

BSc PHY – Angewandte Statistik

LE5 Korrelation und Regression

Patric Eichelberger & Aglaja Busch
aF&E Physiotherapie

✉ patric.eichelberger@bfh.ch | aglaja.busch@bfh.ch
🌐 Moodlekurs

29. Juli 2025



Intro

Korrelation

Pearson Korrelationskoeffizient
Spearman
Korrelationskoeffizient

Einfache lineare
Regression

Regressionsprüfung

Multiple lineare
Regression

Praxis mit R

Zusammenfassung

Intro

Intro

Korrelation

Pearson Korrelationskoeffizient
Spearman
Korrelationskoeffizient

Einfache lineare
Regression

Regressionsprüfung

Multiple lineare
Regression

Praxis mit R

Zusammenfassung

Rückblick Statistische Tests

- ▶ Wir haben gesehen wie ein statistischer Test abläuft.
- ▶ Dabei werden aus einer Fragestellung statistische Hypothesen formuliert.
- ▶ Um die Hypothesen zu testen müssen Daten erhoben werden.
- ▶ Je nach Versuchsdesign und Verteilung der Daten ist ein passender Test auszuwählen.
- ▶ Die Entscheidung anhand eines Tests heisst nicht, dass es in der Wirklichkeit so ist.
- ▶ Statistische Signifikanz heisst nicht klinische Relevanz! Differenzen und Effektgrössen beachten!

Beispiel

Gerichtsprozess vs. klinische Studie – Übernommen von Terry Mills **Legal vs clinical trials: an explanation of sampling errors and sample size**

Intro

Korrelation

Pearson Korrelationskoeffizient
Spearman
Korrelationskoeffizient

Einfache lineare
Regression

Regressionsprüfung

Multiple lineare
Regression

Praxis mit R

Zusammenfassung

Gerichtsprozess

- ▶ Unschuldsvermutung

Klinische Studie

- ▶ Nullhypothese (H_0)
- ▶ Neue Behandlung ist nicht besser als Standardbehandlung

Intro

Korrelation

Pearson Korrelationskoeffizient
Spearman
Korrelationskoeffizient

Einfache lineare
Regression

Regressionsprüfung

Multiple lineare
Regression

Praxis mit R

Zusammenfassung

Gerichtsprozess

- ▶ Alternative: Angeklagter ist schuldig

Klinische Studie

- ▶ Alternativhypothese (H_1)
- ▶ Neue Behandlung ist besser als Standardbehandlung

Intro

Korrelation

Pearson Korrelationskoeffizient
Spearman
Korrelationskoeffizient

Einfache lineare
Regression

Regressionsprüfung

Multiple lineare
Regression

Praxis mit R

Zusammenfassung

Gerichtsprozess

Gerichtsanhörung:

- ▶ Beweise werden gesammelt und präsentiert

Klinische Studie

Durchführung der Studie:

- ▶ Daten werden gesammelt und analysiert

Intro

Korrelation

Pearson Korrelationskoeffizient
Spearman
Korrelationskoeffizient

Einfache lineare
Regression

Regressionsprüfung

Multiple lineare
Regression

Praxis mit R

Zusammenfassung

Entscheidung

Gerichtsprozess

Urteil:

- ▶ „schuldig“
- ▶ oder...
- ▶ „nicht schuldig“ (Gericht entscheidet nicht ob die Person unschuldig ist!)

Klinische Studie

Testergebnis:

- ▶ „Datenlage zeigt Überlegenheit an“ (H_0 verwerfen)
- ▶ oder...
- ▶ „Datenlage zeigt Überlegenheit nicht an“ (H_0 nicht verwerfen; nicht: H_0 annehmen!)

Intro

Korrelation

Pearson Korrelationskoeffizient
Spearman
Korrelationskoeffizient

Einfache lineare
Regression

Regressionsprüfung

Multiple lineare
Regression

Praxis mit R

Zusammenfassung

Mögliche Fehler

Gerichtsprozess

Urteil:

- ▶ Unschuldige Person schuldig gesprochen
- ▶ oder...
- ▶ Schuldige Person nicht schuldig gesprochen

Klinische Studie

Testergebnis:

- ▶ Daten zeigen an, dass die neue Behandlung überlegen ist, obwohl es in Wirklichkeit nicht so ist. Wahre H_0 verwerfen (Fehler 1. Art).
- ▶ oder...
- ▶ Daten zeigen nicht an, dass die neue Behandlung überlegen ist, obwohl es in Wirklichkeit so ist. Falsche H_0 nicht verwerfen (Fehler 2. Art).

Intro

Korrelation

Pearson Korrelationskoeffizient
Spearman
Korrelationskoeffizient

Einfache lineare
Regression

Regressionsprüfung

Multiple lineare
Regression

Praxis mit R

Zusammenfassung

		Realität	
		H_0 wahr	H_1 wahr
Forschung	H_0 wahr	Keine Entdeckung $1-\alpha$ 	Entdeckung verpasst Fehler 2. Art β 
	H_1 wahr	Falsche Entdeckung Fehler 1. Art α 	Entdeckung Power $1-\beta$ 

Intro

Korrelation

Pearson Korrelationskoeffizient

Spearman

Korrelationskoeffizient

Einfache lineare
Regression

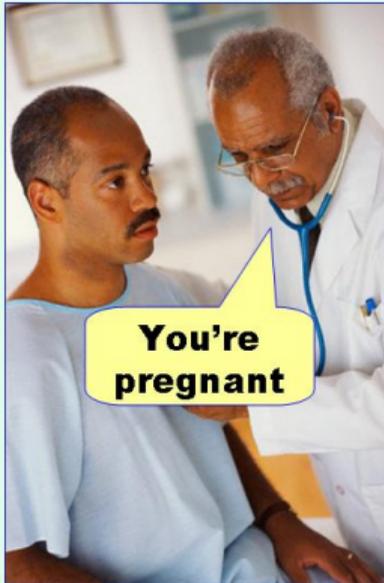
Regressionsprüfung

Multiple lineare
Regression

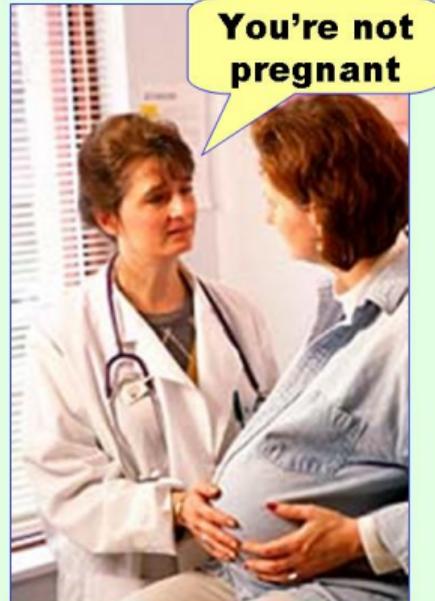
Praxis mit R

Zusammenfassung

Type I error
(false positive)



Type II error
(false negative)



Quelle: effectsizefaq.com

Intro

Korrelation

Pearson Korrelationskoeffizient
Spearman
Korrelationskoeffizient

Einfache lineare
Regression

Regressionsprüfung

Multiple lineare
Regression

Praxis mit R

Zusammenfassung

Rückblick Statistische Tests: p-Wert

Definition p-Wert

Der p-Wert gibt die Wahrscheinlichkeit an, unter Annahme der Nullhypothese ein Ergebnis zu erhalten, das mindestens so extrem ist wie die im Experiment beobachteten Resultate.

- ▶ Angenommen die Nullhypothese ist wahr: Je kleiner der p-Wert, desto unwahrscheinlicher ist es ein Stichprobenresultat per Zufall zu beobachten.
- ▶ Ist diese Wahrscheinlichkeit kleiner als das Signifikanzniveau (oft 5%), wird von einem signifikanten Testergebnis gesprochen.
- ▶ Ob der Schwellenwert unter- oder überschritten wird, entscheidet vor allem die Stichprobengrösse
- ▶ **Der p-Wert ist keine Interpretation, er braucht Interpretation!**

M. Baker, "Statisticians issue warning over misuse of p values," vol. 531, pp. 151–151, 2016.

<https://www.horizonte-magazin.ch/2019/03/07/braucht-es-den-statistischen-schwellenwert-noch/>

<https://www.nzz.ch/karriere/studentenleben/der-problem-wert-ld.133818>

Rückblick Statistische Tests: p-Wert (cont)

Häufige Fehlinterpretationen

- ▶ Keine Signifikanz heisst, es gibt keinen Effekt
- ▶ Wahrscheinlichkeit, dass die Nullhypothese stimmt
- ▶ Abschätzung dafür, wie verlässlich die Ergebnisse sind
- ▶ Mass für die wissenschaftliche Bedeutung der Resultate
- ▶ Kriterium, um wissenschaftliche Studien miteinander zu vergleichen
- ▶ Ist der p-Wert kleiner als ein bestimmter Schwellenwert, darf die Studie veröffentlicht werden. Ist der p-Wert gross, bleibt sie in der Schublade.

M. Baker, "Statisticians issue warning over misuse of p values," vol. 531, pp. 151–151, 2016.

<https://www.horizonte-magazin.ch/2019/03/07/braucht-es-den-statistischen-schwellenwert-noch/>

<https://www.nzz.ch/karriere/studentenleben/der-problem-wert-ld.133818>

Lernziele Korrelation und Regression

- ▶ Die Studierenden berechnen und interpretieren Korrelations- und den Rangkorrelationskoeffizienten.
- ▶ Die Studierenden modellieren den Zusammenhang zwischen zwei Variablen mit Hilfe einfacher linearer Regression.
- ▶ Die Studierenden wenden ein lineares Modell an um Vorhersagen zu treffen.
- ▶ Die Studierenden erläutern die Bedeutung einer Regressionsprüfung und beschreiben die wichtigsten Voraussetzungen (Linearität, Normalverteilung der Residuen, Konstante Variabilität).

Intro

Korrelation

Pearson Korrelationskoeffizient
Spearman
Korrelationskoeffizient

Einfache lineare
Regression

Regressionsprüfung

Multiple lineare
Regression

Praxis mit R

Zusammenfassung

Korrelation

Intro

Korrelation

Pearson Korrelationskoeffizient

Spearman

Korrelationskoeffizient

Einfache lineare
Regression

Regressionsprüfung

Multiple lineare
Regression

Praxis mit R

Zusammenfassung

Zusammenhang zweier Variablen

- ▶ Oft ist der Zusammenhang zwischen zwei stetigen Variablen von Interesse.
- ▶ Dieser Zusammenhang kann mit Hilfe von Streudiagrammen, Korrelations- und Rangkorrelationskoeffizienten beschrieben werden.

Beispiel

Der Zusammenhang zwischen der Rotation der Hüfte INNEN passiv in Flexion und Extension soll untersucht werden.

Intro

Korrelation

Pearson Korrelationskoeffizient

Spearman

Korrelationskoeffizient

Einfache lineare
Regression

Regressionsprüfung

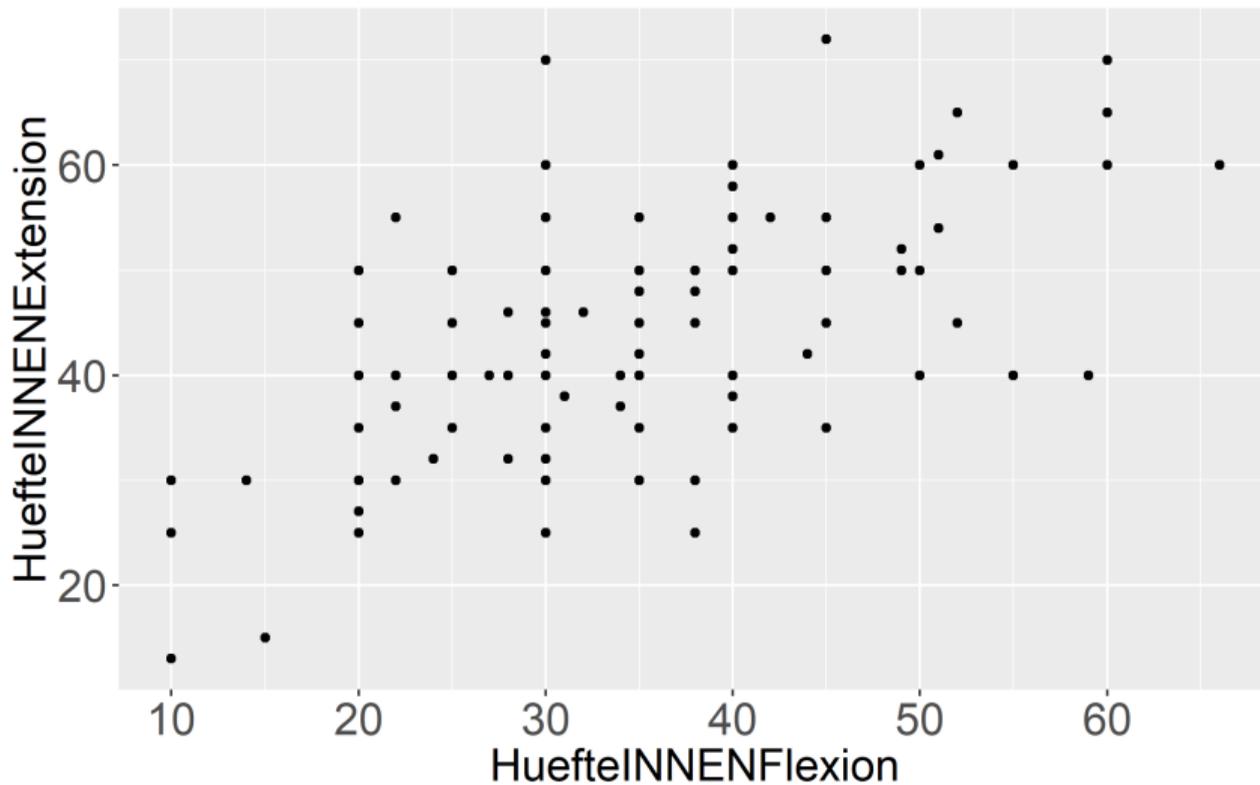
Multiple lineare
Regression

Praxis mit R

Zusammenfassung



Streudiagramm



Intro

Korrelation

Pearson Korrelationskoeffizient

Spearman

Korrelationskoeffizient

Einfache lineare
Regression

Regressionsprüfung

Multiple lineare
Regression

Praxis mit R

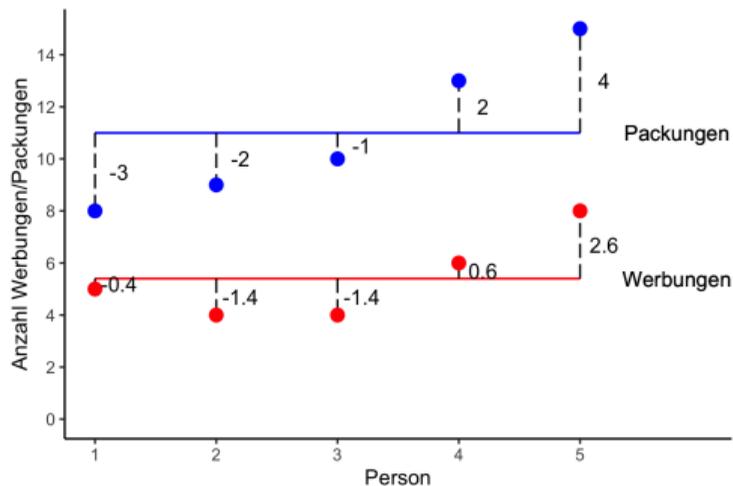
Zusammenfassung

Wie Zusammenhänge messen?

Es wird geschaut ob die zwei Variablen kovariieren. Dafür braucht es die Kovarianz.

Person	1	2	3	4	5	\bar{x}	s
Werbung konsumiert	5	4	4	6	8	5.4	1.67
Packung M&Ms gekauft	8	9	10	13	15	11.0	2.92

Beispiel nach A. Field, Miles und Z. Field [2012](#), p. 207



$$\text{Variance}(s^2) = \frac{\sum (x_i - \bar{x})^2}{N - 1} = \frac{\sum (x_i - \bar{x})(x_i - \bar{x})}{N - 1}$$

$$\text{COV}_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{N - 1}$$

Intro

Korrelation

Pearson Korrelationskoeffizient

Spearman
Korrelationskoeffizient

Einfache lineare
Regression

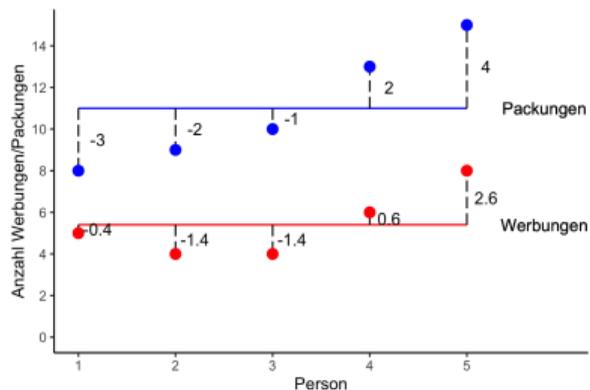
Regressionsprüfung

Multiple lineare
Regression

Praxis mit R

Zusammenfassung

Wie Zusammenhänge messen ? (cont.)



$$\text{COV}_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{N - 1}$$

- ▶ $\text{cov}_{xy}(+)$: Abweichung von x und y vom Mittelwert in gleiche Richtung.
- ▶ $\text{cov}_{xy}(-)$: Abweichung von x und y vom Mittelwert in entgegengesetzte Richtung.

$$\text{cov}_{xy} = \frac{(-0.4)(-3) + (-1.4)(-2) + (-1.4)(-1) + (0.6)(2) + (2.6)(4)}{4} = \frac{1.2 + 2.8 + 1.4 + 1.2 + 10.4}{4} = 4.25$$

- ▶ cov_{xy} hängt von der Messkala ab > Konvertierung in eine Standardeinheit für Vergleichbarkeit.
- ▶ Division durch Produkt der Standardabweichungen.
- ▶ Standardisierte Kovarianz = Korrelationskoeffizient (Pearson)

$$r = \frac{\text{COV}_{xy}}{s_x s_y}$$

Intro

Korrelation

Pearson Korrelationskoeffizient
Spearman
Korrelationskoeffizient

Einfache lineare
Regression

Regressionsprüfung

Multiple lineare
Regression

Praxis mit R

Zusammenfassung

Korrelation

Pearson Korrelationskoeffizient

Intro

Korrelation

Pearson Korrelationskoeffizient

Spearman

Korrelationskoeffizient

Einfache lineare
Regression

Regressionsprüfung

Multiple lineare
Regression

Praxis mit R

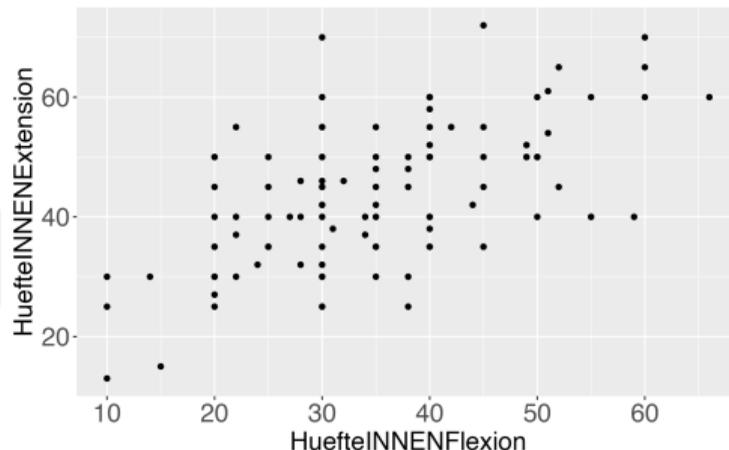
Zusammenfassung

Korrelationskoeffizient nach Pearson

- ▶ Der **lineare Zusammenhang** zwischen zwei Variablen wird mit dem **Korrelationskoeffizient** nach Pearson gemessen.
- ▶ Der Korrelationskoeffizient nach Pearson wird mit r bezeichnet.

Beispiel

In unserem Beispiel ist $r = 0.58$



Intro

Korrelation

Pearson Korrelationskoeffizient

Spearman

Korrelationskoeffizient

Einfache lineare
Regression

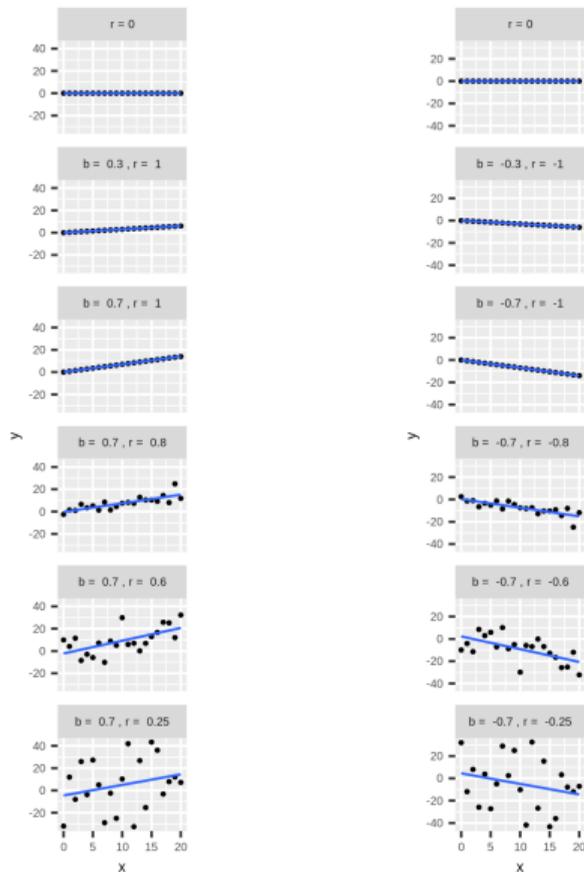
Regressionsprüfung

Multiple lineare
Regression

Praxis mit R

Zusammenfassung

Interpretation



- ▶ Der Korrelationskoeffizient r ist ein Wert zwischen -1 und 1 .
- ▶ Wenn **kein Zusammenhang** besteht zwischen den beiden Variablen, so ist $r = 0$.
- ▶ Je stärker der Zusammenhang zwischen den beiden Variablen ist, desto näher bei 1 oder -1 liegt r .
- ▶ Bei $r = 1$ liegen alle Punkte auf einer Geraden mit einer positiven Steigung » **Positiver vollständig linearer Zusammenhang**.
- ▶ Bei $r = -1$ liegen alle Punkte auf einer Geraden mit einer negativen Steigung » **Negativer vollständig linearer Zusammenhang**.

Intro

Korrelation

Pearson Korrelationskoeffizient

Spearman

Korrelationskoeffizient

Einfache lineare
Regression

Regressionsprüfung

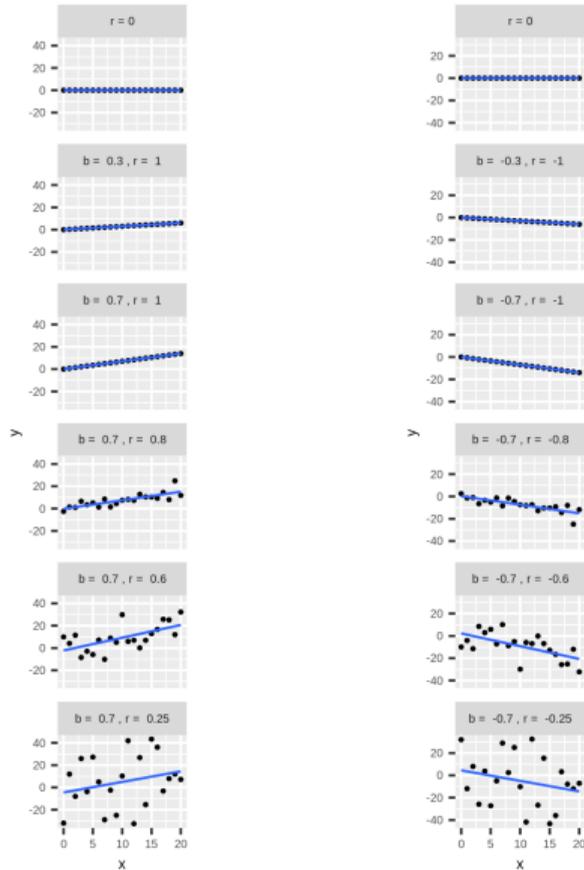
Multiple lineare
Regression

Praxis mit R

Zusammenfassung



Interpretation (cont.)



Man spricht von einem

- ▶ starken linearen Zusammenhang wenn $|r| \geq 0.7$.
- ▶ mittleren linearen Zusammenhang wenn $0.3 \leq |r| < 0.7$.
- ▶ schwachen linearen Zusammenhang wenn $|r| < 0.3$.

learning statistics with jamovi: a tutorial for psychology students and other beginners, Freies E-Book, <https://www.learnstatswithjamovi.com/> (Navarro und Foxcroft 2018)

Intro

Korrelation

Pearson Korrelationskoeffizient

Spearman

Korrelationskoeffizient

Einfache lineare
Regression

Regressionsprüfung

Multiple lineare
Regression

Praxis mit R

Zusammenfassung

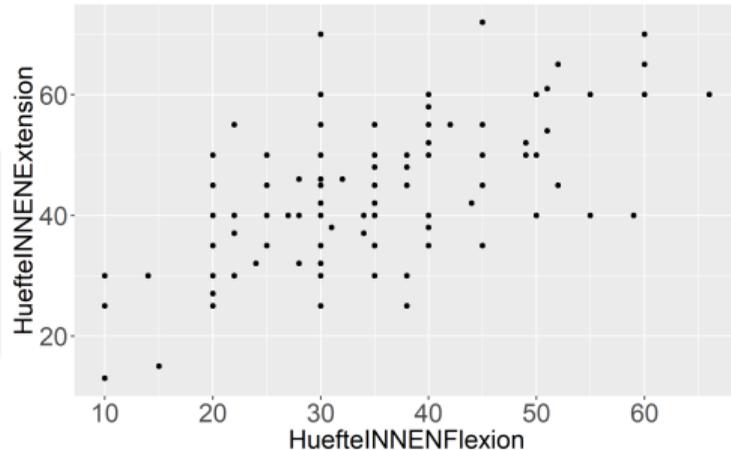
Interpretation (cont.)

Man spricht von einem

- ▶ starken linearen Zusammenhang wenn $|r| \geq 0.7$.
- ▶ mittleren linearen Zusammenhang wenn $0.3 \leq |r| < 0.7$.
- ▶ schwachen linearen Zusammenhang wenn $|r| < 0.3$.

Beispiel

In unserem Beispiel mit $r = 0.58$ liegt also ein mittlerer positiver linearer Zusammenhang vor.

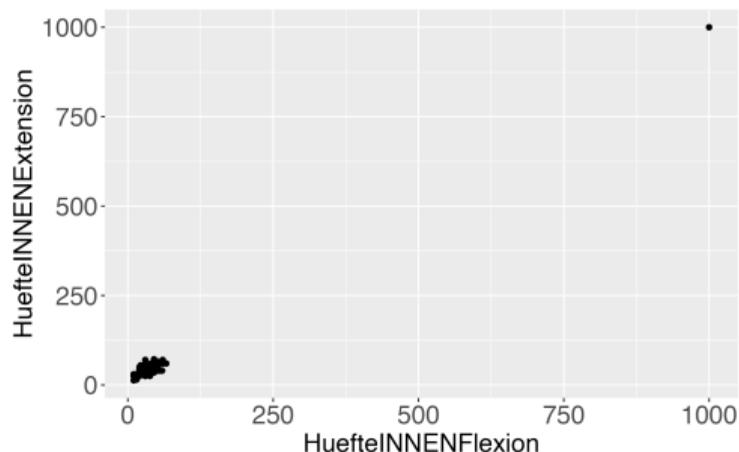


learning statistics with jamovi: a tutorial for psychology students and other beginners, Freies E-Book,
<https://www.learnstatswithjamovi.com/> (Navarro und Foxcroft 2018)

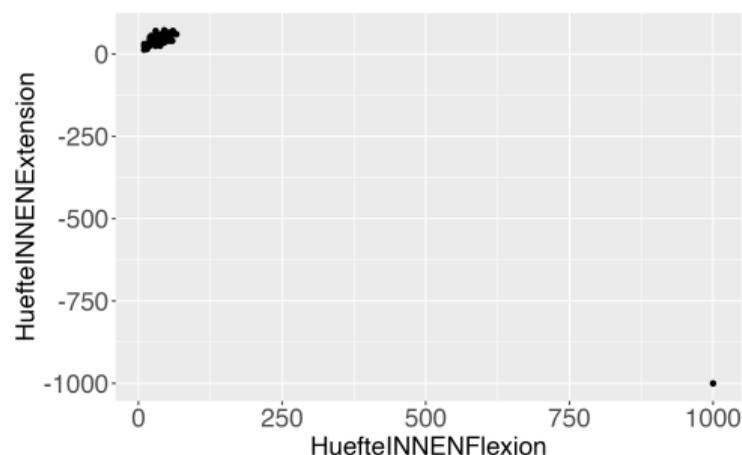
Robustheit

Der Korrelationskoeffizient nach Pearson ist nicht robust gegenüber Ausreißern!

Hinzufügen eines Punktes (1000,1000) führt zu $r = 0.993$.



Hinzufügen eines Punktes (1000,-1000) führt zu $r = -0.975$.



Intro

Korrelation

Pearson Korrelationskoeffizient

Spearman

Korrelationskoeffizient

Einfache lineare
Regression

Regressionsprüfung

Multiple lineare
Regression

Praxis mit R

Zusammenfassung

Korrelation

Spearman Korrelationskoeffizient

Intro

Korrelation

Pearson Korrelationskoeffizient

Spearman

Korrelationskoeffizient

Einfache lineare
Regression

Regressionsprüfung

Multiple lineare
Regression

Praxis mit R

Zusammenfassung

Rangkorrelationskoeffizient nach Spearman

- ▶ Ein **robusterer Korrelationskoeffizient** ist der Rangkorrelationskoeffizient nach **Spearman**.
- ▶ Dieser wird mit r_S bezeichnet und misst...
 - ▶ ...den **linearen Zusammenhang** zwischen den **Rängen von zwei Variablen**.
 - ▶ ...den **monotonen Zusammenhang** zwischen **zwei Variablen**.
- ▶ Den Rangkorrelationskoeffizient nach Spearman erhält man, indem man beide Variablen für sich rangiert und den Korrelationskoeffizienten nach Pearson auf die rangierten Werte anwendet.

Beispiel

In unserem Beispiel ist $r_S = 0.555$ liegt also ein mittlerer positiver monotoner Zusammenhang vor.

Intro

Korrelation

Pearson Korrelationskoeffizient

Spearman
Korrelationskoeffizient

Einfache lineare
Regression

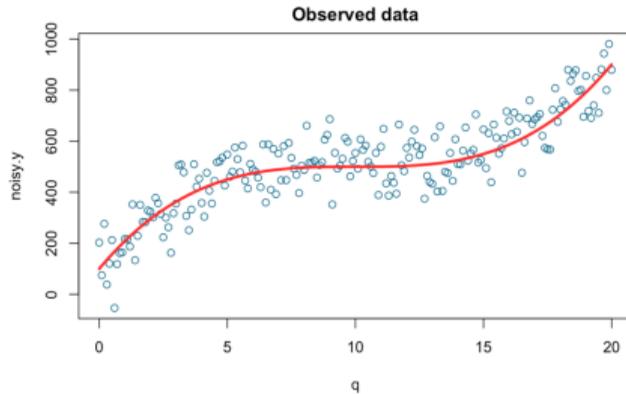
Regressionsprüfung

Multiple lineare
Regression

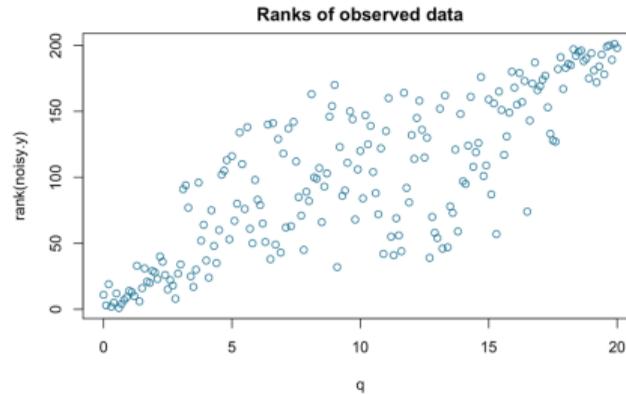
Praxis mit R

Zusammenfassung

Rangkorrelationskoeffizient nach Spearman (cont.)



- ▶ Nicht linearer Zusammenhang
- ▶ Monoton steigend - Variablen bewegen sich in die gleiche Richtung
- ▶ Linearität wird mit r quantifiziert



- ▶ Zusammenhang der Ränge wird linearer
- ▶ Linearität wird mit r_S quantifiziert

Intro

Korrelation

Pearson Korrelationskoeffizient
Spearman
Korrelationskoeffizient

Einfache lineare
Regression

Regressionsprüfung

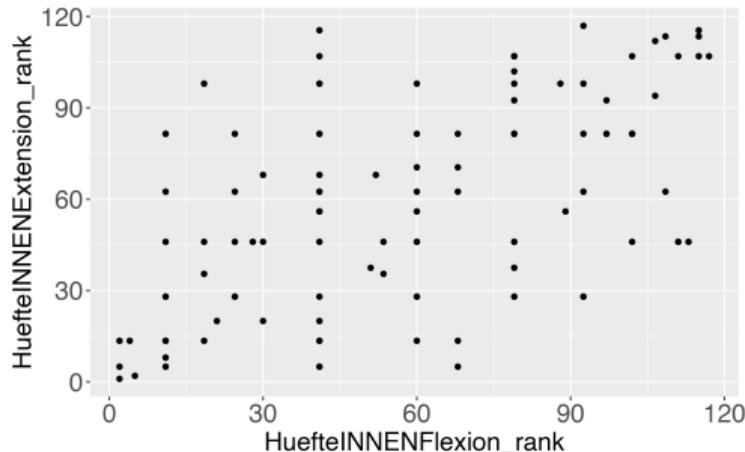
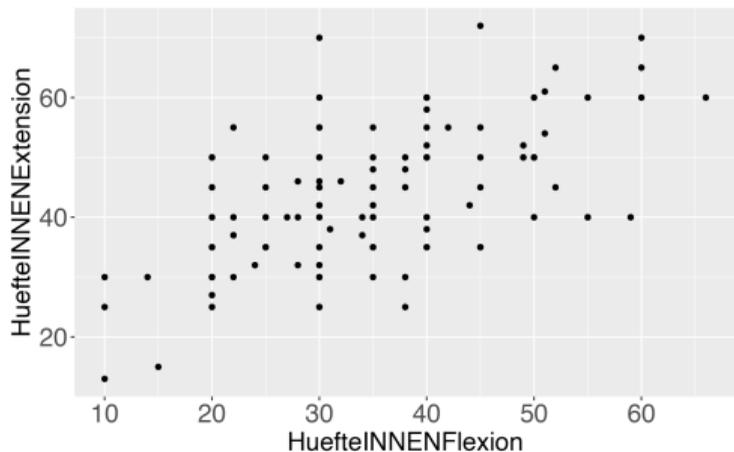
Multiple lineare
Regression

Praxis mit R

Zusammenfassung

Rangierung

HuefteINNENFlexion	HuefteINNENExtension	HuefteINNENFlexion_rank	HuefteINNENExtension_rank
55	40	111	46.0
32	46	52	68.0
30	50	41	81.5
40	50	79	81.5
60	60	115	107.0



Intro

Korrelation

Pearson Korrelationskoeffizient

Spearman

Korrelationskoeffizient

Einfache lineare
Regression

Regressionsprüfung

Multiple lineare
Regression

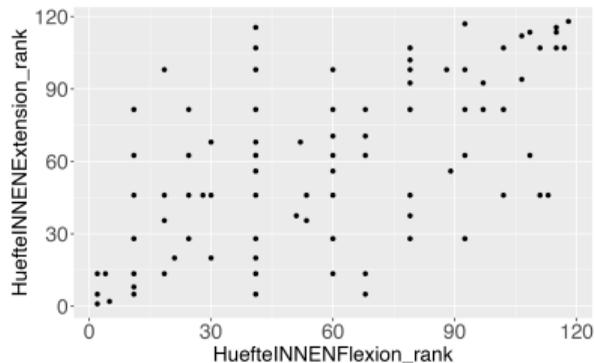
Praxis mit R

Zusammenfassung

Robustheit

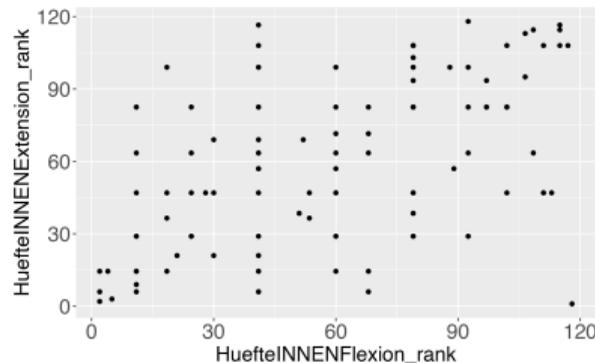
Der Korrelationskoeffizient nach Spearman ist robust gegenüber Ausreißern!

Hinzufügen eines Punktes (1000,1000) führt zu $r_S = 0.567$.



HuefteINNENFlexion	HuefteINNENExtension	HuefteINNENFlexion_rank	HuefteINNENExtension_rank
32	46	52	68.0
30	50	41	81.5
40	50	79	81.5
60	60	115	107.0
1000	1000	118	118.0

Hinzufügen eines Punktes (1000,-1000) führt zu $r_S = 0.516$.



HuefteINNENFlexion	HuefteINNENExtension	HuefteINNENFlexion_rank	HuefteINNENExtension_rank
32	46	52	69.0
30	50	41	82.5
40	50	79	82.5
60	60	115	108.0
1000	-1000	118	1.0

Intro

Korrelation

Pearson Korrelationskoeffizient
Spearman
Korrelationskoeffizient

Einfache lineare
Regression

Regressionsprüfung

Multiple lineare
Regression

Praxis mit R

Zusammenfassung

Man spricht von einem

- ▶ starken **monotonen** Zusammenhang wenn $|r_S| \geq 0.7$.
- ▶ mittleren **monotonen** Zusammenhang wenn $0.3 \leq |r_S| < 0.7$.
- ▶ schwachen **monotonen** Zusammenhang wenn $|r_S| < 0.3$.

learning statistics with jamovi: a tutorial for psychology students and other beginners, Freies E-Book,
<https://www.learnstatswithjamovi.com/> (Navarro und Foxcroft 2018)

Einfache lineare Regression

Intro

Korrelation

Pearson Korrelationskoeffizient

Spearman

Korrelationskoeffizient

Einfache lineare
Regression

Regressionsprüfung

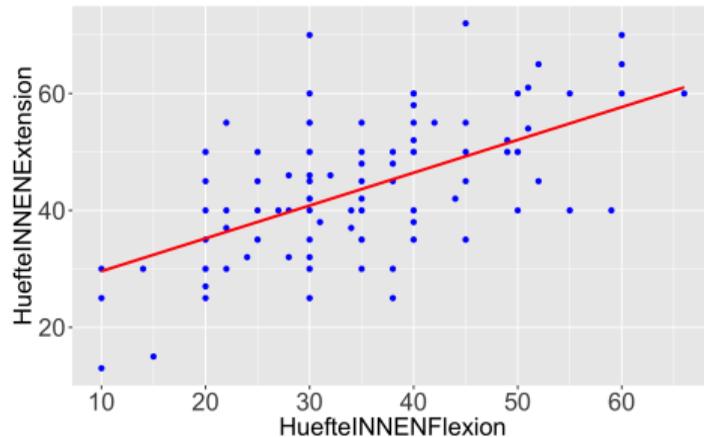
Multiple lineare
Regression

Praxis mit R

Zusammenfassung

Regressionsgerade

- ▶ Wenn man einen linearen Zusammenhang zwischen zwei Zufallsvariablen entdeckt hat, möchte man diesen Zusammenhang natürlich auch beschreiben können.
- ▶ Man sucht die Gerade, die „am besten zu den Daten passt“: **Regressionsgerade**
- ▶ Das Berechnungsverfahren heisst **lineare Regression**.



Intro

Korrelation

Pearson Korrelationskoeffizient

Spearman

Korrelationskoeffizient

Einfache lineare
Regression

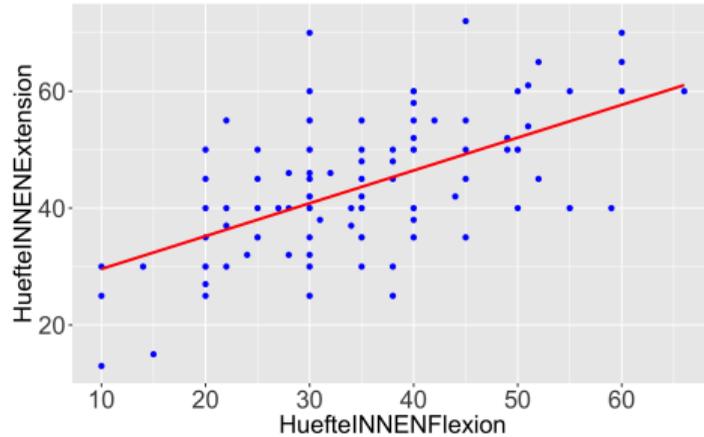
Regressionsprüfung

Multiple lineare
Regression

Praxis mit R

Zusammenfassung

Bezeichnung der Variablen



- ▶ Wir sind interessiert an Hüfte INNEN passiv in Extension in Abhängigkeit von Hüfte INNEN passiv in Flexion.
- ▶ Hüfte INNEN passiv in Extension ist darum die **abhängige** oder erklärte Variable.
- ▶ Hüfte INNEN passiv in Flexion ist die **unabhängige** oder erklärende Variable. Im Englischen oft als predictor bezeichnet.

Intro

Korrelation

Pearson Korrelationskoeffizient

Spearman

Korrelationskoeffizient

Einfache lineare
Regression

Regressionsprüfung

Multiple lineare
Regression

Praxis mit R

Zusammenfassung

Regressionsgleichung

Für die Regressionsgerade gilt die Gleichung

$$y = a + b \cdot x$$

- ▶ a ist der Achsenabschnitt (Wert von y wenn $x = 0$)
- ▶ b ist die Steigung (Zunahme der y -Werte wenn x um 1 vergrößert wird)

Beispiel

Model Coefficients

Predictor	Estimate	SE	95% Confidence Interval		t	p
			Lower	Upper		
Intercept	23.980	2.7546	18.523	29.436	8.71	< .001
HuefteINNENFlexion	0.562	0.0739	0.415	0.708	7.60	< .001

In unserem Beispiel erhalten wir die Gleichung

$$\text{HuefteINNENExtension} = 24 + 0.56 \cdot \text{HuefteINNENFlexion}$$

Intro

Korrelation

Pearson Korrelationskoeffizient
Spearman
Korrelationskoeffizient

Einfache lineare
Regression

Regressionsprüfung

Multiple lineare
Regression

Praxis mit R

Zusammenfassung

Vorhersage

- ▶ Die Gleichung erlaubt nun eine Vorhersage.
- ▶ Achtung: Vorhersagen dürfen nicht ausserhalb des Bereiches, in dem man Daten hat, gemacht werden

Beispiel

Ist zum Beispiel die Rotation in Flexion $x = 30$, so ist die Rotation in Extension $y = 24 + 0.56 \cdot 30 = 40.8^\circ$.

Intro

Korrelation

Pearson Korrelationskoeffizient
Spearman
Korrelationskoeffizient

Einfache lineare
Regression

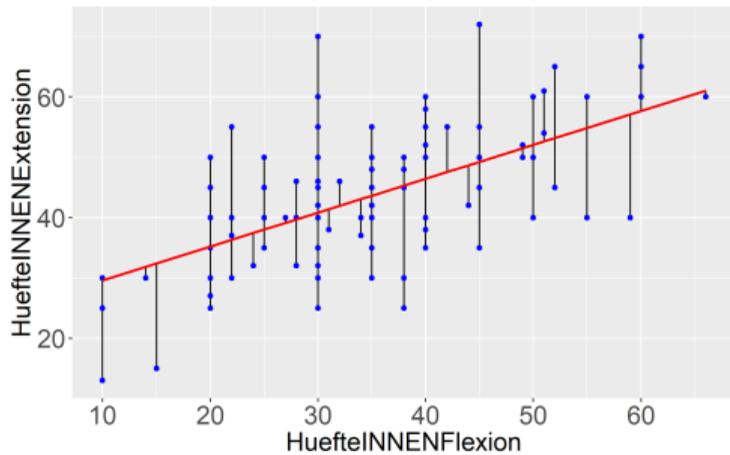
Regressionsprüfung

Multiple lineare
Regression

Praxis mit R

Zusammenfassung

Modelldiagnostik via Residuen



Die Residuen sind die Abweichungen der y -Werte von der Regressionsgeraden.

Bei einer qualitativ guten linearen Regression

- ▶ sind die Residuen normalverteilt => QQ-Plot
- ▶ haben die Residuen eine konstante Variabilität => Streudiagramm

Intro

Korrelation

Pearson Korrelationskoeffizient
Spearman
Korrelationskoeffizient

Einfache lineare
Regression

Regressionsprüfung

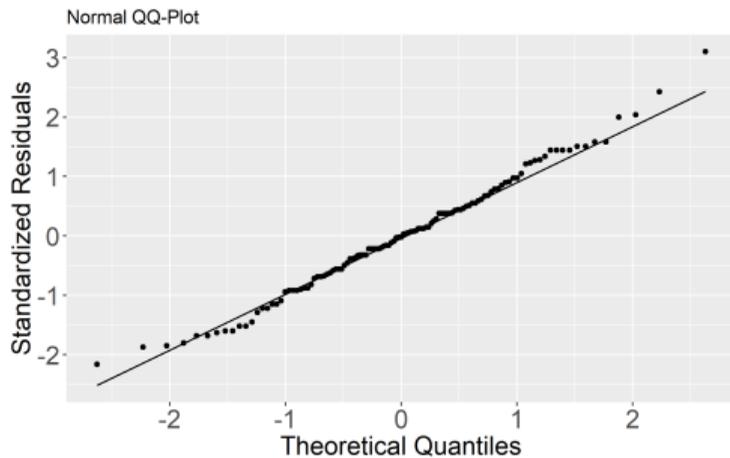
Multiple lineare
Regression

Praxis mit R

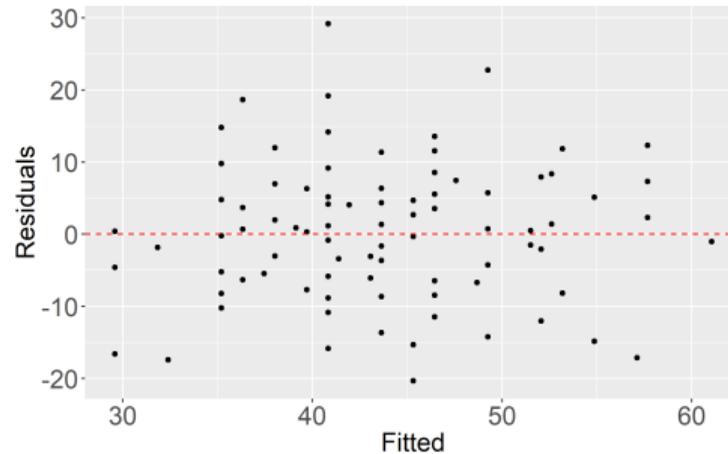
Zusammenfassung

Modelldiagnostik via Residuen für Beispiel

Es sind keine auffälligen Abweichungen von einer Geraden sichtbar. Die Normalverteilungsannahme scheint erfüllt zu sein.



Es gibt keine auffällig grossen Residuen und kein Muster. Die Residuen scheinen also eine konstante, nicht von x abhängige Variabilität aufzuweisen.



Intro

Korrelation

Pearson Korrelationskoeffizient

Spearman

Korrelationskoeffizient

Einfache lineare
Regression

Regressionsprüfung

Multiple lineare
Regression

Praxis mit R

Zusammenfassung

Bestimmtheitsmass

- ▶ Das Bestimmtheitsmass R^2 wird benutzt um die Stärke des Zusammenhangs zwischen der abhängigen und der/den erklärenden Variablen zu beschreiben.
- ▶ In der einfachen linearen Regression entspricht das Bestimmtheitsmass dem Korrelationskoeffizienten nach Pearson im Quadrat.
- ▶ R^2 sagt wie viel % der Variation in y durch x erklärt werden können.
- ▶ Ist $R^2 > 0.5$ spricht man von einem starken Zusammenhang.

Beispiel

Model Fit Measures				
Model	R	R^2	Adjusted R^2	RMSE
1	0.578	0.334	0.329	9.37

In unserem Beispiel ist $R^2 = 0.33$, also ein nicht allzu starker Zusammenhang.

Intro

Korrelation

Pearson Korrelationskoeffizient
Spearman
Korrelationskoeffizient

Einfache lineare
Regression

Regressionsprüfung

Multiple lineare
Regression

Praxis mit R

Zusammenfassung

Regressionsprüfung

Intro

Korrelation

Pearson Korrelationskoeffizient
Spearman
Korrelationskoeffizient

Einfache lineare
Regression

Regressionsprüfung

Multiple lineare
Regression

Praxis mit R

Zusammenfassung

Prüfung der Regressionskoeffizienten

- ▶ Um zu überprüfen, ob ein Zusammenhang zwischen der abhängigen und der erklärenden Variablen besteht, wird ein t -Test durchgeführt mit der Nullhypothese $H_0 : b = 0$.
- ▶ Man kann auch ein Vertrauensintervall für die Steigung bestimmen.
- ▶ Meistens wird ebenfalls mit einem t -Test überprüft, ob der Achsenabschnitt $a = 0$ ist.

Beispiel

Model Coefficients

Predictor	Estimate	SE	95% Confidence Interval		t	p
			Lower	Upper		
Intercept	23.980	2.7546	18.523	29.436	8.71	< .001
HuefteINNENFlexion	0.562	0.0739	0.415	0.708	7.60	< .001

$p < 0.001$ für b (HuefteINNENFlexion) \Rightarrow Steigung signifikant von 0 verschieden, also besteht ein Zusammenhang.

Intro

Korrelation

Pearson Korrelationskoeffizient

Spearman

Korrelationskoeffizient

Einfache lineare
Regression

Regressionsprüfung

Multiple lineare
Regression

Praxis mit R

Zusammenfassung

Analogie t -Test

Betrachten wir das Beispiel von den Signifikanztests: Vergleich von HüfteINNENFlexion zwischen Männern und Frauen.

Independent Samples T-Test

		statistic	df	p	Mean difference	SE difference
HuefteINNENFlexion	Student's t	-3.52	89.0	0.00068	-9.86	2.80

Model Coefficients

Predictor	Estimate	SE	t	p
Intercept	32.50	1.40	23.22	< .00001
Geschlecht:				
f - m	9.86	2.80	3.52	0.00068

t -Test und lineare Regression mit Geschlecht als erklärender Variable kommen zum gleichen Ergebnis!

Intro

Korrelation

Pearson Korrelationskoeffizient
Spearman
Korrelationskoeffizient

Einfache lineare
Regression

Regressionsprüfung

Multiple lineare
Regression

Praxis mit R

Zusammenfassung

Multiple lineare Regression

Intro

Korrelation

Pearson Korrelationskoeffizient

Spearman

Korrelationskoeffizient

Einfache lineare
Regression

Regressionsprüfung

Multiple lineare
Regression

Praxis mit R

Zusammenfassung

Verschiedene Arten der Regression

- ▶ Einfache lineare Regression: Beschreibt den linearen Zusammenhang zwischen zwei Variablen.
- ▶ Multiple lineare Regression: Beschreibt den linearen Zusammenhang zwischen einer abhängigen und mehreren erklärenden Variablen.
- ▶ Verallgemeinerte lineare Regression: Ist die Zielgrösse nicht stetig, müssen andere Methoden angewendet werden wie z.B. die logistische Regression.

Intro

Korrelation

Pearson Korrelationskoeffizient
Spearman
Korrelationskoeffizient

Einfache lineare
Regression

Regressionsprüfung

Multiple lineare
Regression

Praxis mit R

Zusammenfassung

Multiple lineare Regression: Beispiel

- ▶ Vorheriges Beispiel: HüfteINNENExtension in Abhängigkeit von HüfteINNENFlexion
- ▶ Nun können wir z.B. das Geschlecht als weitere unabhängige Variable ins Modell nehmen.

Regressionsgleichung bei zwei Variablen:

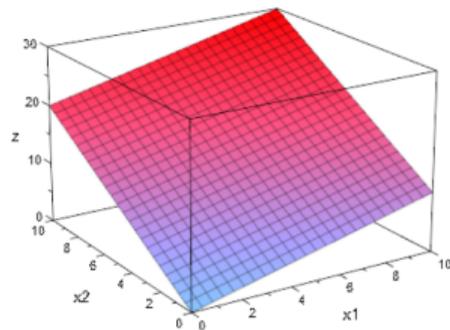
$$y = a + b_1x_1 + b_2x_2$$

In unserem Fall:

$$\text{HüfteINNENExtension} = a + b_1\text{HüfteINNENFlexion} + b_2\text{Geschlecht}$$

Regressionsgleichung bei n Variablen:

$$y = a + b_1x_1 + b_2x_2 + \dots + b_nx_n$$



Intro

Korrelation

Pearson Korrelationskoeffizient
Spearman
Korrelationskoeffizient

Einfache lineare
Regression

Regressionsprüfung

Multiple lineare
Regression

Praxis mit R

Zusammenfassung

Multiple lineare Regression: Beispiel (cont.)

Model Coefficients

Predictor	Estimate	SE	t	p
Intercept	23.534	2.9956	7.86	< .001
HuefteINNENFlexion	0.541	0.0854	6.33	< .001
Geschlecht: f - m	6.199	2.4073	2.58	0.012

Das Modell bestätigt unsere Erkenntnisse von vorher, dass Frauen eine höhere Hüftrotation haben als Männer (Unterschied 6.2 Grad, p-Wert 0.01).

Vorhersage männliche Person, HüfteINNENFlexion = 30 Grad

$$\text{HüfteINNENExtension} = 23.53 + 0.54 \cdot 30 + 6.2 \cdot 0 = 39.7^\circ$$

Vorhersage weibliche Person, HüfteINNENFlexion = 30 Grad

$$\text{HüfteINNENExtension} = 23.53 + 0.54 \cdot 30 + 6.2 \cdot 1 = 45.9^\circ$$

Intro

Korrelation

Pearson Korrelationskoeffizient

Spearman

Korrelationskoeffizient

Einfache lineare
Regression

Regressionsprüfung

Multiple lineare
Regression

Praxis mit R

Zusammenfassung

Bestimmtheitsmass

- ▶ In der multiplen linearen Regression sollte man das adjustierte Bestimmtheitsmass betrachten.
- ▶ Dieses ist nicht von der Anzahl erklärender Variablen abhängig.

Beispiel

Model Fit Measures

Model	R	R ²	Adjusted R ²	RMSE
1	0.655	0.430	0.417	9.19

In unserem Beispiel ist $R_{adj}^2 = 0.42$. Dieses Modell ist also besser als das ohne Geschlecht.

Intro

Korrelation

Pearson Korrelationskoeffizient

Spearman

Korrelationskoeffizient

Einfache lineare
Regression

Regressionsprüfung

Multiple lineare
Regression

Praxis mit R

Zusammenfassung

Praxis mit R

Intro

Korrelation

Pearson Korrelationskoeffizient

Spearman

Korrelationskoeffizient

Einfache lineare
Regression

Regressionsprüfung

Multiple lineare
Regression

Praxis mit R

Zusammenfassung

Streudiagramm in Jamovi

Zusatzmodul scatr installieren

The screenshot displays the Jamovi software interface. At the top, there is a menu bar with 'Data' and 'Analyses' tabs. Below the menu bar is a toolbar with various analysis icons. A red circle highlights a '+' icon labeled 'Modules' in the top right corner. The main window shows a list of installed and available modules. The 'Available' tab is selected, and the 'scatr' module (version 1.1.1) is highlighted with a red box. The 'scatr' module description reads: 'Allows you to produce several types of good-looking explorative plots such as scatter plots and pareto charts. You can find it under the 'Exploration' menu.' Below the 'scatr' module, there are other modules listed: 'MAJOR - Meta Analysis 1.0.0', 'TOSTER 0.3.3', and 'Base R 1.1.2'. In the background, a data table is visible with columns 'code' and 'Geschl...', and a 'Correlation Matrix' plot is shown on the right side of the interface.

Intro

Korrelation

Pearson Korrelationskoeffizient
Spearman
Korrelationskoeffizient

Einfache lineare
Regression

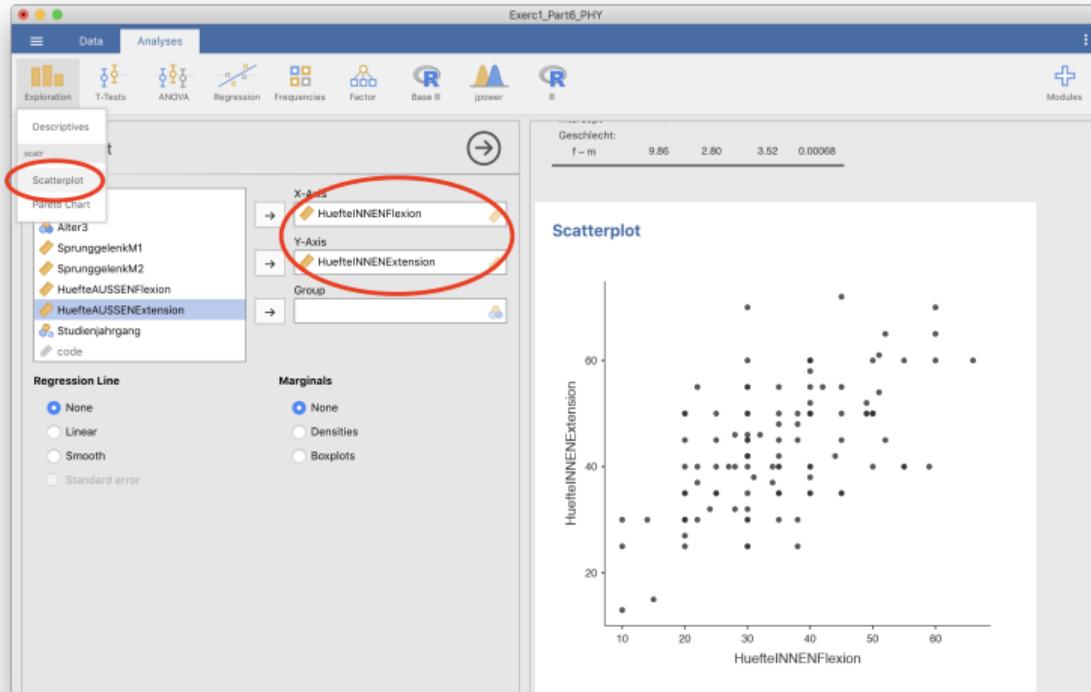
Regressionsprüfung

Multiple lineare
Regression

Praxis mit R

Zusammenfassung

Streudiagramm in Jamovi (cont.)



Intro

Korrelation

Pearson Korrelationskoeffizient
Spearman
Korrelationskoeffizient

Einfache lineare
Regression

Regressionsprüfung

Multiple lineare
Regression

Praxis mit R

Zusammenfassung

Korrelationskoeffizienten in Jamovi

The screenshot shows the Jamovi software interface with the 'Correlation Matrix' analysis selected. The 'Correlation Matrix' menu is open, and the 'Pearson' and 'Spearman' options are selected. The results pane shows a table with Pearson's r and Spearman's rho values for HuefteINNENFlexion and HuefteINNENExtension.

Correlation Matrix

		HuefteINNENFlexion	HuefteINNENExtension
HuefteINNENFlexion	Pearson's r	—	0.578
	p-value	—	<.00001
	Spearman's rho	—	0.555
	p-value	—	<.00001
HuefteINNENExtension	Pearson's r	—	—
	p-value	—	—
	Spearman's rho	—	—
	p-value	—	—

Linear Regression

Model Fit Measures

Model	R	R ²	Adjusted R ²	RMSE
1	0.578	0.334	0.329	9.37

Model Coefficients

Predictor	Estimate	SE	95% Confidence Interval		t	p
			Lower	Upper		
Intercept	23.980	2.7546	18.523	29.436	8.71	<.00001

Intro

Korrelation

Pearson Korrelationskoeffizient
Spearman
Korrelationskoeffizient

Einfache lineare
Regression

Regressionsprüfung

Multiple lineare
Regression

Praxis mit R

Zusammenfassung

Lineare Regression

The screenshot shows the SPSS Linear Regression dialog box and its results. The dependent variable is 'HueftaHNENExtension' and the covariate is 'HueftaHNENFlexion'. The results include Model Fit Measures, Model Coefficients, Assumption Checks, and a Q-Q Plot.

Linear Regression

Dependent Variable: HueftaHNENExtension

Covariates: HueftaHNENFlexion

Model Fit Measures

Model	R	R ²	Adjusted R ²	RMSE
1	0.578	0.334	0.329	9.37

Model Coefficients

Predictor	Estimate	SE	95% Confidence Interval		t	p
			Lower	Upper		
Intercept	23.960	2.7546	18.523	29.436	8.71	<.00001
HueftaHNENFlexion	0.562	0.0739	0.415	0.708	7.60	<.00001

Assumption Checks

Q-Q Plot

Residuals Plots

Intro

Korrelation

Pearson Korrelationskoeffizient
Spearman
Korrelationskoeffizient

Einfache lineare
Regression

Regressionsprüfung

Multiple lineare
Regression

Praxis mit R

Zusammenfassung

Multiple linear Regression

The screenshot shows the SPSS Linear Regression dialog box with the following settings:

- Dependent Variable:** HuefteINNENExtension
- Covariates:** HuefteINNENFlexion
- Factors:** Geschlecht (highlighted with a red circle)

Below the dialog box, the results are displayed:

Linear Regression

Model Fit Measures

Model	R	R ²	Adjusted R ²	RMSE
1	0.655	0.430	0.417	9.19

Model Coefficients

Predictor	Estimate	SE	t	p
Intercept	23.534	2.9956	7.86	<.00001
HuefteINNENFlexion	0.541	0.0854	6.33	<.00001
Geschlecht:				
f - m	6.199	2.4073	2.58	0.01169

Independent Samples T-Test

Independent Samples T-Test

	statistic	df	p	Mean difference	SE difference	
HuefteINNENFlexion	Student's t	-3.52	89.0	0.00068	-9.86	2.80

Linear Regression

Intro

Korrelation

Pearson Korrelationskoeffizient
Spearman
Korrelationskoeffizient

Einfache lineare
Regression

Regressionsprüfung

Multiple lineare
Regression

Praxis mit R

Zusammenfassung

Zusammenfassung

Intro

Korrelation

Pearson Korrelationskoeffizient

Spearman

Korrelationskoeffizient

Einfache lineare
Regression

Regressionsprüfung

Multiple lineare
Regression

Praxis mit R

Zusammenfassung

- ▶ Das einfachste Mittel den Zusammenhang zwischen zwei kontinuierlichen Variablen zu beurteilen sind Korrelationskoeffizienten.
- ▶ Der Korrelationskoeffizient kann nach Pearson oder Spearman (Rangkorrelation) berechnet werden, wobei die Methode nach Pearson stark auf Ausreisser reagiert (nicht robust ist).
- ▶ Interpretation Korrelationskoeffizienten
 - ▶ Pearson: linearer Zusammenhang der Variablen; wenn $r = 1$ dann exakt linear
 - ▶ Spearman: linearer Zusammenhang der Ränge; monotoner Zusammenhang der Variablen

Zusammenfassung (cont.)

- ▶ Die lineare Regression ist ein weiteres Mittel um den linearen Zusammenhang zwischen einer abhängigen und einer oder mehrerer unabhängigen Variablen zu beschreiben.
- ▶ Dabei wird im Fall von zwei Variablen eine Gerade an die Daten „eingepasst“ => Regressionsgleichung. Achtung: Nur sinnvoll wenn wirklich ein linearer Zusammenhang besteht.
- ▶ Mit dem Modell ist es möglich Vorhersagen zu machen, wobei gewisse Randbedingungen zu beachten sind:
 - ▶ Keine Vorhersagen ausserhalb des Wertebereichs.
 - ▶ Residuen müssen normalverteilt sein und eine konstante Streuung haben.
 - ▶ Das Bestimmtheitsmass R^2 gibt über die Güte des Modells Aufschluss.
- ▶ Der Einfluss der Regressionsparameter kann statistisch getestet werden. Man kommt zum gleichen Schluss wie mit einem t -Test wenn das lineare Modell entsprechend gebaut ist.

Intro

Korrelation

Pearson Korrelationskoeffizient

Spearman

Korrelationskoeffizient

Einfache lineare
Regression

Regressionsprüfung

Multiple lineare
Regression

Praxis mit R

Zusammenfassung