# ASPEKTE RAKENDE DIE TOEPASSING VAN KOMPLEKSE STEEKPROEFNEMINGSONTWERPE IN DIE PRAKTYK

deur

ALIDA HERBST

voorgelê te vervulling van die vereistes vir die graad

**MSc (TOEGEPASTE STATISTIEK)**

in die Fakulteit Wis- en Natuurkunde

UNIVERSITEIT VAN PRETORIA,

PRETORIA

MAART 1985

# VOORWOORD

Ek wil graag Professor H.S. Schoeman bedank vir sy bydrae tot hierdie navorsing. Verder wil ek my dank betuig teenoor Professor D.J. Stoker vir die voorstelling van die onderwerp van hierdie navorsing. Sy voortdurende aanmoediging en waardevolle advies was onmisbaar.

Ek wil ook die RGN bedank, wat dit vir my moontlik gemaak het om hierdie navorsing te doen as deel van 'n navorsingsprojek. Erkenning word ook verleen aan die Meningspeilingsentrum (MPS) van die RGN vir die beskikbaarstelling van meerstadium komplekse steekproefdata, en aan Dr. S.H.C. du Toit vir sy hulp met die koppeling van programme.

# INHOUDSOPGAWE

III. VARIANSIEBERAMING IN DIE GEVAL VAN KOMPLEKSE STEEKPROEWE

# SAMEVATTING


ASPEKTE RAKENDE DIE TOEPASSING VAN KOMPLEKSE

STEEKPROEFNEMINGSONTWERPE IN DIE PRAKTYK.

deur

ALIDA HERBST

Leier: PROFESSOR H.S. SCHOEMAN

Medeleier: PROFESSOR D.J. STOKER

Departement: STATISTIEK

GRAAD: MSc (TOEGEPASTE STATISTIEK)

Literatuur oor die beraming van standaardfoute van beramers van populasieparameters in die geval van komplekse steekproewe is nie so algemeen verkrygbaar soos dié in die geval van eenvoudige ewekansige steekproewe met onafhanklike waarnemings, nie.

Die kompleksiteit van 'n steekproef word gevorm deur stratifikasie, trosvorming en moontlike ongelyke seleksiewaarskynlikhede. In hierdie geval is dit belangrik om die kompleksiteit van die steekproefopname in ag te neem by beraming, hierdie aspek word in diepte in die studie ondersoek.

Buiten die standaardfoute van beramers van populasieparameters is daar twee ander belangrike maatstawwe om te beskou by die analisering van 'n

komplekse steekproef, naamlik die ontwerpeffek (deff = 'design effect') en die intratroskorrelasiekoëffisiënt (roh = 'rate of homogeneity').

Die ontwerpeffek is per definisie die verhouding van die variansie van 'n beramer van 'n populasieparameter uit 'n komplekse steekproef teenoor die variansie van dieselfde beramer wanneer die kompleksiteit van die steekproef geïgnoreer word. Die intratroskorrelasiekoëffisiënt, roh, meet die variasie in die veranderlike onder beskouing binne die primêre of eerste stadium steekproefnemingseenhede (PSE's) van die steekproefontwerp teenoor die variasie in die veranderlike oor die hele steekproef.

Die konsep van oordraagbaarheid word bespreek en prakties toegepas. Oordraagbaarheid verwys na die moontlikheid om die waarde van die standaardfout, deff-waarde of roh-waarde van een subklas na 'n ander, van een veranderlike na 'n ander, of van een opname na 'n ander, oor te dra.

Twee belangrike tipes subgroepe is kruisgroepe en gesegregeerde groepe. Kruisgroepe sny dwarsdeur elke PSE in die hele steekproef, byvoorbeeld geslag of ouderdom. Gesegregeerde groepe daarenteen sny nie deur die PSE's nie, maar bevat slegs sekere PSE's in die geheel byvoorbeeld geografiese gebiede en bevolkingsgroepe.

Die deff-waarde as maatstaf van oordraagbaarheid is breër toepasbaar in die praktyk as die standaardfoute, en is veral geskik vir gesegregeerde subgroepe, aangesien die PSE-grootte en steekproefontwerp, wat 'n groot invloed op die deff-waarde het, nie vir sulke subgroepe drasties verskil nie. Van al drie die maatstawwe is die roh-waarde as maatstaf van oordraagbaarheid die algemeenste toepasbaar in die praktyk en is veral geskik vir kruisgroepe aangesien die grootte van die PSE's, wat hier drasties kan verskil, nie die roh-waarde beïnvloed nie.

Die program CLUSTERS (ontwikkel deur die World Fertility Survey) word gebruik om oordraagbaarheid te ondersoek deur gebruikmaking van RGN-steekproefdata. CLUSTERS maak gebruik van die Taylor-linearisasiemetode van variansieberaming wat 'n groot tekortkoming het naamlik dat dit beperk is tot die variansieberaming van lineêre beramers (bv. gemiddeldes of verhoudings) van populasieparameters. Twee herhaalde-replikasiemetodes van variansieberaming is ontwikkel en in die literatuur weergegee om hierdie tekortkoming te oorbrug. Die twee herhaalde-replikasiemetodes, naamlik Gebalanseerde Herhaalde-Replikasie (BRR = 'Balanced Repeated Replication') en Uitsnit Herhaalde-Replikasie (JRR = 'Jackknife Repeated Replication') sowel as die Taylor-linearisasiemetode word bespreek. Die waardes van die variansieberamings vir lineêre beramers van populasieparameters uit die RGN-steekproefopname, word vir al drie die variansieberamingsmetodes deur middel van CLUSTERS en 'n BRRJRR-program, wat deur die outeur ontwikkel is, met mekaar vergelyk.

## SUMMARY

# ASPECTS CONCERNING THE APPLICATION OF COMPLEX SAMPLING DESIGNS IN PRACTICE.

by

ALIDA HERBST

Superviser: PROFESSOR H.S. SCHOEMAN

Co-superviser: PROFESSOR D.J. STOKER

Department: STATISTICS

DEGREE: MSc (APPLIED STATISTICS)

Literature on the estimation of standard errors of estimators of variable values for complex samples is not so commonly obtainable as that concerning simple random sampling with independent observations.

The complexity of a sample is formed by stratification, clustering and possible unequal selection probabilities. It is important to take the complexity of the complex sample into account when estimating variances, this is the main purpose of this study.

Except for the standard errors of estimators, there are two other measures that are worth using during the analysis of the data resulting from a complex sample, namely the design effect (deff) and the intraclass

correlation coefficient (roh = 'rate of homogeneity').

The design effect is per definition the variance of an estimator of population parameters in a complex sample in relation to the variance of the same estimator ignoring the complexity of the sample.

The intraclass correlation coefficient, roh, measures the variation in a variable within the primary or first stage sampling units (PSU's) of the sample in comparison with the variance in the variable in the complete sample.

The concept portability is discussed and practically illustrated. Portability refers to the possibility of carrying over from one subclass to another, from one variable to another or from one survey to another, the conclusions drawn regarding the sampling error, deff value or roh value.

Two important subclasses are cross classes and segregated classes. Cross classes cut across the whole sample and contain a part of each PSU, for example sex and age. Segregated classes do not cut through any PSU's, but contain a few PSU's in totality for example geographical regions and race groups.

The deff value as a measure of portability is more widely applicable in practice than the standard error, and is especially suitable for segregated classes, since the size of the PSU's and the sampling design which may have a drastic effect on the value of the design effect do not change drastically.

Of the three possible measures the roh value as a measure of portability is the widest applicable in practice and is especially suitable for cross classes since the size of the PSU's, which might differ drastically, has little, if any, effect on the roh value.

The programme CLUSTERS (developed by the World Fertility Survey) is used to investigate portability by using HSRC sampling data. CLUSTERS makes use of the Taylor linearisation method for variance estimation, which has a big shortcoming in the sense that it is restricted to the variance estimation of linear estimators (for example means or proportions) of variable values. Two repeated replication methods of variance estimation have been developed and are discussed in the literature to overcome this shortcoming.

The two repeated replication methods, namely Balanced Repeated Replication (BRR) and Jackknife Repeated Replication (JRR) as well as the Taylor linearisation method are discussed. The values of the variance estimates for linear estimators of variable values, using the HSRC sampling survey data, are compared for all three methods of variance estimation with the aid of CLUSTERS and a BRRJRR programme developed by the author.

# I. KOMPLEKSE STEEKPROEFONTWERP

## 1.1 INLEIDING

Oor die jare is sekere standaard statistiese metodes vir die beraming van populasieparameters ontwikkel onder die aanname van eenvoudige ewekansige steekproefneming. Hierdie steekproefnemingsprosedure is gebaseer op die onafhanklike trekking van elemente (onafhanklike waarnemings), uit een of ander eindige populasie wat die verkryging van enige teoretiese resultate heelwat vergemaklik en vereenvoudig.

In die praktyk word dikwels van meerstadiumsteekproefneming gebruik gemaak. Die rede daarvoor is dat dit tyd en koste bespaar en tweedens is die hoeveelheid administratiewe werk daaraan verbonde baie minder as in die geval van 'n eenstadiumsteekproef.

'n Komplekse steekproefontwerp is die resultaat van 'n meerstadium-steekproefontwerp waarin daar van stratifikasie en trossteekproefneming gebruik gemaak word. In 'n komplekse steekproefontwerp word daar in die praktyk in 'n redelike mate afgewyk van die aanname van onafhanklike trekking van steekproefelemente en dit lei outomaties tot 'n sekere mate van onderberaming in die variansies van beramers van populasieparameters. Dit is dus belangrik om variansieberamingstegnieke te gebruik wat die kompleksiteit van die steekproefontwerp in ag neem.

In hoofstuk 2 word die Taylor-linearisasiemetode van variansieberaming beskou. Hierdie metode word gebruik om die standaardfoute van beramers van populasieparameters te bereken en vanaf hierdie standaardfoutwaardes word die ontwerpeffek (deff-waarde) en die intratroskorrelasiekoëffisiënt (roh) bereken. Hierdie drie maatstawwe word vervolgens gebruik om oordraagbaarheid van standaardfoute toe te pas: kyk hoofstuk 2 waar die teorie bespreek en prakties geïllustreer word.

Die Taylor-linearisasiemetode van variansieberaming wat in hoofstuk 2 gebruik word het egter 'n tekortkoming naamlik dat dit beperk is tot die variansieberaming van lineêre beramers (bv. gemiddeldes of verhoudings) van populasieparameters. Hierdie tekortkoming in die Taylor-linearisasiemetode is inherent en twee herhaalde-replikasiemetodes van variansieberaming, wat met gemak nie-lineêre beramers (bv. korrelasie- of regressiekoëffisiënte) kan hanteer, is in die literatuur ontwikkel.

Hierdie twee herhaalde-replikasiemetodes van variansieberaming, naamlik Gebalanseerde Herhaalde-Replikasie (BRR) en Uitsnit Herhaalde-Replikasie (JRR) sowel as die Taylor-linearisasiemetode van variansieberaming, word in detail in hoofstuk 3 bespreek. Hierdie metodes word deur middel van data afkomstig van 'n RGN steekproefopname van huishoudings prakties toegepas deur gebruikmaking van die program CLUSTERS sowel as die BRRJRR program ontwikkel deur die outeur.

## 1.2 KOMPLEKSE MEERSTADIUMSTEEKPROEFNEMING

Komplekse steekproef.

Die beskrywing van die begrippe wat vervolgens bespreek word, word verkry uit die artikel van Verma & Pearce (1978) sowel as 'n geleentheidspublikasie van Stoker (1983). 'n Komplekse steekproef is 'n meerstadiumsteekproef gebaseer op stratifikasie en trosvorming. Seleksiewaarskynlikhede van uiteindelike steekproefnemingseenhede

(USE's) mag gelyk of ongelyk wees.

Stratifikasie word toegepas om 'n heterogene populasie in meer homogene dele (of strata) te verdeel. Veranderlikes wat as stratifikasie-veranderlikes gebruik kan word is byvoorbeeld bevolkingsgroep, stedelike/plattelandse woongebiede, en geografiese streke.

'n Basiese doel van stratifikasie is om die variansies (standaardfoute) van beramers te verklein aangesien die onderskeie strata elk meer homogeen is as wat die populasie in geheel is, en slegs binne-stratum variasie bydra tot die variansie van die beramers. Deur middel van stratifikasie word nie net die variansie van 'n beramer verklein nie, maar groter verteenwoordigendheid van die steekproef ten opsigte van die populasie word ook daardeur verseker.

Na stratifikasie volg trosvorming waardeur elke stratum in verskillende trosse verdeel word, dit is versamelings van populasie-elemente. Die ideaal is dat die trosse so heterogeen as moontlik gevorm moet word, maar omdat trosse gewoonlik uit aangrensende eenhede gevorm word met die oog op kostebesparing, wat immers die doel van trosvorming is, is die trosse gewoonlik meer homogeen as die stratum met die gevolg dat die binnetrosvariasie kleiner as die tussentrosvariasie is. Hierdeur word die variansie van die beramer, wat deur stratifikasie verklein is, weer vergroot. Die vergroting van die variansie van 'n beramer deur trosvorming is in die praktyk gewoonlik meer as die verkleining van die variansie van 'n beramer deur stratifikasie.

Indien die elemente met ongelyke waarskynlikhede getrek word, moet daarvoor gekompenseer word deur gebruikmaking van gewigte.

Stratifikasie word in 'n meerstadiumsteekproef hoofsaaklik gebruik om die volgende twee belangrike redes:

a) om die verteenwoordigendheid van die steekproef te verseker
b) om die variansies van die beramers van die populasieparameters te verklein en sodoende meer presiese beramings met 'n vooraf bepaalde totale steekproefgrootte te verkry, as wat die geval sou wees met 'n eenvoudige ewekansige steekproef.

In 'n komplekse steekproefnemingsituasie word 'n populasie eers gestratifiseer en die populasie-elemente in trosse saamgegroepeer. Die trosse word primêre steekproefeenhede (PSE's) genoem. Die eerste stadium van steekproefneming bestaan uit die ewekansige trek van 'n aantal trosse. Binne elk van die getrekte PSE's word die elemente in kleiner trosse saamgegroepeer, wat dan die sekondêre steekproefeenhede (SSE's) genoem word. In die tweede stadium van steekproefneming word 'n aantal SSE's uit elk van die getrekte PSE's getrek. Die proses kan nou op hierdie wyse voortgesit word. In die derde stadium sal byvoorbeeld tersiêre steekproefeenhede (TSE's) in elk van die getrekte SSE's gevorm word en 'n aantal TSE's sal uit elk van die SSE's ewekansig getrek word. Hierdie TSE's is nog kleiner trosse van elemente van die populasie as die SSE's. In die finale stadium van steekproefneming word enkel populasie-elemente óf klein groepe (trosse) van elemente getrek, wat die uiteindelike steekproefeenhede (USE's) genoem word. Indien die USE's uit klein trosse bestaan is al die populasie-elemente in 'n getrekte USE ingesluit in die steekproef. Indien die SSE's ook die USE's is dit 'n tweestadiumsteekproef.

## Domein van 'n populasie

Wanneer 'n populasie gestratifiseer word, dit wil sê in 'n aantal dele of strata verdeel word, is dit ook moontlik om hierdie strata saam te groepeer in domeins. 'n Domein bevat dus 'n aantal strata en vorm 'n deel van die populasie. Elke stratum (en gevolglik ook PSE, SSE, TSE .... of USE in die stratum) behoort tot een en slegs een domein. As alle huishoudings in die RSA die populasie vorm, dan kan alle stede of dorpe in die RSA as strata beskou word. Hierdie strata kan in twee domeins verdeel word naamlik alle stedelike huishoudings en alle nie-stedelike huishoudings. Elke stratum (huishouding) behoort dus tot slegs een van die twee domeins.

## Subgroepe van 'n steekproef

In die analises van meerstadiumsteekproewe word baie van subgroepe gebruik gemaak. 'n Aantal elemente van 'n steekproef wat een of meer eienskap in gemeen het, vorm 'n subgroep in die steekproef. Voorbeelde van subgroepe is etniese groepe, ouderdomsgroepe, of byvoorbeeld Blankes tussen die ouderdomme 18 en 46 jaar.

Dit is belangrik om tussen twee soorte subgroepe te onderskei:

i) Kruisgroepe is groepe wat in elk van die steekproefnemingstadia min of meer eenvormig versprei is oor die strata en trosse van populasie-elemente. Ouderdomsgroepe, geslag en soms ook inkomste is voorbeelde van kruisgroepe.

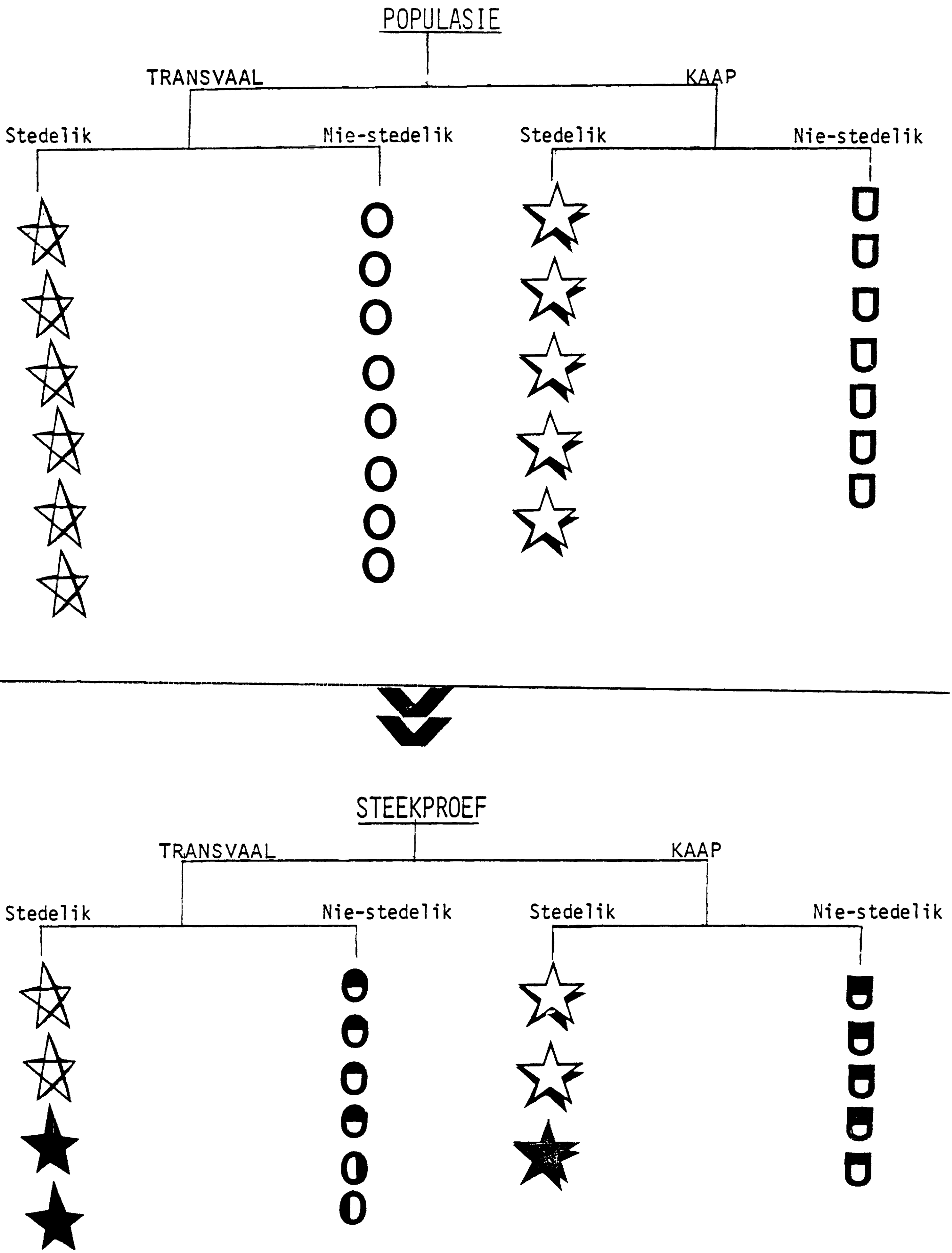
ii) Gesegregeerde groepe is subgroepe wat strata of primêre steekproefnemingseenhede in 'n steekproef in hul geheel bevat. Voorbeeld: geografiese gebiede nl. landelik of stedelik, en etniese groepe.

Die verskillende begrippe wat in die voorafgaande paragrawe bespreek is, word vervolgens met behulp van 'n voorbeeld en 'n skets toegelig en in verband met mekaar gebring.

Beskou die populasie van alle huishoudings in die Transvaal en Kaap-provinsie as teikenpopulasie. Vir sensusdoeleindes is al die huishoudings in opnemersubdistrikte (OSD's) ingedeel. Gebruik hierdie OSD's as PSE's. Groepeer dan die PSE's in strata deur as stratifikasieveranderlikes te neem: provinsie en stedelik/nie-stedelike woongebiede. Enige versameling van strata staan as 'n domein bekend byvoorbeeld die domein van alle stedelike huishoudings. 'n Domein is dus 'n deel van die teikenpopulasie. Gestratifiseerde trossteekproefneming impliseer dan dat uit elke stratum 'n trossteekproef (bestaande uit PSE's) getrek word.

Gesegregeerde subgroepe is deelversamelings van 'n steekproef en besit die eienskap dat dit PSE's in die geheel bevat. 'n Gesegregeerde subgroep in hierdie steekproef kan die verskillende PSE's van 'n bepaalde bevolkingsgroep in die steekproef wees. In die skets is slegs stedelike huishoudings beskou om 'n gesegregeerde subgroep te illustreer.

Kruisgroepe bevat nie gehele PSE's nie, maar sny deur die grense van die PSE's, byvoorbeeld die steekproef van manlike respondente. In die skets is nie-stedelike huishoudings gebruik om die begrip kruisgroepe te illustreer.




Gesegregeerde groep : Alle stedelike Blankes en bevat PSE's in geheel.

Kruisgroep : Alle nie-stedelike mans. en Ingekleurde gedeeltes van elke PSE.

## 1.3. STANDAARDFOUTE IN KOMPLEKSE MEERSTADIUMSTEEKPROEFNEMING

Die beraamde waarde van die standaardfout van 'n beramer van 'n populasieparameter (soos'n totaal, 'n gemiddelde en 'n proporsie) gee 'n aanduiding van die presisie van die beramer wat gebruik is. Om standaardfoutberamings in komplekse meerstadiumsteekproefneming te kan doen is dit nodig om minstens 2 PSE's per stratum te trek. Indien dit nie die geval is nie moet naasliggende strata twee-twee saamgegroepeer word. Hierdie tegniek staan bekend as die geamalgameerde-stratum tegniek.

Indien 'n steekproef nie-selfwegend is, moet die gewigte, dit wil sê die ongelyke seleksiewaarskynlikhede van die steekproefelemente, in die berekening van die standaardfoute in ag geneem word.

Die variansie van 'n beramer word afgelei uit die binne-stratumvariasie van die geweegde PSE-totaalwaardes van die relevante veranderlike.

Om die standaardfout van enige beramer in 'n meerstadiumsteekproef-ontwerp te bereken, is die volgende inligting ten opsigte van elke element noodsaaklik:

i) tot welke PSE die element behoort

ii) tot welke stratum die PSE behoort

iii) die gewigsfaktor geassosieer met die element

iv) die waardes van alle relevante veranderlikes vir die element.

## 1.4 DIE DRIE MAATSTAWWE VAN OORDRAGING

In hoofstuk 2 word die begrip oordraagbaarheid bespreek. Oordraging beteken dat die waarde van 'n berekende standaardfout van 'n beramer soos bepaal vir 'n subgroep, in die plek gestel word van die onbekende standaardfout van dieselfde beramer vir dieselfde of soortgelyke veranderlikes in 'n ander vergelykbare subgroep. Onder 'n soortgelyke veranderlike word 'n veranderlike verstaan wat baie nou verband hou met die oorspronklike veranderlike, sodat beide in wese dieselfde inherente eienskap van die USE net op twee maniere uitdruk. Oordraagbaarheid van standaardfoute kan ook toegepas word binne 'n subgroep. Die tevredenheid van 'n persoon ten opsigte van sy stemreg en die belangrikheid van politieke aangeleenthede vir die individu kan beide gesien word as twee veranderlikes wat sorteer onder politieke aangeleenthede. Hierdie twee veranderlikes kan as soortgelyke veranderlikes beskou word. Die begrip oordraagbaarheid het betrekking op die mate waarin die onbekende waarde van 'n dergelike standaardfout uit die bekende waarde van 'n 'soortgelyke' standaardfout afgelei of verkry kan word. Oordraging geskied deur die proses van imputasie. Imputasie kan beteken dat die standaardfoutwaardes direk oorgedra word, óf dat dit deur middel van deff-waardes, óf roh-waardes kan geskied, waarby soms van 'n imputasiefaktor gebruik gemaak word. Laasgenoemde drie maatstawwe naamlik standaardfout, ontwerpeffek (deff) en intratroskorrelasie-koëffisiënt (roh) word kortliks bespreek.

Die standaardfoute van beramers van populasieparameters kan baie insiggewende inligting omtrent die presisie van 'n steekproefopname verskaf.

Die ontwerpeffekwaarde, deff-waarde, is per definisie die verhouding van die steekproefvariansie van 'n beramer van 'n populasieparameter wanneer die totale kompleksiteit van die steekproef in ag geneem word tot die steekproefvariansie van dieselfde beramer wanneer die totale kompleksiteit van die steekproef geïgnoreer word, dit wil sê wanneer die steekproef as 'n eenvoudige ewekansige steekproef beskou word (kyk paragraaf 2.3.3).

Die intratroskorrelasiekoëffisiënt (roh) is 'n maatstaf van die variasie in die veranderlike onder beskouing binne die PSE's van die steekproef relatief tot die variasie in die veranderlike oor die hele hele steekproef (kyk paragraff 2.3.4).

Die program CLUSTERS (ontwikkel deur die World Fertility Survey) word in die praktiese toepassing van hoofstuk 2 gebruik om oordraagbaarheid deur middel van imputasie van bogenoemde drie maatstawwe onderskeidelik onder sekere omstandighede te illustreer.

## 1.5 BRR EN JRR TEGNIEKE

Dit is nie korrek om die gewone formules vir die berekening van standaardfoute wat berus op die aanname van onafhanklike waarnemings te gebruik wanneer data afkomstig van 'n komplekse meerstadiumsteekproef geanaliseer word nie, aangesien die aanname nie noodwendig vir laasgenoemde steekproefontwerp korrek is nie.

In hoofstuk 3 van hierdie verhandelig word drie variansieberamings-tegnieke vir komplekse steekproefdata in detail bespreek. Hierdie drie variansieberamingsmetodes is die Taylor-linearisasiemetode, die Gebalanseerde Herhaalde-Replikasie of BRR-metode (BRR = 'Balanced Repeated Replication') en die Uitsnit Herhaalde-Replikasie of JRR-metode (JRR = 'Jackknife Repeated Replication').

Die Taylor-linearisasiemetode van variansieberaming het 'n groot tekortkoming naamlik dat dit beperk is tot die beraming van variansies van lineêre beramers (bv. gemiddeldes en verhoudings) van populasieparameters. Toepassing van die metode op nie-lineêre beramers veroorsaak sydigheid in die variansieberaming. Die twee herhaalde-replikasiemetodes besit 'n definitiewe voordeel bo die Taylor-linearisasiemetode van variansieberaming, aangesien BRR en JRR spesiaal ontwikkel is om die variansies van nie-lineêre beramers (bv. regressie- en korrelasiekoëffisiënte) van populasieparameters te kan hanteer.

'n Program BRRJRR wat gebruik maak van die BRR- en JRR-metodes van variansieberaming is ontwikkel en gekombineer met die program CLUSTERS sowel as met 'n program van R.S. Jewett (verkry deur persoonlike kommunikasie). BRRJRR is tot dusver die enigste program van sy soort in Suid-Afrika, en word in hoofstuk 3 behandel waar dit ook toegepas word op 'n RGN steekproefopname van huishoudings.

# II. MAATSTAWWE VAN OORDRAAGBAARHEID

## 2.1 INLEIDING

Die berekening van standaardfoute van beramers van populasieparameters gee vir die navorser 'n baie goeie aanduiding van die presisie van die beramings wat hy verkry het.

Vanaf die standaardfoute kan twee ander baie insiggewende maatstawwe bereken word naamlik die ontwerpeffek, deff, ('design effect') en die intratroskorrelasiekoëffisiënt, roh, ('rate of homogeneity'). Hierdie drie maatstawwe kan gebruik word om die oordraging van standaardfoute te bewerkstellig.

Soos reeds in hoofstuk 1 gedefinieer beteken oordraging dat die waarde van 'n berekende standaardfout van 'n beramer soos bepaal vir 'n subgroep, in die plek gestel word van die onbekende standaardfout van dieselfde beramer vir dieselfde of soortgelyke veranderlikes in 'n vergelykbare subgroep.

Oordraagbaarheid kan geskied deur middel van drie maatstawwe naamlik standaardfoute, deff-waardes en roh-waardes. Hierdie drie maatstawwe word in detail in hierdie hoofstuk bespreek sowel as die verskillende omstandighede waaronder elk oordraagbaar is.

Die teorie van hierdie hoofstuk word aan die einde van die hoofstuk aan die hand van 'n RGN-steeproefopname en met behulp van die program CLUSTERS prakties geïllustreer.

### 2.1.1 Doel van standaardfoute

'n Navorser wat 'n hoeveelheid ingesamelde data het, het 'n sekere kriterium nodig waarvolgens hy die presisie van die resultate van sy steekproefopname kan meet. Hierdie kriterium kan gebaseer word op die groottes van die berekende waardes van die standaardfoute van die verskillende beramers wat gebruik is.

Standaardfoute kan vir die volgende twee doeleindes bereken word:

1) Om die presisie van die resultate van die steekproefopname te bepaal.
2) Om inligting te voorsien vir die beplanning van soortgelyke toekomstige steekproefopnames.

### 2.1.2 Beraming van standaardfoute

Laat $\bar{y}$ die steekproefgemiddelde ('n steekproefproporsie of steekproef-totaal kan ook as 'n gemiddelde geskryf word) wees, wat 'n puntberamer van die waarde van die populasiegemiddelde $\bar{Y}$ is. Die verdeling wat bestaan uit alle moontlike waardes wat $\bar{y}$ kan aanneem by soortgelyke steekproeftrekking uit die eindige populasie tesame met die geassosieerde waarskynlikhede word die steekproefverdeling van $\bar{y}$ genoem. Die verwagte waarde van die puntberamer is die gemiddelde van die steekproefverdeling en word gegee deur:

$$\bar{Y} = E(\bar{y}) = \sum_{s} p_s \cdot \bar{y}_s$$

waarby $\bar{y}_s$ die gemiddeld van die s-de steekproef aandui, $p_s$ die waarskynlikheid om die s-de steekproef te trek aandui, en $\sum_s$ die sommasie oor alle moontlike verskillende steekproewe (van gelyke grootte) aandui.

Die variansie, $\sigma^2_{\bar{y}}$, van $\bar{y}$ is 'n maatstaf van die variasie in die steekproefverdeling van $\bar{y}$ en dit word as volg gedefinieer

$$\sigma^2_{\bar{y}} \equiv \mathrm{Var}(\bar{y}) = p_s[\bar{y}_s - E(\bar{y})]^2$$

Die standaardfout, $\sigma_{\bar{y}}$, is slegs die vierkantswortel van die variansie $\sigma^2_{\bar{y}}$. In die praktyk is $\sigma_{\bar{y}}$ onbekend en moet dit beraam word. Hieraan word aandag gegee in paragraaf 2.1.4 in die geval van komplekse steekproewe.

### 2.1.3 <u>Vertrouensintervalle</u>

Inferensie oor die "waardes" van populasiekarakteristieke word dikwels gemaak in terme van vertrouensintervalle. Intervalberamings vir populasiegemiddeldes word verkry vanuit die steekproefverdeling van $\bar{y}$.

Op grond van die sentrale limietstelling is die steekproefverdeling van $\bar{y}$ asimptoties normaal. Hoe groter die steekproefgrootte, hoe beter is die normaalbenadering. Uit normaalteorie word die 95% vertrouensinterval vir $\bar{Y}$, die populasiegemiddelde, gegee deur:

$$\bar{y} \pm 1{,}96\,\sigma_{\bar{y}}$$

waarby $\bar{y}$ die gemiddelde van 'n spesifieke steekproef aandui. Vir 'n meer gedetailleerde bespreking van die vertrouensintervalle, kyk Verma (1982).

### 2.1.4 Berekening van standaardfoute vir komplekse steekproewe.

In hoofstuk 1 is 'n meerstadiumsteekproefopname bespreek. Aangesien elke steekproefnemingstadium bydra tot die variasie in die steekproefgrootheid, is dit noodsaaklik om hierdie verskillende steekproefnemingstadia in berekening te bring by die berekening van standaardfoute vir meerstadium komplekse steekproewe.

Die basiese aannames waaraan voldoen moet word vir die berekening van standaardfoute in 'n meerstadiumsteekproefopname is die volgende:

a) Twee of meer PSE's (kyk hoofstuk 1) moet uit elke stratum getrek word.

b) Die PSE's moet onafhanklik van mekaar en met teruglegging getrek word.

Beskou 'n tweestadiumsteekproefnemingsontwerp, waarby die populasie uit H strata bestaan. Laat stratum h uit $M_h$ PSE's bestaan waarby elke PSE uit 'n aantal USE's bestaan. Laat vir die i-de PSE in stratum h geld dat:

$y_{hij}$ = die waarde van die veranderlike y vir die j-de getrekte USE in die i-de PSE van die h-de stratum.

$y_{hi}$ = die geweegde som van al die getrekte USE's in die i-de PSE van die h-de stratum.

$= \sum_{j=1}^{m_{hi}} w_{hij} y_{hij}$ waarby $w_{hij}$ die steekproefgewigte vir die getrekte USE's is en $m_{hi}$ die totale aantal getrekte USE's in die i-de getrekte PSE van stratum h is. (Die USE's in die steekproef word geweeg om te kompenseer vir ongelyke seleksie waarskynlikhede of vir ongelyke nie-respons proporsies).

$y_h$ = die som van die waardes van y geneem oor al die getrekte USE's in stratum h.

$= \sum_{i=1}^{m_h} y_{hi}$

Hier is $m_h$ die aantal getrekte PSE's in die h-de stratum.

en

y = die som van die waardes van y geneem oor al die USE's in die steekproef

$= \sum_{h=1}^{H} y_h.$

Soortgelyk vir 'n veranderlike x, wat ook ter sprake mag wees.

Die steekproefvariansie van die $y_{hi}$'s en van die $y_h$'s word gegee deur

$$s_{yh}^2 = [m_h/(m_h - 1)] \sum_{i=1}^{m_h} (y_{hi} - y_h/m_h)^2$$

$$= [m_h/(m_h-1)] [\sum_i y_{hi}^2 - y_h^2/m_h] \ldots\ldots\ldots\ldots\ldots\ldots (2.2.1)$$

en

$$s_y^2 = \sum_h s_{yh}^2$$

waarby $m_h$ die aantal gekose PSE's in stratum h is. $s_{yh}^2$ en $s_y^2$ is onderskeidelik puntberamers van die ooreenkomstige populasievariansies $\sigma_{yh}^2$ en $\sigma_y^2$. Die standaardfoute is die vierkantswortel van die variansies $s_{yh}^2$ en $s_y^2$. Alle voorgaande begrippe kan ook vir x gedefinieer word, waarby x enige veranderlike aandui.

Beskou vervolgens die verhoudingsberamer

$$r = \frac{\sum_h\sum_i\sum_j w_{hij} \cdot y_{hij}}{\sum_h\sum_i\sum_j w_{hij} \cdot x_{hij}}$$

$$= \frac{\sum_h\sum_i y_{hi}}{\sum_h\sum_i x_{hi}}$$

$$= \sum_h y_h / \sum_h x_h$$

$$= y/x$$

Die verhoudingsberamer word baie algemeen in die praktyk gebruik. Spesiale gevalle van verhoudingsberamers is proporsies en gemiddeldes. Vir proporsies sowel as vir gemiddeldes is die waarde van $x_{hij}$ gelyk aan een (indien die eienskap onder beskouing realiseer) andersins gelyk aan nul.

Om die benaderde variansie van r te bereken word gebruik gemaak van 'n bekende resultaat in Kish & Hess (1959), naamlik dat tot die eerste orde van benadering geld vir groot steekproewe:

$$s_r^2 \equiv \text{var}(r)$$

$$\equiv \text{var}(y/x) \equiv (1-f)/x^2.[s_y^2 + r^2 s_x^2 - 2rs_{xy}] \qquad (2.2.2)$$

waarby $f = n/N$ die steekproefverhouding genoem word en

$$s_y^2 \equiv \text{var}(y) = \sum_{h=1}^{H} m_h/(m_h-1)[\quad y_{hi}^2 - y_h^2/m_h] \qquad (2.2.3a)$$

soortgelyk is

$$s_x^2 \equiv var(x) = \sum_h [m_h/(m_h-1)][\sum_i x_{hi}^2 - x_h^2/m_h] \qquad (2.2.3b)$$

Die kovariansie term word as volg gedefinieer:

$$s_{xy} = kov(x,y) = \sum_h [m_h/(m_h-1)][\sum_i y_{hi}x_{hi} - y_h x_h/m_h] \qquad (2.2.4)$$

Deur $s_y^2$, $s_x^2$ en $s_{xy}^2$ se vergelykings in (2.2.2) te vervang volg:

$$s_r^2 \equiv var(r) = (1-f)/x^2 \sum_h [m_h/(m_h-1)][\sum_i y_{hi}^2 - y_h^2/m_h$$
$$+ r^2 \sum_h x_{hi}^2 - r^2 x_h^2/m_h - 2r\{\sum_i y_{hi}x_{hi} - y_h x_h/m_h\}]$$

$$= [(1-f)/x^2] \sum_h \{m_h/(m_h-1)\}[\sum_i y_{hi}^2 - r^2\sum_i x_{hi}^2 - 2r\sum_i y_{hi}x_{hi}$$
$$-\{y_h^2/m_h + r^2 x_h^2/m_h - 2r y_h x_h/m_h\}]$$

$$= [(1-f)/x^2][\sum_h \{m_h/(m_h-1)\}[\sum_i (y_{hi}^2 - r x_{hi})^2 - 1/m_h(y_h - r x_h)^2]]$$

$$= [(1-f)/x^2][\sum_h \{m_h/(m_h-1)\}\{\sum_i z_{hi}^2 - z_h^2/m_h\}] \qquad (2.2.5)$$

waar

$$z_{hi} = y_{hi} - r x_{hi}$$

en

$$z_h = \sum_i z_{hi}$$
$$= y_h - r x_h$$

Let wel

$$z = \sum_h z_h$$
$$= y - rx$$
$$= y - (y/x)x$$
$$= 0$$

Bogenoemde berekeningsformule kan ook gebruik word vir die berekening van die variansie van r vir 'n subgroep van 'n steekproef. In hierdie geval sal die USE's wat nie tot die subgroep behoort nie buite rekening gelaat word. Die noemer in 'n verhoudingsberamer is dikwels 'n telveranderlike, sodat x dan die totale steekproefgrootte is. Die variasie binne die trosse dra by tot die variansie in die verhoudingsberamer r, deur middel van die terme $s_x^2$ en $s_{xy}$ indien gemiddeldes, proporsies of totale beskou word, waar $x_{hij}$ die waarde een of nul kan aanneem.

Dit is bewys (Verma 1982) dat die sydigheid van 'n verhoudingsberamer relatief tot die standaardfout kleiner is as die relatiewe koëffisiënt van variasie, (cv(x)), (kyk paragraaf 2.1.5), en dat dit verminder soos wat die aantal steekproefeenhede verminder. In 'n goed ontwerpte steekproef waar daar min variasie in die trosgrootte voorkom is die sydigheid gewoonlik weglaatbaar klein.

Indien die variansie van die verskil tussen twee waardes van verhoudingsberamers bereken wil word, sal die volgende berekeningsformule toegepas word (Kish (1965)):

$$var(r-r') = var(r) + var(r') - 2kov(r,r') \qquad (2.2.6)$$

waarby

$$r-r' = y/x - y'/x'$$

en

$$kov(r,r') = [1/(xx')][kov(y,y') + rr'kov(x,x') - rkov(y',x) - r'.kov(y,x')] \qquad (2.2.7)$$

Daar is verskeie situasies waar die verskil tussen twee verhoudings beskou word. Twee verskillende eienskappe in dieselfde steekproef kan vergelyk word. 'n Meer algemene verskynsel is die vergelyking van dieselfde eienskap tussen twee subgroepe van dieselfde steekproef waarby die subgroepe uit onderling uitsluitende kategorieë in die noemer bestaan. 'n Voorbeeld van so 'n situasie is om die gemiddelde ouderdom waarby vroue van twee verskillende ouderdomsgroepe (vb. 25-34 en 35-44) in die huwelik tree met mekaar te vergelyk. Die kovariansie term in (2.2.6) sal slegs voorkom indien USE's gemeenskaplik in beide die verhoudings voorkom of indien die steekproefeenhede vanuit dieselfde PSE's voorkom.

### 2.1.5 Koëffisient van relatiewe variasie

Die koëffisiënt van relatiewe variasie van die veranderlike x word aangedui deur cv(x), en as volg gedefinieer:

$$cv(x) \equiv (s_x^2)^{\frac{1}{2}}/\bar{x} = s_x/\bar{x} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2.2.8)$$

Hierdie koëffisiënt van variasie verminder soos wat die aantal USE's per PSE verminder (Kish 1965). Vir gestratifiseerde trossteekproewe word cv(x) beraam deur:

$$cv(x) = (1/x)[\sum_h [m_h/(m_h-1)](\sum_i x_{hi}^2 - (x_h)^2/m_h)]^{\frac{1}{2}} \ldots\ldots (2.2.9)$$

waarby $x_{hi}$ die geweegde waarde vir die veranderlike x in die i-de getrekte PSE van die h-de stratum is, en $x_h$ gelyk is aan die som van die waardes van x geneem oor al die getrekte USE's in stratum h.

'n Klein variasie in die noemer van 'n verhoudingsberamer, $y/x$, beperk die sydigheid in die beraming van die variansie van die verhoudingsberamer. Om die sydigheid klein te hou moet die koëffisiënt van relatiewe variasie wat 'n funksie van die sydigheid is, klein wees. Indien die subgroepe (kruisgroepe of gesegregeerde groepe) se trosgroottes grootliks varieer, of indien die steekproefelemente in die subgroepe min is of swak versprei is oor die verskillende PSE's in die steekproef, dan kan dit 'n groot koëffisiënt van variasie in x tot gevolg hê. Anders gestel, vir 'n goed ontwerpte steekproef met min variasie in die trosgroottes van die subgroepe, sal die koëffisiënt van relatiewe variasie in x ($cv(x)$) relatief klein wees, d.w.s. die sydigheid van x sal weglaatbaar klein wees.

Vir steekproewe met 'n groot aantal strata en 'n beperkte variasie in die trosgroottes (aantal USE's in 'n PSE), is 'n goeie benadering vir die koëffisiënt van relatiewe variasie van x die volgende:

$$cv(x) = 2/g.v.v.$$

waarby g.v.v. die grade van vryheid aandui wat verkry word deur die aantal PSE's in die steekproef minus die aantal strata te bereken (Kish 1965: 289-291).

<u>Voorbeeld</u>.

Indien twee PSE's per stratum getrek word en daar is H strata in die steekproef, dan sal

$$g.v.v. = 2H-H = H$$

$$cv(x) \cong 2/H$$

Vir 100 getrekte PSE's en 50 strata sal

g.v.v.= 100-50 = 50

$cv(x) \cong 2/50$

$= 0{,}04$

terwyl vir 25 getrekte PSE's met 12 strata sal

g.v.v. = 25-12 = 13

$cv(x) \cong 2/13$

$= 0{,}15$

Dit is belangrik om daarop te let dat groot sydighede en onstabiliteit vermy word indien cv(x) kleiner as 0,2 is en verkieslik kleiner as 0,1.

## 2.2 PATRONE VAN VARIASIE EN OORDRAAGBAARHEID

Daar is verskeie redes waarom dit handig is om die patrone van variasie in die waardes van standaardfoute, geneem oor verskeie veranderlikes, subgroepe en steekproefopnames, te ondersoek. Hierdie patrone word dan in verband gebring met die struktuur van 'n steekproef. 'n Paar van hierdie redes word kortliks uiteengesit.

a) Oordraging van berekende resultate.

Beskou 'n groot ondersoek met 'n hele verskeidenheid veranderlikes wat ontleed moet word. Om in so 'n ondersoek die standaardfoute vir die beramers van al die populasieparameters vir al die moontlike veranderlikes te gaan bereken, is 'n byna onmoontlike taak, aangesien ook die standaardfoute van die beramers in verskeie kruis- sowel as gesegregeerde groepe gewoonlik ook bereken moet word.

Vir bogenoemde probleem is 'n praktiese oplossing die oordraging van berekende waardes van standaardfoute van beramers soos verkry in verskillende subgroepe in 'n steekproef, na soortgelyke beramers in dié subgroepe waarvoor die standaardfoute nog nie bereken is nie. Voordat hierdie oordraging van die standaardfoute van die verskillende beramers moontlik is, is 'n deeglike ondersoek, hetsy deur 'n loodsstudie of uit vorige soortgelyke ondersoeke, na die groottes en variasie van die standaardfoute vir beramers van populasieparameters vir verskillende soortgelyke veranderlikes, ten opsigte van verskillende subgroepe noodsaaklik.

b) Weergawe van resultate in 'n verslag.

Dit is belangrik dat resultate in 'n verslag duidelik en maklik leesbaar moet wees. Indien die standaardfoute vir al die beramers vir alle moontlike subgroepkombinasies in tabelle weergegee word kan dit 'n verslag onnodig omvangryk maak. Dit is dus wenslik om 'n eenvoudige formule vir die berekening van benaderde waardes van standaardfoute van beramers wat nie bereken is nie, daar te stel. Weer eens noodsaak dit oordraging van standaardfoute en daarvoor is kennis aangaande die verspreidingsvorme van die standaardfoute nodig.

c) Modellering van berekende resultate.

Indien daar min PSE's in die totale steekproef is, is daar 'n groot moontlikheid dat die standaardfoute van beramers vir populasieparameters grootliks sal varieer. In so 'n geval is dit wenslik om soortgelyke veranderlikes te beskou en dan die gemiddelde standaardfout van beramers wat op hierdie veranderlikes betrekking het te bereken eerder as om op die resultate van standaardfoute van individuele beramers staat te maak. Met modulering word dus bedoel die sodanige berekening van die gemiddeldes van die standaardfoute.

d) Steekproefontwerp en evaluasie.

'n Beskouing van die groottes van die standaardfoute van die beraming van 'n spesifieke populasieparameter (byvoorbeeld populasietotaal) vanuit die onderskeie subgroepe gee 'n aanduiding van:

i) Die presisie van die steekproefopname
ii) Inligting vir die ontwerp van toekomstige steekproefopnames.

Opsomming.

Uit bogenoemde bespreking volg dat inligting wat verkry word deur die patrone van variasie van standaardfoute van beramers deeglik te beskou van groot waarde vir enige navorser kan wees. Die inligting wat deur oordraging en modellering van standaardfoute verkry word, kan tot misleidende inligting aangaande die ware waardes van die variasies in standaardfoute lei indien dit nie met omsigtigheid toegepas word nie. Laasgenoemde probleem kan grootliks oorbrug word indien ekstrapolasie en gladstrykings van standaardfoute van beramers van populasieparameters toegepas word op baie verskillende soortgelyke veranderlikes vir verskillende soortgelyke subgroepe.

## 2.3 OORDRAAGBAARHEID EN IMPUTASIE

### 2.3.1 Inleiding

Oordraagbaarheid is vir die eerste keer deur Kish & Frankel (1974) toegepas en beteken die volgende: Veronderstel die standaardfout van die beramer van 'n populasieparameter in 'n subgroep is bekend en hierdie standaardfout word in die plek gestel van die onbekende standaardfout van dieselfde beramer van dieselfde populasieparameter vir dieselfde veranderlikes in 'n soortgelyke subgroep, dan staan dit as oordraging bekend. Oordraagbaarheid kan ook toegepas word om die berekende standaardfoute van veranderlikes na die onbekende standaardfoute van soortgelyke veranderlikes vir dieselfde of soortgelyke subgroepe oor te dra.

Oordraging beteken dus dat die berekende waarde van die standaardfout van 'n beramer soos bepaal vir 'n subgroep, gebruik word om die onbekende waarde van die standaardfout van dieselfde beramer vir dieselfde of soortgelyke veranderlikes in 'n ander soortgelyke subgroep te verkry. Oordraagbaarheid van standaardfoute kan ook toegepas word in dieselfde subgroep op die beramers van populasieparameters vir verskeie soortgelyke veranderlikes.

Daar is hoofsaaklik drie maatstawwe wat onder verskillende omstandighede gebruik kan word om oordraagbaarheid mee te bewerkstellig. Hierdie maatstawwe is die standaardfoute self, die ontwerpeffek (deff) en die intratroskorrelasiekoëffisiënt (roh). Laasgenoemde twee maatstawwe het 'n breër spektrum van toepasbaarheid as standaardfoute self.

Oordraagbaarheid in terme van hierdie drie maatstawwe word vervolgens bespreek, sowel as die omstandighede waaronder elk vir oordraging gebruik kan word.

### 2.3.2 Standaardfoute

Indien oordraagbaarheid toegepas wil word op beramers van soortgelyke (kyk paragraaf 1.4) veranderlikes vir subgroepe van ongeveer dieselfde grootte dan kan dit direk deur middel van die standaardfoute gedoen word.

Die volgende faktore moet ooreenstem by die toepassing van oordraagbaarheid deur middel van standaardfoute:

a) Die tipe veranderlike wat beraam word.
b) Die eenheid waarin gemeet is.
c) Die aard van die statistiek waardeur die veranderlike beraam word.
d) Die grootte van die subgroepsteekproef (wat grootliks kan varieer vanaf een subgroep na 'n ander).
e) Die grootte van die USE's. (Let wel: die USE's kan uit individuele elemente of uit klein trosse elemente (waarnemings) bestaan).
f) Die tipe waarnemings wat gemaak is.

Bogenoemde faktore speel 'n rol indien standaardfoute direk in oordraging toegepas wil word.

### 2.3.3 Ontwerpeffek (Deff = Design effect)

By die ontwerp van 'n steekproefplan vir 'n steekproefopname van byvoorbeeld huishoudings is daar verskeie faktore wat in ag geneem moet word waaronder:

a) Wat is reeds bekend aangaande die populasie onder beskouing.

b) Is daar kaarte van die steekproefopnamegebiede beskikbaar.

c) Hoe toeganklik is die gebiede.

d) Hoeveel veldwerkers is beskikbaar.

e) Die totale oorhoofse kostes verbonde aan die steekproefopname.

Bogenoemde faktore speel elkeen 'n rol by die ontwerp van 'n steekproefplan. Om 'n steekproef so prakties en ekonomies moontlik te ontwerp is dit dikwels noodsaaklik om van meerstadiumtrossteekproefneming gebruik te maak. In hierdie geval is dit belangrik om te weet wat die invloed op die variasie is deur die toepassing van hierdie meer ekonomiese (in terme van koste) steekproefontwerp. Om hierdie effek te meet, is die konsep van die ontwerpeffek (deff) ingevoer. Deff is per definisie die verhouding van die variansie van die beramer onder beskouing wanneer die kompleksiteit van die steekproef in ag geneem word tot die variansie van dieselfde beramer indien die totale kompleksiteit van die ontwerp geheel en al geïgnoreer word, dit is wanneer die steekproef as 'n eenvoudige ewekansige steekproef beskou word. Indien die variansie van 'n beramer as 'n maatstaf van effektiwiteit beskou word meet deff dus die effektiwiteit van die komplekse steekproefnemingsontwerp teenoor die effektiwiteit van die ewekansige steekproefnemingsontwerp. Die totale kompleksiteit van 'n steekproefontwerp word veroorsaak deur stratifikasie, trosvorming en moontlike ongelyke seleksiewaarskynlikhede.

As $s_{es}^2$ die variansie van 'n beramer aandui in die geval van eenvoudige ewekansige steekproefneming en $s_{ts}^2$ die variansie in die geval 'n komplekse steekproefnemingsontwerp voorstel, waarby die trosse in beide gevalle van gelyke grootte M is, en die steekproefelemente met gelyke waarskynlikhede getrek is, dan geld vir N (die populasiegrootte) groot die volgende verband (Kish 1965):

$$s_{ts}^2 \cong \{1 + (M-1)\rho\}s_{es}^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(2.3.1)$$

of

$$\text{deff} \cong s_{ts}^2/s_{es}^2 \cong 1 + (M-1)\rho \quad \ldots\ldots\ldots\ldots\ldots\ldots(2.3.2)$$

waarby $\rho$ die intratroskorrelasiekoëffisiënt is (kyk paragraaf 2.3.4).

Per definisie is

$$\text{deff} = \frac{\text{variansie van 'n beramer vir 'n komplekse steekproef}}{\text{variansie van 'n beramer vir 'n eenvoudige ewekansige steekproef}}$$

Die vierkantswortel van deff word aangedui deur

$$\text{deft} = \frac{\text{standaardfout van 'n beramer vir 'n komplekse steekproef}}{\text{standaardfout van 'n beramer vir 'n eenvoudige ewekansige steekproef}}$$

$$= s_{ts} / s_{es} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(2.3.3)$$

$$\Rightarrow \quad \text{deft} = (\text{deff})^{\frac{1}{2}}$$

Buiten die ontwerpeffek (deff) is daar ook 'n geweegde ontwerpeffek (weff). Deff en weff verskil van mekaar ten opsigte van die noemer. Die noemer in deff bestaan uit die variansie van 'n eenvoudige ewekansige steekproef waarby die totale kompleksiteit van die steekproef geïgnoreer word. Die noemer in weff daarenteen bestaan uit die variansie van 'n ewekansige steekproef waarby 'n gedeelte van die kompleksiteit van die steekproef geïgnoreer word nl: stratifikasie en trosvorming. In die noemer van deff word ongelyke seleksiewaarskynlikhede geïgnoreer by die berekening van die variansie van die beramer, maar nie in noemer van weff nie.

Deur middel van deff word die effek van die totale kompleksiteit van die steekproef gemeet en vir die doeleindes van die bespreking sal slegs deff verder beskou word.

In die praktyk is ontwerpeffekte in die meerderheid gevalle 'n beter maatstaf vir oordraagbaarheid as standaardfoute en die rede daarvoor is die volgende:

Faktore wat beide $s_{es}^2$ (variansie van 'n beramer vir 'n eenvoudige ewekansige steekproef) en $s_{ts}^2$ (variansie van 'n beramer vir 'n komplekse steekproef) in dieselfde mate beïnvloed, het geen effek op die waarde van deff nie. Uit (2.3.2) is dit duidelik dat sulke faktore mekaar sal uitkanselleer. Bogenoemde faktore behels die volgende:

a) Die waarde van die beramer.
b) Die variansie van die veranderlike in die populasie.
c) Die steekproefgrootte mits trosgroottes dieselfde bly.

Die volgende faktore het wel nog 'n invloed op die waarde van deff:

a) Die tipe beramer dit wil sê gemiddelde, korrelasiekoëffisiënt, regressiekoëffisiënt, ensovoorts.
b) Die ontwerp van die steekproef.

c) Die tipe waarnemings wat ontleed word en die aard van die USE's (dit wil sê bestaan die USE's uit individuele elemente of uit klein trosse).

Elke beramer het sy eie standaardfout wat impliseer dat die ontwerpeffek nie 'n waarde is wat aan die steekproef gekoppel is nie, maar dat vir die beramers van populasieparameters vir elk van die verskillende veranderlikes 'n ontwerpeffek-waarde bestaan.

Die faktore wat die grootste invloed op die waarde van die ontwerpeffek uitoefen is trosvorming en stratifikasie. Die ontwerpeffek is 'n omvattende maatstaf wat poog om die effek van veral trosvorming en stratifikasie weer te gee. In hoofstuk 1 is daarop gewys dat een van die doelstellings van stratifikasie is om die variansie van 'n beramer te verklein. Trosvorming daarenteen neig om die variansie van 'n beramer van 'n populasieparameter te vergroot. Gevolglik sal deur middel van stratifikasie die waarde van deff in die algemeen kleiner as een word, terwyl trosvorming die teenoorgestelde effek sal meebring. Indien die deff-waarde een is impliseer dit dat die komplekse steekproef net so effektief is as 'n eenvoudige ewekansige steekproef. 'n Deff-waarde kleiner as een impliseer dat deur komplekse steekproefneming (a.g.v. stratifikasie) die variansie van 'n beramer kleiner word ten opsigte van die variansie van die beramer in die geval van eenvoudige ewekansige steekproefneming. Omgekeerd beteken 'n deff-waarde groter as een dat die komplekse steekproefnemingsprosedure (a.g.v. trosvorming) die variansie van 'n beramer, teenoor die variansie van die beramer in die geval van eenvoudige ewekansige steekproefneming, vergroot het.

In die praktyk neig naburige elemente om meer homogeen te wees as elemente wat verder van mekaar af lê, wat tot die gevolg het dat elemente wat in dieselfde tros lê geneig sal wees om 'n positiewe korrelasie in 'n veranderlike mee te bring, die effektiwiteit van die steekproefontwerp word dus verlaag en die variansie word groter. Dit is gevolglik van fundamentele belang om te bepaal met hoeveel die effektiwiteit van die steekproefontwerp verminder, teenoor die vermindering in koste van die steekproefopname, indien van trosvorming in 'n steekproefontwerp gebruik gemaak word.

Die ontwerpeffek geassosieer met 'n veranderlike word vir meeste subgroepe kleiner namate die grootte (M) van die PSE's afneem. Kyk paragraaf 2.3.2. Dit is dus nie wenslik om vir byvoorbeeld dieselfde veranderlike die waarde van die ontwerpeffek vir die hele steekproef regstreeks te probeer oordra na dié vir 'n subgroep nie, aangesien eersgenoemde waarde die ontwerpeffek vir die subgroep sal oorberaam.

Die ontwerpeffek kan basies as 'n funksie van twee komponente gesien word naamlik:

a) Die graad van homogeniteit binne die PSE's in 'n steekproef, wat hoofsaaklik bepaal word deur die waarde van die intratros-korrelasiekoëffisiënt (roh).
b) Die groottes van die PSE's in die steekproef (aangedui deur die simbool M)

Vergelyking (2.3.1) vir die berekening van die variansie van 'n beramer in die geval van trossteekproefneming word vervolgens afgelei (kyk STA452-8 Steekproefnemingstegnieke, UNISA) om die roh-waarde se definisie en invloed in 'n steekproef duideliker uit te beeld.

Beskou 'n ewekansige steekproef van n trosse uit 'n populasie van N trosse waarby elke tros M elemente bevat. Die beramer van die populasie totaal Y word gegee deur

$$\hat{Y} = (N/n)\sum_{i=1}^{n} y_i = (N/n)\sum_{ij}^{nM}\sum y_{ij} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(2.3.4)$$

met $\sum_{i=1}^{n} y_i$ die sommasie geneem oor al die PSE's in die steekproef en $\sum_{j=1}^{M} y_{ij}$ die sommasie geneem oor die USE's in die i-de PSE.

Die variansie van $\hat{Y}$ kan direk geskryf word as

$$V(\hat{Y}) = N^2(1-n/N)S_y{}^2/n$$
$$= (N^2/n)(1-n/N)[\sum_{i=1}^{N} (Y_i-\bar{Y})^2/(N-1)] \quad \ldots\ldots\ldots\ldots\ldots\ldots(2.3.5)$$

$$\mathrm{var}(y_{ij}) = \sum_{ij}\sum(y_{ij}-\bar{Y}_e)^2/NM$$

$$\rho = \mathrm{kov}(y_{ij},y_{ik})/\mathrm{var}(y_{ij})$$

$$=[\sum_{ijk}\sum\sum(y_{ij}-\bar{Y}_e)(y_{ik}-\bar{Y}_e)/NM(M-1)]/[NM/\sum_{ij}\sum (y_{ij}-\bar{Y}_e)^2]$$

$$=[\sum_{ijk}\sum\sum (y_{ij}-\bar{Y}_e)(y_{ik}-\bar{Y}_e)]/[(M-1)\sum_{ij}\sum (y_{ij}-\bar{Y}_e)^2]\ldots(2.3.6)$$

Maar

$$(Y_i - \bar{Y})^2 = \sum_{i=1}^{N} (\sum_{j=1}^{M} Y_{ij} - M(\bar{Y}/M))^2$$

Laat $\bar{Y}_e = \bar{Y}/M$

Dus $$(Y_i-\bar{Y})^2 = \sum_i[\sum_j(y_{ij} - \bar{Y}_e)]^2$$
$$= \sum_{ij}\sum(y_{ij}-\bar{Y}_e)^2 + 2 \sum_{ij<k} \sum(y_{ij}-\bar{Y}_e)(y_{ik}-\bar{Y}_e)$$
$$=(NM-1)S^2 + (M-1)(NM-1)\rho S_y{}^2 \quad \ldots\ldots\ldots\ldots\text{uit } (2.3.6)$$

$$=(NM-1)S^2\{1 + (M-1)\rho\}$$

Uit (2.3.5) volg

$$\sigma_{ts}^2 \cong V(\hat{Y}) = (N^2/n)(1-n/N)((NM-1)/(N-1))\, S_y^2[\, 1+ (M-1)\rho]$$

$$\cong (N^2/n)(1-n/N)MS_y^2[1 + (M-1)\rho] \quad \text{............(2.3.7)}$$

waarby $\rho$ die intratroskorrelasiekoëffisiënt is, en

$$S_y^2 = \sum_{ij}\sum (y_{ij}-\bar{Y}_e)^2/(NM - 1)$$

Let op bogenoemde benadering geld vir N groot.

Verder geld dat

$$\sum\sum_{ijk\neq j}\sum(y_{ij}-\bar{Y}_e)(y_{ik}-\bar{Y}_e) = 2 \sum\sum_{ijk}\sum(y_{ij}-\bar{Y}_e)(y_{ik}-\bar{Y}_e)$$

$$\Rightarrow 2\sum\sum_{ijk}\sum(y_{ij}-\bar{Y}_e)(y_{ik}-\bar{Y}_e)$$

$$=(M - 1)\rho\sum_{ij}\sum(y_{ij}-\bar{Y}_e)^2$$

$$=(M - 1)\rho(NM - 1)\, S_y^2$$

Indien 'n eenvoudige ewekansige steekproef van nM elemente direk uit die populasie getrek word volg dat

$$\sigma_{es}^2 = V(\hat{Y}) = \{(NM)^2/(nM)\}(1 - nM/NM)S_y^2$$

$$= (N^2/n)(1- n/N)MS_y^2 \quad \text{..........................(2.3.8)}$$

Indien (2.3.7) in terme van (2.3.8) geskryf word volg dat

$$\sigma_{ts}^2 \cong [1 + (M - 1)\rho]\sigma_{es}^2 \text{..........................(2.3.9)}$$

Hieruit kan die vergelyking vir deff direk afgelei word, nl.

$$deff = \sigma_{ts}^2/\sigma_{es}^2 \cong 1 + (M-1)\rho \text{........................} (2.3.10a)$$

waar $\rho$ die intratroskorrelasiekoëffisiënt aandui. Hierdie intratroskorrelasiekoëffisiënt word aangedui deur dieselfde simbool as die gewone korrelasiekoëffsient (rho). Let egter op die verskil in spelling. Die gewone korrelasiekoëffisiënt word as rho gespel, terwyl die intratroskorrelasiekoëffisiënt as roh gespel word, afkomstig van die eerste letters van die woorde 'rate of homogeneity'.

Indien M=1 in (2.3.10) dan is deff=1 en herlei trossteekproefneming na eenvoudige ewekansige steekproefneming. Uit (2.3.10) is dit duidelik dat, indien roh konstant bly, die deff-waarde groter word namate die trosgrootte toeneem, wat tot 'n vermindering in die doeltreffendheid van die steekproefneming lei. In trossteekproefneming met trosse van verskillende groottes bly die verband in (2.3.9) benaderd geld indien M vervang word met $\bar{M}$, die gemiddelde trosgrootte in die ontwerp, mits die trosgroottes nie te veel onderling varieer nie.

Vergelyking (2.3.10a) kan na aanleiding van bogenoemde bespreking ook as volg geformuleer word:

$$deff = \sigma_{ts}^2/\sigma_{es}^2 \cong 1 + (\bar{M}-1)\rho \text{........................} (2.3.10b)$$

### 2.3.4 Intratroskorrelasiekoëffisiënt (Roh)

Roh is 'n maatstaf van die variasie in die veranderlike onder bekouing binne die PSE's van die steekproefontwerp tot die variasie in die veranderlike oor die hele steekproef. 'n Hoë roh-waarde dui aan dat die waardes van die veranderlike binne 'n tros (d.w.s. in die PSE's) meer homogeen is, dit is sterker onderling gekorreleerd is, as die waardes van die veranderlike in die hele steekproef.

In Kish (1965) word eksplisiet verwys na die feit dat roh minder afhanklik is van die besondere steekproefontwerp of subgroepgroottes as wat die deff-waardes is.

Indien die trosgroottes vir verskillende subgroepe grootliks varieer en van oordraagbaarheid vir standaardfoute tussen subgroepe gebruik gemaak wil word, dan is roh die geskikste maatstaf om te gebruik. Dit impliseer onder andere dat indien die roh-waarde van byvoorbeeld die totale steekproefgemiddelde vir 'n veranderlike oorgedra wil word na die roh-waarde vir die gemiddelde van dieselfde veranderlike vir byvoorbeeld 'n sekere subgroep, dan sal die nuwe standaardfoutwaarde wat uit die nuwe roh-waarde verkry word, 'n stabieler waarde wees as wanneer die deff-waarde of die standaardfout as maatstaf van oordraagbaarheid gebruik word.

Uit (2.3.10b) kan ook vir roh opgelos word, naamlik:

$$\text{roh} = (\text{deff} - 1)/(\bar{M} - 1)$$
$$= (\text{deft}^2 - 1)/(\bar{M} - 1) \quad \ldots\ldots (2.3.10)$$

Hier word M vervang met $\bar{M}$, wat die gemiddelde PSE grootte aandui. Die gebruik van bogenoemde verband tussen deft en roh word vervolgens toegelig:

Voorbeeld:

Beskou 'n tweestadiumtrossteekproef van grootte n = 2500, met 49 getrekte PSE's en waarby elke PSE 'n gemiddelde van $\bar{M}$ = 51 USE's bevat, Veronderstel dat vir 'n spesifieke veranderlike die waarde van deff = deft$^2$ = 2 is. Dit wil sê die variansie van 'n beramer onder trossteekproefneming is presies twee maal so groot as die variansie van dieselfde beramer vir 'n eenvoudige ewekansige steekproef van dieselfde grootte. Dit impliseer dat 'n eenvoudige ewekansige steekproef van grootte n'= n/deff = 1250, dieselfde steekproefpresisie sou gee. (Die hele steekproef word beskou en nie subgroepe nie.)

Die waarde van roh is dan:

$$\begin{aligned} roh &= (deft^2 - 1)/(\bar{M} - 1) \\ &= (2 - 1) / (51 - 1) \\ &= 1/50 \\ &= 0{,}02 \end{aligned}$$

Beskou weer dieselfde aantal PSE's met die gemiddelde getal eenhede per tros nou $\bar{M}$ = 26, soos in die geval van 'n moontlike subgroep, en neem die roh-waarde as konstant. Dan, indien roh as oordraagbaar tussen die steekproef en 'n substeekproef beskou word, sal die ontwerpeffek gelyk wees aan

$$\begin{aligned} deff &= 1 + 0{,}02(26 - 1) \\ &= 1{,}5. \end{aligned}$$

In die geval waar die gemiddele trosgrootte $\bar{M}$ in twee steekproefontwerpe grootliks verskil, is roh 'n beter maatstaf van oordraagbaarheid as deff aangesien die effek van $\bar{M}$ nie die roh-waarde beïvloed nie. Om roh as maatstaf van oordraagbaarheid te gebruik is dit noodsaaklik dat soortgelyke veranderlikes en steekproefontwerpe geneem word en dieselfde beramer beskou word. Wanneer roh as maatstaf van oordraagbaarheid gebruik word, word die roh-waarde nie deur die steekproefgroottes beïnvloed nie.

Dit is nie wenslik om roh as maatstaf van oordraagbaarheid toe te pas indien ongelyksoortige veranderlikes beskou word nie, aangesien in dergelike gevalle waardes van roh waargeneem is wat met faktore van tot 'n 100 verskil het (Verma 1982).

### 2.3.5 Die gebruik van roh en deff in imputasie

Oordraging geskied deur imputasie, wat of regstreeks tussen standaardfoutwaardes of tussen deff-waardes of tussen roh-waardes kan geskied, waarby soms van 'n imputasiefaktor gebruik gemaak word. Die oordra van 'n bekende standaardfout ($stf_0$) na 'n onbekende standaardfout ($stf_1$), of die verkryging van 'n onbekende roh-waarde (of deff-waarde) vanaf 'n bekende roh-waarde (of deff-waarde) staan as imputasie bekend.

Beskou die volgende skematiese illustrasie. Om van A na B te gaan kan roete I gevolg word vir gelyksoortige veranderlikes in dieselfde steekproefopname, óf roete II kan gevolg word in die geval van gesegregeerde subgroepe, óf roete III kan gevolg word in die geval van kruisgroepe. Die oorgang van $stf_0$ na $stf_1$, of van $deff_0$ na $deff_1$, of van $roh_0$ na $roh_1$, staan bekend as imputasie.

```
      A                              B
Bekende stf0 ~ → ~ ~ ~ ~ → ~ ~  Geimputeerde stf1
    ↓  *                          *  ↑
    ↓  ↓                          ↑  ↑
    ↓  *                          *  ↑
  deff0 * →  *  * →  *  *  * →  *   * deff1
    ↓                                ↑
    ↓                                ↑
    ↓                                ↑
   roh0→ → → → → → → → → → → → → → → →roh1
```

ROETE I = ~
ROETE II = *
ROETE III = →

Indien roh as maatstaf vir oordraagbaarheid gebruik word dan sal die geïmputeerde waarde van roh verkry word deur die berekende waarde $roh_0$ en 'n imputasiefaktor $\lambda_1$ te gebruik, naamlik:

$$roh_1 = \lambda_1 \, roh_0 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2.3.11)$$

Daarna kan die onbekende deff-waarde beraam word deur:

$$deff_1 = deft^2{}_1$$

$$= [1 + roh_1(\bar{M}_1 - 1)]$$

$$= [1 + \lambda_1 roh_0(\bar{M}_1 - 1)]$$

Die gemiddelde grootte $\bar{M}_1$ van die PSE's in die steekproef word bereken deur $\bar{M}_1 = n_1/a$, waarby $n_1$ die steekproefgrootte is en a die aantal PSE's wat getrek is. Die waarde van $\lambda_1$ word gebaseer op empiriese studies. Die WFS (World Fertility Survey) het heelwat ondersoeke geloots in verskeie lande na die fertileit van vrouens en het 'n waarde van $\lambda_1$ = 1,2 voorgestel (Verma 1982).

Indien deff as maatstaf vir oordraagbaarheid gebruik word, kan die geïmputeerde waarde

$$deff_1 = deft_1^2 = 1 + k(deft_0^2 - 1) \qquad (2.3.12)$$

verkry word deur middel van die berekende waarde $deff_0$ en deur 'n imputasiefaktor k waarby k 'n konstante of funksie van subgroepgrootte en -tipe is. Weer eens word die waarde van k bepaal deur middel van empiriese studies wat op verskillende soortgelyke veranderlikes uitgevoer word.

Die waarde van die onbekende standaardfout $stf_1$ word ten slotte verkry uit die formule

$$(stf_1)^2 = (deff_1)(s_{es})^2$$

waarby $s_{es}^2$ 'n beraming van $_{es}^2$ is, die variansie van 'n beramer (bv. die steekproefgemiddelde) in die geval van eenvoudige ewekansige steekproefneming.

### 2.3.6 Opsomming

Opsommend kan gesê word dat die standaardfout van byvoorbeeld 'n subgroepgemiddelde vir 'n veranderlike alleen oordraagbaar is na dié van 'n soortgelyke veranderlike indien die subgroepgroottes ewegroot is en die steekproefontwerpe soortgelyk is.

Ontwerpeffekte as maatstaf van oordraagbaarheid is breër toepasbaar in die praktyk as standaardfoute. Die graad van homogeniteit binne die PSE's (gemeet deur roh) van 'n steekproef sowel as die gemiddelde trosgrootte $\bar{M}$ bepaal die waarde van die ontwerpeffek. Deff is die beste maatstaf om te gebruik indien die subgroepgroottes van nagenoeg dieselfde omvang is en wanneer met soortgelyke veranderlikes gewerk word. Deff is dus die beste maatstaf van oordraagbaarheid indien met gesegregeerde subgroepe gewerk word aangesien die PSE-groottes en steekproefontwerp, wat 'n groot invloed op die deff-waarde het, in die algemeen nie te veel vir sulke subgroepe varieer nie.

Indien die trosgroottes grootliks varieer is roh 'n beter maatstaf van oordraagbaarheid as deff. Roh is die beste maatstaf om te gebruik vir die oordra van 'n standaardfout van byvoorbeeld die gemiddelde vir die hele steekproef na die standaardfout van 'n subgroepgemiddelde vir dieselfde of soortgelyke (kyk paragraaf 1.4) veranderlikes. Roh skyn besonder geskik te wees vir die imputasie van die variansie van 'n beramer van 'n populasieparameter in die geval van kruisgroepe.

## 2.4 STANDAARDFOUTE VAN SUBGROEPGEMIDDELDES

### 2.4.1 Inleiding.

Kish (1965) som kortliks die verband tussen die standaardfoute van 'n beramer vir subgroepe ($se_s$) en die vir die hele steekproef ($se_t$) in die volgende drie punte op:

a) As kruisgroepe taamlik gelykmatig versprei is oor die verskillende PSE's dan sal die gemiddelde grootte van 'n PSE vir die kruisgroep heelwat kleiner wees as dié vir die hele steekproef. Hierdie vermindering in grootte is ongeveer proporsioneel tot die subgroepgrootte ten opsigte van die totale steekproefgrootte.

b) Indien die subgroepe beperk is tot slegs 'n subversameling van die PSE's van die steekproef dan is die variansie baie minder stabiel.

c) Vir gesegregeerde subgroepe is die subgroep nie gelykmatig versprei oor die verskillende PSE's nie, en hierdie ongelykmatige verspreiding lei tot 'n hoër koëffisiënt van relatiewe variasie vir die beramer in die subgroep as vir die hele steekproef. As gevolg van die hoër koëffisiënt van relatiewe variasie word die foutevariasie sowel as die sydigheid in die verhoudingsberamer groter.

Opsommend kan gesê word dat die variasie in die standaardfout van 'n beramer vir 'n subgroep direk verband hou met hoe die subgroepe oor die PSE's van die steekproef versprei lê, en dit lei tot die onderskeid tussen kruisgroepe en gesegregeerde groepe (kyk Hoofstuk 1).

### 2.4.2 <u>Standaardfoute vir kruisgroepe</u>

As die kruisgroepe kleiner word, word die PSE's (trosse) ook kleiner en kan verwag word dat die ontwerpeffek vir kruisbgroepe ($deff_s$) kleiner sal wees as die ontwerpeffek vir die hele steekproef ($deff_t$). As die kruisgroepgrootte in 'n selfwegende trossteekproef baie klein word dan sal die steekproefontwerp neig na 'n eenvoudige ewekansige steekproef, en die deff-waarde kom dan al nader aan een.

Om die verband tussen die standaardfout van 'n beramer, bereken vir 'n subgroep, en die standaardfout van dieselfde beramer, vir die hele steekproef, te verkry, kan die volgende formule gebruik word:

$$\frac{deff_s - 1}{deff_t - 1} = \frac{roh\ (\bar{b}_s - 1)}{roh\ (\bar{b}_t - 1)} \qquad (2.4.1)$$

waar $\bar{b}_s$ en $\bar{b}_t$ die gemiddelde trosgroottes vir die subgroep en die hele steekproef onderskeidelik aandui. Veronderstel dat vir kruisgroepe is $roh_t = roh_s$, dit wil sê dat die intratroskorrelasiekoëffisiënt van die hele steekproef gelyk is aan die intratroskorrelasiekoëffisiënt van die subgroep (kruisgroep). Dan sal die regterkant van (2.4.1) gelyk wees aan $(\bar{b}_s-1)/(\bar{b}_t-1)$. Hierdie waarde kan benader word deur $M_s$ wat die grootte van die subgroep in verhouding tot die hele steekproef aandui. Laasgenoemde is slegs moontlik indien die gemiddelde trosgrootte $\bar{b}_s$ heelwat groter as een is. Vergelyking (2.4.1) herlei in hierdie geval as volg:

$$\frac{deft_s^2 - 1}{deft_t^2 - 1} = \frac{\bar{b}_s - 1}{\bar{b}_t - 1}$$

$$\cong M_s$$

Dus

$$deft_s^2 - 1 = M_s(deft_t^2 - 1)$$

$$deft_s^2 = 1 + M_s(deft_t^2 - 1) \qquad (2.4.2).$$

Uit bogenoemde model is dit duidelik dat die waarde van $deft_s^2$, indien dit groter as een word, proporsioneel is tot die waarde van $M_s$ (die grootte van die subgroep in verhouding tot die grootte van die hele steekproef). Dit impliseer dat vir baie klein subgroepe sal die waarde van $M_s$ baie klein word en $deft_s^2$ nader aan een beweeg, dit wil sê die steekproef sal in die rigting van 'n eenvoudige ewekansige steekproef beweeg.

Vergelyking (2.4.2) geld vir selfwegende steekproewe. Nie-selfwegende steekproewe is 'n algemene verskynsel in die praktyk en alhoewel die effek van stratifikasie en trosvorming afneem vir klein kruisgroepe, neem die gewigseffek nie af nie. Daarom is dit nodig om (2.4.2) vir nie-selfwegende steekproewe aan te pas.

Om te kompenseer vir die effek van ongelyke gewigte (ongekorreleerd met die populasievariansie) moet die variansie van alle beramers met die faktor

$$L=[(\sum_{h=1}^{H} n_h w_h^2)(\sum_{h=1}^{H} n_h)]/(\sum_{h=1}^{H} n_h w_h)^2 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2.4.3)$$

vermenigvuldig word (Kish 1965), waarby $n_h$ die aantal PSE's in stratum h met gewig $w_h$ aandui. Daar is gevind dat die deff-waarde vir baie klein kruisgroepe, na die waarde $L^{\frac{1}{2}}$ neig. Indien die effek van weging uitgesluit word kan 'n nuwe deff-waarde (deft') benaderd as volg verkry word:

$$\begin{aligned} \text{deft} &= se/sr \\ &= L^{\frac{1}{2}}\ \text{deft}' \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2.4.4) \end{aligned}$$

waarby se die standaardfout van 'n beramer van 'n populasieparameter in die geval van 'n komplekse steekproef, en sr die standaardfout van dieselfde beramer in die geval van 'n eenvoudige ewekansige steekproef, voorstel.

Vir 'n selfwegende eenvoudige ewekansige steekproef word die variansie van 'n gemiddelde $\bar{y}$ vir 'n steekproef van grootte n beraam deur die vergelyking

$$sr^2(\bar{y}) = (1-f)/n[\sum_j (y_j - \bar{y})^2/(n-1)]$$
$$= [(1-f)/n]s^2.$$

Vir die totale steekproef en vir 'n subgroep kan die variansie van 'n beramer van 'n populasieparameter vir 'n eenvoudige ewekansige steekproef as volg geformuleer word:

Vir die hele steekproef:

$sr^2{}_t = [(1-f)/n]s^2{}_t$

Vir die subgroep:

$sr^2{}_s = [(1-f)/n]s^2{}_s$..................................(2.4.5).

Die verband tussen die standaardfoute van die beramers vir nie-selfwegende komplekse steekproewe vir onderskeidelik die hele steekproef ($se_t$) en vir 'n subgroep ($se_s$) kan dan as volg geskryf word (Verma (1982)):

Vir die hele steekproef:

$se_t{}^2 = (s_t{}^2/n)(L_t)(deft'_t{}^2)$

Vir die subgroep:

$se_s{}^2 = (s_s{}^2/n)(L_s)(deft'_s{}^2)$.....................(2.4.6)

waarby $s_t{}^2$, die variansie van die beramer van 'n populasieparameter vir die hele steekproef, en $s_s{}^2$, die variansie van dieselfde beramer vir 'n subgroep, in (2.4.5) gegee word. Om die waardes van $se_s$ en $se_t$ in verband met mekaar te bring moet die subgroepwaardes en totale steekproefwaardes vir elkeen van die drie hoeveelhede aan die regterkant van (2.4.6) in verband met mekaar gebring word.

Die volgende verwantskappe is kenmerkend vir kruisgroepe:

a) Die verband tussen die deft-waardes ooreenkomstig (2.4.2) is as volg:

$$deft'^2_s = 1 + (deft'^2_t - 1)M_s \qquad (2.4.7)$$

met deft' gedefinieer in (2.4.4).

b) Die aanname $L_s = L_t$ kan geregverdig word aangesien die relatiewe verteenwoordiging van die domeins in die subgroep in die algemeen dieselfde is as in die hele steekproef (kyk Verma 1982).

c) Die waardes van die standaardafwykings $s_s$ (vir subgroepe) en $s_t$ (vir die hele steekproef) kan aansienlik verskil, aangesien die subgroepe van belang kruisgeklassifiseer is volgens faktore wat gekorreleerd is met die veranderlikes wat beraam word.

### 2.4.3 <u>Standaardfoute vir gesegregeerde groepe</u>

Vir gesegregeerde subgroepe kan standaardfoute bereken word op dieselfde wyse as vir die hele steekproef. Die verband tussen die standaardfoute van die beramers van die populasieparameters in die geval van gesegregeerde subgroepe en die standaardfoute van dieselfde tipe beramers in die geval van die hele steekproef word baie algemeen ondersoek. Let op dat die resultate vir byvoorbeeld individuele gesegregeerde subgroepe minder stabiel is as gevolg van die kleiner aantal PSE's in 'n gesegregeerde subgroep as in die hele steekproef. By gesegregeerde subgroepe kan die gewigsfaktor $L_s$ vir die subgroep heelwat verskil van die gewigsfaktor $L_t$ vir die hele steekproef. Gewoonlik is $L_s < L_t$ aangesien die gewigte dikwels van stratum tot stratum verskil terwyl die eenhede binne in 'n stratum selfwegend is.

Indien die steekproefontwerp sodanig is dat die trosgroottes in verskillende gesegregeerde subgroepe baie dieselfde is kan verwag word dat die deff-waarde benaderd dieselfde waarde sal hê. In die praktyk kom dit meer algemeen voor dat die steekproefontwerpe van die verskillende gesegregeerde subgroepe van mekaar verskil. 'n Eenvoudige model wat die ontwerpeffek van die subgroep ($\text{deft}_s$) in terme van die ontwerpeffek van die hele steekproef ($\text{deft}_t$) uitdruk kan as volg geformuleer word (kyk Verma 1982):

$$\text{deft}_s^2 = 1 + c_s(\text{deft}_t^2 - 1) \ldots\ldots\ldots\ldots\ldots(2.4.11)$$

waarby $c_s$ 'n konstante is wat vir die subgroep empiries bepaal word deur

$$c_s = (\text{deft}_s^2 - 1)/(\text{deft}_t^2 - 1)$$

uit (2.4.11) te bereken vir reeds berekende deff-waardes van 'n groep soortgelyke veranderlikes.

### 2.4.4 Standaardfoute vir subgroepverskille

Die variansie van die verskil tussen twee subgroepgemiddeldes $\bar{y}_a$ en

$\bar{y}_b$ kan as volg geskryf word (Kish 1965):

$$var(\bar{y}_a - \bar{y}_b) = var(\bar{y}_a) + var(\bar{y}_b) - 2kov(\bar{y}_a,\bar{y}_b)\text{.........}(2.4.12)$$

Die kovariansieterm is gewoonlik positief behalwe wanneer die twee subgroepe (a en b) van totaal verskillende PSE's kom, in welke geval die kovariansie nie noodwendig positief hoef te wees nie. Die volgende ongelykheid behoort in die algemeen te geld:

$$[(s^2{}_a/n_a)L_a + (s_b{}^2/n_b)L_b]$$
$$< var(\bar{y}_a - \bar{y}_b)$$
$$< [var(\bar{y}_a) + var(\bar{y}_b)]\text{..........................}(2.4.13)$$

waar die linkerkantste uitdrukking die variansie van $(\bar{y}_a - \bar{y}_b)$ vir twee onafhanklike ewekansige steekproewe met ongelyke seleksiewaarskynlikhede voorstel, terwyl die regterkantste uitdrukking die variansie vir die trossteekproef voorstel indien die kovariansieterm positief of nul is. Met ander woorde die variansie van die verskil tussen twee gemiddeldes by trossteekproefneming het 'n ontwerpeffek ($deff_d$) met 'n positiewe intratroskorrelasiekoëffisiënt, dit is $deft_d > 1$.

Soortgelyk as in (2.4.6) geld vir verskille tussen gemiddeldes dat

$$var(\bar{y}_a - \bar{y}_b) = (s_a{}^2/n_a + s_b{}^2/n_b)L_d deft'_d{}^2\text{.........}(2.4.14)$$

Die effek van weging ($L_d$) is dus ook aanwesig in die geval van verskille tussen subgroepe.

Indien die kovariansieterm in (2.4.12) as funksie van $var(\bar{y}_a)$ +

$var(\bar{y}_b)$ beskou word, kan (2.4.12) soos volg geskryf word:

$$var(\bar{y}_a - \bar{y}_b) = \beta^2[var(\bar{y}_a) + var(\bar{y}_b)] \ldots\ldots\ldots\ldots\ldots (2.4.15)$$

waarby $0<\beta<1$. Vir $\beta = 1$ is daar geen kovariansieterm nie. Gewoonlik is $\beta$ tussen 0,9 en 1,0. ( $^2$:0,8 - 1,0). Daar is gevind dat die $\beta$ waarde neig om kleiner te word vir goed verspreide demografiese subgroepe en ook vir groepe veranderlikes met 'n groot deft-waarde (Verma 1982).

Veronderstel $s_a^2 = s_b^2$ (= $s^2$) en die gewigsfaktor (L) is dieselfde vir verskillende subgroepe sowel as vir verskille tussen subgroepe. Indien van (2.4.6) en (2.4.7) gebruik gemaak word met $M_s = n_a/n$ of $M_s = n_b/n$, en $n_d$ die helfte van die harmoniese gemiddelde van die subgroepgroottes $n_a$ en $n_b$ is, dit wil sê as

$$n_d = (n_a n_b)/(n_a + n_b) \ldots\ldots\ldots\ldots\ldots\ldots (2.4.16)$$

is, dan kan (2.4.13) se linker- en regterkante as volg herlei word:

Linkerkant:

$[(s^2{}_a/n_a)L_a + (s_b{}^2/n_b)L_b]$

$=[(s^2/nM_s)L + (s^2/nM_s)L]$

$=L_d[s^2/n_a + s^2/n_b]$

$=L_d[s^2 n_b/n_a n_b + s^2 n_a/n_a n_b]$

$=L_d s^2[(n_b + n_a)/n_a n_b]$

$=L_d s^2/n_d$ uit (2.4.16)

Regterkant:

$[var(\bar{y}_a) + var(\bar{y}_b)]$

$=(s_a{}^2/n_a)L_a\text{deft}'_a{}^2 + (s_b{}^2/n_b)L_b\text{deft}'_b{}^2$

$=(s^2/n_a)L_d[1 + (\text{deft}'_t{}^2 - 1)M_a] + (s^2/n_b)L_d[1 + (\text{deft}'_t{}^2 - 1)M_b]$

$=(s^2/n_a)L_d + (s^2/n_a)L_d(n_a/n)[\text{deft}'_t{}^2 - 1] +$

$(s^2/n_b)L_d + (s^2/n_b)L_d(n_b/n)[\text{deft}'_t{}^2 - 1]$

$=(s^2/n_d)L_d + 2(s^2/n)L_d[\text{deft}'_t{}^2-1]$

$=(s^2/n_d)L_d[1 + 2(n_d/n)(\text{deft}'_t{}^2-1)].$

Uit bogenoemde aannames en die nodige afleidings volg dat (2.4.13) as volg uitgedruk kan word

$$(s^2/n_d)L_d < \text{var}(\bar{y}_a - \bar{y}_b) < (s^2/n_d)L_d[1 + 2(n_d/n)(\text{deft}_t{}^2 - 1)] \quad ..(2.4.17)$$

Selfs vir taamlike groot subgroepe is die afstand tussen die onderste en boonste grense in (2.4.17) in die algemeen klein.

<u>Voorbeeld</u>

As $\text{deft}'_t{}^2 = 2$, $L_d = 1$ en $n_a = n_b = 0{,}2n$, dan is

$$n_d = \frac{(0{,}2n)^2}{0{,}2n + 0{,}2n} = \frac{0{,}04n^2}{0{,}4n} = 0{,}1n$$

en (2.4.17) word

$$10s^2/n < \text{var}(\bar{y}_a - \bar{y}_b) < 12s^2/n.$$

As aangeneem word dat $\text{var}(\bar{y}_a - \bar{y}_b)$ presies halfpad tussen die grense van (2.4.17) lê, dan volg

$$\mathrm{var}(\bar{y}_a - \bar{y}_b) \cong (s^2/n_d)L_d[1 + (n_d/n)(\mathrm{deft}'_t{}^2 - 1)] \quad \ldots\ldots\ldots(2.4.18)$$

Hierdie vergelyking is identies aan vergelykings (2.4.6) en (2.4.7) waarby $n_d$ die effektiewe subgroepgrootte genoem word. Dit impliseer dat die variansie van die verskil tussen twee subgroepgemiddeldes ongeveer gelyk is aan die variansie vir 'n subgroep waarvan die grootte gelyk is aan die helfte van die harmoniese gemiddelde van die twee subgroepgroottes

- kyk (2.4.16).

## 2.5 STRATEGIEË VIR DIE BEREKENING VAN STANDAARDFOUTE

Om al die nodige inligting aangaande die presisie van die beramings vanuit 'n steekproefopname te verskaf kan die volgende riglyne gevolg word:

a) Plaas soortgelyke veranderlikes in groepe bymekaar.

b) Bereken die standaardfoute vir die beramers van verskeie veranderlike waardes in die verskillende groepe.

c) Bepaal watter kruisgroepe, gesegregeerde groepe en verskille tussen subgroepe van belang is om te ontleed.

d) Indien verskillende kruisgroepe beskou word is dit wenslik om van die roh-waardes gebruik te maak vir die toepassing van oordraagbaarheid.

e) Wanneer verskillende gesegregeerde groepe van belang beskou word is die deft-waardes die betere maatstaf van oordraagbaarheid om te gebruik.

f) Bereken die standaardfoute, deft-waardes en roh-waardes nie net vir die steekproef in sy geheel en subgroepe nie, maar ook vir verskille tussen subgroepgemiddeldes.

## 2.6 ONTWERPEFFEKTE VIR KOMPLEKSE STATISTIEKE

Kish en Frankel (1974) het tot die volgende gevolgtrekkings gekom aangaande die verband tussen die ontwerpeffek vir 'n komplekse statistiek ( $\gamma$ ) (bv. 'n korrelasie- of regressiekoëffisiënt) en die ontwerpeffek vir die gemiddelde ($\bar{y}$) van 'n gegewe veranderlike in 'n steekproef of in 'n subgroep daarvan:

a) $deft(\gamma)>1$. In die algemeen is die ontwerpeffekte vir komplekse statistieke groter as een. Die standaardafwykings van komplekse statistieke wat op eenvoudige ewekansige steekproefneming gebaseer is, is geneig om die standaardafwyking van dié komplekse statistieke te onderberaam.

b) $deft(\gamma) < deft(\bar{y})$. Die ontwerpeffek vir 'n komplekse statistiek is geneig om kleiner te wees as die ontwerpeffek vir 'n gemiddelde van 'n gegewe veranderlike in 'n steekproef of in 'n subgroep daarvan.

c) Daar is 'n direkte verband tussen $deft(\bar{y})$ en $deft(\gamma)$ in die opsig dat vir veranderlikes met 'n hoë $deft(\bar{y})$ waarde, neig die waarde van $deft(\gamma)$ om ook hoog te wees.

d) $Deft(\gamma)$ neig om 'n ooreenkoms te toon met $deft(\bar{y}_a - \bar{y}_b)$, die ontwerpeffek vir verskille tussen gemiddeldes.

Die volgende verband word deur Kish en Frankel (1974) gegee:

$$deft^2(\Upsilon)=1 + k(deft_t^2(\bar{y})-1)$$

met $deft(\bar{y})>1$ en k $(0<k<1)$ 'n spesifieke waarde afhangende van die tersake beramer.

## 2.7 EMPIRIESE ONDERSOEKE

### 2.7.1 Inleiding

Die teorie wat in hierdie hoofstuk bespreek is, word vervolgens empiries ondersoek. Hierdie praktiese toepassing is beperk tot 'n steekproef-opname van huishoudings en geskied deur gebruikmaking van 'n RGN steekproef oor lewenskwaliteit. 'n Program met die naam CLUSTERS wat deur die 'World fertility survey', by name Verma, V. en Pearce, M. (1978) ontwikkel is, word gebruik om verskeie statistieke vir verskillende veranderlike-waardes te bereken, onder andere die deft-waarde, roh-waarde en die standaardfoute vir totale, gemiddeldes of proposies. Die doel van hierdie praktiese oefening is om te bepaal watter imputasiewaardes vir verskillende soorte van veranderlikes in die Suid-Afrikaanse situasie gebruik kan word.

### 2.7.2 Die rekenaarprogram CLUSTERS

Die rekenaarprogram CLUSTERS is 'n FORTRAN program wat standaardfoute, deft-waardes en roh-waardes vir beramers van populasieparameters vir meerstadiumtrossteekproewe bereken.

Standaardfoute, ontwerpeffekte en intratroskorrelasiekoëffisiënte vir beramers kan bereken word vir die hele populasie sowel as vir subgroepe van die populasie.

Die program CLUSTERS kan 'n veranderlike regstreeks vanaf 'n individuele rekord gebruik of 'n nuwe veranderlike skep vanuit veranderlikes wat in die rekord voorkom.

Die variasie in die waarde van roh ('rate of homogeneity') tussen verskillende (ongelyksoortige) veranderlikes is gewoonlik aansienlik groter as die variasie in die waarde van roh vir 'n spesifieke veranderlike oor verskillende subgroepe. Die waarde van roh is gewoonlik ook groter tussen subgroepe gebaseer op verskillende kenmerke as die variasie in die waarde van roh tussen subgroepe wat gevorm is deur slegs verskillende kategorieë van dieselfde kenmerke te beskou.

<u>'n Kort oorsig oor die invoer benodig vir CLUSTERS</u>

Die invoerdataleêr moet bestaan uit FORTRAN-leesbare vaste-lengte rekords, een rekord vir elke USE. Hierdie dataleêr word beëindig deur 'n 'dummy' rekord van dieselfde lengte as al die ander rekords: dit dui die einde van die invoerleêr aan.

a) Berekenings word uitgevoer vir elke gespesifiseerde veranderlike oor die hele steekproef sowel as oor alle gespesifiseerde subgroepe van die steekproef. Wanneer domeins gespesifiseer word, word die standaardfoute en ander steekproefgroothede vir elke domein afsonderlik bereken, asook vir alle gespesifiseerde subgroepe in elke domein.

b) Subgroepe kan ook in pare gespesifiseer word in welke geval verskille tussen die waardes van 'n veranderlike in die gepaarde subgroepe outomaties bereken sal word.

c) Veranderlikes kan direk vanaf 'n gespesifiseerde veld op die invoerrekord gelees word of nuwe veranderlikes kan geskep word vanuit verskillende veranderlikes op die invoerrekord deur middel van herkoderingskaarte. Waardes van veranderlikes wat geïgnoreer en nie in berekening gebring moet word nie kan ook gespesifiseer word.

d) Die struktuur of vorm van die steekproef kan ook gespesifiseer word.

e) USE's in die steekproef kan geweeg word om te kompenseer vir ongelyke seleksiewaarskynlikhede. Vir berekenings is slegs die relatiewe waarde van die gewigte belangrik en die gespesifiseerde gewigte kan af- of opgeskaleer word volgens keuse. Om te verhoed dat beraamde groothede te groot word, is dit wenslik en veiliger om die gewigte so te skaleer dat die gemiddelde waarde van die gewigte na aan 1/(gemiddelde steekproefgrootte per PSE) lê.

<u>Berekening van gewigte</u>.

Indien 'n komplekse meerstadiumsteekproef nie proporsioneel getrek is nie, is dit noodsaaklik om van gewigte gebruik te maak om te kompenseer vir die feit dat waarnemings nie met gelyke waarskynlikhede getrek is nie. Hier volg 'n kort illustrerende voorbeeld (Stoker 1983) van hoe te werk gegaan sal word om gewigte in 'n meerstadiumsteekproef te bereken.

Illustrerende voorbeeld.

Veronderstel die steeproef bestaan uit 300 huishoudings en dat die struktuur van die steekproef as volg is:

Daar is ses strata met 100 PSE's in totaal oor al die strata. Net 25 PSE's word getrek en uit elke getrekte PSE word daar 'n aantal USE's getrek. So byvoorbeeld kan 'n stedelike woongebied in strata verdeel word, uit elk kan 'n aantal straatblokke (PSE's) getrek word en uit elke getrekte PSE kan 'n aantal woonhuise (USE's) getrek word.

TABEL A

Skematies aangedui word die volgende veronderstel:

| stratum nr. | 1 | 2 | 3 | 4 | 5 | 6 | Totaal |
|---|---|---|---|---|---|---|---|
| # PSE'S in populsie | 15 | 18 | 10 | 25 | 20 | 12 | 100 |
| # PSE'S in steeproef | 4 | 5 | 3 | 6 | 4 | 3 | 25 |
| # woonhuise in getrekte PSE'S | 49 | 65 | 40 | 70 | 55 | 75 | |
| | 41 | 52 | 55 | 30 | 40 | 60 | |
| | 56 | 56 | 45 | 50 | 45 | 35 | |
| | 48 | 50 | | 65 | 45 | 35 | |
| | | 40 | | 35 | | | |
| | | | | 45 | | | |
| T O T A A L | 194 | 212 | 140 | 295 | 185 | 170 | 1 196 |
| # woonhuise getrek in getrekte PSE's | 12 | 15 | 12 | 15 | 13 | 20 | |
| | 10 | 12 | 15 | 8 | 9 | 14 | |
| | 14 | 13 | 10 | 12 | 10 | 11 | |
| | 12 | 10 | | 15 | 12 | | |
| | | 8 | | 8 | | | |
| | | | | 10 | | | |
| T O T A A L | 48 | 58 | 37 | 68 | 44 | 45 | 300 |

Die gewig geassosieer met die trek van woonhuis een, in PSE een, in stratum een, sal as volg wees:

1/[(4)(205)/2520 x 12/49 ] of ook 1/[(4)(49)/656 x 12/49] = 656/48

waarby 205 die bevolking in die getrekte PSE, 2520 die bevolking in die totale stratum en 656 die aantal wooneenhede in die stratum is (kyk tabel A).

Uitvoerbesonderhede.

Die uitvoer van die program CLUSTERS bestaan uit drie afdelings:

a) Eerstens verskyn 'n uitdruk van die invoerbeheerkaarte.
b) Tweedens verskyn 'n opsomming van besonderhede van die ontwerp van die steekproef.
c) Die derde deel van die uitvoer bevat die standaardfoute en ander relevante statistiese groothede. Hierdie resultate kom voor in blokke wat ooreenstem met die subgroepe of pare-subgroepe wat gespesifiseer is. So byvoorbeeld sal:

Blok 1: Resultate vir die hele steekproef bevat

Blok 2: Resultate vir die subgroep wat eerste gespesifiseer is

.
.
.
.

Blok n: Resultate vir die subgroep wat laaste gespesifiseer is bevat.

Elke blok bestaan uit afdelings wat ooreenstem met elke gespesifiseerde domein. As geen domeins gespesifiseer word nie sal slegs die totale domein voorkom. Dus binne elke blok lyk die uitvoer as volg:

Afdeling 1: Domein 1

Afdeling 2: Domein 2

.
.
.

Afdeling n: Domein 0 = Totaal.

Hierdie afdelings bevat op hulle beurt weer die resultate vir die subgroepe. Dit sal as volg wees:

Resultate vir subgroep 1

Resultate vir subgroep 2

Verskille tussen subgroep 1 en subgroep 2.

Elke veranderlike se resultate word in 'n horisontale lyn uitgedruk en bevat die volgende gegewens:

R - die gemiddelde of proporsie van die veranderlike

SE - die standaardfoute van R, met inagneming van die steekproefontwerp

N - die ongeweegde aantal gevalle wat in die berekenings gebruik is

WN - die geweegde aantal gevalle wat in die berekenings gebruik is

SER - die standaardfout van R, onder die aanname van eenvoudige ewekansige steekproefneming

SD - die standaardafwyking, d.w.s. SER vermeningvuldig met die vierkantswortel van N

DEFT- die ontwerpeffek

ROH - intratroskorrelasiekoëffisiënt = $(DEFT^2 - 1)/(B - 1)$ waar B die gemiddelde trosgrootte is

SE/R- relatiewe standaardfout

R-2SE - die onderste grens van die 95% vertrouensinterval

R+2SE - die boonste grens van die 95% vertrouensinterval

B - die ongeweegde gemiddelde trosgrootte

cv - die relatiewe koëffisiënt van variasie. Hierdie grootheid kom slegs voor vir die kenmerke van subgroepe.

### 2.7.3 Berekening van standaardfoute deur die program CLUSTERS.

Die metode van beraming van standaardfoute wat deur die program CLUSTERS gevolg word, word vervolgens bespreek (Stoker 1983).

Die program CLUSTERS is geskryf vir meerstadium komplekse steekproewe en dus is die enigste vereiste om standaardfoutberamings te kan doen dat daar minstens twee PSE's per stratum is. CLUSTERS het egter die beperking dat dit slegs beraamde standaardfoute van beramings van populasieparameters soos totale, gemiddeldes en proposies (verhoudings) kan bereken. Dit wil sê CLUSTERS is geensins ontwikkel om beramings van standaardfoute van populasieparameters soos korrelasiekoëffisiënte, regressiekoëffisiënte, parsiële korrelasiekoëffisiënte en meervoudige regressiekoëffisiënte te bereken nie.

Beskou die volgende verhoudingsberamer

$$r = \sum_{h}\sum_{i} y_{hi} / \sum_{h}\sum_{i} x_{hi} \quad \text{........................}(2.7.1)$$

waarby daar H strata is, $m_h$ PSE's in die h-de stratum getrek word en $y_{hi}$ en $x_{hi}$ onderskeidelik die berekende totaalwaardes van die veranderlikes (of kenmerke) y en x vir die i-de getrekte PSE in die h-de stratum, verkry uit die steekproefdata, is.

Die berekende totaalwaardes van $y_{hi}$ en $x_{hi}$ van y en x word verkry uit

$$y_{hi} = \sum_j w_{hij}y_{hij} \text{ en } x_{hi} = \sum_j w_{hij}x_{hij}$$

waarby $y_{hij}$ en $x_{hij}$ die waardes van die veranderlikes x en y onderskeidelik vir die j-de steekproefelement behorende tot die i-de getrekte PSE van die h-de stratum is en $w_{hij}$ die gewig geassosieer met hierdie steekproefelement is. Die sommasie van j gaan oor die $r_{hi}n_{hi}$ steekproefelemente in die i-de PSE van die h-de stratum waarvoor daar response is (dit is waarvoor daar waardes van y en x bekom is). Gevolglik dui $n_{hi}$ op die aantal getrekte steekproefelemente en $r_{hi}$ dui die responsfraksie in die spesifieke PSE aan.

Indien 'n steekproef selfwegend is, is al die gewigte gelyk, d.w.s. $w_{hij}=w$, 'n konstante wat deur normalisasie gelyk aan een gestel kan word.

Die berekening van die gewigte word in die illustrerende voorbeeld in paragraaf 2.7.2 van hierdie hoofstuk toegelig. Indien $p_{hij}$ die waarskynlikheid is om in die h-de stratum die i-de PSE en daaruit die j-de steekproefelement te trek is, dan is $w_{hij}$:

$$w_{hij} \cong 1/p_{hij} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2.7.2)$$

Let wel vergelyking (2.7.2) is slegs geldig indien daar geen nie-respons is. Indien daar wel nie-respons is moet (2.7.2) as volg gewysig word:

$$w_{hij} \cong 1/(r_{hij}p_{hij}) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2.7.3)$$

As daar vir byvoorbeeld een PSE geen respondente was, dan word vir daardie PSE $x_{hi} = y_{hi} = 0$ gedefinieer.

Die variansie van r word as volg beraam (kyk Kish (1965) p.192)

$$s_r^2 = \sum_h m_h \sum_i (z_{hi}^2 - z_h^2)/[(m_h-1)(\sum_h x_h)^2] \ldots\ldots (2.7\ .4)$$

waarby $z_h = \sum_i z_{hi}$, $z_{hi} = y_{hi} - rx_{hi}$ en $x_h = \sum_i x_{hi} \ldots\ldots (2.7.5)$

Indien daar deurgaans slegs 2 gekose PSE's per stratum is, dit wil sê $m_h = 2$, sal (2.7.5) reduseer na

$$s_r^2 = \sum_h^H (z_{h1} - z_{h2})^2/(x_{h1} + x_{h2})^2$$
$$= \sum_h \{(y_{h1}-rx_{h1})-(y_{h2}-rx_{h2})\}^2/(x_{h1}+x_{h2})^2$$
$$= \sum_h \{(y_{h1}-y_{h2})^2-2r(y_{h1}-y_{h2})(x_{h1}-x_{h2})+r^2(x_{h1}-x_{h2})^2\}/(x_{h1}+x_{h2})^2$$

of

$$= (s^2{}_{yh}-2rs_{yxh}+r^2s^2{}_{xh})/(x_{h1}+x_{h2})^2$$

aangesien byvoorbeeld

$$s_{yh}^2 = 1/(2-1)\sum_i^2 \{y_{hi}-(y_{h1}+y_{h2})/2\}^2$$
$$= (y_{h1}-y_{h2})^2/2 + (y_{h2}-y_{h1})^2/2$$
$$= (y_{h1} - y_{h2})^2$$

2.7.4 Datastel en analises

Vir die doeleindes van die praktiese analises is gebruik gemaak van 'n RGN-datastel wat handel oor vrae aangaande die lewenskwaliteit van mense. Die steekproefontwerp is 'n meerstadium komplekse steekproef gestratifiseer volgens ekonomiese streke in die platteland en landdrosdistrikte in die stedelike gebiede. OSD's is as PSE's in die steekproef-ontwerp getrek. PSE staan vir die primêre of eerste stadium steeproefeenhede (kyk hoofstuk 1). Die OSD's is opnemersubdistrikte wat vir die sensus gebuik is, en wat duidelik beskryfde en afgemerkte grense op kaarte het. Hierdie OSD's is in strata saamgegroepeer in ag nemende die stratifikasie veranderlikes hierbo genoem en ander tersaaklike oorweginge, en daarna is uit elke stratum 'n voorafbepaalde aantal OSD's as PSE's getrek.

Die OSD's is redelik homogeen in bevolkingsgrootte en het normaalweg gewissel van 120 na 240 huishoudings elk. Kleiner OSD's is bymekaar gegooi en groter OSD's is gehalveer sodat die OSD's deurgaans met nie meer as 'n faktor 2 verskil of gevarieer het nie. Hierdie mate van ongelykheid in gewigte is geïgnoreer.

Die huishoudings in elk van die getrekte OSD's is gelys en in elke OSD is vanaf die lys 'n aantal huishoudings op 'n sistematies of ewekansige wyse getrek. In die RGN steekproef soos hierbo beskryf verteenwoordig die huishoudings die sekondêre steekproefeenhede (SSE's). Die kwalifiserende individu waar toepaslik is op 'n ewekansige wyse getrek.

Daar is gebruik gemaak van sekere kruisgroepe, naamlik:

AG00 - persone met ouderdomme 18 tot 30 jaar

AG31 - persone met ouderdomme 31 tot 45 jaar

AG46 - persone met ouderdomme 46 tot 60 jaar

AG60 - persone met ouderdomme ouer as 60 jaar

SEXM - mans

SEXV - vrouens

OGTR - ongetroude persone

GTR - getroude persone

ST9 - persone met std9 en laer kwalifikasie

ST10 - persone met std10 en hoër kwalifikasie

Die afhanklike veranderlikes waarvan gebruik gemaak is, is in vier groepe verdeel nl:

a) Demografiese veranderlikes.

b) Politiese veranderlikes.

c) Veranderlikes wat oor ekonomiese vooruitsigte en werkstevredenheid handel.

d) Veranderlikes wat oor godsdiens en liefde handel.

DEMOGRAFIESE VERANDERLIKES

OUD - ouderdom

INK - inkomste gekategoriseer as volg:

1: 0-R124 per maand

2: R125-R249 per maand

3: R250-R499 per maand

4: R500-R749 per maand

5: R750-R999 per maand

6: R1 000-R1 249 per maand

7: R1 250-R1 499 per maand

8: R1 500-R1 999 per maand

9: R2 000-R2 499 per maand

10: R2 500-R2 999 per maand

11: R3 000-R3 999 per maand

12: R4 000-R4 999 per maand

13: R5 000 of meer per maand

HUIS - aantal persone in die huishouding

KIND - aantal kinders onder 11 jaar in die huishouding

POLITIESE VERANDERLIKES

BEV1 - Hoe belangrik is u verhouding met ander bevolkingsgroepe.

BEV2 - Hoe tevrede is u met die manier waarop u oor die weg kom met ander bevolkingsgroepe.

BEV3 - Hoe tevrede is u met die respek wat ander bevolkingsgroepe aan u betoon.

STEM - Hoe tevrede is u met u stemreg.

POL - Hoe belangrik is politieke aangeleenthede vir u.

EKONOMIESE VERANDERLIKES

SAL1 - Hoe tevrede is u met u salaris.

SAL2 - Hoe belangrik is u inkomste vir u.

SAL3 - Hoe tevrede is u met die manier waarop u in staat is om vir u gesin te sorg.

WERK - Hoe belangrik is u werk of beroep.

KOS - Hoe tevrede is u met die voedselpryse.

HUUR - Hoe tevrede is u met die huur wat u betaal.

TRAN - Hoe tevrede is u met u vervoerkostes.

VERH - Die manier waarop u behandel word by die werk.

GODSDIENSTIGE VERANDERLIKES

REL1 - Hoe belangrik is u godsdiensoortuigings.

REL2 - Hoe tevrede is u met u godsdienstige lewe.

HUW1 - Hoe belangrik is u huwelik of liefdesverhouding.

GLUK - Hoe tevrede is u met u gesin se geluk en vrede.

OUER - Hoe tevrede is u met hoe 'n goeie ouer u is.

TOEK - Hoe belangrik is u kinders se toekoms vir u.

Die bogenoemde vrae wat handel oor hoe tevrede die respondent met sekere aspekte is, is as volg gekategoriseer:

0: Tevrede  1: Ontevrede

Die vrae wat handel oor hoe belangrik sekere aangeleenthede vir die respondent is, is as volg gekategoriseer:

0: Belangrik  1: Nie belangrik

2.7.5 <u>Bespreking van resultate</u>.

Hier volg 'n paar opsommende tabelle wat verkry is uit die uitvoer van die program CLUSTERS. Alle analises is uitgevoer op BLANKES, ASIËRS en KLEURLINGE. Die analises is uitgevoer volgens die drie groepe veranderlikes naamlik Politiese, Ekonomiese en Godsdienstige veranderlikes. Van die analises is ook gedoen op die subgroepe nl. ouderdomsubgroepe, geslag, huwelikstatus en kwalifikasie. Die opskrifte van die tabelle beteken die volgende:

NAAM: Verwys na die naam van die veranderlike.

R : Dui die gemiddeld van die veranderlike-waardes aan.

SE : Dui die standaardfout van R, met inagneming van die steekproef ontwerp.

DEFT: Dui die vierkantswortel van die ontwerpeffek aan.

ROH : $(DEFT^2 - 1)/(B - 1)$ met B die gemiddelde PSE-grootte

GEM ROH : Die gemiddelde roh-waarde geneem oor al die kruisgroepe per veranderlike.

6/5 : Die waarde in kolom 6 gedeel deur die waarde in kolom 5.

Standaardfoute is bereken vir 23 verskillende veranderlikes. Daar is altesaam 3 identiese ondersoeke vir die verskillende bevolkingsgroepe gedoen nl: Blankes, Asiërs en Kleurlinge. Die berekenings is vir elk van die opnames gedoen vir gemiddeldes van veranderlike-waardes gebaseer op die hele steekproef sowel as op 12 subgroepe en vir verskille tussen 6 pare van subgroepgemiddeldes.

Die veranderlikes in die tabelle 2.1 tot 2.3 is volgens die roh-waardes, wat in 'n dalende volgorde voorkom, gerangskik.
In plaas daarvan om die verskillende groepe soortgelyke veranderlikes in aparte tabelle weer te gee word hulle onderskei deur die kodes 1 tot 4 in tabelle 2.1 tot 2.3. Die kodes, wat heel links in elke ry se eerste kolom aangedui word, beteken die volgende:

1. Demografiese veranderlikes
2. Politiese veranderlikes
3. Ekonomiese veranderlikes
4. Godsdienstige veranderlikes

Sekere verskille in waardes vir dieselfde tipe veranderlikes kan in elke opname en tussen die opnames gemerk word. Hierdie verskille kom egter nie deurgaans voor nie en is te wyte aan 'n groot hoeveelheid

BLANKES TABEL 2.1

| 1 NAAM | 2 R | 3 SE | 4 DEFT | 5 ROH | 6 GEM. ROH | 7 6/5 |
|---|---|---|---|---|---|---|
| 1 INK | 6.500 | 0.183 | 1.867 | 0.366 | 0.562 | 1.536 |
| 1 HUIS | 3.136 | 0.093 | 1.690 | 0.273 | 0.540 | 1.978 |
| 1 KIND | 0.715 | 0.063 | 1.634 | 0.246 | 0.467 | 1.898 |
| 3 KOS | 0.249 | 0.023 | 1.547 | 0.205 | 0.398 | 1.941 |
| 3 SAL3 | 0.659 | 0.025 | 1.501 | 0.185 | 0.326 | 1.762 |
| 4 OUER | 0.744 | 0.022 | 1.468 | 0.170 | 0.370 | 2.176 |
| 4 TOEK | 0.866 | 0.017 | 1.442 | 0.159 | 0.458 | 2.881 |
| 3 HUUR | 0.550 | 0.024 | 1.362 | 0.126 | 0.305 | 2.421 |
| 2 STEM | 0.833 | 0.017 | 1.341 | 0.117 | 0.327 | 2.795 |
| 4 GLUK | 0.904 | 0.014 | 1.336 | 0.116 | 0.286 | 2.466 |
| 3 TRAN | 0.588 | 0.023 | 1.327 | 0.112 | 0.240 | 2.143 |
| 1 OUD | 42.884 | 0.771 | 1.292 | 0.099 | 0.228 | 2.303 |
| 3 SAL2 | 0.963 | 0.009 | 1.281 | 0.095 | 0.232 | 2.442 |
| 2 BEV1 | 0.821 | 0.017 | 1.267 | 0.089 | 0.138 | 1.551 |
| 4 REL1 | 0.919 | 0.012 | 1.253 | 0.084 | 0.182 | 2.167 |
| 3 SAL1 | 0.569 | 0.021 | 1.234 | 0.077 | 0.290 | 3.766 |
| 3 VERH | 0.608 | 0.021 | 1.224 | 0.073 | 0.122 | 1.671 |
| 2 POL | 0.620 | 0.021 | 1.224 | 0.073 | 0.223 | 3.055 |
| 3 WERK | 0.829 | 0.016 | 1.211 | 0.069 | 0.223 | 3.232 |
| 4 HUW1 | 0.882 | 0.013 | 1.153 | 0.049 | 0.223 | 4.551 |
| 4 REL2 | 0.867 | 0.013 | 1.113 | 0.035 | 0.053 | 1.514 |
| 2 BEV3 | 0.847 | 0.013 | 1.007 | 0.002 | 0.049* | 24.500* |
| 2 BEV2 | 0.896 | 0.011 | 1.004 | 0.001 | 0.052* | 52.000* |
| TOTGEM | | | | 0.134 | 0.295 | 2.184 |

steekproefvariasie.

In tabelle 2.1, 2.2 en 2.3 blyk dit dat die patroon van die roh-waardes deurgaans vir die drie bevolkingsgroepe as volg geskets kan word: Die roh-waardes van die Godsdienstige veranderlikes kode(4) kom taamlik laag voor vir al drie die opnames, dit wil sê die roh-waardes neig om oor die algemeen kleiner te wees. Die roh-waardes van die Ekonomiese veranderlikes kode (3) kom wyd verspreid voor. Die roh- en deft-waardes van die Politiese veranderlikes kode(2) neig om baie hoog voor te kom veral vir Asiërs en Kleurlinge.

Ongelykhede tussen opnames kan aan verskeie faktore toegeskryf word soos byvoorbeeld verskillende steekproefeenhede, variërende PSE-groottes in die verskillende stadiums, die verdeling van die veranderlikes in die populasie, sowel as die bydraende variansiefaktor van die ondervraer.

<u>Resultate</u> <u>vir</u> <u>subgroepgemiddeldes</u> <u>en</u> <u>vergelykings</u>.

Die subgroep-roh's in kolom 6 van tabelle 2.1, 2.2 en 2.3 is gemiddeldes vir elke veranderlike individueel geneem oor al die moontlike subgroepe. Die waardes in kolom 6 neig om oor die algemeen te verminder saam met dié van kolom 5, hoewel dit lank nie deurgaans die geval is nie. Die rede hiervoor is dat die gemiddelde PSE-grootte vir sommige subgroepe (veral die wat met * gemerk is) te klein raak om 'n betroubare roh-waarde weer te gee. Indien die gemiddelde PSE-grootte kleiner as 6 is, bereken die program CLUSTERS nie die roh-waarde nie omrede in derglike gevalle die roh-waarde onstabiel is. Die roh-waardes in hierdie gevalle wat in die tabelle aangegee word is egter met die hand bereken vir

ASIËRS TABEL 2.2

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| NAAM | R | SE | DEFT | ROH | GEM. ROH | 6/5 |
| 2 STEM | 0.254 | 0.036 | 2.957 | 0.516 | 0.501 | 0.971 |
| 2 POL | 0.449 | 0.035 | 2.554 | 0.368 | 0.200 | 0.534 |
| 3 KOS | 0.146 | 0.024 | 2.464 | 0.338 | 0.430 | 1.272 |
| 3 TRAN | 0.374 | 0.033 | 2.441 | 0.330 | 0.349 | 1.058 |
| 2 BEV2 | 0.798 | 0.027 | 2.407 | 0.319 | 0.297 | 0.931 |
| 2 BEV3 | 0.739 | 0.028 | 2.254 | 0.272 | 0.183 | 0.673 |
| 2 BEV1 | 0.903 | 0.018 | 2.168 | 0.246 | 0.437 | 1.776 |
| 3 SAL2 | 0.825 | 0.022 | 2.038 | 0.210 | 0.183 | 0.871 |
| 1 INK | 4.557 | 0.133 | 2.009 | 0.202 | 0.155 | 0.767 |
| 3 HUUR | 0.443 | 0.028 | 1.997 | 0.199 | 0.164 | 0.824 |
| 1 KIND | 1.626 | 0.100 | 1.965 | 0.190 | -0.165 | -0.868 |
| 4 HUW1 | 0.827 | 0.019 | 1.815 | 0.153 | 0.036 | 0.235 |
| 3 SAL1 | 0.342 | 0.023 | 1.709 | 0.128 | 0.215 | 1.680 |
| 3 SAL3 | 0.686 | 0.022 | 1.704 | 0.127 | 0.022 | 0.173 |
| 3 WERK | 0.739 | 0.021 | 1.692 | 0.124 | 0.002 | 0.016 |
| 1 HUIS | 5.060 | 0.099 | 1.549 | 0.093 | -0.037 | -0.398 |
| 3 VERH | 0.473 | 0.021 | 1.493 | 0.082 | 0.156 | 1.902 |
| 4 OUER | 0.748 | 0.017 | 1.412 | 0.066 | 0.041 | 0.621 |
| 4 REL2 | 0.960 | 0.007 | 1.277 | 0.042 | 0.117* | 2.786* |
| 4 TOEK | 0.806 | 0.014 | 1.255 | 0.038 | -0.246* | -6.470* |
| 4 GLUK | 0.918 | 0.010 | 1.248 | 0.037 | 0.213 | 5.757* |
| 4 REL1 | 0.918 | 0.004 | 1.073 | 0.010 | 0.036 | 3.600* |
| 1 OUD | 35.388 | 0.365 | 1.971 | 0.004 | 0.008 | 2.000 |
| TOTGEM | | | | 0.169 | 0.143 | 0.900 |

volledigheidshalwe.

Die breukwaardes in kolom 7 van gemiddelde subgroep-roh's teenoor die totale steekproef-roh's (kolom 6/kolom 5) varieer rondom 2.1 vir die Blankes, 0.9 vir die Asiërs en 1.2 vir die Kleurlinge, met uitsluiting van die hoogs onbetroubare gevalle wat met * gemerk is.

Vir die opname onder die Blankes was die gemiddelde PSE-grootte vir meeste subgroepe baie klein terwyl vir die opname onder die Asiërs die gemiddelde PSE-grootte meer bevredigend was. Gevolglik is die roh-waardes van die veranderlikes vir die Blankes oor die algemeen minder stabiel as die vir die Asiërs of Kleurlinge.

Die gemiddeldes van die breuke van die subgroep roh's (kruisgroepe) teenoor die totale roh-waardes is onderskeidelik (vir die drie bevolkingsgroepe) 2.184(Blank), 0.900(Asiër) en 1.180(Kleurling). 'n Ander aanwysing kan bereken word deur die breuk van die gemiddeld van die roh's oor die subgroepe (kolom 6) te beskou teenoor die gemiddeld van die roh's vir die hele steekproef (kolom 5). Hierdie breuke is onderskeidelik 2.201(Blank), 0.846(Asiër) en 0.957(Kleurling) en is 'n beter maatstaf aangesien die groter roh's, wat belangriker is, outomaties 'n groter gewig het.

'n Belangrike aspek naamlik die verwantskappe tussen die standaardfoute, deff-waardes en roh-waardes van die gemiddeldes van die veranderlike-waardes is tot dusver nog nie beskou nie. Dit word vervolgens beskou en vir dié doeleindes is tabelle 2.1 tot 2.3 herrangskik volgens die groepe soortgelyke veranderlikes, dit wil sê volgens die

KLEURLINGE TABEL 2.3

| 1 NAAM | 2 R | 3 SE | 4 DEFT | 5 ROH | 6 GEM. ROH | 7 6/5 |
|---|---|---|---|---|---|---|
| 3 TRAN | 0.418 | 0.031 | 1.931 | 0.307 | 0.269 | 0.876 |
| 1 INK | 3.608 | 0.142 | 1.929 | 0.306 | 0.789 | 2.578 |
| 2 BEV2 | 0.709 | 0.028 | 1.890 | 0.289 | 0.027 | 0.093 |
| 3 HUUR | 0.528 | 0.028 | 1.689 | 0.208 | 0.256 | 1.231 |
| 3 KOS | 0.127 | 0.018 | 1.644 | 0.191 | -0.428 | -2.241 |
| 2 BEV3 | 0.613 | 0.026 | 1.610 | 0.179 | 0.215 | 1.201 |
| 3 SAL3 | 0.669 | 0.024 | 1.572 | 0.165 | -0.151 | -0.915 |
| 2 STEM | 0.219 | 0.021 | 1.549 | 0.157 | 0.056 | -0.357 |
| 2 BEV1 | 0.768 | 0.021 | 1.509 | 0.144 | -0.082 | -0.569 |
| 4 REL2 | 0.930 | 0.013 | 1.497 | 0.140 | 0.238 | 1.700 |
| 2 POL | 0.322 | 0.023 | 1.484 | 0.135 | 0.330 | 2.444 |
| 3 SAL1 | 0.450 | 0.024 | 1.469 | 0.130 | 0.133 | 1.023 |
| 1 HUIS | 5.430 | 0.116 | 1.450 | 0.124 | 0.231 | 1.863 |
| 3 SAL2 | 0.929 | 0.012 | 1.447 | 0.123 | -0.030 | -0.244 |
| 4 GLUK | 0.893 | 0.014 | 1.407 | 0.110 | 0.172 | 1.564 |
| 3 VERH | 0.597 | 0.022 | 1.381 | 0.103 | -0.012 | -0.117 |
| 1 KIND | 1.482 | 0.062 | 1.319 | 0.083 | 0.170 | 2.048 |
| 1 OUD | 38.852 | 0.579 | 1.253 | 0.064 | 0.247 | 3.859 |
| 4 OUER | 0.778 | 0.017 | 1.226 | 0.057 | 0.253 | 4.439 |
| 4 TOEK | 0.829 | 0.015 | 1.201 | 0.050 | 0.445 | 8.900 |
| 3 WERK | 0.798 | 0.016 | 1.194 | 0.048 | -0.140 | -2.917 |
| 4 HUW1 | 0.799 | 0.016 | 1.190 | 0.047 | 0.003 | 0.064 |
| 4 REL1 | 0.983 | 0.005 | 1.142 | 0.034 | 0.055 | 1.618 |
| TOTGEM | | | | 0.138 | 0.132 | 1.180 |

BLANKES TABEL 2.4

| NAAM | R | SE | DEFT | DEFF | ROH | GEM. ROH |
|---|---|---|---|---|---|---|
| INK | 6.500 | 0.183 | 1.867 | 3.485 | 0.366 | 0.562 |
| HUIS | 3.136 | 0.093 | 1.690 | 2.856 | 0.273 | 0.540 |
| KIND | 0.715 | 0.063 | 1.634 | 2.670 | 0.246 | 0.467 |
| OUD | 42.884 | 0.771 | 1.292 | 1.669 | 0.099 | 0.228 |
| TOTGEM | | | | 2.670 | | |
| STEM | 0.833 | 0.017 | 1.341 | 1.798 | 0.117 | 0.327 |
| POL | 0.620 | 0.021 | 1.224 | 1.498 | 0.073 | 0.223 |
| BEV1 | 0.821 | 0.017 | 1.267 | 1.605 | 0.089 | 0.138 |
| BEV2 | 0.896 | 0.011 | 1.004 | 1.008 | 0.001 | 0.052* |
| BEV3 | 0.847 | 0.013 | 1.007 | 1.014 | 0.002 | 0.049* |
| TOTGEM | | | | 1.385 | | |
| KOS | 0.249 | 0.023 | 1.547 | 2.393 | 0.205 | 0.398 |
| SAL1 | 0.569 | 0.021 | 1.234 | 1.523 | 0.077 | 0.290 |
| SAL2 | 0.963 | 0.009 | 1.281 | 1.641 | 0.095 | 0.232 |
| SAL3 | 0.659 | 0.025 | 1.501 | 2.253 | 0.185 | 0.326 |
| HUUR | 0.550 | 0.024 | 1.362 | 1.855 | 0.126 | 0.305 |
| TRAN | 0.588 | 0.023 | 1.327 | 1.761 | 0.112 | 0.240 |
| VERH | 0.608 | 0.021 | 1.224 | 1.498 | 0.073 | 0.122 |
| WERK | 0.829 | 0.016 | 1.211 | 1.467 | 0.069 | 0.223 |
| TOTGEM | | | | 1.799 | | |
| OUER | 0.744 | 0.022 | 1.468 | 2.155 | 0.170 | 0.370 |
| TOEK | 0.866 | 0.017 | 1.442 | 2.079 | 0.159 | 0.458 |
| GLUK | 0.904 | 0.014 | 1.336 | 1.785 | 0.116 | 0.286 |
| HUW1 | 0.882 | 0.013 | 1.153 | 1.329 | 0.049 | 0.223 |
| REL1 | 0.919 | 0.012 | 1.253 | 1.570 | 0.084 | 0.182 |
| REL2 | 0.867 | 0.013 | 1.113 | 1.239 | 0.035 | 0.053 |
| TOTGEM | | | | 1.693 | | |

indentifiserende kode in die eerste kolom.

## Standaardfoute as maatstaf van oordraging.

Indien direk van standaardfoute as maatstaf van oordraging gebruik gemaak word moet dit met groot versigtigheid gedoen word. Die grootte van die steekproef of subgroep mag nie grootliks varieer nie, die eenheid waarmee die veranderlikes onder beskouing gemeet is moet dieselfde wees, die beramer van die populasieparameters moet identies

ASIËRS TABEL 2.5

| NAAM | R | SE | DEFT | DEFF | ROH | GEM. ROH |
|---|---|---|---|---|---|---|
| INK | 4.557 | 0.133 | 2.009 | 4.036 | 0.202 | 0.155 |
| KIND | 1.626 | 0.100 | 1.965 | 3.861 | 0.190 | -0.165 |
| HUIS | 5.060 | 0.099 | 1.549 | 2.399 | 0.093 | -0.037 |
| OUD | 35.388 | 0.365 | 1.971 | 3.885 | 0.004 | 0.008 |
| TOTGEM | | | | 3.545 | | |
| STEM | 0.254 | 0.036 | 2.957 | 8.743 | 0.516 | 0.501 |
| POL | 0.449 | 0.035 | 2.554 | 6.523 | 0.368 | 0.200 |
| BEV1 | 0.903 | 0.018 | 2.168 | 4.700 | 0.246 | 0.437 |
| BEV2 | 0.798 | 0.027 | 2.407 | 5.794 | 0.319 | 0.297 |
| BEV3 | 0.739 | 0.028 | 2.254 | 5.081 | 0.272 | 0.183 |
| TOTGEM | | | | 6.168 | | |
| SAL1 | 0.342 | 0.023 | 1.709 | 2.921 | 0.128 | 0.215 |
| SAL2 | 0.825 | 0.022 | 2.038 | 4.153 | 0.210 | 0.183 |
| SAL3 | 0.686 | 0.022 | 1.704 | 2.904 | 0.127 | 0.022 |
| WERK | 0.739 | 0.021 | 1.692 | 2.863 | 0.124 | 0.002 |
| VERH | 0.473 | 0.021 | 1.493 | 2.229 | 0.082 | 0.156 |
| KOS | 0.146 | 0.024 | 2.464 | 6.071 | 0.338 | 0.430 |
| TRAN | 0.374 | 0.033 | 2.441 | 5.958 | 0.330 | 0.349 |
| HUUR | 0.443 | 0.028 | 1.997 | 3.988 | 0.199 | 0.164 |
| TOTGEM | | | | 3.886 | | |
| HUW1 | 0.827 | 0.019 | 1.815 | 3.294 | 0.153 | 0.036 |
| OUER | 0.748 | 0.017 | 1.412 | 1.994 | 0.066 | 0.041 |
| REL1 | 0.918 | 0.004 | 1.073 | 1.151 | 0.010 | 0.036 |
| REL2 | 0.960 | 0.007 | 1.277 | 1.631 | 0.042 | 0.117* |
| TOEK | 0.806 | 0.014 | 1.255 | 1.575 | 0.038 | -0.246* |
| GLUK | 0.918 | 0.010 | 1.248 | 1.558 | 0.037 | 0.213 |
| TOTGEM | | | | 1.867 | | |

wees en die tipe veranderlikes wat beraam word moet baie soortgelyk wees. Tabelle 2.4 tot 2.6 is gesorteer volgens groepe soortgelyke veranderlikes naamlik Demografies, Polities, Ekonomies en Godsdienstige veranderlikes.

Vir die Blanke bevolkingsgroep lê die SE-waardes, behalwe vir 'n paar uitskieters voldoende naby mekaar om te sê oordraging werk. SAL2 en OUER is twee vrae van 'n baie sensitiewe aard onder die Blanke bevolkingsgroepe. Indien laasgenoemde twee vrae as uitskieters beskou

KLEURLINGE TABEL 2.6

| NAAM | R | SE | DEFT | DEFF | ROH | GEM. ROH |
|---|---|---|---|---|---|---|
| INK | 3.608 | 0.142 | 1.929 | 3.721 | 0.306 | 0.789 |
| HUIS | 5.430 | 0.116 | 1.450 | 2.103 | 0.124 | 0.231 |
| KIND | 1.482 | 0.062 | 1.319 | 1.740 | 0.083 | 0.170 |
| OUD | 38.852 | 0.579 | 1.253 | 1.570 | 0.064 | 0.247 |
| TOTGEM | | | | 2.284 | | |
| BEV1 | 0.768 | 0.021 | 1.509 | 2.274 | 0.144 | -0.082 |
| BEV2 | 0.709 | 0.028 | 1.890 | 3.572 | 0.289 | 0.027 |
| BEV3 | 0.613 | 0.026 | 1.610 | 2.592 | 0.179 | 0.215 |
| STEM | 0.219 | 0.021 | 1.549 | 2.399 | 0.157 | 0.056 |
| POL | 0.322 | 0.023 | 1.484 | 2.202 | 0.135 | 0.330 |
| TOTGEM | | | | 2.608 | | |
| SAL1 | 0.450 | 0.024 | 1.469 | 2.158 | 0.130 | 0.133 |
| SAL2 | 0.929 | 0.012 | 1.447 | 2.094 | 0.123 | -0.030 |
| SAL3 | 0.669 | 0.024 | 1.572 | 2.471 | 0.165 | -0.151 |
| TRAN | 0.418 | 0.031 | 1.931 | 3.729 | 0.307 | 0.269 |
| HUUR | 0.528 | 0.028 | 1.689 | 2.853 | 0.208 | 0.256 |
| KOS | 0.127 | 0.018 | 1.644 | 2.703 | 0.191 | -0.428 |
| WERK | 0.798 | 0.016 | 1.194 | 1.426 | 0.048 | -0.140 |
| VERH | 0.597 | 0.022 | 1.381 | 1.907 | 0.103 | -0.012 |
| TOTGEM | | | | 2.418 | | |
| GLUK | 0.893 | 0.014 | 1.407 | 1.980 | 0.110 | 0.172 |
| OUER | 0.778 | 0.017 | 1.226 | 1.503 | 0.057 | 0.253 |
| TOEK | 0.829 | 0.015 | 1.201 | 1.442 | 0.050 | 0.445 |
| HUW1 | 0.799 | 0.016 | 1.190 | 1.416 | 0.047 | 0.003 |
| REL1 | 0.983 | 0.005 | 1.142 | 1.304 | 0.034 | 0.055 |
| REL2 | 0.930 | 0.013 | 1.497 | 2.241 | 0.140 | 0.238 |
| TOTGEM | | | | 1.648 | | |

word is die variasie tussen soortgelyke veranderlikes slegs 0.011 tot 0.021 wat baie klein is. By die Asiër bevolkingsgroep is godsdienstige vrae baie sensitief. Indien die veranderlikes REL1 en REL2 sowel as TRAN en BEV1 in tabel 2.5 as uitskieters beskou word, lê die SE-waardes vir die verskillende groepe gelyksoortige veranderlikes weer eens naby genoeg aan mekaar om te sê oordraging tussen standaardfoute kan suksesvol toegepas word. Vir die Kleurling bevolkingsgroep kan oordraging ook suksesvol toegepas word, indien REL1 en SAL2 as uitskieters gesien word.

Behalwe vir die uitskieters, varieer die SE-waardes vir soortgelyke veranderlikes in elke bevolkingsgroep met 'n faktor van uiters 0.01. Die SE-waardes lê dus voldoende naby mekaar om te sê dat oordraging met groot sukses toegepas kan word.

Die deff-waarde as maatstaf van oordraging.

In die teoretiese gedeelte van hierdie hoofstuk is gemeld dat deff as maatstaf van oordraging aanbeveel word vir gesegregeerde subgroepe. Binne elk van die drie opname groepe wat geanaliseer is, is daar geen gesegregeerde subgroepe nie. Al drie die opnames onder die verskillende bevolkingsgroepe is egter gedoen volgens dieselfde basiese steekproefontwerp, en die groottes van die steekproefopnames verskil ook nie drasties nie. Gevolglik is die drie opnames onder die verskillende bevolkingsgroepe as gesegregeerde groepe beskou. Indien die model in (2.3.12) naamlik $\mathrm{deft}_1^2 = k(\mathrm{deft}_0^2-1)$ beskou word waarby $\mathrm{deft}_0^2$ die berekende deff-waarde aandui en $\mathrm{deft}_1^2$ die onbekende deff-waarde aandui, dan kan die imputasiewaarde k vir die vier verskillende groepe soortgelyke veranderlikes bereken word deur (2.3.12) as volg te herlei:

$$k=(\mathrm{deft}_1^2-1)/(\mathrm{deft}_0^2-1)$$

$$=(\mathrm{deff}_1-1)/\mathrm{deff}_0-1)$$

In tabel 2.7 volg 'n paar moontlike waardes vir k bereken uit tabelle 2.4 tot 2.6.

Die Demografiese en Godsdienstige veranderlikes is ook individueel beskou (Tabel 2.8) om te kyk wat die omvang van die k-waardes is. Slegs

Tabel 2.7

| | $Deff_0$ | | $Deff_1$ |
|---|---|---|---|
| | ASIËRS | -> | BLANKES |
| DEMOGRAFIES | | 0.656 | |
| POLITIES | | 0.074 | |
| EKONOMIES | | 0.277 | |
| GODSDIENS | | 0.799 | |
| | KLEURLINGE | -> | BLANKES |
| DEMOGRAFIES | | 0.522 | |
| POLITIES | | 0.239 | |
| EKONOMIES | | 0.563 | |
| GODSDIENS | | 1.069 | |
| | ASIËRS | -> | KLEURLINGE |
| DEMOGRAFIES | | 0.505 | |
| POLITIES | | 0.311 | |
| EKONOMIES | | 0.491 | |
| GODSDIENS | | 0.747 | |

die imputasie van Blankes na Asiërs en van Asiërs na Kleurlinge is beskou. Die imputasiewaardes, k, het gewissel tussen 0.1 en 7.0. Die

Tabel 2.8

| BLANKES | | -> | ASIËRS | -> | KLEURLINGE |
|---|---|---|---|---|---|
| NAAM | DEFF | k | DEFF | k | DEFF |
| INK | 4.036 | 0.819 | 3.485 | 1.095 | 3.721 |
| HUIS | 2.856 | 0.754 | 2.399 | 0.788 | 2.103 |
| KIND | 2.670 | 0.713 | 3.861 | 0.259 | 1.740 |
| OUD | 1.669 | 4.312 | 3.885 | 0.198 | 1.570 |
| Std.afw | | 1.667 | | 0.431 | |
| OUER | 2.155 | 0.861 | 1.994 | 0.506 | 1.503 |
| TOEK | 2.079 | 0.533 | 1.575 | 0.768 | 1.442 |
| GLUK | 1.785 | 0.711 | 1.558 | 1.756 | 1.980 |
| HUW1 | 1.329 | 6.973 | 3.294 | 0.181 | 1.416 |
| REL1 | 1.570 | 0.265 | 1.151 | 2.013 | 1.304 |
| REL2 | 1.239 | 2.640 | 1.631 | 1.967 | 2.241 |
| Std.afw | | 2.579 | | 0.808 | |

standaardafwykings vir k het taamlik gevarieer.

Die roh-waarde as maatstaf van oordraging.

Roh as maatstaf van oordraging is veral van belang by kruisgroepe. Alhoewel verskeie kruisgroepe in hierdie drie opnames soos byvoorbeeld geslag, ouderdomsgroepe, kwalifikasie en huwelikstatus beskou is, was die gemiddelde PSE-grootte by meeste kruisgroepe te klein, wat onstabiele roh-waardes tot gevolg het. Gevolglik is daar geen poging aangewend om moontlike imputasiewaardes tussen die roh-waardes vir die hele steekproef na die roh-waardes vir kruisgroepe vir dieselfde tipe veranderlikes te vind nie.

Gemiddelde resultate vir 5 tipe veranderlikes.

Die gemiddelde roh-waardes vir die volgende groepe veranderlikes word bereken.

Sosio-ekonomies: kwalifikasie, huwelikstatus, geslag.

(0.095 + 0.052 + 0.135)/3 = 0.094

(0.096 + 0.029 + 0.054)/3 = 0.060

(0.153 + 0.059 + 0.037)/3 = 0.083

Demografies: # kinders<11jr, # mense in huishouding, ouderdom, Gemiddelde inkomste.

Polities: BEV1, BEV2, BEV3, STEM, POL

Ekonomies: SAL1, SAL2, SAL3, WERK, KOS, HUUR, TRANS, VERH.

Godsdiens: REL1, REL2, HUW1, GLUK, OUER, TOEK.

In tabel 2.9 blyk dit dat vir Blankes, die Demografiese veranderlikes die gemiddeld grootste roh-waarde en die Politiese veranderlikes die gemiddeld kleinste roh-waarde het, terwyl die gemiddeld grootste roh-waarde by die Asiërs en die Kleurlinge by die Politiese veranderlikes voorkom. By die Asiërs en die Kleurlinge kom die gemiddeld kleinste

Tabel 2.9

| | BLANKES | | ASIËRS | | KLEURLINGE | |
|---|---|---|---|---|---|---|
| | N | ROH | N | ROH | N | ROH |
| Sosio-ekonomies | 3 | 0.094 | 3 | 0.060 | 3 | 0.083 |
| Demografies | 4 | 0.246 | 4 | 0.120 | 4 | 0.144 |
| Polities | 5 | 0.049 | 5 | 0.344 | 5 | 0.181 |
| Ekonomies | 8 | 0.118 | 8 | 0.192 | 8 | 0.159 |
| Godsdiens | 6 | 0.102 | 6 | 0.058 | 6 | 0.073 |
| TOTAAL | 26 | 0.122 | 26 | 0.155 | 26 | 0.015 |

roh-waarde by die Godsdienstige veranderlikes voor. Uit die roh-waardes in tabel 2.9 volg dat die Asiërs en Kleurlinge baie homogeen is ten opsigte van die Politiese veranderlikes en heterogeen ten opsigte van die Godienstige veranderlikes. Die Blankes is weer baie heterogeen ten opsigte van die Politiese veranderlikes.

2.7.6 Gevolgtrekking.

Uit die analise van hierdie spesifieke steekproefdata wil dit voorkom asof die beginsel van oordraagbaarheid met groot versigtigheid toegepas moet word.

Om reeds berekende standaardfoute te gebruik om die waardes van die onbekende standaardfoute van beramers van populasieparameters vir soortgelyke veranderlikes in dieselfde steekproefopname te gebruik blyk heel sinvol te wees vir die steekproewe (waar die PSE-grooottes ongeveer dieselfde is) wat hier beskou is.

By die toepassing van oordraagbaarheid deur gebruikmaking van die imputasie van die deff-waardes het die imputasiewaarde k taamlik gevarieer. Uit tabel 2.7 is dit duidelik dat die waarde van k deurgaans die grootste vir Godsdienstige tipe veranderlikes is en die kleinste vir die Politiese tipe veranderlikes. Die grootte van die imputasiewaardes hang af van die gesegregeerde subgroepe wat gebruik is, dit wil sê of vanaf Asiërs na Kleurlinge geïmputeer word, of vanaf Kleurlinge na Blankes en so meer. In tabel 2.8 is die imputasiewaardes, k, vir die Demografies en Godsdienstige veranderlikes individueel beskou, die standaardafwyking van k het gewissel tussen 0.4 en 2.5.

Die roh-waardes is volgens Kish en Frankel (1974) die beste waardes om te gebruik vir oordraging wanneer kruisgroepe beskou word. Alhoewel heelwat kruisgroepe in die analises beskou is, is gevind dat die gemiddelde PSE-groottes by kruisgroepe telkens te klein word en gevolglik tot onstabiele roh-waardes lei. Daar is dus geen poging aangewend om moontlike imputasie-waardes tussen die roh-waardes vir die hele steekproef en die roh-waardes vir verskillende kruisgroepe vir dieselfde tipe veranderlikes te vind nie. Geen uitspraak kan dus gemaak word aangaande die gebruik van die roh-waarde om oordraagbaarheid toe te pas nie.

## 2.8 OPSOMMING.

Indien die data afkomstig van 'n meerdoelige komplekse steekproefontwerp geanaliseer word kan 'n deeglike studie van verskeie vorme van die standaardfoute van beramers baie insiggewende inligting verskaf.

Buiten die standaardfoute van beramers van populasieparameters is daar ook twee ander maatstawwe wat van belang is in die analise van die data van 'n komplekse steekproefopname naamlik, die ontwerpeffek (deff) en die intratroskorrelasiekoëffisiënt (roh). Die ontwerpeffek (deff) wat in die hoofstuk beskou is, is per definisie die verhouding van die steekproefvariansie van 'n beramer wanneer die totale kompleksiteit van die steekproef in ag geneem word teenoor die steekproefvariansie van 'n beramer wanneer die totale kompleksiteit van die steekproef geïgnoreer word, dit wil sê wanneer die steekproef as 'n eenvoudige ewekansige steekproef beskou word.

Die intratroskorrelasiekoëffisiënt (roh) meet die variasie in die veranderlike-waardes onder beskouing binne die PSE's van die steekproefontwerp teenoor die variasie in die veranderlike-waardes beskou oor die hele steekproef.

Al drie die maatstawwe standaardfoute, deff-waardes en roh-waardes kan gebruik word in oordraagbaarheid. Oordraagbaarheid kom kortliks daarop neer dat vanaf enige van die bogenoemde drie maatstawwe word gevolgtrekkings gemaak oor die onbekende waarde van dieselfde maatstaf vir beramers van populasieparameters vir dieselfde of soortgelyke veranderlikes in die hele steekproef of in verskillende subgroepe (kruis- of gesegregeerde subgroepe) van die steekproef, met die uiteindelike doel om gevolgtrekkings oor die standaardfoute van beramers te kan maak.

Kruisgroepe is subgroepe wat in elk van die steekproefnemingstadia min of meer eenvormig versprei is oor die strata en trosse van populasie-elemente (PSE's), terwyl gesegregeerde subgroepe strata of hoëre steekproefeenhede in hul geheel bevat.

Indien die drie maatstawwe van oordraagbaarheid gerangskik word vanaf die minste tot meeste bruikbaar dan sal die rangskikking as volg wees: standaardfoute, deff-waardes en roh-waardes.

Standaardfoute van beramers van populasieparameters is alleen regstreeks oordraagbaar na dié van soortgelyke verandelikes indien die subgroepgroottes dieselfde is en die steekproefontwerpe soortgelyk is.

Ontwerpeffekte (deff) is breër toepasbaar in die praktyk as standaardfoute en is in die algemeen die beste wyse van oordraagbaarheid indien met gesegregeerde subgroepe gewerk word, aangesien die ontwerp van die steekproef en die gemiddelde PSE-grootte, wat 'n groot invloed op die waarde van deff het, nie in die geval van gesegregeerde subgroepe beïnvloed word nie.

Die roh-waarde het die breedste spektrum van toepasbaarheid van al drie die maatstawwe. Die roh-waarde is veral geskik vir gebruik in die geval van kruisgroepe aangesien dit nie geaffekteer word deur die gemiddelde PSE-grootte nie.

Data afkomstig van 'n RGN-opname van huishoudings wat gebaseer is op 'n komplekse steekproef is gebruik vir die toepassing van bogenoemde teorie. Die program CLUSTERS, wat gebruik maak van die Taylor-linearisasiemetode van variansieberaming, is gebruik en verskeie interessante gevolgtrekkings is gemaak.

Die Taylor-linearisasiemetode van variansieberaming is slegs een van verskeie metodes van variansieberaming vir die beramers van populasieparameters vir komplekse steekproewe. Hierdie metode het 'n groot tekortkoming naamlik dat dit beperk is tot die variansieberaming van lineêre beramers (bv. gemiddeldes of proporsies) van populasieparameters, by nie-lineêre beramers veroorsaak hierdie metode groot sydigheid. Daar is twee herhaalde-replikasiemetodes van variansieberaming wat onder andere ontwikkel is om laasgenoemde tekortkoming te oorbrug, dit wil sê wat ook die variansies van nie-lineêre beramers (bv. korrelasiekoëffisiënte en regressiekoëffisiënte) van populasieparameters kan beraam.

Die Taylor-linearisasiemetode van variansieberaming sowel as die twee herhaalde-replikasiemetodes (BRR en JRR) van variansieberaming word in detail in die volgende hoofstuk bespreek.

# III. VARIANSIEBERAMING IN DIE GEVAL VAN KOMPLEKSE STEEKPROEWE

## 3.1 INLEIDING

Die berekening van variansies van beramers van populasieparameters vir ewekansige steekproewe, wat gebaseer is op die aanname van onafhanklike waarnemings, kom algemeen in die literatuur voor. Indien 'n komplekse steekproef beskou word waar waarnemings nie onafhanklik van mekaar getrek is nie, en meer bepaald ook in trosse getrek is, moet 'n meer ingewikkelde metode gevolg word vir die berekening van die variansies van beramers van populasieparameters.

Die doel van hierdie hoofstuk is om drie metodes vir die berekening van die variansies van beramers van populasieparameters vir komplekse steekproewe te beskou. Hierdie drie metodes is die Taylor-linearisasie-metode, die Gebalanseerde Herhaalde-Replikasiemetode of BRR-metode (BRR = 'Balanced Repeated Replication') en die Uitsnit Herhaalde-Replikasiemetode of JRR-metode (JRR = 'Jackknife Repeated Replication').

Robert Jewett het 'n program ontwikkel om ortogonale halfsteekproewe te verkry vir die berekening van variansies volgens die BRR metode. Hierdie program is uitgebrei, gekoppel aan die program CLUSTERS en 'n nuwe program BRRJRR is ontwikkel om variansies volgens die BRR- en JRR-metodes te bereken in die lineêre geval. Verdere uitbreiding van die BRRJRR program na die nie-lineêre geval is regstreeks. In laasgenoemde geval lei die Taylor-linearisasiemetode (CLUSTERS) tot onaanvaarbare resultate.

## 3.2 EERSTE EN TWEEDE ORDE BERAMERS

Enige steekproefberamer (bv. 'n steekproefgemiddelde $\bar{y}$) van 'n populasieparameter (bv. $E(\bar{y})=u$) wat van primêre belang is vir enige ontleder van 'n steekproefopname word beskou as eerste orde beramers. Die beramers wat die variasie van die eerste orde beramers meet word tweede orde beramers genoem.

Eerste orde beramers sluit beramers in soos gemiddeldes, verskille tussen gemiddeldes, korrelasiekoëffisiënte, regressiekoëffisiënte, parsiële korrelasiekoëffisiënte en meervoudige korrelasie-koëffisiënte.

Tweede orde beramers, wat die beramers van die steekproefvariansie van bogenoemde eerste orde beramers is, kan deur verskeie metodes bereken word. Wanneer besluit word op 'n berekeningsmetode vir die tweede orde beramers speel die metode van steekproefneming 'n uiters belangrike rol. Vir eenvoudige ewekansige steekproefneming kan die Taylor-, Skoenlus- of algemene Uitsnitmetode, gebruik word, maar daar bestaan eenvoudiger regstreekse metodes ook. In die geval van 'n komplekse steekproef kan die Taylor, Uitsnit Herhaalde-Replikasie en Gebalanseerde Herhaalde-Replikasie beramingsmetodes gebruik word.

Laasgenoemde drie beramingsmetodes word vervolgens uiteengesit en bespreek.

## 3.3 DIE TAYLOR-LINEARISASIEMETODE VAN VARIANSIEBERAMING

Die Taylor-metode word ook somtyds die 'delta' of $\gamma$-metode genoem. Meeste handboeke behandel die gebruik van die Taylor-metode slegs in die geval van verhoudings of gemiddeldes. Deming (1960) en Kish (1965) propageer die gebruik van die Taylor-metode vir die berekening van die variansie van komplekse eerste orde beramers (d.w.s. funksies anders as verhoudings of gemiddeldes). Die eerste publikasie wat handel oor die gebruik van die Taylor-metode vir meer komplekse eerste orde beramers het in 1968 verskyn (Tepping (1968)).

Die lineêre benadering waarop die Taylor-metode gebaseer is, genoem die Taylor-linearisasiemetode, word verkry deur die eerste orde beramer waarin belanggestel word te skryf in die vorm van die Taylor-uitbreiding en dan slegs die eerste twee terme in die Taylor-uitbreiding te gebruik vir die berekening van die benaderde tweede orde beramer.

Tepping (1968) beskryf die Taylor-linearisasiemetode as volg:

Laat $\underset{\sim}{y} = (y_1, y_2, \ldots\ldots, y_k)$ 'n vektor van k steekproefberamers wees waarby elke element 'n lineêre kombinasie van PSE waardes $y_{iha}$ van die veranderlikes voorstel waarby die indeks, h, die stratum aandui, die indeks, a, die PSE-waarde in daardie stratum aandui, en die indeks i die i-de USE in die a-de PSE van die h-de stratum aandui. Dit wil sê

$$y_i = \sum_{h=1}^{H} \sum_{a=1}^{2} y_{iha} = \sum_{h=1}^{H} y_{ih.} \qquad \text{vir alle } i=1,2,\ldots\ldots,k.$$

Hierby word veronderstel dat daar twee PSE's uit elke stratum in die steekproef is wat met gelyke waarskynlikheid f uit die strata getrek is, en waarby veronderstel word dat daar altesaam H strata is.

Laat $E(\underset{\sim}{y}) = \underline{u} = (u_1, u_2, \ldots.u_k)$ die ooreenstemmende vektor van waardes van die populasieparameters wees. Laat $g(\underline{u})$ die eerste orde parameter voorstel wat deur die funksie $g(\underline{y})$ beraam word.

Die eerste aaname wat gemaak word is dat die variansie van $g(\underline{y})$ benaderd gelyk is aan die variansie van die eerste term van die meerveranderlike Taylor-benadering van $g(\underline{y})$ naby $\underline{u}$. Dit wil sê:

$$Var[g(\underline{y})] \cong Var[g(\underline{u}) + \sum_{i=1}^{k} (y_i - u_i)\ \partial g(\underline{u})/\partial u_i] \ldots\ldots\ldots (3.3.1)$$

waar die parsiële afgeleides $\partial g(\underline{u})/\partial u_i$ bepaal word in die punt $\underline{y} = \underline{u}$.

Aangesien $g(\underline{u})$ en $u_i\{\partial g(\underline{u})/\partial u_i)$ konstant is, vereenvoudig (3.3.1) na

$$Var[g(\underline{y})] \cong Var[\sum_{i=1}^{k} (\partial g(\underline{u})/\partial u_i)y_i] \ldots\ldots\ldots (3.3.2)$$

Een van die aannames vir die toepassing van die Taylor-linearisasie-metode is dat die trekkings onafhanklik tussen strata geskied, sodat (3.3.2) as volg geskryf kan word:

$$var[g(\underline{y})] \cong \sum_{h=1}^{H} W_h^2 Var[\sum_{i=1}^{k} (\partial g(\underline{u})/\partial u_i)y_{ih\cdot}\ \}\ldots\ldots\ldots (3.3.3)$$

waarby $W_h$ die gewig van die h-de stratum aandui, naamlik die grootte van die h-de stratum tot die grootte van die populasie.

Indien daar twee PSE's sonder teruglegging en met gelyke waarskynlikheid $f_h$ uit die h-de stratum getrek word, $h = 1,2,\ldots H$, dan word die variansie van die $y_{ih}$'s beraam deur

$$var[\sum_{i=1}^{k} y_{ih} \partial g(\underline{u})/\partial u_i]$$

$$= (1-f_h)(\sum_i y_{ih1}\partial g(\underline{u})/\partial u_i - \sum_i y_{ih2}\partial g(\underline{u})/\partial u_i)^2 \ldots\ldots(3.3.4)$$

waarby $y_{ih1}$ en $y_{ih2}$ die steekproeftotale van die waardes van die twee PSE's in die h-de stratum is. Keyfitz (1957) het die aandag op hierdie eenvoudige vorm vir 2 PSE's in 'n stratum gevestig.

Indien $W_h = 1$ en $f_h = f$ vir alle $h=1,2,\ldots H$ is, is die beramer van $Var(g(\underline{y}))$ die volgende:

$$var(g(\underline{y})) = (1-f)\sum_{h=1}^{H}[\sum_{i=1}^{k} y_{ih1}\partial g(\underline{u})/\partial u_i - y_{ih2}\partial g(\underline{u})/\partial u_i]^2 \;\ldots.(3.3.5)$$

Vir die gebruik van hierdie beramer is die ideaal om waardes te hê vir die konstantes $\partial g(\underline{u})/\partial u_i$. Indien hierdie konstantes bekend is dan sal $g(\underline{u})$ heel waarskynlik ook bekend wees en sal dit nie nodig wees om $var(g(\underline{y}))$ te beraam nie. Hierdie konstantes moet dus beraam word uit die beskikbare steekproefdata. Vir die beraming van hierdie konstantes word die aanname, nl. dat die substitusie van die steekproefwaardes vir die populasiewaardes nie die fout in die variansieberaming wesentlik vergroot nie, gemaak.

Vir die toepassing van die Taylor-linearisasiemetode is geeneen van die volgende vereistes nodig nie.

a) Gelyke waarskynlikheid van trekking van die PSE's binne of tussen strata.

b) Gelyke stratumgewigte.

c) Presies twee PSE's per stratum.

Hierdie Taylor-linearisasiemetode is daarop gebaseer dat die variasie in die noemer van die verhoudingsberamer klein is om sydigheid in die beraming te beperk. Gevolglik is dit noodsaaklik dat die koëffisiënt van variasie (cv) (kyk paragraaf 2.1.5) van die veranderlike in die noemer van die verhoudingsberamer klein moet wees. 'n Goeie leidraad vir die grootte van cv is dat dit kleiner as 0,1 of uiters 0,2 moet wees om die sydigheid klein te hou. Die toepasbaarheid van die Taylor-linearisasiemetode vereis dus dat die noemer van die verhoudingsberamer as nagenoeg konstant aanvaar kan word. Indien die noemer van die verhoudingsberamer nie min of meer konstant is nie, dan kan die sydigheid in die beraming van so 'n omvang wees dat die Taylor-linearisasiemetode tot onaanvaarbare resultate kan lei. Voorbeelde van verhoudingsberamers waar die veranderlike in die noemer nie min of meer as konstant aanvaar kan word nie is bv. korrelasiekoëffisiënte en regressiekoëffisiënte. Vir hierdie gevalle is die sogenaamde herhaalde-replikasiemetodes ontwikkel.

Daar is twee herhaalde-replikasiemetodes van variansieberaming nl. Gebalanseerde Herhaalde-Replikasie (BRR) en die Uitsnit Herhaalde-Replikasie (JRR).

## 3.4 DIE TWEE HERHAALDE REPLIKASIEMETODES VAN VARIANSIEBERAMING

### 3.4.1 <u>Inleiding</u>

Beskou 'n komplekse meerstadiumsteekproef, dan geld die aanname van onafhanklikheid tussen die waarnemings van die steekproef nie. Indien die standaardfoute van beramers van populasieparameters in komplekse steekproewe bereken wil word is daar ses alternatiewe moontlikhede wat gevolg kan word, naamlik:

a) Om die berekening en weergawe van standaardfoute te ignoreer.

b) Om die standaardfoute te bereken met die bekende formules wat op eenvoudige ewekansige steekproefneming gebaseer is, en wat in die geval onder beskouing tot onderberamings sal lei aangesien die aanname van onafhanklikheid tussen waarnemings nie geld nie.

c) Om die navorser te verwittig dat hy moet wag vir die moontlike ontwikkeling van die teorie wat vir komplekse steekproewe geld.

d) Om ewekansige steekproewe te neem wat voldoen aan die beskikbare verdelingsteorie. Dit is egter meestal onprakties en nie ekonomies nie.

e) Eenvoudige herhaalde steekproewe is gewoonlik ook nie aanvaarbaar nie, aangesien dit in botsing is met die vereiste stratifikasie en nie genoegsame presisie vir die berekening van standaardfoute verskaf nie.

f) 'n Moontlike alternatief kan die toepassing van die Taylor-linearisasiemetode wees wat reeds vroeër in die hoofstuk bespreek is. Die Taylor-linearisasiemetode kan egter nie vir komplekse eerste orde benaderings gebruik word nie. BRR (Gebalaseerde Herhaalde-Replikasie) en JRR (Uitsnit Herhaalde Replikasie) is twee metodes wat spesifiek ontwerp is vir komplekse steekproewe en komplekse eerste orde beramers en is meer algemeen en prakties toepasbaar as die Taylor-linearisasie metode.

Van hierdie ses alternatiewe is die eerste drie alternatiewe nie bevredigend nie, die vierde en vyfde moontlikhede is onprakties en die sesde moontlikheid is die enigste aanvaarbare oplossing. Gevolglik word in die res van die hoofstuk op die BRR- en JRR-metodes van variansieberaming gekonsentreer.

### 3.4.2 Die Gebalanseerde Herhaalde-Replikasiemetode van variansieberaming

BRR word soms ook pseudoreplikasie of halfsteekproefreplikasie genoem. Die naam BRR ('Balanced Repeated Replication'), of Gebalanseerde Herhaalde Halfsteekproewe, is 'n goeie beskrywende woord vir presies wat dit behels, en die betekenis van die naam van die metode kan as volg verduidelik word.

1) 'n Halfsteekproef ('replication' of 'halfsample') word verkry deur uit elk van H strata in die steekproef een van twee PSE's te kies. Let op elke stratum in die steekproef bevat twee PSE's. Hierdie halfsteekproef wat outomaties kleiner is as die hele steekproef moet die kompleksiteit van die steekproefontwerp weergee. Uit die halfsteekproef word die beramerwaarde onder beskouing bereken.

2) Herhaalde ('repeated') halfsteekproewe word gebruik aangesien die beramer wat gebaseer word op 'n enkele halfsteekproef te veranderlik en varieerbaar is. Die proses in (1) kan herhaal word deur nuwe halfsteekproewe te trek en nuwe beramerwaardes te bereken. Vanuit die k halfsteekproewe kan die gemiddeld van k variansieberamers bereken word. Aangesien elk van die k waardes die variansie beraam, sal die gemiddelde van die variansies ook die variansie beraam, ten spyte van die korrelasie tussen hulle. Die presisie van hierdie gemiddelde sal verbeter hoe meer halfsteekproewe geneem word. Hierdie verbetering in die presisie is relatief stadig indien die herhaalde halfsteekproewe sommer willekeurig gekies word sonder enige definitiewe patroon. Vir maksimum doeltreffendheid is

daar dus 'n sekere patroon nodig waarvolgens die halfsteekproewe getrek moet word.

3) Aangesien H strata in 'n steekproef verteenwoordig is, en elke stratum 2 PSE's bevat, dan is daar $2^{H-1}$ moontlike verskillende maniere om halfsteekproewe te trek. McCarthy (1966) het 'n metode ontwikkel naamlik die ortogonale balanseringsmetode. Deur middel van hierdie metode word 'n ortogonale matriks geskep. Hierdie matriks stel dan die patroon waarvolgens die halfsteekproewe getrek moet word, voor. Die halfsteekproewe wat volgens hierdie patroon getrek word is ortogonaal tot mekaar en sodoende word die aantal moontlike kombinasies van halfsteekproewe, naamlik $2^{H-1}$, drasties verminder tot slegs H ortogonale halfsteekproewe. McCarthy (1966) het ook aangetoon dat in die lineêre geval (dit wil sê vir beramers soos bv. gemiddeldes en verhoudings) sal die variansieberamer presies gelyk wees aan die variansieberamer wat verkry sou word indien al $2^{H-1}$ moontlike halfsteekproewe beskou word.

Die BRR-metode van variansieberaming kan as volg verduidelik word. Beskou 'n gestratifiseerde steekproefontwerp met twee geselekteerde PSE's per stratum wat met gelyke waarskynlikheid f gekies is. Let op dat so 'n model wat bestaan uit twee onafhanklike PSE's uit elke stratum, die mees basiese ontwerp is.

Laat S die hele steekproef voorstel en laat $H_i$ die i-de halfsteekproef voorstel wat gevorm word deur een van die twee PSE's uit elk van die H strata te kies. Laat $C_i$ die komplementêre halfsteekproef van $H_i$ wees wat gevorm word deur al die PSE's in S wat nie in $H_i$ is nie.

Indien daar nou k halfsteekproewe $H_1, \ldots, H_k$ gevorm word met hulle ooreenstemmende komplementêre halfsteekproewe $C_1, \ldots, C_k$, dan kan die steekproefberamers van die variansie van die eerste orde beramers g(S) verkry word deur middel van die volgende vergelykings:

Beramer I (BRR-H) - Half minus totaal

$$var_{BRR-H}(g(S)) = ((1-f)/k) \sum_{i=1}^{k} (g(H_i)-g(S))^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots(3.4.1)$$

Beramer II (BRR-C) - Komplement minus totaal

$$var_{BRR-C}(g(S)) = ((1-f)/k) \sum_{i=1}^{k} (g(C_i)-g(S))^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots(3.4.2)$$

Beramer III (BRR-S) - Som van BRR-H en BRR-C

$$var_{BRR-S}(g(S)) = [var_{BRR-H}(g(S)) + var_{BRR-C}(g(S))]/2 \quad \ldots.(3.4.3)$$

Beramer IV (BRR-D) Verskil van BRR-H en BRR-C

$$var_{BRR-D}(g(S)) = ((1-f)/4k) \sum_{i=1}^{k} (g(H_i) - g(C_i))^2 \ldots\ldots\ldots\ldots\ldots(3.4.4)$$

Deur k herhaalde halfsteekproewe te selekteer en dan die gemiddelde van die k berekende variansies te kry word die presisie van die variansieberamer verbeter.

Daar is verskeie metodes om die verskillende PSE's, wat die half- en komplementêre halfsteekproewe $H_i$ en $C_i$ vorm, te verkry. Een metode is die ortogonale balanseringsmetode van McCarthy (1966). Hierdie metode word in die praktiese gedeelte van hierdie hoofstuk bespreek.

Elkeen van die bogenoemde vier tweede orde beramers beraam die variansie van 'n gelineariseerde vorm van die eerste orde beramers. Hierdie linearisasie van g(S) is as volg:

$$1/2k \sum_{i=1}^{k} (g(H_i) + g(C_i)) \quad \text{(3.4.5)}$$

Omdat $H_i$ en $C_i$ maklik omruilbaar is, besit die vergelykings BRR-H en BRR-C dieselfde verwagte waardes. Die BRR-H en BRR-C vorme kan gesien word as beramers van BRR-S wat goedkoper (in terme van rekenaartyd) is om te bereken. Omdat BRR-S meer presies is en beide BRR-H en BRR-C in BRR-S geïnkorporeer is word BRR-H en BRR-C in die res van die bespreking uitgelaat.

### 3.4.3 <u>Basiese eienskappe van BRR variansieberamers en teoretiese grondslag</u>

Die BRR-tegniek van variansieberaming gee goeie benaderings vir 'n groot verskeidenheid beramers. Die doel van hierdie paragraaf is om goeie regverdiging vir die gebruik van die BRR variansieberamingsmetode te gee.

Vir eenvoudigheidshalwe word 'n steekproef, S, bestaande uit twee PSE's per stratum weer eens beskou, waarby die PSE's onafhanklik van mekaar en met teruglegging getrek is. Vir elkeen van die H strata word die twee PSE's deur dieselfde seleksieprosedure verkry. In die steekproef S wat uit 2H PSE's bestaan word op ewekansige wyse een PSE uit elkeen van die strata getrek sodat dit 'n halfsteekproef $H_j$ voorstel. Die oorblywende PSE's in elk van die H strata stel dan die komplementêre halfsteekproef $C_j$ voor. Let op dat $H_j$ en $C_j$ twee halfsteekproewe voorstel wat gebaseer is op dieselfde seleksieprosedures. S stel ook dieselfde seleksieprosedure voor, maar is net verdubbel in elke opsig.

Indien die steekproefberamingsfunksie f op die hele steekproef toegepas word, word die beramer $\hat{b}=f(S)$ verkry. Indien hierdie selfde beramingsfunksie op halfsteekproef $H_j$ toegepas word, word die beramer $b_j = f(H_j)$ verkry en indien dit toegepas word op die komplementêre halfsteekproef $C_j$ word die beramer $b_j' = f(C_j)$ verkry, sodat $\hat{b}$, $b_j$ en $b_j'$ al drie beramers van dieselfde populasieparameter B is. Die doel is nou om $Var(\hat{b})$ te beraam deur gebruikmaking van $b_j$ en $b_j'$ en om die beramer te verbeter deur middel van herhaalde waardes vir $b_j$ en $b_j'$ wat uit die $2^{H-1}$ moontlike halfsteekproewe in S verkry kan word.

Hier volg 'n lys van die simbole en terminologie wat in die verdere bespreking gebruik word:

B = die populasiewaarde wat beraam word.

f(.) = 'n beramingsfunksie

$\hat{b} = f(S)$ = die beramer wat deur die navorser gebruik word om B te beraam, en waarvoor die variansie $Var(\hat{b})$ beraam moet word.

$b_j = f(H_j)$ is 'n beramer van b gebaseer op die j-de halfsteekproef, $H_j$.

$b_j' = f(C_j)$ is 'n beramer van b gebaseer op die j-de komplementêre halfsteekproef, $C_j$.

$\bar{b}_j = (b_j + b_j')/2$ is die gemiddeld van $b_j$ en $b_j'$

$var(\bar{b}_j) = (b_j - b_j')^2/4 = (b_j - \bar{b}_j)^2$ dui die beramer van die variansie van $\bar{b}_j$ aan.

$\overline{var_k}(\bar{b}_j)$ en $\overline{ste_k}(\bar{b}_j)$ wat respektiewelik die beramers van die variansie en van die standaardfout van $\bar{b}_j$, oor k halfsteekproewe, aandui.

Die hoofletter V in Var dui op die variansie van die populasie en die kleinletter v in var dui op die variansie van die steekproef.

In die lineêre geval dit wil sê wanneer die beramingsfunksie lineêr is, is $\hat{b} = \bar{b}_j$, maar ongelukkig is dit nie altyd so in die nie-lineêre geval nie. Die hoofbelangstelling lê egter in beramings vir $Var(\hat{b})$ uit gemiddeldes van $var(\bar{b}_j)$ vir nie-lineêre gevalle.

Laat $E(b_j)$ die verwagte waarde van $b_j = f(H_j)$ vir die j-de halfsteekproef $H_j$ aandui. Aangesien $H_j$ en $C_j$ komplementêr is hoef die formule slegs in die geval van $H_j$ afgelei word. Die ooreenkomstige formule vir $C_j$ kan effektief verkry word deur net die rolle om te ruil vir alle moontlike steekproewe i en j.

$E(b_j) = E(b_j') = E(\bar{b}_j) = E(\bar{b}_i)$..............................(3.4.6a)

In die algemeen is $E(\hat{b}) \neq E(\bar{b}_j)$ vir die nie-linieêre geval.

Verder is die variansie van die populasie

$$\begin{aligned} Var(\bar{b}_j) &= E[(b_j - b_j')^2/4] \\ &= E(b_j - \bar{b}_j)^2 \\ &= E(b_j' - \bar{b}_j)^2 \quad \text{.................................(3.4.6b)} \end{aligned}$$

en

$Var(\bar{b}_i) = Var(\bar{b}_j)$ vir alle i en j

Soos reeds aangedui beraam $var(\bar{b}_j) = (b_j - b_j')^2/4$ die variansie van $\bar{b}_j$. Hierdie beramer van die variansie is baie onstabiel (Kish 1965) en deur middel van herhalings kan die presisie verbeter word. Deur k halfsteekproewe te neem en die beramingsproses k keer te herhaal word die gemiddelde variansieberamer

$$\overline{var_k}(\bar{b}_j) = \sum_j (b_j - b_j')^2/4k \text{ ..........................(3.4.6c)}$$

verkry, wat die presisie baie verbeter. Omdat die beraamde variansie vir elke herhaling dieselfde verwagte waarde het uit (3.4.6a) sal

$$E[var(\bar{b}_j)] = E[\overline{var_k}(\bar{b}_j)] = Var(\bar{b}_j) \text{.................(3.4.6d)}$$

McCarthy (1966) het die lineêre geval beskou en aangetoon dat deur gebruikmaking van 'n ortogonale balanseringspatroon vir die keuse van herhaalde halfsteekproewe, kan variansieberamers deur middel van H halfsteekproewe verkry word wat gelyk sal wees aan die variansieberamers wat verkry sou word indien al $2^{H-1}$ moontlike halfsteekproewe gebruik word. Hierdie argument is egter slegs geldig in die lineêre geval en nie in die nie-lineêre geval nie. Hierdie ortogonaal gebalanseerde metode is deur Kish (1966) in die nie-lineêre geval ook suksesvol gebruik aangesien hy glo dat dit die beste metode is.

Die ideaal is dat deur middel van BRR $\overline{var_k}(\bar{b}_j)$ 'n presiese beramer van $Var(\bar{b}_j)$ is, en dat hierdie beramer so na as moontlik aan $Var(\hat{b})$ is. Alhoewel $\bar{b}_j \neq \hat{b}$ in die nie-lineêre geval is, het Kish (1966) gevind dat die verskil gering is vir baie nie-lineêre beramers, en in hierdie geval is die BRR-tegniek baie bruikbaar.

Kish en Frankel (1974) het in verskeie opnames deur middel van empiriese ondersoeke gevind dat vir verskeie beramers en vir elke herhaling die waarde van $\bar{b}_j = (b_j + b_j')/2$ baie na aan dié van $\hat{b}$ was.

Definieer $e_j = \bar{b}_j - \hat{b}$. Let op $(b_j - \bar{b}_j) = (b_j - b_j')/2$, dan volg:

$$(b_j - \hat{b})^2 = (b_j - \bar{b}_j + \bar{b}_j - \hat{b})^2$$
$$= (b_j - b_j')^2/4 + e_j^2 + e_j(b_j - b_j') \ldots\ldots\ldots\ldots\ldots\ldots (3.4.7a)$$

Indien die gemiddeldes geneem word oor k halfsteekproewe vir $(b_j - \hat{b})^2$ en $(b_j' - \hat{b})^2$ volg:

$$\sum_j (b_j - \hat{b})^2/k - \overline{var_k}(\bar{b}_j) = \sum_j e_j^2/k + \sum_j e_j(b_j - b_j')/k$$
$$\sum_j (b_j' - \hat{b})^2/k - \overline{var_k}(\bar{b}_j) = \sum_j e_j^2/k + \sum_j e_j(b_j - b_j')/k \ldots\ldots (3.4.7b)$$

In die lineêre geval is $e_j = (\bar{b}_j - \hat{b}) = 0$ vir elke herhaling, dus sal beide terme aan die regterkant wegval, en

$$\overline{var_k}(\bar{b}_j) = \sum_j (b_j - b_j')^2/4k$$
$$= \sum_j (b_j - \hat{b})^2/k$$
$$= \sum_j (b_j' - \hat{b})^2/k$$

Alhoewel $e_j = (\bar{b}_j - \hat{b}) = 0$ nie presies geld vir die nie-lineêre geval nie, geld dit wel benaderd (Kish (1966) het dit empiries aangetoon).

Kish (1965) het verskeie ondersoeke uitgevoer op die gekwadreerde afwykings $(\bar{b}_j - \hat{b})^2$ van die beramers $(b_j - b_j')^2/4$ van die variansie van $\bar{b}_j$ en empiriese antwoorde bekom vir die volgende twee basies verwante vrae:

a) Is herhaalde waardes van $\bar{b}_j$ voortdurend naby die waarde van

$\hat{b}$?

b) Is die twee beramers $\sum_j [(b_j - \hat{b})^2 + (b_j' - \hat{b})^2]/2k$ en $\sum_j (b_j - b_j')^2/4k$ naby mekaar?

Beide bogenoemde vrae is beantwoord deur die berekening van die waardes van die gekwadreerde verskille $\sum_j (\bar{b}_j - \hat{b})^2$ as breuk van die variansieberamers $\sum_j (b_j - b_j')^2/4$ te neem. Hierdie berekenings is in drie studies vir verskeie beramers byvoorbeeld gemiddeldes, regressiekoëffisiënte, eenvoudige en parsiële korrelasie-koëffisiënte ondersoek deur Kish (1965) en daar is gevind dat die breuke oor die algemeen kleiner was as 0,005 en almal was kleiner as 0,02.

Die individuele waardes van $\bar{b}_j$ was voortdurend naby aan $\hat{b}$, in vergelyking met die steekproefvariansie van $b_j$ en $b_j'$. Die variansies wat bereken is vir $\bar{b}_j$ is dus goeie benaderings vir die variansie van $\hat{b}$.

Aangesien

$$\sum_j [(b_j - \hat{b})^2 + (b_j' - \hat{b})^2]/2k$$
$$= \overline{var_k}(\bar{b}_j) + \sum_j e_j^2/k$$

is die linkerkant 'n goeie beramer vir $\overline{var_k}(\bar{b}_j)$ en kan in die vervolg gebruik word aangesien Kish (1966) gevind het dat die term $\sum_j e_j^2/k$ baie klein is.

'n Ander belangrike vraag is of dit nodig is om $b_j$ sowel as sy komplement $b_j'$ te bereken. Is dit nie moontlik dat net $b_j$, dit wil sê $\sum_j (b_j - \hat{b})^2/k$, alleen gebruik word vir die beraming van $Var(\hat{b})$ nie?

In die lineêre geval is $(b_j' - b)^2 = (b_j - \hat{b})^2$ sodat die berekenings vir die komplement oorbodig sal wees. Hierdie gelykheid geld egter nie in die nie-lineêre geval nie, en $b_j$ met sy komplement $b_j'$ sal wel meer inligting verskaf en sodoende die presisie van die variansieberamer verbeter. Die vraag is of hierdie verbetering in die presisie die verdubbeling van die aantal berekenings regverdig.

Om 'n direkte antwoord ja of nee vir bogenoemde vraag te gee is nie moontlik nie, maar die volgende punte is hierby van belang.

In (3.4.7b) verdwyn die term $\sum_j e_j(b_j - b_j')$ wanneer die gemiddeldes vir die ooreenstemmende komplemente $(b_j - \hat{b})^2$ en $(b_j' - \hat{b})^2$ bereken word. Kish (1965) het die waardes van die relatiewe verskil van die twee beramers bereken, naamlik

$$
\begin{aligned}
&2\sum_j[(b_j - \hat{b})^2 - (b_j' - \hat{b})^2]/\sum_j(b_j - b_j')^2 \\
&= 4\sum_j(\bar{b}_j - \hat{b})(b_j - b_j')/\sum_j(b_j - b_j')^2
\end{aligned}
$$

Hulle gemiddelde waarde was versprei om 0,02. Dit dui op 'n klein verbetering in die presisie van die variansieberamers, klein genoeg om miskien slegs die gebruik van $\sum_j (b_j - \hat{b})^2/k$ te regverdig en sodoende die ekstra kostes verbonde daaraan om die beramers van die ooreenstemmende komplementêre halfsteekproewe te bereken, te bespaar.

### 3.4.4 <u>Die Uitsnit Herhaalde-Replikasiemetode van variansieberaming</u>.

Die Uitsnit Herhaalde-Replikasiemetode (JRR) van variansieberaming is 'n uitvloeisel van die Tuckey-uitsnitberamingsprosedure gekombineer met die

BRR metode. Voordat die JRR metode bespreek word is dit noodsaaklik om eers kortliks na die algemene Tuckey-uitsnitmetode van variansieberaming te kyk.

Beskou 'n ewekansige steekproef $\underline{X} = (X_1,\ldots,X_n)$ van grootte n uit 'n waarskynlikheidsverdeling met onbekende verdelingsfunksie F.

Dui dit aan as $\underline{X} \sim F$

Laat $(x_1,\ldots,x_n)$ 'n realisasie van $(X_1,\ldots,X_n)$ aandui, dan kan die steekproefgemiddelde $\bar{x} = \sum_{i=1}^{n} x_i/n$ bereken word. Die waarde van die standaardfout, wat 'n beraming van $\sigma_{\bar{x}}$ is, kan bereken word uit

$$s_{\bar{x}} = [1/(n(n-1)) \sum_{i=1}^{n} (x_i - \bar{x})^2]^{\frac{1}{2}} \ldots\ldots\ldots\ldots\ldots\ldots (3.4.8)$$

Vergelyking (3.4.8) het egter 'n groot tekortkoming. Dit geld naamlik net vir die gemiddeld $\bar{x}$ en kan nie op 'n ooglopende manier uitgebrei word na ander beramers soos byvoorbeeld die mediaan nie. Die Uitsnitmetode van variansieberaming is 'n metode wat hierdie probleem die hoof bied.

Die Uitsnitmetode in bogenoemde eenvoudige voorbeeld behels kortliks die volgende:

$$\text{Laat } \bar{x}_{(i)} = (n\bar{x} - x_i)/(n-1)$$
$$= 1/(n-1) \sum_{j \neq i} x_j \ldots\ldots\ldots\ldots\ldots\ldots (3.4.9)$$

die steekproefgemiddelde van die datastel wees met die i-de waarneming uitgesluit.

Laat $\bar{x}_{(.)} = \sum_{i=1}^{n} \bar{x}_{(i)}/n$ ..................................(3.4.10)

die gemiddelde van die gemiddeldes $\bar{x}_{(i)}$ gegee in (3.4.9), wees.

(Let op $\bar{x}_{(.)} = \bar{x}$)

Die Uitsnitberamer van die standaardfout word as volg uitgedruk.

$$\hat{\sigma}_j = [\{(n-1)/n\} \sum_{i=1}^{n} (\bar{x}_{(i)} - \bar{x}_{(.)})^2]^{\frac{1}{2}} \quad \text{........................(3.4.11)}$$

Vergelyking (3.4.~~10~~11) kan vir die gemiddelde as beramer as volg herlei word na (3.4.8):

$$\hat{\sigma}_j = [\{(n-1)/n\} \sum_{i=1}^{n} \{(n\bar{x} - x_i)/(n-1) - \sum_{i=1}^{n} (n\bar{x} - x_i)/(n(n-1))\}^2]^{\frac{1}{2}}$$

$$= [\{(n-1)/n\} \sum_i [(n\bar{x}/(n-1) ) - x_i/(n-1) - n^2\bar{x}/n(n-1) + \bar{x}/(n-1)]^2]^{\frac{1}{2}}$$

uit (3.4.9) en (3.4.10)

$$= [\{(n-1)/n\} 1/(n-1)^2 \sum_i (n\bar{x} - x_i - n\bar{x} + \bar{x})^2]^{\frac{1}{2}}$$

$$= [\{1/(n-1)n\} \sum_i (x_i - \bar{x})^2]^{\frac{1}{2}}$$

Die voordeel van (3.4.11) is dat dit maklik veralgemeen kan word na ander beramers. Laat hiertoe

$$\hat{\theta}_X \equiv \hat{\theta}(X_1, X_2, \ldots., X_n)$$

enige beramer wees wat gebaseer is op die steekproefelemente $X_1$ .... $X_n$.

Laat $\hat{\theta}_x = \hat{\theta}(x_1, \ldots, x_n)$ 'n realisasie van $\hat{\theta}_X$ wees.

Dui die realisasie van $\hat{\theta}_x$ met $x_i$ weggelaat, aan deur

$$\hat{\theta}_{(i)} = \hat{\theta}(x_1 \ldots\ldots x_{(i-1)}, x_{(i+1)} \ldots\ldots x_n)$$

en definieer

$\hat{\theta}_{(.)} = \hat{\theta}_{(i)}/n.$

Met die algemene Uitsnitmetode van beraming word nou bedoel dat $\bar{x}_{(.)}$ (kyk 3.4.10) van die Uitsnitmetode hierbo bespreek, vervang word deur $\hat{\theta}_{(.)}$. Hierdie algemene Uitsnitmetode kan toegepas word indien daar 'n ewekansige steekproef bestaan en belanggestel word in die tweede orde beramers van veral komplekse eerste orde beramers. Indien 'n komplekse steekproef, wat nie gebaseer is op die aanname van onafhanklikheid tussen waarnemings nie, beskou word, dan sal heeltemal verkeerde resultate verkry word indien gebruik gemaak word van die algemene Uitsnitmetode wat nie die kompleksiteit van die steekproefontwerp in ag neem nie.

As gevolg van die behoefte aan 'n variansieberamingsmetode wat gebruik kan word vir komplekse steekproefontwerpe, is die algemene Uitsnitmetode aangepas en het die Uitsnit Herhaalde-Replikasiemetode (JRR) ontstaan.

Hierdie metode word vervolgens kortliks uiteengesit:
Beskou 'n meerstadium komplekse steekproefontwerp wat bestaan uit H strata en waarby uit elke stratum slegs 2 PSE's met gelyke waarskynlikheid f getrek is. Laat S die hele steekproef voorstel. Om die JRR metode toe te pas moet daar eerstens H JRR-steekproewe uit die steekproef S gevorm word. Laat $J_i$ (i=1,2 ....H) die i-de steekproef voorstel wat uit S gevorm word. $J_i$ word verkry deur een van die twee PSE's in stratum i te dupliseer, die ander PSE in stratum i heeltemal weg te laat en die res van die steekproef, S, net so in te sluit.

$CJ_i$ (i=1,2 ....H) sal dan die komplementêre steekproef van $J_i$

voorstel en word gevorm deur die PSE's in die i-de stratum wat weggelaat en gedupliseer is om te ruil en weer eens die res van die steekproef net so in te sluit.

Die steekproefberamers van die variansie van die eerste orde beramers g(S) word bereken deur die volgende vergelykings:

<u>Beramer</u> <u>I</u> (<u>JRR</u>-<u>H</u>)

$$var_{JRR-H}(g(S)) = (1-f)\sum_{i=1}^{H}(g(J_i)-g(S))^2 \quad \text{..................(3.4.12)}$$

<u>Beramer</u> <u>II</u> (<u>JRR</u>-<u>C</u>)

$$var_{JRR-C}(g(S)) = (1-f)\sum_{i=1}^{H}(g(CJ_i)-g(S))^2 \quad \text{.................(3.4.13)}$$

<u>Beramer</u> <u>III</u> (<u>JRR</u>-<u>S</u>)

$$var_{JRR-S}(g(S)) = [var_{JRR-H}(g(S))+ var_{JRR-C}(g(S))]/2 \quad \text{.....(3.4.14)}$$

<u>Beramer</u> <u>IV</u> (<u>JRR</u>-<u>D</u>)

$$var_{JRR-D}(g(S)) = (1-f)/4 \sum_{i=1}^{H}(g(J_i) - g(CJ_i))^2 \quad \text{............(3.4.15)}$$

Die lineêriteit wat geassosieer word met die JRR beramers is van die vorm

$$1/2H\sum_{i} (g(J_i) + g(CJ_i)) \text{.................................(3.4.16)}$$

JRR-H en JRR-C besit dieselfde verwagte waarde as JRR-S en is goekoper (in terme van rekenaartyd) om te bereken as JRR-S, maar JRR-H en JRR-C is minder presiese terme as JRR-S. Vir die res van die hoofstuk sal JRR-H en JRR-C gevolglik uitgelaat word.

### 3.4.5 Uitbreidings op BRR en JRR

Twee PSE's per stratum is 'n basiese steekproefontwerp. Dit voldoen aan die vereistes dat variansieberekening moontlik moet wees met meer as een onafhanklike herhaling per stratum. Indien daar in die praktyk belang gestel word in 'n fyner gestratifiseerde ontwerp dan sal 'n modifikasie van die basiese steekproefontwerp nodig wees. Beskou die volgende seleksiemetodes, een PSE per stratum, óf beheerde seleksie óf sistematiese steekproeftrekking. Die variansie onder hierdie seleksiemetodes kan gewoonlik bereken word deur strata saam te voeg om die basiese gepaarde seleksiemodel te gee. Tot dusver is BRR bespreek in die raamwerk van 'n halveringstegniek wat gebaseer is op twee PSE's per stratum. Die BRR metode is egter nie beperk tot gepaarde seleksies of tot halfsteekproef herhalings nie.

In sommige steekproefontwerpe mag daar dalk te veel strata wees om met gemak te kan hanteer. Daar kan byvoorbeeld 800 PSE's wees in 400 strata, maar 100 gebalanseerde herhalings kan voldoende wees om BRR op toe te pas indien presisie en koste teen mekaar opgeweeg word (Kish (1965)). Indien enigsins moontlik moet probeer word om situasies te vermy waar die groottes van die strata grootliks varieer aangesien dit 'n nadelige effek op die beraming van die variansie het.

Indien die aantal PSE's per stratum 'n ewe getal is, kan dit altyd op ewekansige wyse in twee ewegroot groepe verdeel word. Die BRR en JRR metodes kan dan toegepas word op die PSE's wat in pare afgepaar is. Indien die strata uit 'n onewe aantal PSE's bestaan kan die strata paarsgewys gekombineer word, om 'n ewe aantal PSE's per stratum te verkry (solank dit sinvol is ), alhoewel dit die variansie ietwat vergroot.

Ander moontlike skemas kan uitgewerk word, gebaseer op die idee dat vir enige aantal seleksies n geld:

$$\sum_{j}(y_j - \bar{y})^2/(n-1) = \sum_{i<j}(y_i - y_j)^2/n(n-1)$$

Bogenoemde eienskap kan gebruik word indien die aantal strata klein is en die aantal PSE's per stratum groot is. Stel artifisiële pare saam deur hulle ewekansig te kies uit die stratum en pas dan BRR en JRR op hierdie pare toe (Kish & Frankel (1972)).

Indien daar byvoorbeeld vier PSE's per stratum is, kan die PSE's per stratum op drie verskillende maniere afgepaar word. (1 + 2; 1 + 3; 1 + 4; die ander drie pare sal dan die komplementêre halfsteekproewe verteenwoordig). BRR en JRR kan dan op die bogenoemde gekose pare toegepas word.

Die doel van die BRR-metode is om uit k onafhanklike steekproewe 'n tweede orde beramer van die gemiddelde van 'n stel van k onafhanklike eerste orde beramers $b_1$, ... $b_k$ te verkry. Van die onafhanklike beramers $b_i$ word die gemiddelde $\bar{b} = \sum_i b_i/k$ verkry waarvoor die gemiddelde variansie (tweede orde beraming) bereken kan word uit die volgende formule (kyk Deming 1960):

$$\operatorname{var}(\bar{b}) = \sum_i (b_i - \bar{b})^2/k(k-1)$$

Om te bepaal hoe groot k moet wees is nogal moeilik. Indien k te klein is, is die beramings van die variansie so onstabiel en veranderlik dat dit basies nutteloos is. Sommige navorsers gebruik selfs so min as vier halfsteekproewe. Jones (1956) gee redes waarom hy 25 tot 50 halfsteekproewe aanbeveel. Ander persone verkies weer meer as 50 halfsteekproewe in meeste gevalle (Kish 1965).

'n Groot aantal halfsteekproewe k het egter ook sekere probleme nl:

a) Die eenvoud van die metode gaan verlore.

b) In meeste gevalle gaan te veel van die stratifikasie van die steekproef verlore.

c) Deur k te vergroot verswak die inferensie na $\hat{b}$, wat gebaseer is op die hele steekproef van $\bar{b} = \sum_i b_i/k$, die gemiddeld van die statistieke gebaseer op die aantal halfsteekproewe.

Hoe nader $\bar{b}$ aan $\hat{b}$ is, en $\operatorname{var}(\bar{b})$ aan $\operatorname{var}(\hat{b})$ is, hoe beter is die BRR en JRR-variansieberamers. Indien die aantal halfsteekproewe te veel word sal $\bar{b}$ verder weg beweeg van $\hat{b}$ en sal $\operatorname{var}(\bar{b})$ 'n swakker beramer van $\operatorname{var}(\hat{b})$ word.

Standaardfoute is van groot belang in steekproefopnames maar BRR kan ook vir ander statistieke gebruik word. Een ooglopende uitbreiding is na die toets vir betekenisvolheid van die nulhipotese (kyk Verma 1982).

Die BRR metode is baie buigbaar, maar die vereiste van onafhanklike halfsteekproewe in die ontwerp is noodsaaklik.

### 3.4.6 Toepassing van die BRR- en JRR- metodes van variansieberaming

Vir al drie die variansieberamingsmetodes (Taylor, BRR en JRR) hou die berekeningskoste direk verband met die aantal strata en nie met die totale steekproefgrootte nie. By al drie die metodes van variansieberaming hoef die rekenaar slegs eenmaal deur die individuele rekords (USE's) te gaan. Tydens hierdie eerste lopie word benodigde statistieke ( y, $y^2$, yx) vir elkeen van die 2 x H PSE's in die hele steekproef en vir alle veranderlikes in die steekproef bereken. Hierdie statistieke vir al die PSE's verskaf die intermediêre statistieke vir die hele steekproef. Enige verdere berekenings word dan gedoen vanuit hierdie intermediêre statistieke.

Indien die BRR-metode van variansieberaming gebruik word, is slegs een lopie oor die 2H intermediêre statistieke nodig om die halfsteekproewe en die komplementêre halfsteekproewe te vorm. Vervolgens word die intermediêre statistieke vir die halfsteekproewe gevorm deur die intermediêre statistieke van elkeen van die getrekte PSE's uit die twee PSE's in elk van die H strata, bymekaar te tel. Die komplementêre halfsteekproef word dan gevorm deur die intermediêre statistiek van die halfsteekproef af te trek vanaf die intermediêre statistiek vir die totale steekproef. Die benodigde eerste orde beramers g(S), $g(H_i)$ en $g(C_i)$ word dan bereken vanuit die intermediêre statistieke van die totale steekproef, die i-de halfsteekproef en die i-de komplementêre halfsteekproef. Hierdie terme word dan gemanipuleer soos in (3.4.3) en (3.4.4) om die BRR-S of BRR-D beramers te vorm.

Die berekening wat vir die JRR-beramers benodig word, is basies dieselfde as die berekening wat vir die BRR-beramers nodig is. Die enigste verskil tussen die twee metodes kom in by die vorming van die steekproewe en komplementêre steekproewe, uit die hele steekproef. Om byvoorbeeld die i-de JRR-steekproef te vorm, word die intermediêre statistiek van een van die PSE's in die i-de stratum afgetrek van die intermediêre statistiek van die totale steekproef, en die intermediêre statistiek van die ander PSE in die stratum word dan by die resultaat bygetel. Deur nou die twee PSE's se intermediêre statistieke wat afgetrek en bygetel is in die i-de stratum om te ruil word die prosedure herhaal om die i-de komplementêre JRR-steekproef te vorm.

Om die variansies van relatief eenvoudige eerste orde beramers te bereken is die Taylor-linearisasiemetode van variansieberaming die populêrste. Hierdie eerste orde beramers sluit bv. verhoudings of gemiddeldes en verskille van verhoudings of gemiddeldes in. Sodra die berekenings van die parsiële afgeleides van die steekproefberamers meer tydrowend word as om die beramers van die halfsteekproewe en komplementêre halfsteekproewe te bereken verloor die Taylor-linearisasiemetode sy tydvoordeel (in terme van rekenaartyd).

Vir meer komplekse eerste orde beramers word die uitdrukking van die parsiële afgeleides in geslote vorm matematies te ingewikkeld om te bereken en dan moet BRR of JRR gebruik word.

Aangesien die afgeleides vir parsiële- en meervoudige korrelasie-koëffisiënte te ingewikkeld is, besit die BRR- en JRR-metodes 'n definitiewe voordeel bo die Taylor-linearisasiemetode.

## 3.5 DIE METODE WAT DEUR BRR GEBRUIK WORD OM ORTOGONALE MATRIKSE TE BEREKEN.

Om 'n ortogonale matriks te stoor in die rekenaargeheue kan problematies wees aangesien dit baie stoorruimte in die rekenaar benodig. Vir 'n matriks van grootte 256 by 256 is 65536 grepe nodig om die matriks te stoor en huidiglik is dit te veel ruimte om deur 'n enkele program in beslag geneem te word. Robert S. Jewett het 'n program ontwikkel wat 'n ortogonale matriks vorm met die eienskap dat die hele matriks gegenereer word vanaf die eerste kolom daarvan. Dit los die stoorruimte probleem op, aangesien slegs een kolom van die matriks in die rekenaargeheue gestoor hoef te word.

Die metode wat gevolg word om hierdie ortogonale matriks te genereer word gegee in die artikel 'Constructing Orthogonal Replications for Variance Estimation (Gurney & Jewett (1975)). Hierdie metode genereer 'n ortogonale matriks deur twee parameters p en n te gebruik waarby p 'n priemgetal en n enige positiewe heelgetal is. Die rekenaarprogram wat deur Jewett ontwikkel is gebruik die waarde p=2 en kan enige heelgetalwaarde vir n tot en met 12 hanteer.

Vervolgens word 'n algemene beskrywing van die stappe wat gevolg word om 'n $2^n$ by $2^n$ ortogonale matriks te genereer, gegee:

Stap 1: Beskou 'n ring polinome wat koëffisiënte in die liggaam van heelgetalle modulo 2 besit. Vind dan 'n polinoom van graad n wat nie in 'n laer graad voorgestel kan word nie.

<u>Byvoorbeeld</u>

$X^3 + X + 1$ kan nie verder onderverdeel word nie.

Stap 2: Hierdie ring van polinome vorm 'n 'Galios' liggaam wat $2^n$ elemente bevat.

'n Generator van hierdie liggaam is 'n polinoom $p(x)^k=1$ as en slegs as $k=2^n-1$.

'n Generator vir die liggaam moet nou gevind word.

Byvoorbeeld

$X + 1$

Stap 3: Skryf die nie-nul elemente van die liggaam as magte van die generator.

Byvoorbeeld

$(X + 1)^1 = X + 1$

$(X + 1)^2 = X^2 + 1$

$(X + 1)^3 = X^2$

$(X + 1)^4 = X^2 + X + 1$

$(X + 1)^5 = X$

$(X + 1)^6 = X^2 + X$

$(X + 1)^7 = 1$

Stap 4. In elkeen van die bogenoemde vergelykings is die koëffisiënt van die ene óf een óf nul. Vorm nou 'n vektor van lengte $2^n-1$ wat bestaan uit hierdie koëffisiënte van die ene in die volgorde wat hulle voorkom.

Byvoorbeeld

Die vektor van lengte 7 is (1,1,0,1,0,0,1).

Hierdie vektor is die een wat deur die rekenaarprogram gegenereer word, vir n=3. Soos reeds genoem kan die program enige waarde van n tot en met 12 inlees en die toepaslike vektor van lengte $2^n-1$ bereken.

Hierdie vektor (generator) word dan as volg gebruik om 'n ortogonale matriks te genereer.

Stap 5. Laat hierdie vektor die eerste kolom van die matriks vorm. Die volgende kolom word gevorm deur elke element van die oorspronklike kolom een plek op(boontoe) te skuif en die heel boonste element na heel onder te skuif. Hierdie proses word voortgesit totdat 'n $2^n -1$ by $2^n -1$ matriks gevind word.

<u>Byvoorbeeld</u>

| | | | | | | |
|---|---|---|---|---|---|---|
| 1 | 1 | 0 | 1 | 0 | 0 | 1 |
| 1 | 0 | 1 | 0 | 0 | 1 | 1 |
| 0 | 1 | 0 | 0 | 1 | 1 | 1 |
| 1 | 0 | 0 | 1 | 1 | 1 | 0 |
| 0 | 0 | 1 | 1 | 1 | 0 | 1 |
| 0 | 1 | 1 | 1 | 0 | 1 | 0 |
| 1 | 1 | 1 | 0 | 1 | 0 | 0 |

Stap 6. Verander vervolgens al die nulle in die matriks na -1. Voeg 'n ry van -1,-1..-1 heel onder by en 'n ekstra kolom van ene heel regs. Hierdie matriks vorm dan 'n ortogonale matriks van grootte $2^n$ by $2^n$.

<u>Byvoorbeeld</u>

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1 | 1 | -1 | 1 | -1 | -1 | 1 | 1 |
| 1 | -1 | 1 | -1 | -1 | 1 | 1 | 1 |
| -1 | 1 | -1 | -1 | 1 | 1 | 1 | 1 |
| 1 | -1 | -1 | 1 | 1 | 1 | -1 | 1 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| -1 | -1 | 1 | 1 | 1 | -1 | 1 | 1 |
| -1 | 1 | 1 | 1 | -1 | 1 | -1 | 1 |
| 1 | 1 | 1 | -1 | 1 | -1 | -1 | 1 |
| -1 | -1 | -1 | -1 | -1 | -1 | -1 | 1 |

Uit hierdie laaste twee stappe is dit duidelik dat indien 'n vektor van lengte $2^n-1$ in die rekenaarkerngeheue gestoor word is dit maklik om enige ander element van die matriks te bereken. Hier is 'n paar vektore wat bereken is vir sekere waardes van n nl:

Vir n = 3 is die vektor gelyk aan 1 1 0 1 0 0 1

Vir n = 4 is die vektor gelyk aan 1 1 1 0 0 0 1
0 0 1 1 0 1 0 1

Vir n = 5 is die vektor gelyk aan 1 1 1 1 0 1 0
0 0 1 0 0 1 0 1 0 1 1 0 0 0 0 1 1 1
0 0 1 1 0 1

Die rekenaarprogram wat hierdie generators bereken word gegee in bylae A.

## 3.6 EMPIRIESE ONDERSOEKE

### 3.6.1 Die BRRJRR program

Na aanleiding van die teorie bespreek in hierdie hoofstuk is besluit om 'n rekenaarprogram te skryf. 'n BRRJRR-program is vervolgens by die RGN ontwikkel om die BRR- en JRR-variansieberamers vir komplekse meerstadiumsteekproewe te bereken. Hierdie program is huidiglik beperk tot gemiddeldes, proporsies en verhoudings. Die program BRRJRR kan die onderskeie statistieke bereken vir komplekse meerstadiumsteekproewe met ongelyke seleksiewaarskynlikhede en of responsproporsies sowel as vir selfwegende komplekse steekproewe.

Vir beide die berekening van die BRR- en JRR-statistieke word daar deur middel van die rekenaar slegs een keer deur die individuele rekords (elemente) gegaan. Tydens hierdie lopie word die som, som van kwadrate en som van kruisprodukte vir elk van die 2H PSE's , waaruit die hele steekproef bestaan, bereken. Hierdie totale wat vir elk van die PSE's verkry word, staan bekend as die intermediêre statistieke vir die hele steekproef. Alle verdere berekenings word vanuit hierdie totale bereken. Die program BRRJRR bereken nie hierdie intermediêre statistieke nie aangesien die program CLUSTERS reeds hierdie statistieke bereken. Voordat die program BRRJRR gebruik word, moet die intermediêre statistieke vanaf die program CLUSTERS verkry word.

Wanneer die BRR-metode van variansieberaming benodig word, word die intermediêre statistieke vanaf die program CLUSTERS verkry, daarna word Jewett se program, wat in BRRJRR geïnkorporeer is, uitgevoer om die generator (dit is die eerste kolom) vir die berekening van die ortogonale matriks te kry. Vervolgens word al die elemente van die ortogonale matriks bereken en die halfsteekproewe daarvolgens getrek. Direk daarna word die statistieke vir elk van die halfsteekproewe bereken en vervang in die BRR vergelykings 3.4.1 tot 3.4.4.

Vir die gebruik van die JRR-metode van variansieontleding word weereens van die CLUSTERS program gebruik gemaak vir die berekening van die intermediêre statistieke. Daarna word deur middel van 'n ewekansige getal generator, getalle ewekansig gegenereer om die JRR steekproewe en komplementêre steekproewe te verkry. Sodra die JRR-steekproewe en komplementêre steekproewe getrek is en die benodigde statistieke uit die intermediêre statistieke bereken is, word berekende waardes in die onderskeie JRR vergelykings 3.4.12 tot 3.4.15 vervang.

Die BRRJRR-program is 'n interaktiewe program en tydens die uitvoering van die program word die volgende vrae gevra:

i) U MATRIKS SAL 2**A BY 2**A WEES , WAT MOET DIE WAARDE VAN A WEES?

ii) WIL U SLEGS N BEPERKTE UITVOER HE? JA=1,NEE=0.

iii) WIL U DIE BRR-STATISTIEK BEREKEN? JA=1,NEE=0.

Indien die antwoord op vraag (iii) JA is volg vraag (iv) anders volg vraag (v)

iv) MAAK U GEBRUIK VAN GEWIGTE? JA=1,NEE=0.

v) WIL U DIE JRR-STATISTIEK BEREKEN? JA=1,NEE=0.

Indien die antwoord op vraag (v) JA is volg vraag (vi) anders eindig die program

vi) MAAK U GEBRUIK VAN GEWIGTE? JA=1,NEE=0.

Indien die antwoord op vraag (ii) NEE is, d.w.s indien 'n volledige uitvoer vir BRR sowel as vir die JRR variansieberamingsmetodes aangevra word sal die uitvoer as volg wees:

<u>EERSTE DEEL VAN UITVOER</u>:

N = 3

GENERATOR T=

1 1 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

COLUMN 1= 1101001

*** THE MATRIX IS ORTHOGONAL ***

Hierdie eerste deel van die uitvoer is die uitvoer van die gedeelte van die program van Jewett wat geïnkorporeer is in die BRRJRR program. In hierdie geval is die waarde van n=3. Dit wil sê 'n $2^3$ by $2^3$ ortogonale matriks kan vanaf hierdie generator verkry word.

<u>TWEEDE DEEL VAN UITVOER</u>:

INRERMEDIÊRE STATISTIEKE

| | | | |
|---|---|---|---|
| 1.00 | 30.00 | 10.00 | 4.00 |
| -1.00 | 16.22 | 13.51 | 5.00 |
| 1.00 | 5.55 | 5.55 | 5.00 |
| -1.00 | 20.00 | 6.67 | 4.00 |
| 1.00 | 17.78 | 11.11 | 5.00 |
| -1.00 | 17.78 | 17.78 | 8.00 |
| 1.00 | 10.00 | 10.00 | 5.00 |
| -1.00 | 10.00 | 6.00 | 3.00 |
| 1.00 | 26.67 | 20.00 | 3.00 |
| -1.00 | 9.37 | 9.37 | 3.00 |
| 1.00 | 18.52 | 14.82 | 4.00 |

| | | | |
|---|---|---|---|
| -1.00 | 25.00 | 16.67 | 6.00 |
| 1.00 | 76.66 | 16.66 | 5.00 |
| -1.00 | 83.33 | 16.67 | 2.00 |
| 1.00 | 43.33 | 16.66 | 5.00 |
| -1.00 | 9.26 | 9.26 | 5.00 |

Hierdie deel van die uitvoer word verkry uit die program CLUSTERS. Bostaande intermediêre statistieke verteenwoordig die totale verkry per PSE. Hierdie spesifieke voorbeeld bestaan uit 16 PSE's en 8 strata. Elke stratum in die steekproef bestaan uit twee PSE's en hulle word deur 'n 1 of -1 geïdentifiseer. Die tweede kolom bevat die totale van 'n spesifieke veranderlike vir die USE's in elke PSE. Die derde kolom bevat die totale gewigte. Die vierde kolom gee die totale aantal USE's in elke PSE aan.

DERDE DEEL VAN DIE UITVOER:

VIR DIE HELE STEEKPROEF:

GEWEEGDE SOMY = 2.0897

# USE = 72.0000

VIERDE DEEL VAN DIE UITVOER:

```
COL NR    1    1 1-1 1-1-1 1-1
    1.00   30.00     4.00
    1.00    5.55     5.00
   -1.00   17.78     8.00
    1.00   10.00     5.00
   -1.00    9.37     3.00
   -1.00   25.00     6.00
    1.00   76.66     5.00
   -1.00    9.26     5.00

COL NR    2    1-1 1-1-1 1 1-1
    1.00   30.00     4.00
   -1.00   20.00     4.00
    1.00   17.78     5.00
```

```
   -1.00   10.00    3.00
   -1.00    9.37    3.00
    1.00   18.52    4.00
    1.00   76.66    5.00
   -1.00    9.26    5.00

COL NR   3  -1 1-1-1 1 1 1-1
   -1.00   16.22    5.00
    1.00    5.55    5.00
   -1.00   17.78    8.00
   -1.00   10.00    3.00
    1.00   26.67    3.00
    1.00   18.52    4.00
    1.00   76.66    5.00
   -1.00    9.26    5.00

COL NR   4   1-1-1 1 1-1-1-1
    1.00   30.00    4.00
   -1.00   20.00    4.00
   -1.00   17.78    8.00
    1.00   10.00    5.00
    1.00   26.67    3.00
   -1.00   25.00    6.00
   -1.00   83.33    2.00
   -1.00    9.26    5.00

COL NR   5  -1-1 1 1-1 1 1-1
   -1.00   16.22    5.00
   -1.00   20.00    4.00
    1.00   17.78    5.00
    1.00   10.00    5.00
   -1.00    9.37    3.00
    1.00   18.52    4.00
    1.00   76.66    5.00
   -1.00    9.26    5.00

COL NR   6  -1 1 1-1 1-1-1-1
   -1.00   16.22    5.00
    1.00    5.55    5.00
    1.00   17.78    5.00
   -1.00   10.00    3.00
    1.00   26.67    3.00
   -1.00   25.00    6.00
   -1.00   83.33    2.00
   -1.00    9.26    5.00

COL NR   7   1 1-1 1-1-1-1-1
    1.00   30.00    4.00
    1.00    5.55    5.00
   -1.00   17.78    8.00
    1.00   10.00    5.00
   -1.00    9.37    3.00
   -1.00   25.00    6.00
   -1.00   83.33    2.00
   -1.00    9.26    5.00

COL NR   8   1 1 1 1 1 1 1 1
```

| | | |
|---|---|---|
| 1.00 | 30.00 | 4.00 |
| 1.00 | 5.55 | 5.00 |
| 1.00 | 17.78 | 5.00 |
| 1.00 | 10.00 | 5.00 |
| 1.00 | 26.67 | 3.00 |
| 1.00 | 18.52 | 4.00 |
| 1.00 | 76.66 | 5.00 |
| 1.00 | 43.33 | 5.00 |

Aangesien daar 8 strata is, is 'n 8 by 8 ortogonale matriks geskep. Eerstens word die kolomnommers aangedui, dan die onderskeie kolomme wat aantoon watter van die PSE's uit elke stratum in die halfsteekproef verteenwoordig word. Die waardes van die intermediêre statistieke word in kolomme twee en drie aangedui.

VYFDE DEEL VAN UITVOER:

GEBALANSEERDE HERHAALDE REPLIKASIE UITVOER

```
HALFSP NR    1:
   SOMY GEWEEG GEM=          1.9270
   SOM1 =                   41.0000
KOMPL  NR    1:
   SOMCY GEWEEG GEM=         2.2368
   SOMC1 =                  31.0000
HALFSP NR    2:
   SOMY GEWEEG GEM=          2.2838
   SOM1 =                   33.0000
KOMPL  NR    2:
   SOMCY GEWEEG GEM=         1.9504
   SOMC1 =                  39.0000
HALFSP NR    3:
   SOMY GEWEEG GEM=          1.7442
   SOM1 =                   38.0000
KOMPL  NR    3:
   SOMCY GEWEEG GEM=         2.4581
   SOMC1 =                  34.0000
HALFSP NR    4:
   SOMY GEWEEG GEM=          2.0742
   SOM1 =                   37.0000
KOMPL  NR    4:
   SOMCY GEWEEG GEM=         2.1075
   SOMC1 =                  35.0000
HALFSP NR    5:
   SOMY GEWEEG GEM=          1.9454
   SOM1 =                   36.0000
KOMPL  NR    5:
   SOMCY GEWEEG GEM=         2.2104
```

```
    SOMC1 =                 36.0000
HALFSP NR    6:
    SOMY GEWEEG GEM=         1.9622
    SOM1 =                  34.0000
KOMPL   NR    6:
    SOMCY GEWEEG GEM=        2.2132
    SOMC1 =                 38.0000
HALFSP NR    7:
    SOMY GEWEEG GEM=         1.9967
    SOM1 =                  38.0000
KOMPL   NR    7:
    SOMCY GEWEEG GEM=        2.1738
    SOMC1 =                 34.0000
HALFSP NR    8:
    SOMY GEWEEG GEM=         2.1804
    SOM1 =                  36.0000
KOMPL   NR    8:
    SOMCY GEWEEG GEM=        1.9906
    SOMC1 =                 36.0000
```

In hierdie deel van die uitvoer word eerstens (indien met ongelyke seleksiewaarskynlikhede gewerk word) die geweegde gemiddelde per halfsteekproef gegee en daarna word aangetoon hoeveel USE's in daardie halfsteekproef is.

<u>SESDE</u> <u>DEEL</u> <u>VAN</u> <u>UITVOER</u>:

```
DIE BRR-H STATISTIEK IS          0.12190

DIE BRR-C STATISTIEK IS          0.11827

DIE BRR-S STATISTIEK IS          0.12010

DIE BRR-D STATISTIEK IS          0.11981
```

<u>SEWENDE</u> <u>DEEL</u> <u>VAN</u> <u>UITVOER</u>:

UITSNIT HERHAALDE REPLIKASIE UITVOER

```
HALFSP NR    1:
    SOMY GEWEEG GEM =         1.9863
               SOM1 =        73.0000
KOMPL   NR    1:
    SOMCY GEWEEG GEM=        2.1968
              SOMC1 =       71.0000
HALFSP NR    2:
    SOMY GEWEEG GEM =         2.0291
```

```
              SOM1 =        73.0000
KOMPL  NR   2:
   SOMCY GEWEEG GEM=       2.1497
             SOMC1 =      71.0000
HALFSP NR   3:
   SOMY GEWEEG GEM =        2.1615
              SOM1 =       69.0000
KOMPL  NR   3:
   SOMCY GEWEEG GEM=       2.0225
             SOMC1 =      75.0000
HALFSP NR   4:
   SOMY GEWEEG GEM =        2.0489
              SOM1 =       74.0000
KOMPL  NR   4:
   SOMCY GEWEEG GEM=       2.1322
             SOMC1 =      70.0000
HALFSP NR   5:
   SOMY GEWEEG GEM =        2.0665
              SOM1 =       72.0000
KOMPL  NR   5:
   SOMCY GEWEEG GEM=       2.1156
             SOMC1 =      72.0000
HALFSP NR   6:
   SOMY GEWEEG GEM =        2.0766
              SOM1 =       70.0000
KOMPL  NR   6:
   SOMCY GEWEEG GEM=       2.1026
             SOMC1 =      74.0000
HALFSP NR   7:
   SOMY GEWEEG GEM =        2.0566
              SOM1 =       75.0000
KOMPL  NR   7:
   SOMCY GEWEEG GEM=       2.1228
             SOMC1 =      69.0000
HALFSP NR   8:
   SOMY GEWEEG GEM =        1.9935
              SOM1 =       72.0000
KOMPL  NR   8:
   SOMCY GEWEEG GEM=       2.1791
             SOMC1 =      72.0000
```

In hierdie deel van die uitvoer word eerstens (indien met ongelyke seleksiewaarskynlikhede gewerk word) die geweegde gemiddelde per JRR-steekproef gegee en daarna word aangetoon hoeveel USE's in daardie JRR-steekproef is.

<u>AGSTE DEEL VAN UITVOER</u>:

DIE JRR-H STATISTIEK IS          0.12702

DIE JRR-C STATISTIEK IS          0.12512

DIE JRR-S STATISTIEK IS 0.12607

DIE JRR-D STATISTIEK IS 0.12602

Indien die antwoord op vraag (ii) JA is, d.w.s. indien slegs 'n beperkte uitvoer vir beide BRR en JRR benodig word, dan sal slegs die volgende uittreksels van die volledige uitvoer gegee word.

eerste deel van uitvoer

sesde deel van uitvoer

agste deel van uitvoer

### 3.6.2 Analises

Dieselfde datastel wat in die praktiese gedeelte van hoofstuk 2 vir analises gebruik is word weer eens beskou. In hierdie analises word van die BRRJRR program en die program CLUSTERS gebruik gemaak om die onderskeie variansies vir gemiddeldes van verskillende veranderlikes met mekaar te vergelyk. Die uitvoer van die analises volg in die volgende tabelle 3.1 tot 3.3.

Dit is baie duidelik dat daar feitlik geen verskil tussen die waardes van die Taylor-, BRR- of JRR-statistieke is nie. Die rede hiervoor is dat gemiddeldes wat lineêre statistieke is, beskou is. Die geringe verskille wat wel voorkom kan toegeskryf word aan benaderingsverskille.

BLANKES TABEL 3.1

| | CLUSTERS - UITVOER | | | BRRJRR - UITVOER | | |
|---|---|---|---|---|---|---|
| | R | SE | SER | R | BRR | JRR |
| OUD | 43.225 | 0.745 | 0.597 | 43.227 | 0.779 | 0.745 |
| INK | 6.472 | 0.205 | 0.099 | 6.473 | 0.214 | 0.205 |
| HUIS | 3.126 | 0.092 | 0.056 | 3.127 | 0.090 | 0.092 |
| KIND | 0.672 | 0.051 | 0.035 | 0.672 | 0.048 | 0.051 |
| BEV1 | 0.831 | 0.017 | 0.013 | 0.831 | 0.015 | 0.017 |
| BEV2 | 0.899 | 0.011 | 0.011 | 0.899 | 0.012 | 0.011 |
| BEV3 | 0.855 | 0.014 | 0.012 | 0.855 | 0.015 | 0.014 |
| STEM | 0.832 | 0.018 | 0.013 | 0.832 | 0.018 | 0.018 |
| POL | 0.617 | 0.021 | 0.017 | 0.617 | 0.021 | 0.021 |
| SAL1 | 0.572 | 0.022 | 0.018 | 0.572 | 0.019 | 0.022 |
| SAL2 | 0.961 | 0.009 | 0.007 | 0.961 | 0.009 | 0.009 |
| SAL3 | 0.670 | 0.025 | 0.017 | 0.670 | 0.024 | 0.025 |
| WERK | 0.821 | 0.019 | 0.014 | 0.821 | 0.019 | 0.019 |
| KOS | 0.265 | 0.027 | 0.016 | 0.265 | 0.026 | 0.027 |
| HUUR | 0.563 | 0.026 | 0.018 | 0.563 | 0.023 | 0.026 |
| TRAN | 0.583 | 0.030 | 0.017 | 0.583 | 0.027 | 0.030 |
| VERH | 0.610 | 0.024 | 0.017 | 0.610 | 0.023 | 0.024 |
| REL1 | 0.931 | 0.012 | 0.009 | 0.931 | 0.011 | 0.012 |
| REL2 | 0.878 | 0.020 | 0.012 | 0.878 | 0.019 | 0.020 |
| HUW1 | 0.882 | 0.014 | 0.011 | 0.882 | 0.015 | 0.014 |
| GLUK | 0.908 | 0.017 | 0.010 | 0.908 | 0.018 | 0.017 |
| OUER | 0.761 | 0.024 | 0.015 | 0.761 | 0.025 | 0.024 |
| TOEK | 0.880 | 0.021 | 0.012 | 0.880 | 0.021 | 0.021 |

ASIËRS TABEL 3.2

| | CLUSTERS - UITVOER | | | BRRJRR - UITVOER | | |
|---|---|---|---|---|---|---|
| | R | SE | SER | R | BRR | JRR |
| OUD | 36.125 | 0.409 | 0.384 | 36.125 | 0.433 | 0.409 |
| INK | 4.488 | 0.144 | 0.066 | 4.488 | 0.155 | 0.144 |
| HUIS | 5.152 | 0.089 | 0.065 | 5.152 | 0.085 | 0.089 |
| KIND | 1.410 | 0.041 | 0.040 | 1.410 | 0.036 | 0.041 |
| BEV1 | 0.894 | 0.017 | 0.009 | 0.894 | 0.018 | 0.017 |
| BEV2 | 0.792 | 0.028 | 0.011 | 0.792 | 0.028 | 0.028 |
| BEV3 | 0.737 | 0.029 | 0.012 | 0.737 | 0.028 | 0.029 |
| STEM | 0.244 | 0.030 | 0.012 | 0.244 | 0.029 | 0.030 |
| POL | 0.438 | 0.036 | 0.014 | 0.438 | 0.034 | 0.036 |
| SAL1 | 0.344 | 0.024 | 0.013 | 0.344 | 0.023 | 0.024 |
| SAL2 | 0.809 | 0.029 | 0.011 | 0.809 | 0.028 | 0.029 |
| SAL3 | 0.683 | 0.019 | 0.013 | 0.683 | 0.019 | 0.019 |
| WERK | 0.721 | 0.025 | 0.013 | 0.721 | 0.023 | 0.025 |
| KOS | 0.147 | 0.029 | 0.010 | 0.147 | 0.031 | 0.029 |
| HUUR | 0.444 | 0.033 | 0.014 | 0.444 | 0.034 | 0.033 |
| TRAN | 0.372 | 0.047 | 0.014 | 0.372 | 0.047 | 0.047 |
| VERH | 0.473 | 0.024 | 0.014 | 0.473 | 0.022 | 0.024 |
| REL1 | 0.977 | 0.006 | 0.004 | 0.977 | 0.007 | 0.006 |
| REL2 | 0.955 | 0.009 | 0.006 | 0.956 | 0.009 | 0.009 |
| HUW1 | 0.820 | 0.017 | 0.011 | 0.820 | 0.017 | 0.017 |
| GLUK | 0.920 | 0.009 | 0.008 | 0.920 | 0.009 | 0.009 |
| OUER | 0.748 | 0.018 | 0.012 | 0.748 | 0.017 | 0.018 |
| TOEK | 0.804 | 0.015 | 0.011 | 0.804 | 0.016 | 0.015 |

KLEURLINGE TABEL 3.3

| | CLUSTERS - UITVOER | | | BRRJRR - UITVOER | | |
|---|---|---|---|---|---|---|
| | R | SE | SER | R | BRR | JRR |
| OUD | 39.127 | 0.565 | 0.454 | 39.127 | 0.578 | 0.565 |
| INK | 3.605 | 0.114 | 0.073 | 3.605 | 0.110 | 0.114 |
| HUIS | 5.391 | 0.114 | 0.080 | 5.391 | 0.125 | 0.114 |
| KIND | 1.449 | 0.062 | 0.045 | 1.449 | 0.062 | 0.062 |
| BEV1 | 0.770 | 0.022 | 0.014 | 0.770 | 0.021 | 0.022 |
| BEV2 | 0.724 | 0.031 | 0.014 | 0.724 | 0.030 | 0.031 |
| BEV3 | 0.632 | 0.028 | 0.016 | 0.632 | 0.026 | 0.028 |
| STEM | 0.225 | 0.023 | 0.014 | 0.225 | 0.022 | 0.023 |
| POL | 0.311 | 0.025 | 0.015 | 0.311 | 0.025 | 0.025 |
| SAL1 | 0.450 | 0.024 | 0.016 | 0.450 | 0.026 | 0.024 |
| SAL2 | 0.930 | 0.013 | 0.008 | 0.930 | 0.014 | 0.013 |
| SAL3 | 0.677 | 0.021 | 0.015 | 0.677 | 0.021 | 0.021 |
| WERK | 0.793 | 0.019 | 0.013 | 0.793 | 0.019 | 0.019 |
| KOS | 0.135 | 0.019 | 0.011 | 0.135 | 0.018 | 0.019 |
| HUUR | 0.534 | 0.030 | 0.016 | 0.534 | 0.033 | 0.030 |
| TRAN | 0.431 | 0.033 | 0.016 | 0.431 | 0.035 | 0.032 |
| VERH | 0.584 | 0.024 | 0.016 | 0.584 | 0.023 | 0.024 |
| REL1 | 0.984 | 0.005 | 0.004 | 0.984 | 0.005 | 0.005 |
| REL2 | 0.934 | 0.012 | 0.008 | 0.934 | 0.013 | 0.012 |
| HUW1 | 0.796 | 0.016 | 0.013 | 0.796 | 0.016 | 0.016 |
| GLUK | 0.894 | 0.013 | 0.010 | 0.894 | 0.014 | 0.013 |
| OUER | 0.791 | 0.017 | 0.013 | 0.791 | 0.017 | 0.017 |
| TOEK | 0.823 | 0.015 | 0.012 | 0.823 | 0.013 | 0.015 |

## 3.7 OPSOMMING

Vir eenvoudige ewekansige steekproefneming is die berekenings van variansies van beramers van populasieparameters regstreeks.

Daar is heelwat minder literatuur beskikbaar oor die beraming van variansies van beramers van populasieparameters vir komplekse meerstadiumsteekproewe. Die hoofdoel van hierdie hoofstuk was om drie variansieberamingsmetodes vir komplekse meerstadiumsteekproewe te bespreek, nl. die Taylor-linearisasiemetode, die BRR- (Gebalanseerde Herhaalde-Replikasie) metode en die JRR- (Uitsnit Herhaalde-Replikasie) metode. Al drie hierdie metodes is breedvoerig bespreek.

'n Groot tekortkoming van die Taylor-linearisasiemetode is dat dit beperk is tot hoofsaaklik lineêre beramers (bv. gemiddeldes of verhoudings). Die BRR- en JRR-metodes is onder andere ontwikkel om hierdie tekortkoming te oorbrug en besit sodoende 'n groot voordeel bo die Taylor-linearisasiemetode.

Die baie interessante benadering van die twee herhaalde replikasiemetodes vir die berekening van variansies van beramers van populasieparameters uit komplekse steekproewe, het daartoe gelei dat 'n program BRRJRR deur die outeur hier by die RGN ontwikkel is. Hierdie program bereken BRR- en JRR-variansieberamers vir komplekse steekproewe.

Aangesien die program CLUSTERS wat gebruik maak van die Taylor-linearisasiemetode, reeds by die RGN beskikbaar is, is die eerste gedeelte van die program, wat die intermediêre statistieke vir al die PSE's bereken, aan die BRRJRR program gekoppel.

Jewett het 'n program ontwikkel om 'n ortogonale matriks te genereer. Deur middel van hierdie program kan McCarthy (1966) se ortogonale balanseringsmetode gebruik word vir die trek van halfsteekproewe vir die BRR-metode van variansieberaming.

Hierdie program van Jewett is uitgebrei en ook aan die BRRJRR program gekoppel. 'n Gedeelte van die program SAMPLE ontwikkel deur S.H.C. du Toit is as 'n subroetine in die BRRJRR-program geïnkorporeer om op 'n ewekansige wyse die JRR-steekproewe te trek.

Deur hierdie vier programme (CLUSTERS, SAMPLE, JEWETT se program en BRRJRR) aan mekaar te koppel is 'n program deur die outeur daargestel wat die enigste van sy soort in Suid-Afrika is en beslis een van die weinige programme in die wêreld is wat van McCarthy (1966) se ortogonale balanseringsmetode gebruik maak vir die berekening van die BRR-variansieberamers, en wat van die ewekansige pseudo getallegenerator gebruik maak vir die daarstel van die JRR-steekproewe en komplementêre JRR-steekproewe. Hierdie program is huidiglik beperk tot lineêre beramers (bv. gemiddeldes of verhoudings), maar dit kan relatief maklik uitgebrei word na nie-lineêre beramers (bv. korrelasie- of regressiekoëffisiënte).

Deur van die data van 'n RGN komplekse steekproef oor lewenskwaliteit gebruik te maak is die BRRJRR-program gebruik om die BRR en JRR se standaardfoute vir gemiddeldes van verskeie veranderlikewaardes te vergelyk met die standaardfoute van die Taylor-linearisasie soos verkry uit die program CLUSTERS. Die resultate was feitlik identies en gevolglik kan die BRR- en JRR- variansieberamingsmetode met vrymoedigheid toegepas word.

# BRONNELYS

Cochran, W.E. (1977). Sampling Techniques. John Wiley and Sons, Londen.

Deming, W.E. (1960). Some theory of sampling. John Wiley and Sons, New York.

Efron, B., & Gong G. (1983). A Leisurely Look at the Bootstrap, the Jackknife, and Cross-Validation. The American Statistician. Vol.37, No.1, pp.36-49. Washington.

Frankel, M.R. (1971) Inference from Clustered Samples. Institute for social Research, University of Michigan, Ann Arbor, Mich.

Frankel, M.R. (1971). Inference from Survey Samples: An Empirical Investigation. Institute for Social Research, University of Michigan, Ann Arbor, Mich.

Gurney, M. (1970). The variance of the Replication method for estimating variances for Grouped Strata. Technical Notes No.3, U.S. Bureau of the Census, Washington D.C..

Gurney, M. & Jewett, R. (1975). Constructing Orthogonal Replications for Variance Estimation. JASA. Vol. 70, pp. 819-821.

Jewett, R.S., Variance Estimation using Half-Sample Replications. Ongepubliseerde manuskrip U.S. Buro van Sensus.

Jones, H.L. (1956). Investigating the Properties of a Sample Mean by Employing Random Sub-Sample Means. American Statistical Association, No. 51, Washington.

Kish, L., and Frankel, M.R. (1968) Balanced Repeated Replications for Analytical Statistics. Proceedings of the Social Statistics Section of ASA.

Kish, L., and Frankel, M.R. (1970) Balanced Repeated Replications for Standard Errors. JASA, No. 65, pp 1071-1094. Washington.

Kish, L., and Frankel, M.R. (1974) Inference from Complex Samples. J. Royal Statistical Society(B), Vol. 36, pp.4537.

Kish, L. (1965). Survey Sampling. John Wiley and Sons, New York.

Kish, L. (1964) Generalization for Complex Probability Sampling. Proceedings of the Social Statistics Section, American Statistical Association.

Kish, L. (1972) Sampling Errors fo Large Scale Multipurpose Surveys.

Kish, L. (1980) Design and Estimation for Domains. Prepared for the Conference on Censuses and Sample Surveys, organized by the Institute of Statisticians, Cambridge University, England.

Kish, L., Groves, R.M. & Krotki, K.P. (1976). Sampling Errors for Fertility Surveys. WFS Occasional Papers No.17. Voorburg, Nederland.

Kish, L., & Hess, I. (1959). variances of Ratios and their Differences in Multi-Stage Samples. JASA No. 54, 416-446. Washington.

Keyfitz, N. (1957). Estimates of Sampling Variance. When two units are selected from each stratum. JASA, 52, 503-510.

Little, R.J.A. (1978). Generalized Linear Models for Cross-Classified Data from the WFS. WFS Technical Bulletins no 5. New York.

McCarthy, P.J. (1966). Replication: An approach to the Analysis of Data from Complex Surveys. National Center for Health Statistics, Series 2, No. 14, Washington.

McCarthy, P.J. (1966). Pseudo-Replication: Further Evaluation and Application the Balanced Half-Sample Technique. National Center for Health Statistics, Series 2, No.31, Washington.

McCarthy, P.J. (1969). Pseudo-Replication: Half Samples. Review of the International Statistical Institute, No.37, pp. 239-64.

Plackett, R.L. & Burman, P.J. (1946) The Design of optimum Multi-Factorial Experiments. Biometrika, No.33, 305-325.

Tepping, B.J. (1968). The estimation of Variance in Complex Surveys. Proceedings of the Social Statistics Section of the American Statistical Association.

Stoker, D.J. (1983). Steekproefneming in die Praktyk. Geleentheidspublikasie No. 4, RGN, Pretoria.

Stoker, D.J. (saamsteller)(1983). Steekproefnemingstegnieke. Kursus STA452-8, UNISA, Pretoria.

Verma, V. and Pearce, M. C. (1978). Users' Manual for CLUSTERS. WFS Technical Paper No.770. New York.

Verma, V. (1982). The Estimation and Presentation of Sampling Errors. WFS Technical Bulletins No.11. New York.

# BYLAE A

## <u>Die Program BRRJRR FORTRAN</u>

```
C     HIERDIE PROGRAM IS ONTWIKKEL DEUR ALIDA HERBST, BY DIE INSTITUUT
C     VIR STATISTIESE NAVORSING, RAAD VIR GEESTESWETENSKAPLIKE
C     NAVORSING (RGN), PRETORIA, TEL 283944 X 13. HIERDIE PROGRAM IS
C     ONTWERP OM TE LOOP OP 'N IBM REKENAAR. DIE EERSTE GEDEELTE VAN
C     DIE PROGRAM BEREKEN 'N GENERATOR VIR DIE ORTOGONALE MATRIKS
C     WAT GEBRUIK WORD OM DIE HALFSTEEKPROEWE TE BEREKEN. HIERDIE
C     GEDEELTE VAN DIE PROGRAM IS ONTWIKKEL DEUR ROBERT JEWETT VAN
C     DIE STATISTIESE METODES AFDELING, BURO VIR SENSUS, SUITLAND,
C     MARYLAND, TEL 301-763-2306 . DIE SUBROETINE CREATE IS HEELWAT
C     GEWYSIG.
C
      INTEGER P(30),Q(30),R(30),T(30),MODULO(30,15),ISOL
      INTEGER S(30),COL(5000),SUM,A,A2I
      INTEGER A2,APLUS1,AMIN1
      COMMON/SOMME/STOTY,STOTYY,STOTXY,STOT1,SOMY(300),SOMYY(300),NVAR,
     +SOMCY(300),SOMCYY(300),SOMCXY(300),SOMC1(300),
     +SOMXY(300),SOM1(300),TOTY,TOTYY,NOM
      COMMON/INV/DATA(300,30),IROW,JCOL
      COMMON/IJR/JR(100),IANT1
C
C     THE PROGRAM READS IN THE VALUE FOR A, AND GENERATES THE FIRST
C     COLUMN OF THE 2**A BY 2**A ORTHOGONAL MATRIX. THE WAY THE PROGRAM
C     IS SET UP, THE VALUE FOR A SHOULD BE LESS THAN OR EQUAL TO 12.
C
      DO 10 I=1,30
      P(I)=0
      Q(I)=0
      R(I)=0
      S(I)=0
      T(I)=0
      DO 10 J=1,15
      MODULO(I,J)=0
   10 CONTINUE
      DO 15 I=1,5000
      COL(I)=0
   15 CONTINUE
C
      WRITE(1,20)
   20 FORMAT ('U MATRIKS SAL 2**A BY 2**A WEES , WATTER MOET DIE WAARDE
     +VAN A WEES?')
      READ(1,* )  A
C 50  FORMAT( )
      WRITE(6,60) A
   60 FORMAT ('1 A= ',I5///)
      APLUS1=A+1
      AMIN1=A-1
      A2=2**A-1
C
C     THIS PART OF THE PROGRAM FINDS AN IRREDUCIBLE POLYNOMIAL P(X)
```

```
C      OF DEGREE A. THIS POYNOMIAL IS STORED IN THE ARRAY P SUCH THAT
C      P(I) = THE COEFFICIENT OF THE X**(I-1) TERM. IN FACT, THIS IS HOW
C      ALL POLYNOMIALS ARE STORED IN THE PROGRAM.
C
       NUM=A2+1
  100  NUM=NUM+1
       IF(NUM.GT.2*A2-1) STOP 1
       CALL CREATE(P,NUM,APLUS1)
       CALL SOLVE(P,ISOL,APLUS1)
       IF (ISOL.EQ.1) GO TO 100
       LIM=A/2
       DO 200 N=1,LIM
       LIM1=2**N+1
       LIM2=2**(N+1)-1
       DO 175 NUM1=LIM1,LIM2
       CALL CREATE(Q,NUM1,APLUS1)
       CALL SOLVE(Q,ISOL,APLUS1)
       IF (ISOL.EQ.1) GO TO 175
       LIM3=2**(A-N)+1
       LIM4=2**(A-N+1)-1
       DO 150 NUM2=LIM3,LIM4
       CALL CREATE(R,NUM2,APLUS1)
       CALL SOLVE(R,ISOL,APLUS1)
       IF (ISOL.EQ.1) GO TO 150
       CALL MULT(Q,R,T,A,APLUS1)
       DO 125 I=1,APLUS1
       IF(T(I).NE.P(I)) GO TO, 150
  125  CONTINUE
       WRITE(6,130) Q,R,P
  130  FORMAT(' P IS NOT PRIME: Q*R=P'/' Q=',30I2/' R=',30I2/' P=',30I2/)
       GO TO 100
  150  CONTINUE
  175  CONTINUE
  200  CONTINUE
       WRITE(8,230) P
  230  FORMAT (/' IRREDUCIBLE POLYNOMIAL P='/' P=',30I2/)
C
C      THIS PART CONSTRUCTS THE ARRAY MODULO. ROW N OF THE ARRAY GIVES
C      THE POLYNOMIAL (OF DEGREE LESS THAN A) THAT THE POLYNOMIAL X**N IS
C      EQUIVLAENT TO MODULO THE PRIME POLYNOMIAL P(X)
C
       DO 400 I=1,A
       MODULO(A,I)=P(I)
  400  CONTINUE
       LIM1=2*A-2
       DO 600 N=APLUS1,LIM1
       NUM=2**(N-A)-1
       LIM2=2**(N-A+1)-1
  450  NUM=NUM+1
       IF (NUM.GT.LIM2) STOP 2
       CALL CREATE(Q,NUM,APLUS1)
       CALL MULT(Q,P,R,A,APLUS1)
       R(N+1)=0
       DO 500 J=APLUS1,N
       IF (R(J).NE.0) GO TO 450
  500  CONTINUE
       WRITE(8,550) N,Q,R
```

```
 550  FORMAT (' EQUIVALENT POLYNOMIALS : X**N=Q*P+R : N= ',I5/
     *    ' Q= ',30I2/' R=',30I2/)
      DO 575 I=1,A
      MODULO(N,I)=R(I)
 575  CONTINUE
 600  CONTINUE
      WRITE (8,605) (I1,(MODULO(I1,I2),I2=1,15),I1=1,30)
 605  FORMAT(//'  X**N MODULE THE PRIME POLYNOMIAL'//30(' N=',I4,6X,
     C   15I2/)///)
C
C     THIS PART TESTS A NUMBER OF POLYNOMIALS TO SEE IF THEY COULD
C     GENERATE THE FIELD.
C     A GENERATOR IS FOUND, AND THE VECTOR IS CREATED THAT CAN BE USED
C     TO CONSTRUCT THE ORTHOGONAL MATRIX
C
      NUM=2
 620  NUM=NUM+1
      IF (NUM.GT.A2) STOP 3
      CALL CREATE(T,NUM,APLUS1)
      DO 640 I=1,A
      Q(I)=T(I)
 640  CONTINUE
      K=1
      COL(1)=Q(1)
 650  K=K+1
      IF (K.GT.A2) STOP 4
      CALL MULT(Q,T,S,A,APLUS1)
      CALL MODPOL(S,R,A,MODULO)
      DO 680 I=1,A
      Q(I)=R(I)
 680  CONTINUE
      COL(K)=Q(1)
      DO 690 I=2,A
      IF (Q(I).NE.0) GO TO 650
 690  CONTINUE
      IF (K.NE.A2) GO TO 620
      WRITE (6,695) T
 695  FORMAT (//'  GENERATOR T='/2X,30I2//)
C
C     THIS PART OF THE PROGRAM TESTS THE MATRIX TO SEE IF THE
C     COLUMNS ARE ORTHOGONAL.
C
      WRITE(6,720) (COL(I),I=1,A2)
 720  FORMAT(' COLUMN 1= ',100I1//(12X,100I1//))
      DO 740 I=1,A2
      IF (COL(I).EQ.0) COL(I)=-1
 740  CONTINUE
      DO 800 J=2,A2
      SUM=0
      DO 760 K=1,A2
      N=K+J-1
      IF (N.GT.A2) N=MOD(N,A2+1)+1
      SUM=SUM+COL(K)*COL(N)
 760  CONTINUE
      IF (SUM.NE.-1) WRITE(6,790)
 790  FORMAT (///' *** THE MATRIX IS NOT ORTHOGONAL *** ')
      IF (SUM.NE.-1) STOP 5
```

```
 800  CONTINUE
      WRITE (6,820)
 820  FORMAT(////' *** THE MATRIX IS ORTHOGONAL *** ')
      READ (5,*) IROW,JCOL
      I2=IROW/2
      READ (5,*) ((DATA(I,J),J=1,JCOL),I=1,IROW)
      WRITE(1,821)
 821  FORMAT (/'WIL U SLEGS N BEPERKTE UITVOER HE? JA=1,NEE=0.')
      READ(1,*) IBEP
      READ (5,*)(JR(I),I=1,I2)
 825  DO 826 I=1,IROW
      IF(DATA(I,2).EQ.0.0) DATA(I,5)=0.0
      STOTY=DATA(I,2) +STOTY
      STOTXY=DATA(I,3) +STOTXY
      STOT1 =DATA(I,5) +STOT1
      IF (IBEP.EQ.1) GO TO 826
      WRITE(6,830)  DATA(I,1),DATA(I,2),DATA(I,3),DATA(I,5)
  826 CONTINUE
  830 FORMAT (4F8.2)
      WRITE (1,835)
 835  FORMAT ('WIL U DIE BRR STATISTIEK BEREKEN? JA=1,NEE=0')
      READ (1,*) IANT1
      IF (IANT1.NE.1) GO TO 840
      CALL KOLBRR (A2,COL,IBEP)
      CALL BRR
 840  WRITE (1,850)
 850  FORMAT ('WIL U DIE JRR STATISTIEK BEREKEN? JA=1,NEE=0')
      READ (1,*) IANT2
      IF(IANT2.NE.1) GOTO  860
      CALL KOLJRR (A2,COL,IBEP)
      CALL JRR
 860  STOP 7
      END
C
C     THIS SUBROUTINE MULTIPLIES TWO POLYNOMIALS OF DEGREE LESS THAN OR
C     EQUAL TO (A-1). THE POLYNOMIALS W & X ARE MULTIPLIED, AND
C     THE RESULT IS Z.
C
      SUBROUTINE MULT(W,X,Z,A,APLUS1)
      INTEGER W(30),X(30),Z(30)
      INTEGER A,APLUS1
      DO 1100 I=1,30
      Z(I)=0
 1100 CONTINUE
      DO 1200 I=1,APLUS1
      DO 1200 J=1,APLUS1
      IF (W(I).EQ.1.AND.X(J).EQ.1) Z(I+J-1)=Z(I+J-1)+1
 1200 CONTINUE
      LIMZ=2*A
      DO 1250 I=1,LIMZ
      Z(I)=MOD(Z(I),2)
 1250 CONTINUE
      RETURN
      END
C
C     THE ARRAY MODULO IS USED TO TAKE A POLYNOMIAL MODULO THE
C     PRIME POLYNOMIAL P. WE FIND THE POLYNOMIAL Z (OF DEGREE LESS
```

```
C      THAN A) SUCH THAT Z=W(MODULE P).
C
       SUBROUTINE MODPOL(W,Z,A,MODULO)
       INTEGER W(30),Z(30),A,MODULO(30,15)
       DO 1300 I=1,A
       Z(I)=W(I)
 1300  CONTINUE
       LIMZ=2*A-2
       DO 1400 N=A,LIMZ
       IF (W(N+1).EQ.0) GO TO 1400
       DO 1350 J=1,A
       Z(J)=Z(J)+MODULO(N,J)
 1350  CONTINUE
 1400  CONTINUE
       DO 1430 J=1,A
       Z(J)=MOD(Z(J),2)
 1430  CONTINUE
       RETURN
       END
C
C      THIS SUBROUTINE CHECKS TO SEE IF THE POLYNOMIAL Z HAS A SOLUTION
C
       SUBROUTINE SOLVE(Z,ISOL,APLUS1)
       INTEGER Z(30),ISOL,APLUS1,SUM
       SUM=0
       DO 1500 I=1,APLUS1
       SUM=SUM+Z(I)
 1500  CONTINUE
       SUM=MOD(SUM,2)
       IF (SUM.EQ.0) ISOL=1
       IF (SUM.NE.0) ISOL=0
       IF (Z(1).EQ.0) ISOL=1
       RETURN
       END
C
C      THIS SUBROUTINE IS USED TO CREATE A POLYNOMIAL Z CORRESPONDING
C      TO THE INTEGER N. THE POLYNOMIAL IS OF ORDER LESS THAN OR EQUAL
C      TO A+1.
C
C      NOTE THAT THIS IS WHERE THE UNIVAC FORTRAN FIELD (FLD) FUNCTION
C      IS USED TO EXTRACT THE BINARY BITS FROM THE NUMBER N AND CREATE
C      A POLYNOMIAL Z. FOR EXAMPLE, IF N=11, THEN THIS IS EQUAL TO 1101
C      IN BINARY, AND THIS GIVES RISE TO THE POLYNOMIAL X**3+X**2+1.
C
C      SUBROUTINE CREATE(Z,N)
C      INTEGER Z(30)
C      DO 1600 I=1,APLUS1
C      Z(I)=FLD(36-I,1,N)
C1600  CONTINUE
C      RETURN
C      END
       SUBROUTINE CREATE(Z,N,APLUS1)
       INTEGER Z(30),APLUS1
       DO 1600 I=1,APLUS1
       Z(I)=N
       IF (I.EQ.1) GOTO 1100
       K=I-1
```

```
      DO  600 J=1,K
      Z(I)=Z(I)/2
  600 CONTINUE
 1100 CONTINUE
      Z(I)=MOD(Z(I),2)
 1600 CONTINUE
      RETURN
      END
C
C     HIERDIE SUBROUTINE WERK DIE KOLOMME UIT VANAF DIE GENERATOR
C     EN BEREKEN DIE SOM TOTALE VIR ELKE HALFSTEEKPROEF
C
C***********************************************************************
      SUBROUTINE KOLBRR(A2,COL,IBEP)
      INTEGER COL(1),A2,A2I,K1,R,L,S,W
      DIMENSION RMATR(300,30)
      COMMON/SOMME/STOTY,STOTYY,STOTXY,STOT1,SOMY(300),SOMYY(300),NVAR,
     +SOMCY(300),SOMCYY(300),SOMCXY(300),SOMC1(300),
     +SOMXY(300),SOM1(300),TOTY,TOTYY,NOM
      COMMON/INV/DATA(300,30),IROW,JCOL
         WRITE (6,173)
         WRITE (1,4)
    4 FORMAT  (// 'WERK U MET GEWIGTE? JA=1,NEE=0')
      READ (1,*) IANT
      IR2=IROW-1
      NOM=IROW/2
      IRD2=IROW/2
      STOTY=0.0
      STOTYY=0.0
      STOTXY=0.0
      STOT1 =0.0
      TOTY =0.0
      DO 14 II=1,IRD2
      SOMY(II)=0.0
      SOMYY(II)=0.0
      SOMXY(II)=0.0
      SOM1(II)=0.0
   14 CONTINUE
      IF (IANT.EQ.1) GOTO  400
C
C    HIER WORD DIE TOTALE VAN DIE STEEKPROEF BEREKEN
C
      DO 16 I=1,IROW
C     IF(DATA(I,2).EQ.0.0) DATA(I,5)=0.0
      STOTY=DATA(I,2) +STOTY
      STOTXY=DATA(I,3) +STOTXY
      STOTYY=DATA(I,4) +STOTYY
      STOT1 =DATA(I,5) +STOT1
   16 CONTINUE
      IN=A2+1
C
C   DIE GEMIDDELD VAN DIE HELE STEEKPROEF EN DIE GEMIDDELDE VAN DIE
C   SOM VAN KWADRATE WORD IN TOTY EN TOTYY GESTOOR

      TOTY=STOTY/STOT1
      STOT1=STOT1
      WRITE (6,18) TOTY,STOT1
```

```
   18  FORMAT('VIR DIE HELE STEEKPROEF:'/3X,'SOMY =',3X,F12.2
      +/3X,'SOM1 =',3X,F12.2)

C BEREKEN DIE EERSTE HALFSTEEKPROEF SE STATISTIEKE
C
       J=1
       A2I=A2+1
       COL(A2I)=-1
       WRITE (6,100) J,(COL(L),L=1,IRD2)
        DO 24 K=1,IR2,2
        L=(K+1)/2
        K1=K-1
         DO 23 I=1,2
             R=K1+I
          DO 22 J=1,JCOL
          COLL=COL(L)
             IF (COLL.EQ.DATA(R,1))
      +               RMATR(L,J)=DATA(R,J)
             IF (COLL.NE.DATA(R,1)) GO TO 23
   22     CONTINUE
   23     CONTINUE
   24     CONTINUE
          IRD2=IROW/2
          DO 28 L=1,IRD2
           IF(RMATR(L,2).EQ.0.0) RMATR(L,5) =0.0
           SOMY(1)=RMATR(L,2)+SOMY(1)
           SOMXY(1)=RMATR(L,3)+SOMXY(1)
           SOMYY(1)=RMATR(L,4)+SOMYY(1)
           SOM1(1)=RMATR(L,5)+SOM1(1)
           IF (IBEP.EQ.1) GO TO 28
           WRITE (6,149) (RMATR(L,1),RMATR(L,2),RMATR(L,5))
   28      CONTINUE
           SOMCY(1)=STOTY-SOMY(1)
           SOMCXY(1)=STOTXY-SOMXY(1)
           SOMCYY(1)=STOTYY-SOMYY(1)
           SOMC1(1)=STOT1-SOM1(1)
C
C      DIE MIDDELSTE GEDEELTE VAN DIE HALFSTEEKPROEWE WORD NOU BEREKEN
C
       NOM1=NOM-1
       DO 60 J=2,NOM1
       COL(A2)=COL(1)
       DO 40 I=2,A2
       COL(I-1)=COL(I)
   40  CONTINUE
       IF (IBEP.EQ.1) GO TO 41
       WRITE (6,100) J,(COL(L),L=1,IRD2)
   41  IR2=IROW-1
        DO 44 K=1,IR2,2
        L=(K+1)/2
        K1=K-1
         DO 43 I=1,2
             R=K1+I
          DO 42 S=1,JCOL
             COLL=COL(L)
             IF (COLL.EQ.DATA(R,1))
      +               RMATR(L,S)=DATA(R,S)
```

```
          IF (COLL.NE.DATA(R,1)) GO TO 43
   42    CONTINUE
   43    CONTINUE
   44    CONTINUE
         DO 48 L=1,IRD2
          IF(RMATR(L,2).EQ.0.0) RMATR(L,5) =0.0
          SOMY(J)=RMATR(L,2)+SOMY(J)
          SOMXY(J)=RMATR(L,3)+SOMXY(J)
          SOMYY(J)=RMATR(L,4)+SOMYY(J)
          SOM1(J)=RMATR(L,5)+SOM1(J)
          IF (IBEP.EQ.1) GO TO 48
          WRITE (6,149) (RMATR(L,1),RMATR(L,2),RMATR(L,5))
   48   CONTINUE
          SOMCY(J)=STOTY-SOMY(J)
          SOMCXY(J)=STOTXY-SOMXY(J)
          SOMCYY(J)=STOTYY-SOMYY(J)
          SOMC1(J)=STOT1-SOM1(J)
   60 CONTINUE
  100 FORMAT  (/,'COL NR',2X,I2,2X,50I2,/,60I2)
C
C  NOU WORD DIE LAASTE HALFSTEEKPROEF BEREKEN
C
  110 DO 120 I=1,IRD2
      COL(I)=1
  120 CONTINUE
      IF (IBEP.EQ.1) GO TO 121
      WRITE(6,100) IRD2,(COL(L),L=1,IRD2)
  121 IR2=IROW-1
       DO 144 K=1,IR2,2
       L=(K+1)/2
       K1=K-1
        DO 143 I=1,2
            R=K1+I
         DO 142 S=1,JCOL
           COLL=COL(L)
            IF (COLL.EQ.DATA(R,1))
     +           RMATR(L,S)=DATA(R,S)
            IF (COLL.NE.DATA(R,1)) GO TO 143
  142    CONTINUE
  143    CONTINUE
  144    CONTINUE
         J=NOM
         DO 148 L=1,IRD2
          IF(RMATR(L,2).EQ.0.0) RMATR(L,5) =0.0
          SOMY(J)=RMATR(L,2)+SOMY(J)
          SOMXY(J)=RMATR(L,3)+SOMXY(J)
          SOMYY(J)=RMATR(L,4)+SOMYY(J)
          SOM1(J)=RMATR(L,5)+SOM1(J)
          IF (IBEP.EQ.1) GO TO 148
          WRITE (6,149) (RMATR(L,1),RMATR(L,2),RMATR(L,5))
  148   CONTINUE
  149     FORMAT (3F8.2)
          SOMCY(J)=STOTY-SOMY(J)
C         SOMCXY(J)=STOTXY-SOMXY(J)
C         SOMCYY(J)=STOTYY-SOMYY(J)
          SOMC1(J)=STOT1-SOM1(J)
         NOM  =IROW/2
```

```
C
C   AL DIE HALFSTEEKPROEWE IS NOU BEREKEN EN UITGESKRYF
C   NOU WORD DIE SOMME BEREKEN VIR AL DIE HALFSTEEKPROEWE SAAM
C
         DO 160 J=1,NOM
         SOMY(J)=SOMY(J)/SOM1(J)
         SOMCY(J)=SOMCY(J)/SOMC1(J)
C        SOMXY(J)=SOMXY(J)
C        SOMYY(J)=SOMYY(J)/SOM1(J)
C        SOMCYY(J)=SOMCYY(J)/SOMC1(J)
          IF (IBEP.EQ.1) GO TO 160
         WRITE (6,170) J,SOMY(J),SOM1(J)
         WRITE (6,171) J,SOMCY(J),SOMC1(J)
  160    CONTINUE
  170    FORMAT ('HALFSP NR ',I3,':'/3X,'SOMY =',3X,F12.2/3X,
     +'SOM1 =',3X,F12.2)
  171    FORMAT ('KOMPL  NR ',I3,':'/3X,'SOMCY =',2X,F12.2/3X,
     +'SOMC1 =',2X,F12.2)
  173    FORMAT(//'BALANCED REPEATED REPLICATION UITVOER')
         IF (IANT.NE.1) GOTO 600
C
C    HIER WORD DIE TOTALE VAN DIE STEEKPROEF BEREKEN
C    INDIEN DIE STEEKPROEF GEWIGTE HET
  400 DO 416 I=1,IROW
      IF(DATA(I,2).EQ.0.0) DATA(I,5)=0.0
      STOTY=DATA(I,2) +STOTY
      STOTXY=DATA(I,3) +STOTXY
      STOT1 =DATA(I,5) +STOT1
  416 CONTINUE
      IN=A2+1
C
C   DIE GEMIDDELD VAN DIE HELE STEEKPROEF EN DIE GEMIDDELDE VAN DIE
C   SOM VAN KWADRATE WORD IN TOTY EN TOTYY GESTOOR

      TOTY=STOTY/STOTXY
      WRITE (6,418) TOTY,STOT1
  418  FORMAT('VIR DIE HELE STEEKPROEF:'/3X,'GEWEEGDE SOMY =',3X,F12.4/
     +/3X,'# USE=',3X,F12.4)

C BEREKEN DIE EERSTE HALFSTEEKPROEF SE STATISTIEKE
C
      J=1
      A2I=A2+1
      COL(A2I)=-1
      IF (IBEP.EQ.1) GO TO 419
      WRITE (6,500) J,(COL(L),L=1,IRD2)
  419  DO 424 K=1,IR2,2
       L=(K+1)/2
       K1=K-1
        DO 423 I=1,2
            R=K1+I
         DO 422 J=1,JCOL
         COLL=COL(L)
            IF (COLL.EQ.DATA(R,1))
     +           RMATR(L,J)=DATA(R,J)
            IF (COLL.NE.DATA(R,1)) GO TO 423
  422    CONTINUE
```

```
  423     CONTINUE
  424     CONTINUE
          IRD2=IROW/2
          DO 428 L=1,IRD2
           SOMY(1)=RMATR(L,2)+SOMY(1)
           SOMXY(1)=RMATR(L,3)+SOMXY(1)
           SOM1(1)=RMATR(L,5)+SOM1(1)
           IF (IBEP.EQ.1) GO TO 428
           WRITE (6,549) (RMATR(L,1),RMATR(L,2),RMATR(L,5))
  428      CONTINUE
           SOMCY(1)=STOTY-SOMY(1)
           SOMCXY(1)=STOTXY-SOMXY(1)
           SOMC1(1)=STOT1-SOM1(1)
C
C      DIE MIDDELSTE GEDEELTE VAN DIE HALFSTEEKPROEWE WORD NOU BEREKEN
C
      NOM1=NOM-1
      DO 460 J=2,NOM1
      COL(A2)=COL(1)
      DO 440 I=2,A2
      COL(I-1)=COL(I)
  440 CONTINUE
      IF (IBEP.EQ.1) GO TO 441
      WRITE (6,500) J,(COL(L),L=1,IRD2)
  441 IR2=IROW-1
       DO 444 K=1,IR2,2
       L=(K+1)/2
       K1=K-1
        DO 443 I=1,2
            R=K1+I
         DO 442 S=1,JCOL
            COLL=COL(L)
            IF (COLL.EQ.DATA(R,1))
     +           RMATR(L,S)=DATA(R,S)
            IF (COLL.NE.DATA(R,1)) GO TO 443
  442    CONTINUE
  443    CONTINUE
  444    CONTINUE
         DO 448 L=1,IRD2
          IF(RMATR(L,2).EQ.0.0) RMATR(L,5) =0.0
          SOMY(J)=RMATR(L,2)+SOMY(J)
          SOMXY(J)=RMATR(L,3)+SOMXY(J)
          SOM1(J)=RMATR(L,5)+SOM1(J)
          IF (IBEP.EQ.1) GO TO 448
          WRITE (6,549) (RMATR(L,1),RMATR(L,2),RMATR(L,5))
  448   CONTINUE
          SOMCY(J)=STOTY-SOMY(J)
          SOMC1(J)=STOT1-SOM1(J)
          SOMCXY(J)=STOTXY-SOMXY(J)
  460 CONTINUE
  500 FORMAT  (/,'COL NR',2X,I2,2X,50I2,/,60I2)
C
C  NOU WORD DIE LAASTE HALFSTEEKPROEF BEREKEN
C
  510 DO 520 I=1,IRD2
      COL(I)=1
  520 CONTINUE
```

```
      IF (IBEP.EQ.1) GO TO 521
      WRITE(6,500) IRD2,(COL(L),L=1,IRD2)
  521 IR2=IROW-1
       DO 544 K=1,IR2,2
       L=(K+1)/2
       K1=K-1
        DO 543 I=1,2
            R=K1+I
         DO 542 S=1,JCOL
           COLL=COL(L)
            IF (COLL.EQ.DATA(R,1))
     +           RMATR(L,S)=DATA(R,S)
            IF (COLL.NE.DATA(R,1)) GO TO 543
  542    CONTINUE
  543    CONTINUE
  544    CONTINUE
         J=NOM
         DO 548 L=1,IRD2
          RMATR(L,4)=RMATR(L,2)/RMATR(L,3)
          SOMY(J)=RMATR(L,2)+SOMY(J)
          SOMXY(J)=RMATR(L,3)+SOMXY(J)
          SOM1(J)=RMATR(L,5)+SOM1(J)
          IF (IBEP.EQ.1) GO TO 548
          WRITE (6,549) (RMATR(L,1),RMATR(L,2),RMATR(L,5))
  548    CONTINUE
  549     FORMAT (3F8.2)
          SOMCY(J)=STOTY-SOMY(J)
          SOMCXY(J)=STOTXY-SOMXY(J)
          SOMC1(J)=STOT1-SOM1(J)
         NOM  =IROW/2
C
C   AL DIE HALFSTEEKPROEWE IS NOU BEREKEN EN UITGESKRYF
C   NOU WORD DIE SOMME BEREKEN VIR AL DIE HALFSTEEKPROEWE SAAM
C
         DO 560 J=1,NOM
         SOMY(J)=SOMY(J)/SOMXY(J)
         SOMCY(J)=SOMCY(J)/SOMCXY(J)
         IF (IBEP.EQ.1) GO TO 560
         WRITE (6,570) J,SOMY(J),SOM1(J)
         WRITE (6,571) J,SOMCY(J),SOMC1(J)
  560    CONTINUE
  570    FORMAT ('HALFSP NR ',I3,':'/3X,'SOMY GEWEEG GEM=',3X,F12.4/3X,
     +'SOM1 =',13X,F12.4)
  571    FORMAT ('KOMPL  NR ',I3,':'/3X,'SOMCY GEWEEG GEM=',2X,F12.4/3X,
     +'SOMC1 =',12X,F12.4)
  600    RETURN
         END
C**********************************************************************
C   HIERDIE SUBROUTINE BEPAAL DIE JACKKNIFE STEEKPROEWE EN BEREKEN
C   DIE ONDERSKEIE TOTALE PER STEEKPROEF
C**********************************************************************
      SUBROUTINE KOLJRR(A2,COL,IBEP)
      INTEGER COL(1),A2,A2I,K1,R,L,S,W
      DIMENSION
     +SOMY1(300),SXY1(300),SYY1(300),S11(300),
     +SCY1(300),SCXY1(300),SCYY1(300),SC11(300),
     +SOMY2(300),SXY2(300),SYY2(300),S12(300),
```

```
     +SCY2(300),SCXY2(300),SCYY2(300),SC12(300),
     +SOMY3(300),SXY3(300),SYY3(300),S13(300),
     +SCY3(300),SCXY3(300),SCYY3(300),SC13(300)
      COMMON/SOMME/STOTY,STOTYY,STOTXY,STOT1,SOMY(300),SOMYY(300),NVAR,
     +SOMCY(300),SOMCYY(300),SOMCXY(300),SOMC1(300),
     +SOMXY(300),SOM1(300),TOTY,TOTYY,NOM
      COMMON/INV/DATA(300,30),IROW,JCOL
      COMMON/IJR/JR(100),IANT1
C     READ (5,*) IROW,JCOL
C     READ (5,*) ((DATA(I,J),J=1,JCOL),I=1,IROW)
C     WRITE(1,12)
C  12 FORMAT  (// 'HOEVEEL VERANDERLIKES GEBRUIK U?')
C     READ (1,*) NVAR
       WRITE (6,50)
       WRITE (1,4)
    4 FORMAT  (// 'WERK U MET GEWIGTE? JA=1,NEE=0')
      READ (1,*) IANT3
      IRD2=IROW/2
      STOTY=0.0
      STOTYY=0.0
      STOTXY=0.0
      STOT1 =0.0
      TOTY=0.0
      DO 14 II=1,IRD2
      SOMY(II)=0.0
      SOMYY(II)=0.0
      SOMXY(II)=0.0
      SOM1(II)=0.0
      SOMCY(II)=0.0
      SOMCYY(II)=0.0
      SOMCXY(II)=0.0
      SOMC1(II)=0.0
   14 CONTINUE
C
C  DIE STATISTIEKE VIR DIE HELE STEEKPROEF WORD BEREKEN
C
      IF(IANT3.EQ.1) GOTO 500
      DO 16 I=1,IROW
      IF(DATA(I,2).EQ.0.0) DATA(I,5)=0.0
      STOTY=DATA(I,2) +STOTY
      STOTXY=DATA(I,3) +STOTXY
      STOTYY=DATA(I,4) +STOTYY
      STOT1 =DATA(I,5) +STOT1
   16 CONTINUE
      IN=A2+1
C
C  DIE GEMIDDELDE WORD GESTOOR IN TOTY EN TOTYY
      IF (IANT1.EQ.1) GO TO 19
      TOTY=STOTY/STOT1
      WRITE (6,18) TOTY,STOT1
   18 FORMAT('VIR DIE HELE STEEKPROEF:'/3X,'SOMY =',3X,F12.2/3X,
     +'SOM1 =',3X,F12.2)

C INDIEN DIE JRR STATISTIEKE BEREKEN WORD WORD DIE STEEKPROEWE ANDERS
C SAAMGESTEL.
C
   19 J=1
```

```
      A2I=A2+1
      COL(A2I)=-1
         NOM  =IROW/2
         DO 20 L=1,IROW
         SOMY1(L)=0.0
         SOMY2(L)=0.0
         SOMY3(L)=0.0
         SXY1(L)=0.0
         SXY2(L)=0.0
         SXY3(L)=0.0
         SYY1(L)=0.0
         SYY2(L)=0.0
         SYY3(L)=0.0
         S11(L)=0.0
         S12(L)=0.0
         S13(L)=0.0
         SCY1(L)=0.0
         SCY2(L)=0.0
         SCY3(L)=0.0
         SCXY1(L)=0.0
         SCXY2(L)=0.0
         SCXY3(L)=0.0
         SCYY1(L)=0.0
         SCYY2(L)=0.0
         SCYY3(L)=0.0
         SC11(L)=0.0
         SC12(L)=0.0
         SC13(L)=0.0
  20     CONTINUE
C
C NOU WORD DIE JRR STEEKPROEWE GETREK DIE HELE STEEKPROEF WORD GETREK
C BEHALWE IN EEN STRATUM WORD EEN VAN DIE PSE'S GEDUPLISEER EN DIE
C ANDER PSE WEGGELAAT
C
         NOM  =IROW/2
         DO 40 L=1,NOM
         IK2=L*2
         IK1=IK2-1
         IKO=IK1-1
         IKB=IK2+1
         COLL=JR(L)
         IF (COLL.NE.DATA(IK1,1)) GOTO 22
             SOMY1(L)=2*DATA(IK1,2)
             SXY1(L)=2*DATA(IK1,3)
             SYY1(L)=2*DATA(IK1,4)
             S11(L)=2*DATA(IK1,5)
             SCY1(L)=2*DATA(IK2,2)
             SCXY1(L)=2*DATA(IK2,3)
             SCYY1(L)=2*DATA(IK2,4)
             SC11(L)=2*DATA(IK2,5)
   22    IF (COLL.EQ.DATA(IK1,1))  GOTO 24
             SOMY1(L)=2*DATA(IK2,2)
             SXY1(L)=2*DATA(IK2,3)
             SYY1(L)=2*DATA(IK2,4)
             S11(L)=2*DATA(IK2,5)
             SCY1(L)=2*DATA(IK1,2)
             SCXY1(L)=2*DATA(IK1,3)
```

```
            SCYY1(L)=2*DATA(IK1,4)
            SC11(L)=2*DATA(IK1,5)
        IF (IK1.EQ.1) GO TO 31
 24          DO  30 I=1,IKO
             SOMY2(L)=DATA(I,2) + SOMY2(L)
             SXY2(L)=DATA(I,3) + SXY2(L)
             SYY2(L)=DATA(I,4) + SYY2(L)
             S12(L)=DATA(I,5) + S12(L)
 30          CONTINUE
 31       IF (IK2.EQ.IROW) GO TO  35
            DO 34 I=IKB,IROW
            SOMY3(L)=DATA(I,2) + SOMY3(L)
            SXY3(L)=DATA(I,3) + SXY3(L)
            SYY3(L)=DATA(I,4) + SYY3(L)
            S13(L)=DATA(I,5) + S13(L)
  34       CONTINUE
  35      SOMY(L)=SOMY1(L)+SOMY2(L) + SOMY3(L)
          SOMYY(L)=SYY1(L)+SYY2(L) + SYY3(L)
          SOMXY(L)=SXY1(L)+SXY2(L) + SXY3(L)
          SOM1(L)=S11(L)+S12(L) + S13(L)
          SOMCY(L)=SCY1(L)+SOMY2(L) + SOMY3(L)
          SOMCXY(L)=SCXY1(L)+SXY2(L) + SXY3(L)
          SOMCYY(L)=SCYY1(L)+SYY2(L) + SYY3(L)
          SOMC1(L)=SC11(L)+S12(L) + S13(L)
  40    CONTINUE
C
C  DIE TOTALE VIR AL DIE HALFSTEEKPROEWE WORD NOU BEREKEN
C  DAAR IS SOVEEL HALFSTEEKPROEWE SOOS WAT DAAR STRATA IS
C
        DO 45 J=1,NOM
        SOMY(J)=SOMY(J)/SOM1(J)
        SOMCY(J)=SOMCY(J)/SOMC1(J)
C       SOMYY(J)=SOMYY(J)/SOM1(J)
C       SOMCYY(J)=SOMCYY(J)/SOMC1(J)
        IF (IBEP.EQ.1) GO TO 45
        WRITE (6,51) J,SOMY(J),SOM1(J)
        WRITE (6,52) J,SOMCY(J),SOMC1(J)
  45    CONTINUE
  51    FORMAT ('HALFSP NR ',I3,':'/3X,'SOMY =',3X,F12.2/3X,
     +'SOM1 =',3X,F12.2)
  52    FORMAT ('KOMPL  NR ',I3,':'/3X,'SOMCY =',2X,F12.2/3X,
     +'SOMC1 =',2X,F12.2)
  50    FORMAT(//'JACKNIFE REPEATED REPLICATION UITVOER')
      IF (IANT3.NE.1) GOTO 600
 500  DO 516 I=1,IROW
      IF(DATA(I,2).EQ.0.0) DATA(I,5)=0.0
      STOTY=DATA(I,2)+STOTY
      STOTXY=DATA(I,3)+STOTXY
      STOT1 =DATA(I,5) +STOT1
      IF (IBEP.EQ.1) GO TO 516
      WRITE(6,517) DATA(I,2),DATA(I,3)
 516  CONTINUE
 517  FORMAT (2F12.2)
      NOM=IROW/2
      IN=A2+1
C
C  DIE GEMIDDELDE WORD GESTOOR IN TOTY EN TOTYY
```

```
C     IF (IANT1.EQ.1) GO TO 519
      TOTY=STOTY/STOTXY
      WRITE (6,518) TOTY,STOT1
 518  FORMAT('VIR DIE HELE STEEKPROEF:'/3X,'SOMY GEWEEG GEM=',3X,F12.4
     +/3X,'SOM1 =',3X,F12.4)

C INDIEN DIE JRR STATISTIEKE BEREKEN WORD WORD DIE STEEKPROEWE ANDERS
C SAAMGESTEL.
C
 519  J=1
      A2I=A2+1
      COL(A2I)=-1
         NOM  =IROW/2
         DO 520 L=1,IROW
         SOMY1(L)=0.0
         SOMY2(L)=0.0
         SOMY3(L)=0.0
         SXY1(L)=0.0
         SXY2(L)=0.0
         SXY3(L)=0.0
         SYY1(L)=0.0
         SYY2(L)=0.0
         SYY3(L)=0.0
         S11(L)=0.0
         S12(L)=0.0
         S13(L)=0.0
         SCY1(L)=0.0
         SCY2(L)=0.0
         SCY3(L)=0.0
         SCXY1(L)=0.0
         SCXY2(L)=0.0
         SCXY3(L)=0.0
         SCYY1(L)=0.0
         SCYY2(L)=0.0
         SCYY3(L)=0.0
         SC11(L)=0.0
         SC12(L)=0.0
         SC13(L)=0.0
 520     CONTINUE
C
C NOU WORD DIE JRR STEEKPROEWE GETREK DIE HELE STEEKPROEF WORD GETREK
C BEHALWE IN EEN STRATUM WORD EEN VAN DIE PSE'S GEDUPLISEER EN DIE
C ANDER PSE WEGGELAAT
C
         NOM  =IROW/2
         DO 540 L=1,NOM
         IK2=L*2
         IK1=IK2-1
         IKO=IK1-1
         IKB=IK2+1
         COLL=JR(L)
         IF(COLL.NE.DATA(IK1,1))GOTO 522
             SOMY1(L)=2*DATA(IK1,2)
             SXY1(L)=2*DATA(IK1,3)
             S11(L)=2*DATA(IK1,5)
             SCY1(L)=2*DATA(IK2,2)
             SCXY1(L)=2*DATA(IK2,3)
```

```
              SC11(L)=2*DATA(IK2,5)
  522     IF (COLL.EQ.DATA(IK1,1))  GOTO 524
              SOMY1(L)=2*DATA(IK2,2)
              SXY1(L)=2*DATA(IK2,3)
              S11(L)=2*DATA(IK2,5)
              SCY1(L)=2*DATA(IK1,2)
              SCXY1(L)=2*DATA(IK1,3)
              SC11(L)=2*DATA(IK1,5)
  524     IF (IK1.EQ.1) GO TO 531
               DO  530 I=1,IKO
               SOMY2(L)=DATA(I,2) + SOMY2(L)
               S12(L)=DATA(I,5) + S12(L)
               SXY2(L)=DATA(I,3) + SXY2(L)
  530          CONTINUE
  531       IF (IK2.EQ.IROW) GO TO 535
              DO 534 I=IKB,IROW
              SOMY3(L)=DATA(I,2) + SOMY3(L)
              S13(L)=DATA(I,5) + S13(L)
              SXY3(L)=DATA(I,3) + SXY3(L)
  534         CONTINUE
  535       SOMY(L)=SOMY1(L)+SOMY2(L) + SOMY3(L)
            SOM1(L)=S11(L)+S12(L) + S13(L)
            SOMXY(L)=SXY1(L)+SXY2(L) + SXY3(L)
            SOMCY(L)=SCY1(L)+SOMY2(L) + SOMY3(L)
            SOMC1(L)=SC11(L)+S12(L) + S13(L)
            SOMCXY(L)=SCXY1(L)+SXY2(L) + SXY3(L)
  540     CONTINUE
C
C  DIE TOTALE VIR AL DIE HALFSTEEKPROEWE WORD NOU BEREKEN
C  DAAR IS SOVEEL HALFSTEEKPROEWE SOOS WAT DAAR STRATA IS
C
          DO 545 J=1,NOM
          SOMY(J)=SOMY(J)/SOMXY(J)
          SOMCY(J)=SOMCY(J)/SOMCXY(J)
          IF (IBEP.EQ.1) GO TO 545
          WRITE (6,551) J,SOMY(J),SOM1(J)
          WRITE (6,552) J,SOMCY(J),SOMC1(J)
  545     CONTINUE
  551     FORMAT ('HALFSP NR ',I3,':'/3X,'SOMY GEWEEG GEM =',3X,F12.4
     +/13X,'SOM1 =',3X,F12.4)
  552     FORMAT ('KOMPL  NR ',I3,':'/3X,'SOMCY GEWEEG GEM=',2X,F12.4
     +/13X,'SOMC1 =',2X,F12.4)
  600 RETURN
      END
C
C  HIER WORD DIE BRR STATISTIEKE BEREKEN
C
      SUBROUTINE BRR
      INTEGER NVAR,NOM
      COMMON/SOMME/STOTY,STOTYY,STOTXY,STOT1,SOMY(300),SOMYY(300),NVAR,
     +SOMCY(300),SOMCYY(300),SOMCXY(300),SOMC1(300),
     +SOMXY(300),SOM1(300),TOTY,TOTYY,NOM
C
C     BEREKEN NOU DIE BRR STATISTIEKE
C     X(I)=1-F/K*SOM(G(HI)-G(S))**2
C     EERSTENS WORD DIE BRR-C EN BRR-H  STATISTIEKE BEREKEN
C
```

```
      XVAR=0.0
      XCVAR=0.0
      BRRK=0.0
      DO 7 J=1,NOM
      SDIFF=SOMY(J)-TOTY
      SCDIFF=SOMCY(J)-TOTY
C          WRITE(6,9) SOMY(J),SOMCY(J),SDIFF,SCDIFF
      XVAR=(SDIFF**2) + XVAR
      XCVAR=(SCDIFF**2) + XCVAR
    7 CONTINUE
      XVARJ=XVAR/NOM
      XCVARJ=XCVAR/NOM
      BRRH=(1.0)*XVARJ
      BRRC=(1.0)*XCVARJ
      BRRH1=SQRT(BRRH)
      BRRC1=SQRT(BRRC)
      WRITE(6,10) BRRH1
      WRITE(6,12) BRRC1
    9 FORMAT ('SOMY(J)',2X,F8.3,2X,'SOMCY(J)',2X,F8.3,'SDIFF',2X,F10.8,
     +2X,'SCDIFF',2X,F10.8)
   10 FORMAT ('DIE BRR-H STATISTIEK IS  ',F15.5)
   12 FORMAT ('DIE BRR-C STATISTIEK IS  ',F15.5)
C
C     DIE BRR-S STATISTIEK WORD NOU BEREKEN
C     BRRS=(BRRH+BRRC)/2
C
      BRRS=(BRRH+BRRC)/2
      BRRS=SQRT(BRRS)
      WRITE(6,13) BRRS
   13 FORMAT ('DIE BRR-S STATISTIEK IS  ',F15.5)
C
C     DIE BRR-D STATISTIEK WORD NOU BEREKEN
C
      DO 14 J=1,NOM
      VERSK=SOMY(J)-SOMCY(J)
      DVAR=VERSK**2
      BRRK=DVAR + BRRK
   14 CONTINUE
      BRRDT=BRRK/(NOM*4)
      BRRD=(1.0)*BRRDT
      BRRD=SQRT(BRRD)
      WRITE(6,16) BRRD
   16 FORMAT ('DIE BRR-D STATISTIEK IS  ',F15.5)
      RETURN
      END
      SUBROUTINE JRR
      INTEGER NVAR,NOM
      REAL JRRK,JRRH,JRRH1,JRRS,JRRD,JRRC,JRRC1,JRRDT
      COMMON/SOMME/STOTY,STOTYY,STOTXY,STOT1,SOMY(300),SOMYY(300),NVAR,
     +SOMCY(300),SOMCYY(300),SOMCXY(300),SOMC1(300),
     +SOMXY(300),SOM1(300),TOTY,TOTYY,NOM
C
C     BEREKEN NOU DIE  JRR STATISTIEKE
C     X(I)=1-F/K*SOM(G(HI)-G(S))**2
C
      XVAR=0.0
      XCVAR=0.0
```

```
      JRRK=0.0                                                          BRR0962
      DO 7 J=1,NOM                                                      BRR0963
      SDIFF=SOMY(J)-TOTY                                                BRR0964
      SCDIFF=SOMCY(J)-TOTY                                              BRR0965
C     WRITE(6,9) TOTY,SOMY(J),SOMCY(J),SDIFF,SCDIFF                     BRR0966
      XVAR=(SDIFF**2) + XVAR                                            BRR0967
      XCVAR=(SCDIFF**2) + XCVAR                                         BRR0968
    7 CONTINUE                                                          BRR0969
      JRRH=XVAR                                                         BRR0970
      JRRH1=SQRT(JRRH)                                                  BRR0971
      WRITE(6,10) JRRH1                                                 BRR0972
      JRRC=XCVAR                                                        BRR0973
      JRRC1=SQRT(JRRC)                                                  BRR0974
      WRITE(6,12) JRRC1                                                 BRR0975
   10 FORMAT ('DIE JRR-H STATISTIEK IS  ',F15.5)                        BRR0976
   12 FORMAT ('DIE JRR-C STATISTIEK IS  ',F15.5)                        BRR0977
C                                                                       BRR0978
C     DIE JRR-S STATISTIEK WORD NOU BEREKEN                             BRR0979
C     JRRS=(JRRH+JRRC)/2                                                BRR0980
C                                                                       BRR0981
      JRRS=(JRRH+JRRC)/2                                                BRR0982
      JRRS=SQRT(JRRS)                                                   BRR0983
      WRITE(6,13) JRRS                                                  BRR0984
   13 FORMAT ('DIE JRR-S STATISTIEK IS  ',F15.5)                        BRR0985
C                                                                       BRR0986
C     DIE JRR-D STATISTIEK WORD NOU BEREKEN                             BRR0987
C                                                                       BRR0988
      DO 14 J=1,NOM                                                     BRR0989
      VERSK=SOMY(J)-SOMCY(J)                                            BRR0990
      DVAR=VERSK**2                                                     BRR0991
      JRRK=DVAR + JRRK                                                  BRR0992
   14 CONTINUE                                                          BRR0993
      JRRDT=JRRK/4                                                      BRR0994
      JRRD=JRRDT                                                        BRR0995
      JRRD=SQRT(JRRD)                                                   BRR0996
      WRITE(6,16) JRRD                                                  BRR0997
   16 FORMAT ('DIE JRR-D STATISTIEK IS  ',F15.5)                        BRR0998
      RETURN                                                            BRR0999
      END                                                               BRR1000
```