

Sammenligning av metoder for beregning av årsdøgntrafikk



Notatnr
Forfattere

SAMBA/06/15
Ola Haug
Magne Aldrin

Dato

3. juli 2015

Norsk Regnesentral

Norsk Regnesentral (NR) er en privat, uavhengig stiftelse som utfører oppdragsforskning for bedrifter og det offentlige i det norske og internasjonale markedet. NR ble etablert i 1952 og har kontorer i Kristen Nygaards hus ved Universitetet i Oslo. NR er et av Europas største miljøer innen anvendt statistisk-matematisk modellering og har et senter for forskningsdrevet innovasjon, Statistics for Innovation – (sfi)², med finansiering fra Norges forskningsråd. Det jobbes med et bredt spekter av problemstillinger, for eksempel finansiell risiko, jordobservasjon, estimering av fiskebestander og beskrivelse av geologien i petroleumsreservoarer. NR er ledende i Norge innen utvalgte deler av informasjons- og kommunikasjonsteknologi. Innen IKT-området har NR innsatsområdene e-inkludering, informasjonssikkerhet og smarte informasjonssystemer.

NRs visjon er forskningsresultater som brukes og synes.

Tittel	Sammenligning av metoder for beregning av årsdøgntrafikk
Forfattere	Ola Haug <ola.haug@nr.no> Magne Aldrin <magne.aldrin@nr.no>
Dato	3. juli 2015
Publikasjonsnummer	SAMBA/06/15

Sammendrag

Faktormetoden og basiskurvemetoden er to ulike teknikker som begge beregner gjennomsnittlig døgntrafikk over året (ÅDT) basert på registrering av antall kjøretøy over en avgrenset tidsperiode. Vi har vurdert de to metodene opp mot faktisk ÅDT for ulike kategorier av simulerte korttidstillinger. Ut fra simuleringsdataene som danner grunnlag for analysen, viser det seg at faktormetoden er den universelt beste av de to idet den presterer best både for tellinger som begrenser seg til noen timer, og i situasjoner hvor telleperioden strekker seg over flere uker.

Forsidebilde: ©iStock.

Emneord	Faktormetoden, basiskurvemetoden, relativ feil, absolutt feil
Målgruppe	SINTEF Teknologi og samfunn, Statens Vegvesen Vegdirektoratet
Tilgjengelighet	Åpen
Prosjekt	Metodikk for ÅDT-estimering
Prosjektnummer	220533
Satsningsområde	Statistikk for teknologi, industri og forvaltning
Antall sider	20
© Copyright	Norsk Regnesentral

Innhold

1	Innledning	7
2	Simuleringseksperiment	7
3	Analyse	8
4	Resultater	9
5	Diskusjon	12
6	Konklusjon	15
	Referanser	16
A	Metodisk usikkerhet i faktormetoden - formelverk	18
	A.1 Bakgrunn	18
	A.2 Teori	19
	A.3 En anvendelse	20

1 Innledning

På oppdrag fra Statens Vegvesen Vegdirektoratet har Norsk Regnesentral (NR) i samarbeid med SINTEF Teknologi og samfunn sammenlignet to ulike metoder for beregning av årsdøgntrafikk (ÅDT) fra korttidstillinger av vegtrafikk. De to metodene, kalt faktormetoden og basiskurvemetoden, er begge per i dag oppført som alternativer for beregning av trafikkparametere i Veileder i trafikkdata, Håndbok V714 (Vegdirektoratet, 2014). Vegdirektoratet ønsker en vurdering av om det finnes en universelt optimal metode blant de to, eller eventuelt om det for gitte tellesituasjoner er slik at den ene er å foretrekke framfor den andre.

2 Simuleringseksperiment

Sammenligningsstudien er bygget opp som et simuleringseksperiment hvor det for tellepunktår med høy telledekning (>95%) genereres en rekke syntetiske korttidstillinger i henhold til totalt 15 ulike kategorier av forhåndsdefinerte telleoppsett (case). For hver simulering av korttidstillinger beregnes ÅDT med de to metodene.

De 15 kategoriene av korttidstillinger varierer fra enkelttimer via døgn til registreringer over én til flere uker. Kategoriene er definert av SINTEF Teknologi og samfunn og Vegdirektoratet i fellesskap og beskrevet i tabell 1.

Case	Beskrivelse
1	1 uke
2	2 uker
3	3 uker
4	1 + 1 + 1 + 1 uker (1 uke i hvert kvartal)
5	1 + 1 + 1 + 1 + (1) uker, ekstra uke ved høy SDT (1 uke i hvert kvartal + 1 uke i juli)
6	4 uker i starten av året
7	2 + 2 timer (0700 – 0900 og 1500 – 1700)
8	2 + 2 + 2 timer (0700 – 0900, 1100 – 1300 og 1500 – 1700)
9	1 time (makstime)
10	6 timer (start kl 0700 eller kl 1500)
11	1 døgn
12	1 døgn i januar
13	2 + 2 uker
14	1 måned (30 dager etter hverandre)
15	8 uker i starten av året

Tabell 1. Kategorier av korttidstilleoppsett.

Datasettet som produseres fra simuleringseksperimentet er videre å oppfatte som todelt med hensyn på valg av faktor- eller basiskurver gjennom den logiske variabelen 'Ran-

dom': For Random=FALSE brukes kurveklasser som bestemt av Vegdirektoratet for hvert registreringspunkt ut fra dets trafikkprofil, mens for Random=TRUE er det trukket en tilfeldig kurveklasse som brukes gjennomgående over et helt case.

For hver kombinasjon av case (1, 2, ..., 15) og Random (FALSE/TRUE) genereres 13×1000 korttidstellesituasjoner. Mellom ulike simuleringer varieres registreringspunktnummeret, og telleperioden trekkes i henhold til det aktuelle caset. For hver simulering estimeres ÅDT for faktor- og basiskurvemetoden, samt *faktisk* ÅDT (beregnet som telledøgntrafikk, TDT, dvs gjennomsnittlig døgntrafikk over alle talte døgn i hvert tellepunktår). Ved å la registreringspunktene inngå som en variabel i simuleringen, vil ulike registreringspunkter opptre med ulik frekvens i simuleringseksperimentet.

Programvaren som brukes i simuleringseksperimentet er utviklet av Triona. Faktormetoden er en manuell metode bygget rundt et sett av standard variasjonskurver for ulike vegtyper. Basiskurvemetoden er implementert og i rutinemessig drift som en del av trafikkdatabanken NorTraf. I bunnen av denne metoden ligger funksjoner opprinnelig skrevet av NR i C#, men som i dette prosjektet er sydd sammen av Triona til et javabibliotek.

3 Analyse

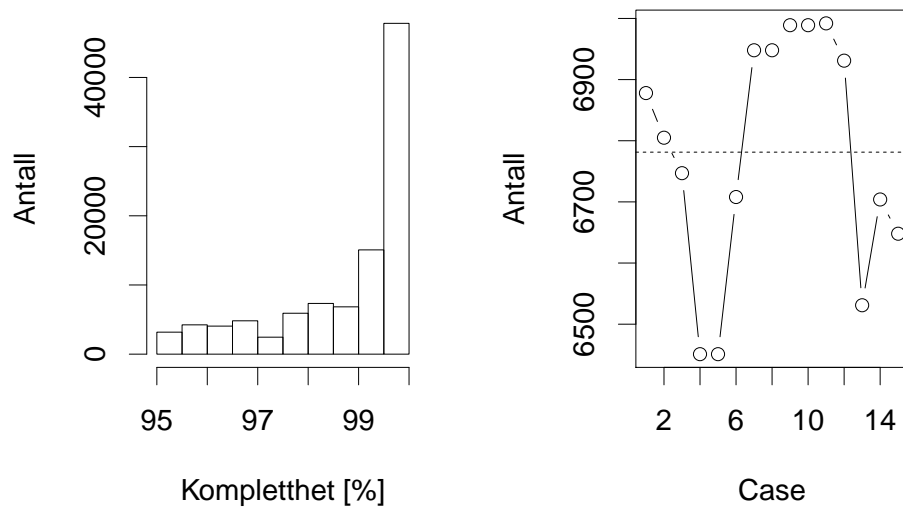
I analysen av resultatene fra simuleringseksperimentet har vi utelatt tellepunktår med kompletthetsverdi mindre enn 95%. Dette reduserer effekten av den skjevheten som introduseres ved at fasit ÅDT er beregnet som TDT, mens faktor- og basiskurvemetoden begge leverer estimater for gjennomsnittlig døgntrafikk over alle årets dager, ÅDT.

I det opprinnelige simuleringsmaterialet fra Triona foreligger det enkelte ekstreme avvik mellom fasitverdi og estimatene for ÅDT. Vegdirektoratet har funnet at avvikene går igjen for bestemte registreringspunkter, og at de gjerne er koblet til en tidsforskyvning i grunnlagsdataene. Totalt dreier det seg om 57 punkter av i alt 390. Alle simuleringer for disse punktene er tatt ut av den videre analysen. De gjenværende 333 registreringspunktene opptre i fra 60 til 1294 enkeltsimuleringer i sammenligningsstudien. I figur 1 er fordelingen av kompletthetsverdier i det endelige datamaterialet vist som et histogram (venstre panel). Høyre panel viser hvor mange simuleringer som inngår i hvert case i det endelige datamaterialet. I gjennomsnitt består et case av 6781 enkeltsimuleringer, varierende fra 6451 for case 4 og 5 til 6992 for case 11.

Vi har vurdert de to metodene ut fra avviksmålene gjennomsnittlig relativ absoluttfeil (*MRAE*) og gjennomsnittlig relativ kvadratfeil (*RMRSE*), definert som

$$MRAE = \frac{1}{N} \sum_{i=1}^N \frac{|y_i - y_{i,fasit}|}{y_{i,fasit}} \quad (1)$$

$$RMRSE = \sqrt{\frac{1}{N} \sum_{i=1}^N \left(\frac{y_i - y_{i,fasit}}{y_{i,fasit}} \right)^2} \quad (2)$$



Figur 1. Telledekning (venstre panel) og antall simuleringer for hvert case (høyre panel) etter fjerning av tellepunkter med feilaktige data og tellepunktår med lav dekning.

hvor N er antall simuleringer (enten for et case eller totalt), $y_{i,fasit}$ er fasit ÅDT for simulering i og y_i representerer ÅDT-estimer enten fra faktormetoden eller basiskurve-metoden. Lavest verdi indikerer minst avvik fra fasit og er dermed best. I tillegg har vi brukt relativ feil (RE_i) og absolutt feil (AE_i) for hver simulering i for å vurdere hvorvidt metodene over- eller underestimerer fasit ÅDT,

$$RE_i = \frac{y_i - y_{i,fasit}}{y_{i,fasit}} \quad (3)$$

$$AE_i = y_i - y_{i,fasit} \quad (4)$$

Det er gjort separate analyser for tilfellene Random=TRUE og Random=FALSE.

4 Resultater

Verdier for avviksmålene $MRAE$ og $RMRSE$ er vist i tabellene 2 og 3 for tilfellene Random=FALSE og Random=TRUE. Om vi først betrakter tilfellet hvor valg av kurveklasser er gjort på bakgrunn av kunnskap om registreringspunktene (Random=FALSE), ser vi hvordan telleoppsettet påvirker estimatene. Ser vi spesifikt på $MRAE$ for faktormetoden i tabell 2, ser vi hvordan avviket avtar fra case 1 (1 telleuke sammenhengende) til case 3 (3 telleuker sammenhengende), og videre til case 4 hvor nedgangen er betydelig når 4 telleuker spres over året. Dette er i tråd med hva vi ville forvente. Ved å legge til ytterligere en telleuke i juli (case 5) oppnås kun en marginal forbedring i forhold til case 4. Sammenhengende telling i 4 uker (case 6) eller 30 dager (case 14) gir langt dårligere

resultater enn når tellingene spres (case 4), og det hjelper heller ikke å øke den sammenhengende telleperioden til 8 uker (case 15). I case 13 telles to uker første halvår og to uker andre halvår, og tallene viser igjen at det er bedre med spredte tellinger enn med en kontinuerlig telleperiode av tilsvarende lengde. Tendensen er tilsvarende for oppsettene med times- og døgntellinger (case 7-12), selv om avvikene som sådan naturlig nok er høyere her. Tellinger kun i makstimen (case 9) gir høyest avvik etterfulgt av 2 sammenhengende timer i morgen- og ettermiddagsrushet (case 7) som i sin tur er litt verre enn å telle 6 sammenhengende timer med start enten kl 0700 eller kl 1500 (case 10). Langt bedre er det å fordele de 6 timene på tre bolker utover dagen (0700-0900, 1100-1300 og 1500-1700) (case 8), noe som gir et resultat på linje med å telle i 24 påfølgende timer (case 11).

Case	<i>MRAE</i> -faktor	<i>MRAE</i> -basis	<i>RMRSE</i> -faktor	<i>RMRSE</i> -basis
1	0.061	0.087	0.096	0.137
2	0.057	0.081	0.092	0.129
3	0.054	0.085	0.087	0.193
4	0.028	0.029	0.043	0.041
5	0.026	0.028	0.040	0.040
6	0.059	0.125	0.092	0.359
7	0.142	0.179	0.189	0.228
8	0.116	0.151	0.159	0.195
9	0.284	0.286	0.370	0.360
10	0.131	0.144	0.182	0.199
11	0.116	0.105	0.159	0.148
12	0.126	0.139	0.189	0.196
13	0.040	0.052	0.064	0.103
14	0.052	0.077	0.080	0.171
15	0.054	0.102	0.085	0.244
Totalt	0.091	0.113	0.154	0.205

Tabell 2. Avviksmål *MRAE* og *RMRSE* for faktor- og basiskurve metoden for ulike oppsett av korttidstallinger i tilfellet Random=FALSE.

Ser vi så på avvikene mellom faktor- og basiskurve metoden, finner vi at faktormetoden er best i 14 og 12 av de totalt 15 telleoppsettene målt etter henholdsvis *MRAE* og *RMRSE*. Faktormetoden kommer også best ut etter begge kriteriene når vi ser på avvik for alle simuleringene under ett ('Totalt' i tabellen). Riktignok er ikke utslagene like store for alle telleoppsettene, men den gjennomgående tendensen er klar.

For simulering med tilfeldig tilordning av kurveklasser (Random=TRUE) er bildet mer nyansert, se tabell 3. Nå er faktormetoden best etter henholdsvis *MRAE* og *RMRSE* i bare 7 og 6 av de totalt 15 telleoppsettene, men den har fortsatt lavest avvik for alle simuleringene sett under ett. I tråd med hva vi ville forvente, er avvikene generelt større enn når kurveklassene velges ut fra egenskapene til registreringspunktene.

For å vurdere om de to metodene over- eller underestimerer årsdøgntrafikken, har vi be-

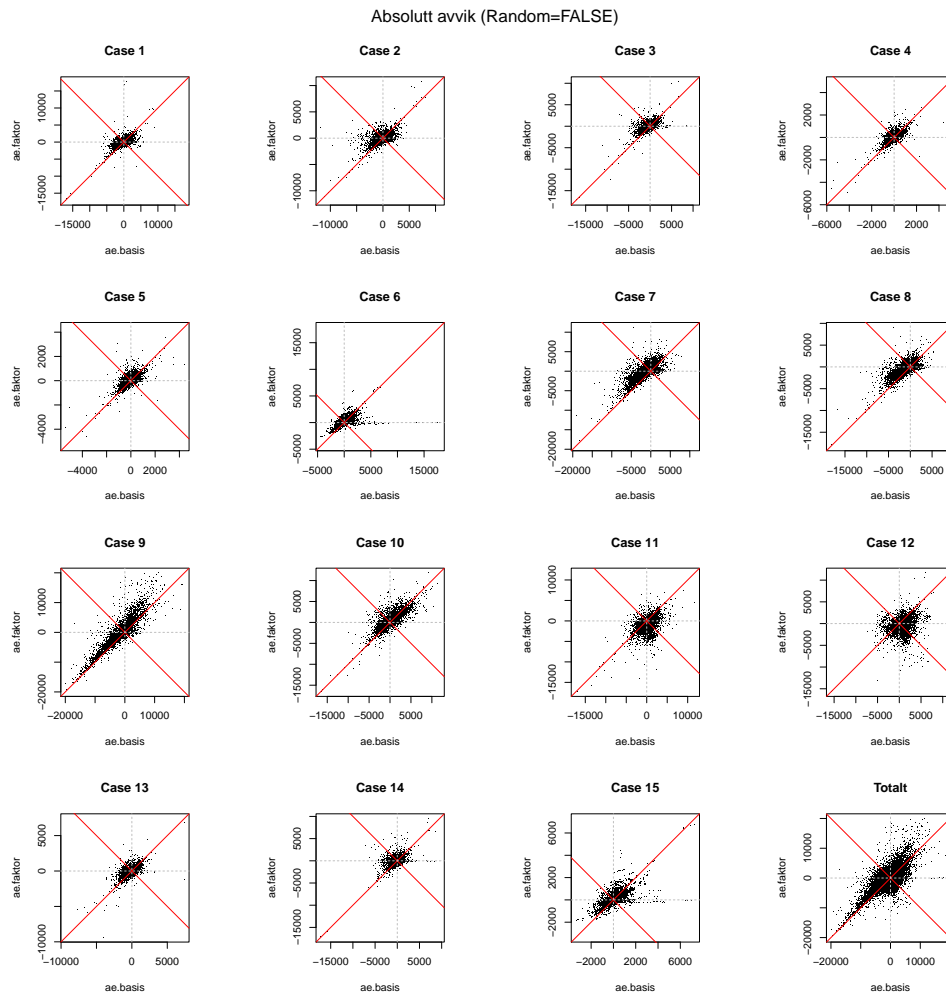
Case	<i>MRAE</i> -faktor	<i>MRAE</i> -basis	<i>RMRSE</i> -faktor	<i>RMRSE</i> -basis
1	0.125	0.153	0.179	0.199
2	0.166	0.087	0.318	0.136
3	0.246	0.103	0.304	0.143
4	0.041	0.071	0.061	0.091
5	0.041	0.031	0.061	0.044
6	0.104	0.110	0.153	0.149
7	0.194	0.219	0.265	0.276
8	0.160	0.192	0.230	0.252
9	0.322	0.860	0.414	1.222
10	0.774	0.228	1.021	0.293
11	0.319	0.447	0.408	0.566
12	0.213	0.186	0.278	0.248
13	0.118	0.070	0.202	0.098
14	0.170	0.077	0.317	0.122
15	0.093	0.086	0.141	0.135
Totalt	0.209	0.198	0.369	0.393

Tabell 3. Avviksmål *MRAE* og *RMRSE* for faktor- og basiskurve metoden for ulike oppsett av korttidstillinger i tilfellet Random=TRUE.

regnet relativt og absolutt avvik (RE_i og AE_i). Figurene 2 og 3 viser spredningsdiagram for AE_i for tilfellene Random=FALSE og Random=TRUE. Det er for hvert panel snakk om veldig mange punkter som dels er plottet oppå hverandre, så figurene skal tolkes med forsiktighet. Det er heller ikke felles akseverdier mellom de ulike panelene. Vi ser likevel at det for begge metodene og alle casene er enkelte svært store avvik fra fasit ÅDT, både positive og negative. De røde linjene deler panelene i fire kvadranter. Punkter som faller innenfor de to kvadrantene langs positiv og negativ x-akse representerer enkeltsimuleringer hvor faktormetoden har lavest avvik, mens for punkter som faller i kvadrantene langs y-aksen er basiskurve metoden best.

Den relative feilen gir også innblikk i hvordan estimatene plasserer seg i forhold til fasit ÅDT. Figurene 4 og 5 viser fordelingen til RE_i for henholdsvis faktor- og basiskurve metoden i tilfellet Random=FALSE, mens figurene 6 og 7 gir tilsvarende oversikt for Random=TRUE. Her opereres det med samme utstrekning på x-aksen for sammenhørende kategorier av telleoppsett i de to metodene, og de vertikale stiplede linjene markerer yttergrensene i det aktuelle datasettet. Siden RE_i angir det relative avviket, er histogrammene også sammenlignbare mellom de ulike kategoriene. Ut fra definisjonen av RE_i vil -1 være en nedre grense for denne størrelsen.

Om vi først betrakter telleoppsettene for tilfellet Random=FALSE, synes det som om histogrammene i form er relativt like for de to metodene. De vertikale stiplede linjene avslører imidlertid at basiskurve metoden oftere enn faktormetoden er den som definerer utstrekningen av x-aksen ved å spenne ut hele denne, dvs at avvikene er større for denne metoden. Dette er i tråd med resultatene vist i *MRAE*- og *RMRSE*-kolonnene i tabell 2.



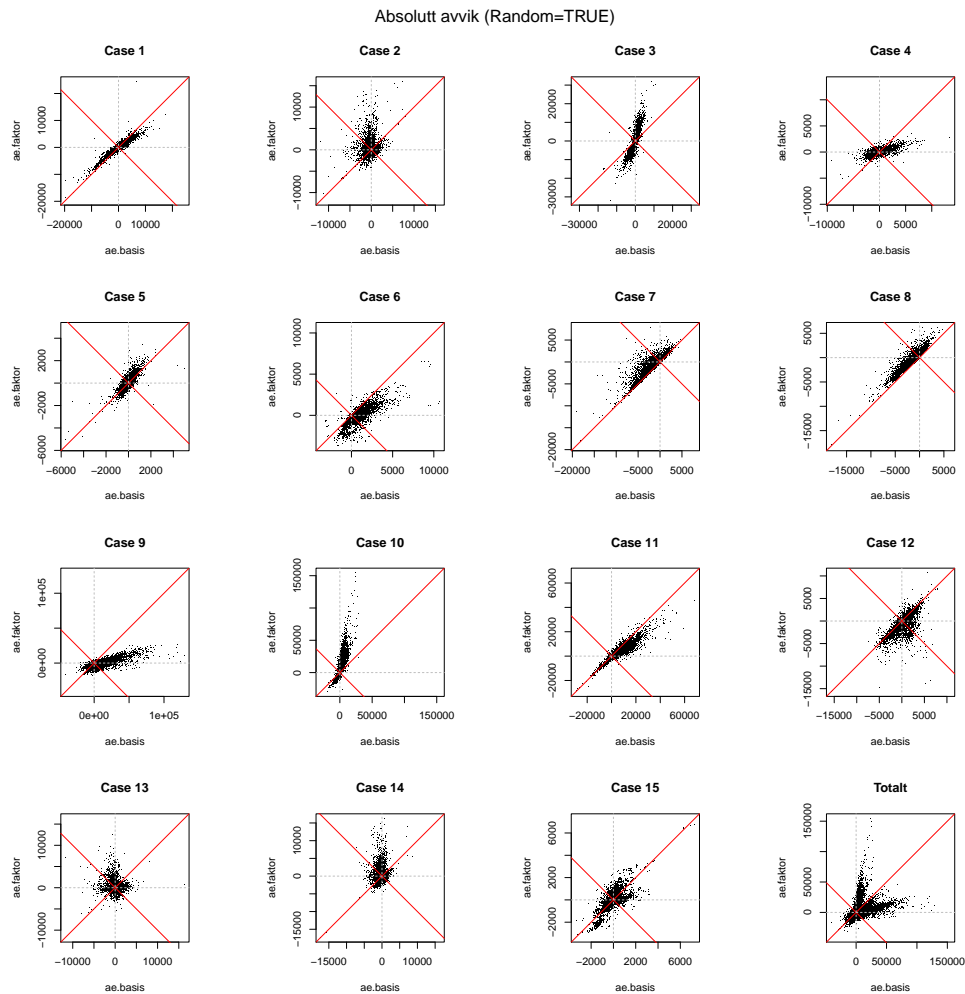
Figur 2. Absolutt avvik for tilfellet Random=FALSE. Basiskurve metoden langs x-aksen, faktormetoden langs y-aksen.

For Random=TRUE er bildet et annet. Her avviker de to metodene med hensyn på formen til histogrammene, men også når det gjelder hvilket feildatasett som har størst utstrekning (de vertikale linjene). Denne prestasjonsutjevningen mellom metodene er i tråd med hva vi fant for *MRAE* og *RMRSE* gjengitt i tabell 3.

5 Diskusjon

Gjennom simuleringseksperimentet er det avdekket ulike typer uregelmessigheter. Dette gjelder både resultater fra programvaren som Triona leverer og feil knyttet til telledataene som går inn i beregningene.

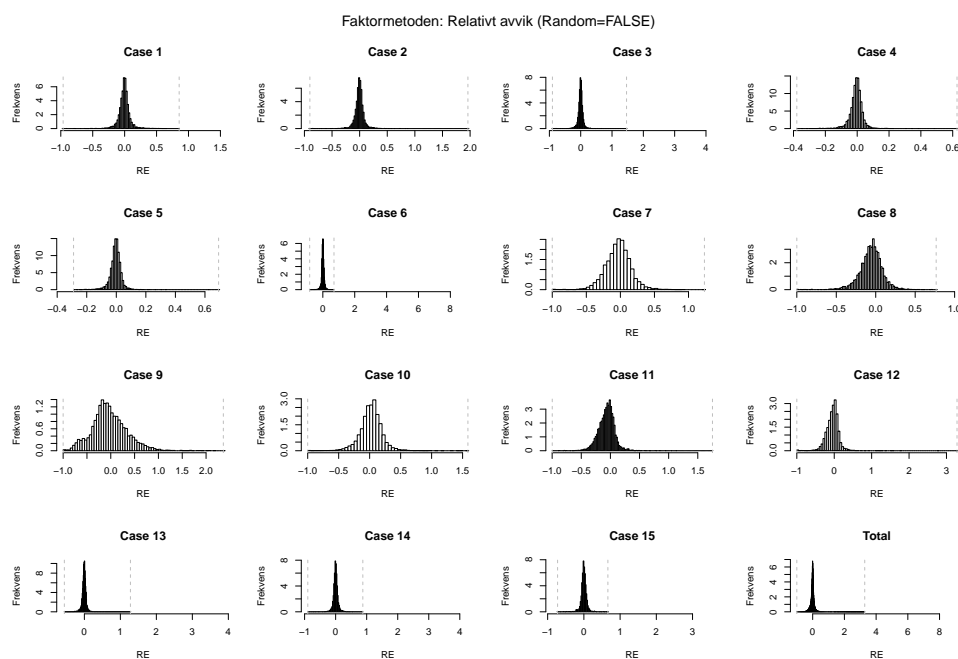
Trionas beregning av telledekning (komplette dager) er inkonsistent for enkelte av tellepunktårene. For eksempel er det for et spesifikt tellepunkt (2000006) registrert en dekning på 2.73% i 2008. Dette tilsvarer om lag 10 dager med tellinger, mens nærmere undersøkelser foretatt av Vegdirektoratet viser at tellinger foreligger for 60 dager i januar og februar



Figur 3. Absolutt avvik for tilfellet Random=TRUE. Basiskurve metoden langs x-aksen, faktormetoden langs y-aksen.

2008 svarende til en kompletthet på rundt 16%. Triona på sin side opererer med 62 telle dager for dette tellepunktåret i sitt datamateriale, så her er det åpenbart noe som ikke henger helt sammen. Flere av telleoppsettene strekker seg utover 10 dager (case 2, 3, 4, 5, 6, 13, 14, 15), og noen av dem krever også en fordeling av tellingene over hele året (case 4, 5, 13). Selv med en telledekning på 16% vil disse siste casene umulig kunne genereres for et tellepunkt som kun har tellinger i årets to første måneder. Likevel produserer programvaren til Triona ÅDT-estimer for begge metodene i tillegg til faktisk ÅDT-verdi. Dette gjør oss i tvil om algoritmen Triona har implementert for å trekke ut simulerte korttidstellingene fra faktiske telldata virker slik den er tenkt. Spesielt ville det være interessant å få greie på hvilke telleverdier som faktisk blir brukt for tidspunkter hvor det ikke foreligger tellinger.

Vegdirektoratet har sett nærmere på beregning av telledekningen med utgangspunkt i egne data. For de fleste tellepunkter og år later det til å være overensstemmelse mellom verdiene som produseres av programvaren til Triona og deres tall. For enkelte tellepunktår er det imidlertid til dels store avvik (opptil 20 prosentpoeng). Betydningen av avviken-

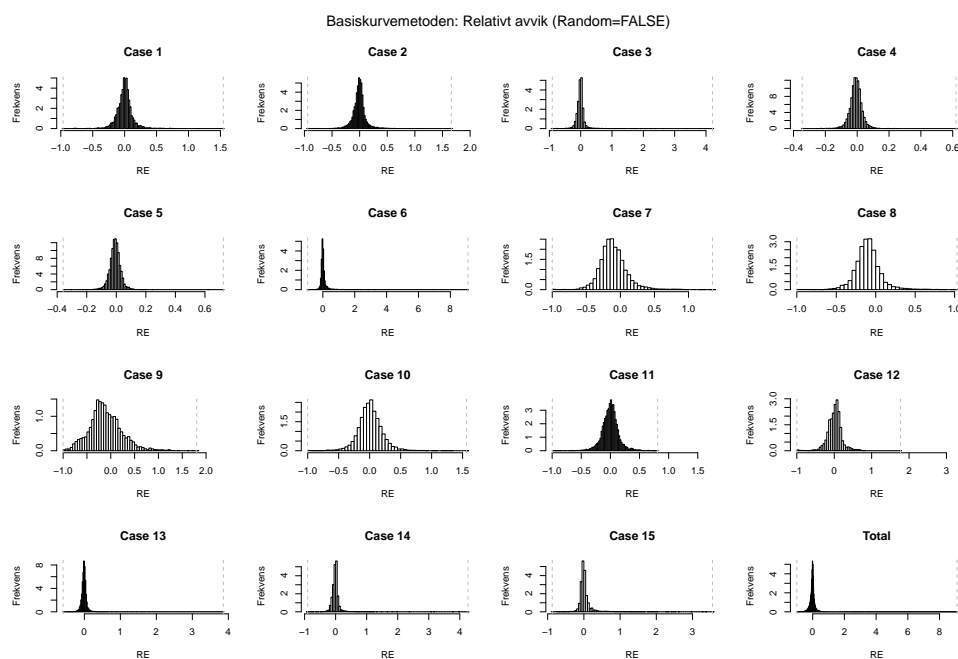


Figur 4. Relativt avvik for ulike kategorier av telleoppsett samt 'Total' ved bruk av faktormetoden for tilfellet Random=FALSE.

de beregning av telledekning i Trionas programvare er trolig liten for tilnærmet komplette telleserier. Et lite forbehold må vi likevel ta all den tid vi ikke vet hvilke telleverdier programmet bruker for eventuelt manglende telletimer. For 100% komplette telleserier er problemstillingen ikke aktuell. En stikkprøveanalyse gjort av Vegdirektoratet for tellepunktår med kompletthetsverdi over 80% viser neglisjerbare avvik mellom Trionas verdi for telledekning og Vegdirektoratets tall.

Når det gjelder telldataene i seg selv, har Vegdirektoratet tatt for seg enkeltsimuleringer hvor avvikene mellom estimert fasitverdi for ÅDT (TDT) og ÅDT-estimatene fra faktor- eller basiskurvemetoden er betydelige. De mest ekstreme avvikene er gjerne knyttet til bestemte tellepunkter, og det viser seg at det ofte er en tidsforskyvning i grunnlagsdataene i disse tilfellene. Slike tidsforskyvninger oppstår gjerne i overgangen mellom sommer- og vintertid (og motsatt) og vedvarer inntil telleutstyret igjen blir synkronisert mot faktisk kalendertid. For et tellepunkt på Lysaker utenfor Oslo viser et plott av gjennomsnittlig årlig trafikkvolum for alle døgnets timer at rushet i 2008 er anslagsvis 5 timer forsinket i forhold til årene 2009 og 2010, se figur 8. Slike feilregistreringer av telletimer er kritisk negative for beregningen av trafikkparametere.

Store avvik i AE_i eller RE_i for både faktor- og basiskurvemetoden kan være en indikasjon på at noe er feil i telldataene (for eksempel er tidsforskyvning fortsatt mulig selv om telledekningen er høy). Det er imidlertid vanskelig å trekke noen slutning om hvor store avvik som må påregnes for en gitt simulering all den tid vi ikke har tilgjengelig tilhørende estimater for det teoretiske standardavviket.



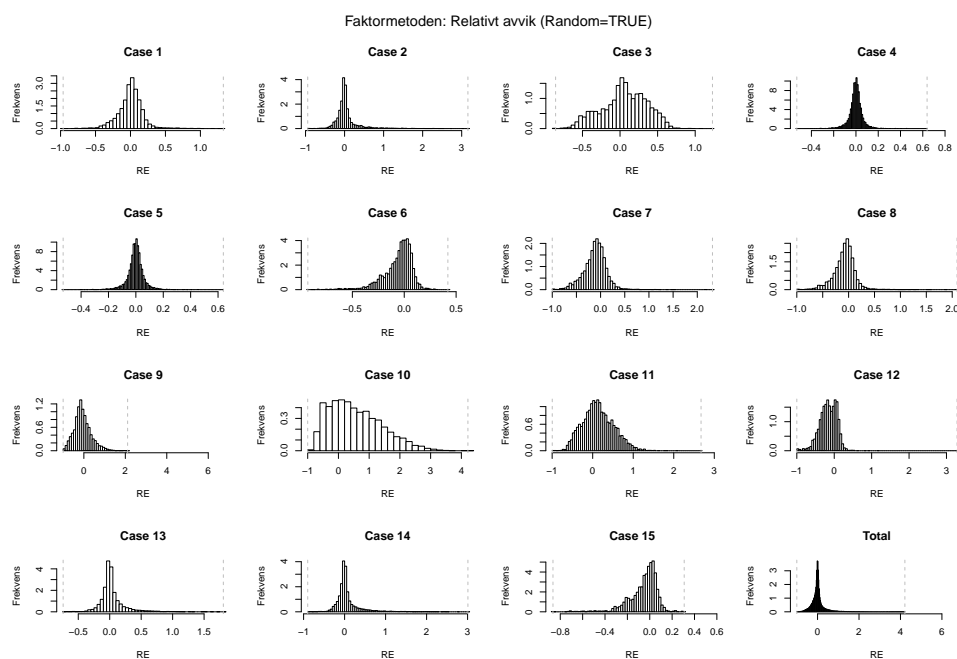
Figur 5. Relativt avvik for ulike kategorier av telleoppsett samt 'Total' ved bruk av basiskurvevremetoden for tilfellet Random=FALSE.

6 Konklusjon

Resultatene i kapittel 4 viser at faktormetoden gjennomgående er bedre enn basiskurvevremetoden til estimering av årstdøgntrafikk, ÅDT. I tilfellet hvor det velges kurveklasser ut fra kjennskap til trafikkprofilen til hvert enkelt tellepunkt (Random=FALSE), er faktormetoden bedre enn basiskurvevremetoden for et klart flertall av de forhåndsdefinerte korttidstillekategoriene målt både etter *MRAE* og *RMRSE*. Og for de få tilfellene hvor basiskurvevremetoden kommer best ut, er denne kun marginalt bedre enn faktormetoden. For simuleringer hvor kurveklassene velges tilfeldig (Random=TRUE) presterer de to metodene omtrent jevngodt, men dette tilfellet er ikke så interessant ettersom estimering av trafikkparametere typisk vil foregå med bruk av den kunnskap som er tilgjengelig om tellepunktene. Kategoriene av tellesituasjoner som er undersøkt i denne studien dekker både situasjoner med svært kort tellehistorikk (timer) og situasjoner hvor telldata er tilgjengelig over lengre perioder (uker). Det er derfor grunnlag for å si at faktormetoden utgjør en universelt beste metode for beregning av årstdøgntrafikk.

Grunnen til at basiskurvevremetoden kommer dårlig ut vurdert etter *MRAE* og *RMRSE*, er ikke fullt ut kartlagt. Grafiske framstillinger av absolutt og relativt avvik for enkeltsimuleringer antyder imidlertid at basiskurvevremetoden produserer flere store avvik fra fasit ÅDT enn hva som er tilfellet for faktormetoden. Disse store enkeltavvikene for gitte korttidstillekategorier bidrar kritisk til å ødelegge for det samlede avviket (*MRAE* og *RMRSE*).

For å finne årsaken til hvorfor basiskurvevremetoden gir slike utslag, må man gå inn i de



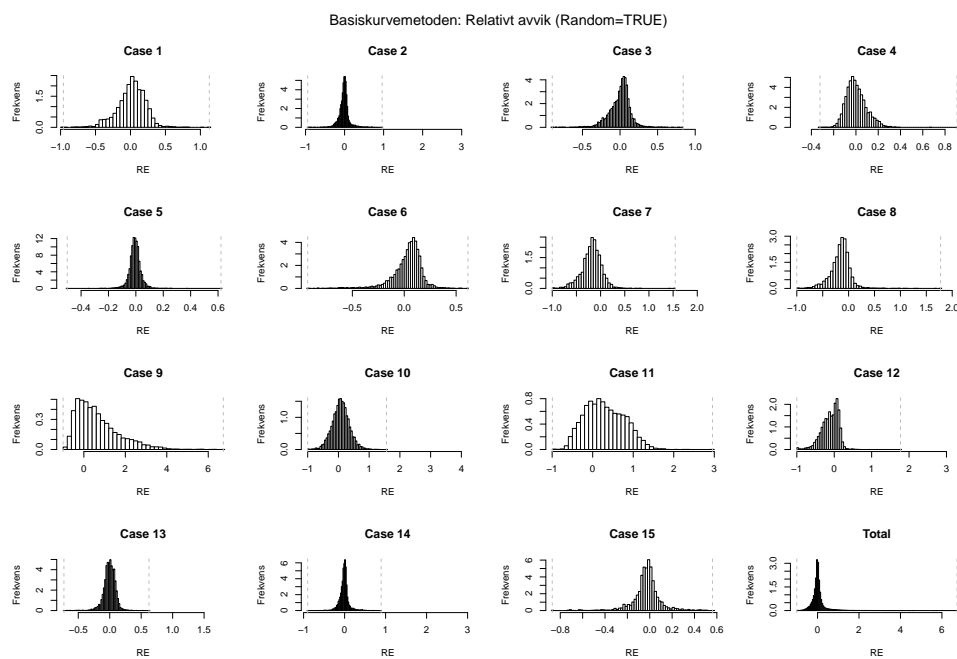
Figur 6. Relativt avvik for ulike kategorier av telleoppsett samt 'Total' ved bruk av faktormetoden for tilfellet Random=TRUE.

aktuelle tellesituasjonene og analysere nærmere hvordan basiskurvetoden håndterer disse. Det kan for eksempel ligge uavdekkede metodiske svakheter bak, eller det kan være feil i telldataene som slår uheldig ut. Det bør også nevnes at den versjonen av basiskurvetoden som er brukt i simuleringseksperimentet ble kodet på nytt i et annet programmeringsspråk av Triona i forbindelse med denne sammenligningsstudien.

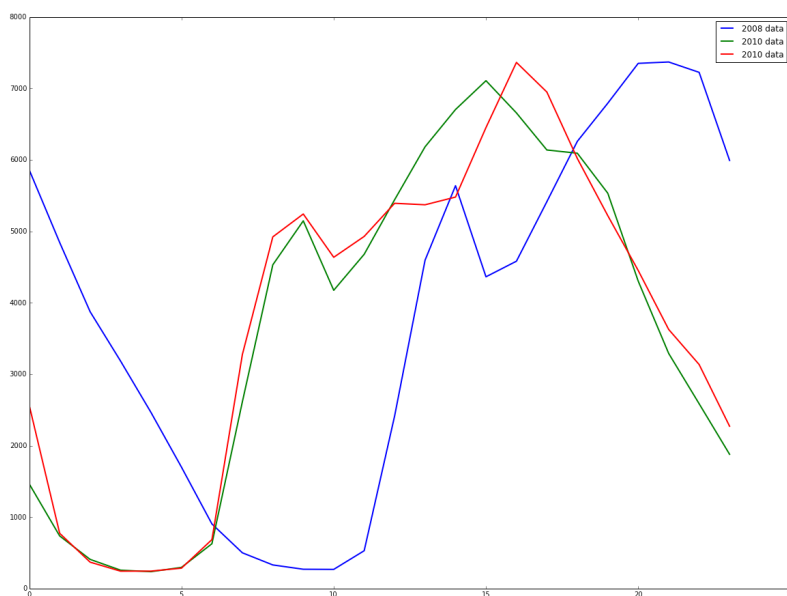
Referanser

Vegdirektoratet (2011). *Veileder innsamling og beregning av trafikkdata til støykartlegging*. Statens Vegvesen Vegdirektoratet. VD rapport nr 48.

Vegdirektoratet (2014). *Veileder i trafikkdata, håndbok V714*. Statens Vegvesen Vegdirektoratet. ISBN: 978-82-7207-630-5.



Figur 7. Relativt avvik for ulike kategorier av telleoppsett samt 'Total' ved bruk av basiskurvemetoden for tilfellet Random=TRUE.



Figur 8. Tidsforskyvning i telledata for registreringspunkt på Lysaker som beregnet av SINTEF. Kurvene viser gjennomsnittlig årlig trafikkvolum over døgnet for årene 2008, 2009 og 2010.

A Metodisk usikkerhet i faktormetoden - formelverk

A.1 Bakgrunn

Registrering og beregning av trafikkparametre (som ÅDT) er forbundet med ulike former for feil og usikkerhet. For det første er det begrensninger i nøyaktigheten ved registrering av kjøretøy. Måletekniske feil opptrer såvel for automatisk måleutstyr som ved manuelle tellinger. Dernest vil tidspunktet for trafikkregistreringen kunne medføre det vi kaller tilfeldige feil. Trafikkuhell, været og ulike arrangementer er alle eksempler på faktorer som kan påvirke det aktuelle trafikkbildet og dermed også tellingene. Til sist er det slik at selv om kjøretøyregistreringene var eksakte og uten feil, ville tallene for ÅDT likevel være forbundet med en metodisk usikkerhet. Størrelsen på denne usikkerheten er knyttet til den aktuelle beregningsmodellen (for eksempel faktor- eller basiskurve metoden), men generelt er det slik at usikkerheten avtar desto større andel av trafikken som telles. I ekstremtilfellet at alle årets timer er talt, blir den metodiske usikkerheten for ÅDT lik null.

De ulike feiltypene nevnt over er omtalt i kapittel 4.5.4 i Veileder for innsamling og beregning av trafikkdata til støykartlegging (Vegdirektoratet, 2011). Der uttrykkes den metodiske usikkerheten ved beregning av ÅDT fra faktormetoden som

$$U_i = \sqrt{(x_i(n))^2 + (y_i(m))^2 + (z_i(k))^2} \quad (\text{A.1})$$

hvor i angir faktorkurvesettet og

$$\begin{aligned} x_i &= \text{usikkerheten ved beregning av døgntrafikk fra } n \text{ timer} \\ y_i &= \text{usikkerheten ved beregning av ukedøgntrafikk fra } m \text{ dager} \\ z_i &= \text{usikkerheten ved beregning av ÅDT fra } k \text{ uker.} \end{aligned} \quad (\text{A.2})$$

Tabeller over $x_i(n)$, $y_i(m)$ og $z_i(k)$ er gitt i vedlegg 3 i Vegdirektoratet (2014). Ifølge tabellteksten angir verdiene utvalgsusikkerhet i prosent for beregning av 95% konfidensintervall. Det betyr at $x_i(n)$, $y_i(m)$ og $z_i(k)$ representerer relative usikkerheter. Utover dette er vi ikke kjent med hvordan tallene er framkommet og vil derfor måtte ta verdiene for gitt. Generelt er den metodiske usikkerheten i ÅDT-estimatene bestemt ut fra informasjonsinnholdet i telledataene. Dette er igjen påvirket både av tidspunktet for registreringen og lengden på registreringsperioden. Ved å telle flere timer (eller døgn eller uker), så vil usikkerheten gå ned. Men ikke alle timer bidrar med like mye informasjon. For eksempel vil registreringer i morgen- eller ettermiddagsrushet føre til en lavere usikkerhet enn registreringer tatt over en like lang periode på natta når trafikken typisk utgjør en mindre andel av det totale døgnvolumet. Tabellverdiene i Vegdirektoratet (2014) er basert på at kjøretøyregistreringene er gjennomført i perioder med mest trafikk. Har vi registrering for kun én time skal det være den timen i døgnet med mest trafikk, har vi registrering i to timer skal det være de to timene i døgnet med mest trafikk, og så videre.

A.2 Teori

Utgangspunktet for formelen (A.1) for usikkerhet er uttrykket for beregning av ÅDT fra korttidstillinger av trafikken. Siden kjøretøy typisk registreres over en begrenset tidsperiode, må det gjøres en skalering av tellingene for å få estimater som er representative for et gjennomsnittsdøgn slik ÅDT er ment å beskrive. Den samlede skaleringsfaktoren er uttrykt gjennom kurver for døgnvariasjon (timesoppløsning), ukevariasjon (ukedagsoppløsning) og årsvariasjon (ukenummeroppløsning). Ulike typer av trafikkvariasjonsmønstre lar seg representere gjennom i alt syv ulike sett av slike faktorkurver. Årsdøgntrafikken for et gitt sett av faktorkurver beregnes ut fra dette som

$$\text{ÅDT} = \frac{V}{d \cdot u \cdot a} \quad (\text{A.3})$$

hvor V er registrert antall kjøretøy i telleperioden og hvor produktet $d \cdot u \cdot a$ utgjør den samlede skaleringsfaktoren. Faktoren d angir talletimenes andel av døgnvolumet (ut fra normert trafikk), u representerer telledøgnets normerte trafikk relativt til gjennomsnittlig ukedøgntrafikk, og a angir telleukens normerte trafikk i forhold til gjennomsnittlig uketraffikk over året. Med normert trafikk forstås historisk trafikkvariasjonsmønster for veger som tilhører den aktuelle gruppen av faktorkurver.

Når vi i det følgende utleder bakgrunnen for formelen (A.1), ser vi kun på den usikkerheten som ligger i at d , u og a har en usikkerhet knyttet til seg. Usikkerhet i registrert antall kjøretøy V , derimot, sorterer under det som innledningsvis og i tråd med Vegdirektoratet (2011) er definert som tilfeldige feil og inngår *ikke* i beregningene. Det er likevel viktig å være oppmerksom på at bidraget fra denne komponenten kan være vel så stort som den metodiske usikkerheten, og dette må tas hensyn til når man vurderer den totale usikkerheten i ÅDT.

Vi betrakter altså V som gitt (konstant) og representerer usikkerheten i faktorene d , u og a ved å betrakte disse størrelsene som stokastiske variabler med tilhørende forventningsverdier μ_d , μ_u og μ_a og standardavvik σ_d , σ_u og σ_a . Generelt gjelder for en funksjon av stokastiske variabler at uttrykk for forventningsverdi og varians kan tilnærmes via en Taylorutvikling rundt forventningsverdiene til enkeltvariablene. For variansen er det oftest tilstrekkelig å ta med første ordens ledd i Taylorrekka. For en funksjon $f(\mathbf{X})$ bestående av p stokastiske variabler, $\mathbf{X} = (X_1, X_2, \dots, X_p)$, kan variansen dermed skrives

$$\sigma_f^2 = \text{Var}(f(\mathbf{X})) \simeq \sum_{i=1}^p \sum_{j=1}^p \left(\frac{\partial f}{\partial x_i} \right) \left(\frac{\partial f}{\partial x_j} \right) \sigma_{i,j} \quad (\text{A.4})$$

hvor $\sigma_{i,j}$ betegner kovariansen mellom X_i og X_j og hvor tilfellet $i = j$ representerer variansen. Av spesiell interesse for ÅDT gitt som i (A.3) er uttrykket for variansen til den inverse av en stokastisk variabel X_1 med standardavvik σ_1 . Fra (A.4) følger at det relative standardavviket til $f(\mathbf{X}) = 1/X_1$ er gitt som

$$\frac{\sigma_f}{|f(\mathbf{X})|} \simeq \sqrt{\left(\frac{\sigma_1}{X_1} \right)^2} \quad (\text{A.5})$$

Forutsetter vi at variablene d , u og a i (A.3) er ukorrelerte slik at $\sigma_{i,j} = 0$ for $i \neq j$ i (A.4), kan uttrykket i (A.5) utvides til

$$\frac{\sigma_{\hat{\text{ÅDT}}}}{|\hat{\text{ÅDT}}|} = \sqrt{\left(\frac{\sigma_d}{\mu_d}\right)^2 + \left(\frac{\sigma_u}{\mu_u}\right)^2 + \left(\frac{\sigma_a}{\mu_a}\right)^2}. \quad (\text{A.6})$$

Sammenholder vi (A.6) med den opprinnelige formelen for usikkerhet gitt i (A.1), gjenkjenner vi U_i som den relative usikkerheten $\frac{\sigma_{\hat{\text{ÅDT}}}}{|\hat{\text{ÅDT}}|}$. Videre tilsvarer $x_i(n)$ den relative usikkerheten $\frac{\sigma_d}{\mu_d}$, mens $y_i(m) = \frac{\sigma_u}{\mu_u}$ og $z_i(k) = \frac{\sigma_a}{\mu_a}$. Med en slik fortolkning betegner dermed usikkerhetene i (A.2) det man innenfor statistikkfaget mer presist omtaler som *variasjonskoeffisienter* definert som forholdet mellom standardavvik og forventningsverdi.

A.3 En anvendelse

Årsdøgntrafikken gitt i (A.3) er uttrykk for den gjennomsnittlige døgntrafikken over året. I henhold til det såkalte sentralgrenseteoremet innenfor statistikken kan det derfor argumenteres for at ÅDT asymptotisk følger en normalfordeling. For en normalfordelt stokastisk variabel X med ukjent forventning μ og kjent standardavvik σ_0 gjelder generelt at et konfidensintervall for μ på nivå α er gitt som

$$(\hat{\mu} - u_{\frac{\alpha}{2}} \cdot \sigma_0, \hat{\mu} + u_{\frac{\alpha}{2}} \cdot \sigma_0) \quad (\text{A.7})$$

hvor $\hat{\mu}$ er en punkt estimator for μ og $u_{\frac{\alpha}{2}}$ er $\frac{\alpha}{2}$ -kvantilen i standard normalfordelingen. Et 95% konfidensintervall tilsvarer $\alpha = 5\%$ med $u_{0.025} = 1.96$. Som en tilnærming angis ofte et 95% konfidensintervall som punkttestimatet for parameteren ± 2 standardavvik.

På side 62 i avsnitt 3.1.2 i Veileder i trafikkdata, Håndbok V714 (Vegdirektoratet, 2014) er det i Eksempel 4 vist hvordan et 95% konfidensintervall for ÅDT-verdien kan bestemmes som $\widehat{\text{ÅDT}} \pm U_i \cdot \widehat{\text{ÅDT}}$. En sammenligning med uttrykket for et nivå- α konfidensintervall i (A.7) viser at faktoren tilsvarende $u_{\frac{\alpha}{2}}$ -kvantilen ikke inngår eksplisitt i disse beregningene. Imidlertid kan det være at $u_{\frac{\alpha}{2}}$ -kvantilen er bakt inn i verdiene for $x_i(n)$, $y_i(m)$ og $z_i(k)$ i (A.1). Fortolkningen av disse størrelsene blir da en litt annen enn i avsnitt A.2 ved at $u_{\frac{\alpha}{2}}$ også kommer inn som en faktor ($x_i(n) = u_{\frac{\alpha}{2}} \cdot \frac{\sigma_d}{\mu_d}$, etc). Ved å innlemme kvantilen i tabellverdiene mister man fleksibilitet i usikkerhetsberegningene ved at alle konfidensintervaller låses til et fast nivå ($\alpha = 5\%$ gir kun mulighet for å beregne 95% intervaller).

Som nevnt i avsnitt A.1, er vi ikke kjent med hvordan verdiene i tabellene for beregning av usikkerhet er framkommet. Vi kan dermed heller ikke si nøyaktig hva de representerer. Men dersom $\frac{\alpha}{2}$ -kvantilen *ikke* er trukket inn i tabellverdiene og vi samtidig forutsetter at usikkerheten U_i i (A.1) faktisk representerer en variasjonskoeffisient slik utledningen i avsnitt A.1 gir grunn til å tro, så innebærer det at konfidensintervallet i Eksempel 4 beregnes som $\widehat{\text{ÅDT}} \pm \hat{\sigma}_{\widehat{\text{ÅDT}}}$. Dette tilsvarer i så fall et 68% konfidensintervall (punkttestimat pluss/minus ett standardavvik), og ikke 95%.