2006/70 Notater

Statistics Norway

Statistisk sentralbyrå

Jan Bjørnstad

Notater

Imputering i AKU for undersysselsetting

Stabsavdelingen/Seksjon for statistiske metoder og standarder

Innhold

1. Innledning

Det finnes to typer frafall i arbeidskraftundersøkelsene (AKU). Enhetsfrafall (også kalt totalt frafall) har vi når intervjuobjektet (IO) ikke har besvart spørreskjemaet. Partielt frafall forekommer når IO har besvart noen, men ikke alle spørsmålene vedkommende skulle ha besvart (i henhold til pilhenvisningene). Enhetsfrafall justeres ved å benytte kalibreringsvekting i estimeringsmetoden. Det justeres for partielt frafall ved hjelp av *imputering* for de variable som har en høy andel uoppgitte svar. Dette gjelder variablene «undersysselsetting», «ønsket arbeidstid for undersysselsatte», «ønsket arbeidstid for ledige» og «faktisk arbeidstid». Hovedårsaken til partielt frafall for disse variablene er at de tilhørende spørsmålene ikke skal stilles ved indirekte intervju (et nært familiemedlem svarer på vegne av IO). Imputeringsmetodene som brukes er tidligere beskrevet «praktisk og ikke-statistisk» i notatene av Hobæk (1993) og Håland, Hobæk og Bø (1993). Vedø, Lie og Bjørnstad (2000, avsnitt 3.1.3) beskriver hovedtrekkene ved den implisitte modellen for responsmekanismen (RM) som AKU's imputeringsrutiner er basert på. I dette notatet konsentrerer vi oss om imputeringen for undersysselsettingen og utdyper beskrivelsen i Vedø et al (2000). Hovedformålet er å gi en presis statistisk og sannsynlighetsteoretisk beskrivelse av imputeringen for undersysselsetting og den underliggende modellen av responsmekanismen som implisitt benyttes. Dette vil gjøre det lettere å vurdere dagens imputeringsrutine samt foreta sammenligninger med rene modell-baserte imputeringsmetoder.

2. Imputering for partielt frafall i undersysselsetting

2.1 AKUs imputeringsmetode

Grovt sett er det to hovedårsaker til partielt frafall: Indirekte intervju og feilkoding / pilhenvisningsfeil. Spørsmålene om undersysselsetting skal ikke stilles ved indirekte intervju. Skjemaer fra indirekte intervjuer har alltid partielt frafall på disse spørsmålene. Innføring av dataassistert intervjuing i AKU i 1996 gir færre pilhenvisningsfeil, siden teknikken gjør det vanskeligere for intervjueren å gå feil i skjemaet. Spørsmålnummeringen brukt i dette notatet refererer til spørreskjema fra 1996, se Bø og Håland (2002)

Undersysselsetting forekommer pr. definisjon blant deltidssysselsatte. Definer undersysselsettingsvariablen y ved

$$y = 1 \text{ hvis person er undersysselsatt,} \\ \quad = 0 \text{ ellers}$$

Deltidssysselsatte er personer med arbeidstid høyst 31 timer pr. uke eller arbeidstid fra 32 til 36 timer med timetall som er lavere enn det som er vanlig heltidsarbeid i yrket eller bedriften.

For å bli definert som undersysselsatt må man svare ja på følgende tre spørsmål:

Sp. 36: Ønsker du en lengre avtalt/gjennomsnittlig arbeidstid, under forutsetning av at inntekten øker tilsvarende ?

Sp. 37a. Har du forsøkt å få lengre arbeidstid ?

Sp. 39. Kan du starte med økt arbeidstid med en gang eller innen en måned ?
(Det egentlige spørsmål 39 i AKU : Hvor raskt kunne du starte med økt arbeidstid?)

Disse spørsmålene gis bare til deltidssysselsatte og ved direkte intervju av IO. Dessuten skal alle deltidssysselsatte først besvare spørsmål 34b om hvilken hovedvirksomhet de har.

La $y_{36} = 1$ hvis «ja» på spørsmål 36, og 0 ellers. Vi definerer y_{37} og y_{39} på samme måte, hvor nå spørsmål 37 står for 37a. Vi ser da at

$$y = y_{36} y_{37} y_{39} \text{ eller ekvivalent } y = 1 \Leftrightarrow y_{36} = y_{37} = y_{39} = 1.$$

Vi bemerker at det imputeres kun for y, ikke separat for de enkelte spørsmålvariablene. Partielt frafall for undersysselsetting kan opptre i hvert av de tre spørsmålene. Dvs. *frafallsmønsteret* kan beskrives ved responsvektoren $\mathbf{r} = (r_{36}, r_{37}, r_{39})$ hvor

$$r_j = 1 \text{ hvis IO gir svar på spørsmål } j \text{, og 0 ellers.}$$

Ved indirekte intervju har vi alltid $\mathbf{r} = \mathbf{0} = (0,0,0)$. Spørsmål 37a blir bare stilt til IO hvis $y_{36} = 1$, og $r_{37} = 0 \Rightarrow r_{39} = 0$. Dette gir at vi har følgende mulige konfigurasjoner av $\mathbf{r}$ ved direkte intervju: $\mathbf{0} = (0,0,0)$, $\mathbf{1} = (1,1,1)$, $(1,0,0)$ og $(1,1,0)$.

For å illustrere hvor mange tilfeller vi kan ha av de forskjellige verdier av $\mathbf{r}$, i 1. kvartal 1992 var det i alt 325 personer med frafall på undersysselsetting hvorav 311 med $\mathbf{r} = \mathbf{0}$, 5 personer med $\mathbf{r} = (1,0,0)$ og 9 personer med $\mathbf{r} = (1,1,0)$. I alt 3757 personer var aktuelle for spørsmål 36. I 2. kvartal 1992 var det ialt 357 personer med partielt frafall på undersysselsetting, alle med $\mathbf{r} = \mathbf{0}$, hvorav 166 ved indirekte intervju.

Imputeringsmetoden for undersysselsetting som er i bruk i AKU i dag, er delvis basert på etterstratifisering etter kjønn x_1 og aldersgruppe x_2. Her er $x_1 = 1$ hvis kvinne og 0 hvis mann, og x_2 tar verdiene 1 - 5 etter aldersgruppene 16-19, 20-24, 25-39, 40-54 og 55-74. La $\mathbf{x} = (x_1, x_2)$.

La r_u være responsvariabelen med hensyn til variabelen «undersysselsetting», med $r_{ui} = 1$ hvis person i gir svar på undersysselsetting, og 0 ellers. Vi lar $\mathbf{r}_i$ betegne responsvektoren for person i. Da har vi at: $(r_{ui} = 1) \Leftrightarrow (\mathbf{r}_i = \mathbf{1}) \cup (r_{36i} = 1, y_{36i} = 0) \cup (r_{36i} = r_{37i} = 1, y_{36i} = 1, y_{37i} = 0)$. La

s_{del} = totale utvalg av deltidssysselsatte
$s = \{i \in s_{del} : r_{ui} = 0\}$ = frafallsgruppen for undersysselsetting

La s_{dir}= frafallsgruppen for direkte intervjuer med $m = |s_{dir}|$ lik antall personer i s_{dir}, og s_{indir} = frafallsgruppen på grunn av indirekte intervjuer og $m_{indir} = |s_{indir}|$.

Responsvariabelen for spørsmål 34b betegnes med r_{34}. Alle med direkte intervju som har $r_{34} = 1$ og $r_{36} = 0$ antas å ha verdien $y_{36} = 0$ (man antar at intervjueren har glemt å krysse av for nei i spørsmål 36). Disse personene får derfor imputeringsverdi $y^* = 0$. I perioden 1988-1992 utgjorde denne gruppen stort sett ca. 60 prosent av det partielle frafallet med direkte intervju. I 2. kvartal 1992 var det 101 av totalt 191 med partielt frafall ved direkte intervju. Legg merke til at for å kunne besvare spørsmål 36 så må spørsmål 34b være besvart. Det betyr at $r_{34} = 0 \Rightarrow \mathbf{r} = \mathbf{0}$. For å skille mellom de forskjellige tilfellene for imputeringen kan vi dele opp s_{dir} etter verdien av r_{34} og $\mathbf{r}$ med $s_r = \{i \in s_{dir}\text{:}\ r_{34i} = 1 \;\&\; \mathbf{r} = \mathbf{0}\}$ og $s_f = \{i \in s_{dir} : r_{34i} = 0 \text{ eller } (r_{34i} = 1 \;\&\; r_{36i} = 1)\}$. La $m_r = |s_r|$ og $m_f = |s_f|$.

De forskjellige frafallsutvalgene kan oppdeles etter verdien på etterstratifiseringsvariablene $\mathbf{x}$:

$$s_{dir}(\mathbf{x}),\ s_{indir}(\mathbf{x}),\ s_f(\mathbf{x}) \text{ og } s_r(\mathbf{x}) \text{ med størrelser } m(\mathbf{x}), m_{indir}(\mathbf{x}),\ m_f(\mathbf{x}) \text{ og } m_r(\mathbf{x}).$$

Vi nevnte at $y^* = 0$ i s_r. For direkte intervjugruppen s_f avhenger imputeringsmetoden av verdien til $\mathbf{r}$. Det er derfor hensiktsmessig å dele opp utvalget av deltidssysselsatte slik det grafisk er vist i Figur 1. Figuren viser også relevante data. Her er $q(\mathbf{x})$ lik observert andel undersysselsatte blant deltidsyssel-

satte i stratum **x**. I tillegg er $s_{f0} = \{i \in s_f : r_{34i} = 0, \mathbf{r}_i = \mathbf{0}\} = \{i \in s_f : r_{34i} = 0\}$,
$s_{f1} = \{i \in s_f : \mathbf{r}_i = (1{,}0{,}0)\}$, og $s_{f2} = \{i \in s_f : \mathbf{r}_i = (1{,}1{,}0)\}$ med størrelser m_{f0}, m_{f1}, og m_{f2}. Disse delutvalgene kan igjen oppdeles etter verdien på etterstratifiseringsvariablene **x**, betegnet med $s_{fk}(\mathbf{x})$ med størrelse $m_{fk}(\mathbf{x})$ for $k = 0,1,2$. Vi legger merke til at $s_{f1} = \{i \in s_{del} : \mathbf{r}_i = (1{,}0{,}0) \text{ og } r_{ui} = 0\}$ og $s_{f2} = \{i \in s_{del} : \mathbf{r}_i = (1{,}1{,}0) \text{ og } r_{ui} = 0\}$.

Figur 1. Utvalget av deltidssysselsatte med observerte data og frafallsmønster i parentes

Stratum	*Svarut-valget*	*Frafallsutvalget s*				
	$r_u = 1$	s_{indir} antall: m_{indir}	s_{dir} antall: m			
			s_f[r_{34} = 0 eller r_{34} = 1& **r** = (1,0,0), (1,1,0)] antall: m_f			s_r antall: m_r
			s_{f0} - antall: m_{f0}	s_{f1} - antall: m_{f1}	s_{f2} - antall: m_{f2}	
.	.	.	.	.	.	.
.	.	.	.	.	.	.
x	$q(\mathbf{x})$	$s_{indir}(\mathbf{x})$ [**r** = **0**]	$s_{f0}(\mathbf{x})$ [r_{34} = 0, **r** = **0**]	$s_{f1}(\mathbf{x})$ [r_{34}= 1; r_u =0, **r** = (1,0,0)]	$s_{f2}(\mathbf{x})$ [r_{34} = 1; r_u =0, **r** = (1,1,0)]	$s_r(\mathbf{x})$ [r_{34}= 1, **r** = **0**]
.	.	.	.	.	.	.
.	.	.	.	.	.	.

I s_r er imputeringsverdien lik $y^* = 0$. For de resterende enhetene i frafallsgruppen s trekkes for hvert stratum **x** et tilfeldig utvalg av imputerte undersysselsatte slik at den imputerte andelen undersysselsatte i $s(\mathbf{x})$ - $s_r(\mathbf{x})$ er lik $q(\mathbf{x})$, med det tilleggskravet at frafallsgruppene $s_{f1}(\mathbf{x})$, $s_{f2}(\mathbf{x})$ har fortrinnsrett. Dette begrunnes med at de som faller fra i spørsmål 37a eller 39 regnes for å ha en større sannsynlighet for å være undersysselsatt enn de som faller fra i spørsmål 36.

Sannsynlighetsfordelingen til Y er relativt til populasjonen av deltidssysselsatte. Dvs. at $P(Y = 1|\mathbf{x})$ er sannsynligheten for at en deltidssysselsatt i populasjonsstratum **x** er undersysselsatt. En presis formulering av AKU's imputeringsmetode:

1) For $i \in s_r : y_i^* = 0$.

2) La $n^*(\mathbf{x})$ være det antall personer i $s_{indir}(\mathbf{x}) \cup s_f(\mathbf{x})$ som skal imputeres som undersysselsatt i frafallsstratum **x**. Da er $n^*(\mathbf{x}) = [m_{indir}(\mathbf{x}) + m_f(\mathbf{x})]\hat{P}(Y = 1 \mid \mathbf{x}, r_u = 1)$

$$\text{hvor } \hat{P}(Y = 1 \mid \mathbf{x}, r_u = 1) = q(\mathbf{x})$$

3) La $n_0^*(\mathbf{x})$, $n_1^*(\mathbf{x})$, $n_2^*(\mathbf{x})$ være antall personer med imputert verdi $y^* = 1$ i frafallsgruppene $s_{indir}(\mathbf{x}) \cup s_{f0}(\mathbf{x}), s_{f1}(\mathbf{x}), s_{f2}(\mathbf{x})$ henholdsvis, slik at $n^*(\mathbf{x}) = n_0^*(\mathbf{x}) + n_1^*(\mathbf{x}) + n_2^*(\mathbf{x})$.

Utvalget av imputerte undersysselsatte trekkes først fra $s_{f2}(\mathbf{x})$, deretter fra $s_{f1}(\mathbf{x})$, og tilslutt fra $s_{indir}(\mathbf{x}) \cup s_{f0}(\mathbf{x})$, på følgende måte :

(i) $n_2^*(\mathbf{x}) = |s_{f2}(\mathbf{x})| = m_{f2}(\mathbf{x})$ hvis $m_{f2}(\mathbf{x}) \leq n^*(\mathbf{x})$, dvs. alle $y_i^* = 1$ i $s_{f2}(\mathbf{x})$.
$n_2^*(\mathbf{x}) = n^*(\mathbf{x})$ hvis $m_{f2}(\mathbf{x}) > n^*(\mathbf{x})$, trukket som et tilfeldig utvalg fra $s_{f2}(\mathbf{x})$.

Hvis $n_2^*(\mathbf{x}) < n^*(\mathbf{x})$:

(ii) $n_1^*(\mathbf{x}) = |s_{f1}(\mathbf{x})| = m_{f1}(\mathbf{x})$ hvis $m_{f1}(\mathbf{x}) \leq n^*(\mathbf{x}) - m_{f2}(\mathbf{x})$, dvs. alle $y_i^* = 1$ i $s_{f1}(\mathbf{x})$.
$n_1^*(\mathbf{x}) = n^*(\mathbf{x}) - m_{f2}(\mathbf{x})$, hvis $m_{f1}(\mathbf{x}) > n^*(\mathbf{x}) - m_{f2}(\mathbf{x})$, trukket som et tilfeldig utvalg fra $s_{f1}(\mathbf{x})$.
Hvis $n_1^*(\mathbf{x}) + n_2^*(\mathbf{x}) < n^*(\mathbf{x})$, dvs. $n_1^*(\mathbf{x}) + n_2^*(\mathbf{x}) = m_{f1}(\mathbf{x}) + m_{f2}(\mathbf{x}) < n^*(\mathbf{x})$:

(iii) $n_0^*(\mathbf{x}) = n^*(\mathbf{x}) - \{m_{f1}(\mathbf{x}) + m_{f2}(\mathbf{x})\}$, trukket som et enkelt tilfeldig utvalg fra $s_{indir}(\mathbf{x}) \cup s_{f0}(\mathbf{x})$.

Vanligvis er $m_{f1} + m_{f2}$ et lite antall personer. Som nevnt tidligere, i 1. og 2. kvartal av 1992 så var $m_{f1} + m_{f2}$ = 14 og 0. Det betyr at vi typisk (dvs. **alltid**) har $m_{f1}(\mathbf{x}) + m_{f2}(\mathbf{x}) < n^*(\mathbf{x})$. For dette tilfellet så kan imputeringsmetoden illustreres grafisk som i Figur 2.

Figur 2. Utvalget av deltidssysselsatte med observerte og imputerte data, for tilfellet $m_{f1}(\mathbf{x}) + m_{f2}(\mathbf{x}) < n^*(\mathbf{x})$

Stratum	*Svarut-valget*	s				
		s_{indir}	s_{dir}			
			s_f (r_{34} = 0 eller r_{34} = 1& $\mathbf{r}$ = (1,0,0) , (1,1,0))			s_r
	$r_u = 1$	$\mathbf{r} = \mathbf{0}$	s_{f0}(r_{34} = 0, $\mathbf{r} = \mathbf{0}$)	s_{f1}	s_{f2}	r_{34} = 1, $\mathbf{r} = \mathbf{0}$
.	.	.		$\mathbf{r}$ = (1,0,0)	$\mathbf{r}$ = (1,1,0)	.
				$r_u = 0$	$r_u = 0$	
.	.	antall: $m_0(\mathbf{x}) = m_{indir}(\mathbf{x}) + m_f(\mathbf{x})$				
$\mathbf{x}$	$q(\mathbf{x})$			antall: $m_{f1}(\mathbf{x})$	antall: $m_{f2}(\mathbf{x})$	antall: $m_r(\mathbf{x})$
		Tilfeldig utvalg av personer gis $y^* = 1$ Antall: $m_0(\mathbf{x})\cdot q(\mathbf{x}) - (m_{f1}(\mathbf{x}) + m_{f2}(\mathbf{x}))$		$y^* = 1$	$y^* = 1$	$y^* = 0$
.	.	.		.	.	.
.	.	.		.	.	.

2.2. Implisitt imputeringsmodell og konsekvenser for responsmekanismen

For å kunne se hva slags imputeringsmodell og modell for responsmekanismen (RM) AKU implisitt benytter seg av skal vi bruke en generell modellbasert måte å *beskrive* stort sett en hvilken som helst imputeringsmetode for undersysselsetting. Hovedresultatet er:

Teorem 1. Vi har følgende RM-modell for AKU med hensyn til undersysselsetting:

a) $P(R_u = 1 \mid \mathbf{x}, \bar{s}_r, y) = P(R_u = 1 \mid \mathbf{x}, \bar{s}_r)$ for $y = 0,1$
b) $P(R_u = 0 \mid \mathbf{x}, \bar{s}_r, y) = P(R_u = 0 \mid \mathbf{x}, \bar{s}_r)$ for $y = 0,1$

<u>*Kommentar*</u> : Utenfor s_r er R_u uavhengig av Y, gitt $\mathbf{x}$. Dvs., AKUs imputeringsrutine antar implisitt at innenfor hvert stratum $\mathbf{x}$ i $\bar{s}_r$ så er sannsynligheten for frafall den samme for undersysselsatte som de andre deltidssysselsatte, dvs. uavhengig av om de er undersysselsatt eller ikke. Det betyr at det partielle frafallet for undersysselsettingsvariabelen antas ignorerbart utenfor s_r.

Korollar 1. $P(\mathbf{R}=\mathbf{1}\mid \mathbf{x},\bar{s}_r, y=1)=P(R_u=1\mid \mathbf{x},\bar{s}_r, y=1)=P(R_u=1\mid \mathbf{x},\bar{s}_r)$

Bevis. Siden $(R_u=1)\Leftrightarrow(\mathbf{R}=\mathbf{1})\cup(R_{36}=1,Y_{36}=0)\cup(R_{36}=1,R_{37}=1,Y_{36}=1,Y_{37}=0)$, så has at $(R_u=1,Y=1)\Leftrightarrow(\mathbf{R}=1,Y=1)$ og dermed følger den første likheten. Resultatet følger nå fra Teorem 1. q.e.d.

Dette betyr at utenfor s_r så er sannsynligheten for å få svar på alle undersysselsettingsspørsmål blant de undersysselsatte i stratum **x** er lik sannsynligheten for svar om undersysselsetting i hele stratum **x**.

For å vise Teorem 1, starter vi med å uttrykke det totale antall imputerte undersysselsatte $n^*(\mathbf{x})$ for en hvilken som helst imputeringsmetode på følgende vis:

$$n^*(\mathbf{x}) = n_{00}^*(\mathbf{x}) + n_{01}^*(\mathbf{x}) + n_1^*(\mathbf{x}) + n_2^*(\mathbf{x}) + n_{01r}^*(\mathbf{x}).$$

Her er $n_1^*(\mathbf{x})$ og $n_2^*(\mathbf{x})$ definert som før. De andre størrelsene $n_{00}^*(\mathbf{x})$, $n_{01}^*(\mathbf{x})$ og $n_{01r}^*(\mathbf{x})$ er antall personer som skal imputeres som undersysselsatt i $s_{indir}(\mathbf{x})$, $s_{f0}(\mathbf{x})$, $s_r(\mathbf{x})$ henholdsvis. AKU-metoden har $n_{01r}^*(\mathbf{x}) = 0$, men det kan godt tenkes at det er for enkelt. La oss innføre d for å indikere om vi har direkte eller indirekte intervju :

$$d=\begin{cases}1 \text{ hvis direkte intervju}\\ 0 \text{ hvis indirekte intervju}\end{cases}$$

Med en gitt imputeringsmodell, $P(Y=1|\text{data})$ for frafallsgruppen, så vil en metodebasert imputeringsmetode bestemme andel undersysselsatte i en gitt frafallsgruppe til å være lik den estimerte $\hat{P}(Y_i=1\mid data)$, for $i\in$ frafallsgruppen. Dvs., antall imputerte undersysselsatte i de forskjellige frafallsgruppene bestemmes på følgende måte:

(1) $n_{01r}^*(\mathbf{x}) = m_r(\mathbf{x})\cdot\ \hat{P}(Y=1\mid \mathbf{x}, d=1, r_{34}=1, \mathbf{r}=\mathbf{0})$

(2) $n_2^*(\mathbf{x}) = m_{f2}(\mathbf{x})\cdot \hat{P}(Y=1\mid \mathbf{x}, d=1, r_{34}=1, \mathbf{r}=(1,1,0), r_u=0)$

(3) $n_1^*(\mathbf{x}) = m_{f1}(\mathbf{x})\cdot \hat{P}(Y=1\mid \mathbf{x}, d=1, r_{34}=1, \mathbf{r}=(1,0,0), r_u=0)$

(4) $n_{00}^*(\mathbf{x}) = m_{indir}(\mathbf{x})\cdot \hat{P}(Y=1\mid \mathbf{x}, d=0)$ (siden $d = 0 \Rightarrow \mathbf{r} = \mathbf{0}$)

(5) $n_{01}^*(\mathbf{x}) = m_{f0}(\mathbf{x})\cdot \hat{P}(Y=1\mid \mathbf{x}, d=1, r_{34}=0, \mathbf{r}=\mathbf{0})$.

I (2) og (3) er det egentlig unødvendig å ta med at ($d = 1$, $r_{34} = 1$), siden $\mathbf{r} = (1,1,0)$ og $\mathbf{r} = (1,0,0)$ begge impliserer at ($d = 1$, $r_{34} = 1$). I (5) er det unødvendig å ta med at $\mathbf{r} = \mathbf{0}$, siden $r_{34} = 0 \Rightarrow \mathbf{r} = \mathbf{0}$. Ved å sammenligne (1) - (5) med beskrivelsen av AKU kan vi se hva slags imputeringsmodell, dvs., betinget modell for undersysselsetting gitt forskjellige frafallskonfigurasjoner, AKU forutsetter, og av det også utlede den responsmodellen AKU faktisk har antatt for det partielle frafallet om undersysselsetting.

Vi ser følgende konsekvenser av AKU's imputeringsrutine, når det gjelder estimerte sannsynligheter for undersysselsetting gitt forskjellige frafallskonfigurasjoner, med $m_0(\mathbf{x}) = = m_{indir}(\mathbf{x}) + m_f(\mathbf{x})$:

$\hat{P}(Y=1\mid \mathbf{x}, d=1, r_{34}=1, \mathbf{r}=\mathbf{0}) = 0$

$\hat{P}(Y=1\mid \mathbf{x}, d=1, r_{34}=1, \mathbf{r}=(1,1,0), r_u=0) = \min(1, m_0(\mathbf{x})q(\mathbf{x})/ m_{f2}(\mathbf{x}))$

$= 1, \qquad$ hvis $m_{f2}(\mathbf{x}) \le m_0(\mathbf{x})q(\mathbf{x})$
$= m_0(\mathbf{x})q(\mathbf{x}) / m_{f2}(\mathbf{x})$, hvis $m_{f2}(\mathbf{x}) > m_0(\mathbf{x})q(\mathbf{x})$

$\hat{P}(Y = 1 \mid \mathbf{x}, d = 1, r_{34} = 1, \mathbf{r} = (1,0,0), r_u = 0)$
$= 1, \qquad$ hvis $m_{f1}(\mathbf{x}) + m_{f2}(\mathbf{x}) \le m_0(\mathbf{x})q(\mathbf{x})$
$= [m_0(\mathbf{x})q(\mathbf{x}) - m_{f2}(\mathbf{x})] / m_{f1}(\mathbf{x})$, hvis $m_{f1}(\mathbf{x}) + m_{f2}(\mathbf{x}) > m_0(\mathbf{x})q(\mathbf{x}) \ge m_{f2}(\mathbf{x})$
$= 0, \qquad$ hvis $m_0(\mathbf{x})q(\mathbf{x}) < m_{f2}(\mathbf{x})$

$\hat{P}(Y = 1 \mid \mathbf{x}, d = 0, \mathbf{r} = \mathbf{0}) = \hat{P}(Y = 1 \mid \mathbf{x}, d = 1, r_{34} = 0, \mathbf{r} = \mathbf{0})$
$= [m_0(\mathbf{x})q(\mathbf{x}) - (m_{f1}(\mathbf{x}) + m_{f2}(\mathbf{x}))] / [m_{indir}(\mathbf{x}) + m_{f0}(\mathbf{x})]$, hvis $m_{f1}(\mathbf{x}) + m_{f2}(\mathbf{x}) < m_0(\mathbf{x})q(\mathbf{x})$
$= 0, \qquad$ ellers

$= q(\mathbf{x}) - (1 - q(\mathbf{x})) \dfrac{m_{f1}(\mathbf{x}) + m_{f2}(\mathbf{x})}{m_{indir}(\mathbf{x}) + m_{f0}(\mathbf{x})} = p_0(\mathbf{x}), \qquad$ hvis $p_0(\mathbf{x}) > 0$
$= 0, \qquad$ hvis $p_0(\mathbf{x}) \le 0$.

Vi ser at $\hat{P}(Y = 1 \mid \mathbf{x}, d = 0, \mathbf{r} = \mathbf{0}) \le q(\mathbf{x})$, og $\hat{P}(Y = 1 \mid \mathbf{x}, d = 0, \mathbf{r} = \mathbf{0}) = q(\mathbf{x})$ hvis og bare hvis $m_{f1}(\mathbf{x}) + m_{f2}(\mathbf{x}) = 0$.

Den <u>ekte</u> frafallsgruppen for stratum $\mathbf{x}$ i henhold til AKU's imputeringsopplegg er gitt ved

$$s_{Ef}(\mathbf{x}) = s_{indir}(\mathbf{x}) \cup s_f(\mathbf{x}).$$

Vi kan nå *innføre* følgende estimerte responssannsynligheter, betinget gitt at personen er i den ekte frafallsgruppen, med definisjonen $\hat{P}(\mathbf{R} = \mathbf{r} \mid s_{Ef}(\mathbf{x})) = \hat{P}(\mathbf{R}_i = \mathbf{r} \mid i \in s_{Ef}(\mathbf{x}))$:

$\hat{P}(\mathbf{R} = \mathbf{0} \mid s_{Ef}(\mathbf{x})) = [m_{indir}(\mathbf{x}) + m_{f0}(\mathbf{x})] / m_0(\mathbf{x})$, $m_0(\mathbf{x}) = = m_{indir}(\mathbf{x}) + m_f(\mathbf{x}) = |s_{Ef}(\mathbf{x})|$

$\hat{P}(\mathbf{R} = (1,1,0) \mid s_{Ef}(\mathbf{x})) = m_{f2}(\mathbf{x}) / m_0(\mathbf{x})$

$\hat{P}(\mathbf{R} = (1,0,0) \mid s_{Ef}(\mathbf{x})) = m_{f1}(\mathbf{x}) / m_0(\mathbf{x})$

Vi har da: $\hat{P}(\mathbf{R} = \mathbf{0} \mid s_{Ef}(\mathbf{x})) + \hat{P}(\mathbf{R} = (1,1,0) \mid s_{Ef}(\mathbf{x})) + \hat{P}(\mathbf{R} = (1,0,0) \mid s_{Ef}(\mathbf{x})) = 1$.

Det vanligste er at $\hat{P}(\mathbf{R} = \mathbf{0} \mid s_{Ef}(\mathbf{x}))$ er nær 1 (I 2. kvartal 1992 var denne frafallssannsynligheten lik 1). Vi kan nå uttrykke de estimerte sannsynligheter for undersysselsetting gitt forskjellige frafallskonfigurasjoner på følgende form, ved å benytte at $q(\mathbf{x}) = \hat{P}(Y = 1 \mid \mathbf{x}, r_u = 1)$:

<u>AKUs estimerte imputeringsmodell:</u>

(AKU-I) $\hat{P}(Y = 1 \mid \mathbf{x}, \mathbf{r} = (1,1,0), r_u = 0) = \min\left(1, \dfrac{\hat{P}(Y = 1 \mid \mathbf{x}, r_u = 1)}{\hat{P}(\mathbf{R} = (1,1,0) \mid s_{Ef}(\mathbf{x})}\right)$

(AKU-II) $\hat{P}(Y = 1 \mid \mathbf{x}, \mathbf{r} = (1,0,0), r_u = 0) = \max\left\{0, \min\left(1, \dfrac{\hat{P}(Y = 1 \mid \mathbf{x}, r_u = 1) - \hat{P}(\mathbf{R} = (1,1,0) \mid s_{Ef}(\mathbf{x}))}{\hat{P}(\mathbf{R} = (1,0,0) \mid s_{Ef}(\mathbf{x}))}\right)\right\}$

Dvs. : $\hat{P}(Y = 1 \mid \mathbf{x}, \mathbf{r} = (1,0,0), r_u = 0) = 0$, hvis $\hat{P}(Y = 1 \mid \mathbf{x}, r_u = 1) < \hat{P}(\mathbf{R} = (1,1,0) \mid s_{Ef}(\mathbf{x}))$.

Ellers så er

$$\hat{P}(Y=1\mid \mathbf{x},\mathbf{r}=(1,0,0),r_u=0)=\min\left(1,\frac{\hat{P}(Y=1\mid \mathbf{x},r_u=1)-\hat{P}(\mathbf{R}=(1,1,0)\mid s_{Ef}(\mathbf{x})}{\hat{P}(\mathbf{R}=(1,0,0)\mid s_{Ef}(\mathbf{x}))}\right)$$

(AKU-III) $\hat{P}(Y=1\mid \mathbf{x},d=0)=\hat{P}(Y=1\mid \mathbf{x},d=1,r_{34}=0,\mathbf{r}=\mathbf{0})$

$$=\max\left(0,1-\frac{\hat{P}(Y=0\mid \mathbf{x},r_u=1)}{\hat{P}(\mathbf{R}=\mathbf{0}\mid s_{Ef}(\mathbf{x}))}\right)=\max\left(0,\frac{\hat{P}(Y=1\mid \mathbf{x},r_u=1)-\hat{P}(\mathbf{R}\neq\mathbf{0}\mid s_{Ef}(\mathbf{x}))}{\hat{P}(\mathbf{R}=\mathbf{0}\mid s_{Ef}(\mathbf{x}))}\right)$$

(AKU-IV) $\hat{P}(Y=1\mid \mathbf{x},d=1,r_{34}=1,\mathbf{r}=\mathbf{0})=\hat{P}(Y=1\mid s_r(\mathbf{x}))=0$

Dette betyr at AKU's imputeringsmodell er gitt ved (AKU-I) - (AKU-IV) med P istedenfor $\hat{P}$.

Den betingede fordelingen til Y gitt at frafallsvariabelen $r_u = 0$ er gitt i Lemma 1. Bevis er i Appendix.

Lemma 1.

a) $P(Y=1\mid \mathbf{x},r_u=0,s_{Ef})=P(Y=1\mid \mathbf{x},r_u=1)$

som kan uttrykkes

b) $P(Y=1\mid \mathbf{x},\bar{s}_r,r_u=0)=P(Y=1\mid \mathbf{x},\bar{s}_r,r_u=1)=P(Y=1\mid \mathbf{x},\bar{s}_r)(=P(Y=1\mid \mathbf{x})/P(\bar{s}_r\mid \mathbf{x}))$

<u>Kommentar</u>. Vi ser at Lemma 1 holder når $P(\mathbf{R}=(1,1,0),R_u=0\mid \mathbf{x})=P(\mathbf{R}=(1,0,0),R_u=0\mid \mathbf{x})=0$, dvs. $P(\mathbf{R}\neq\mathbf{0}\cap s_{Ef}\mid \mathbf{x})=0$, og Lemma 1 holder derfor også når $\mathbf{r}=\mathbf{0}$ i hele frafallsgruppen.

Lemma 1b) sier at Y og R_u er stokastisk uavhengige, gitt $(\mathbf{x},\bar{s}_r)$. Herav følger Teorem 1. Merk at to 0/1 variable $R_u$og Y er stokastisk uavhengige hvis og bare hvis P(Y=1|r_u=1) = P(Y=1).

Neste avsnitt skal se på "tilfellet som alltid holder", dvs. $m_{f1}(\mathbf{x}) + m_{f2}(\mathbf{x}) < n^*(\mathbf{x}) \Leftrightarrow$ $\hat{P}(Y=1\mid \mathbf{x},r_u=1)>\hat{P}(\mathbf{R}\neq\mathbf{0}\mid s_{Ef}(\mathbf{x}))$. Det vil antas at det samme holder for sannsynlighetene.

2.3. Modellkonsekvenser for responsmekanismen av AKUs imputeringsmetode. Tilfellet som alltid holder- $P(Y=1\mid \mathbf{x},r_u=1)>P(\mathbf{R}\neq\mathbf{0}\mid s_{Ef}(\mathbf{x}))$

Fra (AKU-I) -(AKU-IV), modellen som ligger til grunn for AKU's imputeringsmetode er nå:

<u>Imputeringsmodellen</u>:

(1) $P(Y=1\mid \mathbf{x},\mathbf{r}=(1,1,0),r_u=0)=1$

(2) $P(Y=1\mid \mathbf{x},\mathbf{r}=(1,0,0),r_u=0)=1$

(3) $P(Y=1\mid \mathbf{x},d=0)=P(Y=1\mid \mathbf{x},d=1,r_{34}=0,\mathbf{r}=\mathbf{0})=\dfrac{P(Y=1\mid \mathbf{x},r_u=1)-P(\mathbf{R}\neq\mathbf{0}\mid s_{Ef}(\mathbf{x}))}{P(\mathbf{R}=\mathbf{0}\mid s_{Ef}(\mathbf{x}))}$

(4) $P(Y=1\mid \mathbf{x},d=1,r_{34}=1,\mathbf{r}=\mathbf{0})=P(Y=1\mid \mathbf{x},s_r)=0\Leftrightarrow P(\bar{s}_r\mid \mathbf{x},y=1)=1$

Vi skal nå se på *frafallsmønsteret* for undersysselsettingsvariabelen (bevis i Appendix).

Korollar 2. Sannsynlighetene $P(\mathbf{R}=\mathbf{r},R_u=0\mid \mathbf{x},y)$:

(i) $P(\mathbf{R}=(1,0,0),R_u=0\mid \mathbf{x},y=1)=P(\mathbf{R}=(1,0,0)\mid \mathbf{x},y=1)=P(\mathbf{R}=(1,0,0),R_u=0\mid \mathbf{x})/P(Y=1\mid \mathbf{x})$

(ii) $P(\mathbf{R}=(1,1,0),R_u=0\mid \mathbf{x},y=1)=P(\mathbf{R}=(1,1,0)\mid \mathbf{x},y=1)=P(\mathbf{R}=(1,1,0),R_u=0\mid \mathbf{x})/P(Y=1\mid \mathbf{x})$

(iii) $P(\mathbf{R}=(1{,}0{,}0), R_u=0 \mid \mathbf{x}, y=0) = P(\mathbf{R}=(1{,}1{,}0), R_u=0 \mid \mathbf{x}, y=0) = 0$

$P(\mathbf{R}=\mathbf{0}, R_u=0 \mid \mathbf{x}, y) = P(\mathbf{R}=\mathbf{0} \mid \mathbf{x}, y)$ og:

$$\text{(iv)}\quad P(\mathbf{R}=\mathbf{0} \mid \mathbf{x}, y=1) = \frac{P(\mathbf{R}=\mathbf{0} \mid x) - P(s_r \mid \mathbf{x})}{P(\bar{s}_r \mid \mathbf{x})} - P(\mathbf{R}\neq\mathbf{0} \cap s_{Ef} \mid \mathbf{x})\left(\frac{1}{P(Y=1 \mid \mathbf{x})} - \frac{1}{P(\bar{s}_r \mid \mathbf{x})}\right)$$

$$= \frac{P(R_u=0 \mid \mathbf{x}) - P(s_r \mid \mathbf{x})}{P(\bar{s}_r \mid \mathbf{x})} - \frac{P(\mathbf{R}\neq\mathbf{0} \cap s_{Ef} \mid \mathbf{x})}{P(Y=1 \mid \mathbf{x})}$$

$$\text{(v)}\quad P(\mathbf{R}=\mathbf{0} \mid \mathbf{x}, y=0) = P(R_u=0 \mid \mathbf{x}, y=0) = 1 - P(R_u=1 \mid \mathbf{x})\frac{P(\bar{s}_r \mid \mathbf{x}, y=0)}{P(\bar{s}_r \mid \mathbf{x})}$$

$$= P(\mathbf{R}=\mathbf{0} \mid \mathbf{x})\frac{P(\bar{s}_r \mid \mathbf{x}, y=0)}{P(\bar{s}_r \mid \mathbf{x})} + 1 - \frac{P(\bar{s}_r \mid \mathbf{x}, y=0)}{P(\bar{s}_r \mid \mathbf{x})}(1 - P(\mathbf{R}\neq\mathbf{0} \cap s_{Ef} \mid \mathbf{x}))$$

Kommentar : Når det gjelder frafallstype har vi ikke ignorerbarhet, heller ikke utenfor s_r . Frafalls-sannsynlighetene (i),(ii) for $\mathbf{r}\neq\mathbf{0}$ er nær 0 (ca. 0,001 -0,002), og dermed ikke så interessante.

Vi legger også merke til at $P(\mathbf{R}=\mathbf{0} \mid \mathbf{x}, \bar{s}_r, y=0) = P(R_u=0 \mid \mathbf{x}, \bar{s}_r, y=0) = P(R_u=0 \mid \mathbf{x}, \bar{s}_r)$.

For å illustrere "graden" av ikke–ignorerbarhet så skal vi se på at modelleksempel.

Typisk modell, relativt til populasjonen av deltidssysselsatte

(T-0) $P(Y=0 \mid \mathbf{x})$ varierer mellom 0,72 og 0,90

(T-1) $P(s_r \mid \mathbf{x}) = 0{,}025$

(T-2) $P(\mathbf{R}\neq\mathbf{0} \mid s_{Ef}(\mathbf{x})) = 0{,}05$

(T-3) $P(s_{Ef} \mid \mathbf{x}) = 0{,}05$

(T-4) $P(\bar{s}_r \mid \mathbf{x}, y=0)/P(\bar{s}_r \mid \mathbf{x}) = 0{,}99 - 0{,}997$

Disse tilnærmingene er basert på observerte AKU-utvalg. For eksempel, angående (T-1), så vil ca. 100 av ca. 4000 deltidssysselsatte være i denne gruppen. (T-4) gjelder når (T-0) og (T-1) holder.

Fra Teorem 1, $P(R_u=1 \mid \mathbf{x}, \bar{s}_r, y) = P(R_u=1 \mid \mathbf{x}, \bar{s}_r)$, som er ekvivalent med

$$P(R_u=0 \mid \mathbf{x}, y) = P(R_u=0 \mid \mathbf{x})\frac{P(\bar{s}_r \mid \mathbf{x}, y)}{P(\bar{s}_r \mid \mathbf{x})} + 1 - \frac{P(\bar{s}_r \mid \mathbf{x}, y)}{P(\bar{s}_r \mid \mathbf{x})}.$$

Dette medfører at under typisk så blir

$$P(R_u=0 \mid \mathbf{x}, y=0) = P(R_u=0 \mid \mathbf{x}) + \varepsilon \cdot P(R_u=1 \mid \mathbf{x}) \text{ hvor } 0{,}003 \le \varepsilon \le 0{,}01.$$

$$P(R_u=0 \mid \mathbf{x}, y=1) = P(R_u=0 \mid \mathbf{x}) - 0{,}026 \cdot P(R_u=1 \mid \mathbf{x}).$$

Legg merke til at $P(R_u=0 \mid \mathbf{x})$ er av størrelsesorden 0,10. Det betyr at graden av avhengighet mellom R_u og Y er stor, når vi ikke betinger med hensyn på $\bar{s}_r$. Sannsynligheten for frafall er betydelig mindre blant de undersysselsatte enn de andre deltidssysselsatte.

Fra Korollar 2 (iv) og (v) fås at

$$P(\mathbf{R}=\mathbf{0} \mid \mathbf{x}, y=1) = 1{,}026 \cdot P(\mathbf{R}=\mathbf{0} \mid \mathbf{x}) - 0{,}048 \text{ til } 1{,}026 \cdot P(\mathbf{R}=\mathbf{0} \mid \mathbf{x}) - 0{,}032$$

$$P(\mathbf{R}=\mathbf{0} \mid \mathbf{x}, y=0) = P(R_u=0 \mid \mathbf{x}, y=0) = P(\mathbf{R}=\mathbf{0} \mid \mathbf{x}) + 0{,}0025 + \varepsilon \cdot P(\mathbf{R}\neq\mathbf{0} \mid \mathbf{x}),\quad 0{,}003 \le \varepsilon \le 0{,}01.$$

2.4 Forenklet tilfelle: r = 0 i hele frafallsgruppen

Nå er $P(\mathbf{R} \neq \mathbf{0} \cap s_{Ef} \mid \mathbf{x}, y) = 0$, eller ekvivalent $P(\mathbf{R} = (1{,}0{,}0), R_u = 0 \mid \mathbf{x}, y) =$ $P(\mathbf{R} = (1{,}1{,}0), R_u = 0 \mid \mathbf{x}, y) = 0$. Vi har da selvsagt også $P(\mathbf{R} \neq \mathbf{0} \cap s_{Ef} \mid \mathbf{x}) = 0$ og dermed $P(\mathbf{R} \neq \mathbf{0} \mid s_{Ef}(\mathbf{x})) = 0$. Det gir følgende konsekvenser for imputeringsmodell og RM, direkte fra (3),(4), Korollar 2 (iv) og (v):

(Imp-1) $P(Y = 1 \mid \mathbf{x}, d = 0) = P(Y = 1 \mid \mathbf{x}, d = 1, r_{34} = 0, \mathbf{r} = \mathbf{0}) = P(Y = 1 \mid \mathbf{x}, r_u = 1)$
(Imp-2) $P(Y = 1 \mid \mathbf{x}, d = 1, r_{34} = 1, \mathbf{r} = \mathbf{0}) = P(Y = 1 \mid \mathbf{x}, s_r) = 0 \Leftrightarrow P(\bar{s}_r \mid \mathbf{x}, y = 1) = 1$

(RM-1) $P(\mathbf{R} = \mathbf{0} \mid \mathbf{x}, y = 1) = \dfrac{P(\mathbf{R} = \mathbf{0} \mid \mathbf{x}) - P(s_r \mid \mathbf{x})}{P(\bar{s}_r \mid \mathbf{x})} = \dfrac{P(R_u = 0 \mid \mathbf{x}) - P(s_r \mid \mathbf{x})}{P(\bar{s}_r \mid \mathbf{x})} = P(R_u = 0 \mid \mathbf{x}, \bar{s}_r)$

(RM-2) $P(\mathbf{R} = \mathbf{0} \mid \mathbf{x}, y = 0) = P(R_u = 0 \mid \mathbf{x}, y = 0) = P(\mathbf{R} = \mathbf{0} \mid \mathbf{x}) \dfrac{P(\bar{s}_r \mid \mathbf{x}, y = 0)}{P(\bar{s}_r \mid \mathbf{x})} + 1 - \dfrac{P(\bar{s}_r \mid \mathbf{x}, y = 0)}{P(\bar{s}_r \mid \mathbf{x})}$

Selv om vi har **r =0** i hele frafallsgruppen så er sannsynligheten for frafall fremdeles avhengig av *y*-verdien. Vi har også fra Teorem 1b), siden {**R=0} = {**R_u = 0}, $P(\mathbf{R} = \mathbf{0} \mid \mathbf{x}, \bar{s}_r, y) = P(\mathbf{R} = \mathbf{0} \mid \mathbf{x}, \bar{s}_r)$. Dvs., utenfor s_r så er **R** og *Y* uavhengige. I tillegg har vi følgende (bevis i Appendix):

(RM-3) $P(\mathbf{R} = \mathbf{1} \mid \mathbf{x}, y = 1) = P(R_u = 1 \mid \mathbf{x}, y = 1) = P(R_u = 1 \mid \mathbf{x}) / P(\bar{s}_r \mid \mathbf{x}) = P(R_u = 1 \mid \mathbf{x}, \bar{s}_r)$

Her er $P(\bar{s}_r \mid \mathbf{x}) = 1 - P(s_r \mid \mathbf{x}) = 1 - P(D = 1, R_{34} = 1, \mathbf{R} = \mathbf{0} \mid \mathbf{x})$

(RM-4) $P(R_u = 1 \mid \mathbf{x}, y = 0) = P(R_u = 1 \mid \mathbf{x}) \dfrac{P(\bar{s}_r \mid \mathbf{x}, y = 0)}{P(\bar{s}_r \mid \mathbf{x})} = P(R_u = 1 \mid \mathbf{x}, \bar{s}_r) P(\bar{s}_r \mid \mathbf{x}, y = 0)$
og $P(\bar{s}_r \mid \mathbf{x}, y = 0) = 1 - \{P(s_r \mid \mathbf{x}) / P(Y = 0 \mid \mathbf{x})\}$.

2.5. Sannsynligheten for partielt frafall på grunn av indirekte intervju

Først ser vi at $P(D = 0 \mid \mathbf{x}, \bar{s}_r) = P(D = 0 \mid \mathbf{x}) / P(\bar{s}_r \mid \mathbf{x})$. Vi har følgende resultater (bevis i Appendix).

Teorem 2
(a) Anta tilfellet som alltid holder**-** $P(Y = 1 \mid \mathbf{x}, r_u = 1) > P(\mathbf{R} \neq \mathbf{0} \mid s_{Ef}(\mathbf{x}))$**.** Da has at

(i) $P(D = 0 \mid \mathbf{x}, y = 1) = \dfrac{P(D = 0 \mid \mathbf{x})}{P(\mathbf{R} = \mathbf{0} \mid s_{Ef}(\mathbf{x}))} \left(\dfrac{1}{P(\bar{s}_r \mid \mathbf{x})} - \dfrac{P(\mathbf{R} \neq \mathbf{0} \mid s_{Ef}(\mathbf{x}))}{P(Y = 1 \mid \mathbf{x})} \right)$

(ii) $P(D = 0 \mid \mathbf{x}, y = 0) = \dfrac{P(D = 0 \mid \mathbf{x})}{P(\mathbf{R} = \mathbf{0} \mid s_{Ef}(\mathbf{x}))} \cdot \dfrac{P(\bar{s}_r \mid \mathbf{x}, y = 0)}{P(\bar{s}_r \mid \mathbf{x})}$

(iii) $P(D = 0 \mid \mathbf{x}, \bar{s}_r, y) = P(D = 0 \mid \mathbf{x}, y) / P(\bar{s}_r \mid \mathbf{x}, y)$

slik at $P(D = 0 \mid \mathbf{x}, \bar{s}_r, y = 1) = P(D = 0 \mid \mathbf{x}, \bar{s}_r) \left(\dfrac{1}{P(\mathbf{R} = \mathbf{0} \mid s_{Ef}(\mathbf{x}))} - \dfrac{P(\bar{s}_r \mid \mathbf{x}) P(\mathbf{R} \neq \mathbf{0} \mid s_{Ef}(\mathbf{x}))}{P(Y = 1 \mid \mathbf{x})} \right)$

og $P(D = 0 \mid \mathbf{x}, \bar{s}_r, y = 0) = \dfrac{P(D = 0 \mid \mathbf{x}, \bar{s}_r)}{P(\mathbf{R} = \mathbf{0} \mid s_{Ef}(\mathbf{x}))}$.

(b) Anta **r =0** i frafallsgruppen. Da has at

(i) $P(D = 0 \mid \mathbf{x}, y = 1) = P(D = 0 \mid \mathbf{x}, \bar{s}_r)$
|(ii) $P(D = 0 \mid \mathbf{x}, y = 0) = P(D = 0 \mid \mathbf{x}, \bar{s}_r) \cdot P(\bar{s}_r \mid \mathbf{x}, y = 0)$.

(iii) $P(D=0\,|\,\mathbf{x},\bar{s}_r,y)=P(D=0\,|\,\mathbf{x},\bar{s}_r)$
Dvs., D, Y er stokastisk uavhengige utenfor s_r.

Kommentar: Generelt, gitt stratum **x** i $\bar{s}_r$, så antas i AKU at det partielle frafallet pga indirekte intervju er ikke-ignorerbart. Derimot så har vi ignorerbarhet for det partielle frafallet pga indirekte intervju utenfor s_r når **r = 0** er eneste mulighet i frafallsgruppen.

Som illustrasjon, under typisk modell i kap. 2.3 så får vi fra Teorem 2a:

(i) $P(D=0\,|\,\mathbf{x},y=1)=$ $0{,}55\cdot P(D=0\,|\,\mathbf{x})$ til $0{,}89\cdot P(D=0\,|\,\mathbf{x})$
(ii) $P(D=0\,|\,\mathbf{x},y=0)=$ $1{,}04\cdot P(D=0\,|\,\mathbf{x})$ til $1{,}05\cdot P(D=0\,|\,\mathbf{x})$
(iii) $P(D=0\,|\,\mathbf{x},\bar{s}_r,y=1)=$ $0{,}57\cdot P(D=0\,|\,\mathbf{x},\bar{s}_r)$ til $0{,}88\cdot P(D=0\,|\,\mathbf{x},\bar{s}_r)$
og $P(D=0\,|\,\mathbf{x},\bar{s}_r,y=0)=1{,}053\cdot P(D=0\,|\,\mathbf{x},\bar{s}_r)$

AKUs imputeringsmetode antar implisitt en sterk avhengighet mellom undersysselsetting og indirekte intervju. Sannsynligheten for indirekte intervju antas å være betraktelig mindre blant undersysselsatte enn for de andre deltidsysselsatte.

2.6. Frafalls-sannsynligheter for "direkte intervju"-gruppen

Vi antar forenklet tilfelle **r =0** i frafallsgruppen, og skal se på $P(\mathbf{R}=\mathbf{0}\,|\,\mathbf{x},y,d=1)$ og $P(\mathbf{R}=\mathbf{0}\,|\,\mathbf{x},\bar{s}_r,y,d=1)$. Først noterer vi at

$$P(\bar{s}_r\,|\,\mathbf{x},y=1,d=1)=\frac{P(\bar{s}_r\,|\,\mathbf{x},y=1)-P(D=0\,|\,\mathbf{x},y=1)}{P(D=1\,|\,\mathbf{x},y=1)}.$$

Derfor har vi, fra (4) i kap. 2.3,

$$\text{(I)}\quad P(\bar{s}_r\,|\,\mathbf{x},y=1,d=1)=1.$$

Teorem 3
(a) $P(\mathbf{R}=\mathbf{0}\,|\,\mathbf{x},\bar{s}_r,y,d=1)=P(\mathbf{R}=\mathbf{0}\,|\,\mathbf{x},\bar{s}_r,d=1)$
(b) $P(\mathbf{R}=\mathbf{0}\,|\,\mathbf{x},y,d=1)=P(\mathbf{R}=\mathbf{0}\,|\,\mathbf{x},\bar{s}_r,d=1)P(\bar{s}_r\,|\,\mathbf{x},y,d=1)+P(s_r\,|\,\mathbf{x},y,d=1)$

Teorem 3 (a) er hovedresultatet her. Det partielle frafallet for undersysselsetting i "direkte" gruppen er ignorerbart utenfor s_r. Vi viser til Teorem 1 for lignende resultatet uavhengig av direkte eller indirekte intervju. Beviset for Teorem 3 er i Appendix. Direkte fra Teorem 3 følger:

Korollar 3
(a) $P(\mathbf{R}=\mathbf{0}\,|\,\mathbf{x},y=1,d=1)=P(\mathbf{R}=\mathbf{0}\,|\,\mathbf{x},\bar{s}_r,d=1)=\dfrac{P(\mathbf{R}=\mathbf{0}\,|\,\mathbf{x},d=1)-P(s_r\,|\,\mathbf{x},d=1)}{P(\bar{s}_r\,|\,\mathbf{x},d=1)}$
(b) $P(\mathbf{R}=\mathbf{0}\,|\,\mathbf{x},y=0,d=1)=$ $P(\mathbf{R}=\mathbf{0}\,|\,\mathbf{x},d=1)\dfrac{P(\bar{s}_r\,|\,\mathbf{x},y=0,d=1)}{P(\bar{s}_r\,|\,\mathbf{x},d=1)}+1-\dfrac{P(\bar{s}_r\,|\,\mathbf{x},y=0,d=1)}{P(\bar{s}_r\,|\,\mathbf{x},d=1)}$

Også i direkte intervju gruppen er sannsynligheten for frafall minst blant de undersysselsatte.

2.7. En sammenligning mellom stratifisert hot-deck og AKUs imputering

AKU's imputeringsmetode er ikke vanlig (stratifisert) hot-deck imputering. I den observerte delen av utvalget så er alltid r_{34} = 1. Vi kan derfor ikke stratifisere etter verdien av r_{34}. Vi kan heller ikke stratifisere etter verdien av *d,* siden vi har ingen observasjoner i utvalget med *d* = 0. Stratifisert hot-deck imputering kan foretas etter variablene **x** som ville tilsvart å bruke

$$\hat{P}(Y=1\,|\,\mathbf{x},r_u=0)=\hat{P}(Y=1\,|\,\mathbf{x},r_u=1)=q(\mathbf{x})\,,$$

(uavhengig av frafallskonfigurasjonen på de tre tilhørende spørsmålene)

hvor q ($\mathbf{x}$) er observert andel undersysselsatte i stratum $\mathbf{x}$. Vi legger merke til at

$$\hat{P}(Y=1\,|\,\mathbf{x},r_u=1)=\hat{P}(Y=1\,|\,\mathbf{x},\mathbf{r}=\mathbf{1})\frac{\hat{P}(\mathbf{R}=\mathbf{1}\,|\,\mathbf{x})}{\hat{P}(R_u=1\,|\,\mathbf{x})}\,.$$

Følgende tabell viser sammenligning mellom stratifisert hot-deck og AKUs imputeringmetode for det "tilfellet som alltid holder", $P(Y=1\,|\,\mathbf{x},r_u=1)>P(\mathbf{R}\neq\mathbf{0}\,|\,s_{Ef}(\mathbf{x}))$:

Frafalls-gruppe	Imputeringssannnsynlighet	Stratifisert hot-deck	AKU
s_{indir}	$\hat{P}(Y=1\,|\,\mathbf{x},d=0,\mathbf{r}=\mathbf{0}\,|\,\mathbf{x})$	$q(\mathbf{x})$	$q(\mathbf{x})-(1-q(\mathbf{x}))\dfrac{m_{f1}(\mathbf{x})+m_{f2}(\mathbf{x})}{m_{indir}(\mathbf{x})+m_{f0}(\mathbf{x})}$
s_{f0}	$\hat{P}(Y=1\,|\,\mathbf{x},d=1,r_{34}=0,\mathbf{r}=\mathbf{0}\,|\,\mathbf{x})$		
s_{f1}	$\hat{P}(Y=1\,|\,\mathbf{x},r_u=0,r_{34}=1,\mathbf{r}=(1{,}0{,}0)\,|\,\mathbf{x})$	$q(\mathbf{x})$	1 (antall ca. 0)
s_{f2}	$\hat{P}(Y=1\,|\,\mathbf{x},r_u=0,r_{34}=1,\mathbf{r}=(1{,}1{,}0)\,|\,\mathbf{x})$	$q(\mathbf{x})$	1 (antall ca. 0)
s_r	$\hat{P}(Y=1\,|\,\mathbf{x},d=1,r_{34}=1,\mathbf{r}=\mathbf{0}\,|\,\mathbf{x})$	$q(\mathbf{x})$	0

3. Konklusjoner

En generell hovedkonklusjon for AKUs imputeringsmetode er at utenfor s_r så antas undersysselsetting og frafall for undersysselsettingsvariabelen stokastisk uavhengige innenfor et gitt stratum $\mathbf{x}$. Det betyr at sannsynligheten for frafall i ett gitt stratum utenfor s_r er den samme for undersysselsatte som for andre deltidssysselsatte. Dette gjelder også innenfor frafallsgruppene fra direkte intervju og indirekte intervju.

Når det gjelder *frafallstype* så er frafallssannsynligheten i ett gitt stratum avhengig av undersysselsettingsstatus både utenfor s_r og generelt.

Generelt for AKUs metode antas, imidlertid, frafallsannsynligheten for undersysselsetting å være avhengig av om person er undersysselsatt eller ikke, og frafallssannsynligheten er betydelig mindre blant de undersysselsatte. Dette synes å være en urimelig antakelse og er en konsekvens av at ingen i frafallsgruppen s_r imputeres til å være undersysselsatt. Denne gruppen utgjør vanligivs 25-30% av den totale frafallsgruppen. Det ville vært en rimeligere antakelse at frafallssannsynligheten er *større* blant de undersysselsatte enn de andre deltidssysselsatte. En konsekvens av "AKUs antakelse" er at antall undersysselsatte høyst sannsynlig blir merkbart *underestimert* .

Også når vi deler opp frafallsgruppen i "indirekte intervju" og "direkte intervju" så har vi samme egenskap. AKUs imputeringsmetode antar implisitt en sterk avhengighet mellom undersysselsetting og indirekte intervju, og sannsynligheten for indirekte intervju er betraktelig mindre blant undersysselsatte enn for de andre deltidsysselsatte. Også i direkte intervju gruppen er sannsynligheten for frafall minst blant de undersysselsatte.

På bakgrunn av disse vurderingene anbefales det at stratifisert hot-deck imputering og modellbasert imputering (kommer i separat notat) utprøves som alternativer til dagens imputeringsmetode.

4. Appendix

<u>Bevis for Lemma 1.</u> La oss først vise at b) er ekivalent med a) :
$P(Y=1\,|\,\mathbf{x},\bar{s}_r,r_u=0)=P(Y=1\,|\,x,s_{Ef},r_u=0)$ siden begivenhetene $\{\bar{s}_r,R_u=0\}$ og $\{s_{Ef},R_u=0\}$ er identiske. Siden $\{R_u=1\}\Rightarrow\bar{s}_r$ har vi trivielt at $P(Y=1\,|\,\mathbf{x},r_u=1)=P(Y=1\,|\,\mathbf{x},\bar{s}_r,r_u=1)$. Vi har

dermed vist at a) er ekvivalent med første likhet i b). Den siste identiteten følger nå fra

$$P(Y=1\mid \mathbf{x},\bar{s}_r) = P(R_u=0\mid \mathbf{x},\bar{s}_r)P(Y=1\mid \mathbf{x},\bar{s}_r,r_u=0)+P(R_u=1\mid \mathbf{x},\bar{s}_r)P(Y=1\mid \mathbf{x},\bar{s}_r,r_u=1)$$

$$= P(R_u=0\mid \mathbf{x},\bar{s}_r)P(Y=1\mid \mathbf{x},\bar{s}_r,r_u=1)+P(R_u=1\mid \mathbf{x},\bar{s}_r)P(Y=1\mid \mathbf{x},\bar{s}_r,r_u=1)= P(Y=1\mid \mathbf{x},\bar{s}_r,r_u=1)$$

Parentesen i pkt. b): Fra (AKU-IV): $\hat{P}(Y=1\mid \mathbf{x},s_r)=0$ og dermed er $P(Y=1\mid \mathbf{x}) = P(Y=1\cap \bar{s}_r\mid \mathbf{x})$.

Det gjenstår å bevise punkt a): Fra (AKU-I) - (AKU-III): $P(Y=1\mid \mathbf{x},r_u=0,s_{Ef})P(s_{Ef}\mid \mathbf{x})=$

$$P(R_u=0,\mathbf{R}=(1{,}1{,}0)\mid \mathbf{x})P(Y=1\mid \mathbf{x},r_u=0,\mathbf{r}=(1{,}1{,}0))+P(R_u=0,\mathbf{R}=(1{,}0{,}0)\mid \mathbf{x})P(Y=1\mid \mathbf{x},\mathbf{r}=(1{,}0{,}0))$$
$$+\ P(D=1,R_{34}=0,\mathbf{R}=\mathbf{0}\mid \mathbf{x})P(Y=1\mid \mathbf{x},d=1,R_{34}=0,\mathbf{r}=\mathbf{0})+P(D=0\mid \mathbf{x})P(Y=1\mid \mathbf{x},d=0)$$

$$\overset{\text{AKU-I}}{=} \min\{P(R_u=0,\mathbf{R}=(1{,}1{,}0)\mid \mathbf{x}),P(Y=1\mid \mathbf{x},r_u=1)P(s_{Ef}\mid \mathbf{x})\}$$
$$\overset{\text{AKU-II}}{+}$$
$$\max\{0,\min(P(\mathbf{R}=(1{,}0{,}0),R_u=0\mid \mathbf{x}),P(Y=1\mid \mathbf{x},r_u=1)P(s_{Ef}\mid \mathbf{x})-P(\mathbf{R}=(1{,}1{,}0),R_u=0\mid \mathbf{x}))\}$$
$$\overset{\text{AKU-III}}{+} \max\{0,P(Y=1\mid \mathbf{x},r_u=1)P(s_{Ef}\mid \mathbf{x})-P(\mathbf{R}\neq\mathbf{0}\cap s_{Ef}\mid \mathbf{x})\}$$

Denne summen er av formen:

$$\text{G} = \min(a_1,b)+\max(0,\min(a_2,b-a_1))+\max(0,b-(a_1+a_2))$$

Vi skal vise at G = b, ved å betrakte de tre følgende mulige tilfellene hver for seg:

(I) $b\le a_1$: G = $b+0+0=b$
(II) $a_1<b\le a_1+a_2$: G = $a_1+(b-a_1)+0=b$
(III) $b>a_1+a_2$: G = $a_1+a_2+(b-a_1-a_2)=b$

Dermed : $P(Y=1\mid \mathbf{x},r_u=0,s_{Ef})P(s_{Ef}\mid \mathbf{x}) = P(Y=1\mid \mathbf{x},r_u=1)P(s_{Ef}\mid \mathbf{x})$ q.e.d.

<u>*Bevis for Korollar 2*</u>. Tilfellet (iii) følger direkte fra (1) og (2).

(i) $P(\mathbf{R}=(1{,}0{,}0),R_u=0\mid \mathbf{x},y=1)=\dfrac{P(\mathbf{R}=(1{,}0{,}0),Y=1,R_u=0\mid \mathbf{x})}{P(Y=1\mid \mathbf{x})}\overset{(2)}{=}\dfrac{P(\mathbf{R}=(1{,}0{,}0),R_u=0\mid \mathbf{x})}{P(Y=1\mid \mathbf{x})}$

(ii) $P(\mathbf{R}=(1{,}1{,}0),R_u=0\mid \mathbf{x},y=1)=\dfrac{P(\mathbf{R}=(1{,}1{,}0),R_u=0,Y=1\mid \mathbf{x})}{P(Y=1\mid \mathbf{x})}\overset{(1)}{=}\dfrac{P(\mathbf{R}=(1{,}1{,}0),R_u=0\mid \mathbf{x})}{P(Y=1\mid \mathbf{x})}$

(iv) $P(R_u=0\mid \mathbf{x},y=1)=P(\mathbf{R}=\mathbf{0}\mid \mathbf{x},y=1)+P(\mathbf{R}\neq\mathbf{0}\cap s_{Ef}\mid \mathbf{x},y=1)$ og

$$P(R_u=0\mid \mathbf{x},y=1)=P(s_r\mid \mathbf{x},y=1)+P(R_u=0\cap \bar{s}_r\mid \mathbf{x},y=1)$$
$$\overset{(4)}{=} 0+P(R_u=0\mid \mathbf{x},y=1)\overset{Tm.1b}{=}P(R_u=0\mid \mathbf{x},\bar{s}_r)$$

$P(\mathbf{R}\neq\mathbf{0}\cap s_{Ef}\mid \mathbf{x},y=1)\overset{(1),(2)}{=}P(\mathbf{R}\neq\mathbf{0}\cap s_{Ef}\mid \mathbf{x})/P(Y=1\mid \mathbf{x})$ og andre del av (iv) følger.

Første del følger av at $P(R_u=0\mid \mathbf{x})=P(\mathbf{R}=\mathbf{0}\mid \mathbf{x})+P(\mathbf{R}\neq\mathbf{0}\cap s_{Ef}\mid \mathbf{x})$.

(v)

$$P(\mathbf{R}=\mathbf{0}\mid \mathbf{x},y=0)=P(R_u=0\mid \mathbf{x},y=0)$$
$$-P(\mathbf{R}=(1,1,0),R_u=0\mid \mathbf{x},y=0)-P(\mathbf{R}=(1,0,0),R_u=0\mid \mathbf{x},y=0)=P(R_u=0\mid \mathbf{x},y=0)$$

Now:

$$P(R_u=0\mid \mathbf{x},y=0)=1-P(R_u=1\mid \mathbf{x},y=0)=1-P(R_u=1\cap \bar{s}_r\mid \mathbf{x},y=0)$$
$$=1-P(\bar{s}_r\mid \mathbf{x},y=0)P(R_u=1\mid \mathbf{x},\bar{s}_r,y=0)=1-P(\bar{s}_r\mid \mathbf{x},y=0)\frac{[P(R_u=1\mid \mathbf{x},\bar{s}_r)-P(Y=1,R_u=1\mid \mathbf{x},\bar{s}_r)]}{P(Y=0\mid \mathbf{x},\bar{s}_r)}$$
$$\overset{Kor.1}{=}1-P(\bar{s}_r\mid \mathbf{x},y=0)\frac{[P(R_u=1\mid \mathbf{x},\bar{s}_r)-P(Y=1\mid \mathbf{x},\bar{s}_r)R_u=1\mid \mathbf{x},\bar{s}_r)]}{P(Y=0\mid \mathbf{x},\bar{s}_r)}$$
$$=1-P(\bar{s}_r\mid \mathbf{x},y=0)\frac{P(R_u=1\mid \mathbf{x})}{P(\bar{s}_r\mid \mathbf{x})}$$

Andre del følger av at $P(R_u=0\mid \mathbf{x})=P(\mathbf{R}=\mathbf{0}\mid \mathbf{x})+P(\mathbf{R}\neq\mathbf{0}\cap s_{Ef}\mid \mathbf{x})$. Q.e.d.

<u>Bevis for (RM-3) og (RM-4) i kap. 2.4</u>

(RM-3):

Første identitet følger av at $P(\mathbf{R}=\mathbf{1},Y=1\mid \mathbf{x},y=1)=P(R_u=1,Y=1\mid \mathbf{x})$. Den andre identiteten:

$$P(R_u=1\mid \mathbf{x},y=1)=P(R_u=1\cap \bar{s}_r\mid \mathbf{x},y=1)=P(\bar{s}_r\mid \mathbf{x},y=1)P(R_u=1\mid \mathbf{x},\bar{s}_r,y=1)$$
$$\overset{(4),Kor.1}{=}1\cdot P(R_u=1\mid \mathbf{x},\bar{s}_r).$$

(RM-4):

$$P(R_u=1\mid \mathbf{x},y=0)=P(R_u=1\cap \bar{s}_r\mid \mathbf{x},y=0)=P(\bar{s}_r\mid \mathbf{x},y=0)P(R_u=1\mid \mathbf{x},\bar{s}_r,y=0)$$
$$\overset{Tm.1,a}{=}P(\bar{s}_r\mid \mathbf{x},y=0)P(R_u=1\mid \mathbf{x})/P(\bar{s}_r\mid \mathbf{x}).$$

q.e.d.

<u>Bevis for Teorem 2</u>. Punkt (a-i):

$$P(D=0\mid \mathbf{x},y=1)=\frac{P(Y=1\mid x,d=0)P(D=0\mid \mathbf{x})}{P(Y=1\mid \mathbf{x})}$$ og fra (3) i delkapittel 2.3:

$$P(Y=1\mid \mathbf{x},d=0)=\frac{P(Y=1\mid \mathbf{x},r_u=1)-P(\mathbf{R}\neq\mathbf{0}\mid s_{Ef}(\mathbf{x}))}{P(\mathbf{R}=\mathbf{0}\mid s_{Ef}(\mathbf{x}))}.$$

I tillegg,

$$P(Y=1\mid \mathbf{x},r_u=1)=\frac{P(Y=1\mid \mathbf{x})P(R_u=1\mid \mathbf{x},y=1)}{P(R_u=1\mid \mathbf{x})}\overset{Tm.1,a}{=}\frac{P(Y=1\mid \mathbf{x})}{P(\bar{s}_r\mid \mathbf{x})}$$ og (a-i) følger

Punkt (a-ii) tilsvarende:

$$P(D=0\mid \mathbf{x},y=0)=\frac{P(Y=0\mid \mathbf{x},d=0)P(D=0\mid \mathbf{x})}{P(Y=0\mid \mathbf{x})}$$ og

$$P(Y=0\mid \mathbf{x},d=0)=$$
$$1-\frac{P(Y=1\mid \mathbf{x},r_u=1)-P(\mathbf{R}\neq\mathbf{0}\mid s_{Ef}(\mathbf{x}))}{P(\mathbf{R}=\mathbf{0}\mid s_{Ef}(\mathbf{x}))}=\frac{P(Y=0\mid \mathbf{x},r_u=1)}{P(\mathbf{R}=\mathbf{0}\mid s_{Ef}(\mathbf{x}))}=\frac{P(Y=0\mid \mathbf{x})-P(s_r\mid \mathbf{x})}{P(\bar{s}_r\mid \mathbf{x})P(\mathbf{R}=\mathbf{0}\mid s_{Ef}(\mathbf{x}))}$$

$$P(D=0\mid \mathbf{x},y=0)=\frac{P(D=0\mid \mathbf{x})}{P(\mathbf{R}=\mathbf{0}\mid s_{Ef}(\mathbf{x}))}\cdot\frac{1-\{P(s_r\mid \mathbf{x})/P(Y=0\mid \mathbf{x})\}}{P(\bar{s}_r\mid \mathbf{x})}$$, og (a-ii) følger.

Punkt (a-iii) følger direkte av : $D=0 \Rightarrow \bar{s}_r$.
Punkt (b) er en direkte konsekvens av (a). q.e.d.

Bevis for Teorem 3:
(a)

$$P(\mathbf{R}=\mathbf{0} \mid \mathbf{x},\bar{s}_r, y=1, d=1) = \frac{P(\mathbf{R}=\mathbf{0} \cap \bar{s}_r \cap Y=1 \cap D=1 \mid \mathbf{x})}{P(\bar{s}_r \cap Y=1 \cap D=1 \mid \mathbf{x})}$$

$$= \frac{P(\mathbf{R}=\mathbf{0}, Y=1, D=1 \mid \mathbf{x}) - P(s_r \cap Y=1 \mid \mathbf{x})}{P(\bar{s}_r \cap Y=1 \cap D=1 \mid \mathbf{x})} = \frac{P(\mathbf{R}=\mathbf{0} \mid \mathbf{x}, y=1, d=1)P(Y=1, D=1 \mid \mathbf{x}) - P(s_r \mid \mathbf{x}, y=1)}{P(\bar{s}_r \mid \mathbf{x}, y=1, d=1)P(Y=1, D=1 \mid \mathbf{x})}$$

$$\overset{(4),I}{=} P(\mathbf{R}=\mathbf{0} \mid \mathbf{x}, y=1, d=1) = 1 - P(R_u=1 \mid \mathbf{x}, y=1, d=1) = 1 - \frac{P(R_u=1 \mid \mathbf{x}, y=1)}{P(D=1 \mid \mathbf{x}, y=1)}$$

$$\overset{(RM-3),Tm.2b}{=} 1 - \frac{P(R_u=1 \mid \mathbf{x},\bar{s}_r)}{P(D=1 \mid \mathbf{x},\bar{s}_r)} = 1 - P(R_u=1 \mid \mathbf{x},\bar{s}_r, d=1) = P(R_u=0 \mid \mathbf{x},\bar{s}_r, d=1) = P(\mathbf{R}=\mathbf{0} \mid \mathbf{x},\bar{s}_r, d=1).$$

Herav følger at

$$P(\mathbf{R}=\mathbf{0} \mid \mathbf{x},\bar{s}_r, y=0, d=1) = \frac{P(\mathbf{R}=\mathbf{0}, Y=0 \mid \mathbf{x},\bar{s}_r, d=1)}{P(Y=0 \mid \mathbf{x},\bar{s}_r, d=1)}$$

$$= \frac{P(\mathbf{R}=\mathbf{0} \mid \mathbf{x},\bar{s}_r, d=1) - P(Y=1 \mid \mathbf{x},\bar{s}_r, d=1)P(\mathbf{R}=\mathbf{0} \mid \mathbf{x},\bar{s}_r, d=1)}{P(Y=0 \mid \mathbf{x},\bar{s}_r, d=1)} = P(\mathbf{R}=\mathbf{0} \mid \mathbf{x},\bar{s}_r, d=1).$$

(b) Først, fra Teorem 2b (iii), $P(D=1 \mid \mathbf{x}, y) = P(s_r \mid \mathbf{x}, y) + P(\bar{s}_r \mid \mathbf{x}, y)P(D=1 \mid \mathbf{x},\bar{s}_r)$.

$$P(\mathbf{R}=\mathbf{0} \mid \mathbf{x}, y, d=1) = 1 - \frac{P(R_u=1 \mid \mathbf{x}, y)}{P(D=1 \mid \mathbf{x}, y)} \overset{RM-3,RM-4}{=} 1 - \frac{P(R_u=1 \mid \mathbf{x},\bar{s}_r)P(\bar{s}_r \mid \mathbf{x}, y)}{P(D=1 \mid \mathbf{x}, y)}$$

$$= \frac{[P(D=1 \mid \mathbf{x}, y) - P(D=1 \mid \mathbf{x},\bar{s}_r)P(\bar{s}_r \mid \mathbf{x}, y)] + P(\mathbf{R}=\mathbf{0} \mid \mathbf{x},\bar{s}_r, d=1)P(D=1 \mid \mathbf{x},\bar{s}_r)P(\bar{s}_r \mid \mathbf{x}, y)}{P(D=1 \mid \mathbf{x}, y)}$$

$$= \frac{P(s_r \mid \mathbf{x}, y)}{P(D=1 \mid \mathbf{x}, y)} + P(\mathbf{R}=\mathbf{0} \mid \mathbf{x},\bar{s}_r, d=1)\left(1 - \frac{P(s_r \mid \mathbf{x}, y)}{P(D=1 \mid \mathbf{x}, y)}\right)$$

Og resultatet følger.

5. Referanser

Bø, T.P. og Håland, I. (2002). Dokumentasjon av arbeidskraftundersøkelsen (AKU). Notater 2002/24

Hobæk, T. (1993). Justering av partielt frafall i arbeidskraftundersøkelsen (AKU). Notater 1993/34

Håland, I., Hobæk, T. og Bø, T.P. (1993). Dokumentasjon av arbeidskraftundersøkelsen (AKU). Notater 1993/27

Vedø, A., Lie, J-A. og Bjørnstad, J. (2000). Statistisk modellering i AKU. Modellstudier og modell-estimering. Notater 2000/21

De sist utgitte publikasjonene i serien Notater

2006/41 KOSTRA. Arbeidsgrupperapporter 2006. 169s.

2006/42 T. Gulbrandsen: Levekårsundersøkelse blant studenter. Dokumentasjonsrapport. 66s.

2006/43 A-G. Jørstad: Overvåkingssystemet for bedrifter i Bof. 19s.

2006/44 M. Høstmark og B.O. Lagerstrøm: Undersøkelse om Arbeidsmiljø: Destruktiv atferd i arbeidslivet. Dokumentasjonsrapport. 43s.

2006/45 T.K. Schjerven og K.Å. Wass: Faglig modell og rammeverk i StatRes. 67s.

2006/46 R. Sønsterudbråten: FOB2001. Dokumentasjon av logistikk og svartjeneste. 68s.

2006/47 K. Henriksen: Utvalgsplan til konsumprisindeksens nye matvareindeks - Basert på strekkodedata. 23s.

2006/48 A.B. Thorud, D. Rafat, S. Ferstad og E. Vinju: Tverrgående revisjon i KOSTRA - Bedring av påliteligheten i nøkkeltallene. 65s.

2006/49 T. Granseth: Grensehandel. En analyse av kvaliteten av data. 48s.

2006/50 E. Engelien, H. Høie og M. Steinnes: Bygging i strandsona. Metode og resultater. 18s.

2006/51 A. Akselsen, K.I. Bøe og Ø. Sivertstøl: FD - Trygd. Dokumentasjonsrapport. Arbeidssøkere, 1.1.1992-30.4.2001. 75s.

2006/52 L. Østby: Bruk av velferdsordninger blant nyankomne innvandrere fra de nye EØS-landene i 2005. 34s.

2006/53 G. Claus: Inntekts- og formuesundersøkelsen for personlig næringsdrivende 2004. Dokumentasjon. 28s.

2006/54 J. Heldal: Logistisk regresjon - kurskompendium i byråskolens kurs SM507. 51s.

2006/55 L.H. Thingstad: Varehandelsstatistikk 2002 - omsetning etter varegruppe. 59s.

2006/56 H.Kull Brofoss og A. Barstad: Internasjonale erfaringer med områderettede tiltak i storbyer. En litteraturstudie. 101s.

2006/57 B. Bye og I. Ringdal: Disaggregering av helse-, omsorg- og utdanningstjenester i MSG6-modellen. 39s.

2006/59 Leiemarkedsundersøkelsen 2006. Dokumentasjonsrapport. 43s.

2006/60 J. Hamre og A. Vedø: Utvalgsundersøkelse om egenmeldt sykefravær. Dokumentasjon av utvalgsplanen, utvalget for 2006 og standardfeilberegninger. 50 s.

2006/61 E. C. Rauan: Undersøking om foreldrebetaling i barnehagar, august 2006. 45s.

2006/62 Indikatorer på kjemikalieområdet - Risiko for skade på helse og miljø grunnet bruk av kjemiske stoffer, fase 2. 100s.

2006/63. Lønnsstatistikk 2006. Etablering av populasjon og utvalg. Dokumentasjonsnotat. 51s.

2006/64. Bygg, anlegg og eiendomsdrift - tall og metode. 53s.

2006/65: O. Villund: Forsøk med imputering av utførte timeverk i Arbeidskraftundersøkelsen. 58 s.

2006/66. FD - Trygd Dokumentasjonsrapport. Arbeidssøkere 1.5.2001-31.12.2004. 60s.

2006/67: E. Holmøy: Non-Ponzi-Game betingelser og lukking av anvendte intertemporale likevektsmodeller. 38s.

2006/68. Kirkelig rapportering 2006 Felles- og menighetsråd. 19s.

2006/69. FD-Trygd Dokumentasjonsrapport. Stønader til enslig forsørger. 1992-2005. 45s.