

## Opgavebesvarelse, Basalkursus, uge 3

### Opgave 1: 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  $p_0=20\%$ , 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, 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 se, at dette er tilfældet ved at udføre testet i SAS

```
data operation;
input udfald antal;
datalines;
0 10
1 0
;
run;

proc freq data=operation;
    tables udfald / binomial(p=0.8);
exact binomial;
weight antal;
run;
```

hvorved vi får

The FREQ Procedure

udfald	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	10	100.00	10	100.00

Binomial Proportion for udfald = 0

Proportion (P)	1.0000
ASE	0.0000
95% Lower Conf Limit	1.0000
95% Upper Conf Limit	1.0000

Exact Conf Limits	
95% Lower Conf Limit	0.6915
95% Upper Conf Limit	1.0000

Test of H0: Proportion = 0.8

```

ASE under H0          0.1265
Z                     1.5811
One-sided Pr > Z      0.0569
Two-sided Pr > |Z|    0.1138

Exact Test
One-sided Pr >= P     0.1074
Two-sided = 2 * One-sided 0.2147

Sample Size = 10

```

Vi ser nederst i dette output, at P-værdien for det eksakte test netop er 21.5%, 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, ses at være (1.00,1.00), hvilket naturligvis er helt vanvittigt og skyldes, at standard error her estimeres til 0.

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 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.

## Opgave 2: 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  $2 \times 2$ -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  $\chi^2$ -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 grupperne (**treat**) først og responstypen (**udskrevet**) sidst i **table**-sætningen. Samtidig udregner vi  $\chi^2$ -testet for uafhængighed (option **chisq**), samt de forventede værdier (option **expected**) i hver celle:

```
data astma;
input udskrevet$ treat$ antal;
datalines;
nej pred 47
nej plac 71
ja pred 20
ja plac 2
;
run;

proc freq data = astma;
weight antal;
table treat * udskrevet / chisq nopercnt expected;
run;
```

hvorved vi får outputtet

Table of treat by udskrevet

treat		udskrevet		
Frequency		Expected		
Row	Pct	ja	nej	Total
-----+-----+-----+				
plac		2	71	73
		11.471	61.529	
		2.74	97.26	
-----+-----+-----+				
pred		20	47	67
		10.529	56.471	
		29.85	70.15	
-----+-----+-----+				
Total		22	118	140

Statistics for Table of treat by udskrevet

Statistic	DF	Value	Prob
Chi-Square	1	19.3871	<.0001

Fisher's Exact Test

Cell (1,1) Frequency (F)	2
Two-sided Pr <= P	9.282E-06

Vi ser, at der er en stærkt signifikant forskel på de to behandlinger, idet  $\chi^2$ -størrelsen er 19.39, med  $P < 0.0001$ .

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

Når man skal vurdere hvorvidt det er nødvendigt at bruge Fisher's eksakte test, skal man se om de *forventede* værdier nogen steder er mindre end 5. De forventede værdier fremgår af tabellen ovenfor, og det ses, at de alle er en del større end 5, så der intet til hinder for at bruge det sædvanlige  $\chi^2$ -test. (Det, der har forvirret forfatterne, er selvfølgelig, at et af de *observerede* tal er mindre end 5).

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 normalfordelingsapproksimation 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 ses i outputtet længere nede.

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%, dvs. med en forskel på 27.1%.

Vi kan få SAS til at beregne et sikkerhedsinterval på denne differens ved at tilføje en option `riskdiff` i `table`-sætningen:

```
proc freq data = astma;
weight antal;
table treat * udskrevet / riskdiff;
run;
```

hvorved vi får (her kun medtaget halvdelen, svarende til “Column 1 Risk”, da kolonne 1 angiver de udskrevne

Statistics for Table of treat by udskrevet

Column 1 Risk Estimates						
	Risk	ASE	(Asymptotic) 95% Confidence Limits		(Exact) 95% Confidence Limits	
Row 1	0.0274	0.0191	0.0000	0.0648	0.0033	0.0955
Row 2	0.2985	0.0559	0.1889	0.4081	0.1928	0.4227
Total	0.1571	0.0308	0.0969	0.2174	0.1012	0.2282
Difference	-0.2711	0.0591	-0.3869	-0.1553		
Difference is (Row 1 - Row 2)						
Sample Size = 140						

Vi ser, at forskellen på ca. 27%, eller 0.27 har et konfidensinterval på (0.155, 0.387), altså en forskel på mellem 15.5% og 38,7%.

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

Vi kan få SAS til at beregne odds ratio med sikkerhedsinterval ved at tilføje option `relrisk` i `table`-sætningen:

```
proc freq data = astma;
weight antal;
table treat * udskrevet / relrisk;
run;
```

hvorved vi får

Statistics for Table of treat by udskrevet

Estimates of the Relative Risk (Row1/Row2)

Type of Study	Value	95% Confidence Limits	
Case-Control (Odds Ratio)	0.0662	0.0148	0.2965
Cohort (Col1 Risk)	0.0918	0.0223	0.3779
Cohort (Col2 Risk)	1.3865	1.1804	1.6285

Sample Size = 140

Bemærk, at alle de ovenfor angivne størrelser til sammenligning af de to grupper refererer til (Row1/Row2), dvs. svarende til placebo vs. prednisolon, fordi det relaterer sig til den alfabetiske rækkefølge. For at svare på spørgsmålene nedenfor, skal alle størrelser derfor inverteres. Alternativt kunne man omdøbe navnene, f.eks. til 1:prednisolon og 2:placebo.

Odds-ratio for udskrivning mellem placebo og prednisolone er 0.0662 med et 95% konfidensinterval på (0.0148;0.2965). Den omvendte odds-ratio fås ved at tage den inverse, den bliver 15.1 (3.37;67.7).

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

Den relative risiko for udskrivning er  $\frac{1}{0.0918} = 10.9$  med et 95% CI på  $(\frac{1}{0.3779}, \frac{1}{0.0223}) = (2.65; 44.7)$ .

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

Den relative risiko for hospitalisering er 0.72 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 prednisolon-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 parameteren.



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% konfidensinterval for denne andel er (16%; 39%).
- 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,  $\chi^2 = 17$ ,  $p=0.0000$ .

### 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**

	$\leq 7.0$	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 (drengene hhv. piger) skal sammenlignes m.h.t. fordelingen af antal timer i sengen pr. døgn. Den naturlige sammenligning vil basere sig på den kumulative fordeling for hver af de to grupper, og vi benytter derfor option `cumcol` i tabel opsætningen nedenfor.

Men først skal vi indlæse data og opstille en tabel med relevante procenter. Samtidig laver vi lige et  $\chi^2$ -test for uafhængighed (mellem køn og timer i sengen, dvs. et test for om piger og drenge sover lige længe):

```
data sov;
input timer n sex $;
datalines;
7 88 D
7.5 109 D
8 210 D
8.5 324 D
9 359 D
9.5 313 D
10 182 D
10.5 85 D
7 92 P
7.5 108 P
8 217 P
8.5 349 P
9 436 P
9.5 334 P
10 198 P
10.5 65 P
;
run;
proc freq data = sov;
weight n;
table timer * sex / chisq cumcol nopercnt norow;
run;
```

og vi finder hermed

```
The FREQ Procedure

Table of timer by sex
```

timer		sex	
Frequency			
Col Pct			
Cumulative Col%	D	P	Total
7	88	92	180
	5.27	5.11	
	5.27	5.11	5.19
7.5	109	108	217
	6.53	6.00	
	11.80	11.12	11.44
8	210	217	427
	12.57	12.06	
	24.37	23.18	23.75
8.5	324	349	673
	19.40	19.40	
	43.77	42.58	43.15
9	359	436	795
	21.50	24.24	
	65.27	66.81	66.07
9.5	313	334	647
	18.74	18.57	
	84.01	85.38	84.72
10	182	198	380
	10.90	11.01	
	94.91	96.39	95.68
10.5	85	65	150
	5.09	3.61	
	100.00	100.00	100.00
Total	1670	1799	3469

Statistics for Table of timer by sex

Statistic	DF	Value	Prob
Chi-Square	7	7.8305	0.3478

Hvis man sammenligner tallene i tabellen under cumulative col %, ser men at der ikke er megen forskel på søvnfordelingen mellem drenge og piger. Samme konklusion giver  $\chi^2$ -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  $\chi^2$ -testet jo ikke særligt stærkt, når det drejer sig om så store tabeller.

Man kunne få den tanke at anvende et **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. Sammenlignet med kejsersnitseksemplet fra forelæsningerne går inddelingen her den anden vej, hvilket gør anvendelsen af trend-testet meningsløst.

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 regressionsanalyse hvor man anvender sex som forklarende variabel.

```
proc ttest data = sov;
  class sex;
  freq n;
  var timer;
run;
```

hvorved vi finder resultatet

The TTEST Procedure

Variable: timer  
Frequency: n

sex	N	Mean	Std Dev	Std Err	Minimum	Maximum
D	1670	8.8530	0.8838	0.0216	7.0000	10.5000
P	1799	8.8471	0.8506	0.0201	7.0000	10.5000
Diff (1-2)		0.00586	0.8668	0.0295		

sex	Method	Mean	95% CL Mean	Std Dev
D		8.8530	8.8106 8.8954	0.8838
P		8.8471	8.8078 8.8865	0.8506
Diff (1-2)	Pooled	0.00586	-0.0519 0.0636	0.8668
Diff (1-2)	Satterthwaite	0.00586	-0.0520 0.0637	

sex	Method	95% CL	Std Dev
D		0.8548	0.9148
P		0.8237	0.8794
Diff (1-2)	Pooled	0.8468	0.8876
Diff (1-2)	Satterthwaite		

Method	Variances	DF	t Value	Pr >  t
Pooled	Equal	3467	0.20	0.8424
Satterthwaite	Unequal	3423.6	0.20	0.8426

#### Equality of Variances

Method	Num DF	Den DF	F Value	Pr > F
Folded F	1669	1798	1.08	0.1112

Forskellen mellem det antal timer drenge og piger tilbringer i sengen er 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.