

Eksami näidisülesanded

Ülesanne 1

Vaata alltoodud R'i väljundit ja pane kirja matemaatiline mudel meid huvitava sündmuse ($Y=1$) toimumise tõenäosuse leidmiseks. Leia ka meid huvitava sündmuse toimumise tõenäosuse väärtus juhul kui $x_1=2$ ja $x_2=\text{"rikas"}$.

```
> mudel=glm(y~x1+factor(x2), family=binomial())
> summary(mudel)

Call:
glm(formula = y ~ x1 + factor(x2), family = binomial())

Deviance Residuals:
    Min       1Q   Median       3Q      Max 
-1.7706  -1.0737  -0.7354   1.1177   1.8250 

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   0.005514   0.226984   0.024  0.98062
x1            0.462044   0.159305   2.900  0.00373 **
factor(x2)rikas 0.340206   0.359161   0.947  0.34353
factor(x2)vaene -0.663399   0.355215  -1.868  0.06182 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 276.76  on 199  degrees of freedom
Residual deviance: 261.15  on 196  degrees of freedom
AIC: 269.15
```

Vastused:

Arvutusvalem:

$P(Y=1 \mid X_1; X_2) = \dots\dots\dots$

Numbriline väärtus:

$P(Y=1 \mid X_1=2; X_2=\text{"rikas"}) = \dots\dots\dots$

Ülesanne 2

Metsikus läänes sooviti uurida imerohu mõju – kas imerohi aitab kauem elada? Vaadati imerohu tarvitajaid ja mittetarvitajaid ning uuriti, kas nad järgmise kümne aasta jooksul surid või ei surnud:

	imerohi	
	+	-
suri	101	200
ei surnud	2000	10100

Küsimus 1: Kuidas imerohu tarvitamine mõjutab inimeste surmašansse? Leia šansside suhe (OR) imerohu mõjule koos 95%-usaldusintervalliga!

Kuna imerohtu kippusid enam tarvitama vanad inimesed, siis koguti ka täpsemat informatsiooni imerohu tarvitajate vanuse kohta. All on toodud imerohu tarvitajate/mittetarvitajate vanusejaotus nii ellujääjate (kes 10 aasta jooksul ei surnud) kui ka ära surnud inimeste jaoks:

Ellujääjad:

	imerohi	
	+	-
Vana	1000	100
noor	1000	10000

Need kes surid:

	imerohi	
	+	-
Vana	100	100
Noor	1	100

Küsimus 2

Kas peale vanuse mõju arvessevõtmist on imerohul mõju? Milline see mõju on? Testi imerohu mõju kasutades Cochran-Mantel-Haenzeli testi, leia Cochran-Mantel-Haenzeli hinnang ühisele šansside suhtele (OR-le)!

Küsimus 3

Millised tunnused võiks lisaks vanusele imerohu mõju analüüsi praegu segada (kirjelda võimalikke segavaid faktoreid/tunnuseid – miks nad sinu arvates on segavad faktorid). Kuidas saaks ka sinu poolt kirjeldatud tunnuste mõju analüüsi käigus arvesse võtta?

Ülesanne 3

Õlle tarbimise (mitu pudelit õlut tudeng nädalas ära joob) ja soo ühisjaotust iseloomustab järgmine sagedustabel:

	<i>Õlu</i>		
<i>sugu</i>	<1	1–5	>5
naine	463	43	6
mees	68	49	31

Iseloomusta nende kahe tunnuse vahelise seose tugevust entroopiakordaja (Thail'i U ehk uncertainty coefficient) abil

$$U(\tilde{O}lu | Sugu) =$$

Leia ka tunnuse *Õlu* Shannoni entroopia (möödetuna bittides):

$$H(\tilde{O}lu) = ...$$

Ülesanne 4

Tuleta usaldusvahemik riskide vahele.

Ülesanne 5

Oletame, et Ronald Fisher oleks tema meelerahu häirinud daami maitsemeelt katsetanud 10 tassi abil – viide neist oleks kallatud enne piima ja siis lisatud teed, viide tassi oleks aga kõigepealt valatud teed ja siis alles lisatud piima. Juhul kui kehtiks nullhüpotees (daam, kelle maitsemeelt testitakse ei tee tegelikult vahet kumba koostisosa on tassi esimesena valatud), siis milline oleks tõenäosus, et katsetaja suudaks õigesti ära arvata kõigi 10 tassi kohta mida neisse on esimesena valatud? Milline tuleks kirjeldatud olukorras Fisheri täpse testi p-väärtus, kui testitav daam tõepoolest suudaks kõigi 10 tassi kohta teha õige otsuse?