

9 1970/110 b

Per Ottestad

STATISTIKK

(Del II)

Utgave 1970



**Norges landbrukshøgskoles
bibliotek**

q1970/110b

Per Ottestad

STATISTIKK

(Del II)

Utgave 1970



0000-10-0000

0000-10-0000

0000-10-0000

0000-10-0000

Innholdsfortegnelse

	side
F. Om estimering	112
F.1. Punktestimering	112
F.2. Konfidensgrensene	121
F.3. Konfidensgrensen for en sannsynlighet	128
G. Testing av hypoteser	130
G.1. Om hypoteser og testing av dem	130
G.2. Kji-kvadrat testen	137
G.3. F-test og variansanalyse for en-veis gruppering ..	146
H. Metoder i eksperimentalforskningen	151
H.1. Om formålet med et forsøk	151
H.2. Fri randomisering	161
H.3. Blokkplanen	168
H.4. Rangering og gruppering av forsøksledd	175
H.5. Forsøk med flere faktorer	181
H.6. Split-plot i blokkforsøk	191
H.7. Andre planer for lokale forsøk	194
H.8. Utvidede forsøk	195

F. Om estimering.

F.1. Punkttestimering.

En av de aller viktigste oppgaver en står overfor i forskningsarbeid, går ut på å estimere ukjente størrelser eller parametere i et univers, slik som forklart i avsnitt B.8. For å nevne et eksempel, kan oppgaven være å estimere den meravling en venter å oppnå ved å øke mengden av kvelstoffgjødsel til en kløversort med et visst antall kilo pr. dekar. En liknende oppgave går ut på å estimere forskjellen i avlingsmengde mellom to sorter poteter.

En tilsynelatende enkel oppgave går ut på å estimere avstanden mellom to punkter i terrenget. De vansker en står overfor i et slikt tilfelle, kommer av at det ikke er mulig å utføre målingene uten å gjøre feil. I Tab. F.1 er gjengitt $n = 10$ observasjoner av en slik avstand, tatt med stålbånd. Vi ser at det er variasjon i observasjonene, og denne variasjon kan jo ikke skyldes annet enn observasjonsfeil.

Tabell F.1.

901,375	901,465	901,500	901,460	901,395
901,405	901,430	901,460	901,480	901,435

Det er nødvendig å skjelne mellom to slags observasjonsfeil, systematiske feil og tilfeldige eller random feil. Betegner vi observasjonene med x_i ($i=1,2,\dots,n$), parameteren med θ , den systematiske feil med c og den random feil med e_i , har vi at

$$x_i = \theta + c + e_i$$

Den totale feil er da

$$x_i - \theta = c + e_i$$

Den systematiske feil (c) skyldes årsaker som virker på observasjonene i samme retning og med samme styrke. Virkningen er der-

for at observasjonene blir enten systematisk for store eller systematisk for små ettersom c har positivt eller negativt fortegn. Denne feilen skyldes meget ofte feil ved de hjelpemidler en bruker. Et målebånd av stål kan være justert for f.eks. 10°C . Hvis en da bruker det ved 20°C , vil en få observasjoner som er systematisk for små.

Det er alminnelig antatt at dersom det finnes en systematisk feil, kan den bestemmes på en eller annen måte og elimineres. Under forutsetning av at dette lar seg gjøre, kan vi i vår modell sette $c = 0$, og vi har da at

$$x_i = \theta + e_i$$

De random feil kan vi imidlertid ikke kvitte oss med. Disse skyldes nemlig årsaker som en ikke kan skaffe seg full kontroll over. Dette betyr naturligvis ikke at vi ikke har noen kontroll med dem. Ved å arbeide presist og bruke gode hjelpemidler, kan en skaffe seg bedre observasjoner enn dem en får ved unøyaktig arbeidsteknikk og mindreverdige hjelpemidler.

Den random feil varierer mellom gjentakene, og vi må derfor oppfatte e som en random variabel. I teorien antar en at den skyldes et stort antall årsaker og en kan derfor oppfatte den som en sum av et stort antall småfeil. En antar også at e har tilnærmet normal fordelingsfunksjon.

Det kan vises at den random feil (e) har forventningen null, altså at $E(e) = 0$. Hvis nemlig $E(e) \neq 0$, inneholder e_i et felles element som opptrer som en systematisk feil og hører inn under leddet c i modellen.

Siden $E(e) = 0$, er også (se avsnitt E.2) $E(\bar{e}) = 0$. Og siden nå $\bar{x} = \theta + \bar{e}$, har vi i følge D.4 at $E(\bar{x}) = \theta$. Vi sier da at \bar{x} er en forventningsrett estimator av parameteren θ . Det aktuelle gjen-

nomsnitt er estimatet av θ . For vårt eksempel i Tab.F.1 finner vi at $\bar{x} = 901,441$ som er estimatet av avstanden mellom de to punktene. Vi forutsetter da at observasjonene ikke er beheftet med systematisk feil.

Å forutsette at det ikke finnes systematisk feil, eller at den er helt eliminert, kan nok mange ganger være en dristig forutsetning. Skaffes observasjonene til veie ved forsøk, kan en unngå systematiske feil ved å benytte seg av randomiseringsteknikk slik som antydnet i avsnitt A.2. I andre tilfelle er vanskene i hvert fall meget større. Det ville imidlertid føre for langt her om vi skulle gå nærmere inn på hvordan en skal gå fram for å unngå systematiske feil eller i det minste gjøre slike feil så små at de ikke spiller noen rolle. Vanskene som oppstår under observasjonsarbeidet er nemlig så forskjellige fra det ene tilfelle til det andre at det som er felles kan være vanskelig å oppdage. I alminnelighet kan en si at det gjelder at de tekniske hjelpemidlene brukes riktig. En må også som oftest sørge for at den teknikk en bruker, er blitt testet. Med dette mener vi at det er utført en undersøkelse som går ut på å observere parametere med kjente verdier. Gjør en det, kan en kanskje få påvist at teknikken fører til systematiske feil. Observasjonene kan så nyttes til å bestemme eller estimere denne feilen. I nye tilfelle kan en så bruke dette estimatet til å korrigere observasjonene med.

Det er sikkert fornuftig å regne med at det finnes systematiske feil i observasjoner som ikke er skaffet til veie ved forsøk. Når vi derfor i det følgende går ut fra at det ikke finnes systematisk feil i de observasjonene vi bygger på, må dette oppfattes som uttrykk for at den systematiske feil er så liten at vi kan se bort fra den. Det har jo ingen hensikt å bruke gjennomsnittet som

estimator av θ dersom \bar{x} er en forventningsrett estimator av $(\theta+c)$ vi og/ikke har noe middel til å estimere c .

Gjennomsnittet (\bar{x}) er naturligvis å oppfatte som en funksjon av observasjonene. En estimator er alltid det. Men før vi godtar en funksjon av observasjonene som estimator av en parameter θ , må vi ha brakt på det rene om den er en brukbar estimator. Med dette mener vi at den må ha visse egenskaper som f.eks. at den er forventningsrett. Gjennomsnittet er alltid en forventningsrett estimator av forventningen $E(\bar{x}) = \mu$, hvor μ er forventningen for den observerte random variable. På samme måte har variansen $V = s^2$, som jo også er en funksjon av observasjonene, den estimeringsegenskap at den er en forventningsrett estimator av σ^2 *).

La oss nå i sin alminnelighet tenke oss at vi ønsker å estimere en parameter θ . Vi må da skaffe oss ett eller flere sett observasjoner som direkte (som i vårt eksempel) eller indirekte kan danne grunnlaget for estimeringen. Estimeringen av θ utføres så ved hjelp av en funksjon av disse observasjonene, forutsatt at $\hat{\theta}$ har de egenskaper vi må kreve av en estimator. Av en god estimator krever en vanligvis at den skal være forventningsrett, altså at $E(\hat{\theta}) = \theta$. Det er imidlertid også andre egenskaper en estimator bør ha for å kunne regnes som en god estimator. Men disse andre egenskaper kan vi ikke komme inn på her.

*) For mange vil det kanskje virke noe forvirrende at s ikke er en forventningsrett estimator av σ . Det kan nemlig vises at $E(s) < \sigma$. Dette skyldes den matematiske operasjon som går ut på at en tar kvadratroten av $V = s^2$.

Som all annen statistisk metodikk forutsetter også estimering at det samplet av gjentak vi har hentet våre observasjoner fra, er et random sampel. Vi oppfatter da samplet som en random representant for et univers som i de aller fleste tilfelle er en abstraksjon. Dette kan mange ganger komme til å volde oss store vansker. Gjelder det å estimere noe slikt som avstanden mellom to punkter, er saken nokså enkel. Gjentar vi målingen av avstanden på uavhengig måte og etter samme oppskrift vil gjentakene være et random sampel. Uavhengighet betyr her at en måling av avstanden ikke har noe å si for resultatet av senere målinger.

I andre tilfelle kan vi imidlertid komme opp i store vansker. Vi skal forsøke å belyse dette ved hjelp av et par eksempler.

La oss tenke oss at et gruveselskap driver to kullgruver, og at en av en eller annen grunn er interessert i å estimere forskjellen i askeinnholdet i kullet fra de to gruvene. I praksis må en da ta ut et antall prøver av kullet fra gruve T_1 og et antall prøver av kullet fra gruve T_2 . Askeinnholdet bestemmes så for hver av disse prøvene. Resultatet fra en slik undersøkelse er gjengitt i Tab. F.2. Den observerte random variable er her x = prosent aske. Gjennomsnittene er $\bar{x}_1=21,5$ og $\bar{x}_2=18,0$.

Tabell F.2

T_1	T_2
24,3	18,2
20,8	16,9
23,7	20,2
17,4	16,7
21,3	
107,5	72,0

Når problemet er slik vi har beskrevet det, har vi to konkrete universer. Det ene (U_1) består av alle prøvene fra gruve T_1 , det andre (U_2) av alle prøvene fra gruve T_2 . Antall gjentak i disse

universene vil naturligvis være meget store. Er nå de to samplene på $n_1=5$ og $n_2=4$ random sampler fra de to universene, er det en liketil sak å estimere forskjellen.

La oss si at den observerte random variable har forventningen $E(x) = \mu_1$ i U_1 og $E(x) = \mu_2$ i U_2 . Den parameteren gruveselskapet er interessert i å få estimert, er differensen $\theta = \mu_1 - \mu_2$. Forutsettes det at observasjonene er fri for systematiske feil og at samplene er random sampler, vil gjennomsnittene \bar{x}_1 og \bar{x}_2 være forventningsrette estimatorer av de to forventningene. Det kan vises, noe vi kommer tilbake til, at $\hat{\theta} = \bar{x}_1 - \bar{x}_2$ er en forventningsrett estimator av differensen θ . Estimattet blir da $21,5 - 18,0 = 3,5$.

Det meget vanskelige spørsmål som melder seg i dette og mange liknende tilfelle, er hvordan en skal gå fram for å skaffe seg random sampler. I en kullgruve finnes kullet i lag, og i lag er det igjen lagdeling og annen form for heterogenitet. Under slike forhold er det neppe mulig å skaffe seg et sampel som kan oppfattes som en random representant for hele gruve. Dette forutsetter nemlig i det minste at alle forekomster av kull er kjent. Enklere er det naturligvis, enda også det kan volde atskillig hodebry, å skaffe seg et random sampel av kull fra den forekomst som står for tur til å bli drevet ut.

Estimeringen er meget enklere når observasjonene skaffes til veie ved et forsøk og reglene for planlegging og utførelse av forsøket er respektert. Med disse reglene tenker vi da først og fremst på at uttak av forsøksenheter til forsøksleddene skal gjøres ved hjelp av loddtrekning. I avsnitt A.2 har vi gitt en foreløpig forklaring på at det er nødvendig å randomisere, og vi skal komme tilbake til saken senere. Vi skal her tenke oss at forsøket utføres på et felt, f.eks. en åker, at det er to forsøksledd (T_1 og T_2)

og at en har brukt den plan som går under navn av blokkplanen. Som forklart i A.2 må vi da først dele forsøksfeltet i et antall (n) blokker og hver blokk i to ruter eller forsøksenheter. I hver av blokkene skal så den forsøksenhet som skal brukes til T_1 , tas ut ved loddtrekning. De n blokkene må vi oppfatte som et sampel som i egenskap av et random sampel, representerer et abstrakt univers. Det vi da oppnår ved hjelp av randomiseringen, er at de n enhetene som er tatt ut til T_1 og de n enhetene som er tatt ut til T_2 , er random sampler fra det samme universet.

Reglene for utførelsen av et forsøk omfatter imidlertid mer enn dette. Å utføre forsøket betyr jo at det må utføres en rekke handlinger: jorda skal bearbeides og gjødsles, frøet skal såes ut eller plantene plantes ut, feltet skal pleies (f.eks. renses for ugras) i vekstperioden, avlingen skal høstes, tørkes og veies. I praksis vil noen av disse handlingene bli utført felles for hele feltet og noen individuelt for de enkelte forsøksenhetene. Som vi har pekt på foran, kan ikke en handling utføres helt likt i to eller flere tilfelle. Dette gjelder naturligvis også her. Det en kan sørge for, er at en handling blir gjentatt etter samme oppskrift. Noe mer kan en ikke ta sikte på. Men naturligvis er det viktig, også i et tilfelle som dette, at arbeidet utføres samvittighetsfullt og mest mulig presist.

De observasjonene som er gjengitt i Tab. F.3 stammer fra et feltforsøk etter blokkplanen. Forsøksleddene T_1 og T_2 er her to sorter førmargkål, og den observerte random variable er vekten (kg pr. rute) av tørrstoffavlingen.

Tabell F.3

Blokk nr. i	$T_1 (x_{1i})$	$T_2 (x_{2i})$	$d_i = x_{1i} - x_{2i}$
1	65	48	17
2	49	45	4
3	63	42	21
4	57	48	9
5	45	39	6
6	52	52	0
	331	274	57

Hensikten med et forsøk som dette er å skaffe observasjoner som kan brukes som grunnlag for bl.a. estimering av avlingsdifferensen mellom sortene. Vi må da tenke oss at i det universet som er representert av de n gjentakene, er forventningen $E(x_1) = \mu_1$ for T_1 og $E(x_2) = \mu_2$ for T_2 , og at den parameteren som ønskes estimert er $\theta = \mu_1 - \mu_2$. Som i foregående eksempel er estimatoren $\bar{x}_1 - \bar{x}_2$ og estimatet er $55,17 - 45,67 = 9,5$ (kg).

Til estimering av θ kan vi også bruke differensene $d_i = x_{1i} - x_{2i}$. For vårt eksempel er gjennomsnittet av disse lik $\bar{d} = 57/6 = 9,5$, dvs. lik differensen $\bar{x}_1 - \bar{x}_2$. Dette må være riktig i alle tilfelle. Vi har nemlig at

$$n \bar{d} = \sum d_i = \sum (x_{1i} - x_{2i}) = \sum x_{1i} - \sum x_{2i} = n(\bar{x}_1 - \bar{x}_2)$$

Gjennomsnittet av differensene d_i er derfor også en estimator av $\theta = \mu_1 - \mu_2$. Vi kan imidlertid ikke uten videre gå ut fra at \bar{d} eller $\bar{x}_1 - \bar{x}_2$ er en forventningsrett estimator. Også i slike tilfelle som dette kan det oppstå systematiske feil, bl.a. på grunn av nabopåvirkninger forsøksenhetene imellom. Men en vet nå så meget om slike effekter at en kan unngå dem. Skaffes observasjonene til veie ved riktig planlagte og godt gjennomførte forsøk, kan en i hvert fall regne med at den systematiske feil er så liten at den ikke har noen betydning for resultatet.

De vansker en står overfor i slike tilfelle, henger sammen med spørsmålet om hva det er for et univers en opererer med. Vi har nevnt foran at det er et abstrakt univers og er det som de n gjentakene (i vårt eksempel: blokkene) representerer i egenskap av et random sampel. Dette er imidlertid bare en rent formell beskrivelse, og i praksis kan det ha betydning å kunne si noe mer om det. Det en da først og fremst bør feste seg ved, er de ytre vilkår for forsøket. I vårt eksempel er det særlig kvaliteten av dyrkingsjorda på forsøksfeltet og det er temperaturen og nedbørmengden i vekstperioden. Dessuten er det hva slags gjødsel og hvilke mengder som er brukt, kanskje også jordbearbeidingsmåten. Disse ytre vilkårene er, kan vi si, karakteristikk på universet og setter grenser for gyldigheten av den konklusjon en kommer til. Er forsøket utført på f.eks. et felt hvor jorda har et relativt høyt innhold av leire og i et år med lavt temperaturnivå og stor nedbør, har ikke estimatet av θ gyldighet for et felt hvor jorda har stort sandinnhold og i en vekstsesong med relativt høyt temperaturnivå og lite nedbør. Grunnen til dette er at en må regne med at det finnes samspill mellom sorter, eller i sin alminnelighet: forsøksledd, og slike ytre faktorer.

Universet er imidlertid også karakterisert ved den variasjon det er mellom gjentakene. I vårt eksempel er det først og fremst variasjon i kvaliteten av jorda en vil tenke på, men det kan også være andre faktorer som spiller en rolle. Vi kan si at jo større denne variabiliteten er, jo mer generelt er universet og jo større gyldighet har resultatet. At en sjelden legger noe større vekt på denne karakteristikken av universet, kommer vel av at det er så snevert at en ikke kan godta resultatet som et tilstrekkelig empirisk grunnlag for praktiske handlingsregler. Skal en skaffe seg et slikt

grunnlag, må en utføre forsøket med gjentak av en annen karakter. Men dette skal vi komme inn på senere.

F.2. Konfidensgrensene.

Vi vil nå forutsette at vi har en forventningsrett estimator $\hat{\theta}$ av parameteren θ . Er f.eks. θ avstanden mellom to punkter i terrenget og observasjonene av den er fri for systematiske feil, er $\hat{\theta} = \bar{x}$ en forventningsrett estimator. I tilfelle hvor vi har bruk for verdien av θ til f.eks. visse beregninger, kan vi som oftest ikke gjøre noe bedre enn å bruke den verdi $\hat{\theta}$ har, dvs. bruke estimatet av θ . Estimaten er imidlertid bare en mer eller mindre god tilnæringsverdi for θ , og i praksis blir det derfor spørsmål om estimaten er godt nok for formålet. Vi har m.a.o. bruk for en størrelse som viser hvilken presisjon estimeringen av θ har.

Vi har vært inne på dette spørsmålet i avsnitt B.10 og forklarte der at det vi kan oppnå, er å beregne grensene for et intervall - konfidensintervallet - og så regne med eller påstå at verdien av θ er å finne i dette intervallet. I noen meget viktige tilfelle kan disse konfidensgrensene bestemmes ved hjelp av Students fordelingsfunksjon som er beskrevet i avsnitt E.3.

Vi vil tenke oss at θ er avstanden mellom to punkter og at vi har skaffet oss n uavhengige observasjoner av den. Vi vil forutsette at disse observasjonene er fri for systematiske feil slik at vi kan bruke gjennomsnittet \bar{x} som estimator. I formelen for t (side 109) kan vi derfor erstatte μ med θ . Forutsetter vi så at den observerte random variable $x = \theta + e$, hvor e er den random feil, har normal fordelingsfunksjon, vet vi at

$$t = \frac{\bar{x} - \theta}{s} \sqrt{n}$$

er en random variabel med kjent fordelingsfunksjon, nemlig den som er kjent som Student's. Denne funksjonen har bare en parameter, nemlig antall frihetsgrader, som i det tilfelle vi har for oss her, er lik $f = n-1$.

Er nå P sannsynligheten for $|t| \geq a$, kan vi av Tabell I (bak i boka) avlese verdien av a når f er kjent og P er valt. Sannsynligheten for $|t| \leq a$ er da naturligvis $Q = 1-P$. Dette er da også sannsynligheten for

$$|\bar{x} - \theta| \leq n.s/\sqrt{n}$$

En omskriving av denne ulikheten fører til

$$\bar{x} - a.s/\sqrt{n} \leq \theta \leq \bar{x} + a.s/\sqrt{n}$$

De to uttrykkene til venstre og til høyre gir oss konfidensgrensene for θ . Bredden av konfidensintervallet er derfor lik $2a.s/\sqrt{n}$.

For eksemplet i Tab. F.1 er $\bar{x} = 901,441$ og $s = 0,040$. Her er $n = 10$ og $f = 9$. Velger vi å bruke konfidenssannsynligheten $Q=0,95$, er $P = 0,05$, og vi ser da av Tabell I at vi skal sette $a = 2,262$.

Vi finner at $a.s/\sqrt{n} = 0,028$. Konfidensgrensene er derfor

$$\bar{x} - a.s/\sqrt{n} = 901,441 - 0,028 = 901,413$$

$$\text{og } \bar{x} + a.s/\sqrt{n} = 901,441 + 0,028 = 901,469$$

Sier vi nå at θ har en verdi mellom disse to grensene, er dette en påstand eller et utsagn om θ . Vi kan ikke være helt sikre på at dette utsagnet er riktig, fordi t kan ha en verdi større enn $a = 2,262$. Sannsynligheten for dette er den valte verdi av P , her $P = 0,05$. Den valte verdi av P blir dermed et tallmessig uttrykk for risikoen for at utsagnet om θ ikke er riktig.

Konfidenssannsynligheten $Q = 1-P$ blir derfor sannsynligheten for at utsagnet $\bar{x} - a.s/\sqrt{n} \leq \theta \leq \bar{x} + a.s/\sqrt{n}$ er korrekt. Den er sannsynligheten for utsagnet i et univers hvor hvert gjentak består av n enkeltgjentak av en observasjon av θ . Dette betyr at hvis vi

velger f.eks. $Q = 0,95$, beregner konfidensgrensene for en kjent θ i et meget stort antall tilfelle og hver gang eller for hvert gjentak påstår at θ har en verdi innen dette intervall, vil vi finne at utsagnet er korrekt i 95% av tilfellene.

Vi innser kanskje lettest at dette er riktig ved å feste oss ved at de to ulikhetene vi opererer med, $|\bar{x} - \theta| \leq a.s/\sqrt{n}$ og $\bar{x} - a.s/\sqrt{n} \leq \theta \leq \bar{x} + a.s/\sqrt{n}$, er den samme. Den siste er bare en matematisk omskriving av den første. Vi vet at sannsynligheten for den første er Q , og da må Q også være sannsynligheten for den siste. Grunnen til at en har festet betegnelsen "konfidens" til den, er at brukt i sammenheng med konfidensgrensene er den et tallmessig uttrykk for vår tiltro til at utsagnet om θ er korrekt.

Vi ser av Tabell I at hvis vi øker konfidenssannsynligheten fra f.eks. 0,95 til 0,99, dvs. at vi senker risikoen for feil fra 0,05 til 0,01, blir verdien av a økt. For $f=9$ frihetsgrader betyr dette en økning av verdien av a fra 2,262 til 3,250, og for vårt eksempel vil det bety at bredden av konfidensintervallet økes fra 0,056 til 0,085. Konsekvensen er derfor at jo mindre risiko for feil vi tar, jo bredere konfidensintervall vil vi få.

Forutsatt at våre observasjoner er fri for systematiske feil, er i vårt eksempel $\bar{x} = 901,441$ det beste estimatet vi kan skaffe oss ved hjelp av våre $n=10$ observasjoner. Det er denne verdien vi må bruke dersom θ skal nyttes til videre beregninger, f.eks. til arealberegninger. Bredden av konfidensintervallet vil da vise hvor presist estimatet er. Et bredt intervall kan vise at estimatet ikke er godt nok til det vi ønsker å bruke det til.

Som nevnt i avsnitt E.3 er utledningen av Student's fordelingsfunksjon basert på den forutsetningen at fordelingsfunksjonen for den observerte random variable er normal. Dette er en lite realis-

tisk forutsetning. Det har derfor vært et problem for undersøkelse om denne forutsetningen er nødvendig når en tenker på den praktiske bruk av t . Resultatet av mange undersøkelser viser at en kan bruke Student's fordelingsfunksjon selv om fordelingsfunksjonen for den observerte random variable avviker svært meget fra den normale. Disse undersøkelsene har vært nødvendige fordi en aldri, eller i hvert fall sjelden, vet hvilken fordelingsfunksjon den observerte random variable har.

Vi skal her nøye oss med å gjengi resultatene av en slik undersøkelse. Det ble forutsatt at de verdier den random variable kan ha i et gjentak er $x = 0, 1, 2, \dots, 9$ og at sannsynligheten var den samme for alle x , dvs. at $f(x) = 0,1$. Det kan da lett vises at forventningen er $E(x) = 4,5$. Det ble så trukket random sampler på $n=25$ gjentak og

$$|t| = \frac{|\bar{x} - 4,5|}{s} \sqrt{25}$$

ble beregnet for hvert sampel. Dette ble gjentatt 100 ganger og en fikk 100 verdier av t . Antall frihetsgrader for t er i dette tilfelle lik $f=n-1 = 24$. Ved opptelling fant en så antallet (z) av t -verdier større enn de verdier av a en har i Tabell I.

Resultatet ble følgende:

P	a	z	$z/100$
0,1	1,711	12	0,12
0,05	2,064	4	0,04
0,025	2,392	2	0,02
0,01	2,797	2	0,02

Her er naturligvis z en random variabel og avvik mellom $z/100$ og P må en derfor vente. Det kan vises at de avvik vi har her, ikke er større enn at de kan tolereres, og en kan derfor si at overensstemmelsen mellom $z/100$ og P er tilfredsstillende.

Det er utført en rekke slike undersøkelser og da med mange forskjellige former for $f(x)$ som utgangspunkt. Resultatene av disse

er at fordelingsfunksjonen for den observerte random variable ingen praktisk betydning har. Flere matematiske undersøkelser er også utført, og også resultatene av disse peker i samme retning. Student's fordelingsfunksjon er derfor blitt kalt en robust fordelingsfunksjon.

Vi kan derfor også bruke Student's fordelingsfunksjon som grunnlag for beregning av konfidensgrensene for parameteren $\theta = \mu_1 - \mu_2$ for eksemplet i Tab. F.3. Vi benytter oss da av differensene $d_i = x_{1i} - x_{2i}$. For disse finner vi at $\bar{d} = 9,5$ og $s = 8,02$ og vi har at $n=6$. Velger vi å bruke konfidenssannsynligheten $Q = 0,95$, ser vi av Tabell I at vi skal sette $a = 2,571$. Vi finner da at $a \cdot s / \sqrt{n} = 8,42$ og at konfidensgrensene er

$$\bar{d} - a \cdot s / \sqrt{n} = 9,5 - 8,42 = 1,08$$

$$\text{og } \bar{d} + a \cdot s / \sqrt{n} = 9,5 + 8,42 = 17,92$$

Bruken av Student's fordelingsfunksjon er noe mer komplisert når det gjelder beregningen av konfidensgrensene for den parameteren vi ønsker å estimere i en situasjon som den som er beskrevet i sammenheng med eksemplet i Tab. F.2. Som vi skal se senere, er dette en situasjon vi også står overfor når observasjonene er skaffet til veie ved et forsøk. Den parameteren vi ønsker å estimere er her $\theta = \mu_1 - \mu_2$, og estimatoren er $\hat{\theta} = \bar{x}_1 - \bar{x}_2$. Det har naturligvis ikke noen hensikt å beregne konfidensgrenser for en slik parameter med mindre estimatoren er forventningsrett, altså at $E(\hat{\theta}) = \theta$. Vi må m.a.o. forutsette eller sørge for at $\hat{\theta}$ er fri for systematisk feil.

Vi har i dette tilfelle to standardavvik: σ_1 for T_1 og σ_2 for T_2 . Standardavvikene for de to gjennomsnittene \bar{x}_1 og \bar{x}_2 er da (se avsnitt E.2) lik $\sigma_1 / \sqrt{n_1}$ og $\sigma_2 / \sqrt{n_2}$, hvor n_1 og n_2 er antall

gjentak for T_1 og T_2 . I avsnitt E.2 er det vist at da er standardavviket for differensen $\hat{\theta} = \bar{x}_1 - \bar{x}_2$ lik

$$\sigma_{\hat{\theta}} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Er $\sigma_1 = \sigma_2 = \sigma$, finner vi at

$$\sigma_{\hat{\theta}} = \sigma \sqrt{\frac{n_1 + n_2}{n_1 n_2}}$$

Vi har også to middelavvik, gitt ved

$$(n_1 - 1)s_1^2 = \sum (x_{1i} - \bar{x}_1)^2 \quad \text{og} \quad (n_2 - 1)s_2^2 = \sum (x_{2i} - \bar{x}_2)^2$$

Av disse danner vi så et veid gjennomsnitt, nemlig

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)} = \frac{\sum (x_{1i} - \bar{x}_1)^2 + \sum (x_{2i} - \bar{x}_2)^2}{n_1 + n_2 - 2}$$

og bruker s som estimator for σ i formelen for $\sigma_{\hat{\theta}}$.

Forutsetter vi nå at observasjonene er uavhengige, at fordelingsfunksjonen for den observerte random variable er normal, og at $\sigma_1 = \sigma_2 = \sigma$, kan vi vise at

$$t = \frac{\hat{\theta} - \theta}{s} \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}}$$

er en Student's t med $f = n_1 + n_2 - 2$ frihetsgrader. Dette gir oss så midlet til beregning av kofidensgrenser for parameteren $\theta = \mu_1 - \mu_2$. Vi finner lett at disse grensene er

$$\hat{\theta} - a.s. \sqrt{\frac{n_1 + n_2}{n_1 n_2}} \quad \text{og} \quad \hat{\theta} + a.s. \sqrt{\frac{n_1 + n_2}{n_1 n_2}}$$

hvor verdien av a finnes i Tabell I for $f = n_1 + n_2 - 2$ frihetsgrader og den konfidenssannsynlighet $Q = 1 - P$ vi ønsker å bruke.

For vårt eksempel i Tab. F.2 har vi at $\hat{\theta} = 3,5$, og vi finner at $s = 2,324$ og $\sqrt{(n_1 + n_2)/n_1 n_2} = 0,671$. Antall frihetsgrader er her

$f = 7$, og velger vi så konfidenssannsynligheten $Q = 0,95$, ser vi at $a = 2,365$. Konfidensgrensene blir derfor lik

$$3.5 - 3,69 = - 0,19 \quad \text{og} \quad 3.5 + 3,69 = 7,19$$

Vi ser her at konfidensintervallet omfatter verdien null. Dette vil da si at $\theta = 0$ er en av de verdier av θ som vi må akseptere. Vi skal vise senere at dette er i samsvar med at de observasjoner vi har skaffet oss, ikke er et tilstrekkelig grunnlag for påvisning av forskjell mellom de to gruvene når det gjelder kulletts askeinnhold.

Også i tilfelle som dette står en naturligvis overfor spørsmålet om det er forsvarlig å bruke Student's fordelingsfunksjon. En av forutsetningene er som nevnt at fordelingsfunksjonen for den observerte random variable er normal. Det viser seg at avvik fra denne fordelingsfunksjonen heller ikke her har noen nevneverdig betydning. Vi har imidlertid også forutsatt at $\sigma_1 = \sigma_2$, og dette er en forutsetning det er noe vanskeligere å komme forbi. Vi kan jo ikke forutsette at likhet mellom de to standardavvik er realisert. Det viser seg imidlertid at heller ikke denne forutsetningen må tas alvorlig hvis $n_1 = n_2$ eller det er liten forskjell mellom disse antall. En har derfor gjennom planleggingen av en undersøkelse et middel til å legge forholdene til rette for bruken av Student's t til beregning av konfidensgrensene for differensen $\theta = \mu_1 - \mu_2$.

Vi skal senere vise at Student's fordelingsfunksjon også kan brukes til beregning av konfidensgrensene for andre parametre.

F.3. Konfidensgrenser for en sannsynlighet.

La oss tenke oss at vi har skaffet oss et random sampel på n gjentak og at vi ved opptelling har funnet at z av disse har kjennetegnet E . Vi har f.eks. sådd $n = 100$ gulrotfrø og etter en tid funnet at $z = 78$ av disse har spirt. I slike tilfelle kan en være interessert i å beregne konfidensgrenser for sannsynligheten for E , altså for $p(E;U) = p$. Ved universet U forstår vi vanligvis et abstrakt univers representert av samplet i egenskap av et random sampel.

I avsnitt D.4 er det vist at den relative frekvens for E , $y = z/n$, er en forventningsrett estimator av p og at standardavviket for y er $\sigma_y = \sqrt{\frac{p(1-p)}{n}}$. For sampler som ikke er for små har en funnet at fordelingsfunksjonen for y er tilnærmet normal. Dette vil si at sannsynligheten for

$$|y - p| \leq a \sqrt{\frac{p(1-p)}{n}}$$

er $Q = 1-P$ når vi bruker slike korresponderende verdier av a og Q som gitt i Tab. D.4. I stedet for denne tabell kan vi bruke Tab. D.5 eller Tabell I. I siste fall må vi bruke de verdier av a som er oppført i nederste rekke, dvs. for $f \rightarrow \infty$.

Det kan vises at konfidensgrensene svarende til konfidenssannsynligheten Q , kan finnes ved å løse ligningen

$$(y-p)^2 = a^2 \frac{p(1-p)}{n}$$

eller omskrevet:

$$(n+a^2) p^2 - (2ny+a^2) p + ny^2 = 0$$

Denne ligningen har alltid to reelle løsninger, og disse er nedre og øvre konfidensgrense for p .

For vårt eksempel hvor $n = 100$ og $y = 0,78$ finner vi at de konfidensgrensene som svarer til konfidenssannsynligheten $Q = 0,95$,

er 0,69 og 0,85. Det er altså en sannsynlighet lik $Q = 0,95$ for at $p =$ sannsynligheten for spiring er å finne i intervallet fra 0,69 til 0,85.

Denne metoden for beregning av konfidensgrensene for $P(E;U) = p$ bygger på den forutsetningen at fordelingsfunksjonen for $y = z/n$ er tilnærmet normal. Spørsmålet er da naturligvis om denne approksimasjonen er god nok i alle tilfelle. Det er vist at det er den ikke og at en for å kunne bruke metoden må stille visse minimumskrav til samplets størrelse. Den praktiske regel en er kommet fram til, er at det må kreves at np må være lik eller større enn 5. Dette vil si at hvis p har en liten verdi, må det kreves at samplet er større enn om verdien av p er stor. I praksis er det ikke lett å bedømme hvor stort sampel en har bruk for. I regelen vil en ha tidligere erfaringer å ta utgangspunkt i.

G. Testing av hypoteser.

G.1. Om hypoteser og testing av dem.

I avsnitt A.5 har vi prøvd å forklare hva en forstår med en hypotese. Vi viser nå til dette avsnittet og nøyer oss her med å si at en hypotese er en foreløpig forklaring eller en foreløpig beskrivelse. En hypotese er derfor også noe vi tar sikte på å få satt på prøve ved å konfrontere den med observasjoner, og helst da med observasjoner som er skaffet til veie for dette spesielle formål. Metoder for slike konfrontasjoner eller prøvinger er en meget viktig del av metodikken i empirisk forskning.

Siden en i empiriske tilfelle aldri kan være helt sikker på at et standpunkt for (eller imot) et utsagn er riktig, må vi også regne med at vi kan ta feil når vi forkaster en hypotese. Dette kan være nyttig å merke seg allerede her. Det en tar sikte på er å innrette seg slik at en vet, iallfall så noenlunde, hvor stor risikoen er for å forkaste en treffende hypotese. Denne risikoen, uttrykt som en sannsynlighet, både kan og skal den bestemme som tar standpunkt.

La oss ta for oss igjen det eksemplet vi benyttet i foregående avsnitt. Vi hadde her en observasjon å bygge på, nemlig at 78 av 100 gulrotfrø hadde spirt. Noen hypotese om verdien av sannsynligheten for spiring hadde vi ikke. Men vi kom fram til at sannsynligheten måtte være å finne innen et intervall fra 0,69 til 0,85. Vi kan derfor si at vi aksepterer alle de hypoteser som kunne ha vært fremsatt om $P(E;U)$ innen dette intervallet. Alle andre hypoteser om $P(E;U)$, dvs. alle verdier mellom 0 og 0,69 og alle verdier mellom 0,85 og 1, ville vi derfor ikke kunne akseptere. Hvis det hadde vært fremsatt som hypotese at $P(E;U) = 0,6$ ville vi måtte forkaste den.

Det resultat vi kom til ved beregning av konfidensgrensene, er naturligvis betinget av valget av konfidenssannsynlighet, nemlig

$Q = 0,95$. Hvis vi senket risikoen for feil og valte $Q = 0,99$, måtte vi sette $a = 2,576$. Konfidensgrensene ville da bli $0,66$ og $0,87$, og vi måtte akseptere et noe videre intervall av mulige verdier.

Utgangspunktet for våre beregninger av konfidensgrensene var

$$|y-p| \leq a \sqrt{\frac{p(1-p)}{n}}$$

eller

$$\frac{|y-p|}{\sqrt{p(1-p)}} \sqrt{n} \leq a$$

Venstre siden i den siste ulikheten er en random variabel fordi y er det. La oss betegne den med u . Vi kan så tenke oss at vi velger hypotetiske verdier for p og beregner u for hver av disse verdier og den observasjon av y vi har (i vårt eks. $y = 0,78$). Vi vil da innse at for alle de verdier av p som hører med til konfidensintervallet (for konfidenssannsynligheten $0,95$, intervallet fra $0,69$ til $0,85$), ville vi fått en verdi av u som er mindre enn $a = 1,96$. For alle verdier av p utenfor dette intervallet ville vi fått en verdi av u som er større enn $a = 1,96$. Og hvis så disse siste verdiene oppfattes som hypoteser om $P(E;U)$, ville alle disse bli forkastet på nivået $P = 0,05$.

Dette viser at vi kan bruke u som et middel til å prøve eller teste en hypotese som går ut på en bestemt verdi av $P(E;U)$. Vi sier da at u er en testvariabel. La oss tenke oss at det er fremsatt en hypotese som går ut på at spiringssannsynligheten er lik $P(E;U) = p = 0,6$. Setter vi inn denne p -verdien sammen med $y = 0,78$, finner vi $u = 3,67$ som er større enn $a = 1,96$. Den er også større enn $a = 2,576$ som svarer til $Q = 0,99$ og $P = 0,01$. Konklusjonen må derfor bli at hypotesen $p = 0,6$ forkastes.

La oss også prøve med $p = 0,75$. Vi må da tenke oss at det er fremsatt en hypotese som går ut på at spiringssannsynligheten er lik $0,75$. Denne verdien av p og $y = 0,75$ gir $u = 0,69$, en verdi mindre enn $a = 1,96$. Dette vil da si at det ikke er grunnlag for å forkaste denne hypotesen. Dette er imidlertid ikke ensbetydende med at vi aksepterer den. Konfidensgrensene forteller oss jo at vi bør akseptere en verdi av p mellom $0,69$ og $0,85$. Blant disse finner vi også $p = 0,75$.

Vi ser av dette at det faktisk at det ikke er grunnlag nok for å forkaste den testede hypotesen ikke uten videre betyr at vi kan akseptere den. Vi kan iallfall bare akseptere den som en av et stort antall muligheter. Vilkåret for at vi kan akseptere den er derfor ikke bare at det ikke er grunnlag for å forkaste den. Det må foreligge noe mer å bygge på.

Slik **vi har** fremstilt saken er hypotesen $p = 0,75$ helt vilkårlig valgt. Det er ikke noe holdepunkt på forhånd for valget av den. La oss derfor ta for oss et nytt eksempel.

Hos bananfluen forekommer karakteren $E = \text{"sepia"}$ som er en øyefarge mutant. La oss tenke oss at det er fremsatt en hypotese som går ut på at ved en bestemt krysning er sannsynligheten for at et avkom skal få E lik $P(E;U) = 0,25$. Til grunn for en slik hypotese kan en da ha et visst erfaringsmateriale og dessuten de Mendelske arvelovene. Denne hypotesen skiller seg derfor fra den vi tok for oss i forrige eksempel ved at det kan være et godt holdepunkt for den før en eventuell prøvning og at det derfor også kan være god mening i å akseptere den dersom prøvingen faller ut til gunst for den.

For den krysningen det er tale om var en imidlertid ikke villig til å akseptere den uten prøvning. Det ble derfor utført et

krysningsforsøk som gav $n = 886$ avkom ialt og blant disse $z = 204$ med E. Vi har derfor $y = z/n = 0,23$ og med $p = 0,25$ at

$$u = \frac{|y-p|}{\sqrt{p(1-p)}} \sqrt{n} = 1,36$$

altså en verdi mindre enn $a = 1,96$. Vi kan derfor si at verdien av u er en bekreftelse på at hypotesen $P(E;U) = 0,25$ er treffende.

I noen tilfelle kan vi ved å prøve en hypotese komme til det resultat at hypotesen er treffende. Det må da som sagt foreligge noe mer enn bare selve utfallet av den statistiske testen. I slike tilfelle er det derfor to mulige og alternative utfall av testingen: vi forkaster hypotesen eller vi aksepterer den. I svært mange tilfelle, kanskje de aller fleste tilfelle, er det imidlertid bare tale om et mulig utfall, nemlig forkastelse. Dette er vanligvis tilfelle der hvor hypotesen er det vi i avsnitt A.5 har kalt en null-hypotese. Hvis vi f.eks. påstår at to kornsorter gir samme mengde kornavling, kan en slik påstand oppfattes som en hypotese, og kalles da en null-hypotese. Tar vi for oss eksemplet i Tab. F.2 på nytt, kan vi si at vi ønsker å prøve en hypotese som går ut på at $E(x)$, hvor x er prosent aske, er den samme i de to universene det er tale om. Den null-hypotesen vi derfor ville være interesserte i å sette på prøve, er en som sier at $\mu_1 = \mu_2$. I avsnitt F.2 fant vi at konfidensgrensene for denne differensen er $-0,19$ og $7,19$. Det vi kan tillate oss å akseptere i dette tilfelle er derfor en differens mellom forventningene lik et eller annet tall mellom disse to grensene og blant disse da også $\mu_1 - \mu_2 = 0$. I dette tilfelle kan vi derfor ikke forkaste null-hypotesen, men vi kan heller ikke akseptere den.

Dette er den situasjon vi som oftest står overfor når det er en null-hypotese som testes og utfallet er slik at vi ikke kan for-

kaste den. Det er sannsynligvis sjelden vi da kan akseptere nullhypotesen. Regelen er at det er et stort antall andre hypoteser i nabolaget av den vi har testet, som av like stor grunn kan aksepteres. Resultatet er naturligvis da at vi ikke kan akseptere noen av dem. Konklusjonen blir derfor i disse tilfelle at de data vi har, ikke er tilstrekkelige som grunnlag for et standpunkt. De data vi har i Tab. F.2 er således ikke tilstrekkelige som grunnlag for å ta standpunkt til spørsmålet om det er forskjell i askeinnholdet i kull fra de to kullgruvene. Det er godt mulig at en gruveingeniør kan si at han ut fra generelle synspunkter vil mene at det er en forskjell, og det er godt mulig at han har rett i det. Men de data vi har er ikke tilstrekkelige som grunnlag for en bekreftelse av denne meningen. Vi kan ikke engang uttale oss om i hvilken retning ulikheten eventuelt går.

Vi står overfor en lignende situasjon når prøvingen av nullhypotesen faller ut slik at vi forkaster den. Dette må jo da føre til at vi aksepterer et eller annet alternativ. Vi skal illustrere situasjonen ved å ta for oss eksemplet i Tab. F.3. Det var her en sammenligning mellom to sorter førmargkål det gjaldt og den observerte random variable var mengde tørrstoffavling. I avsnitt F.2 fant vi at konfidensgrensene for differensene mellom de to forventningene var 1,08 og 17,92. Nullhypotesen går i dette tilfelle ut på det samme som i foregående eksempel, nemlig at $\mu_1 = \mu_2$, eller at $\mu_1 - \mu_2 = 0$. Som vi skal vise om litt, kan vi forkaste denne nullhypotesen, og konklusjonen må da bli at differensen mellom forventningene har en verdi mellom de to konfidensgrensene. Det er klart at her er estimeringen av avlingsdifferensen lite tilfredsstillende fordi vi har et så bredt konfidensintervall. Men vi kan likevel si at vi har kommet fram til et nyttig resultat. Det er

nyttig å vite at det er en påvist ulikhet og at det er sort T_1 som -
under de vilkår forsøket er utført - gir størst tørrstoff-
avling. Det vi har villet forklare med dette eksemplet, er at det
er ikke et bestemt alternativ til null-hypotesen vi aksepterer.

La oss ta for oss den ulikhet som i avsnitt F.2 ble benyttet
som utgangspunkt for beregning av konfidensgrensene for parameteren
 θ for vårt siste eksempel. Denne ulikheten var

$$|\bar{x} - \theta| \leq a \cdot s / \sqrt{n}$$

Vi kan også skrive dette slik

$$\frac{|\bar{x} - \theta|}{s} \sqrt{n} \leq a$$

Vi ser at venstre side av denne ulikheten er en random variabel.
Den har fått betegnelsen t (sml. avsnitt E.3). Det er verdier av
 θ som er innesluttet i konfidensintervallet for θ , som tilfreds-
stiller ulikheten $|t| \leq a$. Alle andre verdier av θ vil da naturlig-
vis tilfredsstille ulikheten $|t| \geq a$. Siden konfidensgrensene for θ
for vårt eksempel i Tab. F.3 er 1,08 og 17,92, vil $\theta = 0$ være en
av de verdier av θ som finnes utenfor konfidensintervallet. Inn-
setter vi $\theta = 0$ i formelen for $|t|$ og regner ut, finner vi at
 $|t| = 2,90$, dvs. større enn den verdi av a som etter Tabell I for
 $f = 5$ frihetsgrader svarer til $P = 0,05$.

Dette leder da til at vi kan bruke verdien av a i Tabell I
som en kritisk verdi når vi ønsker å teste en hypotese om θ .

Ønsker vi f.eks. å teste null-hypotesen $\theta = 0$, beregner vi

$$|t| = \frac{\bar{x}}{s} \sqrt{n}$$

og undersøker om denne verdi er større enn den a som er gitt i
Tabell I for det antall frihetsgrader (f) som gjelder for den
situasjon vi har for oss. Er verdien av $|t|$ større enn a , forkaster
vi null-hypotesen. For $\theta = 0$ finner vi for vårt eksempel $|t| = 2,90$
med $f=5$ frihetsgrader. Verdien av $|t|$ er som vi ser større enn

$a = 2,571$ for $P = 0,05$. Betrakter vi derfor risiko-nivået $P = 0,05$ som trygt nok, må konklusjonen gå ut på at null-hypotesen $\theta = 0$ forkastes.

Den verdi av P vi bruker i et aktuelt tilfelle, er her sannsynligheten for at en sann, riktig eller treffende hypotese vil bli forkastet. Den er altså et tallmessig uttrykk for en risiko for å begå denne spesielle feil. Det er meget vanlig at en setter verdien til $P = 0,05$. En bør imidlertid ikke hefte seg for meget ved en slik konvensjonell verdi. En bør heller hver gang tenke over hvilke konsekvenser det kan ha at en kommer til en konklusjon som ikke er holdbar.

Til en valt verdi av P og et gitt antall frihetsgrader (f) svarer altså en verdi av a bestemt slik at

$$2 \int_a^{\infty} f(t) dt = P$$

hvor $f(t)$ er gitt i avsnitt E.3. Brukes $|t|$ som testvariabel, kalles a gjerne en kritisk verdi av $|t|$. Intervallet $|t| \geq a$ kalles det kritiske område for $|t|$ eller også forkastningsområdet for hypotesen. For $f = 4$ og $P = 0,05$ vil en finne (se Tabell I) at $a = 2,776$. Dette er da den kritiske verdi for $|t|$, og forkastningsområdet for den testede hypotesen er området $|t| \geq 2,776$. Har en et tilfelle hvor $f = 4$ og en velger $P = 0,05$, sier en at den funne verdi av $|t|$ er signifikant på nivået $0,05$ hvis $|t| \geq 2,776$. Senker en verdien av P til f.eks. $0,01$, ser en at for $f = 4$ blir den kritiske verdien økt til $4,604$.

Det kan være av betydning at en allerede her merker seg at signifikansnivået, altså P -verdien, bør velges forut for testingen. Det nivå en velger å bruke, bør nemlig være resultat av en overveielse av hvilke konsekvenser det kan ha at en treffende hypotese

blir forkastet. Som oftest er vel disse konsekvensene ikke av noen særlig alvorlig natur, og dessuten har mangelfull planlegging meget oftere skylden for uholdbare konklusjoner.

Det er foran pekt på en forskjell mellom en vanlig hypotese og en null-hypotese. Denne forskjellen gjelder imidlertid bare tolkningen av det resultat av testingen vi kommer til. Det er ingen forskjell når det gjelder selve testingsteknikken. I det følgende skal vi derfor bruke betegnelsen H_0 på det som testes, en vanlig hypotese eller en null-hypotese.

G.2. Kji-kvadrat testen.

La oss igjen ta for oss det tilfelle at vi har observert antall gjentak (z) med et bestemt kjennetegn (E) i et sampel på n gjentak. Den hypotesen vi ønsker å teste i dette tilfelle er en verdi av sannsynligheten for E , $P(E;U) = p$. Som testvariabel benyttet vi i foregående avsnitt

$$u = \frac{|y-p|}{\sqrt{p(1-p)}} \sqrt{n}$$

hvor $y = z/n$. Vi regnet da med at fordelingsfunksjonen for y er tilnærmet normal med $E(y) = p$ og $\text{var}(y) = p(1-p)/n$.

En har imidlertid funnet at u^2 har visse fordeler framfor u som testvariabel. For u^2 finner vi lett at

$$u^2 = \frac{(y-p)^2}{p(1-p)} n = \frac{(z-np)^2}{np(1-p)}$$

Denne siste variable går under betegnelsen χ^2 (kji-kvadrat). Fordelingsfunksjonen for χ^2 kan lett utledes av $f(y)$ ved substitusjon. En må da huske på (se avsnitt D.2) at det uttrykket som skal omformes er $f(y) dy$. Da nå $dy = \frac{1}{2} \sqrt{\frac{p(1-p)}{n}} (\chi^2)^{-\frac{1}{2}} d(\chi^2)$, finner en

at

$$f(\chi^2) = \frac{1}{\sqrt{2\pi}} (\chi^2)^{-\frac{1}{2}} e^{-\frac{1}{2}\chi^2} = \frac{1}{\sqrt{2\pi}} (\chi^2)^{\frac{1}{2}(1-2)} e^{-\frac{1}{2}\chi^2}$$

En ser at dette er en Gamma fordelingsfunksjon som er beskrevet i avsnitt D.2, og det er den samme fordelingsfunksjon for χ^2 som er nevnt der når en setter antall frihetsgrader lik $f = 1$.

Den fordelingsfunksjonen for χ^2 som er utledet på denne måten, er et spesialtilfelle av den mer alminnelige fordelingsfunksjon

$$f(\chi^2) = K \cdot (\chi^2)^{\frac{1}{2}(f-2)} e^{-\frac{1}{2}\chi^2}$$

hvor f er antall frihetsgrader. Hvordan dette antall skal bestemmes i de forskjellige tilfelle, skal vi komme tilbake til.

Vi ser at f er den eneste parameter i denne funksjonen. Vi kan derfor velge signifikansnivå, dvs. verdi av P , og for $f = 1, 2, 3, 4, 5, \dots$ bestemme den kritiske verdi a av χ^2 slik at

$$\int_a^{\infty} f(\chi^2) d(\chi^2) = P$$

Disse kritiske verdier vil en finne i Tabell II for $P = 0,05$, $0,025$ og $0,01$.

Teknikken for testing av hypotesen H_0 foregår i praksis på samme måte som vi i foregående avsnitt har nyttet den testvariable u . En velger signifikansnivå (f.eks. $P=0,05$), beregner χ^2 og ser etter om denne verdien er større eller mindre enn den kritiske verdi (a). Er verdien større enn a , vil konklusjonen gå ut på forkastelse av H_0 .

Denne fordelingsfunksjonen er naturligvis den som gjelder for χ^2 under forutsetning av at H_0 er treffende. Under denne forutsetningen kan vi lett finne forventningen og variansen for χ^2 . Vi anvender de formler for forventning og varians som er gitt i avsnitt D.2, og har da at

$$E(\chi^2) = f \quad \text{og} \quad \text{var}(\chi^2) = 2f$$

Er derfor den beregnede verdi av χ^2 i et aktuelt tilfelle meget større enn antall frihetsgrader, er det et tegn på at det er noe i veien med den testede hypotesen.

Benyttes χ^2 -testen for det eksemplet fra foregående avsnitt hvor $P(E;U) = p = 0,25$, $z = 204$ og $n = 886$, finner vi at $np = 221,5$, $np(1-p) = 166,125$ og

$$\chi^2 = \frac{(204-221,5)^2}{166,125} = 1,84$$

Antall frihetsgrader er i dette tilfelle lik $f = 1$, og vi ser av Tabell II at for $P = 0,05$ er den kritiske verdi lik $a = 3,841$. Konklusjonen må derfor bli at H_0 - dvs. $P(E;U) = 0,25$ - kan ikke forkastes.

I avsnitt C.2 benyttet vi et eksempel hvor E = normale børster, $z = 2211$ og $n = 2835$. La oss så tenke oss at H_0 går ut på at $P(E;U) = p = 0,75$. Vi finner da at $np = 2126,25$, $np(1-p) = 531,5625$ og $\chi^2 = 13,512$. Antall frihetsgrader er også her $f = 1$, og for forkastningsnivået $P = 0,05$ er derfor den kritiske verdi den samme (3,841) som i foregående eksempel. Konklusjonen må derfor bli at H_0 forkastes.

La oss tenke oss at E_1, E_2, \dots, E_m er m kjennetegn som kan inntreffe og som utelukker hverandre i et gjentak. Dette betyr at et gjentak nødvendigvis må ha et av disse kjennetegn og bare ett av dem. La oss betegne sannsynlighetene for disse kjennetegn i et univers med p_1, p_2, \dots, p_m og frekvensene i et random sampel på n gjentak med z_1, z_2, \dots, z_m . Da er naturligvis

$$\sum p_i = 1 \quad \text{og} \quad \sum z_i = n$$

Er $m = 2$, har vi med disse nye betegnelsene at

$$\chi^2 = \frac{(z_1 - np_1)^2}{np_1 p_2} \quad \text{eller} \quad \chi^2 = \frac{(z_2 - np_2)^2}{np_1 p_2}$$

hvor $p_2 = 1 - p_1$.

Fp 202/105
|
MB

Den formel for χ^2 som vanligvis brukes, er imidlertid en annen. Det kan lett bevises at

$$\chi^2 = \frac{(z_1 - np_1)^2}{np_1 p_2} = \frac{(z_1 - np_1)^2}{np_1} + \frac{(z_2 - np_2)^2}{np_2} .$$

For vårt eksempel har vi når $E_1 =$ "normale børster" og $E_2 =$ "reduerte børster", at $z_1 = 2211$, $z_2 = n - z_1 = 2835 - 2211 = 624$, $p_1 = 0,75$ og $p_2 = 0,25$. Vi har da at $np_1 = 2126,25$ og $np_2 = 708,75$. Dette innsatt i den siste formel for χ^2 gir

$$\chi^2 = 3,378 + 10,134 = 13,512.$$

Den siste formelen for χ^2 er utvidet til å omfatte et hvilket som helst antall alternative kjennetegn som utelukker hverandre i et gjentak. La disse kjennetegn være E_i ($i = 1, 2, \dots, m$), frekvensene i et random sampel på n gjentak z_i ($i = 1, 2, \dots, m$) og sannsynlighetene p_i ($i = 1, 2, \dots, m$). Da er

$$\chi^2 = \sum \frac{(z_i - np_i)^2}{np_i} .$$

La oss ta for oss et nytt eksempel. En krysset røde og elfenbensfargede torskemunn og fikk $n = 97$ avkom. Blant disse fant en $z_1 = 22$ med $E_1 =$ "rød", $z_2 = 52$ med $E_2 =$ "lyserød" og $z_3 = 23$ med $E_3 =$ "elfenbensfarge". Sett nå at en har en hypotese (H_0) som går ut på at $p_1 = 0,25$, $p_2 = 0,50$ og $p_3 = 0,25$. Vi har da $np_1 = 24,25$, $np_2 = 48,50$ og $np_3 = 24,25$. Dette innsatt i formelen for χ^2 gir:

$$\chi^2 = 0,209 + 0,253 + 0,064 = 0,526.$$

Det er vist at forutsatt H_0 , dvs. forutsatt at de p -verdier vi bruker til beregning av χ^2 er sannsynlighetene for E 'ene i universet, er fordelingsfunksjonen for χ^2 tilnærmet identisk med den fordelingsfunksjonen som er gitt foran.

For å kunne bruke Tabell II må vi vite hvordan vi skal bestemme antall frihetsgrader. Vi må her nøye oss med å gi noen regler for dette uten å gi noen grunn for disse reglene.

Den første av disse reglene går ut på følgende. Hvis de verdiene av sannsynlighetene p_1, p_2, \dots, p_m vi bruker til beregning av χ^2 , er inkludert i H_0 , altså at de ikke er estimert ved hjelp av de frekvensene vi bruker til beregning av χ^2 , går regelen ut på at f er lik antall frekvenser vi bruker redusert med antall ligninger disse frekvensene må tilfredsstille. Ved en slik ligning forstår vi en ligning som uttrykker at sampelstørrelsen (n) er et gitt tall. I et tilfelle der det er m alternative kjennetegn E_1, E_2, \dots, E_m og vi bruker frekvensene z_1, z_2, \dots, z_m til beregning av χ^2 gir denne regelen $f = m - 1$. Reduksjonen med en enhet skyldes at frekvensene skal tilfredsstille ligning $\sum z_i = n$.

I et tilfelle der det er bare to alternative kjennetegn og dermed to frekvenser, er $m = 2$ og dermed $f = 1$. Dette gjelder vårt eksempel hvor $E_1 =$ "normale børster" og $E_2 =$ "reduserte børster" og vi fant at $\chi^2 = 13,512$. I vårt andre eksempel har vi $m = 3$ og derfor $f = 2$. Vi ser derfor at den funne χ^2 -verdien, $\chi^2 = 0,526$, er mindre enn den nedre grensen for det kritiske område som svarer til $P = 0,05$, nemlig $a = 5,991$. Den funne χ^2 -verdi er derfor ikke signifikant på 5% nivået og konklusjonen må bli at H_0 forkastes ikke.

Vi har forutsatt foran at verdiene av sannsynlighetene for E_i ($i = 1, 2, \dots, m$), altså p_i ($i = 1, 2, \dots, m$) er inkludert i den hypotesen vi ønsker å teste. Det er i virkeligheten verdiene av disse sannsynlighetene som utgjør hypotesen. I mange tilfelle der vi bruker χ^2 -testen, er imidlertid problemstillingen en annen. Det vi er ute etter, er å finne ut om det kan sies at sannsynlig-

heten for et kjennetegn E , altså $P(E;U) = p$, er ulik i to eller flere tilfelle. Vi kan også si det slik at vi har random sampler fra to eller flere universer, og vi er interessert i spørsmålet om sannsynligheten for et kjennetegn E - eller det kan være flere kjennetegn som utelukker hverandre - er forskjellig i disse to universene. Den null-hypotesen vi ønsker å teste, går derfor ut på at sannsynlighetene for hvert av de alternative kjennetegn det er tale om (E_1, E_2, \dots, E_m), er de samme i de universene vi sammenligner ved hjelp av de random sampler vi har. Til beregningen av χ^2 bruker vi da ikke hypotetiske verdier av p 'ene. Vi estimerer disse p 'ene ved hjelp av våre sampler.

La oss ta for oss et eksempel. De personer som utvandrer fra Norge i en periode, f.eks. ett år, må vi oppfatte som et sampel av gjentak. Universet av gjentak er da det universet samplet representerer i egenskap av et random sampel. Gjentakene i universet og i samplet kan nå grupperes på flere måter, f.eks. etter de alternative kjennetegn $E_1 = \text{"mann"}$ og $E_2 = \text{"kvinne"}$. La $P(E_1;U) = p_1$ og $P(E_2;U) = p_2$. Da er naturligvis $p_1 + p_2 = 1$.

La oss nå tenke oss at vi har to slike sampler, ett som omfatter alle utvandrede personer i årene 1934-35 og ett som omfatter alle utvandrede personer i årene 1946-47. Vi har da også to universer, og vi kan være interesserte i å undersøke om p_1 er forskjellig i disse universene.

I tabell G.1 er oppgitt antall gjentak med $E_1 = \text{mann}$ og antall gjentak med $E_2 = \text{kvinne}$ for to sampler av utvandrede personer.

Tabell G.1

Periode	E ₁	E ₂	n
1934-35	372	575	947
1946-47	1069	1381	2450
Sum	1441	1956	3397
p'	0,4242	0,5758	1

$$p'_1 = \frac{1441}{3397} = 0,4242$$
$$p'_2 = \frac{1956}{3397} = 0,5758$$

For å kunne beregne χ^2 må vi ha estimater av de to sannsynlighetene p_1 og p_2 . Forutsetter vi nå at p'ene er de samme i de to universene, er det ingen grunn til når vi skal estimere dem, å skille mellom de to samplene. Vi slår derfor samplene sammen til et sampel på $n = 3397$ og bruker de relative frekvensene for E_1 og E_2 i dette sampel som estimatorer. Estimatenes er i tabellen betegnet med p'.

Vi beregner så χ^2 ved hjelp av p'_1 og p'_2 som om disse estimatene hadde vært sannsynligheter inkludert i hypotesen. For hver av de to periodene beregner vi følgende

$$\frac{(z_1 - np'_1)^2}{np'_1} + \frac{(z_2 - np'_2)^2}{np'_2}$$

Resultatet for hver av de to periodene er

1934-35	3,819
1946-47	1,475
Sum	5,294

Det er vist at forutsatt at null-hypotesen er treffende, er summen av disse to tall med tilstrekkelig tilnærming et kji-kvadrat med $f = 1$ frihetsgrad. Vi ser av Tabell II at $\chi^2 = 5,294$ er signifikant på 5% nivået. Forkaster vi så null-hypotesen på dette grunnlag må vår konklusjon gå ut på at det i årene fra 1934-35 til 1946-47 har foregått en omstilling til utvandring og at denne omstilling har vært forskjellig for kvinner og menn. De relative

frekvensene for menn i de to periodene er 0,39 for 1934-35 og 0,44 for 1946-47. Den omstilling som har skjedd har m.a.o. gitt seg utslag i en økt sannsynlighet for menn.

La oss nå i alminnelighet tenke oss at det er m alternative kjennetegn E_i ($i = 1, 2, \dots, m$) og at vi har k sampler. Vår null-hypotese går da ut på at sannsynligheten for E_i (p_i) er den samme i de k universene som våre k sampler representerer i egenskap av random sampler. Vi estimerer så p_i ved å slå sammen de k samplene og bruker den relative frekvens for E_i (p'_i) i det sammenslåtte sampel som estimator. La j være numret på samlet ($j = 1, 2, \dots, k$). Den enkelte frekvens kan vi da betegne med z_{ij} . For hvert sampel beregnes

$$w_j = \sum_1^m \frac{(z_{ij} - n_j p'_i)^2}{n_j p'_i}$$

og deretter $\chi^2 = \sum_1^k w_j$. Dette kji-kvadrat har da $f = (k-1)(m-1)$ frihetsgrader.

I alle tilfelle der vi bruker estimator til beregning av χ^2 , må antall frihetsgrader reduseres med antallet av disse. Vi kan sette at

$$f = a - b - c$$

der a er antallet av frekvenser vi har brukt til beregning av χ^2 , b er antallet av ligninger disse frekvensene må tilfredsstille og c er antall estimator vi har brukt. I det tilfelle vi har for oss er det naturligvis km frekvenser vi bruker, altså $a = km$. Vi betrakter så størrelsen av samplene (n_j) som gitte tall. For hvert sampel har vi derfor ligningen $\sum z_{ij} = n_j$, dvs. k slike ligninger. Følgelig er $b = k$. Antallet av estimator er tilsynelatende m , p'_i ($i = 1, 2, \dots, m$). Men siden $\sum p'_i = 1$, er det i virkeligheten tilstrekkelig å beregne $(m-1)$ av disse estimatene.

Vi har derfor at $c = m-1$. Dette innsatt gir

$$f = km - k - (m-1) = (k-1)(m-1).$$

I Tabell G.2 er gitt frekvensene for $E_1 =$ udyktige, $E_2 =$ hjelpedyktige og $E_3 =$ stridsdyktige for to sampler av norske rekrutter, 1935 og 1936. Estimatene av sannsynlighetene p_i ($i = 1, 2, 3$) er gitt. Videre er oppgitt verdiene av w_j definert ovenfor. Vi har derfor at

$$\chi^2 = \sum w_j = w_1 + w_2 = 5,860 + 5,865 = 11,725$$

Siden $k = 2$ og $m = 3$, er antall frihetsgrader lik $f = (2-1)(3-1) = 2$. Vi ser av Tabell II at denne χ^2 -verdi er signifikant på 1% nivået.

Tabell G.2

År	E_1	E_2	E_3	n_j	w_j
1935	6057	2176	17988	26221	5,860
1936	6243	1994	17666	25903	5,865
Sum	12300	4170	35654	52124	11,725
p'	0,236	0,080	0,684		

Våre tall gir oss ikke noe grunnlag for å finne ut hva dette kommer av. Men det er vel rimelig å tro at det skyldes at legebedømmelsen av rekruttene ikke har vært ens i de to årene.

Kji-kvadrat testen har andre anvendelser som vi ikke har anledning til å ta med her. Vi skal nøye oss med å nevne at en også kan bruke χ^2 til en nærmere vurdering av sammenligninger mellom de relative frekvenser for en random variabel i et sampel og sannsynlighetene etter en valt fordelingsfunksjon.

Som nevnt foran er den fordelingsfunksjonen som er tabulert i Tabell II, bare en approksimasjon til den egentlige fordelingsfunksjon for χ^2 . For praktiske formål er approksimasjonen god nok forutsatt at de forventninger en bruker under beregningen av χ^2

alle er i det minste lik 5. Dette gjelder både når sannsynlighetene (p) er inkludert i den testede hypotesen og når de er estimert (p'). I et tilfelle der np eller np' er mindre enn 5, må en slå sammen kjennetegn slik at np eller np' for det sammenslåtte kjennetegn blir lik eller større enn 5. Som eksempel kan nevnes at dersom np' for E₁ eller/og for E₂ i eksemplet i Tabell G.2 ble mindre enn 5 - dette kunne f.eks. ha inntruffet hvis sampelstørrelse hadde vært meget mindre - kunne vi slå sammen E₁ og E₂ (og da også summert frekvensene for disse to kjennetegn) til et nytt kjennetegn E' = ikke-stridsdyktig. Vi ville da ha bare 4 frekvenser til beregning av χ^2 , og antall frihetsgrader ville bli f = 1.

G.3. F-test og variansanalyse for en-veis gruppering.

I avsnitt F.2 benyttet vi Student's t som hjelpemiddel til beregning av konfidensgrensene for differensen mellom forventningene for en random variabel (x) i to universer, U₁ og U₂. Vi forutsatte da at vi hadde et sampel på n₁ gjentak fra U₁, et sampel på n₂ gjentak fra U₂ og dermed n₁ og n₂ observasjoner av x. Vi satte så $\hat{\theta} = \bar{x}_1 - \bar{x}_2$, $\theta = \mu_1 - \mu_2$ og

$$t = \frac{\hat{\theta} - \theta}{s} \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}}$$

hvor

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

Under visse forutsetninger som vi skal komme tilbake til, har det lyktes å vise at denne t har den fordelingsfunksjonen som er beskrevet i avsnitt E.3. Antall frihetsgrader er da f = n₁ + n₂ - 2.

I avsnitt G.1 er det så videre pekt på hvordan t kan brukes til å teste null-hypotesen $\theta = 0$. Visetter $\theta = 0$ og får da

$$|t| = \frac{|\bar{x}_1 - \bar{x}_2|}{s} \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}}$$

For eksemplet i Tab. F.2 har vi at $n_1=5$, $n_2=4$, $\bar{x}_1=21,5$, $\bar{x}_2=18,0$ og $s = 2,324$. Ved innsetning finner vi så at

$$|t| = \frac{3,5}{2,324} \sqrt{20/9} = 2,244$$

For $f = 7$ frihetsgrader og $P = 0,05$ er etter Tabell I den kritiske verdi for $|t|$ lik $a = 2,365$. Den funne verdi av $|t|$ er altså noe mindre. Har vi derfor valt forkastningsnivået $P = 0,05$, må konklusjonen bli at null-hypotesen $\theta = 0$ forkastes ikke.

Vi skal nå vise hvordan vi kan teste en null-hypotese som går ut på at

$$\mu_1 = \mu_2 = \mu_3 = \dots = \mu_k$$

Vi kan tenke oss som eksempel at vi har skaffet oss observasjoner av $x = \%$ aske fra k gruver (T_1, T_2, \dots, T_k) og at vi er interessert i å teste en null-hypotese som går ut på at forventningen for x er den samme i de k universene.

Vi ønsker å benytte en test som gjør bruk av observasjonene fra alle gruvene under ett, en test som tester null-hypotesen i sin helhet og ikke stykkevis som f.eks. $\mu_1 = \mu_2, \mu_3 = \mu_4$ osv.

Til illustrasjon skal vi bruke eksemplet i Tab. G.3 hvor observasjonene er observasjoner av $x =$ vekten (i kg) av 56 dager gamle grisunger i $k = 5$ kull. Disse grisungene har samme far, og det er $k = 5$ mødre betegnet med T_1, T_2, \dots, T_5 . Vi ser at antallet av gjentak (grisunger) varierer noe mødrene imellom. Vi vil betegne antall gjentak, antall observasjoner, med n_j ($j = 1, 2, \dots, k = 5$). Videre er i tabellen oppført summene av observasjonene (S_j), gjennomsnittene (\bar{x}_j) og variansene ($V_j = s_j^2$) for hver av de 5 gruppene.

Tabell G.3

	T_1	T_2	T_3	T_4	T_5
	12	16	11	15	17
	18	17	8	16	17
	16	12	9	11	14
	13	10	12	14	14
			10		13
			10		
n_j	4	4	6	4	5
s_j	59	55	60	56	75
\bar{x}_j	14,75	13,75	10,00	14,00	15,00
V_j	7,58	10,92	2,00	4,67	3,50

Vi må nå forestille oss at gjentakene (grisungene) i hvert av disse kullene representerer et univers i egenskap av et random sampel.

I vårt eksempel har vi altså $k=5$ slike universer, og vi er interesserte i å undersøke om det kan sies at forventningen for den observerte random variable (i vårt eksempel $x =$ vekten) er forskjellig i noen av disse universene. Vår null-hypotese går derfor ut på at de k forventningene er like.

Den testen en bruker til testing av denne null-hypotesen, går under navn av F-testen. Det er i hovedtrekkene det samme som Students t -test, og en vil da også finne at for det tilfelle at $k = 2$ er $F = t^2$.

La oss nå si at vi har k grupper av observasjoner, gruppene betegnet med T_j ($j = 1, 2, \dots, k$). I den j 'te gruppen har vi n_j observasjoner som vi vil betegne med x_{ji} . Fotskriften j sier oss her hvilken gruppe observasjonen hører med til og fotskriften i hvilket nummer den har innen gruppen.

For hver gruppe har vi naturligvis et gjennomsnitt og en varians definert på vanlig måte. Disse vil vi betegne med \bar{x}_j og V_j (el. s_j^2). Vi har da

$$\bar{x}_j = \frac{1}{n_j} \sum x_{ji} = \frac{S_j}{n_j}$$

og

$$V_j = s_j^2 = \frac{1}{n_j - 1} \sum (x_{ji} - \bar{x}_j)^2.$$

Varianskvotienten F er lik forholdet mellom to varianser. Den ene av disse er avhengig av variasjonen i de k gruppegjennomsnittene. Denne variansen kalles derfor ofte mellom-gruppe variansen. Vi skal betegne den med V_t . Den andre av de to variansene som vi skal betegne med V_R , kalles innen-gruppe variansen. Den er avhengig av variasjonen i observasjonene innen gruppene og er et gjennomsnitt (et veid gjennomsnitt) av gruppevariansene V_j .

Vi har naturligvis også et gjennomsnitt for alle observasjonene tatt under ett. Dette vil vi betegne med \bar{x} . Er $N = \sum n_j$ antallet av observasjoner i det hele, er

$$\bar{x} = \frac{1}{N} \sum \sum x_{ji} = \frac{1}{N} \sum S_j = \frac{1}{N} \sum n_j \bar{x}_j.$$

Er antall observasjoner det samme i alle grupper, altså $n_1 = n_2 = n_3 \dots = n_k = n$, er $N = nk$ og

$$\bar{x} = \frac{1}{k} \sum \bar{x}_j$$

altså et vanlig gjennomsnitt av gruppegjennomsnittene.

Innen-gruppe variansen V_R er som nevnt, et veid gjennomsnitt av gruppevariansene V_j . Vektene er lik antall frihetsgrader for V_j , altså $(n_j - 1)$. Vi har derfor at

$$V_R = \frac{1}{N - k} \sum (n_j - 1) V_j.$$

Nevneren $N - k$ er summen av vektene: $N - k = \sum (n_j - 1)$. Siden gruppevariansen er $V_j = \frac{1}{n_j - 1} \sum (x_{ji} - \bar{x}_j)^2$, har vi også at

$$V_R = \frac{1}{N - k} \sum \sum (x_{ji} - \bar{x}_j)^2.$$

Mellom-gruppe variansen V_t er, kan vi si, en varians for gruppegjennomsnittene. Når vi ser bort fra at vi her må bruke vektor (n_j) er den dannet på samme måte som den varians vi definerte i B.5.

Vi har

$$V_T = \frac{1}{k-1} \sum n_j (\bar{x}_j - \bar{x})^2.$$

Er antallet av observasjoner det samme i alle k gruppene, dvs. at $n_j = n$, forenkles formlene til

$$V_R = \frac{1}{k} \sum V_j$$

og

$$V_T = \frac{n}{k-1} \sum (\bar{x}_j - \bar{x})^2.$$

Disse to variansene kan naturligvis beregnes direkte ved de gitte formlene. For vårt eksempel i Tab. G.3 finner vi at $V_R = 5,19$ og $V_T = 22,73$. Det er imidlertid vanlig å utføre beregningene på en annen måte, ved at en trekker inn i beregningene det en kan kalle den totale kvadratsummen. Dette er summen av kvadratene av enkeltobservasjonenes avvik fra gjennomsnittet av alle, altså enkeltobservasjonenes avvik fra \bar{x} . Denne kvadratsummen er derfor $\sum \sum (x_{ji} - \bar{x})^2$. Under denne summeringen må vi ta med alle N observasjonene.

Det kan nå lett bevises at

$$\sum \sum (x_{ji} - \bar{x})^2 = \sum n_j (\bar{x}_j - \bar{x})^2 + \sum \sum (x_{ji} - \bar{x}_j)^2.$$

Vi ser at de to leddene på høyre side av ligningen er tellerne i formlene for V_T og V_R . Vi skal kalle dem for korthets skyld kvadratsummene for V_T og V_R .

Det er denne oppdelingen av den totale kvadratsummen i to additive kvadratsummer og deretter beregning av V_T og V_R som er det vi kaller en varians-analyse. Varians-analysen har en videre målsetting enn den vi skal beskjefte oss med her. Vi tenker her bare på analysen som grunnlag for beregning av de to variansene.

Beregningsmessig bruker en oppdelingen av den totale kvadratsummen på følgende måte. En beregner den totale kvadratsummen og

kvadratsummen for V_T . Differensen mellom disse to er da kvadratsummen for V_R . I det tilfelle vi har for oss her, det vi kaller en-veis gruppering, betyr dette ikke noen stor reduksjon av regnearbeidet fordi en alltid bør beregne gruppevariansene hver for seg. Men vi skal se at i andre tilfelle betyr denne teknikken en betydelig forenkling av regnearbeidet.

I de aller fleste tilfelle er det en fordel å beregne kvadratsummen for V_T , altså $\sum n_j (\bar{x}_j - \bar{x})^2$, ved hjelp av gruppesommene S_j (jfr. Tab. G.3) i stedet for ved hjelp av gruppegjennomsnittene (\bar{x}_j). Sammenhengen mellom gruppesummer og gruppegjennomsnitt er naturligvis $S_j = n_j \bar{x}_j$. La så S være summen av alle N observasjonene, altså $S = \sum \sum x_{ji} = \sum S_j$.

En kan da vise at

$$\sum \sum (x_{ji} - \bar{x})^2 = \sum \sum x_{ji}^2 - S^2/N$$

og at

$$\sum n_j (\bar{x}_j - \bar{x})^2 = \sum S_j^2/n_j - S^2/N$$

For vårt eksempel i Tab. G.3 har vi at $N = 23$, $S = 305$ og derfor at $S^2/N = 4044,5652$. Videre finnes at $\sum \sum x_{ji}^2 = 4229$ og at $\sum S_j^2/n_j = 4135,5$. Dette innsatt gir da

$$\begin{aligned} \sum \sum (x_{ji} - \bar{x})^2 &= 4229 - 4044,5652 = 184,4348 \\ \sum n_j (\bar{x}_j - \bar{x})^2 &= \frac{4135,5 - 4044,5652}{1} = 90,9348 \\ \underline{\underline{\sum \sum (x_{ji} - \bar{x}_j)^2}} &= \underline{\underline{93,5000}} \end{aligned}$$

Vi har videre at $k-1 = 4$ og $N-k = 18$. Følgelig er

$$V_T = \frac{90,9348}{4} = 22,73$$

og

$$V_R = \frac{93,5}{18} = 5,19.$$

De to divisorene, $k-1$ og $N-k$, er antall frihetsgrader for variansene V_T og V_R . Antall frihetsgrader for V_T er da $f = k-1 = 4$ og antall

frihetsgrader for V_R er $f = N - k = 18$. Legg merke til at summen er lik $(N-1)$.

Det er vanlig å samle resultatene av beregningene slik:

	Kvadratsum	f	V
Total	184,4348	22	
Mellom grupper	90,9348	4	22,73
Innen "	93,5000	18	5,19

Til testing av null-hypotesen $\mu_j = \text{konstant}$ bruker en så

$$F = V_T / V_R$$

Vi må da kjenne fordelingsfunksjonen for F under forutsetningen $\mu_j = \text{konstant}$. Denne funksjonen er

$$f(F) = K \frac{F^{\frac{1}{2}(f_1-2)}}{(f_1 F + f_2)^{\frac{1}{2}(f_1+f_2)}}$$

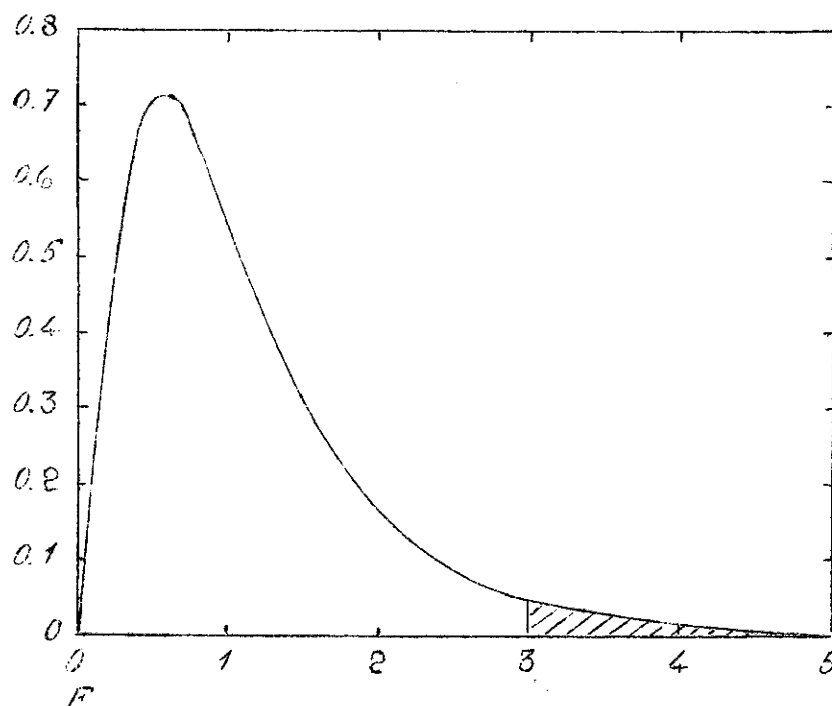
hvor K er en konstant. I det tilfelle vi har for oss her, er $f_1 = k - 1$ og $f_2 = N - k$.

For å kunne utlede denne funksjonen har en måttet forutsette at fordelingsfunksjonen for den observerte random variable (x) er normal og at standardavviket er det samme i de k universene. Det er imidlertid vist at fordelingsfunksjonen for F er lite påvirket av disse forutsetningene, slik at F -testen kan regnes å være en alminnelig brukbar test av null-hypotesen $\mu_j = \text{konstant}$.

Under forutsetning av at $\mu_j = \text{konstant}$ kan en vise at de to variansene har samme forventning, altså at $E(V_T) = E(V_R)$. For vårt eksempel har vi funnet at V_T er betydelig større enn V_R , og dette tyder da på at det er noe i veien med null-hypotesen.

Vi ser at fordelingsfunksjonen for F har bare to parametere, nemlig antall frihetsgrader (f_1 og f_2). I Fig. G.1 er inntegnet grafen for funksjonen for det tilfelle at $f_1 = 6$ og $f_2 = 12$. Forventningen for F er noe større enn abscissen til grafens maksimum.

Det skraverte felt til høyre er lik $P = 0,05$. Det avgrenses til venstre av ordinaten til $F = a = 3$. Denne verdien $a = 3$ er da den nedre grensen for det kritiske område for F som svarer til $P = 0,05$. En f -verdi lik eller større enn $a = 3$ er altså signifikant på



Figur G.1.

5% nivået. Det er verdier i dette området vi oppfatter som "påviselig" i strid med null-hypotesen. Nå vil imidlertid en meget liten F -verdi, altså når V_T er liten i forhold til V_R , også bety et stort avvik fra $E(F)$. Vi har derfor også et kritisk område for F til venstre i figuren, det er et område som inneholder F -verdier i nærheten av null. For antall frihetsgrader lik $k-1 = 6$ og $N-k = 12$, strekker dette kritiske område seg fra $F = 0$ til $F = 0,25$. Det hender i praksis at en finner slike signifikante F -verdier, men oftest er det da nokså uklart hva det kommer av, dvs. også uklart hvordan et slikt resultat skal tolkes. Vi skal derfor i dette kurset

innskrenke oss til diskusjonen av de signifikante F-verdier som vi har når V_T er større enn V_R .

I tabell III finner vi grenseverdiene for de kritiske områder for F som svarer til $P = 0,05$. Vi ser at antall frihetsgrader for den største av de to variansene er gitt i hodet av tabellen, antall frihetsgrader for den minste i venstre kolonne. Betrakter vi nå bare de tilfelle der V_T er større enn V_R , vil det si at vi finner antall frihetsgrader for V_T i hodet av tabellen, antall frihetsgrader for V_R i venstre kolonne. Den tilsvarende tabell for $P = 0,025$ og $P = 0,01$ er gjengitt i Tabell IV og Tabell V.

For vårt eksempel i Tab. G.3 finner vi at $F = 22,73/5,19 = 4,38$. Vi ser av Tabell III at det kritiske område for F som svarer til $F = 0,05$, er begrenset til venstre av $a = 2,93$. $F = 4,38$ er derfor signifikant på 5% nivået.

Siden null-hypotesen er $\mu_j = \text{konstant}$, må forkastelse av null-hypotesen bety at det er påviselig variasjon i forventningen for den observerte random variable (x) mellom de k universene. En signifikant $F = V_T/V_R$ kommer av at gruppegjennomsnittene varierer mer gruppene imellom enn vi med rimelighet kan godta dersom forventningene er like.

For vårt eksempel i Tab. G.3 må den signifikante F-verdi vi har fått, bety at forventningen for vekten av grisunger varierer påviselig mødrene imellom. Naturligvis kan dette tilfelle være ett av de tilfelle der vi gjør feil ved å forkaste null-hypotesen. Men som forklart foran, er vi nødt til å ta sjansen på feil forkastelse for i det hele tatt å kunne oppdage noe. Dette er ikke noe som er karakteristisk for statistiske tester. Det er noe som gjelder induksjonen i sin alminnelighet. Når vi sier at det er påviselig variasjon i vekten av grisungene mødrene imellom, må vi imidlertid for-

utsette at det forsøket som har gitt observasjonene som resultat, er slik planlagt og gjennomført at det er ulikheter i mødre-egenskaper (vi tenker kanskje særlig på arvefaktorer) som har slått ut i den signifikante F-verdi vi har fått. Vår tolking ville f.eks. ikke være holdbar dersom kullene hadde vært foret ulikt eller de hadde vokst opp under ulike vekstvilkår ellers.

I forsøk betyr T_1, T_2, \dots, T_k alternativer av det vi kaller forsøksfaktoren. T'ene kan f.eks. bety alternativer av kvantitativt eller/og kvalitativt variert gjødsling. Eller, de kan stå for alternative behandlingsmåter av goudaost. Vi må i slike tilfelle sørge for å planlegge og gjennomføre forsøket slik at det er ulikheter i effekten av de forskjellige alternativer av forsøksfaktoren som eventuelt slår ut i en signifikant F. Vi kan naturligvis bruke F-testet også i andre tilfelle, men da blir nok i regelen tolkingen av en signifikant F-verdi flertydig.

La oss f.eks. tenke oss at T_1-T_5 i Tab. G.3 betyr fem forskjellige år og at den observerte random variable (x) er karakteren i matematikk for en bestemt kategori av studenter. La oss videre tenke oss at resultatet av F-testen er det samme som for eksemplet i Tab. G.3. Vår konklusjon måtte i så fall gå ut på at det er påviselig ulikheter i karakternivået årene imellom. Et slikt resultat ville naturligvis kunne bli til nytte, men vi ville ikke kunne si bestemt hva grunnen til ulikheten er. Det kan være flere årsaker til slike ulikheter: variasjon i kvaliteten av undervisningen, variasjon i studentmaterialet (dvs. at de k samplene av studenter ikke er random sampler fra samme univers), variasjon i vanskelighetsgraden av eksamensspørsmålene m.m. Hvis observasjonene ikke er skaffet til veie gjennom et godt planlagt forsøk, må en være meget forsiktig med tolkingen.

Vi har tatt for oss variansanalysen og F-testen for det vi kan kalle en-veis grupperte observasjoner, dvs. observasjoner gruppert bare etter T_j ($j = 1, 2, \dots, k$). To-veis gruppering vil vi ha hvis observasjonene er skaffet til veie ved et forsøk etter blokkplanen. Som eksemplet i Tab. A1 viser, har vi da gruppering både etter T_j ($j=1, 2, \dots, k$) og etter blokk. Variansanalysen og F-testen for slike tilfelle skal vi ta for oss i sammenheng med analysen av forsøksdata. Vi skal også vise senere at vi har bruk for F-testen i andre tilfelle.

H. Metoder i eksperimentalforskningen.

H.1 Om formålet med et forsøk.

I avsnittene A.2 og A.3 er det allerede nevnt at formålet med et forsøk er å skaffe fram observasjoner som kan gi grunnlag for svar på spørsmål som er stilt. Av flere grunner er det nødvendig at disse spørsmålene er formulert på forhånd. En må ikke vente med spørsmålene til etter at forsøket er gjennomført. Grunnen til dette er bl.a. at typen av spørsmål vil være avgjørende for hvordan en skal planlegge forsøket og også for den analysen en vil utføre når observasjonene foreligger.

Svar på spørsmål vil en prøve å skaffe seg ved å sammenligne observasjonene for to eller flere forsøksledd. Ved et forsøksledd forstår vi da f.eks. en valt mengde av et vitamin tilsatt en eller annen standardkost til smågriser eller et valt antall kilo kaliumgjødsel til en sort poteter. Et spørsmål kan da være: hvor stor avlingsøkning vil en kunne vente å oppnå ved å øke mengden av kaliumgjødsel med så og så mange kilo pr. dekar? For å få svar på et slikt spørsmål må en naturligvis planlegge og utføre forsøket med i det minste to forsøksledd.

I dette kurset vil vi forutsette at det er to eller flere forsøksledd med. Betegnelsen forsøk eller eksperiment brukes også når det er bare ett ledd. Dette vil da si at undersøkelsen går ut på at det utføres en handling under gitte vilkår og at en observerer hva som hender som en konsekvens av handlingen. For å trekke et skille mellom denne typen eksperiment og forsøk med i det minste to forsøksledd, kan en betegne den siste typen som sammenlignende forsøk.

Oppstillingen av spørsmål og valg av forsøksledd er naturligvis forsøkslederens sak. Men i grove trekk kan vi likevel ordne

spørsmålene og typen av forsøksledd i tre grupper:

(1) Forsøksleddene er kvalitative og hovedspørsmålet gjelder rangering av leddene. Dette er i regelen situasjonen når forsøksleddene er sorter av en vekst (f.eks. poteter), kvalitativt variert føring til smågriser og lignende.

(2) Forsøksleddene er kvalitative eller/og kvantitative og det gjelder spørsmål om differensen mellom to forsøksledd som er utpekt på forhånd. En ønsker f.eks. å estimere den forskjell i avlingsmengde en kan vente ved å øke mengden av kaliumgjødning til en sort poteter.

(3) Forsøksleddene er kvantiteter og en ønsker å finne en regel for hvordan effekten avhenger av disse. Som eksempel kan vi tenke oss forsøksleddene 0 kg (T_1), 25 kg (T_2) og 50 kg (T_3) kaliumgjødning til en vekst og en ønsker å finne en regel for hvordan effekten på f.eks. avlingsmengden avhenger av mengden av kaliumgjødning.

I praktisk forsøksvirksomhet vil disse gruppene kunne forekomme i blanding. Dette vil en finne at de kan være i f.eks. såkalte faktorielle forsøk. Grunnen til at vi stiller opp disse gruppene er først og fremst at vi dermed kan få litt orden på de statistiske analysemetodene vi har bruk for.

Statistiske metoder som svarer til disse tre problemgruppene vil bli behandlet i særskilte avsnitt: gruppe (1) i H.4, gruppe (2) i H.2 og H.3, og gruppe (3) i I.4.

Til et forsøk må det velges et forsøksmateriale, f.eks. et større eller mindre stykke av en åker. Et forsøksmateriale er enten delt eller må deles i forsøksenheter og i gjentak. Bruker en blokkplanen (se Fig.A.1) i et feltforsøk, må en først dele feltet i blokker og blokkene igjen i forsøksenheter. I dette tilfelle er det

da blokkene som oppfattes som gjentak. En kan imidlertid sløyfe oppdelingen i blokker, og da er det forsøksenhetene som må oppfattes som gjentak. I et fôringsforsøk til smågriser kan en bruke de enkelte individene som gjentak eller en kan ordne det slik at det er kull av smågriser som er gjentak.

I avsnitt A.3 har vi forsøkt å forklare at de observasjoner som vi skaffer oss ved slike forsøk, er det grunnlaget vi har å bygge på når vi skal prøve å gi svar på de oppstilte spørsmål. Svarene vil da alltid gis i form av utsagn og disse er utsagn om en større mangfoldighet av gjentak enn de gjentak som er benyttet i forsøket. Våre utsagn eller konklusjoner gjelder for eller kan appliseres på et univers eller en populasjon av gjentak, og når det gjelder forsøk er dette universet alltid en abstraksjon. Siden det er samplet av gjentak som er representant for universet, vil bredden av universet og dermed gyldighetsområdet for konklusjonene være avhengig av graden av heterogeniteten i forsøksmaterialet. Sett fra dette synspunkt er et mer heterogent forsøksmateriale mer verdifullt enn et mindre heterogent. På den annen side vil en få mindre presise estimater når heterogeniteten er stor enn når den er mindre, forutsatt at antallet av gjentak er det samme.

Forsøk av den typen som vi her har brukt for å illustrere tankegangen, er slike som en kan kalle lokale. De konklusjoner en kommer til ved å basere seg på slike forsøk, er alltid betinget av de spesielle vilkår forsøket er utført under. På et forsøksfelt som består av en del av en åker, er som oftest jordtypen omtrent den samme overalt. Jorda kan f.eks. over hele feltet være sterkt leirholdig. Det er vel umiddelbart innlysende at gyldigheten av et utsagn som bygger på observasjoner fra et slikt forsøk, må være betinget av jordtypen. Vi kan derfor ikke uten videre applisere våre

utsagn på et univers hvor gjentakene er karakterisert ved at jorda er sterkt sandholdig.

At de utsagn som bygger på observasjoner fra slike lokale forsøk er betinget, er det neppe unntak fra. Vi må derfor se på resultatene fra slike forsøk som noe foreløpig. Dette er iallfall hovedregelen. Det kan naturligvis vise seg at disse betingelsene har liten betydning, men det kan en jo ikke vite uten å utvide rammen for forsøket. Det finnes imidlertid også eksempler på at utsagn fra et lokalt forsøk har tilstrekkelig gyldighet.

I de aller fleste tilfelle er det mer enn en random variabel som blir observert. Er T_j ($j=1,2,..k$) ulike mengder av et gjødselslag til en hvetesort, kan det være at størst interesse knytter seg til størrelsen av avlingen. En vil imidlertid også samtidig være interessert i slike egenskaper som stråstivhet og bakeevne. Det kan vise seg at den eller de forsøksledd som kan ventes å gi de største avlingene, kan ha uheldige virkninger på f.eks. stråstivheten. På samme måte kan det vise seg at en økning av planteavstanden under planting av gran gir økt tilvekst, samtidig som den fører til mindre verdifullt trevirke.

Svært mange forsøksledd er av en slik karakter at det er nødvendig å skaffe seg observasjoner av mer enn en random variabel. Dette er naturligvis særlig viktig hvis det resultat en kommer til blir omsatt i handling, for da er det jo nødvendig at en kjenner så langt det er mulig alle de konsekvenser handlingen kan ha. Vi kan imidlertid her ikke komme nærmere inn på hvordan en skal kombinere observasjoner av flere random variable. I det følgende vil vi derfor beskjefte oss med observasjoner av bare en random variabel, som f.eks. avlingsmengde, vektøkning hos smågriser, tilvekst i skog og vekttap under lagring av goudaost.

H.2 Fri randomisering.

Både i avsnitt A.2 og i avsnitt G.3 har vi vært inne på hva en forstår med fri randomisering eller randomisering uten restriksjoner. La oss nå som eksempel tenke oss at en ønsker å finne ut hva det betyr for vektøkningen hos smågriser at en tilsetter føret en viss mengde vitamin A og hva det betyr at en tilsetter en viss mengde kalsium. For å få greie på dette må en utføre et forsøk med $k = 3$ forsøksledd: $T_1 =$ normalt fôr, $T_2 =$ normalt fôr tilsatt en valt mengde vitamin A, og $T_3 =$ normalt fôr tilsatt en valt mengde kalsium.

Fra en eller flere besetninger tar en så ut ved loddtrekning de forsøksdyr en mener å ha bruk for. Det må da naturligvis først tas standpunkt til hva en vil forstå med "smågriser". La oss si at dette betyr griser i alderen mellom 5 og 10 uker. Blant de griser i denne aldersgruppen som er til disposisjon, tas det ut n dyr til hver av de tre forsøksleddene, alt i alt $N = 3n$. Det er rimelig og vanlig i praksis at det tas ut det samme antall forsøksenheter til hvert forsøksledd. Dette vil da si, når antall forsøksledd er k , at antall forsøksenheter som må skaffes til veie er $N = nk$.

En må også ta standpunkt til størrelsen av n . Det som blir avgjørende da er 1) hvilken presisjon en tar sikte på, 2) hvor mange forsøksenheter en har å velge blant, og 3) hvilke midler en har til rådighet. Med midler forstås både pengemidler, teknisk assistanse og i noen tilfelle plass til gjennomføring av forsøket. Det er viktig når en skal ta standpunkt til størrelsen av n at en har erfaringer fra lignende forsøk å bygge på. Fra et statistisk synspunkt er det om å gjøre at n er så stor som mulig.

Det hender imidlertid ikke sjelden at noen forsøksenheter faller bort i løpet av den tiden forsøket varer. Forsøksdyr blir syke

og må derfor tas ut. Det samme gjelder plantekultur-forsøk hvor det ikke sjelden hender et eller annet med forsøksenheter (ruter) som gjør at de må skjalttes ut. En må imidlertid vise stor forsiktighet og - en kan kanskje også si - tilbakeholdenhet når det må tas standpunkt til om en forsøksenhet skal tas ut. Består forsøksmaterialet av smågriser og den random variable en har under observasjon er vektøkningen, vil det at et forsøksdyr får lungebetennelse sikkert være god nok grunn til å ta det ut av forsøksmaterialet. Derimot må ikke et forsøksdyr tas ut fordi det vokser særlig lite eller særlig meget. Er en ikke kritisk nok kan en komme til å forvanske det observasjonsmateriale som skal brukes som grunnlag for å gi svar på de spørsmål som er årsaken til at forsøket ble satt i gang.

Vi må imidlertid regne med at når forsøket er utført, vil det ikke sjelden vise seg at det antall forsøksenheter en har observasjoner for, ikke er det samme for alle forsøksledd. antallet må vi da betegne med n_j ($j=1,2,\dots,k$), og det totale antall observasjoner blir lik $N = \sum n_j$. Det observasjonsmateriale som er gjengitt i Tab. G.3 kan tjene som et aktuelt eksempel.

Å utføre et forsøk i praksis er en vanskelig oppgave. Både under planleggingen og i forsøksperioden vil det som oftest melde seg spørsmål som det ikke er lett å ta standpunkt til. I en fremstilling som handler om forsøksmetode i sin alminnelighet, må en gå forbi disse spørsmål.

Hvis et forsøk planlegges og utføres etter prinsippet fri randomisering, er det de enkelte forsøksenhetene som er gjentak. De N gjentakene er da å oppfatte som representant for et abstrakt univers. For blant annet å få greie på hvordan vi skal analysere de N observasjonene er det nødvendig at vi prøver å bli klar over

hvilke elementer en observasjon (x_{ji}) består av for et tilfeldig gjentak i universet. Dette vil da si at vi må danne oss en modell for x_{ji} . En slik modell er en ligning som uttrykker x_{ji} som en sum av ledd med en bestemt betydning. Vi påstår ikke at alle leddene eksisterer i alle aktuelle situasjoner. Vi må imidlertid regne med den mulighet at de eksisterer og innrette oss etter det.

For et forsøk etter prinsippet fri randomisering er denne modellen meget enkel. Vi må tenke oss at noe av variasjonen i x_{ji} kommer av at den observerte random variable har ulik forventning forsøksleddene imellom, altså at $E(x_{ji}) = \mu_j$. Resten av variasjonen kommer av at forsøksmaterialet er heterogent. I et feltforsøk skyldes variasjonen at vekstvilkårene ikke er identiske rutene i mellom. En tredje årsak til variasjon i observasjonene er at vi gjør feil under observasjonsarbeidet. Den variasjon i observasjonene som disse feil fører til, kan vi imidlertid slå sammen med den variasjon som skyldes heterogeniteten i forsøksmaterialet, til en random variabel e . Modellen blir derfor

$$x_{ji} = \mu_j + e_{ji}$$

Det universet vi tenker på her, er det som de N forsøksenheter representerer i egenskap av et random sampel. En realistisk forutsetning om e er at den i dette universet har forventningen null. På grunn av at forsøksenhetene blir fordelt på de k forsøksleddene (T_j) ved loddtrekning, vil de k samplene av forsøksenheter være random sampler fra samme univers, slik at forventningen for e er lik null for hvert forsøksledd. Vi har derfor for hvert T_j at

$$E(x_{ji}) = \mu_j$$

Vi ser da også hva hensikten med fordelingen av forsøksenhetene på forsøksleddene ved loddtrekning er. Hvis vi ikke bruker lodd-

trekning, har vi ingen garanti for at de k samplene av forsøksenheter er random sampler fra samme univers. Bruk av andre prinsipper for fordelingen av forsøksenhetene gir ingen garanti for at $E(e) = 0$ for hvert T_j - og derfor heller ikke for $E(x_{ji}) = \mu_j$.

Om e må vi forutsette at e_{ji} er innbyrdes uavhengige verdier både innen hvert av de k samplene og mellom samplene. Denne forutsetningen er vel neppe alltid helt realistisk, men vi skal ikke komme nærmere inn på denne saken her fordi en nærmere behandling av den hører mer naturlig med til forsøksteknikk. Vi vil derfor her gå ut fra at forutsetningen er tilfredsstilt.

Om e er det dessuten vanlig å forutsette at den er en random variabel med normal fordelingsfunksjon og konstant varians (σ_e^2). Ingen av disse forutsetningene er realistiske. La oss ta for oss forutsetningen om at e har konstant varians.

La oss tenke oss $k = 2$ og at T_1 og T_2 er behandlingsmåter, f.eks. to mengder gjødsel, før til griser uten og med tilsetning av et vitamin, planteavstand under planting av gran, to behandlingsmåter av ost osv. Hvis vi da i modellen $x_{ji} = \mu_j + e_{ji}$ forutsetter at e har samme varians for de to T_j , vil ulikheten i forventningene være den eneste forskjell i effektene av T_1 og T_2 . Blant de årsaker som i et feltforsøk er bestemmende for heterogeniteten i forsøksmaterialet, kan vi nevne: variasjon i jordas fysiske tilstand, variasjonen i dens innhold av næringsemner (P,K,N...) og variasjon i nedbørmengden og den mengde varme som tilføres røttene i vekstsesongen. Er da T_1 og T_2 to mengder av et gjødselslag, vil det å akseptere at e har konstant varians være ensbetydende med å akseptere at den samlede effekt av T_j og de mange miljøfaktorene er en sum av effekten av forsøksfaktoren og effekten av miljøfaktorene. Vi kan vel si at det å akseptere noe slikt må bero på et sterkt for-

enklet og lite realistisk syn på hvordan årsaker virker sammen. Vi skal derfor i det følgende regne med den muligheten at σ_e er forskjellig for de forskjellige T_j . Som vi skal se har dette den konsekvens at f.eks. $\mu_1 - \mu_2$ og $\mu_2 - \mu_3$ ikke kan estimeres med samme presisjon.

Av modellen finnes at

$$\bar{x}_j = \mu_j + \bar{e}_j$$

og

$$\bar{x} = \mu + \bar{e}$$

hvor μ er gjennomsnittet av de k μ_j ; $\mu = \frac{1}{k} \sum \mu_j$. For to valte T_j , f.eks. T_p og T_q , har vi at

$$\bar{x}_p = \mu_p + \bar{e}_p$$

$$\bar{x}_q = \mu_q + \bar{e}_q$$

og derfor at

$$\bar{x}_p - \bar{x}_q = (\mu_p - \mu_q) + (\bar{e}_p - \bar{e}_q)$$

Siden $E(e) = 0$ for hver T_j , er også $E(\bar{e}_p) = E(\bar{e}_q) = 0$ og følgelig er

$$E(\bar{x}_p - \bar{x}_q) = \mu_p - \mu_q$$

dvs. at $(\bar{x}_p - \bar{x}_q)$ er en forventningsrett estimator av $(\mu_p - \mu_q)$.

Videre finner vi (sml. side 108) at

$$\text{var}(\bar{x}_p - \bar{x}_q) = \frac{\sigma_p^2}{n_p} + \frac{\sigma_q^2}{n_q}$$

Teknikken for beregning av konfidensgrensene for $(\mu_p - \mu_q)$ er beskrevet side 127. Vi erstatter bare fotskriftene 1 og 2 med p og q .

En differens mellom forventninger, f.eks. $\mu_p - \mu_q$, kaller vi en kontrast. Er $k = 2$, er det bare en kontrast, nemlig $\mu_1 - \mu_2$ (el. $\mu_2 - \mu_1$). Med k forsøksledd er det flere kontraster som ønskes estimert og det behøver ikke være differensen mellom to forventninger. Vi kan ha kontraster som $(\mu_p - \mu_q)$ og $(\mu_r - \mu_s)$, men vi

kan også ha slike kontraster som $(\mu_p + \mu_q - 2\mu_r)$ og $(\mu_p + \mu_q - \mu_r - \mu_s)$.

Det må imidlertid være fullt samsvar mellom de kontrastene en opererer med under analysen og de spørsmål som forsøket er planlagt å gi svar på. Både spørsmålene og de kontrastene som svarer til disse, må altså være formulert før forsøket settes i gang. Vi kan si at de kontrastene en ønsker estimert, skal være nevnt i planen for forsøket.

La oss nå tenke oss at $k = 4$, altså at det er 4 forsøksledd T_1, T_2, T_3 og T_4 . Vi kan som eksempel tenke oss 4 mengder av et gjødselslag, f.eks. 0 kg (T_1), 50 kg (T_2), 100 kg (T_3) og 200 kg (T_4) pr. dekar. La oss også tenke oss at den observerte random variable (x) er mengde avling av f.eks. en sort bygg.

To kontraster som sikkert har interesse i et slikt tilfelle er $(\mu_2 - \mu_1)$ og $(\mu_4 - \mu_3)$. Estimatorene er $(\bar{x}_2 - \bar{x}_1)$ og $(\bar{x}_4 - \bar{x}_3)$ og konfidensgrensene for hver av kontrastene kan beregnes slik som beskrevet side 127. Bruker vi til beregning av disse grensene den α -verdi i Tabell I som svarer til $P = 0,05$, vil konfidenssannsynligheten for hvert av de to intervallene være $Q = 1 - P = 0,95$.

La oss tenke oss at planen for forsøket også omfatter estimering av kontrasten $(\mu_3 - \mu_2)$. Estimatoren for denne kontrasten er naturligvis $(\bar{x}_3 - \bar{x}_2)$. Men når vi så skal beregne konfidensgrenser for $(\mu_3 - \mu_2)$, kommer vi opp i vansker fordi $(\bar{x}_3 - \bar{x}_2)$ ikke er uavhengig av $(\bar{x}_2 - \bar{x}_1)$ og $(\bar{x}_4 - \bar{x}_3)$. Dessuten vil det også være korrelasjon mellom de s 'ene som brukes. Under beregningen av konfidensgrensene for $(\mu_2 - \mu_1)$ bruker vi

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

for $(\mu_4 - \mu_3)$

$$s^2 = \frac{(n_3 - 1)s_3^2 + (n_4 - 1)s_4^2}{n_3 + n_4 - 2}$$

og for $(\mu_3 - \mu_2)$ må vi så bruke

$$s^2 = \frac{(n_2-1)s_2^2 + (n_3-1)s_3^2}{n_2 + n_3 - 2}$$

Fordi vi bruker s_2 og s_3 om igjen er det korrelasjon mellom den siste og de to første.

Undersøkelser har vist at trass i disse korrelasjonene er konfidenssannsynligheten i hvert tilfelle tilnærmet lik $Q = 1-P$. Den effekten som korrelasjonene har på konfidensnivået er så liten at den ikke har noen betydning i praksis. ^{*)}

Det antall spørsmål som en eventuelt kan få svar på med k forsøksledd er $k-1$. Dette antallet svarer til antall frihetsgrader for V_T (mellom-gruppe variansen) i variansanalysen (se side 151 o.flg.). Det er imidlertid spørsmålene som er det primære. Har en m spørsmål, må en utføre et forsøk med $k = m+1$ forsøksledd. En må da først skrive om spørsmålene til kontraster, og når det er gjort, vil en også vite hvilke forsøksledd en skal ta med.

Vi har hittil ikke nevnt noe om hvilken funksjon variansanalysen har når det gjelder slike forsøk. Omfatter planen for forsøket alle de kontrastene en ønsker å få estimert, kan en vel si at variansanalysen er overflødig. Den kan da neppe gi noen opplysninger om kontrastene (og dermed svar på spørsmålene) som ikke konfidensgrensene for kontrastene gir. Men det er mange eksempler på forsøk hvor planen ikke gir beskjed om kontrastene. De fleste sortsforsøk hører med blant disse. Formålet med forsøket er å finne fram til en rangering av forsøksleddene og da er variansanalysen

*) Vi må her nøye oss med å vise til P. Ottestad: Statistical Models and their Experimental Application. Charles Griffin & Company, London 1970.

og F-testen nyttige hjelpemidler. Vi skal imidlertid ta opp denne problemstilling i et eget avsnitt, H.4.

La oss tenke oss at forsøket er planlagt med n gjentak for hvert forsøksledd. Da har vi at $\text{var}(\bar{x}_p - \bar{x}_q) = (\sigma_p^2 + \sigma_q^2)/n$. Forutsetter vi at ingen av gjentakene må tas ut i forsøksperioden, vil bredden av konfidensintervallet for kontrasten $(\mu_p - \mu_q)$ være proporsjonal med $\sqrt{(s_p^2 + s_q^2)/n}$. I forsøk etter prinsippet fri randomisering gjør derfor heterogeniteten i forsøksmaterialet seg gjeldende med full tyngde. Vi kan redusere intervallbredden og dermed øke presisjonen ved å øke antall (n) gjentak for hvert forsøksledd. Men dette betyr jo straks et tillegg i kostnader, og i mange tilfelle er det heller ikke lett å skaffe nye forsøksenheter.

H.3 Blokkplanen.

Som vi har forklart i avsnitt A.2 går blokkplanen i et feltforsøk ut på at forsøksfeltet først deles i et antall (n) blokker og deretter deles hver blokk i like mange (k) ruter eller forsøksenheter som det er forsøksledd. Ved loddtrekning fordeles så forsøksleddene på rutene innen hver blokk. Det en oppnår med dette er at forsøksleddene kommer gjennomgående nærmere hverandre på feltet enn om en fordelte leddene på alle $N = nk$ rutene etter prinsippet fri randomisering. En kan derfor vente at heterogeniteten i forsøksmaterialet vil gjøre seg svakere gjeldende under sammenligningen av to forsøksledd enn om fri randomisering blir brukt. Om det i et aktuelt tilfelle faktisk er slik avhenger av hvor stor heterogeniteten er, om en har vært heldig med oppdelingen av feltet i blokker, og av antall forsøksledd. Er heterogeniteten svak er det ikke rimelig

å tro at presisjonen ved estimeringen av kontrastene vil bli økt nevneverdig, om noe i det hele tatt, sammenlignet med presisjonen ved fri randomisering. Det samme er tilfelle hvis antall forsøksledd er stort.

På samme måte og av samme grunn er det rimelig å regne med at en vil oppnå noe større presisjon ved estimeringen av kontraster ved å bruke kull som blokker i et forsøk med alternative fôringer til smågriser. Stort sett vil vel to griser fra samme kull ligne hverandre (genetisk, alder, tidligere fôring og stell) mer enn to griser fra forskjellige kull. Heterogeniteten innen kull er rimeligvis svakere enn heterogeniteten mellom kull.

Bruker en kull som blokker i et fôringsforsøk med smågriser som forsøksmateriale, må en ved loddtrekning ta ut k griser fra hvert av n kull slik at en også i dette tilfelle har k forsøksenheter i hver blokk. De k forsøksleddene fordeles så ved loddtrekning på de k forsøksenheterne innen hver blokk.

Når et slikt blokkforsøk er utført og alt har gått etter planen, vil vi ha $N = nk$ observasjoner av den observerte random variable (x). Disse kan så ordnes etter forsøksledd (T_j) og blokk slik som eksemplet i Tab. H.1 viser. Observasjonene er her avlingsmengde (gitt i antall kg pr. forsøksenhet eller rute) for tomatplanter. Disse plantene er utsatt for en sykdom som går under navn av "brune røtter" og som har den virkning at den reduserer avlingsmengden. En har prøvd å finne kjemiske midler å behandle jorda med for å motvirke sykdommen, og i sammenheng med det ble det utført et blokkforsøk ($n=4$ blokker) med $k = 7$ slike desinfeksjonsmidler som forsøksledd. Vi skal nøye oss med å ta med resultatene for $k = 3$ av dem, i tabellen betegnet med T_1 , T_2 og T_3 .

Tabell H1

Blokk	Forsøksledd			S_i
	T_1	T_2	T_3	
B_1	39	36	41	116
B_2	33	42	44	119
B_3	41	47	45	133
B_4	39	44	46	129
S_j	152	169	176	497

I tabellen er også oppført summen (S_j) for hvert forsøksledd, summen (S_i) for hver blokk og summen av alle $N = nk$ observasjonene ($S = 497$).

Vi skal nå ta for oss spørsmålet om hvordan observasjonene kan brukes til estimering av kontraster og beregning av konfidensgrenser. I neste avsnitt (H.4) skal vi så komme inn på variansanalysen og F-testen. Først må vi imidlertid prøve å bli klar over hva en skal forstå med universet, og vi må lage en modell for den observerte random variable.

Med universet må vi her forstå et som er representert av de n blokkene oppfattet som et random sampel. I samsvar med dette bruker en betegnelsen gjentak om blokkene. Dette universet er naturligvis ikke identisk med det som er representert av de N forsøksenhetene slik situasjonen er når prinsippet fri randomisering blir nyttet. Som oftest er vel det siste universet bredere eller mer generelt enn det første, men forskjellen kan ikke være av betydning for valget av forsøksplan.

Verdiene av den observerte random variable og observasjonene selv vil vi også her betegne med x_{ji} slik at $j = 1, 2, \dots, k$ og $i = 1, 2, \dots, n$. Den første fotskriften (j) er altså numerert på forsøksleddet og den siste (i) numerert på blokken. Som for fri randomisering må vi ^{i} modellene for x_{ji} ha med et ledd (μ_j) som skyldes at

forventningen for x er forskjellig forsøksleddene imellom. Fordi det betyr en liten forenkling skal vi imidlertid skrive $\mu_j = \mu + a_j$. Her er da μ et felles nivå og a_j står for avviket av μ_j fra dette nivået. Uten å innskrenke gyldigheten av modellen kan vi sette at $\sum a_j = 0$. Er nemlig denne summen forskjellig fra null, betyr det at a_j inneholder en konstant addend som rettelig hører med til μ . Siden forsøksleddene er valte alternativer er både μ og a_j ($j=1,2,\dots,k$) å oppfatte som parametere.

Som for fri randomisering må vi ha med et ledd (e_{ji}) som skyldes de faktorer som er årsak til heterogeniteten mellom forsøksenheterne innen blokkene. Dette leddet vil dessuten inneholde observasjonsfeil og samspillet mellom forsøksledd og disse heterogenitetsfaktorene.

Det er imidlertid også heterogenitet blokkene imellom. I den effekt på x som disse heterogenitetsfaktorene har, kan vi skille ut en effekt som er felles for alle forsøksledd. Denne effekten skal vi betegne med z_i . Dessuten må en jo regne med at det er samspill mellom forsøksledd og disse heterogenitetsfaktorene. Denne samspilleffekt skal vi betegne med w_{ji} .

Modellen blir derfor

$$x_{ji} = \mu + a_j + z_i + w_{ji} + e_{ji}$$

Her er da z_i n verdier av en random variabel z som gir en beskrivelse av effekten av f.eks. vekstvilkårene og som er felles for alle forsøksledd. Uten å innskrenke gyldigheten av modellen kan vi sette $E(z) = 0$, slik at z_i er verdier som varierer omkring null. Begrunnelsen for at vi kan sette $E(z) = 0$ er den samme som for $\sum a_j = 0$. Vi vil også om z_i forutsette at z_i er n innbyrdes uavhengige verdier av z . Dette er i samsvar med at vi oppfatter blokkene som gjentak i et univers.

Som oftest settes $w = 0$. Modellen blir da

$$x_{ji} = \mu + a_j + z_i + e_{ji}$$

La oss undersøke hva dette betyr. La oss ta for oss to forsøksledd T_p og T_q . For blokknr. i har vi da

$$x_{pi} = \mu + a_p + z_i + e_{pi}$$

og

$$x_{qi} = \mu + a_q + z_i + e_{qi}$$

Differensen er

$$x_{pi} - x_{qi} = (a_p - a_q) + (e_{pi} - e_{qi}) = (\mu_p - \mu_q) + (e_{pi} - e_{qi})$$

Ser vi bort fra den random differensen $(e_{pi} - e_{qi})$, er forskjellen $(x_{pi} - x_{qi})$ nøyaktig den samme for alle blokkene. Tenker vi oss et feltforsøk der T_q og T_p er to ulike mengder av et gjødselslag, har vekstvilkårene, slik de varierer blokkene imellom, ingen betydning. Er x avlingsmengder, vil det si at det vi kan vente å oppnå ved å øke gjødselmengden fra T_p til T_q er uavhengig av vekstvilkårene. Vi må vel derfor kunne si at en påstand om at $w = 0$ må bero på et sterkt forenklet og lite realistisk syn på hvordan årsaker virker sammen.

Det er godt mulig at w spiller liten eller ingen rolle i mange tilfelle. Men vi kan ikke vite om det er slik, verken når vi skal planlegge forsøket eller når vi skal ta fatt på den statistiske analysen. Vi bør derfor helst regne med at w eksisterer.

I et forsøk med griser der kull er nyttet som blokker, vil z være effekten av slike faktorer (miljøfaktorer) som genetisk konstitusjon og alder. Er da f.eks. T_1 en standardføring og T_2 samme standardføring med tilsetning av en viss mengde av et vitamin, vil $w = 0$ bety at effekten av vitamintilsetningen er helt uavhengig av miljøfaktorene. Å forutsette en slik uavhengighet er lite realistisk. Derfor må modellen også i dette tilfelle inneholde leddet w .

Som nevnt er z_i å oppfatte som n uavhengige verdier av en random variabel z , med $E(z) = 0$. På samme måte må vi oppfatte w_{ji} som n uavhengige verdier av en random variabel w_j . I modellen står derfor w_{ji} for n uavhengige verdier av hver av k random variable. Uten å innskrenke gyldigheten av modellen kan vi for hver T_j sette $E(w_j) = 0$. For forsøksleddet T_j er effekten av miljøfaktorene $(z_i + w_{ji})$, og følgelig ulik for de forskjellige T_j . Siden nå både z og w er effekter av miljøfaktorene (ikke nødvendigvis de samme miljøfaktorene) må vi regne med korrelasjon mellom z på den ene siden og w 'ene på den andre. Av samme grunn må vi også regne med korrelasjon mellom w 'ene innbyrdes.

La oss så ta for oss estimeringen av kontrastene, f.eks. kontrasten $(\mu_p - \mu_q) = (a_p - a_q)$. Av modellen finner vi at gjennomsnittet for T_j er

$$\bar{x}_j = \mu + a_j + \bar{z} + \bar{w}_j + \bar{e}_j$$

og for estimatoren ($\bar{x}_p = \bar{x}_q$) at

$$\bar{x}_p - \bar{x}_q = (a_p - a_q) + (\bar{w}_p - \bar{w}_q) + (\bar{e}_p - \bar{e}_q)$$

Siden e står for variasjonen innen blokkene, er det realistisk å gå ut fra at det ikke finnes korrelasjon mellom e og w 'ene.

Vi har derfor at

$$\text{var}(\bar{x}_p - \bar{x}_q) = \text{var}(\bar{w}_p - \bar{w}_q) + \text{var}(\bar{e}_p - \bar{e}_q)$$

Og her er $\text{var}(\bar{w}_p - \bar{w}_q)$ ikke den samme for alle kontrastene. Estimeringen av de forskjellige kontrastene er derfor ikke like presis.

Konsekvensen av dette er at en til beregning av konfidensgrensene må bruke et middelavvik (s) beregnet særskilt for hver kontrast. For kontrasten $(\mu_p - \mu_q)$ må vi derfor bruke differensene $d_i = x_{pi} - x_{qi}$. Konfidensgrensene for kontrasten blir da

$$\bar{d} \pm a.s. / \sqrt{n}$$

hvor a er den verdi vi finner i Tabell I for $f = n-1$ frihetsgrader. Gjennomsnittet \bar{d} er jo her lik $\bar{d} = \bar{x}_p - \bar{x}_q$ og er derfor etter det som er vist foran, en forventningsrett estimator av kontrasten. Beregningen av konfidensgrensene er vist ved eksemplet i Tab.H.2. Observasjonene er fra et blokkforsøk med $n=5$ blokker og $k=4$ forsøksledd. Det er her forutsatt at det i planen for forsøket er bestemt at de tre kontrastene en ønsker estimert er $(\mu_2 - \mu_1)$, $(\mu_3 - \mu_2)$ og $(\mu_4 - \mu_3)$. Gjennomsnittene for d_{1i} , d_{2i} og d_{3i} er forventningsrette estimatører av disse kontrastene.

Tab. H.2. Konfidenssannsynlighet $Q = 0,95$

Blokk	T ₁	T ₂	T ₃	T ₄	d_{1i}	d_{2i}	d_{3i}
					$x_{2i} - x_{1i}$	$x_{3i} - x_{2i}$	$x_{4i} - x_{3i}$
B ₁	33	48	51	62	15	3	11
B ₂	30	39	50	57	9	11	7
B ₃	34	43	58	62	9	15	4
B ₄	37	51	52	60	14	1	8
B ₅	29	46	51	56	17	5	5
\bar{d}					12,8	7,0	7,0
s/\sqrt{n}					1,625	2,608	1,225
$a \cdot s/\sqrt{n} = 2,776 \cdot s/\sqrt{n}$					4,51	7,24	3,40
$\bar{d} - a \cdot s/\sqrt{n}$					8,29	-0,24	3,60
$\bar{d} + a \cdot s/\sqrt{n}$					17,31	14,24	10,40

På grunn av samspillseffektene (w) er det korrelasjon mellom estimatorene, f.eks. mellom $\bar{d} = \bar{x}_p - \bar{x}_q$ og $\bar{d} = \bar{x}_r - \bar{x}_s$. Det er også korrelasjon mellom middelavvikene for differensene d_i . Etter resultatene av de undersøkelser som er gjort, er imidlertid effekten på konfidenssannsynligheten av disse korrelasjonene uten betydning i praktiske anvendelser. Hvert av de tre konfidensintervallene i Tab.H.2 svarer derfor meget nær til konfidenssannsynligheten $Q = 0,95$.

H.4 Rangering og gruppering av forsøksledd.

I avsnitt H.1 er nevnt at et hovedformål med forsøk er å skaffe seg observasjoner som grunnlag for rangering av forsøksleddene, eventuelt også å gruppere dem. Dette er særlig aktuelt i slike tilfelle hvor forsøksleddene er kvalitative. Et eksempel har vi i forsøk hvor forsøksleddene er sorter, f.eks. k potetsorter. Det forekommer naturligvis også da at det i planen for forsøket er tatt med enkelte bestemte kontraster som ønskes estimert, men vi må se bort fra dette her. Spørsmålet om rangering kan også forekomme i tilfelle hvor forsøksleddene er kvantitative, men dette skyldes vel som oftest mangler ved forsøksplanleggingen.

En rangering av et antall sorter forutsetter i regelen at en tar hensyn til mer enn en egenskap. Gjelder det f.eks. hvetesorter, bør en ta hensyn til avlingsmengde, stråstivhet, bakeevne, sykdomsresistens m.m. Dette er imidlertid en meget komplisert sak som vi ikke kan ta opp her. Vi må nøye oss med å bruke observasjonene av en enkelt random variabel som f.eks. avlingsmengde.

I praksis vil en da naturligvis rangere etter verdien av gjennomsnittet \bar{x}_j . For eksemplet i Tab.H.1 finner vi disse verdiene: 38,00 for T_1 , 42,25 for T_2 og 44,00 for T_3 , og rangeringen blir $T_1 - T_2 - T_3$. Men en slik rangering kan jo bygge på sviktende grunnlag, nemlig hvis det ikke kan påvises ulikheter mellom forventningene μ_j ($j=1,2,\dots,k$). Før en tar i bruk gjennomsnittene for rangering av T_j , må det utføres en test av null-hypotesen

$$\mu_1 = \mu_2 = \mu_3 = \dots = \mu_k$$

Det er en forutsetning for en forsvarlig rangering av T_j ved hjelp av \bar{x}_j at de observasjoner forsøket har gitt, er tilstrekkelig grunnlag for å forkaste denne null-hypotesen.

Den testen som brukes oftest er F-testen. I avsnitt G.3 er det vist hvordan variansanalyse og F-test nyttes til testing av null-hypotesen $\mu_j = \text{konstant}$ for det tilfelle at forsøket er utført etter prinsippet fri randomisering. Vi skal nå ta for oss variansanalysen for observasjoner fra et blokk-forsøk. Også i dette tilfelle går variansanalysen ut på å dele opp kvadratsummen $\sum \sum (x_{ji} - \bar{x})^2$, hvor \bar{x} er gjennomsnittet av alle $N = nk$ observasjonene. Det er enkelt å vise at vi kan sette

$$\sum \sum (x_{ji} - \bar{x})^2 = k \sum (\bar{x}_i - \bar{x})^2 + n \sum (\bar{x}_j - \bar{x})^2 + \sum \sum (x_{ji} - \bar{x}_i - \bar{x}_j + \bar{x})^2$$

hvor \bar{x}_i ($i=1,2,\dots,n$) betyr blokkgjennomsnittene, $\bar{x}_i = S_i/k$.

Disse tre kvadratsummene danner så igjen grunnlaget for beregning av tre varianser, nemlig

$$V_B = \frac{k}{n-1} \sum (\bar{x}_i - \bar{x})^2$$

$$V_T = \frac{n}{k-1} \sum (\bar{x}_j - \bar{x})^2$$

$$V_R = \frac{1}{(n-1)(k-1)} \sum \sum (x_{ji} - \bar{x}_i - \bar{x}_j + \bar{x})^2.$$

Antall frihetsgrader for disse tre variansene er $f = n-1$ for V_B , $f = k-1$ for V_T og $f = (n-1)(k-1)$ for V_R .

Det er denne oppdeling av den totale kvadratsummen i tre kvadratsummer og beregningen av de tre tilsvarende varianser vi kaller en variansanalyse. De tre kvadratsummene kan beregnes direkte etter formlene. Men i praksis utføres arbeidet ved at en først beregner den totale kvadratsummen, kvadratsummen for V_B og kvadratsummen for V_T . Kvadratsummen for V_R finnes så som differensen mellom den første og summen for de to siste.

De beregningene/utfører i praksis er gitt ved følgende form-
ler, hvor S er summen av alle N observasjonene:

$$\begin{aligned} \sum \sum (x_{ji} - \bar{x})^2 &= \sum \sum x_{ji}^2 - S^2/N, \\ k \sum (\bar{x}_i - \bar{x})^2 &= \frac{1}{k} \sum S_i^2 - S^2/N \\ n \sum (\bar{x}_j - \bar{x})^2 &= \frac{1}{n} \sum S_j^2 - S^2/N. \end{aligned}$$

For vårt eksempel i Tab.H.1 har vi $n = 4$, $k = 3$ og $N = 12$. Vi har $S = 497$, og derfor $S^2/N = 20\,584,08$. Videre finnes $\sum \sum x_{ji}^2 = 20\,775$, $\sum S_i^2 = 61\,947$ og $\sum S_j^2 = 82\,641$. Dette innsatt gir:

$$\begin{aligned} \sum \sum (x_{ji} - \bar{x})^2 &= \dots\dots\dots 190,92 \\ k \sum (\bar{x}_i - \bar{x})^2 &= 64,92 \\ n \sum (\bar{x}_j - \bar{x})^2 &= \underline{76,17} \qquad \qquad \qquad 141,09 \\ &\qquad \qquad \qquad \underline{\sum \sum (x_{ji} - \bar{x}_i - \bar{x}_j + \bar{x})^2} \qquad \qquad \qquad 49,83 \end{aligned}$$

Resultatet stilles ofte opp i følgende skjema, der f betyr antall frihetsgrader:

Årsak	Kvadratsum	f	V
Blokk	64,92	3	
T	76,17	2	38,085
Rest	49,83	6	8,305
Total	190,92	11	

Vi har altså at $V_T = 38,085$ og $V_R = 8,305$.

Som alt nevnt er antall frihetsgrader for de tre variansene lik $(n-1)$, $(k-1)$ og $(n-1)(k-1)$. Vi konstaterer lett at

$$(n-1) + (k-1) + (n-1)(k-1) = nk-1 = N-1.$$

Det en tok sikte på med den undersøkelsen vi har brukt som eksempel, var å få brakt på det rene om de tre desinfeksjonsmidlene har ulik effekt på avlingsmengden. Nå kan naturligvis effekten av et desinfeksjonsmiddel gi seg utslag på flere måter, men det er i dette tilfelle den effekt det har på forventningen for den observerte random variable vi først og fremst er interessert i. Den null-hypo-

tesen vi i samsvar med dette må stille opp til testing, er nullhypotesen $\mu_1 = \mu_2 = \mu_3$.

En kan teste denne nullhypotesen ved F-testen. De to variansene en bruker er V_T og V_R . Variansen for blokk, altså V_B , har ingen interesse i denne sammenhengen. For vårt eksempel finner vi at

$$F = \frac{38,085}{8,305} = 4,59$$

med $f=2$ frihetsgrader for V_T og $f=6$ frihetsgrader for V_R . Vi ser da av Tabell III at denne F-verdi/er ^{ikke}signifikant på 5% nivået. Tabellen for $P = 0,05$ viser at den nedre grensen for det kritiske område for F i dette tilfelle er $a = 5,14$.

I samsvar med dette må konklusjonen gå ut på at forsøket ikke har gitt det nødvendige grunnlag for å ta standpunkt til hvilke av de tre desinfeksjonsmidlene er å foretrekke. Vi kan derfor heller ikke foreta noen rangering av forsøksleddene. Hvis vi bygger på signifikansnivået $P = 0,05$, er grunnlaget for å ta standpunkt ikke godt nok.

Tar vi for oss eksemplet i Tab. H.2, kommer vi til følgende variansanalyse:

Årsak	Kvadratsum	f	V
Blokk	106,20	4	
T	1960,15	3	653,38
Rest	92,60	12	7,72
Total	2158,95	19	

For dette tilfelle har vi $F = 653,38/7,72 = 84,63$, dvs. en verdi som er så stor at vi trygt kan forkaste nullhypotesen $\mu_j =$ konstant. Vi har derfor også et godt grunnlag å basere rangeringen av forsøksleddene på. For gjennomsnittene \bar{x}_j finner vi her 32,6 for T_1 , 45,4 for T_2 , 52,4 for T_3 og 59,4 for T_4 . Rangeringen blir derfor $T_1 - T_2 - T_3 - T_4$. Hvis våre observasjoner hadde vært

observasjoner av avlingsmengde, ville vi kunne si at T_4 er å foretrekke fremfor T_1 . Men vi ville ikke kunne si at T_4 er å foretrekke fremfor T_3 . Heller ikke kan vi si at T_3 er å foretrekke fremfor T_2 og T_1 og T_2 fremfor T_1 . En videregående undersøkelse er nødvendig før vi har grunnlag for slike konklusjoner. Vi kan imidlertid ikke her komme inn på de metodene vi da har bruk for.

For forsøk etter prinsippet fri randomisering er fremgangsmåten for rangering av forsøksleddene den samme som for blokkforsøk. Også da må det naturligvis kreves at observasjonene gir tilstrekkelig grunnlag for å forkaste null-hypotesen $\mu_j = \text{konstant}$.

Vi må til slutt komme noe inn på gyldigheten av F-testen. For det tilfelle at forsøket er utført etter prinsippet fri randomisering, er det modellen

$$x_{ji} = \mu + e_{ji}$$

som ligger til grunn for utledning av fordelingsfunksjonen for F og dermed for F-testen. Det er dessuten en forutsetning at e_{ji} er nk innbyrdes uavhengige verdier av en normal random variabel med forventningen $E(e) = 0$ og $\text{var}(e) = \text{konstant}$. Den tilsvarende modellen for forsøk etter blokkplanen er

$$x_{ji} = \mu + z_i + e_{ji}$$

med de samme forutsetningene om e_{ji} .

Det er derfor forutsatt i begge tilfelle at e ikke inneholder samspill mellom heterogenitetsfaktorer og forsøksledd. Dette er ikke realistisk. Det er heller ikke realistisk å forutsette at e har normal fordelingsfunksjon.

Sampelundersøkelser har vist at i første tilfelle (fri randomisering) har disse urealistiske forutsetningene liten betydning for F-testen. Dette vil si at hvis a er den verdi som avgrenser

det forkastningsområde for null-hypotesen som etter fordelingsfunksjonen for F svarer til f.eks. $P = 0,05$, er sannsynligheten for $F \geq a$ under realistiske forutsetninger (og forutsatt at null-hypotesen er treffende) ikke nevneverdig forskjellig fra $P = 0,05$.

I modellen for blokkforsøk ser vi at leddet (w) for samspill mellom forsøksledd og blokk mangler. Sampelundersøkelser har vist at dette har den betydning at F -testen er noe for følsom. Det kan bety så meget at de verdier av a som vi finner i Tabell III, altså for $P = 0,05$, kan svare til et signifikansnivå på noe slikt som $P = 0,10$. En må derfor være noe forsiktig med bruken av F -testen. Når det gjelder blokkforsøk og en har bestemt seg for å bruke signifikansnivået $P = 0,05$, er det sikkert fornuftig å bruke Tabell IV (for $P = 0,025$) i stedet for Tabell III.

H.5. Forsøk med flere faktorer.

I praktisk virksomhet kan en som oftest skille ut et større eller mindre antall enkeltoperasjoner. Er virksomheten planlagt, vil en ha beskrivelse av hver enkelt operasjon og disse beskrivelsene danner så sammen en oppskrift for det hele. En gardbruker som har bestemt seg til å bruke en bestemt åker til dyrking av hvete, må velge kornsort, ta standpunkt til hvordan det skal gjødsles, ta standpunkt til hvordan jorda skal bearbeides m.m. Oppskriften eller planen for hele operasjonen er et resultat av en mer eller mindre sammensatt beslutningsprosess.

Dette viser at skal en utføre et forsøk som kan gi grunnlag for veiledning for praktisk virksomhet, må en i regelen utføre forsøket med flere forsøksfaktorer. Det nytter ikke da å utføre isolerte forsøk med hver enkelt faktor om gangen fordi det kan finnes samspill mellom dem. Det er for eksempel tenkelig at effekten av en økning av mengden av kaliumgjødsel til en vekst kan være avhengig av mengden av kvelstoffgjødsel, av hvilken sort en dyrker, av måten å bearbeide jorda på osv. Det er derfor nødvendig at faktorene tas med i samme forsøket.

Slike forsøk kan bli meget kompliserte. Vi må her nøye oss med to faktorer og skal ta for oss et lokalt forsøk etter blokkplanen. Hvis en behersker planlegging og statistisk analyse for et slikt forsøk, er det ikke vanskelig å utvide til mer enn to faktorer.

La oss betegne de to faktorene med P_p ($p = 1, 2, 3 \dots r$) og Q_q ($q = 1, 2, 3 \dots s$). Vi kan tenke på et forsøk der P_p er r mengder av et gjødselslag P og Q_q s mengder av et annet gjødselslag Q til f.eks. en sort bygg. Men både P_p og Q_q kan naturligvis være alternative kvaliteter.

Forsøksleddet T_j ($j = 1, 2, 3, \dots, k$) er da sammensatt av en mengde P og en mengde Q . Det vanlige er at alle kombinasjoner av P_p og Q_q tas med. Forsøksleddene er kombinasjonene

$$\begin{array}{cccc} P_1 Q_1 & P_1 Q_2 & \dots & P_1 Q_s \\ P_2 Q_1 & P_2 Q_2 & \dots & P_2 Q_s \\ \vdots & \vdots & & \vdots \\ P_r Q_1 & P_r Q_2 & \dots & P_r Q_s \end{array}$$

Vi sier da at forsøksleddene er sammensatt ortogonalt av P og Q . Er k betegnelsen på antallet av forsøksledd, er $k = r \cdot s$.

I den effekt som forsøksleddet har på den observerte random variable (x) kan vi nå skille ut tre komponenter. Den første komponenten er en selvstendig effekt av P , dvs. en effekt som ikke er avhengig av eller betinget av Q . Vi kaller dette hovedeffekten av P . Den andre komponenten er en tilsvarende hovedeffekt av Q . Men det kan også være en blandet effekt, det vi kaller samspillet mellom P og Q . Dette kommer av at den endring av forventningen for den observerte random variable (f.eks. avlingsmengde) som skyldes en økning eller en minskning av mengden av P , kan være avhengig av mengden av Q . Er P_1 og P_2 to mengder av kaliumgjødsel ($P_1 < P_2$) og x er mengde avling, vil samspill mellom P og Q (Q f.eks. kvelstoffgjødsel) bety at den meravling en kan vente som et resultat av en økning av kaliumgjødsel fra P_1 til P_2 , er avhengig av Q .

I vår modell for blokkforsøk

$$x_{ji} = \mu + a_j + z_i + w_{ji} + e_{ji}$$

betyr dette at a_j er sammensatt additivt av tre komponenter:

$$a_j = b_p + c_q + d_{pq}$$

Her er da b_p og c_q hovedeffektene av P og Q og d_{pq} er samspillet.

Av samme grunn som vi foran satte $\sum a_j = 0$, kan vi her sette $\sum b_p = 0$, $\sum c_q = 0$ og $\sum d_{pq} = 0$ både når vi summerer over p og når vi summerer over q.

På samme måte består w_{ji} , som er samspillet mellom forsøksledd (T_j) og blokk, av tre komponenter;

$$w_{ji} = u_{pi} + v_{qi} + t_{pqi}$$

Her er da u_{pi} samspillet mellom P og blokk, v_{qi} er samspillet mellom Q og blokk og t_{pqi} er samspillet mellom PQ-samspillet og blokk. Til disse kan vi, uten at modellen taper noe i gyldighet, sette som krav at

$$E(u) = 0 \quad \text{for hver } p$$

$$E(v) = 0 \quad \text{for hver } q$$

og

$$E(t) = 0 \quad \text{for hver } p \text{ og } q.$$

Modellen blir da

$$x_{pqi} = \mu + b_p + c_q + d_{pq} + z_i + u_{pi} + v_{qi} + t_{pqi} + e_{pqi}$$

($p = 1, 2, \dots, r$, $q = 1, 2, \dots, s$ og $i = 1, 2, \dots, n$). Det siste leddet, e_{pqi} , står her (som foran) for $N = nrs$ innbyrdes uavhengige verdier av en random variabel e med $E(e) = 0$. Men $\text{var}(e)$ kan også her være avhengig av P_p og Q_q .

Et forsøk med flere faktorer kan - som forsøk med en faktor - ha til hovedformål å gi grunnlag for estimering av kontraster. En slik kontrast er f.eks.

$$\begin{aligned} \mu_{32} - \mu_{22} &= (b_3 + c_2 + d_{32}) - (b_2 + c_2 + d_{22}) \\ &= (b_3 - b_2) + (d_{32} - d_{22}) \end{aligned}$$

og er den endring i $E(x)$ som følger av at mengden av P endres fra P_2 til P_3 samtidig som mengden av Q er lik Q_2 .

En annen kontrast er

$$\mu_{22} - \mu_{11} = (b_2 + c_2 + d_{22}) - (b_1 + c_1 + d_{11})$$

og er den endring i $E(x)$ som følger av en samtidig endring av P fra P_1 til P_2 og av Q fra Q_1 til Q_2 .

For hver kontrast har vi n observasjoner, en for hver blokk. For den siste kontrasten har vi differensene $(x_{22i} - x_{11i})$ som observasjoner. Gjennomsnittet er $(\bar{x}_{22} - \bar{x}_{11})$, og som for forsøk med en faktor vil vi finne at

$$E(\bar{x}_{22} - \bar{x}_{11}) = \mu_{22} - \mu_{11}$$

Disse n observasjonene brukes som vist i avsnitt H.3 til beregning av konfidensgrensene for kontrastene. Kontrastene må da være utpekt på forhånd, dvs. tatt med i forsøksplanen. Arbeidet med å sette opp kontrastene slik at de stemmer overens med de spørsmål en venter svar på, kan være meget vanskelig. Vi har imidlertid ikke tid til å ta med noe mer om dette her.

Vi skal så vise hvordan variansanalysen tar seg ut for et to-faktor forsøk. Det er ingen større vansker med dette når T_j er satt sammen ortogonalt av P_p og Q_q , slik at antall forsøksledd er $k = rs$. Da vil nemlig gjennomsnittene $\bar{x}_j = \bar{x}_{pq}$ kunne settes opp i en 2-veis gruppering etter P_p ($p = 1, 2, \dots, r$) og Q_q ($q = 1, 2, \dots, s$). Vi kan derfor bruke den samme oppdeling som vi brukte i avsnitt H.3 og vil ha at

$$\begin{aligned} n \sum (\bar{x}_j - \bar{x})^2 &= n \sum \sum (\bar{x}_{pq} - \bar{x})^2 \\ &= ns \sum (\bar{x}_p - \bar{x})^2 + nr \sum (\bar{x}_q - \bar{x})^2 + n \sum \sum (\bar{x}_{pq} - \bar{x}_p - \bar{x}_q + \bar{x})^2 \end{aligned}$$

Kvadratsummen for V_R kan deles opp på tilsvarende måte. Vi kan vise at

$$\begin{aligned} \sum \sum (x_{j_i} - \bar{x}_i - \bar{x}_j + \bar{x})^2 &= \sum \sum \sum (x_{pqi} - \bar{x}_i - \bar{x}_{pq} + \bar{x})^2 \\ &= s \sum \sum (\bar{x}_{pi} - \bar{x}_i - \bar{x}_p + \bar{x})^2 \\ &+ r \sum \sum (\bar{x}_{qi} - \bar{x}_i - \bar{x}_q + \bar{x})^2 \\ &+ \sum \sum \sum (x_{pqi} - \bar{x}_{pq} - \bar{x}_{pi} - \bar{x}_{qi} + \bar{x}_p + \bar{x}_q + \bar{x}_i - \bar{x})^2 \end{aligned}$$

Antall frihetsgrader (f) deles opp på tilsvarende måte. Vi har at

$$k-1 = rs-1 = (r-1) + (s-1) + (r-1)(s-1)$$

og at

$$(k-1)(n-1) = (rs-1)(n-1) = (r-1)(n-1) + (s-1)(n-1) + (r-1)(s-1)(n-1)$$

Variansanalysen blir derfor slik som vist i Tab. H.3.

Tabell H.3

	f	Kvadratsum
Blokk	n-1	$rs \sum (\bar{x}_i - \bar{x})^2$
P	r-1	$ns \sum (\bar{x}_p - \bar{x})^2$
Q	s-1	$nr \sum (\bar{x}_q - \bar{x})^2$
PQ	(r-1)(s-1)	$n \sum \sum (\bar{x}_{pq} - \bar{x}_p - \bar{x}_q + \bar{x})^2$
PB	(r-1)(n-1)	$s \sum \sum (\bar{x}_{pi} - \bar{x}_i - \bar{x}_p + \bar{x})^2$
QB	(s-1)(n-1)	$r \sum \sum (\bar{x}_{qi} - \bar{x}_i - \bar{x}_q + \bar{x})^2$
PQB	(r-1)(s-1)(n-1)	$\sum \sum \sum (x_{pqi} - \bar{x}_{pq} - \bar{x}_{pi} - \bar{x}_{qi} + \bar{x}_p + \bar{x}_q + \bar{x}_i - \bar{x})^2$
Sum	nrs-1	$\sum \sum \sum (x_{pqi} - \bar{x})^2$

Kvadratsummen for P, Q, ... PQB dividert med antall frihetsgrader gir oss så 6 varianser: $V_P, V_Q, V_{PQ}, V_{PB}, V_{QB}$ og V_{PQB} , som vi kan bruke til testing av 3 null-hypoteser. Disse null-hypotesene er $b = 0, c = 0$ og $d = 0$. Den første ($b=0$) testes ved $F = V_P/V_{PB}$, den andre ($c=0$) ved $F = V_Q/V_{QB}$ og den siste ($d=0$) ved $F = V_{PQ}/V_{PQB}$.

De matematiske forutsetningene for disse testene er 1) at vi i modellen kan sette $u = v = t = 0$ og 2) at e_{pqi} er $N = nrs$ innbyrdes uavhengige verdier av en normal random variabel e , at $E(e) = 0$ og at $\text{var}(e)$ er uavhengig av P og Q.

I aktuelle tilfelle er disse forutsetningene ikke realistiske. Dette virker vanligvis slik at F-testen blir noe for følsom slik som beskrevet i avsnitt H.4.

Til beregning av kvadratsummene bruker vi i praksis summer i stedet for gjennomsnitt. La oss først ta for oss kvadratsummen for V_P , V_Q og V_{PQ} . Ved å summere over i (dvs. over blokkene) får vi summene S_{PQ} . Disse er da ordnet etter P_p og Q_q slik :

	P_1	P_2	...	P_r	Summer
Q_1	S_{11}	S_{21}	...	S_{r1}	
Q_2	S_{12}	S_{22}	...	S_{r2}	S_q
:	:	:		:	
Q_s	S_{1s}	S_{2s}	...	S_{rs}	
Summer		S_p			S

hvor S er summen av alle N observasjonene.

Vi har da følgende regneskjema:

$$\begin{array}{rcl}
 n \sum \sum (\bar{x}_{pq} - \bar{x})^2 & = & \frac{1}{n} \sum \sum S_{pq}^2 - S^2/N \\
 ns \sum (\bar{x}_p - \bar{x})^2 & = & \frac{1}{ns} \sum S_p^2 - S^2/N \\
 nr \sum (\bar{x}_q - \bar{x})^2 & = & \frac{1}{nr} \sum S_q^2 - S^2/N
 \end{array}$$

Sum

$$\text{Differens} \quad n \sum \sum (\bar{x}_{pq} - \bar{x}_p - \bar{x}_q + \bar{x})^2$$

På samme måte fås S_{ip} ved å summere over q (dvs. over alternativene av Q). Disse nyttes så på samme måte til beregning av $ns \sum (\bar{x}_p - \bar{x})^2$ (på ny), $rs \sum (\bar{x}_i - \bar{x})^2$ og $s \sum \sum (\bar{x}_{pi} - \bar{x}_i - \bar{x}_p + \bar{x})^2$. På tilsvarende måte brukes summene S_{qi} til beregning av $nr \sum (\bar{x}_q - \bar{x})^2$ (på ny), $rs \sum (\bar{x}_i - \bar{x})^2$ (på ny) og $r \sum \sum (\bar{x}_{qi} - \bar{x}_i - \bar{x}_q + \bar{x})^2$.

Til slutt finner vi så kvadratsummen for PQB som differensen Total kv.sum - (kv.sum for B + kv.sum for P + ... + kv.sum for QB).

La oss så ta for oss et fingert eksempel med $r = s = 2$ og $n = 6$. Vi vil for illustrasjonens skyld tenke oss at P_1 og P_2 er to sorter (f.eks. byggsorter) og Q_1 og Q_2 to mengder ($Q_2 > Q_1$)

av et gjødselslag (f.eks. kvelstoff). Vi har da $N = nrs = 24$ observasjoner av, la oss si $x =$ mengde avling.

Observasjonene er:

Blokk nr.	P_1Q_1	P_2Q_1	P_1Q_2	P_2Q_2	S_i
1	42	49	62	61	214
2	30	42	48	50	170
3	31	40	47	41	159
4	43	50	53	62	208
5	29	38	44	46	157
6	43	46	52	55	196
Sum S_j	218	265	306	315	1104 (S)

Summene S_{pq} er:

	P_1	P_2	S_q
Q_1	218	265	483
Q_2	306	315	621
S_p	524	580	1104

Vi har da $S^2/N = 1104^2/24 = 50784$ og

$$n \sum \sum (\bar{x}_{pq} - \bar{x})^2 = \frac{1}{n} \sum \sum S_{pq}^2 - S^2/N = 51768,33 - 50784 = 984,33$$

$$ns \sum (\bar{x}_p - \bar{x})^2 = \frac{1}{ns} \sum S_p^2 - S^2/N = 50914,67 - 50784 = 130,67$$

$$nr \sum (\bar{x}_q - \bar{x})^2 = \frac{1}{nr} \sum S_q^2 - S^2/N = 51577,50 - 50784 = 793,50 \quad 924,17$$

$$n \sum \sum (\bar{x}_{pq} - \bar{x}_p - \bar{x}_q + \bar{x})^2 = 60,16$$

Vi bruker så S_{ip} og S_{iq} slik som vist foran og får følgende kvadratsummer:

$$\begin{aligned} \sum\sum\sum (x_{ipq} - \bar{x})^2 &= \sum\sum\sum x_{ipq}^2 - S^2/N = && 1898,00 \\ rs\sum(\bar{x}_i - \bar{x})^2 &= && 792,50 \\ ns\sum(\bar{x}_p - \bar{x})^2 &= && 130,67 \\ nr\sum(\bar{x}_q - \bar{x})^2 &= && 793,50 \\ n\sum(\bar{x}_{pq} - \bar{x}_p - \bar{x}_q + \bar{x})^2 &= && 60,16 \\ s\sum(\bar{x}_{pi} - \bar{x}_i - \bar{x}_p + \bar{x})^2 &= && 32,83 \\ r\sum(\bar{x}_{qi} - \bar{x}_i - \bar{x}_q + \bar{x})^2 &= && 38,00 && 1847,66 \\ \hline \sum\sum\sum (x_{pqi} - \bar{x}_{pq} - \bar{x}_{pi} - \bar{x}_{qi} + \bar{x}_p + \bar{x}_q + \bar{x}_i - \bar{x})^2 &= && 50,34 \end{aligned}$$

De 6 kvadratsummene vi har bruk for, med antall (f) frihetsgrader, er stilt sammen i Tab. H.4.

Tabell H.4			
	Kv.sum	f	V
P	130,67	1	130,67
Q	793,50	1	793,50
PQ	60,16	1	60,16
PB	32,83	5	6,57
QB	38,00	5	7,60
PQB	50,34	5	10,07

De tre F-verdiene blir så
 for test av $b = 0$: $F = 130,67/6,57 = 19,90$
 - " - $c = 0$: $F = 793,50/7,60 = 104,41$
 - " - $d = 0$: $F = 60,16/10,07 = 5,97$

Velger vi å bruke 5% nivået, ser vi av Tabell III at det kritiske område for F som svarer til 1 og 5 frihetsgrader, er avgrenset til venstre av 6,61. Vi forkaster derfor null-hypotesene $b = 0$ og $c = 0$.

I tilfelle der $r = s = 2$ er det ikke vanskelig å se hvilke kontraster det kan være tale om. Det kan imidlertid også da være

flere sett av kontraster. Det ene av disse er $(b_2 - b_1)$, $(c_2 - c_1)$ og $(d_{11} + d_{22} - d_{12} - d_{21}) = (d_{22} - d_{21}) - (d_{12} - d_{11})$. Hvis P_1 og P_2 er to sorter og Q_1 og Q_2 to mengder av et gjødselslag ($Q_2 > Q_1$), ser vi at den siste kontrasten er et uttrykk for den ulikhet det er mellom de to sortene i den effekt økningen av gjødselmengden fra Q_1 til Q_2 har.

Vi har observasjoner av alle tre kontrastene for hver blokk. Setter vi

$$y_{1i} = x_{22i} + x_{21i} - x_{12i} - x_{11i}$$

$$y_{2i} = x_{22i} + x_{12i} - x_{21i} - x_{11i}$$

og
$$y_{3i} = x_{22i} + x_{11i} - x_{21i} - x_{12i}$$

vil vi finne at

$$E(y_1) = 2(b_2 - b_1)$$

$$E(y_2) = 2(c_2 - c_1)$$

og

$$E(y_3) = d_{11} + d_{22} - d_{21} - d_{12}$$

Det følger videre av dette at \bar{y}_1 , \bar{y}_2 og \bar{y}_3 er forventningsrette estimatorene av de tre kontrastene.

Konfidensgrensene for kontrasten $(b_2 - b_1)$ er da lik

hvor
$$s^2 = \frac{1}{n-1} \sum (y_{1i} - \bar{y}_1)^2$$

og der a tas ut av Tabell I for den P -verdi vi ønsker å bruke og med $(n-1)$ frihetsgrader.

Nøyaktig samme teknikk brukes for de to andre kontrastene. Konfidensgrensene for den siste kontrasten er $\bar{y}_3 \pm as_3/\sqrt{n}$.

Settes konfidenssannsynligheten til $Q = 0,95$, finner vi for vårt eksempel at konfidensgrensene er :

$$\begin{aligned} \text{for } b_2 - b_1 &: \frac{1}{2}(9,33 \mp 2,571.2,09) = \begin{cases} 1,98 \\ 7,35 \end{cases} \\ \text{for } c_2 - c_1 &: \frac{1}{2}(23,00 \mp 2,571.2,24) = \begin{cases} 8,62 \\ 14,38 \end{cases} \\ \text{og for } d_{11} + d_{22} - d_{12} - d_{21} &: 6,33 \mp 2,571.2,59 = \begin{cases} -0,32 \\ 12,98 \end{cases} \end{aligned}$$

Brukes signifikansnivået $P = 0,05$ er det i dette eksemplet ikke påviselig samspill mellom de to faktorene. Dette betyr imidlertid ikke at det ikke finnes slikt samspill, det betyr bare at vi ikke har vært i stand til å påvise det. La oss nå forutsette at det er samspill mellom P og Q og undersøke hva det betyr. Vi vil da først tenke oss at P_1 og P_2 er to sorter og at Q_1 og Q_2 er to mengder ($Q_1 < Q_2$) av et gjødselslag.

Fra tabellen over summene S_{pq} finner vi ved divisjon med $n=6$ at gjennomsnittene \bar{x}_{pq} er:

	P_1	P_2
Q_1	36,7	44,2
Q_2	51,0	52,7

Den økning i gjennomsnittene vi registrerer for hver av de to sortene når Q økes fra Q_1 til Q_2 er

$$P_1 : 51,0 - 36,7 = 14,3$$

$$P_2 : 52,7 - 44,2 = 8,5$$

Forskjellen er altså noe større for P_1 enn for P_2 . Disse differensene er nå estimater av $(d_{12} - d_{11})$ og $(d_{22} - d_{21})$. Og regner vi med et påviselig PQ-samspill, betyr det derfor at $d_{12} - d_{11} > d_{22} - d_{21}$. Vi kan si at det betyr at sort P_1 reagerer sterkere på økningen av gjødselmengden enn P_2 gjør.

Tenker vi oss at P_1 og P_2 er to mengder ($P_2 > P_1$) av f.eks. kvelstoff-gjødsel og Q_1 og Q_2 to mengder ($Q_2 > Q_1$) av kalium-gjødsel, vil samspillet bety at den større mengde kvelstoff (altså P_2) demper

effekten av en økning av kalium-gjødsel fra Q_1 til Q_2 i sammenligning med den mindre mengden kvelstoff (P_1).

H.6. Split-plot i blokkforsøk.

La oss tenke oss samme situasjon som vi behandlet i foregående avsnitt, at vi vil utføre et blokkforsøk med to faktorer P_p ($p = 1, 2, 3, \dots, r$) og Q_q ($q = 1, 2, 3, \dots, s$). Men vi vil nå tenke oss at det er knyttet større interesse til estimeringen av kontrastene mellom P 'ene og samspillkontrastene enn til kontrastene mellom Q '-ene. Det kan til eksempel tenkes at en er lite interessert i alternative kvelstoff-gjødslinger og at hovedinteressen gjelder alternative kalium-gjødslinger og samspillet. Vi kan da utføre forsøket på en slik måte at vi får en bedre sammenligning mellom P_p enn mellom Q_q .

I et feltforsøk etter blokkplanen kan vi ofte oppnå dette ved først å dele blokkene i s hovedruter (el. storruter) og hver hovedrute i r delruter (el. småruter). Vi fordeler så de s Q_q på random måte på hovedrutene og de r P_p på random måte på delrutene innen hver hovedrute.

En slik forsøksplan kan vi også bruke når alternativene av q -faktoren (Q_q) krever større forsøksenheter enn alternativene av P -faktoren (P_p). La oss tenke oss at Q_q er s alternative måter å bearbeide jorda på og P_p er r alternative gjødslinger eller r sorter. Det er da mulig at ruter på f.eks. 10 m^2 er store nok for P_p , men at ruter på denne størrelsen er for små til at en kan komme til med redskapene for jordbearbeidingen. Da kan vi bruke split-plot planen og de store rutene eller hovedrutene til Q_q og de små rutene eller delrutene til P_p . Er P_p $r = 5$ sorter og vi bruker ruter på 10 m^2 til disse, vil rutene til Q_q bli på 50 m^2 .

Et annet eksempel har vi når Q_q er s valte temperaturer i et veksthusforsøk. Av bygningstekniske grunner kan det da bli nødvendig å bruke veksthus som hovedruter. Innen hvert veksthus kan en så dele opp i delruter til P_q .

Variansanalysen blir den samme som den vi gav en beskrivelse av i foregående avsnitt. Er n antallet av blokker ($i = 1, 2, \dots, n$), vil vi også nå få $N = nrs$ observasjoner av f.eks. mengde avling. Ved å summere over i (dvs. over blokkene) får vi rs summer S_{pq} som ordnes etter P_p og Q_q . Ved å summere over q , får vi nr summer S_{pi} ordnet etter B_i og P_p . Og ved å summere over p , får vi ns summer S_{qi} ordnet etter B_i og S_q . Vi får derfor også de samme seks variansene: V_Q , V_P , V_{PQ} , V_{QB} , og V_{PQB} .

I modellen i foregående avsnitt må vi imidlertid gjøre en viktig forandring. Leddet e_{pqi} må nå erstattes av $e'_{qi} + e''_{pqi}$. Her betyr da e'_{qi} den random variasjon som skyldes ulikheten mellom hovedrutene innen blokkene og e''_{pqi} den random variasjon som skyldes ulikheten mellom delrutene innen hovedrutene og blokkene. En må regne med den muligheten at $\text{var}(e')$ kan være forskjellig for de forskjellige Q_q og at $\text{var}(e'')$ kan være forskjellig for de forskjellige kombinasjoner av P_p og Q_q .

Også i dette tilfelle er det vel som oftest bestemte kontraster en er interessert i å estimere. Det er kontraster mellom Q_q og det er kontraster mellom P_p som er utpekt på forhånd. Vi har imidlertid ikke tid til å gå noe nærmere inn på dette. Vi må nøye oss med å nevne til slutt at null-hypotesen $b = 0$, $c = 0$ og $d = 0$ også i dette tilfelle testes ved hjelp av varianskvotientene $F = V_Q/V_{QB}$, $F = V_P/V_{PB}$ og $F = V_{PQ}/V_{PQB}$.

Til grunn for de kvadratsummer og varianser som er vist i Tab. H.5 hadde en $N = 60$ observasjoner av mengde kornavling fra et faktorielt forsøk med $s = 3$ gjødslinger (Q_q) og $r = 5$ vårkveite-sorter (P_p). De tre gjødslingsalternativene var fordelt på hovedrutene og var Q_1 uten kvelstoffgjødning, $Q_2 = 200$ kg salpeter og $Q_3 = 400$ kg salpeter. Sortene var fordelt på delrutene. Antall blokker var $n = 4$.

Tabell H.5

	Kv.sum	f	V
B (blokk)	28,40	3	
Q	39,30	2	19,65
P	178,38	4	44,59
PQ	2,63	8	0,33
QB	6,17	6	1,03
PB	10,07	12	0,84
PQB	6,68	24	0,28
Sum	271,63	59	

Vi har så $F = V_Q/V_{QB} = 19,65/1,03 = 19,1$, $F = V_P/V_{PB} = 44,59/0,84 = 53,2$ og $F = V_{PQ}/V_{PQB} = 0,33/0,28 = 1,18$. Sammenligner vi med Tabell III vil vi se at null-hypotesen $b = 0$ og $c = 0$ forkastes på 5% nivået, derimot ikke null-hypotesen $d = 0$. Det er m.a.o. påviselig ulikhet mellom effekten av de tre gjødslingene og mellom de 5 sortene, men det er ikke påviselig samspill mellom gjødsling og sort.

H.7. Andre planer for lokale forsøk.

Arbeid med utvikling av planer for forsøk ble for alvor tatt opp i 1930-årene. Professor R. A. Fisher var da i flere år knyttet til Rothamsted Experimental Station, og i denne tiden utformet han prinsippene for det de fleste i dag forstår med forsøksmetodikk. Hans bok "The Design of Experiments" som ble utgitt i 1935, har vært betraktet som grunnleggende.

Fisher oppfattet ikke forsøksmaterialet som tilfeldig, slik vi har gjort i de foregående avsnittene. Heller ikke regnet en den gangen med samspill mellom forsøksledd og heterogenitetsfaktorene. Modellene ble derfor enklere enn det vi har operert med.

Siden da har det vært arbeidet intenst med å utvikle nye planer for forsøk. Dels har en vært opptatt med planer som tar sikte på å øke presisjonen for sammenligningen mellom forsøksledd, og dels har en beskjeftiget seg med problemer som melder seg når antallet av forsøksledd er særlig stort. For å øke presisjonen er det foreslått bl.a. å bruke eneggede tvillingkalver som gjentak i et blokkforsøk for sammenligning av to fôringsmidler. En har også utarbeidet planer som går ut på å bruke forsøksenhetene om igjen et antall ganger. Og for å mestre problemene med stort antall forsøksledd er det utarbeidet planer hvor blokkene ikke har forsøksenheter nok til alle forsøksleddene.

Alt dette arbeidet har imidlertid dreiet seg om slike forsøk som vi foran har kalt lokale. Det er allerede nevnt at resultatene fra slike forsøk bare i unntakstilfelle er tilstrekkelige som grunnlag for veiledning for praktisk virksomhet. Likevel har lokale forsøk stor betydning i forsøksvirksomheten. I et større forskningsprogram kan alle slags forsøk være nyttige forutsatt at de er skikkelig planlagt og utført. Men de forsøk det til syvende og sist kommer an på, skal vi ta for oss i det følgende avsnitt.

H.8. Utvidede forsøk.

Et råd som gis for praktisk virksomhet, bygger alltid på en prognose. Et slikt råd er jo et utsagn som sier hvordan en skal handle i fremtiden, f.eks. neste år. Det er derfor viktig at en prøver å bli klar over hvordan et forsøk må utføres for at det skal kunne gi grunnlag for prognoser.

Vi vet at resultatet av en handling som oftest er avhengig av flere faktorer. Noen av disse kan naturligvis kontrolleres, iallfall til en viss grad. Som eksempel kan nevnes at effekten av at en sort poteter gjødsles med et visst kvantum kaliumgjødsel kan være avhengig av kvaliteten av dyrkingsjorda og av temperaturnivået i vekstperioden. Av disse faktorer kan kvaliteten av dyrkingsjorda kontrolleres ved at den kan beskrives og klassifiseres. Råd om gjødsling kan en så gi særskilt for hver av disse klassene. Det er imidlertid klart at også innen en slik klasse er det variasjon i jordkvaliteten, og vi har derfor ikke full kontroll. For klimaelementene mangler vi ennå brukbare prognoser, og dette betyr at klimafaktorene må betraktes som heterogenitetsfaktorer vi ikke har noen kontroll med. Derfor må et forsøk som tar sikte på å gi grunnlag for praktiske råd, dekke en noenlunde normal klimavariasjon.

Vi kan si at dersom et forsøk skal utføres i den hensikt å gi grunnlag for veiledning om praktiske handlinger, må det planlegges slik at ingen av de faktorer blir kontrollert som ikke kan bli eller blir kontrollert i den praktiske virksomheten. Gjelder det f.eks. plantedyrking (valg av sort, valg av gjødsling o.l.) er det derfor klart at gjentakene må dekke en viss variasjon i jordkvalitet og variasjon i klimaelementene. Og siden et råd av økonomiske grunner må gis for et geografisk område, må gjentakene dekke både geografisk variasjon og variasjon fra år til år.

I praksis kan en til et slikt forsøk skaffe seg gjentak på en av to måter: 1) Det velges et antall (n_1) lokaliteter (bygd, grend e.l.) og innen hver lokalitet et felt hvor gjentaket plasseres. De n_1 geografisk fordelte gjentak gjentas så i n_2 år, og en har dermed ialt $n = n_1 n_2$ gjentak. Siden et felt ikke må brukes om igjen må hver lokalitet være så stor at en kan velge nytt felt fra år til år. 2) Det velges fritt n_1 felter for hvert år. Etter n_2 år har en også da $n = n_1 n_2$ gjentak.

Det vil kanskje bli hevdet at den siste fremgangsmåten er den som skal brukes og, dessuten, at de n_1 feltene skal tas ut ved loddtrekning hvert år. Forutsetningen for loddtrekning er imidlertid at en har en fortegnelse eller en liste over alle aktuelle felter innen området. Hvis f.eks. forsøket går ut på sammenligning av et antall hvete sorter, måtte denne listen omfatte alle de feltene innen området som i praksis kan tenkes brukt til hvetedyrking. Slike lister mangler i de aller fleste tilfelle. I praksis vil dette bety at feltene og dermed gjentakene må tas ut på annen måte enn ved loddtrekning.

I noen tilfelle er det sikkert mulig å ta ut et sampel på n_1 lokaliteter i et område ved loddtrekning. Det er sikkert også mulig å bruke loddtrekning når en skal ta ut det feltet (innen en lokalitet) hvor gjentaket skal plasseres i et bestemt år. Det som til slutt blir avgjørende for om resultatene kan legges til grunn for veiledning, er imidlertid om heterogeniteten i forsøksmateriale er tilstrekkelig representativ. Den bør jo være representativ for den heterogenitet en må regne/i ^{med} fremtiden.

Det er derfor ikke lett å ta standpunkt til hvilken av de to fremgangsmåtene som bør brukes. Det en kan si er at den planen som

bygger på valg av lokaliteter og felter innen lokaliteter, er den enkleste å bruke i praksis. På den annen side er det rimelig å regne med at den andre planen vil gi et mer representativt forsøksmateriale.

Sett fra et teoretisk synspunkt er det tilstrekkelig at et gjentak består av en forsøksenhet for hvert forsøksledd. I praksis hender det imidlertid at noen enheter må skjultes ut av en eller annen grunn. For å sikre seg at det i alle gjentak er observasjoner for hvert forsøksledd bør derfor hvert gjentak omfatte i det minste to forsøksenheter for hvert forsøksledd. For hvert enkelt forsøksledd vil en da ha i det minste en observasjon. I analysen er det imidlertid gjennomsnittet en bruker som observasjon. Er antall forsøksledd lik k , blir antallet av observasjoner lik $N = nk$. De metodene en skal bruke er derfor de samme som for et lokalt blokkforsøk med k forsøksledd og n gjentak, slik som beskrevet i avsnittene H.3, H.4 og I.4.

Dersom noen av forsøksenhetene må skjultes ut, vil en ha et noe varierende antall observasjoner bak disse gjennomsnittene, altså \bar{x}_{ji} ($j=1,2,\dots,k$, $i=1,2,\dots,n$). I variansanalysen kan en ta hensyn til dette, men beregningsarbeidet blir da noe mer komplisert. Effekten av et noe varierende antall har imidlertid liten betydning for resultatet, og vi skal derfor ikke komme inn på saken.

Vi har valt å bruke et plantedyrkingsforsøk som eksempel fordi vi da kan se hvilke hovedfaktorer heterogeniteten i forsøksmaterialet beror på. Situasjonen er imidlertid neppe annerledes i andre forskningssektorer.

Det er viktig at en sørger for at antallet av forsøksledd i et slikt utvidet forsøk ikke blir stort. Derfor kan lokale forsøk

bli meget nyttige fordi en på grunnlag av resultater fra disse kan skjulte ut en del forsøksledd som en regner for mindre verdifulle. En risikerer da at verdifulle ledd kan bli borte, men det kan en neppe gjøre noe ved. Viktig er det imidlertid at spørsmålet om hvilke forsøksledd en skal ha med i det utvidede forsøket, blir grundig drøftet under planleggingen.

På samme måte som i et lokalt blokkforsøk er det her samplet av gjentak som representerer det universet som eventuelle konklusjoner kan appliseres på. En blir derfor nødt til å skaffe seg et eller annet holdepunkt for å dømme om hvorvidt dette universet er generelt nok til at konklusjonene kan legges til grunn for veiledning for praktisk virksomhet. I et utvidet plantekulturforsøk er det jordkvalitet og klimaelementene som er de viktigste årsaker til heterogeniteten. Spørsmålet blir da om samplet dekker så stor variasjon i disse faktorene at heterogeniteten kan betraktes som realistisk. En kan f.eks. kreve at forsøket fortsettes i så mange år at en har fått med en noenlunde normal variasjon i temperatur og nedbør fra år til år. Viser det seg f.eks. at det har vært et relativt høyt temperaturnivå og relativt lite nedbør i de vekstsesongene som har vært med i forsøket, kan ikke resultatene av forsøket sies å gi tilstrekkelig grunnlag for veiledning for praktisk virksomhet. I et slikt tilfelle kan en ikke betrakte forsøket som avsluttet. Det må da fortsette inntil en har fått med en noenlunde normal variasjon i både temperatur og nedbørmengde. Det er også en forutsetning for å bruke resultatene som grunnlag for veiledning at samplet av gjentak dekker en noenlunde normal geografisk variasjon i jordkvaliteten. Å bringe på det rene om gjentakene tilfredstiller dette kravet, er meget vanskelig fordi en da må ta hensyn til et nokså stort antall egenskaper (fysiske og kjemiske) ved

dyrkingsjorda. I praksis blir vel derfor dette spørsmålet avgjort ved skjønn.

Da det vi nå forstår med moderne naturvitenskap ble grunnlagt i det 17. århundre, var det problemer som nærer inn under fysikken en tok fatt på. Dette har historisk sett ført til enkelte uheldige konsekvenser for biologisk forskning som ble tatt opp senere. Innen den fenomenkrets som fysikerne beskjeffiger seg med, er det nemlig en langt mer utpreget uniformitet enn det er innen den biologiske verden. Det vi her særlig sikter til er at fysikerne ikke har slike betydningsfulle faktorsamspill å ta hensyn til som biologene. En forsøkte imidlertid å etterligne fysikernes metodikk, modeller og tenkesett. Mye av denne påvirkningen henger fremdeles igjen. Ennå finnes det ikke få forskere innen biologi som mener at observasjoner er mindreverdige hvis de stammer fra forsøk hvor forsøksmaterialet er meget heterogent. Men dette er å sette saken på hodet. Heterogenitet er noe som eksisterer i den virkelige verden. Og i den grad det er mulig må derfor forskerne ta sikte på å basere sine forsøk på materialer som har en mest mulig realistisk heterogenitet. Dette er vilkåret for at en skal komme fram til virkelighetsnære beskrivelser og til data som kan legges til grunn for utarbeidelse av prognoser.

Det finnes sikkert tilfelle hvor en realistisk heterogenitet er tilsynelatende svak, tilfelle som synes å ligne meget på dem fysikerne steller med. Som eksempel kan vi nevne en undersøkelse av vekttap under lagring av goudaost, hvor forsøksleddene er et antall forskjellige behandlingsmåter. Forsøk med sikte på å finne den behandlingsmåten som gir minst vekttap, må naturligvis utføres i ostelagre som har en nærmere bestemt teknisk standard. Det kan da være fristende å velge et bestemt lager og skaffe seg gjentak

innen dette i et enkelt år. Resultatet kan en så tenke seg applisert på et univers av ostelagre som har denne bestemte tekniske standard. Men dette er neppe tilfredsstillende. Det er godt mulig at heterogeniteten ostelagre imellom er liten. En må imidlertid regne med geografisk heterogenitet i det råstoffet osten lages av. I råstoffet må en også regne med heterogenitet i tid. Dessuten har en å gjøre med heterogenitet på grunn av ulikheter mellom de personer som steller med osten i lagringsperioden. Tilsammen kan disse og andre faktorer skape en ikke ubetydelig heterogenitet og føre til samspill med forsøksfaktoren. Tar en derfor sikte på å skaffe^{seg}/et grunnlag for en konklusjon som kan appliseres på et univers av ostelagre av en viss standard, bør en planlegge forsøket etter samme prinsipp som et plantekulturforsøk.

Det finnes tilfelle hvor slike omfattende forsøk neppe er nødvendige. Men disse tilfelle er så spesielle at vi ikke skal komme inn på eksempler her.

Til slutt må vi også komme inn på spørsmålet om hvilke konsekvenser det har at en i praktisk virksomhet handler i samsvar med det resultat et slikt forsøk har ført til. La oss tenke oss at skogeierne innen et geografisk område går sammen om et forsøk som tar sikte på å finne ut om tilveksten i granskog av en bestemt bonitet blir økt - og eventuelt med hvor meget - ved at en gjødsler med et visst kvantum pr. arealenhet av en type nitrogengjødsel. Til sammenligning må en naturligvis ha med forsøksleddet "ugjødslet". Forsøket må da planlegges og utføres slik som beskrevet for et plantekulturforsøk. En må ha gjentak som representerer skogeiernes skoger og et antall år.

La oss tenke oss at observasjonene viser at gjødslingen fører til økt tilvekst og at økningen er mer enn stor nok til å kompen-

sere for kostnadene med gjødslingen. La oss videre tenke oss at skogeierne handler i samsvar med dette. Hva kan en så vente å oppnå?

Det en kan vente er at for alle skogene sett under ett, vil tilveksten øke i det minste så mye at utgiftene til gjødslingen blir dekket. Men dette betyr ikke uten videre at alle skogeierne vil vinne noe ved å gjødsle. På grunn av samspillet med heterogenitetsfaktorene vil noen skogeiere kanskje oppnå meget, andre lite, og kan hende vil noen stå seg bedre på å la være å gjødsle. Det er denne situasjonen de forskerne som gir råd om praktiske handlinger og de praktikere som må overveie om de skal følge rådet eller ikke, står overfor.

Disse forsøkene som plantekulturforskerne kaller spredte og som vi her har kalt utvidede, er av flere grunner meget vanskelige å utføre i praksis. Når plante- og jordkultur unntas, er en vel ennå ikke kommet særlig langt med å gjennomføre slike planer. Det er imidlertid neppe tvil om at denne typen forsøk vil få en dominerende plass i fremtidens forsøksvirksomhet. Dette gjelder ikke bare når målsettingen er rent praktisk. Det gjelder alle forsøk hvor målsettingen er å skaffe opplysninger om virkningen av inngrep i naturen.