

Vekter i De nasjonale reisevaneundersøkelsene: Et historisk overblikk



Vekter i De nasjonale reisevaneundersøkelsene

Et historisk overblikk

Fredrik Alexander Gregersen

Forsideillustrasjon: Shutterstock

Transportøkonomisk institutt (TØI) har opphavsrett til hele rapporten og dens enkelte deler. Innholdet kan brukes som underlagsmateriale. Når rapporten siteres eller omtales, skal TØI oppgis som kilde med navn og rapportnummer. Rapporten kan ikke endres. Ved eventuell annen bruk må forhåndssamtykke fra TØI innhentes. For øvrig gjelder [åndsverklovens](#) bestemmelser.

Tittel:	Vekter i De nasjonale reisevaneundersøkelsene: Et historisk overblikk	Title:	Weights in the Norwegian national travel survey
Forfatter:	Fredrik Alexander Gregersen	Author:	Fredrik Alexander Gregersen
Dato:	01.2017	Date:	01.2017
TØI rapport:	1548/2017	TØI Report:	1548/2017
Sider:	14	Pages:	14
ISBN elektronisk:	978-82-480-1856-8	ISBN Electronic:	978-82-480-1856-8
ISSN:	0808-1190	ISSN:	0808-1190
Finansieringskilde:	Grunnbevilgning (SIS) -TØI	Financed by:	Institute of Transport Economics
Prosjekt:	3665 – Reisevane og mobilitet - program	Project:	3665 – Klikk eller trykk her for å skrive inn tekst.Travel behaviour and mobility
Prosjektleder:	Fredrik Alexander Gregersen	Project Manager:	Fredrik A. Gregersen
Kvalitetsansvarlig:	Tom Erik Julsrud	Quality Manager:	Tom Erik Julsrud
Fagfelt:	Reisevaner og mobilitet	Research Area:	Travel behaviour and Mobility
Emneord:	Nasjonal RVU Reisevaneundersøkelsen – reisevanespørreskjema Reisevaner RVU Spørreskjemaundersøkelse Statistikk Vekter	Keywords:	National travel survey Statistics Surveys Weights

Sammendrag:

Denne rapporten gir en kort innføring i hvordan vektene i De nasjonale reisevaneundersøkelsene (RVU) er bygget opp og råd angående bruk.

Summary:

The Norwegian national travel survey do apply weights to get unbiased estimates. This report presents a short overview over the weights applied in the surveys.

Language of report: Norwegian

*Transportøkonomisk Institutt
Gaustadalleen 21, 0349 Oslo
Telefon 22 57 38 00 - www.toi.no*

*Institute of Transport Economics
Gaustadalleen 21, 0349 Oslo, Norway
Telefon 22 57 38 00 - www.toi.no*

Forord

Denne rapporten er finansiert som en del av grunnbevilgningen til TØI. Formålet med denne rapporten er at brukere som ønsker å bruke De nasjonale reisevaneundersøkelsene (RVU) skal få en oversikt over hvilke vekter som er benyttet i de ulike årgangene av undersøkelsen. Videre få noen generelle tips om bruk og hvor man kan finne mer informasjon om vektene.

Jeg vil takke Øystein Engebretsen og Ronny Klæboe for nyttige innspill og kommentarer på en tidligere versjon av denne rapporten. Tom Erik Julsrud har kvalitetssikret rapporten. Trude Kvalsvik har stått for den endelige utformingen av rapporten. Hvis man har innspill på følgende dokument ta gjerne kontakt på fag@toi.no.

Oslo, januar 2017
Transportøkonomisk institutt

Gunnar Lindberg
direktør

Frode Longva
avdelingsleder

Innhold

Sammendrag

Summary

1	Litt generell informasjon om vekter i RVU	1
2	Et regneeksempel på bruk av vekter i en spørreundersøkelse	3
3	Kort innføring i oppbygningen av vekter i RVU.....	5
4	Historisk overblikk over vekter i RVU.....	6
5	Noen tips ved vekting	9
6	Referanser.....	12
7	Vedlegg – Utregning av tabell 2.2.....	13

Sammendrag

Vekter i De nasjonale reisevaneundersøkelsene: - Et historisk overblikk

*TØI rapport 1548/2017
Forfatter: Fredrik Alexander Gregersen
Oslo 2017 14 sider*

Denne rapporten gir en kort innføring i hvordan vektene i De nasjonale reisevaneundersøkelsene (RVU) er bygget opp og råd angående bruk.

Siden den første reisevaneundersøkelsen ble foretatt i 1985 har både datainnsamlingsteknikker og bruk av undersøkelsen endret seg betydelig. Tidligere var det eksempelvis mer vanlig å se på hele landet under ett, mens man i de senere år har blitt mer opptatt av regionale forskjeller. Dette stiller andre krav til bruk av vekter for å få forventingsrette estimater. Med forventningsrett estimator menes det at en estimator ikke har systematiske avvik fra den parameteren vi ønsker å estimere.

Vektene i RVU vil i mange tilfeller bidra til og være avgjørende for å få forventingsrette estimater ved analyser basert på RVU. Imidlertid finnes det tilfeller der ukritisk bruk av vektene kan bidra til å gi forventningsskjev estimater. Det er derfor viktig at forskere som anvender RVU setter seg inn i hvordan vektene i RVU er bygget opp for den eller de årgangene av RVU man skal benytte. Denne rapporten gir en kort innføring i hvordan vektene i RVU er bygget opp i de ulike årgangene av RVU.

Summary

Weights in the Norwegian national travel survey

TØI Report 1548/2017

Author: Fredrik Alexander Gregersen

Oslo 2017 14 pages Norwegian language

The Norwegian national travel survey do apply weights to get unbiased estimates. This report presents a short overview over the weights applied in the surveys.

The Norwegian national travel survey has been conducted in 1985, 1992, 1998, 2001, 2005, 2009 and 2013/14. Each time apart from 1985 the results are presented with weights. How the weights have been estimated have varied substantially over time. This report presents a short overview over the weights for each year.

1 Litt generell informasjon om vekter i RVU

Etter det amerikanske presidentvalget i 2016 ble det debattert hvorvidt feil bruk av vekter gjorde at man ikke klarte å forutse valgresultatet. Resultatet av valget ble at Donald Trump vant på tross av at omtrent samtlige meningsmålinger predikerte at Hillary Clinton kom til å vinne valget (nrk.no/urix/usa-valg--meningsmalinger-1.13207272).

Ved spørreundersøkelser ønsker man som oftest å si noe om en hel populasjon basert på et lite utvalg av populasjonen. Ved det amerikanske valget ønsket man å gjette hvordan hele befolkningen kom til å stemme basert på et lite utvalg av befolkningen. I reisevaneundersøkelsene ønsker man å kartlegge hvordan innbyggere i Norge reiser basert på et utvalg av disse innbyggerne. Denne rapporten vil kort gå gjennom hvordan man i De nasjonale reisevaneundersøkelsene (RVU) i Norge har benyttet vekter og bør benytte dem. Vekter er ofte avgjørende for å få forventingsrette estimater basert på spørreundersøkelser. Med forventningsrett estimator menes det at en estimator ikke har systematiske avvik fra den parameteren vi ønsker å estimere.

Dette innebærer at vektene i RVU vekter opp grupper som er underrepresentert i utvalget i forhold til i populasjonen og tilsvarende vekter ned grupper som er overrepresentert. Eksempelvis har man i RVU 2013/14 regionale tilleggsutvalg, dette innebærer at man har flere respondenter fra enkelte regioner, blant annet hovedstadsområdet, enn størrelsen på populasjonen skulle tilsi. Hvis man ønsker å presentere gjennomsnittstall for hele landet vil med andre ord disse områdene være overrepresentert hvis man ikke vekter resultatene. Dette vil føre til at man ikke får forventingsrette estimater. Eksempelvis vil trolig bruk av transportmidler som ikke finnes i hele landet slik som t-bane bli overrepresentert.

Tilsvarende er det store kjønnsforskjeller i bruk av transportmidler. Flere studier både fra Norge og utlandet viser eksempelvis at kvinner kjører mindre bil enn menn (Hjorthol og Vågane 2014, Hjorthol, m. fl. 2010, Prashker, m. fl. 2008). Det kan derfor være avgjørende at man har samme prosentvise fordeling av kjønn i utvalget som i populasjonen for å estimere bruk av transportmidler i hele populasjonen basert på utvalget. Hvis den prosentvise fordelingen i utvalget og populasjonen er ulike kan man veie opp dette ved å vekte resultatene. Jeg vil komme tilbake til et regneeksempel ved bruk av vekter i kapittel 2.

RVU 2013/14 er den sjuende reisevaneundersøkelsen som er gjennomført i Norge. De seks foregående ble gjennomført i 1985, 1992, 1998, 2001, 2005 og 2009. RVU 2013/14 er den største nasjonale reisevaneundersøkelsen som har vært gjennomført i Norge, med ca 60000 intervjuer. Undersøkelsen består av et landsrepresentativt basisutvalg på 10000 intervjuer og regionale tilleggsutvalg på ca 50000 intervjuer. Det pågår i år en ny RVU som skal gå løpende frem til 2019. Hvilke vekter som vil bli benyttet i denne undersøkelsen er ikke avklart. Denne undersøkelsen er derfor ikke en del av denne rapporten.

I reisevaneundersøkelsene innhentes bakgrunnsopplysninger om intervjupersonen og husholdningen vedkommende tilhører, hvilke reiser han/hun har foretatt på registreringsdagen (daglige reiser) og lengre reiser (100 km og lengre, samt reiser til/fra Norge) som er foretatt siste måned. Undersøkelsen gir detaljerte opplysninger om daglige og lange reiser, i tillegg til en rekke bakgrunnsopplysninger om intervjupersonen og

vedkommendes husholdning (for flere detaljer se Vågane, m. fl. (2011) og Hjorthol, m. fl. (2014)). Dette er en omfattende spørreundersøkelse, som gir et rikt datamateriale og et vell av analysemuligheter og muligheter til å besvare mange problemstillinger. I hovedsak har det samme spørreskjemaet vært brukt på alle de nasjonale RVUene. Dette datamaterialet gir derfor også gode muligheter til å studere sosial endring, hva som har skjedd med befolkningens reisevaner over tid og viktige forklaringsfaktorer på eventuell endring. RVU gir også, slik det er påpekt i Gundersen (2015), innsikt i formålet med reisen, en innsikt man ikke får av rene trafikktegninger. I slike tellinger vil man kun få antall passeringer av eksempelvis bil eller sykkel.

Som i de fleste spørreundersøkelser er det mange feilkilder. I grove trekk kan man dele feilkildene inn i to grupper; 1) fra utvalget til populasjonen man studerer og 2) feilkilder i svarene fra respondentene (Fowler Jr 2013). Av feilkilder i gruppe 1 finner vi eksempelvis at enkelte aldersgrupper er underrepresentert i utvalget. Dette kan eksempelvis skyldes at enkelte aldersgrupper i mindre grad enn andre ønsker å svare på spørreundersøkelsen. Videre kan eksempelvis enkelte geografiske områder være overrepresentert i utvalget. I RVU 2013/14 er det tilleggsutvalg for enkelte byområder. I gruppe 2 er feil slik som at respondentene kan glemme hvilke reiser de har foretatt, misforstå spørsmålene eller bevisst svarer feil på enkelte spørsmål.

Hovedhensikten med vektene er å lage et representativt utvalg av hele befolkningen basert på det utvalget man har tilgjengelig. Med andre ord er vekting rettet mot enkelte feil i gruppe 1 og vil ikke rette opp feil i gruppe 2. Videre er det viktig å merke seg at ikke alle feil i gruppe 1 kan bli rettet opp av vekting. Eksempelvis kan ikke vekting rette opp systematiske forskjeller mellom respondenter som svarer på spørreundersøkelsen og de som unnlater å svare. Gitt at personer som svarer på spørreundersøkelsen eksempelvis reiser lengre enn andre, vil dette føre til forventingsskjev utvalg (Groves Robert, m. fl. 2004). Videre vil det være slik at vekting av praktiske årsaker ikke kan ta hensyn til alle faktorer som påvirker reisemønster til respondentene. Eksempelvis vet man fra tidligere forskning at inntekt påvirker valg av reisemiddel (Hsiao 2007). Dette er ikke en faktor som er inkludert i vektingen i RVU.

Hvordan vektene er utformet varierer noe over tid. For de fleste årganger finnes informasjon om hvordan vektene er utformet i nøkkelrapporten for den respektive årgangen av RVU. Merk imidlertid at vekter for årgangene 1992 og 1998 har blitt laget i ettertid på grunn av utvalgsproblematikk i disse årgangene. Dermed kan man for disse årgangene ikke finne informasjon i nøkkelrapportene om vektingen som benyttes i dag (se kapittel 4).

Denne rapporten vil i kapittel 2 presentere et regneeksempel for å forklare forskjeller i gjennomsnitt og standardavvik ved bruk av og ikke bruk av vekter i en tenkt spørreundersøkelse. Dette eksempelet er ment å gi en kort innføring i intuisjonen bak vekter i spørreundersøkelser. Deretter følger en kort innføring i oppbygging av vekter i RVU. I kapittel 4 finnes en oversikt over variabelnavn og litt informasjon om de ulike komponentene i vektene som er benyttet i de ulike årgangene av RVU. Til slutt følger noen tips ved bruk av vekter i RVU.

2 Et regneeksempel på bruk av vekter i en spørreundersøkelse

Jeg vil nå ved hjelp av et svært forenklet eksempel kort forklare hva vekter gjør med resultatene hvis man anvender pakken Weight i SPSS. Dette er gjort ettersom de fleste som analyserer RVU benytter denne pakken i SPSS. Andre pakker i SPSS eller programmer kan gi noe andre estimater, særlig gjelder dette for standardavviket. Eksempelvis kan programmer som anvender bootstrap få andre standardavvik. For en nærmere beskrivelse av bootstrap se Efron (1992). For å lese resten av dette denne rapporten trenger man ikke å sette seg inn i hva bootstrap er. For en nærmere beskrivelse av programpakker for vekter i Stata gir Kreuter og Valliant (2007) en god oversikt både over prinsipper ved vekting generelt og muligheter for vekting i Stata.

Tenke deg at du ønsker å studere kjørte kilometer en gitt dag for personer som jobber i en bedrift ved å trekke et utvalg av ansatte. Fra tidligere studier vet man at menn i gjennomsnitt kjører lengre enn kvinner. Det jobber 100 personer i bedriften og 60 % er menn og 40 % er kvinner. Du klarer imidlertid bare å inkludere 2 menn og 4 kvinner. Med andre ord «representerer» hver kvinne i undersøkelsen 10 personer mens hver mann «representerer» 30 personer. Ved bruk av oppblåste-vekter blir dermed vekten for hver mann 30 og 10 for hver kvinne. Hvis man derimot bruker ikke-oppblåste vekter blir vekten 1,8 for menn og 0,6 for hver kvinne. Ved ikke-oppblåste vekter er med andre ord summen av vekter lik antallet respondenter. Se tabell 2.1 for oversikt over vektene.

Tabell 2.1: Oversikt over utvalg og vekter

Antall personer i bedriften (populasjonen)	Antallet i utvalget	Kjønn	Andel inkludert i utvalget av populasjonen	Oppblåste-vekter	Ikke-oppblåste-vekter (proporsjonale-vekter)
60	2	M	2/60	60/2=30	(60/100)/(2/6)=1,8
40	4	K	4/40	40/4=10	(40/100)/(4/6)=0,6
Sum (n)	6			(30*2)+(10*4)=100	(2*1,8)+(0,6*4)=6

Hvis man så regner ut gjennomsnittlig kjørelengde med oppblåstvekt og ikke-oppblåstvekt får man samme resultatet altså 63 km. Derimot er standardavviket i disse tilfellene forskjellige, se tabell 2.2. Videre er også gjennomsnittet i utvalget med og uten vekter ulikt. Uten vekt er gjennomsnittlig kjørelengde 55 km. Tabell 2.2 presenterer standardavviket med oppblåst vekt, ikke-oppblåstvekt og uten vekt. For en nærmere beskrivelse av utregningene i denne tabellen, se vedlegg. For leser som ikke er kjent med pakken Weight i SPSS kan det være en fordel å regne gjennom eksempelet i vedlegg før man leser videre i denne rapporten.

Tabell 2.2: Standardavvik og gjennomsnitt

	Uten vekter	Oppblåste-vekter	Ikke-oppblåste-vekter
Standardavvik	24,29	25,44	27,73
Gjennomsnitt	55	63	63

Vi tenker oss videre at man har informasjon om bosted til arbeidstakerne som enten bor i by A eller B. Vi vet at fordelingen av kvinner og menn i disse byene fra denne bedriften er lik som i bedriften altså 60% menn og 40% kvinner i hver by. Videre tenker vi oss at vi har intervjuet en kvinne og en mann fra by A og resten kommer fra by B (se tabell 2.3).

Fortsatt er vi interessert i å studere antall kjørte kilometer i hver by. Hvis man så bruker vekten som er laget over vil man ende med både galt standardavvik og gjennomsnitt. Dette gjelder både om man velger oppblåste-vekter eller ikke-oppblåste-vekter. Grunnen til dette er at det prosentvise forholdet mellom menn og kvinner innenfor hver by i utvalget er ulikt fra resten av populasjonen.

Tabell 2.3: Fordeling på bosted blant de spurte og kjørte kilometer.

Bosted (by)	KM	Kjønn
A	100	M
B	50	M
B	50	K
B	60	K
A	40	K
B	30	K

Som man ser av dette veldig enkle eksempelet er altså vektens utforming sentralt for om den bidrar til å forbedre resultatene i analysen man ønsker å utføre. I dette eksempelet ville altså ikke vekten slik den var utformet bidratt til å forbedre resultatene hvis man studere kjøre lengde for personer i hvert av bostedene isolert.

3 Kort innføring i oppbygningen av vekter i RVU

Hovedhensikten med vekter i RVU er å lage et representativt utvalg av hele befolkningen basert på det utvalget man har fra befolkningen. De betyr i korte trekk at man prøver å få utvalget som er inkludert i en gitt årgang av RVU til å representere den faktiske befolkningen ut fra de faktorene man inkluderer i vekten. Hvis vekten, slik som i 2013/14, er basert på geografi, alder, årstid og ukedag er altså hovedhensikten med vekten å lage et representativt utvalg av hele befolkningen basert på disse faktorene. Det er imidlertid viktig å merke seg, som nevnt over, at også andre faktorer som ikke er med i vekten kan ha stor betydning for valg av reisemiddel og lengde på reisene, eksempelvis inntekt eller utdanning. Merk videre at man kan ta hensyn til lavere svarprosent i enkelte aldersgrupper ved å inkludere flere personer i grupper som historisk sett har hatt lav svarprosent. I RVU 2005 ble eksempelvis dette gjort og dermed har man i ettertid ikke vektet for alder, men kun for geografi.

Tradisjonelt har man ved hjelp av RVU primært studert trender i hele befolkningen i hele landet. Imidlertid har man i de senere år blitt mer opptatt av regionale forskjeller eller forskjeller mellom eksempelvis urbane strøk og ikke-urbane strøk innenfor en region. Vektene er i utgangspunktet ikke laget for dette formålet. Man skal derfor når man ser på delområder eller undergrupper bruke vektene med forsiktighet. Gitt at menn i en region som er brukt som enhet for vekting er underrepresentert, trenger det imidlertid ikke å være tilfellet i hele regionen. Eksempelvis kan de være overrepresentert i sentrumsområder. Hvis man deretter starter å sammenligne sentrumsområder og andre områder innenfor disse regionene kan det gi svært misvisende estimater. Man bør derfor så langt det lar seg gjøre undersøke utvalget opp mot den faktiske befolkningen hvis man undersøker delområder eller grupper.

Vekten i RVU blir presentert på to former. I den ene formen er summen av vekter lik antall respondenter i hele datasettet. Dette er det man kaller ikke-opplåste-vekter eller proporsjonale-vekter. Den andre måten å presentere vektene på er det man kaller opplåste-vekter. I dette tilfellet er summen av vekter lik antallet i personer i den befolkningen man ønsker å studere. Om vekten blir presentert som opplåst og/eller ikke-opplåst varierer fra årgang til årgang av RVU. Dette har særlig stor betydning når man regner ut standardavvik eller når man summerer kjørte kilometer eller antall reiser. Det er derfor svært viktig å merke seg hvilken type vekt som blir brukt.

Videre er det for enkelte årganger slik at man presenterer de ulike vektingskomponentene i RVU. Altså hvis vekten består av flere komponenter er delkomponentene tilgjengelig. I disse tilfellene er det viktig å merke seg at hvilken vekt man skal bruke avhenger av formålet med studien. Hvis man ønsker å sammenligne forskjeller mellom to aldersgrupper skal man eksempelvis som hovedregel ikke vekte for alder. Dette fordi det ikke tilfører noe informasjon som kan forbedre resultatene ved å gjøre dette. Hvis man bare beregner gjennomsnitt vil imidlertid ikke resultatene bli påvirket av denne vektingen. Derimot vil både standardavvik, beregning av p-verdier etc. bli påvirket av en eventuell vekting.

4 Historisk overblikk over vekter i RVU

Det har skjedd en stor utvikling innen teknologi fra 1985, da den første RVUen ble foretatt og frem til i dag. I de første RVUene ble det foretatt intervjuer ansikt til ansikt. I 2013/14 ble intervjuene foretatt via telefon. I neste RVU er det planlagt at den delvis skal baseres på internettbaserte løsninger. Denne utviklingen har også påvirket hvordan man kan få tak i respondenter. Når intervjuene ble utført ansikt til ansikt var det større krav til at grupper av de som ble intervjuet var samlet geografisk, altså det man kaller klynger (clusters). I de siste RVUene er dette ikke nødvendig. Imidlertid har svarprosenten blant respondentene vært fallende over tid. Dette har vært med å bidra til at man fortsatt ikke i dag kan si om man har et tilfeldig utvalg noe som taler for at man bør benytte vekter. Videre er det i de senere år utført tilleggsutvalg. Det vil si at man bevisst har intervjuet flere personer innenfor enkelte geografiske områder, såkalte regionale tilleggsutvalg. Dette forutsetter vekting for å sikre geografisk representativitet.

Tabell 4.1 gir en oversikt over de viktigste vektene for ulike årganger av RVU. Vekten det er anbefalt å bruke er markert med grønn tekst. Merk imidlertid at hvilken vekt man skal bruke avhenger av formålet med studien. Vekten som er markert **fet skrift** er den som gir best representativitet for hele befolkningen samlet sett. *Gråtekst* indikerer at vekten finnes, men ikke er tilgjengelig i de offentlige tilgjengelige filene.

Tabell 4.1: Oversikt over vekter i RVU for hver årgang av undersøkelsen

RVU-årgang	Vekter	Innhold (vektet for)	Kommentarer til variabel
RVU-1984-85 N=4320	Ingen vekt		
Merknader: Utvalget er relativt lite og det finnes ingen vekt for denne RVUen.			
RVU 1992 N=6000	aldf_w	Alder og fylke	Oppblåst
	aldf_ww	Alderf_w* husholdningsstørrelse	Oppblåstvekt
	aldf_v	Tilsvarende som aldf_w	Ikke oppblåstvekt
	aldf_v_v	Tilsvarende som aldf_ww. Vektet for alder, husholdnings- størrelse og fylke.	Ikke oppblåst vekt. Se Vibe (1993).
Merknader: Husholdningsutvalg vil si at man ringer til en husholdning å spør etter den personen som sist hadde bursdag i husholdningen. Det gjør at enslige er overrepresentert i utvalget. Dette kan føre til forventingsskjev utvalg (se Klæboe og Grue (1999)). Vektene her er laget i ettertid slik at resultater presentert i nøkkelrapporten avviker fra resultater basert på denne vekten. Man har i vekten man har laget i ettertid i større grad enn ved den opprinnelige vekten tatt hensyn til problematikk knyttet til husholdningsstørrelse.			
RVU 1998 N=8838	Huhw	Vekt basert på husholdningsstørrelse (hmover13) se under	
	hmover13	Antall i husholdningen over 13 og eldre	

RVU-årgang	Vekter	Innhold (vektet for)	Kommentarer til variabel
	fw	Korreksjonsfaktor for fylker	Denne er basert på antall innbyggere i hvert fylke.
	wvar	Vekt (husholdningsutvalg *korrelasjon faktor for fylker)	Se Klæboe og Grue (1999). Merk at det er en annen vekt som er brukt i nøkkelrapporten for denne RVUen (Stangeby, m. fl. 1999). wvar er ikke-opplåstvekt.
<p>Merknader: Samme metode som 1992, det vil si husholdningsutvalg. Merk at bruk av wvar til en viss grad retter opp at det er et forventingsskjevt utvalg som følge av husholdningsutvalg (se Klæboe og Grue (1999)). Merk at vektene her er laget i ettertid slik at resultater presentert i nøkkelrapporten avviker fra resultater basert på denne vekten. Man har i vekten man har laget i ettertid i større grad enn ved den opprinnelige vekten tatt hensyn til problematikk knyttet til husholdningsstørrelse.</p>			
RVU 2001 N=20752	Vekt	Kommune	Se Denstadli og Hjorthol (2002)
<p>Merknader: Laget egen tilleggsundersøkelse for byer se felles RVU. Vektet for geografiske områder, ikke vektet etter alder og kjønn.</p>			
Felles RVU 2001 N=52990	vekt_1	Utvalgsvekt - etter utvalgsprosent og ukedag	Ikke opplåst
	vekt_2	Alle ukedager: Oppblåsing til antall bosatte 13 år+	Oppblåst vekt
	vekt_3	Virkedager: Oppblåsing til antall bosatte 13 år+ (folk uten reise mand-ferd vekt 0)	Oppblåst vekt
<p>Merknader: I tillegg til den ordinære RVUen fra 2001 ble det laget en egen tilleggsundersøkelse for Oslo og Akershus i 2001. De fleste spørsmålene var overlappende med spørsmålene i den ordinære RVUen, mens enkelte spørsmål ikke er helt identiske. Det er ProSam som har foretatt denne undersøkelsen. Felles RVU er en betegnelse på begge disse undersøkelsene slått sammen.</p>			
RVU 2005 N=17514	utvalgsvekt	Vektet for geografi (vektsoner)	Ikke opplåst. Se Denstadli, m. fl. (2006)
	befolkningsvekt	Tilsvarende som utvalgsvekt	Oppblåst
<p>Merknader: Ingen</p>			
RVU 2009 N=28922	utvalgsvekt	Kjønn, alder, geografi (vektingssoner) og intervjudtidspunkt	Se Vågane, m. fl. (2011)
	utvalgsvekt_trondheim	Utvalgsvekt som brukes for spørsmål som ikke er stilt Trondheimsutvalget, bl.a. lange reiser	
<p>Merknader: Ingen</p>			
RVU 2013/2014 N=60813 (28390 lange reiser)	utvalgsvekt	Vekting etter geografi, alder, årstid og ukedag. Ikke opplåst vekt.	Se Hjorthol, m. fl. (2014)
	hoved_vekt	Oppblåsningsvekt etter geografi, alder, årstid og ukedag - vektet sum=bosatte 13 år+ i Norge 2014	

RVU-årgang	Vekter	Innhold (vektet for)	Kommentarer til variabel
	utvalgsvekt_lang	Vekting (for lange reiser) etter geografi, alder, årstid og ukedag - vektet sum=antall intervjuer med spm om lange reiser	
	hoved_vekt_lang	Oppblåsningsvekt (for intervjuer om lange reiser) etter geografi, alder, årstid og ukedag - vektet sum=bosatte 13 år+ i Norge 2014	
Merknader: Merk at alder, årstid og ukedag er justert innenfor hver geografisk sone. Dette er ikke gjort i tidligere RVUer.			

5 Noen tips ved vekting

Som nevnt flere ganger tidligere i denne rapporten, er hovedhensikten med vekten i RVU å lage et representativt utvalg av hele befolkningen. Hvis man studerer hele befolkningen i en gitt årgang av RVU skal man som hovedregel bruke vekten markert med **fet skrift** i tabell 4.1. Hvis derimot man ser på deler av befolkningen eller ulike geografiske delområder, skal man bruke vekten med stor grad av forsiktighet.¹ Det betyr imidlertid ikke at man bør anvende RVU-data uten vekter eller andre tilpasninger. I disse tilfellene kan man eksempelvis vurdere å lage egne vekter, eller kjøre regresjoner der man heller legger inn ulike vektingsparametere som forklaringsvariabler. Hvilken løsning som er best avhenger av problemstillingen man ønsker å besvare.

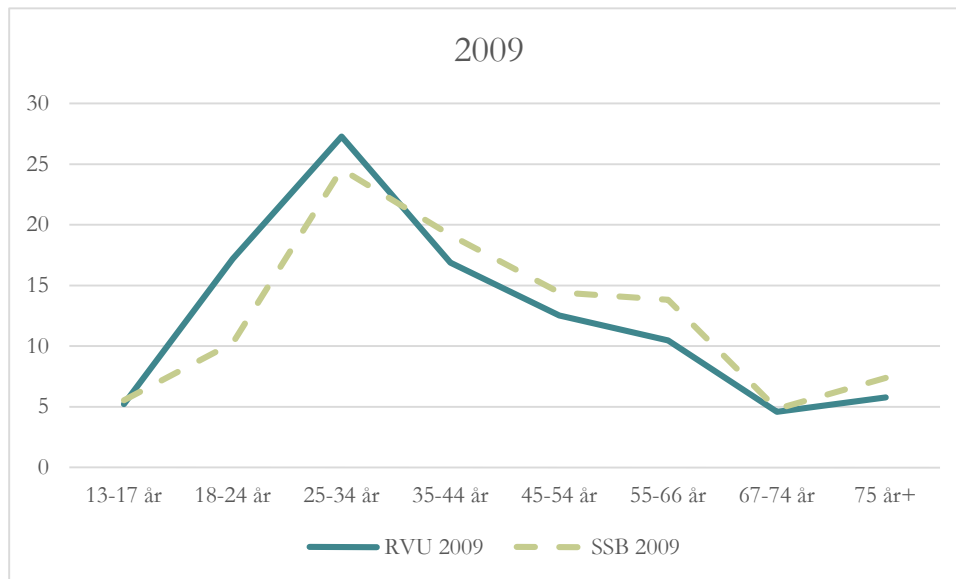
Uavhengig om man ser på delutvalg av RVU eller ikke bør man som hovedregel ikke vekte for de faktorene man studerer. Om man eksempelvis ønsker å studere forskjellen mellom to kommuner for RVU 2001, vil vekten ikke hjelpe til å forbedre resultatene. Dette fordi utvalgsvekten kun justerer for forskjeller mellom kommuner for å lage et representativt utvalg av hele befolkningen. Det er nettopp dette man ønsker å studere, altså forskjellene mellom kommuner. Videre skal man heller ikke vekte hvis man er interessert i å studere forskjeller innenfor en kommune. Det er derfor viktig i enhver studie der man benytter RVU-data å undersøke hva vekten i RVU faktisk justerer for. Videre bør man lese over nøkkelrapporten for RVUen for å se om det er foretatt grep i selve undersøkelsen som reduserer behovet for vekter. Et eksempel på slike tiltak, også fra RVU 2001, er at man i selve undersøkelsen har prøvd å intervju flere i grupper som oftere ikke svarer på spørreundersøkelser. Slike tiltak kan gjøre at det blir mindre behov for vekting.

I så stor grad som mulig bør man også sjekke om fordelingen av sentrale variabler er i overensstemmelse med de faktiske verdier. Eksempelvis, hvordan fordelingen av alder og kjønn er innenfor det området man studerer sammenlignet med den faktiske befolkningen oppgitt i SSB sin befolkningsdatabase. Et eksempel på dette er hentet fra RVU 2009 og 2013. I årgangene før 2009 ble aldersvekter laget for hele landet. I 2013/14 ble alle vektningskomponentene justert slik at de var riktige innenfor hver vektningszone. Figur 5.1 a) viser den stiplede linjen den faktiske befolkningen i 2009 i Oslo. Videre viser den blå linjen aldersfordelingen i RVU vektet for Oslo. Som figuren viser er unge overrepresentert i RVU 2009 i Oslo. Dette peker i retning av at hvis man skal studere Oslo for RVU i 2009 bør man vurdere å lage en egen vekt som retter opp denne skjevheten. Figur 5.1 b) viser aldersfordelingen for RVU 2013 i Oslo etter vekting. I motsetning til i figur a) er aldersfordelingen i RVU og SSB her sammenfallende. Dette tyder altså på at aldersfordelingen i Oslo i RVU etter vekting er korrekt og man trenger derfor ikke å lage nye vekter. Merk imidlertid at avhengig av problemstillingen man skal studere, kan det være andre faktorer enn alder man bør justere for i sine analyser. Dette kan igjen peke mot at man må lage nye vekter også for RVU 2013, når man ser på delutvalget Oslo.

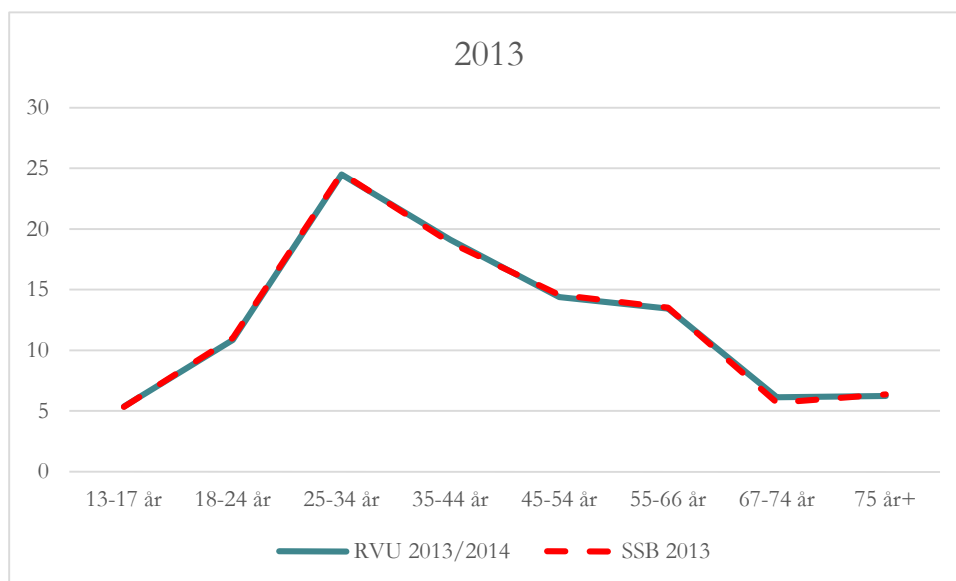
¹ Merk at det er store endringer i vektingen RVU2013/14 i forhold til tidligere årganger som gjør den bedre egnet til å se regionale forskjeller i forhold til tidligere årganger av RVU.

Figur 5.1: Eksempel fra Oslo RVU 2009 og 2013/14:

a) Aldersfordeling RVU 2009



b) Aldersfordeling RVU 2013/14



Videre er viktig å skille mellom utvalgsstørrelse og vekten. Ved proporsjonale (ikke oppblåste) vekter er summen av vekter lik antall respondenter (i hele utvalget), men det er ikke tilfellet i delutvalg. Man bør derfor alltid oppgi antall respondenter i tillegg til eventuell vekt. Ved bruk av reisefilen bør man også skille mellom antall reiser, antall respondenter og vekt. Merk at enkelte respondenter i utvalget kan ha foretatt ingen reiser og dermed ikke kommer med i reisefilen. Hvis man skal ha med alle respondenter bør man dermed anvende personfilen og se om det er noen respondenter som ikke har foretatt noen reiser, og eventuelt legge disse til for å få korrekt antall respondenter.

Kort oppsummert vil vektene i RVU i mange tilfeller bidra til å gi forventingsrette estimater. Imidlertid finnes det tilfeller der ukritisk bruk av vektene kan bidra til å gi forventingsskjevne estimater. Det er derfor svært viktig at forskere som anvender RVU setter seg inn i hvordan vektene i RVU er bygget opp for den eller de årgangene av RVU man skal benytte. Nøkkelrapporten for den gitte årgangen av RVU gir en oversikt over hvordan vektene er bygget opp. For årgangene 1992 og 1998 har det blitt utarbeidet nye vekter etter at nøkkelrapporten har blitt trykket og det kan for disse årgangene være hensiktsmessig å også lese Klæboe og Grue (1999).

6 Referanser

- J. M. Denstadli, Ø. Engebretsen, R. Hjorthol og L. Vågane. (2006). Den nasjonale reisevaneundersøkelsen 2005 - nøkkelrapport *TØI* (Vol. 844/2006).
- J. M. Denstadli og R. Hjorthol. (2002). Den nasjonale reisevaneundersøkelsen 2001 - nøkkelrapport *TØI* (Vol. 588/2002).
- B. Efron. (1992). *Bootstrap methods: another look at the jackknife*: Springer.
- F. J. Fowler Jr. (2013). *Survey research methods*: Sage publications.
- M. Groves Robert, J. Fowler Floyd, P. Couper Mick Jr, M. Lepkowski James og T. R. Singer Eleanor. (2004). *Survey Methodology*. Hoboken: NJ: John Wiley & Sons.
- F. Gundersen. (2015). Sammenligning av nasjonale reisevaneundersøkelser, regionale reisevaneundersøkelser og trafikktellinger *TØI* (Vol. 1430/2015).
- R. Hjorthol, Ø. Engebretsen og T. P. Uteng. (2014). Den nasjonale reisevaneundersøkelsen 2013/14 - nøkkelrapport *TØI* (Vol. 1383/2014).
- R. Hjorthol og L. Vågane. (2014). Allocation of tasks, arrangement of working hours and commuting in different Norwegian households *Journal of transport geography* (Vol. 35, pp. 75-83).
- R. J. Hjorthol, L. Levin og A. Sirén. (2010). Mobility in different generations of older persons: The development of daily travel in different cohorts in Denmark, Norway and Sweden *Journal of transport geography* (Vol. 18, pp. 624-33).
- M.-H. Hsiao. (2007). Long-Term Aggregate Travel Demand Analysis: A Comparison between Japan and Taiwan *Proceedings of the Eastern Asia Society for Transportation Studies* (Vol. 2007, pp. 121-): Eastern Asia Society for Transportation Studies.
- R. Klæboe og B. Grue. (1999). The Sampling Method in the 1998 Norwegian Travel Survey. An analysis of estimator properties using WesVar *TOI* (Vol. 1135/1999).
- F. Kreuter og R. Valliant. (2007). A survey on survey statistics: What is done and can be done in Stata *Stata Journal* (Vol. 7, pp. 1).
- J. Prashker, Y. Shiftan og P. Hershkovitch-Sarusi. (2008). Residential choice location, gender and the commute trip to work in Tel Aviv *Journal of transport geography* (Vol. 16, pp. 332-41).
- I. Stangeby, J. V. Haukeland og A. Skogli. (1999). Reisevaner i Norge i 1998 *TØI* (Vol. 418/1999).
- N. Vibe. (1993). Norske reisevaner, Dokumentasjonsrapport for den landsomfattende reisevaneundersøkelsen 1991-92 *TØI* (Vol. 1983/1993).
- L. Vågane, I. Brechan og R. Hjorthol. (2011). Den nasjonale reisevaneundersøkelsen 2009 - nøkkelrapport *TØI* (Vol. 1130/2011).
- L. Vågane, I. Brechan og R. Hjorthol. (2011). Den nasjonale reisevaneundersøkelsen 2009 - nøkkelrapport *TØI* (Vol. TØI rapport 1130/2011).

7 Vedlegg – Utrekning av tabell 2.2

Jeg vil nå gå gjennom hvordan pakken Weight i SPSS har regnet ut standardavvik og gjennomsnitt i tabell 2.2. I SPSS vil en vekt på 2 ved bruk av pakken Weight si at man dupliserer en observasjon. Med andre ord vil tabell A.1 a) og b) være like hvis man anvender syntaksen «weight by WeightX» på tabell A.1 a) nedenfor.

Tabell A.1 a)

<i>Km</i>	<i>Kjønn</i>	<i>WeightX</i>
10	0	2
20	1	3

Tabell A.1b)

<i>Km</i>	<i>Kjønn</i>
10	0
10	0
20	1
20	1
20	1

Formelen for utregning av gjennomsnitt blir dermed gitt av formel A.1. Der w_i er vekten for observasjon i , mens x_i er verdien på observasjon i altså km i tabell A.1 a). Hvis man ikke vektet er vekten 1, altså blir vekten lik antall observasjoner. I formelen A.1 ville dermed nevneren blitt antall observasjoner (n).

Formel A.1: Gjennomsnitt

$$\bar{x} = \frac{\sum_{i=1}^W w_i \times x_i}{\sum_{i=1}^W w_i}$$

Standardformel for standardavvik er gitt ved formel A.2, der n er antall observasjoner, x_i er observasjon i og gjennomsnittet av alle observasjoner er gitt ved \bar{x} .

Formel A.2: Standardavvik

$$\sigma_x = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$

Vektet standardavvik blir dermed gitt av formel A.3.

Formel A.3: Vektet standardavvik

$$\sigma_w = \sqrt{\frac{\sum_{i=1}^n w_i \left(x_i - \frac{\sum_{i=1}^n w_i \times x_i}{\sum_{i=1}^n w_i} \right)^2}{\sum_{i=1}^n w_i - 1}}$$

Ved hjelp av formel A.1 og A.3 blir standardavvik og gjennomsnitt regnet ut for tabell 2.2, tabell 2.2 inkludert noe mellomregning er presentert i tabell A.2.

Tabell A.2: Utregning av tabell 2.2: Gjennomsnitt og standardavvik

	<i>KM</i>	<i>Kjønn</i>	<i>Oppblåste-vekter</i>	<i>Ikke - oppblåstevekter</i>	<i>Oppblåste-vekter</i>	<i>Ikke - oppblåstevekter</i>
					$\sum_{i=1}^n w_i \left(x_i - \frac{\sum_{i=1}^n w_i \times x_i}{\sum_{i=1}^n w_i} \right)^2$	
	100	M	30	1.8	41070	2464.2
	50	M	30	1.8	5070	304.2
	50	K	10	0.6	1690	101.4
	60	K	10	0.6	90	5.4
	40	K	10	0.6	5290	317.4
	30	K	10	0.6	10890	653.4
Sum			100	6	64100	3846
Standardavvik	24,29				25,44	27,73
Gjennomsnitt						
$\bar{x} = \frac{\sum_{i=1}^w w_i \times x_i}{\sum_{i=1}^w w_i}$	55				63	63

Transportøkonomisk institutt (TØI)

Stiftelsen Norsk senter for samferdselsforskning

TØI er et anvendt forskningsinstitutt, som mottar basisbevilgning fra Norges forskningsråd og gjennomfører forsknings- og utredningsoppdrag for næringsliv og offentlige etater. TØI ble opprettet i 1964 og er organisert som uavhengig stiftelse.

TØI utvikler og formidler kunnskap om samferdsel med vitenskapelig kvalitet og praktisk anvendelse. Instituttet har et tverrfaglig miljø med rundt 70 høyt spesialiserte forskere.

Instituttet utgir tidsskriftet Samferdsel med 10 nummer i året og driver også forskningsformidling gjennom TØI-rapporter, artikler i vitenskapelige tidsskrifter, samt innlegg og intervjuer i media. TØI-rapportene er gratis tilgjengelige på instituttets hjemmeside www.toi.no.

TØI er partner i CIENS Forskningscenter for miljø og samfunn, lokalisert i Forskningsparken nær Universitetet i Oslo (se www.ciens.no). Instituttet deltar aktivt i internasjonalt forsknings-samarbeid, med særlig vekt på EUs rammeprogrammer.

TØI dekker alle transportmidler og temaområder innen samferdsel, inkludert trafiksikkerhet, kollektivtransport, klima og miljø, reiseliv, reisevaner og reiseetterspørsel, arealplanlegging, offentlige beslutningsprosesser, næringslivets transport og generell transportøkonomi.

Transportøkonomisk institutt krever opphavsrett til egne arbeider og legger vekt på å opptre uavhengig av oppdragsgiverne i alle faglige analyser og vurderinger.

Besøks- og postadresse:

Transportøkonomisk institutt
Gautstadalléen 21
NO-0349 Oslo

22 57 38 00
toi@toi.no
www.toi.no