Faculty of Health Sciences

Basal Statistik

Kategorisk outcome. Tabeller.

Lene Theil Skovgaard

24. februar 2020

Kategorisk outcome

- Sandsynligheder og odds
- Binomialfordelingen
- 2×2 tabeller, relativ risiko og odds ratio
- Case-control studier
- Større tabeller
- Confounding, Simpsons paradox
- Parrede binære data

Home pages:
`http://publicifsv.sund.ku.dk/~sr/BasicStatistics`
E-mail: `ltsk@sund.ku.dk`

$*$: Siden er lidt teknisk

Sandsynligheder

En sandsynlighed er en talværdi mellem 0 og 1, der udtrykker usikkerhed omkring et udfald/hændelse

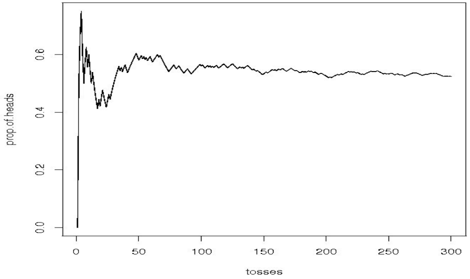
- p $\sim \frac{1}{2}$: stor usikkerhed
  (møntkast)
- p $\sim$ 0 eller 1: lille usikkerhed
  (vinde i lotto / overleve næste dag)
- *realistiske* værdier:
  - sandsynlighed for at være farveblindhed
  - for at udvikle cancer inden 60 år
  - for at få en komplikation efter en operation

Bestemmelse af sandsynligheder

- Logik
  En terning har 6 sider, som ender opad med lige stor sandsynlighed, nemlig $\frac{1}{6}$
  sandsynlighed$=\sharp$gunstige$/\sharp$mulige
- Erfaring
  Hvis vinden kommer fra øst om sommeren, bliver det med stor sandsynlighed (dvs. sædvanligvis) varmt –
  frekvensfortolkningen eller evidensbaseret
- Subjektivt
  Fornemmelser, hensigter, vurderinger,
  måske baseret på tidligere (vage) erfaringer

Frekvensfortolkningen

Store Tals Lov: Møntkast



Når vi udfører eksperimentet mange gange, vil frekvensen stabilisere sig omkring sandsynligheden for plat/krone

Hvis mønten er "fair", er dette også simpel logik.

Frekvensfortolkning, fortsat

Dette matematiske begreb involverer uafhængige og identiske gentagelser af samme eksperiment.

Hvad betyder det?

- Hvis de *virkelig* var identiske, ville de vel give det samme?
- De skal være identiske mht. de betingelser, som vi ved (eller mener at vide) har indflydelse på resultatet
- Vi har ingen mulighed for at skelne mellem dem, de er ombyttelige (interchangeable)

Børnefødsler

To mulige udfald: Bliver det en dreng eller en pige?

Hvad er gentagelserne her? Søskende ...

Ofte, betyder 'identiske gentagelser' i praksis

- situationer, der 'ser ens ud'
  Her: fødsler blandt andre kvinder

Det antages, at alle kvinder har den samme sandsynlighed p for at få en dreng (resp. pige), eller vi kan i hvert fald ikke på forhånd sige, *hvem* der skulle have større eller mindre sandsynlighed ...

Hvad er sandsynligheden for, at det næste barn født på Rigshospitalet er en dreng?

Eksempler

1. I en kasse ligger kugler nummereret 1-9, en af hver. Hvis man trækker en kugle tilfældigt, hvad er så sandsynligheden for, at den har et nummer større end 5?
2. Forestil dig, at du har kastet en mønt 20 gange og har fået 19 kroner og en plat.
   Hvad er sandsynligheden for at få krone i næste kast?
3. Hvad er sandsynligheden for, at du får en overordnet stilling inden for de næste 10 år?
4. Et par blå og et par sorte sokker ligger enkeltvis sammenrodet i en skuffe. Hvis du tilfældigt tager to sokker op, hvad er så sandsynligheden for, at de har samme farve? .....I kan lige tænke i pausen

"Enten-eller" sandsynligheder

I løbet af ens levetid er man udsat for en række risici, f.eks. at pådrage sig forskellige kræftformer:

- lungekræft
- brystkræft
- ...

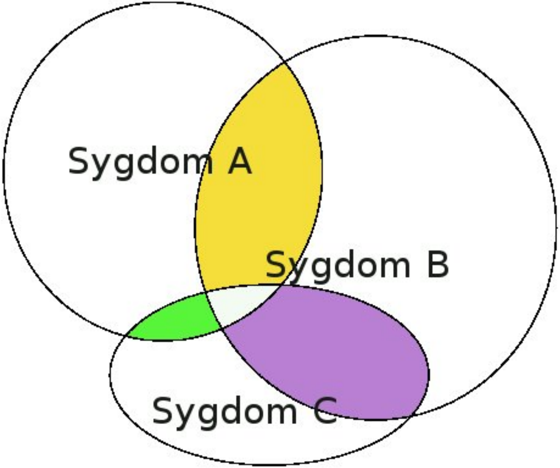
hver med en vis sandsynlighed.

Hvad er sandsynligheden for at få "en eller anden" cancer?

- Det er ikke summen af de enkelte sandsynligheder?
  Hvorfor ikke?
- Fordi den ene form ikke udelukker de øvrige!

Sandsynlighed for foreningsmængde

Enten-eller: $P(A \cup B \cup C)$?



Sandsynlighed for fællesmængde

Både-og: $P(A \cap B \cap C)$?
Hvad er sandsynligheden for at få *både* **L**ungekræft *og* **B**rystkræft?

Hvis de to kræftformer er uafhængige af hinanden, gælder der den simple formel

$$P(L \cap B) = P(L) \times P(B)$$

Ofte er det netop spørgsmålet om uafhængighed, der er i fokus:

- Uafhængighed mellem køn og farveblindhed
- Uafhængighed mellem behandling og forekomst af komplikation

Eksempel: Farveblindhed og køn

	Farveblindhed?		
	nej	ja	Total
Piger	119	1	120
Drenge	144	6	150
Total	263	7	270

Outcome Y: Farveblindhed
dikotom, 0/1, nej/ja
Kovariat: Køn:
dikotom, 0/1, pige/dreng

Nulhypotese: H_0: **Farveblindhed er uafhængig af køn**

Hyppighed af farveblindhed blandt drenge:
$P(farveblind|dreng) = \frac{6}{150} = 0.04$

Hyppighed af farveblindhed blandt piger:
$P(farveblind|pige) = \frac{1}{120} = 0.0083$

De ser jo ikke helt ens ud...

Fordeling af antal farveblinde under H_0

Antag, at sandsynligheden for farveblindhed er 2.6% (svarende til de observerede 7 tilfælde ud af 270 børn) for *både* drenge og piger, altså uafhængig af køn

Hvor mange farveblinde ville vi så forvente blandt

- 120 nyfødte piger: `120*0.026=3.1`
- 150 nyfødte drenge: `150*0.026=3.9`

Vi finder i stedet 1 hhv. 6.

- Er det mærkeligt?
- Eller er det bare tilfældighedernes spil?

Fordeling under H_0, fortsat

Fordelingen af antal farveblinde, dvs:
De mulige udfald, og deres respektive sandsynligheder.

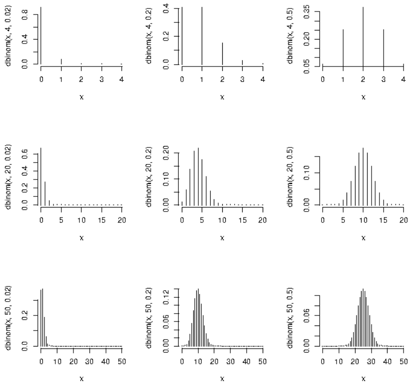
F.eks. for pigerne:

- Mulige udfald: $X = 0, 1, 2, \ldots, 119, 120$
- Sandsynligheder (punktsandsynligheder) ?:
  - $P(X = 0) = (1 - 0.026)^{120} = 0.042$
  - $P(X = 1) =?$
  -
  -
  - $P(X = 120) = 0.026^{120} = 6.3 \cdot 10^{-191}$

Vi kalder denne fordeling en Binomialfordeling,
med sandsynlighedsparameter $p = 0.026$

Eksempler på binomialfordelinger

n=4, 20 og 50; p=0.02, 0.2 og 0.5



Binomialfordelingen, generelt

Binomialfordelingen fremkommer som sum af
N uafhængige binære variable,
hver med sandsynlighed p for et 1-tal
(og dermed med sandsynlighed $1 - p$ for et 0).

Vi skriver: $X \sim \text{Bin}(N, p)$

Eksempel: Afkom blandt 4-barns mødre:

$$X_m = U_{m1} + U_{m2} + U_{m3} + U_{m4}$$

hvor U_{mj} er indikatoren for, at barn j for mor m er en dreng.

Antal drenge for den m'te mor: $X_m \sim \text{Bin}(4, p)$

*Binomialfordelingen, teknisk

Fordelingen Bin(N, p) har punktsandsynligheder

$$P(X = x) = \binom{N}{x} p^x (1-p)^{N-x}$$

Størrelsen $\binom{N}{x}$ kaldes Binomialkoefficienten og angiver antallet af måder hvorpå man kan vælge x ud af N:

$$\binom{N}{x} = \frac{N!}{x!(N-x)!}$$

$$N! = N(N-1)(N-2)\cdots 2 \cdot 1$$

Desuden defineres: $0! = 1$

Eksempel: 4-barns mødre

- 4 binære (0/1) variable for hver mor, $U_{m1}, U_{m2}, U_{m3}, U_{m4}$:

$$U_{mi} = \begin{cases} 1, & \text{hvis fødsel } i \text{ for mor } m \\ & \text{resulterer i en dreng} \\ 0, & \text{hvis det bliver en pige} \end{cases}$$

- Sandsynlighed for drengefødsel: p
- $X_m = U_{m1} + U_{m2} + U_{m3} + U_{m4}$,
  antal drenge for den m'te mor
- Hvilke kombinationsmuligheder er der?

Kønsfordeling i 4-barns familier

X_m	antal drenge	antal piger	sandsynlighed
0	0	4	$(1-p)^4$
1	1	3	$4 \times p \times (1-p)^3$
2	2	2	$6 \times p^2 \times (1-p)^2$
3	3	1	$4 \times p^3 \times (1-p)$
4	4	0	p^4

Hvordan ser de fordelinger ud?

Det så vi yderst til højre foroven på s. 15

Når man har store antal....

f.eks. når man tæller

- antal farveblinde født i Danmark pr. år
- antal indlæggelser i danske kommuner pr. år

er det meget besværligt, eller endda umuligt, at udregne sandsynligheder, baseret på Binomialfordelingen....

Heldigvis findes der så approksimationer....
fordi fordelingen ligner

- Normalfordelingen, for moderate p
- Poissonfordelingen, for små p

Passer Binomialfordelingen i praksis?

Norsk undersøgelse af kønsfordelingen blandt søskende:

Tabell 2 Antall barn pr. familie og deres kjønnsfordeling, ca. 1950–1984

	Antall kvinner Absolutt	%	Antall kvinner som har fått ett barn til (%)
Ettbarnsmødre			
1 jente	299 991	48,6	74,7
1 gutt	317 528	51,4	74,8
I alt	617 519	100,0	74,7
Tobarnsmødre			
2 jenter	109 566	23,7	44,6
1 jente og 1 gutt	230 111	49,9	39,6
2 gutter	121 886	26,4	45,0
I alt	461 563	100,0	42,2
Trebarnsmødre			
3 jenter	24 072	12,4	32,9
2 jenter og 1 gutt	69 084	35,5	29,4
1 jente og 2 gutter	73 262	37,6	29,3
3 gutter	28 429	14,6	31,6
I alt	194 847	100,1	30,1
Firebarnsmødre			
4 jenter	3 969	6,8	30,8
3 jenter og 1 gutt	13 901	23,7	28,6
2 jenter og 2 gutter	20 806	35,5	27,0
1 jente og 3 gutter	15 251	26,0	27,9
4 gutter	4 699	8,0	28,5
I alt	58 626	100,0	28,0

Forventede sandsynligheder - og antal

for 4-barns familier,
baseret på en Binomialfordeling med $p = 0.514$
og formlerne s. 17

x	P(X=x)	i %	Forventet antal familier
0	0.05578855	5.6	3271
1	0.23601082	23.6	13836
2	0.37441223	37.4	21950
3	0.26398887	26.4	15477
4	0.06979953	7.0	4092

Modelkontrol: Observeret vs. forventet

Vi sammenligner den *forventede* fordeling fra s. 22 med den *observerede* fordeling fra s. 21 af antal drenge i 4-barnsfamilier.

x	obs. antal	forv. antal	obs.-forv.
0	3969	3271	+698
1	13901	13836	+65
2	20806	21950	-1144
3	15251	15477	-226
4	4699	4092	+607

Goodness-of-fit: $\sum \frac{(obs.-forv.)^2}{forv.} = 302.2 \sim \chi^2(4)$
som helt klart viser, at Binomialfordelingen *ikke* passer
(det formodes I dog ikke at forstå...)

Bemærkninger til modeltilpasning

- Der er for mange familier med 4 enskønnede børn og for få med 2 af hver.
- Modellen passer nogenlunde for familier med 1 af den ene slags og 3 af den anden slags.
- Sammenholdt med tabel 1 (næste side), hvoraf man ser at sandsynligheden for en drengefødsel afhænger af hvor mange drenge, man har i forvejen, må vi konkludere, at

  Det ser ud til, at nogle kvinder har tendens til at føde drenge og andre til at føde piger.

Mere information fra den norske undersøgelse

Tabell 1 Sannsynligheten for at neste barn er gutt, gitt antall tidligere barn og deres kjønnsfordeling

	Paritet (antall tidligere levendefødte barn)	Sannsynlighet for å få en gutt %	Antall gutter fra før	Sannsynlighet for å få en gutt %
Første barn	0	51,4	0	51,4
Andre barn	1	51,2	0	51,1
			1	51,3
Tredje barn	2	51,4	0	50,7
			1	51,4
			2	51,9
Fjerde barn	3	51,1	0	49,9
			1	50,9
			2	51,2
			3	52,3
Femte barn	4	50,6	0	49,6
			1	50,9
			2	50,4
			3	50,1
			4	52,7

Bemærkninger til modeltilpasning, fortsat

- Tabel 2 (s. 21) viser, at sandsynligheden for at få et barn mere afhænger af kønsfordelingen blandt de børn, man har i forvejen, idet kvinder med enskønnede børn har større tendens til at få et barn mere.
- Men Tabel 1 (s. 25) viser så, at disse med overvejende sandsynlighed (sat lidt på spidsen) bare får en mere af den slags, de allerede har i forvejen, og så *forstærker* det den ovenfor fundne effekt.

Der er selektion: De kvinder, der har tendens til at få samme slags børn hver gang, får generelt flere børn.

En del af den fundne overhyppighed af enskønnede søskende kan dog også skyldes forekomst af (enæggede) flerfoldsfødsler....

Et eksempel i detaljer

Påstand: Der fødes flere drenge end piger,
evt. bare *i min familie*

En mulig undersøgelse:
for forståelsens skyld lavet rigtig lille, f.eks:

- 8 konsekutive fødsler på Rigshospitalet
eller
- 8 børnebørn

Hvilken størrelse skal observeres?
X: Antal (af de 8), der viser sig at være drenge

Undersøgelsens (hypotetiske) udfald:

$x = 7$, altså 7 drenge (og dermed kun 1 pige)

Ukendt parameter:

$p =$ sandsynligheden for at en tilfældig fødsel resulterer i et drengebarn

Vores bedste gæt på p (maximum likelihood estimatet p-hat) er nu (naturligvis) *andelen* af drengebørn

$$\hat{p} = \frac{x}{n} = \frac{7}{8}$$

Umiddelbart ser det ud til, at der fødes flest drenge

Men: Det er jo *små tal*,
så kunne det ikke blot være sket ved en *tilfældighed*?
Vi opstiller (nul)hypotesen:

$H_0 : p = \frac{1}{2}$ (lige stor sandsynlighed for dreng og pige)

mod alternativet:

$H_A : p \neq \frac{1}{2}$
(Sandsynligheden for en drengefødsel er enten større eller mindre end for en pigefødsel)

Hvis vi kan afkræfte hypotesen H_0, har vi sandsynliggjort, at der er størst sandsynlighed for at føde drengebørn (fordi $\frac{7}{8} > \frac{1}{2}$)

Fremgangsmåde

Vi *forestiller os*, at H_0 er sand og ser om det fører til noget, der ligner en modstrid, dvs. noget som er meget/ekstremt usandsynligt.

P-værdien er netop sandsynligheden for at finde noget, der er mindst ligeså ekstremt - *hvis H_0 faktisk er sand.*

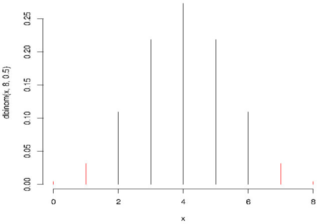
Hvis H_0 er sand, hvilke X'er vil vi da forvente at observere?

Formentlig nogle omkring 4 $(=\frac{8}{2})$.

Vi udregner fordelingen af X (antal drengebørn) under H_0

Fordelingen af X under H_0

$$X \sim \text{Bin}(n = 8, p = 0.5)$$




```
> x<-0:8
> cbind(x,cumsum(dbinom(x,8,0.5)))
      x
 [1,] 0 0.00390625
 [2,] 1 0.03515625
 [3,] 2 0.14453125
 [4,] 3 0.36328125
 [5,] 4 0.63671875
 [6,] 5 0.85546875
 [7,] 6 0.96484375
 [8,] 7 0.99609375
 [9,] 8 1.00000000
```

Har vi observeret noget ekstremt?

P-værdien for H_0 er *halesandsynligheden*

$$P(X \geq 7|H_0) + P(X \leq 1|H_0) = 0.07$$

Konklusion

Hvis H_0 er sand, har vi observeret noget *ret langt ude i højre hale*, men man vil dog i 7% af tilfældene observere noget mindst ligeså ekstremt, ved tilfældighedernes spil alene

Vi har altså ikke tilstrækkelig evidens for at forkaste H_0.

Måske fordi vi har for lidt information (for få børnebørn), eller *måske* fordi der ikke *er* nogen kønspræference i familien...

Vi skal huske **konfidensinterval**/**sikkerhedsinterval** for p
– men det lader vi programmet udregne:

$$\text{Eksakt:} \quad (0.474, 0.997)$$
$$\text{Approksimativt:} \quad (0.646, 1.000)$$

Datastruktur til brug for tabel-analyser

Enten 8 linier = 8 observationer, og 1 variabel:

```
D
D
P
D
D
D
D
D
```

eller 2 linier =
2 observationer og 2 variable

```
D 7
P 1
```

Hvis man i forvejen har et datasæt, man arbejder med, har man *den lange* version, men hvis man lige skal checke en tabel, er det lettest at skrive det ind som nogle enkelte linier (den korte version).

Hvordan gør man så i praksis?

Koden til indlæsning (af højre alternativ)
(med kommentarer til indlæsning af *det lange* alternativ):

```
data gender;
input type$ antal;
datalines;
D 7              /*  kunne ogsaa skrives som 8 linier  */
P 1              /*  7 linier med D'er og 1 med et P  */
;
run;
```

og så selve analysen:

```
proc freq data=gender;
   tables type / binomial(p=0.5);
exact binomial;
weight antal;   /*  fordi vi ikke skriver alle 8 linier  */
run;
```

Output fra Binomial-test af $p = 0.5$

```
                                    Cumulative    Cumulative
type     Frequency      Percent      Frequency      Percent
---------------------------------------------------------
D               7        87.50              7        87.50
P               1        12.50              8       100.00
```

```
 Binomial Proportion for type = D
-----------------------------------
Proportion (P)                0.8750
ASE                           0.1169
95% Lower Conf Limit          0.6458
95% Upper Conf Limit          1.0000

Exact Conf Limits
95% Lower Conf Limit          0.4735
95% Upper Conf Limit          0.9968

   Test of H0: Proportion = 0.5

ASE under H0                  0.1768
Z                             2.1213
One-sided Pr >  Z             0.0169
Two-sided Pr > |Z|            0.0339
```

```
Exact Test
One-sided Pr >=  P            0.0352
Two-sided = 2 * One-sided     0.0703

Sample Size = 8
```

Bemærk, at der er både et approksimativt og et eksakt konfidensinterval (se s. 32)

Konklusioner baseret på konfidensintervaller

Konfidensintervallet udtrykker de *trolige* værdier af den ukendte parameter, dvs.

"Intervallet indeholder med stor sandsynlighed den sande værdi"

Her ligger værdien $\frac{1}{2}$ i konfidensintervallet, og derfor er der ikke signifikant forskel på sandsynligheden for pigefødsel og drengefødsel. Men pas på:
Konklusionen afhænger også af intervallets størrelse!

- Er det snævert, kan man konkludere
- Er det bredt, kan man slet ikke konkludere noget!!

I eksemplet med de 8 børnebørn

Det eksakte konfidensinterval for sandsynligheden for et drengebarnebarn blev fundet til (0.474, 0.997).
Er sandsynlighederne for dreng og pige lige store?

- Måske, men det kan vi i hvert fald ikke slutte ud af denne undersøgelse!!
- Vi kan nemlig ikke afvise, at sandsynligheden for et drengebarnebarn er op imod 99.7%!

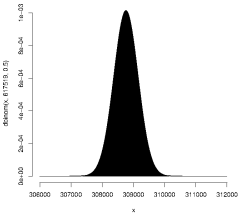
P-værdien (sandsynligheden for at få mindst så stor en skævhed bare ved et tilfælde, *hvis H_0 var sand* var 0.07, men konfidensintervallet er bredt, såe.....
Resultatet er inkonklusivt

I det store norske materiale

finder vi blandt i alt 617519 fødsler

- 299991 pigefødsler
- 317528 drengefødsler

og så skulle vi arbejde i fordelingen Bin$(617519, 0.5)$,
som har det meget normalfordelingslignende udseende:



Konfidensintervaller:

Eksakt:	$(0.5130, 0.5154)$
Approksimativt:	$(0.5130, 0.5154)$

Estimation i Binomialfordelingen

Hvis x ud af n er farveblinde,
estimerer vi sandsynligheden p for farveblindhed ved

Estimat: $\hat{p} = \frac{x}{n}$

For store n, og moderate p'er er s.e.$(\hat{p}) = \sqrt{\frac{p(1-p)}{n}}$
Med denne standard error finder vi nedenstående estimater for sandsynligheden for farveblindhed, med konfidensintervaller:

- Drenge: $\hat{p} = 0.04(0.016)$, CI=(0.0086, 0.0714)
- Piger: $\hat{p} = 0.0083(0.0083)$, CI=(-0.0080, 0.0246)

Disse er baseret på en normalfordelings-approksimation
og er ikke anvendelige for små forventede værdier (< 10)

Estimation i Binomialfordelingen, II

Eksakte konfidensintervaller

Hvis n ikke er så stor, eller hvis p er ret lille,
bør man ikke benytte formlen for standard error fra s. 39,
men i stedet benytte eksakte konfidensintervaller

For sandsynligheden for farveblindhed finder vi disse til

Piger: (0.0002, 0.0456)
Drenge: (0.0148, 0.0850)

Men her er det forskellen, der er interessant, så
vi skal se på en 2-gange-2-tabel

Farveblindhed vs. køn

Er farveblindhed lige hyppigt blandt drenge og piger?, dvs.
Er farveblindhed uafhængig af køn?

p_d Sandsynlighed for farveblindhed blandt drenge

p_p Sandsynlighed for farveblindhed blandt piger

Hypotese H_0: $p_d = p_p$

testes med

- Chi-i-anden test (χ^2-test),
  med mindre tabellerne er ret *tynde*. Så bruges i stedet
- Fishers eksakte test

som altså er test for uafhængighed

Hvad er en tynd tabel?

Det er en tabel med mindst én *forventet* værdi under 5...
men hvad er så en forventet værdi under H_0?

Det så vi på s. 13
Den totale frekvens af farveblindhed var 2.6% (7/270),
dvs. *vi forventer under* H_0:

- blandt 120 nyfødte piger: `120*0.026=3.1` farveblinde
- blandt 150 nyfødte drenge: `150*0.026=3.9` farveblinde

Begge disse er under 5, så vi bør ikke bruge χ^2-testet

Brug Fishers eksakte test i stedet for
– det skader aldrig

Datastruktur for 2*2-tabeller

Egentlig skal vi i eksemplet om farveblindhed have 270 linier, en for hvert barn, men da der kun findes 4 forskellige *typer* af børn, kan vi nøjes med

- 4 observationer=linier
- 3 kolonner
  - kønnet (pige/dreng)
  - farveblind (ja/nej)
  - antallet af den pågældende type barn

```
pige  nej 119
pige   ja   1
dreng nej 144
dreng  ja   6
```

Vægtning eller ej?

Når man taster data ind på denne *korte facon*, skal alle analyser herefter udføres vægtet, se kode s. 45

Hvis alle 270 linier haves (typisk, når man har data fra sit eget studie), skal man *ikke* vægte.

Se mere s. 45

To-gange-to tabeller i praksis

Kode til *den korte* data-facon, inclusive diverse options:

```
proc freq data=farveblind;
   tables gender*farveblind
   / nopercent nocol chisq expected riskdiffc relrisk;
exact riskdiff;
weight antal;
run;
```

- `nopercent nocol`: Vi vil kun have række-procenter
- `chisq`: giver et χ^2-test for uafhængighed, se s. 47
- `expected`: giver forventede antal i tabellen (s. 46) (så vi kan checke, at disse ikke er for små)
- `riskdiff`: giver estimat for forskellen $p_d - p_p$, (s. 50)
- `relrisk`: giver relativ risiko og odds ratio, (s. 53)

Tabel med observerede og forventede værdier

samt rækkeprocenter (kode s. 45:

```
gender     farveblind

Frequency|
Expected |
Row Pct  |ja      |nej     |  Total
---------+--------+--------+
dreng    |      6 |    144 |    150
         | 3.8889 | 146.11 |
         |   4.00 |  96.00 |
---------+--------+--------+
pige     |      1 |    119 |    120
         | 3.1111 | 116.89 |
         |   0.83 |  99.17 |
---------+--------+--------+
Total           7      263      270
```

Sørg for, at grupperne er rækker, og outcomes er søjler

Output: sammenligning af hyppigheder

```
Statistics for Table of gender by farveblind

Statistic                      DF       Value      Prob
------------------------------------------------------
Chi-Square                      1      2.6472    0.1037
Likelihood Ratio Chi-Square     1      3.0022    0.0832
Continuity Adj. Chi-Square      1      1.5418    0.2144
                                                      /_______________
 WARNING: 50% of the cells have expected counts less /
          than 5. Chi-Square may not be a valid test. \_______________
                                                      \

       Fisher's Exact Test
----------------------------------
Cell (1,1) Frequency (F)         6
Left-sided Pr <= F          0.9847
Right-sided Pr >= F         0.1047

Table Probability (P)       0.0894     /
Two-sided Pr <= P           0.1364    --------------P-value
                                       \
```

Der skal altid anvendes 2-sidede tests, med mindre man har en *afsindig god* begrundelse.

*Teknisk note

Der findes forskellige raffinementer til χ^2-teststørrelsen

- Continuity adjusted:
  Forbedret approksimation for ret tynde tabeller
- Likelihood ratio test:
  som jo er det, vi plejer at bruge
  men som regneteknisk er sværere at regne ud
  (betyder ikke så meget mere)
- Fishers eksakte test (se s. 42),
  som *altid* kan bruges
-

Her er alle teststørrelser rimeligt enige om konklusionen →

Foreløbig konklusion

Der er *ikke* signifikant forskel på drenge og piger mht farveblindhed.

Kan vi så konkludere, at farveblindhed ikke afhænger af køn?

Nej..., der kunne jo være tale om en Type 2 fejl.

Vi skal kvantificere den ukendte forskel

Kvantificering af differensen $p_d - p_p$

Estimatet er naturligvis $\hat{p}_d - \hat{p}_p = 0.0400 - 0.0083 = 0.0317$, men vi skal også have et konfidensinterval (CI), se option `riskdiffc` og linien `exact riskdiff` i koden s. 45:

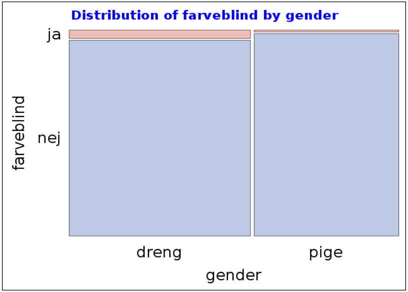
```
                    Column 1 Risk Estimates

                                 (Asymptotic) 95%      (Exact) 95%
                 Risk       ASE  Confidence Limits    Confidence Limits
-------------------------------------------------------------------------
Row 1          0.0400    0.0160    0.0053    0.0747    0.0148    0.0850
Row 2          0.0083    0.0083    0.0000    0.0288    0.0002    0.0456
Total          0.0259    0.0097    0.0051    0.0467    0.0105    0.0527

Difference     0.0317    0.0180   -0.0112    0.0745   -0.0890    0.1517
                      Difference is (Row 1 - Row 2)
```

Estimeret forskel: 0.0317 (svarende til 3.17 procentpoint), med konfidensinterval CI=(-0.0112, 0.0745)

Ud af 100 nyfødte er der rimeligvis et sted mellem 1.12 flere piger end drenge, men op til 7.5 flere drenge end piger, der er farveblind

Alternative mål for forskel

- Relativ risiko for farveblindhed: $\frac{\hat{p}_d}{\hat{p}_p} = \frac{0.0400}{0.0083} = 4.82$
- Odds ratio: $\frac{\hat{p}_d/(1-\hat{p}_d)}{\hat{p}_p/(1-\hat{p}_p)} == \frac{0.0400/0.9600}{0.0083/0.9917} = \frac{0.0417}{0.0084} = 4.96$



Relativ risiko for ikke-farveblindhed?

Relativ risiko

er forholdet eller ratio mellem de to estimerede sandsynligheder:

$$\frac{\hat{p}_d}{\hat{p}_p} = \frac{0.0400}{0.0083} = 4.82 \ \ (\text{faktisk } 4.80)$$

Det er næsten 5 gange så hyppigt for en dreng at være farveblind i forhold til for en pige

Relativ risiko afhænger af, hvad der udvælges til at være 1-kategorien:

$$\text{RR(respons=1)} \neq \frac{1}{\text{RR(respons=0)}}$$

Her er relativ "risiko" (chance) for at være normalt-seende: $\frac{1-0.0400}{1-0.0083} = 0.97$, for drenge vs. piger

Konfidensgrænser

er besværlige at regne i hånden, brug option `relrisk` fra koden s. 45, der giver outputtet

```
          Estimates of the Relative Risk (Row1/Row2)

Type of Study                   Value       95% Confidence Limits
-----------------------------------------------------------------
Case-Control (Odds Ratio)      4.9583        0.5887       41.7605
Cohort (Col1 Risk)             4.8000        0.5858       39.3291
Cohort (Col2 Risk)             0.9681        0.9333        1.0041
```

Vi ved fra s. 46, at
`Row1` er drenge, `Row2` er piger.
`Col1` er "ja", `Col2` er "nej".

Bemærk, at usikkerheden er stor (brede konfidensgrænser) for værdierne langt fra 1.

* Odds

Hvis p angiver sandsynligheden for en begivenhed, defineres den tilsvarende odds som $\frac{p}{1-p}$, altså forholdet mellem sandsynligheden for "at det sker" og sandsynligheden for "at det ikke sker".

Odds angiver det forventede forhold mellem antal vundne og antal tabte væddemål (f.eks. i et hesteväddeløb) og samtidig også hvor meget man ville vinde ved en given satsning
(hvis ellers det var *fair game*).

Sandsynlighed p	Odds p/(1-p)	Sats, vind(+sats)
0.01	1/99 - 1:99	Sats 1, vind 99(+1)
0.05	1/19 - 1:19	Sats 1, vind 19(+1)
0.1	1/9 - 1:9	Sats 1, vind 9(+1)
0.2	1/4 - 1:4	Sats 1, vind 4(+1)
0.25	1/3 - 1:3	Sats 1, vind 3(+1)
0.5	1 - 1:1	Sats 1, vind 1(+1)

Odds ratio

Odds for farveblindhed er, for

Drenge: $p_d = 0.0400 \Rightarrow \frac{p_d}{1-p_d} = 0.0417$

Piger: $p_p = 0.0083 \Rightarrow \frac{p_p}{1-p_p} = 0.0084$

Odds ratio for farveblindhed, for drenge vs. piger:

$$\frac{p_d/(1-p_d)}{p_p/(1-p_p)} = \frac{0.0400/0.9600}{0.0083/0.9917} = \frac{0.0417}{0.0084} = 4.96$$

Odds for at en dreng er farveblind er næsten 5 gange så høj som for en pige (output s. 53).

Odds ratio, fortsat

Odds for ikke-farveblindhed er, for

Drenge: $1 - p_d = 0.9600, \quad \frac{1-p_d}{p_d} = 24$

Piger: $1 - p_p = 0.9917, \quad \frac{1-p_p}{p_p} = 119.5$

Odds ratio for ikke-farveblindhed, for drenge vs. piger:

$$\frac{(1 - p_d)/p_d}{(1 - p_p)/p_p} = \frac{0.9600/0.0400}{0.9917/0.0083} = \frac{24}{119.5} = 0.20$$

Odds for at en dreng er ikke-farveblind er ca. en femtedel af den tilsvarende for en pige

Vi konstaterer, at odds ratio er symmetrisk:

$\text{OR(respons=1)} = \frac{1}{\text{OR(respons=0)}}$

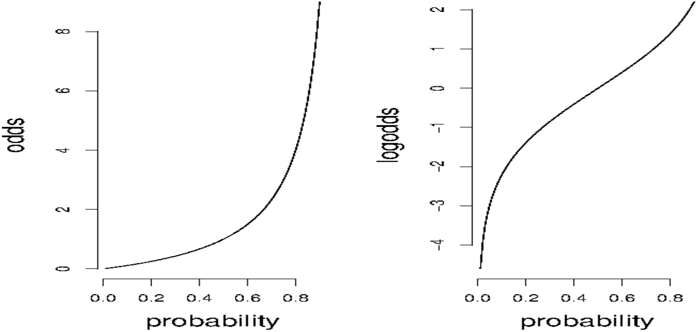
Hvad skal vi med odds?

når de nu er lidt svære at forstå...

De har nogle gode egenskaber
– i hvert fald, når man logaritmerer dem:

- Odds kan blive vilkårligt store
  de er ikke begrænset af 1 som sandsynligheder
  men de er begrænset af 0 nedadtil
- Log odds: $\log(\frac{p}{1-p})$ =logit(p)
- Log odds er ubegrænset, både opadtil og nedadtil
  og er derfor gode at anvende, når man skal modellere lineære effekter (logistisk regression, mere om det senere)
- De er *helt essentielle* i forbindelse med case-control studier

Odds og Logodds



- Odds er nedadtil begrænset af 0
- Log odds er ubegrænset

Konklusion om farveblindhed

- Måske tendens til større forekomst af farveblindhed hos drenge
- men vi kan ikke konkludere med noget, der ligner sikkerhed ud fra så lille et materiale
- Forskellen *kan* tænkes at være ganske betragtelig (op til ca. en faktor 40)

Måske skulle vi designe en ny (og noget større) undersøgelse.... gerne i form af en case-control undersøgelse, som er meget stærkere (se s. 62-67)

Dimensionering af 2-gange-2 tabel

Hvis nu faktisk *de sande* frekvenser af farveblindhed for piger og drenge er 1% hhv. 4%, hvor mange børn skal vi da undersøge for at kunne påvise dette med en rimelig styrke (power)?
Se kode s. 91

```
The POWER Procedure
Fisher's Exact Conditional Test for Two Proportions

         Computed N Per Group
            Nominal     Actual     N Per
   Index      Power      Power     Group
       1        0.8      0.801       455
       2        0.9      0.900       585
```

Altså omkring 500 børn af hvert køn
Hvad, hvis forskellen ikke er helt så stor, er det så helt uinteressant?

Hvilken forskel vil vi nødigt overse?

F.eks. en 50% øget hyppighed blandt drengebørn?
Eller rettere (for senere sammenligning) en odds ratio på 1.5 (kode s. 91)

```
The POWER Procedure
Fisher's Exact Conditional Test for Two Proportions

Reference (Group 1) Proportion        0.01
Odds Ratio                            1.5

          Computed N per Group

            Nominal     Actual     N per
   Index      Power      Power     Group
       1        0.8      0.800      8291
       2        0.9      0.900     10964
```

Altså helt oppe omkring 10000 børn af hvert køn!

Case-control studier

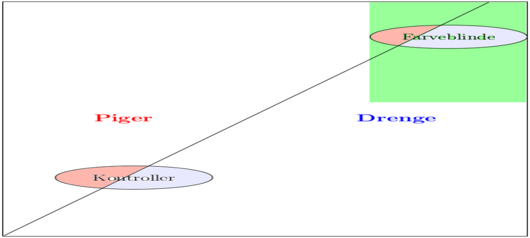
Da der er meget få farveblinde, vil det være en fordel at *vende problemstillingen om*:

- Registrer alle tilfælde af farveblindhed over f.eks. 1 år, og studer kønsfordelingen
- Udvælg kontroller (fra samme år/område) og sammenlign med kønsfordelingen blandt disse
  (eller sammenlign med en kendt kønsratio....)

Tænkt eksempel: 120 farveblinde og 120 kontroller
(et lidt mindre omfang end det nuværende,
men selvfølgelig med *langt flere* farveblinde)

Case-control sampling

Den grønne firkant symboliserer farveblinde børn,
og ellipserne symboliserer vores sample:



- Der er generelt lidt flere drenge end piger
- Blandt de farveblinde er der en *stor overvægt* af drenge, og det opdager vi med større styrke, fordi vi sampler mange farveblinde

Case-control studier, tænkt tabel

```
Table of farveblind by gender

farveblind     gender

Frequency|
Row Pct  |
Col Pct  |dreng   |pige    |  Total
---------+--------+--------+
ja       |     99 |     21 |    120
         |  82.50 |  17.50 |
         |  61.49 |  26.58 |
---------+--------+--------+
nej      |     62 |     58 |    120
         |  51.67 |  48.33 |
         |  38.51 |  73.42 |
---------+--------+--------+
Total         161       79      240
```

Bemærk, at det nu er farveblindhed, der er sat som rækker

Case-control studier, II

Koder svarer helt til de tidligere, se s. 45

```
Statistics for Table of farveblind by gender

                  Odds Ratio and Relative Risks

Statistic                          Value       95% Confidence Limits
--------------------------------------------------------------------
Odds Ratio                        4.4101        2.4409        7.9681
Relative Risk (Column 1)          1.5968        1.3183        1.9341
Relative Risk (Column 2)          0.3621        0.2355        0.5567

Sample Size = 240
```

Sammelign til det større studie, s. 53. Her får vi:

- Væsentligt smallere CI for odds ratio
- Relativ risiko har ingen mening her:
  Det er *relativ risiko for at være dreng...*

Studier af sjældne fænomener (p lille)

Når fænomenet er sjældent har man brug for at lave case-control studier for at undgå vanvittigt store sample sizes

og så kan man ikke mere estimere relativ risiko,
men kun odds ratio.

Heldigvis gælder det, at:

Når p er lille, er Odds $\approx p$

og dermed Odds Ratio $\approx$ Relativ Risiko

Dimensionering af case-control studier

"Udfaldet" er så at betragte som "*at være dreng*"
og frekvensen af dette udfald under H_0 er ca 50%=0.5.
Hvis vi igen siger, at vi nødigt vil overse en odds ratio på 1.5 (se s. 61), kan vi dimensionere, som det ses på koden s. 92

```
The POWER Procedure
Fisher's Exact Conditional Test for Two Proportions
Reference (Group 1) Proportion                     0.5
Odds Ratio                                         1.5

         Computed N per Group

            Nominal     Actual     N per
   Index      Power      Power     Group
       1        0.8      0.800       407
       2        0.9      0.900       539
```

Det hjalp jo en hel del: Ca. 500 mod ca. 10000 på s. 61

Kategoriske variable (Class variable)

Vi skelner mellem forskellige typer:

- Dikotom (binær): To kategorier
  - Komplikation (ja/nej), Farveblindhed (ja/nej)
- Nominal: Flere kategorier
  - Behandling, Operationstype
  - Øjenfarve: blå, brune, grønne
- Ordinal: Flere ordnede kategorier
  - Smertegrad (ingen, let, moderat, kraftig), tumorstadium

Kan optræde som såvel outcome som kovariater,
nominal dog oftest som kovariat

Større tabeller (nominal kovariat vs. binært outcome)

Komplikationer ved forskellige typer af operationer:

	Komplikation?			
Operationstype	Nej	Ja	Total	Risiko (SE)
Gynækologisk	235	5	240	0.021 (0.009)
Abdominal	210	35	245	0.143 (0.022)
Ortopædisk	200	6	206	0.029 (0.012)
Total	645	46	691	0.067 (0.009)

Er der forskel på komplikationssandsynlighederne?

Spørgsmålet er noget vagere end for 2 gange 2 tabeller:
Hvis der er forskel, hvor er det så?
Testet bliver ikke så stærkt, og forskelle kan *gemme sig*

Datastruktur - igen

- Hvis man arbejder med individ-data med i alt 691 linier, kan man umiddelbart lave sin tabel-analyse:

  ```
  proc freq data=langt;
     tables type*komplikation /
            nopercent nocol chisq cellchisq expected;
  run;
  ```

  Husk, at grupperne=operationstyperne skal være rækker!
  – og at det er *rækkeprocenterne*, der er interessante
- men hvis man vil lave analysen ud fra en allerede eksisterende tabel, f.eks. fra en artikel, må man skrive som angivet på næste side.

Datastruktur - når man bare har tabellen

- 6 linier, svarende til de 3*2 forskellige typer af patienter
- 3 kolonner:
  - operationstype
  - komplikation ja/nej
  - antal patienter af denne type

```
data operation;
input type$ komplikation$ antal;
datalines;
Gynecological ja 5
Gynecological nej 235
Abdominal ja 35
Abdominal nej 210
Orthopedic ja 6
Orthopedic nej 200
;
run;

proc freq data=operation;
   tables type*komplikation /
          nopercent nocol chisq cellchisq expected;
weight antal;
run;
```

Output (kode fra forrige side)

```
type            komplikation

Frequency       |
Expected        |
Cell Chi-Square|
Row Pct         |0       |1       |  Total
----------------+--------+--------+
Abdomina        |    210 |     35 |    245
                | 228.69 |  16.31 |
                | 1.5275 | 21.418 |
                |  85.71 |  14.29 |
----------------+--------+--------+
Gynecolo        |    235 |      5 |    240
                | 224.02 | 15.977 |
                | 0.5379 | 7.5416 |
                |  97.92 |   2.08 |
----------------+--------+--------+
Orthoped        |    200 |      6 |    206
                | 192.29 | 13.713 |
                | 0.3094 | 4.3386 |
                |  97.09 |   2.91 |
----------------+--------+--------+
Total                645       46      691
```

Bemærk:

Ingen forventede værdier under 13

Output, fortsat

(kode s. 71)

```
Statistics for Table of type by komplikation

Statistic                     DF       Value      Prob
------------------------------------------------------
Chi-Square                     2     35.6734    <.0001
Likelihood Ratio Chi-Square    2     34.3203    <.0001

Sample Size = 691
```

χ^2-testet giver en kraftig forkastelse af hypotesen om uafhængighed, dvs. der er en forskel på komplikationssandsynlighederne for de tre operationstyper.

Bidragene til χ^2-størrelsen (`Cell Chi-Square` på s. 72) viser, at der specielt er
for mange komplikationer i abdominal-gruppen.

Eksempel om behandling af nyresten

Outcome: Succes?
Kovariat: Behandling...(*og en mere lige om lidt*)

	Succes?		
	nej	ja	Total
Behandling A	77	273	350
Behandling B	61	289	350
Total	138	562	700

Odds ratio for succes for behandling B vs. behandling A: 1.34 (0.92, 1.94)

Odds ratio for succes for behandling A vs. behandling B: 0.75 (0.51, 1.09)

Behandling B ser ud til at være bedre end behandling A

Nu opdeler vi efter nyrestenens størrelse

svarende til 2 kovariater: Behandling og stenens størrelse

Små sten:

	Succes?		
	nej	ja	Total
Behandling A	6	81	87
Behandling B	36	234	270
Total	42	315	357

Store sten:

	Succes?		
	nej	ja	Total
Behandling A	71	192	263
Behandling B	25	55	80
Total	96	247	343

Odds ratio for succes for behandling A vs. behandling B:

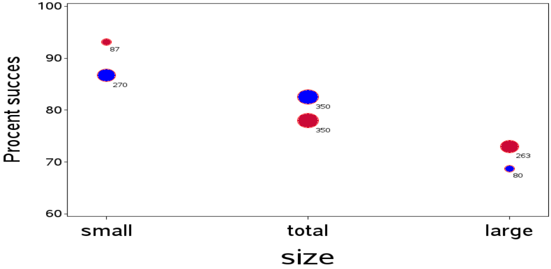
- små sten: 2.08 (0.84, 5.11)
- store sten: 1.23 (0.71, 2.12)

Behandling A ser ud til at være bedre end behandling B

Simpsons paradoks

Behandling A er bedst for både små og store sten
- men *ikke*, når vi ser på den totale population!!

Confounding:
Der er sammenhæng mellem stenstørrelse og succesrate: De store sten behandles fortrinsvis med A, og derfor ser A værst ud, totalt set:



Nyt eksempel: Påvisning af tuberkulose

Spytprøver fra 210 patienter (mistænkt for tuberkulose) dyrkes i et substrat (A). Hvis der er vækst, kaldes prøven positiv.
Vi har også et facit: syg/rask

	Positiv dyrkning?		
	nej	ja	Total
Tuberkulose facit			
syg	18	32	50
rask	137	23	160
Total	155	55	210

Test for uafhængighed?

- Det er der næppe
- — og det er ikke relevant!

Hvad er det så, vi vil vide?

Sensitivitet, specificitet mv.

Sensitivitet: 32/50=0.64
Hvor godt detekteres de syge?
Vi misser 36%

Specificitet: 137/160=0.86
Hvor ofte frikendes de raske?
Vi har 14% risiko for at få et falsk positivt svar

Positiv prediktiv værdi: 32/55=0.58
Hvor ofte kan vi stole på en positiv test?
En positiv test er kun udtryk for sygdom i lidt over halvdelen af tilfældene

Negativ prediktiv værdi: 137/155=0.88
Hvor ofte kan vi stole på en negativ test?
Vi kan ikke føle os helt sikre, selv om prøven er negativ, der er 12% risiko for, at det er en falsk negativ

Sammenligning af to substrater, A og B

Vi ser nu kun på spytprøver fra de 50 tuberkulosepatienter (de syge fra tabel s. 77). Disse dyrkes i både substrat A og B.

A	B	Antal patienter/prøver
+	+	20
+	-	12
-	+	2
-	-	16
I alt		50

Er substraterne lige effektive til at finde de tilstedeværende tuberkelbakterier?

Sensitivitet for substrat B: 22/50=0.44, altså lavere end for A

men det er parrede data!

Hvad er spørgsmålet?

Er der forskel på de to substraters sensitivitet:

altså evnen til at detektere bakterierne i prøven, dvs. sandsynligheden for en positiv prøve.

```
data tuberkulose;
input A$ B$ antal;
datalines;
+ + 20
+ - 12
- + 2
- - 16
;
run;
```

Test af ens sensitiviteter

- Vi skal undersøge, om der er ens marginal-procenter = sensitiviteter
- Der er *ikke* tale om et test for uafhængighed!
  men et Mac Nemar test

```
proc freq data=tuberkulose;
   tables A*B / norow nocol;   /
   exact mcnem;               -------------------------
weight antal;                  \
run;
```

Output fra analysen

fra koden på forrige side

```
The FREQ Procedure

Table of A by B

A         B

Frequency|
Percent  |+       |-       |  Total
---------+--------+--------+
+        |     20 |     12 |     32
         |  40.00 |  24.00 |  64.00
---------+--------+--------+
-        |      2 |     16 |     18
         |   4.00 |  32.00 |  36.00
---------+--------+--------+
Total          22       28       50
            44.00    56.00   100.00
```

Sensitiviteter:

$$p_A = 0.64$$

$$p_B = 0.44$$

Output, fortsat

McNemar test, hypotese $p_A = p_B$:

```
Statistics for Table of A by B

       McNemar's Test
----------------------------
Statistic (S)         7.1429
DF                         1
Asymptotic Pr >  S    0.0075       /
Exact      Pr >= S    0.0129      ------------------
                                   \
Sample Size = 50
```

Bemærk, at dette er noget helt andet end et test for uafhængighed (som slet ikke er relevant i denne sammenhæng).

Det ville ikke være så godt, hvis de to metoder var uafhængige....

Konklusion

Der er signifikant forskel på de dyrkningsmetoder:
A er signifikant bedre end B ($P = 0.0129$)

A-metoden opdager 20% flere bakterier ($0.64 - 0.44 = 0.20$), med konfidensgrænser (0.064, 0.336), dvs. svarende til et sted mellem 6% og 34% større chance for detektion.

Disse konfidensgrænser kræver en ret kompliceret beregning, som ligger udenfor dette kursus.

APPENDIX

med SAS-programbidder svarende til nogle af slides

- Test i Binomialfordelingen, s. 86
- 2×2-tabeller, s. 87-90
- Dimensionering, s. 91-92
- 3×2-tabel, s. 93
- Parrede binære data, s. 94

Test i Binomialfordelingen

Slide 33-34

```
data gender;
input type$ antal;
datalines;
D 7              /*  kunne ogsaa skrives som 8 linier  */
P 1              /*  7 linier med D'er og 1 med et P  */
;
run;

proc freq data=gender;
   tables type / binomial(p=0.5);
exact binomial;
weight antal;    /*  fordi vi ikke skriver alle 8 linier  */
run;
```

Datastruktur for tabeller

Slide 43 og 45

Egentlig skal vi i eksemplet om farveblindhed have 270 linier, en for hvert barn, men da der kun findes 4 forskellige *typer* af børn, kan vi nøjes med 4 linier, samt en kolonne med antallene:

```
data farveblind;
input gender$ farveblind$ antal;
datalines;
pige nej 119
pige ja 1
dreng nej 144
dreng ja 6
;
run;
```

I alle analyser skal man så bare huske at benytte vægtning:

```
weight antal;
```

Tabel med observerede og forventede værdier

samt

- rækkeprocenter
- test for uafhængighed

Slide 46-47

```
proc freq data=farveblind;
   tables gender*farveblind /
          chisq nopercent nocol expected;
weight antal;
run;
```

Kvantificering af differensen $p_d - p_p$

Slide 50

Brug option `riskdiff` i `tables`-sætningen:

```
proc freq data=farveblind;
   tables gender*farveblind / riskdiff;
weight antal;
run;
```

Relativ risiko og odds ratio i SAS

Slide 53

Brug option `relrisk` i `tables`-sætningen:

```
proc freq data=farveblind;
   tables gender*farveblind / relrisk;
weight antal;
run;
```

Dimensionering af 2-gange-2 tabel

Slide 60

```
proc power;
        twosamplefreq test=fisher
        groupproportions = (.01 .04)
        npergroup = .
        power = 0.8,0.9;
   run;
```

Slide 60

```
proc power;
        twosamplefreq test=fisher
        relativerisk = 1.5
        refproportion = 0.01
        npergroup = .
        power = 0.8,0.9;
   run;
```

Dimensionering af case-control undersøgelse

Slide 67

```
proc power;
        twosamplefreq test=fisher
        oddsratio = 1.5
        refproportion = 0.5
        npergroup = .
        power = 0.8,0.9;
   run;
```

Man kan evt. benytte denne hjemmeside:
http://sampsize.sourceforge.net/iface/s3.html#cc

Større tabel - 3×2

Slide 71-73

```
data operation;
input type$ komplikation$ antal;
datalines;
Gynecological ja 5
Gynecological nej 235
Abdominal ja 35
Abdominal nej 210
Orthopedic ja 6
Orthopedic nej 200
;
run;

proc freq data=operation;
   tables type*komplikation /
          nopercent nocol chisq cellchisq expected;
weight antal;
run;
```

Bemærk brugen af `cellchisq`, der udskriver bidraget til χ^2-størrelsen for hver enkelt celle

Parrede binære data

Slide 79ff

```
data tuberkulose;
input A$ B$ antal;
datalines;
+ + 20
+ - 12
- + 2
- - 16
;
run;

proc freq data=tuberkulose;
   tables A*B / norow nocol;   /
   exact mcnem;               --------------------------
weight antal;                  \
run;
```