



LUNDS UNIVERSITET

Poisson-modellering av dödsbränder i Sverige

Elna Lindahl
Jonatan Hedberg

Kandidatuppsats i Statistik (15 hp)
Handledd av Björn Holmquist
HT 2016

SAMMANFATTNING

Denna uppsats är skriven på uppdrag av Lunds tekniska högskola och Myndigheten för samhällsskydd och beredskap. Uppdraget består av att modellera antalet dödsbränder som sker i Sverige över tid, samt undersöka ifall det har skett någon signifikant förändring. Detta görs utifrån ett antagande om att dödsbränder är Poisson-fördelade, därmed betraktas antalet dödsbränder över tid som en Poisson-process. Frågeställningen kan då besvaras genom att studera om, och i så fall hur, processens intensitet förändras över tid. Intensiteten modelleras med Poisson-regression med tidskovariat. Vi använder en modell med kvartalsdata som innehåller säsongskomponenter, en linjär trend och en brytpunkt i mitten av 2012. Den linjära trenden har en negativ koefficient, antalet dödsbränder minskar alltså succesivt under perioden. Brytpunkten år 2012 indikerar att intensiteten minskar med ca 20%. En nedgång konfirmeras även med ett t-test. Vi finner också att modellen, med ett 99% konfidensintervall, fångar upp variationen i processen med undantag för extremvärden.

SUMMARY

The authors of this thesis were tasked with modelling fatal fires in Sweden to answer the question, whether the quarterly number of fires has changed over time. The usage of the Poisson distribution in modelling accidents and mortalities is well attested. Problems arise when measuring the number of fire related mortalities, as several deaths can be attributed to one fire. This violates the assumption of independence of the Poisson distribution. To avoid this, the number of fatal fires is modelled.

The number of fires leading to mortalities is approached as a Poisson point process. This allows us to answer the question posed by looking at if and how the intensity of the process varies, which is done by applying Poisson regression with time covariates. The data used comes from the Swedish Civil Contingency Agency. It contains, among others, monthly data of all fatal fires from 1999 to 2016. The monthly data varies to a great extent, which makes modelling difficult. For that reason, the months are summed into quarters. Regressing on the quarterly data, a model containing seasonal components, a linear trend and a breakpoint in mid-2012 is successfully fitted. Winter has, according to the model a positive effect on the incident ratio, while summer has a negative effect. The linear trend acts negatively, meaning that the rate of fatal fires is decreasing. Lastly, the breakpoint, pushes the incident rate down with approximately 20%. With a 99% confidence interval, the model predicts the rate of fatal fires, except for extreme values. To corroborate a decline in the number of fatal fires after mid-2012, a t-test is performed.

1 INLEDNING

Att förstå och kartlägga kunskap om tidigare olycksfall är en förutsättning för ett effektivt förebyggande arbete mot dödsbränder. Sedan slutet av 90-talet har Myndigheten för samhällsskydd och beredskap (MSB) vidtagit åtgärder för en allt tydligare preventionsinriktning i sin insamling av statistik. År 1999 startades Dödsbrandsdatabasen (DBDB) där MSB sammanställer och redovisar uppgifter om bränder i Sverige som leder till dödsfall (Bergqvist & Jonsson, 2012). För att kunna utvärdera det förebyggande arbetet är det viktigt att förstå hur förekomsten av dödsbränder förändras över tid. På uppdrag av MSB och Brandteknik vid Lunds tekniska högskola analyseras därför i denna uppsats dödsbrandsstatistiken i Sverige från 1999 till och med 2015, i syfte att besvara frågan om, och i så fall hur, intensiteten förändras över tid. En Poisson-regression med tidskovariat anpassas till datamaterialet, med antalet dödsbränder per kvartal som beroende variabel.

Poisson-modeller har en lång historia inom modellering av olycksstatistik. Det första kända exemplet kommer från den ryske ekonomen och statistikern von Bortkewitsch (1898) som applicerade Poisson-fördelningen på ett datamaterial över antalet soldater i den Preussiska armén som dödats av hästsparkar. Rose (1990) tillämpar Poisson-regression på data över flygplansolyckor hos 35 amerikanska flygbolag och Johansson (1996) använder utökad Poisson-regression för att skatta effekten av sänkt hastighetsbegränsning på antalet motorvägsolyckor i Sverige.

I von Bortkewitsch exempel sammanfaller varje olycka med ett dödsfall. I de fall då en olycka istället kan orsaka flera dödsfall uppstår det problem. Det beror på att de dödsfall som sker till följd av en och samma olycka är kopplade till varandra och således inte kan ses som oberoende händelser. Detta bryter mot ett av Poisson-fördelningens grundantaganden, vilket gör att dödsfallen inte kan vara Poisson-fördelade. Michener och Tighe (1992) menar att det rentutav är olämpligt att använda dödsfallen vid Poisson-modellering av den här typen av olyckor, och väljer därför att tillämpa Poisson-regression på dödsolyckor istället för dödsfall i sin studie av olyckor på Colombias motorvägar. Fridstrøm et al. (1995) finner att Poisson-regressioner på antal dödsfall i trafiken har betydligt mer överspridning än regressioner på antalet dödsolyckor. De anser dock att Poisson-fördelningen kan ses som en ad hoc approximation av den sanna fördelningen för dödsfall, men att överspridning alltid är att vänta. Denna överspridning kan i vissa fall korrigeras, exempelvis genom att istället använda negativ binomialregression. Mer komplicerade metoder kan dock behövas ifall tidsserier med både överspridning och autokorrelation studeras (Johansson, 1996).

Även specifikt inom brandområdet har Poisson-fördelningen tillämpats på olycksstatistik. Mulvaney et al. (2008) tillämpar Poisson-regression på paneldata

över antal omkomna i brand per 100 000 invånare i England under perioden 1995 till 2004. I de fall då överspridning förekommer använder de negativ binomialregression. I studien *Smoke alarm ownership and house fire death rates in children* används Poisson-regression för att analysera hur antalet dödsfall i brand hos barn, per 100 000 invånare, beror på tiden samt på förekomst av brandvarnare, också detta i England (DiGiuseppi, et al., 1998). Jonsson, Runefors och Särdaqvist (2016) analyserar dödsbrandsstatistiken i Sverige under åren 1952 till 2013 med Poisson-regression. Den beroende variabeln är här antalet döda i brand per millioner invånare, med standardiserad åldersfördelning. Enligt resonemanget ovan kan valet av beroende variabel i dessa studier väntas ge upphov till överspridning i modellen.

I den här uppsatsen anpassas en Poisson-regressionsmodell som förklarar hur förekomsten av dödsbränder har förändrats över tid. Den beroende variabeln är antalet dödsbränder, detta för att kunna tillämpa Poisson-fördelningens oberoendeantagande. Vi finner att en modell med binära säsongvariabler och en linjär tidskomponent fångar upp mycket av variationen, med undantag för extremvärden. Den nedåtgående trenden som löper genom hela datamaterialet tyder på en successiv minskning i antalet dödsbränder. Det visar sig vidare att det sker en påtaglig förändring i processen efter sommaren 2012, då intensiteten sjunker, även detta inkluderas i modellen. Samtliga analyser genomförs i programvaran R.

Uppsatsen har följande struktur: i avsnitt 2 introduceras Poisson-processen och dess egenskaper, därefter beskrivs metoden för parameterskattning vid Poisson-regression, detta i avsnitt 3. I avsnitt 4 presenteras datamaterialet, och i avsnitt 5 redovisas resultatet. Slutligen, i avsnitt 6, diskuteras resultatet och potentiella bakomliggande orsaker till detta.

2 POISSON-PROCESSEN

För att definiera Poisson-processen behöver några av Poisson-fördelningens egenskaper först beskrivas. En slumpvariabel Y är Poisson-fördelad, $Po(\mu)$, om dess möjliga värden är positiva heltal och om den följer sannolikhetsfunktionen

$$P(Y = y) = \frac{\mu^y e^{-\mu}}{y!}. \quad (2.1)$$

Poisson-fördelningen uppträder då man räknar antalet företeelser som inträffar slumpmässigt i tiden eller i rummet. Det kan gälla antalet inkommande samtal till en telefonväxel, antalet partiklar i ett preparat eller som i vårt fall antalet dödsbränder per tidsenhet (Kingman, 1993). Antagandet om slumpmässighet är grundläggande och medför att händelserna inträffar oberoende av varandra.

Vidare gäller att

- (i) Om $Y \sim Po(\mu)$ så är $E(Y) = \mu$ och $V(Y) = \mu$.
- (ii) Om $Y \sim Po(\mu)$ och $X \sim Po(\nu)$ så är $Y + X \sim Po(\mu + \nu)$, givet att Y och X är oberoende

(Blom, et al., 2005).

När vi studerar Poisson-händelser som sker under ett tidsintervall $(0, t)$, av längden t , är det kutym att parametrisera väntevärdet som $E(Y) = \lambda t$. Vi mäter alltså fördelningens intensitet λ , det vill säga genomsnittligt antal händelser per tidsenhet. När vi studerar antalet Poisson-händelser över tid kan vi se det som en Poisson-process. Nedan definieras förutsättningarna för en Poisson-process.

Låt $N(s)$ vara ett positivt heltal som indikerar antalet inträffade dödsbränder i intervallet $(0, s]$. Vi definierar vidare att T_n är tidpunkten för den n :te dödsbranden så att $N(s) = 3$ när $T_3 \leq s < T_4$, det vill säga när den tredje branden har inträffat men ännu inte den fjärde. Notera vidare att om $s < r$, så är $N(s) \leq N(r)$ och $N(r) - N(s)$ antalet händelser som inträffar i intervallet $(s, r]$. Räkneprocessen $N(s)$ definieras då som en Poisson-process med intensiteten λ om följande kriterier är uppfyllda (se t.ex. Ross (1983) eller Durrett (2010)):

- (iii) $N(0) = 0$.
- (iv) Antalet händelser som inträffar i icke överlappande intervall är oberoende.
- (v) Antalet händelser i ett intervall av längden h är Poisson-fördelade med väntevärdet λh .

Då vi undersöker flera på varandra följande tidsintervall av lika längd, kan vi generalisera Poisson-processen till situationer där intensiteten förändras över tid. Intensiteten skrivs då som en funktion av tiden $\lambda(t)$, där $\mu(s, r) = \int_s^r \lambda(t) dt$ mäter det förväntade antalet händelser i intervallet $(s, r]$. Poisson-processen definieras också av den ackumulerade intensitetsfunktionen

$$\Lambda(s) = \int_0^s \lambda(t) dt, \quad (2.2)$$

vilket är väntevärdet av $N(s)$ (Fierro & Tapia, 2011). Om intensitetsfunktionen $\lambda(t)$ är konstant över tid betecknas processen som en homogen Poisson-process, och i de fall då intensiteten varierar med tiden betecknas den som inhomogen. I det inhomogena fallet kan alltså intensitet variera från ett tidsintervall till ett annat (Durrett, 2010). I denna studie betraktas varje kvartal som ett nytt tidsintervall, därmed skattas en ny intensitet för varje kvartal. Notera att det är intensitetsfunktionen som skattas i uppsatsen, vilket inte skall förväxlas med den ackumulerade intensitetsfunktionen.

3 POISSON-REGRESSION

För att undersöka om Poisson-processen är inhomogen och hur intensiteten i så fall förändras över tid, kan en Poisson-regression anpassas. Då skattas ett värde på $\mu(s, r) = \int_s^r \lambda(t) dt$ för varje tidsintervall $(s, r]$, i detta fall för varje kvartal. Parameterskattningen baserar sig på generaliserade linjära modeller (GLM), en flexibel generalisering av linjär regression som tillåter andra fördelningar från exponentialfamiljen än normalfördelningen. GLM är ett ramverk, introducerat av Nelder och Wedderburn (1972), som förenar ett flertal maximum likelihood-skattningar, däribland Poisson-regression. GLM bygger på att responsvariabler från fördelningar i exponentialfamiljen kopplas samman med en linjär modell via en länkfunktion. Istället för att hitta transformationer som simultant genererar normalfördelade variabler och linjära parametrar, tillämpas alltså en länkfunktion vars invers är den betingade medelvärdesfunktionen (McCulloch, et al., 2008).

Modellen består alltså av

1. En sannolikhets- eller täthetsfunktion från exponentialfamiljen.
2. En linjär prediktor $\eta = \mathbf{X}'\boldsymbol{\beta}$
3. En länkfunktion g sådan att $E(\mathbf{Y}) = g^{-1}(\eta)$.

Där \mathbf{X}' är en $(1 \times k)$ vektor bestående av de förklarande variablerna och $\boldsymbol{\beta}$ är den tillhörande $(k \times 1)$ parametervektorn.

En täthetsfunktion tillhör exponentialfamiljen om den kan skrivas på formen

$$f(y|\theta, \phi) = \exp\left[\frac{\theta y - b(\theta)}{a(\phi)} + c(y, \phi)\right] \quad (3.1)$$

där a , b och c är funktioner som beror av vilken fördelning som används, θ och ϕ (spridningsparametern) är parametrar kopplade till väntevärde respektive varians. Olika fördelningar ger således olika loglikelihoodfunktioner. För en generell framställning se exempelvis McCullagh & Nelder (1983). Poisson-fördelningens sannolikhetsfunktion (2.1) kan skrivas på samma form som ekvation (3.1). Sannolikhetsfunktionen logaritmeras och exponentieras, därtill låter vi $\theta = \ln\mu$.

Det ger följande sannolikhetsfunktion:

$$P(Y = y|\theta) = \exp[\theta y - \exp(\theta) - \ln(y!)] \quad (3.2)$$

För Poisson-fördelningen definieras därmed funktionerna a , b och c enligt följande:

$$a(\phi) = 1 \quad (3.3)$$

$$b(\theta) = \mu = \exp(\theta) \quad (3.4)$$

$$c(y, \phi) = -\ln(y!) \quad (3.5)$$

Dessutom är η definierat så att $\eta = \theta$, alltså är länkfunktionen

$$g = \ln(\mathbf{X}'_i \boldsymbol{\beta}), \quad (3.6)$$

vilket betyder att det betingade väntevärdet ges av

$$g^{-1} = E[Y_i | \mathbf{X}'_i] = \mu_i = \exp(\mathbf{X}'_i \boldsymbol{\beta}). \quad (3.7)$$

Vi får då loglikelihoodfunktionen

$$l(\mathbf{Y}, \boldsymbol{\beta}) = \sum_{i=1}^n \mathbf{X}'_i \boldsymbol{\beta} y_i - \exp(\mathbf{X}'_i \boldsymbol{\beta}) + \ln(y_i!). \quad (3.8)$$

Denna funktion maximeras med avseende på $\boldsymbol{\beta}$, vilket görs numeriskt då funktionen inte är linjär. Den metod som används här är 'Fisher scoring', som likt Newton-Raphsons metod går ut på att man itererar sig fram till lösningen (McCullagh & Nelder, 1983). Det innebär att ett startvärde på $\boldsymbol{\beta}^m$ väljs. Scorefunktionen tillsammans med Hessianen, d.v.s. en matris innehållande alla andraderivator av loglikelihoodfunktionen med avseende på de olika $\beta_{i,j}$, används sedan för att hitta ett nytt värde $\boldsymbol{\beta}^{m+1}$ som ligger närmare lösningen på maximeringsproblemet. Denna procedur fortsätter tills $\boldsymbol{\beta}^m = \boldsymbol{\beta}^{m+1}$, vilket kan visas vara detsamma som att iterativt skatta koefficienterna via vägd minstakvadratmetod (Dobson & Barnett, 2008).

3.1 Förekomst av överspridning i Poisson-regression

Som nämnt ovan karakteriseras Poisson-fördelningen av att väntevärde och varians är lika (i). I praktiken är det dock inte ovanligt att räknedata som förväntas kunna modelleras med Poisson-regression, har en varians som överstiger väntevärdet, en så kallad överspridning. Om ursprungliga data har överspridning kan källorna till detta ibland fångas upp i en korrekt specificerad modell och eliminera problemet (Fridström & Ingebrigtsen, 1991). När detta inte är möjligt finns det flera alternativ för att inkludera överspridningen i modellen, ofta används negativ binomialregression eller quasi Poisson där spridningsparametern ϕ inte är fix utan tillåts variera (Zeileis, et al., 2008). Källan till förhöjd varians kan vara att det finns ett beroende, det vill säga att flera

observationer i räkneprocessen framkallats av samma händelse. Överspridningen kan också bero på att det betingade väntevärdet varierar för olika delar i materialet, och materialet består därmed av flera olika Poisson-fördelningar. Intensiteten kan då inte skattas väntevärdesriktigt för alla observationer simultant, men kan istället skattas betingat på t.ex. ålder eller kön. Notera att i de fall då förhöjd varians är en konsekvens av systematiska fel i modellen eller att förklarande variabler som borde funnits med saknas, kan detta inte lösas genom att modellera överspridningen, modellen behöver istället korrigeras (Berk & MacDonald, 2008). Förekomst av överspridning i modellen utvärderas med hjälp av ett överspridningstest. Testet baserar sig på den linjära variansfunktion som kan användas vid negativ binomialregression (Cameron & Trivedi, 2013).

$$\text{Var}(y_i|x_i) = (1 + \alpha)\mu_i = (\text{dispersion})\mu_i. \quad (3.9)$$

Nollhypotesen, $\alpha = 0$, det vill säga att varken över- eller underspridning förekommer, testas mot mothypotesen om överspridning, detta med ett Z-test.

3.2 Modellvalidering

Vad gäller anpassningsgraden för en Poisson-regression kan måttet 'deviance' användas. 'Deviance' är loglikelihoodkvoten mellan en fullständig modell med en parameter för varje observation och den skattade modellen, multiplicerad med två

$$D = 2[l(y, b_{fullständig}) - l(y, b_{skattad})]. \quad (3.10)$$

Förutsatt att den skattade modellen är korrekt specificerad, är 'deviance' approximativt χ^2 -fördelad med frihetsgraderna $n - k$, där n är antalet observationer och k antalet parametrar som skattas (Dobson & Barnett, 2008). 'Deviance' kan således användas för att testa nollhypotesen att det inte finns någon skillnad mellan den fullständiga och den skattade modellen.

För modellvalidering studeras även Pearson-residualerna. De räknas ut genom att dividera residualerna med standardavvikelsen hos den oberoende variabeln. För Poisson-regression blir detta

$$r_i = \frac{y_i - \hat{y}_i}{\sqrt{\hat{y}_i}}. \quad (3.11)$$

Residualerna plottas mot de predikterade värdena, det vill säga mot \hat{y} och analyseras grafiskt. Om något mönster kan observeras betyder det i de flesta fall att överspridning förekommer eller att modellen är felspecificerad (Hilbe, 2014). Residualerna plottas även mot index för att utvärdera i vilka tidsperioder modellen ger bäst anpassning.

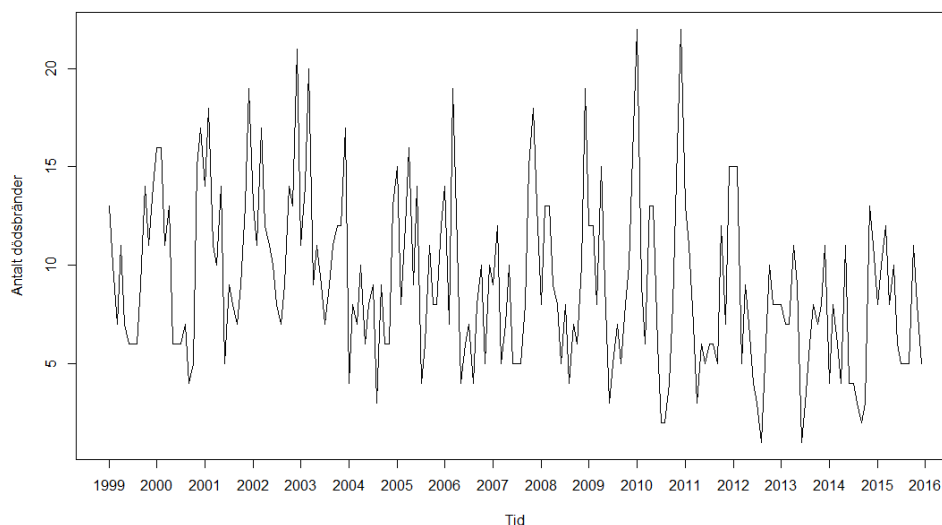
Slutligen valideras modellen grafiskt genom att jämföra observerade värden med skattade värden och tillhörande 99 % konfidensintervall. Medelfelen beräknas som en linjärkombination av koefficienternas medelfel vid varje observation. Med utgångspunkt från likelihoodskattningens approximativa normalfördelningsegenskaper används ett tabellvärde från normalfördelningen, vilket tillsammans med medelfelet ger konfidensintervallet.

4 DATAMATERIAL

Det datamaterial som används i denna rapport kommer från Dödsbrandsdatabasen (DBDB) som Myndigheten för samhällsskydd och beredskap (MSB) ansvarar för och tillhandahåller. Databasen innehåller bland annat ett register över antalet dödsbränder i Sverige för varje månad från 1999 till 2014. Preliminär statistik finns även för år 2015. Statistiken baseras framför allt på uppgifter från kommunal räddningstjänst, men kompletteras av uppgifter från polis och sjukvård. För att säkerställa kvaliteten i statistiken görs även kontroller mot Rättsmedicinalverket, för att konfirmera att dödsfallen sker till följd av brand, samt för att upptäcka dödsfall som inte rapporterats av de andra myndigheterna. En dödsbrand definieras i databasen som en brand som lett till en eller flera personers död, där en brand definieras som en okontrollerad låga, glöd eller rökutveckling som orsakar skada. För att en person ska räknas som död i brand måste ett antal kriterier uppfyllas:

1. Offret har omkommit till följd av skador som orsakats av brand eller explosionsartad förbränning. Både avsiktliga bränder och oavsiktliga bränder finns med i statistiken. Dödsfall i trafikolyckor är med endast i de fall då det framgår att offret var vid liv vid det tillfälle då brand eller brandgaser nådde kroppen.
2. Dödsfallet inträffade inom 30 dagar efter branden.
3. Dödsbranden skedde i Sverige.

Insamlad statistik över antal dödsfall och antal dödsbränder anses vara jämförbar mellan åren 1999 och 2014 (McIntyre, 2015).



Figur 1: Antal dödsbränder per månad, januari 1999 till och med december 2015.

TABELL 1: Deskriptiv statistik

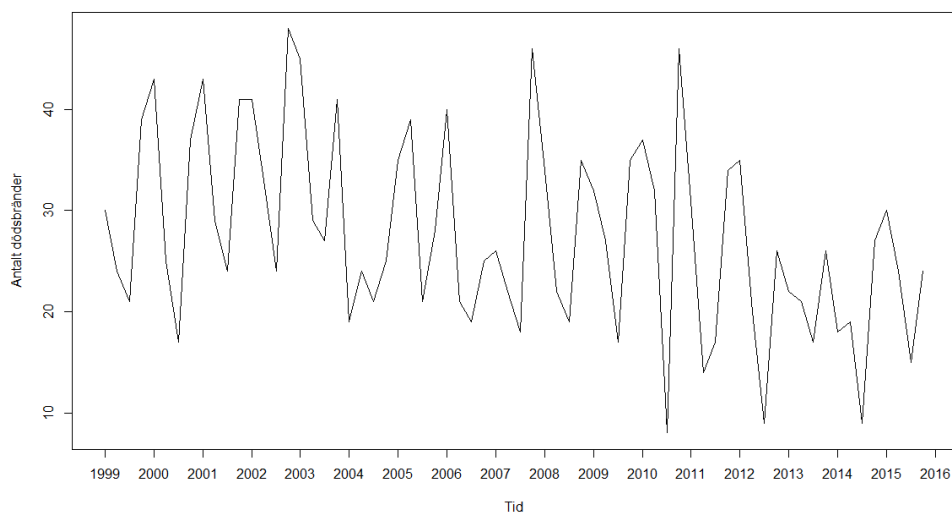
Antal dödsbränder per kvartal

Medelvärde	27,5
Varians	92,9
Min.	8
Max.	48

En betydande variation kvarstår emellertid även efter att tidsserien har aggregerats från månader till kvartal, vilket framgår av figur 2 som presenteras i nästa avsnitt. I tabell 1 ser vi att variansen är mer än tre gånger så stor som medelvärdet. Vi har således en överspridning i kvartalsdatamaterialet. Skillnaden mellan det största och minsta värdet är också stor.

5 RESULTAT

I figur 2 presenteras observerat antal bränder per kvartal, processen uppvisar en tydlig periodicitet med fler dödsbränder under vinterhalvåret än under sommaren. Nedan utvärderas några alternativ för hur detta, samt förändringar i processen kan modelleras. Därefter redovisas regressionsmodellen samt dess skattade koefficienter.



Figur 2: Antalet dödsbränder per kvartal

Centralt för analysen är att specificera om, och i så fall hur, intensiteten i processen förändras gradvis över tid, därför inkluderas en linjär trend tid ($tid=1, \dots, 68$) i modellen. Säsongsvariationen inkluderas i modellen i form av två binära variabler, D_{vinter} och D_{sommar} , där vinter representerar kvartal ett och fyra, och sommar indikerar kvartal tre. Den alternativa framställningen, att låta varje kvartal utgöra en egen binär variabel, utvärderas också. Här förloras dock en frihetsgrad utan ökad anpassningsgrad; kvartal ett är inte signifikant skilt från kvartal fyra.

Den grafiska analysen pekar även på en förändring från och med sommaren 2012, då det genomsnittliga antalet dödsbränder tycks sjunka. Hur detta samband skall modelleras undersöks med hjälp av ett antal alternativa modeller. En tes är att det främst är den höga förekomsten av dödsbränder under vintrarna som dämpas efter denna brytpunkt, och en dummy för tiden från och med kvartal tre 2012, D_{2012} , inkluderas därför initialt som en interaktion med D_{vinter} . Interaktionen utgör, enligt LR-testet, ett signifikant bidrag till en modell som innehåller tid , D_{vinter} och D_{sommar} . Även Akaïkes informationskriterium (AIC) och värdet på 'deviance' förbättras. Att låta D_{2012} finnas med som en binär variabel under alla årstider ger dock en ännu bättre anpassning då förändringen efter denna brytpunkt visar sig vara signifikant även för vår- och sommarmånaderna. Ett tredje alternativ är att modellera tiden efter juni 2012 med en separat trend; att låta D_{2012} interagera med variabeln tid . Detta ger nästan samma resultat som att inkludera D_{2012} som en fristående variabel. Resultatet är alltså tvetydigt, från och med kvartal tre 2012 kan vi se en genomgående lägre genomsnittsnivå i antal dödsbränder per kvartal, men vi kan även välja att tolka detta som att den nedåtgående trenden förstärks efter denna brytpunkt. Sett till hur mycket enskilda extremvärden påverkar skattningen och

den grafiska anpassningen, är de två modellerna likvärdiga. Att modellera en nivåskillnad, istället för en skillnad i trend, ger dock en marginellt bättre anpassning enligt AIC och 'deviance'. Därmed, anpassas följande Poisson-regression:

$$\ln\lambda = \eta = \beta_0 + \beta_1\text{tid} + \beta_2D_{\text{vinter}} + \beta_3D_{\text{sommar}} + \beta_4D_{2012}.$$

De skattade koefficienterna presenteras i tabell 2. Notera att koefficienten β_1 visar på förändringen i det logaritmerade antalet dödsbränder per kvartal, då vi går från ett kvartal till nästa. För de binära variablerna tolkas β_i som förändringen i den logaritmerade responsvariabeln då värdet hos prediktorn går från 0 till 1. För att låta en enhets förändring i respektive oberoende variabel spegla förändringen direkt i den beroende variabeln beräknas e^{β_i} . Då

$$\lambda = e^{\beta_0 + \beta_1\text{tid} + \beta_2D_{\text{vinter}} + \beta_3D_{\text{sommar}} + \beta_4D_{2012}} = e^{\beta_0} e^{\beta_1\text{tid}} e^{\beta_2D_{\text{vinter}}} e^{\beta_3D_{\text{sommar}}} e^{\beta_4D_{2012}},$$

följer att de exponentierade koefficienterna kan tolkas som den relativa effekten av respektive prediktor, på den skattade intensiteten. Dessa relativa effekter återfinns i tabell 3 med tillhörande 95 % konfidensintervall. Konfidensintervallen är likelihoodbaserade, vilket innebär att de beräknas med utgångspunkt från Poisson-fördelningen (Hilbe, 2014). Låt nu interceptet, det vill säga kvartal två, fungera som jämförelsenivå. Jämfört med denna basnivå förväntas antal dödsbränder per kvartal vara i genomsnitt mellan 21 och 52 % högre under vinterhalvåret, om alla andra faktorer hålls konstanta. På motsvarande sätt sjunker antalet dödsbränder med mellan 16 och 38 % då vi går från vår till sommar. Den förändring över tid som processen uppvisar, framträder bland annat som en nivåskillnad efter sommaren 2012. Genomsnittligt antal dödsbränder per kvartal ligger här på en signifikant lägre nivå, när man tar hänsyn till säsongsvariationer. Som framgår av tabell 3 rör det sig om en relativ minskning om mellan 4,5 och 31,6 %, med 95 % konfidens. Notera att den binära variabeln D_{2012} bör betraktas som ett sätt att modellera den förändring som sker, och koefficientens konfidensintervall fungerar då som ett mått på hur stor förändringen är. Konfidensintervallet och koefficientens signifikans kan däremot inte betraktas som ett test för att brytpunkten inte är en slump händelse. Skillnaden mellan antalet dödsbränder före och efter brytpunkten år 2012 kan bekräftas med ett Welch's t-test. Medelvärde under perioden 1999 till och med kvartal två 2012 ($\hat{\mu}_1$) beräknas till 29,35 dödsbränder per kvartal, och motsvarande värde för resterande period ($\hat{\mu}_2$) är 20,50. Nollhypotesen $H_0: \mu_1 = \mu_2$ testas mot mothypotesen $H_1: \mu_1 \neq \mu_2$. Med stöd i Centrala gränsvärdesatsen är de skattade väntevärdena approximativt normalfördelade, och testfunktionen därmed approximativt t-fördelad. Testfunktionens värde beräknas till 4,12 med 29,73 frihetsgrader, vilket motsvarar ett p-värde på 0,00028.

$$T = \frac{\hat{\mu}_1 - \hat{\mu}_2}{\sqrt{\frac{\hat{\sigma}_1^2}{n_1} + \frac{\hat{\sigma}_2^2}{n_2}}} \quad (5.1)$$

Den nedåtgående trend som löper genom hela processen är även den signifikant skild från noll och konfidensintervallet i tabell 3 visar på en nedgång med mellan 0,08 och 0,7 % per kvartal. Sett över hela processen innebär det en minskning om ca 23 %.

TABELL 2: RESULTAT POISSON-REGRESSION

BEROENDE VARIABEL: ANTAL DÖDSBRÄNDER PER KVARTAL

Variabel	Koefficient	Medelfel	P-värde
Intercept	3,3797	0,0655	2×10^{-16} ***
Tid	-0,0039	0,0016	0,0145*
D _{vinter}	0,3043	0,0568	$8,56 \times 10^{-8}$ ***
D _{sommar}	-0,3237	0,0753	$1,70 \times 10^{-5}$ ***
D ₂₀₁₂	-0,2128	0,0851	0,0124*

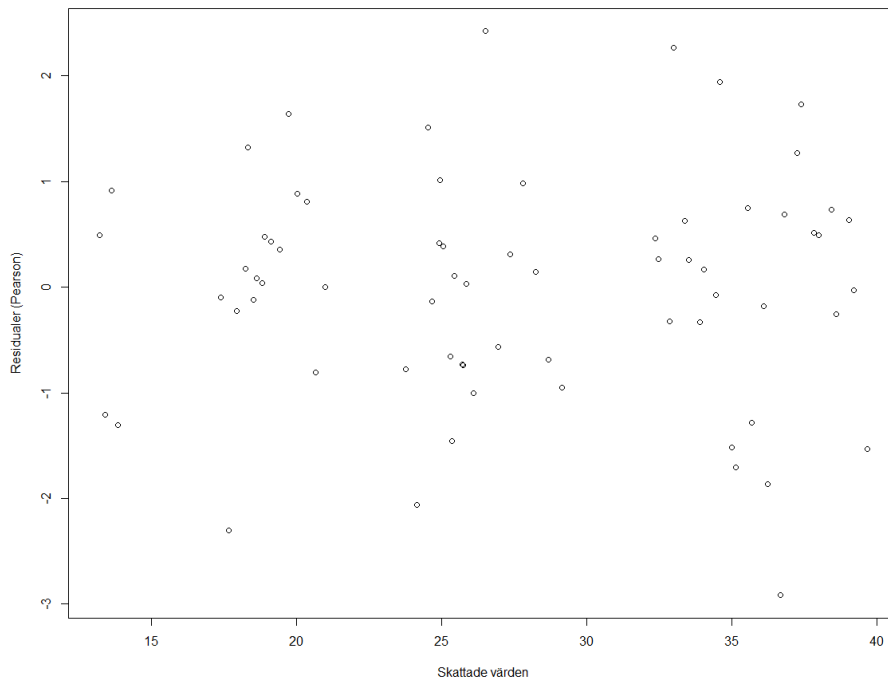
Kommentar: Tabellen visar tvåsidigt p-värde för test av nollhypotesen att den verkliga koefficienten är noll. Antal observationer: 68. Residual deviance: 80.104, 63 frihetsgrader, vilket ger p-värde 0.0718. Överspridningsparameter: 1.1309, p-värde: 0.2648.

TABELL 3: RESULTAT POISSON-REGRESSION

EXPONENTIERADE KOEFFICIENTER

Variabel	$e^{\hat{\beta}}$	95 % Konfidensintervall för $e^{\hat{\beta}}$	
Intercept	29,3613	25,7885	33,3379
Tid	0,9961	0,9930	0,9992
D _{vinter}	1,3557	1,2139	1,5169
D _{sommar}	0,7235	0,6238	0,8380
D ₂₀₁₂	0,8083	0,6837	0,9547

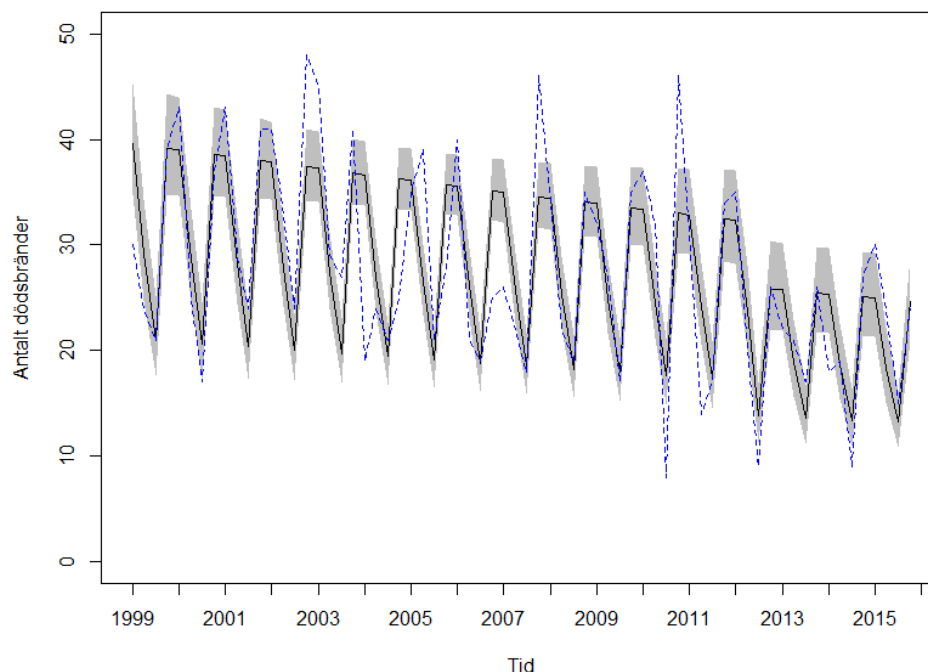
Modellen är en tillfredställande anpassning till materialet baserat på LR-testet av deviance. Residualerna uppvisar inga tecken på beroendestruktur, vilket kan ses i figur 3. När residualerna plottas över tid, kan det även konstateras att modellen ger en bättre anpassning efter brytpunkten år 2012.



Figur 3: Residualer mot skattade värden

Den överspridning som finns i processen fångas upp väl i modellen; spridningsparametern är ej signifikant skild från 1 och korrekta konfidensintervall kan således beräknas. I figur 4 presenteras observerade samt skattade värden med tillhörande 99 % konfidensintervall. Det kan konstateras att modellen fångar upp variationen på ett tillfredställande vis, med undantag för extremvärden.

Som nämnt i avsnitt 3.1 kan överspridning förekomma som en följd av att processen består av mer än en Poisson-process, med olika värden på λ . Utifrån detta resonemang kan resultatet tolkas som att sex olika Poisson-processer har identifierats: vår, sommar och vinter, alla tre med en process före sista juni 2012 och en efter. I dessa sex Poisson-processer har även intensiteten ett svagt tidsberoende.



Figur 4: Skattade värden heldragna i svart med konfidensintervall i grått, och faktiska värden streckade i blått.

6 DISKUSSION

I den här uppsatsen modelleras antalet dödsbränder i Sverige mellan år 1999 och år 2015 med Poisson-regression. När säsongsvariationen fångats upp i binära variabler finner vi en nedåtgående trend under hela perioden. Dessutom sker en trappstegsliknande nedgång efter kvartal två år 2012; antalet dödsbränder per kvartal sjunker med mellan 4,5 och 31,6 %, med 95 % konfidens. I detta avsnitt diskuteras modellen samt nedgången som skedde efter mitten av år 2012.

Som bekant kan Poisson-processen betraktas som en sekvens av Bernoulliförsök, vilket i detta fall skulle innebära att varje brand ses som ett försök med två möjliga utfall, antingen leder branden till dödsfall eller gör den det ej. För en djupare förståelse för processen skulle den alltså kunna analyseras med utgångspunkt i antalet bränder under perioden. Sannolikheten för att en brand leder till dödsfall skulle då kunna skattas och man skulle även kunna undersöka hur denna sannolikhet förändras med tiden eller med andra kovariat. Detta skulle kräva en mycket noggrann avvägning vid val av definition av 'brand'. Somliga bränder, till exempel brand i en papperskorg på en öppen plats, leder i stort sett aldrig till dödsfall och bör därför inte betraktas som ett 'försök' i detta sammanhang. Man skulle kunna tänka sig att låta alla bränder som leder till uttryckning finnas med i materialet men det skulle ändå innebära en över- eller

underteckning i antalet bränder; överteckning i samband med falskt larm och underteckning i de fall då räddningstjänsten inte kontaktas i samband med en brand som kunde ha lett till dödsfall. Statistiken för dödsbränder är mycket mer exakt definierad och fullständig vilket innebär att antalet dödsbränder per brand troligtvis skulle överskattas eller underskattas systematiskt.

Valet av beroende variabel i modellen är som nämnt inte helt självklart. Vid modellering med Poisson-regression är det en fördel att analysera antalet dödsbränder, mot bakgrund av fördelningens oberoendeantagande. Men att även studera utvecklingen över tid med avseende på hur många individer som omkommer i varje brand, är naturligtvis angeläget. Ett alternativ som berörts inledningsvis är att använda Poisson-fördelningen som en approximation för antalet dödsfall i brand, och justera för eventuell överspridning med hjälp av negativ binomialregression eller quasi Poisson-regression. Ett mer idealt, men möjligen mer tidskrävande alternativ, är att modellera antalet dödsbränder med Poisson-fördelningen och sedan inom dödsbränderna hitta ett sätt att modellera hur många individer som omkommer i varje brand. Antal dödsfall per dödsbrand betraktas därmed som en slumpvariabel som beskriver dödsbrändernas omfattning.

Uppsatsens resultat väcker frågan om vad som har förorsakat minskningen i antalet dödsbränder efter år 2012. Då detta är en uppsats i ämnet statistik saknas sakkunskap inom brandområdet för att besvara frågan på ett välinformerat vis. Med detta sagt kan några spekulationer och förslag ändå presenteras. Mulvaney et al. (2008) beskriver hur studier i England visat att barn till långtidsarbetslösa föräldrar har en förhöjd risk att dö i brandrelaterade olyckor jämfört med barn vars föräldrar har sysselsättning. Av statistiken från IDA framgår att de flesta dödsbränderna inträffar i hemmet (Myndigheten för samhällsskydd och beredskap, 2016). Jonsson et al. (2016) menar att det tycks finnas ett samband mellan att små barn, från mitten av 60-talet och framåt, spenderat mindre tid hemma och att risken att dö i brand har minskat i samma grupp. I Sverige tog arbetslösheten fart i samband med finanskrisen 2008 och nådde sin topp på 8,6 % år 2010, för att sedan sjunka och stabilisera sig på omkring 8 % från år 2011 och framåt (Statistiska centralbyrån, 2016). Detta skulle kunna förklara en del av den minskning i antalet dödsbränder som sker år 2012. Notera dock att antalet dödsbränder inte ökar på samma markanta vis när arbetslösheten stiger mellan år 2008 och år 2009. Arbetslösheten efter år 2012 ligger på en högre nivå än vad den gjorde före finanskrisen, detsamma gäller dock inte för det genomsnittliga antalet dödsbränder.

Det är inte omöjligt att antalet dödsbränder har ett inneboende cykliskt mönster, det vill säga att det finns samhällsmekanismer som gör att toppar följs av dalar och vice versa. En ökad förekomst av dödsbränder leder till exempel till tidningsrubriker och ökad medvetenhet om brandsäkerhet hos allmänheten. När

sedan antalet allvarliga brandolyckor dämpas riktas oro och uppmärksamhet till andra samhällsproblem som för stunden upplevs som mer akuta än brandsäkerhet.

Myndigheters agerande kan även ha bidragit till förändringen. Exempelvis startade MSB kampanjen ”Aktiv mot brand” år 2012. Kampanjen syftar till att öka enskilda individers kunskap om brandrisker, framförallt i hemmen, samt om brandskyddsutrustning och hur man bäst agerande vid brand. Kampanjen har sedan dess aktualiserats fyra gånger per år (Myndigheten för samhällsskydd och beredskap, 2012). Räddningstjänsten Syd har sedan år 2010 gjort hembesök för att förebygga uppkomsten av bostadsbränder. De uppger att antalet bostadsbränder har minskat med 30 % mellan år 2010 och 2015 (Räddningstjänsten Syd, 2016). Hembesök har även skett i andra län så som Västra Götalands län (Borås Tidning, 2014) och Värmlands län (SVT Nyheter, 2010).

Nedgången efter kvartal två, år 2012 skulle slutligen kunna bero på slumpen. Då modelleringen av brytpunkten baseras på en grafisk analys av processen, kan inte p-värdet som presenteras i tabell 2 användas som argument för att nedgången inte är en avvikelse. Detta eftersom vi testar en företeelse som vi redan har observerat. Det som istället talar för att denna nedgång är en verklig förändring i intensiteten, är dock att extremvärdena blir färre och variansen minskar efter sommaren 2012. Poisson-fördelningens väntevärde och varians är lika stora, vilket gör att en minskning av intensitet måste innebära en minskning av varians och med detta en minskad sannolikhet för stora avvikelser från väntevärdet. Därmed är det osannolikt att nedgången endast beror på slumpen.

LITTERATURFÖRTECKNING

- Bergqvist, A., & Jonsson, A. (2012). *Dödsbränder i Sverige - Kvalitetsgranskning av MSBs dödsbrandsdatabas*. Myndigheten för samhällsskydd och beredskap. Hämtat från <https://www.msb.se/sv/Produkter--tjanster/Publikationer/Publikationer-fran-MSB/Dodsbrander-i-Sverige/> den 17 12 2016
- Berk, R., & MacDonald, J. (2008). Overdispersion and Poisson Regression. *Journal Quantitative of Criminologi*, 24(3), ss. 269-284.
- Blom, G., Enger, J., Englund, G., Grandell, J., & Holst, L. (2005). *Sannolikhets teori och statistikteori med tillämpningar*. Lund: Studentlitteratur AB.
- Bortkewitsch, L. (1898). *Das Gesetz der Kleinen Zahlen*. Leipzig: B. G. Teubner.
- Borås Tidning. (den 20 Mars 2014). Räddningstjänsten knackar dörr för att informera om brandrisk. Borås, Västra Götalands län, Sverige. Hämtat från www.bt.se/boras/raddningstjansten-knackar-dorr-for-att-informera-om-brandrisk/ den 29 12 2016
- Cameron, C. A., & Trivedi, P. K. (2013). *Regression Analysis of Count Data* (2 uppl.). Cambridge: Cambridge University Press.
- DiGuseppi, C., Roberts, I., & Li, L. (1998). Smoke alarm ownership and house fire death rates in children. *Journal of Epidemiology and Community Health*, 52(11), ss. 760-761.
- Dobson, A. J., & Barnett, A. G. (2008). *An Introduction to Generalized Linear Models* (3:e uppl.). Boca Raton: Chapman & Hall.
- Durrett, D. (2010). *Essentials of Stochastic Processes*. New York: Springer-Verlag.
- Fierro, R., & Tapia, A. (2011). Testing Homogeneity for Poisson Processes. *Revista Colombiana de Estadística*, 34(3), ss. 421-432.
- Fridstrøm, L., & Ingebrigtsen, S. (1991). An Aggregate Accident Model Based on Pooled, Regional Time-Series Data. *Accid. Anal. and Prev.*, 23, ss. 363-378.
- Fridstrøm, L., Ifver, J., Ingebrigtsen, S., Kulmala, R., & Krogsgård Thomsen, L. (1995). Measuring the Contribution of Randomness, Exposure, Weather, and Daylight to the Variation in Road Accident Counts. *Accident Analysis and Prevention*, 27(1), ss. 1-20.
- Hilbe, J. M. (2014). *Modeling Count Data*. New York: Cambridge University Press.
- Johansson, P. (1996). Speed Limitation and Motorway Casualties: A Time Series Count Data Regression Approach. *Accident Analysis and Prevention*, 28(1), ss. 73-87.

- Jonsson, A., Runefors, M., & Särndqvist, S. (2016). Fire-Related Mortality in Sweden: Temporal Trends 1952 to 2013. *Fire Technology*, 52(6), ss. 1697-1707.
- Kingman, J. F. (1993). *Poisson Processes*. New York: Oxford University Press.
- McCullagh, P., & Nelder, J. A. (1983). *Generalized Linear Models*. Cambridge: University Press.
- McCulloch, C. E., Searle, S. R., & Neuhaus, J. M. (2008). *Generalized, Linear, and Mixed Models*. New Jersey: John Wiley & Sons, Inc.
- McIntyre, C. (2015). *Dödsbrandsstatistiken*. Karlstad: Myndigheten för samhällsskydd och beredskap.
- Michener, R., & Tighe, C. (1992). A Poisson Regression Model of Highway Fatalities. *The American Economic Review*, 82(2), ss. 452-456.
- Mulvaney, C., Kendrick, D., Towner, E., Brussoni, M., Hayes, M., & Powell, J. (2008). Fatal and non-fatal fire injuries in England 1995-2004: time trends and inequalities by age, sex and area deprivation. *Journal of Public Health*, 31(1), ss. 154-161.
- Myndigheten för samhällsskydd och beredskap. (2012). *Aktiv mot brand*. Hämtat från <https://www.msb.se/sv/Forebyggande/Brandskydd/Kampanj-Aktiv-mot-brand/Om-kampanjen/> den 28 12 2016
- Myndigheten för samhällsskydd och beredskap. (2016). *Statistikdatabasen*. Hämtat från <https://ida.msb.se/ida2#page=a0002> den 26 10 2016
- Nelder, J. A., & Wedderburn, R. W. (1972). Generalized Linear Models. *Journal of the Royal Statistical Society*, 135(3), 370-384.
- Rose, N. L. (1990). Profitability and Product Quality: Economic Determinants of Airline Safety Performance. *Journal of Political Economy*, 98(5), ss. 944-964.
- Ross, S. M. (1983). *Stochastic Processes*. Berkeley: John Wiley & Sons, Inc.
- Räddningstjänsten Syd. (2016). *Bostadsbränderna minskar i Räddningstjänsten Syds område*. Hämtat från <http://www.rsyd.se/om-oss/press/pressmeddelanden/2016/bostadsbranderna-minskar-i-raddningstjansten-syds-omrade/> den 27 12 2016
- Statistiska centralbyrån. (2016). *Arbetslösheten tog fart under finanskrisen*. Hämtat från <http://www.scb.se/hitta-statistik/sverige-i-siffror/samhallets-ekonomi/arbetslosheten/> den 23 12 2016
- SVT Nyheter. (den 10 maj 2010). Karlstadregionen: Brandsäkerheten kontrollerad. Karlstad, Värmlands län, Sverige. Hämtat från <http://www.svt.se/nyheter/lokalt/varmland/karlstadregionen-brandsakerheten-kontrollerad?gmenu=open> den 29 1 2016

Zeileis, A., Kleiber, C., & Jackman, S. (2008). Regression Models for Count Data in R. *Journal of Statistical Software*, 27(8).