

Studienplan für den CAS-Studiengang Advanced Statistical Data Science



b
UNIVERSITÄT
BERN

17. Oktober 2019

Der CAS-Studiengang Advanced Statistical Data Science ist eine universitäre Weiterbildung, die zur Erteilung des „Certificate of Advanced Studies in Advanced Statistical Data Science, Universität Bern (CAS ASDS Unibe)“ führt. Rechtsgrundlage ist das Reglement der philosophisch-naturwissenschaftlichen Fakultät für die Weiterbildungsstudiengänge in Statistical Data Science vom 14.11.2019.

1. Studiengangsziele

Ziele

Die Teilnehmenden erwerben, aufbauend auf den Vorkenntnissen aus dem CAS-Studiengang „Statistical Data Science“, Kompetenzen in angewandter Statistik und Data Science: Die Teilnehmenden können

- a. die im Studiengang vorgestellten fortgeschrittenen und spezialisierten Methoden und Modelle der angewandten Statistik erläutern;
- b. die theoretischen Konzepte, die den Methoden und Modellen gemäss a. zugrunde liegen, erläutern;
- c. für konkrete Probleme der Datenanalyse geeignete Methoden und Modelle gemäss a. auswählen und diesen Entscheid begründen;
- d. die Methoden und Modelle gemäss a. korrekt auf entsprechende Problemstellungen anwenden – einerseits unter Verwendung einer vorgegebenen Statistik-Software, andererseits in einfachen Fällen auch ohne Software;
- e. die Ergebnisse von Analysen gemäss a. mit Bezug auf die ursprüngliche Fragestellung korrekt interpretieren.

2. Umfang, Ziele und Inhalte der Studiengangselemente

Umfang

Der Studiengang umfasst 21.5 Kurstage (ca. 146 Präsenzstunden, inklusive 2 Tage mit Leistungskontrollen) und 16 ECTS-Credits (ca. 400 bis 480 Arbeitsstunden insgesamt). Davon entfallen 9 ECTS-Credits auf die Pflichtmodule P2.1 bis P2.6, 1 ECTS-Credit auf das Praxismodul und 6 ECTS-Credits auf drei der Wahlmodule W2.1 bis W2.8, von denen pro Durchführung des Studiengangs jeweils mindestens fünf angeboten werden.

Ziele und Inhalte	<p>Der Umfang, die spezifischen Ziele und die grundlegenden Inhalte der Studiengangselemente werden nachfolgend beschrieben. Stichwortartige Angaben zum genaueren Inhalt der Pflicht- und Wahlmodule sind im Anhang aufgeführt.</p>
Modul P2.1	<p>Unüberwachtes Lernen und Dimensionsreduktion Umfang: 4 Halbtage (1.5 ECTS-Credits)</p> <p>In diesem Modul geht es darum, Strukturen von Daten herauszuarbeiten und so zu neuen Erkenntnissen zu gelangen. Beim unüberwachten Lernen versucht man (z. B. mit einer Clusteranalyse), Beobachtungen zu sinnvollen Gruppen zusammenzufassen. Bei der Hauptkomponenten- und der Faktoranalyse ist das Ziel, komplexes multivariates Datenmaterial übersichtlich und leichter interpretierbar aufzubereiten, indem höherdimensionale Daten mit möglichst wenig Informationsverlust in wenigen Dimensionen repräsentiert werden.</p>
Modul P2.2	<p>Lineare gemischte Modelle Umfang: 4 Halbtage (1.5 ECTS-Credits)</p> <p>Gemischte Modelle beinhalten sowohl feste Effekte, wie sie in klassischen linearen Modellen verwendet werden, als auch zufällige Effekte. Zufällige Effekte verwendet man, wenn in einer Untersuchung nur eine zufällige Auswahl der theoretisch denkbaren Stufen eines Faktors vorkommt. Gemischte Modelle sind in vielen Situationen nützlich, in denen Abhängigkeiten zwischen verschiedenen Beobachtungen bestehen.</p>
Modul P2.3	<p>Analyse kategorieller Daten Umfang: 4 Halbtage (1.5 ECTS-Credits)</p> <p>Kategorielle Daten können nicht normalverteilt sein, so dass viele bekannte statistische Verfahren nicht verwendet werden können. Dieses Modul stellt passende Modelle für die Analyse kategorieller Daten vor, mit denen u. a. Abhängigkeiten zwischen kategoriellen Größen untersucht werden können. Es baut dabei auf Ansätzen aus vorangehenden Modulen auf.</p>
Modul P2.4	<p>Rechenintensive Verfahren Umfang: 4 Halbtage (1.5 ECTS-Credits)</p> <p>Neuere statistische Methoden sind nur dank der Entwicklung passender Software und genügend schneller Rechner entstanden. Einfache exakte statistische Verfahren, wie der exakte Test von Fisher, erreichten erst dank dem Computer ihre heutige Verbreitung. Neben den exakten Tests werden in diesem Modul weitere rechenintensive Methoden behandelt, wie das Bootstrap-Verfahren, das Jackknife-Verfahren oder die multiple Imputation als Verfahren zur Analyse von Daten mit fehlenden Werten.</p>

Modul P2.5

Bayes-Statistik

Umfang: 4 Halbtage (1.5 ECTS-Credits)

Nach der Durchführung eines Experiments oder einer Studie sind die Daten bekannt, aber der interessierende Parameter unbekannt. In der Bayes-Statistik soll die Unsicherheit über diesen Parameter durch eine Wahrscheinlichkeitsverteilung quantifiziert werden. Vorkenntnisse (oder zumindest Vermutungen) über den Parameter, die bereits vor der Durchführung vorliegen, werden mit einer a-priori-Wahrscheinlichkeitsverteilung erfasst. Die Kombination der a-priori-Verteilung und der Information der Daten mittels Bayes-Theorem liefert die a-posteriori-Verteilung des unbekanntes Parameters. Daraus lassen sich Schlüsse ziehen, ähnlich wie mit den klassischen (frequentistischen) Methoden.

Modul P2.6

Prädiktion

Umfang: 4 Halbtage (1.5 ECTS-Credits)

Vorhersagen für unsichere zukünftige Ereignisse sind allgegenwärtig. Normalerweise werden Vorhersagen als Punktvorhersagen gegeben, die in irgendeinem Sinne eine „beste Vermutung“ darstellen. Um der Unsicherheit Rechnung zu tragen, sind probabilistische Vorhersagen zu bevorzugen, das heißt es wird eine Wahrscheinlichkeitsverteilung für das zukünftige Ereignis angegeben. Bei Punktvorhersagen ist es unabdingbar zu spezifizieren, was genau eine „beste Vermutung“ bedeutet. In diesem Modul werden die entscheidungstheoretischen Grundlagen behandelt, um sinnvolle Vorhersagen zu machen, die Qualität gegebener Vorhersagen zu evaluieren und Vorhersagen zu vergleichen.

Praxismodul

Praxismodul

Umfang: 3 Halbtage (1 ECTS-Credit)

Dieses Modul unterstützt die Teilnehmenden bei der Vernetzung der Inhalte der vorangehenden Module und bei der Übertragung in die Praxis. Die Teilnehmenden stellen eigene Daten und dazugehörige statistische Herausforderungen im Plenum vor. Basierend auf den Modulen der beiden CAS-Studiengänge (CAS in Statistical Data Science und CAS in Advanced Statistical Data Science) werden mögliche Analysemethoden im Plenum besprochen. Anschliessend dokumentieren die Teilnehmenden diese Diskussion in einem schriftlichen Bericht.

Modul W2.1

Analyse hochdimensionaler Daten

Umfang: 4 Halbtage (2 ECTS-Credits)

Von hochdimensionalen Daten spricht man insbesondere dann, wenn mehr Variablen als Beobachtungseinheiten vorliegen. Klassische statistische Verfahren können in solchen Situationen oft gar nicht angewendet werden, oder aber sie liefern Ergebnisse, die sich viel zu eng an die vorliegenden Daten anlehnen und deshalb nicht verallgemeinerbar sind. Um eine solche Überanpassung („overfitting“) zu vermeiden, kann man beispielsweise das Kriterium zur

Anpassung eines linearen Modells so verändern, dass einfachere Modelle bevorzugt werden.

Modul W2.2

Zeitreihenanalyse

Umfang: 4 Halbtage (2 ECTS-Credits)

Die Wissenschaft der Zeitreihen entspringt dem alten Wunsch, die Zukunft voraussagen zu können. Das Wetter oder der Verlauf der Aktienkurse von morgen hängen erfahrungsgemäss mit den Werten von heute zusammen. Zeitreihen-Modelle berücksichtigen diese Korrelation aufeinander folgender Werte und ermöglichen dadurch eine Analyse der Zusammenhänge und eine Prognose zukünftiger Werte.

Modul W2.3

Analyse von Ereigniszeiten

Umfang: 4 Halbtage (2 ECTS-Credits)

In vielen Anwendungsbereichen ist die Zeitdauer bis zum Eintreten eines bestimmten Ereignisses von Interesse. Je nach Kontext spricht man auch von einer Verweildauer (in einem bestimmten Zustand) oder einer Überlebenszeit; in technischen Anwendungen hat sich für entsprechende Methoden der Begriff der Zuverlässigkeitsanalyse etabliert. Die interessierende Zeitdauer ist zum Zeitpunkt der Datenauswertung oft nicht bei allen Beobachtungseinheiten bekannt, weil bei einigen das entsprechende Ereignis noch nicht eingetreten ist. Solche Daten sind zensiert. Dieses Modul bietet eine Einführung in die Analyse von Ereigniszeiten und umfasst u. a. Verfahren zum Vergleich von Gruppen und zur Beurteilung des Zusammenhangs von quantitativen oder qualitativen Einflussgrössen mit einer Ereigniszeit als Zielgrösse.

Modul W2.4

Versuchsplanung

Umfang: 4 Halbtage (2 ECTS-Credits)

Je nach Fachgebiet und Fragestellung kann man nicht nur in Beobachtungsstudien erheben, was ohnehin geschehen würde, sondern man kann auch in experimentellen Studien gezielt die Werte gewisser Grössen festlegen. Dieses Modul befasst sich mehrheitlich mit solchen Versuchen, mit verschiedenen gebräuchlichen Versuchsplänen und der Analyse der damit erhobenen Daten. Daneben werden Themen behandelt, die auch für die Planung einer Beobachtungsstudie relevant sein können, wie etwa verschiedene Arten von Hypothesen und die Bestimmung des benötigten Stichprobenumfangs.

Modul W2.5

Fortgeschrittene nichtparametrische Methoden

Umfang: 4 Halbtage (2 ECTS-Credits)

Dieses Modul setzt das Modul „Nichtparametrische Methoden“ aus dem CAS in Statistical Data Science fort, wobei hauptsächlich nichtparametrische Methoden bei multivariaten Daten und bei zeitlich abhängigen Variablen diskutiert werden. Zudem werden weitere allgemeinere, nichtparametrische Methoden ausführlich behandelt.

Modul W2.6

Stichprobenerhebungen

Umfang: 4 Halbtage (2 ECTS-Credits)

Stichprobenerhebungen werden insbesondere in der Wirtschafts- und Sozialstatistik intensiv verwendet. Dieses Modul führt die Methode ein und geht näher auf die Planung und Auswertung von einfachen, geschichteten, geklumpten oder räumlichen Zufallsstichproben in endlichen Populationen ein. Auch praktische Probleme wie Antwortausfälle und Ausreisser werden angesprochen.

Modul W2.7

Risiko- und Extremwertanalyse

Umfang: 4 Halbtage (2 ECTS-Credits)

Die Extremwerte – Minimum und Maximum – sind beispielsweise in ökologischen Fragestellungen wie auch bei Risikobeurteilungen von zentraler Bedeutung. Ihre statistische Modellierung und Analyse erfordert spezifische Verfahren und Verteilungen wie die Gumbel- und Paretoverteilungen. Das Ziel besteht darin, zukünftige Extremwerte abzuschätzen und Überschreitungswahrscheinlichkeiten von gegebenen Grenzwerten zu bestimmen. Risikobeurteilungen sind das Ziel dieser Methoden.

Modul W2.8

Weitere aktuelle Themen (gemäss Ausschreibung)

Umfang: 4 Halbtage (2 ECTS-Credits)

Die Programmleitung kann bei der Ausschreibung einer Durchführung des Studiengangs weitere aktuelle Themen festlegen, die im Rahmen eines Wahlmoduls behandelt werden.

3. Leistungsnachweise im Studiengang

Leistungsnachweise

Der Leistungsnachweis zum Praxismodul umfasst die Präsentation der eigenen Daten und statistischen Herausforderungen sowie den schriftlichen Bericht zur Diskussion im Plenum.

Die Leistungsnachweise zu den übrigen Modulen werden in der Regel in Form von schriftlichen Prüfungen erbracht. Wiederholungen von Leistungskontrollen können auf Beschluss der Programmleitung auch mündlich durchgeführt werden. Die Programmleitung kann vor Beginn des Studiengangs festlegen, dass für gewisse Module die Leistungsnachweise in Form von schriftlichen Hausarbeiten erbracht werden.

Die Programmleitung entscheidet aufgrund der Bewertung der Leistungsnachweise und der Erfüllung der weiteren Leistungsanforderungen über das Bestehen und die Erteilung des Zertifikats.

Näheres regeln die Richtlinien der Programmleitung zu den Leistungskontrollen.

4. Schlussbestimmungen

Änderungen des Studienplans	Änderungen des Studienplans unterliegen der Genehmigung durch die philosophisch-naturwissenschaftliche Fakultät. Ausgenommen sind Änderungen des Anhangs, die von der Programmleitung beschlossen werden und der Genehmigung des Studienausschusses der Fakultät unterliegen.
Inkrafttreten	Unter der Bedingung, dass das Reglement der philosophisch-naturwissenschaftlichen Fakultät für die Weiterbildungsstudiengänge in Statistical Data Science vom 14.11.2019 bis zu diesem Zeitpunkt vom Senat genehmigt worden ist, tritt dieser Studienplan auf den 01.03.2020 in Kraft.
17.10.2019	Von der Programmleitung beschlossen: Die Vorsitzende Prof. Dr. Johanna F. Ziegel
14.11.2019	Von der philosophisch-naturwissenschaftlichen Fakultät genehmigt: Der Dekan Prof. Dr. Zoltan Balogh

Anhang: Stichworte zu den Inhalten der Pflicht- und Wahlmodule

Modul P2.1	Unüberwachtes Lernen und Dimensionsreduktion Clusteranalyse, Ähnlichkeits- und Distanzmasse, Hauptkomponentenanalyse, Faktoranalyse, Rotation, latente Variablen
Modul P2.2	Lineare gemischte Modelle Feste und zufällige Effekte, Abhängigkeitsstrukturen, Modellbildung, Modelldiagnostik, Mehrebenen-Modelle/hierarchische Modelle
Modul P2.3	Analyse kategoriieller Daten Zusammenhangs- und Übereinstimmungsmasse für kategorielle Variablen, stratifizierte Tabellen, verbundene Merkmale, abhängige Beobachtungen, verallgemeinerte lineare gemischte Modelle
Modul P2.4	Rechenintensive Verfahren Monte-Carlo-Methode, exakte Tests, Bootstrap, Jackknife, fehlende Werte, Imputation
Modul P2.5	Bayes-Statistik Satz von Bayes, konjugierte Verteilungen, Bayes-Schätzung, Kreditabilitätsbereich, Test, Simulationsmethoden, Markov Chain Monte Carlo, hierarchische Modelle
Modul P2.6	Prädiktion Punkt- und Verteilungsvorhersagen, Evaluation und Vergleich von Vorhersagen, Prädiktionsbereiche vs. Konfidenzbereiche
Modul W2.1	Analyse hochdimensionaler Daten Ridge Regression, Lasso, Elastic Net, Boosting, Variablenselektion
Modul W2.2	Zeitreihenanalyse Zerlegung einer Zeitreihe in Trend, zyklische Komponente und Rest, saisonale Effekte, autoregressive (AR-)Modelle, Moving-Average-(MA-)Modelle, ARMA-Modelle, Vorhersage
Modul W2.3	Analyse von Ereigniszeiten Ausfallrate, Überlebenskurve, Kaplan-Meier-Schätzung, Logrank-Test, Cox-Regression, parametrische Modelle, Stratifizierung, konkurrierende Risiken, Frailty-Modelle

Modul W2.4

Versuchsplanung

Beobachtungsstudien und experimentelle Studien, Nicht-Unterlegenheit/Äquivalenz, Randomisierung, Stichprobenumfang, klassische Versuchspläne und deren Analyse, unvollständige Blöcke, Wirkungsflächen

Modul W2.5

Fortgeschrittene nichtparametrische Methoden

Allgemeine nichtparametrische Varianzanalyse, relative Marginal-Effekte, Daten-Alignment, longitudinale Daten, multivariate Verfahren

Modul W2.6

Stichprobenerhebungen

Einfache, geschichtete, geklumpfte Stichproben, endliche Populationen, räumliche Stichprobenpläne, Quantifizierung der Unsicherheit

Modul W2.7

Risiko- und Extremwertanalyse

Maxima, Minima, Exzesse, Extremwert- und (verallgemeinerte) Pareto-Verteilungen, Gumbel-, Fréchet- und Weibull-Verteilungen, Risikobeurteilungen extremer oder rarer Ereignisse, Value-at-Risk und andere Risikomasse