LINEAARISET MALLIT (806308A)

Loppukoe 20.6.2011

Leena Pasanen

Vastaa kaikkiin kysymyksiin!

1. Tarkastellaan "perusmuotoisen" lineaarisen regressiomallin

(A*)
$$Y_i = \beta_1 x_{i1} + \beta_m x_{im} + \varepsilon_i = \beta' X_i + \varepsilon_i$$
$$\varepsilon_i \perp\!\!\!\perp X_i \ , \ \varepsilon_i \sim NID(0, \sigma^2) \ , \ i = 1, \ldots, n$$

matriisimuotoista versiota

(A)
$$Y = X\beta + \varepsilon \ , \ \varepsilon \sim N_n(0, \sigma^2 I) \ ,$$
$$Y = (Y_1 \cdots Y_n)' \ , \ X = \begin{pmatrix} x_{11} & \ldots & x_{1m} \\ \cdot & \ldots & \cdot \\ \cdot & \ldots & \cdot \\ x_{n1} & \ldots & x_{nm} \end{pmatrix}$$

regressiokerroinvektorin $\beta = (\beta_1 \cdots \beta_m)'$ pienimmän neliösumman estimaattoria (johon päädytään sekä momenttiperiaatetta että ML-periaatetta käytettäessä) $\widehat{\beta} = (X'X)^{-1}X'Y$ sekä vastaavaa sovitevektoria $\widehat{Y} = X\widehat{\beta}$ ja jäännöstermivektoria $E = Y - \widehat{Y}$. (Huom: Tavallisesti $x_{i1} \equiv 1$.)

- Jos vastemuuttujasta saaduista havainnoista koostuvaa vektoria $Y \in \mathcal{R}^n$ ja selittävistä muuttujista saaduista havainnoista koostuvia vektoreita $X_j^* = (x_{1j} \cdots x_{nj})' \in \mathcal{R}^n$ $(j = 1, 2, \cdots, m)$ ajatellaan "*havaintovektoreina*" (ja $\mathcal{R}^n$:n alkioina), niin millaisia tulkintoja sovitevektorille $\widehat{Y}$ voidaan antaa
  - $\widehat{Y}$:n ja Y:n välisen *etäisyyden* $\| Y - \widehat{Y} \|$ kannalta?
  - $\widehat{Y}$:n ja Y:n välisen *kulman* $\phi(Y, \widehat{Y})$ kannalta?
  - Y:n lineaarikuvauksena (projektiona)?

  Miten jäännöstermivektori E voidaan tulkita Y:n lineaarikuvauksena (projektiona)?
- Millainen on estimaattorin $\widehat{\beta}$ otantajakauma? Entä sen yksittäisten komponenttien $\widehat{\beta}_j$ $(j = 1, \cdots, m)$ otantajakaumat?

Elävöitä vastaustasi liitteessä 1 esitetyillä laskelmilla, jotka liittyvät 104:n Kustavin kunnassa (kunta Turun ulkosaaristossa) kasvaneen männyn rinnankorkeudelta (1.4 metrin korkeudelta juurenniskasta lukien) ja kuuden metrin korkeudelta mitattujen läpimittojen (d_i ja $d6_i$, senttimetreinä) ja puiden tilavuuksien (Y_i, litroina) välisen yhteyden tutkimiseen. (Nämä läpimittatiedot on helppo mitata ennen puun kaatamista.

Edellinen yksinkertaisesti mittanauhan, jälkimmäinen pitkän puukepin päähän asetetun kaulaimen avulla.)

Voidaan todeta, että läpimittatietojen d_i ja $d6_i$ mukaisen tasaisesti kapenevan pyörähdyskartion tilavuus olisi (litroina)

$$x_i = \frac{1}{3} \cdot \pi \cdot \left(0.5 \cdot d_i \cdot \frac{\text{htot}_i}{\text{hrinta}_i}\right)^2 \cdot \text{htot}_i \cdot \frac{1}{1000},$$

jossa

$$\text{hrinta}_i = d_i \cdot \frac{600 - 140}{d_i - d6_i} \quad \text{ja} \quad \text{htot}_i = 140 + \text{hrinta}_i$$

Tästä syystä liitteesä 1 on kokeiltu mallia

$$Y_i = \beta_1 + \beta_2 x_i + \epsilon_i \quad , \quad \epsilon_i \sim NID(0, \sigma^2) \quad , \quad i = 1, \ldots, 104.$$

Kommentoi estimointituloksia erityisesti parametrin β_2 (ja sen estimaattorin hajontaarvion) osalta. (Etukäteen voisi nimittäin olettaa, että β_2:n ei pitäisi olla "kovin" kaukana ykkösestä.)

Näyttääkö mallin vakioparametri β_1 estimoituneen järkevästi?

Kuinka suureksi muodostui jäännösneliösuma $E'E$?

2. Tarkastellaan jatkuvien satunnaismuuttujien Y ja $X = (X_1, \ldots, X_m)'$ yhteiskäyttäytymistä (ts. yhteisjakaumaa)

   - Mitä tarkoitetaan Y:n regressiofunktiolla X:n suhteen?
   - Millaisissa sovellustilanteissa on järkevää ja hyödyllistä keskittyä tutkimaan (ja estimoimaan) Y:n regressiofunktiota X:n suhteen?
   - Millaisten tilanteiden kuvaamiseen regressiomallit eivät lainkaan sovellu? Anna esimerkki tällaisesta tilanteesta.

Elävöitä vastaustasi edellisessä tehtävässä esitellyn ongelman avulla. Toisen vaihtoehdon kuvata puun tilavuuden Y_i riippuvuutta läpimittatiedoista d_i ja $d6_i$ tarjoaisi esimerkiksi malli

(B)
$$Y = \beta_1 \cdot (d_i)^{\beta_2} \cdot (d6_i)^{\beta_3} \cdot \kappa_i, \;,$$
$$\log \kappa_i \sim NID(0, \sigma^2 I), \; i = 1, \ldots, 104$$

Onko regressiomallin käyttö ylipäätään järkevää tätä yhteyttä kuvatessa? (Perustele mielipiteesi.)

Miten malli (B) voitaisiin tulkita lineaariseksi regressiomalliksi?

Liitteessä 2 on esitetty parametrien β_2 ja β_3 estimoimiseksi laaditut SAS- ja R-ohjelmat sekä ohjelmien tuottamat tulostukset.

Miten näiden estimointitulosten perusteella voidaan laskea alkuperäisten tilavuuksien odotusarvosovitteet $\widehat{Y}_i = E(Y_i | d_i, d6_i)$? Liitteessä 2 on tulostettu mallin (B) *linearisoidun*

version antamat sovitteet ja transformoitu ne sitten takaisin alkuperäiseen litraskaalaan. Lisäksi on laskettu "jäännösneliösumma" $\sum_{i=1}^{104}(Y_i - \widehat{Y}_i)^2$. Vertaa mallin (B) toimintatarkkuutta liitteessä 1 saatuihin tuloksiin. Kumpi malli näyttää paremmin hyödyntävän läpimittatietoihin sisältyvän, puun tilavuutta koskevan informaation?

3. Tarkastellaan tavallista lineaarista regressiomallia (A*). Miten havaintoaineiston Y_i, $X_i = (1\, x_{i2} \ldots x_{im})'$ $(i = 1, \ldots, n)$ perusteella voidaan konstruoida $(1-\alpha)$-tasoinen (α on pieni positiiviluku esim. $\alpha = 0.05$) luottamusväli

   (a) regressiokertoimelle β_j

   (b) regressiokertoimien lineaarikombinaatiolle $C'\beta = c_1\beta_1 + \cdots + c_m\beta_m$?

   (c) uuden havainnon odotusarvolle $EY_{n+1} = \beta' X_{n+1}$?

   (d) uudelle havainnolle Y_{n+1}

   (Kohdissa c ja d on ajateltu, että uuteen havaintoyksikköön $n+1$ liittyvät selittävien tekijöiden arvot on koottu vektoriin X_{n+1} ja että sama malli (A) pätee myös tässä uudessa havaintoyksikössä.)

   Kohdassa b on tarkoitus ensin päätellä estimaattorin $C'\widehat{\beta}$ otantajakauman muoto.

   Elävöitä vastaustasi seuraavalla esimerkillä. Tehtävässä 2 mallin (B) mukainen regressiofunktio olisi astetta $\gamma = \beta_2 + \beta_3$ oleva homogeenifunktio, ts. 10 %:n lisäyksen molemmissa mittatiedoissa d_i ja $d6_i$, pitäisi aiheuttaa kertoimen $(1.10)^\gamma$ mukaisen kasvun puun tilavuuteen. Muodosta liitteessä 2 esitettyjen tulosten perusteella 95%-tasoinen luottamusväli skaalaparametrille $\gamma = \beta_2 + \beta_3$. (Huom, SAS:in proc reg:in optiota covb ja R:n funktiota vcov on käytetty tulostamaan matriisi $S^2(X'X)^{-1}$, missä $S^2 = \frac{1}{n-m}E'E$)

4. Miten perusmuotoiseen lineaariseen regressiomalliin (A) sisältyvien oletusten

   - virhetermien ϵ_i varianssit σ_i^2 oletettiin kaikki yhtä suuriksi
   - eri havaintoyksiköihin sisältyvät viehetermit oletettiin toisistaan riippumattomiksi
   - virhetermit oletettiin normaalisesti jakautuneiksi

   realistisuutta voidaan tutkia

   (a) visuaalisesti

   (b) testien avulla

   Mikäli jokin oletuksista osoittautuu epärealistiseksi, mitä asian korjaamiseksi (tai huomioon ottamiseksi) voidaan tehdä?

   Elävöitä vastaustasi liitteessä 3 esitellyillä kuvioilla ja laskelmilla, joiden tarkoituksena on tutkia tehtävissä 1 ja 2 (ts. liitteissä 1 ja 2) estimoitujen mallien realistisuutta. Huomaatko liitteen 3 tulosten perusteella syyn liitteessä 1 saatuihin, jossain määrin outoihin tuloksiin?

5. Miten saman lineaarisen mallin puitteissa voidaan tutkia sekä kvalitatiivisten että kvantitatiivisten selittäjien vaikutuksia jatkuvaan vastemuuttujaan? Millaisia kysymyksenasetteluja tällaisten mallien puitteissa voidaan tutkia? Millaisia parametrirajoitteita tavallisimmat tulkinnallisesti mielenkiintoiset hypoteesit vastaisivat?

Elävöitä vastaustasi liitteessä 4 esitellyn sovellusesimerkin avulla. Datatiedostossa SYNTPAIN on esitetty eräitä tietoja Oulun- ja Lapin lääneissä vuoden 1966 aikana sekä 1.7.1985-30.6.1986 välisenä aikana syntyneistä lapsista, heidän äideistään ja heidän syntymäpainoistaan. Tiedostossa esitetty aineisto on umpimähkäinen otos kaikista mainittuina aikaväleinä syntyneistä lapsista, ja se käsittää runsaat 2100 lasta. Halutaan tutkia äidin raskauden aikaisen tupakoinnin vaikutuksia keskimääräisiin syntymäpainoihin (muuttuja $SYNTPAIN$). Toisaalta on ilmeistä, että keskimääräinen syntymäpaino vaihtelee myös sukupuolittain.

Lisäksi on selvää, että raskauden kesto (ns. gestaatioikä, muuttuja $NEUVVIIK$) vaikuttaa oleellisesti lapsen syntymäpainoon. (Aineistossa olivat mukana vain yksisikiöiset raskaudet.) Plottauskuvioiden avulla on helppo huomata, ettei yhteys ole lineaarinen, vaan että sitä voidaan parhaiten approksimoida kolmannen asteen polynomin avulla, esimerkiksi mallin

$$\text{(C)} \quad \begin{aligned} &SYNTPAIN_i = \beta_1 + \beta_2 ga_i + \beta_3 ga_i^2 + \beta_4 ga^3 + \beta_5 sex_i + \varepsilon_i \\ &\varepsilon \sim N_n(0, \sigma^2 I),\ i = 1, \ldots, n\ , \end{aligned}$$

mukaisesti, kun $SYNTPAIN_i$ vastaa i. lapsen syntymäpainoa (10 grammoissa mitattuna), ga_i lapsen gestaatioikää (viikoissa mitattuna) ja sex_i sukupuolta (1=tyttö, 0=poika). Parametreja $\beta_1, \ldots, \beta_5$ pidetään rajoitteettomina, mutta $\sigma^2 > 0$.

Vertaa mallia (C) malliin

$$\text{(D)} \quad \begin{aligned} &SYNTPAIN_i = \beta_1^* + \beta_2^*(ga_i - 40) + \beta_3^*(ga_i - 40)^2 + \beta_4^*(ga_i - 40)^3 + \beta_5^* sex_i + \varepsilon_i \\ &\varepsilon \sim N_n(0, \sigma^2 I),\ i = 1, \ldots, n, \\ &\beta_1^*, \ldots, \beta_5^* \text{ rajoitteettomia, } \sigma^2 \end{aligned}$$

Millainen yhteys vallitsee muotoilun (C) parametrin β_5 ja muotoilun (D) parametrin β_5^* välillä? Miten parametrit β_5 ja β_5^* voidaan tulkita? Mikä olisi parametrin β_1^* selkokielinen tulkinta? Entä parametrin β_1?

Täydennetään mallia siten, että se pyrkii ottamaan huomioon myös äidin tupakoinnin vaikutuksen lapsen syntymäpainoon. Tiedot äidin tupakoinnista kirjattiin muuttujaan $AIDINTUP$ muodossa 0 $\leftrightarrow$ äiti ei tupakoinut lainkaan,1 $\leftrightarrow$ äiti poltti korkeintaan 9 savuketta päivässä, 2 $\leftrightarrow$ 'äiti poltti 10 savuketta päivässä tai sitä enemmän. Karkeistetaan nämä tiedot muotoon $tup = 0 \leftrightarrow$ äiti poltti korkeintaan 9 savuketta päivässä tai ei tupakoinut lainkaan, $tup = 1 \leftrightarrow$ äiti poltti vähintään 10 savuketta päivässä. Näyttääkö gestaatioiän vaikutus syntymäpainoon samanlaiselta tupakoivien ja tupakoimattomien äitien lapsilla? (Tutki tätä kysymystä sopivaa F-testiä käyttäen)

Näyttääkö äidin tupakoinnilla ja lapsen sukupuolella olevan "yhdysvaikutusta" lapsen syntymäpainoon? Näyttääkö äidin tupakointi ylipäätään vaikuttavan lapsen syntymäpainoon? (Perustele vastauksesi huolellisesti liitteen 4 laskelmiin tukeutuen.)

Kuinka suuri näyttäisi olevan täysiaikaisen ($ga = 40$) poikalapsen keskimääräinen syntymäpaino, mikäli hänen äitinsä ei ole tupakoinut lainkaan?

Liite 1

SAS:

```
data mantil; infile 'c:\data\kustavi.dat';
 input d d6 h tilav;
 hrinta=d*(600-140)/(d-d6);
 htot=hrinta+140; d0=(htot/hrinta)*d;
 x=3.14*d0*d0*htot/(12*1000); run;
proc reg data=mantil;
 model tilav = x  ;  /* Malli 1 */
 output out=mantil predicted=m1sovite residual=m1e; run;
```

The REG Procedure
Model: MODEL1
Dependent Variable: tilav

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	6817250	6817250	1854.40	<.0001
Error	102	374978	3676.25620		
Corrected Total	103	7192228			

Root MSE	60.63214	R-Square	0.9479
Dependent Mean	488.08654	Adj R-Sq	0.9474
Coeff Var	12.42242		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	29.69937	12.19249	2.44	0.0166
x	1	0.87220	0.02025	43.06	<.0001

R:

```
> mantil<-read.table("c:/data/kustavi.dtr",header=TRUE)
> attach(mantil)
> hrinta<-d*(600-140)/(d-d6)
> htot<-hrinta+140
> d0<-(htot/hrinta)*d
> mantil$x<-(3.14*d0*d0*htot)/(12*1000)
> attach(mantil)
> m1<-lm(tilav~x , data=mantil)
> summary(m1)

Call:
lm(formula = tilav ~ x, data = mantil)

Residuals:
     Min       1Q   Median       3Q      Max
-188.249  -22.376   -1.737   27.297  168.354

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 29.69937   12.19249   2.436   0.0166 *
x            0.87220    0.02025  43.063   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 60.63 on 102 degrees of freedom
Multiple R-Squared: 0.9479,     Adjusted R-squared: 0.9474
F-statistic:  1854 on 1 and 102 DF,  p-value: < 2.2e-16

> sum(m1$residual^2)
[1] 374978.136
```

Liite 2

SAS:

```
data mantil; infile 'c:\data\kustavi.dat';
 input d d6 h tilav;
 ltilav=log(tilav); ld=log(d);
 ld6=log(d6); run;
proc reg data=mantil;
 model ltilav = ld ld6 / covb ;  /* Malli 2 */
 output out=mantil predicted=m2sovite residual=logm2e;
run;
```

The REG Procedure
Model: MODEL1
Dependent Variable: ltilav

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	54.92962	27.46481	1756.24	<.0001
Error	101	1.57948	0.01564		
Corrected Total	103	56.50910			

Root MSE	0.12505	R-Square	0.9720
Dependent Mean	5.99035	Adj R-Sq	0.9715
Coeff Var	2.08759		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-0.64848	0.17800	-3.64	0.0004
ld	1	1.07197	0.12475	8.59	<.0001
ld6	1	1.05228	0.09197	11.44	<.0001

Covariance of Estimates

Variable	Intercept	ld	ld6
Intercept	0.0316826771	-0.019033278	0.0104446641
ld	-0.019033278	0.0155629731	-0.0108174
ld6	0.0104446641	-0.0108174	0.0084583703

Sovitteiden transformointi takaisin litraskaalaan:
==========

```
data mantil; set mantil;
 m2e=tilav-exp(m2sovite+0.5*0.01564);
 m1e2=m1e*m1e;
 m2e2=m2e*m2e; run;
```

Jäännöstermien neliöiden keskiarvojen laskeminen:
==========

```
proc means data=mantil; var m1e2 m2e2; run;
```

The MEANS Procedure

Variable	N	Mean	Std Dev	Minimum	Maximum
m1e2	104	3605.56	7044.06	1.5230729E-6	35437.54
m2e2	104	3505.26	6545.71	0.3293467	39318.74

R:

```
> m2<-lm(log(tilav)~log(d)+log(d6), data=mantil)
> summary(m2)
Call:
lm(formula = log(tilav) ~ log(d) + log(d6), data = mantil)

Residuals:
     Min       1Q   Median       3Q      Max
-0.34931 -0.07893 -0.01098  0.07369  0.51801

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.64848    0.17800  -3.643 0.000427 ***
log(d)       1.07197    0.12475   8.593 1.11e-13 ***
log(d6)      1.05228    0.09197  11.442  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1251 on 101 degrees of freedom
Multiple R-Squared: 0.972,      Adjusted R-squared: 0.9715
F-statistic:  1756 on 2 and 101 DF,  p-value: < 2.2e-16
```

```
> vcov(m2)
            (Intercept)      log(d)     log(d6)
(Intercept)  0.03168268 -0.01903328  0.01044466
log(d)      -0.01903328  0.01556297 -0.01081740
log(d6)      0.01044466 -0.01081740  0.00845837
>
> m2s2<-sum(m2$residual^2)/(length(tilav)-3)
> jaannosm2<-tilav-exp(m2$fitted)*exp(0.5*m2s2)
> sum(jaannosm2^2)/length(tilav)
[1] 3505.254
> sum(m1$residual^2)/length(tilav)
[1] 3605.559
```

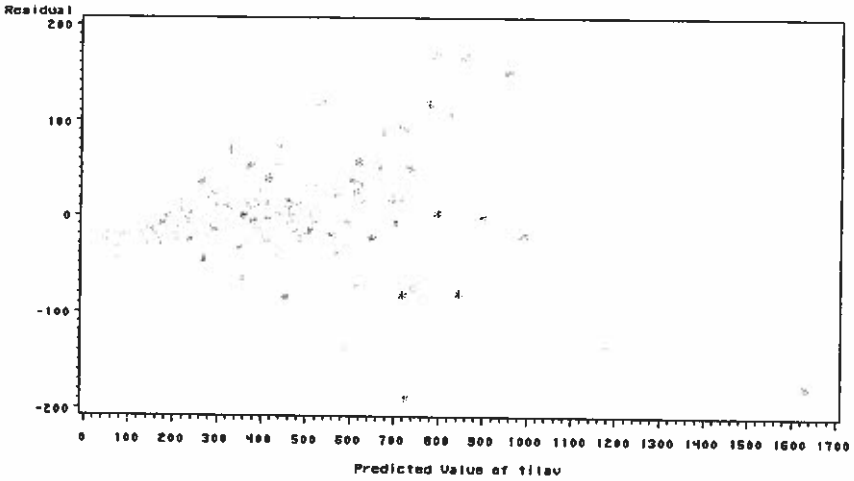
Liite 3

SAS:

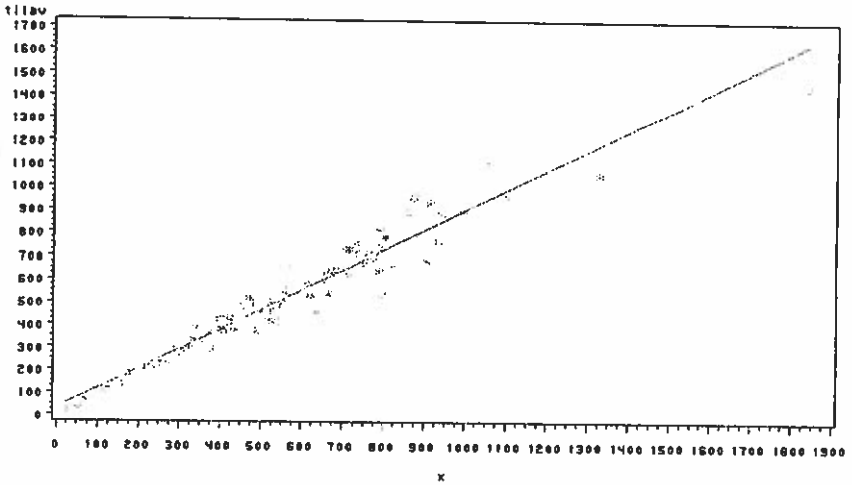
```
data mantil; infile 'c:\data\kustavi.dat';
 input d d6 h tilav;
 hrinta=d*(600-140)/(d-d6);
 htot=hrinta+140; d0=(htot/hrinta)*d;
 x=3.14*d0*d0*htot/(12*1000);
 ltilav=log(tilav); ld=log(d); ld6=log(d6);
 painot=1/x**2; havno=_n_; run;
proc reg data=mantil;
 model tilav = x  ;  /* Malli 1 */
 output out=mantil predicted=m1sovite residual=m1e; run;
proc gplot data=mantil;
 plot m1e*m1sovite;
 symbol v=star c=green i=none; run;
proc gplot data=mantil;
 plot (tilav m1sovite)*x / overlay;
 symbol1 v=star c=green i=none;
 symbol2 v=none c=red i=join; run;
```

Huom.: Estimointitulokset löytyvät Liitteestä 1.

Malli 1: Jäännöstermit vastaan sovitteet



Malli 1: Havaintojen sirontakuvio sekä sovitesuora




```
data BP; set mantil; u=(m1e**2)/(374978/104); run;
proc reg data=BP;
 model u=x ; run;
```

The REG Procedure
Dependent Variable: u
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	133.21974	133.21974	52.28	<.0001
Error	102	259.91189	2.54816		
Corrected Total	103	393.13164			

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-1.02634	0.32100	-3.20	0.0018
x	1	0.00386	0.00053324	7.23	<.0001

```
proc reg data=mantil;
 model tilav = x  ;  /* Malli 1 painottaen estimoituna */
 weight painot; run;
```

The REG Procedure
Dependent Variable: tilav

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	62.61115	62.61115	5847.44	<.0001
Error	102	1.09216	0.01071		
Corrected Total	103	63.70331			

Root MSE	0.10348	R-Square	0.9829
Dependent Mean	85.32562	Adj R-Sq	0.9827
Coeff Var	0.12127		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.66109	1.98795	0.33	0.7402
x	1	0.93421	0.01222	76.47	<.0001

.....

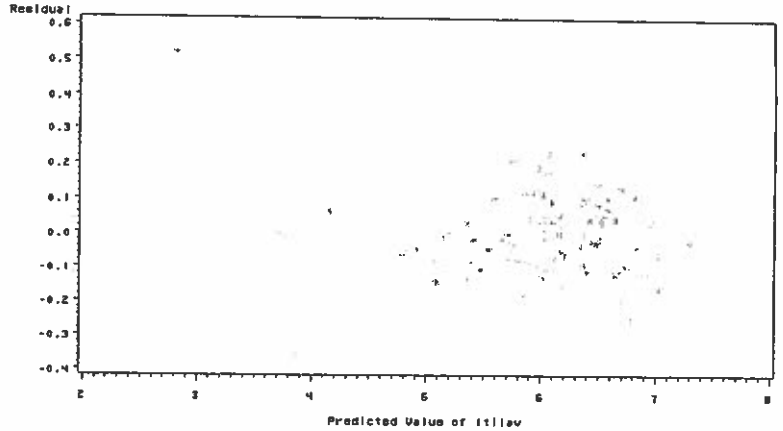
Liitteen 2 malli:

```
proc reg data=mantil;
 model ltilav = ld ld6 ;  /* Malli 2 */
 output out=mantil predicted=m2sovite residual=logm2e; run;
```

Huom.: Estimointitulokset löytyvät Liitteestä 2.

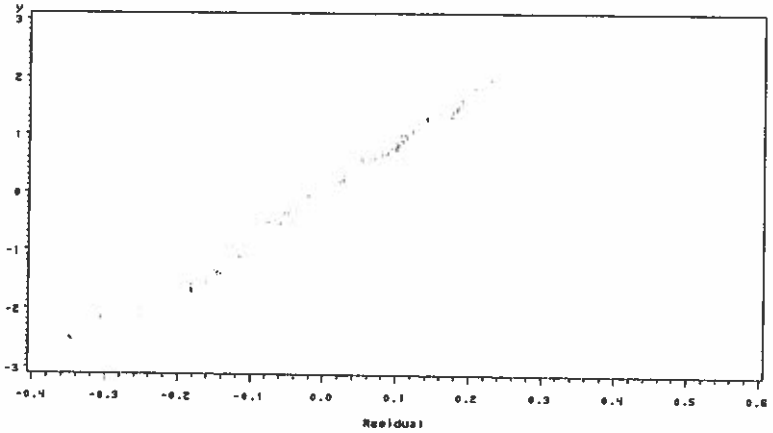
```
proc gplot data=mantil;
 plot logm2e*m2sovite;
 symbol v=star c=green i=none; run;
```


Malli 2: Jäännöstermit vastaan sovitteet




```
proc sort data=mantil out=sortmant; by logm2e; run;
data sortmant; set sortmant; j=_n_;
 y=probit((j-0.375)/(104+0.25)); run;
proc gplot data=sortmant; plot y*logm2e; symbol v=star; run;
```

Malli 2: Jäännöstermien probit-kuvio



```
proc univariate data=mantil normal; var logm2e;
 output out=JB n=frekv skewness=g1 kurtosis=g2; run;
```

The UNIVARIATE Procedure
Variable: logm2e (Residual)

Moments

N	104	Sum Weights	104
Mean	0	Sum Observations	0
Std Deviation	0.12383357	Variance	0.01533475
Skewness	0.48146231	Kurtosis	2.39953675
Uncorrected SS	1.57947946	Corrected SS	1.57947946
Coeff Variation	.	Std Error Mean	0.01214288

Tests for Normality

Test	--Statistic---		-----p Value------	
Shapiro-Wilk	W	0.969373	Pr < W	0.0164

```
data JB; set JB; JB=frekv*((g1*g1/6)+(g2*g2/24));
 p=1-probchi(JB,2); run;
proc print data=JB; var JB p; run;
```

Obs	JB	p
1	28.9683	.000000512

.....

Poistetaan kokeeksi havainto numero 63, jonka mittatiedot ovat seuraavat:

d	d6	h	tilav
8.600	3.000	7.690	28

Puu siis kapenee aluksi *erittäin* voimakkaasti, mutta on siitä huolimatta varsin korkea, ja tilavuuttakin kertyy 28 litraa.

```
proc reg data=mantil;
 model ltilav = ld ld6 ;  /* Malli 2 ilman havaintoa no. 63 */
 where havno^=63;
 output out=mantil predicted=m2sovite residual=logm2e; run;
proc gplot data=mantil;
 plot logm2e*m2sovite;
 symbol v=star c=green i=none; run;
```


The REG Procedure
Dependent Variable: ltilav
Analysis of Variance

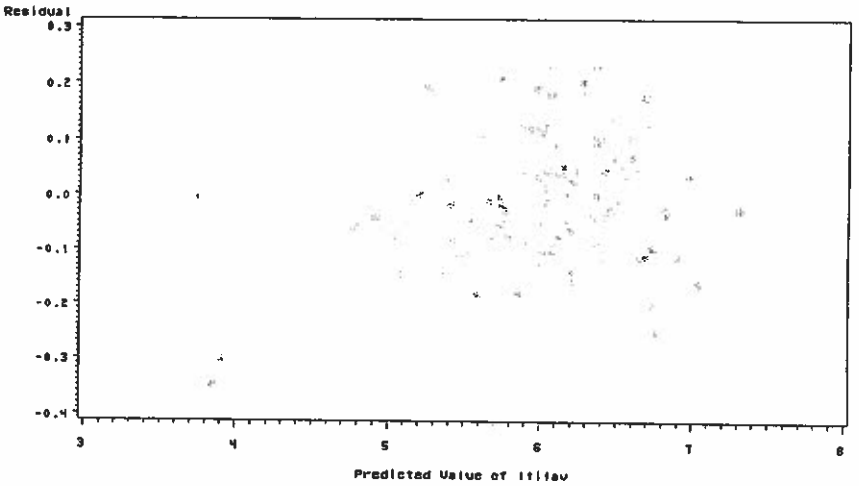
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	48.15110	24.07555	1967.48	<.0001
Error	100	1.22367	0.01224		
Corrected Total	102	49.37477			

Root MSE	0.11062	R-Square	0.9752
Dependent Mean	6.01615	Adj R-Sq	0.9747
Coeff Var	1.83871		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-0.74520	0.15847	-4.70	<.0001
ld	1	0.95907	0.11232	8.54	<.0001
ld6	1	1.20774	0.08631	13.99	<.0001

Malli 2: Jäännöstermit vastaan sovitteet (ilman havaintoa 63)



```
proc sort data=mantil out=sortmant; by logm2e; run;
data sortmant; set sortmant; j=_n_;
 y=probit((j-0.375)/(103+0.25)); run;
proc gplot data=sortmant;
 plot y*logm2e;
 symbol v=star; run;
proc univariate data=mantil normal; var logm2e;
 output out=JB n=frekv skewness=g1 kurtosis=g2; run;
data JB; set JB; JB=frekv*((g1*g1/6)+(g2*g2/24));
 p=1-probchi(JB,2); run;
proc print data=JB; var JB p; run;
```

The UNIVARIATE Procedure
Variable: logm2e (Residual)

Moments

N	103	Sum Weights	103
Mean	-0.0050292	Sum Observations	-0.518006
Std Deviation	0.11326458	Variance	0.01282886
Skewness	-0.1849626	Kurtosis	0.23814289
Uncorrected SS	1.31114929	Corrected SS	1.30854414
Coeff Variation	-2252.1462	Std Error Mean	0.01116029

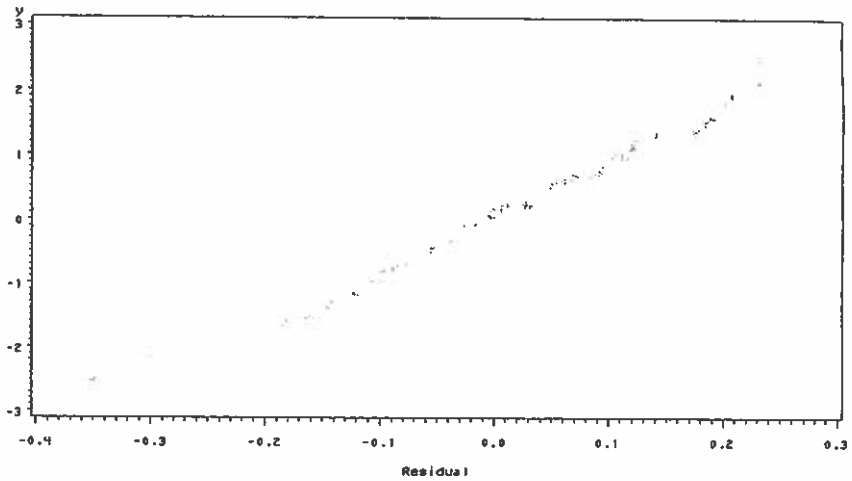
Tests for Normality

Test	--Statistic---		-----p Value------	
Shapiro-Wilk	W	0.987975	Pr < W	0.4845

```
data JB; set JB; JB=frekv*((g1*g1/6)+(g2*g2/24));
 p=1-probchi(JB,2); run;
proc print data=JB; var JB p; run;
```

Obs	JB	p
1	0.83068	0.66012

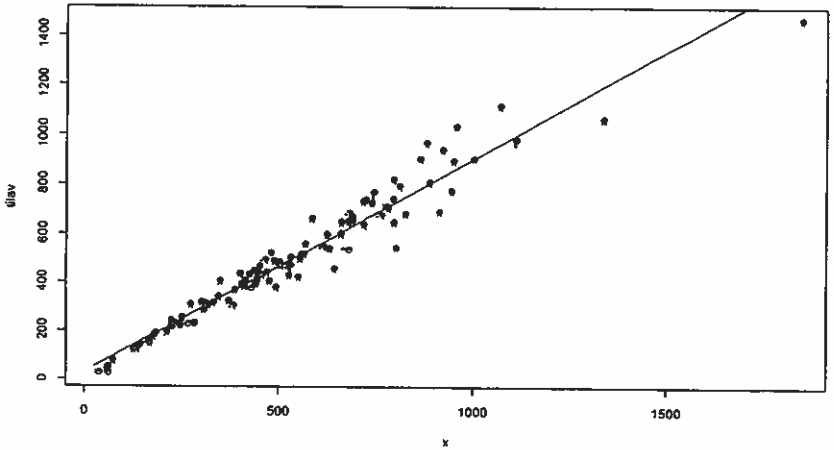
Malli 2: Jäännöstermien probit-kuvio (ilman havaintoa 63)



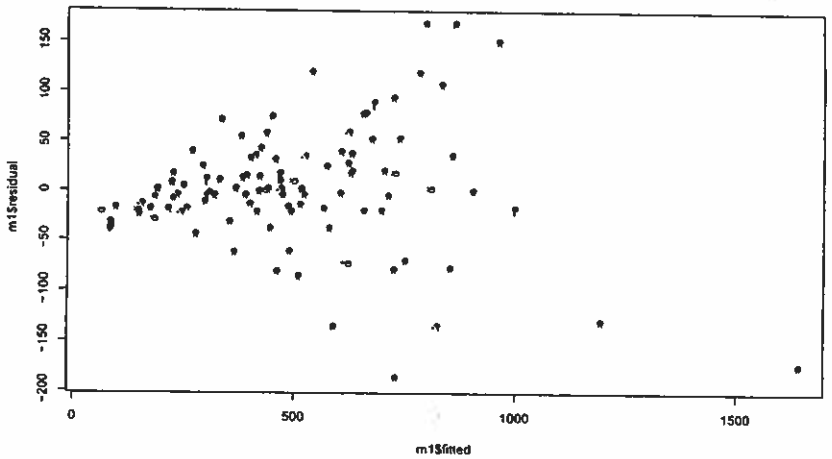
R:

```
> mantil<-read.table("c:/data/kustavi.dtr",header=TRUE)
> attach(mantil)
> hrinta<-d*(600-140)/(d-d6)
> h<-hrinta+140
> d0<-(h/hrinta)*d
> mantil$x<-3.14*d0*d0*h/(12*1000)
> mantil$painot<-1/mantil$x^2
> attach(mantil)
> m1<-lm(tilav~x , data=mantil)
> summary(m1)

Call:
lm(formula = tilav ~ x, data = mantil)

Residuals:
     Min       1Q   Median       3Q      Max
-188.249  -22.376   -1.737   27.297  168.354
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 29.69937   12.19249   2.436   0.0166 *
x            0.87220    0.02025  43.063   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 60.63 on 102 degrees of freedom
Multiple R-Squared: 0.9479,     Adjusted R-squared: 0.9474
F-statistic:  1854 on 1 and 102 DF,  p-value: < 2.2e-16

> sum(m1$residual^2)/length(tilav)
[1] 3605.559
> plot(tilav~x)
> lines(fitted.values(m1)[order(x)]~sort(x))
```

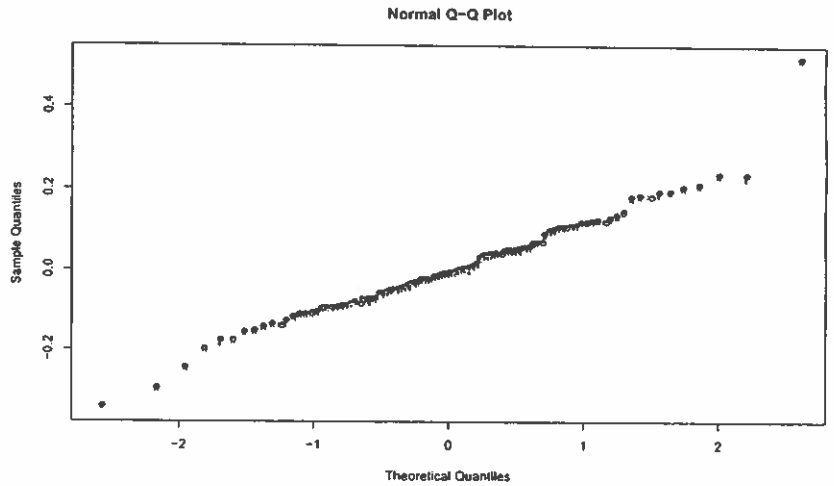


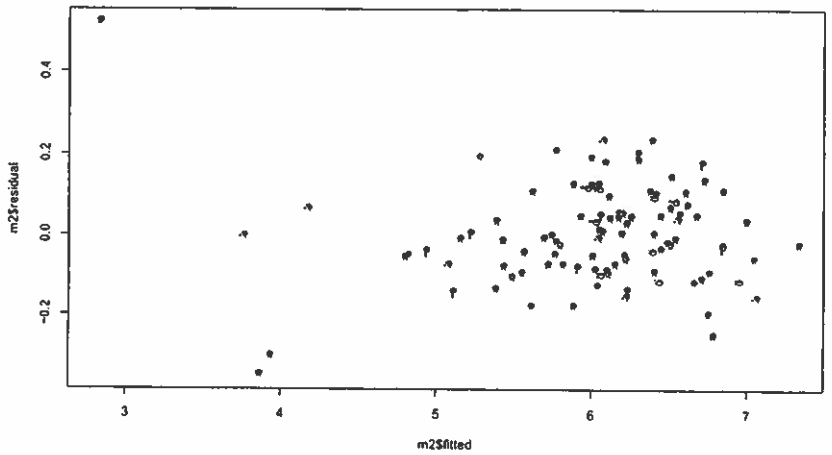

```
> plot(m1$residual~m1$fitted)
> library(zoo)
> library(lmtest)
> bptest(tilav~x,varformula=~x,data=mantil,studentize=FALSE)

        Breusch-Pagan test

data:  tilav ~ x
BP = 66.6098, df = 1, p-value = 3.309e-16
```


```
> m1<-lm(tilav~x , data=mantil,weights=painot)
> summary(m1)

Call:
lm(formula = tilav ~ x, data = mantil, weights = painot)

Residuals:
    Min      1Q  Median      3Q     Max
-146762  -20102  -12169    5564  195752

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 118.21412   19.16668   6.168 1.41e-08 ***
x             0.75626    0.02057  36.767  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 49390 on 102 degrees of freedom
Multiple R-Squared: 0.9298,     Adjusted R-squared: 0.9292
F-statistic:  1352 on 1 and 102 DF,  p-value: < 2.2e-16
>
>
> m2<-lm(log(tilav)~log(d)+log(d6), data=mantil)
> summary(m2)
```

```
Call:
lm(formula = log(tilav) ~ log(d) + log(d6), data = mantil)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.34931 -0.07893 -0.01098  0.07369  0.51801 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -0.64848    0.17800  -3.643 0.000427 ***
log(d)       1.07197    0.12475   8.593 1.11e-13 ***
log(d6)      1.05228    0.09197  11.442  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1251 on 101 degrees of freedom
Multiple R-Squared: 0.972,      Adjusted R-squared: 0.9715 
F-statistic:  1756 on 2 and 101 DF,  p-value: < 2.2e-16 

> e<-m2$residual
> g1<-mean(e^3)/mean(e^2)^(1.5)
> g2<-(mean(e^4)/(mean(e^2)^2))-3
> g1
[1] 0.4744901
> g2
[1] 2.228574
> JB<-length(e)*((g1^2/6)+(g2^2/24))
> JB
[1] 25.42412
> p<-1-pchisq(JB,df=2,ncp=0,lower.tail=TRUE,log.p=FALSE)
> p
[1] 3.014544e-06
>
> shapiro.test(m2$residual)

        Shapiro-Wilk normality test

data:  m2$residual
W = 0.9694, p-value = 0.01637
> qqnorm(m2$residual)
> plot(m2$residual~m2$fitted)
```

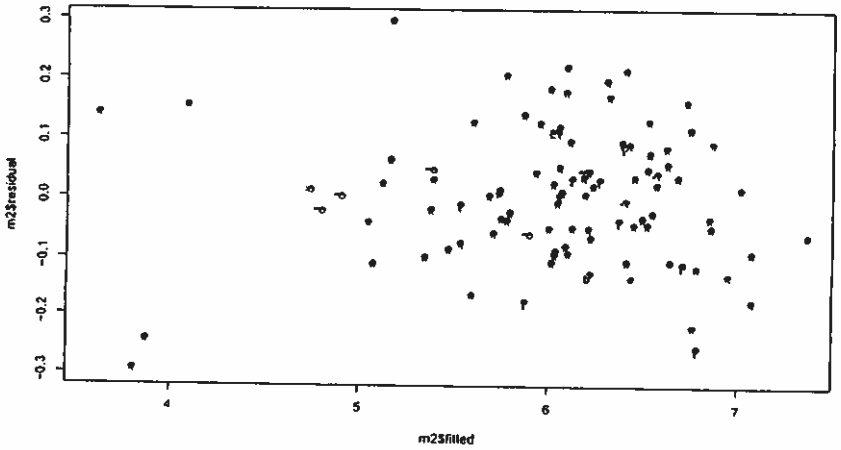





Poistetaan havainto numero 63:

```
> mantil2<-mantil[-63,]
> m2<-lm(log(tilav)~log(d)+log(d6), data=mantil2)
> summary(m2)
```

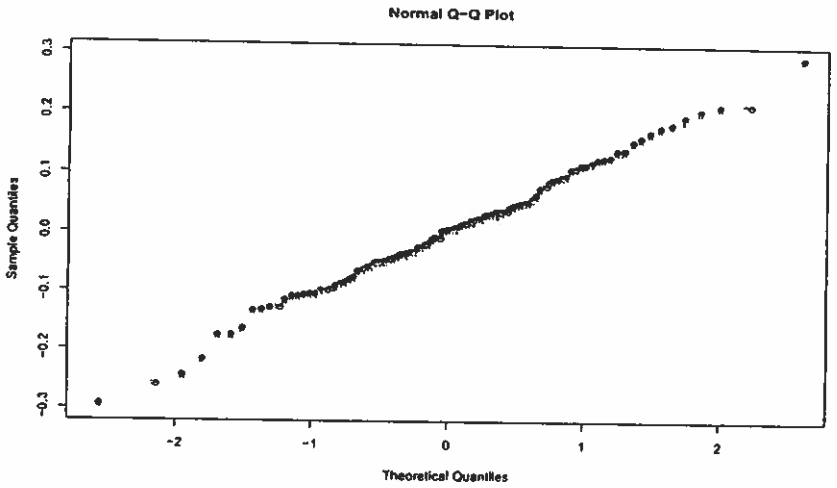
```
Call:
lm(formula = log(tilav) ~ log(d) + log(d6), data = mantil2)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.297406 -0.067217  0.002033  0.075306  0.292880 


Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -0.74520    0.15847  -4.702 8.25e-06 ***
log(d)       0.95907    0.11232   8.539 1.56e-13 ***
log(d6)      1.20774    0.08631  13.993  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1106 on 100 degrees of freedom
Multiple R-Squared: 0.9752,     Adjusted R-squared: 0.9747 
F-statistic:  1967 on 2 and 100 DF,  p-value: < 2.2e-16

> plot(m2$residual~m2$fitted)
```




```
> e<-m2$residual
> shapiro.test(e)
```



```
        Shapiro-Wilk normality test

data:  e
W = 0.9948, p-value = 0.9666

> qqnorm(e)
```



```
data sp;
 infile 'c:\data\syntpain.dat';
 input SYNTPAIN    KOHORTTI   SUKUPUOL  NEUVVIIK  AIDINTUP ;
 ga=NEUVVIIK;
 sex=SUKUPUOL-1;
 tup=0; if AIDINTUP=2 then tup=1;
 sextup=sex*tup;
 enna=ga-40;
 enna2=enna*enna;
 enna3=enna*enna2;
 tupenna=tup*enna;
 tupenna2=tup*enna2;
 tupenna3=tup*enna3;
run;
proc mixed data=sp;
 class sex tup;
 model SYNTPAIN = sex|tup enna enna2 enna3 tup*enna tup*enna2 tup*enna3 / s;
 contrast 'gest. ian vaik. samanlaista'
              enna*tup   1 -1 ,
              enna2*tup  1 -1 ,
              enna3*tup  1 -1 ;
run;
```

The Mixed Procedure

Class Level Information

Class	Levels	Values
sex	2	0 1
tup	2	0 1

Var Parm	Estimate
Residual	2171.02

Fit Statistics

-2 Res Log Likelihood	21769.6
AIC (smaller is better)	21771.6
AICC (smaller is better)	21771.6
BIC (smaller is better)	21777.3

Solution for Fixed Effects

Effect	sex	tup	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept			353.27	6.0890	2061	58.02	<.0001
sex	0		6.7555	7.9730	2061	0.85	0.3969
sex	1		0	.	.	.	.
tup		0	5.4019	6.3042	2061	0.86	0.3916
tup		1	0	.	.	.	.
sex*tup	0	0	6.8342	8.2511	2061	0.83	0.4076
sex*tup	0	1	0	.	.	.	.
sex*tup	1	0	0	.	.	.	.
sex*tup	1	1	0	.	.	.	.
enna			9.0406	4.1999	2061	2.15	0.0315
enna2			-3.3337	1.6915	2061	-1.97	0.0489
enna3			-0.2087	0.1474	2061	-1.42	0.1570
enna*tup		0	-0.3450	4.2854	2061	-0.08	0.9358
enna*tup		1	0	.	.	.	.
enna2*tup		0	1.6807	1.7155	2061	0.98	0.3273
enna2*tup		1	0	.	.	.	.
enna3*tup		0	0.1433	0.1489	2061	0.96	0.3361
enna3*tup		1	0	.	.	.	.

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
sex	1	2061	6.08	0.0138
tup	1	2061	3.16	0.0757
sex*tup	1	2061	0.69	0.4076
enna	1	2061	17.13	<.0001
enna2	1	2061	8.45	0.0037
enna3	1	2061	3.39	0.0659
enna*tup	1	2061	0.01	0.9358
enna2*tup	1	2061	0.96	0.3273
enna3*tup	1	2061	0.93	0.3361

Contrasts

Label	Num DF	Den DF	F Value	Pr > F
gest. ian vaik. samanlaista	3	2061	1.47	0.2210

```
*****************************************************************************

proc reg data=sp;
 model SYNTPAIN = sex tup sextup enna enna2 enna3 tupenna tupenna2 tupenna3 ;
 test  tupenna=0 , tupenna2=0 ,  tupenna3=0 ;
run;
```

The REG Procedure
Model: MODEL1
Dependent Variable: SYNTPAIN

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	2194675	243853	112.32	<.0001
Error	2061	4474471	2171.01924		
Corrected Total	2070	6669146			

Root MSE	46.59420	R-Square	0.3291
Dependent Mean	350.79479	Adj R-Sq	0.3261
Coeff Var	13.28247		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	372.26517	1.61892	229.95	<.0001
sex	1	-13.58971	2.12407	-6.40	<.0001
tup	1	-12.23614	6.59839	-1.85	0.0638
sextup	1	6.83419	8.25112	0.83	0.4076
enna	1	8.69551	0.85148	10.21	<.0001
enna2	1	-1.65302	0.28590	-5.78	<.0001
enna3	1	-0.06538	0.02125	-3.08	0.0021
tupenna	1	0.34504	4.28539	0.08	0.9358
tupenna2	1	-1.68072	1.71551	-0.98	0.3273
tupenna3	1	-0.14330	0.14894	-0.96	0.3361

Liite 4

The REG Procedure
Model: MODEL1

Test 1 Results for Dependent Variable SYNTPAIN

Source	DF	Mean Square	F Value	Pr > F
Numerator	3	3189.41837	1.47	0.2210
Denominator	2061	2171.01924		

****

```
proc reg data=sp;
 model SYNTPAIN = sex tup sextup enna enna2 enna3 ;
run;
```

The REG Procedure
Model: MODEL1
Dependent Variable: SYNTPAIN

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	2185107	364184	167.63	<.0001
Error	2064	4484039	2172.49947		
Corrected Total	2070	6669146			

Root MSE	46.61008	R-Square	0.3276
Dependent Mean	350.79479	Adj R-Sq	0.3257
Coeff Var	13.28699		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	372.69175	1.60624	232.03	<.0001
sex	1	-13.59531	2.12470	-6.40	<.0001
tup	1	-19.19402	5.67476	-3.38	0.0007
sextup	1	9.47396	8.12644	1.17	0.2438
enna	1	8.78625	0.83184	10.56	<.0001
enna2	1	-1.77665	0.27721	-6.41	<.0001
enna3	1	-0.07421	0.02077	-3.57	0.0004

****

```
proc reg data=sp;
 model SYNTPAIN = sex tup  enna enna2 enna3 ;
run;
```

The REG Procedure
Model: MODEL1
Dependent Variable: SYNTPAIN

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	2182154	436431	200.85	<.0001
Error	2065	4486992	2172.87730		
Corrected Total	2070	6669146			

Root MSE	46.61413	R-Square	0.3272
Dependent Mean	350.79479	Adj R-Sq	0.3256
Coeff Var	13.28815		

Parameter Estimates

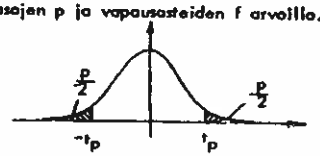
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	372.41057	1.58816	234.49	<.0001
sex	1	-12.94750	2.05093	-6.31	<.0001
tup	1	-14.57106	4.05971	-3.59	0.0003
enna	1	8.77653	0.83187	10.55	<.0001
enna2	1	-1.79142	0.27694	-6.47	<.0001
enna3	1	-0.07507	0.02076	-3.62	0.0003

TAULUKKO 2 : Studentin t – jakauma :

Kriittisiä arvoja t_p joillakin merkitsevyystasojen p ja vapausasteiden f arvoilla.

Esim. Kaksisuuntainen testi merkitsevyystasolla p

Jos f=3 ja p=0.05 on $P(|\underline{t}|>3.182)=0.05$



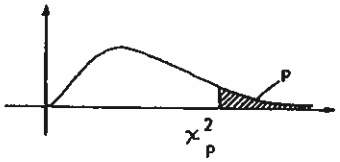
f	merkitsevyystaso p yksisuuntaisessa testissä					
	0.05	0.025	0.01	0.005	0.001	0.0005
	merkitsevyystaso p kaksisuuntaisessa testissä					
	0.10	0.05	0.02	0.01	0.002	0.001
1	6.314	12.706	31.821	63.657	318.311	636.619
2	2.920	4.303	6.965	9.925	22.326	31.598
3	2.353	3.182	4.541	5.841	10.213	12.941
4	2.132	2.776	3.747	4.604	7.173	8.610
5	2.015	2.571	3.365	4.032	5.893	6.859
6	1.943	2.447	3.143	3.707	5.208	5.959
7	1.895	2.365	2.998	3.499	4.785	5.405
8	1.860	2.306	2.896	3.355	4.501	5.041
9	1.833	2.262	2.821	3.250	4.297	4.781
10	1.812	2.228	2.764	3.169	4.144	4.587
11	1.796	2.201	2.718	3.106	4.025	4.437
12	1.782	2.179	2.681	3.055	3.930	4.318
13	1.771	2.160	2.650	3.012	3.852	4.221
14	1.761	2.145	2.624	2.977	3.787	4.140
15	1.753	2.131	2.602	2.947	3.733	4.073
16	1.746	2.120	2.583	2.921	3.686	4.015
17	1.740	2.110	2.567	2.898	3.646	3.965
18	1.734	2.101	2.552	2.878	3.611	3.922
19	1.729	2.093	2.539	2.861	3.579	3.883
20	1.725	2.086	2.528	2.845	3.552	3.850
21	1.721	2.080	2.518	2.831	3.527	3.819
22	1.717	2.074	2.508	2.819	3.505	3.792
23	1.714	2.069	2.500	2.807	3.485	3.767
24	1.711	2.064	2.492	2.797	3.467	3.745
25	1.708	2.060	2.485	2.787	3.450	3.725
26	1.706	2.056	2.479	2.779	3.435	3.707
27	1.703	2.052	2.473	2.771	3.421	3.690
28	1.701	2.048	2.467	2.763	3.408	3.674
29	1.699	2.045	2.462	2.756	3.396	3.659
30	1.697	2.042	2.457	2.750	3.385	3.646
40	1.684	2.021	2.423	2.704	3.307	3.551
50	1.676	2.009	2.403	2.678	3.262	3.495
60	1.671	2.000	2.390	2.660	3.232	3.460
80	1.664	1.990	2.374	2.639	3.195	3.415
100	1.660	1.984	2.365	2.626	3.174	3.389
200	1.653	1.972	2.345	2.601	3.131	3.339
500	1.648	1.965	2.334	2.586	3.106	3.310
∞	1.645	1.960	2.326	2.576	3.090	3.291

Huom.
Viimeisellä rivillä olevat luvut ovat normaalijakauman kriittisiä arvoja z_p.

TAULUKKO 3 : χ^2 – jakauma :

Yksisuuntaiseen testiin liittyviä kriittisiä arvoja χ^2_p joillakin merkitsevyystasojen p ja vapausasteiden f arvoilla.

Esim. Jos f = 15 ja p = 0.05 on $P(\underline{\chi}^2 > 24.996) = 0.05$

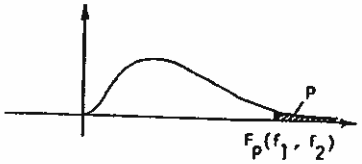


f \ p	0.99	0.95	0.10	0.05	0.01	0.001
1	0.000	0.004	2.706	3.841	6.635	10.828
2	0.020	0.103	4.605	5.991	9.210	13.816
3	0.115	0.352	6.251	7.815	11.345	16.266
4	0.297	0.711	7.779	9.488	13.277	18.467
5	0.554	1.145	9.236	11.070	15.086	20.515
6	0.872	1.635	10.645	12.592	16.812	22.458
7	1.239	2.167	12.017	14.067	18.475	24.322
8	1.646	2.733	13.362	15.507	20.090	26.125
9	2.088	3.325	14.684	16.919	21.666	27.877
10	2.558	3.940	15.987	18.307	23.209	29.588
11	3.053	4.575	17.275	19.675	24.725	31.264
12	3.571	5.226	18.549	21.026	26.217	32.909
13	4.107	5.892	19.812	22.362	27.688	34.528
14	4.660	6.571	21.064	23.685	29.141	36.123
15	5.229	7.261	22.307	24.996	30.578	37.697
16	5.812	7.962	23.542	26.296	32.000	39.252
17	6.408	8.672	24.769	27.587	33.409	40.790
18	7.015	9.390	25.989	28.869	34.805	42.312
19	7.633	10.117	27.204	30.144	36.191	43.820
20	8.260	10.851	28.412	31.410	37.566	45.315
21	8.897	11.591	29.615	32.671	38.932	46.797
22	9.542	12.338	30.813	33.924	40.289	48.268
23	10.196	13.091	32.007	35.172	41.638	49.728
24	10.856	13.848	33.196	36.415	42.980	51.179
25	11.524	14.611	34.382	37.652	44.314	52.620
26	12.198	15.379	35.563	38.885	45.642	54.052
27	12.879	16.151	36.741	40.113	46.963	55.476
28	13.565	16.928	37.916	41.337	48.278	56.892
29	14.256	17.708	39.087	42.557	49.588	58.302
30	14.953	18.493	40.256	43.773	50.892	59.703
40	22.164	26.509	51.805	55.758	63.691	73.402
50	29.707	34.764	63.167	67.505	76.154	86.661
60	37.485	43.188	74.397	79.082	88.379	99.607
70	45.442	51.739	85.527	90.531	100.425	112.317
80	53.540	60.391	96.578	101.879	112.329	124.839
90	61.754	69.126	107.565	113.145	124.116	137.208
100	70.065	77.930	118.498	124.342	135.806	149.448

TAULUKKO 4.1.: Fisherin F - jakauma :

Yksisuuntaiseen testiin liittyviä kriittisiä arvoja $F_p(f_1, f_2)$ merkitsevyystasolla p vapausasteiden f_1 ja f_2 eri kombinaatioilla.

$$F(f_1, f_2) = \frac{\chi_1^2/f_1}{\chi_2^2/f_2} = \frac{s_1^2}{s_2^2}$$



f_2 \ f_1	1	2	3	4	5	6	7	8	12	16	24	∞
	merkitsevyystaso p = 0.05											
1	161	200	216	225	230	234	237	239	244	246	249	254
2	18.5	19.0	19.2	19.2	19.3	19.3	19.3	19.4	19.4	19.4	19.5	19.5
3	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.74	8.69	8.64	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	5.91	5.84	5.77	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.68	4.60	4.53	4.37
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.00	3.92	3.84	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.57	3.49	3.41	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.28	3.20	3.12	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.07	2.99	2.90	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	2.91	2.83	2.74	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.79	2.70	2.61	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.69	2.60	2.51	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.60	2.51	2.42	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.53	2.44	2.35	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.48	2.38	2.29	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.42	2.33	2.24	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.38	2.29	2.19	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.34	2.25	2.15	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.31	2.21	2.11	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.28	2.18	2.08	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.25	2.16	2.05	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.23	2.13	2.03	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.20	2.11	2.00	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.18	2.09	1.98	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.16	2.07	1.96	1.71
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.15	2.05	1.95	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.13	2.04	1.93	1.67
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.12	2.02	1.91	1.65
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.10	2.01	1.90	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.09	1.99	1.89	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.00	1.90	1.79	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	1.92	1.82	1.70	1.39
100	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.85	1.75	1.63	1.28
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.75	1.64	1.52	1.00

Esim. $F_{0.05}(3, 15) = 3.29$ eli $P(\underline{F}(3, 15) > 3.29) = 0.05$

TAULUKKO 4.2.: Fisherin F - jakauma

Yksisuuntaiseen testiin liittyviä kriittisiä arvoja $F_p(f_1, f_2)$ merkitsevyystasolla p vapausasteiden f_1 ja f_2 eri kombinaatioilla.

$$F(f_1, f_2) = \frac{\chi_1^2/f_1}{\chi_2^2/f_2} = \frac{s_1^2}{s_2^2}$$

f_2 \ f_1	1	2	3	4	5	6	7	8	12	16	24	∞
	merkitsevyystaso p = 0.01											
1	4050	5000	5400	5630	5760	5860	5930	5980	6110	6170	6230	6370
2	98.5	99.0	99.2	99.2	99.3	99.3	99.4	99.4	99.4	99.4	99.5	99.5
3	34.1	30.8	29.5	28.7	28.2	27.9	27.7	27.5	27.1	26.8	26.6	26.1
4	21.2	18.0	16.7	16.0	15.5	15.2	15.0	14.8	14.4	14.2	13.9	13.5
5	16.3	13.3	12.1	11.4	11.0	10.7	10.5	10.3	9.89	9.68	9.47	9.02
6	13.7	10.9	9.78	9.15	8.75	8.47	8.26	8.10	7.72	7.52	7.31	6.88
7	12.2	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.47	6.27	6.07	5.65
8	11.3	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.67	5.48	5.28	4.86
9	10.6	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.11	4.92	4.73	4.31
10	10.0	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.71	4.52	4.33	3.91
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.40	4.21	4.02	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.16	3.97	3.78	3.36
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	3.96	3.78	3.59	3.17
14	8.86	6.51	5.56	5.04	4.70	4.46	4.28	4.14	3.80	3.62	3.43	3.00
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.67	3.49	3.29	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.55	3.37	3.18	2.75
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.46	3.27	3.08	2.65
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.37	3.19	3.00	2.57
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.30	3.12	2.92	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.23	3.05	2.86	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.17	2.99	2.80	2.36
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.12	2.94	2.75	2.31
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.07	2.89	2.70	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.03	2.85	2.66	2.21
25	7.77	5.57	4.68	4.18	3.86	3.63	3.46	3.32	2.99	2.81	2.62	2.17
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	2.96	2.78	2.58	2.13
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	2.93	2.75	2.55	2.10
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	2.90	2.72	2.52	2.06
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	2.87	2.69	2.49	2.03
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	2.84	2.66	2.47	2.01
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.66	2.48	2.29	1.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.50	2.31	2.12	1.60
100	6.90	4.82	3.98	3.51	3.21	2.99	2.82	2.69	2.37	2.19	1.98	1.43
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.18	2.00	1.79	1.00

Esim. $F_{0.01}(3, 15) = 5.42$ eli $P(\underline{F}(3, 15) > 5.42) = 0.01$