

Dokumentasjon døgnfordelinger

Dette notatet dokumenterer døgnfordelingsmodellene som leveres i forbindelse med prosjekt 4356 *Fordeling Trenklin*. For mer detaljerte beskrivelser refereres det til TØI-rapport 1558/2017.

Tilpassing og uttesting av døgnfordelingene er gjort med programvaren R. De endelige modellene er implementert i Microsoft Excel.

Det leveres:

1. Excel-fil med inndata til døgnfordelingene.
2. Excel-filer med beregning av døgnfordelingene (en fil for hver reisehensikt).
3. CSV-filer med døgnfordelinger for alle relasjoner.

1 Inndata

Inndata for døgnfordelingene er reisetid mellom togstasjoner (fra NSBs tilbudsmatriser) og en relativ pendlerindeks som er beregnet fra SSBs sysselsettingsstatistikk for 2014.

Den relative pendlerindeksen er beregnet ved en transformasjon av pendlerdata på grunnkrets nivå til togstasjonsrelasjoner, og angir forholdet mellom pendlerstrømmen fra stasjon O til D og fra stasjon D til O.

Filen **inndata_dognfordelingene.xlsx** består av data i 11 kolonner. De fire første kolonnene er ID og navn for start/ende-stasjon i relasjonen. Deretter kommer ombordtid, antall bytter, ventetid ved togbytte og den totale reisetiden (ombordtid + ventetid).

Pendlerindeks.OD og Pendlerindeks.DO er mål på omfanget av pendling i de respektive retningene, og den relative pendlerindeksen er gitt ved $\text{Pendlerindeks.OD} / \text{Pendlerindeks.DO}$.

Det er variablene **Reisetid** og **Relativ.Pendlerindeks** som er inndata i modellene.

Under vises et uttrekk av de første radene i datasettet.

Fra.ID	Til.ID	Fra.Navn	Til.Navn	Ombordtid	Bytter	Bytventetid	Reisetid	Relativ.Pendlerindeks	Pendlerindeks.OD	Pendlerindeks.DO
2	1	Ski	Oslo S	21.1899	0	0	21.1899	7.7276	558.47	72.27
3	1	Moss	Oslo S	43.7395	0	0	43.7395	13.4002	513.63	38.33
4	1	Rygge	Oslo S	52.7481	0	0	52.7481	29.68	59.36	2
5	1	Råde	Oslo S	58.8069	0	0	58.8069	109.825	43.93	0.4
6	1	Fredrikstad	Oslo S	70.8628	0	0	70.8628	10.7192	288.24	26.89
7	1	Sarpsborg	Oslo S	84.9984	0	0	84.9984	17.6857	148.56	8.4

For stasjonsrelasjoner der vi enten mangler pendlerstatistikk, eller det er null pendlere, har vi satt pendlerindeksen til 0.0001. Mangler det data i begge retninger vil den relative pendlerindeksen bli 1, altså antar vi like mange pendlere i begge retninger ved manglende informasjon.

Den største verdien av den relative pendlerindeksen for relasjoner med pendlerdata er 159. Ved å dele på 0.0001 kan vi få veldig store verdier for enkelte relasjoner, men dette blir håndtert i modellberegningen fordi hvis den relative pendlerindeksen er veldig stor vil all

vekt legges på morgen-peaken, på samme måte vil all vekt legges på ettermiddags-peaken hvis den relative pendlerindeksen er veldig liten.

Reisetid mangler for enkelte relasjoner, manglene verdi indikeres med NA.

2 Modeller og modellfiler

2.1 Matematisk definisjon og parameterverdier

For stasjonsrelasjon (i, j) og reisehensikt H (arbeid, tjeneste eller andre reise-hensikter), har vi at døgnfordelingen over ankomsttid ved endestasjonen er en mixture av normalfordelinger (diskretisert og normalisert):

$$f_{i,j}^H(t) = \frac{\lambda_1^H N(t; \mu_1^H, \sigma_1^H) + \lambda_2^H N(t; \mu_1^H, \sigma_1^H)}{C_{i,j}^H}, t \in \{0, 1, 2, \dots, 1439\}, \quad (1)$$

der $\lambda_2^H = 1 - \lambda_1^H$ og normaliseringskonstanten er gitt ved

$$C_{i,j}^H = \sum_{t=0}^{1439} \lambda_1^H N(t; \mu_1^H, \sigma_1^H) + \lambda_2^H N(t; \mu_1^H, \sigma_1^H).$$

Her betegner $N(t; \mu, \sigma)$ en normalfordelt sannsynlighetstetthet med forventning μ og standardavvik σ , evaluert i punktet t .

For enklere notasjon i uttrykket over har vi ikke tatt med at mixture-vektene og lokasjonsparameterne er relasjonsavhengige. Vi oppsummerer fordelingene for de ulike hensiktene i avsnittene som følger.

I valget av parameterverdier er det sett på parameterestimatene med og uten reisetid som forklaringsvariabel basert på RVU, og estimatene basert på telldata, og gjort et kompromiss mellom disse, samt en avrunding til hele timer, eksempelvis kl. 16:00 i stedet for 15:57. Denne vurderingen bør etterprøves ved å validere fordelinger for viktige relasjoner med enten passasjertellinger eller etablert kunnskap.

Datagrunnlaget har vært størst for arbeidsreisene. For de to andre reisehensiktene var estimeringsgrunnlaget tynnere, og især tjenestereisefordelingen er det knyttet stor usikkerhet til. Vi anbefaler å innhente ytterligere data for å kunne etablere bedre døgnfordelinger for tjenestereiser med tog.

2.1.1 Arbeidsreiser

For arbeidsreiser bestemmes mixture-vektene av den relative pendlerindeksen, og lokasjonsparameteren for ettermiddagskomponenten er avhengig av reisetiden. Vi har valgt å la morgenkomponenten være konstant. Vi har at

$$\lambda_1^{arbeid}(i, j) = \min \left\{ 1, \max \{ 0, \gamma_0 + \gamma_1 \ln \rho_{ij} \} \right\}.$$

og at $\mu_2^{arbeid} = \alpha^{arbeid} + \beta^{arbeid} \times T_{ij}$, der T_{ij} er reisetiden fra stasjon i til stasjon j . Parameterne som benyttes i den implementerte modellen vises i Tabell 1.

Tabell 1: Parameterverdier for døgnfordelingsmodellen for arbeidsreiser.

Parameter	Verdi
γ_0	0,43
γ_1	0,1
μ_1^{arbeid}	480 (kl. 08:00)
α^{arbeid}	960 (kl. 16:00)
β^{arbeid}	0,5
σ_1^{arbeid}	60
σ_2^{arbeid}	120

Vi har lagt inn en begrensning slik at lokasjonsparameteren for ettermiddagskomponenten ikke kan bli større enn 1080 (kl. 18:00).

2.1.2 Tjenestereiser

På grunn av svært lite datagrunnlag for tjenestereiser med tog foreslår vi å benytte døgnfordelingene for arbeidsreiser med en liten justering slik at andelen reiser midt på dagen blir høyere.

Dette oppnås ved å innføre en tredje komponent med lokasjonsparameter kl. 12. Vi har satt $\mu_0^{tjeneste} = 720$ og $\sigma_0^{tjeneste} = 300$.

De fire andre lokasjon- og skalaparameterne beholdes som i Tabell 1, men vektene må justeres for å ta høyde for en tredje komponent. Vi har satt $\lambda_0^{tjeneste} = 0,3$.

Tjenestereisefordelingene er svært like arbeidsreisefordelingene, men med litt større etterspørsel midt på dagen. Hvis det er ønskelig (pga. eksisterende kunnskap eller nye data) med høyere etterspørsel midt på dagen kan variansen justeres ned og/eller vekten opp.

2.1.3 Andre reiser

For de andre reisehensiktene (besøk, handel, etc.), er ikke pendlerindeksen viktig da dette er en type reiser som i stor grad foretas etter arbeidstid uavhengig av relasjon (dette skyldes delvis hvordan reisehensikt defineres). Mixture-vektene er estimert til henholdsvis 0,35 og 0,65.

På samme måte som for de andre hensiktene har vi valgt å la lokasjonsparameteren for morgenkomponenten være konstant, mens lokasjonsparameteren for ettermiddagskomponenten er gitt ved:

$$\mu_2^{andre} = \alpha^{andre} + \beta^{andre} \times T_{ij},$$

der T_{ij} er reisetiden fra stasjon i til stasjon j .

En oppsummering av parameterverdiene finnes i Tabell 2.

Tabell 2: Parameterverdier for døgnfordelingsmodellen for øvrige reiser.

Parameter	Verdi
λ_1	0,35
λ_2	0,65
μ_1^{andre}	720 (kl. 12:00)
α^{andre}	1070 (kl. 17:50)
β^{andre}	0,1
σ_1^{andre}	270
σ_2^{andre}	170

Vi har lagt inn en begrensning slik at lokasjonsparameteren for ettermiddagskomponenten ikke kan bli større enn 1140 (kl. 19:00) for lange reisetider.

Døgnfordelingen for andre reisehensikter framstår som homogen over stasjonsrelasjoner.

2.2 Implementasjonen

Modellene er implementert i Excel og finnes i filene:

- Døgnfordeling_Arbeidsreiser.xlsx
- Døgnfordeling_Tjenestereiser.xlsx
- Døgnfordeling_Andre_reiser.xlsx

Alle filene består av fem ark:

1. Parameterverdier
2. Restriksjoner
3. Inndata
4. Beregninger
5. Døgnfordeling

Parameterarket inneholder verdiene fra Tabell 1 og Tabell 2. Restriksjonsarket legger begrensning på forventningsverdien om ettermiddagen. Inndata inneholder reisetid og den relative pendlerindeksen for valgt stasjonsrelasjon. I beregningsarket beregnes vektene og forventningsverdiene. Her sikres det at alle vekter ligger i intervallet $[0,1]$ og summerer seg til 1. I arket med døgnfordelingen ligger en ikke-normalisert og en normalisert døgnfordeling, den normaliserte er markert med grønn farge.

Ved å endre inndata genereres det en ny døgnfordeling for den aktuelle stasjonsrelasjonen.

Videre uttesting og kalibrering mot nye data anbefales, især for tjenestereiser.

3 Fordelinger for alle relasjoner

Det leveres også tre CSV-filer som inneholder døgnfordelinger for alle stasjonsrelasjoner.

Filene er bygget opp som følger:

- Kolonne 1 til 4: ID og navn for start- og endestasjon i stasjonsrelasjonen,
- Kolonne 5 til 1444: døgnfordelingen for samme relasjon for hvert minutt i døgnet.