

Statistik & Mathematik in Anwendung

Ein Kurs zur
Marktanalyse in der Glücksspielindustrie

Leseprobe



ELENA
AKA Fu

für potentielle

Verleger/
Sponsoren

und alle anderen
Interessierten

Die Wissenschaft de

FUSSBALL

VORHERSAG

Basiswissen Sp

Über die Autorin

Elena Schälke ist Finanzexpertin, Betriebswirtschaftlerin und Gründerin von Fussballwitwe.com.

Noch als Jugendliche wurde sie 1989 in ihrem ersten Projekt aktiv, wo sie eine führende Rolle bei der Bildung eines erstklassigen Bildungs- und Kulturzentrums in Dresden übernahm.

Nach ihrem Umzug nach England spielte sie eine entscheidende Rolle bei der Abwendung eines signifikanten Haushaltsdefizits eines Schulverbandes.

Elena arbeitete auch als Fundraising Beraterin für den Deutschen Entwicklungsdienst in Uganda.

Es war in England und später in Afrika, dass Elena begann, den schädlichen Einfluss des Glücksspiels auf die Menschen zu bemerken. Sie war erstaunt über das Unvermögen von Wettlern, auch nur einfache Finanzkonzepte wie Prozentsätze und Wahrscheinlichkeiten, zu verstehen und anzuwenden, kombiniert mit einem allgemeinen Mangel an Kenntnissen des Buchmacher-Marktes.

Im Jahr 2010 rief sie Fussballwitwe.com ins Leben, um die Menschen über den Wettmarkt, Finanzmanagement, Statistik und Quotenberechnung aufzuklären.

Elena studierte Pädagogik und hält auch ein Diplom in Wirtschaftswissenschaften mit den Schwerpunkten quantitative Analyse, Mathematik und Finanzmanagement.

Ihre Leidenschaften sind neben der Mathematik Familie und Freunde, Reisen, entdecken verschiedener Kulturen, Wandern, Fotografieren und Computertechnologien.





Die Wissenschaft der Fußballvorhersage

Mehr als/Weniger als 'X' Tore

EINFÜHRUNG DER AUTORIN

GRUNDLAGEN: WAHRSCHEINLICHKEITEN, WETTQUOTEN, VALUE, RENDITE, PROFITABILITÄT

1. GRUNDLEGENDE STATISTISCHE TERMINOLOGIE

1.1. HÄUFIGKEITSVERTEILUNG DER TORE IN DER BUNDESLIGA

- 1.1.1. Auszählen der Bundesliga-Tore
- 1.1.2. Mittelwert (Erwartungswert)

1.2. STREUUNG UM DEN MITTELWERT

- 1.2.1. Stichprobengröße ist sehr wichtig!
- 1.2.2. Mittlere Absolute Abweichung
- 1.2.3. Prozentuale Abweichung (Mittlere Relative Abweichung)

1.3. STANDARDABWEICHUNG - DAS HAUPTMAß FÜR VARIABILITÄT

- 1.3.1. Standardabweichung in Formeln
- 1.3.2. Übungen: Mittlere Absolute Abweichung und Standardabweichung
- 1.3.3. Die Relative Standardabweichung: Variationskoeffizient

1.4. DIE VERTEILUNGSFUNKTION: HÄUFIGKEITSVERTEILUNG IN PROZENT

- 1.4.1. Übungen: Prozentrechnung
- 1.4.2. Graphische Darstellung von Verteilungen
- 1.4.3. Variationskoeffizient bei Wahrscheinlichkeitsfunktionen
- 1.4.4. Übungen: Standardabweichungen und Häufigkeitsverteilung

1.5. KOVARIANZ UND KORRELATION

- 1.5.1. Form: Die Empfundene Stärke der Teams (HO/AO Quotient)
- 1.5.2. Empirischer Korrelationskoeffizient

1.6. DER STANDARDFEHLER DES MITTELWERTS

1.7. SIGNIFIKANZ, SIGNIFIKANZNIVEAU, KONFIDENZINTERVALLE

1.8. REVISIONSAUFGABEN: TORVERTEILUNGEN, ABWEICHUNGEN, KORRELATION

2. WETTQUOTENBERECHNUNG

2.1. WAHRSCHEINLICHKEIT UND WETTQUOTEN

- 2.1.1. Vorhersage der durchschnittlichen Wettquoten für die Bundesliga
- 2.1.2. Durchschnittliche Wettquoten der Buchmacher
- 2.1.3. Übungen: Wahrscheinlichkeiten, Wettquoten, Spannweiten

2.2. ERÖFFNUNGSWETTQUOTEN SPANNWEITE

- 2.2.1. Wettbörse (Betfair) Eröffnungswettquoten
- 2.2.2. Wettquotenberechnung: Mannschaft vs. Mannschaft
- 2.2.3. Übungen: Prozentrechnung, Standardabweichung, Wettquoten

2.3. BERECHNUNG DER 'NULL' (FAIREN) WETTQUOTEN

- 2.3.1. Was sind 'Null' - Quoten? Was sind faire Wettquoten?
- 2.3.2. Markteröffnungs-Wettquoten
- 2.3.3. Eigene Wettquoten Berechnen
- 2.3.4. ‚Tendenzen‘ beim Setzen von Wettquoten

2.4. CLUSTER TABELLEN: ÜBER/UNTER 'X' TORE

- 2.4.1. Was sind Clustergruppen?
- 2.4.2. Gruppierung nach Stärke der Teams
- 2.4.3. Torverteilung nach Mannschaften
- 2.4.4. Kalkulationstabellen nach Mannschaften
- 2.4.5. Standardabweichung / Kovarianz Tabellen
- 2.4.6. Übungen: Verwendung der Clustertabellen zur Wettquotenberechnung

2.5. WETTQUOTEN MIT HILFE DES VALUEKALKULATORS

2.6. REVISIONSAUFGABEN: WETTQUOTENBERECHNUNG

3. RISIKOMANAGEMENT

3.1. FINANZWIRTSCHAFTLICHE BEGRIFFE: RENDITE, ROI UND RENTABILITÄT

- 3.1.1. Einsatz
- 3.1.2. Rendite (Yield)
- 3.1.3. Return on Investment (ROI)
- 3.1.4. Rentabilität (Profitabilität)
- 3.1.5. Übungen: Rendite, ROI und Profitabilität

3.2. RISIKOPROGNOSE UND RISIKOBEWERTUNG: WERTIGKEIT (VALUE) EINES HANDELS (WETTE)

- 3.2.1. Was ist Wertigkeit (Value)? Was ist ein mathematischer Vorteil?
- 3.2.2. Wetten auf Wertigkeit (Value) in der Praxis
- 3.2.3. Übungen: Wertigkeit (Value)

3.3. VORBEUGENDE MAßNAHMEN: ANFANGSBANK KALKULIEREN

- 3.3.1. Länge von Gewinn- und Verlustreihen
- 3.3.2. Zeitpunkt von Siegesserien und Niederlagen
- 3.3.3. Anfangsbank - Faustregel

3.3.4. Übungen: Gewinn- & Verlustserien, Anfangsbank

3.4. REVISIONSAUFGABEN: RISIKOMANAGEMENT

MARKTDYNAMIK UND PROGNOSE

4. MARKTDYNAMIK

- 4.1. TORLINIEN VERSTEHEN IN ÜBER/UNTER TORWETTEN
- 4.2. „OVERROUND“ – EINE FORM VON ARBITRAGE
- 4.3. „HEBELWIRKUNG“ VON WETTQUOTEN
- 4.4. KOMBINATIONSWETTEN: DES BUCHMACHERS LIEBLING
- 4.5. WETTQUOTEN SIND PREISE DER WETTEN
- 4.6. FALLENDE WETTQUOTEN STEHEN FÜR PREISSTEIGERUNG
- 4.7. BUCHMACHER PASSEN WETTQUOTEN AN DIE ÖFFENTLICHE MEINUNG AN
- 4.8. STATISTISCHE PHÄNOMENE
- 4.9. PRÄZISION, RICHTIGKEIT UND GENAUIGKEIT

5. WETTEN IN DER PRAXIS

- 5.1. NICHT ALLES, WAS GLÄNZT, IST GOLD
- 5.2. SIMULATION 2,5 TORE WETTEN/ INFLECTIONPOINTS
- 5.3. METHODE I: WETTEN FINDEN MIT HILFE DER CLUSTERGRUPPENTABELLEN
- 5.4. METHODE II: VALUEWETTEN FINDEN MIT HILFE DES VALUEKALKULATORS
- 5.5. ÜBUNGEN: CLUSTERGRUPPEN, VALUEBERECHNUNG UND IDENTIFIKATION VON WETTEN MIT WERTIGKEIT

BERÜHMTE ABSCHLIEßENDE WORTE

LÖSUNGEN ZU DEN ÜBUNGEN

ABKÜRZUNGSVERZEICHNIS

TABELLEN UND GRAFIKEN INDEX

WEITERFÜHRENDE LITERATUR

Einführung der Autorin

Dieser Kurs vermittelt die für das Verständnis der Quotenkalkulation und des Buchmacher-Marktes notwendigen Kenntnisse.

Es behandelt relativ einfache deskriptive Statistik und zeigt, wie Datensätze ausgewertet, Wahrscheinlichkeiten und Wettquoten berechnet, Marktchancen analysiert und fundierte Entscheidungen getroffen werden können, um Fußballergebnisse vorherzusagen.

Zu den Themen, die behandelt werden, gehören unter anderem: Verteilungen, Abweichungen, Graphiken und Diagramme, Wettquotenberechnung, Finanzterminologien, Risikomanagement und natürlich die Frage, ob und wie "Value" im Wettmarkt gefunden werden kann.

Mit Vermittlung eines tieferen Verständnisses der vielen verschiedenen Elemente, die erforderlich sind, um den Buchmacher-Markt und die Wettquotenberechnung zu verstehen, werden in diesem Buch viele falsche Überzeugungen, die die Mehrheit der Wettfreunde und Fußballfans, entlarvt.

Denken Sie wirklich, dass ein Buchmacher seine Wettquoten auf Mannschaftsnachrichten, Verletzungen, Suspendierungen, Form, etc. basiert?

Ziel dieses Kurses ist es, Mathematik und grundlegende Statistiken auf praktische Weise zu vermitteln, um Fußballergebnisse vorherzusagen. Sie werden überrascht sein zu lernen, dass dies nicht nur möglich, sondern sogar ganz einfach ist, wenn man etwas Statistik anwendet.

Sie werden auch erstaunt sein, zu lernen, wie Buchmacher ihre Wettquoten steuern und handhaben. Sie werden herausfinden, wie die Wettquoten während der Ante-Post-Phase manipuliert werden, jedoch nicht von den Buchmachern, sondern von den Spielern selbst.

Am Ende dieses Kurses sollten Sie in der Lage sein, Über/Unter-Tore mit augenfälliger Genauigkeit vorherzusagen, Ihre Freunde zu beeindrucken und sich selbst zu verblüffen.

Der Startschuss ist gefallen...

Das Spiel läuft!

Über dieses Buch

Dieser Kurs weicht von anderen Büchern und Veröffentlichungen, die Fußballprognosen und Glückspieltipps anbieten, auf folgende Weise ab:

- Er vermittelt, wie **mit Hilfe von Mathematik und Statistik Fußballergebnisse vorhergesagt** werden. Es gibt keine Annahmen, Vermutungen oder emotionale Erklärungen; alles basiert auf Zahlen und Fakten und ist durch diese belegt.
- Das Hauptziel dieses Kurses ist es, **Wettquotenberechnung zu lehren** und denjenigen, die bisher durch Mathematik entfremdet wurden, wieder einen Enthusiasmus für Zahlen zu vermitteln.
- Dieses Buch erklärt im Detail, wie Buchmacher ihre Wettquoten einstellen; die Begrenzungen, die sie haben, welche „Preiskalkulationsfehler“ verursachen.

Dieses Buch kann auch ergänzend fürs Studium von Grundlagenstatistik verwendet werden. Es behandelt einige statistische Konzepte, die in den ersten Semestern an Universitäten gelehrt werden.

- **Leicht verständliche Beispiele stärken das Verständnis** und helfen, statistische Konzepte in der Praxis anzuwenden.
- **Klare und verständliche Sprache** mit zahlreichen Fußnoten zur weiteren Verbesserung des Lernens durch externe Quellen

Annahmen

Um dieses Buch auf eine überschaubare Anzahl von Seiten zu beschränken, wird angenommen, dass der Leser in der Vergangenheit ein wenig Erfahrung mit Statistik gemacht hat. Zum Beispiel sind "Mittelwert" oder "Median" vertraute Begriffe, Diagramme können bearbeitet werden und man erinnert sich vage daran, eine "Verteilung" gesehen zu haben.

Auch wird davon ausgegangen, dass der Leser über Grundkenntnisse in mathematischen Operationen verfügt, beispielsweise Zeichen wie 'x', 'y' und auch das Summierungszeichen erkennt.

Eine weitere Annahme ist, dass Grundlagen wie das Finden der Quadratwurzel einer Zahl oder die Quadratur einer Zahl bekannt sind, oder man zumindest weiß, wo die richtigen Knöpfe auf dem Taschenrechner zu finden sind.

Aber bitte keine Sorge, wenn Sie Mathematik als Herausforderung sehen oder sich als extrem rostig empfinden; alles wird so gründlich und anschaulich wie nur möglich erklärt.

Wie dieses Buch organisiert ist

Dieses Buch ist in zwei Teile gegliedert. Im ersten Teil werden einige statistische Grundkonzepte vorgestellt, gefolgt von einer Einführung in die Wettquotenberechnung und Risikomanagementkontrolle.

Der zweite Teil erklärt Marktdynamik, vertieft Wettquotenberechnung, erklärt die Verwendung von Finanzkennzahlen als Maßstab für Performance und Erfolgsprognose. In diesem Teil wird das im ersten Teil vermittelte Wissen angewendet, um ein Portfolio von Wetten zu konstruieren.

Beide Teile sind in Kapitel unterteilt, die die wichtigsten Ziele der einzelnen Teile in leicht verdauliche Stücke aufteilen.

1. TEIL: Grundlagen: Wahrscheinlichkeiten, Wettquoten, Value, Rendite, Profitabilität

Das Ziel des ersten Teils ist, einige der grundlegendsten mathematischen und statistischen Konzepte zu vermitteln, die für das Wetten von Bedeutung sind.

Das Verstehen von Wettquoten ist hier das Schlüsselement. Es wird ein umfassendes Spektrum mathematischer Formeln, Erklärungen, Beispiele und Übungen präsentiert.

Grundlegende Statistische Terminologie

Dieses Kapitel stellt Häufigkeitsverteilungen und Abweichungen vom Mittelwert im Zusammenhang mit Toren vor, die in einer Liga während einer Spielzeit oder einer Folge von Spielzeiten erzielt wurden.

„Form“ (empfundene Stärke der Teams) wird als Variable präsentiert und Korrelation nahegeführt.

Wettquotenberechnung

Hier behandeln wir, wie Wettquoten kalkuliert werden und in welchen Spannweiten diese im Markt erscheinen. Die Kunst der Quotenberechnung wird in grafische Details zerlegt und die statistische Terminologie des ersten Abschnitts findet seine erste Anwendung.

Risikomanagement

In diesem Abschnitt werden wichtige Finanzkennzahlen wie Rendite, ROI und Rentabilität erläutert. Darüber hinaus werden die Begriffe "mathematischer Vorteil" und "Value" einer Wette sowie die Berechnung der Größe einer Startbank abgedeckt.

2. TEIL: Marktdynamik und Prognose

In diesem Teil dreht sich alles um die praktische Anwendung von Statistik und Mathematik auf Wetten und wie Buchmacher ihre Quoten setzen, um Geld zu verdienen.

Anhand von Beispielen aus der Praxis wird das im ersten Teil gewonnene Wissen genutzt, um zu zeigen, wie man "Fehler" in den Quoten der Buchmachermärkte aufspüren kann.

Marktdynamik

Ein Exkurs in die Marktdynamik der Buchmacherindustrie – Verstehen von Marktpreisen (Wettquoten), Torlinien, Kombinationswetten...

Es werden einige statistische Phänomene erklärt, welche einen psychologischen Einfluss auf das Wettverhalten des Spielers haben.

Wetten in der Praxis

Das wohl spannendste Kapitel für Wettfreunde; denn alles, was man bisher im Kurs gelernt hat, wird hier zusammengefasst und in praktische Tools umgewandelt.

Der Leser ist inzwischen in der Lage, Fußballspielergebnisse mit erstaunlicher Genauigkeit zu prognostizieren und lernt im letzten Kapitel, Simulationen zu erstellen, Wettquoten im Markt zu seinen Gunsten zu interpretieren und die Marktdynamik der Buchmacher auszunutzen.

Dieser Kurs verwendet eine didaktische Lehrmethode, die eine pädagogische Technik ist. Es ist ein sehr strukturierter Lernstil, und als solcher sollte jeder Abschnitt des Kurses gemeistert werden, bevor man zum nächsten übergeht.

Zur weiteren Ermutigung gibt es eine Fülle von Übungen, um das Gelernte mit den am Ende des Buches gefundenen Lösungen zu vergleichen, manchmal mit weiteren Erläuterungen.

Es gibt auch regelmäßig "Übungsboxen", die den Lerner dazu auffordern, das bis dahin Gelernte zu wiederholen und zu festigen.

Am Ende des Buches findet sich ein Abkürzungsverzeichnis und ein Tabellen- und Grafikenverzeichnis. Außerdem finden sich am Ende des Buches einige nützliche Links zu weiterführenden Artikeln, die sich auf die im Kurs enthaltenen Konzepte beziehen.

Und letztendlich das Kleingedruckte! Um den Fluss des Kurses nicht zu unterbrechen, wurden einige Erläuterungen auf die Fußnoten ausgedehnt. Wer neugierig geworden ist, findet in den Fußnoten auch Hinweise zu externen Lernquellen. Bitte achten Sie daher auf die Fußnoten!

Verwendete Wegweiser

Dieser Kurs enthält visuelle Hilfsmittel, die helfen, fokussiert zu bleiben und die Aufmerksamkeit auf das Geschehen zu lenken.

Hier sind die Wegweiser, denen man auf der Reise durch diesen Kurs begegnen wird:



Heureka!

Ein Schrei der Genugtuung oder Entdeckungsfreude, den Sie hoffentlich erleben werden, wenn im entscheidenden Moment des Verstehens alles zusammenkommt!



Wissenschaft in Aktion

Dieses Zeichen zeigt spannende und manchmal überraschende Fakten, wo Mathematik, Statistik und Psychologie sichtbar werden, wenn Buchmacher ihre Angebote kalkulieren und Profite planen.



Technischer Kram

Wenn Sie dieses Symbol sehen, können Sie den betreffenden Abschnitt überspringen, wenn Sie Ihr Gehirn nicht mit zu vielen technischen Details noch mehr durcheinanderbringen möchten. Die Erklärungen und Ableitungen der Formeln existieren vor allem für die Puristen, die ein besonderes mathematisches Interesse haben und mehr über die Hintergründe der Schlussfolgerungen wissen möchten.



Merke!

Dieses Zeichen wird verwendet, um bestimmte Ideen zu verstärken, die kritisch sind. Dinge, welche für das Verständnis des Kurses sehr wichtig sind.

Dieses Symbol findet sich auch dort, wo Ideen aus mehreren Kapiteln des Kurses zusammengetragen und Schlussfolgerungen gezogen werden.



Megaphone!!!

Dieses Symbol findet sich an Stellen, wo allgemeine Missverständnisse und Gefahren adressiert und falsche Überzeugungen von Fußballfans herausgefordert werden.

An diesen Stellen des Kurses sind die Informationen von großer Bedeutung, und Sie müssen genau hinschauen, um alles zu verstehen, um auf dem richtigen Weg zu bleiben. Sie sollen es laut und deutlich hören!

Grundlagen: Wahrscheinlichkeiten, Wettquoten, Value, Rendite, Profitabilität

1. Grundlegende Statistische Terminologie

- Torverteilung und Prozentrechnung
- Spannweiten: Abweichung vom Mittelwert und Standardabweichung
- Präzision, Richtigkeit und Genauigkeit

Dieses Kapitel ist eine kurze Einführung in *Deskriptive Statistik*. Die Kenntnis der Verteilung und Spannweite eines Datensatzes liefert Informationen, die eine grobe Vorstellung davon vermitteln, was in Zukunft geschehen wird.

Sie lernen hier, vergangene Distributionen zu analysieren und Spannweiten zu berechnen und mit diesem Wissen die voraussichtliche Verteilung für die nächste Spielzeit zu prognostizieren. Sie vergleichen Ihre eigenen Berechnungen mit den tatsächlichen Ergebnissen, und werden überrascht sein, wie berechenbar Fußballergebnisse sind, wenn man nur ein wenig von Statistik versteht.

1.1. Häufigkeitsverteilung der Tore in der Bundesliga

Wettquoten¹ stellen Wahrscheinlichkeiten dar, und Wahrscheinlichkeiten sind ein Teilbereich der Statistik. Die Grundlage von Statistik und damit von Wettquoten ist das Auszählen von Datensätzen.

Deshalb beginnen wir zunächst einmal mit dem Auszählen von Toren in fünf Spielzeiten der Bundesliga, und zählen die Zahl der Spiele mit genau "null" Toren, oder "einem" Tor oder "zwei" Toren, und so weiter, bis wir "sieben oder mehr" Tore erreicht haben.

¹ Der Kurs verwendet europäische Wettquoten (eins geteilt durch die Wahrscheinlichkeit). Wer mehr mit einem anderen Quotenformat vertraut ist und/oder wissen möchte, wie man andere Quoten umwandelt, ist hier ein Artikel zum Thema: [Understanding Betting Odds – Moneyline, Fractional Odds, Decimal Odds, Hong Kong Odds, IN Odds, MA Odds](#)

1.1.1. Auszählen der Bundesliga-Tore

Eine Bundesliga-Spielzeit umfasst 306 Spiele, in welcher sich 18 Mannschaften auswärts und zu Hause treffen. Multipliziert man diese Zahl mit fünf (Spielzeiten), ergibt sich eine Zahl von 1.530 Spielen, die Summe der acht Zahlen in der blauen Spalte 'Total' unten.

Deutsche Bundesliga 1 Gesamttore							
	2012-13	2013-14	2014-15	2015-16	2016-17	Summe	Durchschn. (\bar{x})
0 Tore	20	13	28	24	19	104	20.8
1 Tore	36	34	49	45	47	211	42.2
2 Tore	80	72	72	66	72	362	72.4
3 Tore	69	68	58	67	79	341	68.2
4 Tore	47	55	53	58	43	256	51.2
5 Tore	33	35	27	25	20	140	28.0
6 Tore	12	18	11	18	13	72	14.4
7 + Tore	9	11	8	3	13	44	8.8
	306	306	306	306	306	1530	

Abb. 1: Auszählung der Tore Deutsche Bundesliga 1 - 2012-13 bis 2016-17

Machen Sie sich mit dem Inhalt der obigen Tabelle und deren Aufbau vertraut.

Es sei Ihnen strengstens ans Herz gelegt, dies mit jeder Grafik und Tabelle in jeder Etappe zu tun. Während dieses Kurses sollten Sie immer versuchen, eine Lektion zu "meistern", einschließlich der Übungen, bevor Sie zur nächsten Lektion übergehen.

Als Teil Ihres Kaufs haben Sie eine zusätzliche Excel-Arbeitsmappe mit allen Tabellen und Grafiken erhalten, die für diesen Kurs verwendet wurde, einschließlich der Formeln. Keine Sorge – im Kursbuch wird die Mechanik hinter jeder Formel, die im Kursbuch verwendet wird, im Detail erklärt.

Buchmachermathematik ist ein Thema, das für viele Menschen komplex und schwierig ist.

Daher, selbst wenn Sie bereits einige der statistischen Terminologien kennen, die behandelt werden, sollten Sie dem Kurs Schritt-für-Schritt folgen und trotz eventueller Vorkenntnisse jede Aufgabe ausführen.

Statistische Konzepte werden oftmals in einem ungewohnten Zusammenhang vorgeführt und auf die Voraussage von Sportergebnissen angewandt, Sie werden garantiert viel Neues dazulernen...

Viel Glück und viel Spaß!

1.1.2. Mittelwert (Erwartungswert)

Wenn man von Mittelwerten (Mittel oder Durchschnitt) in der Statistik spricht, meint man meistens das arithmetische Mittel.

Das arithmetische Mittel \bar{x} erhält man, indem man die betrachteten Zahlen summiert und dann durch deren Anzahl dividiert.

Die torlosen Spiele der Bundesliga (0 Tore) haben beispielsweise einen Mittelwert von 20,8.

$$\frac{20 + 13 + 28 + 24 + 19}{5} = 20,8$$

1. Aufgabe:

Download von <http://www.football-data.co.uk/data.php> und speichern Sie alle Rohdaten der Bundesliga der vergangenen fünf Spielzeiten (2012-2017).

Tipp: Öffnen Sie jede .csv-Datei mit Excel, je nachdem, mit welcher Version von Excel Sie arbeiten, und speichern Sie dann entweder eine .xls-Datei (v8.0 - v11.0: Excel 1997-2003) oder eine .xlsx Excel-Arbeitsmappe (v12.0 und aufwärts: 2007).

Als nächstes sollten Sie Excel verwenden, um die Anzahl der Tore pro Spiel für jede Spielzeit auszuzählen und eine Tabelle ähnlich der Tabelle wie oben in **Abb. 1** (Auszählung der Tore Deutsche Bundesliga 1 - 2012-13 bis 2016-17) zu erstellen.

Berechnen Sie auch die Mittelwerte.

1.2. Streuung um den Mittelwert

Für Wettquotenberechnung muss man den Begriff "Abweichung vom Mittelwert" nicht nur verstehen, sondern sich auch vollständig verinnerlichen!

Im oberen Kapitel haben Sie Tore ausgezählt und "Mittelwerte" gebildet. Jedoch nur einen Mittelwert zu bilden ist leider nicht genug, denn die tatsächlichen Werte (*die beobachteten Werte der Vergangenheit wie auch die prognostizierten Werte der Zukunft*) streuen um diesen.

Man muss sich dies einprägen! Wenn man prognostiziert, passiert es eher selten, dass die neue Messung genau den erwarteten Wert (historischen Mittelwert) ergibt.



So war beispielsweise das arithmetische Mittel (der Durchschnitt) für Spiele mit genau 2 Toren für die Bundesliga Spielzeiten 2010-11 bis 2014-15: 72,2

Wer in der folgenden Spielzeit (2015-16) genau 72 Tore erwartet hätte, wäre enttäuscht worden, denn in Wirklichkeit gab es „nur“ 69 Spiele mit genau 2 Toren.

Dies war die „Abweichung vom Mittelwert“: In der 2015-16 Spielzeit gab es in dieser Gruppe von 2-Tore Spielen drei Spiele weniger als im Durchschnitt der vorhergehenden fünf Jahre.

Als nächstes beschäftigen wir uns mit der Definition der "Abweichung":

In Mathematik und Statistik ist die Abweichung ein Maß für die Differenz zwischen dem beobachteten Wert einer Variablen und einem anderen Wert. (In diesem Kurs werden beispielsweise die Zahl der Tore einer Spielzeit zum Durchschnitt der Tore der letzten fünf Spielzeiten ins Verhältnis gesetzt).

Der Begriff "Abweichung" bezeichnet die Streuung von Werten einer Häufigkeits- oder Wahrscheinlichkeitsverteilung um einen Standortparameter (in der Regel um den Mittelwert).

Im Kapitel **1.1.** haben Sie Tore gezählt und Mittelwerte berechnet.

Wir schauen uns nun den Durchschnitt für 'Unter 1,5 Tore'² Spiele in den fünf Spielzeiten an: Im Durchschnitt endeten 63,0 Spiele pro Spielzeit mit weniger als zwei Toren (waren somit „unter 1,5 Tore“ Spiele).

63,0 ist der 'Mittelwert' - der sogenannte 'Standortparameter'. Dies ist der Erwartungswert (Mittelwert), um welchen herum sich eine 'Häufigkeitsverteilung' befindet, mit einzelnen Werten, die sich um den Mittelwert ‚streu‘en‘.

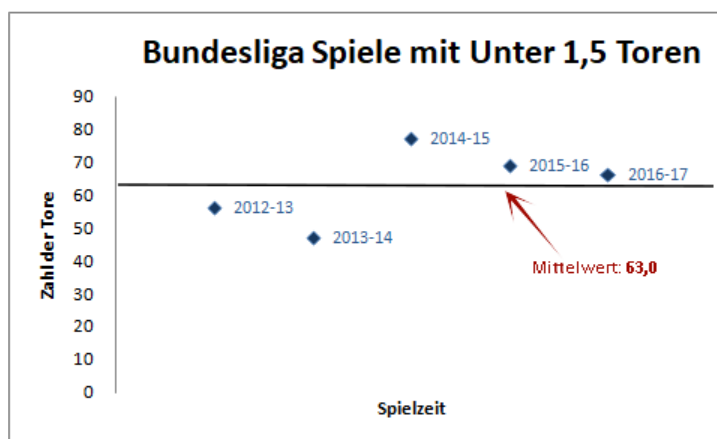


Abb. 2: Streuung Unter 1,5 Tore - Deutsche Bundesliga 2012-13 bis 2016-17

² „Unter 1,5 Tore“ sind weniger als 2, also die Summe der Spiele mit 0 Toren und 1 Tor – siehe auch: 4.1. Torlinien Verstehen in Über/Unter Torwetten, Seite 108

Abb. 2 zeigt die **Streuung um den Mittelwert**.

So endeten in der 2013-14 Spielzeit 47 Spiele mit "Unter 1,5 Toren", während in der 2014-15 Spielzeit es 77 Spiele mit "Weniger als 1,5 Toren" waren. Dies waren die „Außenseiter“; die anderen drei Spielzeiten lagen näher am Mittelwert. Jedoch in keiner der Spielzeiten wurde **genau** der Mittelwert gespielt.

Der nächste Begriff ist: **Spannweite**

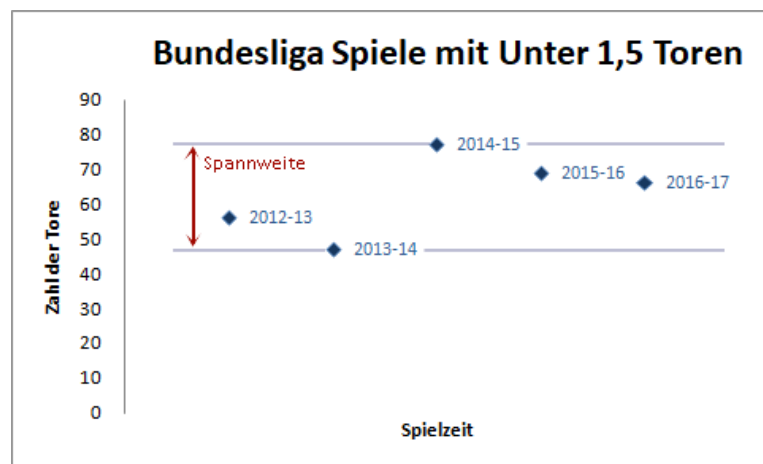


Abb. 3: Spannweite Unter 1,5 Tore - Deutsche Bundesliga 2012-13 bis 2016-17

Um die Spannweite zu ermitteln, findet man das Jahr mit der höchsten Zahl an Unter 1,5 Tore Spielen (hier: 2014-15) und auch das Jahr mit der kleinsten Zahl (hier: 2013-14). Diese beiden Werte werden dann einfach subtrahiert:

2013-14: **47** Spiele mit weniger als 1,5 Toren

2014-15: **77** Spiele mit weniger als 1,5 Toren

Spannweite: $77 - 47 = 30$

Die 'Spannweite' der beobachteten fünf Spielzeiten beträgt 30.

Die Werte der anderen drei Spielzeiten liegen zwischen den niedrigsten und höchsten Zählungen:

2012-13: **56** Spiele mit weniger als 1,5 Toren

2015-16: **69** Spiele mit weniger als 1,5 Toren

2016-17: **66** Spiele mit weniger als 1,5 Toren

Bitte einprägen!

Die beiden Begriffe ‚Spannweite‘ und ‚Streuung‘ werden noch ganz viel im Kurs gebraucht werden!

Nochmals zur Erinnerung:

Die **Spannweite** (englisch: *range*) **R** berechnet sich als Differenz zwischen dem größten und dem kleinsten Messwert.

Streuung beschreibt die **Streubreite von Werten** um einen Lageparameter herum.

Spannweite (R) = maximale Zahl der Messwerte (in unserem Beispiel: die Zahl der Spiele in einer Gruppe, wie beispielsweise ‚Unter 1,5 Tore‘) minus die minimale Zahl der Messwerte

$$R = x_{\max} - x_{\min}$$

Die ‚Spannweite‘ im Beispiel:

$$77 - 47 = 30$$

Nun vergleichen wir diese Zahlen mit dem Mittelwert oder ‚Standortparameter‘ und führen damit den Begriff der **Abweichung** ein.

‚**Abweichung**‘ ist eine **Maßzahl der Streuung** (auch Dispersion).

Um Streuung zu beschreiben gibt es in der deskriptiven Statistik **verschiedene Maßzahlen, welche die Streubreite von Werten um einen Lageparameter herum beschreiben.**

Es gibt **verschiedene Berechnungsmethoden**, welche sich prinzipiell durch ihre **Beeinflussbarkeit beziehungsweise Empfindlichkeit gegenüber Ausreißern** unterscheiden.

Die **Streuung der Häufigkeitsverteilung wird als Standardfehler** bezeichnet.³

³ Wikipedia: [Streuung \(Statistik\)](#)

1.5. Kovarianz und Korrelation

In den vorhergehenden Kapiteln haben wir Tore ausgezählt, Häufigkeitsverteilungen ermittelt, sowie deren Standardabweichungen und Variationskoeffizienten berechnet.

Als ob das nicht schon genug Mathe war, nun befassen wir uns als Nächstes mit der Kovarianz. Denn mit dieser lässt sich erkennen, ob es einen Zusammenhang zwischen Variablen gibt oder nicht.

Dass Buchmacher für Torwetten verschiedene Wettquoten setzen, ist bekannt und wir werden weiter im Kurs behandeln, wie mit Hilfe der Torauszahlungen und Standardabweichungen Wettquoten berechnet werden.

Die Frage, die wir uns in diesem Kapitel stellen, ist, ob es vielleicht noch andere Variablen gibt, welche man ebenfalls benutzen kann, um die Genauigkeit der Voraussage zu erhöhen, wie viele Tore es in einem Spiel geben könnte.

Wir tauchen daher nun in das Thema von Korrelation und Kovarianz ein! Kaffee holen, durchatmen, stark sein! Kovarianz setzt Varianz nochmal eines drauf!

Kovarianz ist ein Maß für den linearen Zusammenhang zweier Variablen.

Ein positives Vorzeichen gibt an, dass sich beide Variablen in dieselbe Richtung bewegen (*steigt der Wert einer Variablen an, steigt auch der Wert der anderen*).

Ein negatives Vorzeichen sagt das Gegenteil über den Zusammenhang aus (*wenn der Wert einer Variablen steigt, fällt der Wert der anderen*).

Ein Wert von Null oder nahe Null deutet darauf hin, dass kein Zusammenhang besteht.

Wir werden an dieser Stelle die Herleitung, Formeln und detaillierte Erklärung der Kovarianz überspringen.

In Effekt ähnelt diese der Abweichung und Varianz, nur dass man mit zwei Variablen rechnet, \mathbf{x} und \mathbf{y} .

Die Variable \mathbf{x} sind unsere Tore, mit denen wir im Kurs bereits hoch- und runtergespielt haben und die Variable \mathbf{y} ist die Stärke der Teams. Wie wir das genau machen, wird auf den folgenden Seiten im Detail erklärt.

Die Kovarianz zeigt die Stärke des Zusammenhangs von zwei Variablen, es ist allerdings (*wie wir es auch bei der Standardabweichung gesehen hatten*) schwierig, aus einem unstandardisierten Wert konkrete Schlüsse zu ziehen.

Hier liegt der Unterschied zur Korrelation: **Korrelation ist die standardisierte Kovarianz**

Wir überspringen daher die Kovarianz und konzentrieren uns auf den Korrelationskoeffizienten.

1.5.1. Form: Die Empfundene Stärke der Teams (HO/AO Quotient)

Soweit haben wir im Kurs Tore ausgezählt und uns Häufigkeitsverteilungen angesehen.

Wir hatten Daten nach Jahren sortiert und dann daraus Mittelwerte gebildet und festgestellt, dass in der Bundesliga eine ähnliche Anzahl von Tore jedes Jahr gespielt werden; natürlich plus/minus der Standardabweichung.

An dieser Stelle kommt nun endlich ‚Form‘ ins Spiel, die empfundene Stärke der Teams und diese ist besonders gut durch die Heim-Unentschieden-Auswärts Wettquoten reflektiert.

Wir kommen im Kurs später noch zu einem Kapitel, in welchem beschrieben und gezeigt wird, dass Buchmacher ihre Preise an die öffentliche Meinung anpassen³⁰, dies spielt uns in die Hände, da dadurch die 1x2 Wettquoten weniger die wahren Wahrscheinlichkeiten widerspiegeln, sondern die öffentliche Meinung.

Wird ein Team von der Öffentlichkeit als besonders stark empfunden, dann ist deren Wettquote entsprechend niedrig und vice versa; wird ein Team als besonders schwach gegen den Gegner empfunden, gibt es eine hohe Wettquote auf einen Sieg dieses Teams.

So spielte beispielsweise am 14. Oktober 2017 Dortmund gegen Leipzig und die Wettquoten waren wie folgt:

Dortmund: **1,57** vs Leipzig: **7,0**

Dortmund ward der eindeutige Favorit, Leipzig der Underdog.

Hier ist ein Screenshot mit allen Wettquoten für dieses Wochenende:

Soccer » Germany » Bundesliga					
13 Oct 2017		1	X	2	B's
18:30	VfB Stuttgart - FC Köln	2.01	3.49	3.87	15
14 Oct 2017		1	X	2	B's
13:30	Bayern Munich - SC Freiburg	1.08	13.06	26.57	14
13:30	Hannover - Eintracht Frankfurt	2.46	3.20	3.08	14
13:30	Hertha Berlin - Schalke	2.61	3.22	2.85	14
13:30	Hoffenheim - FC Augsburg	1.79	3.79	4.52	14
13:30	1. FSV Mainz 05 - Hamburger SV	1.70	3.79	5.17	14
16:30	Dortmund - RB Leipzig	1.57	4.41	5.58	14
15 Oct 2017		1	X	2	B's
13:30	Bayer Leverkusen - Wolfsburg	1.59	4.29	5.60	14
16:00	SV Werder Bremen - B. Monchengladbach	2.69	3.46	2.62	14

Abb. 14: Bundesliga Heim – Unentschieden – Auswärts Wettquoten für das WE 13.10. bis 15.10.2017
(Oddsportal, durchschnittliche Wettquoten, 4 Tage vor Spielbeginn)

Wie man sehen kann, Bayern München war ein ganz eindeutiger Favorit gegen den SC Freiburg und hatte eine kleine Wettquote von gerade mal 1,08. Hertha Berlin und Schalke

³⁰ Kapitel 4.7. Buchmacher passen Wettquoten an die öffentliche Meinung an, Seite 123

wurden als einigermaßen gleichstark bewertet; so hatte Berlin eine Wettquote i.H.v. 2,61 und Schalke 2,85.

Wie die Spiele letztendlich endeten, kann ja jeder selbst nachsehen. Garantiert nicht, wie der Fußballfreund vom Bauchgefühl her vermutet hätte. 😊

Das Schöne an diesen Quoten ist, dass diese gut archiviert sind. Man kommt an diese leicht ran und kann mit ihnen rechnen.

Sie ändern sich nicht mehr und sie geben die öffentlich wahrgenommene Gegnerstärke in Zahlen wider. Was Besseres kann einem Statistiker gar nicht passieren!

Um die Stärke der Gegner in Zahlen auszudrücken, braucht man nur die Wettquoten für den Heim- und Auswärtssieg. Man nimmt diese beiden und dividiert diese.

Dortmund: **1,57** vs Leipzig: **7,0**

$$1,57 \text{ geteilt durch } 7,0 = 0,224$$

Das war's!

Wir haben unsere unabhängige Variable **y**, welche wir später nutzen werden, um zu prüfen, ob es eine Korrelation zwischen der empfundenen Stärke der Teams und den Torergebnissen der Spiele gibt. Ebenfalls interessiert uns natürlich, ob die Wettquoten für Über/Unter Wetten von der empfundenen Stärke der Teams abhängen.

Nutzen des HO/AO Quotienten

HO: Home Odds (Heimwettquote)

AO: Away Odds (Gastwettquote)

Durch die Division von HO durch AO erhält man einen Quotienten, welcher die von der Öffentlichkeit empfundene „Stärke“ der Mannschaften widerspiegelt.

Gibt es beispielsweise ein Spiel mit „gleich starken“ Mannschaften, dann liegt der HO/AO-Quotient um 1,0 (mehr oder weniger) wie beispielsweise im Spiel Hertha Berlin gegen Schalke (siehe **Abb. 14**).

Herthas Heimspiel Wettquote: 2,61 und Schalkes Auswärts Wettquote: 2,85

$$2,61 / 2,85 = 0,9158$$

Wir stellen uns nicht die Frage: „Sind diese Teams wirklich gleichwertig?“

Wir rechnen die 1x2-Wettquoten auch nicht nach, sondern wir verwenden einfach die Marktpreise und gehen davon aus, dass der Markt die öffentliche Meinung so gut wie nur möglich widergespiegelt hat.³¹

Teamranking und öffentliche Meinung wird permanent in den 1x2 Wettquoten berücksichtigt, das müssen Sie hier an dieser Stelle einfach mal glauben. Denn so verdienen Buchmacher ihr Geld!

Clustergruppenbildung mit Hilfe des HO/AO Quotienten

In den Beispielen bisher hatten wir immer mit Saisons gerechnet und die Verteilungen der einzelnen Spielzeiten betrachtet. Das Wort ‚Cluster‘³² ist in den vorhergehenden Kapiteln noch nicht gefallen, aber mathematisch gesprochen sind die Spielzeiten (Saisons) ‚Clustergruppen‘.

Die Gesamtmenge von 1.530 Einzeldaten, welche wir betrachten, sind in Teilmengen von Jahren (Spielzeiten) aufgeteilt.

Mit dem HO/AO Quotienten kann man ebenfalls Clustergruppen bilden, nur dass deren Interpretation (ihre Bedeutung) nicht ganz so offensichtlich ist.

Abb. 15 zeigt die HO/AO Quotienten für das im Kurs betrachtete Beispielwochenende 13.10. bis 15.10.2017:

		Heim	Unentsch	Auswärts	HO/AO
Bayern München	SC Freiburg	1.08	13.06	26.57	0.041
Dortmund	RB Leipzig	1.57	4.41	5.58	0.281
Bayer Leverkusen	Wolfsburg	1.59	4.29	5.60	0.284
1. FSV Mainz 05	Hamburger SV	1.70	3.79	5.17	0.329
Hoffenheim	FC Augsburg	1.79	3.79	4.52	0.396
VfB Stuttgart	FC Köln	2.01	3.49	3.87	0.519
Hannover	Eintracht Frankfurt	2.46	3.20	3.08	0.799
Hertha Berlin	Schalke	2.61	3.22	2.85	0.916
SV Werder Bremen	Mönchengladbach	2.69	3.46	2.62	1.027

Abb. 15: HO/AO Quotienten für das WE 13.10. bis 15.10.2017

8. Aufgabe:

Erstellen Sie eine Exceltabelle, welche alle 1.530 Spielen der 2012-13 bis 2016-17 Spielzeiten enthält, und berechnen Sie deren HO/AO Quotienten.

In welche fünf gleichgroße Clustergruppen lassen sich die Spiele einteilen?

³¹ Hier findet sich ein Artikel dazu auf der Fussbalwitwe.com Webseite: [Buchmacher passen Wettquoten an Erwartungen der Öffentlichkeit an](#)

Bitte lesen Sie diesen Artikel, wenn Sie Zweifel haben. In diesem Artikel analysierte ich das Spiel Heart of Midlothian vs. Aberdeen am 24.08.2013 (Scottish Premier League) und zeigte, dass die Buchmacher die Wahrscheinlichkeiten entsprechend der öffentlichen Meinung und Erwartungen einfach umgekehrt hatten.

³² Cluster leitet sich ab aus dem Englischen Wort cluster = ‚Traube‘, ‚Bündel‘, ‚Schwarm‘, oder ‚Ballung‘

Um die Aufgabe: 8 zu lösen, brauchen Sie folgende Formel in Excel:

PERCENTILE(Array;Alpha)
QUANTIL(Array;Alpha)

Diese Excel Formel gibt das Alpha-Quantil einer Gruppe von Daten zurück. Mit Hilfe dieser Funktion legt man einen Akzeptanzschwellenwert fest.

Möchte man beispielweise eine große Datenmenge in fünf gleichgroße Teile teilen, dann sind die Quantile (Alphawerte): 20% (0,2), 40% (0,4), 60% (0,6), 80% (0,8) und 100% (1,0).

Die Formel gibt die Zahl wieder, wo der „Wendepunkt“ ist, d.h. beispielsweise =QUANTIL(A1:A1530,0.2) gibt diejenige Zahl wieder, bei welcher 20% der Zahlen unterhalb des Schwellenwerts liegen und die anderen 80% oberhalb.

Der Datensatz muss für die Benutzung der Formel nicht sortiert sein.

Quantile für die 1.530 Spiele der Bundesliga

20%: 0,2918	Das Heimteam ist klarer Favorit mit einer sehr guten Chance auf einen Sieg (<i>große gefühlte Sicherheit für den Ausgang des Spiels</i>)
40%: 0,5292	Das Heimteam ist definitiv stärker als der Gast, aber es gibt im Spiel auch durchaus eine gute Chance für ein Unentschieden (<i>schwankende Meinung ob Heim oder Unentschieden</i>)
60%: 0,7830	Das Heimteam könnte durchaus gewinnen, aber es könnte auch eventuell ein Unentschieden oder ein Gast Sieg werden (<i>nicht wirklich klar, in welche Richtung sich das Spiel entwickeln wird</i>)
80%: 1,3172	Die Chance auf ein Unentschieden ist recht hoch, denn die Mannschaften sind ungefähr gleichstark (<i>gefühltes Unentschieden im Spiel</i>)
100%: 23,9400	Das Heimteam ist schwächer als der Gast; es könnte ein Auswärts-sieg werden (<i>der Favorit ist die Gastmannschaft</i>)

9. Aufgabe:

Analysieren Sie das Wochenende 13.10. bis 15.10.2017 (**Abb. 15**):

- ▶ Auf welche Clustergruppen hatten sich die 9 Spiele aufgeteilt?
- ▶ Sind Sie einverstanden (gefühlsmäßig) mit der Clusterung? Spiegeln Sie Ihrer Meinung nach die gefühlten Chancen dieses Wochenendes wider?
- ▶ Checken Sie die Ergebnisse der 9 Spiele. Sind diese tatsächlich so ausgefallen wie die Öffentlichkeit dies vorausgesagt hatte?

Die Spiele fielen an dem betrachteten Wochenende in drei Cluster:

bis 0,2918

Das Heimteam ist klarer Favorit mit einer sehr guten Chance auf einen Sieg (*große gefühlte Sicherheit für den Ausgang des Spiels*)

Bayern München vs SC Freiburg: **5:0**

Dortmund vs RB Leipzig: **2:3**

Bayer Leverkusen vs Wolfsburg: **2:2**

0,2918 bis 0,5292

Das Heimteam ist definitiv stärker als der Gast, aber es gibt im Spiel auch durchaus eine gute Chance für ein Unentschieden (*schwankende Meinung ob Heim oder Unentschieden*)

1. FSV Mainz 05 vs Hamburger SV: **3:2**

Hoffenheim vs FC Augsburg: **2:2**

VfB Stuttgart vs FC Köln: **2:1**

0,7830 bis 1,3172

Die Chance auf ein Unentschieden ist recht hoch, denn die Mannschaften sind ungefähr gleichstark (*gefühltes Unentschieden im Spiel*)

Hannover vs Eintracht Frankfurt: **1:2**

Hertha Berlin vs Schalke: **0:2**

SV Werder Bremen vs Mönchengladbach: **0:2**

Außer in der zweiten Gruppe (*HO/AO Quotient: 0,2918 bis 0,5292*) lag die Öffentlichkeit mit ihrer Meinung so ziemlich daneben, denn außer dass Bayern einen klaren Heimsieg einfuhr, enden die anderen beiden Favoritenspiele mit einem Unentschieden und einem Gast Sieg. Die drei gefühlten Unentschieden trafen ebenfalls nicht ein.

Der Kurs dreht sich jedoch nicht darum, wie man am besten einen Heimsieg, ein Unentschieden oder einen Gast Sieg voraussagt, sondern wie man Tore prognostiziert. Der Ausflug in die 1x2 Wettquoten ist an dieser Stelle nur, um die von der Öffentlichkeit gefühlte „Form“ als eine Variable in die Kalkulationen mit einzubeziehen.

Wir werden später im Kurs sehen, ob die gefühlte „Form“ einen Einfluss auf die Torergebnisse hatte; oder nicht.

Wie genau ist der HO/AO Quotient?

In diesem Kurs werden die Daten von **football-data.co.uk** verwendet. Die Wettquoten werden für Wochenendspiele an Freitagabenden und für die anderen Wochentage an Montagabenden (UK-Zeit) erfasst.

Sie sind genau zum Zeitpunkt der Erfassung, aber selbstverständlich sind sie am nächsten Tag anders, vor allem in den Stunden vor dem Anpfiff, nachdem die Teamnachrichten veröffentlicht wurden.

Normalerweise bewegen sich jedoch Wettquoten nicht so gewaltig, dass ein Team, das anfänglich im "klaren" Favoritencluster war, plötzlich zum Underdog wird.

Was jedoch passieren kann, ist, dass ein Spiel, dessen Wettquoten am Vortag erfasst wurden, näher zum Anpfiff in ein anderes Cluster fällt, insbesondere wenn der HO/AO-Quotient des Spiels am Rande des jeweiligen Clusters platziert ist.

Um den Korrelationskoeffizienten auszurechnen, brauchen wir uns darum keine Sorge zu machen. Jedoch später im Kurs wird diese Frage erneut auftauchen und diskutiert werden, wenn wir anfangen, die Bundesliga in deren 18 Mannschaften aufzuteilen und auch nach Heim und Auswärts zu unterscheiden.

So, nun kommen wir endlich zur Berechnung des Korrelationskoeffizienten...

1.5.2. Empirischer Korrelationskoeffizient

Für eine Messreihe von gepaarten Ausprägungen $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ wird der empirische Korrelationskoeffizient berechnet, indem man die empirische Kovarianz und empirischen Varianzen der jeweiligen Zufallsvariablen verwendet.

Dies führt zu folgender Formel:

$$\text{Kor}_e(x, y) := \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

$\text{Kor}_e(x, y)$ empirischer **K**orrelationskoeffizient der Variablen **x** und **y**

$:=$ "wird per Definition gleich ... gesetzt"

$\sqrt{\quad}$ Quadratwurzel

$(\dots)^2$ Quadrierung des Ergebnisses innerhalb der Klammer

\bullet Multiplikationszeichen

Σ Summe von (*was auch immer auf der rechten Seite des Σ Sigma Zeichen steht*)³³

$i=1$ Dies ist der Startwert des Laufindex, Summationsvariable bzw. Lauf- oder Zählvariable, je nachdem welchen Term man bevorzugt oder in der Schule gelernt hat.

Im unserem Beispiel ist das die erste Variable des Datensatzes

n Dies ist der Endwert, bei welchem wir aufhören.

³³ [Summenzeichen in Wikipedia](#)

In den vorhergehenden Aufgaben und Übungen war $n = 5$ (Spielzeiten), für die Berechnung des Korrelationskoeffizienten ist $n = 1.530$ (jedes einzelne Spiel)

x_i	x beschreibt die erste veränderliche Variable (z.B. die Zahl der Tore), und i ist der dazugehörige Laufindex (wie in: 1. Spiel, 2. Spiel, ..., n . Spiel)
\bar{x}	Arithmetisches Mittel (Mittelwert) der x Verteilung
y_i	y beschreibt die zweite veränderliche Variable (z.B. HO/AO Quotient), und i ist der dazugehörige Laufindex (wie in: 1. Spiel, 2. Spiel, ..., n . Spiel)
\bar{y}	Arithmetisches Mittel (Mittelwert) der y Verteilung

Dies ist eine Formel, die sich nicht auf einem Stück Papier einfach mal schnell ausrechnen lässt. Wir rechnen nun statt mit nur 5 Variablen (*in den Beispielen bis jetzt waren es immer: fünf Saisons*) mit der Gesamtmenge aller Spiele (*1.530 Einzelspiele*).

Um den Beispielberechnungen in diesem Kapitel zu folgen, muss unbedingt ein Tabellenkalkulationsprogramm, wie beispielsweise Excel, zum Einsatz kommen.

10. Aufgabe:³⁴

Berechnen Sie die Korrelationskoeffizienten für die fünf HO/AO Cluster für jedes der Über/ Unter X Tore Ergebnisse (Über/Unter 0,5 Tore, Über/Unter 1,5 Tore, etc.)

Stellen Sie die Korrelationskoeffizienten für das erste HO/AO Cluster (x-Achse) für die Unter/ Über 2,5 Tore in Streudiagrammen graphisch dar:

- ▶ Beobachtete Ergebnisse: Zahl der Tore
- ▶ Buchmacher Wettquoten für die Unter 2,5 Wetten
- ▶ Buchmacher Wettquoten für die Über 2,5 Wetten

Tipps zur Lösung:

Die Berechnung der Korrelationskoeffizienten erfolgt auf Grundlage der einzelnen Werte für jedes einzelne Spiel.

\bar{x} und \bar{y} sind die Mittelwerte aller 1.530 Spiele.

Die Buchmacher Wettquoten für die Über/Unter 2,5 Tore finden sich in dem von Ihnen bereits von football-data.co.uk heruntergeladenem Datensatz

Die Buchmacher Wettquoten müssen vor der graphischen Darstellung in deren implizierte Wahrscheinlichkeiten umgewandelt werden.

Fügen Sie in die Graphiken eine lineare ‚Trendlinie‘ hinzu und visualisieren Sie damit den Korrelationskoeffizienten.

³⁴ Sollten Sie Schwierigkeiten haben, diese Aufgabe auf Anhieb selbstständig zu lösen, bitte nicht verzweifeln! Wir arbeiten uns durch sie auf den nächsten Seiten, Schritt für Schritt, gemeinsam durch.

Interpretation des Korrelationskoeffizienten

Der Korrelationskoeffizient (KK) kann Werte zwischen -1 und +1 annehmen und für dessen Interpretation muss man sich einprägen, dass der Korrelationskoeffizient ein Maß für den linearen Zusammenhang ist.

Es existieren viele verschiedene Empfehlungen für die Interpretation des Korrelationskoeffizienten und zu den meistverwendeten zählen die Faustregeln von Cohen. Diese besagen:

- Wenn ein Zusammenhang zwischen zwei Dingen als sehr, sehr unwahrscheinlich galt, kann eine Korrelation von $KK = .10$ bereits als "merklich", $.30$ als "hoch" und $.50$ als "sehr hoch" betrachtet werden.
- Wenn jedoch eigentlich ein hoher oder nahezu perfekter Zusammenhang erwartet wird, müssen andere Maßstäbe angehalten werden.

Man könnte dann sagen, dass nur ein KK über $.90$ einen perfekten Zusammenhang anzeigt, und ein KK kleiner als $.50$ zum Nachdenken führen sollte, ob es da überhaupt einen Zusammenhang gibt.

Ein negativer Korrelationskoeffizient bedeutet eine negative (inverse) Korrelation. Bei einer negativen Korrelation steigt eine Variable während die andere fällt.

Beispielweise der Zusammenhang zwischen HO/AO Quotient und der Zahl der Tore: Je stärker das Auswärtsteam, desto kleiner die Wettquote für Über 2,5 Wetten (höhere Wahrscheinlichkeit von vielen Toren).

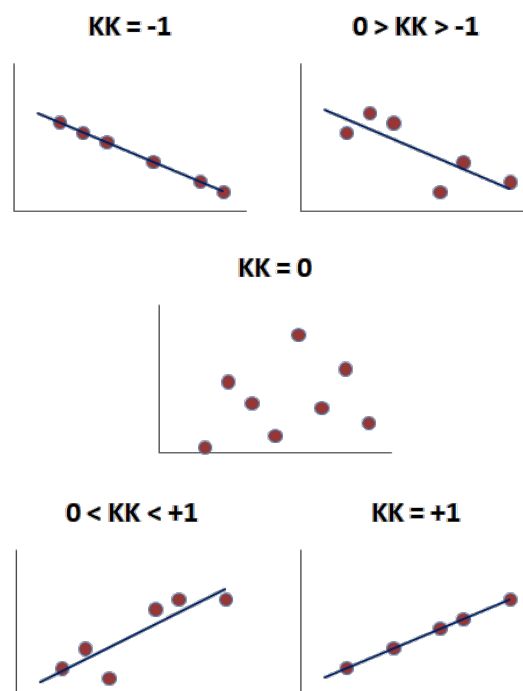


Abb. 16: Visualisierung der verschiedenen Werte, die ein Korrelationskoeffizient annehmen kann

Ist der Korrelationskoeffizient Null oder nahe Null, dann deutet das darauf hin, dass es keinen Zusammenhang gibt.

Und zuletzt...

Ein positiver Korrelationskoeffizient bedeutet eine positive Korrelation. Bei einer positiven Korrelation steigt eine Variable, wenn die andere auch steigt.

Ein weiteres Beispiel im Zusammenhang zwischen HO/AO Quotient und der Zahl der Tore: Je stärker das Heimteam, desto höher die Chance auf wenigstens ein Tor im Spiel (also eine sehr kleine Chance auf ein 0:0).

Zur genaueren Beurteilung der Korrelationskoeffizienten vom HO/AO Quotienten und Torergebnissen kommen wir später.

Korrelationskoeffizienten: HO/AO Cluster - Über/ Unter X Tore

Es wird ernst! Wir berechnen die Korrelationskoeffizienten und benutzen die Formel von Seite 47 (*ganz am Anfang des Kapitels 1.5.2. Empirischer Korrelationskoeffizient*).

Wir sind im Besitz aller Variablen (Download von footballdata.co.uk) und haben auch bereits die fünf Quantile für die 1.530 Spiele der Bundesliga (Clustergruppen) ausgerechnet (*Seite 45*).

Nochmals zur Erinnerung:

$i=1$	Dies ist der Startwert des Laufindexes <i>Jeweils die erste Variable des Datensatzes der jeweiligen Clustergruppe</i>
n	Dies ist der Endwert, bei welchem wir aufhören. <i>Jeweils die letzte Variable des Datensatzes der jeweiligen Clustergruppe</i>
x_i	x beschreibt die erste veränderliche Variable (<i>HO/AO Quotient</i>), und i ist der dazugehörige Laufindex (<i>wie in: 1. Spiel, 2. Spiel, ..., n. Spiel</i>)
\bar{x}	Arithmetisches Mittel (Mittelwert) der x Verteilung
y_i	y beschreibt die zweite veränderliche Variable (<i>z.B. Zahl der Unter/ Über Ergebnisse</i>), und i ist der dazugehörige Laufindex (<i>wie in: 1. Spiel, 2. Spiel, ..., n. Spiel</i>)
\bar{y}	Arithmetisches Mittel (Mittelwert) der y Verteilung

Bitte beachten Sie, dass wir im Vergleich zu Seite 47 (*Formel und Erklärung der mathematischen Symbole in 1.5.2. Empirischer Korrelationskoeffizient*) die Variablen x und y „umgedreht“ haben.

x ist unser HO/AO Quotient, da wir später, wenn wir Graphen produzieren, die HO/AO Werte in der x -Achse verwenden werden und y ist diejenige Variable, von welcher wir jeweils den Zusammenhang mit dem HO/AO Quotienten wissen wollen (*wie beispielsweise die Anzahl der ausgezählten Über/Unter Tore, Buchmacherwettquoten, etc.*).

$\bar{x} = 1,27$ (durchschnittlicher HO/AO Quotient von fünf Jahren Bundesliga)

$\bar{y} = 2,91$ (durchschnittliche Zahl der Tore in fünf Jahren Bundesliga)

Nochmals zur Erinnerung; dies ist die Formel für den Korrelationskoeffizienten:

$$Kor_e(x, y) := \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

Um die Formel in Excel zur Anwendung zu bringen und die untenstehende Tabelle (**Abb. 17**) zu replizieren wie auch die Berechnungen nachzuvollziehen, ist es empfehlenswert, sich auf Grundlage der footballdata.co.uk Daten eine Hilfstabelle zu erstellen.

In dieser werden ein paar Spalten für die spätere Verwendung in der Formel eingerichtet:

- Eine Spalte, in welcher die Clustergruppen nummeriert sind (1 für 1. Clustergruppe, 2 für 2. Clustergruppe, etc.)
- Weitere Spalten sind:
 - $(x_i - \bar{x})$
 - $(y_i - \bar{y})$
- Folgende 3 Spalten erleichtern das Nutzen der Formelfunktion in Excel (die Formel wird kürzer):
 - $(x_i - \bar{x}) \cdot (y_i - \bar{y})$
 - $(x_i - \bar{x})^2$
 - $(y_i - \bar{y})^2$

Hier ist die Excelformel, um die in **Abb. 17** (Seite 52) **blau** markierte Zelle (Korrelationskoeffizient in der ersten Clustergruppe für Unter 2,5 Tore) zu berechnen

```
=SUMPRODUCT((ArrayHOOAQuotienten=1)*(ArrayTorergebnis<=2),
ArraySpalte(x_i - x̄) • (y_i - ȳ ))/
(SQRT(SUMPRODUCT((ArrayHOOAQuotienten=1)*(ArrayTorergebnis<=2),
ArraySpalte(x_i - x̄)^2))*
SQRT(SUMPRODUCT((ArrayHOOAQuotienten=1)*(ArrayTorergebnis<=2),
ArraySpalte(y_i - ȳ )^2)))
```

```
= SUMMENPRODUKT((ArrayHOOAQuotienten=1)*(ArrayTorergebnis<=2),
ArraySpalte(x_i - x̄) • (y_i - ȳ ))/
(WURZEL(SUMMENPRODUKT((ArrayHOOAQuotienten=1)*(ArrayTorergebnis<=2),
ArraySpalte(x_i - x̄)^2))*
WURZEL(SUMMENPRODUKT((ArrayHOOAQuotienten=1)*(ArrayTorergebnis<=2),
ArraySpalte(y_i - ȳ )^2)))
```

Um die Formel in Aktion zu sehen, öffnen Sie die dem Kursbuch beiliegende Exceltabelle, und gehen da zur Registerkarte „HO+AO“.

In dieser finden Sie die **Abb. 17** mit allen Formeln. Dies sollte Ihr letzter Ausweg sein (*Bitte nicht schummeln! Erst selbst versuchen!*) und wird Ihnen auf jeden Fall helfen, Ihre eigenen Kalkulationen erfolgreich durchzuführen.

	Bundesliga		Korrelationskoeffizient		
	0.2918	0.5292	0.783	1.3172	23.94
U 0.5	0.999	0.996	0.994	0.856	(0.569)
U 1.5	0.975	0.975	0.976	0.856	(0.535)
U 2.5	0.897	0.902	0.894	0.797	(0.557)
U 3.5	0.663	0.729	0.707	0.652	(0.429)
U 4.5	0.292	0.488	0.424	0.453	(0.234)
U 5.5	0.022	0.309	0.236	0.242	(0.164)
U 6.5	(0.122)	0.198	0.114	0.136	(0.085)

Abb. 17: Korrelationskoeffizienten Bundesliga Unter X Tore

In **Abb. 17** kann man sehen, dass es einen starken linearen Zusammenhang gibt zwischen, beispielsweise, den Unter 2,5 Tore Ergebnissen und der ersten HO/AO Clustergruppe (*das Heimteam ist klarer Favorit*).

Dieser Zusammenhang wird immer schwächer, je stärker das Heimteam wahrgenommen wird und kippt in einen negativen Korrelationskoeffizienten in der fünften Clustergruppe (*das Heimteam ist schwächer als die Gastmannschaft*).

Ein positiver Korrelationskoeffizient bedeutet eine positive Korrelation. Bei einer positiven Korrelation steigt eine Variable, wenn die andere auch steigt.

Für die Unter 2,5 Tore Ergebnisse, beispielsweise, heißt das konkret:

Je höher der HO/AO Quotient (Stärke des Heimteams fällt im Vergleich zum Gast), desto mehr Unter 2,5 Tore Spiele sind zu beobachten.

Ein negativer Korrelationskoeffizient bedeutet eine negative (inverse) Korrelation. Bei einer negativen Korrelation steigt eine Variable während die andere fällt.

In der fünften Clustergruppe bei den Unter 2,5 Tore Ergebnissen, bedeutet der negative Korrelationskoeffizient:

Je höher der HO/AO Quotient (das Heimteam wird immer schwächer), desto weniger Unter 2,5 Tore Spiele sind zu beobachten.

Die Korrelationskoeffizienten für die Über X Tore sind erwartungsgemäß (*mehr oder weniger*) das Gegenteil der Unter X Tore Korrelationskoeffizienten:

	Bundesliga		Korrelationskoeffizient		
	0.2918	0.5292	0.783	1.3172	23.94
O 0.5	(0.283)	(0.032)	(0.117)	(0.119)	0.042
O 1.5	(0.468)	(0.318)	(0.343)	(0.362)	0.230
O 2.5	(0.707)	(0.678)	(0.684)	(0.660)	0.410
O 3.5	(0.861)	(0.868)	(0.863)	(0.822)	0.543
O 4.5	(0.931)	(0.945)	(0.943)	(0.886)	0.616
O 5.5	(0.960)	(0.971)	(0.973)	(0.936)	0.603
O 6.5	(0.974)	(0.986)	(0.990)	(0.970)	0.722

Abb. 18: Korrelationskoeffizienten Bundesliga Über X Tore

Abb. 18 zeigt die Korrelationskoeffizienten für die Über X Tore Ergebnisse.

Auch hier kann man sehen, dass es einen starken linearen Zusammenhang gibt zwischen, dieses Mal einen negativen, bei den Über 2,5 Tore Ergebnissen und der ersten HO/AO Clustergruppe (*das Heimteam ist klarer Favorit*).

Dieser negative Zusammenhang kippt in einen positiven Korrelationskoeffizienten in der fünften Clustergruppe (*das Heimteam ist schwächer als die Gastmannschaft*).

Ein negativer Korrelationskoeffizient bedeutet eine negative (inverse) Korrelation. Bei einer negativen Korrelation steigt eine Variable während die andere fällt.

In der ersten Clustergruppe bei den ‚Über 2,5 Tore‘ Ergebnissen, bedeutet der negative Korrelationskoeffizient:

Je höher der HO/AO Quotient (die Gastmannschaft wird immer stärker), desto weniger Über 2,5 Tore Spiele sind zu beobachten.

Ein positiver Korrelationskoeffizient bedeutet eine positive Korrelation. Bei einer positiven Korrelation steigt eine Variable, während die andere auch steigt.

Für die Unter 2,5 Tore Ergebnisse in der fünften Clustergruppe heißt das konkret:

Je höher der HO/AO Quotient (Stärke des Heimteams fällt im Vergleich zum Gast), desto mehr Über 2,5 Tore Spiele sind zu beobachten.

Wir führen uns diesen Zusammenhang im folgenden Kapitel graphisch vor Augen.

Streudiagramme und deren Interpretation

Die erste Graphik, welche wir erstellen, ist ein Streudiagramm, in welchem die einzelnen Werte (Torergebnisse der Spiele) eingetragen werden.

Die x-Achse zeigt die HO/AO Quotienten, die y-Achse die Zahl der Tore im jeweiligen Spiel.

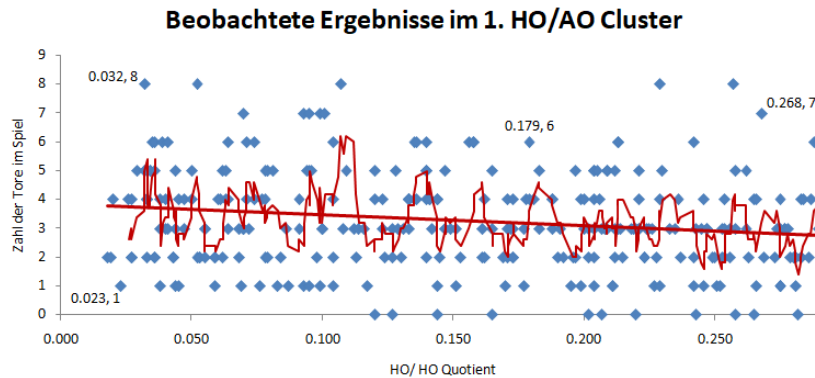


Abb. 19: Beobachtete Tor - Ergebnisse im 1. HO/AO Cluster (Streudiagramm)

Jeder blaue Punkt ist ein im Graphen eingetragener Datensatz. An einigen der Punkte sind Zahlen eingefügt, um vorzuführen wie das Streudiagramm erstellt ist.

Beispielsweise der Punkt links oben:

Das ist der Datensatz für HO/AO Quotienten: 0,032 und es gab 8 Tore im Spiel. Konkret war es das Spiel von Bayern München gegen Hamburg am 14.02.2015.

Der Datensatz unten links:

Das ist der Datensatz für HO/AO Quotienten: 0,023 und es gab 1 Tor im Spiel. Konkret war es das Spiel von Bayern München gegen Darmstadt am 06.05.2017.

11. Aufgabe:

Es gibt zwei weitere Punkte im Streudiagramm (**Abb. 19**), welche ein Label haben.

Suchen Sie die dazugehörigen Spiele in Ihrem Datensatz heraus und falls Sie es noch nicht gemacht haben, replizieren Sie das Streudiagramm selbst.

Tabellenkalkulationsprogramme wie Excel haben eine eingebaute Funktion, um Graphen zu erstellen. Experimentieren Sie solange rum, bis Ihnen ein ähnlicher Graph gelingt.

Schaut man sich die Trendlinie in **Abb. 19** an, kann man sehen, dass die Gesamtzahl der Tore in einem Spiel von durchschnittlich ca. 3,8 bei einem HO/AO Quotienten unter 0,05 auf durchschnittlich 2,8 am Ende des Clusters (HO/AO: 0,2918) gefallen ist.

Ohne die rote Trendlinie wäre es in der **Abb. 19** kaum erkennbar, dass eine Korrelation (lineare Abhängigkeit) von der steigenden Stärke des Heimteams und Zahl der Tore existiert.

Wir haben dies ja auch nur deswegen gewusst, da wir im vorhergehenden Kapitel die Korrelationskoeffizienten berechnet und deren Bedeutung analysiert hatten.

Um es sich visuell etwas besser vor Augen zu führen, kann man auch ein Streudiagramm erstellen, in welchem nur zwei Werte dargestellt werden:

1 für ein Ergebnis mit weniger als 2,5 Toren und

2 für ein Ergebnis mit mehr als 2,5 Toren.

Hierfür muss man in seiner Excel Tabelle eine Spalte hinzufügen mit folgender Formel:

=IF(I8<3,1,2) – oder - =WENN(I8<3,1,2)

Nutzt man dann diese neu erstellte Spalte als Grundlage für das Streudiagramm, erhält man folgende Graphik:

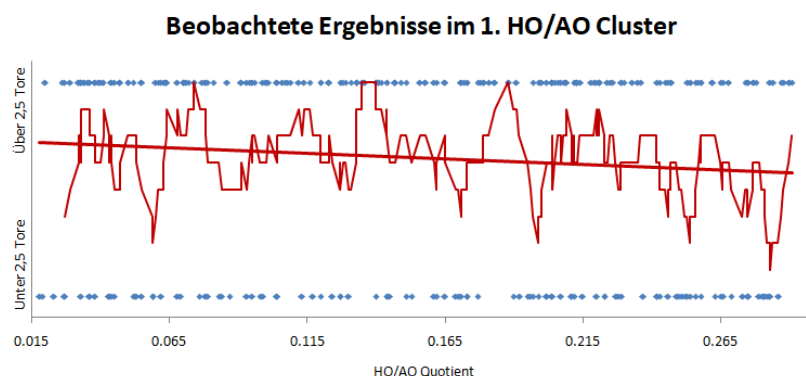


Abb. 20: Ja/Nein Streudiagramm Unter/Über X Tore Spiele – 1. HO/AO Clustergruppe

Bei genauen Anschauen der **Abb. 20**, kann man sehen, dass die Unter 2,5 Tore Ergebnisse auf der linken Seite (sehr kleine HO/AO Quotienten) spärlicher sind als auf der rechten Seite. Das heißt, dass die Zahl der Unter 2,5 Ergebnisse zunimmt, je stärker die Heimmannschaft wird.

Dies ist genau die Aussage, die uns der positive Korrelationskoeffizient gegeben hat (eine Variable steigt, während die andere auch steigt).

Je höher der HO/AO Quotient (Stärke des Heimteams fällt im Vergleich zum Gast), desto mehr Unter 2,5 Tore Spiele sind zu beobachten.

Dass die Buchmacher diese Korrelation beim Setzen ihrer Wettquoten berücksichtigen, zeigt das folgende Streudiagramm:

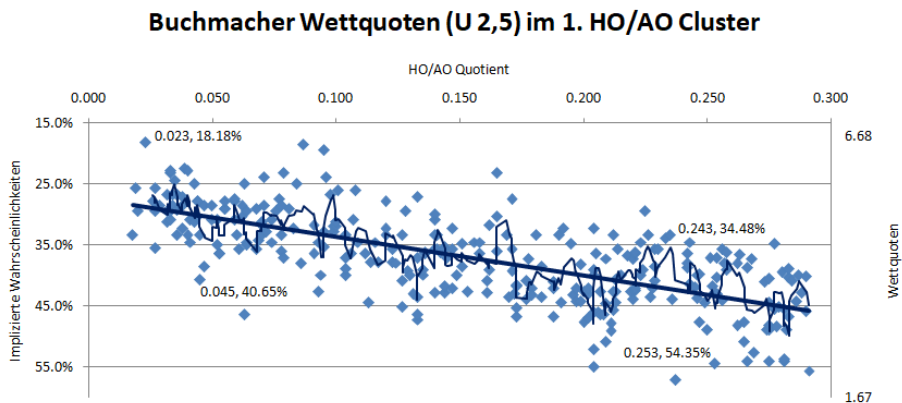


Abb. 21: Buchmacher Wettquoten für Unter 2,5 Tore – HO/AO Quotienten 1. Cluster

Die Neigung der Trendlinie in **Abb. 21** sieht deutlich stärker aus als bei den beobachteten Ergebnissen.

Grund hierfür ist, dass wir die Wettquoten in Form derer implizierten Wahrscheinlichkeiten graphisch darstellen während die **Abb. 20** die Trendlinie aufgrund simpler Ja/Nein Antworten zeigt – Gab es im Spiel über 2,5 Tore? Ja oder Nein.

Einige der blauen Punkte in **Abb. 21** enthalten ihre genauen Koordinaten (x und y). Wir werden diese Graphik nutzen, um ein Gefühl für Buchmacher Wettquoten und deren Spannweiten zu entwickeln.

Beispielsweise der Punkt links oben:

Das ist der Datensatz für HO/AO Quotienten 0,023 und die Buchmacher boten für dieses Spiel eine Wettquote für die Unter 2,5 Tore Wette für eine implizierte Wahrscheinlichkeit von 18,18% an.

Diese Wahrscheinlichkeit umgerechnet in deren korrespondierende Wettquote beträgt: 5,5.

Konkret war es das Spiel von Bayern München gegen Darmstadt am 6.5.2017.

Der Punkt links unten:

Das ist der Datensatz für HO/AO Quotienten 0,045 und die Buchmacher boten für dieses Spiel eine Wettquote für die Unter 2,5 Tore Wette für eine implizierte Wahrscheinlichkeit von 40,65% an.

Diese Wahrscheinlichkeit umgerechnet in deren korrespondierende Wettquote beträgt: 2,46.

Konkret war es das Spiel von Bayern München gegen Hertha Berlin am 25.04.2015.

Die Spannweite der Wettquoten am linken Ende des 1. HO/AO Clusters hat von einer 2,46 Wettquote (40,65% Wahrscheinlichkeit) bis zu einer 5,5 Wettquote (18,18% Wahrscheinlichkeit) gereicht.³⁵

12. Aufgabe:

Es gibt zwei weitere Punkte im Streudiagramm (**Abb. 21**), welche gekennzeichnet sind.

Berechnen Sie die dazugehörigen Wettquoten, suchen Sie dann die dazugehörigen Spiele in Ihrem Datensatz heraus und checken Sie Ihre Ergebnisse mit den Buchmacherwettquoten.

Welche Spannweite hatten die Unter 2,5 Wettquoten am rechten Ende des 1. HO/AO Clusters?

Falls Sie es noch nicht gemacht haben, replizieren Sie das Streudiagramm selbst.

Erstellen Sie auch andere Streudiagramme, für andere Clustergruppen. Rechnen Sie die Wettquoten an deren linken und rechten Enden der x-Achse aus.

Entwickeln Sie ein Gefühl für Wettquoten! Sie werden dieses später im Kurs gut gebrauchen können.

Der Korrelationskoeffizient wird später im Kurs wiederauftauchen, da dieser von großer Bedeutung ist, um Wettquoten zu kalkulieren und den Markt zu verstehen.

Jedoch an dieser Stelle sei erstmal genug zum Thema Kovarianz und Korrelation und wir gehen zum nächsten Kapitel über...

³⁵ Beide Spiele endeten übrigens jeweils mit nur 1 Tor. Jedoch bitte daraus keine voreiligen Schlüsse ziehen! Dies ist totaler Zufall.