

Examen November 2019 - SVAR

Opgave 1 - Sensorisk profilering af Øl (60%)

Til kvalitetskontrol af fødevarer benyttes sensoriske bedømmere. I dette studie er formålet 1) at sensorisk sammenligne forskellige produkter, samt 2) at sammenligne to paneler: Et generelt sensorisk panel (Panel1) og et panel af øl kendere (Panel2).

Til dette formål er to forskellige øl, **Heineken** og **Tuborg**, blevet vurderet af disse to paneler på ialt 11 sensoriske parametre. For begge paneler er denne profilering foretaget i triplicater.

NOTE: forkortelsen før attributen henfører til: A: aftertaste, B: basic taste, F: flavor og M: mouthfeel.

Dette giver et forsøgsdesign med ialt 32 prøver.

Designvariablene fortolkes som følger:

- **Panel** - paneltype (*Panel1*, *Panel2*)
- **Judge** - dommer nummer i de respektive paneler. Der er klart nok intet overlap mellem paneler.
- **Beer** - Øl type (*Tuborg*, *Heineken*)
- **Panel_Beer** - Kombinationen af panel og dommer.

Til at svare på spørgsmålene benyttes to versioner af dette dataset:

- **BeerMN** - som indeholder gennemsnit for hver dommer / øl, således at triplikaterne er fjernet.
- **BeerSD** - som indeholder triplikat spredningen for hver dommer / øl.

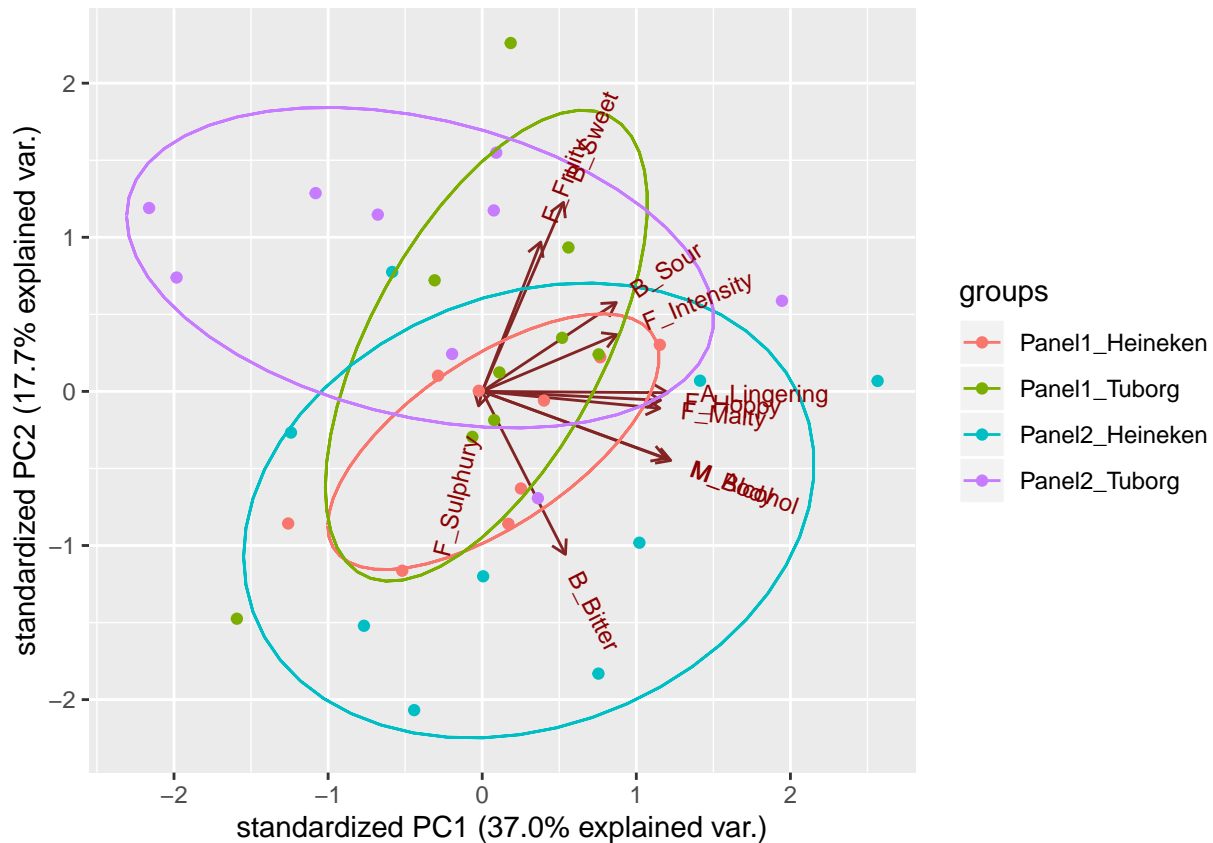
Det anbefales at indlæse disse data ved `load('Beer_Sensorik.RData')`, hvorved begge data.frames bliver tilgængelige i R. Som service er data også tilgængelig som en excel fil (`Beer_Sensorik.xlsx`) med to faner/sheets.

1

Lav en PCA baseret på de sensoriske attributter, og visualiser den således at både panel og øltype fremgår.

SVAR

```
library(ggplot2)
library(ggbiplot)
load('Beer_Sensorik.RData')
PCAm1 <- prcomp(BeerMN[,5:15], scale. = T)
ggbiplot(PCAm1, groups = BeerMN$Panel_Beer, ellipse = T)
```



2

Fortolk denne PCA model med henblik på sammenligning af sensoriske attributter; Hvilke der diskriminere (og ikke diskriminerer) øltyperne og om disse er generelle for begge paneler.

SVAR

Det registreres at **Panel2** noget mere konsekvent adskiller de to produkter (ifht **Panel1**). Heldigvis ses det dog at retningerne for adskildelse af de to produkter er ensartet, således at **Tuborg** scorer højere på PC2 (højere i F_Fruit, B_Sweet samt lavere i B_Bitter), men også M_Alcohol og M_Body synes at diskriminere. B_Sour, F_intensity, A_lingering, F_Malty og F_Hoppy er descriptorer, som er korreleret, beskriver PC1, men som ikke diskriminerer produkter og ej heller paneler specielt tydeligt. F_Sulphury, er en attribut, som er ikke passer specielt godt med de resterende attributer.

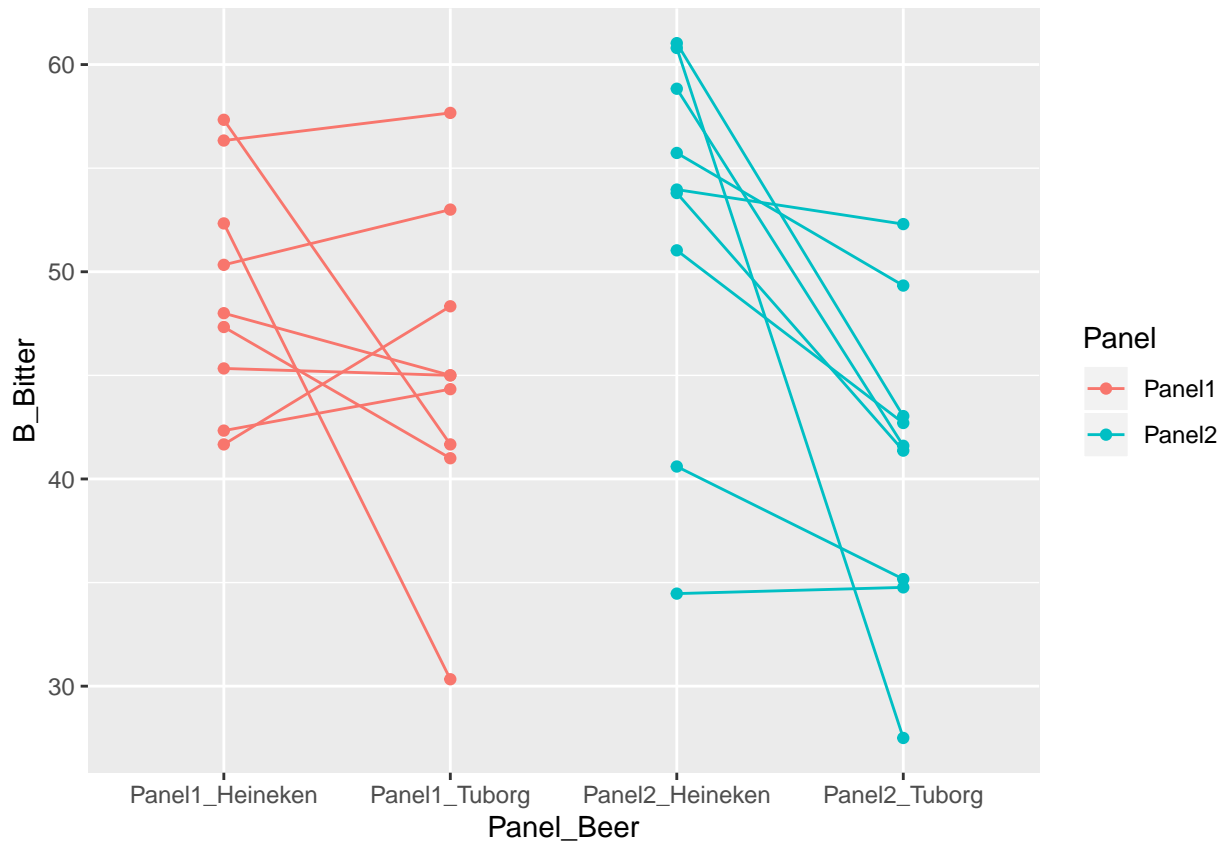
3

Udvælg **en** sensorisk attribut som kandidat til adskildelse øltyperne, og lav et tilhørende plot af denne hvor øltype og panel fremgår.

SVAR

Attributterne: F_Fruit, B_Sweet, B_Bitter samt M_Alcohol og M_Body er gode bud på enkelt variable til diskriminering af produkterne. Herunder et plot af B_Bitter.

```
ggplot(data = BeerMN, aes(Panel_Beer, B_Bitter, color = Panel, group = Judge)) +
  geom_point() +
  geom_line()
```



4

Baseret på denne attribut, estimer da to konfidensintervaller på forskellen mellem ø-ltyper; en for hvert panel.

SVAR

Dette kan gøres på to måder.

To parede t-tests

```
# panel 1
r1 <- t.test(data = BeerMN[BeerMN$Panel=='Panel1'], B_Bitter ~ Beer, paired = T)
# panel 2
r2 <- t.test(data = BeerMN[BeerMN$Panel=='Panel2'], B_Bitter ~ Beer, paired = T)
```

```
knitr::kable(rbind(data.frame(panel = 'Panel1', broom::tidy(r1)),
  data.frame(panel = 'Panel2', broom::tidy(r2))), digits = 3, caption = 'Konfidenseinterval for forskel me
```

Table 1: Konfidenseinterval for forskel mellem produkter i hvert panel

panel	estimate	statistic	p.value	parameter	conf.low	conf.high	method	alternative
Panel1	3.852	1.233	0.253	8	-3.351	11.055	Paired t-test	two.sided
Panel2	11.389	3.296	0.011	8	3.420	19.358	Paired t-test	two.sided

De to konfidensintervaller for forskellen mellem Heinicken og Tuborg på attributen B_Bitter viser at Heineken generelt er mere bitter end Tuborg.

For Panel2 er denne forskel omkring 11 med konfidensgrænser fra 3 til 19, og statistisk signifikant $p = 0.01$.

For Panel1 er forskellen mindre; 4 med konfidensgrænser overlappende 0; -3 til 11 og dermed heller ikke statistisk signifikant $p = 0.25$.

5

Vurder om denne attribut 1) er en god diskriminator til adskildelse af øltyper og 2) om den er lige stærk for de to paneler. Du kan eventuelt underbygge resultatet med tilhørende t-tests.

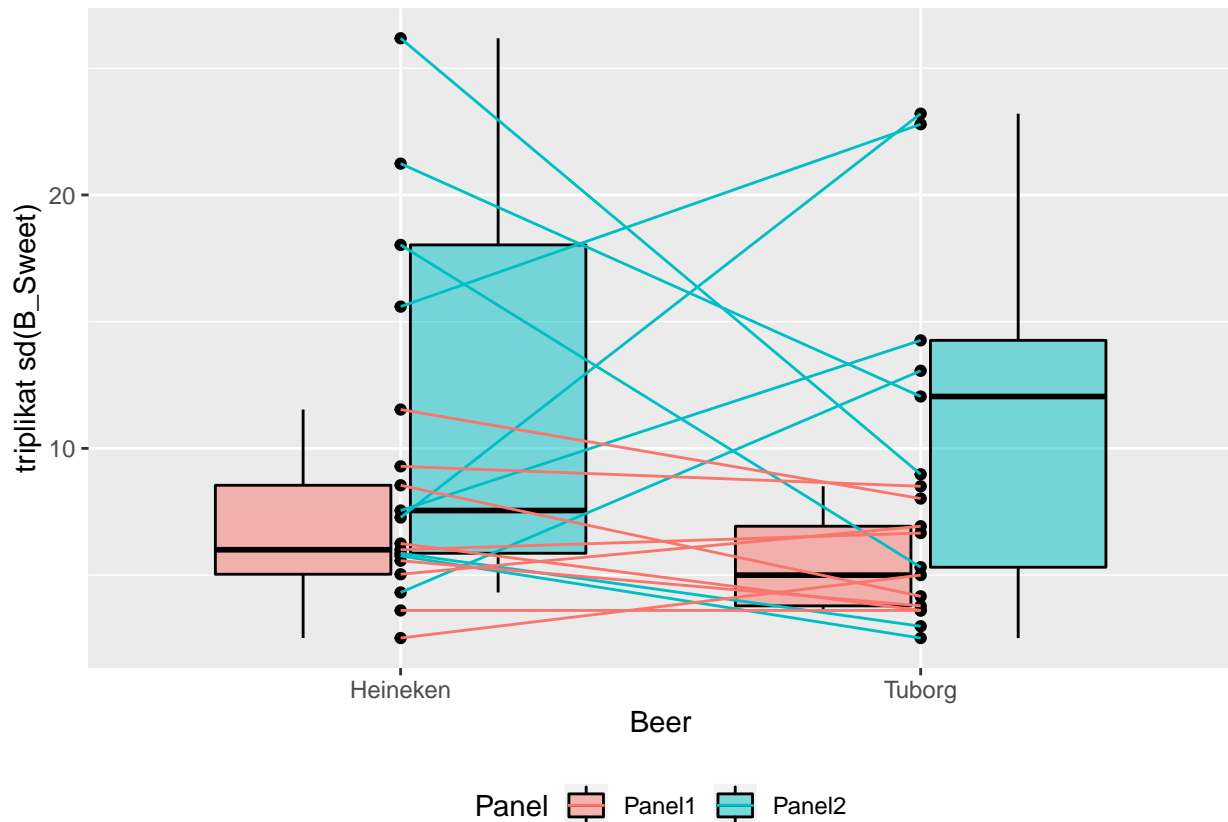
SVAR

I pkt 4 ses det at Panel 2 ville kunne benytte B_Bitter som diskriminerende attribut, mens Panel1 skal trænes mere for at kunne benytte B_Bitter. PCAen viser at det ikke rigtig synes muligt at udvælge en anden attribut som gør jobbet for Panel1.

I det ovenstående kigges der på **gennemsnit** af triplikater. Imidlertid indeholder triplikat bestemmelserne information om den enkelte dommers præcision for hver attribut, hver øltype. I det følgende benyttes **BeerSD** dataset.

Følgende kommando laver et plot over triplikat-spredningen for attributten B_Sweet som funktion af øltype, panel og de individuelle dommere.

```
ggplot(data = BeerSD, aes(Beer,B_Sweet,group = Judge, color = Panel)) +
  geom_boxplot(aes(group = Panel_Beer, fill = Panel), color = 'black', alpha = 0.5) +
  geom_point(color = 'black') +
  geom_line() +
  ylab('triplikat sd(B_Sweet)') +
  theme(legend.position = 'bottom')
```



6

Fortolk plottet med henblik på at sammenligne præcisionen mellem de to paneler.

SVAR

Y-aksen viser præcision hvor jo-lavere jo-bedre gælder. Det ses at Panel1 har markant lavere triplikat usikkerhed i sammenligning med Panel2. Plottet indikerer *ikke* at denne usikkerhed er forskellig mellem produkterne. I.e. det er lige svært / let at vurdere Heineken / Tuborg.

En statistisk analyse af disse data kan udføres med en ANOVA type model med tre forklarende variable. Koden herunder viser hvorledes denne model bygges.

```
mdl <- lm(data = BeerSD, B_Sweet ~ Beer + Panel + Judge)
```

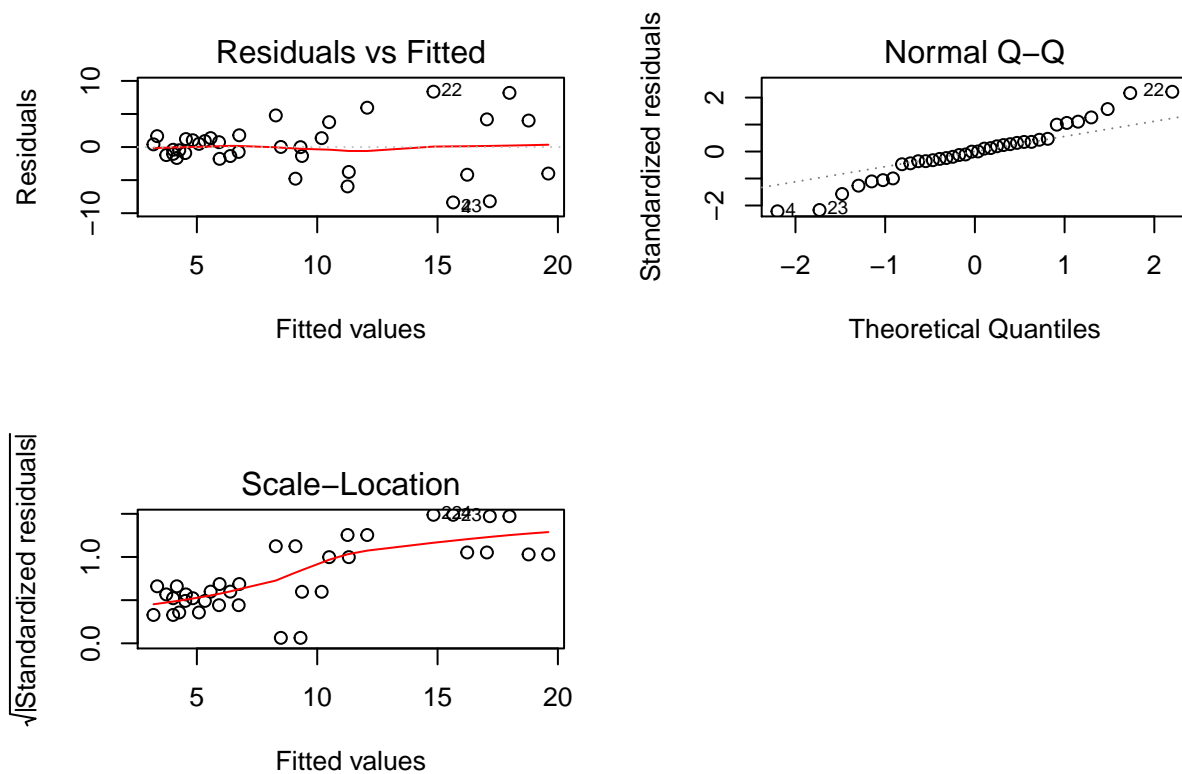
7

Tjek om modellens antagelser passer og foretag evt. justeringer.

SVAR

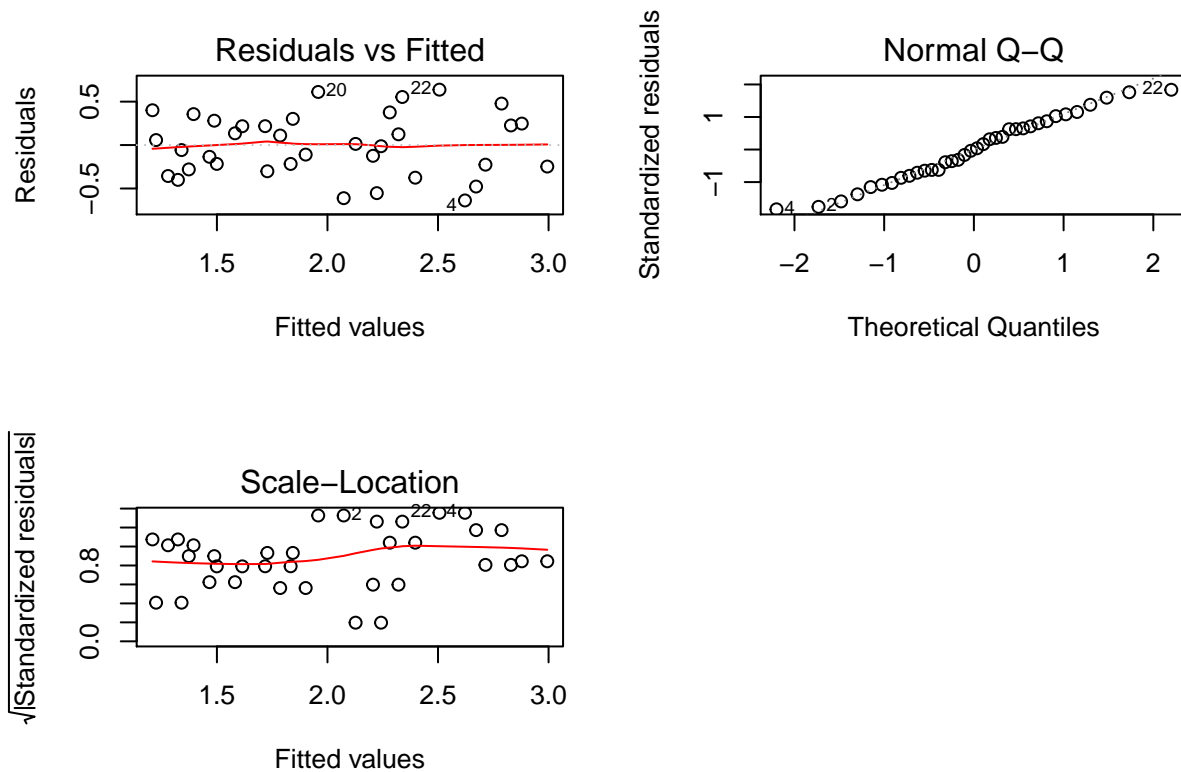
Modellen antager, at residualerne kommer fra samme normalfordeling og er uafhængige.

```
par(mfrow = c(2,2))
plot(mdl)
```



Det ses at der synes større varians på residual med højere forventede værdi (trompet i plot 1, samt haler i top og bund af qqplot - plot 2). Derfor justeres modellen med `log()`. Justeres med `sqrt()` afhjælder det noget af problemet, dog ikke fuldstændig, hvorfor `log` vælges.

```
mdl2 <- lm(data = BeerSD, log(B_Sweet) ~ Beer + Panel + Judge)
par(mfrow = c(2,2))
plot(mdl2)
```



8

Lav en ANOVA evaluering af (den evt. justerede) model og fortolk denne, specielt med henblik på at sammenligne præcisionen af de to paneller. HINT: Passer den umiddelbare observation i plottet med den statistiske tolkning?

SVAR

```
anova(md12)
```

```
## Analysis of Variance Table
##
## Response: log(B_Sweet)
##          Df Sum Sq Mean Sq F value    Pr(>F)
## Beer      1  0.1196  0.11959    0.4686  0.502878
## Panel     1  2.6922  2.69221   10.5480  0.004735 **
## Judge    16  7.2735  0.45459    1.7811  0.124097
## Residuals 17  4.3390  0.25523
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Det ses at faktoren *Panel* er signifikant ($p = 0.004$) hvilket passer med observationen i plottet; at Panel1 har lavere triplakvariation end Panel2. Yderligere, er faktoren *Beer* ikke signifikant ($p = 0.6$), hvilket også passer med observationen i plottet, hvor usikkerheden synes uafhængig af øltype. Faktoren *Judge* er ikke signifikant, hvilket indikerer at en dommer med en høj triplikatvariation for et produkt, ikke også ligger med høj for det andet produkt. Dette ses i plottet ved at linierne krydser.

En Note: Det er interessant, at det panel som kan adskille produkterne (jvf pkt 1.1 - 1.5) (*Panel2*) er det panel med den dårligste triplikat variation. Dette er i umiddelbar modstrid, men skyldes nok, at *Panel1* bruger langt mindre af skalaen og dermed har sværere ved at diskriminere, men også opnår bedre (lavere) triplikat usikkerhed.

Opgave 2 - Triangeltest af øl (40%)

Sensorisk evaluering af produkter umiddelbart før frigivelse til salg, foretages af et trænet panel. Som en del af den interne akkreditering udfører sådanne paneller blind-tests i form af en adskildelse af to forskellige produkter der dog er valgt til at være meget ens. Denne blind-test udføres via en triangeltest.

Et øl-trænet sensorisk panel på $n = 12$ dommere har udført en blind-test mellem Carlsberg og Tuborg. Af de $n = 12$ dommere var $x = 10$ istand til at diskriminere de to produkter. % Data

1

Vurder panelets performance i forhold til tilfældige svar ved at formulere **model** for data, opstille **null-** og **alternativ** hypotese samt en **statistiks test**.

SVAR

Lad X være antallet af korrekte svar.

Model for antal korrekte svar er: $X \sim \text{bin}(n = 12, p)$

Hvis dommerne gætter er ssh for korrekt svar $1/3$. Det giver følgende **hypotester**.

$H_0 : p = p_0 = 1/3$ og $H_A : p > p_0$

10 korrekte svar er noget mere end forventet ($n * p_0 = 12 * 1/3 = 4$), men er det usansynligt meget mere??

$p = P(X \geq 10) = 1 - P(X \leq 10 - 1) = 0.0005$

Givet testen sluttet det at dommerne godt kan diskriminere produkterne.

```
n <- 12
x <- 10
p0 <- 1/3
1 - pbinom(x-1,n,p0)
```

```
## [1] 0.0005438045
```

2

Estimer modelens parameter samt angiv et konfidensinterval for denne.

SVAR

Da vi *IKKE* kan antage H_0 gives nu det bedste bud på ssh parameteren:

$$\hat{p} = x/n$$

og tilhørende konfidensinterval

$$p_{low} = \hat{p} - z * s_p \text{ til } p_{high} = \hat{p} + z * s_p.$$

Table 2: Panel størrelse (n), antal korrekte- (x) og forkerte- ($n - x$) svar for 3 sensoriske øl-paneler

	n	x	$n + x$
Panel 1	20	12	8
Panel 2	18	14	4
Panel 3	27	26	1

Hvor $s_p = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

```
phat <- x/n
sp <- sqrt( phat*(1-phat) / n )
z <- 1.96
p_low <- phat - z*sp
p_high <- phat + z*sp
c(p_low,phat,p_high)
```

```
## [1] 0.6224709 0.8333333 1.0441958
```

Således er den mest sandsynlige frekvens 0.83 (men værdier fra 0.62 til 1 er realistiske).

Bemærk at øvre grænse afrundes til 1, da p ikke kan være større end 1 (hvh mindre end 0).

Som en del af en brede generel akkreditering, sammenlignes 3 forskellige paneller med samme metode.

Givet at alle panelerne er trænede, er det sekundært om de enkelte paneler er bedre en tilfældig diskriminering. Derimod er det interessant om 1) Panelerne performer *forskelligt* samt 2) hvilket panel som er *bedst*.

Herunder er angivet triangeltestresultater for tre paneler.

3

Baseret på data fra de 3 paneler, formuler tilsvarende 3 modeller, og en H_0 - og H_A hypotese til sammenligning af panelerne.

SVAR

Hver panel følger en binomial fordeling:

$X_i \sim \text{bin}(n_i, p_i)$ hvor $i = (\text{Panel1}, \text{Panel2}, \text{Panel3})$

Hypotesen er at panelerne er lige gode: $H_0 : p_{\text{panel1}} = p_{\text{panel2}} = p_{\text{panel3}}$ med alternativet at *mindst eet panel skilder sig ud*

4

Test denne hypotese og kommenter på om panelerne kan antages at have forskellig (diskriminerende) performance.

SVAR

Dette er test for antalstabeller:

```
M <- matrix(c(12,8,14,4,26,1),nrow = 2 )
chisq.test(M)
```

```
##
## Pearson's Chi-squared test
##
## data: M
## X-squared = 9.537, df = 2, p-value = 0.008493
```

H_0 må forkastes, og det sluttet at panelerne er forskellige ifht diskriminering.

5

Giv et bud på om alle 3 paneler kan antages forskellige, eller om der er et panel som er afvigende (bedre eller dårligere end de to andre).

SVAR

Frekvensen af korrekte svar beregnes

```
ph <- M[1,] / colSums(M)
ph
```

```
## [1] 0.6000000 0.7777778 0.9629630
```

Det ses at især er *Panel 3* med 26 korrekte (ud af 27) overlegne, mens *Panel 1* skal træne noget mere.

Det er uklart givet frekvenserne om alle tre grupper er forskellige. Derfor laves der tilhørende 3 konfidensintervaller:

```
z <- 1.96
n <- colSums(M)
sp <- sqrt(ph*(1-ph) / n)

plow <- ph - z*sp
phigh <- ph + z*sp

knitr::kable(data.frame(Panel = c(1,2,3),plow,ph,phigh),
              digits = 2,
              caption = 'p hat og CI for de tre paneler')
```

Table 3: p hat og CI for de tre paneler

Panel	plow	ph	phigh
1	0.39	0.60	0.81
2	0.59	0.78	0.97
3	0.89	0.96	1.03

Det sluttet at Panel 1 og Panel 3 er forskellige, mens det ikke rigtig er muligt at adskilde Panel 2 fra hverken Panel 1 og Panel 3.

OBS: Der gives fuldt point for at vurdere panel-forskellighed baseret på de observerede frekvenser (\hat{p}).

Beskrivelse af triangeltest

I en triangeltest præsenteres hver dommer for tre produkter; to ens (**A,A**) og en afvigende (**B**) og skal udpege det afvigende produkt **B**.