4EK211 Základy ekonometrie

Odhad klasického lineárního regresního modelu II

Cvičení 3

Klasický lineární regresní model - zadání příkladu

Soubor: CV3_PR1.xls

Data: ***y*** = maloobchodní obrat potřeb pro domácnost v mld. CZK
$\boldsymbol{x_1}$ = disponibilní příjem v mld. CZK
$\boldsymbol{x_2}$ = cenový index

Zadání: Odhadněte závislost maloobchodního obratu (***y***) na disponibilním příjmu ($\boldsymbol{x_1}$) a cenovém indexu ($\boldsymbol{x_2}$).
Proveďte specifikaci, kvantifikaci, verifikaci a aplikaci EM.

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i,\ i = 1, 2, \ldots, 8$$

Klasický lineární regresní model - specifikace

Specifikace EM

- **určení proměnných**

  y = endogenní (vysvětlovaná) proměnná

  $\boldsymbol{x_1}$ = exogenní (vysvětlující) proměnná

  $\boldsymbol{x_2}$ = exogenní (vysvětlující) proměnná

- **určení vzájemných vazeb mezi proměnnými (forma závislosti) + formulace hypotéz**

  $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i, \qquad i = 1, 2, ..., 8$

- **předpokládané znaménka, očekávané hodnoty odhadnutých parametrů**

  $\boldsymbol{\beta_1}$ = v intervalu (0,1) pokud nepracujeme s úsporami nebo > 0 s úsporami

  $\boldsymbol{\beta_2}$ = by mělo být < 0

Klasický lineární regresní model - kvantifikace

Kvantifikace EM

- odhad modelu MNČ, MS Excel nebo GiveWin (PcGive)

```
EQ( 1) Modelling y by OLS (using CV3_PR1.xls)
       The estimation sample is: 1966 to 1973

                  Coefficient  Std.Error  t-value  t-prob Part.R^2
Constant              3.01620      1.032     2.92   0.033   0.6308
x1                   0.103550   0.004550     22.8   0.000   0.9904
x2                 -0.0979638    0.01583    -6.19   0.002   0.8845

sigma                0.120682  RSS                 0.0728202076
R^2                  0.997094  F(2,5) =      857.8 [0.000]**
log-likelihood        7.44531  DW                          1.95
no. of observations         8  no. of parameters              3
mean(y)                  10.5  var(y)                    3.1325
```

Klasický lineární regresní model - kvantifikace

Kvantifikace EM

- zápis odhadnutého regresního modelu (regresní nadroviny)

	Coefficient
Constant	3.01620
x1	0.103550
x2	-0.0979638

  - napozorované hodnoty:

    $$Y_i = 3{,}016 + 0{,}104X_{1i} - 0{,}098X_{2i} + e_i$$

  - vyrovnané hodnoty:

    $$\hat{Y}_i = 3{,}016 + 0{,}104X_{1i} - 0{,}098X_{2i}$$

- je to tzv. **bodový odhad**

Klasický lineární regresní model - verifikace

Verifikace ekonomická

- **předpokládaná znaménka, očekávané hodnoty odhadnutých parametrů**

  $\boldsymbol{\beta_1}$ = v intervalu (0,1) pokud nepracujeme s úsporami nebo > 0 s úsporami → b_1 splňuje předpoklad

  $\boldsymbol{\beta_2}$ = by mělo být < 0 → b_2 splňuje předpoklad

- **ekonomická interpretace**

  b_0 – bez interpretace

  b_1 – absolutní (příjmová) pružnost $b_1 = \frac{\partial Y}{\partial X_1}$

  b_1 = 0,104 → vzroste-li disponibilní příjem x_1 o 1 jednotku (tj. o 1 mld. CZK) a <u>x_2 se nezmění</u>, vzroste maloobchodní obrat potřeb pro domácnost <u>v průměru</u> o 0,104 mld. CZK

Klasický lineární regresní model - verifikace

Verifikace ekonomická

b_2 – absolutní (cenová) pružnost $b_2 = \dfrac{\partial Y}{\partial X_2}$

b_2 = -0,098 → vzroste-li cenový index x_2 o jeden procentní bod a x_1 se nezmění, klesne maloobchodní obrat potřeb pro domácnost v průměru o 0,098 mld. CZK

- b_1 i b_2 jsou definovány v jednotkách pozorovaných proměnných.

- **koeficienty relativní pružnosti – *q***
- počítá se vždy vzhledem ke konkrétnímu pozorování
- Koeficient (výsledek přímo) interpretujeme v %.

- koeficient příjmové pružnosti $q_{x_1} = \dfrac{\partial Y}{\partial X_1}\dfrac{X_1}{Y} = b_1 \dfrac{X_1}{Y}$

- koeficient cenové pružnosti $q_{x_2} = \dfrac{\partial Y}{\partial X_2}\dfrac{X_2}{Y} = b_2 \dfrac{X_2}{Y}$

Klasický lineární regresní model - verifikace

Verifikace ekonomická

- relativní pružnost pro rok 1973

  $Y_{(73)} = 13{,}6 \qquad X_{1(73)} = 209 \qquad X_{2(73)} = 113$

- **koeficient příjmové pružnosti** $q_{x_1(73)} = b_1 \dfrac{x_1}{y} = 0{,}104 \dfrac{209}{13{,}6} = 1{,}60\ldots = 1{,}60\ \%$

  – zvýší-li se v roce 1973 disponibilní příjem x_1 o 1 % a cenový index x_2 se nezmění, vzroste maloobchodní obrat potřeb pro domácnost y v průměru o 1,60 %

- **koeficient cenové pružnosti** $q_{x_2(73)} = b_2 \dfrac{x_2}{y} = -0{,}098 \dfrac{113}{13{,}6} = -0{,}81\ldots = -0{,}81\%$

  – zvýši-li se v roce 1973 cenový index x_2 o 1 % a disponibilní příjem x_1 se nezmění, klesne maloobchodní obrat potřeb pro domácnost y v průměru o 0,81 %

Klasický lineární regresní model - verifikace

Verifikace statistická

	Coefficient	Std.Error	t-value	t-prob	Part.R^2
Constant	3.01620	1.032	2.92	0.033	0.6308
x1	0.103550	0.004550	22.8	0.000	0.9904
x2	-0.0979638	0.01583	-6.19	0.002	0.8845

sigma 0.120682 RSS 0.0728202076

Standard error

- standardní chyba regresního koeficientu
- slouží k určení významnosti parametrů, k intervalovým odhadům

$$s_{b_i} = s\sqrt{(X^T X)^{-1}_{ii}}$$

kde $s = \sqrt{\frac{1}{(n-(k+1))}\sum e^T e}$ nebo **sigma** ve výstupu a $e^T e$ = **RSS**

ii – prvek z diagonály momentové matice, pro náš příklad:

$$(X^T X)^{-1} = \begin{pmatrix} 73{,}111 & 0{,}234 & -1{,}061 \\ 0{,}234 & 0{,}001 & -0{,}004 \\ -1{,}061 & -0{,}004 & 0{,}017 \end{pmatrix} \qquad s_{b_0} = \sqrt{\frac{1}{(8-3)}0{,}073}\sqrt{73{,}111} = 1{,}032$$

Klasický lineární regresní model - verifikace

Verifikace statistická

	Coefficient	Std.Error	t-value	t-prob	Part.R^2
Constant	3.01620	1.032	2.92	0.033	0.6308
x1	0.103550	0.004550	22.8	0.000	0.9904
x2	-0.0979638	0.01583	-6.19	0.002	0.8845

t-value = *t*-statistika, t-poměr

- *t*-statistika slouží k určení významnosti jednotlivých parametrů v modelu
- testuje se hypotéza: H_0: $\beta_j = 0$, H_1: $\beta_j \neq 0$

$$t_j = \frac{b_j - \beta_j}{s_{b_j}}$$

- obecně pro *t*-statistiku platí

$$|t_j| = \left|\frac{b_j}{s_{b_j}}\right| = \frac{Koefcient}{Sm.\,odchylka}$$

- $|t_j| > t^*_{1-\alpha/2\,(n-k-1)} \rightarrow$ nezamítám hypotézu H_1 o významnosti proměnné v modelu: předpokládám, že proměnná je **významná**
- $|t_j| \leq t^*_{1-\alpha/2\,(n-k-1)} \rightarrow$ nezamítám hypotézu H_0 o nevýznamnosti proměnné v modelu: předpokládám, že proměnná je **nevýznamná**

Klasický lineární regresní model - verifikace

Verifikace statistická

	Coefficient	Std.Error	t-value	t-prob	Part.R^2
Constant	3.01620	1.032	2.92	0.033	0.6308
x1	0.103550	0.004550	22.8	0.000	0.9904
x2	-0.0979638	0.01583	-6.19	0.002	0.8845

t-prob (v Gretlu: p-value) t-pravděpodobnost, p-hodnota

- pravděpodobnost, že nulová hypotéza je pravdivá (tj. daná vysvětlující proměnná je v modelu nevýznamná)
- t-prob < 0,05 → proměnná je statisticky významná na 5% hladině
- t-prob < 0,01 → proměnná je statisticky významná na 1% hladině
- používáme místo práce s tabulkovými hodnotami *t*-statistiky

Part. R^2

- Parciální korelační koeficient: pro **dané ostatní exogenní proměnné**.
- pozor na odlišení od párového korelačního koeficientu typu: r_{XY} a vícenásobného koeficientu determinace R^2
- Pro X, Y, a $Z = \{Z_1, Z_2, \dots, Z_k\}$: Part. $R^2_{X,Y\,.\,Z}$ je r_{e1e2} kde *e1* a *e2* jsou rezidua z regrese X na Z a Y na Z (X,Y jako endog. proměnné LRM)
- viz Gujarati, D.: Basic Econometrics, 4th ed.: kap. 7.3 THE MEANING OF PARTIAL REGRESSION COEFFICIENTS

Klasický lineární regresní model - verifikace

Verifikace statistická

sigma	0.120682	RSS	0.0728202076
R^2	0.997094	F(2,5) =	857.8 [0.000]**
log-likelihood	7.44531	DW	1.95
no. of observations	8	no. of parameters	3
mean(y)	10.5	var(y)	3.1325

sigma

- standardní chyba regrese [$u \sim N(0,\sigma^2)$]
- charakteristika výběrového rozptylu, který dostaneme po kvantifikaci abstraktního modelu
- vzorec $$\boldsymbol{s} = \sqrt{\frac{1}{(\boldsymbol{n}-(\boldsymbol{k}+1))}\sum \boldsymbol{e}^T\boldsymbol{e}} = \sqrt{\frac{1}{(8-3)}0{,}073} = 0{,}1206$$
- užívá se při výpočtu standardní chyby regresního koeficientu

Klasický lineární regresní model - verifikace

Verifikace statistická

sigma	0.120682	RSS	0.0728202076
R^2	0.997094	F(2,5) =	857.8 [0.000]**
log-likelihood	7.44531	DW	1.95
no. of observations	8	no. of parameters	3
mean(y)	10.5	var(y)	3.1325

RSS

- součet čtverců reziduí = $\sum e_i^2 = \sum e^T e \rightarrow \min$
- užívá se při výpočtu sigma nebo R^2 (koeficientu vícenásobné determinace)

R^2 – koeficient vícenásobné determinace

- hodnotí celkovou kvalitu modelu, určuje, jak se model shoduje s daty

$$R^2 = \frac{VS\check{C}}{CS\check{C}} = 1 - \frac{NS\check{C}}{CS\check{C}} \qquad R^2 \in \langle 0,1 \rangle$$

$$CS\check{C} = \sum (Y_i - \bar{Y})^2 \quad VS\check{C} = \sum (\hat{Y}_i - \bar{Y})^2 \quad NS\check{C} = \sum (Y_i - \hat{Y}_i)^2$$

Klasický lineární regresní model - verifikace

Verifikace statistická

R² – koeficient vícenásobné determinace

- je-li NSČ = 0, pak R^2 = 1 → dokonalá shoda modelu s daty
- v případě, že koeficient R^2 není statisticky významný (viz *F*-poměr), doporučují se úpravy:
  - přidání další vysvětlující proměnné
  - zvýšení počtu pozorování
  - změna funkčního tvaru regresní rovnice
- nezohledňuje počet vysvětlujících proměnných – hodnota R^2 nikdy neklesne přidáním dalších vysvětlujících proměnných do modelu
- proto existuje korigovaný koeficient determinace (tj. R^2_{adj} nebo $\overline{R}^2$)

$$\overline{R}^2 = 1 - (1 - R^2)\frac{n-1}{n-(k+1)}$$

- rovnost jen pokud R^2 = 1 nebo k = 0

- ***R*² „PLATÍ“ JEN PRO LRM S ÚROVŇOVOU KONSTANTOU!**

Klasický lineární regresní model - verifikace

Verifikace statistická

sigma	0.120682	RSS	0.0728202076
R^2	0.997094	F(2,5) =	857.8 [0.000]**
log-likelihood	7.44531	DW	1.95
no. of observations	8	no. of parameters	3
mean(y)	10.5	var(y)	3.1325

***F*(*k*, *n*-*k*-1), zde: *F*(2,5)**

- F-poměr – testuje statistickou významnost modelu

  H_0: R^2 se statisticky významně neliší od nuly,
  altern.: LRM jako celek je statisticky nevýznamný,
  altern.: $\beta_1 = \ldots \beta_k = 0$; LRM dán pouze úrovňovou konstantou β_0.

  H_1: R^2 se statisticky významně liší od nuly,
  alt.: LRM jako celek je statisticky významný.

$$F = \frac{R^2}{1 - R^2} \frac{n - (k+1)}{k}$$

- $F > F^*_{(k,n-k-1)}$ … zamítáme H_0 ve prospěch H_1
- $F \leq F^*_{(k,n-k-1)}$ … nezamítáme H_0
- [p-value]:
  - [p-value] $\leq \alpha$ $\rightarrow$ zamítám hypotézu H_0, model je tedy významný
  - [p-value] $> \alpha$ $\rightarrow$ nezamítám H_0, model je tedy nevýznamný

Klasický lineární regresní model - verifikace

Verifikace statistická

sigma	0.120682	RSS	0.0728202076
R^2	0.997094	F(2,5) =	857.8 [0.000]**
log-likelihood	7.44531	DW	1.95
no. of observations	8	no. of parameters	3
mean(y)	10.5	var(y)	3.1325

log-likelihood

- hodnota věrohodnostní funkce (log-pravděpodobnost)

- Durbinova-Watsonova (DW) statistika *d*
- užívá se pro testování vlastností náhodných složek
- test autokorelace prvního řádu

Klasický lineární regresní model - verifikace

Verifikace statistická

sigma	0.120682	RSS	0.0728202076
R^2	0.997094	F(2,5) =	857.8 [0.000]**
log-likelihood	7.44531	DW	1.95
no. of observations	8	no. of parameters	3
mean(y)	10.5	var(y)	3.1325

no. of observations

- počet pozorování

no. of parameters

- počet regresních parametrů – tj. ***k*+1** (vč. konstanty)
  (počet vysvětlujících proměnných = ***k***)

mean (y)

- průměr vysvětlované (tj. endogenní) proměnné

var (y)

- rozptyl vysvětlované (tj. endogenní) proměnné

Klasický lineární regresní model - verifikace

Verifikace ekonometrická

- ověřuje splnění podmínek pro použití MNČ
- testuje se:

- Základní ekonometrická verifikace:

  HETEROSKEDASTICITA (náhodné složky)

  AUTOKORELACE (náhodné složky)

  MULTIKOLINEARITA (vysvětlujících proměnných)

- Další možnosti ekonometrické verifikace:

  Normalita reziduí (Jarque-Berra test)

  Test chybné specifikace modelu (Ramseyho RESET test)

  atd. (množství testů, obecných i pro vybrané typy dat či odhadů)

Klasický lineární regresní model - aplikace

Aplikace

- predikce apod., ukládání vyrovnaných hodnot, reziduí...
- predikce – dosazení konkrétních hodnot do regresní funkce

Klasický lineární regresní model - kvantifikace

Kvantifikace EM

- intervalový odhad

$$P\left\{b_i - s_{b_i} t^*_{1-\alpha/2(n-k-1)} \le \beta_i \le b_i + s_{b_i} t^*_{1-\alpha/2(n-k-1)}\right\} = 1-\alpha$$

$$t^*_{1-\alpha/2(n-k-1)} = t^*_{1-0,05/2(8-2-1)} = 2,571$$

$$s_{b_1} = 0,0046$$

$$b_1 \pm s_{b_1} t^*_{1-\alpha/2(8-2-1)} = 0,104 \pm 0,0046 * 2,571$$

$$\beta_1 \in \langle 0,0922\,;0,1158 \rangle$$

n = počet pozorování

k = počet vysvětlujících proměnných

α = hladina významnosti, např. 5 % (oboustranný interval)