Overlevelsestavler

Anna-Vera Jørring Pallesen, Johan Sebastian Ohlendorff, Laust Hvas Mortensen and Thomas Alexander Gerds

1 Introduktion

Overlevelsestavlen repræsenterer en matematisk model, der kvantificeres ved hjælp af konkrete demografiske data. Modellen genererer derefter en omfattende beskrivelse af dødelighedsforholdene i den specifikke befolkning. De forskellige mål for dødelighed konstrueres på baggrund af overlevelsestavlen. Dette gælder også for målene vedrørende forekomsten af vielser, skilsmisser, vandringer og i vis grad forskellige fertilitets- og reproduktionsmål. Overlevelsestavlen kunne derfor gennemgås på en ret abstrakt måde og fortolkes forskelligt, afhængigt af om den skal anvendes til at beskrive dødelighed, vielser, skilsmisser eller fertilitet. I det følgende vil vi dog fokusere på at opbygge modellen omkring målingen af befolkningens dødelighed for at gøre det lettere at forstå modellens umiddelbare anvendelighed.

1.1 Middellevetid

Hvor mange år kan en nyfødt i dag forvente at leve? Dette spørgsmål er umuligt at besvare korrekt fordi svaret umiddelbart afhænger hvad der sker i fremtiden. Alligevel er middellevetid, altså den forventede gennemsnitlige levetid af en nyfødt, et demografisk værktøj som anvendes hyppigt til belysning af befolkningens nuværende dødelighedsniveau. Middellevetid bruges også som sammenligningsgrundlag på tværs af befolkninger og tid. Tallet angiver det gennemsnitlige antal år, som en nyfødt kan forventes at leve *under den forudsætning*, at de nuværende mortalitetsrater for alle grupperinger af køn og alderstrin holder sig på det samme niveau i al fremtid. Med middellevetiden har man et relativt simpelt begreb, som gør det muligt at sammenligne forskellige befolkningers dødelighed. I praksis vil de nuværende dødshyppigheder formentlig ikke holde sig på et konstant niveau i al fremtid.

I gennem mange år har der været en tendens til faldende mortalitetsrater, og der er meget, som tyder på, at det er en udvikling som fortsætter. Den konkrete fortolkning af middellevetiden for 0-årige som det gennemsnitlige antal år, som en nyfødt kan forventes at leve, vil derfor formentlig undervurdere den faktiske middellevetid. Men formålet med middellevetiden er heller ikke at forudsige præcist, hvor længe nyfødte vil leve. Formålet er at have et simpelt begreb, der kan sammenlignes på tværs af befolkninger og tid.

1.2 Andre dødelighedsmål

Middellevetiden er et mål som resulterer fra den såkaldte overlevelsestavle. En klassisk overlevelsestavle beskriver også en række andre dødelighedsmål, såsom den forventede restlevetid fra alder x, sandsynligheden for at dø inden alder x og sandsynligheden for at være i live ved alder x.

2 Konstruktion af overlevelsestavle

Overlevelsestavler beskriver hvordan en tænkt lukket fødselskohorte reduceres med stigende alder alene på grund af dødsfald. Fordi kohorten er lukket, er død den eneste mulige afgang fra kohorten. Der tages udgangspunkt i en fiktiv tabelbefolkning bestående af ℓ_0 personer, som antages at være født på nøjagtig samme tidspunkt. Antallet af fiktive tabelpersoner ℓ_0 kaldes for 'radix', og radix sættes typisk til $\ell_0 = 100.000$.

2.1 Eksempel

```
x <- hent_mortalitetsrate_data(breaks = c(0,1,5,10,seq(20,90,10),Inf)
    ,
                               tid = 2019,
                               right = FALSE,
                               alder = "all_no_total")
x <- mutate(x,M = Dod/R)
x_kvinder <- filter(x,KŒN == "Kvinder")
x_kvinder <- mutate(x_kvinder,a = c(0.1,2,rep(5,10)),n = c(1,4,rep
    (10,10)))
tavle_kvinder <- overlevelsestavle(x_kvinder,
                                   mortalitet = "M",
                                   alder = "aldersinterval")
print(tavle_kvinder,digits = 2)
```

```
   Alder      l     d    p       q      L       T    e
 1:     0 100000   251 1.00 0.00251  99774 8829735 88.3
 2:   1-4  99749    40 1.00 0.00040 398917 8729961 87.5
 3:   5-9  99709    79 1.00 0.00080 996695 8331044 83.6
 4: 10-19  99630    96 1.00 0.00096 995819 7334349 73.6
 5: 20-29  99534   189 1.00 0.00190 994393 6338530 63.7
 6: 30-39  99345   377 1.00 0.00380 991560 5344137 53.8
 7: 40-49  98967   889 0.99 0.00899 985228 4352577 44.0
 8: 50-59  98078  2820 0.97 0.02876 966680 3367349 34.3
 9: 60-69  95258  7748 0.92 0.08133 913840 2400669 25.2
10: 70-79  87510 16271 0.81 0.18593 793745 1486829 17.0
11: 80-89  71239 36282 0.49 0.50930 530981  693084  9.7
12:   90+  34957 34957 0.00 1.00000 162103  162103  4.6
```

2.2 Dekrementfunktion

Funktionen ℓ angiver hvor mange tabelpersoner er stadigvæk i live ved alder x og beskriver hvordan tabelbefolkningen reduceres på grund af dødsfald. Startværdien ℓ_0 angiver hvor mange tabelpersoner der er i tabelbefolkningen helt i begyndelsen hvor alder er lige med 0, og ℓ_30) angiver hvor mange tabelpersoner er i livet ved alder 30. Fordi ℓ er monoton faldende som funktion af alder, altså $\ell_x \geq \ell_{x+1}$, kalder man den for dekrementfunktion. For eksempel, $\ell_30 = 91.047$ betyder at ud af $\ell_0 = 100.000$ tabelpersoner 91.047 personer er stadigvæk i live ved alder 30. I dette eksempel er overlevelsessandsynligheden i tabelbefolkningen ved alder 30 lige med

$$o(30) = \frac{\ell_30}{\ell_0} = \frac{91.047}{100.000} = 91,1\%.$$

Overlevelsesfunktionen er defineret som

$$o(x) = \frac{\ell_x}{\ell_0}.$$

Under konstruktionen af overlevelsestavler er opgaven at beregne dekrementfunktionens værdier ℓ_x for alle alderstrin $x = 0, 1, \ldots, x^{max}$ hvor x^{max} er det sidste alderstrin. Per konstruktion, der bliver forklaret i detaljer nedenfor, dør alle resterende tabelpersoner i det sidste alderstrin, det vil sige $\ell_{x^{max}+1} = 0$, og dermed har vi også $o(x^{max} + 1) = 0$.

2.3 Dødshyppigheder

Dødshyppigheder ${}_kq_x$ beskriver sandsynligheden for at dø i aldersintervallet $(x, x + k]$ given overlevelse indtil alder x. Dødshyppigheder forbinder den ægte, åbne befolkning, som man interesserer sig for, med den tænkte, lukkede tabelbefolkning der definerer overlevelsestavlen. Den underliggende idé er at mortalitetsraterne er det samme i disse to befolkninger for begge køn og alle alderstrin.

2.3.1 Approksimationsformel

For at beregne dødshyppighederne ${}_kq_x$ bruger vi data fra den ægte befolkning, typisk fra en kort kalenderperiode. Problemet er at den ægte befolkning er *åben*. Dødsfald bliver ikke bliver registreret for personer som udvandrer i perioden og både udvandrer og indvandrer i perioden bidrager ikke med risikotid til hele perioden. Derfor er det uklar hvor mange personer der er i aldersintervallet. Aldersspecifikke summariske mortalitetsrater løser problemet ved at dividere antal dødsfald med antal risikotid hvor indvandrer og udvandrer bidrager med den tid de nu har været i befolkningen. Nøglen til en approksimation af dødshyppigheder baseret på mortalitetsrater er følgende centrale formel:

$${}_kq_x = \frac{k \cdot {}_kM_x}{1 + (k - {}_ka_x) \cdot {}_kM_x} \tag{K3.1}$$

Formlen afhænger aldersspecifikke mortalitetsrater ${}_kM_x$, længden af aldersintervallet k, og også en konstant ${}_ka_x$, som kaldes Chiang's a. Konstanten ${}_ka_x$ beskriver den gennemsnitlige levetid i aldersintervallet $(x, x+k]$ for personer, der døde mellem alder x og alder $x+k$. Dermed beskriver $(k - {}_ka_x)$ den gennemsnitlige tid som en person der døde i aldersintervallet var død.

2.3.2 Chiang's a

For at beregne dødshyppigheder med den centrale formel (K3.1) har vi brug for at specificere Chiang's a for alle aldersintervaller. Chiang's a skal approksimere det forventede antal år levet i intervallet af en person, som dør i intervallet. Hvis Chiang's a opfylder dette, bliver

$$\text{Dødstid i aldersinterval} = (k - {}_ka_x) \cdot {}_kD_x,$$

en god approksimation af antal dødstid som alle personer der døde i aldersintervallet har været død.

Hvis vi antager at dødtider er jævnt fordelt i aldersintervallet, altså at det er lige sandsynligt at dø i starten som det er i slutningen af aldersintervallet, er det rimeligt at vælge

$${}_ka_x = \frac{k}{2}.$$

Det første og sidste aldersinterval vil dog altid kræve særlige værdier af ${}_ka_x$. I det første leveår er dødstiderne meget skævt fordelt over året, de fleste dødstider inden 1-års fødselsdagen ligger kort efter fødslen. Derfor sætter vi ${}_1a_0 = 0,1$. For det sidste interval x^{max} vil man typisk vælge ${}_\infty a_{x^{max}}$, så dødshyppigheden bliver 1. Det svarer til at vælge ${}_\infty a_{x^{max}} = \frac{1}{{}_\infty M_{x^{max}}}$.

	5-års aldersintervaller	10-års aldersintervaller
første leveår	${}_1a_0 = 0,1$	${}_1a_0 = 0,1$
aldersinterval 1-5 år	${}_4a_1 = 4 \cdot 0,5 = 2$	${}_9a_1 = 9 \cdot 0,5 = 4,5$
alle andre intervaller	${}_ka_5 = 5 \cdot 0,5$=2,5	${}_ka_{10} = 10 \cdot 0,5 = 5$
sidste aldersinterval	$a_{x^{max}} = \frac{1}{{}_\infty M_{x^{max}}}$	$a_{x^{max}} = \frac{1}{{}_\infty M_{x^{max}}}$

2.3.3 Forklaring af den centrale formel

I det følgende skal vi på en uformelt måde forklare formel (K3.1). Hvis den ægte befolkning var lukket, altså uden forekomst af ind- og udvandring, ville man kunne beregne dødshyppighederne simpelt som antal dødsfald i aldersintervallet divideret med antal personer i starten af aldersintervallet:

$$\text{Dødshyppighed} = \frac{\text{Antal dødsfald i aldersinterval}}{\text{Antal personer i starten}}.$$

Fordi aldersintervallet er k år lang gælder

$$\text{Antal personer i starten} = \frac{\text{Risikotid} + \text{Dødstid}}{k}.$$

Her er risikotid antal år som befolkningens personer har livet (i aldersintervallet) og dødstid antal år som befolkningens personer var døde. Med denne formel kan dødshyppigheden skrives som

$$\text{Dødshyppighed} = \frac{k \cdot \text{Antal dødsfald i aldersinterval}}{\text{Risikotid} + \text{Dødstid}}. \tag{K3.2}$$

Vi sætter Chiang's a sådan at

$$\text{Dødstid i aldersinterval} = (k - {}_k a_x) \cdot {}_k D_x,$$

er en god approksimation af antal dødstid som alle personer der døde i aldersintervallet har været død (c.f., afsnit 2.3.2). Hvis vi nu anvender formlen for den aldersspecifikke mortalitetsrate

$${}_k M_x = \frac{{}_k D_x}{{}_k R_x}$$

ser vi at den centrale formel (K3.1) er faktisk lige med formel (K3.2):

$$\begin{aligned}
\frac{k \cdot {}_k M_x}{1 + (k - {}_k a_x) \cdot {}_k M_x} &= \frac{k \cdot \frac{{}_k D_x}{{}_k R_x}}{1 + (k - {}_k a_x) \cdot \frac{{}_k D_x}{{}_k R_x}} \\
&= \frac{k \cdot {}_k D_x}{{}_k R_x \cdot (1 + (k - {}_k a_x) \cdot \frac{{}_k D_x}{{}_k R_x})} \\
&= \frac{k \cdot {}_k D_x}{{}_k R_x + (k - {}_k a_x) \cdot {}_k D_x}.
\end{aligned}$$

2.3.4 Beregning af antal dødsfald og overlever

For at beregne antal tabelpersoner som overlever indtil det første alderstrin, $x = 1$, skal vi beregne hvor mange tabelpersoner dør mellem alder $x = 0$ og alder $x = 1$. For at beregne hvor mange tabelpersoner overlever alder $x + k$ skal vi beregne hvor mange af de resterende ℓ_x tabelpersoner dør i aldersintervallet $(x, x + k]$. Vi betegner med ${}_k d_x$ antal tabelpersoner som dør mellem alder x og alder $x + k$. Dermed er ${}_1 d_x$ antal tabelpersoner som dør mellem alder x og alder $x + 1$. Sandsynligheden for at dø mellem to alderstrin (dødshyppighederne) er som sagt det centrale element ved konstruktionen af overlevelsestavlen. Vi beregner antal dødsfald i aldersintervallet $(x, x + k]$ ved at multiplicere antal tabelpersoner i starten af intervallet med dødshyppigheden:

$${}_k d_x = {}_k q_x \cdot \ell_x. \tag{K3.3}$$

Det er vigtig at skelne mellem antal døde ${}_k D_x$ i den ægte befolkning og antal døde ${}_k d_x$ i tabelbefolkningen. Baseret på formel (K3.3) kan vi beregne hvor mange tabelpersoner er i livet i starten af det næste aldersinterval:

$$\ell_{x+k} = \ell_x - {}_k d_x.$$

Alternativt kan vi starte med at beregne dekrementfunktionen baseret på dødshyppigheden

$$\ell_{x+k} = \ell_x \cdot (1 - q_x),$$

og bagefter beregne antal dødsfald som

$$_k d_x = l_x - l_{x+k}.$$

2.3.5 Beregning af restlevetid og middellevetid

Vi betegner med $_kL_x$ den samlede gennemlevede tid i tabelbefolkningen i en alder mellem x og $x+k$. Da dødsfald er eneste afgangsårsag i tabelbefolkningen har vi

$$\begin{aligned} {_kL_x} &= \text{bidrag fra overlevende} + \text{bidrag fra døde} \\ &= k \cdot \ell_{x+k} + {_ka_x} \cdot {_kd_x} \\ &= {_ka_x} \cdot \ell_x + (k - {_ka_x}) \cdot \ell_{x+k}. \end{aligned}$$

Vi skal nu beregne den forventede restlevetid for en x-årig tabelperson. For en nyfødt er $x = 0$ og dermed bliver den forventede restlevetid til den forventede levetid, som betegnes med middellevetid. Lad T_x angive den samlede levetid i tabelbefolkningen efter x-års fødselsdagen, specielt er T_0 den samlede levetid i tabelbefolkningen. Vi beregner

$$\begin{aligned} T_x &= {_kL_x} + \cdots + {_kL_{x^{max}}} \\ &= {_ka_x} \cdot \ell_x + (k - {_ka_x}) \cdot \ell_{x+k} + \cdots + {_\infty a_{x^{max}}} \cdot \ell_{x^{max}}. \end{aligned}$$

I tabelbefolkning oplever ℓ_x personer deres x-års fødselsdag, så den gennemsnitlige levetid efter x-års fødselsdagen bliver

$$e_x = \frac{T_x}{\ell_x} = \text{gennemsnitlige restlevetid.}$$

Dette gennemsnit kaldes den forventede restlevetid for en x-årig tabelperson. På tilsvarende vis bliver middellevetid beregnet som

$$e_0 = \frac{T_0}{\ell_0} = \text{middellevetid.}$$

2.4 Danmark statistik

Med etableringen af den personstatistiske database har Danmarks Statistik fået nye muligheder for at beregne dødshyppighederne mere korrekt, idet databasen for alle personer i Danmark indeholder eksakt information om eventuel dødsdato og ind- og udvandringstidsdatoer. Der kan således for hver enkelt person udregnes nøjagtigt, hvor mange dage personen i en årsperiode har været i Danmark og hvor mange af dagene i årsperioden, personen har været død. Den søgte dødshyppighed skal præcist angive sandsynligheden for at dø på et bestemt alderstrin – det vil sige mellem to fødselsdage. For at opnå denne hyppighed laves der en særlig beregning for hver enkelt person fra fødselsdag til fødselsdag i en periode, der omfatter to kalenderår. I offentliggørelsen af middellevetid fra 19. marts 2010 er det kalenderårene 2008 og 2009, der ligger til grund for beregningerne. For alle personer, der var i den danske befolkning på et eller andet tidspunkt mellem deres fødselsdag i 2008 og i 2009, er der lavet en beregning for antallet af dage, personen var i Danmark og antallet af dage personen var død i perioden mellem de to fødselsdage. For personer, der ikke dør mellem to fødselsdage, vil antallet af dage som død naturligvis være 0. Efterfølgende laves der en sammenlægning for personer med samme køn og alderstrin for at få det samlede antal levedage og dødedage. Personer vil placeres på det alderstrin, som svarer til det antal år, de fyldte i startåret, hvilket i eksemplet vil sige 2008. En person, som fyldte 60 år 1. januar 2008 vil fx tilhøre de 60-årige. Det samme vil en person, der fyldte 60 år 31. december 2008. Der kan altså i yderste konsekvens være næsten et års forskel mellem den periode, som personer på samme alderstrin følges.