**Lucia de B**
Hauwert, A.G.

# Lucia de B

**Gonny Hauwert**

**12 september 2007**

## Inhoudsopgave

1 **Inleiding** ***2***
2 **Berekeningen voor de rechtszaak** ***3***
2.1 *Opmerkingen over deze methode* *5*
3 **Statistische toetsen** ***6***
3.1 *Bespreking van de toetsen* *7*
3.2 *Vergelijkingen van de toetsen* *11*
3.3 *Data samenvoegen* *12*
4 **Conclusie** ***13***

## 1 Inleiding

Deze scriptie gaat over de statistische berekeningen die gebruikt zijn voor de rechtszaak van Lucia de B. Zij is in juni 2004 door het gerechtshof in Den Haag veroordeeld voor 7 moorden en 3 pogingen tot moord. Zij heeft hiervoor levenslang en TBS gekregen. Zij heeft in verschillende ziekenhuizen gewerkt waaronder het Juliana Kinderziekenhuis (JKZ) en het Rode Kruis Ziekenhuis (RKZ). De statisticus dr. Elffers is door de rechter gevraagd om een statistisch rapport te schrijven over de zaak. In eerste instantie heeft hij alleen berekeningen gedaan voor het JKZ, omdat alleen van dat ziekenhuis de gegevens vrij gegeven waren. In dit ziekenhuis zijn ze haar gaan verdenken, omdat er erg veel incidenten tijdens haar diensten plaats vonden (onder incidenten worden sterfgevallen en reanimaties verstaan). Op verzoek van de rechter zijn later ook berekeningen gedaan voor twee afdelingen van het RKZ waar ze in dezelfde periode gewerkt heeft. Men vroeg zich af of het toeval zou kunnen zijn dat Lucia betrokken was bij al die incidenten, terwijl ze onschuldig was. De rechter heeft tijdens de rechtszaak aan dr. Elffers gevraagd wat de kans is dat het toeval zou kunnen zijn dat Lucia bij zoveel incidenten aanwezig was. Dr. Elffers heeft uitgerekend wat de kans is dat een willekeurig persoon betrokken kan zijn bij zoveel incidenten, als de incidenten volgens toeval gebeuren. Waarom zijn deze berekeningen zo belangrijk en welke invoeld hebben ze tijdens de rechtszaak gehad? Statistici geloofden dat er medisch bewijs was voor de moorden en de medici geloofden dat daar statistisch bewijs voor was. Dit heeft er toe geleid dat de rechter Lucia schuldig heeft bevonden.

In deze scriptie zullen we kijken naar de berekeningen die dr. Elffers gedaan heeft, de aanmerkingen daarop en mogelijke verbeteringen. Hiervoor worden alternatieve statistische toetsingsgrootheden besproken en met elkaar vergeleken.

## 2 Berekeningen voor de rechtszaak

Tijdens de rechtszaak van Lucia wilde de rechter weten of er sprake was van toeval of geen toeval, dat zij bij zoveel incidenten aanwezig was. Om deze rede is er aan Dr. Elffers gevraagd om dit te berekenen. Hij heeft een methode gebruikt die per afdeling vraagt of het denkbaar is, dat gegeven het aantal incidenten, de verdeling van de incidenten over de verpleegkundige vergelijkbaar is met toeval. De methode neemt het aantal incidenten en het totaal aantal diensten dat alle verpleegkundige gedraaid hebben als gegeven. Voor deze methode heeft hij twee aannamen gedaan. Ten eerste neemt hij aan dat de kans dat Lucia aanwezig was tijdens een incident even groot is als voor elke andere verpleegkundige. Ten tweede heeft hij aangenomen dat de incidenten onafhankelijk zijn voor verschillende diensten. Het aantal diensten per verpleegkundige is hiermee hypergeometrisch verdeeld.

In de volgende tabel staan de gegeven die bekend gemaakt zijn voor deze rechtszaak:

| **Tabel1: JKZ** | *met incident* | *zonder incident* | totaal |
|---|---|---|---|
| Dienst waar Lucia aanwezig was | 8 | 134 | 142 |
| Dienst waar Lucia niet aanwezig was | 0 | 887 | 887 |
| Totaal | 8 | 1021 | 1029 |

| **Tabel2: RKZ-42** | *met incident* | *zonder incident* | totaal |
|---|---|---|---|
| Dienst waar Lucia aanwezig was | 5 | 53 | 58 |
| Dienst waar Lucia niet aanwezig was | 9 | 272 | 281 |
| Totaal | 14 | 325 | 339 |

| **Tabel3: RKZ-41** | *met incident* | *zonder incident* | totaal |
|---|---|---|---|
| Dienst waar Lucia aanwezig was | 1 | 0 | 1 |
| Dienst waar Lucia niet aanwezig was | 4 | 361 | 365 |
| Totaal | 5 | 361 | 366 |

Omdat dit de enige bekende gegevens zijn, heeft dr. Elffers verder nog een aantal aannamen gedaan. Namelijk dat er een vaste kans $p$ is op een incident tijdens een dienst (de kans hangt er dus niet vanaf of het een dag-, avond- of nachtdienst was). Daarnaast nam hij aan dat de incidenten onafhankelijk van elkaar gebeurd zijn. Door te conditioneren op het totaal aantal diensten en het aantal diensten dat Lucia wel en niet heeft gedraaid, kun je de kans berekenen dat Lucia precies bij 0,1,2. . . ,9 incidenten aanwezig zou zijn. Dit is mogelijk met de formule voor de hypergeometrische verdeling. De waarden in de onderstaande tabel worden als volgt benoemd:
$m$ is het totaal aantal diensten van Lucia
$n$ is het totaal aantal diensten waar Lucia niet aanwezig was
$x$ is het aantal diensten van Lucia waarin een incidenten plaats heeft gevonden
$y$ is het aantal diensten waar Lucia niet aanwezig was en een incidenten plaats heeft gevonden

| Tabel: | *met incident* | *zonder incident* | totaal |
|---|---|---|---|
| Dienst waar Lucia aanwezig was | $x$ | $m-x$ | $m$ |
| Dienst waar Lucia niet aanwezig was | $y$ | $n-y$ | $n$ |
| Totaal | $z=x+y$ | $m+n-x-y$ | $m+n$ |

De hypergeometrische verdelingsfunctie zier er in dit geval als volgt uit:

$$h(x) = \frac{\binom{m}{x}\binom{n}{y}}{\binom{m+n}{z}}$$

Door het gebruik van een conditionele verdeling hangt de kansverdeling van de incidenten tijdens Lucia haar diensten niet meer af van de onbekende parameter $p$.

Dit wiskundige model kunnen we beschouwen als het vaasmodel. Voor het JKZ heb je een vaas met 1029 knikkers, wat staat voor het totaal aantal diensten in het ziekenhuis. Acht van deze knikkers zijn zwart, deze staan voor de diensten met incidenten. 1021 zijn wit en staan voor de diensten zonder incidenten. We willen de kans weten dat alle acht de incidenten tijdens Lucia haar diensten per toeval plaats vonden. Dit kunnen we berekenen door 142 knikkers uit de vaas te pakken, welke staan voor het aantal diensten dat Lucia gedraaid heeft. We kijken dan naar het aantal mogelijkheden waarop je 134 witte knikkers en de 8 zwarte kunt pakken.

Dr. Elffers heeft de volgende nulhypothese genomen: als de incidenten allemaal per toeval zijn gebeurd dan zijn ze hypergeometrisch verdeeld. Hij verwerpt deze hypothese als de p-waarde kleiner is dan 10000. In hoofdstuk 3 gaan we dieper in op statistische toetsen. We kunnen nu de p-waarde en kans uitrekenen dat Lucia aanwezig was bij minstens het aantal incidenten dat ze heeft meegemaakt, verdeeld over haar diensten op de verschillende afdelingen. We doen dit door de waarde van de hypegeometrische verdeling van het aantal incidenten waar ze aanwezig was te berekenen. Hierbij tellen we de waarde van de hypergeometrische verdeling van elke mogelijke groter aantal incidenten op. We willen namelijk de kans dat ze minstens bij zoveel incidenten aanwezig was bepalen.

Voor tabel 1 van het JKZ geldt de volgende berekening voor de p-waarde:

$$h(8) = \frac{\binom{142}{134}\binom{887}{887}}{\binom{1029}{1021}} = 0.000000110572$$

Dit geeft ook gelijk de p-waarde, omdat ze bij alle incidenten aanwezig was. Er is een correctie factor toegevoegd aan de p-waarde van het JKZ, omdat we de kans willlen weten dat een willekeurige zuster aanwezig is bij zoveel van de incidenten. De waarde van het JKZ is daarom vermenigvuldigd met het aantal zusters dat er werken, namelijk 27. Zo krijg je de kans dat een willekeurige zuster acht van de acht incidenten meemaakt. De p-waarde wordt dan $0.000000110572 \cdot 27 = 0.0000029854$. Dit geeft aan dat de kans dat Lucia aanwezig was bij acht van de acht incidenten ongeveer 1 op de 3.000.000 is.

Voor tabel 2 van het RKZ afdeling 42 geldt:

$$h(5) = \frac{\binom{58}{53}\binom{281}{272}}{\binom{339}{325}} = 0.01868453$$

Deze tellen we op bij $h(6)$ tot en met $h(14)$. Dit geeft een p-waarde van 0.07155922. De kans dat Lucia betrokken was bij minstens vijf van de 14 incidenten is ongeveer 1 op de 14.

Lucia heeft maar één dienst op afdeling 41 van het RKZ gedraaid. Tijdens deze dienst vond een stefgeval plaats. In de tijd dat zij er gewerkt heeft waren er totaal 366 diensten en 5 incidenten. Omdat zij er een erg korte periode heeft gewerkt kunnen we de p-waarde niet op dezelfde manier berekenen. Dr. Elffers heeft gekeken naar de kans, dat het toeval is dat Lucia tijdens die ene dienst één van de vijf sterfgevallen heeft meegemaakt. Hij kwam uit op $\frac{5}{366} = 0.013661$, oftewel een kans van ongeveer 1 op de 73. Bij het RKZ is de correctie factor niet toegepast, omdat Lucia toen al verdacht was.

Nu zijn alle kansen van de verschillende afdelingen berekend en moeten ze nog samengevoegd worden om een uitspraak te kunnen doen. Dr. Elffers berekende de kans dat Lucia toevallig aanwezig was bij zoveel incidenten, door de verschillende p-waarde te vermenigvuldigen onder de gegeven condities. De uiteindelijke p-waarde is volgens hem gelijk aan $0.0000029854 \cdot 0.07155922 \cdot 0.013661 = 0.00000000292$, oftewel een kans van ongeveer 1 op 342 miljoen. Dr. Elffers heeft er wel bij gezegd dat dit niet bewijst dat Lucia schuldig is.

### 2.1 Opmerkingen over deze methode

Het staat niet vast hoe je statistiek in deze zaak kunt gebruiken, maar we kunnen wel laten zien dat er op een heleboel punten aan de berekeningen van dr. Elffers getwijfeld kan worden. Een aantal wordt hieronder genoemd.

- Hij heeft aangenomen dat de kans dat er iemand overlijd in één van de ziekenhuizen overal even groot is. Hierdoor vallen de kansen in zijn berekeningen tegen elkaar weg. Maar zijn de kansen wel aan elkaar gelijk? In het JKZ werkte ze op de medium care waar redelijk ernstig zieke kinderen lagen. In het RKZ heeft ze op twee verschillende afdelingen gewerkt waar oudere mensen lagen. Is de kans dat een kind in het JKZ overlijd even groot als de kans dat er een ouder iemand overlijd in het RKZ? Dit is iets waar we niet zeker over kunnen zijn.

- De gegevens van het JKZ wordt twee keer gebruikt in zijn berekeningen. Een keer om de verdachte te identificeren en daarna om de p-waarden mee te berekenen. Hij gebruikt dus eerst de data om een hypothese op te zetten en daarna test hij de hypothese met dezelfde data. Vanwege deze reden heeft hij ook de correctie factor van 27 toegevoegd zoals beschreven wordt in de vorige paragraaf, maar is die correctie wel genoeg?

- Of deze methode heel realistisch is hangt af van de realiteit van de aannamen die gemaakt zijn. Hoe realistisch is bijvoorbeeld de aanname dat de incidenten willekeurig verdeeld zijn over de diensten van de zusters. Je kunt bijvoorbeeld bedenken dat er een correlatie bestaat tussen de zusters en de incidenten. Misschien overlijden mensen vaker 's nachts dan overdag. Een zuster die vaak nachtdiensten draait heeft dan een grotere kans om een incident mee te maken. Een goede zuster zou alerter kunnen zijn op ontwikkelingen van crisissen en wordt eerder ingedeeld op moeilijke diensten zoals bijvoorbeeld de nachtdiensten. Een aantal opeenvolgende incidenten binnen een korte tijd zouden misschien een gevolg kunnen zijn van veranderingen binnen een afdeling. Een zuster zou in een bepaalde tijd meer kunnen werken door bijvoorbeeld vakanties van andere zusters. Een incident in dienst $n+1$ zou een gevolg kunnen zijn van iets dat in dienst $n$ gebeurt is. Als je naar kortere periodes gaat kijken hoeven de diensten dus niet meer geheel willekeurig ingevuld te worden. Er kunnen dus onschuldige redenen zijn voor correlatie tussen incidenten en zusters.

- Hij vermenigvuldigt de drie p-waarden van de ziekenhuizen met elkaar. Dit betekent dat hij aanneemt dat onder zijn nulhypothese de incidenten totaal willekeurig op elk van de drie afdelingen voorkomen en dat ze onafhankelijk over de afdelingen verdeeld zijn, maar met mogelijke verschillende aantallen binnen de afdelingen. Door het vermenigvuldigen van de p-waarden van verschillende afdelingen, kan je elke willekeurige zuster die in genoeg verschillende ziekenhuizen of op genoeg verschillende afdelingen werkt verdacht maken. De p-waarde wordt namelijk na elke vermenigvuldiging kleiner.

## 3 Statistische toetsen

Een statistische toets is een methode om na te gaan of een bepaalde veronderstelling, de nulhypothese, met de waargenomen data verworpen dient te worden. Als de nulhypothese niet verworpen kan worden dan wordt deze geaccepteren, zij het "bij gebrek aan bewijs". De gemaakte veronderstelling wordt verworpen als de waargenomen verschillen met wat verwacht was niet meer op toeval lijken te berusten.

De p-waarde geeft aan hoe extreem de gevonden waarde voor de toetsingsgrootheid in de verdeling onder de nulhypothese is. Hoe kleiner de p-waarde, hoe extremer de uitkomst. In de praktijk worden waarden van 5% en 1% aangehouden als grens; is de p-waarde kleiner, dan spreekt men van een significante, resp. sterk significante uitkomst.

Er zijn verschillende statistische toetsen om je hypothese te toetsen. Hier onder zal ik een aantal relevante toetsen beschrijven, die mogelijk bij het toetsten van de hypothese aan de hand van de data voor de rechtzaak gebruikt kunnen worden.

- De chi-kwadraat toets. Deze toets wordt gebruikt om te zien of waargenomen aantallen systematisch afwijken van verwachte aantallen. Deze toets wordt vaak gebruikt om kruistabellen te analyseren. In het geval van deze zaak kijkt de toets naar hoe moordlustig Lucia in het RKZ was en in de andere ziekenhuizen.

- Mantel-Haenszel toets. Deze toets is ontworpen om de alternatieve hypothese met een grote p-waarde te verwerpen. Hiermee bereken je het onderscheidend vermogen van de onafhankelijkheid van de data.

- Fishers methode voor het combineren van verschillende tabellen. Fisher heeft een methode bedacht om verschillende p-waarden te combineren. Dit kun je bijvoorbeeld doen voor p-waarden die gebaseerd zijn op de Fisher's exact test.

- Er is ook nog de mogelijkheid om de data te combineren van de verschillende ziekenhuizen. Dus als het ware alle diensten op één hoop te gooien alsof alles in één ziekenhuis heeft plaats gevonden en vervolgens de p-waarde berekenen met de chi-kwadraat toets.

### 3.1 Bespreking van de toetsen

Om een uitleg te geven over de verschillende toetsen die in de vorige paragraaf genoemd zijn, beginnen we met het formuleren van de hypothese die we willen toetsen.

Er zijn drie afdelingen waar Lucia gewerkt heeft en waar we de gegevens van hebben. Op deze afdelingen hebben meerdere incidenten plaats gevonden in de tijd dat zij er gewerkt heeft. We zijn geïntereseerd in de volgende nulhypothese: in alle drie de afdelingen is er geen verband tussen de aanwezigheid van Lucia en de incidenten.

We gaan kijken naar de eigenschappen van de verschillende toetsingsgroothe-den onder de nulhypothese. Hiervoor gebruiken we de volgende 2x2 tabel:

| | *met incident* | *zonder incident* | totaal |
|---|---|---|---|
| Dienst waar Lucia aanw. was | $\hat{p}_i n_i = x_i$ | $n_i - x_i$ | $n_i$ |
| Dienst waar Lucia niet aanw. was | $\hat{q}_i m_i = y_i$ | $m_i - y_i$ | $m_i$ |
| Totaal | $x_i + y_i$ | $n_i + m_i - x_i - y_i$ | $n_i + m_i$ |

met $i = 1, 2, 3$ omdat er gegevens bekend zijn van drie afdelingen. Veronderstel dat $X_i \sim \text{Bin}(m_i, p_i)$ en $Y_i \sim \text{Bin}(n_i, q_i)$ onafhankelijk van elkaar. Hierbij zijn $X_i$ en $Y_i$ het aantal incidenten die tijdens respectivelijk de diensten van Lucia $m_i$ en de diensten van de andere verpleegkundige $n_i$ plaats hebben gevonden. Verder is $p_i$ de kans op een incident als Lucia dienst heeft en $q_i$ de kans op een incident als ze geen dienst heeft. De $\hat{p}_i$ en $\hat{q}_i$ zijn de kansen die uit de gegeven data volgen.

We beginnen met de chi-kwadraat toetsingsgrootheid. Deze grootheid ziet er als volgt uit:

$$\chi^2 = \sum_{j=1}^{12} \frac{(N_j - E_j)^2}{E_j}$$

met $j$ de drie keer vier cellen van de drie 2x2 tabellen. $E_j$ geeft het rijtotaal keer het kolomtotaal gedeeld door het totaal aantal trekkingen en $N_j$ geeft de waarneming. In ons geval geeft dit de volgende formule met $i$ de verschillende tabellen:

$$\begin{aligned}\chi^2 \quad = \quad & \sum_{i=1}^{3} \left( \frac{(x_i - \frac{n_i(x_i+y_i)}{n_i+m_i})^2}{\frac{n_i(x_i+y_i)}{n_i+m_i}} + \frac{(n_i - x_i) - \frac{(n_i-x_i)(n_i-(x_i+y_i))}{n_i+m_i}}{\frac{(n_i-x_i)(n_i-(x_i+y_i))}{n_i+m_i}} \right. \\ & \left. + \frac{y_i - \frac{m_i(x_i+y_i)}{n_i+m_i}}{\frac{m_i(x_i+y_i)}{n_i+m_i}} + \frac{(m_i - y_i) - \frac{(m_i-y_i)(m_i-(x_i+y_i))}{n_i+m_i}}{\frac{(m_i-y_i)(m_i-(x_i+y_i))}{n_i+m_i}} \right)\end{aligned}$$

Onder de nulhypothese is $\chi^2$ bij benadering verdeeld als $\chi^2_3$.

Ten tweede de Mantel-Haenszel toetsingsgrootheid. Deze wordt gegeven door de volgende formule:

$$\chi^2_{\text{MH}} = \frac{(|\sum_{i=1}^{3} \frac{n_i m_i}{n_i+m_i}(\hat{p}_i - \hat{q}_i)| - 0.5)^2}{\sum_{i=1}^{3} \frac{n_i m_i}{n_i+m_i-1}\hat{r}_i(1-\hat{r}_i)}$$

Om te begrijpen wat deze grootheid doet, gaan we kijken naar de verschillende onderdelen. Hiervoor hebben we de odds ratio ($\kappa$) nodig. De odds ratio is de verhouding tussen twee odds. De odds is een quotiënt van waarschijnlijkheden: de kans dat iets plaats vindt gedeeld door de kans dat het niet plaats vindt. De Mantel-Haenszel toets is ontworpen om een goed onderscheidings vermogen te hebben, dat wil zeggen dat $\kappa$ voor alle afdelingen gelijk is. De odds ratio wordt gegeven door:

$$\kappa = \frac{\frac{p_i}{1-p_i}}{\frac{q_i}{1-q_i}} \qquad \text{voor alle } i.$$

Hierbij is $p_i$ de kans op een incident als Lucia dienst heeft en $q_i$ de kans op een incident als Lucia geen dienst heeft. Als de odds ratio gelijk is aan 1 dan is de kans dat er een incident gebeurd tijdens de diensten van Lucia gelijk aan de kans op een incident als zij niet aanwezig is. We toetsen nu $\kappa = 1$ versus $\kappa > 1$.

Het is mogelijk om $q_i$ uit te drukken in $p_i$, voor de helderheid hebben we de index $i$ weg gelaten:

$$\begin{aligned}
\kappa &= \frac{\frac{p}{1-p}}{\frac{q}{1-q}} = \frac{p(1-q)}{q(1-p)} \\
&\Rightarrow q = \frac{p(1-q)}{\kappa(1-p)} = \frac{p-pq}{\kappa(1-p)} \\
&\Rightarrow q + \frac{pq}{\kappa(1-p)} = q(1+\frac{p}{\kappa(1-p)}) = \frac{p}{\kappa(1-p)} \\
&\Rightarrow q = \frac{\frac{p}{\kappa(1-p)}}{\frac{\kappa(1-p)+p}{\kappa(1-p)}} = \frac{\kappa(1-p)p}{\kappa(1-p)(\kappa(1-p)+p)} \\
&= \frac{p}{\kappa(1-p)+p}
\end{aligned}$$

Hieraan kun je zien dat als $\kappa$ groter wordt, wordt $q$ kleiner. Dit geeft aan dat voor een grotere $\kappa$ de kans op een incident tijdens Lucia haar dienst groter is dan de kans op een incident tijdens de diensten van andere verpleegkundigen.

Neem $S = \sum_{i=1}^{3} \frac{n_i m_i}{n_i+m_i}(p_i - q_i)$. We bekijken een vast aantal tabellen en nemen aan dat de tabellen onderling onafhankelijk zijn. Neem $m_i$ en $n_i$ groot en $\kappa = 1$, omdat de nulhypothese ervan uitgaat dat $p_i = q_i$. Noem de kans $p_i = q_i$ gelijk aan $r_i$. Vervolgens kunnen we de variantie van S bepalen:

$$\begin{aligned}
\mathrm{Var}(S) &= \sum_i (\frac{n_i m_i}{n_i+m_i})^2 (\frac{p_i(1-p_i)}{n_i} + \frac{q_i(1-q_i)}{m_i}) \\
&= \sum_i (\frac{n_i m_i}{n_i+m_i})^2 \frac{n_i+m_i}{n_i m_i} r_i(1-r_i) \\
&= \sum_i \frac{n_i m_i}{n_i+m_i} r_i(1-r_i)
\end{aligned}$$

Schat $p_i$ en $q_i$ met $\hat{p}_i$ en $\hat{q}_i$, omdat $p_i$ en $q_i$ onbekend zijn. Dit geeft dat $r_i$, $\hat{r}_i$ wordt. De verwachtingswaarde van $r_i$ veranderd als volgt:

$$\begin{aligned}
E[\hat{r}_i(1-\hat{r}_i)] &= E[\hat{r}_i] - E[\hat{r}_i^2] = r_i - (r_i^2 + \frac{r_i(1-r_i)}{n_i+m_i}) = r_i(1 - r_i - \frac{1-r_i}{n_i+m_i}) \\
&= \frac{n_i+m_i-1}{n_i+m_i} r_i(1-r_i)
\end{aligned}$$

Door het vervangen van $r_i$ door $\hat{r}_i$ wordt de verwachtingswaarde iets kleiner. Dit verklaart de $n_i - 1$ in de noemer van de Mantel-Haenszel toets. De Mantel-Haenzel toetsingsgrootheid geeft de som over de absolute waarde van $S$ in het kwadaart met een kleine correctie voor kleine steekproef aantallen. Vervolgens word dit gedeeld door de variantie van $S$.

Om wat meer over deze toets te kunnen zeggen gaan we de variantie en de verwachtingswaarde van de toets uitrekenen. Dit kunnen we doen door te conditioneren op de kolomtotalen $x_i + y_i$, wat ervoor zorgt dat alleen $x$ variabel is, om vervolgens de score test toe te passen. De score toets is uniform de meest krachtige toets om de nulhypothese $\kappa = 1$ versus $\kappa > 1$ te toetsen. Deze toets houd in dat je de afgeleide van het logaritme van $P(x)$ neemt.
Waar

$$P(x) = P(X = x|X + Y = x + y) = \frac{\kappa^x \binom{m}{x}\binom{n}{y}}{\sum_{x'=0,x+y-y'}^{x'=x+y,m} \kappa^{x'} \binom{m}{x'}\binom{n}{x+y-x'}}.$$

We nemen eerst het logaritme van $P(x)$

$$\log(P(x)) = x\log(\kappa) + \log(\binom{m}{x}\binom{n}{y}) - \sum_{x'=0,x+y-y'}^{x'=x+y,m} x'\log(\kappa\binom{m}{x'}\binom{n}{x+y-x'}))$$

leiden dit vervolgens af naar $\kappa$:

$$\frac{\partial}{\partial\kappa}\log P(x) = \frac{x}{\kappa} - \frac{\sum \frac{x'}{\kappa}\binom{m}{x'}\binom{n}{x+y-x'}}{\sum \binom{m}{x'}\binom{n}{x+y-x'}}$$

en bekijken dit onder de nulhypothese, waar geldt dat $\kappa = 1$

$$\frac{\partial}{\partial\kappa}\log(P(x)) = x - \frac{\sum x'\binom{m}{x'}\binom{n}{x+y-x'}}{\sum \binom{m+n}{x+y}} = x - E_{\kappa=1}[X] = x - \frac{x+y}{m+n}m.$$

Voor alle $i$ tabellen samen volgt dan:

$$\begin{aligned}
\frac{\partial}{\partial\kappa}\log(P(x_i)) &= \sum_i x_i - \frac{x_i + y_i}{m_i + n_i}m_i \\
&= \sum_i \frac{m_i x_i + n_i x_i - m_i x_i - m_i y_i}{m_i + n_i} \\
&= \sum_i \frac{n_i x_i - m_i y_i}{m_i + n_i} \\
&= \sum_i \frac{n_i m_i}{m_i + n_i}(\frac{x_i}{m_i} - \frac{y_i}{n_i}) \\
&= \sum_i \frac{n_i m_i}{m_i + n_i}(p_i - q_i)
\end{aligned}$$

Dit geeft voor de score toets: $T = \frac{\partial}{\partial\kappa}\log P(x_i)_{\kappa=1} = \sum_i \frac{n_i m_i}{m_i + n_i}(p_i - q_i)$

We gaan de variantie en verwachtingswaarde van deze toets T bekijken. Neem hiervoor aan dat de steekproefgrootte $N$ is met $N = \sum_i N_i$ en $N_i = n_i + m_i$, $m_i = \alpha_i N_i$ en $n_i = \beta_i N_i$ met $\alpha_i$ en $\beta_i$ vast en $\alpha_i + \beta_i = 1$. De kansen $p_i$ en $q_i$ hangen af van N, wat we als volgt noteren $p_{Ni}$ en $q_{Ni}$ en er geldt dat $p_{Ni} \to p_i$ en

$q_{N\,i} \to p_i$ voor $N \to \infty$. We willen $\kappa$ dichtbij 1, daarom kijken we naar $\kappa - 1 = \frac{\delta}{\sqrt{N}}$. De nulhypothese wordt: $\delta = 0$ versus $\delta > 0$.

$$\begin{aligned}
\mathrm{Var}(\frac{S}{\sqrt{N}}) &= \frac{1}{N}\sum_i (\frac{n_i m_i}{n_i + m_i})^2 (\frac{p_i(1-p_i)}{m_i} + \frac{q_i(1-q_i)}{n_i}) \\
&= \sum_i \frac{N_i}{N}\frac{1}{N_i}(\frac{\alpha_i\beta_i N_i^2}{\beta_i N_i + \alpha_i N_i})^2 (\frac{p_i(1-p_i)}{\beta_i N_i} + \frac{q_i(1-q_i)}{\alpha_i N_i}) \\
&= \sum_i \frac{N_i}{N}\frac{1}{N_i}(\frac{\alpha_i\beta_i N_i^2}{\alpha_i N_i + \beta_i N_i})^2 (\frac{\alpha_i N_i + \beta_i N_i}{\alpha_i N_i \beta_i N_i} p_i(1-p_i)) \\
&= \sum_i \frac{N_i}{N}\frac{\alpha_i\beta_i}{\alpha_i + \beta_i} p_i(1-p_i) \\
&= \sum_i \frac{N_i}{N}\alpha_i\beta_i p_i(1-p_i)
\end{aligned}$$

$$\begin{aligned}
E[\frac{S}{\sqrt{N}}] &= \frac{1}{\sqrt{N}}\sum_i \frac{n_i m_i p_i}{n_i + m_i} - \frac{m_i n_i q_i}{n_i + m_i} \\
&= \frac{1}{\sqrt{N}}\sum_i \frac{\alpha_i\beta_i N_i^2 p_i}{\alpha_i N_i + \beta_i N_i} - \frac{\alpha_i\beta_i N_i^2 \frac{p_i}{p_i + \kappa(1-p_i)}}{\alpha_i N_i + \beta_i N_i} \\
&= \frac{1}{\sqrt{N}}\sum_i \frac{N_i^2 \alpha_i\beta_i p_i (1 - \frac{1}{p_i + \kappa(1-p_i)})}{N_i} \\
&= \frac{1}{\sqrt{N}}\sum_i N_i\alpha_i\beta_i p_i (1 - \frac{1}{p_i + \frac{\delta(1-p_i)}{\sqrt{N}}} \\
&\approx \frac{1}{\sqrt{N}}\sum_i N_i\alpha_i\beta_i p_i (1 - (1 - \frac{\delta(1-p_i)}{\sqrt{N}} + \mathcal{O}(\frac{1}{N}))) \\
&\approx \frac{1}{\sqrt{N}}\sum_i \frac{N_i}{\sqrt{N}}\alpha_i\beta_i\delta p_i(1-p_i) \\
&= \sum_i \frac{N_i}{N}\alpha_i\beta_i\delta p_i(1-p_i)
\end{aligned}$$

Voor $N \to \infty$ geldt: $\frac{T}{\sqrt{N}} \sim \mathcal{N}(\delta \sum_i \frac{N_i}{N}\alpha_i\beta_i p_i(1-p_i), \sum_i \frac{N_i}{N}\alpha_i\beta_i p_i(1-p_i))$

Ten derde de methode van Fisher om p-waarden te combineren. Om te zien wat deze grootheid doet, kijken we naar zijn verdeling. Stel alle verschillende nulhypotheses zijn waar, voor alle $k$ met $k$ onafhankelijke toetsingsgrootheden $T_i$. Onder de nulhypothese is het product van alle p-waarden ongeveer uniform $(0,1)$ verdeeld. Dan is $-2\sum_{i=1}^{k} \log(\text{p-waarde}) \sim 2\text{Gamma}(k,1) = \text{Gamma}(\frac{2k}{2}, \frac{1}{2}) = \chi^2_{2k}$ verdeeld. Kijk nu naar de verdeling van $\frac{T_i}{\sigma_i} = Z_i$ dan is $Z_i \sim \mathcal{N}(\delta\sigma_i, 1) \sim Z_i + \delta\sigma_i$. Dit geeft $-2\sum_{i=1}^{k} \log(1 - \Phi^{-1}(Z_i + \delta\sigma_i))$ met $\Phi$ de standaardnormaleverdeling.

We kunnen niet veel over deze methode zeggen. Alleen dat deze minder goed is dan die van Mantel-Haenszel, die is namelijk gebaseerd op $\sum_i T_i$, gegeven uniform de meest krachtige toets.

Ten vierde gaan we alle data bij elkaar optellen. Om dit te mogen doen nemen we aan dat alle $p_i$'s van de verschillende afdelingen gelijk aan elkaar zijn. Dit is hetzelfde als het kijken naar één ziekenhuis waarvoor we de volgende tabel hebben:

| | *met incident* | *zonder incident* | Totaal |
|---|---|---|---|
| Dienst waar Lucia aanwezig was | 14 | 187 | 201 |
| Dienst waar Lucia niet aanwezig was | 13 | 1520 | 1533 |
| Totaal | 27 | 1707 | 1734 |

Dit heeft een $\chi_1^2$ verdeling.

### 3.2 Vergelijkingen van de toetsen

We willen weten wat een goede toetsingsgrootheid zou zijn om de data van de drie afdelingen waar Lucia gewerkt heeft te combineren. Om dit te bepalen gaan we de alternatievehypotheses en de verwachtingswaarde vergelijken. We doen dit omdat we de toets willen gebruiken die de meeste zekerheid geeft dat als we Lucia schuldig bevinden, dat ze ook daadwerkelijk schuldig is. Een toetsingsgrootheid met een grote verwachtingswaarde en een niet te grote alternatievehypothese geeft dit resultaat.

We weten dat de nulhypothese voor alle toetsen $p_i = q_i$ is. Eerst gaan we kijken naar de verschillende voorwaarden onder de alternatievehypothese. Voor de $\chi^2$ grootheid is de alternatievehypothese $p_i \neq q_i$. Deze hypothese kijkt naar alle verschillende mogelijkheden voor $p_i$ en $q_i$. Dit zijn alle mogelijke kansen op incidenten bij alle verpleegkundige. De methode die er vanuit gaat dat alles in één ziekenhuis gebeurd, is heeft als alternatievehypothese $p \neq q$. Dit is ongeveer hetzelfde als de $\chi^2$ methode. De alternatievehypothese van de Mantel-Haenszel grootheid is $\frac{p_i}{1-p_i} = \eta \frac{q_i}{1-q_i}$ met $\eta > 0$. Deze hypothese kijkt heel specifiek naar de verhouding tussen het wel of niet betrokken zijn bij incidenten van Lucia ten opzichte van de andere verpleegkundige. Voor de Fishers combinatie methode is de alternatieve hypothese dat één of meer nulhypothesen niet waar zijn zodanig dat de p-waarde van de toets de neiging hebben om heel klein te zijn. Dit is een eenzijdige toets en heeft dus een grote mogelijkheid voor het alternatief.

Nu bekijken we de verwachtingswaarde van de veschillende grootheden. De chi-kwadraat methode is onder de nulhypothese bij benadering verdeeld als $\chi_3^2$. Dit geeft met dezelfde voorwaarde die we bij de Mantel-Haenszel grootheid in de vorige paragraaf gegeven zijn dat:

$$\begin{aligned}
\chi_3^2 \quad &\sim \quad \mathcal{N}\left(\delta \sum_i \frac{N_i}{N}\alpha_i \sum_i \frac{N_i}{N}\beta_i p(1-p), \sum_i \frac{N_i}{N}\alpha_i \sum_i \frac{N_i}{N}\beta_i p(1-p)\right) \\
&= \mathcal{N}\left(\delta \sqrt{\sum_i \frac{N_i}{N}\alpha_i \sum_i \frac{N_i}{N}\beta_i} \sqrt{p(1-p)}, 1\right)^2
\end{aligned}$$

Als we de verwachtingswaarde en variantie uit de vorige paragraaf invullen in de Mantel-Haenszel toetsingsgrootheid dan geldt:

$$\begin{aligned}
\text{MH} \quad &\sim \quad \frac{\sum_{i=1}^3 \left|\mathcal{N}\left(\sum_i \frac{N_i}{N}\alpha_i\beta_i\delta p_i(1-p_i), \sum_i \frac{N_i}{N}\alpha_i\beta_i p_i(1-p_i)\right) - 0.5\right|^2}{\sum_i \frac{N_i}{N}\alpha_i\beta_i p_i(1-p_i)} \\
&= \mathcal{N}\left(\frac{\sum_i \frac{N_i}{N}\alpha_i\beta_i\delta p_i(1-p_i)}{\sqrt{\sum_i \frac{N_i}{N}\alpha_i\beta_i p_i(1-p_i)}}, 1\right)^2
\end{aligned}$$

$$= \mathcal{N}\left(\delta\sqrt{\sum_i \frac{N_i}{N}\alpha_i\beta_i}\sqrt{p(1-p)}, 1\right)^2$$

Om te weten te komen welke verwachtingswaarde het grootst is. Gaan we $\sum_i \frac{N_i}{N}\alpha_i \sum_i \frac{N_i}{N}\beta_i$ en $\sum_i \frac{N_i}{N}\alpha_i\beta_i$ met elkaar vergelijken. Neem $\frac{N_i}{N} = \mu_i$, $\alpha_i + \beta_i = \gamma_i = 1$ en $\frac{\alpha_i}{\alpha_i+\beta_i} = \varphi_i$. Dit geeft voor de verwachtingswaarde van $\chi^2_3$ het volgende:

$$\begin{aligned}
\sum_i \frac{N_i}{N}\alpha_i \sum_i \frac{N_i}{N}\beta_i &= \sum_i \mu_i\gamma_i\varphi_i \sum_i \mu_i\gamma_i(1-\varphi_i) \\
&= \sum_i \mu_i\gamma_i\varphi_i \sum_i \mu_i\gamma_i - \sum_i \mu_i\gamma_i\varphi_i \sum_i \mu_i\gamma_i\varphi_i \\
&= \sum_i \mu_i\gamma_i\varphi_i \sum_i \mu_i\gamma_i - (\sum_i \mu_i\gamma_i\varphi_i)^2 \\
&= \sum_i \mu_i\gamma_i\varphi_i - (\sum_i \mu_i\gamma_i\varphi_i)^2
\end{aligned}$$

en voor Mantel-Haenszel

$$\begin{aligned}
\sum_i \frac{N_i}{N}\alpha_i\beta_i &= \sum_i \mu_i\gamma_i\varphi_i\gamma_i(1-\varphi_i) \\
&= \sum_i \mu_i\gamma_i^2\varphi_i(1-\varphi_i) \\
&= \sum_i \mu_i\gamma_i^2\varphi_i - \sum_i \mu_i\gamma_i^2\varphi_i^2 \\
&= \sum_i \mu_i\gamma_i\varphi_i - \sum_i \mu_i\gamma_i^2\varphi_i^2
\end{aligned}$$

Voor $\alpha_i$ en $\beta_i$ geldt dat ze beide tussen 0 en 1 in liggen. We kunnen ze dus vergelijken met kansen, daarvoor is bekend dat $E[X^2] \leq E[X]^2$ dus $\sum_i \gamma_i^2\varphi_i^2 \leq (\sum_i \gamma_i\varphi_i)^2$. Wat aangeeft dat de $\chi^2$ grootheid een betere toets is om te gebruiken, onder de voorwaarden dat de gegeven $p_i$'s allemaal gelijk aan elkaar zijn en dat de $\alpha_i$'s allemaal verschillend zijn. Deze waarde van de $\chi^2$ grootheid is kleiner dan die van de Mantel-Haenszel dus de verwachtingswaarde is groter, wat een betere 'verwerpingskans' geeft. Als alle $\alpha_i$'s hetzelfde zijn geldt de gelijkheid. In paragraag 3.1 hebben we laten zien dat we niet zo veel over de verwachtingswaarde van de methode van Fisher kunnen zeggen. Voor de methode om alle data bij elkaar op te tellen is de verwachtingswaarde gelijk aan die van de $\chi^2_3$ verdeling, alleen dan zonder de index $i$.

### 3.3 Data samenvoegen

We hebben voor een aantal berekeningen gebruik gemaakt van de aanname dat $p_i = q_i$ voor grote $N$, maar hoe groot is de fout die we maken als $p_i \neq q_i$. Dit hebben we gebruikt bij het berekenen van de verwachtingswaarde en variantie ven de Mantel-Haenszel toets en de Chi-kwadraat toets. Meer informatie en deze berekeningen zijn te vinden in het artikel "The Cochran-Mantel-Haenszel test and the Lucia the Berk case"van Prof. dr. Richard D. Gill. Hierin laat hij zien dat als we $p_i$ en $q_i$ laten af hangen van een kleine $\theta$ we een fout krijgen die afhangt van $\theta$. Dit betekent dat de verwachtingswaarde van één van de toetsen gelijk aan nul is, dus $\delta = 0$

de verwachtingswaarde ongelijk aan nul zou kunnen zijn. Het kan fout gaan bij de afdelingen waar de kans op succes groot is. Dus als Lucia relatief meer uren werkt op een afdeling waar veel incidenten plaats vinden.

## 4 Conclusie

Lucia de B. is veroordeeld voor 7 moorden en 3 pogingen tot moord. Voor haar rechtszaak zijn statistische berekeningen gedaan. De vraag is hoe goed waren deze berekeingen. Op een groot aantal punten zijn er aanmerkingen op de methode die gebruikt is. We hebben vier mogelijke andere methode/statistische toetsingsgrootheden besproken die mogelijk beter zouden zijn om in het geval van Lucia te gebruiken. De methode van Fischer om de data van de verschillende afdelingen te combineren, lijkt mij het minst geschikte alternatief. Deze is minder goed dan de Mantel-Haenszel toets en we hebben er weinig gegevens over. De methode om alle data bij elkaar op te tellen en er vanuit te gaan dat alles in één ziekenhuis is gebeurd vind ik niet zo realistisch. Je gaat er dan vanuit dat de kans $p_i$ op een incident tijdens een dienst van Lucia op alle verschillende afdelingen hetzelfde is. Dit is een aanname waarvan we niet weten hoe realistisch deze is. Verder hebben we nog gekeken naar de Chi-kwadraat toets en de Mantel-Haenszel toets. De verwachtingswaarde van de Chi-kwadraat toets geeft een beter resultaat onder de aannamen die we gedaan hebben, maar de alternatievehypothese van de Mantel-Haenszel toets is meer gericht op wat we nodig hebben voor de rechtszaak van Lucia. Deze hypothese kijkt namelijk heel specifiek naar de verhouding tussen het wel of niet betrokken zijn bij incidenten van Lucia ten opzichte van de andere verpleegkundige. Dus onder de aannamen dat de steekproefgrote groot is lijkt de Mantel-Haenszel toetsingsgrootheid de beste statisitische toets om te gebruiken.

## Referenties


[1] T. Derksen, *Lucia de B.*, Veen Magazines, Diemen (2006).

[2] Dr.H. Elffers, *Verdeling reanimatie- en overlijdensgevallen in het Juliana Kinderziekenhuis en Rode Kruisziekenhuis*, (2002).

[3] R.D. Gill, *Elffers' methode and Ellfers'mistake*, (2007).

[4] John A. Rice, *Mathematical Statistics and Data Analysis*, Seccond Editioan, Duxbury Press, Belmont, California, (1995).

[5] Joseph L. Fleiss, Bruce Levin, Myunghee Cho Paik, *Statistical Methods for Rates and Proportions*, Third Edition, Wiley, (2003).

[6] R.D. Gill, *The Cochran-Mantel-Haenszel test and the Lucia the Berk case*, (2007).