

Fynbus

DS815: Kvalitet og generaliserbarhed af data om social adfærd

Alexander Dernild
Syddansk Universitet

Sat 06 Mar 2021

Hypotese: Er det oftere folk med visse typer rejsekort (personligt, anonymt eller flex), der glemmer at checke ud, eller sker det uafhængigt?

For at teste denne hypotese kan vi bruge en χ^2 -test for uafhængighed, der tester om to eller flere kategoriske variabler er uafhængige af hinanden.

Table 1: Observeret frekvens: Manglende check ud og korttype 2020

	Anonymt	Flex	Personligt	Total
Check Udvej	0	102	434	536
Ja	104	373	183.760	184.237
Nej	7.591	44.076	491.642	543.309
Total	7.695	44.551	675.836	728.082

Table 2: Observeret frekvens (%): Manglende check ud og korttype 2020

	Anonymt	Flex	Personligt	Total
Check Udvej	0,00	0,23	0,06	0,07
Ja	1,35	0,84	27,19	25,30
Nej	98,65	98,93	72,75	74,62
Total	100,00	100,00	100,00	100,00

De ovenstående tabeller viser frekvensen af glemte udcheckninger, fordelt på de tre rejsekortstyper *Anonymt*,

Flex og *Personligt*. Umiddelbart ser det ud til, at det oftest er personer med rejsekortstypen *Personligt*, der glemmer at checke ud, dernæst *Flex* og til sidst *Anonymt*. Ydermere tyder det på, at personer med et *Flex*-rejsekort oftere benytter *Check Udvej*-appen.

I tabel 3 nedenfor, fremgår den forventede frekvens af glemte udcheckninger, såfremt at glemte udcheckninger, sker uafhængigt af rejsekortstypen. Heri er det tydeligt, at personer med et *Personligt* rejsekort oftere glemmer at checke ud, end personer med de andre rejsekortstyper. Derudover viser det sig, at personer med et *Flex*-rejsekort oftere benytter *Check Udvej*-appen, end forventet.

Table 3: Forventet frekvens: Manglende check ud og korttype 2020

	Anonymt	Flex	Personligt	Total
Check Udvej	5,66	32,80	497,54	536
Ja	1.947,18	11.273,38	171.016,45	184.237
Nej	5.742,16	33.244,83	504.322,01	543.309
Total	7.695,00	44.551,00	675.836,00	728.082

Note:

Pearsons chi-i-anden-test

¹ χ^2 (4, N=728.082) 17.836,74

² p-værdi: 0 ***

³ * p < 0.05, ** p < 0.01, *** p < 0.001

For at blive klogere på hvordan, det glemte check ud afhænger af rejsekortstypen, kan der laves en post-hoc test med de standardiserede chi-i-anden residualer, der viser forskellen mellem den observerede og forventede frekvens. Et positivt residual betyder, at der er flere end forventet i en bestemt kategori, og omvendt hvis residualet er negativt.

Table 4: Post-hoc: Manglende check ud og korttype 2020

Svarmulighed	Værdi	Anonymt	Flex	Personligt
Check Udvej	Residual	-2,393667	12,47588	-10,63753
	p-værdi	0,150128	0,00000	0,00000
Ja	Residual	-48,587436	-122,59651	133,10298
	p-værdi	0,000000	0,00000	0,00000
Nej	Residual	48,690026	121,70136	-132,31233
	p-værdi	0,000000	0,00000	0,00000

Note:

Standardiserede chi-i-anden residualer og p-værdier med Bonferoni korrektion

¹ Signifikans: $p < 0.05$, Stærk signifikans: $p < 0.01$, Meget stærk signifikans: $p < 0.001$

I den ovenstående tabel kan det ses, at der er en højere andel end forventet af rejsekortstypen *Anonymt* og *Flex*, der ikke har glemt at checke ud. Modsat viser tabellen, at der er flere med et *Personligt* rejsekort end forventet, der har glemt at checke ud. Derudover er personer med et *Flex* rejsekort, bedre til at benytte *Check Udvej*-appen end de med et *Personlig* kort, hvor residualet for dem med et *Anonymt* kort ikke er signifikant.

På næste side er der en visualisering, af de standardiserede residualer, hvor en blå nuance betyder at der er flere end forventet, og en rød nuance det modsatte.

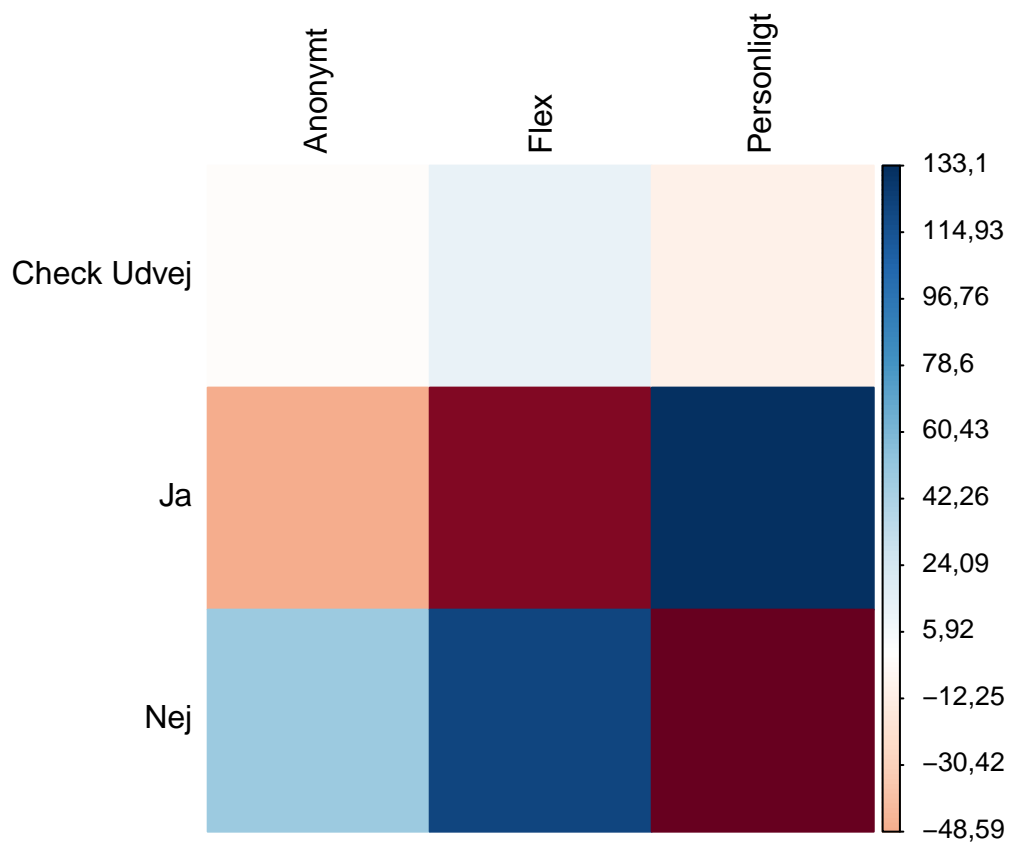


Figure 1: Visualisering af standardiserede residualer 2020

For at se om det samme mønster gentager sig i 2021, gentager vi de samme tests som før. I dataene fra 2021 er der dog den forskel, at brug af *Check-Udvej*-appen ikke er registreret.

Table 5: Observeret frekvens: Manglende check ud og korttype 2021

	Anonymt	Flex	Personligt	Total
Ja	67	128	19.148	19.343
Nej	3.885	12.885	138.993	155.763
Total	3.952	13.013	158.141	175.106

Table 6: Forventet frekvens: Manglende check ud og korttype 2021

	Anonymt	Flex	Personligt	Total
Ja	436,56	1.437,47	17.468,97	19.343
Nej	3.515,44	11.575,53	140.672,03	155.763
Total	3.952,00	13.013,00	158.141,00	175.106

Note:

Pearsons chi-i-anden-test

¹ χ^2 (2, N=175.106) 1.874,114

² p-værdi: 0 ***

³ * p < 0.05, ** p < 0.01, *** p < 0.001

Der er stadig i 2021, en meget stærk signifikant afhængighed, mellem rejsekortstype og glemte udcheckninger. Dog ser det ikke ud til at forskellen, mellem de forskellige rejsekortstyper, er lige så stor i 2021, som den var i 2020. Igen kan vi lave en post-hoc test med de standardiserede residualer, for at se hvor stor forskellen mellem korttyper er, og sammenligne det med forskellen i 2020.

Table 7: Post-hoc: Manglende check ud og korttype 2021

Svarmulighed	Værdi	Anonymt	Flex	Personligt
Ja	Residual	-18,96861	-38,06131	43,27299
	p-værdi	0,00000	0,00000	0,00000
Nej	Residual	18,96861	38,06131	-43,27299
	p-værdi	0,00000	0,00000	0,00000

Note:

Standardiserede chi-i-anden residualer og p-værdier med Bonferroni korrektion

¹ Signifikans: p < 0.05, Stærk signifikans: p < 0.01, Meget stærk signifikans: p < 0.001

Det tyder på at afhængigheden mellem rejsekortstype og glemte udcheckninger, er blevet mindre fra 2020 til 2021, når man sammeligner post-hoc testene. Dog er afhængigheden stadig meget stærk signifikant, og forskellen mellem disse år kan evt. tilskrives, at der er mindre rejsende i 2021 end i 2020.

De standardiserede residualer fra 2021 testen, kan ligeledes 2020 residualerne visualiseres:

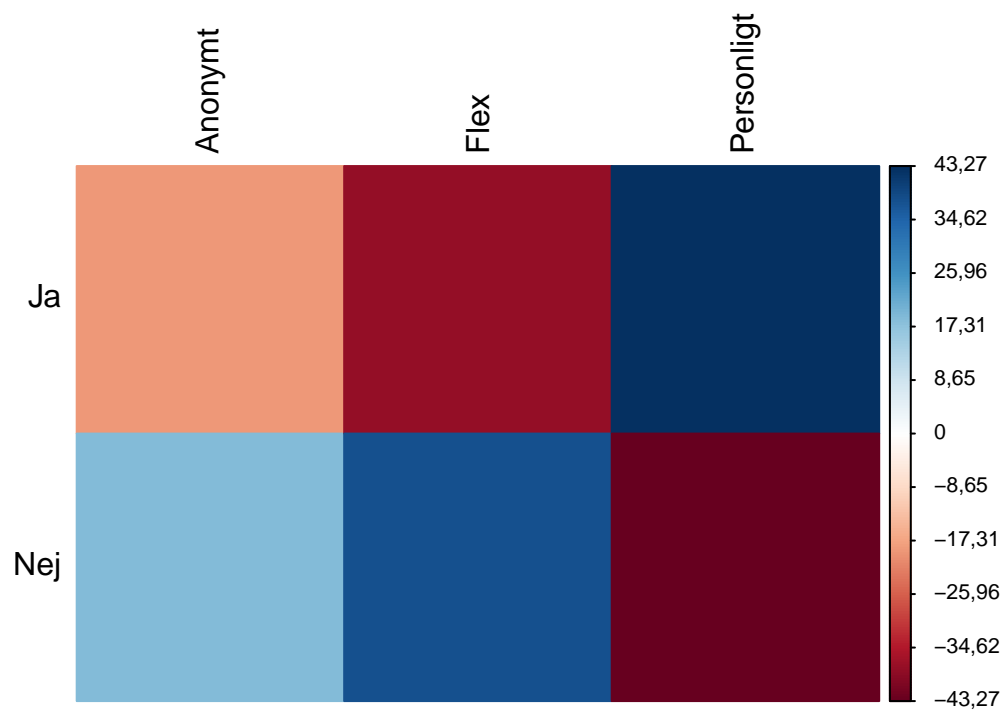


Figure 2: Visualisering af standardiserede residualer 2021