

Functie	Junior Data Scientist / Missing Data Specialist
Naam	R.M.Schouten
Voornaam	Rianne
Geboren	11/05/1991
Geslacht	Vrouw
Woonplaats	Utrecht
Website	RianneSchouten.github.io



Motivatie

Vanwege een overvloed aan enthousiasme en doorzettingsvermogen en vanwege mijn leergierige karakter ben ik met name geschikt voor opdrachten gericht op de ontwikkeling van benodigde en geschikte analysemethoden. Mijn interesse ligt op het snijvlak van het bedenken en ontwikkelen van software applicaties en het toepassen van deze hulpmiddelen om maatschappelijke vraagstukken te beantwoorden.

Mijn specialisatie richt zich op missing data, een veelvoorkomend probleem voor iedereen die werkt met datasets. Omdat missende waarden kunnen zorgen voor onbetrouwbare uitkomsten van analyses en modellen, en omdat dit tot ongefundeerde beleidskeuzes kan leiden, is een goede omgang met missing data essentieel. Mijn onderzoek richt zich op de ontwikkeling en evaluatie van deze missing data methoden.

Opleidingen

- | | |
|-------------------|---|
| 04/2017 – nu | Excellence Program for Data Science – BigData Republic
<i>1 dag per week een training van Senior Data Scientists van BigData Republic</i>

Het doel van het traject is om te leren hoe ik mijn vaardigheden op het gebied van data science in kan zetten om bedrijfsprocessen te optimaliseren en verbeteren. Enerzijds leer ik meer over IT-engineering (zoals asana, python, linux, git, virtual machines, SQL), data science technieken (zoals statistische modellen, machine learning technieken, clustering, feature engineering, visualisatie, neurale netwerken, image processing) en het toepassen van deze methoden op use cases. Anderzijds ontwikkel ik mij tot een professionele consultant die bedrijven kan helpen in het vertalen van de businesswens naar data science oplossingen.

<i>More info: www.dpadatascience.nl</i> |
| 09/2014 – 05/2017 | Research Master Methodology and Statistics for the Behavioural, Biomedical and Social Sciences, Universiteit Utrecht
<i>Master of Science (MSc), GPA: 4.0</i>

Tijdens dit masterprogramma heb ik geleerd welke methodologische en statistische aspecten van belang zijn voor het goed uitvoeren van wetenschappelijk onderzoek. Ik heb kennis vergaard over vele statistische analysemethoden, de assumpties die eraan voorafgaan en hoe statistiek gebruikt kan worden om allerlei soorten vragen te beantwoorden. Met name ben ik enthousiast over de wiskundige achtergrond van deze methoden en hoe je met behulp van wiskundige formules de computer het werk kunt laten doen. |
| 09/2013 – 08/2014 | Preparatory Programme Social Policy and Social Interventions, Universiteit Utrecht
<i>30 ECTS</i> |
| 02/2013 – 06/2013 | Preparatory Programme Social Sciences, University van Amsterdam |



30 ECTS

09/2009 – 08/2012

Bachelor Geneeskunde, University of Utrecht
Bachelor of Science (BSc), GPA: 3.24

Alhoewel ik bij het starten van deze opleiding niet beseftte dat ik eigenlijk geen arts wilde zijn, heb ik het bachelor programma met plezier gevolgd. Ik heb geleerd hoe het menselijk lichaam werkt en welke onderliggende natuurkundige en scheikundige aspecten een belangrijke rol vervullen in een gezond lichaam. Bovendien heb ik mij ontwikkeld op het gebied van communicatieve en sociale vaardigheden.

Onderzoek

01/2016 - nu

Department of Methodology and Statistics, Universiteit Utrecht
Missing Data Research

Eenieder die wetenschappelijk onderzoek verricht krijgt te maken met missende data. Het ontbreken van gegevens is een probleem voor statistische analysemethoden, aangezien deze methoden meestal alleen werken met complete datasets. Daarom ontwikkelen we methoden die de geobserveerde data optimaal gebruiken om de missende waarden te schatten. Dit noemen we imputatie.

Voordat nieuwe imputatietechnieken gelanceerd kunnen worden, moeten we ze testen. Dit doen we door middel van een simulatiestudie. In zo'n studie simuleren we een complete dataset, vervolgens schieten we daar gaten in, en dan passen we de ontwikkelde missing data methode toe. Door de toepassing te vergelijken met de oorspronkelijke complete dataset, weten we of een imputatietechniek goed werkt.

Ik heb een methode ontwikkeld die op een valide en legitieme manier missende waarden creëert in een complete dataset. Deze multivariate amputatiemethode heb ik geïmplementeerd in een R-functie, welke inmiddels is getest en vrijgegeven is. Ik heb uitgebreide simulaties gedaan om te laten zien in welke situaties mijn functie beter werkt dan de huidige amputatietechnieken. Met de multivariate methode worden de missende waarden gecreëerd zoals je intuïtief verwacht en wenst.

Voor een draft van mijn onderzoek, zie: [Manuscript MultivariateAmputation](#)
Voor een uitleg van R-functie `ampute`, zie: [Vignette Ampute](#)

Ik heb mijn onderzoek gepresenteerd op de UseR!2017 conferentie. Voor een samenvatting van mijn talk, zie: [UseR!2017Talk RianneSchouten](#)

11/2016 - 12/2016

Department of Statistics, Columbia University in the City of New York
Staff Associate

Gedurende twee maanden heb ik dr. Gerko Vink geassisteerd bij zijn onderzoek op het gebied van missing data. We waren onderdeel van de STAN-groep, een team van software developers onder leiding van prof. Andrew Gelman. Mijn werk varieerde van het reviewen van literatuur tot het ontwikkelen van R-code en het doen van simulaties. Mijn samenwerking met dr. Gerko Vink heeft geleid tot mooie inzichten, die we binnenkort hopen te publiceren.

09/2015 – 12/2015

Dutch Institute for Space Research (SRON)
Internship Material Science, Grade: 9.0
The Use of Principle Component Analysis in Analyzing Fe-55 X-Ray Pulses



Bij SRON heb ik me verdiept in de wiskundige techniek achter PCA en deze toegepast op hun experimentele datasets. Voor de ontwikkeling van een telescoop voor de ATHENA-missie waren de onderzoekers bij SRON op zoek naar een statistische methode die de temperatuur van X-ray photons zo precies mogelijk weergaf. Ik heb de traditionele methode van matched filtering vergeleken met de nieuwe methode van PCA. Het werk heeft een SPIE conference paper opgeleverd ([directe link](#)):

de Vries, C. P., Schouten, R. M., van der Kuur, J., Gottardi, L., & Akamatsu, H. (2016, July). Microcalorimeter pulse analysis by means of principle component decomposition. In *SPIE Astronomical Telescopes+ Instrumentation* (pp. 99055I-99055I). International Society for Optics and Photonics.

Software tools

R (incl. Rmarkdown)

>3 jaar

Voor de ontwikkeling van R-functie `ampute` heb ik geleerd hoe officiële R-code eruit ziet, aan welke eisen die moet voldoen, hoe je een classificatie meegeeft aan objecten, hoe je documentatie ontwikkelt met RMarkdown en Package Knitr, hoe je een vignette vormgeeft en nog meer.

Python

1 jaar

Gedurende mijn tijd bij SRON heb ik gewerkt met Python. Alhoewel ik niet getraind was in het gebruiken van Python, maakte ik mij de taal gemakkelijk eigen, zonder extra hulp. Bekende modules zoals NumPy, SciPy en ScikitLearn heb ik gebruikt om mijn analyses te kunnen uitvoeren. Door de trainingen van BDR kan ik mij nu verder ontwikkelen.

LateX

2 jaar

Mijn scriptiewerk, twee artikelen en verschillende presentaties heb ik gemaakt met LateX. Ik vind het een zeer prettig middel om mijn werk overzichtelijk te presenteren.

SPSS

>3 jaar

Voor het uitvoeren van statistiek is SPSS de basis van alle software tools. Niet alleen weet ik goed hoe het werkt, ik heb ook meerdere malen les gegeven aan studenten.

Andere statistische software

2 jaar

Gedurende mijn opleiding leerden we hoe sommige analyses het beste uit te voeren zijn met specifieke software, zoals multilevel analyse met HLM, structural equation modeling met MPlus en Bayesiaanse statistiek met OpenBugs.

Werkervaring

09/2014 – 05/2017

Self-employed statistical consultant for WO-students

Vele bachelor, master en PhD studenten hebben problemen met het uitvoeren van hun statistische analyses, vaak simpelweg omdat ze er niet goed les in krijgen. Ik heb voor menig student gefunctioneerd als adviseur, leraar, gids en zelfs als co-auteur voor een wetenschappelijk artikel ([link](#)).

01/2012 – 05/2012

Student assistant Universiteit Utrecht

Courses:

Survey Research: Design, implementation and data processing

Survey Research: Statistical analysis and estimation

Introduction to research methods and statistics for social scientists'



Als student assistent was ik verantwoordelijk voor twee dingen. Allereerst zorgde ik voor de praktische zaken: alle materialen zijn op orde, de studenten weten waar ze moeten zijn, de koffie en thee is geregeld, en ik beantwoord willekeurige vragen over waar de supermarkt is en waar een fiets te huren is (het waren veelal buitenlandse studenten). Anderzijds was ik inhoudelijk betrokken, ik organiseerde sommige van de practica of ik assisteerde tijdens de practica en legde de vragen en opdrachten aan de studenten uit. Mijn duidelijke uitleg en vriendelijke houding heeft me een goede beoordeling opgeleverd.

09/2012 – 02/2013

Bloed prikken bij Saltro Diagnostics Center

Met de auto reed ik langs de patiënten die te oud of te ziek waren om naar de prikpost te komen. Het was altijd weer een uitdaging om al het bloed op tijd verzameld en bij het lab te krijgen.

Talenkennis

Taal	Woord	Schrift
Engels Nederlands	vloeiend moedertaal	vloeiend moedertaal