

Lineaarsed mudelid

Kordamisküsimused

Eksami üldine korraldus

Ülesannete lahendamisel võib kasutada kõikmõeldavaid materjale (kaasa arvatud R-i kuid mitte kaastudengeid ega teisi statistikaeksperte), kuid tuleb silmas pidada, et kirjaliku eksami aeg on äärmiselt piiratud (2h). Vastused tuleb saata õppejõu e-mailile. Soovi korral võib ülesannete lahendused/vastused kirjutada paberile ja siis saata pilt eksamitööst õppejõule – valemite ja tõestuste puhul on paberile kirjutamine ja seejärel tehtud töö pildistamine lihtsalt märkimisväärselt kiirem muudest variantidest. Soovi korral võib muidugi osade ülesannete vastused otse e-maili kirjutada ja osade ülesannete lahendused foto kujul edastada.

Küsimuste tekkimisel võin lahenduskäikude kohta küsida hilisema videovestluse käigus.

Teoreetiliste küsimuste teemad

Küsida võib kõike, mida kursuse käigus tutvustatud on. Siiski on mõned teemad õppejõule eriti hingelähedased, sestap võib alltoodud teemadel teoreetilisi küsimusi rohkem (või tõenäolisemalt) esineda:

1. Lineaarse mudeli parameetrite hindamine vähimruutude meetodil ja suurima tõepära meetodil.
2. Maatriksi P_X omaduste tõestamine: $P_X X = X$; P_X ei sõltu üldistatud pöördmaatriksi valikust; P_X on sümmeetriline.
3. Parim lineaarne nihketa hinnang; Gauss-Markovi teoreem.
4. Hinnatava parameeterfunktsiooni hinnangu jaotus, t-test hinnatava parameeterfunktsiooni kohta hüpoteeside kontrollimiseks, usaldusintervall parameetrite hinnatavale lineaarsele funktsioonile.
5. Mudelite võrdlemine F-testi abil. Põhimõtte selgitus, tõestused.
6. Prognoosiintervall uuele vaatlusele.
7. Determinatsioonikordaja ja kohandatud determinatsioonikordaja. Mida näitavad, kuidas arvutada, miks rikkama mudeli determinatsioonikordaja on (samade vaatluste korral) alati suurem (või samasuur) kui vaesema mudeli determinatsioonikordaja; kuidas leida kohandatud determinatsioonikordajate tavalise determinatsioonikordaja kaudu.
8. Mallows'i C_p -kriteerium
9. Scheffé meetod mitmese võrdlemise teostamiseks.
10. Tukey (ja Tukey-Kramer'i) mitmese võrdlemise meetod

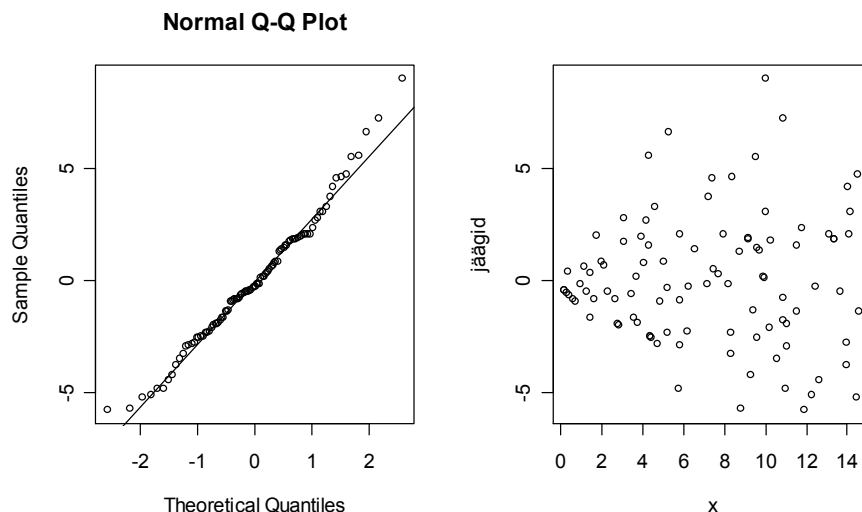
Kui mõni konkreetne teema pole esindatud ülaltoodud loetelus, ei tähenda see seda, et te ei peaks antud loengu materjalidega tutvuma. Lihtsalt ülaltoodud teemade puhul peaksite esitatud tõestustest piisavalt hästi aru saama, et suudaksite neid tõestuseid modifitseerida, ära tundma tõestustest esinenud suuruseid või suutma kasutada tõestuse põhiideed uudes olukorras.

Näiteid ülesannete kohta

Ülesanne 1

Vaata alljärgnevaid jooniseid. Iga joonise korral otsusta, kas lineaarse mudeli eeldustega on kõik korras ning vajadusel maini, kuidas võiks probleemi (kui see esineb) lahendada - maini vaid ühte, sinu arvates kõige kindlamini töötavat „parandust“.

Mudel 1



Lineaarse mudeli eeldustega kõik korras?

Mida teeksid olukorra parandamiseks:.....

Ülesanne 2

Hinnati dispersioonanalüüsi mudel, kus sees kaks faktortunnust (ja nende koosmõju). Faktortunnusel sugu on kaks taset (F- naine; M- mees), faktortunnusel Treatment on samuti kaks taset: + (said ravi) ja – (kontrollgrupp). Analüüsi tulemused on toodud all:

```
> m1=lm(spb~Treatment+Gender+Treatment:Gender)
> summary(m1)
```

Residuals:

Min	1Q	Median	3Q	Max
-36.9906	-6.4563	0.2498	7.0918	28.6907

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	180.0365	0.7008	256.905	< 2e-16 ***
Treatment+	-0.4365	0.9642	-0.453	0.651
Gender2F	5.6875	0.9141	6.222	7.21e-10 ***
Treatment+:Gender2F	-9.5548	1.2925	-7.393	3.05e-13 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.13 on 996 degrees of freedom

Multiple R-Squared: 0.1224, Adjusted R-squared: 0.1197

F-statistic: 46.29 on 3 and 996 DF, p-value: < 2.2e-16

1. Kas ravimil on mõju?
2. Milline on ravi saanud ja ravi mittesaanud (kontrollgrupp) naiste vaheline erinevus?

Ülesanne 3

Sooviti uurida inimeste sissetulekuid. Üle eestimaa küsitleti 1500 inimest, igast maakonnast 100. Uuringu tegija leidis kaks sind huvitada võivat numbrit: eestimaalaste palkade dispersiooniks (nihketa hinnang) saadi 10000 ja maakondade kaupa arvatud dispersioonide keskmine oli 8000. Soovitakse kontrollida kas inimeste sissetulekud on maakonniti erinevad. Leia asjasepuutuva F-statistiku väärtus ja vabadusastmete arvud. Kas olulisuse nivool 0,05 saab nullhüpoteesi (kõigis maakondades on keskmine palk sama) kummutada? F-jaotuse kriitilisi väärtuseid saad R-is leida käsu $qf(0.95, df1, df2)$ abil

Ülesanne 4

Antud on järgmine andmestik:

sort	saak (Y)
1	10
1	12
2	14
2	18
3	20
3	24

Kaalutakse mudelit

$$Y_{ij} = \mu_i + \varepsilon_{ij}$$

Kus Y_{ij} on saak j . põllul kus kasvatati sorti i .

Leia parameetervektori hinnang järgides klassikalist β hindamiseks kasutatavat valemit (näita arvutusi), testi kas sort 1 keskmine saagikus võib olla kaks korda väiksem kui sort 2 keskmine saagikus (näita arvutusi ja vahetulemusi).

Ülesanne 5

Täida lüngad!

```
> summary(mudel)
```

```
[...]
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-1194.395	89.815	-13.30	8.29e-15 ***
Pikkus	47.369	2.328		< 2e-16 ***

Residual standard error: on 33 degrees of freedom

Multiple R-Squared: 0.9261, Adjusted R-squared:

F-statistic: 413.8 on and 33 DF, p-value: < 2.2e-16

Ülesanne 6

Millise jaotusega on suurus

$$\frac{\left(\lambda^T \hat{\beta} - \lambda^T \beta_0\right)^T \left(\lambda^T \hat{\beta} - \lambda^T \beta_0\right)}{\lambda^T (\mathbf{X}^T \mathbf{X})^{-1} \lambda \cdot MSE}$$

ja millepärast on ta sinu poolt arvatud jaotusega? Esita põhjendus (tõesta oma väidet!)

Ülesanne 7

Mida võidakse testida järgmise statistiku abil:

$$t = \frac{Y_1 - x_1^T (X^T X)^{-1} X^T Y}{\sqrt{Y^T (I - P_X) Y (1 + x_1^T (X^T X)^{-1} x_1) / (n - \text{rank}(X))}}?$$

Selgita, miks sobib ülaltoodud teststatistik sinu poolt pakutud hüpoteesi kontrollimiseks!

Ülesanne 8

Kasutatakse mudelit

$$Y_{ij} = \mu_i + \varepsilon_{ij},$$

kus $i=1..k$ ja $j=1..n_i$. Kui $n_3=4$, siis milline on vaatluse Y_{32} mõjukus?