

Regresní analýza

Organizačně

- Do 13:30
 - Co je to regresní analýza?
 - Kdy se používá?
 - Na jaké otázky může nabídnout odpověď?
 - Základní principy
- Přestávka
- Praktické procvičení

- **KDYŽ NĚČEMU NEPOROZUMÍTE, OZVĚTE SE !!!**

Použití

- TESTOVÁNÍ TEORIÍ !!!
- Zjištění vlivu nezávisle proměnné na závisle proměnnou
 - Při kontrole dalších možných faktorů
 - (predikce: jakou hodnotu bude mít závisle proměnná při určité kombinaci nezávisle proměnných)

Příklady otázek ze závěrečných prací v ISu

- Existuje vztah mezi kvalitou a náročností výuky a mírou stresu u studentů?
 - Studenti vyšších ročníků zažívají menší míru stresu oproti studentům nižších ročníků
 - Kvalita online výuky má kladný vliv na míru stresu
 - nároky vyučujících mají vliv opačný
- Jak závisí stupeň glomerulární filtrace na biochemických, demografických a antropometrických údajích pacientů?
 - Mezi nezávislé faktory asociované s nižší glomerulární filtrací patří vyšší hladina sérového kreatininu, vyšší věk, ženské pohlaví, jiný než Afroamerický etnický původ, vyšší koncentrace sérové urey a nižší koncentrace sérového albuminu
- Jak závisí rychlost plavání na stylu?

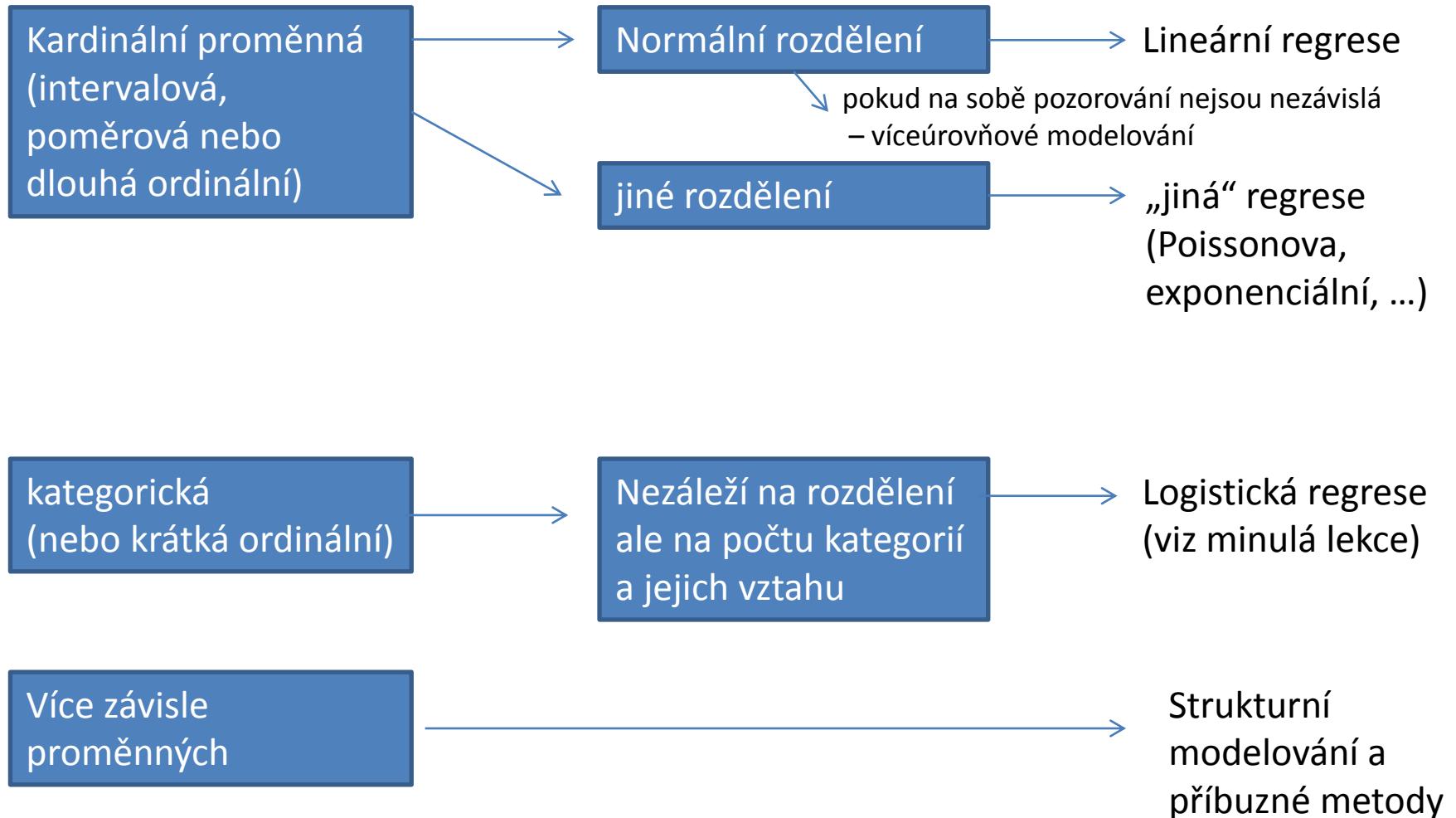
Příklady otázek ze závěrečných politologických prací v ISu

- Desítky volebně-geografických prací
- Co ovlivňuje jednotu českých poslaneckých klubů?
 - H1a: Jednotnost hlasování je vyšší u vládních stran.
 - H1b: Jednotnost hlasování vládní strany je vyšší, čím těsnější je většina, kterou disponuje.
- Je míra korupce ovlivněna i používaným volebním systémem?
 - 1) Korupce roste s rostoucími volebními obvody v systémech s otevřenými kandidátkami.
 - 2) Korupce klesá s rostoucími volebními obvody v systémech s uzavřenými kandidátkami.
- Co ovlivňuje (ne)účast poslanců na hlasování v Poslanecké sněmovně PČR?
 - účast na hlasování se bude zvyšovat s rostoucí pravděpodobností, že daný poslanec, či poslankyně bude pivotálním ... hlasem ...
 - poslanci ze vzdálenějších obvodů budou mít vyšší míru absencí při hlasováních ve Sněmovně než poslanci, kteří jsou přímo z Prahy, nebo blízkého okolí

Podmínky

- **Jedna závisle proměnná**
 - + (Jedna nebo) **více** nezávisle proměnných
- (Normálně rozdělená) **kardinální** závisle proměnná
 - Ve skutečnosti se podmínka týká rozdělení reziduí
 - To ale před provedením analýzy neznáme
 - Pro některá rozdělení jsou lepší jiné typy regresní analýzy
 - Rozdělení a typ nezávisle proměnných může být jakékoli
- Několik dalších různě důležitých podmínek
 - Nezávislost pozorování
 - Předpoklad lineárního vztahu
 - Nezávislost nezávisle proměnných mezi sebou
 - Homogenní rozptyl reziduí

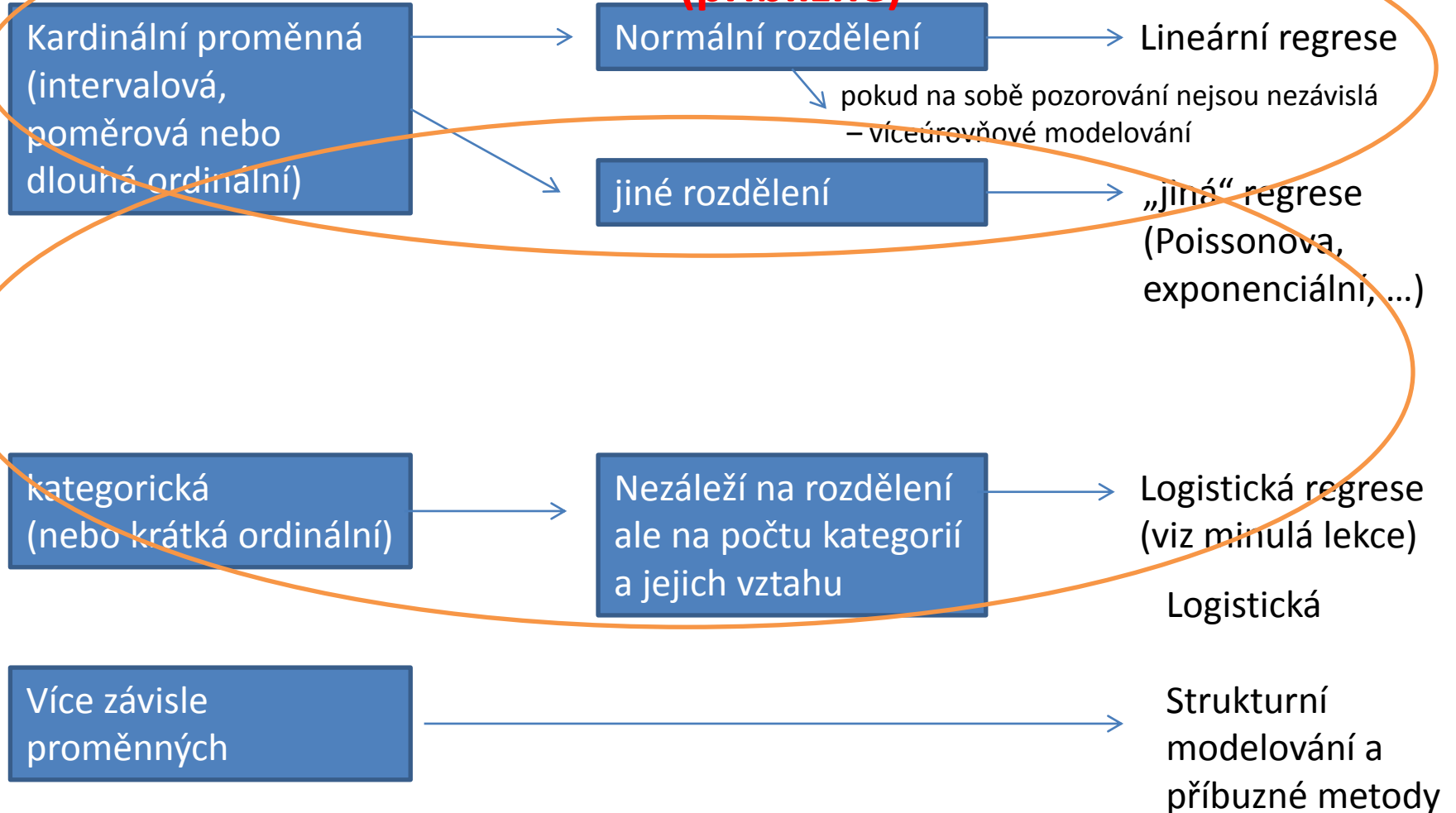
Rozhodovací strom



Rozhodovací strom

Lineární

(přibližně)

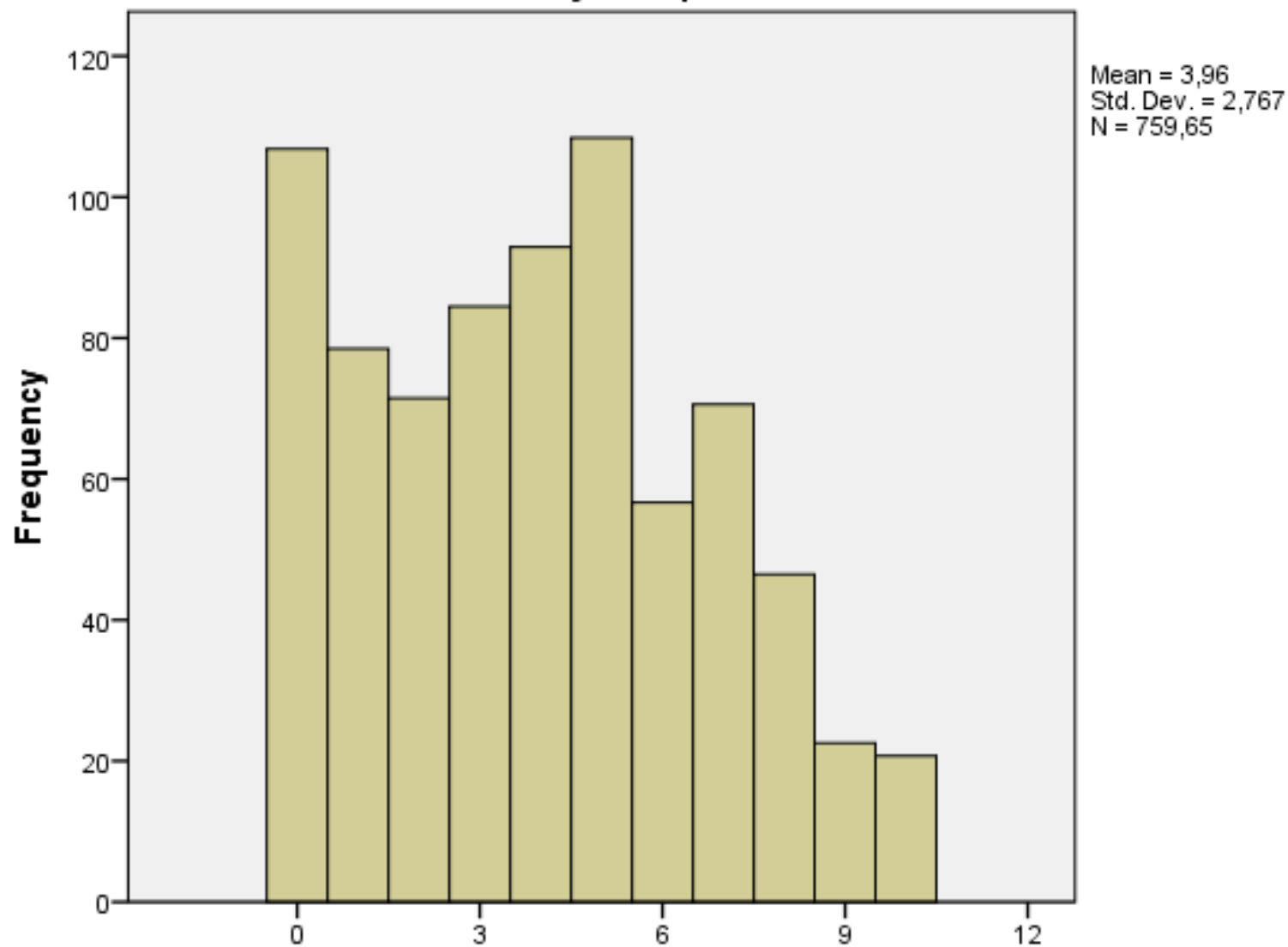


Logistická

Strukturní modelování a příbuzné metody

| | Stata | SAS | SPSS | Mplus | R |
|---|-----------------------|---------------------------------------|----------------------|-----------------------|-------------------|
| Regression Models | | | | | |
| Robust Regression | Stata | SAS | | | R |
| Models for Binary and Categorical Outcomes | | | | | |
| Logistic Regression | Stata | SAS | SPSS | Mplus | R |
| Exact Logistic Regression | Stata | SAS | | | R |
| Multinomial Logistic Regression | Stata | SAS | SPSS | Mplus | R |
| Ordinal Logistic Regression | Stata | SAS | SPSS | Mplus | R |
| Probit Regression | Stata | SAS | SPSS | Mplus | R |
| Count Models | | | | | |
| Poisson Regression | Stata | SAS | SPSS | Mplus | R |
| Negative Binomial Regression | Stata | SAS | SPSS | Mplus | R |
| Zero-inflated Poisson Regression | Stata | SAS | | Mplus | R |
| Zero-inflated Negative Binomial Regression | Stata | SAS | | Mplus | R |
| Zero-truncated Poisson | Stata | SAS | | | R |
| Zero-truncated Negative Binomial | Stata | SAS | | Mplus | R |
| Censored and Truncated Regression | | | | | |
| Tobit Regression | Stata | SAS | | Mplus | R |
| Truncated Regression | Stata | SAS | | | R |
| Interval Regression | Stata | SAS | | | R |
| Multivariate Analysis | | | | | |
| One-way MANOVA | Stata | SAS | SPSS | | |
| Discriminant Function Analysis | Stata | SAS | SPSS | | |
| Canonical Correlation Analysis | Stata | SAS | SPSS | | R |
| Multivariate Multiple Regression | Stata | SAS | | Mplus | |
| Mixed Effects Models | | | | | |
| Generalized Linear Mixed Models | | Introduction to GLMMs | | | |
| Mixed Effects Logistic Regression | Stata | | | | R |
| Other | | | | | |
| Latent Class Analysis | | | | Mplus | |

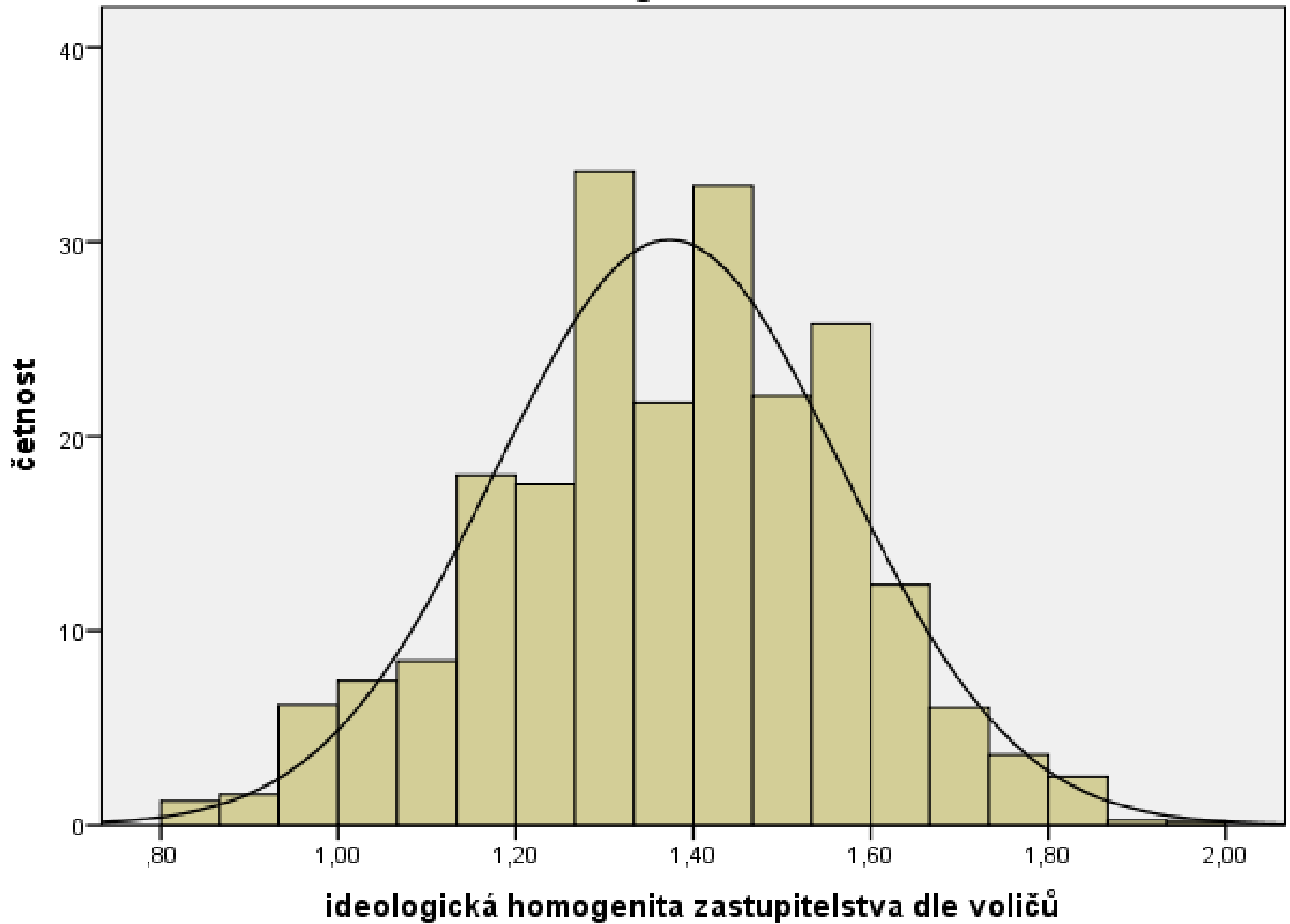
zájem o politiku: EU



zájem o politiku: EU

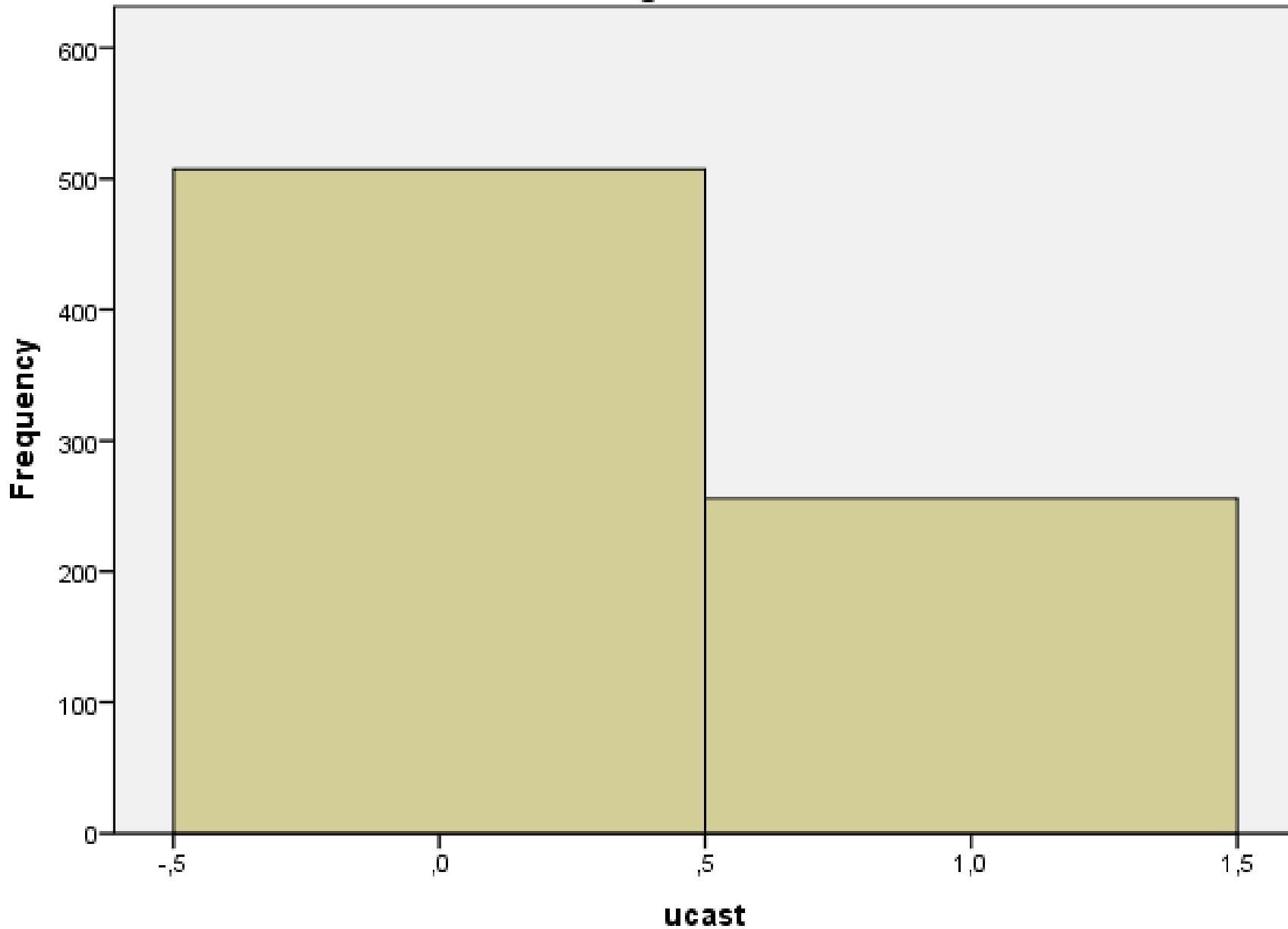
Cases weighted by vaha

Histogram 1



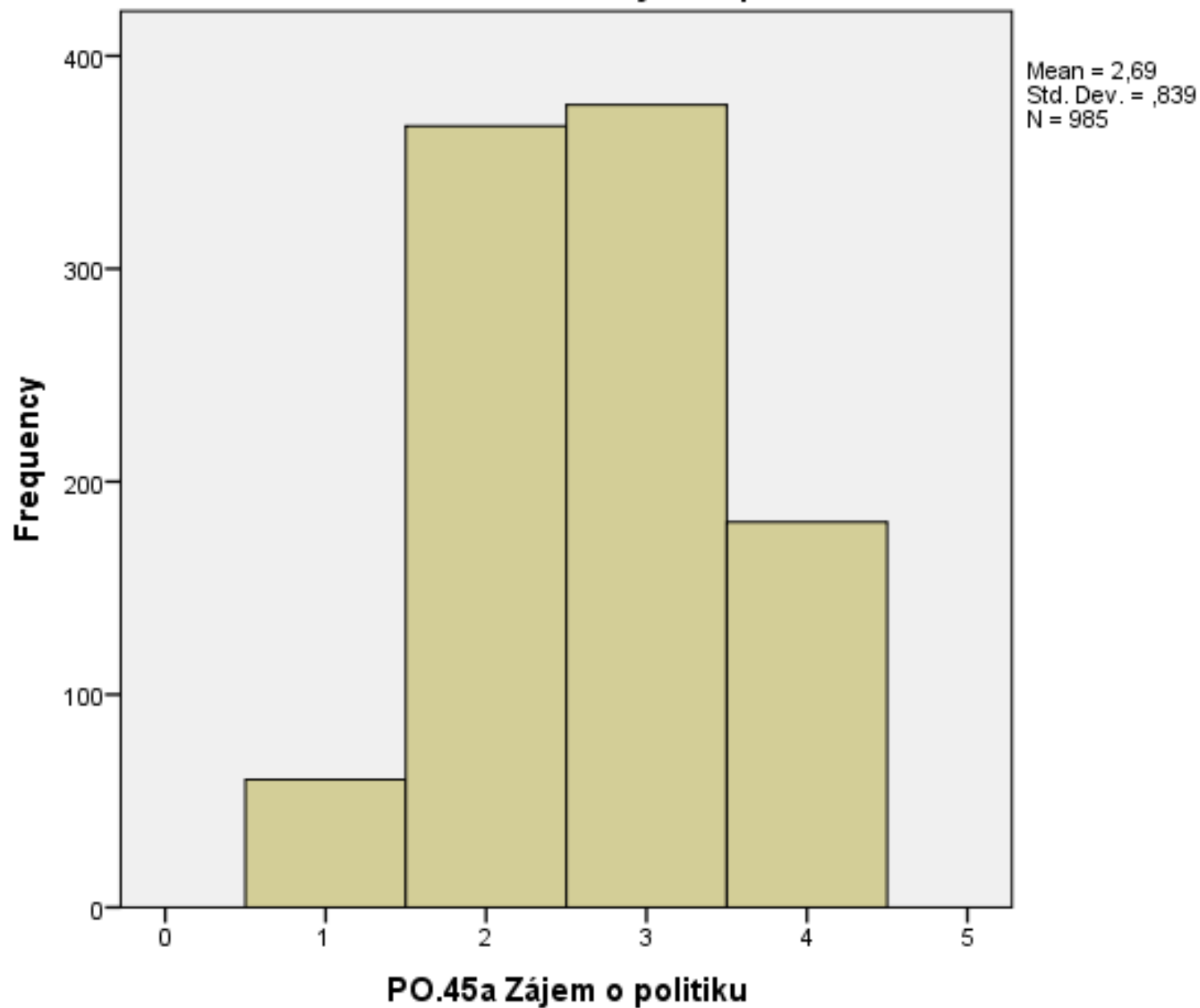
Cases weighted by vaha

Histogram 2

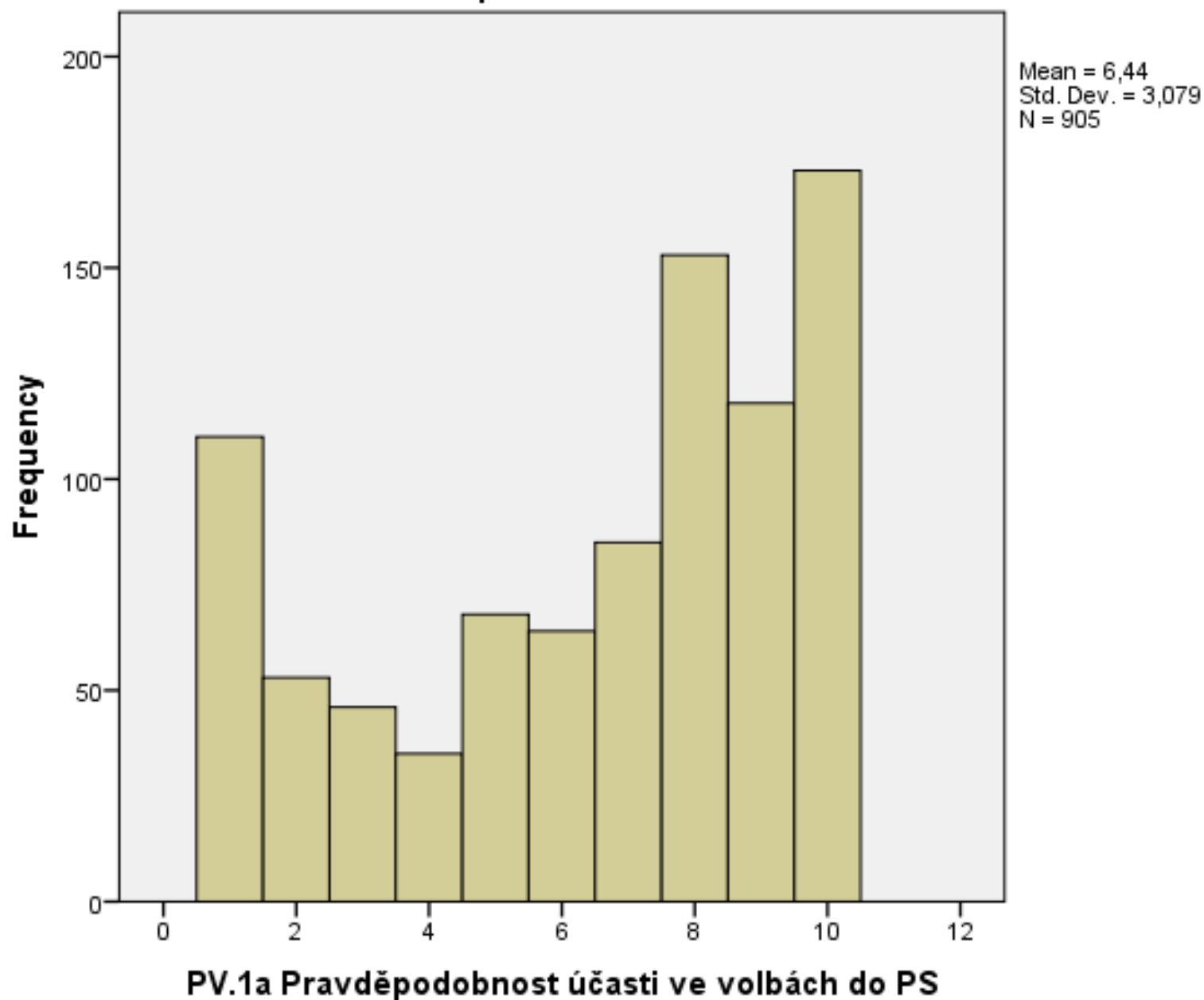


Cases weighted by vaha

PO.45a Zájem o politiku



PV.1a Pravděpodobnost účasti ve volbách do PS



Co regrese dělá

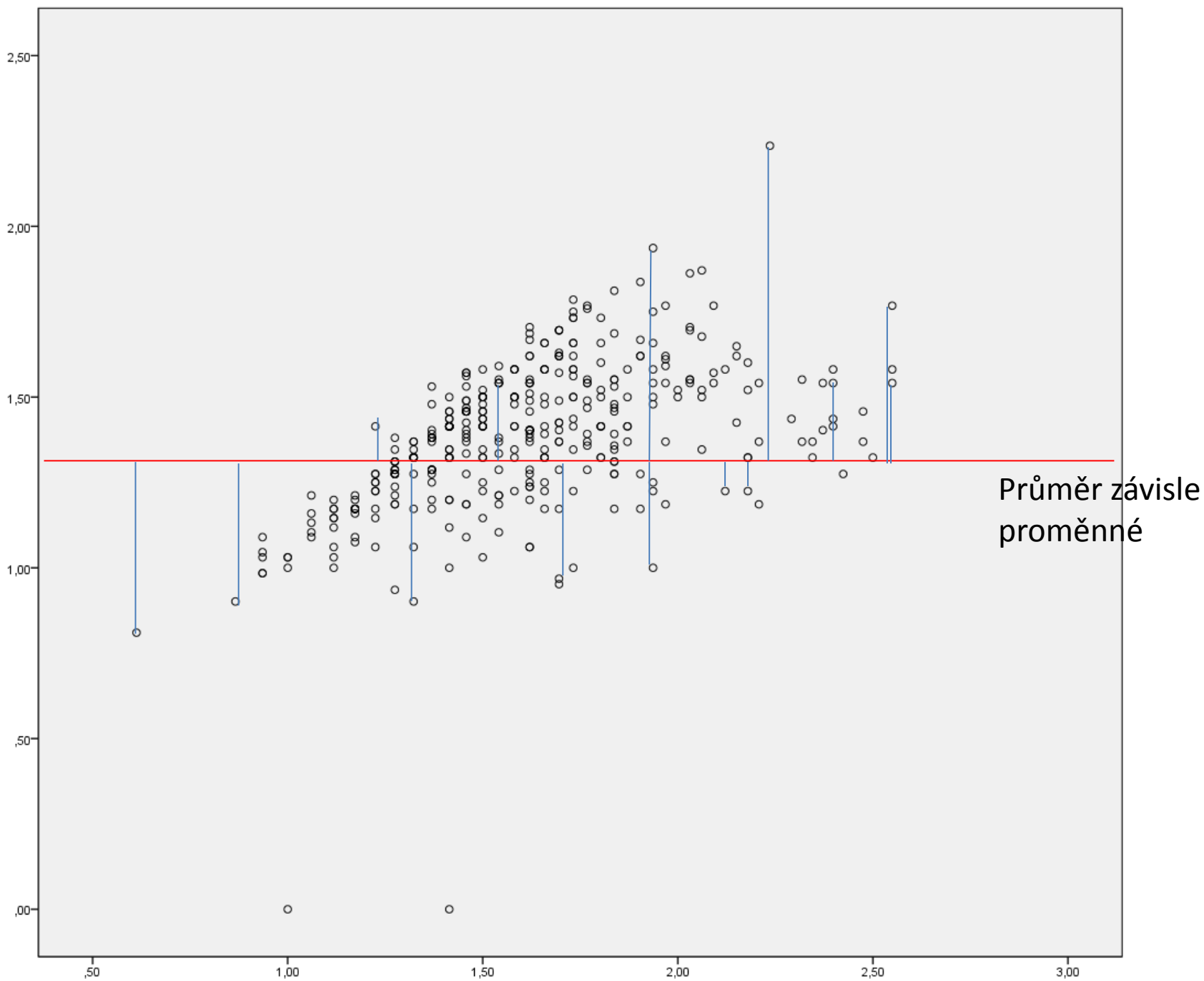
- **Odhad parametrů** přímky (při 1 nezávisle proměnné), roviny (při 2) či nadroviny (při více)
- Parametry: **sklon** (pro každou proměnnou) a ***konstanta*** (jedna pro celý model)
- Parametry popisují vztah mezi nezávisle a závisle proměnnou
- Hodnota závisle proměnné (y) = konstanta (a) + sklon(b)*hodnota nezávisle proměnné (x)
- $y = a + b * x$
- $y = a + b_1 * x + b_2 * x + b_3 * x + \dots$

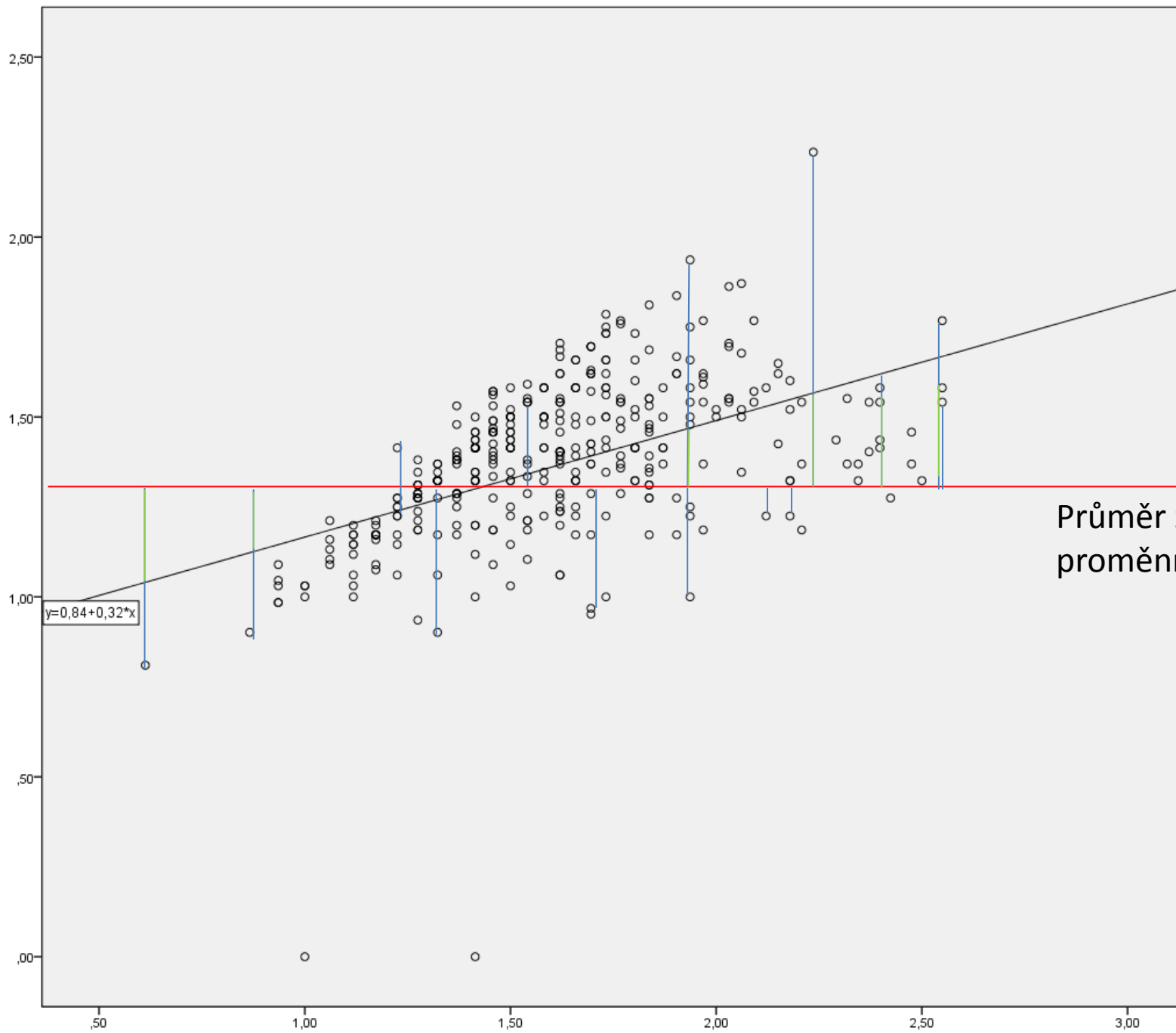
Co nám výpočet poskytne?

- R-square (česky index determinace)
 - Ukazuje jak dobře model sedí na data
- Parametry
 - Unstandardized beta (nestandardizovaný beta koeficient)
 - Constant (konstanta)
- Hodnoty signifikance

Co je to R-square?

- Ukazuje, kolik procent rozptylu závisle proměnné je vysvětleno přidáním nezávisle proměnných
- Původní rozptyl je vypočten jako suma kvadratických odchylek mezi průměrem a jednotlivými hodnotami závisle proměnné
- „nový“ rozptyl je vypočten jako suma odchylek od regresní přímky/roviny
- Rozdíl mezi původním a novým rozptylem vydělený původní variabilitou = R-square
- Čím víc proměnných, tím nižší R-square
 - Řešeno pomocí adjusted R-square



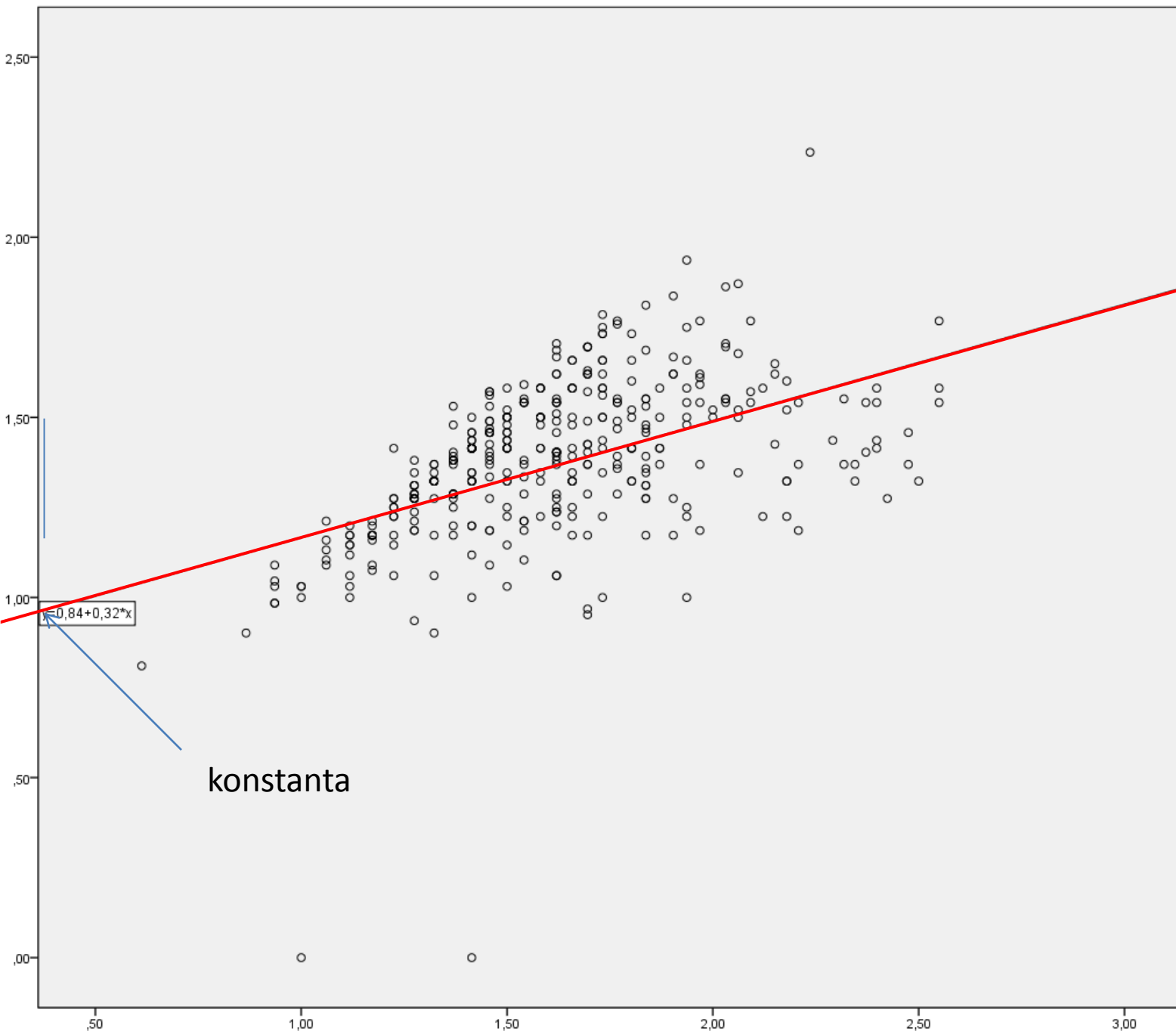


Průměr závisle
proměnné

Konstanta

- Jaká je očekávaná hodnota nezávisle proměnné, pokud jsou hodnoty všech nezávisle proměnných 0
- Pro smysluplnou interpretaci je často potřeba rekódovat proměnné
 - Každý má nějaký věk, pohlaví, výšku, váhu, ...
 - Důležitá pro tvorbu grafů (není předmětem tohoto kurzu)
- **Není důležitá pro interpretaci výsledků**

Y



X

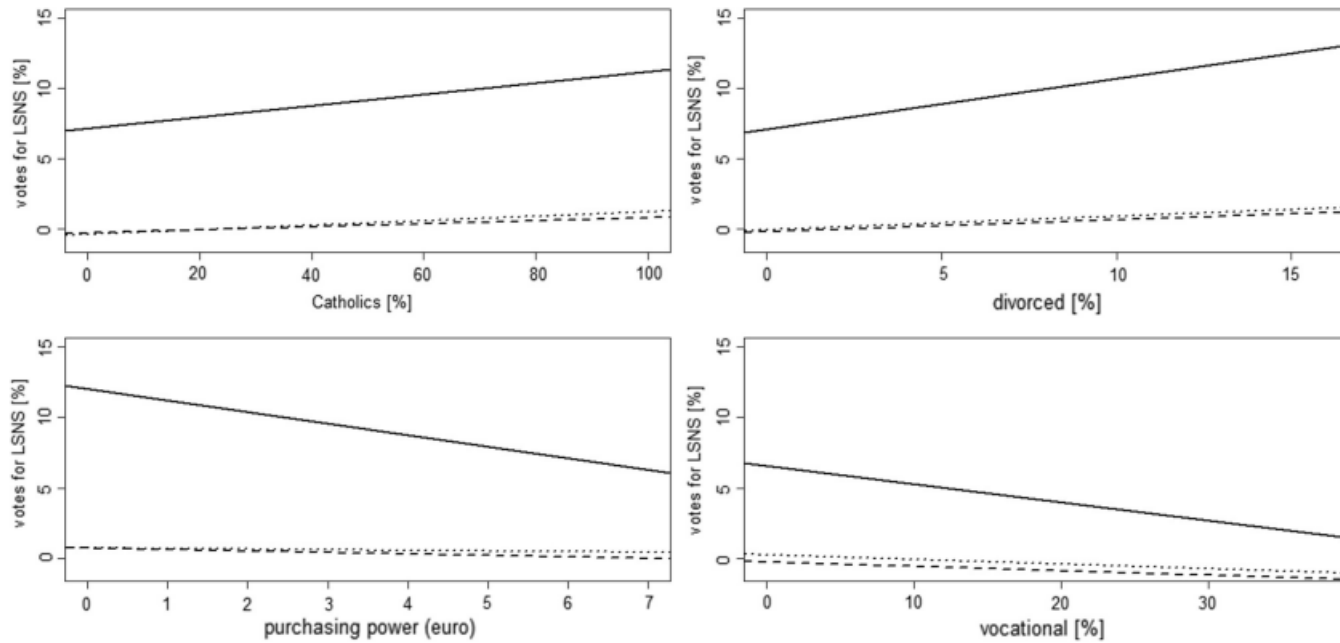
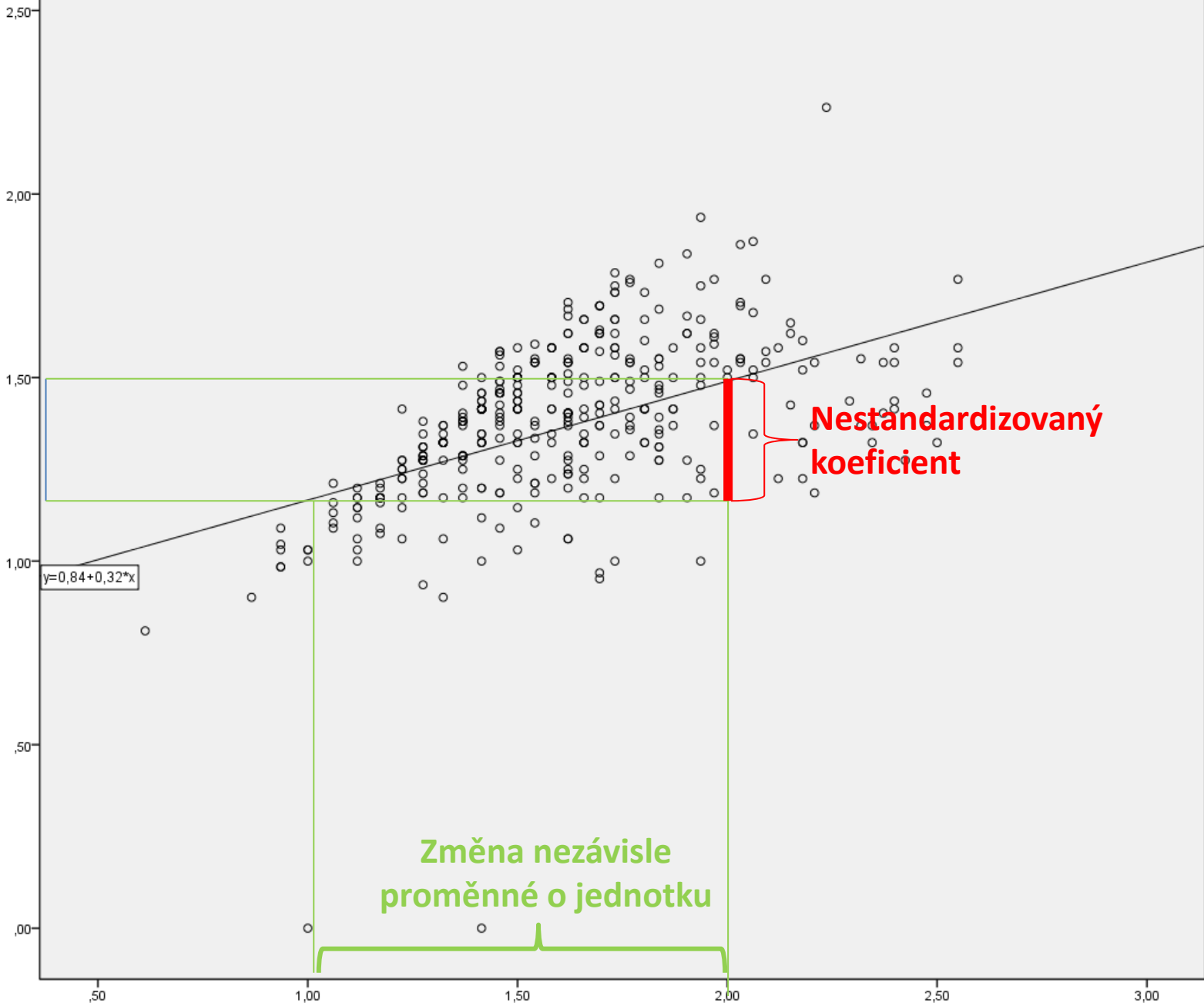


Figure 2. Effects of continuous variables on support for LSNS. — 2016, ··· 2012, - - 2010

Nestandardizovaný Beta koeficient

- **Nejdůležitější věc z celé regresní analýzy!!!**
- efekt nezávisle proměnné na závisle proměnnou
- **„o kolik se změní hodnota závisle proměnné, pokud se hodnota nezávisle proměnné změní o jednotku“** pokud vše ostatní zůstává shodné
- Různé proměnné se mohou změnit o různý počet jednotek
 - Pro srovnání síly proměnných v modelu – standardizovaný koeficient beta (jakou změnu v počtu směrodatných odchylek závisle proměnné způsobí změna o směrodatnou odchylku nezávisle proměnné)

Y

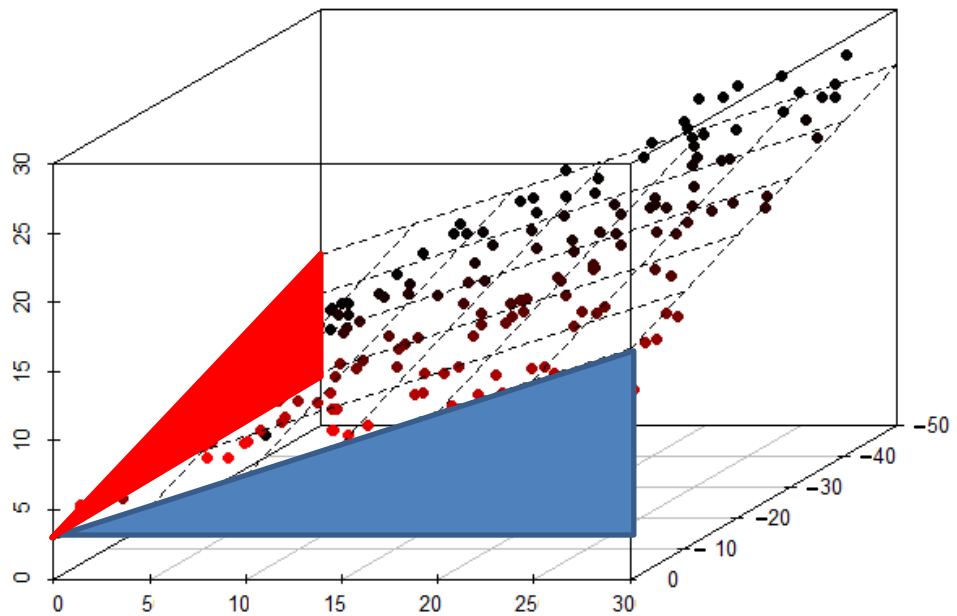


**Nestandardizovaný
koeficient**

**Změna nezávisle
proměnné o jednotku**

$y=0,84+0,32*x$

X



Příklad

- Téma: Názory na zasahování státu do ekonomiky
- Popis problému:
 - Občané mají různé názory na to, zda a jak by měl stát zasahovat do hospodářství
- Otázka: Co způsobuje rozdílné názory na zásahy státu do ekonomiky mezi občany?

Postup

- Nadefinování modelu pomocí hypotéz vycházejících z teorie
- Sestavení datasetu obsahujícího závisle a nezávisle proměnné dle specifikace
- Zkontrolování rozdělení závisle proměnné
- Zkontrolování vlastností nezávisle proměnných

Teorie

- Politické hodnoty
- Hodnoty jsou preferovanými stavy věcí (svoboda x sociální spravedlnost)
- Hodnoty se utváří v průběhu socializace – role věku
- Hodnoty jsou ovlivněny aktuální situací jedince (adaptace) - role příjmu
- Role vzdělání a třídy

Hypotézy

- H1: starší voliči budou preferovat vyšší míru zasahování státu do ekonomiky
- H2: s rostoucím příjmem poroste preference vyšší ekonomické svobody.
- H3: s vyšším vzděláním poroste preference vyšší ekonomické svobody
- H3X: s vyšším vzděláním poroste preference vyšší míry zasahování státu do ekonomiky
- H4: lidé se zkušeností s nezaměstnaností budou více preferovat zásahy do ekonomiky než lidé bez takové zkušenosti

- H0 proměnná nemá vliv

Proměnné

- Závisle proměnná: Míra zasahování státu do ekonomiky
 - Vytvořeno jako faktorové skóre na základě proměnných
 - Hodnoty 0 – 10 (0 – zasahování, 10 – svoboda)
 - Ke kterému z každé dvojice následujících výroků byste se spíše přiklonil?
 - Rozvoj hospodářství má být ponechán vlastnímu vývoji/má být usměrňován státem
 - Stát má zaručit, aby ten, kdo chce pracovat, dostal práci/
Kdo chce pracovat, musí se o získání práce postarat sám
 - Velkým hospodářským podnikům má stát umožnit co největší samostatnost/
Na velké hospodářské podniky má stát co nejvíce dohlížet
 - Velikost soukromého vlastnictví by nijak být omezována neměla/by nějakým způsobem být omezována měla.

Nezávisle proměnné

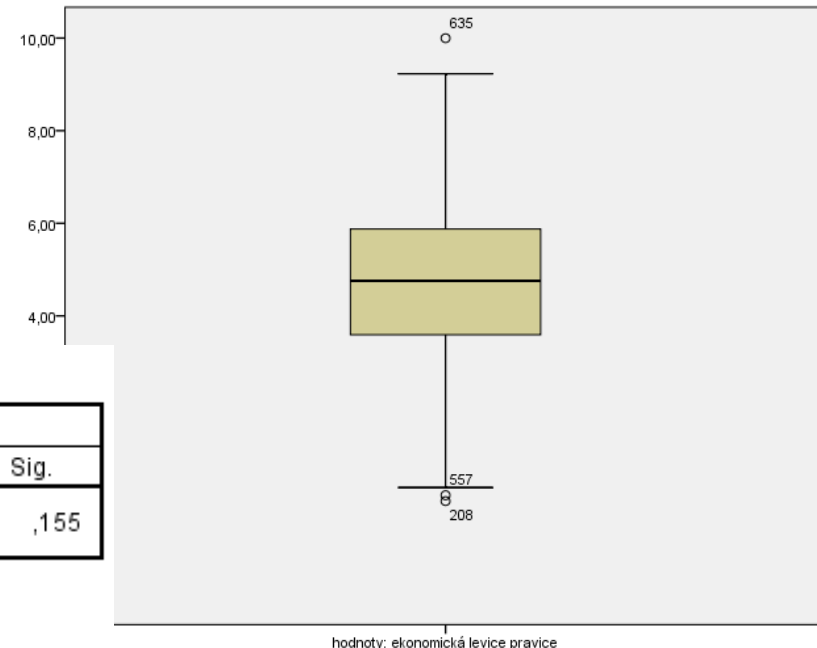
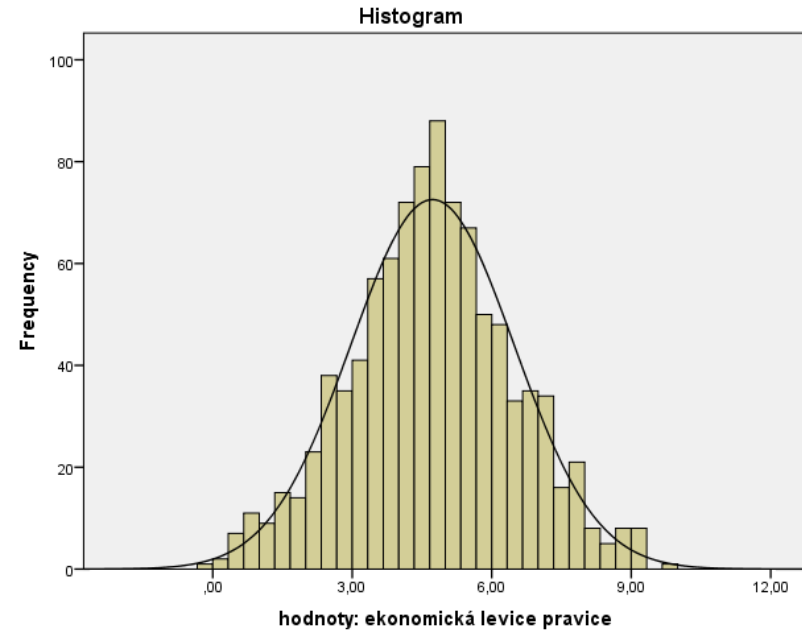
- Nezávisle / kontrolní proměnné
- Příjem: čistý příjem domácnosti
- Subjektivní hodnocení příjmu (dichotomická proměnná)
- Věk
- Nespokojenost s vnějšími podmínkami: součet proměnných ptajících se na hodnocení ekonomické a politické situace (od 0 do 10)
- Vzdělání: kategorická proměnná rekódovaná na dummy proměnné
 - ZŠ vzdělání referenční kategorií
- Nezaměstnanost: kategorická proměnná rekódovaná na dummy proměnné
 - Bez zkušenosti s nezaměstnaností jako referenční kategorie
- Zájem o politiku: 0-10

Normalita závisle proměnné

- Jinakost rozdělení
 - ovlivňuje především hodnoty signifikance
 - Zkresluje odhady parametrů
- Prvně vizuální zhodnocení pomocí histogramu
- Testy nejsou úplně nezbytné, ale jejich provedením nic nezkazíte
 - K-S a S-W
 - Ve velkých souborech lze brát s rezervou
 - **Šikmost a strmost není větší než 3x SE**

Test normality závisle proměnné

- Histogram
 - Analyze- descriptive stat- frequencies – plots
- Kolmogorův-Smirnovův test
 - Analyze – descriptive stat – explore – plots – normality plots with tests



Tests of Normality

| | Kolmogorov-Smirnov ^a | | | Shapiro-Wilk | | |
|------------------------------------|---------------------------------|-----|-------------------|--------------|-----|------|
| | Statistic | df | Sig. | Statistic | df | Sig. |
| hodnoty: ekonomická levice pravece | ,021 | 959 | ,200 [*] | ,998 | 959 | ,155 |

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Další postup

- Rekódování nezávisle proměnných
- Kontrola multikolinearity nezávisle proměnných
 - Nezávisle proměnné by mezi sebou neměly příliš souviset
 - První kontrola pomocí korelačního koeficientu
 - Není nezbytné
 - Další kontrola přímo v modelu
 - **Je nezbytné**
- Výpočet

Kontrola multikolinearity

- Analyze – correlate - bivariate

Correlations

| | | IDE.2 Věk | prijem | nespokojeno st s vnějšími podmínkami | IDE.10a Osobní čistý měsíční příjem |
|--|---------------------|-----------|---------|--|--|
| IDE.2 Věk | Pearson Correlation | 1 | -,330** | ,155** | ,163** |
| | Sig. (2-tailed) | | ,000 | ,000 | ,000 |
| | N | 1043 | 651 | 989 | 764 |
| prijem | Pearson Correlation | -,330** | 1 | -,259** | ,545** |
| | Sig. (2-tailed) | ,000 | | ,000 | ,000 |
| | N | 651 | 652 | 634 | 639 |
| nespokojenost s vnějšími podmínkami | Pearson Correlation | ,155** | -,259** | 1 | -,103** |
| | Sig. (2-tailed) | ,000 | ,000 | | ,006 |
| | N | 989 | 634 | 991 | 718 |
| IDE.10a Osobní čistý měsíční příjem | Pearson Correlation | ,163** | ,545** | -,103** | 1 |
| | Sig. (2-tailed) | ,000 | ,000 | ,006 | |
| | N | 764 | 639 | 718 | 765 |

** . Correlation is significant at the 0.01 level (2-tailed).

Naklikání modelu

- Analyze – regression – linear
- Dependent: lp_ekonom
- Independent: vek, nespokojenost, prijem, chudi_subj, učeň, sš, vš, zkus_nezam, zajem, muž, mesto
- Statistics: colinearity diagnostics, casewise diagnostics >2,5
- Plots: Y:*ZRESID, X:*ZPRED

OK

Interpretace R^2 a adj. R^2

Model Summary^b

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate | Durbin-Watson |
|-------|-------------------|----------|-------------------|----------------------------|---------------|
| 1 | ,306 ^a | ,094 | ,077 | 1,69542 | 1,605 |

a. Predictors: (Constant), příjem, vek, ...

b. Dependent Variable: hodnoty: ekonomická levice pralice

- Model vysvětluje 9,4 % variability závisle proměnné

ANOVA^a

| Model | | Sum of Squares | df | Mean Square | F | Sig. |
|-------|------------|----------------|-----|-------------|-------|-------------------|
| 1 | Regression | 172,310 | 11 | 15,665 | 5,450 | ,000 ^b |
| | Residual | 1664,315 | 579 | 2,874 | | |
| | Total | 1836,625 | 590 | | | |

a. Dependent Variable: hodnoty: ekonomická levice pralice

b. Predictors: (Constant), příjem, vek, ...

- Model je statisticky významný (tj. můžeme jeho výstupy zobecnit na populaci)

Interpretace R^2 a adj. R^2

- **neukazuje**, nakolik jsou výsledky platné v celém souboru,
- **neukazuje**, pro jaké procento voličů vztah platí
- ukazuje jak moc model vysvětluje rozptyl v závisle proměnné.
- Jak dobře model popisuje realitu (zaznamenanou v datech)
- Když je model nesignifikantní (tj. žádná z proměnných nepřispívá k vysvětlení rozptylu), tak použité nezávisle proměnné nemají efekt na závisle proměnnou,
 - nikoli, že k analýze proměnné není regrese použitelná
 - To závisí na naplnění předpokladů
 - Dobrý výsledek (vyvrací teorii)

Interpretace konstanty

- **Pro testování teorií nepodstatné**
- Nesmyslná, protože nikdo ve vzorku nemá věk 0
- Proto proměnnou věk rekódujeme
 - Odečteme 15
- V novém modelu je konstantu možné interpretovat:
- hodnota závisle proměnné očekávaná pro nejmladší občanky, spokojené s podmínkami, bez příjmu, ale subjektivně bohaté, se zš vzděláním, bez zájmu o politiku, a zkušenosti s nezaměstnaností žijící ve vsi (= 5,1)
 - Jde jen o informaci, co konstanta ukazuje, v textech ji obvykle **není potřeba interpretovat**

Coefficients^a

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. | C |
|-------|--|-----------------------------|------------|---------------------------|--------|------|---|
| | | B | Std. Error | Beta | | | |
| 1 | (Constant) | 5,135 | ,426 | | 12,064 | ,000 | |
| | vek | -,018 | ,005 | -,180 | -4,057 | ,000 | |
| | nespokojenost s vnějšími podmínkami | -,029 | ,044 | -,028 | -,651 | ,515 | |
| | příjem | ,010 | ,005 | ,090 | 1,932 | ,054 | |
| | subjektivně chudá domácnost | -,367 | ,174 | -,094 | -2,102 | ,036 | |
| | učňovské vzdělání nebo sš bez maturity | ,086 | ,210 | ,024 | ,412 | ,681 | |
| | sš s maturitou | ,181 | ,223 | ,048 | ,815 | ,415 | |
| | vysokoškolské vzdělání | ,272 | ,264 | ,055 | 1,030 | ,303 | |
| | zkušenost s nezaměstnaností | -,273 | ,186 | -,062 | -1,470 | ,142 | |
| | zajímá se o politiku | -,053 | ,149 | -,015 | -,353 | ,724 | |
| | muž | ,111 | ,145 | ,031 | ,768 | ,443 | |
| | obec nad 5000 obyvatel | ,069 | ,151 | ,018 | ,458 | ,647 | |

a. Dependent Variable: hodnoty: ekonomická levice pravice

Interpretace nestandardizovaného beta koeficientu

- **Klíčový výstup pro interpretaci**
- 2 situace
- Dummy proměnné x kardinální proměnné
- Interpretace efektu dummy proměnné:
 - nestandardizovaný koeficient ukazuje rozdíl dané kategorie oproti referenční kategorii
- Interpretace efektu kardinální proměnné
 - Při změně nezávisle proměnné o jednotku se hodnota závisle proměnné změní o hodnotu nestandardizovaného koeficient

Interpretace efektu dummy proměnné

Coefficients^a

| Model | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. | Total |
|-------|--|------------|---------------------------|-------|--------|-------|
| | B | Std. Error | Beta | | | |
| 1 | (Constant) | 5,135 | ,426 | | | |
| | vek | -,018 | ,005 | -,180 | -,4057 | ,000 |
| | nespokojenost s vnějšími podmínkami | -,029 | ,044 | -,028 | -,651 | ,515 |
| | příjem | ,010 | ,005 | ,090 | 1,932 | ,054 |
| | subjektivně chudá domácnost | -,367 | ,174 | -,094 | -2,102 | ,036 |
| | učňovské vzdělání nebo sš bez maturity | ,086 | ,210 | ,024 | ,412 | ,681 |
| | sš s maturitou | ,181 | ,223 | ,048 | ,815 | ,415 |
| | vysokoškolské vzdělání | ,272 | ,264 | ,055 | 1,030 | ,303 |
| | zkušenost s nezaměstnaností | -,273 | ,186 | -,062 | -1,470 | ,142 |
| | zajímá se o politiku | -,053 | ,149 | -,015 | -,353 | ,724 |
| | muž | ,111 | ,145 | ,031 | ,768 | ,443 |
| | obec nad 5000 obyvatel | ,069 | ,151 | ,018 | ,458 | ,647 |

a. Dependent Variable: hodnoty: ekonomická levice pravice

Interpretace efektu dummy proměnné

- Subjektivně chudý občan preferuje zásahy do ekonomiky více než subjektivně bohatý občan volič (pokud jsou ostatní sledované charakteristiky stejné) a to o 0,36 bodu
- Nebo též
- Pokud je vše ostatní shodné, pak rozdíl na škále ekonomických hodnot mezi subjektivně bohatým a chudým občanem je 0,36 bodu. Chudý občan více preferuje zásahy do ekonomiky.

- **Kategorie K se liší o $\pm XX$ od referenční kategorie** (pokud je vše ostatní shodné)
- Interpretace musí obsahovat:
 - Identifikaci kategorie proměnné, ke které je koeficient vztažen
 - Identifikaci referenční kategorie
 - Informaci o velikosti rozdílu

Interpretace efektu kardinální proměnné

Coefficients^a

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. | Co Tot |
|-------|--|-----------------------------|------------|---------------------------|--------|------|-----------|
| | | B | Std. Error | Beta | | | |
| 1 | (Constant) | 5,135 | ,426 | | 12,064 | ,000 | |
| | vek | -,018 | ,005 | -,180 | -4,057 | ,000 | |
| | nespokojenost s vnějšími podmínkami | -,029 | ,044 | -,028 | -,651 | ,515 | |
| | příjem | ,010 | ,005 | ,090 | 1,932 | ,054 | |
| | subjektivně chudá domácnost | -,367 | ,174 | -,094 | -2,102 | ,036 | |
| | učňovské vzdělání nebo sš bez maturity | ,086 | ,210 | ,024 | ,412 | ,681 | |
| | sš s maturitou | ,181 | ,223 | ,048 | ,815 | ,415 | |
| | vysokoškolské vzdělání | ,272 | ,264 | ,055 | 1,030 | ,303 | |
| | zkušenost s nezaměstnaností | -,273 | ,186 | -,062 | -1,470 | ,142 | |
| | zajímá se o politiku | -,053 | ,149 | -,015 | -,353 | ,724 | |
| | muž | ,111 | ,145 | ,031 | ,768 | ,443 | |
| | obec nad 5000 obyvatel | ,069 | ,151 | ,018 | ,458 | ,647 | |

a. Dependent Variable: hodnoty: ekonomická levice pravice

Interpretace efektu kardinální proměnné

- Pokud má občan A o 1 000 Kč vyšší příjem než volič B a vše ostatní je shodné, pak volič A preferuje o 0,01 bodu ekonomickou svobodu
- **Nebo lépe**
- S růstem příjmu o 1000 Kč (pokud vše ostatní zůstává shodné) preference ekonomické svobody vzroste o 0,01 bodu
- Lze násobit
 - Pokud příjem vzroste o 10 000 Kč , pak preference ekonomické svobody vzroste o 0,1 bodu
 - Pokud příjem vzroste o 100 000 Kč , pak preference ekonomické svobody vzroste o 1 bod

Coefficients^a

| Model | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
|-------|--|------------|---------------------------|-------|--------|
| | B | Std. Error | Beta | | |
| 1 | (Constant) | 5,135 | ,426 | | |
| | vek | -,018 | ,005 | -,180 | -4,057 |
| | nespokojenost s vnějšími podmínkami | -,029 | ,044 | -,028 | -,651 |
| | příjem | ,010 | ,005 | ,090 | 1,932 |
| | subjektivně chudá domácnost | -,367 | ,174 | -,094 | -2,102 |
| | učňovské vzdělání nebo sš bez maturity | ,086 | ,210 | ,024 | ,412 |
| | sš s maturitou | ,181 | ,223 | ,048 | ,815 |
| | vysokoškolské vzdělání | ,272 | ,264 | ,055 | 1,030 |
| | zkušenost s nezaměstnaností | -,273 | ,186 | -,062 | -1,470 |
| | zajímá se o politiku | -,053 | ,149 | -,015 | -,353 |
| | muž | ,111 | ,145 | ,031 | ,768 |
| | obec nad 5000 obyvatel | ,069 | ,151 | ,018 | ,458 |

a. Dependent Variable: hodnoty: ekonomická levice pravice

Hodnocení signifikance

- Zobecňování výsledků na populaci
- Obvyklá hranice sig. $< 0,05$
- Potom považujeme efekt za signifikantní na hladině významnosti 95 %
- Nic nám nebrání zvolit si jinou hladinu významnosti (např. 90%, 99% nebo 99,99%)
- S nižší hladinou roste riziko, že budeme za platný považovat i efekt který v populaci neplatí
- S vyšší hladinou vyšší riziko že budeme za neplatný považovat i efekt, který v populaci platí

Následná kontrola

- Outlieři
- Homogenita rozptylu reziduí (homoskedascita)
- multikolinearita

Honocení multikolinearity

- VIF
- Arbitární hranice: 3 (může být problém) – 5 (jistý problém)
- A zároveň podobné hodnoty v dimenzích
- Proměnné levice a pravice
 - V pořádku, neboť se jedná o dummy proměnné vytvořené z jediné kategorické proměnné

Coefficients^a

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. | Collinearity Statistics | |
|-------|--|-----------------------------|------------|---------------------------|--------|------|-------------------------|-------|
| | | B | Std. Error | Beta | | | Tolerance | VIF |
| 1 | (Constant) | 5,135 | ,426 | | 12,064 | ,000 | | |
| | vek | -,018 | ,005 | -,180 | -4,057 | ,000 | ,793 | 1,262 |
| | nespokojenost s vnějšími podmínkami | -,029 | ,044 | -,028 | -,651 | ,515 | ,859 | 1,161 |
| | příjem | ,010 | ,005 | ,090 | 1,932 | ,054 | ,723 | |
| | subjektivně chudá domácnost | -,367 | ,174 | -,094 | -2,102 | ,036 | ,790 | |
| | učňovské vzdělání nebo sš bez maturity | ,086 | ,210 | ,024 | ,412 | ,681 | ,471 | |
| | sš s maturitou | ,181 | ,223 | ,048 | ,815 | ,415 | ,457 | |
| | vysokoškolské vzdělání | ,272 | ,264 | ,055 | 1,030 | ,303 | ,555 | |
| | zkušenost s nezaměstnaností | -,273 | ,186 | -,062 | -1,470 | ,142 | ,890 | |
| | zajímá se o politiku | -,053 | ,149 | -,015 | -,353 | ,724 | ,879 | 1,138 |
| | muž | ,111 | ,145 | ,031 | ,768 | ,443 | ,931 | 1,074 |
| | obec nad 5000 obyvatel | ,069 | ,151 | ,018 | ,458 | ,647 | ,960 | 1,042 |

a. Dependent Variable: hodnoty: ekonomická levice pravice

Outlieři

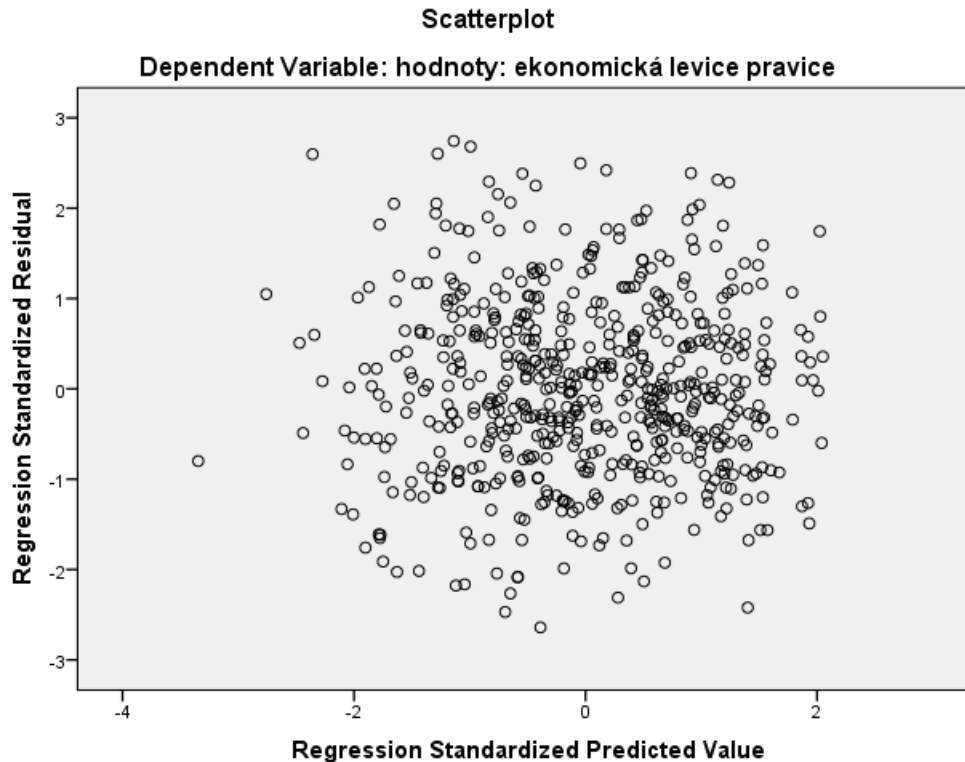
- Např. pro případ č. 70 očekáváme, že nebude mít vyhraněný názor, ale přitom reálně jde o velmi levicového občana
- Podobně případ 105, ten je ale pravicový
- Můžeme vyřadit a zjistit, co to udělá s výsledky

Casewise Diagnostics^a

| Case Number | Std. Residual | hodnoty: ekonomická levice pravice | Predicted Value | Residual |
|-------------|---------------|--|--------------------|----------|
| 70 | -2,177 | ,39 | 4,0863 | -3,69304 |
| 105 | 2,283 | 9,23 | 5,3563 | 3,87237 |
| 127 | -2,420 | 1,34 | 5,4433 | -4,10519 |
| 156 | 2,015 | 60 | 3,0146 | 3,41904 |

Homoskedascita

- V reziduích by neměl být žádný zřetelný vzorec



heteroskedascita

- Příklad situace kdy homoskedascita není v pořádku

