STATISTIK FÜR DIE SOZIALWISSENSCHAFTEN

Trivariate lineare Regressionen

Asymmetrische Zusammenhänge zwischen drei metrischen Variablen

Meine Notizen:

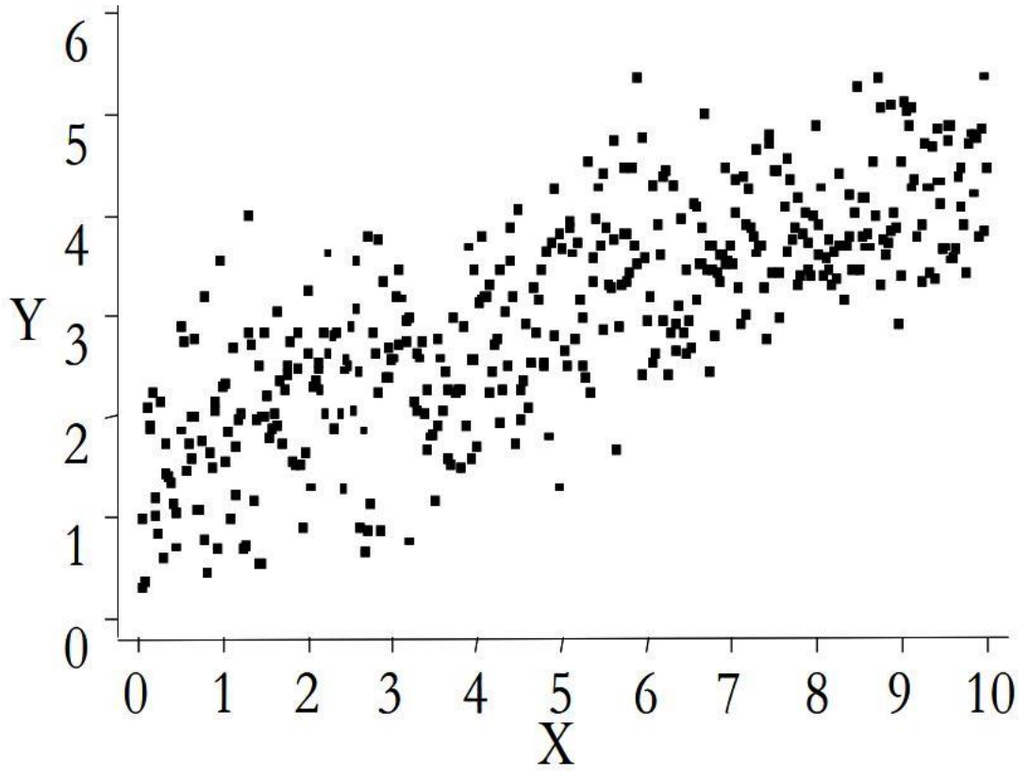
Themen der Woche

- Zusammenhänge zweier Variablen darstellen
- Symmetrische Zusammenhänge berechnen
- Die lineare Regression
- Schätzfunktionen
- Der Determinationskoeffizient R^2
- PRE-Maße
- Erweiterung der Regression auf drei Variablen

Meine Notizen:

Zwei metrische Variablen graphisch darstellen

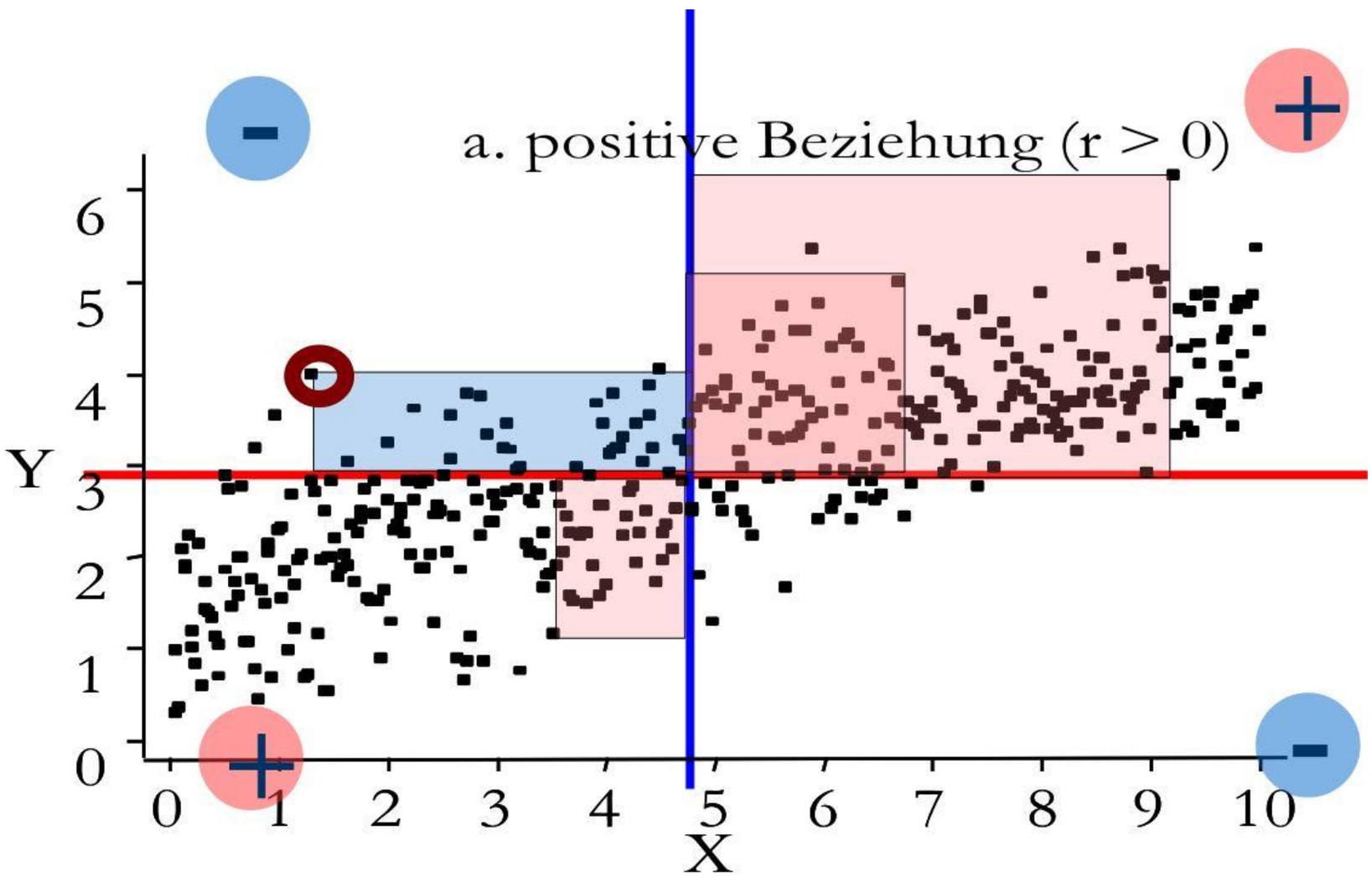
- Tabellarisch können zwei metrische Variablen kaum dargestellt werden.
- Eine graphische Darstellung durch ein Koordinatensystem ist aber sehr günstig.
- Für jeden Fall wird ein Punkt eingetragen. Die Ausprägung der einen Variablen wird auf die x-Achse, die der anderen auf der y-Achse aufgetragen. So entsteht ein 2-dimensionales Streudiagramm:



Meine Notizen:

Zusammenhänge am Streudiagramm messen

- Es werden Rechtecke eingezeichnet, die die Differenzen zum Schwerpunkt abbilden:



Meine Notizen:

Zusammenhang metrischer Variablen

- Der Zusammenhang wird zunächst gemessen als Summe der Flächen der Rechtecke.
- Diese Summe wird Kovariation genannt und mit SP abgekürzt.
- Sie berechnet sich nach:

$$SP_{XY} = \sum_{i=1}^{n} (x_i - \bar{x}) * (y_i - \bar{y})$$

- Die Kovarianz ist entsprechend:

$$s_{XY} = \frac{SP_{XY}}{n} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x}) * (y_i - \bar{y})$$

$$= \left(\frac{1}{n} \sum_{i=1}^{n} x_i * y_i \right) - \bar{x} * \bar{y}$$

Meine Notizen:

Korrelation (nach Pearson)

- Als Standardisierung wird die Kovarianz geteilt durch die Standardabweichungen der Variablen X und Y:

$$r_{XY} = \frac{s_{XY}}{s_x * s_Y} = \frac{SP_{XY}}{\sqrt{SS_X * SS_Y}}$$

- Diese Korrelationen können ihrer Größe nach bewertet werden. Dazu gelten folgende Faustregeln:

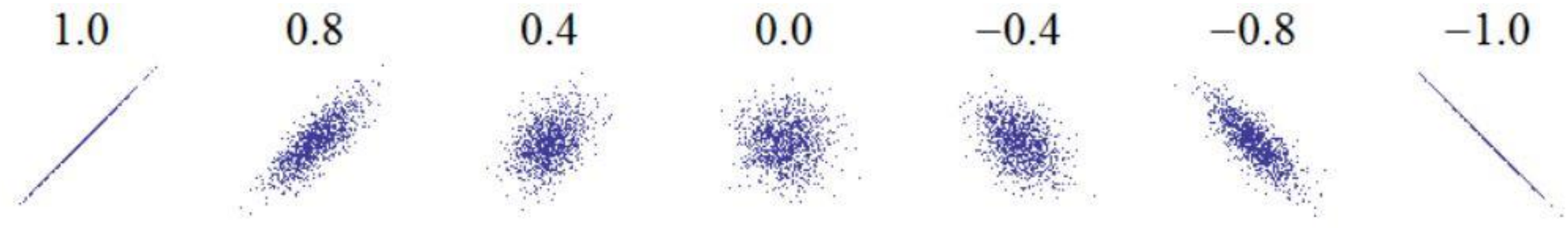
-0.05	$\leq r_{xy} <$	0	vernachlässigbare Korrelation	0	$< r_{xy} \leq$ 0,05
-0,20	$\leq r_{xy} <$	-0,05	geringe Korrelation	0,05	$< r_{xy} \leq$ 0,20
-0,50	$\leq r_{xy} <$	-0,20	mittlere Korrelation	0,20	$< r_{xy} \leq$ 0,50
-0,70	$\leq r_{xy} <$	-0,50	hohe Korrelation	0,50	$< r_{xy} \leq$ 0,70
-1,00	$< r_{xy} <$	-0,70	sehr hohe Korrelation	0,70	$< r_{xy} <$ 1,00
-1,00	$= r_{xy}$		perfekte Korrelation	1,00	$= r_{xy}$

-

+

Meine Notizen:

Lineare Zusammenhänge am Streudiagramm



Meine Notizen:

Zusammenhang zweier metrischer Variablen

- Die gemeinsame Variation zweier metrischer Variablen heißt ihre Kovariation SP_{XY}:

$$SP_{XY} = \sum_{i=1}^{n}(x_i - \bar{x}) * (y_i - \bar{y})$$

- Die Kovarianz s_{XY} ist entsprechend die geteilte Varianz zweier metrischer Variablen:

$$s_{XY} = \frac{SP_{XY}}{n} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}) * (y_i - \bar{y}) = \left(\frac{1}{n}\sum_{i=1}^{n} x_i * y_i\right) - \bar{x} * \bar{y}$$

- Der relative Anteil geteilter Varianz/Variation heißt Korrelation r_{XY}:

$$r_{XY} = \frac{s_{XY}}{s_X * s_Y} = \frac{SP_{XY}}{\sqrt{SS_X * SS_Y}} = \frac{s_{XY}}{\sqrt{s_X^2 * s_Y^2}}$$

Meine Notizen:

Begrenztheit der Korrelation

- Eine Korrelation beschreibt einen Zusammenhang ohne diesen zu erklären.
  - Bei der Korrelation handelt es sich um nur einen Zahlwert, der den Zusammenhang oder Nicht-Zusammenhang misst.
  - Eine Korrelation beschreibt keinen Mechanismus aus Ursache und Wirkung.
    Führt Schokoladenkonsum zu Nobelpreisen?
    Führen Nobelpreise zu mehr Schokoladenkonsum?
    Gibt es überhaupt eine direkte Wirkung zwischen Nobelpreisen und Schokolade?
- Deshalb ist eine Korrelation für Prognosen nicht geeignet.
  - Eine Korrelation hilft nicht zu beurteilen, welchen Einfluss eine Veränderung in einer Variablen auf eine andere Variable hat.
  - Eine Korrelation hilft nicht zu schätzen, welche Ausprägungen ein weiterer Fall hätte.
  - Eine Korrelation hilft nicht zu entscheiden, welches Ausprägungspaar typisch wäre.
    Wie viel Schokolade muss ich für einen weiteren Nobelpreis essen?
    Wie viele Nobelpreise gibt es für 7 kg Schokolade?
    Wie viele Nobelpreise sind typisch für ein Land mit 4 kg Schokolade?
- Außerdem gibt eine Korrelation keine Auskunft über andere als die vorhandenen Ausprägungspaare, Zwischen- oder Randwerte.

Meine Notizen:

Lösungsansatz zu den Grenzen der Korrelation

- Da der Anteil gemeinsamer Streuung an der Gesamtstreuung keine Erklärung des Zusammenhangs bietet, versuchen wir die Streuung einer Variablen zu erklären, indem wir uns die Streuung einer anderen anschauen.
  (*Streuung der Nobelpreise erklären durch Streuung des Schokoladenkonsums*)
- Die Streuung der Variable, die wir als Ursache ansehen, soll erklären, warum die andere Variable Streuung hat. (*Schokoladenkonsum erklärt Nobelpreise*)
- Die Variable der Ursache heißt deshalb erklärende Variable (*Schokolade*) oder auch unabhängige Variable (UV). Die andere Variable heißt dann erklärte Variable (*Nobelpreise*) oder auch abhängige Variable.
- Der Zusammenhang wird als asymmetrisch aufgefasst, während die Korrelation ein symmetrisches Maß ist.

Meine Notizen:

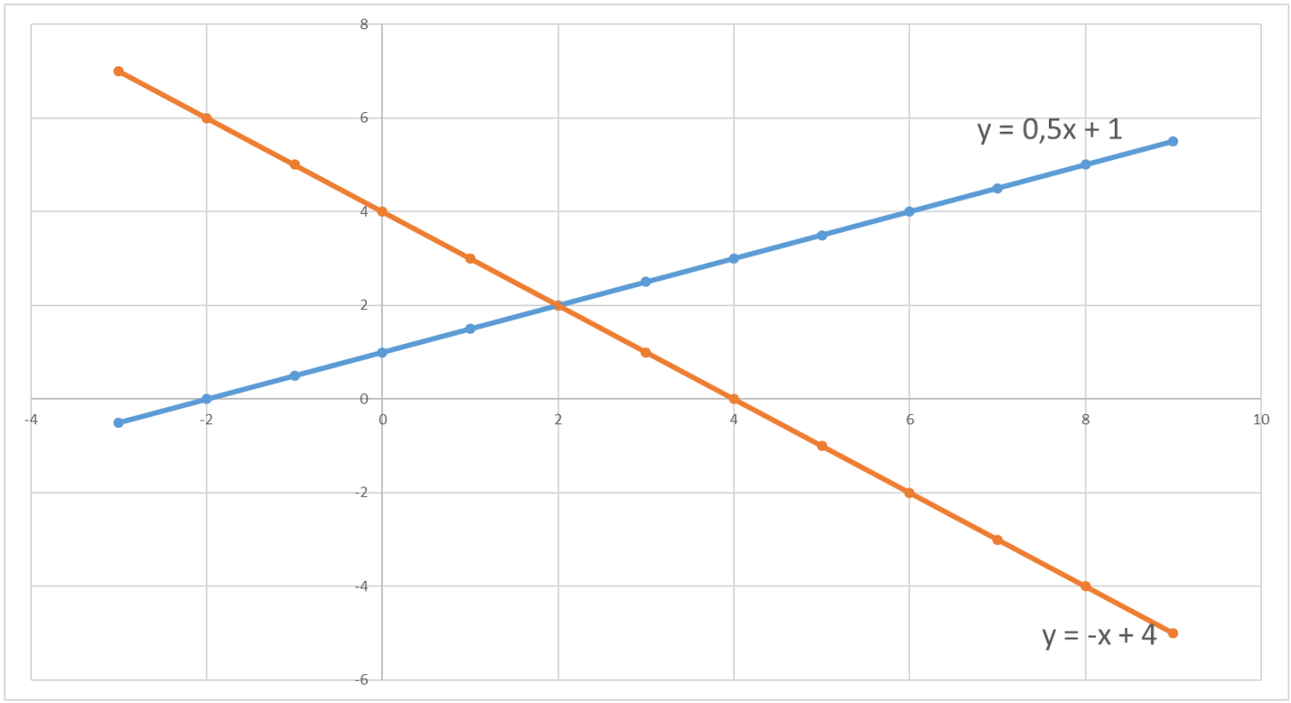
Lösungsansatz zu den Grenzen der Korrelation

- Während die Korrelation als symmetrisches Zusammenhangsmaß nur einen Wert des Zusammenhangs ausgibt, kann bei Annahme eines asymmetrischen Zusammenhangs (/eines Ursachs-Wirkungs-Prinzips) eine Funktion gesucht werden, die (Prognose-) Werte für die erklärte Variable gibt in Abhängigkeit von der erklärenden Variablen.
- Für einen asymmetrischen Zusammenhang muss also eine Funktion gesucht werden, die die Wirkung gut beschreibt.
- Die einfachste Form von Funktion ist die lineare Funktion. Sie wird deshalb in diesem Semester genutzt, um Zusammenhänge zu beschreiben.
  Andere Funktionsformen folgen in Statistik III.

Meine Notizen:

Lineare Funktionen

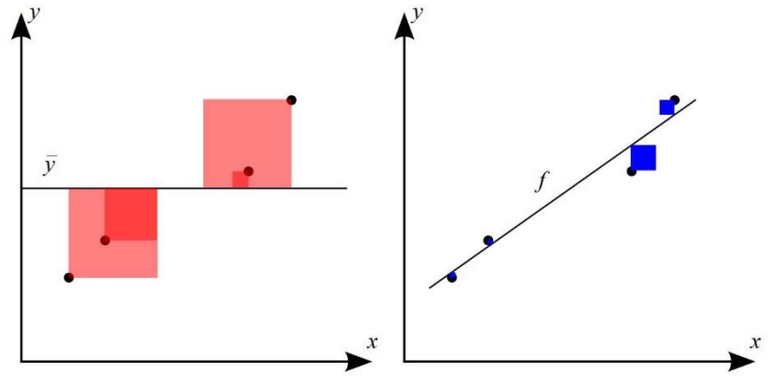
- Lineare Funktionen zeichnen sich durch ihre Steigung aus und durch ihren Schnittpunkt mit der y-Achse.
- Die Steigung wird mit b_1 abgekürzt, der y-Achsenschnittpunkt mit b_0.
- Ihre allgemeine Form ist also: $y = b_0 + b_1 * x$



Meine Notizen:

Beste lineare Funktion suchen

- In die Wolke des Streudiagramms soll eine Gerade hineingelegt werden:



- Es wird die Gerade gewählt, die die quadrierten Abweichungen der Punkte von der Geraden minimiert. Dieses Prinzip nennt sich OLS und ist vom Mittelwert bekannt.
- Dies ist deshalb gleichbedeutend damit, dass die Vorhersagewerte der Geraden (möglichst viele) Eigenschaften von Mittelwerten aufweisen:
  - Summe der Residuen = Mittelwert der Residuen = 0
  - keine Korrelation zw. Vorhersagewerten und Residuen
  - Summe der quadrierten Differenzen zwischen Y und den Y-Dach soll minimal sein.

Meine Notizen:

Beste lineare Funktion suchen

Wie lässt sich nun diese lineare Funktion rechnerisch finden?

- Eine Erklärung von oben hilft uns hier weiter:
  - Es soll die Streuung einer Variablen durch die Streuung einer anderen Variablen erklärt werden.
  - Dies kann offensichtlich nur über die gemeinsame Streuung geschehen.
  - Die Steigung der gesuchten Funktion ist deshalb der Anteil der gemeinsamen Streuung an der Streuung der erklärenden/unabhängigen Variablen.
  - Anders als für die Korrelation (Anteil an der mittleren Streuung der beiden Variablen) ist für den asymmetrischen Zusammenhang nur der Anteil an der unabhängigen Variablen entscheidend.
  - Die Steigung berechnet sich dann nach:

$$b_1 = \frac{s_{XY}}{s_X^2} = \frac{\frac{1}{n}\left(\sum_{i=1}^{n}(x_i - \bar{x}) * (y_i - \bar{y})\right)}{\frac{1}{n}\left(\sum_{i=1}^{n}(x_i - \bar{x})^2\right)}$$

Meine Notizen:

Beste lineare Funktion suchen

- Wie lässt sich der Schnittpunkt mit der y-Achse finden?
- Die gefundene Funktion kann unteranderem aufgefasst werden als eine Vorhersage von Mittelwerten, d.h. ist y = 6 für x = 4 kann das interpretiert werden als: Werden viele Fälle mit x = 4 gezogen, werden sie im Mittelwert ein y von 6 haben.
- Das heißt insbesondere, dass $y(\bar{x}) = b_0 + b_1 * \bar{x} = \bar{y}$.
- Daraus ergibt sich:

$$b_0 = \bar{y} - b_1 * \bar{x}$$

- Die Steigung b_1 heißt Regressionsgewicht, der y-Achsen-abschnitt b_0 heißt Regressionskonstante.

Meine Notizen:

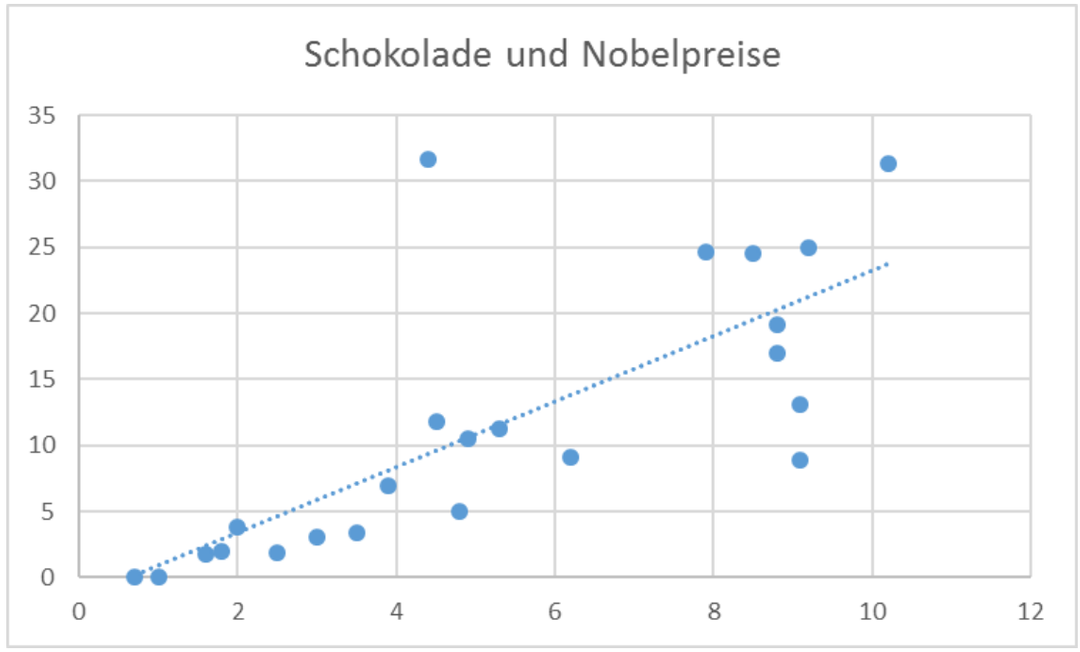
Lineare Regression

- Die beste lineare Funktion zur Annäherung des Zusammenhangs heißt lineare Regression.
- Die lineare Regression gibt nur den linearen oder zumindest nur den monotonen Zusammenhang an, z.B. also keine U-förmigen Zusammenhänge.
- Lineare Funktionen erklären eine Variable Y durch eine Variable X. Die Annahmen über Ursache und Wirkung müssen aber aus der Theorie kommen.
- Die lineare Regression gibt für jedes beliebige x eine Schätzung für y, das $\hat{y}$.
- Die lineare Regression gibt an, wie (stark) eine Änderung von x sich auf y auswirkt, nämlich um einmal das Regressionsgewicht b_1 pro Einheit von X.
- Die Prognose der linearen Regression kann als mittlerer y-Wert verstanden werden, der sich einstellt, wenn viele Fälle mit demselben x-Wert gefunden werden.

Meine Notizen:

Beste lineare Funktion: am Beispiel

- Mittlerer Schokokonsum: 5,291 kg
- Mittlere Zahl Nobelpreise: 11,568
- Kovarianz: 22,38
- Varianz (Schokokonsum): 9,00
- Gesucht: $\hat{y} = b_0 + b_1 * x$



$$b_1 = \frac{s_{XY}}{s_X^2} \approx \frac{22{,}38}{9{,}00} \approx 2{,}486$$

$$b_0 = \bar{y} - b_1 * \bar{x} \approx 11{,}568 - 2{,}486 * 5{,}291 \approx -1{,}585$$

$$\hat{y} = -1{,}585 + 2{,}486 * x$$

Meine Notizen:

Lineare Regression: am Beispiel

- Bei der Funktion $\hat{y} = -1{,}585 + 2{,}486 * x$ handelt es sich um eine Schätzfunktion von y, die die Information der x-Werte zur Schätzung nutzt.
- Wenn $\hat{y} = -1{,}585 + 2{,}486 * 4 = 8{,}358$ bedeutet dies zweierlei:
  - Finden wir einen Fall/ein Land mit 4 kg Schokoladenkonsum, ist es die bestmögliche Vorhersage 8,358 Nobelpreise pro 10 Mio. Einwohner vorherzusagen.
  - Finden wir sehr viele Fälle/Länder mit 4 kg Schokoladenkonsum, wird sich der Mittelwert ihrer Nobelpreise mit steigender Zahl von Ländern an 8,358 annähern.
- b_1 = 2,486 bedeutet, dass im Mittel davon auszugehen ist, dass für jedes weitere Kilo Schokoladenkonsum die Zahl der Nobelpreise pro 10 Mio. Einwohner um 2,486 steigt.
- Die Regression misst die Wirkung von Schokolade auf Nobelpreise. Statistisch ist nur die Messung der Stärke einer solchen Wirkung. Ihre Existenz und Richtung muss theoretisch begründet werden.

Meine Notizen:

Fehler in linearen Regressionen

- Natürlich trifft die Schätzfunktion der linearen Regression nicht jeden Punkt genau.
- Trotz Minimierung dieser Abweichungen durch OLS bleiben Fehler.
- Die genauen Ausprägungen von y lassen sich in diesem Sinne darstellen als

$$y_i = b_0 + b_1 * x + e_i = \hat{y} + e_i$$

- e_i heißt Fehlerterm oder Residuum für den i-ten Fall bzw. die Abweichung des tatsächlichen Falls von seiner Schätzung.
- Hätten wir die Information der X-Variablen nicht, lägen wir mit einer Schätzung des y-Wertes noch schlechter. Eine solche Schätzung wäre der Mittelwert von Y, der Fehler wäre, wie in VL 3 gesehen, genau die Variation von Y

Meine Notizen:

Fehler in linearen Regressionen

- Dieses Prinzip führt dazu, dass wir den Schätzfehler von Y, also die Variation von Y, aufteilen können in einen Teil, der nun durch X zu erklären ist und einen anderen, der als Fehler bleibt und keine Erklärung findet.
- Diese Idee nennt sich das Prinzip der Variationsteilung:

$$\underbrace{\sum_{i=1}^{n}(y_i - \bar{y})^2}_{\text{Variation von } y} = \underbrace{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}_{\text{Variation der Residuen}} + \underbrace{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}_{\text{Variation der Regresswerte}}$$

- Das Prinzip der Variationsteilung erlaubt uns insbesondere die Frage zu stellen, wie viel Prozent der Variation von Y durch X erklärt werden kann.

Meine Notizen:

Determinationskoeffizient R^2

- Der Anteil der Variation von Y, der durch Variation von X erklärt werden kann, ist ein Qualitätskriterium für Regressionen. Er heißt Determinationskoeffizient R^2 oder auch Bestimmtheitsmaß.

- Da: Gesamtvariation von Y = erklärte Variation + nicht erklärte Variation gilt:

$$R^2 = \frac{E_1 - E_2}{E_1} = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2 - \sum_{i=1}^{n}(y_i - \hat{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

- Für bivariate Regressionen gilt glücklicherweise außerdem:

$$R^2 = r_{XY}^2$$

Meine Notizen:

Determinationskoeffizient R^2: am Beispiel

- Die Korrelation zwischen Schokoladenkonsum und Nobelpreisen lag bei $r_{XY} = 0{,}757$.
- Der Einfluss von Schokolade auf Nobelpreise kann beschrieben werden mit

$$\hat{y} = -1{,}585 + 2{,}486 * x$$

- Der Determinationskoeffizient für diese lineare Regression ist

$$R^2 = r_{XY}^2 = 0{,}757^2 = 0{,}573$$

- Dies kann interpretiert werden als:

  57,3% der Streuung von Nobelpreisen kann durch den Schokoladenkonsum erklärt werden.

Meine Notizen:

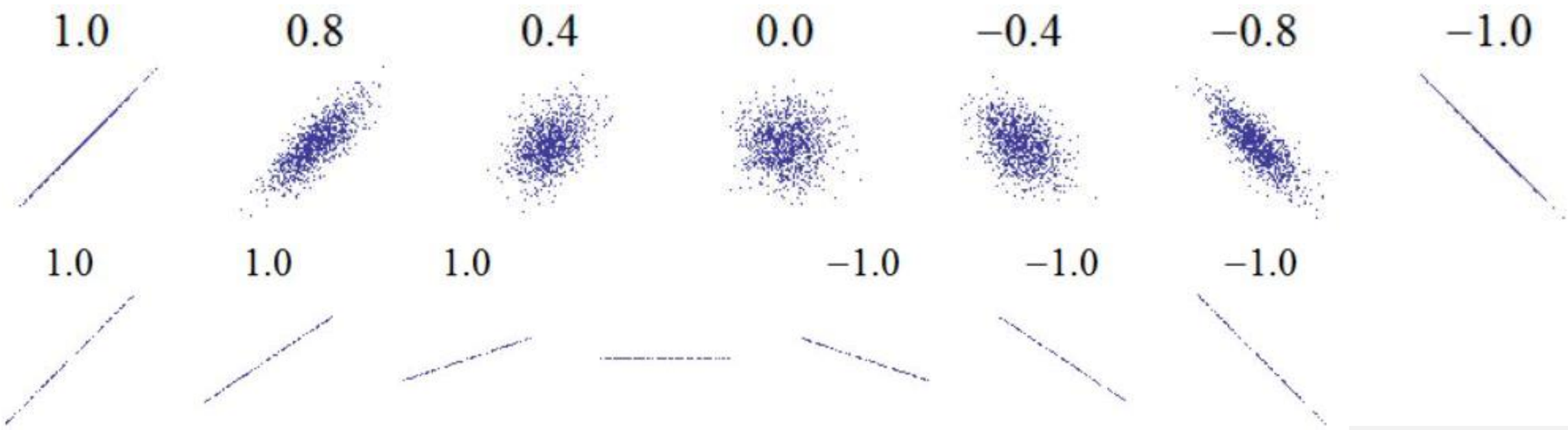
PRE-Maße

- Der Determinationskoeffizient ist ein Beispiel für ein PRE-Maß.
- PRE steht für „Proportional Reduktion of Error“.
- Die Idee hinter allen PRE-Maßen ist:
  - Wir betrachten den Vorhersagefehler, den Schätzung durch ein Lagemaß produziert.
  - Dann schauen wir auf das Zusammenhangsmaß oder das Modell. Das PRE-Maß gibt aus, wie sehr das Modell die Schätzung durch Lagemaße verbessert.
- Für PRE-Maße gilt $0 \leq PRE \leq 1$.
- PRE-Maße können als eigenes Zusammenhangsmaß betrachtet werden. Vor allem aber sind sie ein Gütekriterium für ein asymmetrisches Erklärungsmodell.

Meine Notizen:

Grenzen der Regression I: Linearität

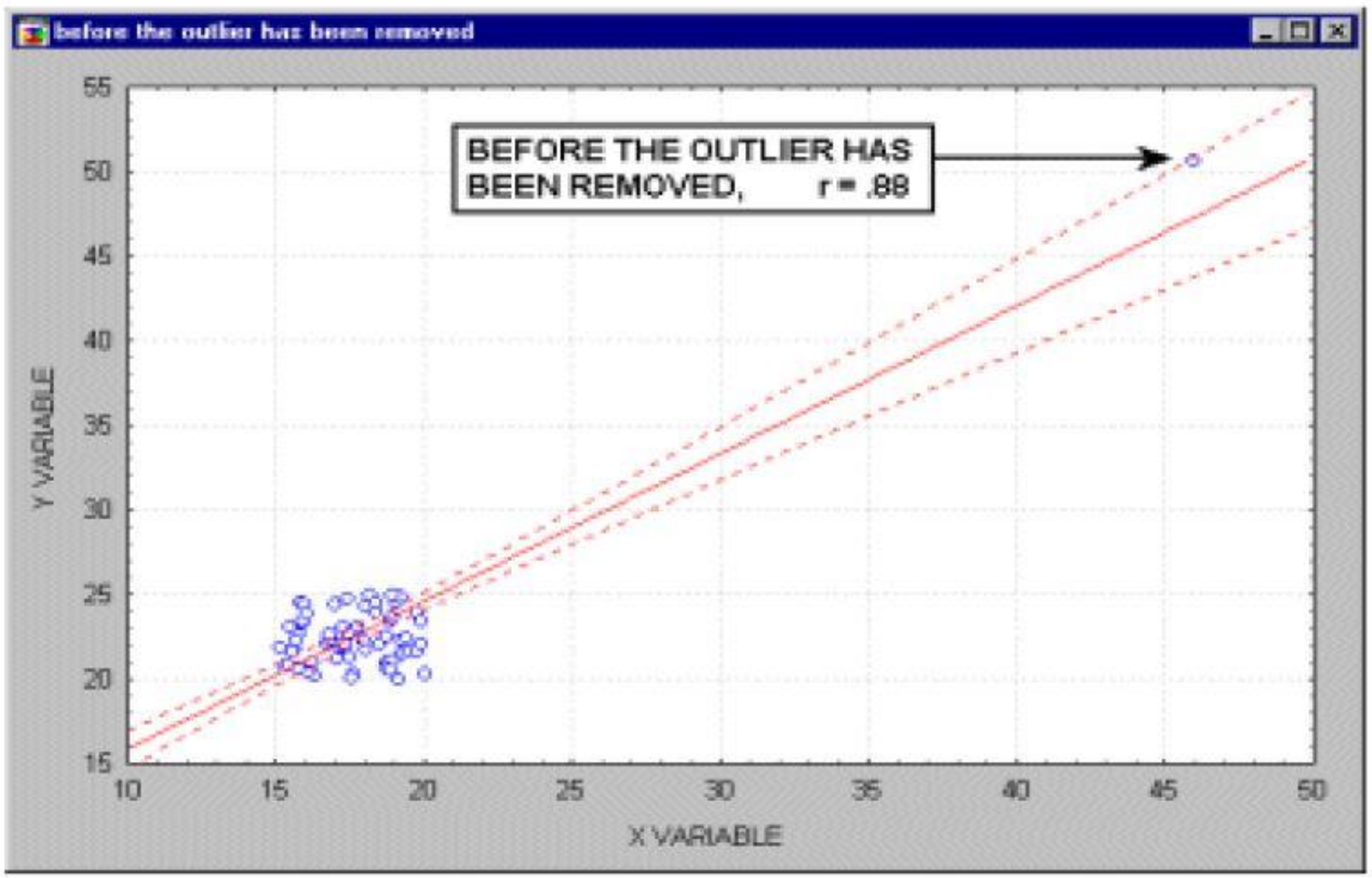
- Regressionen fassen nur monotone Zusammenhänge.



Meine Notizen:

Grenzen der Regression II

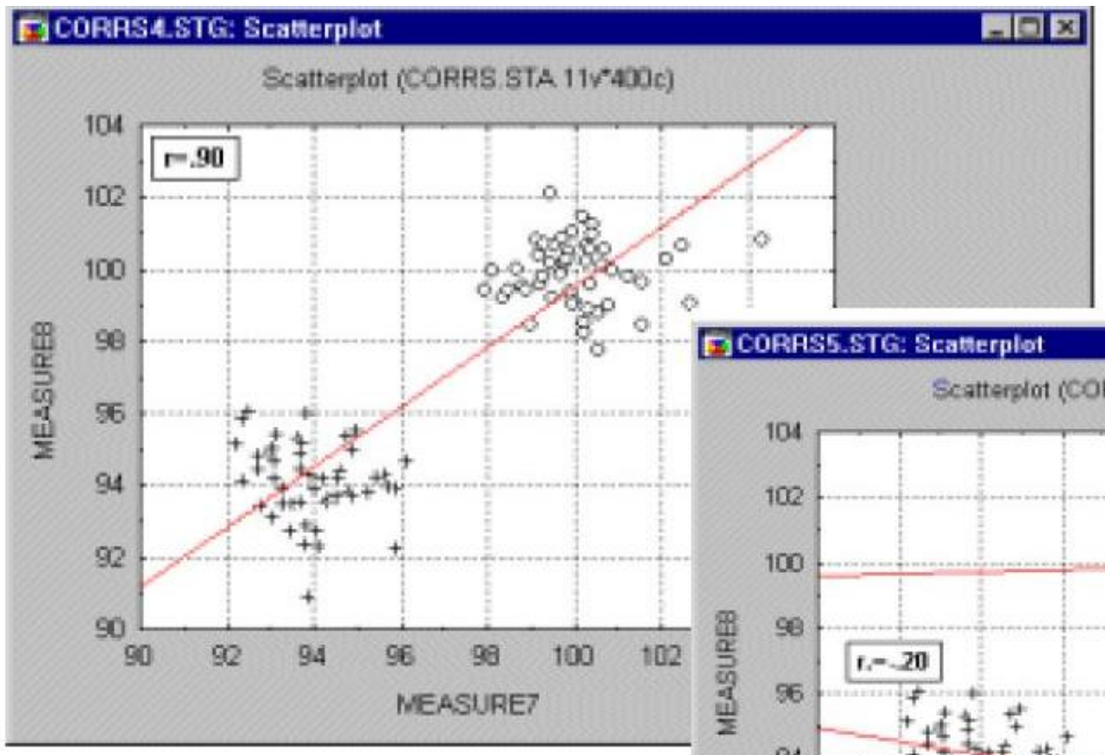
- Bivariate Regressionen brauchen annähernd normalverteilte Daten.

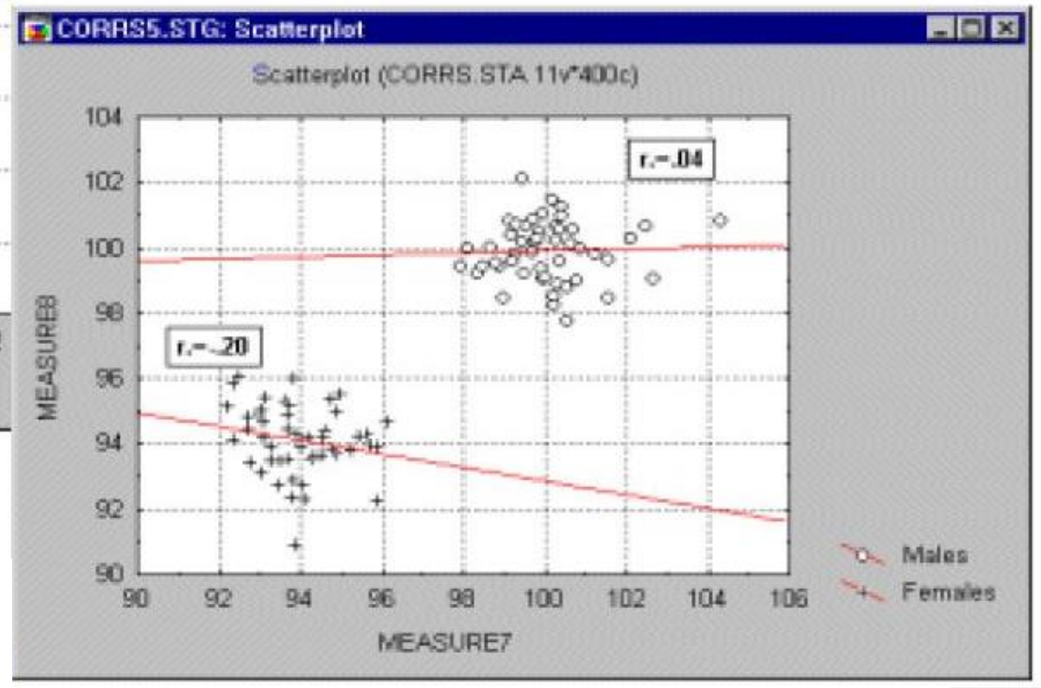


Meine Notizen:

Grenzen der Regression III

- Regressionen können durch Drittvariablen gestört werden.





eine Notizen:

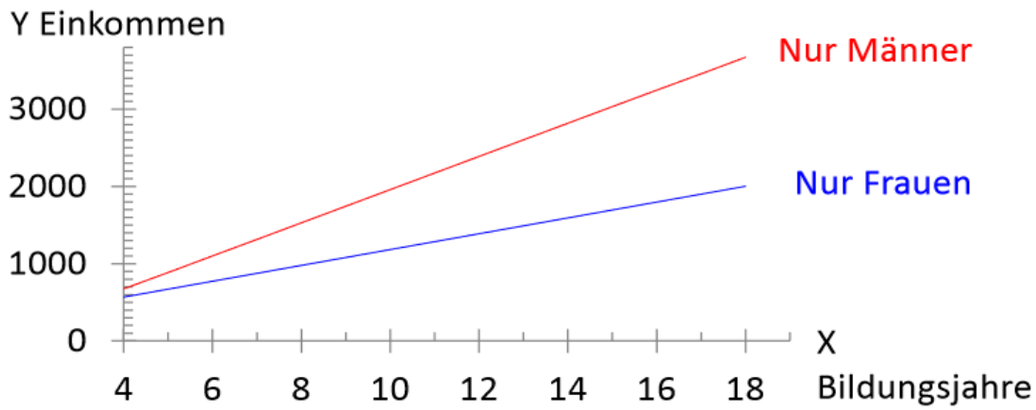
Die Grenze der Bivariatheit überkommen: Trivariate lineare Regressionen

- Im obigen Fall haben wir zwei metrische Variablen, für die eine Regression berechnet werden kann. Die scheinbare Wirkbeziehung ist aber durch eine dritte Variable verursacht, die hier dichotom ist.
- Wollen wir den Einfluss dieser dritten Variablen hier prüfen, ist es noch leicht einzelne Regressionen für männlich und weiblich zu rechnen, wir nennen sie konditionale Regressionsmodelle. Es entstehen dabei aber mehrere Probleme:
  - Wir haben nicht mehr ein Modell zur Beschreibung von Zusammenhängen, sondern zwei.
  - Die beiden Modelle sind nur schwer zu vergleichen.
  - Wie sollen Drittvariablen berücksichtigt werden, die nicht Dichotom, vielleicht sogar metrisch sind?

Meine Notizen:

Die Grenze der Bivariatheit überkommen: Konditionale Regression am Beispiel

Bsp: Ist Geschlecht eine relevante Drittvariable für den Einfluss von Bildung auf Einkommen?



Bivariate Regr.: nur Männer Y = Einkommen	b
Konstante	−184.31
Bildungsjahre	214.36

Daten: Allbus 2012, n=916

Bivariate Regr.: nur Frauen Y = Einkommen	b
Konstante	158.32
Bildungsjahre	102.47

Daten: Allbus 2012, n=781

- Es bleibt offen:
  - Wie groß ist der Einfluss von Geschlecht auf den Zusammenhang Bildung-Einkommen?
  - Ist der Einfluss des Geschlechts relevant?
  - Wie gut erklärt das konditionale Modell das Einkommen einer Person?

Meine Notizen:

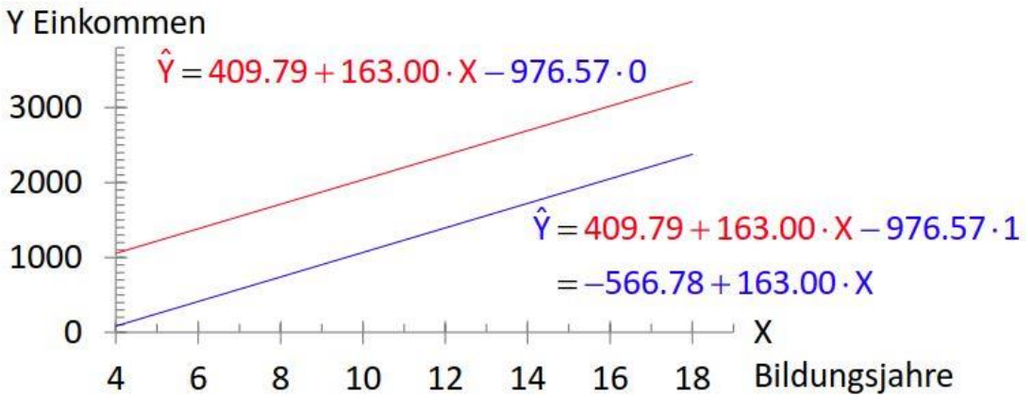
Trivariate lineare Regressionen

- Zur Behebung dieser Probleme lassen sich die beiden konditionalen Regressions-modelle zu einem gemeinsamen trivariaten Regressionsmodell zusammenführen.
- In der trivariaten Regression wird eine abhängige Variable (AV) durch zwei unabhängige Variablen (UVs) erklärt:

  Bedingte Mittelwerte / Vorhersagewerte der abhängigen Variable Y

  = lineare Funktion von 2 (erklärenden) Variablen X und W.

$$Y = \underbrace{b_0 + b_1 \cdot X + b_2 \cdot W}_{=\hat{Y}} + E$$



Trivariate Regression Y = Einkommen	b
Konstante	409.79
Bildungsjahre	163.00
Geschlecht	–976.57

Daten: Allbus 2012, n=1701

Meine Notizen:

Vorhersagen durch Trivariate lin. Regressionen

- Durch Einsetzen in die Regressionsgleichung können wie im bivariaten Fall Vorhersagen für bestimmte Personen oder Gruppen getroffen werden:

z. B.: *männliche (W=0) Person mit Realschulabschluss (X= 10):*

$\hat{Y}$ = 409.79 +163.00·**10** –976.57·**0** = 2039.79 € *= prognostiziertes Einkommen bei männlichen Realschulabsolventen*

$\hat{Y}$ = 409.79 +163.00· **9** –976.57·**0** = 1876.79 € *= prognostiziertes Einkommen bei männlichen Hauptschulabsolventen*

Differenz: **163.00** € = b_1

→ Interpretation Regressionsgewicht: **b_1 gibt Veränderung an, wenn X um +1 Einheit ansteigt**!

$\hat{Y}$ = 409.79 +163.00·10 –976.57·**1** = 1063.22 € *= prognostiziertes Einkommen bei **weiblichen** Realschulabsolventinnen*

$\hat{Y}$ = 409.79 +163.00·10 –976.57·**0** = 2039.79 €*= prognostiziertes Einkommen bei **männlichen** Realschulabsolventen*

Differenz: **–976.57** € = b_2

→ Interpretation Regressionsgewicht: **b_2 gibt Veränderung an, wenn W um +1 Einheit ansteigt**!

Meine Notizen:

Lineradditivität in Trivariater lin. Regressionen

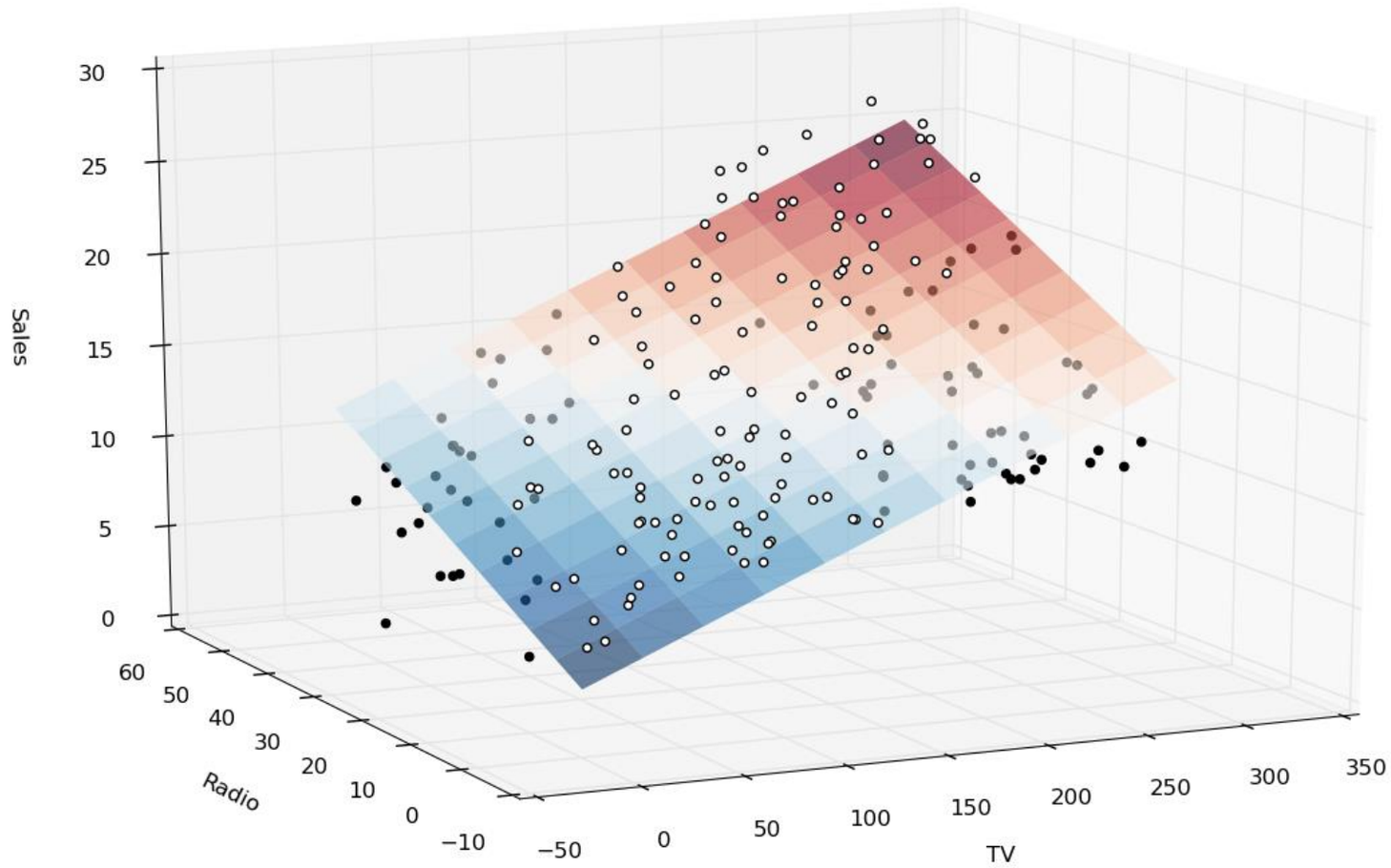
- Jede unabhängige Variable ist jeweils Drittvariable für die andere unabhängige Variable.
- Bivariate Regressionsmodelle einer unabhängigen Variable (gegeben ein Wert der jeweils anderen unabhängigen Variable) unterscheiden sich nur bei der Regressions-konstanten. Im Beispiel bedeutet das: Mehr Bildung wirkt in beiden Geschlechtern gleich. Sollte Bildung unterschiedliche Effekte haben, kann das hier nicht gemessen werden (Linearadditivität).
- die bedingten Regressionsgewichte für eine gegebene Drittvariable sind i.A. ≠ bivariate Regressionsgewichte

Meine Notizen:

Trivariate lin. Regressionen graphisch darstellen

- Ist die zu prüfende Drittvariable metrisch, funktioniert alles wie gesehen.

  Aufwendiger ist nur die graphische Darstellung:



Meine Notizen:

Trivariate Regressionsgewichte bestimmen

- Die Bestimmung der Regressionsgewichte folgt der Grundidee des bivariaten Falles.

  Jedoch müssen nun alle Kovariationen berücksichtigt werden:

$$b_1 = \frac{SS_W * SP_{YX} - SP_{YW} * SP_{XW}}{SS_W * SS_X - (SP_{XW})^2} = \frac{s_W^2 * s_{YX} - s_{YW} * s_{XW}}{s_W^2 * s_X^2 - (s_{XW})^2} = \frac{\hat{\sigma}_W^2 * \hat{\sigma}_{YX} - \hat{\sigma}_{YW} * \hat{\sigma}_{XW}}{\hat{\sigma}_W^2 * \hat{\sigma}_X^2 - (\hat{\sigma}_{XW})^2}$$

$$b_2 = \frac{SS_X * SP_{YW} - SP_{YX} * SP_{XW}}{SS_W * SS_X - (SP_{XW})^2} = \frac{s_X^2 * s_{YW} - s_{YX} * s_{XW}}{s_W^2 * s_X^2 - (s_{XW})^2} = \frac{\hat{\sigma}_X^2 * \hat{\sigma}_{YW} - \hat{\sigma}_{YX} * \hat{\sigma}_{XW}}{\hat{\sigma}_W^2 * \hat{\sigma}_X^2 - (\hat{\sigma}_{XW})^2}$$

$$b_0 = \bar{y} - b_1 * \bar{x} - b_2 * \bar{w}$$

Meine Notizen:

Trivariate Regressionsgewichte bestimmen

- Als Beispiel wird die Wirkung von Bildung und Geschlecht auf das Einkommen berechnet:

Variable		Mittelwert	Varianzen und Kovarianzen		
Bildungsjahre	(X)	11.75663	11.1807		
Geschlecht	(W)	0.46022	0.10264	.24856	
Einkommen	(Y)	1876.652	1722.2	–226.01	4161753.6
			(X)	(W)	(Y)

Daten: Allbus 2012 (n=1647, Berechnungen mit STATA)

Y = Einkommen	b
Konstante	409.79
Bildungsjahre	163.00
Geschlecht	–976.57
R^2	0.1205

Daten: Allbus 2012, n=1697

$$b_1 = \frac{.24856 \cdot 1722.2 - (-226.01) \cdot .10264}{0.24856 \cdot 11.1807 - .10264^2} = 163.00 \quad b_2 = \frac{11.1807 \cdot (-226.01) - 1722.2 \cdot .10264}{0.24856 \cdot 11.1807 - .10264^2} = -976.59$$

$$b_0 = 1876.652 - 163 \cdot 11.75663 + 976.59 \cdot .46022 = 409.77$$

Meine Notizen:

Was Sie am Ende der Woche können sollten

- Kern: Sie nuten Regressionsmodelle, um für zwei unabhängige Variablen deren kontrollierte Effekte auf eine Abhängige zu analysieren.
- Sie berechnen und interpretieren symmetrische bivariate Zusammenhänge.
- Sie kennen Vorteile und Einschränkungen von Regressionen.
- Sie bestimmen eine Regressionsgleichung und interpretieren diese.
- Sie erklären die Idee der Variationszerlegung.
- Sie berechnen und interpretieren den Determinationskoeffizienten.
- Sie erklären die Logik von PRE-Maßen.
- Sie erweitern Regressionen auf drei Variablen.
- Sie erklären das Prinzip der Linearadditivität.

Meine Notizen: