



**EUNIVERSITETET
I OSLO**

Avdeling for ernæringsvitenskap

DET MEDISINSKE FAKULTET

Eksamen

ERN 4010

Ernæringsepidemiologi

Fredag 4. oktober 2024 kl 09:00-11:30

Oppgavesettet består av 5 oppgaver.

Poengfordeling mellom oppgavene:

Oppgave 1 – 13 poeng

Oppgave 2 – 30 poeng

Oppgave 3 – 32 poeng

Oppgave 4 – 20 poeng

Oppgave 5 – 5 poeng

I alt 100 poeng

Bruk kalkulator i Inspira.

Lykke til!

Oppgave 1

Du er vitenskapelig assistent i et stort epidemiologisk prosjekt som studerer risikoen for irritable tarmsyndrom (IBS) og inntak av sukker. Det er sendt ut et selvadministrert spørreskjema til 78 000 kvinner og menn i alderen 25-75 år. Av de inviterte har 35 000 personer svart på spørsmål om de har IBS (svarkategorier «ja» og «nei»), og 35 600 har svart på tilstrekkelig mange kostholdsspørsmål til at du kan beregne inntak av sukker. Av de som har svart på IBS-spørsmålet og har data for sukkerinntak har 34 500 gitt samtykke til å delta i undersøkelsen.

I alt 1700 personer har svart «ja» på spørsmålet om de har IBS. I utgangspunktet er sukkerinntaket en kontinuerlig variabel, men du dikotomiserer den i ≥ 50 gram/dag og < 50 gram/dag. Av de 1700 personene som har svart at de har IBS har 1000 personer et sukkerinntak ≥ 50 gram/dag. Blant de som ikke har IBS er det 12 800 som har et sukkerinntak på ≥ 50 gram/dag.

Dette gir en odds ratio på 2,23 for å ha IBS for de som har et sukkerinntak ≥ 50 gram/dag i forhold til de som har et daglig inntak av sukker < 50 gram.

- a) Basert på denne odds ratioen, hva vil du si om risikoen for å ha IBS når det daglige sukkerinntaket er ≥ 50 gram i forhold til om sukkerinntaket er < 50 gram/dag. **Maks 2 poeng**

Svar: OR viser at de som har et daglig sukkerinntak på ≥ 50 gram har mer enn dobbel så høy odds for IBS som de som har et daglig sukkerinntak < 50 gram.

- b) For å få vurdert kvaliteten på IBS-variabelen evaluerer du denne opp mot et høykvalitetsregister som har 100 % dekning av IBS (fiktivt register). Du vurderer opplysningene i registret som «gullstandard» for IBS forekomst. Når du sammenlikner de selvrapporterte opplysningene om IBS med det som er registrert i registeret, viser det seg at 10 % av deltakerne som har oppgitt ikke å ha IBS er oppført med dette i registeret. Motsvarende er det 5 % av deltakerne som har oppgitt å ha IBS som ikke står i registeret. Du har ikke mulighet til å validere kostdataene, og må anta at de er valide for alle deltakerne.

Sett opp en 2 x 2 tabell basert på de nye opplysningene om IBS og regn ut den nye odds ratioen.

Svar:

	Sukkerinntak ≥ 50 gram/dag	Sukkerinntak < 50 gram/dag	
Har IBS	2230	2665	4895
Har ikke IBS	11 570	18 035	29 605

	13 800	20 700	34 500
--	--------	--------	--------

OR = 1,30

Maks 8 poeng

- c) Er feilrapporteringen av IBS i spørreskjemaene differensiell eller ikke-differensiell? Begrunn svaret kort.

Svar: Ikke-differensiell. Graden av feilrapportering er lik hos de som har et sukkerinntak ≥ 50 gram/dag og de som har et sukkerinntak < 50 gram/dag. Målefeilstrukturen i IBS-variabelen er ikke avhengig av eksponeringen.

Maks 3 poeng

Oppgave 2

En svensk kollega er interessert i om vegetarianere er spesielt beskyttet mot tarmkreft. Han skal skrive en søknad til etisk komité for å kontakte alle i alderen 60-79 år som fikk tarmkreft i perioden 2020-2023 i tre områder i Sverige. Du skal hjelpe ham å designe studien.

- a) (2 poeng) Hvilket studiedesign vil du anbefale ham å bruke? Forklar svaret.

Kasus kontroll. Her med prevalente kasus (2020-2023).

Begrunnelse: Oppgaven gir ingen andre muligheter ettersom det ikke er beskrevet noen kohort hvor dette kan ses på. Kasus-kontroll egner seg ettersom tarmkreft er en relativt sjelden sykdom, og dette blir da et effektivt studiedesign.

- b) (3 poeng) Hvem skal han inkludere i studien? Forklar svaret.

Kasuser og kontroller. Bør defineres mhp kjønn og alder. Likeledes forklare hvor kasusene kommer fra og hvor kontrollene skal rekrutteres fra. Bør beskrive populasjonsbasert rekruttering av begge (registerbasert gjerne).

- c) (2 poeng) Hva er eksponeringsvariabel (variabler) og hva er utfallsvariabel?

Vegetarkosthold (bør forklare hvordan dette skal spørres om) er eksponeringsvariabel og utfall = tarmkreft

- d) (3 poeng) Beskriv hvordan han må gå frem for å få innhentet informasjonen i 2b.

Bør beskrive hvor kontroller skal hentes fra for å sørge for at kontroller kommer fra samme basepopulasjon som kasus. Deretter viktig å invitere kasus og kontroller til å delta og be dem svare på spørsmål om kosthold osv. FØR tarmkrefttilfellene ble syke. Gjerne beskrive tiltak for å sørge for høy responsrate.

- e) (2 poeng) Sett opp en enkel 2 x 2 tabell og vis hvordan du vil beregne effektmålet

	Ka	ktr
Eksp (vegetar)	A	B
Ikke eksp (ikke veg)	C	D

OR = ad/bc

- f) (3 poeng) Hva er mulige konfunderende faktorer?

Alder, kjønn, røyking, alkohol, mulig utdanning og inntekt, mulig også fødeland.

- g) (4 poeng) Hvordan kan du justere for de konfunderende faktorene?

Justere for så mye detaljer som mulig i multivariat logistisk regresjon

h) (5 poeng) Hvilke mulige seleksjonsfeil er du bekymret for i studien? Forklar hvorfor det kan inntreffe og hvordan du kan prøve å unngå det.

Svaret avhenger av svaret på d. Gitt at det er en populasjonsbasert gruppe som inviteres, og at de inviteres på en «nøytral» måte: en bekymring er særlig selektiv deltakelse, dvs. at det er særlig vegetarianere som deltar blant kontrollgruppen. Dette kan da gi en skjevhet (her tendere til å overestimere den beskyttende effekten av et vegetarkosthold). Kan prøve å unngå dette ved å ikke kunngjøre hypotesen (!), kan ellers være vanskelig å gjøre så mye med det i praksis med dette designet.

Kollegaen din utfører studien. Han holder nå på med analysene og lurere på hvordan han skal se på røyking samtidig med vegetarkosthold. Han tror de ikke-røykende vegetarianerne har lavest risiko og at de røykende kjøttspiserne har høyest risiko.

i) (3 poeng) Hva heter det epidemiologiske begrepet han her ser på?

Effekt modifikasjon (evt statistisk interaksjon).

j) (3 poeng) Dersom han ikke finner denne gradienten i risiko – hva tror du årsakene kan være?

At det ikke er noen interaksjon/effekt modifikasjon. Alternativt at det er residual konfundering (eller noen andre seleksjons- eller andre skjevheter for eksempel i den ene ekstreme kategorien) som medfører bias.

Oppgave 3

Sabir og medarbeidere har studert sammenhengen mellom inntak av protein i kosten og muskelmasse og muskelstyrke i Helseundersøkelsen i Hordaland (HUSK). Kostdata ble samlet inn ved hjelp av et omfattende, validert matvarefrekvensskjema (FFQ) på to tidspunkt (HUSK2 og HUSK3). Muskelstyrke ble målt ved håndgripestyrke (HGS), og muskelmasse ble estimert ved appendikulær skjelettmuskelmasse (ASMM) målt ved bioelektrisk impedans (BIA). Muskelmasse og muskelstyrke ble kun målt ved ett tidspunkt (HUSK3). Forskerne studerte sammenhengen mellom proteininntak og muskelmasse og muskelstyrke både i tverrsnittsanalyser (HUSK3) og i oppfølgingsanalyser (longitudinelt fra HUSK2 til HUSK3).

Tittelen på artikkelen er «*The association of isocaloric substitution of dietary protein in middle age with muscle mass and strength in old age: The Hordaland Health Study*».

3a) Hva menes med «isocaloric substitution of dietary protein»? (3 poeng)

Svar: At man bytter ut en mengde protein (f.eks. målt som g/d) med en mengde av andre energigivende næringsstoffer (g/d) med tilsvarende energiinnhold.

3b) Forklar kort forskjellen på en regresjonsanalyse der man har en substitueringsmodell og en regresjonsanalyse der man har en adderingsmodell. Bruk proteininntak, inntak av andre makronæringsstoffer og energiinntak til å forklare forskjellen. (7 poeng)

Svar: Substitueringsmodell: viser effekten når totalt energiinntak blir holdt konstant, slik at en økning i inntak av andre makronæringsstoffer (ikke protein) blir oppveid av en tilsvarende reduksjon i inntak av protein. Effekten av å bytte ut protein med andre energigivende næringsstoffer.

Adderingsmodell: viser den uavhengige effekten av å øke inntaket av protein, mens inntaket av de andre makronæringsstoffene (ikke protein) holdes konstant. Energiinntaket vil da nødvendigvis øke.

3c) To mye brukte modeller for energijustering i ernæringsepidemiologiske studier er den såkalt standardmodellen og residualmodellen. For hver av modellene oppgi om det er en substitueringsmodell eller en adderingsmodell. (3 poeng).

Svar: Begge er substitueringsmodeller.

3d) Se på vedlagte tabell 3. Alle påfølgende spørsmål i denne oppgaven omhandler resultatene for appendikulær skjelettmuskelmasse (ASMM).

Hva fant forskerne i tverrsnittsanalysene av ASMM når man byttet ut protein med fett, og ikke justerte for potensielle konfunderende faktorer? Skriv opp estimatene og gi en kort fortolkning av resultatene. (5 poeng)

Svar:

Menn: beta-estimat -0.099 95% CI (-0.175, -0.023), $p = 0.011$. Statistisk signifikant sammenheng. I tabell 3 står det feilaktig 0.023 (det rette er -0.023). Ingen blir trukket pga. denne skrivefeilen hos tidsskriftet.

Kvinner: beta-estimat -0.094 95% CI (-0.148, -0.040), $p < 0.001$. Statistisk signifikant sammenheng.

3e) Hva fant forskerne i tverrsnittsanalysene av ASMM når man byttet ut protein med fett, og justerte for potensielle konfunderende faktorer? Skriv opp estimatene og gi en kort fortolkning av resultatene.

Spilte det noen rolle om man justerte for potensielle konfunderende faktorer eller ikke? (5 poeng)

Svar:

Menn: beta-estimat -0.012 95% CI (-0.068, 0.044), $p = 0.672$. Ikke statistisk signifikant sammenheng.

Kvinner: beta-estimat -0.042 95% CI (-0.080, -0.004), $p = 0.032$. Statistisk signifikant sammenheng.

Etter justering for potensielle konfunderende faktorer var det ikke lengre signifikant sammenheng for menn. For kvinner var sammenhengen fremdeles signifikant, men svakere.

3f) Hva fant forskerne i de longitudinelle analysene av ASMM når man byttet ut protein med fett, og justerte for potensielle konfunderende faktorer? (5 poeng)

Svar:

Menn: beta-estimat -0.009 95% CI (-0.091, 0.073), $p = 0.828$. Ikke statistisk signifikant sammenheng.

Kvinner: beta-estimat 0.042 95% CI (-0.009, 0.093), $p = 0.109$. Ikke statistisk signifikant sammenheng.

I de longitudinelle analysene der man justerte for potensielle konfunderende faktorer var det ikke lenger signifikant sammenheng hverken for menn eller kvinner.

3g) Oppsummer sammenhengen mellom total protein og ASMM når totalt proteininntak blir substituert med fett. Baser svaret på tabell 3 og resultatene i de foregående spørsmålene. (4 poeng)

Svar: I multivariate tverrsnittsanalyser fant man en negativ assosiasjon mellom substituerings av totalt proteininntak med fett på ASMM for kvinner, men ikke for menn. I de longitudinelle multivariate analysene fant man ingen statistisk signifikant sammenheng verken for menn eller kvinner. Vite at man legger mer vekt på de longitudinelle analysene, enn på tverrsnittsanalysene.

Tabell 3.

Z. Sabir et al.

Current Developments in Nutrition 8 (2024) 102052

TABLE 3

Sex-specific cross-sectional and longitudinal associations of amount and source of protein with ASMM and HGS in the Hordaland Health Study. Estimates obtained from multiple linear regression analysis with substitution of macronutrients (E%).

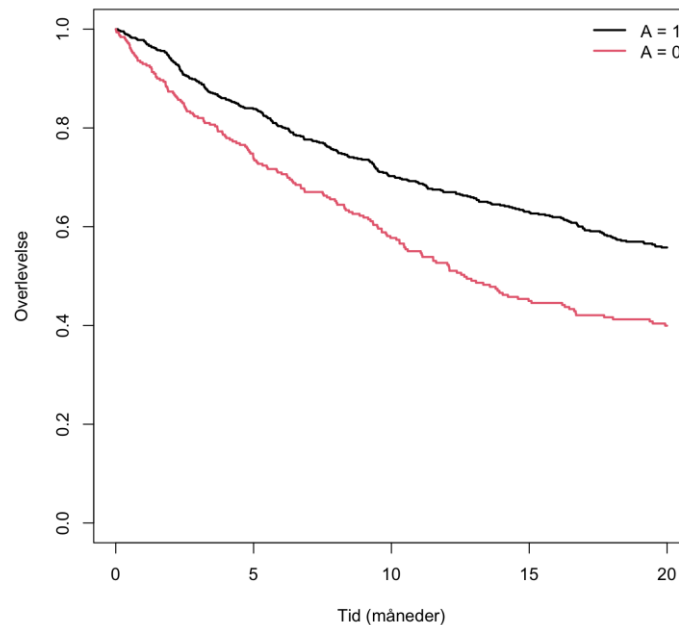
Appendicular skeletal muscle mass (kg)				Handgrip strength (kg)				
Cross-sectional associations								
Men (n = 632)		Women (n = 783)		Men (n = 621)		Women (n = 734)		
Estimate (95% CI)	P value	Estimate (95% CI)	P value	Estimate (95% CI)	P value	Estimate (95% CI)	P value	
Total protein								
Substituted by fat								
M1	-0.099 (-0.175, 0.023)	0.011	-0.094 (-0.148, -0.040)	<0.001	-0.048 (-0.259, 0.162)	0.651	0.018 (-0.101, 0.136)	0.770
M2 ^a	-0.012 (-0.068, 0.044)	0.672	-0.042 (-0.080, -0.004)	0.032	-0.035 (-0.247, 0.177)	0.746	0.030 (-0.091, 0.150)	0.627
Substituted by carbohydrates								
M1	-0.090 (-0.160, -0.020)	0.012	-0.111 (-0.162, -0.061)	<0.001	-0.061 (-0.259, 0.138)	0.550	-0.059 (-0.170, 0.052)	0.295
M1 ^a	0.000 (-0.051, 0.052)	0.986	-0.046 (-0.083, -0.010)	0.012	-0.059 (-0.260, 0.142)	0.563	-0.045 (-0.158, 0.068)	0.436
Animal protein								
Substituted by plant protein								
M1	-0.119 (-0.309, 0.070)	0.217	-0.015 (-0.137, 0.107)	0.806	0.096 (-0.430, 0.623)	0.720	0.129 (-0.141, 0.399)	0.348
M2 ^a	0.006 (-0.133, 0.144)	0.933	0.029 (-0.057, 0.115)	0.508	0.068 (-0.461, 0.598)	0.800	0.131 (-0.139, 0.402)	0.341
Substituted by fat								
M1	-0.111 (-0.188, -0.033)	0.005	-0.106 (-0.161, -0.051)	<0.001	-0.022 (-0.238, 0.194)	0.840	0.021 (-0.100, 0.142)	0.732
M2 ^a	-0.017 (-0.074, 0.040)	0.564	-0.043 (-0.082, -0.004)	0.031	-0.012 (-0.230, 0.206)	0.914	0.033 (-0.090, 0.156)	0.594
Substituted by carbohydrates								
M1	-0.094 (-0.166, -0.022)	0.010	-0.125 (-0.177, -0.074)	<0.001	-0.040 (-0.243, 0.162)	0.696	-0.059 (-0.172, 0.055)	0.311
M2 ^a	-0.005 (-0.058, 0.048)	0.855	-0.049 (-0.086, -0.012)	0.010	-0.040 (-0.245, 0.165)	0.701	-0.045 (-0.161, 0.072)	0.453
Longitudinal associations								
Men (n = 770)		Women (n = 1030)		Men (n = 751)		Women (n = 960)		
Estimate (95% CI)	P	Estimate (95% CI)	P	Estimate (95% CI)	P	Estimate (95% CI)	P	
Total protein								
Substituted by fat								
M1	-0.141 (-0.238, -0.045)	0.004	-0.058 (-0.122, 0.006)	0.073	-0.144 (-0.421, 0.133)	0.307	-0.162 (-0.301, -0.024)	0.021
M2 ^b	-0.009 (-0.091, 0.073)	0.828	0.042 (-0.009, 0.093)	0.109	-0.108 (-0.389, 0.173)	0.450	-0.116 (-0.255, 0.024)	0.103
Substituted by carbohydrates								
M1	-0.174 (-0.264, -0.085)	<0.001	-0.038 (-0.095, 0.020)	0.203	-0.117 (-0.372, 0.138)	0.367	-0.121 (-0.246, 0.005)	0.060
M2 ^b	-0.036 (-0.112, 0.040)	0.355	0.045 (-0.001, 0.091)	0.056	-0.091 (-0.351, 0.169)	0.492	-0.104 (-0.230, 0.022)	0.105
Animal protein								
Substituted by plant protein								
M1	-0.244 (-0.455, -0.032)	0.024	0.049 (-0.099, 0.197)	0.515	-0.210 (-0.837, 0.417)	0.510	0.091 (-0.234, 0.417)	0.582
M2 ^b	0.019 (-0.160, 0.197)	0.837	0.047 (-0.070, 0.164)	0.429	-0.167 (-0.801, 0.467)	0.604	0.062 (-0.260, 0.385)	0.705
Substituted by fat								
M1	-0.168 (-0.267, -0.069)	P<0.001	-0.096 (-0.163, -0.029)	0.005	-0.167 (-0.453, 0.119)	0.252	-0.178 (-0.324, -0.031)	0.018
M2 ^b	-0.015 (-0.099, 0.069)	0.723	0.023 (-0.031, 0.077)	0.394	-0.126 (-0.417, 0.165)	0.394	-0.125 (-0.274, 0.023)	0.098
Substituted by carbohydrates								
M1	-0.178 (-0.269, -0.087)	<0.001	-0.069 (-0.128, -0.010)	0.021	-0.119 (-0.378, 0.140)	0.368	-0.138 (-0.267, -0.009)	0.036
M2 ^b	-0.042 (-0.119, 0.035)	0.287	0.028 (-0.019, 0.076)	0.236	-0.093 (-0.357, 0.171)	0.490	-0.115 (-0.244, 0.015)	0.084

Abbreviations: M1, Unadjusted model; M2^a, Model adjusted for BMI (kg/m²), smoking (yes/no), and self-reported physical activity (inactive/active) in HUSK3; M2^b, Model adjusted for BMI (kg/m²), smoking (yes/no), and self-reported physical activity (inactive/active) in HUSK2.

Oppgave 4

Vi er interessert i å sammenligne levetider (tid til død, målt fra behandlingsstart) for pasienter som har fått en ny eksperimentell behandling ($A=1$) med pasienter som har fått tradisjonell/standard behandling ($A=0$). Det er 400 pasienter i hver gruppe.

Figuren under viser det resulterende Kaplan-Meier-plottet.



4a) Hvordan kan vi fortolke kurvene i figuren over?

(2 poeng)

Svar: Kurvene viser estimert sannsynlighet for å overleve over tid i de to gruppene.

4b) Ved tid = 10 måneder kan man lese av at (omtrentlig) verdi for de to kurvene er på 0,6 og 0,7. Hvor mange har dødd i hver gruppe og hvordan tolker du forskjellen mellom gruppene?

(2 poeng)

Svar: 160 og 120 personer har dødd i hhv. gruppe 0 og gruppe 1. Forskjellen i 10 måneders dødelighet (og overlevelse) mellom gruppene er på 10 prosentpoeng (eller 40 personer)

4c) Hvorfor trenger vi egne metoder for levetidsanalyse for å analysere denne typen data? Hva er den sentrale antagelsen for å analysere disse dataene med Kaplan-Meier-estimatoren som over?

(4 poeng)

Svar: På grunn av sensurering/sensurerte levetider. Vi må kunne anta uavhengig sensurering.

Sammen med Kaplan-Meier-figuren har det også blitt gjort en Logrank-test, hvor resultatet blir oppgitt å være en p-verdi på $<0,001$.

4 d) Hva forteller resultatet fra Logrank-testen?

(2 poeng)

Svar: Logrank-testen sier at det er en signifikant forskjell mellom de to gruppene (og gitt figuren, vil det si at gruppe 1 har bedre overlevelse enn gruppe 0).

I tillegg til en Logrank-test er det også blitt gjort en Cox-regresjon, hvor utfallet er tid til død og den eneste kovariaten er behandling (en binær variabel som kan ta verdiene 0 eller 1, ettersom hvilken behandlingsgruppe individene tilhører). I resultatet fra Cox-regresjonen kan man lese at $\exp(\beta)=1,6$ og at tilhørende p-verdi er $<0,001$.

4 e) Hva er fortolkningen til $\exp(\beta)$ og hva sier dette estimatet om forskjellen mellom gruppene over? Hvordan vil du sammenligne dette resultatet med resultatet fra Logrank-testen over? Hva er den sentrale antagelsen i denne estimeringen av $\exp(\beta)$?

(6 poeng)

Svar: Den oppgitte $\exp(\beta)$ er et estimat på hasardratioen mellom gruppene og vil her si at hasarden er 1,6 ganger så høy i gruppe 0 som i gruppe 1. P-verdien fra denne Cox-regresjonen svarer til å teste det samme som Logrank-testen. Den sentrale antagelsen er at hasardfunksjonene i de to gruppene er proporsjonale.

I tillegg til behandlingsvariabelen A har vi informasjon om også en kontinuerlig kovariat L, som er en score som sier noe om alvorlighetsgraden til pasientens sykdom, målt rett før behandlingstidspunktet.

4 f) Oppgi en grunn til hvorfor vi skulle ønske å også inkludere L i Cox-regresjonsmodellen.

(2 poeng)

Svar: For å justere for (tilfeldige eller systematiske) forskjeller mellom de to gruppene.

Når vi i en ny Cox-regresjon justerer for både variablene A og L, endrer estimatet av $\exp(\beta)$ for behandling seg, og er nå 1,3 i stedet for 1,6.

4 g) Hva sier dette om fordelingen til L-variabelen i de to gruppene?

(2 poeng)

Svar: Dette indikerer at L, som sier noe om risiko for død, er ulikt fordelt i de to behandlingsgruppene ved tid null (gruppe 0 har flere pasienter med høy alvorlighetsgrad).

Oppgave 5

5a) Hvilke muligheter gir det å analysere kosthold som kostholdsmønstre, fremfor analyser av enkelte næringsstoffer/matvarer? (1 p)

- Å syntetisere store mengder data til noen få, men nyttige indikatorer, gjøre noe komplekst mere simpelt. Beskrive kostholdsvaner.

5b) Nevn 2 metodiske utfordringer ved kostholdsmønstre som analysemetode. (2 p)

1 poeng per riktig utfordring, maks 2 poeng.

- Endringer over tid
- Ingen gullstandard, mange subjektive avgjørelser
- målefeil i kostholdsdata
- Validitet og reproduserbarhet

5c) Hvilken av de to hovedtilnærmingene innen analyser av kostholdsmønstre tilhører Principal Components Analysis (PCA)? (1p)

- Empirisk/data driven tilnærming/a posteriori

5d) Nevn én begrensning ved tradisjonell kostholds tilnærming. (1p)

1 poeng for en riktig begrensning, maks 1 poeng.

- Man tar ikke hensyn til de kompliserte interaksjoner det er mellom næringsstoffer/matvarer.
- Vanskelig å utforske deres unike effekt separat
- Effekten av ett enkelt næringsstoff/matvare kan være for liten til å kunne påvise effekt
- Analyse av et stort antall næringsstoffer/matvarer kan lede til at man finner tilfeldige sammenhenger.

- Oppgavesettet slutt-