



Certificate of Advanced Studies
Datenanalyse

Daten fallen heute in grossen Mengen an. Im CAS Datenanalyse (CAS DA) erlernen Sie Methoden und Werkzeuge, um diese gezielt für Ihre Zwecke zu nutzen – beispielsweise für die Optimierung und Planung von Prozessen und Dienstleistungen, für das Qualitätsmanagement, oder um neue Erkenntnisse für strategische Entscheide zu gewinnen.



Inhaltsverzeichnis

1	Umfeld	3
2	Zielpublikum	3
3	Ausbildungsziele	3
4	Voraussetzungen	3
5	Termine, Anmeldung und Durchführungsort	3
6	Kompetenzprofil	4
7	Kursübersicht	5
8	Kursbeschreibungen	6
	8.1 Tooling und Datenmanagement	6
	8.2 Deskriptive Statistik und mathematische Grundlagen	7
	8.3 Statistisches Testen	7
	8.4 Grafische Datenexploration und Datenvisualisierung	8
	8.5 Regressionsanalyse	8
	8.6 Zeitreihen und Prognosen	9
	8.7 Data Mining	10
	8.8 Wahltag Open Data	10
	8.9 Wahltag Kausalanalyse	11
	8.10 Wahltag Zeitreihen und Regression	11
	8.11 Wahltag Zeitreihen und Regression	11
9	Kompetenznachweis	12
10	Lehrmittel	12
11	Dozierende	13
12	Organisation	14

1 Umfeld

Daten sammeln allein bringt noch keine Erkenntnisse. Entscheidend ist das «Making Sense out of Data»: Wie können Daten beschrieben und analysiert werden, welche Aussagen kann man mit ihnen machen und welche Schlussfolgerungen kann man aus ihnen ziehen? Auf dem Markt stehen leicht bedienbare Software-Tools zur Aufbereitung von Daten, zur Analyse und zur Visualisierung zur Verfügung. Das CAS DA vermittelt einen praktischen Grundstock an Wissen, wissenschaftlichen Vorgehensweisen und Werkzeugen für die Datenanalyse.

2 Zielpublikum

Das CAS DA richtet sich an Verantwortliche für Datenanalyse-Projekte: an Informatikerinnen und Informatiker, die in Datenanalyse-Projekten arbeiten sowie an wissenschaftliche Mitarbeitende, welche beim Erstellen von Analysen und Studien beraten und selbst Daten auswerten.

3 Ausbildungsziele

- Sie überblicken das Gebiet der Datenanalyse.
- Sie können Daten methodisch aufbereiten, analysieren und visualisieren.
- Sie können die Regressionsanalyse und die Analyse von Zeitreihen als Prognose-Instrumente anwenden.
- Sie lernen das Arbeiten mit der Sprache R (r-project.org), einem der wichtigsten und verbreitetsten Werkzeuge der Data Analyse.
- Sie bekommen einen Einblick in modernes Data Mining und können ausgewählte Methoden anwenden.

4 Voraussetzungen

Sie können mit mathematischen Gesetzen umgehen und haben die Bereitschaft zur anwendungsorientierten Arbeit mit statistischer Software.

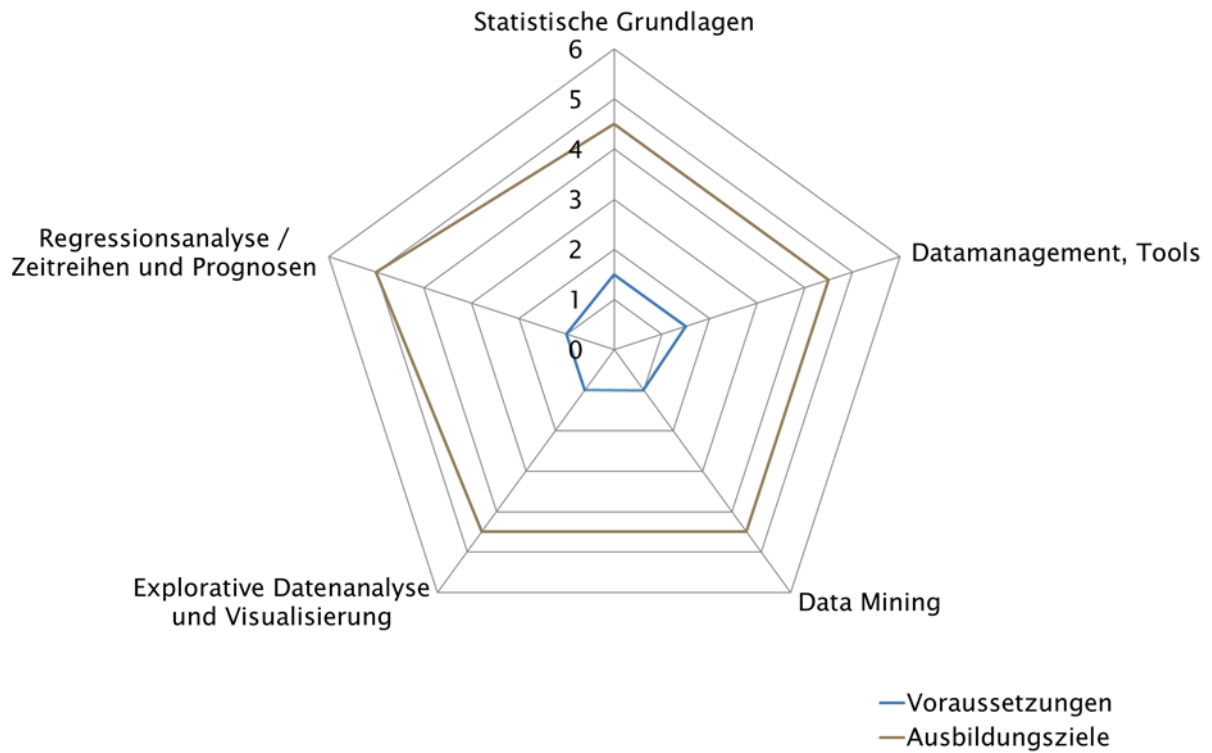
5 Termine, Anmeldung und Durchführungsort

Kursstart Kalenderwoche 43/2019, Anmeldeschluss Ende Kalenderwoche 40/2019.

Das CAS dauert ein Semester und findet an einem Tag pro Woche von 08:30 Uhr bis 16:15 Uhr statt.

Berner Fachhochschule, Weiterbildung, Wankdorffeldstrasse 102, 3014 Bern,
Telefon +41 31 848 31 11, E-Mail office.ti-be@bfh.ch.

6 Kompetenzprofil



Kompetenzstufen

1. Kenntnisse/Wissen
2. Verstehen
3. Anwenden
4. Analyse
5. Synthese
6. Beurteilung

7 Kursübersicht

Kurs / Lehreinheit	Lektionen	Dozierende
Tooling und Datenmanagement	24	Rudolf Farys Lukas Hobi
Deskriptive Statistik und mathematische Grundlagen	12	Michel Krebs
Statistisches Testen	28	Michel Krebs
Grafische Datenexploration und Datenvisualisierung	20	Oliver Hümbelin Benjamin Wiederkehr
Regressionsanalyse	16	Raul Gimeno
Zeitreihenanalyse und Prognosen	16	Raul Gimeno
Data Mining	24	Andrea Giovannini
Optionale Wahltage – Kausalanalyse – Open Data – Neuronale Netze – Zeitreihen und ARIMA, Regression von Qualitativen Variablen	je 8	Diverse
Total obligatorische Kurse	140	

Das CAS umfasst insgesamt 12 ECTS-Punkte. Für die einzelnen Kurse ist entsprechend Zeit für Selbststudium, Prüfungsvorbereitung etc. einzurechnen.

8 Kursbeschreibungen

Nachfolgend sind die einzelnen Kurse dieses Studienganges beschrieben.

Der Begriff Kurs schliesst alle Veranstaltungstypen ein, es ist ein zusammenfassender Begriff für verschiedene Veranstaltungstypen wie Vorlesung, Lehrveranstaltung, Fallstudie, Living Case, Fach, Studienreise, Semesterarbeiten usw.

8.1 Tooling und Datenmanagement

Lernziele	<p>Einführung in das Statistiksoftwarepaket R, welches sich zunehmend zu einer Standardsprache der Datenanalyse entwickelt. Folgende Themen sollen behandelt werden: Grundlegende Funktionsweise von R, Datenmanagement, einfache Auswertungen sowie die Einbindung von R in den persönlichen Workflow (Umgang mit unterschiedlichen Datenquellen/-formaten und Einbindung von Resultaten in die Textverarbeitung (Word/Latex/HTML).</p> <p>Die Teilnehmenden werden befähigt, R für eigene Anwendungen einzusetzen, und kennen die wichtigsten „Anlaufstellen“ (Literatur und Onlinehilfen) um das bestehende Wissen weiterzuentwickeln und auf neue Anwendungen auszuweiten.</p>
Themen und Inhalte	<p>Einführung in R</p> <ul style="list-style-type: none">– Grundlegende Funktionen der Datenmanipulation und Objektsprache R.– Benutzung von R-Studio.– Einlesen und Aufbereitung von Daten mit R Studio: Datensätze laden, verbinden, umformen, aggregieren (u.a. mit dplyr/data.table, reshape2) und exportieren.– Berechnung von Verteilungs- und Zusammenhangs-Massen, statistischer Unsicherheit, Darstellung in Tabellen.– Erstellen und Interpretation von Kennzahlen.– Grafische Darstellung von Ergebnissen: built-in plot Befehle, Library ggplot2.
Lehrmittel	<ul style="list-style-type: none">– Skript/Readings auf eLearning Plattform– Literaturempfehlungen Nr. 2, 4– http://tryr.codeschool.com/– http://shiny.rstudio.com/– R Instructor on Android, S. Murphy– Onlinehilfen:<ul style="list-style-type: none">– Generell: Suchmaschinen– Spezifische Websites (stackoverflow, Quick R (statmethods.net), R mailing lists, google groups, you tube channels...)

8.2 Deskriptive Statistik und mathematische Grundlagen

Lernziele	Die Teilnehmenden: <ul style="list-style-type: none">– Erlernen die grundlegenden Konzepte der deskriptiven Statistik. Nach Absolvierung des Moduls sind sie in der Lage, Daten aufzubereiten und zu präsentieren.– Kennen Matrizen und sind in der Lage, elementare Matrizenoperationen korrekt durchzuführen.
Themen und Inhalte	<ul style="list-style-type: none">– Statistische Kennzahlen– Verteilungen– Lage- und Streuungsmasse– Quantile– Bivariate Datenanalyse– Matrizen und Matrizenoperationen
Lehrmittel	<ul style="list-style-type: none">– Folien/Skript/Readings auf eLearning Plattform– Literaturempfehlung Nr.1

8.3 Statistisches Testen

Lernziele	Die Teilnehmenden: <ul style="list-style-type: none">– Erlernen die Grundlagen der Wahrscheinlichkeitsrechnung und der schliessenden Statistik.– Kennen insbesondere die statistischen Konzepte der Schätzung, des Hypothesentests sowie des Vertrauensintervalls und können diese in der Praxis anwenden.
Themen und Inhalte	<ul style="list-style-type: none">– Wahrscheinlichkeitsrechnung– Zufallsvariable– Summen von Zufallsvariablen– Vertrauensintervalle und Hypothesentests– Lineare Einfachregression– Schätzen– Bestimmtheitsmass– Prognose
Lehrmittel	<ul style="list-style-type: none">– Folien/Skript/Readings auf eLearning Plattform– Literaturempfehlung Nr. 1

8.4 Grafische Datenexploration und Datenvisualisierung

Lernziele	<p>Die Teilnehmenden:</p> <ul style="list-style-type: none"> – Können den Nutzen von explorativer Datenanalyse und Datenvisualisierungen im Prozess der Datenanalyse einschätzen. – Sind mit den zentralen Techniken der Datenexploration vertraut und können diese mit R umsetzen. – Sind fähig basierend auf den Gestaltungs-Prinzipien der Datenvisualisierung, anschauliche Graphiken zu erstellen. – Lernen die Möglichkeiten von interaktiven Datenvisualisierungen kennen und können eigene, einfache Applikationen programmieren.
Themen und Inhalte	<ul style="list-style-type: none"> – Bedeutung und Funktion von explorativer Datenanalyse und Datenvisualisierungen im „Epicycles of Analysis“ – Techniken der Datenexploration mit R und ggplot2 <ul style="list-style-type: none"> – „Grammar of Graphics“ – Klassische univariate Techniken: Balken- und Kuchendiagramm, Histogramm, Boxplots – Bi- und multivariate Techniken: Streudiagramme, Techniken zum Vergleich von Verteilungen und zur Visualisierung von Entwicklungen über die Zeit, Mosaikplot, Correlogram und generalisierte Scatterplot Matrizen, small multiples mit Trellis-Plots, Radar-Charts – Erkennen räumlicher Muster, Choroplethenkarte und Punkteverteilungskarten – Visualisierungen als Mittel der Kommunikation <ul style="list-style-type: none"> – Gestalt-Prinzipien der Datenvisualisierung – Interaktive Graphiken als Webapplikationen
Lehrmittel	<ul style="list-style-type: none"> – Folien/R-Skripte und über e-learning bereitgestellte Texte – Literaturempfehlungen Nr. 3, 6, 7

8.5 Regressionsanalyse

Lernziele	<p>Die Teilnehmenden:</p> <ul style="list-style-type: none"> – Lernen die Regressionsanalyse als vielseitiges und klassisches Instrument kennen, mit dem Beziehungen zwischen abhängigen und unabhängigen Grössen hergestellt und Prognosen erstellt werden können. – Können Methoden und Kriterien zur Überprüfung eines Modells, möglicher Einschränkungen, möglicher Modellfehler und zur Einschätzung der Prognosequalität anwenden.
Themen und Inhalte	<ul style="list-style-type: none"> – Lineare Regression – Multiple lineare Regression – Lineare Restriktionen – Prognose und Prognosequalität – Analyse der Modellstruktur: <ul style="list-style-type: none"> – Multikollinearität – Heteroskedastizität – Autokorrelation – Mehrgleichungssysteme – Nichtlineare Zusammenhänge
Lehrmittel	<ul style="list-style-type: none"> – Folien/Buch/Readings auf eLearning Plattform – Literaturempfehlung Nr. 5

8.6 Zeitreihen und Prognosen

Lernziele	<p>Die Teilnehmenden:</p> <ul style="list-style-type: none">– Kennen die Eigenschaften und Charakteristika von Moving-Average und Autoregressiven.– Können eine Zeitreihe anhand verschiedener Methoden glätten.– Können zwischen trend-stationären und differenz-stationären Prozessen unterscheiden.– Verstehen die Problematik des sogenannten spurious-regression-Problem.– Können die Unit-root-Tests (Einheitswurzel-Test) anwenden.
Themen und Inhalte	<ul style="list-style-type: none">– Glättungsverfahren:<ul style="list-style-type: none">– Gleitende Durchschnitte– Exponentielle Glättung– Holt-Verfahren– Saisonbereinigung:<ul style="list-style-type: none">– Additives und multiplikatives Modell– Regressionsverfahren– Holt-Winters Verfahren– Stochastische Prozesse– Unit roots-Tests– Korrelogramm
Lehrmittel	<ul style="list-style-type: none">– Folien/Readings auf eLearning Plattform– Literaturempfehlung Nr. 5

8.7 Data Mining

Lernziele	<p>Die Teilnehmenden:</p> <ul style="list-style-type: none"> – Können selbständig ein unstrukturiertes Datenset bereinigen und in eine strukturierte Form bringen. – Können sich selbständig anhand der vorhandenen Daten und der aktuellen Fragestellung für die richtige Data Mining Disziplin sowie einen geeigneten Algorithmus entscheiden. – Verstehen quantitative Kennzahlen um die Genauigkeit der Ergebnisse zu überprüfen. – Vertiefen die Anwendung von Standard-Software zur Datenanalyse (R und ApacheSpark). – Kennen Alternativen um grosse Datenmengen, die nicht auf eine einzelne Maschine passen, zu analysieren.
Themen und Inhalte	<ul style="list-style-type: none"> – Überblick Supervised vs. Non-Supervised Machine Learning – Pre-Processing / Feature Extraction / Dimension Reduction – Wiederholung: Regression – Association Rule Mining – Clustering – Classification – Bewertung von Ergebnissen – Muster in Daten erkennen, Explorative Datenanalyse – Parallelisierung von Algorithmen mit ApacheSpark – Ausflug 1: Erkennung von handschriftlichen Zahlensymbolen mit neuronalen Netzen – Ausflug 2: Neural-Art: Wie neuronale Netze lernen berühmte Maler zu imitieren – Ausflug 3: Kaggle Wettbewerb: Treten Sie gegen die internationalen Meister des Fachs an
Lehrmittel	<ul style="list-style-type: none"> – Folien/Skript/Readings auf eLearning Plattform

8.8 Wahltag Open Data

Lernziele	<p>Die Teilnehmenden setzen sich mit den Möglichkeiten global verfügbarer Daten auseinander.</p>
Themen und Inhalte	<ul style="list-style-type: none"> – Die Relevanz von Daten in einer zunehmend digitalisierten Welt <ul style="list-style-type: none"> – für ein globales Wissenssystem – für Ökonomie und Wirtschaft – Open Data in der Schweiz – Öffentlich zugängliche Daten <ul style="list-style-type: none"> – Open Data nutzen (Gruppenarbeit)

8.9 Wahltag Kausalanalyse

Allgemein	Datenanalysten sind in der Regel weniger an reinen Korrelationen interessiert, sondern an Fragen zu Ursache und Wirkung. Kausale Zusammenhänge sind jedoch mit Ausnahme experimentell erhobener Daten nur schwer nachzuweisen. Experimente sind jedoch in vielen Fällen nicht durchführbar, z.B. weil sie zu teuer oder unethisch sind. Auch lassen sich experimentelle Ergebnisse oft nicht auf die reale Welt eins zu eins übertragen. Basierend auf dem Potential Outcomes Framework wurden in den letzten zwei Jahrzehnten verschiedene Methoden zur Herleitung von Ursache/Wirkungs-Beziehungen für nicht experimentelle Daten entwickelt bzw. wiederentdeckt und erfreuen sich zunehmender Beliebtheit in der Datenanalyse.
Lernziele	Nach Abschluss des Moduls sind die Teilnehmenden sensibilisiert für typische Probleme bei der Datenanalyse und haben ein Grundverständnis, wann und wie kausale Methoden zum Einsatz kommen können. Praktische Erfahrungen mit diesen Methoden werden anhand von Hands-on-Sessions mit R geübt.
Themen und Inhalt	<ul style="list-style-type: none"> – Potential Outcomes Framework – Gerichtete azyklischen Graphen – Difference-in-Differenz Schätzer – Regression Discontinuity Design – Überblick zu weiteren Verfahren wie z.B. Matching und Instrumentalvariablen
Lehrmittel	<ul style="list-style-type: none"> – Folien/Skript/Readings auf eLearning Plattform – Bücher zur Orientierung und Vertiefung: <ul style="list-style-type: none"> – Angrist/Pischke, 2008: Mostly Harmless Econometrics – Morgan/Winship, 2011: Counterfactuals and Causal Inference

8.10 Wahltag Zeitreihen und Regression

Themen und Inhalte	Ergänzungen zu den Themen Zeitreihen mit ARIMA (4 Lektionen) und Regression von qualitativen Variablen (4 Lektionen).
--------------------	---

8.11 Wahltag Zeitreihen und Regression

Themen und Inhalte	Ergänzungen zum Fach Data Mining zu neuronalen Netze und Deep Learning.
--------------------	---

9 Kompetenznachweis

Für die Anrechnung der 12 ECTS-Punkte ist das erfolgreiche Bestehen der Qualifikationsnachweise (Prüfungen, Projektarbeiten) erforderlich, gemäss folgender Aufstellung:

Kompetenznachweis	Gewicht	Art der Qualifikation	Erfolgsquote Studierende
Tooling und Datenmanagement	2	Übungen / Hausaufgabe	0 - 100 %
Deskriptive Statistik und statistisches Testen	2	Übungen + Schriftlich 60' / Open Book, Laptop	0 - 100 %
Grafische Datenexploration und Datenvisualisierung	2	Übungen + Schriftlich 30' / Open Book	0 - 100 %
Regressionsanalyse, Zeitreihen und Prognosen	2	Schriftlich 60' / Closed Book, Laptop	0 - 100 %
Data Mining	2	Übungen + Schriftlich 45' / Open Book, Laptop	0 - 100 %
Wahltag	0	Nicht Teil des Kompetenznachweises	
Gesamtgewicht / Erfolgsquote	10		0 - 100 %
ECTS-Note			A - F

Jeder Studierende kann in einem Kompetenznachweis eine Erfolgsquote von 0 bis 100% erreichen. Die gewichtete Summe aus den Erfolgsquoten pro Thema und dem Gewicht des Themas ergibt eine Gesamterfolgsquote zwischen 0 und 100%. Die Gesamterfolgsquote wird in eine ECTS Note A bis E umgerechnet, gemäss Studienreglement. Weniger als 50% Gesamterfolgsquote ergibt eine ungenügende Note F.

10 Lehrmittel

Die nachfolgend aufgeführten, obligatorischen Lehrmittel sind wesentlich für das Lernen während des geführten Unterrichtes. Sie sind durch die Studierenden zu beschaffen.

Nr	Titel	Autoren	Verlag	Jahr	ISBN Nr.
1.	Statistik ohne Angst vor Formeln	Andreas Quatember	Pearson Studium	2017	978-3-86894-320-7

Für das Einlesen und als Begleitmaterial werden nachfolgend aufgeführte Bücher empfohlen. Die Beschaffung liegt im Ermessen der Studierenden.

Nr	Titel	Autoren	Verlag	Jahr	ISBN Nr.
2.	R in Action – data analysis and graphics with R	Robert I. Kabacoff	Manning	2016	978-1-61729-138-8
3.	Visualize This: The Flowing Data Guide to Design, Visualization, and Statistics	Nathan Yau	John Wiley & Sons Inc	2011	978-0-470-94488-2
4.	An R Companion to Applied Regression	John Fox, Harvey Sanford Weisberg	SAGE Publications Inc	2011	9781412975148
5.	Introduction to Modern Time Series Analysis	Uwe Hassler, Gebhard Kirchgässner, Jürgen Wolters	Springer	2013	978-3-642-44029-8
6.	R Graphics Cookbook, 2nd Edition Practical Recipes for Visualizing Data	Winston Chang	O'Reilly UK Ltd.	2018	9781491978597
7.	Storytelling with Data: A Data Visualization Guide for Business Professionals	Cole Nussbaumer Knaflic	Wiley	2015	978-1119002253

Weitere Empfehlungen und Hinweise bei den einzelnen Lehrveranstaltungen.

11 Dozierende

Vorname Name	Firma	E-Mail
Michel Krebs	BFH	michel.krebs@bfh.ch
Oliver Hümbelin	BFH	oliver.huembelin@bfh.ch
Rudolf Farys	UniBe	rudolf.farys@soz.unibe.ch
Raul Gimeno	BFH	rauldiego.gimeno@bfh.ch
Andrea Giovannini	IBM	andrea.giovannini@ch.ibm.com
Lukas Hobi	BFH	lukas.hobi@bfh.ch
Oleg Lavrovsky	datalets.ch	oleg@datalets.ch

12 Organisation

CAS-Leitung:

Prof. Dr. Arno Schmidhauser, Departement Technik und Informatik

Tel: +41 31 84 83 275

E-Mail: arno.schmidhauser@bfh.ch

Dr. Oliver Hümbelin, Departement Soziale Arbeit

Tel: +41 31 848 36 97

E-Mail: oliver.huembelin@bfh.ch

CAS-Administration:

Andrea Moser

Tel: +41 31 84 83 211

E-Mail: andrea.moser@bfh.ch

Dokumenteninformation

Study Guide CAS Datenanalyse
Stand: 3. September 2019

Dieser Study Guide gilt für die Publikation ab Herbstsemester 2019.

Während der Durchführung des CAS können sich Anpassungen bezüglich Inhalten, Lernzielen, Dozierenden und Kompetenznachweisen ergeben. Es liegt in der Kompetenz der Dozierenden und der Studienleitung, aufgrund der aktuellen Entwicklungen in einem Fachgebiet, der konkreten Vorkenntnisse und Interessenslage der Teilnehmenden, sowie aus didaktischen und organisatorischen Gründen Anpassungen im Ablauf eines CAS vorzunehmen.

Berner Fachhochschule
Technik und Informatik
Weiterbildung
Wankdorffeldstrasse 102
CH-3014 Bern

Telefon +41 31 848 31 11
Email: office.ti-be@bfh.ch

bfh.ch/ti/weiterbildung
ti.bfh.ch/cas-da