MI 140

ISBN 87-7701-630-0

TALKUNNEN

2

Allan C. Malmberg

Tre test:

- Modeltest
- Forskelstest
- Sammenhængstest

INFA Matematik - 1998

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

TALKUNNEN

2

Allan C. Malmberg

Tre test:

- Modeltest
- Forskelstest
- Sammenhængstest

Talkunnen

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

Indholdsfortegnelse TALKUNNEN 2

Modeltest 3
1. Er terningen god nok? 3
2. Hvad kan vi opnå med en modeltest? 9
3. Opgaver 11

Forskelstest 19
1. Et eksempel 19
2. Programmet Forskelstest 24
3. Opgaver 29

Sammenhængstest:
1. Et eksempel 37
2. Vi anvender programmet TEST 39
3. Sammenhæng og forklaring 42
4. Opgaver 43

Modeltest:

Stemmer data med modellen?

1. Er terningen god nok?

Sanne udfører et eksperiment der består i 60 kast med en terning. Hun fører regnskab med hvilke øjental der forekommer i de 60 kast. Her er hendes tabel over kasteresultaterne:

Øjental	1	2	3	4	5	6
Hyppighed	6	12	9	13	13	7

Hun er lidt forbavset over disse tal. Hun ved jo at der ved et kast med en terning er samme chance for hvert af de seks øjental, det undrer hende derfor at der er så stor forskel på hvor tit de enkelte øjental forekommer: Øjentallet 1 er kun forekommet 6 gange, men øjentallene 4 og 5 er begge forekommet 13 gange i de 60 kast.

Sanne spekulerer på om der kan være noget galt med terningen. Hvis den er skæv på en eller anden måde, så kan det jo være forklaringen på at tallene afviger så meget fra det forventede.

Hvis de seks øjental har samme chance for at forekomme, ville Sanne vente at alle de observerede hyppigheder vil ligge tæt på 10. Der er jo 60 kast, og hvis de skulle fordele sig ligeligt på de seks muligheder, så måtte der være 10 af hver slags.

Sanne arbejder her med en tænkt *model*, et billede af hvordan kastene ville fordele sig i en ideel virkelighed. Men hun ved godt fra tidligere eksperimenter med terningkast at virkeligheden kan afvige en del fra modellen. Kaster hun seks kast, er det jo ikke særligt ofte at det vil forekomme at hvert af de seks øjental forekommer lige præcis én gang. Og i en serie på 60 kast er det selvfølgelig heller ikke at forvente at hvert øjental forekommer lige præcis 10 gange.

Men hvor meget må virkelighedens data nu afvige fra modellen før Sanne bør få mistanke om at der er noget galt?

Til at besvare dette spørgsmål gør vi brug af en statistisk test, *en modeltest*. Ved hjælp af den kan vi få oplysning om hvor godt tallene fra eksperimentet stemmer med den opstillede model.

Vi gentager her Sannes tabel over de opnåede resultater, men vi tilføjer de tal som vi ville forvente efter den opstillede model:

Øjental	1	2	3	4	5	6
Hyppighed	6	12	9	13	13	7
Model	10	10	10	10	10	10

Der er her tale om en model som giver samme chance for de alle de foreliggende resultater. En sådan model kalder vi *en jævn fordeling.*

Når vi skal vurdere hvor meget Sannes tal afviger fra den opstillede model, så kigger vi på forskellen mellem de observerede antal og dem der forudsiges af modellen. Ved øjentallet 1 tager vi altså forskellen mellem 6 og 10, ved øjentallet 2 tager vi forskellen mellem 12 og 10, osv.

Disse forskelle kvadreres og adderes:

$$(6-10)^2 + (12-10)^2 + \ldots + (7-10)^2 = 48$$

Vi kan se at hvis Sanne havde fået tal der præcist passede med modellen, så ville alle tal i parenteserne være 0, og så ville den samlede sum også være 0.

Vi bruger denne sum af de kvadrerede tal som et udtryk for hvor meget de opnåede resultater afviger fra modellens tal. Jo større summen er, desto større er afvigelsen mellem de observerede data og modellens tal.

Vi anvender nu modeltesten på Sannes tal. Den foreligger som et edb-program, *Modeltest*, så vi er fri for at foretage beregninger ved håndkraft. Modeltest er et af tre programmer der findes i INFA-programmet TEST.

Sanne fortæller først programmet at der er tale om seks forskellige resultattyper. Der åbnes nu seks felter på skærmen, og Sanne indtaster her de kasteantal hun opnåede ved eksperimentet: 6, 12, 9, 13, 13, 7.

Derefter fortæller hun at den model der skal testes er en jævn fordeling. I modellinien indsætter programmet nu tallet 10 i hver rubrik.

Programmet giver herefter oplysninger om hvor godt de observerede data stemmer med en opstillede model:

Testresultat:
Ingen signifikant afvigelse mellem model og data.

Ordet „signifikant“ betyder *påfaldende* eller *betydende*.

Programmet melder altså her at der intet påfaldende er ved de foreliggende data, og der er derfor ikke grund til at tvivle på den opstillede model.

Vi kan fortælle dig at der ved en kasteserie af den foreliggende art er ca. 45% chance for at få resultater som afviger lige så meget - eller endnu mere - end dem Sanne fik.

Med andre ord: I næsten hver anden kasteserie kan der forventes en så stor afvigelse som den der var i Sannes kasteserie.

Modeltesten viser således at Sanne ikke har nogen grund til at mistænke sin terning for at være skæv. Afvigelser som den hun fik, er ganske almindelige.

En ny kasteserie

Malene fremlægger følgende resultater fra 60 kast med en terning:

Øjental	1	2	3	4	5	6
Hyppighed	9	12	10	10	11	8

Vi kan hurtigt se at Malenes tal stemmer bedre med en jævn fordeling end Sannes tal gjorde.

Vi udfører en modeltest på Malenes tal. Vi får følgende oplysninger fra programmet:

Testresultat:
Signifikant lille afvigelse mellem model og data.

Programmet udskriver endvidere:

Testresultatet signifikant på 5%-niveauet:							
10%	5%	x	2.5%	1%	0.5%	0.1%	0%

Krydset på chancelinien viser at når den opstillede model er korrekt, så er der under 5% chance for at opnå et testresultat der stemmer så godt med modellen.

På chancelinien er der seks områder hvori et kryds kan afsættes. Det første svarer til at testresultatet er signifikant på 10%-niveauet, det sidste svarer til at testresultatet er signifikant på 0.1%-niveauet.

Malenes data giver ikke nogen som helst grund til at tvivle på den opstillede model. Hvis man indsamlede data fra 100 serier af kast, ville vi kunne forvente at der i mindre end 5 af serierne ville forekomme en bedre overensstemmelse mellem model og data end dem vi ser i Malenes kasteserie.

Men Sanne kommer nu i tvivl: Er Malenes data ægte? Har hun virkelig udført de 60 kast, eller har hun blot digtet sine kasteresultater? De er jo næsten for gode.

Modeltesten viser at der kun i 5 tilfælde ud af 100 kan forventes så „pæne resultater“ som dem Malene har fremlagt. Sanne har derfor et godt grundlag for at udspørge Malene lidtnærmere om de udførte kast, men hvis Malene fastholder at de er gode nok, må Sanne tage det til efterretning. Der er jo dog mellem 2.5% og 5% chance for at Malenes data er ægte.

Modellen forkastes

Sanne har mistanke til en terning som ser ud til at være lidt ekstra tung i den ene ende. Hun tester den ved en kasteserie på 60 kast. Her er hendes resultater:

Øjental	1	2	3	4	5	6
Hyppighed	16	5	9	7	6	17

Som model anvender Sanne en jævn fordeling.

En modeltest giver følgende resultat:

Testresultat:

Signifikant stor afvigelse mellem model og data

Programmet udskriver endvidere:

Testresultatet signifikant på 2.5%-niveauet:

10% 5% 2.5% x 1% 0.5% 0.1% 0%

Ud fra dette konkluderer Sanne at der nok er tale om en skæv terning. Hvis terningen var ægte, ville der jo være under 2.5% chance for at få så stor en afvigelse som den der findes i de givne data. Den opstillede model, en jævn fordeling, forkastes. Men nye data kan selvfølgelig ændre beslutningen.

Krydset på chancelinien er her sat i området mellem 2.5% og 1%. Jo længere krydset er placeret mod højre på chancelinien, jo mindre chance er der for at de observerede data kan forekomme når den opstillede model er korrekt. Vi kan derfor sige at jo lavere testresultatets signifikans-niveau er når der foreligger en signifikant stor afvigelse mellem data og model, jo mere overbevist er vi om at den opstillede model ikke er den korrekte.

Modeltestens beregninger

Vi kan forestille os at modeltesten beregner testchancerne på følgende måde: Programmet udfører 1000 eksperimenter af den foreliggende art, og det opstiller en statistik over de afvigelser fra den given model der forekommer i de 1000 eksperimenter.

Hvis det fx viser sig at der er 90 eksperimenter af de 1000 der giver en afvigelse som er af mindst samme størrelse som den der er i de givne data, så fortæller programmet at der er en chance på 90/1000, dvs. 9% for at få en så stor afvigelse som den der foreligger.

Hvis der blev udført en ny serie på 1000 eksperimenter, kunne vi få lidt andre tal, men i praksis ville der ikke være store forskelle. Så vi opfatter modeltestens chancer som nogle der er fastlagt på grundlag af et stort antal af eksperimenter, der er altså tale om statistisk sandsynlighed.

2. Hvad kan vi opnå med en modeltest?

Når vi anvender en modeltest kan vi bruge den på følgende måde:

1. Hvis testen viser at afvigelsen mellem model og data er *signifikant stor*, vil vi være på vagt over for den opstillede model: Vi er i tvivl om den opstillede model kan accepteres.

2. Hvis testen viser at afvigelsen mellem model og data er *signifikant lille*, så kan vi være i tvivl om der er fusket med de givne data.

3. Hvis testen viser at der ingen *signifikant afvigelse* er mellem model og data, så har vi ingen grund til at tvivle på den opstillede model eller på ægtheden af de benyttede data.

I de to tilfælde, 1 og 2, kan vi måske ønske os nye data indsamlet før vi tager endelig stilling.

Øvelse 1
Udfør en modeltest på følgende resultater fra 60 terningkast:

Øjental	1	2	3	4	5	6
Hyppighed	9	11	9	11	11	9

Hvad er din konklusion?

Andre modeller
Ved modeltesten kan vi også teste modeller der ikke bygger på en jævn fordeling. Lad os tænke os at vi ved terningkast slår øjentallene 1 og 6 sammen i én gruppe, og lader de andre øjental udgøre hver sin gruppe.

Med tallene fra en af Sannes kasteserier har vi da:

Øjental	1 og 6	2	3	4	5
Hyppighed	13	12	9	13	13
Model	20	10	10	10	10

Ved anvendelse af modeltesten skal vi først meddele at der er tale om 5 grupper af data. Dernæst indtaster vi de opnåede hyppigheder, og dernæst de tal der er anført under Model. Modellen er jo denne gang *ikke* en jævn fordeling, der forventes ikke samme antal kast i de fem grupper: Første gruppe forventes at indeholde 20 kast, de øvrige fire hver 10 kast.

Modeltesten giver følgende resultat:

Testresultat:

Ingen signifikant afvigelse mellem model og data.

Modeltesten viser altså at der ikke er grund til at forkaste den opstillede model.

Modellens tal
Ved opstilling af en model skal du indtaste modellens tal i rubrikkerne i nederste linie. Disse tal må gerne indtastes som decimaltal, de angiver jo de forventede gennemsnitstal i et stort antal forsøg og behøver derfor ikke at være hele tal.

Af hensyn til modeltestens beregninger bør tallene i model-rubrikkerne ikke have værdier under 5. Hvis du opstiller en model med en eller flere rubrikker som indeholder en talværdi under 5, vil programmet opfordre dig til at slå nogle grupper sammen, således at du i alle rubrikker kommer op på værdier der ligger på 5 og derover.

3. Opgaver

I arbejdet med opgaverne bør du indlede med at tage et overblik over de givne data og give dit gæt på om der foreligger et signifikant resultat. Først derefter anvender du modeltesten.

Når du bliver mere øvet, kan du prøve at gætte på hvor stærk en signifikans der vil være tale om ved det foreliggende datasæt.

1. Dit eget eksperiment
Udfør 60 kast med en terning og lav en tabel over forekomsten af øjental.

Undersøg dine data ved hjælp af en modeltest.

2. Gentag eksperimentet
Gentag eksperimentet fra opgave 1 og udfør også en modeltest på det nye datasæt.

Slå derefter de to datasæt fra opgave 1 og 2 sammen til ét sæt og udfør en modeltest på dette sæt.

3. Et opdigtet datasæt
Få en kammerat til at digte et datasæt som viser hvordan 60 terningkast fordeler sig på de seks mulige øjental.

Undersøg datasættet ved en modeltest.

4. En falsk terning
En terning er forfalsket således at der er en chance på 50% for at få en sekser ved et terningkast, medens der er en chance på 10% for hvert af de øvrige øjental.

Terningen undersøges i en kasteserie på 100 kast. Der opnås følgende øjental:

Øjental	1	2	3	4	5	6
Hyppighed	6	9	6	10	8	61

Undersøg om disse data er overensstemmelse med den givne model.

5. Data fra LOD
Lad programmet LOD udskrive 100 tal fra området 1..10.

Undersøg ved en modeltest om de givne resultater er i overensstemmelse med en jævn fordeling over de ti mulige talværdier.

6. Mendels data
Den verdensberømte arvelighedsforsker Gregor Mendel fremlagde i 1865 data fra et forsøg med nedarvede egenskaber hos ærter. De fremlagte data skulle bekræfte den model som Mendel havde opstillet efter teoretiske overvejelser.

Ærterne var opdelt i fire kategorier efter udseende og farve.

	Observationer	Mendels model
Glat gul	315	313
Rynket gul	101	104
Glat grøn	108	104
Rynket grøn	32	35

Undersøg Mendels data ved en modeltest. Hvad er din konklusion om Mendels eksperimentelle data?

7. Er fagene lige populære?
I en undersøgelse deltager 300 skoleelever. De bliver spurgt om hvilket skolefag de synes bedst om. Svarene fordeler sig på fem fag:

Dansk	65
Matematik	88
Idræt	41
Engelsk	54
Fysik/kemi	52

Undersøg ved en modeltest om disse data stemmer med at der ikke er nogen væsentlig forskel på fagenes popularitet.

8. At slå to ens øjental
Tag dine resultater fra et eksperiment hvor du har undersøgt hvor svært det er at slå *To ens øjental* med to terninger.

Undersøg ved en modeltest om dine resultater passer med følgende model:

Antal kast	1-2	3-5	6-10	over 10
Chance	30%	30%	24%	16%

9. En anden model
Undersøg ligesom i foregående opgave om dit datasæt fra undersøgelsen af hvor svært det er at slå *To ens* passer med følgende model:

Antal kast	1-3	4-6	7-10	over 10
Chance	44%	26%	18%	12%

10. Kast med en tegnestift
Ved kast med en tegnestift kan der opnås to resultater:

Tegnestiften ender „på ryg“ eller den ender „på spids“.

Udfør mindst 50 kast med en tegnestift og undersøg ved en modeltest om det kan afvises at de to udfald har samme chance for at forekomme.

11. Er modellen god nok?
Når man kaster med to mønter, kan der forekomme følgende tre resultater:

Begge mønter viser krone
Begge mønter viser plat
En mønt viser krone, en viser plat

Udfør 60 kast med to mønter (brug små mønter i et raflebæger) og noter dine resultater i kasteserien.

Undersøg ved en modeltest om resultaterne stemmer med en jævn fordeling over de tre mulige resultater.

Udfør dernæst yderligere 60 kast og anvend en modeltest på det samlede datasæt fra de to kasteserier.

12. Tal fra virkeligheden
En optælling af cifferfordelingen i befolkningstallene for Danmarks amter giver følgende fordeling:

1:6 2:16 3:10 4:14 5:8 6:16 7:10 8:7 9:4

Cifferet 0 er ikke taget med, da det jo ikke kan stå forrest i et tal.

Undersøg ved en modeltest om der kan være tale om en jævn cifferfordeling.

13. Fysikkens tal
En optælling af cifferfordelingen i atomvægtene for de første 12 grundstoffer i det periodiske system giver følgende fordeling:

1: 13 2: 8 3: 2 4: 5 5: 3 6: 3 7: 4 8: 5 9: 12

Cifferet 0 er ikke taget med, da det jo ikke kan være første ciffer i et tal.

Undersøg ved en modeltest om der kan være tale om en jævn cifferfordeling.

14. En afvigelse på 10%
Undersøg ved en modeltest om der kan være tale om en jævn fordeling af krone og plat når der observeres en fordeling af krone- og platkast som ligger 10% fra det forventede, dvs. når der forekommer 40% krone og 60% plat:

Antal kast: 50	Krone: 20	Plat: 30
Antal kast: 100	Krone: 40	Plat: 60
Antal kast: 150	Krone: 60	Plat: 90
Antal kast: 200	Krone: 80	Plat: 120

Hvilken konklusion kan du give vedrørende testens følsomhed over for en afvigelse på en fast procentdel?

15. En afvigelse på 10
Undersøg ved en modeltest om der kan være tale om en jævn fordeling af krone og plat når der observeres en fordeling af krone- og platkast som ligger 10 fra det forventede, dvs. når antallet af kronekast er 10 mindre end halvdelen af kasteantallet.

Antal kast: 50	Krone: 15	Plat: 35
Antal kast: 100	Krone: 40	Plat: 60
Antal kast: 150	Krone: 65	Plat: 85
Antal kast: 200	Krone: 90	Plat: 110

Hvilken konklusion kan du give vedrørende testens følsomhed over for en afvigelse på et fast antal?

16. Hvis skævheden forstørres
I 60 kast med en terning opnås følgende resultater:

Øjental	1	2	3	4	5	6
Hyppighed	7	13	9	13	12	6

Undersøg ved en modeltest om disse resultater er i overensstemmelse med en jævn fordeling.

Vi tænker os nu at der udføres 60 ekstra kast med præcis den samme fordeling: „Vi ganger de første resultater med 2". Vi har herefter 120 kast der fordeler sig således på de enkelte øjental:

Øjental	1	2	3	4	5	6
Hyppighed	14	26	18	26	24	12

Undersøg fordelingen ved en modeltest.

Prøv derefter at gange de første resultater med 3 og udfør en modeltest på denne fordeling.

17. Hjælper det med træning?
24 skoleelever deltager i en konditest. Derefter går de ind i en periode med træning for at forsøge at forbedre deres kondital, og efter træningsperioden bliver deres kondital målt igen.

Det viser sig 16 af de 24 elever har forbedret deres kondital, medens 8 elever får et ringere resultat end ved første prøve.

Anvend en modeltest på disse tal: Er der nogen signifikant afvigelse fra en jævn fordeling mellem bedre og ringere kondital?

18. Historiske terninger
Terninger har været anvendt til spil i flere tusinde år. Der har dog ikke altid været tale om terninger af den kvalitet vi kender i dag, mange var uregelmæssige og kunne give skæve resultater.

Her er resultaterne fra 235 kast foretaget med en terning fra det 16. århundrede før vor tidsregning:

1: 37   2: 17   3: 49   4: 59   5: 28   6:45
Undersøg disse data ved en modeltest: Er der grundlag for at påstå at terningen var skæv?

Og her er resultater fra 240 kast med en terning fra omkring år 300:

1: 38   2: 47   3: 36   4: 31   5: 33   6: 55.

Undersøg også disse resultater ved en modeltest.

Forskelstest:

Er der forskel mellem de to datasæt?

1. Et eksempel

Vi skal nu se på en test som kan hjælpe os med at vurdere om der er en væsentlig forskel på to datasæt.

Vi ser på et eksempel:

Tre piger og fem drenge deltager i en matematikprøve. Deres point i prøven er følgende:

Pigerne: 15, 16, 30

Drengene: 18, 23, 27, 31, 33

Vi vil nu anvende en forskelstest på dette materiale. Den kræver ikke beregninger af gennemsnit eller andre statistiske mærketal. Den bygger alene på *en rangorden* af de opnåede pointtal.

Her er alle pointtal for de 8 elever opstillet efter størrelse:

15	16	18	23	27	30	31	33
P	P	D	D	D	P	D	D

I tilknytning til tallene har vi anført om pointtallet er opnået af en dreng eller en pige.

Et overblik over fordelingen af point på drenge og piger giver det indtryk at pigernes og drengenes placering i rækkefølgen ikke ser ud til at være ganske tilfældig. Det ser ud som om der er en skævhed med størst koncentration af pigerne på de lavere pointtal: „Pigerne klarer sig dårligere ved prøven end drengene."

Ved forskelstesten vil vi undersøge om der er statistisk grundlag for at fremsætte denne påstand.

Hvert af pointtallene giver vi nu det nummer, „den rang", det har i den opstillede rækkefølge:

15	16	18	23	27	30	31	33	Pointtal
P	P	D	D	D	P	D	D	
1	2	3	4	5	6	7	8	Rang

Herefter ser vi ikke mere på de foreliggende pointtal, men udelukkende på *rangtallene*, tallene i den nederste linie.

Vi ser at pigernes pointtal har rangene 1, 2 og 6. Disse tal har summen 9. Det vil sige at pigernes *rangsum* i den forelagte rækkefølge er 9.

Det er dette tal vi vil se nærmere på i forskelstesten. Vi vil undersøge om denne rangsum på 9 kan siges at være udtryk for en skævhed. En skævhed som er så udtalt, så *signifikant*, at den får os til at forkaste den hypotese at pigernes og drengenes placering i rangordenen er tilfældig.

Hvis det er rigtigt at pigerne klarer sig dårligere ved prøven end drengene, vil vi forvente at pigernes pointtal vil findes i den nedre ende af opstillingen, vi vil altså vente at finde at pigernes rangsum er „for lille".

Oversigt over rangsummerne

Testen udfører vi ved at sammenligne den observerede rangsum på 9 med alle de rangsummer der kan forekomme for de tre pigers pointtal.

Den laveste rangsum der kunne forekomme for de tre piger, er 6. Den forekommer jo hvis de tre piger er placeret på de tre første pladser:

Placering: 1, 2, 3 Rangsum: 1+2+3 = 6.

Den højeste rangsum der kan forekomme er 21. Den forekommer hvis de tre piger er placeret på de tre øverste pladser:

Placering: 6, 7, 8 Rangsum: 6+7+8 = 21

Rangsummerne kan altså i det forelagte tilfælde variere fra 6 op til 21:

6____21

Den rangsum pigerne opnåede ved prøven var 9. Spørgsmålet er nu hvor skævt en rangsum på 9 er placeret i forhold til alle mulige rangsummer.

Vi laver en lille oversigt over rangsummerne fra 6 op til 9:

Rangsum 6: 1+2+3

Rangsum 7: 1+2+4

Rangsum 8: 1+2+5, 1+3+4

Rangsum 9: 1+2+6, 1+3+5, 2+3+4

Ved optælling kan vi se at der er 7 tilfælde hvor rangsummen er på 6, 7, 8 eller 9.

Vi har altså at der ligger 7 rangsummer i området fra 6 op til 9:

6________9________________________________21
7 stk.

Spørgsmålet er nu: Hvor stor en del udgør de 7 af alle de rangsummer der findes i det samlede felt fra 6 op til 21.

Dette antal kan bestemmes ved en tællemetode: På hvor mange måder kan der udvælges 3 tal blandt 8. Hver gang vi opstiller en rangsum tager vi jo tre af tallene fra talmængden 1,2,3,4,5,6,7,8.

Det søgte antal er givet ved udtrykket:

$$K(8,3) = \frac{8 \cdot 7 \cdot 6}{1 \cdot 2 \cdot 3} = 56$$

Der er altså i alt 56 mulige måder at beregne rangsummen på i det forelagte eksempel.

Hver af disse 56 kombinationer af pigernes range tillægger vi nu samme sandsynlighed. Hvis vi antager at piger og drenge er lige dygtige til prøven, så vil ethvert sæt af range for de tre piger have samme chance for at forekomme, hvad enten det er rangene (1,2,3), rangene (2,4,6) eller fx rangene (6,7,8). Så hvis den opstillede hypotese er sand, vil enhver af de 56 mulige rangsæt have sandsynligheden 1/56.

De 7 muligheder i området fra rangsum 6 til rangsum 9 skal nu sammenlignes med det samlede antal på 56.

Vi kan derfor sige: Chancen for at få en rangsum i området fra 6 til 9 er 7/56 = 1/8 = 12.5%. Testen har dermed beregnet en testchance på 12.5%, dvs. der er 12.5% chance for at få en rangsum som er så lille som den foreliggende, eller endnu mindre.

En testchance på 12.5% er almindeligvis ikke noget der giver grund til forkastelse af en opstillet hypotese. Vi vil derfor konkludere at der ikke er opnået så skævt et resultat at der er grund til at forkaste den opstillede hypotese:

Testresultatet viser ikke nogen signifikant forskel mellem de to datasæt, pigernes og drengenes.

Tallene giver os altså ikke grundlag for at hævde at prøven viser at drengene er bedre end pigerne.

Hvis pigerne i stedet havde været placeret som nr. 1, 2 og 4, så havde deres rangsum været 7. Af oversigten ser vi at der kun er to måder hvorpå vi kan få en rangsum på 7 eller derunder. Chancen for en rangsum i området fra 6 til 7 er derfor: 2/56 = 3.6%.

Denne chance på 3.6% ville måske få os til at sige at den opstillede hypotese skulle forkastes. Vi kan derfor i denne situation konkludere: Prøveresultatet tyder på at drenge og piger ikke klarer sig lige godt ved prøven. Da pigernes resultat ligger i det venstre område, „deres rangsum er for lille", kan vi mere tydeligt sige: *Det ser ud til at pigerne klarer sig dårligere end drengene*.

Testen ville i dette tilfælde føre til resultatet:

> *Drengene har klaret sig signifikant bedre end pigerne.*

Og der ville blive tilføjet:

> *Testresultatet er signifikant på 5%-niveauet.*

Det betyder at der er mindre end 5% chance for at et så godt resultat for drengene kan forekomme når der foretages en tilfældig fordeling af range til de to grupper.

Vi anvender kun rangene

Den opstillede forskelstest kaldes også en *rangsumtest*. Den bygger jo sine beregninger på de rangsummer der kan forekomme.

Som du har set, benytter en rangsumtest kun rangene for de forelagte data, derimod ikke talværdierne. Det gør rangsumtesten til et lettilgængeligt værktøj. Men dette er selvfølgelig ikke uden konsekvenser. Ved kun at benytte rangene for de forelagte data, går en række informationer tabt, idet der jo ofte ligger væsentlige oplysninger om de givne data i deres talværdier.

Der findes da også forskelstest som gør brug af de forelagte talværdier og ikke kun af deres range. Nogle af disse test har imidlertid en række forudsætninger som ikke altid kan opfyldes, og det kan i øvrigt være svært at afgøre om de pågældende forudsætninger er opfyldt eller ej. Man vil derfor ofte stille sig tilfreds med at benytte en rangsumtest som alene forudsætter at der foreligger data der meningsfyldt kan opstilles i en rangorden.

2. Programmet Forskelstest

Vi skal nu se hvordan programmet Forskelstest kan benyttes ved undersøgelse af om der foreligger signifikante forskelle mellem to datasæt. Programmet findes i INFA-programmet TEST.

I programmet åbner der sig en menu med følgende muligheder:

> Enkeltdata
> Gruppedata

Vi vælger punktet *Enkeltdata*. Programmet beder nu om oplysninger om hvor mange data der er i de to sæt. I vort eksempel består det ene sæt af data for pigerne, det andet sæt af data for drengene.

Vi svarer derfor at det ene sæt består af 3 data og det andet af 5 data. Herefter spørger programmet hvilke data der foreligger for de to sæt. Vi svarer med først at indtaste tallene 15, 16 og 30. Derefter indtaster vi tallene for drengene: 18, 23, 27, 31 og 33.

Vi går nu til *Udfør test*. Her får vi følgende oversigt på skærmen:

> *Testresultat: Ingen signifikant forskel mellem A og B*

Vi anvender herefter programmet på den situation at pigernes pointtal ligger som nr. 1, 2 og 4 i rangfølgen. Vi kan fx benytte pointtallene 15, 16 og 20 for pigerne, medens vi for drengene bruger de samme pointtal som før.

Vi får da følgende udskrift fra programmet:

Testresultat: B har signifikant højere range end A

Og der udskrives endvidere:

Testresultatet er signifikant på 5%-niveauet:

10%   5% x   2.5%   1%   0.5%   0.1%   0%

Vi kan så overveje om vi ud fra dette tør konkludere at pigerne klarer sig dårligere ved prøven end drengene. Men nu har vi i hvert fald et statistisk grundlag for at fremsætte en sådan påstand.

Programmet benytter betegnelserne A og B for de to datasæt. På skærmen kan du se hvilke datasæt de to bogstaver står for.

En forskelstest på de to datasæt A og B kan føre til tre mulige resultater:

Ingen signifikant forskel mellem A og B

A har signifikant højere range end B

B har signifikant højere range end A

Øvelse 1
Lad programmet foretage en beregning for den situation hvor de tre piger havde opnået pointtallene 19, 20 og 25, medens drengene har de samme pointtal som før.

Øvelse 2
Lad programmet foretage en beregning for den situation hvor de tre piger havde opnået rangene 3, 7 og 8. Digt selv nogle pointtal der passer til disse range. Hvilken konklusion ville du give i dette tilfælde om drengenes og pigernes dygtighed?

Sammenfald af talværdier
Lad os antage at pointtallene for de tre piger er 15, 17 og 30, og drengenes pointtal er 17, 23, 27, 31 og 33. I dette tilfælde er der pointsammenfald, idet pointtallet 17 er opnået af to elever.

I en sådan situation fordeler man rangene således at de point der har pointsammenfald deler de pågældende range:

15	17	17	23	27	30	31	33
P	P	D	D	D	P	D	D
1	2.5	2.5	4	5	6	7	8

De to elever der begge har opnået 17 point har delt rangene 2 og 3 og derfor begge fået 2.5 som rang.

Rangsummen for pigerne er i denne situation: 1 + 2.5 + 6 = 9.5.

Øvelse 3
Afprøv programmet: Vælg *Enkeltdata* og indtast de 8 pointtal der er nævnt ovenfor.

*

Programmet kan klare alle de pointsammenfald der måtte forekomme. Hver gang fordeles de pågældende range mellem de elever der har pointsammenfald, og beregningerne foregår derefter på samme måde som i eksemplet vi lige har set på.

Øvelse 4
Afprøv en situation med nogle flere pointsammenfald, og lad programmet udføre forskelstesten.

Gruppedata
Under dette menupunkt gemmer sig en mulighed for en let indtastning af data som foreligger med en stor mængde af pointsammenfald.

Lad os se på et eksempel hvor 40 elever deltager i et forsøg med en ny undervisningsmetode i matematik. 40 andre elever udgør kontrolgruppen. I en afsluttende prøve inddeles besvarelserne i 4 kategorier:

Kategori 1: Under middel
Kategori 2: Middel
Kategori 3: God
Kategori 4: Særdeles god

Vi kan fx tildele talværdierne 1, 2, 3 og 4 til disse fire kategorier. Den afsluttende prøve giver følgende resultater:

	Kat. 1	Kat. 2	Kat. 3	Kat. 4
Forsøgsgruppen	5	7	16	12
Kontrolgruppen	7	9	15	9

Det kunne se ud som om forsøgsgruppen har klaret sig bedre end kontrolgruppen. Vi kan nu anvende en rangsumtest på disse data. Talværdien 1 er tildelt 12 personer. Det betyder at rangene 1-12 må deles mellem disse 12 personer. De får derfor hver rangen 6.5, nemlig (1+12)/2 = 6.5.

På tilsvarende måde tildeles range til de øvrige personer i skemaet.

Denne tildeling af range foretager programmet for os når vi går ind under menupunktet *Gruppedata*.

Her spørges først om hvor mange grupper der er tale om. I vort eksempel er der 4 grupper, svarende til de fire kategorier ved prøven. Vi svarer derfor: 4.

Dernæst beder programmet om de antal der er placeret i hvert af skemaets felter. Når de er indtastet, ser skærmen således ud:

Antal i datasæt	A	5	7	16	12
Antal i datasæt	B	7	9	15	9

Derefter går vi til Udfør test. Her får vi nu følgende udskrift fra programmet:

Testresultat: Ingen signifikant forskel mellem A og B.

Vi kan nævne at testchancen kan beregnes til knap 17%. Det betyder at hvis det er rigtigt at de to grupper er lige dygtige, så er der ca. 17% chance for at opnå data med en skævhed som den foreliggende.

Nu vil en testchance på 17% ikke føre til forkastelse af den opstillede hypotese. Vi må derfor konstatere at der ikke i de opnåede resultater er tilstrækkeligt statistisk grundlag for at tro at forsøgsgruppen klarer sig bedre til prøven end kontrolgruppen. Med andre ord: Den nye undervisningsmetode giver ikke signifikant bedre resultater end den der er anvendt i kontrolgruppen.

I et skema som det der foreligger i denne undersøgelse, er det ligegyldigt om vi tildeler andre talværdier til de fire kategorier. I stedet for 1, 2, 3 og 4 kunne vi fx have benyttet 10, 20, 50 og 100. Det eneste afgørende er at til bedre præstationer svarer større talværdier. Talværdierne skal altså afspejle en rangorden. (Vi kan også vende sagen om og lade bedre præstationer svare til mindre talværdier. Der skal blot være sådan at bedre præstationer svarer til at talværdierne ændres i én og samme retning hver gang). I programmet antages det at data indtastes i *stigende rangorden*, altså således at det første felt svarer til de laveste range. -

Hvis du indtaster i *faldende rangorden*, kan du blot fjerne tegnet i feltet som fastlægger den stigende rangorden.

Hvis vi havde lavet en undersøgelse i en klasse der viste hvor mange drenge og piger der havde matematik, engelsk og dansk som deres bedste fag, kunne vi fx få et skema som ser sådan ud:

	Matematik	Engelsk	Dansk
Drenge	10	5	7
Piger	12	7	3

Her er imidlertid ikke nogen naturlig rangorden mellem de tre grupper af data. Vi kunne således lige så godt have placeret Dansk eller Matematik i midten af skemaet. Det har derfor ingen mening at anvende en rangsumtest på de foreliggende data.

Ved et skema med kun to grupper er der ingen forvirrende ombytningsmuligheder. Enten står den ene gruppe forrest eller også den anden, og i begge tilfælde vil en rangsumtest føre til samme resultat. I et skema af den art kan derfor godt udføre en rangsumtest, selv om der ikke umiddelbart foreligger nogen naturlig rangfølge af de to grupper af data.

Øvelse 5
I en klasse er der 10 drenge og 12 piger der har matematik som deres bedste fag, og der er 8 drenge og 3 piger der har dansk som deres bedste fag. Undersøg disse data ved en forskelstest.

3. Opgaver

I opgaverne forudsættes det at der opstilles den hypotese at de to sæt af data stammer fra samme materiale, og at den fordeling af rangene der foreligger i to datasæt kan tænkes at være foretaget ved lodtrækning. Forskelstesten vil give en statistisk vurdering af denne antagelse.

Når testen fører til et signifikant resultat, kan du konkludere hvilket datasæt „der er bedst“. Men du kan ikke sige noget om „hvor meget bedre“ det ene sæt er i forhold til det andet. Bemærk at det ikke altid er datasættet med de højeste range der er bedst. Hvis undersøgelsen fx drejer sig om bilisters reaktionstid, så er det jo bedst at have korte reaktionstider, dvs. lave range i en rangordnet opstilling.

Når testens resultat er Ingen signifikant forskel, vil du ikke have nogen grund til at forkaste den opstillede hypotese om at de to datasæt kunne være fremkommet ved tilfældig udtagelse fra samme talmateriale. Eller med andre ord: *Der er ikke påvist nogen væsentlig forskel mellem de to grupper*.

Opstil i hver af opgaverne den hypotese du vil teste, og giv din fortolkning af testens resultat.

Et forslag: Inden du benytter testen, så tag et overblik over de givne data og giv dit gæt på hvad testen vil føre frem til.

1. Konditest
Ved en konditest af to grupper af idrætsmænd opnås følgende kondital for gruppens medlemmer:

Gruppe 1: 63, 71, 74, 69, 69, 76, 71, 72, 70.

Gruppe 2: 58, 72, 61, 67, 59, 69, 71, 61.

Undersøg de forelagte data ved hjælp af en forskelstest.

2. Mål i fodboldkampe
I en spillerunde i en fodboldturnering scores der følgende antal mål i kampene i to divisioner:

1. division: 3, 2, 1, 0, 2, 3, 2, 0, 1, 4.

2. division: 2, 5, 2, 4, 3, 1, 4, 2, 0, 3.

Undersøg de forelagte data ved hjælp af en forskelstest.

3. Driftssikkerhed

Maskiner af to fabrikater undersøges for deres driftssikkerhed. Her er antallet af dage med fejlfri kørsel:

Type 1:

195, 178, 250, 238, 252, 236, 234, 200, 212, 237, 214.

Type 2:

262, 242, 253, 212, 276, 266, 158, 301, 140.

Undersøg de forelagte data ved hjælp af en forskelstest.

4. Fejltastninger

To tasteoperatørers fejlprocenter undersøges. Der foreligger følgende data i 12 prøver:

Operatør 1:

2.05, 2.18, 3.04, 2.35, 2.37, 2.23, 2.23, 2.56, 1.02, 1.57, 1.45, 1.88.

Operatør 2:

1.31, 2.39, 0.24, 1.93, 2.38, 1.22, 0.35, 1.14, 1.43, 1.93, 2.48, 1.59.

Undersøg de forelagte data ved hjælp af en forskelstest.

5. C-vitamin

I en klasse på 20 skoleelever udvælges ved lodtrækning 10 som hver får et stort tilskud af C-vitamin. I løbet af skoleåret noteres følgende antal sygedage i de to grupper:

C-vitamin: 0, 0, 2, 3, 5, 10 13, 18, 20, 22.

De øvrige: 0, 2, 3, 6, 9, 15, 17, 18, 21, 24.

Undersøg de forelagte data ved hjælp af en forskelstest.

6. En færdighedsprøve

I en færdighedsprøve opnår to elevgrupper følgende point ud af 100:

Hold 1: 81, 71, 81, 69, 78, 53, 63, 45, 83, 81.

Hold 2: 85, 84, 86, 98, 94, 68, 93, 78, 74, 73.

Undersøg de forelagte data ved hjælp af en forskelstest.

7. Reaktionstider

I et psykologisk eksperiment foreligger følgende reaktionstider for forsøgsgruppen og for en kontrolgruppe:

Forsøgsgruppen:

12, 15, 14, 15, 17, 18, 8, 16, 20, 14, 19, 18, 7, 12.

Kontrolgruppen:

15, 12, 21, 16, 11, 20, 16, 14, 13, 18, 27, 23, 12, 22.

Undersøg de forelagte data ved hjælp af en forskelstest.

8. Et eksperiment

Lad to elever udføre følgende eksperiment: Der afmærkes et punkt på gulvet. Fra 3 meters afstand kastes en mønt som skal komme så tæt på det afmærkede punkt som muligt. Lad hver elev få 5-10 kast (lige mange til hver), og mål op hvor langt mønten lander fra målet.

Undersøg ved en forskelstest om der er noget der tyder på at de to elever ikke er lige dygtige.

9. Højre eller venstre hånd

Udfør eksperimentet fra opgave 8 således at du lader samme person foretage kast med skiftevis venstre og højre hånd.

Undersøg de forelagte data med en forskelstest: Er der noget der tyder på at personen rammer bedre med den ene hånd end med den anden?

10. Kopimaskiners sammenbrud

Her er antallet af timer mellem tilkaldelse af service til to fabrikater af kopimaskiner:

Maskine 1: 408, 127, 215, 555, 510.

Maskine 2: 88, 93, 214, 102, 365, 515.

Undersøg de forelagte data ved en forskelstest.

11. Hvem reagerer hurtigst?

I en måling af reaktionstider deltager to grupper af idrætsudøvere: en gruppe af sprintere og en gruppe af bordtennisspillere.

Sprintere:

0.215, 0.234, 0.239, 0.243, 0.248, 0,248, 0.249, 0.252, 0.259, 0.268, 0.273, 0.296.

Bordtennisspillere:

0.202, 0.211, 0.216, 0.229, 0.230, 0.241, 0.242, 0.246, 0.249, 0.256, 0.276, 0.285.

Undersøg de forelagte data ved hjælp af en forskelstest.

12. Antal kilometer pr. liter

For biler af to fabrikater måles benzinforbruget ved kørsel på en motorbane.

Fabrikat P:

16.10, 16.12, 16.20, 16.22, 16.30, 16.32, 16.50, 16.53, 16.67, 16.70, 16.75, 17.10, 17.15, 17.32, 17.40, 17.90.

Fabrikat R:

15.99, 16.02, 16.03, 16.10, 16.12, 16.14, 16.23, 16.25, 16.51, 16.67, 16.78, 16.92, 17.14, 17.16, 17.25, 17.70.

Undersøg de forelagte data ved en forskelstest.

13. Reaktionstid

I INFA-programmet BILER er indlagt et program til måling af reaktionstider. Lad to personer afprøve programmet og afgør ved hjælp af en test om der er en signifikant forskel på de to personers reaktionstider.

14. En prøve

Ved en prøve foreligger følgende resultater:

	Ikke bestået	Bestået
Gruppe 1	5	12
Gruppe 2	9	4

Undersøg de forelagte data ved en forskelstest.

15. En klar forskel?

I en klasse udtrækkes ved lodtrækning 3 piger og 3 drenge. De deltager i samme opgaveløsning. Her er resultaterne:

	Løser ikke opgaven	Løser opgaven
Drenge	3	0
Piger	0	3

Undersøg disse data ved en forskelstest.

16. En klarere forskel?

I en klasse udtrækkes ved lodtrækning 4 piger og 4 drenge. De deltager i samme opgaveløsning: Her er resultaterne:

	Løser ikke opgaven	Løser opgaven
Drenge	4	0
Piger	0	4

Undersøg disse data ved en forskelstest.

17. Prøv selv

Opstil nogle skemaer som dem i opgave 15 og 16, og undersøg hvornår der foreligger en forskel som kan påvises gennem en forskelstest.

18. Tre grupper af data

I en prøve deltager hold 1 og 2. Deres præstationer bliver opdelt i tre kategorier: Lav, Middel, Høj:

	Lav	Middel	Høj
Hold 1:	10	12	4
Hold 2:	6	10	12

Undersøg de forelagte data ved hjælp af en forskelstest.

Sammenhængstest:

Er der en sammenhæng mellem to datasæt?

1. Et eksempel

Ni elever deltager i to prøver, en matematikprøve og en danskprøve. Her er deres point i de to prøver:

Elev	1	2	3	4	5	6	7	8	9
Matematik	80	53	58	64	44	84	76	82	83
Dansk	48	38	24	36	28	50	33	40	37

Vi er interesseret i at undersøge om der er en positiv sammenhæng mellem resultaterne i de to prøver: Når man scorer højt i den ene prøve, scorer man så også højt i den anden? Eller sagt på en anden måde: Når man er god til matematik, er man så også god til dansk?

Der kunne også forekomme en anden form for sammenhæng mellem to prøver: Når man scorer højt i den en prøve, scorer man så lavt i den anden prøve?

For at belyse spørgsmålet om sammenhæng mellem de to prøver ser vi på de enkelte præstationers *rangorden*. Vi erstatter de opnåede pointtal med deres nummer i rangfølgen i de to prøver.

I matematikprøven er 44 det laveste pointtal. Det er elev nr.5 der har dette pointtal. Denne elev får derfor rangen 1 i matematikprøven. Det næstbedste pointtal er 53, det svarer til rangen 2.

Her er prøveresultaterne omsat til range:

Elev	1	2	3	4	5	6	7	8	9
Matematik	6	2	3	4	1	9	5	7	8
Dansk	8	6	1	4	2	9	3	7	5

Vi kan nu lettere overskue om der er tale om en positiv sammenhæng: Det ser ud til at en høj rang i den ene prøve passer sammen med en høj rang i den anden. Med andre ord: De der har klaret sig godt i matematikprøven, har også klaret sig godt i danskprøven.

Men vi kan af rangtabellen også se at der er udsving: Elev nr. 2 har således klaret sig godt i danskprøven, men ikke så godt i matematikprøven.

Hvis der var total positiv sammenhæng, så ville enhver af de ni elever have opnået den samme rang i de to prøver: Den der var bedst i matematikprøven, ville også være bedst i danskprøven, den der var næstbedst i matematikprøven, ville også være næstbedst i danskprøven, osv.

Sådan er det ikke i den foreliggende tabel. Vi ser derfor nærmere på tallene. For hver elev ser vi på rangene i de to prøver.

Når vi skal vurdere hvor stor sammenhæng der er mellem resultaterne i de to prøver, så kigger vi på forskellen mellem rangen i matematikprøven og rangen i danskprøven. For elev nr. 1 tager vi altså forskellen mellem 6 og 8, ved elev nr. 2 tager vi forskellen mellem 2 og 6, osv.

Disse forskelle kvadreres og adderes:

$$(6-8)^2 + (2-6)^2 + \ldots + (8-5)^2 = 38$$

Vi kan se at hvis der havde være en total positiv sammenhæng mellem resultaterne i de to prøver, så ville alle tal i parenteserne være 0, og så ville den samlede sum også være 0.

Vi bruger denne sum af de kvadrerede tal som et udtryk for hvor stærk en positiv sammenhæng der er mellem resultaterne i de to prøver. Jo stærkere sammenhængen er, jo mindre er den beregnede sum.

2. Vi anvender programmet TEST

Vi anvender nu en statistisk test på de to prøver. Den foreligger som et edb-program, *Sammenhængstest*, så vi er fri for at foretage beregninger ved håndkraft. Sammenhængstest finder du i INFA-programmet TEST.

Vi indtaster først til programmet at der er tale om to datasæt der hver indeholder ni observationer (der skal altid være samme antal observationer i de to sæt). Vi indtaster herefter de ni pointtal i den rækkefølge de har i tabellen.

Programmet tildeler selv range til pointtallene og giver derefter følgende udskrift:

Testresultat: Der er en signifikant positiv sammenhæng

Og der udskrives endvidere:

Testresultatet er signifikant på 5%-niveauet:

10%	5%	x	2.5%	1%	0.5%	0.1%	0%

Det betyder at sammenhængen mellem de to datasæt er så stærk at der er under 5% chance for at den ville forekomme hvis rangene i de to prøver var blevet fordelt tilfældigt.

Et andet eksempel

Ti elever deltager i to prøver. Her er deres rangorden i de to prøver A og B:

Elev	1	2	3	4	5	6	7	8	9	10
Prøve A	7	4	3	10	6	2	9	8	1	5
Prøve B	5	7	10	3	1	9	6	2	8	4

Vi kan næsten af tabellen se at når man klarer sig godt i den ene prøve, så klarere man sig mindre godt i den anden. Men det passer dog ikke på alle elever.

Vi udfører sammenhængstesten på tallene. Det giver følgende udskrift:

Testresultat: Der er en signifikant negativ sammenhæng

Og der udskrives endvidere:

Testresultatet er signifikant på 2.5%-niveauet:

10%	5%	2.5%	x	1%	0.5%	0.1%	0%

Der er altså under 2.5% chance for at den observerede negative sammenhæng forekommer ved en tilfældig fordeling af rangene i de to prøver.

Ved en sammenhængstest er der tre mulige resultater:

Ingen signifikant sammenhæng

En signifikant positiv sammenhæng

En signifikant negativ sammenhæng

Øvelse 1
Prøv at opstille to datasæt for hvilke sammenhængstesten giver følgende resultat: Ingen signifikant sammenhæng.

Sammenfald af talværdier
Ved sammenfald af talværdier fordeler programmet rangene på samme måde som i rangsumtesten. Du kan derfor indtaste data til programmet uden at tage hensyn til om samme pointtal forekommer flere gange.

Signifikant sammenhæng?
Lad os se på det eksempel at der kun indgår tre tal i hvert datasæt. Hvis der er tale om total positiv sammenhæng, så vil fordelingen af range være sådan:

Prøve A	1	2	3
Prøve B	1	2	3

Hvis vi udregner summen af de kvadrerede forskelle mellem rangene får vi 0:

$$(1\text{-}1)^2 + (2\text{-}2)^2 + (3\text{-}3)^2 = 0$$

Lad os nu se på hvordan rangene kunne fordeles hvis der var tale om en tilfældig fordeling.

Vi fastholder rangene i prøve A, og opstiller de mulige fordelinger af rangene i B. Der er seks muligheder:

Prøve A	1	2	3
Prøve B	1	2	3
	1	3	2
	2	1	3
	2	3	1
	3	1	2
	3	2	1

Disse seks fordelinger af rangene i prøve B er alle lige sandsynlige når vi fordeler rangene på tilfældig vis. De har altså hver en sandsynlighed på 1/6, dvs. ca. 17%.

Når vi derfor har et datasæt for prøve B som giver helt samme rangfordeling som i prøve A, så er der ingen grund til at vi undrer os. Der er jo ca. 17% chance for at en sådan fordeling kan opstå ved en tilfældighed.

Den observerede totale positive sammenhæng mellem prøve A og prøve B er derfor *ikke signifikant*. Hvis den skulle have været signifikant i en sammenhængstest, måtte den jo ikke have en chance på over 10% for at forekomme ved en tilfældig fordeling.

Øvelse 2
Undersøg to prøver som hver indeholder 4 data. Er en total positiv sammenhæng mellem de to prøver signifikant ved en sammenhængstest?

3. Sammenhæng og forklaring

Du skal være opmærksom på at en positiv sammenhæng mellem to datasæt ikke betyder at gode resultater i det ene datasæt *er årsag til* gode resultater i det andet. Sammenhængstesten kan kun fortælle om der er en signifikant positiv sammenhæng mellem talværdierne i de to sæt. *Den kan ikke give nogen forklaring på hvorfor denne sammenhæng er til stede*.

Således kan en positiv sammenhæng mellem en prøve i matematik og en prøve i dansk ikke udlægges på den måde at gode kundskaber i matematik er *årsagen* til gode kundskaber i dansk. Der er en positiv sammenhæng mellem præstationerne i de to prøver, men der kan jo være en bagved liggende forklaring på at præstationerne følges pænt ad.

En undersøgelse viste at der for elever i 8. klasse var en negativ sammenhæng mellem det antal timer de benyttede til hjemmearbejde i matematik og de resultater de opnåede i en skriftlig prøve i faget. Det betyder ikke at et beskedent hjemmearbejde er *årsag* til gode matematikresultater.

En statistik over salget af sodavand på sommersøndage viste at der var en signifikant positiv sammenhæng mellem salget af sodavand og antallet af trafikuheld. Men dette kan ikke udlægges sådan at det er de mange sodavand der er årsag til trafikuheldene. Kan du selv give en mulig forklaring på den positive sammenhæng?

En sammenhængstest kan kun give oplysninger om eventuelle signifikante sammenhænge mellem de foreliggende datasæt, *den kan ikke levere forklaringen på sammenhængen*. Her må du selv komme med forslag.

4. Opgaver

I opgaverne bør du først tage et overblik over de givne data og derefter give et gæt på om du tror der foreligger en signifikant sammenhæng mellem resultaterne. Derefter kan du belyse situationen nærmere ved at anvende en sammenhængstest.

1. Dansk stil
To censorer i dansk bedømmer begge 10 stile. De stiller dem op i rangorden:

1. censor: Stil nr. 3, 9, 5, 4, 7, 6, 2, 8, 1, 10

2. censor: Stil nr. 7, 4, 9, 2, 1, 3, 5, 6, 8, 10

Undersøg ved en sammenhængstest hvor enige de to bedømmere er.

2. Eksperter og skoleelever
Et udvalg af sprogkyndige eksperter vurderer 10 forelagte bøgers sproglige sværhedsgrad. En gruppe skoleelever vurderer de samme bøger ved at tage stikprøver fra teksten og tælle op hvor mange procent af ordene i teksten der har over 6 bogstaver.

Her er de to gruppers vurdering af bøgernes sproglige sværhedsgrad. De opstiller bøgerne i følgende rangorden:

Eksperter : Bog nr. 10, 1, 6, 4, 3, 2, 7, 8, 5, 9

Skoleelever : Bog nr. 10, 6, 1, 4, 3, 2, 5, 7, 9, 8

Undersøg de to gruppers vurderinger ved hjælp af en sammenhængstest.

3. To prøver
12 elever opnår følgende pointtal ved to prøver:

Prøve A : 42, 58, 47, 32, 75, 72, 70, 43, 45, 86, 67, 78

Prøve B : 22, 26, 19, 17, 45, 68, 66, 36, 42, 72, 34, 60

Undersøg om der er nogen signifikant sammenhæng mellem resultaterne i de to prøver.

4. To vurderinger
To lærere vurderer ti elevers færdigheder i engelsk. De opstiller eleverne i følgende rangorden:

1. lærer : Elev nr. 3, 6, 5, 1, 4, 10, 8, 7, 9, 2

2. lærer : Elev nr. 1, 5, 6, 2, 4, 8, 10, 9, 7, 3

Undersøg lærernes enighed ved hjælp af en sammenhængstest.

5. Hvem vurderer bedst?
Fem personer A, B, C, D og E skal vurdere 10 metallodder og placere dem i rangorden efter vægt. Her er den rækkefølge der blev valgt af de fem (lod nr. 1 er det letteste, nr. 2 det næstletteste, osv.):

A: 1, 6, 2, 7, 3, 8, 4, 9, 5, 10
B: 2, 1, 6, 4, 3, 7, 5, 10, 8, 9
C: 6, 1, 2, 3, 4, 5, 10, 7, 8, 9
D: 3, 2, 1, 6, 4, 5, 7, 9, 8, 10
E: 4, 2, 1, 3, 7, 6, 10, 5, 9, 8

Undersøg ved hjælp af en sammenhængstest hvor gode de fem personer er til at vurdere loddernes korrekte vægtrangfølge.

6. Længdespring
En længdespringer deltager i 8 konkurrencer i løbet af sæsonen. Her er hendes resultater i de 8 konkurrencer:

5.68, 5.99, 5.60, 5.88, 5.72, 6.08, 6.02, 5.96

Er det rigtigt at hendes præstationer bliver bedre i løbet af sæsonen? Belys problemet ved en sammenhængstest.

7. Din forklaring
Giv dine kommentarer til følgende signifikante sammenhænge:

1. En undersøgelse i en skoles 1.- 9. klasse viser at der er en signifikant positiv sammenhæng mellem elevernes legemsvægt og deres regnefærdighed.

2. En undersøgelse i en række kommuner viser at der er en signifikant positiv sammenhæng mellem antallet af lærere i kommunen og antallet af cykeltyverier.

3. En undersøgelse i en række lande viser at der er en signifikant positiv sammenhæng mellem antallet af tv-apparater pr. 1000 indbyggere og salget af hovedpinepiller.

8. Eksamensresultater i matematik
14 skoleelever opstilles i rangorden efter de resultater de opnåede ved folkeskolens afgangsprøve (FA) og senere ved studentereksamen (SE):

FA:	13,	8,	12,	10,	14,	11,	9,	6,	3,	4,	5,	1,	2,	7
SE:	13,	2,	11,	14,	12,	7,	6,	1,	8,	4,	5,	10,	3,	9

Undersøg om der er nogen signifikant sammenhæng mellem resultaterne ved de to eksamener.

9. Regning og matematik
En klasses 24 elever deltager i to prøver, en regneprøve og en matematikprøve. Her er deres resultater:

Elev	Matematik	Regning	Elev	Matematik	Regning
1	60	45	13	37	34
2	52	48	14	38	30
3	40	39	15	48	43
4	27	27	16	44	40
5	20	41	17	58	31
6	43	32	18	64	58
7	50	33	19	39	22
8	56	39	20	54	38
9	25	26	21	34	37
10	49	36	22	51	24
11	71	62	23	40	45
12	40	35	24	35	46

Undersøg ved en test om der foreligger nogen signifikant sammenhæng mellem resultaterne i de to prøver.

10. Konditest

12 elever får målt deres kondital samt deres hvilepuls. Her er rangfølgen for eleverne ved de to målinger:

Konditest: 7, 2, 1, 5, 4, 12, 8, 10, 9, 6, 3, 11

Hvilepuls: 8, 7, 12, 4, 11, 5, 9, 6, 3, 1, 10, 2

Undersøg om der er en signifikant sammenhæng mellem de to datasæt.

11. Bedre med tiden?

En studerende får et job hvor han skal klargøre computere til firmaets kunder. Han klarer følgende antal i de første 15 uger:

20, 22, 26, 23, 25, 31, 30, 27, 29, 40, 43, 50, 52, 46, 56

Er der grundlag for at hævde at hans præstationer bliver bedre med tiden?

12. Hjertesygdomme hos kvinder

Statistisk Tiårsoversigt giver følgende tal for dødsfald forårsaget af hjertesygdomme:

1985: 8753	1986: 8762	1987: 8720	1988: 8785
1989: 8558	1990: 8938	1991: 8349	1992: 8259
1993: 8547	1994: 7895		

Undersøg om der er en signifikant nedadgående tendens i disse tal.

13. Ulykkestilfælde

Fra Danmarks Statistik foreligger følgende tal for mænd der omkommer ved ulykkestilfælde:

1985: 1457	1986: 1396	1987: 1367	1988: 1498
1989: 1477	1990: 1376	1991: 1267	1992: 1229
1993: 1245	1994: 1273		

Undersøg om der er en signifikant nedadgående tendens i disse tal.

14. Mon der er en sammenhæng?

Statistisk Tiårsoversigt giver følgende tal for anmeldte voldsforbrydelser og for antallet af besøg på kunstmuseer.

	Voldsforbrydelser	Besøg på kunstmuseer (i tusind)
1987	9005	1978
1988	9463	2102
1989	10291	2305
1990	10651	2524
1991	11119	2313
1992	12258	2432
1993	13487	2671
1994	14208	2619
1995	13357	2472

Undersøg om der er en signifikant sammenhæng mellem de to datasæt.

15. Biler og færdselsuheld

En statistik giver følgende oplysninger om antallet af indregistrerede biler i Danmark og antallet af færdselsuheld med personskade:

	Indreg. biler (i tusind)	Færdselsuheld m. personskade
1986	1841	11170
1987	1882	10164
1988	1897	9978
1989	1901	9922
1990	1892	9155
1991	1903	8757
1992	1921	8965
1993	1942	8513
1994	1946	8279
1995	2020	8373

Undersøg om der er en signifikant sammenhæng mellem de to datasæt.

16. Vejrdata

Her er en oversigt over tre vejrvariable for sommermånederne i Danmark i perioden 1983-97. Tallene er gennemsnit for danske målestationer.

År	Middeltemperatur	Nedbør (mm.)	Antal soltimer
1983	16.0	55	799
1984	14.9	160	612
1985	14.7	227	638
1986	14.7	152	679
1987	13.4	251	476
1988	15.7	233	584
1989	15.6	150	767
1990	15.7	196	687
1991	15.3	180	664
1992	16.9	161	803
1993	14.0	215	674
1994	16.5	194	822
1995	16.5	110	850
1996	15.3	120	731
1997	17.5	164	873

Undersøg om der findes signifikante sammenhænge mellem nogle af de anførte datasæt.