



Tentissä saa käyttää kirjallista materiaalia oman vaihtaman mukaan.

1. Sairaan hallinto haluaa tutkia, miten poitetaan ikä x_1 , sairauden vakavuus x_2 ja tuska-
suusaste x_3 vaikuttavat poitteen käyttäisyyteen y . Potilasaineiston perustiedoilla saadaan
regressioanalyysillä seuraavat tulokset.

ANALYSIS OF THE MODEL
PARAMETER ESTIMATES

Variable	Estimate	Std	t	p(2-tail)
Constant	162.88	25.78	6.32	0.00
Variable 1:	-1.21	0.30	-4.01	0.00
Variable 2:	-0.67	0.82	-0.81	0.43
Variable 3:	-8.61	12.24	-0.70	0.49

R-squared = 0.67266 R(adj)-squared = 0.62097

ANOVA table

Source	df	SS	MS	F	P
Regression	3	4133.63	1377.88	13.014	0.000075
Residual	19	2011.58	105.87		
Total	22	6145.22			

Oletetaan, että regressioanalyysin oletukset ovat voimassa.

Tarvitaanko mallissa ollenkaan muuttujia vakiotermin lisäksi? Testaa, voidaanko muuttujia
 x_3 poistaa mallista.

Haluamme poistaa mallista samalla kertaa molemmat muuttujat x_2 ja x_3 . Selitä, miten
testaisit tätä kysymystä, sanoitään riskitasolla $\alpha = 0.05$. Mikä on tässä tapauksessa nolli-
hypoteesi?

2. Tarkastelemme regressioanalyysin perusmallia

$$y = X\beta + e,$$

missä $X \in \mathbb{R}^n \times p$, $p < n$, ja $\mathbb{E}(e) = 0$.

- Olkoon X^T jokin matriisin X vaakarivi. Osoita, että $X^T\beta$ on estimoitava.
- Oletetaan, että $\text{rank}(X) = p$. Osoita, että $X^T\beta$ on estimoitava kaikilla $\lambda \in \mathbb{R}^p$.
- Olkoon nyt $\text{rank}(X) < p$ ja oletetaan, että $X^T\beta$, $i = 1, \dots, k$, ovat estimoitavia.
Osoita, että $\sum_{i=1}^k d_i X_i^T \beta$ on estimoitava, missä $d_i \in \mathbb{R}$.

Käynnä

3. Olkoon $x \in \mathbb{R}^p$ satunnaismuuttuja kovarianssmatriisina Σ ja $Y_k = q_k^T x$ populaation k .
pääkomponentti. Lausu satunnaismuuttujien Y_k ja x_k , joka on x :n k . komponentti, välinen
kovarianssi $\text{Cov}(x_k, Y_k) = \Sigma$ n ominaisarvojen ja -vektoreiden avulla.

4. Tarkastelemme logistista regressioanalyysistä. Meillä on kaksi huokkaa R_0 , jonka havain-
not on koodattu 0:ksi ja R_1 , jonka havainnot on koodattu 1:ksi. Käytössämme on myös
mittausdata x_1, \dots, x_n , missä jokaisen mittauspisteen $x_i \in \mathbb{R}^p$ huokka tunnetaan. Logisti-
nen regressioanalyysi antaa kerronestimaatin β .

Saamme uuden havainnon x_n ja tehtävämme on huokkella se toiseen huokasta R_0 ja R_1 .
Osoita, että x_n huokkellaan huokkaan R_1 , jos $x_n^T \hat{\beta} > 0$.