MATHEMATISCH CENTRUM
AMSTERDAM

Statistische Afdeling
Rapport S 80
door
Dr J.Hemelrijk Jr

Maart 1952.

De invloed van vaatverwijdende middelen bij vaatonderbrekingen in de darmsector bij caviae.

Bij een aantal caviae werden op operatieve wijze bepaalde vaatonderbrekingen in de darmsector aangebracht. Het gevolg hiervan is gewoonlijk, dat het proefdier sterft, slechts zelden treedt spontaan herstel op. Er zijn aanwijzingen, dat het herstel bevorderd wordt door toediening van vaatverwijdende middelen. Om dit nader te onderzoeken werden de volgende experimenten verricht.

Aan een contrôlegroep van 20 dieren werden na de operatie geen medicamenten toegediend. Daarnaast werden aan 4 andere groepen direct na de operatie en op vaste latere tijden 4 verschillende medicamenten (één voor iedere groep) intramusculair ingespoten. Van ieder der dieren werd genoteerd of het aan de operatie stierf of zich hiervan herstelde. De resultaten van deze proeven zijn in tabel I samengevat.

Tabel I
Waarnemingsresultaten

experiment	n	hersteld	niet hersteld
A. Contrôle	20	2	18
B. 0,25 mgr acetylcholine	20	7	13
C. 0,01 mgr nitroglycerine	20	9	11
D. 0,2 mgr mechochyl	10	2	8
E. 0,25 mgr acetylcholine + 1 mgr neo-antergan	20	6	14

n = aantal dieren der proef.
De op een aantal tijdstippen na de operatie ingespoten stof staat bij ieder experiment vermeld.

Alle gebruikte middelen (behalve de bij experiment E toegevoegde neo-antergan) werken vaatverwijdend. Het ligt daarom voor de hand allereerst na te gaan of van deze groep middelen als geheel een genezende werking uitgaat. Daartoe nemen wij de experimenten B, C, D en E tezamen en verkrijgen dan tabel II.

Tabel II

	hersteld	niet hersteld	totaal
Contrôle	2	18	20
B, C, D en E	24	46	70
totaal	26	64	90

Met behulp van de methode der 2 x 2-tabel (vgl. de aan dit rapport toegevoegde memoranda S 47 (M 6) en S 53 (M 23)) kunnen wij nu de hypothese toetsen, dat de vaatverwijdende middelen het herstel niet beinvloeden. Exacte berekening van de bij deze toets behorende tweezijdige overschrijdingskans geeft de waarde 0,048. Op grond van een overschrijdingskans van deze grootte (of kleiner), gaat men gewoonlijk over tot verwerping van de getoetste hypothese (men gebruikt n.l. zeer vaak als onbetrouwbaarheidsdrempel de waarde 0,05). Deze verwerping berust dan op het principe, dat men, aldus handelende, gemiddeld in ongeveer één van de 20 gevallen een hypothese, die juist is, toch zal verwerpen. Volgens deze gedragslijn handelende zou men dus in dit geval tot een genezende werking van de gebruikte groep van vaatverwijdende middelen besluiten. Men dient zich hierbij echter bewust te zijn van het feit, dat de gevonden overschrijdingskans niet klein genoeg is om deze conclusie met enige mate van stelligheid te trekken. Indien er bij de contrôleproef één dier meer hersteld was, of indien er bij één der overige groepen één dier meer gestorven was, zou de overschrijdingskans groter dan 0,05 zijn geweest. Hieruit blijkt wel, dat het trekken van een scherpe grens nauwelijks verantwoord moet worden geacht en dat men, op grond van het bovenstaande experiment alleen, de genezende werking van de gebruikte groep vaatverwijdende middelen, niet als ondubbelzinnig aangetoond kan beschouwen. Het is echter wel gerechtvaardigd om van een redelijke aanwijzing in deze richting te spreken.

Vergelijken wij de experimenten B, C, D en E ieder apart met A, gebruikmakende van dezelfde exacte methode der 2 x 2-tabel, dan vinden wij de volgende tweezijdige overschrijdingskansen:

Voor A tegen B: 0,13;
Voor A tegen C: 0,033;
Voor A tegen D: 0,58;
Voor A tegen E: 0,24.

Hieruit ziet men, dat, zoals ook uit tabel I reeds blijkt, de inspuitingen der nitroglycerine de beste resultaten hebben gegeven (de overschrijdingskans is bij deze proef nl. verreweg het kleinst). Indien uitsluitend de experimenten A en C zouden zijn uitgevoerd, zou het resultaat zelfs overtuigender zijn geweest, dan nu met alle experimenten tezamen het geval is. Hoewel men dientengevolge kan vermoeden, dat de nitroglycerine in de eerste plaats in aanmerking komt om als eventueel werkzaam te worden beschouwd, blijkt dit vermoeden door verdere toetsing, op grond van de in tabel I vermelde resultaten, niet te kunnen worden bevestigd, hetgeen ten gevolge heeft, dat de genezende werking van de nitroglycerine apart niet als ondubbelzinnig aangetoond kan worden beschouwd. Vergelijken wij experiment C nl. met ieder der experimenten B, D en E, wederom met de methode der 2 x 2-tabel, dan blijken de daarbij verkregen uitkomsten geen aanleiding te geven om een verschil tussen de werkzaamheden van nitroglycerine en van één van de andere drie middelen aan te nemen. De kleinste bij deze 3 toetsingen gevonden overschrijdingskans wordt gevonden bij vergelijking van C met D en bedraagt 0,25. Deze uitkomst kan echter het gevolg zijn van het betrekkelijk geringe aantal proefdieren in ieder experiment. Het is zeer wel mogelijk, dat bij een proef met meer dieren wel een verschil zou zijn gevonden. Hoewel wij dus geen verschil tussen de middelen kunnen aantonen, wijst de uitslag van het experiment er wel op dat het gewenst is bij verdere onderzoekingen speciaal aandacht aan de nitroglycerine te besteden.

Samenvattend komen wij dus tot de volgende

Conclusie:

Het uitgevoerde experiment geeft een redelijke aanwijzing voor een genezende werking van de gebruikte groep van vaatverwijdende middelen. Hoewel bij het onderhavige experiment de beste resultaten werden behaald met herhaalde inspuitingen van 0,01 mgr nitroglycerine, kon een verschil in werking van de 4 middelen - wellicht ten gevolge van de kleine aantallen proefdieren - niet worden aangetoond. Echter moet toch wel rekening worden gehouden met de mogelijkheid, dat er van één of meer der gebruikte middelen geen genezenden werking uitgaat.

Scherpere conclusies kunnen slechts verwacht worden op grond van verdere onderzoekingen, waarbij het aanbevelendwaardig is speciaal de uitwerking van de nitroglycerine nader te beschouwen.

MATHEMATISCH CENTRUM
Amsterdam
Statistische Afdeling.

S47(M6)

Algemene gang van zaken bij het toetsen van een hypothese.[1]

De toetsing van een hypothese $\mathcal{H}_0$ berust steeds op een aantal waarnemingen $x_1, x_2, \ldots, x_n$ van één of meer stochastische grootheden[2], of op enige groepen van waarnemingen (bv. twee steekproeven).

Bij een toets behoort een toetsingsgrootheid $\underline{u}$ (soms meer dan één), die een functie is van bovengenoemde stochastische grootheden en die, voor de waargenomen waarden $x_1, x_2, \ldots, x_n$ een waarde aanneemt, die berekend kan worden (bv.: het gemiddelde der waarnemingen, of de spreiding, of het verschil van de gemiddelden van twee waarnemingen).

De toetsingsgrootheid wordt steeds zo gekozen, dat men, op grond van de onderstelling, dat $\mathcal{H}_0$ juist is, de waarschijnlijkheidsverdeling van deze grootheid kan berekenen.

Vervolgens kiest men een verzameling Z van mogelijke uitkomsten van $\underline{u}$, en wel op zodanige wijze, dat de kans, dat $\underline{u}$ een in Z gelegen waarde aanneemt, onder de hypothese $\mathcal{H}_0$, gelijk is aan een gegeven getal α, zodat Z dus van α afhankelijk is[3]. Z heet de kritieke zône van de toets, α de onbetrouwbaarheidsdrempel (Engels: level of significance). Voor α neemt men veelal de waarde 0,05 of 0,01.

Men verwerpt nu $\mathcal{H}_0$ op grond van de waarnemingen $x_1, x_2, \ldots, x_n$, indien de bij deze waarnemingen behorende waarde van $\underline{u}$ in Z ligt. Dit wordt vaak uitgedrukt door te zeggen, dat het resultaat van het experiment "significant" is. De waarde van α moet dan echter worden vermeld. De kans, dat dit zal gebeuren, is, indien $\mathcal{H}_0$ juist is, gelijk aan α. Derhalve is α de kans op ten onrechte verwerping van de juiste hypothese, ook de kans op een fout van de eerste soort genoemd. Indien men deze methode toepast, met α = 0,05 resp. 0,01, zal men in gemiddeld ongeveer één op 20 resp. op 100 van de gevallen, waarin de hypothese die men toetst juist is, deze toch verwerpen.

1) Dit memorandum is slechts bedoeld ter oriëntatie en streeft niet naar volledigheid of volledige exactheid.

2) Een stochastische grootheid is een grootheid, die een waarschijnlijkheidsverdeling bezit, of, anders gezegd, een grootheid, die voor de elementen van een collectie (universum, populatie) gedefinieerd is en daarop allerlei waarden aanneemt. Stochastische grootheden worden aangegeven door onderstreepte letters.

3) Soms kan men slechts bereiken, dat deze kans $\leq \alpha$ is.

De toetsingstheorie biedt in het algemeen geen mogelijkheid om tot aanvaarding van een hypothese te komen. Indien een bepaalde hypothese $\mathcal{H}_0$ niet verworpen kan worden, is dit gewoonlijk met een hele verzameling van hypothesen tegelijk het geval. Niet-verwerpen staat dus niet gelijk met aanvaarden.

Wel zal men vaak in de loop van een statistische analyse bepaalde onderstellingen, die plausibel schijnen en voor de verdere analyse van nut zijn, toetsen, alvorens ze bij de verdere bewerking van het materiaal te gebruiken. Worden zij dan op grond van de toets niet verworpen, dan houdt dit in zo verre een rechtvaardiging van die onderstellingen in, dat een grote afwijking door de toets veelal wel zou zijn ontdekt. Indien men dan verder de onderstellingen gebruikt, verwaarloost men eventueel aanwezige afwijkingen van onbekende grootte, die echter niet zo groot zijn, dat zij door de toets zijn ontdekt.

Vele toetsen gelden zelf alleen onder bepaalde onderstellingen omtrent de waarschijnlijkheidsverdelingen der stochastische grootheden, waarvan waarnemingen zijn verricht. Deze nevenvoorwaarden dienen steeds uitdrukkelijk te worden vermeld en, zo mogelijk, zelf te worden getoetst.

In plaats van de onbetrouwbaarheidsdrempel α wordt vaak bij de uitslag van een toetsing de overschrijdingskans k opgegeven; dit is de kleinste waarde van α, waarbij in het betrokken geval, nog tot verwerping van $\mathcal{H}_0$ zou zijn overgegaan; anders gezegd: de kleinste α, waarvoor de gevonden waarde der toetsingsgrootheid nog juist in de (bij α behorende) kritieke zône Z ligt. Wordt dus de waarde k opgegeven en werkt men met onbetrouwbaarheidsdrempel α, dan wordt verworpen, indien $k \leq \alpha$ is.

Voor het onderscheid tussen één- en tweezijdige toetsing en de keuze tussen deze twee mogelijkheden vergelijke men bv. de tweede hieronder gegeven litteratuurplaats. Wij moeten hier volstaan met de opmerking, dat éénzijdige toetsing veelal eerder tot verwerping van $\mathcal{H}_0$ leidt, maar dat deze slechts onder bijzondere omstandigheden kan worden toegepast.

Litteratuur:

J.Neyman, First course in probability and statistics, New York, 1950, Chapter 5.

J.Hemelrijk en H.R. van der Vaart, Het gebruik van één- en tweezijdige overschrijdingskansen voor het toetsen van hypothesen, Statistica 4 (1950) p.54-66.

MATHEMATISCH CENTRUM,
2de Boerhaavestr. 49,
A m s t e r d a m - O.

Statistische Afdeling.
S 53 (M 23).

Toetsing van de hypothese $p_1 = p_2$ met behulp van een 2 x 2-tabel[1].

Wij beschouwen twee reeksen van onafhankelijke experimenten, waarbij ieder experiment van de ene reeks één van de twee resultaten A of $\bar{A}$ (non-A) heeft en ieder experiment van de tweede reeks één van de beide resultaten B of $\bar{B}$ (hierbij kan A=B zijn). Daarbij wordt ondersteld, dat bij ieder der experimenten van de ene reeks de kans op A gelijk aan p_1 (en dus de kans op $\bar{A}$ gelijk aan $1-p_1$) is en bij ieder der experimenten van de tweede reeks de kans op B gelijk aan p_2 (en dus de kans op $\bar{B}$ gelijk aan $1-p_2$). De te toetsen hypothese luidt nu:

$$H_0: \; p_1 = p_2.$$

Indien de eerste reeks uit n en de tweede reeks uit m waarnemingen bestaat, waaronder n_1 (resp. m_1) maal A (resp. B) voorkomt, kunnen deze gegevens in de volgende 2 x 2-tabel worden samengevat:

	A resp. B	$\bar{A}$ resp. $\bar{B}$	totaal
eerste reeks	n_1	$n-n_1$	n
tweede reeks	m_1	$m-m_1$	m
totaal	r	n+m-r	n+m

Als toetsingsgrootheid wordt n_1, het aantal malen A in de eerste reeks waarnemingen, gebruikt. Indien H_0 juist is bezit deze grootheid onder de voorwaarde, dat r de bij het experiment gevonden waarde aanneemt, de volgende waarschijnlijkheidsverdeling: de kans, dat een bepaalde waarde n_1 aangenomen wordt, is gelijk aan:

$$\frac{\binom{n}{n_1}\binom{m}{m_1}}{\binom{n+m}{r}}$$

Als kritieke zône worden de waarden van n_1 met de kleinste waarschijnlijkheden bijeengezocht, tot de gekozen betrouwbaarheidsdrempel het toevoegen van een nieuwe waarde verhindert (bij éénzijdige toetsing bestaat de kritieke zône uitsluitend uit grote of uitsluitend uit kleine waarden van n_1).

---

1) Dit memorandum is slechts bedoeld ter oriëntatie en streeft niet naar volledigheid of volledige exactheid.

De overschrijdingskans, behorende bij de gevonden waarde van n_1, is gedefiniëerd als de som van alle waarschijnlijkheden van bovenstaande verdeling, die hoogstens gelijk aan de waarschijnlijkheid van de gevonden waarde zijn (bij éénzijdige toetsing echter gelijk aan de som van de waarschijnlijkheden van alle waarden die groter of gelijk aan de gevondene, of van alle waarden, die kleiner of gelijk aan de gevondene zijn). Deze exacte toetsingsmethode voor H_0 is afkomstig van R.A. Fisher.

Indien n en m zo groot zijn, dat deze exacte berekening te omslachtig wordt, maakt men gebruik van de volgende benadering:

Gemiddelde en spreiding van de grootheid n_1 zijn (indien H_0 juist is):

$$\frac{nr}{n+m} \quad \text{resp.} \quad \sqrt{\frac{nmrs}{(n+m)^2(n+m-1)}} \qquad s = n+m-r$$

Men gebruikt dan in plaats van de exacte waarschijnlijkheidsverdeling van n_1 de normale verdeling met hetzelfde gemiddelde en dezelfde spreiding en in plaats van de gevonden waarde van n_1 neemt men het getal, dat $\frac{1}{2}$ dichter bij het gemiddelde ligt dan deze gevonden waarde (dit laatste is de z.g. "continuïteitscorrectie", die bij toenemende n en m weldra verwaarloosd kan worden). Met behulp van de benadering gaat men dan verder te werk als boven beschreven, daarbij gebruik makende van een tabel van de normale verdeling.

<u>Litteratuur</u>:

R.A.Fisher, Statistical Methods for Research Workers, London 1948, p. 96. <u>Opmerking</u>: Fisher gebruikt hier de éénzijdige overschrijdingskans.

J.Hemelrijk, Waarschijnlijkheidsrekening en Statistiek, Vacantiecursus Mathematisch Centrum, Amsterdam 1950, § 4.