

wat ik nu toch weer zag ...

Het doel van deze rubriek is allerlei publikaties onder de aandacht van onze lezers te brengen, waarbij, gebruik makend van methoden en technieken op ons vakgebied, onjuiste of dubieuze conclusies zijn getrokken of een verkeerde voorstelling van zaken werd gegeven. Deze kunnen dan, behalve tot 'lering ende vermaeck', gebruikt kunnen worden bij het overdragen van de rol van ons vakgebied.

Het ligt in de bedoeling om deze stukjes, die kort moeten zijn, toe te spitsen op het misbruik in algemeen toegankelijke media als radio, televisie en (vooral) kranten en algemene tijdschriften. Het is zeker niet de bedoeling om via deze rubriek een discussie te gaan voeren over allerlei beweringen in academische publikaties. Het gaat om het 'algemeen maatschappelijk belang' van het juiste gebruik van de informatieoverdracht en de problemen die daarbij op kunnen treden.

De redactie berust bij Arend Oosterhoorn (adres zie voorblad). De stukjes dienen aan hem te worden gezonden, waarbij

- ◇ de tekst het liefst wordt aangeleverd op een floppy in WordPerfect of naar WordPerfect te converteren formaat;
- ◇ mogelijke grafieken of andere originelen zodanig worden meegestuurd dat er geen al te groot kwaliteitsverlies optreedt door kopieën van kopieën van Uiteraard worden die originelen geretourneerd na publikatie, tenzij anders afgesproken;
- ◇ de lengte van de bijdrage zich beperkt tot maximaal twee zijden A4 volgens de aanwijzingen van Kwantitatieve Methoden.

In het geval dat er meerdere inzendingen zijn naar aanleiding van dezelfde publikatie, zal de redactie de mogelijkheid bekijken om één reactie samen te stellen waarin de visies van alle inzenders herkenbaar is.

In de loop der tijd zullen we wellicht komen met bijstellingen met betrekking tot deze rubriek. In de praktijk zal zich een optimale vorm aandienen.

De redactie beslist over de uiteindelijke vorm en plaatsing.

Overschattingen

Chiel Bakkeren

In de Automatiseringsgids van 20 mei 1994 stond een heel artikel over 'Steekproef is methode voor foutencorrectie'. Het ging over de controle van fouten in een database. Volgens het artikel leert de praktijk dat 6 à 10 procent fouten eerder regel is dan uitzondering. Het artikel beschrijft een steekproefmethode, gebaseerd op het AOQL principe, waarmee deze databases kunnen worden onderzocht en zonodig opgeschoond.

Het artikel heeft een kader waarin de volgende tekst staat

Zoeken naar fouten

Twee voorbeelden

In een direct-mail-database bevatte 10 procent van de adressen waarschijnlijk een of meerdere fouten. Deze fouten zaten in de spelling van namen, huisnummers en postcodes. Het totaal aantal adressen in de database was circa 6000. Op basis van een risico-analyse kon met het bestand verdelen in 20 deelbestanden van 300 adressen. Per deelbestand trok men een steekproef van 8 adressen.

Men trof in 4 van de 20 steekproeven een fout aan. In totaal controleerde, en waar

nodig corrigeerde, men 1328 adressen. Men kreeg zo 100 procent zekerheid dat er minder dan 5 procent fouten in het bestand achterbleven. De werkelijke fout zal hier beneden liggen (bij 95 procent betrouwbaarheid maximaal 113 adressen).

Een tweede voorbeeld.

Een organisatie was bezig met de voorbereiding om haar database (ongeveer 150.000 records), over te zetten naar een nieuw systeem. Voor de conversie moest het systeem stilgelegd worden met alle kosten van dien. Om de gegevens zo snel en goed

mogelijk over te krijgen wilde men de zekerheid dat maximaal 0,1 procent van de gegevens afweek van de vereisten van het nieuwe systeem.

De database werd gesplitst in 15 deelbestanden. Per deelbestand trok men een steekproef van 370 records. In 1 van de steekproeven werd een fout ontdekt. Men controleerde dit deelbestand (5000 records). In totaal controleerde men 10.180 records. De conversie verliep soepel. Slechts 10 records moesten na de conversie handmatig worden bijgewerkt.

In het artikel schorten meerdere dingen, maar het meest opvallend is de misser in het betreffende kader. De manier waarop de daar vermelde waarschijnlijkheden precies berekend worden, wordt niet gespecificeerd, maar de claim '100 procent zekerheid' (2e kolom) lijkt wat te veel van het goede.

De enige manier om 100% zekerheid te verkrijgen, dat het percentage resterende fouten kleiner is dan 5%, is immers om tenminste 95% van het bestand te controleren en zo nodig te corrigeren. Wordt minder dan 95% gecontroleerd, dan bestaat er altijd de kans dat alle niet gecontroleerde elementen fout zijn.

In het onderhavige geval is $1328/6000 \approx 22\%$ van het bestand gecontroleerd/gecorrigeerd; er bestaat dus zelfs een (zeer kleine) kans dat 78% van het bestand nog fouten bevat!

Er is nog veel werk aan de winkel

Ruud H. Koning

Sommige misverstanden krijg je waarschijnlijk nooit uit de wereld. Neem nu het verhaaltje in *Ad Valvas* van 2 juni 1994.

Huisgenoot Joris Venrooij (21), eerstejaars student medicijnen aan de VU, verbleef in India toen hij in een lokale Indiase krant las over de ramp. Hij kon toen nog niet vermoeden dat hij niet veel later bij Brian zou gaan wonen, in dezelfde flat waar het vliegtuig een krater had geslagen. De ruimte, het vele groen en de lage huur van de kamer trokken Joris vier maanden geleden naar de Bijlmer toe. De ramp lijkt hij laconiek naast zich neer te leggen: 'Dit is de veiligste flat van de wereld: zo'n vliegtuig stort niet voor de tweede keer op dezelfde plek neer.'

...

Wie biedt er meer indices

Joop Hox, Peter Swanborn

In een artikel (van voor de redactie onbekende herkomst) staat het verslag van een onderzoek naar discriminatie:

HET ONDERZOEK

In totaal hebben drie onderzoeksduo's, elk bestaande uit een Marokkaan en een Nederlander met gelijklopende kwalificaties, telefonisch gereflecteerd op 257 vacatures. In 92 gevallen kregen zowel de Nederlandse als de Marokkaanse kandidaat te horen dat in de vacature al was voorzien. Daar kunnen we dus geen discriminatie vaststellen. Bij de overgebleven 165 pogingen kwam er 122 maal met *beide* sollicitanten een sollicitatiegesprek tot stand. Maar bij overige 43 vacatures kreeg de Marokkaan meteen te horen dat er voor de baan al iemand

was - of er werd een andere uitvlucht verzonden - terwijl de Nederlander wèl de gelegenheid tot een telefonisch sollicitatiegesprek kreeg. Het omgekeerde - dat de Nederlander werd afgescheept en de Marokkaan niet - kwam niet voor. Na de telefonische ronde volgt in een sollicitatieprocedure een persoonlijk gesprek. In 60 van de 122 gevallen kregen beide kandidaten te horen dat ze de tweede ronde niet hadden gehaald. In 43 gevallen mochten we wel allebei komen. In 17 gevallen kreeg alleen Ruud de Wit een uitnodiging en Mustafa El

Mansouri niet, in 2 gevallen lag dat andersom.

Bij de 43 uitnodigingen tot nader gesprek voor beide kandidaten was het niet steeds zo dat de nep-kandidaten ook zonder meer konden gaan. Bijvoorbeeld wanneer ze ter plekke een proeve van bekwaamheid moesten afleggen (vis fileren, verkopen) of over een bijzondere kennis moesten beschikken (tropische vissoorten, gordijnstoffen). In 17 gevallen was de kans op ontmaskering gering en konden beide het sollicitatiegesprek aangaan.

Die 17 gesprekken resulteerden in 9 dubbele afwijzingen, één dubbele aanname en verder werd de Hollander 6 maal en de Marokkaan één maal gekozen.

Tellen we alle gevallen van voordeel in de verschillende fasen van het sollicitatieproces bij elkaar op, dan wordt de Hollander 66 keer uitgekozen tegen 3 maal de Marokkaanse kandidaat. Het netto-getal 63 levert de maat voor discriminatie en die wordt gedeeld op het aantal van 165 bruikbare observaties. In tenminste 38 procent van de gevallen is er dus sprake van discriminatie.

Als je dit artikel zo leest wordt de verlciding wel erg groot om een prijsvraag uit te schrijven voor het construeren de meest creatieve index.

