

Viktige modeller og begreper i statistikk

Når du skal lese forskningsartikler, kan det være nyttig at du kjenner navnet på noen viktige modeller og begreper:

Forfattere

Hugo Hammer
førsteamanuensis

Kjetil Gundro Brurberg
Førsteamanuensis
Senter for kunnskapsbasert praksis, HiB og forsker ved Nasjonalt kunnskapssenter for
helsetjenesten

Sykepleien Forskning 2014 9(1)(84-88)
DOI: <https://doi.org/10.4220/sykepleief.2014.0075>

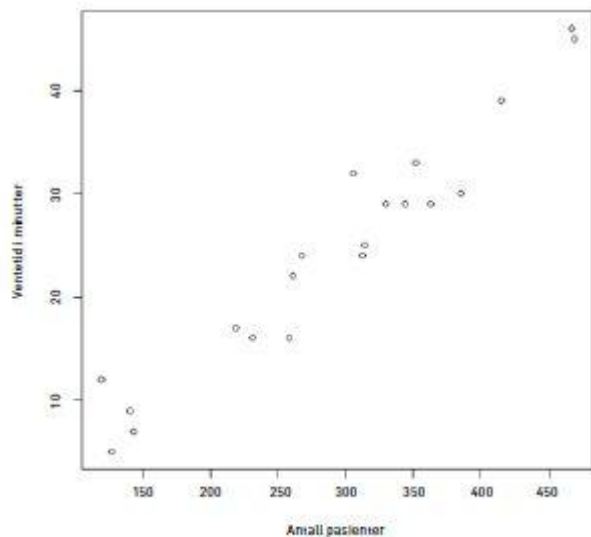
Her vil vi se på enkelte begreper og modeller i statistikk som du fra tid til annen vil støte på når du leser forskningsartikler.

TABELL 1: Antallet besøkende pasienter og gjennomsnittlig ventetid i minutter (fiktive data).

Dag	Antallet besøkende	Gjennomsnittlig ventetid
Dag 1	306	32
Dag 2	330	29
Dag 3	344	29
Dag 4	363	29
Dag 5	313	24
Dag 6	268	24
Dag 7	386	30
Dag 8	140	9
Dag 9	219	17
Dag 10	416	39
Dag 11	258	16
Dag 12	127	5
Dag 13	119	12
Dag 14	352	33
Dag 15	231	16
Dag 16	314	25
Dag 17	466	46
Dag 18	468	45
Dag 19	261	22
Dag 20	143	7

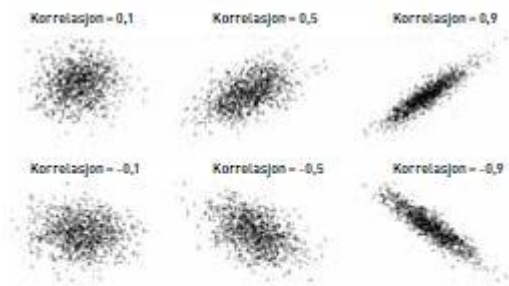
Korrelasjon

Utforskning av sammenhenger står svært sentralt i statistikken, og vi har tidligere sett på sammenhengen mellom statiner og kolesterolreduksjon (1). Tabell 1 viser antall besøkende på et akuttmottak sammen med gjennomsnittlig ventetid per pasient (fiktive data). Av tabellen ser vi sterke indikasjoner på en sammenheng mellom antall besøkende og ventetid. For dagene 10, 17 og 18 er det over 400 besøkende, og gjennomsnittlig ventetid er nærmere tre kvarter. På dag 8, 12 og 20 er det få besøk, og ventetiden er under ti minutter. Sammenhengen blir enda lettere å se når vi plotter data som vist i figur 1.



Figur 1: Sammenhengen mellom antallet besøkende pasienter på akuttmottak og gjennomsnittlig ventetiden per pasient.

Det går an å regne på graden av samsvar mellom to variabler. Et mye brukt mål er lineær korrelasjon (Pearson korrelasjon), et korrelasjonsmål som alltid ligger mellom -1 og 1. Dess nærmere korrelasjonen er -1 eller 1, dess sterkere er sammenhengen mellom de to variablene. Fortegnet viser retningen på sammenhengen. For data i figur 1 er korrelasjonsmålet 0,967 som betyr at vi har en svært sterk sammenheng mellom antall besøk og ventetid. Vi sier at korrelasjonen er positiv fordi flere besøk gir økt ventetid. Et eksempel på negativ korrelasjon kan være sammenhengen mellom antallet timer en person trener kondisjon og tiden han/hun trenger for å løpe 3000 meter. Korrelasjonen er negativ idet økt treningsmengde fører til at tiden man bruker går ned. Figur 2 illustrerer hvordan data kan se ut for ulike korrelasjoner. Ventetidsdataene i figur 1 likner mest på dataene oppe til høyre i figur 2.



Figur 2: Eksempler på data ved ulike korrelasjoner. Sterkere korrelasjon mot høyre, positiv korrelasjon øverst og negativ korrelasjon nederst.

Korrelasjon er ikke kausalitet

Når vi observerer en sammenheng (signifikant korrelasjon) mellom to variable er det lett å tro at det finnes årsakssammenhenger, men dette kan føre til alvorlige feilslutninger. Flere studier har for eksempel vist at personer som røyker mye har en tendens til å skåre dårligere på intelligens tester enn personer som røyker lite eller ingenting (4). En person som røyker vil kanskje være fristet til å konkludere med at intelligensen vil øke i det øyeblikket røyken legges på hylla, mon det? Forklaringer for sammenhengen kan for eksempel være at personer med høyere intelligens er mindre fristet til å begynne å røyke, eller at sosioøkonomisk status henger sammen med både økt risiko for røykestart og redusert intelligensskår. Hvis en av disse to forklaringene er årsaken, hjelper det neppe mye på intelligensen å slutte å røyke. Korrelasjonsstudier forteller altså ikke hva som forårsaker funnene. Til det trengs andre typer observasjoner og analyser.

Lineær regresjon

Lineær regresjon støter du oftest på i studier der forskerne observerer en rekke individer over tid og senere utfører statistiske analyser. Vi kan for eksempel være interessert i å studere sammenhengen mellom røyking og lungefunksjon. Siden vi ikke kan lage en studie der deltakerne i en gruppe pålegges å røyke, må vi nøye oss med observasjonsdata av lungefunksjon målt i en gruppe mennesker bestående av både røykere og ikke-røykere. Da kan det være fristende å sammenlikne lungefunksjon i de to gruppene direkte, men som vi skal se kan en slik tilnærming gi svært gale resultater.

TABELL 2: Lungefunksjonsmål (FEV) målt i liter og andre data for et utvalg av pasientene som opprinnelig ble publisert i [6]

Pasient	FEV	Alder	Høyde	Kjønn	Røyker?
1	1,708	9	145	K	Nei
2	1,735	8	137	M	Nei
3	0,839	4	122	K	Nei
4	1,776	5	130	M	Nei
5	1,703	8	138	M	Nei
6	1,577	4	124	K	Nei
7	3,816	13	161	K	Nei
8	2,236	14	168	K	Ja
9	3,411	11	161	K	Nei
10	4,309	14	175	M	Ja
11	2,691	10	170	K	Nei
12	4,271	14	184	M	Nei
13	2,387	11	154	M	Nei
14	3,120	15	154	K	Ja
15	4,086	18	170	M	Ja

Tabell 2 viser et utvalg av data som opprinnelig ble analysert i (5). FEV er et lungefunksjonsmål som viser hvor mye luft en pasient evner å puste ut i løpet av et sekund. I tillegg til FEV har vi for hvert barn/ungdom informasjon om alder, høyde, kjønn og røykestatus. Totalt ble det gjennomført målinger for 654 barn og unge mellom 3 og 19 år. Hvis vi beregner korrelasjonen mellom FEV og røykestatus får vi 0,25 med 95 prosent konfidensintervall fra 0,17 til 0,32. Ifølge disse resultatene har barn/unge som røyker bedre lungefunksjon enn de som ikke røyker. Et slikt resultat er overraskende og skyldes at andre faktorer lurar i kulissene. Videre analyser viser nemlig at det er en korrelasjon både mellom alder og FEV og mellom alder og røyking. Det betyr – ikke overraskende – at lungefunksjonen bedres etter hvert som barna vokser, og at det i all hovedsak er de eldste i datasettet som røyker. Den egentlige (og negative) sammenhengen mellom røyking og lungefunksjon er totalt overskygget av den sterke positive sammenhengen mellom alder og FEV. Vi sier at alder er en konfunderende (confounding) variabel for sammenhengen mellom FEV og røyking.

TABELL 3: Resultater av en regresjonsanalyse av FEV data slik de ofte presenteres i artikler

	Estimat	Standardfeil	T-verdi	p-verdi
Skjæringspunkt	-4,5442	0,232	-19,58	<0,001
Alder	0,0655	0,009	6,90	<0,001
Høyde	0,0410	0,002	21,90	<0,001
Kjønn [M]	0,1571	0,033	4,31	<0,001
Røyker [ja]	-0,0872	0,059	1,47	0,141

Ved bruk av lineær regresjon kan man i større grad isolere effekten av en og en faktor. Vi ønsker nå å bruke lineær regresjon til å undersøke hvilke faktorer som kan forutse (predikere) lungefunksjon (FEV). Vi sier at FEV er den avhengige variabelen i analysen. Faktorene vi bruker til å predikere FEV, her vil det si alder, høyde, kjønn og røykevaner, kalles uavhengige variabler. Tabell 3 viser resultatet av en regresjonsanalyse slik de gjerne presenteres i forskningsartikler. Lengst til høyre i tabellen ser vi en kolonne med p-verdier, og vi ser at alder, høyde og kjønn har svært små p-verdier. Det betyr at vi har en sterk sammenheng mellom FEV og disse faktorene. I kolonnen «Estimat» ser vi at verdiene er positive for alle tre variablene, som betyr at vi har en positiv sammenheng mellom FEV og disse variablene, for eksempel at FEV blir høyere med alder. Verdien 0,0655 sier oss at hvis vi sammenlikner to barn/unge med ett års aldersforskjell, men som har samme høyde, kjønn og røykevaner, vil FEV i gjennomsnitt være 0,0655 liter høyere hos den eldste. Vi har altså isolert effekten av alder. Tilsvarende ser vi at en gutt vil ha en FEV som i gjennomsnitt er 0,157 liter høyere enn ei jente så lenge de to har samme alder, høyde og røykevaner. En person som røyker vil derimot ha en FEV som i gjennomsnitt er 0,0872 liter lavere enn eller lik en ikke-røyker. Rett nok er ikke denne sammenhengen statistisk signifikant (p-verdi 0,141), men dette resultatet er mer i tråd med det vi vet om røykingens negative effekter enn da vi kun så på korrelasjonen mellom røyking og FEV. Vi kan si at regresjonsanalysen har hjulpet oss med å kontrollere for alder, høyde og kjønn.

Logistisk regresjon

Vi har sett hvordan vi kan bruke lineær regresjon å finne ut hvilke faktorer som predikerer lungefunksjon i form av FEV. FEV er det vi kaller et kontinuerlig mål (3). Logistisk regresjon er en nær slektning av lineær regresjon som brukes når den avhengige variabelen ikke er kontinuerlig, men består av kategorier som død versus levende. Ved hjelp av logistisk regresjon kan vi eksempelvis beregne hvordan faktorer som kjønn og alder (uavhengige variable) påvirker sannsynligheten for overlevelse (avhengig variabel). I likhet med lineær regresjon kan vi bruke logistisk regresjon til å isolere ulike effekter og kontrollere for forskjeller mellom grupper.

ANOVA

Vi har tidligere vist hvordan statistikk kan brukes til å sammenlikne to grupper, for eksempel for å påvise forskjeller mellom statiner og placebo (1). Det er lett å forestille seg naturlige utvidelser av et slikt eksperiment som involverer mer enn to grupper. Kanskje ønsker vi å sammenlikne effekten av ulike doseringer eller medisiner mot hverandre. Når vi har behov for å sammenlikne flere enn to grupper, er ANOVA (variansanalyse) et mye brukt verktøy. I variansanalyser er det, som navnet tilsier, den observerte variansen og ikke gjennomsnittsverdier som danner utgangspunkt for analysene.

TABELL 4: Gruppeinndeling ved toveis ANOVA

	Placebo	Fluvastatin
Treningsprogram	Gruppe 1	Gruppe 2
Ingen trening	Gruppe 3	Gruppe 4

Ved bruk av ANOVA kan vi altså studere effekten av flere faktorer samtidig. Tenk at vi ønsker å studere den kombinerte effekten av statiner og trening på reduksjon av kolesterol. For å undersøke dette kan vi lage oss et oppsett der forsøkspersonene plasseres tilfeldig i fire grupper (tabell 4). Vi ser at pasientene i gruppe 2 får statiner i tillegg til et spesifikt treningsprogram. Personene i gruppe 3 får placebomedisin og ikke noe treningsprogram. Ved bruk av ANOVA kan vi nå undersøke effekten av statiner og trening, og samtidig kan vi finne ut om kombinasjonen av statiner og trening har ytterligere effekt enn statiner eller trening alene. På tilsvarende måte som når vi sammenlikner to grupper (2), er signifikans, p-verdi og konfidensintervall viktige tall å se etter i artikler hvor ANOVA benyttes.

Ikke-parametriske tester

Vi har tidligere sett eksempler på sammenlikning av to grupper (2), og i denne artikkelen har vi beskrevet hvordan vi kan sammenlikne flere grupper (ANOVA). Disse testene baserer seg på antakelsen om at dataene i de ulike gruppene er tilnærmet normalfordelte. I mange situasjoner kan vi ikke gjøre denne antakelsen om dataene. Heldigvis finnes det tester for disse situasjonene også. Når du skal lese forskningsartikler, kan det være nyttig at du kjenner navnet på de mest vanlige testene. Skal to grupper sammenliknes er det vanlig å benytte Mann–Whitney U-test som også ofte refereres til som Wilcoxon rank-sum-test eller Wilcoxon-Mann-Whitney-test. Ved sammenlikning av mer enn to grupper benyttes vanligvis Kruskal-Wallis test. Også for disse testene presenteres resultatene i form av p-verdier, så det er nyttig å se etter disse verdiene i teksten.

Referanser

1. Brurberg KG, Hammer HL. Hvordan sammenlikne statistisk? Sykepleien Forskning. 2013;8:174–7.

2. Brurberg KG, Hammer HL. Hypotesetesting. Sykepleien Forskning. 2013;8:267–9.
3. Brurberg KG, Hammer HL. Variabeltyper og dikotome effektmål. Sykepleien Forskning 2013; 8:372–4.
4. Weiser M, Zarka S, Werbeloff N, Kravitz E, Lubin G. Cognitive test scores in male adolescent cigarette smokers compared to non-smokers: a population-based study. *Addiction* 2009;105:358–63.
5. Tager, IB, Weiss, ST, Rosner, B, & Speizer, FE. Effect of parental cigarette smoking on pulmonary function in children. *American Journal of Epidemiology*. 1979;110:15–26.