



universität
wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Der Einfluss der Corona-Pandemie auf Studierenden-
leistungen an der Fakultät für Wirtschaftswissenschaften
der Universität Wien“

verfasst von / submitted by

Sarah Kaltenhauser, BSc (WU)

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of
Master of Science (MSc)

Wien, 2021 / Vienna 2021

Studienkennzahl lt. Studienblatt /
degree programme code as it appears on
the student record sheet:

UA 066 915

Studienrichtung lt. Studienblatt /
degree programme as it appears on
the student record sheet:

Masterstudium Betriebswirtschaft

Betreut von / Supervisor:

ao. Univ.-Prof. Mag. Dr. Michaela Schaffhauser-Linzatti

Inhaltsverzeichnis

1	Einleitung.....	1
2	Theoretischer Hintergrund.....	2
2.1	Lehr- und Lernerfolg	2
2.1.1	Konzeptdefinition	2
2.1.2	Einflüsse auf Lehr- und Lernerfolg	5
2.2	Soziale Eingebundenheit und die Bedeutung von Netzwerken	7
2.2.1	Soziale Eingebundenheit durch Peer-Gruppen	7
2.2.2	Soziale Eingebundenheit durch Netzwerke	8
2.3	Psychische Belastungen.....	11
2.4	Zeitmanagement.....	13
2.4.1	Prokrastination.....	14
2.4.2	Berufstätigkeit von Studierenden.....	15
2.5	Onlinelehre	17
2.5.1	Herausforderungen der Onlinelehre	17
2.5.2	Verwendete Technologien im SS 2020 und WS 2020/21 an der Fakultät für Wirtschaftswissenschaften der Universität Wien.....	19
2.5.3	Anforderungen an Studierende	19
2.5.4	Anforderungen an Lehrende.....	21
2.6	Zusammenfassung der Einflüsse und Entwicklung der Arbeitshypothese	22
3	Methodische Vorgehensweise	23
3.1	Stichprobenbeschreibung	24
3.1.1	Ursprünglich gelieferte Stichprobe	24
3.1.2	Stichprobeneinschränkung	26
3.1.3	Datennachlieferung	29
3.2	Forschungsmethode	30
3.2.1	Binäre logistische Regression	30
3.2.2	Multinomiale logistische Regression	32
3.2.3	Ordinales logistisches Modell	33
3.2.4	Generalisiertes ordinales logistisches Modell.....	35

3.2.5	Prüfung des Modells	36
3.3	Analyse	38
3.3.1	Auswahl der unabhängigen Variablen im Modell	39
3.3.2	Binäre logistische Regression	42
3.3.3	Ordinale logistische Regression	46
3.3.4	Multinomiale logistische Regression	49
4	Empirische Ergebnisse und Diskussion	50
4.1	Resultate der Voranalyse	51
4.2	Resultate der binären logistischen Regression für Bachelorstudierende	55
4.3	Resultate der binären logistischen Regression für Masterstudierende	57
4.4	Resultate der ordinalen logistischen Regression für Bachelorstudierende	59
4.5	Resultate der multinomialen logistischen Regression für Bachelorstudierende	66
5	Diskussion der Resultate.....	69
5.1	Conclusio	69
5.2	Limitationen und weitere Forschungsmöglichkeiten	70
	Literaturverzeichnis	72
	Appendix	81
	Abstract.....	81
	Binäre Analyse.....	81
	Ordinale Analyse	87
	Multinomiale Analyse	93

Abbildungsverzeichnis

Abbildung 1: Altersverteilung in der Stichprobe	24
Abbildung 2: Durchschnittliches Alter pro Jahr, gesamte Stichprobe	24
Abbildung 3: Durchschnittliches Alter pro Jahr, eingeschränkte Stichprobe	28
Abbildung 4: Korrelationsmatrix potenziell unabhängiger Variablen	41
Abbildung 5: Durchschnittliche Erfolgswahrscheinlichkeit in den Alterskategorien, Bachelorstudium	43
Abbildung 6: Durchschnittliche Erfolgswahrscheinlichkeit in den Alterskategorien, Masterstudium	44
Abbildung 7: Durchschnittliche Note pro Alterskategorie, Masterstudium	47
Abbildung 8: Durchschnittliche Note pro Alterskategorie, Bachelorstudium	47
Abbildung 9: Residuen des einfachen binären logistischen Modells	52
Abbildung 10: Residuen des binären logistischen Modells unter Berücksichtigung von Nichtlinearität	53
Abbildung 11: Vorhergesagte Werte Modell (1) und Modell (2)	54
Abbildung 12: Vorhergesagte Werte des binären Modells für Bachelorstudierende	56
Abbildung 13: Residuen des Modells für Masterstudierende	59

Tabellenverzeichnis

Tabelle 1: Datenfelder aus der Datenlieferung	23
Tabelle 2: Darstellung Prüfungsverteilung über Semester und Jahre	25
Tabelle 3: Notenverteilung pro Semester	26
Tabelle 4: Darstellung Prüfungsverteilung über Semester und Jahre, eingeschränkte Stichprobe	27
Tabelle 5: Notenverteilung pro Semester, eingeschränkte Stichprobe	28
Tabelle 6: Manuell erstellte Datenfelder	38
Tabelle 7: Ergebnisse χ^2 -Tests	40
Tabelle 8: Übersicht Datensätze binäre Modelle	42
Tabelle 9: Übersicht Datensätze ordinale Modelle	46
Tabelle 10: Resultate der binären logistischen Regression	51
Tabelle 11: Resultate der binären logistischen Regression für Bachelorstudierende für Daten bis SS 20 sowie bis WS 20/21	55
Tabelle 12: Resultate der binären logistischen Regression für Masterstudierende für Daten bis SS 20 sowie bis WS 20/21	57
Tabelle 13: Resultate der ordinalen logistischen Regression für Bachelorstudierende	60

Tabelle 14: Test der Modelle auf proportionale OR in den Kategorien	62
Tabelle 15: Resultate der generalisierten ordinalen logistischen Regression für Bachelorstudierende.....	63
Tabelle 16: Resultate der multinomialen logistischen Regression für Bachelorstudierende..	67

Abkürzungsverzeichnis

OR	Odds Ratio
RRR	Relative Risk Ratio
AIC	Akaike'sches Informationskriterium
BIC	Bayes'sches Informationskriterium

1 Einleitung

Mit der Einführung von umfangreichen Schutzmaßnahmen nach Auftreten der ersten COVID-19-Fälle in Österreich im März 2020 veränderte sich der Universitätsbetrieb maßgeblich. Im Sommersemester 2020 wurden Lehrveranstaltungen des allgemeinen Lehrbetriebs auf virtuelle Vorlesungen, Übungen und auch virtuellen Prüfungsbetrieb umgelegt (Rektorat Universität Wien 2020a). Während das Wintersemester 2020/21 im hybriden Modus geplant war und auch so begann (Rektorat Universität Wien 2020b), wurden die Lehrveranstaltungen aufgrund steigender Infektionszahlen ab 3. November 2020 wieder soweit möglich auf Onlinelehre umgestellt (Rektorat Universität Wien 2020c).

Solche Änderungen bedeuten für alle Betroffenen der Bildungseinrichtung Umstellungen – sowohl Lehrende als auch Studierende waren gleichermaßen gefordert, sich auf diese Rahmenbedingungen einzustellen, um einen möglichst erfolgreichen Ablauf der Onlinelehre zu gewährleisten. Verschiedenste Bereiche des Lernumfelds von Studierenden wurden durch die Pandemie beeinflusst – soziale Interaktionen waren physisch nicht oder nur eingeschränkt möglich (anfangs verordnet aufgrund § 2 COVID-19-Maßnahmengesetz BGBl I 12/2020 StF mit VO des Gesundheitsministers BGBl II 98/2020), unter jüngeren Personen kam es zu einem starken Anstieg an psychischen Belastungen (Braun et al. 2020) und eine rein virtuelle Lernumgebung erfordert ein hohes Maß an intrinsischer Motivation und selbstständiges Lernen von Studierenden (Hao 2016, Yilmaz 2017). Gleichzeitig erhöht die Onlinelehre allerdings auch die Flexibilität des Studiums, ein Umstand, der gerade berufstätigen Studierenden entgegenkommt (Hall 2010). Die vorliegende Arbeit untersucht, wie sich akademische Studierendenleistungen in Semestern, welche von der Corona-Pandemie beeinflusst wurden, im Vergleich zu vorherigen Semestern verändert haben. Die Forschungsfrage dieser Arbeit lautet daher wie folgt:

Wie haben sich die akademischen Studierendenleistungen in Semestern, welche von der Corona-Pandemie betroffen waren (Sommersemester 2020 sowie Wintersemester 2020/21), im Vergleich zu vorhergehenden Semestern verändert?

Um diese Frage zu beantworten, wurden von der Abteilung für Datenmanagement der Universität Wien Daten betreffend Prüfungsergebnisse von Studierenden, beginnend im Sommersemester 2013, zur Verfügung gestellt. Diese Daten umfassen alle Prüfungen, bei welchen Frau ao. Univ.-Prof. Mag. Dr. Schaffhauser-Linzatti als Prüferin beteiligt war und ergänzen die Prüfungsergebnisse um weitere Indikatoren (wie beispielsweise Alter und Geschlecht der Studierenden). Die vorhandene Datenmenge wird in einem ersten Schritt gesichtet und auf eine Un-

termenge eingeschränkt, um eine etwaige Beeinflussung der statistischen Modelle zu vermeiden. Außerdem wird eine Indikatorvariable eingeführt, welche angibt, ob ein Semester durch die Corona-Pandemie beeinflusst wurde (Sommersemester 2020 sowie Wintersemester 2020/21) oder nicht. Anschließend werden die Daten in zwei unterschiedlichen Herangehensweisen analysiert: In einem ersten Schritt wird die Erfolgswahrscheinlichkeit betrachtet, eine Prüfung zu bestehen. Dazu wird eine binäre abhängige Variable Erfolg definiert, welche misst, ob eine Prüfung bestanden wurde oder nicht. Anschließend wird eine logistische Regression durchgeführt, um signifikante Einflüsse auf die Wahrscheinlichkeit, eine Prüfung zu bestehen, festzustellen. In der zweiten Analyse werden direkt die Noten von Studierenden in einer Prüfung betrachtet (von 1 = *Sehr Gut* bis 5 = *Nicht Genügend*). Da diese ordinal skaliert sind, wird zuerst eine ordinale logistische Regression sowie nach Prüfung der Voraussetzungen eine generalisierte ordinale logistische Regression durchgeführt. Dabei wird die Wahrscheinlichkeit betrachtet, in unterschiedliche Notenkategorien zu fallen (eine höhere Kategorie entspricht schlechteren Noten).

Der Aufbau der Arbeit ist wie folgt: Zuerst erfolgt eine theoretische Ausarbeitung der Einflüsse auf das Lernumfeld von Studierenden in Kapitel 2. Aufgrund der Aktualität der Arbeit wird versucht, bereits bestehende Studien betreffend die Auswirkungen der Corona-Pandemie auf Studierende und die Bevölkerung im Allgemeinen miteinzubeziehen. Die Veränderungen der Onlinelehre werden dabei in den Kontext des Lehr- und Lernumfelds gestellt und Voraussetzungen für erfolgreiche Onlinelehre (siehe etwa Nebel 2017, Hao 2016 oder Yilmaz 2017) besprochen. Kapitel 0 beschreibt die vorhandenen Daten deskriptiv sowie die gewählte Vorgehensweise und die theoretischen Konzepte hinter der binären logistischen Regression, der ordinalen logistischen Regression, der generalisierten ordinalen logistischen Regression sowie der multinomialen logistischen Regression. Die Ergebnisse der unterschiedlichen statistischen Modelle werden in Kapitel 4 besprochen und miteinander verglichen – dadurch soll auch das statistisch geeignetste Modell ausgewählt werden. Die Diskussion der Ergebnisse sowie die Conclusio sind anschließend in Kapitel 5 zu finden, in welchem auch nochmals auf die Limitationen und weiteren Forschungsmöglichkeiten in diesem Gebiet hingewiesen wird.

2 Theoretischer Hintergrund

2.1 Lehr- und Lernerfolg

2.1.1 Konzeptdefinition

Lern- und Lehrerfolg kann durch unterschiedliche Konzepte und Herangehensweisen definiert werden. Hattie (2015) hält als Problematik fest, dass auch an Universitäten, ähnlich wie in schulischer Bildung (Hattie 2009), ein oberflächliches Wissen als wichtiger angesehen wird als

sogenanntes *deep learning* (Hattie 2015, S. 80). Die nachfolgende Aufzählung berücksichtigt daher, dass nicht nur fachbezogene Kompetenzen eine wichtige Rolle spielen, sondern genauso eine persönliche Weiterentwicklung der Studierenden als Lernziel angesehen wird (Hattie 2015). Auch Wuttke (2000) fordert, dass im Rahmen der akademischen Lehre nicht nur die wissensbasierten Anforderungen erfüllt, sondern auch die individuelle Persönlichkeitsbildung forciert werden sollen. Wie individuell das Verständnis von Studienerfolg sein kann, zeigt Gold (1988) auf, der den Studienerfolg als Zufriedenheit der Studierenden mit der eigenen Leistung definiert.

a) Noten von Studierenden

Jirjahn (2007) führt an, dass die Noten von Studierenden zu einer Differenzierung am Arbeitsmarkt beitragen können, da diese einen wichtigen Hinweis auf die Befähigung von Studierenden bieten. Helm (2017) nutzt unter anderem die Noten von Studierenden, um den Erfolg eines Lehrkonzepts zu überprüfen. Klenke (2018) spricht allerdings auch offen an, dass die Noten von Studierenden nicht immer mit Erfolg gleichzusetzen sind – so kann eine schlechte positive Note mit wenig Aufwand für jemanden genauso ein Erfolgserlebnis darstellen wie eine sehr gute Note für jemand anderen. Die Abschlussnote eines Studiums hat jedoch auch im Rekrutierungsprozess nach Abschluss des Studiums einen wichtigen Einfluss: Da die Personalauswahl von Unternehmen bei Berufsanfängern oftmals von Unsicherheit geprägt ist, kommt der Signalwirkung verschiedenster Kriterien besondere Bedeutung zu (Neeß 2015). Neeß (2015) konnte hier zeigen, dass die Abschlussnote eines Studiums von Unternehmen in die Personalauswahl miteinbezogen wird, jedoch in ungefähr gleichem Ausmaß wie weitere Kriterien (zum Beispiel Praxis- und Auslandserfahrung).

b) Studiendauer und Studienabschluss

In Bezug auf die Differenzierung der Studierenden am Arbeitsmarkt werden von Jirjahn (2007) nicht nur die Noten als Signal genannt, sondern auch die Studiendauer. Oftmals wird bereits der Studienabschluss selbst als Lehrerfolg gezählt, ohne auf die Dauer des Studiums explizit einzugehen. So haben etwa Mitterauer et al. (2012) als erfolgreiche Studierende all jene Personen klassifiziert, die innerhalb von acht Jahren nach Studienbeginn das Medizinstudium (Regelstudiendauer von sechs Jahren) abgeschlossen hatten. Müller und Braun (2018) nennen Studienabbruchquoten (wie viele Studierende also ihr Studium abgeschlossen und wie viele es abgebrochen haben) als Indikator für das Qualitätsmanagement an Hochschulen.

c) Kompetenz erreichen

Kompetenz beschreibt nicht nur fachliches Wissen – welches sich unter anderem auch über Noten messen lässt –, sondern ebenso die Fähigkeit, neue Informationen zu analysieren und zu verstehen (Chickering 1969, zitiert nach Hattie 2015). Auch interpersonelle Kompetenzen, oftmals als Teamwork zusammengefasst, sind wichtige Fähigkeiten, welche sich Studierende im Laufe des Studiums aneignen sollen (Chickering 1969, zitiert nach Hattie 2015).

d) Emotionen handhaben und von Autonomie zu Unabhängigkeit entwickeln

Emotionen wie etwa Angst und Ärger erschweren das Lernen (Hattie 2015). Hofmann (2010) hält fest, dass Ängste das Lernen und den Studienerfolg negativ beeinflussen können. Nach Hattie (2015) können positive Emotionen wie etwa Optimismus und Hoffnung das Lernen unterstützen. Diese Emotionen handzuhaben und auszugleichen, ist laut Chickering (1969, zitiert nach Hattie 2015) ebenso ein wichtiger Bestandteil eines erfolgreichen Studienabschlusses. Die Entwicklung von Autonomie zu Unabhängigkeit wird explizit von Chickering (1969, zitiert nach Hattie 2015) angeführt und von Hattie (2015) aufgegriffen, der betont, dass dies auch inkludiert, dass Studierende ein Selbstwertgefühl erreichen und sich selbst genügen. Dies meint auch, dass Zusicherung von anderen Personen nicht mehr benötigt wird (Hattie 2015).

e) Reife Beziehungen zu anderen Personen aufbauen

Wie bereits unter Punkt c) mit interpersonellen Kompetenzen angesprochen, hält Chickering die Beziehungen zu anderen Personen nochmals explizit als wichtigen Lehr- und Lernerfolg im Rahmen eines Studiums fest (Chickering 1969, zitiert nach Hattie 2015). Dazu gehört, dass Unterschiede akzeptiert werden und man mit anderen Gruppenmitgliedern erfolgreich zusammenarbeiten kann (Hattie 2015). Nach Pfaff und Huddleston (2003) wird dieser Aspekt auch in der Geschäftswelt immer wichtiger, sodass das Erlernen von erfolgreicher Gruppenarbeit an Universitäten einer zeitgemäßen Ausbildung entspricht.

f) Probleme lösen und Entscheidungen zu treffen

Mayer (1998) greift den Aspekt des Problemlösens auf und hält fest, dass viele Studierende zwar bekannte Probleme lösen können, jedoch Schwierigkeiten dabei haben, die gelernten Konzepte auf neue Problemstellungen anzuwenden. Auch Hattie (2015) erwähnt das Problemlösen explizit als Lehr- und Lernerfolg, welcher bereits von Chickering (1969, zitiert nach Hattie 2015) definiert wurde. Für das Erlernen dieser Fähigkeit identifizierte Mayer (1998) den

Willen der Studierenden, also die Motivation, etwas zu lernen und Arbeit in das Lernen zu investieren, als wichtigen Indikator.

g) Eigene Identität und Integrität entwickeln

Akademische Integrität ist nach Hanbidge et al. (2017) wie folgt definiert: „a set of conventions that scholars follow in their work, and which generates credibility, trust, and respect within the academic community“ (Hanbidge et al. 2017, S. 135). Nicht nur Hanbidge et al. (2017), auch bereits Chickering (1969, zitiert nach Hattie 2015) und Hattie (2015) nennen die Entwicklung von akademischer Integrität als wichtigen Lehrerfolg im Rahmen eines akademischen Studiums. Das Selbstbewusstsein und der Glaube an die eigenen Fähigkeiten stellt einen wichtigen Entwicklungsschritt für Studierende und somit einen Lernerfolg dar (Chickering 1969, zitiert nach Hattie 2015).

h) Berufliche Situation von Absolventen eines Hochschulstudiums

Im Rahmen von Absolventenbefragungen werden vor allem quantitativ messbare Konzepte von Absolventen im Rahmen des Hochschulmanagements abgefragt – diese umfassen etwa das Einstiegsgehalt nach Studienabschluss sowie die Dauer, bis eine Stelle gefunden wurde (Welpé et al. 2020). Nach Flöther et al. (2015) ist das Interesse an dem beruflichen Werdegang von Absolventen für Hochschulen im Rahmen von Diskussionen über zukünftige Entwicklungen sehr hoch. Die Informationen dazu werden von Hochschulen unter anderem für Qualitätssicherung, Hochschulentwicklungsplanung und auch Evaluation oder die Qualifizierung von Wissenschaftlern benötigt und können daher ein wichtiges Instrument zur Messung von Lehrerfolg darstellen (Flöther et al. 2015). Solche Absolventenstudien werden nicht nur von vielen Universitäten eingesetzt, sondern auch national und international durchgeführt – wie etwa die Studie CHEERS (Careers after Higher Education: Graduation – an European Research Study), welche 1999 durchgeführt wurde und Informationen zum jeweiligen Studium der Absolventen, aber auch der beruflichen Situation, wie etwa Einkommen und Stellung, Tätigkeiten etc. abfragte (Welpé et al. 2020).

Wie anhand der oben angeführten Punkte a) bis h) ersichtlich ist, wird Lehr- und Lernerfolg sehr unterschiedlich definiert. Im gegebenen Kontext bezieht sich das Konzept der (akademischen) Studierendenleistungen – auch in Hinblick auf die verfügbaren Daten – jedoch immer auf die erreichten Noten von Studierenden in den untersuchten Prüfungen.

2.1.2 Einflüsse auf Lehr- und Lernerfolg

Erfolg (beziehungsweise Leistung) kann vereinfachend als eine Funktion der Fähigkeiten eines jeden Studenten sowie der Motivation zu und Anzahl an studienrelevanten Tätigkeiten

dargestellt werden (Rizzuto et al. 2009). Zusätzlich beeinflussen verschiedenste Kontextvariablen den Lehr- und Lernerfolg (Rizzuto et al. 2009). Untersucht man die Einflüsse darauf, lassen sich grundsätzlich zwei verschiedene Arten der Untersuchungen feststellen: Müller und Braun (2018) unterscheiden zwischen den Konzepten der „Studierendenzentrierung“ und der „Lehrendenzentrierung“. Die Lehrendenzentrierung legt den Fokus dabei auf den Einfluss der Lehre, etwa der Darbietung des Lehrmaterials und des Lehrstils auf den Studienerfolg. Die Studierendenzentrierung hingegen betrachtet Einflussfaktoren vonseiten der Studenten und deren Tätigkeiten (Müller und Braun 2018).

Einflüsse aus dem Konzept der Lehrendenzentrierung

Nach Hattie (2015) sollten Lehrende an Hochschulen vor allem vier Eckpunkte einer Strategie berücksichtigen: (1) Die Unterrichtseinheiten strategisch zu planen und selbst zu bewerten, um Feedback in weitere Einheiten einfließen lassen zu können, (2) bewusst Strategien zur Erhöhung des Lehrendeneinflusses einzubauen, (3) den Studierenden eine Stimme zu geben, um ihre Sicht der Lehre zu erfahren, und (4) Daten über den Einfluss zu sammeln und zu analysieren (Hattie 2015). Prosser und Trigwell (2006) halten fest, dass der Lernerfolg von Studierenden auch von der Beziehung dieser zu dem Lehr- und Lernkontext abhängt. Auch Braun und Hannover (2008) können feststellen, dass die Lehrorientierung von Lehrenden an Hochschulen einen Einfluss auf den Kompetenzgewinn der Studierenden hat. Die studierendenfokussierte Lehrorientierung, welche sich vor allem durch aktivierende und sozial-interaktive Lehrmethoden auszeichnet, weist dabei einen höheren positiven Einfluss auf den Kompetenzgewinn der Studierenden auf als die vergleichbare lehrendenfokussierte Lehrorientierung, welche sich durch häufigeren Vortrag der Lehrenden auszeichnet (Braun und Hannover 2008).

Einflüsse aus dem Konzept der Studierendenzentrierung

Wigfield und Eccles (2000) halten in ihrem Erwartung-mal-Wert-Modell die Motivation und die Erwartungen von Studierenden an eigene Leistungen für einen der wichtigsten Einflussfaktoren auf akademische Leistungen. Müller und Braun (2018) heben hingegen hervor, dass vor allem die Aktivität eines Studierenden, also wie viel Zeit und Bemühungen dieser in studienbezogene Tätigkeiten investiert, für den Studienerfolg ausschlaggebend ist. Winteler und Forster (2008) betonen, dass diese Aktivität als Einflussfaktor sogar wichtiger als die Herkunft der Studierenden oder die Universität selbst ist. Eine ähnliche Feststellung trifft Hattie (2009), der auf Basis der Studie von O'Connor und Paunonen (2007) zwischen dem Einfluss von kognitiven Fähigkeiten und Persönlichkeitsmerkmalen unterscheidet. Während nämlich kognitive Fähigkeiten bestimmen, was eine Person leisten kann, setzen Persönlichkeitsmerkmale fest, was diese Person voraussichtlich tun wird (Hattie 2009) – dies betont also zusätzlich zu den

kognitiven Voraussetzungen die Wichtigkeit der Aktivität eines Lernenden. Auch Chamorro-Premuzic und Furnham (2008) halten fest, dass kognitive Fähigkeiten (gemessen durch IQ-Tests) nur etwa 50% der Varianz in akademischen Leistungen erklären können und es daher weitere Faktoren gibt, die diese Leistung beeinflussen. Ähnlich geben Furnham et al. (2003) an, dass die Persönlichkeitsindikatoren akademische Leistungen besser vorhersagen als die Intelligenz von Studierenden. Ergänzend zu den genannten Faktoren konnte Jirjahn (2007) herausfinden, dass sich ein höheres verfügbares Einkommen, ein fester Partner, das Alter von Studierenden, gutes Zeitmanagement und auch eine erhöhte Freizeitaktivität positiv auf den Studienerfolg auswirken. Nach Hattie (2015) werden rund 50% der Varianz des Lernerfolgs von solchen Faktoren, die von den Bildungseinrichtungen nicht beeinflussbar sind, bestimmt.

Die COVID-19-Pandemie beeinflusste die Lehr- und Lernumgebung auf vielfältige Weise: So führte sie etwa zu einer starken physischen Distanz. Sowohl die Anwesenheit an der Universität als auch sonstige Treffen mit haushaltsfremden Personen waren erschwert oder sogar unmöglich (anfangs verordnet aufgrund § 2 COVID-19-Maßnahmengesetz BGBl I 12/2020 StF mit VO des Gesundheitsministers BGBl II 98/2020). Dies beeinflusst verschiedenste Kontextvariablen, welche wiederum den akademischen Erfolg von Studierenden bestimmen: Die soziale Eingebundenheit von Studierenden wurde etwa durch die räumliche Distanz und Online-Tools als Kommunikationsmittel beeinflusst (Schober et al. 2020). Weiters ist eine globale Pandemie kein alltägliches Ereignis und verändert damit psychische Belastungen von Studierenden (Hasan und Bao 2020). Das Zeitmanagement wurde nicht von der Pandemie selbst beeinflusst, sehr wohl jedoch die Art der Lehre, da von Präsenz- auf Onlinelehre umgestellt werden musste. Da sich die Voraussetzungen für erfolgreiche Lehre in der Online- und der Präsenzlehre unterscheiden, wird auch dieser Einflussfaktor in der vorliegenden Arbeit untersucht (Nebel 2017, Yilmaz 2017).

2.2 Soziale Eingebundenheit und die Bedeutung von Netzwerken

2.2.1 Soziale Eingebundenheit durch Peer-Gruppen

Bereits 1990 konnten Wang et al. (1990) die Lernumgebung als wichtigen Einflussfaktor auf den Lernerfolg identifizieren, wobei Peer-Gruppen in der Lernumgebung die einflussreichste Variable darstellen. Sallee und Tierney (2007) untersuchen den Einfluss von Peer-Gruppen auf akademische Leistungen. Dabei werden Peer-Gruppen klar von anderen sozialen Netzwerken unterschieden: Während bei sozialen Netzwerken jegliche soziale Beziehungen von Personen umfasst sein können, bestehen Peer-Gruppen aus Personen mit demselben hierarchischen Status und ähnlichen Interessen. Dies bedeutet, dass alle Teilnehmer auf derselben Ebene miteinander kommunizieren – im Gegensatz zu etwa Eltern-Kind-Beziehungen. Peer-Gruppen sind also *„any set of same-age peers linked by a common interest or identity who*

engage in sustained interaction“ (Sallee und Tierney 2007, S. 7). Im Rahmen von Peer-Learning an der Universität bezeichnet der Begriff „Peer“ ähnlich *„students [who] learn with and from each other without the immediate intervention of a teacher“* (Boud et al. 1999, S. 413). Peer-Gruppen können also eine Gruppe im Rahmen der Universität oder eines Kurses sein, genauso allerdings Personen ohne Lernkontext umfassen, zum Beispiel eine Gruppe von Freunden aus der Nachbarschaft (Wang et al. 1990). Basierend auf Bordieus (1986) und Coleman (1990) Theorien von sozialem Kapital führen Sallee und Tierney (2007) aus, dass Mitglieder von Peer-Gruppen akademische Leistungen herabwürdigen und damit negativ beeinflussen können, genauso können sie aber auch zu solchen ermuntern. Um Vorteile von Peer-Gruppen zu nutzen, wurden verschiedenste Lehrkonzepte damit umgesetzt – durch Gruppenprojekte, Studiengruppen, Peer-Feedback und ähnliche kollaborative Lehrmethoden wird der Peer-Effekt innerhalb eines Kurses beziehungsweise innerhalb der Universität genutzt (Boud et al. 1999). Helm (2017) betont jedoch, dass kollaboratives Lernen per se noch nicht zwingend zu höheren Lernerfolgen führt, da auch negative Zusammenhänge zwischen kollaborativen Methoden und dem Lernerfolg bestehen. Bei kooperativem oder kollaborativem Lernen muss nämlich auch berücksichtigt werden, dass heterogene Gruppenmitglieder (zum Beispiel gute und schlechte Studierende) nicht nur voneinander profitieren: Die besseren Studierenden können dominieren und zu passivem Verhalten bei schwächeren Mitgliedern der Gruppe führen (Krause et al. 2009). Je mehr Mitglieder sich jedoch einbringen, desto größer kann der Erfolg der gesamten Gruppe sein, da alle Ressourcen genutzt werden (Krause et al. 2009). Gleichzeitig mussten Krause et al. (2009) bei von ihnen untersuchten homogenen Gruppen von Studierenden feststellen, dass sich die Gruppenleistung in der Lernphase durch kooperatives Lernen zwar verbessert, das Lernen der einzelnen Gruppenmitglieder davon jedoch nicht (weder positiv noch negativ) beeinflusst wurde.

2.2.2 Soziale Eingebundenheit durch Netzwerke

Im Widerspruch zu Krause et al. (2009) stellen Wang et al. (1990) und ähnlich auch Helm (2017) fest, dass die kognitive und soziale Entwicklung von Personen nicht nur vom eigenen Verhalten und den genetischen Voraussetzungen beeinflusst wird, sondern genauso von der Interaktion mit der sozialen Umgebung. Dies entspricht auch der Social Learning Theory von Bandura (1977), wonach durch erfolgreiche Gruppenarbeiten individuelle Lernziele leichter erreicht werden. Die soziale Umgebung kann dabei auch als soziales Netzwerk gesehen werden, welches als *„one’s relationships with others“* (Carolan 2013, S. 4) beschrieben wird und damit deutlich weiter definiert ist als eine Peer-Gruppe. Während nämlich in einer Peer-Gruppe nach der Definition von Sallee und Tierney (2007) ähnliche Ansichten und Identitäten sowie eine wiederkehrende Interaktion vorhanden sind, muss dies in einem sozialen Netzwerk nicht

zwingend eintreten. Ähnlich wie bei Peer-Gruppen können Einflüsse auf den Lernerfolg festgestellt werden: Nach Helm (2017) kann das soziale Netzwerk sogar mehr Einfluss als individuelle Charakteristika einer Person haben; diese Charakteristika beeinflussen aber oftmals wiederum das eigene soziale Netzwerk. Soziale Netzwerke innerhalb eines Kurses bilden für Studierende soziales Kapital, mit dessen Hilfe ein Kurs leichter bewältigbar ist und durch welches Studierende selbst resilienter gegenüber negativen Einflüssen sind (Rizzuto et al. 2009). Dies wurde auch von Pascarella und Terenzini (1991, S. 620) festgehalten: „[...] *part of the impact of college is determined by the extent and content of one's interactions with [...] faculty members and student peers*“. Im Unterschied zu reinen Peer-Gruppen finden die Interaktionen hier also auch zwischen Studierenden und Lehrenden sowie zwischen Studierenden untereinander statt. Diese Interaktionen führen nicht nur zur höheren Beständigkeit von Studierenden in einem Studium, sondern auch zu verbesserten Studienleistungen (Beattie und Thiele 2016). Beattie und Thiele (2016) erweitern daher das Konzept von sozialem Kapital um die Facette „akademisches soziales Kapital“, welches sich aus Beziehungen von Studierenden zu Fakultätsmitgliedern und auch untereinander zusammensetzt. Bereits eine einzelne Beziehung zu einem Fakultätsmitglied oder zu anderen Studierenden kann einen positiven Einfluss auf den akademischen Erfolg haben (Chambliss und Takacs 2014). Diese sozialen Beziehungen bestimmen auch, wen man um Rat fragen würde, und können somit einen großen Einfluss auf die weitere Entwicklung und Entscheidungen (Kurswahl, etc.) haben (Chambliss und Takacs 2014). Rizzuto et al. (2009) führen dabei gleich drei hauptsächliche Faktoren an, durch welche soziale Netzwerke innerhalb von Kursen beziehungsweise an der Universität selbst Studierenden dabei helfen, akademische Ziele leichter zu erreichen: (1) Der Zugang zu Information (einerseits das Wissen, andererseits allerdings auch die Fähigkeiten von Kollegen) kann es Studierenden erleichtern, Aufgaben zu erfüllen. (2) Eine gegenseitige Erwartungshaltung innerhalb des Netzwerkes bestimmt Verpflichtungen und Vertrauen, wie etwa das Mitschreiben oder Ausleihen von Notizen für einen abwesenden Kollegen (Rizzuto et al. 2009). (3) Soziale Normen bestimmen, was innerhalb des Netzwerkes erwünscht und erwartet wird – wie etwa die oben genannte Mitschrift. Die Autoren konnten auch empirisch nachweisen, dass die Tiefe eines sozialen Netzwerkes die akademischen Leistungen von Studierenden positiv beeinflusst (Rizzuto et al. 2009). Während Beattie und Thiele (2016) die Klassengröße von Kursen untersuchen und feststellen, dass zu große Kurse (mit zu vielen Studenten) zu einer geringeren Interaktion von Studierenden untereinander und von Studierenden zu Lehrenden führt, hat Hiltz (1994) auch für e-Learning-Kurse eine maximale Teilnehmerzahl von 30 Personen vorgeschlagen, um die sozialen Interaktionen noch gewährleisten zu können. Nun muss festgehalten werden, dass sich die Technik zur Nutzung von e-Learning-Materialien seit 1994 bedeutend verbessert hat – so konnten etwa Alqudah et al. (2020) in einer Befragung von Lehrenden an medizinischen Universitäten in Jordanien zur e-Learning-Situation im Rahmen von

COVID-19 feststellen, dass die Interaktion mit den Studierenden von den Lehrenden als positiv empfunden wurde. Auch Lu et al. (2013) halten fest, dass die Bildung von sozialem Kapital ebenso virtuell möglich ist, da e-Learning-Methoden Möglichkeiten zur sozialen Interaktion bieten und viele Studierende sich in Online-Foren und Chatgruppen austauschen. Die Autoren untersuchen, ob Studierende auch bei Online-Learning von sozialem Kapital profitieren, ob also Studierendenleistungen, welche in der Form des e-Learnings erbracht werden, davon beeinflusst werden (Lu et al. 2013). Sie finden dabei auch im Kontext des e-Learning soziales Kapital bei Gruppen von Studierenden, welches vor allem die Aspekte Kooperation, Kollaboration und Inklusion beziehungsweise Soziabilität von Studierenden beeinflusst (Lu et al. 2013). Dies lässt sich im Rahmen der Studie jedoch auch auf die Art der Gruppenarbeit in der Studie, welche verpflichtend war und ein hohes Ausmaß an Zusammenarbeit der Mitglieder voraussetzte, zurückführen. Im Gegensatz zu Krause et al. (2009), welche im Zusammenhang mit Peer-Gruppen und Gruppenarbeiten keinen Einfluss von sozialem Kapital auf die Lernleistung des einzelnen Studierenden fanden, stellen Lu et al. (2013) fest, dass soziales Kapital die Noten auf die Gruppenarbeiten und ebenso die Gesamtnote der Studierenden in einem Kurs – wenn auch weniger stark als bei der Gruppenarbeit – positiv beeinflusst. Dies geht konform mit der Netzwerktheorie, welche besagt, dass dichtere Netzwerke (etwa enge Zusammenarbeit bei Gruppenarbeiten) zu stärkeren Effekten durch soziales Kapital führen können (Lu et al. 2013).

Zur Stärkung von sozialem Kapital durch Gruppenarbeiten in der Onlinelehre muss angemerkt werden, dass Gruppenarbeiten auch in der Präsenzlehre stark vertreten sind (Boud et al. 1999) und somit die Bildung von sozialem Kapital sowohl in einem e-Learning- als auch im Präsenzkontext gefördert wird. Die durch COVID-19 bedingte Umstellung auf reine Onlinelehre erschwert aber jedenfalls die Kontaktaufnahme von Studierenden vor oder nach einem Kurs im Rahmen von informellen Gesprächen. Dies betrifft auch kurze Frageeinheiten vor oder nach Präsenzeinheiten zwischen Studierenden und Lehrenden. In einer österreichweiten Befragung von Studierenden wurde die Bedeutung von sozialen Kontakten im Rahmen eines Studiums betont: Motivationsprobleme von vielen Studierenden wurden auf fehlenden fachlichen Austausch sowohl von Studierenden untereinander als auch zwischen Studierenden und Lehrenden zurückgeführt (Schober et al. 2020). Aufgrund der Wichtigkeit von sozialen Netzwerken, sozialem Kapital, dem klar positiven Einfluss von kollaborativen Lehrmethoden sowie der Vertiefung von sozialen Netzwerken auf akademische Studierendenleistungen wird Folgendes festgehalten:

„Es ist davon auszugehen, dass die physische Absenz von Studierenden und Lehrenden im reinen e-Learning-Kontext zu einer geringeren Dichte der sozialen Netzwerke führt und Studierendenleistungen tendenziell negativ beeinflusst werden.“

2.3 Psychische Belastungen

Hofmann (2010) hält fest, dass psychische Belastungen, etwa in Form von Ängsten und Stresszuständen, die akademischen Leistungen von Studierenden negativ beeinflussen. Luthans et al. (2012) untersuchen das gegenteilige Phänomen, nämlich sogenanntes „psychologisches Kapital“ (welches aus den Ressourcen Hoffnung, Wirksamkeit, Resilienz und Optimismus besteht) und können dieses positiv mit akademischen Leistungen von Studierenden verbinden.

a) Allgemeine psychische Belastungen von Studierenden

Wie bereits in Kapitel 2.1 besprochen, stellt ein Studium an Studierende nicht nur die Anforderung des Erlernens von Wissen, sondern soll ebenso zur persönlichen Weiterentwicklung beitragen (siehe etwa Hattie 2015; Wuttke 2000). Beratungsstellen an Universitäten werden dabei oftmals von Studierenden genutzt, welche diese Weiterentwicklung gerade zu Beginn des Studiums als Überforderung wahrnehmen (Hofmann et al. 2017). Gründe für psychische Belastungen können persönliche, soziale und finanzielle Probleme sein, wobei die akademischen Folgen dieser bis zum Studienabbruch führen können (Hofmann et al. 2017). Psychische Belastungen oder Störungen waren bereits vor Eintritt der COVID-19-Pandemie unter Studierenden verbreitet – zwischen 23% und 60% der Studierenden erfüllten Kriterien von verschiedensten psychischen Krankheiten, von depressiven Symptomen bis hin zu Angststörungen (Hofmann et al. 2017).

b) Psychische Belastungen in der COVID-19-Pandemie

Obwohl die COVID-19-Pandemie im ersten Halbjahr 2021 aller Voraussicht nach nur eingedämmt und nicht gestoppt werden kann, wurden bereits erste Studien zu den psychischen Einflüssen der Pandemie veröffentlicht. So konnte etwa bereits in einer Befragung im Zeitraum von März 2020 bis April 2020 festgestellt werden, dass rund 25% aller Befragten einer Studie auf den Philippinen unter mittelstarken bis starken Angstzuständen und rund 16% an mittelstarken bis starken Depressionen und psychischen Beeinträchtigungen litt (Tee et al. 2020). Dabei muss festgehalten werden, dass die Bevölkerung der Philippinen bereits zuvor mit Ausbrüchen gefährlicher Krankheiten konfrontiert wurde (Tee et al. 2020). Die breiter angelegte Meta-Analyse von Salari et al. (2020) betreffend Befragungen und Studien zu psychischen Belastungen der Bevölkerung in Asien und Europa während der Corona-Pandemie zeigt, dass

das Vorkommen von Stresssymptomen bei 29,6%, Angstzuständen bei 31,9% und Depressionen bei 33,7% der Bevölkerung von Asien und Europa liegt. Weiters konnte bereits vor Auftreten von COVID-19 festgestellt werden, dass Frauen von Depressionen häufiger betroffen sind als Männer (Lim et al. 2018) – eine Tatsache, welche sich auch in Studien betreffend COVID-19 widerspiegelt (Salari et al. 2020). Interessanterweise konnten Salari et al. (2020) auch herausfinden, dass die Altersgruppe von 21 bis 40 Jahren am stärksten von psychischen Belastungen wie Angst, Stress und Depressionen betroffen ist, obwohl die Sterblichkeit und Gefährlichkeit von COVID-19 mit dem Alter ansteigt. Die Gründe dafür sind laut Salari et al. (2020), dass sich diese Altersgruppe auch um die eigene Zukunft und zukünftige wirtschaftliche Belastungen sorgt. Außerdem sind jüngere Personen in sozialen Netzwerken aktiv und haben dadurch einen größeren Zugang zu Informationen, was zu einem höheren Stresslevel führen kann (Salari et al. 2020). Zu ähnlichen Ergebnissen kommen auch Hakami et al. (2020), welcher bei Studierenden der Zahnmedizin in Saudi-Arabien ein Level an Depressionen von 60,6%, Angstzuständen von 37,0% und Stress von 34,9% misst. Entsprechend den Resultaten von Lim et al. (2018) weisen Frauen eine höhere Wahrscheinlichkeit auf, bedingt durch COVID-19 unter psychischen Belastungen zu leiden (Hakami et al. 2020). Auch bei Studierenden in Bangladesch wurde psychischer Stress festgestellt und auf ineffektive e-Learning-Systeme sowie die Angst vor dem Verlust eines Studienjahres zurückgeführt (Hasan und Bao 2020). Husky et al. (2020) unterscheiden in ihrer Untersuchung von französischen Studierenden zwischen jenen Personen, welche aufgrund des ersten Lockdowns (von Husky et al., 2020, S. 2, auch *confinement period* genannt) umzogen (zumeist wieder zurück zu den Eltern) und jenen Personen, welche am Studienort blieben. Rund 60,2% aller befragten Studierenden geben an, dass das Angstlevel gestiegen sei – dabei sind jene Studierende, welche am Studienort verblieben waren, stärker davon betroffen (Husky et al. 2020). Auch das generelle Stresslevel – vor allem betreffend finanzielle Probleme und die eigene Gesundheit – ist für rund 61,6% der befragten Studierenden erhöht, wobei auch hier jene Personen, welche nicht kurzfristig umgezogen waren, stärker davon betroffen sind (Husky et al. 2020). In den Ergebnissen einer Befragungswelle in Österreich, durchgeführt vom Zentrum für Public Health der Medizinischen Universität Wien, lassen sich ebenso durchwegs höhere Depressions- und Angstlevel bei jüngeren Personen feststellen (Braun et al. 2020). Bei der jüngsten Bevölkerungsgruppe (16 bis 29 Jahre) wurde zwischen den ersten beiden Befragungswellen (23.04.-05.05.2020 und 15.05.-28.05.2020) ein Anstieg der Suizidalität und der berichteten psychischen Gewalt verzeichnet, wobei die Werte zumindest bis zur Befragungswelle im August 2020 erhöht geblieben sind (Braun et al. 2020).

c) Folgen von psychischen Belastungen

Ängste und depressive Symptome beeinflussen nicht nur das Privatleben von Studierenden, sondern auch ihre akademischen Leistungen negativ (Hofmann 2010). Dieser negative Einfluss auf akademische Leistungen und zusätzlich eine negative Beeinflussung der beruflichen Entwicklung sowie eine höhere Wahrscheinlichkeit für Plagiarismus wurde von Dyrbye et al. (2006) festgestellt. Dies entspricht auch den Resultaten von Luthans et al. (2012): Die Autoren untersuchen nicht den negativen Einfluss von psychischen Belastungen, sondern das gegenteilige Konzept des positiven Einflusses auf akademische Leistungen durch psychologisches Kapital (Luthans et al. 2012). Dieses Konzept wird dabei wie folgt definiert:

„an individual's positive psychological state of development characterized by: (1) having confidence (self-efficacy) to take on and put in the necessary effort to succeed at challenging tasks; (2) making a positive attribution (optimism) about succeeding now and in the future; (3) persevering toward the goals, and when necessary, redirecting paths to goals (hope) in order to succeed; and (4) when beset by problems and adversity, sustaining and bouncing back and even beyond (resilience) to attain success“ (Luthans et al. 2007, S. 3).

Durch verschiedenste psychische Belastungen werden unterschiedliche Konstrukte des psychologischen Kapitals negativ beeinflusst (Hoffnung, Wirksamkeit, Resilienz, Optimismus). Der positive Effekt, welcher durch psychologisches Kapital erreicht werden kann, wird dadurch gemindert. Hofmann et al. (2017) betonen die Wichtigkeit von psychologischen Beratungsstellen an Universitäten, um möglichst vielen jungen Menschen einen Studienabschluss zu ermöglichen. Solche Beratungen sollen die psychischen Belastungen und damit einhergehende erschwerte Studienbedingungen adressieren und Studienabbrüche vermeiden (Hofmann et al. 2017).

Aus den oben angeführten Verstärkungen von psychischen Belastungen und den Folgen dieser lässt sich aufgrund der psychischen Einflüsse der Corona-Pandemie folgende Erwartung festhalten:

„Durch die steigende psychische Belastung von Studierenden in Zeiten der COVID-19-Pandemie ist davon auszugehen, dass akademische Leistungen negativ beeinflusst werden.“

2.4 Zeitmanagement

Nach Schiefele und Wild (1994) lässt sich Zeitmanagement den ressourcenbasierten Lernstrategien zuordnen, zu welchen auch Anstrengung, Konzentration und Lernumgebung zählen. Gemeinsam mit anderen Prädiktoren erlaubt Zeitmanagement, bis zu 25% der Varianz des Lernerfolgs von Studierenden festzulegen (Boerner et al. 2005). Schiefele und Wild (1994)

beziehen Zeitmanagement auf die Planung und Kontrolle der Arbeitszeit, wobei Michinov et al. (2011) Prokrastination als einen wichtigen Einfluss auf Zeitmanagement feststellen. Lohbeck et al. (2017) definieren Prokrastination als „*Aufschieben von [...] wichtigen Aufgaben trotz der damit einhergehenden negativen Konsequenzen [...]*“ (Lohbeck et al. 2017, S. 522), was sich negativ auf die Komponente Zeitmanagement, wie von Schiefele und Wild (1994) verstanden, auswirkt. Die Wichtigkeit von Zeitmanagement im speziellen Fall von Onlinelehre wird auch in Kapitel 2.5.3 besprochen. In diesem Abschnitt sollen zwei Faktoren erläutert werden, welche erfolgreiches Zeitmanagement beeinflussen – einerseits die Prokrastination, andererseits aber auch der Konflikt mit anderweitigen Verpflichtungen wie etwa der Berufstätigkeit von Studierenden.

2.4.1 Prokrastination

Nach Schouwenburg (2004) geben rund 70% der Studierenden an, im Laufe ihres Studiums zu prokrastinieren. Ferrari und Tice (2000) haben Motive hinter dem Aufschieben von wichtigen Tätigkeiten erforscht und herausgefunden, dass dies oft bei Versagensangst eingesetzt wird: Im Falle von schlechten Ergebnissen werden dann nämlich nicht eigene fehlende Fähigkeiten verantwortlich gemacht, sondern die Behinderung (zu wenig Zeit infolge zu langen Zuwartens) selbst. Gute Ergebnisse hingegen zählen durch den knappen Zeitrahmen noch mehr; daher wird dies oft von Personen ausgeführt, welche Angst vor negativer Wahrnehmung haben (Ferrari und Tice 2000). Nach Lohbeck et al. (2017), welche Prokrastination unter Studierenden der Erziehungswissenschaften untersucht haben, neigen Studenten mit einem geringeren Selbstkonzept und geringerem Studieninteresse sowie dem Ziel der Arbeitsvermeidung öfters zu Prokrastination. Außerdem sind von diesem Phänomen häufiger Männer als Frauen betroffen (siehe u.a. Lohbeck et al. 2017; Ferrari und Tice 2000).

Schraw et al. (2007) haben – im Gegensatz zu den oben genannten Theorien – herausgefunden, dass Prokrastination unter Studierenden ein weit verbreitetes Phänomen ist; wobei es häufiger auftritt, je selbstbestimmter Studierende arbeiten und je weiter diese im Studium fortgeschritten sind. Daher führen sie das Konstrukt der *akademischen Prokrastination* (Schraw et al. 2007, S. 12) ein, welches beschreibt, dass zu erledigende Arbeit absichtlich aufgeschoben und überdacht wird. Das Prokrastinieren von Aufgaben wird dabei als Lernstrategie zur Selbstmotivation angesehen (Ferrari et al. 1995). Schraw et al. (2007) konnten in ihrer Studie auch feststellen, dass viele Studierende durch Prokrastination ein Stresslevel erreichen, das ihnen optimales Arbeiten erlaubt. Im Gegensatz zu vorherigen Studien und den daraus resultierenden Annahmen gab kein Proband in der Studie von Schraw et al. (2007) an, ausschließlich aufgrund von Faulheit oder Versagensangst zu prokrastinieren – wobei die Autoren durch-

aus bei den meisten Probanden ein gewisses Level an Langeweile, Faulheit, Angst oder Versagensangst feststellen konnten. Basierend auf Studien, welche auch positive Effekte von Prokrastination feststellen konnten, unterscheiden Chu und Choi (2005) zwischen aktiver und passiver Prokrastination. Aktive Prokrastination lässt sich dabei als aktive Entscheidung verstehen, das Fertigstellen von Arbeiten aufzuschieben, um dadurch optimal arbeiten zu können (Chu und Choi 2005). Passive Prokrastination hingegen wird zumeist unabsichtlich aufgrund von Entscheidungsschwäche begangen und beeinflusst den (akademischen) Erfolg negativ (Chu und Choi 2005).

Im Kontext von e-Learning sind Studierende in ihrer Entscheidung, wann und wo sie lernen möchten, freier – es benötigt daher mehr Selbstmotivation und -regulation vonseiten der Studierenden (Goda et al. 2015). Selbstregulation beschreibt dabei einen aktiven Lernprozess, wobei Studierende sich selbst Ziele setzen und die eigenen Lernfortschritte überwachen (Rakes und Dunn 2009). Goda et al. (2015) konnten feststellen, dass ein konstantes Lernverhalten (also das Vermeiden des Aufschiebens von Aufgaben) von Studierenden im Rahmen der Onlinelehre zu besseren akademischen Leistungen führt. Diskussionsforen bieten die Möglichkeit für Studierende, sich aktiv zu melden und Fragen zu stellen – Michinov et al. (2011) konnten diesbezüglich nachweisen, dass Personen, welche sich häufiger in Diskussionsforen einbringen, auch höhere akademische Leistungen erbringen. Dies lässt sich auch dadurch begründen, dass Personen, die Aufgaben aufschieben, erst später und dadurch weniger stark in Diskussionen involviert sind und sich in Foren melden (Michinov et al. 2011). Das reine Angebot an Diskussionsforen in einem Kursmanagement-System führt also nicht per se zu mehr Interaktionen von Studierenden (Michinov et al. 2011).

Zusammenfassend lässt sich sagen, dass Studierende in der Onlinelehre zur Vermeidung von Prokrastination und den damit einhergehenden negativen Folgen ein höheres Maß an Selbstregulation und -motivation benötigen (Goda et al. 2015). Die COVID-19-bedingte Umstellung auf Onlinelehre betrifft nicht nur Studierende, welche sich aktiv für ein Online-Studium oder Online-Kurse entschieden, sondern auch jene, welche Präsenz-Kurse belegt haben. Es kann daher nicht davon ausgegangen werden, dass alle Studierende das für Onlinelehre notwendige Maß an Selbstregulation und intrinsischer Motivation aufweisen. Dem entspricht auch, dass an der Universität Wien im Rahmen einer Befragung von Lehrenden und von Studierenden die Kommunikation im Sommersemester 2020 als tendenziell schlechter eingestuft wurde (Karall et al. 2020).

2.4.2 Berufstätigkeit von Studierenden

In Österreich hat ein hoher Anteil (über 70%) von Studierenden kein volles Zeitbudget zur Verfügung, sondern gilt als *faktisch Teilzeitstudierender* (Isensee und Wolter 2017, S. 17).

Dieser Begriff beschreibt Studierende, welche zwar in Vollzeitstudiengängen inskribiert sind, jedoch ein Zeitbudget unter der Referenzschwelle für ein solches Studium aufweisen (Isensee und Wolter 2017). Da die Masterstudiengänge an der Universität Wien als Vollzeitstudien ausgelegt sind, fallen berufstätige Studierende in die Kategorie *faktisch Teilzeitstudierende*. *Faktisch Teilzeitstudierende* nach dieser Definition gibt es auch in anderen Ländern – von rund 50% in Großbritannien (Callender 2008; Metcalf 2003) bis zu über 70% in Australien (Ryan et al. 2011). Über 80% der befragten *faktisch Teilzeitstudierenden* arbeiten neben dem Studium, um ihre Lebenshaltungskosten zu decken (Hall 2010; Callender 2008), für rund 50% der befragten Studierenden ist dies der einzige Grund, um neben dem Studium zu arbeiten (Hall 2010). 30% führten außerdem die Sammlung von Arbeitserfahrung an, wobei dies von rund 20% der befragten Personen der einzige Grund war, um zu arbeiten (Hall 2010). Die meisten der befragten *faktisch Teilzeitstudierenden* (über 90%) haben außerdem Zeitprobleme betreffend die Kombination von Studium und Arbeit, weiters wird die Motivation und Konzentration für das Studium von der Arbeit oftmals negativ beeinflusst (Hall 2010). 31% der befragten Studierenden geben an, dass eine erhöhte Flexibilität der Universität, etwa durch das Anbieten von Online-Lehrveranstaltungen oder mehr Flexibilität bei Abgabedeadlines, helfen würde (Hall 2010). Callender (2008) zeigt, dass zwischen der Berufstätigkeit neben dem Studium und dem akademischen Studienerfolg ein negativer Zusammenhang besteht – so führt etwa eine Teilzeitstelle von im Schnitt 15 Stunden pro Woche dazu, dass ein Studierender im Vergleich zu einem nicht-arbeitenden Studierenden eine um 40% geringere Wahrscheinlichkeit aufweist, eine gute (*first oder upper-second degree*) Abschlussnote zu haben.

Ähnlich wie bei der Onlinelehre müssen berufstätige Studierende eine hohe Selbstdisziplin sowie eine hohe intrinsische Motivation aufweisen (Seeber et al. 2006). Grundsätzlich kam der Umstieg auf Onlinelehre einer der Aufforderungen der *faktisch Teilzeitstudierenden*, nämlich einer erhöhten Flexibilität im Studium (Hall 2010) nach. Gleichzeitig wurde jedoch die Arbeit von rund 50% der befragten Studierenden in den USA direkt durch die COVID-19-Pandemie beeinflusst (Owens et al. 2020). Dies entspricht auch den Ergebnissen von Cohen et al. (2020), nach welchen etwa die Hälfte der befragten Studierenden in den USA, welche im Februar 2020 arbeiteten, im April 2020 arbeitslos waren. Von denjenigen Studierenden, welche im April noch Arbeit hatten, verdienten wiederum rund 50% weniger als noch im Februar (Cohen et al. 2020). Nach kürzlich erschienenen Medienberichten – aktuelle Studien sind nach Wissensstand der Autorin derzeit nicht verfügbar – ist die Situation in Deutschland und Österreich ähnlich: Rund 40% der Studierenden haben ihren Job verloren (Thaler 2020; Redaktionsnetzwerk Deutschland 2020). Da – wie oben angesprochen – die finanzielle Situation für viele Studie-

rende der Hauptgrund für die Arbeit neben dem Studium ist, führt eine Verschlechterung derselben zu einer Erhöhung von psychischen Belastungen und den daraus folgenden Beeinträchtigungen der akademischen Leistungen.

Durch die wahrgenommenen geringeren Interaktionen in der Onlinelehre und die damit einhergehende Problematik der Prokrastination wird eine negative Entwicklung der akademischen Studierendenleistungen erwartet. Der erhöhten Flexibilität, welche berufstätigen Studierenden entgegenkommt, stehen erhöhte psychische Belastungen aufgrund von Jobverlust und finanziellen Problemen entgegen. Die Erwartungshaltung an die Effekte der Umstellung von Präsenz- auf Onlinelehre ist daher wie folgt:

„Die Umstellung von Präsenz- auf Onlinelehre und die damit einhergehenden negativen Einflüsse auf das Zeitmanagement beeinflussen Studierendenleistungen negativ.“

2.5 Onlinelehre

2.5.1 Herausforderungen der Onlinelehre

Bereits vor der pandemiebedingten abrupten Umstellung der Präsenz- auf Onlinelehre wurden Onlinestudiengänge, welche hauptsächlich für ihre Möglichkeit der berufsbegleitenden Ausbildung wahrgenommen wurden, ausgebaut (Nebel 2017). Dabei wurde auch bereits die Problematik des Einzelkämpfertums erkannt – durch fehlende persönliche Kontakte sind Studierende, welche ein reines Onlinestudium beginnen, isoliert (Nebel 2017). Ebenso wird die fehlende Peer-Gruppe von Nebel (2017) angesprochen: Die Motivation zur Erledigung von Aufgaben und Ähnlichem muss daher von den Studierenden selbst kommen und zu einem hohen Grad intrinsischer Natur sein. Im Rahmen einer österreichweiten Befragung von Studierenden konnte festgestellt werden, dass sich im ersten Lockdown in Österreich (15.03.2020-01.05.2020) etwa 7% der Studierenden sehr oder ziemlich erfolgreich im Distance-Learning bezeichneten, wohingegen 16% von Problemen hierbei berichteten (Schober et al. 2020). Studierende gaben an, dass ihnen der Kontakt zu Kollegen am meisten fehlt (Schober et al. 2020) – das Fehlen der sozialen Eingebundenheit durch Präsenzeinheiten konnte durch Distance-Learning-Angebote nicht kompensiert werden. Nebel (2017) formuliert wichtige Voraussetzungen vonseiten der Studierenden und auch vonseiten der Universität für einen erfolgreichen Verlauf und Abschluss eines Onlinestudiums: (1) enge und persönliche Unterstützung beim Studieneinstieg, (2) Toleranz und Flexibilität des sozialen Umfelds von Studierenden, (3) technische Affinität der Studierenden (dies impliziert ausreichend ausgestattete technische Infrastruktur sowie das Wissen um den Umgang mit derselben), (4) Möglichkeiten für die Studierenden, sich miteinander zu vernetzen und der Isolation vorzubeugen, (5) funktionierende

Peer-to-Peer-Netzwerke für gemeinsame Lernerfahrungen, (6) eine Abfrage der Vorkenntnisse von Studierenden und (7) ein kompetenter Umgang mit Phasen der Spannung und Entspannung sowie Drucksituationen.

In Bezug auf die Lehrmethoden, welche in der Onlinelehre eingesetzt werden, kann zwischen synchronen und asynchronen Methoden unterschieden werden. Sogenannte „Course Management Systems“, auch CMS abgekürzt (Fita et al. 2016, S. 982), wie etwa das von der Universität Wien eingesetzte System *Moodle*, unterstützen asynchrones Lernen (Fita et al. 2016). Dazu gehört, dass Studierende keine fix vorgegebenen Zeiten haben, in welchen bestimmte Aufgaben erledigt werden müssen, sondern selbst entscheiden können, wann Aufgaben erledigt oder Aufzeichnungen angesehen werden (Fita et al. 2016). Gleichzeitig führt dies jedoch zu einer Abnahme der direkten Kommunikation zwischen Studierenden und Lehrenden (Fita et al. 2016). Sogenannte „Synchronous Virtual e-Learning tools“, als SVL abgekürzt (Fita et al. 2016, S. 983), funktionieren hingegen nach dem Prinzip eines Webinars und schaffen eine kursähnliche Atmosphäre. Die von der Universität Wien eingesetzten Tools *Collaborate* oder *Big Blue Button* zählen zu solchen Anwendungen.

Von den oben genannten Voraussetzungen nach Nebel (2017) sind vor allem Toleranz und Flexibilität (2), Möglichkeiten zur Vernetzung (4) sowie die Ermöglichung von Peer-to-Peer-Netzwerken (5) vonseiten der Universität oder Bildungseinrichtung für einen erfolgreichen COVID-19-bedingten Umstieg von Präsenz- auf Onlinelehre zu erbringen, die restlichen Voraussetzungen beziehen sich auf die Studierenden. Voraussetzung (1), nämlich enge Unterstützung beim Studieneinstieg, wird im Sommersemester nur wenige Studierende betreffen, da der reguläre Studienbeginn im Wintersemester liegt. Ebenso wird Voraussetzung (6) in dieser Arbeit nicht näher betrachtet, da das Vorwissen von Studierenden vor allem in berufsbegleitenden Studiengängen relevant ist und durch die Pandemiesituation nicht beeinflusst wurde. Die anderen Voraussetzungen lassen sich jedoch auf die COVID-19-bedingte Absenz und Onlinelehre an der Universität Wien im Sommersemester 2020 sowie im Wintersemester 2020/2021 umlegen. Im Folgenden wird nochmals detailliert auf die Voraussetzungen von Studierenden und Lehrenden eingegangen, welche für erfolgreiche Onlinelehre im akademischen Kontext notwendig sind. Außerdem wird ausgeführt, welche Technologien an der Universität Wien im Rahmen des Distance-Learning aufgrund von COVID-19 eingesetzt wurden und welche Voraussetzungen und technisches Equipment die beteiligten Personen dafür aufweisen müssen.

2.5.2 Verwendete Technologien im SS 2020 und WS 2020/21 an der Fakultät für Wirtschaftswissenschaften der Universität Wien

Mit März 2020 wurden die Lehrveranstaltungen an der Universität Wien auf das Format Distance-Learning umgestellt – aufgrund des ersten Lockdowns (verordnet aufgrund § 2 COVID-19-Maßnahmengesetz BGBl I 12/2020 StF mit VO des Gesundheitsministers BGBl II 98/2020) war es nicht mehr möglich, vor Ort zu studieren. Daher wurden unterschiedlichste Lern- und Lehrformate von Studierenden sowie Lehrenden genutzt. Laut *u:wiki* der Universität Wien werden sieben verschiedene Videokonferenz-Tools angewendet: *Jitsi*, *BigBlueButton (BBB)*, *Blackboard Collaborate*, *MS Teams*, *u:stream* und *Kaltura* (Karall und Pfeiffer 2020), wobei alle Tools unterschiedliche Teilnehmerbegrenzungen aufweisen. Außerdem werden nur die Tools *BigBlueButton* und *Blackboard Collaborate* für Übungen und Seminare sowie zusätzlich *u:stream* und *Kaltura* für Vorlesungen empfohlen (Karall und Pfeiffer 2020). Zu jedem der Tools werden den Lehrenden außerdem Anleitungen geboten, um die technologischen Aspekte zu vereinfachen – dies entspricht auch Voraussetzung (3) nach Nebel (2017) in Bezug auf Lehrkräfte. Diese Videokonferenz-Tools konnten auch für mündliche Prüfungen eingesetzt werden (Karall 2020).

Weiters wird *Moodle* als Kursmanagement-System verwendet; wie auch die meisten anderen Tools war *Moodle* bereits vor Eintritt der Ausgangsbeschränkungen in Verwendung. So war im Sommersemester 2019 bereits für 5.896 Kurse ein *Moodle*-Kurs angelegt, im Sommersemester 2020 stieg diese Anzahl auf 8.673. Dieser starke Anstieg relativiert sich jedoch, wenn beachtet wird, dass bereits im Wintersemester 2019 7.019 Kurse angelegt wurden (Karall et al. 2020). Dass dieses Tool auch von den Studierenden angenommen wurde, zeigt sich in der Zahl von 45.000 *Moodle*-Nutzern im Sommersemester 2020. Auch alle schriftlichen Prüfungen wurden über das Kursmanagement-System *Moodle* abgehalten (Karall 2020). Mit Stand Juli 2020 nutzten bereits 91,1% aller Lehrveranstaltungen einen *Moodle*-Kurs, um mit Studierenden zu kommunizieren oder Unterlagen zur Verfügung zu stellen. Rund 59% der Lehrenden verwendeten im Sommersemester 2020 neu erstellte Lehrmaterialien, 41,7% auch neu zusammengestellte Lehrmaterialien – nur rund 1,1% der Lehrenden haben ihre Lehrmaterialien nicht angepasst (Karall et al. 2020). Generelles Fazit der Lehrendenbefragung an der Universität Wien ist, dass Lehrende Distance-Learning als Ergänzung, keinesfalls aber als Ersatz für Präsenzlehre sehen (Karall et al. 2020).

2.5.3 Anforderungen an Studierende

Nach Hao (2016) sollen drei Punkte bei Studierenden vorhanden sein, damit diese gut auf das e-Learning vorbereitet sind: Technologisches Wissen und Kompetenz, die Fähigkeit, selbstständig zu lernen und Kommunikationsfähigkeiten. Yilmaz (2017) verwendet zur Messung die

E-Learning Readiness Scale, bestehend aus Skalen für die Kompetenz in Bezug auf Computer, Internet, Online-Kommunikation sowie selbst-angeführtes Lernen, Lernkontrolle und Motivation. Auch wenn diese Kompetenzen in den Studien von Yilmaz (2017) und Hao (2016) mit dem Flipped-Classroom-Konzept in Verbindung gebracht werden, können diese auch auf e-Learning angewendet werden. Das Vorhandensein solcher Kompetenzen führt zu einer höheren Zufriedenheit der Studierenden mit der e-Learning-Methode (Yilmaz 2017).

Technische Anforderungen

Die technische Affinität von Studierenden wird sowohl von Nebel (2017) als auch von Hao (2016) und Yilmaz (2017) als wichtige Voraussetzung angesprochen. Durch den Einsatz von zahlreichen Online-Tools müssen Studierende einerseits überhaupt die Möglichkeit haben, auf diese zugreifen zu können (mittels passenden technischen Equipments wie etwa Laptop, Lautsprecher, eventuell Videokamera und Mikrofon), andererseits muss auch das Wissen betreffend der Funktionsweise verschiedener Tools und die entsprechenden Computereinstellungen vorhanden sein. In einer Studie mit 84 Undergraduate-Studenten konnte Hao (2016) feststellen, dass die technologischen Komponenten bei den Studierenden am meisten fortgeschritten waren und neue Technologien daher für diese kein Problem darstellten. Der Autor führt dies dabei auf den Aspekt der „digital natives“ zurück, da viele Studierende bereits mit neuen Technologien aufgewachsen sind und keine Schwierigkeiten haben, diese anzuwenden (Hao 2016).

Zeit- und Stressmanagement

Ein kompetentes Zeit- sowie Stressmanagement vonseiten der Studierenden erweist sich in Zeiten von erhöhten psychischen Belastungen (siehe Kapitel 2.3) als weitere wichtige Voraussetzung nach Nebel (2017) für erfolgreiche akademische Leistungen. Auch dieser Punkt ist in der *E-Learning Readiness Scale* von Yilmaz (2017) sowie in den Anforderungen von Hao (2016) implizit enthalten, da sich selbstorganisiertes und diszipliniertes Lernen aus der Organisation des eigenen Studienalltags sowie aus dem Umgang mit herausfordernden Situationen zusammensetzt. Das psychologische Kapital, wie von Luthans et al. (2012) definiert, kann diesbezüglich eine wichtige Grundlage darstellen, um auch in Zeiten der COVID-19-Pandemie Stresssituationen zu bewältigen, damit akademische Leistungen nicht negativ beeinflusst werden. Gerade das Selbstbewusstsein (*self-efficacy*) in Verbindung mit Optimismus als Bestandteile des psychologischen Kapitals von Studierenden stellt dabei einen elementaren Teil von erfolgreichem Zeit- und Stressmanagement dar. E-Learning ist also für jene Studierenden leichter, welche ein hohes psychologisches Kapital aufweisen. Wie auch bereits in Kapitel 2.4

angesprochen, sind die Aspekte Lernkontrolle und Motivation wichtige Bestandteile von erfolgreichem Stress- und Zeitmanagement.

2.5.4 Anforderungen an Lehrende

Technische Anforderungen

Ähnlich wie Studierende müssen auch Lehrende eine entsprechende technische Ausstattung und technisches Grundwissen zur Bedienung der Online-Tools besitzen (Nebel 2017). Durch die umfassende Betroffenheit von COVID-19 und die damit einhergehenden *Social-Distancing*-Maßnahmen mussten auch Lehrende, welche normalerweise kein Interesse an technologischen Hilfsmitteln oder e-Learning zeigen, Onlinelehre einsetzen. Auch hier muss also diese Voraussetzung erfüllt werden, um erfolgreiches e-Learning zu gewährleisten.

Toleranz und Flexibilität

Während sich Nebel (2017) auf das soziale Umfeld des Studierenden bezieht, müssen Toleranz und Flexibilität auch vonseiten der Lehrenden gegeben sein. Studierenden mit Kindern kann die Konzentration auf das Studium und die Anwesenheit bei Lehrveranstaltungen erschwert werden, da zu bestimmten Zeiten Schulen und Kinderbetreuungseinrichtungen geschlossen waren (Fassmann 2020) und eine Betreuung durch Großeltern ebenfalls oftmals nicht erwünscht und empfohlen wurde (Leonhartsberger 2020). Daher müssen in dieser Situation nicht nur das soziale Umfeld der Studierenden, sondern genauso Lehrende Toleranz und Flexibilität zeigen. Nicht alle Studierende waren zu den ursprünglich vereinbarten Lehrveranstaltungszeiten verfügbar (durch private genauso wie universitäre Gründe, wie etwa Kursverschiebungen).

Möglichkeiten zur Vernetzung von Studierenden

In Präsenzform gab es sowohl im Sommer- als auch im Wintersemester nur begrenzte oder überhaupt keine Möglichkeiten, um der Isolation von Studierenden vorzubeugen. Nichtsdestoweniger ist – wie bereits in Kapitel 2.2 besprochen – die Interaktion von Studierenden für den Studienerfolg äußerst wichtig und kann als Voraussetzung für erfolgreiche Onlinelehre angesehen werden. Eine Möglichkeit zur Förderung der Interaktion und der Bildung eines Peer-to-Peer-Netzwerkes sind peer-basierte Lernaufgaben wie etwa Gruppenaufgaben. Um Prokrastination in der Onlinelehre und damit einhergehende schlechtere akademische Leistungen (siehe Kapitel 2.4.1) zu vermeiden, haben Artino und Stephens (2009) einige Aufgaben von Lehrenden identifiziert: Lehrende sollen explizite Instruktionen und Strukturen vorgeben, Online-Diskussionen und damit auch Interaktionen fördern und sooft möglich Feedback geben (Artino und Stephens 2009, Michinov et al. 2011).

2.6 Zusammenfassung der Einflüsse und Entwicklung der Arbeitshypothese

Durch die plötzliche Umstellung von Präsenz- auf Onlinelehre hatten viele Studierende keine Möglichkeit, im Rahmen von Kursen an der Universität ihr soziales Kapital auszubauen und Beziehungen mit anderen Studierenden aufzubauen (Beattie und Thiele 2016, Chambliss und Takacs 2014). Onlinelehre bietet zwar die Möglichkeit, Gruppenarbeiten und ähnliches durchzuführen, informelle Möglichkeiten zur Kommunikation fehlen jedoch größtenteils. Es ist daher davon auszugehen, dass Studierendenleistungen negativ beeinflusst werden. Das Fazit der literarischen Untersuchung des Themengebietes der sozialen Eingebundenheit von Studierenden lautet wie folgt:

„Es ist davon auszugehen, dass die physische Absenz von Studierenden und Lehrenden im reinen e-Learning-Kontext zu einer geringeren Dichte der sozialen Netzwerke führt und Studierendenleistungen tendenziell negativ beeinflusst werden.“

Nicht nur finanzielle Belastungen durch etwaigen Arbeitsverlust, sondern auch soziale Distanz und gesundheitliche Sorgen führten zu einem spürbaren Anstieg an psychischen Belastungen von österreichischen Studierenden (Braun et al. 2020). Dies entspricht auch Beobachtungen von Studierenden in anderen Ländern (siehe etwa Hakami et al. 2020, Husky et al. 2020, Hasan und Bao 2020).

„Durch die steigende psychische Belastung von Studierenden in Zeiten der COVID-19-Pandemie ist davon auszugehen, dass akademische Leistungen negativ beeinflusst werden.“

Im Rahmen von Onlinelehre benötigen Studierende ein höheres Ausmaß an Selbstdisziplin und intrinsischer Motivation (Goda et al. 2015). Es ist – auch im Hinblick auf den Anstieg an psychischen Belastungen und Unsicherheiten – nicht davon auszugehen, dass alle Studierenden die notwendige Selbstmotivation und -regulation zur erfolgreichen Onlinelehre aufweisen. Außerdem wurde in einer Umfrage von Lehrenden und Studierenden an der Universität Wien bereits festgestellt, dass die Kommunikation im Sommersemester 2020 tendenziell schlechter zu bewerten ist, was auf geringere Interaktionen von Studierenden schließen lässt (Karall et al. 2020). Das Fazit der plötzlichen Umstellung auf Onlinelehre lautet daher:

„Die Umstellung von Präsenz- auf Onlinelehre und die damit einhergehenden negativen Einflüsse auf das Zeitmanagement beeinflussen Studierendenleistungen negativ.“

Die Literaturanalyse und bereits bestehende Studien lassen darauf schließen, dass die Corona-Pandemie eine Verschlechterung der Lernsituation vieler Studierenden nach sich zieht. Das Fazit der Themengebiete deutet auf einen negativen Einfluss der Corona-Pandemie

hinsichtlich der akademischen Studierendenleistungen hin. Es wird daher folgende Arbeitshypothese entwickelt:

„Prüfungsleistungen von Studierenden, welche in Semestern während der Corona-Pandemie erbracht wurden, sind schlechter als Prüfungsleistungen von Studierenden, welche in Semestern vor der Corona-Pandemie erbracht wurden.“

3 Methodische Vorgehensweise

Von der Datenabteilung der Universität Wien Ergebnisse von Prüfungen im Zeitrahmen vom Wintersemester 2013/14 bis zum Wintersemester 2020/21 zur Verfügung gestellt werden. Da die Prüfungsergebnisse für das Wintersemester 2020/21 erst im Rahmen einer Datennachlieferung im April 2021 zur Verfügung gestellt werden konnten, behandelt das vorliegende Kapitel zuerst die ursprüngliche Datenlieferung und anschließend den umfangreichen Datensatz. Die gelieferten Daten betreffen alle Prüfungen, welche unter Mitwirkung von ao. Univ.-Prof. Mag. Dr. Schaffhauser-Linzatti als Prüferin durchgeführt wurden. Während die deskriptiven Analysen und einfache Chi²-Berechnungen im Statistikprogramm *R* durchgeführt werden, wurde für die Berechnung der komplexeren Logit-Regressionen das Statistikprogramm *Stata* verwendet. Das vorliegende Kapitel beschreibt die Stichprobe sowie die Vorgehensweise der Datenanalyse – die Ergebnisse derselben werden anschließend in Kapitel 4 diskutiert.

Datenfeld	Beschreibung
RND	Eindeutige anonymisierte Identifikationsnummer für jeden Studierenden
Geschlecht	Das Geschlecht in den Kategorien männlich und weiblich (<i>m/w</i>)
Alter	Das Alter des Studierenden
Antritt Nr	Die Nummer des Antritts zur Prüfung
Spalte 1	Erreichte Note auf die Prüfung als Zahl (1, 2, 3, 4, 5, leer)
Note	Erreichte Note auf die Prüfung ausgeschrieben (<i>Sehr Gut, Gut, Befriedigend, Genügend, Nicht Genügend, mit Erfolg teilgenommen, ohne Erfolg teilgenommen</i>)
Prüferin	Indikator für die Prüfer der jeweiligen Prüfung. Entweder Frau ao. Univ.-Prof. Mag. Dr. Schaffhauser-Linzatti alleine oder gemeinsam mit weiteren nicht genannten Prüfern
Prüfungsdatum	Durchführungsdatum der Prüfung
Semester	Semester, in welchem die Prüfung durchgeführt wurde (codiert als JahrW oder JahrS zur Unterscheidung von Winter- und Sommersemester)
Typ	Art der Prüfung, entweder <i>LV-Prüfung, Fachprüfung</i> oder <i>Prüfungsteil</i>
Status	Status der Prüfungsbeurteilung, entweder <i>Freigegeben</i> oder <i>Bewertet</i>
Studienrichtung	Studienrichtung, in welche der Studierende inskribiert ist
Veranstaltung_Leistungsnummer&Name	Kombination aus den Informationen der Lehrveranstaltung und des Semesters
Prüfung	Beschreibung der Prüfung (Prüfungstitel)

Tabelle 1: Datenfelder aus der Datenlieferung

3.1 Stichprobenbeschreibung

3.1.1 Ursprünglich gelieferte Stichprobe

Die Daten umfassen 17.473 Prüfungsergebnisse im Zeitraum Wintersemester 2013/14 bis Wintersemester 2020/21 von insgesamt 9.128 Studierenden. Die gelieferten Datenfelder und Beschreibungen dazu sind in Tabelle 1 zu finden.

Von den 9.128 Studierenden in der Stichprobe sind 45,02% (4.109 Studierende) männlich und 54,98% (5.019 Studierende) weiblich. Im Durchschnitt sind die Studierenden 26,38 Jahre alt, männliche Studierende sind dabei unwesentlich älter (im Durchschnitt 26,77 Jahre) als weibliche Studierende (im Durchschnitt 26,06 Jahre). Dies entspricht auch dem Median der Stichprobe, welcher bei 26 Jahren liegt. Wie in Abbildung 1 ersichtlich ist, ist der Großteil der Studierenden zwischen 20 und 30 Jahre alt, wobei der jüngste Studierende in der Stichprobe 17 Jahre und der älteste Studierende in der Stichprobe 76 Jahre alt ist.

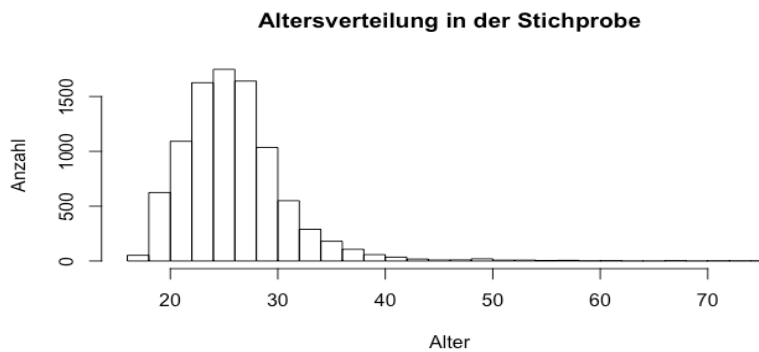


Abbildung 1: Altersverteilung in der Stichprobe

Aus der Stichprobe lässt sich jedoch eine klare Verschiebung der Altersstruktur der Studierenden herauslesen. In Abbildung 2 ist das durchschnittliche Alter pro Jahr dargestellt – während es im Jahr 2013 noch bei 29,7 Jahren liegt, sinkt das durchschnittliche Alter bis zum Jahr 2020

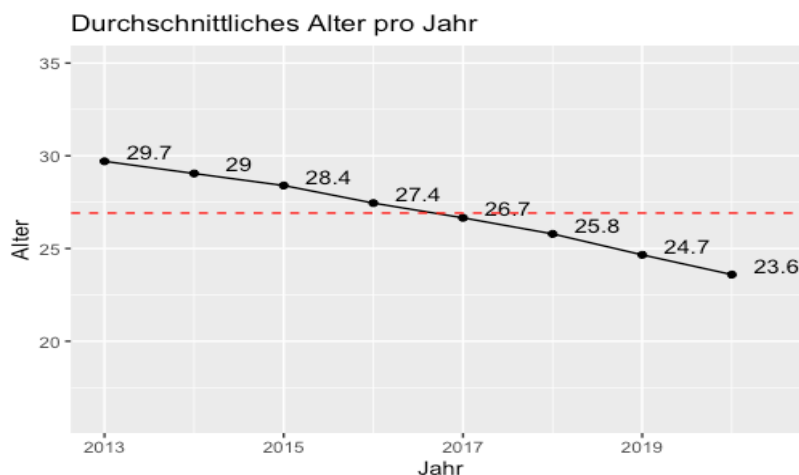


Abbildung 2: Durchschnittliches Alter pro Jahr, gesamte Stichprobe

stark ab und beträgt nur noch 23,6 Jahre. Dies scheint also nicht durch die Corona-Pandemie beeinflusst zu sein, sondern zeigt einen kontinuierlichen Abfall des durchschnittlichen Alters pro Jahr der Prüflinge.

Der Großteil der abgelegten Prüfungen waren LV-Prüfungen (9.119 Prüfungen, 52,19%) und Fachprüfungen (8.253 Prüfungen, 47,23%), nur ein kleiner Teil der Prüfungen entfällt auf Prüfungsteile (101 Prüfungen, 0,6%). Von den LV-Prüfungen wurden wiederum der Großteil von Frau ao. Univ.-Prof. Mag. Dr. Schaffhauser-Linzatti als einziger Prüferin durchgeführt (8.019 Prüfungen, das entspricht 87,9% der LV-Prüfungen), wohingegen die Fachprüfungen und Prüfungsteile zum Großteil von Frau ao. Univ.-Prof. Mag. Dr. Schaffhauser-Linzatti gemeinsam mit weiteren Prüfern geprüft wurden (8.206 Fachprüfungen, das entspricht 99,4% aller Fachprüfungen sowie 97 Prüfungsteile, das entspricht 96% aller Prüfungsteile in den Daten).

12.969 (74,2%) der Prüfungen in den Daten sind Erstantritte, davon sind 7.038 (54,3%) LV-Prüfungen, 5.831 (45%) Fachprüfungen und 100 (0,7%) Prüfungsteile. Von den restlichen Prüfungen sind 3.170 (18,1%) Zweitantritte, 3.170 (5,5%) Drittantritte und 337 (1,9%) Prüfungen entsprechen einem weiteren (vierten oder höheren) Antritt.

Weiters wurde der Großteil der untersuchten Prüfungen im Rahmen eines Bachelorstudiums (15.466 Prüfungen, das entspricht 88,5%) oder eines Masterstudiums (1.205 Prüfungen, das entspricht 6,9%) abgelegt. Weitere Studiengänge, welche in den Daten vorhanden sind, sind Lehramtsstudien (224 Prüfungen), Universitätslehrgänge (162 Prüfungen), Diplomstudien (105 Prüfungen), Doktoratsstudien (11 Prüfungen) sowie andere Studiengänge (300 Prüfungen, davon fallen 297 unter „Besuch einzelner Lehrveranstaltungen UniStG“). Die Anzahl der durchgeführten Prüfungen verteilt sich auf die Semester wie in Tabelle 2 dargestellt.

Semester	Anzahl Prüfungen	Prüfungen insgesamt
2013S 2013W	913 834	1.747
2014S 2014W	989 908	1.897
2015S 2015W	531 1.226	1.757
2016S 2016W	669 1.477	2.146
2017S 2017W	969 1.767	2.736
2018S 2018W	1.065 1.319	2.384
2019S 2019W	957 1.703	2.660
2020S 2020W	1.018 1.128	2.146

Tabelle 2: Darstellung Prüfungsverteilung über Semester und Jahre

Dabei muss festgehalten werden, dass die Leistungen von 2020W nicht alle Prüfungen in diesem Semester beinhalten können, da die Datenlieferung bereits Mitte Jänner 2021 – und damit noch vor Ende des Semesters – erfolgte. Erst durch die Datennachlieferung wurden die Ergebnisse des Wintersemesters 2020/21 vollständig erfasst (siehe Kapitel 3.1.3).

Die erreichten Noten auf die Prüfungen sind nicht normalverteilt. Es wurden 841 *Sehr Gut* (4,8%), 2.131 *Gut* (12,2%), 3.032 *Befriedigend* (17,4%), 4.259 *Genügend* (24,4%), 5.782 *Nicht Genügend* (33%) sowie 1.343 *mit Erfolg teilgenommen* (7,7%) und 85 *ohne Erfolg teilgenommen* (0,5%) vergeben. Tabelle 3 stellt die Notenverteilung über den Semesterverlauf hinweg dar.

Note	Semester															
	2013 S	2013 W	2014 S	2014 W	2015 S	2015 W	2016 S	2016 W	2017 S	2017 W	2018 S	2018 W	2019 S	2019 W	2020 S	2020 W
1	29	36	87	46	9	54	29	59	39	27	39	91	33	92	88	83
2	53	91	115	111	36	151	82	128	99	85	206	165	120	252	197	240
3	109	109	173	167	81	231	108	204	181	217	200	218	234	320	262	218
4	190	219	268	243	157	278	168	384	265	460	253	321	250	416	222	165
5	409	379	306	341	247	512	264	604	335	740	270	402	254	457	172	90
Mit E. tg.	96	0	15	0	1	0	18	98	50	238	97	122	66	166	77	299
Ohne E. tg.	27	0	25	0	0	0	0	0	0	0	0	0	0	0	0	33

Tabelle 3: Notenverteilung pro Semester

3.1.2 Stichprobeneinschränkung

Da keine weiteren Informationen über die zusätzlichen Prüfer geliefert werden konnten, wird die Stichprobe auf Prüfungen, welche von Frau ao. Univ.-Prof. Mag. Dr. Schaffhauser-Linzatti durchgeführt wurden, eingeschränkt. Somit kann sowohl eine Beeinflussung der Ergebnisse durch andere Prüfer als auch durch einen gegebenenfalls anderen Prüfungsmodus ausgeschlossen werden. Daraus resultierend ergibt sich eine Stichprobe von 8.035 Prüfungen von 5.345 Studierenden. Da der Großteil dieser Prüfungen (99,8%) vom Typus LV-Prüfung ist, wird die Stichprobe auf diesen Typus eingeschränkt. Außerdem werden nur jene Prüfungen betrachtet, welche im Rahmen eines Bachelorstudiums (7.150 Prüfungen) oder eines Masterstudiums (510 Prüfungen) abgelegt wurden. Weiters werden in einem ersten Schritt nur die Erstantritte von Studierenden betrachtet, um Wiederholungseffekte oder andere Einflüsse bei Zweit- und Mehrantritten auszuschließen.

Aus diesen Einschränkungen resultiert eine Stichprobe von 5.658 Prüfungen, welche von insgesamt 4.940 Studierenden abgelegt wurden. 2.100 Studierende, das entspricht 42,5%, sind männlich, 2.840 Studierende (57,5%) sind weiblich. Die Studierenden sind im Durchschnitt 26,8 Jahre alt, wobei männliche Studierende geringfügig älter (27,3 Jahre) als weibliche Studierende (26,5 Jahre) sind. Der jüngste Studierende in der Stichprobe ist 19 Jahre alt, der älteste Studierende 71 Jahre. Der Großteil der betrachteten Prüfungen wurden im Rahmen eines Bachelorstudiums abgelegt (5.188 Prüfungen, also 91,7%), die restlichen betrachteten Prüfungen (470 Prüfungen, 8,3%) wurden im Rahmen eines Masterstudiums abgelegt. Alle betrachteten Prüfungen sind Erstantritte der Studierenden. Die Verteilung über die unterschiedlichen Semester wird in Tabelle 4 dargestellt. Bei dieser Darstellung wird besonders deutlich, dass die Daten aus dem Wintersemester 2020 (7 Prüfungen) nur zu einem marginalen Teil vorhanden sind.

Semester	Anzahl Prüfungen	Prüfungen insgesamt
2013S	275	488
2013W	213	
2014S	432	789
2014W	357	
2015S	189	627
2015W	438	
2016S	270	635
2016W	365	
2017S	343	966
2017W	623	
2018S	411	868
2018W	457	
2019S	407	876
2019W	469	
2020S	402	409
2020W	7	

Tabelle 4: Darstellung Prüfungsverteilung über Semester und Jahre, eingeschränkte Stichprobe

Die Notenverteilung der betrachteten Prüfungen ist wie folgt: 300 (5,3%) *Sehr Gut*, 649 (11,5%) *Gut*, 944 (16,7%) *Befriedigend*, 1.226 (21,7%) *Genügend* und 2.022 (35,7%) *Nicht Genügend* sowie 475 (8,4%) *mit Erfolg teilgenommen* und 42 (0,7%) *ohne Erfolg teilgenommen*. Tabelle 5 stellt die Notenverteilung aufgeteilt auf Semester in der eingeschränkten Stichprobe dar:

Note	Semester															
	2013 S	2013 W	2014 S	2014 W	2015 S	2015 W	2016 S	2016 W	2017 S	2017 W	2018 S	2018 W	2019 S	2019 W	2020 S	2020 W
1	10	12	60	29	2	16	14	18	10	16	18	28	17	23	24	3
2	11	19	47	38	19	31	29	47	33	31	127	37	66	62	52	0
3	14	23	66	59	23	57	43	56	77	81	78	47	114	78	128	0
4	36	67	88	89	58	101	59	76	89	102	75	78	94	108	106	0
5	100	92	134	142	87	233	107	158	134	184	109	148	116	182	92	4
Mit E. tg.	85	0	14	0	0	0	18	10	0	209	4	119	0	16	0	0
Ohne E. tg.	19	0	23	0	0	0	0	0	0	0	0	0	0	0	0	0

Tabelle 5: Notenverteilung pro Semester, eingeschränkte Stichprobe

Ähnlich wie bei der gesamten Stichprobe ist auch in der eingeschränkten Stichprobe ein starker Abfall des durchschnittlichen Alters pro Jahr zu beobachten. Die Entwicklung der Altersstruktur wird in Abbildung 3 dargestellt. Während das durchschnittliche Alter der Studierenden im Jahr 2013 noch bei 30,5 Jahren liegt, sinkt es bis zum Jahr 2019 auf 24,3 Jahre und steigt im Jahr 2020 wieder auf 24,6 Jahre an. Vergleicht man also die zwei Zeitperioden vor Eintritt der Corona-Pandemie (Jahre 2013 bis 2019) und während der Corona-Pandemie (Sommersemester 2020), so unterscheidet sich das durchschnittliche Alter deutlich: Das durchschnittliche Alter von Studierenden vor der Corona-Pandemie liegt bei 26,96 Jahren, während der Corona-Pandemie hingegen bei 24,61 Jahren. Die älteste Person im Zeitraum vor der Corona-

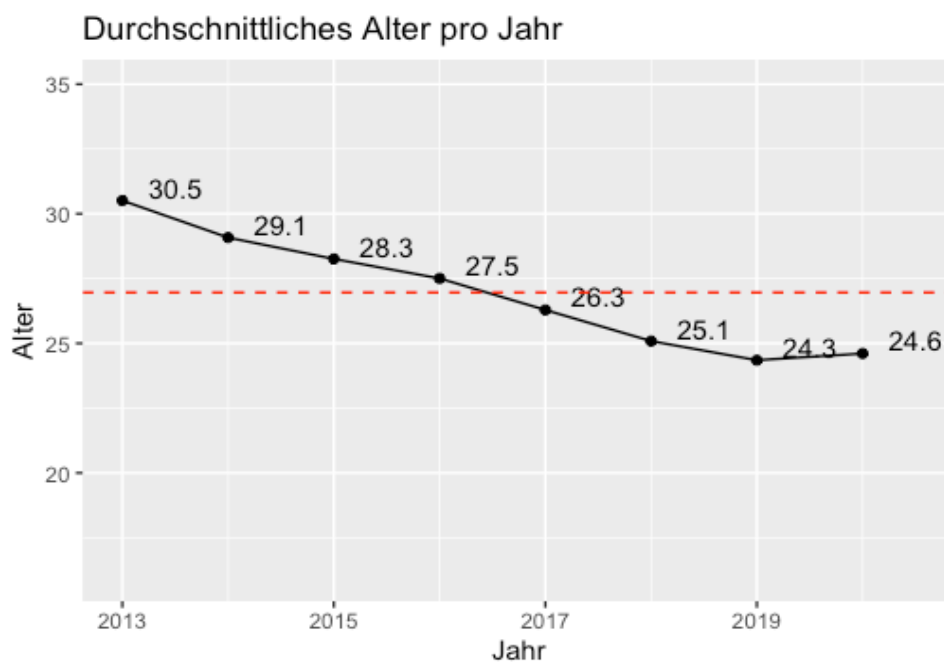


Abbildung 3: Durchschnittliches Alter pro Jahr, eingeschränkte Stichprobe

Pandemie ist 71 Jahre alt, während der Corona-Pandemie hingegen 56 Jahre. Dementsprechend ist auch eine starke Verschiebung der Altersstruktur zu beobachten: Im Zeitraum Sommersemester 2013 bis Wintersemester 2019 waren 38,7% der Studierenden in der Alterskategorie 16 bis 25 Jahre zu finden, 57,9% in der Alterskategorie 26 bis 35 Jahre, rund 3% in der Alterskategorie 36 bis 45 Jahre und die restlichen Studierenden (rund 0,4%) waren älter als 45 Jahre. In den beiden Semestern Sommersemester und Wintersemester 2020 sind in der angegebenen Stichprobe jedoch 67% der Studierenden der Alterskategorie 16 bis 25, 30,3% der Alterskategorie 26 bis 35 sowie rund 2% der Kategorie 36 bis 45 Jahre zuzuweisen und die restlichen 0,7% sind älter als 45 Jahre.

3.1.3 Datennachlieferung

Die Datenmenge konnte im April 2021 um die finalen Prüfungsergebnisse des Wintersemesters 2020/21 ergänzt werden. Somit wurden insgesamt 18.304 Prüfungsergebnisse von 9.335 Studierenden geliefert (davon 831 Beobachtungen in der Nachlieferung). In dieser erweiterten Grundgesamtheit sind 45,02% der Studierenden männlich und 54,98% weiblich. Das durchschnittliche Alter der Studierenden beträgt 26,3 Jahre, wobei männliche Studierende zu dem Zeitpunkt, als die Prüfung abgelegt wird, im Schnitt 26,9 Jahre alt sind und weibliche Studierende 26,2 Jahre.

Auch in dieser erweiterten Stichprobe entfällt der Großteil der Prüfungen auf LV-Prüfungen (9.743 Prüfungen, 53,2%) und Fachprüfungen (8.446 Prüfungen, 46,1%), wobei von den LV-Prüfungen 8.643 von Frau ao. Univ.-Prof. Mag. Dr. Schaffhauser-Linzatti als alleinige Prüferin durchgeführt wurden (das entspricht 88,7%). Wie bereits oben beschrieben, wird auch diese erweiterte Stichprobe zur Analyse eingeschränkt – es werden daher nur Erstantritte zu LV-Prüfungen, welche von Frau ao. Univ.-Prof. Mag. Dr. Schaffhauser-Linzatti durchgeführt wurden, analysiert. Dies resultiert in einer Stichprobe von 6.494 Prüfungen, welche infrage kommen würden. Von diesen Prüfungen wurden 5.646 LV-Prüfungen im Rahmen eines Bachelorstudiums und 529 LV-Prüfungen im Rahmen eines Masterstudiums absolviert.

Da erst mit der Datennachlieferung die Prüfungsergebnisse des Wintersemesters 2020 als vollständig erfasst gelten und eine ausreichende Datengrundlage bilden, werden die Analysen in dieser Arbeit einerseits mit dem zuerst gelieferten Datensatz (Aussagekraft nur für das Sommersemester 2020) und anschließend mit dem gesamten Datensatz inklusive Nachlieferung berechnet (Aussagekraft inklusive Wintersemester 2020/21).

3.2 Forschungsmethode

3.2.1 Binäre logistische Regression

In einem ersten Schritt wird als binäre abhängige Variable die Variable *Erfolg* definiert. Wie in Tabelle 6 (siehe Kapitel 3.3) dargestellt, nimmt diese den Wert 1 an, wenn ein Studierender die Prüfung bestanden hat, ansonsten den Wert 0. Die vorhergesagten Werte (predicted values) dieses Modells sind aufgrund der binären abhängigen Variablen Wahrscheinlichkeiten und können nicht direkt interpretiert werden (Stock und Watson 2011). Eine grobe Beschreibung des Modells der logistischen Regression nach Backhaus et al. (2018) ist

$$\pi(x) = f(x_1, \dots, x_n)$$

wobei $\pi(x) = \text{Prob}(Y = 1|x)$ darstellt. Dies ist also die bedingte Wahrscheinlichkeit für ein erfolgreiches Bestehen der Prüfung für die gegebenen Werte der unabhängigen Variablen, welche im Modell inkludiert sind (Backhaus et al. 2018). Die Regressionskoeffizienten zeigen dabei die Änderung der Wahrscheinlichkeit bei der Änderung einer Einheit der unabhängigen Variablen (Stock und Watson 2011). Um die Werte der Regression korrekterweise auf den Bereich $[0,1]$ einzugrenzen, muss eine Probit- beziehungsweise Logit-Regression durchgeführt werden. Die Eingrenzung auf den eingeschränkten Wertebereich ist notwendig, da Wahrscheinlichkeiten nur Werte zwischen 0 und 1 (also zwischen 0% und 100% Wahrscheinlichkeit) annehmen können – negative Werte oder Werte über 1 können somit nicht interpretiert werden. Während dem Probit-Regressionsmodell eine kumulierte Normalverteilung zugrunde liegt, ist beim Logit-Modell eine kumulierte logistische Verteilung die Basis – beide unterscheiden sich hauptsächlich in den Rändern der Verteilung, wodurch es nur bei Extremwerten zu stärkeren Abweichungen der Modelle kommt (Komlos und Süßmuth, 2010). Es wird daher ein Logit-Regressionsmodell gewählt, um reellwertige Variablen in eine Wahrscheinlichkeit zu transformieren (also vom Wertebereich $[-\infty, +\infty]$ zu $[0,1]$). Die Linearkombination der unabhängigen Variablen lässt sich daher nach Backhaus et al. (2018) für alle Regressoren wie folgt darstellen:

$$z(x) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n$$

Diese z-Werte fließen dabei in die logistische Regressionsfunktion ein:

$$\pi(x) = \frac{1}{1 + e^{-z(x)}}, \text{ wobei für } x_k \text{ das Modell wie folgt aussieht: } \pi(x_k) = \frac{1}{1 + e^{-(\alpha + \beta x_k)}}$$

Da in das Modell, welches in dieser Arbeit verwendet wird, mehrere unabhängige Variablen einfließen, muss die logistische Regressionsfunktion nochmals erweitert werden (für die unabhängigen Variablen x_1 bis x_j):

$$\pi(x) = \frac{1}{1 + e^{-(\alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_j x_j)}}$$

Der Koeffizient einer unabhängigen Variable in diesem Modell stellt die Änderung des z-Wertes bei der Änderung um eine Einheit der Variable dar (Stock und Watson 2011). Der z-Wert bildet im Modell also einen Prädiktor für die Wahrscheinlichkeit $\pi(x)$, wobei gilt, dass $\pi(x)$ wächst, wenn $z(x)$ wächst (Backhaus et al. 2018). Um die Werte des Modells korrekt zu interpretieren, müssen daher die marginalen Effekte berechnet werden (Stock und Watson 2011). Nach diesem Modell existiert für jede Prüfung, die abgelegt wird, eine Wahrscheinlichkeit, dass diese bestanden wird, welche wiederum durch die unabhängigen Variablen im Modell beeinflusst wird (Backhaus et al. 2018). Die Schätzmethode, welche für diese Variante der Regressionsanalyse angewandt wird, ist die *Maximum-Likelihood-Methode*. Das Prinzip dieser Methode lässt sich nach Backhaus et al. (2018, S. 288) wie folgt zusammenfassen: „Bestimme die Schätzwerte für die unbekannten Parameter so, dass die realisierten Daten maximale Plausibilität (Likelihood) erlangen“. Die zu maximierende Likelihood-Funktion, welche durch die Anwendung des Prinzips entsteht, ist dabei

$$L(a, b) = \prod_{k=1}^K p(x_k)^{y_k} * [1 - p(x_k)]^{1-y_k}$$

wobei in dem Modell dieser Arbeit $y_k = 1$ bei erfolgreich bestandener Prüfung (positive Note) und $y_k = 0$ bei nicht bestandener Prüfung eingesetzt wird. Aus dieser Likelihood-Funktion kann die zu maximierende Log-Likelihood-Funktion durch Logarithmieren und Umwandlung des Produktes in eine Summe gewonnen werden (Backhaus et al. 2018).

$$LL(a, b) = \sum_{k=1}^K \ln[p(x_k)] * y_k + \ln[1 - p(x_k)] * (1 - y_k)$$

Hier ist zu beachten, dass das Maximum der Log-Likelihood-Funktion jedenfalls kleiner oder gleich 0 sein wird, wobei man sich durch das Maximieren dem Wert 0 annähern möchte. Es werden iterative Algorithmen angewandt, welche aufgrund der Konvexität der Log-Likelihood-Funktion jedenfalls das globale Optimum finden (Backhaus et al. 2018). Die Interpretation kann über sogenannte *Odds* oder das *Odds Ratio* (OR) erfolgen, was Chancenverhältnissen entspricht (Hoffmann 2016). Während nämlich *Odds* das Verhältnis einer Wahrscheinlichkeit zu ihrer Gegenwahrscheinlichkeit beschreiben, also $odds = \frac{p}{1-p}$, beschreibt das Chancenverhältnis das Verhältnis von zwei *Odds*-Werten zueinander (Backhaus et al. 2018).

$$OR = \frac{odds(x+1)}{odds(x)} = e^b$$

e^b beschreibt also den Faktor, um welchen sich die *Odds* erhöhen, wenn die Variable x um eine Einheit erhöht wird und wird somit auch als Effekt-Koeffizient beschrieben (Backhaus et al. 2018). Ist e^b größer als 1, so steigen die Chancen für ein Ereignis um das e^b -fache, wenn sich x um eine Einheit vergrößert. Bei binären unabhängigen Variablen kann auch – ergänzend zum OR – das relative Risiko zur Interpretation herangezogen werden (Backhaus et al. 2018). Um die Interpretation von OR etwas zu erleichtern, wird ein kurzes Beispiel angeführt:

Angenommen, die *Odds* (Verhältnis Wahrscheinlichkeit zu Gegenwahrscheinlichkeit) für weibliche Teilnehmer, dass sie Produkt x ablehnen, anstatt es zu befürworten, beträgt 0,11, also 11%. Dieselbe *Odds* für männliche Teilnehmer beträgt 0,07, also 7%. Damit beträgt das OR von weiblichen zu männlichen Teilnehmern $0,11/0,07 = 1,57$. Diese Zahl besagt, dass die Wahrscheinlichkeit, dass weibliche Teilnehmer das Produkt ablehnen, 1,57-mal so hoch ist wie die Wahrscheinlichkeit, dass männliche Teilnehmer das Produkt ablehnen (Hoffmann 2016).¹

3.2.2 Multinomiale logistische Regression

Im Gegensatz zur binären logistischen Regression kann die abhängige Variable im multinomialen Modell mehr als zwei Ausprägungen annehmen. Im Folgenden wird theoretisch das multinomiale Modell nach Backhaus et al. (2018) durchbesprochen.

In diesem Modell stellt $\pi_g(x_k) = \text{Prob}(Y_k = g|x_k)$ dar, wobei g für eine Ausprägung der multinomial skalierten Zufallsvariablen, also der abhängigen Variablen, steht. Eine Voraussetzung des Modells ist daher auch, dass $\sum_{g=1}^G \pi_g(x_k) = 1$ gilt, wobei x_k als Zeilenvektor für $(x_{1k}, x_{2k}, \dots, x_{jk})$ mit j unabhängigen Variablen steht. Im Modell wird dabei eine der Kategorien der abhängigen Variablen als Referenzkategorie identifiziert, wobei die Parameter dieser Kategorie auf 0 gesetzt werden und die Parameter der übrigen Kategorien den relativen Effekt in Bezug auf diese Kategorie beschreiben (Backhaus et al. 2018). Daher gilt für alle Kategorien außer der Referenzkategorie (welche zumeist die letzte Kategorie G ist):

$$\pi_g(x) = \frac{e^{\alpha_g + \beta_{g1}x_1 + \dots + \beta_{gj}x_j}}{1 + \sum_{h=1}^{G-1} e^{\alpha_h + \beta_{h1}x_1 + \dots + \beta_{hj}x_j}}$$

¹ Die Zahlen des Beispiels (nicht jedoch die abhängige Variable) wurden aus Hoffmann (2016, S. 92) übernommen

wobei $g = 1, \dots, G-1$, also alle Kategorien außer die Referenzkategorie annimmt. Für die Referenzkategorie gilt dann

$$\pi_G(x) = \frac{1}{1 + \sum_{h=1}^{G-1} e^{\alpha_h + \beta_{h1}x_1 + \dots + \beta_{hj}x_j}}$$

Dies wird implizit auch bei der binären logistischen Regression umgesetzt, wobei immer der Wert 0 als Referenzkategorie bei der binären abhängigen Variablen gewählt wird. Die Wahrscheinlichkeit $p_g(x_k)$ ist somit wie folgt:

$$p_g(x_k) = \frac{e^{a_g + b_{g1}x_{1k} + \dots + b_{gj}x_{jk}}}{1 + \sum_{h=1}^{G-1} e^{a_h + b_{h1}x_{1k} + \dots + b_{hj}x_{jk}}} \text{ für } (g = 1, \dots, G)$$

wobei für die Referenzkategorie G $a_G = b_{G1} = \dots = b_{Gj} = 0$ gilt (Backhaus et al. 2018).

Die Verwendung einer nicht-binären kategorialen abhängigen Variable hat Auswirkungen auf die Log-Likelihood-Funktion, welche auch in diesem Modell zu maximieren ist (Backhaus et al. 2018):

$$LL = \sum_{k=1}^K \sum_{g=1}^G \ln[p_g(x_k)] * y_{gk}$$

($y_{gk} = 1$ wenn die Kategorie g im Fall k auftritt).

Die Interpretation der Ergebnisse des Modells, zum Beispiel die OR, erfolgt damit immer relativ zur Referenzkategorie. Generell kann aber für die Berechnung der *Odds* jede Kategorie als Referenzkategorie gewählt und mithilfe des Baseline-Logits die *Odds* bestimmt werden (Backhaus et al. 2018).

3.2.3 Ordinales logistisches Modell

Da Noten ordinal skaliert sind, können auch diese als Ausprägungen von akademischem Studienerfolg gelten. Im multinomialen logistischen Modell wird eine Ordnung der Kategorien der abhängigen Variablen nicht berücksichtigt, Noten können allerdings jedenfalls geordnet werden und sollten daher – wenn möglich – in einem ordinalen logistischen Modell analysiert werden. Die Variante, welche in dieser Arbeit näher untersucht wird, ist jene der kumulierten Logits – die möglichen Ausprägungen (G Kategorien) der abhängigen Variable Y werden für jede Kategorie g in die zwei Kategorien $Y \leq g$ sowie $Y > g$ eingeteilt (Agresti 2010). Die kumulierten Logits sind dabei für jede der G Kategorien der abhängigen Variablen mit den entsprechenden Wahrscheinlichkeiten $\pi_1, \dots, \pi_G$ wie folgt definiert (Agresti 2010):

$$\text{logit}[P(Y \leq g)] = \log \frac{P(Y \leq g)}{1 - P(Y \leq g)} = \log \frac{\pi_1 + \dots + \pi_g}{\pi_{g+1} + \dots + \pi_G}$$

Dadurch können für eine abhängige Variable mit G Kategorien insgesamt $G-1$ Logits gebildet werden, welche im ordinalen Modell vereint sind (Denham 2017). Es wird also immer die Wahrscheinlichkeit, in eine Kategorie, welche niedriger oder gleich der Kategorie g ist, zu fallen, der Wahrscheinlichkeit, in eine höhere Kategorie als g zu fallen, gegenübergestellt (Agresti 2010). Hierbei werden immer alle Kategorien in die Berechnung des kumulierten Logits miteinbezogen (Denham 2017).

In diesem Modell stellt $P(Y \leq g) = \text{Prob}(Y_k \leq g | x_k)$ dar, wobei g für eine Ausprägung der ordinal skalierten abhängigen Variablen und x_k als Spaltenvektor für die k unabhängigen Variablen steht (Agresti 2010). Aufgrund des kumulativen Ansatzes gilt das Modell auch hier für $g = 1, 2, \dots, G-1$. Das Modell für j erklärende Variablen sieht dabei wie folgt aus:

$$P(Y \leq g) = \frac{e^{\alpha_g + \beta_1 x_1 + \dots + \beta_j x_j}}{1 + e^{\alpha_g + \beta_1 x_1 + \dots + \beta_j x_j}}$$

β beschreibt hier die Effekte der jeweiligen unabhängigen Variablen, wobei diese Effekte für alle kumulativen Logit-Ergebnisse gleich sind (Agresti 2010). Während also im multinomialen Modell β für alle unabhängigen Variablen und hier auch für alle Kategorien der abhängigen Variablen berücksichtigt werden muss, sind im ordinalen Modell die Effekte der unabhängigen Variablen in allen Kategorien der abhängigen Variablen gleich (Hoffmann 2016). Hier ist wichtig zu erwähnen, dass ein ordinale logistisches Modell – wie von Hoffmann (2016) beschrieben – proportionale Odds voraussetzt. Dies bedeutet, dass in allen Kategorien der unabhängigen Variablen die OR konstant (also nicht signifikant unterschiedlich) sind (Hoffmann 2016).

Die Wahrscheinlichkeitswerte für einzelne Beobachtungen und Zellen sind dabei (Agresti 2010), mit $\alpha_0 = -\infty$ und $\alpha_G = \infty$:

$$p_{gj} = P(y_j = g) = \frac{e^{\alpha_g + \beta_1 x_1 + \dots + \beta_j x_j}}{1 + e^{\alpha_g + \beta_1 x_1 + \dots + \beta_j x_j}} - \frac{e^{\alpha_{g-1} + \beta_1 x_1 + \dots + \beta_j x_j}}{1 + e^{\alpha_{g-1} + \beta_1 x_1 + \dots + \beta_j x_j}}$$

Die zu maximierende Log-Likelihood-Funktion für das ordinale logistische Regressionsmodell sieht dabei wie folgt aus (Stata 2021b):

$$LL = \sum_{j=1}^N \sum_{g=1}^G \ln(p_{gj}) * y_{gj}$$

($y_{gj} = 1$ wenn die Kategorie g im Fall k auftritt) (siehe oben). Dabei steht j für jede einzelne Beobachtung und g beschreibt die Kategorien der abhängigen Variablen.

3.2.4 Generalisiertes ordinales logistisches Modell

Zusätzlich zu dem multinomialen und dem ordinalen Modell hat Williams (2006) das sogenannte generalisierte ordinale logistische Modell entwickelt. Dieses ist vor allem dann anzuwenden, wenn die Voraussetzungen des ordinalen logistischen Modells nicht erfüllt sind (zu den Voraussetzungen siehe Kapitel 3.2.5). Der Vorteil des generalisierten ordinalen logistischen Modells im Vergleich zum multinomialen Modell ist vor allem jener, dass die Information über die Ordnung der Kategorien der abhängigen Variablen nicht verloren geht (Williams 2016). Bevor jedoch das generalisierte ordinale logistische Modell zur Anwendung kommt, rät Williams (2016) dazu, die Daten noch auf weitere Gründe für die fehlenden Voraussetzungen zu überprüfen, etwa durch Inklusion eines quadratischen Terms einer unabhängigen Variable.

In diesem Modell stellt $\pi(x_k) = \text{Prob}(Y_k > g)$ dar, wobei g für eine Ausprägung der ordinal skalierten Zufallsvariablen (der abhängigen Variablen) steht (Williams 2006). Für eine ordinal abhängige Variable mit M Kategorien sieht das Modell also wie folgt aus:

$$\pi(x_k) = \frac{e^{\alpha_g + X_k \beta_g}}{1 + e^{\alpha_g + X_k \beta_g}} \text{ wobei } g = 1, 2, \dots, M-1$$

Vergleicht man dieses Modell mit dem ordinalen logistischen Modell, so ist der Unterschied in den Werten der β zu finden – während hier jedes β einen Index aufweist, enthält das ordinale logistische Modell dieselben Werte, die β sind jedoch ohne Index zu finden (Williams 2016). Auch hier stellt X_k einen Zeilenvektor für $(x_{1k}, x_{2k}, \dots, x_{jk})$ mit j unabhängigen Variablen dar. Das ordinale logistische Modell ist daher eine Sonderform des generalisierten ordinalen logistischen Modells, in welchem die β für alle Kategorien gleich sind (daher auch die Voraussetzung der proportionalen OR in allen Kategorien) (Williams 2016). Bei der Interpretation der Ergebnisse müssen dabei auch die unterschiedlichen Kategorien der abhängigen Variablen berücksichtigt werden. Bei M Kategorien gibt es $M-1$ Koeffizienten und Signifikanzwerte, wobei sich diese kumulativ verhalten (Williams 2016). In der ersten Kategorie wird das Ergebnis des Vergleichs von $g = 1$ gegenüber $g = 2, \dots, M$ dargestellt. In der zweiten Kategorie entspricht das Ergebnis einer Gegenüberstellung von $g = 1, 2$ und $g = 3, \dots, M$. Dies setzt sich bis zur letzten Kategorie fort, von welcher die Ergebnisse dargestellt werden, hier erfolgt eine Gegenüberstellung von $g = 1, \dots, M-1$ und $g = M$ (Williams 2016).

Trotz der Vorzüge des generalisierten ordinalen logistischen Modells im Vergleich zum multinomialen Modell hält Williams (2016) fest, dass bei einer Beeinträchtigung der Annahme von

proportionalen OR in den Kategorien der abhängigen Variablen durch viele unabhängige Variablen ein multinomiales Modell sinnvoll sein kann.

3.2.5 Prüfung des Modells

Um die Güte des entwickelten Modells zu überprüfen, kann der Wert der maximierten Log-Likelihood-Funktion verwendet werden (Backhaus et al. 2018). Da dieser immer kleiner als 0 ist, wird der Wert $-2LL$ herangezogen, wobei kleinere Werte eine bessere Anpassung bedeuten. Eine perfekte Anpassung wäre erreicht, wenn $-2LL$ den Wert 0 annimmt (Backhaus et al. 2018).

Der Wert $-2LL$ ist schließlich auch für die Berechnung des Likelihood-Ratio-Tests (LR-Test) notwendig. Die Likelihood-Ratio-Statistik ist nach Backhaus et al. (2018) wie folgt definiert:

$$LLR = -2 * (LL_0 - LL_v)$$

Hier entspricht LL_0 dem maximierten Log-Likelihood-Wert für das Modell, in welchem nur der konstante Term als unabhängige Variable enthalten ist und LL_v dem maximierten Log-Likelihood-Wert für das vollständige Modell, welches alle unabhängigen Variablen enthält (Backhaus et al. 2018). Dieser Wert kann mithilfe einer χ^2 -Verteilung überprüft werden, da davon ausgegangen wird, dass LLR χ^2 -verteilt mit j Freiheitsgraden ist, wobei j der Anzahl an unabhängigen Variablen im Modell entspricht (Backhaus et al. 2018). Wenn nämlich $LLR = \chi^2_{emp} > \chi^2_{theor}$ ist, so kann das Modell als statistisch signifikant gewertet werden und die Nullhypothese, nämlich $\beta_1 = \beta_2 = \dots = \beta_j$ kann verworfen werden (Backhaus et al. 2018). Außerdem kann die Likelihood-Ratio-Statistik auch zur Überprüfung verwendet werden, ob weitere unabhängige Variablen im Modell tatsächlich zu einem Erkenntnisgewinn führen – anstelle von LL_0 wird also das vollständige Modell einfach mit einem kleineren Modell (weniger unabhängige Variablen), LL_r , verglichen.

Im Gegensatz zu dem Bestimmtheitsmaß R^2 der linearen Regression kann für die logistische Regression nur ein Pseudo- R^2 errechnet werden, welches nicht den Anteil von erklärter Streuung an gesamter Streuung im Modell darstellt, sondern das Verhältnis der LL_0 zum LL_v (Backhaus et al. 2018). Hier wird McFadden's Pseudo- R^2 verwendet, welches einen höheren Wert (zwischen 0 und 1) annimmt, je näher LL_v an 0 liegt, was dem maximalen Log-Likelihood-Wert entspricht (Veall und Zimmermann 1994). Zusätzlich zu diesen Prüfungen der Güte des Modells können auch Klassifizierungstabellen verwendet werden (Denham 2017). Die Klassifizierungstabellen können dabei als Indikator für Prognosen des Modells gelten, wobei die Trefferquote den Anteil der korrekt prognostizierten Werte von allen Werten beschreibt; die Sensiti-

vität den Anteil der korrekt prognostizierten positiven Einordnungen (1, zum Beispiel die korrekt prognostizierten bestandenen Prüfungen) und die Spezifität den Anteil der korrekt prognostizierten negativen Einordnungen (0, zum Beispiel die korrekt prognostizierten nicht bestandenen Prüfungen) (Backhaus et al. 2018). Für die Erstellung der Klassifizierungstabelle muss allerdings zuvor noch ein Trennwert festgelegt werden, welcher beschreibt, ab welchem Wert des y_k in die Alternativen 0 oder 1 eingeteilt wird (wobei zumeist der Trennwert mit 0.5 festgesetzt wird). Die ROC-Kurve (Receiver-Operating-Characteristic-Kurve) bildet dabei eine Verallgemeinerung der Klassifizierungstabelle über unterschiedliche Trennwerte hinweg und gilt nach Christensen et al. (2014) als stabil (Backhaus et al. 2018).

Wie bei der linearen Regression werden auch bei der logistischen Regression die Residuen des Modells überprüft. Diese werden für ein logistisches Regressionsmodell als $e_k = y_k - p_k$ dargestellt, also als Differenz von beobachteten Werten und geschätzten Wahrscheinlichkeiten (Backhaus et al. 2018). Die Residuen sind daher im Intervall $[-1, 1]$ zu finden, wobei die Summe über alle Residuen 0 ergibt (Backhaus et al. 2018). Dies kann auch hilfreich sein, um Ausreißer im Modell zu identifizieren und gegebenenfalls zu exkludieren – durch Standardisieren der Residuen können Ausreißer leicht festgestellt (Wert > 2) und anschließend das Modell ohne diese Ausreißer nochmals gerechnet werden (Denham 2017).

Weiters werden die unabhängigen Variablen des Modells auf Multikollinearität überprüft. Dies würde bedeuten, dass sich die Streuungen der unabhängigen Variablen überschneiden und somit keine eindeutige Zuordnung der Information zu den Variablen möglich ist (Backhaus et al. 2018). Das Problem dabei ist, dass die unabhängigen Variablen eines Modells stark miteinander korrelieren, also voneinander abhängig sind (Komlos und Süssmuth 2010). Dies würde eine korrekte Berechnung der Koeffizienten eines Modells verunmöglichen und sollte daher ausgeschlossen werden. Um Multikollinearität zu entdecken, wird das Konstrukt des Varianzinflationsfaktors (VIF) verwendet – dieser kann für jede der unabhängigen Variablen separat berechnet werden. Je höher der Varianzinflationsfaktor ausfällt, desto stärker ist Multikollinearität in dem Modell vorhanden – Komlos und Süssmuth (2010) geben dabei als Daumenregel den Wert 4 an, Backhaus et al. (2018) nennen sowohl 5 als auch 10 als mögliche Grenzwerte. Alles unter diesen Grenzwerten deutet nicht auf Multikollinearität hin, darüber muss eine solche jedoch bereits angenommen werden (Komlos und Süssmuth 2010).

Zusätzlich werden die Werte des Akaike'schen Informationskriteriums (AIC) und des Bayes'schen Informationskriteriums (BIC) für die jeweiligen Modelle berechnet. Hiermit können die entwickelten Modelle untereinander verglichen und zwischen diesen das beste Modell ausgewählt werden (Backhaus et al. 2018). Dabei wird zusätzliche Modellkomplexität bestraft

– die Modelle mit den kleinsten Werten entsprechen daher den besseren Modellen (Backhaus et al. 2018).

Zusätzlich zu diesen Gütekriterien existieren weitere Kriterien, welche explizit ordinale beziehungsweise multinomiale logistische Regressionsmodelle überprüfen können. Die Devianz vergleicht das zu prüfende Modell mit dem bestmöglichen Modell, auch saturiertes Modell genannt. Dies stellt auch den größten Gegensatz zur LR-Statistik dar, welche das zu prüfende Modell mit dem schlechtesten Modell (nur der konstante Faktor als erklärende Variable) vergleicht (Backhaus et al. 2018). Das bestmögliche Modell beschreibt dabei perfekte Anpassung, da es für jede Beobachtung Parameter besitzt und somit im Vergleich zur Realität nichts vereinfacht (und daher auch nicht als tatsächliches Modell angesehen werden kann) (Backhaus et al. 2018). Die Formel für die Devianz lautet $D = -2(LL - LL_s)$, wobei kleinere Werte für eine bessere Anpassung des zu prüfenden Modells stehen (Backhaus et al. 2018). Bei der Verwendung eines ordinalen logistischen Regressionsmodells muss auch die Annahme der proportionalen Odds überprüft werden – dazu kann ein LR-Test durchgeführt werden, welcher überprüft, ob die OR in allen Kategorien gleich sind (Hoffmann 2016). Die Nullhypothese, die besagt, dass die OR in allen Kategorien gleich sind, sollte hier nicht verworfen werden – das Zielbild wäre also ein p-Wert größer als das Signifikanzniveau von 5% (Hoffmann 2016).

Schließlich kann noch die Güte der einzelnen unabhängigen Variablen überprüft werden, wobei hierzu die p-Werte der jeweiligen Koeffizienten interpretiert werden. Die Annahme im Modell dieser Arbeit ist, dass die p-Werte anhand von $\alpha = 5\%$ beurteilt werden, sodass p-Werte kleiner als 0,05 als statistisch signifikant angesehen werden.

3.3 Analyse

Manuell erstelltes Datenfeld	Beschreibung
Jahr	Das Jahr, in welchem die Prüfung abgelegt wurde. Wird jeweils aus der Semesterinformation gewonnen (2013S und 2013W zählen zum Jahr 2013).
Erfolg	Ein binärer Indikator, welcher angibt, ob die Prüfung positiv bestanden wurde. „Nicht genügend“ und „ohne Erfolg teilgenommen“ führen zum Wert 0, alle anderen zum Wert 1.
Corona_Indikator	Ein binärer Indikator, welcher angibt, ob eine Prüfung vor der Corona-Pandemie (2013S bis 2019W) abgelegt wurde (Wert 0) oder während der Corona-Pandemie (Wert 1, 2020S und 2020W).
Wahl_Pflicht	Diese Variable nimmt drei Ausprägungen an: „Pflicht“, „Wahlfach“ oder „Wahlpflichtfach“.

Tabelle 6: Manuell erstellte Datenfelder

Zusätzlich zu den bereits vorhandenen Datenfeldern wurden vier weitere Datenfelder erstellt. Diese sind mit Beschreibung in Tabelle 6 angeführt und sollen die Analyse der Stichprobe erleichtern.

Die Prüfungsergebnisse werden als ordinalskalierte Daten behandelt: Sie können sinnvoll in eine Ordnung gebracht werden, Werte wie Standardabweichung oder Mittelwert können jedoch nicht gebildet werden (Denham 2017). Dies liegt daran, dass Abstände zwischen unterschiedlichen Notenwerten nicht als gleich betrachtet werden können (so kann etwa nicht beurteilt werden, ob ein *Gut* gleich nahe an einem *Sehr Gut* liegt wie ein *Genügend* an einem *Nicht Genügend*). Aufgrund der Skalierung der abhängigen Variablen wird eine binäre logistische Regression sowie eine ordinale logistische Regression durchgeführt und beschrieben.

3.3.1 Auswahl der unabhängigen Variablen im Modell

In einer Erstanalyse des vorhandenen Datensatzes kann festgestellt werden, dass das Studium (Bachelor- oder Masterstudium) einen großen Einfluss auf die jeweilige abhängige Variable aufweist. Daher werden die Modelle für Bachelor- und Masterstudierende getrennt berechnet, um genauere Aussagen treffen zu können. Zur Analyse der Auswahl der unabhängigen Variablen werden dennoch in einem ersten Schritt alle verfügbaren erklärenden Variablen des gesamten Datensatzes miteinbezogen. Die Auswahl berücksichtigt dann die Variable *Studium* nicht weiter, da anhand dieser der Datensatz bereits separiert betrachtet wird.

Zur Festlegung der unabhängigen Variablen im Modell wurde zuerst ein χ^2 -Test für die Zusammenhänge zwischen der binären Variablen *Erfolg* beziehungsweise der ordinal skalierten Variable *Noten* und weiteren vorhandenen Daten durchgeführt. Der χ^2 -Test wurde sowohl für den Datensatz bis Sommersemester 2020 (Grundgesamtheit 5.658 Prüfungen, davon 5.141 mit einer Beurteilung von *Sehr Gut* bis *Nicht Genügend*) als auch für den erweiterten Datensatz mit den Daten bis Wintersemester 2020/21 durchgeführt (Grundgesamtheit 6.175 Prüfungen im Bachelor- und Masterstudium, davon 5.685 Prüfungen mit einer Beurteilung von *Sehr Gut* bis *Nicht Genügend*). Die Resultate sind in Tabelle 7 dargestellt.

Der Einfluss der Corona-Pandemie auf Studierendenleistungen an der Fakultät für Wirtschaftswissenschaften der Universität Wien

Variable 1	Variable 2	p-Wert	p-Wert (inkl. WS20)	Anmerkung
Noten (1-5)	Semester (2013S-2020W)	<0,001**	<0,001**	
Noten (1-5)	Corona_Indikator (0/1)	<0,001**	<0,001**	
Noten (1-5)	Geschlecht (w/m)	<0,05*	>0,05	
Noten (1-5)	Studium (Bachelor/Master)	<0,001**	<0,001**	
Noten (1-5)	Studiumsart (Anderes, Lehramt, Politik-, Sprach-, Wirtschaftswissenschaften)	<0,001**	<0,001**	Aufgrund weniger Einträge ungenau
Noten (1-5)	Alterskategorie (<25, 25-35, 35-45, 45-55, 55-65, 65- 100)	<0,001**	<0,001**	Aufgrund weniger Einträge ungenau
Noten (1-5)	Prüfungstyp (Bachelorseminar, BWL, Buchhaltung, FIS, Kooperatives Wirt- schaften, Masterarbeitskonversato- rium, Rechnungswesen, Wirtschafts- kunde, Solidarische Ökonomie)	<0,001**	<0,001**	Aufgrund weniger Einträge ungenau
Noten (1-5)	Jahr	<0,001**	<0,001**	
Noten (1-5)	Wahl_Pflicht (Pflichtfach, Wahlfach, Wahlpflichtfach)	<0,001**	<0,001**	
Erfolg (0/1)	Semester (2013S-2020S)	<0,001**	<0,001**	
Erfolg (0/1)	Corona_Indikator (0/1)	<0,001**	<0,001**	
Erfolg (0/1)	Geschlecht (w/m)	<0,05*	<0,05*	
Erfolg (0/1)	Studium (Bachelor/Master)	<0,001**	<0,001**	
Erfolg (0/1)	Studiumsart (Anderes, Lehramt, Politik-, Sprach-, Wirtschaftswissenschaften)	<0,001**	<0,001**	
Erfolg (0/1)	Alterskategorie (<25, 25-35, 35-45, 45-55, 55-65, 65- 100)	<0,001**	<0,001**	Aufgrund weniger Einträge ungenau
Erfolg (0/1)	Prüfungstyp (Bachelorseminar, BWL, Buchhaltung, FIS, Kooperatives Wirt- schaften, Masterarbeitskonversato- rium, Rechnungswesen, Wirtschafts- kunde, Solidarische Ökonomie)	<0,001**	<0,001**	Aufgrund weniger Einträge ungenau
Erfolg (0/1)	Jahr	<0,001**	<0,001**	
Erfolg (0/1)	Wahl_Pflicht (Pflichtfach, Wahlfach, Wahlpflichtfach)	<0,001**	<0,001**	

Tabelle 7: Ergebnisse χ^2 -Tests

Wie in Tabelle 7 ersichtlich, besteht sowohl zwischen den *Noten* in ordinal skaliert Form als auch der binären Variable *Erfolg* und den infrage kommenden unabhängigen Variablen immer ein signifikanter Zusammenhang (größtenteils sogar auf dem 1%-Niveau signifikant). Hier muss festgehalten werden, dass vor allem die Zusammenhänge zu den Variablen *Alterskategorie* sowie *Prüfungstyp* und auch der *Art des Studiums* ungenau sein können, da in gewissen Grup-

pen nur eine geringe Anzahl an Beobachtungen registriert werden kann. Die Variable *Geschlecht* weist einen weniger stark signifikanten Zusammenhang zu *Erfolg* und den *Noten* auf, nichtsdestoweniger ist der Zusammenhang auf dem 5%-Level signifikant. Die Variable *Alter* liegt in diskreter Form vor und wird in dieser für das Modell herangezogen. Um eine Beeinflussung der Variablen untereinander zu vermeiden, wurde eine Korrelationsmatrix erstellt – diese ist in Abbildung 4 zu sehen.

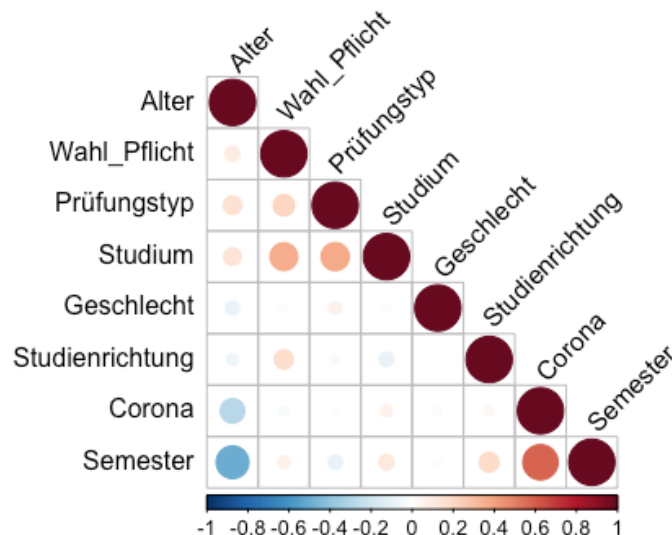


Abbildung 4: Korrelationsmatrix potenziell unabhängiger Variablen

Als Basisdaten für die Korrelationsmatrix wurde der gesamte Datensatz herangezogen – also 6.175 Erstantritte zu Prüfungen von Studierenden im Bachelor- oder Masterstudium. Hier ist klar ersichtlich, dass zwischen den Variablen nur ein geringer Einfluss besteht. Die Variable *Semester* hängt hingegen stark mit der Variable *Corona* zusammen (was nicht verwunderlich ist, da die Indikatorvariable für die Corona-Pandemie aus der Variable *Semester* berechnet wurde) – auch der p-Wert des Zusammenhangs ist signifikant ($p\text{-Wert} < 0,001$). Weiters kann ein Zusammenhang zwischen den Variablen *Studium* (in den Ausprägungen Bachelorstudium und Masterstudium) sowie *Studienrichtung* (in den Ausprägungen Wirtschafts-, Politik- und Sprachwissenschaften, Lehramt und Anderes) und auch *Studium* sowie *Wahl_Pflicht* festgestellt werden (da die Unterscheidung in Wahl-, Pflicht und Wahlpflichtfächer nach Studium und auch Studienrichtung variiert). Auch der *Prüfungstyp* (in den Ausprägungen Bachelorseminar, Wirtschaftslehre, Buchhaltung, Führungsinformationssysteme, Kooperatives Wirtschaften und Kooperationsmanagement, Masterarbeitskonversatorium, Mentoring & Wegweiser, Rechnungswesen, Seminar aus Wirtschaftskunde, Servicemodul und Solidarische Ökonomie) korreliert stark mit der Variable *Wahl_Pflicht* und dem *Studium*. Die Variablen *Prüfungstyp* und *Semester* werden daher im Modell vorerst nicht inkludiert, um Multikollinearität auszuschließen. Die gewählten unabhängigen Variablen für das Modell sind also *Corona_Indikator*, *Alter*,

Wahl_Pflicht, *Studienrichtung* und *Geschlecht*. Wie bereits zuvor erwähnt, wird auch die Variable *Studium* nicht berücksichtigt, da der Datensatz getrennt für Bachelor- und Masterstudierende betrachtet wird.

3.3.2 Binäre logistische Regression

Im binären Logit-Regressionsmodell wird die Variable *Erfolg* als abhängige Variable herangezogen, die Variablen *Corona_Indikator*, *Alter*, *Wahl_Pflicht*, *Studienrichtung* und *Geschlecht* sind unabhängige Variablen in dem Modell. Die Daten verteilen sich über die Modelle wie folgt:

Modell	Studium	Semester	# Daten
Binär	Bachelorstudierende	bis SS 20	5.188 Resultate 4.628 Studierende
		bis WS 20/21	5.646 Resultate 5.052 Studierende
	Masterstudierende	bis SS 20	470 Resultate 422 Studierende
		bis WS 20/21	529 Resultate 478 Studierende

Tabelle 8: Übersicht Datensätze binäre Modelle

Die Regressionsgleichung sieht dementsprechend wie folgt aus:

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + e_i$$

$$Erfolg_i = \beta_0 + \beta_1 Corona_Indikator + \beta_2 Alter + \beta_3 Geschlecht + \beta_4 Wahl_Pflicht + \beta_5 Studienrichtung + e_i \quad (1)$$

Erfolg steht also für die abhängige Variable (Y_i), β_0 stellt die Konstante des Modells dar, β_a ($a > 0$) ist der Koeffizient der jeweiligen unabhängigen Variablen und e_i fungiert als Maß für den Fehlerterm der Regressionsgleichung. In Stata wird diese Regression mit dem folgenden Befehl ausgeführt: *logit Erfolg i.corona i.geschlecht Alter i.wahlpflicht i.studienrichtung, robust cluster(RND)*

Wie bereits angeführt sind die vorhergesagten Werte dieses Modells die z-Werte der Funktion. Um also interpretierbare Werte zu erlangen, muss noch der Befehl *margins, dydx(*)* ausgeführt werden. Die vorhergesagten Werte des Modells werden mittels *predict* gespeichert. Der letzte Teil des Befehls, *cluster(RND)*, führt zu geclusterten Standardfehlern pro Studierendem (Angrist und Pischke 2008). Die Clusterung ist notwendig, da die Werte von 5.658 Prüfungen (beziehungsweise 6.175 Prüfungen im Datensatz bis WS 2020/21), aber nur von 4.940 Studierenden (5.715 Studierende bis WS 2020/21) vorhanden sind, was darauf hinweist, dass manche Studierende mehrere Prüfungen absolviert haben. Um personenbezogene Effekte zu berücksichtigen, wird nach diesen gruppiert (Angrist und Pischke 2008). Mit dem Befehl *estat*

classification werden anschließend auch die Klassifizierungstabelle sowie damit verbundene Statistiken für das jeweilige Modell ausgegeben.

Imbens und Lemieux (2008) betonen die Wichtigkeit der Visualisierung während der Durchführung von Regressionsanalysen. Daher wird auch in diesem Modell der Zusammenhang zwischen den unabhängigen Variablen und der abhängigen Variable auf potenzielle nichtlineare Zusammenhänge sowohl im Datensatz der Bachelorstudierenden als auch in jenem der

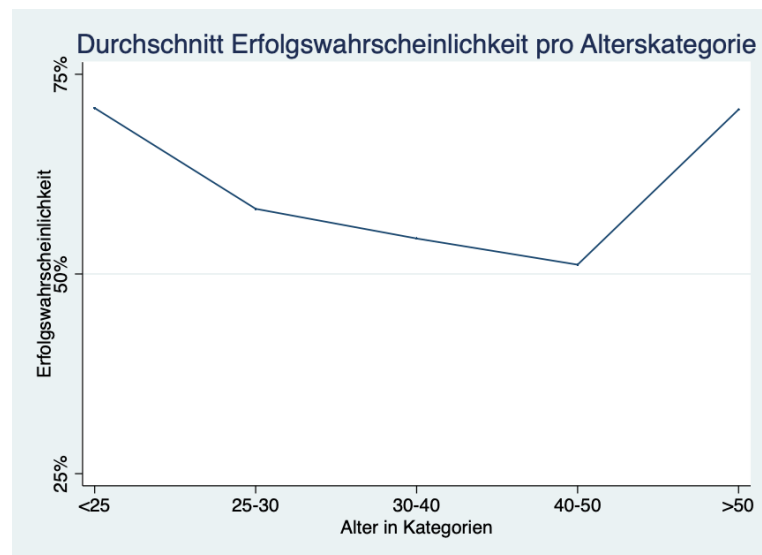


Abbildung 5: Durchschnittliche Erfolgswahrscheinlichkeit in den Alterskategorien, Bachelorstudium

Masterstudierenden untersucht. Da alle verwendeten unabhängigen Variablen bis auf die Variable *Alter* binär skaliert sind, wird nur diese auf potenzielle nichtlineare Zusammenhänge überprüft. Aufgrund der geringen Anzahl an Beobachtungen bei steigendem Alter werden die Beobachtungen in Alterskategorien gruppiert. Die Gruppierung unterscheidet sich hier nach Bachelor- und Masterstudierenden. Da die Anzahl an Datenpunkten bei Bachelorstudierenden in der Spanne von 20 bis 30 Jahren am höchsten ist, wird diese Spanne in zwei Kategorien unterteilt (unter 25 Jahre und 25 bis 30 Jahre), die restlichen Daten werden in Zehnjahresschritten dargestellt und alle Beobachtungen über 50 Jahre werden in einer Kategorie zusammengefasst. In Abbildung 5 ist ersichtlich, dass sich die Erfolgswahrscheinlichkeit (der Durchschnitt aus der Variable Erfolg * 100 pro Kategorie) in den unterschiedlichen Kategorien jedenfalls nichtlinear verhält – während die Erfolgswahrscheinlichkeit in der Alterskategorie der Unter-25-Jährigen sehr hoch ist, sinkt diese bis zur Altersgruppe der 40- bis 50-Jährigen ab und steigt für die Kategorie der Über-50-Jährigen wieder an. Für Masterstudierende ergibt sich eine Häufung der Datenpunkte in der Alterskategorie der 25- bis 35-Jährigen mit nur wenigen Beobachtungen von Personen über 35 Jahren. Die Alterskategorien werden daher definiert als jünger als 25 Jahre, 25 bis 30 Jahre, 30 bis 35 Jahre und älter als 35 Jahre. In Abbildung 6

ist dabei ersichtlich, dass es in den Daten der Masterstudierenden keine offensichtliche nicht-lineare Beziehung zu geben scheint.

Um die potenzielle Nichtlinearität in Form des quadratischen Anstiegs der Erfolgswahrschein-

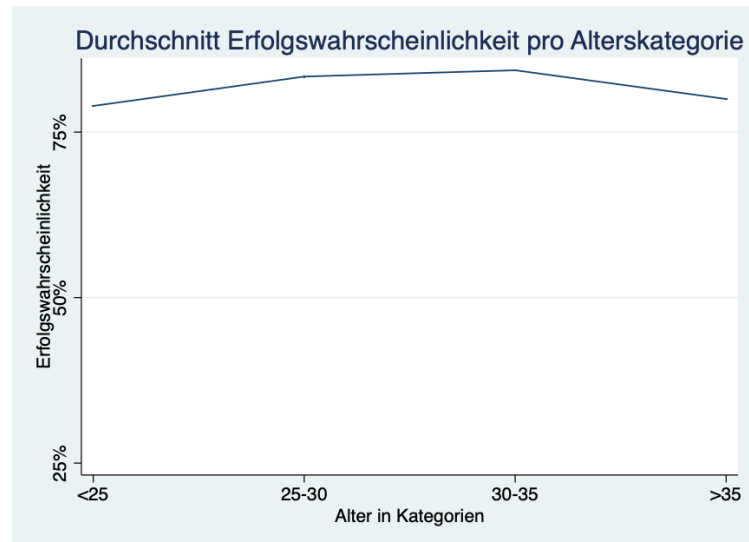


Abbildung 6: Durchschnittliche Erfolgswahrscheinlichkeit in den Alterskategorien, Masterstudium
lichkeit im Modell für Bachelorstudierende zu berücksichtigen, wird in diesem Modell die Variable $Alter^2$ eingeführt. Um die Interpretation zu erleichtern, wird das Alter zentriert (von jedem Wert der Variable $Alter$ wird der Mittelwert dieser Variablen abgezogen) und der zentrierte Wert wird anschließend quadriert. Dies wird in Stata mit folgenden Befehlen erreicht:

```
generate alter_cen = Alter-26.79
```

```
generate alter2 = alter_cen^2
```

Hier entspricht $Alter$ der Variablen und 26,79 stellt den Mittelwert derselben dar. Dadurch werden in der Variablen $alter_cen$ die zentrierten Werte gespeichert, welche wiederum für die Variable $alter2$ quadriert werden können (Angrist und Pischke 2008). Nach Backhaus et al. (2018) kann das Quadrieren der Variable deshalb durchgeführt werden, da das lineare Regressionsmodell Linearität in der Beziehung in den Parametern erfordert und eine unabhängige Variable X durch das Pendant $X' = f(X)$ ersetzt werden kann, wobei f eine nichtlineare Funktion darstellt. Die Regressionsgleichung sieht dementsprechend wie folgt aus:

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_2^2 + \beta_4 X_3 + \beta_5 X_4 + e_i$$

$$Erfolg_i = \beta_0 + \beta_1 Corona_Indikator + \beta_2 Alter_zentriert + \beta_3 Alter^2 + \beta_4 Geschlecht + \quad (2)$$

$$\beta_5 \text{Wahl_Pflicht} + \beta_6 \text{Studienrichtung} + e_i$$

Wie zuvor stellt *Erfolg* die abhängige Variable dar, im Gegensatz zum ersten Modell liegt die Variable *Alter* nun allerdings in zentrierter Form (der Mittelwert wurde von allen Werten subtrahiert) und in quadrierter Form vor. β_0 stellt die Konstante des Modells dar, β_a ($a > 0$) ist der Koeffizient der jeweiligen unabhängigen Variablen und e_i fungiert als Maß für den Fehlerterm der Regressionsgleichung. In Stata wird diese Regression mit dem folgenden Befehl ausgeführt: `logit Erfolg i.corona i.geschlecht alter_cen alter2 i.wahlpflicht i.studienrichtung, robust cluster(RND)`

Um interpretierbare Werte zu erlangen, wird auch hier der Befehl `margins, dydx(*)` ausgeführt.

Im Anschluss an die jeweilige Modellberechnung wird die Güte des Modells anhand der in Kapitel 3.2.5 diskutierten Indikatoren überprüft. Die Log-Likelihood-Werte der jeweiligen Modelle werden dabei von Stata am Beginn jedes errechneten Modells angezeigt und können somit sofort interpretiert werden. Außerdem wird für jedes Modell ein Wald- χ^2 -Test durchgeführt und der Wert sowie der dazugehörige p-Wert angezeigt, anhand dessen beurteilt werden kann, ob das Modell eine statistische Aussagekraft aufweist (Denham 2017). Auch der Pseudo- R^2 -Wert wird bei jeder Berechnung angezeigt, wobei Stata hier die Berechnungsmethode von McFaddens R^2 nutzt (Williams 2020). Zur Überprüfung der Residuen können diese mit dem Befehl `predict stdres, rstand` erzeugt werden und anschließend in einem Scatterplot dargestellt werden. Da die Residuen bereits standardisiert sind, lassen sich Ausreißer (kleiner -2 oder größer 2) leicht feststellen und können gegebenenfalls ausgeschlossen werden. Weiters kann ein goodness-of-fit Test des Modells mit dem Befehl `lfit, group(10) table` ausgeführt werden (UCLA 2021a). Mittels `fitstat` werden weitere wichtige Modellgrößen angezeigt, diese werden im Kapitel 4 mit den jeweiligen Ergebnissen diskutiert. Die Klassifikationstabelle kann außerdem mit dem Befehl `estat classification` aufgerufen und weiter interpretiert werden. Um das Modell auf Multikollinearität überprüfen zu können, wird es nochmals gerechnet. Trotz der binären abhängigen Variablen wird zur Berechnung des Varianzinflationsfaktors der Befehl

`regress Erfolg i.corona i.geschlecht Alter i.studium, robust cluster(RND)`

sowie direkt im Anschluss der Befehl `vif` ausgeführt. Die Verwendung von `regress` anstelle von `logit` kann für die Berechnung des Varianzinflationsfaktors deshalb erfolgen, da für die Bestimmung der Multikollinearität nur die unabhängigen Variablen betrachtet werden.

3.3.3 Ordinale logistische Regression

Im ordinalen Logit-Modell wird nicht nur das Bestehen einer Prüfung als Erfolgsindikator herangezogen, sondern es werden auch die Noten als ordinale Daten berücksichtigt. Dazu müssen jene Beobachtungen aus der Stichprobe ausgeschlossen werden, welche nur die Beurteilungen „mit Erfolg teilgenommen“ und „ohne Erfolg teilgenommen“ beinhalten. Damit bleiben folgende Daten übrig:

Modell	Studium	Semester	# Daten
Ordinal	Bachelorstudierende	bis SS 20	4,793 Resultate 4,517 Studierende
		bis WS 20/21	5,251 Resultate 4,944 Studierende
	Masterstudierende	bis SS 20	348 Resultate 324 Studierende
		bis WS 20/21	407 Resultate 380 Studierende

Tabelle 9: Übersicht Datensätze ordinale Modelle

Die Regressionsgleichung sieht daher wie folgt aus:

$$Note_i = \beta_0 + \beta_1 Corona_Indikator + \beta_2 Alter + \beta_3 Geschlecht + \beta_4 Wahl_Pflicht + \beta_5 Studienrichtung + e_i \quad (3)$$

Wie auch im binären logistischen Regressionsmodell wird zuerst eine Regression durchgeführt, in welcher potenzielle nichtlineare Effekte nicht berücksichtigt werden. Aufgrund der Tatsache, dass sich die Werte in eine Ordnung bringen lassen (von *Sehr Gut* bis *Nicht Genügend*) und mehr als zwei Ausprägungen haben, muss der Befehl *ologit* in Stata verwendet werden. Die Befehle *logit* und *mlogit* sind bei unterschiedlichen Szenarien anzuwenden, nämlich einerseits bei binärer Ausprägung der abhängigen Variablen und andererseits bei Ausprägungen, die nicht in eine Ordnung gebracht werden können (Stata 2021b). Dabei wurde die Anwendung des Befehls *logit* in Kapitel 3.3.2 besprochen und die Anwendung des Befehls *mlogit* wird in Kapitel 3.3.4 ausgeführt.

ologit Spalte1 i.corona i.geschlecht Alter i.studium, robust cluster(RND)

Im Vergleich zur binären logistischen Regression unterscheidet sich einerseits der Befehl *ologit* von jenem der binären Regression (*logit*) und die unabhängige Variable ändert sich. Das Resultat dieses Befehls sind die geschätzten Koeffizienten. Durch den Befehl

ologit Spalte1 i.corona i.geschlecht Alter i.studium, robust cluster(RND) or

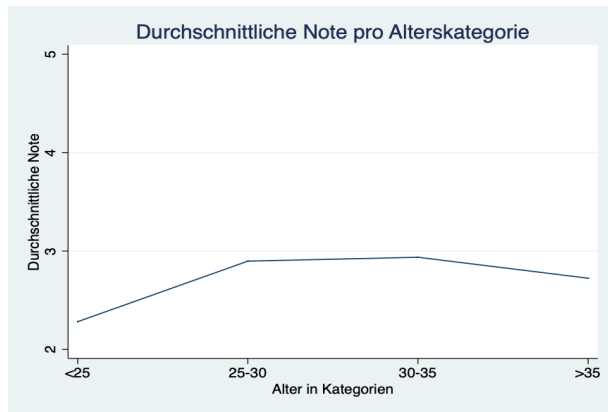


Abbildung 7: Durchschnittliche Note pro Alterskategorie, Masterstudium

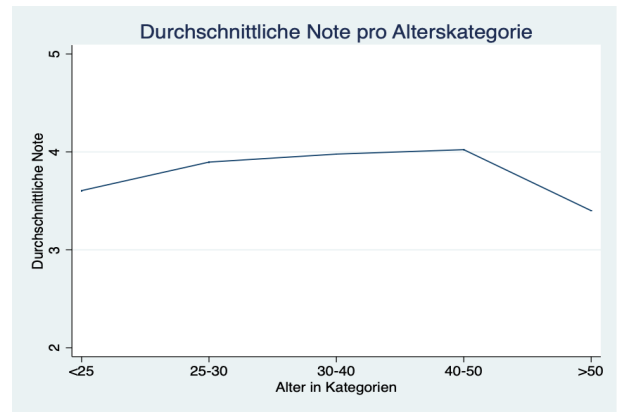


Abbildung 8: Durchschnittliche Note pro Alterskategorie, Bachelorstudium

werden nicht die logarithmierten *Odds* angegeben, sondern die OR ausgegeben. Dies entspricht also dem e^b anstatt des b (siehe Kapitel 3.2.1) und erleichtert dadurch die Interpretation. Der einzige Unterschied ist die Darstellung der Ergebnisse, die Berechnung und Schätzung unterscheidet sich nicht (Stata 2021b). Die OR werden dabei immer relativ zu der Referenzkategorie interpretiert. Die Interpretation der OR wird bei den Resultaten vorgenommen. Auch in diesem Modell wird die Variable *Alter* im Gesamtdatensatz auf potenzielle Nichtlinearität untersucht und die durchschnittliche Note pro Alterskategorie in Abbildung 8 (Bachelorstudierende) sowie in Abbildung 7 (Masterstudierende) dargestellt. Während sich die durchschnittlichen Noten bei Bachelorstudierenden zwischen *Befriedigend* und *Genügend* und bei Masterstudierenden zwischen *Gut* und *Befriedigend* bewegen, ist ersichtlich, dass der Notendurchschnitt in den mittleren Alterskategorien ansteigt und in der ältesten Alterskategorie wieder sinkt. Trotz unterschiedlicher Alterskategorien ist der grundsätzliche Trend für Master- und Bachelorstudierende hier gleich. Daher wird ein potenzieller quadratischer Effekt des Alters auf die Noten vermutet und in beide statistische Modelle (sowohl für Bachelorstudierende als auch für Masterstudierende) entsprechend inkludiert. Da die Werte allerdings im Vergleich zur binären logistischen Analyse verringert sind, wird auch der Altersdurchschnitt für das Zentrieren der Variablen nochmals je Datensatz neu berechnet. Um die Interpretation zu erleichtern, wird das Alter zentriert (von jedem Wert der Variable *Alter* wird der Mittelwert dieser Variablen abgezogen) und der zentrierte Wert wird anschließend quadriert. Dies wird in Stata mit folgenden Befehlen erreicht:

```
generate alter_cen_ordinal = Alter-26.60
```

```
generate alter_ordinal2 = alter_cen_ordinal^2
```

Hier entspricht *Alter* der Variablen und 26,60 stellt den Mittelwert für Bachelorstudierende derselben dar (für Masterstudierende wird analog vorgegangen). Dadurch werden in der Variab-

len *alter_cen_ordinal* die zentrierten Werte für das ordinale Modell gespeichert, welche wiederum für die Variable *alter_ordinal2* quadriert werden können (Angrist und Pischke 2008). Das Modell wird daher um die quadrierte Altersvariable erweitert:

```
ologit Spalte1 i.corona i.geschlecht alter_cen_ordinal alter_ordinal2 i.studienrichtung i.wahl-  
pflicht, robust cluster(RND)
```

beziehungsweise

```
ologit Spalte1 i.corona i.geschlecht alter_cen_ordinal alter_ordinal2 i.studienrichtung i.wahl-  
pflicht, robust cluster(RND) or
```

Die Regressionsgleichung für das erweiterte Modell ist daher

$$Note_i = \beta_0 + \beta_1 Corona_Indikator + \beta_2 Alter_zentriert + \beta_3 Alter^2 + \beta_4 Geschlecht + \beta_5 Studienrichtung + \beta_6 Wahl_Pflicht + e_i \quad (4)$$

Ähnlich wie beim binären logistischen Modell wird die Güte des Modells im Anschluss an die Berechnung anhand der in Kapitel 3.2.5 diskutierten Indikatoren überprüft. Auch für die ordinale logistische Regression gilt, dass die Log-Likelihood-Werte sowie der Wert und p-Wert des Wald- χ^2 -Tests und Pseudo- R^2 -Wert für jedes Modell direkt von Stata angezeigt werden. Nun können allerdings die Residuen nicht mehr mittels *predict* erzeugt werden, es muss daher auf andere Mittel der Überprüfung des Modells zurückgegriffen werden (Stata 2021a). Es ist weiterhin möglich, wichtige Modellgrößen mittels *fitstat* anzeigen zu lassen, welche später diskutiert werden.

Mittels *margins*, *predict(outcome(1)) atmeans post* – und weiteren angepassten Befehlen – kann die Wahrscheinlichkeit angezeigt werden, eine bestimmte Note (in diesem Fall ein *Sehr Gut*, da beim outcome explizit 1 angegeben wurde) zu erreichen, während die anderen Variablen mit ihren Mittelwerten angenommen werden. Weiters muss in diesem ordinalen Modell nicht nochmals die Multikollinearität von den unabhängigen Variablen überprüft werden, da diese jenen des binären logistischen Modells entsprechen.

Zusätzlich muss im Rahmen der Berechnung des ordinalen logistischen Modells die Grundannahme der proportionalen OR in den unterschiedlichen Kategorien überprüft werden (Hoffmann 2016). Dies erfolgt durch die Ausführung des *ologit* Befehls und anschließender Überprüfung durch den Befehl *oparallel*. Bei Entdeckung einer Verletzung der Annahme kann weiter überprüft werden, ob sich die OR nur in den Kategorien einer unabhängigen Variablen nicht proportional verhalten oder ob dies mehrere unabhängige Variablen des Modells betrifft

(Hoffmann 2016). Dazu wird die Regression wie oben beschrieben mit dem *ologit*-Befehl erneut ausgeführt und direkt danach der Befehl *brant*. Hier wird die Annahme der proportionalen OR für jede unabhängige Variable im Modell getestet und angezeigt. Mithilfe dieser Tests können die manuelle Berechnung und Überprüfung der OR pro Kategorie für jede unabhängige Variable erspart bleiben (Hoffmann 2016).

Nach Williams (2016) ist es keine Seltenheit, dass die oben genannten Annahmen verletzt sind und die Modelle damit an Aussagekraft verlieren. Ist dies der Fall, gibt es zwei Möglichkeiten, um auch verletzte Annahmen im Modell zu berücksichtigen. Einerseits kann ein multinomiales Modell – wie auch in Kapitel 3.3.4 beschrieben – gerechnet werden, andererseits gibt es die Möglichkeit, auf den Befehl *gologit2* (Williams 2006) auszuweichen. Dieser Befehl führt ein generalisiertes ordinales logistisches Regressionsmodell aus und erlaubt abweichende Koeffizienten, wenn die Annahme der proportionalen OR für die unabhängige Variable verletzt ist (Hoffmann 2016). Im Gegensatz zum multinomialen Modell bietet das generalisierte ordinale logistische Regressionsmodell dabei auch einen Mehrwert an Information, da es die Ordnungsstruktur der abhängigen Variablen berücksichtigt – eine Information, die im multinomialen Modell verloren geht (Williams 2016). Es wird daher noch zusätzlich zu den oben angeführten Modellen (3) und (4) ein weiteres Modell (5) erstellt. Die Regressionsgleichung gleicht jener von Modell (4).

$$\text{Note}_i = \beta_0 + \beta_1 \text{Corona_Indikator} + \beta_2 \text{Alter_zentriert} + \beta_3 \text{Alter}^2 + \beta_4 \text{Geschlecht} + \beta_5 \text{Studienrichtung} + \beta_6 \text{Wahl_Pflicht} + e_i \quad (5)$$

Dieses Modell wird allerdings nun mit folgendem Befehl ausgeführt:

```
gologit2 Spalte1 i.corona i.geschlecht alter_cen_ordinal alter_ordinal2 i.studienrichtung  
i.wahlpflicht, robust cluster(RND) or autofit
```

Der Vorteil dieser Vorgehensweise ist dabei, dass das Modell nun für jede unabhängige Variable überprüft, ob die Voraussetzung der proportionalen OR in den unterschiedlichen Kategorien der abhängigen Variablen erfüllt und gegebenenfalls das Modell anzupassen ist (Hoffmann 2016).

3.3.4 Multinomiale logistische Regression

Im multinomialen logistischen Modell wird davon ausgegangen, dass die abhängige Variable keine Ordnungsstruktur aufweist (Williams 2016). Dies ist bei der abhängigen Variablen in dieser Analyse zwar nicht der Fall, die multinomiale logistische Regression kann allerdings auch dann eingesetzt werden, wenn das ordinale logistische Modell die Voraussetzungen wie etwa proportionale OR in den Kategorien bei mehreren unabhängigen Variablen nicht erfüllt

(Williams 2016). Die Vorgehensweise bleibt ähnlich zu der ordinalen logistischen Regression – die abhängige Variable setzt sich aus den Noten der Studierenden bei den Prüfungen zusammen und auch die unabhängigen Variablen bleiben gleich wie im ordinalen logistischen Modell.

Daher wird dieses Modell anhand der Variablen in Modell (5) ausgeführt. Das Resultat dieses Befehls *mlogit* sind die geschätzten Koeffizienten. Im Gegensatz zur ordinalen logistischen Regression stellt die multinomiale logistische Regression keine OR dar, sondern sogenannte Relative Risk Ratios (RRR). Der Befehl dazu endet daher nicht mit *or*, sondern mit *rrr*. Das relative Risiko vergleicht dabei die Wahrscheinlichkeiten (nicht die *Odds*, welche ja dem Verhältnis von Wahrscheinlichkeit zu Gegenwahrscheinlichkeit entsprechen), dass ein Ereignis auftritt zwischen zwei Szenarien (Denham 2017). Die RRR können ebenso aus den Koeffizienten des oben angeführten Modells, also durch e^{coeff} gewonnen werden (UCLA 2021b). Das RRR ist dabei standardmäßig als Verhältnis von $e^{\text{coeff}}(x)$ und $e^{\text{coeff}}(x+1)$, beide jeweils relativ zur Referenzkategorie, definiert (UCLA 2021b). Alle anderen Faktoren werden dabei konstant gehalten. Unter Berücksichtigung der Nichtlinearität des Alters lauten die Befehle für die geschätzten Koeffizienten beziehungsweise die RRR der Regressionsgleichung (5):

```
mlogit Spalte1 i.corona i.geschlecht alter_cen_ordinal alter_ordinal2 i.studium, robust cluster(RND)
```

und

```
mlogit Spalte1 i.corona i.geschlecht alter_cen_ordinal alter_ordinal2 i.studium, robust cluster(RND) rrr
```

4 Empirische Ergebnisse und Diskussion

Die Resultate der analytischen Untersuchung der Daten werden in derselben Reihenfolge wie auch die Vorgehensweise selbst dargestellt. Kapitel 4.1 enthält die Ergebnisse der allgemeinen Voranalyse, aus welcher geschlossen wurde, dass in den weiteren Schritten zwischen Bachelor- und Masterstudierenden unterschieden wird. Das genaue Skript, welches die Befehle für die Analyse und die Erstellung der Graphiken enthält, ist im Anhang zu finden. Aufgrund von Fehlermeldungen in den Ergebnissen der ordinalen Logit-Regression für den eingeschränkten Datensatz von Masterstudierenden und der geringen Anzahl an Datenpunkten wird diese Analyse sowie die multinomiale Logit-Regression nur für Bachelorstudierende durchgeführt. Nicht signifikante Koeffizienten werden in den nachfolgenden Tabellen immer mit *n.s.* dargestellt.

4.1 Resultate der Voranalyse

Effekte auf akademische Studierendenleistungen		
	Abhängige Variable: Erfolg = 1 falls Prüfung bestanden	
	(1)	(2)
Corona_Indikator (<i>vor Corona</i>) <i>während Corona</i>	0,1314** (0,0169)	0,1018** (0,0251)
Geschlecht (<i>männlich</i>) <i>weiblich</i>	0,0267* (0,0133)	n.s. (n.s.)
Alter	-0,0090** (0,0015)	-0,0169** (0,0018)
Alter (quadriert)		0,0008** (0,0001)
Studium	0,2278** (0,0180)	0,2374** (0,0174)
Wahl_Pflicht (<i>Pflicht</i>) <i>Wahlfach</i>	0,0657** (0,0165)	0,0698** (0,0160)
<i>Wahlpflichtfach</i>	0,1140** (0,0232)	0,1166** (0,0228)
Pseudo-R-squared	0,046	0,0485
p-Wert des Modells	<0,001	<0,001
Anzahl an Datenpunkten	4.944	5.603
VIF	1,11	1,26
% korrekt klassifizierter Werte	66,4%	65,73%
Log-Likelihood Modell	-3.006,23	-3.427,35
Log-Likelihood nur Konstante	-3.151,07	-3.602,22

Tabelle 10: Resultate der binären logistischen Regression

Anmerkung: Diese Tabelle gibt die marginalen Effekte der Variablen sowie die robusten Standardfehler des Logit-Modells an und entspricht daher nicht direkt den Koeffizienten. Die abhängige Variable (Erfolg) ist ein Indikator, welcher angibt, ob eine Prüfung bestanden wurde (1) oder nicht (0). Die *kursiv* angeführten Kategorien unter jeder unabhängigen Variablen geben an, auf welche Ausprägung sich die marginalen Effekte beziehen – die in Klammern angeführte Kategorie ist die entsprechende Referenzkategorie.

**Signifikanz 1%-Level, *Signifikanz 5%-Level

Die Voranalyse wurde anhand des gesamten Datensatzes durchgeführt und dient als Ausgangsbasis für die weitere Spezifizierung der Modelle. In diesem Modell, welches auf der ersten Datenlieferung basiert, fließen 5.658 Prüfungsergebnisse ein. Die Ergebnisse sind in Tabelle 10 dargestellt. Modell (1) (entsprechend der Regressionsgleichung (1) aus Kapitel 3.3.2) beschreibt dabei das Modell ohne die Berücksichtigung von nichtlinearen Effekten der Variable *Alter*, diese Effekte werden sodann in Modell (2) miteinbezogen. Es ist dabei klar ersichtlich, dass alle Variablen bis auf das Geschlecht einen signifikanten Einfluss auf die abhängige Variable Erfolg haben. Die Referenzwerte sind bei den unabhängigen Variablen in Klammern angegeben – so wird etwa der Wert der Variablen *Corona_Indikator* als Anstieg oder Abfall der Wahrscheinlichkeit, in Zeiten der Corona-Pandemie eine Prüfung zu bestehen im Vergleich zu der Wahrscheinlichkeit in den Semestern davor (Referenzwert: vor Corona) interpretiert. Da in Tabelle 10 bereits die marginalen Effekte angeführt sind, können die Werte direkt interpretiert werden. So bedeutet etwa der positive Wert für die Variable *Corona_Indikator* in Modell (1), dass in Zeiten der Corona-Pandemie (im Vergleich zu Prüfungen, welche vor dem

Sommersemester 2020 abgelegt wurden) die Wahrscheinlichkeit, eine Prüfung zu bestehen, um 13,14% steigt. Das Alter hat einen geringfügigen, aber signifikanten Effekt: So ist ein marginaler Anstieg des Alters mit einer Senkung der Wahrscheinlichkeit, eine Prüfung zu bestehen um 0,9% verbunden. Auch das Studium hat offensichtlich einen signifikanten Einfluss auf die Erfolgswahrscheinlichkeit bei Prüfungen. Da der Wert der Variable *Studium* für Bachelorstudiengänge 0 und für Masterstudiengänge 1 beträgt, kann festgehalten werden, dass die Wahrscheinlichkeit, eine Prüfung zu bestehen, bei Studierenden in Masterstudiengängen um 22,8% höher ist als bei Studierenden in Bachelorstudiengängen. Außerdem ist die Wahrscheinlichkeit, eine Prüfung zu bestehen, in Wahlfächern um 6,6% und in Wahlpflichtfächern um 11,4% höher als in Pflichtfächern. In Tabelle 10 sind in Bezug auf Modell (1) klare Einschränkungen des Modells ersichtlich. So beträgt der Wert des Pseudo- R^2 (nach McFadden) nur 4,6% und deutet somit auf einen schwachen Erkenntnisgewinn des Modells hin. Aus der Klassifikationstabelle kann mit dem Prozentsatz an korrekt klassifizierten Werten ein geringer Erkenntnisgewinn geschlossen werden – rund 66% der Werte werden korrekt zugeordnet, was nur leicht über einer zufälligen Zuordnung von rund 50% liegt (Denham 2017).

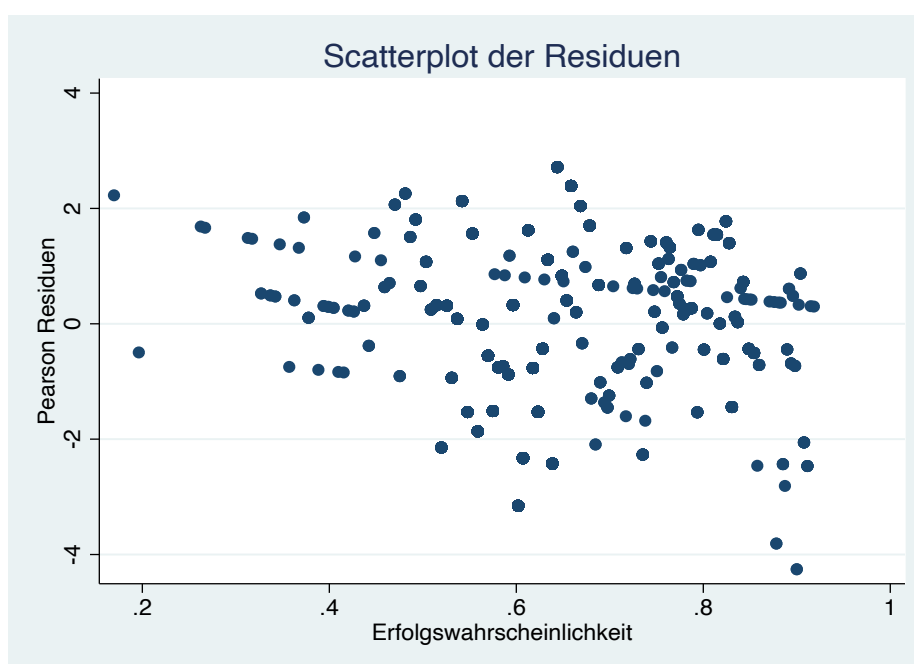


Abbildung 9: Residuen des einfachen binären logistischen Modells

In Abbildung 9 werden die standardisierten Residuen des Modells dargestellt – hier ist ersichtlich, dass es einige Ausreißer (Residuenwert über 2 oder unter -2) gibt, welche das Ergebnis beeinflussen könnten. Diese werden jedoch im weiteren Verlauf der Berechnung ausgeschlossen und sind daher in den oben angeführten Werten nicht mehr enthalten. Daher unterscheidet sich auch die Anzahl an Datenpunkten wie in Tabelle 10 dargestellt von der Gesamtzahl an verfügbaren Beobachtungen. Der p-Wert der Wald-Statistik ist signifikant und zeigt damit an,

dass das Modell zumindest eine unabhängige Variable enthält, welche nützlich zur Bestimmung der abhängigen Variablen ist (Field, Miles und Field 2012). Der Varianzinflationsfaktor zeigt an, dass kein Problem mit Multikollinearität der erklärenden Variablen gegeben ist, da der Wert von 1,11 weit unter problematischen Grenzwerten wie 5 oder 10 liegt (Backhaus et al. 2018).

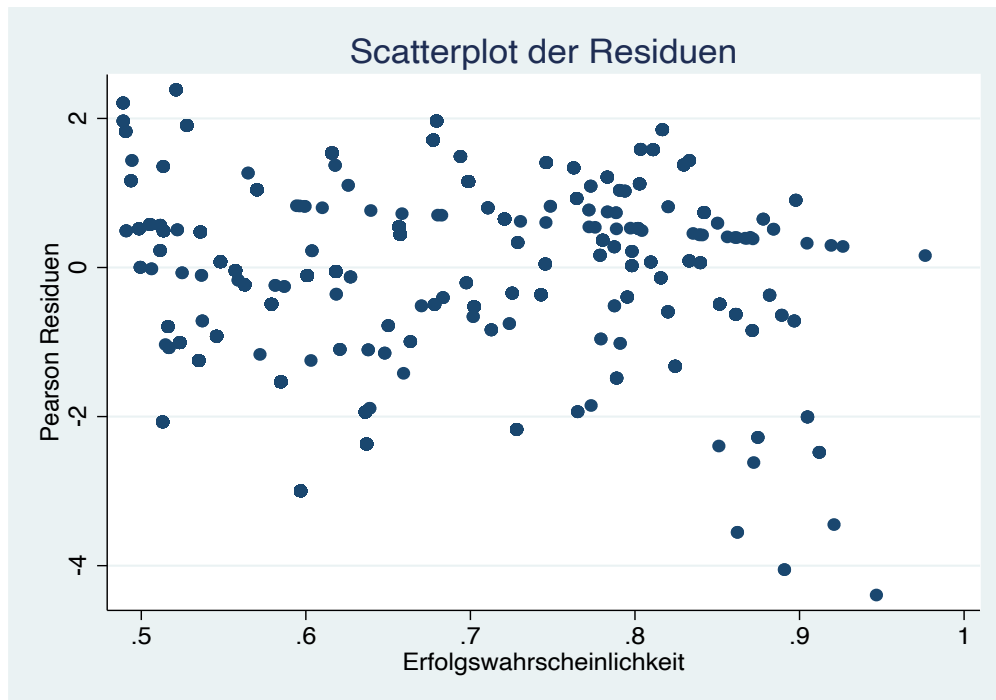


Abbildung 10: Residuen des binären logistischen Modells unter Berücksichtigung von Nichtlinearität

Modell (2) entspricht prinzipiell Modell (1), es werden jedoch auch nichtlineare Effekte der Variable *Alter* zugelassen, wodurch die Variable *Alter* selbst in zentrierter und in quadrierter Form im Modell enthalten ist. Das Hinzufügen der nichtlinearen Information führt zu einer leichten Änderung der Größenordnung der beobachteten Effekte – an der Signifikanz der Variablen selbst ändert sich jedoch nichts. Dieses Modell gibt also an, dass beim Ablegen einer Prüfung in Zeiten der Corona-Pandemie im Vergleich zu den Semestern davor die Wahrscheinlichkeit, eine Prüfung zu bestehen, um 10,2% höher ist. Das *Geschlecht* hat keinen signifikanten Einfluss, in Bezug auf das *Alter* sind jedoch beide Variablen signifikant – dies ist auf die Nichtlinearität der Variable *Alter* zurückzuführen. Während also ausgehend vom Mittelwert des Alters ein zusätzliches Lebensjahr zu einer geringeren Erfolgswahrscheinlichkeit führt, so nimmt diese Erfolgswahrscheinlichkeit mit zunehmender Entfernung zum Mittelwert wieder zu. Die Effektgrößen sind eher gering, jedoch sind beide Variablen auf dem 1%-Level signifikant. Der Effekt des Studiums bleibt ähnlich – in Masterstudiengängen ist (im Vergleich zu Bachelorstudien) die Erfolgswahrscheinlichkeit um rund 23,8% höher. Auch die Variable *Wahl_Pflicht* hat einen signifikanten Einfluss, wobei die Effektgrößen ungefähr gleich bleiben wie bei Modell (1). Auch in diesem Modell zeigen jedoch die Werte zur Überprüfung der Modellgüte, dass nur

eine eingeschränkte Aussagekraft vorliegt. Das Pseudo- R^2 nach McFadden ist nur minimal angestiegen (auf 4,85%), auch der Prozentsatz der korrekt eingeordneten Werte ist kaum gestiegen und beträgt immer noch rund 66%. Im Vergleich zu einer zufälligen Einteilung der Werte mit einer korrekten Klassifizierung von rund 50% ist dieser Wert bedingt zufriedenstellend. Es scheint jedoch keine systematischen Fehler im Modell zu geben – so ist der p-Wert der Wald-Statistik signifikant und zeigt an, dass das Modell selbst eine Aussagekraft hat (Denham 2017).

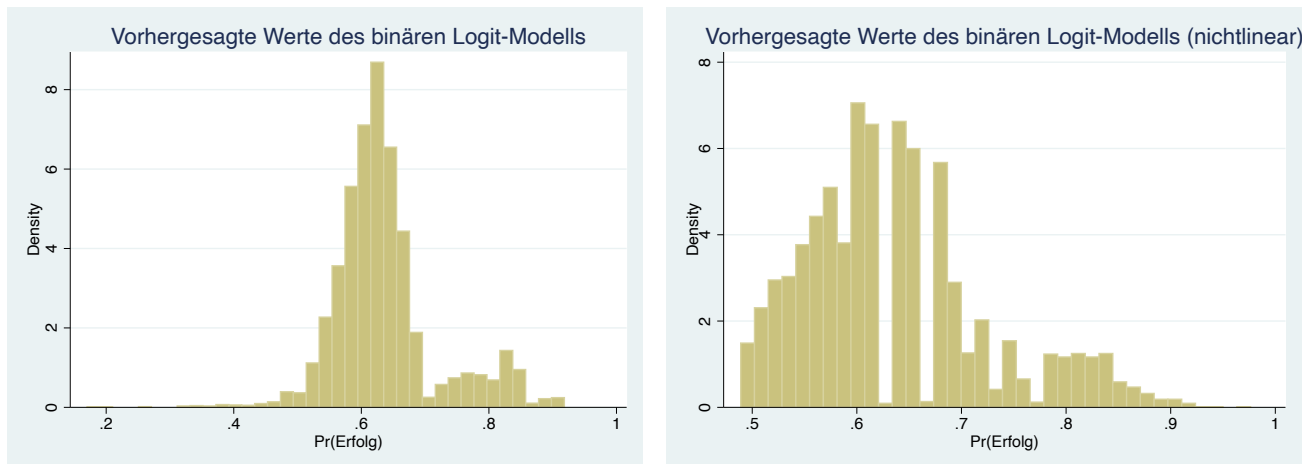


Abbildung 11: Vorhergesagte Werte Modell (1) und Modell (2)

Auch der Varianzinflationsfaktor ist gering und zeigt keine Anzeichen von Multikollinearität zwischen den unabhängigen Variablen. In Abbildung 10 sind die Residuen des Modells dargestellt – auch hier lässt sich kein Muster erkennen. Es scheint daher keine systematischen Abweichungen zu geben. Die Log-Likelihood-Werte von beiden Modellen zeigen eine leichte Verbesserung des LL_0 -Wertes und des Log-Likelihood-Wertes des aktuellen Modells. Interessanterweise gibt es jedoch bei beiden Modellen klare Unterschiede in den vorhergesagten Werten – Abbildung 11 stellt diese gegenüber. Während Modell (1) den Großteil der vorhergesagten Werte zwischen einer Erfolgswahrscheinlichkeit von 0,5 (50%) und 0,7 (70%) produziert, findet sich der Großteil der vorhergesagten Werte von Modell (2) zwischen 0,5 (50%) und 0,6 (60%). Obwohl der Mittelwert der vorhergesagten Werte für beide Modelle bei 0,6352 liegt, ist die Spannweite sehr unterschiedlich: Die Werte von Modell (1) haben eine Spannweite von 0,7486 (Minimum: 0,1697, Maximum: 0,9183), jene von Modell (2) jedoch nur eine Spannweite von 0,4874 (Minimum: 0,4890, Maximum: 0,9764). Die vorhergesagten Werte beschreiben dabei die Wahrscheinlichkeit, dass eine Beobachtung in die höhere Kategorie der abhängigen Variable fällt (d. h. in diesem Fall die Prüfung besteht) (Denham 2017).

Aufgrund der klaren Unterschiede zwischen Studierenden eines Bachelor- und eines Masterstudiums werden diese Datensätze in den weiteren Modellen voneinander getrennt betrachtet.

4.2 Resultate der binären logistischen Regression für Bachelorstudierende

Die Ergebnisse der unterschiedlichen Modelle sind in Tabelle 11 dargestellt. Beide berechneten Modelle entsprechen der Regressionsgleichung (4), einmal werden die Daten bis Sommersemester 2020 berücksichtigt (erste Datenlieferung), einmal werden die Gesamtdaten bis inklusive Wintersemester 2020/21 berücksichtigt (inklusive Nachlieferung). Die Daten in Tabelle 11 entsprechen den Werten der Modelle ohne Ausreißer (Residuen, welche betragsmäßig größer als 2 sind, wurden entfernt).

Effekte auf akademische Studierendenleistungen		
	Abhängige Variable: Erfolg = 1 falls Prüfung bestanden	
	(4, bis SS20)	(4, bis WS20/21)
Corona_Indikator (<i>vor Corona</i>) <i>während Corona</i>	0,115** (0,029)	0,138** (0,019)
Geschlecht (<i>männlich</i>) <i>weiblich</i>	0,037** (0,014)	0,034* (0,014)
Alter	-0,011** (0,002)	-0,010** (0,002)
Alter (quadriert)	0,0004** (0,0001)	0,0004** (0,0001)
Wahl_Pflicht (<i>Pflichtfach</i>) <i>Wahlfach</i>	0,127** (0,017)	0,103** (0,017)
<i>Wahlpflichtfach</i>	0,166** (0,026)	0,156** (0,025)
Studienrichtung (<i>Anderes</i>) <i>Lehramt</i>	0,309** (0,223)	0,294** (0,021)
<i>Politikw.</i>	-0,126* (0,064)	-0,157** (0,060)
<i>Sprachw.</i>	n.s. (n.s.)	n.s. (n.s.)
<i>Wirtschaftsw.</i>	0,083** (0,020)	0,070** (0,019)
Pseudo-R-squared	0,061	0,062
p-Wert des Modells	< 0,001	< 0,001
Anzahl an Datenpunkten	4.697	5.086
VIF	1,34	1,35
% korrekt klassifizierter Werte	63,25%	64,06%
Log-Likelihood Modell	-2.911,25	-3.127,1
Log-Likelihood nur Konstante	-3.100,45	-3.332,4
BIC	5.915,5	6.276,3
AIC	5.844,5	6.348,2

Tabelle 11: Resultate der binären logistischen Regression für Bachelorstudierende für Daten bis SS 20 sowie bis WS 20/21

Anmerkung: Diese Tabelle gibt die marginalen Effekte der Variablen sowie die robusten Standardfehler des Logit-Modells an und entspricht daher nicht direkt den Koeffizienten. Die abhängige Variable (Erfolg) ist ein Indikator, welcher angibt, ob eine Prüfung bestanden wurde (1) oder nicht (0). Die *kursiv* angeführten Kategorien unter jeder unabhängigen Variablen geben an, auf welche Ausprägung sich die marginalen Effekte beziehen – die in Klammern angeführte Kategorie ist die entsprechende Referenzkategorie.

**Signifikanz 1%-Level, *Signifikanz 5%-Level

In Tabelle 11 werden die marginalen Effekte der unabhängigen Variablen dargestellt. Der Großteil der Variablen ist sowohl für den Datensatz bis Sommersemester 2020 als auch für

den Datensatz bis Wintersemester 2020/21 signifikant. Die Unterscheidung dieser beiden Modelle ist besonders für die Variable *Corona_Indikator* interessant: Im Sommersemester 2020 stieg für Bachelorstudierende die Wahrscheinlichkeit, eine Prüfung zu bestehen, um 11,5% im

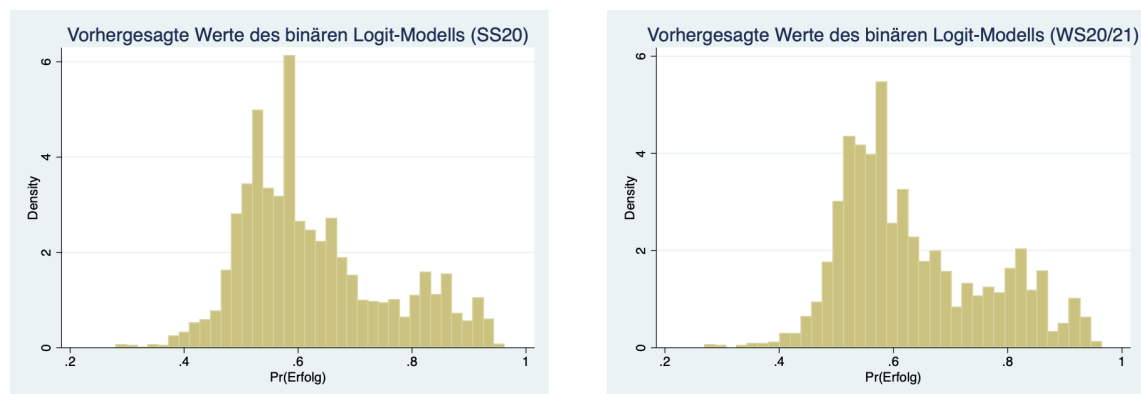


Abbildung 12: Vorhergesagte Werte des binären Modells für Bachelorstudierende

Vergleich zu den Vorsemestern. Im Sommer- und Wintersemester gemeinsam stieg die Wahrscheinlichkeit, eine Prüfung zu bestehen, um 13,8% im Vergleich zu den Semestern vor der Corona-Pandemie. Der Effekt des virtuellen beziehungsweise hybriden Lernens und der Corona-Pandemie vergrößert sich also im Zeitverlauf. Weibliche Studierende weisen eine um 3,4% beziehungsweise 3,7% höhere Wahrscheinlichkeit auf, eine Prüfung zu bestehen, als männliche Studierende. Das Alter hat wiederum einen leicht negativen Einfluss auf die Wahrscheinlichkeit, eine Prüfung zu bestehen, was jedoch durch die Nichtlinearität abgeschwächt wird. Im Vergleich zu einem Pflichtfach steigt die Wahrscheinlichkeit, eine Prüfung zu bestehen, um 12,7% (beziehungsweise 10,3%) für Wahlfächer und um 16,6% (beziehungsweise 15,6%) für Wahlpflichtfächer. Im Vergleich zu nicht kategorisierten Studienrichtungen haben Studierende des Lehramts eine um 30,9% (29,4%) höhere Wahrscheinlichkeit, Studierende der Politikwissenschaften eine um 12,6% (15,7%) geringere und Studierende der Wirtschaftswissenschaften eine um 8,3% (7,0%) höhere Wahrscheinlichkeit, eine Prüfung zu bestehen. Beide Modelle weisen einen signifikanten p-Wert der Wald-Statistik auf, das Pseudo- R^2 beträgt 6,1% beziehungsweise 6,2%. Es gibt abermals kein Problem mit dem Varianzinflationsfaktor, welcher mit einem Wert von 1,34 beziehungsweise 1,35 unter den problematischen Grenzwerten liegt. Der Prozentsatz der korrekt klassifizierten Werte hat sich im Vergleich zum allgemeinen Modell leicht verringert und beträgt nun 63% beziehungsweise 64% (inklusive Wintersemester 2020/21). Das Modell hat jedenfalls statistische Aussagekraft, es scheinen jedoch noch einige weitergehende Einflüsse auf die abhängige Variable zu fehlen.

In Abbildung 12 sind die vorhergesagten Werte des jeweiligen Modells dargestellt. Hier ist klar ersichtlich, dass die Erweiterung der Daten zu keiner Trendumkehr in den Daten führt – der Großteil der Werte ist in beiden Modellen bei einer Erfolgswahrscheinlichkeit von 0,5 (50%)

und 0,8 (80%) zu finden. Auch die Spannweite ist ähnlich – für Daten bis inklusive Wintersemester 2020/21 liegt das Minimum bei 0,27 (27%) und das Maximum bei 0,96 (96%). Bei den Daten bis inklusive Sommersemester 2020 liegt das Minimum bei 28%, das Maximum bei 96%.

Die Log-Likelihood-Werte sowie AIC und BIC der beiden Modelle können nicht verglichen werden, da die Datenbasis unterschiedlich ist. Bei beiden Modellen ist jedoch ersichtlich, dass der Log-Likelihood-Wert des Modells jedenfalls näher am angestrebten Wert 0 liegt als der Log-Likelihood-Wert der reinen Konstante des Modells.

4.3 Resultate der binären logistischen Regression für Masterstudierende

Effekte auf akademische Studierendenleistungen		
	Abhängige Variable: Erfolg = 1 falls Prüfung bestanden (4, bis SS20) (4, bis WS20/21)	
Corona_Indikator (<i>vor Corona</i>) <i>während Corona</i>	n.s. (n.s.)	n.s. (n.s.)
Geschlecht (<i>männlich</i>) <i>weiblich</i>	-0,074* (0,030)	-0,046* (0,031)
Alter	n.s. (n.s.)	n.s. (n.s.)
Wahl_Pflicht (<i>Pflichtfach</i>) <i>Wahlfach</i>	n.s. (n.s.)	n.s. (n.s.)
<i>Wahlpflichtfach</i>	-0,13** (0,044)	-0,10* (0,043)
Studienrichtung (<i>Anderes</i>) <i>Lehramt</i>	0,267* (0,127)	. (.)
<i>Politikw.</i>	-0,472**	.
<i>Sprachw.</i>	n.s. (n.s.)	(.) n.s.
<i>Wirtschaftsw.</i>	(0,130) 0,344** (0,079)	(n.s.) 0,389** (0,078)
Pseudo-R-squared	0,215	0,171
p-Wert des Modells	< 0,001	< 0,001
Anzahl an Datenpunkten	438	453
VIF	1,37	1,50
% korrekt klassifizierter Werte	87,44%	87,64%
Log-Likelihood Modell	-137,33	-140,50
Log-Likelihood nur Konstante	-174,96	-198,84
BIC	335,5	329,9
AIC	294,7	297,0

Tabelle 12: Resultate der binären logistischen Regression für Masterstudierende für Daten bis SS 20 sowie bis WS 20/21

Anmerkung: Diese Tabelle gibt die marginalen Effekte der Variablen sowie die robusten Standardfehler des Logit-Modells an und entspricht daher nicht direkt den Koeffizienten. Die abhängige Variable (Erfolg) ist ein Indikator, welcher angibt, ob eine Prüfung bestanden wurde (1) oder nicht (0). Die *kursiv* angeführten Kategorien unter jeder unabhängigen Variablen geben an, auf welche Ausprägung sich die marginalen Effekte beziehen – die in Klammern angeführte Kategorie ist die entsprechende Referenzkategorie.

**Signifikanz 1%-Level, *Signifikanz 5%-Level

Die Ergebnisse der unterschiedlichen Modelle sind in Tabelle 12 dargestellt. Wie auch bei den Bachelorstudierenden entsprechen beide Modelle der Regressionsgleichung (4), wobei einmal die Daten bis Sommersemester 2020 (erste Datenlieferung) und einmal die Gesamtdaten bis inklusive Wintersemester 2020/21 (inklusive Nachlieferung) berücksichtigt werden. Die Daten in Tabelle 12 entsprechen auch hier den Werten der Modelle ohne Ausreißer (Residuen, welche betragsmäßig größer als 2 sind, wurden entfernt). Tabelle 12 stellt die marginalen Effekte der unabhängigen Variablen dar, wobei hier klare Unterschiede zu den Ergebnissen des Datensatzes für Bachelorstudierende erkennbar sind. Einige unabhängige Variablen (darunter auch der Indikator für Semester, welche von der Corona-Pandemie betroffen sind) sind in diesem Datensatz nicht mehr signifikant – die marginalen Effekte können somit auch nicht interpretiert werden. Nichtsdestoweniger hat auch dieses Modell einen signifikanten p-Wert der Wald-Statistik, das Pseudo- R^2 beträgt 21,5% (bis Sommersemester 2020) beziehungsweise 17,1% (bis Wintersemester 2020/21). In diesem Datensatz haben weibliche Studierende eine um 7,4% (Sommersemester 2020) beziehungsweise 4,6% (bis Wintersemester 2020/21) geringere Wahrscheinlichkeit, eine Prüfung zu bestehen als männliche Studierende. Im Vergleich zu einem Pflichtfach werden Prüfungen in einem Wahlpflichtfach mit einer um 13% (bis Sommersemester 2020) beziehungsweise 10,3% (bis Wintersemester 2020/21) geringeren Wahrscheinlichkeit bestanden. Im Vergleich zu einer nicht kategorisierten Studienrichtung weisen Studierende des Lehramts im Datensatz bis Sommersemester 2020 eine um 26,7% höhere Wahrscheinlichkeit auf, Studierende der Politikwissenschaften eine um 47,2% geringere Wahrscheinlichkeit und Studierende der Wirtschaftswissenschaften eine um 34,4% höhere Wahrscheinlichkeit auf, eine Prüfung zu bestehen. Im Datensatz bis Wintersemester 2020/21 kann das Modell für Studierende des Lehramts und der Politikwissenschaften keine marginalen Effekte berechnen. Dies ist bereits der erste Hinweis auf fehlende statistische Aussagekraft des Modells – die Anzahl der berücksichtigten Datenpunkte (ohne Ausreißer) ist mit 438 (bis Sommersemester 2020) beziehungsweise 453 (bis Wintersemester 2020/21) sehr gering. Der Varianzinflationsfaktor deutet auf kein Problem mit Multikollinearität hin, auch die Klassifizierung der Werte ist mit rund 88% sehr gut. Hier muss allerdings berücksichtigt werden, dass im Grunddatensatz – zum Beispiel bis Wintersemester 2020/21 – nur 56 Prüfungen enthalten sind, die negativ beurteilt wurden. Das Modell erkennt nur 3 negative Prüfungen korrekt und weist damit eine Spezifität von 5,36% auf, die Sensitivität des Modells (99,24%, 394 aus 397

tatsächlich positiv bewerteten Prüfungen korrekt erkannt) bedingt die derart hohe korrekte Klassifizierung der Werte.

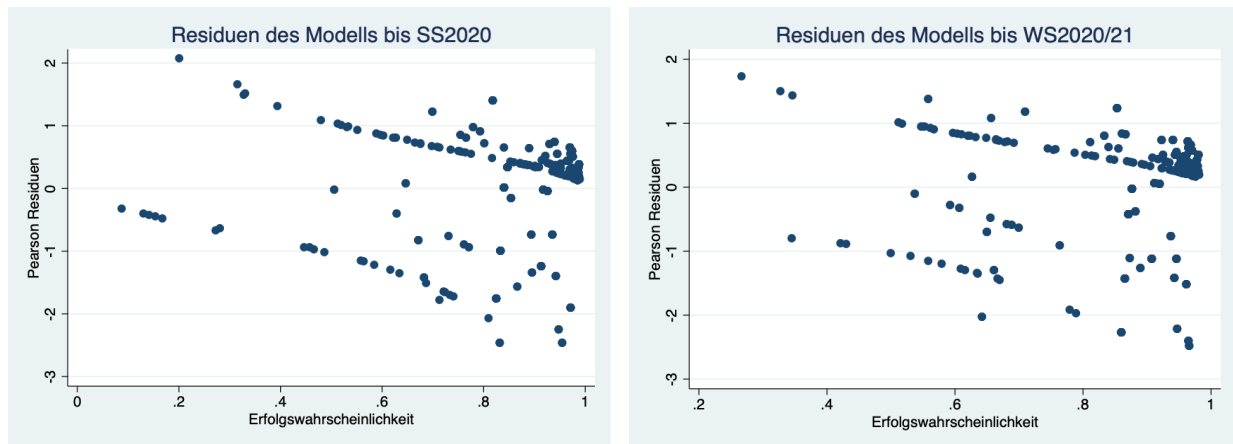


Abbildung 13: Residuen des Modells für Masterstudierende

Auch in Abbildung 13 ist klar ersichtlich, dass dieses Modell durch den geringen Datensatz nur eine begrenzte Aussagekraft aufweist. Die Residuen sind in beiden Datensätzen sehr klar bei einer hohen Erfolgswahrscheinlichkeit konzentriert und Muster sind in beiden Datensätzen erkennbar. Trotz der hohen Werte für das Pseudo- R^2 und der guten Klassifizierung des Modells sollten die Werte dieses Modells nur mit Vorsicht interpretiert werden – vor allem unter dem Aspekt der Spezifität, welche nur drei Prüfungsleistungen tatsächlich korrekt klassifiziert.

4.4 Resultate der ordinalen logistischen Regression für Bachelorstudierende

Wie in Kapitel 3.3.3 ausgeführt, ist in den Daten der Bachelor- und Masterstudierenden ein nichtlinearer Zusammenhang zwischen den Alterskategorien und den erreichten Noten feststellbar. Es wird daher in diesem Kapitel zuerst Modell (4) dargestellt (siehe Tabelle 13), wobei dieses in Folge auf die Voraussetzung der parallelen OR in den verschiedenen Kategorien der abhängigen Variablen überprüft wird. Da dies nicht erfüllt ist, wird dasselbe Modell – wie in Regressionsgleichung (5) beschrieben – als generalisierte ordinale Logit-Regression durchgeführt und dargestellt. Die Interpretation der Ergebnisse in Tabelle 13 muss daher mit Vorsicht erfolgen und es wird darauf hingewiesen, dass die Werte, welche sich aus Modell (5) ergeben, zur weitergehenden Interpretation verwendet werden sollten. Für ein besseres Verständnis werden jedoch in diesem Kapitel auch die Werte aus Modell (4), wie dargestellt in Tabelle 13, durchbesprochen und die Interpretation der OR erklärt.

Effekte auf akademische Studierendenleistungen		
	Abhängige Variable: Noten (Skala von 1 bis 5)	
	(4, bis SS20)	(4, bis WS20/21)
Corona_Indikator (<i>vor Corona</i>) <i>während Corona</i>	n.s. (n.s.)	0,616** (0,044)
Geschlecht (<i>männlich</i>) <i>weiblich</i>	0,815** (0,045)	0,832** (0,044)
Alter	1,035** (0,008)	1,040** (0,008)
Alter (quadriert)	0,998* (0,001)	0,998** (0,001)
Wahl_Pflicht (<i>Pflichtfach</i>) <i>Wahlfach</i>	n.s. (n.s.)	n.s. (n.s.)
<i>Wahlpflichtfach</i>	0,413** (0,047)	0,402** (0,046)
Studienrichtung (<i>Anderes</i>) <i>Lehramt</i>	0,318** (0,030)	0,335** (0,030)
<i>Politikw.</i>	n.s. (n.s.)	n.s. (n.s.)
<i>Sprachw.</i>	n.s. (n.s.)	n.s. (n.s.)
<i>Wirtschaftsw.</i>	0,805** (0,066)	0,778** (0,059)
Pseudo-R-squared	0,0277	0,0292
p-Wert des Modells	< 0,001	< 0,001
Anzahl an Datenpunkten	4.793	5.251
VIF	1,37	1,38
Log-Likelihood Modell	-7.295,1	-7.322,5
Log-Likelihood nur Konstante	-7.439,8	-7.542,9
BIC	14.761,0	14.765,5
AIC	14.630,1	14.608,4

Tabelle 13: Resultate der ordinalen logistischen Regression für Bachelorstudierende

Anmerkung: Diese Tabelle gibt die OR der Variablen sowie die robusten Standardfehler des ordinalen Logit-Modells an und entspricht daher nicht direkt den Koeffizienten. Die abhängige Variable (Noten) sind ordinal skaliert von 1 (Sehr gut) bis 5 (Nicht Genügend). Die *kursiv* angeführten Kategorien unter jeder unabhängigen Variablen geben an, auf welche Ausprägung sich das OR bezieht – die in Klammern angeführte Kategorie ist die entsprechende Referenzkategorie.

**Signifikanz 1%-Level, *Signifikanz 5%-Level

Im Hinblick auf die Signifikanz von Variablen ähneln die Ergebnisse von Modell (4) für Bachelorstudierende bis Sommersemester 2020 sowie bis Wintersemester 2020/21 den Resultaten der jeweiligen binären logistischen Regression. Ein großer Unterschied ist die Signifikanz der Variable *Corona_Indikator* – während im Datensatz der Werte bis Sommersemester 2020 kein signifikanter Einfluss festgestellt werden kann, so ist die Variable im Datensatz mit den erweiterten Werten bis Wintersemester 2020/21 auf dem 1%-Level signifikant. Das OR für die Variable *Corona_Indikator* in Modell (4, bis WS 20/21) für die Ausprägung *während Corona* beträgt 0,616. Da keine explizite Referenzkategorie der abhängigen Variablen angegeben ist, kann dies wie folgt interpretiert werden: Die Wahrscheinlichkeit, eine schlechtere (=höhere) Note auf eine Prüfung in Semestern der Corona-Pandemie zu erhalten, ist 0,616-mal der

Wahrscheinlichkeit, eine schlechtere (=höhere) Note auf eine Prüfung in einem Semester, welches nicht von der Corona-Pandemie beeinflusst wurde, zu erhalten. Für die Interpretation in Bezug auf akademische Studienerfolge kann daher festgehalten werden, dass die Wahrscheinlichkeit, eine schlechtere Note zu erhalten, vor der Corona-Pandemie deutlich höher war und im vorliegenden Datensatz bedeutend gesunken ist. Dieselbe Variable weist im Datensatz bis zum Sommersemester 2020 keinen signifikanten Einfluss auf und die OR sollen daher auch nicht interpretiert werden. Das Geschlecht und auch das Alter sind in beiden Datensätzen signifikant: Die Wahrscheinlichkeit, eine schlechtere (=höhere) Note auf eine Prüfung zu erhalten, ist für weibliche Bachelorstudierende 0,815-mal (beziehungsweise 0,832-mal) derselben Wahrscheinlichkeit für männliche Bachelorstudierende. Das Alter zeigt ein OR von gerundet 1,04 in beiden Datensätzen und ist auf dem 1%-Level signifikant. Es kann daher festgehalten werden, dass die Wahrscheinlichkeit, eine schlechtere (=höhere) Note zu erhalten, mit einem Anstieg des Alters um ein Jahr um das 0,04-fache steigt. Der Effekt ist also – auch gerade im Vergleich mit dem OR der Variable *Corona_Indikator* – sehr klein, aber vorhanden und signifikant, wobei dieser durch den leichten nichtlinearen Zusammenhang abgeschwächt wird.

Auch die Prüfungsart beeinflusst die Noten von Studierenden signifikant – hier kann festgehalten werden, dass die Wahrscheinlichkeit für Bachelorstudierende bei einer Prüfung im Rahmen eines Wahlpflichtfachs eine schlechtere (=höhere) Note zu erhalten, nur 0,413-mal (beziehungsweise 0,402-mal) derselben Wahrscheinlichkeit in einem Pflichtfach beträgt. Ähnlich die Interpretation der Studienrichtung, welche nur für zwei Ausprägungen im Vergleich zu Kategorie „Anderes“ (also nicht kategorisierte Studienrichtungen) signifikant ist: Studierende des Lehramts haben nur 0,318-mal (beziehungsweise 0,335-mal) die Wahrscheinlichkeit, auf eine Prüfung eine schlechtere Note zu erhalten, als Studierende von nicht kategorisierten Studienrichtungen. Studierende der Wirtschaftswissenschaften haben nur 0,805-mal (beziehungsweise 0,778-mal) die Wahrscheinlichkeit, auf eine Prüfung eine schlechtere Note zu erhalten, als Studierende von nicht kategorisierten Studienrichtungen. Die Modelle weisen nur einen sehr geringen Pseudo- R^2 -Wert von 2,7% beziehungsweise 2,9% auf. Dies ist auch an der relativ geringen Verbesserung der Log-Likelihood-Werte der Modelle im Vergleich zum jeweiligen Log-Likelihood-Wert der Konstanten ersichtlich. Der p-Wert der Wald-Statistik in den Modellen ist jedoch am 1%-Level signifikant und zeigt daher an, dass diese zur Erklärung der abhängigen Variablen geeignet sind. Auch der Varianzinflationsfaktor deutet auf keine Schwierigkeiten hin, da der Wert von 1,37 beziehungsweise 1,38 weit unter den Grenzwerten von 5 beziehungsweise 10 liegt (Backhaus et al. 2018).

Test der gesamten Modelle in einem Likelihood-Ratio-Test			
	Chi ² -Wert	p-Wert des Chi ² -Modells	
Modell (4, bis SS 20)	189,7	<0,001	
Modell (4, bis WS 20/21)	204,4	<0,001	
Test der einzelnen Variablen von Modell (4, bis WS 20/21) in einem Brant-Test			
	Chi ² -Wert	p-Wert des Chi ² -Modells	Freiheitsgrade
Corona_Indikator	31,66	<0,01	3
Geschlecht	6,36	0,095	3
Alter (zentriert)	12,51	<0,01	3
Alter ²	4,42	0,219	3
Wahl_Pflicht (<i>Pflichtfach</i>)	12,56	<0,01	3
<i>Wahlfach</i>		<0,01	3
<i>Wahlpflichtfach</i>	14,67	<0,01	3
Studienrichtung (<i>Anderes</i>)	20,91	<0,01	3
<i>Lehramt</i>		<0,01	3
<i>Politikw.</i>		0,797	3
<i>Sprachw.</i>		0,878	3
<i>Wirtschaftsw.</i>	16,91	<0,01	3

Tabelle 14: Test der Modelle auf proportionale OR in den Kategorien

Wie von Hoffmann (2016) vorgeschlagen, werden die beiden Modelle auf die Voraussetzung von proportionalen OR in den unterschiedlichen Kategorien der abhängigen Variablen überprüft. Wie in Tabelle 14 dargestellt, sind die Werte des Likelihood-Ratio-Tests auf die Verhältnismäßigkeit für beide Modelle signifikant. In diesem Fall bedeutet das, dass beide Modelle die Voraussetzung nicht erfüllen können, da die Nullhypothese einer Verhältnismäßigkeit entspricht und wir diese aufgrund der Testergebnisse verwerfen müssen (Hoffmann 2016). Um zu überprüfen, ob dies auf eine einzelne Variable im Modell zurückzuführen ist, wird das umfangreichere Modell (4, bis WS20/21) zusätzlich einem Brant-Test unterzogen. Hier ist ersichtlich, dass nur wenige Variablen die Voraussetzung der proportionalen OR erfüllen – das Problem erstreckt sich also auf mehrere Variablen. Um eine falsche Interpretation der Ergebnisse aus Modell (4) zu verhindern, wird in einem ersten Schritt Modell (5) gerechnet. Dieses beschreibt Modell (4) und enthält daher auch alle Variablen des Modells, wird allerdings – wie in Kapitel 3.3.3 beschrieben – mit dem Befehl *gologit2* und der Ergänzung *autofit* durchgeführt. Dadurch wird ermöglicht, dass die Voraussetzung der proportionalen OR für jede unabhängige Variable geprüft und das Modell bei Nichterfüllung dementsprechend angepasst wird (Hoffmann 2016). Die Ergebnisse sind in Tabelle 15 als Modell (5) dargestellt. Der Unterschied

Effekte auf akademische Studierendenleistungen		
	Abhängige Variable: Noten (Skala von 1 bis 5) (5, bis SS20) (5, bis WS20/21)	
Corona_Indikator (<i>vor Corona</i>) <i>während Corona</i>	1 ^a 2,308* 2 ^a 1,721** 3 ^a n.s. 4 ^a 0,612**	1 ^a n.s. 2 ^a n.s. 3 ^a 0,564** 4 ^a 0,500**
Geschlecht (<i>männlich</i>) <i>weiblich</i>	1 ^a 0,805** 2 ^a 1,722** 3 ^a 0,805** 4 ^a 0,805**	1 ^a 0,826** 2 ^a 0,826** 3 ^a 0,826** 4 ^a 0,826**
Alter	1 ^a n.s. 2 ^a 1,026* 3 ^a 1,034** 4 ^a 1,041**	1 ^a n.s. 2 ^a 1,033** 3 ^a 1,041** 4 ^a 1,044**
Alter (quadriert)	1 ^a 0,998** 2 ^a 0,998** 3 ^a 0,998** 4 ^a 0,998**	1 ^a 0,998** 2 ^a 0,998** 3 ^a 0,998** 4 ^a 0,998**
Wahl_Pflicht (<i>Pflichtfach</i>) <i>Wahlfach</i>	n.s.	1 ^{a-3^a} n.s. 4 ^a 1,266*
<i>Wahlpflichtfach</i>	1 ^a 0,212** 2 ^a 0,328** 3 ^a 0,399** 4 ^a 0,441**	1 ^a 0,190** 2 ^a 0,299** 3 ^a 0,391** 4 ^a 0,450**
Studienrichtung (<i>Anderes</i>) <i>Lehramt</i>	1 ^a 0,212** 2 ^a 0,328** 3 ^a 0,399** 4 ^a 0,441**	1 ^a n.s. 2 ^a 0,390** 3 ^a 0,247** 4 ^a 0,283**
<i>Politikw.</i>	n.s.	n.s.
<i>Sprachw.</i>	n.s.	n.s.
<i>Wirtschaftsw.</i>	1 ^a 0,480** 2 ^a 0,738* 3 ^a 0,711** 4 ^a n.s.	1 ^a 0,397** 2 ^a 0,645* 3 ^a 0,673** 4 ^a n.s.
Pseudo-R-squared	0,0411	0,0421
p-Wert des Modells	< 0,001	< 0,001
Anzahl an Datenpunkten	4.793	5.251
VIF	1,37	1,38
Log-Likelihood Modell	-6.530,1	-7.255,5
Log-Likelihood nur Konstante	-6.810,1	-7.542,9
BIC	13.331,4	14.725,0
AIC	13.124,2	14.514,9

Tabelle 15: Resultate der generalisierten ordinalen logistischen Regression für Bachelorstudierende
Anmerkung: Diese Tabelle gibt die OR der Variablen des generalisierten ordinalen Logit-Modells an und entspricht daher nicht direkt den Koeffizienten. Die abhängige Variable (Noten) sind ordinal skaliert von 1 (Sehr gut) bis 5 (Nicht Genügend). Die *kursiv* angeführten Kategorien unter jeder unabhängigen Variablen geben an, auf welche Ausprägung sich das OR bezieht – die in Klammern angeführte Kategorie ist die entsprechende Referenzkategorie. Im generalisierten ordinalen Logit-Modell werden die Kategorien der abhängigen Variablen explizit einander gegenübergestellt. 1 entspricht Note 1 vs 2,3,4,5. 2 entspricht Note 1,2 vs Noten 3,4,5. 3 entspricht Noten 1,2,3 vs Noten 4,5. 4 entspricht Noten 1,2,3,4 vs 5.

**Signifikanz 1%-Level, *Signifikanz 5%-Level

zu den vorhergehenden Modellen (4) ist dabei klar ersichtlich – da Modell (5) für unterschiedliche OR in den Kategorien der abhängigen Variablen für jede unabhängige Variable erlaubt, werden diese auch für jede unabhängige Variable angeführt. Dies ergibt zwar einen umfangreicheren Modell-Output, erleichtert allerdings auch die Interpretation der OR.

Bereits bei der Variable *Corona_Indikator* ist der Unterschied zwischen den Modellen – und die Wichtigkeit der Unterscheidung der Modelle – ersichtlich: Während Modell (4) den gesamten Effekt der unabhängigen Variablen betrachtet, wird in Modell (5) klar, dass die Signifikanz nicht für alle Kategorien der abhängigen Variablen gilt. So ist etwa Kategorie 3, also die Gegenüberstellung der Wahrscheinlichkeit der Noten *Sehr Gut*, *Gut* und *Befriedigend* gegen alle anderen Noten (*Genügend*, *Nicht Genügend*) in Bezug auf die Variable *Corona_Indikator* im Datensatz bis zum Sommersemester 2020 für Bachelorstudierende nicht signifikant. Die anderen Kategorien sind jedoch sehr wohl signifikant – so beträgt die Wahrscheinlichkeit, eine schlechtere Note als ein *Sehr Gut* im Sommersemester 2020 zu erhalten, 2,3-mal die Wahrscheinlichkeit aus Vorsemestern und ist somit gestiegen. Umgekehrt ist die Wahrscheinlichkeit, im Sommersemester 2020 während der Corona-Pandemie auf eine Prüfung ein *Nicht Genügend* zu erhalten, für Bachelorstudierende nur 0,662-mal der Wahrscheinlichkeit aus den Vorsemestern. Im erweiterten Datensatz bis Wintersemester 2020/21 zeigt sich, dass nur noch die Vergleiche der letzten Kategorien signifikant bleiben – so beträgt die Wahrscheinlichkeit, ein *Genügend* oder ein *Nicht Genügend* in den Semestern zu erhalten, welche von der Corona-Pandemie beeinflusst wurden, nur 0,564-mal die Wahrscheinlichkeit aus Vorsemestern. Die Wahrscheinlichkeit, ein *Nicht Genügend* anstelle einer beliebigen anderen Note zu erhalten, sinkt sogar auf das 0,5-Fache der Wahrscheinlichkeit aus Vorsemestern.

Die Variable *Geschlecht* ist statistisch signifikant, interessanterweise sind die OR im erweiterten Datenmodell proportional in den unterschiedlichen Kategorien. Für das Modell basierend auf dem Datensatz bis Sommersemester 2020 ist etwa die Wahrscheinlichkeit, eine schlechtere (=höhere) Note als ein *Sehr Gut* zu erhalten, für weibliche Studierende nur 0,805-mal der Wahrscheinlichkeit von männlichen Studierenden. Die Wahrscheinlichkeit, eine höhere Note als ein *Gut* zu erhalten, ist jedoch bei weiblichen Studierenden 1,7-mal der Wahrscheinlichkeit von männlichen Studierenden. Im Modell basierend auf dem Datensatz bis Wintersemester 2020/21 lässt sich jedoch feststellen, dass die OR in allen Notenkategorien gleich sind – die Wahrscheinlichkeit, eine schlechtere Note zu erhalten, beträgt bei weiblichen Bachelorstudierenden nur 0,826-mal der Wahrscheinlichkeit von männlichen Bachelorstudierenden.

Der Effekt des Alters bleibt signifikant, ist jedoch sehr gering – während ein zusätzliches Lebensjahr die Wahrscheinlichkeit für eine schlechtere Note in beinahe allen Kategorien geringfügig erhöht, wird dieser Effekt durch eine Abschwächung, welche mithilfe der nichtlinearen

Variable gemessen wurde, moderiert. Für die Variable *Wahl_Pflicht* zeigt sich nun, dass einige Unterschiede nicht mehr signifikant sind: So gibt es etwa im Datensatz bis Sommersemester 2020 keinen signifikanten Unterschied in den erreichten Noten zwischen Prüfungen in Pflicht- und Wahlfächern. Auch im erweiterten Datensatz bis Wintersemester 2020/21 ist nur die letzte Kategorie, welche alle positiven Noten einem *Nicht Genügend* gegenüberstellt, auf dem 5%-Level signifikant. Der Effekt eines Wahlpflichtfachs im Vergleich zu einem Pflichtfach ist jedoch sehr groß – und außerdem in allen Kategorien der abhängigen Variable signifikant. So beträgt beispielsweise die Wahrscheinlichkeit, eine schlechtere Note als ein Sehr Gut zu erhalten, bei Prüfungen in Wahlpflichtfächern für Bachelorstudierende nur 0,212-mal (bis Sommersemester 2020) beziehungsweise nur 0,19-mal (bis Wintersemester 2020/21) der Wahrscheinlichkeit von Prüfungen in Pflichtfächern. Der Effekt schwächt sich zwar größtmäßig ab, jedoch beträgt die Wahrscheinlichkeit, ein *Nicht Genügend* auf eine Prüfung in einem Wahlpflichtfach zu erhalten, nur 0,44-mal (Sommersemester 2020) beziehungsweise 0,45-mal (Wintersemester 2020/21) die Wahrscheinlichkeit bei Prüfungen in Pflichtfächern.

Der Vergleich der unterschiedlichen Studienrichtungen liefert keine signifikanten Unterschiede zwischen den Ergebnissen von Studierenden der Politikwissenschaften, der Sprachwissenschaften und anderen Studienrichtungen. Bei Studierenden des Lehramts kann ein signifikanter Unterschied zu Studierenden aus anderen Studienrichtungen festgestellt werden, es wird jedoch darauf hingewiesen, dass sich die Effektgrößen in dem Modell bis Sommersemester 2020 und dem Modell bis Wintersemester 2020/21 sehr stark unterscheiden. Die Interpretation dieser Effekte sollte daher besonders vorsichtig erfolgen. Für Unterschiede in den erreichten Noten zwischen Studierenden der Wirtschaftswissenschaften und anderer Studienrichtungen kann festgehalten werden, dass die Effektgrößen nicht so stark variieren und die Effekte daher gefestigter zu sein scheinen. So beträgt etwa die Wahrscheinlichkeit, eine schlechtere Note als ein Sehr Gut auf eine Prüfung zu erhalten, für Bachelorstudierende der Wirtschaftswissenschaften nur 0,48-mal (Sommersemester 2020) beziehungsweise 0,397-mal (Wintersemester 2020/21) der Wahrscheinlichkeit von Bachelorstudierenden anderer Studienrichtungen.

Die Modellkennzahlen der Modelle (5) verbessern sich im Vergleich zu den Modellen (4) leicht, so beträgt der Wert des Pseudo- R^2 rund 4% und der Log-Likelihood-Wert des Modells verringert sich auf -7.255 (beziehungsweise -6.530). Auch AIC und BIC geben diesem Modell im Vergleich zu den Modellen (4) den Vorzug, da die Werte für die Modelle (5) am kleinsten sind. An dieser Stelle ist es auch wichtig zu erwähnen, dass das *gologit2*-Modell, also das generalisierte ordinale logistische Modell, im Falle einer einzelnen Beobachtung im Datensatz bis Wintersemester 2020/21 zu einer negativen vorhergesagten Wahrscheinlichkeit führt. Bereits McCullagh und Nelder (1989) haben festgestellt, dass dies bei einem generalisierten ordinalen

logistischen Modell nicht ausgeschlossen werden kann. Bei einem geringen Vorkommen stellt dies jedoch keinen Modellfehler dar (McCullagh und Nelder 1989). Da es nur eine der 5.251 Beobachtungen betrifft, wird betont, dass dies die Aussagekraft des Modells nicht beeinträchtigt. Williams (2016) hält jedoch fest, dass bei Verletzung der Annahmen des ordinalen logistischen Modells durch mehrere unabhängige Variablen auch die Verwendung eines multinomialen logistischen Modells in Betracht gezogen werden sollte. Die Berechnungen werden also, wie in Kapitel 3.3.4 beschrieben, erneut durchgeführt. Dies liegt daran, dass in dem hier verwendeten Modell jedenfalls die Mehrheit der unabhängigen Variablen die Voraussetzung der proportionalen OR in den Kategorien der abhängigen Variablen verletzt.

4.5 Resultate der multinomialen logistischen Regression für Bachelorstudierende

Die Ergebnisse der unterschiedlichen Modelle sind in Tabelle 16 dargestellt. Die Modelle (6), wiederum separat bis Sommersemester 2020 und bis Wintersemester 2020/21 betrachtet, beschreibt das Modell unter der Berücksichtigung von nichtlinearen Effekten der Variable *Alter*.

Die Werte, welche in Tabelle 16 zu finden sind, stellen die RRR der jeweiligen Modelle dar. Im Unterschied zu Modell (5) sind die Voraussetzungen für die multinomiale logistische Regression geringer und – wie in Kapitel 3.3.4 beschrieben – berücksichtigen die β -Werte des Modells sowohl die unabhängige Variable als auch die Kategorie der abhängigen Variablen. Alle Werte werden gegen die Referenzkategorie 5 (*Nicht Genügend*) analysiert und müssen auch dementsprechend interpretiert werden. Zur Interpretation der Werte aus den Modellen (6) ist für den Datensatz bis Sommersemester 2020 wie folgt auszuführen: Die Variable *Corona_Indikator* ist für zwei der vier errechneten Logit-Werte signifikant. So ist das relative Risiko der Kategorie 3 (*Befriedigend*) im Vergleich zur Referenzkategorie (*Nicht Genügend*) 2,259 ceteris paribus. Wenn also die Variable *Corona_Indikator* auf 1 gestellt wird (also in Semestern, welche von der Corona-Pandemie beeinflusst wurden), steigt das relative Risiko von *Befriedigend* zu *Nicht Genügend* um 1,259. Das bedeutet, dass das relative Risiko, mit *Befriedigend* anstatt *Nicht Genügend* beurteilt zu werden, um das 1,259-fache wahrscheinlicher ist. Das relative Risiko eines *Genügend* zu einem *Nicht Genügend* steigt in Zeiten der Corona-Pandemie um das 1,016-fache. In diesem Modell ist auch die Variable *Geschlecht* in zwei Kategorie signifikant. Das relative Risiko eines *Gut* zu einem *Nicht Genügend* steigt bei Frauen um das 1,511-fache im Vergleich zu Männern und jenes eines *Genügend* zu einem *Nicht Genügend* um das 1,259-fache. Das *Alter* hingegen ist nun in der Vergleichskategorie von *Sehr Gut* und *Nicht Genügend* nicht mehr signifikant. Für ein *Gut* beträgt das relative Risiko im Vergleich zu einem *Nicht Genügend* bei einem zusätzlichen Altersjahr nur mehr das 0,949-fache, für ein *Befriedigend* sinkt das relative Risiko auf das 0,940-fache und für ein *Genügend* auf das 0,973-fache;

Effekte auf akademische Studierendenleistungen		
	Abhängige Variable: Noten (Skala von 1 bis 5) (6, bis SS20) (6, bis WS20/21)	
Corona_Indikator (<i>vor Corona</i>) <i>während Corona</i>	1 ^a -2 ^a n.s. 3 ^a 2,259** 4 ^a 2,016**	1 ^a 1,585* 2 ^a 1,964** 3 ^a 2,697** 4 ^a 1,902**
Geschlecht (<i>männlich</i>) <i>weiblich</i>	1 ^a n.s. 2 ^a 1,511** 3 ^a n.s. 4 ^a 1,259**	1 ^a n.s. 2 ^a 1,417** 3 ^a n.s. 4 ^a 1,268**
Alter	1 ^a n.s. 2 ^a 0,949** 3 ^a 0,940** 4 ^a 0,973**	1 ^a n.s. 2 ^a 0,935** 3 ^a 0,939** 4 ^a n.s.
Alter (quadriert)	1 ^a n.s. 2 ^a 1,002* 3 ^a 1,002** 4 ^a n.s.	1 ^a n.s. 2 ^a 1,003** 3 ^a 1,003** 4 ^a n.s.
Wahl_Pflicht (<i>Pflichtfach</i>) <i>Wahlfach</i> <i>Wahlpflichtfach</i>	1 ^a -3 ^a n.s. 4 ^a 0,679** 1 ^a 6,339** 2 ^a 3,013** 3 ^a 1,560* 4 ^a 1,434*	1 ^a n.s. 2 ^a 0,687* 3 ^a n.s. 4 ^a 0,620** 1 ^a 6,689** 2 ^a 3,247** 3 ^a n.s. 4 ^a n.s.
Studienrichtung (<i>Anderes</i>) <i>Lehramt</i> <i>Politikw.</i> <i>Sprachw.</i> <i>Wirtschaftsw.</i>	1 ^a 3,479** 2 ^a 6,455** 3 ^a 5,408** 4 ^a 1,916** n.s. n.s. 1 ^a 2,265** 2 ^a n.s. 3 ^a n.s. 4 ^a 1,433*	1 ^a 3,575** 2 ^a 5,980** 3 ^a 4,994** 4 ^a 1,896** n.s. n.s. 1 ^a 2,450** 2 ^a n.s. 3 ^a n.s. 4 ^a n.s.
Pseudo-R-squared	0,0417	0,0427
p-Wert des Modells	< 0,001	< 0,001
Anzahl an Datenpunkten	4.793	5.251
VIF	1,37	1,38
Log-Likelihood Modell	-6.526,2	-7.220,8
Log-Likelihood nur Konstante	-6.810,1	-7.542,9
BIC	13.425,3	14.818,4
AIC	13.140,4	14.529,5

Tabelle 16: Resultate der multinomialen logistischen Regression für Bachelorstudierende

Anmerkung: Diese Tabelle gibt die RRR der Variablen des multinomialen Logit-Modells an und entspricht daher nicht direkt den Koeffizienten. Die abhängige Variable (Noten) sind ordinal skaliert von 1 (Sehr gut) bis 5 (Nicht Genügend). Die *kursiv* angeführten Kategorien unter jeder unabhängigen Variablen geben an, auf welche Ausprägung sich das RRR bezieht – die nicht angeführte Kategorie ist die entsprechende Referenzkategorie. **Signifikanz 1%-Level, *Signifikanz 5%-Level

^aIm multinomialen Logit-Modell wird die höchste Kategorie (=5, Nicht Genügend) als Referenzkategorie gewählt.

Dies jeweils bei einem zusätzlichen Lebensjahr. Auch hier muss festgehalten werden, dass der Effekt durch die Nichtlinearität abgeschwächt wird.

Die meisten Kategorien bei Wahlfächern im Vergleich zu Pflichtfächern sind nicht signifikant, jedoch ist das Risiko eines *Genügend* zu einem *Nicht Genügend* in einer Prüfung, welche in einem Wahlfach abgelegt wird, das 0,679-fache Risiko einer Prüfung in einem Pflichtfach. Für Wahlpflichtfächer sind in allen Kategorien signifikante Effekte erkennbar – so ist etwa das relative Risiko, mit *Sehr Gut* anstatt *Nicht Genügend* beurteilt zu werden, 5.339-fach wahrscheinlicher bei Wahlpflichtfächern als bei Pflichtfächern.

Bei der Variable *Studienrichtung* sind wiederum Politikwissenschaften und Sprachwissenschaften in allen Kategorien nicht signifikant. Für Lehramtsstudierende jedoch lässt sich festhalten, dass das relative Risiko eines *Sehr Gut* im Vergleich zu einem *Nicht Genügend* das 3,479-fache desselben Risikos bei Studierenden anderer Studienrichtungen beträgt. Dieser positive Effekt lässt sich über alle Notenkategorien hinweg feststellen. Bei Studierenden der Wirtschaftswissenschaften sind nur das relative Risiko eines *Sehr Gut* zu einem *Nicht Genügend* und eines *Genügend* zu einem *Nicht Genügend* signifikant. Für Bachelorstudierende der Wirtschaftswissenschaften lässt sich auch festhalten, dass das relative Risiko eines *Sehr Gut* im Vergleich zu einem *Nicht Genügend* das 2,265-fache desselben Risikos bei Studierenden anderer Studienrichtungen beträgt.

Der Pseudo- R^2 -Wert beträgt 0,0417 und ist damit immer noch gering. Der p-Wert der Wald-Statistik ist jedoch signifikant und zeigt an, dass das Modell erklärende Information über die abhängige Variable beinhaltet. Der Log-Likelihood-Wert des Modells (6) bis Sommersemester 2020 beträgt -6.526, was beinahe gleich gut ist wie der Wert des generalisierten ordinalen logistischen Modells.

Bei Modell (6) bis Wintersemester 2020/21 wird nun der erweiterte Datensatz analysiert, dabei sind nun alle Kategorien bei der Variable *Corona_Indikator* signifikant: Das relative Risiko eines *Sehr Gut* zu einem *Nicht Genügend* steigt in Zeiten der Corona-Pandemie auf das 1,585-fache, jenes eines *Gut* auf das 1,964-fache, eines *Befriedigend* auf das 2,697-fache und eines *Genügend* auf das 1,902-fache. Generell lässt sich daher sagen, dass bei Prüfungen, welche während der Corona-Pandemie abgelegt wurden, eher erwartet wird, dass Studierende eine bessere Note als ein *Nicht Genügend* erhalten. In dem erweiterten Modell bleibt die Variable *Geschlecht* in zwei Kategorien signifikant: Das relative Risiko eines *Gut* zu einem *Nicht Genügend* steigt bei Frauen auf das 1,417-fache im Vergleich zu Männern und jenes eines *Genügend* zu einem *Nicht Genügend* auf das 1,268-fache. Beim Alter bleibt die Effektgröße beinahe gleich, nur in der Kategorie 4 (Gegenüberstellung *Genügend* zu *Nicht Genügend*) ist nun kein

signifikanter Effekt feststellbar. Auch hier wird der Effekt des Alters wieder durch den nichtlinearen Einfluss moderiert. Eine ähnliche Beobachtung ergibt sich bei der Variable *Studienrichtung*, da auch hier die Signifikanz und Effektgrößen sehr ähnlich bleiben. Ein Gegensatz zu dem Modell basierend auf dem Datensatz bis Sommersemester 2020 ist die Variable *Wahl_Pflicht*: Im erweiterten Modell ist für den Vergleich von Wahl- und Pflichtfach auch die Kategorie 2 (Gegenüberstellung *Gut* zu *Nicht Genügend*) signifikant. Für die Gegenüberstellung von Wahlpflicht- und Pflichtfach hingegen sind nun nur mehr die ersten beiden Kategorien signifikant, die Effektgrößen bleiben jedoch ähnlich wie im kleineren Modell.

Der Pseudo- R^2 -Wert beträgt 0,0427, der p-Wert der Wald-Statistik ist signifikant. Der Log-Likelihood-Wert des Modells (6) bis Wintersemester 2020/21 beträgt -7.221, etwas besser als der Wert des generalisierten ordinalen logistischen Modells.

5 Diskussion der Resultate

5.1 Conclusio

Um den Einfluss der Corona-Pandemie auf Studierendenleistungen festzustellen, wurde im Rahmen dieser Arbeit einerseits eine theoretische Untersuchung der Einflüsse auf das Lernumfeld von Studierenden vorgenommen, andererseits wurden vorhandene Daten statistisch analysiert. Dazu wurde die logistische Regression als Methode ausgewählt, wobei unterschiedliche abhängige Variablen zu unterschiedlichen Varianten der logistischen Regression führen – einerseits wird eine binäre logistische Regression zur Analyse der Erfolgswahrscheinlichkeit durchgeführt, andererseits werden auch die Noten als abhängige Variable im Rahmen einer ordinalen beziehungsweise multinomialen logistischen Regression analysiert.

In der theoretischen Untersuchung des Einflusses der Corona-Pandemie auf Studierendenleistungen an der Universität Wien ist klar ersichtlich, dass sich das Lernumfeld von Studierenden verschlechtert hat: Gestiegene psychische Belastungen aufgrund der Pandemie-Situation, verringerte Möglichkeiten zur sozialen Einbindung von Studierenden und zum Aufbau von Kontakten an der Universität sowie komplexere Anforderungen und Herausforderungen in der Onlinelehre lassen auf eine Verschlechterung der akademischen Studierendenleistungen schließen.

In der empirischen Untersuchung der zur Verfügung gestellten Daten konnte jedoch das Gegenteil festgestellt werden – die Erfolgswahrscheinlichkeit von Bachelorstudierenden bei Prüfungen ist in den Semestern, welche von der Corona-Pandemie beeinflusst wurden, signifikant angestiegen. Dies betrifft einerseits die Wahrscheinlichkeit, eine Prüfung zu bestehen (also die reine Unterscheidung in Bestehen und Nicht-Bestehen einer Prüfung), andererseits jedoch

auch die erreichten Noten in einer Prüfung. Im generalisierten logistischen Modell sowie auch im multinomialen Modell lässt sich feststellen, dass die Wahrscheinlichkeit für bessere Noten in Zeiten der Corona-Pandemie ebenfalls gestiegen ist. Aufgrund der starken Beeinflussung des Modells wurde dieses für Bachelor- und Masterstudierende getrennt berechnet. Eine geringe Anzahl an Datenpunkten führte zu Inkonsistenzen bei dem generalisierten ordinalen und dem multinomialen logistischen Modell für Masterstudierende, sodass für diese nur das binäre logistische Modell ausgeführt wurde. Der signifikante Effekt wurde dabei nur für Bachelorstudierende festgestellt, wobei der Datensatz von diesen auch um ein Vielfaches umfangreicher ist als jener der Masterstudierenden. Weiters wird festgehalten, dass die Ergebnisse dieser Modelle für die Resultate von Erstantritten zu LV-Prüfungen gültig sind und – wie auch bereits ausgeführt – nur die Resultate von Prüfungen, welche von Frau ao. Univ.-Prof. Mag. Dr. Schaffhauser-Linzatti als (Einzel-)Prüferin durchgeführt wurden, betrachtet.

Diese Einschränkung stellt zwar einerseits einen Nachteil im Bereich der Generalisierbarkeit dar, andererseits kann somit leichter auf die Umstellungen in Zeiten der Corona-Pandemie eingegangen werden. So lassen sich etwa nach Auskunft von Frau ao. Univ.-Prof. Mag. Dr. Schaffhauser-Linzatti die Ergebnisse auch dadurch erklären, dass im Rahmen der Onlinelehre der Prüfungsmodus leicht verändert wurde, um auf die Bedürfnisse von Studierenden besser einzugehen. Genau diese Flexibilität von Lehrenden und Studierenden ist ein Punkt, welcher von Nebel (2017) auch als Voraussetzung für erfolgreiche Onlinelehre angesprochen wird.

5.2 Limitationen und weitere Forschungsmöglichkeiten

Wie bereits unter Kapitel 5.1 angesprochen, ist eine klare Limitation dieser Arbeit die fehlende Generalisierbarkeit der Ergebnisse. Da im Rahmen der Umstellung auf Onlinelehre jeder Lehrende gefordert war, sich darauf einzustellen und auch die Kursmodalitäten im Wintersemester 2020/21 sowie im Sommersemester 2020 sehr unterschiedlich ausgestaltet waren, können die Ergebnisse der Betrachtung einer einzelnen Prüferin nicht auf andere Prüfer und Lehrende angewandt werden. Gleichzeitig stellt diese Limitation jedoch auch eine vielversprechende weiterführende Forschungsmöglichkeit dar, mit welcher umfangreichere Datensätze analysiert werden könnten.

Eine weitere Limitation ist das fehlende direkte Feedback von Studierenden. Die Daten dieser Arbeit wurden von der Abteilung für Datenmanagement der Universität Wien geliefert und betrachten tatsächlich nur die Leistungen der Studierenden bei Prüfungen. Es wäre auch durchaus interessant gewesen, weitere Daten primär bei Studierenden abzufragen und in das Modell zu inkludieren – dadurch hätte sich eventuell auch die statistische Aussagekraft der Mo-

delle verbessert. Da die Prüfungsleistungen bereits im Sommersemester 2020 und im Wintersemester 2020/21 erbracht wurden, ist es im Nachhinein nicht mehr möglich, Studierende zu befragen. Im Rahmen der Corona-Pandemie wurde die Situation jedoch von vielen Lehrenden dazu genutzt, auch Feedback von Studierenden einzuholen – hier ergibt sich die Möglichkeit, Ergebnisse von solchen Befragungen mit Prüfungsleistungen zu verknüpfen und weiter zu analysieren.

Literaturverzeichnis

- Agresti, Alan (2010): Analysis of ordinal categorical data. John Wiley & Sons.
- Angrist, Joshua D.; Pischke, Jörn-Steffen (2008): Mostly harmless econometrics: An empiricist's companion. Princeton University Press.
- Alqudah, Noor M.; Jammal, Hisham M.; Saleh, Omar; Khader, Yousef; Obeidat, Nail; Alqudah, Jumana (2020): Perception and experience of academic Jordanian ophthalmologists with E-Learning for undergraduate course during the COVID-19 pandemic. In: *Annals of medicine and surgery* (2012) 59, S. 44–47.
- Artino, Anthony R.; Stephens, Jason M. (2009): Academic motivation and self-regulation: A comparative analysis of undergraduate and graduate students learning online. In: *The Internet and Higher Education* 12 (3-4), S. 146–151.
- Backhaus, Klaus; Erichson, Bernd; Plinke, Wulff; Weiber, Rolf (2018): Regressionsanalyse. In *Multivariate Analysemethoden*, S. 57-124. Springer Gabler, Berlin, Heidelberg.
- Bandura, Albert (1977): Social learning theory. 1. Aufl. Englewood cliffs (Prentice-Hall series in social learning theory).
- Beattie, Irene R.; Thiele, Megan (2016): Connecting in Class?: College Class Size and Inequality in Academic Social Capital. In: *The Journal of Higher Education* 87 (3), S. 332–362.
- Boerner, Sabine; Seeber, Günther; Keller, Helmut; Beinborn, Peter (2005): Lernstrategien und Lernerfolg im Studium. Zur Validierung des LIST bei berufstätigen Studierenden. In: *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie* 37 (1), S. 17–26.
- Boud, David; Cohen, Ruth; Sampson, Jane (1999): Peer Learning and Assessment. In: *Assessment & Evaluation in Higher Education* 24 (4), S. 413–426.
- Braun, Edith; Hannover, Bettina (2009): Zum Zusammenhang zwischen Lehr-Orientierung und Lehr-Gestaltung von Hochschuldozierenden und subjektivem Kompetenzzuwachs bei Studierenden. In: *Perspektiven der Didaktik*, S. 277-291. VS Verlag für Sozialwissenschaften.
- Braun, Marlies; Kirchner, Stefanie; Niederkrotenthaler, Thomas; Till, Benedikt (2020): Sars CoV-2: Mental Health in Österreich. Ausgewählte Ergebnisse zur sechsten Befragungswelle, 07.08.-22.08.2020. Hg. v. Unit Suizidforschung und Mental Health Promotion, Abteilung für Sozial- und Präventivmedizin, Zentrum für Public Health, Medizinische Universität Wien. Online verfügbar unter <http://www.suizidforschung.at/wp-content/uploads/2020/10/Bericht-Welle-6.pdf>, zuletzt aktualisiert am August 2020, zuletzt geprüft am 07.01.2021.

- Callender, Claire (2008): The impact of term-time employment on higher education students' academic attainment and achievement. In: *Journal of Education Policy* 23 (4), S. 359–377.
- Carolan, Brian V. (2013): Social network analysis and education. Theory, methods & applications. Los Angeles, Calif.: Sage Publ.
- Chambliss, Daniel F.; Takacs, Christopher G. (2014): How college works. Cambridge, Mass: Harvard University Press.
- Chamorro-Premuzic, Tomas; Furnham, Adrian (2004): A possible model for understanding the personality-intelligence interface. In: *British Journal of Psychology* 95(2), S. 249-264.
- Chickering, Arthur Wright (1969): Education and identity. San Francisco: Jossey-Bass (Jossey-Bass series in higher education).
- Chu, Angela Hsin Chun; Choi, Jin Nam (2005): Rethinking procrastination: positive effects of "active" procrastination behavior on attitudes and performance. In: *The Journal of social psychology* 145 (3), S. 245–264.
- Cohen, Alison K.; Hoyt, Lindsay T.; Dull, Brandon (2020): A Descriptive Study of COVID-19-Related Experiences and Perspectives of a National Sample of College Students in Spring 2020. In: *Journal of Adolescent Health* 67 (3), S. 369–375.
- Christensen, Björn; Papies, D., Proppe, D.; Clement, M. (2014): Gütemaße der logistischen Regression bei unbalancierten Stichproben. In: *WiSt-Wirtschaftswissenschaftliches Studium* 43 (4), S. 211-213.
- Denham, Bryan E. (2017): Categorical Statistics for Communication Research. 1. Auflage. Malden, Oxford: John Wiley & Sons, Inc.
- Dyrbye, Liselotte N.; Thomas, Matthew R.; Shanafelt, Tait D. (2006): Systematic Review of Depression, Anxiety, and Other Indicators of Psychological Distress Among U.S. and Canadian Medical Students. In: *Academic Medicine* 81 (4), S. 354–373.
- Fassmann, Heinz (2020): Umgang des Bildungssystems mit dem Coronavirus - Erlass. Online verfügbar unter https://www.bmbwf.gv.at/dam/jcr:224ef638-49dc-44b1-b58e-e6420a8e11bd/corona_erlass.pdf, zuletzt aktualisiert am 12.03.2020, zuletzt geprüft am 10.01.2021.
- Ferrari, Joseph R.; Johnson, Judith L.; McCown, William G. (1995): Procrastination Research. In: Joseph R. Ferrari, Judith L. Johnson und William G. McCown (Hg.): Procrastination and Task Avoidance. Boston, MA: Springer US, S. 21–46.

Ferrari, Joseph R.; Tice, Dianne M. (2000): Procrastination as a Self-Handicap for Men and Women: A Task-Avoidance Strategy in a Laboratory Setting. In: *Journal of Research in Personality* 34 (1), S. 73–83.

Field, Andy; Miles, Jeremy; Field, Zoe (2012): *Discovering statistics using R*. Sage publications.

Fita, Ana; Monserrat, Jose F.; Moltó, Germán; Mestre, Eva M.; Rodriguez-Burruezo, Adrián (2016): Use of Synchronous e-Learning at University Degrees. In: *Comput Appl Eng Educ* 24 (6), S. 982–993.

Flöther, Choni; Janson, Kerstin; Krücken, Georg; Müller, Lars; Plasa, Tim Niels; Teichler, Ulrich; Wolf, Vera (2015): *Generation Hochschulabschluss: Vielfältige Perspektiven auf Studium und Berufseinstieg*. Münster: Waxmann.

Furnham, Adrian; Chamorro-Premuzic, Tomas; McDougall, Fiona (2003): Personality, cognitive ability, and beliefs about intelligence as predictors of academic performance. In: *Learning and Individual Differences* 14(1), S. 47-64.

Goda, Yoshiko; Yamada, Masanori; Kato, Hiroshi; Matsuda, Takeshi; Saito, Yutaka; Miyagawa, Hiroyuki (2015): Procrastination and other learning behavioral types in e-learning and their relationship with learning outcomes. In: *Learning and Individual Differences* 37, S. 72–80.

Gold, A. (1988). *Studienabbruch, Abbruchneigung und Studienerfolg: Vergleichende Bedingungsanalysen des Studienverlaufs*. Lang.

Hakami, Zaki; Khanagar, Sanjeev B.; Vishwanathaiah, Satish; Hakami, Abrar; Bokhari, Ahmed M.; Jabali, Ahmad H. et al. (2020): Psychological impact of the coronavirus disease 2019 (COVID-19) pandemic on dental students: A nationwide study. In: *Journal of dental education*, S. 1–10.

Hall, Ralph (2010): The work–study relationship: experiences of full-time university students undertaking part-time employment. In: *Journal of Education and Work* 23 (5), S. 439–449.

Hanbidge, Alice S., Mackenzie, Amanda, Sanderson, Nicole, Scholz, Kyle, & Tin, Tony (2017): Academic Success Foundation: Enhancing Academic Integrity through Mobile Learning. *International Association for Development of the Information Society*.

Hao, Yungwei (2016): Exploring undergraduates' perspectives and flipped learning readiness in their flipped classrooms. In: *Computers in Human Behavior* 59, S. 82–92.

Hasan, Najmul; Bao, Yukun (2020): Impact of “e-Learning crack-up” perception on psychological distress among college students during COVID-19 pandemic: A mediating role of “fear of academic year loss”. In: *Children and youth services review* 118, S. 1–9.

Hattie, John (2009): Visible learning. A synthesis of over 800 meta-analyses relating to achievement. 1. publ: Routledge (Educational research).

Hattie, John (2015): The applicability of Visible Learning to Higher Education. In: *Scholarship of Teaching and Learning in Psychology* 1 (1), S. 79–91.

Helm, Christoph (2017): Effects of social learning networks on student academic achievement and pro-social behavior in accounting. Effekte sozialer Lernnetzwerke auf Schülerleistungen und das Sozialverhalten im Fach Rechnungswesen. In: *Journal for Education Research Online* (9), S. 52–76.

Hiltz, Starr Roxanne (1994): The virtual classroom. Learning without limits via computer networks. Norwood, NJ: Ablex Publ. Corp (Human/computer interaction).

Hoffmann, John P. (2016): Regression Models for Categorical, Count, and Related Variables: An Applied Approach. University of California Press.

Hofmann, Frank-Hagen (2010): Kreativität und Krise. Heidelberg, Univ., Diss., 2010. Universität Heidelberg, Heidelberg.

Hofmann, Frank-Hagen; Sperth, Michael; Holm-Hadulla, Rainer Matthias (2017): Psychische Belastungen und Probleme Studierender. Entwicklungen, Beratungs- und Behandlungsmöglichkeiten. In: *Psychotherapeut* 62 (5), S. 395–402.

Husky, Mathilde M.; Kovess-Masfety, Viviane; Swendsen, Joel D. (2020): Stress and anxiety among university students in France during Covid-19 mandatory confinement. In: *Comprehensive psychiatry* 102, S. 1–3.

Imbens, Guido W.; Lemieux, Thomas (2008). Regression discontinuity designs: A guide to practice. In: *Journal of Econometrics* 142(2), S. 615-635.

Isensee, Fanny; Wolter, Andrä (2017): Nicht-traditionelle Studierende in internationaler Perspektive. Eine vergleichende Untersuchung. In: *Zeitschrift Hochschule und Weiterbildung* (1), S. 13–23.

Jirjahn, Uwe (2007): Welche Faktoren beeinflussen den Erfolg im wirtschaftswissenschaftlichen Studium? In: *Schmalenbachs Zeitschrift für betriebswirtschaftliche Forschung : ZfbF* 59 (3), S. 286–313.

Karall, Eva Maria (2020): Digitales Prüfen. Universität Wien. Online verfügbar unter <https://wiki.univie.ac.at/pages/viewpage.action?pageId=105517385>, zuletzt aktualisiert am 04.12.2020, zuletzt geprüft am 07.01.2021.

Karall, Eva Maria; Mitterauer, Lukas; Zwiauer, Charlotte (2020): Distance Learning - Lessons Learned. Transformation der Lehre in digitale Räume - vom Schock zur Selbstverständlichkeit? Universität Wien. Online verfügbar unter https://www.bmbwf.gv.at/dam/jcr:43d4f32e-2b62-4ad9-80d8-b8c17e06486d/Pr%C3%A4sentation_der_Uni_Wien_von_Eva_Karall_ZID,_Lukas_Mitterauer_BEfQ_Charlotte_Zwiauer_CTL.pdf, zuletzt aktualisiert am 10.09.2020, zuletzt geprüft am 10.01.2021.

Karall, Eva Maria; Pfeiffer, Daniel (2020): Vergleich: Collaborate - Collaborate 250+ - BBB - Jitsi - MS Teams - u:stream - Kaltura. Universität Wien. Online verfügbar unter <https://wiki.univie.ac.at/pages/viewpage.action?pageId=101498793>, zuletzt aktualisiert am 30.10.2020, zuletzt geprüft am 07.01.2021.

Klenke, Kira (2018). Studieren kann man lernen: Mit weniger Mühe zu mehr Erfolg. (5. Aufl.) Wiesbaden: Springer Fachmedien.

Komlos, John; Süßmuth, Bernd (2010): Empirische Ökonomie: Eine Einführung in Methoden und Anwendungen. Springer-Verlag.

Krause, Ulrike-Marie; Stark, Robin; Mandl, Heinz (2009): The effects of cooperative learning and feedback on e-learning in statistics. In: *Learning and Instruction* 19 (2), S. 158–170.

Leenhartsberger, Simone (2020): Kontakt zu Großeltern als Gratwanderung. ORF.at. Online verfügbar unter <https://orf.at/stories/3163605/>, zuletzt aktualisiert am 04.05.2020, zuletzt geprüft am 10.01.2021.

Lim, Grace Y.; Tam, Wilson W.; Lu, Yanxia; Ho, Cyrus S.; Zhang, Melvyn W.; Ho, Roger C. (2018): Prevalence of Depression in the Community from 30 Countries between 1994 and 2014. In: *Scientific reports* 8 (1), S. 1–10.

Lohbeck, Annette; Hagenauer, Gerda; Mühlig, Andrea; Moschner, Barbara; Gläser-Zikuda, Michaela (2017): Prokrastination bei Studierenden des Lehramts und der Erziehungswissenschaften. In: *Z Erziehungswiss* 20 (3), S. 521–536.

Lu, June; Yang, Jun; Yu, Chun-Sheng (2013): Is social capital effective for online learning? In: *Information & Management* 50 (7), S. 507–522.

Luthans, Brett Carl; Luthans, Kyle William; Jensen, Susan M. (2012): The Impact of Business School Students' Psychological Capital on Academic Performance. In: *Journal of Education for Business* 87 (5), S. 253–259.

- Luthans, Fred; Youssef, Carolyn M.; Avolio, Bruce J. (2007): Psychological capital. Developing the human competitive edge. Oxford, New York: Oxford University Press.
- Mayer, Richard E. (1998): Cognitive, metacognitive, and motivational aspects of problem solving. In: *Instructional science* 26(1), S. 49-63.
- McCullagh, P.; Nelder, J. A. (1989): Binary data. In: *Generalized linear models*, S. 98-148. Springer US.
- Metcalfe, Hilary (2003): Increasing Inequality in Higher Education: The role of term-time working. In: *Oxford Review of Education* 29 (3), S. 315–329.
- Michinov, Nicolas; Brunot, Sophie; Le Bohec, Olivier; Juhel, Jacques; Delaval, Marine (2011): Procrastination, participation, and performance in online learning environments. In: *Computers & Education* 56 (1), S. 243–252.
- Mitterauer, Lukas; Haidinger, Gerald; Frischenschlager, Oskar (2012): Prädiktoren des Studienabschlusses im 2002 reformierten Curriculum der Medizinischen Universität Wien. In: *Wien Med Wochenschr* 162, S. 74–88.
- Müller, Lars; Braun, Edith (2018): Student Engagement. In: *Z Erziehungswiss* 21 (3), S. 649–670.
- Nebel, Eckart (2017): Möglichkeiten und Herausforderungen akademischer Lernprozesse in Online-Studiengängen. In: Hedwig Rosa Griesehop und Edith Bauer (Hg.): *Lehren und Lernen online*. Wiesbaden: Springer Fachmedien Wiesbaden, S. 55–66.
- Neeß, Christina (2015): Worauf achten Arbeitgeber im Auswahlprozess von Absolventen wirtschaftswissenschaftlicher Studiengänge? Ergebnisse eines faktoriellen Surveys. In: *Journal for Labour Market Research* 48(4), S.305-323.
- O'Connor, Melissa C.; Paunonen, Sampo V. (2007): Big Five personality predictors of post-secondary academic performance. In: *Personality and Individual Differences* 43 (5), S. 971–990.
- Owens, Meghan R.; Brito-Silva, Francilia; Kirkland, Tracie; Moore, Carolyn E.; Davis, Kathleen E.; Patterson, Mindy A. et al. (2020): Prevalence and Social Determinants of Food Insecurity among College Students during the COVID-19 Pandemic. In: *Nutrients* 12 (9), S. 1–17.
- Pascarella, Ernest T.; Terenzini, Patrick T. (1991): How college affects students. Findings and insights from twenty years of research. 1st ed. San Francisco: Jossey-Bass Publishers (The Jossey-Bass higher and adult education series).
- Pfaff, Elizabeth; Huddleston, Patricia (2003): Does it matter if I hate teamwork? What impacts student attitudes toward teamwork. In: *Journal of Marketing Education* 25 (1), S. 37-45.

Prosser, Michael; Trigwell, Keith (2006): Confirmatory factor analysis of the approaches to teaching inventory. In: *British Journal of Educational Psychology* 76(2), S. 405-419.

Radovan, Marko; Makovec, Danijela (2015): Relations between students' motivation, and perceptions of the learning environment. Povezave med motiviranostjo studentov in zaznavanjem ucnega okolja. In: *Center for Educational Policy Studies Journal* 5 (2), S. 115–138.

Rakes, Glenda; Dunn, Karee (2009): The Impact of Online Graduate Students' Motivation and Self-Regulation on Academic Procrastination. In: *Journal of Interactive Online Learning* 9 (1), S. 78–93.

Redaktionsnetzwerk Deutschland (2020): Umfrage: 40 Prozent der Studenten haben wegen Corona-Krise Job verloren. Online verfügbar unter <https://www.rnd.de/politik/studenten-in-corona-krise-40-prozent-haben-ihren-job-verloren-R6IJD2ROBQVQEB6FH5KFJ3EHKM.html>, zuletzt aktualisiert am 08.06.2020, zuletzt geprüft am 09.01.2021.

Rektorat Universität Wien (2020a). Aktuelle Informationen / current information. [E-Mail]. [rektorat@univie.ac.at; 08.04.2020]

Rektorat Universität Wien (2020b). Informationen zum Semesterstart. Aktuelle Regelungen unter COVID-19 Bedingungen. [E-Mail]. [rektorat@univie.ac.at; 25.09.2020]

Rektorat Universität Wien (2020c). Unibetrieb ab 3. November. [E-Mail]. [rektorat@univie.ac.at; 03.11.2020]

Rizzuto, Tracey E.; LeDoux, Jared; Hatala, John Paul (2009): It's not just what you know, it's who you know: Testing a model of the relative importance of social networks to academic performance. In: *Soc Psychol Educ* 12 (2), S. 175–189.

Ryan, Martin; Barns, Angela; McAuliffe, Donna (2011): Part-time Employment and Effects on Australian Social Work Students: A Report on a National Study. In: *Australian Social Work* 64 (3), S. 313–329.

Salari, Nader; Hosseinian-Far, Amin; Jalali, Rostam; Vaisi-Raygani, Aliakbar; Rasoulpoor, Shna; Mohammadi, Masoud et al. (2020): Prevalence of stress, anxiety, depression among the general population during the COVID-19 pandemic: a systematic review and meta-analysis. In: *Globalization and health* 16 (1), S. 1–11.

Sallee, Margaret W.; Tierney, William G. (2007): The Influence of Peer Groups on Academic Success. In: *College and university* 82, S. 7–14.

Schiefele, Ulrich; Wild, Klaus Peter (1994): Lernstrategien im Studium : Ergebnisse zur Faktorenstruktur und Reliabilität eines neuen Fragebogens. In: *Zeitschrift für Differentielle und Diagnostische Psychologie* 15 (4), S. 185–200.

Schober, Barbara; Lüftenegger, Marko; Spiel, Christiane (2020): Lernen unter COVID-19-Bedingungen. Erste Ergebnisse - Studierende. Fakultät für Psychologie, Universität Wien. Online verfügbar unter https://lernencovid19.univie.ac.at/fileadmin/user_upload/p_lernencovid19/Zwischenergebnisse_Studierende.pdf, zuletzt aktualisiert am 24.04.2020.

Schouwenburg, Henri C. (2004): Procrastination in Academic Settings: General Introduction. In: Henri C. Schouwenburg, Clarry H. Lay, Timothy A. Pynchyl und Joseph R. Ferrari (Hg.): *Counseling the procrastinator in academic settings*. Washington: American Psychological Association, S. 3–17.

Schraw, Gregory; Wadkins, Theresa; Olafson, Lori (2007): Doing the things we do: A grounded theory of academic procrastination. In: *Journal of Educational Psychology* 99 (1), S. 12–25.

Seeber, Günther; Boerner, Sabine; Keller, Helmut; Beinborn, Peter (2006): Strategien selbstorganisierten Lernens bei berufstätigen Studierenden. Ausgewählte Ergebnisse einer empirischen Untersuchung. In: *Journal für Sozialwissenschaften und ihre Didaktik* 2.

Stata (2021a): mlogit postestimation - Postestimation tools for mlogit. Online verfügbar unter <https://www.stata.com/manuals/rmlogitpostestimation.pdf>, zuletzt geprüft am 09.03.2021

Stata (2021b): ologit - Ordered logistic regression. Online verfügbar unter <https://www.stata.com/manuals15/rologit.pdf>, zuletzt geprüft am 09.03.2021.

Stock, James H., and Mark Watson. (2011): *Dynamic factor models*. Oxford Handbooks Online.

Tee, Michael L.; Tee, Cherica A.; Anlacan, Joseph P.; Aligam, Katrina Joy G.; Reyes, Patrick Wincy C.; Kuruchittham, Vipat; Ho, Roger C. (2020): Psychological impact of COVID-19 pandemic in the Philippines. In: *Journal of affective disorders* 277, S. 379–391.

Thaler, Selina (2020): Coronavirus: Finanzielle Hilfe für Studierende in Notlage. Online verfügbar unter <https://www.derstandard.at/story/2000116103278/coronavirus-finanzielle-hilfe-fuer-studierende-in-notlage>, zuletzt aktualisiert am 24.03.2020, zuletzt geprüft am 09.01.2020.

UCLA (2021a): Lesson 3 Logistic Regression Diagnostics. Online verfügbar unter <https://stats.idre.ucla.edu/stata/webbooks/logistic/chapter3/lesson-3-logistic-regression-diagnostics-2/>, zuletzt geprüft am 09.03.2021

UCLA (2021b): Multinomial Logistic Regression | STATA annotated output. Online verfügbar unter <https://stats.idre.ucla.edu/stata/output/multinomial-logistic-regression-2/>, zuletzt geprüft am 09.03.2021

Veall, Michael R.; Zimmermann, Klaus F. (1994): Evaluating Pseudo-R²'s for Binary Probit Models. In: *Quality and Quantity* 28 (2), S. 151-164.

Wang, Margaret C.; Haertel, Geneva D.; Walberg, Herbert J. (1990): What Influences Learning? A Content Analysis of Review Literature. In: *The Journal of Educational Research* 84 (1), S. 30–43.

Welp, Isabell M.; Stumpf-Wollersheim, Jutta; Folger, Nicholas; Prenzel, Manfred (2020): Leistungsbewertung in wissenschaftlichen Institutionen und Universitäten. Berlin, Boston: De Gruyter Oldenbourg.

Wigfield, A.; Eccles, J. S. (2000): Expectancy–value theory of achievement motivation. In: *Contemporary Educational Psychology* 25(1), S. 68-81.

Williams, Richard (2006): Generalized ordered logit/partial proportional odds models for ordinal dependent variables. In: *The Stata Journal* 6(1), S. 58-82.

Williams, Richard (2016). Understanding and interpreting generalized ordered logit models. In: *The Journal of Mathematical Sociology* 40(1), S. 7-20.

Williams, Richard (2020). Using Stata 11 & higher for Logistic Regression. Online verfügbar unter <https://www3.nd.edu/~rwilliam/stats3/Logistic-Stata.pdf>, zuletzt aktualisiert am 12.01.2021, zuletzt geprüft am 09.03.2021

Winteler, Adi; Forster, Peter (2008): Lern-Engagement der Studierenden: Indikator für die Qualität und Effektivität von Lehre und Studium. In: *Das Hochschulwesen* 56, S. 170.

Wuttke, Eveline (2000): Lernstrategien im Lernprozess. In: *Z Erziehungswiss* 3 (1), S. 97–110.

Yilmaz, Ramazan (2017): Exploring the role of e-learning readiness on student satisfaction and motivation in flipped classroom. In: *Computers in Human Behavior* 70, S. 251–260.

Appendix

Abstract

Aufgrund der Corona-Pandemie mussten die Lehrveranstaltungen der Universität Wien im Sommersemester 2020 zur Gänze und im Wintersemester 2020/21 teilweise auf Online-Lehre umgestellt werden. Der Lehrbetrieb wurde dadurch beeinflusst und die Lehr- und Lernumgebung von Studierenden hat sich stark verändert. Die vorliegende Arbeit untersucht, welche Einflussfaktoren sich durch die Corona-Pandemie geändert haben und welche Auswirkungen dieser Änderungen auf akademische Studierendenleistungen festgestellt werden können. Dazu wird einerseits eine theoretische Literaturanalyse durchgeführt, andererseits werden auch Prüfungsergebnisse von Studierenden der Universität Wien analysiert. Durch logistische Regressionen werden sowohl die Noten als auch die Erfolgswahrscheinlichkeit von Studierenden bei Prüfungen, welche von Frau ao. Univ.-Prof. Mag. Dr. Schaffhauser-Linzatti durchgeführt wurden, untersucht.

Binäre Analyse

```
*          ANALYSIS MASTER THESIS KALTENHAUSER          *
*          AUTHOR: SARAH KALTENHAUSER                   *
*          APRIL 2021                                     *
*****

*****

*          Erstanalyse BA & MA gemeinsam                  *
*****

clear all
import excel "/Users/sarahkaltenhauser/Dokumente/Uni/WU/Masterstudium/Masterarbeit/neu/Datenanalyse/Update_data_Prufer_ersterAntritt_Wahl_Pflicht.xlsx", sheet("Sheet1") firstrow
drop if (Studium!="Bachelorstudium" & Studium!="Masterstudium" )

encode Corona_Indikator, generate(corona)
encode Studium, generate(studium)
encode Semester, generate(semester)
encode Wahl_Pflicht, generate(wahlpflicht)

table(Alter)
*drop Alterskategorie
generate Alterskategorie1 =.
replace Alterskategorie1=1 if Alter<25
replace Alterskategorie1=2 if Alter>=25 & Alter<30
replace Alterskategorie1=3 if Alter>=30 & Alter<40
replace Alterskategorie1=4 if Alter>=40 & Alter<50
replace Alterskategorie1=5 if Alter>=50
table(Alterskategorie1)

egen mean1 = mean(100*Erfolg), by(Alterskategorie1)
graph twoway connected mean1 Alterskategorie1, sort(Alterskategorie1)
```

```
* --> quadratischer Zusammenhang erkennbar --> quadratische Altersvariable wird inkludiert
* --> zentriere die Altersvariable indem der mean immer abgezogen wird
summ Alter
generate alter_cen = Alter-26.44729
generate alter2 = alter_cen^2

* Logit Regression basierend auf binv$rer Variable
*****
*      1) Test nach Multikollinearität vor Modellerstellung
regress Erfolg i.corona i.Geschlecht alter_cen alter2 i.wahlpflicht
i.studium, robust cluster(RND)
vif

*      2) Binv$re Regression 1. Runde
logit Erfolg i.corona i.Geschlecht alter_cen alter2 i.wahlpflicht
i.studium, robust cluster(RND)
margins, dydx(*)
predict winfitlin
predict stdres, rstand
generate exclude =.
replace exclude=0 if stdres>-2 & stdres<2
replace exclude=1 if stdres<=-2
replace exclude=1 if stdres>=2

* 3) Binv$re Regression ohne Ausreißer
logit Erfolg i.corona i.Geschlecht alter_cen alter2 i.wahlpflicht
i.studium if exclude==0, robust cluster(RND)
margins, dydx(*)
predict winfitlin2
hist winfitlin2, title("Vorhergesagte Werte des binv$ren Logit-Modells")
predict stdres2, rstand
scatter stdres2 winfitlin2
fitstat
estat classification

*      4) Gleiches Modell ohne quadrierte Altersvariable
*      4a) Binv$re Regression 1. Runde
regress Erfolg i.corona i.Geschlecht Alter i.wahlpflicht i.studium, robust cluster(RND)
vif
logit Erfolg i.corona i.Geschlecht Alter i.wahlpflicht i.studium, robust cluster(RND)
margins, dydx(*)
predict winfitlin_1
predict stdres1, rstand
generate excludel =.
replace excludel=0 if stdres1>-2 & stdres1<2
replace excludel=1 if stdres1<=-2
replace excludel=1 if stdres1>=2

* 4b) Binv$re Regression ohne Ausreißer
logit Erfolg i.corona i.Geschlecht Alter i.wahlpflicht i.studium if excludel==0, robust cluster(RND)
margins, dydx(*)
predict winfitlin3
```

```
hist winfitlin3, title("Vorhergesagte Werte des bin\$\ren Logit-Mo-
dells")
predict stdres3, rstand
scatter stdres3 winfitlin2
fitstat
estat classification

*****
*               Analyse Datensatz bis SS2020               *
*****

*****
*               Bachelorstudierende only                   *
*****

clear all
import excel "/Users/sarahkaltenhauser/Dokumente/Uni/WU/Masterstu-
dium/Masterarbeit/neu/Datenanalyse/data_PruÃäfer_ersterAn-
tritt_Wahl_Pflicht.xlsx", sheet("Sheet1") firstrow
drop if (Studium!="Bachelorstudium")

encode Corona_Indikator, generate(corona)
encode Studiumsart, generate(studienrichtung)
encode Semester, generate(semester)
encode Wahl_Pflicht, generate(wahlpflicht)

table(Alter)
*drop Alterskategorie
generate Alterskategorie1 =.
replace Alterskategorie1=1 if Alter<25
replace Alterskategorie1=2 if Alter>=25 & Alter<30
replace Alterskategorie1=3 if Alter>=30 & Alter<40
replace Alterskategorie1=4 if Alter>=40 & Alter<50
replace Alterskategorie1=5 if Alter>=50
table(Alterskategorie1)

egen mean1 = mean(100*Erfolg), by(Alterskategorie1)
graph twoway line mean1 Alterskategorie1, sort(Alterskategorie1)
title("Durchschnitt Erfolgswahrscheinlichkeit pro Alterskategorie")
xlabel(1 "<25" 2 "25-30" 3 "30-40" 4 "40-50" 5 ">50") ylabel(25 "25%"
50 "50%" 75 "75%") xtitle("Alter in Kategorien") ytitle("Erfolgswahr-
scheinlichkeit")
* --> quadratischer Zusammenhang erkennbar --> quadratische Altersvari-
able wird inkludiert
* --> zentriere die Altersvariable indem der mean immer abgezogen wird
summ Alter
generate alter_cen = Alter-26.60563
generate alter2 = alter_cen^2

* Logit Regression basierend auf bin\$\rer Variable
*****
*      1) Test nach Multikollinearit\st vor Modellerstellung
regress Erfolg i.corona i.Geschlecht alter_cen alter2 i.wahlpflicht
i.studienrichtung, robust cluster(RND)
vif

*      2) Bin\$\re Regression 1. Runde
```

```
logit Erfolg i.corona i.Geschlecht alter_cen alter2 i.wahlpflicht
i.studienrichtung, robust cluster(RND)
margins, dydx(*)
predict winfitlin
predict stdres, rstand
generate exclude =.
replace exclude=0 if stdres>-2 & stdres<2
replace exclude=1 if stdres<=-2
replace exclude=1 if stdres>=2

* 3) Binäre Regression ohne Ausreißer
logit Erfolg i.corona i.Geschlecht alter_cen alter2 i.wahlpflicht
i.studienrichtung if exclude==0, robust cluster(RND)
margins, dydx(*)
predict winfitlin2
hist winfitlin2, title("Vorhergesagte Werte des binären Logit-Modells
(SS20)")
predict stdres2, rstand
scatter stdres2 winfitlin2
fitstat
estat classification

*****
*                               *
*               Masterstudierende only                               *
*****

clear all
import excel "/Users/sarahkaltenhauser/Dokumente/Uni/WU/Masterstudium/Masterarbeit/neu/Datenanalyse/data_Prufer_ersterAntritt_Wahl_Pflicht.xlsx", sheet("Sheet1") firstrow
drop if (Studium!="Masterstudium")

encode Corona_Indikator, generate(corona)
encode Studiumsart, generate(studienrichtung)
encode Semester, generate(semester)
encode Wahl_Pflicht, generate(wahlpflicht)

table(Alterskategorie)
table(Spaltel)
*drop Alterskategorie1
generate Alterskategorie1 =.
replace Alterskategorie1=1 if Alter<25
replace Alterskategorie1=2 if Alter>=25 & Alter<30
replace Alterskategorie1=3 if Alter>=30 & Alter<35
replace Alterskategorie1=4 if Alter>=35
table(Alterskategorie1)

egen mean1 = mean(100*Erfolg), by(Alterskategorie1)
graph twoway line mean1 Alterskategorie1, sort(Alterskategorie1)
title("Durchschnitt Erfolgswahrscheinlichkeit pro Alterskategorie")
xlabel(1 "<25" 2 "25-30" 3 "30-35" 4 ">35") ylabel(25 "25%" 50 "50%" 75 "75%")
xtitle("Alter in Kategorien") ytitle("Erfolgswahrscheinlichkeit")

* --> kein quadratischer Zusammenhang erkennbar

* Logit Regression basierend auf binärer Variable
*****
*      1) Test nach Multikollinearität vor Modellerstellung
```

```
regress Erfolg i.corona i.Geschlecht Alter i.wahlpflicht i.studienrichtung, robust cluster(RND)
vif
```

```
*      2) Binäre Regression 1. Runde
logit Erfolg i.corona i.Geschlecht Alter i.wahlpflicht i.studienrichtung, robust cluster(RND)
margins, dydx(*)
predict winfitlin
predict stdres, rstand
generate exclude =.
replace exclude=0 if stdres>-2 & stdres<2
replace exclude=1 if stdres<=-2
replace exclude=1 if stdres>=2
```

```
* 3) Binäre Regression ohne Ausreißer
logit Erfolg i.corona i.Geschlecht Alter i.wahlpflicht i.studienrichtung if exclude==0, robust cluster(RND)
margins, dydx(*)
predict winfitlin2
hist winfitlin2, title("Vorhergesagte Werte des binären Logit-Modells")
predict stdres2, rstand
scatter stdres2 winfitlin2, title("Residuen des Modells bis SS2020")
xtitle("Erfolgswahrscheinlichkeit") ytitle("Pearson Residuen")
fitstat
estat classification
```

```
*****
*      Analyse Datensatz bis WS2020      *
*****
```

```
*****
*      Bachelorstudierende only      *
*****
```

```
clear all
import excel "/Users/sarahkaltenhauser/Dokumente/Uni/WU/Masterstudium/Masterarbeit/neu/Datenanalyse/Update_data_Prufer_ersterAntritt_Wahl_Pflicht.xlsx", sheet("Sheet1") firstrow
drop if (Studium!="Bachelorstudium")
```

```
encode Corona_Indikator, generate(corona)
encode Studiumsart, generate(studienrichtung)
encode Semester, generate(semester)
encode Wahl_Pflicht, generate(wahlpflicht)
```

```
table(Alter)
*drop Alterskategorie
generate Alterskategorie1 =.
replace Alterskategorie1=1 if Alter<25
replace Alterskategorie1=2 if Alter>=25 & Alter<30
replace Alterskategorie1=3 if Alter>=30 & Alter<40
replace Alterskategorie1=4 if Alter>=40 & Alter<50
replace Alterskategorie1=5 if Alter>=50
table(Alterskategorie1)
```

```
egen mean1 = mean(100*Erfolg), by(Alterskategorie1)
```

```
graph twoway line mean1 Alterskategorie1, sort(Alterskategorie1)
title("Durchschnitt Erfolgswahrscheinlichkeit pro Alterskategorie")
xlabel(1 "<25" 2 "25-30" 3 "30-40" 4 "40-50" 5 ">50") ylabel(25 "25%"
50 "50%" 75 "75%") xtitle("Alter in Kategorien") ytitle("Erfolgswahr-
scheinlichkeit")
* --> quadratischer Zusammenhang erkennbar --> quadratische Altersvari-
able wird inkludiert
* --> zentriere die Altersvariable indem der mean immer abgezogen wird
summ Alter
generate alter_cen = Alter-26.25558
generate alter2 = alter_cen^2

* Logit Regression basierend auf binv$rer Variable
*****
*      1) Test nach Multikollinearität vor Modellerstellung
regress Erfolg i.corona i.Geschlecht alter_cen alter2 i.wahlpflicht
i.studienrichtung, robust cluster(RND)
vif

*      2) Binv$re Regression 1. Runde
logit Erfolg i.corona i.Geschlecht alter_cen alter2 i.wahlpflicht
i.studienrichtung, robust cluster(RND)
margins, dydx(*)
predict winfitlin
predict stdres, rstand
generate exclude =.
replace exclude=0 if stdres>-2 & stdres<2
replace exclude=1 if stdres<=-2
replace exclude=1 if stdres>=2

* 3) Binv$re Regression ohne Ausreißer
logit Erfolg i.corona i.Geschlecht alter_cen alter2 i.wahlpflicht
i.studienrichtung if exclude==0, robust cluster(RND)
margins, dydx(*)
predict winfitlin2
hist winfitlin2, title("Vorhergesagte Werte des binv$ren Logit-Modells
(WS20/21)")
predict stdres2, rstand
scatter stdres2 winfitlin2
fitstat
estat classification

*****
*                               Masterstudierende only                               *
*****
clear all
import excel "/Users/sarahkaltenhauser/Dokumente/Uni/WU/Masterstu-
dium/Masterarbeit/neu/Datenanalyse/Update_data_Prufer_ersterAn-
tritt_Wahl_Pflicht.xlsx", sheet("Sheet1") firstrow
drop if (Studium!="Masterstudium")

encode Corona_Indikator, generate(corona)
encode Studiumsart, generate(studienrichtung)
encode Semester, generate(semester)
encode Wahl_Pflicht, generate(wahlpflicht)

table(Alterskategorie)
table(Spaltel)
```

```
*drop Alterskategorie1
generate Alterskategorie1 =.
replace Alterskategorie1=1 if Alter<25
replace Alterskategorie1=2 if Alter>=25 & Alter<30
replace Alterskategorie1=3 if Alter>=30 & Alter<35
replace Alterskategorie1=4 if Alter>=35
table(Alterskategorie1)

egen mean1 = mean(100*Erfolg), by(Alterskategorie1)
graph twoway line mean1 Alterskategorie1, sort(Alterskategorie1)
title("Durchschnitt Erfolgswahrscheinlichkeit pro Alterskategorie")
xlabel(1 "<25" 2 "25-30" 3 "30-35" 4 ">35") ylabel(25 "25%" 50 "50%" 75 "75%")
xtitle("Alter in Kategorien") ytitle("Erfolgswahrscheinlichkeit")

* --> kein quadratischer Zusammenhang erkennbar

* Logit Regression basierend auf binomialer Variable
*****
*      1) Test nach Multikollinearität vor Modellerstellung
regress Erfolg i.corona i.Geschlecht Alter i.wahlpflicht i.studienrichtung, robust cluster(RND)
vif

*      2) Binomial Regression 1. Runde
logit Erfolg i.corona i.Geschlecht Alter i.wahlpflicht i.studienrichtung, robust cluster(RND)
margins, dydx(*)
predict winfitlin
predict stdres, rstand
generate exclude =.
replace exclude=0 if stdres>-2 & stdres<2
replace exclude=1 if stdres<=-2
replace exclude=1 if stdres>=2

* 3) Binomial Regression ohne Ausreißer
logit Erfolg i.corona i.Geschlecht Alter i.wahlpflicht i.studienrichtung if exclude==0, robust cluster(RND)
margins, dydx(*)
predict winfitlin2
hist winfitlin2, title("Vorhergesagte Werte des binomialen Logit-Modells")
predict stdres2, rstand
scatter stdres2 winfitlin2, title("Residuen des Modells bis WS2020/21")
xtitle("Erfolgswahrscheinlichkeit") ytitle("Pearson Residuen")
fitstat
estat classification
```

Ordinale Analyse

```
*      ANALYSIS MASTER THESIS KALTENHAUSER      *
*      AUTHOR: SARAH KALTENHAUSER                *
*      APRIL 2021                                *
*****
*****
*      Erstanalyse BA & MA gemeinsam              *
```

```
clear all
import excel "/Users/sarahkaltenhauser/Dokumente/Uni/WU/Masterstudium/Masterarbeit/neu/Datenanalyse/Update_data_Prufer_ersterAntritt_Wahl_Pflicht.xlsx", sheet("Sheet1") firstrow
drop if (Studium!="Bachelorstudium" & Studium!="Masterstudium" )
drop if Spalte1 == .

encode Corona_Indikator, generate(corona)
encode Studium, generate(studium)
encode Semester, generate(semester)
encode Wahl_Pflicht, generate(wahlpflicht)

table(Alter)
*drop Alterskategorie
generate Alterskategorie1 =.
replace Alterskategorie1=1 if Alter<25
replace Alterskategorie1=2 if Alter>=25 & Alter<30
replace Alterskategorie1=3 if Alter>=30 & Alter<40
replace Alterskategorie1=4 if Alter>=40 & Alter<50
replace Alterskategorie1=5 if Alter>=50
table(Alterskategorie1)

egen mean1 = mean(Spalte1), by(Alterskategorie1)
graph twoway connected mean1 Alterskategorie1, sort(Alterskategorie1)
graph twoway line mean1 Alterskategorie1, sort(Alterskategorie1)
title("Durchschnittliche Note pro Alterskategorie") xlabel(1 "<25" 2 "25-30" 3 "30-40" 4 "40-50" 5 ">50") ylabel(3 "3" 4 "4" 5 "5")
xtitle("Alter in Kategorien") ytitle("Durchschnittliche Note")
* --> quadratischer Zusammenhang erkennbar --> quadratische Altersvariable wird inkludiert
* --> zentriere die Altersvariable indem der mean immer abgezogen wird
summ Alter
generate alter_cen = Alter-26.75647
generate alter2 = alter_cen^2

* Logit Regression basierend auf ordinaler Variable
*****
*      1) Test nach Multikollinearität vor Modellerstellung
regress Spalte1 i.corona i.Geschlecht alter_cen alter2 i.wahlpflicht i.studium, robust cluster(RND)
vif

*      2) Ordinale Regression 1. Runde --> Voraussetzung: proportionale Odds Ratios
ologit Spalte1 i.corona i.Geschlecht alter_cen alter2 i.studium i.wahlpflicht, robust cluster(RND) or
oparallel
*      test shows that assumption is violated*
*      check if violated due to 1 variable or more*
ologit Spalte1 i.corona i.Geschlecht alter_cen alter2 i.studium i.wahlpflicht, robust cluster(RND) or
brant
*      violated by nearly all variables*
*      --> use alternative option*

* 3) Ordinale generalisierte Regression
```

```
gologit2 Spalte1 i.corona i.Geschlecht alter_cen alter2 i.studium
i.wahlpflicht, robust cluster(RND) or
estat ic
fitstat

* große Differenzen bei Master- und Bachelorstudierenden --> auch hier
aufteilen

*****
*           Analyse Datensatz bis SS2020           *
*****

*****
*           Bachelorstudierende only               *
*****

clear all
import excel "/Users/sarahkaltenhauser/Dokumente/Uni/WU/Masterstu-
dium/Masterarbeit/neu/Datenanalyse/data_Prufer_ersterAn-
tritt_Wahl_Pflicht.xlsx", sheet("Sheet1") firstrow
drop if (Studium!="Bachelorstudium")
drop if Spalte1 == .

encode Corona_Indikator, generate(corona)
encode Studiumsart, generate(studienrichtung)
encode Semester, generate(semester)
encode Wahl_Pflicht, generate(wahlpflicht)

table(Alter)
table(Studium)
unique(RND)
*drop Alterskategorie
generate Alterskategorie1 =.
replace Alterskategorie1=1 if Alter<25
replace Alterskategorie1=2 if Alter>=25 & Alter<30
replace Alterskategorie1=3 if Alter>=30 & Alter<40
replace Alterskategorie1=4 if Alter>=40 & Alter<50
replace Alterskategorie1=5 if Alter>=50
table(Alterskategorie1)

egen mean1 = mean(Spalte1), by(Alterskategorie1)
graph twoway line mean1 Alterskategorie1, sort(Alterskategorie1)
title("Durchschnittliche Note pro Alterskategorie") xlabel(1 "<25" 2
"25-30" 3 "30-40" 4 "40-50" 5 ">50") ylabel(2 "2" 3 "3" 4 "4" 5 "5")
xtitle("Alter in Kategorien") ytitle("Durchschnittliche Note")
* --> quadratischer Zusammenhang erkennbar --> quadratische Altersvari-
able wird inkludiert
* --> zentriere die Altersvariable indem der mean immer abgezogen wird
summ Alter
generate alter_cen = Alter-26.60213
generate alter2 = alter_cen^2

* Logit Regression basierend auf ordinaler Variable
*****
*           1) Test nach Multikollinearität vor Modellerstellung
regress Spalte1 i.corona i.Geschlecht alter_cen alter2 i.wahlpflicht
i.studienrichtung, robust cluster(RND)
vif
```

```
*      2) Ordinale Regression 1. Runde --> Voraussetzung: proportionale
Odds Ratios
ologit Spalte1 i.corona i.Geschlecht alter_cen alter2 i.studienrichtung
i.wahlpflicht, robust cluster(RND) or
oparallel
*      test shows that assumption is violated*
*      check if violated due to 1 variable or more*
ologit Spalte1 i.corona i.Geschlecht alter_cen alter2 i.studienrichtung
i.wahlpflicht, robust cluster(RND) or
brant
*      violated by nearly all variables*
*      --> use alternative option*

* 3) Ordinale generalisierte Regression
gologit2 Spalte1 i.corona i.Geschlecht alter_cen alter2 i.studienrichtung
i.wahlpflicht, robust cluster(RND) or autofit
estat ic
fitstat

*****
*      Masterstudierende only      *
*****
clear all
import excel "/Users/sarahkaltenhauser/Dokumente/Uni/WU/Masterstudium/Masterarbeit/neu/Datenanalyse/data_Prufer_ersterAntritt_Wahl_Pflicht.xlsx", sheet("Sheet1") firstrow
drop if (Studium!="Masterstudium")
drop if Spalte1 == .

encode Corona_Indikator, generate(corona)
encode Studiumsart, generate(studienrichtung)
encode Semester, generate(semester)
encode Wahl_Pflicht, generate(wahlpflicht)
unique(RND)

table(Alterskategorie)
table(Spalte1)
*drop Alterskategorie1
generate Alterskategorie1 =.
replace Alterskategorie1=1 if Alter<25
replace Alterskategorie1=2 if Alter>=25 & Alter<30
replace Alterskategorie1=3 if Alter>=30 & Alter<35
replace Alterskategorie1=4 if Alter>=35
table(Alterskategorie1)

egen mean1 = mean(Spalte1), by(Alterskategorie1)
graph twoway line mean1 Alterskategorie1, sort(Alterskategorie1)
title("Durchschnittliche Note pro Alterskategorie") xlabel(1 "<25" 2
"25-30" 3 "30-35" 4 ">35") ylabel(1 "1" 2 "2" 3 "3" 4 "4" 5 "5")
xtitle("Alter in Kategorien") ytitle("Durchschnittliche Note")

* --> kein quadratischer Zusammenhang erkennbar

* Logit Regression basierend auf ordinaler Variable
*****
*      1) Test nach Multikollinearität vor Modellerstellung
regress Spalte1 i.corona i.Geschlecht Alter i.wahlpflicht i.studienrichtung, robust cluster(RND)
```

vif

```
*      2) Ordinale Regression 1. Runde --> Voraussetzung: proportionale
Odds Ratios
ologit Spaltel i.corona i.Geschlecht Alter i.studienrichtung i.wahl-
pflicht, robust cluster(RND) or
*      perfect prediction --> Datensatz wohl zu
klein/nichtssagend
```

```
*****
```

```
*      Analyse Datensatz bis WS2020      *
*****
```

```
*****
```

```
*      Bachelorstudierende only      *
*****
```

```
clear all
import excel "/Users/sarahkaltenhauser/Dokumente/Uni/WU/Masterstu-
dium/Masterarbeit/neu/Datenanalyse/Update_data_PruÃfer_ersterAn-
tritt_Wahl_Pflicht.xlsx", sheet("Sheet1") firstrow
drop if Studium != "Bachelorstudium"
drop if Spaltel == .
```

```
encode Corona_Indikator, generate(corona)
encode Studiumsart, generate(studienrichtung)
encode Semester, generate(semester)
encode Wahl_Pflicht, generate(wahlpflicht)
```

```
table(Alter)
table(Studium)
unique(RND)
*drop Alterskategorie
generate Alterskategorie1 =.
replace Alterskategorie1=1 if Alter<25
replace Alterskategorie1=2 if Alter>=25 & Alter<30
replace Alterskategorie1=3 if Alter>=30 & Alter<40
replace Alterskategorie1=4 if Alter>=40 & Alter<50
replace Alterskategorie1=5 if Alter>=50
table(Alterskategorie1)
```

```
egen mean1 = mean(Spaltel), by(Alterskategorie1)
graph twoway line mean1 Alterskategorie1, sort(Alterskategorie1)
title("Durchschnittliche Note pro Alterskategorie") xlabel(1 "<25" 2
"25-30" 3 "30-40" 4 "40-50" 5 ">50") ylabel(2 "2" 3 "3" 4 "4" 5 "5")
xtitle("Alter in Kategorien") ytitle("Durchschnittliche Note")
* --> quadratischer Zusammenhang erkennbar --> quadratische Altersvari-
able wird inkludiert
* --> zentriere die Altersvariable indem der mean immer abgezogen wird
summ Alter
generate alter_cen = Alter-26.22605
generate alter2 = alter_cen^2
```

```
* Logit Regression basierend auf ordinaler Variable
*****
```

```
*      1) Test nach Multikollinearität vor Modellerstellung
regress Spaltel i.corona i.Geschlecht alter_cen alter2 i.wahlpflicht
i.studienrichtung, robust cluster(RND)
```

vif

```
*      2) Ordinale Regression 1. Runde --> Voraussetzung: proportionale  
Odds Ratios  
ologit Spaltel i.corona i.Geschlecht alter_cen alter2 i.studienrichtung  
i.wahlpflicht, robust cluster(RND) or  
oparallel
```

```
          *      test shows that assumption is violated*  
          *      check if violated due to 1 variable or more*  
ologit Spaltel i.corona i.Geschlecht alter_cen alter2 i.studienrichtung  
i.wahlpflicht, robust cluster(RND) or  
brant
```

```
          *      violated by nearly all variables*  
          *      --> use alternative option*
```

```
* 3) Ordinale generalisierte Regression  
gologit2 Spaltel i.corona i.Geschlecht alter_cen alter2 i.studienrich-  
tung i.wahlpflicht, robust cluster(RND) or autofit  
estat ic  
fitstat
```

```
*****
```

```
*      Masterstudierende only      *
```

```
*****
```

```
clear all  
import excel "/Users/sarahkaltenhauser/Dokumente/Uni/WU/Masterstu-  
dium/Masterarbeit/neu/Datenanalyse/Update_data_Prufer_ersterAn-  
tritt_Wahl_Pflicht.xlsx", sheet("Sheet1") firstrow  
drop if (Studium!="Masterstudium")  
drop if Spaltel == .
```

```
encode Corona_Indikator, generate(corona)  
encode Studiumsart, generate(studienrichtung)  
encode Semester, generate(semester)  
encode Wahl_Pflicht, generate(wahlpflicht)  
unique(RND)
```

```
table(Spaltel)  
*drop Alterskategorie1  
generate Alterskategorie1 =.  
replace Alterskategorie1=1 if Alter<25  
replace Alterskategorie1=2 if Alter>=25 & Alter<30  
replace Alterskategorie1=3 if Alter>=30 & Alter<35  
replace Alterskategorie1=4 if Alter>=35  
table(Alterskategorie1)
```

```
egen mean1 = mean(Spaltel), by(Alterskategorie1)  
graph twoway line mean1 Alterskategorie1, sort(Alterskategorie1)  
title("Durchschnittliche Note pro Alterskategorie") xlabel(1 "<25" 2  
"25-30" 3 "30-35" 4 ">35") ylabel(2 "2" 3 "3" 4 "4" 5 "5") xtitle("Al-  
ter in Kategorien") ytitle("Durchschnittliche Note")
```

```
* --> kein quadratischer Zusammenhang erkennbar
```

```
* Logit Regression basierend auf ordinaler Variable
```

```
summ Alter  
generate alter_cen = Alter-28.44226
```

```
generate alter2 = alter_cen^2

* Logit Regression basierend auf ordinaler Variable
*****
*      1) Test nach Multikollinearität vor Modellerstellung
regress Spalte1 i.corona i.Geschlecht alter_cen alter2 i.wahlpflicht
i.studienrichtung, robust cluster(RND)
vif

*      2) Ordinale Regression 1. Runde --> Voraussetzung: proportionale
Odds Ratios
ologit Spalte1 i.corona i.Geschlecht alter_cen alter2 i.studienrichtung
i.wahlpflicht, robust cluster(RND) or
oparallel
                *      test shows that assumption is violated*
                *      check if violated due to 1 variable or more*
ologit Spalte1 i.corona i.Geschlecht alter_cen alter2 i.studienrichtung
i.wahlpflicht, robust cluster(RND) or
brant
                *      violated by nearly all variables*
                *      --> use alternative option*

* 3) Ordinale generalisierte Regression
gologit2 Spalte1 i.corona i.Geschlecht alter_cen alter2 i.studienrichtung
i.wahlpflicht, robust cluster(RND) or autofit
estat ic
fitstat
```

Multinomiale Analyse

```
*      ANALYSIS MASTER THESIS KALTENHAUSER      *
*      AUTHOR: SARAH KALTENHAUSER                *
*      MAY 2021                                  *
*****

*****
*      Analyse Datensatz bis SS2020                *
*****

*****
*      Bachelorstudierende only                    *
*****

clear all
import excel "/Users/sarahkaltenhauser/Dokumente/Uni/WU/Masterstudium/Masterarbeit/neu/Datenanalyse/data_Prufer_ersterAntritt_Wahl_Pflicht.xlsx", sheet("Sheet1") firstrow
drop if (Studium!="Bachelorstudium")
drop if Spalte1 == .

encode Corona_Indikator, generate(corona)
encode Studiumsart, generate(studienrichtung)
encode Semester, generate(semester)
encode Wahl_Pflicht, generate(wahlpflicht)

table(Alter)
table(Studium)
unique(RND)
*drop Alterskategorie
```

```
generate Alterskategorie1 =.
replace Alterskategorie1=1 if Alter<25
replace Alterskategorie1=2 if Alter>=25 & Alter<30
replace Alterskategorie1=3 if Alter>=30 & Alter<40
replace Alterskategorie1=4 if Alter>=40 & Alter<50
replace Alterskategorie1=5 if Alter>=50
table(Alterskategorie1)

egen mean1 = mean(Spaltel), by(Alterskategorie1)
graph twoway line mean1 Alterskategorie1, sort(Alterskategorie1)
title("Durchschnittliche Note pro Alterskategorie") xlabel(1 "<25" 2
"25-30" 3 "30-40" 4 "40-50" 5 ">50") ylabel(2 "2" 3 "3" 4 "4" 5 "5")
xtitle("Alter in Kategorien") ytitle("Durchschnittliche Note")
* --> quadratischer Zusammenhang erkennbar --> quadratische Altersvari-
able wird inkludiert
* --> zentriere die Altersvariable indem der mean immer abgezogen wird
summ Alter
generate alter_cen = Alter-26.60213
generate alter2 = alter_cen^2

* Logit Regression basierend auf ordinaler Variable
*****
*      1) Test nach Multikollinearität vor Modellerstellung
regress Spaltel i.corona i.Geschlecht alter_cen alter2 i.wahlpflicht
i.studienrichtung, robust cluster(RND)
vif

*      2) Ordinale Regression 1. Runde --> Voraussetzung: proportionale
Odds Ratios
mlogit Spaltel i.corona i.Geschlecht alter_cen alter2 i.studienrichtung
i.wahlpflicht, robust cluster(RND) rr

*****
*      Analyse Datensatz bis WS2020      *
*****

*****
*      Bachelorstudierende only      *
*****

clear all
import excel "/Users/sarahkaltenhauser/Dokumente/Uni/WU/Masterstu-
dium/Masterarbeit/neu/Datenanalyse/Update_data_Prufer_ersterAn-
tritt_Wahl_Pflicht.xlsx", sheet("Sheet1") firstrow
drop if Studium != "Bachelorstudium"
drop if Spaltel == .

encode Corona_Indikator, generate(corona)
encode Studiumsart, generate(studienrichtung)
encode Semester, generate(semester)
encode Wahl_Pflicht, generate(wahlpflicht)

table(Alter)
table(Studium)
unique(RND)
*drop Alterskategorie
generate Alterskategorie1 =.
replace Alterskategorie1=1 if Alter<25
replace Alterskategorie1=2 if Alter>=25 & Alter<30
```

```
replace Alterskategorie1=3 if Alter>=30 & Alter<40
replace Alterskategorie1=4 if Alter>=40 & Alter<50
replace Alterskategorie1=5 if Alter>=50
table(Alterskategorie1)

egen mean1 = mean(Spaltel), by(Alterskategorie1)
graph twoway line mean1 Alterskategorie1, sort(Alterskategorie1)
title("Durchschnittliche Note pro Alterskategorie") xlabel(1 "<25" 2
"25-30" 3 "30-40" 4 "40-50" 5 ">50") ylabel(2 "2" 3 "3" 4 "4" 5 "5")
xtitle("Alter in Kategorien") ytitle("Durchschnittliche Note")
* --> quadratischer Zusammenhang erkennbar --> quadratische Altersvari-
able wird inkludiert
* --> zentriere die Altersvariable indem der mean immer abgezogen wird
summ Alter
generate alter_cen = Alter-26.22605
generate alter2 = alter_cen^2

* Logit Regression basierend auf ordinaler Variable
*****
*      1) Test nach Multikollinearität vor Modellerstellung
regress Spaltel i.corona i.Geschlecht alter_cen alter2 i.wahlpflicht
i.studienrichtung, robust cluster(RND)
vif

*      2) Ordinale Regression 1. Runde --> Voraussetzung: proportionale
Odds Ratios
mlogit Spaltel i.corona i.Geschlecht alter_cen alter2 i.studienrichtung
i.wahlpflicht, robust cluster(RND) rr
```