Problem Set

## Mindestlöhne und Unternehmensgewinne - Eine interaktive Analyse mit R

## Inhaltsübersicht

Herzlich Willkommen zu diesem interaktiven RTutor Problem Set. Im Rahmen meiner Bachelorarbeit gehen wir den Zusammenhängen von Mindestlöhnen und der Unternehmensprofitabilität auf den Grund. Die Basis bildet der Artikel [Minimum Wages and Firm Profitability](https://www.aeaweb.org/articles?id=10.1257/app.3.1.129) von Mirko Draca, Stephen Machin und John Van Reenen, erschienen im Januar 2011 im American Economic Journal: Applied Economics.

Kommen Sie mit auf eine Reise ins Vereinigte Königreich und seine Arbeitswelt.

Durch die Verknüpfung von inhaltlichen Hintergründen, fachlicher Analyse von Ergebnissen, persönlichen Einschätzungen und technischer Arbeit mit der Statistiksoftware R sollen Sie durch dieses Problem Set von einem fundierten Wissenszuwachs profitieren.  
Es ist das Ziel, dass Sie sich nach der geführten Erarbeitung des Themas kritisch und differenziert mit politischen Mindestlohnforderungen auseinandersetzen können und dazu in der Lage sind, Ihre Meinungsbildung ökonomisch zu begründen.  
Zudem wird Ihnen die Möglichkeit geboten, sich kritisch mit Fachliteratur auseinanderzusetzen und Analysen einzuordnen.

### Inhalt

1. Einführung in den Mindestlohn
2. Datenüberblick
3. Difference in Differences

* 3.1 Wahl der Gruppen
* 3.2 Händische DiD Berechnung
* 3.3 DiD-Schätzung mittels Regression
* 3.4 Weiterführende Analysen

1. Einordnung des Artikels

## 1. Einführung in den Mindestlohn

**“Jetzt 12€ Mindestlohn wählen.”**

Ein zentrales Wahlkampfthema der SPD zur Bundestagswahl 2021. Es war nicht das erste Mal und wird auch nicht das letzte Mal gewesen sein, dass die Versprechung zur Verbesserung des Lohns ein zentrales Instrument der Arbeitsmarktpolitik ist, um eine große Wählerschaft zu mobilisieren.  
Hinter diesem leicht verständlichen Versprechen steckt allerdings die komplexe Frage, wie die Lohnerhöhungen getragen werden.

Häufig wird in der Forschung die Frage der Auswirkungen auf die Beschäftigung betrachtet. Dabei treten Kontroversen auf und es herrscht keine Einigkeit in den Schlussfolgerungen.  
Daher wählen wir mit unserer Analyse nach Draca et. al. (2011) einen anderen Ansatz.  
Anstatt die Beschäftigung in den Vordergrund zu rücken, widmen wir uns der vorgelagerteten Frage, wie die Lohnsteigerungen innerhalb von Unternehmen kompensiert werden.  
Eine solche Betrachtungsweise ist nach den Autoren in der Forschung bisher weitestgehend unbeachtet geblieben (vgl. Draca et. al (2011), S. 129f.)

Die Basis der Analyse bilden Unternehmensdaten aus dem Vereinigten Königreich (UK).

Quiz: Wie können Unternehmen die Lohnsteigerung kompensieren?

* Verkleinerung der Gewinnmarge [x]
* Weitergabe der Kostensteigerung an Verbraucher [ ]
* Effizienzsteigerung [ ]

**Hintergrund der Einführung**

Wie in Deutschland, wurden und werden Löhne im Vereinigten Königreich teilweise durch Tarifverhandlungen festgelegt.  
Die Notwendigkeit der Einführung eines nationalen Mindestlohns wurde in Großbritannien durch den sinkenden Einfluss von Gewerkschaften gesehen.  
Haben im Jahr 1970 noch 80% der Arbeitnehmer von Tarifverträgen profitiert, so sank dieser Anteil bis 2000 um 50 Prozentpunkte. Mit einem Machtwechsel im britischen Unterhaus, hin zur sozialdemokratischen Labour-Partei, kommt es zur Veränderung in der Arbeitsmarktpolitik. Zum 1. April 1999 wird der nationale Mindestlohn eingeführt.

*(vgl. Brown (2007), S.2f.)*

Quiz: Auf welche Höhe schätzen Sie den gesetzlichen Mindestlohn im Vereinigten Königreich zur Einführung 1999?

* 3.60£ [x]
* 5.70£ [ ]
* 7.80£ [ ]

*Hinweis*: Wir verwenden einheitlich den Punkt als Dezimaltrennzeichen, da sich unsere Daten und die Software ebenfalls auf diese Konvention stützen.

Weiter gilt zu erwähnen, dass der nationale Mindestlohn (NMW) nicht die erste Art des Mindestlohns im Vereinigten Königreich war.  
Von 1909 bis zum Jahr 1993 an wurden gesetzliche Mindestlöhne durch *Wage Councils* innerhalb von Branchen bestimmt.  
Das Novum des NMW liegt in der Reichweite über alle Branchen hinweg. Ähnlich wie die *Wage Councils*, gibt die unabhängige Niedriglohnkommission (LPC) eine Empfehlung zur Mindestlohnhöhe.  
Von 1999 an wird die Lohnhöhe fortlaufend von der Niedriglohnkomission angepasst.

Quiz: Auf welche Höhe schätzen Sie den gesetzlichen Mindestlohn im UK im Jahr 2024?

* 9.11£ [ ]
* 11.22£ [x]
* 14.44£ [ ]

### Info: Aufgliederung des Mindestlohns

Der volle Mindestlohn im Vereinigten Königreich wird erst ab dem Alter von 21 Jahren gezahlt. Die genaue Aufgliederung, sowie die künftig geplanten Anpassungen können sie sich [auf der Internetseite der Regierung](https://www.gov.uk/national-minimum-wage-rates) ansehen.

### Theorie der Unternehmensgewinne

Den theoretischen Ausgangspunkt für die Analyse bietet die Idee der Profitmaximierung nach Ashenfelter & Smith (1979).  
Dabei wird von einem gewinnmaximierenden Unternehmen ausgegangen, dessen Gewinn anhand von Variablen des Lohnsatzes , der Kosten und des Outputpreises gemessen wird.

Aus der abgeleiteten Funktion der Gewinnmaximierung erhalten wir die negative Arbeitsnachfrage .

In unserem Fall wird der Lohnsatz auf die Höhe des Mindestlohns M angepasst. Dadurch entsteht zum Zeitpunkt der Mindestlohneinführung eine Gewinnveränderung von

Unter Gebrauch einer Taylor-Reihe, kann die Gewinnveränderung approximiert dargestellt werden:

Zusätzlich führen wir die Variable *S* ein, die den Umsatz eines Unternehmens definiert. Dadurch können wir die Gewinnmarge darstellen und erhalten

wobei .

Gehen wir nun wie oben bereits erläutert von einer Lohnveränderung aus.

Im Falle eines negativen können wir auf Basis dieser Gleichung von einer Verringerung der Gewinnmarge ausgehen.  
Diese Grundlage wird später für unsere Difference-in-Difference Analyse und der einhergehenden Wahl der Gruppen relevant.  
*(vgl. Draca et al. (2011), S.131f.)*

## 2. Datenüberblick

Quiz: Um in den Datensatz einzutauchen eine kleine Frage zum Einstieg. Was glauben Sie, wieso wir Unternehmensdaten aus dem Vereinigten Königreich beziehen?

* Keine Kosten der Datenbeschaffung im Vereinigten Königreich [ ]
* Strengere Transparenzregeln in der Bilanzierung und eine zentrale Datenspeicherung [x]
* Die Daten sind anonymisiert und lassen eine unternehmensunabhängige Analyse zu [ ]

**Aufgabe**: Lesen Sie den Datensatz main\_fame.dta mit der Funktion read\_dta ein und speichern Sie den Datensatz unter dat. Führen Sie den Code aus, indem Sie check drücken.

dat = read\_dta('main\_fame.dta')

**Aufgabe:** Um einen ersten Überblick über den Datensatz zu bekommen, geben sie die ersten Zeilen des Datensatzes dat mittels der Funktion head aus.

head(dat)

Wir erhalten eine Vielzahl an Variablen, die mehr oder minder relevant für unsere Analysen sind.  
Widmen wir uns den wesentlichsten Variablen für unsere Analyse:

**regno**: Hier handelt es sich um eine bestimmte Nummer, mit der die Daten jedem bestimmten Unternehmen zugeordnet werden können.

**year**: Gibt das Jahr an, aus dem die Daten stammen. Dabei wird das zurückliegende Geschäftsjahr betrachtet.

**ln\_avwage**: Benennt den logarithmierten Durchschnittslohn im Unternehmen.

**net\_pcm**: Die Gewinnmarge eines Unternehmens wir durch die Division vom Nettogewinn durch den Gesamtumsatz im Geschäftsjahr angegeben.

**sic2**: Die Unternehmen werden in Großbritannien nach Branchen in unterschiedliche Branchen unterteilt. Der bis zu vierstellige UKsic (UK Standard Industrial Classification of Economic Activities) beschreibt dabei die Branchen.  
Das oben ausgegebene Unternehmen besitzt den UKsic 7020. Die sic2-Zahl ist somit 70 und beschreibt Immobilienaktivitäten.  
Eine Übersicht über alle Branchen ist [hier](https://webarchive.nationalarchives.gov.uk/ukgwa/20160105160709/http://www.ons.gov.uk/ons/guide-method/classifications/archived-standard-classifications/uk-standard-industrial-classification-1992--sic92-/uk-sic-2003.pdf) zu finden.

**unionmem**: Beschreibt den Anteil an Gewerkschaftsmitgliedern innerhalb des Unternehmens.

**ptwk**: Anteil an Teilzeitarbeitern innerhalb der vierstelligen UKsic-Branche (sic4).

**female**: Frauenanteil bei Arbeitnehmern innerhalb des sic4.

**gorwk**: Gibt den Regierungsbezirk des jeweiligen Unternehmens an. Dabei handelt es sich um Bezirke in England (1-17), Wales(18) und Schottland(19,20).

Werfen wir einen Blick auf die Variable *month*. Wir wollen sehen, wann die Unternehmen ihren Jahresbericht abgeben und damit ihr Geschäftsjahr beenden.

**Aufgabe:** Gruppieren Sie hierfür den Datensatz dat nach dem Berichtsmonat month. Geben Sie anschließend die Anzahl der Einträge mittels der Funktion length() an.

# dat %>%   
# group\_by(\_\_\_) %>%  
# summarize("Einträge" = \_\_\_(month))  
  
dat %>%   
 group\_by(month) %>%  
 summarize(length(month))

## month length.month.  
## 1 3 20914

Quiz: Alle unsere Daten weisen auf ein Ende des Geschäftsjahres im März hin. Können Sie sich erklären, warum?

* Anders als in Deutschland, endet das Geschäftsjahr im UK bei allen Unternehmen im März. [ ]
* Im Frühjahr gibt es im UK wenig Bewegung auf dem Arbeitsmarkt. Dadurch vermeiden wir unbekannte externe Effekte. [ ]
* Die Mindestlohneinführung findet bei diesen Unternehmen nicht mitten im Geschäftsjahr statt. [x]

## 3. DiD – Difference in Differences

Die Auswirkungen des Mindestlohns auf die Unternehmensgewinne können mit Hilfe der Difference-in-Differences-Methode - kurz **DiD** - beobachtet werden.

Die DiD-Schätzung ist häufig geeignet, um kausale Effekte politischer Maßnahmen empirisch zu beurteilen.  
Dabei werden die Daten in zwei zeitliche Perioden und zwei Gruppen eingeteilt.  
*(vgl. Callaway et al. (2021), S. 2)*

Gerade mit Hinblick auf Lohnentwicklungen stellte die Methode bereits in der Vergangenheit eine wichtige Grundlage.  
So können beispielsweise die Einflüsse von Schocks und fixer Effekte abgefedert werden und so die Auswirkungen einzelner Maßnahmen besser beurteilbar machen.  
*(vgl. Angrist et al. (2009), S. 169)*

## 3.1 Wahl der Gruppen

Bei der Difference in Differences Methode werden die Daten anhand von zwei Kriterien unterteilt.  
Zum einen wird entschieden, ob es sich um eine Pre-Treatment oder eine Post-Treatment Gruppe handelt.

### Info: zeitliche Treatmentgruppen

Bei der DiD-Schätzung ist ein genauer Zeitpunkt ermittelbar, zu dem es ein Treatment gibt.  
In unserem Fall ist dieses Treatment die Einführung des nationalen Mindestlohns.  
Alle Daten, die aus einer Zeit vor dem Treatment stammen, sind Teil der Pre-Treatment Gruppe.  
Jene Daten, die seit dem Treatment erfasst wurden, werden der Post-Treatment Gruppe zugeordnet.  
*(vgl. Angrist et al. (2009), S. 169ff.)*

Zur Erinnerung: Der gesetzliche Mindestlohn wurde am 01. April 1999 eingeführt.

Quiz: Ab welchem Zeitpunkt werden Daten der Post-Treatment Gruppe zugewiesen?

* 1999 [ ]
* 2000 [x]

Für eine DiD-Schätzung muss weiter eine Kontroll- und eine Treatmentgruppe definiert werden.  
In der Treatmentgruppe sollen jene Unternehmen zusammengefasst werden, bei denen der Durchschnittslohn (avwage) im Jahr unter 12000 Pfund lag.  
Die Kontrollgruppe umfasst Unternehmensdaten mit einem Durchschnittslohn zwischen 12000 und 20000 Pfund.  
Eine genauere Bewertung zur Wahl dieser Gruppen erfolgt zu einem späteren Zeitpunkt.

### Parallele Trends

Als relevante Voraussetzung für das Anwenden der DiD-Methode wird die Parallel-Trends-Assumption (PTA) gehandelt.  
Diese besagt, dass ohne die Einführung eines Treatments, die Linien der Kontroll- und Treatmentgruppe weitestgehend parallel weiterliefen.  
Annähernd kann das überprüft werden, indem man sich die Trendlinien der beiden Gruppen vor Einführung des Treatments ansieht.  
*(vgl. Angrist et al. (2009), S. 171f.)*

**Aufgabe**: Lesen Sie hierfür zunächst den Datensatz main\_fame ein und speichern ihn unter dat.

# \_\_\_ = read\_dta('main\_fame.dta')  
  
dat = read\_dta('main\_fame.dta')

Um die Daten jeden Jahres zusammenzufassen, kann die Funktion group\_by genutzt werden.

### Info: group\_by

Die group\_by Funktion aus dem dplyr Paket fasst die Daten anhand einer zu bestimmenden Variablen zusammen und ordnet die Ergebnisse aufsteigend an.  
Um die Gruppierung in einem Output sichtbar zu machen, kann die Pipe um den summarise Befehl erweitert werden, in dem dann die Werte jeder Einheit zusammengefasst werden können.  
(<https://dplyr.tidyverse.org/reference/group_by.html>)

**Aufgabe**: Fügen Sie die passende Variable in group\_by ein und geben Sie für jedes Jahr den durchschnittlichen logarithmierten Durchschnittslohn aus.

# dat\_trend\_treat = dat %>%   
# filter(ctreat1 == 1) %>%  
# group\_by(\_\_\_)%>%  
# summarise("treatment"=\_\_\_(ln\_avwage, na.rm = TRUE))   
  
dat\_trend\_treat = dat %>%   
 filter(ctreat1 == 1) %>%  
 group\_by(year)%>%  
 summarise("treatment"=mean(ln\_avwage, na.rm = TRUE))

Analog dazu erstellen wir auch noch einen Datensatz der Kontrollgruppe.

**Aufgabe**: Führen Sie dazu den Code mittels *check* aus.

dat\_trend\_control = dat %>%  
 filter(ctreat1 == 0) %>%  
 group\_by(year) %>%  
 summarise("control"=mean(ln\_avwage, na.rm = TRUE))

Um die beiden erstellten Datensätze miteinander zu verbinden, nutzen wir den left\_join.

**Aufgabe**: Führen Sie den Code mittels *check* aus und sehen sich den Datensatz an.

dat\_trend = left\_join(dat\_trend\_treat, dat\_trend\_control)

## Joining with `by = join\_by(year)`

dat\_trend

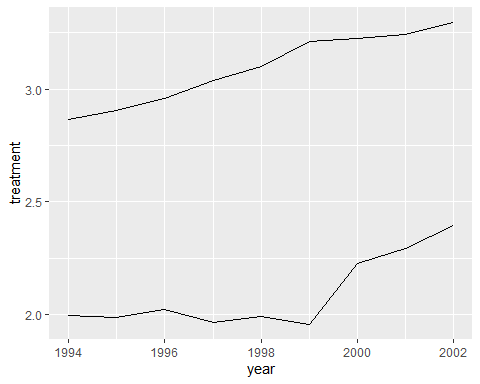
## year treatment control  
## 1 1994 1.99712 2.86534  
## 2 1995 1.98982 2.90611  
## 3 1996 2.02437 2.95668  
## 4 1997 1.96379 3.04003  
## 5 1998 1.99206 3.10183  
## 6 1999 1.95751 3.20962  
## 7 2000 2.22806 3.22456  
## 8 2001 2.29497 3.24137  
## 9 2002 2.39626 3.29615

Jetzt können wir die Daten **visualisieren**, indem wir uns in einem Liniendiagramm die Trends beider Gruppen ausgeben lassen.

Die Basis dafür bildet der oben erstellte Datensatz dat\_trend.

**Aufgabe**: Fügen Sie die Trendlinien der beiden Gruppen hinzu, indem Sie den Code um die fehlende Variable auf der Y-Achse ergänzen.

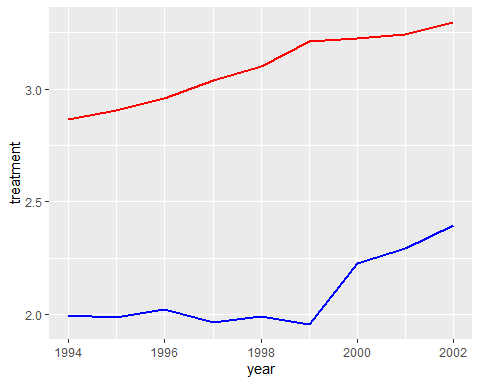
# ggplot(dat\_trend)+  
# geom\_line(aes(x = year, y = treatment))+  
# geom\_line(aes(x = year, y = \_\_\_))  
  
ggplot(dat\_trend)+  
 geom\_line(aes(x = year, y = treatment))+  
 geom\_line(aes(x = year, y = control))



Wir können schon etwas ablesen, die Graphik kann allerdings noch ansehnlicher gestaltet werden um Unterschiede deutlicher zu machen.  
Dafür können wir die Linien kolorieren und sowohl die Lininenstärke, als auch den Linientyp verändern.

**Aufgabe**: Färben Sie die Linie der Treatmentgruppe in blau und die der Kontrollgruppe in rot.  
*Hinweis*: Der englische Farbenname muss in Anführungszeichen angegeben werden.

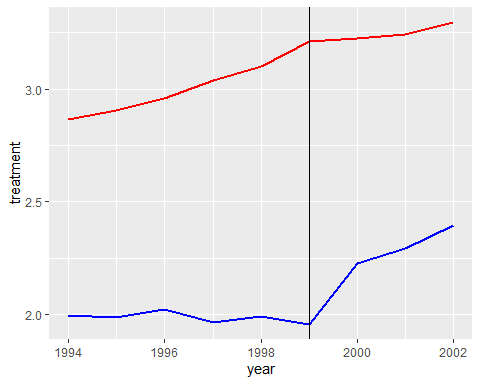
# ggplot(dat\_trend)+  
# geom\_line(aes(x = year, y = treatment), colour= \_\_\_, linetype = 1, linewidth = 0.8)+  
# geom\_line(aes(x = year, y = control), colour = \_\_\_, linetype = 1, linewidth = 0.8)  
  
ggplot(dat\_trend)+  
 geom\_line(aes(x = year, y = treatment), colour= "blue", linetype = 1, linewidth = 0.8)+  
 geom\_line(aes(x = year, y = control), colour = "red", linetype = 1, linewidth = 0.8)



Um klarzumachen, in welcher Periode wir uns befinden, kann der Zeitpunkt der Einführung des nationalen Mindestlohns markiert werden.

**Aufgabe**: Fügen Sie hierfür das letzte Jahr ein, in dem der Mindestlohn nicht galt.  
*Erinnerung:* Das Jahr gibt das Ende des letzten Geschäftsjahres an.

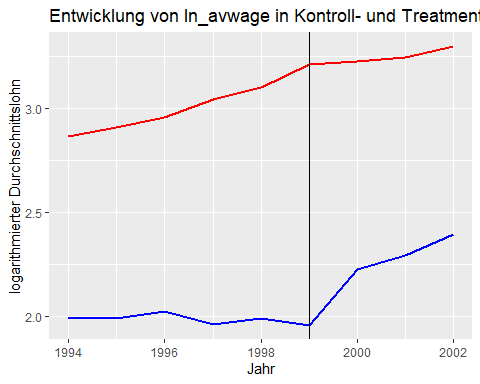
# ggplot(dat\_trend)+  
# geom\_line(aes(x = year, y = treatment), colour= "blue", linetype = 1, linewidth = 0.8)+  
# geom\_line(aes(x = year, y = control), colour = "red", linetype = 1, linewidth = 0.8)+  
# geom\_vline(xintercept = \_\_\_)  
  
ggplot(dat\_trend)+  
 geom\_line(aes(x = year, y = treatment), colour= "blue", linetype = 1, linewidth = 0.8)+  
 geom\_line(aes(x = year, y = control), colour = "red", linetype = 1, linewidth = 0.8)+  
 geom\_vline(xintercept = 1999)



Zum Abschluss geben wir dem Graphen einen Titel, sowie eine passende Achsenbeschriftung.

**Aufgabe**: Führen Sie den Code mittels *check* aus.

ggplot(dat\_trend)+  
 geom\_line(aes(x = year, y = treatment), colour= "blue", linetype = 1, linewidth = 0.8)+  
 geom\_line(aes(x = year, y = control), colour = "red", linetype = 1, linewidth = 0.8)+  
 geom\_vline(xintercept = 1999)+  
 xlab("Jahr")+  
 ylab("logarithmierter Durchschnittslohn")+  
 ggtitle("Entwicklung von ln\_avwage in Kontroll- und Treatmentgruppe")



### Award: Prince(ss) of Arts

Sie haben erfolgreich ein Liniendiagramm mit *ggplot2* entwickelt und bewiesen, dass Sie mit Graphiken umgehen können

Quiz: Erkennen Sie in der Graphik einen parallelen Trend?

* Ja [ ]
* Nein [x]

In unserer Betrachtungsweise ist kein eindeutig paralleler Trend zu erkennen. Ist deshalb die DiD-Schätzung eine falsche Methode, um die Effekte der Mindestlohneinführung zu betrachten?

Aufgrund des kurzen Betrachtungszeitraums des Trends sollte diese Einschätzung auch mit Vorsicht genossen werden: Es wird sich auf Trends bezogen, die sich auf sechs Datenpunkte (1994-1999) stützen.  
Die Betrachtung eines größeren Zeitraums könnte hier eine genauere Betrachtung zulassen.  
Bezüglich der PTA gibt es seitens der Autoren keine weiterführenden Analysen, eine Erwähnung wird innerhalb des Artikels vergeblich gesucht.

### Treatmentwahl

Nachdem wir um die Komplexität der allgemeinen Voraussetzungen für eine DiD-Schätzung wissen, können wir uns weiter der Relevanz der Gruppenwahl widmen.  
In der erstellten Graphik haben wir die Treatment- und Kontrollgruppen zunächst als gegeben angenommen, ohne die Entstehung zu hinterfragen.  
Der Frage der Entstehung gehen wir nun nach.

Als Einstieg sehen wir uns dafür die Veränderung der Löhne über die Jahre an.  
Um die Veränderung der Löhne von Geschäftsjahr zu Geschäftsjahr herauszufinden, muss der Datensatz nach Jahren aufgeteilt werden.  
Das gelingt, indem für jedes Jahr ein eigener Datensatz erstellt wird. Dafür muss der bereits eingelesene Datensatz dat\_main zunächst gefiltert werden.

**Aufgabe**: Wählen Sie die passende Variable für das leere Feld und überprüfen Sie Ihre Eingabe durch *check*.

# pcw95 = dat %>%  
# filter(\_\_\_ == 1995) %>%  
# filter(avwage >= 3)  
  
#  
pcw95 = dat %>%  
 filter(year == 1995)%>%  
 filter(avwage >= 3)

Nun befinden sich in pcw95 alle Daten aus dat\_main, die aus dem Jahr 1995 stammen.  
Neben der Unterscheidung in den Jahren sollen zudem die Unterschiede der Ausgangslöhne herausgestellt werden.  
Dazu werden die Löhne in Perzentile aufgeteilt.

**Aufgabe**: Erstellen Sie einen Vektor percent95, der die logarithmierten Durchschnittslöhne in 100 Perzentile aufteilt. Nutzen Sie dafür die Funktion quantile(dat,seq()). Setzen Sie den fehlenden Wert in seq() ein, um einen Vektor der Länge 100 zu erhalten.

# percent95 = quantile(pcw95$ln\_avwage, seq(0, 1, \_\_\_))  
  
percent95 = quantile(pcw95$ln\_avwage, seq(0, 1, 0.01))

Diese Aufteilung muss nun für jedes Jahr des Datensatzes wiederholt werden.

**Aufgabe**: Führen sie dafür den Chunk durch *check* aus.

pcw96 = dat %>%  
 filter(year == 1996)%>%  
 filter(avwage >= 3)  
percent96 = quantile(pcw96$ln\_avwage, seq(0, 1, 0.01))  
  
pcw97 = dat %>%  
 filter(year == 1997)%>%  
 filter(avwage >= 3)  
percent97 = quantile(pcw97$ln\_avwage, seq(0, 1, 0.01))  
  
pcw98 = dat %>%  
 filter(year == 1998)%>%  
 filter(avwage >= 3)  
percent98 = quantile(pcw98$ln\_avwage, seq(0, 1, 0.01))  
  
pcw99 = dat %>%  
 filter(year == 1999)%>%  
 filter(avwage >= 3)  
percent99 = quantile(pcw99$ln\_avwage, seq(0, 1, 0.01))  
  
pcw00 = dat %>%  
 filter(year == 2000)%>%  
 filter(avwage >= 3)  
percent00 = quantile(pcw00$ln\_avwage, seq(0, 1, 0.01))  
  
pcw01 = dat %>%  
 filter(year == 2001)%>%  
 filter(avwage >= 3)  
percent01 = quantile(pcw01$ln\_avwage, seq(0, 1, 0.01))  
  
pcw02 = dat %>%  
 filter(year == 2002)%>%  
 filter(avwage >= 3)  
percent02 = quantile(pcw02$ln\_avwage, seq(0, 1, 0.01))

Nun können die Unterschiede in jedem einzelnen Perzentil von Jahr zu Jahr ermittelt werden. Die Differenzen sollen in einem Datensatz gespeichert werden.  
Dafür wird zunächst ein Datensatz erstellt, der sich nach den Werten der Perzentile aus dem ersten Beobachtungsjahr 1995 richtet.

**Aufgabe**: Erstellen Sie einen Datensatz diffperc mittels der Funktion data.frame. Dieser soll zunächst die Werte aus percent95 enthalten, welche durch sort() geordnet sind.

diffperc = data.frame(sort(percent95))

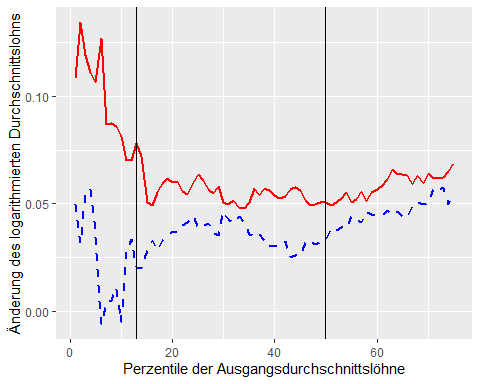
Der Datensatz diffperc kann jetzt um die Differenzen innerhalb der Perzentile erweitert werden.

**Aufgabe**: Führen Sie den Code mit *check* aus.

diffperc = diffperc %>%  
 mutate("percentile" = c(0:100),  
 "diff96" = percent96-percent95,  
 "diff97" = percent97-percent96,  
 "diff98" = percent98-percent97,  
 "diff99" = percent99-percent98,  
 "diff00" = percent00-percent99,  
 "diff01" = percent01-percent00,  
 "diff02" = percent02-percent01)

Für die Betrachtungsweise der Auswirkungen durch die Einführung des Mindestlohns sind zwei Differenzen von besonderer Bedeutung.  
Daher sollen die Änderung des Lohns im Jahr **vor** der Einführung (diff99) mit der Änderung im Jahr **nach** der Einführung (diff00) in einem Liniendiagramm gegenübergestellt werden.

diffperc = filter(diffperc, percentile < 76 & percentile > 0)  
ggplot(diffperc)+  
 geom\_line(aes(x = percentile, y = diff99), colour = "blue", linetype = 2, linewidth = 0.8)+  
 geom\_line(aes(x = percentile, y = diff00), colour = "red", linewidth = 0.8)+  
 geom\_vline(xintercept = 13)+  
 geom\_vline(xintercept = 50)+  
 xlab("Perzentile der Ausgangsdurchschnittslöhne")+  
 ylab("Änderung des logarithmierten Durchschnittslohns")



Es wird deutlich, dass vorallem in den unteren 13 Perzentilen (markiert durch die erste vertikale Linie) von der Mindestlohneinführung profitiert wird und die Unterschiede in der Lohnerhöhung am Größten sind.

Die graphische Veranschaulichung möchten wir auch technisch nachvollziehen.  
Erinnern wir uns hierfür zunächst an den theoretischen Zusammenhang von Gewinnmargen und Lohnerhöhungen:

Quiz: Wann kann ein negativer Einfluss auf die Gewinnmarge entstehen?

* Wenn der Mindestlohn höher als der bisherige Lohn ist, also [x]
* Wenn der Mindestlohn niedrier als der bisherige Lohn ist, also [ ]

Im Mittelpunkt der Mindestlohneinführung steht der Wert von 3.60 Pfund pro Stunde.  
Wir sollten also in unsere Treatmentgruppe jene Mitarbeiter einschließen, die vor der Mindestlohneinführung einen Lohn unter 3.60£ erhielten.  
Da unser Datensatz allerdings keine Daten zu jedem Arbeitnehmer enthält, müssen wir uns der Wahl der Treatmentgruppe nähern.  
Die Variable, die den Lohn innerhalb eines Unternehmens am Besten widerspiegelt ist hierbei avwage, worüber sich auch im Artikel genähert wird.

Dabei werden alle Unternehmen mit einem durchschnittlichen Lohn kleiner gleich **12,000£** in die Treatmentgruppe einsortiert.  
Um zu verstehen, warum sich für diesen Wert als Obergrenze entschieden wird ist es hilfreich, den Wert in Zusammenhang mit dem **Stundenlohn** zu bringen.

Dazu soll zunächst der durchschnittliche Lohn aller Unternehmen aus der Treatmentgruppe in der Pre-Policy Periode im Jahr **1999** ermittelt werden.

**Aufgabe**: Setzen Sie die passenden Werte in die Lücken ein.   
*Hinweis*: avwage wird im Datensatz in 1000 Pfund angegeben.

# mean\_treat = dat %>%  
# filter(avwage <= \_\_\_) %>%  
# filter(year == \_\_\_) %>%  
# summarise(round(mean(avwage)\*1000,2))  
# mean\_treat  
  
mean\_treat = dat %>%  
 filter(avwage <= 12) %>%  
 filter(year == 1999) %>%  
 summarise(round(mean(avwage)\*1000,2))  
mean\_treat

## round(mean(avwage)\*1000, 2)  
## **7998.04**

Dieser Wert gibt den durchschnittlichen Jahreslohn der Treatmentgruppe wieder.  
Um nun auf den Stundenlohn zu kommen, müssen wir ermitteln, in welchem Bereich sich die Jahresarbeitszeit im Geschäftsjahr mit dem Ende in 1999 befindet.

Daten aus dem [OECD Employment Outlook 2002](https://doi.org/10.1787/empl_outlook-2002-en) zeigen, dass die durchschnittliche Jahresarbeitszeit 1998 bei **1731 Stunden** und 1999 bei **1719 Stunden** liegt.  
Dabei werden alle Beschäftigten gleichermaßen berücksichtigt, unabhängig von der Art der Anstellung.

Daraus können wir folgende Jahresarbeitszeit abschätzen, wobei wir approximiert die Daten aus den Jahren 1998 und 1999 verwenden:

**Aufgabe**: Summieren Sie anteilsmäßig 3/4 der durchschnittlichen Arbeitszeit aus dem Jahr **1998** und 1/4 aus dem Jahr **1999** und speichern den Wert in der entsprechenden Variablen.

# annual\_hours = 0.75\*\_\_\_ + 0.25\*\_\_\_  
  
annual\_hours = 0.75\*1731 + 0.25\*1719

Nun haben wir den genäherten Wert, wollen ihn allerdings auf eine Vollzeitbeschäftigung anwenden.

**Aufgabe:** Ermitteln Sie den Anteil an Teilzeitarbeitern (ptwk) in unserem Datensatz dat.  
(*Hinweis*: Da es immer sein kann, dass Daten nicht vollständig sind, fügen wir na.rm = TRUE hinzu, um fehlende Einträge bei der Berechnung zu ignorieren.)

# ptwk\_share = mean(\_\_\_, na.rm = TRUE)  
# ptwk\_share  
  
ptwk\_share = mean(dat$ptwk, na.rm = TRUE)  
ptwk\_share

## [1] 0.1643418

Der Anteil an Teilzeitarbeitern beträgt in unserem Datensatz ca. 16.4%.  
Um eine imaginäre Vollzeitbeschäftigung darzustellen, verdoppeln wir jene Jahresarbeitszeiten der Teilzeitarbeiter und fügen diese der bereits errechneten durchschnittlichen Jahresarbeitszeit hinzu.

**Aufgabe:** Lassen Sie sich die approximierte durchschnittliche Jahresarbeitszeit mittels *check* ausgeben.

annual\_hours = annual\_hours + ptwk\_share \* annual\_hours  
  
annual\_hours

## [1] 2011.983

Nun können wir den durchschnittlichen Stundenlohn innerhalb der Treatmentgruppe darstellen.

**Aufgabe:** Berechnen Sie den durchschnittlichen Stundenlohn der Treatmentgruppe.

# treat\_wage\_hour = mean\_treat / \_\_\_  
# treat\_wage\_hour  
  
treat\_wage\_hour = mean\_treat / annual\_hours  
treat\_wage\_hour

## round(mean(avwage)\*1000, 2)  
## **3.9752**

Mit dieser Näherung kommt man in die Nähe des Mindestlohns. Dennoch ist dies auch mit Vorsicht zu genießen.  
Da die Wahl der Treatmentgruppe auf Durchschnittslöhne eines gesamten Unternehmens beruht, können große Ausreißer das Ergebnis beeinflussen.  
Unsere Näherung beruht auf mehreren Annahmen, die nicht vollumfänglich belegt werden können. Nicht alle vom Mindestlohn profitierenden Arbeitnehmer werden in der Treatmentgruppe abgebildet.  
Dennoch kommen die Autoren zu dem Schluss, dass mit 87% jener Arbeitnehmer Teil der Treatmentgruppe sind.  
*(vgl. Draca et al., S. 137f.)*

Nun haben wir im ersten Schritt die Gruppen und Perioden definiert. Als nächstes werden wir die Daten aktiv den jeweiligen Gruppen zuordnen und zusammenfassen.

## 3.2 händische DiD-Berechnung

Mit einem Blick auf den Datensatz erkennt man, dass für die unterschiedlichen Gruppen bereits Dummy-Variablen erstellt worden sind.

**Aufgabe**: Laden Sie erneut den bereits bekannten Datensatz mittels *check*.

dat = read\_dta('main\_fame.dta')

Erstellen Sie auf Basis des Datensatzes dat eine Tabelle, die alle relevanten Daten für eine händische Berechnung enthält.

**Aufgabe:** Führen Sie den folgenden Code aus, indem Sie den Datensatz nach den für die DiD-Schätzung relevanten Variablen gruppieren.

# dat\_table\_DiD = dat %>%  
# \_\_\_(ctreat1, NMW)%>%  
# filter(pp == 1) %>%  
# summarise(mean\_avwage = mean(avwage), mean\_ln\_avwage = mean(ln\_avwage), mean\_net\_pcm = mean(net\_pcm))  
  
dat\_table\_DiD = dat %>%  
 group\_by(ctreat1, NMW)%>%  
 filter(pp == 1) %>%  
 summarise(mean\_avwage = mean(avwage), mean\_ln\_avwage = mean(ln\_avwage), mean\_net\_pcm = mean(net\_pcm))

Quiz: Wieso wird der Mittelwert verwendet und nicht nur die beiden Jahre vor und nach der Mindestlohneinführung?

* Durch die Betrachtung können kurzfristige Schwankungen und zufällige Effekte abgefedert werden. [x]
* Um einen deutlicheren Unterschied zwischen den beiden Perioden zu erhalten. [ ]

Wenn Sie sich den obigen Code nochmals ansehen, können Sie eine noch nicht eingeführte Variable erkennen.  
Der Datensatz wird hier anhand der Variablen “pp” gefiltert.

Quiz: Wieso und nach welchem Kriterium könnte hier gefiltert werden?

* Es gibt starke Unterschiede in der Berichterstattung zwischen den Ländern im UK. Daher fokussieren wir uns auf England. [ ]
* Aufgrund einer höheren Teilzeitquote bei Frauen, filtern wir hier nach Geschlecht und legen den Fokus auf die Männer. [ ]
* Da kleinere Firmen von manchen Teilen der Berichterstattung befreit sind, sind ihre Daten unvollständig. Daher werden sie nicht weiter berücksichtigt. [x]

Im Vereinigten Königreich sind kleinere Unternehmen von der Veröffentlichung bestimmter Daten befreit.  
Um eine vollumfängliche Analyse zu gewährleisten, entscheiden sich die Autoren daher ein Unterbeispiel zu erstellen, in denen Daten vollständig erfasst werden *(vgl. Draca et al., S. 135, Verw. 13)*.

Das schränkt allerdings auch die Bewertung ein, lenkt den Blick weg von kleineren Unternehmen und muss bei der Betrachtung der Ergebnisse bedacht werden.  
Wie die Autoren die Auswahl des Unterbeispiels treffen wird nicht leider vollumfänglich bekannt und auch eigene Überlegungen lassen nur Mutmaßungen zu (vgl. Anhang 1).

Nachdem die relevanten Daten des DiD zusammengefasst in der Tabelle dat\_table\_DiD vorhanden sind, können die Differenzen nun händisch berechnet werden.

Die Formel hierfür lautet

*(vgl. Draca et al., S. 133)*

Der erste Teil beschreibt die Veränderung in der Treatmentgruppe durch das Treatment der Mindestlohneinführung. Um andere Einflussfaktoren wie saisonale Effekte herauszurechnen, wird im zweiten Teil die Veränderung in der Kontrollgruppe berechnet, welche im Idealfall nicht vom Treatment betroffen ist.

Werfen Sie zunächst einen Blick auf die erzeugte Tabelle und überlegen Sie, welcher Gruppe die jeweiligen Werte angehören.

**Aufgabe:** Lassen Sie sich die oben erstellte Tabelle dat\_table\_DiD anzeigen und überprüfen Sie Ihre Eingabe mittels *check*.

dat\_table\_DiD

## ctreat1 NMW mean\_avwage mean\_ln\_avwage mean\_net\_pcm  
## 1 0 0 16.30307 2.77520 0.06990  
## 2 0 1 18.45086 2.89313 0.05803  
## 3 1 0 9.02691 2.14930 0.12793  
## 4 1 1 11.99248 2.37813 0.08866

Quiz: Welche Spalte beschreibt die Kontroll- bzw. Treatmentgruppe?

* ctreat1 [x]
* NMW [ ]

Quiz: Von welcher Gruppe ist die Rede, wenn NMW = 1 ?

* Pre-Treatment [ ]
* Post-Treatment [x]

**Aufgabe**: Weisen Sie den für die DiD Schätzung relevanten Variablen die passenden Tabelleneinträge aus dat\_table\_DiD zu.

# y1.post.treat = dat\_table\_DiD$mean\_ln\_avwage[\_\_\_]  
# y1.post.control = dat\_table\_DiD$mean\_ln\_avwage[\_\_\_]  
# y1.pre.treat = dat\_table\_DiD$mean\_ln\_avwage[\_\_\_]  
# y1.pre.control = dat\_table\_DiD$mean\_ln\_avwage[\_\_\_]  
#   
  
y1.post.treat = dat\_table\_DiD$mean\_ln\_avwage[4]  
y1.post.control = dat\_table\_DiD$mean\_ln\_avwage[2]  
y1.pre.treat = dat\_table\_DiD$mean\_ln\_avwage[3]  
y1.pre.control = dat\_table\_DiD$mean\_ln\_avwage[1]

Neben des logarithmierten Durchschnittslohns wird auch der DiD-Schätzer der Nettorendite eines Unternehmens je Geschäftsjahr (net\_pcm) berechnet.  
Das geschieht analog zur obigen Schätzung.

**Aufgabe**: Werfen Sie erneut einen Blick auf die erstellte Tabelle und geben Sie nun die Zeilen ein, in denen der passende Wert steht.

# y2.post.treat = dat\_table\_DiD$mean\_net\_pcm[\_\_\_]  
# y2.post.control = dat\_table\_DiD$mean\_net\_pcm[\_\_\_]  
# y2.pre.treat = dat\_table\_DiD$mean\_net\_pcm[\_\_\_]  
# y2.pre.control = dat\_table\_DiD$mean\_net\_pcm[\_\_\_]  
  
y2.post.treat = dat\_table\_DiD$mean\_net\_pcm[4]  
y2.post.control = dat\_table\_DiD$mean\_net\_pcm[2]  
y2.pre.treat = dat\_table\_DiD$mean\_net\_pcm[3]  
y2.pre.control = dat\_table\_DiD$mean\_net\_pcm[1]

Die Werte sind nun gespeichert und die DiD-Schätzer können berechnet werden.

**Aufgabe**: Fügen Sie dem Code das Anzeigen der Ergebnisse hinzu und führen Sie den Code mittels *check* aus.

# DiD\_ln\_avwage = (y1.post.treat - y1.pre.treat) - (y1.post.control - y1.pre.control)  
# DiD\_net\_pcm = (y2.post.treat - y2.pre.treat) - (y2.post.control - y2.pre.control)  
#   
# #Ergebnisse anzeigen  
#   
  
DiD\_ln\_avwage = (y1.post.treat - y1.post.control) - (y1.pre.treat - y1.pre.control)  
DiD\_net\_pcm = (y2.post.treat - y2.post.control) - (y2.pre.treat - y2.pre.control)  
   
#Ergebnisse anzeigen  
DiD\_ln\_avwage

## [1] 0.1109065

DiD\_net\_pcm

## [1] -0.02739991

### Award: DiD Starter-Set

Sie haben erfolgreich eine erste Difference-in-Differences Schätzung durchgeführt und können die unterschiedlichen Gruppen in richtiger Relation zueinander betrachten.

Durch die Einführung des Mindestlohns können wir sagen, dass dadurch in den beobachteten Unternehmen der Durchschnittslohn gestiegen ist.  
Zugleich ist zu erkennen, dass in jenen Unternehmen der Nettogewinn zurück gegangen ist.

Um diese Veränderung genauer bewerten zu können und tiefergehende Analysen zu ermöglichen, machen wir uns die Methode der Regression zu eigen.

## 3.3 DiD-Schätzung mittels Regression

**DAG**

Zu Beginn widmen wir uns der Veränderung im Durchschnittslohn durch die Einführung des Mindestlohns. Um uns einen ersten Überblick über die Zusammenhänge der Variablen in unserer DiD-Schätzung zu verschaffen, sind die Directed Acyclic Graphs (DAGs) von Vorteil.

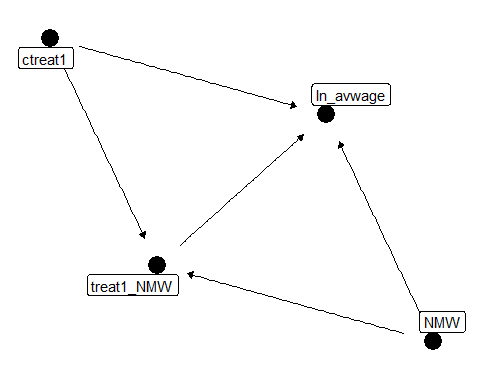
### Info: DAG

In einem DAG werden unterschiedliche Variablen durch Pfade in Form von Pfeildarstellungen in Relation zueinander gesetzt.  
Eine solchen Darstellung kann die Aufstellung von Regressionen veraunschaulichen und nachvollziehbar machen.  
Interessant dabei sind die sogenannten Confounder, die sowohl die abhängige, als auch eine beschreibende Variable beeinflussen.  
(vgl. International Journal of Epidemiology, 2021, Vol. 50, No. 2, S. 622) (<https://academic.oup.com/ije/article/50/2/620/6012812>)

Erklärungen, wie ein DAG in R zustande kommt:  
<https://cran.r-project.org/web/packages/ggdag/vignettes/intro-to-dags.html>  
<https://r-causal.github.io/ggdag/>

**Aufgabe**: Führen Sie hierfür den Code mittels *check* aus und sehen Sie sich den Graphen an.

dag <- dagify(  
 ln\_avwage ~ ctreat1 + treat1\_NMW + NMW,   
 treat1\_NMW ~ ctreat1 + NMW,  
 labels = c(ctreat1 = "ctreat1", treat1\_NMW = "treat1\_NMW", NMW = "NMW", ln\_avwage = "ln\_avwage")  
)  
  
ggdag(dag, node\_size = 6, text = FALSE, edge\_type = "link", use\_labels = "label")+  
 theme\_dag\_blank()



Hier sieht man wie die Variablen, mit denen wir bisher gearbeitet haben, zusammenhängen.

Quiz: Wählen Sie den/die Confounder aus unserem DAG aus?

* ln\_avwage [ ]
* treat1\_NMW & ln\_avwage [ ]
* NMW & ctreat1 [x]
* NMW & treat1\_NMW & ctreat1 [ ]

Wir können aus dem DAG eine Regressionsgleichung der Form  
 ableiten.

Dabei kann jedem Schätzer eine beschreibende Eigenschaft zugewiesen werden.

| Schätzer | Erklärende Variable | Eigenschaft | Confounder |
| --- | --- | --- | --- |
|  |  | Generelle Unterschiede vor dem Treatment | Ja |
|  |  | DiD-Schätzer | Nein |
|  |  | Saisonale Effekte | Ja |

**Aufgabe:** Aus dieser Darstellung soll nun eine quantitative Analyse ermöglicht werden. Lesen Sie dafür einen Auszug des bekannten Datensatzes main\_fame ein, indem Sie den Code mit *check* ausführen.

dat\_raw = read\_dta('main\_fame.dta')  
dat = filter(dat\_raw, pp == 1)

Zu Beginn unserer Analyse wollen wir mehrere Regressionen auf Basis der oben im DAG dargestellten Variablen laufen lassen.  
Dazu sollen die einzelnen Variablen, die ln\_avwage beeinflussen, beobachtet werden.

### Info: stargazer

Die Stargazer Funktion ermöglicht uns einen schön formatierten Output mit der Möglichkeit, mehrere Regressionen parallel darzustellen.

(<https://www.rdocumentation.org/packages/stargazer/versions/5.2.3/topics/stargazer>)

**Aufgabe**: Führen Sie die Regression mit unserem DiD-Schätzer aus und stellen Sie das Ergebnis mittels der stargazer Funktion dar.

# regi = lm(ln\_avwage ~ treat1\_NMW, data = dat)  
#   
# stargazer(\_\_\_, type = "text")  
  
regi = lm(ln\_avwage ~ treat1\_NMW, data = dat)  
  
stargazer(regi, type = "text")

##   
## ===============================================  
## Dependent variable:   
## ---------------------------  
## ln\_avwage   
## -----------------------------------------------  
## treat1\_NMW -0.365\*\*\*   
## (0.016)   
##   
## Constant 2.743\*\*\*   
## (0.006)   
##   
## -----------------------------------------------  
## Observations 4,112   
## R2 0.110   
## Adjusted R2 0.109   
## Residual Std. Error 0.338 (df = 4110)   
## F Statistic 505.677\*\*\* (df = 1; 4110)   
## ===============================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Quiz: Wieso ist der Einfluss der Treatmentvariablen im Gegensatz zur händischen Berechnung negativ?

* Die Regression ist genauer als die händische Berechnung. [ ]
* Die Regression ist unvollständig. [x]

Wir müssen in unserer Regression also sowohl die Interaktionsvariable als auch die dazugehörigen Confounder berücksichtigen.

Daraus folgt dann folgende Regressionsgleichung:

**Aufgabe:** Berechnen Sie auf Basis dieser Gleichung die Einflüsse der Variablen und das damit zusammenhängende DiD-Ergebnis. Stellen Sie das Ergebnis wieder in stargazer dar.

# reg1 = lm(ln\_avwage ~ \_\_\_ + \_\_\_ + \_\_\_, data = dat)  
# stargazer(reg1, type = "text")  
  
reg1 = lm(ln\_avwage ~ ctreat1 + treat1\_NMW + NMW, data = dat)  
stargazer(reg1, type = "text")

##   
## ===============================================  
## Dependent variable:   
## ---------------------------  
## ln\_avwage   
## -----------------------------------------------  
## ctreat1 -0.626\*\*\*   
## (0.013)   
##   
## treat1\_NMW 0.111\*\*\*   
## (0.019)   
##   
## NMW 0.118\*\*\*   
## (0.009)   
##   
## Constant 2.775\*\*\*   
## (0.006)   
##   
## -----------------------------------------------  
## Observations 4,112   
## R2 0.501   
## Adjusted R2 0.500   
## Residual Std. Error 0.253 (df = 4108)   
## F Statistic 1,373.397\*\*\* (df = 3; 4108)  
## ===============================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Der Wert der Interaktionsvariablen hat sich im Vergleich zur unvollständigen Regression stark verändert und erinnert an die händische Berechnung.  
Um diese starke Veränderung besser zu verstehen, können wir die gemeinsame Betrachtung der erklärenden Variablen mit der einzelnen Betrachtung vergleichen.

**Aufgabe**: Führen Sie dazu den Code aus und sehen sich die Ausgabe an.

regii = lm(ln\_avwage ~ ctreat1, data = dat)  
  
regiii = lm(ln\_avwage ~ NMW, data = dat)  
  
stargazer(regi, regii, regiii, reg1, type = "text")

===============================================================================================================================  
## Dependent variable:   
## -----------------------------------------------------------------------------------------------------------  
## ln\_avwage   
## (1) (2) (3) (4)   
## -------------------------------------------------------------------------------------------------------------------------------  
## treat1\_NMW -0.365\*\*\* 0.111\*\*\*   
## (0.016) (0.019)   
##   
## ctreat1 -0.569\*\*\* -0.626\*\*\*   
## (0.010) (0.013)   
##   
## NMW 0.143\*\*\* 0.118\*\*\*   
## (0.011) (0.009)   
##   
## Constant 2.743\*\*\* 2.834\*\*\* 2.628\*\*\* 2.775\*\*\*   
## (0.006) (0.005) (0.008) (0.006)   
##   
## -------------------------------------------------------------------------------------------------------------------------------  
## Observations 4,112 4,112 4,112 4,112   
## R2 0.110 0.456 0.040 0.501   
## Adjusted R2 0.109 0.456 0.039 0.500   
## Residual Std. Error 0.338 (df = 4110) 0.264 (df = 4110) 0.351 (df = 4110) 0.253 (df = 4108)   
## F Statistic 505.677\*\*\* (df = 1; 4110) 3,444.700\*\*\* (df = 1; 4110) 169.547\*\*\* (df = 1; 4110) 1,373.397\*\*\* (df = 3; 4108)  
## ===============================================================================================================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Im Gegensatz zur händischen Berechnung können wir durch die Regression die Signifikanz unserer Schätzung beurteilen.

Quiz: Wie beurteilen Sie die Signifikanz unseres DiD-Schätzers in der dargestellten Regression?

* hoch [x]
* niedrig [ ]

Die Werte der Variablen sind im obigen Fall also hoch signifikant gekennzeichnet.  
Allerdings können die p-Werte fehlerhaft sein, da es sich bei den Daten teilweise um **genestete Daten** handelt.

Genestete Daten liegen vor, wenn mehrere Daten zu einer übergeordneten Einheit zählen.  
*(vgl. Aarts et al. (2014), S. 491)*

In unserem Fall haben wir Daten zum logarithmierten Durchschnittslohn (ln\_avwage) über mehrere Jahre. Diese haben ein Unternehmen (regno) als übergeordnete Einheit.  
Um Fehler bezüglich genesteter Daten zu verhindern, führen wir **Clusterrobuste Standardfehler** ein.

Quiz: Nach welcher Variablen soll geclustert werden?

* ln\_avwage [ ]
* regno [x]

Nun fügen wir die Clusterrobusten Standardfehler in unsere Regression ein und schauen, was sich dadurch verändert.  
Führen Sie hierzu die Regression mithilfe der felm Funktion durch und vergleichen die Regression ohne Clustering (reg1) mit der Regression mit Clustering.

### Info: felm

Im Vergleich zu einem einfachen Regressionsmodell mit lm, kann durch die felm Funktion eine spezifiziertere Regression durchgeführt werden.  
Der Grundaufbau des formulas sieht wie folgt aus:  
**felm(y ~ abhängige Variablen | ausgeschlossene Variablen | Spezifikation einer Instrumentalviariablen | Clusterspezifikation)**  
(<https://www.rdocumentation.org/packages/lfe/versions/2.9-0/topics/felm>)

**Aufgabe**: Tragen Sie die Clusterspezifikation nach der Variablen regno an die passende Stelle ein und setzen die ungenutzten Spezifikationen gleich 0.

# reg2 = felm(ln\_avwage ~ ctreat1 + treat1\_NMW + NMW | \_\_\_ | \_\_\_ | \_\_\_, data=dat)  
# stargazer(reg1, reg2, type ="text")  
  
reg2 = felm(ln\_avwage ~ ctreat1 + treat1\_NMW + NMW | 0 | 0 | regno, data=dat)  
stargazer(reg1, reg2, type ="text")

##

## =====================================================================  
## Dependent variable:   
## -------------------------------------  
## ln\_avwage   
## OLS felm   
## (1) (2)   
## ---------------------------------------------------------------------  
## ctreat1 -0.626\*\*\* -0.626\*\*\*  
## (0.013) (0.024)   
##   
## treat1\_NMW 0.111\*\*\* 0.111\*\*\*   
## (0.019) (0.029)   
##   
## NMW 0.118\*\*\* 0.118\*\*\*   
## (0.009) (0.007)   
##   
## Constant 2.775\*\*\* 2.775\*\*\*   
## (0.006) (0.007)   
##   
## ---------------------------------------------------------------------  
## Observations 4,112 4,112   
## R2 0.501 0.501   
## Adjusted R2 0.500 0.500   
## Residual Std. Error (df = 4108) 0.253 0.253   
## F Statistic 1,373.397\*\*\* (df = 3; 4108)   
## =====================================================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

An den Werten der Schätzer ist keine Veränderung zu erkennen. Allerdings haben sich die Werte der Standardfehler verändert.  
Während man für die Schätzer von ctreat1 und der Interaktionsvariablen treat1\_NMW eine Erhöhung des p-Wertes vorfindet, ist eine Verringerung beim Schätzer der Variablen NMW zu verzeichnen.

Durch die Verwendung von clusterrobusten Standardfehlern wird eine Korrelation der Fehlerterme innerhalb einer Gruppe hergestellt.  
Wenn eine beschreibende Variable innerhalb der Gruppe nicht so stark variiert, kann es sein, dass der Standardfehler in einer Regression ohne clusterrobuste Standardfehler höher ausfällt als er eigentlich ist.  
*(vgl. Angrist et al. (2008), S. 231ff.)*

**Aufgabe:** Betrachten wir nochmal das Ergebnis der aktuellsten Regression, ausführbar mittels *check*

stargazer(reg2, type = "text")

##   
## ===============================================  
## Dependent variable:   
## ---------------------------  
## ln\_avwage   
## -----------------------------------------------  
## ctreat1 -0.626\*\*\*   
## (0.024)   
##   
## treat1\_NMW 0.111\*\*\*   
## (0.029)   
##   
## NMW 0.118\*\*\*   
## (0.007)   
##   
## Constant 2.775\*\*\*   
## (0.007)   
##   
## -----------------------------------------------  
## Observations 4,112   
## R2 0.501   
## Adjusted R2 0.500   
## Residual Std. Error 0.253 (df = 4108)   
## ===============================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Die Schätzer der Variablen beschreiben die Veränderungen im Vergleich zum Pre-Policy Jahr, welche mit Constant beschrieben wird.

Quiz: Wie beschreiben Sie den Wert des Schätzers der DiD-Variablen treat1\_NMW?

* Der Anstieg des Lohns in der Treatmentgruppe ist um ca. 0.11 Pfund höher als in der Kontrollgruppe. [ ]
* Der Anstieg des Lohns in der Treatmentgruppe ist um ca. 11 Prozent höher als in der Kontrollgruppe. [ ]
* Der Anstieg des Lohns in der Treatmentgruppe ist um ca. 11 Prozentpunkte höher als in der Kontrollgruppe. [x]

Wir haben in Erfahrung gebracht, dass es einen Effekt durch die Einführung des gesetzlichen Mindestlohns gibt.  
Um herauszufinden, wie die Unternehmen auf diese Veränderung reagieren, sehen wir uns spiegelbildlich zur obigen Schätzung die Veränderung auf die Gewinnmarge net\_pcm an.

**Aufgabe**: Führen Sie den untenstehenden bekannten Code mit einer angepassten abhängigen Variablen aus und lassen Sie sich das Ergebnis mittels stargazer ausgeben.

# reg3 = felm(\_\_\_ ~ ctreat1 + treat1\_NMW + NMW | 0 | 0 | regno, data = dat)  
#   
# stargazer(\_\_\_, type = "text")  
  
reg3 = felm(net\_pcm ~ ctreat1 + treat1\_NMW + NMW | 0 | 0 | regno, data = dat)  
  
stargazer(reg3, type = "text")

##   
## ===============================================  
## Dependent variable:   
## ---------------------------  
## net\_pcm   
## -----------------------------------------------  
## ctreat1 0.058\*\*\*   
## (0.014)   
##   
## treat1\_NMW -0.027\*\*   
## (0.014)   
##   
## NMW -0.012\*\*\*   
## (0.004)   
##   
## Constant 0.070\*\*\*   
## (0.005)   
##   
## -----------------------------------------------  
## Observations 4,112   
## R2 0.020   
## Adjusted R2 0.019   
## Residual Std. Error 0.152 (df = 4108)   
## ===============================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Wir erkennen, dass die Schätzer signifikante Werte aufweisen und sehen somit, dass unsere zu Beginn von Hand berechneten Werte auch statistisch belastbar sind.  
Auch hier dürfen Sie den Schätzer unserer DiD-Vairablen wieder interpretieren.

Quiz: Was passiert mit der Gewinnmarge, wenn man die Schätzung reg3 als Basis wählt?

* Durch die Mindestlohneinführung ist die Gewinnmarge in Unternehmen der Kontrollgruppe um 2.7 Prozent stärker gesunken. [ ]
* Durch die Mindestlohneinführung ist die Gewinnmarge in Unternehmen der Treatmentgruppe um 2.7 Prozentpunkte stärker gesunken. [x]
* Durch die Mindestlohneinführung ist die Gewinnmarge in Unternehmen der Kontrollgruppe um 2.7 Prozentpunkte stärker gesunken. [ ]

### Award: Earl of DiD

Sie haben die händische Schätzung erfolgreich in eine statistisch belastbare Schätzung überführt und können die Ergebnisse der Regressionen bewerten.

Als nächsten Schritt sehen wir uns weiter an ob es zu Veränderungen kommt, wenn wir auf bestimmte Effekte der Unternehmenswelt kontrollieren.  
Dafür können wir uns die Funktion der **Faktorvariablen** zunutze machen.

Da wir bei nominalskalierten Daten keine Möglichkeit zur messbaren Unterscheidung finden, müssen wir Faktorvariablen nutzen, um diskrete Daten in die Regression miteinzubeziehen.  
Technisch umsetzbar wird diese Faktorisierung, indem eine Gruppe die Referenz bildet. Statistisch gesehen geht diese Wahl willkürlich vonstatten, inhaltlich gibt es nach Hardy (1993) drei Grundüberlegungen zur Wahl der Referenzgruppe:

So soll die Referenzgruppe

**1. Einen sinnvollen Vergleich bieten**  
**2. Klar definiert sein**  
**3. Eine ausreichend große Stichprobengröße im Vergleich zu anderen Gruppen besitzen**

*(vgl. Cohen et al. (2002), S.302ff.)*

### Info: factor

In R können wir entweder die Variablen innerhalb der Regression faktorisieren, oder bereits im Datensatz.

**Aufgabe:** Die Variable year möchten wir in unserem Datensatz faktorisieren und erstellen hierfür eine neue Variable factor\_year. Vergleichen Sie weiter die Werte der neuen mit der bereits existierenden Variablen, indem Sie sich mittels head() die ersten Einträge anzeigen lassen.

# dat = dat %>%  
# mutate("\_\_\_" = factor(year))  
#   
# \_\_\_(dat$year)  
# \_\_\_(dat$factor\_year)  
  
dat = dat %>%  
 mutate("factor\_year" = factor(year))  
  
head(dat$year)

## [1] 1999 2000 2001 2002 1997 1998

head(dat$factor\_year)

## [1] 1999 2000 2001 2002 1997 1998  
## Levels: 1997 1998 1999 2000 2001 2002

Die Einträge scheinen zunächst identisch, dennoch ist bei unserer faktorisierten Variablen der Eintrag *Levels* zu finden.  
*Levels* gibt die unterschiedlichen Kategorien innerhalb der Variablen an. Dabei stellt die erste Kategorie die Referenzkategorie dar.

Quiz: Wie bewerten Sie die gewählte Referenzkategorie? (Denken Sie dabei an die oben beschriebenen Grundüberlegungen)

* Das Jahr 1997 ist sinnvoll gewählt, da es den Startpunkt unserer Analyse markiert. [ ]
* Es gibt andere Jahre, die auf Basis unserer Analyse einen wichtigeren Punkt markieren [x]
* Alle Jahre sind gleich relevant und die Grundüberlegungen treffen auf jedes Jahr zu. Daher kann das Jahr beliebig gewählt werden. [ ]

Eine besser geeignete Referenzkategorie können wir selbst festlegen.

**Aufgabe**: Wählen Sie das Jahr 2000 als Referenzwert für unsere Faktorisierung.

# dat$factor\_year = relevel(dat$factor\_year, ref = "\_\_\_")  
  
dat$factor\_year = relevel(dat$factor\_year, ref = "2000")

Sehen wir uns nun an, was diese Faktorisierung für die Regression bedeutet, indem wir die Regression einmal mit year und einmal mit factor\_year ergänzen.

**Aufgabe**: Geben Sie den Vergleich der unterschiedlichen Regressionen aus, indem Sie den Code mittels *check* ausführen.

reg3 = felm(net\_pcm ~ ctreat1 + treat1\_NMW + NMW | 0 | 0 | regno, data = dat)  
reg3.1 = felm(net\_pcm ~ ctreat1 + treat1\_NMW + NMW + year| 0 | 0 | regno, data = dat)  
reg3.2 = felm(net\_pcm ~ ctreat1 + treat1\_NMW + NMW + factor\_year| 0 | 0 | regno, data = dat)

## Warning in chol.default(mat, pivot = TRUE, tol = tol): die Matrix hat entweder  
## nicht vollen Rang oder ist indefinit

reg3.3 = felm(net\_pcm ~ ctreat1 + treat1\_NMW + NMW + factor(year)| 0 | 0 | regno, data = dat)

## Warning in chol.default(mat, pivot = TRUE, tol = tol): die Matrix hat entweder  
## nicht vollen Rang oder ist indefinit

stargazer(reg3, reg3.1, reg3.2, reg3.3, type= "text")

##   
## ===========================================================================================  
## Dependent variable:   
## -----------------------------------------------------------------------  
## net\_pcm   
## (1) (2) (3) (4)   
## -------------------------------------------------------------------------------------------  
## ctreat1 0.058\*\*\* 0.058\*\*\* 0.058\*\*\* 0.058\*\*\*   
## (0.014) (0.014) (0.014) (0.014)   
##   
## treat1\_NMW -0.027\*\* -0.027\*\* -0.028\*\* -0.028\*\*   
## (0.014) (0.014) (0.014) (0.014)   
##   
## NMW -0.012\*\*\* -0.001 -0.012\*\* -0.015\*\*   
## (0.004) (0.007) (0.006) (0.007)   
##   
## year -0.004\*   
## (0.002)   
##   
## factor\_year1997 -0.008\*   
## (0.005)   
##   
## factor\_year1998   
## (0.000)   
##   
## factor\_year1999 -0.011\*\*   
## (0.005)   
##   
## factor\_year2001 -0.010\*   
## (0.006)   
##   
## factor\_year2002 -0.011\*   
## (0.007)   
##   
## factor(year)1998 0.008\*   
## (0.005)   
##   
## factor(year)1999 -0.002   
## (0.006)   
##   
## factor(year)2000 0.011\*   
## (0.007)   
##   
## factor(year)2001 0.001   
## (0.006)   
##   
## factor(year)2002   
## (0.000)   
##   
## Constant 0.070\*\*\* 7.549\* 0.076\*\*\* 0.068\*\*\*   
## (0.005) (4.295) (0.006) (0.006)   
##   
## -------------------------------------------------------------------------------------------  
## Observations 4,112 4,112 4,112 4,112   
## R2 0.020 0.020 0.021 0.021   
## Adjusted R2 0.019 0.019 0.019 0.019   
## Residual Std. Error 0.152 (df = 4108) 0.152 (df = 4107) 0.152 (df = 4104) 0.152 (df = 4104)  
## ===========================================================================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Man sieht, dass sich die Ergebnisse durch eine Faktorisierung verändern.  
Auch durch die Veränderung der Referenzkategorie entstehen Unterschiede, wenn auch im überschaubaren Rahmen.

Dabei gilt dennoch zu erwähnen, dass die Autoren des Artikels bei der Wahl der Referenzkategorie wohl einer anderen Idee der Wahl gefolgt sind und das Jahr 1997 als Referenz gewählt haben.

Da durch die Darstellung der Faktorvariablen eine längere Ausgabe entsteht, kann es beim Hinzufügen mehrerer Faktorvariablen aus Gründen der Übersichtlichkeit und Vergleichbarkeit sinnvoll sein, unsere Ausgabe auf das Relevanteste reduzieren.

Hierbei lernen Sie eine alternative Darstellung der Regressionsergebnisse zu stargazer kennen.

**Aufgabe:** Laden Sie zunächst das Paket broom.

# library(\_\_\_)  
  
library(broom)

**Aufgabe**: Führen Sie den Code aus, um aus der Liste - die uns durch die felmFunktion erstellt wird - eine Tabelle der Regression zu kreieren. Geben Sie dabei anschließend die Tabelle aus.

# table\_reg3.3 = tidy(reg3.3)  
#   
# #Tabelle ausgeben  
#   
  
table\_reg3.3 = tidy(reg3.3)

## Warning in chol.default(mat, pivot = TRUE, tol = tol): die Matrix hat entweder  
## nicht vollen Rang oder ist indefinit

#Tabelle ausgeben  
table\_reg3.3

## term estimate std.error statistic p.value  
## 1 (Intercept) 0.06796 0.00616 11.03903 0.00000  
## 2 ctreat1 0.05820 0.01365 4.26394 0.00002  
## 3 treat1\_NMW -0.02751 0.01371 -2.00661 0.04486  
## 4 NMW -0.01460 0.00710 -2.05636 0.03981  
## 5 factor(year)1998 0.00844 0.00477 1.76730 0.07725  
## 6 factor(year)1999 -0.00244 0.00589 -0.41373 0.67909  
## 7 factor(year)2000 0.01122 0.00682 1.64576 0.09989  
## 8 factor(year)2001 0.00127 0.00574 0.22196 0.82435  
## 9 factor(year)2002 NaN 0.00000 NaN NaN

Nun haben wir die relevanten Werte der Regression in der Tabelle gespeichert und können darauf wie auf jeden anderen Data Frame zurückgreifen.

### Award: tidy broomer

Glückwunsch, Sie haben die Regressionsausgaben erfolgreich in einer Tabelle gespeichert.   
Um mehr über das broom Paket zu erfahren, geht es [hier](https://cran.r-project.org/web/packages/broom/vignettes/broom.html) entlang.

**Hinweis:** Das Tool wird genutzt, um einen besseren Überblick über die Veränderung durch Variablen zu erhalten. Es dient als Hilfsmittel und nicht als genereller Mechanismus, um Regressionsergebnisse zu analysieren.  
In Anhang 2 können Sie sich die vollständigen Ausgaben der Regressionsergebnisse ausgeben lassen.

Nachdem wir nun die technischen Voraussetzungen geschaffen haben, können wir fachlich fortfahren und unsere Schätzung durch das Hinzufügen weiterer möglich beeinflussende Variablen präzisieren.

Es geht darum zu überprüfen, ob die Veränderung der Unternehmensgewinne wirklich auf die Einführung des Mindestlohns zurückgeht und ob weitere Faktoren die Gewinnmarge eines Unternehmens beeinflussen können.  
Dabei prüfen wir auf unternehmens-, branchen- und ortsspezifische Größen.

**Aufgabe:** Führen Sie die Regressionen mittels *check* aus.

reg4 = felm(ln\_avwage ~ ctreat1 + treat1\_NMW + NMW + grad2 + unionmem + ptwk + female +   
 factor(sic2) + factor(year) + factor(gorwk)| 0 | 0 | regno, data = dat)

reg5 = felm(net\_pcm ~ ctreat1 + treat1\_NMW + NMW + grad2 + unionmem + ptwk + female +   
 factor(sic2) + factor(year) + factor(gorwk)| 0 | 0 | regno, data = dat)

Das waren sehr viele Regressionen, da kann es manchmal kompliziert sein den Überblick zu behalten.  
Um die wichtigsten Ergebnisse knapp zusammenfassen zu können, gibt es die Möglichkeit einzelne Werte aus den jeweiligen Regressionen auszuwählen und sie dann in einer Tabelle gegenüberzustellen.  
Dabei sehen wir uns die Veränderung des Kernergebnisses an.

**Aufgabe:** Um Ergebnisse auslesen zu können, wenden Sie ihre vorher erworbenen Kenntnisse zur Überführung von Regressionen in eine Tabelle an.

# tidy\_reg1 = \_\_\_(reg1)  
# tidy\_reg2 = \_\_\_(reg2)  
# tidy\_reg3 = \_\_\_(reg3)  
# tidy\_reg4 = \_\_\_(reg4)  
# tidy\_reg5 = \_\_\_(reg5)  
  
tidy\_reg1 = tidy(reg1)  
tidy\_reg2 = tidy(reg2)  
tidy\_reg3 = tidy(reg3)  
tidy\_reg4 = tidy(reg4)

tidy\_reg5 = tidy(reg5)

**Aufgabe**: Erzeugen Sie zunächst die gegenüberstellende Tabelle der Regression ohne und mit Kontrollvariablen mittels *check* und lassen Sie sie ausgeben. Runden Sie die Ergebniss dabei auf drei Nachkommastellen.

# reg\_table\_ln\_avwage = data.frame("Variable" = "treat1\_NMW","Schätzer\_reg2" = round(tidy\_reg2$estimate[3],\_\_\_),"se\_reg2"= round(tidy\_reg2$std.error[3],\_\_\_), "Schätzer\_reg4" = round(tidy\_reg4$estimate[3],\_\_\_), "se\_reg4"= round(tidy\_reg4$std.error[3],\_\_\_))  
#   
# reg\_table\_ln\_avwage

reg\_table\_ln\_avwage = data.frame("Variable" = "treat1\_NMW","Schätzer\_reg2" = round(tidy\_reg2$estimate[3],3),"se\_reg2"= round(tidy\_reg2$std.error[3],3), "Schätzer\_reg4" = round(tidy\_reg4$estimate[3],3), "se\_reg4"= round(tidy\_reg4$std.error[3],3))  
  
reg\_table\_ln\_avwage

## Variable Schätzer\_reg2 se\_reg2 Schätzer\_reg4 se\_reg4  
## 1 treat1\_NMW 0.111 0.029 0.09 0.026

Der positive Effekt der Mindestlohneinführung auf die Löhne der Treatmentgruppe ist weiterhin erkennbar, hat sich etwas verringert und ist weiterhin signifikant.  
Zudem wirkt die Wahl der erklärenden Variablen sinnvoll, da sie einen signifikanten Einfluss auf den Lohn haben. *(vgl. Anhang 2.1)*

Sehen wir uns das Ganze auch noch für die Veränderung in der Gewinnmarge an.

**Aufgabe:** Führen Sie den Code mittels *check* aus.

reg\_table\_net\_pcm = data.frame("Variable" = "treat1\_NMW","Schätzer\_reg3" = round(tidy\_reg3$estimate[3],3),"se\_reg3"= round(tidy\_reg3$std.error[3],3), "Schätzer\_reg5" = round(tidy\_reg5$estimate[3],3), "se\_reg5"= round(tidy\_reg5$std.error[3],3))  
  
reg\_table\_net\_pcm

## Variable Schätzer\_reg3 se\_reg3 Schätzer\_reg5 se\_reg5  
## 1 treat1\_NMW -0.027 0.014 -0.029 0.012

Auch hier können wir feststellen, dass unsere Schätzung weiter an Robustheit gewinnt und dadurch unsere These der Verringerung des Gewinnmarge durch die Mindestlohneinführung weiter aufrechterhalten werden kann.

## 3.4 Weiterführende Analysen

Neben den von uns dargestellten Ergebnissen, gibt es im untersuchten Artikel weitere interessante Darstellungen, die unsere bisherige Analyse zu den Auswirkungen auf Unternehmen im Niedriglohnsektor untermauern und erweitern.

### Placebo-Experiment

Durch die Einführung eines Placebo-Experiments wird festgestellt, dass es vor der Mindestlohneinführung keine relevanten Zusammenhänge zwischen der Treatment- und Kontrollgruppe bestanden.  
Teil dieses Experiments ist die Prüfung einer imaginären Mindestlohneinführung 3 Jahre vor der wahren Einführung.  
Dabei wurde kein Trend erkennbar, der auf einen bereits bestehenden Zusammenhang der beiden Gruppen schließen würde.  
*(vgl. Draca et al., S.142ff.)*

### Unterschied zwischen Industrien

Die Autoren vergleichen die Wirkung der Mindestlohneinführung Wettbewerben mit hoher und niedriger Marktmacht.  
Dabei kommen die Autoren zu dem Schluss, dass es Unterschiede im Umgang mit Mindestlohneinführungen gibt.  
Während Unternehmen mit einer verhältnismäßig geringen Marktmacht die Erhöhung der Löhne hauptsächlich über die Verringerung der Gewinnmarge kompensieren, ist zu erkennen, dass marktmächtige Unternehmen die Erhöhung durch Preisanpassungen an Kunden weitergeben.  
*(vgl. Draca et al., S.146f.)*

### Auswirkungen auf weitere Unternehmensfaktoren

Wie zu Beginn erwähnt könnten neben des Gewinns auch noch weitere Faktoren von der Einführung des nationalen Mindestlohns betroffen sein.  
Sowohl auf eine Veränderung am Arbeitsmarkt, als auch auf eine Veränderung in der Produktionskraft von nNternehmen werden keine signifikanten Einflüsse erkannt.  
Derweil ist ein langfristiger Trend zu erkennen der die Vermutung zulässt, dass es zu weniger Markteintritten in den stark vom der Mindestlohneinführung betroffen Unternehmen kommt.  
*(vgl. Draca et al., S.148f.)*

## 4 Einordnung des Artikels

Die wissenschaftliche Arbeit von Draca et al. macht sich die besondere Eigenschaft des quasi-experimentellen Experiments zu eigen und wählt dazu passenderweise die Methodik der Difference-in-Differences Schätzung. Dabei zeigen die Autoren gut auf, dass die Einführung von Mindeslöhnen zu einer Veränderung der Unternehmensgewinne führt.

Hinzu kommt, dass für die häufig untersuchte Frage der Veränderung in der Beschäftigung kein relevanter Effekt erkannt wird. Durch die Wahl von plausiblen Kontrollvariablen aus unterschiedlichen Bereichen der Unternehmenswelt wirkt die Argumentation fundiert.  
Der zunächst nicht evidenten Einteilung in Kontroll- und Treatmentgruppe wird sich genähert und anhand von Lohnstrukturen werden die Daten logisch nachvollziehbar eingeteilt.

Kritisch anzumerken ist die Wahl der Beispieldatensätze. Hier wird wenig transparent argumentiert und eine klare Struktur der Idee der Selektion ist nicht erkennbar.  
Da wir keine genauen Daten für jeden Arbeitnehmer haben, ist unsere Analyse vor allem auf den Niedriglohnsektor anwendbar, wobei mit einem Anteil von 87% auch ein relativ großer Teil der Arbeiter abgebildet wird.  
Begleitend zeigt das eine Unternehmensstruktur auf, in der Lohnunterschiede vorallem zwischen Sektoren variieren und weniger innerhalb eines Unternehmens.  
Eine weitere Komponente, die im Rahmen der Analyse vermisst wird, ist eine finanzielle Rahmenbedingung als Kontrollvariable.  
Der Output des Unternehmens bleibt weitestgehend unberücksichtigt, da im Datensatz keine Daten zu Produkten wie Preise oder Qualität vorhanden sind. Diese Problematik ist auch den Autoren bewusst und sie weisen auf diese Einschränkung in der Bewertung hin.

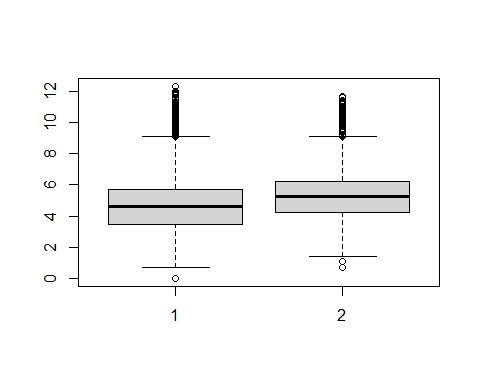
* Weitere Forschung Höherer Rückgang in der Marge bei Unternehmen mit großer Marktmacht entdeckt
* Rückbezug zur Einleitung
* Ist ein Mindestlohn zu befürworten

## Anhang

### Anhang 1

Werfen wir einen Blick auf die aussortierten Daten und überprüfen anhand der Beschäftigungszahlen, ob es Aufffälligkeiten gibt:

dat = read\_dta('main\_fame.dta')  
dat\_no\_pp = dat %>%  
 filter(pp == 0)  
  
dat\_pp = dat %>%  
 filter(pp == 1)  
  
boxplot(log(dat\_no\_pp$emp),log(dat\_pp$emp))



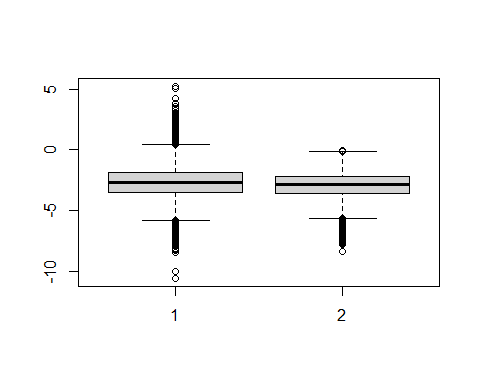
boxplot(log(dat\_no\_pp$net\_pcm),log(dat\_pp$net\_pcm))

## Warning in log(dat\_no\_pp$net\_pcm): NaNs wurden erzeugt

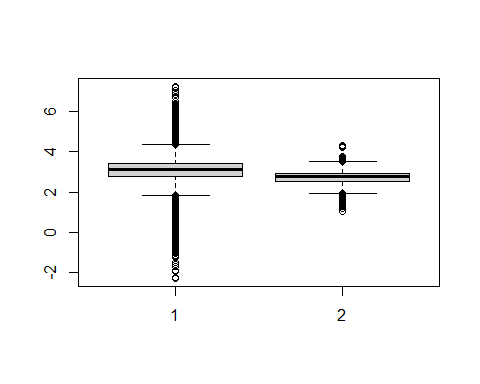
## Warning in log(dat\_pp$net\_pcm): NaNs wurden erzeugt

## Warning in bplt(at[i], wid = width[i], stats = z$stats[, i], out =  
## z$out[z$group == : Ausreißer (-Inf) im Boxplot 1 wird nicht gezeichnet

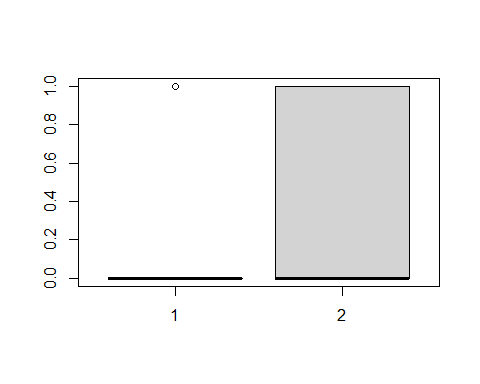
## Warning in bplt(at[i], wid = width[i], stats = z$stats[, i], out =  
## z$out[z$group == : Ausreißer (-Inf) im Boxplot 2 wird nicht gezeichnet



boxplot(log(dat\_no\_pp$avwage),log(dat\_pp$avwage))



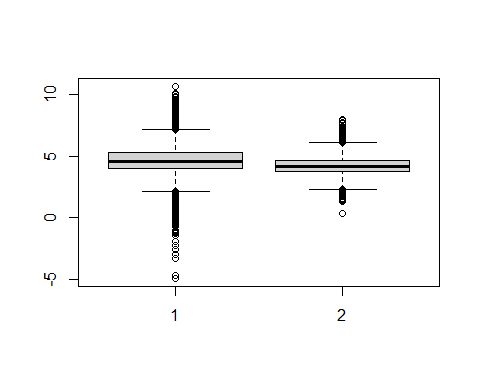
boxplot((dat\_no\_pp$manuf),(dat\_pp$manuf))



boxplot(log(dat\_no\_pp$turnemp),log(dat\_pp$turnemp))

## Warning in log(dat\_no\_pp$turnemp): NaNs wurden erzeugt

## Warning in bplt(at[i], wid = width[i], stats = z$stats[, i], out =  
## z$out[z$group == : Ausreißer (-Inf) im Boxplot 1 wird nicht gezeichnet



Im Gesamten ist zu erkennen, dass die Daten, die sich nicht in der Auswahl befinden, breiter gestreut sind.  
Man erkennt einen Unterschied in den Beschäftigungszahlen der beiden Gruppen.

## Anhang 2

### Anhang 2.1

Regressionsergebnisse zum logarithmierten Durchschnittslohn:

reg2 = felm(ln\_avwage ~ ctreat1 + treat1\_NMW + NMW | 0 | 0 | regno, data=dat\_pp)  
  
reg4 = felm(ln\_avwage ~ ctreat1 + treat1\_NMW + NMW + grad2 + unionmem + ptwk + female +   
 factor(sic2) + factor(year) + factor(gorwk)| 0 | 0 | regno, data = dat\_pp)

## Warning in chol.default(mat, pivot = TRUE, tol = tol): die Matrix hat entweder  
## nicht vollen Rang oder ist indefinit

stargazer(reg2, reg4, type ="text")

##   
## =======================================================  
## Dependent variable:   
## -----------------------------------  
## ln\_avwage   
## (1) (2)   
## -------------------------------------------------------  
## ctreat1 -0.626\*\*\* -0.548\*\*\*   
## (0.024) (0.025)   
##   
## treat1\_NMW 0.111\*\*\* 0.090\*\*\*   
## (0.029) (0.026)   
##   
## NMW 0.118\*\*\* 0.212\*\*\*   
## (0.007) (0.012)   
##   
## grad2 0.339\*\*\*   
## (0.114)   
##   
## unionmem -0.335\*\*\*   
## (0.096)   
##   
## ptwk -1.002\*\*\*   
## (0.187)   
##   
## female 0.123   
## (0.086)   
##   
## factor(sic2)5 -0.342\*\*\*   
## (0.075)   
##   
## factor(sic2)10 0.089   
## (0.096)   
##   
## factor(sic2)14 -0.155   
## (0.194)   
##   
## factor(sic2)15 -0.143\*   
## (0.082)   
##   
## factor(sic2)17 -0.156\*   
## (0.089)   
##   
## factor(sic2)18 -0.189   
## (0.164)   
##   
## factor(sic2)20 -0.186\*\*   
## (0.083)   
##   
## factor(sic2)21 -0.039   
## (0.084)   
##   
## factor(sic2)22 -0.143   
## (0.089)   
##   
## factor(sic2)23 -0.089   
## (0.089)   
##   
## factor(sic2)24 -0.164\*   
## (0.093)   
##   
## factor(sic2)25 -0.157\*   
## (0.091)   
##   
## factor(sic2)26 -0.076   
## (0.088)   
##   
## factor(sic2)27 -0.016   
## (0.095)   
##   
## factor(sic2)28 -0.097   
## (0.082)   
##   
## factor(sic2)29 -0.095   
## (0.086)   
##   
## factor(sic2)30 -0.424\*\*\*   
## (0.096)   
##   
## factor(sic2)31 -0.192\*\*   
## (0.089)   
##   
## factor(sic2)32 -0.230\*\*\*   
## (0.085)   
##   
## factor(sic2)33 -0.214\*   
## (0.124)   
##   
## factor(sic2)34 -0.051   
## (0.117)   
##   
## factor(sic2)35 -0.025   
## (0.095)   
##   
## factor(sic2)36 -0.139\*   
## (0.081)   
##   
## factor(sic2)40 -0.030   
## (0.163)   
##   
## factor(sic2)41 0.034   
## (0.093)   
##   
## factor(sic2)45 -0.061   
## (0.083)   
##   
## factor(sic2)50 -0.098   
## (0.076)   
##   
## factor(sic2)51 -0.150\*   
## (0.077)   
##   
## factor(sic2)52 0.108   
## (0.098)   
##   
## factor(sic2)55 -0.032   
## (0.091)   
##   
## factor(sic2)60 -0.067   
## (0.085)   
##   
## factor(sic2)61 -0.211\*\*   
## (0.100)   
##   
## factor(sic2)62 0.136   
## (0.096)   
##   
## factor(sic2)63 -0.091   
## (0.097)   
##   
## factor(sic2)64 -0.071   
## (0.113)   
##   
## factor(sic2)65 -0.130   
## (0.087)   
##   
## factor(sic2)66 -0.117   
## (0.093)   
##   
## factor(sic2)67 -0.203\*\*   
## (0.083)   
##   
## factor(sic2)70 -0.120   
## (0.088)   
##   
## factor(sic2)71 -0.116   
## (0.090)   
##   
## factor(sic2)72 0.794\*   
## (0.425)   
##   
## factor(sic2)74 -0.174\*\*   
## (0.084)   
##   
## factor(sic2)80 0.036   
## (0.150)   
##   
## factor(sic2)85 0.029   
## (0.116)   
##   
## factor(sic2)91 -0.124   
## (0.083)   
##   
## factor(sic2)92 -0.088   
## (0.085)   
##   
## factor(sic2)93 -0.056   
## (0.089)   
##   
## factor(sic2)95 0.326\*\*\*   
## (0.105)   
##   
## factor(year)1998 0.020\*\*\*   
## (0.007)   
##   
## factor(year)1999 0.018\*\*   
## (0.009)   
##   
## factor(year)2000 -0.153\*\*\*   
## (0.011)   
##   
## factor(year)2001 -0.060\*\*\*   
## (0.009)   
##   
## factor(year)2002   
## (0.000)   
##   
## factor(gorwk)2 0.113\*   
## (0.064)   
##   
## factor(gorwk)3 0.022   
## (0.061)   
##   
## factor(gorwk)4 -0.001   
## (0.072)   
##   
## factor(gorwk)5 -0.069   
## (0.059)   
##   
## factor(gorwk)6 -0.062   
## (0.068)   
##   
## factor(gorwk)7 -0.021   
## (0.058)   
##   
## factor(gorwk)8 0.032   
## (0.063)   
##   
## factor(gorwk)9 -0.017   
## (0.055)   
##   
## factor(gorwk)10 -0.003   
## (0.057)   
##   
## factor(gorwk)11 0.038   
## (0.055)   
##   
## factor(gorwk)13 -0.082   
## (0.062)   
##   
## factor(gorwk)14 0.004   
## (0.059)   
##   
## factor(gorwk)15 0.029   
## (0.056)   
##   
## factor(gorwk)16 0.001   
## (0.054)   
##   
## factor(gorwk)17 0.042   
## (0.061)   
##   
## factor(gorwk)19 -0.006   
## (0.056)   
##   
## factor(gorwk)20 0.053   
## (0.067)   
##   
## Constant 2.775\*\*\* 3.017\*\*\*   
## (0.007) (0.101)   
##   
## -------------------------------------------------------  
## Observations 4,112 4,112   
## R2 0.501 0.590   
## Adjusted R2 0.500 0.582   
## Residual Std. Error 0.253 (df = 4108) 0.232 (df = 4035)  
## =======================================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

### Anhang 2.2

Regressionsergebnisse zur Gewinnmarge:

reg3 = felm(net\_pcm ~ ctreat1 + treat1\_NMW + NMW | 0 | 0 | regno, data = dat\_pp)  
  
reg5 = felm(net\_pcm ~ ctreat1 + treat1\_NMW + NMW + grad2 + unionmem + ptwk + female +   
 factor(sic2) + factor(year) + factor(gorwk)| 0 | 0 | regno, data = dat\_pp)

## Warning in chol.default(mat, pivot = TRUE, tol = tol): die Matrix hat entweder  
## nicht vollen Rang oder ist indefinit

stargazer(reg3, reg5, type = "text")

## =======================================================  
## Dependent variable:   
## -----------------------------------  
## net\_pcm   
## (1) (2)   
## -------------------------------------------------------  
## ctreat1 0.058\*\*\* 0.037\*\*\*   
## (0.014) (0.013)   
##   
## treat1\_NMW -0.027\*\* -0.029\*\*   
## (0.014) (0.012)   
##   
## NMW -0.012\*\*\* -0.012   
## (0.004) (0.007)   
##   
## grad2 -0.024   
## (0.063)   
##   
## unionmem 0.115   
## (0.098)   
##   
## ptwk -0.129\*\*   
## (0.054)   
##   
## female 0.063   
## (0.051)   
##   
## factor(sic2)5 -0.219\*\*\*   
## (0.013)   
##   
## factor(sic2)10 -0.074   
## (0.059)   
##   
## factor(sic2)14 -0.067   
## (0.092)   
##   
## factor(sic2)15 -0.020   
## (0.032)   
##   
## factor(sic2)17 -0.058\*   
## (0.031)   
##   
## factor(sic2)18 -0.053   
## (0.040)   
##   
## factor(sic2)20 0.015   
## (0.020)   
##   
## factor(sic2)21 0.001   
## (0.042)   
##   
## factor(sic2)22 0.006   
## (0.036)   
##   
## factor(sic2)23 0.026   
## (0.051)   
##   
## factor(sic2)24 -0.043   
## (0.038)   
##   
## factor(sic2)25 -0.024   
## (0.022)   
##   
## factor(sic2)26 -0.010   
## (0.042)   
##   
## factor(sic2)27 -0.051   
## (0.045)   
##   
## factor(sic2)28 -0.024   
## (0.021)   
##   
## factor(sic2)29 -0.025   
## (0.028)   
##   
## factor(sic2)30 -0.085\*\*   
## (0.034)   
##   
## factor(sic2)31 -0.037   
## (0.032)   
##   
## factor(sic2)32 -0.014   
## (0.026)   
##   
## factor(sic2)33 -0.014   
## (0.020)   
##   
## factor(sic2)34 -0.096\*   
## (0.050)   
##   
## factor(sic2)35 -0.054   
## (0.051)   
##   
## factor(sic2)36 -0.003   
## (0.020)   
##   
## factor(sic2)40 0.216   
## (0.163)   
##   
## factor(sic2)41 0.184\*\*\*   
## (0.065)   
##   
## factor(sic2)45 -0.009   
## (0.022)   
##   
## factor(sic2)50 -0.016   
## (0.012)   
##   
## factor(sic2)51 -0.008   
## (0.012)   
##   
## factor(sic2)52 0.024   
## (0.022)   
##   
## factor(sic2)55 0.092\*\*\*   
## (0.022)   
##   
## factor(sic2)60 -0.036   
## (0.033)   
##   
## factor(sic2)61 0.041   
## (0.089)   
##   
## factor(sic2)62 -0.102   
## (0.065)   
##   
## factor(sic2)63 -0.010   
## (0.030)   
##   
## factor(sic2)64 0.010   
## (0.057)   
##   
## factor(sic2)65 0.056   
## (0.041)   
##   
## factor(sic2)66 0.063   
## (0.049)   
##   
## factor(sic2)67 -0.056\*   
## (0.033)   
##   
## factor(sic2)70 0.262\*\*\*   
## (0.056)   
##   
## factor(sic2)71 0.068\*   
## (0.039)   
##   
## factor(sic2)72 -0.112   
## (0.124)   
##   
## factor(sic2)74 0.005   
## (0.023)   
##   
## factor(sic2)80 -0.063   
## (0.067)   
##   
## factor(sic2)85 -0.055   
## (0.063)   
##   
## factor(sic2)91 -0.022   
## (0.029)   
##   
## factor(sic2)92 0.019   
## (0.029)   
##   
## factor(sic2)93 0.033   
## (0.026)   
##   
## factor(sic2)95 0.074\*\*   
## (0.034)   
##   
## factor(year)1998 0.008\*   
## (0.004)   
##   
## factor(year)1999 -0.004   
## (0.006)   
##   
## factor(year)2000 0.008   
## (0.006)   
##   
## factor(year)2001 0.002   
## (0.005)   
##   
## factor(year)2002   
## (0.000)   
##   
## factor(gorwk)2 -0.004   
## (0.028)   
##   
## factor(gorwk)3 0.013   
## (0.029)   
##   
## factor(gorwk)4 0.021   
## (0.030)   
##   
## factor(gorwk)5 0.024   
## (0.026)   
##   
## factor(gorwk)6 0.009   
## (0.026)   
##   
## factor(gorwk)7 -0.021   
## (0.027)   
##   
## factor(gorwk)8 -0.011   
## (0.040)   
##   
## factor(gorwk)9 0.026   
## (0.025)   
##   
## factor(gorwk)10 0.007   
## (0.028)   
##   
## factor(gorwk)11 0.008   
## (0.025)   
##   
## factor(gorwk)13 0.028   
## (0.032)   
##   
## factor(gorwk)14 0.031   
## (0.029)   
##   
## factor(gorwk)15 0.020   
## (0.026)   
##   
## factor(gorwk)16 0.016   
## (0.025)   
##   
## factor(gorwk)17 0.013   
## (0.032)   
##   
## factor(gorwk)19 0.014   
## (0.025)   
##   
## factor(gorwk)20 0.023   
## (0.032)   
##   
## Constant 0.070\*\*\* 0.018   
## (0.005) (0.033)   
##   
## -------------------------------------------------------  
## Observations 4,112 4,112   
## R2 0.020 0.294   
## Adjusted R2 0.019 0.281   
## Residual Std. Error 0.152 (df = 4108) 0.130 (df = 4035)  
## =======================================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

## 

## Referenzen

**Literatur**

* Aarts, E., Verhage, M., Veenvliet, J. V., Dolan, C. V., & Van Der Sluis, S. (2014). A solution to dependency: using multilevel analysis to accommodate nested data. Nature neuroscience, 17(4), 491-496.
* Abadie, A., Athey, S., Imbens, G. W., & Wooldridge, J. M. (2023). When should you adjust standard errors for clustering?. The Quarterly Journal of Economics, 138(1), 1-35.
* Angrist, J. D., & Pischke, J. S. (2009). Mostly harmless econometrics: An empiricist’s companion. Princeton university press.
* Callaway, B., & Sant’Anna, P. H. (2021). Difference-in-differences with multiple time periods. Journal of econometrics, 225(2), 200-230.
* Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2013). Applied multiple regression/correlation analysis for the behavioral sciences. Routledge.; Abrufbar unter: <https://doi.org/10.4324/9780203774441>
* Draca, M., Machin, S., & Van Reenen, J. (2011). Minimum wages and firm profitability. American economic journal: applied economics, 3(1), 129-151.
* Brown, William (2007), Friedrich-Ebert-Stiftung, Büro London, London; abrufbar unter: <https://library.fes.de/pdf-files/bueros/london/04855-20071107.pdf>
* OECD (2002), OECD Employment Outlook 2002, OECD Publishing, Paris
* <https://webarchive.nationalarchives.gov.uk/ukgwa/20160105160709/http://www.ons.gov.uk/ons/guide-method/classifications/archived-standard-classifications/uk-standard-industrial-classification-1992--sic92-/uk-sic-2003.pdf> (Letzter Aufruf: 05.02.2025)

**R Pakete und Darstellungen**

* <https://dplyr.tidyverse.org/reference/group_by.html> (Letzter Aufruf: 05.02.2025)
* <https://cran.r-project.org/web/packages/ggdag/vignettes/intro-to-dags.html> (Letzter Aufruf: 05.02.2025)
* <https://r-causal.github.io/ggdag/> (Letzter Aufruf: 05.02.2025)
* <https://markdown-syntax.de/Syntax-GFM/Tabellen/> (Letzter Aufruf: 05.02.2025)
* <https://www.rdocumentation.org/packages/stargazer/versions/5.2.3/topics/stargazer> (Letzter Aufruf: 05.02.2025)
* <https://www.rdocumentation.org/packages/lfe/versions/2.9-0/topics/felm> (Letzter Aufruf: 05.02.2025)
* <https://cran.r-project.org/web/packages/broom/vignettes/broom.html> (Letzter Aufruf: 05.02.2025)

## Einreichen Ihrer Lösung

Um Ihre Lösung abzugeben, gehen Sie wie folgt vor:

1. Scrollen Sie nach oben und drücken auf das Balkensymbol um zu sehen, wie viele Punkte Sie bisher erreicht haben. Um eine höhere Punktzahl zu erlangen, können Sie sich an den noch ungelösten Aufgaben versuchen.
2. Wenn Sie Ihre Lösung abgeben wollen, drücken Sie auf den Download ganz rechts.
3. Im geöffneten Tab wählen Sie die Schaltfläche “Download Submission File”. Ihr Browser sollte daraufhin eine Datei mit der Endung .sub herunterladen.
4. Laden Sie Ihre heruntergeladene Datei auf einer entsprechenden Lernplattform wie Moodle hoch, genaue Anweisungen erhalten Sie von Ihrem Dozenten.