

Statistische Afdeling.
Rapport S 38, door
J. Hemelrijk.

Statistische aspecten van een onderzoek, naar
de erfelijkheid van tweelinggeboorten.

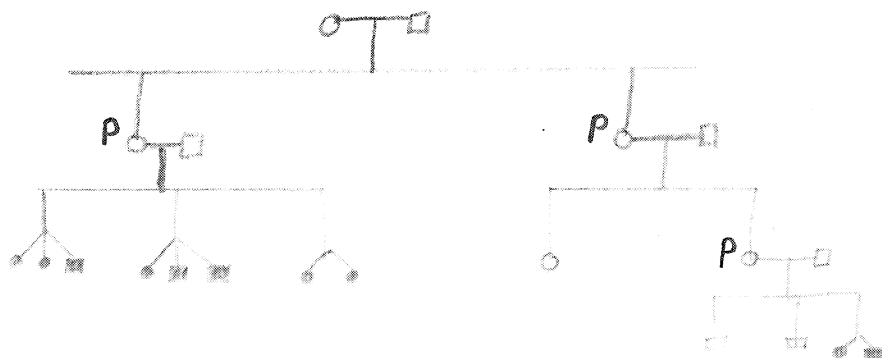
1. Inleiding. Het onderzoek heeft betrekking op dizygote tweelingen. Er zijn 72 min of meer uitgebreide stambomen beschikbaar, waarin minstens één dizygote tweeling voorkomt. Het materiaal wekt sterk de indruk, dat een eventuele erfelijke factor evenzeer in de man als in de vrouw aanwezig kan zijn; deze hypothese komt, op grond van de verzamelde stambomen, zeker niet voor verwerping in aanmerking. Met deze kwestie zullen wij ons verder in dit rapport niet bezighouden. In het volgende wordt bovendien de aanwezigheid van minstens één erfelijke factor aangenomen.

Het verdere onderzoek betreft de wijze van passage van de erfelijke factor door de stamboom. Hierbij wordt het begrip passage als volgt gedefinieerd:

Een passage door persoon A heeft plaats gevonden, indien 1^e twee van de directe nakomelingen van A dizygote tweelingen zijn, en

2^e een van A verschillende directe nakomeling van directe voorouders van A een dizygote tweeling heeft verkregen.

Beschouwen wij b.v. de volgende stamboom:



○ = vrouw

□ = man

●● = tweeling

P = passage

Passageschema: v v v

In deze stamboom hebben volgens bovenstaande definitie 3 passages plaats gevonden, alle door vrouwelijke personen.

Uit de 72 stambomen werden die stambomen gezocht, waarin minstens twee dizygote tweelingen voorkomen, en wel in minstens twee ver schillende gezinnen. Dit waren er 55.

Men kan nu op verschillende wijzen trachten een ~~uit-~~ *in-*zicht te verkrijgen in de wijze van passage van de erfelijke factor. Wij zullen deze methoden in de volgende punten nader beschouwen.

2. De onderzoekers hebben van ieder der 55 stambomen het passageschema opgesteld (voor bovenstaande stamboom is dit v v v) en verkregen onderstaande tabel.

Tabel I. Passageschema's in 55 stambomen.

v	26	vv	26	vvv	5	vvvv	1
m	19	mm	25 ⁴	mmm	1	mmmm	1
		vm	16	vvm	3	vmmm	1
		mv	19	vmm	1	mmvv	1
				vmv	2	mvvm	1
				mvm	4		
				mmv	1		

Uit deze getallen blijkt geen enkele voorkeur voor een bepaalde wijze van passage. Het totale aantal schema's is groter dan 55 (het aantal stambomen), daar in een aantal van deze stambomen verschillende groepen tweelingen optreden, zodat meer dan één serie passages in één stamboom gevonden wordt. Telt men in deze stambomen het totale aantal passages door vrouw en door man, dan blijken deze aantallen 153 resp. 129 te zijn (deze aantallen zijn niet uit tabel I af te leiden, daar op sommige stambomen passages door elkaar lopen of soms gedeeltelijk samenvallen zij zijn daarom weer uit de stambomen geteld). Deze uitkomst is niet in strijd met de hypothese, dat een passage door vrouw even waarschijnlijk is als een passage door man. De overschrijdingskans, d.i. de kans op het gevonden resultaat, of een onwaarschijnlijker resultaat, is onder aanname van genoemde hypothese ongeveer $\frac{1}{4}$. Dit betekent dus, dat als de hypothese juist is, in ongeveer één van de vier gevallen een dergelijk resultaat als het gevondene, of een voor de hypothese nog ongunstiger, zal optreden.

Deze uitkomst biedt dus geen mogelijkheid, tot een voorkeur voor passage door vrouw of door man te concluderen, in overeenstemming met hetgeen in de inleiding is gezegd.

3. Vervolgens werd onderzocht of misschien binnen iedere stamboom een voorkeur, soms voor mannelijke en soms voor vrouwelijke passages bestond. Dit zou er nl. op wijzen, dat er twee erfelijke factoren in het spel zijn, één werkende via de vrouw en één via de man. Daar een statistische toetsingsmethode niet kan leiden tot het aanvaarden maar wel tot het verwerpen van een hypothese, zal men in dit geval de hypothese toetsen, dat er ook binnen iedere stamboom geen voorkeur bestaat voor passages door vrouw of door man, in de hoop, deze hypothese te kunnen verwerpen.

De onderzoekers stelden daartoe een tabel op van de passageschema's in de stambomen, verdeeld in twee groepen: bij de ene groep was de laatste passage een passage door vrouw, bij de tweede door man.

Deze methode bezit twee nadelen:

- a. Men kan vaak niet uitmaken welke passage de laatste is, daar de twee (of meer) tweelingen in dezelfde generatie kunnen voorkomen.

In dat geval werd als laatste passage genomen de vader of moeder van het gezin, waar een bezoek werd gebracht, om inlichtingen over de stamboom te verkrijgen. Dit is uiteraard een volkomen willekeurig criterium en men kan er niet zeker van zijn, dat door toepassing daarvan geen onzuiverheden in het onderzoek binnensluipen.

- b. Indien men alle passageschema's opstelt en deze vervolgens in twee groepen indeelt, selecterende op de laatste passage, zal men in de groep, waarbij deze laatste passage een passage door vrouw was een overmaat van passages door vrouw verkrijgen en in de andere groep een overmaat van passages door man. Wij zullen dit aan een voorbeeld toelichten.

Stel men beschouwt een aantal passages van de lengte twee, b.v.

v v	100
m m	100
v m	100
m v	100

Het is duidelijk, dat deze getallen zeker niet tot verwerping van de in het begin van deze paragraaf genoemde hypothese kunnen leiden. Selecteert men nu naar aanleiding van de laatste passage, dan worden de twee

groepen:

	v v	100		m m	100
	m v	100		v m	100
totaal	v	300		m	300
	m	100		v	100

Door de selectie ontstaat dus automatisch, bij schema's van de lengte 2, een verhouding van 1 op 3 (resp. 3 op 1) van het aantal mannelijke en vrouwelijke passages.

Nog sterker is het effect van de selectie bij de passageschema's van lengte 1, daar selectie in dat geval eenvoudig alle vrouwelijke passages in de ene en alle mannelijke in de andere groep brengt.

Indien men nu deze selectiemethode toch wél toepassen, zal men de laatste passage, dit is dus de passage, die voor de selectie is gebruikt, bij het toetsen van de gestelde hypothese buiten beschouwing moeten laten. Dit heeft echter het grote nadeel, dat de schema's daardoor korter worden (en, als zij de lengte 1 bezitten, zelfs geheel wegvallen), zodat de omvang van het materiaal zodanig vermindert, dat geen resultaat meer mag worden verwacht. Er werd dan ook met deze methode geen resultaat gevonden.

De hierboven opgesomde bezwaren gelden voor iedere groepering van de gegevens, waarbij geselecteerd wordt op één van de passages. Daarom hebben wij gezocht naar een ^{andere} methode ter toetsing van de in het begin genoemde [van deze paragraaf) hypothese, waarbij wij ernaar gestreefd hebben, zo weinig mogelijk van het verzamelde materiaal buiten beschouwing te laten. Deze methode wordt in de volgende paragraaf besproken.

4. Toetsing van de hypothese, dat er binnen de stambomen geen voorkeur bestaat voor passage door man of door vrouw.

De aan de toetsingsmethode ten grondslag liggende gedachte is de volgende:

Indien de te toetsen hypothese juist is, zal het aantal mannelijke en vrouwelijke passages binnen ieder der passageschema's verdeeld zijn volgens een waarschijnlijkheidsverdeling, die volgt uit de hypothese, dat de kans op een vrouwelijke passage gelijk is aan die op een mannelijke.

Deze waarschijnlijkheidsverdeling is gemakkelijk te berekenen. Men kan dan precies nagaan, hoe vaak bepaalde schema's voor kunnen komen, zonder dat de hypothese verworpen behoeft te worden.

Indien daarentegen de hypothese onjuist is en er binnen ieder schema een uitgesproken voorkeur voor vrouwelijke of voor mannelijke passages bestaat, zal het aantal schema's, waarin te veel vrouwelijke passages gevonden worden evenals het aantal schema's, waarin teveel mannelijke passages voorkomen, groter zijn, dan in overeenstemming is met de hypothese.

Hebben wij b.v. een schema, waarin in totaal 6 passages voorkomen, dan is de kans, dat dit alle vrouwelijke of alle mannelijke passages zijn, gelijk aan $2^{-6} + 2^{-6} = 2^{-5} = \frac{1}{32}$. Zou men te veel van dergelijke schema's vinden, dan zou daaruit volgen, dat de hypothese verworpen moest worden.

Wij kunnen de methode verder als volgt aanduiden: zij x het verschil van het aantal vrouwelijke en het aantal mannelijke passages in één passageschema (positief genomen), ^{gedeelte door de lengte van het passageschema} dan neemt x voor ieder passageschema een tussen 0 en 1 gelegen waarde aan. Indien de hypothese, welke men wil verwerpen, juist is zullen deze waarden in het algemeen niet te ver van 0 verwijderd zijn, indien echter de alternatieve hypothese, inhoudende, dat er wel voorkeur binnen de schema's bestaat, juist is, zullen zij niet ver van de waarde 1 verwijderd zijn. Deze beschrijving dient echter slechts om de gedachtengang van de methode aan te geven. Bij de uitvoering van de toets, die in de Appendix (§6) beschreven is, wordt niet met de grootte x gewerkt, maar met een analoge grootte, die wiskundig beter hanteerbaar is.

Daar het materiaal bij deze methode niet in twee groepen behoeft te worden verdeeld, treedt geen selectie naar een bepaalde passage op, zodat de schema's niet behoeven te worden verkort.

Korte schema's moeten echter buiten beschouwing worden gelaten. Dit vindt zijn oorzaak in het feit, dat een schema van lengte twee, waarin alleen vrouwelijke of alleen mannelijke passages voorkomen (dus v of m), ook als de hypothese juist is, in het geheel

niet onwaarschijnlijk zijn, zoals het ook niet onwaarschijnlijk is, om bij het werpen met een munt tweemaal hetzelfde (dus tweemaal achtereen kruis of twee maal achtereen munt) te krijgen. Deze kans is, om precies te zijn, gelijk aan $\frac{1}{4}$, dus even zeer van 0 als van 1 verwijderd. Bij schema's van de lengte 1 neemt de bovengenoemde grootheid x zelfs noodzakelijkerwijze de waarde 1 aan. Wij moeten ons daarom beperken tot schema's, die minstens de lengte 3 bezitten.

Verder valt het aan te bevelen, stambomen, waarbij zowel in de familie van de vader als van de moeder van een tweeling verdere tweelingen voorkomen, buiten beschouwing te laten, daar hier, indien er twee erfelijke factoren zijn, vermoedelijke beide factoren werkzaam zijn, die door hun gelijktijdig optreden het effect, dat beide afzonderlijk zouden hebben, zouden vertroebelen. Het gewenste resultaat, t.w. het verwerpen van de gestelde hypothese, als deze onjuist is, zou dus worden bemoeilijkt, indien deze stambomen mede in de toetsing worden betrokken. Het is derhalve beter, deze stambomen buiten beschouwing te laten.

Dat deze selectie statistisch verantwoord is, is daaraan gelegen, dat het te toetsen aspect, de onderscheiding tussen mannelijke en vrouwelijke passages, niet in de selectie betrokken is.

Wij houden dan 22 stambomen over, die in tabel II zijn samengevat. (N.B. de telling van de passages heeft nu zo plaats gevonden, dat geen enkele passage meer dan één maal is geteld).

Tabel II

Aantal vrouwelijke en mannelijke passages in 22 stambomen

nummer stamboom	aantal passages	
	v	m
8	3	0
11	5	9
13	2	1
14 A	3	0
14 B	3/	4
15 A	1	2
18	3	1
22	1	2
26	2	6
31	3	1
32	3	4
42	3	0
47	2	1
48	6	4
50	3	0
51	2	1

Tabel II (vervolg)

Aantal vrouwelijke en mannelijke passages in 22 stamboom-
men.

nummer stamboom	aantal passages	
	v	m
53	1	3
58	0	6
59	3	1
60	5	0
66	1	3
70	2	1

Passen wij nu bovengenoemde toetsingsmethoden toe, dan vinden wij een overschrijdingskans van 0,71. Dit betekent, dat ook als de hypothese juist is dat binnen iedere stamboom mannelijke en vrouwelijke passages even waarschijnlijk zijn, ~~4~~ 7 van 10 dergelijke onderzoeken een minstens even sterk afwijkend resultaat zouden geven, zodat er geen enkele reden is, de hypothese te verwerpen. Wij hebben ook dezelfde toetsing nog toegepast op de passageschema's met lengte 4 of groter (een procedure, die, indien zij tot een significant resultaat zou leiden, bedenkelijk zou moeten worden genoemd) en vonden dan voor de overschrijdingskans 0,5. De berekeningen zijn in de Appendix samengevat.

5. Conclusie.

De negatieve resultaten van § 3 en 4 leiden tot de conclusie, dat het waarnemingsmateriaal geen mogelijkheid biedt de hypothese, dat er geen voorkeur voor mannelijke of vrouwelijke passages bestaat, te verwerpen. Dit ~~betekent~~ betekent, dat er, statistisch gezien, geen reden is, om aan te nemen, dat er meer dan één erfelijke factor aanwezig is. Ook indien ~~er~~ niet-statistische gronden zijn, om dit aan te nemen, kan men uit het verzamelde materiaal dus geen steun voor deze aanname ontlenen.

Onze persoonlijke indruk is, dat het gezochte effect (2 factoren, dus voorkeur voor mannelijke of vrouwelijke passages binnen iedere stamboom) niet aanwezig is. Deze indruk berust, uiteraard, op de statistische aspecten van het onderzochte materiaal. Indien het effect wel bestaat, is het in ieder geval zo klein, dat

het zeer moeilijk te constateren zal zijn. De enige opvallende stambomen in tabel II zijn no. 58 en 60. De kans, dat in een stamboom met 6 passages alle passages van één geslacht zullen zijn, is $\frac{1}{32}$ en de analoge kans voor 5 passages is $\frac{1}{16}$. Daar wij echter 22 stambomen hebben, waarbij geen van de overigen enige aanwijzing geeft in dezelfde richting als no. 58 en 60, mag hieraan geen betekenis worden toegekend. Het is geenszins abnormaal, dat in een serie van 22 waarnemingen twee waarnemingen met dergelijke waarschijnlijkheden optreden. Dit alles wordt mathematisch uitgedrukt door de in de vorige paragraaf vermelde overschrijdingskans van 0,71.

6. Appendix.

In deze paragraaf wordt de gevolgde statistische methode kort uiteengezet. Volgens

E.S. Pearson, The probability integral transformation and its use for testing goodness of fit and combining independent tests of significance, Biometrika 30 (1938) p. 134-148 (zie pag. 137),

kan men een aantal onafhankelijke toetsingen als volgt combineren: hebben de toetsingen tweezijdige overschrijdingskansen $\varepsilon_1, \dots, \varepsilon_n$ opgeleverd, dan heeft

$$(1) \quad Q = -2 \log \sqrt[n]{\varepsilon_i}$$

een χ^2 -verdeling met $2n$ vrijheidsgraden. Als kritieke zone voor deze χ^2 -verdeling neme men de eenzijdige rechter-zone: $\chi^2 \geq \chi_0^2$.

Uit ieder van de stambomen van tabel II kan nu een dubbele overschrijdingskans berekend worden volgens de formule:

$$(2) \quad \varepsilon = 2^{-m-v} \sum_{i=0}^v \binom{m}{i}$$

als m het totaal aantal passages is en er v ($\leq m-v$) passages van het ene geslacht en $m-v$ passages van het andere geslacht onder zijn.

In tabel III vindt men de berekeningen samengevat;

paren	v	$m-v$
-------	-----	-------

en	$m-v$	v
----	-------	-----

uit tabel II zijn daarbij niet meer onderscheiden, daar zij dezelfde ε bezitten.

Tabel III

Significantieberekening voor de gegevens van tabel II.

passages	n_i	ε_i	$-\log \varepsilon_i$	$-n_i \log \varepsilon_i$
9 - 5	1	0,424	0,858	0,858
6 - 4	1	0,754	0,282	0,282
5 - 2	1	0,289	1,241	1,241
4 - 3	1	1	0	0
6 - 0	1	0,031	3,474	3,474
5 - 0	1	0,063	2,764	2,764
4 - 1	1	0,375	0,981	0,981
3 - 1	5	0,625	0,430	2,150
3 - 0	4	0,25	1,386	5,544
2 - 1	6	1	0	0
totaal	22			17,294

n_i = aantal maal, dat een paar, als in de eerste kolom genoemd, in tabel II voorkomt.

$$\begin{aligned} \text{Derhalve is: } \varphi &= -2 \log \sqrt[n]{\prod \varepsilon_i} = 34,588 \\ \text{Aantal vrijheidsgraden} &44 \\ \therefore P &= 0,71. \end{aligned}$$

Laten we de laatste twee regels van tabel II weg, dan komt er:

$$\begin{aligned} \varphi &= 23,5 \\ \text{Aantal vrijheidsgraden} &24 \\ \therefore P &\approx 0,5 \end{aligned}$$