

Základy ekonometrie

I. Úvod do ekonometrie a práce s daty

Obsah tématu

- 1 Úvod do ekonometrie
 - Ekonomický model
 - Ekonometrický model
- 2 Práce s daty
 - Typy ekonomických dat
 - Grafické metody práce s daty
 - Popisné statistiky a korelace
- 3 Ekonometrický software a studijní literatura

Obsah tématu

1 Úvod do ekonometrie

- Ekonomický model
- Ekonometrický model

2 Práce s daty

- Typy ekonomických dat
- Grafické metody práce s daty
- Popisné statistiky a korelace

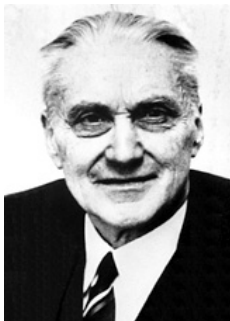
3 Ekonometrický software a studijní literatura

Co je ekonometrie?

- Není jednotná definice.
- „Ekonometrie je vše, co dělají ekonometři.“
- „Ekonometrie je věda zabývající se aplikací statistických metod pro analýzu ekonomických fenoménů.“
- Propojení a rozšíření poznatků ekonomické teorie, matematické ekonomie, ekonomické statistiky a matematické statistiky.
- Počátek 30. let 20. století: založení Ekonometrické společnosti (Econometric Society) a časopisu *Econometrica*.

„Nobelova cena“ za ekonomii

- 1969 – první „Nobelova cena“ za ekonomii (**The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel**) pro Ragnara Frische a Jana Tinbergena za „**rozvoj a aplikaci dynamických modelů pro analýzu ekonomických procesů.**“



Ragnar Frisch (1895–1973)

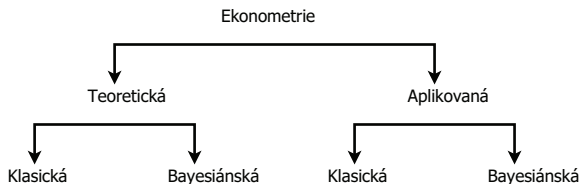


Jan Tinbergen (1903–1994)

Kdo je ekonometr?

- **Ekonom** – využívá ekonomickou teorii pro podporu svých empirických analýz.
- **Matematik** – formuluje ekonomickou teorii způsobem vhodným pro statistické testování.
- **Účetní** – hledá a sbírá ekonomická data a propojuje teoretické proměnné s pozorovatelnými.
- **Aplikovaný statistik** – odhaduje ekonomické vztahy a predikuje možné výstupy.
- **Teoretický statistik** – rozvíjí statistické techniky s ohledem na specifika empirických problémů v ekonomii.

Ekonometrie



- Klasická – prevažujúci prístup (Základy ekonometrie, Časové řady, Ekonometrie).
- Bayesiánská – súčasný rozvoj díky pokrokům ve výpočetní technice (Bayesiánská analýza).

K čemu je ekonometrie dobrá?

- Jak prakticky, efektivně, systematicky a vědecky používat data k zodpovězení různých ekonomických otázek a problémů.
 - Ovlivní devizové intervence centrální bankou ekonomiku? A jak hodně?
 - Mají dlouhodobě nezaměstnaní větší problém nalézt zaměstnání než krátkodobě nezaměstnaní?
 - ovlivní zvýšení minimální mzdy zaměstnanost a nezaměstnanost?
 - Předchází růst akciových trhů růstu ekonomiky nebo je tomu naopak?
 - Jsou akciové trhy v různých částech světa propojeny? Jak silně? A ovlivnila toto propojení nějak finanční krize z roku 2008?
 - Jaký vliv na trestnou činnost bude mít dodatečná finanční injekce na vyslání více uniformovaných policistů do ulic zkoumaného města?
 - Jaký efekt má investice do určitého typu reklamní kampaně na prodejnost konkrétního výrobku?
 - Co a jak ovlivňuje rozhodnutí zákazníka o nákupu některého z řady vzájemně si konkurujících výrobků?
 - Jsou spolu trhy vlakové a autobusové přepravy na konkrétní trase propojeny? Jak silně reagují zákazníci na změny cen jednoho z druhů přepravy?

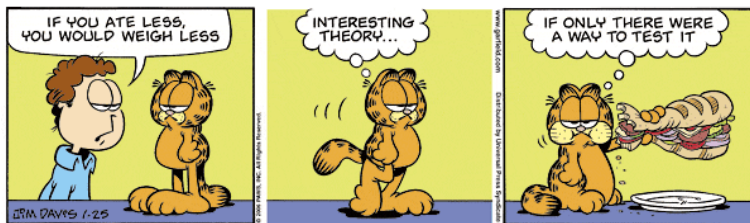
Jaké další otázky může ekonometrie zodpovědět?

- Nástroje velmi univerzální (viz např. témata projektů ze Základů ekonometrie).
 - Jaké faktory ovlivňují výši spropitného, které dávají hosté v restauracích?
 - Jaké dopady má zavedení či zvýšení ekologické daně na registraci ojetých automobilů?
 - Jaké faktory ovlivňují počet sexuální partnerů studentů?
 - Co může ovlivňovat výkon běžce na 10 km?
 - Co rozhoduje o spotřebě alkoholu mezi studenty?
 - Jaké charakteristiky filmů ovlivňují pravděpodobnost získání ocenění Český lev?

Jaké odpovědi nám může ekonometrie pomoci najít?

- *Zahálení u televize a počítače má na studijní průměr negativní dopady. S každou hodinou, kterou student průměrně denně stráví těmito aktivitami, se mu průměr zhorší o 0,036.*
- *Při rozhodování, který cvičící se zdá být nejlepším nám model vyhodnotil Ing. Michala Kvasničku, Ph.D.*
- *Pokud počet narozených dětí stoupne o jedno, ovlivní to celkovou populaci asi o 3,63881 jedince.*
- *Další člověk, který se přistěhuje, způsobí pokles populace o -1,30902 jedince.*
- *Pokud se respondent pokoušel osvojit si elfské písmo či trpasličí runy, pravděpodobnost návštěvy promítání filmu Hobit se snižuje o celých 42 %.*
- *V práci jsem zjistil, že jediným významným faktorem ovlivňujícím dobu běhu je uběhnutá vzdálenost.*
- *Bohužel jsem z mého výzkumu nic nezjistila, ale byla to pro mě zajímavá zkušenost.*

Ekonometrie a testování hypotéz



Obsah tématu

1 Úvod do ekonometrie

- Ekonomický model
- Ekonometrický model

2 Práce s daty

- Typy ekonomických dat
- Grafické metody práce s daty
- Popisné statistiky a korelace

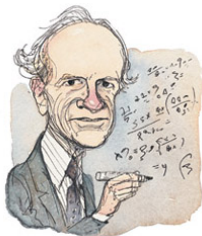
3 Ekonometrický software a studijní literatura

Příklady ekonomických modelů

- Empirická analýza – využití dat k testování teorie nebo k odhadu vztahů mezi veličinami.
- Pro testování ekonomických teorií – konstrukce ekonomického modelu.
- Příklady:
 - Ekonomický model kriminality.
 - Pracovní školení a produktivita pracovníků.
- Formální ekonomický model = základ ekonometrické analýzy.

Ekonomický model kriminality – úvod

- Gary Becker – nositel Nobelovy ceny, model kriminálního chování (1968).



Gary S. Becker (*1930)

- Maximalizace užitku – výnosy a náklady kriminální činnosti.
- Množství času věnovaného trestné činnosti jako funkce různých faktorů.

Ekonomický model kriminality

$$y = f(x_1, x_2, x_3, x_4, x_5, x_6, x_7)$$

y = hodiny strávené kriminálními aktivitami,

x_1 = „hodinová mzda“ z kriminální aktivity,

x_2 = hodinová mzda v legálním zaměstnání,

x_3 = další příjem (mimo zaměstnání a trestnou činnost),

x_4 = pravděpodobnost dopadení,

x_5 = pravděpodobnost odsouzení v případě dopadení,

x_6 = očekávaný trest v případě odsouzení,

x_7 = věk.

Ekonomický model kriminality – specifikace

- Reprezentativní faktory ovlivňující rozhodování o páchání trestné činnosti.
- Funkce $f(\cdot)$ obvykle nespecifikována, závisí na podobě užitkové funkce (málokdy známa).
- Ekonometrická analýza – odhad vlivu jednotlivých faktorů na páchání trestné činnosti jednotlivců.

Pracovní školení a produktivita pracovníků – úvod

- Vládou financovaný program školení pracovníků (práce s počítačem, 20 týdnů, účast na jednotlivých částech dobrovolná) – analýza efektivity.
- Zkoumání vlivu školení na produktivitu pracovníka.
- Menší potřeba ekonomické teorie – dostačující ekonomická intuice, že vzdělání, pracovní zkušenosti a školení ovlivňují produktivitu.
- Ekonomický předpoklad – pracovníci placeni v závislosti na své produktivitě.

Pracovní školení a produktivita pracovníků – model

$$wage = f(educ, exper, training)$$

wage = hodinová mzda,

educ = roky strávené ve škole,

exper = roky pracovních zkušeností,

training = počet týdnů zapojení do programu školení pracovníků..

- Možné další faktory ovlivňující mzdu × rovnice dostatečně pokrývá řešený problém.

Obsah tématu

1 Úvod do ekonometrie

- Ekonomický model
- **Ekonometrický model**

2 Práce s daty

- Typy ekonomických dat
- Grafické metody práce s daty
- Popisné statistiky a korelace

3 Ekonometrický software a studijní literatura

Ekonomický a ekonometrický model

- Specifikace ekonomického modelu \Rightarrow ekonometrický model.
- Specifikace funkční závislosti mezi proměnnými, $f(\cdot)$.
- Vypořádání se s přímo nepozorovanými proměnnými, např.:
 - „mzda“ jednotlivců z trestné činnosti – proměnná dobře definována, ale obtížně pozorovatelná;
 - pravděpodobnost zatčení – pro jednotlivce obtížně zjistitelná, možno využít globální statistiku zatčení a odvodit odpovídající „proxy“ proměnnou.

Ekonometrický model kriminality

$$\begin{aligned} \text{crime} = & \alpha + \beta_1 \text{wage}_m + \beta_2 \text{othinc} + \beta_3 \text{freqarr} + \beta_4 \text{freqconv} \\ & + \beta_5 \text{avgsen} + \beta_6 \text{age} + \epsilon \end{aligned}$$

crime = nějaké měřítko četnosti kriminální aktivity,

wage_m = mzda z legálních aktivit,

othinc = příjem z jiných zdrojů (akcie apod.),

freqarr = četnost zatčení před vykonáním trestného činu
aproximuje pravděpodobnost zatčení,

freqconv = četnost odsouzení,

avgsen = průměrná délka trestu po odsouzení,

age = věk.

Ekonometrický model kriminality – vysvětlení

- Volba proměnných dle ekonomické teorie.
- **náhodná složka**: ϵ = nepozorované faktory, „mzda“ z trestné činnosti, rodinné zázemí, chyby v měření kriminální aktivity, pravděpodobnosti zatčení apod.
- **parametry**: $\alpha, \beta_1, \dots, \beta_6$ = měřítko směru a síly vztahu mezi **vysvětlovanou proměnnou**, *crime*, a **vysvětlujícími proměnnými**, faktory determinujícími kriminalitu v modelu.

Ekonometrický model pracovního školení a produktivity pracovníků

$$wage = \alpha + \beta_1 educ + \beta_2 exper + \beta_3 training + \epsilon$$

wage = hodinová mzda,

educ = roky strávené ve škole,

exper = roky pracovních zkušeností,

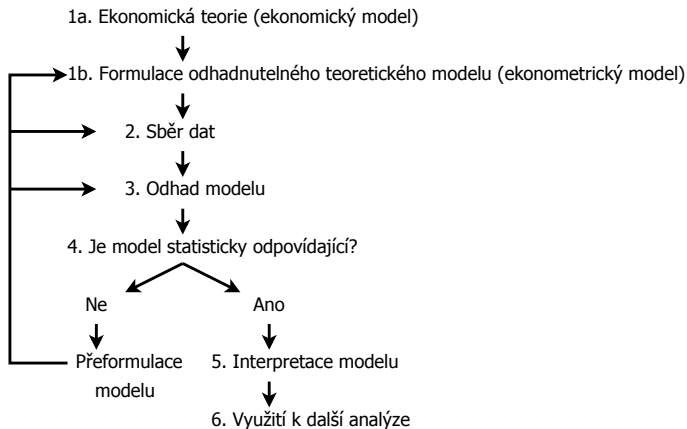
training = počet týdnů zapojení do programu školení pracovníků.

- ϵ = faktory kvality vzdělání, rodinného zázemí a další, které ovlivňují mzdu.
- β_3 = parametr vyjadřující vliv školení na mzdu.

Další otázky

- Ekonometrická analýza i bez detailů konstrukce ekonomického modelu (odvození mnohdy náročné).
- Přímé vyjádření ekonomického modelu skrze ekonometrický model (na základě ekonomických úvah o možných proměnných).
- Obvyklý postup \times omezení ekonomické analýzy.
- Stanovení **hypotéz** prostřednictvím neznámých parametrů (např. mzda z legálního zaměstnání $wage_m$ nemá vliv na kriminální chování $\Leftrightarrow \beta_1 = 0$).

Schéma postupu ekonometrického modelování



Postup ekonometrického modelování

- 1 Formulace ekonometrického modelu – na základě ekonomické teorie či ekonomické intuice o vztahu mezi zkoumanými proměnnými.
- 2 Sběr relevantních dat – z dostupných zdrojů nebo vlastní tvorba (např. dotazníky).
- 3 Volba vhodné odhadové metody – např. pro jednorovnicový nebo víceroovnicový model.
- 4 Statistické hodnocení modelu – ověření předpokladů kladených na odhadové techniky, data nebo model.
- 5 Teoretické hodnocení modelu – např. soulad parametrů (znamének) s ekonomickou teorií.
- 6 Využití modelu – testování ekonomických hypotéz, predikce, návrhy pro praktické rozhodování (na vládní nebo soukromé úrovni; dle povahy řešeného problému).

Obsah tématu

- 1 Úvod do ekonometrie
 - Ekonomický model
 - Ekonometrický model
- 2 Práce s daty
 - Typy ekonomických dat
 - Grafické metody práce s daty
 - Popisné statistiky a korelace
- 3 Ekonometrický software a studijní literatura

Obsah tématu

- 1 Úvod do ekonometrie
 - Ekonomický model
 - Ekonometrický model
- 2 Práce s daty
 - Typy ekonomických dat
 - Grafické metody práce s daty
 - Popisné statistiky a korelace
- 3 Ekonometrický software a studijní literatura

Časové řady

- Pozorování proměnných pro nějakou jednotku (stát, firma, apod.) v čase.

$$Y_t \quad t = 1, \dots, T$$

- Mállokdy splněn předpoklad o nezávislosti pozorování v čase.
- Frekvence dat.
- Sezónní charakter – potřeba očištění.

Časové řady – příklad

Tabulka: Minimální mzda, nezaměstnanost a další data pro Puerto Rico

| Obsno | Year | Avgmin | Avgcov | Unemp | GNP |
|-------|------|--------|--------|-------|--------|
| 1 | 1950 | 0.20 | 20.1 | 15.4 | 878.7 |
| 2 | 1951 | 0.21 | 20.7 | 16.0 | 925.0 |
| 3 | 1952 | 0.23 | 22.6 | 14.8 | 1015.9 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 37 | 1986 | 3.35 | 58.1 | 18.9 | 4281.6 |
| 38 | 1987 | 3.35 | 58.2 | 16.8 | 4496.7 |

Časové řady – příklad (komentář)

- Data z Wooldridge (2009), dostupná v gretlu (*prminwge.gdt*).
- Původně: Castillo-Freeman and Freeman (1992) – vliv minimální mzdy v Puerto Rico.

Obsno = pozorování;

Year = rok;

Avgmin = průměrná (hodinová) minimální mzda v roce;

Avgcov = průměrná míra pokrytí

procento pracovníků beroucích minimální mzdu;

Unemp = míra nezaměstnanosti;

GNP = hrubý národní produkt.

Průřezová data

- Data pro řadu různých jednotek (jednotlivci, domácnosti, firmy, města, státy) v daném časovém okamžiku.

$$Y_i \quad i = 1, \dots, N$$

- Charakter náhodného výběru – někdy problém.
- Souhrnná průřezová data (pooled cross sections) – **různé** jednotky pro více časových okamžiků (vhodné pro analýzu strukturálních změn mezi obdobími, např. změna daňových sazeb).

Průřezová data – příklad 1

Tabulka: Průřezová data o mzdách a dalších individuálních charakteristikách

| Obsno | Wage | Educ | Exper | Female | Married |
|-------|-------|------|-------|--------|---------|
| 1 | 3.10 | 11 | 2 | 1 | 0 |
| 2 | 3.24 | 12 | 22 | 1 | 1 |
| 3 | 3.00 | 11 | 2 | 0 | 0 |
| 4 | 6.00 | 8 | 44 | 0 | 1 |
| 5 | 5.30 | 12 | 7 | 0 | 1 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 525 | 11.56 | 16 | 5 | 0 | 1 |
| 526 | 3.50 | 14 | 5 | 1 | 0 |

Průřezová data – příklad 1 (komentář)

- Data z Wooldridge (2009), dostupná v gretlu (*wage1.gdt*).
- Zkrácená data o 526 pracovnících za rok 1976.

Obsno = pozorování;

Wage = hodinová mzda;

Educ = počet let vzdělání;

Exper = počet let pracovních zkušeností;

Female = indikátor pohlaví;

Married = indikátor rodinného vztahu.

Průřezová data – příklad 2

Tabulka: Data o mírách růstu a dalších charakteristikách

| Obsno. | Country | Gpcrgdp | Govcons60 | Second60 |
|--------|-----------|---------|-----------|----------|
| 1 | Argentina | 0.89 | 9 | 32 |
| 2 | Austria | 3.32 | 16 | 50 |
| 3 | Belgium | 2.56 | 13 | 69 |
| 4 | Bolivia | 1.24 | 18 | 12 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 61 | Zimbabwe | 2.30 | 17 | 6 |

Průřezová data – příklad 2 (komentář)

- Data z Wooldridge (2009).
- Původně: De Long a Summers (1991) – studie měř růstu mezi zeměmi.

Obsno = pozorování;

Country = stát;

Gpcrgdp = průměrný růst HDP na osobu mezi lety 1960–1958;

Govcons60 = vládní spotřeba jako procento HDP;

Second60 = procento dospělých se sekundárním vzděláním.

Souhrnná průřezová data – příklad

Tabulka: Ceny domů a další charakteristiky ve dvou letech

| Obsno. | Year | Hprice | Proptax | Sqrfc | Bdrms |
|--------|------|--------|---------|-------|-------|
| 1 | 1993 | 85500 | 42 | 1600 | 3 |
| 2 | 1993 | 67300 | 36 | 1440 | 3 |
| 3 | 1993 | 134000 | 38 | 2000 | 4 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 250 | 1993 | 243600 | 41 | 2600 | 4 |
| 251 | 1995 | 65000 | 16 | 1250 | 2 |
| 252 | 1995 | 182400 | 20 | 2200 | 4 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 520 | 1995 | 57200 | 16 | 1100 | 2 |

Souhrnná průřezová data – příklad (komentář)

- Data z Wooldridge (2009).
- Data o prodejních cenách domů v letech 1993 (250 pozorování) a 1995 (270 pozorování).

Obsno = pozorování;

Year = rok;

Hprice = cena domu;

Proptax = daň z nemovitosti;

Sqrft = rozloha domu ve čtverečních stopách;

Bdrms = počet ložnic.

Problém náhodného výběru - příklad

- Zkreslení odhadu při nenáhodném výběru.
- Prezidentské volby USA (1936); průzkum veřejného mínění *Literary Gazette* → Alf M. London drtivě porazí Franklina D. Roosevelta (57 % × 43 %).
- Skutečnost: Roosevelt drtivě zvítězil (59 % × 41 %).
- Proč rozpor mezi průzkumem a realitou?
- Výběr náhodného vzorku na základě telefonních seznamů a seznamů majitelů automobilů.
- V roce 1936 – málo vlastníků automobilů a telefonů; vlastníci obvykle bohatší vrstvy volící Republikány ⇒ náhodný vzorek podstřelil počet Demokratů ⇒ zkreslení odhadů.
- Co internetové průzkumy dnes?

Panelová data

- Časová i prostorová dimenze.
- Data o stejných jednotkách (státech, městech, firmách) a jejich charakteristikách v průběhu několika let.

$$Y_{it} \quad i = 1, \dots, N \quad t = 1, \dots, T$$

- Náročnější na získávání.
- Možnost kontroly dodatečných vlivů.
- Informace o změnách chování v čase.

Panelová data – příklad

Tabulka: Panelová data kriminality ve městech

| Obsno. | Město | Rok | Murders | Population | Unem | Police |
|--------|-------|------|---------|------------|------|--------|
| 1 | 1 | 1986 | 5 | 350000 | 8.7 | 440 |
| 2 | 1 | 1990 | 8 | 359200 | 7.2 | 471 |
| 3 | 2 | 1986 | 2 | 64300 | 5.4 | 75 |
| 4 | 2 | 1990 | 1 | 65100 | 5.5 | 75 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 297 | 149 | 1986 | 10 | 260700 | 9.6 | 286 |
| 298 | 149 | 1990 | 6 | 245000 | 9.8 | 334 |
| 299 | 150 | 1986 | 25 | 543000 | 4.3 | 520 |
| 300 | 150 | 1990 | 32 | 546200 | 5.2 | 493 |

Panelová data – příklad (komentář)

- Data z Wooldridge (2009).
- Data o kriminalitě ve 150 městech ve Spojených státech v průběhu dvou let (1986 a 1990).

Obsno = pozorování;

City = město;

Year = rok;

Murders = počet vražd;

Population = počet obyvatel;

Unem = míra nezaměstnanosti;

Police = počet příslušníků policie.

Transformace dat

- Dle potřeby empirické analýzy.
- Diference: $\Delta Y_t = Y_t - Y_{t-1}$.
- Tempo růstu (procentní změna):

$$\% \Delta Y_t = \frac{Y_t}{Y_{t-1}} - 1 = \frac{Y_t - Y_{t-1}}{Y_{t-1}} \quad (\times 100[\%])$$

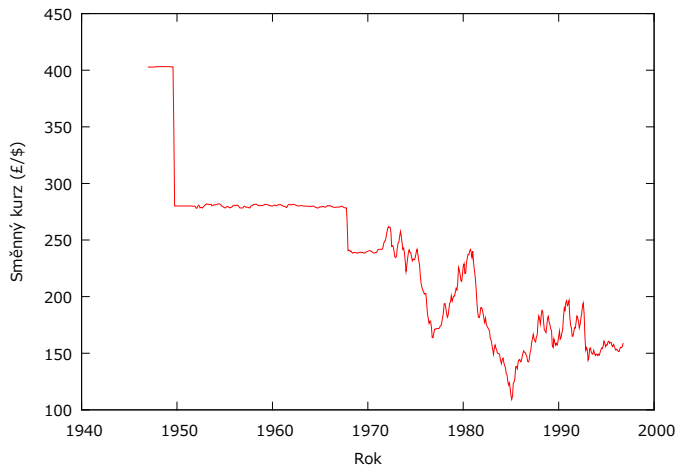
- Diference logaritmů (aproximace tempa růstu):

$$\% \Delta Y_t \approx \ln(Y_t) - \ln(Y_{t-1}) \quad (\times 100[\%])$$

Obsah tématu

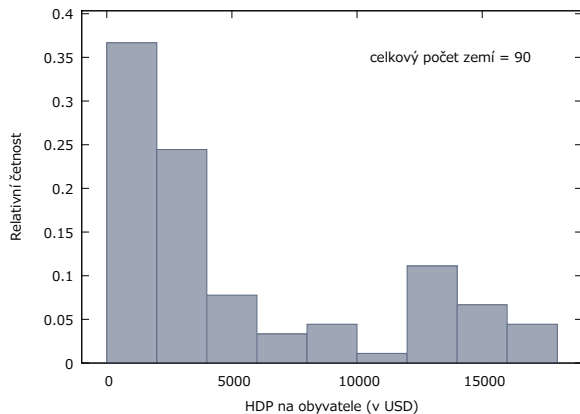
- 1 Úvod do ekonometrie
 - Ekonomický model
 - Ekonometrický model
- 2 **Práce s daty**
 - Typy ekonomických dat
 - **Grafické metody práce s daty**
 - Popisné statistiky a korelace
- 3 Ekonometrický software a studijní literatura

Spojnicové grafy



Obrázek: Časová řada vývoje směnného kurzu GBP vzhledem k USD.

Histogramy



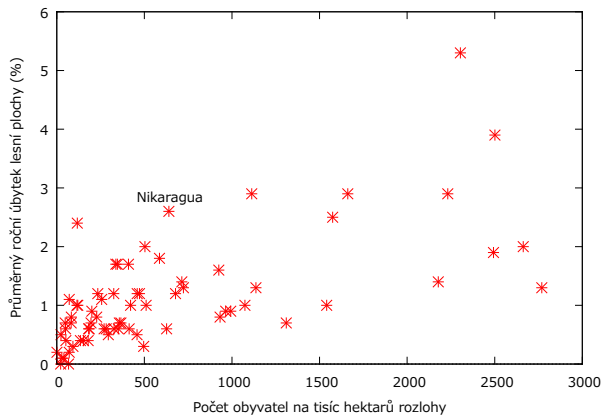
Obrázek: Histogram HDP na osobu pro 90 zemí světa.

Absolutní a relativní četnost

Tabulka: Tabulka četnosti pro HDP na osobu.

| Interval (USD) | Četnost | |
|----------------|-----------|-----------|
| | Absolutní | Relativní |
| 0-2000 | 33 | 36.67 % |
| 2001-4000 | 22 | 24.44 % |
| 4001-6000 | 7 | 7.78 % |
| 6001-8000 | 3 | 3.33 % |
| 8001-10000 | 4 | 4.44 % |
| 10001-12000 | 2 | 2.22 % |
| 12001-14000 | 9 | 10.00 % |
| 14001-16000 | 6 | 6.67 % |
| 16001-18000 | 4 | 4.44 % |

Bodové grafy (grafy XY)



Obrázek: Bodový graf odlesnění vzhledem k hustotě obyvatelstva.

Obsah tématu

1 Úvod do ekonometrie

- Ekonomický model
- Ekonometrický model

2 Práce s daty

- Typy ekonomických dat
- Grafické metody práce s daty
- **Popisné statistiky a korelace**

3 Ekonometrický software a studijní literatura

Výběrová střední hodnota a rozptyl

- Náhodný výběr: Y_1, \dots, Y_N .
- Výběrový průměr (výběrová střední hodnota):

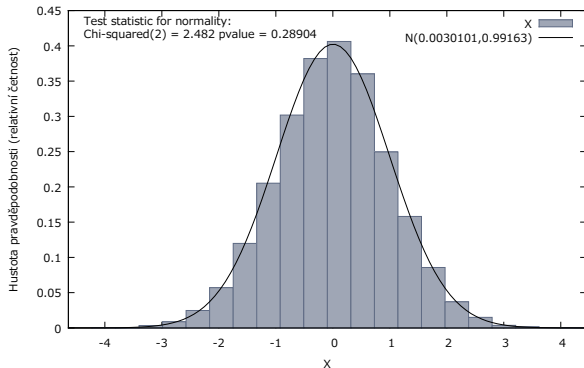
$$\bar{Y} = \frac{\sum_{i=1}^N Y_i}{N}$$

- Výběrová směrodatná odchylka:

$$s_Y = \sqrt{\frac{\sum_{i=1}^N (Y_i - \bar{Y})^2}{N - 1}}$$

- Výběrový rozptyl: s_Y^2 .

Histogram a zvonové rozdělení



Obrázek: Histogram pro zvonové rozdělení.

Očekávaná hodnota a rozptyl

- Náhodná veličina Y .
- Očekávaná hodnota a populační střední hodnota: $E(Y) \equiv \mu$

$$E(Y) = \sum_{i=1}^N y_i p(y_i) \quad \text{diskrétní případ, výběrový prostor } \{y_1, \dots, y_N\}$$

$$E(Y) = \int_{-\infty}^{\infty} yf(y)dy \quad \text{spojitý případ}$$

- Populační rozptyl: $\text{var}(Y) = \sigma^2$

$$\text{var}(Y) = E[(Y - \mu)^2] = E(Y^2) - \mu^2$$

Příklad – očekávaný výnos akcie

- S pravděpodobností 70 % (0.7) budou trhy stabilní (výnos 1 %); s 10 % pravděpodobností dojde k pádu trhů (výnos -10 %); 20% pravděpodobnost růstu nálady na trzích (výnos 5 %).
- $p(y_i) = \Pr(Y = y_i)$.

$$\begin{aligned} E(Y) &= p(0.05)0.05 + p(0.01)0.01 + p(-0.10)(-0.10) \\ &= 0.20 \times 0.05 + 0.70 \times 0.01 + 0.10 \times (-0.10) \\ &= 0.007. \end{aligned}$$

- Očekávaná hodnota výnosu akcie v příštím měsíci 0.7% (o něco méně než 1%).

Příklad – očekávaná nejistota výnosu akcie

- Pro příklad rizika očekávaného výnosu akcie máme možné realizace pro Y^2 : $(0.05)^2 = 0.0025$, $(0.01)^2 = 0.0001$ a $(-0.10)^2 = 0.01$.

$$\begin{aligned} E(Y^2) &= p(0.0025) \times 0.0025 + p(0.0001) \times 0.0001 \\ &\quad + p(0.01) \times 0.01 \\ &= 0.20 \times 0.0025 + 0.70 \times 0.0001 + 0.10 \times 0.01 \\ &= 0.00157. \end{aligned}$$

$$\begin{aligned} \text{var}(Y) &= E(Y^2) - [E(Y)]^2 \\ &= 0.00157 - (0.007)^2 \\ &= 0.001521. \end{aligned}$$

- Odmocněním získáme hodnotu 0.039. Očekávaný výnos je tedy 0.7% s nejistotou odpovídající $\pm 3.9\%$.

Korelace

- Vztah mezi dvěma proměnnými, X a Y .
- Korelace (korelační koeficient):

$$r = \frac{\sum_{i=1}^N (Y_i - \bar{Y})(X_i - \bar{X})}{\sqrt{\sum_{i=1}^N (Y_i - \bar{Y})^2} \sqrt{\sum_{i=1}^N (X_i - \bar{X})^2}}.$$

Vlastnosti korelace

- 1 $r \in \langle -1, 1 \rangle$.
- 2 $r > 0$... pozitivní korelace; $r < 0$... negativní korelace; $r = 0$... nekorelovanost; $r = 1$ nebo $r = -1$... perfektní korelace.
- 3 $r_{XY} = r_{YX}$.
- 4 $r_{XX} = 1$.

Možné interpretace hodnot korelace

- De Vaus (2002) – Analyzing Social Science Data: 50 Key Problems in Data Analysis.
- Nutno brát s rezervou.

| Hodnota korelace (v abs. hodnotě) | Interpretace |
|-----------------------------------|--------------------------|
| 0.01 – 0.09 | triviální, žádná |
| 0.10 – 0.29 | nízká až střední |
| 0.30 – 0.49 | střední až podstatná |
| 0.50 – 0.69 | podstatná až velmi silná |
| 0.70 – 0.89 | velmi silná |
| 0.90 – 1.00 | téměř perfektní |

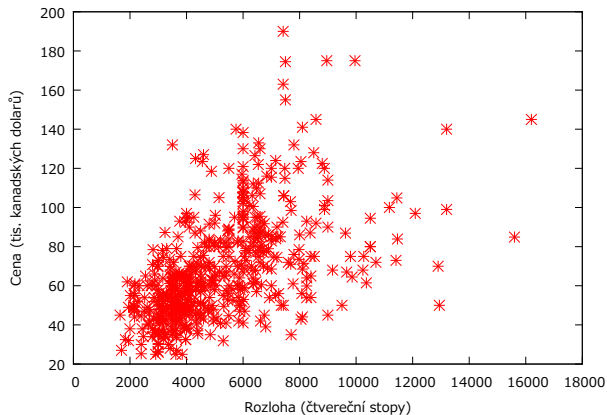
Příklad – ceny domů

- Datový soubor *hprice.gdt*, Koop (2008)
- Data o $N = 546$ domech prodaných ve Windsoru, v Kanadě, v průběhu léta roku 1987.
- $Y =$ prodejní cena domu, $X =$ rozloha domu ve čtverečních stopách
 $\Rightarrow r_{XY} = 0.54$
 - 1 Domy s větší rozlohou mají tendenci mít vyšší hodnotu než domy s rozlohou malou.
 - 2 Existuje pozitivní souvislost mezi rozlohou a prodejní cenou.
 - 3 Variabilita v rozloze domů vysvětluje 29 % (tzn. $0.54^2 = 0.29$) variability cen domů.
- $Z =$ počet ložnic $\Rightarrow r_{YZ} = 0.37$: domy s větším počtem ložnic mají tendenci mít vyšší hodnotu než domy s menším počtem ložnic.
- $r_{XZ} = 0.15 \Rightarrow$ domy s vyšší rozlohou mají tendenci mít vyšší počet ložnic (korelační koeficient relativně nízký \Rightarrow slabý vztah mezi velikostí domu a počtem ložnic).

Korelace a kauzalita

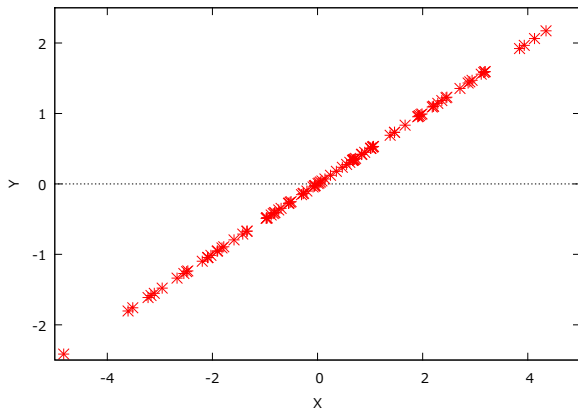
- Korelace \neq kauzalita.
- Možná existence třetí proměnné.
- Přímá a nepřímá kauzalita.
- Důležité znalosti o zkoumaném problému.

Korelace a bodový graf – příklad 1



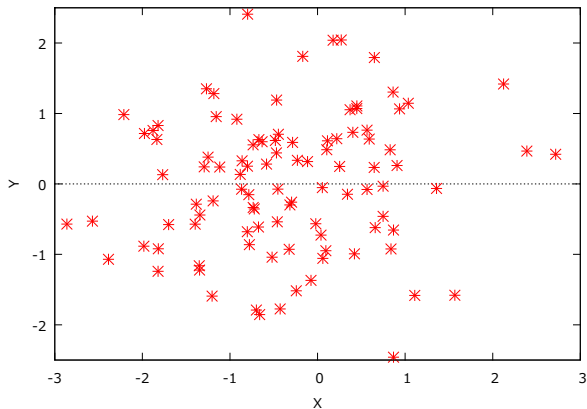
Obrázek: Bodový graf rozlohy domů vzhledem k jejich ceně.

Korelace a bodový graf – příklad 2



Obrázek: Bodový graf dvou dokonale korelovaných proměnných.

Korelace a bodový graf – příklad 3



Obrázek: Bodový graf nekorelovaných proměnných.

Korelační matice

- Korelace pro více proměnných.

| | X | Y | Z |
|-----|--------|-------|-------|
| X | 1.000 | | |
| Y | 0.318 | 1.000 | |
| Z | -0.131 | 0.097 | 1.000 |

Populační korelace a kovariance

- Příklad: rizikovost portfolia; investice do akcií společnosti na výrobu deštníků a na výrobu zmrzliny.
- Populační kovariance:

$$\text{cov}(X, Y) = E(XY) - E(X)E(Y)$$

- Populační korelace:

$$\text{corr}(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)\text{var}(Y)}}$$

- Výběrové statistiky jako odhad populačních protějšků.

Obsah tématu

- 1 Úvod do ekonometrie
 - Ekonomický model
 - Ekonometrický model

- 2 Práce s daty
 - Typy ekonomických dat
 - Grafické metody práce s daty
 - Popisné statistiky a korelace

- 3 Ekonometrický software a studijní literatura

Software a literatura

- Software: komerční vs. open source; „klikací“ vs. skriptové nebo kombinované.
- Literatura: spousta učebnic, odlišnosti v rozsahu témat a náročnosti; různý styl výkladu.
- Přehled – viz první kapitola učebního textu.