



Module Handleiding

# Master Applied Statistics

## Module 6: Introduction to Data Processing & Programming in R

Code: AS\_06

HANDLEIDING

## Inhoudsopgave

1. Cursusgegevens .....	3
2. Inleiding.....	4
3. Eindkwalificaties & Leerdoelen .....	4
4. Omschrijving van de vakinhoud.....	5
5. Programmaoverzicht.....	6
6.Onderwijsvormgeving .....	8
7.Verwachtingen van studenten en voorkennis .....	8
8. Module belasting .....	9
9.Wijze van toetsen & vaststellen eindcijfer .....	9
10.Collegemateriaal .....	10

## 1. CURSUSGEGEVENS

<b>Naam onderdeel</b>	Introduction to Data Processing & Programming in R
<b>Vakcode</b>	AS_06
<b>Opleiding</b>	Master Applied Statistics
<b>Leerlijn/Vakgroep/Discipline</b>	Data-Analyse
<b>Aantal studiepunten</b>	6 ECTS
<b>Semester en studiefase</b>	Jaar 1 semester 2
<b>Contacturen per semester</b>	40
<b>College periode</b> <b>Werkcolleges+studievrije periode</b> <b>Tentamen</b> <b>Hertentamenperiode</b>	2 juni – 13 juni 2025 16 juni – 10 juli 2025 11 juli 2025 25 augustus – 5 september 2025
<b>Vereiste voorkennis</b>	Afgeronde Wetenschappelijke Bachelor/ Premaster Wetenschappelijke Vorming IGSR
<b>Docent(en) + contactgegevens</b>	Gerko Vink, <a href="mailto:g.vink@uu.nl">g.vink@uu.nl</a>
<b>Spreekuur docent(en)</b>	Op afspraak

## 2. INLEIDING

Dit is de handleiding voor de module Introduction to Data Processing & Programming in R van de Masteropleiding Applied Statistics. Het doel van deze module is om inzicht te bieden in zowel de basisprincipes als de basis dataverwerking- en analysetechnieken in R.

Deze module richt zich op het ontwikkelen van praktische vaardigheden in R, een statistisch softwarepakket, dat breed wordt ingezet binnen de academische wereld en de beroepspraktijk bij het uitvoeren van statistische analyses. De module biedt daarnaast verdiepende inzichten in de belangrijkste functies bij het programmeren in R, variërend van het importeren en beheren van datasets tot fundamentele analyses en visualisaties en presentatie van documenten.

De leerdoelen van de module worden ook in deze handleiding opgenomen evenals een uitgebreide omschrijving van de vakinhoud, het programmaoverzicht en de lesonderdelen. Tenslotte volgt er een uitwerking van de wijze van toetsen en de vaststelling van het eindcijfer.

## 3. EINDKWALIFICATIES & LEERDOELEN

Dit vak draagt bij aan de volgende eindkwalificaties van de Masteropleiding:

### I. Kennis en inzicht

EK3. Verdiepende kennis over statistische softwarepakketten en de standaard terminologie horende bij calculatie-en programmeer vaardigheden in R

### II. Toepassen kennis en inzicht

De afgestudeerde is in staat om:

EK5. Secundaire of primaire data van geavanceerde statistische technieken in R te combineren en bewerken.

### III. Communicatie

De afgestudeerde is in staat om:

EK9. Als statisticus binnen multidisciplinaire teams te werken en te communiceren.

## Leerdoelen

Leerdoelen voor deze cursus zijn aangegeven met tussen haakjes aan welke eindkwalificatie die gerelateerd zijn.

Aan het eind van deze module kan de student:

- de software R gebruiken voor reproduceerbare statistische analyse van gegevens en programmeren (EK3)
- met behulp van syntax en code grafieken te (re)produceren, verwerken en aanpassen (EK5)
- in teamverband werken om data uit verschillende bronnen te verkennen, samen te voegen en aan te passen, en inferentiële, predictieve en beschrijvende data-analyse te presenteren en uit te voeren (EK9)

## 4. OMSCHRIJVING VAN DE VAKINHOUD

De hoofdonderwerpen die hierbij aan de orde komen zijn:

- databestanden beheren, data uit verschillende softwarebronnen exporteren en inlezen (importeren)
- databaseer en -hercodering, data mutaties en data throughput
- creëren van grafieken en visualisaties (visual storytelling)
- syntax en code lezen en schrijven (programmeren)
- tools om data te verkennen
- bestanden en gegevens uit verschillende bronnen samenvoegen (mergen)
- reproduceerbare codes schrijven om inferentiële, predictieve en beschrijvende statistieken en visualisaties te produceren voor twee of meer variabelen.

## 5. PROGRAMMAOVERZICHT

In de tabel hieronder staat een voorlopige planning van de lesonderdelen. Bij aanvang van de module zal een overzicht met onderwijsactiviteiten per college, de te bestuderen/voor te bereiden stof, lesdoelen, opdrachten, werkvormen en toets momenten beschikbaar gesteld worden.

**Tabel 1. Programmaoverzicht en lesonderdelen**

Collegedag (=4 uren)	Datum	Lesonderdeel
Gecombineerde HC & WC	Periode: 2 juni – 13 juni 2025	
1	Ma 2 juni 2025	<b>Introductie R &amp; RStudio</b> <ul style="list-style-type: none"> <li>- Overzicht RStudio-interface</li> <li>- Werken met R-scripts, Quarto documenten en de R-console</li> <li>- Basis R-commando's voor beschrijvende statistische analyse</li> <li>- Werken met projecten en scripts</li> <li>- Begrijpen en toepassen van verschillende eenvoudige data containers</li> </ul>
2	Di 3 juni 2025	<b>Data management</b> <ul style="list-style-type: none"> <li>- Importeren en bestuderen van datasets</li> <li>- Begrijpen en toepassen van verschillende datatypes en database formats</li> <li>- Variabelen labelen en (her)coderen</li> <li>- De blauwdruk van R: frames en environments</li> <li>- Pipes</li> <li>- Formules gebruiken in functies</li> </ul>
3	Wo 4 juni 2025	<b>Data Manipulatie</b> <ul style="list-style-type: none"> <li>- Het combineren van datasets</li> <li>- Groeperen en aggregeren</li> <li>- Nieuwe variabelen creëren</li> <li>- Filteren en sorteren van gegevens</li> <li>- Het maken en aanpassen van datagroepen</li> <li>- Clustering van gegevens</li> </ul>
4	Do 5 juni 2025	<b>Statistische analyse</b> <ul style="list-style-type: none"> <li>- Beschrijvende statistiek</li> <li>- Kruistabellen en frequentieverdelingen</li> <li>- <math>X^2</math>-toets en associatiematen</li> <li>- Simpele lineaire regressie</li> <li>- Analyses draaien op groepen</li> </ul>
5	Vr 6 juni 2025	<b>Geen hoorcollege</b>
6	Ma 9 juni 2025	<b>Functioneel programmeren</b> <ul style="list-style-type: none"> <li>- Zelf functies ontwikkelen, gebruiken en debuggen</li> <li>- Map / Reduce workflows</li> <li>- Binaire operators</li> <li>- Trekken uit verdelingen</li> <li>- Random number generation</li> </ul>

7	Di 10 juni 2025	<b>Geavanceerde data technieken</b> <ul style="list-style-type: none"> <li>- Data-cleaning technieken: missing values (ontbrekende data) identificeren en corrigeren</li> <li>- Synthetische data maken</li> </ul>
8	Wo 11 juni 2025	<b><i>Geen hoorcollege</i></b>
9	Do 12 juni 2025	<b>Visualisatie van Data</b> <ul style="list-style-type: none"> <li>- Basisplots: histogrammen, scatterplots en boxplots</li> <li>- Geavanceerde plots met ggplot2</li> <li>- Aanpassen van grafieken voor publicatie</li> <li>- Exporteren van grafieken en resultaten</li> </ul>
10	Vr 13 juni 2025	<b>Ontwikkelen &amp; programmeren in R</b> <ul style="list-style-type: none"> <li>- Zelf R-packages maken, testen, documenteren en publiceren</li> <li>- Samenvatting + Q&amp;A</li> </ul>
<b>Werkcolleges</b>	Ma 16 – Vr 27 juni 2025	Maken van werkopdrachten m.b.v. R
<b>Onderwijsvrij</b>	30 juni – 10 juli 2025	
<b>Tentamen</b>	<b>Vrijdag 11 juli 2025</b>	

## 6. Onderwijsvormgeving

De *leeractiviteiten* in deze module zijn: Hoorcolleges (HC): Theorie en voorbeelden (studenten bestuderen de slides vóór de colleges); Werkcolleges (WC): Tijdens de werkcolleges krijgen studenten verdere instructies/begeleiding voor het maken van de opdrachten; Zelfstudie: Het uitwerken van de gegeven opdrachten en analyses als voorbereiding op de WC (individueel of als groep).

In deze module wordt gewerkt met ‘samenwerkend leren’ als didactische werkvorm. Hierbij worden de studenten in groepen ingedeeld waarbij ze, in dit geval, complementair zelfstandig aan de opdrachten werken. Tijdens de werkcolleges worden de opdrachten gepresenteerd en bediscussieerd in groepsverband volgens een beurtensysteem. Na het college wordt door de groep de uitgewerkte opdrachten en aantekeningen geüpload op Moodle in de hiervoor aangemaakte mappen. Deze werkvorm stelt de studenten in staat kennis met elkaar te delen en vaardigheden en inzichten op te doen, tijdens zowel het groepswerk als de presentaties.

## 7. VERWACHTINGEN VAN STUDENTEN EN VOORKENNIS

Van de studenten wordt verwacht dat ze zich voorbereiden op de werkcolleges. De studenten dienen ook een actieve en zelfsturende leer houding te demonstreren waarbij ze nieuwe kennis willen/kunnen uitzoeken. Verder wordt er verwacht dat ze naast de colleges wekelijks gemiddeld 20 uur aan het vak besteden (zie Module belasting).

De verwachtingen van de student voor het goed doorlopen van deze module zijn verder:

1. Basiskennis wiskunde en statistiek (BSc niveau)
2. Het tijdig downloaden van college slides, documenten (via Moodle) voor zelfstudie, collegevoorbereiding en het maken van de opdrachten
3. Controleren van de Module/cursus Moodle page voor aankondigingen over colleges en toetsing
4. Het hebben van een Laptop of PC/ microfoon / webcam (eventueel) / Goede Internetverbinding voor het volgen van sommige colleges



## 8. MODULE BELASTING

De module belasting telt in totaal 6 ECTs (SP) wat gemiddeld neerkomt op het reserveren van 168 studieuren. In de onderstaande tabel vind je de spreiding van de uren over de diverse activiteiten die betrokken zijn bij het behalen van dit vak met een opsplitsing van contacturen, zelfstudie uren en het tentamen.

**TABEL 2. STUDIEPUNTEN UITGEWERKT IN CONTACT EN ZELF-STUDIEUREN**

**Totaal = 6 ECTs: 6 x 28 = 168 uren**

Activiteit	Aantal weken/dagen	Totaal aantal uren
<b>Contacturen</b>		
Gecombineerde hoor- en werkcolleges	2 weken	8 dagen x 4 uren = 40 uren
Werkcolleges	2 dagen	2 dagen x 4 uren = 8 uren
Responsiecollege	1 dag	3,5 uren
<b>Zelfstudie uren</b>		
Studieplanning maken	2 weken	2 uren/week = 4 uren
Literatuurstudie	2 weken	8 uren/week = 16 uren
Uitwerken van vraagstukken/opdrachten	2 weken	20 uren/week = 40 uren
Vorbereiding middels ondersteunend materiaal (collegeslides, video's)	2 weken	15 uren/week = 30 uren
Herhaling	2 weken	7,5 uren/week = 15 uren
<b>Tentamen</b>	1 dag	3,5 uren

## 9. WIJZE VAN TOETSEN & VASTSTELLEN EINDCIJFER

Het tentamen van Module 5 wordt voor 100% getoetst op basis van opdrachten die fysiek en individueel worden afgenomen in 3-uren tijd. Het toetswerk is een **Statistische Analyse kennis- en vaardigheidstoets** waarbij de focus gelegd wordt op het uitvoeren van analyses op gecombineerde en bewerkte datafiles (met incomplete informatie). Deze datafiles worden gebruikt voor het beantwoorden van de probleemstellingen van gegeven casussen waarbij ook kennis m.b.t. interpretatie van resultaten getoetst zal worden. Beoordeling geschiedt d.m.v. een antwoordmodel.

### CESUUR

De student haalt deze module met een cijfer van 5.5 of meer. Indien een student de toets niet heeft gehaald, dan komt die in aanmerking voor een herkansing. Voor het deelnemen zijn er geen bijkomende voorwaarden dan die vermeld zijn in de OER waaronder het tijdig intekenen voor deelname.

## 10. COLLEGEMATERIAAL

- Collegeslides, gepubliceerd Moodle en op [www.gerkovink.com/sur](http://www.gerkovink.com/sur)
- Wickham, H., Çetinkaya-Rundel, M., & Grolemund, G. (2023, 2nd edition). *R for data science*. "O'Reilly Media, Inc.". Open Access Book freely available from <https://r4ds.hadley.nz/>.
- Wickham, H. (2019). *Advanced R*. Chapman and Hall/CRC. Open Access Book freely available from <https://adv-r.hadley.nz>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2023, 2nd edition, corrected print). *An introduction to statistical learning with applications in R*. New York: springer. Open Access Book freely available from <https://www.statlearning.com/>
- Van Buuren, S. (2018). *Flexible Imputation of Missing Data*. CRC Press. Book freely available from <https://stefvanbuuren.name/fimd/>
- Een selectie van relevante en vrij beschikbare open source materialen (BUKU platform)