

# Základy ekonometrie

## II. Netechnický úvod do regrese

# Obsah tématu

- 1 Regrese
  - Úvod do regrese
  - Příklady
- 2 Jednoduchý regresní model
  - Příklady a základní pojmy
  - Odhady parametrů modelu
  - Statistické aspekty
- 3 Model vícenásobné regrese
  - Odhad a interpretace
  - Specifické problémy vícenásobné regrese
- 4 Umělé proměnné

# Obsah tématu

- 1 Regrese
  - Úvod do regrese
  - Příklady
- 2 Jednoduchý regresní model
  - Příklady a základní pojmy
  - Odhady parametrů modelu
  - Statistické aspekty
- 3 Model vícenásobné regrese
  - Odhad a interpretace
  - Specifické problémy vícenásobné regrese
- 4 Umělé proměnné

# Obsah tématu

## 1 Regrese

- Úvod do regrese
- Příklady

## 2 Jednoduchý regresní model

- Příklady a základní pojmy
- Odhady parametrů modelu
- Statistické aspekty

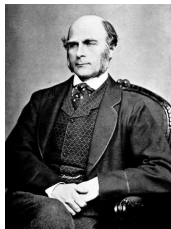
## 3 Model vícenásobné regrese

- Odhad a interpretace
- Specifické problémy vícenásobné regrese

## 4 Umělé proměnné

# Historie pojmu „regrese“?

- Sir Francis Galton – bratranec Charlese Darwina, školitel Karla Pearsona.



Francis Galton (1822–1911)



Karl Pearson (1857–1936)

- Empirické zjištění (1886), potvrzena Pearsonem:
  - tendence vyšších rodičů mít vyšší děti a tendence menších rodičů mít menší děti;
  - v průměru však děti vyšších rodičů jsou menší než jejich rodiče a děti menších rodičů jsou větší než jejich rodiče.



# Pojetí regrese v současnosti

- Obecně závislost mezi veličinami (lineární, nelineární, dvě i více veličin).
- Regresní analýza – studium závislosti jedné proměnné, **závisle proměnné**, na jedné či více ostatních proměnných, **vysvětlujících proměnných**.
- Účel – odhad a/nebo predikce populační střední hodnoty či průměrné hodnoty závisle proměnné v závislosti na známých (pevně daných) hodnotách vysvětlujících proměnných (v opakujících se výběrech).
- $E(Y|X)$  jako funkce  $x$  ( $Y$  závisle proměnná,  $X$  vysvětlující proměnné,  $x$  jejich realizace).

# Obsah tématu

## 1 Regrese

- Úvod do regrese
- Příklady

## 2 Jednoduchý regresní model

- Příklady a základní pojmy
- Odhady parametrů modelu
- Statistické aspekty

## 3 Model vícenásobné regrese

- Odhad a interpretace
- Specifické problémy vícenásobné regrese

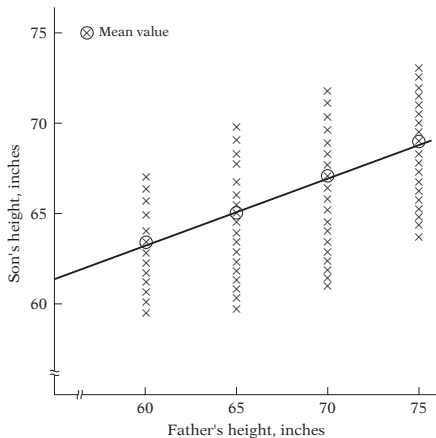
## 4 Umělé proměnné



## Příklad 1 – Galtonova regrese

- Galton – otázka stability v rozdělení výšky populace.
- V současném pohledu – otázka jak se např. *průměrná* výška chlapců mění v závislosti na dané výšce otců.
- Predikce průměrné výšky (očekávaná výška) synů při znalosti výšky jejich otců.
- Obrázek – rozdělení výšky synů v hypotetické populaci vzhledem k daným či *fixním* hodnotám výšky otců.
- **Regresní přímka** – ukazuje jak průměrná výška synů roste s výškou otců (obecné tendence).

## Příklad 1 – Galtonova regrese (obrázek s daty)

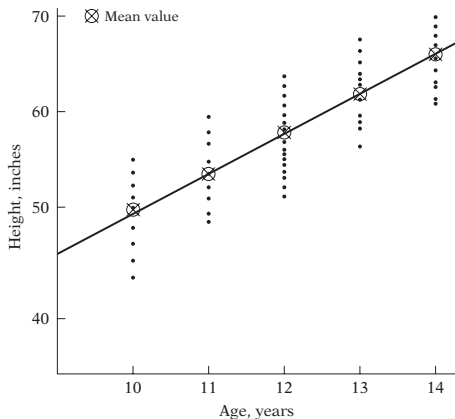


Zdroj: Gujarati, Porter (2009) – Basic econometrics.

## Příklad 2 – Výška a věk chlapců v populaci

- Obrázek — rozdělení výšky chlapců v hypotetické populaci měřené v *pevně* daném věku.
- Ne všichni chlapci v daném věku mají stejnou výšku  $\times$  v průměru se výška zvyšuje s věkem (do určitého věku).
- Zobrazení závislosti – odpovídající regresní přímka.
- Při znalosti věku chlapce,  $X = x$ , můžeme predikovat průměrnou (očekávanou) výšku,  $Y$ , tedy  $E(Y|X = x)$ .

## Příklad 2 – Výška a věk chlapců v populaci (obrázek s daty)



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

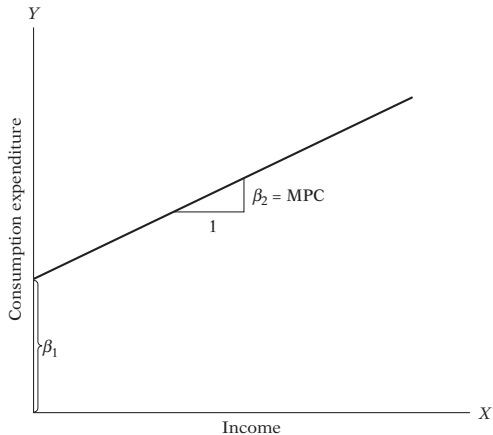
## Příklad 3 – Spotřební funkce

- Analýza závislosti výdajů na osobní spotřebu,  $Y$ , na osobním důchodu,  $X$ .
- Ekonomický model (Keynesiánská spotřební funkce):

$$Y = \beta_1 + \beta_2 X.$$

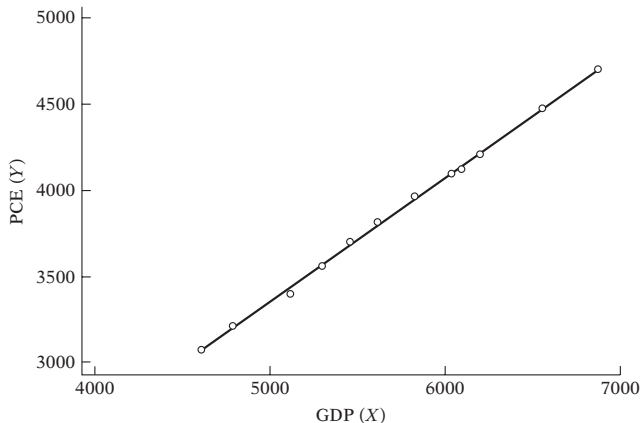
- Odhad mezního sklonu ke spotřebě (MPC) – odhad parametru  $\beta_2$ .
- MPC = průměrná (očekávaná) změna ve výdajích na osobní spotřebu při změně reálného důchodu o jednotku.
- Data – roční.

# Příklad 3 – Spotřební funkce (model)



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

## Příklad 3 – Spotřební funkce (obrázek s daty)



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

## Příklad 4 – Monopol a cenová elasticita

- Monopol – analýza reakce poptávky na změnu ceny.
- Experiment pro odhad *cenové elasticity* poptávky po produktu.
- Nalezení „nejziskovější“ ceny.



## Příklad 5 – Phillipsova křivka

- Analýza míry změny mezd ve vztahu k míře nezaměstnanosti.
- Predikce průměrné změny mezd při dané úrovni nezaměstnanosti.
- Pohled na inflační procesy v ekonomice (mzdová a cenová inflace).
- Ilustrativní příklad – v realitě komplexnější modely.

# Příklad 5 – Phillipsova křivka (obrázek s daty)

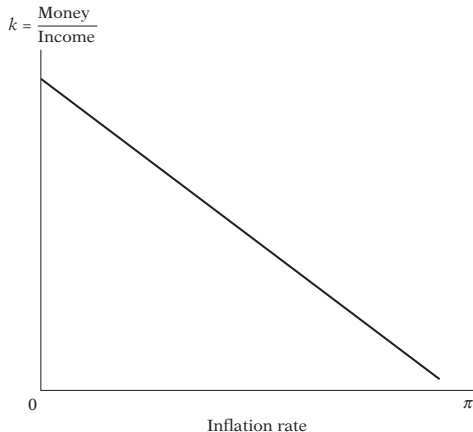


Zdroj: Gujarati, Porter (2009) – Basic econometrics.

## Příklad 6 – Poptávka po penězích

- Monetární ekonomie – za jinak neměnných podmínek s růstem inflace,  $\pi$ , se snižuje poměr,  $k$ , důchodu drženého v podobě peněz.
- Obrázek – model závislosti.
- Kvantitativní analýza vztahu  $\rightarrow$  predikce množství peněz (v podílu k důchodu), které lidé chtějí držet při různých mírách inflace.

# Příklad 6 – Poptávka po penězích (model)



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

## Příklad 7 – Výdaje na reklamu

- Marketingové oddělení – vztah výdajů na reklamu a poptávky po produktu společnosti.
- Hledání *elasticity poptávky* na výdaje na reklamu (procentní změna poptávky v závislosti na jednoprocentní změně výdajů na reklamu).
- Nalezení „optimálního“ rozpočtu na reklamu.

## Příklad 8 – Výnosy v zemědělství

- Závislost výnosů (např. pšenice nebo jiného obilí) na teplotách, množství srážek, slunečním svitu a hnojení.
- Predikce průměrného výnosu při daných informacích o vysvětlujících proměnných.

# Obsah tématu

- 1 Regrese
  - Úvod do regrese
  - Příklady
- 2 Jednoduchý regresní model
  - Příklady a základní pojmy
  - Odhady parametrů modelu
  - Statistické aspekty
- 3 Model vícenásobné regrese
  - Odhad a interpretace
  - Specifické problémy vícenásobné regrese
- 4 Umělé proměnné

# Obsah tématu

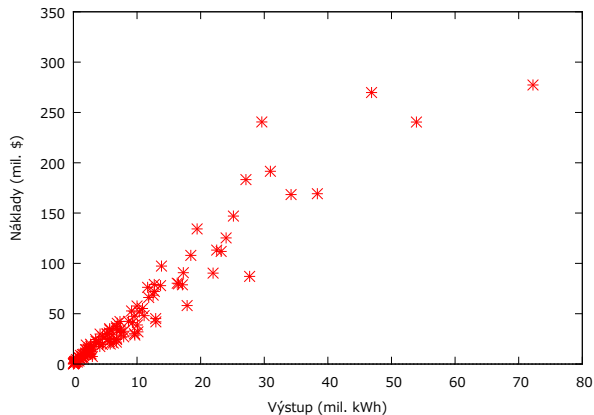
- 1 Regrese
  - Úvod do regrese
  - Příklady
- 2 **Jednoduchý regresní model**
  - **Příklady a základní pojmy**
  - Odhady parametrů modelu
  - Statistické aspekty
- 3 Model vícenásobné regrese
  - Odhad a interpretace
  - Specifické problémy vícenásobné regrese
- 4 Umělé proměnné



## Příklad nákladů produkce v odvětví výroby elektřiny

- Průřezová data pro 123 společností v USA v roce 1970.
- Analýza nákladové funkce a faktorů ovlivňujících celkové náklady.
- Závislost nákladů a výstupu firem.

# Příklad nákladů produkce – data



Obrázek: Bodový graf výstupu vzhledem k nákladům.

# Lineární regresní model

- Lineární vztah mezi náklady,  $Y$ , a výstupem,  $X$ :

$$Y = \alpha + \beta X.$$

- Neznámé **koeficienty (parametry)** modelu:  $\alpha$  ... úroňová konstanta,  $\beta$  ... parametr sklonu (vliv proměnné  $X$  na  $Y$ ).
- **Náhodná složka (chybový člen)**,  $\epsilon$  – chyby měření, opomenutí vysvětlující proměnné, neměřitelné veličiny ovlivňující náklady  $\Rightarrow$  pozorování neleží přesně na teoretické přímce.

$$Y = \alpha + \beta X + \epsilon$$

# Lineární regresní model – terminologie

Tabulka: Popis proměnných v jednoduchém LRM

$Y$	$X$
závisle proměnná	nezávisle proměnná
vysvětlovaná proměnná	vysvětlující proměnná
kontrolovaná proměnná	kontrolní (řídící) proměnná
predikovaná proměnná	prediktor
regresant	regresor

# Volba směru závislosti

- Volba závisle proměnné a vysvětlujících proměnných.
- Problém kauzality a interpretace výsledků.
- Kauzalita z ekonomické teorie (mikroekonomie a nákladová funkce) nebo intuice (ceny domů a jejich charakteristiky).
- Nejasná kauzalita – mzdová a cenová inflace (ekonometrický přístup pomocí Grangerovských kauzalit).

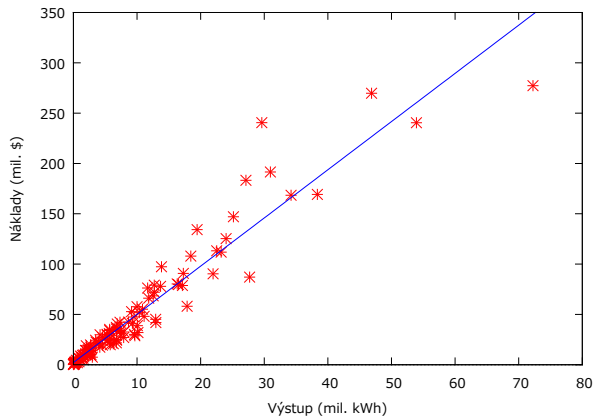
# Obsah tématu

- 1 Regrese
  - Úvod do regrese
  - Příklady
- 2 **Jednoduchý regresní model**
  - Příklady a základní pojmy
  - **Odhady parametrů modelu**
  - Statistické aspekty
- 3 Model vícenásobné regrese
  - Odhad a interpretace
  - Specifické problémy vícenásobné regrese
- 4 Umělé proměnné

# Odhad parametrů a přímka vyrovnání

- Regresní analýza – odhady neznámých parametrů:  $\hat{\alpha}$ ,  $\hat{\beta}$ .
- Spousta možností odhadů  $\Rightarrow$  spousta prokládajících přímek.
- Nejlepší proložení dat?

# Příklad nákladů produkce – data a nejlepší vyrovnání



**Obrázek:** Bodový graf výstupu vzhledem k nákladům s regresní přímkou vyrovnání.



# Náhodná složka a reziduum

- Jednotlivá pozorování:

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

- Náhodná složka:

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

- **Reziduum:**

$$\hat{\epsilon}_i = Y_i - \hat{\alpha} - \hat{\beta} X_i.$$

- Regresní přímka vyrovnání:

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta} X_i.$$

- **Vyrovnané hodnoty:**

$$\hat{Y}_i.$$

# Regrese jako přímka nejlepšího vyrovnání

- Součet čtverců reziduí (SSR).

$$\begin{aligned} SSR &= \sum_{i=1}^N \hat{\epsilon}_i^2 \\ &= \sum_{i=1}^N (Y_i - \hat{\alpha} - \hat{\beta}X_i)^2 \\ &= \sum_{i=1}^N (Y_i - \hat{Y}_i)^2. \end{aligned}$$

- Odhady minimalizující tento součet: metoda nejmenších čtverců (ordinary least squares – OLS).

# Interpretace OLS odhadů

- Úrovňová konstanta – ne vždy interpretovatelná × informace o efektivitě trhů v modelu CAPM.
- $\hat{\alpha} = 2.19$  ... odhad fixních nákladů v odvětví.

- Parametr sklonu:

$$\frac{d\hat{Y}_i}{d\hat{X}_i} = \hat{\beta}.$$

- $\hat{\beta} = 4.79$  ... odhad mezních nákladů v odvětví.

# Interpretace OLS odhadů – obecný přehled

**Tabulka:** Interpretace parametrů dle funkční podoby proměnných.

Model	Vysvětlovaná	Vysvětlující	Interpretace $\beta$
Level-Level	$Y$	$X$	$\Delta Y = \beta \Delta X$
Level-Log	$Y$	$\ln X$	$\Delta Y = (\beta/100)\% \Delta X$
Log-Level	$\ln Y$	$X$	$\% \Delta Y = (100\beta) \Delta X$
Log-Log	$\ln Y$	$\ln X$	$\% \Delta Y = \beta \% \Delta X$

# Problémy interpretace odhadů ekonometrických modelů



- Interpretovat lze téměř vše × otázka věrohodnosti závěrů a fundovanosti zdůvodnění.
- Důležité pojmy:
  - *gee* = jé(žíš), jémine (hovorová am. angličtina) × *gee (up)* = hot! (příkaz zvířeti zabočit vpravo).
  - *alien* = mimozemšťan, vetřelec, návštěvník (z vesmíru).
  - *abduction* = únos.

# Kvalita vyrovnání – koeficient determinace

- *Celkový součet čtverců:*

$$TSS = \sum_{i=1}^N (Y_i - \bar{Y})^2.$$

- *Regresní součet čtverců:*

$$RSS = \sum_{i=1}^N (\hat{Y}_i - \bar{Y})^2.$$

- Celková variabilita  $Y$ :

$$TSS = RSS + SSR.$$

- **Koeficient determinace,  $R^2$**  ( $0 \leq R^2 \leq 1$ ):

$$R^2 = \frac{RSS}{TSS} = 1 - \frac{SSR}{TSS}.$$

# Obsah tématu

- 1 Regrese
  - Úvod do regrese
  - Příklady
- 2 **Jednoduchý regresní model**
  - Příklady a základní pojmy
  - Odhady parametrů modelu
  - **Statistické aspekty**
- 3 Model vícenásobné regrese
  - Odhad a interpretace
  - Specifické problémy vícenásobné regrese
- 4 Umělé proměnné

# Intervaly spolehlivosti

- **Intervaly spolehlivosti** odhadu parametru – nejistota spojená s bodovým odhadem.

$$Pr(Int_D < \beta < Int_H) = 0.95$$

- Hladina spolehlivosti – míra důvěry ke zvolenému intervalu (např. 95 %).
- Obvykle 0.99 = 99 %, 0.95 = 95 %, 0.90 = 90 %.



# Testování hypotéz – příklad



- Ekonometrie jako nástroj k hledání odpovědí na zajímavé otázky života skrze formulaci testovatelných hypotéz.
- Důležité pojmy:
  - *stare* = zírat, čučet.
  - *cookie* = sušenka, keks (hl. am. angličtina).
  - *theory* = teorie, hypotéza.
  - *test* = ověřit, testovat (teorie a hypotézy se obvykle ověřují).

# Testování hypotéz o parametru – příklady

- „Zvyšuje úroveň vzdělání jednotlivce jeho potenciál, pokud jde o výši jeho pracovních příjmů?“
- „Zvyšuje určitý typ reklamní strategie nebo kampaně celkové tržby?“
- „Sníží nový vládní systém rekvalifikačních programů nezaměstnanost?“
- Většina otázek typu: „Má vysvětlující proměnná vliv na závisle proměnnou?“, resp. „Je  $\beta \neq 0$  v regresi  $Y$  na  $X$ ?“.

# Testování hypotéz o parametru

- Nulová a alternativní hypotéza:  $H_0 : \beta = 0$  oproti  $H_1 : \beta \neq 0$ .
- Testová statistika pro statistickou významnost parametru:

$$t = \frac{\hat{\beta}}{s_b}.$$

- **Hladina významnosti**: obvykle 0.01, 0.05, 0.10  $\Rightarrow$  (1-hladina spolehlivosti) = pravděpodobnost nutná k nezamítnutí nulové hypotézy (na základě pozorování).
- **Kritická hodnota** testu – z tabulek, z počítačového software; na základě zvolené hladiny významnosti definuje oblast zamítnutí nulové hypotézy.
- **p-hodnota**: porovnáváme se zvolenou hladinou významnosti; pravděpodobnost získání výsledku stejně extrémního jako pozorujeme v datech (čím nižší, tím méně pravděpodobnější nulová hypotéza).
- Testování hypotéz možné s využitím intervalu spolehlivosti.

# Testování hypotéz a typy chyb

- Rozhodování o zamítnutí a nezamítnutí nulové hypotézy → dva typy chyb:
  - zamítnutí  $H_0$ , i když je pravdivá – *chyba prvního druhu (type I error)*;
  - nezamítnutí  $H_0$  při její nepravdivosti – *chyba druhého druhu (type II error)*.
- Testování hypotéz – pouze prostředek k rozhodnutí, jestli se „můžeme tvářit“ jakoby nulová hypotéza platila.
- Pro danou velikost vzorku nelze minimalizovat obě chyby současně ⇒ chyba prvního druhu důležitější (Neyman a Pearson) → chybu prvního druhu volíme na nízké úrovni pravděpodobnosti (0.01, 0.05) a minimalizujeme chybu druhého druhu (síla testu).
- *Hladina významnosti* – pravděpodobnost chyby prvního druhu.
- *Síla testu* – pravděpodobnost nedopuštění se chyby druhého druhu.

## Výstup počítačových programů.

- $\hat{\beta}$ : bodový odhad metodou nejmenších čtverců jako nejlepší odhad skutečné hodnoty  $\beta$ .
- 95% interval spolehlivosti: informace o intervalu, ve kterém s 95% pravděpodobností leží skutečná hodnota parametru  $\beta$ .
- Směrodatná odchylka (standardní chyba) odhadu parametru ( $\hat{\beta}$ ),  $s_b$ : měřítko toho, jak přesný náš odhad je.
- Testová  $t$ -statistika pro testování  $H_0 : \beta = 0$ .
- $p$ -hodnota pro testování  $H_0 : \beta = 0$ .

# Výsledky pro model nákladů energetických podniků.

**Tabulka:** Jednoduchý regresní model pro data o spotřebě elektřiny

Proměnná	Koef.	Sm. odch.	$t$ -stat.	$p$ -hodnota	95% int. spol.
Konstanta	2.19	1.88	1.16	0.25	[-1.53;5.91]
Výstup	4.79	0.13	36.36	0.00	[4.53;5.05]

# Testování hypotézy o $R^2$ .

- $H_0 : R^2 = 0$ ,  $H_1 : R^2 \neq 0 \rightarrow$  má vysvětlující proměnná statisticky významnou sílu k vysvětlení závisle proměnné?

- $F$ -statistika:

$$F = \frac{(N - 2)R^2}{1 - R^2}.$$

- Porovnání s kritickou hodnotou nebo  $p$ -hodnota a naše hladina významnosti.

# Obsah tématu

- 1 Regrese
  - Úvod do regrese
  - Příklady
- 2 Jednoduchý regresní model
  - Příklady a základní pojmy
  - Odhady parametrů modelu
  - Statistické aspekty
- 3 Model vícenásobné regrese
  - Odhad a interpretace
  - Specifické problémy vícenásobné regrese
- 4 Umělé proměnné



# Obsah tématu

- 1 Regrese
  - Úvod do regrese
  - Příklady
- 2 Jednoduchý regresní model
  - Příklady a základní pojmy
  - Odhady parametrů modelu
  - Statistické aspekty
- 3 **Model vícenásobné regrese**
  - **Odhad a interpretace**
  - Specifické problémy vícenásobné regrese
- 4 Umělé proměnné

# Model a OLS odhad

- Model:

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

- Součet čtverců reziduí:

$$SSR = \sum_{i=1}^N \left( Y_i - \hat{\alpha} - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i} - \dots - \hat{\beta}_k X_{ki} \right)^2.$$

- Koeficient determinace – vliv všech proměnných.
- $F$ -test – schopnost *všech* vysvětlujících proměnných vysvětlit chování vysvětlované proměnné:

$$F = \frac{N - k - 1}{k} \frac{R^2}{1 - R^2}.$$

# Interpretace výsledků OLS odhadů

- Parametr – mezní vliv vysvětlující proměnné na vysvětlovanou, pokud ostatní proměnné zůstávají neměnné.

**Tabulka:** Regresní model pro data o cenách domů.

Proměnná	Koef.	Sm. odch.	t-stat.	p-hodnota	95% int. spol.
Konstanta	-4009.55	3603.11	-1.11	0.27	[-11087.3;3068.25]
Rozloha	5.43	0.37	14.70	0.00	[4.70;6.15]
<i>Počet</i>					
ložnic	2824.61	1214.81	2.33	0.02	[438.30;5210.93]
koupelen	17105.17	1734.43	9.86	0.00	[13698.12;20512.22]
pater	7634.90	1007.97	7.57	0.00	[5654.87;9614.92]

# Obsah tématu

- 1 Regrese
  - Úvod do regrese
  - Příklady
- 2 Jednoduchý regresní model
  - Příklady a základní pojmy
  - Odhady parametrů modelu
  - Statistické aspekty
- 3 Model vícenásobné regrese
  - Odhad a interpretace
  - **Specifické problémy vícenásobné regrese**
- 4 Umělé proměnné

# Volba vysvětlujících proměnných

- Dva protichůdné požadavky:
  - 1 Zahrnutí co nejvíce proměnných schopných vysvětlit chování závisle proměnné.
  - 2 Vyhnout se zahrnutí irelevantních proměnných, které snižují statistickou významnost ostatních vysvětlujících proměnných.

# Zkreslení při opomenutí důležité vysvětlující proměnné

- Omitted (missing) variable bias.
- Problém hodný pozornosti z důvodů vysoce pravděpodobného zkreslení odhadů parametrů modelu.



## Příklad zkreslení při opomenutí důležité vysvětlující proměnné

- Původní model (sm. odchylky v závorkách):

$$\hat{Y} = \frac{-4009.55}{(-1.11)} + \frac{5.43}{(14.70)} X_1 + \frac{2824.61}{(2.33)} X_2 + \frac{17105.17}{(9.86)} X_3 + \frac{7634.90}{(7.57)} X_4.$$

- $X_1$  = celková rozloha domu (ve čtverečních stopách),  $X_2$  = počet ložnic,  $X_3$  = počet koupelen,  $X_4$  = počet pater (kromě přízemí).
- Jednoduché regrese ceny domů ( $Y$ ) na počet ložnic ( $X$ ):

$$\hat{Y} = 28773.43 + 13269.98X.$$

## Praktický postup

- V praxi nemožné zahrnout všechny vlivy (vždy existuje zkreslení při nezahrnutí důležité vysvětlující proměnné) – víra v to, že mají minimální vysvětlující sílu a nejsou korelovány se zahrnutými proměnnými.
- Začít s co možná největším počtem vysvětlujících proměnných, odstranit statisticky nevýznamné a zkoušet různé varianty modelu.
- Výsledná regrese – jen statisticky významné proměnné (s výjimkou úrovně konstanty), případně proměnné u kterých zkoumáme statisticky významný vliv.
- Kvalita různě specifikovaných modelů – např. dle koeficientu determinace (nemusí být jediným kritériem!).



# Multikolinearita

- Problém silné nebo úplné vzájemné korelace dvou nebo více **vysvětlujících** proměnných.



# Multikolinearita – důsledky

- Přibližná multikolinearita – vysoké směrodatné odchylky odhadu parametrů (koeficient determinace vysoký  $\times$  všechny parametry statisticky nevýznamné).
- Dokonalá multikolinearita – nemožnost odhadu (intuitivně z interpretace regresního koeficientu, tj. jak se změní vysvětlovaná proměnná, pokud se změní vysvětlující proměnná o jednotku, při neměnnosti ostatních vysvětlujících proměnných  $\times$  nemůže nastat při dokonalé korelaci!).
- Obvyklé řešení – vypuštění proměnné způsobující multikolinearitu.
- Jedna z možností „testování“ – korelační matice.

# Obsah tématu

- 1 Regrese
  - Úvod do regrese
  - Příklady
- 2 Jednoduchý regresní model
  - Příklady a základní pojmy
  - Odhady parametrů modelu
  - Statistické aspekty
- 3 Model vícenásobné regrese
  - Odhad a interpretace
  - Specifické problémy vícenásobné regrese
- 4 **Umělé proměnné**

# Práce s umělými proměnnými

- Způsob převodu kvalitativní vysvětlující proměnné (pohlaví, členství v odborech) do podoby kvantitativní vysvětlující proměnné (1 nebo 0).
- Interpretace jako v případě „normálních“ proměnných.
- Porovnávání se *základní skupinou (kategorií)*  $\Rightarrow$  potřeba uvědomit si, která to je.
- Rozdělení pozorování do jednotlivých kategorií (pozor na multikolinearitu).
- Možnost různých sklonů regresních přímků pro různé kategorie pozorování – nová veličina jako součin umělé a další „normální“ vysvětlující proměnné.
- Umělé závisle proměnné = jiný typ modelů a odhadových technik (logit, probit)!

# Hodinová mzda vzhledem k rodinnému stavu a regionu (odhad)

- Vzorek 528 osob z května 1985, příklad z Gujarati, Porter (2009), odhad modelu ( $R^2 = 0.0322$ ):

	$\hat{Y}_i$	=	8.8148	+	1.0997 $D_{2i}$	-	1.6729 $D_{3i}$
sm. odch.			(0.4015)		(0.4642)		(0.4854)
t-stat			(21.9528)		(2.3688)		(-3.4462)
p-value			(0.0000)		(0.0182)		(0.0006)

- $Y$  = hodinová mzda (v dolarech).
- $D_2$  = rodinný status, (1 = ženatý/vdaná, 0=jiný).
- $D_3$  = místo rezidence, (1 = Jih, 0=jinak).

# Hodinová mzda vzhledem k rodinnému stavu a regionu (interpretace)

- Dvě kvalitativní vysvětlující proměnné (každá dvě kategorie).
- Základní kategorie (srovnávací): neženatá (nevdaná) osoba, nesídlící na Jihu.
- Veškerá srovnání vzhledem k této základní kategorii (odhady statisticky významné  $\times$  nízká hodnota  $R^2$ ):
  - Průměrná (očekávaná) hodinová mzda základní kategorie = 8.81.
  - Mzda ženatých: vyšší o 1.10 dolarů.
  - Mzda Jižanů: nižší o 1.67 dolarů.
- V případě více kvalitativních proměnných nutné chápat výsledky vzhledem k základní kategorii!

## Příklad – prodejní ceny domů

- Regrese cen domů,  $Y$ , na umělou proměnnou označující vybavení klimatizací,  $D$ , proměnnou udávající rozlohu domu,  $X \Rightarrow$  model

$$Y_i = \alpha + \beta_1 D_i + \beta_2 X_i + \epsilon_i.$$

- Odhady:  $\hat{\alpha} = 32693$ ,  $\hat{\beta}_1 = 20175$  a  $\hat{\beta}_2 = 5.638$ .
- Dvě různé regresní přímky (se stejným sklonem):

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta}_1 + \hat{\beta}_2 X_i = 52868 + 5.638 X_i,$$

pro  $D_i = 1$  (dům s klimatizací) a

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta}_2 X_i = 32693 + 5.638 X_i,$$

pro  $D_i = 0$  (dům bez klimatizace).

# Odhad různých mezních efektů pro různé kategorie domů.

- Regrese cen domů,  $Y$ , na umělou proměnnou označující vybavení klimatizací,  $D$ , proměnnou udávající rozlohu domu,  $X$ , a novou proměnnou  $Z = DX \Rightarrow$  model

$$Y_i = \alpha + \beta_1 D_i + \beta_2 X_i + \beta_3 Z_i + \epsilon_i.$$

- Odhady:  $\hat{\alpha} = 35684$ ,  $\hat{\beta}_1 = 7613$  a  $\hat{\beta}_2 = 5.02$  a  $\hat{\beta}_3 = 2.25$ .
- Marginální vliv celkové rozlohy domu: 7.27 dolarů pro domy s klimatizací a 5.02 dolarů pro domy bez klimatizace.
- $p$ -hodnota pro  $\beta_3 = 0.02$  (marginální vliv je statisticky významný).
- Růst rozlohy domu má tendenci více zhodnocovat dům vybavený klimatizací než dům bez klimatizace.
- $\beta_1$  statisticky nevýznamný ( $p$ -hodnota 0.17)  $\Rightarrow$  není dodatečná „fixní prémie“ za klimatizaci.



## Mezní vliv při logaritmu vysvětlované proměnné.

- $\ln Y$  (logaritmus vysvětlované proměnné).
- Interpretace koeficientu umělé vysvětlující proměnné (násobeného 100) – procentní změna  $Y$  při neměnných ostatních proměnných.
- Pokud koeficient naznačuje velké procentní změny je dobré získat přesné vyjádření.
- $\hat{\beta}$  (koeficient umělé proměnné  $X$ ), potom přesná predikovaná hodnota  $Y$  pro  $X = 1$  oproti základní skupině  $X = 0$  je

$$100 \cdot [\exp(\hat{\beta}) - 1].$$

- Pozor na znaménka parametrů (chápání základní skupiny).

## Vliv počítačů na mzdy (model)

- Efekt využití počítačů na mzdy.
- Krueger, A. B. (1993): „How Computers Have Changed the Wage Structure: Evidence from Microdata, 1984–1989,“ *Quarterly Journal of Economics* 108, 33–60.
- $compwork = 1$  pokud využívá osoba počítač v práci (0 jinak) a  $comphome = 1$  pokud osoba používá počítač doma (0 jinak).
- Standardní zahrnutí dalších faktorů (*other factors*): vzdělání, zkušenosti, pohlaví apod.
- Krueger neukazuje odhad úrovně konstanty (nemá vliv na interpretaci výsledků)  $\Rightarrow$  základní skupina lidí nepoužívající počítač doma ani v práci.

## Vliv počítačů na mzdy (odhad a interpretace)

- Odhad na vzorku 13 379 osob z roku 1989 (Current Population Survey), v závorce sm. odchylky odhadu parametrů:

$$\ln(\widehat{wage}) = \hat{\alpha} + \frac{.177}{(0.009)} compwork + \frac{.070}{(.019)} comphome \\ + \frac{.017}{(.023)} compwork \cdot comphome + other\ factors.$$

- Zisk z využití počítače v práci (ne doma) = 17.7 % (přesnější odhad 19.4 %).
- Lidé využívající počítač jen doma = mzdová prémie 7 % oproti základní skupině.
- Rozdíl mezd pracovníků nevyužívajících počítač 26.4 % (přesněji 30.2 %) × člen interakce statisticky (i ekonomicky) nevýznamný ⇒ není zde synergický efekt.