

Palautus4

Timo Järvinen
592042

January 2019

1 Tehtava A: Tutustuminen R:n

Vektorissa s ($s = c("1", "2", "3", "4")$) olevat numerot ovat lainausmerkeissä, eli ne käyttäytyvät kuin merkit. Komento `as.numeric(s)` tekee näistä normaaleja numeroja, eli poistaa lainausmerkit numeroiden ympäriltä. Se ei kuitenkaan automaattisesti tallenna tätä tietoa vektoriin. Mikäli siis komennon jälkeen laitat konsoliin `"s"`, sieltä tulee sama alkuperäinen merkkivektori.

Matriisien `A(1,2;9,10)` ja `B(4,5;7,8)` matriisitulo saadaan lasketuksi kaavalla `A%%B`. Tulokseksi saadaan `(18,21;106,125)`.

Listan l_1 jälkimmäisen alilistan viimeisen alkion arvo saadaan arvoksi `FALSE` komennolla `l1[[2]][-1] = FALSE`

Vektori `[1, 3, 5, ..., 99]` saadaan luoduksi esimerkiksi komennolla `seq(from = 1, to = 99, by = 2)`. Toinen vaihtoehto olisi tietysti näpytellä numerot manuaalisesti (`c(1, 3, 5, ..., 99)`).

Lista palauttaa sen sisällä olevat vektorit erikseen. Se voi sisältää esimerkiksi vektoreita eri mittauksista. Data frame taas yhdistää nämä vektorit matriisiksi, jolloin sitä voidaan lukea kuin taulukkoa: esimerkiksi samalla rivillä on tulokset samasta mittauksesta, kun taas samalla sarakkeella on saman koehenkilön eri mittaukset.

Listan (tässä tapauksessa nimeltään `list`) ensimmäisen alilistan alkioihin voi viitata esimerkiksi komennolla `list[1]`. Data framessa (tässä esimerkissä nimeltään `data`) saman asian ajaa komento `data[,1]`.

Komento `head()` palauttaa data framen tai vektorin ensimmäiset arvot, `names()` taas antaa data framen sarakkeiden nimet, `summary()` nimensä mukaisesti kiteyttää datan tulokset palauttaen esimerkiksi minimin, maksimin ja keskiarvon. Komento `rep()` muuttaa matriisin vektoriksi, `seq()` luo vektorin sille annetuilla määreillä, esimerkiksi `seq(from = 1, to = 99, by = 2)` luo vektorin, joka alkaa numerosta yksi, päättyy numeroon 99 ja peräkkäisten numeroiden erotus on kaksi.

Matriisin rivien keskiarvojen laskeminen

```
1 dat <-  
2   matrix(rnorm(n = 15, mean = 100, sd = 30), nrow = 3, ncol = 5)  
3 #For-loop  
4 i = 1  
5 for(row in 1:nrow(dat)) {  
6   sum = 0  
7   amount = 0  
8   for(col in 1:ncol(dat)) {  
9     sum = sum + dat[row,col]  
10    print(dat[row,col])  
11    print(sum)  
12    amount = amount + 1  
13  }  
14  ka = sum/amount  
15  cat(sprintf("Rivin %s keskiarvo: %s\n", i, ka))  
16  i = i+1  
17 }  
18 #Apply-funktio  
19 row.sums = apply(dat, 1, sum)  
20 row.means = row.sums/5  
21 print(row.means)
```

2 Tehtävä B: Useamman selittäjän lineaarinen regressio

Uskon asuntojen mediaanihintoihin vaikuttavan erityisesti asumistiheyden sekä asunnoissa olevien huoneiden määrään.

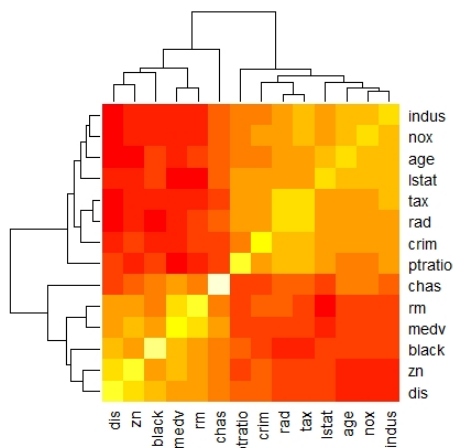


Figure 1: Korrelaatiokuva

Kuvaajasta nähdään, että mediaanihinnan kanssa parhaiten korreloi huoneiden määrä asunnossa, mutta sitten taas negatiivisesti vahvasti korreloi lstat ja oppilaiden määrän suhde opettajiin. Näitä käytin myös selittävinä muuttujina. Regressiokertoimet ovat: $lstat = -0.616$, $rm = 4.38$, $ptratio = -0.820$.

Mallin keskineliövirhe $MSE = 19.8206$. Ennustuskkyky on mallissa melko huono, siinä voisi olla esimerkiksi enemmän muuttujia tai suurempi datasetti.

3 Kotitehtävä

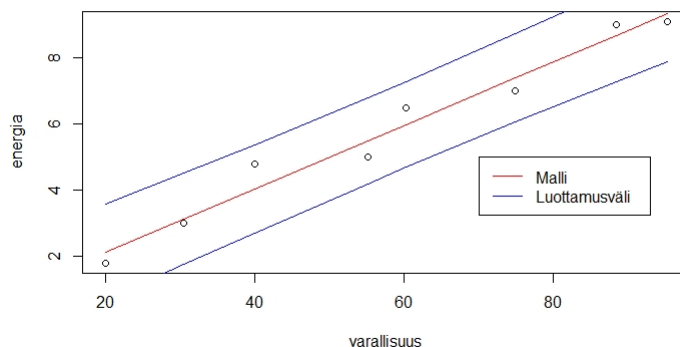


Figure 2: Lineaarinen energiakulutusmalli

```

1 varallisuus = c(20.0, 30.5, 40.0, 55.1, 60.3, 74.9, 88.4, 95.2)
2 energia = c(1.8, 3.0, 4.8, 5.0, 6.5, 7.0, 9.0, 9.1)
3
4 energiadata = data.frame(varallisuus, energia) #Tehdä data frame
5
6 raha = data.frame(varallisuus) #Varallisuudesta oma data frame
7
8 malli = lm(energia ~ varallisuus, data=energiadata) #Lineaarinen malli
9 ennustetut_pisteet = predict(malli, raha) #Lasketaan ennustetut
   pisteet
10 plot(energiadata) #Piiirret n mittauspisteet
11
12 lines(varallisuus, ennustetut_pisteet, col="red") #Piiirret n
   malli
13
14 luottamusvali = predict(malli, raha, interval='prediction', level
   =0.95) #Lasketaan luottamusväli
15

```

```

16 lines(varallisuus, luottamusvali[,2], col='blue')#piirretään
    luottamusv li
17 lines(varallisuus, luottamusvali[,3], col='blue')
18
19 #Listataan selitteet
20 legend(70,5,legend=c("Malli", "Luottamusv li"), col=c("red", "blue
    "), lty=1:1, cex=1.0)

```

Erona matlab-versioon on etenkin, että tässä versiossa pitää luoda erikseen vektoreista data frameja. Riveinä R:n koodi on silti vain noin 3/4 Matlab-versiostani, vaikka tämä tuntuikin vaikeamalta. Prediction-funktiot olivat melko samanlaisia.