

3

Allan C. Malmberg

Datasæt i samspil

- Krydstabeller
- Grafer
- Mærketal

INFA-Matematik:

Informatik i matematikundervisningen

Et delprojekt under

INFA: Informatik i skolens fag
Et forskningsprogram på Danmarks Lærerhøjskole

Projektledelse:

Allan C. Malmberg
Inge B. Larsen
Viggo Sadolin

Distribution af programmer og tekster:

INFA, Danmarks Lærerhøjskole
Emdrupvej 115B, 2400 NV

*

Tekst: Allan C. Malmberg

Layout: Leif Glud Holm

© INFA 1999

Talkunnen

Serien Talkunnen indeholder fremstillinger af en række emner som kan give ideer til behandlingen af tal i en elementær undervisning.

Indholdsfortegnelse TALKUNNEN 3

Datasæt i samspil	4
MATTEST	4
INFAKRYDS åbnes	5
Matematiktesten	6
Aktiviteter	9
Udelad skemaer	9
Sammenhæng	12
Grupperinger	14
Nyt svar defineres	16

Datasæt i samspil

Vi skal i det følgende gøre brug af INFA-programmet INFAKRYDS (med programnavnet INFAX). Dette program kan benyttes i arbejdet med talmaterialer hvor man ønsker at belyse samspillet mellem to datasæt.

Programmet kan fx benyttes ved spørgeskema-undersøgelser. Her indhentes for en række personer deres svar på de opstillede spørgsmål. Svarene indlægges i INFAKRYDS, og der kan derefter foretages en række statistiske analyser af talmaterialet.

Til INFAKRYDS er udarbejdet en brugervejledning (INFA 1994, MI 95). Den bør du have til rådighed indtil du er fuldt fortrolig med programmet. - Der er også i programmet indlagt en hjælpefunktion som kan være til nytte undervejs i arbejdet.

MATTEST

Som eksempel materiale vil vi benytte resultater fra en matematiktest. Den består af to prøver, den ene med talopgaver, den anden med tekstopgaver. I testen deltager elever fra 9. klasser på fem skoler, i alt er der data fra 130 elever.

Resultaterne er indlagt i en datafil MATTEST som kan hentes ind i INFAKRYDS til bearbejdning. For hver af de 130 elever er der givet oplysninger om følgende:

- Dreng eller pige (0 for dreng, 1 for pige)
- Skole (nummereret 1-5)
- Point i talopgaverne (0-42 point)
- Point i tekstopgaverne (0-24 point)
- Samlet pointtal (0-66 point)

For hver elev foreligger altså fem tal. Disse tal findes indlagt i datafilen MATTEST. Her findes 130 skemaer, ét for hver af eleverne.

INFAKRYDS åbnes

Gå ind i programmet INFAKRYDS og klik på menupunktet *Disk*. Her vælger du „Hent“, og der åbnes en liste over de datafiler som er til rådighed for INFAKRYDS. Udpeg MATTEST og klik på feltet *Hent*. Matematiktestens data er nu til rådighed i INFAKRYDS.

På skærmen vil du se det første af de 130 skemaer med elevernes data. Du kan blade i skemaerne ved hjælp af piletasterne for nedenu på skærmen. Men der er ingen grund til at se nærmere på de enkelte skemaer, det er der mere behov for hvis der skal rettes i nogle af de data der er indlagt.

Beskriv. Gå ind i menupunktet *Beskriv* og vælg *Et spørgsmål*. Du skal nu udpege et af de fem spørgsmål på svarskemaet. Vælg fx *Skole*.

Du får på skærmens højre side en oversigt over hvor mange elever der deltager fra hver skole. Oversigten gives i form af en sumtabel. Du ser at der fra skole 1 deltager 27 elever, fra skole 2 deltager 29 elever. Fra skole 5 er der kun 22 elever.

		Skole			
Svar		h	f	H	F
1		27	0.208	27	0.208
2		29	0.223	56	0.431
3		27	0.208	83	0.638
4		25	0.192	108	0.831
5		22	0.169	130	1.000

Lad os se hvordan de fordeler sig på drenge og piger. Klik igen på *Beskriv* og vælg igen *Et spørgsmål*. Udpeg denne gang *Køn*.

Du får nu en oversigt over hvordan fordelingen på drenge og piger er: 67 drenge og 63 piger.

Vi vil nu se hvordan fordelingen på drenge og piger er på de enkelte skoler. Vælg *Beskriv* og klik denne gang på *To spørgsmål*. Klik nu på *Køn* og derefter på *Skole*.

På skærmen kommer nu en *krydstabel* som viser hvordan fordelingen på drenge og piger er på de fem skoler. I tabellen kan du aflæse:

Skole	0 (dreng)	1 (pige)	I alt
1	15	12	27
2	16	13	29
3	14	13	27
4	13	12	25
5	9	13	22
	67	63	130

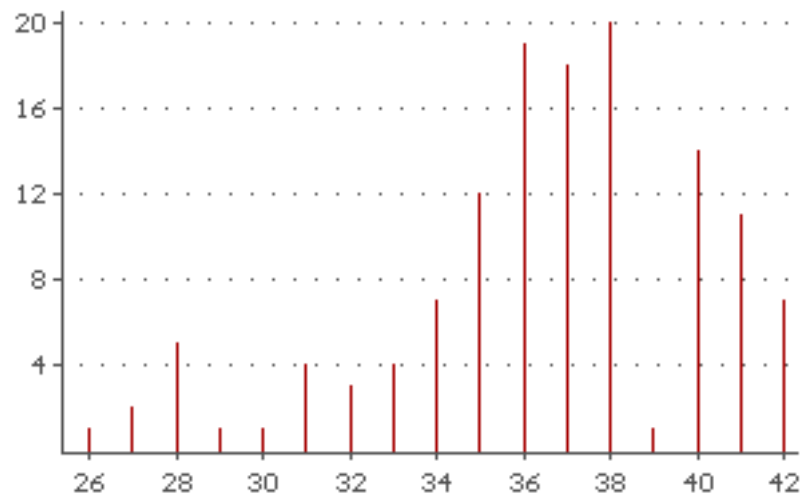
Tabellen hedder en krydstabel fordi den kombinerer eller „krydser“ to spørgsmål, her spørgsmålene *Køn* og *Skole*.

Matematiktesten

Vi vil nu se på resultaterne fra matematiktesten. Gå ind i *Beskriv* og vælg *Et spørgsmål*. Klik dernæst på *Talopgaver*.

På skærmen kommer en tabel over de 130 elevers resultater i talopgaverne. Tabellen foreligger som en sumtabel. Hele tabellen kan ikke være på skærmen på en gang, men du kan klikke på piletasterne i tabellens bund og dermed få tabellen til at rulle op og ned på skærmen.

Nederst på skærmen kommer et grafisk billede af resultaternes fordeling. Billedet er givet som et pindediagram.



Af tabel og diagram kan du se at pointtallene i talopgaverne spred sig fra 26 til 42 point. Ingen af de 130 elever har altså fået under 26 point, og du kan se at syv elever har fået maksimumpoint, nemlig 42.

Nogle klik-muligheder. Afprøv følgende: Klik på bogstavet *f* i tabellens hoved og undersøg hvad der sker med den lodrette akse i det grafiske billede. Klik derefter på *F*, og du får et nyt grafisk billede.

Prøv også med et klik på en linie i tabellen. Og prøv med et klik på et tal på grafens vandrette akse.

Klik endvidere på lup-symbolet midt på skærmen i den nederste kant. Du får da et grafisk billede i fuld skærmbredde. Du vender tilbage til det oprindelige billede ved et klik på *Ok*.

Udskriv. Gå til menupunktet *Udskriv* og vælg en udskrift af tabellen. Gå til *Udskriv* igen og vælg denne gang en udskrift af grafen.

Mærketal. Klik på feltet *Mærketal*. Du får nu en liste over statistiske mærketal. Du vil kunne se hvor mange svar der indgår i stati-

stikken, nemlig 130, og derefter vil du kunne se hvad det mindste og det største pointtal er. Du vil også kunne aflæse at typetallet i fordelingen er 38, altså er 38 det pointtal der forekommer hyppigst blandt de 130 deltagere.

Endvidere vil du kunne se hvad kvartilsættet er: Nedre kvartil, median og øvre kvartil. Endelig er talsættets middeltal anført.

Mærketal		
Antal svar	:	130
Mindste svar	:	26
Største svar	:	42
Variationsbredde	:	16
Type svar	:	38
Nedre kvartil	:	35
Median	:	37
Øvre kvartil	:	39
Middeltal	:	36.485
Varians	:	13.219
Standardafvigelse	:	3.636

Ved et klik på *Udskriv* kan du få en printerudskrift af mærketallene.

* * *

Vi vil nu gennem en række aktiviteter give dig en indførelse i programmets muligheder. Alle aktiviteter knytter sig til datafilen MAT-TEST.

Aktivitet 1: Tekstopgaver

Undersøg hvordan resultaterne i tekstopgaverne fordeler sig for de 130 elever. Udskriv tabel, graf og mærketal og giv dine kommentarer.

Aktivitet 2: Vi ser på én skole

Vi vil se hvordan eleverne fra en af de fem skoler har klaret sig i matematiktesten. Lad os vælge skole 2.

Vi skal nu ikke bruge alle de 130 skemaer i vor statistik, men kun dem der hører til skole 2.

Gå ind i menupunktet *Tilret* og vælg *Udelad skemaer*. Der åbnes nu for en indtastning ud for hvert af skemaets spørgsmål. Vi skal her indtaste ud for *Skole*. Vi skal angive at alle skemaer der ikke hører til skole 2, skal udelades.

Dette kan vi angive på forskellig vis. Lettest er det blot at skrive

$<>2$

De to ulighedstegn betyder „mindre end eller større end“, dvs. „forskellig fra“. Med andre ord: Vi har skrevet at alle de skemaer der ikke hører til skole 2, skal udelades.

Afslut indtastningen med *Enter*. Hvis du nu blader i skemaerne, vil du se at alle de skemaer der ikke hører til skole 2 har fået en overstregning i øverste højre hjørne. De indgår altså ikke i statistikken.

Vælg herefter *Beskriv*, klik på *Et spørgsmål* og udpeg SUM. Der kommer nu på skærmen en statistik over hvordan eleverne på skole 2 har klaret sig i den samlede matematiktest.

Kontroller at materialet omfatter det antal skemaer som svarer til antallet af elever fra skole 2.

Udskriv tabel, graf og mærketal og studer materialet nærmere.

Hvis du vil have en udskrift af de skemaer der indgår i undersøgelsen, kan du gå til *Udskriv* og vælge *Svar*. Du får da udskrevet indholdet af alle de skemaer der er omfattet af statistikken.

Aktivitet 3: Vælg en anden skole

Udskift skole 2 med en anden af de fem skoler: Gå ind under *Tilret*, vælg *Udelad skemaer* og indtast den nye skoles nummer.

Udskriv tabel, graf og mærketal. Benyt udskrifterne til at foretage en sammenligning af resultaterne fra de to skoler.

NB. Når du har fået udskrevet en statistik hvor nogle skemaer er udeladt, så bør du gå ind og slette betingelserne, ellers gælder de automatisk i din næste undersøgelse: Gå ind under *Tilret*, vælg *Udelad skemaer* og slet de anførte betingelser.

I brugervejledningen kan du på side 8 se hvordan betingelser om udeladelse af skemaer kan udformes. Ved udeladelse af skole 1,3,4, og 5 kunne vi også have skrevet:

$=1 \vee =3 \vee =4 \vee =5$

idet „eller“ her er angivet ved bogstavet \vee . - En anden mulighed for at angive at skole 1, 3, 4 og 5 skal udelades havde været:

$=1 \vee >2$

Husk at vi skal angive de skemaer der *udelades*, ikke dem der skal indgå i undersøgelse.

Er du i tvivl om en angivet betingelse udtrykker det ønskede, så blad i skemaerne og kontroller at det er de rigtige skemaer der udelades.

Aktivitet 4: Dreng og piger på skole 3

Vi vil nu sammenligne pointtallene for drenge og piger på skole 3.

Indsæt under *Skole* en betingelse så kun skole 3 indgår i statistikken.

Indsæt endvidere under *Køn* en betingelse som bevirker at pigernes skemaer udgår.

Vælg dernæst *Beskriv*, klik på *Et spørgsmål* og vælg SUM.

Udskriv derefter tabel, graf og mærketal over drengenes præstationer.

Udskift betingelsen under *Køn*, så det nu er drengenes skemaer der udelades. Udskriv derefter en statistik over pigernes præstationer.

Foretag på grundlag af de udskrevne statistikker en sammenligning af drengenes og pigerne præstationer ved matematiktesten.

Aktivitet 5: Benyt en forskelstest

Benyt de udskrevne pointtal for drenge og piger i aktivitet 4 til at undersøge om der foreligger en signifikant forskel mellem drengenes og pigernes præstationer ved matematiktesten på skole 3. (Vedr. forskelstest: se Talkunnen 2).

Aktivitet 6: Hvem klarer sig bedst i talopgaver?

Udskriv en statistik over drengenes resultater i talopgaverne på skole 3, og udskriv derefter en tilsvarende statistik for pigerne.

Studer de to statistikker nærmere og overvej om du mener der er nogen signifikant forskel mellem de to gruppers præstationer.

Udfør derefter en forskelstest på de to datasæt.

Aktivitet 7: Sammenhæng?

Vi vil nu se på om der er en sammenhæng mellem de resultater der opnås i talopgaverne og dem der opnås i tekstopgaverne.

Vi holder os til skole 1. I betingelserne anfører vi derfor under *Skole* at vi udelader skemaerne for skolerne 2, 3, 4 og 5.

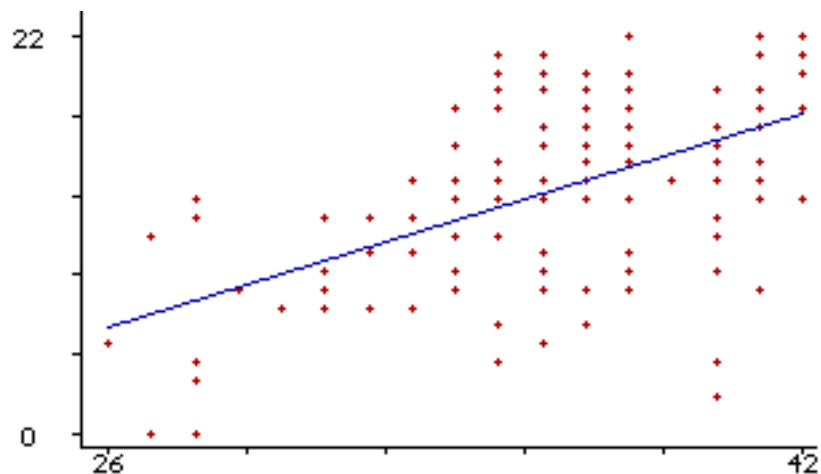
Dernæst går vi ind under *Beskriv* og vælger *To spørgsmål*.

Vi klikker på Talopgaver og på Tekstopgaver. Der kommer nu på skærmen en krydstabel som viser hvordan pointtallene i de to prøver er kombineret. Det kan være svært at overskue tabellen, selv om vi lader den rulle op og ned på skærmen.

Også en udskrift vil give et uoverskueligt billede, men vi skal senere se hvordan der kan opnås en mere håndterlig tabel.

Det interessante er imidlertid det grafiske billede nederst på skærmen. Her er i et koordinatsystem afsat punkter som viser hvordan pointtallene er kombineret. Ud ad den vandrette akse har vi afsat pointtallene i talopgaverne, og ud ad den lodrette akse har vi afsat pointtallene i tekstopgaverne.

Ved et klik på feltet *Linie* bliver der indtegnet en linie i det grafiske billede. Ved et klik på lup-symbolet midt på skærmens nederste del får vi tegnet et grafisk billede i fuld skærbredde. Den indtegnede linie er konstrueret således at det er den linie der stemmer bedst med de punkter der er afsat i koordinatsystemet: Vi tager alle de afsatte punkters lodrette afstande fra linien og kvadrerer og adderer disse afstande. Denne sum bruger vi som et udtryk for punkternes afvigelse fra linien. Den tegnede linie er den der giver den mindste afvigelse mellem punkter og linie. Denne linie kaldes i statistikken for *regressionslinien* for det kombinerede datasæt.



Klik på Ok og gå over til *Mærketal*. Her findes nogle oplysninger om den indtegnede regressionslinie. Dens ligning vil være anført: $Y = aX + b$, hvor værdierne af a og b er oplyst. Endvidere er der anført et tal for den såkaldte *korrelationskoefficient*, r . Dens værdi er et udtryk for hvor godt punkternes beliggenhed stemmer overens med linien, altså hvor stærk en sammenhæng der er tale om. Hvis værdien af r er positiv, så er der tale om en positiv sammenhæng mellem de to datasæt, dvs. høje pointtal i den ene prøve svarer til høje pointtal i den anden. Er r negativ, er der tale om en negativ sammenhæng mellem pointtallene i de to prøver. Jo nærmere r er på 1 eller -1, jo stærkere sammenhæng er der tale om. Hvis alle punkter ligger på regressionslinien vil r være 1 eller -1, alt efter om linien har en positiv eller en negativ hældning.

Vi kan ikke umiddelbart ud fra værdien af r afgøre om der foreligger en signifikant positiv eller negativ sammenhæng mellem de to datasæt. Dette spørgsmål må afgøres ved en sammenhængstest.

I den foreliggende situation viser oversigten over mærketal at korrelationskoefficienten har en værdi på over 0.70. Dette er et udtryk for en positiv sammenhæng mellem resultaterne i talopgaverne og resultaterne i tekstopgaverne.

Dit valg af rækkefølge for udpegning af de to spørgsmål er afgørende for hvad der afsættes på de to akser. Det er det først udpegede spørgsmål som lægger beslag på den vandrette akse. Ved et passende valg kan du altså selv bestemme hvad der skal afsættes ud ad de to koordinatakser.

Aktivitet 8: En sammenhængstest

Udfør en sammenhængstest (se Talkunnen 2) på resultaterne i talopgaverne og resultaterne i tekstopgaverne for de 27 elever på skole 1. De nødvendige data vil du kunne aflæse af krydstabellen i aktivitet 7.

Aktivitet 9: Resultaterne grupperes

Vi vil nu have en mere overskuelig udgave af krydstabellen fra aktivitet 7. Vi holder os stadig til skole 1, og går nu ind i menupunktet *Tilret*. Her vælger vi *Grupper svar*. Ud for Talopgaver anfører vi følgende:

21:30/31:40/41:50/51:60/61:70

Hermed får vi resultaterne opdelt i fem grupper. Den første indeholder pointtallene fra 21 til 30, den næste pointtallene fra 31 til 40, osv.

Tilsvarende inddeler vi pointtallene under Tekstopgaver i grupper. Her kan værdierne ikke overstige 24, så vi vælger følgende gruppeinddeling:

0:4/5:9/10:14/15:19/20:24

Herefter vælger vi under menupunktet *Beskriv* de samme to spørgsmål som i aktivitet 7.

Udskriv krydstabellen og bemærk hvordan der nu opdeles i grupper for de to datasæt. Kig på mærketal, og bemærk at de ikke stemmer helt overens med dem vi fik i aktivitet 7.

Heller ikke det grafiske billede er helt det samme som før. Det skyldes at programmet ved grupperede data benytter de enkelte grupperes midtpunkt som værdi for de data der ligger i gruppen.

Aktivitet 10: Hele materialet

Fjern betingelserne om valg af skole, men behold gruppeinddelingen fra aktivitet 9. Udfør en statistik for hele materialet: Lav en krydstabel for alle 130 elevers resultater i Talopgaver og i Tekstopgaver. Undersøg hvad korrelationskoefficienten er.

Fjern dernæst gruppeinddelingen og udskriv et grafisk billede af de 130 punkter og den tilhørende regressionslinie.

Aktivitet 11: En modeltest

Udskriv en sumtabel som viser fordelingen af point i tekstopgaverne. Materialet skal omfatte alle 130 elever.

Udskriv endvidere mærketal for fordelingen af point.

Du vil her se at kvartilsættet for datamaterialet er opgivet til:

Nedre kvartil: 10 Median: 14 Øvre kvartil: 18

Det betyder at der i hvert af de fire områder 0-10 point, 11-14 point, 15-18 point, og 19-24 point vil findes ca. 25% af eleverne.

Vi vil benytte denne opdeling i en modeltest af de enkelte skolars resultater.

Vælg skole 1 og find ud af hvor mange af skolens elever der har opnået point i hver af de fire områder. Indtast derefter disse antal i en modeltest (se Talkunnen 2) og benyt som model en jævn fordeling. Undersøg om der er nogen signifikant afvigelse mellem model og data.

Foretag den samme modeltest for hver af de øvrige fire skoler.

Aktivitet 12: En modeltest for SUM

Undersøg hvad kvartilsættet er i det samlede materiale for summen af pointtallene i de to prøver. Udfør dernæst en modeltest for én eller to af de fem skoler.

Aktivitet 13: En ny sum

Tilrettelæggerne af matematiktesten har en ide om at de to pointtal for talopgaverne og tekstopgaverne ikke bør vægtes lige meget. De vil derfor undersøge hvilken betydning det har at de lader pointtallene i tekstopgaverne tælle dobbelt.

Dette kan vi nemmest undersøge ved at indføre et nyt spørgsmål i de 130 skemaer. Vi kan da lade programmet udføre alt regnearbejdet.

Gå ind i menupunktet *Tilret* og vælg *Afled nyt svar*. Gå ind i det første frie felt i skemaet, dvs. ved spørgsmål 6, og indskriv her en tekst, fx NYSUM. Afslut med et tryk på *Enter*.

I feltet tilhøjre på skærmen indskrives du nu:

$s3 + 2*s4$

Dermed udtrykker du at pointtallet for det ekstra spørgsmål beregnes som summen af svaret ved spørgsmål 3 og det dobbelte af svaret ved spørgsmål 4. Afslut med et klik på højre musetast. Du vil da i skemaerne kunne se at det nye spørgsmål og de tilhørende svar er indlagt.

Herefter kan du i anvendelsen af MATTEST inddrage det nye spørgsmål:

Undersøg hvordan pointtallene fordeler sig i NYSUM for hele materialet.

Undersøg hvilken sammenhæng der er mellem resultaterne i SUM og resultaterne i NYSUM. Udskriv et diagram der viser regressionslinien.

NB Når du har indført et nyt spørgsmål i materialet, må du gå ind under menupunktet *Disk* og vælge *Gem*. Du gemmer da skemaerne med de nye spørgsmål og svar. Som navn for datafilen kan du blot vælge det allerede benyttede navn, MATTEST.

Aktivitet 14: Brug NYSUM

Vælg en af de fem skoler. Udskriv en statistik som viser drengenes og pigernes resultater i NYSUM.

Undersøg ved en forskelstest om der er en signifikant forskel mellem drengenes og pigernes resultater.

Aktivitet 15: Forskellen mellem de to prøver

Indlæg i programmet som et nyt spørgsmål forskellen i pointtal mellem prøven i Talopgaver og prøven i Tekstopgaver.

Benyt som afledt spørgsmål:

FORSKEL s3 - s4

Undersøg hvordan denne forskel fordeler sig i det samlede materiale. Udskriv tabel og graf.

Aktivitet 16: Find eleverne

Hvordan finder vi de elever, eller rettere de skemaer, som svarer til bestemte præstationer?

Overvej hvordan du vil finde den eller de elever der har klaret sig bedst i Tekstopgaverne. Du vil gerne vide hvilken skole de kommer fra, og du vil vide om der er tale om drenge eller piger.

Overvej hvordan du kan foretage denne søgning. Udfør den ikke ved at gennemblade alle 130 skemaer, men prøv om du på en smart måde kan bruge menupunktet *Tilret* med *Udelad skemaer*.

Find også den eller de elever der har den største forskel mellem pointtallene i Talopgaver og Tekstopgaver.

Aktivitet 17: Dit eget datasæt

Opbyg dit eget datasæt og indlæg det i INFAKRYDS. Som eksempel kan nævnes et datasæt der indeholder oplysninger om en konditest i en elevgruppe. Om hver elev kan fx indlægges følgende oplysninger:

Køn, Alder, Højde, Vægt, Kondital, Cigaretforbrug

Datasættet bør omfatte mindst 25 personer.

Overvej eventuelle andre oplysninger der kan inddrages.

Aktivitet 18: Andre ideer til datasæt

Giv forslag til datasæt som kunne benyttes i arbejdet med INFAKRYDS.

MI 144

ISBN 87-7701-680-7
ISSN 1398-6716