

Opgavebesvarelse, Basalkursus, uge 3

Opgave 1: Udskrivning af astma patienter (DGA s. 273)

I en randomiseret undersøgelse foretaget af Storr et. al. (Lancet, i, 1987) sammenlignes effekten af en enkelt dosis prednisolone med placebo for børn med akut astma.

Der var 73 børn i placebo gruppen og 67 i prednisolone gruppen.

Resultatafsnittets første sætning lyder: “2 patients in the placebo group (3%, 95% confidence interval -1 to 6%) and 20 in the prednisolone group (30%, 19 to 41%) were discharged at first examination ($P < 0.0001$)”

Metodeafsnittet forklarer, at ovenstående P-værdi er udregnet ved hjælp af Fishers eksakte test.

1. *Skriv 2×2 -tabellen op.*

Der var 2 ud af 73 placebo-randomiserede der blev udskrevet, dvs. 71 der ikke blev, og tilsvarende var der 20 af de 67 behandlede, der blev udskrevet, og 47, der ikke blev.

To-gange-to tabellen ser derfor således ud:

Respons	Gruppe	
	Placebo	Prednisolone
Udskrevet	2	20
Hospitaliseret	71	47
	73	67

2. *Udregn χ^2 -teststørrelsen for test af uafhængighed.*

Først skal vi indlæse tabellen svarende til ovenstående. For at få den letteste fortolkning af de forskellige størrelser til sammenligning af de to grupper, er det smartest at **vende den om** i forhold til den ovenstående tabel, så hver behandling kommer til at svare til en række, medens responserne (udskrevet eller ej) bliver søjler. Dette opnås ved at skrive nedenstående

```

astma = matrix(c(20,2,47,71), nrow=2)
rownames(astma) = c("pred","plac")
colnames(astma) = c("udskriv","hosp")
astma

> astma
      udskriv hosp
pred      20   47
plac       2   71

```

For at få rækkeprocenter (1. marginal) på, skriver vi:

```

> prop.table(astma,1)*100
      udskriv      hosp
pred 29.850746 70.14925
plac  2.739726 97.26027

```

χ^2 -teststørrelsen for test af uafhængighed skal nu sammenline de to procenter for udskrivning, altså 29.6% og 2.7%. Det gør vi ved at skrive

```

> chisq.test(astma)

```

Pearson's Chi-squared test with Yates' continuity correction

```

data:  astma
X-squared = 17.394, df = 1, p-value = 3.037e-05

```

```

> chisq.test(astma)$expected
      udskriv      hosp
pred 10.52857 56.47143
plac 11.47143 61.52857

```

Dette test benytter en kontinuitetskorrektion og er derfor ikke “*det sædvanlige*” test. Dette fås ved i stedet at benytte `correct=F`, hvorved vi får

```
> chisq.test(astma,correct=F)

Pearson's Chi-squared test

data:  astma
X-squared = 19.387, df = 1, p-value = 1.067e-05
```

Vi ser, at der er en stærkt signifikant forskel på de to behandlinger, idet χ^2 -størrelsen er 17.39 (eller 19.39), med $P < 0.0001$.

3. *Var det nødvendigt at anvende Fishers eksakte test?*

Der er i artiklen anvendt et eksakt test for uafhængighed, og dette er aldrig *forkert*. Det er dog unødvendigt, når sample er tilstrækkelig stor. Tommelfingerreglen siger, at de *forventede* værdier (hvis hypotesen om uafhængighed er korrekt) skal være større 5. Vi udregner derfor de forventede værdier

```
> chisq.test(astma)$expected
      udskriv      hosp
pred 10.52857 56.47143
plac 11.47143 61.52857
```

og det ses, at de alle er en del større end 5, så der intet til hinder for at bruge det sædvanlige χ^2 -test. (Det, der har forvirret forfatterne, er selvfølgelig, at et af de *observerede* tal er mindre end 5).

Fishers eksakte test giver da også en tilsvarende lille P-værdi, og samtidig nogle oplysninger om odds ratio, som vil blive kommenteret nedenfor:

```
> fisher.test(astma)

Fisher's Exact Test for Count Data

data:  astma
```

```

p-value = 9.282e-06
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
  3.356496 137.298909
sample estimates:
odds ratio
 14.85492

```

4. *Kommenter de i resultatafsnittet angivne sikkerhedsintervaller.*

Sikkerhedsintervallet for udskrivningssandsynligheden i placebo gruppen er lidt tosset, idet det er udregnet som estimatet plus minus 2 gange den spredning, der er baseret på en normalfordelingsapproximation til Binomialfordelingen, uanset at de nedre grænser derved falder under 0. For placebo-gruppen kunne det være relevant at lave et eksakt konfidensinterval som ville blive (0.0033;0.0955), som det ses i outputtet nedenfor:

```

> binom.test(2,73)

Exact binomial test

data:  2 and 73
number of successes = 2, number of trials = 73, p-value < 2.2e-16
alternative hypothesis: true probability of success is not equal to 0.5
95 percent confidence interval:
 0.003335329 0.095489164
sample estimates:
probability of success
 0.02739726

```

5. *Udregn/estimer (med 95% sikkerhedsgrænser):*

- *Forskellen i udskrivningssandsynligheder.*

Vi ser af rækkeprocenterne i tabellen ovenfor, at de to udskrivningssandsynligheder er hhv. 2.7% og 29.9%, med en forskel på 27.1%.

Vi kan få R til at beregne et sikkerhedsinterval på denne differens ved at benytte `prop.test`:

```
> 0.29850746-0.02739726  
[1] 0.2711102
```

```
>prop.test(astma)
```

2-sample test for equality of proportions with continuity correction

```
data:  astma  
X-squared = 17.394, df = 1, p-value = 3.037e-05  
alternative hypothesis: two.sided  
95 percent confidence interval:  
 0.1410042 0.4012162  
sample estimates:  
   prop 1   prop 2  
0.29850746 0.02739726
```

Vi ser, at forskellen på ca. 27%, eller 0.27 har et konfidensinterval på (0.141, 0.401), altså en forskel på mellem 14.1% og 40.1%.

- *Odds-ratio for udskrivning mellem prednisolon og placebo.*

Fra Fishers eksakte test ovenfor fik vi estimeret odds ratio til 14.85, med konfidensgrænserne (3.36, 137.30), altså ganske upræcist.

- *Relativ risiko for udskrivning mellem prednisolon og placebo.*

På dette tidspunkt er det smart at benytte epidemiologi-pakken `epitools`, og hvis den ikke findes i forvejen, må man skrive

```
install.packages("epitools")  
library("epitools")
```

Herefter kan man (ved at vende sin tabel på passende vis) få den ønskede relative risiko:

```

> astma2 = matrix(c(71,47,2,20), nrow=2)

> rownames(astma2) = c("plac", "pred")
> colnames(astma2) = c("hosp", "udskriv")

> astma2
      hosp udskriv
plac   71      2
pred   47     20

```

Vi kan nu bruge funktionen `epitab`, som giver nedenstående output:

```

> epitab(astma2,method="riskratio")
$tab
      hosp      p0 udskriv      p1 riskratio      lower      upper      p.value
plac   71 0.9726027      2 0.02739726      1.00000      NA      NA      NA
pred   47 0.7014925     20 0.29850746     10.89552  2.646222 44.86109 9.282006e-06

$measure
[1] "wald"

$conf.level
[1] 0.95

$pvalue
[1] "fisher.exact"

```

Den relative “risiko” for udskrivning for prednisolon vs. placebo ses således at være 10.9 med et 95% CI på (2.65;44.9).

- *Relativ risiko for hospitalisering mellem prednisolon og placebo.*

Da en lavere procentdel af de prednisolon-behandlede bliver hospitaliseret, får vi her en relativ risiko under 1. Vi kan få den frem ved at vende tabellen på endnu en ny måde, nemlig ved at bytte om på søjlerne:

```

> astma3 = matrix(c(2,20,71,47), nrow=2)
> rownames(astma3) = c("plac","pred")
> colnames(astma3) = c("udskriv","hosp")

```

```
> astma3
      udskriv hosp
plac      2    71
pred     20    47
```

Nu finder vi i stedet:

```
> epitab(astma3,method="riskratio")
$tab
      udskriv      p0 hosp      p1 riskratio      lower      upper
plac      2 0.02739726   71 0.9726027 1.0000000      NA      NA
pred     20 0.29850746   47 0.7014925 0.7212529 0.6140745 0.8471378
      p.value
plac      NA
pred 9.282006e-06

$measure
[1] "wald"

$conf.level
[1] 0.95

$pvalue
[1] "fisher.exact"
```

Den relative risiko for hospitalisering ses altså at være 0.72, med med et 95% konfidensinterval på (0.61,0.85). De inverse af disse tal, 1.39 (1.18;1.63) — dvs. den relative risiko for hospitalisering mellem placebo- og prednisolol-grupperne, afviger betragteligt fra både odds-ratio og relativ risiko for udskrivning.

Bemærk at den relative risiko for hospitalisering har et konfidensinterval der er meget snævrere end intervallerne for OR og den relative risiko for udskrivning. Det skyldes i det væsentlige 2-tallet som for OR og RR for udskrivning er afgørende for nøjagtigheden af parameterestimatet.

6. *Formuler en konklusion i ord ud fra hver af de 3 typer udregninger.*

Sammenfattende kan man sige.

- Andelen af patienter der profiterer fra prednisolone er 27%. Et 95% konfidensintervallet for denne andel er (16%; 39%4).

- Forholdet mellem antallet af patienter der udskrives hhv. hospitaliseres er 15 gange større i prednisolone-gruppen end i placebo-gruppen. Et konfidensinterval for dette forhold er fra 3 til 68 gange.
- Den relative risiko (chance) for at blive udskrevet for prednisolone-behandlede i forhold til placebo-behandlede er 11. Et 95% konfidensinterval for denne relative risiko er fra 2.6 til 45.
- Den relative risiko for at blive hospitaliseret for placebo-behandlede i forhold til prednisolone-behandlede er 1.4. Et 95% konfidensinterval for denne relative risiko er fra 1.2 til 1.6.
- Der er signifikant forskel på placebo- og prednisolone-behandling, $\chi^2 = 17.4$, $P < 0.000$.

Opgave 2: Postoperative komplikationer

En traditionelt anvendt operationsprocedure har en **kendt** risiko for postoperative komplikationer på 20%.

1. *Hvad er sandsynligheden for at operere 10 konsekutive patienter med den traditionelle metode, uden at se nogen tilfælde af postoperative komplikationer?*

Da komplikationssandsynligheden for den traditionelle metode vides at være $p_0 = 0.2$ (20%), må sandsynligheden for en komplikationsfri operation være $1 - 0.2 = 0.8$. Da udfaldene af de enkelte operationer antages at være uafhængige af hinanden, må sandsynligheden for 10 konsekutive operationer uden komplikationer være

$$0.8^{10} = 0.107$$

altså 10.7%. Det sker altså jævnligt.

En ny metode er nu blevet foreslået, og den er netop blevet afprøvet på de første 10 patienter, uden at dette har givet anledning til nogen postoperative komplikationer.

2. *Er der tilstrækkelig evidens for at den nye operationsteknik er bedre end den traditionelle?*

I lyset af svaret på spørgsmål 1 ovenfor, må vi sige, at det ikke er særligt spektakulært at observere 10 konsekutive komplikationsfri operationer med den gamle metode, og der kan derfor ikke være evidens for, at den nye metode er bedre end den gamle.

Faktisk vil de 10.7% netop være P-værdien for test af hypotesen

$$H_0 : p_1 = p_0$$

hvor p_1 betegner den ukendte sandsynlighed for komplikationer med den nye metode, **vel at mærke, hvis vi tester ensidigt**, dvs. hvis vi på forhånd er sikre på, at den nye metode i hvert fald ikke kan være *værre* end den gamle. P-værdien er jo netop *halesandsynligheden*, dvs. sandsynligheden for at observere *dette eller noget endnu mere ekstremt*, hvis hypotesen er sand. Og der *er* jo ikke noget, der er endnu mere ekstremt, hvis vi kun ser på den ene hale.

Det er sædvanligvis ikke rimeligt at teste ensidigt, så derfor er det mere rimeligt at sige, at P-værdi er det dobbelte, altså 0.214, eller 21.4%. Vi kan få et eksakt konfidensinterval med tilhørende P-værdi ved at skrive:

```
> binom.test(x=10, n=10, p=0.8)
```

```
Exact binomial test
```

```
data: 10 and 10
number of successes = 10, number of trials = 10, p-value = 0.2282
alternative hypothesis: true probability of success is not equal to 0.8
95 percent confidence interval:
 0.6915029 1.0000000
sample estimates:
probability of success
1
```

Vi ser i dette output, at P-værdien for det eksakte test er 22.8%, og vi kan derfor ikke forkaste vores hypotese, og der er dermed *ikke* evidens for, at den nye metode er bedre end den gamle.

Vi benytter naturligvis det eksakte test til sådanne små datamaterialer, og vi kan også aflæse det eksakte sikkerhedsinterval til (0.69,1.00). Det

tilsvarende approksimative sikkerhedsinterval, baseret på en approksimation af Binomialfordelingen til Normalfordelingen bør *ikke* benyttes i sådan en situation.

3. *Hvor mange patienter skal man operere uden postoperative komplikationer før man med rimelig sikkerhed kan sige, at den nye metode er bedre end den traditionelle?*

Vi så ovenfor, at P-værdien for hypotesen om identitet af de to behandlingsmetoder kunne udregnes approksimativt som

$$2 \times (1 - 0.2)^n$$

hvor vi med n betegner det antal personer, der observeres konsekutivt uden komplikationer. Hvis denne P-værdi er mindre end 0.05, vil vi forkaste vores hypotese. Vi skal således løse uligheden

$$2 \times (1 - 0.2)^n < 0.05$$

Dette kan gøres ved “*trial and error*” metoden, eller ved at tage logaritmer på begge sider af ulighedstegnet, hvorved vi får

$$\begin{aligned} \log(2) + n \times \log(0.8) &< \log(0.05) \\ n &> \frac{\log(0.05) - \log(2)}{\log(0.8)} = 16.53 \end{aligned}$$

Vi skal altså op på at observere mindst 17 konsekutive patienter uden at se en eneste komplikation, førend vi med noget, der ligner rimelig sikkerhed at kunne konkludere, at den nye metode er bedre end den gamle.

Bemærk i øvrigt, at vi i udregningen ovenfor skiftede ulighedstegn fra $<$ til $>$, da vi dividerede med $\log(0.8)$. Det gjorde vi fordi $\log(0.8) = -0.22$, altså negativt.

Vi checker lige disse udregninger i R ved at afprøve hhv. $n = 16$ og $n = 17$:

```
> binom.test(x=16, n=16, p=0.8)

Exact binomial test

data: 16 and 16
number of successes = 16, number of trials = 16, p-value = 0.0548
alternative hypothesis: true probability of success is not equal to 0.8
95 percent confidence interval:
 0.7940928 1.0000000
sample estimates:
probability of success
                  1
```

.....

```
> binom.test(x=17, n=17, p=0.8)

Exact binomial test

data: 17 and 17
number of successes = 17, number of trials = 17, p-value = 0.03345
alternative hypothesis: true probability of success is not equal to 0.8
95 percent confidence interval:
 0.8049357 1.0000000
sample estimates:
probability of success
                  1
```

og konstaterer, at $n = 17$ lige netop er tilstrækkeligt til at få signifikans ($P=0.033$), hvorimod $n = 16$ er for lidt ($P=0.055$).

Bemærk, at dette *ikke* er en traditionel dimensionering, da vi jo forudsætter, at der ikke kommer nogen komplikationer.

Opgave 3: Børns sovevaner (DGA s. 274)

For 3469 børn på 14 år har man registreret antal timer tilbragt i sengen i løbet af et døgn (Macgregor & Balding, Ann. Hum. Biol., 15, 1988).

I nedenstående tabel ses data opdelt efter køn, idet tiderne er afrundet til nærmeste halve time.

Antal timer tilbragt i sengen

	≤ 7.0	7.5	8.0	8.5	9.0	9.5	10.0	> 10.0	Total
Drenge	88	109	210	324	359	313	182	85	1670
Piger	92	108	217	349	436	334	198	65	1799
Total	180	217	427	673	795	647	380	150	3469

1. *Hvilke metoder kunne anvendes til at sammenligne fordelingen for piger og drenge?*

De to grupper (drenge hhv. piger) skal sammenlignes m.h.t. fordelingen af antal timer i sengen pr. døgn.

Måden, som data er skrevet op på, lægger klart op til, at man foretager et χ^2 -test, så vi starter med at indlæse en 2*8-matrix:

```
> timer = matrix(c(88,109,210,324,359,313,182,85,
+                  92,108,217,349,436,334,198,65),
+                  ncol = 8, byrow=T)
```

```
> timer
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
[1,]   88  109  210  324  359  313  182   85
[2,]   92  108  217  349  436  334  198   65
```

hvorefter vi udregner rækkeprocenter:

```
> prop.table(timer,1)*100
      [,1]      [,2]      [,3]      [,4]      [,5]      [,6]      [,7]      [,8]
[1,] 5.269461 6.526946 12.57485 19.40120 21.49701 18.74251 10.89820 5.089820
[2,] 5.113952 6.003335 12.06226 19.39967 24.23569 18.56587 11.00611 3.613118
```

Disse rækkeprocenter ser meget ens ud for de to køn, men vi checker lige med et χ^2 -test:

```
> chisq.test(timer)
```

Pearson's Chi-squared test

```
data: timer  
X-squared = 7.8305, df = 7, p-value = 0.3478
```

Vi finder altså ingen signifikant forskel med χ^2 -testet ($P=0.35$), ret overbevisende i betragtning af de meget store tal, det drejer sig om.

2. *Er der nogen forskel på piger og drenge i denne henseende?*

Det ser altså ikke sådan ud, men nu er χ^2 -testet jo ikke særligt stærkt, når det drejer sig om så store tabeller.

Man kunne få den tanke at anvende det såkaldte **trend test**, da en af siderne i tabellen er en ordinal inddeling (tid i sengen). Det ville imidlertid svare til at vende årsagsammenhængen og undersøge om sandsynligheden for at være en dreng afhang af hvor lang tid man tilbragte i sengen.

Sammenlign f.eks. til en situation, hvor et trend test *ville* være relevant, nemlig hvis inddelingerne var skostørrelse vs. risikoen for få kejsersnit. Humlen er: Hvad er outcome, og hvad er forklarende variabel?

Alternativt kunne man lave et t-test, hvor man så må lade som om alle personer i samme celle i tabellen have samme værdi for antal timer i sengen. Det svarer blot til at have timer som responsvariabel i en analyse hvor man anvender køn som forklarende variabel. Nu opfatter vi altså timerne som en kvantitativ information, selv om den er afrundet (og derfor ikke bliver flot afbildet i histogrammet nedenfor).

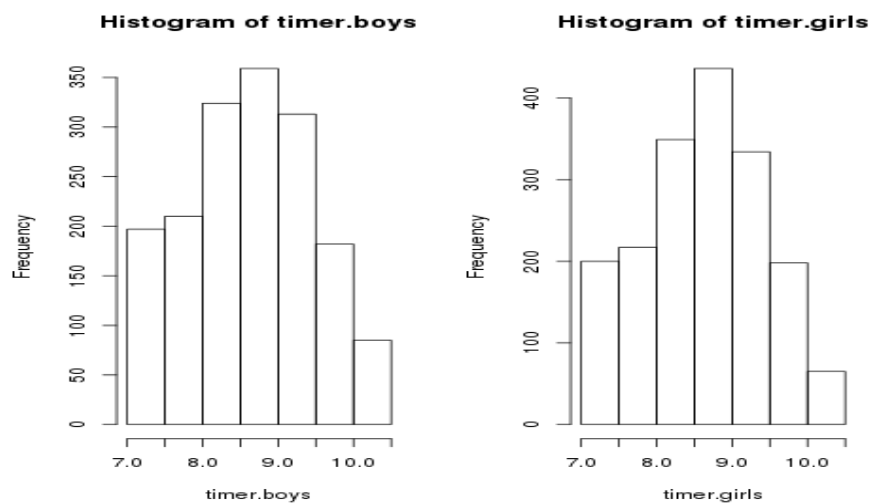
Vi definerer derfor de to vektorer, svarende til hhv. drenge- og pige-observationer:

```
> timer.boys = c(rep(7, 88), rep(7.5, 109), rep(8, 210), rep(8.5, 324),
+               rep(9, 359), rep(9.5, 313), rep(10, 182), rep(10.5, 85))
> table(timer.boys)
timer.boys
  7  7.5   8  8.5   9  9.5  10 10.5
88 109 210 324 359 313 182  85
```

```
> timer.girls = c(rep(7, 92), rep(7.5, 108), rep(8, 217), rep(8.5, 349),
+               rep(9, 436), rep(9.5, 334), rep(10, 198), rep(10.5, 65))
> table(timer.girls)
timer.girls
  7  7.5   8  8.5   9  9.5  10 10.5
92 108 217 349 436 334 198  65
```

hvorefter vi kan tegne histogrammer:

```
> par(mfrow=c(1,2))
> hist(timer.boys)
> hist(timer.girls)
```



Vi kan nu foretage et T-test ved at skrive

```
> t.test(timer.boys, timer.girls)

Welch Two Sample t-test

data:  timer.boys and timer.girls
t = 0.19857, df = 3423.6, p-value = 0.8426
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.05197197  0.06368539
sample estimates:
mean of x mean of y
 8.852994  8.847137

> mean(timer.boys) - mean(timer.girls)
[1] 0.005856713
```

Forskellen mellem det antal timer drenge og piger tilbringer i sengen er altså gennemsnitligt 0.00586 timer eller ca. 21 sekunder. Et konfidensinterval for dette middeltal er (-0.0520, 0.0637) timer, dvs. (-3.1, 3.8) minutter. Af testet ses at der ikke er nogen signifikant forskel på drenge og piger hvad angår den gennemsnitlige tid de tilbringer i sengen. Af tabellen ses desuden at der heller ikke er nogen reel forskel i fordelingen af den tid de to populationer tilbringer i sengen.