

Projekt nr. 1:

Tælling af S-togs passagerer.

Fordelingen af billetindtægterne (ca. 2.5 mia årligt) fra Hovedstadsområdets offentlige transportsystem mellem de fem operatører (DSB regional, DSB S-tog, Movia, Metroen og Lokalbanelne) er baseret på passagertællinger, der foretages af operatørerne selv under overvågelse af det såkaldte *Model- og tælleudvalg*. Dette er et udvalg nedsat af trafikstyrelsen, hvori de fem operatører er repræsenteret. Hvordan disse tællinger foretages, og hvordan de indgår i fordelingsalgoritmen er en meget lang og kompliceret historie, som ikke vil blive fortalt her. Vi skal kun beskæftige os med vurdering af DSB S-togs nye vejningsbaserede tællinger, som beskrives i det følgende.

Indtil for nylig (til og med 2008, sikkert også til dels 2009) har S-togenes passagertællinger været foretaget af perrontællere. På udvalgte stationer (typisk ved en stor station på hver gren samt ved Østerport) tælles ca. en dag om måneden hvor mange passagerer der befinder sig i hvert eneste tog som afgår fra perronen. Hver vogns passagertal vurderes af en perrontæller. Det drejer sig om folk med en vis rutine, som er gode til at skønne hvor mange personer der sidder eller står inde i toget.

Denne metode bliver løbende kontrolleret ved sammenligning med passagertællinger, der foretages af en person inde i toget. Den har vist sig at være ret nøjagtig i det lange løb. Så nøjagtig, at vi ikke vil stille spørgsmålstegn ved den i forbindelse med denne opgave, som først og fremmest handler om muligheden for passagertælling ved vejning af S-togene. Her vil perrontællingerne overtage rollen som kontrol-tællinger.

Af tekniske grunde, som har at gøre med kalibreringen af bremserne og togets affjedring, vejes alle S-togets moduler i forvejen ved hver eneste afgang. Det er disse vejninger, der registreres via togets computer. De nulstilles, således at et tomt modul i princippet vejer 0 kg. Vejningerne sker på en heltalsskala, hvor enheden er 256.4 kg. I datasættet er vægtene dog omregnet til kg.

I foråret 2008 er vejningsmetoden løbende anvendt, samtidig med at de gamle perrontællinger er fortsat. Disse perrontællinger er fortsat kontrolleret ved tællinger inde i toget, og disse kontrol-tællinger udviste (stadig) god overenstemmelse med perrontællingerne.

Data findes på datafilen 10-1DAT.TXT, og også som ISUW datasættet 10-1DAT.SUD. Tekstfilen består af en overskriftsline med variabel-

navne og 15020 datalinier. Hver linie svarer til et S-togs afgang fra en station. Betydningen af de enkelte variable er følgende:

L1 Litra for togmodul 1
L2 Litra for togmodul 2, -1 hvis modul 2 ikke findes
L3 Litra for togmodul 3, -1 hvis modul 3 ikke findes
V1 Vægt af første halvdel af modul 1 (i kg)
V2 Vægt af anden halvdel af modul 1 (i kg)
V3 Vægt af første halvdel af modul 2 (i kg)
V4 Vægt af anden halvdel af modul 2 (i kg)
V5 Vægt af første halvdel af modul 3 (i kg)
V6 Vægt af anden halvdel af modul 3 (i kg)
APTS Antal PTS vogne (2, 4, 6 eller 8)
TPTS Antal PTS vogne faktisk talt (2, 4, 6 eller 8)
PTS Antal passagerer ifølge PTS

... hvor PTS står for “perron tælle system”.

Bemærk, at de to variable APTS og TPTS åbenbart opererer med vogne der udgør en ottendedel af et tog af maksimal længde, og at et tog altid består af 2, 4, 6 eller 8 vogne. I andre sammenhænge opererer DSB med enheder som er 1/16 af et langt tog (det gælder konventionen omkring “litra”, se nedenfor). For en sikkerheds skyld vil vi fortsætte med at definere en “vogn” som en ottendedel af et langt tog, selv om denne definition i virkeligheden refererer til tidligere generationer af S-tog, som havde egentlige vogne (ligesom et regional- eller kystbanetog idag).

Desværre bliver det mere kompliceret nu: Et S-tog er sammensat af *moduler*, som hver for sig kan bestå af 2 eller 4 vogne. Normalt består et tog af et eller to, i sjældne tilfælde tre moduler (selv om man i princippet kunne sammensætte et langt tog af fire tovognsmoduler). De tre første variable, L1, L2 og L3 fortæller hvordan toget er sammensat. L1 er “litra” for første modul. L2 er litra for andet modul, -1 hvis der kun er et modul. Og L3 er litra for tredje modul, -1 hvis der kun er et eller to. Litra (som normalt er en bogstavkombination, efterfulgt af et tal) er i denne sammenhæng et firecifret nummer, som identificerer det pågældende modul entydigt. Dette nummer fortæller også, om et modul er langt eller kort. Et tovognsmodul hedder 4*** (hvor stjernerne står for cifre), medens et firevogns modul har litra af formen 8***.

Ved vejning af et firevogns modul vejes de to halvdele hver for sig. Dette er baggrunden for den lidt komplicerede datastruktur, som kan beskrives således. De seks vægtangivelser V1, ..., V6 knytter sig til litra på følgende måde:

V1 er vægten af første halvdel af litra1
V2 er vægten af anden halvdel af litra1
V3 er vægten af første halvdel af litra2
osv. osv. ,

hvor -1 er koden for en vejning som ikke findes, enten fordi det pågældende modul ikke findes eller fordi vejningen refererer til anden halvdel af et modul som kun er to vogne langt. I princippet burde der således gælde følgende:

V1 er aldrig -1 .

V2 er -1 hvis og kun hvis `litra1` er et tovognsmodul.

V3 er -1 hvis og kun hvis toget kun består af ét modul.

V4 er -1 hvis og kun hvis toget enten kun består af ét modul, eller modul 2 er et tovognsmodul.

Osv. osv.

Desværre er datafilen fuld af inkonsistenser og fejl. En stor del af opgaven består i at tage stilling til, hvordan disse skal håndteres.

Eksempler. Første datalinie i `10-1DAT.TXT`, sammen med linien af variabelnavne, ser sådan ud:

L1	L2	L3	V1	V2	V3	V4	V5	V6	APTS	TPTS	PTS
4102	8201	-1	1538	0	-1	-1	-1	-1	6	4	24

Her burde V1, V3 og V4 være angivet, de øvrige burde være -1 . Samtidig viser APTS og TPTS at der (korrekt) skulle være 6 vogne, men kun 4 er talt. Så der er noget helt galt, og datalinien bør rimeligvis kasseres.

Linie 80 (datalinie 79) ser sådan ud:

L1	L2	L3	V1	V2	V3	V4	V5	V6	APTS	TPTS	PTS
8105	8122	-1	0	0	769	1538	-1	-1	8	8	35

Her er der formelt ikke noget galt. Det virker bare underligt, at halvdel af toget har været tomt, medens der i den anden halvdel — ifølge perrontællingen — har siddet 35 passagerer (som i så fald skulle have en gennemsnitsvægt på 66 kg). Man kunne overveje at kassere datalinien alene af denne grund. Men den tvivl man måtte have her forsvinder, når man får at vide (se nedenfor) at modulet 8105 er blandt de “black-listede”.

Linie 431 (datalinie 430) ser sådan ud:

L1	L2	L3	V1	V2	V3	V4	V5	V6	APTS	TPTS	PTS
4112	8133	-1	7948	0	5128	6923	-1	-1	6	6	174

Her er der ikke andet galt end at V2 burde være -1 , ikke 0. Men det er en let gennemskuelig fejl ved kodningen (den går sikkert igen), og det vil nok være det mest rimelige at beholde denne datalinie. Det er de rigtige vægte som er positive, og der er nogenlunde overensstemmelse med perrontællingen (de angivne 174 passagerer skulle så veje ca. 115 kg i snit).

Som det ses er der behov for en omfattende datarensning før den egentlige statistiske analyse. Resultaterne af en analyse vil sikkert i høj grad afhænge af hvilke observationer der kasseres, så det kan meget vel

være nødvendigt at gentage de samme analyser flere gange for mere eller mindre kritisk fejlrettede datasæt.

Af yderligere information, der kan være relevant for datarensningen, har vi følgende. Der er kendte, eksperimentelt konstaterede vejningsfejl ved følgende litra:

8101 , . . . , 8110.

8115.

8118.

8166.

8176.

8181.

8188.

8192.

Opgavens formål er at analysere det foreliggende datamateriale med særligt henblik på besvarelse af følgende spørgsmål:

1. Hvordan kan de løbende vejninger bedst muligt bruges til udregning af estimerede passagertal, der kan erstatte perrontællingerne som DSB S-togs input til fordelingsnøglen?

2. Diskuter de fejlkilder, som indgår i denne metode. Det kan oplyses, at antallet af S-togsrejser pr. år er af størrelsesorden 90 millioner. Det er ikke helt det antal, som vejetællingerne giver os (det er snarere antal rejser mellem enkeltstationer), men omregningfaktorerne mellem disse antal kan opfattes som kendte, således at den relative fejl på disse antal er den samme. Endvidere kan oplyses, at perrontællingerne ifølge tidligere foretagne kontroltællinger (inde i toget) rammer ca. 1.2% for lavt, og på årsbasis er estimeret med en usikkerhed som svarer til at man i formlen $\text{perrontælling} = 0.988 \times \text{kontroltælling}$

skal forsyne tallet 0.988 med en standardafvigelse på 0.002.

3. I hvor høj grad kan vejninger fra de litra, som vides at veje forkert, alligevel bruges til passagertælling, hvis man indfører en passende (litra-specifik) korrektion?

Opgave 1 afleveres tirsdag d. 23. marts kl. 12 i kombisekretariatet.