JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN

# Vorlesung: Statistik I

Prof. Dr. Simone Abendschön

9. Einheit

# Plan für heute



- **Zusammenhänge zwischen (pseudo-)metrischen Merkmalen**
  - Allgemein
  - Hypothesen
  - Grafische Veranschaulichung
  - Zusammenhangsmaß Pearson‘s r
- **Zusammenhangsmaß zwischen metrischem und gruppiertem Merkmal: PRE-Maß $\eta$2 (Eta-Quadrat)**

# Zusammenhangsmaße



- Maße zur Berechnung, ob und wie stark zwei Merkmale miteinander **statistisch zusammenhängen** oder „**korrelieren**“

- Bei ordinalskalierten und metrisch skalierten Maßen auch Aussagen über die **Richtung des Zusammenhangs** möglich (positiv oder negativ)

- **Pearson‘s r** als Zusammenhangsmaß für zwei metrisch skalierte Merkmale (**Korrelationskoeffizient**)

# Allgemein



**Folgende Fragen stellen sich bei der statistischen Analyse von Merkmalszusammenhängen:**

- Wie lässt sich die **Form** des Zusammenhangs zwischen X und Y beschreiben?
- Welche **Richtung** hat der Zusammenhang zwischen X und Y, d.h. ist er **negativ** oder **positiv**?
- Welche **Stärke** hat der Zusammenhang zwischen X und Y?
- (Lässt sich der in der Stichprobe ermittelte Zusammenhang auf die Population übertragen? → Inferenzstatik → nächstes Semester)

# Zusammenhangsmaße



- Quantifizierung des Zusammenhangs durch **Korrelationskoeffizienten**
- **Positive Korrelation**:
    - Hohe Werte in der einen Variablen gehen mit hohen Werten in der anderen Variablen einher
    - Niedrige Werte in der einen gehen mit niedrigen Werten in der anderen Variablen einher
- **Negative Korrelation**:
    - Hohe (niedrige) Werte in der einen Variablen gehen mit niedrigen (hohen) Werten in der anderen Variablen einher
- Ggfs. auch **kein Zusammenhang (Korrelation um 0)**

# Zusammenhangsmaße



**Wann ist ein Messwert „hoch“? Wann ist ein Messwert „niedrig“?**

- Vergleich anhand des arithmetischen Mittels der jeweiligen Variablen
    - **Hohe Messwerte** entsprechen Werten über dem arithmetischen Mittel
    - **Niedrige Messwerte** entsprechen Werten unter arithmet. Mittel

**Stärke des Zusammenhangs zwischen zwei metrischen Variablen ergibt sich durch die Abweichung der Messwerte vom jeweiligen Mittelwert**

# Hypothesen



**In der quantitativen Sozialforschung entwickeln wir Hypothesen, die Zusammenhänge zwischen (mind.) 2 Merkmalen postulieren**

- ***Zusammenhangshypothesen***

  Beispiel: Je geringer der soziale Status desto schlechter der Gesundheitszustand im Rentenalter → **linearer Zusammenhang**

- ***Unterschiedshypothesen***

  Beispiel: Rentner*innen mit hohem sozialen Status haben einen besseren Gesundheitszustand als Rentner*innen mit niedrigem sozialen Status

# Korrelationen



- Eine **lineare Korrelation** beschreibt einen **linearen** Zusammenhang zwischen zwei Variablen
- D.h. eine Veränderung von X geht mit einer dazu proportionalen Veränderung von Y einher
  - Beispiel: Je größer ein Mensch ist, desto schwerer ist er (Pro cm mehr Größe, bestimmte Grammzahl mehr)
- Korrelation trifft **keine Aussage** zur Kausalität
- Korrelationsanalyse misst nur, ob sich zwei Merkmale **im Gleichklang** bewegen

# Untersuchung von Hypothesen



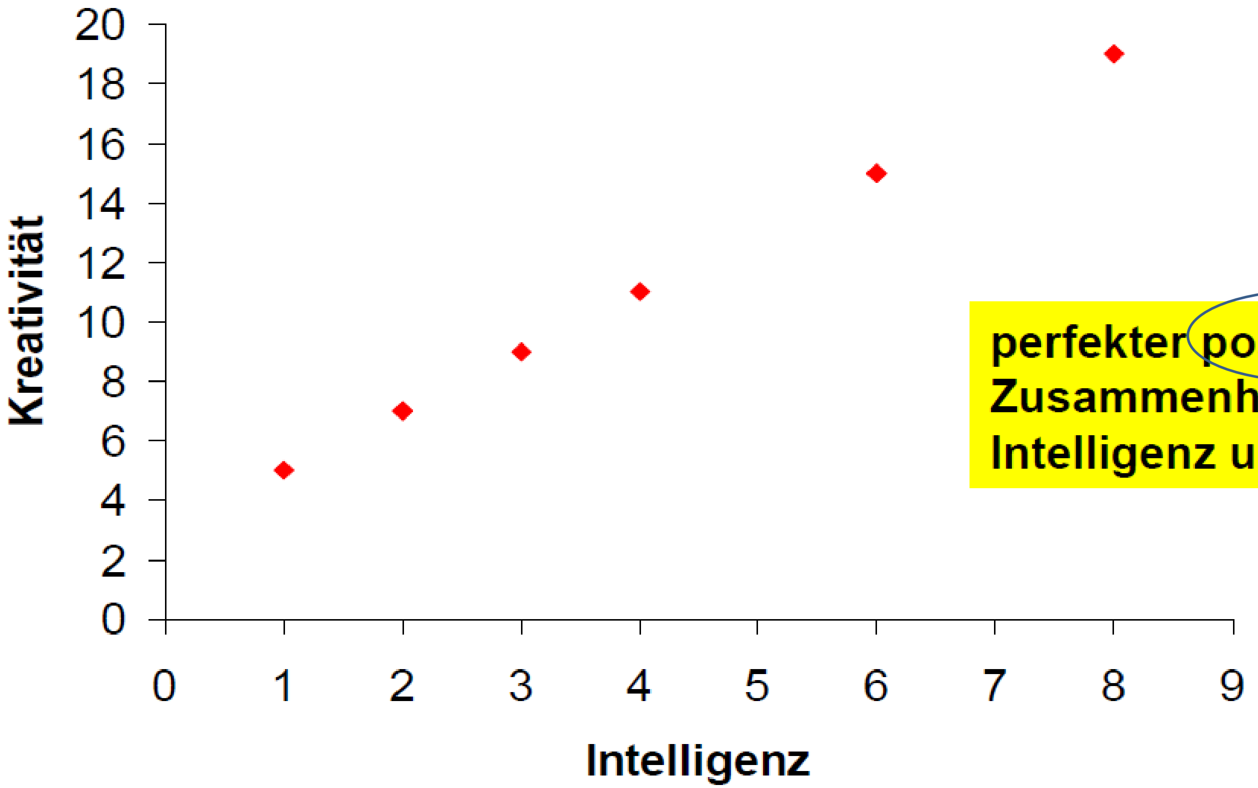
- Bi- (und multi)variate Analysen, um Zusammenhänge zwischen Merkmalen zu untersuchen
- Bei **metrisch** skalierten Daten:
  - **Korrelationsanalyse** (Pearson‘s r)
  - zunächst bietet sich eine **grafische Darstellung** an: **Streudiagramm** (auch Punktwolkendiagramm, Scatterplot)

- **Zusammenhänge zwischen (pseudo-)metrischen Merkmalen**
  - Allgemein
  - Hypothesen
  - Grafische Veranschaulichung
  - Zusammenhangsmaß Pearson‘s r
- **Zusammenhangsmaß zwischen metrischem und gruppiertem Merkmal: PRE-Maß $\eta$2 (Eta-Quadrat)**

# Beispiel: Streudiagramm



**Zusammenhangshypothese: „Je intelligenter eine Person ist, desto kreativer ist sie auch“.**




**perfekter positiver Zusammenhang zwischen Intelligenz und Kreativität**

# Beispiel 2: Streudiagramm



**Zusammenhangshypothese (S. Lehrbrief S. 63/64): „Je intelligenter eine Person ist, desto besser kann sie auch räumlich denken“**

Tabelle 34: IQ und Testergebnis beim räumlichen Denken – Urliste

| ID | IQ | Testergebnis |
|---|---|---|
| 1 | 104 | 98 |
| 2 | 90 | 80 |
| 3 | 103 | 85 |
| 4 | 115 | 99 |
| 5 | 105 | 95 |
| 6 | 118 | 106 |
| 7 | 109 | 100 |
| 8 | 95 | 88 |

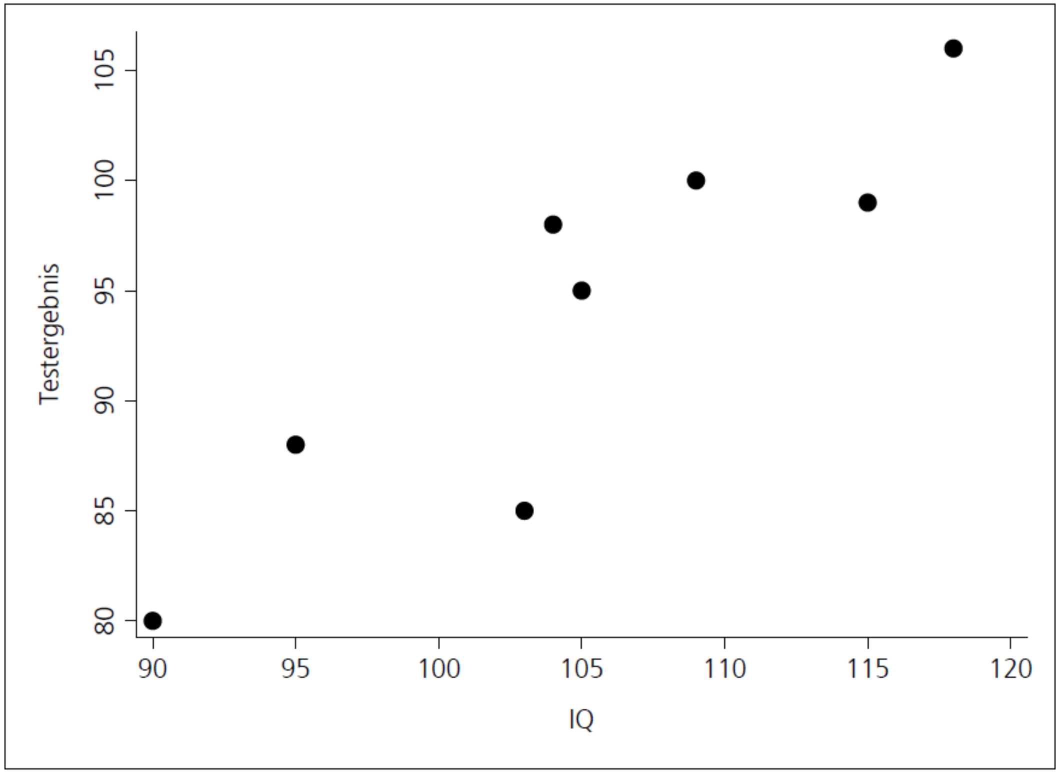
Quelle: Eigene Darstellung

# Beispiel 2: Streudiagramm



**Zusammenhangshypothese (S. Lehrbrief S. 63/64): „Je intelligenter eine Person ist, desto besser kann sie auch räumlich denken“**

Abbildung 13: IQ und Testergebnis beim räumlichen Denken – Streudiagramm

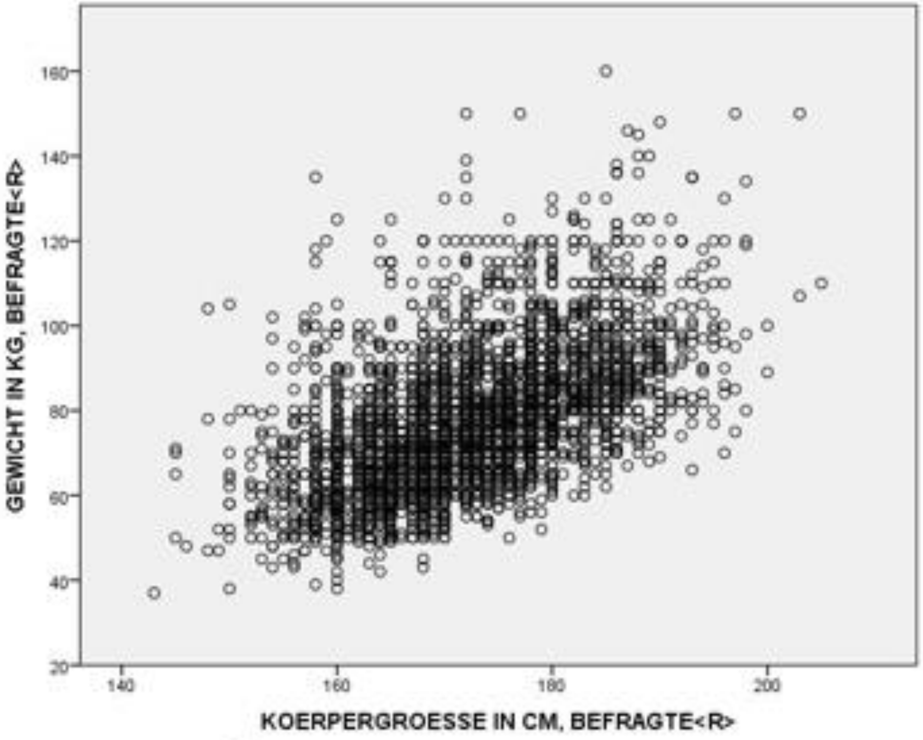



Quelle: Eigene Darstellung

# Beispiel 3: Streudiagramm

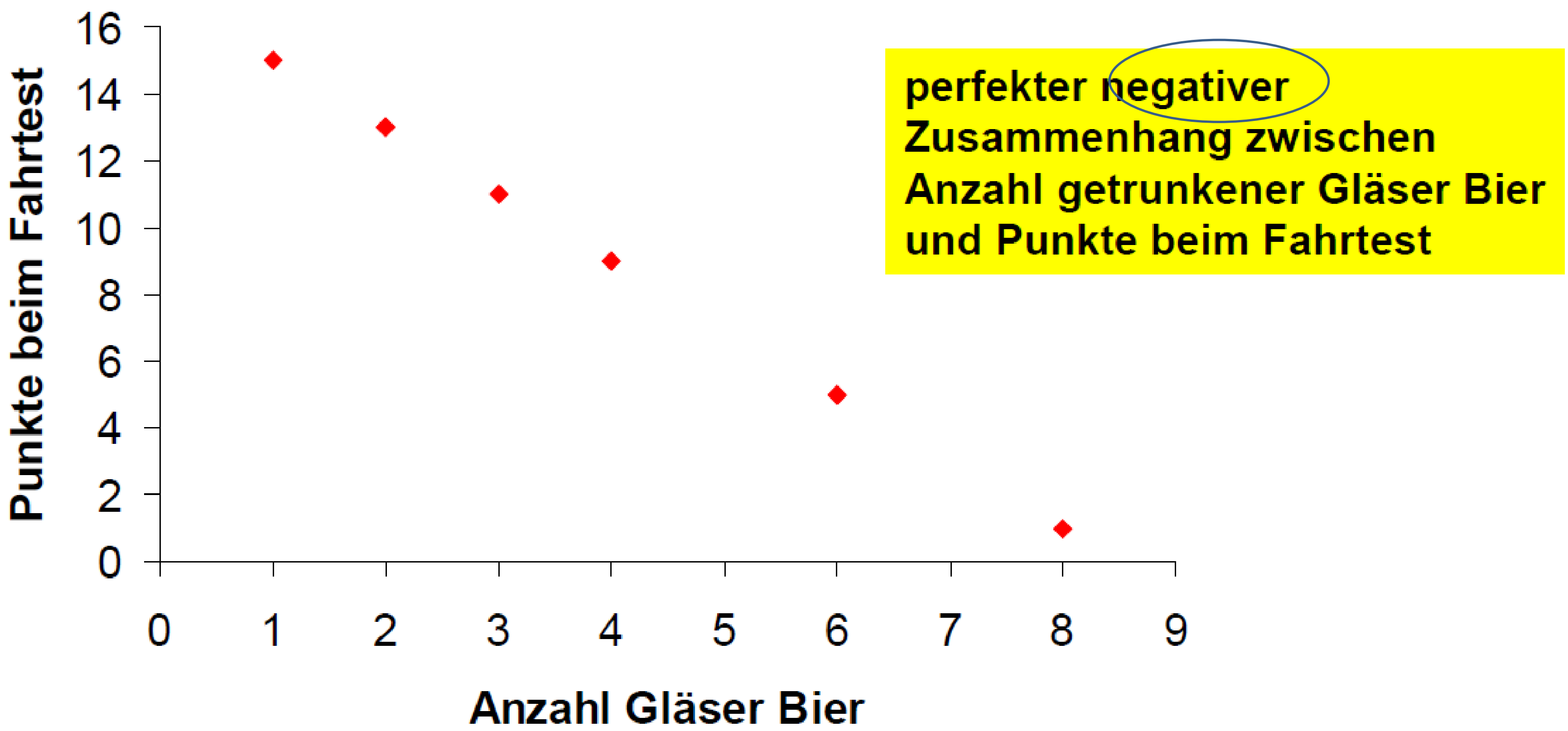
**Zusammenhangshypothese: „Je größer man ist, desto schwerer ist man auch“ (Streudiagramm erstellt auf Basis des Allbus 2016)**

r=0,547



# Beispiel 4: Streudiagramm

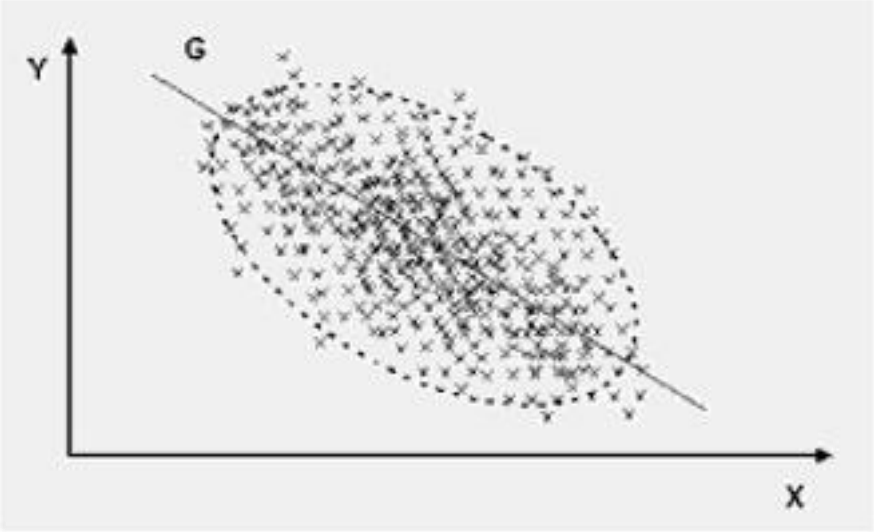
**Zusammenhangshypothese: „Je mehr Alkohol man trinkt,, desto schlechter fährt man Auto“**



# Beispiel 5: Streudiagramm



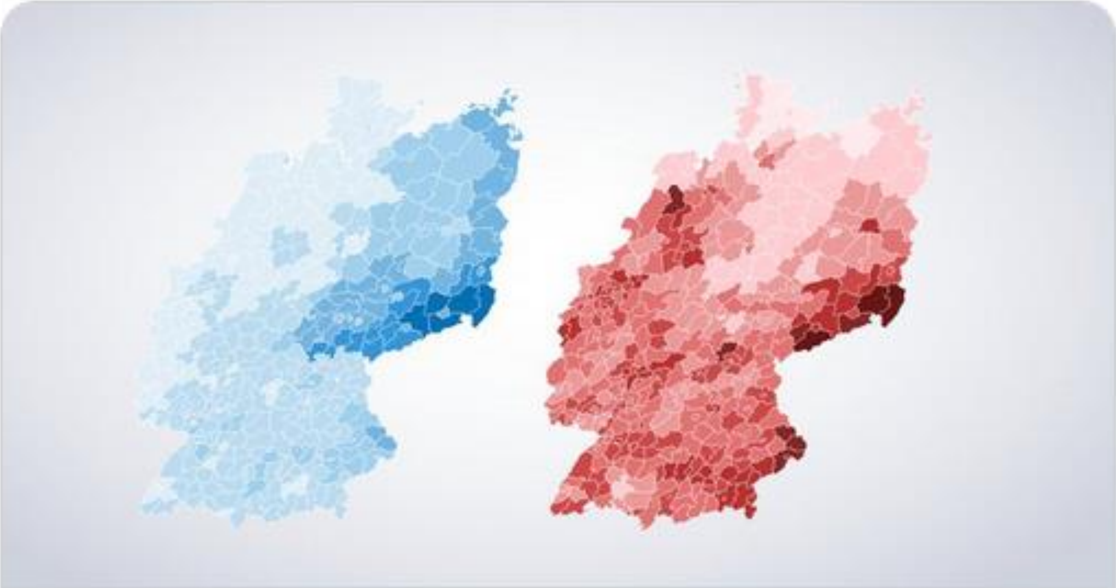
**Zusammenhangshypothese: „Je älter ein Auto ist, desto niedriger sein Verkaufswert“ – negativer Zusammenhang**



# Aktuelles Beispiel

**Matthias Quent** @Matthias_Quent · 12. Dez. 2020

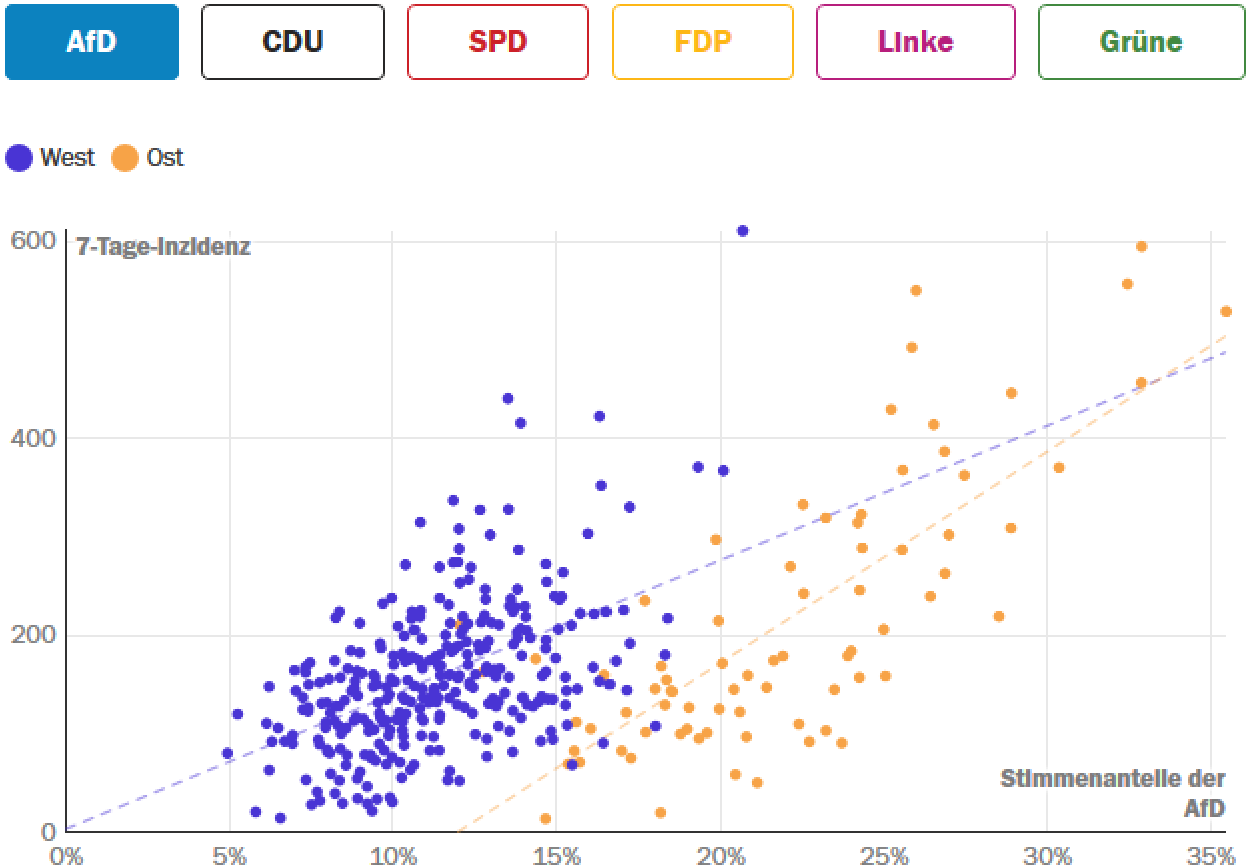
Hängen #AfD-Hochburgen und hohe #Coronazahlen zusammen? @plateauton @Tagesspiegel hat unsere Berechnungen @IDZ_Jena überprüft und ausgeweitet. Die Ergebnisse bestätigen die Korrelation. interaktiv.tagesspiegel.de/lab/hotspots-u... via @tagesspiegel



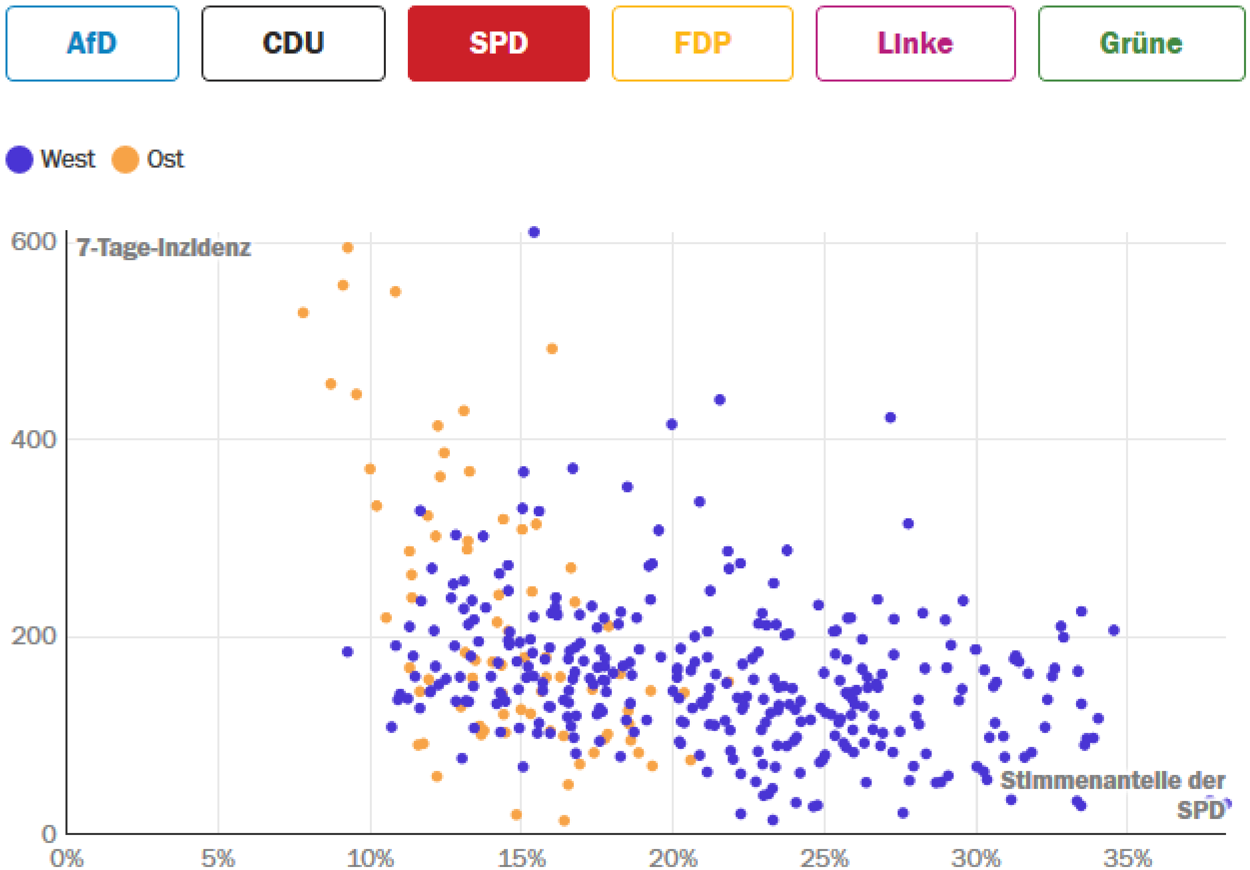
Hängen AfD-Hochburgen und hohe Coronazahlen zusammen?

Eine Analyse legt nahe, dass in Landkreisen mit großer AfD-Wählerschaft auch die Fallzahlen höher sind. Was ist dran? Wir rechnen nach.

interaktiv.tagesspiegel.de




https://interaktiv.tagesspiegel.de/lab/hotspots-und-rechte-haengen-afd-hochburgen-und-corona-hotspots-zusammen/




https://interaktiv.tagesspiegel.de/lab/hotspots-und-rechte-haengen-afd-hochburgen-und-corona-hotspots-zusammen/

# Pearson‘s r



- Auch **Bravais-Pearson Produktkorrelation**
- Berechnet die Stärke des **linearen Zusammenhangs** zwischen zwei (pseudo-)metrisch skalierten Variablen
- Zwischenschritt zur Berechnung: **Kovarianz**

# Kovarianz - Berechnung

Kovarianz beschreibt die **gemeinsame Streuung zweier Merkmale**

1) Abweichung vom Mittelwert für jedes Messwertepaar bestimmen

2) Gemeinsame Abweichung beider Messwerte von ihren Mittelwerten durch Multiplikation berechnen

3) Berechnung der Summe der Abweichungsprodukte

4) Berechnung des durchschnittliche Abweichungsprodukt (mittels Division durch n)

$$\mathrm{cov}\,(x,y) = \frac{\sum_{i=1}^{n}(x_i-\bar{x})\ (y_i-\bar{y})}{n}$$

Wiederholung Varianz $$\sigma^2 = \frac{1}{n} * \sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}$$

# Kovarianz



- hoch **positiv**, wenn hohe positive Abweichungen mit hohen positiven Abweichungen einhergehen bzw. hohe negative Abweichungen mit hohen negativen Abweichungen
- hoch **negativ**, wenn hohe positive Abweichungen mit hohen negativen Abweichungen einhergehen und umgekehrt.
- gleich **null**, wenn die Richtung der Abweichungen vom Mittelwert in X nicht systematisch mit einer bestimmten Richtung der Abweichungen vom Mittelwert in Y einhergeht.

# Kovarianz



- Aber: **unstandardisiertes** Maß, Größe ist abhängig von den jeweiligen Maßeinheiten der beiden Merkmale

- Vergleich zwischen Kovarianzen wird erschwert

→ **Standardisierung mit dem Korrelationskoeffizienten Pearson's r** anhand der Division durch das Produkt der Standardabweichungen beider Merkmale

# Pearson‘s r



Pearson‘s r entspricht der anhand des Produkts der Standardabweichungen standardisierten Kovarianz

$$r = \frac{cov(x;y)}{s_x * s_y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x}) * (y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 * \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

# Pearson‘s r



- **Wertebereich von -1 bis +1** (siehe auch Spearman‘s Rho!),
- Vorzeichen zeigt **Richtung** der Korrelation an, Betrag die **Stärke** des Zusammenhanges
    - Negatives Vorzeichen: negativer Zusammenhang
    - Positives Vorzeichen: positiver Zusammenhang

# Interpretation Pearson‘s r - Faustregel



Tabelle 36: Interpretation von Pearson’s r

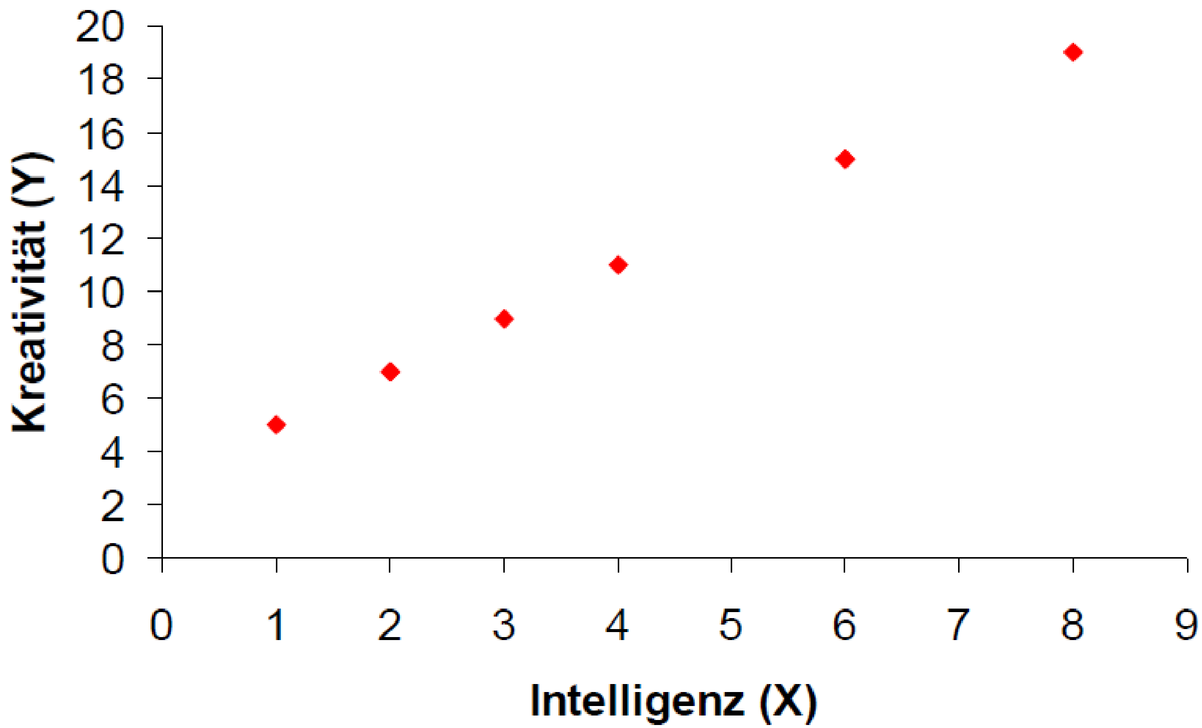
| Korrelationskoeffizient ($\|r\|$) | Interpretation |
|---|---|
| $\leq 0{,}05$ | kein Zusammenhang |
| $> 0{,}05$ bis $\leq 0{,}20$ | schwacher Zusammenhang |
| $> 0{,}20$ bis $\leq 0{,}50$ | mittelstarker Zusammenhang |
| $> 0{,}50$ bis $\leq 0{,}70$ | starker Zusammenhang |
| $> 0{,}70$ | sehr starker Zusammenhang |

Quelle: Eigene Darstellung

Aber: Die Beurteilung der Höhe einer Korrelation hängt immer von der zugrunde liegenden Fragestellung ab!

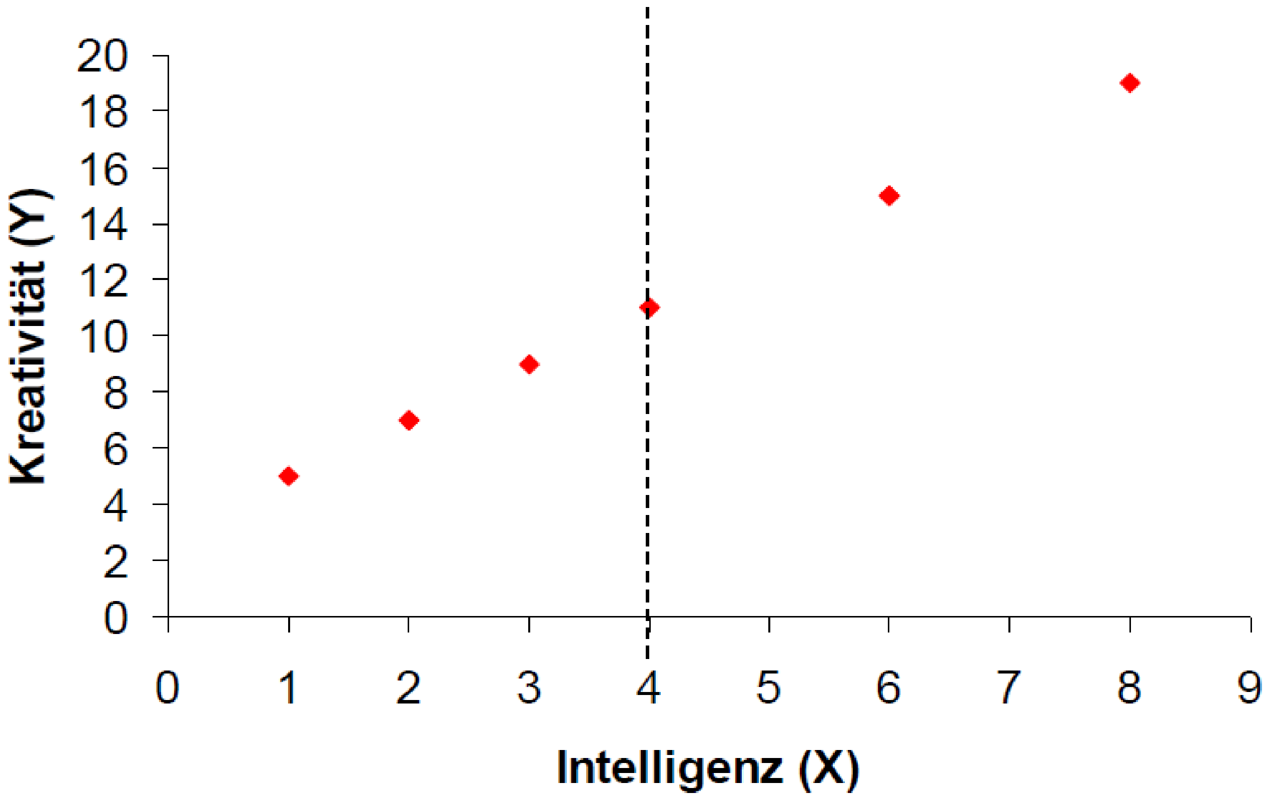
# Pearson‘s r: Schritt für Schritt





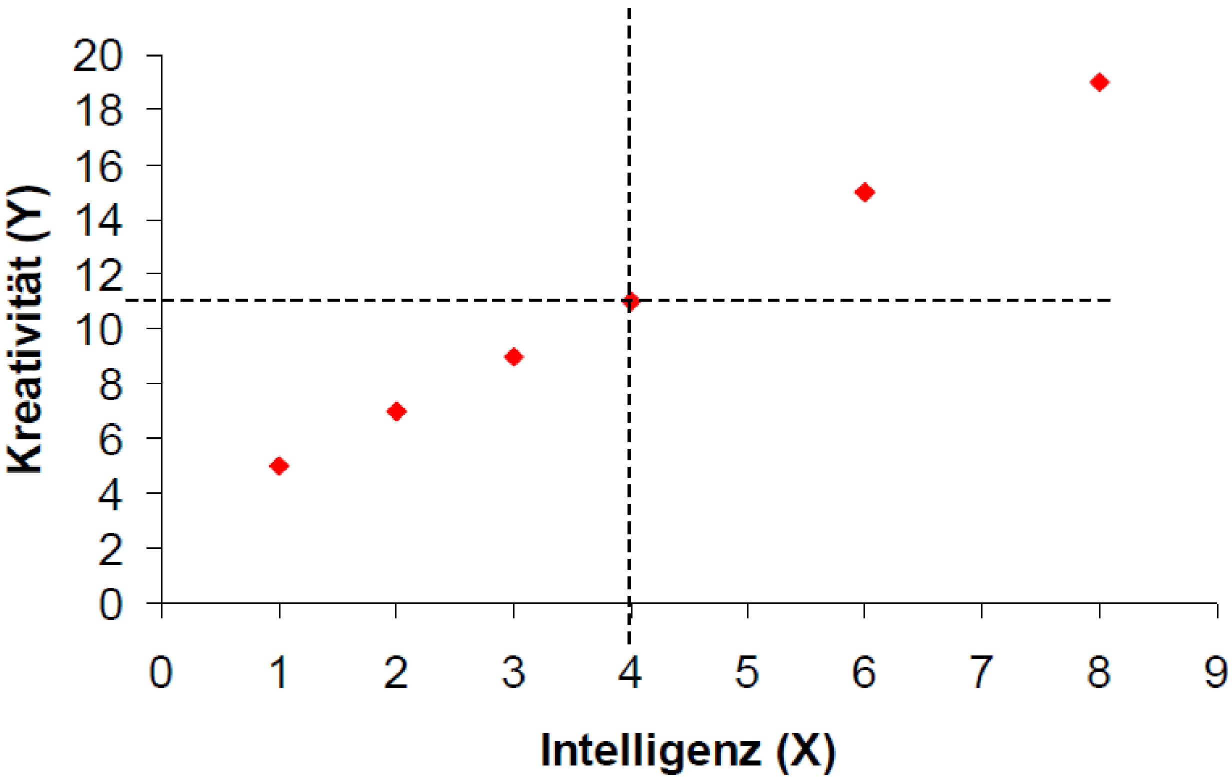

| | *X* | *Y* |
|---|---|---|
| | 1 | 5 |
| | 2 | 7 |
| | 2 | 7 |
| | 3 | 9 |
| | 4 | 11 |
| | 6 | 15 |
| | 6 | 15 |
| | 8 | 19 |
| *M* = | 4 | 11 |
| $s^2$ = | 5,25 | 21 |

# Pearson's r: Schritt für Schritt




| | *X* | *Y* |
|---|---|---|
| | 1 | 5 |
| | 2 | 7 |
| | 2 | 7 |
| | 3 | 9 |
| | 4 | 11 |
| | 6 | 15 |
| | 6 | 15 |
| | 8 | 19 |
| *M* = | 4 | 11 |
| $s^2$ = | 5,25 | 21 |

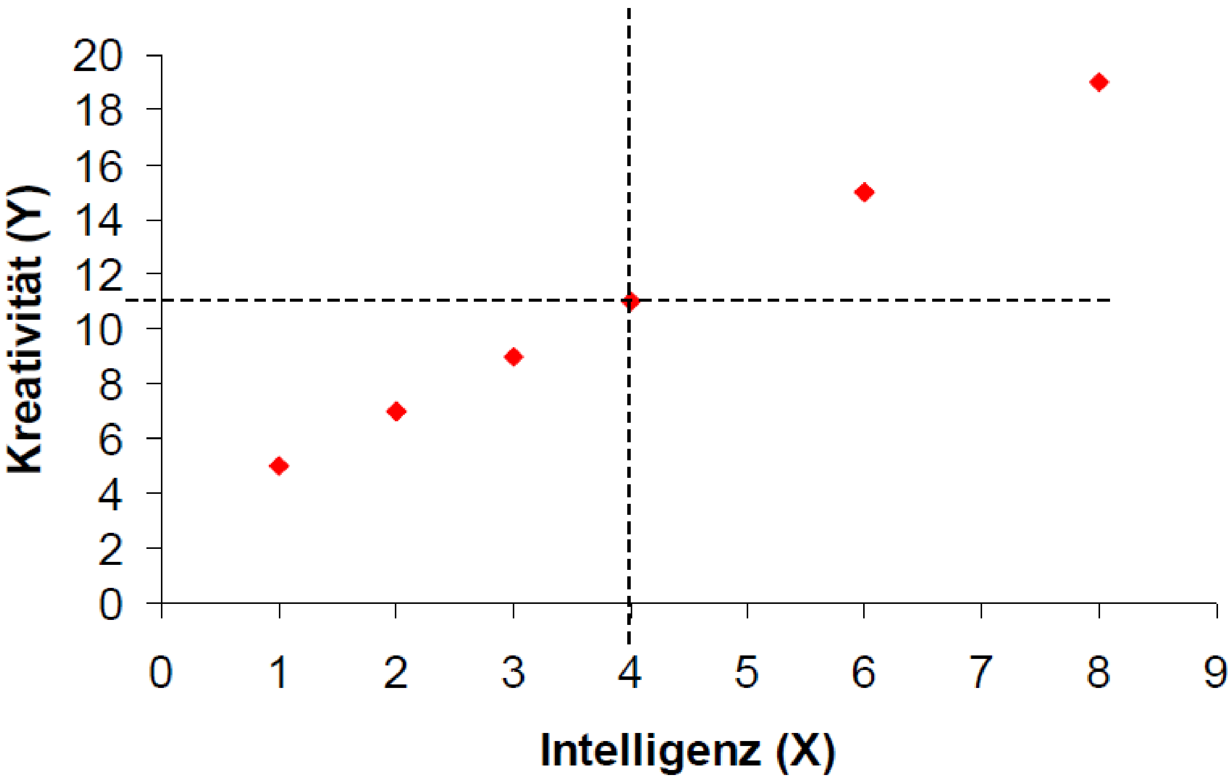
# Pearson‘s r: Schritt für Schritt




| | *X* | *Y* |
|---|---|---|
| | 1 | 5 |
| | 2 | 7 |
| | 2 | 7 |
| | 3 | 9 |
| | 4 | 11 |
| | 6 | 15 |
| | 6 | 15 |
| | 8 | 19 |
| $M$ = | 4 | 11 |
| $s^2$ = | 5,25 | 21 |

# Pearson‘s r: Schritt für Schritt



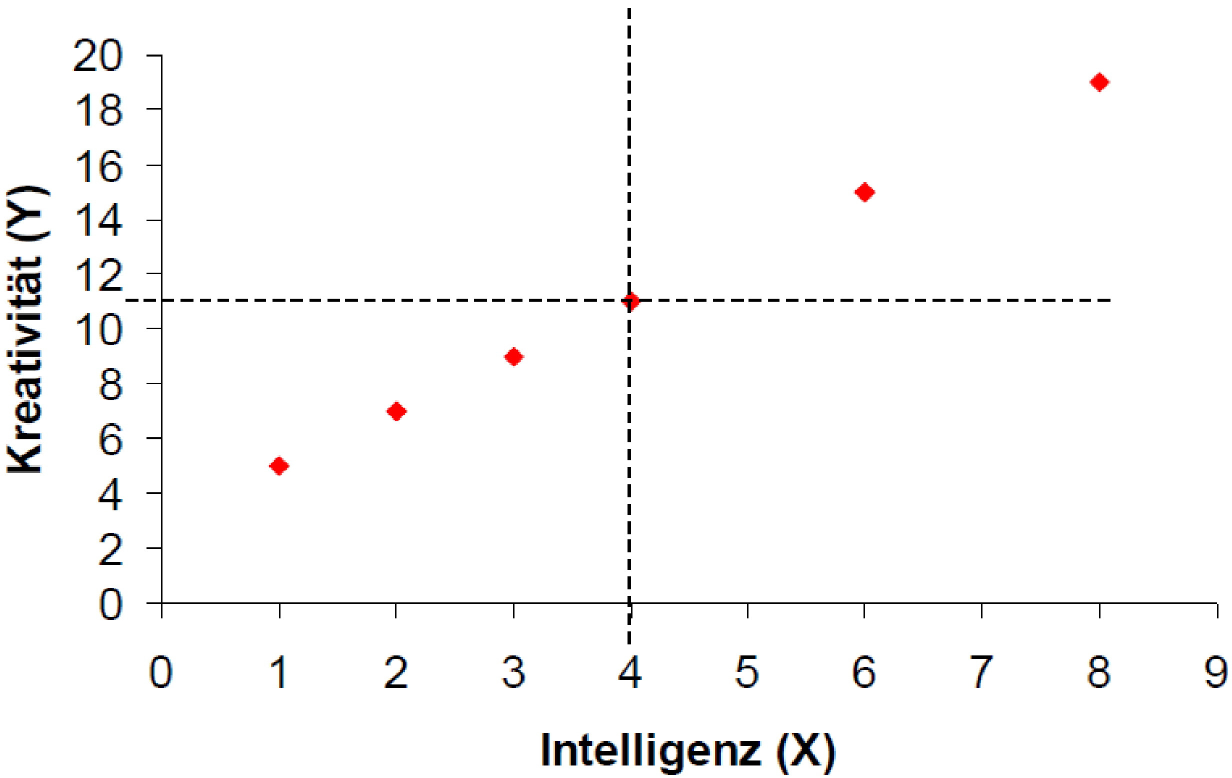



| | *X* | *Y* |
|---|---|---|
| | 1 | 5 |
| | 2 | 7 |
| | 2 | 7 |
| | 3 | 9 |
| | 4 | 11 |
| | 6 | 15 |
| | 6 | 15 |
| | 8 | 19 |
| $M$ = | 4 | 11 |
| $s^2$ = | 5,25 | 21 |

**Wann ist der Zusammenhang zwischen zwei Variablen X und Y positiv?**

# Pearson's r: Schritt für Schritt



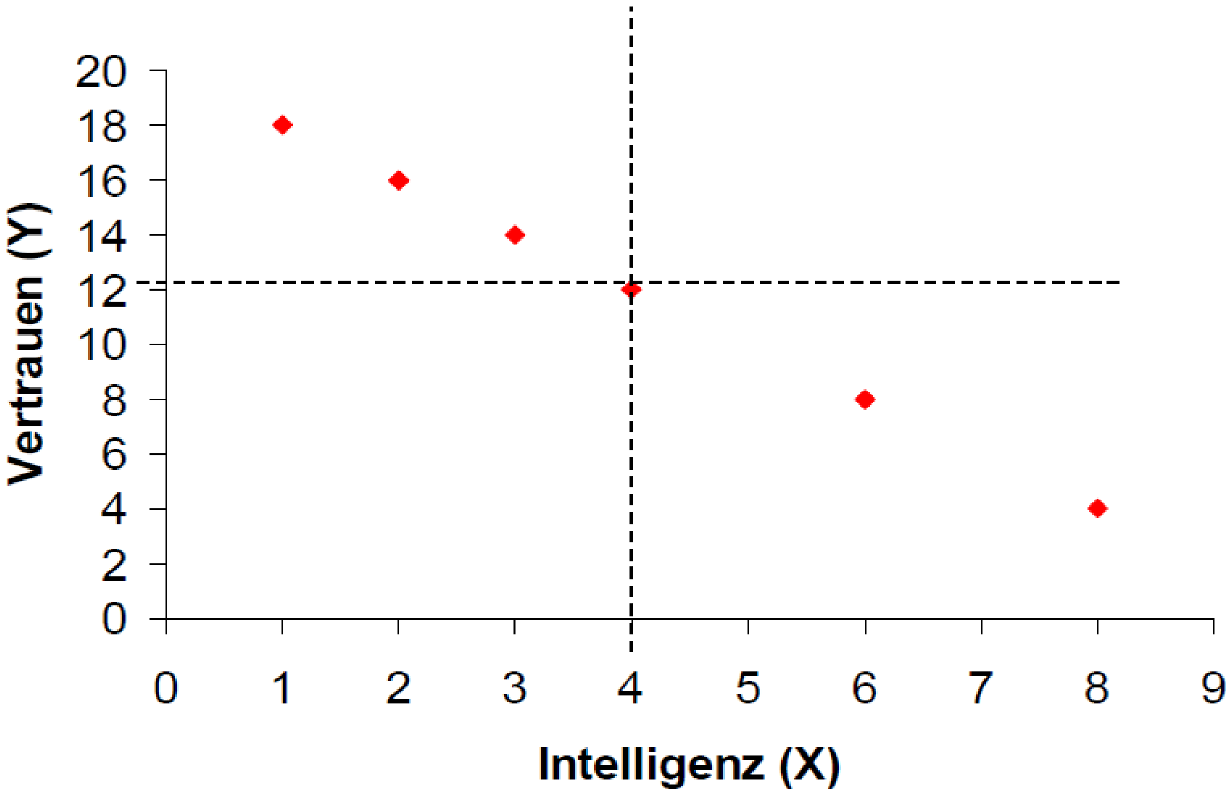



| | X | Y |
|---|---|---|
| | 1 | 5 |
| | 2 | 7 |
| | 2 | 7 |
| | 3 | 9 |
| | 4 | 11 |
| | 6 | 15 |
| | 6 | 15 |
| | 8 | 19 |
| $M$ = | 4 | 11 |
| $s^2$ = | 5,25 | 21 |

**Zusammenhang X und Y dann positiv, wenn x-Werte $> \bar{x}$ liegen, mit y-Werte $> \bar{y}$ einhergehen (und umgekehrt)**

# Pearson's r: Schritt für Schritt

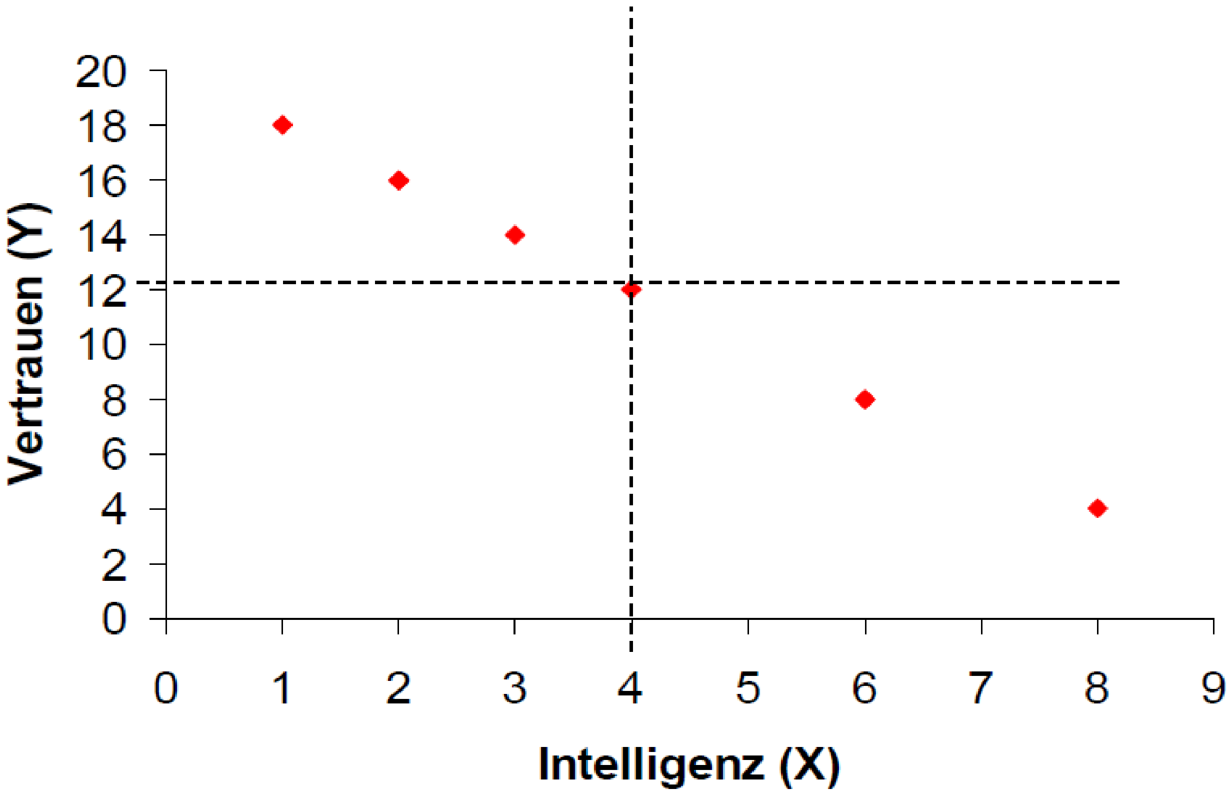





| | *X* | *Y* |
|---|---|---|
| | 1 | 18 |
| | 2 | 16 |
| | 2 | 16 |
| | 3 | 14 |
| | 4 | 12 |
| | 6 | 8 |
| | 6 | 8 |
| | 8 | 4 |
| $M$ = | 4 | 12 |
| $s^2$ = | 5,25 | 21 |

**Negativer Zusammenhang zwischen 2 Variablen X und Y?**

# Pearson‘s r: Schritt für Schritt

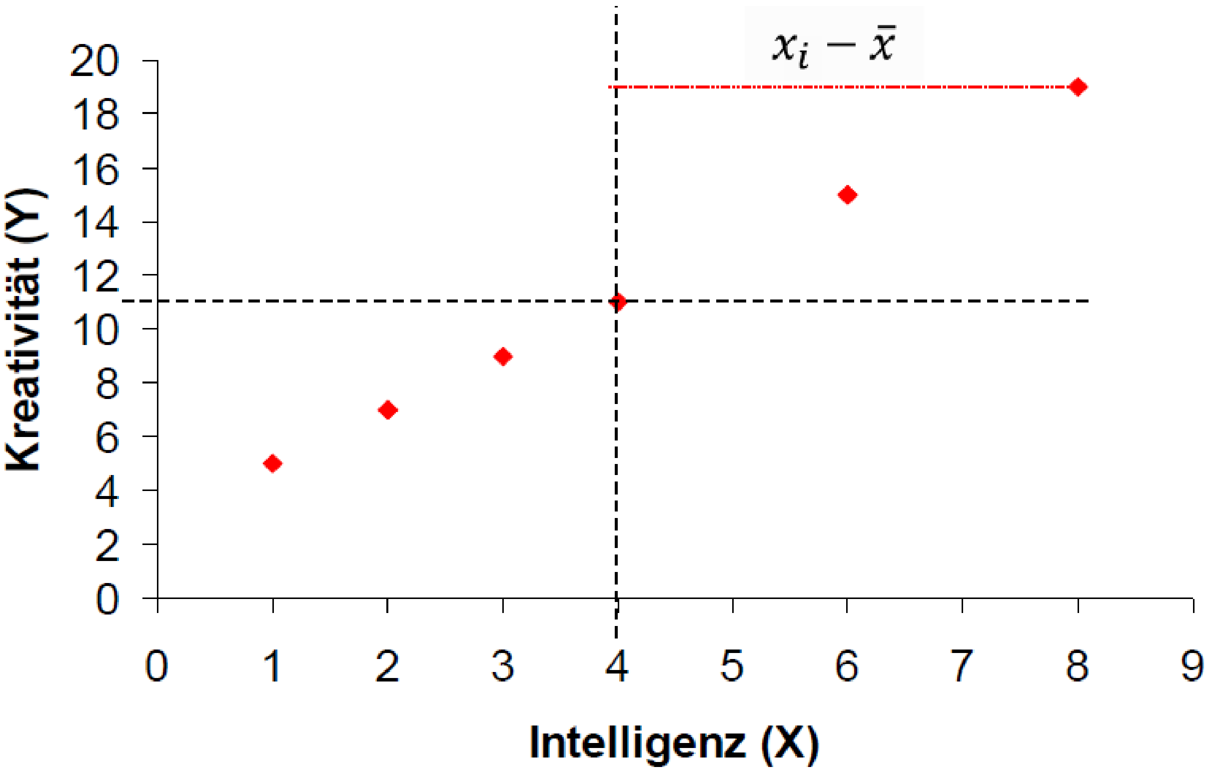



| | *X* | *Y* |
|---|---|---|
| | 1 | 18 |
| | 2 | 16 |
| | 2 | 16 |
| | 3 | 14 |
| | 4 | 12 |
| | 6 | 8 |
| | 6 | 8 |
| | 8 | 4 |
| $M$ = | 4 | 12 |
| $s^2$ = | 5,25 | 21 |

**Zusammenhang dann negativ, wenn zwischen 2 Variablen x-Werte >$\overline{x}$ einhergehen mit y-Werte <$\overline{y}$**

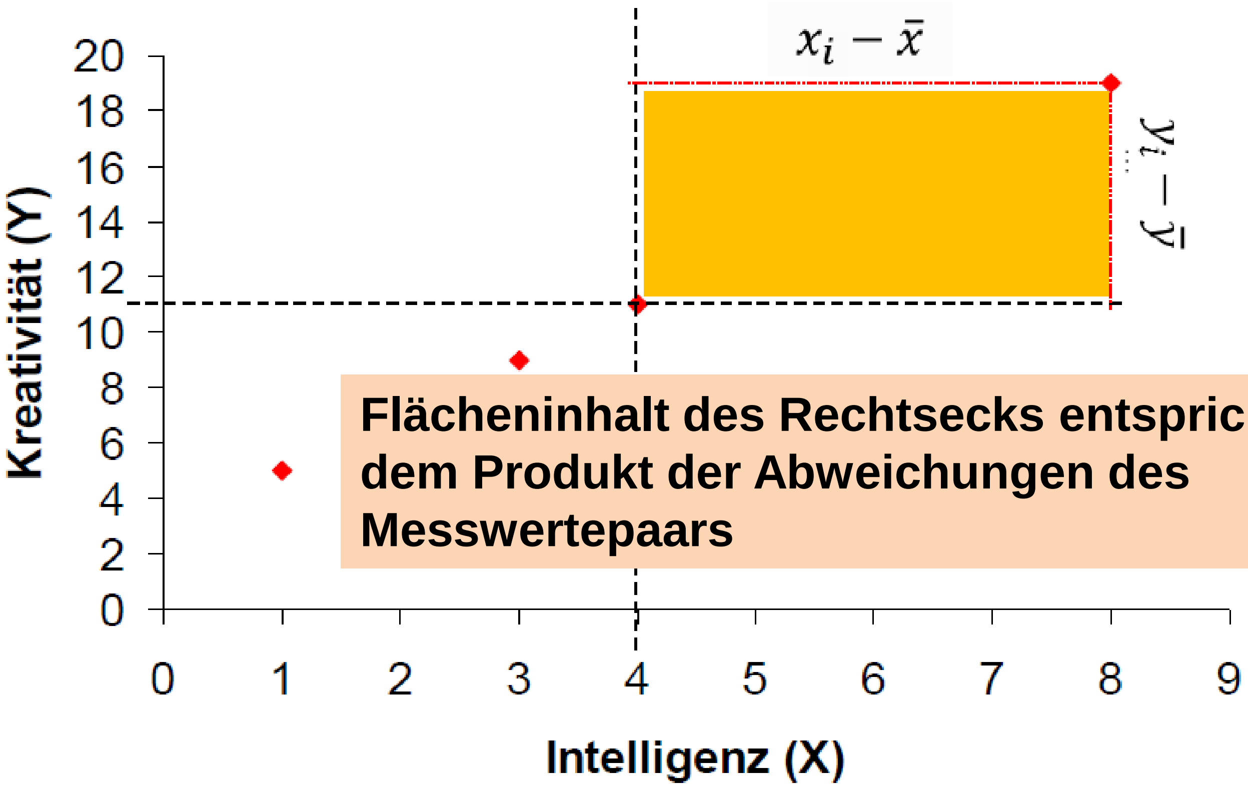
# Pearson's r: Schritt für Schritt






| | *X* | *Y* |
|---|---|---|
| | 1 | 5 |
| | 2 | 7 |
| | 2 | 7 |
| | 3 | 9 |
| | 4 | 11 |
| | 6 | 15 |
| | 6 | 15 |
| | 8 | 19 |
| *M* = | 4 | 11 |
| $s^2$ = | 5,25 | 21 |

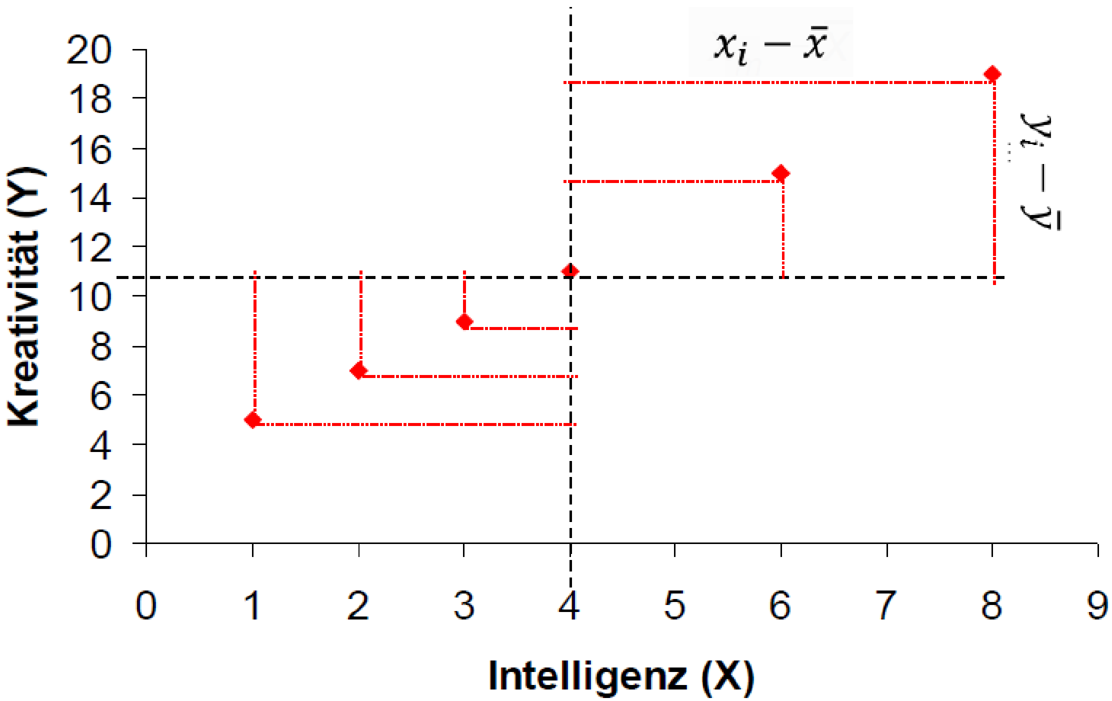
# Pearson‘s r: Schritt für Schritt




**Flächeninhalt des Rechtsecks entspricht dem Produkt der Abweichungen des Messwertepaars**

| | *X* | *Y* |
|---|---|---|
| | 1 | 5 |
| | 2 | 7 |
| | 2 | 7 |
| | 3 | 9 |
| | 4 | 11 |
| | 6 | 15 |
| | 6 | 15 |
| | 8 | 19 |
| $M$ = | 4 | 11 |
| $s^2$ = | 5,25 | 21 |

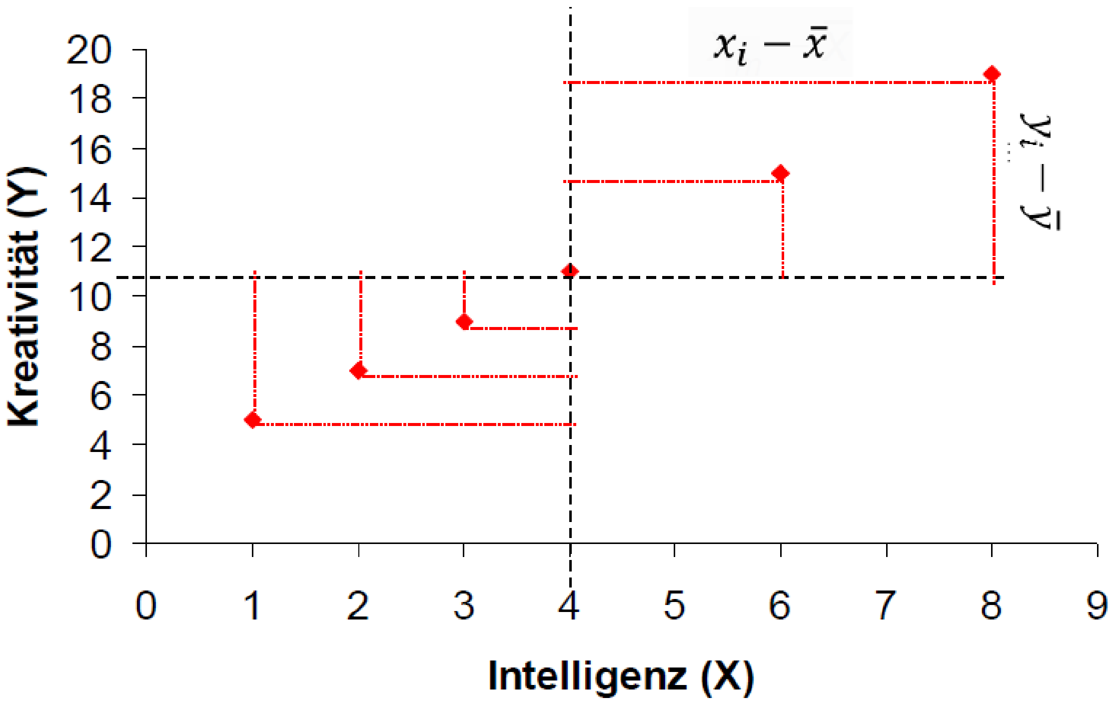
# Pearson‘s r: Schritt für Schritt




| $x_i - \bar{x}$ | $y_i - \bar{y}$ |
|---|---|
| 1 – 4 = -3 | 5 – 11 = -6 |
| 2 – 4 = -2 | 7 – 11 = -4 |
| 2 – 4 = -2 | 7 – 11 = -4 |
| 3 – 4 = -1 | 9 – 11 = -2 |
| 4 – 4 = 0 | 11 – 11 = 0 |
| 6 – 4 = 2 | 15 – 11 = 4 |
| 6 – 4 = 2 | 15 – 11 = 4 |
| 8 – 4 = 4 | 19 – 11 = 8 |

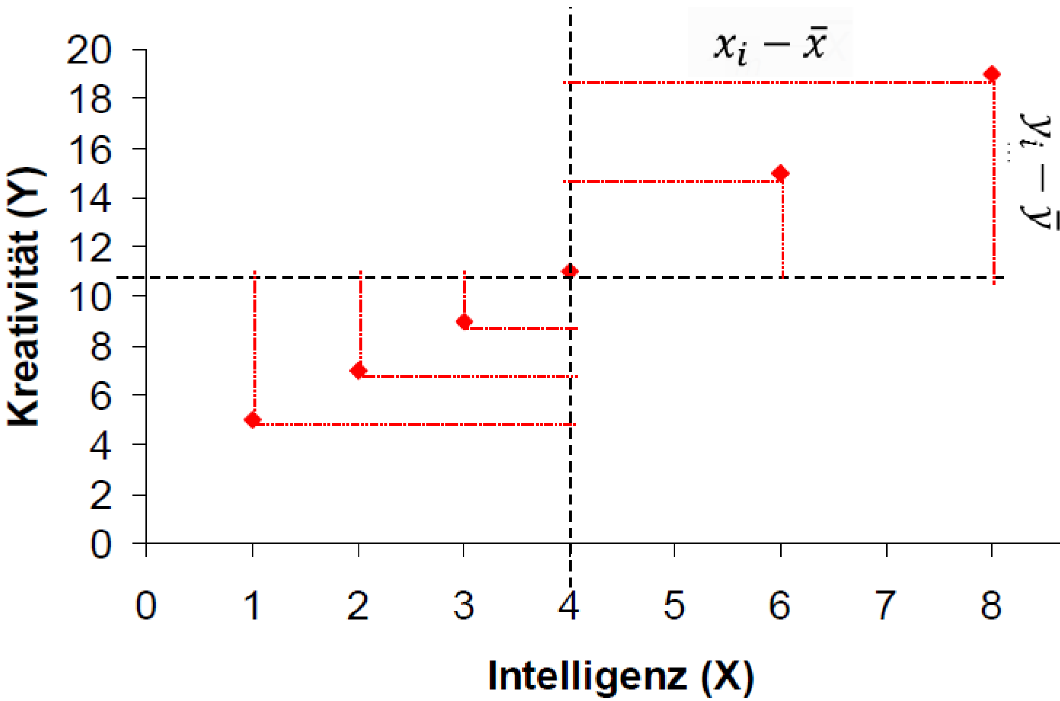
**Schritt 1: Für jedes $x_i$ sowie $y_i$ wird Differenz vom jeweiligen arithmetischen Mittel berechnet**

# Pearson‘s r: Schritt für Schritt




| $(x_i - \bar{x})(y_i - \bar{y})$ | | |
|---|---|---|
| -3 | -6 | 18 |
| -2 | -4 | 8 |
| -2 | -4 | 8 |
| -1 | -2 | 2 |
| 0 | 0 | 0 |
| 2 | 4 | 8 |
| 2 | 4 | 8 |
| 4 | 8 | 32 |

**Schritt 2: Für jedes Wertepaar $xy_i$ wird das *Kreuzprodukt*, d.h. das Produkt der Mittelwertsabweichung berechnet**

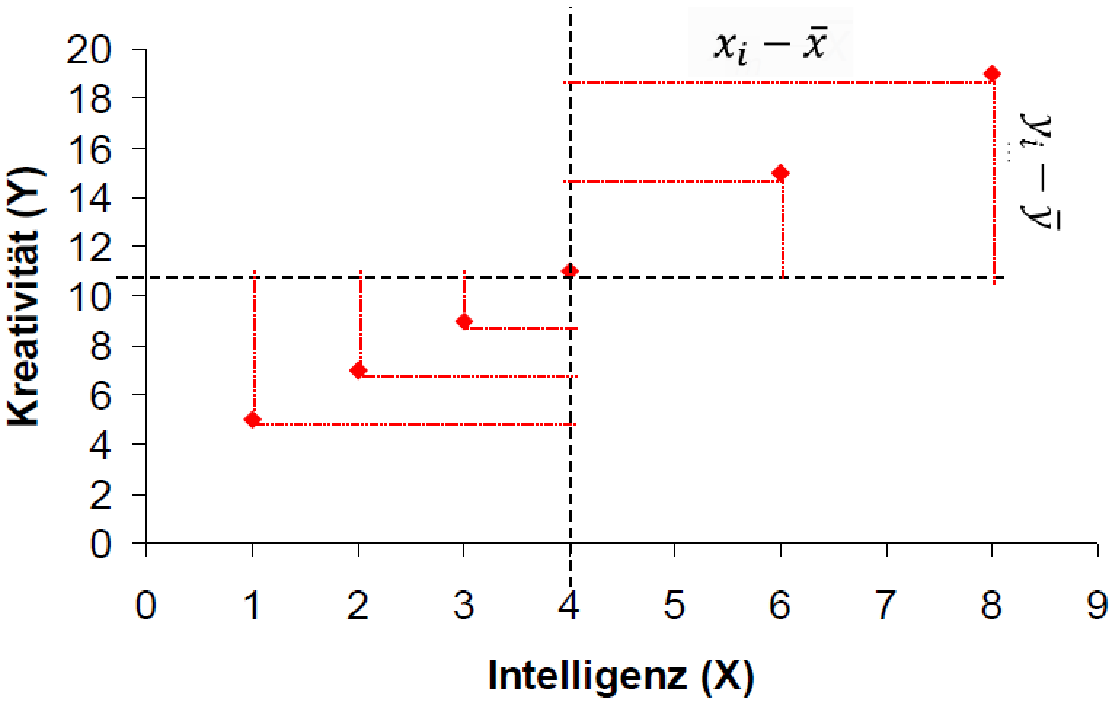



| $(x_i - \bar{x})(y_i - \bar{y})$ | | |
|---|---|---|
| -3 | -6 | 18 |
| -2 | -4 | 8 |
| -2 | -4 | 8 |
| -1 | -2 | 2 |
| 0 | 0 | 0 |
| 2 | 4 | 8 |
| 2 | 4 | 8 |
| 4 | 8 | 32 |
| Summe: | | 84 |

**Schritt 3: Berechnung der Kreuzproduktsumme, d.h. die Summe aller Kreuzprodukte von i = 1 bis n**

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

# Pearson's r: Schritt für Schritt



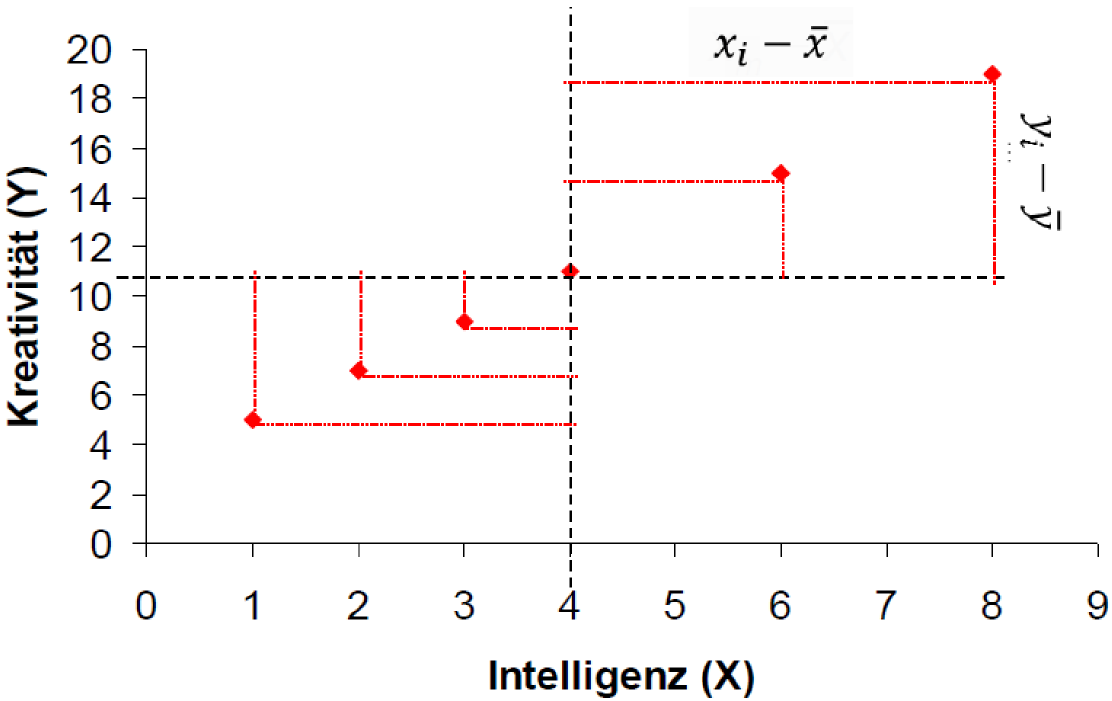

| $(x_i - \bar{x})(y_i - \bar{y})$ | | |
|---|---|---|
| -3 | -6 | 18 |
| -2 | -4 | 8 |
| -2 | -4 | 8 |
| -1 | -2 | 2 |
| 0 | 0 | 0 |
| 2 | 4 | 8 |
| 2 | 4 | 8 |
| 4 | 8 | 32 |

| | |
|---|---|
| Summe: | 84 |

| | |
|---|---|
| Kovarianz: | 10,5 |

**Schritt 4: Berechnung Kovarianz, indem durch n geteilt wird („mittleres Kreuzprodukt")**

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




$$s_X^2 = 5{,}25 \qquad s_X = 2{,}29$$

$$s_Y^2 = 21 \qquad s_Y = 4{,}58$$

$$s_X \cdot s_Y = 2{,}29 \cdot 4{,}58 = 10{,}5$$

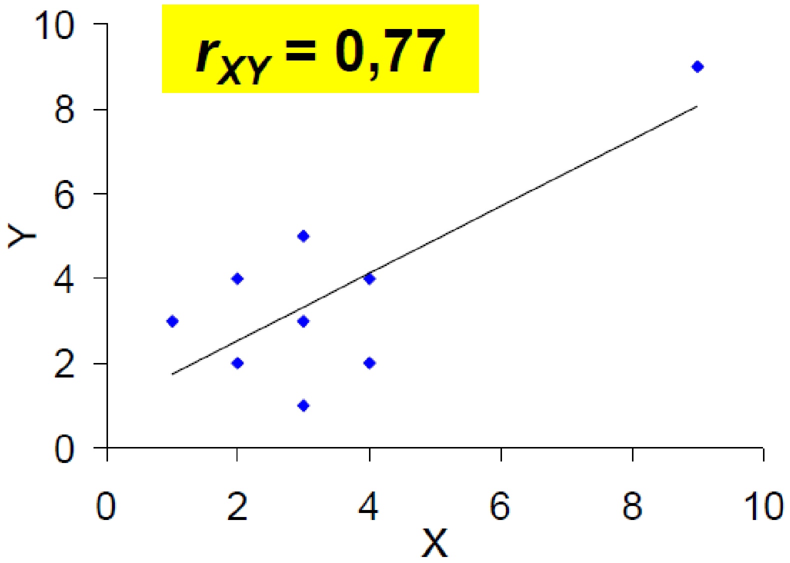
$$r = \frac{Cov_{xy}}{s_x s_y}$$

$$\mathrm{r} = \frac{10{,}5}{10{,}5} = 1$$

**Schritt 5: Berechnung Pearson‘s r durch Relativierung der empirischen Kovarianz an der maximalen Kovarianz**

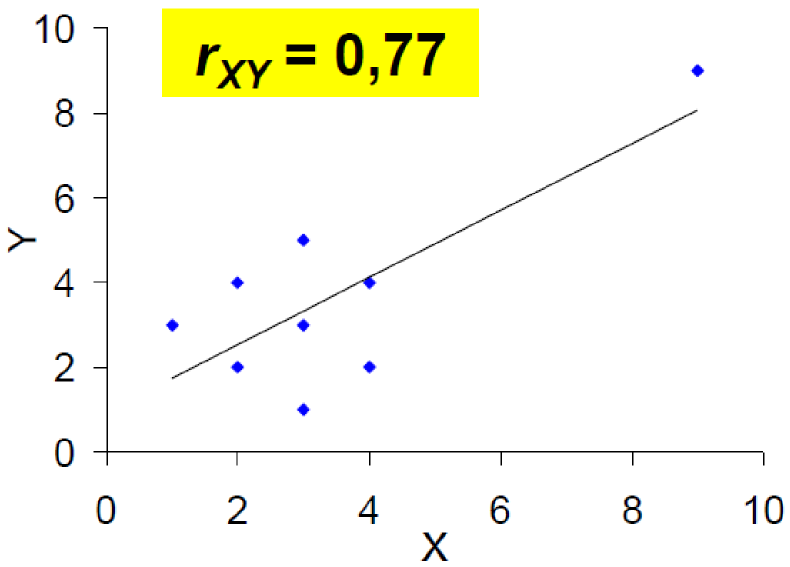
# Pearson‘s r: Ergänzung

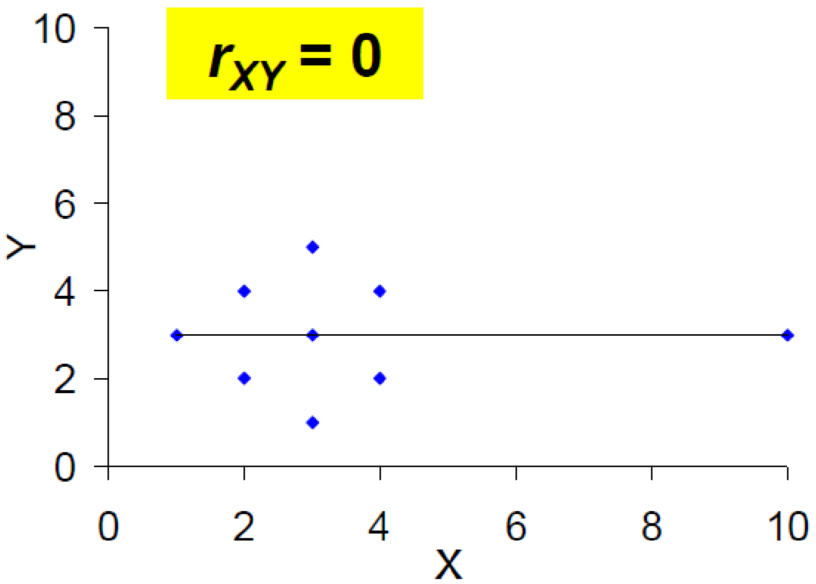
**Korrelationskoeffizienten sind sensitiv gegenüber Ausreißern und Extremwerten, v.a. bei kleinen Stichproben**



# Pearson‘s r: Ergänzung

**Korrelationskoeffizienten sind sensitiv gegenüber Ausreißern und Extremwerten, v.a. bei kleinen Stichproben**

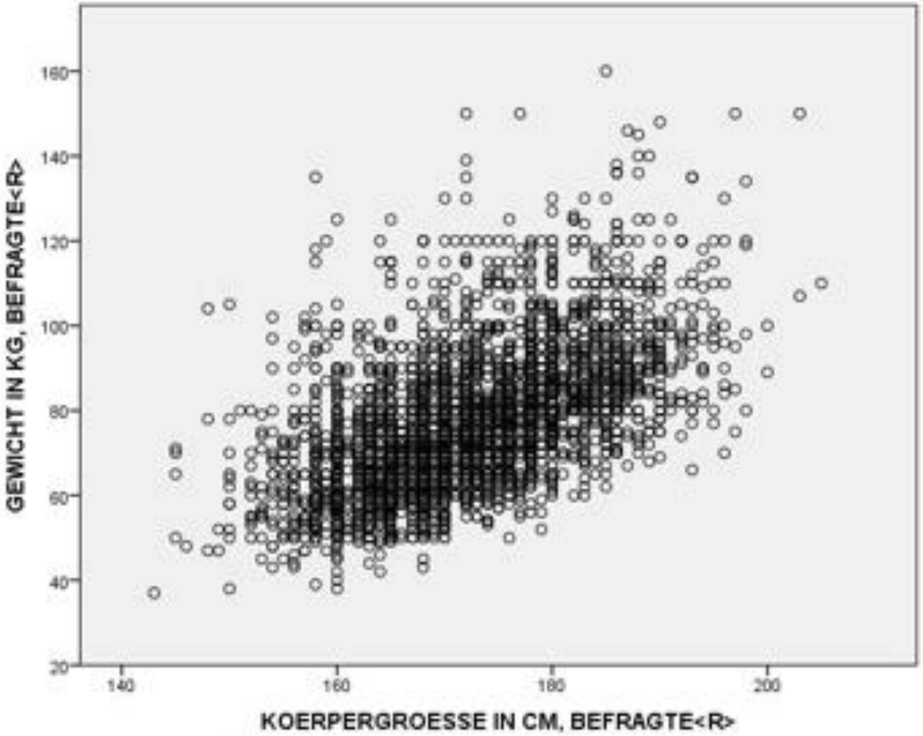


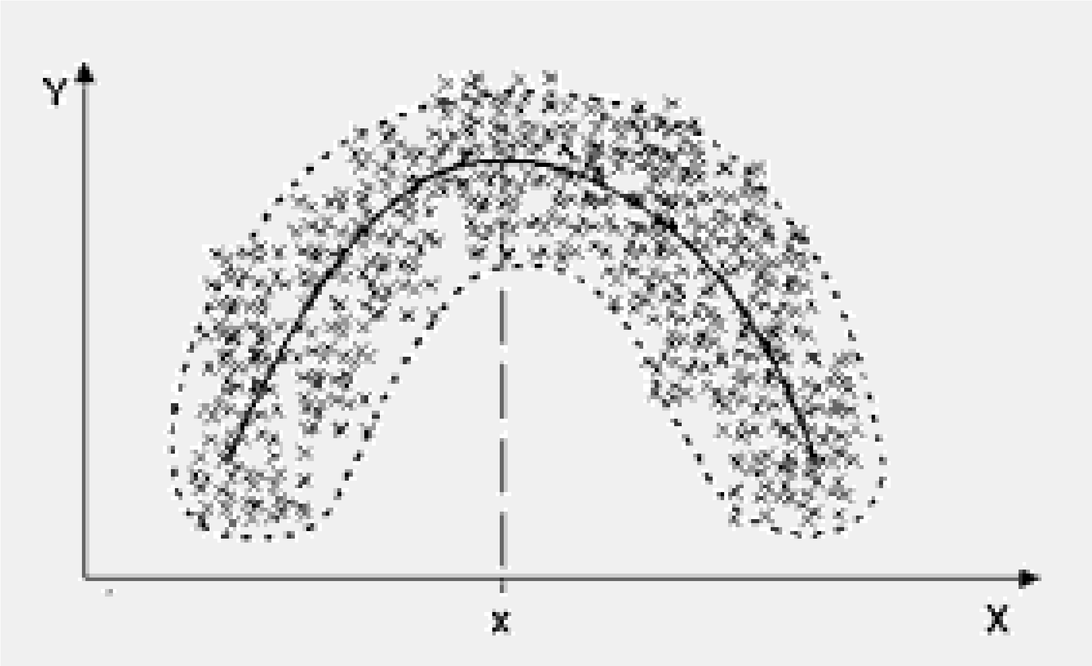
# Pearson‘s r Ergänzung

**Zusammenhangshypothese: „Je größer man ist, desto schwerer ist man auch“ (Streudiagramm erstellt auf Basis des Allbus 2016)**

r=0,547



**Keine lineare Korrelation! r=0** (Bsp. Leistungsfähigkeit und Anspannung während Klausur)



# Pearson‘s r: Übung

1) Zeichnen Sie ein Streudiagramm
2) Berechnen Sie Pearson‘s r
3) Interpretieren Sie das Ergebnis

| | $x_i$ | $y_i$ | $(x_i-\bar{x})$ | $(y_i-\bar{y})$ | $(x_i-\bar{x})(y_i-\bar{y})$ | $(x_i-\bar{x})^2$ | $(y_i-\bar{y})^2$ |
|---|---|---|---|---|---|---|---|
| A | 0 | 2 | | | | | |
| B | 10 | 6 | | | | | |
| C | 4 | 2 | | | | | |
| D | 8 | 4 | | | | | |
| E | 8 | 6 | | | | | |
| | | | | | | | |

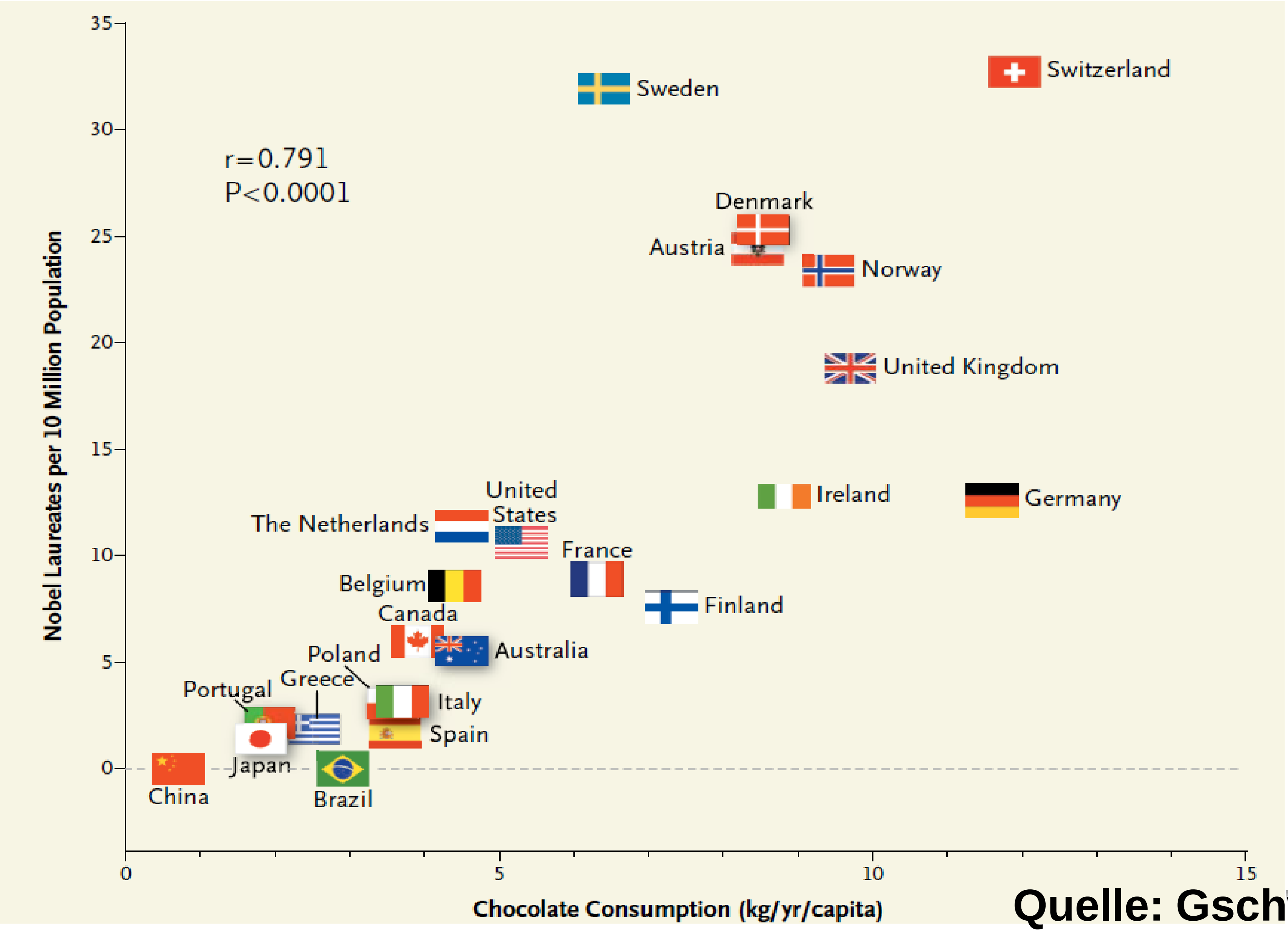
# Wichtig



- **Korrelationsberechnung braucht inhaltliche Theorie bzw. Plausibilität!**
- **Korrelation ≠ Kausalität!**

# Theorie???Plausibilität???






Quelle: Gschwend 2013

# Plan für heute



- **Zusammenhänge zwischen (pseudo-)metrischen Merkmalen**
  - Allgemein
  - Hypothesen
  - Grafische Veranschaulichung
  - Zusammenhangsmaß Pearson‘s Rho
- **Zusammenhangsmaß zwischen metrischem und gruppiertem Merkmal: PRE-Maß $\eta 2$ (Eta-Quadrat)**

# Wiederholung PRE-Maße



- PRE: „Proportional Reduction of Error“
- Verschiedene PRE-Maße in der Statistik
- **Ausgangsfrage:** Wie gut können die Werte einer abhängigen Variable durch die Werte einer unabhängigen Variable vorhergesagt werden?
- Folgen der **gleichen Logik**
- Setzen eine **„gerichtete“ Hypothese** voraus, d.h. wir kennen abhängige und unabhängige Variable

# Wiederholung PRE-Maße

**Schrittweises Vorgehen:**

1) Wie lautet die Prognose des Wertes der abhängigen Variable **ohne** Kenntnis der unabhängigen Variablen? (Vorhersagefehler E1)
2) Prognose des Wertes der abhängigen Variable **mit** Kenntnis der Verteilung der unabhängigen Variable (Vorhersagefehler E2)
3) Ermittlung des PRE-Maßes und Aussage, ob die Vorhersage durch die unabhängige Variable verbessert wurde PRE = (E1 - E2) / E1

**Je nach PRE-Maß werden die Fehler dabei unterschiedlich berechnet**

# PRE-Maß $\eta$2 (Eta-Quadrat)



- Auch **Eta-Koeffizient**
- **Abhängige Variable** muss mind. (pseudo-) Intervallskalenniveau aufweisen
- **Unabhängige Variable** kann beliebiges (aber gruppiertes) anderes Niveau aufweisen

# PRE-Maß $\eta 2$ (Eta-Quadrat)



- Berechnung basiert auf **Quadratsummen**
- Quadratsumme stellt die **Summe der quadrierten Abweichungen vom arithmetischen Mittelwert** eines Merkmals dar
- **Verschiedene Quadratsummen werden unterschieden:**
  - **„Quadratsumme Gesamt“:** Entspricht der Summe aller quadrierten Abweichungen vom Mittelwert der (abhängigen) Variable **(Vorhersagefehler E1)**.
  - **„Quadratsumme innerhalb“**: Wird die unabhängige Gruppenvariable berücksichtigt, wird innerhalb der Gruppen berechnet, wie stark die Gruppenmittelwerte vom Gesamtmittelwert abweichen **(Vorhersagefehler E2)**.
  - **„Quadratsumme zwischen“:** „Quadratsumme Gesamt“- „Quadratsumme innerhalb“

# PRE-Maß $\eta 2$ (Eta-Quadrat)



PRE = (E1 - E2) / E1

$$\eta 2 = \frac{\text{Quadratsumme Gesamt} - \text{Quadratsumme innerhalb}}{\text{Quadratsumme Gesamt}} = \frac{\text{Quadratsumme Zwischen}}{\text{Quadratsumme Gesamt}}$$

# Beispiel PRE-Maß $\eta 2$ (Eta-Quadrat)



## Politisches Wissen von Schüler*innen mit und ohne Migrationshintergrund

| ID | Migrationshintergrund | Politisches Wissen ($y_i$) |
|---|---|---|
| 1 | 1 (Nein) | 15 |
| 2 | 2 (Ja) | 10 |
| 3 | 2 (Ja) | 2 |
| 4 | 2 (Ja) | 9 |
| 5 | 1 (Nein) | 7 |
| 6 | 1 (Nein) | 16 |
| 7 | 1 (Nein) | 14 |
| 8 | 2 (Ja) | 6 |
| 9 | 2 (Ja) | 4 |
| 10 | 1 (Nein) | 12 |

Quelle: Eigene Darstellung

Arithmetisches Mittel: 9,5

# PRE-Maß $\eta 2$ (Eta-Quadrat)



Schritt 1

- Arithmetisches Mittel als bester Vorhersagewert, wenn keine unabhängige Variable berücksichtigt wird

Schritt 2

- Berücksichtigung uV

Schritt 3

- Ermittlung Eta-Koeffizient

# PRE-Maß $\eta 2$ (Eta-Quadrat)



Tabelle 43: Arbeitstabelle Migrationshintergrund und politisches Wissen

| ID | Migrations-hintergrund | Politisches Wissen ($y_i$) | $y_i - \bar{y}$ | $(y_i - \bar{y})^2$ |
|---|---|---|---|---|
| 1 | 1 (Nein) | 15 | 5,5 | 30,25 |
| 2 | 2 (Ja) | 10 | 0,5 | 0,25 |
| 3 | 2 (Ja) | 2 | –7,5 | 56,25 |
| 4 | 2 (Ja) | 9 | –0,5 | 0,25 |
| 5 | 1 (Nein) | 7 | –2,5 | 6,25 |
| 6 | 1 (Nein) | 16 | 6,5 | 42,25 |
| 7 | 1 (Nein) | 14 | 4,5 | 20,25 |
| 8 | 2 (Ja) | 6 | –3,5 | 12,25 |
| 9 | 2 (Ja) | 4 | –5,5 | 30,25 |
| 10 | 1 (Nein) | 12 | 2,5 | 6,25 |
| | | $\sum = 95$ | | $\sum = 204{,}5$ |
| | | $\bar{y} = 9{,}5$ | | |

Quelle: Eigene Darstellung

**E1: Quadratsumme Gesamt**

# PRE-Maß $\eta 2$ (Eta-Quadrat)

| ID | Migrations-hintergrund | Politisches Wissen ($y_i$) | $y_i - \bar{y}$ | $(y_i - \bar{y})^2$ |
|---|---|---|---|---|
| 1 | 1 (Nein) | 15 | 2,2 | 4,84 |
| 5 | 1 (Nein) | 7 | –5,8 | 33,64 |
| 6 | 1 (Nein) | 16 | 3,2 | 10,24 |
| 7 | 1 (Nein) | 14 | 1,2 | 1,44 |
| 10 | 1 (Nein) | 12 | –0,8 | 0,64 |
| | | $\sum = 64$ | | $\sum = 50,8$ |
| | | $\bar{y} = 12,8$ | | |

Quelle: Eigene Darstellung

**Quadratsumme Gruppe „Nein“**

# PRE-Maß $\eta 2$ (Eta-Quadrat)

| ID | Migrations-hintergrund | Politisches Wissen ($y_i$) | $y_i - \bar{y}$ | $(y_i - \bar{y})^2$ |
|---|---|---|---|---|
| 2 | 2 (Ja) | 10 | 3,8 | 14,44 |
| 3 | 2 (Ja) | 2 | –4,2 | 17,64 |
| 4 | 2 (Ja) | 9 | 2,8 | 7,84 |
| 8 | 2 (Ja) | 6 | –0,2 | 0,04 |
| 9 | 2 (Ja) | 4 | –2,2 | 4,84 |
| | | $\sum = 31$ | | $\sum = 44,8$ |
| | | $\bar{y} = 6,2$ | | |

Quelle: Eigene Darstellung

**Quadratsumme Gruppe „Ja“**

# PRE-Maß $\eta 2$ (Eta-Quadrat)

$$\eta 2 = \frac{\text{Quadratsumme Gesamt} - \text{Quadratsumme innerhalb}}{\text{Quadratsumme Gesamt}} = \frac{\text{Quadratsumme Zwischen}}{\text{Quadratsumme Gesamt}}$$

**E2: 44,8+50,8=95,6**

$$\eta 2 = \frac{E1 - E2}{E1} = \frac{204{,}5 - (50{,}8 + 44{,}8)}{204{,}5} = \frac{204{,}5 - 95{,}6}{204{,}5} = \frac{108{,}9}{204{,}5} = 0{,}53$$

Schätzfehler kann um 53% vermindert werden, wenn wir die uV berücksichtigen

# PRE-Maß $\eta 2$ (Eta-Quadrat)



- Kann Werte zwischen 0 und 1 annehmen

Tabelle 41: Interpretation von Eta-Quadrat

| Eta-Quadrat | Interpretation |
|---|---|
| < 0,01 | kein Effekt |
| 0,01 bis < 0,06 | kleiner Effekt |
| 0,06 bis < 0,14 | mittlerer Effekt |
| ≥ 0,14 | großer Effekt |

Quelle: Eigene Darstellung

# Übung: PRE-Maß $\eta 2$ (Eta-Quadrat)



| ID | Taschengeld pro Monat in Euro | Gruppe |
|---|---|---|
| 1 | 3 | 1 |
| 2 | 9 | 2 |
| 3 | 10 | 1 |
| 4 | 4 | 2 |
| 5 | 2 | 1 |
| 6 | 4 | 1 |
| 7 | 3 | 2 |
| 8 | 6 | 1 |
| 9 | 5 | 2 |
| 10 | 6 | 2 |

Sie wollen untersuchen, ob ältere Grundschulkinder tatsächlich mehr Taschengeld bekommen als jüngere und haben dazu folgende Daten ermittelt und zwei Kinder-Gruppen gebildet: 1: Klasse 1 und 2 vs. 2: Klasse 3 und 4 – Berechnen Sie Eta-Quadrat und interpretieren Sie das Ergebnis!

# Zusammenfassung bivariate Analysen



**Berechnung**

| Messniveau | nominal | ordinal | metrisch |
|---|---|---|---|
| nominal | **Chi-Quadrat, Cramer‘s V Lambda C** | **Cramer‘s V Lambda C** | **Eta-Koeffizient** Mittelwertvergleich (t-test) |
| ordinal | **Cramer‘s V Lambda C** | **Spearman‘s Rho;** (Kendalls Tau B) gamma | **Spearman‘s Rho;** (Kendalls Tau B) gamma |
| metrisch | **Eta-Koeffizient** Mittelwertvergleich (t-test) | **Spearman‘s Rho;** (Kendalls Tau B) gamma | **Pearson‘s r** |

# Zusammenfassung bivariate Analyse



**Darstellung**

| Messniveau | nominal | ordinal | metrisch |
|---|---|---|---|
| nominal | Kreuztabelle | Kreuztabelle | (gruppierte) Boxplots |
| ordinal | | Kreuztabelle | (gruppierte) Boxplots |
| metrisch | | | Streudiagramm |

# Zusammenfassung bivariate Analysen



**Was können bivariate Analysen leisten?**

- Beschreibung der **Zusammenhänge** zwischen 2 Variablen
- **Kreuztabellen** und die damit verbundenen Maße (Chi-Quadrat, Cramers V) können Auskunft über Zusammenhänge zwischen nominalen Variablen liefern
- **Korrelationskoeffizienten** (Spearman/Pearson) beschreiben die Stärke „gleichsinniger“ oder gegenläufiger Zusammenhänge zweier Variablen
- **PRE-Maße** (Lambda, Eta) geben an, inwieweit eine unabhängige Variable durch Einbezug einer weiteren Variable „besser“ erklärt wird

# Was können bivariate Analysen nicht leisten?



- Überprüfen, inwieweit die Zusammenhänge zwischen zwei Variablen womöglich noch von weiteren Variablen abhängig sind
- Mehrere unabhängige Variablen in ihrem gemeinsamen Einfluss auf eine abhängige Variable untersuchen
- Kausalitätsrichtung bestimmen