

1. Statistisk analyse af et datasæt

1.1 Datapræsentation

1.

Opgavenummer 39

2.

Tabel 1:

Karakteristika	Deltagere (125)
Farveændring* Median Inter quartile range	24,21 12,33-33,90
Farveændring* Gennemsnit Spredning	23,13488 13,14096
Koncentration af hydrogenperoxid - % Median Inter quartile range	7,87 4,25-11,32
Koncentration af hydrogenperoxid - % Gennemsnit Spredning	7,84522 4,106054
Brug af lys under behandling - antal Ja Nej	78 (62,4) 47 (37,6)
Ætsning på gummer efter behandling - antal (%) Ja Nej	67 (53,6) 58 (46,4)
Tilfredshed af produktet - antal (%) Ville ikke bruge produktet igen Ville måske bruge produktet igen Ville bruge produktet igen	47 (37,6) 46 (36,8) 32 (25,6)

*Farveændring er ændringen af farven på tænderne før og efter behandling. Farven bedømmes ud fra en "lightness" skala, hvor værdien 0=sort og 100=hvid.

Da der opgives to numeriske variable (farveændring og koncentration), har jeg valgt at vise både gennemsnit og spredning, og median og IQR, idet de tilsammen giver et tæt referat af deltagernes karakteristika. Gennemsnit og spredning benyttes på symmetriske fordelinger, og da mit datasæt er tilnærmelsesvist symmetrisk, ligner gennemsnittet og medianen hinanden meget, mens spredning og IQR også ser nogenlunde ens ud. Jeg valgte at benytte begge for at vise netop at det var symmetrisk, og for at tage højde for hvad man personligt bedst kan lide at betragte og analysere på. Gennemsnit og median giver et hurtigt indblik i den koncentration og farveændring, som dataen centrerer om, mens spredning og IQR giver et overskueligt billede af hvordan resten af dataen forholder sig til disse midtværdier.

Derudover er der opgivet tre kategoriske variable (brug af lys, ætsning på gummer og tilfredshed), hvoraf de to første er binære, mens den sidste har tre grupper. Da populationen, der deltog i undersøgelsen, er på 125, gav det mening for mig at inkludere hvor mange af den totale population, der befandt sig i hver af grupperne. Derudover inkluderede jeg en procent, for igen at danne et overblik. Dog var populationsantallet så nært 100 at det næsten var irrelevant, men jeg valgte at tage det med, fordi det kan være rart at se på, hvis man hurtigt ser ned over tabellen, og ikke bruger så meget tid på at sætte sig ind i undersøgelsen.

1.2 Overordnet effekt af tandblegningsproduktet

1.

Jeg har valgt at lave t-test for én population, idet der kun betragtes én afhængig variabel, og ingen uafhængig. Den afhængige variabel er altså farveændringen, som jo er en difference mellem start- og slutfarven på tænderne. T-test foretages ved sammenligning af individers differencer i en population. Farveændringen afhænger af brug af produktet, og da alle i undersøgelsen har benyttet blegeproduktet, ses der altså ikke på nogen uafhængig variabel, som dette tilfælde ville være om man har brugt produktet eller ej, hvor længe man har brugt produktet eller lignende. En forudsætning for at kunne lave t-test er at dataen skal være nogenlunde normalfordelt. Der tjekkes for normalfordeling, før der udføres t-test ved at lave et histogram i RStudio, og det ses at dataen er nogenlunde normalfordelt.

Efter at have lavet t-test i RStudio blev det oplyst at den gennemsnitlige farveændring for populationen er 23,13.

Når der udføres t-test i RStudio udregner programmet også p-værdi for følgende nulhypotese:

H_0 : "Det sande gennemsnit er lig med 0".

Ud fra denne nulhypotese kan det altså vises om produktet har været effektivt, idet et positivt gennemsnit vil vise at produktet har haft en positiv effekt på populationens tænder. Nulhypotesen kan derefter be- eller afkræftes ved brug af p-værdi.

2.

RStudio oplyser $p < 2,2 \cdot 10^{-16}$. Da denne p-værdi er meget mindre end 0,05, svarende til 5%, forkastes nulhypotesen. Det er altså meget sandsynligt at det sande gennemsnit ikke er lig med 0, og tandblegningsmidlet altså har haft en effekt på populationen. P-værdien siger dog ikke noget om effektens fortegn, blot at gennemsnittet ikke er 0.

3.

Når der udføres t-test i RStudio, oplyses der også et 95% konfidensinterval. I dette tilfælde er det [20,81 ; 25,46]. Jeg er altså 95% sikker på at intervallet [20,81 ; 25,46] indeholder det sande gennemsnit for farveændringen af populationens tænder efter at have benyttet tandblegningsproduktet.

Dette betyder altså at tandblegningsproduktet har en positiv effekt på gennemsnittet af populationen, og at det er meget sandsynligt at den gennemsnitlige ændring af farven på deres tænder ligger mellem 20,81 og 25,46.

4.

For at udregne et referenceinterval tager man gennemsnitsværdien for variabelen og trækker 2-spredningen fra for at finde den nedre grænse, og tager gennemsnitsværdien og lægger 2-spredning til for at finde den øvre.

Referenceintervallet for populationens farveændring er [-3,15 ; 49,4].

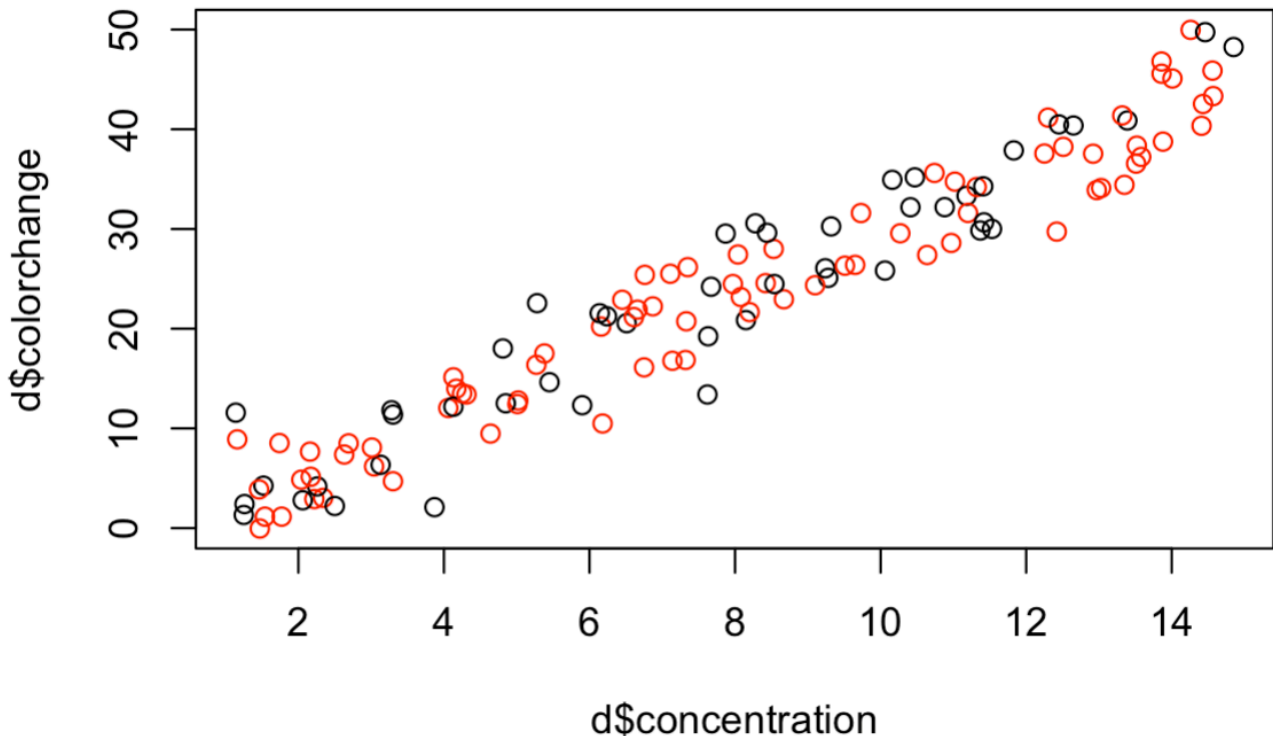
Et 95% referenceinterval betyder at 95% af dataen ligger i intervallet. Altså ligger 95% af populationens farveændringer mellem værdierne -3,15 og 49,4. For at dette er helt sandt, skal dataen være symmetrisk fordelt. Da den laveste værdi for farveændring i mit datasæt er -0,04, må det altså betyde at datasættet ikke er helt symmetrisk, da intervallet rækker lidt udover dette. Det indeholder dog to værdier der er over 49,4, som er den øvre grænse for intervallet, så det passer stadig nogenlunde med at 95% af dataen ligger i intervallet, selvom det i grunden er lidt mere end 95% der ligger i intervallet.

Dette referenceinterval viser at tandblegningsproduktet primært dækker positive resultater. Da et interval ikke i sig selv fortæller noget om fordelingen af dataen, vil man ikke normalt kunne konkludere at størstedelen af populationen oplever en positiv effekt, idet mange målinger for eksempel kunne ligge under 0, og der så bare var en enkelt som var langt større end 0. Men idet man bruger spredning og gennemsnit til at udregne dette interval, kan man godt konkludere at langt størstedelen af populationen oplever at deres tænder bliver hvidere efter brug af produktet. Altså fortæller dette referenceinterval at tandblegningsproduktet har en positiv effekt på langt størstedelen af populationen.

5.

Det 95% konfidensinterval er [20,81 ; 25,46] og det 95% referenceinterval er [-3,15 ; 49,4].

I langt de fleste tilfælde vil referenceintervallet være meget bredere end konfidensintervallet, idet referenceintervallet afspejler laveste og højeste værdi (ideelt set), mens konfidensintervallet afspejler gennemsnittet af værdierne. Idet referenceintervallet indeholder spredning og gennemsnit, mens konfidensintervallet indeholder standardfejlen, vil forskellen på de to yderværdier være forskellen mellem standardfejl og spredning. Hvis spredningen derfor er lig med standardfejlen, vil de to intervaller være ens. Dette er dog ikke sandsynligt.



1.3 Effekt af koncentration og lys på ændring i tandfarve

1.

For at undersøge hvordan tandfarven afhænger af koncentrationen af hydrogenperoxid samt brug af lys, benyttes multipel lineær regression. Tandfarven er altså den afhængige variabel, som vil ses op ad y-aksen. Koncentrationen vil være uafhængig i forhold til tandfarven, den vil ses ud ad x-aksen. Da sammenhængen mellem farveændring og koncentration afhænger af brug af lys, og da brug af lys er en kategorisk variabel, farves de datapunkter, hvor der er brugt lys, røde, mens dem, hvor lys ikke er brugt, farves sorte. Hvis det ikke skulle undersøges, hvad brug af lys havde af effekt på farveændringen, ville der bare være tale om almindelig lineær regression, men da jeg også er interesseret i at undersøge effekten af brug af lys (interaktionseffekten) udover sammenhængen mellem koncentration og farveændring, er multipel lineær regression den mest brugbare og informative statistiske model.

Modellen tegnes i RStudio:

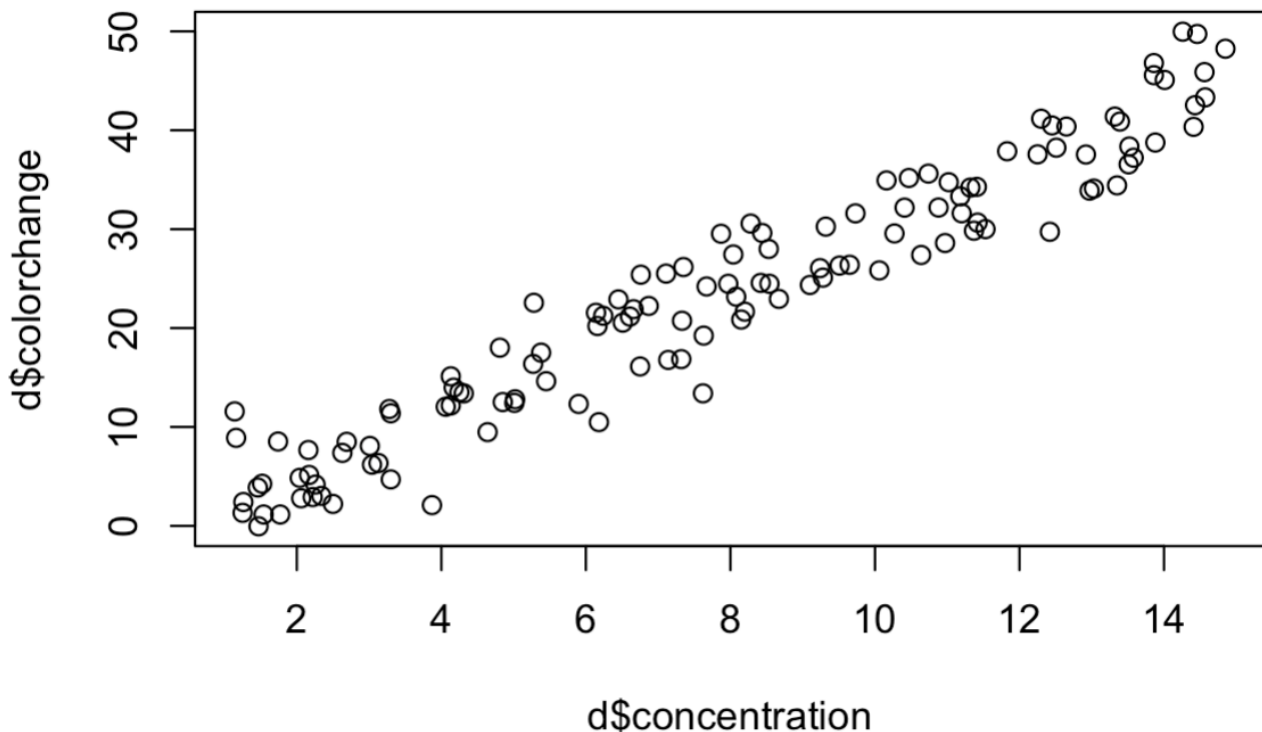
De røde prikker viser de data der har benyttet lys under behandlingen, mens de sorte viser dem, der ikke benyttede lys.

2.

Når man laver en summary i RStudio af modellen, får man oplyst estimater, som bruges til at opstille ligningerne:

Ligningen for den estimerede sammenhæng ved brug af lys er altså:
 $\text{farveændring} = 3,19 \cdot \text{koncentration} - 1,46$

Ligningen for den estimerede sammenhæng uden brug af lys er:
 $\text{farveændring} = (3,19 + (-0,16)) \cdot \text{koncentration} + (-1,46 + 0,56) \Leftrightarrow$
 $\text{farveændring} = 3,03 \cdot \text{koncentration} - 0,9$



3.

Nulhypotesen lyder som følger:

H_0 : "Der er ingen forskel på tandfarvning når der bruges lys og når der ikke bruges lys."

P-værdien aflæses i RStudio til $p=0,35$. Da denne p-værdi er meget større end 0,05, svarende til 5%, derfor forkastes den ikke. Altså er der ingen forskel på farveændringen af tænderne hos de deltagere som benyttede lys og de deltagere som ikke benyttede lys.

4.

Idet det lige er blevet vist, at der ikke er forskel på farveændringen af tænderne hos deltagerne, der brugte lys, og dem som ikke gjorde, kan man, i stedet for at lave multipel regression med tre variable, forsimpler ved at se bort fra lysbrug, og nøjes med en simpel lineær regression med de to variable koncentration og farveændring.

Idet farveændringen stadig afhænger af koncentration af hydrogenperoxid, bevares farveændring som den afhængige variabel, som ses op ad y-aksen, mens koncentration fortsat er uafhængig, og ses ud ad x-aksen.

5.

Da jeg nu blot ser på en simpel lineær regression, kan jeg opskrive en model for den og køre en summary i RStudio. Der får jeg de forsimplerede estimater, som benyttes til at opskrive ligningen for linjen, der viser sammenhængen mellem koncentration og farveændring:

$\text{farveændring} = 3,08 \cdot \text{koncentration} - 1,04$

6.

Parameteren "intercept", som vises i summary i RStudio, er et udtryk for linjens skæring med y-aksen. I mit tilfælde er estimeret, at linjen skærer y-aksen i punktet $(0; -1,04)$. Dette betyder at der, til koncentrationen 0% hydrogenperoxid, vil være en negativ ændring af tandfarven, ifølge denne simple model opstillet ud fra datasættet. Altså vil man, ifølge estimeret, sænke sin tandfarve på "lightness" skalaen med 1,04 hvis man ikke benytter produktet. Dette er dog igen et estimat, og det er sandsynligvis ikke sådan i virkeligheden, når man ikke ser på en større tidsændring.

7.

Følgende nulhypotese er testet for:

H_0 : "Farveændringen til koncentrationen 0% hydrogenperoxid er lig med 0".

P-værdien der oplyses i RStudio er $p=0,14$. Da denne p-værdi er meget større end 0,05, forkaster vi ikke nulhypotesen. Altså er der ingen farveændring ved en koncentration på 0% hydrogenperoxid, svarende til man ikke benytter tandblegningsproduktet.

Dette giver fin mening, idet man ikke kan forvente en blegning af tænderne hvis man ikke benytter nogle midler. Det er relevant i forhold til den videnskabelige problemstilling, da man ser at

farveændringen afhænger af koncentrationen af hydrogenperoxid, og dermed kan bekræfte af tandblegningsproduktet har en effekt.

8.

Når man undersøger konfidensinterval på 95% i RStudio, får man oplyst følgende intervaller for mit datasæt:

95% konfidensinterval for farveændringen pr 1 procentpoint hydrogenperoxid:
[2,93 ; 3,24]

95% konfidensinterval for farveændringen til koncentrationen 0% hydrogenperoxid (start ændringen, eller ændringen uden brug af produktet):
[-2,40 ; 0,33]

Altså er jeg 95% sikker på at den sande farveændring, hver gang man behandler med 1 procentpoint hydrogenperoxid mere, ligger i intervallet [2,93 ; 3,24] i populationen. Derudover er jeg 95% sikker på at den sande farveændring ved en koncentration på 0% hydrogenperoxid befinder sig i intervallet [-2,40 ; 0,33] i populationen.

9.

Ud fra modellen forventes en farveændring på 35,9 ud fra "lightness" skalaen ved behandling med en hydrogenperoxid koncentration på 12%. Bemærk at dette er et gennemsnit, idet ændringen prædikeres ud fra en estimeret linje.

10.

Farveændringen til koncentrationen 10% er gennemsnitligt 29,8. Dette, samt spredningen, aflæses i RStudio, og referenceintervallet opstilles.

95% referenceinterval:

[22,6 ; 36,9]

Altså forventes 95% af dataen ved brug af koncentration 10% hydrogenperoxid at have en farveændring mellem 22,6 og 36,9.

1.4 Sammenhæng mellem ætsning og brugertilfredshed

1.

Jeg vælger at benytte χ^2 -test, idet brugertilfredshed afhænger af ætsningen, og brugertilfredsheden er en kategorisk variabel med tre grupper, mens ætsningen er binær. Denne type test anvendes for at påvise hvordan en kategorisk variabel afhænger af en anden kategorisk variabel. Man vil altså kunne betragte dataens interaktion mellem de forskellige kategorier.

2.

Tilfredshed og ætsning	Vil ikke bruge produktet igen	Vil måske bruge produktet igen	Vil bruge produktet igen
Ingen ætsning	8	20	30
Ætsning	39	26	2

H_0 : "Gruppernes (ætsning og ingen ætsning) gennemsnit er ens". Altså vil jeg gerne se om ætsning har en indflydelse på populationens tilfredshedsfølelse. Hvis gennemsnitne i gruppen med ætsning var den samme som i gruppen uden ætsning, ville ætsningen ikke have nogen indflydelse på tilfredsheden. Dette er dog ikke tilfældet, idet p-værdien tilhørende denne nulhypotese er $p=1,44 \cdot 10^{-10}$, hvilket er meget mindre end 0,05. Altså forkastes nulhypotesen, og det konkluderes af gruppen, som oplevede ætsning, ikke har samme gennemsnit som gruppen, der ikke oplevede ætsning. Altså har ætsning en påvirkning på populationens tilfredshedsfølelse.

3.

Sandsynligheden for at ville bruge produktet igen givet man har oplevet ætsninger er altså $2/67 = 0,0299 = 2,99\%$.

RStudio bruges til at finde et 95% konfidensinterval tilhørende denne sandsynlighed:
[0,005 ; 0,113]

Altså er jeg 95% sikker på at den sande sandsynlighed for at benytte produktet igen, givet man har oplevet ætsninger, ligger i intervallet [0,005 ; 0,113].

4.

Den relative risiko er ikke lig med 1, altså er de to risici ikke ens. Da den relative risiko her er 4,87, vil det sige at hver gang én person, af dem der ikke vil bruge produktet igen, ikke oplever ætsninger, vil 4,87 personer, af dem der ikke vil bruge produktet igen, opleve ætsninger. Altså er det langt flere i kategorien, af dem som ikke vil bruge produktet igen, som har oplevet ætsninger, som dem der ikke har oplevet ætsninger.

5.

For at kunne finde en odds ratio, må man skulle sammenligne flere grupper. Derfor opstilles nu følgende grupper:

	Vil måske bruge produktet igen	Vil måske ikke bruge produktet igen (dem der definitivt ikke vil + dem der definitivt vil)
Ingen ætsning	20	38
Ætsning	26	41

De to gruppers odds udregnes, og odds ratioen udregnes som 1,20.

Da odds ratioen ikke er lig med 0, er grupperne ikke ens, og forholdet mellem dem der oplevede ætsninger og dem som ikke gjorde i de to grupper, dem der måske vil bruge produktet igen, samt den sammenslåede gruppe af dem der definitivt enten vil eller ikke vil bruge produktet igen, er ikke ens.

Altså er forholdet mellem personer der har ætsninger og personer der ikke oplevede ætsninger i gruppen, med personer der måske ville bruge produktet igen, 1,20 gange større, end forholdet mellem personer der har ætsninger og personer der ikke oplevede ætsninger i den anden gruppe.

1.5 Sammenhæng mellem koncentration og ætsning

1.

Da jeg er interesseret i at undersøge forholdet mellem koncentration af hydrogenperoxid og ætsning af gummerne, vælger jeg at benytte mig af logistisk regression. Det gør jeg fordi ætsning afhænger af koncentrationen her, og ætsning er en binær, kategorisk variabel, mens koncentration, som er en uafhængig variabel, er kontinuer. Logistisk regression benyttes til at se på hvordan sandsynligheden for ætsning af gummerne afhænger af koncentrationen af hydrogenperoxid.

2.

Ud fra de estimerede værdier jeg finder i RStudio, kan udtrykket for den logistiske regression opstilles:

$\log \text{Odds} = 0,63 \cdot \text{koncentration} - 4,63$

3.

Odds ratioen findes i RStudio og viser sig at være 1,88. Da denne ikke er 1, må koncentrationen af hydrogenperoxid altså have en effekt på om ens gummer ætser eller ej.

95% konfidensinterval:

[1,56 ; 2,39]

Jeg er altså 95% sikker på at den sande odds ratio for ætsning af gummerne ligger i intervallet [1,56 ; 2,39] for populationen.

4.

Jeg er uenig, idet odds ratio $\neq 1$, altså har koncentrationen en effekt på ætsningen.

H_0 : "Koncentrationen af hydrogenperoxid har ikke en effekt på ætsning".

I RStudio fås $p=3,76 \cdot 10^{-9}$, hvilket er meget mindre end 0,05. Altså forkastes nulhypotesen, og igen bekræftes det, at koncentrationen har en effekt på ætsning af gummerne.

5.

Ved brug af modellen prædikteres det, at oddsene for at få ætsning på gummerne ved brug af 6% hydrogenperoxid koncentration, er 0,43. Altså er der, ud fra modellen, en 43% sandsynlighed for at få ætsede gummer ved brug af denne koncentration.

6.

Sandsynligheden for at få ætsninger på gummerne ved en koncentration op 6% prædikteres til at være 0,30. Altså er der en 30% sandsynlighed for at have ætsninger hvis man benytter en hydrogenperoxid koncentration på 6%.

7.

Jeg benytter CAS til at solve for x i følgende ligning:

$$0,8 = \frac{e^{-4,63+0,63 \cdot x}}{1 + e^{-4,63+0,63 \cdot x}}$$

$$\Leftrightarrow x = 9,52$$

Altså overstiger sandsynligheden for at få ætsning på gummerne 80% ved en koncentration på 9,52% hydrogenperoxid.

2. Læsning og forståelse af videnskabelig artikel

1.

Det forventes at alle tænderne for alle grupperne, efter at have ligget i te i 7 dage, vil blive mørkere, idet de alle er klargjort ens. Altså vil ΔL være et negativt tal som ikke er 0. Det forventes at $\Delta L < 0$.

2.

95% konfidensinterval for ΔL for produktet CR udregnes i RStudio baseret på data fra artiklen: [-6,94 ; -9,76]

Altså er jeg 95% sikker på at den sande ændring i lysintensiteten (ΔL) ligger i intervallet [-6,94 ; -9,76] for tandblegningsproduktet CR.

3.

Nulhypotesen undersøger om de forskellige gruppers tænder er ens i farveparameter. De er altså interesserede i at undersøge om alle tænder er samme farve, således at der er et ens udgangspunkt for de forskellige produkter, så man kan teste deres effektivitet som blegemidler. Det oplyses at $p > 0,05$, så nulhypotesen forkastes ikke, og man konkluderer at tænderne, mere eller mindre, er samme farve i forhold til de parametre de tester ud fra.

4.

En p -værdi på 0,78 er et udtryk for, at der er en 78% sandsynlighed for at tændernes farveparametre er ens. Altså er det meget sandsynligt at alle gruppernes tænder har samme farve.

5.

Værdien 12,42 er et udtryk for den gennemsnitlige totale farveændring efter 56 dages brug af blegemidlet CR. Altså er de 10 tænder gennemsnitligt steget 12,42 i hvidhed ud fra farveparametrene. Værdien 4,47 er spredningen, altså den gennemsnitlige afstand til gennemsnittet. Størstedelen af tændernes farveændring forventes altså at ligge i referenceintervallet [3,48 ; 21,36], hvilket altsammen er et udtryk for en forhvidning af tandfarven. Det er interessant at betragte, fordi man ser at mundskyllet har en effekt, og man får hvidere tænder hvis man benytter det korrekt.

6.

For at analysere effekten af CR til "Time 3" betragtes en farveændring, der vil være den afhængige variabel og som er kontinuer, som afhænger af tid, som er den uafhængige variabel. Da jeg kun ser på effekten til ét bestemt tidspunkt ("Time 3"), ville jeg benytte t-test for én population til at analysere effekten af CR. I t-test bruger man differensen som mål, og det er lige netop farveændringen jeg er interesseret i at betragte.

7.

Da figur 2 viser hvor røde eller grønne tænderne er, siger den ikke i sig selv noget om blegemidlernes samlede blegningseffekt. Δa for alle fem gruppers tænder ligger mellem 3,37 og 5,30 før behandling med produkterne starter, men efter de har ligget i te i syv dage (se tabel 2). Men i figur 2 har alle fem grupper en negativ Δa , hvilket betyder at tænderne i alle grupperne er blevet mere grønne efter behandling. Da AS ligger tættest på y-værdien 0, må denne være mindst effektiv til at foretage farveændring mod grøn, men da dette ikke i sig selv er et mål for hvor hvide tænderne er, kan figur 2 ikke alene benyttes til at konkludere om blegeprodukterne er effektive og bleger tænderne mere end hvis man ikke benytter dem.

8.

Når man betragter figur 2, ser man at tid har en tydelig effekt på farveændringen hos blegemidlerne. Det forventes altså at der er en interaktion mellem farveændring og tid. Dog ses det, at gruppen AS ikke har nogen markant hældning, hvorimod de fire blegemidler alle har en tydelig negativ hældning som funktion af tid. Altså må der være en interaktion mellem grupperne og tid, idet der ikke sker nogen udvikling af gruppe ASs rød-grøn farveændring henover tid, men der ses en ændring hos alle produkterne.

9.

Jeg er enig i denne påstand, idet alle tre mundskyl ligger mellem 12,42 og 14,70 i farveændring efter 56 dages behandling. Sammenlignet med gelens (OP) farveændring på 20,28 og de tænder der ikke blev behandlet (AS) hvor ændringen var 2,28, ligger de tre mundskyls ΔE meget samlet blandt sig selv og med afstand til de to ydergrupper. Derudover kan man se på figur 1, som viser ΔL som funktion af tid, igen at mundskyllene ligger nærmest oveni hinanden, og ikke et er signifikant bedre end et andet. Man kan nævne at figur 3 viser hvor gule eller blå tænderne bliver efter brug af de forskellige blegemidler, og at mundskyllene her ligger lidt mere spredt for sig og ikke så grupperet som i figur 1. Her ses det at midlet SC resulterer i mere blålige tænder efter behandlingsperioden, og da blåligt skær kan få tænder til at fremstå hvidere, kan dette give en lille forskel. Men overordnet set er de tre mundskyllemidler tilnærmelsesvist ens i resultater af tandblegning efter 56 dage.

3. Generelle spørgsmål

1.

a

2.

c, d, e

3.

a.

Da der er en øget relativ risiko på 25% for at få huller i tænderne hvis man drikker den sukkerholdige drik, som hvis man ikke gør, må det betyde at der er en stigning på 125%. Det kan udtrykkes som:

$$1,25 = (\text{kariesrisiko for dem som drikker den sukkerholdige drik}) / (\text{kariesrisiko for dem som ikke drikker den sukkerholdige drik})$$

Det oplyses at kariesrisikoen, for dem som drikker den sukkerholdige drik, er 10%, altså 0,1, og udtrykket kan nu omskrives til:

$$1,25 = 0,1 / (\text{kariesrisiko for dem som ikke drikker den sukkerholdige drik})$$

Nu kan man isolere kariesrisiko, for dem som ikke drikker den sukkerholdige drik:

$$1,25 \cdot (\text{kariesrisiko for dem som ikke drikker den sukkerholdige drik}) = 0,1 \Leftrightarrow$$

$$\text{kariesrisiko for dem som ikke drikker den sukkerholdige drik} = 0,1 / 1,25 \Leftrightarrow$$

$$\text{kariesrisiko for dem som ikke drikker den sukkerholdige drik} = 0,08$$

Altså er risikoen for at få huller i tænderne blandt dem der ikke drikker den sukkerholdige drik 8%.

b.

Risikoen for at få huller i tænder, hvis man drikker produktet er 10%, risikoen for at få huller i tænderne, hvis man ikke drikker produktet er 8%, altså er forskellen på de to 10-8=2 procentpoint.

Appendiks

```
d <- read.csv("http://causal.sund.ku.dk/e19/data39.csv", header=TRUE)
```

```
#opg 1
```

```
#1.1
```

```
#2
```

```
table(d$light)
```

```
table(d$erosion)
```

```
table(d$satisfaction)
```

```
mean(d$colorchange)
```

```
sd(d$colorchange)
```

```
median(d$colorchange)
```

```
quantile(d$colorchange, 1/4)
```

```
quantile(d$colorchange, 3/4)
```

```
boxplot(d$colorchange)
```

```
median(d$concentration)
```

```
quantile(d$concentration, 1/4)
```

```
quantile(d$concentration, 3/4)
```

```
boxplot(d$concentration)
```

```
mean(d$concentration)
```

```
sd(d$concentration)
```

```
(78/125)*100
```

```
(47/125)*100
```

```
(67/125)*100
```

```
(58/125)*100
```

```
(47/125)*100
```

```
(46/125)*100
```

```
(32/125)*100
```

```
#1.2
```

```
#1
```

```
hist(d$colorchange)
```

```
t.test(d$colorchange)
```

```
#2
```

```
t.test(d$colorchange)
```

```
#3
```

```
t.test(d$colorchange)
```

```
#4
```

```
mean(d$colorchange)
```

```
sd(d$colorchange)
```

```
23.13488-2*13.14096
```

```
23.13488+2*13.14096
```

```
#1.3
```

```
#1
```

```
plot(d$concentration, d$colorchange, col=(d$light+1))
```

(grunden til der står d\$light+1 er for at punkterne ikke blev hvide idet værdien "1" i R koder for hvid, og "2" koder for rød)

```
#2
```

```
effekt<-lm(d$colorchange~d$concentration*d$light)
```

```
summary(effekt)
```

```
abline(-1.4625,3.1908, col="red")
```

```
abline(-1.4625+0.5623, 3.1908-0.1573)
```

```
-1.4625+0.5623
```

```
3.1908-0.1573
```

```

#3
summary(effekt)

#4
plot(d$concentration, d$colorchange)

#5
simpler<-lm(d$colorchange~d$concentration)
summary(simpler)
abline(-1.03611,3.08087)

#7
summary(simpler)

#8
confint(simpler)

#9
3.08087*12-1.03611

#10
3.08087*10-1.03611
29.77259-2*3.572
29.77259+2*3.572

#1.4
#2
(jeg har talt tilfældene i datasættet for at opstille tabellen i opgaven)
8+39+20+26+30+2
tilfreds<-matrix(c(8, 39, 20, 26, 30, 2), nrow=2, ncol=3)
chisq.test(tilfreds)

#3
2+26+39
(2/(2+26+39))
(2/(2+26+39))*100
prop.test(2, 67)

#4
8+39
8/47
39/47
0.8297872/0.1702128

#5
26/20
(39+2)/(30+8)
1.3/1.078947

#1.5
#2
aetsning<-glm(d$erosion~d$concentration, family=binomial)
aetsning

#3
exp(0.6315)
exp(confint(aetsning))

#4

```

summary(ætsning)

#5

$\exp(0.6315 \cdot 6 - 4.6279)$

#6

$\exp(0.6315 \cdot 6 - 4.6279) / (1 + (\exp(0.6315 \cdot 6 - 4.6279)))$

#7

$0.8 = \exp(0.6315 \cdot x - 4.6279) / (1 + (\exp(0.6315 \cdot x - 4.6279)))$

#opg 2

#2

$2.23 / \sqrt{10}$

$(-8.35) + 2 \cdot 0.7051879$

$(-8.35) - 2 \cdot 0.7051879$

#5

$12.42 + 2 \cdot 4.47$

$12.42 - 2 \cdot 4.47$