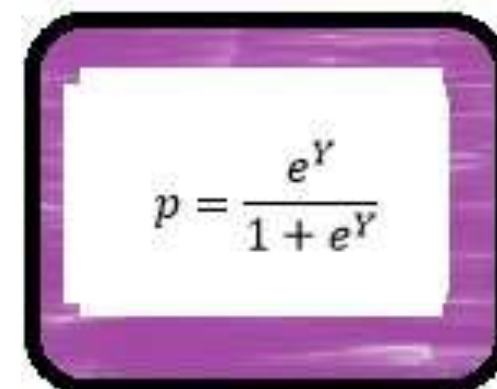
# Logit model - jak funguje (odvození „p“)

Odvození pravděpodobnosti „p“ ze vztahu  $\ln(p/(1-p)) = Y$  ve 4 krocích

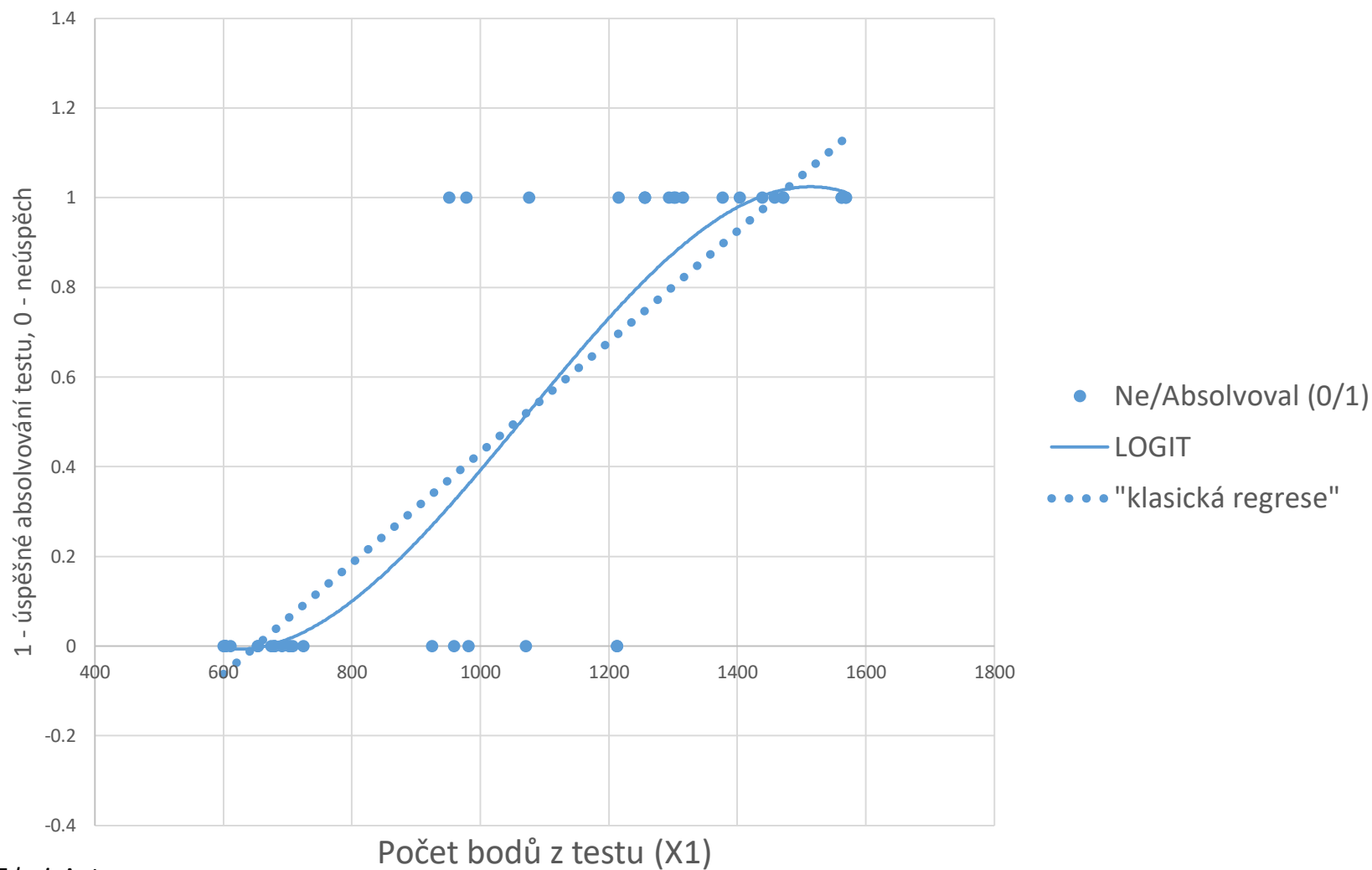
1.  $p/(1-p) = \exp(y)$
2.  $p = \exp(y) - p \cdot \exp(y)$
3.  $p + p \cdot \exp(y) = \exp(y)$
4.  $p \cdot (1 + \exp(y)) = \exp(y)$
5.  $p = \exp(y) / (1 + \exp(y))$


$$y = \ln\left(\frac{p}{1-p}\right)$$




$$p = \frac{e^y}{1 + e^y}$$

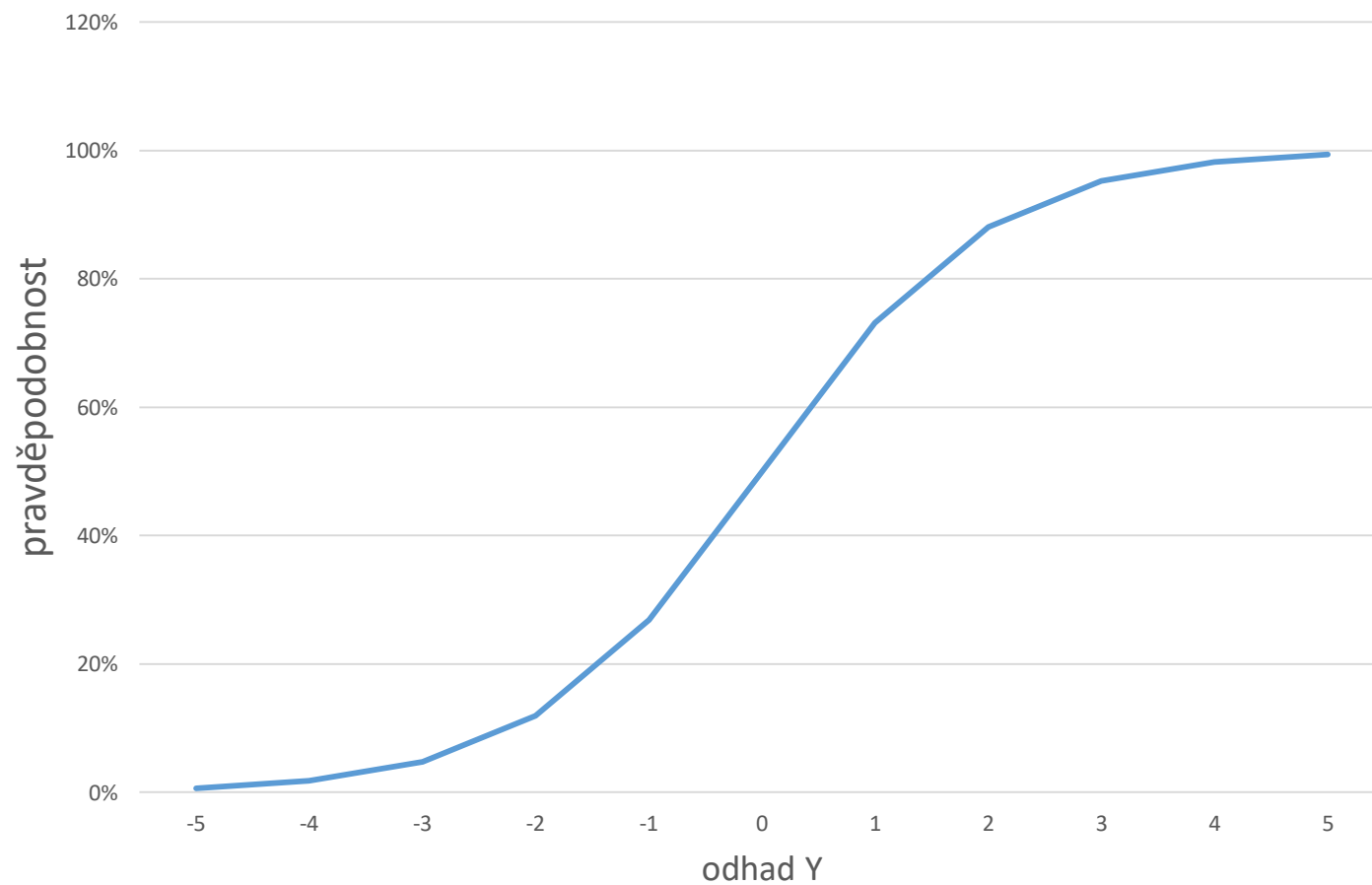
# Logit model vs „klasická regrese“ - graficky



Zdroj: Autor

Kurz „1VF343 Ekonomická data a jejich analýza“

# Od Y ku „ $p=(0, 100\%)$ “ graficky



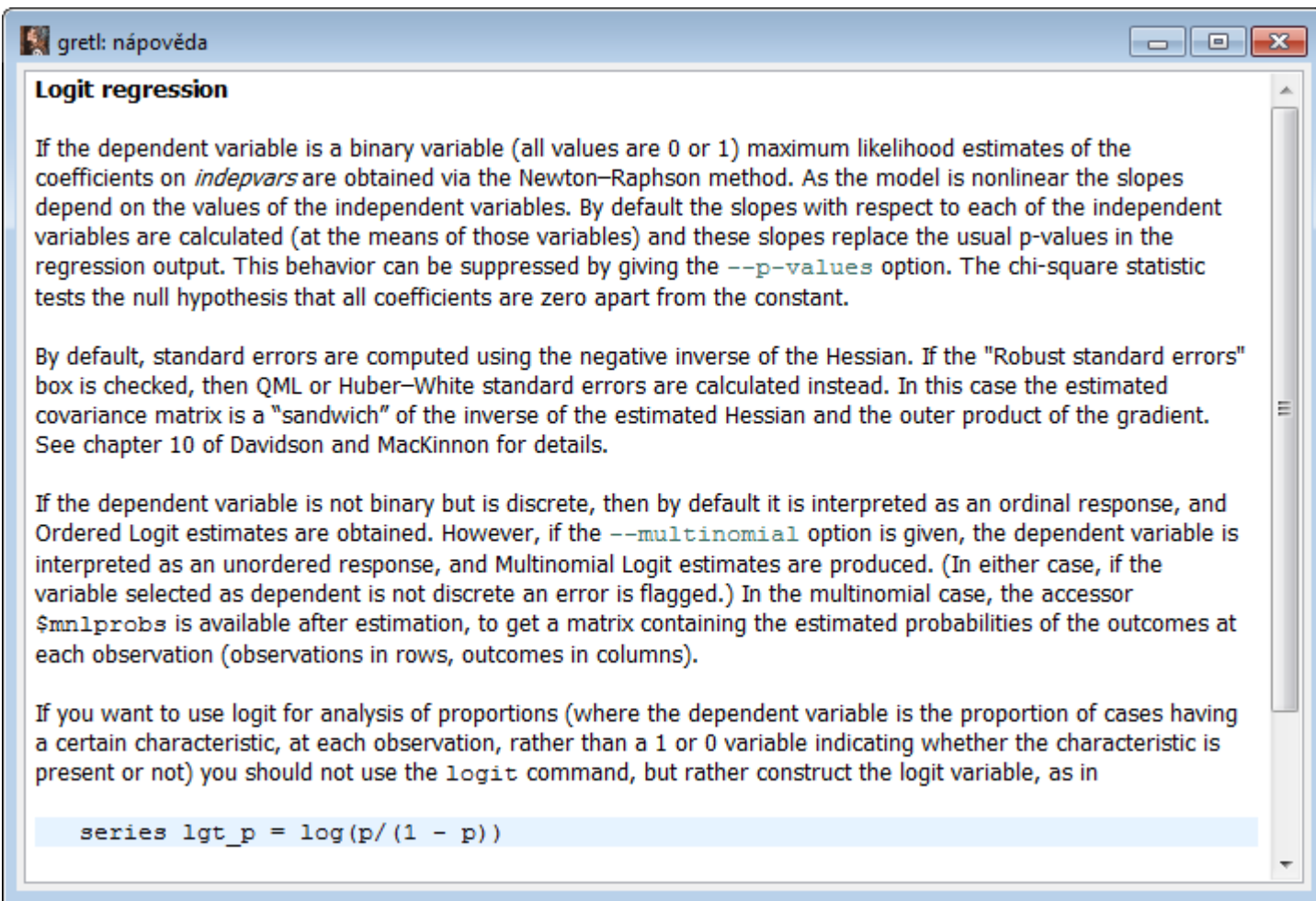
Zdroj: Autor

# Logit v praxi – vstupní data

Týdny přípravy	0,50	0,75	1,00	1,25	1,50	1,75	1,75	2,00	2,25	2,50
úspěch 0/1	0	0	0	0	0	0	1	0	1	0
Týdny	2,75	3,00	3,25	3,50	4,00	4,25	4,50	4,75	5,00	5,50
úspěch 0/1	1	0	1	0	1	1	1	1	1	1

Zdroj: Autor

# Gretl – Logit – nápověda a vstup



**gretl: nápověda**

### Logit regression

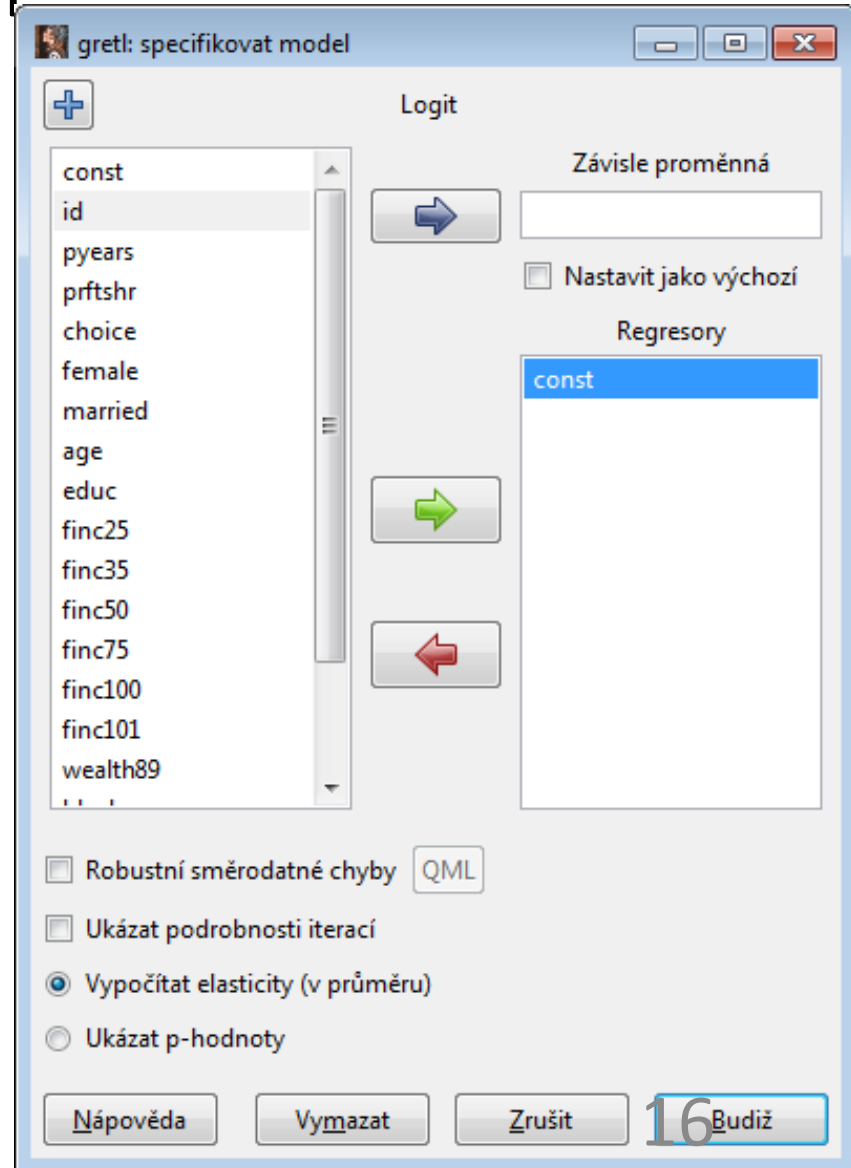
If the dependent variable is a binary variable (all values are 0 or 1) maximum likelihood estimates of the coefficients on *indepvars* are obtained via the Newton–Raphson method. As the model is nonlinear the slopes depend on the values of the independent variables. By default the slopes with respect to each of the independent variables are calculated (at the means of those variables) and these slopes replace the usual p-values in the regression output. This behavior can be suppressed by giving the `--p-values` option. The chi-square statistic tests the null hypothesis that all coefficients are zero apart from the constant.

By default, standard errors are computed using the negative inverse of the Hessian. If the "Robust standard errors" box is checked, then QML or Huber–White standard errors are calculated instead. In this case the estimated covariance matrix is a "sandwich" of the inverse of the estimated Hessian and the outer product of the gradient. See chapter 10 of Davidson and MacKinnon for details.

If the dependent variable is not binary but is discrete, then by default it is interpreted as an ordinal response, and Ordered Logit estimates are obtained. However, if the `--multinomial` option is given, the dependent variable is interpreted as an unordered response, and Multinomial Logit estimates are produced. (In either case, if the variable selected as dependent is not discrete an error is flagged.) In the multinomial case, the accessor `$mnlprobs` is available after estimation, to get a matrix containing the estimated probabilities of the outcomes at each observation (observations in rows, outcomes in columns).

If you want to use logit for analysis of proportions (where the dependent variable is the proportion of cases having a certain characteristic, at each observation, rather than a 1 or 0 variable indicating whether the characteristic is present or not) you should not use the `logit` command, but rather construct the logit variable, as in

```
series lgt_p = log(p/(1 - p))
```



**gretl: specifikovat model**

### Logit

Závisle proměnná

Nastavit jako výchozí

Regresory

const

Robustní směrodatné chyby QML

Ukázat podrobnosti iterací

Vypočítat elasticity (v průměru)

Ukázat p-hodnoty

Nápověda Vymazat Zrušit 16 Budič



# Logit v praxi – výstup 1 (gretl)

- Odvozena rovnice  $Y = a_0 + a_1X_1$
- Tedy  $Y = -4 + 1,5 \cdot \text{počet týdnů}$
- Zároveň je určena i p-hodnota
- Celková kvalita testu viz na konci
- Korektní předpověď v 16 případech
- Chybná ve 4 případech

```
gretl: model 2
Soubor Upravit Testy Uložit Grafy Analýza LaTeX
Model 2: Logit, za použití pozorování 1-20
Závisle proměnná: x0
Směrodatné chyby založené na Hessiánu

-----
                koeficient   směr. chyba   z       p-hodnota
-----
const          -4,07771      1,76099      -2,316   0,0206   **
MIN             1,50465      0,628721     2,393   0,0167   **

Střední hodnota závisle proměnné      0,500000
Sm. odchylka závisle proměnné         0,512989
McFaddenův koeficient determinace      0,420767
Adjustovaný koeficient determinace     0,276497
Logaritmus věrohodnosti                 -8,029878
Akaikovo kritérium                       20,05976
Schwarzovo kritérium                     22,05122
Hannan-Quinnovo kritérium                20,44851
zde je poznámka o zkratkách statistik modelu

Počet 'správně předpovězených' případů = 16 (80,0%)
f(beta'x) pro střední hodnotu nezávisle proměnných = 0,249
Test poměru věrohodnosti: Chi-kvadrát(1) = 11,6661 [0,0006]

                Předpovězené
                0   1
Skutečné 0   8   2
           1   2   8
```

Zdroj: Autor (Gretl)

# Logit v praxi – výstup 2 (excel logits ukol.xls) - pravděpodobnosti

- Proveden přepočet na „p“ v procentech
- Interpretace vztahu mezi změnou v počtu týdnů přípravy (X1) a změnou v pravděpodobnosti úspěšného absolvování předmětu.

	min	min	min	min	min	min	min	min
1	3,25	3,5	4	4,25	4,5	4,75	5	5,5
1	1	1	1	1	1	1	1	1
1	3,25	3,5	4	4,25	4,5	4,75	5	5,5
0,8	0,812403	1,188565	1,94089	2,317053	2,693215	3,069378	3,44554	4,197865
%	69,26%	76,65%	87,44%	91,03%	93,66%	95,56%	96,91%	98,52%
%		7,39%		3,58%		1,90%		1,61%
e v průměru					17,01%			

směrnice  
-----  
0,374888

Student se původně zabýval studiem 3,5 hod., což je pravděpodobnost úspěchu 76,65 %. Rozhodne se celkovou dobu studia prodloužit na 4,5 hod, což se zvýší pravděpodobnost na 93,66 %. Změna pravděpodobnosti úspěšného absolvování kurzu u studenta je **17,01 %**.

Zdroj: Autor (Excel)



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.*

