Arbeidsnotater

STATISTISK SENTRALBYRÅ

Dronningensgt. 16, Dep, Oslo 1. Tlf.*(02) 41 38 20

IO 77/41

15. november 1977

TESTING I TABELLER

av

Tor Haldorsen*)

INNHOLD

*) Takk til Gunnvor Iversen og Hans Viggo Sæbø for nyttige kommentarer til manuskriptet.

1. INNLEDNING

I dette notatet vil vi se på hvordan vi kan nytte teorien om hypoteseprøving når vi skal vurdere et tallmateriale presentert i tabellform.

De vanligste tabellene ved Underavdelingen for intervjuundersøkelser viser hvordan enheter i ulike grupper fordeler seg m.h.t. en variabel. Fordelingene gis ved prosenttall (andeler). For å angi den usikkerhet slike tall er beheftet med, er det vanlig å presentere tabeller med anslag for utvalgsvarians/standardavvik. Disse tabellene viser hvordan usikkerheten varierer med prosenttall og tallet på enheter i tilsvarende spesifiserte gruppe. I publikasjonene tar en ofte også med en oppskrift på å lage konfidensintervall for prosenttallene (intervall som med stor sikkerhet dekker den korrekte verdi for vedkommende andel). Dette er en grei praksis. En opplyser om usikkerheten og overlater til leseren å avgjøre hvordan denne informasjonen kan kombineres med annen kunnskap om emnet som behandles.

I flere situasjoner må vi imidlertid selv ta stilling til den usikkerhet tallene er beheftet med. Det kan være når vi utarbeider kommentarer til tabeller som publiseres, vurderer resultatene av en forundersøkelse eller skal velge ut tabeller for publisering. Vi må da avgjøre om observerte forskjeller eller mønstre i tallmaterialet gir uttrykk for realiteter, eller om disse tendensene kan forklares ved usikkerhet alene. I denne situasjon kan hypoteseprøving være et hjelpemiddel som sammen med faglig innsikt gir grunnlag for riktige avgjørelser.

Når vi analyserer en tabell og vil bruke testing som hjelpemiddel, er det nødvendig å skille mellom to situasjoner. Hvis vi vil teste forhold som vi har oppdaget i det konkrete tallmaterialet, så er det prinsipielt en annen situasjon enn å teste hypoteser som er formulert før vi ser på materialet. I notatet behandles metoder for begge situasjoner.

Metodene som vi skal omtale, er for situasjoner der variablene er målt på nominalnivå. Vår framstilling bygger for det meste på Sverdrup (1964), (1973) og (1975) og Miller (1966).

I kapittel 2 presenteres modellen som legges til grunn for metodene. Kapittel 3 og 4 omhandler testing av hypoteser som formuleres på forhånd. I kapittel 5 beskrives en metode for å teste forskjeller/mønstre som er blitt aktuelle etter en har gransket tallmaterialet.

2. MODELL

En typisk tabell fra Underavdelingen er:

Tabell 2.1. Lønnstakere i grupper for pendlerstatus etter bostedstype. Prosent

	I alt	Bostedsstrøk				Tallet på personer
		Spredtbygd	Tettbygd, mindre enn 10 000 innbyggere	Tettbygd, 10 000-50 000 innbyggere	Tettbygd, over 50 000 innbyggere	
Dagpendlere	100	25	37	5	33	60
Uke- og langtidspendlere	100	61	28	3	8	36
Ikke fast frammøtested	100	45	24	8	24	105
Øvrige lønnstakere	100	21	27	15	37	1 114
Alle lønnstakere.	100	24	27	14	35	1 315

Denne er tatt fra Iversen (1976) og er laget ved horisontal prosentuering i tabellen:

Tabell 2.2. Lønnstakere i grupper for pendlerstatus etter bostedstype

	Bostedsstrøk				I alt
	Spredtbygd	Tettbygd, mindre enn 10 000 innbyggere	Tettbygd, 10 000-50 000 innbyggere	Tettbygd, over 50 000 innbyggere	
Dagpendlere	15	22	3	20	60
Uke- og langtidspendlere	22	10	1	3	36
Ikke fast frammøtested	47	25	8	25	105
Øvrige lønnstakere	230	304	166	414	1 114
Alle lønnstakere.	314	361	178	462	1 315

Dette tallmaterialet vil bli brukt som gjennomgangseksempel og er utgangspunkt for den modell vi behandler.

I tabell 2.1 og 2.2 har vi 4 ulike, spesifiserte grupper i forspalten og 4 kategorier av variabelen bostedsstrøk i hodet. Vi vil se på tabeller av samme slag med r grupper i forspalten og k kategorier i hodet. Vi vil se på slike tabeller som et resultat av r uavhengige multinomiske forsøksrekker med underliggende sannsynligheter.

$$\begin{array}{cccc} p_{11}, & p_{12}, & \ldots, & p_{1k} \\ p_{21}, & p_{22}, & \ldots, & p_{2k} \\ \vdots & \vdots & & \vdots \\ p_{r1}, & p_{r2}, & \ldots, & p_{rk} \end{array}$$

der $\sum_{j=1}^{k} p_{ij} = 1$ for alle i .

For å betegne tabeller som tabell 2.2 vil vi bruke symboler.

$$\begin{array}{cccc|c} X_{11} & X_{12} & \ldots & X_{1k} & n_1 \\ X_{21} & X_{22} & \ldots & X_{2k} & n_2 \\ \vdots & \vdots & & \vdots & \\ X_{r1} & X_{r2} & \ldots & X_{rk} & n_r \\ \hline X_{.1} & X_{.2} & \ldots & X_{.k} & n \end{array}$$

Uten ytterligere forutsetninger er $\hat{p}_{ij} = X_{ij} / n_i$ en naturlig estimator for p_{ij} for $i = 1,2,\ldots,r$ og $j = 1,2,\ldots,k$. Tabell 2.1 inneholder estimater for de underliggende p_{ij}-ene presentert i form av prosenter, dvs. som $\hat{p}_{ij} \times 100$.

Modellen kan brukes når enhetene i hver av forspaltegruppene er trukket som enkle tilfeldige utvalg. Vi vil også bruke den når alle enhetene er trukket som ett slikt utvalg, eller etter en mer komplisert utvalgsplan som den vi bruker i Byrået. I det siste tilfellet kan tallene være beheftet med større usikkerhet enn den som inngår i modellen. Vi regner ikke med at dette har noen betydning for testing av sammenhenger.

3. SAMMENLIGNING AV TO PROSENTTALL

Vi vil bruke tabell 2.1 som eksempel. Den viser at i materialet er andelen bosatt i spredtbygde strøk 61 prosent blant uke- og langtidspendlere og 21 prosent blant øvrige lønnstakere. Dette skal vi bruke til å uttale oss om de underliggende p_{21} og p_{41}.

Hvis vi på forhånd har bestemt at det vi vil kommentere/vurdere i tabellen er forskjellen mellom disse to andelene, kan vi gå fram på følgende måte. Vi baserer oss på observatoren

$$V = \frac{\frac{X_{21}}{n_2} - \frac{X_{41}}{n_4}}{\sqrt{\left(\frac{1}{n_2}+\frac{1}{n_4}\right) \cdot \frac{X_{21}+X_{41}}{n_2+n_4} \cdot \left(1 - \frac{X_{21}+X_{41}}{n_2+n_4}\right)}} . \qquad (3.1)$$

Vi vil bestemme et tall g $(g > 0)$ slik at vi kan påstå $p_{21} > p_{41}$ dersom V er større enn g, eller $p_{21} < p_{41}$ dersom V er mindre enn $-g$. Hvis nå i virkeligheten $p_{21} = p_{41}$, så vil vi helst unngå å påstå $p_{21} > p_{41}$ eller $p_{21} < p_{41}$. Det gir oss en mulighet til å bestemme g. Vi setter H_0: $p_{21} = p_{41}$ som vår "nullhypotese" og krever at hvis H_0 gjelder, så skal sannsynligheten være høyst ε for å påstå $p_{21} > p_{41}$ eller $p_{21} < p_{41}$.
Ut fra forutsetningene (modellen) kan nå g bestemmes. Hvis H_0 gjelder, så er V tilnærmet normalfordelt med forventning 0 og varians 1. Med $g = g_{1-\varepsilon/2}$, $1 - \varepsilon/2$ - fraktilen i normalfordelingen, så er kravet til metoden oppfylt. Med $\varepsilon = 0,1$ gir tallene i tabell 2.2

$$V = 5,78 > 1,65 \quad (= g_{1-0,1/2}) ,$$

så vi påstår $p_{21} > p_{41}$, dvs. andelen bosatt i spredtbygde strøk er større blant uke- og langtidspendlere enn blant øvrige lønnstakere.

ε er nivået for metoden. Det er sannsynligheten for å forkaste H_0 når H_0 gjelder. Denne størrelsen må vi selv bestemme. Vanlige valg er $\varepsilon = 0,05$ og $\varepsilon = 0,01$. Vi ser at jo mindre ε velges, dess større blir $g_{1-\varepsilon/2}$. Det betyr at vi minsker muligheten for at vi kommer fram til en av påstandene $p_{21} > p_{41}$ eller $p_{21} < p_{41}$ uansett hva de underliggende p_{21}, p_{41} er. Hvis det er viktig å få påpekt alle forskjeller i et materiale og ikke så alvorlig å komme med påstander om forskjeller der det ikke er noen, så bør vi velge ε "stor". I eksemplet valgte vi $\varepsilon = 0,1$.

At vi bruker betegnelsen "nullhypotese" om "$p_{21} = p_{41}$", betyr ikke at vi på forhånd tror eller mener at slik er det. Det er mer naturlig å se på den som et utgangspunkt, en basis, som vi sammen med modellen bruker for å vurdere det observerte resultat.

Hvis vi innfører p for den felles verdi som p_{21} og p_{41} har under nullhypotesen, så har vi:

$$V = \frac{\hat{p}_{21} - \hat{p}_{41}}{\sqrt{(\frac{1}{n_2} + \frac{1}{n_4})\,\hat{p}(1-\hat{p})}} .$$

$\hat{p}_{21} - \hat{p}_{41}$ har varians $\frac{1}{n_2} p_{21}(1 - p_{21}) + \frac{1}{n_4} p_{41}(1 - p_{41})$. Hvis H_0 gjelder, så er denne lik $(\frac{1}{n_2} + \frac{1}{n_4})\, p(1-p)$ som estimeres med $(\frac{1}{n_2} + \frac{1}{n_4})\, \hat{p}(1-\hat{p})$ der $\hat{p} = (X_{21} + X_{41}) / (n_2 + n_4)$. Vi ser da at observatoren V er lik differansen vi vurderer dividert på kvadratroten av den estimator vi ville bruke for variansen når H_0 gjelder.

Foran brukte vi at V under H_0 var tilnærmet normalfordelt. Dette gjelder bare hvis $n_2 p_{21}$, $n_2(1-p_{21})$, $n_4 p_{41}$ og $n_4 \cdot (1-p_{41})$ er tilstrekkelig store. Vanlig er å sikre seg dette ved å bruke tilnærmelsen bare hvis X_{21}, $n_2 - X_{21}$, X_{41} og $n_4 - X_{41}$ alle større eller lik 10. Noen tillater seg også å bruke tilnærmelsen selv om noen av de aktuelle antall observasjoner er mindre.

For gitte n_1, n_2, $\hat{p}$ og ε kan en beregne hvor stor den observerte forskjell minst må være for at vi med metoden vil erklære den signifikant, dvs. påstå at den gjelder de underliggende p-er . I tabell 3.2 har vi satt opp noen slike minsteverdier for ulike kombinasjoner av n_1, n_2, ε og $\hat{p}$.

Tabell 3.2. Minste observerte forskjell som vil erklæres signifikant

n_1	n_2	ε	$\hat{p}$ 0,1	0,3	0,5
20	20	0,01	-	0,37	0,41
		0,05	-	0,28	0,31
		0,25	-	0,17	0,18
20	100	0,01	-	0,29	0,32
		0,05	-	0,22	0,24
		0,25	-	0,13	0,14
20	1000	0,01	-	0,27	0,29
		0,05	-	0,20	0,22
		0,25	-	0,12	0,13
100	100	0,01	0,11	0,17	0,18
		0,05	0,08	0,13	0,14
		0,25	0,05	0,07	0,08
100	1000	0,01	0,08	0,12	0,14
		0,05	0,06	0,09	0,10
		0,25	0,04	0,06	0,06
1000	1000	0,01	0,03	0,05	0,06
		0,05	0,03	0,04	0,04
		0,25	0,02	0,02	0,03

I tabellen har vi utelatt noen verdier for $\hat{p} = 0,1$ p.g.a. at forventet antall observasjoner blir svært lite.

Bruker vi nivå ε=0,05 og har 100 enheter i hver av de to gruppene vi sammenligner og den gjennomsnittlige andel for disse anslås å være 0,3, så viser tabellen at forskjellen mellom andelene må være minst 0.13 for at vi skal påstå den er signifikant. Er de observerte andelene 0,25 og 0,35 så er forskjellen for liten til at vi kan påstå det er forskjell på de underliggende p-er, mens forskjellen mellom 0,22 og 0,38 er stor nok. Tabellen viser at kravet på den observerte forskjell avtar når n_1 og/eller n_2 øker, avtar når ε øker, øker når $\hat{p}$ øker mot 0,5.

Vi kan bruke tabeller som 3.2 på flere måter. Hvis det i en undersøkelse fins forspaltegrupper som går igjen i de fleste tabeller, så lager vi tabellen for disse aktuelle n_i-ene. Vi kan da som regel nøye oss med ett valg av ε, men bør ha mange verdier av $\hat{p}$. Observerte forskjeller kan da vurderes i forhold til tabellen, i stedet for å beregne observatoren (3.1) hver eneste gang. Tabell 3.2 gir også inntrykk av hvor store underliggende forskjeller må være før vi kan regne med å få påvist dem for ulike antall i forspaltgruppene. Vi kan også finne direkte uttrykk for dette hvis vi beregner styrken til metoden. For gitte n_i-er angir styrken sannsynligheten for å påstå signifikant forskjell. Styrken avhenger av de underliggende p-er. Tabell 3.3 inneholder anslag for styrken når vi har valgt ε=0,05.

Tabell 3.3 Anslag for metodens styrke når nivået er lik 0,05

n_1	n_2	p_1	p_2	Sannsynlighet for å påvise at forskjell
50	50	0,2	0,1	0,28
		0,3	0,2	0,21
		0,5	0,4	0,17
50	1 000	0,2	0,1	0,59
		0,5	0,4	0,29
1 000	50	0,2	0,1	0,39
		0,5	0,4	0,28
100	100	0,2	0,1	0,51
		0,5	0,4	0,29
500	500	0,2	0,1	0,99
		0,5	0,4	0,89
50	50	0,3	0,1	0,71
		0,6	0,4	0,52
100	100	0,3	0,1	0,95
		0,6	0,4	0,81
50	50	0,4	0,1	0,95
		0,6	0,3	0,87
100	100	0,4	0,1	1,00
		0,6	0,3	0,99

Tabellen viser at hvis vi har 100 i hver av gruppene og de underliggende p-er er 0,2 og 0,1, så er sannsynligheten omtrent 0,51 for å få påvist at det er forskjell. Vi kan lese ut av tabellen hvor store gruppene må være for at vi skal ha en rimelig sjanse til å påvise ulike forskjeller. Vi ser at det har liten hensikt å arbeide med grupper på 50 enheter hvis en søker å påvise forskjell når den underliggende forskjell er omlag 0,1.

En bør merke seg at alt som beregnes i kapittel 3 forutsetter at vi på forhånd har bestemt oss for det ene paret av andeler som skal sammenlignes. Hypotesen $p_{21} = p_{41}$ en en apriori hypotese.

4. SAMMENLIGNING AV FLERE PROSENTTALL

Ofte vil vi foreta flere enn én sammenligning i en tabell. La oss anta at vi på forhånd bestemmer oss for å sammenligne hver av de tre pendlergruppene med "øvrige ansatte" m.h.t. andel bosatt i spredtbygde strøk. Vi skal altså kommentere/vurdere p_{11} og p_{41}, p_{21} og p_{41}, p_{31} og p_{41}. Som basis for å vurdere resultatet velger vi "nullhypotesen" H_0: $p_{11} = p_{41}$, $p_{21} = p_{41}$, $p_{31} = p_{41}$. Denne er sammensatt av tre delhypoteser og kan også formuleres som H_0': $p_{11} = p_{21} = p_{31} = p_{41}$ (=p), dvs. alle fire andelene er like og er lik en ukjent p. En framgangsmåte er nå å se på observatorer som i kapittel 3 for hver av de tre delhypotesene. Vi bruker observatorene

$$V_1 = (\hat{p}_{11} - \hat{p}_{41}) / \sqrt{(1/n_1 + 1/n_4)\, \hat{p}(1 - \hat{p})}$$

$$V_2 = (\hat{p}_{21} - \hat{p}_{41}) / \sqrt{(1/n_2 + 1/n_4)\, \hat{p}(1 - \hat{p})}$$

$$V_3 = (\hat{p}_{31} - \hat{p}_{41}) / \sqrt{(1/n_3 + 1/n_4)\, \hat{p}(1 - \hat{p})} ,$$

$$\text{der } \hat{p} = (X_{11}+X_{21}+X_{31}+X_{41}) / (n_1+n_2+n_3+n_4)$$

$$= X_{.1}/n$$

som er en naturlig estimator for den felles verdi p under H_0. Når vi setter inn de observerte verdier, velger vi konstant g og ser om $V_i < -g$ eller $V_i > g$ for i = 1, 2, 3, og kommer med påstander om de underliggende

p_{i1}-er etter resultatet av denne sammenligningen. Får vi f.eks. $V_1 < -g$, $-g < V_2 < g$ og $V_3 > g$, vil vi påstå $p_{11} < p_{41}$ og $p_{31} > p_{41}$. Det er fristende å velge g som i kapittel 3, dvs. sette $g = g_{1-\varepsilon/2}$, $(1 - \varepsilon/2)$-fraktilen i normalforedelingen. Da H_0 nå består av 3 delhypoteser, kan vi regne ut at hvis H_0 gjelder, så gir metoden sannsynlighet mindre eller lik <u>3 ε</u> for å komme med en eller flere feilaktige påstander. Hvis vi vil at metoden skal gi sannsynlighet høyst lik ε for å komme med en eller flere feilaktige påstander når H_0 gjelder, så må vi velge $g = g_{1-\varepsilon/2.3}$, ($1 - \varepsilon/6$ - fraktilen i normalfordelingen). ε må vi velge ut fra vurderinger som nevnt i kapittel 3.

Gjennomfører vi dette med tallene fra tabell 2.1, får vi:

$$V_1 = (0{,}25 - 0{,}21) \,/\, \sqrt{(1/60 + 1/1114) \cdot 0{,}24 \cdot (1 - 0{,}24)} = 0{,}71$$

$$V_2 = (0{,}61 - 0{,}21) \,/\, \sqrt{(1/36 + 1/1114) \cdot 0{,}24 \cdot (1 - 0{,}24)} = 5{,}53$$

$$V_3 = (0{,}45 - 0{,}21) \,/\, \sqrt{(1/105 + 1/1114) \cdot 0{,}24 \cdot (1 - 0{,}24)} = 5{,}50$$

Når vi velger $\varepsilon = 0{,}1$, får vi $g = g_{1-0,1/6} = g_{0,9833} = 2.13$. Vi har altså $-g < V_1 < g$, $V_2 > g$ og $V_3 > g$, så konklusjonen blir $p_{21} > p_{41}$ og $p_{31} > p_{41}$.

Vi baserer oss også her på normalfordelingen. Hvis vi ser bort fra denne tilnærmingen, så vet vi allikevel bare at sannsynligheten er høyst ε for å komme med en eller flere feilaktige påstander når H_0 gjelder. Den eksakte verdi av sannsynligheten, det eksakte nivå, kjenner vi ikke, den vil variere med avhengigheten mellom V_1, V_2 og V_3. I eksemplet vil det eksakte nivå være ubetydelig mindre enn ε.

Tilsvarende tabell 3.2, kan en også lage en tabell som viser minste observerte forskjell som vil bli erklært signifikant, når en foretar 3 parvise sammenligninger. En slik tabell vil vise at de observerte forskjellene må være større enn de i tabell 3.2. Nå vil noen kanskje mene at allerede forskjellene som kreves i tabell 3.2 er nokså store, og det er naturlig å spørre om nødvendigheten av å foreta en simultan vurdering av de tre delhypotesene. Et argument for dette er at det gir den beste beskrivelse av hva vi av og til gjør når vi vurderer en tabell. Vi ser på de observerte tallene og velger ut de forskjeller som på en eller annen måte virker slående, ofte er det forskjeller som vi ser er store. Nå er det slik at med V_1, V_2 og V_3 som definert først i kapitlet, så er under H_0 hver av disse tilnærmet normalfordelt med forventning 0 og varians 1. Men den største av disse, $\max(V_1, V_2, V_3)$, har forventning på om lag 0,85. Derfor får vi ikke nivå ε hvis vi ser på de tre forskjellene, velger ut den største og sammenligner med $g_{1-\varepsilon/2}$. Dess flere grupper vi har i forspalten, dess større blir sannsynligheten under H_0 for at den største V-observatoren blir større enn $g_{1-\varepsilon/2}$.

Så lenge sammenligningene er spesifisert på forhånd, kan vi bruke tilsvarende metoder for å gjennomføre et større antall sammenligninger i en tabell. Vi skal komme nærmere inn på dette siden, men vil først presentere en metode som ikke krever at vi a priori spesifiserer hvilke sammenligninger vi vil gjøre.

5. TESTING AV KONTRASTER

I dette kapitlet skal vi behandle en metode som tillater oss å teste forskjeller og mønstre som vi først har funnet interessante etter å ha studert tallmaterialet. For en mer utførlig og generell framstilling henviser vi til Sverdrup (1975).

Vi bruker fortsatt modellen fra kapittel 2 og velger oss en "nulltilstand" for å vurdere resultatene. Vi bruker H: $p_{ij} = p_j$, for $i = 1,\ldots,r$ og $j = 1,\ldots,k$, dvs. vi antar at for alle gruppene i forspaltene så vil enhetene fordele seg etter én og samme fordeling m.h.t. variabelen i hodet. Sagt på en annen måte, de underliggende p-er er like for de r gruppene. Metoden gjør det mulig å teste om <u>lineære kontraster</u> er signifikant større enn 0. Lineære kontraster er uttrykk av typen
$f = \sum_{i=1}^{r} \sum_{j=1}^{k} f_{ij} p_{ij}$ der f_{ij} er konstanter som tilfredsstiller $\sum_{i=1}^{r} f_{ij} = 0$ for alle j. Det er altså en lineærkombinasjon der p_{ij}-ene er multiplisert med ulike f_{ij}-er. f_{ij}-ene må være slik at "for hver kolonne" må f_{ij}-ene summere seg til 0.

Et eksempel er $f = p_{21} - p_{41}$, her er $f_{21} = 1$ og $f_{41} = -1$ og de andre f_{ij}-ene lik 0. Et annet eksempel er $f = \frac{1}{3}(p_{11} + p_{21} + p_{31}) - p_{41}$, denne kontrasten gir oss en sammenligning mellom gjennomsnittlig andel for pendlergruppene og andel for "øvrige lønnstakere". Vil vi veie pendlergruppene relativt til antall personer, kan vi se på kontrasten
$f = \frac{n_1}{n_1 + n_2 + n_3} p_{11} + \frac{n_2}{n_1 + n_2 + n_3} p_{21} + \frac{n_3}{n_1 + n_2 + n_3} p_{31} - p_{41}$. Med en kontrast som $f = (p_{21} + p_{22}) - (p_{41} + p_{42})$ kan vi sammenligne to av gruppene når kategoriene 1 og 2 i hodet er slått sammen.

Setter vi inn de observerte andelene, $\hat{p}_{ij}$-ene, for p_{ij}-ene i kontrasten, får vi en observert kontrast, $\hat{f} = \sum_i \sum_j f_{ij} \hat{p}_{ij}$. Denne vil ha en viss varians, estimatoren for denne under H betegnes "$\mathrm{est}_H \operatorname{var} \hat{f}$". (Vi skal senere vise hva denne blir.)

Metoden er nå: For kontraster vi finner interessante (muligens først etter å ha sett materialet), påstår vi

$$\sum_i \sum_j f_{ij} p_{ij} > 0 \quad \text{hvis}$$

$$\frac{\sum_i \sum_j f_{ij} \hat{p}_{ij}}{\sqrt{\mathrm{est}_H \operatorname{var} \hat{f}}} > \sqrt{Z_{1-\varepsilon}}$$

der $Z_{1-\varepsilon}$ er $(1-\varepsilon)$-fraktilen i kji-kvadratfordelingen med $(r-1)\cdot(k-1)$ frihetsgrader. Metoden har den egenskap at når antall observasjoner er stort, så vil sannsynligheten være høyst ε for å komme med en feilaktig påstand. Merk at sannsynligheten ε ikke gjelder en enkel kontrast, men gjelder simultant for alle kontraster vi kunne komme til å vurdere.

For å kunne bruke metoden må vi finne $\text{est}_H \text{var}\, \hat{f}$ for de ulike kontrastene vi vurderer. For en gitt estimert kontrast, $\sum_i \sum_j f_{ij} \hat{p}_{ij}$ der $\hat{p}_{ij} = X_{ij}/n_i$, har vi:

$$\text{var}\,(\sum_i \sum_j f_{ij} \hat{p}_{ij}) = \sum_i \text{var}\,(\sum_j f_{ij} \hat{p}_{ij})$$

$$= \sum_i \frac{1}{n_i} \,(\sum_j f_{ij}^2\, p_{ij}(1-p_{ij}) - 2 \sum_{j<k} f_{ij}\, f_{ik}\, p_{ij}\, p_{ik}).$$

Når vi er i "nulltilstanden", dvs. at $p_{ij} = p_j$ for $i = 1,\ldots,r$ og $j = 1,\ldots,k$, har vi:

$$\text{var}_H(\sum_i \sum_j f_{ij} \hat{p}_{ij}) = \sum_i \frac{1}{n_i} (\sum_j f_{ij}^2\, p_j(1-p_j) - 2 \sum_{j<k} f_{ij}\, f_{ik}\, p_j\, p_k).$$

En estimator for denne får vi ved å sette inn $\hat{p}_j = X_{\cdot j}/n$ for p_j-ene, så

$$\text{est}_H \text{var}\, \hat{f} = \sum_i \frac{1}{n_i} (\sum_j f_{ij}^2\, \hat{p}_j(1-\hat{p}_j) - 2 \sum_{j<k} f_{ij}\, f_{ik}\, \hat{p}_j\, \hat{p}_k).$$

Vi skal bruke tabell 2.1 til å gi noen eksempler. I denne er $r = 4$ og $k = 4$, så $(r-1)\cdot(k-1) = 9$. Velger vi som før $\varepsilon = 0{,}1$, finner vi av en tabell over kji-kvadratfordelingen at $\sqrt{Z_{1-\varepsilon}} = \sqrt{Z_{0,9}} = \sqrt{14.684} =$ 3,83. Det betyr at kontraster erklæres signifikant større enn 0 hvis den observerte kontrast normert med standardavviket er større enn 3,83.

I kapittel 3 så vi på forskjellen mellom p_{21} og p_{41}. Nå er $f = p_{21} - p_{41}$ en kontrast og $\text{est}_H \text{var}\, \hat{f} = (1/n_2 + 1/n_4)\, \hat{p}_1\, (1-\hat{p}_1)$. Vi får

$$\frac{\hat{p}_{21} - \hat{p}_{41}}{\sqrt{\text{est}_H \text{var}\, \hat{f}}} = \frac{0{,}4}{0{,}072} = 5{,}53 > 3{,}83,$$

så metoden gir $p_{21} - p_{41}$ signifikant større enn 0. Merk at vi her ikke behøvde å formulere noen hypotese om forholdet mellom p_{21} og p_{41} på forhånd. Metoden garanterer at nivået er mindre eller lik 0,1, selv om vi først etter å ha observert den store differansen på 0,4 , bestemmer oss for å teste forskjellen.

Noen ganger vil vi sammenligne for ulike aggregerte nivåer av inndelingene vi bruker. Det kan være at vi etter å ha sett på tabell 2.1, finner ut at vi vil betrakte de tre pendlergruppene samlet i forhold til "øvrige lønnstakere". Metoden gjør det mulig å foreta slike sammenligninger med bakgrunn i den opprinnelige modell. La oss f.eks. fortsatt se på andelen bosatt i spredtbygd strøk. Aktuelt kan da være å se på kontrasten
$f = \frac{n_1}{n_1+n_2+n_3} p_{11} + \frac{n_2}{n_1+n_2+n_3} p_{21} + \frac{n_3}{n_1+n_2+n_3} p_{31} - p_{41}$ der pendlergruppene veies med relativt antall i gruppene. Vi får

$$\text{est}_H \operatorname{var} \hat{f} = \left(\frac{1}{n_1+n_2+n_3} + \frac{1}{n_4}\right) \hat{p}_1(1-\hat{p}_1) \quad \text{og}$$

$$\hat{f} / \sqrt{\text{est}_H \operatorname{var} \hat{f}} = 0{,}21 / 0{,}033 = 6{,}42 > 3{,}83 ,$$

så forskjellen er signifikant.

Pendlergruppene kan også gis samme vekt. Da ser vi på kontrasten
$f = \frac{1}{3}(p_{11}+p_{21}+p_{31}) - p_{41}$. For denne finner vi

$$\text{est}_H \operatorname{var} \hat{f} = \left[\frac{1}{9}\left(\frac{1}{n_1} + \frac{1}{n_2} + \frac{1}{n_3}\right) + \frac{1}{n_4}\right] \cdot \hat{p}_1 \cdot (1-\hat{p}_1) \quad \text{og}$$

$$\frac{\frac{1}{3}(\hat{p}_{11}+\hat{p}_{21}+\hat{p}_{31}) - \hat{p}_{41}}{\sqrt{\text{est}_H \operatorname{var} f}} = \frac{0{,}23}{0{,}035} = 6{,}57 > 3{,}83 ,$$

så også målt på denne måten er forskjellen signifikant.

På tilsvarende måte kan vi også formulere kontraster som innebærer at vi "slår sammen" kategorier i hodet. La oss anta vi så på kontrastene $p_{11} - p_{41}$ og $p_{12} - p_{42}$ og fant

$$\frac{\hat{p}_{11} - \hat{p}_{41}}{\sqrt{\text{est}_H \operatorname{var} \hat{f}}} = 0{,}707 \quad \text{og}$$

$$\frac{\hat{p}_{12} - \hat{p}_{42}}{\sqrt{\text{est}_H \operatorname{var} \hat{f}}} = 1{,}700 .$$

Vi kan ikke påstå signifikant forskjell. Men vi ser avvikene går samme vei for begge kategoriene, da kan vi "slå sammen" kategoriene, dvs. se på kontrasten $f = (p_{11} + p_{12}) - (p_{41} + p_{42})$. Vi finner

$$\begin{aligned} \mathrm{var}_H \hat{f} &= \frac{1}{n_1} (p_1(1-p_1) + p_2(1-p_2) - 2\,p_1\,p_2) \\ &\qquad + \frac{1}{n_4} (p_1(1-p_1) + p_2(1-p_2) - 2\,p_1\,p_2) \\ &= (\frac{1}{n_1} + \frac{1}{n_4}) (\,(p_1+p_2)(1-(p_1+p_2))\,). \end{aligned}$$

Da blir

$$\hat{f} \,/ \sqrt{\mathrm{est}_H \mathrm{var}\, \hat{f}} = 0{,}14 \,/\, 0{,}066 = 2{,}11 .$$

Nå er $2{,}11 < 3{,}83$ så vi kan heller ikke påstå at denne kontrasten er signifikant større enn 0 , men det er verdt å merke seg at vi oppnådde en større verdi på observatoren og at metoden garanterer oss et bestemt nivå ε (= 0,1 i vårt eksempel) når vi foretar slike "manipulasjoner" etter vi har sett på observasjonsmaterialet.

Nå kan det være tidkrevende å lete i et materiale etter signifikante kontraster, men Sverdrup (1975) gir også en regel for å avgjøre om det overhodet finnes signifikante kontraster i materialet. En tester homogenitetshypotesen, H: $p_{ij} = p_j$ for alle (i, j) , med observatoren

$$Z = \sum_{i,j} \frac{(X_{ij} - n_i \cdot \frac{X_{\cdot j}}{n})^2}{n_i \cdot \frac{X_{\cdot j}}{n}} .$$

Hvis Z større enn $Z_{1-\varepsilon}$, $(1-\varepsilon)$ - fraktilen i kji-kvadratfordelingen med $(r-1)\cdot(k-1)$ frihetsgrader, forkastes hypotesen, Sverdrup (1964), s. 219. Hvis en i et materiale ikke oppnår forkastning med denne testen, så finnes det ingen signifikante kontraster i materialet. Får vi forkastning, finnes det signifikante kontraster, men det er selvsagt ikke gitt at dette er de kontrastene vi finner interessante.

Vi har hele tiden hatt homogenitetshypotesen som "nulltilstand" og sett på kontraster relativt til denne. I Sverdrup (1975) finnes eksempler med andre valg av "nulltilstand" og dermed andre mengder av kontraster som kan vurderes. Der fins også eksempler der det demonstreres hvordan en ved hjelp av kontrastene kan undersøke sammenhengen mellom resultatene og ekstern informasjon.

6. SAMMENLIGNING AV METODENE I KAPITTEL 4 OG 5

Selv om metoden i kapittel 5 gir oss mange muligheter, så har den sine ulemper. Metoden kan ha liten styrke, dvs. sannsynligheten er liten for å avsløre signifikans. Dette er en naturlig konsekvens av at den sikrer oss et nivå uansett hvilken kontrast vi tester. Sverdrup (1975, s. 17) gir eksempel på hvordan en kan begrense mengden av kontraster og dermed oppnå økt styrke for de kontraster som er av interesse.

Økt styrke kan det også gi å bruke metoden fra kapittel 4. Vi spesifiserer på forhånd en romslig mengde lineære kontraster som omfatter alle vi overhodet kan tenke oss å undersøke. La oss si det er n stykker i alt. For kontrastene beregner vi

$$V_f = \frac{\sum_i \sum_j f_{ij} \hat{p}_{ij}}{\sqrt{\text{est}_H \operatorname{var} \hat{f}}}$$

og sammenligner med $\pm g_{1-\varepsilon/2 \cdot n}$. Vi er da sikret nivå ε. Siden $V_{\hat{f}}$ blir den samme ved begge metoder, blir metoden fra kapittel 4 sterkest hvis $g_{1-\varepsilon/2\cdot n} < \sqrt{Z_{1-\varepsilon}}$. Vi skal bruke tabell 2.1 som eksempel. Der er tre pendlergrupper og en stor gruppe øvrige lønnstakere. Hvis vi på forhånd bestemte oss for å bruke den sistnevnte som referansegruppe og foreta separate sammenligninger mellom andelene for denne og pendlergruppene, så blir det tre sammenligninger for hver kategori i hodet. Vi ser her for eksemplets skyld bort fra at det er en slags ordning mellom kategoriene i hodet og vil også sammenligne andelene når vi slår sammen kolonnene 1 og 2, 1 og 3, 1 og 4. Det gir i alt $3 \times (4+3) = 21$ forhold som skal vurderes. Med $\varepsilon = 0{,}1$ finner vi at med metoden i kapittel 4 skal observatorene sammenlignes med $\pm g_{1-0,1/2\cdot 21} = \pm g_{0,99762} = \pm 2{,}82$, metoden i kapittel 5 gir 3,83. Siden observatorene er de samme i begge tilfelle, gir metoden fra kapittel 4 størst styrke for kontrastene av interesse.

Nå vil en i Iversen (1976) også finne eksempler på at pendlergruppene sammenlignes innbyrdes. Hvis vi også på forhånd skal ta dette i betraktning, blir det $\binom{4}{2} \times (4+3) = 42$ forhold som vi muligens vil vurdere. Da må observatorene sammenlignes med $\pm g_{1-0,1/2\cdot42}$ = $\pm g_{0,99881} \approx \pm 3{,}04$.

Det synes derfor som metoden i kapittel 4 er å foretrekke hvis en på forhånd kan angi et moderat antall sammenligninger en vil foreta i tabellen. Overslag over hva som gir størst styrke kan en gjøre i det enkelte tilfelle. Men det er vesentlig at slike overslag gjøres på forhånd, har en først gitt seg i kast med observasjonsmaterialet, er det bare metoden i kapittel 5 som sikrer at en opererer med et spesifisert nivå.

7. BRUK AV METODENE

Vi vil antyde noen poenger i forbindelse med bruk av metodene på Underavdelingen.

Det kreves en viss regnemessig innsats. En har i dag ingen mulighet til å få utført noen av testene i samband med utkjøring av tabeller med standardprogrammer (STP). Et minstekrav for å bruke metoden fra Sverdrup (1975) er at tabellprogrammet gir observatoren for homogenitetstesten. Tabellprogrammene i DDPP vil gi denne. Men skal det bli noen praktisk løsning, så er det nødvendig at en ellers også finner det lønnsomt å kjøre det meste av tabellene i en undersøkelse med DDPP.

Så lenge en holder seg til relativt enkle kontraster, er det lett å utføre beregningene på en lommekalkulator. Er denne programmerbar, som en HP-25, vil regnearbeidet etter hvert innskrenke seg til å trykke inn de aktuelle tall fra tabellen.

Mange prosjekter på Underavdelingen behandler komplekse sosiale prosesser. De omfatter problemer der mange variable inngår. Noen av disse variable er ikke direkte observerbare, i analysen tar en hensyn til dem ved å bringe inn indikatorer og andre observerbare variable som en antar har sammenheng med de førstnevnte. Praksis i dag er å avbilde disse forhold med nokså enkle tabeller der kanskje bare to og tre variable inngår av gangen. Signifikans eller ikke signifikans i slike tabeller er ikke så greie å tolke som det kan synes av de metoder som er omtalt. En god kommentar til slike tabeller må trekke inn teori og kunnskap om andre variable enn de som inngår i tabellene. Vi anbefaler derfor at en kun ser på metodene i dette notatet som veiledende.

Det ideelle ville være at saksbehandleren på forhånd brukte teori og kunnskaper til å formulere en mer spesifisert modell. En ville da oppnå bedre estimater og kunne bruke testing som et mer direkte hjelpemiddel. For mange prosjekter er dette neppe gjennomførbart.

Før de tar fatt på kommentararbeidet bør saksbehandlerne lage seg en tabell som tabell 3.2. Dette uansett om de baserer den på at en, flere eller alle kontraster i tabellene skal testes. Det vil være en forsikring mot å basere sine kommentarer på tendenser som er altfor usikre. Når en bruker testing som hjelpemiddel på denne måten, tror vi ikke det er nødvendig å endre den "beskrivende" stil som brukes i de fleste arbeider på Underavdelingen. For den jevne leser vil det neppe være særlig opplysende å få vite at "følgende forskjeller er signifikante på 5 % nivå". Selv med et teknisk appendiks og mindre tekniske formuleringer kan dette bli mer villedende enn veiledende. På den annen side kan det ofte være på sin plass med forbehold om at "påpekte forskjeller er små og/eller basert på få observasjoner" når en viser til forskjeller som er testet og funnet ikke-signifikante.

REFERANSER

Iversen, Gunvor (1976): "Pendlernes levekår." Utkast. Maskinskrevet notat 4/11-76.

Miller, Rupert G. (1966): "Simultaneous statistical inference." McGraw-Hill, New York.

Sverdrup, Erling (1964): "Lov og tilfeldighet", bind II. Universitetsforlaget, Oslo.

Sverdrup, Erling (1973): "Lov og tilfeldighet", bind I, 2. utgave. Universitetsforlaget, Oslo.

Sverdrup, Erling (1975): "Multiple comparisons by binary and multinary observations." Artikkel 75, Statistisk Sentralbyrå, Oslo.