

Cvičení 12

V následujících příkladech předpokládejte, že data jsou pořízena jako náhodné výběry z populací sledujících normální rozdělení. K tomuto předpokladu však nepřistupujte nekriticky! Pokud budete přesvědčeni, že zde musíte použít jiný test, učiňte to.

1. Pomocí metody nejmenších čtverců odvoďte odhady koeficientů lineárního regresního modelu mezi proměnnými x (nezávislá proměnná) a y (závislá proměnná)¹.
2. Zjistěte závislost počtu pracovních hodin za měsíc spojených s provozováním anesteziologické služby na velikosti spádové populace nemocnice (viz tabulka 1). Údaje byly získány ve 12 nemocnicích ve Spojených státech.

OT
Pracné

Tabulka 1: Nemocnice

| Nemocnice | Celkový počet pracovních hodin | Spádová populace (v tisících osob) |
|-----------|--------------------------------|------------------------------------|
| 1 | 304,37 | 25,5 |
| 2 | 2616,32 | 294,3 |
| 3 | 1139,12 | 83,7 |
| 4 | 285,43 | 30,7 |
| 5 | 1413,77 | 129,8 |
| 6 | 1555,68 | 180,8 |
| 7 | 383,78 | 43,4 |
| 8 | 2174,27 | 165,2 |
| 9 | 845,3 | 74,3 |
| 10 | 1125,28 | 60,8 |
| 11 | 3462,6 | 319,2 |
| 12 | 3682,33 | 376,2 |

- (a) Rozhodněte, která proměnná je závislá a která nezávislá, respektive, která proměnná je vysvětlována respektive vysvětlující.
- (b) Zobrazte korelační pole a i s ohledem na věcnou stránku věci rozhodněte, který regresní model nejlépe vystihuje vzájemnou vazbu mezi sledovanými ukazateli. Modelů můžete uvažovat několik.
- (c) Vypočítejte parametry regresního modelu(ů).
- (d) Prostřednictvím testu zjistěte zda je(jsou) modely celkově statisticky významné. Dále pracujte jen s významnými modely.
- (e) Prostřednictvím testu zjistěte zda lze regresní model(y) zjednodušit. Pokud ano, zjednodušte.

¹Pro snazší kontrolu výsledků označte absolutní člen β_0 a regresní koeficient i lineárního členu β_1

- (f) Pokud jste mimo jiné uvažovali lineární regresní model, interpretujte hodnotu regresního koeficientu u tohoto modelu.
 - (g) Vypočítejte a interpretujte hodnoty koeficientů respektive indexů korelace a determinace. Na základě těchto hodnot vyberte „nejlepší“ model a dále pracujte jen s ním.
 - (h) Odhadněte v závislosti na spádové populaci celkový počet pracovních hodin pro 100 000 lidí velkou spádovou oblast.
 - (i) Určete 95% interval spolehlivosti celkového počtu pracovních hodin pro 100 000 lidí velkou spádovou oblast.
 - (j) Odhadněte v závislosti na spádové populaci **průměrný** celkový počet pracovních hodin pro 100 000 lidí velkou spádovou oblast.
 - (k) Určete 95% interval spolehlivosti pro odhad **průměrného** celkového počtu pracovních hodin pro 100 000 lidí velkou spádovou oblast.
 - (l) Graficky znázorněte empirické a teoretické (tj. odhadnuté na základě zvoleného modelu) hodnoty závisle proměnné. Vypočítejte rezidua.
3. Byly sledovány výdaje (v tisících Kč za 3 měsíce) za potraviny v jednotlivých domácnostech v závislosti jednak na počtu členů domácnosti a na celkovém čistém příjmu domácnosti (v tisících Kč za 3 měsíce). Údaje jsou v následující tabulce (viz tabulka 2).

Tabulka 2: Závislost výdajů v jednotlivých domácnostech

| | | | | | | | |
|--|----|----|----|----|----|----|----|
| Početnost domácnosti | 4 | 3 | 5 | 3 | 6 | 2 | 5 |
| Čistý příjem (Kč/3 měsíce) | 60 | 42 | 60 | 75 | 63 | 36 | 66 |
| Výdaje za potraviny (Kč/3 měsíce) | 21 | 5 | 24 | 30 | 27 | 12 | 27 |

- (a) Rozhodněte, která proměnná je závislá a která nezávislá, respektive, která proměnná je vysvětlovaná respektive vysvětlující.
- (b) Zobrazte korelační pole a i s ohledem na věcnou stránku věci rozhodněte, který regresní model nejlépe vystihuje vzájemnou vazbu mezi sledovanými ukazateli. Modelů můžete uvažovat několik.
- (c) Vypočítejte parametry regresního modelu(ů).
- (d) Prostřednictvím testu zjistěte zda je(jsou) modely celkově statisticky významné. Dále pracujte jen s významnými modely.
- (e) Prostřednictvím testu zjistěte zda lze regresní model(y) zjednodušit. Pokud ano, zjednodušte.
- (f) Pokud jste mimo jiné uvažovali lineární regresní model, interpretujte hodnotu regresního koeficientu u tohoto modelu.

- (g) Vypočítejte a interpretujte hodnoty koeficientů respektive indexů korelace a determinace. Na základě těchto hodnot vyberte „nejlepší“ model a dále pracujte jen s ním.
- (h) Odhadněte bodově i intervalově (95% interval spolehlivosti) v závislosti na čistém příjmu domácnosti výdaje za potraviny pro příjmy Kč 45 000,- a 65 000,-.
- (i) Odhadněte bodově i intervalově (99% interval spolehlivosti) v závislosti na čistém příjmu domácnosti **průměrné** výdaje za potraviny pro příjmy Kč 45 000,- a 65 000,-.
- (j) Graficky znázorněte empirické a teoretické (tj. odhadnuté na základě zvoleného modelu) hodnoty závisle proměnné. Vypočítejte rezidua.
4. U 44 náhodně vybraných studentek a studentů byla zaznamenáno pohlaví, velikost nohy (UK) a tělesná výška (cm), viz tabulka 3.

Tabulka 3: Velikosti nohou a výšky studentek a studentů – OP 2007/2008 ZS

| Pohlaví | Číslo nohy (UK) | Výška (cm) | Pohlaví | Číslo nohy (UK) | Výška (cm) |
|---------|-----------------|------------|---------|-----------------|------------|
| žena | 38 | 163 | žena | 38 | 165 |
| žena | 38 | 168 | žena | 37 | 167 |
| žena | 38 | 164 | žena | 40 | 165 |
| žena | 37 | 162 | muž | 45 | 191 |
| žena | 38 | 168 | žena | 41 | 171 |
| žena | 41 | 165 | žena | 38 | 168 |
| žena | 39 | 173 | žena | 40 | 165 |
| žena | 38 | 172 | žena | 39 | 170 |
| žena | 38 | 168 | muž | 48 | 197 |
| žena | 38 | 165 | žena | 39 | 179,5 |
| žena | 39 | 170 | žena | 39 | 175 |
| žena | 39 | 165 | muž | 43 | 186 |
| žena | 41 | 170 | muž | 42 | 180 |
| žena | 40 | 168 | muž | 46 | 192 |
| žena | 41 | 168 | žena | 40 | 173 |
| žena | 38 | 170 | žena | 39 | 170 |
| žena | 40 | 173 | žena | 40 | 174 |
| žena | 40 | 172 | žena | 40 | 175 |
| žena | 39 | 162 | žena | 38 | 168 |
| žena | 38 | 159 | žena | 38 | 174 |
| žena | 39 | 173 | žena | 38 | 170 |
| žena | 40 | 172 | žena | 38 | 167 |

- (a) Rozhodněte, která proměnná je závislá a která nezávislá, respektive, která proměnná je vysvětlovaná respektive vysvětlující.

- (b) Zobrazte korelační pole a i s ohledem na věcnou stránku věci rozhodněte, který regresní model nejlépe vystihuje vzájemnou vazbu mezi sledovanými ukazateli. Modelů můžete uvažovat několik.
 - (c) Vypočítejte parametry regresního modelu(ů).
 - (d) Prostřednictvím testu zjistěte zda je(jsou) modely celkově statisticky významné. Dále pracujte jen s významnými modely.
 - (e) Prostřednictvím testu zjistěte zda lze regresní model(y) zjednodušit. Pokud ano, zjednodušte.
 - (f) Pokud jste mimo jiné uvažovali lineární regresní model, interpretujte hodnotu regresního koeficientu u tohoto modelu.
 - (g) Vypočítejte a interpretujte hodnoty koeficientů respektive indexů korelace a determinace. Na základě těchto hodnot vyberte „nejlepší“ model a dále pracujte jen s ním.
 - (h) Odhadněte v závislosti na velikosti nohy tělesnou výšku postavy s nohou číslo 38 a 46.
 - (i) Určete 95% interval spolehlivosti pro tento odhad.
 - (j) Odhadněte v závislosti na velikosti nohy **průměrnou** tělesnou výšku postavy s nohou číslo 39.
 - (k) Určete 95% interval spolehlivosti pro odhad **průměrné** tělesnou výšku postavy s nohou číslo 38 a 46.
 - (l) Graficky znázorněte empirické a teoretické (tj. odhadnuté na základě zvoleného modelu) hodnoty závisle proměnné. Vypočítejte rezidua.
 - (m) S ohledem na vybočující hodnoty, které ovlivňují výsledky celé analýzy (hodnoty pro muže), vyzkoušejte rozdělit soubor na muže a ženy a odpovězte znovu na otázky 4b–4l. O výsledcích přemýšlejte!
5. U patnácti náhodně vybraných jatečných prasat byla při kontrole jatečné užitkovosti zjišťována délka jatečného trupu a výška špeku. Výsledky jsou uvedeny v tabulce 4).
- (a) Rozhodněte, která proměnná je závislá a která nezávislá, respektive, která proměnná je vysvětlovaná respektive vysvětlující.
 - (b) Zobrazte korelační pole a i s ohledem na věcnou stránku věci rozhodněte, který regresní model nejlépe vystihuje vzájemnou vazbu mezi sledovanými ukazateli. Modelů můžete uvažovat několik.
 - (c) Vypočítejte parametry regresního modelu(ů).
 - (d) Prostřednictvím testu zjistěte zda je(jsou) modely celkově statisticky významné. Dále pracujte jen s významnými modely.
 - (e) Prostřednictvím testu zjistěte zda lze regresní model(y) zjednodušit. Pokud ano, zjednodušte.

Tabulka 4: Výsledky kontroly jateční užitkovosti prasat

| Zvíře | Délka jatečného trupu (cm) | Výška špeku (cm) |
|-------|----------------------------|------------------|
| 1 | 92,7 | 3,80 |
| 2 | 93,2 | 3,90 |
| 3 | 94,0 | 3,70 |
| 4 | 95,0 | 3,65 |
| 5 | 93,5 | 3,80 |
| 6 | 96,5 | 3,65 |
| 7 | 95,5 | 3,70 |
| 8 | 92,7 | 3,70 |
| 9 | 93,5 | 3,60 |
| 10 | 95,5 | 3,65 |
| 11 | 95,2 | 3,60 |
| 12 | 94,7 | 3,70 |
| 13 | 96,5 | 3,55 |
| 14 | 97,2 | 3,50 |
| 15 | 98,0 | 3,48 |

- (f) Pokud jste mimo jiné uvažovali lineární regresní model, interpretujte hodnotu regresního koeficientu u tohoto modelu.
- (g) Vypočítejte a interpretujte hodnoty koeficientů respektive indexů korelace a determinace. Na základě těchto hodnot vyberte „nejlepší“ model a dále pracujte jen s ním.
- (h) Určete jakou výšku špeku lze v průměru očekávat u prasete, jehož délka trupu je 96 cm.
- (i) Určete 95% interval spolehlivosti pro tyto odhady.
- (j) Graficky znázorněte empirické a teoretické (tj. odhadnuté na základě zvoleného modelu) hodnoty závisle proměnné. Vypočítejte rezidua.