

Modul 3: Multiple Lineare Regression

PI 6250 – Ökonometrie I

Max Heinze (mheinze@wu.ac.at)

Department für Volkswirtschaftslehre, WU
Wien

Basierend auf einem Foliensatz von [Simon Heß](#)

10. April 2025

Einführung

Vektor- und Matrixnotation

Multivariate vs. bivariate Modelle

Praktisches Beispiel

Was wir bisher getan haben

- Wir haben bisher **bivariate Modelle** betrachtet, also Modelle, die **zwei Variablen** haben: eine **abhängige Variable y** und eine **unabhängige Variable x** .
 - Die Anwendungsbereiche für solche Modelle sind relativ beschränkt.
 - Oft haben Variablen mehrere Einflussfaktoren und wir wollen das modellieren.
 - Wir haben über Dummy-Variablen gesprochen, hatten aber noch keine Möglichkeit, *mehrere* Dummy-Variablen im selben Modell zu berücksichtigen.
- Daher gehen wir jetzt weiter zu **multivariaten Modellen**: Modellen mit **mehr als zwei Variablen**. Wir haben nach wie vor **eine abhängige Variable y** , aber jetzt **mehrere unabhängige Variablen x_1, x_2, \dots, x_K** .
 - In diesem Kapitel gehen wir ähnliche Themen durch wie im vorigen Kapitel, aber auf diesen allgemeineren Fall angepasst.

Das einfachste multivariate Modell

Fangen wir mit dem einfachstmöglichen Modell mit mehr als zwei Variablen an:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i$$

- Die Beobachtungen unserer erklärenden Variablen (x_{ik}) haben jetzt **zwei Subskripte**:
 - Das erste Subskript, $i = 1, \dots, N$, bezeichnet nach wie vor die einzelne Beobachtung (z.B. das Individuum).
 - Das zweite Subskript, $k = 1, \dots, K$, indiziert die einzelnen erklärenden Variablen.
 - Die Reihenfolge ist dann wichtig, wenn wir das Modell in Matrixnotation anschreiben.

Wie **interpretieren** wir die **Parameter** in so einem Modell?

Parameter in multivariaten Modellen

Wie **interpretieren** wir die **Parameter** in so einem Modell?

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i$$

Der Parameter

$$\beta_1 = \frac{\partial E(y_i | x_{i1}, x_{i2})}{\partial x_{i1}}$$

- misst den **erwarteten** Unterschied der Variable y ,
- wenn wir die **Variable x_1** um eine Einheit ändern,
- und gleichzeitig **alle anderen beobachteten Variablen unverändert** lassen.

Diese Interpretation wird oft als **Ceteris-Paribus-Interpretation** bezeichnet; dabei ist aber wichtig, dass nur die **beobachteten** und im Modell mit einbezogenen Variablen tatsächlich fixiert werden.

Ein Beispiel für ein multivariates Modell

Wie sieht das in einem Beispiel aus?

$$\text{Lohn}_i = \beta_0 + \beta_1 \text{Bildung}_i + \beta_2 \text{Erfahrung}_i + u_i$$

In diesem Modell misst der Parameter

$$\beta_1 = \frac{\partial E(\text{Lohn}_i \mid \text{Bildung}_i, \text{Erfahrung}_i)}{\partial \text{Bildung}_i}$$

die **erwartete Änderung** des **Lohns**, wenn sich die **Bildung** um **eine Einheit ändert, und wir die Erfahrung gleich halten.**

- Wir modellieren nach wie vor den **Erwartungswert**. Es ist daher wichtig, dass wir von einer erwarteten oder durchschnittlichen Änderung sprechen.
- Wir interpretieren diesen partiellen Effekt als die Änderung, die wir beobachten, wenn wir **alle anderen (beobachteten) Faktoren gleich halten** (ceteris paribus; engl. holding all else fixed).

Ein komplexeres Modell

Wir können so viele Variablen hinzufügen, wie wir wollen:

$$\text{Lohn}_i = \beta_0 + \beta_1 \text{Bildung}_i + \beta_2 \text{Erfahrung}_i + \beta_3 \text{Alter}_i + \beta_4 \text{Karrierejahre}_i + \beta_5 \text{Gewerkschaft}_i + u_i$$

- **Gewerkschaft_i** ist eine Dummy-Variable, die Gewerkschaftszugehörigkeit angibt.
 - Warum fügen wir keine zusätzliche Variable **Nicht Gewerkschaft_i** hinzu?
 - Die Variablen wären direkt, mechanisch, invers miteinander verbunden. Es würde keinen Sinn machen, einen Effekt auf die beiden Variablen aufzuteilen; und es wäre mathematisch auch nicht möglich.
 - Wenn wir Dummy-Variablen zu verschiedenen Kategorien in unserem Modell haben, müssen wir immer eine Kategorie als Referenzkategorie weglassen.
- **Erfahrung_i** und **Karrierejahre_i** sind eng miteinander korreliert.
 - Die „ceteris-paribus“-Interpretation wird hier zunehmend schwierig. Wie kann man die Anzahl der Karrierejahre erhöhen, aber die Erfahrung unverändert halten?

Herleitung der OLS-Schätzer: Versuch 1

Versuchen wir, die OLS-Schätzer so herzuleiten, wie wir das im bivariaten Fall gemacht haben. Wir beginnen damit, eine Verlustfunktion aufzustellen:

$$\left(\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_K \right) = \arg \min_{\left(\tilde{\beta}_0, \tilde{\beta}_1, \dots, \tilde{\beta}_K \right)} \sum_{i=1}^N \left(y_i - \tilde{\beta}_0 - \tilde{\beta}_1 x_{i1} - \cdots - \tilde{\beta}_K x_{iK} \right)^2.$$

Wir können das ableiten und gleich Null setzen und erhalten ein System von Bedingungen erster Ordnung:

$$\begin{aligned} -2 \sum_{i=1}^N \left(y_i - \tilde{\beta}_0 - \tilde{\beta}_1 x_{i1} - \cdots - \tilde{\beta}_K x_{iK} \right) &= 0 \\ -2 \sum_{i=1}^N x_{i1} \left(y_i - \tilde{\beta}_0 - \tilde{\beta}_1 x_{i1} - \cdots - \tilde{\beta}_K x_{iK} \right) &= 0 \\ &\vdots \\ -2 \sum_{i=1}^N x_{iK} \left(y_i - \tilde{\beta}_0 - \tilde{\beta}_1 x_{i1} - \cdots - \tilde{\beta}_K x_{iK} \right) &= 0 \end{aligned}$$

Herleitung der OLS-Schätzer: Versuch 1

Wir können das ableiten und gleich Null setzen und erhalten ein System von Bedingungen erster Ordnung:

$$\begin{aligned} -2 \sum_{i=1}^N (y_i - \tilde{\beta}_0 - \tilde{\beta}_1 x_{i1} - \cdots - \tilde{\beta}_K x_{iK}) &= 0 \\ -2 \sum_{i=1}^N x_{i1} (y_i - \tilde{\beta}_0 - \tilde{\beta}_1 x_{i1} - \cdots - \tilde{\beta}_K x_{iK}) &= 0 \\ &\vdots \\ -2 \sum_{i=1}^N x_{iK} (y_i - \tilde{\beta}_0 - \tilde{\beta}_1 x_{i1} - \cdots - \tilde{\beta}_K x_{iK}) &= 0 \end{aligned}$$

Dieses Gleichungssystem ist lösbar, da es linear ist und $K + 1$ Gleichungen und $K + 1$ Variablen hat. Allerdings können wir **ohne Matrixalgebra keine Lösung** für $\hat{\beta}_k$ finden.

Interpretation der Bedingungen erster Ordnung

Wir können, wie im bivariaten Fall, diese Bedingungen erster Ordnung als Momentenbedingungen interpretieren:

- Die **Bedingung für die Konstante** $\hat{\beta}_0$ besagt:
 - Das Stichprobenmittel der OLS-Residuen muss 0 sein.
 - Entsprechende Momentenbedingung der Grundgesamtheit: $E(u_i) = 0$.
- Die **Bedingung für die Steigungsparameter** $\hat{\beta}_k, k = 1, \dots, K$ besagt:
 - Die Stichproben-Kovarianz zwischen Residuen und einem Regressor x_{ik} muss 0 sein.
 - Entsprechende Momentenbedingung der Grundgesamtheit: $\text{Cov}(x_{ik}, u_i) = E(x_{ik}u_i) = 0$.
- Wir haben also wieder **zwei Varianten**, um die **OLS-Schätzer** herzuleiten:
 - Als Bedingungen erster Ordnung aus einem Optimierungsproblem,
 - oder als Momentenbedingungen ausgehend von $E(u_i) = 0$ und $\text{Cov}(x_{ik}, u_i) = E(x_{ik}u_i) = 0$.

Einführung

Vektor- und Matrixnotation

Multivariate vs. bivariate Modelle

Praktisches Beispiel

OLS-Annahmen 1 bis 4

Vektornotation

Wenn wir **mehrere Variablen** haben, wird die im letzten Kapitel benutzte **Summenschreibweise** unhandlich. Wir benutzen daher **Vektoren** und **Matrizen**, um Modelle mit mehr als zwei Variablen anzuschreiben.

Sehen wir uns noch einmal dieses Modell an (etwas kompakter als vorher):

$$\text{Lohn}_i = \beta_0 \cdot 1 + \beta_1 \text{Bildung}_i + \beta_2 \text{Erfahrung}_i + \beta_3 \text{Alter}_i + u_i$$

Beachte, dass wir beim konstanten Parameter eine **1** hinzugefügt haben. Wenn wir diese **1** wie eine Variable betrachten, können wir β_0 wie die anderen Parameter betrachten und einen **Vektor aus Variablen** sowie einen **Vektor aus Parametern** anschreiben:

$$\boldsymbol{x}_i = \begin{pmatrix} 1 \\ \text{Bildung}_i \\ \text{Erfahrung}_i \\ \text{Alter}_i \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}$$

Vektornotation

$$\boldsymbol{x}_i = \begin{pmatrix} 1 \\ \text{Bildung}_i \\ \text{Erfahrung}_i \\ \text{Alter}_i \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}$$

Wir können den **Variablen-Vektor transponieren**, also Spalten und Zeilen vertauschen.
Wir markieren das mit einem kleinen Strich:

$$\boldsymbol{x}'_i = (1, \text{Bildung}_i, \text{Erfahrung}_i, \text{Alter}_i)$$

Jetzt können wir uns die Regeln der **Matrixmultiplikation** zunutze machen:

$$\boldsymbol{x}'_i \boldsymbol{\beta} = 1 \cdot \beta_0 + \text{Bildung}_i \beta_1 + \text{Erfahrung}_i \beta_2 + \text{Alter}_i \beta_3$$

Vektornotation

$$\mathbf{x}_i' \boldsymbol{\beta} = 1 \cdot \beta_0 + \text{Bildung}_i \beta_1 + \text{Erfahrung}_i \beta_2 + \text{Alter}_i \beta_3$$

Wir können unser Regressionsmodell also **extrem kompakt** schreiben, egal wie **viele Variablen** wir haben:

$$y_i = \mathbf{x}_i' \boldsymbol{\beta} + u_i.$$

Wir können auch das **OLS-Optimierungsproblem** so anschreiben:

$$\hat{\boldsymbol{\beta}} = \arg \min_{\tilde{\boldsymbol{\beta}}} \sum_{i=1}^N (y_i - \mathbf{x}_i' \boldsymbol{\beta})^2.$$

Die Lösung für dieses Problem ist:

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^N \mathbf{x}_i \mathbf{x}_i' \right)^{-1} \left(\sum_{i=1}^N \mathbf{x}_i y_i \right)$$



Übungsaufgabe

Löse dieses Optimierungsproblem!

Matrixnotation

Diese Gleichung beschreibt das Regressionsmodell für eine Beobachtung i .

$$y_i = \mathbf{x}'_i \boldsymbol{\beta} + u_i.$$

Wir können das Modell aber **noch kompakter** darstellen: mit **einer Gleichung für alle Beobachtungen**. Dazu definieren wir:

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & x_{11} & \dots & x_{1K} \\ 1 & x_{21} & \dots & x_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N1} & \dots & x_{NK} \end{pmatrix}, \quad \mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{pmatrix}.$$

Matrixnotation

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & x_{11} & \dots & x_{1K} \\ 1 & x_{21} & \dots & x_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N1} & \dots & x_{NK} \end{pmatrix}, \quad \mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{pmatrix}.$$

Unser Modell schaut jetzt so aus:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}.$$



Übungsaufgabe

Was sind die Dimensionen von \mathbf{y} , $\boldsymbol{\beta}$, \mathbf{X} , und \mathbf{u} ?

Was ist die Varianz eines Vektors?

\mathbf{u} ist der Vektor der Fehlerterme. Er ist also ein **Vektor aus Zufallsvariablen**, die alle den Mittelwert 0 und eine Varianz von σ^2 haben. **Was ist also die Varianz von \mathbf{u} ?**

Die **Varianz eines Vektors** ist eine Matrix. Die **diagonalen Elemente** dieser Matrix sind die Varianzen der einzelnen Elemente des Vektors. Die **Elemente abseits der Diagonale** sind die Kovarianzen zwischen den einzelnen Elementen.

Wir nennen so eine Matrix auch **Varianz-Kovarianz-Matrix** (engl. **variance-covariance matrix, VCM**):

$$\text{Var}(\mathbf{u}) = \begin{pmatrix} \text{Cov}(u_1, u_1) & \dots & \text{Cov}(u_1, u_N) \\ \vdots & \ddots & \vdots \\ \text{Cov}(u_N, u_1) & \dots & \text{Cov}(u_N, u_N) \end{pmatrix} = \begin{pmatrix} \text{Var}(u_1) & \dots & \text{Cov}(u_1, u_N) \\ \vdots & \ddots & \vdots \\ \text{Cov}(u_N, u_1) & \dots & \text{Var}(u_N) \end{pmatrix}$$

Herleitung der OLS-Schätzer: Versuch 2

In Matrixnotation ist die **Summe der Residuenquadrate**:

$$\hat{u}'\hat{u} = (\mathbf{y} - \mathbf{X}\tilde{\beta})'(\mathbf{y} - \mathbf{X}\tilde{\beta}),$$

und das OLS-Optimierungsproblem ist:

$$\hat{\beta} = \arg \min_{\tilde{\beta}} \hat{u}'\hat{u}.$$

Wenn wir dieses Problem lösen, erhalten wir den Schätzer

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y},$$

wir müssen aber Matrizen ableiten. Um das zu vermeiden, verwenden wir die **Momentenmethode**.

Herleitung der OLS-Schätzer: Versuch 2

In Matrixschreibweise haben wir **eine Momentenbedingung** für unser Modell:

$$E(\mathbf{X}' \mathbf{u}) = 0.$$

Wenn unser \mathbf{X} wie vorher definiert eine 1er-Spalte hat, impliziert diese Bedingung nämlich auch, dass $E(\mathbf{u}) = 0$.

Wir beginnen wieder damit, die **Momente der Grundgesamtheit** durch **Stichprobenmomente** zu ersetzen. Aus $E(\mathbf{X}' \mathbf{u}) = 0$ wird also:

$$\mathbf{X}' \hat{\mathbf{u}} = 0.$$

Wir können unseren **OLS-Schätzer** jetzt sehr einfach herleiten:

Herleitung der OLS-Schätzer: Versuch 2

Wir können unseren **OLS-Schätzer** jetzt sehr einfach herleiten:

$$\mathbf{X}'\hat{\mathbf{u}} = \mathbf{X}'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = 0$$

$$\mathbf{X}'\mathbf{y} - \mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = 0$$

$$\mathbf{X}'\mathbf{y} = \mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}$$

$$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}$$

$$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \hat{\boldsymbol{\beta}}$$

Wir erhalten denselben Schätzer wie bei der Herleitung mittels Optimierungsproblem,

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}.$$

Ein Schätzer, drei Schreibweisen

Wir haben unseren OLS-Schätzer in drei verschiedenen Schreibweisen kennen gelernt:

In Summenschreibweise (für den bivariaten Fall):

$$\hat{\beta}_1 = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^N (x_i - \bar{x})^2},$$

in Vektorschreibweise:

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^N \mathbf{x}_i \mathbf{x}_i' \right)^{-1} \left(\sum_{i=1}^N \mathbf{x}_i y_i \right),$$

und in Matrixschreibweise:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{y}.$$

Einführung Vektor- und Matrixnotation

Multivariate vs. bivariate Modelle

Praktisches Beispiel

OLS-Annahmen 1 bis 4

Erwartungswert des OLS-Schätzers

Residuen und vorhergesagte Werte

Genau wie bei **einfacher linearer Regression** teilen wir die beobachteten y -Werte bei **multipler linearer Regression** auf in einen

- **erklärten Teil** \hat{y} (die vorhergesagten Werte), und
- **nicht erklärten Teil** \hat{u} (die Residuen):

$$y_i = \underbrace{\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_K x_{iK}}_{\hat{y}_i} + \hat{u}_i.$$

- Das **Stichprobenmittel** der Residuen ist 0.
- Das **Stichprobenmittel** der **vorhergesagten Werte** ist demnach \bar{y} .
- Bei **einfacher linearer Regression** befindet sich der Punkt (\bar{x}, \bar{y}) auf der **Regressionsgerade**. Bei **multipler linearer Regression** befindet sich der Punkt $(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_K, \bar{y})$ auf der **Regressions-Hyperebene** (im Prinzip eine Gerade in $K + 1$ Dimensionen. Bei 3 Dimensionen wäre das eine Fläche.)

Regressionskoeffizienten

Wenn wir das **einfache lineare Regressionsmodell**

$$y_i = \beta_0^* + \beta_1^* x_{i1} + u_i$$

und das **multiple lineare Regressionsmodell**

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_K x_{iK} + u_i$$

schätzen, dann werden die Schätzungen für $\hat{\beta}_1^*$ und $\hat{\beta}_1$ **grundsätzlich nicht übereinstimmen**. Nur in zwei **Spezialfällen** wären die Schätzungen gleich:

- Wenn $\text{Cov}(x_{i1}, x_{ik}) = 0$ für alle $k \neq 1$ (was selten ist, Variablen sind meistens zumindest ein bisschen korreliert), und
- wenn $\beta_k = 0$ für alle $k \notin \{0, 1\}$ (also wenn alle anderen Variablen nicht relevant sind).

Wann sind Regressionskoeffizienten gleich?

Wir können diese Eigenschaft für den einfachen Fall von einem bzw. zwei Regressoren genauer ansehen. Wir betrachten das **einfache lineare Regressionsmodell**

$$y_i = \beta_0^* + \beta_1^* x_{i1} + u_i$$

und das **multiple lineare Regressionsmodell** mit zwei Regressoren

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i.$$

Hier kann man zeigen, dass

$$\hat{\beta}_1^* = \hat{\beta}_1 + \hat{\beta}_2 \hat{\delta} \quad \Rightarrow \quad E(\hat{\beta}_1^*) = \beta_1^* = \beta_1 + \beta_2 \delta,$$

wobei δ den Steigungsparameter aus einer **Regression von x_2 auf x_1 bezeichnet**. Wir sehen: $\hat{\beta}_1^*$ und $\hat{\beta}_1$ sind nur dann gleich, wenn $\hat{\beta}_2$ oder $\hat{\delta}$ gleich 0 ist.

Korrekte und nicht korrekte Modelle, zwei Arten von Bias

- Wenn das **Modell mit zwei Regressoren** das „**korrekte Modell**“ ist, wir aber **stattdessen** das **Modell mit einem Regressor** schätzen:
 - Der korrekte Parameter wäre β_1 ,
 - wir schätzen aber $\beta_1^* = \beta_1 + \beta_2\delta$.
 - Das nennen wir Verzerrung durch ausgelassene Variablen (engl. **omitted variable bias, OVB**).
- Wenn das **Modell mit einem Regressor** das „**korrekte Modell**“ ist, wir aber **stattdessen** das **Modell mit zwei Regressoren** schätzen:
 - Der korrekte Parameter wäre β_1^* ,
 - wir schätzen aber $\beta_1 = \beta_1^* - \beta_2\delta$
 - Das nennen wir **collider bias**.

Wir **wissen nicht**, ob das eine oder andere Modell „korrekt“ ist. **Insbesondere muss das größere Modell nicht automatisch besser als das kleinere Modell sein.** Wir müssen uns überlegen, welches Modell besser zu unseren Annahmen passt.

Anpassungsgüte im multivariaten Modell

Genau wie im bivariaten Fall können wir die **Variation in y** in einen **erklärten Teil**, also **Variation, die von Variation in x ausgeht**; und in einen **nicht erklärten Teil**, also einen Teil, der von **unbeobachteten Faktoren** ausgeht, aufteilen:

$$\sum_{i=1}^N (y_i - \bar{y})^2 = \sum_{i=1}^N (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^N \hat{u}_i^2$$
$$\text{SST} = \text{SSE} + \text{SSR}$$

Und genau wie im bivariaten Fall ist das **Bestimmtheitsmaß R^2** (engl. **coefficient of determination**) eine Maßzahl zur **Anpassungsgüte** (engl. **goodness of fit**), gibt an, welcher Anteil der Variation durch unser Modell erklärt wird, und hat genau dieselben Probleme, die wir auch im letzten Kapitel besprochen haben. Außerdem wird das R^2 immer ansteigen, wenn wir Variablen hinzufügen.

$$R^2 = \frac{\text{SSE}}{\text{SST}} = 1 - \frac{\text{SSR}}{\text{SST}}.$$

Einführung
Vektor- und Matrixnotation
Multivariate vs. bivariate Modelle
Praktisches Beispiel
OLS-Annahmen 1 bis 4
Erwartungswert des OLS-Schätzers
Varianz des OLS-Schätzers

Juhu, echte Daten

Schauen wir uns das, was wir theoretisch besprochen haben, einmal in einem Anwendungsbeispiel an. Wir beginnen damit, den Datensatz zu laden und die Variablen, die wir benötigen, zu selektieren:

R Code [Start Over](#) [Run Code](#)

```
1 # Pakete laden
2 library(AER) # Enthält unseren Datensatz
3 library(dplyr) # Enthält select()
4
5 # Daten laden
6 data("CPSSW8") # Datensatz mit Stundenlöhnen
7
8 # Nur die Variablen behalten, die uns interessieren
9 CPSSW8 <- CPSSW8 |>
10   select(earnings, gender, age, education)
```



Übungsaufgabe

Wähle andere Variablen aus und rechne Regressionen so wie auf den nächsten Folien.

Juhu, echte Daten

Als nächstes können wir uns eine Zusammenfassung unserer Variablen ansehen.

R Code

⟳ Start Over

▷ Run Code

```
1 # Zusammenfassung unserer Variablen:  
2 summary(CPSSW8)
```

Eine Regression!

R Code

↻ Start Over

▷ Run Code

```
1 lm(log(earnings) ~ education, data = CPSSW8) |>  
2 summary()
```

Der **Steigungskoeffizient** ist 0.094.

- Da wir ein **Log-Level-Modell** haben, bedeutet das: für Leute mit *einem zusätzlichen Jahr Bildung erwarten wir einen um 9.4 Prozent höheren Lohn.*

Die **Konstante** ist 1.485.

- Wir schätzen, dass Personen *ohne Bildung im Durchschnitt* ein logarithmiertes Einkommen von 1.485 haben.
- Das entspricht $\exp(1.485) = 4.41\$$. Das ist allerdings **nicht** der durchschnittliche Lohn dieser Gruppe, da $E(\log(\cdot)) \neq \log(E(\cdot))$.

Was passiert, wenn wir zusätzliche Variablen hinzufügen?

Mehr Variablen

R Code

⟳ Start Over

▷ Run Code

```
1 lm(log(earnings) ~ education + gender + age, data  
2   summary())
```

Der **Steigungskoeffizient** für `education` ist 0.094.

- Das bedeutet: für Leute mit einem zusätzlichen Jahr Bildung erwarten wir einen um 9.4 Prozent höheren Lohn, **nachdem wir für Geschlecht und Alter kontrolliert haben.**
- Der Koeffizient ist zwar in diesem Fall **sehr nahe** an dem aus dem vorherigen Modell, aber **nicht gleich**.

Der Koeffizient für `gender==female` ist -0.234.

- Frauen haben in diesem Datensatz im Mittel einen um 23 Prozent geringeren Lohn, **nachdem wir für Bildung und Alter kontrolliert haben.**

Der Koeffizient für `age` ist 0.0089.

- Personen, die ein Jahr älter waren, haben im Mittel um 0.89 Prozent höhere Löhne, **nachdem wir für Bildung und Geschlecht kontrolliert haben.**

Die Konstante ist 1.22.

Verschiedene Darstellungen von Resultaten

Wenn wir eine Regression in R rechnen und die Funktion `summary()` anwenden, bekommen wir **folgenden Output:**

```
Call:  
lm(formula = log(earnings) ~ education + gender + age, data =  
CPSSW8)  
  
Residuals:  
    Min      1Q  Median      3Q     Max  
-2.79472 -0.28807  0.02562  0.32439  1.63195  
  
Coefficients:  
            Estimate Std. Error t value Pr(>|t|)  
(Intercept) 1.2206890  0.0130145   93.80 <2e-16 ***  
education    0.0941281  0.0007922  118.82 <2e-16 ***  
genderfemale -0.2338747  0.0039207  -59.65 <2e-16 ***  
age          0.0088690  0.0001839   48.22 <2e-16 ***  
---  
Signif. codes:  0 '****' 0.001 '***' 0.01 '**' 0.05 '*' 0.1 '.' 1  
  
Residual standard error: 0.4816 on 61391 degrees of freedom  
Multiple R-squared:  0.2452,    Adjusted R-squared:  0.2452  
F-statistic: 6649 on 3 and 61391 DF,  p-value: < 2.2e-16
```

In einem **Paper** oder einer **Bachelorarbeit** präsentiert man die Resultate gewöhnlicherweise in so einer Tabelle:

Dependent variable: log(earnings)		
	(1)	(2)
education	0.094*** (0.001)	0.094*** (0.001)
genderfemale		-0.234*** (0.003)
age		0.009*** (0.0002)
Constant	1.485*** (0.011)	1.221*** (0.013)
Observations	61,395	61,395
R ²	0.174	0.245
Note: *p<0.1; **p<0.05; ***p<0.01		
Numbers in parentheses are standard errors.		

Vektor- und Matrixnotation

Multivariate vs. bivariate Modelle

Praktisches Beispiel

OLS-Annahmen 1 bis 4

Erwartungswert des OLS-Schätzers

Varianz des OLS-Schätzers

Frisch-Waugh-Lovell-Theorem

Von den SLR- zu den MLR-Annahmen

Um Aussagen über **Erwartungswert** und **Varianz** treffen zu können, brauchen wir wieder eine Reihe von **Annahmen**. Diese Annahmen **MLR.1 bis MLR.4** sind generalisierte Versionen der Annahmen **SLR.1 bis SLR.4** aus dem vorigen Kapitel.

Satz von Gauß-Markow: Annahmen für Multiple Lineare Regression (MLR)

- (1) Linearität in Parametern
- (2) Zufallsstichprobe
- (3) ***Keine perfekte Multikollinearität***
- (4) Exogener Fehlerterm

(MLR.1) Linearität in Parametern

Die Regressionsfunktion der Grundgesamtheit (PRF) muss **linear** in ihren **Parametern** sein:

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_K x_{iK} + u_i$$

- Transformationen (z.B. logarithmische) sind natürlich auch hier kein Problem, da die PRF trotzdem eine lineare Kombination **der Parameter** bleibt.
- In Matrixnotation können wir dieses Modell anschreiben als:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}.$$

(MLR.2) Zufallsstichprobe

Unsere **Stichprobe** mit N Beobachtungen, $\{(y_i, x_{i1}, \dots, x_{iK}), i = 1, 2, \dots, N\}$ muss **zufällig** aus der Grundgesamtheit gezogen werden. Die Wahrscheinlichkeit, eine Beobachtung in die Stichprobe aufzunehmen, muss für alle gleich sein, und darf nicht davon abhängen, wen wir zuerst „gezogen“ haben.

- Wenn diese Annahme erfüllt ist, können wir Beobachtungen und Fehlerterme als unabhängig voneinander ansehen.

(MLR.3) Keine perfekte Multikollinearität

Im bivariaten Fall haben wir **Variation in den x -Werten** an dieser Stelle angenommen. Hier, im multivariaten Fall, benötigen wir eine weiterreichende Annahme: **kein Regressor darf eine Linearkombination aus anderen Regressoren sein**. Formell können wir sagen:

Die Matrix der Regressoren \mathbf{X} enthält **keine Spalte, die eine Linearkombination anderer Spalten ist**. \mathbf{X} hat also vollen Rang.

- Wir können die SLR-Version der Annahme auch in diesem Kontext betrachten, wenn wir uns eine Matrix mit zwei Spalten, einer konstanten 1er-Spalte und einer Spalte mit x -Werten ohne Variation vorstellen. Dann ist die zweite Spalte eine Linearkombination der ersten Spalte.
- Im SLR-Fall haben wir diese Annahme gebraucht, um durch $\sum_{i=1}^N (x_i - \bar{x})^2$ **dividieren** zu können. Jetzt benötigen wir diese Annahme, damit $(\mathbf{X}' \mathbf{X})$ **invertierbar** ist.
- Sowohl im SLR- als auch im MLR-Fall brauchen wir diese Annahme, da wir den Schätzer sonst mathematisch **nicht berechnen können**.

Wann ist MLR.3 verletzt?

- Die Annahme MLR.3 ist **verletzt**, wenn ein Regressor eine (gewichtete) Summe oder Differenz anderer Regressoren ist.
 - **Beispiel:** Wir untersuchen die Beziehung zwischen Einkommen der Eltern und eigenem Einkommen, indem wir das einkommen einer Person auf das Einkommen beider Elternteile (x_{i1} und x_{i2}) regressieren. Wenn wir die Summe des Einkommens der Eltern als x_{i3} hinzufügen, ist die Annahme verletzt. Wenn wir den Unterschied im Einkommen der beiden Elternteile als x_{i3} hinzufügen, ist die Annahme auch verletzt.
 - Bei **Dummy-Variablen** ist diese Annahme verletzt, wenn wir alle möglichen Kategorien in die Regression mit einbeziehen. Wenn wir z.B. Daten über Personen in allen 27 EU-Ländern haben und einen Länder-Dummy mit einbeziehen wollen, müssen wir ein Land als Referenzkategorie aussparen.

Wann ist MLR.3 nicht verletzt?

- Die Annahme MLR.3 ist **nicht verletzt**, wenn ein Regressor eine **nicht-lineare** Kombination anderer Regressionen ist.
 - **Beispiel:** Wir inkludieren das **quadrierte** Einkommen der einzelnen Elternteile: $x_{i3} = x_{i1}^2$ und $x_{i4} = x_{i2}^2$. Auf diese Weise können wir parabelförmige Beziehungen modellieren.
 - **Beispiel:** Wir inkludieren $x_{i3} = x_{i1} \times x_{i2}$. Das nennen wir einen **Interaktionseffekt**. Damit können wir Effekte, die voneinander abhängen, modellieren – es macht aber die Interpretation schwieriger (da eine einfache „Ceteris-Paribus-Interpretation“ nicht mehr möglich ist).

Wir besprechen **quadrierte Regressoren** und **Interaktionsterme** in einem späteren Modul detaillierter. An diesem Punkt ist nur wichtig, zu wissen, dass es sie gibt, und dass sie nicht zu einer Verletzung von MLR.3 führen.

Wann ist MLR.3 auch nicht verletzt?

- Die Annahme MLR.3 ist auch **nicht verletzt**, wenn zwei Regressoren (stark) miteinander korreliert sind.
 - Je **stärker** zwei Regressoren **korreliert** sind, desto **weniger präzise** werden unsere OLS-Schätzungen.
 - Sie sind aber immer noch **BLUE**, da wir nirgends voraussetzen, dass Regressoren nicht korreliert sind. Korrelation zwischen Regressionen verletzt keine der Gauß-Markov-Annahmen, **solange sie nicht perfekt ist**.
 - **Perfekte Korrelation** wird von Statistiksoftware automatisch erkannt, **nahezu perfekte Korrelation** nicht. Daher ist sie als Problem schwieriger zu erkennen. Ob zwei Variablen „zu stark“ miteinander korrelieren, muss in jedem Fall individuell beurteilt werden.

(MLR.4) Exogene Fehler

Der Erwartungswert des Fehlerterms u ist für **jeden Regressor** 0:

$$E(u_i | x_{i1}, \dots, x_{iK}) = 0$$

In Matrixnotation (diese Annahme ist noch etwas stärker, da sie nicht nur Regressoren, sondern auch deren Linearkombinationen umfasst):

$$E(\mathbf{u} | \mathbf{X}) = \mathbf{0}$$

- Wie im SLR-Fall impliziert diese Annahme:
 - $E(u_i) = 0$, und
 - $\text{Cov}(x_{ik}, u_i) = 0$ und somit $E(x_{ik}u_i) = 0$ für alle Regressoren x_{ik} .

Wann ist MLR.4 verletzt?

Wir nehmen an, dass **Regressoren** und **unbeobachtete Faktoren** unabhängig voneinander sind. Das ist in Experimenten einfach zu erreichen, bei Beobachtungsdaten aber weit weniger trivial. Wir nennen den Fall, in dem MLR.4 verletzt ist, **Endogenität**.

Wir nennen in dem Fall, dass $E(x_{ik}u_i) \neq 0$, x_{ik} einen **endogenen Regressor**. Das kann der Fall sein, wenn:

- wir eine **Variable, die mit Regressoren korreliert und relevant in der Erklärung** unserer abhängigen Variable ist, **auslassen**. Dann haben wir **omitted variable bias**.
 - **Beispiel:** Wir regressieren Lohn auf Ausbildung und berücksichtigen Talent/Motivation nicht.
- die **abhängige Variable** ihrerseits Einfluss auf einen **Regressor** hat. Diesen Fall nennen wir **Rückwärtskausalität** (engl. **reverse causality**).
 - **Beispiel:** In der Covid-Pandemie hatte verstärktes Maskentragen einen Einfluss auf die Anzahl der Infektionen, aber die Anzahl der Infektionen hatte gleichzeitig einen Einfluss darauf, wie viele Leute eine Maske tragen.
- die wahre Beziehung **nicht-linear** ist.

Multivariate vs. bivariate Modelle

Praktisches Beispiel

OLS-Annahmen 1 bis 4

Erwartungswert des OLS-Schätzers

Varianz des OLS-Schätzers

Frisch-Waugh-Lovell-Theorem

Wie viele Variablen?

OLS ist unverzerrt

Wenn die vier Annahmen MLR.1 bis MLR.4 erfüllt sind, können wir **beweisen**, dass der OLS-Schätzer **unverzerrt** (engl. **unbiased**) ist. Formell:

Unter den Annahmen MLR.1 bis MLR.4 gilt: $E(\hat{\beta}_k) = \beta_k, \quad k = 0, 1, \dots, K,$
für jeden Wert der Parameter β_j . In Matrixnotation:

$$E(\hat{\beta}) = \beta,$$

wobei β die Dimension $(K + 1) \times 1$ hat.

Der OLS-Schätzer ist also ein **unverzerrter**/erwartungstreuer (engl. unbiased) **Schätzer** für die **Konstante** und alle **Steigungsparameter**. Wir können das wieder beweisen, indem wir den Schätzer in den wahren Koeffizienten und eine Stichprobenfehler-Komponente aufteilen.

Beweis: OLS ist unverzerrt

Wir beginnen mit der Aufteilung von $\hat{\beta}$:

$$\begin{aligned}\hat{\beta} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\beta + \mathbf{u}) \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} \\ &= \underbrace{\beta}_{\text{wahrer Parameter}} + \underbrace{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}}_{\text{Stichprobenfehler}}.\end{aligned}$$

Wir können diesen Schritt vergleichen mit dem Schritt im Beweis für den SLR-Fall, in dem wir $\hat{\beta}_1$ so aufgeteilt haben:

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^N (x_i - \bar{x})u_i}{\sum_{i=1}^N (x_i - \bar{x})x_i}.$$

Beweis: OLS ist unverzerrt

Mit dieser Aufteilung können wir im Beweis fortfahren:

$$\begin{aligned} E(\hat{\beta} | X) &= E(\beta + (X'X)^{-1}X'u | X) \\ &= \beta + E((X'X)^{-1}X'u | X) \\ &= \beta + (X'X)^{-1}X' \underbrace{E(u | X)}_{=0 \text{ (MLR.4)}} \\ &= \beta. \end{aligned}$$

Da $E(\hat{\beta} | X) = \beta \Rightarrow E(\hat{\beta}) = \beta$ (Satz der iterierten Erwartungen), ist der **OLS-Schätzer unverzerrt.**

Praktisches Beispiel
OLS-Annahmen 1 bis 4
Erwartungswert des OLS-Schätzers

Varianz des OLS-Schätzers

Frisch-Waugh-Lovell-Theorem
Wie viele Variablen?
Visualisierung

(MLR.5) Homoskedastizität

Die Varianz des Fehlerterms u_i ist für alle x_{ik} gleich:

$$\text{Var}(u_i \mid x_{i1}, \dots, x_{iK}) = \text{Var}(u_i) = \sigma^2,$$

bzw. in Matrixschreibweise:

$$\text{Var}(\mathbf{u} \mid \mathbf{X}) = \sigma^2 \mathbf{I}_N,$$

wobei \mathbf{I}_N die Identitätsmatrix mit Dimension $n \times n$ ist.

- Die Interpretation dieser Annahme ist analog zur entsprechenden SLR-Annahme.
- Die Annahme ist z.B. verletzt, wenn Personen mit mehr Bildung eine höhere Varianz im Einkommen haben. Solche Verletzungen kommen entsprechend häufig vor.

Varianz des OLS-Schätzers

Unter den Annahmen **MLR.1 bis MLR.5** ist die Varianz des OLS-Schätzers

$$\text{Var} \left(\hat{\boldsymbol{\beta}} \right) = \sigma^2 (\mathbf{X}' \mathbf{X})^{-1},$$

wobei sich $\text{Var}(\cdot)$ hier auf eine Varianz-Kovarianz-Matrix bezieht.



Übungsaufgabe

Zeige, wie man zu diesem Ausdruck für die Varianz kommt. In welchem Schritt braucht man dazu welche der Annahmen MLR.1 bis MLR.5? • Herleitung

Mehr zur Varianz des OLS-Schätzers

Sehen wir uns das bivariate Modell in Matrixnotation an:

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix}, \quad \text{Var}(\hat{\beta}) = \begin{pmatrix} \text{Var}(\hat{\beta}_0) & \text{Cov}(\hat{\beta}_0, \hat{\beta}_1) \\ \text{Cov}(\hat{\beta}_1, \hat{\beta}_0) & \text{Var}(\hat{\beta}_1) \end{pmatrix}$$

Wir haben bisher immer nur die Stichproben-**Varianz** eines Schätzers besprochen, aber nicht die Stichproben-**Kovarianz**. Statistiksoftware schätzt für gewöhnlich nur die Varianzen der Parameter, und nicht die Kovarianzen, also nur die **Diagonale** der Varianz-Kovarianz-Matrix. Wir benötigen die Kovarianz später für bestimmte statistische Tests.

Varianz einzelner Koeffizienten

Analog zur expliziten Formel für die Varianz im bivariaten Fall können wir aus der vorherigen Formel für die Varianz folgende Formel für die **Varianz eines einzelnen Koeffizienten** herleiten:

$$\text{Var}(\hat{\beta}_k | \mathbf{X}) = \frac{\sigma^2}{\sum_{i=1}^N (x_{ik} - \bar{x}_k)^2} \times \frac{1}{1 - R_k^2},$$

wobei R_k^2 das R^2 einer Regression von x_k auf alle anderen Regressoren $x_j, j \neq k$ ist.

- Wir sehen, dass ein **großes** σ^2 die Varianz **erhöht** (weniger präzise Schätzung),
- dass ein **großes** N die Varianz **verringert** (präzisere Schätzung),
- und dass **starke Variation** im Regressor x_k und **schwache Korrelation** von x_k mit $x_j, j \neq k$ die Varianz **verringern** (präzisere Schätzung).
- Stark korrelierte Regressoren machen unsere Schätzungen also weniger präzise.

Ein Schätzer für die Varianz

Genau wie im bivariaten Fall **kennen** wir aber die **Varianz σ^2 nicht**, sondern benötigen einen Schätzer.

Man kann zeigen (wir lassen den Beweis aus), dass der folgende Schätzer:

$$E\left(\frac{\sum_{i=1}^N \hat{u}_i^2}{N - K - 1}\right) = E\left(\frac{\hat{u}'\hat{u}}{N - K - 1}\right) = E(\hat{\sigma}^2) = \sigma^2$$

unter den **Annahmen MLR.1 bis MLR.5** ein **unverzerrter Schätzer** der Varianz des Fehlers ist.

Wir dividieren durch $N - K - 1$ (und nicht durch N), um für die Freiheitsgrade zu korrigieren: Unsere Schätzung ergibt aus N Beobachtungen $K - 1$ Koeffizienten, also bleiben $N - K - 1$ Freiheitsgrade. Wir haben dieselbe Korrektur im bivariaten Fall vorgenommen.

Satz von Gauß-Markov

Jetzt können wir, analog zum bivariaten Fall, den **Satz von Gauß-Markov** für den multivariaten Fall formulieren:

Unter den **Annahmen MLR.1 bis MLR.5** ist der **OLS-Schätzer**

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_K \end{pmatrix}$$

der **beste lineare unverzerrte Schätzer** (engl. **best linear unbiased estimator, BLUE**)
der Parameter $\beta = (\beta_0, \beta_1, \dots, \beta_K)'$.

OLS-Annahmen 1 bis 4

Erwartungswert des OLS-Schätzers

Varianz des OLS-Schätzers

Frisch-Waugh-Lovell-Theorem

Wie viele Variablen?

Visualisierung

Appendix

Mehr Ceteris Paribus

Es ist intuitiv nicht leicht, zu verstehen, was die **Koeffizienten in einem multivariaten Modell** tatsächlich messen. Das **Frisch-Waugh-Lovell-Theorem** gibt uns eine zusätzliche Herangehensweise.

Wir betrachten folgendes Modell:

$$y_i = x_{i1}\beta_1 + \mathbf{x}'_{i2}\boldsymbol{\beta}_2 + u_i, \quad E\left(\begin{pmatrix} x_{i1} \\ \mathbf{x}_{i2} \end{pmatrix} u_i\right) = 0.$$

Wir können annehmen, dass y_i der Lohn ist, x_{i1} Geschlecht, und \mathbf{x}_{i2} ein Vektor aus einer 1er-Spalte, Bildung, und dem Alter. Wir nehmen an, dass wir primär an β_1 interessiert sind und heben x_{i1} daher hervor und fassen den Rest des Modells in Vektorschreibweise zusammen.

Die Variablen, an denen wir nicht primär interessiert sind, nennen wir üblicherweise **Kontrollvariablen** (engl. **control variables** oder **controls**). Wir inkludieren sie, damit das Modell vollständig ist.

Frisch-Waugh-Lovell-Theorem

Wir beginnen damit, y_i **nur auf den Vektor \mathbf{x}_{i2}** (und nicht auf x_1) zu regressieren. Wir „behalten“ uns aus dieser Regression die **Vorhersagefehler** und bezeichnen sie als $\underline{y}_i^{(R)}$.

$$y_i = \mathbf{x}'_{i2} \boldsymbol{\alpha} + \underbrace{\underline{y}_i^{(R)}}_{\text{Fehler}}$$

Als nächstes regressieren wir unsere **Variable von Interesse**, x_{i1} , **auf den Vektor \mathbf{x}_{i2}** , „behalten“ uns auch hier die **Vorhersagefehler** und bezeichnen sie als $\underline{x}_{i1}^{(R)}$.

$$x_{i1} = \mathbf{x}'_{i2} \boldsymbol{\gamma} + \underbrace{\underline{x}_{i1}^{(R)}}_{\text{Fehler}}$$

Vereinfacht gesprochen haben wir jetzt eine „Version“ von y_i , bei der wir den Einfluss von \mathbf{x}_{i2} „herausgefiltert“ haben, und eine „Version“ von x_{i1} , bei der wir den Einfluss von \mathbf{x}_{i2} „herausgefiltert“ haben.

Frisch-Waugh-Lovell-Theorem

$$y_i = \mathbf{x}'_{i2} \boldsymbol{\alpha} + \underbrace{y_i^{(R)}}_{\text{Fehler}}$$

$$x_{i1} = \mathbf{x}'_{i2} \boldsymbol{\gamma} + \underbrace{x_{i1}^{(R)}}_{\text{Fehler}}$$

Interessanterweise können wir **denselben Parameter β_1 auf zwei unterschiedliche Weisen erhalten:**

- wenn wir y_i auf x_{i1} und \mathbf{x}'_{i2} regressieren (ursprüngliche Regression), und
- wenn wir $y_i^{(R)}$ auf $x_{i1}^{(R)}$ regressieren (FWL-Regression der Residuen der beiden Hilfs-Regressionen).

Frisch-Waugh-Lovell-Theorem in der Stichprobe

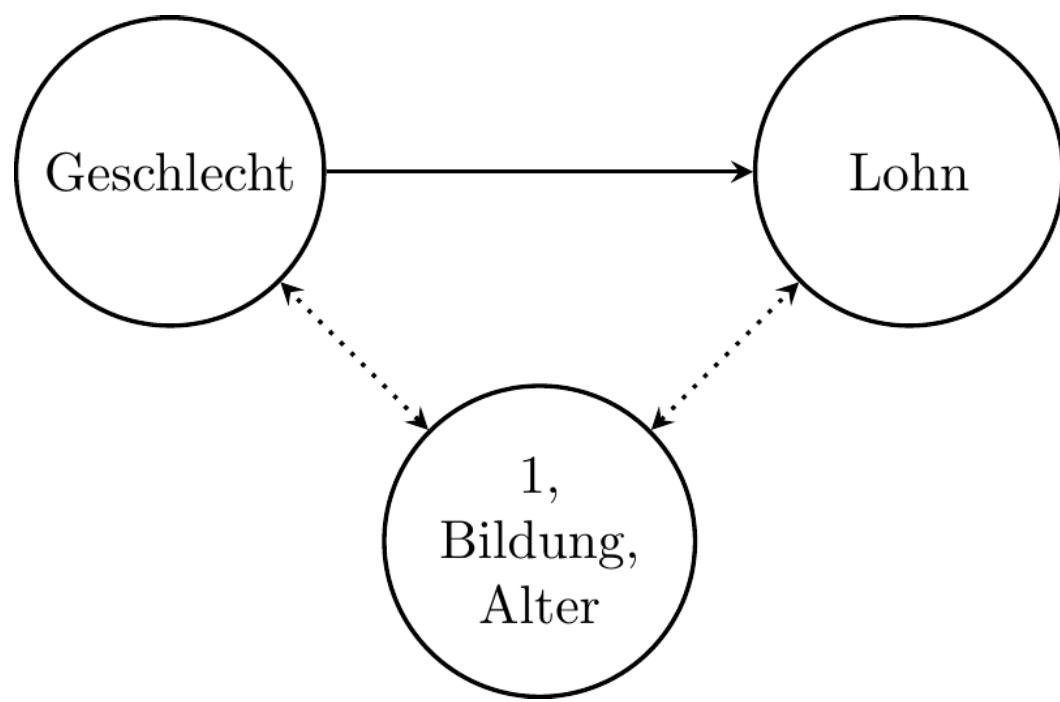
Wenn wir eine Stichprobe an Daten haben, können wir wie folgt vorgehen, um unseren Schätzer $\hat{\beta}_1$ auf diese Weise zu erhalten:

- (1) Wir regressieren y_i auf x'_{i2} und erhalten die **Residuen** $\hat{y}_i^{(R)}$.
- (2) Wir regressieren x_{i1} auf x'_{i2} und erhalten die **Residuen** $\hat{x}_{i1}^{(R)}$.
- (3) Wir regressieren $\hat{y}_i^{(R)}$ auf $\hat{x}_{i1}^{(R)}$ und erhalten den **OLS-Schätzer** $\hat{\beta}_1$. Dieser Schätzer entspricht dem Schätzer aus der ursprünglichen Regression.

Dieses Ergebnis bezeichnen wir als **Frisch-Waugh-Lovell-Theorem**, nach Frisch und Waugh (1933) sowie Lovell (1963). Es kann uns helfen, die Parameter des **multivariaten Modells intuitiv** zu verstehen.

Frisch-Waugh-Lovell und Interpretation

Wir können das vorherige Beispiel mit folgendem kausalen **Graphen** illustrieren:
Wir nehmen an, das Geschlecht hat einen Einfluss auf den Lohn, aber es gibt auch eine Korrelation zwischen den Variablen in x'_{i2} und sowohl Geschlecht als auch Lohn.



- Wir können den Fehler $y_i^{(R)}$ als die **Variation in y_i** interpretieren, die nicht durch x'_{i2} erklärt werden kann.
- Ebenso können wir den Fehler $\hat{x}_{i1}^{(R)}$ als die **Variation in x_{i1}** interpretieren, die nicht durch x'_{i2} erklärt werden kann.
- Wir „filtern“ also den strichlierten Effekt in der Grafik heraus.
- Wir können den Effekt von Interesse **nach dem „Herausfiltern“** durch eine einfache bivariate Regression finden.
- Daher interpretieren wir Effekte in multivariaten Regressionsmodellen als **Ceteris-Paribus-Effekte**.

FWL in der Praxis

R Code

[Start Over](#)

[Run Code](#)

```
1 # Pakete laden
2 library(AER) # Enthält unseren Datensatz
3 library(dplyr) # Enthält select()
4
5 # Daten laden
6 data("CPSSW8") # Datensatz mit Stundenlöhnen
7
8 # Nur die Variablen behalten, die uns interessieren
9 CPSSW8 <- CPSSW8 |>
10   select(earnings, gender, age, education)
11
12 CPSSW8$female <- CPSSW8$gender=="female"
```

FWL in der Praxis

R Code

[Start Over](#)

[Run Code](#)

```
1 # Ursprüngliches Modell schätzen
2 full_model <- lm(log(earnings) ~ female + education + age, data = CPSSW8)
3
4 # Residuen für y_i
5 log_earning_residuals <- lm(log(earnings) ~ education + age, data = CPSSW8) |>
6   residuals()
7
8 # Residuen für x_{i1}
9 female_residuals <- lm(female ~ education + age, data = CPSSW8) |>
10  residuals()
11
12 # FWL-Modell schätzen
13 partialled_out_model <- lm(log_earning_residuals ~ 0 + female_residuals, data = CPSSW8)
```

FWL in der Praxis

R Code

[⟳ Start Over](#)

[▷ Run Code](#)

```
1 # Output ursprüngliches Modell  
2 summary(full_model)
```

R Code

[⟳ Start Over](#)

[▷ Run Code](#)

```
1 # Output FWL-Modell  
2 summary(partialled_out_model)
```

Erwartungswert des OLS-Schätzers

Varianz des OLS-Schätzers

Frisch-Waugh-Lovell-Theorem

Wie viele Variablen?

Visualisierung

Appendix

Zu viele oder zu wenige Variablen?

Nachdem wir jetzt die Möglichkeit haben, **so viele Variablen, wie wir wollen**, in unserer Regression zu berücksichtigen, stellt sich die Frage:

- Wie viele Variablen sind **zu viele** Variablen?
- Wie viele Variablen sind **zu wenige** Variablen?

Natürlich gibt es **keine „Faustregel“** oder **allgemein gültige Antwort** auf diese Frage. Stattdessen müssen wir für jedes Modell und jede Variable **individuell entscheiden**, ob es sinnvoll ist, sie zu berücksichtigen.

Wenn wir **relevante Variablen auslassen**, bekommen wir ein **omitted variable bias**-Problem. In diesem Fall wird der Effekt, der eigentlich zu der ausgelassenen Variable gehört, inkorrekt erweise den im Modell enthaltenen Variablen zugeschrieben.

Omitted Variable Bias

Was passiert, wenn wir relevante **Variablen** in unserem Modell **auslassen**? Unser **Schätzer** wird **nicht länger unverzerrt** sein, und wir können das beweisen.

Angenommen, das ist das „wahre“ Modell. Wir haben die Regressoren „aufgeteilt“ in zwei Matrizen, im Prinzip ist das aber das gleiche Modell, das wir im Rahmen dieses Kapitels immer behandelt haben:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \mathbf{u}$$

Was passiert, wenn wir **stattdessen dieses Modell** schätzen?

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$$

Omitted Variable Bias

Wir teilen wieder $\hat{\beta}$ auf, allerdings verwenden wir für y das **wahre Modell**.

$$\begin{aligned}\hat{\beta} &= (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{y} \\ &= (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' (\mathbf{X} \beta + \mathbf{Z} \gamma + \mathbf{u}) \\ &= (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{X} \beta + (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{Z} \gamma + (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{u} \\ &= \beta + (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{Z} \gamma + (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{u}\end{aligned}$$

Wenn wir jetzt den **Erwartungswert** dieses Ausdrucks nehmen, sehen wir, dass der Schätzer **nicht mehr unverzerrt** ist.

$$\begin{aligned}\mathrm{E}(\hat{\beta} | \mathbf{X}) &= \mathrm{E}(\beta + (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{Z} \gamma + (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{u} | \mathbf{X}) \\ &= \beta + \mathrm{E}((\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{Z} \gamma | \mathbf{X}) + \mathrm{E}((\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{u} | \mathbf{X}) \\ &= \beta + \underbrace{(\mathbf{X}' \mathbf{X})^{-1} \mathrm{E}(\mathbf{X}' \mathbf{Z} | \mathbf{X}) \gamma}_{\text{Bias (Verzerrung)}} + \mathbf{0}\end{aligned}$$

Omitted Variable Bias

Wir können **sehr einfach** sehen, wovon die „Richtung“ des Bias abhängt:

$$E(\hat{\beta} | \mathbf{X}) = \boldsymbol{\beta} + \underbrace{(\mathbf{X}' \mathbf{X})^{-1} E(\mathbf{X}' \mathbf{Z} | \mathbf{X}) \gamma}_{\text{Bias (Verzerrung)}} + \mathbf{0}$$

	$E(\mathbf{X}' \mathbf{Z} \mathbf{X})$ positiv	$E(\mathbf{X}' \mathbf{Z} \mathbf{X})$ negativ
γ positiv	Positiver Bias	Negativer Bias
γ negativ	Negativer Bias	Positiver Bias

- $E(\mathbf{X}' \mathbf{Z} | \mathbf{X})$ sagt uns, ob die Variablen in \mathbf{X} mit den Variablen in \mathbf{Z} **korreliert** sind.
- Wir sehen, dass der **Bias nur dann Null** ist, wenn einer der Faktoren Null ist:
 - $E(\mathbf{X}' \mathbf{Z} | \mathbf{X}) = \mathbf{0}$: Die Variablen in \mathbf{X} sind nicht mit den Variablen in \mathbf{Z} korreliert.
 - $\gamma = 0$: Die Variablen in \mathbf{Z} waren von vorne herein nicht relevant in der Erklärung von y .

Wann zu viele Variablen ein Problem sind

Aber auch **zu viele Variablen** können ein Problem darstellen, insbesondere in den folgenden Fällen und aus den folgenden Gründen:

- Wenn wir zu viele Variablen haben und diese **stark korreliert** sind, werden unsere Schätzungen **weniger präzise**.
- Wenn wir **mehr Parameter als Beobachtungen haben** ($N < K$), ist **Annahme MLR.3** verletzt und wir können keine Schätzung berechnen.
- Unnötige Variablen können auch zu einer Verletzung von MLR.4 führen.

Wann zu viele Variablen ein Problem sind

In folgendem Beispiel wird durch das Hinzufügen einer zusätzlichen Variable MLR.4 verletzt:

- Angenommen, wir wollen den **Effekt von Dünger auf landwirtschaftliche Erträge** herausfinden und führen ein Experiment durch, bei dem wir die **Düngernutzung korrekt randomisieren**.
 - Wenn wir das Modell $\text{Erträge}_i = \beta_0 + \beta_1 \text{Dünger}_i + u_i$ schätzen, ist es gerechtfertigt, anzunehmen, dass $E(u_i | \text{Dünger}_i) = 0$, da wir Düngernutzung ja randomisiert haben.
 - Wenn wir aber $\text{Erträge}_i = \beta_0 + \beta_1 \text{Dünger}_i + \beta_2 \text{Unkraut} + u_i$ schätzen, ist die Annahme $E(u_i | \text{Dünger}_i, \text{Unkraut}_i) = 0$ **wahrscheinlich nicht mehr erfüllt**, da Unkraut nicht randomisiert ist und wahrscheinlich mit unbeobachteten Faktoren korreliert.
 - **Vorsicht:** Wenn wir R^2 zur **Modellauswahl verwenden würden**, würden wir **das „falsche“ Modell** auswählen, da R^2 bei Hinzufügen einer weiteren Variable niemals sinkt.

Varianz des OLS-Schätzers
Frisch-Waugh-Lovell-Theorem
Wie viele Variablen?

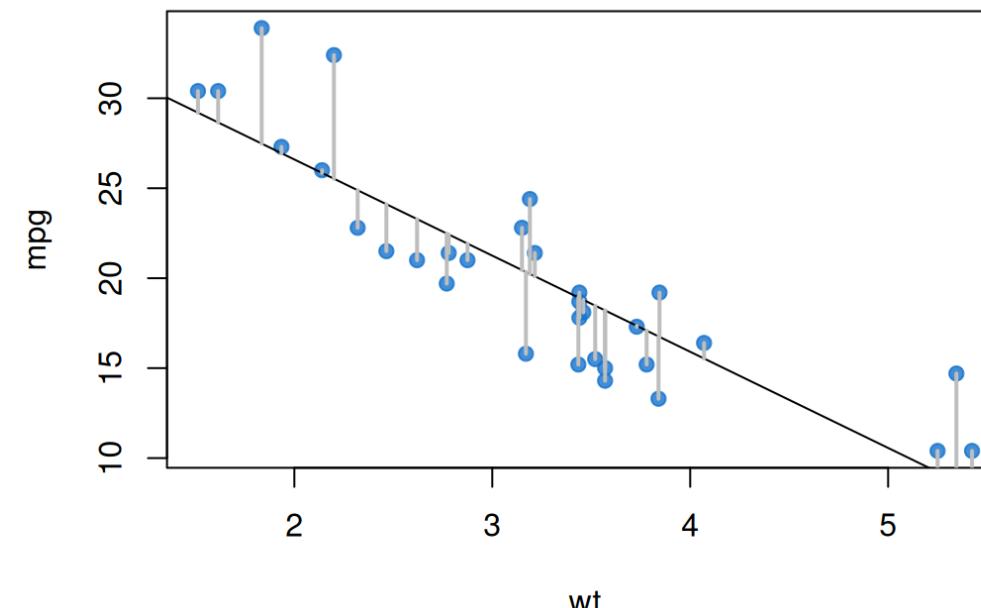
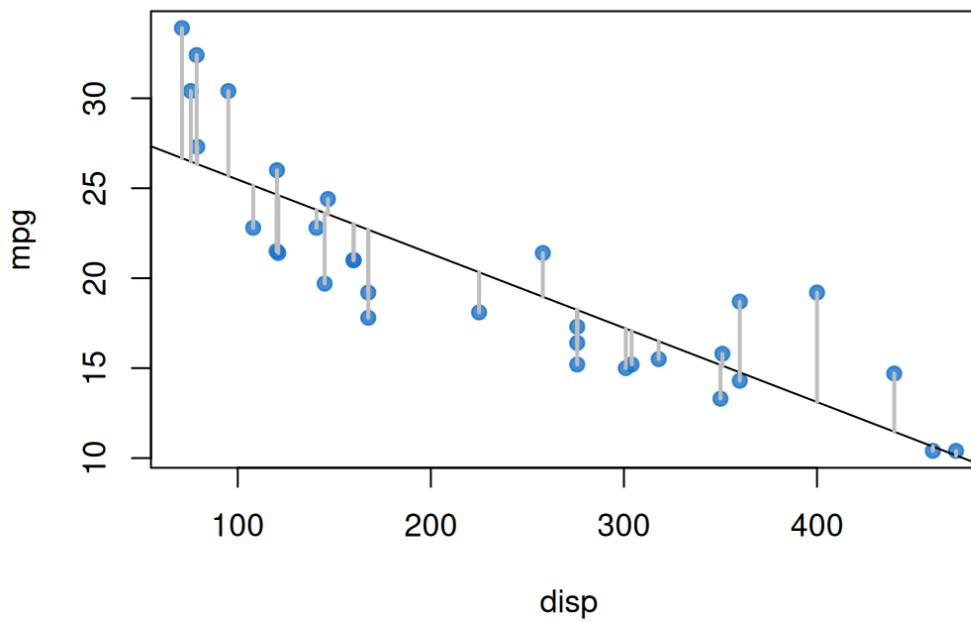
Visualisierung

Appendix

Beispiel 1: Autos

- Der Datensatz `mtcars` enthält 32 **Automodelle** (1973–74) und ihr(e)
- Effizienz in Meilen pro Gallone Treibstoff (`mpg`),
- Gewicht (`wt`), Hubraum (`disp`), etc.

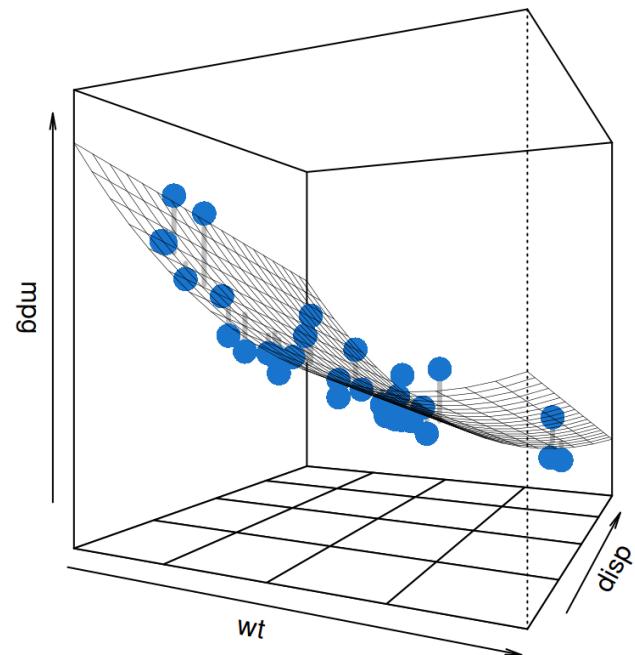
Beginnen wir mit einer **einfachen linearen Regression**:



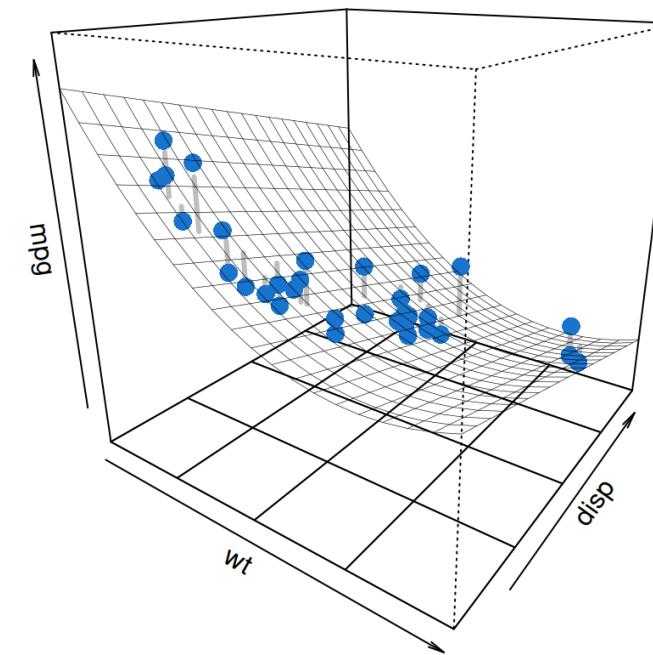
Multivariate Regression mit quadratischem Term

$$\text{mpg}_i = \beta_0 + \beta_1 \text{wt}_i + \beta_2 \text{wt}_i^2 + \beta_3 \text{disp}_i + u_i$$

Perspektive 1

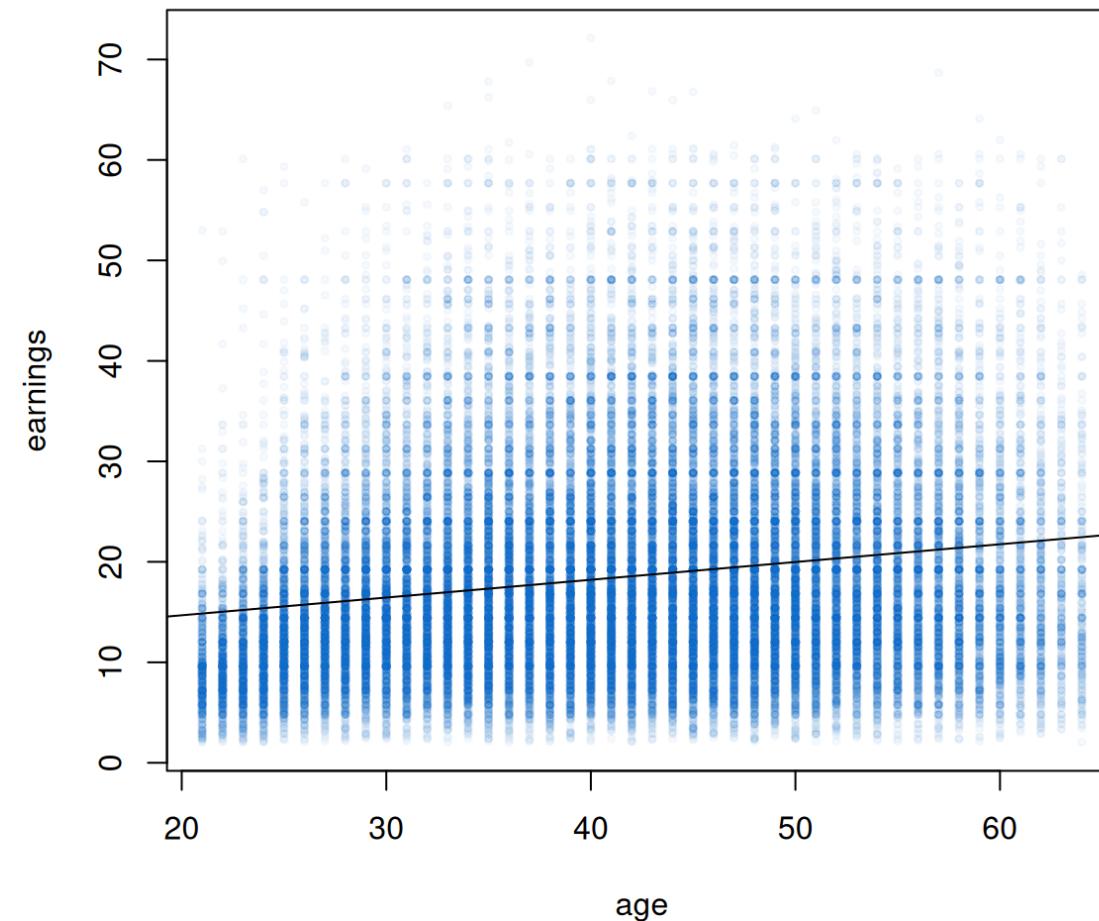
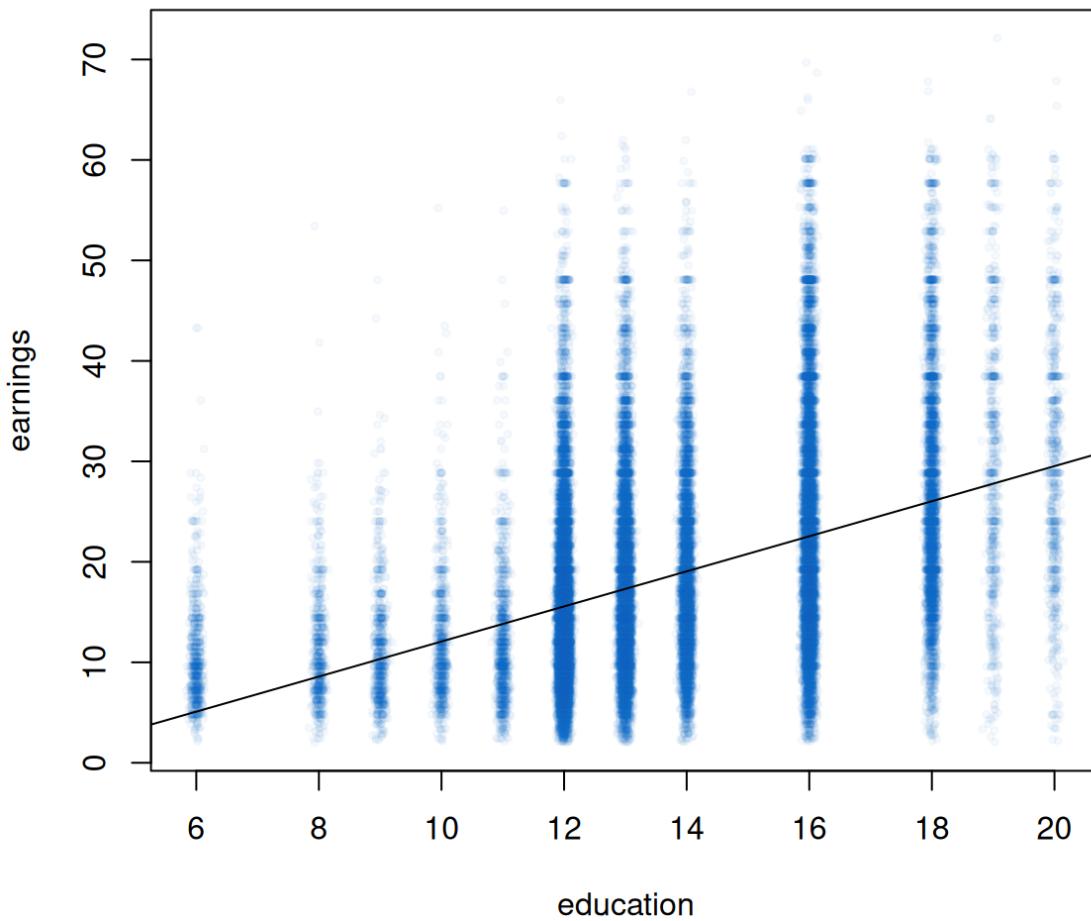


Perspektive 2



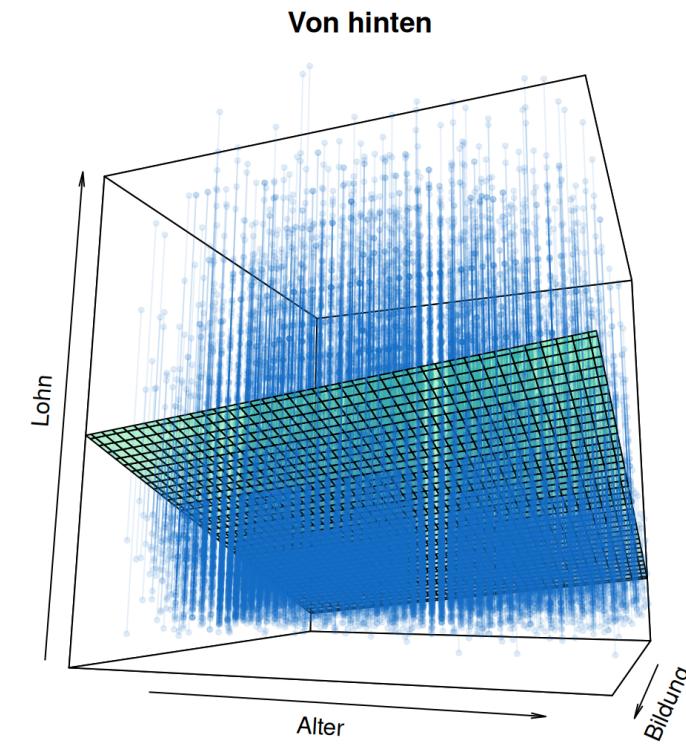
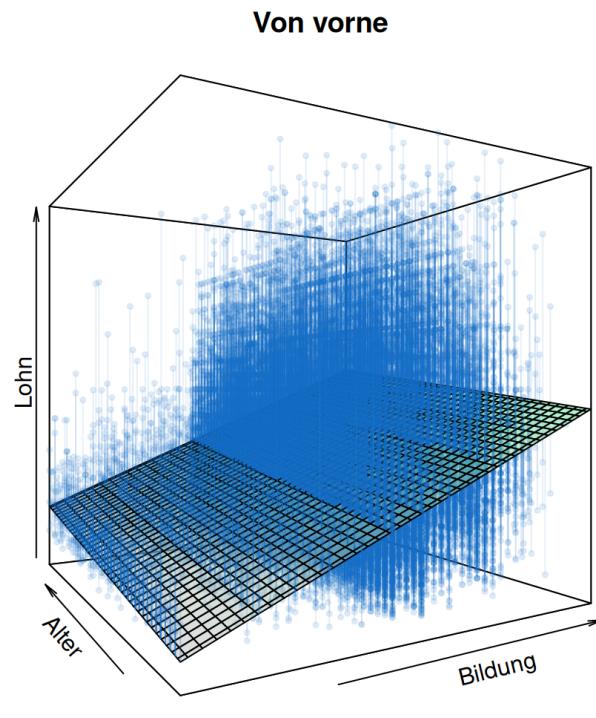
Beispiel 2: Lohn, Alter und Bildung

Sehen wir uns wieder die CPS-Daten an; zuerst wieder als bivariate Regression (von Lohn auf Bildung).



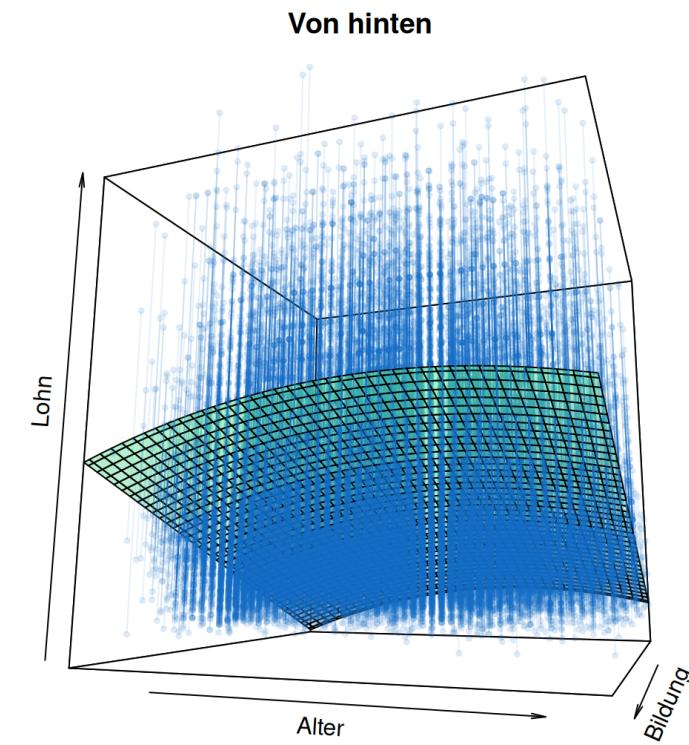
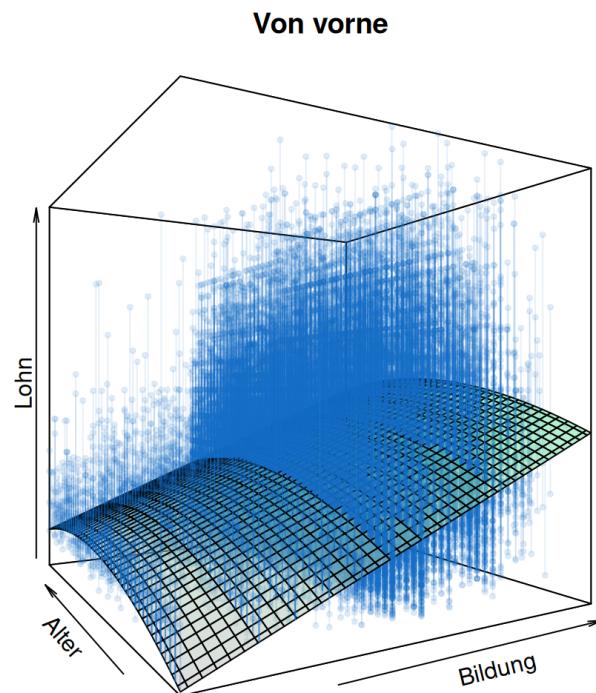
Linear in Alter und Bildung

$$y_i = \beta_0 + \beta_1 \text{Bildung}_i + \beta_2 \text{Alter}_i + u_i$$



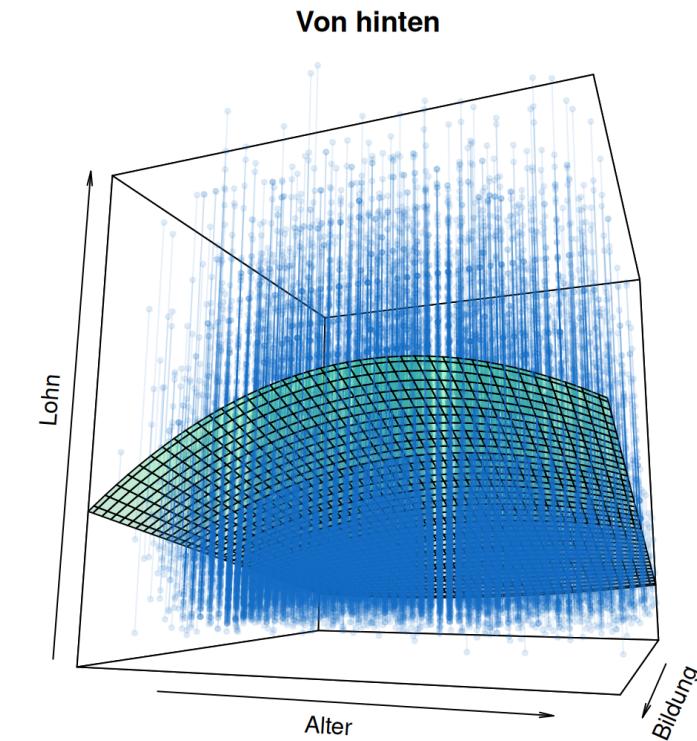
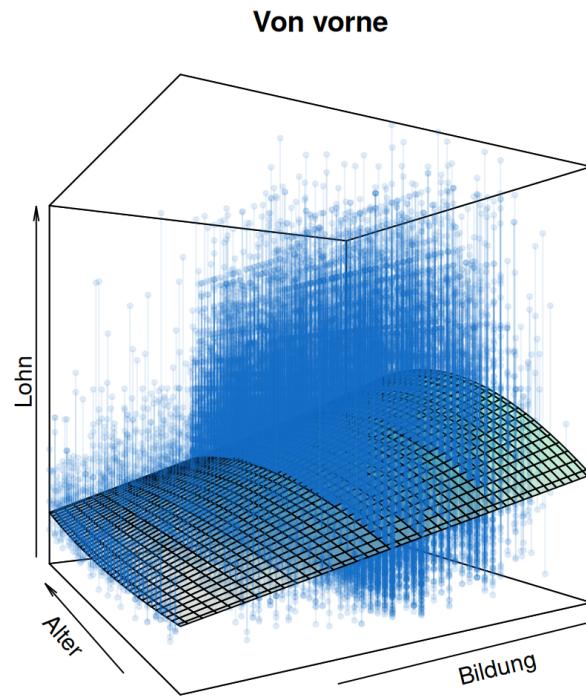
Alter quadriert, linear in Bildung

$$y_i = \beta_0 + \beta_1 \text{Bildung}_i + \beta_2 \text{Alter}_i + \beta_3 \text{Alter}_i^2 + u_i$$



Alter Quadriert, Interaktion mit Bildung

$$y_i = \beta_0 + \beta_1 \text{Bildung}_i + \beta_2 \text{Alter}_i + \beta_3 \text{Alter}_i^2 \\ + \beta_4 \text{Bildung}_i \times \text{Alter}_i + \beta_5 \text{Bildung}_i \times \text{Alter}_i^2 + u_i$$



Literatur

- Frisch, R., & Waugh, F. V. (1933). Partial time regressions as compared with individual trends. *Econometrica: Journal of the Econometric Society*, 387–401.
- Lovell, M. C. (1963). Seasonal adjustment of economic time series and multiple regression analysis. *Journal of the American Statistical Association*, 58(304), 993–1010.
- Wooldridge, J. M. (2020). *Introductory econometrics : a modern approach* (Seventh edition, S. xxii, 826 Seiten). Cengage. <https://permalink.obvsg.at/wuw/AC15200792>

Frisch-Waugh-Lovell-Theorem

Wie viele Variablen?

Visualisierung

Appendix

Herleitung OLS-Schätzer aus Optimierungsproblem

Wir beginnen mit einer Umformung.

$$\begin{aligned} \mathbf{u}'\mathbf{u} &= (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\ &= \mathbf{y}'\mathbf{y} - \boldsymbol{\beta}'\mathbf{X}'\mathbf{y} - \mathbf{y}'\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta} \\ &= \mathbf{y}'\mathbf{y} - 2\boldsymbol{\beta}'\mathbf{X}'\mathbf{y} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta} \end{aligned}$$

Im dritten Schritt machen wir uns den Fakt zunutze, dass $\boldsymbol{\beta}'\mathbf{X}'\mathbf{y} = \mathbf{y}'\mathbf{X}\boldsymbol{\beta}$, da es sich um einen Skalar handelt. Jetzt müssen wir ableiten:

$$\frac{\partial \mathbf{u}'\mathbf{u}}{\partial \boldsymbol{\beta}} = -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\boldsymbol{\beta} \stackrel{!}{=} 0,$$

woraus wir den Schätzer erhalten:

• Zurück

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}.$$

Herleitung Varianz

$$\begin{aligned}\text{Var}(\hat{\boldsymbol{\beta}} \mid \mathbf{X}) &= \text{Var}\left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \mid \mathbf{X}\right) \\ &= \text{Var}\left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \mathbf{u}) \mid \mathbf{X}\right) \\ &= \text{Var}\left(\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} \mid \mathbf{X}\right) \\ &= \text{Var}\left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} \mid \mathbf{X}\right) \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\text{Var}(\mathbf{u} \mid \mathbf{X})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{I}\sigma^2\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\end{aligned}$$

• Zurück