Aalto-universitetet

MS-A0509 Grundkurs i sannolikhetskalkyl och statistik Gripenberg/Metsalo
Övning 6
Vecka 7, 10-14.2.2014

Returnera lösningarna till I-uppgifterna senast 14.2.2014 kl. 18.00 OBS!!
Kom ihåg att skriva ditt namn och studentnummer!

I1. De observerade värdena av slumpvariablern X och Y är:

x	1	3	4	6	8	9	11	14
y	1	2	4	4	5	7	8	9

(a) Bestäm estimat b_0 och b_1 för koefficienterna i regressionsmodellen $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ med minsta kvadratmetoden.
(b) Testa nollhypotesen $\mathrm{H}_0 : \ \beta_1 = 0$ på signifikansnivån 5%

I2. Beträffande ett observerat stickprov (x_j, y_j), $j = 1, \dots, n$ har vi följande uppgifter: $n = 6$, $\overline{x} = 54.1667$, $\overline{y} = 143.8333$, $s_x^2 = 194.1667$, $s_y^2 = 285.7667$, $s_{xy} = 228.8333$, $r_{xy} = 0.9715$ och $s^2 = 20.0966$,

(a) Räkna ut estimaten b_0 och b_1 för koefficienterna β_0 och β_1 i regressionsmodellen $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ och estimatet r_{xy} för korrelationskoefficienten.
(b) Testa nollhypotesen $\mathrm{H}_0 : \beta_0 = 0$ på 1% signifikansnivå.
(c) Testa nollhypotesen $\mathrm{H}_0 : \ \rho_{xy} = 0$ på 1% signifikansnivå.

I3. Antag att punkterna (x_j, y_j), $j = 1, \dots, n$ är givna och att vi med minsta kvadratmetoden räknat ut b_0 och b_1 dvs.

$$b_1 = \frac{s_{xy}}{s_x^2} \quad \text{och} \quad b_0 = \overline{y} - b_1 \overline{x}.$$

Visa att eftersom

$$s^2 = \frac{1}{n-2} \sum_{i=1}^{n} \big(b_0 + b_1 x_j - y_j\big)^2,$$

så gäller

$$s^2 = \frac{n-1}{n-2}\, s_y^2\, (1 - r_{xy}^2),$$

och

$$\frac{b_1}{\sqrt{\frac{s^2}{(n-1)s_x^2}}} = \frac{r_{xy}\sqrt{n-2}}{\sqrt{1 - r_{xy}^2}}.$$

I4. Enligt statistikcentralen var antalet omkomna i trafikolyckor under åren 2000–2011 följande:

2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
396	433	415	379	375	379	336	380	344	279	272	292

Nu kan man använda dessa siffror för att med hjälp av linjär regression uppskatta hur många som kan befaras omkomma i trafikolyckor år 2014 men här gäller det att använda följande metod för att få en uppfattnig om med vilken noggrannhet man får svaret.

Använd tex. `matlab/octave` och låt `x` vara årtalen minus 2014 (dvs. -14, -13, osv.) och låt `y` vara antalen omkomna (och skriv dem som radvektorer). Det som man sedan gör är att slumpmässigt välja 12 av dessa tex. med kommandona

```
jj=floor(12*rand(1,12)+1); xx=x(jj); yy=y(jj);
```

Sedan räknar man ut koefficienterna i regressionslinjen med

```
M=[0*xx'+1, xx']; cc=(M'*M)^(-1)*M'*yy';
```

Det estimerade antalet omkomna 2014 blir då

```
f=cc(1);
```

eftersom modellen är $y \approx b_0 + b_1 x$ och vi valt 2014 som nollpunkt. Upprepa nu detta tex. 200 gånger så att du skriver alla dessa kommandon i en fil tex. `b.m` och före alla dessa räkningar skriver `for j=1:200`, sedan ändrar det sista kommandot till `f(j)=cc(1);` och till sist skriver `end`. När du sparat filen ger du kommandot `b`
(om du nu kallat den `b.m`).

För att rita ett histogram kan du ge kommandot `hist(f)` men ge som svar ett intervall så att 5% av talen ligger till vänster och 5% till höger om detta intervall. Detta kan du göra genom att först ordna vektorn med `f=sort(f);` och sedan räkna ut `f(round(0.05*n)+1)` och `f(round(0.95*n)-1)` om du räknat ut estimatet `n` gånger.

I5. Med hjälp av observationerna (x_j, y_j), $j = 1, \ldots, n$ har vi räknat it koefficienterna i regressionslinjen $y = b_0 + b_1 x$ i det fall att x är den ”förklarande” variabeln (dvs. då vi antar att $Y_j = \beta_0 + \beta_1 x_j + \varepsilon_j$) och också för den ”inversa” regressionslinjen $x = a_0 + a_1 y$ där y är den ”förklarande” variabeln (dvs. då vi antar att $X_j = \alpha_0 + \alpha_1 y_j + \varepsilon_j$). Dessutom har vi räknat ut korrelationskoefficienten r_{xy} och konstaterat att nollhypotesen $\mathrm{H}_0 : \beta_1 = 0$ kan i detta fall förkastas på sigifikansnivån 0.01. Vad kan man säga om följande påståenden (då vi antar att $s_x > 0$ och $s_y > 0$):

(a) Linjerna $y = b_0 + b_1 x$ och $x = a_0 + a_1 y$ sammanfaller om och endast om $|r_{xy}| = 1$.
(b) Också hypotesen $\mathrm{H}_0 : \alpha_1 = 0$ kan förkastas på signifikansnivån 0.01.

Motivera dina svar!

Besvara Stack-uppgifterna (`stack3.aalto.fi/course/view.php?id=18`)
senast 15.2.2014 kl. 22.00 OBS!!
