

Statistické zpracování dat 7.přednáška

Mgr. Radmila Krkošková, Ph.D.



**SLEZSKÁ
UNIVERZITA**

OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ



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



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY



Vícenásobná lineární regresní analýza (2)



Obsah přednášky



**SLEZSKÁ
UNIVERZITA**
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

- **Heteroskedasticita (H-S)**
- **Testy H-S a její odstraňování**
- **Autokorelace**
- **Nominální proměnné**



Heteroskedasticita



SLEZSKÁ
UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

Rozptyl náhodné chyby u je konstantní, tj.

$$\text{Var}(u) = \sigma^2$$

Graficky: Hodnoty jsou rozptýleny ve stejně širokém pásu kolem regresní funkce (regresní nadroviny)

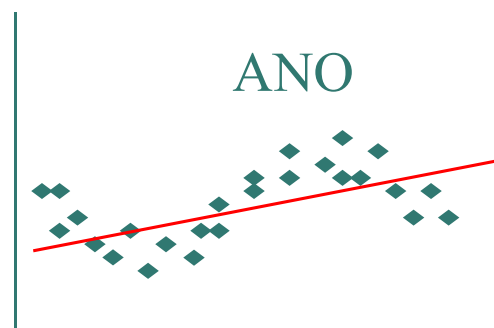
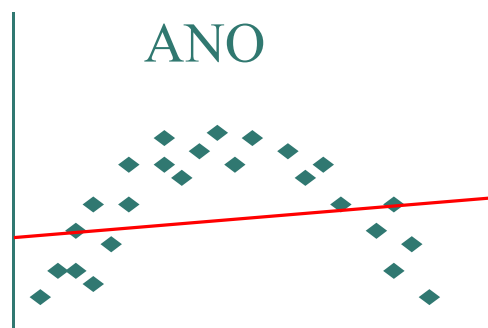
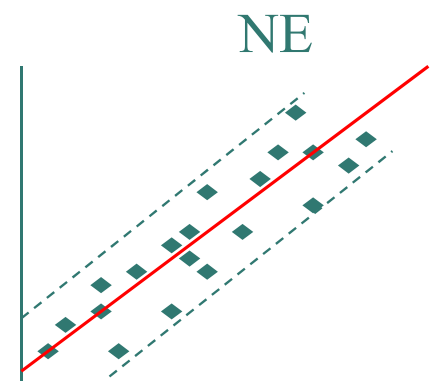
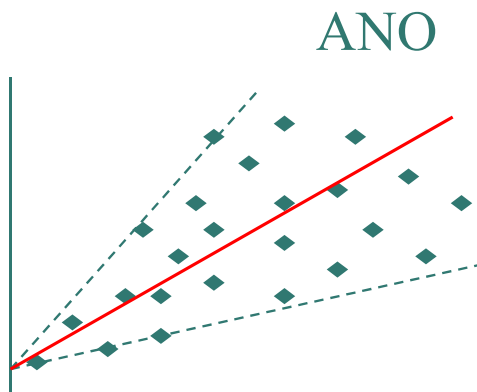
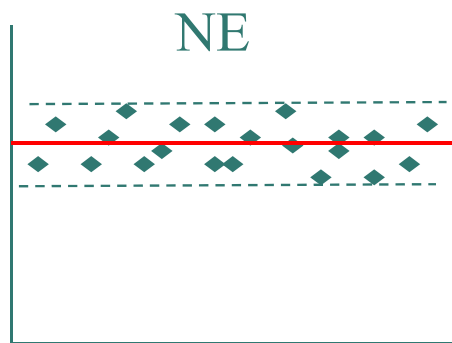
Otázky:

1. Co je podstatou heteroskedasticity (H-S)?
2. Jaké jsou důsledky H-S?
3. Jak zjišťovat H-S v dané situaci?
4. Jak odstraňovat H-S?



Jak vypadá H-S?

Grafická analýza reziduí:



Co je podstatou H-S?



SLEZSKÁ
UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

Jedná se o rozptyl náhodné chyby u_i v regresním (populačním)

modelu, např. $Y_i = B_0 + B_1 X_{i1} + B_2 X_{i2} + u_i$



Co je podstatou H-S?



SLEZSKÁ
UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

Některé důvody nekonstantnosti rozptylu:

1. Učení se z chyb: rozptyl počtu chyb se s rostoucím časem zmenšuje
2. S rostoucím věkem roste rozptyl příjmů zaměstnanců
3. S lepšími technikami sběru dat klesá rozptyl chyb v datech
4. S přítomností odlehlých hodnot roste rozptyl



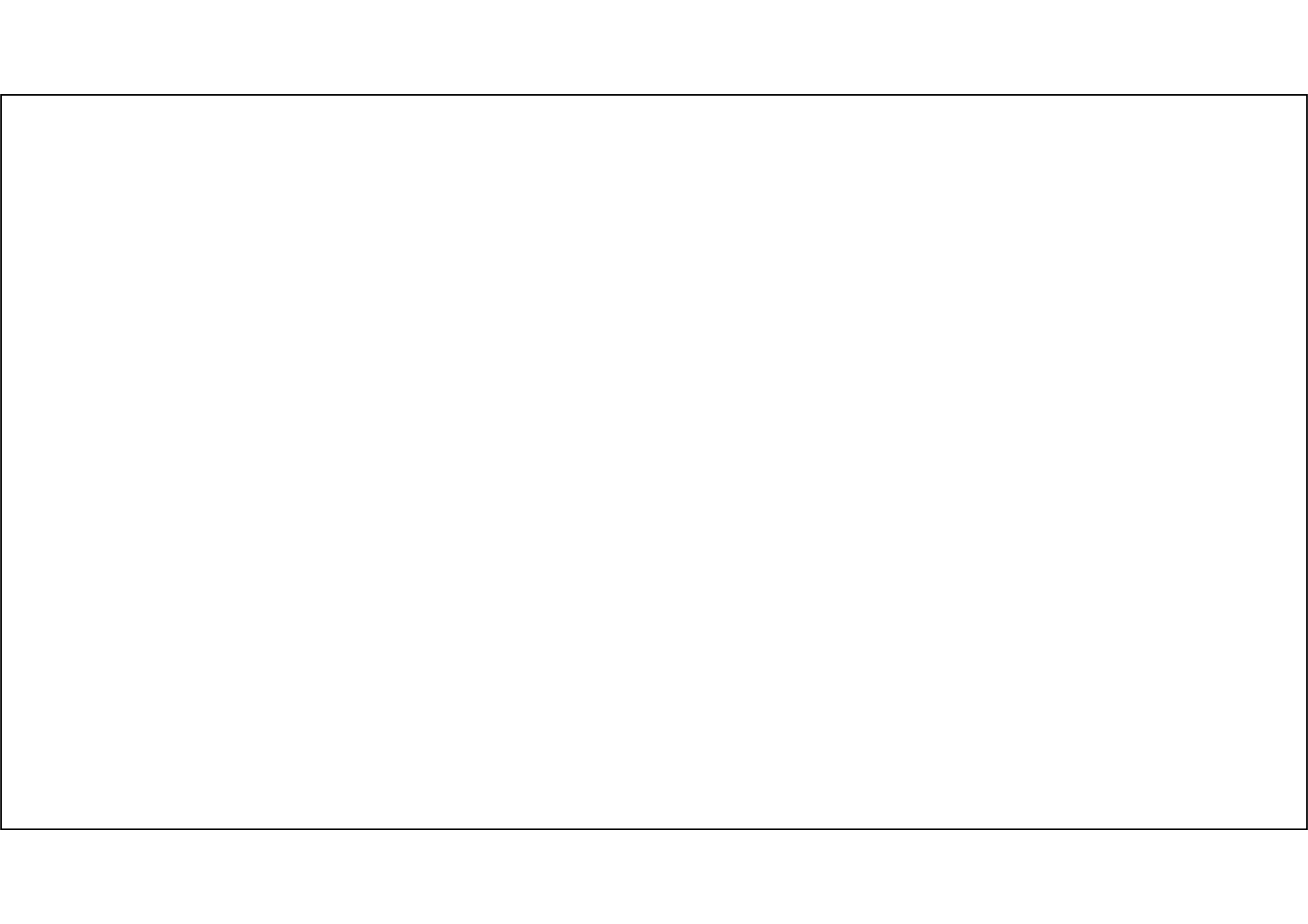
Co je podstatou H-S?



**SLEZSKÁ
UNIVERZITA**
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

5. U špatně specifikovaného modelu dochází k proměnlivosti rozptylu
6. Šikmost rozdělení vysvětlujících proměnných zvětšuje rozptyl
7. Panelová (průřezová) data mívají proměnlivý rozptyl





Jaké jsou důsledky H-S?



SLEZSKÁ
UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

- Odhady regresních koeficientů již nejsou nevychýlené a nemusí mít nejmenší rozptyl
- Testy regresních koeficientů jsou nespolehlivé (nepravdivé) → nebezpečí špatných (zavádějících) závěrů
- Nesplnění předpokladu homoskedasticity činí obvyklé závěry o regresním modelu ***nepoužitelnými !!!***



Testy H-S



SLEZSKÁ
UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

Bartletův test H-S: (zjednodušený G-Q test)

Vychází z rozdělení dat podle velikosti proměnné X do dvou částí (vzorků): $X_i \leq D$ a $X_i > D$... data uspořádána podle X , $D \cong \bar{X}$

Testuje se hypotéza o rovnosti rozptylů u_i v obou vzorcích (v Excelu: Analýza dat, Dvouvýběrový F-test pro rozptyl,...)

Pokud se hypotéza o rovnosti rozptylu zamítá, potom se hypotéza o H-S přijímá (a obráceně)



Příklad – postup při testování H-S



SLEZSKÁ
UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

Testujte H-S v datech Příjmů (X) a Spotřebních výdajů (Y)
(viz soubor [B test H-S nový.xls](#))

Krok 1: Sestavte lin. regresní model: $Y = 9,29 + 0,64 * X + e$
(v Excelu: Nástroje → Analýza dat... → Regrese...)

Krok 2: Uspořádejte sloupce X a e podle rostoucího X



Příklad – postup při testování H-S



SLEZSKÁ
UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

Krok 3: Rozdělte soubor na dva stejně veliké (event. prostřední prvek vypusťte): oba výběry mají po 15 údajích

Krok 4: Proveďte Dvouvýběrový F test pro rozptyl reziduí e

(v Excelu: Nástroje → Analýza dat → Dvouvýběrový F-test pro rozptyl)



Příklad - výsledek



SLEZSKÁ
UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

Výsledek: $p\text{-hodnota} = 0,01 < 0,05 \Rightarrow$

oba výběry mají různý rozptyl,

tzn. nulovou hypotézu zamítáme (na hladině 5%).

Je přítomna H-S !



Metody odstranění H-S



SLEZSKÁ
UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

- Myšlenku transformace H-S dat na data bez H-S ukážeme na **jednoduchém regresním modelu**
- **Regresní model:** $Y_i = B_0 + B_1 X_i + u_i$ (*)
- Transformujeme vhodně H-S data na data bez H-S
- Řešíme nový **homoskedastický** regr. model (tj. bez H-S) !!!



Rozptyl náhodné chyby závisí na regresoru



SLEZSKÁ
UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

Případ 1: Předpokládáme $E(u_i^2) = \sigma_i^2 = \sigma^2 X_i$
(σ^2 je neznámá konstanta!)

Provedeme transformaci (substituci) do (*):

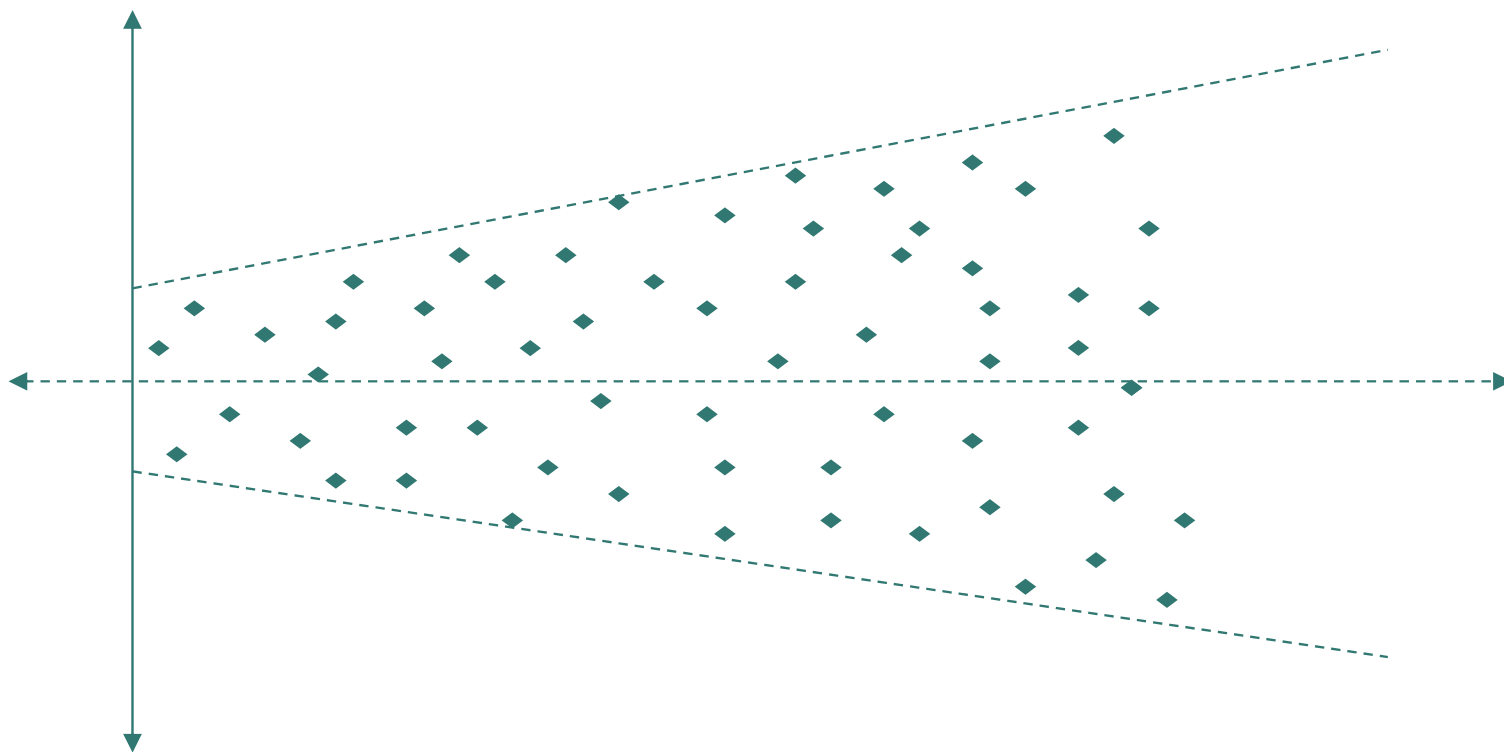
$$\frac{Y_i}{\sqrt{X_i}} = B_0 \left(\frac{1}{\sqrt{X_i}} \right) + B_1 \sqrt{X_i} + \frac{u_i}{\sqrt{X_i}} \quad (**)$$

Transformovaný model (**) je model bez H-S,
neboť

$$E(v_i^2) = E\left(\frac{u_i^2}{X_i}\right) = \frac{E(u_i^2)}{X_i} = \frac{\sigma^2 X_i}{X_i} = \sigma^2$$



Případ 1: $\sigma_i^2 = \sigma^2 X_i$



Úprava dat



- Úprava vstupních dat, např.: $Y_i = \text{Obrat}_i$, $X_i = \text{VaV}_i$

$$Y_i' = \frac{Y_i}{\sqrt{X_i}} \quad X_i' = \frac{1}{\sqrt{X_i}} \quad X_i'' = \sqrt{X_i}$$

- V programu (Excel, SPSS) se zaklikne volba:
„konstanta je nula“, neboť model má tvar

$$Y_i' = B_0 X_i' + B_1 X_i'' + v_i \quad (+)$$

- Odhady koeficientů modelu (+) jsou již v pořádku, v (+) lze pak pro X predikovat Y' a zpětnou transformací též:

$$Y = Y' \sqrt{X}$$



Rozptyl náhodné chyby závisí na kvadrátu regresoru



SLEZSKÁ
UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

Případ 2: Předpokládáme $E(u_i^2) = \sigma_i^2 = \sigma^2 X_i^2$

Provedeme transformaci (substituci) do (*):

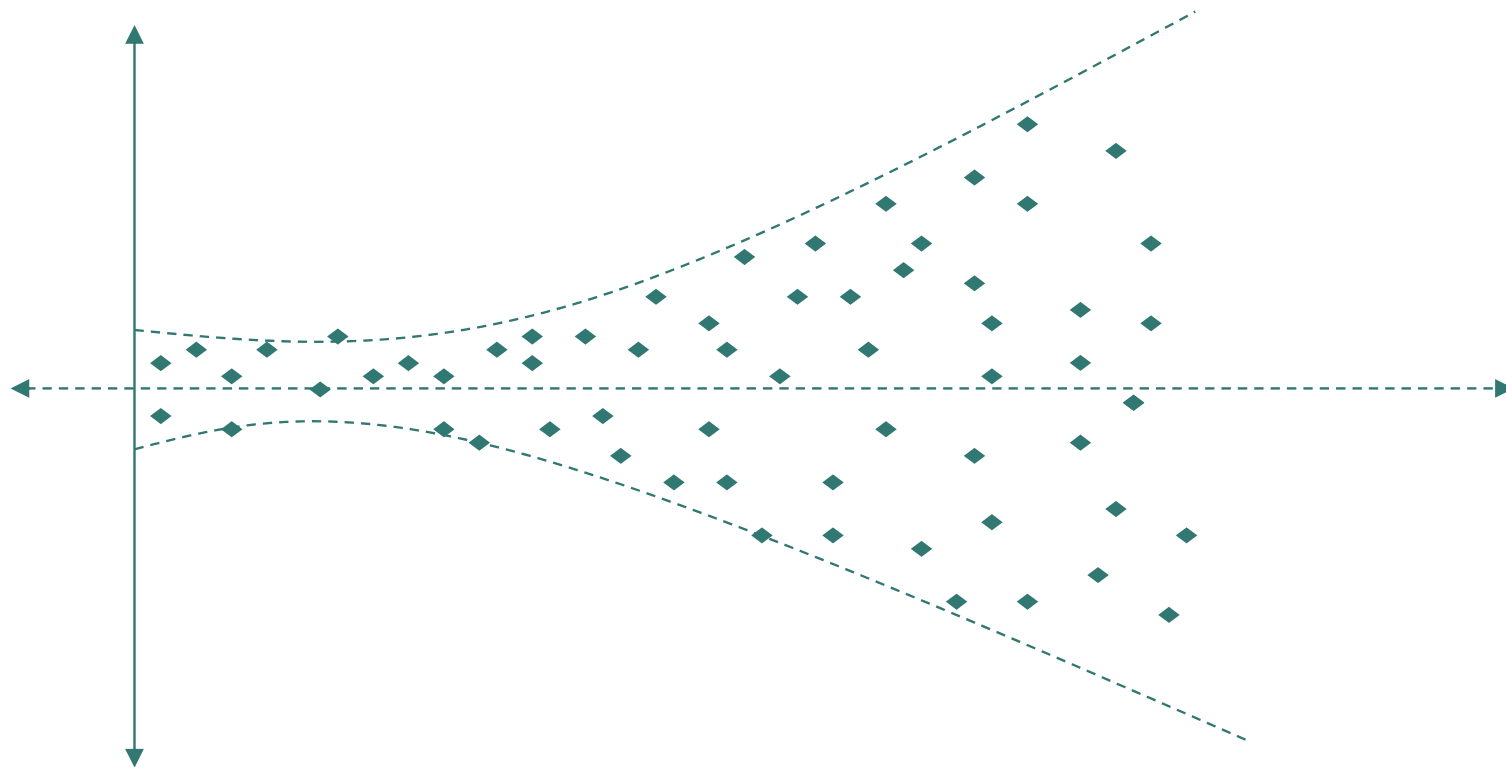
$$\frac{Y_i}{X_i} = B_0 \left(\frac{1}{X_i} \right) + B_1 + \frac{u_i}{X_i} \quad (***)$$

Transformovaný model (***) je opět model bez H-S, neboť

$$E(v_i^2) = E\left(\frac{u_i^2}{X_i^2}\right) = \frac{E(u_i^2)}{X_i^2} = \frac{\sigma^2 X_i^2}{X_i^2} = \sigma^2$$



Případ 2: $\sigma_i^2 = \sigma^2 X_i^2$



Log-lineární model



SLEZSKÁ
UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

- Namísto spekulací o tvaru závislostí σ_i^2 je někdy výhodnější provést reformulaci modelu
- Osvědčená reformulace na log-lineární model:

$$Y_i = B_0 + B_1 \ln X_i + u_i$$

- Důvod reformulace: Logaritmická funkce zmenšuje měřítko variability modelu!



Postup při identifikaci a odstraňování H-S



**SLEZSKÁ
UNIVERZITA**
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

1. Zobrazit data $X (X_i)$ vers. Y (bodový graf)
2. Vybrat vhodný regresní model VLR (resp. JLRM)
3. Řešit JLRM + vypočíst rezidua



Postup při identifikaci a odstraňování H-S



SLEZSKÁ
UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

4. Provést Bartletův test HS:

- Seřadit data podle X , rozdělit na 2 stejné části (horní-dolní, ev, vynechat prostřední hodnotu)
- Provést Dvouvýběrový F-test pro rozptyl
- Přijmout/zamítnout H_0 o rovnosti rozptylů
- V případě zamítnutí H_0 **přijmout hypotézu o přítomnosti HS v modelu.**



Postup při identifikaci a odstraňování H-S



**SLEZSKÁ
UNIVERZITA**
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

5. Zvolit model pro odstranění HS
(lineární, kvadratický, log.)
6. Transformovat data podle zvoleného modelu
7. Provést JLR analýzu pro zvolený model, vypočítat
8. Provést Bartletův test HS pro zvolený model



Postup při identifikaci a odstraňování H-S



SLEZSKÁ
UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

9. V případě výsledku Bartlet. testu:

„Nepřítomnost HS“ ukončit analýzu s výsledkem:

Ve zvoleném modelu není HS přítomna.

Model lze pak použít pro analytické účely, např. pro predikci apod. Po zpětné transformaci výsledků též pro původní model.

10. V případě výsledku Bartlet. testu: „Přítomnost HS“ zvolit nový model pro odstranění HS.



Autokorelace



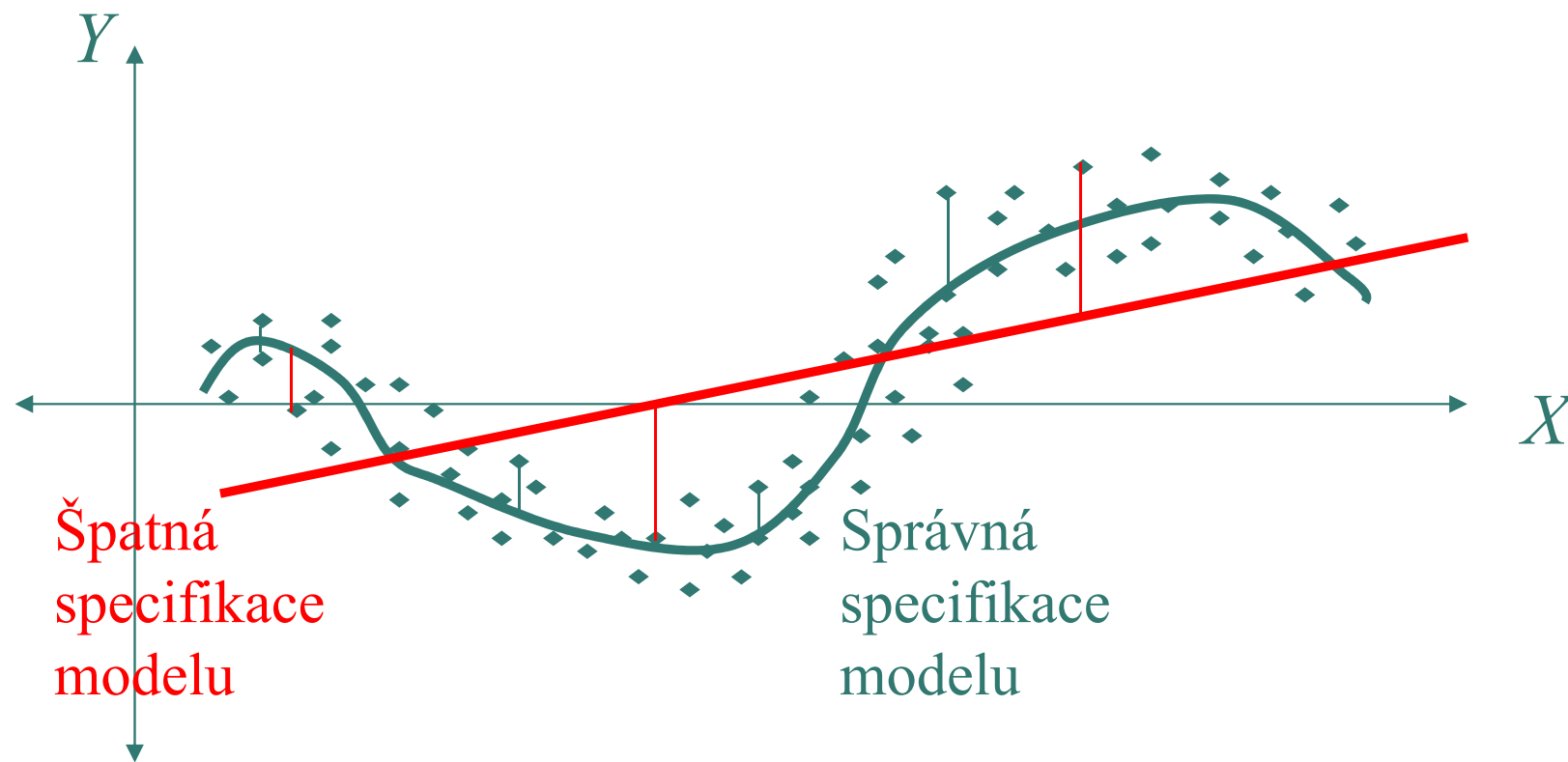
SLEZSKÁ
UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

AC = **korelace** mezi pozorováními uspořádanými **v čase** (data jsou časové řady) nebo **v prostoru** (data jsou průřezová, tj. v jednom časovém okamžiku/intervalu)

$$E(u_i, u_j) = 0, \quad i \neq j$$



Autokorelace



Test nulovosti autokorelací reziduí



- **Autokorelační funkce ρ_k :** $\rho_k = Cor(e_t, e_{t-k})$

Má platit: $\rho_k = 0$ pro $k \neq 0$

- **Odhady autokorelační funkce r_k :**

$$r_k = \frac{\sum_t e_t e_{t-k}}{\sum_t e_t^2}$$

Nulová hypotéza $H_0: \rho_k = 0$

Testové kritérium: $T_k = \frac{r_k \sqrt{n-2}}{\sqrt{1-r_k^2}}$

Obor přijetí: $A = (-u_{1-\alpha/2}, u_{1-\alpha/2})$

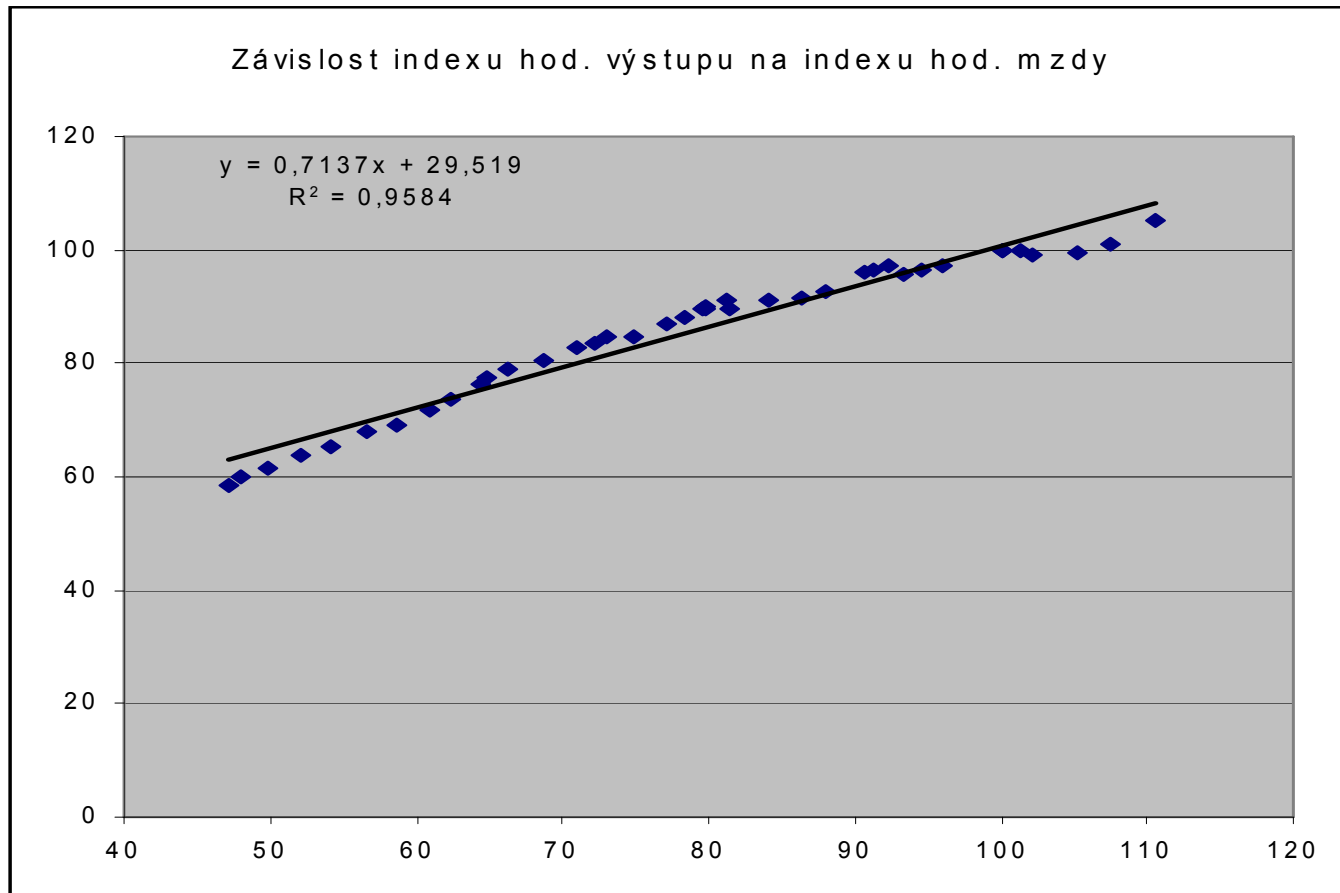
Má platit: $T_k \in A$ pro $k > 0$



Příklad



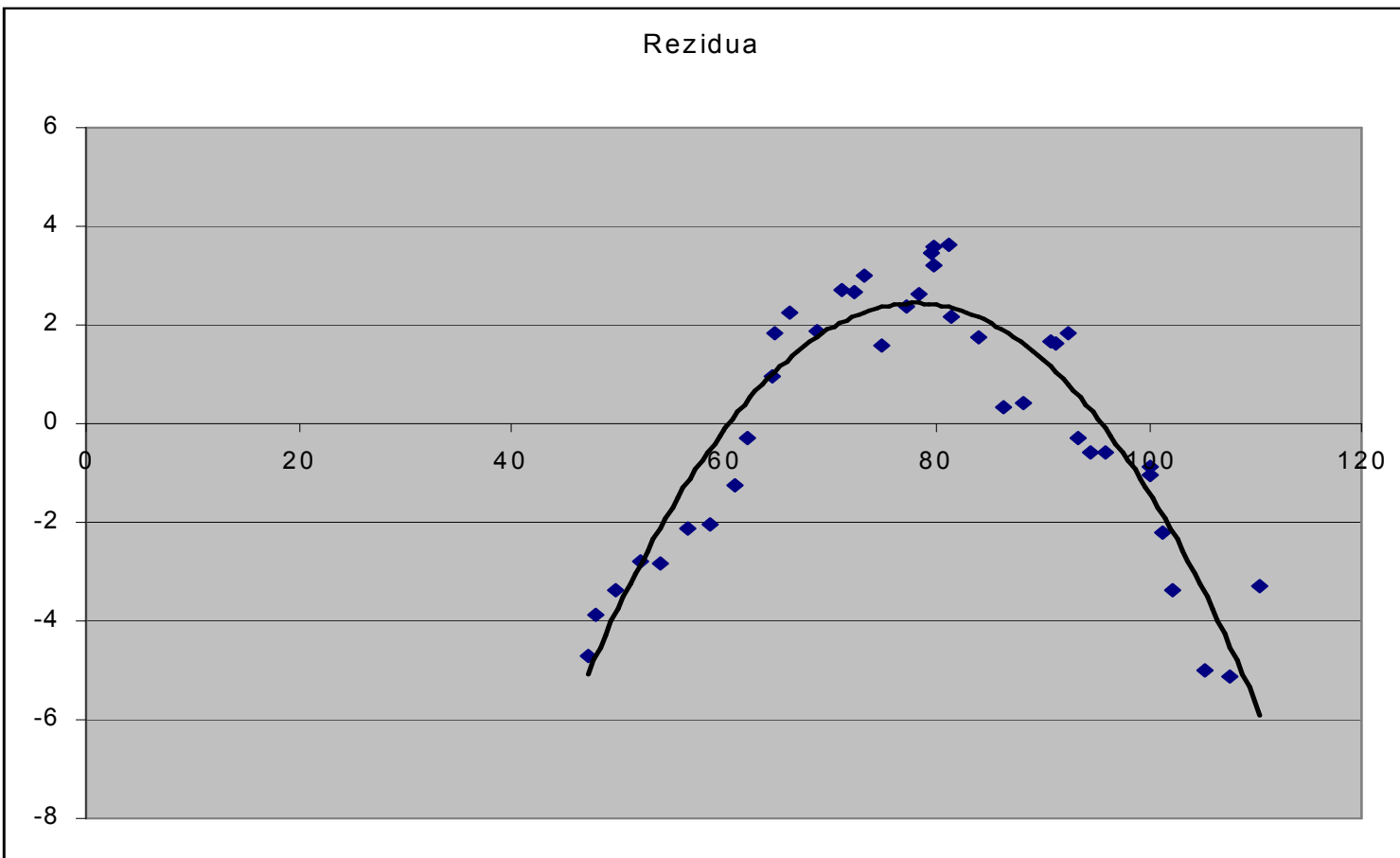
**SLEZSKÁ
UNIVERZITA**
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ



Příklad – testy autokorelačních koeficientů



**SLEZSKÁ
UNIVERZITA**
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ



Příklad – testy autokorelačních koeficientů



SLEZSKÁ
UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

r1	r2	r3	r4	r5	r6	r7	r8	r9	r10
0,929586	0,838779	0,740086	0,647914	0,497907	0,317624	0,069975	-0,087644	-0,200395	-0,284443

T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
13,10423	8,004856	5,718251	4,419849	2,983293	1,740557	0,364495	-0,457173	-1,062842	-1,54169

Výsledek testu H0:

z	z	z	z	z	p	p	p	p	p
---	---	---	---	---	---	---	---	---	---

A = (-1,96, +1,96)

z = zamítáme

p = přijímáme





Děkuji Vám za pozornost!!!

