
Rapportage analyse onderzoek

Eenheid/Eenheden	:	Management / Korpsleiding
Lijst/zaaknummer	:	2014-407 ID-wisselingen
Bijlage	:	Behoort bij Rapport (Vervolg onderzoek)
Datum	:	25-11-2014

Inleiding

In dit document wordt de methode beschreven die door het NVIK is gebruikt voor het uitvoeren van de steekproef voor de aanvang van de analyse op het onderwerp van de legale (Oost-Europese) identiteitswisselingen. De door het NVIK toegepaste onderzoeksmethode is, vooraf, afgestemd met de faculteit criminologie van de Erasmus Universiteit te Rotterdam.

Steekproefgrootte

Er bestaan verschillende steekproefmethodes, een veelgebruikte is de aselechte (of random) steekproef, waarbij alle eenheden uit de populatie evenveel kans hebben om in de steekproef terecht te komen. Dit gebeurt aan de hand van een aantal parameters die u dient in te voeren:

1. De grootte van de populatie

De totale populatie bestaat uit 49.153 regels met persoonsgegevens verdeeld over negen nationaliteiten (waarbij de landen van voormalig Joegoslavië als één geboorteland werden beschouwd), te weten: Albanese (2%), Burgers van Bosnië-Herzegovina (14%), Joegoslavische (9%), Burgers van Kosovo (0,4%), Burgers van Kroatië (2%), Macedonische (5%), Burgers van Montenegro (0,2%), Burgers van Oekraïne (67%) en Burgers van Servië en Montenegro (0,3%). Uitgangspunt voor de populatie is zicht hebben op personen die bovenstaande nationaliteit hebben, alsmede beschikken over éénzelfde geboortedatum en geboorteplaats. De opgevraagde persoonsinformatie bevat: V-nummer, achternaam, voorvoegsel, voornamen, geslacht, geboortedatum, geboorteplaats, geboorteland en datum eerste registratie BVV. Daarnaast nummer(s) van eventueel bij de betreffende vreemdeling geregistreerde ID-documenten. Indien geen geboorteplaats bekend is, kon de record genegeerd worden.

2. De gewenste foutenmarge

Dit is de positieve of negatieve afwijking die u toestaat op het bekomen steekproefresultaat, d.w.z. de nauwkeurigheid die u vooropstelt. Stel dat in uw steekproefonderzoek 40% van de respondenten positief antwoordt op een bepaalde stelling, dan heeft u bij een foutenmarge van 2% de zekerheid dat tussen de 38% en 42% positief zou geantwoord hebben indien u de vraag aan de totale populatie had gesteld. Hoe kleiner de toegestane foutenmarge, hoe groter uw steekproef zal moeten zijn. Voor deze populatie is gekozen voor een foutenmarge van 5%. Deze foutmarge is structureel toegepast op alle weergaven van resultaten. Als blijkt dat bijvoorbeeld 12 procent in de steekproef een dubbele identiteit heeft, zal dit in deze rapportage verwoord worden als een marge tussen de 7 en 17 procent.

3. Het gewenste betrouwbaarheidsniveau

Het betrouwbaarheidsniveau geeft aan hoe zeker u kunt zijn van de foutenmarge, en vertelt m.a.w. hoe vaak het werkelijke percentage van de populatie dat een bepaald

antwoord zou kiezen, binnen de foutenmarge ligt. In marktonderzoek worden de foutenmarges over het algemeen berekend op een betrouwbaarheidsniveau van 95%. Dit betekent dat een onderzoeksuitkomst in 19 van de 20 gevallen conform de realiteit is. Een hoger betrouwbaarheidsniveau (bv 99%) vereist logischerwijze een grotere steekproef. Het gewenste betrouwbaarheidsniveau voor deze analyse is 95%.

Uit de berekening blijkt dat de steekproef, bij een betrouwbaarheidsniveau van 95% en een foutenmarge van 5%, moet bestaan uit 382 personen.

Gestratificeerde steekproeftrekking met aselekt begin

Voor deze analyse is gekozen voor een gestratificeerde steekproeftrekking. De reden hiervoor is dat elementen in de populatie qua verhouding op een kenmerk nogal scheef verdeeld zijn. In dit bestand, waarin negen nationaliteiten vertegenwoordigd zitten, is nationaliteit zo'n kenmerk dat zeer scheef verdeeld is (zie reeds vermelde percentages). Dit zou er toe kunnen leiden dat er te weinig of geen personen van een bepaalde nationaliteit worden geselecteerd. Geslacht is eveneens scheef verdeeld. De populatie bestaat voor 61% uit mannen, 38% uit vrouwen en van 1% is het geslacht onbekend. Om hiervoor te corrigeren is de populatie ingedeeld in verschillende nationaliteitsgroepen (hier wordt meestal de Latijnse term voor groep gebruikt, namelijk stratum) en opgedeeld naar geslacht. In onderstaande tabel zitten de aantallen verdeeld naar nationaliteit en geslacht. Voor het juist aantal personen in de steekproef is het percentage geselecteerd dat overeenkomt met de steekproef over de hele populatie. De verhoudingen zijn hierin gelijk gebleven. Van ieder stratum is nu willekeurig een x-aantal elementen geselecteerd. Helaas moet meteen gerealiseerd worden dat bepaalde strata te weinig cases bevat om te analyseren. Met de verkregen steekproef is het alleen mogelijk verantwoorde uitspraken te doen over de totale populatie en in het bijzonder over de nationaliteiten: Burgers van Bosnie-Herzegovina, Joegoslavische en Burgers van Oekraïne. De overige nationaliteiten bevatten te weinig cases om verantwoorde uitspraken over te doen.

Met aselekt begin

De wijze waarop binnen een stratum de elementen getrokken worden kan volstrekt aselekt zijn (gestratificeerde aselekte steekproeftrekking), dan wel volgens een aselekt begin (gestratificeerde steekproeftrekking met aselekt begin). Bij deze steekproeftrekking is gekozen voor een aselekt begin. Deze wijze van steekproef trekken verschilt van de standaard aselekte steekproeftrekking. Ten eerste: de eenheden van de populatie worden netjes naar kenmerk in een rij gezet. Zo kunnen bijvoorbeeld alle namen van de personen die tot de populatie behoren op alfabetische volgorde worden geplaatst. Voor deze populatie is gekozen om de personen op geboortedatum te rangschikken. Vervolgens, en dat is het tweede verschil, trekt men uit deze lijst een element om de zoveel elementen, afhankelijk hoe groot de stratum is en hoeveel elementen er van dat stratum nodig zijn voor in de steekproef. Het enige wat nu nog aan het toeval wordt overgelaten is het eerste nummer dat als beginpunt wordt gekozen, in dit geval is steeds gekozen voor een willekeurig (rond) getal. Het voordeel is dat de steekproef een verdeling heeft die beter overeenkomt met de populatie: van de personen zitten er verhoudingsgewijs evenveel jongere mensen als oudere mensen. In de bijlage is een tabel opgenomen waarin populatie en steekproef afgezet zijn tegen geslacht en nationaliteit.

Bevindingen, conclusie en resultaten

In het steekproefbestand staan 382 sets bestaande uit een 933 records (regels). Records met gelijke geboortedatum, geslacht en geboorteland zijn tot één set gerekend. Een aantal sets (113) bevatten v-nummers met persoonsgegevens die exact gelijkkluidend waren. Gemiddeld zitten 2,4 records in een setje. Deze verhouding is door te trekken naar het

totale bestand met een betrouwbaarheidsniveau van 95% en een foutmarge van 5%, dan zitten er in het totaalbestand tussen de 19.119 en 20.075 setjes. Om een schatting te maken voor het aantal dubbelingen zullen de percentages in de resultaten omgezet worden naar aantallen gebaseerd op deze waardes. Alle 382 sets zijn zogenaamd "handmatig" uitgelopen door de informatiemakelaars van het NVIK, waarbij de bronsystemen BVV en PSH-v zijn geraadpleegd. Om uit te sluiten of vast te stellen of sprake is of was van een uniek persoon met meerdere namen, is gekeken naar de beschikbare foto's in genoemde systemen. Tevens zijn de geregistreerde bijzonderheden bekeken om vast te stellen of sprake is of was van een uniek persoon. In de data van de steekproef bleken 55 personen, die, als uniek, konden worden aangeduid en met een andere naam en/of tweede V-nummer in de BVV voorkwamen. Wanneer dit resultaat wordt gegeneraliseerd naar het totale bestand, dan is 14 procent, tussen de 2677 en 2810 personen (wederom met een foutmarge van 5%), met twee of meer V-nummers in de BVV zijn ingevoerd.

Op basis van het bovengenoemde onderzoek, fotomatching en geregistreerde informatie, bleek, dat 23 van de 55 personen onder twee of meerdere namen voorkwamen. Met andere woorden, bij 1147 tot 1204 personen (met een foutmarge van 5%) in het totaalbestand zou sprake zijn van een dubbele identiteit. In de meeste gevallen waren foto's aanwezig die de vaststelling van een dubbele identiteit op basis van fotomatching mogelijk maakte. Meerdere oorzaken kunnen ten grondslag liggen aan de koppeling van een tweede identiteit aan één en dezelfde persoon. Zo kan het aannemen van de naam van de partner bij een huwelijksovereenkomst zorgen voor een dubbele identiteit. Daarnaast kunnen namen binnen registraties verschillen op één of twee letters waardoor er sprake lijkt van een dubbele identiteit. In beiden gevallen kan er sprake zijn van bewuste danwel onbewuste aanpassing in de naam. Bij de overige 32 personen, dus tussen 1529 en 1606 personen (5% foutmarge) in het totaalbestand, ligt mogelijk een systeem- of invoerfout ten grondslag. De personen zijn dubbel opgevoerd en hebben meerdere vreemdelingsnummers gekregen. Ook het gebruik van diakrieten in de naam zorgde voor dubbelingen. N.b. in het totaalbestand bleken de nationaliteit "Burger van Oekraïne" de grootste vertegenwoordiging te zijn. Opvallend was dat bij de ingevoerde gegevens, gerelateerd aan de geboorteplaats, veelvuldig UKR, zijnde de landcode voor Oekraïne, was ingevoerd. Personen die met deze geboorteplaats waren geregistreerd zijn niet in het onderzoek meegenomen.

Van 5162 tot 5420 personen (5% foutmarge) in het totaalbestand is niet vast te stellen of er sprake is van een dubbeling, omdat er te weinig identificerende gegevens van de betrokkenen aanwezig zijn om te kunnen vergelijken. Er zou sprake kunnen zijn van een mogelijke dubbele nationaliteit maar dit is aan de hand van de ten dienste van het NVIK beschikbare informatie niet vast te stellen of uit te sluiten. Dit komt in de meeste gevallen door het ontbreken van een profielfoto (en biometrische gegevens) waarmee matching van identiteiten zou kunnen plaatsvinden. Het ontbreken van deze gegevens kan veroorzaakt zijn door onvoldoende toegang tot de systemen waarin de gegevens zijn opgeslagen. Door het ontbreken van de juiste gegevens is het niet mogelijk vast te stellen of uit te sluiten of er daadwerkelijk sprake is van een dubbele identiteit. Ook is het door het onvoldoende kunnen beschikken over gegevens niet uit te sluiten of er in specifieke gevallen sprake is van eventuele tweelingbroers of -zussen die exact dezelfde geslachtsnaam, geboortedatum en -plaats hebben en daardoor als mogelijke dubbeling naar voren zijn gekomen. In enkele gevallen bestaat het vermoeden dat het om een tweeling gaat en niet om een dubbeling van identiteit. Ook lijkt in enkele gevallen, ondanks voldoende biometrische gegevens, genoeg twijfel te bestaan bij het vaststellen van een eventuele dubbeling aan de hand van een fotomatch.

Bijlage

Tabel personen in populatie en steekproef afgezet naar nationaliteit

	Totale populatie				Steekproef			
	Totaal	Man	Vrouw	Onb	Totaal	Man	Vrouw	Onb
<i>Nationaliteit</i>	<i>aantal</i>	<i>aantal</i>	<i>aantal</i>	<i>aantal</i>	<i>aantal</i>	<i>aantal</i>	<i>aantal</i>	<i>aantal</i>
Albanese	1009	702	296	11	8	6	2	0
B v Bosnië-Herz	6833	3465	3320	48	53	27	26	0
Joegoslavische	4460	2408	1857	195	35	19	14	2
Burger v Kosovo	203	91	112	0	2	1	1	0
Burger v Kroatië	1097	571	507	19	8	4	4	0
Macedonische	2487	1329	1146	12	19	10	9	0
B v Montenegro	106	49	57	0	1	0	1	0
B v Oekraïne	32819	21496	11280	43	255	167	88	0
B v Servië en M	139	93	45	1	1	1	0	0
Totaal	49153	30204	18620	329	382	235	145	2