Háskóli Íslands	09.10.16 Línuleg algebra og tölfræði	Raunvísindadeild
Laugardagur	11. desember 2004	kl 09:00-12:00

Leyfileg hjálpargögn: Glósur, bækur og reiknivélar	**Athugið** að fartölvur eru ekki leyfðar.	**Vægi** dæma er gefið: 100 stig teljast full lausn, en alls eru stigin fleiri.

Lausnir

Notið 5% marktæknikröfu nema annað sé tekið fram. Munið að taka skýrt fram núlltilgátur og gagntilgátur þar sem það á við.

1. (30) Látum $\mathbf{X}$ vera fylkið sem hefur $\mathbf{a} = (2, 0, 2, 1, 0)'$, $\mathbf{b} = (0, 3, 0, 0, 4)'$, og $\mathbf{c} = (0, 2, -3, 0, 1)'$ sem dálkvigra. Látum síðan $\mathbf{y}$ vera dálkvigurinn $\mathbf{y} = (4, 1, 4, 0, 3)'$. Látum $\mathbf{V}$ vera það hlutrúm í $\mathbb{R}^5$, sem spannast af $\mathbf{a}$, $\mathbf{b}$ og $\mathbf{c}$.

(a) Finnið ofanvarp $\mathbf{y}$ á $\mathbf{V}$.
(b) Finnið tölur, b_1, b_2 og b_3 sem eru þannig að $b_1\mathbf{a} + b_2\mathbf{b} + b_3\mathbf{c}$ gefi sem besta nálgun að $\mathbf{y}$ í merkingu minnstu kvaðrata.
(c) Finnið kvaðratsummu frávika (SSE) í aðhvarfsgreiningaverkefninu $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$.
(d) Lýsið í orðum og með jöfnum, hvernig þið getið gefið ykkur tilheyrandi forsendur og prófað núlltilgátu um að $\mathbf{b}$ og $\mathbf{c}$ séu óþarfir við að spá um gildi $\mathbf{y}$.

Lausn:

Ath: Hér er einfaldast að byrjar á að leysa lið (b) því ef $\hat{\boldsymbol{\beta}} = (b_1, b_2, b_3)'$ ef fundið má reikna $\hat{\mathbf{y}}$ með $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$.

(a) Þótt leysa megi þennan lið með því að leysa (b) fyrst, má líka leysa hann með því að finna fyrst hornréttan einingargrunn með aðferð Gram-Schmidt og varpa síðan $\mathbf{y}$ á þann grunn, eins og í eftirfarandi. Hin leiðin er hins vegar **miklu** auðveldari.

Fyrsti einingavigurinn verður $\mathbf{v}_1 = \mathbf{a}/||a|| = (1/3)(2, 0, 1, 2, 0)'$.

Til að finna næsta einingavigur byrjum við á að ofanvarpa $\mathbf{b}$ á $\mathbf{v}_1$ og finna leifina. Nú eru $\mathbf{a}$ og $\mathbf{b}$ greinilega hornréttir og $\mathbf{b}$ og $\mathbf{v}_1$ það því líka svo ofanvarpið $\mathbf{w}_2$ er núll. Þar með er leifin bara $\mathbf{b}$ sjálft og stöðluð leif er því $\mathbf{v}_2 = \mathbf{r}_2/||\mathbf{r}_2|| = \mathbf{b}/||\mathbf{b}||$, þ.e. $\mathbf{v}_2 = (1/5)(0, 3, 0, 0, 4)'$.

Næst þarf að finna ofanvarp $\mathbf{c}$ á $sp\{\mathbf{v}_1, \mathbf{v}_2\}$, en það er gefið með

$$\mathbf{w}_3 = (\mathbf{c} \cdot \mathbf{v}_1)\mathbf{v}_1 + (\mathbf{c} \cdot \mathbf{v}_2)\mathbf{v}_2$$

Athugum að

$$(\mathbf{c} \cdot \mathbf{v}_1) = (0, 2, -3, 0, 1)' \cdot \frac{1}{3}(2, 0, 2, 1, 0)' = \frac{-6}{3} = -2$$

og

$$(\mathbf{c} \cdot \mathbf{v}_2) = (0, 2, -3, 0, 1)' \cdot \frac{1}{5}(0, 3, 0, 0, 4)' = \frac{10}{5} = 2$$

svo

$$\mathbf{w}_3 = -2\mathbf{v}_1 + 2\mathbf{v}_2 = \frac{-2}{3}(2, 0, 2, 1, 0)' + \frac{2}{5}(0, 3, 0, 0, 3)'$$

þ.e.

$$\mathbf{w}_3 = (\frac{-4}{3}, \frac{6}{5}, \frac{-4}{3}, \frac{-2}{3}, \frac{6}{5})' = \frac{1}{15}(-20, 18, -20, -10, 18)'.$$

Leifin eftir þetta ofanvarp er

$$\mathbf{r}_3 = \mathbf{c} - \mathbf{w}_3 = (0, 2, -3, 0, 1)' - \frac{1}{15}(-20, 18, -20, -10, 18)' = \frac{1}{15}(0+20, 30-18, -45+20, 0+10, 15-18)' = \frac{1}{15}(20, 12, -$$

sem hefur lengdina $||\mathbf{r}_3|| = ...$ og einingavigurinn er staðlaða leifin, þ.e.

$$\mathbf{v}_3 = \frac{\mathbf{r}_3}{||\mathbf{r}_3||} = ...$$

Síðan reiknast

$$\hat{\mathbf{y}} = \sum_i (\mathbf{y} \cdot \mathbf{v}_i)\mathbf{v}_i = \ldots$$

(Athugið að kvaðratrótin, sem virðist erfið viðureignar, hverfur þegar reiknað er $(\mathbf{y} \cdot \mathbf{v}_3)\mathbf{v}_3$, en dæmið er vissulega þunglamalegt).

(b) Við finnum $\hat{\boldsymbol{\beta}}$ með því að leysa normaljöfnurnar, $\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\hat{\mathbf{y}}$.

Hér er

$$\mathbf{X} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 2 \\ 2 & 0 & -3 \\ 1 & 0 & 0 \\ 0 & 4 & 1 \end{bmatrix}, \quad \text{og} \quad \mathbf{y} = \begin{bmatrix} 4 \\ 1 \\ 4 \\ 0 \\ 3 \end{bmatrix}$$

svo

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 9 & 0 & -6 \\ 0 & 25 & 10 \\ -6 & 10 & 14 \end{bmatrix}, \text{ og } \quad \mathbf{X}'\mathbf{y} = \begin{bmatrix} 16 \\ 15 \\ -7 \end{bmatrix}$$

svo normaljöfnurnar verða

$$\begin{aligned} 9b_1 \qquad\quad - 6b_3 &= 16 \\ 25b_2 + 10b_3 &= 15 \\ -6b_1 + 10b_2 + 14b_3 &= -7 \end{aligned}$$

Við athugum, að það er hægt að einfalda miðjöfnuna með því að deila í gegn með 5, margfalda svo með 2 og draga hana svo frá þeirri þriðju:

$$\begin{array}{rcl} -6b_1 + 10b_2 + 14b_3 &=& -7 \\ 10b_2 + 4b_3 &=& 6 \\ \hline -6b_1 + 10b_3 &=& -13 \end{array}$$

Þá erum við komin með tvær jöfnur í tveimur óþekktum

$$\begin{aligned} 9b_1 - 6b_3 &= 16 \\ -6b_1 + 10b_3 &= -13 \end{aligned}$$

sem má umrita til að geta lagt þær saman:

$$\begin{array}{rcl} 18b_1 - 12b_3 &=& 32 \\ -18b_1 + 30b_3 &=& -39 \\ \hline 18b_1 &=& -7 \end{array}$$

þ.e.a.s. $b_1 = -\frac{7}{18}$ og þá má nota ísetningu til að finna hin gildin og fá

$$\hat{\boldsymbol{\beta}} = (b_1, b_2, b_3)' = \frac{1}{270}(410, 204, -105)'$$

Nú má reikna beint $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \frac{1}{270}(820, 402, 1135, 410, 711)'$.

(c) SSE er alltaf skilgreint sem kvaðratsumma frávika gagna frá líkani og er reiknað á ýmsan hátt:

$$||\mathbf{y} - \hat{\mathbf{y}}||^2 = ||\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}||^2 = \sum_i (y_i - \hat{y}_i)^2 = \ldots = 3.648$$

(d) Við þekkjum einingagrunninn $\{\mathbf{v}_1\}$ fyrir rúmið $W = sp\{\mathbf{a}\}$ og $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ fyrir rúmið V. Við vitum því að við getum prófað núlltilgátuna $H_0 : \beta_2 =$

$\beta_3 = 0$ vs H_a : ekki H_0 með því að bera saman SSE reiknað sem frávik $\mathbf{y}$ frá ofanvarpinu á hvort rúm fyrir sig, þ.e. $SSE(R)$ og $SSE(F)$.

Við myndum reikna $SSE(F)$ eins og að ofan. Í því eru $n - p = 4 - 3$ frígráður. Síðan myndum við reikna $SSE(R)$ byggt á frávikinu milli $\mathbf{y}$ og $\mathbf{Z}\boldsymbol{\gamma}$ þar sem $\mathbf{Z} = \mathbf{a}$, en þetta ofanvarp er það sama og ofanvarp $\mathbf{y}$ á $\mathbf{v}_1$. Síðan fæst F-próf með því að taka hefðbundið hlutfall

$$\frac{(SSE(R) - SSE(F)) \,/\, (p - r)}{SSE(F) / \,(n - p)}$$

þar sem $r = 1$.

Athugum að hér er

$$\mathbf{Z}\boldsymbol{\gamma} = (\mathbf{y} \cdot \mathbf{v}_1)\mathbf{v}_1 = \frac{1}{9}(8, 0, 8, 0, 0)'$$

og þá fæst beint

$$SSE(R) = 29.358$$

og þá er allt komið sem þarf til að framkvæma F-prófið á tilgátunni

$$H_0 : X\boldsymbol{\beta} = Z\boldsymbol{\gamma}$$

sem er jafngild tilgátunni $H_0 : \beta_2 = \beta_3 = 0$ í grunnlíkaninu $E[y_i] = \beta_1 a_i + \beta_2 b_i + \beta_3 c_i$.

2. (10) Af 32 nemendum í námskeiði á háskólastigi voru 18 konur. Af þeim höfðu 16 lokið heimaverkefnum en 6 karlanna höfðu lokið sömu verkefnum.
(a) Er marktækur munur á skilahlutfalli kynjanna?
(b) Nemandi fullyrðir að það sé alveg upp og ofan hvort fólk skili verkefninu: "þetta gerir ekki nema annar hver nemandi". Setjið þessa fullyrðingu fram sem núlltilgátu og prófið hana.

Lausn: Byrjum á að setja dæmið snyrtilega upp:

	Karlar	Konur	Alls
Skilað	6	16	22
Ekki skilað	8	2	10
Alls	14	18	32

(a) Látum nú p_1 vera hlutfall karla sem skilað hafa heimaverkefnum og p_2 vera tilsvarandi hlutfall kvenna (áður en tilraunin er gerð). Látum ennfremur X vera hendinguna sem lýsir fjölda karla sem skila og Y tilsvarandi hendingu fyrir konur. Áður en tilraunin er gerð er eðlilegt að líta svo á að X og Y séu tvíkostadreifðar hendingar: $X \sim b(n_1, p_1)$ og $Y \sim b(n_2, p_2)$ með $n_1 = 14$ og $n_2 = 18$.

Leggjum nú fram núlltilgátuna $H_0 : p_1 = p_2$ vs $H_a : p_1 \neq p_2$. Ef núlltilgátan er rétt ætti

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

að vera útkoma úr $n(0,1)$ og við höfnum því H_0 ef reiknað gildi á z er þ.a. $|z| > z_{1-\alpha/2}$. Hér er

$$\hat{p} = \frac{x+y}{n_1+n_2}$$

Hér reiknast $z = -2.7869$ svo $|z| > 1.96 = z_{1-\alpha/2}$ (m.v. $\alpha = 0.05$) og við höfnum því H_0. Munurinn á skilum kynjanna er marktækur.

(b) Nú er ekki spurt um kynbundinn mun heldur aðeins hvort þetta geti alfarið stafað af tilviljun, m.ö.o. annar hver nemandi skilar verkefninu. Setjum því fram tilgáturnar

$$H_0 : p = \frac{1}{2} \text{ vs } H_a : p \neq \frac{1}{2}$$

Ef H_0 er rétt kemur

$$z = \frac{\hat{p} - \frac{1}{2}}{\sqrt{\frac{1}{2}(1-\frac{1}{2})/n}}$$

(nokkurnvegin) úr $n(0,1)$ og við höfnum því núlltilgátunni ef reiknað gildi á z er þ.a. $|z| > z_{1-\alpha/2}$.

Hér fáum við $z = 2.121$ svo við höfnum H_0. Fullyrðingin á ekki rétt á sér.

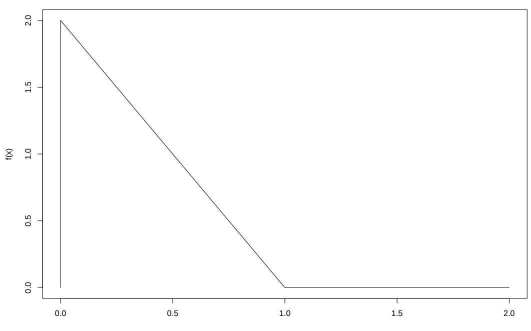
3. (10) Tiltekinni líkindadreifingu er lýst með samfelldu þéttifalli, f, sem er núll utan bilsins (0,1) en innan þessa bils minnkar þéttifallið línulega frá $f(0) = k$ niður í $f(1) = 0$.

(a) Hvað þarf k að vera og hver er jafna (stæða) f?

Látum nú X vera hendingu með þéttifallið f.
(b) Hvað er $P[X \geq 2]$?
(c) Hvað er $P[X \leq \frac{1}{2}]$?

Lausn: Hér hefði átt að nota [0,1] því vitanlega er átt við lokaða bilið, þ.e. núll og einn eru meðtalin.

(a) Þegar við byrjum á að teikna fallið sjáum við að ferill þess myndar þríhyrning með x-ásnum.



Þríhyrningur hefur flatarmál sem nemur hálfri hæð sinnum grunnlína. Hæðin hér er k og grunnlínan 1 þannig að flatarmálið er $k/2$. En flatarmálið er líka líkurnar sem verða að vera samtals 1 og því þarf að gilda $k = 2$.

Fallið er núll utan bilsins [0,1] en þar á milli má lýsa því með líkingu beinnar línu. Auljgóst er að hallinn er neikvæður og skurðpunkturinn við y-ásinn er í $k = 2$. Fallið minnkar frá 2 í 0 á bilinu frá 0 ti l1, þ.e. lækkar um 2 einingar við aukningu um eina einingu á x-ás og hefur því hallann -2 á því bili:

$$f(x) = \begin{cases} 0 & x < 0 \\ 2 - 2x & 0 \leq x \leq 1 \\ 0 & x > 0 \end{cases}$$

(b) Þéttifallið er núll fyrir ofan 1 og því er flatarmálið 0 fyrir ofan 2. Líkurnar eru því núll.

(c) Einfaldara er að reikna fyrst líkurnar fyrir ofan hálfan, en þar er um flatarmál rétthyrnds þríhyrnings að ræða. Það flatarmál er einn fjórði og líkurnar fyrir neðan hálfan eru því $\frac{3}{4}$.

4. (15) Endingartími ryksugu sem nota skal í skóla er uppgefinn 60 mánuðir (5 ár). Skólakerfið hefur notað svipaðar ryksugur áður og sannfært sig um að staðalfrávik endingartímans er 12 mánuðir.

Gefið ykkur forsendur sem þarf til að leysa eftirfarandi liði (og lýsið forsendunum). Gerið ráð fyrir að endingartíminn sé óháður notkun.

(a) Hverjar eru líkurnar á að tiltekin ryksuga endist a.m.k. 6 ár?

Hér þarf að gefa sér forsendu um líkindadreifingu. Gerum ráð fyrir að endingartíminn sé normaldreifður. Þá er hægt að líta á hendingu $Y \sim n(\mu = 5, \sigma^2 = 1^2)$ sem lýsingu á endingartímanum. Verkefnið er þá að reikna $P[Y \geq 6]$ sem felst í að staðla og fletta svo upp í töflu.

(b) Skóli kaupir tvær ryksugur og notar þær jöfnum höndum. Hverjar eru líkurnar á að skólinn verði ryksugulaus eftir 5 ár?

Skólinn verður ryksugulaus ef báðar ryksugurnar bila innan 5 ára. Meðaltalið í normaldreifingunni er 5 svo líkurnar eru $\frac{1}{2}$ á að tiltekin ryksuga bili á 5 árum. Líkur á að báðar bili verða þá $\frac{1}{2}\frac{1}{2} = \frac{1}{4}$ en hér höfum við gert ráð fyrir að bilanirnar séu óháðir atburðir og að líkurnar breytist ekkert hvort sem eru ein eða tvær í gangi.

(c) Grunnskólakerfið ætlar að samræma fjárfestingar sínar og fær tilboð í 100 ryksugur. Hverjar eru líkurnar á að a.m.k. 20 þeirra endist í a.m.k. 6 ár?

Látum Y og p vera eins og í (a)-lið. Ef X er fjöldi ryksuga sem endist í 6 ár, þá er X af gerðinni "fjöldi jákvæðra útkoma" í n tilraunum með líkum p á jákvæðri útkomu í hvert sinn. Hér þarf því að nota tvíkostadreifingu.

(d) Hverjar eru líkurnar á að engin ryksuga af 100 endist í 5 ár?

Aftur tvíkostadreifing, en nú er fjöldinn svo hár að nota þarf normaldreifingarnálgun.

(5): (20) Mælingar á hitastigi á þremur stöðum á landinu gáfu eftirfarandi niðurstöður:

Staður	A	B	C
n	10	5	14
$\bar{x}$	12.2	20.5	11.3
s	1.3	1.1	1.4

(a) Reiknið öryggismörk fyrir meðalhitann á stað A.
(b) Er marktækur munur á hitastiginu á stað A og B?
(c) Hvaða forsendur voru notaðar í hverjum lið?

Í lið (a) er á ferðinni meðaltal fyrir óþekkt hópmeðaltal. Staðalfrávík hópsins er vitanlega óþekkt en gefin er mæling á því. Gerum því ráð fyrir óháðum normaldreifðum mælingum og notum t-dreifingu.

Í lið (b) má gera ráð fyrir óháðum mælingum úr normaldreifingum frá A og B, gera ráð fyrir sama staðalfráviki og bera svo saman hópmeðaltölin með t-prófi fyrir tvo hópa.

Athugið að gögnin frá C eru hér ekki notuð, en það hefði ekki verið neitt galið að bæta matið á staðalfrávikinu með því að nota mælingarnar í C til að fá fleiri frígráður.

6. (15) (a) Teiknið mynd og nýtið hana til að finna þau gildi, x og y, sem hámarka $z = 2x + y$ með tilliti til

$$\begin{aligned} x + 3y &\geq 3 \\ x - y &\leq 2 \\ x + y &\leq 4 \\ x &\geq 0 \\ y &\geq 0 \end{aligned}$$

Munið að rökstyðja svarið.

(b) Hver verður lausnin ef á að lágmarka z?

Athugið, að myndin **verður** að vera rétt og það **verður** að afmarka rétt svæði. Það **verður** líka að **færa rök** fyrir lausninni, þannig að greinilega sé ekki um ágiskun að ræða. “Rök” þýðir annað hvort að teikna greinilega hvernig z breytist innan svæðisins eða að reikna gildin í öllum skurðpunktum sem falla innan svæðisins (ekki gleyma skurðpunktum við ása).

7. (30) Niðurstöður (y) tiltekinna prófana á sjúklingum voru fengnar með þremur mælitækjum ($m = 1, 2, 3$) og eftir mismunandi marga daga (x) í meðferð. Prófanirnar felast í því að mæla eftirstöðvar tiltekins efnis. Þær eru gerðar á log-kvarða og því er reiknað með að línulegt samband gildi við tíma á þeim kvarða, þ.e. milli y og x.

Þessi gögn voru sett inn í tölfræðipakka til að kanna, hvaða atriði hafa hugsanlega áhrif á mælingarnar. Skilgreint var líkanið $y = \mu + \alpha_m + \beta x + e$ og hér fyrir neðan kemur tilsvarandi úttak úr R.

Munið að rökstyðja svörin með tilvísun í tiltekna(r) tölu(r) í úttakinu. Ekki dugar t.d. að vísa almennt í "P-gildi" heldur þarf að skrifa sjálft gildið!

(a) Má sleppa x?
(b) Hvað útskýrir líkanið mikinn hluta breytileikans í gögnunum?
(c) Hvað er $SSTOT$?
(d) Lítið á núlltilgátuna $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = 0$. Yrði henni hafnað ef ekki væri tekið tillit til x?
(e) Útskýrir líkanið marktækan hluta breytileikans?
(f) Hvert er óvissumatið á β (þ.e. $\hat{\sigma}_{\hat{\beta}}$)?
(g) Hvað tapast margar frígráður við að hafa α_m í líkaninu?
(h) Hvað er MSE í líkaninu?
(i) Hver er spáin um prófniðurstöðu ($\hat{y}$) úr mælitæki 2 eftir 4 daga?

Gögnin:

```
> dat
  x  y m
1 2  8 1
2 2  4 1
3 3 21 1
4 4 10 2
5 4 12 2
6 7 17 2
7 7 28 3
8 9 31 3
9 9 36 3
```

Þessi spurning byggir einfaldlega á að geta lesið úttak úr tölfræðipakka. Athugasemdin að ofan segir allt sem þarf.

> **summary(lm(y ∼ m+x,data=dat))**

```
 Call:
lm(formula = y ~ m + x, data = dat)

Residuals:
      1       2       3       4       5       6       7       8       9
-1.9524 -5.9524  7.9048  0.1429  2.1429 -2.2857  0.5238 -2.7619  2.2381

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    3.667      4.777   0.768    0.477
m2            -6.381      5.964  -1.070    0.334
m3             1.810     10.612   0.171    0.871
x              3.143      1.633   1.924    0.112

Residual standard error: 4.99 on 5 degrees of freedom
Multiple R-Squared: 0.8751,     Adjusted R-squared: 0.8001
F-statistic: 11.67 on 3 and 5 DF,  p-value: 0.01073
```

> **anova(lm(y ∼ m+x,data=dat))**

```
> anova(lm(y~m+x,data=dat))
Analysis of Variance Table

Response: y
          Df Sum Sq Mean Sq F value   Pr(>F)
m          2 779.56  389.78 15.6567 0.007035 **
x          1  92.19   92.19  3.7031 0.112300
Residuals  5 124.48   24.90
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

> drop1(lm(y ~ m+x,data=dat),test="F")

```
> drop1(lm(y~m+x,data=dat),test="F")
Single term deletions

Model:
y ~ m + x
       Df Sum of Sq     RSS     AIC F value  Pr(F)
<none>              124.476  31.642
m       2   104.041 228.517  33.109  2.0896 0.2190
x       1    92.190 216.667  34.630  3.7031 0.1123
```

> anova(lm(y ~ 1,data=dat),lm(y ~ m+x,data=dat))

```
> anova(lm(y~1,data=dat),lm(y~m+x,data=dat))
Analysis of Variance Table

Model 1: y ~ 1
Model 2: y ~ m + x
  Res.Df    RSS Df Sum of Sq      F  Pr(>F)
1      8 996.22
2      5 124.48  3    871.75 11.672 0.01073 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```