97/22 Notater 1997

Statistics Norway

Statistisk sentralbyrå

Håkan Lövkvist

Standardiserte rater - en metodebeskrivelse med eksempler fra dødsårsaksstatistikken

Notater

Avdeling for personstatistikk/Kontor for EDB

Forord

Dette notatet er tenkt å brukes som en veiledning, eller håndbok, for beregning, tolking og forståelse av standardiserte rater innenfor epidemiologien og samfunnsmedisinen. Det innledes med en presentasjon av noen sentrale begreper i kapittel 2. I kapittel 3 tar vi oss av selve beregningen av de standardiserte ratene, mens kapittel 4 bør bli betraktet som en introduksjon til variansberegning og testing av hypoteser. Kapittel 4 er egnet for den som er særlig interessert i å utføre mer dypgående analyser av et datamaterial og som ønsker et enkelt verktøy for måling av den statistiske usikkerheten rundt ratene. I kapittel 5 ser vi på to konkrete eksempler fra dødsårsaksstatistikken. I tillegg til å sette tall inn i formlene presenterer vi også de programmene i statistikkverktøyet SAS som vi har brukt for å beregne de standardiserte ratene.

Det praktiske arbeidet som opprinnelig ligger bak dette notatet er til store deler blitt utført i samarbeid med Seksjon for helsestatistikk. I denne forbindelsen har Finn Gjertsens fagkunnskap innenfor dødsårsaksstatistikken vært til stor hjelp. Når det gjelder det metodemessige inneholdet har Siri Størmer og Leiv Solheim, Seksjon for statistiske metoder og standarder, samt Nico Keilman, Seksjon for demografi og levekårsforskning, bidratt med interessante og konstruktive synspunkter og kommentarer.

Forfatteren vil rette en takk til alle som har medvirket i forbindelse med arbeidet med dette notatet.

Oslo, 21. mars 1997

Håkan Lövkvist

Innhold

1 INNLEDNING 5

2 NOEN EPIDEMIOLOGISKE BEGREPER 6

2.1 DESKRIPTIVE EPIDEMIOLOGISKE MÅL 6
2.2 POPULASJONER 6
2.3 UJUSTERT OG JUSTERT RATE 7

3 EFFEKTMÅLING VED INDIREKTE OG DIREKTE STANDARDISERING 9

3.1 RÅRATE OG STRATUMSPESIFIKK INSIDENSRATE 9
3.2 INDIREKTE STANDARDISERING 10
3.2.1 Noen ord om råraten i forbindelse med indirekte standardisering 10
3.2.2 SMR (Standardized Mortality Ratio eller Standardized Morbidity Ratio) 10
3.3 DIREKTE STANDARDISERING 11
3.3.1 Direkte standardisert insidensrate 11
3.3.2 SRR (Standardized Risk Ratio) 12

4 STATISTISK USIKKERHET, KONFIDENSINTERVALLER OG HYPOTESETESTING 14

4.1 KONFIDENSINTERVALLER OG HYPOTESETESTING - EN KORT TEORETISK INNFØRING 14
4.1.1 Konfidensintervaller 14
4.1.2 Ensidige konfidensintervaller 15
4.1.3 Hypotesetesting 16
4.2 POISSONFORDELING, BINOMISK FORDELING OG NORMALFORDELINGSTILNÆRMING 17
4.3 VARIANSBEREGNING AV USTANDARDISERTE OG STANDARDISERTE RATER 19
4.3.1 Noen ord om hypotesetesting av rater under normalfordelingstilnærming 19
4.3.2 Stratumspesifikk insidensrate 19
4.3.3 Rårate og SMR 20
4.3.4 Direkte standardiserte rater 21

5 EKSEMPLER FRA DØDSÅRSAKSSTATISTIKKEN 23

5.1 INNLEDNING 23
5.2 EKSEMPEL 1: INDIREKTE STANDARDISERING AV DØDELIGHET I HJERTE- OG KARSYKDOMMER I OSLO BYDELER23
5.3 EKSEMPEL 2: DIREKTE STANDARDISERING AV DØDELIGHET IALT I OSLO BYDELER 30
5.4 TOLKING AV RESULTATENE 35
5.5 SAS-PROGRAMMER FOR BEREGNINGENE 38

6 LITTERATUR 44

Tabellregister

TABELL 4.1. OVERSIKT OVER FORMULERING AV NULL- OG MOTHYPOTESE FOR TRE HYPOTESETESTMODELLER..... 17

TABELL 4.2. MODELLER FOR HYPOTESETESTING UNDER NORMALFORDELING, HVOR VI HAR EN STOKASTISK VARIABEL X SOM ER N(μ,V(μ)) (NORMALFORDELT MED EN PARAMETER SOM HAR EN FORVENTET VERDI μ OG EN VARIANS V(μ)). TESTFUNKSJONER OG BESLUTNINGSGRUNNLAG FOR EN-, TO- OG TRESTJERNERS SIGNIFIKANSE. 19

TABELL 5.1. ESTIMATER TIL RÅRATEN OG SMR, MED KONFIDENSGRENSER, FOR DIAGNOSTISERT DØDELIGHET I HJERTE- OG KARSYKDOMMER. BYDELER I OSLO. 28

TABELL 5.2. STRATUMSPESIFIKKE DATA SOM LIGGER BAK ESTIMERINGEN AV RÅRATEN OG SMR FOR BYDEL 01 (BYGDØY-FROGNER). 29

TABELL 5.3. ESTIMATER TIL DEN DIREKTE STANDARDISERTE INSIDENSRATEN OG SRR, MED KONFIDENSGRENSER, FOR DIAGNOSTISERT TOTAL DØDELIGHET. BYDELER I OSLO. 33

TABELL 5.4. STRATUMSPESIFIKKE DATA SOM LIGGER BAK ESTIMERINGEN AV DEN DIREKTE STANDARDISERTE INSIDENSRATEN OG SRR FOR DEN UEKSPONERTE KONTROLLBYDELEN (23 VINDEREN) OG EN EKSPONERT BYDEL (06 GAMLE OSLO) 34

Figurregister

FIGUR 5.1. TEMAKARTFREMSTILLING SOM VISER HVORDAN DET ESTIMERTE INDIREKTE STANDARDISERTE DØDELIGHETSTALLET (SMRG) FOR DØDELIGHET I HJERTE- OG KARSYKDOMMER, FRA EKSEMPEL 1, FORDELER SEG I DE FORSKJELLIGE BYDELENE I OSLO. 36

FIGUR 5.2. TEMAKARTFREMSTILLING SOM VISER HVORDAN DET ESTIMERTE DIREKTE STANDARDISERTE RISIKOFORHOLDET ($SRR_{E|U}$) FOR TOTAL DØDELIGHET, FRA EKSEMPEL 2, FORDELER SEG I DE FORSKJELLIGE BYDELENE I OSLO. 37

1 Innledning

I mange tilfeller ønsker man å kunne se på og sammenligne helsetilstanden blant befolkningen i forskjellige geografiske områder. Hensiktene kan være mange, det kan f.eks. handle om å påvise miljøgifters innvirking på folkehelsen eller å få underlag for planlegging av sykehus, utarbeiding av folkehelsekampanjer o.l. En formålstjenlig presentasjon av et statistiskt material som på en enkel og tydelig måte forteller om folkehelsetilstanden i den aktuelle populasjonen vil her få en sentral betydning.

Ved å bruke standardiserte rater vil man få en slik enkel og ryddig presentasjon. Metodene som presenteres i dette notatet innebærer at man sammenligner flere geografiske områder med hensyn på noen sykdoms- eller dødsårsaksdiagnose ved å korrigere for ulikheter i befolkningssammensetningen. Vanligvis velger man å korrigere med hensyn på alder og kjønn, ettersom antallet dødsfall eller sykdomstilfeller (i en eller annen gitt diagnose) vanligvis er sterkt relatert til disse kjennemerkene.

Man skiller mellom indirekte og direkte standardisering. I indirekte standardisering sammenligner man det studerte, «egne» geografiske området med en referansepopulasjon som f.eks. kan omfatte hele landet. Det gjør man ved å beregne SMR-forholdet, et effektmål som forteller om eventuell overdødelighet eller forhøyet sykdomsforekomst i det «egne» geografiske området, gitt at befolkningssammensetningen for referansepopulasjonen er korrigert etter det «egne» geografiske områdets befolkningssammensetning. Direkte standardisering innebærer at man i utgangspunktet lager korreksjonsvekter etter opplysninger om befolkningssammensetningen i en bestemt standardpopulasjon. Dette medfører at man direkte kan sammenligne flere studerte geografiske områder med hverandre, ettersom de derved antas å få en lik, «kunstig» befolkningssammensetning. Et vanlig effektmål i denne sammenhengen er SRR-forholdet, som forteller om hvor mye større risikoen er for å bli syk eller dø dersom man tilhører befolkningsgruppe A, sammenlignet med om man tilhører befolkningsgruppe B.

I dette notatet har vi brukt dødsårsaksdata fra bydeler i Oslo som eksempel. Formålet med dette er rent pedagogisk, slik at tallene som blir presentert i tabellene og figurene ikke representerer offisiell statistikk.

2 Noen epidemiologiske begreper

2.1 Deskriptive epidemiologiske mål

Det finnes et antall måter å beskrive forekomsten av f.eks. sykdom eller død på. Vi er mest interesserte i to slike deskriptive epidemiologiske mål, nemlig *insidenstallet* og den *kumulative insidensen*. Insidenstallet beskriver *antallet nye tilfeller* av en bestemt sykdom under løpet av en måleperiode på t år som vi benevner *risikotid*. Den kumulative insidensen beskriver på tilsvarende måte antallet, eller *andelen individer som er, blir eller har vært syke* i denne perioden. La oss betrakte de deskriptive epidemiologiske målene som stokastiske variabler[1]. Da vil det vanligvis vise seg at det førstnevnte målet er *poissonfordelt* (ettersom populasjonen som de *nye tilfellene* er en del av antas å være tilnærmet uendelig) mens det andre, den kumulative insidensen, er *binomisk* fordelt (da det i dette tilfellet handler om andelen syke *individer* av en populasjon med endelig størrelse som vi kjenner)[2].

Enda et eksempel på deskriptive epidemiologiske mål er *prevalens*, som innebærer at man måler *andelen individer* med sykdom i en populasjon på et *bestemt tidspunkt*. Prevalensen defineres også som nettoeffekten av kumulativ insidens og avgang (det vil si, at individet enten dør eller blir friskt igjen). I likhet med den kumulative insidensen kan prevalensen antas å være binomisk fordelt. Forskjellen mellom de to måtene å beskrive en befolkning på er at man ikke tar hensyn til noen risikotid når man måler prevalensen.

I det praktiske eksempelet med dødelighetsrater lengre frem i dette notatet bruker vi kumulativ insidens som deskriptivt epidemiologiskt mål. Grunnen til dette er at døden er noe som bare kan ramme en individ en gang. Antallet dødsfall kan bli satt i proporsjon til populasjonsstørrelsen og det bør derved være naturlig å anta at antallet dødsfall under måleperioden er binomisk fordelt.

2.2 Populasjoner

Vi taler om flere forskjellige typer av populasjoner i dette notatet. En *eksponert populasjon* er en befolkningsgruppe som vi er interesserte i å undersøke. Vanligvis er den eksponerte populasjonen lik befolkningen i et geografisk avgrenset område, f.eks. en kommune eller et fylke. En *ueksponert populasjon* er en gruppe av individer som vi ønsker å sammenligne med den eksponerte populasjonen ved direkte standardisering (som vi kommer tilbake til). Et konkret eksempel er følgende: det finnes god

[1] En stokastisk variabel kan anta tilfeldige verdier med en gitt sannsynlighetsfordeling.

[2] Dette er en vanlig «tommelfingerregel» som man bør behandle med forsiktighet. I mange tilfeller går det faktisk an å bruke poissonfordelingen, også til mål av kumulativ insidens (og prevalens). Dette forutsetter imidlertid at antallet syke eller døde i en diagnose er svært lite i forhold til gjennomsnittspopulasjonen. Omvendt finnes det situasjoner hvor man kan, eller bør, bruke binomisk fordeling ved estimering av insidenstall. Dette forutsetter derimot på sin side at man kjenner populasjonen som insidenstallet er en andel av (og f.eks. kan besvare spørsmålet «Hvor mange tilfeller av forkjølelse kan maksimalt bli innmeldt til trygdekontoret under vinterhalvåret 1995/96» dersom man ønsker å studere antallet påmeldte forkjølelser under denne risikotiden). Populasjonstallsparameteren i den binomiske fordelingen er i enkelte tilfeller fullt mulig å estimere. Dette vil vi imidlertid ikke gå inn på i dette notatet. Poissonfordeling og binomisk fordeling kommer vi tilbake til i kapittel 4.

grunn til å mistenke at befolkningen i Oslo/Akershus er mer predisponerte for å bli syke i lungekreft enn befolkningen i Sogn og Fjordane på grunn av luftforurensningene i storbymiljøet. Vi sier da at Osloinnbyggerne tilhører den eksponerte populasjonen mens innbyggerne i Sogn og Fjordane er en ueksponert populasjon, som vi ønsker å sammenligne Oslotallene for lungekreft med. En *referansepopulasjon*, slik vi har valgt å definere den her, er en større gruppe av individer som den eksponerte populasjonen inngår i. Totalbefolkningen i Norge er et eksempel.

Begrepene *standardpopulasjon* og *gjennomsnittspopulasjon* brukes i forbindelse med definisjon av folkemengden og standardiseringsvektene til selve justeringen av ratene. *En standardpopulasjon* er en befolkning med en sammensetning som vanligvis, men ikke nødvendigvis, er «nøytral» i forhold til befolkningsgruppene som skal studeres, og den brukes for å lage standardiseringsvekter til direkte standardisering. Et eksempel på en standardpopulasjon er folkemengden i Norge i et bestemt år, fordelt på alder og kjønn (som er standardiseringsstrata, se under). En standardpopulasjon kan også være en matematisk formel som bygger på teoretiske antakelser om sannsynlighet for overlevelse blant befolkningen. En *gjennomsnittspopulasjon* er en befolkning med en sammensetning som bygger på et gjennomsnitt av folkemengden under måle- eller risikoperioden. Vanligvis velger man å bruke folkemengdetallene for et år mitt i perioden. Gjennomsnittspopulasjonen brukes for å lage vekter til indirekte standardisering, og det er den enkelte befolkningsgruppens «egne» gjennomsnittlige folkemengdstall som vil bli aktuelle her.

2.3 Ujustert og justert rate

Et tall som forteller om graden av sykelighet, eller om dødeligheten i et samfunn, (det kan være i ujustert eller justert form) gis ofte betegnelsen *rate*. Et annet ord er *effektmål*. Avhengig av hva det er man vil fortelle finnes det flere forskjellige typer av rater. Vi vil ta opp begrepene *rårate*, *stratumspesifikk insidensrate*, *standardisert insidensrate*, *SMR* og *SRR* i kapittel 3.

I mange tilfeller ønsker man å kunne sammenligne forskjellige befolkningsgrupper med hensyn på forekomst av sykdom, dødelighet o.l. Det kan handle om å sammenligne land, fylker, kommuner eller, som i eksemplene lengre frem i dette notatet, bydeler. Ettersom land, fylker og bydeler vanligvis har forskjellig befolkningsstruktur må man lage noen form av førbearbeiding av dataene før man kan sammenligne tallene. En måte å gjøre dette på er å standardisere.

Standardisering innebærer at man lager en justering av en rate ved å beregne en veiet sum hvor man tar hensyn til befolkningsstrukturen. Man bruker da en eller flere *standardiseringsvariabler* som inneholder bakgrunnsopplysninger i gruppert form om befolkningen. Gruppene kalles *standardiseringsstrata*. Vanligvis velger man å standardisere etter aldersintervaller og kjønn. Man utfører med andre ord en slags etterstratifisering[3].

---

[3] Begrepet «confounding» (fra engelsk: forvirring, «rot») blir ofte brukt i denne sammenhengen. «Confounding» forklares av at en risikofaktor fordeler seg ulikt i forskjellige befolkningsgrupper. La oss som eksempel ta alderssammensetningen i en populasjon: Risikoen for å dø er betydelig større blant eldre personer enn blant yngre, og en vanlig aldersrelatert dødsårsak er død som følge av hjerte- og karsykdommer. Hvis vi sammenligner populasjonen i et geografiskt område hvor gjennomsnitts-

Man skiller mellom *indirekte* og *direkte standardisering*. Indirekte standardisering innebærer at man går ut fra befolkningsfordelingen til gjennomsnittspopulasjonen for den studerte (eksponerte) populasjonen. Når man så beregner SMR (som vi kommer tilbake til i avsnitt 3.2.2) så vil man finne et mål for sammenligning av den eksponerte populasjonen med en referansepopulasjon, som f.eks. kan omfatte hele landet, og som er justert etter den eksponerte populasjonens vekter. I den direkte standardiseringen sammenligner man to eller flere populasjoner med hverandre ved å justere etter en standardpopulasjons vekter.

Begrepet *intensitet* brukes av og til i dette notatet og bør derfor forklares. Intensiteten er et tall for antallet tilfeller av den studerte sykdommen (eller antallet studerte dødsfall) under *hele risikotiden*. De ubearbeidete, aggregerte rådataene fra register o.l. beskriver ofte intensitetene (dette gjelder f.eks. for dødlighetsdataene til eksemplene lengst bak i dette notatet, som har blitt hentet inn fra dødsårsaksregistrene for årene 1991-94, og som rett og slett forteller om antallet døde i en gitt diagnose under denne tidsperioden).

---

alderen er høy med populasjonen i et annet område hvor det bor en større andel unge mennesker, så vil råraten for dødelighet i hjerte- og karsykdommer sannsynligvis være større i området hvor «de eldre» bor. Ved å standardisere etter alder vil man eliminere denne «confoundingeffekten», og det vil derved bli fullt mulig å sammenligne de to populasjonene med hverandre, til tross for at aldersstrukturen er forskjellig. Da vil man kanskje komme frem til at dødeligheten i området hvor det bor større andel ungdommer i virkeligheten er høyere, og derved få grunnlag for å trekke konklusjoner om andre, utenforliggende faktorers innvirking på dødeligheten i denne populasjonen (f.eks. større eksponering av risikofaktorer som høy luftfororensing, dårlig arbeidsmiljø, stress,dårlige kostvaner og røykevaner blant befolkningen etc.).

3 Effektmåling ved indirekte og direkte standardisering

3.1 Rårate og stratumspesifikk insidensrate

Insidensraten uttrykker det forventete antallet sykdomstilfeller pr. innbygger og tidsenhet under en gitt risikotid. Tidsenheten er vanligvis år, og det er tid målt i år som vi heretter kommer til å bruke. Insidensraten for en eksponert/ueksponert populasjon g, som vi vil studere under en risikotid på t år, benevnes ofte *summarisk rate* eller *rårate* og forkortes vanligvis *CR* (etter «*crude rate*» på engelsk). Den beregnes som[4]

$$\boxed{\hat{CR}_g = \frac{\hat{m}_{g.}}{\overline{N}_{g.}t}} \tag{1}$$

hvor

$\hat{m}_{g.}$ = intensiteten i populasjon g, det vil si antallet tilfeller av en sykdom, eller antallet syke eller døde individer i denne populasjonen, under hele risikotiden på t år

og

$\overline{N}_{g.}$ = gjennomsnittspopulasjonen for populasjon g.

For et bestemt standardiseringsstratum k vil den *stratumspesifikke insidensraten* på tilsvarende måte bli beregnet som

$$\hat{h}_{gk} = \frac{\hat{m}_{gk}}{\overline{N}_{gk}t} \tag{2}$$

hvor

$\hat{m}_{gk}$ = intensiteten i standardiseringsstratum k og populasjon g

og

$\overline{N}_{gk}$ = gjennomsnittspopulasjonen for standardiseringsstratum k i populasjon g.

Vanligvis taler man om insidensratene på dette nivået som *alders- og kjønnsspesifikke*, dersom standardiseringen er blitt utført etter alder og kjønn.

[4] Størrelser med «hatt-tegn» («^») representerer estimatorer av stokastiske variabler. Det finnes med andre ord en statistisk usikkerhetsfaktor rundt disse størrelsene og de bør derved bli betraktet som estimater. Måling av den statistiske usikkerheten blir behandlet i kapittel 4.

3.2 Indirekte standardisering

3.2.1 *Noen ord om råraten i forbindelse med indirekte standardisering*

Råraten kan også beregnes som

$$\hat{CR}_g = \sum_{k=1}^{K} \frac{\overline{N}_{gk}}{\overline{N}_{g.}} \hat{h}_{gk} \tag{3}$$

(jfr. med (1))

hvor K er det totale antallet standardiseringsstrata

det vil si som summen av de stratumspesifikke insidensratene, vektet ved den eksponerte gjennomsnittspopulasjonens størrelse innenfor de ulike strataene, ettersom

$$\sum_{k=1}^{K} \frac{\overline{N}_{gk}}{\overline{N}_{g.}} \hat{h}_{gk} = \sum_{k=1}^{K} \frac{\overline{N}_{gk} \hat{m}_{gk}}{\overline{N}_{g.} \overline{N}_{gk} t} = \sum_{k=1}^{K} \frac{\hat{m}_{gk}}{\overline{N}_{g.} t} = \frac{\hat{m}_{g.}}{\overline{N}_{g.} t} = \hat{CR}_g . \tag{4}$$

Poenget med denne utledningen av formelen til råraten vil vi se når denne raten blir satt i relasjon til raten av en referansepopulasjon som er etterstratifisert ved vekting med gjennomsnittspopulasjonen til den studerte befolkningsgruppen (termen $\overline{N}_{gk} / \overline{N}_{g.}$). Denne påstanden leder oss frem til neste avsnitt, som handler om definisjon og beregning av SMR.

3.2.2 *SMR (Standardized Mortality Ratio eller Standardized Morbidity Ratio)*

Formålet med SMR er å studere sykelighet eller dødelighet i en befolkningsgruppe ved å sammenligne med tilsvarende insidenser fra en referansepopulasjon. La oss vende tilbake til råraten: CR_g er et mål for sykeligheten eller dødeligheten i en eksponert populasjon, f.eks. en bestemt kommune eller et bestemt fylke som vi ønsker å studere. Hvordan ser sykelighets- eller dødelighetstallene ut i kommunen eller fylket i forhold til situasjonen i hele landet? For å få svar på dette spørsmålet betrakter vi hele landet som vår referansepopulasjon og beregner insidensraten, $\hat{II}_{RIg}$, for denne:

$$\boxed{\hat{II}_{RIg} = \sum_{k=1}^{K} \frac{\overline{N}_{gk}}{\overline{N}_{g.}} \hat{h}_{Rk}}, \tag{5}$$

(Bemerk at vi bruker størrelsene til *den studerte befolkningsgruppens gjennomsnittspopulasjon* som standardiseringsvekter, poenget med dette er at vi ønsker konsistens mellom telleren og nevneren når vi beregner og tolker SMR (se formlene under))

hvor

$\hat{h}_{Rk}$ =den stratumspesifikke insidensraten for standardiseringsstratum k i referansepopulasjonen, det vil si

$$\hat{h}_{Rk} = \frac{\sum_{g=1}^{G} \hat{m}_{gk}}{\overline{N}_{.k} t} = \frac{\hat{m}_{.k}}{\overline{N}_{.k} t}, \tag{6}$$

hvor mengden {1,...,g,...,G} er alle de (studerte eller ikke-studerte) befolkningsgruppene som inngår i referansepopulasjonen.

$\hat{\Pi}_{Rlg}$ forteller oss altså om det forventete antallet sykdoms- eller dødsfall dersom vi hadde hatt referansepopulasjonens intensiteter i «vårt» geografiske område g.

Dersom vi dividerer råraten med insidensraten for referansepopulasjonen så vil vi få et effektmål som benevnes SMR (som er en forkortelse for Standardized Mortality Ratio *eller* Standardized Morbidity Ratio), det vil si

$$\boxed{\hat{SMR}_g = \frac{\hat{CR}_g}{\hat{\Pi}_{Rlg}}} \tag{7}$$

Formelen for estimering av SMR kan også skrives

$$\hat{SMR}_g = \frac{\sum_{k=1}^{K} \overline{N}_{gk} \hat{h}_{gk}}{\sum_{k=1}^{K} \overline{N}_{gk} \hat{h}_{Rk}}. \tag{8}$$

Verbalt defineres SMR som det faktiske antallet dødsfall (eller sykdomstilfeller) dividert på det forventete antallet dødsfall. Og det forventete antallet dødsfall gis av referansepopulasjonens stratumspesifikke insidensrater justerte ved den «egne», studerte befolkningsgruppens sammensetning (definert som gjennomsnittspopulasjonen under måleperioden, eller risikotiden).

3.3 Direkte standardisering

3.3.1 Direkte standardisert insidensrate

Som vi nevnte i avsnitt 2.3 innebærer direkte standardisering at man lager en justering med hensyn til sammensetningen av en *standardpopulasjon* i hensikt å kunne sammenligne to eller flere populasjoner med hverandre. Som standardpopulasjon kan man f.eks. bruke hele landets befolkning fordelt på standardiseringsstrataene under et bestemt referanseår.

Den *direkte standardiserte insidensraten*, DI_g, angir insidensmålet slik det ville sett ut dersom befolkningssammensetningen i populasjon g var densamme som i standardpopulasjonen. De egne insidensratene for populasjon g blir med andre ord etterjustert med hensyn på standardpopulasjonens befolkningssammensetning. Dette medfører at man direkte kan sammenligne to eller flere populasjoner i både tid og rom. Estimatet for den direkte standardiserte insidensraten beregnes som

$$\boxed{\hat{DI}_g = \frac{1}{N}\sum_{k=1}^{K} N_k \hat{h}_{gk}} \tag{9}$$

hvor

N_k = folkemengden i gruppe k i standardpopulasjonen

og

N = folkemengden i standardpopulasjonen totalt, det vil si $\sum_{k=1}^{K} N_k$.

Den direkte standardiserte insidensraten kan med fordel brukes i tabeller hvor man f.eks. vil sammenligne tall fra forskjellige fylker eller kommuner med hverandre (DI målt som dødelighet pr. 100 000 innbyggere brukes f.eks. i SSB's NOS-presentasjon av dødsårsaksstatistikken). Den kan også brukes for å sammenligne en populasjon som er eksponert for risikoen å få en bestemt sykdom med en tilsvarende ueksponert befolkningsgruppe. Dette leder oss til SRR-tallet, som vi behandler i neste avsnitt.

3.3.2 SRR (Standardized Risk Ratio)

Det standardiserte risikoforholdet, SRR, forteller om hvor stor risikoen for sykdom eller død er dersom man tilhører en befolkningsgruppe som er eksponert for en faktor som eventuelt kan ligge bak utviklingen av en diagnostisert sykdom sammenlignet med om man tilhører en ueksponert kontrollpopulasjon. Det kan f.eks. handle om å bestemme overdødeligheten i lungekreft blant dagligrøykere ved å sammenligne disse med ikkerøykere[5]. Det standardiserte risikoforholdet beregnes som

$$\boxed{\hat{SRR}_{E|U} = \frac{\hat{DI}_E}{\hat{DI}_U} = \frac{\sum_{k=1}^{K} N_k \hat{h}_{Ek}}{\sum_{k=1}^{K} N_k \hat{h}_{Uk}}} \tag{10}$$

hvor

$\hat{DI}_E$ = den direkte standardiserte insidensraten for den *eksponerte* populasjonen *E*,

[5] Analyser som går ut på at man sammenligner en eksponert og en ueksponert populasjon benevnes ofte *kasus-kontroll-studier*. Den relative risken, enten den er ustandardisert eller standardisert, benevnes ofte *risk-ratio*. Det finnes en lang rekke måter å utføre slike analyser på, og de er blant annet avhengige av om man vil se på sykdomsforekomsten i en befolkning eller utføre på forhånd planlagde, kliniske studier (av f.eks. et legemiddels effekt på et bestemt sykdomssymptom) på en forsøksgruppe (som får et virksomt legemiddel) og en kontrollgruppe (som får piller med s.k. placeboeffekt). Vi går ikke nærmere inn på dette her, den interesserte henvises isteden til referanselitteraturen (f.eks. Ahlbom(1990)).

$D\hat{I}_U$ = den direkte standardiserte insidensraten for den *ueksponerte* kontrollpopulasjonen *U*,

$\hat{h}_{Ek}$ = den stratumspesifikke insidensraten for standardiseringsstratumet *k* og den *eksponerte* populasjonen *E*,

og

$\hat{h}_{Uk}$ = den stratumspesifikke insidensraten for standardiseringsstratumet *k* og den *ueksponerte* kontrollpopulasjonen *U*[6].

[6] Indeksering med *E* og *U* (*E* for eksponert populasjon, *U* for ueksponert kontrollpopulasjon) har vi valgt for å markere at vi med den standardiserte risikoraten $SRR_{E|U}$ lager en form av kasus-kontrollstudie (se også fotnote 5) hvor vi sammenligner de begge populasjonene med hverandre.

4 Statistisk usikkerhet, konfidensintervaller og hypotesetesting

4.1 Konfidensintervaller og hypotesetesting - en kort teoretisk innføring

4.1.1 Konfidensintervaller

I våre anvendelser i dette notatet betrakter vi det oppmålte antallet tilfeller av sykdom, eller antallet individer som rammes av sykdom eller død, $m_{<indeks>}$, som en stokastisk variabel mens alt annet er å betrakte som ikke-stokastiske konstanter. Vi er interesserte i å bruke den stokastiske variabelen for å beregne konfidensintervaller og for å lage hypotesetesting.

Et konfidensintervall forteller hvilke verdier en stokastisk variabel kan anta med en sannsynlighet på (1-α)*100 %. Dersom vi har estimert tallet $\hat{\mu}$ for parameteren μ, så vil den nedre konfidensgrensen få betegnelsen $x_{(\alpha/2)}$ mens den øvre konfidensgrensen benevnes $x_{(1-\alpha/2)}$. Intervallet kan da skrives som

$$x_{(\alpha/2)} \le \mu \le x_{(1-\alpha/2)}. \tag{11}$$

La, som vi allerede har nevnt, $\hat{\mu}$ være en estimator for μ. La oss også anta at $\hat{\mu}$ har en fordeling f eller g. Vi bestemmer da konfidensgrensene $x_{(\alpha/2)}$ og $x_{(1-\alpha/2)}$, slik at

$$\Pr(\hat{\mu} \le x_{(\alpha/2)}) = \sum_{j=x_0}^{x_{(\alpha/2)}} f(j) = \alpha/2 \tag{12}$$

og

$$\Pr(\hat{\mu} \le x_{(1-\alpha/2)}) = \sum_{j=x_0}^{x_{(1-\alpha/2)}} f(j) = (1-\alpha/2), \tag{13}$$

> *hvor*
> *f(j)* er en punktsannsynlighet,
> *og*
> x_0 er den laveste verdien som den stokastiske variabelen kan anta (er lik 0 dersom det handler om en variabel som er binomisk fordelt eller poissonfordelt) ,

eller

$$\Pr(\hat{\mu} \le x_{(\alpha/2)}) = \int_{x_0}^{x_{(\alpha/2)}} g(x)dx = \alpha/2 \tag{14}$$

og

$$\Pr(\hat{\mu} \le x_{(1-\alpha/2)}) = \int_{x_0}^{x_{(1-\alpha/2)}} g(x)dx = (1-\alpha/2)\,, \tag{15}$$

hvor
g(x) er en sannsynlighetstetthet,
og
x_0 er den laveste verdien som den stokastiske variabelen kan anta (er lik -∞ dersom den er normalfordelt).

4.1.2 Ensidige konfidensintervaller

En spesiell form for konfidensintervaller, s.k. ensidige konfidensintervaller, brukes for å teste om et estimat enten er større enn eller mindre enn en gitt verdi. Slik ensidig hypotesetesting går vi nærmere inn på i neste avsnitt. Det ensidige konfidensintervallet kan sies å være «åpent» på den ene siden og «lukket» på den andre, mens konvensjonelle 2-sidige konfidensintervaller er «åpne» på begge sider. Det ensidige konfidensintervallet kan skrives som

$$\mu \le x_{(1-\alpha)}\,, \tag{16}$$

og kan da kalles et ensidig konfidensintervall som er «åpent på høyre side». Vi har en konfidensgrense, og den vil vi få ved å bestemme størrelsen på $x_{(1-\alpha)}$, slik at

$$\Pr(\hat{\mu} \le x_{(1-\alpha)}) = \sum_{j=x_0}^{x_{(1-\alpha)}} f(j) = (1-\alpha) \tag{17}$$

dersom $\hat{\mu}$ er en estimator for μ med en diskret sannsynlighetsfordeling, eller

$$\Pr(\hat{\mu} \le x_{(1-\alpha)}) = \int_{x_0}^{x_{(1-\alpha)}} g(x)dx = (1-\alpha) \tag{18}$$

dersom $\hat{\mu}$ er kontinuerlig fordelt.

Tilsvarende intervall som er «åpent på venstre side» kan skrives som

$$\mu \ge x_{(\alpha)} \tag{19}$$

og konfidensgrensen for hhv. diskret og kontinuerlig fordeling bestemmes ved

$$\Pr(\hat{\mu} \le x_{(\alpha)}) = \sum_{j=x_0}^{x_{(\alpha)}} f(j) = \alpha \tag{20}$$

og

$$\Pr(\hat{\mu} \le x_{(\alpha)}) = \int_{x_0}^{x_{(\alpha)}} g(x)dx = \alpha \,. \tag{21}$$

En viktig forskjell mellom de 2-sidige konfidensintervallene og de ensidige konfidensintervallene er at de førstnevnte kan illustreres som sannsynlighetsfordelinger med to «haler» med sannsynlighet $\alpha/2$ (2,5 % ligger «utenfor» på hver side, dersom det handler om 95 % konfidens), mens de ensidige konfidensintervallene har en «hale» med sannsynlighet α (det vil si 5 % ligger utenfor på bare den ene siden under 95 % konfidens).

4.1.3 Hypotesetesting

Formålet med hypotesetesting er å gi informasjon som ligger til grunn for konklusjoner om tall som foreligger for presentasjon. Nullhypotesen, som vi betegner med H_0 og som er en antagelse om at det ikke foreligger noe avvik fra en gitt verdi, bør forkastes dersom sannsynligheten for at mothypotesen H_1 er riktig er stor, f.eks. 95 % eller mer. Da sier man at H_1 er signifikant på 5 %-nivået.[7] Hypotesetestingen foregår ved hjelp av en s.k. testfunksjon. Denne brukar man for å beregne sannsynligheten for at H_0 er riktig, den s.k. *p-verdien* (fra ordet *prob-value* på engelsk).[8]

Vanligvis er det tre forskjellige modeller for hypotesetesting som vil bli aktuelle. Modell (I) tar utgangspunkt i om det foreligger noe signifikant avvik fra en bestemt normgivende nullhypoteseverdi, f.eks. insidensen lik 0 eller SMR lik 1. Modell (II) og (III) har samme utgangspunkt, men her tester vi om estimatet henholdsvis er *større enn* eller *mindre enn* nullhypoteseverdien. Til dette formålet bruker vi et ensidig konfidensintervall. **Tabell 4.1** gir en oversikt over hypoteseformuleringene for de tre modellene, testfunksjonene for beregning av p-verdien samt en verbal forklaring av hvordan hypotesene skal tolkes.

[7] La oss i fortsettelsen bruke størrelsen α for testens signifikansnivå, slik at α=0,05 ved tester på 5%-nivået (s.k. enstjernerstest), α=0,01 ved tester på 1%-nivået (to-stjernerstest) eller α=0,001 ved 0,1% signifikanstest (tre-stjernerstest).

[8] P-verdien må ikke bli forvekslet med verdien for signifikansnivået, α. P-verdien er en verdi som blir beregnet ved hjelp av testfunksjonen mens α er et på forhånd gitt kriterium for akseptanse eller forkastelse av nullhypotesen.

Tabell 4.1. Oversikt over formulering av null- og mothypotese for tre hypotesetestmodeller.

Modell:	Hypoteseformulering:	Verbal tolking:	Testfunksjon:
I	**H₀**: $\mu = \mu_0$ **H₁**: $\mu \neq \mu_0$ (2-sidig test)	**H₀**: Man kan ikke påvise at estimatet avviker fra nullhypoteseverdien μ_0. **H₁**: Man kan påvise et slikt avvik.	$p = \begin{cases} Pr(\hat{\mu} \geq \mu_0) \\ Pr(\hat{\mu} \leq \mu_0) \end{cases}$ Forkast H_0 dersom $p<\alpha/2$.
II	**H₀**: $\mu = \mu_0$ **H₁**: $\mu > \mu_0$ (ensidig test)	**H₀**: Man kan ikke påvise at estimatet er større enn nullhypoteseverdien. **H₁**: Man kan påvise at estimatet er større.	$p = Pr(\hat{\mu} \geq \mu_0)$ Forkast H_0 dersom $p<\alpha$.
III	**H₀**: $\mu = \mu_0$ **H₁**: $\mu < \mu_0$ (ensidig test)	**H₀**: Man kan ikke påvise at estimatet er mindre enn nullhypoteseverdien. **H₁**: Man kan påvise at estimatet er mindre.	$p = Pr(\hat{\mu} \leq \mu_0)$ Forkast H_0 dersom $p<\alpha$.

4.2 Poissonfordeling, binomisk fordeling og normalfordelingstilnærming

Som vi allerede har nevnt i kapittel 2 er det to typer av sannsynlighetsfordelinger som er sentrale i forbindelse med estimering av de epidemiologiske målene insidens, kumulativ insidens og prevalens, nemlig poissonfordeling og binomisk fordeling.

Poissonfordelingen er en diskret sannsynlighetsfordeling med en parameter, som faktisk uttrykker både forventet verdi og varians. Vi lar den «sanne» intensiteten m_{gk} betegne denne parameteren. En poissonfordelt variabel kan inneholde alle positive heltall $\{0,1,2,3,....\}$. Det finnes en rekke eksempler på stokastiske variabler som er tilnærmet poissonfordelte, blant annet antallet personer som står i køen på posten eller antallet flyulykker i løpet av et år. I de aller fleste tilfellene er insidensene, som deskriptive epidemiologiske måleenheter betraktet, også tilnærmet poissonfordelt. Man bør være oppmerksom på at vi bruker ordet «tilnærmet» ettersom poissonfordelingen forutsetter at populasjonen er uendelig stor (i det virkelige livet er dette en umulighet, men teoretisk vil antakelsen om en tilnærmet uendelig populasjon by på en del praktiske fordeler).

Den binomiske fordelingen er en sannsynlighetsfordeling med to parametrer, som vi kan kalle π og $\overline{N}_{gk}$. Vi lar π være lik $(m_{gk}/\overline{N}_{gk})$. En binomisk fordelt variabel kan inneholde alle heltall i intervallet $\{0,1,2,...,\overline{N}_{gk}\}$. I situasjoner hvor vi har en endelig populasjon vil sannsynlighetsfunksjonen til effektmålet være binomisk fordelt. Dette vil være aktuelt når vi måler kumulativ insidens og prevalens.

Vi har i dette notatet valgt å begrense oss til å bruke normalfordelingstilnærming når vi beregner konfidensgrenser og lager hypotesetesting. Dette skyldes at rutinene for den «eksakte» bestemmelsen av konfidensintervaller rundt de estimerte ratene oftest er unødvendig kompliserte. Dersom befolkningsgruppene som skal måles ikke er altfor små bør en slik approksimering ikke være noe problem:

Ifølge det *sentrale grenseverditeoremet* («*Central Limit Theorem*», se Hogg & Tanis(1977)) finnes det en «tommelfingerregel» som sier at det går an å bruke normalfordelingstilnærming dersom intensitetstallet, $\hat{m}_{gk}$, ikke er mindre enn 5.

For et materiale som er *poissonfordelt* med parameteren m_{gk} gjelder, at den forventete verdien[9]

$$E(\hat{m}_{gk}) = m_{gk} \tag{22}$$

og variansen

$$v(\hat{m}_{gk}) = m_{gk} \tag{23}$$

(det vil si, variansen er lik den forventete verdien).

For et materiale som er *binomisk fordelt* med parametrene π, m_{gk} og $\overline{N}_{gk}$ gjelder på tilsvarende måte, at den forventete verdien

$$E(\hat{m}_{gk}) = \pi\overline{N}_{gk} = m_{gk} \tag{24}$$

mens variansen

$$v(\hat{m}_{gk}) = \overline{N}_{gk}\pi(1-\pi) = m_{gk}(1-\frac{m_{gk}}{\overline{N}_{gk}}) \tag{25}$$

For parameteren m_{gk} beregnes så et 95 % konfidensintervall lik

$$\boxed{\hat{m}_{gk} - 1{,}96\sqrt{v(\hat{m}_{gk})} \le m_{gk} \le \hat{m}_{gk} + 1{,}96\sqrt{v(\hat{m}_{gk})}}, \tag{26}$$

hvor 1,96 er verdien for en normalfordelt variabel z, med forventet verdi lik 0 og varians lik 1, som svarer mot $1-\alpha/2 = 0{,}975$.

[9] Vi bruker her de stratumspesifikke størrelsene som parametrer og estimatorer. Det går selvsagt også an å bruke tilsvarende ustandardiserte størrelser i forbindelse med beregning av ustandardiserte rater. Når det gjelder de standardiserte ratene så betrakter vi disse, for enkelhetens skull, på samme måte som de ustandardiserte ratene i variansberegningene i avsnittene 4.3.3 og 4.3.4. La oss derved si at «tommelfingerregelen» som sier at intensitetstallet må være større enn eller lik 5 også gjelder når vi indekserer dette tallet som $\hat{m}_{g.}$.

4.3 Variansberegning av ustandardiserte og standardiserte rater

4.3.1 Noen ord om hypotesetesting av rater under normalfordelingstilnærming

Vi har allerede vært inne på normalfordelingstilnærmet beregning av konfidensintervaller for intensiteter.

To nøkkelbegreper i forbindelse med beregning av konfidensintervaller og hypotesetesting, når vi ønsker å normalfordelingstilnærme, er estimatene av den forventete verdien og variansen, $\hat{\mu}$ og $\hat{v}(\hat{\mu})$. Dersom vi kjenner disse to estimatene er det i stort sett bare å plassere inn dem i en modell og utføre hypotesetesten. **Tabell 4.2**, som kan sammenlignes direkte med **tabell 4.1** fra avsnitt 4.1, forteller oss om hypoteseformulering, testobservator og signifikansnivåer for akseptanse/forkastelse av hhv. en-, to- og trestjerners signifikanstest.

I de følgende avsnittene skal vi se på hvordan man estimerer variansene til de effektmålene som vi allerede har behandlet i kapittel 3.

Tabell 4.2. Modeller for hypotesetesting under normalfordeling, hvor vi har en stokastisk variabel x som er N(μ,v(μ)) (normalfordelt med en parameter som har en forventet verdi μ og en varians v(μ)). Testobservatorer og beslutningsgrunnlag for en-, to- og trestjerners signifikanse.

Modell:	**Hypoteseformulering:**	**Test-observator:** *(formel nr. 27)*	**Forkast H_0 på 5 %-nivået dersom**	**Forkast H_0 på 1 %-nivået dersom**	**Forkast H_0 på 0,1 %-nivået dersom**
I	**H_0**: $\mu = \mu_0$ **H_1**: $\mu \neq \mu_0$	$z_p = \frac{\hat{\mu} - \mu_0}{\sqrt{\hat{v}(\hat{\mu})}}$	$\lvert z_p \rvert > 1{,}96$	$\lvert z_p \rvert > 2{,}576$	$\lvert z_p \rvert > 3{,}29$
II	**H_0**: $\mu = \mu_0$ **H_1**: $\mu > \mu_0$	$z_p = \frac{\hat{\mu} - \mu_0}{\sqrt{\hat{v}(\hat{\mu})}}$	$z_p > 1{,}645$	$z_p > 2{,}326$	$z_p > 3{,}09$
III	**H_0**: $\mu = \mu_0$ **H_1**: $\mu < \mu_0$	$z_p = \frac{\hat{\mu} - \mu_0}{\sqrt{\hat{v}(\hat{\mu})}}$	$z_p < -1{,}645$	$z_p < -2{,}326$	$z_p < -3{,}09$

4.3.2 Stratumspesifikk insidensrate

Variansen til poissonfordelte og binomisk fordelte stokastiske variabler under normalfordelingstilnærming er allerede blitt overfladisk behandlet i avsnitt 4.2. Fra det utgangspunktet vil vi finne at den estimerte variansen til den stratumspesifikke insidensraten er

$$\boxed{\hat{v}(\hat{h}_{gk}) = \frac{\hat{m}_{gk}}{\overline{N}_{gk}^{\,2} t^2}} \qquad (28)$$

(jfr. med (23))

dersom insidensraten er beregnet med grunnlag fra et poissonfordelt intensitetstall[10], mens tilsvarende varians for den binomisk fordelte kumulative insidensen vil bli

$$\boxed{\hat{v}(\hat{h}_{gk}) = \frac{\frac{\hat{m}_{gk}}{\overline{N}_{gk}}(1-\frac{\hat{m}_{gk}}{\overline{N}_{gk}})}{\overline{N}_{gk}t^2}} \tag{29}$$

(jfr. med (25))

4.3.3 Rårate og SMR

Råraten har blitt beregnet på samme måte som den stratumspesifikke insidensraten. Dette fremgår tydelig i avsnitt 3.1. Forskjellen er at råraten har blitt beregnet for hele populasjon g mens den stratumspesifikke insidensraten bare representerer et utsnitt av denne populasjonen, nemlig de observasjonene som tilhører stratum k.

Med dette utgangspunktet kan vi, på samme måte som i forrige avsnittet, approksimativt beregne variansen til $\hat{CR}_g$ som

$$\boxed{\hat{v}(\hat{CR}_g) = \frac{\overline{N}_{g.}\hat{CR}_g t}{\overline{N}_{g.}{}^2 t^2} = \frac{\hat{CR}_g}{\overline{N}_{g.}t}} \tag{30}$$

(jfr. med (28))

dersom det dreier som poissonfordelt materiale, og som

$$\boxed{\hat{v}(\hat{CR}_g) = \frac{\hat{\Pi}_g t(1-\hat{CR}_g t)}{\overline{N}_{g.}t^2}} \tag{31}$$

(jfr. med (29))

dersom materialet antas å være binomisk fordelt. Vi bør i sammenhengen nevne at påstanden at

$$\frac{m_{g.}}{\overline{N}_{g.}} = E(CR_g t) \tag{32}$$

er blitt benyttet ved utledningen av formlene.

---

[10] Bevis: Ettersom $\hat{v}(\hat{m}_{gk}) = \hat{m}_{gk}$ under poissonfordeling, og $\hat{h}_{gk} = \frac{\hat{m}_{gk}}{\overline{N}_{gk}t}$, så må

$$\hat{v}(\hat{h}_{gk}) = \hat{v}(\frac{\hat{m}_{gk}}{\overline{N}_{gk}t}) = \frac{1}{\overline{N}^2{}_{gk}t^2}\hat{v}(\hat{m}_{gk}) = \frac{\hat{m}_{gk}}{\overline{N}^2{}_{gk}t^2}.$$

Når vi beregner variansen til SMR, som vi skrev om i avsnitt 3.3.2, så kan vi, for våre behov, se bort fra variansen til $\hat{II}_{RIg}$ i nevneren. Denne beskriver nemlig en referansepopulasjon som i de aller fleste tilfellene er stor (referansepopulasjonen vil vanligvis omfatte f.eks. hele landet, hele fylket etc.) og nærmest kan betraktes som en konstant. Den estimerte variansen til $\hat{SMR}_g$ vil derved bli

$$\boxed{\hat{v}(\hat{SMR}_g) = \frac{\hat{v}(\hat{CR}_g)}{\hat{II}_{RIg}^{\ 2}}} \tag{33}$$

4.3.4 Direkte standardiserte rater

Variansen til den direkte standardiserte insidensraten bør kunne bli approksimert ved at man i prinsippet bruker samme opplegg som ved beregning av variansen til råraten. Det er en konservativ tilnærming men den byr blant annet på den fordelen at vi unngår problemet med skjevheter i estimatene som skyldes små strata, noe vi skulle fått dersom vi brukte et eller annet opplegg for beregning av stratumvariansen. La oss anta at avvikene til forventningen til DI_g og avvikene til forventningen til CR_g har samme fordeling. Da bør det være forsvarlig å approksimativt beregne den estimerte variansen til $\hat{DI}_g$ ved å bruke

$$\boxed{\hat{v}(\hat{DI}_g) = \frac{\overline{N}_{g.}\hat{DI}_g t}{\overline{N}_{g.}^{\ 2} t^2} = \frac{\hat{DI}_g}{\overline{N}_{g.} t}}$$

(34) *(jfr. med (30))*

dersom det dreier seg om et poissonfordelt materiale, og

$$\boxed{\hat{v}(\hat{DI}_g) = \frac{\hat{DI}_g t(1-\hat{DI}_g t)}{\overline{N}_{g.} t^2}}$$

(35) *(jfr. med (31))*

dersom materialet antas å være binomisk fordelt.

Den estimerte variansen til det estimerte standardiserte risikoforholdet $\hat{SRR}_{E|U}$ kan approksimeres[11] til

[11] Til approksimeringen har vi brukt følgende teorem: dersom x_1 og x_2 er to stokastiske variabler og y er en funksjon av disse ($y=f(x_1, x_2)$), så gjelder, at

$$\mathrm{var}(\hat{y}) \approx \left(\frac{\partial \hat{y}}{\partial \hat{x}_1}\right)^2 \mathrm{var}(\hat{x}_1) + 2\left(\frac{\partial \hat{y}}{\partial \hat{x}_1}\right)\left(\frac{\partial \hat{y}}{\partial \hat{x}_2}\right) \mathrm{cov}(\hat{x}_1, \hat{x}_2) + \left(\frac{\partial \hat{y}}{\partial \hat{x}_2}\right)^2 \mathrm{var}(\hat{x}_2)$$

Ved å betrakte $SRR_{E|U}$ som en funksjon av DI_E og DI_U, og dessuten anta at kovariansen er lik null (det vil si, uavhengighet rår), har vi anslått variansen til

$$\boxed{\hat{v}(S\hat{R}R_{E|U}) = \frac{D\hat{I}_U{}^2\hat{v}(D\hat{I}_E) + D\hat{I}_E{}^2\hat{v}(D\hat{I}_U)}{D\hat{I}_U{}^4}}. \tag{36}$$

hvor

$\hat{v}(D\hat{I}_E) =$ den estimerte variansen til den den direkte standardiserte insidensraten for den/de eksponerte populasjonen(/e) (beregnet ved formel (34) eller (35))

og

$\hat{v}(D\hat{I}_U) =$ den tilsvarende estimerte variansen til den den direkte standardiserte insidensraten for den ueksponerte kontrollpopulasjonenen (også den beregnet ved formel (34) eller (35)).

$$\mathrm{var}(S\hat{R}R_{E|U}) \approx \left(\frac{1}{\partial D\hat{I}_E}\partial\left(\frac{D\hat{I}_E}{D\hat{I}_U}\right)\right)^2 \mathrm{var}(D\hat{I}_E) + \left(\frac{1}{\partial D\hat{I}_U}\partial\left(\frac{D\hat{I}_E}{D\hat{I}_U}\right)\right)^2 \mathrm{var}(D\hat{I}_U) =$$

$$= \frac{1}{D\hat{I}_U{}^2}\mathrm{var}(D\hat{I}_E) + \frac{D\hat{I}_E{}^2}{D\hat{I}_U{}^4}\mathrm{var}(D\hat{I}_U) = \frac{D\hat{I}_U{}^2 v(D\hat{I}_E) + D\hat{I}_E{}^2 v(D\hat{I}_U)}{D\hat{I}_U{}^4}$$

5 Eksempler fra dødsårsaksstatistikken

5.1 Innledning

Vi har valgt å se på dødelighetsstatistikk for Oslo bydeler. Begrunnelsen for dette er at ulikhetene i levekår i de forskjellige bydelene i Oslo er vel kjent, og at dette materialet derved bør være vel egnet for en pedagogisk verdifull fremstilling av konkrete eksempler på bruk av standardiserte rater.

To forskjellige typer av spørsmål vil bli besvart ved hjelp av presentasjon av estimatene til hhv. SMR og SRR. I avsnitt 5.2 ser vi på dødelighetsmønsteret i Oslo bydeler i forhold til hele kommunen. Vi har her valgt dødelighet i hjerte- og karsykdommer (ifølge tresiffer ICD-kode 390-459) som kjennetegn og vi bruker indirekte standardisering for å se på dødeligheten i de enkelte bydelene sammenlignet med kommunen som helhet. I avsnitt 5.3, hvor vi bruker direkte standardisering, går vi ut fra den økonomisk og sosialt best stilte bydelen på Oslo vestkant, Vinderen, som vi sammenligner med de andre bydelene i kommunen med hensyn på den totale dødeligheten (ICD-kode 000-999). «Hvor mye større risiko for å dø under løpet av et år vil en gjennomsnittlig innbyggere i f.eks. Gamle Oslo løpe dersom man sammenligner med en person som bor på Vinderen?», er et konkret spørsmål som avsnitt 5.3 gir (et kraftig forenklet) svar på.

Som epidemiologisk mål betraktet ser vi på dødelighetsstatistikken som et eksempel på kumulativ insidens. Vi går derved ut fra antakelsen at binomisk fordeling foreligger i de begge eksemplene som vi presenterer i de følgende avsnittene.

5.2 Eksempel 1: Indirekte standardisering av dødelighet i hjerte- og karsykdommer i Oslo bydeler

I eksempel 1 ser vi på dødeligheten i hjerte- og karsykdommer i Oslo bydeler i perioden 1991-94. Hele Oslo kommune er vår referansepopulasjon. Vi ønsker å se hvor stor dødeligheten i de enkelte bydelene er i forhold til hele Oslo kommune, gitt at befolkningssammensetningen i hele kommunen er den samme som den var i den enkelte studerte bydelen i midten av måleperioden. Derved kan vi oppdage eventuell overdødelighet i noen bydeler. Til den indirekte standardiseringen har vi til å begynne med følgende opplysninger:

1. Vi kjenner antallet dødsfall for årene 1991-94 (risikotiden er altså fire år) på bydels- og stratumnivå
2. Vi kjenner gjennomsnittspopulasjonens størrelse på bydels- og stratumnivå, den er definert som folkemengden pr. 1.1.1993

3. Vi bruker en inndeling i 18 aldersgrupper (stratum 1-17 bygger på 5-årsintervaller fra 0 til 85 år, stratum 18 er personer som er 85 år og eldre) og kjønn til stratuminndelingen som danner grunnlag for standardiseringene[12]

Vi har valgt å vise to forskjellige effektmål. De er

- det forventete antallet dødsfall pr. år i perioden, det vil si den estimerte råraten $\hat{CR}_g$
- et effektmål som forteller om dødeligheten i enkelte bydeler i forhold til Oslo totalt, det vil si $\hat{SMR}_g$ (hvor Oslo totalt er en referansepopulasjon)

I tillegg har vi valgt å presentere den statistiske usikkerheten til $\hat{CR}_g$ og $\hat{SMR}_g$ i form av 95-prosents konfidensintervaller.

Resultatene fremgår av **tabell 5.1**. Kolonne (2) viser de ujusterte intensitetene, det vil si antallet døde i den enkelte bydelen under hele fireårsperioden. I kolonne (3) presenteres gjennomsnittspopulasjonen pr. 1.1.1993. Råraten, og grensene for et 95 prosents konfidensintervall til denne, gis av kolonnene (4), (5) og (6). SMR presenteres i kolonne (7), mens hhv. kolonne (8) og (10) gir den nedre og øvre grensen for et 95 prosents konfidensintervall til denne rate. Kolonne (9) er tenkt å gi (den nedre) grensen for et 95 prosents ensidig konfidensintervall, og hensikten er her å teste hypotesen om dødelighetsratene som er større enn 1 kan bli betraktet som signifikante på 5 %-nivået (en-stjerners ensidig test). Vi vil se at dette er tilfelle for åtte bydeler.

Hvis vi ser på de ustandardiserte råratene (kolonne (4)) til **tabell 5.1** vil vi finne at Søndre Nordstrand har lavest dødelighet. Ifølge våre beregnete SMR-tall i kolonne (7) vil imidlertid tre andre bydeler, nemlig Vinderen, Bygdøy-Frogner og Røa, komme «bedre» ut. Årsaken til dette er at det i Søndre Nordstrand bor forholdsvis mange unge mennesker. Nå bør man ikke sammenligne bydelene direkte med hverandre ettersom vi har brukt indirekte standardisering. Det er den enkelte bydelens «egne» befolkningssammensetning som ligger til grunn for beregningen av SMR. Man bør isteden sammenligne Søndre Nordstrand med hele Oslo kommune, og her kan vi altså trekke konklusjonen at standardiseringen vil moderere dødelighetstallene noe for denne bydelen i forhold til kommunen som helhet. Dette gir oss anledning til å si at befolkningssammensetningen *vil* påvirke tallene i betydelig grad, og at alders- og kjønnsstandardisering derved vil ha stor betydning for rett tolking av resultatene.

Tabell 5.2 gjengir de stratumspesifikke dataene for en bestemt bydel, 01 Bygdøy-Frogner. Disse dataene ligger i vårt eksempel bak beregningene av råraten og SMR. Her er det kolonne (6) og (10) som ligger i sentrum. Kolonne (6), som gjengir uttrykket $\frac{\overline{N}_{gk}}{\overline{N}_{g.}}\hat{h}_{gk}$, brukes for å beregne

---

[12] Ved å kontrollere gjennomsnittspopulasjonsverdiene for Oslo bydeler i 1993 (og standardpopulasjonsverdiene i 1981 som vi bruker i eksempelet med direkte standardisering i neste kapittel) har vi funnet at vi ved å bruke inndelingen i 2*18 strata vil unngå å få nullverdier i nevneren når vi beregner insidensratene, samtidig som stratuminndelingen bør være tilstrekkelig detaljert for å gi et godt standardiseringsgrunnlag.

$$\hat{CR}_g = \sum_{k=1}^{K} \frac{\overline{N}_{gk}}{\overline{N}_{g.}} \hat{h}_{gk} = \sum_{k=1}^{K} \frac{(\text{kol.}3)}{\sum(\text{kol.}3)} (\text{kol.}5) =$$
$$= \frac{1.083}{19.115} 0{,}00023 + \frac{699}{19.115} 0{,}00107 + \ldots + \frac{493}{19115} 0{,}04919 = 0{,}0048653$$

(37) *(tilpassning av (3))*

som blir presentert i oppblåst form i kolonne (4) i **tabell 5.1** (0,0048653*100.000=486,53). Kolonne (10) gir tilsvarende uttrykk $\frac{\overline{N}_{gk}}{\overline{N}_{g.}} \hat{h}_{Rk}$, slik at

$$\hat{\Pi}_{Rlg} = \sum_{k=1}^{K} \frac{\overline{N}_{gk}}{\overline{N}_{g.}} \hat{h}_{Rk} = \sum_{k=1}^{K} \frac{(\text{kol.}3)}{\sum(\text{kol.}3)} (\text{kol.}9) =$$
$$= \frac{451}{19.115} 0{,}00009 + \frac{319}{19.115} 0{,}00002 + \ldots + \frac{493}{19115} 0{,}07741 = 0{,}0066737.$$

(38) *(tilpassning av (5))*

SMR for bydel 01 (kolonne 7 i **tabell 5.1**) vil deretter bli beregnet ved

$$\hat{SMR}_g = \frac{\hat{CR}_g}{\hat{\Pi}_{Rlg}} = \frac{0{,}0048653}{0{,}0066737} \approx 0{,}73.$$

(39) *(tilpassning av (7))*

Den stratumspesifikke insidensraten for bydel 01 (kolonne 5 i **tabell 5.2**) beregnes ved

$$\hat{h}_{gk} = \frac{\hat{m}_{gk}}{\overline{N}_{gk} t} = \frac{(\text{kol.}4)}{(\text{kol.}3) * 4}.$$

(40) *(jfr. med (2))*

For kvinner i aldersgruppen 85 år og eldre vil vi altså få

$$\frac{(\text{kol.}4)}{(\text{kol.}3) * 4} = \frac{97}{493 * 4} \approx 0{,}04919.$$

(41) *(tilpassning av (40))*

Insidensraten for referansepopulasjonen (i dette tilfellet hele Oslo kommune, se kolonne 9 i **tabell 5.2**) beregnes på tilsvarende måte ved

$$\hat{h}_{Rk} = \frac{\sum_{g=1}^{G} \hat{m}_{gk}}{\overline{N}_{.k} t} = \frac{(\text{kol.}8)}{(\text{kol.}7) * 4} = \frac{2.461}{7.948 * 4} \approx 0{,}07741 \tag{42}$$

(tilpassning av (6))

dersom vi holder oss til standardiseringsstratum «KV:18» (kvinner i alderen 85 år og eldre).

La oss vende tilbake til **tabell 5.1** og konfidensgrensene. Disse er blitt beregnet etter følgende oppskrift:

For konfidensgrensene $CR_{g(2,5\%)}$ og $CR_{g(97,5\%)}$ (kolonnene (5) og (6)) trenger vi variansen til råraten. Den har vi beregnet som

$$\hat{v}(\hat{CR}_g) = \frac{\hat{CR}_g t(1 - \hat{CR}_g t)}{\overline{N}_{g.} t^2} = \frac{0{,}0048653 * 4(1 - 0{,}0048653 * 4)}{19.115 * 16} = \\ = 0{,}00000006239. \tag{43}$$

(tilpassning av (31))

Den nedre konfidensgrensen har vi så beregnet som

$$CR_{g(2,5\%)} = \hat{CR}_g - 1{,}96\sqrt{\hat{v}(\hat{CR}_g)} = 0{,}0048653 - 1{,}96\sqrt{0{,}00000006239} = \\ = 0{,}0043757. \tag{44}$$

(jfr. med (26))

Ved oppblåsing med 100.000 er den blitt presentert som 437,57 i tabellen. For den øvre grensen gjelder tilsvarende

$$CR_{g(97,5\%)} = \hat{CR}_g + 1{,}96\sqrt{\hat{v}(\hat{CR}_g)} = 0{,}0048653 + 1{,}96\sqrt{0{,}00000006239} = \\ = 0{,}00535{,}49 \\ \Rightarrow 0{,}00535{,}49 * 100.000 = 535{,}49. \tag{45}$$

For SMR-tallet gjelder i prinsippet samme sak. Variansen til dette,

$$\hat{v}(\hat{SMR})_g = \frac{\hat{v}(\hat{CR}_g)}{\hat{\Pi}_{Rlg}^{\ 2}} = \frac{0{,}00000006239}{0{,}0066737^2} \approx 0{,}0014008. \tag{46}$$

(tilpassning av (33))

Konfidensgrensene (kolonnene (8), (9) og (10) i **tabell 5.1**) beregnes hhv. som

$$SMR_{g(2,5\%)} = \hat{SMR}_g - 1{,}96\sqrt{\hat{v}(\hat{SMR}_g)} = 0{,}73 - 1{,}96\sqrt{0{,}0014008} \approx 0{,}66, \tag{47}$$

$$SMR_{g(5\%)} = S\hat{M}R_g - 1{,}645\sqrt{\hat{v}(S\hat{M}R_g)} = 0{,}73 - 1{,}645\sqrt{0{,}0014008} \approx 0{,}67 \tag{48}$$

og

$$SMR_{g(97,5\%)} = S\hat{M}R_g + 1{,}96\sqrt{\hat{v}(S\hat{M}R_g)} = 0{,}73 + 1{,}96\sqrt{0{,}0014008} \approx 0{,}80. \tag{49}$$

Konfidensgrensene forteller oss at vi skal forkaste en nullhypotese med en-stjerners signifikans dersom den sanne verdien ligger utenfor disse grensene. La oss nå «snu» på spørsmålet og bruke testobservatoren i **tabell 4.2** (modell I) for hypotesetesting av påstanden at det estimerte SMR-tallet for Bygdøy-Frogner er forskjellig fra 1. Vi vil da få

$$|z_p| = \left| \frac{S\hat{M}R_g - 1}{\sqrt{\hat{v}(S\hat{M}R_g)}} \right| = \left| \frac{0{,}73 - 1}{\sqrt{0{,}0014008}} \right| = |-7{,}21| = 7{,}21.$$

(50) *(tilpassning av (27))*

Ettersom $|z_p|$ er langt større enn 3,29 så kan vi påvise at det sanne SMR-tallet bør være klart mindre enn 1 med trestjerners signifikans.

Tabell 5.1. Estimater til råraten og SMR, med konfidensgrenser, for diagnostisert dødelighet i hjerte- og karsykdommer. Bydeler i Oslo.

Bydel	$\hat{m}_{g.}$	$\overline{N}_{g.}$	$\hat{CR}_g$	$CR_{g(2,5\%)}$	$CR_{g(97,5\%)}$	$\hat{SMR}_g$	$SMR_{g(2,5\%)}$	$SMR_{g(5\%)}$	$SMR_{g(97,5\%)}$
(kol.1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
01	372	19.115	**486,53**	437,57	535,49	**0,73**	0,66	0,67	0,80
02	634	22.299	**710,79**	656,26	765,33	**0,92**	0,85	0,86	0,99
03	712	26.421	**673,71**	624,89	722,52	**0,96**	0,89	0,90	1,03
04	949	26.515	**894,78**	838,88	950,68	**1,13**	1,06	1,07	1,20
05	595	23.495	**633,11**	582,89	683,34	**1,16**	1,07	1,09	1,25
06	455	20.628	**551,44**	501,33	601,54	**1,22**	1,11	1,13	1,33
07	335	14.966	**559,60**	500,35	618,85	**1,02**	0,92	0,93	1,13
08	295	16.311	**452,15**	401,02	503,28	**0,91**	0,81	0,82	1,01
09	113	26.588	**106,25**	86,70	125,80	**0,85**	0,69	0,72	1,00
10	325	10.468	**776,18**	693,11	859,24	**1,06**	0,94	0,96	1,17
11	261	12.916	**505,19**	444,52	565,85	**1,14**	1,00	1,02	1,28
12	271	11.900	**569,33**	502,32	636,34	**0,93**	0,82	0,84	1,04
13	335	14.881	**562,80**	503,21	622,38	**0,99**	0,89	0,90	1,10
14	758	19.684	**962,71**	895,51	1.029,91	**1,10**	1,03	1,04	1,18
15	229	15.292	**374,38**	326,25	422,50	**0,91**	0,79	0,81	1,03
16	381	27.993	**340,26**	306,33	374,20	**1,23**	1,11	1,13	1,35
17	302	20.730	**364,21**	323,43	404,98	**1,19**	1,06	1,08	1,33
18	107	6.638	**402,98**	327,24	478,72	**1,35**	1,09	1,13	1,60
19	334	15.919	**524,53**	468,87	580,19	**1,06**	0,94	0,96	1,17
20	466	22.118	**526,72**	479,40	574,04	**1,04**	0,95	0,96	1,14
21	394	16.487	**597,44**	539,16	655,72	**1,01**	0,91	0,92	1,11
22	378	15.445	**611,85**	550,93	672,77	**0,92**	0,82	0,84	1,01
23	162	17.464	**231,91**	196,36	267,45	**0,57**	0,48	0,50	0,66
24	312	20.575	**379,10**	337,36	420,85	**0,74**	0,66	0,67	0,82
25	547	24.905	**549,09**	503,58	594,59	**0,99**	0,91	0,92	1,07
26	12	981	**305,81**	133,84	477,78	**0,96**	0,42	0,51	1,49
27	22	1.433	**383,81**	224,66	542,96	**1,03**	0,60	0,67	1,46

Kommentar: *Bydelsnummereringen svarer mot følgende bydelsnavn:*

01 Bygdøy-Frogner
02 Uranienborg
03 St.Hanshaugen
04 Sagene
05 Grünerløkka
06 Gamle Oslo
07 Ekeberg
08 Nordstrand
09 Søndre Nordstrand
10 Lambertseter
11 Bøler
12 Manglerud
13 Østensjø
14 Helsfyr
15 Hellerud
16 Furuset
17 Stovner
18 Romsås
19 Grorud
20 Bjerke
21 Grefsen
22 Sogn
23 Vinderen
24 Røa
25 Ullern
26 Sentrum
27 (administrativ, ikke-geog. bydel)

Tabell 5.2. Stratumspesifikke data som ligger bak estimeringen av råraten og SMR for bydel 01 (Bygdøy-Frogner).

Stratum		Data fra bydel 01 Bygdøy-Frogner				Data fra referansepop. (hele Oslo)			
Kjønn	Aldersgruppe	$\bar{N}_{gk}$	$\hat{m}_{gk}$	$\hat{h}_{gk}$	$\frac{\bar{N}_{gk}}{\bar{N}_{g.}}\hat{h}_{gk}$	$\bar{N}_{.k}$	$\hat{m}_{.k}$	$\hat{h}_{Rk}$	$\frac{\bar{N}_{gk}}{\bar{N}_{g.}}\hat{h}_{Rk}$
(kol.1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
M	1	451	-	-	-	16.512	6	0,00009	0,0000021
M	2	247	-	-	-	11.505	-	-	-
M	3	239	-	-	-	10.175	-	-	-
M	4	319	-	-	-	10.489	1	0,00002	0,0000004
M	5	622	-	-	-	15.734	-	-	-
M	6	1.248	-	-	-	24.243	3	0,00003	0,0000020
M	7	1.083	1	0,00023	0,0000131	23.284	9	0,00010	0,0000055
M	8	762	-	-	-	18.761	21	0,00028	0,0000112
M	9	641	-	-	-	16.600	23	0,00035	0,0000116
M	10	699	3	0,00107	0,0000392	16.550	71	0,00107	0,0000392
M	11	536	3	0,00140	0,0000392	10.799	95	0,00220	0,0000617
M	12	389	4	0,00257	0,0000523	8.620	121	0,00351	0,0000714
M	13	347	6	0,00432	0,0000785	9.393	264	0,00703	0,0001276
M	14	358	15	0,01047	0,0001962	9.560	521	0,01362	0,0002552
M	15	320	18	0,01406	0,0002354	8.982	735	0,02046	0,0003425
M	16	263	30	0,02852	0,0003924	6.404	876	0,03420	0,0004705
M	17	172	42	0,06105	0,0005493	3.880	856	0,05515	0,0004963
M	18	139	41	0,07374	0,0005362	2.209	811	0,09178	0,0006674
KV	1	438	1	0,00057	0,0000131	15.667	26	0,00041	0,0000095
KV	2	242	-	-	-	10.998	2	0,00005	0,0000006
KV	3	261	-	-	-	9.663	-	-	-
KV	4	331	-	-	-	10.403	1	0,00002	0,0000004
KV	5	849	-	-	-	18.317	1	0,00001	0,0000006
KV	6	1.331	-	-	-	26.155	3	0,00003	0,0000020
KV	7	955	-	-	-	21.761	1	0,00001	0,0000006
KV	8	566	-	-	-	17.605	9	0,00013	0,0000038
KV	9	573	-	-	-	15.488	7	0,00011	0,0000034
KV	10	707	2	0,00071	0,0000262	15.886	21	0,00033	0,0000122
KV	11	512	-	-	-	10.969	21	0,00048	0,0000128
KV	12	411	1	0,00061	0,0000131	9.694	35	0,00090	0,0000194
KV	13	402	4	0,00249	0,0000523	10.931	83	0,00190	0,0000399
KV	14	470	2	0,00106	0,0000262	12.298	242	0,00492	0,0001210
KV	15	612	16	0,00654	0,0002093	13.657	510	0,00934	0,0002989
KV	16	606	30	0,01238	0,0003924	11.896	864	0,01816	0,0005756
KV	17	521	56	0,02687	0,0007324	9.131	1.356	0,03713	0,0010119
KV	18	493	97	0,04919	0,0012686	7.948	2.461	0,07741	0,0019965
Totalsum:		19.115	372	0,29785	**0,0048653** *100.000 **=486,53**	472.167	10.056	0,38123	**0,0066737**

5.3 Eksempel 2: Direkte standardisering av dødelighet ialt i Oslo bydeler

I eksempel 2 har vi tatt utgangspunkt i den totale dødeligheten i Oslo bydeler. Vi har til å begynne med samme opplysninger som i eksempel 1, dessuten kjenner vi populasjonsstrukturen på stratumnivå for hele landet pr. 1.1.1981. Sistnevnte opplysninger er fra vår standardpopulasjon, som vi bruker til den direkte standardiseringen.

Vi ønsker å sammenligne bydelene i Oslo med hverandre direkte. For dette har vi valgt å justere ratene slik, at kjønns- og alderssammensetningen er «lik» i alle bydelene. I tillegg ønsker vi å sammenligne Vinderen med alle de andre bydelene. Grunnen til dette er at Vinderen er den bydelen i Oslo hvor den totale dødeligheten er lavest. Dette vil gi SRR-tall som f.eks. kan brukes for å fortelle at Oslo, sett fra en sosialmedisinsk vinkling, er en segregert by.

Det er altså to effektmål vi er interesserte i her, nemlig

- det forventete antallet dødsfall pr. år i perioden gitt standardpopulasjonen, det vil si den direkte standardiserte insidensraten $D\hat{I}_g$
- det standardiserte risikoforholdet, $S\hat{R}R_{E|U}$, som forteller om dødeligheten i de eksponerte bydelene i forhold til en ueksponert kontrollbydel

Dessuten har vi her, på samme måte som i eksempel 1, valgt å presentere 95-prosents konfidensintervaller til effektmålene.

I **tabell 5.3** ser vi presentasjonen av resultatene. Kolonne (4) gjengir den direkte standardiserte insidensraten, mens kolonnene (5) og (6) gir et 95 prosents konfidensintervall til den samme. Kolonne (7) viser på tilsvarende måte det standardiserte risikoforholdet, SRR, og kolonne (8) og (10) gjengir grensene til et 95 prosents konfidensintervall. Kolonne (9) gir den nedre (og eneste) grensen til et 95-prosents ensidig konfidensintervall.

Ut fra den ensidige konfidensgrensen til SRR kan vi allerede nå si at dødeligheten med en-stjerners signifikans er lavere for bydel 23 (Vinderen, som er vår kontrollpopulasjon) enn hva den er for alle de andre bydelene. Konfidensgrensen er nemlig større enn 1 for disse. Vi kan også si at dødeligheten er omtrent dobbelt så høy for Romsås, som kommer «dårligst» ut, som den er for kontrollbydelen Vinderen. SRR-tallet er her 2,23.

Tabell 5.4 går ner på et mer detaljert nivå. Her har vi valgt å se på opplysninger om standardpopulasjonen i 1981 (kolonne (3)) samt data for Vinderen bydel (kolonnene (4), (5) og (6)) og Gamle Oslo (kolonnene (7), (8) og (9)) for hvert enkelt standardiseringsstratum.

Den direkte standardiserte insidensraten for kontrollbydelen Vinderen er blitt beregnet ved

$$\hat{DI}_U = \frac{1}{N}\sum_{k=1}^{K} N_k \hat{h}_{Uk} = \frac{1}{\sum(\text{kol.3})}\sum_{k=1}^{K}(\text{kol.3})(\text{kol.5}) = \sum_{k=1}^{K}\frac{(\text{kol.6})}{\sum(\text{kol.3})} =$$ (51) *(tilpassning av (9))*

$$= \frac{1}{4.092.340}(132.546 * 0,00105 + \ldots + 30.333 * 0,12270) \approx$$

$$\approx (0,0000341 + \ldots + 0,0009095) = 0,0064299$$

$$\Rightarrow 0,0064299 * 100.000 = 642,99$$

mens tilsvarende effektmål for den eksponerte bydelen Gamle Oslo er blitt beregnet ved

$$\hat{DI}_E = \frac{1}{N}\sum_{k=1}^{K} N_k \hat{h}_{Ek} = \frac{1}{\sum(\text{kol.3})}\sum_{k=1}^{K}(\text{kol.3})(\text{kol.8}) = \sum_{k=1}^{K}\frac{(\text{kol.9})}{\sum(\text{kol.3})} =$$ (52) *(tilpassning av (9))*

$$= \frac{1}{4.092.340}(132.546 * 0,00330 + \ldots + 30.333 * 0,19143) \approx$$

$$\approx (0,0001070 + \ldots + 0,0014190) = 0,0135060$$

$$\Rightarrow 0,0135060 * 100.000 = 1350,6.$$

Estimatet til det standardiserte risikoforholdet har vi deretter fått ved å dividere insidensraten til den eksponerte bydelen med insidensraten til kontrollbydelen, slik at

$$\hat{SRR}_{E|U} = \frac{\hat{DI}_E}{\hat{DI}_U} = \frac{0,0135060}{0,0064299} \approx 2,10.$$ (53) *(tilpassning av (10))*

Prinsippet for beregning av konfidensgrenser er den samme som i eksempel 1. Vi trenger først variansen til effektmålene. For den direkte standardiserte insidensraten får vi, hvis vi holder oss til Gamle Oslo bydel, variansen

$$\hat{v}(\hat{DI}_E) = \frac{\hat{DI}_E t(1 - \hat{DI}_E t)}{\overline{N}_{E.} t^2} = \frac{0,0135060 * 4 * (1 - 0,0135060 * 4)}{20.628 * 16} \approx 0,00000015484$$ (54) *(tilpassning av (35))*

mens kontrollbydelen Vinderen på tilsvarende måte vil få variansen

$$\hat{v}(\hat{DI}_U) = \frac{\hat{DI}_U t(1 - \hat{DI}_U t)}{\overline{N}_{U.} t^2} = \frac{0,0064299 * 4 * (1 - 0,0064299 * 4)}{17.464 * 16} \approx 0,00000008967.$$ (55) *(tilpassning av (35))*

Variansen til det standardiserte risikoforholdet vil bli

$$\hat{v}(S\hat{R}R_{E|U}) = \frac{D\hat{I}_U{}^2 \hat{v}(D\hat{I}_E) + D\hat{I}_E{}^2 \hat{v}(D\hat{I}_U)}{D\hat{I}_U{}^4} =$$

$$= \frac{0{,}0064299^2 * 0{,}00000015484 + 0{,}0135060^2 * 0{,}00000008967}{0{,}0064299^4} \approx$$

$$\approx 0{,}0133146.$$

(56) *(tilpass-ning av (36))*

Variansmålene har gitt oss konfidensgrensene i **tabell 5.3.** For Gamle Oslo har disse blitt beregnet på følgende måte:

$$DI_{E(2,5\%)} = D\hat{I}_E - 1{,}96\sqrt{\hat{v}(D\hat{I}_E)} = 0{,}0135060 - 1{,}96\sqrt{0{,}00000015484} =$$

$$= 0{,}0127346$$

$$\Rightarrow 0{,}0127346 * 100.000 = 1.273{,}46. \quad (57)$$

og

$$DI_{E(97,5\%)} = D\hat{I}_E + 1{,}96\sqrt{\hat{v}(D\hat{I}_E)} = 0{,}0135060 + 1{,}96\sqrt{0{,}00000015484} =$$

$$= 0{,}0142771$$

$$\Rightarrow 0{,}0142771 * 100.000 = 1.427{,}71. \quad (58)$$

Konfidensgrensene til den direkte standardiserte insidensraten fremgår av kolonnene (5) og (6) i **tabell 5.3**. For SRR-tallet har konfidensgrensene (kolonnene (8), (9) og (10)) hhv. blitt beregnet som

$$SRR_{E|U(2,5\%)} = S\hat{R}R_{E|U} - 1{,}96\sqrt{\hat{v}(S\hat{R}R_{E|U})} = 2{,}10 - 1{,}96\sqrt{0{,}0133146} \approx 1{,}87, \quad (59)$$

$$SRR_{E|U(5\%)} = S\hat{R}R_{E|U} - 1{,}645\sqrt{\hat{v}(S\hat{R}R_{E|U})} = 2{,}10 - 1{,}645\sqrt{0{,}0133146} \approx 1{,}91 \quad (60)$$

og

$$SRR_{E|U(97,5\%)} = S\hat{R}R_{E|U} + 1{,}96\sqrt{\hat{v}(S\hat{R}R_{E|U})} = 2{,}10 + 1{,}96\sqrt{0{,}0133146} \approx 2{,}33. \quad (61)$$

Tabell 5.3. Estimater til den direkte standardiserte insidensraten og SRR, med konfidensgrenser, for diagnostisert total dødelighet. Bydeler i Oslo.

Bydel	$\hat{m}_{g.}$	$\overline{N}_{g.}$	$\hat{DI}_{g}$	$DI_{g(2,5\%)}$	$DI_{g(97,5\%)}$	$\hat{SRR}_{EIU}$	$SRR_{EIU(2,5\%)}$	$SRR_{EIU(5\%)}$	$SRR_{EIU(97,5\%)}$
(kol.1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
01	872	19.115	**772,29**	710,97	833,61	**1,20**	1,06	1,08	1,35
02	1.517	22.299	**968,14**	904,83	1.031,45	**1,51**	1,34	1,36	1,67
03	1.656	26.421	**1.012,86**	953,43	1.072,30	**1,58**	1,40	1,43	1,75
04	2.093	26.515	**1.228,36**	1.163,31	1.293,40	**1,91**	1,71	1,74	2,11
05	1.492	23.495	**1.328,51**	1.256,81	1.400,22	**2,07**	1,85	1,88	2,29
06	1.170	20.628	**1.350,58**	1.273,46	1.427,71	**2,10**	1,87	1,91	2,33
07	726	14.966	**895,82**	821,37	970,27	**1,39**	1,22	1,25	1,57
08	689	16.311	**854,72**	785,00	924,43	**1,33**	1,17	1,19	1,49
09	320	26.588	**801,65**	748,71	854,60	**1,25**	1,11	1,13	1,39
10	711	10.468	**1.012,74**	918,32	1.107,16	**1,58**	1,37	1,40	1,78
11	650	12.916	**1.139,22**	1.049,30	1.229,14	**1,77**	1,56	1,59	1,99
12	609	11.900	**870,61**	788,25	952,96	**1,35**	1,18	1,20	1,53
13	773	14.881	**943,86**	867,30	1.020,42	**1,47**	1,29	1,32	1,65
14	1.733	19.684	**1.123,77**	1.051,41	1.196,14	**1,75**	1,55	1,58	1,94
15	540	15.292	**850,07**	778,25	921,88	**1,32**	1,16	1,18	1,49
16	981	27.993	**1.202,02**	1.139,37	1.264,68	**1,87**	1,67	1,70	2,07
17	752	20.730	**1.166,67**	1.094,88	1.238,45	**1,81**	1,61	1,65	2,01
18	297	6.638	**1.431,06**	1.291,35	1.570,78	**2,23**	1,93	1,98	2,52
19	829	15.919	**1.094,96**	1.015,48	1.174,43	**1,70**	1,50	1,54	1,90
20	1.123	22.118	**1.023,58**	958,29	1.088,87	**1,59**	1,41	1,44	1,77
21	872	16.487	**920,20**	848,35	992,06	**1,43**	1,26	1,29	1,60
22	886	15.445	**877,25**	804,70	949,80	**1,36**	1,20	1,22	1,53
23	459	17.464	**642,99**	584,29	701,68	**1,00**	.	.	.
24	778	20.575	**755,65**	697,17	814,14	**1,18**	1,03	1,06	1,32
25	1.370	24.905	**959,94**	900,27	1.019,60	**1,49**	1,33	1,35	1,66
26	40	981	**1.081,46**	763,19	1.399,73	**1,68**	1,16	1,25	2,20
27	49	1.433	**950,95**	703,35	1.198,56	**1,48**	1,07	1,14	1,89

Tabell 5.4. Stratumspesifikke data som ligger bak estimeringen av den direkte standardiserte insidensraten og SRR for den ueksponerte kontrollbydelen (23 Vinderen) og en eksponert bydel (06 Gamle Oslo)

Stratum: Kjønn	Aldersgruppe	N_k	Bydel 23 -- Vinderen $\hat{m}_{Uk}$	$\hat{h}_{Uk}$	$\frac{N_k}{N}\hat{h}_{Uk}$	Bydel 06 -- Gamle Oslo $\hat{m}_{Ek}$	$\hat{h}_{Ek}$	$\frac{N_k}{N}\hat{h}_{Ek}$
(kol.1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
M	01	132.546	3	0,00105	0,0000341	11	0,00330	0,0001070
M	02	157.217	-	-	-	1	0,00051	0,0000200
M	03	170.755	-	-	-	-	-	-
M	04	162.448	1	0,00045	0,0000177	1	0,00066	0,0000260
M	05	158.039	1	0,00050	0,0000194	2	0,00053	0,0000200
M	06	156.202	3	0,00144	0,0000548	8	0,00121	0,0000460
M	07	163.295	-	-	-	10	0,00179	0,0000710
M	08	134.546	3	0,00101	0,0000333	13	0,00329	0,0001080
M	09	104.569	2	0,00077	0,0000196	8	0,00246	0,0000630
M	10	97.283	5	0,00158	0,0000375	15	0,00589	0,0001400
M	11	106.656	4	0,00222	0,0000579	13	0,00699	0,0001820
M	12	116.635	7	0,00538	0,0001535	20	0,01597	0,0004550
M	13	109.397	5	0,00414	0,0001106	34	0,03058	0,0008170
M	14	93.352	24	0,01863	0,0004251	51	0,04740	0,0010810
M	15	72.209	30	0,02475	0,0004368	60	0,05976	0,0010540
M	16	48.349	47	0,04776	0,0005643	75	0,09766	0,0011540
M	17	28.006	32	0,05594	0,0003829	90	0,14331	0,0009810
M	18	16.076	65	0,21959	0,0008626	95	0,30063	0,0011810
KV	01	125.969	2	0,00074	0,0000229	4	0,00139	0,0000430
KV	02	150.600	-	-	-	-	-	-
KV	03	161.799	1	0,00046	0,0000181	1	0,00074	0,0000290
KV	04	153.947	1	0,00045	0,0000168	1	0,00069	0,0000260
KV	05	151.220	2	0,00085	0,0000316	2	0,00042	0,0000160
KV	06	148.243	-	-	-	6	0,00101	0,0000370
KV	07	152.656	-	-	-	6	0,00139	0,0000520
KV	08	127.085	2	0,00071	0,0000220	6	0,00189	0,0000590
KV	09	101.678	6	0,00207	0,0000514	4	0,00188	0,0000470
KV	10	96.006	2	0,00066	0,0000156	5	0,00252	0,0000590
KV	11	106.283	8	0,00477	0,0001240	10	0,00767	0,0001990
KV	12	120.223	4	0,00278	0,0000816	5	0,00504	0,0001480
KV	13	119.210	12	0,00860	0,0002504	11	0,00935	0,0002720
KV	14	107.237	11	0,00731	0,0001917	34	0,02648	0,0006940
KV	15	92.445	28	0,01768	0,0003993	50	0,03397	0,0007670
KV	16	72.572	33	0,02806	0,0004976	95	0,05835	0,0010350
KV	17	47.254	35	0,05087	0,0005874	155	0,09498	0,0010970
KV	18	30.333	80	0,12270	0,0009095	268	0,19143	0,0014190
Totalsum:		4.092.340	459	0,63394	**0,0064299**	1170	1,16114	**0,0135060**
					*100.000			*100.000
					=642,99			**=1.350,60**

5.4 Tolking av resultatene

Hvordan skal man tolke ratene som ble estimert i eksemplene? Når det gjelder de indirekte standardiserte effektmålene ifølge eksempel 1 så går vi ut fra alders- og kjønnsfordelingen for den enkelte bydel. Denne sammenligner vi med hele Oslo ved å lage en «tenkt» populasjon med eksakt samme alders- og kjønnsfordeling som bydelen vi studerer, men med hele Oslo kommunes dødelighetstall. SMR forteller oss derved om overdødeligheten (eller underdødeligheten) i den studerte bydelen i forhold til hele Oslo, og holder befolkningssammensetningen konstant slik at en sammenligning er mulig.

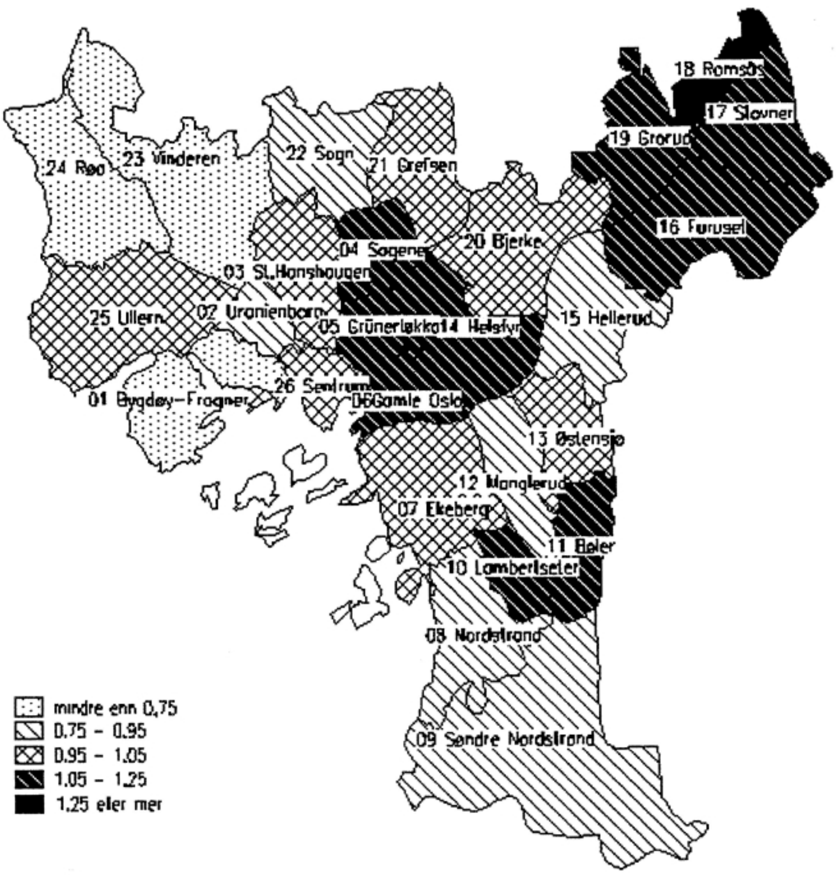
De direkte standardiserte effektmålene skal tolkes noe annerledes. Her gir vi alle bydelene i undersøkelsen samme alders- og kjønnssammensetning. Effektmålet som vi estimerer, den direkte standardiserte insidensraten DI_g som i eksempel 2 beskriver det forventete totale antallet dødsfall, vil da bli sammenlignbart for alle bydelene. En bydel, vi har valgt Vinderen i eksempelet, sammenligner vi med alle de andre bydelene. Vi ser på hvor stor «risiko» et individ løper for å dø dersom vedkommende bor i de andre bydelene i Oslo sammenlignet med om han eller hun bor på Vinderen. Dette gjør vi ved å beregne SRR, som er definert som forholdet mellom den direkte standardiserte insidensraten for hver enkelte studerte, eller eksponerte bydel, og tilsvarende insidensrate for den ueksponerte kontrollbydelen, som i dette tilfellet altså er Vinderen bydel.

Den direkte standardiserte insidensraten, DI_g, brukes i SSB's NOS-publikasjoner om dødsårsaker. Der er den vanligvis «oppblåst» med tallet 100.000. Den forteller om det forventete antallet dødsfall pr. 100.000 innbyggere, gitt alders- og kjønnsfordelingen til en definert standardpopulasjon.

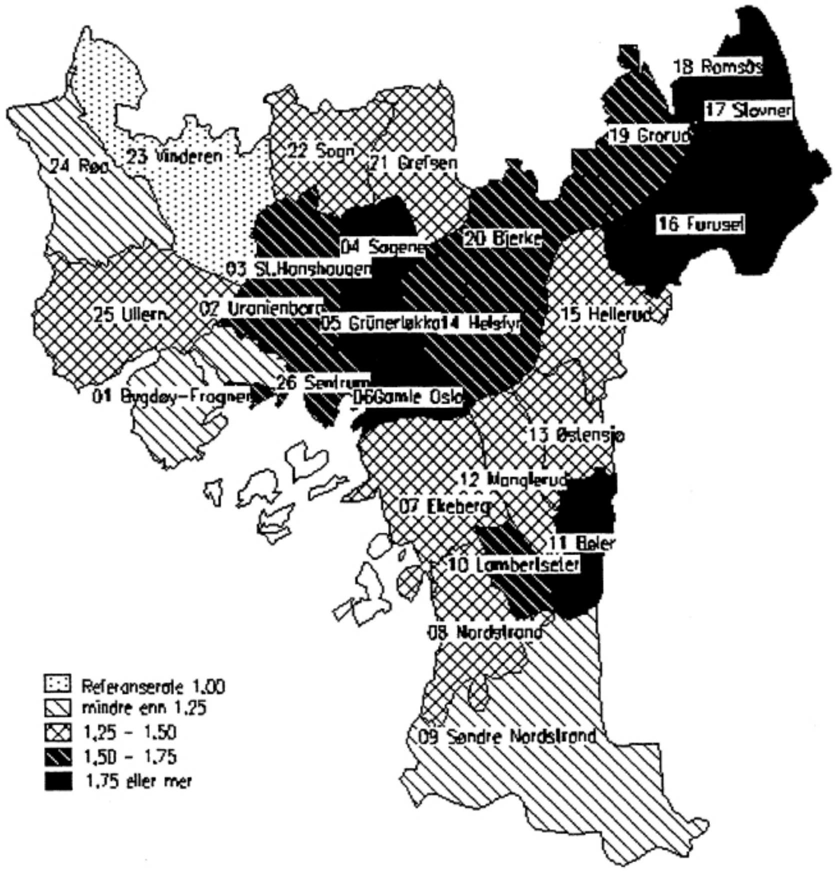
Et helt annet spørsmål er hvordan man skal tolke resultatene fra eksemplene ut fra en mer generell, samfunnsmedisinsk tilnærming. Figurene 5.1 og 5.2 gir en grafisk fremstilling av hvordan hhv. SMR-tallene fra eksempel 1 og SRR-tallene fra eksempel 2 fordeler seg i de forskjellige bydelene. Vi kan her se en tydelig tendens. Dødelighetstallene er i begge tilfellene høyere i bydeler som er preget av lav sosioøkonomisk status. Typiske «østkantstrøk» som f.eks. Gamle Oslo, Grünerløkka, Sagene og Helsfyr kommer dårligere ut enn de mer velbestilte områdene i Uranienborg, Bygdøy-Frogner og Vinderen. Også enkelte innvandrertette områder i de store drabantbyene, f.eks. Romsås og Stovner, har høye SMR- og SRR-tall.

Man bør kanskje ikke være alt for rask med å trekke konklusjoner ut fra denne siste iakttagelsen. Materialet må kompletteres med flere, og mer omfattende studier. Hva man derimot bør fremholde er at standardiserte rater innenfor den samfunnsmedisinske forskningen generelt sett er interessante som verktøy for undersøkelse og presentasjon av et datamateriale.

Figur 5.1. Temakartfremstilling som viser hvordan det estimerte indirekte standardiserte dødelighetstallet (SMRg) for dødelighet i hjerte- og karsykdommer, fra eksempel 1, fordeler seg i de forskjellige bydelene i Oslo.



Figur 5.2. Temakartfremstilling som viser hvordan det estimerte direkte standardiserte risikoforholdet (SRR_{EIU}) for total dødelighet, fra eksempel 2, fordeler seg i de forskjellige bydelene i Oslo.



5.5 SAS-programmer for beregningene

De to SAS-programmene som presenteres her ble brukt til beregningene som ligger bak **tabellene 5.1** og **5.3**. Vi tror at de kan være en god «hint» for den som står i begrep å beregne justerte rater. Programmene er skrevet for SAS-versjon 6.11 men kan, med eller uten mindre justeringer, også brukes til eldre SAS-versjoner.

La oss i utgangspunktet bruke tre datasett som inneholder alle de (ubearbeidete) opplysningene vi trenger. De er

- **sd3.gjoslo93**, som inneholder opplysninger om gjennomsnittspopulasjonen pr. 1.1.1993,
- **sd3.stan81**, som inneholder opplysninger om standardpopulasjonen pr. 1.1.1981, og
- **sd3.oslo9194**, som inneholder aggregerte data (intensiteter, eller dødelighetstall) fra dødsårsaksregisteret for årene 1991-94.

La oss dessuten gå ut fra at SAS-katalogreferansen **sd3** allerede er definert ved kommandoen

```
libname sd3 'katalognavn';
```

Dataene i de tre SAS-datasettene ligger organisert på følgende måte:

sd3.gjoslo93:

Datasettet inneholder to numeriske variabler: NG93 og NGTOT93, og tre grupperingsvariabler: BYDEL, ALDGR og KJONN.

NG93 uttrykker antallet innbyggere pr. 1.1.1993 fordelt på grupperingsvariablene BYDEL (bydel i Oslo, to-siffer kode), ALDGR (to-siffer kode for aldersgruppe, gruppering er blitt utført etter 5-års aldersintervaller) og KJONN (en-siffer kode for kjønn). NG93 = $\overline{N}_{gk}$ ifølge notasjonen som vi bruker til formlene og tabellene.

NGTOT93 uttrykker også antallet innbyggere pr. 1.1.1993, men her er tallet summert over BYDEL og altså fordelt på ALDGR og KJONN. NGTOT93 = $\overline{N}_{.k}$ ifølge vår notasjon i formler og tabeller.

sd3.stan81:

Datasettet inneholder to numeriske variabler: NS81 og NSTOT81, og to grupperingsvariabler: ALDGR og KJONN.

NS81 uttrykker antallet innbyggere i standardpopulasjonen, som er definert som *hele landet* pr. 1.1.1981, fordelt på ALDGR og KJONN. NS81 = N_k ifølge vår notasjon i formler og tabeller.

NSTOT81 uttrykker standardpopulasjonen summert over ALDGR og KJONN, det vil si standardpopulasjonen ialt (som var 4.092.340 innbyggere). NSTOT81 = N ifølge vår notasjon i formler og tabeller.

sd3.oslo9194:

Datasettet inneholder et antall numeriske variabler: DA1, DA2,...,DA*i*,...,DA*n*, og tre grupperingsvariabler: BYDEL, ALDGR og KJONN.

DA1 uttrykker det totale antallet dødsfall i årene 1991-94 fordelt på BYDEL, ALDGR og KJONN.

DA2 uttrykker antallet dødsfall i hjerte- og karsykdommer i årene 1991-94 og er også fordelt på grupperingsvariablene.

DA3-DA*n* uttrykker på tilsvarende måte antallet dødsfall i andre diagnoser. I eksemplene brukte vi bare DA1 og DA2.

For alle DA-variablene gjelder, at $\mathrm{DA}i = \hat{\mathrm{m}}_{\mathrm{gk}}$ ifølge vår notasjon i formler og tabeller.

SAS-programmet for beregning av tallene til **tabell 5.1** vil derved se ut på følgende måte:

```
/*
    SAS-program som beregner råraten(CR_g) og det
    indirekte standardiserte dødlighetstallet (SMR_g),
    ifølge eksempel 1. Programmet produserer tabell 5.1.
*/

*** Beregner først ng_ref, da_ref og h_ref for referansepop.;

proc means data=sd3.gjoslo93 noprint nway;
var ng93;
class kjonn aldgr;
output out=gj sum=ng_ref;

proc means data=sd3.oslo9194 noprint nway;
var da2;
class kjonn aldgr;
output out=os sum=da_ref;

data a_ref(keep=kjonn aldgr h_ref);
  merge gj os;
  by kjonn aldgr;
  h_ref=da_ref/(ng_ref*4);
```

```
*** Kobler a_ref med data på bydelsnivå,
*** beregner h_hatt, wh_hatt og wh_ref;

data a;
  merge sd3.gjoslo93 sd3.oslo9194;
  by bydel kjonn aldgr;

proc sort data=a;
  by kjonn aldgr;

data a;
  merge a a_ref;
  by kjonn aldgr;

proc sort data=a;
  by bydel kjonn aldgr;

data a;              /* lager alders- og kjønnsspesifikk insidensrate */
  set a;
  by bydel kjonn aldgr;
  h_hatt=da2/(ng93*4);
  wh_hatt=h_hatt*(ng93/ngtot93);
  wh_ref=h_ref*(ng93/ngtot93);

proc means data=a noprint nway; /* lager rårate */
var wh_hatt wh_ref da2 ng93;
class bydel;
output out=b(drop=_type_ _freq_) sum=;

data c;              /* lager smr */
  set b;
  smr=wh_hatt/wh_ref;

  var_wh=wh_hatt*4*(1-wh_hatt*4)/(ng93*16); /* variansen til wh_hatt */
  wh_025=wh_hatt-1.96*sqrt(var_wh);         /* konfidens-    */
  wh_975=wh_hatt+1.96*sqrt(var_wh);         /* grenser       */

  var_smr=var_wh/(wh_ref**2);               /* variansen til smr */
  smr_025=smr-1.96*sqrt(var_smr);           /* konfi-        */
  smr_050=smr-1.645*sqrt(var_smr);          /* dens-         */
  smr_975=smr+1.96*sqrt(var_smr);           /* grenser       */

  cr_e=wh_hatt*100000;
  cr_025=wh_025*100000;
  cr_975=wh_975*100000;


proc print data=c; /* Utprinting av tabell 5.1 */
  var da2 ng93 cr_e cr_025 cr_975 smr smr_025 smr_050 smr_975;
  id bydel;
run;
```

Utskriften blir identisk med **tabell 5.1** når det gjelder variablenes rekkefølge. Til slutt er altså variablene definert som

BYDEL = kolonne (1) i **tabell 5.1**,
DA2 = kolonne (2),
NG93 = kolonne (3),
CR_E = kolonne (4),
CR_025 = kolonne (5),
CR_975 = kolonne (6),
SMR = kolonne (7),
SMR_025 = kolonne (8),
SMR_050 = kolonne (9) og
SMR_975 = kolonne (10).

Tilsvarende SAS-program for beregning av tallene i eksempel 2, som blir presentert i **tabell 5.3**, vil se ut på følgende måte:

```
/*
    SAS-program som beregner den direkte standardiserte insidensraten
    (DI_g) og det direkte standardiserte risikoforholdet (SRR_E|U),
    ifølge eksempel 2. Programmet produserer tabell 5.3.
*/

*** Hent inn da1, ng93, ngtot93, ns81 og nstot81;

data a;
  merge sd3.gjoslo93 sd3.oslo9194;
  by bydel kjonn aldgr;

proc sort data=a;
  by kjonn aldgr;

data uu;
  set sd3.oslo9194;
  if bydel='23';
data a;
  merge a sd3.stan81 uu(drop=bydel da1-da9);
  by kjonn aldgr;

proc sort data=a;
  by bydel kjonn aldgr;
```

```
data a b;           /* Lager alders- og kjønnsspesifikk insidensrate */
  set a;            /* Forbereder variansberegningene                  */
  by bydel kjonn aldgr;
  h_e=da1/(ng93*4);
  wh_e=h_e*(ns81/nstot81);

  varh_e=(da1/ng93)*(1-da1/ng93)/(ng93*16);
  w_ve=(ns81/nstot81)*varh_e;

  output a;                        /* a = eksponert populasjon     */
  if bydel='23' then output b;     /* b = ueksponert kontrollpop. */

data b(keep=wh_u w_vu kjonn aldgr ng93);
  set b;
    wh_u=wh_e;
    w_vu=w_ve;

proc means data=a noprint nway;    /* grunnlag for variansberegn   */
var ng93  da1  wh_e w_ve;          /* - eksponert populasjon       */
class bydel;
output out=a2(drop=_type_ _freq_) sum=ne93 da1 wh_e w_ve;

proc means data=b noprint nway;    /* grunnlag for variansberegn   */
var ng93  wh_u w_vu;               /* - ueksponert pop.            */
output out=b2(drop=_type_ _freq_) sum=nu93 wh_u w_vu;

data a2;
  set a2;
  koble=1;
data b2;
  set b2;
  koble=1;
data c;
  merge a2 b2;
  by koble;

  var_di=wh_e*4*(1-wh_e*4)/(ne93*16);
  var_di_u=wh_u*4*(1-wh_u*4)/(nu93*16);

  di_e=wh_e*100000;
  di025=(wh_e-1.96*sqrt(var_di)) *100000;
  di975=(wh_e+1.96*sqrt(var_di)) *100000;

  srr=wh_e/wh_u;
  ln_srr=log(srr);

  v_srr=((wh_u**2)*var_di+(wh_e**2)*var_di_u)/(wh_u**4);
  srr025=srr-1.96*sqrt(v_srr);
  srr050=srr-1.645*sqrt(v_srr);
  srr975=srr+1.96*sqrt(v_srr);
```

```
proc print data=c;   /* Utprinting av tabell 5.3 */
var da1 ne93 di_e di025 di975 srr srr025 srr050 srr975;
id bydel;

run;
```

Her er variablene til slutt definert som

BYDEL = kolonne (1) i **tabell 5.3**,
DA1 = kolonne (2),
NE93 = kolonne (3),
DI_E = kolonne (4),
DI025 = kolonne (5),
DI975 = kolonne (6),
SRR = kolonne (7),
SRR025 = kolonne (8),
SRR050 = kolonne (9) og
SRR975 = kolonne (10).

6 Litteratur

- AHLBOM, ANDERS (1990) *Biostatistik för epidemiologer* (Studentlitteratur, Lund)
- GUSTAFSSON, LENNART & HÅLLBERG, BENGT & WALL, STIG (1980) *Epidemiologi, grundläggande principer och metoder* (Umeå universitet, Umeå)
- GRIMMETT, GEOFFREY & STIRZAKER, DAVID (1982) *Probability and Random Processes* (Oxford Science Publications, New York)
- HOGG, ROBERT V. & TANIS, ELLIOT A. (1977) *Probability & Statistical Inference* (Macmillan Publishing Co., New York)
- KLEINBAUM, DAVID G. & KUPPER, LAWRENCE L & MORGENSTERN, HAL (1982) *Epidemiologic Research, Principles and Quantitative Methods* (Lifetime Learning Publications, Belmont, California)
- LANCASTER, HENRY OLIVER (1990) *Expectations of Life* (Springer-Verlag, New York)
- ROTHMAN, KENNETH J (1986) *Modern Epidemiology* (Little, Brown and Company, Boston)
- THOMSEN, IB (1981) *Standardisering som analyseteknikk* (Interne notater 81/24, SSB)

De sist utgitte publikasjonene i serien Notater

96/53 F.R. Aune: Konsekvenser av en nordisk avgiftsharmonisering på elektrisitetsområdet. 22s.

96/54 M.V. Dysterud og P. Schøning: SSB-AVLØP. 187s.

96/55 E. Vassnes og I. Tuveng: Datagrunnlag for analyse av personers overgang fra utdanning til arbeid: Dokumentasjon. 58s.

96/56 K. Flugsrud, O.K. Hunnes og E. Lasson: Metode for beregning av energivarebruk og utslipp på grunnkretser: Beregninger for 1992 og 1993 for kommunene Oslo, Drammen, Bergen og Trondheim. 61s.

96/57 T. Kalve: Bedre barnevernsdata på edb-lesbart medium. 42s.

96/58 E. Midtlyng og A.A. Ritland: Leseferdigheter i den voksne befolkningen i Norge: Pilot-undersøkelse: Dokumentasjonsrapport. 53s.

96/59 A. Sundvoll og L. Solheim: Undersøkelse om kopiering på universiteter og høgskoler: Pilot-undersøkelse: Dokumentasjonsrapport. 48s.

96/60 A. Sundvoll: Undersøkelse om levekår og nærmiljø i Bergen: Dokumentasjonsrapport. 53s.

96/61 A. Bråten: Populasjon og utvalg - konsum-prisindeksen. 58s.

96/62 M. Kjelsrud og A. Torstensen: Innvandreres tilknytning til arbeidsmarkedet. Situasjonen i november 1994. Bruttoendringer mellom november 1993 og november 1994: Dokumentasjon og analyse. 170s.

96/63 H.M. Teigum: Samordnet levekårsunder-søkelse 1996 - tverrsnittsundersøkelsen: Dokumentasjonsrapport. 57s.

96/64 Å. Kaurin: Emballasjestatistikk: Utprøving av metoder og forslag til metode for innhenting av data til en nasjonal statistikk over emballasjeavfall. 46s.

97/1 S. Opdahl: Levekårsundersøkelse blant mottakere av grunnstønad: Dokumentasjons-rapport med tabeller. 138s.

97/2 E. Berg og K. Rypdal: Historisk utvikling og fremskrivning av forbruket av noen miljø-skadelige produkter. 23s.

97/3 A. Sundvoll: Undersøkelse om velferdsstatens gleder og byrder: Dokumentasjonsrapport. 88s.

97/4 M.S. Bjerkseth: Evaluering av ny metode for utarbeidelse av strukturstatistikk ved Seksjon 460. 145s.

97/6 E. Gulløy, S. Blom og A.A. Ritland: Levekår blant innvandrere 1996: Dokumentasjons-rapport med tabeller. 205s.

97/7 S. Blom og A.A. Ritland: Levekår blant innvandrere 1996: Del 2: Tabeller for nordmenn. 1997. 222s.

97/8 T.C. Mykkelbost: Resultater fra bruker-undersøkelse i forbindelse med NOS 306: Utslipp til luft i norske kommuner 1993. 21s.

97/9 H.M. Teigum: Omnibusundersøkelsene 1996: Dokumentasjonsrapport. 136s.

97/10 P.O. Lande og T. Heimdal: GERIX START: System- og brukardokumentasjon. 49s.

97/11 A. Barstad: Frihetens århundre? Levekår i Norge i et 100-årsperspektiv. 37s.

97/12 G. Sparby: Inntekts- og formuesunder-søkelsen 1992: Dokumentasjon. 101s.

97/13 V. Pedersen: Inntekts- og formuesunder-søkelsen 1993: Dokumentasjon. 94s.

97/14 V. Pedersen: Inntekts- og formuesunder-søkelsen 1994: Dokumentasjon. 93s.

97/15 Metodevalg og kostnader ved etablering og drift av et boligregister. 29s.

97/16 K. Vassenden: Innvandrerstatistikkprosjektet: Styringsgruppas evaluering. 34s.

97/19 H.M. Teigum: Verdiundersøkelsen 1996: Dokumentasjonsrapport. 84s.

97/20 T. Ouren og T. Vik: Prosjektrapport: Voksenopplæringsprosjektet 1995-1996. 24s.

Notater

B *Returadresse:*
Statistisk sentralbyrå
Postboks 8131 Dep.
N-0033 Oslo

Statistisk sentralbyrå

Oslo:
Postboks 8131 Dep.
0033 Oslo

Telefon: 22 86 45 00
Telefaks: 22 86 49 73

Kongsvinger:
Postboks 1260
2201 Kongsvinger

Telefon: 62 88 50 00
Telefaks: 62 88 50 30

ISSN 0806-3745

Statistisk sentralbyrå
Statistics Norway