Det Natur- og Biovidenskabelige Fakultet

Binomialfordelingen: egenskaber, estimation og test

Anders Tolver
Institut for Matematiske Fag

I dag

Dagens forelæsning dækkes af lærebogens kapitel 11

- Hvilken slags problemer skal vi se på?
- Binomialfordelingen (med kendt sandsynlighed)
- Statistik for en enkelt binomialfordeling
- Statistik for to binomialfordelinger (estimation og KI)

Generel info:

- **Afleveringsopgaver:** opgave 3 afleveres elektronisk senest onsdag (26/10). Husk, at du kan spørge hjælpelærerne om rettelser, som du ikke forstår (også til opgave 1+2).
- **Kursusevaluering:** Udfyld som minimum *multiple choice* rubrikker, men skriv gerne kommentarer, hvis du har noget på hjerte.

Hvilken slags problemer skal vi se på?

Eksempel: Binomialdata

Antag at vi ved at en bestemt slags frø spirer med ssh. 60%.

- Hvis vi ser på 8 frø, hvor stor er så sandsynligheden for at mindst 5 af dem spirer?
- Skal designe et forsøg, hvor der skal bruges mindst 10 planter.

  Hvor mange frø skal der plantes, hvis vi vil være 90% sikre på at mindst 10 frø bliver til noget?

Vi skal bruge **binomialfordelingen.**

Men som regel vil sandsynligheden ikke være kendt! Data $\rightarrow$ estimat for sandsynlighed, konfidensinterval, hypotesetest.

Eksempel vedr. forelæserens fritidsinteresser:

- Implicit bias over for undervisere i statistik!

Eksempel: Tabeller

Data fra 100 mus: Har kastrerede mus større risiko for at udvikle diabetes end ikke-kastrerede mus? Mere om det i dag og onsdag.

	Diabetes	Ikke diabetes	Total
Katrerede mus	26	24	50
Ikke-kastrerede mus	12	38	50

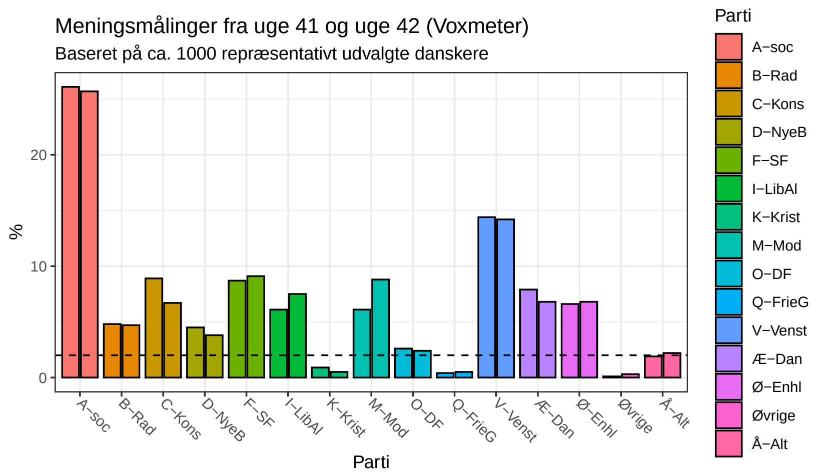
Svar fra 1000 personer vedr. politisk ståsted og foretrukket finansøkonomisk redskab: Er der sammenhæng? På onsdag.

	Demokrat	Republikaner	Uafhængig
Begrænse udgifter	101	282	61
Øge skatter	38	67	25
Øge offentlige invest.	131	88	31
Lade underskuddet vokse	61	90	25

Politiske meningsmålinger (kilde: voxmeter.dk)

```
##       parti uge41 (%) uge41 (n) uge42 (%) uge42 (n)
## 1     A-soc      26.1       263      25.7       242
## 2     B-Rad       4.8        48       4.7        44
## 3    C-Kons       8.9        90       6.7        63
## 4    D-NyeB       4.5        45       3.8        36
## 5      F-SF       8.7        88       9.1        86
## 6   I-LibAl       6.1        61       7.5        71
## 7   K-Krist       0.9         9       0.5         5
## 8     M-Mod       6.1        61       8.8        83
## 9      O-DF       2.6        26       2.4        23
## 10  Q-FrieG       0.4         4       0.5         5
## 11  V-Venst      14.4       145      14.2       134
## 12    Æ-Dan       7.9        80       6.8        64
## 13   Ø-Enhl       6.6        67       6.8        64
## 14   Øvrige       0.1         1       0.3         3
## 15    Å-Alt       1.9        19       2.2        21
```


Politiske meningsmålinger



Valgtema: politiske meningsmålinger

Hvad er den aktuelle vælgertilslutning til de forskellige politiske partier?

- n: antal personer i stikprøven
- p: ukendt vælgertilslutning i befolkningen
- Usikkerhed og konfidensinterval (KI) for p
- Test for hypotesen $H_0 : p = p_0$ (fast værdi/tidligere valgresultat)

Er der forskel på vælgertilslutningen på to forskellige tidspunkter?

- n_1, n_2: størrelse af to (uafhængige) stikprøver
- p_1, p_2: ukendt vælgertilslutning på to tidspunkter
- Usikkerhed/KI for p_1 og p_2
- Usikkerhed/KI for forskel/ændring: $p_2 - p_1$
- Test af hypotesen $H_0 : p_1 = p_2$

Fællestræk

Fælles for eksemplerne er at data består af **antal**.

Vi kan ikke bruge normalfordelingen. I stedet:

- I dag: **Binomialfordelingen**
- Onsdag: **Tabeldata**

I har nok set en del allerede i gymnasiet, men nu har I bedre forudsætninger for at forstå hvad der foregår og hvorfor.

På mange måder meget **nemmere** end normalfordelingsanalyserne!

Måske lidt forvirrrende fordi man ofte kan gøre flere forskellige ting i R, som alle er fornuftige men ikke giver præcis samme resultater.

Binomialfordelingen

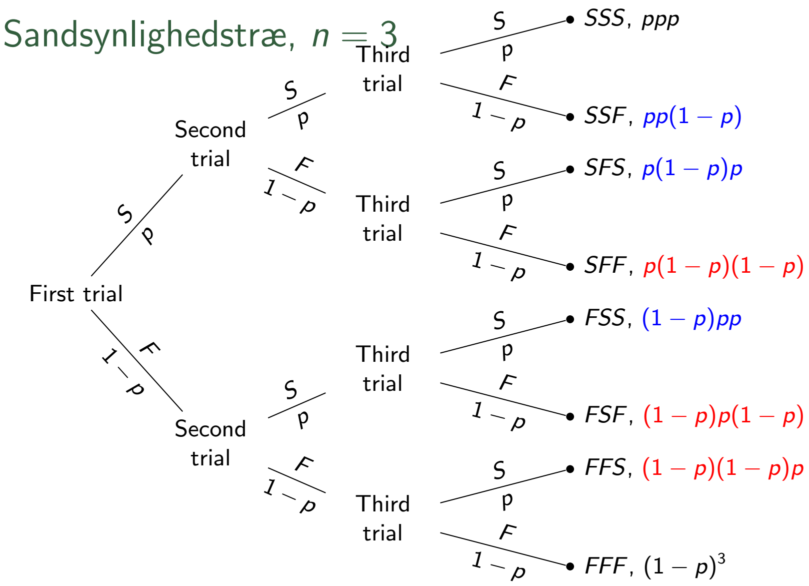
Eksempel: spiring af frø

Antag at vi ved, at et frø har en sandsynlighed på 60% for at spire.

Betragt tre frø.

- Hvad er sandsynligheden for at netop et frø spirer?
- Hvad er sandsynligheden for at mindst et frø spirer?

Sandsynlighedstræ, $n = 3$



Spørgsmål

- Hvilke antagelser lå egentlig bag beregningerne?
- Hvordan formaliserer vi sandsynlighedsberegningerne, så vi også kan klare et større antal frø?
- Hvordan beregner vi sandsynligheder i R?

Independent trials

Independent trials / uafhængige gentagelser:

- ***n* gentagelser** af simpelt eksperiment
- Hver gentagelser har **to mulige udfald**: succes/fiasko

  Kan være hvad som helst: død/levende, spiret eller ikke, prisen stiger/falder, korrekt/forkert, osv.
- **Samme sandsynlighed** for succes i hver gentagelse: p
- Gentagelserne er **uafhængige**

Binomialfordelingen

Lad Y betegne antallet af succeser fra n uafhængige forsøg med samme successandsynlighed p.

Så er Y **binomialfordelt** med antalsparameter (engelsk: size) n og sandsynlighedsparameter p. Vi skriver $Y \sim \text{bin}(n, p)$.

Binomialsandsynlighederne er givet ved

$$P(j \text{ ``succeser''}) = P(Y = j) = \binom{n}{j} \cdot p^j \cdot (1-p)^{n-j},$$

hvor **binomialkoefficienten** — antal måder man kan vælge j dimser ud af n dimser — er givet ved

$$\binom{n}{j} = \frac{n!}{j!(n-j)!}$$

Eksempel: Spiring af frø

$n = 3$ frø der hver især har en sandsynlighed på 60% for at spire.

- $P(Y = 1)$ og $P(Y \geq 1)$, nu vha. formlen og i R
- Hvad er ssh. for at højst to frø spirer, altså $P(Y \leq 2)$?

Hvis der i stedet er 8 frø: Hvad er sandsynlighederne så?

R

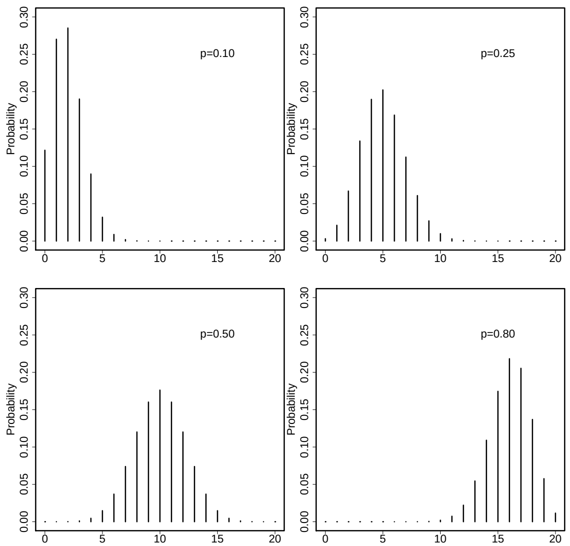
Binomialsandsynligheder i R.

- Sandsynligheder $P(Y = j)$ beregnes med `dbinom`
- Sandsynligheder $P(Y \leq j)$ beregnes med `pbinom`.

I begge dele skal `size` (dvs. n) og `prob` (dvs. p) angives.

```
dbinom(1, size=3, prob=0.6)    ## P(Y=1)

## [1] 0.288

pbinom(2, size=3, prob=0.6)    ## P(Y<=2)

## [1] 0.784

dbinom(0, size=3, prob=0.6)    ## P(Y=0)

## [1] 0.064

1-dbinom(0, size=3, prob=0.6)  ## P(Y>=1)

## [1] 0.936
```


Binomialfordelinger, her med $n = 20$



Eksempel: Spiring af frø

Nyt spørgsmål:

- Skal designe et forsøg, hvor der skal bruges mindst 10 planter.
- Hvor mange frø skal der plantes, hvis vi vil være mindst 90% sikre på at mindst 10 frø bliver til noget?

  Se dagens R-kode!

Middelværdi og varians for binomialfordelinger

For en binomialfordelt variabel $Y \sim \text{bin}(n, p)$ gælder:

Middelværdien er

$$EY = n \cdot p$$

Spredningen

$$\text{sd}(Y) = \sqrt{n \cdot p \cdot (1 - p)}$$

Se figurerne fra før.

Normalfordelingsapproksimation

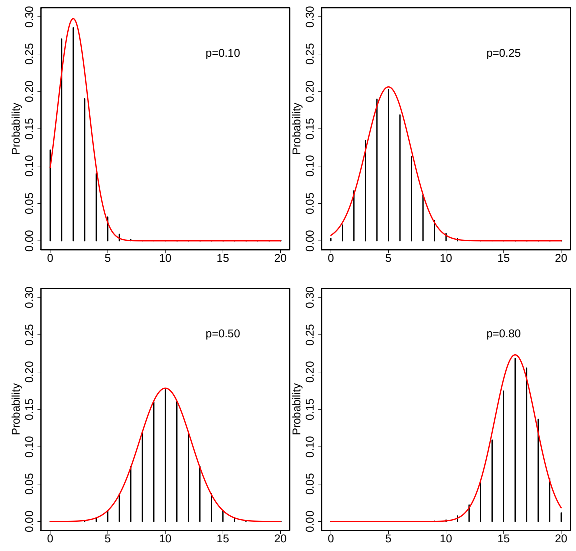
En binomialfordelt variabel er en sum af n 0/1-variable.

Den centrale grænseværdisætning giver **normalfordelingsapprokimation:**

- $\text{bin}(n, p)$ kan approksimeres med $N\big(np, np(1-p)\big)$, dvs. normalfordelingen med den korrekte middelværdi og spredning
- Tommelfingerregel: Approksimation er "god" hvis både $np \geq 5$ og $n(1-p) \geq 5$.

I bogen bliver dette bla. brugt til at beregne diverse binomialsandsynligheder — men brug blot `pbinom` og `dbinom`.

Nogle normalfordelingsapproksimationer



Statistik for en enkelt binomialfordeling

Statistik

Indtil nu har vi lavet beregninger når successsh. er kendt.

Men det er den sjældent i videnskabelige sammenhænge. Faktisk udfører vi snarere forsøg for at undersøge hvad sandsynligheden er!

Givet data vil vi gøre noget af „det sædvanlige":

- **Estimere** sandsynligheden
- Lave et **konfidensinterval** for sandsynligheden
- Lave **hypotesetest** for om p er noget bestemt (hvis relevant)

Senere: Sammenligning af to eller flere binomialsandsynligheder.

Eksempel: Forelæserens fritidsinteresser

Ved forelæsningen d. 7/9-2022 blev I bedt om at gætte på forelæserens fritidsinteresser

- I havde ingen relevant information at basere jeres gæt på
- Jeg kunne gøre hvad jeg ville for at snyde jer
- Jeres gæt kan dog bruges til at estimere **implicit bias** omkring stereotypen: **statistik-forelæseren**!

Implicit bias over for forelæsere i statistik

Baseret på en stikprøve fra 137 studerende

```
## # A tibble: 6 x 4
##   aktivitet      `FALSE` `TRUE`   pct
##   <chr>            <int>  <int> <dbl>
## 1 fodbold             36    101  73.7
## 2 haekle             122     15  10.9
## 3 kantarel            86     51  37.2
## 4 klatre             101     36  26.3
## 5 renovere_bolig      75     62  45.3
## 6 rulleski            71     66  48.2
```

Breaking news: Fordomme lever i bedste velgående blandt studerende på KU. Forelæsere forventes at løbe rundt på rulleski, når de ikke renoverer bolig eller sidder og ser Liverpool spille fodbold.

Fritidsinteresser: Set-up og spørgsmål

- **Statistisk model:** $Y \sim \text{bin}(137, p)$ hvor p er ukendt.

  Fortolkning af p: Andel af studerende, som angiver aktivitet som fritidsinteresse for forelæseren.

  Er antagelserne for at Y er binomialfordelt egentlig OK?
- **Estimat** for p? Tilhørende **standard error** (SE)?
- **Konfidensinterval?**
- Antag at I på forhånd vidste, at forelæseren dyrkede 3 af de angivne aktiviteter i fritiden. Hvilken værdi af p svarer til at I bare gættede? **Hypotesetest.**

Generel teori: Statistisk model, estimation, SE

Statistisk model: $Y \sim \text{bin}(n, p)$ med kendt n og ukendt p.

Observation, y

Estimation: Naturligt estimat for p (når n er kendt):

$$\hat{p} = \frac{\text{antal succeser}}{\text{antal forsøg}} = \frac{y}{n}$$

Standard error: Husk at SE for $\hat{p}$ er (estimeret) spredning for $\hat{p}$.

$$\text{sd}(Y) = \sqrt{np(1-p)}$$

$$\text{sd}(\hat{p}) = \text{sd}\left(\frac{Y}{n}\right) = \frac{1}{n}\sqrt{np(1-p)} = \sqrt{\frac{p(1-p)}{n}}$$

$$\text{SE}(\hat{p}) = \frac{\sqrt{n\hat{p}(1-\hat{p})}}{n} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Generel teori: Konfidensinterval

Pga. normalfordelingsapproksimationen kan vi lave et **95% konfidensinterval** for p som

$$\hat{p} \pm 1.96 \cdot \mathrm{SE}(\hat{p}) = \hat{p} \pm 1.96 \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Bemærk, at vi bruger 97.5%-fraktilen i standardnormalfordelingen $N(0,1)$, nemlig 1.96.

Hvis vi i stedet ønsker 90% KI: Udskift 1.96 med 1.645 som er 95% fraktilen i standardnormalfordelingen.

Generel teori: Forbedret konfidensinterval

KI på forrige slide bygger på N-approx. som kun er OK hvis $np \geq 5$ og $n(1-p) \geq 5$, altså hvis p ikke er for tæt på 0 eller 1 eller n er for lille.

Ellers kan vi risikere at konfidensgraden for vores KI slet ikke er 95% som vi troede, og det kan indeholde værdier udenfor $(0,1)$.

Kan i stedet bruge følgende **forbedrede KI**:

$$\tilde{p} \pm 1.96 \cdot \sqrt{\frac{\tilde{p}(1-\tilde{p})}{n+4}} \quad \text{med} \quad \tilde{p} = \frac{y+2}{n+4}$$

Bemærk at $\tilde{p} = \frac{y+2}{n+4}$ er **"rykket væk" fra 0 og 1** ift. $\hat{p} = \frac{y}{n}$.

Eksempel: fritidsinteresser

Observation: $y = 51$ gætter på at forelæseren kan lide at plukke kantareller.

Statistisk model: $Y \sim \mathrm{bin}(137, p)$ hvor p er ukendt.

Estimation:

$$\hat{p} = \frac{51}{137} = 0.372, \quad \mathrm{SE}(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{137}} = 0.041$$

Simpelt 95% konfidensinterval:

$$\hat{p} \pm 1.96 \cdot \mathrm{SE}(\hat{p}) = 0.372 \pm 1.96 \cdot 0.041 = (0.291\,,\ 0.453)$$

Forbedret 95% KI: $\tilde{p} = 0.376$, 95% KI

$$\tilde{p} \pm 1.96 \cdot \mathrm{SE}(\tilde{p}) = 0.376 \pm 1.96 \cdot 0.041 = (0.296\,,\ 0.456)$$

R: Det simple KI

Det simple KI beregnes "manuelt", ved indsættelse i formler:

```
p <- 51/137
p

## [1] 0.3722628

SE <- sqrt(p * (1-p) / 137)
SE

## [1] 0.04130032

p - 1.96 * SE

## [1] 0.2913141

p + 1.96 * SE

## [1] 0.4532114
```

R: Forbedret KI

Det forbedrede KI beregnes "manuelt", ved indsættelse i formler:

```
p <- (51 + 2)/(137 + 4)
p

## [1] 0.3758865

SE <- sqrt(p * (1-p) / 141)
SE

## [1] 0.04078971

p - 1.96 * SE

## [1] 0.2959387

p + 1.96 * SE

## [1] 0.4558344
```


Fritidsinteresse: Test af hypotese

Observation: $y = 51$ gætter på at forelæseren kan lide at plukke kantareller.

Statistisk model: $Y \sim \text{bin}(137, p)$ hvor n er kendt og p er ukendt.

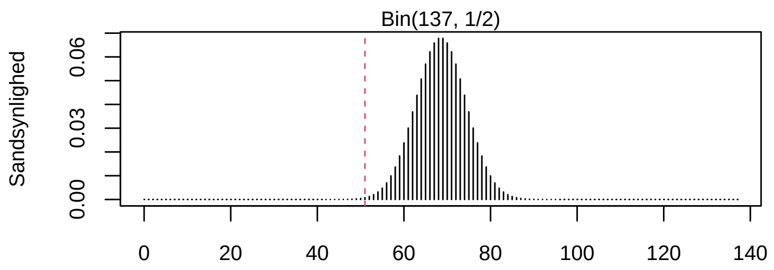
Hypotese: Hvis studerende gætter, så er $p = 1/2 = 0.500$. Vi tester derfor hypotesen $H_0 : p = 1/2$.

"Løsning" via KI: Check om 1/2 ligger i konfidensintervallet.

Men vi kan også lave et egentligt hypotesetest:

- Under hypotesen, dvs. hvis H_0 er sand, kender vi fordelingen af Y fuldstændigt: $Y \sim \text{bin}(137, 1/2)$
- p-værdi: ssh. for — hvis H_0 er sand — at få data der passer lige så dårligt eller dårligere med hypotesen, som $y = 51$.

Fritidsinteresser: Fordeling af Y under hypotesen



- Store/små værdier passer dårligt med H_0
- Værdier „længere væk fra midten" end 51 passer dårligere. Skal formuleres lidt mere præcist...

Fritidsinteresser: p-værdi

Husk at vi kender fordelingen under hypotesen: $Y \sim \text{bin}(137, 1/2)$

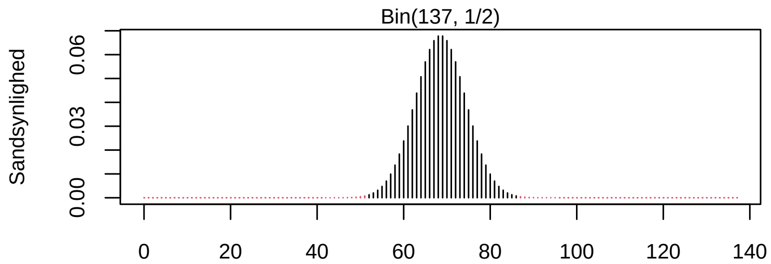
Bruger observationen $y = 51$ selv som **teststørrelse**.

- Observationer med $P(Y = y) \leq P(Y = 51)$ passer mindst lige så dårligt med hypotesen som værdien 51
- Læg punktsandsynlighederne for disse y sammen

Formelt:

$$p\text{-værdi} = \sum_{y: P(Y=y) \leq P(Y=51)} P(Y = y).$$

Fritidsinteresser: p-værdi



- Røde punkter har samme eller mindre ssh. end $y = 51$.
- p-værdi $=$ **sum af røde sandsynligheder** $= 0.0035$
- Vi forkaster H_0; Ikke rimeligt at tro på, at vi blot gættede

R: `binom.test`

p-værdien kan beregnes med `binom.test` (eller manuelt)

```
binom.test(51, 137, p=1/2)

##
##  Exact binomial test
##
## data:  51 and 137
## number of successes = 51, number of trials = 137, p-value = 0.00352
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
##  0.2912522 0.4589238
## sample estimates:
## probability of success
##              0.3722628
```


R: `prop.test`

Bemærk at `prop.test` (-se nedenfor) også kan bruges til beregning af en p-værdi. Men dette test benytter en anden teststørrelse (og givet et lidt andet resultat).

Et forbedret KI kan dog **næsten** beregnes med `prop.test` og option `correct=FALSE`. Ikke helt det samme som på slide 30, men I må gerne bruge det!

```
prop.test(51, 137, p=1/2, correct=FALSE)

##
##  1-sample proportions test without continuity correction
##
## data:  51 out of 137, null probability 1/2
## X-squared = 8.9416, df = 1, p-value = 0.002787
## alternative hypothesis: true p is not equal to 0.5
## 95 percent confidence interval:
##  0.2958352 0.4556584
## sample estimates:
##         p
## 0.3722628
```


Generel teori: Hypotesetest

- **Stat. model:** $Y \sim \text{bin}(n, p)$ hvor n er kendt og p er ukendt.
- **Observation,** $y = y_0$
- **Hypotese,** $H_0 : p = p_0$ for hypoteseværdi p_0. Under hypotesen er $Y \sim \text{bin}(n, p_0)$. Kan tegne denne fordeling!
- ***p*-værdi** $=$ sandsynligheden for at få observationer der passer lige så dårligt eller dårligere med hypotesen end y_0, dvs.

$$p\text{-værdi} = \sum_{y: P(Y=y) \leq P(Y=y_0)} P(Y = y).$$

- **Konklusion** som sædvanlig

Eksempel: Tilslutning til politisk parti

Repetition af centrale begreber:

- **Data:** y antal personer som vil stemme på parti baseret på meningsmåling med n personer
- **Stat. model:** y opfattes som udfaldet af $Y \sim \text{bin}(n, p)$
- **Parametre:** p andelen i population som vil stemme på partiet
- **Estimat:** $\hat{p} = y/n$
- (Simpelt) **95 %-KI:** $\hat{p} \pm \underbrace{1.96 \cdot \sqrt{\hat{p}(1-\hat{p})/n}}_{\text{statistiske usikkerhed}}$

I meningsmålinger fremhæves ofte den **statistiske usikkerhed** i %-point.

Med stikprøve af størrelse $n = 1000$:

- For store partier ($p \approx 0.5$) er stat. usikkerhed ± 3.1 %-point
- For mindre partier ($p \approx 0.1$) er stat. usikkerhed ± 1.9 %-point
- For små partier ($p \approx 0.02$) er stat. usikkerhed ± 0.9 %-point

Statistik for to binomialfordelinger

Eksempel: Kastrering og diabetes

Mistanke om at tidlig kastrering øger risikoen for diabetes.

Forsøg:

- 100 mus inddelt tilfældigt i to grupper (50+50).
- Den ene gruppe mus blev kastreret dagen efter fødsel; den anden gruppe mus blev ikke kastreret
- Efter 112 dage undersøgte man om musene havde udviklet diabetes

	Diabetes	Ikke diabetes	Total
Katrerede mus	26	24	50
Ikke-kastrerede mus	12	38	50

Er der forskel er der på risikoen for at udvikle diabetes? Hvor stor?

Kastrering og diabetes: Statistisk model og formål

Statistisk model:

- Data fra de to grupper er uafhængige
- Kastrerede mus: Observation $y = 26$ fra $\text{bin}(50, p)$
- Ikke-kastrerede mus: Observation $x = 12$ fra $\text{bin}(50, q)$

Interesseret i **forskellen mellem de to grupper:**

- Estimat og konfidensinterval for $p - q$
- Test for hypotesen $H_0 : p = q$.

Generel teori: Statistisk model, estimation

Generelt set-up:

- Statistisk model: $Y \sim \text{bin}(n,p)$ og $X \sim \text{bin}(m,q)$, uafhængige
- Observationer y og x
- Interesseret i forskellen $p - q$.

Estimat for forskel:

$$\widehat{p-q} = \hat{p} - \hat{q} = \frac{y}{n} - \frac{x}{m}$$

Generel teori: Standard error (SE) for forskel

Vi ved godt hvordan vi beregner SE for $\hat{p}$ og $\hat{q}$:

$$\text{SE}(\hat{p}) = \sqrt{\frac{\hat{p}\cdot(1-\hat{p})}{n}}, \quad \text{SE}(\hat{q}) = \sqrt{\frac{\hat{q}\cdot(1-\hat{q})}{m}},$$

Regneregler for varianser/spredninger giver **SE for forskel:**

$$\text{SE}(\hat{p}-\hat{q}) = \sqrt{\text{SE}(\hat{p})^2 + \text{SE}(\hat{q})^2} = \sqrt{\frac{\hat{p}\cdot(1-\hat{p})}{n} + \frac{\hat{q}\cdot(1-\hat{q})}{m}}$$

Generel teori: KI for forskel mellem sandsynligheder

95% KI for differensen mellem de to sandsynligheder, $p - q$:

$$\hat{p} - \hat{q} \pm 1.96 \cdot \text{SE}(\hat{p}-\hat{q})$$

Altså:

$$\hat{p} - \hat{q} \pm 1.96 \cdot \sqrt{\frac{\hat{p}\cdot(1-\hat{p})}{n} + \frac{\hat{q}\cdot(1-\hat{q})}{m}}$$

Vi kan bruge dette konfidensinterval til at lave et "groft" test for hypotesen $H_0 : p = q$. På onsdag laves et egentligt hypotesetest.

Kastrering og diabetes: Estimat og KI

	Diabetes	Ikke diabetes	Total
Katrerede mus	26	24	50
Ikke-kastrerede mus	12	38	50

Estimater for hver gruppe:

$$\hat{p} = \frac{26}{50} = 0.52 \text{ (SE 0.071)}, \quad \hat{q} = \frac{12}{50} = 0.24 \text{ (SE 0.060)}$$

Estimat for forskel: $\hat{p} - \hat{q} = 0.28$ (SE 0.093)

95% KI for differens $p - q$:

$$0.28 \pm 1.96 \cdot 0.093 = (0.098\,,\, 0.462)$$

Nul ligger ikke i KI, så risikoen er større blandt de kastrerede mus.

R

SE for forskel kan beregnes manuelt; se dagens R-kode.

KI for forskel kan beregnes med `prop.test`:

```
prop.test(c(26,12), c(50,50), correct=FALSE)

##
##  2-sample test for equality of proportions without continuity
##  correction
##
## data:  c(26, 12) out of c(50, 50)
## X-squared = 8.3192, df = 1, p-value = 0.003923
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  0.09781821 0.46218179
## sample estimates:
## prop 1 prop 2
##   0.52   0.24
```


Opsummering vedr. R

En enkelt binomialfordeling

- Simpelt KI skal laves "i hånden", dvs. beregn selv de forskellige størrelser
- `prop.test` giver næsten (men ikke helt) det forbedrede KI.
  Med/uden "kontinuitetskorrektion": Bogens formler svarer til *ikke* at bruge korrektionen.
- `binom.test` giver p-værdi for hypotesen $H_0 : p = p_0$
- `prop.test` giver også en p-værdi, men en anden end vi har beregnet. Fornuftig nok så længe np og $n(1-p)$ er ≥ 5

I må selv vælge metoden medmindre I bliver spurgt om noget eksplicit.

Opsummering vedr. R

To binomialfordelinger

- `prop.test` giver estimater for hver ssh. samt KI for forskel.
  Med/uden "kontinuitetskorrektion": Bogens formler svarer til *ikke* at bruge korrektionen.
- SE for forskel skal beregnes manuelt, hvis den bruges
- Giver også en p-værdi. Mere om det på onsdag.

I må selv vælge metoden medmindre I bliver spurgt om noget eksplicit.

Hjemmeopgave: valgkamp

Voxmeter offentliggør hver uge politiske meningsmålinger baseret på telefoninterview med ca. 1000 repræsentativt udvalgte danskere.

- Moderaterne opnåede en tilslutning på 8.8 % ved meningsmålingen i uge 42 ($n = 942$).
- Moderaterne opnåede en tilslutning på 6.1 % ved meningsmålingen i uge 41 ($n = 1008$).

Medierne fokuserede i weekenden kraftigt på, at partiet Moderaterne oplever medvind i valgkampen.

- Er det rimeligt at hævde, at de to meningsmålinger tyder på, at tilslutningen for Moderaterne har ændret sig fra uge 41 til uge 42?
- Angiv et estimat og et 95 %-KI for ændringen i Moderaternes vælgertilslutning fra uge 41 til uge 42.