

Statistiek voor tandartsen (II)

Steekproeven

Samenvatting

In onderzoek wordt vaak een steekproef gebruikt, omdat het niet mogelijk – en ook niet nodig – is de gehele populatie te onderzoeken. Welke waarde mag men hechten aan conclusies op grond van steekproefbevindingen en hoe komen die conclusies tot stand? Hier is beschreven hoe een steekproef kan worden getrokken, waarna uiteengezet wordt wat het begrip 'hypothese' inhoudt en wat bedoeld wordt met verwerpen van een hypothese. Twee soorten (typen) fouten zijn bij het verwerpen van belang, namelijk een hypothese ten onrechte verwerpen en een hypothese ten onrechte niet verwerpen. Kort wordt het belang van de omvang van de steekproef besproken. Het al dan niet verwerpen van een hypothese berust op de berekening van overschrijdingskansen.

SCHUURS AHB, DUIVENVOORDEN HJ, VAN 'T HOF MA. Statistiek voor tandartsen (II). Steekproeven. Ned Tijdschr Tandheelkd 1990; 97: 418-20.

A.H.B. Schuurs, tandarts*
H.J. Duivenvoorden, methodoloog**
M.A. van 't Hof, statisticus***

Uit de *)vakgroep Cariologie en Endodontologie van het Academisch Centrum Tandheelkunde Amsterdam (ACTA), **)vakgroep Medische Psychologie en Psychotherapie, Faculteit der Geneeskunde van de Erasmus Universiteit te Rotterdam en ***)de Medisch Statistische Afdeling van de Faculteit der Geneeskunde en Tandheelkunde van de Katholieke Universiteit te Nijmegen.

Trefwoord: Statistiek

Datum van acceptatie: 1 augustus 1990.

Adres: A.H.B. Schuurs, Louwesweg 1, 1066 EA Amsterdam.

1 INLEIDING

Een populatie is een welomschreven groep, bijvoorbeeld alle Nederlanders, maar ook de Belgische tandartsen, alle patiënten met een volledige prothese in Groningen of de occlusale amalgaamvullingen bij vierdeklassers van een bepaalde school. Een populatie is een verzameling van duidelijk gedefinieerde eenheden (aangeduid als 'elementen') waarop men conclusies van een onderzoek wil betrekken.

Het is gewoonlijk niet zinvol, te duur en tijdrovend de gehele populatie te onderzoeken. Wil men iets weten over een populatie, dan volstaat een onderzoek aan een deel daaruit, een steekproef, mits een getrouwe afspiegeling van die populatie is gewaarborgd. Met afspiegeling wordt bedoeld dat men niet in de bevindingen bij bijvoorbeeld 100 onderzochten is geïnteresseerd, maar een uitspraak wenst over de populatie: men wil de statistische bevindingen van de steekproef generaliseren naar de populatie. Overigens wordt onderscheid gemaakt tussen 'onderzoekspopulatie', bijvoorbeeld Nederland, en 'doelpopulatie', bijvoorbeeld een uitgekozen stad, waaruit dan een steekproef wordt getrokken. De statistiek kan nu worden gebruikt om te generaliseren naar de onderzoekspopulatie. Voor verdere generalisatie naar de onderzoekspopulatie is de statistiek niet toereikend en moet men een beroep doen op gezond verstand, ervaringen, theorieën, enzovoorts.

1.1 Variabiliteit

In een steekproef is altijd sprake van toevalsfluctuaties, van steekproefvariabiliteit. Vanwege deze toevalsbevindingen (Eng.: sampling errors) is bij generalisering

enige reserve nodig, juist omdat een steekproef niet voor 100% de eigenschappen van de populatie weergeeft en zelfs toevallig een onjuiste weergave ervan kan zijn. Door een adequate steekproef (paragraaf 2) van de juiste grootte (paragraaf 4) te nemen wordt het risico van foutieve generalisering beperkt.

2 HOE EEN STEEKPROEF SAMEN TE STELLEN

2.1 Verkeerde steekproef

Een onderzoek naar tandartsbezoek van de Nederlandse populatie kan niet in de Kalverstraat worden gedaan. Daar zal men vermoedelijk relatief veel vrouwen en weinig mannen aantreffen en meer stedelingen dan plattelanders, enzovoorts. Kortom, zulk een 'steekproef' weerspiegelt niet de Nederlandse bevolking; er zal waarschijnlijk vertekening door selectie zijn. Bovendien bestaat de kans dat de enquêteur de gehaaste voorbijganger niet staande wil houden en personen met een afwijzende houding over wil slaan. De enquêteur gaat dan niet aselekt te werk. Generaliseren is door deze inadequate steekproef niet mogelijk.

2.2 Aselecte en bijzondere steekproeven¹

Selectie in de steekproef moet dus worden uitgeschakeld. Dat kan, formeel, aan de hand van tabellen met reeksen in willekeurige volgorde staande getallen, waarmee men bijvoorbeeld uit het bevolkingsregister het gewenste aantal namen kiest. Bij het trekken van zulk een steekproef moet elk element van de gedefinieerde populatie (bijvoorbeeld alle 18-50-jarigen van de on-

derzoeksstad) bij elke trekking een *even grote kans* hebben om in de steekproef terecht te komen: de steekproef moet *aselect* (Eng.: random) zijn. Bovendien moeten de steekproefelementen onafhankelijk worden getrokken. Als iemand in de steekproef geloot wordt, mogen bijvoorbeeld zijn huisgenoten geen verhoogde of verlaagde kans hebben ook in de steekproef te komen. Toch kan ook op deze manier een probleem ontstaan. Wil men bijvoorbeeld vullingen evalueren, dan worden meestal personen getrokken, dat wil zeggen men gaat van de ene eenheid (vullingen) over op een andere (mensen), hetgeen een (selectie)probleem vormt.

Als de elementen van de populatie genummerd zijn, kan men ook uit de eerste *m* elementen aselekt één element trekken en vervolgens elke *m-de* element. Is *m* bijvoorbeeld gelijk aan 11 (dat wil zeggen, men wil een steekproef van 9% van de populatie) en is het aselekt getrokken element uit de eerste 11 personen nummer 9, dan neme men als eerste element nummer 9, als tweede 911, als derde 91111, enzovoorts.

Soms is het nuttig een aselekte steekproef (random sample) uit een *deel* van de populatie te nemen. Is de populatie op voorhand ingedeeld op kenmerken zoals geslacht, en wil men alleen iets weten over vrouwen, dan volstaat een aselekte steekproef uit de subpopulatie vrouwen. Een subpopulatie is in statistische zin een gewone populatie.

– *Gelede (gestratificeerde) aselekte steekproef.* Als wordt verwacht dat een bepaald kenmerk van de populatie, bijvoorbeeld leeftijd, de relatie tussen bijvoorbeeld tandenpoetsen en parodontitis beïnvloedt, dan is het vanwege de efficiëntie aan te bevelen subpopulaties naar leeftijd te onderscheiden (men 'stratificeert' voor leef-

tijd) en uit iedere subpopulatie een aselechte steekproef te trekken. Zo kan het 'effect' van tandenpoetsen met leeftijd efficiënt worden nagegaan.

De samenstelling van een steekproef hangt samen met hoe men een onderzoek opzet (Eng.: design).

3 CONCLUSIE UIT STEEKPROEF-GEGEVENS

Meestal wil men op grond van de verkregen gegevens uit een steekproef weten of een bepaalde veronderstelling juist is en voor de populatie opgaat. Statistische toetsen zijn geformaliseerde procedures, met bepaalde gevolgtrekkingen (inferenties) voor de populatie. Het probleem van de steekproef is dat deze conclusies oplevert die van de werkelijkheid kunnen afwijken. Een tweede steekproef zou hoogstwaarschijnlijk niet exact dezelfde uitkomsten opleveren en kan leiden tot andere conclusies. Bij het trekken van conclusies verdienen twee typen fouten nadere aandacht; zij worden besproken nadat eerst iets over hypothesen is gezegd.

3.1 Hypothesen

In de toetsende statistiek gaat men na of er een relatie, verband, effect enzovoorts bestaat. Dat gebeurt door de hypothese, waarin het *tegengestelde* van het verwachte wordt geponeerd, te verwerpen. De te verwerpen hypothese heet nulhypothese (aangeduid als H_0) en de andere wordt alternatieve hypothese (H_1) genoemd en vertegenwoordigt de onderzoekshypothese. Indien uitgaande van H_0 de onderzoeksgegevens onwaarschijnlijk vreemde resultaten laten zien, wordt H_0 verworpen en H_1 geaccepteerd. De waarschijnlijkheid (de 'kans') van de vreemde bevindingen is gelijk aan de P-waarde (P van Probability). Als de P-waarde onder een bepaalde grens (onbetrouwbaarheidsdrempel) verdwijnt, wordt het resultaat significant genoemd. Ter illustratie van H_0 en H_1 een voorbeeld.

Een onderzoeker wil de relatie tussen cariës en fluoride bestuderen. 1. Als vóór het onderzoek onbekend is of fluoride cariës bevordert of remt, kan H_1 luiden: 'er bestaat een relatie tussen cariës en fluoride'. In H_0 wordt dan gesteld dat er 'geen relatie bestaat'. Uit onderzoek zou kunnen blijken dat alle mensen die fluoride gebruiken zeer weinig cariës hebben en anderen erg veel. Als op grond van statistische toetsing H_0 wordt verworpen, wordt geconcludeerd dat beide variabelen zijn gerelateerd. Vervolgens kan worden nagegaan hoe die relatie ligt. 2. Als op grond van de literatuur en/of eerdere waarnemingen het idee bestaat dat fluoride cariës tegengaat, dan kan H_1 luiden: 'de DMFT index van perso-

nen die fluoride gebruiken is lager dan van personen die niet aan fluoride zijn blootgesteld'. Van belang is dat de formulering onder 1. *geén richting* aangeeft en de formulering onder 2. *wél*.

Het al dan niet aanwezig zijn van een richting in de formulering bepaalt of respectievelijk één- en tweezijdig wordt getoetst en beïnvloedt daarmee de kans op verwerping van H_0 . Bij eenzijdige toetsing kan hoogstens een foutieve conclusie in één richting worden gemaakt. Dit houdt in dat bij eenzijdige toetsing eerder een significant resultaat wordt behaald dan bij tweezijdige toetsing. Overigens, eenzijdige toetsing is vaak discutabel en wordt alleen dan toegepast als het als het ware dwingend naar voren komt. De lezer zal beide keuzen in de vakliteratuur tegenkomen.

3.2 Type I fout (α fout) en verwerpen nulhypothese²

Statistische toetsing is een methode om rekening houdend met de steekproefvariabiliteit tot uitspraken te komen. H_0 wordt (als 'waar') geaccepteerd indien statistische toetsing een hoge P-waarde oplevert. De onderzoeker stelt (tevorens) vast hoe groot de kans mag zijn dat H_0 ten onrechte wordt verworpen ten gevolge van de steekproefvariabiliteit. De kans om H_0 ten onrechte te verwerpen wordt klein gehouden (in feite hebben de wetenschappelijke tijdschriften in deze het laatste woord) en is vaak op $\alpha = 0,05$ gesteld. Dit laatste betekent dat als een onderzoek honderd keer wordt uitgevoerd, H_0 op grond van de bevindingen in gemiddeld vijf onderzoeken wordt verworpen, waar dat in feite niet had moeten. Ernstige gevolgen, bijvoorbeeld in geneesmiddelenonderzoek, van H_0 ten onrechte verwerpen, kunnen reden zijn α op 0,01 of lager te stellen. Het is duidelijk dat de kans op onterechte verwerping van H_0 minder groot is bij $\alpha = 0,01$ dan bij $\alpha = 0,05$.

De gevonden kans P op incorrecte ver-

werping van H_0 wordt doorgaans precies vermeld. P is minimaal 0 en maximaal 1. De term 'significante bevinding' houdt in dat H_0 is verworpen met een bepaalde kans dat die verwerping fout is.

3.3 Type II fout (β fout)

Een tweede mogelijkheid is dat H_0 ten onrechte *niet* wordt verworpen, met andere woorden H_1 moest eigenlijk worden geaccepteerd. De grootte van deze tweede fout, β , is onbekend en hangt af van het werkelijke verschil. De tweede fout wordt nogal eens van minder belang geacht dan de eerste. In feite hangt het van de onderzoeksdoelstelling af of het ten onrechte verwerpen van H_0 erger is dan het ten onrechte niet verwerpen van H_0 . Dat geldt bijvoorbeeld voor een nieuw geneesmiddel dat meer bijwerkingen heeft dan een oud. Als H_0 , behelzende dat er geen verschil in bijwerkingen tussen oud en nieuw middel bestaat, ten onrechte niet wordt verworpen, ondervinden patiënten schade. Maar als in dit voorbeeld H_0 ten onrechte wordt verworpen, dan raakt de geneeskunde achterop, hetgeen meestal minder ernstig is dan mensen schade berokkenen.

Samenvattend, α is de kans op valse significantie en β is de kans op valse non-significantie. In tabel I is de relatie tussen de werkelijkheid en het verwerpen van H_0 nog eens verbeeld.

De zojuist besproken twee typen fouten zijn sterk verwant met twee bekende begrippen in de geneeskundige diagnostiek. De *sensitiviteit* van een diagnostische test houdt in dat met de test een ziek individu inderdaad als 'ziek' wordt gediagnostiseerd en een gezonde als 'gezond' (= *specificiteit*).³ Sensitiviteit heeft te maken met 'valse negatieven' en specificiteit met 'valse positieven'. Dit onderwerp valt verder buiten het bestek van deze artikelenserie.

Tabel I. Relatie tussen de werkelijkheid en de statistische toetsing resultaten (naar Nyman & Pearson, 1939, geciteerd door Bolle c.s., 1974)⁵

		H_0 is in werkelijkheid	
		Juist (= H_1 geaccepteerd)	Onjuist
H_0 wordt op grond van de steekproef:	Verworpen*)	α fout (type I)	correct verworpen
	Niet verworpen	correct geaccepteerd	β fout (type II)

H_0 verwerpen impliceert dat H_1 wordt geaccepteerd.

4 OMVANG VAN EEN STEEKPROEF

Hoe groter een steekproef, hoe nauwkeuriger in het algemeen de informatie over de populatie. Een steekproef kan groter zijn dan nodig, dat wil zeggen dat met een kleiner aantal de benodigde informatie ook was verkregen. Belangrijker nog is dat een (te) grote steekproef zeer kleine, onbelangrijke verschillen aan het licht kan brengen, die dan statistisch significant zijn. Nadeel van een te kleine steekproef is dat men te weinig informatie kan krijgen om uitspraken over de populatie te doen, met andere woorden het zou wel eens kunnen dat relevante bevindingen worden gemist. Een niet-significant resultaat is alleen dan informatief als de steekproefomvang voldoende groot is.

Het is gewenst de steekproefgrootte vooraf te bepalen, door rekening te houden met type I fout (α) en type II fout (β) en met de grootte van het 'effect', dat nog als klinisch of epidemiologisch relevant wordt geacht. Daarbij wordt α op bijvoorbeeld

0,05 gesteld en wordt voor β , die onbekend is, een zekere waarde aangenomen, bijvoorbeeld $\beta = 0,20$. Heeft men reden om β kleiner te kiezen, bijvoorbeeld 0,10, dan

wordt de steekproefomvang aanmerkelijk groter. Hier wordt volstaan met een verwijzing naar elders voor de wijze van uitvoering.⁴

SUMMARY

SAMPLES

Key word: Statistics

Sampling methods and the consequences thereof in regard to the value of the conclusions drawn from the results are discussed. The term 'hypothesis' is clarified, followed by a description (of the importance) of Type I and II errors in statistical testing, which is based on P-values. In a general sense, the sample size is discussed.

LITERATUUR

- ¹DE JONGE H, RUMKE CL, VAN STRIK R. Medische Statistiek. Vakgroep Medische Statistiek, Vrije Universiteit te Amsterdam, 1984.
- ²SIEGEL S, CASTELLAN NJ. Nonparametric statistics. New York: McGraw-Hill book company, 1988: 11.
- ³HAYNES RB. Hoe moeten medische tijdschriften worden gelezen? II. Het beoordelen van een diagnostische test. Ned Tijdschr Geneesk 1983; 127: 2331-7.
- ⁴COHEN L. Statistical power analysis for the behavioral sciences. New York: Academic Press, 1977: par. 1.4.
- ⁵BOLLE EAW, LENOIR JHM, VAN LOON JNM. Wiskundige statistiek. Deventer: Kluwer, 1974: 19.

Bladvulling

Intraveneuze sedatie

Het risico voor de tandarts

Tandheelkundige behandelingen worden door veel patiënten als onplezierig ervaren. Door het tegenwoordige gebruik van benzodiazepines is deze algehele anesthesie vaak niet meer nodig en wordt de behandeling een minder vervelende en veiliger ervaring. Toch kunnen ook deze middelen ongelukkige neveneffecten veroorzaken. Van diazepam, fentanyl en methohexitone worden in verschillende studies verschijnselen als hysterische reacties, hallucinaties en seksuele opwindning gemeld. Uit voorbeelden van rechtzaken in Groot-Brittannië, blijkt dat dit ook een risico inhoudt voor de tandarts; het risico van carrièreverlies en verlies van persoonlijke achting. Als gevolg van een opzienbare zaak in Manchester waar een tandarts schuldig werd bevonden aan het aanranden van patiënten terwijl ze onder invloed waren van sedativa, werd een schriftelijke enquête gehouden in de medische en tandheelkundige pers. De incidenten die gerapporteerd werden van behandelingen met kalmerende middelen melden o.a. seksuele opwindning waaronder masturbatie. De klachten van de patiënten liepen uiteen van het naakt zien van de tandarts tot het idee hebben

verkracht te zijn. Opvallend is dat in bijna alle gevallen tandarts en patiënt van een verschillend geslacht zijn. Een aantal suggesties om problemen te voorkomen zijn:

1. Overweeg of de toediening van kalmerende middelen het risico waard is.
2. Overweeg om alle vrouwelijke patiënten die kalmerende middelen vragen naar een vrouwelijke collega te verwijzen.
3. Ben er zeker van dat de tandartsassistent(e) die als getuige fungeert, aanwezig is vanaf het moment van de toediening tot het moment dat de patiënt de praktijk verlaat. Zorg ervoor dat de naam van de assistent(e) vermeld wordt in de administratie van de patiënt.
4. Overweeg een familielid de procedure te laten observeren en de hele tijd bij de patiënt te laten blijven.

Op 13 maart 1990 ging ook Prof. Dunning in een artikel in de NRC op deze kwestie in. Hij benadrukt in dit artikel 'dat er alle reden blijft om de klachten van vrouwen over seksuele toenadering van hulpverleners volstrekt serieus te nemen en de toch al hoge drempel niet verder op te trekken

door scepsis en wantrouwen'.

Hij beaamt echter ook 'dat het mogelijk is dat seksuele fantasieën, al of niet geïnduceerd door bewustzijnsverlagende farmaca, een bescheiden rol spelen in de klachten'. Ook hij geeft als remedie tegen onterechte klachten 'een derde persoon, verpleegkundige, assistent of collega aanwezig te laten zijn'.

Verder gaat hij in zijn artikel in op de vraag of patiënten tijdens narcose kunnen waarnemen wat zich om hen afspeelt. 'Ongeveer 1% van alle patiënten heeft bewustzijnservaringen tijdens de narcose. De wisselende narcosediepte zet wellicht de poort van herinnering op een kier. De oppervlakkige roes van ingespoten benzodiazepines zou signalen, hun vervorming en herinnering zo kunnen omsmeden tot een erotische fantasie, een ontremde droom met een groot werkelijkheidskarakter bij het ontwaken'.

Mevr. M. van Kerkhoven-Weijs

Bron
FIELDS M. Intravenous sedation. The risk to the dentist. Br Dent J 1990; 169: 4-5.