HTW Chur

Hochschule für Technik und Wirtschaft
University of Applied Sciences

Multivariate Regression

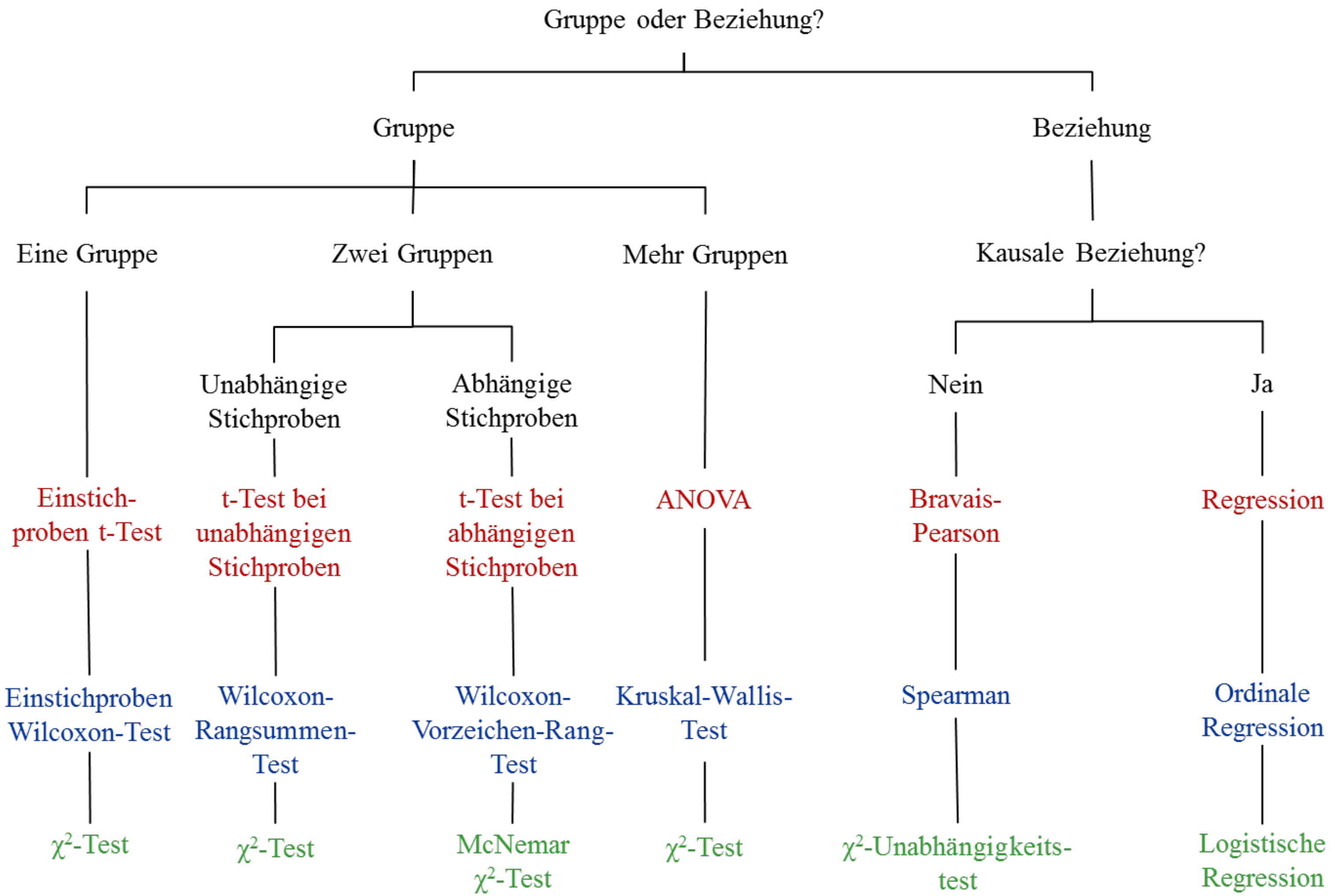
Quantitative Forschungsmethoden

Prof. Dr. Franz Kronthaler

Lernziele

- verstehen, dass mit der Regressionsanalyse ein kausaler Zusammenhang zwischen einer abhängigen und unabhängigen Variablen geschätzt wird
- verstehen, dass die Regressionsanalyse ein theoriegeleitetes Vorgehen erfordert
- wissen was die Methode der Kleinsten-Quadrate ist
- wissen wie eine Regressionsanalyse berechnet wird und wie die Ergebnisse zu interpretieren sind
- verstehen, dass wir den marginalen Effekt einer unabhängigen Variable angeben können, unter Voraussetzung, dass alle anderen Variablen konstant bleiben
- wissen was unter Kontrolle bedeutet
- wissen was eine Dummy-Variable ist
- wissen wie der Fit der Regressionsanalyse geprüft wird

Einleitung



Variable(n) von Interesse sind:
- metrisch und normal verteilt
- ordinal oder metrisch und nicht-normal verteilt
- nominal

Einleitung

- die Regressionsanalyse ist ein äusserst beliebtes Instrument, welches bei sehr vielen Fragestellungen Anwendung findet
- Beispiele:
  - Welchen Einfluss haben der Preis und Marketingaufwendungen auf meine Verkaufszahlen?
  - Welche Faktoren beeinflussen die Innovationsrate eines Landes? Welchen Einfluss haben Ausgaben für Grundlagenforschung?
  - Welche Faktoren beeinflussen das wirtschaftliche Wachstum von Entwicklungsländern? Welche Rolle spielt die Entwicklungshilfe?
  - Welche Faktoren haben einen Einfluss auf die Gewaltbereitschaft von Jugendlichen? Was passiert, wenn Fernsehsendungen gewaltreicher werden?
  - Welche Faktoren beeinflussen die Abtreibungsrate? Welche Rolle spielt die Religionszugehörigkeit?
  - Welche Faktoren beeinflussen den Schutz der Intellektuellen Eigentumsrechte in Ländern? Welche Rolle spielt die kulturelle Identität?

Ziel der Regressionsanalyse

- das Ziel der Regressionsanalyse ist es
  - zu beschreiben, welchen Einfluss unabhängige Variablen X_j auf eine abhängige Variable Y ausüben
  - zu bestimmen, was mit Y passiert, wenn sich eine Variable X_j verändert oder verändert wird

$$Y \longleftarrow X_1, X_2, \ldots, X_J$$

abhängige Variable	unabhängige Variablen
(metrisch)	(metrisch und Dummy-Variablen)

- Dummy-Variablen sind nominale Variablen mit zwei Ausprägungen (z. B. Mann, Frau)

Ziel der Regressionsanalyse

- neben den Bezeichnungen abhängige und unabhängige Variable werden noch zahlreiche andere Bezeichnungen verwendet:

abhängige Variable	*unabhängige Variable*
endogene Variable	*exogene Variable*
erklärte Variable	*erklärende Variable*
Regressand	*Regressor*
Prognosevariable	*Prädiktorvariable*

- die Einteilung in abhängige und unabhängige Variablen ist keine statistische, auf Daten basierende Einteilung, die Einteilung basiert ausschliesslich aufgrund einer theoretischen Begründung
- die theoretische Fundierung ist zentral für die Regressionsanalyse
- fehlt die Fundierung ist das Ergebnis der Regressionsanalyse wertlos, wir können nicht sagen, was mit Y geschieht, wenn wir X_j verändern

«unter Kontrolle»

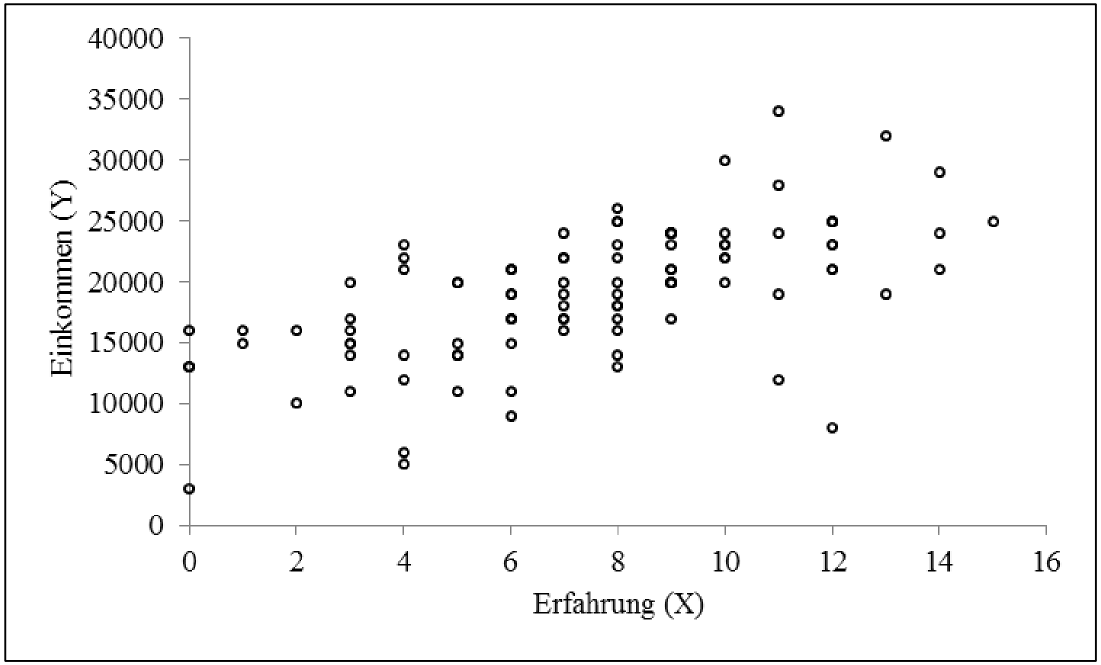
- die Bedeutung von «unter Kontrolle»
  - obwohl wir mit der Regressionsanalyse gleichzeitig mehrere unabhängige Variablen simultan untersuchen, sind wir in der Regel trotzdem nur an einer dieser Variablen interessiert
  - die anderen unabhängigen Variablen nehmen wir in die Regressionsanalyse auf, da auch sie einen Einfluss auf die abhängige Variable ausüben
  - um den echten Einfluss der Variable, die von Interesse ist, zu ermitteln, müssen wir für den Einfluss der anderen Variablen kontrollieren
  - technisch sprechen wir daher auch von Kontrollvariablen
  - wir ermitteln den Einfluss einer unabhängigen Variable X_j auf Y unter Kontrolle aller anderen für den Sachverhalt relevanten Variablen

«unter Kontrolle»

- wir wollen dies noch einmal an einem Beispiel verdeutlichen
  - uns interessiert der Einfluss der Branchenberufserfahrung auf das Unternehmenswachstum
  - die Theorie zeigt an, dass die Branchenberufserfahrung einen Einfluss ausübt
  - gleichzeitig entnehmen wir der Theorie aber auch, dass nicht nur die Branchenberufserfahrung wichtig für das Unternehmenswachstum ist, sondern ebenso Marketing, Forschung und Entwicklung, Branche, etc. eine Rolle spielen könnten
  - um den echten Einfluss der Branchenberufserfahrung zu ermitteln, müssen wir daher für diese Faktoren kontrollieren

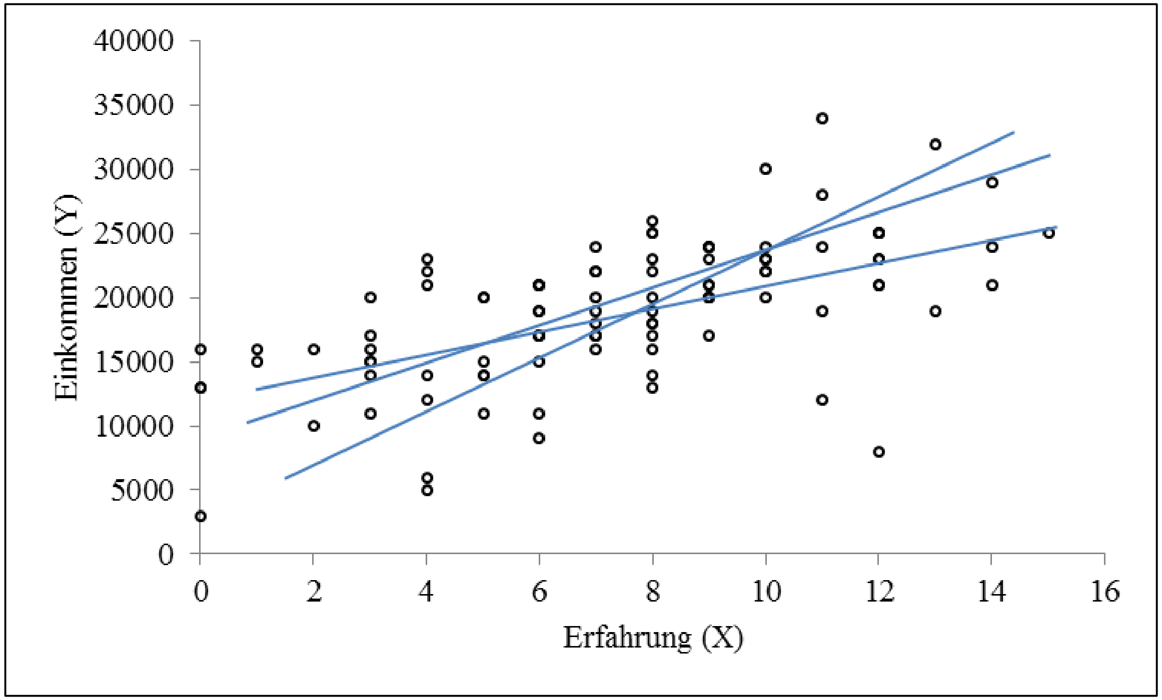
Methode der Kleinsten-Quadrate

- eine wichtige Schätztechnik zur Berechnung der Regressionsgeraden ist die Methode der Kleinsten-Quadrate (Ordinary-Least-Squares OLS)
- die Methode der Kleinsten-Quadrate ist unter bestimmten Voraussetzungen eine der besten Schätztechniken, die wir kennen
- Ausgangspunkt der Überlegungen ist das Streudiagramm zwischen abhängiger und unabhängiger Variable, z. B. Berufserfahrung und Einkommen



Methode der Kleinsten-Quadrate

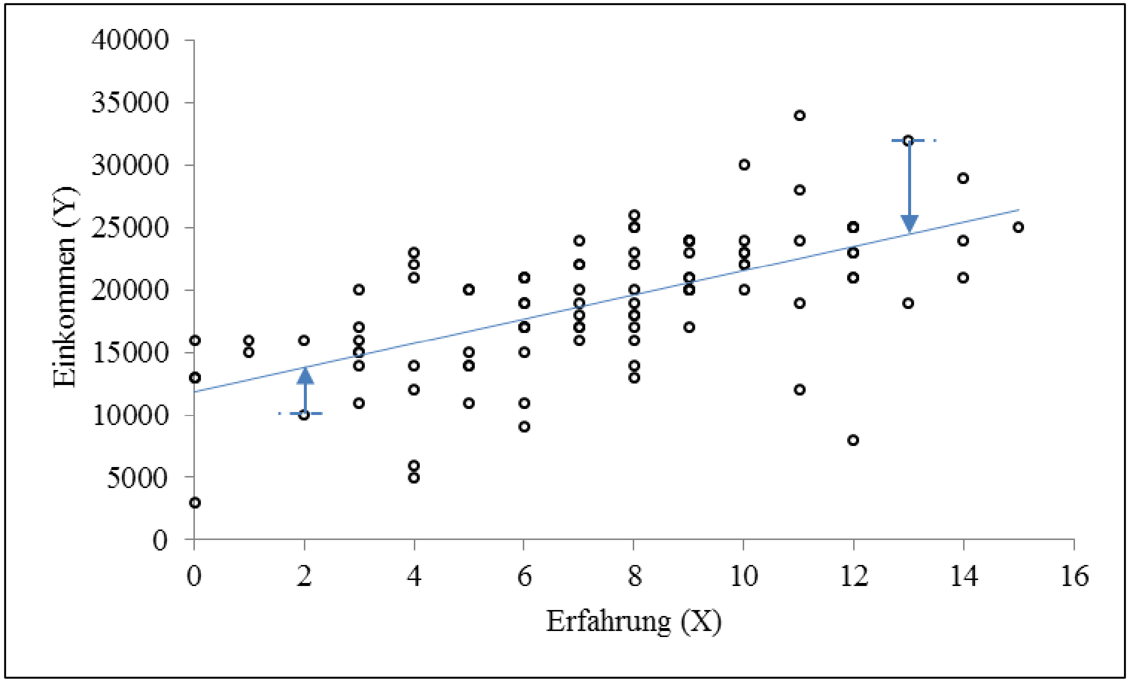
- Ziel ist es, die Punktwolke mit Hilfe einer Geraden optimal zu beschreiben
- folgende Abbildung zeigt drei Versuche freihändig eine Gerade einzuzeichnen



- welche Gerade die richtige Gerade ist, lässt sich visuell schwer entscheiden

Methode der Kleinsten-Quadrate

- ein besseres Verfahren ist die Minimierung der Abstände zwischen den beobachteten Werten und der Geraden



- die Methode der Kleinsten-Quadrate minimiert den Abstand zwischen beobachteten Werten und Gerade

Methode der Kleinsten-Quadrate

- mathematisch lässt sich eine Gerade wie folgt darstellen

  $$Y = b_0 + b_1 X$$

  – b_0 ist der Achsenabschnitt, d. h. der Schnittpunkt der Linie mit der Y-Achse

  – b_1 ist die Steigung der Geraden

- mit Hilfe der Kleinsten-Quadrate-Methode schätzen wir die Gerade, wir schreiben daher

  $$\hat{Y} = b_0 + b_1 X$$

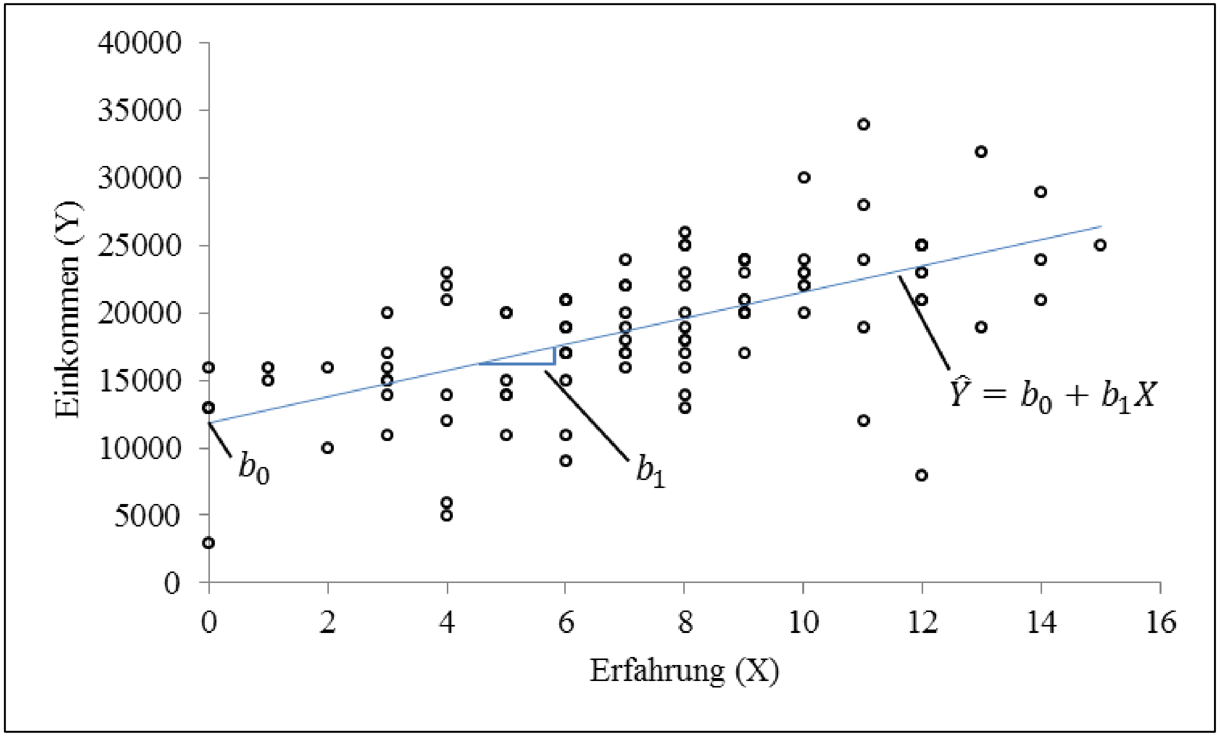
  – $\hat{Y}$ (Y Hut) sind die mit Hilfe der Regressionsgeraden geschätzten Y-Werte von X

  – b_0 ist der Achsenabschnitt

  – b_1 die Steigung

Methode der Kleinsten-Quadrate

- Achsenabschnitt, Steigung und Regressionsgerade sind in folgender Abbildung dargestellt



Methode der Kleinsten-Quadrate

- jetzt können wir uns der Minimierungsregel zuwenden
- Ziel des Kleinsten-Quadrate-Verfahrens ist es, die Summe des Abstandes zwischen beobachteten y_i Werten und geschätzten $\hat{y}_i$ Werten zu minimieren

  $$e_i = y_i - \hat{y}_i$$

  – e_i ist der Abstand zwischen beobachteten und geschätzten Werten
- wenn wir alle Abstände aufsummieren, kommen wir zu folgender Formel

  $$\sum e_i = \sum (y_i - \hat{y}_i)$$

- zusätzlich können wir $\hat{y}_i$ durch die Geradengleichung $\hat{y}_i = b_0 + b_1 x_i$ ersetzen

  $$\sum e_i = \sum \left(y_i - (b_0 + b_1 x_i)\right)$$

Methode der Kleinsten-Quadrate

- der letzte Gedankenschritt zeigt, warum das Verfahren die Methode der kleinsten Quadrate genannt wird
- die Linie, die die Punktwolke am besten beschreibt, ist die durchschnittliche Gerade
- da sie die durchschnittliche Gerade ist, ist die Summe der Abweichungen immer null
- wir behelfen uns daher damit, dass wir nicht die Abweichungen minimieren, sondern die quadrierten Abweichungen

$$\sum {e_i}^2 = \sum \left(y_i - (b_0 + b_1 x_i)\right)^2 \rightarrow min!$$

Methode der Kleinsten-Quadrate

- dieses Minimum finden wir, indem wir die partiellen Ableitungen nach b_0 und b_1 bilden und diese nach b_0 und b_1 auflösen

$$b_0 = \frac{\sum {x_i}^2 \sum y_i - \sum x_i \sum x_i y_i}{n \sum {x_i}^2 - (\sum x_i)^2}$$

$$b_1 = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum {x_i}^2 - (\sum x_i)^2}$$

- die geschätzte Regressionsgerade ist

$$\hat{Y} = b_0 + b_1 X$$

Methode der Kleinsten-Quadrate

- bei mehr als einer unabhängigen Variable ist die Vorgehensweise identisch
- die Zielfunktion ist dann

$$\sum {e_i}^2 = \sum \left(y_i - \left(b_0 + b_1 x_{1i} + b_2 x_{2i} + \ldots + b_j x_{ji}\right)\right)^2 \rightarrow min!$$

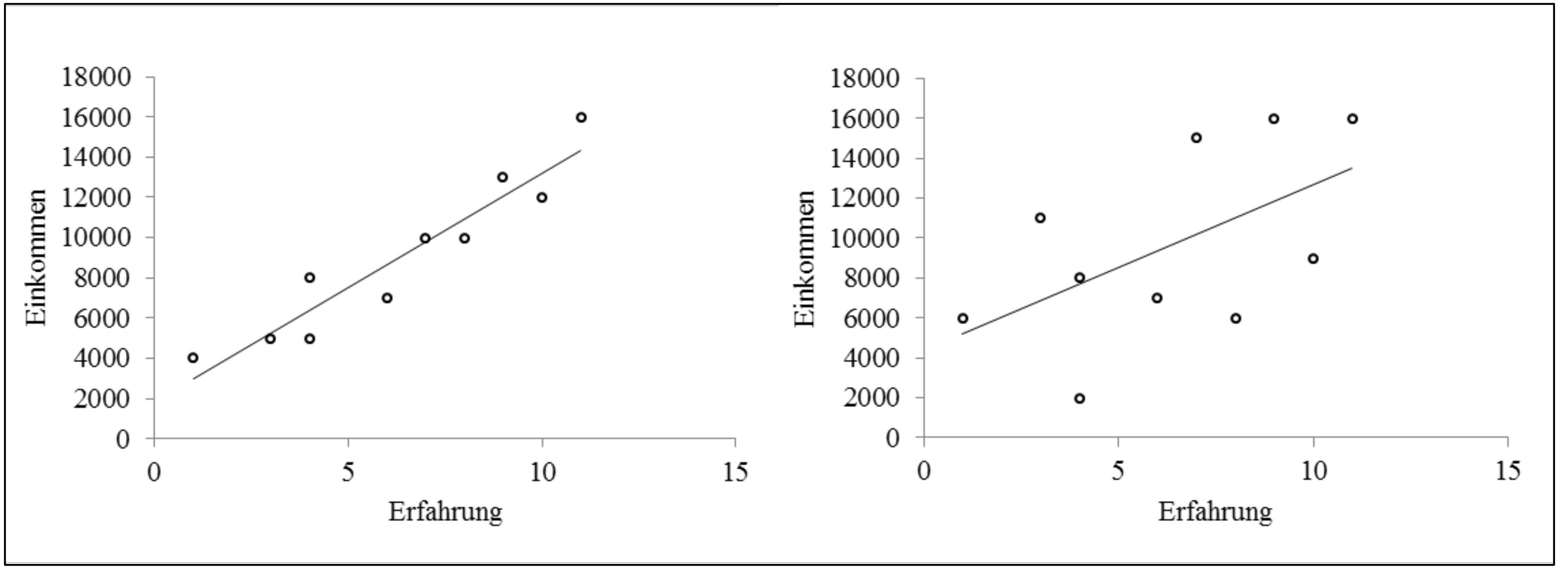
- und die geschätzte Regressionsfunktion ist

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2 + \ldots + b_J X_J$$

  - b_0 ist das Absolutglied
  - b_1 bis b_J sind die geschätzten Koeffizienten für die unabhängigen Variablen X_j
  - J ist die Anzahl der unabhängigen Variablen
  - $\hat{Y}$ sind die mit Hilfe der Regressionsfunktion geschätzten Werte

Das Bestimmtheitsmass R^2

- ein Problem der Regressionsgerade ist, dass wir für jede Punktwolke eine optimale Gerade berechnen können, egal wie gut die Gerade den Zusammenhang erklärt



- es wird ein Mass benötigt, welches uns sagt, wie gut die Gerade zur Punktwolke passt bzw. welchen Erklärungsgehalt die Regressionsgerade hat

Das Bestimmtheitsmass R²

- das Mass, welches wir nutzen ist das Bestimmtheitsmass R^2

$$R^2 = 1 - \frac{unerklärte\ Varianz}{Gesamtvarianz} = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$$

- es gibt an, wie viel Prozent der Veränderung der abhängigen Variable Y durch die unabhängige Variable X erklärt werden
- definiert ist das Bestimmtheitsmass zwischen 0 und 1
- 0 bedeutet, dass die Regressionsgerade 0 % der Bewegung (Varianz) von Y erklärt
- 1 bedeutet, dass sie 100 % der Bewegung (Varianz) von Y erklärt
- 0.3 bedeutet, dass die Regressionsgerade 30 % der Bewegung von Y erklärt

Anwendungen

- Gegeben sind folgende Werte:

Person	Einkommen	Erfahrung
1	7000	6
2	9000	10
3	16000	11
4	2000	4
5	6000	1
6	8000	4

  - Berechne von Hand die Regressionsgerade.
  - Berechne von Hand das R^2.
  - Wie gut erklärt die Regressionsgerade den Zusammenhang?
  - Was passiert, wenn die Berufserfahrung um ein Jahr steigt?

F-Test, t-Test und Adjusted-R^2

- die Regressionsfunktion wird aus einer Stichprobe geschätzt

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2 + \ldots + b_J X_J$$

- sie lässt sich als Realisation einer «wahren» Funktion auffassen mit den unbekannten Parametern β_1, β_2,..., β_J

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_J X_J + u$$

  - Y ist die abhängige Variable
  - β_0 ist das Absolutglied der Regressionsfunktion
  - β_{j} sind die wahren Regressionskoeffizienten
  - X_{j} sind die unabhängigen Variablen
  - u ist die Störgrösse
- in u ist die Vielzahl zufälliger Einflüsse auf Y zusammengefasst und ist entsprechend eine Zufallsvariable

F-Test, t-Test und Adjusted-R^2

- wenn zwischen abhängiger Variable Y und unabhängigen Variablen X_j kein Zusammenhang besteht, dann müssen die wahren Regressionskoeffizienten null sein
  - H_0: $\beta_1 = \beta_2 = \cdots = \beta_J = 0$
  - H_A: $\beta_j \neq 0$ für mindestens ein j
- getestet wird die Nullhypothese mit Hilfe des F-Tests zum Signifikanzniveau von in der Regel $\alpha = 1\%$ bzw. $\alpha = 0.01$

$$F = \frac{R^2/(J)}{(1-R^2)/(N-J-1)}$$

  - R^2 ist das Bestimmtheitsmass
  - J die Anzahl der unabhängigen Variablen
  - N die Anzahl der Beobachtungen
- der Prüfwert ist F-verteilt mit $\mathrm{F}_{\mathrm{g}_1} = \mathrm{J}$ und $\mathrm{F}_{\mathrm{g}_2} = \mathrm{N} — \mathrm{J} — 1$ Freiheitsgraden

F-Test, t-Test und Adjusted-R^2

- für die unabhängigen Variablen stellen wir jeweils eine eigene Null- und Alternativhypothese auf
  - H_0: $\beta_j = 0$
  - H_A: $\beta_j \neq 0$
  - das Signifikanzniveau spezifizieren wir im Vorfeld der Untersuchung mit $\alpha = 10\%, 5\%, 1\%$ bzw. $\alpha = 0.1, 0.05, 0.01$
- die Teststatistik ist

$$t = \frac{b_j}{s_{b_j}}$$

  - b_j ist der geschätzte Koeffizient für die unabhängige Variable j
  - s_{b_j} der Standardfehler für den geschätzten Koeffizienten

F-Test, t-Test und Adjusted-R^2

- ein Problem der Multivariaten Regression ist, dass es möglich ist, das R^2 zu erhöhen, indem man mehr unabhängige Variablen aufnimmt
- eine Hinzunahme einer weiteren unabhängige Variable führt aber nie zu einer Verkleinerung des R^2
- damit können wir durch Hinzunahme von unabhängigen Variablen das R^2 theoretisch beliebig erhöhen und eine erhöhte Erklärungskraft des Modells vortäuschen
- das ${R_{adj}}^2$korrigiert für diesen Umstand

$$
{R_{adj}}^2 = R^2 - \frac{J \times (1 - R^2)}{N - J - 1}
$$

- es ist genauso zu interpretieren wie das einfache R^2, wird aber insbesondere beim Vergleich verschiedener Modelle angewandt

Stichprobengrösse

- bei einer unabhängigen Variable sollte die Stichprobengrösse mindestens 20 Beobachtungen gross sein
- wenn mehrere unabhängige Variablen vorliegen sollte das Verhältnis Beobachtungen zu Variablen mindestens 5 zu 1 besser 20 zu 1 sein
- wenn die Stichprobe gross ist, z. B. n=1000, reagieren die Signifikanztests übersensibel
  - in der Regel werden alle Koeffizienten statistisch signifikant
  - neben der statistischen Signifikanz sollte daher immer auch die praktische Signifikanz berücksichtigt werden

Multinominale und ordinale unabhängige Variablen

- in die Regressionsanalyse können als unabhängige Variablen metrische Variablen und sogenannte Dummy-Variablen (0/1) aufgenommen werden
- was ist zu tun, wenn die unabhängigen multinominal oder ordinal sind
  - multinominale und ordinale Variablen lassen sich in Dummy-Variablen umkodieren
  - aus einer multinominalen oder ordinalen Variable werden K Dummy-Variablen, wobei K die Anzahl möglicher Ausprägungen ist
  - in die Regressionsanalyse können $K-1$ Dummy-Variablen aufgenommen werden, wobei die weggelassene Variable die Referenzkategorie ist

Beispiel Multivariate Regression

- um die Multivariate Regression mit R zu berechnen nutzen wir das folgende Beispiel
  - wir wollen den Einfluss des Preises, der Anzahl Vertreterbesuche und der Werbeausgaben auf die Absatzmenge untersuchen
  - der Datensatz ist reg_sales.RData (das Beispiel stammt aus Backhaus et al. 2011, die Daten sind leicht verändert)
  - die abhängige Variable ist verkaufte Menge pro Periode in einem Absatzgebiet
  - die unabhängigen Variablen sind der Preis des Produktes, die Ausgaben für Verkaufsförderung und die Zahl der Vertreterbesuche (jeweils in einem Absatzgebiet)
  - die Regressionsfunktion ist

$$\hat{Y} = b_0 + b_1 \times price + b_2 \times \mathrm{expenses} + b_3 \times visits$$

- # lade den Datensatz «reg_sales.RData»

Beispiel Multivariate Regression

- Schritt 1:
  - Haben die unabhängigen Variablen Preis, Ausgaben für Verkaufsförderung und Anzahl Vertreterbesuche einen Einfluss auf die Verkaufszahlen?
  - Null- und Alternativhypothese für das Regressionsmodell
    - ❖ H_0: $\beta_1 = \beta_2 = \cdots = \beta_J = 0$
    - ❖ H_A: $\beta_j \neq 0$ für mindestens ein j
    - ❖ $\alpha = 0.01$
  - Null- und Alternativhypothese für die unabhängigen Variablen
    - ❖ H_0: $\beta_j = 0$
    - ❖ H_A: $\beta_j \neq 0$
    - ❖ $\alpha = 0.05$

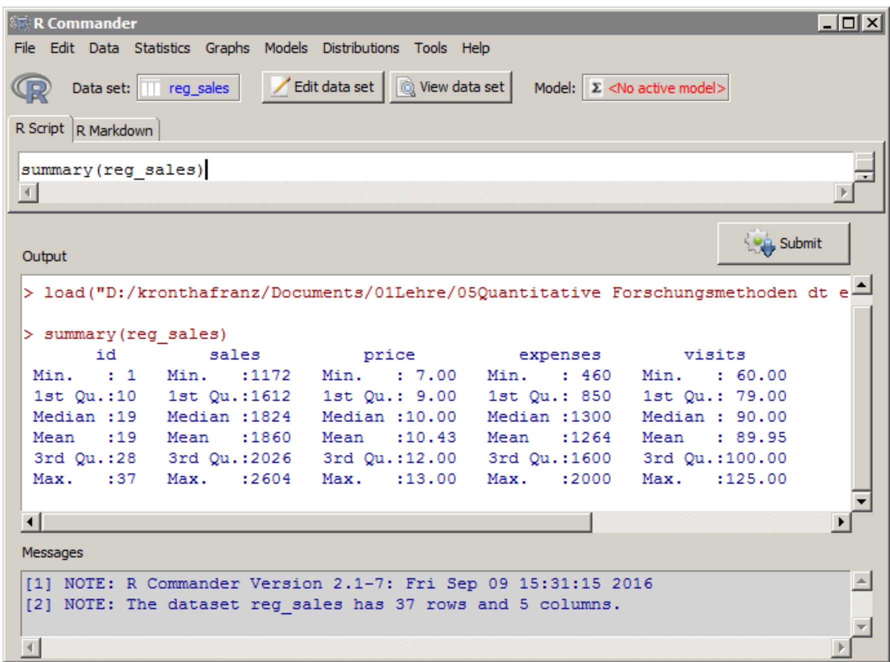
Beispiel Multivariate Regression

– Schritt 2:

- bevor wir die Regressionsanalyse berechnen ist es sinnvoll die deskriptiven Statistiken, die Streudiagramme sowie die Korrelationskoeffizienten zwischen abhängiger und unabhängigen Variablen zu betrachten

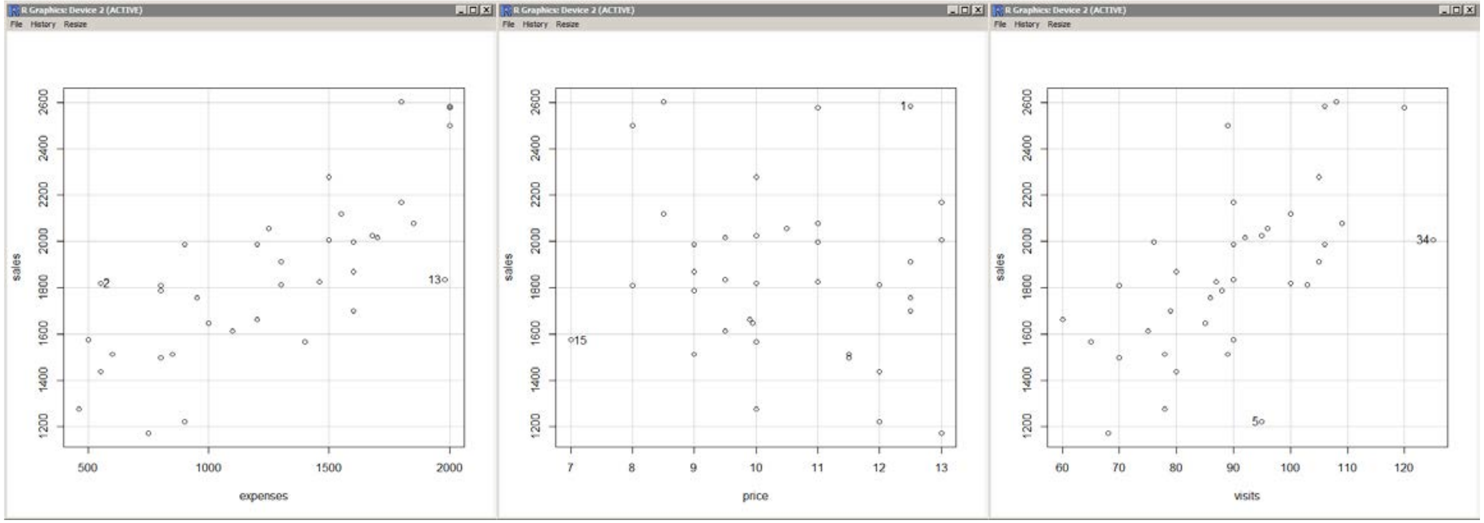
  # nutze die Benutzeroberfläche des R Commanders oder die folgenden Befehle

  summary(reg_sales)



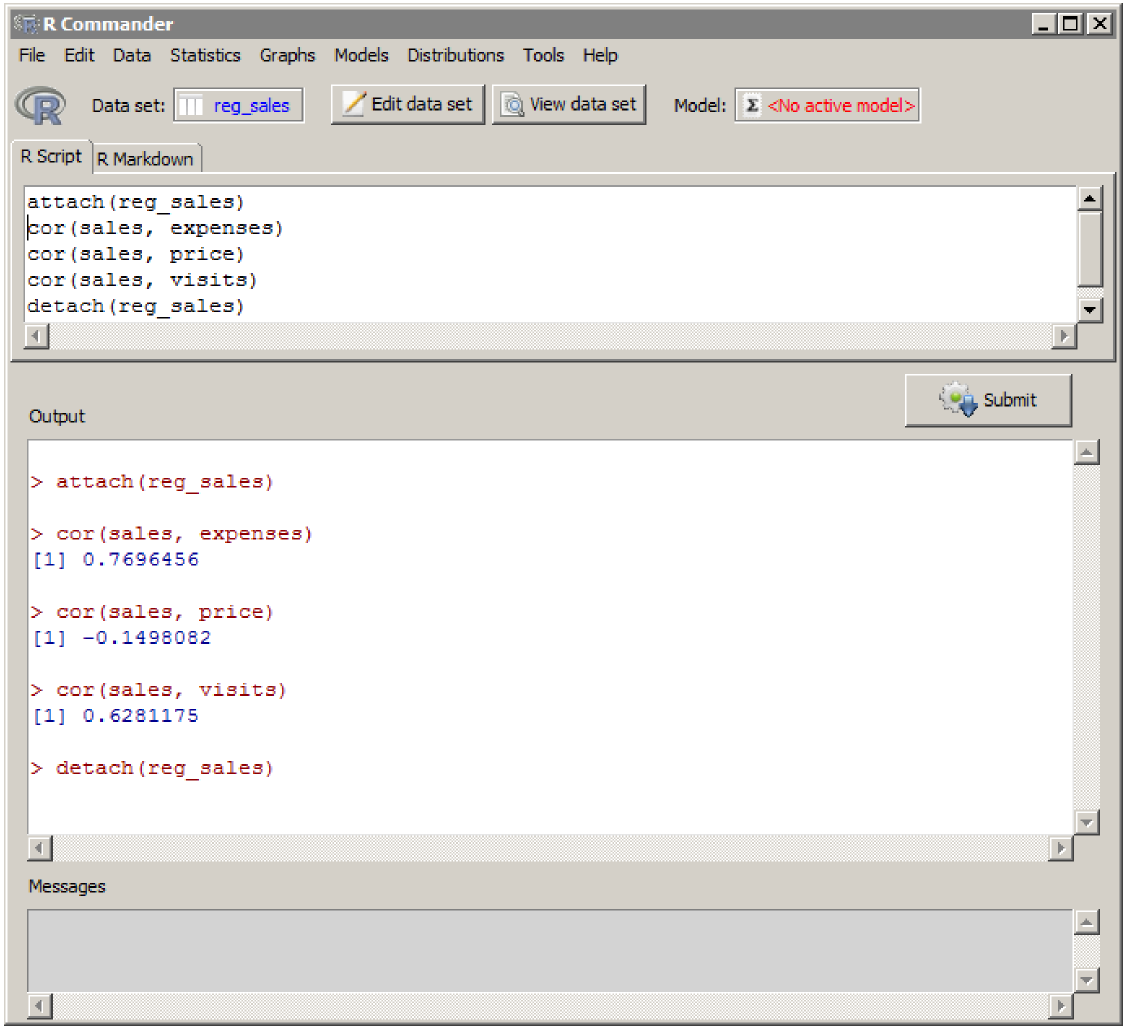
Beispiel Multivariate Regression

```
scatterplot(sales~expenses, reg.line=FALSE, smooth=FALSE, spread=FALSE,
  id.method='mahal', id.n = 2, boxplots=FALSE, span=0.5, data=reg_sales)
scatterplot(sales~price, reg.line=FALSE, smooth=FALSE, spread=FALSE,
  id.method='mahal', id.n = 2, boxplots=FALSE, span=0.5, data=reg_sales)
scatterplot(sales~visits, reg.line=FALSE, smooth=FALSE, spread=FALSE,
  id.method='mahal', id.n = 2, boxplots=FALSE, span=0.5, data=reg_sales)
```



Beispiel Multivariate Regression

```
attach(reg_sales)
cor(sales, expenses)
cor(sales, price)
cor(sales, visits)
detach(reg_sales)
```



Beispiel Multivariate Regression

- Stichprobengrösse
  - ❖ n=37 für alle Variablen (keine fehlenden Werte)
  - ❖ Stichprobe eher klein
- rechne die Regressionsanalyse

```
# nutze die Benutzeroberfläche des R Commanders oder die folgenden Befehle
RegModel.1 <- lm(sales~expenses+price+visits, data=reg_sales)
summary(RegModel.1)
```

Beispiel Multivariate Regression

– Schritt 3: Interpretation der Ergebnisse – F-Statistik

Submit

Output

```
Call:
lm(formula = sales ~ expenses + price + visits, data = reg_sales)

Residuals:
    Min      1Q  Median      3Q     Max
-428.37 -109.88    0.78  133.52  334.56

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 945.26435  247.61619   3.817 0.000563 ***
expenses      0.46890    0.06744   6.953 6.02e-08 ***
price       -55.39382   18.60710  -2.977 0.005417 **
visits        9.99817    2.16386   4.621 5.62e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 178 on 33 degrees of freedom
Multiple R-squared:  0.7771, Adjusted R-squared:  0.7568
F-statistic: 38.34 on 3 and 33 DF,  p-value: 7.324e-11
```

Beispiel Multivariate Regression

– Schritt 3: Interpretation der Ergebnisse – R^2

Output

Submit

```
Call:
lm(formula = sales ~ expenses + price + visits, data = reg_sales)

Residuals:
    Min      1Q  Median      3Q     Max
-428.37 -109.88    0.78  133.52  334.56

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 945.26435  247.61619   3.817 0.000563 ***
expenses      0.46890    0.06744   6.953 6.02e-08 ***
price       -55.39382   18.60710  -2.977 0.005417 **
visits        9.99817    2.16386   4.621 5.62e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 178 on 33 degrees of freedom
Multiple R-squared:  0.7771, Adjusted R-squared:  0.7568
F-statistic: 38.34 on 3 and 33 DF,  p-value: 7.324e-11
```

Beispiel Multivariate Regression

– Schritt 3: Interpretation der Ergebnisse – t-Statistik

Submit

Output

```
Call:
lm(formula = sales ~ expenses + price + visits, data = reg_sales)

Residuals:
    Min      1Q  Median      3Q     Max
-428.37 -109.88    0.78  133.52  334.56

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 945.26435  247.61619   3.817 0.000563 ***
expenses      0.46890    0.06744   6.953 6.02e-08 ***
price       -55.39382   18.60710  -2.977 0.005417 **
visits        9.99817    2.16386   4.621 5.62e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 178 on 33 degrees of freedom
Multiple R-squared:  0.7771, Adjusted R-squared:  0.7568
F-statistic: 38.34 on 3 and 33 DF,  p-value: 7.324e-11
```

Beispiel Multivariate Regression

– Schritt 3: Interpretation der Ergebnisse – Koeffizienten

Output

Submit

```
Call:
lm(formula = sales ~ expenses + price + visits, data = reg_sales)

Residuals:
    Min      1Q  Median      3Q     Max 
-428.37 -109.88    0.78  133.52  334.56 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 945.26435  247.61619   3.817 0.000563 ***
expenses      0.46890    0.06744   6.953 6.02e-08 ***
price       -55.39382   18.60710  -2.977 0.005417 ** 
visits        9.99817    2.16386   4.621 5.62e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 178 on 33 degrees of freedom
Multiple R-squared:  0.7771, Adjusted R-squared:  0.7568 
F-statistic: 38.34 on 3 and 33 DF,  p-value: 7.324e-11
```

Beispiel Multivariate Regression

- Schritt 3: Interpretation der Ergebnisse – Geschätzte Regressionsfunktion
  - die geschätzte Regressionsfunktion ist

    $\hat{Y} = 945.26 + 0.47 * X_1 - 55.39 * X_2 + 9.99 * X_3$

    bzw.

    $\widehat{sales} = 945.26 + 0.47 * expenses - 55.39 * price + 9.99 * visits$

    ❖ das Absolutglied den Wert der abhängigen Variable, wenn alle unabhängigen Variablen den Wert null annehmen

    ❖ die einzelnen Koeffizienten sind die marginale Effekte der jeweiligen Variablen, wenn alle anderen Variablen konstant bleiben

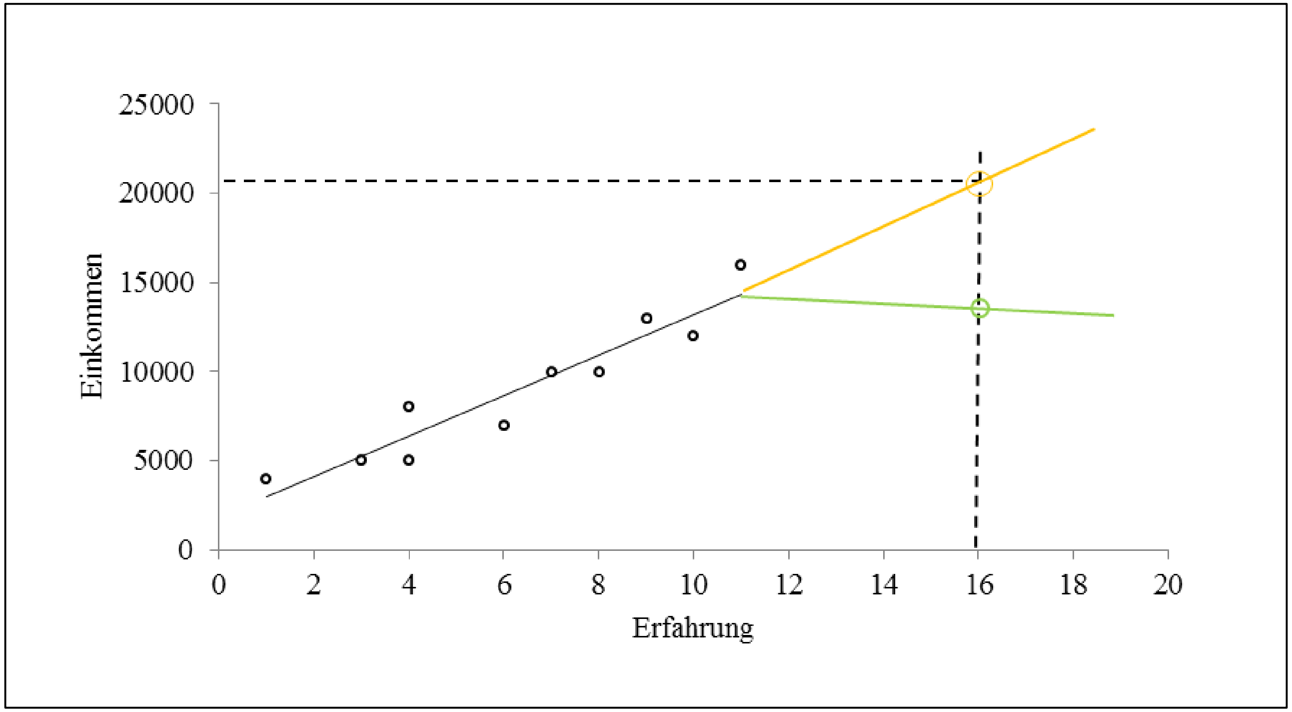
Beispiel Multivariate Regression

– Schritt 3: Interpretation der Ergebnisse – Zusammenfassung

- wir können jetzt angeben, welchen Effekt eine unabhängige Variable auf die abhängige Variable ausübt, unter der Voraussetzung, dass alle anderen Variablen konstant bleiben
- wir können Werte der abhängigen Variable prognostizieren, wenn sich die unabhängigen Variablen verändern oder verändert werten
- Beispiel:
  - ❖ Wie hoch ist die verkaufte Menge pro Periode für ein Absatzgebiet in welchem der Preis des Produktes 15 Euro beträgt, die Ausgaben für Werbung bei 50000 Euro liegen und 200 Vertreterbesuche pro Periode durchgeführt werden?
  - ❖ Wie hoch ist die verkaufte Menge pro Periode für ein Absatzgebiet, wenn der Preis von 15 Euro auf 16 Euro steigt, unter Voraussetzung, dass die Ausgaben für Werbung und die Vertreterbesuche pro Periode gleich bleiben?

Out-of-Sample Vorhersagen

- bei unserem Beispiel haben wir eine Vorhersage getroffen für 200 Vertreterbesuche pro Periode
- die Zahl ist weit von den in der Stichprobe enthaltenen Werte entfernt (vgl. Datensatz)
- solche Prognosen (Out-of Sample Vorhersagen) sind problematisch (vgl. Abbildung)



Anwendungen

- Nutze den Datensatz reg_country.RData (aus dem Lehrbuch von Norusis 2008). Analysiere mit Hilfe der Regressionsanalyse, ob und welchen Einfluss die unabhängigen Variablen «urban, doc, bed, gdp, rad» auf die abhängige Variable «expect» haben.
  - Schreibe die Regressfunktion auf.
  - Berechne die deskriptiven Statistiken.
  - Schätze die Regressfunktion.
  - Interpretiere die Ergebnisse.
- Nutze den Datensatz abortion.RData (zur Verfügung gestellt von Prof. Kahane). Analysiere den Einfluss der unabhängigen Variablen «religion, price, laws, funds, educ, income, picket» auf die abhängige Variable «abortion».
  - Schreibe die Regressfunktion auf.
  - Berechne die deskriptiven Statistiken.
  - Schätze die Regressfunktion.
  - Interpretiere die Ergebnisse.