

Til

Uddybende notat vedr. brugen af AKU-mikrodata

Kort om notatet I forbindelse med udarbejdelsen af notatet om *Retningslinjer for eksterne databrugere: Retningslinjer vedrørende publicering af AKU-data* opstod behovet for en uddybende beskrivelse om baggrunden for de udstukne retningslinjer. I dette notat redegøres for, hvordan mikrodata anvendes hensigtsmæssigt – det henvender sig således til specialiserede brugere.

De udstukne retningslinjer Først redegøres for de udstukne retningslinjer og deres baggrund. Derefter gennemgås kvartals- og årsestimater, samt opregningen med fokus på de to elementer: designvægten og bortfaldskorrektionen.

Retningslinjerne beskrives således i *Retningslinjer vedrørende publicering af AKU-data*:

	Resultat	Anvisning
Kvartalsvise tal	Under 4.000 personer (vægtede tal)	Tal publiceres ikke
	4.000-7.000 personer (vægtede tal)	Tal kan publiceres med en advarsel om at tallene er usikre*
Årlige tal	Under 2.000 personer (vægtede tal)	Tal publiceres ikke
	2.000 – 4.000 personer (vægtede tal)	Tal kan publiceres med en advarsel om at tallene er usikre*

* Tal indenfor den gruppe, der kan publiceres med advarsel, bør kun publiceres i særlige tilfælde. Danmarks Statistik anbefaler, at dette drøftes med medarbejdere fra Arbejdskraftundersøgelsen inden publicering.

Der ligger i høj grad pædagogiske og praktiske hensyn bag den valgte formidlingsmetode. Således lå den nedre grænse for kvartalsvise tal oprindeligt på 3.500, men dette harmonerede dårligt med anvisningen om kun at publicere i hele tusinder. Usikkerheden er for opregnede estimater langtfra altid den samme, hvorfor to tal på 4.000 ikke nødvendigvis er lige sikre.

Grænserne bestemmes af antal interview de bygger på – jo flere interview, des større sikkerhed. Da der er flere interview bag årstallene er den opregnede grænseværdi mindre. Estimaterne er baseret på beskæftigede i forskellige stillingskategorier, da der her typisk er relativt stor usikkerhed.

Begrundelsen for de valgte grænser forsøges illustreret i bilagets figur 1 og 2.

Kvartals-estimer Først ses på kvartals-estimer. Overordnet peger estimerne på, at varianskoefficienten (CV) runder 20 pct. omkring en estimat-størrelse på 7.000, mens de passerer 25 pct. ved en størrelse på 4.000. Med 95 procents sikkerhed ligger et resultat på 7.000 mellem 4.200 og 9.800, mens et resultat på 4.000 ligger mellem 2.000 og 6.000. Afhængigt af estimat-størrelsen kan resultaterne potentielt variere med henholdsvis faktor 2,3 (ved 7.000) og faktor 3 (ved 4.000).

Års-estimer Ved års-estimerne er usikkerheden noget lavere og det er udgivelsesgrænserne også. Af samme årsag kan disse resultater bedre anvendes til struktur-statistiske formål, om end der selvfølgelig stadig er stikprøveusikkerhed. Resultater omkring 4.000 ligger med en CV på ca. 12 pct., mens CV for resultater omkring 2.000 ligger på 18 pct. Med 95 procents sikkerhed ligger estimer på 4.000 altså mellem godt 3.000 og 5.000, mens estimer på 2.000 ligger mellem 1.200 og 2.800. For både den øvre og nedre grænse er tallene væsentligt sikrere end kvartalsestimerne, hvilket selvfølgelig også er formålet med årsdata.

Designvægt Hvert interview i AKU har i gennemsnit en endelig vægt på 200 for kvartalsdata og på 50 for årsdata. Det enkelte interviews vægt udgøres primært af designvægten, der afhænger af størrelsen af de enkelte grupper, der bliver udtrukket til undersøgelsen.

På kvartalsdatasættet fra 4. kvartal 2011 varierede designvægtene mellem 108 og 600, gennemsnittet var 199, mens medianen var 220. Et kvartalsestimat på 7.000 opregnet, består dermed (hvis man ser bort fra bortfaldskorrekturen) af mellem 12 og 65 interview.

Dette skal ses ift. at det normalt anbefales af Danmarks Statistiks metodekontor, at udsagn bør baseres på mindst 20-25 observationer. Det vil også sjældent være tilfældet, at 7.000 opregnede personer bygger på mindre end 20 observationer – derfor er 7.000 en god tommelfingerregel.

Omvendt kan et estimat på 4.000 bygge på mellem 7 og 37 observationer. Der vil således være tilfælde, hvor estimer under 4.000 indeholder tilpas mange observationer til at resultatet bør være sikkert nok til at kunne blive offentliggjort. Her kan ekspertbrugere med adgang til mikrodata teste antallet af observationer bag et bestemt estimat, for at få et fingerpeg om sikkerheden.

Bortfald Hidtil har vi alene set på den rene stikprøveusikkerhed. Men der kan også være et problem med skævheden som følge af bortfald. Overordnet set, er det kun lidt mere end hver anden person i stikprøven, der deltager i undersøgelsen. Bortfaldet kan imidlertid variere ganske kraftigt mellem forskellige grupper i samfundet. Man kan analysere bortfaldet ved at se på det andet element i den endelige opregning – kalibreringsvægten eller bortfaldskorrekturen. Denne korrektionsvægt afhænger dels af den valgte opregningsmodel, dels af det konkrete bortfald for forskellige grupper af respondenter. Den har et gennemsnit på 1, men varierer fra 0,1 til 4 og illustrerer bl.a. hvor stort bortfald, der er i en bestemt gruppe respondenter ift. det gennemsnitlige bortfald blandt alle.

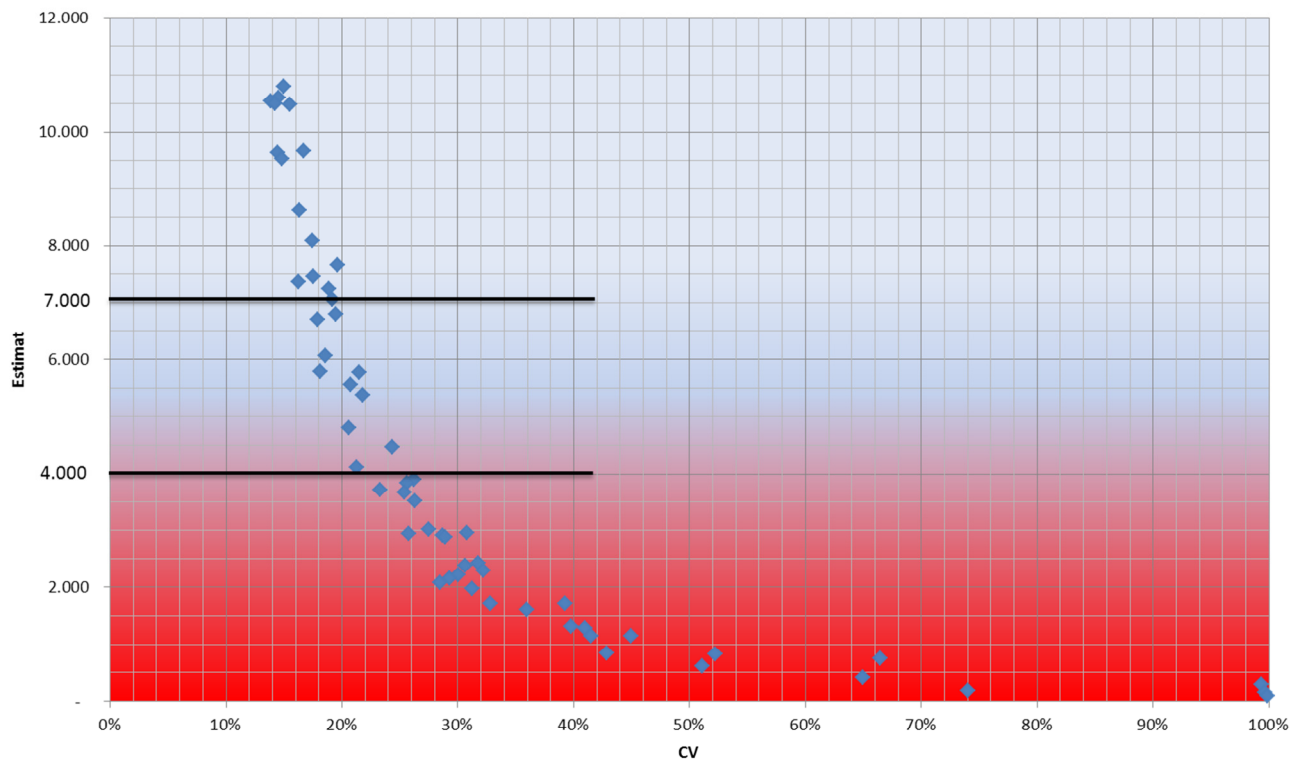
I tabellen er der et eksempel på tre estimer, der umiddelbart ligger meget tæt på hinanden, både vægten, uvægtet samt når man ser på varianskoefficienten. Bortfaldskorrekturen nuancerer imidlertid dette billede. Man ser således, at den nederste gruppe har et estimeret bortfald, der er 11 pct. højere end det gennemsnitlige bortfald i undersøgelsen – hvilket indikerer en højere risiko for skævhed i besvarelsene. Den øverste gruppe har et estimeret bortfald, der er 11 pct. lavere end gennemsnittet, og risikoen for skævhed er dermed lavere for denne gruppe.

3-cifret ISCO	Vægtet estimat	Uvægtet estimat	CV	Bortfaldskorrektion
Ledelse af hovedaktiviteten inden for fremstillingsvirksomhed, råstofudvinding, byggeri, forsyning, distribution mv.	10.539	56	14%	-11%
Arbejde inden for nærings- og nydelsesmiddelområdet	10.493	56	14%	0%
Arbejde med plantevækst	10.796	52	15%	11%

Det er særligt vigtigt at holde øje med bortfaldskorrekturen, når man ser på små grupper, der ligger tæt på udgivelsesgrænserne, for at undersøge risikoen for systematisk bias i de foreliggende besvarelses. Et eksempel til at illustrere problematikken med bortfaldskorrekturen kan ses hos gruppen af indvandrere. Hvor bortfaldskorrekturen for danske statsborgere er -5,3 pct., dvs. de deltager i lidt højere grad end gennemsnittet, er den for ikke-vestlige udenlandske statsborgere 46,2 pct. Risikoen for systematisk skævhed i resultaterne er dermed væsentligt større for ikke-vestlige udenlandske statsborgere end for danske statsborgere. Dette er væsentligt at holde for øje, når man vurderer robustheden i sine resultater. Er man i tvivl om robustheden, er man altid meget velkommen til at kontakte en medarbejder ved Arbejdskraftundersøgelsen.

Bilag: Sammenhængen mellem forskellige stillingskategoriers estimerede størrelse og varianskoeficienter

Figur 1 - Varianskoeficienter for forskellige estimater - kvartalsdata



Figur 2 - Varianskoeficienter for forskellige estimater - årsdata

