

NOTAT

Til Trafikdage 2009

Vedr. Modernisering af Transportvaneundersøgelsen

Fra Hjalmar Christiansen, DTU Transport

august 2009

1 Kort intro om Transportvaneundersøgelsen

Transportvaneundersøgelsen (ofte forkortet som TU), har til formål at kortlægge danskernes transportvaner, principielt defineret som al persontransport indenfor landets grænser.

Metoden er et stort antal interview med danskere (10-84 år) om transportadfærden "dagen i går". Interview gennemføres både pr internet (ca. 20 % af data) og via telefon (80 %). Interviewpersonerne udvælges repræsentativt ved hjælp af CPR-registret og resultaterne opregnes efter geografi, alder og køn.

Undersøgelsen er unik, fordi det er den eneste store, danske undersøgelse med kobling af faktisk transportadfærd til en lang række baggrundsvariable. I international sammenhæng er undersøgelsen unik, fordi den kortlægger alle ture med koordinater for hvert rejsemål.

Transportvaneundersøgelsen er gennemført med stort set samme indhold siden 1992, dog med en afbrydelse i 2004-5. Ofte omtales data fra årene 1992-2003 (172.000 interview) som det "gamle datasæt", og data efter 2005 som det "nye datasæt" (p.t. ca. 46.000 interview).

Siden 1992 er der sket mange forbedringer og andre ændringer i spørgeskemaet, men der er en hovedkerne af spørgsmål, som er med i alle årene.

2 Nyheder i Transportvaneundersøgelsen

I løbet af det seneste år har vi gennemført en større modernisering af transportvaneundersøgelsen.

Blandt nyhederne kan nævnes:

- Nyt spørgeskema (3. februar 2009) med bl.a. kortlægning af rejseruter for kollektiv trafik.
- Dobbelt stikprøve fra 1. juni 2009 (til ca. 26 000 interview pr år)
- Helt ny efterbehandling, opbygget i et rent databasemiljø. (foråret 2009)
- Nyt introduktionsbrev til respondenter (januar 2009)

Den dobbelte stikprøve betyder, at vi forventer at nå 100.000 interview i den nye serie i løbet af sommeren 2011.

3 Ny efterbehandling – nyt variabelkatalog

Den største delaktivitet i moderniseringen har været en komplet ny efterbehandling. Der arbejdes nu konsekvent med data i relationsdatabaser og alle variable er dokumenteret med oprindelse og værdisæt.

I den forbindelse er variabelkataloget opdelt i 3 klasser, efter relevans og fortrolighed:

- **Fortrolige** variable, som ikke må anvendes udenfor DTU Transport, og der kun med særlig tilladelse. Omfatter bl.a. koordinater og præcise adresseoplysninger.
- **Udeladte** variable, som ikke medtages i data, fordi de vurderes at være uinteressante i analysesammenhæng. Eksempler: Varighed af interview. Klokketallet for interview. Valg af De/du i spørgeskema.
- **Ordinære** variable, som omfatter flertallet af variablene.

I det følgende vil jeg give en kort oversigt over de vigtigste variable i det nye variabelkatalog, udvalgt blandt de ordinære variable.

Blandt nyhederne er at der nu som hovedprincip anvendes engelske navne. Årsagen er at data i stigende grad bliver brugt i internationale sammenhænge.

3.1 Data på sessionsniveau

Interviewsession (eller kort: session) er den bærende tabel for data på interviewniveau. Principielt opstår tabellen med 1 post pr påbegyndt interview, men som en del af efterbehandlingen frasorteres alle dubletter, således at der i datasættet højst er 1 (gennemført) interview pr respondent. Sessionstabellen svarer dermed i grundtræk til IP i de gamle data.

Sessionstabellen indeholder grundlæggende 2 slags data:

- Baggrundsoplysninger: køn, alder, bilejerskab, etc.
- Aggregater fra turdagbogen: antal ture, samlet km-tal, etc.

Udvalgte variable (session):

Feltnavn	Beskrivelse	Kommentar
SessionId	Primærnøglen	
DiaryYear, DiaryMonth, DiaryWeekday, DiaryDaytype	År, måned, ugedag, dagtype for turdagbogen	
TotalNumTrips, TotalLen	Aggregater af turdagbogen	
RespYearBorn, RespSex	Fødeår, køn	
RespMainOccup, RespEduLevel	Hovedbeskæftigelse, uddannelse	

<i>Feltnavn</i>	<i>Beskrivelse</i>	<i>Kommentar</i>
HomeAdrMuncode Bopælsk	ommune (Zoneinddelinger kan dannes på baggrund af koordinater).	Koordinater og præcis adresse haves som fortrolige variable
WorkplMuncode Arb	ejdsplads/uddannelseskommune	Koordinater og præcis adresse haves som fortrolige variable
WorkPubPriv O	ffentlig/privat ansat	Virksomhedens navn haves ofte i de fortrolige variable
WorkHoursPw, WorkHourType, WorkatHomeDayspM	Arbejdstid, arbejdstidsform, hjemmearbejde	
HWdaysPw, HWdaysReason	Pendlerdage pr uge, evt. årsag til færre dage.	
WorkParkPoss Park	eringsforhold ved arbpl/uddannelsessted	
RespHasBicycle, RespHasSeasonTicket, ResplsMemCarshare	Cykel, bus/togkort, delebil	
RespHasDrivlic, RespDrivlicYear	Kørekort og år for kørekort	
HousehNumCars	Antal biler, som husstanden råder over	
HousehAccommodation, HousehAccOwnOrRent	Boligtype, ejerforhold	
IncRespondent, IncSpouse, IncFamily, IncHouseh	Indkomst: Egen, ægtefælle, familie, husstand	
NuclFamType, PosInFamily	Familietype og respondentens position i familien	
NuclFamNumPers, NuclFamNumAdults, NuclFamNumDrivLic	Antal personer i kernefamilie, antal voxne, antal med kørekort	Far, mor og børn
FamNumPers, FamNumAdults, FamNumDrivLic	Antal personer i familie, antal voxne, antal med kørekort	Alle familierelaterede i husstanden
HousehNumPers, HousehNumAdults, HousehNumDrivlic	Antal personer i husstand, antal voxne, antal med kørekort	
SessionWeight Væg	tfaktor til opregning	

3.2 Data på turniveau

En tur er principielt rejsen fra et ophold/formål til det næste. Tabellen indeholder derfor logisk data som er knyttet til de enkelte ture og er dermed helt parallel til den tidligere TUR tabel.

Udvalgte variable (tur):

<i>Feltnavn</i>	<i>Beskrivelse</i>	<i>Kommentar</i>
OrigMuncode, OrigOldMuncode	Startpunkt for turen, kommunekode ny/gl	Koordinater og præcis adresse haves som fortrolige variable
DestMuncode, DestOldMuncode	Slutpunkt for turen	
GISdist Lu	ftlinieafstand orig-dest	
OrigPurp	Formål med ophold før turen	
DestPurp	Formål med ophold efter turen	
TripPurp, TripPurpGroup	Formål aggregeret	
DepartHH:MM Afrejs	etidspunkt	
ArrivalHH:MM Ankomsttidspun	kt	
DestDwetime	Opholdstid efter turen (rejsemålet)	
NumModes, SumLen, SumMin	Optælling af transportmidler undervejs som antal, længde og tid	
SumMotorLen, SumMotorMin	Optælling, hvor kun motoriserede transportmidler medregnes	
PrimMode, SecMode	Primært/sekundært transportmiddel	Transportmiddel med hhv. længste og næstlængste afstand
PrimModeLen, SecModeLen	Længde i prim/sek trm	
FirstMode, LastMode	Første/sidste transportmiddel, bortset fra gang	
PartyNumU10, PartyNum1017, PartyNumAdults	Medrejsende, antal i aldersgrupper	
PTPrimMode, PTAccMode, PTEgrMode	Kollektivt hovedtransportmiddel, tilbringer, frabringer	
PTAccLen, PTEgrLen	Til/frabringer længde til kollektiv trafik	
PTticketType, PTprice	Billettype, billetpris	
TrainMode, TrainAccMode, TrainEgrMode	Tog, tilbringer, frabringer	

<i>Feltnavn</i>	<i>Beskrivelse</i>	<i>Kommentar</i>
TrainAccLen, TrainEgrLen	Til/frabringer længde til tog	
FirstStation, LastStation	Frastation, tilstation	
TrainAccDist, TrainEgrDist	til/frabringer luftlinieafstand til fra/tilstation	

3.3 Data for turkæder

Turkæder dannes i efterbehandlingen som hele rejsen fra hjem tilbage til samme hjem. Hjem defineres i denne sammenhæng som samme geografiske sted som enten boligadressen eller udgangspunktet. På denne måde fås samme funktionalitet som den hidtidige KAED-tabel, men definitioner og fremgangsmåde er noget anderledes.

Kædebegrebet er nyttigt, fordi på denne måde kan fx. en bolig-arbejdsstedsrejse analyseres som helhed, uanset om der foretages indkøb eller andre ærinder på vejen. Samtidig har turkædebegrebet den fordel, at det er robust overfor mindre fejl i data.

Udvalgte variable (turkæde):

<i>Feltnavn</i>	<i>Beskrivelse</i>	<i>Kommentar</i>
JourneyType Åb	en/lukket turkæde	Lukket turkæde = hjem-hjem, Halvåben har 1 endepunkt hjemme. Helåben har 2 endepunkter ude.
PrimTargetPurp,	Formål, opholdstid på	Primærophold
PrimTargetDwetime,	primærophold, kommunkode	defineres som ophold
PrimTargetMuncode		med længste opholdstid
SecTargetPurp	Formål på sekundærophold	næstlængste ophold
SumLen, SumMin	Summer af ture på turkæden	
PrimMode, PrimModeLen	Primært transportmiddel, længde heri	Transportmidlet med længste afstand på turkæden
OutBPrimMode, OutBLen,	Primært transportmiddel på	Udturen defineres som
HomeBPrimMode,	udtur/hjemtur, længde af	turen frem til
HomeBLen	udtur/hjemtur	primæropholdet, hjemturen er derefter.

4 Kvalitetsstrategi

4.1 De enkelte interview

Første prioritet er at få respondent / interviewer til at inddatere data i en god kvalitet. Den primære årsag er at på det tidspunkt er det muligt at få opklaret evt. tvivlsspørgsmål. En behagelig sideeffekt er at det sparer på efterbehandlingen.

Vi anvender 2 metoder:

- Spørgeskemaet er programmeret med fejlkontroller for en meget lang liste af mulige fejl. Vi har 2 niveauer: Det "umulige", hvor respondenter tvinges til at ændre sine svar, så de er fysisk mulige. Og det "usandsynlige", hvor respondenter bliver spurgt en extra gang om svaret er korrekt, fordi det ligger udenfor hvad der normalt kan accepteres.
- Vi overvåger en lang række nøgletal, opdelt for de enkelte interviewere, sammenholdt med de tilsvarende værdier for internetinterview.

Samlet betyder dette, at de rå interviewdata har en høj kvalitet og for en stor dels vedkommende kan gå direkte igennem kvalitetskontrollen og ind i det færdige datasæt.

Selve kvalitetskontrollen består i et stort antal fejlfiltre, hvor de enkelte variable kontrolleres enkeltvist og i sammenhæng. De enkelte anmærkninger lægges i en særlig tabel, som herefter analyseres for forskellige mønstre, og hvor efterbehandlingsindsatsen prioriteres efter hvor der forventes størst virkning i form af enten kasserede interview eller fejlrettelser. Ca. 15 % af de gennemførte interview har mindst 1 anmærkning, men store dele af disse er harmløse og giver ikke anledning til videre aktion. Ca. 5 % af de gennemførte interview underkastes varierende grader af behandling, fra tilretning af enkelte værdier over total omkalfatring af data til egentlig kassation.

4.2 Adresseefterkodning

Ca. 95 % af alle steder (turendepunkter, boligadresser, arbejdspladser etc.) geokodes direkte af respondenter ved søg & vælg i spørgeskemaet. I de resterende tilfælde, udfylder respondenter en beskrivelse af stedet som fri tekst. Disse steder efterkodes parallelt til den øvrige efterbehandling. Adresserne behandles til et niveau, hvor 98 % kan stedfæstes på koordinatniveau, og 99,9 % på zoneniveau.

4.3 Repræsentativitet af undersøgelsen

Interviewpersonerne udvælges tilfældigt og repræsentativt fra CPR registret. Svarprocenten (netto) ligger på i gennemsnit ca. 62 %, hvilket anses for ganske tilfredsstillende for en undersøgelse af denne type.

Det færdige resultat opregnes efter køn, aldersgruppe og geografi, således at data tilsammen repræsenterer alle 10-84 årige med bopæl i Danmark.

5 Fremtiden .. ?

Det er svært at spå om fremtiden, men her er en oversigt over hvad der lige nu er på tegnebrædtet:

- Tillægsspørgsmål om hvilken af husstandens biler, der anvendes på bilture. (planlagt sept 2009)
- Tillægsspørgsmål om varighed (antal dage) for rejser med overnatning. (planlagt sept. 2009)
- Storesøster i form af en fjernrejseundersøgelse med fokus på ture af flere dages varighed og på sammenhænge mellem dage. (planlagt start dec. 2009)
- Ny klasse af afledte variable, baseret på genberegning af rejseruter i rejseplanen og som bil i vejnettet. Giver mulighed for at analysere på kollektiv trafik som alternativ, hvor den er fravalgt. – Og modsat: bil som alternativ til kollektiv. (efterår 2009)
- Automatisk udtræksværktøj til hjemmesiden, således at simple tabeller kan udtrækkes som dan-selv. (efterår 2009)