

**Obs: denne transkripsjonen er automatisk generert og kan derfor inneholde små feil.**

00:00.000 --> 00:02.000

Nakne tall

00:13.600 --> 00:15.600

Verden flommer over av tall.

00:15.600 --> 00:21.600

Om du er opptatt av helse eller økonomi, ingeniørfag eller samfunnsfag spiller liten rolle.

00:21.600 --> 00:28.600

Det finnes knapt det fagområdet igjen, der ikke det å samle inn og analysere tall er en helt sentral del

00:28.600 --> 00:33.600

av det å skaffe til vei ny kunnskap og innsikt for å bringe fagområdet fremover.

00:33.600 --> 00:37.600

Det betyr dermed at det å kunne noe om det å analysere tall informasjon,

00:37.600 --> 00:41.600

hvordan man bør gå frem og hva man helst ikke skal gjøre,

00:41.600 --> 00:47.600

har gått fra å være kunnskap for de spesielt interesserte til å bli noe de aller fleste må kunne noe om.

00:47.600 --> 00:50.600

Samme hva slags fag du ønsker å fordype deg i.

00:50.600 --> 00:54.600

I denne podkasten skal vi snakke om akkurat det.

00:54.600 --> 00:58.600

På gjestelisten står et kobbelt av Norges mest kjente formidlere av vitenskap,

00:58.600 --> 01:03.600

fra Andreas Wahl og Ingeborg Sennesett til Inga Strumke og Harald Eia.

01:03.600 --> 01:08.600

Alle har de entusiastisk taket ja til å være med og snakke om statistikk.

01:08.600 --> 01:10.600

Og det er kanskje ikke så rart.

01:10.600 --> 01:17.600

Statistiske analysemetoder havner kanskje ikke på forsiden av de store avisene eller er første sak i nyhetene.

01:17.600 --> 01:22.600

Det er gjerne resultatene av de statistiske analysene som får mest oppmerksomhet.

01:22.600 --> 01:27.600

Samtidig, hvis den statistiske analysen er dårlig, blir resultatene deretter

01:27.600 --> 01:32.600

og den tilhørende diskusjonen om hva alt dette betyr ganske verdiløs.

01:32.600 --> 01:35.600

Jeg heter Jo Røislien og er statistiker.

01:35.600 --> 01:40.600

Og i en seilbåt i skjergården på Sørlandet, omgitt av puter og dyner med blomstermønster,

01:40.600 --> 01:44.600

sitter fysiker og vitenskapsformidler Andreas Wahl.

01:44.600 --> 01:47.600

Jeg har alltid hatt et ganske avslappet forhold til tall.

01:47.600 --> 01:56.600

Det er en del som stresser litt, men det betyr ikke at jeg alltid har fått til skole ting med tall.

01:56.600 --> 01:58.600

Tall er informasjon.

01:58.600 --> 02:04.600

Men tall er også matematikk, så det finnes regler for hva man kan og hva man ikke kan gjøre

02:04.600 --> 02:06.600

for at tallanalysen skal bli rett.

02:06.600 --> 02:08.600

Så la oss starte forsiktig.

02:08.600 --> 02:12.600

La oss se på hva vi må tenke på når vi har tall på ett enkelt fenomen,

02:12.600 --> 02:17.600

en størrelse eller en variabel, som det heter, og vil si noe om denne.

02:17.600 --> 02:21.600

Hvor mye tjener håndverkere? Hvor vanlig er det egentlig med elbil?

02:21.600 --> 02:24.600

Og hvor mye drikker ungdommer, som omtrent?

02:24.600 --> 02:30.600

Det å analysere en tallstørrelse, en variabel, kalles univariatanalyse.

02:30.600 --> 02:35.600

Siden jeg er statistiker, så har jeg selvsagt mange statistikerevenner.

02:35.600 --> 02:40.600

Så jeg har ringt fire av dem og spurt dem, hva er univariatanalyse?

02:40.600 --> 02:42.600

Og jeg er veldig glad for at du spør om.

02:42.600 --> 02:48.600

For det er altså å oppsummere en hel haug med enkelt historier og fortelle historien om de mange.

02:48.600 --> 02:55.600

Univariatanalyse er at vi bruker statistiske metoder til å forstå data samlet inn på en egenskap.

02:55.600 --> 03:00.600

Du er bare interessert i å studere personens blodtrykk,

03:00.600 --> 03:04.600

eller du er bare interessert i å studere huspriser,

03:04.600 --> 03:10.600

eller du er bare interessert i hva er prisen på en standard kog,

03:10.600 --> 03:12.600

som er handlet inn på ulike butikker.

03:12.600 --> 03:16.600

Så hvis jeg har lyst til å beskrive det systoliske blodtrykket til kvinner med en spesiell sykdom,

03:16.600 --> 03:20.600

så velger jeg ut tilfeldig 100 kvinner, måler systolisk blodtrykk,

03:20.600 --> 03:23.600

og så vil jeg starte med å se på datene.

03:23.600 --> 03:26.600

Faktisk se på alle tallene, og så vil jeg lage ulike plott.

03:26.600 --> 03:30.600

Beskrivende statistikk er en slags første steg i en univariatanalyse.

03:30.600 --> 03:34.600

Beskrivende statistikk, eller deskriptiv statistikk som vi ofte kaller det,

03:34.600 --> 03:40.600

handler om å oppsummere et sett med observasjoner, et sett med målinger,

03:40.600 --> 03:42.600

og trekke ut essensen i det.

03:42.600 --> 03:46.600

De mest kjente målene er for eksempel gjennomsnitt og median,

03:46.600 --> 03:49.600

som sier noe om hva de fleste har svart,

03:49.600 --> 03:54.600

og det vi kaller for spredningsmål, som sier noe om hvor stor variasjon det er i svaret.

03:54.600 --> 03:59.600

Vi skal beskrive ting med noen få oppsummeringstall, og gjøre det på en god måte.

03:59.600 --> 04:01.600

Det er beskrivende statistikk.

04:01.600 --> 04:05.600

Ved å sette tall på ulike fenomener i den virkelige verden,

04:05.600 --> 04:08.600

kan vi få viktig informasjon om verden vi lever i.

04:08.600 --> 04:12.600

Men for at vi virkelig skal få nyttegjort av alle tallene som er samlet inn,

04:12.600 --> 04:15.600

må vi få oppsummert dem på en effektiv måte.

04:15.600 --> 04:18.600

Vi trenger noen gode oppsummeringstall.

04:18.600 --> 04:22.600

Grovt sett finnes det to hovedtyper av tall.

04:22.600 --> 04:25.600

Kontinuerlige tall og kategoriske tall.

04:25.600 --> 04:28.600

Kontinuerlige tall er de vi gjerne har målt med et måleapparat,

04:28.600 --> 04:31.600

og som kan ta en hel rekke ulike verdier.

04:31.600 --> 04:36.600

Sånn som høyde og vekt, blodtrykk, temperatur, tid og penger.

04:36.600 --> 04:41.600

Kategoriske tall får vi når vi har delt opp et fenomen i ulike kategorier,

04:41.600 --> 04:44.600

som vi så kan putte observasjonene våre oppi.

04:44.600 --> 04:47.600

Kvinner i en kategori, og menn i en annen.

04:47.600 --> 04:50.600

Farlige matvarer her, og ufarlige matvarer der.

04:50.600 --> 04:56.600

Nordmenn i en kategori, svensker i en annen, dansker, finner og isleninger, og resten.

04:57.600 --> 05:01.600

Nøyaktig hvordan kategoriene bør lages, er det verdt å tenke nøye gjennom.

05:01.600 --> 05:05.600

Men når vi har bestemt oss for hva kategoriene skal være,

05:05.600 --> 05:08.600

er det å oppsummere slike kategoriske tall ganske enkelt.

05:08.600 --> 05:12.600

Det er bare å telle hvor mange som er i hver kategori.

05:12.600 --> 05:15.600

Å oppsummere kontinuerlige tall derimot, krever litt mer.

05:15.600 --> 05:18.600

Det er det mer som må tas hensyn til.

05:18.600 --> 05:23.600

Andreas Wahl er fysiker, og har vært programleder for TV-programmer

05:23.600 --> 05:26.600

som Folkeopplysningen og Med livet som innsats på NRK.

05:26.600 --> 05:31.600

Han har laget podcaster og skrevet bøker, og formidlet kunnskap og vitenskap i en årrekke.

05:32.600 --> 05:41.600

Nå, de siste ti årene, har tall vært en mulighet til å forstå verden, og ikke bli lurt.

05:41.600 --> 05:47.600

Jeg ser for meg at de som har mest bruk for statistikk, er de som virkelig ønsker å lure folk.

05:48.600 --> 05:55.600

Så kan du jo bruke tall ganske godt for å lure folk.

05:55.600 --> 06:00.600

Det er mange som ikke helt forstår litt statistikk, men ikke så mye.

06:00.600 --> 06:08.600

Så bare ved å ta det fra trinn 1 til trinn 2 med statistikk, så har du funnet nøklene til å ikke bli lurt.

06:08.600 --> 06:11.600

Har du blitt lurt av tall og statistikk noen gang?

06:11.600 --> 06:21.600

Absolutt. Ofte når jeg leser nyheter, så presenterer de grafer der det ser ut som det har vært en voldsom økning,

06:21.600 --> 06:24.600

eller noe som har gått voldsomt ned de siste par årene.

06:24.600 --> 06:30.600

Helt til jeg nå etter hvert har lært meg å stoppe opp og kikke litt på disse stolpediagrammene,

06:30.600 --> 06:34.600

og sett at hele bildet er jo ikke med.

06:34.600 --> 06:40.600

Det er ned fra 44 i forfjor, og så var det 42 i fjor, og så er det 40 i år.

06:40.600 --> 06:45.600

Og så ser du som det er en halvering, fordi de har bare kuttet båndene av stolpediagrammene.

06:45.600 --> 06:49.600

Det er en måte jeg da kan bli lurt på.

06:49.600 --> 06:56.600

Det ser ut som at da den regjeringen har kommet til makten,

06:56.600 --> 06:58.600

se hva som har skjedd med grafen etter det,

06:58.600 --> 07:03.600

eller at hver gang vi gjorde sånn og sånn på arbeidsplassen, så gikk produksjonen sånn og sånn.

07:04.600 --> 07:11.600

Og så har vi prøvd å terpe inn at det at ting skjer samtidig ikke nødvendigvis trenger å bety at det ene avhenger av det andre.

07:11.600 --> 07:13.600

At de ikke henger sammen.

07:13.600 --> 07:16.600

Ting kan skje samtidig helt tilfeldig også.

07:16.600 --> 07:19.600

Jeg pleier jo litt sånn grovt å si at det finnes to forskjellige typer tall.

07:19.600 --> 07:24.600

Det ene er sånne kontinuerlige tall, hvor vi måler noe med et måleapparat.

07:24.600 --> 07:28.600

Men så har vi en andre type tall også, hvor vi har laget noen kategorier,

07:28.600 --> 07:30.600

og så teller vi hvor mange som er i hver kategori.

07:30.600 --> 07:32.600

Hva tenker du om en sånn indeling?

07:32.600 --> 07:35.600

Til syn i at det kan være veldig greit å forholde seg til.

07:35.600 --> 07:43.600

Det er lett å bruke de dataene etterpå til å si noe om at det er litt flere menn som opplever sånn og sånn sykdom.

07:43.600 --> 07:48.600

Blå er den beste fargen å ha på klær.

07:48.600 --> 07:52.600

Men det ligger jo mer bak da. Ofte.

07:52.600 --> 07:54.600

I helsefagene driver det om om det er syk eller frisk.

07:54.600 --> 08:00.600

Hvis vi måler temperaturen på noen, så finnes det jo ikke noen tydelig grense for hva som gjør det frisk.

08:01.600 --> 08:04.600

Akkurat der så bikker du og blir syk. Ja, det er ikke så tydelig grense.

08:04.600 --> 08:11.600

Det der i kategoriene er jo en måte å lage system i verden på, som også er tellbar og fin.

08:11.600 --> 08:18.600

Jeg husker jo da kjæresten min studerte antropologi, kulturfag og sånt, så driver han og studerte kulhet.

08:18.600 --> 08:21.600

Og alle er enige om at kategorien kul, den finnes.

08:21.600 --> 08:26.600

Men ingen er enige om hva som er inni den kategorien.

08:26.600 --> 08:28.600

Da har du plutselig noe subjektivt.

08:28.600 --> 08:34.600

Hvis du skal plassere en ... Hvis du får presentert 40 mennesker med bilder,

08:34.600 --> 08:39.600

så skal du putte dem i sekken kul eller ikke så kul.

08:39.600 --> 08:41.600

Da er det mye synsing.

08:41.600 --> 08:54.600

Det ville kanskje være i en sykehussetting også, i form av at samme pasient kan bli kategorisert som sjuk eller ikke sjuk av to ulike mennesker.

08:55.600 --> 09:02.600

Tror du vi er kritiske nok til å vurdere tallene som vi driver og diskuterer?

09:02.600 --> 09:08.600



Det er lett å si at vi diskuterer gjennomsnitt for menn og kvinner, men diskuterer vi tallene i seg selv kritisk nok, tror du?

09:08.600 --> 09:15.600

Ikke når jeg leser den rapporten etterpå, og ser at menn var litt lengre på sykehuset,

09:15.600 --> 09:26.600

eller ble vurdert til å være litt sykere av denne infeksjonen, så leser jeg det som hare tallet, på sannhet.

09:26.600 --> 09:35.600

Men du burde jo tenke, hvem er det som har vurdert dette her, og satt det i ulike kategorier?

09:35.600 --> 09:41.600

Jeg spør fordi at du er fysiker, og dere er en gjeng som har ganske hare tall.

09:41.600 --> 09:44.600

Det er åpenbart hva noe veier, og hvor langt det er.

09:44.600 --> 09:49.600

Det er klart at tallene kommer litt derfra, fra matematikken og fysikken, hvor det er veldig tydelig hva det er.

09:49.600 --> 09:56.600

Men når vi flyttes over til medisin, helse og samfunn, så blir tallene litt mer mushy.

09:56.600 --> 10:00.600

Men likevel, de er besnærende, de der tallene.

10:00.600 --> 10:05.600

Jeg tror jeg har lett for å tenke at noen har funnet ut det, ja, men da er det vel sånn.

10:05.600 --> 10:10.600

Og så er det når jeg har overskudd, at jeg tenker, stopp en halv.

10:10.600 --> 10:13.600

Hvordan har de kommet frem til dette?

10:13.600 --> 10:17.600

Eller hva er det som ligger bak tallene? Hva er det tallene representerer?

10:17.600 --> 10:21.600

Det gjør jeg ikke alltid, men jeg burde gjøre det.

10:21.600 --> 10:23.600

Statistikk handler jo om å oppsummere tallene.

10:23.600 --> 10:29.600

Alle som har hatt bitt litt i statistikk, og lærer litt statistikk, vet at det finnes flere måter å oppsummere de samme tallene på.

10:29.600 --> 10:36.600

Og hver gang vi for eksempel snakker om inntekt, SSB, kommer ut med hvor mye det tjener Ola og Kari Nordman hver høst.

10:36.600 --> 10:39.600

Er det gjennomsnitt, eller er det median?

10:39.600 --> 10:47.600

Ja, er det det du har merket at de alltid bruker median inntekt, eller alltid i hvert fall ofte?

10:49.600 --> 10:51.600

Hvorfor gjør de det?

10:52.600 --> 10:55.600

Ja, ikke sant? For dette er jo spørsmålet om hvorfor ofte, ikke sant?

10:55.600 --> 11:01.600

For alle har hørt den der flotte gjennomsnittet som liksom er en sånn matematisk flott beregning av hvor det typiske ligger.

11:01.600 --> 11:06.600

Median går jo mer rett på tallene, ikke sant? Hva er det som er de midterste tallene?

11:06.600 --> 11:13.600

Og jeg elsker jo det der i casene hvor det finnes en eller annen sånn ekstremt rik fyr, i ei lita bygd,

11:13.600 --> 11:20.600

som gjør at gjennomsnittsinntekten trekkes jo opp og vekk fra der det typiske ligger.

11:20.600 --> 11:28.600

Det er ofte sånn, når vi snakker om de rikeste kommunene i Norge, er det han Gunnar Kjumlid som omtaler sin egen hjemsted,

11:28.600 --> 11:31.600

Sirdal, som er en sånn kvalmende rik kommune.

11:31.600 --> 11:37.600

Der bor det jo ikke så mange, men det bor noen rikinger som drar opp snittet noe voldsomt,

11:37.600 --> 11:45.600

og at det blir litt sånn urettferdig å fremstille som at sirdøler generelt har veldig mye penger.

11:45.600 --> 11:49.600

Ja, og det er jo problematisk med en del statistikk, synes jeg da, ikke sant?

11:49.600 --> 11:54.600

At den er jo ikke egentlig oppsummerende for det fenomenet den skal oppsummere.

11:55.600 --> 11:58.600

Og så tenk på et sykehus da, masse sykepleiere, masse leger.

11:58.600 --> 12:02.600

Er det meningsfullt å gi opp gjennomsnittsinntekten, tenker du?

12:02.600 --> 12:13.600

Snittet av lønnene til alle på sykehuset vil føles urettferdig høy for sykepleiere antageligvis,

12:13.600 --> 12:22.600

fordi de overlegende som er i stort mindretall, men som tjener veldig mye, de påvirker snittet mye.

12:24.600 --> 12:28.600

Hva ville du gjort på et sykehus hvis det var masse sykepleiere og overleger?

12:28.600 --> 12:39.600

Hvis jeg skulle funnet den personen som har en typisk lønn, så ville det jo være mer nyttig å bruke medianinntekt.

12:39.600 --> 12:45.600

Hvis jeg skulle vært riktig, så ville jeg da stilt opp alle de ansatte på sykehuset på en lang rekke,

12:45.600 --> 12:49.600

fra lavest lønn til høyest lønn, litt dårlig gjort kanskje,

12:49.600 --> 12:58.600

og så talt meg halveis på rekka. Og den personen som står i midten, det er den med median inntekt. Er det riktig?

12:58.600 --> 13:03.600

Ja, det er helt riktig. Det er et ganske fint TV-bilde da, med alle sammen stilt opp ved siden av hverandre.

13:03.600 --> 13:08.600

Kjempefint. Også nærbildet av både antageligvis han som står helt øverst og tjener mest,

13:08.600 --> 13:18.600

og hundre han som tjener aller minst også, får det her ubehagelige nærbildet av vedkommende som kan se hele rekka oppover,

13:18.600 --> 13:25.600

og som tjener aller, aller minst på sykehuset. Gode TV-bilder, og et interessant eksperiment.

13:25.600 --> 13:34.600

Men jeg vil også si at median inntektene, altså den personen som står midt i denne rekka her, er mer representativ enn gjennomsnittet,

13:34.600 --> 13:41.600

som da vil være en person litt høyere opp, altså en med litt høyere lønn, vil jeg antage da.

13:41.600 --> 13:49.600

Da nærmer vi oss litt mer i retning direktør enn av denne rekka.

13:50.600 --> 13:57.600

Eller skal vi si at det er ikke meningsfullt å lage et oppsummerings-tal for hele sykehuset,

13:57.600 --> 14:03.600

men vi må ta de ulike yrkesgrupperne, de ulike kategoriene, og gi laget gjennomsnitt for dem? Er det bedre?

14:03.600 --> 14:12.600

Ja, det vil jo være mer nyansert å kanskje gi et mer nyttig bilde.

14:12.600 --> 14:18.600

Det spørres litt hvordan du ser det. Hvis du ser at alle går på jobb og gjør en viktig jobb på sykehuset,

14:18.600 --> 14:26.600

så er det kanskje det første bildet viktig. Hvis du skal ta med hvor stort studielån og hvor mange år av livet folk har brukt før de kom dit,

14:26.600 --> 14:32.600

så er det kanskje mer nyttig å spritte det opp. Blir det plutselig politikk i dette? Lønnspolitikk?

14:32.600 --> 14:39.600

Ja, jeg er litt opptatt av at statistikk ikke er viktig i seg selv. Det er jo for å belyse et fenomen.

14:39.600 --> 14:46.600

Det sier statistiker-Jo at statistikken i seg selv ikke har noen egenverdi. At det er bare når du kan beskrive noe annet. Hører jeg riktig?

14:46.600 --> 14:52.600

Jeg som matematiker synes at statistikk er gøy i seg selv. Men Norsk Statistisk Forening er 300 mennesker.

14:53.600 --> 15:03.600

Dere koser dere med statistikk alene, men for resten av oss er statistikk først verdifullt eller spennende når det beskriver noe rundt oss.

15:03.600 --> 15:13.600

Ja, og beskriver noe som du er opptatt av. For deg ville kanskje være landet på Mars, eller hvordan muttere bør se ut for å lage en god maskine.

15:13.600 --> 15:16.600

Nå er det du som tenker at tallet her var skikkelig digg.

15:16.600 --> 15:28.600

Oi. Tall er jo deilig når du får den der «Aaah!» Så det betyr at derfor er det ikke «Aaah!» når du får den «Aha!»-opplevelsen.

15:28.600 --> 15:39.600

Da er det veldig deilig. Og så tall er jo lettere å bruke når vi skal prøve å sette sammen store mengder med informasjon.

15:39.600 --> 15:46.600

Det er nyttig å velge en representant for sykepleierne, for eksempel på sykesfører, og snakke om lønn.

15:46.600 --> 15:53.600

Det kan være veldig nyttig. Men hvis vi skulle prøve å konkludere på vegne av alle sykepleiere i hele landet, for eksempel,

15:53.600 --> 16:02.600

da må vi overpåk det. Må ikke. Men jeg vil synes det er mye lettere å få en oversikt som er tallfesta.

16:02.600 --> 16:07.600

Og da kan det også være tallfesta hvor fornøyd de er med lønnen sin på en skala fra 1 til 5,

16:07.600 --> 16:16.600

eller hva slags arbeidstid de har gjort om til tall. For de store mengder med informasjon blir litt sånn uoversiktelige.

16:16.600 --> 16:23.600

Hvis jeg skulle hatt et intervju med alle sykepleiere i Norge og prøvde å få et inntrykk, få en oversikt,

16:23.600 --> 16:34.600

så hadde det vært grusomt, vil jeg si. Prøve å få noe ut av det. Så tallene er jo ikke på bekostning av den samtalen,

16:34.600 --> 16:44.600

eller det kvalitative, opplevelsen. Tallene er jo i tillegg et veldig nyttig verktøy.

16:44.600 --> 16:51.600

Og så dukker vi veldig ofte opp i den diskusjonen om hva som er det typiske. Kari og Ola Nordman og sånn.

16:51.600 --> 17:01.600

Samtidig må jeg jo innrømme at det er ofte litt kjedelig med gjennomsnitt. Jeg har ikke lyst til å gå på kino og se en film om en gjennomsnittlig fyr på en gjennomsnittlig dag.

17:01.600 --> 17:03.600

Nei, det er sant.

17:03.600 --> 17:09.600

Jeg vil jo helt se på noe litt mer ekstremt. Disse ekstreme tilfellene, det som i statistikk ofte kalles outliers, uteliggere.

17:09.600 --> 17:16.600

Ekstreme tilfellene. Det er jo morsommere, Andreas. Vil ikke du? Liker du å se gjennomsnittlige filmer om gjennomsnittlige folk?

17:17.600 --> 17:20.600

Nei, nei. Jeg synes jo det er ... Nei.

17:20.600 --> 17:33.600

Det er både Netflixen og NRK TV og ikke minst algoritmene i min Instagram feed og TikTok feed.

17:33.600 --> 17:41.600

Har jo skjønt at det er de overraskende, litt ekstreme tingene som treffer og som jeg blir vel litt lengre for.

17:41.600 --> 17:44.600

Det er litt overraskende på et eller annet nivå.

17:44.600 --> 17:52.600

Den veldig spesielle hunden, for eksempel, som ligger langt utenfor normalen, uansett hvilken hunderase du snakker om.

17:52.600 --> 17:59.600

Den får meg til å stoppe opp fordi den er så snårlig og rar og fin eller stygg eller et eller annet som er veldig spesielt for den.

17:59.600 --> 18:10.600

Kinofilmen som får meg til å gå på kino er jo ikke hverken gjennomsnitts- eller medianlivet som haker meg inn.

18:10.600 --> 18:16.600

Det er når det er noe langt utenfor, men det kan være på ganske ...

18:16.600 --> 18:24.600

Det skal ikke så mye til før jeg ... Før hjernen min sier, ja, dette var noe ... Hæ? Vent litt, hva er det som skjer her?

18:24.600 --> 18:30.600

Noe som avviker litt fra grøten i midten fanger i hvert fall oppmerksomheten min veldig lett.

18:30.600 --> 18:33.600

Litt for lett, vil jeg konkludere med algoritmen.

18:33.600 --> 18:43.600

Jeg kaster bort mye tid på de uteliggende, som ligger utenfor den store massen i midten.

18:43.600 --> 18:46.600

Sånn sett er det litt utaktnemmelig med statistikk.

18:46.600 --> 18:51.600

For dette handler jo om å beskrive det typiske, på det midterste treet og alt dette.

18:51.600 --> 18:55.600

Men tror du vi kunne klart det som samfunn uten tall og statistikk?

18:55.600 --> 19:03.600

Jeg tror vi hadde klart en del ting, det hadde tatt oss en del ti år tilbake.

19:03.600 --> 19:12.600

Det hadde vært vanskelig å oppsummere hvilke medisiner og behandlinger som virker, fordi vi hadde blitt lurt av våre egne oppfatninger.

19:13.600 --> 19:25.600

Det er så mange ting som spiller inn, at jeg tror vi hadde, helsemessig og overlevnessmessig, tatt mange steg tilbake uten statistikk.

19:25.600 --> 19:32.600

Underholdningsmessig hadde vi tatt mange steg tilbake uten statistikk, og samfunnet generelt.

19:32.600 --> 19:37.600

Nå prøvde jeg å lete etter et område som hadde vært samme. Matlaging! Matlaging hadde sikkert vært det samme.

19:38.600 --> 19:48.600

Det er jo søren meg mye innenfor matlaging, og hvordan vi tilbereder og konserverer, som også er bygget på statistikk.

19:48.600 --> 19:56.600

Så nå sliter jeg litt med å finne noen områder i livet som ikke ville vært påvirket hvis jeg hadde fjernet, tatt bort statistikken.

19:56.600 --> 19:58.600

Det blir jeg glad for å høre, selvfølgelig.

19:58.600 --> 19:59.600

Selvfølgelig.

19:59.600 --> 20:02.600

Men hva er ditt verste statistikkøyeblikk?

20:03.600 --> 20:16.600

Jeg har jo tatt statistisk analyse på universitetet, og synes jo at noen av eksemplene var veldig gode og lærerike,

20:16.600 --> 20:19.600

og følte at jeg hadde fått en ny superpower i livet.

20:19.600 --> 20:24.600

Men det var også noen frustrerende, rive seg i håret øyeblikk der, helt klart.

20:24.600 --> 20:28.600

Jeg på en måte ikke forstå hvordan det hang sammen helt.

20:28.600 --> 20:36.600

Og det var veldig frustrerende, men jeg tror det matcher den gleden da det til slutt løsnet, og du fikk den der

20:36.600 --> 20:40.600

«Ja, nå!» faller liksom brikkene plass.

20:40.600 --> 20:45.600

Og det tok noen ganger veldig lang tid, flere uker, ikke sant?

20:45.600 --> 20:52.600

For jeg satt jo ikke hver dag, men jeg satt kanskje i over noen uker, så satt jeg kanskje to-tre ganger i uka,

20:52.600 --> 20:59.600

og liksom prøvde å komme meg igjennom oppgaver og greier, men jeg skjønnte liksom ikke hva er greia her.

20:59.600 --> 21:04.600



Jeg kan få oppskriften, siden man kan hjelpe meg litt, eller jeg kunne få en veilede til å hjelpe meg,

21:04.600 --> 21:06.600  
men skjønnte ikke helt hva det var.

21:06.600 --> 21:11.600  
Og så til slutt da, etter all denne frustrasjonen, så får du den der «Ja!»

21:11.600 --> 21:12.600  
«Sånn ja!»

21:12.600 --> 21:13.600  
«Ja!»

21:13.600 --> 21:23.600  
Ok, så hvis jeg bare gjør «Ja!» og derfor, ja, så får du den der deilige forløsende læringsleden,

21:23.600 --> 21:26.600  
og du oppdager at noen brikke faller på plass.

21:26.600 --> 21:35.600  
For å oppsummere en samling med kontinuerlige tall, så er det først det vi trenger et tall som sier noe om det typiske i tallene,

21:35.600 --> 21:38.600  
sånn som gjennomsnittet eller median.

21:38.600 --> 21:41.600  
Men vi trenger også et tall som sier noe om variasjonen.

21:41.600 --> 21:45.600  
Det er forskjell på om alle tallene vi har samlet inn er nesten helt like,

21:45.600 --> 21:51.600  
eller om det er stor spredning, slik at gjennomsnittet kanskje ikke er fullt så representativt for alle tallene,

21:51.600 --> 21:53.600  
som vi ofte gjerne har lett for å tenke.

21:53.600 --> 21:59.600  
Det finnes en rekke forskjellige oppsummeringstall som på ulikt vis kan oppsummere tall vi har samlet inn.

21:59.600 --> 22:01.600

Men er det egentlig så farlig?

22:01.600 --> 22:07.600

Hvor galt kan det egentlig gå om man velger feil oppsummeringstall for å beskrive et eller annet fenomen med statistikk?

22:07.600 --> 22:12.600

Det kan gå helt styrgalet. Da gir du et feil inntrykk av det du har sett.

22:12.600 --> 22:18.600

Og hvis det du presenterer som fakta blir brukt for å ta viktige beslutninger, så kan det også gå spesielt dårlig.

22:18.600 --> 22:26.600

Og tenk om økonomiske investeringer eller helse spørsmål, for eksempel liv og død, ble basert på at du oppgav feil oppsummeringstall.

22:26.600 --> 22:35.600

Hvis man velger å beskrive et tallmateriale på feil måte, så vil man jo gi et feilaktig inntrykk av hva det tallmateriale består av.

22:35.600 --> 22:42.600

Det er jo et litt klassisk eksempel at hvis man for eksempel oppgir et gjennomsnitt i motsetning til en median,

22:42.600 --> 22:47.600

i et tallmateriale hvor man har noen ekstremverdier, altså enten veldig lave eller veldig høye verdier,

22:47.600 --> 22:51.600

så vil det jo gi et skjevt svar.

22:52.600 --> 22:56.600

Om det ikke er direkte og lyg, så er det i hvert fall veldig, veldig nært og lyg.

22:56.600 --> 22:58.600

Ja, og lyve, det vil vi jo ikke.

22:58.600 --> 23:02.600

Tallene kan ikke snakke selv, så vi må snakke for dem.

23:02.600 --> 23:07.600

Og da er det den historien som tallene faktisk forteller, vi bør formidle.

23:07.600 --> 23:10.600

Ikke den historien vi selv har lyst til å fortelle.

23:10.600 --> 23:14.600

Det kalles dikning, og er noe annet enn statistisk analyse.

23:15.600 --> 23:20.600

I denne første episoden av Nakne tall har vi snakket om univariatanalyse,

23:20.600 --> 23:26.600

altså det å beskrive én variabel, et fenomen i den virkelige verden, med tall.

23:26.600 --> 23:29.600

I neste episode tar vi det ett skritt videre.

23:29.600 --> 23:36.600

Heller enn å bare se på én ting om gangen, skal vi snakke om det å studere sammenhengen mellom to tallstørrelser.

23:36.600 --> 23:38.600

For det er jo gjerne ofte det vi vil, ikke sant?

23:38.600 --> 23:41.600

Å avdekke hvordan verden rundt oss faktisk henger sammen.

23:42.600 --> 23:46.600

Jeg heter Jo Røislien og er professor i statistikk.

23:46.600 --> 23:52.600

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

23:52.600 --> 23:58.600

mens ekspertpanelet bestod av Kathrine Frey Frøslie, Mette Langeås, Marianne Riksheim og Jan Terje Kvaløy.

23:58.600 --> 24:02.600

Produsent for denne podcasten har vært Mari Linn Atterås Larsen og teknikere

24:02.600 --> 24:06.600

Elizabeth Mary Armstrong, Arne Thomas Lund Nilsen og Sebastian Fitjar.

24:06.600 --> 24:10.600

Nakne tall er produsert av Universitetet i Stavanger.