Test for uendret populasjon mhp ulike kategorier

Vi står ofte overfor oppgaven å kontrollere om to (eller flere) utvalg er trukket fra samme populasjon. Vi ønsker å kontrollere at personene i de to utvalgene ikke har forskjellig alders- eller yrkessammensetning mv. Spesielt for reisevaneundersøkelser eller undersøkelser der telle-enheten er turer til og fra ulike soner (turmatriser ønsker vi å kontrollere at turene ikke fordeler seg forskjellig mhp reisehensikt, reisetidspunkt mv ved hvert av de to ulike tidspunktene.

Dersom utvalgene ikke er små, og de enkelte observasjonene er uavhengige[1] kan vi anta at fordelingen på aldersgrupper eller andre kategorier er uavhengig multinomisk fordelt. Vi står således overfor oppgaven med å sammenligne to multinomiske fordelinger.[2]

Mer presist ønsker vi å teste

$H_o: p_{1,j} = p_{2,j}$ $\mathrm{j} = 1..\mathrm{w}$ der w er antall kategorier mot

$\mathrm{H}_1:$ minst en $p_{1,j} \neq p_{2,j}$

Eller med andre ord om fordelingene er homogene). Under null-hypotesen kan det vises (Sverdrup 1964) at observatoren:

$$Z = \sum_{i,j} \frac{\left(X_{i,j} - X_{\cdot j} \dfrac{X_{i\cdot}}{X_{\cdot\cdot}} \right)^2}{X_{\cdot j} \dfrac{X_{i\cdot}}{X_{\cdot\cdot}}}$$ er χ^2 fordelt med w - 1 frihetsgrader. i=1..2 j=1..w

Vi forkaster null-hypotesen når Z er større enn 95% fraktilen i Kji-kvadratfordelingen med w-1 frihetsgrader.

Tabell over fraktilene i Kji-kvadrafordelingen for 1 til 10 frihetsgrader

Frihets-grader	**5% nivå**	**10% nivå**
1	3,84	2,71
2	5,99	4,61
3	7,81	6,25
4	9,49	7,78
5	11,07	9,24
6	12,59	10,64
7	14,07	12,02
8	15,51	13,36
9	16,92	14,68
10	18,31	15,99

[1] Når vi eksempelvis sammenligner turenes fordeling på reiseformål eller antall turer som er foretatt i ulike tidsperioder er det strengt tatt ikke korrekt å regne med at de enkelte reisene foretas uavhengig av hverandre. Turene foretatt av en enkelt person (husholdning) ved hvert av tidspunktene vil være innbyrdes avhengige.

[2] Testen er gyldig selv om vi har flere enn v=2 tidsperioder.

Eksempel:.
Vi ønsker å teste om to aldersfordelinger fra to forskjellige år er "like". Aldersfordelingene er gitt som to prosentfordelinger, men vi kjenner utvalgsstørrelsen. Første skritt er da å regne ut antall personer i de enkelte gruppene: Vi fører radsummene opp til høyre og kolonnesummene nederst. Totalt antall i de to undersøkelsene **X..** framgår også av tabellen.

	1991	**1992**
10-20	12,5%	13,8%
21-30	35,0%	32,5%
31-40	24,0%	28,0%
41-50	21,0%	19,5%
61+	7,5%	6,3%
n	**1000**	**2000**

	1991	**1992**	**Sum 91+92**	**Rad-summer**
10-20	125	275	400	$X_{.1}$
21-30	350	650	1000	$X_{.2}$
31-40	240	560	800	$X_{.3}$
41-50	210	390	600	$X_{.4}$
61+	75	125	200	$X_{.5}$
n	1000	2000	3000	
K.summer	$X_{1.}$	$X_{2.}$	**X..**	**Totalt**

Dersom aldersfordelingen i 1991 ikke er vesentlig forskjellig fra aldersfordelingen i 1992, bør de 400 personene vi totalt har i aldersgruppen 10-20 år, fordele seg i forhold til utvalgsstørrelsen de to årene. Vi får da 400*1000/3000=133.33 og 400*2000/3000=266.67.

Vi lager en tabell over disse forventete verdier ($X_{.j}$*$X_{i.}$/$X_{..}$). Deretter lager vi en tabell over avviket mellom det antallet vi har observert og det vi har beregnet.

Beregnete verdier under forutsetning av samme populasjon i 1991 og 1992 =(Xi.*X.j/X..)

	1991	**1992**	
10-20	133	267	**400**
21-30	333	667	**1000**
31-40	267	533	**800**
41-50	200	400	**600**
61+	67	133	**200**
n	**1000**	**2000**	**3000**

Avvik mellom observerte og forventete verdier

	1991	**1992**	
10-20	-8	8	400
21-30	17	-17	1000
31-40	-27	27	800
41-50	10	-10	600
61+	8	-8	200
n	**1000**	**2000**	**3000**

Testobservatorene er nå:

$$\frac{(-8)^2}{133}+\frac{(17)^2}{333}+\frac{(-27)^2}{267}+\frac{(10)^2}{200}+\frac{(8)^2}{67}+\frac{(8)^2}{267}+\frac{(-17)^2}{667}+\frac{(27)^2}{533}+\frac{(-10)^2}{400}+\frac{(-8)^2}{133}\approx 8{,}3$$

Da vi har 5-1=4 frihetsgrader ser vi at datamaterialet ikke gir grunnlag for å forkaste nullhypotesen om at de to aldersfordelingene er like (8.3 er mindre enn 9.49).

Hvorfor kan vi ikke nøye oss med å teste aldersgruppene hver for seg??
Dersom en tester hver av aldersgruppene for seg, og angir hvilke som er signifikante, vil vi ofte forkaste nullhypotesen om at fordelingene er like, selv om fordelingene faktisk kommer fra samme populasjon..

Vi foretar 5 tester - en test for hver aldersgruppe. Dersom hver av testene er på 5% nivå vil vi under 0-hypotesen ha 95% sannsynlighet for ikke å forkaste. Sannsynligheten for å få forkastning på minst en av testene er lik sannsynligheten for at ikke alle skal resultere i ikke-forkastning = 1-(95%)^5 eller ca. 22.6%.

I vårt eksempel ville en slik framgangsmåte betydd at vi hadde utført 5 tester der vi så på::

$$X = \frac{p_1 - p_2}{\sqrt{\overline{p} \cdot (1 - \overline{p}) \cdot \frac{n_1 + n_2}{n_1 \cdot n_2}}} \quad \text{der } \overline{p} = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$$

p_1 og p_2 er de to prosent - tallene vi skal sammenligne
og n_1 og n_2 er utvalgstørrelsene.

og forkastet 0-hypotesen dersom X er større enn 1.96.

Vi ville da fått forkasning i 3.dje aldersgruppe.ettersom X=2.35.

Nå er vårt eksempel konstruert, og vi har heller ikke tatt hensyn til at det blir en viss avhengighet mellom testene siden vi må ende opp med 100%.

Bruk av mange enkelttester når det er riktig å benytte en samlet test gir vesentlig større sjanse for å komme fram til en feilaktig konklusjon.

SVERDRUP ERLING 1964
Lov og tilfeldighet Bind II, Den praktiske statistikks metode og teknikk, Universitetsforlaget Oslo.