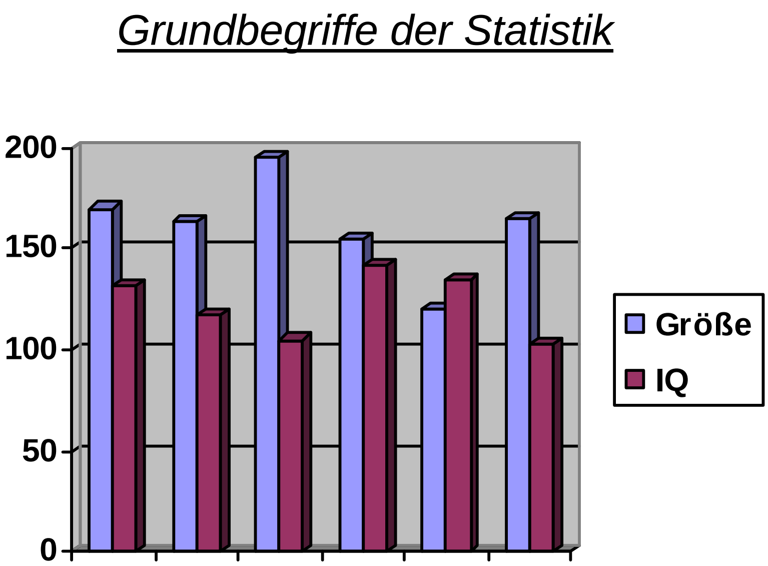
Grundbegriffe der Statistik – Skriptum

Quelle: SACHS, L.: Statistische Methoden: Planung und Auswertung. Berlin Heidelberg 1993.



1. Einleitung

Zahlen bestimmen unseren Alltag. Einerseits ermöglichen sie manchmal erst die **Kommunikation** über abstrakte, der direkten Anschauung weitgehend entzogene Wesenheiten („kommunikativer Aspekt") und andererseits liefern sie oft die Grundlage für individuelle oder kollektive **Entscheidungen** („pragmatischer Aspekt"). Für diese beiden Funktionen von **Quantifizierung** (= Erfassung durch Zahlen) lassen sich leicht Beispiele finden – man denke etwa an Wirtschaftswachstum, Intelligenz, Schulleistung, Blutdruck, Preisstabilität (Inflation).

Aufgabe 1:

- Finde weitere Beispiele für die Verwendung von Zahlen im Alltagsleben.
- Diskutiere die kommunikative und pragmatische Bedeutung von Zahlen am Beispiel von Schulnoten:
  (a) Wie erleichtern bzw. ermöglichen Schulnoten die Kommunikation über Schulleistung?
  (b) Welche (individuelle oder kollektive) Entscheidungen gründen auf Schulnoten?

Allgemein versteht man unter **Messung** die Zuordnung von **Skalenwerten** (nicht unbedingt immer Zahlen) zu Ausprägungsgraden einer **Messgröße**.[1]

Beispiele:

(1) Gewichtsklassen bzw. Güteklassen von Eiern

(2) Verschiedene Temperaturskalen (Celsius, Fahrenheit, ...)

(3) Schulnotenskala von „Sehr gut“ bis „Nicht genügend“

(4) Intelligenzquotient wird als ganze Zahl ausgedrückt

(5) Messung verschiedener physikalischer Größen (Zeit, Masse, Kraft, ...)

Messung setzt ein **Messinstrument** (Uhr, Waage, Test, Maßstab, ...) und genaue **Messvorschriften** (= Regeln zur Anwendung des Messinstruments) voraus. Darüber hinaus sind mit einer Messung bestimmte **Qualitätsansprüche** („Gütekriterien der Messung“) verbunden:

- <u>Objektivität</u> meint die **Unabhängigkeit** des Messergebnisses **von der Person des Messenden** – unter sonst gleichen Bedingungen sollen verschiedene Anwender des Messinstruments zu übereinstimmenden Ergebnissen kommen.
- <u>Reliabilität</u> ist die **Zuverlässigkeit** eines Messinstruments – bei wiederholter Anwendung unter gleichen Bedingungen sollen sich gleiche Ergebnisse einstellen (unabhängig davon, ob tatsächlich das gemessen wird, was erfasst werden soll). Gemeint ist also die „Stabilität“ der Messergebnisse bei gleichen Bedingungen. Die Reliabilität eines Messverfahrens beruht auf dem Vergleich der Messergebnisse untereinander.
- <u>Validität</u> bezeichnet die **Gültigkeit** einer Messung – hohe Validität eines Messergebnisses bedeutet, dass tatsächlich das erfasst wurde, was gemessen werden sollte. Die Validität eines Messverfahrens beruht auf dem Vergleich der Messergebnisse mit einem Außenkriterium (anderer Test, Beobachtung, Modellvorstellung, ...).

<u>Aufgabe 2</u>:

(a) Diskutiere die Objektivität, Reliabilität und Validität von Schulnoten.

(b) Welcher Zusammenhang besteht zwischen Reliabilität und Validität? Diskutiere diese Frage an folgenden Beispielen:

- Körpergröße als Maß der Intelligenz
- Intelligenztest mit hoher Validität, aber geringer Reliabilität.

2. <u>Aufgaben der Statistik</u>

- <u>Beschreibende (deskriptive) Statistik</u>: Erfassung von **Lage** (Mittelwerte), **Verteilung** (Varianz, Standardabweichung) und **Zusammenhang** von Datenmengen.
- <u>Beurteilende bzw. schließende Statistik (Inferenzstatistik)</u>: **Schluss** von einer **Stichprobe** auf eine **Grundgesamtheit** – Anwendung statistischer **Tests** zur **Signifikanzprüfung**.

3. <u>Skalenniveaus</u>

Daten können auf verschiedenen **Skalenniveaus** erfasst werden („**Messung**“):

- <u>**Nominalskala**</u>: Gleichheit / Ungleichheit von Objekten – Klassifizierung; ergibt mit Namen versehene ungeordnete Werte $\Rightarrow$ A = B; A $\neq$ B

  <u>Beispiele</u>: Geschlecht; Beruf; Postleitzahl; Farbe.
- <u>**Ordinalskala**</u>: Geordnete Werte / Größer-Kleiner-Relation – Rangreihe; ergibt Werte, die ihrer Größe entsprechend geordnet sind $\Rightarrow$ A = B; A $\neq$ B; A $<$ B

Beispiele: Schulnoten; Güteklassen (etwa von Eiern); Ränge (Sport, Militär).

- **Intervallskala**: Skala mit konstanten Abständen und willkürlichem Nullpunkt – Gleichheit von Differenzen bzw. Intervallen erfassbar ⇒ A = B; A ≠ B; A < B; A ± B
  Beispiele: Temperatur in Grad Celsius; Kalenderdatum; Punktewert beim Intelligenztest.
- **Verhältnisskala**: Intervallskala mit echtem („natürlichem") Nullpunkt – exakt vergleichbare Verhältnisse ⇒ A = B; A ≠ B; A < B; A ± B; A . B; A : B
  Beispiele: Länge; Gewicht; Alter; Kosten; Phon.

4. Statistische Kennzahlen

(1) Maßzahlen der **Lage** („Mittelwerte"):

- **Median** m:
  Wert in der Mitte der ihrer Größe nach geordneten Daten – er teilt die Datenmenge in zwei gleich große Teilmengen.
  Beispiel:
  (a) 5, 11, 3, 21, 17, 45, 12 ⇒ 3, 5, 11, **12**, 17, 21, 45 ⇒ Median m = 12
  (b) 2, 6, 9, 15, 4, 47 ⇒ 2, 4, **6**, **9**, 15, 47 ⇒ Median m = 7,5 (arithmetisches Mittel (s.u.) der beiden „mittleren" Werte).
  Bemerkung:
  (a) Der Median ist sinnvoll für zumindest **ordinalskalierte Daten**.
  (b) Der Median ist **unempfindlich gegenüber „Ausreißern"** – die beiden Datenmengen {2, 4, 6, 9, 15, 47} und {-1234, 4, 6, 9, 15, 234567} haben denselben Median (nämlich 7,5).

- **Arithmetisches Mittel** $\overline{x}$:
  $$\overline{x} = \frac{x_1 + x_2 + ... + x_n}{n}$$ (für Messwerte $x_1, x_2, ... x_n$)
  Beispiel:
  Arithmetisches Mittel von 4, 7, 2, 9, 1, 3 = ?
  $$\overline{x} = \frac{4+7+2+9+1+3}{6} = \frac{26}{6} \approx 4{,}33$$

(2) Maßzahlen der **Verteilung** („Streuungsmaße"):

Streuungsmaße sagen etwas über die *Verteilung* der Daten aus (eng konzentriert oder über einen weiten Bereich verteilt).

- **Varianz** s^2:
  $$s^2 = \frac{1}{n} \cdot \left[(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + ... + (x_n - \overline{x})^2\right]$$ ($\overline{x}$ = arithmetisches Mittel)

  Beispiel:

  (a) Varianz von 1, 1, 1, 5, 5, 5 ⇒ $\overline{x}$ = 3 ⇒

$$\Rightarrow s^2 = \frac{1}{6} \cdot \left[(1-3)^2 + (1-3)^2 + (1-3)^2 + (5-3)^2 + (5-3)^2 + (5-3)^2\right] = 0$$

(b) Varianz von 3, 3, 3, 3, 3, 3 $\Rightarrow \bar{x} = 3 \Rightarrow$

$$\Rightarrow s^2 = \frac{1}{6} \cdot \left[(3-3)^2 + (3-3)^2 + (3-3)^2 + (3-3)^2 + (3-3)^2 + (3-3)^2\right] = 0$$

- <u>**Standardabweichung**</u> s:

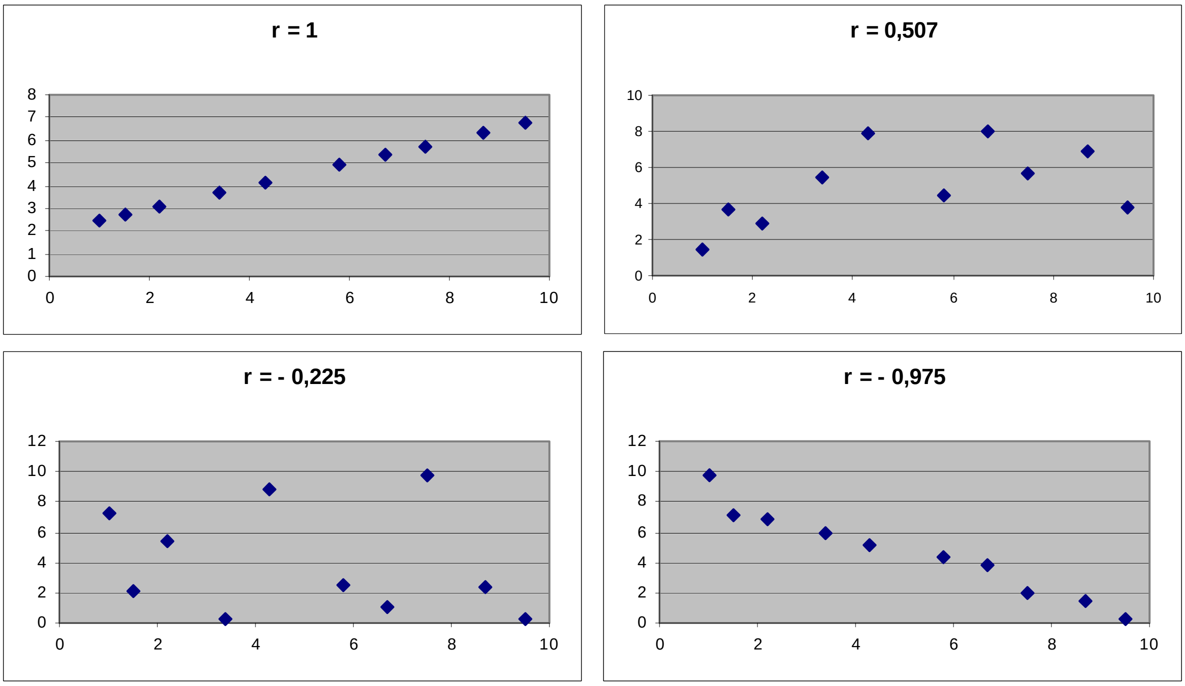
  Die *Standardabweichung* ist die Quadratwurzel aus der Varianz.

(3) <u>Maßzahlen des **Zusammenhangs** („Korrelationskoeffizienten“)</u>:

Der Korrelationskoeffizient r ist ein Maß für den *linearen Zusammenhang* von paarweise gegebenen Daten.

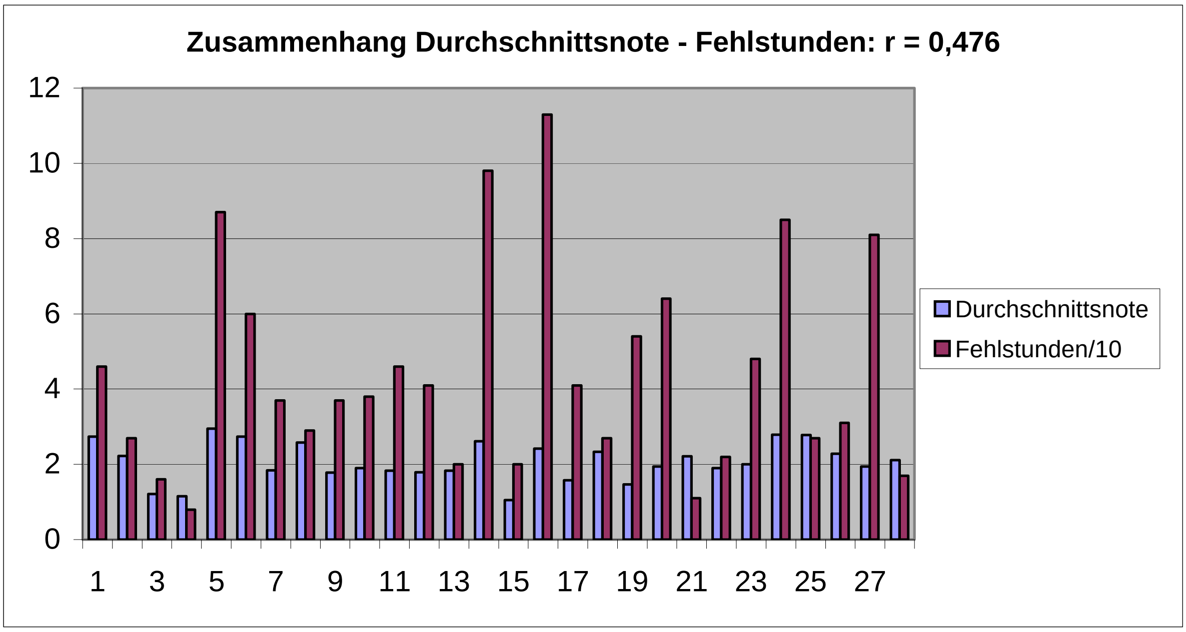
- r = 1: perfekter positiver linearer Zusammenhang
- r = 0: kein linearer Zusammenhang
- r = -1: perfekter negativer linearer Zusammenhang

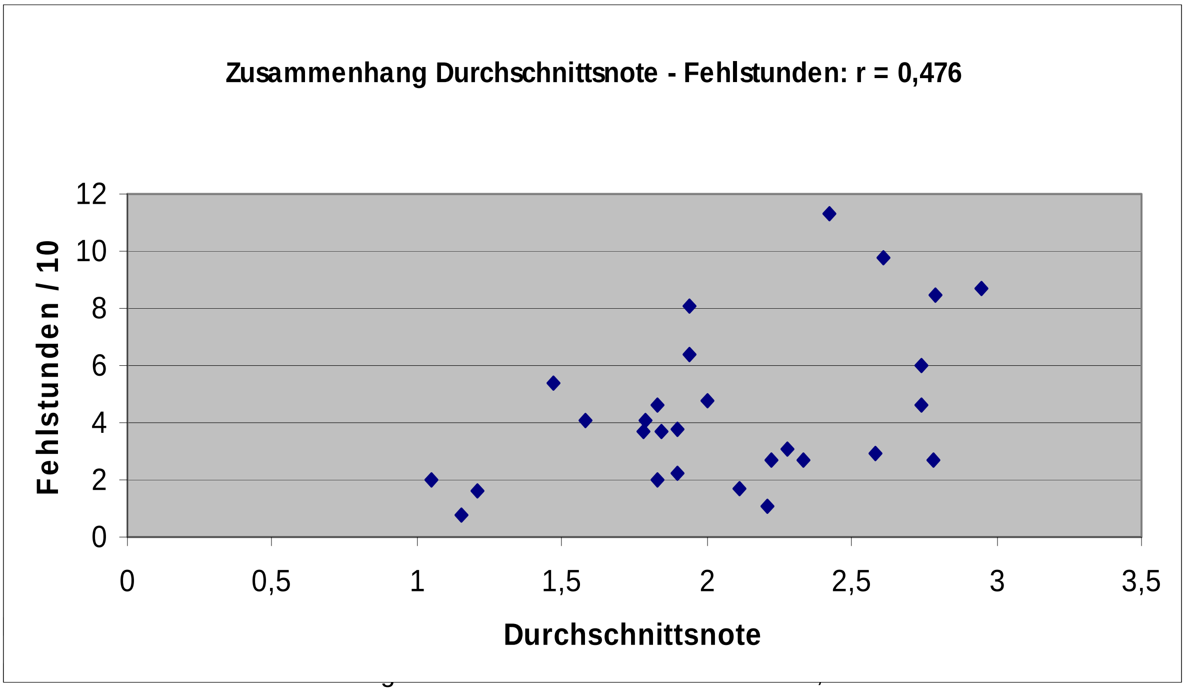
<u>Beispiel für Korrelationskoeffizienten</u>:



<u>Beispiel</u>: <u>Zusammenhang Durchschnittsnote – Fehlstundenanzahl (3c – Schuljahr 2003/04)</u>

Gibt es einen linearen Zusammenhang zwischen der Durchschnittsnote und der Anzahl der Fehlstunden?





Schülerinnen mit einer höheren Fehlstundenzahl auch höhere (also schlechtere) Jahresdurchschnittsnoten. Mit r = 0,476 ist der Zusammenhang mäßig stark ausgeprägt.

<u>Bemerkung</u>: Korrelationen dürfen nicht automatisch kausal interpretiert werden, d.h. auch wenn zwischen Variablen X und Y eine hohe (positive oder negative) Korrelation gefunden wird, so ist nicht unbedingt eine der beiden Variable eine „Ursache" für die andere. Es gibt zunächst immer mehrere Möglichkeiten:

- X beeinflusst Y: $X \Rightarrow Y$
- Y beeinflusst X: $Y \Rightarrow X$
- Es gibt eine dritte Variable Z, die X und Y beeinflusst: $X \Leftarrow Z \Rightarrow Y$

Im Beispiel des Zusammenhangs von Durchschnittsnote und Fehlstundenanzahl bedeutet das:

- Häufiges Fehlen (nicht unbedingt selbstverschuldet) kann die Noten negativ beeinflussen.
- Schlechte Schulnoten können die Motivation zum Schulbesuch reduzieren – man geht dann einfach nicht mehr so gern in die Schule.
- Schulunlust und fehlende Motivation bewirken einerseits häufiges Fernbleiben vom Unterricht und andererseits sind sie auch Ursache für schlechte Leistungen in der Schule.

Voreilige bzw. ungerechtfertigte kausale Interpretation von Korrelationen führt nicht selten zu Fehlschlüssen und unsinnigen Behauptungen. So wurde zum Beispiel tatsächlich eine relativ hohe positive Korrelation zwischen der Anzahl der Störche und der Anzahl der Geburten im Burgenland gefunden – ist damit endlich bewiesen, dass Störche die Babies bringen?

Ein einfaches Beispiel soll zeigen, dass jedes der drei Maße (Lage, Verteilung, Zusammenhang) seine spezifische Bedeutung hat und Informationen liefert, die in den anderen Maßen nicht enthalten sind. Betrachten wir etwa die *Einkommensverhältnisse* in einem Land (oder auch in einer Region oder einer Firma), so können wir uns zunächst für das **Durchschnittseinkommen** interessieren. Dieses sagt durchaus etwas über den Wohlstand einer Region aus und wird auch oft angegeben, um die wirtschaftliche Prosperität grob zu charakterisieren. Was dabei aber „verschwindet" ist Information über die **Verteilung** der Einkommen – wir wissen nicht, ob ein bestimmtes Durchschnittseinkommen dadurch zustande kommt, dass die Einkommen weitgehend gleich verteilt sind oder ob es vielleicht einige wenige Personen mit extrem hohen Einkünften gibt, denen viele Menschen mit sehr geringen Einnahmen gegenüber stehen. Die Frage der Verteilung der Einkommen kann für die Entwicklung und Stabilität einer Gesellschaft bedeutsam sein (man denke an Verteilungskämpfe und den sozialen Frieden). Schließlich kann es interessant sein, die Einkommensverhältnisse in verschiedenen Gesellschaften zu anderen wichtigen Parametern (Bildungsstand, Lebenserwartung, Kindersterblichkeit, Produktivität, Bruttosozialprodukt, ...) in Beziehung zu setzen – also einen **Zusammenhang** zwischen diversen, für eine Gesellschaft bedeutsamen Faktoren und den Einkommensverhältnissen herzustellen. Das ist besonders wichtig im Zuge wirtschaftswissenschaftlicher Modellbildung und bei der empirischen Prüfung entsprechender Hypothesen.

Übungen und Ergänzungen

(1) Bestimme jeweils das Skalenniveau, auf dem „gemessen" wird:

⇒ <u>Autonummern</u>

⇒ <u>Körpergröße</u>

⇒ <u>Schuhgröße</u>

⇒ <u>Luftdruck</u>

⇒ Gesamtkalkül bei der Matura mit den Beurteilungsstufen „*Mit Auszeichnung bestanden*“, „*Mit gutem Erfolg bestanden*“, „*Bestanden*“ und „*Nicht bestanden*“

⇒ Postleitzahlen

⇒ Herzfrequenz

⇒ Tabellenstand der Fußball-Bundesliga

(2) Wie könnte man *Kreativität* messen? Beschreibe eine mögliche Vorgangsweise. Welche Probleme ergeben sich dabei hinsichtlich Objektivität, Reliabilität und Validität?

(3) Ein Psychologe möchte untersuchen, ob ein Zusammenhang zwischen Schulleistung und Intelligenz besteht. Wie könnte er dabei vorgehen? In welcher Hinsicht ist seine Methode problematisch?

(4) Nach einer Eignungsprüfung werden die 90 erfolgreichen KandidatInnen in drei Klassen mit jeweils 30 SchülerInnen eingeteilt. Wie müsste man vorgehen um Klassen zusammenzustellen, die im Hinblick auf ihre Leistungsfähigkeit

(a) möglichst homogen **(b)** möglichst ähnlich sind? (⇒ Wie kann man die Homogenität bzw. die Ähnlichkeit der Leistungsfähigkeit einer Klasse messen?)

Welcher der beiden Alternativen (Homogenität versus Ähnlichkeit) ist der Vorzug zu geben (etwa im Hinblick auf pädagogische Überlegungen)? Begründe die Antwort!

(5) Zwischen der *Körpergröße* und dem *Körpergewicht* wäre eine relativ hohe positive Korrelation zu erwarten. Finde selbst weitere Beispiele, bei denen man **(a)** eine hohe positive Korrelation **(b)** eine stark negative Korrelation erwarten würde.

(6) Ein Arzt findet eine sehr hohe positive Korrelation zwischen dem Körpergewicht und dem Blutdruck. Ist Übergewicht die Ursache für Hypertonie (= Bluthochdruck)?

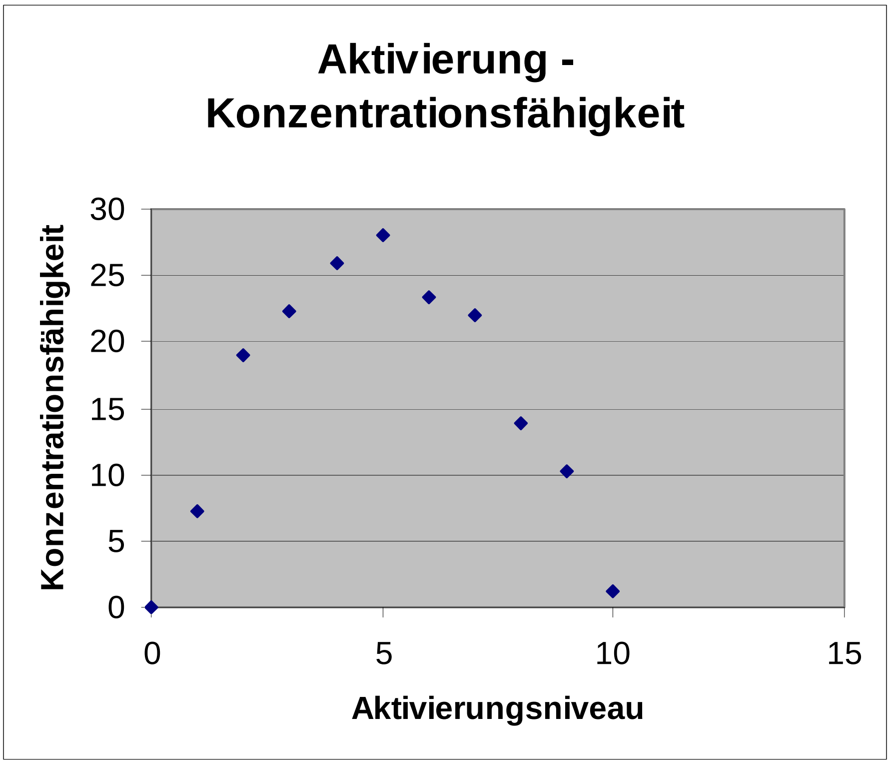
(7) Korrelationskoeffizienten nahe Null bedeuten nicht notwendig das Fehlen jeglichen Zusammenhangs. Betrachten wir dazu das folgende Beispiel:

Ein Psychologe untersuchte die Frage, ob das „Aktivierungsniveau“ (= Grad der „Aufgeregtheit“) einer Person mit deren Konzentrationsfähigkeit zusammenhängt. Dazu wurden (in Vortests) zunächst Personen ausgewählt, die sich hinsichtlich ihrer Konzentrationsfähigkeit in entspanntem Zustand wenig unterschieden. Diese Personen unterzogen sich dann auf verschiedenen Aktivierungsniveaus (von völliger Entspannung bis höchster Aufregung) einem Konzentrationstest. Der Psychologe fand folgende Werte (wobei uns hier Einzelheiten der Messmethoden und der Skalen nicht beschäftigen sollen):

Aktivierung	0	1	2	3	4	5	6	7	8	9	10
Konzentration	0	7,2	19	22,3	26	28	23,4	22	13,8	10,2	1,2

Für diese Daten erhält man einen Korrelationskoeffizienten (linearer Zusammenhang) von r ≈ -0,002 – es scheint also kein Zusammenhang zwischen Aktivierung und Konzentration zu bestehen.

Wenn wir aber die Daten graphisch darstellen, so wird doch ein Effekt deutlich:



Es scheint ein „umgekehrt U-förmiger“ Zusammenhang (statt eines linearen) vorzuliegen. Die Konzentrationsfähigkeit ist bei mittlerer Aktivierung am höchsten und nimmt bei starker Entspannung und großer Aufregung rasch ab. Die Verteilung der Punkte folgt in etwa einer Parabel – es scheint also ein „quadratischer“ Zusammenhang zwischen den beiden Variablen zu bestehen.

Lernziele – Prüfungsfragen

- Welche wichtigen Funktionen erfüllt **Quantifizierung** im Alltag? Erläutere die verwendeten Begriffe und gib Beispiele für jede dieser Funktionen an!
- Was versteht man unter **Messung**? Erläutere den Begriff „Messung“ an Beispielen!
- Nenne drei **Gütekriterien** für Messung!
- Erkläre den Begriff „**Objektivität**“!
- Erkläre den Begriff „**Reliabilität**“!
- Erkläre den Begriff „**Validität**“!
- Was versteht man unter (a) **deskriptiver Statistik** und (b) **Inferenzstatistik**?
- Nenne die vier **Skalenniveaus** von Messung und gib Beispiele dafür an!

- Nenne die Grundtypen **statistischer Kennzahlen** und gib dabei für jeden Typus mindestens ein solches Maß an!
- Lässt sich von einer vorhandenen (positiven oder negativen) Korrelation auf einen **Kausalzusammenhang** schließen? Begründe die Antwort (etwa durch ein Beispiel)!
- Zeige an einem einfachen Beispiel, dass jede der drei **statistischen Kennzahlen** ihre spezielle Bedeutung hat!
- Skizziere an einem Beispiel (Punktwolke!) (a) eine **hohe positive Korrelation** (b) eine **stark negative Korrelation** (c) eine **schwach positive Korrelation** und (d) eine **Korrelation nahe Null**!
- Im Zuge einer empirischen Untersuchung wurde für die Variablen X und Y ein Korrelationskoeffizient von r = 0,013 gefunden. Bedeutet dies, dass keinerlei **Zusammenhang** zwischen X und Y besteht? Begründe die Antwort!
- Gib ein Beispiel für eine **höchst objektive, aber nicht valide „Messung“**!

Anmerkungen:

(1) Diese Definition weicht insofern von der üblichen Begriffsverwendung ab, als bei einer Messung Zahlen als Skalenwerte vorausgesetzt werden (vgl. dazu SACHS (1993), S. 11).