

Biostatistik svar

Opgave 1

- a) Her er det meningen at man skal lave en chi-kvadrat test for at finde ud af om den er statistisk signifikant. Først findes de forventede værdier:

		Svulst type		total
		Ondartet	Godartet	
	0	21.4	2.6	24
Score for Seprase	1	20.5	2.5	23
immunohistokemi	2	20.5	2.5	23
	3	11.6	1.4	13
	Σ	74	9	83

Så skal vi finde chi-kvadrat statistikken ved at beregne summen:

$$\sum \frac{(O-E)^2}{E} = \frac{(17-21,4)^2}{21,4} + \frac{(7-2,6)^2}{2,6} + \frac{(22-20,5)^2}{20,5} + \frac{(1-2,5)^2}{2,5} + \frac{(22-20,5)^2}{20,5} + \frac{(1-2,5)^2}{2,5} + \frac{(13-11,6)^2}{11,6} + \frac{(0-1,4)^2}{1,4} = 11,9$$

Der er 3 frihedsgrader $(\#kolonner-1) * (\#rækker-1)$ og ved at opslag i tabellen herunder ses at $0,01 < p < 0,001$.

Forskerne har konkluderet at chi-testen er signifikant, og de konkluderede også at der er en statistisk signifikant sammenhæng/association mellem svulst type og Seprase tallet.

- b) Brugen af chi-kvadrat testen er ikke en god metode. Fire af de forventede værdier, svarende til 50%, er under 5. Der er derfor ingen garanti for at den beregnede sum følger en chi-kvadrat fordeling. Det brude reviewerne nok have set.
- c) Her fisker jeg efter fisher's exact test (morfar humor alarm). De kunne også have kollapsedet nogle af score grupperne.

Opgave 2

Her er der åbent for en gang snak.

Mange foreslog et eksperimentelt studie hvor man kontrollerer alkohol indtaget for forsøgspersonerne. Det er jo nok i virkeligheden ikke så nemt sikre at folk drikker det man beder dem om, og det er heller ikke etisk forsvarligt at bede folk drikke store mængder, hvis man tænker at lade studiet løbe over en længere periode.

Et observationsstudie ville være bedre.

Man kunne foreslå en tværsnitsundersøgelse (cross-sectional study) hvor man undersøger en gruppe personer én gang. Man kan så undersøge om folk der drikker meget rødvin har kræft hyppigere end folk der ikke drikker meget rødvin. Sådanne studies er relativt billigt, men det giver ingen information om den tidlige udvikling. Man skal også passe på confounding factors. Det kan fx være et folk der drikke meget rødvin også spiser meget ost og at det er ost, der er kræftfremkaldende (find selv på flere).

Man kunne foreslå et kohortestudie (cohort study) hvor en stor gruppe raske personer udvælges. Deres alkohol vaner og alt andet der kunne have interesse kortlægges og personerne følges over tid. Dette er også et prospective design fordi personerne inkluderes inden de har sygdommen. Det er også et longitudinalt studie, da personerne undersøges mere end én gang. Dette design er meget dyrt og tidskrævende, men det giver en masse information. Det kan være vanskeligt at analysere store mængder data for et cohort study. Factorial ANOVA med within subject kan være en mulighed, men ofte kan man også få noget ud af at lave data-mining. Man kan også lave noget survival analyse hen ad vejen.

Man kunne også foreslå en undersøgelse med kontrolpersoner (case-control study), men det er nok ikke den bedste løsning. Her skulle man finde en person der drikker meget rødvin, og en der i alle aspekter ligner denne person, men blot ikke drikker rødvin. Sådanne par skulle der findes en masse af og de skulle følges over tid for at se om de udvikler kræft. Her ligger survival analysen lige til højrebænet. Men hvad med personer der drikker lidt rødvin, og hvor meget er meget rødvin.

I de ovennævnte studier er der en række overvejelser man skal gøre sig om rekruttering. Informeret samtykke er altid nødvendigt. Randomisering er nok ikke mulig. Det er ikke rigtigt muligt at blinde deltagerne i studiet, hvilket nok bringer os til den største vanskelighed. Hvordan ved vi at personerne er ærlige (over for sig selv) om deres alkoholforbrug? Yderligere vanskeliggør det data analysen at alkoholforbruget formegentligt ikke er konstant over tid.

Mange andre ting kunne diskuteres for at hente point hjem.

Opgave 3

Det anerkendte videnskabelige tidsskrift *Circulation Journal* bragte tidligere i år en artikel hvor en stor gruppe forskere havde undersøgt betydningen af et højt plasma niveau af dopamine for patienter med problemer med kranspulsåren. Patienter var opdelt i 2 grupper, en med højt dopamine niveau og en med lavt. Pudsigt nok var antallet af patienter i hver gruppe ens; 105. Forskerne lavede en overlevelsesanalyse af perioden indtil patienterne eventuelt oplevede et nærmere defineret kardiovaskulært tilfælde. Alle patienter blev fulgt i hele perioden på 36 måneder. Se figuren herunder.

- Kaplan-Meier survival curve. Kurven 'opdateres' så snart der er et event, ikke ved bestemt tidsintervaller som ved actual/life table grafer
- Det betyder at der er en signifikant forskel mellem de to kurver og at signifikans niveauet er 0,001

Der er i opgave ikke givet nogle oplysninger om der er dropouts, altså personer der er blevet inkluderet senere end forsøgsstart, og der for ikke har været med i 36 md, eller folk der er droppet ud. Men lad os antage at der ikke er dropouts i datasættet.

- I lav-dopamine gruppen er der ca 98% der overlever de først 12 måneder. Tilsvarende er der 80% for høj-dopamine gruppen
- Den mediane overlevelsestid er den tid hvorefter 50% er døde. Den kan man ikke aflæse på disse kurver, før forsøget har løbet noget længere. Altså et lille trick spørgsmål.
- Her aflæser man blot hvor mange procent der er tilbage efter 36md og ganger med 105, som var med fra starten. Igen under antagelsen af at der ikke er dropouts.
Lav-dopamine: $0,91 \cdot 105 \sim 96$ personer
Høj dopamine: $0,72 \cdot 105 \sim 76$ personer