

Háskóli Íslands	09.10.16 Línuleg algebra og tölfræði	Raunvísindadeild
Laugardagur	11. desember 2004	kl 09:00-12:00
Leyfileg hjálpargögn: Glósur, bækur og reiknivélur	Athugið að fartölvur eru ekki leyfðar.	Vægi dæma er gefið: 100 stig teljast full lausn, en alls eru stigin fleiri.

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 dálkvegurinn $\mathbf{y} = (4, 1, 4, 0, 3)'$. 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 frávíka (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} .

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.

3. (10) Tilttekinni 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 $P[X \leq \frac{1}{2}]$?

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?
- (b) Skóli kaupir tvær ryksugur og notar þær jöfnum höndum. Hverjar eru líkurnar á að skólinn verði ryksugulaus eftir 5 ár?
- (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?
- (d) Hverjar eru líkurnar á að engin ryksuga af 100 endist í 5 ár?

(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?

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 ?

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 tiltekens 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 tilsvareandi 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. $\hat{\sigma}_\beta$)?
- (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
```

```
> 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

```