

Studienplan für den CAS-Studiengang Statistical Data Science



b
UNIVERSITÄT
BERN

17. Oktober 2019

Der CAS-Studiengang Statistical Data Science ist eine universitäre Weiterbildung, die zur Erteilung des „Certificate of Advanced Studies in Statistical Data Science, Universität Bern (CAS SDS 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 Kompetenzen in angewandter Statistik und Data Science: Die Teilnehmenden können

- a. die im Studiengang vorgestellten zentralen Methoden und Modelle der angewandten Statistik erläutern;
- b. die im Studiengang vorgestellten Methoden zur Visualisierung von Daten erläutern;
- c. die theoretischen Konzepte, die den Methoden und Modellen gemäss a. und b. zugrunde liegen, erläutern;
- d. für konkrete Probleme der Datenanalyse geeignete Methoden und Modelle gemäss a. und b. auswählen und diesen Entscheid begründen;
- e. die Methoden und Modelle gemäss a. und b. korrekt auf entsprechende Problemstellungen anwenden – einerseits unter Verwendung einer vorgegebenen Statistik-Software, andererseits in einfachen Fällen auch ohne Software;
- f. die Ergebnisse von Analysen gemäss a. und b. mit Bezug auf die ursprüngliche Fragestellung korrekt interpretieren;
- g. grundlegende Konzepte der verwendeten Statistik-Software erläutern und anwenden, mittels Programmierung effizient Daten aufbereiten, bestehende Funktionen kombinieren sowie einfache eigene Funktionen schreiben.

2. Umfang, Ziele und Inhalte der Studiengangelemente

Umfang	<p>Der Studiengang umfasst 24 Kurstage (ca. 160 Präsenzstunden, inklusive 2 Tage mit Leistungskontrollen) und 16 ECTS-Credits (ca. 400 bis 480 Arbeitsstunden insgesamt, ohne Projektarbeit). Davon entfallen 14 ECTS-Credits auf die Pflichtmodule P1.1 bis P1.6 und 2 ECTS-Credits auf eines der beiden Wahlmodule W1.1 oder W1.2. Darüber hinaus kann freiwillig eine Projektarbeit im Umfang von 1.5 ECTS-Credits verfasst werden.</p>
Ziele und Inhalte	<p>Der Umfang, die spezifischen Ziele und die grundlegenden Inhalte der Studiengangelemente werden nachfolgend beschrieben. Stichwortartige Angaben zum genaueren Inhalt der Pflicht- und Wahlmodule sind im Anhang aufgeführt.</p>
Modul P1.1	<p>Deskriptive Statistik und Wahrscheinlichkeitsrechnung Umfang: 10 Halbtage (3 ECTS-Credits)</p> <p>Dieses Modul dient dazu, eine gemeinsame Basis aus einfachen Begriffen der Statistik und aus den Grundlagen der Wahrscheinlichkeitstheorie zu schaffen. Im ersten Teil wird die deskriptive Statistik repetiert. Insbesondere werden die verschiedenen Arten von Daten, die Kennzahlen einer Stichprobe und die typischen Darstellungsformen behandelt. Der zweite Teil ist der Wahrscheinlichkeitsrechnung gewidmet. Das Hauptaugenmerk gilt den Zufallsvariablen, ihren Verteilungen und ihren Kennzahlen. Ausserdem werden wichtige mathematische Grundlagen repetiert, die in diesem und weiteren Modulen benötigt werden.</p>
Modul P1.2	<p>Statistik-Software und Visualisierung von Daten Umfang: 7 Halbtage (2.5 ECTS-Credits)</p> <p>In diesem Modul erwerben die Teilnehmenden Grundkenntnisse zur Programmierung in einer Statistik-Software, die in den weiteren Modulen verwendet wird. Im Vordergrund stehen die Datenaufbereitung und die deskriptive Statistik. Aus dem vorangehenden Modul bereits bekannte Diagramme sowie fortgeschrittene Visualisierungen werden mit Hilfe der Software erzeugt, und allgemeine Grundregeln der Visualisierung werden eingeführt.</p>
Modul P1.3	<p>Schliessende Statistik Umfang: 6 Halbtage (2.5 ECTS-Credits)</p> <p>Dieses Modul behandelt die Grundfragen der schliessenden Statistik: (i) Welcher Parameter passt am besten zu den Daten? (ii) Sind die Daten mit einem bestimmten Parameterwert vereinbar? (iii) Welche Parameterwerte sind mit den Beobachtungen vereinbar? Die dazugehörigen statistischen Verfahren sind das Schätzen, das Testen und das Bilden eines Vertrauensintervalls. Nach diesem Modul können univariate und bivariate Datensätze unter der klassischen Annahme von normalverteilten Daten ausgewertet werden.</p>

Modul P1.4

Einführung in lineare Modelle

Umfang: 9 Halbtage (3 ECTS-Credits)

In diesem Modul werden lineare Modelle eingeführt, eine Familie von Modellen, die es erlauben, den Wert eines Merkmals durch die Werte mehrerer anderer beobachtbarer Grössen zu erklären und vorauszusagen. Der Zusammenhang zwischen den Einflussgrössen und der Zielgrösse ist dabei in einem zu klärenden Sinn linear. Als wichtige Spezialfälle werden die lineare Regression und die Varianzanalyse (ANOVA) betrachtet. Die linearen Modelle bilden den Ausgangspunkt für die Definition allgemeinerer Modellfamilien, die in verschiedenen späteren Modulen des Weiterbildungsprogramms eingeführt werden. Im Rahmen dieses Moduls werden auch weitere mathematische Grundlagen (Vektor- und Matrixrechnung) aufgefrischt, die für die Arbeit mit linearen Modellen und in späteren Modulen hilfreich sind.

Modul P1.5

Nichtparametrische Methoden

Umfang: 4 Halbtage (1.5 ECTS-Credits)

Parametrische Tests erfordern relativ starke Annahmen über die Verteilung von Messgrössen. Wenn man diese Annahmen abschwächt, wird die Klasse der in Frage kommenden Verteilungen so gross, dass sie keine einfache Parametrisierung mehr hat. Deshalb spricht man von nichtparametrischen Tests. Zwei bekannte Beispiele nichtparametrischer Tests sind der Vorzeichentest und der Wilcoxon-Test. In diesem Modul werden univariate Verfahren präsentiert, die nicht von einer bestimmten Verteilungsfamilie ausgehen.

Modul P1.6

Einführung in die Klassifikation

Umfang: 4 Halbtage (1.5 ECTS-Credits)

Das Ziel der Klassifikation ist es, Beobachtungseinheiten anhand ihrer Eigenschaften mit Hilfe einer geeigneten Regel einer bestimmten Gruppe zuzuordnen, also zu klassifizieren. Die Gruppenzugehörigkeit ist dabei für eine Stichprobe („Trainingsdaten“) bekannt. Das Problem ist auch unter dem Namen „überwachtes Lernen“ (supervised learning) bekannt. In diesem Modul werden grundlegende Begriffe und einfache Klassifikationsmethoden eingeführt. Ausserdem wird für den Fall zweier Gruppen das Verfahren der logistischen Regression betrachtet, das auf den linearen Modellen aufbaut.

Modul W1.1

Vertiefung lineare und verallgemeinerte lineare Modelle

Umfang: 4 Halbtage (2 ECTS-Credits)

In diesem Modul werden einerseits Spezialfälle linearer Modelle vertieft betrachtet, insbesondere die Zweiweg- und Mehrweg-Varianzanalyse, bei welcher der kombinierte Einfluss zweier bzw. mehrerer Gruppenvariablen untersucht werden soll. Andererseits wird die Familie der verallgemeinerten linearen Modelle betrachtet. Bei diesen wird im Gegensatz zu den klassischen linearen Modellen nicht vorausgesetzt, dass die Zielgrösse normalverteilt ist, und es sind flexiblere, nicht-lineare Zusammenhänge zugelassen. Sowohl

das klassische lineare Modell als auch die logistische Regression gehören zu dieser Familie der verallgemeinerten linearen Modelle, aber beispielsweise auch Modelle für absolute Häufigkeiten und für mehrstufige kategorielle Zielgrößen.

Modul W1.2

Vertiefung Klassifikation und maschinelles Lernen

Umfang: 4 Halbtage (2 ECTS-Credits)

Zusätzlich zu den bereits bekannten Methoden zur Klassifikation werden in diesem Modul weitere Methoden verwendet. Klassische statistische Ansätze und solche aus dem Bereich des maschinellen Lernens werden einander gegenübergestellt, und ihre Vor- und Nachteile werden besprochen.

Projektarbeit

Projektarbeit (freiwillig)

Umfang: ca. 1 Arbeitswoche (1.5 ECTS-Credits)

Als freiwillige Ergänzung des Studiengangs kann unter der Leitung einer Dozentin oder eines Dozenten eine Problemstellung aus dem persönlichen, beruflichen bzw. wissenschaftlichen Umfeld bearbeitet werden. Das Vorgehen und die Resultate werden in einer schriftlichen Arbeit dokumentiert und im Studiengang präsentiert. Für die Betreuung wird eine zusätzliche Gebühr verrechnet.

3. Leistungsnachweise im Studiengang

Leistungsnachweise

Die Leistungsnachweise zu den 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.

Die Programmleitung kann Anforderungen an die Projektarbeiten festlegen und Projektarbeiten, die den Anforderungen nicht entsprechen, ablehnen. In diesem Fall wird eine Projektarbeit nicht im Diploma Supplement aufgeführt.

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 P1.1	Deskriptive Statistik und Wahrscheinlichkeitsrechnung Skalenniveaus, Diagramme, wichtige Kennzahlen, Grundbegriffe der Wahrscheinlichkeitsrechnung, wichtige diskrete und absolut stetige Verteilungen und ihre Eigenschaften, mathematische Grundbegriffe
Modul P1.2	Statistik-Software und Visualisierung von Daten Grundlagen der Programmierung in der Software R, Datenaufbereitung in R, beschreibende Statistik in R, fortgeschrittene Visualisierungsmethoden
Modul P1.3	Schliessende Statistik Punktschätzung, Intervallschätzung, Testtheorie, Vertrauensintervalle für Erwartungswerte, Alternativtests und Signifikanztests, Lokationstests für eine und für zwei Stichproben, Quantil-Quantil-Diagramm
Modul P1.4	Einführung in lineare Modelle Einweg-ANOVA, einfache lineare Regression, allgemeines lineares Modell, Vorhersage, Variablenselektion, Überprüfung der Modellvoraussetzungen, Kontraste, Vektoren und Matrizen
Modul P1.5	Nichtparametrische Methoden Tests für den Ein- und Zweistichprobenfall, nichtparametrische Schätzmethoden, nichtparametrische Varianzanalyse, Tests für kategoriale Daten, Korrelation und Abhängigkeit, nichtparametrische Anpassungstests
Modul P1.6	Einführung in die Klassifikation Qualität von Klassifikatoren, Trainingsdaten, Nearest-Neighbor-Verfahren, lineare Diskriminanzanalyse, logistische Regression
Modul W1.1	Vertiefung lineare und verallgemeinerte lineare Modelle Zweiweg-ANOVA, Mehrweg-ANOVA, verallgemeinerte lineare Modelle, Poisson-Regression, Logit-Modelle für mehrstufige Zielgrößen, Negativbinomial- und Gamma-Modelle
Modul W1.2	Vertiefung Klassifikation und maschinelles Lernen Klassifikationsbäume, Random Forests, neuronale Netze