

Eksamen

ERN 4010

Ernæringsepidemiologi

Fredag 4. oktober 2026 kl 09:00-12:00

Oppgavesettet består av 5 oppgaver.

Poengfordeling mellom oppgavene:

Oppgave 1 – 24 poeng

Oppgave 2 – 21 poeng

Oppgave 3 – 30 poeng

Oppgave 4 – 20 poeng

Oppgave 5 – 5 poeng

I alt 100 poeng

Besvar oppgavene kort og konsist.

Lykke til!

Kalkulator finnes i Inspira.

Oppgave 1 (24 poeng)

Du gjennomfører en studie for å undersøke om det er noen sammenheng mellom daglig inntak av energidrikk og høyt blodtrykk blant ungdommer og unge voksne (alder 15 til 25 år). Du har fått rekruttert 2000 personer til å delta i studien.

Bruk av energidrikk er selvrapportert i et spørreskjema, mens blodtrykket er målt av deg og medarbeiderne dine på forskningssenteret der du arbeider.

Deltakerne (n) fordeler seg som vist i tabellen under.

	Høyt blodtrykk	Ikke høyt blodtrykk
Energidrikk daglig	60	940
Ikke energidrikk daglig	30	970

1a) Hva er odds ratioen for høyt blodtrykk for de som daglig inntar energidrikk? Bruk de som ikke drikker energidrikk daglig som referansegruppe. Vis regnestykket og oppgi odds ratioen

(2 poeng).

Svar: $(60 \cdot 970) / (30 \cdot 940) = 2,06$.

1b) Ut fra risikoestimatet i 1a, hva vil du si om sammenhengen mellom daglig inntak av energidrikk og oddsen for høyt blodtrykk?

(1 poeng)

Svar: De som daglig drikker energidrikk har om lag dobbel så høy odds for høyt blodtrykk som de som ikke drikker energidrikk daglig.

1c) Det er ressurskrevende at alle deltakerne må møte på forskningssenteret for å måle blodtrykket. Du vurderer derfor om man kan bruke selvrapportert blodtrykk basert på en enkel nyutviklet målemetode som deltakerne kan få tilsendt ved rekruttering til studien. Den nye målemetoden har sensitivitet på 1,0 og spesifisitet på 0,9.

Sett opp en ny tabell (vis tabellen i svaret ditt) som viser antall personer i hver celle basert på denne informasjonen. Vi antar at sensitiviteten og spesifisiteten til energidrikk-variabelen er uendret (begge deler er 1,0). Regn ut odds ratioen for høyt blodtrykk basert på tallene i den nye tabellen. Bruk de som ikke drikker energidrikk daglig som referansegruppe.

(8 poeng)

Svar: Ny tabell

	Høyt blodtrykk	Ikke høyt blodtrykk
Energidrikk daglig	154	846
Ikke energidrikk daglig	127	873

OR = $(154 \cdot 873) / (127 \cdot 846) = 1,25$

Utrekning for hver celle: $60 + (940 - 940 \cdot 0,9) = 154$, $30 + (970 - 970 \cdot 0,9) = 127$, $940 \cdot 0,9 = 846$ og $970 \cdot 0,9 = 873$

1d) Medfører den reduserte spesifisiteten i 1c differensielle eller ikke-differensielle målefeil? Begrunn svaret kort.

(3 poeng)

Svar: Ikke-differensielle. Feilstrukturen i blodtrykksvariabelen er lik hos de eksponerte (energidrikk daglig) og de ikke-eksponerte (ikke energidrikk daglig).

1e) Dersom du mener målefeilene i oppgaven er ikke-differensielle, gi et eksempel på hvordan feilstrukturen kunne vært dersom målefeilene var differensielle.

Dersom du mener målefeilene i oppgaven er differensielle, gi et eksempel på hvordan feilstrukturen kunne vært dersom målefeilene var ikke-differensielle.

Du skal altså gi et eksempel på en feilstruktur som er motsatt av det du har svart i 1d.

(4 poeng)

Svar: Et eksempel på differensielle målefeil er at spesifisiteten for den nye blodtrykksmålingen er 0,9 blant dem som drikker energidrikk daglig, men bare 0,7 blant dem som ikke drikker energidrikk daglig. Da vil sannsynligheten for feilklassifisering av blodtrykk avhenge av eksponering (energidrikk), og målefeilen er differensiell.

Et eksempel på ikke-differensielle målefeil er at sensitivitet og spesifisitet for blodtrykksmålingen er de samme i gruppen som drikker energidrikk daglig og i gruppen som ikke drikker energidrikk daglig (for eksempel sensitivitet 1,0 og spesifisitet 0,9 i begge grupper). Da er feilstrukturen uavhengig av eksponering.

1f) Drøft kort (maks 3 setninger) om resultatene basert på den nye målemetoden er tilstrekkelig valide til at du kan forsvare å bruke denne metoden hvis du skulle gjentatt undersøkelsen.

(4 poeng)

Svar: Moment som er relevante: OR på 1,25 er betydelig lavere enn 2,01 (selv om begge viser økt risiko), noe som taler imot å bruke den nye målemetoden. Samtidig vet vi ikke noe om konfidensintervallene rundt estimatene, eller hvor stor kostnadsbesparing den nye metoden gir.

1g) I oppgaven er både sensitivitet og spesifisitet til energidrikk-variabelen satt til 1,0. Drøft kort (1-3 setninger) i hvilken grad dette er realistisk.

(2 poeng)

Svar: Selvrapporterte kostholdsdata er svært sårbare for feil. Det er urealistisk at sensitivitet og spesifisitet for denne variabelen er 1,0.

Oppgave 2 (21 poeng)

2a) I en analyse av Sieri og medarbeidere var det overordnede målet å studere assosiasjonen mellom inntak av fett og risiko for brystkreft.

Basert på tallene i tabell 3 (vedlagt), hvordan vil du vurdere inntaket av mettet fett i dette utvalget?

(2 poeng)

Svar: E% fra mettet fett (vist i density modellen) ligger over anbefalt inntak (< 10 E%) i så godt som alle kvintilene.

2b) Sieri og medarbeidere valgte å justere analysen av assosiasjonen mellom inntak av mettet fett og risiko for brystkreft for inntak av energi. Er det et klokt valg? Begrunn svaret kort.

(3 poeng)

Svar: Ja. Energiinntaket er en konfunderende faktor; det samvarierer både med inntak av mettet fett (eksponering) og brystkreft (utfall).

2c) Forskerne delte kvinnene inn i kvintiler etter inntak av ulike typer fett. Se på tabell 3 igjen: Basert på resultatene av analysene der inntaket av mettet fett er delt inn i kvintiler, vil du si at det er en assosiasjon mellom inntak av mettet fett og risiko for brystkreft?

(6 poeng)

Svar: Kvinner i den øverste kvintilen av mettet fett hadde signifikant høyere risiko for brystkreft i ut fra standardmodellen (HR: 1,13; 95% CI 1.00-1.27), næringstetthetsmodellen (HR: 1.10; 95% CI 1.01-1.19) og dekomponeringsmodellen (HR: 1.11; 95% CI 1.00-1.23), men ikke i residualmodellen (HR: 1.06; 95% CI 0.98-1.16).

Trend-test viste signifikant positiv sammenheng for standardmodellen ($p = 0.04$) og dekomponeringsmodellen ($p = 0.045$).

2d) I tillegg til å analysere inntaket av mettet fett som kvintiler, undersøkte forskerne assosiasjonen mellom inntak av mettet fett og risiko for brystkreft når inntaket av mettet fett ble lagt inn i modellene som en kontinuerlig variabel. Hva viser resultatene fra disse modellene?

(3 poeng)

Svar: Når inntak av mettet fett ble analysert som en kontinuerlig variabel, var økt inntak assosiert med økt risiko i alle modellene (HR: 1,02; 95 % KI 1,00–1,04 for standard- og næringstetthetsmodellen; HR: 1,02; 95 % KI 1,00–1,03 for residual- og dekomponeringsmodellen).

2e) Den siste kolonnen i tabell 3 viser sammenhengen mellom inntak av mettet fett (som kontinuerlig variabel) og risiko for brystkreft etter at kostdataene er kalibrert. Hva er formålet med å kalibrere kostdataene?

(2 poeng)

Svar: Å ta hensyn til målefeil.

2f) Hva kalles den statistiske tilnærmingen som ofte blir brukt til kalibrering av kostdata i epidemiologiske studier, og som også er brukt i denne studien?

(2 poeng)

Svar: Regresjonskalibrering.

2g) Med tanke på sammenhengen mellom inntak av mettet fett og risiko for brystkreft: Hva viser resultatene fra modellene der man har kalibrert kostdata (siste kolonne) sammenliknet med modellene der man ikke har kalibrert kostdata (nest siste kolonne)?

(3 poeng)

Svar: Etter kalibrering forble HR-ene for inntak av mettet fett statistisk signifikante i de fleste modellene og var noe høyere enn estimatene fra modellene med kontinuerlig variabel basert på ikke-kalibrerte data.

TABLE 3

Multivariate-adjusted hazard ratios (HRs) (and 95% CIs) of invasive breast cancer risk associated with subtypes of fat intake by quintile (Q) in the cohort of the European Prospective Investigation into Cancer and Nutrition¹

	Intake					P for trend ²	Intake as continuous variable ³	Calibrated data (intake as continuous variable) ⁴
Fat subtype	Q1	Q2	Q3	Q4	Q5			
Saturated fat								
Standard ⁵								
Median (g/d)	16.2	22.6	27.9	34.1	45.0			
Cases/person-years (n)	1160/538 852	1341/545 846	1451/557 165	1536/571 949	1631/598 798			
HR	1	1.08 (0.99, 1.17) ⁶	1.12 (1.02, 1.22)	1.14 (1.03, 1.26)	1.13 (1.00, 1.27)	0.038	1.02 (1.00, 1.04)	1.03 (1.00, 1.07)
Residual ⁷								
Median (g/d)	21.1	26.1	29.3	32.6	38.1			
Cases/person-years (n)	1274/565 391	1342/547 796	1370/547 072	1454/560 043	1679/592 308			
HR	1	1.06 (0.98, 1.15)	1.06 (0.97, 1.15)	1.05 (0.97, 1.15)	1.06 (0.98, 1.16)	0.233	1.02 (1.00, 1.03)	1.03 (0.99, 1.07)
Density ⁸								
Median (% of energy/d)	9.9	12.3	13.9	15.5	18.2			
Cases/person-years (n)	1185/555 735	1374/547 967	1412/552 679	1474/565 272	1674/590 956			
HR	1	1.07 (0.99, 1.16)	1.06 (0.98, 1.15)	1.05 (0.96, 1.13)	1.10 (1.01, 1.19)	0.068	1.02 (1.00, 1.04)	1.04 (1.00, 1.07)
Partition ⁹								
Median (kcal/d)	145.8	203.3	250.9	306.5	404.6			
Cases/person-years (n)	1160/538 852	1341/545 852	1451/557 159	1536/571 961	1631/598 786			
HR	1	1.07 (0.98, 1.16)	1.11 (1.02, 1.21)	1.12 (1.02, 1.23)	1.11 (1.00, 1.23)	0.045	1.02 (1.00, 1.03)	1.03 (1.00, 1.06)
Monounsaturated fat								
Standard ⁵								
Median (g/d)	15.3	21.0	26.0	32.3	44.1			
Cases/person-years (n)	1273/525 283	1352/546 266	1515/568 799	1584/58 914	1395/583 116			
HR	1	0.98 (0.91, 1.07)	1.05 (0.95, 1.15)	1.07 (0.96, 1.18)	1.05 (0.92, 1.20)	0.254	1.02 (0.99, 1.04)	1.05 (1.00, 1.10)
Residual ⁷								
Median (g/d)	20.2	24.2	27.0	30.3	38.1			
Cases/person-years (n)	1310/536 795	1469/551 560	1499/569 583	1552/586 099	1289/568 573			
HR	1	1.09 (1.00, 1.18)	1.03 (0.95, 1.12)	1.10 (1.00, 1.20)	1.08 (0.98, 1.21)	0.150	1.02 (0.99, 1.04)	1.04 (0.99, 1.10)
Density ⁸								
Median (% of energy/d)	9.5	11.3	12.7	14.5	18.2			
Cases/person-years (n)	1291/532 384	1472/548 551	1483/570 103	1579/590 546	1294/571 026			
HR	1	1.07 (0.99, 1.15)	1.02 (0.95, 1.11)	1.06 (0.98, 1.15)	1.05 (0.96, 1.16)	0.323	1.02 (0.99, 1.04)	1.05 (1.00, 1.10)
Partition ⁹								
Median (kcal/d)	137.9	189.0	234.4	291.0	397.2			
Cases/person-years (n)	1273/525 283	1352/546 266	1515/568 799	1584/589 146	1395/583 116			
HR	1	0.98 (0.90, 1.06)	1.04 (0.95, 1.14)	1.06 (0.96, 1.17)	1.04 (0.93, 1.18)	0.253	1.01 (0.99, 1.03)	1.04 (1.00, 1.09)
Polyunsaturated fat								
Standard ⁵								
Median (g/d)	7.2	9.8	12.3	15.4	21.3			
Cases/person-years (n)	1403/550 518	1409/547 583	1376/555 864	1445/570 597	1486/588 046			
HR	1	1.00 (0.93, 1.08)	0.95 (0.88, 1.04)	0.95 (0.87, 1.04)	0.97 (0.88, 1.07)	0.372	0.99 (0.98, 1.01)	0.99 (0.95, 1.03)
Residual ⁷								
Median (g/d)	8.4	10.9	12.7	15.0	19.5			
Cases/person-years (n)	1523/572 752	1390/556 567	1377/552 245	1470/559 334	1359/571 712			
HR	1	0.97 (0.90, 1.06)	1.02 (0.94, 1.10)	1.00 (0.92, 1.08)	0.99 (0.91, 1.08)	0.967	0.99 (0.98, 1.01)	0.99 (0.96, 1.03)
Density ⁸								
Median (% of energy/d)	4.0	5.1	6.1	7.2	9.4			
Cases/person-years (n)	1533/573 440	1389/557 490	1374/552 199	1472/558 845	1351/570 636			
HR	1	0.99 (0.91, 1.07)	0.95 (0.88, 1.03)	0.99 (0.92, 1.07)	0.96 (0.88, 1.04)	0.390	0.99 (0.98, 1.01)	0.99 (0.95, 1.02)
Partition ⁹								
Median (kcal/d)	64.4	88.6	110.7	138.6	191.5			
Cases/person-years (n)	1403/550 518	1409/547 583	1376/555 864	1445/570 598	1486/588 046			
HR	1	1.00 (0.93, 1.08)	0.95 (0.88, 1.04)	0.95 (0.87, 1.04)	0.97 (0.88, 1.07)	0.370	0.99 (0.98, 1.01)	0.99 (0.95, 1.02)

¹ Values were stratified by age and center and were adjusted for educational attainment, smoking status, height, weight, alcohol intake, and menopausal status. *n* = 7119 breast cancer cases.² Test for linear trend was performed on median intake for each quintile.³ Values were log_{1.2} transformed (equivalent to a 20% increase).⁴ Calibrated data were obtained by linear regression models that compare observed nutrient questionnaire measurements with 24-h dietary recall.⁵ Standard models contain log-transformed fat subtype intake and nonalcohol energy intake.⁶ 95% CIs in parentheses (all such values).⁷ Residual models contain residual of regression of log-transformed fat subtype intake and nonalcohol energy intake.⁸ Density models contain log-transformed energy from fat subtypes and nonalcohol energy.⁹ Partition models contain log-transformed fat subtype intake and energy intake from other sources excluding alcohol.

Oppgave 3 (30 poeng)

Din kollega Line har fått tilgang til data fra en stor undersøkelse med kostholdsinformasjon som ble startet i 2015. Det ble brukt et detaljert semikvantitativt matvarefrekvensskjema med over 200 spørsmål samt vel 50 spørsmål om livsstilsfaktorer (røyking, vekt, fysisk aktivitet, søvn, sosiale relasjoner og familieforhold). Skjemaet var validert i den befolkningen det ble brukt og er basert på tidligere undersøkelser.

Skjemaet ble sendt til 250 000 tilfeldig utplukkete innbyggere i Danmark og Norge i alderen 50-69 år. Cirka 80 000 svarte, omtrent likt fordelt på hvert land. Etterpå er befolkningen blitt fulgt med koblinger mot ulike sykdomsrelaterte registre i begge land, mot norsk (og dansk) pasientregister, men også legemiddelregistrene og dødsårsaksregistrene. Fra pasientregisteret har de fått informasjon om diabetes type II.

Du blir bedt om å bidra i analysene av risikofaktorer for diabetes type II.

Koblingene er gjort første gang i 2015 og deretter i 2018, altså før de første GLP-1 medikamentene (semaglutid osv.) ble tatt i bruk. Dere finner 9 000 med diabetes type II i 2015 - og i 2018 er tallet i kohorten blitt 13 300.

Av økonomiske grunner er det ikke gjort noen kobling etter 2018. Det planlegges en ny kobling mot registre i 2027, men for å få midler, må forskerne vise hva de har funnet så langt.

3a) (4 poeng) Hva er insidensen av diabetes type II i studiepopulasjonen på disse tre årene (fra 2015 til 2018)?

SVAR: befolkning at risk: 80 000 minus de med prevalent DM i 2015 (9 000) = 71 000.

Nye tilfeller er 13 300 minus 9 000 = 4 300.

Insidens over 3-års perioden er dermed $4\,300 / 71\,000 = 0.0606$

Alternativt er årlig insidens en tredjedel av dette, så 0.0202 eller 2.02% per år.

3b) (4 poeng) Beskriv hvilke forutsetninger du gjorde for å beregne dette tallet.

SVAR: Dette er nå beregnet som i en lukket (statisk) kohort – altså at ingen har dødd eller flyttet ut av landene i mellomtiden. (Studenten bør forstå at ingen kan komme inn i kohorten, så ikke ta det med.) Studenten bør også komme med tilleggsforutsetninger her mhp. korrekthet av diagnostikk og registre.

3c) (4 poeng) Dere skal nå gjøre analyser for å beregne risikoen for å utvikle diabetes type II forbundet med ulike faktorer, og din oppgave er analyser på kosthold. Du vil se på enkeltfaktorer i kostholdet, men også ulike kostholdsmønstre. Hvilke konfunderende faktorer vil du ha med i analysen?

SVAR: Her avhenger jo svaret noe av hva eksponeringen er. Men generelt sett bør de tenke på totalt energiinntak (spesielt når vi kommer ned i enkeltfaktorer nedenfor). De bør også vite noe om mulige risikofaktorer for DM og få med i alle fall alder, kjønn, land og vekt. De må også vise

at de forstår at dette er risiko for å UTVIKLE DM, så de prevalente kasusene bør ut av analysen – dette skjønner de nok når de leser neste spørsmål og bør i alle fall håndtere der.

3d) (6 poeng) Lag en 2 x 2 tabell hvor du viser fordeling av en risikofaktor. Regn deretter ut effektestimatet for å utvikle diabetes forbundet med denne risikofaktoren basert på hvordan du har satt det opp i tabellen. (Det er ikke viktig hvilken risikofaktor du velger).

SVAR Her må de vise at de kan beregne i alle fall en kumulativ risikoratio (som de gjerne må kalle en insidensrate-ratio). Viktig at de her tar ut de prevalente kasusen fra tabellen sin.

Eksempel

	Syke	Alle
Ekspontert	3 000	35 000
Ikke eksponert	1 300	36 000

Risikoratioen i denne statiske kohorten blir da: $(3000/35\ 000 / 1300/36\ 000) = 0.0857/0.036 = 2,38$. (Det er helt OK om de har faktorer som både gir økt og redusert risiko, bare de kan regne det ut riktig.) (Det er ikke forventet at de legger inn forutsetninger og beregninger om en dynamisk kohort og regner ut persontid osv.)

3e) Du gjør analysene og finner en rekke signifikante risikoestimat. Her er noe av det du finner:

Kostholds faktor:	Effektestimat Relativ risiko (95% CI)
Sukkerholdig brus – 33cl (1 eller flere per dag vs. sjeldnere enn en gang pr md)	0.80 (0.70- 0.98)
Brus med kunstig søtstoff – 33cl (1 eller flere per dag vs. sjeldnere enn en gang pr md)	1.30 (0.90- 1.70)
Alkohol – et glass vin/én enhet (per enhet i uken)	0.92 (0.89-0.96)
Pizza – et stykke (2 eller flere per uke vs. sjeldnere enn en gang pr md)	1.90 (1.70-2.10)
Grønne grønnsaker - en porsjon (ca. 2 dl) (3 eller flere per uke vs. sjeldnere enn en gang pr md)	1.20 (1.01-1.41)

Basilikum - en neve

(3 eller flere per uke vs. sjeldnere enn en gang pr md)

1.70 (0.90- 2,90)

(4 poeng) Er du overrasket over disse resultatene? Hvis ja, hvorfor? Hvis nei, hvorfor tenker du dette er som forventet?

SVAR: Ja dette var ikke helt som forventet. Det kan tyde på at kostholdet er blitt endret pga. tidlige symptomer. For eksempel at de som har tidlige symptomer er begynt å redusere sukkerinntak og gå over til brus med kunstig søtstoff, redusert alkohol, økt grønne grønnsaker. Pizzainntaket (hvis representativt for andre prosesserte varer) er kanskje det eneste som tyder på at noe av kostholdet ikke er endret.

Basilikum-funnet var underlig – men henger for så vidt sammen med grønnsakene og endringer ellers. OG – de bør nevne at noen tilfeldige funn må forventes når man spør 200 spørsmål

3f) (4 poeng) Du skal gi råd til gruppen om fremtidige analyser. Hva vil du anbefale at de ser på etter 2027 koblingen?

SVAR: Gruppen bør være nøye med å ta ut folk med symptomer før de får sykdommen. Dette er ikke helt lett, men de bør i alle fall ta opp dette som en utfordring og foreslå noen løsninger (for eksempel ikke inkludere alle som er diagnostisert innen en viss tid etter utfyllt matvarefrekvensskjema, ta bort de som har brukt antidiabetes-medikamenter evt. GLP-1 tidligere osv.

De bør ta inn noe om GLP-1 bruk og hvordan de skal håndtere dette. Slike medikamenter er blitt brukt i stadig større grad fra 2018 av. Det kan være noen forskjeller mellom Dk og Norge her.

3g) (4 poeng) Foreslå en potensiell effektmodifikator du tenker de bør se på. Forklar hvorfor du mener denne er viktig - og hva du tror de vil finne.

SVAR: Her er et naturlig forslag land, gitt at dette er gjort i Danmark og Norge, men de kan godt komme opp med kjønn, alder, sosioøkonomi eller andre forslag. Hvis de for eksempel nevner land: Det er mulig at effekt av kosthold på utvikling av DM kan være annerledes i DK og Norge. Det er mange grunner til det, inklusivt mulig bruk av medisiner (GLP-1 osv.). I tillegg kan det være underliggende forskjeller som påvirker denne effekten – det kan være forskjeller i annet kosthold vi har som gjør at effekten av en kostholdsfaktor eller mønster er annerledes i ett av landene. Eller det kan være noe annet i miljøet eller helsevesenet eller mikrobiomet vårt. Det viktige her er at de viser at de forstår hva en effektmodifiserende variabel er.

Oppgave 4 (20 poeng)

Sammenhengen mellom potetinntak og risiko for død ble undersøkt i en kohortstudie publisert av forskere ved Avdeling for ernæringsvitenskap i 2024. Studien inkluderte 77 297 menn og kvinner bosatt i Norge. Informasjon om kosthold ble innhentet ved hjelp av et semikvantitativt spørreskjema.

De fleste deltakerne fylte ut sitt første kostholdsskjema i 1974, mens noen fylte ut skjemaet på et senere tidspunkt. Studiens sluttdato var 31/12-2018. Deltakerne ble fulgt fra tidspunktet de fylte ut sitt første kostholdsskjema til død uansett årsak, eller sensurering. Informasjon om emigrasjon og død ble hentet ved kobling mot komplette nasjonale registre.

Tabellen nedenfor viser et utsnitt fra en resultattabell i artikkelen.

Hazard ratios (95% confidence intervals) of all-cause mortality according to potato consumption (n=77,297)

	Weekly potato intake (number of potatoes/wk)		
	≤6	7–13	≥14
Person-years	253,392	1,096,628	1,083,811
All-cause mortality			
Deaths, <i>n</i>	2476	11,603	13,767
Incidence rate per 1000 person-years	9.77	10.58	12.70
Model 1 ³	1 (ref.)	0.92 (0.88, 0.96)	0.89 (0.85, 0.94)
Model 2 ⁴	1 (ref.)	0.94 (0.89, 0.98)	0.89 (0.85, 0.94)
Model 3 ⁵	1 (ref.)	0.93 (0.89, 0.97)	0.88 (0.84, 0.93)

³Cox regression adjusted for age, sex and energy intake (Model 1)

⁴Adjusted for covariates in Model 1 with further adjustment for body mass index, recreational and occupational physical activity, smoking status, marital status, attained education, and self-reported comorbidities.

⁵Adjusted for covariates in Model 2 with further adjustment for weekly servings of meat/processed meat, fish, fish products, and use of melted butter or margarine on meat and fish for dinner.

- a) Hva kalles den typen data som er analysert her? (2 poeng) Hvilken statistisk metode ble brukt for å analysere dataene? (1 poeng)

Svar:

Dette er levetidsdata (tid-til-hendelse data).

Cox regresjon.

- b) Sett opp uttrykket for regresjonsmodellen i modell 1. Det er to alternative formuleringer av modellen, og siden dere skriver i Word er det enklest å skrive den uttrykt ved hjelp av $\ln$ (3 poeng). Bruk b 'er eller β 'er. Forklar alle størrelsene som inngår i regresjonsmodellen (3 poeng).

Svar:

Cox regresjon:

$$\ln[h(t, x_1, x_2, x_3)] = \ln[h_0(t)] + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

eller

$$\ln[h(t, \text{alder}, \text{kjønn}, \text{energi})] = \ln[h_0(t)] + \beta_1 \cdot \text{alder} + \beta_2 \cdot \text{kjønn} + \beta_3 \cdot \text{energi}$$

eller

$$\ln[h(t)] = \ln[h_0(t)] + \beta_1 \cdot \text{alder} + \beta_2 \cdot \text{kj\o nn} + \beta_3 \cdot \text{energi}$$

$$[\text{alternativt } h(t, x_1, x_2, x_3) = h_0(t) \cdot e^{\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3} \text{ eller } h(t, \text{alder}, \text{kj\o nn}, \text{energi}) = h_0(t) \cdot e^{\beta_1 \cdot \text{alder} + \beta_2 \cdot \text{kj\o nn} + \beta_3 \cdot \text{energi}}]$$

$h(t, x_1, x_2, \dots, x_p)$ (eller $h(t)$) er hasardfunksjonen ved tid t

$h_0(t)$ er basis hasarden

$\beta_1, \beta_2, \beta_3$ er regresjonskoeffisienter

x_1, x_2, x_3 er uavhengige variabler/kovariater: alder, kj\o nn, energi

Det er helt greit \aa ikke bruke superscript og skrive f.eks. x1 og b1

- c) I analysene som ga resultatene presentert i tabellen er endepunktet d\o d uansett \aa rsak og det er to muligheter for sensurering av levetiden. Hvilke er disse? (2 poeng) Hva vil det si \aa sensurere levetiden? (3 poeng)

Svar:

Emigrasjon (individet meldte flytting til et annet land) og studiens sluttdato (31/12-2018).

Man vet at personer har overlevd fram til et bestemt tidspunkt, men ikke den faktiske tiden til d\o d. M.a.o. personer som ikke har opplevd hendelsen (d\o d) mens de ble observert, sensureres. Vi vet ikke den faktiske tiden til d\o d for disse. De bidrar med persontid frem til de sensureres (ved emigrasjonsdato eller studieslutt).

- d) Hvor mange regresjonsanalyser er utf\o rt for \aa lage tabellen ovenfor? (2 poeng)

Svar: 3

- e) Er det forskjell p\aa resultatene fra analysene med de ulike modellene i tabellen? (2 poeng)
Beskriv resultatene for modell 3 i tabellen, med estimater og konfidensintervall. (2 poeng)

Svar:

Modell 1, 2 og 3 ga veldig like estimater for hasard ratio med 95% konfidensintervaller.

Modell 3: De som spiste 7-13 poteter per uke eller ≥ 14 poteter/uke hadde signifikant lavere risiko for \aa d\o enn de som spiste ≤ 6 poteter/uke. Hasard ratio (95% konfidensintervall) var henholdsvis 0.93 (0.89, 0.97) og 0.88 (0.84, 0.93). Vi kan si signifikant fordi konfidensintervallene ikke inkluderer 1.

Oppgave 5 (5 poeng)

Metoder for å beskrive og analysere kostholdsmønstre kan deles inn i teoretiske (a priori) og data-drevne (a posteriori) tilnærminger.

a) Beskriv kort hva som skiller en teoretisk og en data-drevet tilnærming.

Svar: Teoretisk metoder (a priori) bygger på tidligere kunnskap/anbefalinger. Data-drevet metoder (a posteriori) bygger på statistisk modellering av de innsamlede kostdataene. (2 p)

b) Gi ett eksempel på en konkret metode eller score for hver av tilnærmingene.

Svar: Eksempler på teoretisk metode: indekser/scorer, men må nevne et konkret f.eks. New Nordic Diet (1 p). Eksempler på empirisk metode: PCA-analyse og clusteranalyse (1 p).

c) Nevn én fordel ved å analysere kostholdsmønstre fremfor enkeltstående matvarer eller næringsstoffer.

Svar: Å syntetisere store mengder data til noen få, men nyttige indikatorer, gjøre noe komplekst mere simpelt. (1 p)