

# **Liechtensteinischer Lohnrechner**

Dokumentation

Version 1.0

Stand: 11. Juli 2019

**Herausgeber und Vertrieb**

Amt für Statistik  
Äulestrasse 51  
9490 Vaduz  
Liechtenstein  
T +423 236 68 76  
F +423 236 69 36

**Auskunft**

Simon Gstöhl T +423 236 68 77  
info.as@llv.li

**Thema  
Copyright**

04 Arbeit und Erwerb  
Wiedergabe unter Angabe des Herausgebers gestattet.  
© Amt für Statistik

## Inhaltsverzeichnis

1.	Einleitung.....	1
2.	Datengrundlage.....	2
2.1	Datenquellen .....	2
2.2	Datenaufbereitung .....	2
2.2.1	Erstellung des Basisdatensatzes.....	2
2.2.2	Verwendete Filter.....	3
3.	Berechnungsmodell.....	5
3.1	Regressionsmodell.....	5
3.2	Variablen.....	6
3.2.1	Abhängige Variable .....	6
3.2.2	Unabhängige Variablen .....	7
3.2.3	Nicht verwendete Variablen .....	11
4.	Resultate.....	12
4.1	Regressionsmodell.....	12
4.1.1	Modellkoeffizienten .....	12
4.1.2	Modellgüte .....	14
4.2	Schätzung der Quartile .....	16
4.3	Multikollinearität .....	17
4.4	Modellüberprüfung .....	17
	Literaturverzeichnis.....	19

## 1. Einleitung

Das Amt für Statistik führt seit 2006 alle zwei Jahre eine Erhebung der Lohndaten durch und publiziert die Ergebnisse in der Lohnstatistik. Seit längerem wurde dabei von den Nutzerinnen und Nutzern der Wunsch nach einer weiteren Aufschlüsselung und der Erstellung von Lohnprognosen geäußert. Da die Erhebung jedoch auf Verwaltungsdaten der Steuerverwaltung basiert, fehlen wichtige Informationen zu verschiedenen Faktoren, welche zwar nicht in den amtlichen Registern erfasst werden, sich aber auf das Lohnniveau auswirken (z.B. berufliche Stellung, Bildungsniveau, etc.).

Die Verknüpfung der Daten aus der Lohnstatistik mit denjenigen aus der Volkszählung erlaubt es nun dem Amt für Statistik den Nutzerinnen und Nutzern erstmals einen Lohnrechner anzubieten. Der Lohnrechner ermöglicht es aufgrund von Eigenschaften der Person und des Unternehmens in dem sie beschäftigt ist, einen mittleren Lohn zu schätzen. Die Berechnungen des Lohnrechners basieren dabei auf einem statistischen Modell, welches anhand von verschiedenen erklärenden Variablen einen Schätzwert für den Medianlohn sowie für die Quartile liefert. Beim Medianlohn handelt es sich um den mittleren Lohn. D.h. das Modell schätzt, dass die eine Hälfte der Beschäftigten mehr verdient als der angegebene Medianlohn, während die andere Hälfte weniger verdient. Auch die Quartile teilen die Beschäftigten in je zwei Gruppen. Ein Viertel der Beschäftigten verdient gemäss dem Modell weniger als das 1. Quartil, drei Viertel mehr. Beim 3. Quartil ist es genau umgekehrt.

Dem Wunsch der Nutzerinnen und Nutzer nach Lohnprognosen kann somit ein Stück weit entgegengekommen werden. Wer erwartet, dass der Lohnrechner eine genaue Lohnangabe für seine Situation liefert, wird jedoch enttäuscht werden. Bei den angegebenen Löhnen handelt es sich um Schätzungen, welche auf einem statistischen Modell basieren, und stellen keinesfalls eine Lohnempfehlung dar. Bei der Schätzung handelt es sich um einen Gruppenmittelwert, und obwohl die Gruppenmitglieder gewisse Gemeinsamkeiten aufweisen, können sie sich in anderen, nicht vom Modell erfassten Eigenschaften unterscheiden. Dabei spielen Faktoren wie Weiterbildungen und individuelle Karriereverläufe genauso eine Rolle wie die Lohnpolitik des Unternehmens oder Verhandlungsgeschick in Lohnverhandlungen. Das Modell trägt diesem Umstand Rechnung indem es neben einem mittleren Lohn auch die Quartile wiedergibt und somit einen Bereich angibt, in dem die Hälfte der Löhne zu finden sein dürfte.

Die verwendete Methodik basiert weitgehend auf der Methodik, die durch das Bundesamt für Statistik (BFS 2008) für die Erstellung des Lohnrechners „Salarium“ verwendet wurde. Dazu wird ein log-lineares Regressionsmodell geschätzt, wobei aufgrund der Residuen eine Anpassung des Mittelwertschätzers gemacht wird, um die Quartilsschätzer zu erhalten.

Die vorliegende Dokumentation richtet sich an die Nutzerinnen und Nutzer des liechtensteinischen Lohnrechners, welche einen vertieften Einblick in den verwendeten Datensatz sowie das verwendete Berechnungsmodell erlangen möchten und strukturiert sich wie folgt: In Kapitel 2 wird zunächst auf die Erstellung des dem Lohnrechner zugrundeliegenden Datensatzes eingegangen. In Kapitel 3 wird das statistische Verfahren theoretisch diskutiert sowie die im Modell verwendeten Variablen präsentiert, bevor in Kapitel 4 auf die Resultate des Modells eingegangen wird.

## **2. Datengrundlage**

Als Grundlage für das Modell wurde zunächst ein geeigneter Datensatz erstellt, der die nötigen Variablen enthält, um Lohnschätzungen zu erstellen. Dieser Abschnitt geht zunächst auf die verwendeten Datenquellen ein, bevor im folgenden Abschnitt die Zusammenführung und Aufbereitung genauer diskutiert werden.

### **2.1 Datenquellen**

Der dem Lohnrechner zugrundeliegende Datensatz basiert auf der Verknüpfung von Daten der Lohnstatistik mit den Daten der Volkszählung.

Die Lohnstatistik wird aus Verwaltungsdaten der liechtensteinischen Steuerverwaltung erstellt. Die Arbeitgeber melden der Steuerverwaltung jeweils Anfang Jahr im Rahmen der Erstellung der Lohnausweise u.a. den an die Arbeitnehmer ausbezahlten Bruttolohn sowie den Zeitraum der Beschäftigung im Unternehmen. Diese Angaben werden vom Amt für Informatik anhand der Arbeitnehmer-Arbeitgeber-Kombination mit Angaben aus dem liechtensteinischen Unternehmensregister per Stichtag 31. Dezember angereichert. Lohnzahlungen derselben Arbeitnehmer-Arbeitgeber-Kombination werden dabei zu einer einzigen Lohnzahlung und Beschäftigungsdauer konsolidiert. Teilzeitpensen werden auf ein einheitliches Vollzeitpensum hochgerechnet. Dieser Datensatz steht dem Amt für Statistik in anonymisierter Form zur Verfügung.

Bei der Lohnstatistik handelt es sich nicht um eine Vollerhebung, sondern um eine Teilerhebung. Der Lohnstatistik 2016 liegen Bruttolohnangaben für 34'558 Arbeitsplätze zugrunde. Die Gesamtzahl der Arbeitsplätze per 31. Dezember 2016 beläuft sich gemäss der Beschäftigungsstatistik auf 39'290 Voll- und Teilzeitarbeitsplätze (Beschäftigungsgrad von 2% und mehr). Hieraus ergibt sich für die Lohnstatistik ein Erfassungsgrad von 88% der Bruttolöhne.

Die Volkszählung wurde vom Amt für Statistik mit Stichtag 31. Dezember 2015 durchgeführt und enthält umfangreiche statistische Informationen zur liechtensteinischen Bevölkerung. Es handelt sich bei der Volkszählung um eine Vollerhebung. Sämtliche Personen, die am Stichtag älter als ein Jahr waren, wurden mittels eines schriftlichen Fragebogens befragt. Die Volkszählung erzielte eine sehr hohe Rücklaufquote von 97%. Dennoch ist es nicht ausgeschlossen, dass bei einer gewissen Anzahl Beobachtungen Angaben zu einzelnen Merkmalen fehlen. Im Rahmen der Datenaufbereitung der Volkszählung wurde eine Reihe von Plausibilitätsprüfungen durchgeführt um un plausible Daten zu entfernen.

### **2.2 Datenaufbereitung**

Nach dieser kurzen Beschreibung der Datenquellen wird in diesem Abschnitt auf die Beschreibung der Datenaufbereitung eingegangen. Zunächst wird dafür die Zusammenführung der Datensätze thematisiert, bevor auf die verwendeten Filter eingegangen wird.

#### **2.2.1 Erstellung des Basisdatensatzes**

Für die Erstellung der Lohnstatistik 2016 führte das Amt für Informatik die Lohnangaben mit den Arbeitgebern und Arbeitnehmermerkmalen aus dem liechtensteinischen Unterneh-

mensregister zusammen. Im Anschluss wurden diese Daten ebenfalls durch das Amt für Informatik mit den Angaben aus der Volkszählung 2015 ergänzt. Beobachtungen, welche keine korrespondierende Beobachtung im Datensatz der Volkszählung aufwiesen, wurden aus dem Datensatz entfernt. Dabei handelte es sich v.a. um Personen, welche im Ausland wohnten oder erst im Laufe des Jahres 2016 zugezogen waren. Während die Grundgesamtheit der Lohnstatistik aus den Beschäftigten in Liechtenstein besteht, sind es in der Volkszählung die Einwohnerinnen und Einwohner Liechtensteins. Im zusammengeführten Datensatz sind daher nur die im Inland erwerbstätigen Einwohnerinnen und Einwohner Liechtensteins enthalten. Das Amt für Statistik erhielt als Ergebnis den anonymisierten Datensatz.

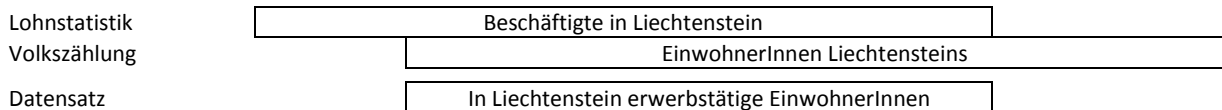


Abbildung 1 Zusammenführung der Datensätze

Die Verknüpfung der beiden Datensätze erfolgte über die AHV-Nummer. Die Angaben aus der Volkszählung wurden per Stichtag 31. Dezember 2015 ausgewiesen. Die Daten zu Bruttolohn und Beschäftigungsdauer aus der Lohnstatistik beziehen sich auf das Kalenderjahr 2016, während die weiteren Angaben zum Arbeitsverhältnis vom Stichtag 31. Dezember 2016 stammen.

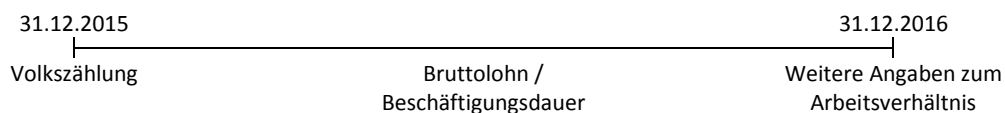


Abbildung 2 Referenzzeitpunkte

Für den Lohnrechner wurden sämtliche arbeitnehmerbezogenen Angaben aus der Volkszählung entnommen. Dadurch können diese Angaben als Ausstattung am 31.12.2015 interpretiert werden. Die Lohnangaben, welche über das Jahr 2016 kumuliert wurden, als Lohn, welcher mit dieser Ausstattung erzielt werden kann. Dabei muss erwähnt werden, dass es innerhalb des Jahres zu Stellenwechseln, Veränderungen des Arbeitspensums, Ausbildungsabschlüssen etc. kommen, welche Auswirkungen auf das Lohnniveau haben, sich jedoch nicht in den erklärenden Variablen widerspiegeln. Im Rahmen der Robustheitsanalyse wurde diese Frage unter anderem aufgenommen, indem die Analyse auf diejenigen Beobachtungen eingeschränkt wurde, die das ganze Jahr beim selben Arbeitgeber beschäftigt waren. So wurde versucht, so gut wie möglich auf die Effekte der unterschiedlichen Stichtage zu kontrollieren. Dabei zeigten sich keine wesentlichen Unterschiede zum Basismodell.

Für die vorliegende Analyse wurden keine fehlenden oder offensichtlich fehlerhaften Löhne imputiert.

### 2.2.2 Verwendete Filter

Für die Erstellung des Datensatzes, welcher dem Modell zugrunde liegt, wurde eine Reihe von Filtern angewandt. Die Filter kamen dabei in drei Schritten zur Anwendung. Tabelle 1 gibt einen Überblick über die betroffenen Fallzahlen.

In einem ersten Schritt wurden die Fälle auf die Hauptbeschäftigung reduziert. Grundlage der Lohnstatistik bilden die Arbeitsverhältnisse. Personen mit mehr als einem Arbeitsver-

hältnis sind daher in den Daten mehr als einmal enthalten. In der Volkszählung wird beim ausgeübten Beruf jedoch nur die Haupttätigkeit erfasst. Aus diesem Grund müssen Nebentätigkeiten aus der Analyse ausgeschlossen werden, da ansonsten der Beruf für die Nebentätigkeit falsch zugeordnet werden könnte, wenn sich dieser von der Haupttätigkeit unterscheidet. Für die Bestimmung der Haupttätigkeit wurde das Arbeitsverhältnis mit einem höheren Beschäftigungsgrad gewählt. Waren die grössten Beschäftigungsverhältnisse gleich gross, wurde das Arbeitsverhältnis mit dem höheren Bruttomonatslohn gewählt. Nach Anwendung dieses Filters sind im Datensatz noch 15'473 Fälle enthalten.

In einem zweiten Schritt wurden die in der Lohnstatistik verwendeten Filter für diesen Datensatz übernommen. Es handelt sich dabei u.a. um den Ausschluss ungewöhnlicher Löhne, tiefer Beschäftigungsverhältnisse, geringer Arbeitspensen sowie von Personen unter 19 Jahren. Eine Beschreibung dieser Filter sowie deren Auswirkungen auf den Medianlohn sind im Abschnitt „1.5 Datenaufbereitung“ der Lohnstatistik 2016 enthalten. Nach Anwendung dieser Filter verbleiben insgesamt 13'211 Beobachtungen im Datensatz.

In einem dritten Schritt werden verschiedene inhaltliche Filter angewandt. Zunächst wurden die Beschäftigten im Sektor „Land- und Forstwirtschaft“ sowie sämtliche Personen, welche der Berufsgruppe „Land- und Forstwirtschaftskräfte“ angehörten, entfernt. Grund für diesen Filter war die Schwierigkeit, die Löhne im Sektor Land- und Forstwirtschaft durch das gewählte Modell effektiv zu beschreiben. Des Weiteren wurden diejenigen Beobachtungen aus dem Datensatz entfernt, welche der beruflichen Stellungen „Selbständig“, „In eigener Kapitalgesellschaft“, „Mitarbeitendes Familienmitglied“ oder „Lehrling“ angehörten. Dieser Filter wurde gesetzt, da diese Angaben nur eine begrenzte Aussagekraft haben. Personen, die in einem eigenen Unternehmen arbeiten, zahlen sich selbst zwar einen Lohn aus, dieser wurde dabei jedoch nicht auf dem Markt erzielt. Ebenso werden Lehrlinge aufgrund der geringen Fallzahl sowie der speziellen Lohnstruktur aus dem Datensatz entfernt. Schliesslich werden auch Personen über 65 Jahren aus dem Datensatz ausgeschlossen. Nach Anwendung dieser Filter verbleiben insgesamt 10'746 Beobachtungen im Datensatz.

Schliesslich fallen 768 Beobachtungen aus dem Datensatz, da bei diesen Daten für mindestens ein verwendetes Merkmal keine Angaben vorliegen.

Der für den Lohnrechner verwendete Datensatz besteht somit aus 9'978 Beobachtungen.

Filtergruppe	Filter	Anzahl Fälle	Verbleibende
<b>Basisdatensatz</b>			<b>17'223</b>
<b>Einschränkung auf Haupttätigkeit</b>			<b>15'473</b>
<b>Anwendung der Filter der Lohnstatistik</b>			<b>13'211</b>
Branche	Land- und Forstwirtschaft	38	
Berufsgruppe	Land- und Forstwirtschaftskräfte	95	
Berufliche Stellung	Selbständig	181	
	In eigener Kapitalgesellschaft	1'389	
	Mitarbeitendes Familienmitglied	616	
	Lehrling	111	
Alter	Alter > 65 Jahre	212	
<b>Anzahl Fälle nach Anwendung der Filter</b>			<b>10'746</b>
Fehlende Werte		768	
<b>Anzahl verbleibender Beobachtungen</b>			<b>9'978</b>

Tabelle 1 Übersicht der verwendeten Filter

### 3. Berechnungsmodell

In diesem Kapitel wird das Berechnungsmodell für die Lohnrechner präsentiert. Es handelt sich dabei um ein log-lineares Regressionsmodell, wobei anschliessend – dem Vorgehen des Bundesamtes für Statistik (BFS 2008) folgend – aufgrund der Residuen eine Anpassung des Mittelwertschätzers gemacht wird, um die Quartilsschätzer zu erhalten. Im zweiten Abschnitt werden anschliessend die im Modell verwendeten Variablen ausführlich diskutiert.

#### 3.1 Regressionsmodell

Für die Modellierung des Vollzeitmonatslohns wird ein log-lineares Modell verwendet. Dabei wird angenommen, dass zwischen dem logarithmierten Vollzeitmonatslohn und den erklärenden Variablen ein linearer Zusammenhang besteht. Das heisst, der logarithmierte Vollzeitmonatslohn lässt sich durch eine Konstante, einer linearen Kombination der  $P$  unabhängigen Variablen sowie einem Residuum erklären:

$$\ln(Y_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_P x_{iP} + \varepsilon_i \quad (1)$$

Dabei bezeichnet  $Y_i$  den Vollzeitmonatslohn der  $i$ -ten Beobachtung,  $x_{ip}$  den Wert der Ausprägung der Variable  $p \in \{1, \dots, P\}$  für diese Beobachtung, und  $\varepsilon_i$  dessen stochastisches Residuum. Für die Verteilung der Residuen wird dabei eine Normalverteilung mit Erwartungswert 0 und einer konstanten Varianz  $\sigma^2$  angenommen. Es gilt daher für alle  $\varepsilon_i$ :

$$\varepsilon_i \sim N(0, \sigma^2) \quad (2)$$

Aus diesen Modellannahmen ergibt sich ein Erwartungswert des logarithmierten Vollzeitmonatslohn der  $i$ -ten Beobachtung:

$$E(\ln(Y_i)) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_P x_{iP} \quad (3)$$

Da die  $P + 1$  Parameter  $\beta_0 \dots \beta_P$  unbekannt sind, müssen diese anhand der Daten geschätzt werden. Diese Schätzung erfolgt mittels dem „Ordinary Least Square“(OLS)-Verfahren bei dem die  $\beta_p$  so bestimmt werden, dass die Summe der quadrierten Residuen minimiert wird.

Sind nun konkrete Ausprägungen gegeben (z.B. Alter = 30, Höchste abgeschlossene Ausbildung = Matura, etc.), kann anhand des Modell der logarithmierte Vollzeitmonatslohn vorhergesagt werden. Dieser lässt sich durch Bildung der Exponentialfunktion in einen Schätzer für den Vollzeitmonatslohn umwandeln.

Dem Vorgehen des Bundesamtes für Statistik (BFS 2008: 25) folgend werden die Quartile anhand der Verteilung der Residuen bestimmt. Für ein gegebenes Profil wird ein Quartil berechnet, indem zum Logarithmus des vom Modell vorhergesagten Gehalts das entsprechende Quartil der Residuen zum Modell addiert wird. Daraus wird der logarithmierte Quartilslohn berechnet, woraus wiederum der Quartilslohn abgeleitet werden kann.

Man erhält also das Quartil  $q$  gemäss der folgenden Formel:



$$\text{Vollzeitmonatslohn}_{i,q} = e^{\hat{Y}_i + \varepsilon_q} \quad (4)$$

wobei  $\hat{Y}_i$  für den aus dem Modell berechneten Schätzwert für die Beobachtung  $i$  steht und  $\varepsilon_q$  für das Quartil der Residuen.

## 3.2 Variablen

In diesem Abschnitt werden die Modellvariablen präsentiert. Zunächst wird auf die abhängige Variable eingegangen, bevor im darauffolgenden Abschnitt die unabhängigen Variablen diskutiert werden. Andere Variablen, auf die in der Modellierung verzichtet wurde, sind im Abschnitt 3.2.3 aufgeführt.

### 3.2.1 Abhängige Variable

Die abhängige Variable ist der logarithmierte, standardisierte Vollzeitmonatslohn. Die Arbeitgeberinnen und Arbeitgeber melden der Steuerverwaltung für alle Beschäftigten den während des vergangenen Jahres ausbezahlten Bruttolohn. Aus den Angaben bezüglich Beschäftigungsdauer und –grad aus dem liechtensteinischen Unternehmensregister werden die Vollzeitmonatslöhne berechnet. Als Basis zur Berechnung des Vollzeitmonatslohns dient der monatliche Bruttolohn gemäss steuerrechtlicher Definition, so wie er in den Lohnmeldungen der Arbeitgeber zu erfassen ist. Der Lohnsteuerpflichtige Bruttolohn umfasst gemäss der liechtensteinischen Steuerverwaltung die folgenden Lohnkomponenten:

- Besoldung, Gehalt, Lohn
- Teuerungszulagen
- Gratifikationen
- Leistungsprämien
- Überzeitvergütungen, Schicht- und Nachtarbeitszulagen, Sonn- und Feiertagsentschädigungen, Ferienentschädigungen
- Treueprämien, Dienstaltersgeschenke, Jubiläumsgaben
- Heirats- und Geburtszulagen
- Mietzinsbeiträge des Arbeitgebers
- Fahrtkostenentschädigungen vom Wohnort zum Arbeitsplatz
- Naturalleistungen (freie Verpflegung, Unterkunft, andere Naturalleistungen)
- Trinkgelder
- Sonstige Vergütungen (Provisionen, Umsatzbeteiligungen, Boni, Gewinnanteile, Verwaltungsrats-, Stiftungsrats und Aufsichtsrats honorare, Tantiemen, Arbeitgeberbeiträge an private Versicherungen des Arbeitnehmers, Abgangsentschädigungen bei Beendigung des Dienstverhältnisses)

Gemäss dieser Definition sind zudem der Lohnsteuerabzug sowie die Sozialbeiträge der Arbeitnehmer an die Sozialversicherungen (AHV/IV/FAK, ALV, NBU, Pensionsversicherung, Krankenkasse) enthalten, die von der Arbeitgebern direkt an die Steuerverwaltung bzw. Sozialversicherungen überwiesen werden. Schliesslich sind ebenso Taggelder, die vom Arbeitgeber an den Arbeitnehmer entrichtet werden, im lohnsteuerpflichtigen Bruttolohn enthalten.

Diese Angaben werden anschliessend auf einen einheitlichen Beschäftigungsgrad von 100% sowie auf eine volle Beschäftigung im ganzen Jahr hochgerechnet und anschliessend durch 12 dividiert. Im Vollzeitmonatslohn ist demnach auch 1/12 des 13. Monatslohns enthalten.

### 3.2.2 Unabhängige Variablen

Die Auswahl der unabhängigen Variablen basiert letztlich auf der Verfügbarkeit der Daten. Um die in Abschnitt 2.2.1 diskutierten zeitlichen Inkonsistenzen so klein wie möglich zu halten, erfolgte die Auswahl nach dem folgenden Prinzip: Personenbezogene Merkmale stammen aus der Volkszählung, arbeitsplatzbezogene Merkmale aus der Lohnstatistik. Tabelle 3 gibt einen Überblick über die verwendeten Variablen.

Variablenname	Art der Variable	Datenquelle
Alter	Diskret	Volkszählung
Höchste abgeschlossene Ausbildung	Kategorial	Volkszählung
Berufliche Stellung	Kategorial	Volkszählung
Berufsgruppe	Kategorial	Volkszählung
Wirtschaftszweig	Kategorial	Lohnstatistik
Unternehmensgrösse	Kategorial	Lohnstatistik
Zugehörigkeitsdauer	Diskret	Lohnstatistik
Beschäftigungsgrad	Stetig	Lohnstatistik

Tabelle 2 Übersicht unabhängige Variablen

Alter: Das Merkmal Alter stammt aus der Volkszählung und stellt somit das Alter am Anfang des Jahres dar. Das Alter wird in abgeschlossenen Altersjahren gemessen und wird durch einen linearen sowie durch einen quadratischen Term modelliert. Dieses Vorgehen wurde so gewählt, da die Verteilung der Vollzeitmonatslöhne nach Alter zunächst einen steigenden Trend aufweist, mit zunehmendem Alter jedoch abflacht. Der Median der Beobachtungen liegt bei 43 Jahren. Angaben zum Alter sind für sämtliche Beobachtungen vorhanden.

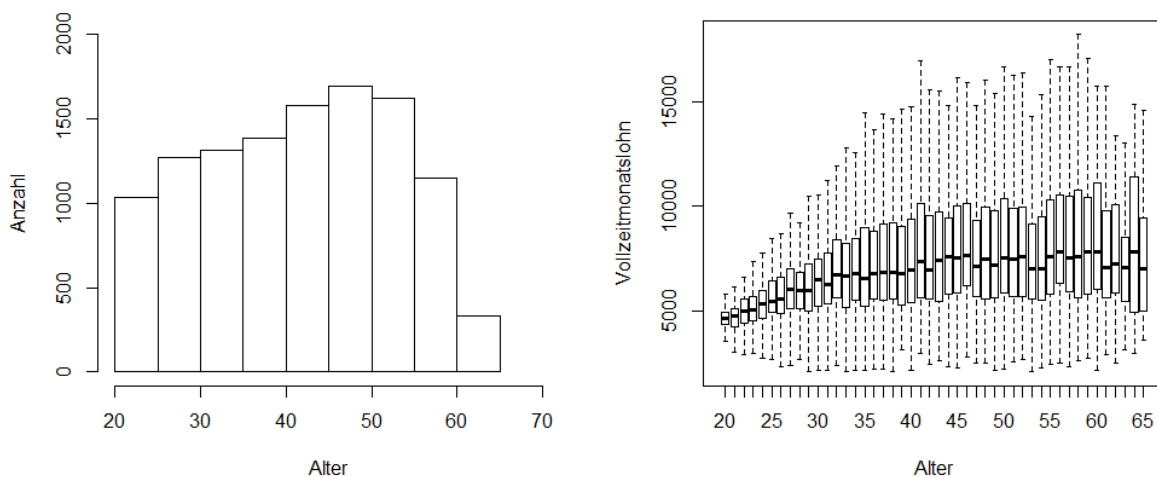


Abbildung 3 Histogramm und Boxplot (ohne Ausreisser) zum Merkmal Alter

**Höchste abgeschlossene Ausbildung:** Das Merkmal Höchste abgeschlossene Ausbildung stammt aus der Volkszählung und repräsentiert das Bildungsniveau der erwerbstätigen Bevölkerung. Um eine solide Referenzgruppe<sup>1</sup> zu erhalten, wurden diese Beobachtungen mit der Ausprägung „Keine Ausbildung „ mit der Gruppe der höchsten abgeschlossenen Ausbildung „Obligatorische Schule“ zusammengelegt. Das Merkmal wird als kategoriale Variable modelliert.

Rund 14% der Beobachtungen für die Angaben zur Ausbildung vorliegen, wiesen als höchste abgeschlossene Ausbildung die Ausprägung „Keine / Obligatorische Schule“ auf. 55% der Beobachtungen haben neben der obligatorischen Schule auch einen weiteren Sekundarschulabschluss als höchste abgeschlossene Ausbildung absolviert, während rund 32% einen tertiären Abschluss aufwiesen. Für 187 Beobachtungen liegen keine Angaben zur höchsten abgeschlossenen Ausbildung vor.

	Anzahl	Anteil	Medianlohn
210 Keine / Obligatorische Schule	1'464	13.9%	5'316
310 Diplommittelschule	543	5.1%	5'489
320 Berufliche Grundbildung	4'471	42.3%	6'381
330 Maturität	752	7.1%	6'799
410 Höhere Fach- /Berufsausbildung	1'169	11.1%	8'565
510 Höhere Fachschule	379	3.6%	10'019
520 Bachelor, Master	1'599	15.1%	10'036
610 Doktorat	182	1.7%	13'594
<i>Keine Angaben</i>	187	-	-

Tabelle 3 Höchste abgeschlossene Ausbildung

**Berufliche Stellung:** Das Merkmal Berufliche Stellung stammt aus der Volkszählung. Es handelt sich dabei um eine Selbstdeklaration der Befragten. Insgesamt gehören rund 76% der Beobachtungen der Kategorie „Mitarbeiter ohne Vorgesetztenfunktion“ an. 19% wiesen eine Vorgesetztenfunktion auf und fast 4% werden der Kategorie Direktionsmitglied zugeordnet. Das Merkmal wird als kategoriale Variable modelliert. Als Referenzgruppe dienen die Mitarbeiter ohne Vorgesetztenfunktion. Angaben zur beruflichen Stellung fehlen für insgesamt 74 Beobachtungen.

	Anzahl	Anteil	Medianlohn
Mitarbeiter ohne Vorgesetztenfunktion	8'211	76.4%	6'357
Mitarbeiter mit Vorgesetztenfunktion	2'066	19.2%	8'730
Direktionsmitglied	395	3.7%	17'466
<i>Keine Angaben</i>	74	-	-

Tabelle 4 Berufliche Stellung

**Berufsgruppe:** Das Merkmal Berufsgruppe stammt aus der Volkszählung 2015. Die Gruppen gemäss der ISCO-08-Klassifikation der ILO wurden aufgrund der Angabe zu gegenwärtigen Beruf abgeleitet. Gruppen, die über weniger als 50 Beobachtungen verfügten wurden in den jeweiligen Hauptgruppen als Sonstige zusammengefasst. Das Merkmal wird als kategoriale Variable modelliert. Als Referenzgruppe dienen die Bürokräfte und verwandte Berufe. Für insgesamt 674 Beobachtungen sind keine Angaben zur Berufsgruppe vorhanden.

<sup>1</sup> Wird in einem linearen Regressionsmodell eine kategoriale Variable verwendet, muss eine der Gruppen als Referenzgruppe definiert werden. Die Koeffizienten der Regressions-schätzung der anderen Gruppen in dieser Variable beziehen sich dabei immer auf den Unterschied zu dieser Referenzgruppe. Die Wahl der Referenzgruppe hat keine Auswirkung auf die Modellschätzung. Es wird jedoch typischerweise bei ordinalen Variablen die „unterste“ Kategorie verwendet. Bei nominalen Variablen wird hingegen häufig eine grosse Gruppe verwendet.

	Anzahl	Anteil	Median- lohn
<b>Führungskräfte</b>			
Geschäftsführer, Vorstände, leitende Verwaltungsbedienstete und Angehörige gesetzgebender Körperschaften	311	3.1%	14'405
Führungskräfte im kaufmännischen Bereich	143	1.4%	10'985
Führungskräfte in der Produktion und bei speziellen Dienstleistungen	153	1.5%	10'052
Sonstige Führungskräfte	229	2.3%	9'642
<b>Akademische Berufe</b>			
Naturwissenschaftler, Mathematiker und Ingenieure	167	1.7%	8'587
Lehrkräfte	498	4.9%	9'222
Betriebswirte und vglb. akademische Berufe	597	5.9%	8'454
Akademische und vglb. Fachkräfte in der Informations- und Kommunikationstechnologie	255	2.5%	9'421
Juristen, Sozialwissenschaftler und Kulturberufe	304	3.0%	9'270
Sonstige akademische Berufe	119	1.2%	8'071
<b>Techniker und gleichrangige nichttechnische Berufe</b>			
Ingenieurtechnische und vglb. Fachkräfte	544	5.4%	7'374
Assistenzberufe im Gesundheitswesen	224	2.2%	7'028
Nicht akademische betriebswirtschaftliche und kaufmännische Fachkräfte und Verwaltungsfachkräfte	1'200	11.9%	7'876
Nicht akademische juristische, sozialpflegerische, kulturelle und verwandte Fachkräfte	171	1.7%	6'667
Sonstige technische und gleichrangige nichttechnische Berufe	295	2.9%	6'735
<b>Bürokräfte und verwandte Berufe</b>			
Bürokräfte und verwandte Berufe	1'457	14.5%	6'835
<b>Dienstleistungsberufe und Verkäufer</b>			
Verkaufskräfte	210	2.1%	4'966
Betreuungsberufe	184	1.8%	5'354
Schutzkräfte und Sicherheitsbedienstete	163	1.6%	8'223
Sonstige Dienstleistungsberufe	414	4.1%	5'643
<b>Handwerks- und verwandte Berufe</b>			
Bau- und Ausbaufachkräfte sowie verwandte Berufe, ausgenommen Elektriker	409	4.1%	5'781
Metallarbeiter, Mechaniker und verwandte Berufe	274	2.7%	6'283
Elektriker und Elektroniker	159	1.6%	6'090
Berufe in der Nahrungsmittelverarbeitung, Holzverarbeitung und Bekleidungsherstellung und verwandte handwerkliche Fachkräfte	145	1.4%	5'390
Sonstige handwerkliche Berufe	76	0.8%	5'903
<b>Bediener von Anlagen und Maschinen und Montageberufe</b>			
Bedienung von Anlagen und Maschinen und Montageberufe	409	4.1%	5'904
<b>Hilfsarbeitskräfte</b>			
Reinigungspersonal und Hilfskräfte	365	3.6%	4'726
Hilfsarbeiter im Bergbau, im Bau, bei der Herstellung von Waren und im Transportwesen	346	3.4%	5'146
Sonstige Hilfsarbeitskräfte	251	2.5%	5'430
<i>Keine Angaben</i>	674	-	-

Tabelle 5 Berufsgruppen

**Wirtschaftszweig:** Die Angaben zum Wirtschaftszweig stammen aus der Lohnstatistik. Es wird dieselbe Zusammenfassung wie in der Lohnstatistik vorgenommen. Die Branchen sind gemäss der schweizerischen NOGA 2008 klassifiziert. Das Merkmal wird als kategoriale Variable modelliert. Als Referenzgruppe wird die Ausprägung „CI-CL Herstellung von elektronischen Erzeugnissen; Maschinen-, Fahrzeugbau“ gewählt. Angaben zur Branche sind für alle Beobachtungen vorhanden.

	Anzahl	Anteil	Medianlohn
B-CB Bergbau, H.v. Nahrung, H.v. Textilien	444	4.1%	5'388
CC H.v. Holzwaren, Papier, Druckerzgn.	130	1.2%	5'660
CD-CG H.v. chem. Erzgn., Glas-, Keramikwa.	237	2.2%	5'710
CH Metallerzeugung u. -bearb., Metallerzgn.	349	3.2%	6'318
CI-CL H.v. elektron. Erzgn. Masch.-, Fahrzeugbau	1'018	9.5%	7'204
CM-E Sonst. Warenh., Energieversorg.	445	4.1%	6'721
F Baugewerbe	889	8.3%	5'936
G Handel, Rep. Fahrzeuge	604	5.6%	5'621
H Verkehr und Lagerei	293	2.7%	6'392
I Gastgewerbe	237	2.2%	4'438
J Medien, Telekommunik., Informatik	221	2.1%	7'738
K Finanz- u. Versicherungsdienstl.	1'232	11.5%	8'759
L, N Wohnungswesen, Sonst. wirtschaftl. Dienstl.	452	4.2%	4'667
MAA Rechts- u. Steuerberat., Wirtschaftsprüfung	937	8.7%	7'955
MAB Verw. v. Unternehmen, Unternehmensberat.	66	0.6%	7'223
MAC-MC Architektur, F&E, snst. techn. Tätigkeiten	267	2.5%	6'879
O, U Öffentliche Verwaltung, Zollbehörden	1'247	11.6%	8'042
P Erziehung u. Unterricht	521	4.8%	9'213
QA Gesundheitswesen	453	4.2%	5'895
QB Heime u. Sozialwesen	354	3.3%	5'964
R, S Unterhaltung, Sonst. Dienstl.	282	2.6%	6'845
T Private Haushalte	68	0.6%	5'327
<i>Keine Angaben</i>	0	-	-

Tabelle 6 Wirtschaftszweig

**Unternehmensgrösse:** Die Angaben zur Unternehmensgrösse stammen aus der Lohnstatistik. Die numerische Variable wurde zu einer kategorialen Variable mit vier Ausprägungen zusammengefasst. Die Unternehmen mit 1-9 Beschäftigten dienen als Referenzgruppe. Es liegen Angaben zu sämtlichen Beobachtungen vor.

	Anzahl	Anteil	Medianlohn
1-9 Beschäftigte	1'622	15.1%	6'175
10-49 Beschäftigte	2'835	26.4%	6'821
50-249 Beschäftigte	2'769	25.8%	6'500
Mehr als 250 Beschäftigte	3'520	32.8%	7'409
<i>Keine Angaben</i>	0	-	-

Tabelle 7 Unternehmensgrösse

**Zugehörigkeitsdauer:** Die Variable Zugehörigkeitsdauer stammt aus den Daten der Lohnstatistik und gibt die Anzahl Jahre wieder, die die Person im entsprechenden Unternehmen beschäftigt ist. Die Zugehörigkeitsdauer wird als diskrete Variable sowie durch einen quadratischen Term modelliert. Dieses Vorgehen wurde so gewählt, da die Verteilung der Vollzeitmonatslöhne nach Zugehörigkeitsdauer zunächst einen steigenden Trend aufweist, mit zunehmendem Alter jedoch abflacht. Der Median der Beobachtungen liegt bei 7 Jahren. Angaben zur Zugehörigkeitsdauer sind für sämtliche Beobachtungen vorhanden. Bei 29 Beobachtungen war die Zugehörigkeitsdauer grösser als das Alter minus 20 Jahre. Diese Beobachtungen wurden aus der Analyse ausgeschlossen.

**Beschäftigungsgrad:** Die Variable Beschäftigungsgrad stammt aus der Lohnstatistik und wird als kontinuierliche Variable modelliert. Der Wertebereich reicht aufgrund der angesetzten Untergrenze von 0.15 bis 1.

### 3.2.3 Nicht verwendete Variablen

Geschlecht: Die Variable Geschlecht wurde nicht in das Modell aufgenommen, weil die Aussagekraft des Koeffizienten im Rahmen des Lohnrechners eingeschränkt ist. Die prozentuale Differenz zwischen den Männern und Frauenlöhnen wäre für jede mögliche Kombination der Variablen konstant. Für die Nutzerinnen und Nutzer des Lohnrechners wäre dies jedoch nicht unbedingt ersichtlich. Der Lohnrechner versucht in diesem Sinne einen gemeinsamen Mittelwert von Frauen- und Männerlöhnen zu berechnen. Wird das Geschlecht als zusätzliche erklärende Variable in das Modell aufgenommen, zeigt sich, dass die Frauen im Mittel ungefähr einen 12.5% tieferen Lohn aufweisen als die Männer. Für eine ausführliche Analyse der Frauen und Männerlöhne wird auf die Publikation „Löhne von Frauen und Männern 2016“ des Amtes für Statistik verwiesen.

Staatsbürgerschaft/ Bewilligungstyp: Der Ausschluss der Nationalität erfolgt aus derselben Begründung wie die Variable Geschlecht. In diesem Fall ist der Koeffizient jedoch deutlich geringer und nur schwach signifikant. Fügt man die Variable liechtensteinische Staatsbürgerschaft als Dummy-Variable hinzu, sagt das dem Lohnrechner zugrundeliegende Modell einen um 1.0% höheren Lohn für Personen mit liechtensteinischer Staatsbürgerschaft voraus. Ebenso wird auf die Unterscheidung zwischen den verschiedenen ausländerrechtlichen Bewilligungstypen verzichtet.

Zivilstand: Das Merkmal Zivilstand stammt aus der Volkszählung 2015. Aus theoretischen Überlegungen gibt es heutzutage keinen Grund, warum eine ledige Person mehr oder weniger verdienen würde als eine verheiratete Person. Das Büro Bass (2017) weist jedoch auf einen Interaktionseffekt zwischen Geschlecht und Zivilstand hin. Da verheiratete Frauen eher ihren Beschäftigungsgrad reduzieren als ledige. Auch in den Daten zeigt sich, dass die verheiratete Personen eher mehr verdienen als ledige. Dies könnte aber auch durch das Alter verursacht werden, da verheiratete Personen im Schnitt älter sind als ledige. Ausschlaggebend für den Ausschluss der Variable war aber letztlich, dass sich trotz solcher theoretischen Zusammenhänge für die Nutzerinnen und Nutzer des Lohnrechners nicht ersichtlich wäre, warum sich der Zivilstand auf den Lohn auswirken sollte. Die Angabe zum Zivilstand ist für alle Beobachtungen vorhanden.

Arbeitsort: Aufgrund der deutlich geringeren Löhne im benachbarten Österreich kann argumentiert werden, dass Personen, welche in einer Gemeinde arbeiten, die näher an der österreichischen Grenze liegt, eher dem Wettbewerbsdruck ausgesetzt sind. Da Liechtenstein jedoch eine sehr kleine geografische Einheit darstellt, wird auf eine weitere Untergliederung verzichtet. Angaben zum Arbeitsort sind für alle Beobachtungen vorhanden.

Sozioprofessionelle Kategorie: Das Merkmal sozioprofessionelle Kategorie stammt aus der Volkszählung 2015. Die Zugehörigkeit zu einer Kategorie wird nicht direkt erfragt, sondern wird aus verschiedenen Merkmalen gebildet. Hierfür werden, die Angaben zur beruflichen Tätigkeit, zur Stellung im Beruf (Erwerbsstatus), zur höchsten abgeschlossenen Ausbildung und bei den Selbstständigen und den Direktoren die Betriebsgrösse genutzt. Die Variable ist demnach eine Kombination von Variablen, die bereits in der Analyse verwendet werden. Sie wurde daher nicht als eigenständige erklärende Variable in das Modell aufgenommen. Die Angaben zu sozioprofessionellen Kategorie sind für alle Beobachtungen vorhanden.

## 4. Resultate

Dieses Kapitel befasst sich mit den Resultaten der Regressionsanalyse. Zunächst werden die Resultate des Regressionsmodells sowie die Quartilsschätzungen vorgestellt. Anschliessend wird kurz auf die Problematik der Multikollinearität eingegangen, bevor die Modellannahmen überprüft werden.

### 4.1 Regressionsmodell

In diesem Abschnitt wird auf die Resultate des Regressionsmodells eingegangen. Dabei werden zunächst die Modellkoeffizienten diskutiert, bevor im darauffolgenden Abschnitt auf die Modellgüte eingegangen wird.

#### 4.1.1 Modellkoeffizienten

In Tabelle 8 sind die Resultate des Regressionsmodells abgebildet. Es zeigt sich, dass die gewählten erklärenden Variablen einen signifikanten Einfluss auf die erklärte Variable haben. Als abhängige Variable dient, wie bereits in Abschnitt 3.2.1 ausführliche diskutiert, der logarithmierte, standardisierte Vollzeitmonatslohn. Die Koeffizienten können daher ungefähr als prozentuale Veränderung interpretiert werden. Der exakte Wert bestimmt sich aus  $e^{\beta} - 1$ , wobei  $\beta$  für den entsprechenden Koeffizienten steht. Die Standardfehler wurden aufgrund beobachteter Heteroskedastizität mittels Huber-White-Korrektur geschätzt<sup>2</sup>.

Der Koeffizient von Alter hat einen positiven Effekt, während der quadrierte Wert ein negatives Vorzeichen aufweist. Dies führt zu der erwarteten konkaven Funktion, d.h. der Lohn steigt in jungen Jahren stärker als in weiteren Verlauf des Erwerbslebens. Das höchste mittlere Einkommen wird demnach im Modell im Alter von 56 Jahren erreicht.

Die Koeffizienten des Faktors Höchste abgeschlossene Ausbildung sind wie erwartet alle positiv. Gegenüber der Referenzgruppe<sup>1</sup> „Ohne Ausbildung/Obligatorische Schule“ führt eine berufliche Grundbildung zu einer Zunahme des Lohns um 9%, eine höhere Fach- und Berufsausbildung um 26% und eine höhere Fachschule sogar um 38%. Auf dem akademischen Bildungsweg führt eine Matura zu einem 14% höheren Lohn, ein Bachelor/Master zu einer Zunahme von 42% und ein Doktorat zu einem um 62% höhere Verdienst. Einzig der Koeffizient Diplommittelschule ist nicht signifikant.

---

<sup>2</sup> Die Verfahren der schliessenden Statistik beruhen unter anderem darauf, dass die Varianz von  $Y|x$  bzw.  $\epsilon|x$  unabhängig von  $x$  also konstant ist. Ist diese Bedingung verletzt, spricht man von Heteroskedastizität. Auch bei Vorliegen von Heteroskedastizität werden die Koeffizienten des Regressionsmodells korrekt geschätzt, allerdings sind die Standardfehler nicht korrekt. Dafür gibt es jedoch Verfahren wie die hier verwendete Huber-White-Korrektur, um die Standardfehler (approximativ) zu korrigieren.

# Dokumentation Lohnrechner

Coefficients:

	Estimate	Std. Error	t-value	Pr(> t )	
(Intercept)	7.6786	0.046	167.907	< 2.2e-16	***
Alter	0.0328	0.002	17.236	< 2.2e-16	***
Alter^2	-0.0003	0.000	-12.513	< 2.2e-16	***
Diplommittelschule	0.0130	0.013	0.973	0.331	
Berufliche Grundbildung	0.0838	0.009	9.196	< 2.2e-16	***
Maturität	0.1318	0.015	8.984	< 2.2e-16	***
Höhere Fach- und Berufsausbildung	0.2305	0.012	18.893	< 2.2e-16	***
Höhere Fachschule	0.3237	0.021	15.218	< 2.2e-16	***
Bachelor, Master	0.3476	0.015	23.944	< 2.2e-16	***
Doktorat	0.4846	0.037	13.152	< 2.2e-16	***
Mitarbeiter mit Vorgesetztenfunktion	0.1664	0.008	20.530	< 2.2e-16	***
Direktionsmitglied	0.5875	0.030	19.339	< 2.2e-16	***
11 Geschäftsführer, Vorstände, leit. Verwaltungsbedienstete	0.2383	0.034	7.033	0.000	***
12 Führungskräfte im kaufmännischen Bereich	0.1714	0.028	6.101	0.000	***
13 Führungskräfte in der Prod. und bei spez. Dienstleistungen	0.0664	0.032	2.080	0.038	*
19 Sonstige Führungskräfte	0.0997	0.020	5.106	0.000	***
21 Naturwissenschaftler, Mathematiker und Ingenieure	0.0327	0.024	1.374	0.169	
23 Lehrkräfte	0.1428	0.031	4.582	0.000	***
24 Betriebswirte und vglb. akademische Berufe	0.1117	0.016	6.815	0.000	***
25 Akad. u. Fachkr. Informations- und Kommunikationstechn.	0.1617	0.017	9.527	< 2.2e-16	***
26 Juristen, Sozialwissenschaftler und Kulturberufe	0.0730	0.024	3.068	0.002	**
29 Sonstige akademische Berufe	0.0519	0.033	1.567	0.117	
31 Ingenieurtechnische und vglb. Fachkräfte	0.0257	0.012	2.079	0.038	*
32 Assistenzberufe im Gesundheitswesen	0.1572	0.021	7.338	0.000	***
33 Nicht akad. betriebsw. und kaufm. Fachkr. und Verw.fachkr.	0.0803	0.011	7.347	0.000	***
34 Nicht akad. jur., sozialpfl., kult. und verwandte Fachkräfte	0.0164	0.025	0.659	0.510	
39 Sonstige techn. und gleichrangige nichttechnische Berufe	0.0052	0.015	0.346	0.729	
52 Verkaufskräfte	-0.1062	0.024	-4.505	0.000	***
53 Betreuungsberufe	0.0253	0.023	1.091	0.275	
54 Schutzkräfte und Sicherheitsbedienstete	0.0588	0.022	2.655	0.008	**
59 Sonstige Dienstleistungsberufe	-0.0674	0.016	-4.316	0.000	***
71 Bau- und Ausbaufachkr. sowie verw. Berufe, ausg. Elektriker	-0.0611	0.015	-4.051	0.000	***
72 Metallarbeiter, Mechaniker und verwandte Berufe	0.0193	0.014	1.382	0.167	
74 Elektriker und Elektroniker	-0.0185	0.019	-0.950	0.342	
75 Nhrngsm., Holzverarb. u. Bekleid.hrst., verw. handw. Fachkr.	-0.1283	0.021	-6.040	0.000	***
79 Sonstige handwerkliche Berufe	-0.0322	0.034	-0.950	0.342	
89 Bedienung von Anlagen und Maschinen und Montageberufe	-0.0662	0.015	-4.558	0.000	***
91 Reinigungspersonal und Hilfskräfte	-0.2456	0.016	-15.407	< 2.2e-16	***
93 Hilfsarb. Bergbau, Bau, Herst. Waren und Transportwesen	-0.1673	0.016	-10.764	< 2.2e-16	***
99 Sonstige Hilfsarbeitskräfte	-0.1395	0.016	-8.579	< 2.2e-16	***
B-CB Bergbau, H.v. Nahrung, H.v. Textilien	-0.1585	0.017	-9.150	< 2.2e-16	***
CC H.v. Holzwaren, Papier, Druckerzgn.	-0.0470	0.022	-2.160	0.031	*
CD-CG H.v. chem. Erzgn., Glas-, Keramikwa.	-0.1346	0.022	-6.133	0.000	***
CH Metallerzeugung u. -bearb., Metallerzgn.	-0.0969	0.015	-6.507	0.000	***
CM-E Sonst. Warenh., Energieversorg.	-0.0734	0.015	-4.779	0.000	***
F Baugewerbe	-0.0506	0.014	-3.570	0.000	***
G Handel, Rep. Fahrzeuge	-0.1162	0.017	-6.698	0.000	***
H Verkehr und Lagerei	-0.1147	0.020	-5.867	0.000	***
I Gastgewerbe	-0.2720	0.024	-11.544	< 2.2e-16	***
J Medien, Telekommunik., Informatik	-0.0341	0.024	-1.428	0.153	
K Finanz- u. Versicherungsdienstl.	0.1216	0.016	7.610	0.000	***
L, N Wohnungswesen, Sonst. wirtschaftl. Dienstl.	-0.1544	0.020	-7.553	0.000	***
MAA Rechts- u. Steuerberat., Wirtschaftsprüfung	0.0975	0.017	5.902	0.000	***
MAB Verw. v. Unternehmen, Unternehmensberat.	-0.0864	0.062	-1.394	0.163	
MAC-MC Architektur, F&E, sonst. techn. Tätigkeiten	-0.0839	0.022	-3.761	0.000	***
O, U Öffentliche Verwaltung, Zollbehörden	-0.0148	0.013	-1.107	0.269	
P Erziehung u. Unterricht	-0.0292	0.032	-0.913	0.361	
QA Gesundheitswesen	-0.1302	0.020	-6.358	0.000	***
QB Heime u. Sozialwesen	-0.2324	0.022	-10.367	< 2.2e-16	***
R, S Unterhaltung, Sonst. Dienstl.	-0.1069	0.024	-4.481	0.000	***
T Private Haushalte	-0.0423	0.045	-0.945	0.345	
10-49 Beschäftigte	0.0473	0.010	4.759	0.000	***
50-249 Beschäftigte	0.0649	0.010	6.327	0.000	***
Mehr als 250 Beschäftigte	0.1053	0.012	9.049	< 2.2e-16	***
Zugehörigkeitsdauer	0.0082	0.001	7.412	0.000	***
Zugehörigkeitsdauer^2	-0.0001	0.000	-2.218	0.027	*
Beschäftigungsgrad	0.0858	0.016	5.356	0.000	***

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2725 on 9911 degrees of freedom

Multiple R-squared: 0.6279, Adjusted R-squared: 0.6254

F-statistic: 214.9 on 66 and 9911 DF, p-value: < 2.2e-16

Tabelle 8 Regressionsmodell



Auch die berufliche Stellung und die Berufsgruppe haben einen signifikanten Einfluss auf die Höhe des Lohns. In Bezug auf die berufliche Stellung zeigt sich, dass Mitarbeiter mit Vorgesetztenfunktion im Mittel 18% und Direktionsmitglieder 80% mehr verdienen als Mitarbeiter ohne Vorgesetztenfunktion. Bezüglich Berufsgruppe schätzt das Modell, dass z.B. in Assistenzberufen im Gesundheitswesen tätige Personen 17% mehr verdienen als in der Referenzgruppe Bürokräfte. Die grösste positive Abweichung ergibt sich bei Geschäftsführenden, Vorständen und leitenden Verwaltungsbediensteten mit einer Zunahme um 27% gegenüber der Referenzgruppe. Die grösste negative Abweichung zeigt sich hingegen beim Reinigungspersonal mit -22%.

Die Koeffizienten der Branchen geben jeweils den Unterschied zur Referenzgruppe „CI-CL Herstellung von elektronischen Erzeugnissen, Maschinen-, Fahrzeugbau“ wieder. Die höchste positive Abweichung findet sich im Wirtschaftszweig „K Finanz- u. Versicherungsdienstl.“, wo die Beschäftigten 13% mehr verdienen als in der Referenzgruppe. Die grösste negative Abweichung findet sich im Gastgewerbe, wo die Beschäftigten 24% weniger verdienen, als in der Referenzgruppe.

Auch die Unternehmensgrösse und die Zugehörigkeitsdauer zum Unternehmen wirken sich positiv auf den Lohn aus. Personen, die in einem Unternehmen mit 10-49 oder 50-249 Beschäftigten arbeiten, verdienen rund 5% bzw. 7% mehr als Arbeitskräfte in kleinen Unternehmen mit bis zu 9 Beschäftigten. In grösseren Unternehmen mit mehr als 250 Beschäftigten sind es sogar 11%. In Bezug auf die Zugehörigkeitsdauer sagt das Modell einen zunehmenden Lohn voraus, wobei die Zunahme aber mit der einer höheren Zugehörigkeitsdauer kleiner wird.

Der Faktor Beschäftigungsgrad weist schliesslich einen Koeffizienten von 0.086 auf. Je zusätzlichem Prozentpunkt an Beschäftigungsgrad steigt der vorausgesagte standardisierte Vollzeitmonatslohn um ca. 0.086%.

Die Koeffizienten weisen somit alle in die zu erwartende Richtung. Bei der Interpretation der Resultate ist jedoch zu beachten, dass die einzelnen Effekte häufig zusammen auftreten. Es wurden keine Interaktionen festgestellt, die zu einem deutlich besseren Modell geführt hätten. Daher wurde deren Modellierung zugunsten eines einfacheren Modells verzichtet.

#### **4.1.2 Modellgüte**

Die Stärke des linearen Zusammenhangs zwischen der abhängigen und den unabhängigen Variablen wird durch das Bestimmtheitsmass  $R^2$  gemessen. Das  $R^2$  entspricht dabei dem Anteil der Varianz der abhängigen Variable, welcher durch die unabhängigen Variablen erklärt wird. Da jede zusätzliche Variable zu einem höheren  $R^2$  führt, wird im korrigierten  $R^2$  versucht diesem Umstand Rechnung zu tragen, indem ein „Strafterm“ für jede zusätzliche Variable eingeführt wird.

Das gewählte Modell erreicht ein  $R^2$  von 0.6279, bzw. ein korrigiertes  $R^2$  von 0.6254. Das bedeutet, dass rund 60% der Varianz des Vollzeitmonatslohns durch die aufgeführten Variablen erklärt werden können. Damit liegt das Modell im selben Bereich wie das vom Bun-

desamt für Statistik (BFS 2008) verwendete. Dort liegt das  $R^2$  im Mittel<sup>3</sup> bei 0.6107 bzw. das adjustierte  $R^2$  bei 0.5983.

Der globale F-Test zeigt, dass auf einem Signifikanzniveau von 5% tatsächlich ein Zusammenhang zwischen den erklärenden Variablen und der erklärten Variable besteht. Tabelle 9 gibt die partiellen F-Tests wieder. Mithilfe der partiellen F-Tests lässt sich prüfen, ob gewisse Modellkoeffizienten zusammen Null sind. Dies ist insbesondere bei kategorialen Variablen sehr wertvoll, da alle Ausprägungen zusammen auf ihren Einfluss geprüft werden. Wie aus Tabelle 9 hervorgeht, weisen sämtliche erklärenden Variablen einen p-Wert kleiner als 0.05 auf, weshalb mit einer Sicherheit von 95% behauptet werden kann, dass bei allen Variablen ein echter Zusammenhang besteht.

	Df	Sum of Sq	RSS	AIC	F value	Pr(>F)	
<none>			736.08	-28'577			
Alter	1	21.362	757.44	-25593	287.625	< 2.2e-16	***
Alter^2	1	11.676	747.75	-25722	157.208	< 2.2e-16	***
Ausbildung	7	87.034	823.11	-24776	167.412	< 2.2e-16	***
Berufliche Stellung	2	104.510	840.59	-24556	703.593	< 2.2e-16	***
Berufsgruppe	28	60.066	796.14	-25150	28.885	< 2.2e-16	***
Wirtschaftszweig	21	68.325	804.40	-25033	43.808	< 2.2e-16	***
Grössenkategorie	3	7.291	743.37	-25784	32.725	< 2.2e-16	***
Zugehörigkeitsdauer	1	4.581	740.66	-25817	61.684	0.000	***
Zugehörigkeitsdauer^2	1	0.417	736.49	-25873	5.613	0.018	*
Beschäftigungsgrad	1	2.722	738.80	-25842	36.650	0.000	***

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Tabelle 9 Partielle F-Tests

Eine weitere Möglichkeit zur Überprüfung der Modellgüte besteht im Vergleich zwischen den vom Modell vorhergesagten sowie den beobachteten Werten. In Abbildung 13 sind die beobachteten Werte gegen die vorhergesagten geplottet. Links sind die logarithmierten Werte zu sehen, rechts die Werte in Franken. Würden sämtliche Werte korrekt vorhergesagt, lägen sämtliche Punkte auf der Diagonalen. Dies kann jedoch realistischweise nicht erreicht werden, da dies voraussetzen würde, dass die Lohnangaben perfekt durch die genannte Variablen vorhergesagt werden. Zur Beurteilung der Modellgüte sollte daher die Streuung der Residuen um die Diagonale betrachtet werden. Liegen die Beobachtungen gleichmässig um die Diagonale, deutet dies darauf hin, dass es keinen systematischen Vorhersagefehler gibt. In Abbildung 5 zeigt sich, dass ein Grossteil der Beobachtungen relativ gleichmässig um die Diagonale liegt. Es ist jedoch eine leichte Schiefelage der vorhergesagten Werte zu erkennen. Die Löhne im untersten Bereich werden dadurch tendenziell leicht überschätzt. Umgekehrt ist es wenig überraschend, dass das Modell die Löhne der Topkader unterschätzt, da diese sich in den erfassten Merkmalen nicht von Kadern in kleineren Unternehmen unterscheiden.

<sup>3</sup> Das Bundesamt für Statistik schätzt für jeden Wirtschaftszweig ein separates Modell. Der hier aufgeführte Wert gibt den Mittelwert der  $R^2$  der verschiedenen Modelle wieder.

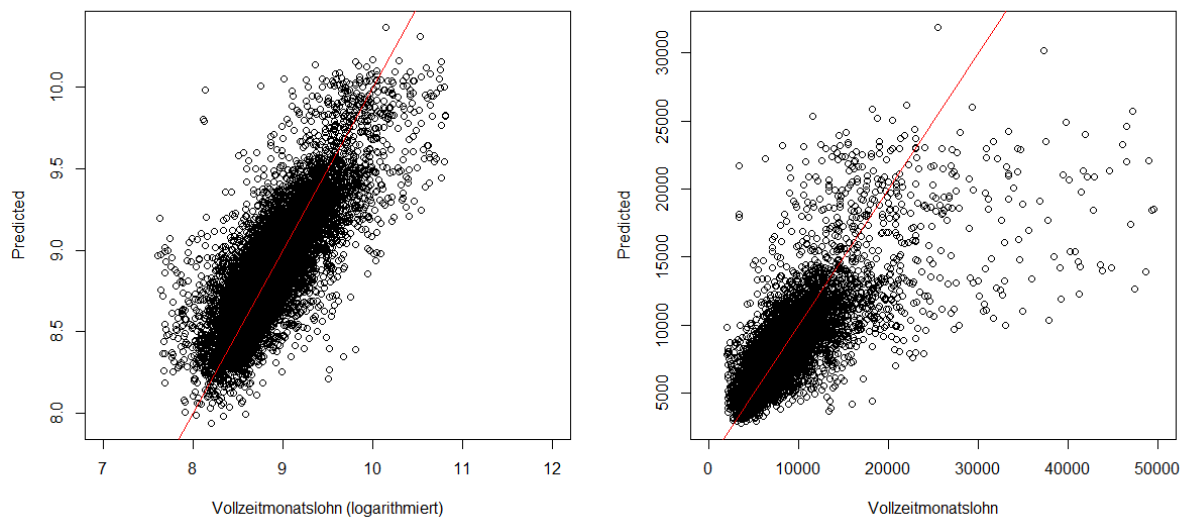


Abbildung 4 Actual-vs-Fitted Plot (ohne beobachtete Vollzeitmonatslöhne über CHF 50'000)

## 4.2 Schätzung der Quartile

Die Quartile werden anhand der in Abschnitt 3.1 beschriebenen Methode angewandt. Für das für den Lohnrechner gewählte Modell ergeben sich die folgenden  $\varepsilon_q$ :

$\varepsilon_{25}$	-0.142343302
$\varepsilon_{50}$	-0.006027509
$\varepsilon_{75}$	0.131991755

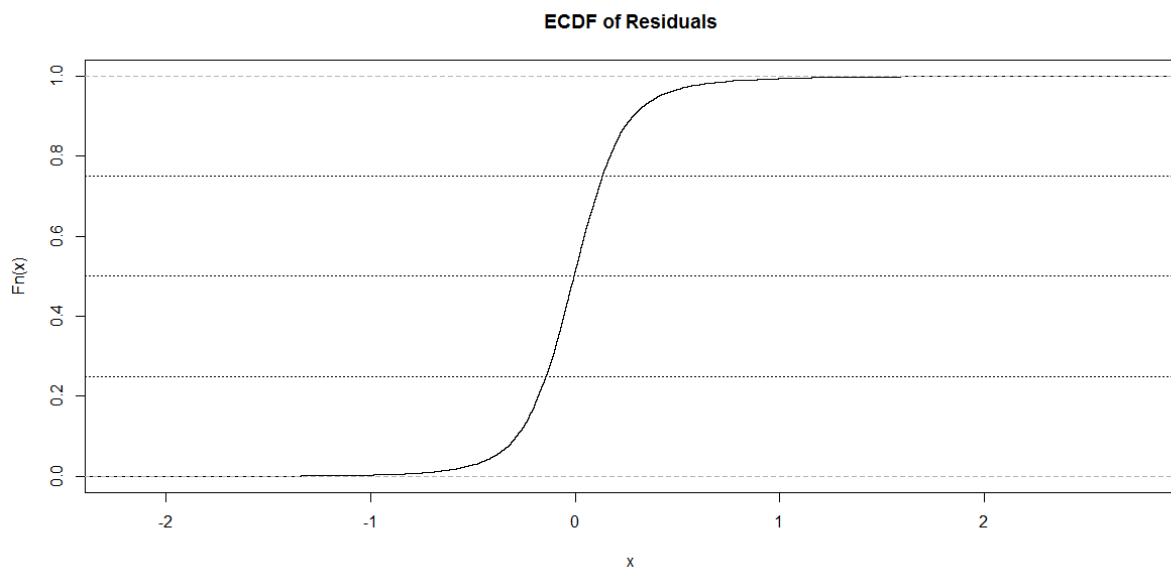


Abbildung 5 Empirical Cumulative Distribution Function of Residuals

### 4.3 Multikollinearität

Eine hohe Korrelation zwischen den unabhängigen Variablen wird als Multikollinearität bezeichnet. Unabhängige Variablen, die untereinander korreliert sind, teilen sich einen gewissen Effekt auf die abhängige Variable. Je stärker dieser Zusammenhang ist, umso eher ist die ceteris-paribus-Interpretation der Effekt in Frage zu stellen. D.h. die Präzision der Koeffizienten des Regressionsmodells ist ungenauer, als wenn die unabhängigen Variablen nicht korreliert wären.

Die Betrachtung der Korrelationsmatrix zeigt, dass nebst den zu erwartenden Faktoren kaum starke paarweise Korrelationen auftreten. Eine sehr hohe Korrelation wird zwischen den Variablen Alter und Alter<sup>2</sup> gemessen (0.99). Die Korrelation dieser Faktoren mit der Zugehörigkeitsdauer ist ebenfalls relativ stark (0.45). Ansonsten zeigt sich eine relativ starke positive Korrelation zwischen der Berufsgruppe Lehrkräfte und dem Wirtschaftszweig Erziehung und Unterricht (0.84), Bau- und Ausbaufachkräfte sowie verwandte Berufe und Baugewerbe (0.57) und Dienstleistungsberufe und Verkäufer und Gastgewerbe (0.40). Die grössten negativen Korrelationen ( $\approx -0.40$ ) bestehen hingegen bei Ausprägungen, derselben kategorialen Variable.

Als zusätzliches Mass zur Bestimmung der Multikollinearität in einem Modell mit kategorialen unabhängigen Variablen werden Generalized Variance Inflation Factors (GVIFs) herangezogen. Hierfür wird in einem ersten Schritt jede unabhängige Variable auf alle anderen unabhängigen Variablen regressiert. Die daraus berechneten Variance Inflation Factors (VIFs) geben an, wie stark das Konfidenzintervall des Koeffizienten  $\beta_j$  im Gegensatz zu Daten anwächst, die unkorreliert wären. Dabei sollten die VIFs nicht grösser als vier sein. Daraus ergibt sich ein kritischer Wert der für Anzahl Merkmalausprägungen  $p$  angepasst GVIFs von 2. Wie aus Tabelle 16 hervorgeht, liegen ausser den offensichtlich korrelierten Variablen Alter, Alter<sup>2</sup>, Zugehörigkeitsdauer und Zugehörigkeitsdauer<sup>2</sup>, die Werte unterhalb dieser kritischen Grenze und sehr nahe an der theoretischen Untergrenze von 1. Aus der Überprüfung der Korrelationsmatrix sowie der GVIFs lässt sich folgern, dass Multikollinearität im Modell keine kritische Rolle spielt.

	GVIF	Df	GVIF <sup>1/(2*Df)</sup>
Alter	63.4571	1	7.9660
Alter <sup>2</sup>	62.5591	1	7.9094
Höchste abgeschlossene Ausbildung	2.8375	7	1.0773
Berufliche Stellung	1.6962	2	1.1412
Berufsgruppe	178.5849	28	1.0970
Branche (NOGA 2008)	138.8908	21	1.1246
Unternehmensgrösse	2.4142	3	1.1582
Zugehörigkeitsdauer	9.6337	1	3.1038
Zugehörigkeitsdauer <sup>2</sup>	8.8179	1	2.9695
Beschäftigungsgrad	1.4226	1	1.1927

Tabelle 10 Variance Inflation Factors

### 4.4 Modellüberprüfung

Das Regressionsmodell basiert auf der Annahme, dass das Modell korrekt spezifiziert wurde und sich daher der Erwartungswert der abhängigen Variable tatsächlich durch die gewählten Regressoren beschreiben lässt. Zur Überprüfung dieser Annahme wird typischerweise der Residual-vs-Fitted-Plot herangezogen. Im Residual-vs-Fitted-Plot werden die Residuen des

Regressionsmodells gegen die vorhergesagten Werte geplottet. Ist die Modellgleichung korrekt spezifiziert, sollte es in diesem Plot keinen systematischen Zusammenhang zwischen den Residuen und den gefitteten Werten geben. Wie aus Abbildung 6 ersichtlich ist, ist dies im gegebenen Modell relativ gut erfüllt. Zwar lässt sich ein leicht u-förmiger Zusammenhang feststellen. Da die Krümmung jedoch relativ gering ist, ist dies noch kein Grund zur Besorgnis. Je stärker die Voraussetzung verletzt ist, umso mehr werden die Ergebnisse in Frage gestellt.

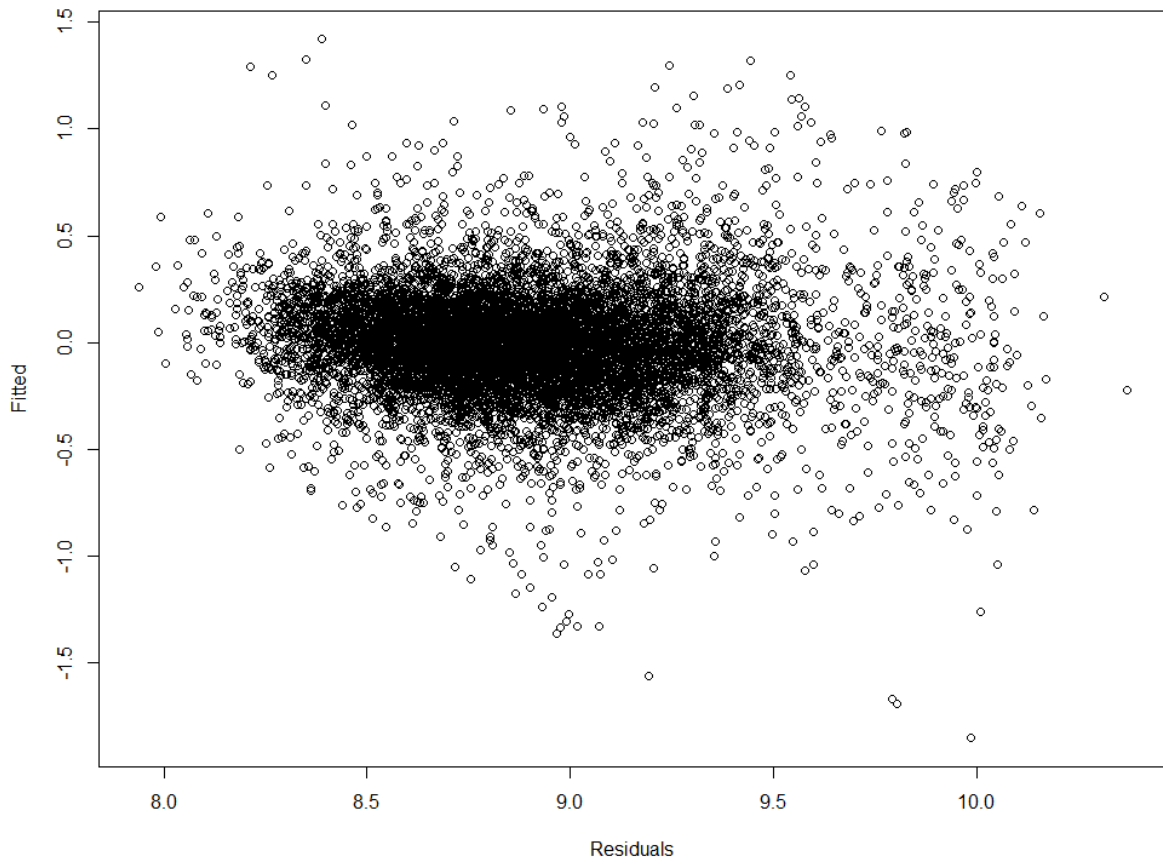


Abbildung 6 Residual-vs-Fitted Plot (ohne beobachtete Vollzeitmonatslöhne über CHF 50'000)

## Literaturverzeichnis

BFS, 2008. *Enquête suisse sur la structure des Salaires 2006. Aspects méthodologiques du modèle des salaires „Salarium“*: Rapport de méthodes. Neuchâtel.

Amt für Statistik, 2018. *Lohnstatistik 2016*. Vaduz.