

Aplikovaná statistika 2.

Veronika Svobodová

email: schneck@email.cz

Organizace výuky

- **11. 3. 2007** 9.20 – 11.50 **tutoriál 1**
- **22. 4. 2007** 9.20 – 11.50 **tutoriál 2**
- **6. 5. 2007** odevzdat POT1 (str. 42)
- **20. 5. 2007** 12.15 – 17.05 **tutoriál 3**
- **3. 6. 2007** odevzdat POT2 (str. 76)

Organizace zkoušky

- vypracování a zaslání POT na adresu:
schneck@email.cz
 - POT 1 do dvou týdnů po 2. tutoriálu
 - POT 2 do dvou týdnů po 3. tutoriálu
- test s použitím programu Excel
 - zadání na webových stránkách
 - příklady (nutné komentáře výsledků, správné spočtení Excelem bez interpretace je nedostačující)
 - teoretické otázky

Literatura

- Špalek, J.: *Aplikovaná statistika II* (Distanční studijní opora), MU Brno 2004
- Seger, J., Hindls, R., Hronová, S.:
Statistika v hospodářství, Praha ETC Publishing, 1998

Obsah předmětu

- Regresní a korelační počet
 - tvorba regresní funkce (1. tutoriál)
 - kvalita regresní funkce (2. tutoriál)
- Časové řady (3. tutoriál)
- Souhrnné cenové indexy (3. tutoriál)

Kvalita regresní funkce

- chceme-li posoudit kvalitu regresní funkce, klademe si vlastně otázku: Jak významné jsou parametry vypočítané regresní funkce?
- k tomuto účelu slouží především popisné statistiky (index determinance, korelační koeficient) a statistické testy (na posouzení vybraného modelu, na posouzení spočítaných koeficientů ve zvoleném modelu), případně konstrukce intervalů spolehlivosti

Kvalita regresní funkce

- abychom mohli definovat index determinance, potřebujeme zavést následující rozptyly:

- rozptyl empirických (naměřených) hodnot okolo empirického průměru

$$s^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n}$$

- rozptyl teoretických (vyrovnaných) hodnot okolo empirického průměru

$$s_t^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{n}$$

- reziduální rozptyl – rozptyl empirických hodnot okolo regresní funkce

$$s_r^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}$$

Kvalita regresní funkce

- vztah mezi uvedenými rozptyly: $s^2 = s_t^2 + s_r^2$
 - případ dokonalé závislosti: naměřené hodnoty leží na zvolené regresní přímce, reziduální rozptyl je roven 0
 - případ naprosté nezávislosti: teoretické hodnoty jsou stejné, teoretický rozptyl je roven 0
 - je přirozené zavést popisnou statistiku $R^2 = \frac{s_t^2}{s^2}$
index determinance
 - odmocnina z indexu determinance se pro lineární regresi nazývá korelační koeficient

Kvalita regresní funkce

- statistický test na výběr regresního modelu (hodnotí typ proložené funkce)

- nulová hypotéza: všechny koeficienty jsou nulové

- testové kritérium F-rozdělení ($p-1$ a $n-p$ stupňů volnosti, n je počet pozorování, p počet parametrů):

$$F = \frac{\frac{s_t^2}{p-1}}{\frac{s_r^2}{n-p}}$$

- kritický obor: $F > F_{1-\alpha}$

- padne-li F do kritického oboru, je vhodné posoudit vhodnost jednotlivých koeficientů

Kvalita regresní funkce

- hodnocení statistické významnosti jednotlivých parametrů zvoleného modelu pomocí intervalu spolehlivosti
 - hledáme interval, ve kterém se nachází hodnota závisle proměnné s danou pravděpodobností
 - v případě, že interval obsahuje nulu (=parametr může nabývat nulové hodnoty), znamená to, že jsme neprokázali, že je nutno jej zahrnout do modelu
 - v případě, že interval neobsahuje nulu, je nutné jej zahrnout do modelu

Kvalita regresní funkce

- interval bude tohoto tvaru (předpokládáme, že regresní parametry mají Studentovo t -rozdělení s $n-p$ stupni volnosti, kde n je počet pozorování, p je počet parametrů):

$$IS_{b_0} = \left(b_0 - s_{b_0} * t_{1-\frac{\alpha}{2}}; b_0 + s_{b_0} * t_{1-\frac{\alpha}{2}} \right)$$

$$IS_{b_1} = \left(b_1 - s_{b_1} * t_{1-\frac{\alpha}{2}}; b_1 + s_{b_1} * t_{1-\frac{\alpha}{2}} \right)$$

Kvalita regresní funkce

- statistický test na nenulovost regresních parametrů
 - nulová hypotéza: regresní koeficient je nulový
 - testové kritérium (t-rozdělení):
$$T = \frac{|b_i|}{s_{b_i}}$$
 - kritický obor:
$$T > t_{1-\frac{\alpha}{2}}$$
 - padne-li testové kritérium do kritického oboru, zamítáme hypotézu o nulovosti koeficientu

Kvalita regresní funkce

- příklad 2.1
- příklad (Seger): V tabulce jsou uvedeny údaje o hodnotě produkce v 100 000 Kč (proměnná y) a o výši investic v 10 000 Kč (x) v roce 1998 v souboru 12 vybraných soukromých strojírenských firem s počtem zaměstnanců větším než 24. Stanovte rovnici regresní přímky modelující závislost y na x . Vypočítejte kritéria její kvality a výsledek zhodnoťte.

Kvalita regresní funkce

- POT1
 - proved'te všemi (třemi) možnostmi, které byly prezentovány v prostředí programu Excel
 - výsledky opatřete vysvětlujícími komentáři
 - v roce 2001 nemá být zadána průměrná hrubá mzda!!!!
 - odhad počítejte s přesnou hodnotou HDP v tabulce!!!!

Děkuji za pozornost.