Tölfræði I (09.10.45)

Vikublað 1

Efni fyrstu og annarrar viku: 103.2, 104.1 og hluti 105.1 af kennsluvef. Aðeins er tekið fyrir líkinda- og tölfræðiefni af vefnum en ekki almennar upplýsingar aðrar.

Dæmi Skiladæmi eru merkt með stjörnu. Dæmum á að skila fyrir kl 12 nk mánudag í hólf kennara (eða sem pdf skjal í tölvupósti, á sama tíma).

1: Reiknið meðaltal og dreifni hendingar X, sem lýtur jafndreifingu á bilinu 0 til 1, $X \sim U(0,1)$.

2: Notið R til að herma 10 000 tölur úr jafndreifingu á bilinu 0 til 1 og reiknið meðaltal og staðalfrávik.

3: Látum $X_1, \ldots \sim U(0,1)$, *i.i.d.* og skilgreinum $Y_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Teiknið líkindadreifingu Y_n fyrir $n = 1, 2$.

Framkvæmið tilsvarandi hermun í R og teiknið stuðlarit dreifinga Y_n fyrir $n = 1, 2, 3, 5, 12$. Hverju líkist dreifing Y_n og hvers vegna?

4:* Gerið einfalda aðhvarfsgreiningu $y = \alpha + \beta x$ **í höndunum** , hvert með sínum gögnum. Gerið þetta með því að setja upp í höndunum töflu með x_i, y_i, $x_i - \bar{x}$, $y_i - \bar{y}$ ásamt tilheyrandi kvaðrötum og krossmargfeldum.

Gögn: Kennitalan, staf fyrir staf áfram eru x-gildin, en staf fyrir staf aftur á bak eru y-gildin.

Endurtakið dæmið í R með

```
fm<-lm(y~x)
summary(fm)
anova(fm)
```

Kvaðratsumma frávika frá línunni er yfirleitt kölluð SSE. Sannfærið ykkur um að "anova()" í R gefi þá niðurstöðu.

5: Lengd (l, cm) og lifandi þyngd (w, g) nokkurra þorska er að finna í töflunni hér fyrir neðan.

Lengd	Óslægt	Lengd	Óslægt	Lengd	Óslægt	Lengd	Óslægt	Lengd	Óslægt
85	6530	95	8000	71	3470	93	9035	99	10130
74	4045	105	10460	96	7225	63	2510	83	5060
73	3620	77	4475	84	6530	12	10	63	2905
71	3225	87	5355	81	5220	95	8600	100	8595
99	9815	95	9650	111	15855				

Sambandi lengdar og þyngdar er oft lýst með $ln(w) = \alpha + \beta ln(l)$. Látum því $x = ln(l)$ og $y = ln(w)$. Gefið er:

$n = 23$, $\sum x = 100.31$, $\sum y = 194.1$, $\sum x^2 = 441.67$, $\sum y^2 = 1682.49$, $\sum x \cdot y = 860.16$.

Metið líkingu bestu línu í gegnum (x, y) með hefðbundinni aðhvarfsgreiningu af y á x (á pappír).

6: Mælingar á aldri lamba í dögum (x) og lifandi þunga (y) við vigtun að hausti gáfu eftirfarandi tölur:

x	y
135	39
137	38
123	34
137	44
125	35
133	36
132	40
121	34
120	33
140	41
129	38
137	41
126	38
130	38
121	34
130	39
125	37
129	36
126	35
137	43

Gefið er: $n = 20$, $\sum x = 2593$, $\sum y = 753$, $\sum x^2 = 336909$, $\sum y^2 = 28533$, $\sum x \cdot y = 97939$.

(a) Metið líkingu bestu línu í gegnum (x, y) með hefðbundinni aðhvarfsgreiningu af y á x - á pappír, án notkunar reikniforrita nema einfaldrar reiknivélar.

(b) Setjið gögnin í í R, endurtakið alla útreikningana í töflunni (með mean()-fallinu). Athugið að gögnin má afrita af vefsíðu, líma inn í skrá og lesa síðan inn í R.

(c) Endurtakið aðhvarfsgreininguna með R (með lm()-fallinu).

7*: Notið R til að herma $n = 10$ normaldreifðar mælingar á línunni $y = \alpha + \beta x$ með $x_i = i$ og $i = 1, \ldots, 10$, með staðalfráviki $\sigma = 2$. Festið $\alpha = 2$ og $\beta = 3$ og finnið minnsta kvaðrata mað á báðum stikunum. Skilgreinið fall í R sem skilar hallastuðlinum.

Endurtakið ofangreint 200 sinnum, takið í hverri keyrslu mat á hallastuðli og gerið stuðlarit sem lýsir dreifingu hallastuðlamatsins.

Lítið á hallastuðlamötin sem niðurstöður 200 mælinga og reiknið meðaltal og staðalfrávik mælinganna.