

- Základní pojmy
- Tvorba regresní funkce
- Přímková (lineární) regrese
- Odvození regresní funkce pomocí metody nejmenších čtverců

1.

## Úvod regresního a korelačního počtu



## Cíl kapitoly

Hledání závislostí mezi dvěma jevy patří k základním nástrojům ekonomické analýzy. V ekonomických úlohách je velmi obvyklé hledání signifikantních souvislostí mezi dvěma, třemi, či více ukazateli. Následující kapitola Vás uvede do jedné z nejdůležitějších metod hledání a kvantifikace těchto závislostí – regresní a korelační analýzy. Na jejím základě budete schopni matematicky popsat vzájemný vztah mezi dvěma jevy a provést či navrhnout předpověď vývoje dané závislosti do budoucna.



## Časová zátěž

6 hodin (3. týden v únoru)

### Úvod

Tato kapitola se zaměřuje na zkoumání a hodnocení souvislostí mezi dvěma a více statistickými znaky. Jedná se o nejčastější oblast statistického zkoumání v ekonomii, stejně jako je možno ji považovat za oblast nejsložitější. Cílem zkoumání vzájemných vztahů a souvislostí je zpravidla snaha o proniknutí do podstaty jevů a procesů a nalezení příčinných (kauzálních) souvislostí.

Regresní a korelační analýza se tedy zaměřuje na situace, kdy existence (výskyt) určitého jevu souvisí (má za následek) s existencí jiného jevu. V praktických úlohách jsou obvyklejší případy, kdy existenci nějakého jevu ovlivňuje skupina jevů.



*Např. míra ekonomického růstu měřená růstem hrubého domácího produktu je zpravidla ovlivněna více než jedním faktorem. Na ekonomický růst má vliv např. míra růstu cenové hladiny, míra nezaměstnanosti, čisté vývozy, apod.*

## 1.1 Základní pojmy

Rozlišujeme závislosti

- pevné – výskytu jednoho jevu nutně odpovídá výskyt druhého jevu. Jedné hodnotě první veličiny odpovídá právě jedna hodnota druhé veličiny.
- volné – výskyt jednoho jevu ovlivňuje výskyt jiného jevu. Jedné hodnotě první veličiny odpovídá více hodnot druhé veličiny, nicméně lze pozorovat nějakou obecnou tendenci těchto hodnot.

Všimněte si, že z hlediska teorie pravděpodobnosti se v případě pevných závislostí jedná o jev jistý. Příkladem pevné závislosti jsou proto například fyzikální zákony. V případě volných závislostí lze z hlediska teorie pravděpodobnosti pouze konstatovat, že se pravděpodobnost nastoupení druhého jevu zvýšila v souvislosti s nastoupením prvního jevu.

V praktických ekonomických úlohách se setkáváme výhradně s volnými závislostmi. Jejich matematickým popisem se zabývají metody regresní a korelační analýzy.

**Regresní analýza** popisuje jednostranné závislosti. Zajímá se tedy o typ závislostí, u kterých je možno rozlišit příčinu (je označována jako nezávisle proměnná) a následek (závisle proměnná). **Korelační analýza** se zabývá vzájemnými závislostmi. V úlohách využívajících korelační analýzy je více kladen důraz na zkoumání intenzity popisované vzájemné závislosti než na zkoumání veličin ve smyslu příčina – následek.

Z hlediska popisných statistických veličin je možno ilustrovat princip regresní a korelační analýzy pomocí tzv. sdružených četností. Jejich podstatu si můžete nastudovat v učebnici SEGER, HINDLS: *Statistické metody v tržním hospodářství* na stranách 229–235.)

### 1.1.1 Cíle regresní a korelační analýzy

Hlavním úkolem je poznání příčinných vztahů mezi statistickými znaky. Výhodiskem tohoto zkoumání jsou statistické údaje, které popisují uvedené jevy.

Tyto statistické údaje možno získat několika způsoby.

- a) Pozorováním  $n$  statistických jednotek s časově prostorově a věcně vymezeným statistickým souborem.
- b) Pozorováním určité statistické jednotky v  $n$  různých časových okamžicích či intervalech.
- c)  $N$ -násobným opakováním určitého pokusu, jež je prováděn za stejných, či přibližně stejných podmínek.

*Příkladem prvního jevu je jakýkoli výsledek statistického průzkumu – například údaje o spotřebním chování domácností v měsíci březnu roku 2002. Statistický soubor druhého typu lze získat například sledováním počtu nezaměstnaných osob registrovaných na úřadech práce v jednotlivých měsících roku 2002. Třetí situaci lze ilustrovat například sledováním výskytu vadných kusů u stejného typu výrobku během jedné pracovní směny.*

Regresní analýza se snaží o nalezení idealizující matematické funkce tak, aby pokud možno co nejlépe vystihla charakter dané závislosti. Tato funkce je potom nazývána **regresní funkcí**. Snažíme se tedy, aby hodnoty vypočítané na základě této regresní funkce co nejvíce odpovídaly hodnotám empirickým (naměřeným, či statisticky zjištěným).

Postupné kroky regresní analýzy

1. Formulovat matematicky představy o charakteru regresní funkce.
2. Posoudit vliv faktorů, které se rozhodneme do dané funkce nezahrnout.
3. Odhadnout konkrétní podobu regresní funkce na základě zjištěných hodnot.
4. Posoudit kvalitu této funkce

Na základě znalosti konkrétní podoby regresní funkce lze provádět odhady vývoje zkoumané veličiny.

Regrese a korelace



Postup regrese

Interpolační a  
extrapolační  
odhad

Tyto odhady mohou být dvou typů:

- a) Interpolační – odhady prováděné uvnitř intervalu hodnot vysvětlující proměnné.
- b) Extrapolační – odhady prováděné mimo oblast měření.

V praktických úlohách je častější provádění extrapolačních odhadů. V ekonomických úlohách se hovoří o predikcích či prognózách vývoje zkoumaného jevu. Extrapolačních odhadů je tedy využíváno ve všech úlohách, kdy nás zajímá vývoj pozorované veličiny v budoucím období stanovíme-li předpoklad vývoje vysvětlujících proměnných. Interpolační odhady se zaměřují na zpětný odhad již známých veličin. Slouží zejména pro posouzení kvality použité regresní funkce, neboť umožňují srovnání známých (empirických) hodnot z jejich vypočítanými protějšky (teoretickými hodnotami).

Jeden z problémů spojený s vytvářením extrapolačních odhadů je riziko, že zkoumaný jev bude mimo interval hodnot vysvětlující proměnné mít jiný průběh než uvnitř tohoto intervalu. Může například dojít k nepředpokládaným změnám chování vysvětlujících veličin, které souvisejí s nečekanými událostmi (živelní pohromy, válka, ropná krize, apod.). Z těchto důvodů jsou extrapolační odhady obvykle interpretovány jako možný vývoj zkoumané skutečnosti.

## 1.2 Tvorba regresní funkce

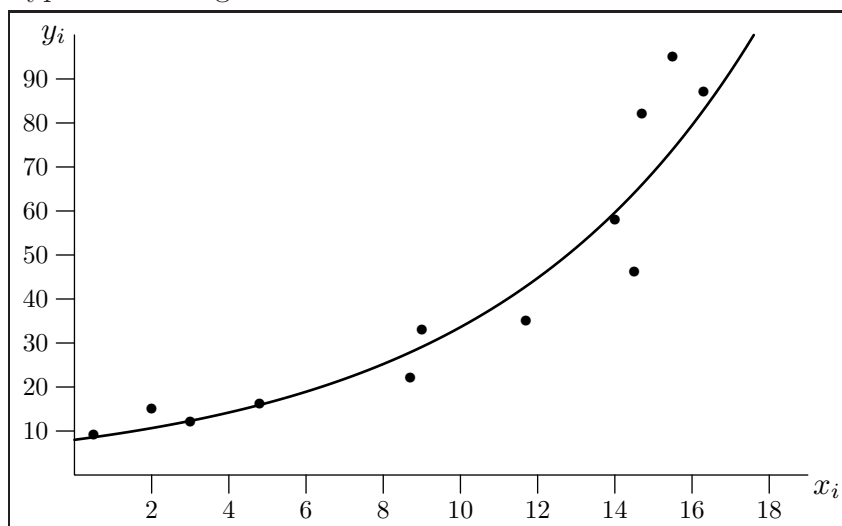
Pro konkrétní volbu tvaru regresní funkce je nutno zohlednit několik faktorů. Zabýváme-li se popisem vztahů mezi ekonomickými veličinami, je nutno vyjít zejména z věcné ekonomické analýzy. Zkoumaná závislost je ve většině případů již teoreticky rozpracována a popsána. Hlavním úkolem analýzy je posoudit možnost aplikace těchto poznatků na konkrétní případ. Teoretické modely obvykle pracují s jistou mírou zjednodušení skutečnosti, je proto nutno posoudit, zda naměřené (zjištěné) hodnoty veličin odpovídají těmto modelům.

V případě, že nelze využít teoreticky odvozených modelů, je nutno přistoupit k odhadu tvaru regresní funkce více intuitivně. Musíme vyjít z rozboru empirického průběhu sledované závislosti. Pro účely tohoto intuitivního přístupu se nejlépe osvědčuje grafické znázornění známých hodnot závisle a nezávisle proměnné.

Oba přístupy s sebou přinášejí významná rizika. Použití teoretických modelů je obvykle omezeno na úzký okruh případů, které vyhovují jeho předpokladům. Naopak využití grafické metody je možno využít ve všech případech. Je však nutno mít na paměti, že grafické znázornění je také určitým zjednodušením skutečnosti. Graficky zachycený průběh závislosti neříká nic o vazbách či vlivech jiných faktorů. Grafické metody je také s úspěchem možno využít zejména pro posouzení dvourozměrné závislosti (jedné závisle a jedné nezávisle proměnné). V případě závislostí vícerozměrných jsou již možnosti grafické analýzy omezené.

Regresní funkci obvykle konstruujeme ve tvaru elementární matematické funkce. Podle typu zvolené funkce poté rozlišujeme následující regrese:

- a) Přímková (lineární) regrese
- b) Parabolická (kvadratická) regrese
- c) Polynomická regrese  $n$ -tého stupně
- d) Logaritmická regrese
- e) Exponenciální regrese
- f) Hyperbolická regrese



Obrázek 1.1: Vyrovnání pomocí exponenciální funkce

### 1.3 Přímková (lineární) regrese

Nejjednodušším a nejužívanějším typem regresní funkce je vyrovnání empirických hodnot přímkou. Lineární regrese vychází z předpokladu, že vývoj závisle proměnné je přímo úměrný vývoji nezávisle proměnné.

Empirickými hodnotami závisle proměnné (obvykle označované jako  $y$ ) se snažíme proložit přímkou tak, aby co nejlépe vystihovala tendenci vývoje těchto hodnot v závislosti na vývoji nezávisle proměnné (obvykle označované jako  $x$ ).

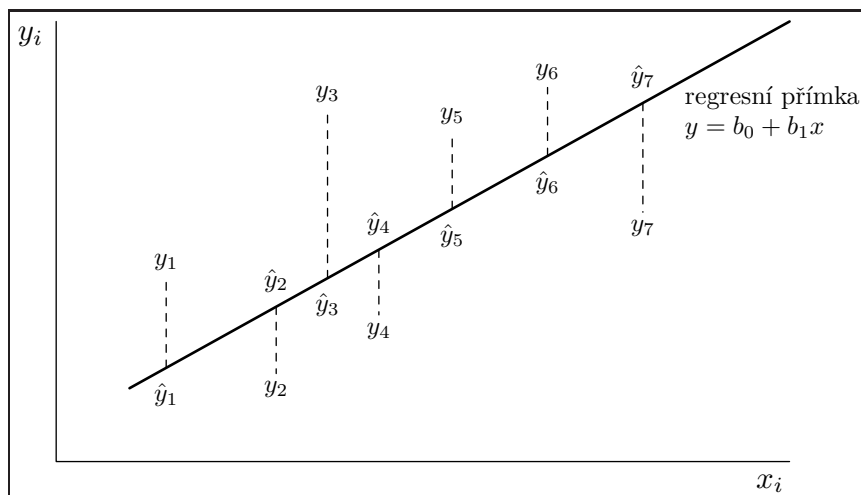
Jelikož je podobných přímek možno pro daný soubor dat možno obvykle nalézt několik (teoreticky nekonečně mnoho) je nutno stanovit kritéria pro volbu té „nejlepší“. Základní používanou podmínkou pro volbu regresní přímky, je požadavek nejnižší celkové odchylky vypočítaných (odhadnutých) a empirických (naměřených) hodnot (viz obrázek 1.1).

Nejznámější metoda, která z tohoto požadavku vychází, je označována jako **metoda nejmenších čtverců**. Metoda vychází z podmínky minimalizace celkové odchylky vyrovnaných a empirických hodnot, kdy je tato celková odchylka počítána jako součet druhých mocnin odchylek jednotlivých datových bodů (hodnot).

# 1. Úvod regresního a korelačního počtu



K odvození regresní funkce jsou užívány i jiné metody než metoda nejmenších čtverců. Tyto metody jsou užívány ve složitějších úlohách, pro potřeby jednodušších analýz je metoda nejmenších čtverců dostatečná. Některé z dalších metod odhadu regresních funkcí naleznete například v učebnici Wonnacot, Wonnacot Statistika pro obchod a hospodářství 397–399.



Obrázek 1.2: Odvození regresní přímky pomocí metody nejmenších čtverců

## 1.4 Odvození regresní funkce pomocí metody nejmenších čtverců

Naším hlavním úkolem je určit konkrétní podobu teoretické regresní funkce a odhadnout její parametry. Označíme-li odhady uvedených parametrů  $b_0, b_1, \dots, b_p$ , pak empirickou regresní funkci můžeme psát (a hledáme) ve formě

$$\hat{y} = f(x_i; b_0, b_1, \dots, b_p).$$

Parametry regresních funkcí se nejčastěji určují metodou nejmenších čtverců.

Mějme  $n$  dvojic pozorovaných hodnot  $(x_i, y_i)$ , kde  $i = 1, 2, 3, \dots, n$ , které vytváří bodový diagram. Množinu těchto bodů nechť popisuje empirická regresní funkce:

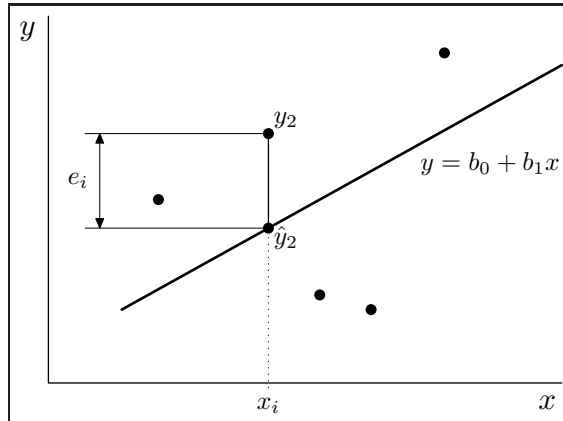
$$\hat{y} = f(x, b_0, b_1).$$

Odchylka  $e_i = y_i - \hat{y}_i$ , kde  $e_i$  je reziduum.

Statistika  $S_r = \sum_{i=1}^n e_i$  udává rozptýlení pozorovaných hodnot závislé proměnné  $y$  kolem empirické regresní funkce. Metoda nejmenších čtverců je založena na *minimalizaci* součtu čtverců reziduálních odchylek. Hledáme tedy minimum funkce:

$$S_r = \sum_{i=1}^n e_i = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - b_0 - b_1 \cdot x_i)^2 = S_r(b_0, b_1) \rightarrow \text{minimum}$$

neboť  $y_i = b_0 + b_1 \cdot x_i$ .



Obrázek 1.3: Defnice odchylky  $e_i$

Pomocí parciálních derivací  $\frac{\partial S_r}{\partial b_0}$ ,  $\frac{\partial S_r}{\partial b_1}$  obdržíme soustavu dvou rovnic o dvou neznámých.

Aby nastal extrém dané funkce (minimum), ještě ve vztahu k druhé derivaci, musí se parciální derivace rovnat *nule*.

$$\frac{\partial S_r}{\partial b_0} = 2 \cdot \sum_{i=1}^n (y_i - b_0 - b_1 \cdot x_i) \cdot (-1) = 0, \quad \frac{\partial S_r}{\partial b_1} = 2 \cdot \sum_{i=1}^n (y_i - b_0 - b_1 \cdot x_i) \cdot (-x_i) = 0.$$

Soustavu upravíme na tvar:

$$\begin{aligned} 2 \cdot \sum_{i=1}^n (y_i - b_0 - b_1 \cdot x_i) \cdot (-1) = 0 &\Rightarrow -\sum_{i=1}^n y_i + n \cdot b_0 + b_1 \sum_{i=1}^n x_i = 0, \\ 2 \cdot \sum_{i=1}^n (y_i - b_0 - b_1 \cdot x_i) \cdot (-x_i) = 0 &\Rightarrow \\ &\Rightarrow -\sum_{i=1}^n (x_i \cdot y_i) + b_0 \sum_{i=1}^n x_i + b_1 \cdot \sum_{i=1}^n x_i^2 = 0. \end{aligned}$$

Znamé hodnoty převedeme na pravou stranu a obdržíme:

$$\begin{aligned} n \cdot b_0 + b_1 \sum_{i=1}^n x_i &= \sum_{i=1}^n y_i, \\ b_0 \sum_{i=1}^n x_i + b_1 \cdot \sum_{i=1}^n x_i^2 &= \sum_{i=1}^n (x_i \cdot y_i). \end{aligned}$$

Řešením této soustavy dvou rovnic, ať již pomocí determinantů, nebo vyloučením jedné neznámé obdržíme:

$$b_1 = \frac{n \cdot \sum_{i=1}^n (x_i \cdot y_i) - \sum_{i=1}^n x_i \cdot \sum_{i=1}^n y_i}{n \cdot \sum_{i=1}^n x_i^2 - \left( \sum_{i=1}^n x_i \right)^2}$$

# 1. Úvod regresního a korelačního počtu

Vydělíme čítec a jmenovatel výrazem  $n^2$  a potom obdržíme:

$$b_1 = \frac{\frac{1}{n} \cdot \sum_{i=1}^n (x_i \cdot y_i) - \frac{1}{n} \sum_{i=1}^n x_i \cdot \frac{1}{n} \sum_{i=1}^n y_i}{\frac{1}{n} \cdot \sum_{i=1}^n x_i^2 - \frac{1}{n^2} \left( \sum_{i=1}^n x_i \right)^2} = \frac{\overline{x \cdot y} - \bar{x} \cdot \bar{y}}{x^2 - (\bar{x})^2} =$$
$$= \frac{\text{cov}(x, y)}{\text{rozptyl}(x)} = \frac{s_{x,y}}{s_x^2} = r_{x,y} \cdot s_x \cdot s_y,$$

neboť:  $\frac{1}{n} \sum_{i=1}^n x_i \cdot \frac{1}{n} \sum_{i=1}^n y_i$  je součin středních hodnot náhodných veličin  $X$  a  $Y$

$\frac{1}{n} \cdot \sum_{i=1}^n (x_i \cdot y_i)$  je střední hodnota součinu hodnot  $x_i, y_i$  náhodných veličin  $X$  a  $Y$

$\frac{1}{n} \cdot \sum_{i=1}^n x_i^2$  je střední hodnota druhých mocnin hodnot náhodné veličiny  $X$

$\frac{1}{n^2} \left( \sum_{i=1}^n x_i \right)^2$  je druhá mocnina střední hodnoty hodnot náhodné veličiny  $X$

Hodnotu  $b_0$  vypočítáme z první rovnice:

$$b_0 = \frac{\sum_{i=1}^n y_i - b_1 \sum_{i=1}^n x_i}{n} = \frac{1}{n} \cdot \sum_{i=1}^n y_i - \frac{b_1}{n} \cdot \sum_{i=1}^n x_i = \bar{y} - b_1 \cdot \bar{x}.$$

Dosazením těchto hodnot do regresní přímky obdržíme:

$$\hat{y} = b_0 + b_1 \cdot x = \bar{y} - b_1 \cdot \bar{x} + b_1 x.$$

Potom  $\hat{y} - \bar{y} = b_1 \cdot (x - \bar{x})$ , kde  $\bar{y}$  a  $\bar{x}$  jsou známé hodnoty. Konečně vyřešíme celou úlohu, tím, že dosadíme i hodnotu  $b_1$  a to:  $b_1 = \frac{s_{x,y}}{s_x^2}$ . Potom získáme rovnici regresní přímky z hodnot náhodné proměnné jednoduchým způsobem:

$$\hat{y} - \bar{y} = \frac{s_{x,y}}{s_x^2} \cdot (x - \bar{x}).$$

Snadno zjistíme střední hodnotu  $\bar{x}$  a  $\bar{y}$ , kovarianci  $s_{x,y}$  a rozptyl  $s_x^2$ , nebo směrodatné odchylky  $s_x, s_y$  a korelaci mezi hodnotami náhodné veličiny  $X$  a  $Y$ .



Kovarianci řadíme mezi popisné statistické veličiny. Její definici i vztah pro výpočet naleznete například v učebnice SEGER, HINDLS: *Statistika v hospodářství*. Výsledkem odvození jsou následující dva koeficienty:

**$b_0$  – absolutní člen (konstanta)** – označuje míru úrovně regresní přímky. Nemá vliv na její sklon, a tedy konkrétní závislost. Je to jenom jakési číslo, které musíme přičíst.

$$b_0 = \frac{\sum_{i=1}^n y_i \sum_{i=1}^n x_i^2 - \sum_{i=1}^n x_i \sum_{i=1}^n y_i x_i}{n \sum_{i=1}^n x_i^2 - \left( \sum_{i=1}^n x_i \right)^2}$$



V regresním modelu chápaném jako měřítko závislosti (jak se  $y$  mění v závislosti na  $x$ ) není jeho výše příliš podstatná.

$b_1$  – **regresní koeficient** – označuje sklon regresní přímky. Neboli o kolik se změní závislá proměnná, změníme-li nezávisle proměnnou (kolikrát  $x$  se rovná  $y$ ). Hodnota regresního parametru je v případě přímkové regrese nejvýznamnější informací.

$b_1 = 0 \implies$  nezávislost

$b_1 = 1 \implies$  100% závislost (kolikrát vzroste  $x$  tolikrát i  $y$ )

$$b_1 = \frac{n \sum_{i=1}^n y_i x_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - \left( \sum_{i=1}^n x_i \right)^2} = \frac{s_{xy}}{s_x^2}$$

### Příklad 1.1

Prozkoumejte vztah mezi výdaji veřejných rozpočtů (VV) a HDP v ČR v letech 1993–2000 (viz tabulka 1.1). Na základě odhadnuté závislosti proveďte odhad výdajů v roce 2001.



| Rok  | HDP     | VV    |
|------|---------|-------|
| 1993 | 1 020,3 | 420,1 |
| 1994 | 1 182,8 | 462,9 |
| 1995 | 1 381,0 | 531,8 |
| 1996 | 1 567,0 | 596,1 |
| 1997 | 1 679,9 | 640,4 |
| 1998 | 1 829,4 | 690,7 |
| 1999 | 1 887,3 | 733,3 |
| 2000 | 1 959,5 | 780,9 |
| 2001 | 2 146,1 |       |

Tabulka 1.1: Vývoj HDP a VV v letech 1993–2001 (mld. Kč)

Snažíme se matematicky popsat závislost výše VV na HDP (lze odhadnout výši VV známe-li objem HDP?)

$x$ ... nezávisle proměnná ... HDP

$y$ ... závisle proměnná ... veřejné výdaje (VV)

Hledáme funkci

$$y = b_0 + b_1 x$$

neboli

$$VV = b_0 + b_1 \text{ HDP.}$$

# 1. Úvod regresního a korelačního počtu

Dosažením do vztahů pro  $b_0$  a  $b_1$  dostáváme následující výraz:

$$b_0 = \frac{\sum_{i=1}^n y_i \sum_{i=1}^n x_i^2 - \sum_{i=1}^n x_i \sum_{i=1}^n y_i x_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2} =$$
$$= \frac{4856 \cdot 20372987 - 12507 \cdot 7898123}{8 \cdot 20372987 - 156430052} = 23,08$$
$$b_1 = \frac{n \sum_{i=1}^n y_i x_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2} = \frac{8 \cdot 7898123 - 12507 \cdot 4856}{8 \cdot 20372987 - 156430052} = 0,37$$
$$b_0 = 23,08$$
$$b_1 = 0,37$$

Regresní přímka má tedy tvar:

$$y = 23,08 + 0,37x$$

neboli

$$VV = 23,11 + 0,37 \cdot \text{HDP}.$$

Celkově lze konstatovat, že výše veřejných výdajů (VV) velmi těsně závisí na objemu HDP. Tuto závislost je možno charakterizovat číslem 0,37. Tedy veřejné výdaje v ČR tvoří přibližně 0,37 (37%) hrubého domácího produktu.

Uvedenou regresní funkci lze použít k vytvoření odhadů. Interpolační odhad je dodatečným odhadem hodnot veřejných výdajů počítaných na základě regresní funkce pro HDP za roky 1993–2000. Interpolační odhady slouží zejména pro porovnání vypočítaných (teoretických) hodnot se skutečností. Jsou tedy do jisté míry měřítkem kvality vypočítané regresní funkce. Hodnoty interpolačních odhadů uvádí následující tabulka:

| Rok  | HDP     | Veřejné výdaje | Interpolační odhad veřejných výdajů |
|------|---------|----------------|-------------------------------------|
| 1993 | 1 020,3 | 420,1          | 404,2                               |
| 1994 | 1 182,8 | 462,9          | 464,9                               |
| 1995 | 1 381,0 | 531,8          | 538,9                               |
| 1996 | 1 567,0 | 596,1          | 608,4                               |
| 1997 | 1 679,9 | 640,4          | 650,5                               |
| 1998 | 1 829,4 | 690,7          | 706,4                               |
| 1999 | 1 887,3 | 733,3          | 728,0                               |
| 2000 | 1 959,5 | 780,9          | 755,0                               |

Extrapolací odhad vypočítáme jako hodnotu veřejných výdajů pro rok 2001 dosazením hodnoty HDP za rok 2001 do regresní funkce.

$$VV_{2001} = 23,08 + 0,37 \cdot 2156 = 824,7$$

## Shrnutí kapitoly

Regresní a korelační počet slouží k matematickému popisu ekonomických závislostí. Ze statistického hlediska jsou pro tyto účely vhodné zejména závislosti silné. Skutečné ekonomické závislosti jsou však velmi zřídka silné. Proto by každé regresní analýze měla předcházet hlubší analýza věcných příčin sledovaného jevu. V opačném případě vzniká reálné nebezpečí, že nadějná regresní úloha zůstane jen v rovině velmi dobrého školního příkladu.

Nejjednodušší funkcí používanou pro popis závislostí je přímka. V tomto případě se dvojicemi bodů označujícími kombinace sledovaných dvou proměnných (obvykle nazývaných závisle a nezávisle proměnná) snažíme proložit přímkou, která co nejvíce vystihovala jejich vzájemný vztah. K odhadu regresní funkce je využíváno celé řady metod. V kapitole je nastíněno použití nejvýznamnější z nich - metody nejmenších čtverců.



## Otázky k zamyšlení

- 1 Vysvětlete rozdíl mezi regresí a korelací.
- 2 Proč není vhodné užití regresního počtu u závislostí, které se nevyznačují příliš velkou těsností?
- 3 Jaký je princip metody nejmenších čtverců? Je jí možno použít i pro jiné než přímkové regresní funkce?



## Příloha kapitoly 1

### Vytvoření regresního modelu v EXCELU

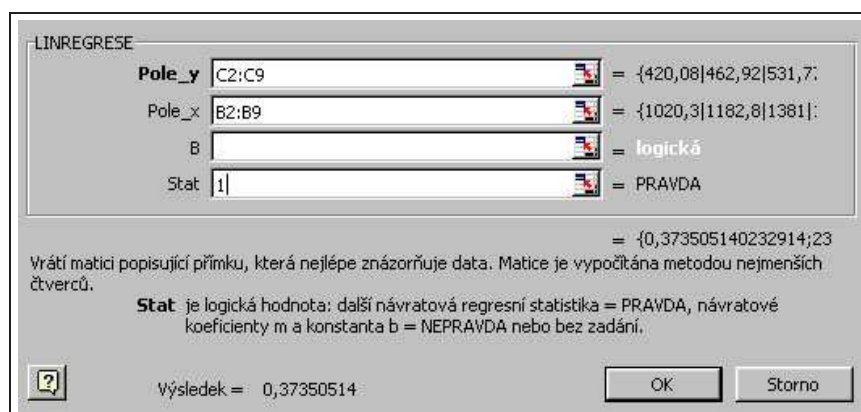
K výpočtu regresní funkce je výhodné využít některého vhodného software. V současné době jsou základní nástroje pro výpočet jednoduchých regresních funkcí běžnou součástí tabulkových procesů. Následující příloha demonstruje možnosti, které při modelování regresní funkce nabízí jeden z nejrozšířenějších programů – program EXCEL.

Postupy uvedené v této i ostatních kapitolách slouží především jako základní shrnutí postupů, jež vycházejí z praktických ukázek prezentovaných během jednotlivých konzultací. Ve spojení se standardními manuály k programu EXCEL však mohou sloužit i jako námět k samostudiu.

### Vytvoření regresní funkce k příkladu uvedenému v kapitole 1.

Pro vytvoření regresní funkce je v EXCELU možno využít funkce LINREGRESE. Postup modelování přímky je následující:

1. Umístíme hodnoty HDP za roky 1993–2000 do buněk B2 až B9 a údaje o výdajích veřejných rozpočtů (VV) do buněk C2 až C9.
2. Do oblasti E3:F7 označíme  $2 \times 5$  buněk jako výstupní oblast pro funkci LINREGRESE
3. Z nabídky zvolíme Vložit/Funkce/Statistické/LINREGRESE
  - **Pole y** ... buňky s hodnotami VV 1993–2000 (C2:C9)
  - **Pole x** ... buňky s hodnotami HDP 1993–2000 (B2:B9)
  - **B**..... nechat prázdné (1 v příp., že nechceme do modelu zahrnout konstantu)
  - **Stat** ..... 1 (1 vyplníme v případě, že chceme vypočítat kromě koeficientů i statistiky kvality)



Obrázek 1.4: Funkce LINREGRESE

4. Po potvrzení tlačítkem OK je nutno klepnout myší do prostoru, kde se zadává funkce a zmáčknout kombinaci kláves CTRL+SHIFT+ENTER. Při správném dodržení postupu se u definovaného vztahu objeví se složené závorky { } a do označených buněk se vyplní hodnoty koeficientů i doplňujících statistik.

5. Výstup fce LINREGRESE V označených deseti buňkách se objeví statistiky, které charakterizují regresní model:

|             |           |
|-------------|-----------|
| 0,373488    | 23,11449  |
| 0,373505    | 23,083314 |
| 0,017526    | 27,968560 |
| 0,986961    | 15,863208 |
| 454,169213  | 6         |
| 114 287,764 | 1 509,848 |

Jednotlivé hodnoty označují následující proměnné:

|                                       |                                       |
|---------------------------------------|---------------------------------------|
| $b_1$                                 | $b_0$                                 |
| směrodatná odchylka koeficientu $b_1$ | směrodatná odchylka koeficientu $b_0$ |
| Index determinace                     | reziduální směrodatná odchylka        |
| hodnota statistiky $F$                | počet stupňů volnosti                 |
| regresní suma čtverců                 | reziduální suma čtverců               |

Hledané koeficienty regresní funkce tedy najdeme v prvním řádku, přičemž jsou uvedeny v pořadí  $b_1$ ,  $b_0$ .

*Význam zbylých charakteristik je rozebrán v následující kapitole.*

Regresní přímka má tedy tvar:

$$y = 23,08 + 0,37x$$

neboli

$$VV = 23,08 + 0,37 \text{ HDP}.$$

Chceme-li vypočítat extrapoláčn odhad, je nutno do další buňky odpovídající HDP v roce 2001 doplnit údaj 2156 mld. Kč. Nadefinujeme-li do sloupce C vypočítanou regresní funkci, obdržíme hodnotu extrapoláčního odhadu pro rok 2001, tj. 824,7 mld. Kč. Dosadíme-li do takto nadefinované regresní funkce i hodnoty HDP pro předchozí roky, dostáváme interpolační odhady výše veřejných výdajů.



