

WEBVTT

00:00.000 --> 00:02.000

'Nakne Tall'

00:02.000 --> 00:11.000

Med tall og statistisk analyse kan vi regne ut om det virkelig er en sammenheng mellom to størrelser eller fenomener i den virkelige verden.

00:11.000 --> 00:16.000

Er det en sammenheng mellom spordybde i bildekkene og bremselengden i regnvær,

00:16.000 --> 00:20.000

mellom høyde og skostørrelse eller prisen på en bolig i en bil?

00:20.000 --> 00:22.000

Hva er tall?

00:22.000 --> 00:27.000

Er det en sammenheng mellom spordybde i bildekkene og bremselengden i regnvær,

00:27.000 --> 00:33.000

mellom høyde og skostørrelse eller prisen på en bolig og hvor mange soverom boligen har?

00:33.000 --> 00:36.000

Dette ordet 'sammenheng' er det lett å bli forført av.

00:36.000 --> 00:41.000

Hjernen vår jakter hele tiden på sammenhenger, årsak og virkning,

00:41.000 --> 00:45.000

men fra statistikkens ståsted er det ikke det ordet sammenheng betyr.

00:45.000 --> 00:50.000

Det er forskjell på en tallmessig sammenheng og en årsakssammenheng.

00:50.000 --> 00:53.000

Jeg heter Jo Røislien og er statistikker.

00:53.000 --> 00:58.000

Og et lite kott innerst i lokalene til forskningsnettstedet forskning.no,

00:58.000 --> 01:04.000

i en gammel murbygning ved Akerselva, sitter molekylærbiolog og forskningsjournalist Eldrid Borgan.

01:05.000 --> 01:08.000

Hva tenker du om å bruke tall for å forstå verden?

01:08.000 --> 01:14.000

Vi som jobber i forskning.no baserer oss på det nesten hele tiden.

01:14.000 --> 01:19.000

Vi slår det opp som store nyheter hver gang noen forskere har funnet noen nye tall,

01:19.000 --> 01:22.000

som sier at sånn her er verden.

01:22.000 --> 01:30.000

Vin er sunt, vin er usunt, kaffe er vidunderlig, men ikke drikk for mye.

01:30.000 --> 01:33.000

Dette er vårt levebrød – tall.

01:33.000 --> 01:40.000

Det matematiske ordet korrelasjon og det beslektede ordet kausalitet er ikke synonymmer.

01:40.000 --> 01:44.000

Ordene ligner kanskje på hverandre, men de betyr ikke det samme.

01:44.000 --> 01:51.000

Jeg har ringt mine fire statistikkervenner og spurt dem om hva som er forskjellen på korrelasjon og kausalitet.

01:51.000 --> 02:00.000

Forskjellen på korrelasjon og kausalitet er at vi i korrelasjon ser på fenomener eller variabler som varierer sammen,

02:00.000 --> 02:04.000

uten at vi egentlig sier noe om at det ene faktisk fører til det andre.

02:04.000 --> 02:08.000

Korrelasjon er bare generelt at det er sammenheng mellom variabler,

02:08.000 --> 02:12.000

at det er sammenheng mellom skoene og høyden til folk.

02:12.000 --> 02:17.000

Kausalitet er at det i tillegg til sammenheng så er det årsak-virkning.

02:17.000 --> 02:24.000

Når vi snakker om kausalitet, så er vi opptatt av å finne et sånt årsaks-virknings-forhold mellom to fenomener.

02:24.000 --> 02:29.000

Vi er interessert i hvordan endringen i det ene fenomenet fører til en endring i noe annet.

02:29.000 --> 02:35.000

Det er mange som har gått i den fella og konkluderer at det er årsak-virkning,

02:35.000 --> 02:40.000

når det egentlig bare er en sammenheng som er forklart av andre mekanismer.

02:41.000 --> 02:45.000

Ja... hvis jeg er lenge ute i sola, så blir jeg solbrent.

02:45.000 --> 02:49.000

Det er på en måte en årsak og effekt. Da har vi kausalitet.

02:49.000 --> 02:52.000

Det er det også med bakterier eller virus som fører til sykdom,

02:52.000 --> 02:55.000

så det er en slags årsakssammenheng, og det er også kausalitet.

02:55.000 --> 03:02.000

Jeg har et eksempel der jeg har gått og sett på karakterer til elever i en skoleklasse.

03:02.000 --> 03:09.000

Jeg har til det gått og kikket inn i hjemmene deres og sett hvor mange bøker de har der.

03:09.000 --> 03:13.000

Så hadde jeg funnet ut at de elevene som har mange bøker hjemme,

03:13.000 --> 03:16.000

de fikk bedre karakterer enn de som har få.

03:16.000 --> 03:20.000

Så det er en samvariasjon mellom antall bøker og karakter.

03:20.000 --> 03:25.000

Men er det fordi alle elevene hadde lest alle disse bøkene hjemme,

03:25.000 --> 03:29.000

eller var det kanskje for at foreldrene i de hjemmene der man hadde bøker,

03:29.000 --> 03:33.000

var kanskje flinkere til å gi barna gode studievaner?

03:33.000 --> 03:37.000

Så var det ikke akkurat det at det var mange bøker hjemme som var årsaken til de gode karakterene,

03:37.000 --> 03:40.000

så var det ikke en kausal sammenheng.

03:40.000 --> 03:43.000

Så når vi har bare samvariasjon og ikke kausalitet,

03:43.000 --> 03:49.000

så kan det være en tredje ting, en tredje faktor, som kan påvirke begge egenskapene vi studerer.

03:49.000 --> 03:53.000

Korrelasjon er altså tallmessig sammenheng mellom to variabler,

03:53.000 --> 03:56.000

mens kausalitet er årsaks-sammenhengen.

03:56.000 --> 04:03.000

Da tror vi at endringen i verdiene til en ene er grunnen til at den andre endrer verdiene sine.

04:04.000 --> 04:11.000

Eldrid Borgan er molekylærbiolog, har jobbet som programleder og manusforfatter for vitenskapsprogrammer på TV,

04:11.000 --> 04:14.000

og er nå forskningsjournalist hos Forskning.no.

04:15.000 --> 04:21.000

Når vi snakker om at det er en sammenheng, så er det ofte statistiske ord som korrelasjon og sånt som ligger i bånd,

04:21.000 --> 04:23.000

at vi har regnet ut en sammenheng.

04:23.000 --> 04:26.000

Men det er forskjell på sammenheng og årsaksammenheng.

04:27.000 --> 04:31.000

Ja, og det der er jo så lett å gå i baret på.

04:31.000 --> 04:39.000

Det er jo ikke bare vi journalister, tenker jeg, men politikere, noen ganger ikke med vilje, noen ganger kanskje med vilje.

04:39.000 --> 04:44.000

Og alle du snakker med gjør den sammenblandingen.

04:44.000 --> 04:49.000

Men også forskere, som gjør jobben vår i forskning.no litt ekstra vanskelig,

04:49.000 --> 05:00.000

fordi forskere kan også finne på å selge inn at grovbrød beskytter mot hjerteinfarkt, slag, tidlig død.

05:00.000 --> 05:02.000

Og så, oi, det er jo fantastisk!

05:02.000 --> 05:05.000

Men, ikke sant, hvordan vet de at det er en årsak?

05:05.000 --> 05:13.000

Har de på en måte latt folk spise grovbrød i 30 år, og noen spiser loff i 30 år, og så sjekket hva som skjedde?

05:13.000 --> 05:17.000

Nei, det har de ikke, ikke sant? Det er veldig sjelden man får til sånne eksperimenter.

05:17.000 --> 05:22.000

Så det er så lett å gå i baret der, altså.

05:22.000 --> 05:28.000

Hvis det virker sant, så er det lettere å trekke sammenhengen mellom korrelasjon og kausalitet.

05:28.000 --> 05:34.000

Jeg tenker på det fordi jeg som bor i Oslo, det er klart det er mye snakk om forskjell på østkanten og vestkanten i Oslo.

05:34.000 --> 05:37.000

Gjennomsnittsinntektene er høyere på vestkanten enn på østkanten.

05:37.000 --> 05:41.000

Det er ulike sosiodemografiske grupper og litt sånt.

05:41.000 --> 05:46.000

Men hvis du flytter fra østkanten i Oslo til vestkanten, så får du jo ikke høyere lønn.

05:46.000 --> 05:48.000

Nei, ikke sant?

05:48.000 --> 05:56.000

Men det der, selv om det høres enkelt ut, så høres det nesten ut som noen tror det noen ganger, på en måte.

05:56.000 --> 05:59.000

Man blander fort sammen, ikke sant?

05:59.000 --> 06:05.000

Det er jo litt det samme som vi snakker veldig ofte om, at de som har høyere utdanning lever lenger og spiser sunnere og sånt.

06:05.000 --> 06:12.000

Men det er jo ikke sikkert at hvis en som ikke hadde høy utdanning gikk og studerte sosiologi på Blindern i fem år,

06:12.000 --> 06:17.000

at de da endelig begynte å trene og spise grovbrød.

06:18.000 --> 06:27.000

Det sier kanskje noe om hvilken bakgrunn du kommer fra, hvordan du har vokst opp med dine foreldre, hvilke vaner du fikk,

06:27.000 --> 06:29.000

hvilke gener du har med deg.

06:29.000 --> 06:31.000

Det er helt andre ting.

06:31.000 --> 06:36.000

Det er ikke vestkant og østkant, eller nødvendigvis den utdanningen i seg selv som har noe å si.

06:36.000 --> 06:40.000

Når du gjør eksperimenter, så kan du jo sette opp.

06:40.000 --> 06:44.000

Dere spiser loff og dere spiser grovbrød, så kan vi sjekke forskjellen.

06:45.000 --> 06:48.000

Da får vi en sånn årsakssammenheng, loff eller grovbrød.

06:48.000 --> 06:50.000

Så hvorfor gjør vi ikke alltid det?

06:51.000 --> 06:53.000

Det hadde vært veldig deilig, tenker jeg.

06:53.000 --> 06:55.000

Da kunne vi fått veldig mange svar.

06:55.000 --> 07:03.000

Men for det første vet jeg ikke hvor mange som ville gått med på å spise loff eller grovbrød i 20 år.

07:03.000 --> 07:05.000

Det hadde vært veldig vanskelig å gjennomføre.

07:05.000 --> 07:08.000

Noen ting er jo direkte uetiske.

07:08.000 --> 07:11.000

Tenk på alle disse alkoholstudiene du har.

07:12.000 --> 07:21.000

Skulle man ha testet, ja dere får drikke fire enheter om dagen i ti år, og så skal dere være avholds.

07:21.000 --> 07:26.000

De fleste synes ikke det er helt etisk riktig å holde på sånn.

07:26.000 --> 07:30.000

Og så blir det fryktelig dyrt også hvis man skal ha eksperimenter på alt.

07:30.000 --> 07:36.000

Veldig mye av forskningen er basert på store befolkningsstudier,

07:36.000 --> 07:39.000

hvor man bare har fulgt mennesker over mange år.

07:39.000 → 07:41.000

Vi har mange av dem i Norge.

07:41.000 --> 07:48.000

Disse som heter helseundersøkelsen i Nord-Trøndelag, HUNT, og vi har flere i Oslo også.

07:48.000 --> 07:54.000

Og de er jo flotte undersøkelser som viser hva har de spist, hvor mye har de trent,

07:54.000 --> 07:56.000

og så hvilke sykdommer får de etter hvert.

07:56.000 --> 08:00.000

Og det er jo sånne undersøkelser hvor vi får vite at

08:00.000 --> 08:04.000

dess hardere du trener, dess lenger lever du, og dess mindre hjerteinfarkt får du.

08:04.000 --> 08:07.000

Eller dess mer grovbrød du spiser, dess bedre er det.

08:07.000 --> 08:11.000

Og kanskje er det sant, men det finnes jo alternative forklaringer.

08:11.000 --> 08:18.000

Hvem i befolkningen er det som klarer å spise bare grovbrød og trene veldig hardt mange ganger i uka?

08:18.000 --> 08:24.000

Kanskje er det folk som har gode jobber, behagelige liv, god råd.

08:24.000 --> 08:30.000

Kanskje er de rett og slett født med bedre gener som gjør at de er friskere som personer.

08:30.000 --> 08:34.000

Det er vanskelig, og vi går i baret hele tiden.

08:34.000 --> 08:38.000

Når vi bare observerer den i virkeligheten og ikke eksperimenterer med den, så får vi jo tall.

08:38.000 --> 08:44.000

Men det er jo noe annet å regne ut en matematisk sammenheng mellom tall, som er det korrelasjonen her.

08:44.000 --> 08:51.000

Og det å få en årsakssammenheng, som er det kausalitet er. Og er også det vi i hjernen helst vil ha.

08:51.000 --> 08:54.000

Det er jo veldig magnetisk den der kausaliteten.

08:54.000 --> 09:00.000

Det er jo det, og jeg tror det er aller lettest å gå i fella når det høres logisk ut.

09:00.000 --> 09:04.000

At å trene mye er sunt, det er jo sånn at ja, selvfølgelig er det det.

09:04.000 --> 09:09.000

Å trene skikkelig hardt er ekstra sunt, ja, selvfølgelig er det det.

09:09.000 --> 09:14.000

Mens når det dukker opp veldig overraskende sammenhenger, så er det kanskje litt lettere å være kritisk.

09:14.000 --> 09:24.000

Under pandemien kom det mange studier som viste at de som hadde høy D-vitaminstatus, nok D-vitamin i kroppen,

09:24.000 --> 09:27.000

de ble ikke like syke av COVID-19.

09:27.000 --> 09:32.000

Det var jo noe som mange hang seg på som en sånn der, vi må ta D-vitamin, ikke sant?

09:32.000 --> 09:37.000

Da gikk det mange diskusjoner rundt i forskning.no på det der, ja, kan det stemme,

09:37.000 --> 09:40.000

eller hva kan være en annen forklaring på at det er sånn?

09:40.000 --> 09:45.000

Helst vil man ha det der eksperimentet som tester det ut, og det gjorde jo faktisk norske forskere,

09:45.000 --> 09:51.000

at de delte mange nordmenn opp i to grupper, som begge fikk en flaske som lignet på tran,

09:51.000 --> 09:57.000

men den ene var ekte tran, som da inneholder D-vitamin, og de andre fikk noe rapsolje med sitronsmak.

09:57.000 --> 10:07.000

Og så holdt de på i flere uker for å se hvem som ble syke i de to gruppene, og de fant ingen effekt av tranen.

10:07.000 --> 10:11.000

Jeg som er statistiker, vi er jo vant med å behandle tall, det er matematikk og det er riktig og galt og sånn,

10:11.000 --> 10:18.000

men det du ikke kan outsource til statistikken, det er jo tolkningen av hva dette her betyr.

10:18.000 --> 10:25.000

Hvor kommer tallene fra? Kan jeg stole på konklusjonen? Hva er det egentlig som er datagrunnlaget?

10:25.000 --> 10:30.000

Hvordan skal jeg tolke dette her? Men du kan ikke outsource alt dette til en statistisk analyse,

10:30.000 --> 10:32.000

du må fortsatt tenke da.

10:32.000 --> 10:41.000

Man må det, og det er jo veldig lett å finne forklaringer som kanskje ikke stemmer, både for forskerne og journalistene.

10:41.000 --> 10:47.000

Men det er jo derfor man må være litt skeptisk når man leser forskningsnyheter, for å si det sånn.

10:47.000 --> 10:57.000

Og så er det å kanskje være litt skeptisk når det bare blir slått fast at noe er fakta.

10:57.000 --> 11:03.000

Tenk på at på grunn av sosiale medier har unge mennesker mer psykiske lidelser.

11:03.000 --> 11:10.000

Det er noen sånne setninger som bare blir slengt ut som om det er sant, og der bør det jo ligge tall bak, ikke sant?

11:10.000 --> 11:16.000

Og så er spørsmålet, gjør det det? Ja, sosiale medier har plutselig kommet inn i livene våre,

11:16.000 --> 11:20.000

og ja, det er noe som tyder på at unge blir mer psykisk syke.

11:20.000 --> 11:26.000

Men da er spørsmålet, er egentlig unge mer psykisk syke enn før?

11:26.000 --> 11:33.000

Eller er det bare at vi fanger det opp, at det er mer åpenhet, at de blir utredet?

11:33.000 --> 11:39.000

Og også, hvordan kan vi i all verden vite at det er sosiale medier som har skylda?

11:39.000 --> 11:46.000

Der finnes det jo en del eksperimenter, som du etterlyser, hvor man gir en gruppe...

11:46.000 --> 11:51.000

Det er nesten alltid studenter som er i sånne psykologiske eksperimenter.

11:51.000 --> 11:58.000

Professoren tar studentene sine og deler dem opp i to grupper, og ber den ene om å skru av telefonen sin,

11:58.000 --> 12:02.000

og den andre om å være på sosiale medier i to timer om dagen, ikke sant?

12:02.000 --> 12:05.000

Men eksperimentene er veldig sprikende.

12:05.000 --> 12:09.000

Det er på en måte at noen viser at man får depressive symptomer,

12:09.000 --> 12:15.000

noen viser at mange blir glad av å bruke sosiale medier, så det er jo ikke så rett fram.

12:15.000 --> 12:18.000

Så det er litt sånn de sure voksne som sitter og...

12:18.000 --> 12:24.000

Det går så dårlig med de unge på grunn av disse forferdelige smarttelefonene og sosiale mediene, ikke sant?

12:24.000 --> 12:29.000

Det digitale får jo skylda for alt. Alt som er gærent.

12:29.000 --> 12:32.000

Og som kanskje ikke er gærent engang.

12:32.000 --> 12:36.000

Man snakker om at barn og unge beveger seg mye mindre enn før,

12:36.000 --> 12:41.000

og det er selvfølgelig også smarttelefoner og gaming og sånn sin skyld.

12:41.000 --> 12:48.000

Men der mangler tallene litt. Vi har ikke gode målinger på hvor mye folk beveget seg før,

12:48.000 --> 12:51.000

og egentlig ikke så veldig gode tall på hvordan de beveger seg nå heller.

12:51.000 --> 12:57.000

Men de tallene vi har tyder på at barn og unge beveger seg ganske mye, egentlig.

12:57.000 --> 12:59.000

Så det er sånne antagelser vi har.

12:59.000 --> 13:02.000

Er dere som journalister flinke nok til å...

13:02.000 --> 13:06.000

Når dere formidler forskning og statistiske resultater,

13:06.000 --> 13:09.000

er dere flinke nok til å skille på den korrelasjonen og kausaliteten,

13:09.000 --> 13:12.000

eller buser dere ut med ting litt for fort?

13:12.000 --> 13:16.000

Svaret er vel nei, vi er ikke flinke nok.

13:16.000 --> 13:19.000

Og ja, vi buser ut med ting litt for fort.

13:19.000 --> 13:24.000

Dessverre, det skulle jeg gjerne sagt.

13:24.000 --> 13:31.000

Jeg håper jo at vi i forskning.no, som er så vant til å jobbe med forskningsartikler og forskere,

13:31.000 --> 13:34.000

er litt flinkere enn Dagbladet og VG.

13:34.000 --> 13:40.000

Men jeg tok en liten titt i arkivet vårt, og jeg ser at vi synder mot det hele tiden.

13:40.000 --> 13:45.000

Det vi prøver på er at vi i hvert fall at et eller annet sted i artikkelen skal skrive dette her med at

13:45.000 --> 13:50.000

forskerne kan ikke vite om det er en årsakssammenheng,

13:50.000 --> 13:56.000

fordi de har ikke testet ut å gi en gruppe grovbrød og en gruppe loffr.

13:56.000 --> 14:02.000

Men det som vi også synder på er at selv om vi kanskje i artikkelen er veldig tydelige på begrensningene,

14:02.000 --> 14:10.000

så headingen antyder kanskje at det er en årsakssammenheng som man egentlig ikke har belegg for.

14:10.000 --> 14:14.000

Vi vet jo det, særlig på Facebook og sånn, når vi deler en artikkel,

14:14.000 --> 14:19.000

så er det kanskje bare en av fem eller enda mindre som faktisk leser artikkelen.

14:19.000 --> 14:27.000

Mens de andre har fått med seg en eller annen fakta om at et glass vin om dagen forlenger livet.

14:31.000 --> 14:33.000

Men det gjør det ikke.

14:33.000 --> 14:40.000

Vi hadde veldig mange sånne saker i 2014, rundt der om hvor sunt det var med alkohol.

14:40.000 --> 14:43.000

Men vi ser jo på de nye kostholdsrådene nå,

14:43.000 --> 14:48.000

og nå tør de jo ikke si en nedre grense som er bra for helsa engang.

14:48.000 --> 14:54.000

Så der har vi helt klart gått i baret mange ganger.

14:54.000 --> 14:58.000

Og der er det jo også en sånn rar liten effekt, for hvis du ser på sammenhengen,

14:58.000 --> 15:03.000

de som drikker ingenting, de som drikker litt, de som drikker litt mer, de som drikker veldig mye,

15:03.000 --> 15:10.000

så har man jo i alle år sett at de der som drikker litt, de er de som lever lengst og er friskest.

15:10.000 --> 15:15.000

Men de der som drikker ingenting, de gjør det faktisk litt dårligere enn de som drikker litt.

15:15.000 --> 15:22.000

Men der kommer det jo igjen, er det alkoholen som har fortjenesten for at de som drikker litt lever lenger.

15:22.000 --> 15:26.000

Og det er jo kanskje ikke det, tror man må.

15:26.000 --> 15:34.000

At de som drikker litt, kan jo hende at det er litt ressurssterke folk som skjønner at de ikke skal drikke for mye,

15:34.000 --> 15:40.000

men likevel vil kose seg med litt vin, så de er ganske friske og lever sunne liv.

15:40.000 --> 15:44.000

Mens de som drikker ingenting, der er det jo selvfølgelig mange forskjellige folk.

15:44.000 --> 15:55.000

Men det kan være litt flere av folk som er syke, har et vanskelig liv, har valgt å ikke drikke på grunn av historie om alkoholisme i familien.

15:55.000 --> 15:59.000

Det kan være mye i den gruppa som gjør at de blir sykere.

16:00.000 --> 16:06.000

Dette er kanskje også noe med måten vi snakker om statistikk og forskningsresultater på.

16:06.000 --> 16:10.000

Du sier at overskriften drar kanskje litt ekstra.

16:10.000 --> 16:14.000

Og jeg vet jo det, det er mange ting jeg leser om i aviser som jeg ikke er sånn super opptatt av.

16:14.000 --> 16:18.000

Jeg er ikke så interessert i det, det er litt sånn lange tekniske avsnitter langt ned i artikkelen.

16:18.000 --> 16:22.000

Men catchy overskrift og kanskje ingressen helt på toppen, det er jo det jeg vil ha.

16:22.000 --> 16:28.000

Det er jo grenser for hvor mange forbehold du kan dytte inn i en catchy overskrift.

16:28.000 --> 16:34.000

Det er grenser for det, og da blir den i hvert fall ikke lest, den artikkelen.

16:34.000 --> 16:43.000

Men det er jo selvfølgelig sånn at man kan prøve å moderere seg litt i den.

16:43.000 --> 16:48.000

Det gjør vi, og det gjør vi i hvert fall i forskning.no, mye mer enn vi gjorde for ti år siden.

16:48.000 --> 16:51.000

Så vi prøver å ikke overselge for mye.

16:52.000 --> 16:58.000

Men tror du folk har et slags, om ikke superbevisst forhold til det,

16:58.000 --> 17:01.000

at man er klar over at sammenheng og årsakssammenheng ikke er det samme,

17:01.000 --> 17:03.000

og at det må tenkes på?

17:03.000 --> 17:08.000

Eller tror du vi mennesker som biologiske vesener er litt kjappere til å trekke den sammenhengen,

17:08.000 --> 17:14.000

fordi vi vil ha årsakssammenhenger? Det er det vi jakter etter?

17:14.000 --> 17:19.000

Ja, jeg tror det. Jeg tror det er sånn hjernen fungerer, at vi skal alltid ha årsak.

17:19.000 --> 17:26.000

Men tilfeldighet er så vanskelig for oss å forstå. Det er vanskelig å tenke at noe er tilfeldig.

17:26.000 --> 17:30.000

Jeg tror vi er bare veldig gode på å se etter mønster.

17:30.000 --> 17:33.000

Det er kanskje det som gjør oss veldig smarte som mennesker.

17:33.000 --> 17:38.000

Vi setter ting i bås, og vi forstår verden rundt oss.

17:38.000 --> 17:44.000

Å nettopp si at det skjedde på grunn av det, det har en årsakssammenheng.

17:44.000 --> 17:47.000

Så jeg tror vi er skikkelig dårlige på det.

17:47.000 --> 17:48.000

Hva skal vi gjøre for å bli bedre?

17:48.000 --> 17:52.000

Øve, øve, øve. Hver gang du hører en nyhet, tenk.

17:52.000 --> 17:56.000

Ja, er det sikkert at det er sånn?

17:56.000 --> 18:01.000

Er det sikkert at det er at elever gjør det bedre på store skoler?

18:01.000 --> 18:08.000

Er det i så fall fordi skolen er stor, eller kan det være noe annet som forklarer det som ikke har noe med saken å gjøre?

18:08.000 --> 18:13.000

Eller plutselig drukner veldig mange mennesker. Er det fordi svømmeundervisningen har blitt dårlig,

18:13.000 --> 18:20.000

eller er det fordi det var veldig fint vær i sommer? At man må sitte og øve litt

18:20.000 --> 18:21.000

Øvekurs for årsakssammenheng?

18:21.000 --> 18:22.000

Ja.

18:23.000 --> 18:25.000

Det er kjempegøy, altså.

18:27.000 --> 18:32.000

Du har jo jobbet som forsker på vei inn i forskningsjournalistikken.

18:32.000 --> 18:36.000

Sittet og sikkert hatt grunnkurs i statistikk, hvordan gikk det?

18:36.000 --> 18:42.000

Jeg må si at jeg syntes det var ganske tørt, det første statistikkurset jeg hadde på universitetet.

18:42.000 --> 18:48.000

Jeg tror noe av grunnen er at da så jeg ikke for meg at dette er noe jeg kommer til å ha bruk for.

18:48.000 --> 18:55.000

Man skjønner på en måte ikke at dette er faktisk: «følg med nå, for dette kommer du til å bruke resten av livet».

18:55.000 --> 19:02.000

Etterhvert valgte jeg frivillig å ta en god del flere statistikkurs, fordi jeg skjønte både at det var veldig spennende,

19:02.000 --> 19:09.000

og at på dette her kan jeg bruke min kunnskap i molekylærbiologi på en helt annen måte.

19:09.000 --> 19:17.000

Hvis jeg kan forstå tall litt mer, da kan jeg sitte og regne på store datasett med tall jeg synes er spennende.

19:17.000 --> 19:23.000

Så havnet jeg i journalistikken, men så er det jo fortsatt veldig nyttig for meg å forstå dette med statistikk.

19:23.000 --> 19:29.000

Så der må jeg bare si til alle studentene som hører på at prøv å bare sette av tid til det faget,

19:29.000 --> 19:34.000

for det er nyttig, og du kommer til å skjonne det om noen år, hvis ikke du skjønner det nå.

19:34.000 --> 19:37.000

Hva er ditt beste statistikkøyeblikk da?

19:37.000 --> 19:40.000

Det var også et spørsmål.

19:40.000 --> 19:50.000

Det må nok ha vært i doktorgraden å få noen skikkelig bra p-verdier.

19:51.000 --> 20:00.000

Det betyr at jeg kan publisere artikkelen min, for jeg har funnet noe som er statistisk signifikant,

20:00.000 --> 20:03.000

som betyr at det ikke er tilfeldig.

20:03.000 --> 20:09.000

Men, som vi har snakket om i dag, det er ikke sikkert at det var årsaks-sammenheng for det.

20:11.000 --> 20:14.000

Hva er ditt verste statistikkøyeblikk?

20:14.000 --> 20:17.000

Dette føler jeg sier mye om statistikere.

20:17.000 --> 20:24.000

Det var mitt verste øyeblikk i en liten stund, hvor jeg hadde tatt en eksamen i et statistikkfag,

20:24.000 --> 20:33.000

hvor vi skulle prøve å lage en statistikkmodell som kunne forutse hvem av disse har kreft.

20:33.000 --> 20:38.000

Det var noe data fra noen blodprøver og noen målinger som var gjort i blodprøver.

20:38.000 --> 20:45.000

Så hadde jeg klart å blande den merkelappen på hvem som hadde kreft og hvem som ikke hadde kreft,

20:45.000 --> 20:47.000

så jeg hadde helt motsatt.

20:47.000 --> 20:52.000

Jeg klarte å forutse veldig godt at alle disse hadde kreft, som var de som var friske.

20:52.000 --> 20:55.000

Så da følte jeg meg som en kronidiot.

20:55.000 --> 20:58.000

Men det som var litt gøy var at – dette er typisk statistikk.

20:58.000 --> 21:02.000

Det syntes ikke de statistikerne var så veldig farlig. Det var ikke så stor feil.

21:02.000 --> 21:05.000

For jeg hadde gjort det riktig, jeg hadde bare blandet det sammen.

21:06.000 --> 21:09.000

Kan du se? Vi statistikere er ikke så opptatt av den virkelige verden.

21:09.000 --> 21:10.000

Nei.

21:11.000 --> 21:15.000

Du, tusen takk, Eldrid. Veldig hyggelig.

21:15.000 --> 21:18.000

Veldig hyggelig å komme og prate om statistikk.

21:18.000 --> 21:20.000

Ja, hvem skulle tro du skulle si det?

21:20.000 --> 21:21.000

Nei.

21:24.000 --> 21:28.000

Korrelasjon er ingen garanti for kausalitet.

21:28.000 --> 21:33.000

Noen ganger er det enkelt å forstå at det må være noe annet som ligger bakom og lurere,

21:33.000 --> 21:39.000

en tredje faktor som fører til den tallmessige sammenhengen mellom de to variablene vi ser på.

21:39.000 --> 21:42.000

Andre ganger er det ikke like enkelt å oppdage.

21:42.000 --> 21:46.000

Og både vanlige folk og forskere, politikere og journalister har gått på en smell.

21:46.000 --> 21:51.000

Men går statistikere også i fella og feiltolker korrelasjon som kausalitet?

21:52.000 --> 21:59.000

Etter min erfaring så er kanskje statistikere de minst bastante og mest forsiktige til å konkludere med kausalitet.

21:59.000 --> 22:01.000

I hele verden, faktisk.

22:01.000 --> 22:04.000

Vi er livredd for å gå i den fella der.

22:04.000 --> 22:05.000

Men har du feilet da?

22:05.000 --> 22:12.000

Jeg liker å tro at jeg er veldig nøye på hvordan jeg legger frem resultater fra statistiske tester.

22:12.000 --> 22:15.000

Jeg vet jo at dette er en stor fallgrube.

22:15.000 --> 22:19.000

Men jeg vet jo at når jeg sitter på kontoret selv og gjør analyser,

22:19.000 --> 22:24.000

så hender det jo at jeg lar meg rive med når jeg oppdager en sterk sammenheng, en sterk korrelasjon.

22:24.000 --> 22:30.000

Og det er jo veldig fristende å tenke at det betyr at noe henger sammen, og at det ene påvirker det andre.

22:30.000 --> 22:34.000

Jeg håper jo ikke at jeg har direkte feilet, men flere ganger har jeg vært veldig forvirret.

22:34.000 --> 22:41.000

Jeg husker en dataanalyse der jeg jobbet med kostholdsdata fra helseundersøkelsen i Trøndelag HUNT.

22:41.000 --> 22:46.000

Der hadde vi delt HUNT-befolkningen inn i ulike grupper basert på grad av overvekt.

22:46.000 --> 22:52.000

Så ville vi se om det var noen spesielle mønster av matinntak, om det var enkelte matvarer som skilte seg ut.

22:52.000 --> 22:59.000

Og fra en sånn beskrivende analyse av dataene så vi det var stort potetinntak i grupper med størst grad av overvekt.

22:59.000 --> 23:03.000

Kan du se overskriften? «Poteter gjør oss overvektige».

23:03.000 --> 23:11.000

Men hvem er det som spiser mest poteter i Norge? Jo, det er de som er eldst, og dessverre blir mange av oss tyngre om årene.

23:11.000 --> 23:17.000

Så alder var nok den underliggende forklaringen på både overvekt og potetinntak.

23:17.000 --> 23:18.000

Men har du feilet da?

23:18.000 --> 23:20.000

Jeg kommer ikke på dem altså.

23:20.000 --> 23:27.000

I så fall så leverer jeg lykkelig uvitenhet om å ha virkelig bastant gått i denne fella.

23:27.000 --> 23:35.000

Men jeg har hatt kjempe mange gode og lange og interessante diskusjoner med forskere for å unngå det.

23:35.000 --> 23:40.000

Jo mer komplekse sammenhenger og datasett er, og analyser som vi gir oss i kast med,

23:40.000 --> 23:45.000

jo større er også sannsynligheten for at noen av de formuleringene som ender på trykk,

23:45.000 --> 23:51.000

kanskje går litt for langt i retning av kausalitet, selv om analysene ikke helt støtter det.

23:51.000 --> 23:54.000

Så noen vil nok kanskje si at jeg har gått i den fella.

23:55.000 --> 24:01.000

Det en statistisk analyse gjør, er å regne ut om det er en matematisk tallmessig sammenheng

24:01.000 --> 24:05.000

mellom to størrelser vi har tallfestet. Korrelasjon.

24:05.000 --> 24:08.000

Men kan man også bruke statistikk for å avdekke kausalitet?

24:08.000 --> 24:10.000

Veldig glad for at du spør om.

24:10.000 --> 24:18.000

Når jeg studerte statistikk så var det et sånn mantra at du kunne aldri bruke statistikk til å finne kausalitet.

24:18.000 --> 24:22.000

Du kan ikke bruke statistikk alene. Det går ikke.

24:22.000 --> 24:27.000

Men ved å kombinere den kunnskapen du har om det du skal studere,

24:27.000 --> 24:32.000

slik at du setter opp gode studiedesign, og at du velger de riktige analysene,

24:32.000 --> 24:34.000

så kan du komme ganske nær kausalitet.

24:34.000 --> 24:37.000

Det er jo en slags heldig gral det å kunne si noe om årsak og virkning.

24:37.000 --> 24:43.000

Det er jo veldig ofte det en har lyst til i mange sammenhenger i vitenskap.

24:43.000 --> 24:47.000

Ja, det er mulig å avdekke kausalitet.

24:47.000 --> 24:53.000

Men det krever ofte gode og nøye planlagte studier med gode studiedesign,

24:53.000 --> 24:55.000

spesielt når man forsker på mennesker.

24:55.000 --> 25:01.000

Det lureste her er kanskje å planlegge et forsøk slik at det er mulig å si noe om årsak og virkning.

25:01.000 --> 25:05.000

Så innen medisin så bruker man jo da randomiserte kliniske forsøk,

25:05.000 --> 25:08.000

som er en slags gullstandard for å avdekke kausalitet.

25:08.000 --> 25:11.000

Spesielt for eksempel når du skal forstå en ny behandling.

25:11.000 --> 25:15.000

Så kort fortalt så tar vi da en haug med personer som er med i et studie,

25:15.000 --> 25:17.000

og deler dem tilfeldig inn i to grupper.

25:17.000 --> 25:20.000

Den ene får en narremedisin, og den andre en ekte medisin.

25:20.000 --> 25:25.000

Nå må hverken personene eller de som analyserer datene, få vite hvem som har fått hva.

25:25.000 --> 25:30.000

Og finner vi en effekt, kan vi da trolig konkludere med at det er effekten av medisinen,

25:30.000 --> 25:32.000

og ikke av andre utenforliggende årsaker.

25:32.000 --> 25:38.000

Og da kan du forsvare av og til å tolke de sammenhengene som du finner som årsakssammenhenger,

25:38.000 --> 25:41.000

selv om statistikken i seg selv ikke gir deg det.

25:42.000 --> 25:44.000

Statistikk er matematikk.

25:44.000 --> 25:49.000

Vi putter tall inn i ligninger, og ut kommer svar med to streker under.

25:49.000 --> 25:53.000

Men statistikk er ikke bare matematikk, det er også kontekst.

25:53.000 --> 25:58.000

Tallene vi analyserer kommer fra et sted. De representerer noe i den virkelige verden.

25:58.000 --> 26:03.000

Og hva akkurat det betyr, kan ikke matematikken uten videre hjelpe oss med.

26:03.000 --> 26:05.000

Matematikken har ingen hjerne.

26:05.000 --> 26:10.000

Den vet ikke, den kan ikke vite, om det er rimelig at sammenhengene vi har funnet i tallene,

26:10.000 --> 26:13.000

også innebærer en årsakssammenheng.

26:13.000 --> 26:17.000

En statistisk analyse skal bidra til å finne ut hva tallene forteller,

26:17.000 --> 26:22.000

og måten tallene er samlet inn på er en viktig del av denne fortellingen.

26:22.000 --> 26:26.000

Så det å tenke, det må vi fortsatt gjøre selv.

26:26.000 --> 26:32.000

I denne episoden av Nakne tall har vi snakket om forskjellen på korrelasjon og kausalitet.

26:32.000 --> 26:35.000

Sammenheng og årsakssammenheng.

26:35.000 --> 26:42.000

Vi har snakket om at måten vi samler inn tallene på vil være med å avgjøre hvor bombastiske konklusjoner vi kan trekke.

26:42.000 --> 26:45.000

Noen tall er bedre enn andre tall.

26:45.000 --> 26:49.000

Det er en grunn til at forskere tviholder på det å eksperimentere,

26:49.000 --> 26:55.000

til tross for at det ofte er både ressurskrevende og tidkrevende, og generelt bare krevende.

26:55.000 --> 27:01.000

I neste episode av Nakne tall skal vi snakke om hele grunnlaget for de statistiske testene vi bruker

27:01.000 --> 27:04.000

for å avdekke sammenhenger i tallene.

27:04.000 --> 27:10.000

Om det å ha en idé om hvordan et eller annet i verden henger sammen, det vi kaller en hypotese,

27:10.000 --> 27:14.000

og så bruke tall for å teste denne hypotesen.

27:15.000 --> 27:18.000

Jeg heter Jo Røislien og er professor i statistikk.

27:18.000 --> 27:24.000

Sammen med meg i redaksjonen var Magnus Nome og Olav Rroen, begge fra produksjonsselskapet Teddy.

27:24.000 --> 27:31.000

Mens ekspertpaneler bestod av Kathrine Frey Frøslie, Mette Langås, Marianne Rikshem Stavset og Jan Terje Kvalheim.

27:31.000 --> 27:34.000

Produsent for denne podkassen har vært Mari Linn Atterås Larsen,

27:34.000 --> 27:39.000

og teknikere Elisabeth Mary Armstrong, Arne Thomas Lund Nilsen og Sebastian Fittjar.

27:39.000 --> 27:43.000

Nakne tall er produsert av Universitetet i Stavanger.