

STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM

MOKKEN R J

Enige ervaringen bij de ontwikkeling van
een standaardtoetsingsprobleem

S 360 (SP 93)

Statistica Neerlandica 19(1965), nr.4,
p 351-354.



1965

Enige ervaringen bij de ontwikkeling van een standaardtoetsingsprogramma *)

door R. J. Mokken **)

UDC 518.5:519.24

S u m m a r y

The continuous improvement in velocity and capacity in the construction of modern computers, the development of programming languages and the possibilities of exchange and wider application of programmes might easily lead the uninitiated to overoptimistic views and expectations considering the saving of time and even the elimination of programming effort.

This paper, scanning the possibilities and restrictions of the applications of exchanged programmes hopes to temper these views.

De stormachtige ontwikkeling op het gebied van de computers, qua reken-capaciteit en snelheid, in- en uitvoerapparatuur en vooral besturing per programmeertaal (ALGOL 60, FORTRAN, e.d.) kan gemakkelijk de indruk wekken, dat voor de verwerkings- en analysefasen van een onderzoeksprogramma tegenwoordig weinig tijd meer behoeft te worden uitgetrokken, wanneer de betrokken werkzaamheden kunnen worden verricht op een moderne rekenautomaat. Met name de opbouw van programma-archieven en de mogelijkheid tot uitwisseling van standaardprogramma's zouden onmiddellijke inschakeling van de rekenautomaat en besparing van de programmeertijd mogelijk maken.

Volgend betoog hoopt duidelijk te maken, dat ondanks de grote tijdsbesparingen die door de genoemde ontwikkeling in vele gevallen kunnen worden gerealiseerd, ook de sociaal-wetenschappelijke onderzoeker er op zal moeten rekenen dat vaak veel tijd en moeite nodig zullen zijn voordat zijn verwerkingsopdracht kan worden uitgevoerd. Dit betoog stoelt niet op een directe ervaring in programmering en bediening van een computer, doch geeft het standpunt en de ervaringen weer van de regelmatige consument van reken-capaciteit.

Het in de titel genoemde standaardtoetsingsprogramma heeft betrekking op een klapper met beschrijvingen van een groot aantal ALGOL-programma's die ten behoeve van de Afdeling Mathematische Statistiek van het Mathematisch Centrum werd samengesteld. Deze programma's worden zo geregeld in de consultatie van de afdeling toegepast, terwijl ook voor de toekomst een herhaald

*) Voordracht gehouden op de Statistische Dag 1965.

***) Wetenschappelijk medewerker Mathematisch Centrum en Seminarium voor de Leer der Communicatiemiddelen aan de Universiteit van Amsterdam.

gebruik mocht worden verwacht, dat het nuttiger leek ze in een standaardvorm voor meer algemeen gebruik geschikt te maken.

Allereerst kwamen uiteraard problemen aan de orde, die door hun omvang alleen bevredigend op de rekenautomaat kunnen worden opgelost. (Factoranalyse, multiple regressie, lineaire programmering, e.d.). Maar ook kwamen andere problemen in aanmerking waarvoor dat minder evident is, omdat zij gemakkelijk met behulp van een tafelrekenmachine zijn op te lossen. Men kan hierbij denken aan eenvoudige statistische toetsingsproblemen (χ^2 -toets voor onafhankelijkheid, parametrische en verdelingsvrije toetsingsproblemen voor één of meer steekproeven, rangcorrelatietoetsen, e.d.).

Wanneer dit eenvoudige rekenwerk veelvuldig voorkomt, zoals vaak in het sociaal-wetenschappelijk onderzoek, dan kan het lonen ook deze berekeningen met de computer uit te voeren. Het volume geeft hier de doorslag.

De in de klapper opgenomen programma's bestrijken dus de meest gebruikelijke procedures op het gebied van de multivariate en univariate statistische analyse.

Hoewel hij werd samengesteld ten behoeve van het eigen gebruik op de statistische afdeling van het Mathematisch Centrum, lenen de opgenomen programma's zich in principe ook tot een gebruik in wijdere kring. Hierbij werd gedacht aan onderzoekers, die voor hun statistische analyse gebruik wensen te maken van de rekenfaciliteiten van het MC, of gebruik kunnen maken van de diensten van een instelling, die is uitgerust met dezelfde rekenapparatuur.

De klapper bevat daartoe naast een schets van de gebruikte statistische procedure, een beschrijving van in- en uitvoer, terwijl het programma zo is ingericht, dat de operator het zonder kennis van het probleem op de machine kan uitvoeren.

Uiteraard bestrijkt deze verzameling programma's, die voornamelijk de statistische analyse van gegevens mogelijk maken, slechts een klein deel van de werkzaamheden waarvoor een onderzoeker op sociaal-wetenschappelijk gebied kan komen te staan. Geconfronteerd met een grote verscheidenheid van variabelen en statistische problemen en naar verhouding grote aantallen waarnemingen worstelt hij met verwerkingsproblemen, die vaak gekenmerkt worden door de relatief eenvoudige rekenbewerkingen en hun grote massa. Taken als het hercoderen van gegevens op ponskaarten, het converteren van uitvoer van een behaalde fase van mechanische verwerking tot invoer voor een andere fase, maar ook het tabelleren van materiaal vallen hieronder. Het feit, dat enkele instanties reeds beschikken over programma's voor eenvoudige problemen van enquête-verwerking, toont dat ook hier de ontwikkeling is ingezet.

De voordelen van het meer toegankelijk maken van programma's voor

ruimere toepassing worden duidelijk wanneer men bedenkt, dat ondanks de enorme vooruitgang bij de constructie van rekenmachines en de ontwikkeling van programmeertalen het programmeren van een enigszins gecompliceerd probleem toch de nodige tijd vraagt, zodat het in bepaalde gevallen wel één à twee maanden kan duren voordat een programma kan worden gebruikt.

De beschikbaarheid van een pasklaar programma betekent daarom een forse besparing van programmeertijd en voor de opdrachtgever dus van wachttijd.

De gebruikswaarde die een computer voor een bepaalde onderzoeker heeft, wordt daarom in niet geringe mate verhoogd, wanneer de "software" ervan een groot aantal voor hem bruikbare programma's bevat. Toch kent de waarde van een bepaald programma-archief voor die onderzoeker zijn beperkingen. Om verschillende redenen.

Standaardprogramma's leggen de oplossing van problemen die regelmatig terugkeren in een bepaalde vorm vast. Vaak echter komt men in de analyse te staan voor problemen waarvan de oplossing geheel of in essentiële details verschilt van de geprogrammeerde uitvoering. De aanpassing daarvan kan zoveel tijd en moeite vragen, dat men beter een nieuw programma kan (laten) schrijven.

Bovendien liggen standaardoplossingen zelden voor altijd vast, omdat zij met de wetenschappelijke vorderingen mee evolueren. Programma-archieven zullen bij die ontwikkeling altijd enigszins achterblijven. Geen programma-archief zal, naar het zich laat aanzien, ooit volledig de behoeften van een bepaald vakgebied volledig dekken. Dit houdt in, dat ook de sociaal-wetenschappelijke onderzoeker eventueel te rade moet gaan bij verschillende programma-archieven.

Hierbij staan hem verschillende wegen open. Allereerst kan hij elders de gewenste programma's opvragen met de bedoeling ze te gebruiken voor de machine waartoe hij toegang heeft. Dit is geen eenvoudige zaak.

Programma's zijn sterk machine-afhankelijk, ook als zij in een programmeertaal zijn gesteld, zodat men ze in het beschouwde geval hoogst zelden zonder meer kan gebruiken. Een opgevraagd programma moet vergezeld zijn van een uitvoerige beschrijving van invoer- en uitvoervereisten en andere gegevens waaruit die technische afhankelijkheid blijkt. Bovendien moet het vervat zijn in een programmeertaal, waarvoor men een compiler (vertaalprogramma) bezit. Zo niet, dan moet het worden omgezet in de programmeertaal, die men gebruikt, werk dat niet veel minder tijd vraagt dan het apart programmeren van het beschouwd probleem.

Is aan deze voorwaarden voldaan, en beschikt de onderzoeker over dezelfde machine, uitgerust met dezelfde hardware, als de programmaleverancier, dan kan het programma in principe meteen worden gebruikt.

Als echter, wat meer voorkomt, het programma geschreven werd voor een

machine van een ander type, of met een andere uitvoering, dan moet het worden aangepast. Zo moeten procedures eigen aan de „vreemde” machine via programmeertaaldefinities worden vertaald in equivalente procedures voor de eigen machine.

Bovendien zijn programmeertalen als ALGOL 60 zeer algemene talen, waarvan soms niet alle mogelijkheden op een bepaalde computer realiseerbaar zijn. De aldus opgelegde restricties variëren van machine tot machine, zodat het ene programma bijv. een andere ALGOL-restrictie kan bevatten dan het andere. Er bestaan dan ook ALGOL-dialecten: de ALGOL-groep in Duitsland; een zeer beperkte versie van ALGOL schijnt onder de naam SMALGOL bekend te staan.

Men zal dus ook moeten nagaan of de programmeertaal van het verkregen programma overeenkomt met de versie die men zelf gebruikt.

Al met al heeft een programmeur in de hier geschetste gevallen toch nog geruime tijd nodig om het verkregen programma te controleren.

De onderzoeker kan wellicht, als andere mogelijkheid, zijn gegevens laten analyseren daar waar het programma aanwezig is. Nog afgezien van communicatie-moeilijkheden kan dit vrij bewerkelijk worden. Wanneer men deelproblemen bij verschillende instellingen laat bewerken, moet men soms verscheidene malen een andere invoer, passend bij de relevante programma's klaarmaken en dat vraagt in het sociaal-wetenschappelijk onderzoek bij grote aantallen (multivariate) waarnemingen vaak veel arbeid.

In de praktijk zal men doorgaans wel trachten het probleem bij de instelling waar men over kan beschikken geprogrammeerd te krijgen. In aanmerking genomen de gebruikelijke zware belasting van de beschikbare programmeurs, moet men daarvoor de nodige tijd uittrekken.

Tenslotte, doch niet in de laatste plaats, moet op een andere mogelijkheid worden gewezen: de onderzoeker leert zelf programmeren. Dit is heel wel mogelijk. De moderne programmeertalen zijn mede met dat doel ontwikkeld. Ook dit vraagt echter tijd en moeite, terwijl zelfs een geroutineerde programmeur een behoorlijk programma niet uit de mouw schudt. Uit een oogpunt van optimale arbeidsverdeling kan dan het bezwaar naar voren komen, dat de onderzoeker, die desgewenst zijn tijd wel beter zal kunnen gebruiken, er op moet zijn voorbereid, heel wat tijd te besteden aan het programmeren van problemen, die hij op papier al heeft opgelost. Een vaak doorslaggevend voordeel is echter, dat de onderzoeker in dit geval voor de uitvoering van dit deel van zijn werk niet meer afhankelijk is van (doorgaans drukbezette) anderen.