



Københavns Universitet



## Erfaringer med eksperimenter der genererer data til brug i undervisningen

Sørensen, Helle

*Published in:*

Symposium i Anvendt Statistik 2019

*Publication date:*

2019

*Document Version*

Også kaldet Forlagets PDF

*Citation for published version (APA):*

Sørensen, H. (2019). Erfaringer med eksperimenter der genererer data til brug i undervisningen. I P. Linde (red.), Symposium i Anvendt Statistik 2019 (s. 224-235). Økonomisk Institut, Københavns Universitet og Det Nationale Forskningscenter for Arbejdsmiljø.

# Erfaringer med eksperimenter der genererer data til brug i undervisningen

Helle Sørensen (helle@math.ku.dk)

Institut for Matematiske Fag, Københavns Universitet

**Resumé** Jeg har de sidste to år udført små eksperimenter på mine kurser og brugt data som cases i undervisningen. I det ene forsøg løser de studerende sudokuer på tid; i det andet gætter de studerende på antallet af punkter i tre punktplot. Jeg har udført begge eksperimenter på to introducerende kurser og har svar fra 100-200 studerende fra hvert kursus. Data fra eksperimenterne har vist sig at være en guldgrube i undervisningen: De giver anledning til at diskutere begreber som randomisering, uafhængighed, transformation, variansanalyse, reproducerbarhed, og de kan bruges i både forelæsninger og opgaver. Og allervigtigst: De studerende synes det er sjovt og kan nemt forholde sig til problemstillinger, data og resultater. Atiklen forklarer hvordan forsøgene er udført og giver eksempler på hvordan jeg har brugt datasættene i undervisningen.

## 1 Introduktion

Jeg har de sidste år leget med at få studerende til at deltage i små eksperimenter og derved genereret data som kan bruges i undervisningen. Et af formålene er naturligvis at frembringe nogle datasæt som de studerende har „et forhold“ til, som alle har lige gode forudsætninger for at forstå (på tværs af studieretninger), og som består af andet/mere end højde, vægt, alder osv. Et andet formål er at give de studerende en lillebitte fornemmelse for forsøgsplanlægning, dataindsamling og dataklargøring. Data har det ellers med at falde ned fra himlen i pæn indpakning på mine kurser, men selvom man i dag vitterligt kan hente store mængder af data ned fra *skyen*, så er planlægning og praktisk udførelse af eksperimenter stadig et vigtigt element i de eksperimentelle fag. For det tredje er det sjovt—både for de studerende (siger de, når de bliver spurgt) og for mig.

I denne artikel beskriver jeg to eksperimenter, som jeg har testet på to kurser med mere end 100 studerende. I det ene forsøg har de studerende løst sudokuer på tid, i det andet har de gættet på antal punkter i tre punktplot. Nedenfor beskriver jeg forsøgene og den praktiske udførelse (afsnit 2), giver eksempler på hvordan jeg har benyttet data i undervisningen (afsnit 3) og rapporterer resultater fra nogle analyser af data (afsnit 4).

Det er ikke raketvidenskab, og masser af undervisere, herunder mange af mine kolleger, har anbefalet og gjort lignende ting før mig (Brophy and Hahn, 2014; Gelman and Nolan, 2002; Pfaff and Weinberg, 2009; Cummiskey *et al.*, 2012). Ikke desto mindre har jeg haft langt større større nytte af datasættene end jeg umiddelbart havde forestillet mig.

## 2 Indsamling af data

Begge eksperimenter er udført på to kurser, og inden jeg beskriver eksperimenterne i detaljer, giver jeg en kort beskrivelse af kurserne.

### 2.1 Kurserne

Kurset *Statistisk Dataanalyse 1* (StatDat1) er obligatorisk på andet eller tredje år på bacheloruddannelserne Husdyrvidenskab, Jordbrugsøkonomi, Naturressourcer og til og med studieåret 2017/18 også på Biologi-Bioteknologi. Der er 150–200 studerende tilmeldt kurset. Kurset er anvendelsesorienteret og handler mest om analyser baseret på normalfordelingen, til og med multipel lineær regression og tosidet variansanalyse.

Kurset *Sandsynlighedsregning og Statistik* (SS) er obligatorisk på første år på bachelordannelserne i Matematik, Matematik-Økonomi og Forsikringsmatematik. Der er cirka 300 studerende tilmeldt kurset. Kurset er matematisk af natur og handler blandt om sandsynlighedsteori for kontinuerte fordelinger. Statistikdelen af kurset handler om en og to stikprøver samt lineær regression.

Eksperimentet med punktplot blev udført i studieåret 2017/18, mens sudokuexperimentet blev lavet i studieåret 2018/19. Begge forsøg blev først lavet på StatDat1.

### 2.2 Sudoku

Sudokuexperimentet er stort set replikeret fra Brophy and Hahn (2014). Der er fire forskellige sudokuer, og hver studerende skal løse en af dem på tid. De fire sudokuer har præcis samme struktur, men adskiller sig ved hvilken type tegn der benyttes: tal, latinske

bogstaver, græske bogstaver, eller symboler. De fire sudokuer er vist i Figur 1, gengivet fra Brophy and Hahn (2014).

Sudokuerne blev printet og sorteret så de fire typer lå skiftevis, således at tildelingen af sudokuerne til studerende var randomiseret. Eftersom strukturen af alle sudokuer var ens, fik de studerende instrukser om ikke at kigge hos hinanden eller snakke sammen i løbet af forsøget. Tidtagning blev taget på et full-screen stopur på projektoren.

2			6		4
		4	5		
5	2	6	1	4	
	3	1	2	5	6
		5	4		
1		2			5

b			f		d
		d	e		
e	b	f	a	d	
	c	a	b	e	f
		e	d		
a		b			e

$\beta$			$\mu$		$\varepsilon$
		$\varepsilon$	$\lambda$		
$\lambda$	$\beta$	$\mu$	$\alpha$	$\varepsilon$	
	$\delta$	$\alpha$	$\beta$	$\lambda$	$\mu$
		$\lambda$	$\varepsilon$		
$\alpha$		$\beta$			$\lambda$

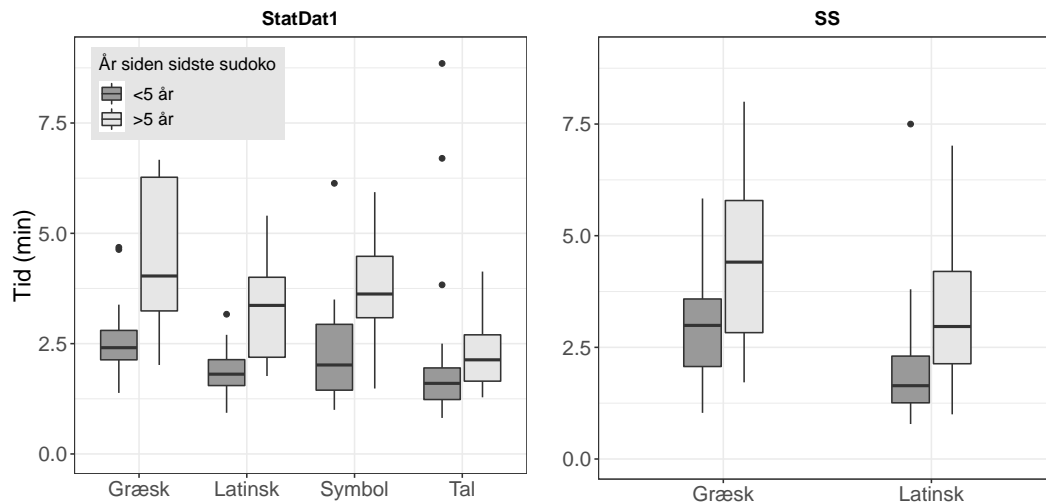
$\Delta$			$\ominus$		$\times$
		$\times$	$\ominus$		
$\ominus$	$\Delta$	$\oplus$	$\blacksquare$	$\times$	
	$\checkmark$	$\blacksquare$	$\Delta$	$\ominus$	$\oplus$
		$\ominus$	$\times$		
$\blacksquare$		$\Delta$			$\ominus$

Figur 1: De fire sudokuer. Bemærk at strukturen er ens, altså at den eneste forskel er typen af tegn der indgår.

På StatDat1 udførte jeg eksperimentet ved en forelæsning og lod forsøget vare ti minutter. Dette viste sig at være lang tid (højeste observation var under ni minutter, næsthøjeste under syv minutter), og jeg lod derfor kun eksperimentet køre otte minutter på SS. På SS blev eksperimentet desuden udført af lærerne på kursets ti øvelsesshold, og jeg brugte kun sudokuer med latinske og græske bogstaver fordi pensum på dette kursus kun omfatter sammenligning af to stikprøver.

Efter de ti henholdsvis otte minutter blev de korrekte løsninger vist til de studerende, så de kunne checke om de havde løst deres sudoku korrekt. Derefter udfyldte deltagerne en elektronisk formular hvor de skulle indtaste hvor lang tid de havde brugt på at løse sudokuen, om de havde løst den korrekt, og desuden svare på en række spørgsmål vedr. studieretning, erfaring med at løse sudokuer, holdnummer (kun SS) mm. Med hensyn til erfaring blev de studerende blev spurgt hvornår de sidst havde løst en sudoku og kunne svare *mindre end 5 år siden* eller *mere end 5 år siden* eller *aldrig*. Meget få svarede at de aldrig havde løst sudokuer, så de sidste to kategorier er slået sammen. Det tog 15–20 minutter at udføre forsøget, inklusive forklaringer og indtastning.

Figur 2 viser boksplot over løsningstiden, opdelt efter kursus (StatDat1 til venstre og SS til højre) samt sudokotype og erfaring. Data ser ud cirka som man kunne forvente: Sudokuer med tal og latinske bogstaver er hurtigere at løse end de øvrige typer, og det hjælper at have løst sudokuer „for nylig“.

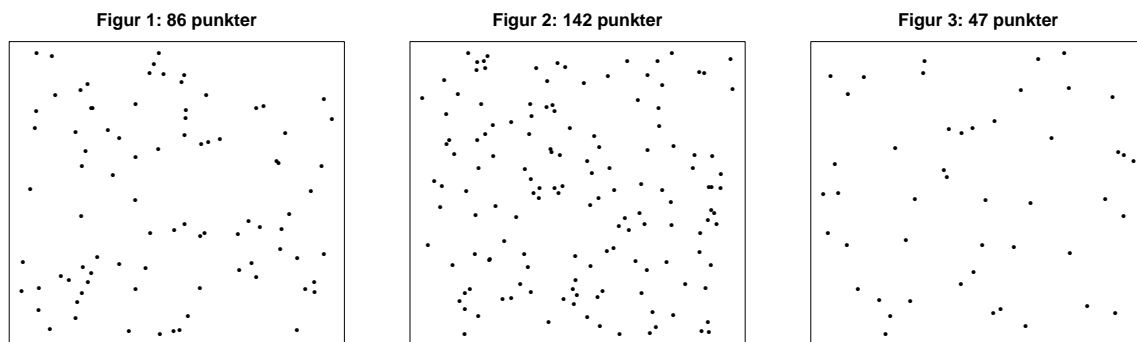


Figur 2: Boksplot over løsningsstiden fra sudokoforsøget opdelt efter kursus, sudokotype og erfaring. Der er data fra 112 studerende på StatDat1 og 192 studerende på SS. Der er cirka lige mange studerende for hver sudokotype (indenfor kursus), mens kategorien  $<5$  år omfatter cirka 60% af deltagerne.

### 2.3 Punktplot

Eksperimentet med punktplot er inspireret af data indsamlet af min tidligere kollega Claus Ekstrøm (Ekstrøm, 2016, afsnit 3.39). Det blev udført ved forelæsninger på begge kurser. Jeg viste tre punktplot på projektoren; et ad gangen, hvert plot i cirka seks sekunder og med en pause mellem plottene. De studerende gættede på antallet af punkter i hvert plot undervejs og angav umiddelbart efter forsøget deres gæt ved at udfylde en elektronisk formular hvor de også svarede på diverse spørgsmål vedr. køn, studieretning mm. De studerende fik strenge instrukser om ikke at snakke sammen undervejs. De sande antal blev først afsløret for de studerende senere i kurset. Det tog under 10 minutter at udføre forsøget, inklusive forklaringer og indtastning af data.

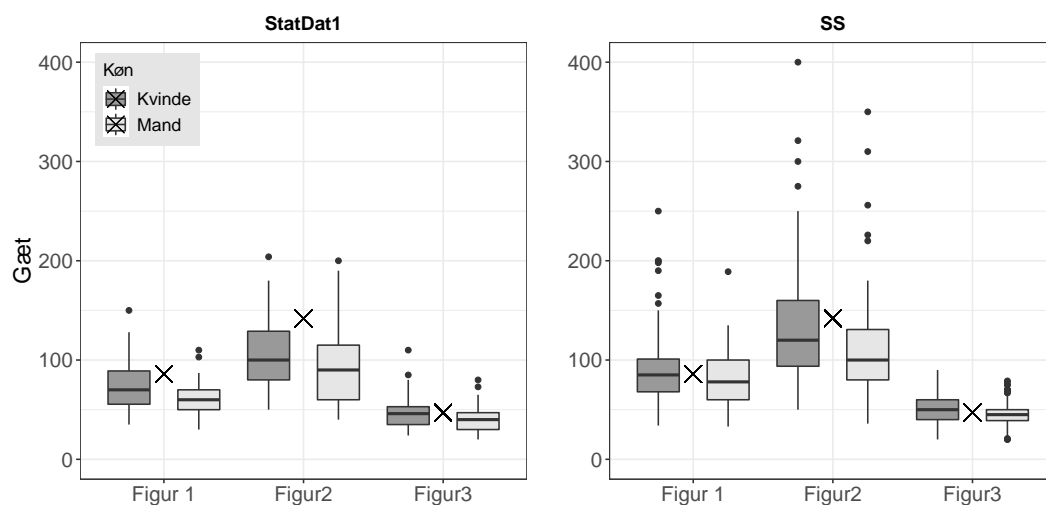
Punktplottene er vist i figur 3. Det sande antal punkter er angivet over punktplottene; det var naturligvis ikke tilfældet da de blev vist til de studerende. Figur 4 viser boksplot over de studerendes gæt, opdelt efter kursus (StatDat1 til venstre og SS til højre) samt figur og køn. Krydserne repræsenterer det sande antal punkter i figurerne. Særligt for figur 2 med mange punkter ses det at det typiske gæt er for lavt.



Figur 3: De tre punktplot som blev vist til de studerende. Det sande antal punkter er angivet øverst; dette blev ikke vist til de studerende.

### 3 Brug af data i undervisningen

Jeg har brugt datasættene som gennemgående cases i kurserne; ved flere forelæsninger og til adskillige øvelsesopgaver og/eller afleveringsopgaver. Sidste år brugte jeg desuden data fra SS til reeksamen på StatDat1, og jeg har desuden lånt data til flere kolleger som har brugt dem til eksamensopgaver på andre kurser.



Figur 4: Boksplot over gættene på antal punkter opdelt efter kursus, figur og køn. Krydsnerne viser det sande antal punkter i de tre figurer (86, 142, 47). Der er data fra 140 studerende på StatDat1 og 191 studerende på SS.

Oprindeligt var tanken først og fremmest at bruge data til illustration af variansanalysemodeller, men data viste sig nyttige til at diskutere en lang række problemstillinger og emner—flere end jeg havde forestillet mig inden jeg havde data. Lad mig give nogle eksempler.

**Randomisering, konfundering** I forbindelse med sudokoeksperimentet diskuterer jeg vigtigheden af at arkene med de fire sudokotyper ligger skiftevis ved uddelingen i lokalet. På StatDat1 er der en tendens til at studerende fra samme studieretning også sidder sammen i lokalet, og de studerende kan nemt forholde sig til konfunderingsproblemer hvis sudokoerne i stedet havde været sorteret så alle af samme type lå sammen. Tilsvarende diskuterer jeg vigtigheden af at de fire sudokoer er helt ens bortset fra typen af tegn, således at en eventuel forskel mellem typerne ikke kan skyldes forskellig sværhedsgrad. Igen noget som de studerende nemt kan forholde sig til—faktisk var der en studerende der bragte det på bane allerede inden jeg var nået til det.

**Uafhængighed** Som nævnt i afsnit 2.2 gjorde jeg i begge eksperimenter et stort nummer ud af at de studerende ikke må snakke sammen i løbet af forsøget. Om det virker ved jeg ikke, og i virkeligheden er det væsentlige er at de studerende kan huske instruksen når vi senere snakker om antagelsen om uafhængighed mellem data fra forskellige personer. Omvendt kan simple scatterplots og korrelationer mellem gættene på de tre figurer bruges til at illustrere at data fra samme person bestemt ikke er uafhængige.

**Dataklargøring, fejl i data, manglende data** De studerende indtastede deres løsningstider/gæt samt baggrundsoplysninger på en elektronisk formular. Adskillige ting kan gå galt i den forbindelse. For eksempel blev de i sudokoforsøget bedt om at indtaste et helt antal minutter samt antal sekunder udover de hele minutter. Hvad betyder det så når en studerende indtaster 1 min og 96 sek? En anden studerende angav tiden 9 min og 46 sek på trods af at forsøget blev stoppet efter 8 min. Første gang jeg kørte punktploteksperimentet, var jeg tilsyneladende ikke helt skarp i mine instrukser; i hvert fald indtastede cirka 10 studerende deres gæt allerede inden jeg havde vist figurerne på projektoren! Disse observationer skulle oplagt fjernes fra datasættet, men hvad gør man med andre gæt på fx 10 eller 1000? Det er næppe seriøse gæt, men hvor går grænsen? Derudover var der studerende der ikke svarede eller ikke svarede entydigt på alle baggrundsspørgsmål; kan man alligevel bruge data fra disse personer?

**Log-transformation** Den problemstilling som datasættene måske har belyst allerbedst, er transformation, herunder fortolkning af estimater når analysen er kørt på log-transformerede data.

Gættene på antal punkter viser sig på smukkeste vis at være log-normalfordelte; der er nogle der gætter alt, alt for højt, men få der gætter meget lavt. QQ-plots for figur 1 fra

StatDat1 er vist i venstre side af figur 5 før og efter log-transformation. På StatDat1 er der mange datasæt der kræver log-transformation, og jeg tror næsten at de studerende antager at det er noget jeg med vilje opsøger (for at genere dem), så det er glimrende at de på egen hånd frembringer data der lægger op til transformation. Log-transformationen fører naturligt til en diskussion af hvordan det tilbagetransformerede estimat skal fortolkes som en estimeret median, og hvordan fx forskellen mellem mænd og kvinder skal fortolkes som en relativ, snarere end en absolut, forskel. Det er også oplagt at diskutere variationskoefficienter: På StatDat1 var de næsten ens for de tre figurer (0.329, 0.336 hhv. 0.319 når der ikke justeres for andre variable).

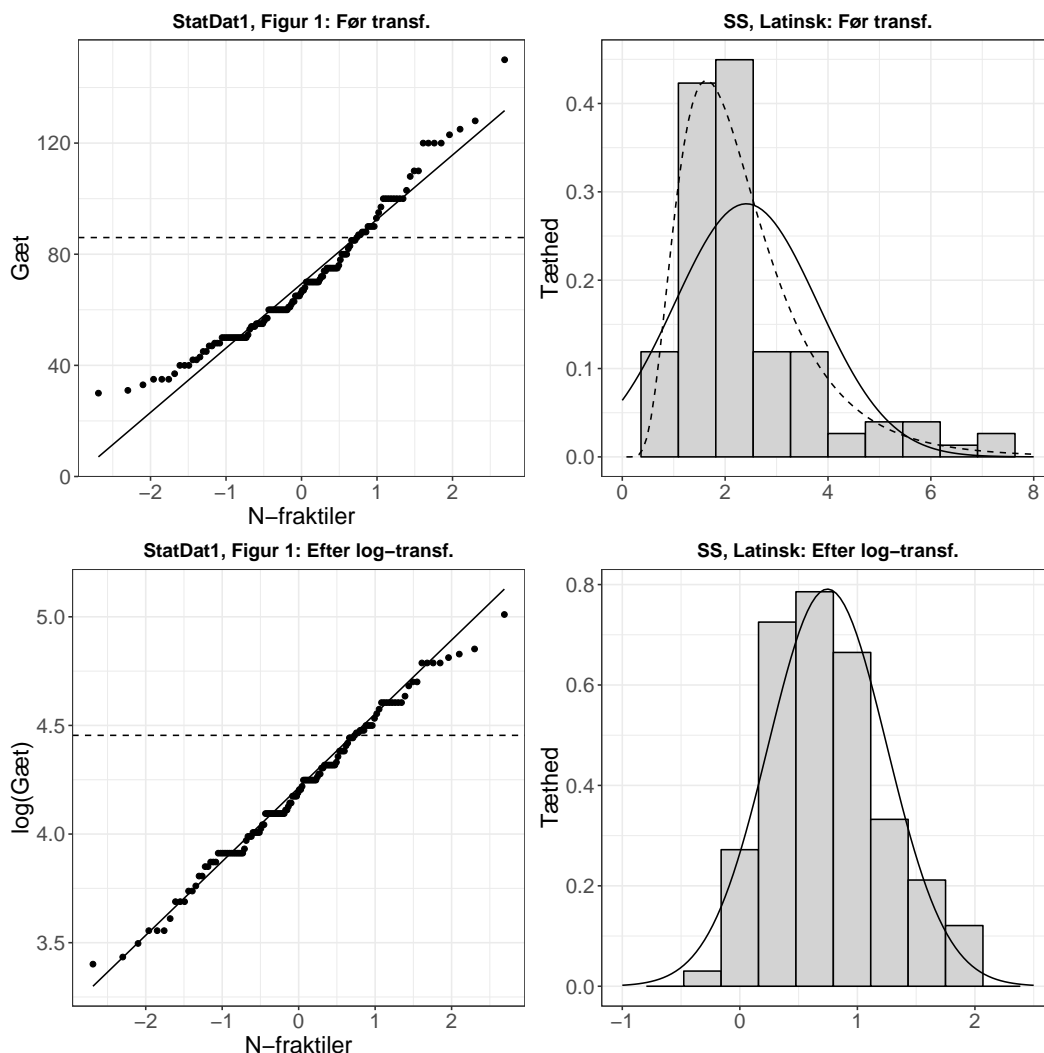
Størstedelen af SS handler om sandsynlighedsregning, og det er først sent i kurset at der er egentlige statistiske analyser, men transformationsproblematikken fungerer udmærket alligevel. En del af den obligatoriske opgave i 2018/19 bestod i at tegne histogrammer og normalfordelingstætheder for løsningsstiden for sudokoer med latinske bogstaver, før og efter transformation; at udlede tætheden for log-normalfordelingen og sammenligne med data; samt bestemme middelværdi og median for log-normalfordelingen. Histogrammer og tætheder er vist i højre side af figur 5.

**Lineære normale modeller** Data fra punktplotforsøget fungerer glimrende som eksempel på en enkelt stikprøve fordi det sande antal punkter kan bruges som reference- eller hypoteseværdi (og ikke er nul). Det viser sig at være særligt sjovt fordi det åbenbart er svært at gætte medmindre der kun er få punkter i plottet. Et lille aber dabei: På dette stadie af undervisningen taler jeg gerne om populationer og stikprøver, men der er ikke for alvor en naturlig population hvorfra vi har udtaget en repræsentativ stikprøve.

For begge eksperimenter ligger det lige til højrebened at køre variansanalyser (på log-transformerede data). På StatDat1 gør jeg det ad flere omgange som kurset skrider frem, dvs. inddrager først en enkelt forklarende variabel (fx sudokotype), derefter flere variable og enkelte vekselvirkninger. Vekselvirkninger er notorisk svært, men bliver mere tilgængeligt i denne indpakning: Er forskellene mellem sudokotyperne fx forskellig for de forskellige niveauer af erfaring? På SS er ambitionsniveauet mindre da pensum kun omfatter en og to stikprøver.

Data er mindre velegnede til regressionsanalyser, men en (måske lidt søgt) opgave om multipel regression på StatDat1 viste dog at gæt på figur 1 og/eller 3 kan inddrages til prædiktion af en persons gæt på antal punkter i figur 2, og man kunne have diskuteret *regression towards the mean* i samme forbindelse.





Figur 5: Illustrationer vedrørende log-transformation. Til venstre ses QQ-plots for gæet på antal punkter i figur 1 før/efter (øverst/nederst) log-transformation. Til højre ses histogrammer for løsningsstider for sudokoen med latinske bogstaver før/ efter log-transformation (øverst/nederst). De ikke-stiplede kurver til højre er normalfordelingstætheder, mens den stiplede kurve i øverste højre graf er tætheden for den estimerede log-normalfordeling.

**Binomialmodeller, logistisk regression** Oprindeligt var det min plan at korrekthed af sudokuen kunne bruges som en binær respons men det viser sig at næsten alle løser sudokuerne korrekt. Til gengæld kan data fra punktplotforsøget bruges binært hvis man udelukkende fokuserer på om gættene er for høje eller lave.

**Reproducerbarhed** I punktplotforsøget på StatDat1 viste der sig at være signifikant kønsforskel; det estimeres at kvinder gætter bedre end mænd (se venstre side af tabel 2). Det havde de (kvindelige) studerende en del morskab af, og det førte naturligt til en diskussion om at man altid bør gentage eksperimenter der har ført til overraskende konklusioner, og derved (forsøge at) konfirmere resultaterne. I dette tilfælde *blev* forsøget jo faktisk gentaget på SS, og med samme konklusion (se højre side af tabel 2), så de kvindelige SS-studerende kunne hovere dobbelt så meget.

Hvis det hørte naturligt hjemme på kurserne, kunne data også illustrere andre temaer. Data fra sudokueksperimentet kunne bruges i forbindelse med overlevelsesanalyse, se evt. Kaplan-Maier plots i Brophy and Hahn (2014), og man kunne afkorte forsøget hvis man gerne vil diskutere censorering i højere grad. Sudokueksperimentet blev udført på øvelsesholdene på SS, hvilket naturligt lægger op til at inddrage en tilfældig holdeffekt, ligesom en tilfældig personeffekt bør inddrages hvis man analyserer gæt på flere figurer i en samlet model. Der er givetvis andre muligheder som jeg ikke har tænkt over.

## 4 Resultater

Resultaterne fra de statistiske analyser er selvsagt ikke det vigtigste for den didaktiske diskussion. Og dog: Det er ikke uvæsentligt for succesoplevelsen at resultaterne er forståelige og måske ligefrem efterviser de teorier/hypoteser man måtte have på forhånd. På kurser med studerende fra eksperimentelle fag er dette en pointe i sig selv. Data er analyseret separat for de to kurser.

### 4.1 Sudoku

For sudokueksperimentet har jeg brugt en tresidet variansanalyse uden vekselvirkninger med logaritmen til løsningstiden som respons og følgende forklarende variable: Sudoku-type (fire typer for StatDat1, to typer for SS, jf. figur 1); studieretning (fire niveauer for hvert kursus), samt erfaring (mere eller mindre end fem år siden sidst løste sudoku).

Tabel 1 viser  $p$ -værdierne for effekten af hver variabel samt estimerer for effekter der kan estimeres på begge kurser, se også figur 1. Estimererne er tilbagetransformerede og angiver dermed den faktor som estimerede medianer vokser med når den givne faktor ændres. Estimererne er tæt på ens for de to kurser: Mediansløsningstiden estimeres til at være 45% hhv. 47% højere for sudokuer med græske bogstaver sammenlignet med latinske bogstaver, og medianløsningstiden estimeres 60% hhv. 64% højere for studerende der ikke har løst sudokuer indenfor de sidste fem år sammenlignet med studerende der faktisk har løst sudokuer indenfor de sidste fem år. På StatDat hvor fire sudokotyper blev testet, var løsningstiden lavest for sudokuer med tal og 64%, 13% hhv. 33% højere for sudokuer med græske bogstaver, latinske bogstaver hhv. symboler.

	StatDat1	SS
Antal studerende	112	192
<i>Test for hovedeffekter</i>		
Sudokotype	$p = 0.00039$	$p = 1.6 \cdot 10^{-9}$
Erfaring	$p = 1.4 \cdot 10^{-7}$	$p = 5.2 \cdot 10^{-13}$
Studieretning	$p = 0.017$	$p = 0.024$
<i>Estimat (95% KI) for faktor</i>		
Græsk vs. Latinsk	1.45 (1.15 – 1.84)	1.47 (1.30 – 1.66)
> 5 år vs. < 5 år siden sidst	1.60 (1.36 – 1.89)	1.64 (1.45 – 1.86)
Variationskoefficient	0.441	0.440

Tabel 1: Resultater fra sudokuexperimentet:  $p$ -værdier og udvalgte estimerer fra en tresidet variansanalyse for hvert kursus. Estimerer er tilbagetransformerede, så de angiver den estimerede faktor som medianen forøges med, når variabelen ændres.

## 4.2 Punktplot

Gættene på antal punkter i figur 1 er analyseret med en tosidet variansanalyse uden vekselvirkning med logaritmen til gættet som respons og køn og studieretning (fire niveauer for hvert kursus) som forklarende variable.

Tabel 2 viser udvalgte resultater, se også figur 3. Mænd estimeres til at gætte 22% hhv. 17% højere end kvinder, og forskellen er statistisk signifikant på begge kurser. Der er 86 punkter i figur 1, og denne værdi er indeholdt i konfidensintervallet for kvinder på SS (gennemsnit over studier), men hverken for mænd på SS eller for kvinder og mænd på

StatDat1.

I de tilsvarende analyser for figur 2 og figur 3 gættede kvinder også signifikant højere end mænd (begge figurer, begge kurser), men analyserne af de tre figurer er oplagt ikke uafhængige så det skal man næppe lægge for meget i. Ikke overraskende gætter man bedre når der er få punkter i figuren: For figur 2 med 142 punkter var det sande antal ikke indeholdt i konfidensintervallet (begge køn, begge kurser), mens det for figur 3 med 47 punkter kun er mænd fra StatDat1 hvor det sande antal ikke er indeholdt i konfidensintervallet.

	StatDat1	SS
Antal studerende	140	191
<i>Test for hovedeffekter</i>		
Køn	$p = 0.0030$	$p = 0.0055$
Studieretning	$p = 0.69$	$p = 0.66$
<i>Estimat (95% KI) for forhold</i>		
Kvinder vs. Mænd	1.22 (1.07 – 1.39)	1.17 (1.05 – 1.30)
<i>Estimeret median (95% KI)</i>		
Kvinder, gennemsnit over studier	70.3 (65.5 – 75.3)	88.0 (80.2 – 96.6)
Mænd, gennemsnit over studier	57.5 (51.8 – 63.7)	75.5 (69.8 – 81.6)
Variationskoefficient	0.322	0.387

Tabel 2: Resultater vedr. figur 1 fra punktploteksperimentet:  $p$ -værdier og udvalgte estimater fra en tosidet variansanalyse for hvert kursus. Estimatet for effekten af køn er tilbagetransformeret, så det angiver estimatet for forholdet mellem mediangæt for kvinder og mænd. Det sande antal punkter i figur 1 er 86.

## 5 Konklusion

Datasættene har vist sig at være guld værd i undervisningen. Det er sjovest at bruge dem sammen med de studerende der har været udsat for eksperimenterne, men de fungerer også fint i efterfølgende kurser fordi de er nemme at forholde sig til. På StatDat1 spurgte jeg de studerende direkte om de synes det var sjovt/irriterende/lige gyldigt at inddrage data som de selv havde været med til at frembringe. Cirka 80% svarede at det var sjovt, resten at det var lige gyldigt, og flere nævnte det sjove i den „konkurrence“ der fx opstod mellem studieretninger.

Min erfaring er altså at de studerende synes det er sjovt, og også at de er villige til at lege med. Jeg forklarer omhyggeligt at det er frivilligt at deltage men at det er sjovest hvis mange deltager og svarer ærligt og seriøst. Der er som regel et par stykker der svarer fjollet/useriøst, men det er ganske få. Selvom der ikke spørges om personfølsomme ting, skal man selvfølgelig sørge for at enkeltpersoner ikke kan genkendes ud fra baggrundsoplysningerne. Og kan man overhovedet spørge til studerendes køn i 2018?

De to eksperimenter kan udvides med flere sudokuer (af forskellig sværhedsgrad), og flere punktplot (gerne også gentagelser af de samme punktplot) således at man får mere komplicerede datastrukturer, men logistikken bliver sværere. Det er en pointe i sig selv at det er praktisk muligt at gennemføre hvert forsøg på 10–20 minutter, også på kurser med mange studerende, og en væsentlig del af succesen stammer fra at (næsten) alle deltager.

Som det forhåbentlig fremgår af artiklen, har jeg selv haft stor fornøjelse af forsøgene. Det er skægt at udføre disse små forsøg og—særligt i sudokoeksperimentet—se hvordan resultaterne i høj grad bliver som man ville forvente.

## Litteratur

Brophy, C. and Hahn, L. (2014). Engaging students in a large lecture: An experiment using sudoku puzzles. *Journal of Statistics Education*, **22**.

Cummiskey, K., Kuiper, S., and Sturdivant, R. (2012). Using classroom data to teach students about data cleaning and testing assumptions. *Frontiers in Psychology*, **3**.

Ekstrøm, C. T. (2016). *The R Primer*. CRC Press, second edition.

Gelman, A. and Nolan, D. (2002). *Teaching Statistics: A bag of tricks*. Oxford University Press.

Pfaff, T. J. and Weinberg, A. (2009). Do hands-on activities increase student understanding?: A case study. *Journal of Statistics Education*, **17**.