



Verdens første p-verdi

MEDISIN OG TALL

KRISTOFFER HERLAND HELLTON

E-post: kristohh@math.uio.no

Kristoffer Herland Hellton (f. 1987) er postdoktor i statistikk ved Matematisk institutt, Universitetet i Oslo, og forsker ved Norsk regnesentral.

Forfatter har fylt ut ICMJE-skjemaet og oppgir ingen interessekonflikter.

JO RØISLIEN

Jo Røislien (f. 1972) er professor i medisinsk statistikk ved Institutt for helsefag, Universitetet i Stavanger, og førsteamanuensis II ved Institutt for matematiske fag, Norges teknisk-naturvitenskapelige universitet. Han er profilert fagformidler og programleder på TV.

Forfatter har fylt ut ICMJE-skjemaet og oppgir ingen interessekonflikter.

En p-verdi er ikke det du tror den er.

Chriftened.			Chriftened.		
Anno.	Males.	Females.	Anno.	Males.	Females.
1629	5218	4683	1648	3363	3181
30	4858	4457	49	3079	2746
31	4422	4102	50	2890	2722
32	4994	4590	51	3231	2840
33	5158	4839	52	3220	2908
34	5035	4820	53	3196	2959
35	5106	4928	54	3441	3179
36	4917	4605	55	3655	3349
37	4703	4457	56	3668	3382
38	5359	4952	57	3396	3289
39	5366	4784	58	3157	3013
40	5518	5332	59	3209	2781
41	5470	5200	60	3724	3247
42	5460	4910	61	4748	4107
43	4793	4617	62	5216	4803
44	4107	3997	63	5411	4881
45	4047	3919	64	6041	5681
46	3768	3395	65	5114	4858
47	3796	3536	66	4678	4319

Figur 1 Utdrag fra det originale datasettet til John Arbuthnot (1667-1735) (2)

Statistikk handler om å kvantifisere usikkerhet. Å tallfeste hvor sikre, eller usikre, vi er på noe. Hvis vi for eksempel observerer en forskjell i måleverdier mellom to grupper, blir det statistiske spørsmålet: Er forskjellen, når vi tar hensyn til usikkerheten, reell eller skyldes den bare tilfeldigheter? Den vanligste måten å besvare dette spørsmålet på er med statistikkens kanskje mest utskjelte størrelse: p-verdien.

Bruken av p-verdier, og hva p-verdien betyr, har fått økende oppmerksomhet og kulminerte i 2016 med at American Statistical Association gjorde noe de nesten aldri gjør. De kom med en uttalelse, om hva p-verdien er – og ikke er – og hvordan den skal brukes (1).

En vanlig forståelse av p-verdien er at den er «sannsynligheten for at resultatet kan ha oppstått tilfeldig». Men dette er fundamentalt galt og en av misforståelsene American Statistical Association (ASA) ville til livs: Hva er feil med utsagnet – og hvorfor er det problematisk?

Sjelden hendelse?

I 1710 satt den matematikkinteresserte legen John Arbuthnot (1667-1735) og tenkte på babyer. Kom gutte- og jentebabyer virkelig til verden med like stor sannsynlighet? Eller var Guds forsyn slik at forholdet mellom kjønnene måtte være noe annet enn 50/50? Arbuthnot trålte kirkebøkene og talte opp antall døpte babyer i London fra 1710 og tilbake til 1629, i alt 82 år med data (fig 1).

Antallet nyfødte varierte fra år til år, og antallet guttebabyer og jentebabyer var heller aldri nøyaktig likt. Det var som forventet. Det bemerkelsesverdige var at det i alle de 82 årene ble født flere guttebabyer enn jentebabyer. Men Arbuthnot ville ikke trekke forhastede slutninger. Denne tilsynelatende systematiske forskjellen kunne fortsatt skyldes naturlig variasjon og tilfeldigheter. Så hvordan kunne han avgjøre om tallene talte for at Gud hadde en finger med i spillet?

Arbuthnot tenkte seg om. Og kom frem til følgende: Det er bare to mulige utfall. Enten er forholdet mellom kjønnene nøyaktig 50/50, og vi har simpelthen observert en sjelden hendelse, eller så er ikke forholdet mellom kjønnene 50/50. Om nullhypotesen er at sjansen for å føde en gutt er 1/2, så kan vi regne ut sannsynligheten for å observere 82 år på rad med flere guttebabyer enn jentebabyer:

$$\left(\frac{1}{2}\right)^{82} = \frac{1}{4\,836\,000\,000\,000\,000\,000\,000\,000} \approx 0,000000000000000000000002$$

Dette er verdens første p-verdi. Den er ekstremt liten. Sannsynligheten er omtrent like stor som å finne et helt bestemt, men tilfeldig, sandkorn i Sahara, og Arbuthnot konkluderte med at en 50/50-fordeling var urimelig (2).

Det er over 300 år siden Arbuthnots utregning, men den matematiske definisjonen av en p-verdi er uforandret. P-verdien er ikke sannsynligheten for at sannheten er 50/50, men sannsynligheten for å få akkurat det resultatet vi har fått *hvis* sannheten er 50/50. Nullhypotesen er en del av p-verdien.

Et målebånd

Det er mai, og du sitter inne med gardinene trukket for. Du lurer på om du skal ta deg en tur ut, men orker ikke å gå bort til døren og kikke ut. Det er klart, hvis det regner, gidder du ikke å gå ut. Men hvis det er sol, kunne det vært fint med en luftetur! Det er bare to muligheter. Enten er det sol ute, eller så er det ikke sol. Det er mai, og da er det gjerne fint vær, så nullhypotesen er at det *er* sol ute, mens alternativet er at det *ikke* er det.

Heldigvis kommer data gående inn døren. Kjæresten din kommer inn i rommet. Klissvåt. Nå har du både en hypotese, og du har data. Helst vil du at p-verdien skal fortelle deg sannsynligheten for at det er sol ute. Men det går ikke. Det p-verdien *kan* si, og sier, er sannsynligheten for at kjæresten din er klissvåt hvis det faktisk er sol ute. Og den p-verdien bør være liten.

I uttalelsen til American Statistical Association står det at "P-values do not measure the probability that the studied hypothesis is true, or the probability that the data were produced by random chance alone. It is a statement about data in relation to a specified hypothetical explanation, and is not a statement about the explanation itself."

P-verdien er ikke et orakel som avslører om vi har funnet sannheten, men et målebånd som måler dataene eller bevisene opp mot én bestemt hypotese.

LITTERATUR:

1. Wasserstein RL, Lazar NA. The ASA's statement on p-values: context, process, and purpose. *Am Stat* 2016; 70: 129 - 33.
 2. Arbuthnot J. An argument for divine providence, taken from the constant regularity observed in the births of both sexes. *Philos Trans R Soc Lond* 1710; 27: 325 - 36.
-

Publisert: 26. juni 2017. Tidsskr Nor Legeforen. DOI: 10.4045/tidsskr.17.0332
© Tidsskrift for Den norske legeforening 2020. Lastet ned fra tidsskriftet.no