Hausarbeit im Rahmen des Studiums

Data Science - Master of Science (M.Sc.)

Lineare Regressionsanalyse

am Beispiel von „A Study of Scarlet“

Alwine Schultze

11. März 2022

Inhaltsverzeichnis

Inhaltsverzeichnis I

Abbildungsverzeichnis II

Tabellenverzeichnis II

1 Einleitung 1

1.1 Fragestellung 1

1.2 Methodik und Zielhierarchie 1

2 Grundlagen 2

2.1 Regressionsanalyse 2

2.2 Terminologie 2

2.3 Mathematische Kennzahlen 3

3 Problembearbeitung 4

3.1 Datenerhebung 4

3.2 Berechnung statistischer Kenngrößen 5

3.2.1 Empirischer Mittelwert 5

3.2.2 Empirische Standardabweichung 6

3.2.3 Korrigierte empirische Standardabweichung 6

3.2.4 Standardfehler des Mittelwerts bzw. Messunsicherheit 7

3.2.5 Relative Standardunsicherheit 7

3.3 Herleitung der Regressionsformel und Berechnung des Bestimmtheitsmaßes 8

3.3.1 Abweichungsquadratsummen 8

3.3.2 Empirische Kovarianz 8

3.3.3 Steigungskoefzient 8

3.3.4 Restvariable 9

3.3.5 Bestimmtheitsmaß 9

3.3.6 Lineare Korrelationskoeffizient 9

4 Ergebnis und Diskussion 10

5 Zusammenfassung 11

Literaturverzeichnis IV

Abbildungsverzeichnis

1 Streudiagramm der Messergebnisse 4
2 Regressionsgerade und Mittelpunkt 10

Tabellenverzeichnis

1 Messungen der Versuchsreihe 5

1 Einleitung

1.1 Fragestellung

Unter Statistik wird die Gesamtheit von Verfahren und Methoden zur Gewinnung von Informationen und Zusammenhängen aus Daten über reale Sachverhalte verstanden [Eck19]. Sie hält Berechnungsmethoden bereit, die helfen, Antworten auf allgemeingültige Fragestellungen zu finden. Sherlock Holmes tätigte zum Beispiel in „A Study of Scarlet“ von Arthur Canon Doyle die Aussage, dass „in neunzig von hundert Fällen lässt sich von der Schrittlänge eines Menschen auf die Körpergröße schließen“. Diese Aussage wird, unter Zuhilfenahme elementarer statistischer Methoden, in der hier zugrunde liegenden Arbeit analysiert und damit die Frage beantwortet, ob tatsächlich eine Abhängigkeit von Schrittlänge zur Körpergröße eines Menschen besteht.

1.2 Methodik und Zielhierarchie

Zur Klärung dieser Frage wird in der empirischen Wirtschaftsforschung die Regressionsanalyse empfohlen. In dieser wissenschaftlichen Ausarbeitung wird zunächst auf die Grundlagen der gewählten Analysemethode näher eingegangen. Zur besseren Verständlichkeit der verwendeten statistischen Formeln zur Berechnung des Bestimmungsmaßes und der Regressionsgeraden wird zudem näher auf die verwendeten mathematischen Parameter eingegangen. Im Hauptteil wird die eigentliche statistische Berechnung durchgeführt. Anschließend werden die berechneten statistischen Werte kritisch bewertet und eine Antwort auf die Forschungsfrage abgeleitet. Ziel der hier vorliegenden Ausarbeitung ist es ein grundlegendes Verständnis der gewählten Analysevariante zu erhalten und unter Zuhilfenahme der Statistik bestimmen zu können, ob Sherlock Holmes Recht oder Unrecht hatte mit seiner Behauptung, dass von der Schrittlänge auf die Körpergröße geschlossen werden kann.

2 Grundlagen

2.1 Regressionsanalyse

Vorangegangene Forschungen haben bereits zahlreiche Varianten von Regressionsanalysen hervor gebracht, von der klassischen linearen Regression bis hin zur logistischen Regression [DU18]. Die am besten geeignetste Variante der Regressionsanalse ist Aufgrund der Voraussetzung, gegeben durch die Messung zweier metrischer Werte, Schrittlänge und Körpergröße, die lineare Regression. Die Regressionsformel, mathematisch ausgedrückt, lautet:

$$y_{\mathrm{i}} = m \cdot x_{\mathrm{i}} + b \tag{1}$$

Um abschließend bewerten zu können, ob Schrittlänge und Körpergröße stark voneinander abhängen, wie von Sherlock Holmes behauptet, benötigt man noch das mathematische Bestimmtheitsmaß r^2. Es liefert die Erkenntnis ob die Regressionsgerade eine sinnvolle Beschreibung der Daten liefert [BC19] und gibt somit Aufschluss darüber, ob eine Korrelation zwischen der Schrittlänge und der Körpergröße besteht. Mathematisch dargestellt als r^2. Das Bestimmtheitsmaß liegt stets zwischen 0 und 1 und wird wie folgt berechnet:

$$r^2 = \frac{S_{\mathrm{xy}}}{\sqrt{S_{\mathrm{xx}} \cdot S_{\mathrm{yy}}}} \tag{2}$$

2.2 Terminologie

Bevor näher auf die Berechnung der linearen Regression und dem Bestimmtheitsmaß eingegangen wird, werden im kommenden Abschnitt vorab noch einige Begrifflichkeiten erläutert.

- Das **Streudiagramm** eignet sich besonders zur Visualisierung zweier von einander abhängiger Messwerte.
- Die **Regressionsgerade** ist eine Gerade, die durch die Messwerte im Streudiagramm verläuft und durch die Regressionsgleichung bestimmt wird.
- Der **Regressand** ist das abhängige Merkmal einer Regression [Eck19]. Mathematisch dargestellt als $y(x)$.
- Der **Regressor** ist das unabhängige Merkmal einer Regression [BC19]. Mathematisch dargestellt als X.
- Die **Residuen** stellen die empirischen Fehler des geschätzen Regressionsmodells dar [BC19]. Mathematisch dargestellt als $u_{rel}(x)$.

2.3 Mathematische Kennzahlen

Neben den allgemeinen Grundbegriffen sind auch die mathematischen Kennzahlen für das bessere Verständnis der aufkommenden statistischen Berechnung des Bestimmungsmaßes und der Regressionsgleichung wichtig und werden in der kommenden Aufzählung näher erläutert.

- **X** stellt den metrischen Wert für die Schrittlänge in cm dar, siehe Regressor.
- **Y** stellt den metrischen Wert für die Körpergröße in cm dar.
- **m** gibt die Steigung der Regressionsgeraden an.
- **b** gibt den Y Wert an, durch den die Regressionsgerade beim X-Wert = 0 verläuft, siehe Restvariable.
- **i** stellt den Werteiterator dar, dieser liegt zwischen 0 und der Anzahl der Messwerte minus 1.
- **n** stellt die Größe der Stichprobe dar (Anzahl Messungen).

Die Grundlagen, die gewählte statistische Methode und alle relevanten mathematischen Parameter, die wichtig sind um die Antwort auf die Kernfrage dieser Arbeit zu finden, wurden in diesem Kapitel erläutert. In dem kommenden Kapitel wird die eigentliche Problembearbeitung Schritt für Schritt durchgeführt und erörtert.

3 Problembearbeitung

3.1 Datenerhebung

Eine jede statistische Auswertung basiert auf Daten. Bevor sich also die Suche nach der Antwort auf die Frage „Besteht ein Zusammenhang zwischen Schrittlänge und Körpergröße?“ begonnen werden kann, wird zunächst eine Stichprobe, die der Berechnung zugrunde gelegt werden kann, benötigt. Für diesen Zweck wurden insgesamt 20 Probanden herangezogen und die Messung an ihnen durchgeführt. Bei der Erhebung der Daten von Schrittlänge und Körpergröße wurden im Vorfeld Regeln definiert, um sowohl systematische[1] als auch statistische[2] Fehler so gering wie möglich zu halten.

- Zur Reduzierung von **systematische Fehler** soll die Messung...
  - ...mit einem genormten Messinstrument, einem Maßband, erfolgen
  - ...der Körpergröße **ohne** Schuhe und mit einem **geraden Rücken an der Wand** vorgenommen werden
  - ...der Schrittlänge von der Ferse bis zum großen Zeh erfolgen
- Zur Reduzierung von **statistischen Fehlern** sollen von jedem Probanden insgesamt zehn Schritte gemessen werden und daraus die **mittlere Schrittweite** errechnet werden (Summe Schrittweiten/Anzahl Schritte).

Unter Berücksichtigung der definierten Messvorschrift ergeben sich, die in Tabelle 1 dargestellten, Messwerte der Stichprobe. In Abbildung 1 wird das Messergebnis in einem Streudiagramm dargestellt, dabei befinden sich die Werte der mittleren Schrittlänge auf der x-Achse und die der Körpergröße auf der y-Achse.

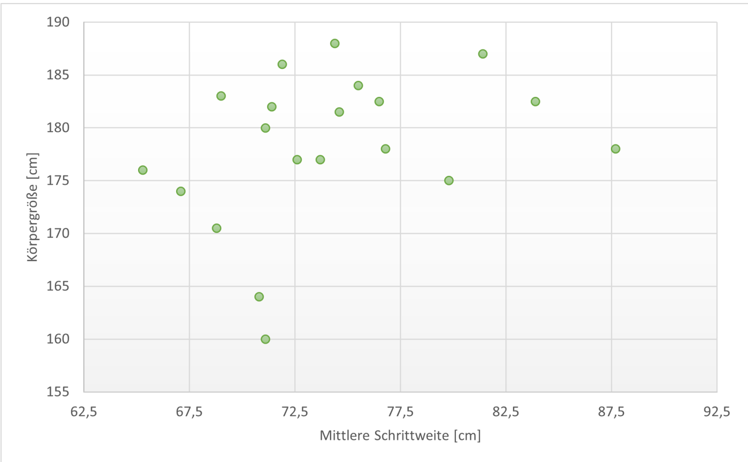

Abbildung 1: Streudiagramm der Messergebnisse

[1]Fehler tritt auf, wenn z.B. der falsche Maßstab verwendet wird [ing22]
[2]Fehler tritt durch zufällige Abweichungen in der Messung auf [ing22]

Tabelle 1: Messungen der Versuchsreihe

Teilnehmer	Körpergröße [cm]	Mittlere Schrittlänge [cm]
Michaela	160,0	71,1
Melek	164,0	70,8
Jenny	170,5	68,8
Patrick	174,0	67,1
Phillip	175,0	79,8
Carsten	176,0	65,3
Joshoua	177,0	73,7
Tim	177,0	72,6
Ferdinand	178,0	87,7
Christian	178,0	76,8
Marvin	180,0	71,1
Sebastian 1	181,5	74,6
Hanna	182,0	71,4
Sven	182,5	83,9
Sebastian 2	182,5	76,5
Eren	183,0	69,0
Marco	184,0	75,5
Marcel	186,0	71,9
Bernd	187,0	81,4
Max	188,0	74,4

Mit den Messwerten kann nun die lineare Regression durchgeführt werden. Zunächst sollte man sich einen ersten Einblick in die gemessenen Daten verschaffen, indem man charakteristische *Kenngrößen* ableitet und diese auswertet [Beu07]. Zum Beispiel die Mittelwerte $\bar{x}$ und $\bar{y}$. Diese und weitere statistische Kenngrößen werden in den folgenden Kapiteln berechnet und ihre Aussage in Bezug auf die Messungen erläutert.

3.2 Berechnung statistischer Kenngrößen

3.2.1 Empirischer Mittelwert

Der empirische Mittelwert, auch *Erwartungswert* genannt, gibt an welcher Wert für X bzw. Y im *Mittel* zu erwarten ist [Beu07]. Diese lassen sich jeweils für die Werte $x_i, 0 \leq i \leq n-1$ mittels der Formel 3 berechnet werden. Äquivalent kann die Formel auch für die Berechnung von $\bar{y}$ herangezogen werden.

$$\bar{x} = \frac{x_0 + x_1 + \ldots + x_{n-1}}{n} = \frac{1}{n} \cdot \sum_{i=0}^{n-1} x_i \tag{3}$$

Unter Anwendung der genannten Formel ergeben sich folgende Erwartungswerte für X und Y. Der durchschnittliche Mensch weist somit eine Schrittlänge von *74,17 cm* und eine Körpergröße von *178,3 cm* auf.

$$\bar{x} = \frac{71,1cm + 70,8cm + \ldots + 74,4cm}{20} = 74,17cm \tag{4}$$

$$\bar{y} = \frac{160,0cm + 164,0cm + \ldots + 188,0cm}{20} = 178,3cm \tag{5}$$

3.2.2 Empirische Standardabweichung

Die empirische Standardabweichung ist die durchschnittliche quadrierte Abweichung vom Mittelwert [BC19] und gibt an wie groß die Streuweite der Messpunkte vom Mittelwert sind. Die Standardabweichung für die Werte $x_i, 0 \leq i \leq n-1$ wird durch die vorliegende Formel 6 beschrieben.

$$s_x = \sqrt{\frac{1}{n} \cdot \sum_{i=0}^{n-1} (x_i - \bar{x})^2} \tag{6}$$

Für die Standardabweichung ergibt sich unter Anwendung der Formel der Wert von *5,50 cm*, d.h. die Schrittweiten sind im Durchschnitt *5,50 cm* vom eigentlichen Mittelwert von *74,17 cm* entfernt.

$$s_x = \sqrt{\frac{1}{20} \cdot ((71,1cm - 74,17cm)^2 + \ldots + (74,4cm - 74,17cm)^2)} = 5,50cm \tag{7}$$

Die gleiche Formel kann auch für die Werte $y_i, 0 \leq i \leq n-1$ verwendet werden. Die Körpergröße liegt im Durchschnitt *7,01 cm* von ihrem Mittelwert entfernt.

$$s_y = \sqrt{\frac{1}{20} \cdot ((160,0cm - 178,3cm)^2 + \ldots + (188,0cm - 178,3cm)^2)} = 7,01cm \tag{8}$$

3.2.3 Korrigierte empirische Standardabweichung

Die korrigierte Standardabweichung, auch Stichprobenvarianz genannt, unterscheidet sich nur minimal von der empirischen Standardabweichung und kommt meist nur bei kleinen Stichprobengrößen zum Tragen. Der einzige Unterschied liegt in der Anzahl der Freiheitsgrade im Nenner der Multiplikation mit der Summe aller Abweichungen. Dieser ist um 1 reduziert und liefert damit laut [BC19] den unverzerrten Schätzer.

$$s_{korr_x} = \sqrt{\frac{1}{n-1} \cdot \sum_{i=0}^{n-1} (x_i - \bar{x})^2} = 5,65cm \tag{9}$$

Die korrigierte Standardabweichung beträgt unter Anwendung der korrigierten Formel 9 *5,65 cm* für die Schrittlänge und *7,20 cm* für die Körpergröße. Damit ist die Weite der Streuung von ihrem Mittel etwas größer geworden.

3.2.4 Standardfehler des Mittelwerts bzw. Messunsicherheit

Der Standardfehler des Mittelwerts wird mit der Formel 10 berechnet und gibt an, wie hoch die Abweichung des Stichprobenmittels vom tatsächlichen Mittelwert ist.

$$s_{\bar{\mathrm{x}}} = \frac{s_{korr}}{\sqrt{n}} = \frac{5,65}{\sqrt{20}} = 1,26cm \tag{10}$$

Die durchschnittliche, errechnete Abweichung vom Mittel für die Schrittlänge beträgt somit *1,26 cm* und für die Körpergröße *1,61 cm*.

$$s_{\bar{\mathrm{y}}} = \frac{7,20}{\sqrt{20}} = 1,61cm \tag{11}$$

Der Standardfehler des Mittelwerts wird auch statistische Messunsicherheit $u(x)$ genannt und kann auch mithilfe der Formel 12 mathematisch berechnet werden. Das Ergebnis kommt auf das gleiche heraus.

$$u(x) = s_{\bar{\mathrm{x}}} = \frac{s}{\sqrt{n}} = \sqrt{\frac{1}{n \cdot (n-1)} \cdot \sum_{i=0}^{n-1} x_{\mathrm{i}}(x_{\mathrm{i}} - \bar{x})^2} \tag{12}$$

3.2.5 Relative Standardunsicherheit

Die relative Standardunsicherheit, auch Residuum genannt, gibt an wie hoch die Unsicherheit relativ zum Mittel $\bar{x}$ ist. Nimmt man die oben berechnete Standardunsicherheit von *1,26 cm* her und stellt das in Relation zu der Breite eines Stuhls und in Relation zu der Breite eines Raumes, so ist die Standardunsicherheit bei dem Stuhl deutlich größer als bei dem Raum. Unter Verwendung der Formel 14 ergibt sich eine relative Messunsicherheit für die Schrittlänge von rund *1,7 Prozent* und für die Körpergröße von rund *0,9 Prozent*.

$$u_{rel}(x) = \frac{u(x)}{\bar{x}} = \frac{1,26cm}{74,17cm} = 0,017 \tag{13}$$

$$u_{rel}(y) = \frac{u(y)}{\bar{y}} = \frac{1,61cm}{178,30cm} = 0,009 \tag{14}$$

Nach der Bestimmung der grundlegenden statistischen Parameter, die bereits eine erste Einsicht in die Messwerte geben konnten, wird nun die Berechnung der linearen Regression und dem Bestimmtheitsmaß durchgeführt, diese gibt schlussendlich Aufschluss über eine Korrelation der Werte X zu Y.

3.3 Herleitung der Regressionsformel und Berechnung des Bestimmtheitsmaßes

3.3.1 Abweichungsquadratsummen

Die Summe der quadratischen Abweichungen der Messwerte von dessen Mittel, auch Abweichungsquadratsumme genannt, wird für die nachfolgenden Berechnungen benötigt, z.B. für die Berechnung der Steigung der Regressionsgerade oder des Bestimmungsmaßes. Die Quadratsummen der Abweichungen werden durch die Formel 15 und Formel 16 bestimmt.

$$S_{xx} = \sum_i (x_i - \bar{x})^2 = \sum_i x_{\mathrm{i}}^2 - n \cdot \bar{x}^2 = 606cm \tag{15}$$

$$S_{yy} = \sum_i (y_i - \bar{y})^2 = \sum_i y_{\mathrm{i}}^2 - n \cdot \bar{y}^2 = 984,20cm \tag{16}$$

Damit ergeben sich für die Quadratsummen der Abweichungen für *X* der Wert *606 cm* und für *Y* der Wert *984,20 cm*.

3.3.2 Empirische Kovarianz

Die Kovarianz S_{xy} gibt an, wie stark der Regressand vom Regressor abhängt. Ein Wert über *0* bedeutet, dass die Variablen stark voneinander anhängen, bei einem Wert unter *0* spricht man von einer gegenläufigen Abhängigkeit [BC19].

$$S_{xy} = \sum_i (x_i - \bar{x}) \cdot (y_{\mathrm{i}} - \bar{y}) = \sum_i x_{\mathrm{i}} \cdot y_{\mathrm{i}} - n \cdot \bar{x} \cdot \bar{y} \tag{17}$$

Die Berechnung, nach Formel 17, ergibt einen Wert von *243,78 cm* für die Kovarianz. Das bedeutet, dass die Veränderung des Regressors einen starken Einfluss auf den Regressanden hat.

$$S_{xy} = (71,1cm \cdot 160,0cm + \ldots + 74,4cm * 188,0cm) - 20 \cdot 74,17cm \cdot 178,3cm = 243,78cm \tag{18}$$

3.3.3 Steigungskoefzient

Nach Vorliegen der Wertes für die Kovarianz und der Abweichungsquadratsumme von *X* kann nun auch die Steigung der Regressionsgeraden, auch Steigungskoefzient genannt, mit der Formel 19 berechnet werden.

$$m = \frac{S_{\mathrm{xy}}}{S_{\mathrm{xx}}} = \frac{243,78cm}{606cm} = 0,4023cm \tag{19}$$

Die Steigung beträgt in diesem Fall *0,4 cm*, d.h. mit einer Erhöhung der Schrittlänge um einen Zentimeter wird die Körpergröße um *0,4 cm* größer geschätzt.

3.3.4 Restvariable

Die Restvariable stellt eine nicht direkt beobachtbare und zufallsbedingte Residualvariable[3] dar [Eck19], d.h. sie ist ein zufälliger Wert, der bei $X = 0$ den Y-Wert angibt. Dieser wird mit der Formel 20 berechnet und beträgt im vorliegenden Fall *148,63 cm* für die Körpergröße bei einer Schrittgröße von *0*.

$$b = -m \cdot \bar{x} + \bar{y} = -0,4 \cdot 74,17 + 178,3cm = 148,46cm \tag{20}$$

3.3.5 Bestimmtheitsmaß

Das Bestimmtheitsmaß r^2, auch Determinationskoeffzient genannt, gibt an wie genau die Schätzung mittels der Regressionsgeraden ist. Die Werte liegen zwischen *0* und *1*. Je höher das Bestimmtheitsmaß ist, desto besser ist die Schätzung des Y-Wertes mittels der Regression, desto näher sind die Messpunkte an der Regressionsgeraden. Berechnet wird das Bestimmungsmaß anhand der Formel 2 und beträgt für das vorliegende Beispiel *0,1*.

$$r^2 = \frac{243,78^2 cm}{606cm \cdot 984,2cm} = 0,0996 \tag{21}$$

3.3.6 Lineare Korrelationskoeffizient

Der wichtigste Indikator zur Bestimmung, ob von einem Wert zuverlässig auf einen anderen Wert geschlossen werden kann, ist der Korrelationskoeffizient r. Dieser nimmt ebenfalls Werte zwischen *0* und *1* an, bei $0,8 < r < 1,0$ liegt eine starke Korrelation vor. Bei einem Wert unter $0,8$ spricht man von einer schwachen Korrelation. Mit der Berechnung 22 ergibt sich ein Wert von rund *0,32*.

$$r = \frac{S_{\text{xy}}}{\sqrt{S_{\text{xx}} \cdot S_{\text{yy}}}} = \frac{243,78cm}{\sqrt{606cm \cdot 984,2cm}} = 0,32. \tag{22}$$

Alle wichtigen statistischen Kenngrößen für die Ergebnisgewinnung liegen nun vor, die Problembearbeitung ist hiermit abgeschlossen. In den folgenden Kapiteln wird auf die daraus resultierenden Ergebnisse näher eingegangen, grafisch dargestellt und abschließend diskutiert.

[3] lat.: *residuum* - Rest

4 Ergebnis und Diskussion

Das gesamte Ergebnis der Regressionsanalyse, mit dem Mittelpunkt (blau), der Regressionsgeraden (blaue Linie) und der Regressionsformel, wird in Abbildung 2 dargestellt. Betrachtet man die Abweichungen der Punkte von ihren Werten auf der Regressionsgerade, so erkennt man, dass die Abweichungen bei der Körpergröße größer sind als bei der Schrittlänge. Dies wird auch durch die Betrachtung der Abweichungsquadratsummen (siehe Kapitel 3.3.1) deutlich. Die Abweichung ist bei den Werten der Körpergröße im Schnitt 1,5x größer als bei den Werten der Schrittlänge.

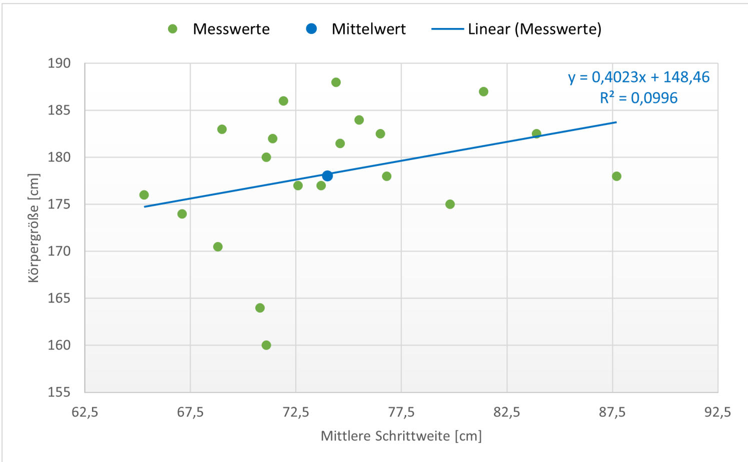

Abbildung 2: Regressionsgerade und Mittelpunkt

Aus der Darstellung geht auch die Regressionsformel $y = 0{,}4023x + 148{,}46\ cm$, der zugrunde liegenden linearen Regression, hervor. Die Restvariable muss hier besonders kritisch betrachtet werden. Diese zeigt auf, dass bei einer Körpergröße von *148,46 cm* eine geschätzte Schrittlänge von *0 cm* vorliegt. Für Personen, die kleiner sind als *148,46 cm* würde die Verwendung der ermittelten Regressionsformel keine realistischen Zahlen schätzen. Zum Beispiel würde bei einer Körpergröße von *120 cm* die Schrittlänge bei *-70 cm* geschätzt werden. Bei Verwendung einer breiter gestreuten Versuchsreihe, in Bezug auf die Körpergröße, würde die lineare Regression ggf. bessere Ergebnisse liefern.

Schlussendlich gibt aber der lineare Korrelationskoeffizient Aufschluss darüber ob Sherlock Holmes mit seiner Behauptung recht hatte, oder nicht. Der Wert für r^2 bei der Berechnung (siehe Kapitel 3.3.6) liegt lediglich bei *0,32*. Für eine starke Korrelation wird jedoch ein Wert von mind. *0,8* benötigt. Es ist davon auszugehen, dass es sich bei der Korrelation von Körpergröße zu Schrittlänge um eine schwache Korrelation handelt.

5 Zusammenfassung

Ziel der hier vorliegenden Ausarbeitung war es, herauszufinden, ob es eine Korrelation zwischen der Schrittlänge und der Körpergröße eines Menschen gibt. Zu diesem Zweck wurde eine lineare Regressionsanalyse durchgeführt und das Bestimmtheitsmaß ermittelt. Dieses gibt Aufschluss darüber, ob eine starke Korrelation der beiden Werte vorliegt, oder nicht. Dabei ergab sich aus der Berechnung ein Wert von *0,32* und, entgegen der Behauptung von Sherlock Holmes, auf eine schwache Abhängigkeit von Körpergröße zu Schrittlänge deutet.

Andererseits musste festgestellt werden, dass die gewählte Stichprobe nicht repräsentativ für die Bevölkerung ist. Die Probanden waren alle bereits im Erwachsenenalter und wiesen lediglich eine Körpergröße zwischen *160,0 cm* und *188,0 cm* auf. Zwar konnte die Behauptung Sherlocks widerlegt werden, jedoch nur unter der Einschränkung, dass lediglich erwachsene Probanden vermessen wurden.

Abschließend kann festgehalten werden, dass die Frage, ob eine Korrelation von Schrittlänge zur Körpergröße vorliegt, differenziert beantwortet werden muss. Einerseits kann die Aussage widerlegt werden. Andererseits wurde die statistische Berechnung mit einer nicht repräsentativen Stichprobe durchgeführt. Somit empfiehlt sich eine Wiederholung der Versuchsreihe mit einer repräsentativeren Stichprobe, bei der die Körpergrößen eine breitere Streuung aufweisen und deutlich mehr Messungen durchgeführt werden.

Literaturverzeichnis

[BC19] Martin Missong Björn Christensen Sören Christensen. *Statistik klipp und klar*. Springer Gabler, 2019. ISBN: 978-3-658-27217-3.

[Beu07] Ottmar Beucher. *Wahrscheinlichkeitsrechnungund Statistik mit MATLAB*. 2., bearbeitete Auflage. Springer, 2007. ISBN: 978-3-540-72155-0.

[DU18] Jochen Mayerl Dieter Urban. *AngewandteRegressionsanalyse:Theorie, Technik undPraxis*. 5., überarbeitete Auflage. Springer VS, 2018. ISBN: 978-3-658-01914-3.

[Eck19] Peter P. Eckstein. *Statistik für Wirtschaftswissenschaftler*. 6., aktualisierte und erweiterte Auflage. Springer Gabler, 2019. ISBN: 978-3-658-24797-3.

[ing22] ingenieurkurse.de. *Systematische und statistische Messfehler*. 2022. URL: `https://www.ingenieurkurse.de/physik/mathematische-grundlagen/systematische-und-statistische-messfehler.html`.