

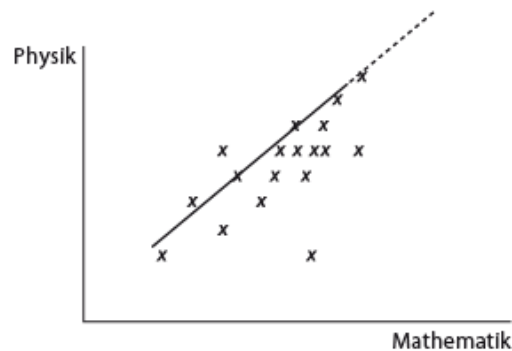
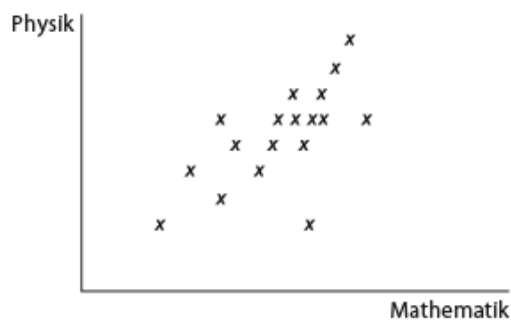
Lineare Regression

Zielsetzung

Es soll geprüft werden, wie eine zweidimensionale Reihe von Daten (zwei Variablen) verwendet werden kann, um den Wert von einer der Variablen für einen gegebenen Wert der anderen zu schätzen.

Erklärung der Aufgabenstellung

Für viele zweidimensionale Verteilungen scheint es eine Beziehung zwischen den zwei Datenmengen zu geben. Wenn wir z. B. die Noten der Kursteilnehmer einer Physikprüfung mit denen vergleichen, die dieser Kurs in der Mathematikprüfung erreicht hat, wird deutlich, dass ein Kursteilnehmer mit guten Noten in Physik häufig auch gute Noten in Mathematik hat. Es scheint eine **positive Wechselbeziehung (positive Korrelation)** zwischen diesen Noten zu geben. Werden die Noten für jeden Kursteilnehmer in einem Diagramm dargestellt, dann kann die Linie, die am besten zur Streuung der Noten passt, benutzt werden, um den Wert einer Note durch eine andere bekannte Note zu schätzen.



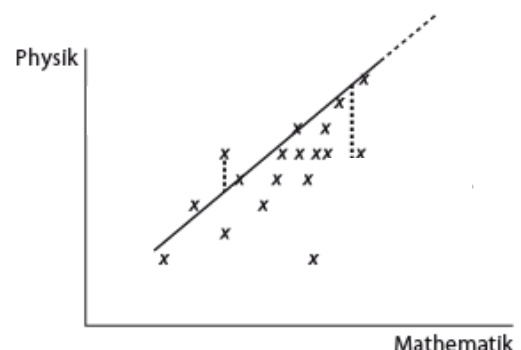
Das Plotten der Prüfungsnoten ergibt ein **Streudiagramm**.

Die Linie ist der Graph der linearen Regression. Wir können den Taschenrechner benutzen, um seine Gleichung zu berechnen.

Die lineare Regressionslinie für die Gleichung einer Geraden lautet:

$$y = a + b x.$$

Die Position der Geraden auf dem Streudiagramm kann so gezeichnet werden, dass, wenn wir alle einzelnen Abweichungen oberhalb der Linie und alle einzelnen Abweichungen unterhalb der Linie quadrieren, die Summe all dieser Quadrate so klein wie möglich ist.



Dieses ist bekannt als die **Methode der kleinsten Quadrate**.

Dieses Verfahren verwendet der Taschenrechner.

Wenn eine Reihe zweidimensionale Daten richtig eingegeben worden ist, dann kann der Taschenrechner benutzt werden, um die Werte von **a** und **b** zu berechnen, die die Gleichung der Regressionslinie festlegen. Damit können wir dann Schätzungen für die Variable y bilden, wenn wir einen Wert für die Variable x kennen.

Diskussionsansatz für den Unterricht


Da die lineare Regression in ihrer Darstellung nicht alle Punkte auf dem Streudiagramm berücksichtigt, wird sie oft als **Ausgleichsgerade** bezeichnet. Darum können nur Annäherungswerte von y für jedes gegebene x berechnet werden.

Das Konzept von zwei abhängigen Variablen führt zu der Idee der positiven (beide steigen oder fallen zusammen) und **negativen Korrelation** (eine steigt, während die andere fällt).

Ein Korrelationskoeffizient ist ein Ansatz, diese Abhängigkeit numerisch darzustellen.

Der EL-W531 kann den **Korrelationskoeffizienten** berechnen:

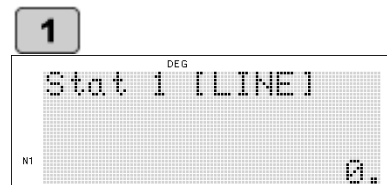
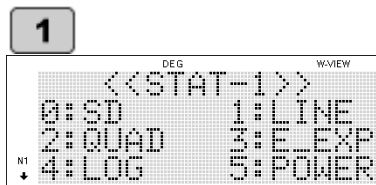
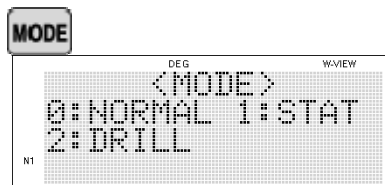
r für die Menge von zweidimensionalen Daten.

Der Wert von **r** wird über die Drittbelegung der Divisionstaste  aktiviert:





Benutzung des Taschenrechners

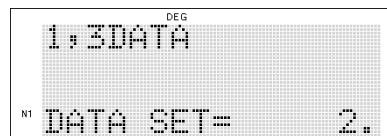
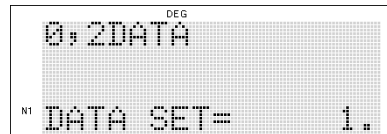
Stellen Sie bei dem Taschenrechner den Statistikmodus ein.



Der Rechner ist jetzt für zweidimensionale Dateneingabe eingestellt.

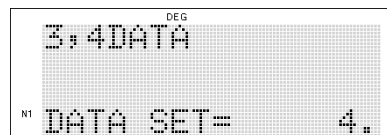
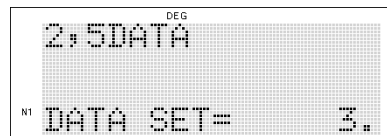
Jedes Paar p, q wird folgendermaßen eingegeben:

p mit der -Taste
DATA
 q mit der -Taste



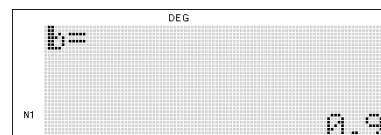
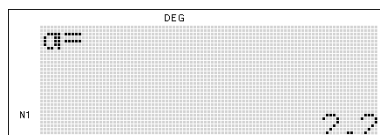
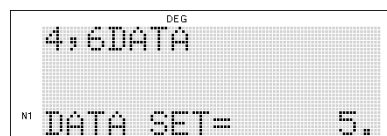
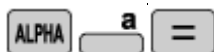
Bitte die zweidimensionalen Datensätze der folgenden Tabelle in den Taschenrechner eingeben. Die einzelnen Displayanzeigen des Taschenrechners sind auf der rechten Seite abgebildet.

p	0	1	2	3	4
q	2	3	5	4	6



Nach der Eingabe der Daten, können die Werte für **a** und **b** in der Gleichung einer Regressionsgraden für $y = a + b x$ mit dem Taschenrechner gefunden werden.

Mit Hilfe der Alpha-Taste können die Werte für **a** und **b** angezeigt werden:



Die Formel für die lineare Regression ist folglich $y = 2,2 + 0,9 x$.

Schätzungen des y -Wertes können durch Einsetzen gegebener x -Werte in die Gleichung errechnet werden.