

# Gewenste kennis en vaardigheden bij analisten op het gebied van data- analyse en dataopslag

---

Resultaten vragenlijsten werkveld 2019 en 2020

## RAPPORTAGE

Auteur: projectgroep 'Datavaardige professional'  
Datum: maart 2021

Domein Applied Science (DAS)  
Postbus 249  
2260 AE Leidschendam  
info@appliedscience.nl

© Domein Applied Science

## Index

---

Inleiding	3
Hoofdstuk 1: Inventarisatie werkzaamheden hbo'ers op het gebied van data-analyse en dataopslag (resultaten 2019)	4
Hoofdstuk 2: Gewenst niveau kennis en vaardigheden hbo'ers op het gebied van data-analyse en dataopslag (resultaten 2020)	7
Bijlage 1: Vragenlijsten 2019	16
Bijlage 2: Vragenlijsten 2020	20

---

## Inleiding

---

Eind 2017 vond de DAS-bijeenkomst 'Professional van de Toekomst' plaats. Centraal stond de vraag: 'Hoe bereiden wij, als DAS-opleidingen, onze studenten voor op dit toekomstige werkveld?'. Een waardevolle bijeenkomst die heeft geleid tot een onderzoek onder de DAS-hogescholen naar het huidige aanbod van het data-onderwijs, eventuele groeigebieden binnen de opleidingen en de wensen die de opleidingen hebben qua ondersteuning vanuit het Domein Applied Science.

De uitkomsten van dit onderzoek zijn opgenomen in het rapport 'Datavaardige professional' dat eind 2018 verscheen. In dit rapport stonden ook adviezen, namelijk:

- Aandacht vergroten voor digitale geletterdheid in de curricula, zodat communicatie binnen multidisciplinaire teams op het gebied van datageneratie, analyse en opslag soepel kan verlopen.
- Faciliteren van uitwisseling tussen met name BML-docenten op het gebied van data-onderwijs, aangezien in dit vakgebied de grootste noodzaak wordt ervaren om afgestudeerden datavaardigheid mee te geven.
- Stimuleren van het werkveld om specifieke wensen op dit gebied duidelijk naar de opleidingen te communiceren.
- Het onderwijs, in samenwerking met het werkveld, voorzien van een geschikte dataset om onderwijs mee te ontwikkelen.
- Data stewardship in beperkte mate opnemen in de onderwijsprogramma's.

In juni 2019 werd de DAS-projectgroep 'datavaardige professional' opgericht en is met deze adviezen aan de slag gegaan. Dit heeft er onder andere toe geleid dat in het najaar van 2019 een inventarisatie is gehouden onder werkgevers en alumni met als doel inzicht te krijgen in wat de behoeftes in het werkveld zijn aan kennis en vaardigheden bij analisten op het gebied van data-analyse en dataopslag. De uitkomsten zijn gebruikt om te onderzoeken of het huidige data-onderwijs in de Applied Science-opleidingen voldoende aansluit bij de behoeftes in het werkveld.

De inventarisatie gaf een goede eerste indruk van de wensen vanuit het werkveld, maar er bleek behoefte aan meer details. In het najaar van 2020 werd dan ook een vervolgvragenlijst opgesteld met als doel in kaart te brengen op welk niveau deze gewenste kennis en vaardigheden door net-afgestudeerden beheerst moeten worden. Hierbij was het niet een vereiste dat de participant de vragenlijst van 2019 had ingevuld.

In totaal hebben 133 werkgevers en 39 alumni de vragenlijst in 2019 ingevuld. De vervolgvragenlijst in het najaar van 2020 vulden 124 werkgevers in.

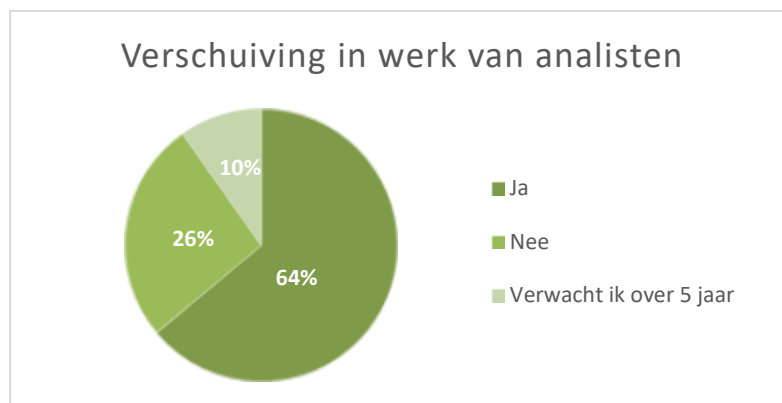
In de bijlagen zijn de opgestelde vragen van zowel de vragenlijst 2019 als 2020 te vinden.

# Hoofdstuk 1: Inventarisatie werkzaamheden hbo'ers op het gebied van data-analyse en dataopslag (resultaten 2019)

In 2019 werd een vragenlijst onder werkgevers en alumni verspreid met als doel de huidige situatie en de verwachte veranderingen op het gebied van werkzaamheden van hbo-afgestudeerden wat betreft data-analyse en dataopslag te inventariseren. In dit hoofdstuk een overzicht van de resultaten.

## 1.1. Is er een verschuiving van het werk van een analist van experimenteel werk naar data-analyses?

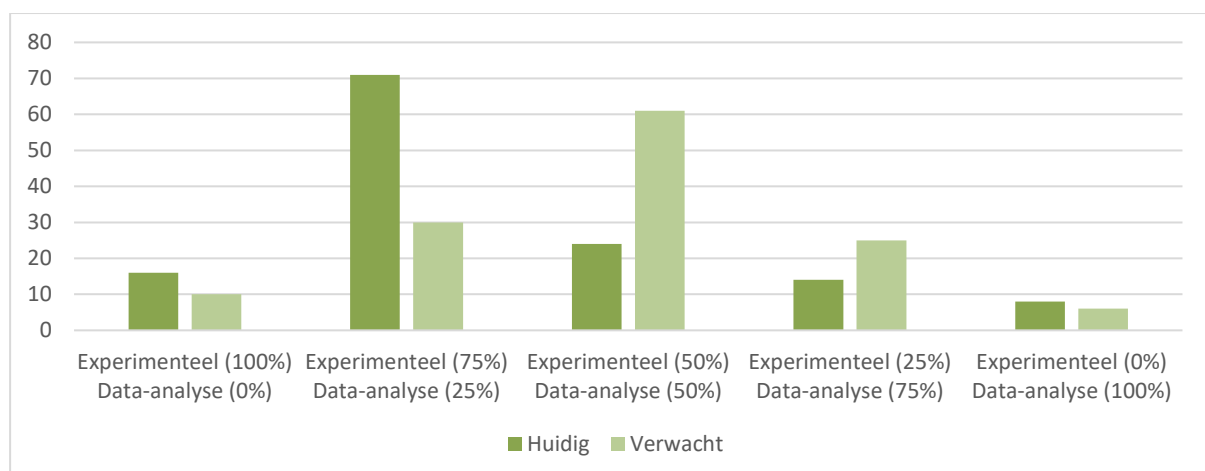
Zoals in figuur 1 weergegeven, geeft 64% van de respondenten aan dat zij op dit moment (2019) een verschuiving in het werk van een analist ervaren van experimenteel werk naar data-analyses. Daarnaast geeft 10% aan dat ze verwachten dat deze verschuiving in werkzaamheden binnen 5 jaar plaatsvindt.



**Figuur 1:** Percentage respondenten dat een verschuiving van experimenteel werk naar data-analyse ziet voor analisten

## 1.2. Hoe groot is de verschuiving?

In de huidige situatie (2019) is bij het merendeel van de respondenten de verhouding in tijdsinspanning van een analist tussen experimenteel werk en data-analyse: 75% experimenteel en 25% data-analyses. Zij verwachten echter een verschuiving binnen vijf jaar naar 50% experimenteel en 50% data-analyses (figuur 2).



**Figuur 2:** Verhouding tijdsinspanning van een analist tussen experimenteel werk en data-analyse in huidige situatie en verwacht over 5 jaar (in aantal werkgevers)

### 1.3. Dataverwerking: wat doet de analist en wat doet de data-expert?

Hieronder een aantal antwoorden dat werd gegeven op de vraag: Welk deel van de dataverwerking wordt uitgevoerd door analisten en vanaf welk punt wordt de analyse overgenomen door data-experts?

- Analisten genereren de data, voeren de kwaliteitscontrole uit en verwerken de data voor interpretatie. Verdere analyses worden uitgevoerd door experts.
- In hoeverre een analist betrokken is bij de dataverwerking hangt af van de expertise van de analist en de analyses.
- Dit hangt sterk af van hoe nieuw de testen/processen zijn waar de data uit verkregen zijn en hoe bekend de analist is met dit type werk.
- Een analist zou prima een stukje van de data-analyse kunnen doen als het niet te complex is. Dan is enige kennis van statistiek en datavisualisatie erg handig en welkom.

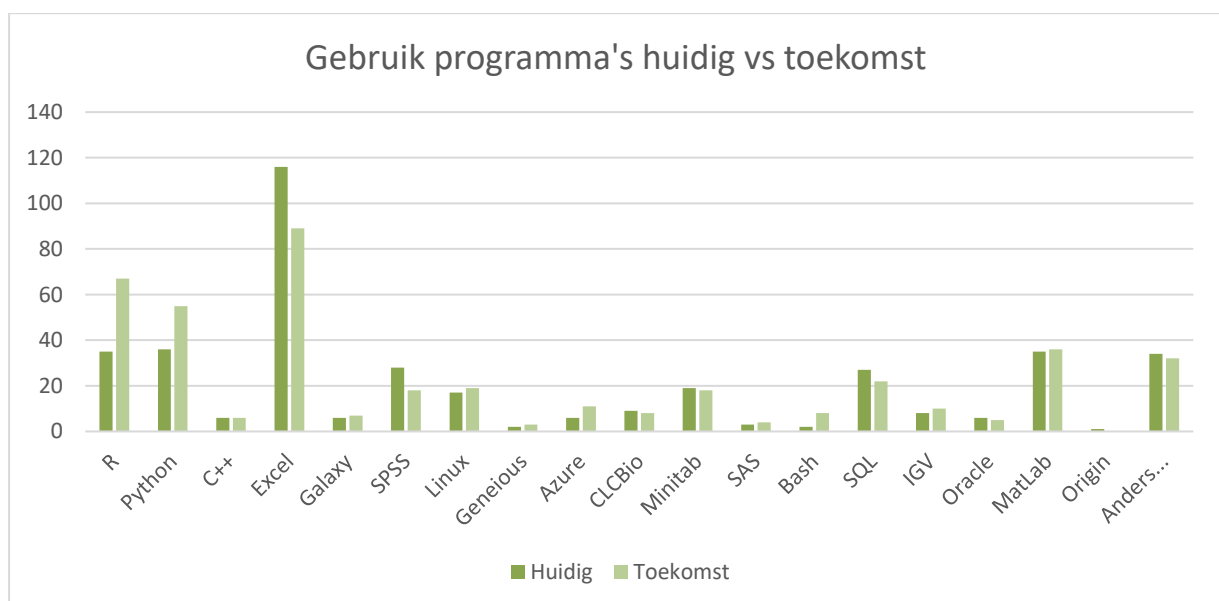
### 1.4. Alumni: wat doet hij/zij zelf en wat een data-expert?

Alumni werd gevraagd om een beeld te geven van de huidige werkzaamheden (2019) van hbo-afgestudeerden op het gebied van data-analyse en dataopslag. Hieronder een aantal omschrijvingen van de werkzaamheden:

- Ordenen van data, figuurvorming en simpele (statistische) analyses worden uitgevoerd door de analist, alles daarboven gebeurt door de experts. (research analist)
- Controleren van de initiële data: komt het monster overeen met de standaarden. Mocht dit niet het geval zijn, gaat het richting de experts. (QC-analist)
- Voor reguliere testen doen de analisten de data-analyses, zoals het bepalen van gemiddelde, standaarddeviatie en %RSD. (associate scientist)
- Statistiek en het maken van grafieken doet de analist, de analyse van sequencing data wordt deels gedaan door bio-informaticus. (assistent onderzoeker)

### 1.5. Gebruikte programma's versus verwacht toekomstig gebruik programma's

In de huidige situatie (2019) is Excel het belangrijkste programma voor de meeste werkgevers bij het uitvoeren van analyses. Daarna volgen R, Python, MatLab en SQL. De respondenten verwachten dat in de toekomst het gebruik van Excel zal afnemen en vooral de inzet van R en Python zal toenemen, zie figuur 3. Sommige respondenten gaven bij 'anders' aan dat naar verwachting in de toekomst er ook programma's op maat gemaakt zullen worden.



**Figuur 3:** Programma's en programmeertalen die op dit moment het belangrijkste zijn voor het werk van een analist vs de belangrijkste programma's en programmeertalen in de toekomst (over 5 jaar) (in aantal werkgevers)

---

### 1.6. Wat missen werkgevers qua kennis en vaardigheden

Om het huidige onderwijs op het gebied van data tegen het licht te houden van de gewenste kennis en vaardigheden in het werkveld, werd de vraag gesteld: Zijn er vaardigheden op het gebied van gegevensverwerking en -analyse die (recent) afgestudeerden op dit moment missen? Zo ja, wat zijn volgens u de belangrijkste? Daarbij kwamen de volgende punten naar voren:

- Vooral statistische kennis kan versterkt worden. Daarbij gaat het niet alleen om de analyse achteraf maar ook voor het opzetten van een Experimental Design (DoE).
- Vergroten programmeervaardigheden (R, Python).
- Verbeteren hoe om te gaan met data: waar sla je iets op, welke codering geef je mee om herleidbaarheid te behouden.
- Veel betere Excel-vaardigheden.

### 1.7. Wat hebben alumni gemist tijdens hun opleiding?

Alumni werd gevraagd of zij nog tips hebben qua aanvullingen en/of aanpassingen in het huidige curriculum van de door hun gevolgde opleiding op het gebied van data-analyse? Hieruit kwamen de volgende punten naar voren:

- Zorg voor mooie praktijkvoorbeelden zodat de waarde van data-analyse duidelijk wordt in een vroeg stadium van de opleiding.
- Meer validaties uitvoeren.
- In research wordt meer en meer verwacht dat je ook big data kan analyseren.
- Ik miste een vak statistiek, het deel in Excel was te beperkt.
- Excel is niet meer voldoende, andere vaardigheden in andere geschikte tools (bv R) zijn handig dan wel noodzakelijk.
- Een betere basis voor het begrijpen van statistische data in wetenschappelijke artikelen.

## Hoofdstuk 2: Gewenst niveau kennis en vaardigheden hbo'ers op het gebied van data-analyse en dataopslag (resultaten 2020)

In dit hoofdstuk zijn de resultaten van de vervolgvragenlijst die in het najaar van 2020 is verspreid, per vraag en per opleiding, weergegeven. Dit hoofdstuk geeft de resultaten weer, het is aan de lezer de resultaten naar eigen inzicht te interpreteren. Doordat de vragenlijsten dit keer opleidingspecifiek waren, komen niet alle vragen voor bij alle opleidingen. Wanneer dit het geval is, is dit aangegeven.

### 2.1. Bij welk soort organisatie bent u werkzaam?

Met deze vraag wilden we in kaart brengen welk soort organisaties de vragenlijst hebben ingevuld. Zoals tabel 1 weergeeft zijn het voor de opleidingen Applied science, Chemie en Chemische technologie voornamelijk profit organisaties geweest en voor de opleidingen BML en Biotechnologie was het een mix van ziekenhuizen, universiteiten en profit organisaties.

**Tabel 1:** Overzicht soort organisaties die de vragenlijst hebben ingevuld

	Ziekenhuis	Universiteit	KNAW of onderzoekslab	Profit organisatie	Anders	Totaal
<b>AS</b>	4	3	2	15	6	30
<b>BML/BT</b>	19	12	2	13	6	52
<b>CH/CT</b>	0	5	1	33	3	42

### 2.2. Hoe groot is uw organisatie?

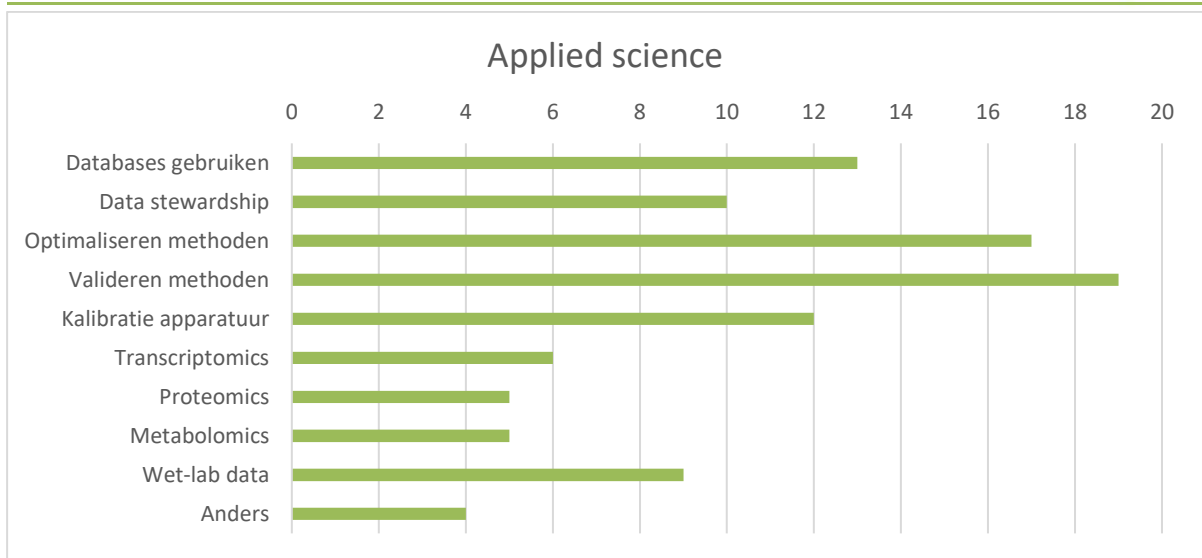
Zoals tabel 2 weergeeft hebben voornamelijk organisaties met meer dan 50 medewerkers meegewerkt aan de vragenlijst.

**Tabel 2:** Overzicht van de grootte van de organisaties die de vragenlijst hebben ingevuld

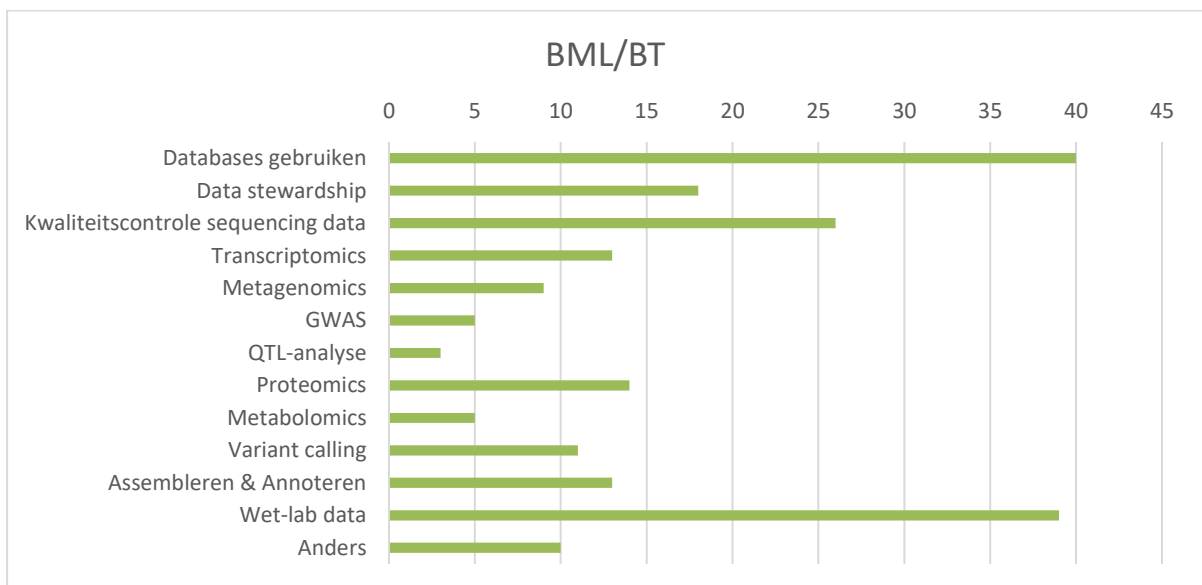
	< 50 medewerkers	> 50 medewerkers	Totaal
<b>AS</b>	2	28	30
<b>BML/BT</b>	8	44	52
<b>CH/CT</b>	7	35	42

### 2.3. Welke typen analyses moeten net-afgestudeerde hbo'ers beheersen?

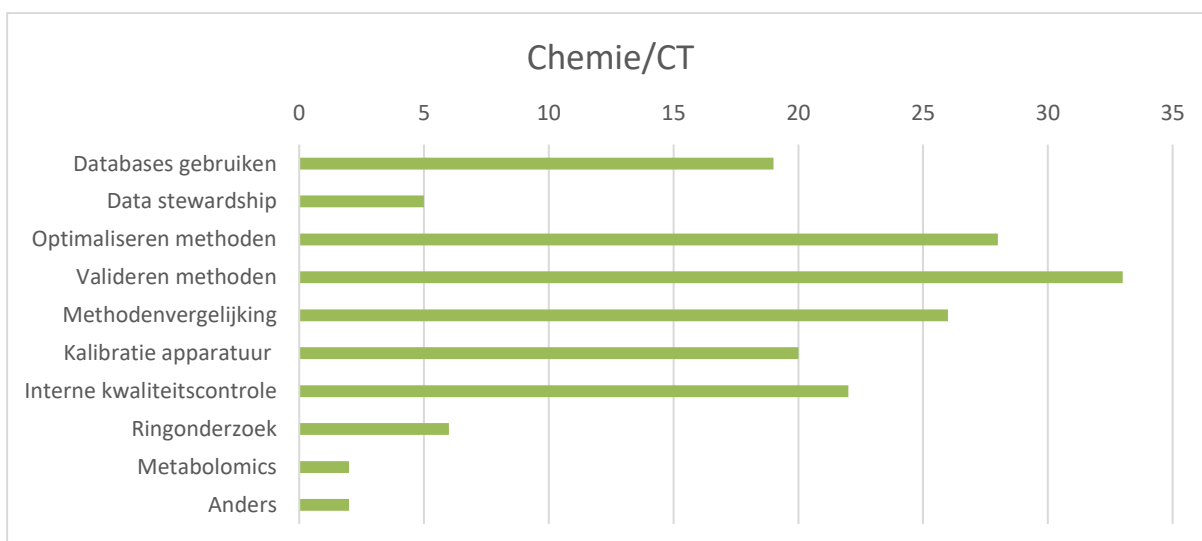
Om inzicht te krijgen in welke typen analyses net afgestudeerde hbo'ers moeten beheersen, konden de respondenten kiezen uit een vooropgestelde lijst met typen analyses. In de figuren 4, 5 en 6 staan de gegeven antwoorden per opleiding weergegeven in aantallen.



**Figuur 4:** De gewenste typen analyses die net-afgestudeerden in de Applied science opleiding moeten beheersen



**Figuur 5:** De gewenste typen analyses die net-afgestudeerden in de BML/BT opleidingen moeten beheersen

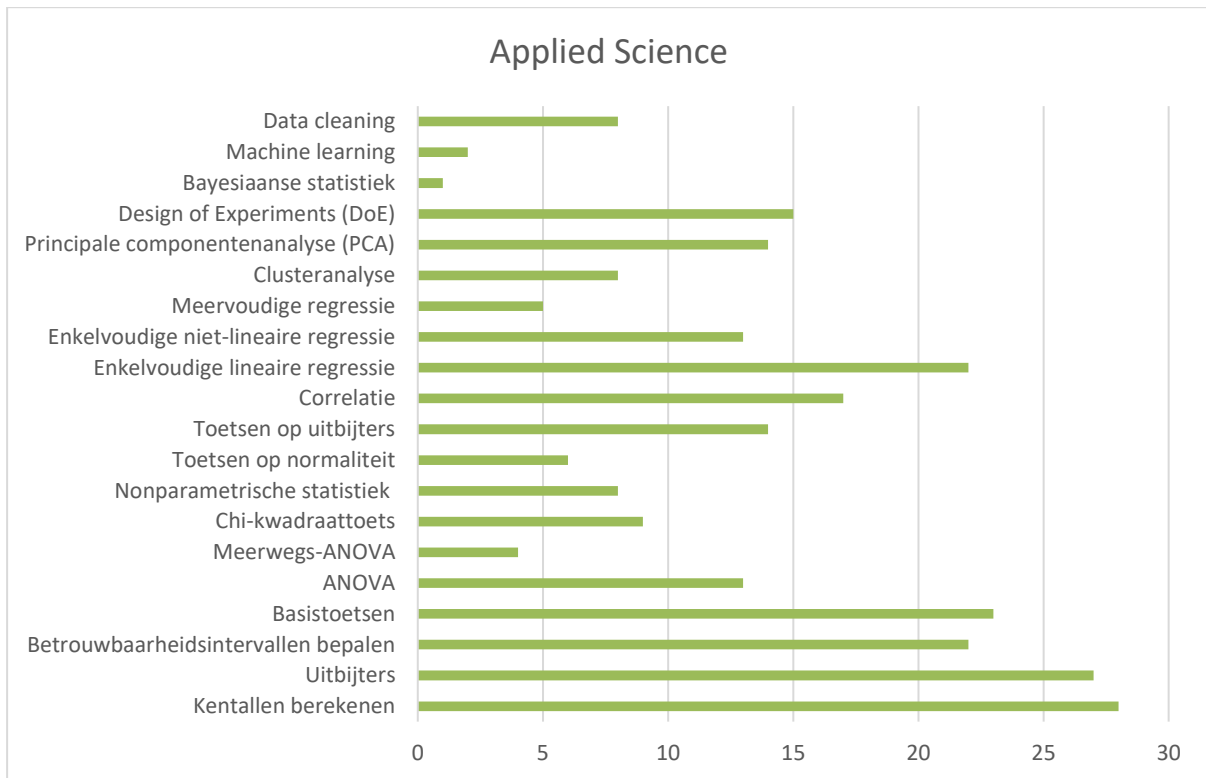


**Figuur 6:** De gewenste typen analyses die net-afgestudeerden in de Chemie/CT opleidingen moeten beheersen

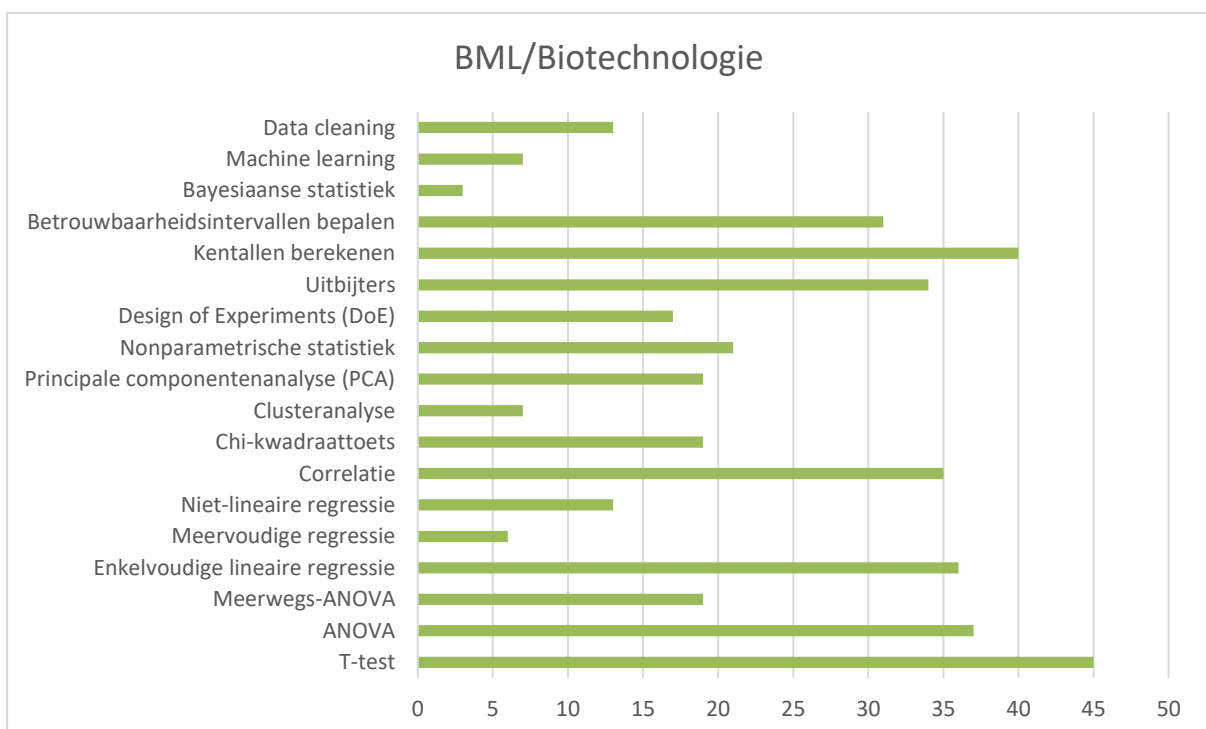


## 2.4. Welke onderdelen op het gebied van statistiek moeten net-afgestudeerde hbo'ers beheersen?

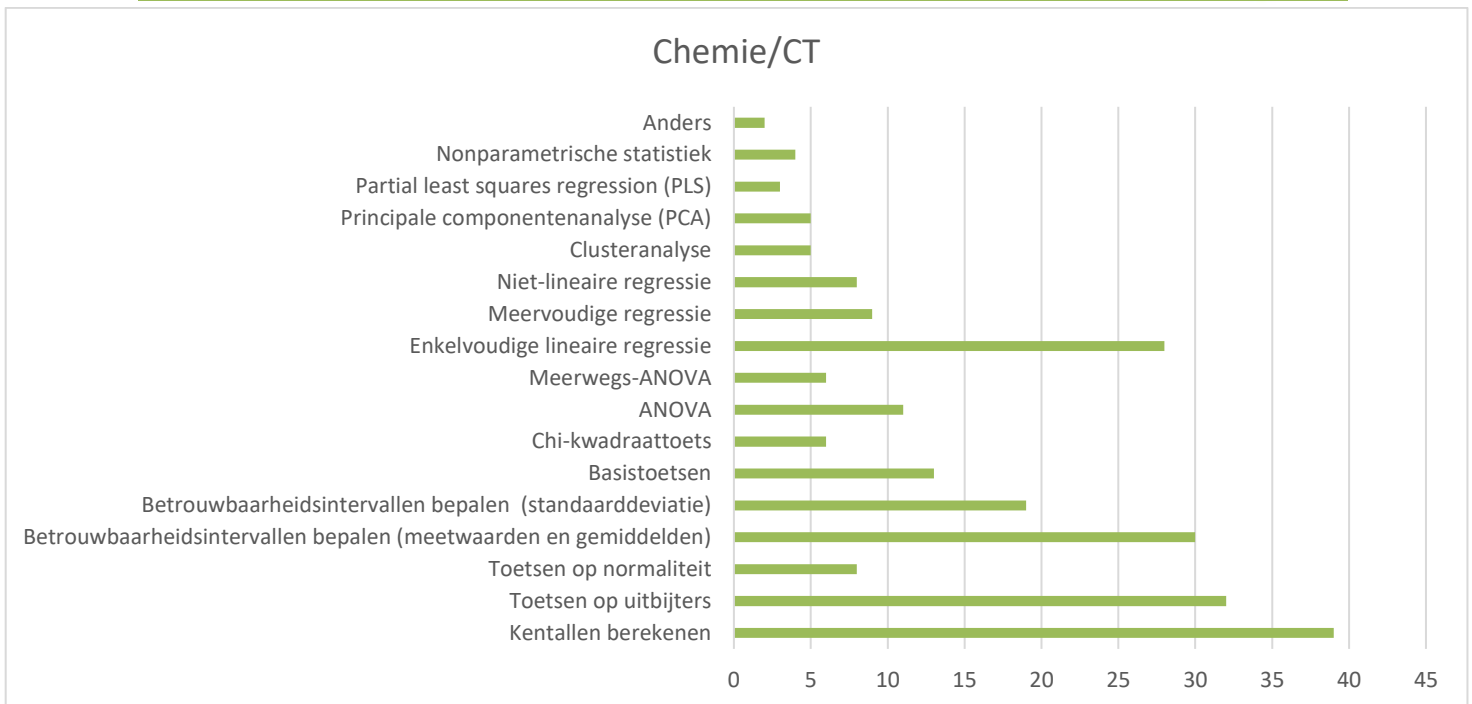
Om een zo goed mogelijke aansluiting bij de werkzaamheden in het werkveld te realiseren, stelden we de vraag welke onderdelen op het gebied van statistiek net-afgestudeerde hbo'ers minimaal moeten beheersen. In figuren 7, 8 en 9 de gegeven antwoorden in aantallen.



**Figuur 7:** De gewenste statistiekkennis bij net-afgestudeerden Applied science (in aantallen)



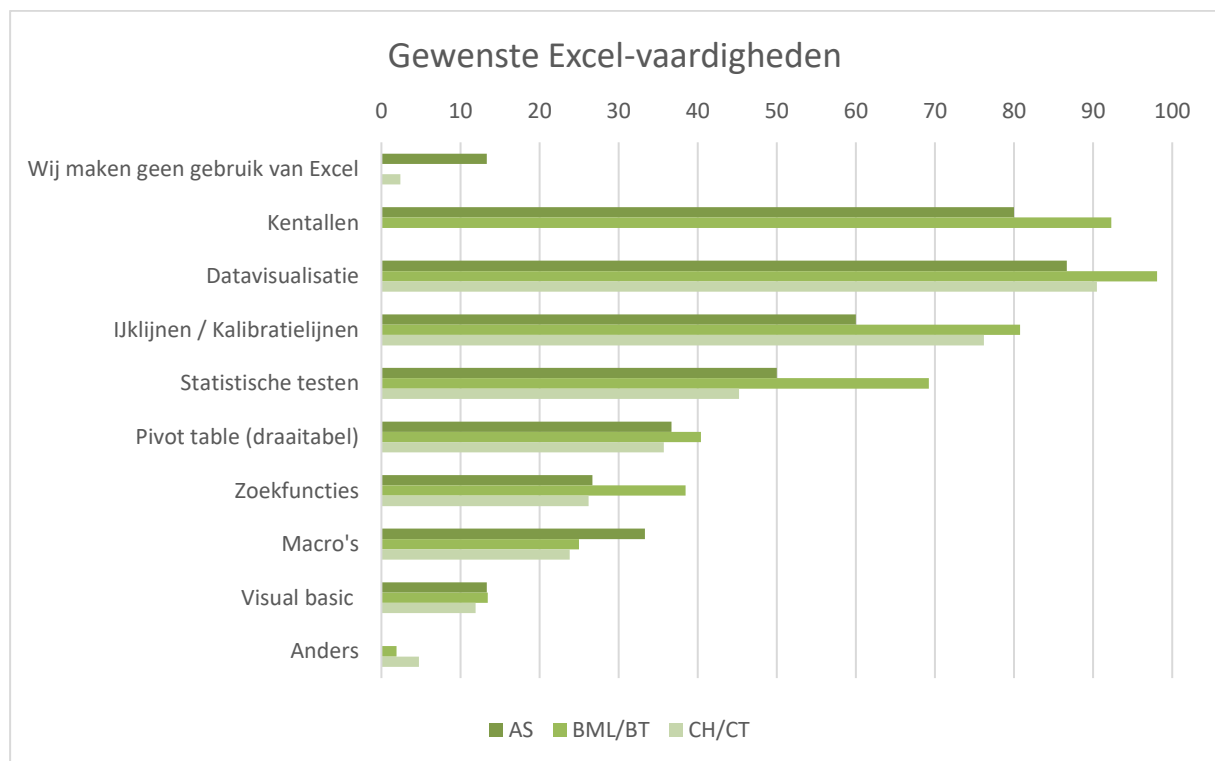
**Figuur 8:** De gewenste statistiekkennis bij net-afgestudeerden BML en Biotechnologie (in aantallen)



**Figuur 9:** De gewenste statistiekennis bij net-afgestudeerden Chemie en Chemische technologie (in aantallen)

## 2.5. Welke Excel-vaardigheden moeten net-afgestudeerde hbo'ers beheersen?

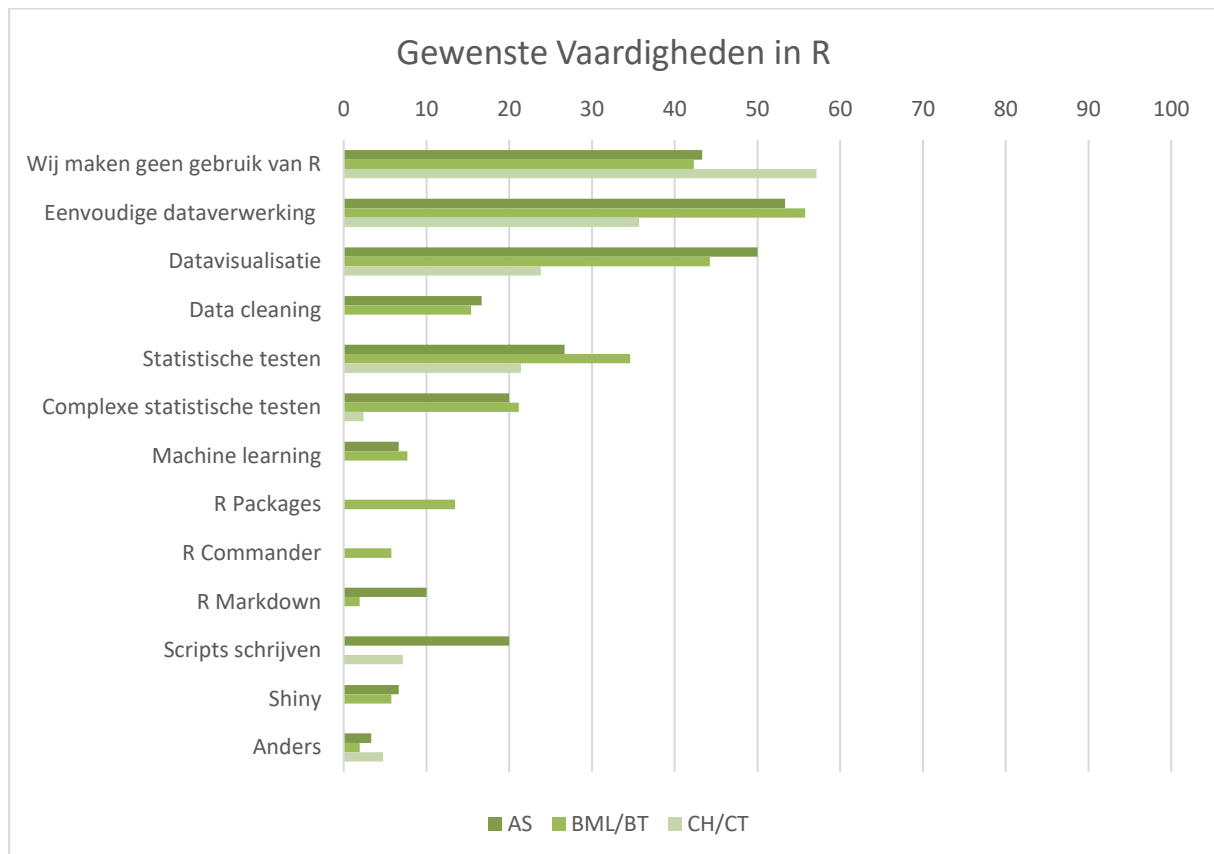
Om meer inzicht te krijgen in het gewenste beheersingsniveau van het programma Excel, is gevraagd welke Excel-vaardigheden net-afgestudeerde hbo'ers minimaal moeten beheersen om een goede aansluiting te hebben op de werkzaamheden in het werkveld. In figuur 10 staan de gegeven antwoorden per opleiding in percentage van het aantal respondenten per opleiding.



**Figuur 10:** Gewenste Excel-vaardigheden bij net-afgestudeerde hbo'ers (in percentage van het aantal respondenten) voor de opleidingen Applied science, BML, Biotechnologie, Chemie en Chemische technologie  
 Noot: In de vragenlijst chemie/CT was de optie kentallen niet meegenomen.

## 2.6. Welke R-vaardigheden moeten net-afgestudeerde hbo'ers beheersen?

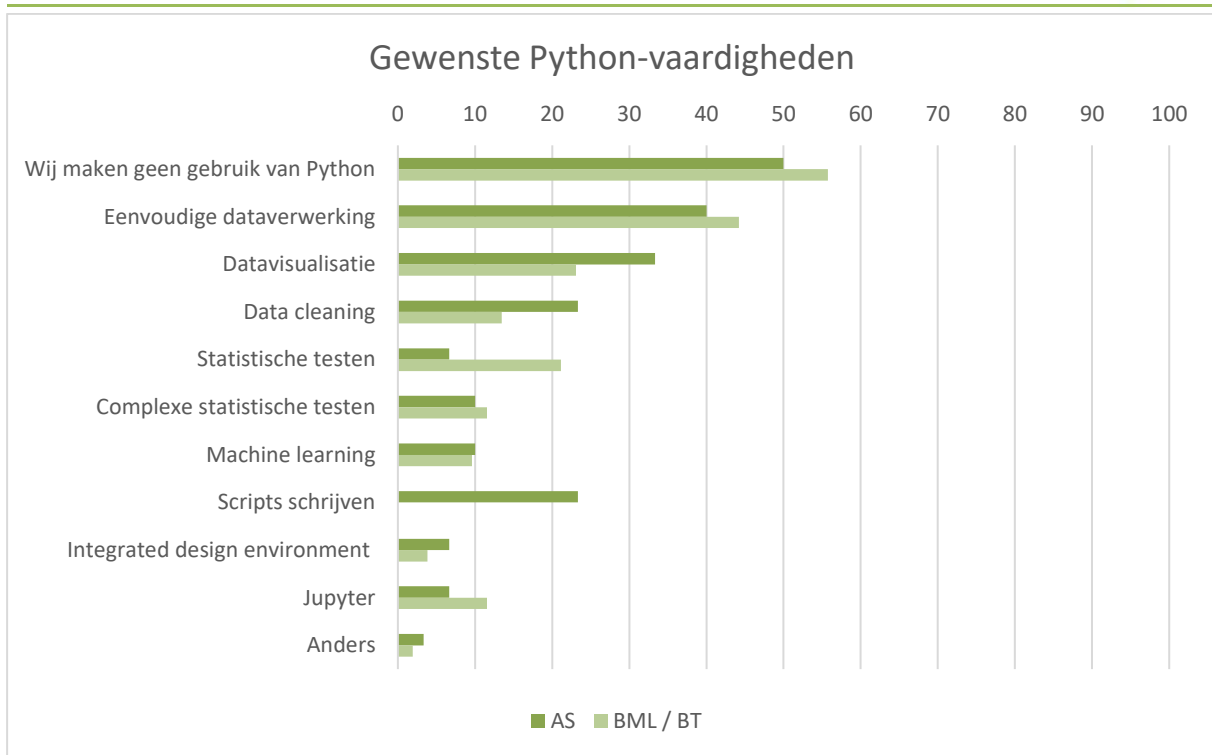
Om inzicht te krijgen in het gewenste beheersingsniveau van de programmeertaal R, stelden we de vraag welke R-vaardigheden net-afgestudeerde hbo-studenten minimaal moeten beheersen voor een goede aansluiting op het werkveld. In figuur 11 staan de gegeven antwoorden per opleiding in percentage van het aantal respondenten per opleiding.



**Figuur 11:** Gewenste R-vaardigheden bij net-afgestudeerde hbo'ers (in percentage van het aantal respondenten) voor de opleidingen Applied science, BML, Biotechnologie, Chemie en Chemische technologie  
*Noot: In de vragenlijst van Applied science was de optie R Packages niet opgenomen, in de vragenlijst voor BML/Biotechnologie was de optie Scripts schrijven niet meegenomen en in de vragenlijst voor Chemie/Chemische technologie de opties Data cleaning, Machine learning, R Packages en Shiny niet.*

## 2.7. Welke Python-vaardigheden moeten net-afgestudeerde hbo'ers beheersen?

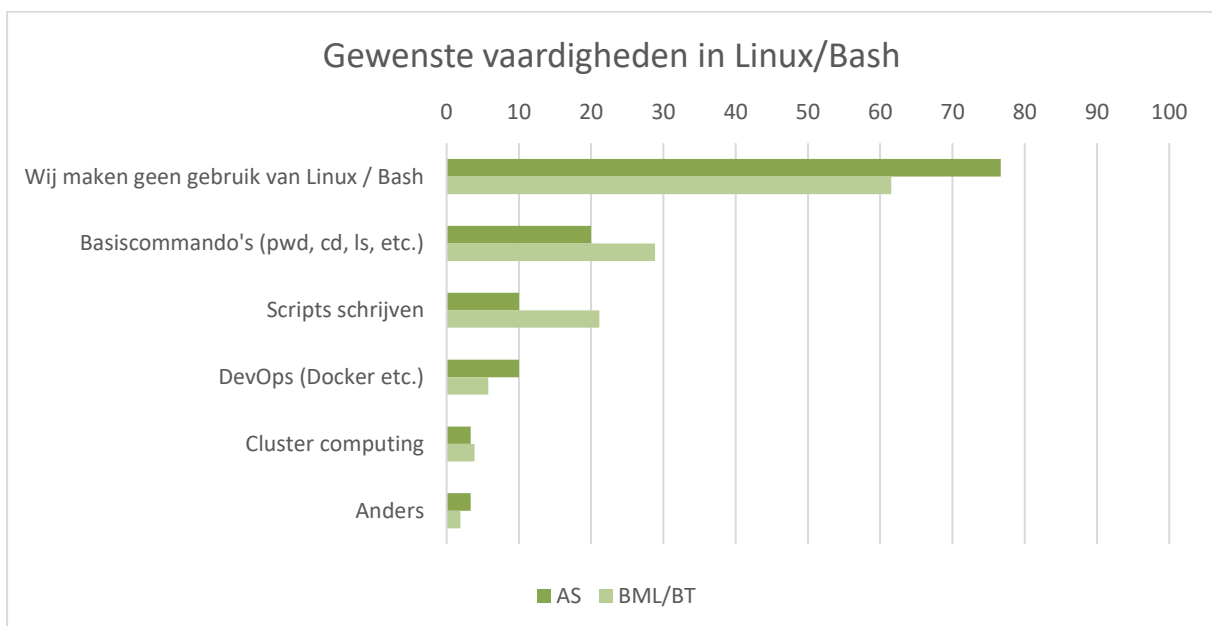
Aangezien in de chemie bijna tot geen gebruik wordt gemaakt van Python is de vraag naar de gewenste vaardigheden met Python bij een net-afgestudeerde hbo'er niet opgenomen in de vragenlijst voor chemie en chemische technologie. In figuur 12 staan de resultaten weergegeven van de opleidingen Applied science, BML en Biotechnologie.



**Figuur 12:** Gewenste vaardigheden in Python bij net-afgestudeerde hbo'ers (in percentage van het aantal respondenten) voor de opleidingen Applied science, BML en Biotechnologie  
*Noot: In de vragenlijst voor BML/Biotechnologie was de optie Scripts schrijven niet meegenomen.*

## 2.8. Welke Linux/Bash-vaardigheden moeten net-afgestudeerde hbo'ers beheersen?

Ook voor het programma Linux/Bash geldt dat de chemie en chemische industrie hier weinig tot geen gebruik van maken. In figuur 13 staat in percentage van het aantal respondenten het resultaat van de opleidingen Applied science, BML en Biotechnologie.



**Figuur 13:** Gewenste vaardigheden Linux/Bash bij net-afgestudeerde hbo'ers (in percentage van het aantal respondenten) voor de opleidingen Applied science, BML en Biotechnologie

## 2.9. Aanvullende opmerkingen

Aan het einde van de vragenlijst hadden de respondenten de mogelijkheid om aanvullende opmerkingen te plaatsen. Hieronder per opleiding de opmerkingen op een rij.

### Applied science

- De vaardigheden binnen onze afdeling lopen erg uiteen. Het hangt sterk van de functie af. We hebben analisten die nooit coderen en onderzoekers die fulltime complexe bio-informatica toepassen. Ik heb deze antwoorden gebaseerd op een wet-lab analist die zijn/haar resultaten kan communiceren met bio-informatici, en die het werk van bio-informatici op een minimaal niveau kan begrijpen.
- Ik verwacht van een student dat hij/zij logisch na kan denken, data op zijn/haar manier duidelijk weer kan geven, daar een goede analyse op kan doen in bijvoorbeeld Excel.
- Statistische perfectie is voor ons niet relevant (of een mooie weergave hiervan), als maar duidelijk is/wordt wat de grote lijn in het verhaal is.
- Veel van de genoemde zaken bevinden zich op het snijvlak analysetechnieken lab/ bio-informatica/ biostatistiek. Zaken als machine learning, Jupyter, Shiny, de wat complexere statistiek, etc. horen naar ons idee thuis in specialistische groepen (Bioinformatics, Biostatistics). Dus is het goed om de studenten een goede basiskennis van een aantal zaken mee te geven maar uiteindelijk kunnen ze bij de specialist terecht die de analyse uitvoert of die ontwikkelt een tool waarmee ze kunnen werken (wel handig dan als je snapt wat er achter zit).
- Analytische instrumenten worden vervangen door nieuwere types en ook andere fabricaten. Voor quality control is method transfer nodig om te bewijzen dat een ander instrument dezelfde resultaten geeft.
- Het zou mooi zijn als net-afgestudeerden zowel Python als R beheersen, maar één van de twee talen is initieel ook afdoende, het zal dan vooral ook afhangen van de domein-interesse waar men terecht kan.
- Zowel techniek als programmeertalen zijn erg vluchtig. Ik vind daarom een focus op 3 principes van groot belang voor een goede data analist:
  1. Reproduceerbaarheid, dit is het enige wat je als data analist kunt garanderen. Er is namelijk altijd discussie mogelijk over de gebruikte modellen/methodes.
  2. Rapporteer volledig en helder, anders heeft je werk geen waarde.
  3. Wees lui en doe geen handmatig repetitief werk. Schrijf een script waardoor het automatisch en reproduceerbaar gaat.
- De vereiste basiskennis is sterk afhankelijk van de functie. Een moleculair biologisch analist die vooral wetlab werk doet, heeft meestal voldoende aan Excel / Graphpad. Een analist die ook bio-informatica als taak/interessesgebied heeft, moet minimaal de basis van R beheersen.
- Analyse van data en de daaruit voortvloeiende grafieken gaan uiteindelijk hand in hand met de verslaglegging. Ik zou dus willen toevoegen dat o.a. het schrijven van een duidelijke caption bij een figuur ook een hele belangrijke datavaardigheid is. Wat zie ik in deze figuur? Welke methode is gebruikt? Etc.
- Hetzelfde geldt ook voor het ontwerp van de figuur zelf: duidelijke kleuren, symbolen, een duidelijke legenda. Goed nadenken over de ranges op de assen. Alles gericht op 'wat wil ik overbrengen op de lezer'.
- Genoemde vaardigheden zijn niet voor al onze projecten nodig.
- Basis voor data-analyse is bij ons: MiniTab. Voor geavanceerdere data-analyse gaan we richting MatLab.
- In aanvulling op Excel is eigenlijk alle ervaring met andere programma's welkom. Python, R en Bash zijn de meest gebruikte talen bij ons op dit moment. Wat dan minimaal aan kennis aanwezig moet zijn en wat ze er "on the job" bij kunnen leren is een beetje lastig. Ervaring met alle programma's is onrealistisch en helemaal geen ervaring niet fijn, maar als ze dus een aantal punten waar ik nu een vinkje gezet heb missen en daarvoor in de plaats een aantal andere die ik

niet aangevinkt had wel hebben, dan denk ik dat ze daarmee ook wel de juiste vaardigheden geleerd hebben om dat wat ze nog niet beheersen snel op te pikken in hun baan.

- Voor ons is het basisbegrip van programmeren het belangrijkste om terug te zien bij startenden. Hierin hoeven ze geen expert te zijn.
- Wij gebruiken Graphpad en SPSS voor veel data-analyses.
- Er is een enorme ontwikkeling in data science land gaande. Het is onmogelijk dat studenten het hele arsenaal leren. Het is belangrijker de beginselen van een aantal programmeertalen (Excel, R, Python) en analysetechnieken te kennen, zodat ze een basis hebben waarop verder gebouwd kan worden. Elke organisatie heeft immers weer net andere eisen.

## BML

- Diverse afdelingen werken wel met Python, R en Linux alleen wij niet.
- Gebruik van microscopie software (Zen / Fiji / ImageJ), of in ieder geval principes van verwerken van microscopie data: hoe is een image opgebouwd (bitdepth, pixels, greyvalues, channels, scale etc), hoe analyseer je signaal of structuren in images (intensities, tresholding, areas etc).
- Wij hebben meerdere projecten lopen, en zijn altijd op zoek naar studenten die kunnen programmeren/coderen in Python en Matlab. Echter, heeft de student best wel een grote basiskennis nodig van deze programma's om zo'n project met hen te kunnen aangaan.
- There is also a need to emphasize presentation and science communication. How to develop a story, what to present and how to give perspective to the audience and more importantly what their internship means for the field; For example, our current intern is working on tuberculosis drug discovery and there was a steep learning curve to understand the background of not only TB but also of drug discovery. Learning relevant background information will be key to develop a well thought out hypothesis driven research and will also help communicate this work to a larger audience.
- Voor statistiek geldt dat het mooi zou zijn als de student weet dat deze testen bestaan en ze kan gebruiken om kwaliteitscontrole te doen (bijv. uitbijters identificeren, hoe groot is de intra vs inter sample variatie, zijn de samples wel normaal verdeeld) en tot een valide conclusie kan komen.
- Afgezien van statistiek is ook normaal labrekenen erg belangrijk, en dit vinden veel studenten heel moeilijk en gaat zeer moeizaam (ook al krijgen ze hier al les in).
- Wij gebruiken SPSS.
- Vragenlijst is ingevuld vanuit (glyco)proteomics onderzoek maar zal sterk overeenkomen voor andere functional omics groepen bij ons (glycomics/metabolomics). Hoewel Python skills in algemene zin erg handig zijn voor omics werkzaamheden is zeker R hogere prioriteit voor een laborant. Tevens zou het goed zijn om ook b.v. bitbucket / versioning / containering basis mee te geven. Qua prioritering zou ik willen aangeven: (i) Algemene kennis van univariate en multivariate analysetechnieken, statistiek, exploratieve (multidimensionale) chemometrische technieken, (un)supervised learning, basis bio-informatica (ii) basis Excel vaardigheden (iii) basis R-vaardigheden. Lagere prioriteit (iv) geavanceerde R kennis en vaardigheden en (v) basis Python-vaardigheden.
- Wanneer laboranten een inzicht en gevoel hebben voor data (pre)processing, analyse en visualisatie gecombineerd met basiskennis/vaardigheden in R dan is dat een prima basis om mee te beginnen. Vanuit deze basis is het mogelijk voor een laborant om verdere vaardigheden die verschillen tussen laboratoria en onderzoeksvelden binnen zijn of haar werk omgeving te leren.
- Bij sommige vragen was het handig geweest om aan te geven hoe belangrijk sommige taken zijn. bijv bij beheersen analyses is data kwaliteit, stewardship en gebruik van databases essentieel terwijl dingen als metagenomics meer nice-to-haves zijn.
- De elektronische systemen en robots die gebruikt worden op ons lab zijn "custom made" en vandaar dat de opleiding niet kan inspelen hierop maar het leren werken hiermee wordt op de werkvloer gedaan. Echter is het van groot belang dat studenten leren omgaan met vele verschillende software programma's voor dataverwerking en kennismaken met programmeren.

- Belangrijk is dat studenten het belang weten van gestructureerde uniforme data, dit zou onder de noemer "Data Stewardship" kunnen vallen.
- Hoewel deze datavaardigheden voor de toekomst zeer gewenst zullen zijn, gaat het nog niet zo'n vaart als ik had gedacht. Veel afdelingen stoeien nog met hoe ze de afweging moeten maken tussen bestaand personeel heropleiden of nieuw personeel aan te nemen, dan wel samen te werken met bioinformatics afdelingen/teams/bedrijven. Ik zou dus niet het halve curriculum omgooien voor volgende jaren, maar geleidelijk aanpassen. Belangrijkste vaardigheden bij de gevraagden: Python/R/Bash voor automatisering en data QC/cleaning. Statistiek is handig en belangrijk, maar wordt vaak beter beheert door meer seniore collega's.
- Geen Excel gebruiken voor statistiek, doe dan GraphPad/Prism, SPSS of R, waarbij één van de eerste twee de voorkeur heeft voor ons type organisatie.
- Binnen onze afdeling zou het nuttig zijn om enige kennis van programmeren te hebben. We werken zelf veel met MATLAB, dat hierboven niet genoemd wordt. Maar het belangrijkste is dat ze ervaring opdoen met een (willekeurige) vorm van programmeren - en dan maakt het wat mij betreft niet uit in welke omgeving dat dan is. Het zal uiteraard niet mogelijk zijn om op alle mogelijke werkwijzen van alle labs volledige aansluiting te verzorgen.
- Een keer een Linux server geïnstalleerd hebben kan zeker geen kwaad. Nginx staat er zo op en dat is nog leuk ook om zelf te kunnen.

### Biotechnologie

Geen aanvullende opmerkingen.

### Chemie / Chemische technologie

- Bij ons zijn de vaardigheden met MS data-analyse erg belangrijk, zowel triplequad MS/MS als HRMS data-analyse, targeted en untargeted, inclusief identificatie van stoffen door spectra met behulp van online databases.
- Het hangt er qua vereiste kennis wel wat vanaf welke functie de persoon zou gaan vervullen. Tevens doen we nog niet heel veel met statistiek, maar gaan we dat wat mij betreft in de toekomst wel meer doen.
- Verder zijn meer algemene vaardigheden zoals abstract kunnen denken en methodisch, gestructureerd werken zeker ook heel belangrijk in ons werk.
- Het kunnen uitvoeren van analyses en data kunnen verwerken is één ding. Het uitleggen en presenteren (in de meest brede zin van het woord) van de resultaten is net zo belangrijk.
- Onze ervaring is dat bij HBO-stagiairs en pas afgestudeerde Hbo'ers het niveau van "schrijven" tegenvalt. Denk daarbij zowel aan het goed presenteren van de resultaten (en dan niet de presentatievaardigheden van de spreker, maar het inhoudelijk opstellen van de presentatie) en het schrijven van een rapport of notitie.
- Naast deze specifieke vaardigheden zou er in het algemeen aandacht moeten zijn voor de wiskundige basiskennis. Verder is programmeren vaak relevant maar ook dat zou niet specifiek op een taal moeten zijn gericht maar meer op het begrijpen van algoritmieken.
- We maken, naast Excel, gebruik van Minitab 18 en gPROMS (modellering en Monte Carlo analyses), voor statistische analyses.
- Vaak is het beheersen niet een vereiste, maar maakt de werkzaamheden wel een stuk makkelijker als de kennis aanwezig is.
- Afgestudeerden moeten handig zijn met een of meer software pakketten die chemische structuren representeren.

## Bijlage 1: Vragenlijst 2019

### Vragenlijst datavaardigheid bedrijven NL

Vraag 1\*

Welk onderwijsprogramma sluit het beste aan bij de werkzaamheden die worden uitgevoerd op uw afdeling?

- Applied science
- Bio-informatica
- Biologie en medisch laboratoriumonderzoek
- Biotechnologie
- Chemie
- Chemische technologie

Vraag 2\*

Wat voor dataverwerking/analyses worden er op uw afdeling uitgevoerd?

Vraag 3\*

Is er volgens u een verschuiving gaande van het werk van een analist van experimenteel werk naar data-analyses?

Vraag 4\*

Welk deel van de dataverwerking wordt uitgevoerd door analisten en vanaf welk punt wordt de analyse overgenomen door data-experts? Bijvoorbeeld: analisten generen de data en voeren de kwaliteitscontrole uit, de verdere analyses worden uitgevoerd door andere experts (bijvoorbeeld door een (bio-)informaticus of statisticus).

Vraag 5a\*

Hoe is de huidige verhouding in tijdsinspanning van een analist tussen experimenteel werk en data-analyse?

Vraag 5b\*

Hoe is de verhouding naar uw verwachting in tijdsinspanning van een analist tussen experimenteel werk en data-analyse over 5 jaar?

Vraag 6a\*

Welke programma's en/of programmeertalen zijn het belangrijkste voor het huidige werk van een analist op uw afdeling?

- R
- Python
- C++
- Excel
- Galaxy
- SPSS
- Linux
- Geneious
- Azure
- CLCBio



- Minitab
- SAS
- Bash
- SQL
- IGV
- Oracle
- MatLab
- Origin
- Anders...

Vraag 6b\*

Welke programma's en/of programmeertalen zijn het belangrijkste voor het werk van een analist op uw afdeling over 5 jaar denkt u?

- R
- Python
- C++
- Excel
- Galaxy
- SPSS
- Linux
- Geneious
- Azure
- CLCBio
- Minitab
- SAS
- Bash
- SQL
- IGV
- Oracle
- MatLab
- Origin
- Anders...

Vraag 7\*

Zijn er vaardigheden op het gebied van gegevensverwerking en -analyse die (recent) afgestudeerden op dit moment missen? Zo ja, wat zijn volgens u de belangrijkste?

Vraag 8\*

Mogen we u benaderen voor een toelichting op uw antwoorden?

---

## Vragenlijst datavaardigheid alumni

Vraag 1\*

Welk onderwijsprogramma heb je gevolgd tijdens je opleiding?

- Applied science
- Bio-informatica
- Biologie en medisch laboratoriumonderzoek
- Biotechnologie
- Chemie
- Chemische technologie

Vraag 2\*

Wat is je huidige functie?

Vraag 3\*

Gebruik je data-analyse voor het uitvoeren van je functie?

Vraag 4\*

Hoe is de verhouding in jouw tijdsinspanning tussen experimenteel werk en data-analyse?

Vraag 5\*

Welk deel van de dataverwerking wordt uitgevoerd door jou en vanaf welk punt wordt de analyse overgenomen door data-experts (bijvoorbeeld door een (bio-)informaticus of statisticus)?

Vraag 6\*

Welke programma's en/of programmeertalen gebruik jij?

- R
- Python
- C++
- Excel
- Galaxy
- SPSS
- Linux
- Geneious
- Azure
- CLCBio
- Minitab
- SAS
- Bash
- SQL
- IGV
- Oracle
- MatLab

- Origin
- Anders...

Vraag 7

Heb je aanvullende cursussen/opleidingen gevolgd op het gebied van dataverwerking? Zo ja, welke?

Vraag 8\*

Welke software voor data-analyse zouden studenten volgens jou minimaal moeten leren gebruiken, gezien vanuit je eigen ervaring?

- R
- Python
- C++
- Excel
- Galaxy
- SPSS
- Linux
- Geneious
- Azure
- CLCBio
- Minitab
- SAS
- Bash
- SQL
- IGV
- Oracle
- MatLab
- Origin
- Anders...

Vraag 9\*

Welke tips heb je nog qua aanvullingen en/of aanpassingen in het huidige curriculum van de door jouw gevolgde opleiding op het gebied van data-analyse?

Vraag 10\*

Mogen we je benaderen voor toelichting op je antwoorden?

## Bijlage 2: Vragenlijst 2020

### Applied science

1. Bij welk soort organisatie bent u werkzaam?\*

- Ziekenhuis
- Universiteit
- KNAW of onderzoekslab
- Profit organisatie
- 

2. Hoe groot is uw organisatie?\*

- < 50 medewerkers
- > 50 medewerkers

3. Op welke afdeling bent u werkzaam?\*

4. Beheersing van analyses\*

Welke typen analyses moeten net-afgestudeerde hbo'ers AS minimaal beheersen voor een goede aansluiting bij de werkzaamheden binnen uw organisatie?

- Databases gebruiken (UniProt, NCBI, BLASTen, etc.)
- Data Stewardship (het beheer van datasets, bijv. volgens FAIR-principes)
- Optimaliseren methoden (bijv. met DoE, simplex)
- Valideren methoden
- Kalibratie apparatuur (m.n. balans, volumetrisch glaswerk)
- Transcriptomics
- Proteomics
- Metabolomics
- Wet-lab data (ELISA, multiplex, cell culture, flowcytometry etc.)
- Anders, namelijk

5. Niveau voor statistiek\*

Welk niveau op het gebied van statistiek moeten net-afgestudeerde hbo'ers AS minimaal beheersen voor een goede aansluiting bij de werkzaamheden binnen uw organisatie?

- Kentallen berekenen (gemiddelden, standaarddeviatie, SEM)
- Uitbijters (ook wel uitschieters, outliers)
- Betrouwbaarheidsintervallen bepalen
- Basistoetsen (t-toets, ANOVA)
- ANOVA
- Meerwegs-ANOVA
- Chi-kwadraattest
- Nonparametrische statistiek (bijv voor toetsen en betrouwbaarheidsinterval)
- Toetsen op normaliteit (bijv. Kolmogorov-Smirnov)
- Toetsen op uitbijters (ook wel uitschieters, outliers)

- Correlatie
- Enkelvoudige lineaire regressie
- Enkelvoudige niet-lineaire regressie
- Meervoudige regressie
- Clusteranalyse
- Principale componentenanalyse (PCA)
- Design of Experiments (DoE)
- Bayesiaanse statistiek
- Machine learning
- Data cleaning
- Anders, namelijk

#### 6. Excel vaardigheden\*

Welke Excel vaardigheden moeten net-afgestudeerde hbo-studenten AS minimaal beheersen om een goede aansluiting te hebben bij de werkzaamheden binnen uw organisatie?

- Datavisualisatie (staafdiagram, boxplot, etc.)
- Kentallen (gemiddelde, mediaan, etc.)
- Statistische testen (T-test, ANOVA, etc.)
- IJklijnen
- Pivot tabel (draaitabellen)
- Zoekfuncties (vlookup, index, match, etc.)
- Macro's
- Visual Basic (VBA)
- Geen, wij maken geen gebruik van Excel
- Anders, namelijk

#### 7. R-vaardigheden\*

Welke R-vaardigheden moeten net-afgestudeerde hbo-studenten AS minimaal beheersen om een goede aansluiting te hebben bij de werkzaamheden binnen uw organisatie?

- Eenvoudige dataverwerking (data inladen, bewerken)
- Datavisualisatie (staafdiagram, boxplot, etc.)
- Statistische testen (T-test, ANOVA, etc.)
- Complexe statistische testen (clusteranalyse, PCA, etc.)
- R Commander
- R Markdown
- Shiny
- Machine learning
- Data cleaning
- Scripts schrijven
- Geen, wij maken geen gebruik van R
- Anders, namelijk

#### 8. Python-vaardigheden\*

Welke Python-vaardigheden moeten net-afgestudeerde hbo-studenten AS minimaal beheersen om een goede aansluiting te hebben bij de werkzaamheden binnen uw organisatie?

- Eenvoudige dataverwerking (data inladen, bewerken)
- Datavisualisatie (staafdiagram, boxplot, etc.)
- Statistische testen (T-test, ANOVA, etc.)
- Complexe statistische testen (clusteranalyse, PCA, etc.)
- Integrated Design Environment (IDE, bijv. PyCharm)
- Jupyter
- Scripts schrijven
- Machine learning
- Data cleaning
- Geen, wij maken geen gebruik van Python
- Anders, nl..

9. Linux / Bash vaardigheden

Welke Linux/Bash-vaardigheden moeten net-afgestudeerde hbo-studenten AS minimaal beheersen om een goede aansluiting te hebben bij de werkzaamheden binnen uw organisatie?

- Basiscommando's (pwd, cd, ls, etc.)
- Scripts schrijven
- Cluster Computing
- DevOps (Docker etc.)
- Geen, wij maken geen gebruik van Linux / Bash
- Anders, nl..

10. Heeft u nog aanvullende opmerkingen?

## BML/Biotechnologie

1. Bij welk soort organisatie bent u werkzaam?\*

- Ziekenhuis
- Universiteit
- KNAW of onderzoekslab
- Profit organisatie
- Anders

2. Hoe groot is uw organisatie?\*

- < 50 medewerkers
- > 50 medewerkers

3. Op welke afdeling bent u werkzaam?\*

4. Beheersing van analyses\*

Welke typen analyses moeten net-afgestudeerde hbo'ers BML minimaal beheersen voor een goede aansluiting bij de werkzaamheden binnen uw organisatie?

- Kwaliteitscontrole Sequencing Data
- Transcriptomics
- Metagenomics
- GWAS
- QTL-analyse
- Proteomics
- Metabolomics
- Variant Calling
- Assembleren & Annoteren
- Databases gebruiken (UniProt, NCBI, BLASTen, etc.)
- Data Stewardship (het beheer van datasets, bijv. volgens FAIR-principes)
- Wet-lab data (ELISA, multiplex, cell culture, flowcytometry etc.)
- Anders, namelijk

#### 5. Niveau voor statistiek\*

Welk niveau op het gebied van statistiek moeten net-afgestudeerde hbo'ers BML minimaal beheersen voor een goede aansluiting bij de werkzaamheden binnen uw organisatie?

- T-test
- ANOVA
- Meerwegs-ANOVA
- Enkelvoudige regressie (lineaire regressie)
- Meervoudige regressie
- Niet-lineaire regressie
- Correlatie
- Chi-kwadraattest
- Clusteranalyse
- Principale componentenanalyse (PCA)
- Nonparametrische statistiek (bijv voor toetsen en betrouwbaarheidsinterval)
- Design of Experiments (DoE)
- Uitbijters (ook wel uitschieters, outliers)
- Kentallen berekenen (gemiddelden, standaarddeviatie, SEM)
- Betrouwbaarheidsintervallen bepalen
- Bayesiaanse statistiek
- Machine learning
- Data cleaning
- Anders, namelijk

#### 6. Excel vaardigheden\*

Welke Excel vaardigheden moeten net-afgestudeerde hbo-studenten BML minimaal beheersen om een goede aansluiting te hebben bij de werkzaamheden binnen uw organisatie?

- Datavisualisatie (staafdiagram, boxplot, etc.)
- Kentallen (gemiddelde, mediaan, etc.)
- Statistische testen (T-test, ANOVA, etc.)

- IJklijnen
- Pivot table (draaitabellen)
- Zoekfuncties (vlookup, index, match, etc.)
- Macro's
- Visual Basic (VBA)
- Geen, wij maken geen gebruik van Excel
- Anders, namelijk

#### 7. R-vaardigheden\*

Welke R-vaardigheden moeten net-afgestudeerde hbo-studenten BML minimaal beheersen om een goede aansluiting te hebben bij de werkzaamheden binnen uw organisatie?

- Eenvoudige dataverwerking (data inladen, bewerken)
- Datavisualisatie (staafdiagram, boxplot, etc.)
- Statistische testen (T-test, ANOVA, etc.)
- Complexe statistische testen (clusteranalyse, PCA, etc.)
- R Commander
- R Markdown
- R Packages
- Shiny
- Machine learning
- Data cleaning
- Geen, wij maken geen gebruik van R
- Anders, namelijk

#### 8. Python-vaardigheden\*

Welke Python-vaardigheden moeten net-afgestudeerde hbo-studenten BML minimaal beheersen om een goede aansluiting te hebben bij de werkzaamheden binnen uw organisatie ?

- Eenvoudige dataverwerking (data inladen, bewerken)
- Datavisualisatie (staafdiagram, boxplot, etc.)
- Statistische testen (T-test, ANOVA, etc.)
- Complexe statistische testen (clusteranalyse, PCA, etc.)
- Integrated Design Environment (IDE, bijv. PyCharm)
- Jupyter
- Machine learning
- Data cleaning
- Geen, wij maken geen gebruik van Python
- Anders, nl..

#### 9. Linux / Bash vaardigheden

Welke Linux/Bash-vaardigheden moeten net-afgestudeerde hbo-studenten BML minimaal beheersen om een goede aansluiting te hebben bij de werkzaamheden binnen uw organisatie ?

- Basiscommando's (pwd, cd, ls, etc.)
- Scripts schrijven



- Cluster Computing
- DevOps (Docker etc.)
- Geen, wij maken geen gebruik van Linux / Bash
- Anders, nl..

10. Heeft u nog aanvullende opmerkingen?

## Chemie / Chemische technologie

1. Bij welk soort organisatie bent u werkzaam?\*

- Ziekenhuis
- Universiteit
- KNAW of Onderzoekslab
- Profit organisatie
- Anders

2. Hoe groot is uw organisatie?\*

- < 50 medewerkers
- > 50 medewerkers

3. Op welke afdeling bent u werkzaam?\*

4. Beheersing van analyses\*

Welke typen analyses moeten net-afgestudeerde hbo'ers Chemie minimaal beheersen voor een goede aansluiting bij de werkzaamheden binnen uw organisatie?

- Optimaliseren methoden (bijv. met DoE, simplex)
- Valideren methoden
- Methodenvergelijking
- Kalibratie apparatuur (m.n. balans, volumetrisch glaswerk)
- Interne kwaliteitscontrole
- Ringonderzoek
- Metabolomics
- Databases gebruiken
- Data Stewardship (het beheer van datasets)
- Anders, namelijk

5. Niveau statistiek\*

Welk niveau op het gebied van statistiek moeten net-afgestudeerde hbo'ers Chemie minimaal beheersen voor een goede aansluiting bij de werkzaamheden binnen uw organisatie?

- Kentallen berekenen (o.a. gemiddelden, standaarddeviatie)
- Toetsen op uitbijters (ook wel uitschieters, outliers)
- Toetsen op normaliteit (bijv. Chi2-toets of Kolmogorov-Smirnov)
- Betrouwbaarheidsintervallen bepalen (voor meetwaarden en gemiddelden)

- Betrouwbaarheidsintervallen bepalen (voor standaarddeviatie)
- Basistoetsen (Z-,t-,F-toets,Chi2-toets)
- Chi-kwadraattest (voor kwaliteit van een fit)
- ANOVA
- Meerwegs-ANOVA
- Enkelvoudige regressie (lineaire regressie)
- Meervoudige regressie
- Niet lineaire regressie
- Clusteranalyse
- Principale componentenanalyse (PCA)
- Partial least squares regression (PLS)
- Nonparametrische statistiek (bijv voor toetsen en betrouwbaarheidsinterval)
- Anders, namelijk

#### 6. Excel vaardigheden\*

Welke Excel-vaardigheden moeten net-afgestudeerde hbo-studenten Chemie minimaal beheersen om een goede aansluiting te hebben bij de werkzaamheden binnen uw organisatie?

- Datavisualisatie (staafdiagram, boxplot, etc.)
- Statistische testen (T-test, ANOVA, etc.)
- Kalibratielijnen
- Pivot table (draaitabellen)
- Zoekfuncties (vlookup, index, match, etc.)
- Macro's
- Visual Basic
- Geen, wij maken geen gebruik van Excel
- Anders, namelijk

#### 7. R-vaardigheden\*

Welke R-vaardigheden moeten net-afgestudeerde hbo-studenten chemie minimaal beheersen om een goede aansluiting te hebben bij de werkzaamheden binnen uw organisatie ?

- Eenvoudige dataverwerking (data inladen, bewerken)
- Datavisualisatie (staafdiagram, boxplot, etc.)
- Statistische testen (T-test, ANOVA, etc.)
- Complexe statistische testen (clusteranalyse, PCA, etc.)
- R Commander
- R Markdown
- Scripts schrijven
- Geen, wij maken geen gebruik van R
- Anders, namelijk

#### 8. Heeft u nog aanvullende opmerkingen?