Johann Wolfgang Goethe-Universität Frankfurt am Main

Fachbereich Mathematik

Institut für Didaktik der Mathematik

L3 Vertiefungsseminar

Anwendungsbezogener Stochastikunterricht

Michael Schneider

Thema 3: Regression kritisch betrachten

Christoph Walther

christoph-walther@gmx.de

Mat.nr.: 3035580

L3 Lehramt (8. Fachsemester)

Tag der Abgabe: 06.05.2009

Inhalt

1 Einleitung

Diese Ausarbeitung soll die Idee für eine Unterrichtssequenz zum Thema „Regression kritisch betrachten" inhaltlich und methodisch darstellen. Die Sequenz wurde nur ansatzweise in der Praxis erprobt, nämlich innerhalb der Seminarstunde „L3 Vertiefungsseminar zum anwendungsbezogenen Stochastikunterricht" vom 30.04.2009 von Mathematikstudenten. Die Seminarstunde war für mich die Basis diese Unterrichtssequenz zu entwickeln.
Im Folgenden werde ich „Lineare Regression" als Sachinhalt darstellen und anschließend die Unterrichtsplanung inklusive Lehrzielen beschreiben. Im letzten Punkt werde ich in einer Reflexion meinen persönlichen Standpunkt zum Verlauf der exemplarischen Stundeninhalte im Seminar darstellen.

2 Lineare Regression als Inhalt

„The simple regression model can be used to study the relationship between two variables." (Woolridge 2006, 24)
Das heißt Regression kann dazu benutzt werden, zwei Variablen in Beziehung zu setzen. Daher können wir den Begriff des Regressionsmodells im Kapitel „funktionale Zusammenhänge" im Bereich der Analysis II im Mathematiklehrplan (Hessen G8) finden. Vor allem im Leistungskurs sollen Möglichkeiten der Annäherung von Funktionen als Anwendungsfeld in Betracht gezogen werden. Hier wird das Regressionsmodell als mathematisches Verfahren herangezogen, bei dem aus konkreten empirischen Daten Näherungsfunktionen gewonnen werden können. Weiter wird die Regression (egal ob quadratisch, linear, exponentiell) als sehr gut für den Unterricht geeignet beschrieben, da diese „theoretisch leicht erarbeitet werden kann und moderne Rechner durchgängig unterschiedliche Regressionsmodelle bereitstellen." (vgl. HKM 2008, 45)
Nach Lehrplan ist dieser Inhalt im Bereich der Anwendung und Vertiefung der Differential- und Integralrechnung nur für den Leistungskurs der Klasse 11 vorgesehen. Wenn man jedoch die Regressionsrechnung dahingehend interpretiert, dass es darum geht generelle Aussagen über Daten zu machen, dann kann dies auch in die Jahrgangsstufe 7 projiziert werden. Hier geht es im Abschnitt der beschreibenden Statistik unter anderem auch um die Interpretation von statistischen Angaben im realen Kontext (Aussagekraft von Statistiken und deren Bewertung) (vgl. HKM 2008, 19). Gerade das Hinterfragen von Statistiken und deren Aussage ist ein wichtiger Bereich im Mathematikunterricht.

Mathematisch soll bei der linearen Regression das Problem gelöst werden, dass bei n Punkten mit den Koordinaten $(x_i, y_i) \in R$, $i = \{1,...,n\}$, $n \geq 2$, die näherungsweise auf einer Geraden liegen, eine Regressionsgerade der Form $y = mx + b$ gefunden werden soll, die die Punkte „am besten" annähert. Auf dem Arbeitsblatt 1 (Lückentext – Lineare Regressionsgerade berechnen) wird die Lösung über die Methode „der kleinsten Quadrate" (ordinary least squares) näher beschrieben.
Schüler können diese Methode entweder mathematisch erfassen, wie es auf dem Arbeitsblatt gemacht wurde, oder experimentell ausprobieren, die Fehlerquadrate über Schiebregler in Excel (vgl. Abbildung 1) zu minimieren, um die Idee zu verstehen. Andere Methoden, wie die logistischer Regression oder die gewichtete lineare Regression werden hier nicht thematisiert.
Beide Ansätze, der mathematisch herleitende, sowie der experimentelle Ansatz haben Vor- und Nachteile. Für Schüler ist die mathematische Herleitung sicher nicht immer direkt einsichtig, da viele Umformungen und Formeln bekannt sein müssen. Beispiele dafür sind allgemeine Formeln für Mittelwert, Varianz und Kovarianz. Außerdem muss das Ableiten von Summen und zusammengesetzen Funktionen beherrscht werden. Ich habe jedoch den mathematischen Ansatz gewählt, um die Formeln für die Steigung und den Achsenabschnitt der Regressionsgeraden nicht aus dem „Nichts" entstehen zu lassen. Dieses Vorgehen ist jedoch höchstens für den Leistungskurs relevant.

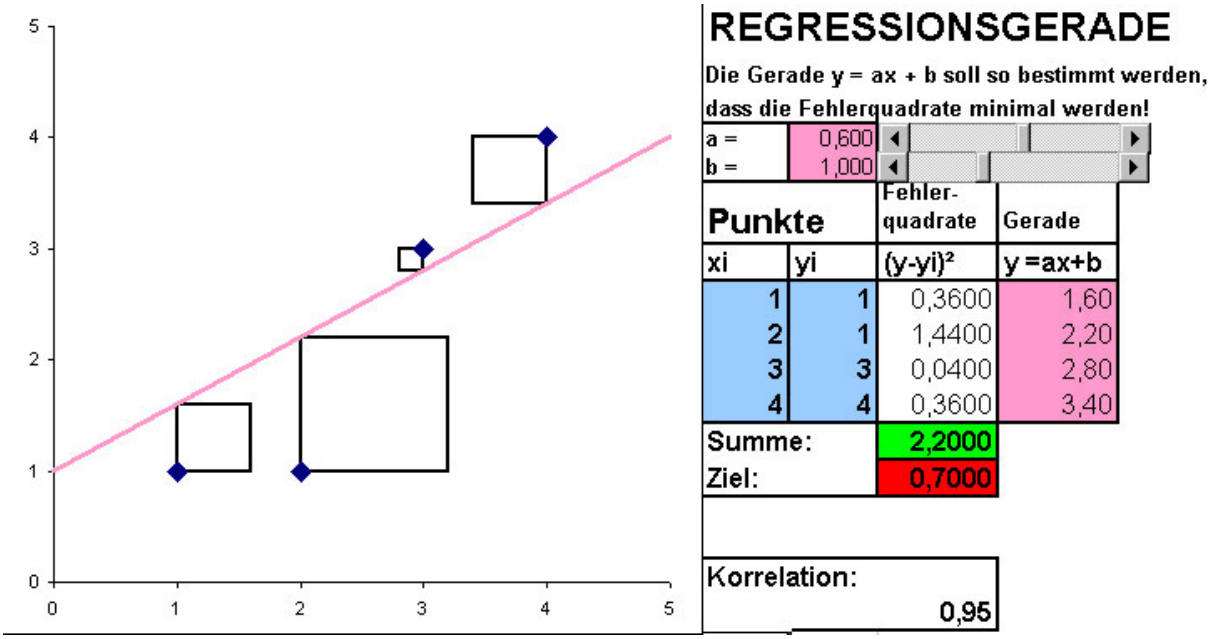


Abbildung 1 Excel-Sheet zum experimentellen entdecken der Regressionsgerade.

Paul stellt in seinem Spiel „Findet die Regressionsgerade" einen experimentellen Zugang zu kleinste Fehlerquadrate Schätzung dar (vgl. Abbildung 1). Dieser

Ansatz verdeutlicht sehr gut, wie der mathematische Ansatz zum Finden einer Regressionsgeraden aussieht.

In einem von mir erstellten Beispiel (siehe Anhang), sollen die mathematischen Kenntnisse, wie man Achsenabschnitt und Steigung berechnet, anschließend praktisch angewandt werden. Es ist eine Nachfrage-Kurve zu erstellen. Das Beispiel ist sowohl als Arbeitsblatt, wo die Rechentabelle händisch ausgefüllt werden kann, oder als Excel-Blatt vorhanden.

Ein weiterer wichtiger Abschnitt der Unterrichtsequenz ist das Interpretieren von Regressionsgeraden, dazu haben die Schüler die Aufgabe Daten vom PC-Konsum in Stunden pro Monat mit den zugehörigen Schulnotendurchschnitten in Beziehung zu setzen. Die Daten sollen visualisiert und mit einer Regressionsgeraden versehen werden. Dazu erfahren die Schüler, wie man die Berechnung der Regressionsgeraden in Excel sehr schnell machen kann (Funktionen: ACHSENABSCHNITT(y-Werte; x-Werte) und STEIGUNG(y-Werte; x-Werte)). Zusätzlich kann mit wenigen Klicks eine lineare Trendgerade im Diagramm angezeigt werden, die auch über das Verfahren der linearen Regression berechnet wird.

Das entstandene Diagramm (siehe Abbildung 2) gibt sicherlich Diskussionsstoff. Fragen an die Schüler könnten sein, welche Schulnoten sie denn für ihren PC-Konsum bekommen würden?

Im Seminar sind sehr schnell die Begriffe „Korrelation" und „Kausalität" gefallen und auch Schüler werden einen Zusammenhang sehr schnell erkennen, aber auch in Frage stellen. Unter „Korrelation" wird im Allgemeinen ein normiertes Maß für den linearen Zusammenhang zweier Variablen verstanden (vgl. Rinne 2003, 76). Eine Assoziation zwischen den beiden Variablen Schulnote und PC-Konsum ist sicher schnell gefunden, d.h. für die Schüler zum Beispiel: „Wenn jemand viel Computer spielt, dann bekommt diese Person schlechte Noten in der Schule". Ein klarer kausaler Zusammenhang im Sinne einer Wenn-Dann-Beziehung wird mit dem linearen Zusammenhang der Variablen gleich gestellt (vgl. Wooldridge 2006, 13). Dennoch wird es schnell Einsprüche gegen diese oben genannte These geben, die es dann zu diskutieren gilt.

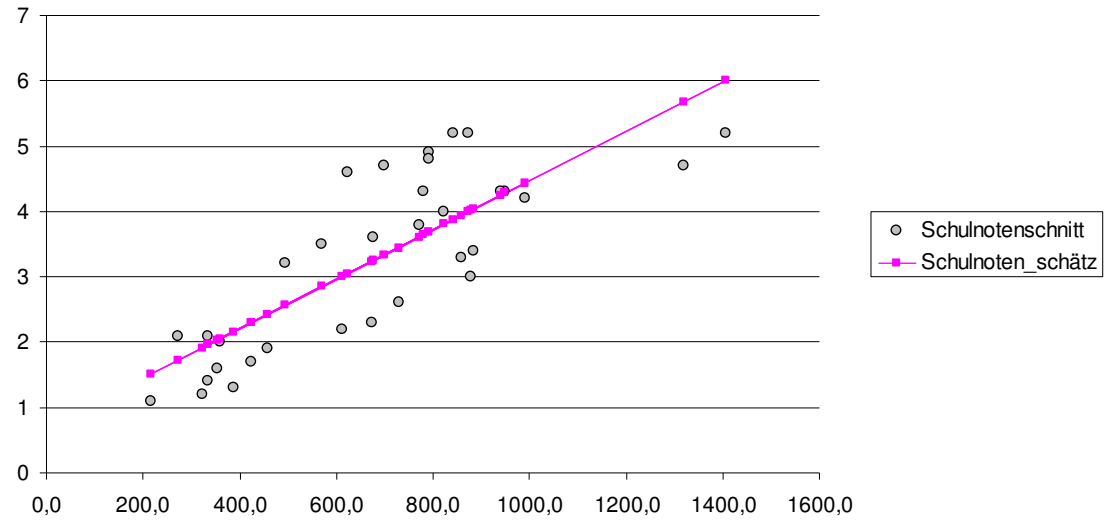


Abbildung 2 Zusammenhang zwischen Notendurchschnitt (y-Achse) und PC-Konsum in Stunden pro Monat (x-Achse).

Dieser Fall von irreführender Statistik gibt Anlass sich weiteren Zusammenhängen zu widmen. Dazu dient das Arbeitsblatt zur „Korrelation vs. Kausalität". Friedrich (2006) hat hier eine Reihe von Diagrammen zusammengestellt, die die Schüler auf ihre Aussagekraft hin beurteilen sollen, diese habe ich auf zwei Arbeitsblätter aufgeteilt, die verschiedene Problematiken betonen: Einerseits das Problem der Inhomogenität von Datensätzen (Männer und Frauen in einer Statistik über die Schuhgröße vs. Einkommen) und andererseits der Einfluss von Hintergrundvariabeln (Kriminalität und Ausländeranteil, welche jedoch beide von der Größe einer Region beeinflusst werden können). Die beiden Aspekte sollen anhand der Diagramme erkannt und verdeutlicht werden. Dies ist besonders wichtig, da der Schluss von der Korrelation zweier Variablen auf einen kausalen Zusammenhang einer der häufigsten und auch trügerischsten in der Statistik ist. (vgl. Krämer 2008, 165)

Einen weiteren Effekt der Regressionsrechnung, der ihr auch ihren Namen verleiht, behandeln Engel und Vogel (2002) – Den Regressionseffekt. Die beiden Autoren zeigen einen handlungsorientieren Ansatz mit einem Kartenexperiment auf, um den Effekt zu verdeutlichen.

Das Kartenexperiment kann auch mit Skibbokarten durchgeführt werden, diese sind von 1-10 nummeriert. Es gibt zwei Stapel mit jeweils vier Karten von eins bis zehn. Der eine Stapel dient als Zufallsfaktor, der bei den jeweiligen Runden hinzuaddiert wird. Zuerst bekommt die eine Hälfte der Klasse eine zufällige Karte von 1 bis 5, die andere Hälfte bekommt einen Zufallswert von 6 bis 10. Dies sind die Ausgangswerte der Schüler. Nun wird jeweils eine Karte vom „Zufallseinfluss-Stapel" gezogen und zum Ausgangswert hinzuaddiert. Es lässt sich die Tendenz

zu einer sehr flachen Regressionsgeraden ausmachen, da auch Schüler mit schlechten Ausgangskarten „zufällig" sehr gut sein können, im Gegensatz zu Schülern mit sehr guten Ausgangskarten, die auch zufällig sehr schlecht sein können. Das Experiment gibt Anlass dieses Phänomen weiter auszuprobieren. Dazu habe ich das Lisp-Programm von Engels und Vogel in Excel umgesetzt. Hier kann zu den wahren Ausgangswerten jeweils ein mehr oder weniger starker Zufallswert hinzuaddiert werden – die Steigung der Regressionsgeraden ändert sich dementsprechend (vgl. Abbildung 3).

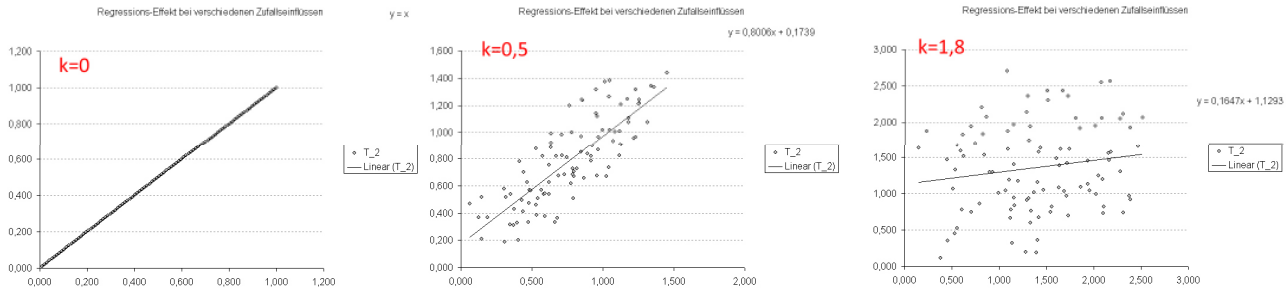


Abbildung 3 Änderung der Steigung der Regressionsgeraden je nach Stärke des Zufallsfaktors (k).

Die Theorie dieses Phänomens kann wiederum nur im Leistungskurs näher betrachtet werden, dazu habe ich ein Blatt zur Theorie des Regressionseffekts im Anhang beigelegt. Im Grundkurs kann dieser Effekt sicher auch betrachtet werden, hier bietet es sich, wie auch im LK, an weitere Beispiele für diesen Effekt zu suchen.

3 Unterrichtsplanung

3.1 Einführung in die Regressionsrechnung (90min)

3.1.1 Lernziele

A Kognitive Lernziele

Die Schüler sollen …die Problemstellung der Regressionsanalyse kennen lernen.
…verstehen das Prinzip der Minimierung der Fehlerquadrate.
…können mit Excel oder vorgegebner Tabelle die Koeffizienten der Regressionsgeraden errechnen.

B Sozial-affektive Lernziele

Die Schüler sollen …Aufgaben in Partnerarbeit lösen und sich dabei unterstützen.

3.1.2 Stundenverlauf (90min)

Phase	Inhalt	Methode/ Sozialform
Einstieg	Ausgangsproblem: Wir wollen die Beziehung zwischen zwei Variablen untersuchen. Ziel: Aussagen machen Idee: Linearen Zusammenhang suchen und darstellen.	Lehrer-Schüler Gespräch
Problemstellung	Spiel: Anpassen der Geraden an die Daten. Mit dem Ziel eine möglichst geringe Fläche an Fehlerquadraten zu erreichen.	Experimenteller Zugang zur Methode der kleinsten Quadrate-Schätzung. Schüler in Partnerarbeit.
Mathematische Herleitung	Lückentext zur linearen Regression	Schüler in Gruppenarbeit (max. 4 Personen)
Ergebniskontrolle	Übung: Eine Zahl 60 soll so in 2 Summanden a,b zerlegt werden, dass das Produkt aus erstem und Quadrat des zweiten Summanden maximal wird.	Schüler schreibt/schreiben an die Tafel. Lehrer-Schüler Gespräch
Vertiefung	Beispiel zur Regression rechenen. Nachfragekurve von Getränken in Flaschen	Schüler arbeiten selbstständig. Lösungen können beim Lehrer erfragt werden. L läuft herum, beantwortet Fragen.
Abschluss/ Reflexion	Was haben wir heute gemacht? Eine Kurzzusammenfassung	Schüler im Plenum

3.2 Interpretation von Regressionsgeraden – Kausalität vs. Korrelation (90min)

3.2.1 Lernziele

A Kognitive Lernziele

Die Schüler sollen …den Unterschied zwischen Korrelation und Kausalität erkennen.

…über graphische Ergebnisse einer Regression diskutieren und diese kritisieren

B Sozial-affektive Lernziele

Die Schüler sollen …Aufgaben in Partnerarbeit lösen und sich dabei unterstützen.

…diskutieren und kritisieren, dabei Kritikfähig sein.

3.2.2 Stundenverlauf (90min)

Phase	Inhalt	Methode/ Sozialform
Einstieg	Wiederholung: Wie errechnet man eine Regressionsgerade (in Excel)	Lehrer-Schüler Gespräch
Problemstellung	Ausgangsproblem: Visualisierung und Schätzung der Regressionsgeraden für die	Handlungsorientierter Zugang zum Problem des kausalen Zusammenhangs vs. der

	Variablen PC-Konsum und Schulnotenschnitt	Korrelation zweier Variablen.
Diskussion	Diskussion der Ergebnisse	Gruppendiskussion in Rollen: Gruppe 1 ist der Meinung, dass Schulnoten und PC-Konsum kausal zusammenhängen, Gruppe 2 nicht.
Arbeitsphase	Bearbeiten der Arbeitsblätter „Korrelation vs. Kausalität".	Schüler in Partnerarbeit. Lehrer ist als Berater unterwegs.
Diskussion	Besprechung der Ergebnisse	Plenum
Abschluss/ Reflexion	Was haben wir heute gemacht? Zusammentragen der Ergebnisse (Inhomogenität, Hintergrundvariablen als Störfaktoren)	Lehrer-Schüler Gespräch, Tafel

3.3 Der Regressionseffekt (45min)

3.3.1 Lernziele

A Kognitive Lernziele

Die Schüler sollen …den Regressionseffekt kennen lernen.

…den Einfluss von Zufallsgröße auf Test in Excel ausprobieren.

…Beispiele für den Regressionseffekt finden.

B Sozial-affektive Lernziele

Die Schüler sollen …ein Kartenexperiment gemeinsam durchführen.

3.3.2 Stundenverlauf (45min)

Phase	Inhalt	Methode/ Sozialform
Einstieg	Wiederholung: Welche Probleme können bei der Interpretation von Regressionsgeraden auftreten?	Lehrer-Schüler Gespräch
Problemstellung	Wieso heißt Regression eigentlich „Regression"? => Es gbit den Regressioneffekt	Lehrervortrag zu geschichtlichen Hintergründen (vgl. Engels & Vogel 2002, 1)
Experiment	Durchführen des Kartenexperiments Und Auswerten der gesammelten Daten in Excel.	Auswertung in Gruppen, um Diskussion über die Ergebnisse anzuregen.
Diskussion	Welchen Einfluss haben die beiden Werte (Ausgangskarte vs. Zufallskarte) auf das Ergebnis?	Diskussion über das Ergebnis/Plenum
Abschluss/ Reflexion	Weitere Beispiele für den Regressionseffekt finden. Zusammentragen der Ergebnisse	Lehrer-Schüler Gespräch, Tafel

4 Reflexion

Lineare Regression wird vor allem in der Ökonometrie verwendet, um Zusammenhänge von Variablen zu modellieren (zum Beispiel Einkommen in Abhängigkeit von Geschlecht, Alter, Ausbildung und Hautfarbe). Hier ist die Unterscheidung zwischen Korrelation und Kausalität besonders wichtig, daher finde ich es wichtig, dass man darüber diskutieren kann.

Krämer (2008) macht in seinem Buch sehr deutlich, wie leicht und überzeugend man mit Statistiken lügen kann, gerade deshalb ist es wichtig schon in der Schule auf diese Probleme aufmerksam zu machen. In der Seminarstunde ist mir aufgefallen, wie schwer es selbst einigen Studenten gefallen ist, Korrelation sauber von Kausalzusammenhängen zu trennen, auch der Lückentext ist einigen sichtlich schwer gefallen. Hier ist auch der größte Kritikpunkt an diesem Stundenentwurf. Er basiert sehr stark auf dem Theorieverständnis, dass bei den Schülern entwickelt werden soll, dazu sind anders als im Lehrplan (siehe Seite 3) beschrieben, nicht immer leichte mathematische Schritte zu gehen. Ich denke nicht, dass ich diesem Lückentext, so wie ich ihn in der Seminarstunde verwendet habe, an eine Schülergruppe treten könnte. Hier ist sicher der experimentelle Ansatz von großer Bedeutung. Dennoch habe ich als Schüler immer „vom Himmel gefallene" Formeln gehasst. Ich hatte das Gefühl etwas nur halb gelernt zu haben.

Ich wäre nun sehr gespannt, ob dieser Stundenentwurf nur näherungsweise in der Praxis bestehen könnte, sicher ist jedoch, dass er noch an einigen Stellen verbessert werden kann.

Leider zu spät für meine Planungen habe ich die Seite http://www.learn-line.nrw.de/angebote/neuemedien/medio/mathe/stochastik/lunge/luvol01.htm entdeckt. Hier ist ein Umfangreiches Projekt zum Thema „Standardisiertes Lungenvolumen als Beispiel für eine statistische Modellbildung" eingestellt. Hier wird sehr schön gezeigt, wie man eine Stunde auf der Basis des Programms Fathom! als Werkzeug machen kann.

5 Literatur

Engel, Joachim, Vogel, Markus (2002). Versinken wir alle im Mittelmaß. Zum Verstehen des Regressionseffekts. Zugriff am 22.04.2009 unter http://www.studienseminare-ge-gym.nrw.de/K/riemer/mathematik/fachseminar/literatur/stochastik/mittelmass-mnu-2002.pdf

Friedrich, Nina (2006). Unterrichtsentwurf für den vierten Unterrichtsbesuchim Fach Mathematik. Zugriff am 30.04.2009 unter http://www.studienseminare-ge-gym.nrw.de/K/riemer/mathematik/fachseminar/entwuerfe/friedrich-regression-gk-11.pdf

Hessisches Kultusministerium (2008). Lehrplan Mathematik Gymnasialer Bildungsgang. Jahrgangsstufen 5G bis 12G. Zugriff am 29.04.2009 unter http://www.kultusministerium.hessen.de/irj/servlet/prt/portal/prtroot/slimp.CMReader/HKM_15/HKM_Internet/med/016/016060d2-2a4e-b115-3a16-e91921321b2c,22222222-2222-2222-2222-222222222222,true.pdf

Krämer, Walter (2008). So lügt man mit Statistik (11. Aufl.). München: Serie Piper

Paul, Markus. Spiel: Finde die Regressionsgerade. Zugriff am 5.05.2009 unter http://www.ammu.at/archiv/18/18_41.htm.

Rinne, Horst (2003). Taschenbuch der Statistik (3. Aufl.) Frankfurt am Main: Verlag Harry Deutsch.

Wooldridge, Jeffrey (2006). Introductory Econometrics. A Modern Approach (3. Aufl. International Student Edition). Toronto: Thomoson South-Western.

6 Anhang

Lückentext – Lineare Regressionsgerade berechnen

Bei n Punkten mit den Koordinaten $(x_i, y_i) \in R$, $i = \{1,...,n\}$, $n \geq 2$ die näherungsweise auf einer Geraden liegen kann eine Regressionsgerade der Form $y = mx + b$ gefunden werden, die die Punkte „am besten" annähert.

Jeder Punkt kann mit der Gleichung $y_i = \beta_0 + \beta_1 x_i + u_i$ beschrieben werden.

Wobei u_i der Fehlerterm ist um den jeder Punkt von der Regressionsgeraden abweicht. Eine wichtige Annahme für das Modell ist, dass $E[u] =$ [ ] und $COV(X,u) = 0$ ist. D.h. eine systematische Verzerrung soll ausgeschlossen sein.

$y_i - \beta_0 - \beta_1 x_i = u_i$ bezeichnet man als Residuum, dies ist der Abstand jedes Datenpunktes von [ ]. Man bestimmt die Gerade, so dass die Abweichung der y_i von den geschätzen Werten $\hat{y}_i$ möglichst gering wird. Die Minimierungsaufgabe für die Methode der kleinsten Quadrate (OLS) lautet:

$$(1)\ \sum_{i=1}^{n} (u_i)^2 = \min_{\beta_0, \beta_1} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$

Wir betrachten β_1 als fest und betrachten die Funktion

$$f(\beta_0) = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 .$$

Aus $f'(\beta_0) =$ [$= 0$] erhält man (β_0 in Abhängigkeit von β_1)

$$\beta_0 = \bar{y} - \beta_1 \bar{x}$$

mit $\bar{x} = \frac{1}{n} \sum_i x_i$ und $\bar{y} = \frac{1}{n} \sum_i y_i$ als arithmetisches Mittel von y_i, x_i. Den Wert für β_0 setzen wir in (1) ein und erhalten

$$(2)\ \sum_{i=1}^{n} (y_i - \beta_1 x_i - \bar{y} + \beta_1 \bar{x})^2$$

Dadurch lässt sich eine Funktion $g(\beta_1) = \sum_{i=1}^{n} (y_i - \bar{y} - \beta_1 (x_i - \bar{x}))^2$ definieren. Davon berechnen wir ebenfalls das Minimum. Durch ausmultiplizieren und sortieren nach den Potzenzen von β_1 ergibt sich

$$g(\beta_1) = \underbrace{\sum_{i=1}^{n} (y_i - \bar{y})^2}_{=:Var(Y)} - 2\beta_1 \underbrace{\sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x})}_{=:COV(X,Y)} + \beta_1^2 \underbrace{\sum_{i=1}^{n} (x_i - \bar{x})^2}_{=:Var(X)}$$

Mit den Abkürzungen für die Varianzen und die Kovarianz erhalten wir

$$g(\beta_1) = Var(Y) - 2\beta_1\, COV(X,Y) + \beta_1^2\, Var(X)$$

Mit der Ableitung von $g(\beta_1)$ erhalten wir $g(\beta_1)' =$ [ ].

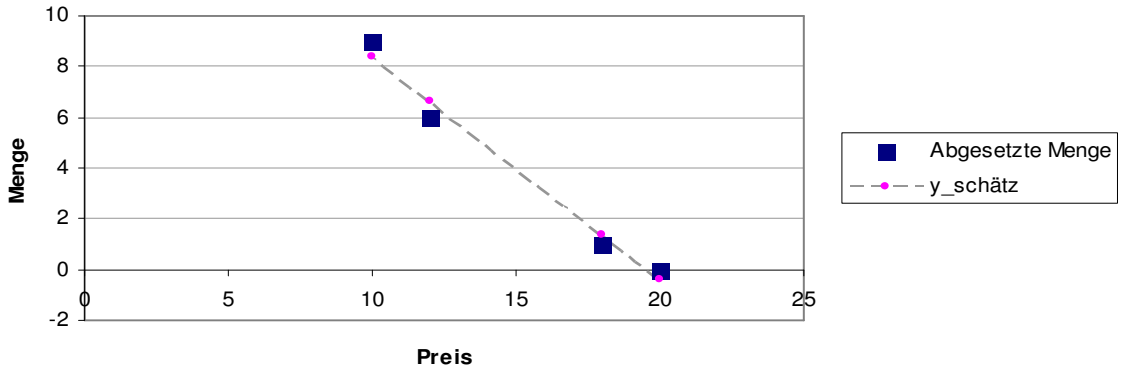
Aus $g(\beta_1)' = 0$ ergibt sich für das Minimum

$$\beta_1 = \quad = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2}$$

Aufgabe: Regressionsgerade errechnen

Errechnet zu den folgenden Daten eine **Regressionsgerade**, die über den Flaschenpreis die abgesetzte Menge vorhersagt.

Wie lautet die **Geradengleichung** ?



i	Preis x_i	Menge y_i	$x-\bar{x}$	$y-\bar{y}$	$(x_i-\bar{x})(y_i-\bar{y})$	$(x_i-\bar{x})^2$
1	20	0	5	-4	-20	25
2	18	1	3	-3	-9	9
5	12	6	-3	2	-6	9
6	10	9	-5	5	-25	25
Summe	60	16	0	0	-60	68
Mittelwert	15	4				

Lösung:

$\beta_1 = -0.88$

$\beta_0 = 17.235$

$\hat{y}_i =$

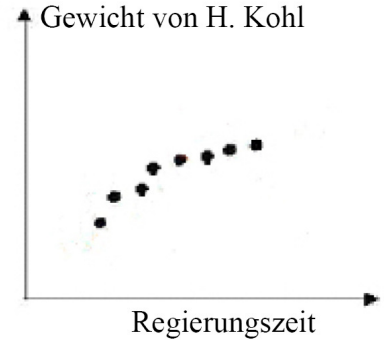
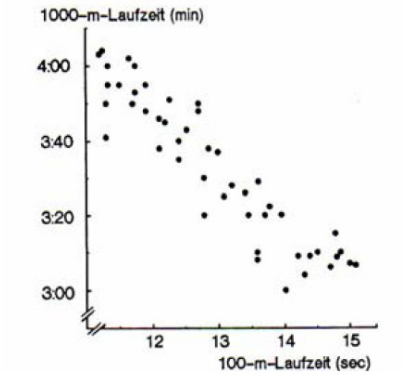
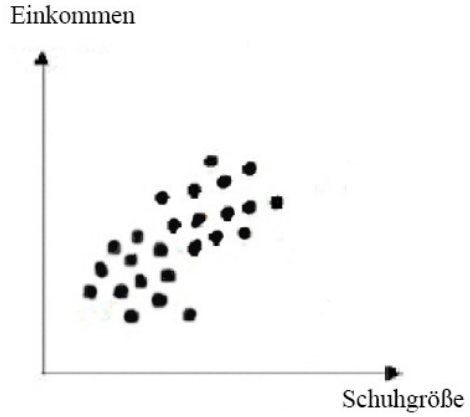
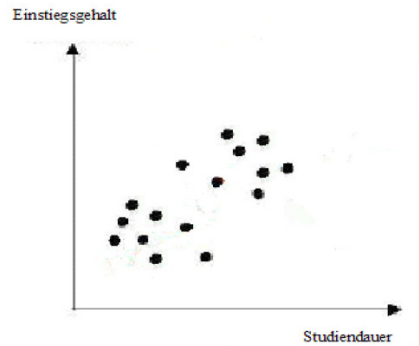
(vgl. http://de.wikipedia.org/wiki/Regressionsanalyse#Berechnung_der_Regressionsgeraden)

Aufgabe: Stellt euch die Daten mit Regressionsgeraden in einem Diagramm dar:
Schätzt die Regressionsgerade mit den Excel-Funktionen Achsenabschnitt() und Steigung().
Interpretiert das Ergebnis: Wie viele Stunden sitzt ihr vor dem PC ? Vergleicht mit eurer Schulnote!

Stichprobe		
PC Konsum (h/Woche)	Schulnotenschnitt	Schulnoten_schätz
270,9	2,1	
791,2	4,9	
858,7	3,3	
883,0	3,4	
728,5	2,6	
214,9	1,1	
1405,8	5,2	
1318,8	4,7	
698,0	4,7	
822,9	4	
771,9	3,8	
991,7	4,2	
624,3	4,6	
333,8	1,4	
673,3	2,3	
950,1	4,3	
841,9	5,2	
611,1	2,2	
422,5	1,7	
877,6	3	
495,1	3,2	
334,0	2,1	
780,4	4,3	
941,6	4,3	
458,9	1,9	
871,8	5,2	
791,0	4,8	
359,7	2	
676,8	3,6	
388,1	1,3	
352,7	1,6	
570,5	3,5	
322,3	1,2	

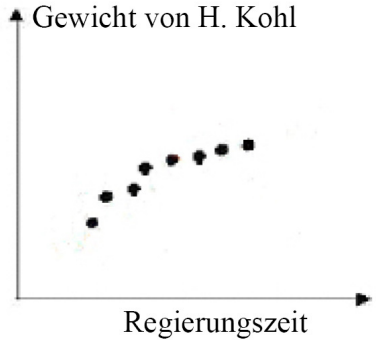
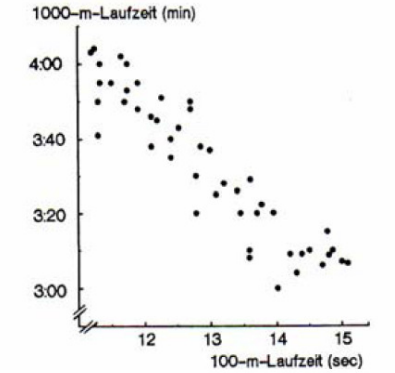
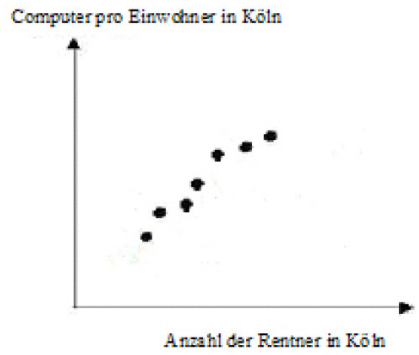
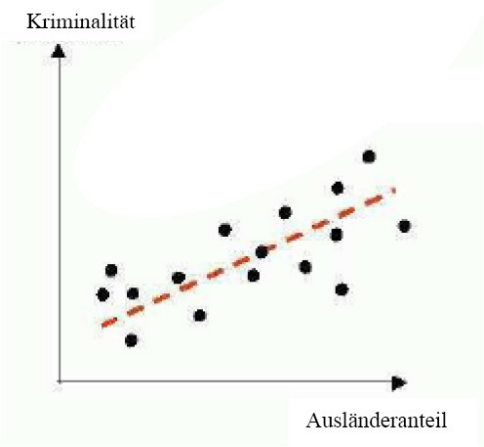
Achsenabschnitt : =ACHSENABSCHNITT (Schulnotenschnitt ;PC-Konsum)
Steigung : =STEIGUNG (Schulnotenschnitt ;PC-Konsum)

Beurteilt die folgenden Darstellungen bezüglich des Problems **„Korrelation vs. Kausalität"**. Schreibt kurz auf, welche Probleme bei den einzelnen Darstellungen auftreten könnten. Veranschaulicht die Probleme ggf. im Diagramm oder durch weitere Diagramme.

http://www.studienseminare-ge-gym.nrw.de/K/riemer/mathematik/fachseminar/entwuerfe/friedrich-regression-gk-11.pdf

Beurteilt die folgenden Darstellungen bezüglich des Problems **„Korrelation vs. Kausalität"**. Schreibt kurz auf, welche Probleme bei den einzelnen Darstellungen auftreten könnten. Veranschaulicht die Probleme ggf. im Diagramm oder durch weitere Diagramme.

http://www.studienseminare-ge-gym.nrw.de/K/riemer/mathematik/fachseminar/entwuerfe/friedrich-regression-gk-11.pdf

Auswertung des Kartenexperiments

			Test		Retest	
	Gruppe	"Wahres Können"	Zufälliger Einfluss	Summe T1	Zufälliger Einfluss	Summe T2
1	A					
2	A					
3	A					
4	A					
5	A					
6	A					
7	A					
8	A					
9	B					
10	B					
11	B					
12	B					
13	B					
14	B					
15	B					
16	B					

	Gruppe A	Gruppe B
Mittelwert "Wahres Können"		
Mittelwert "T1"		
Mittelwert "T2"		

Zur Theorie des Regressionseffektes

Beobachtung: Die Mittelwerte der beiden Extremgruppen bewegen sich aufeinander zu. D.h. die Steigung der Regressionsgerade ist kleiner als Null.
Modell:
Die Zufallsvariablen X, Y, T repräsentieren die Ergebnisse bei Vor- und Nachtest, sowie die wahre Fähigkeit.
$\Rightarrow X = T + u_1$ und $Y = T + u_2$

Mit dem Koeffizienten der der Steigung $\beta_1 = \frac{COV(X,Y)}{Var(X)}$ ergibt sich:

$$Var(x) = Var(T) + Var(u_1)$$

$$\begin{aligned} COV(X,Y) &= E[(X - E[X])(Y - E[Y])] \\ &= E[(T - E[T] + u_1)(T - E[T] + u_2)] \\ &= \underbrace{E[T - E[T]]}_{Var(T)} \left(1 + \underbrace{E[u_1] + E[u_2]}_{=0}\right) \\ &= Var(T) \end{aligned}$$

Für die Steigung der Regressionsgeraden ergibt sich also

$$\beta_1 = \frac{Var(T)}{Var(T) + Var(u_1)}.$$

Bei unterschiedlichen Intensitäten des Zufallseinflusses, d.h. $\Rightarrow X = T + k \cdot u_1$ und $Y = T + k \cdot u_2$ ergibt die Steigung

$$\beta_1 = \beta_1(k) = \frac{Var(T)}{Var(T) + k^2 \cdot Var(u_1)}.$$

Wie entwickelt sich β_1 bei der Veränderung von k? Betrachte $\lim_{k \to \infty} \beta_1(k)$ und $\lim_{k \to 0} \beta_1(k)$.

Engel, Joachim, Vogel, Markus (2002). Versinken wir alle im Mittelmaß. Zum Verstehen des Regressionseffekts. Zugriff am 22.04.2009 unter http://www.studienseminare-ge-gym.nrw.de/K/riemer/mathematik/fachseminar/literatur/stochastik/mittelmass-mnu-2002.pdf