STATISTIK

Example

Alternative Prüfungsleistung

Autor:	Josefine RICHEY
Matrikelnummer:	362984
Korrespondenz:	j.richey@stud.hs-wismar.de
Studiengang:	Bachelor Betriebswirtschaft, 4. Fachsemester
Betreuer:	Prof. Dr. rer. pol. Gerhard Müller,
	Frau Monika Augustyniak

Wismar, 19.6.2022

1 Univariate Analyse

Im folgenden Abschnitt wird das vorliegende Datenmaterial des durchschnittlichen Benzinpreises in Deutschland im Rahmen der univariaten Datenanalyse verdichtet, analysiert und ausgewertet.

1.1 Konkretisierung des Merkmals

Merkmal	durchschnittlicher Benzinpreis in Deutschland für 1 Liter Superbenzin
Merkmalsausprägung (im vorliegenden Datenmaterial)	Centbeträge im Bereich von 117,1 bis 197,0 Cent je Liter
Statistische Masse	Stichprobe von 37 Monaten (April, August, Dezember) im Zeitraum von April 2010 bis April 2022
Statistische Einheit	ein Monat (April, August oder Dezember)
Merkmalstyp	quasi-stetig
Skalenniveau	metrisch verhältnisskaliert

Abbildung 1: Übersicht der Merkmalsbeschreibung

Zur Konkretisierung bzw. näheren Erläuterung des Merkmals seien folgende Punkte ergänzt:

- Bei dem Merkmal Benzinpreis handelt es sich ausdrücklich nicht um einen konkreten Wert, sondern um den Durchschnittspreis für 1 Liter Superbenzin des betrachteten Monats in Deutschland. Im Folgenden wird aus Gründen der Lesbarkeit auf eine Erwähnung des durchschnittlichen Preises verzichtet und lediglich vom Benzinpreis gesprochen; gemeint ist aber der Durchschnittspreis, sofern nicht anders erwähnt.
- Die Stichproben wurden hier jeweils im Abstand von 4 Monaten gewählt; die Werte liegen deshalb für den Monat April, August oder Dezember vor. Eine entsprechende Spezifizierung der statistischen Einheit wurde in der obigen Tabelle vorgenommen.
- Die Merkmalsausprägungen können jenseits der Stichprobe selbstverständlich jeden Wert über Null annehmen, wurden aber aufgrund ihrer Vielfältigkeit für die Analyse auf die Spannweite der Stichprobenwerte begrenzt.

1.2 Verdichtung und Häufigkeitsverteilung

Bevor die Häufigkeitsverteilung des vorliegenden Datenmaterials untersucht werden kann, wird dieses verdichtet. Da eine Gruppierung der Daten aufgrund zahlreicher, unterschiedlicher Merkmalsausprägungen wenig sinnvoll ist, werden die Daten wie folgt klassiert:

y_j **Bpreis in ct** $\cdot$ l^{-1}	h_j	f_j	H_j	F_j	Δx_j	f^*_j
[110;130)	6	0,1622	6	0,1622	20	0,0081
[130;140)	8	0,2162	14	0,3784	10	0,0216
[140;150)	10	0,2703	24	0,6486	10	0,0270
[150;160)	9	0,2432	33	0,8919	10	0,0243
[160;200)	4	0,1081	37	1	40	0,0027
$\sum$	37	1				

Abbildung 2: klassierte Datenaufbereitung des durchschnittlichen Benzinpreises

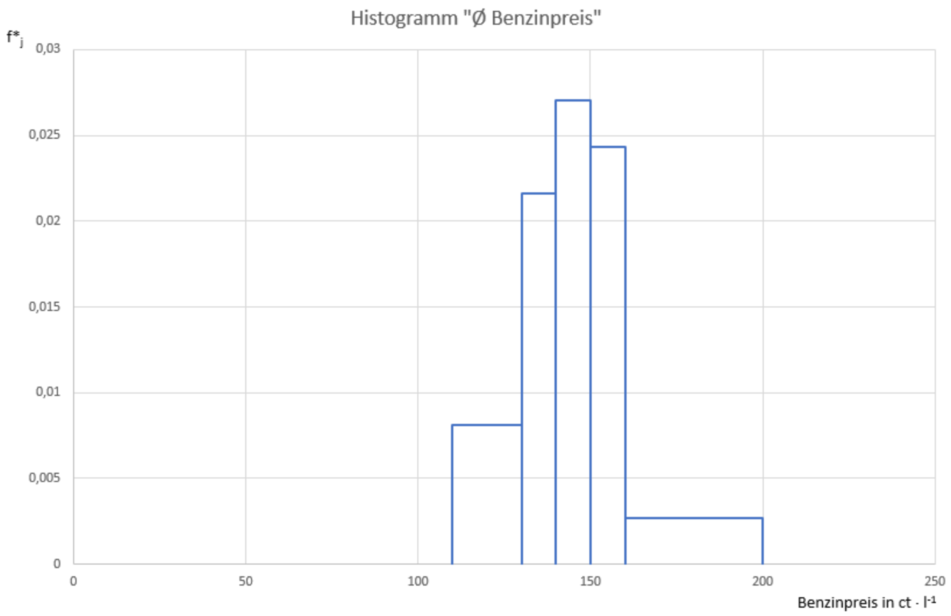

Abbildung 3: Histogramm ”Benzinpreis”

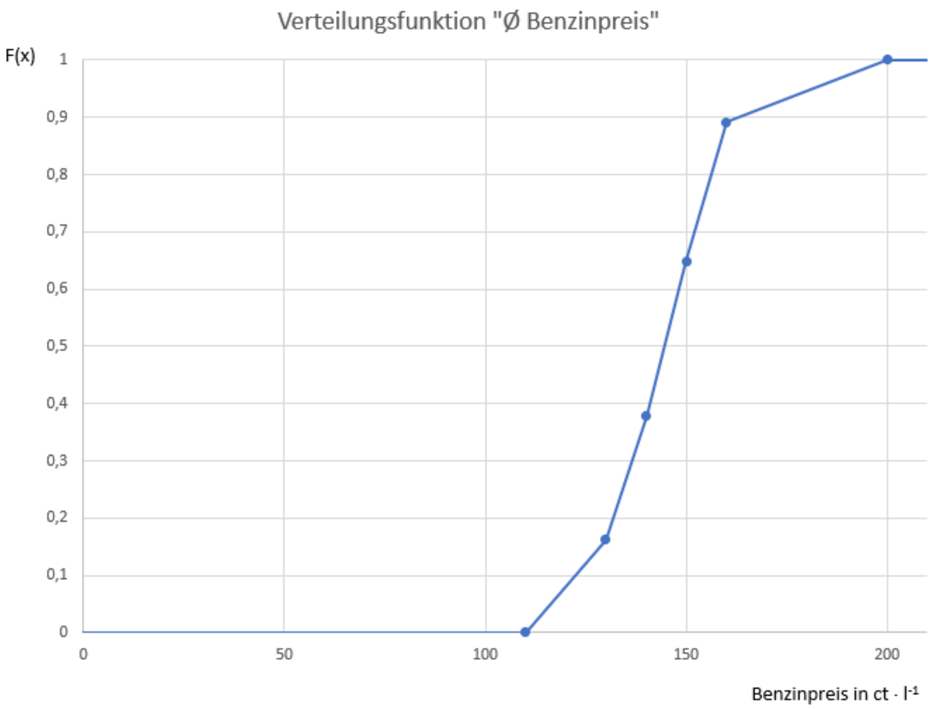

Abbildung 4: Verteilungsfunktion "Benzinpreis"

1.3 Lageparameter

Zur Bestimmung des **Modalwert** werden die Benzinpreis-Daten der Urliste gruppiert und ihre absolute Häufigkeit ermittelt.
Dabei lässt sich kein eindeutiger Modalwert bestimmen. Sowohl 141,2 ct · l^{-1} als auch 165,6 ct · l^{-1} treten mit einer absoluten Häufigkeit von 2 auf.
Der Modalwert verliert damit an Aussagekraft und wird nicht genauer betrachtet.

Um den **Median** (Zentralwert) zu ermitteln, werden die Benzinpreis-Daten der Urliste der Größe nach sortiert.
Da eine Stichprobe im Umfang von 37 Werten vorliegt, ist die Ermittlung ohne viel Aufwand möglich ((37+1)/2 = 19). Der in der Mitte liegende Wert (x_{19}) der Stichprobe ist der Median und beträgt 143,7 ct · l^{-1}.

Da das vorliegende Merkmal verhältnisskaliert und damit metrisch ist, kann auch das **arithmetische Mittel** (1) berechnet werden.
Dafür wird die Summe aller erfassten Werte gebildet und durch den Umfang der Stichprobe geteilt:

$$\bar{x} = \frac{1}{37} \cdot \sum_{i=1}^{37} x_i = 144{,}7\,\text{ct}\cdot\text{l}^{-1} \tag{1}$$

Dabei ergibt sich ein Mittelwert von 144,7 ct · l^{-1}. Durchschnittlich lag der Benzinpreis im Monat damit bei 144,7 ct · l^{-1}.

1.4 Streuungsparameter

In Vorbereitung auf die Analyse der Streuung, werden zunächst das erste und dritte **Quartil** (2 bis 9) sowie der zwischen ihnen bestehende **Quartilsabstand** (10) berechnet.

$$(n+1) \cdot p = (37+1) \cdot 0{,}25 = 9{,}5 \quad (2)$$
$$g(9{,}5) = 9;\ nka(9{,}5) = 0{,}5 \quad (3)$$

$$\begin{aligned} x_{0,25} &= (1-0{,}5) \cdot x_9 + 0{,}5 \cdot x_{10} && (4)\\ &= 0{,}5 \cdot 134{,}5\,\text{ct}\cdot\text{l}^{-1} + 0{,}5 \cdot 135{,}6\,\text{ct}\cdot\text{l}^{-1} && (5)\\ &= 135{,}050\,\text{ct}\cdot\text{l}^{-1} && (6)\end{aligned}$$

$$g(28{,}5) = 28;\ nka(28{,}5) = 0{,}5 \quad (7)$$
$$\begin{aligned} x_{0,75} &= (1-0{,}5) \cdot x_{28} + 0{,}5 \cdot x_{29} && (8)\\ &= 153{,}200\,\text{ct}\cdot\text{l}^{-1} && (9)\end{aligned}$$

$$Q = x_{0,75} - x_{0,25} = 153{,}200 - 135{,}050 = 18{,}150\,\text{ct}\cdot\text{l}^{-1} \quad (10)$$

Die mittleren 50% der untersuchten Monate unterscheiden sich demnach im Benzinpreis um maximal 18,150ct · l^{-1}.

Mithilfe des ermittelten Quartilsabstandes kann die vorliegende Datenreihe nun auf **Ausreißer** untersucht werden. Dafür wird der Quartilsabstand mit dem Faktor 1,5 mutlipliziert und Toleranzgrenzen für die vorliegenden Werte bestimmt:

$$Q \cdot 1{,}5 = 18{,}150\,\text{ct}\cdot\text{l}^{-1} \cdot 1{,}5 = 27{,}225 \quad (11)$$
$$x_{0,25}:\ 135{,}050 - 27{,}225 = 107{,}825\,\text{ct}\cdot\text{l}^{-1} \quad (12)$$
$$x_{0,75}:\ 153{,}200 + 27{,}255 = 180{,}425\,\text{ct}\cdot\text{l}^{-1} \quad (13)$$

Werte außerhalb des so definierten Rahmens lassen sich als Ausreißer einordnen; darunter fällt lediglich der Wert 197,0 ct · l^{-1}.
Dieser Ausreißer wurde auch in die Berechnung des arithmetischen Mittels einbezogen. Um zu prüfen, ob dieses durch den Ausreißerwert maßgeblich verzerrt

wurde, kann das arithmetische Mittel noch einmal nur für die mittleren 50 % berechnet werden. Dabei ergibt sich ein Wert von 144,1 ct · l^{-1}, welcher nur minimal unter dem zuvor berechneten Mittelwert liegt.
Der Ausreißer verzerrt damit das arithmetische Mittel leicht nach oben, was bei einer Differenz von 0,6 ct · l^{-1} durchaus vernachlässigbar scheint.

Aus den ermittelten Daten kann nun ein Boxplot-Diagramm erstellt werden:

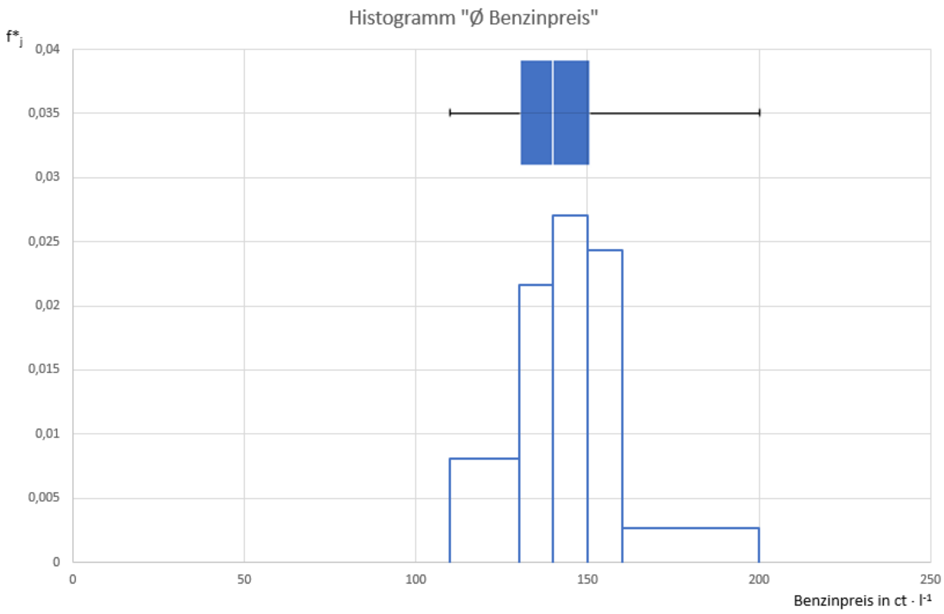

Abbildung 5: Histogramm ”Benzinpreisïnklusive Boxplot-Diagramm

Anhand des Boxplots wird die Verteilung der Werte noch einmal grafisch deutlich. Der Median innerhalb der Box liegt - wenn auch nur minimal - näher am unteren Quartil, die Box ist nicht mittig positioniert, sondern Richtung Minimum verschoben.Die Verteilung ist damit asymmetrisch: **linkssteil** bzw. rechtsschief.
Der Abstand der Merkmalsausprägungen links vom Median ist damit kleiner als der größerer Merkmalsausprägungen rechts vom Median. Zudem fällt das arithmetische Mittel größer als der Zentralwert aus, auch wenn die Differenz nicht sonderlich groß ist (vgl. Median = 143,7 ct · l^{-1}, arithmetisches Mittel = 144,7 ct · l^{-1} und Differenz von 1,0 ct · l^{-1}).

Die Verteilung und damit einhergehende Abweichung der Werte lässt sich noch genauer mit der **Varianz** (14-15) und **Standardabweichung** (16) untersuchen.

$$s_x^2 = \frac{37}{36} \cdot (\frac{782.266,33}{37} - 144,7^2)\,[\text{ct}\cdot\text{l}^{-1}]^2 \tag{14}$$
$$= 222,77697\,[\text{ct}\cdot\text{l}^{-1}]^2 \tag{15}$$

Die **Varianz** s_x^2 ist aufgrund ihrer quadratischen Dimension nur schwer mit den bisher ermittelten Werten vergleichbar und lässt keine Interpretation zu. Sie dient lediglich als Zwischenschritt zur Ermittlung der **Standardabweichung** s_x.

$$s_x = \sqrt{222,77697} = 14,926\,\text{ct}\cdot\text{l}^{-1} \tag{16}$$

Die Standardabweichung gewährt nun Einblick in die Streuung der Werte um den Mittelwert. Durchschnittlich weichen die Benzinpreise der betrachteten Monate in der Stichprobe um 14,926 ct · l^{-1} vom Mittelwert ab.
Das Skalenniveau des Merkmals erlaubt es darüber hinaus, den **Variationskoeffizienten** (17) zu berechnen und damit die Streuung über die Merkmalsgrenzen hinaus zur allgemeinen Vergleichbarkeit zu quantifizieren:

$$V = \frac{s_x}{\bar{x}} = \frac{14,926\,\text{ct}\cdot\text{l}^{-1}}{144,7\,\text{ct}\cdot\text{l}^{-1}} = 0,10318 \tag{17}$$

Prozentual gesehen unterliegen die Benzinpreise damit einer schwachen Streuung von 10,318 %.

1.5 Zwischenfazit univariate Datenanalyse

2 Bivariate Datenanalyse

Der nächste Abschnitt widmet sich der bivariaten Datenanalyse. Dabei wird nicht mehr nur unsere Y-Variable ”Benzinpreis” betrachtet, sondern zusammen mit weiteren Einflussgrößen analysiert, um mögliche Zusammenhänge zu erkennen. Ziel der Analyse ist es, Einflussgrößen zu identifizieren und Zusammenhänge offen zu legen, die in einem weiteren Schritt die Prognose von Werten der beeinflussten Größe ermöglichen.
Die zu betrachtenden Einflussgrößen sind BIP_D, EURUSD, Rohölpreis, Stimmung und OPEC und wurden analog zu den Benzinpreisen pro Monat im Zeitraum von April 2010 bis April 2022 stichprobenartig erfasst.
Im Folgenden werden diese Merkmale zunächst anhand ihres Skalenniveaus kategorisiert und dann nacheinander auf ihren linearen Zusammenhang mit der Y-Variablen untersucht. Ähnliche Vorgehensweisen in den Analyseverfahren werden aufgrund der Redundanz nur einmal ausführlich erklärt und danach mit Verweis analog wiederholt.

2.1 Einflussgröße I: Bruttoinlandsprodukt

BIP_D steht für das **Bruttoinlandsprodukt** in Deutschland und wird im vorliegenden Datenmaterial in MRD. € erfasst.
Es ist ein Indikator für Wohlstand, Wachstum und Leistungsfähigkeit einer Volkswirtschaft und misst die im gesamten Inland entstandene Wirtschaftsleistung innerhalb eines bestimmten Zeitraumes.[1] Da jedes Land im weitesten Sinne *irgendwas* produziert, kann das BIP allenfalls bei Null liegen, jedoch nicht negativ werden, und ist damit **metrisch verhältnisskaliert.** Zur Analyse des Zusammenhangs wird deshalb auf den Korrelationskoeffizienten zurückgegriffen.

[1] vgl. bwl-lexikon.de

2.1.1 Einleitende Analyse

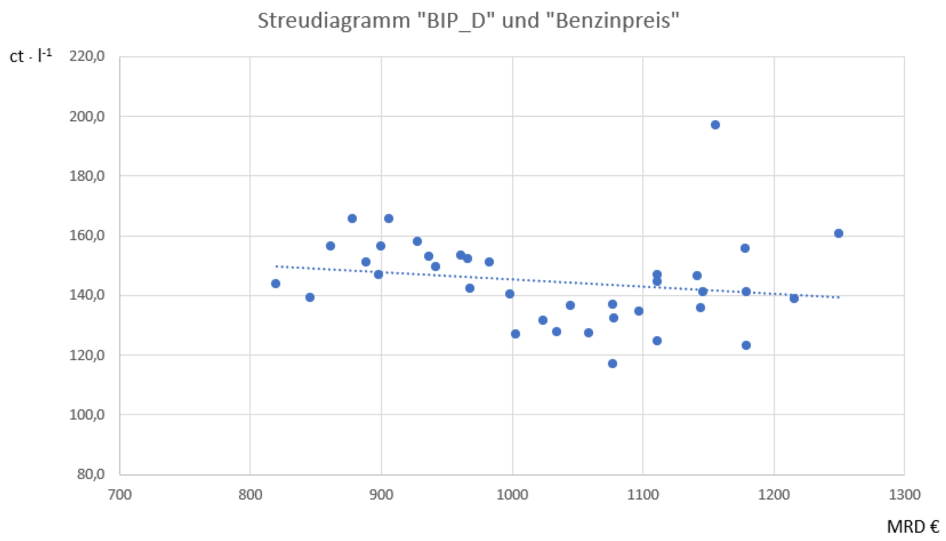

Abbildung 6: Streudiagramm "BIP_D" und "Benzinpreis"

Mithilfe der Trendlinie ist bereits ein schwacher negativer linearer Zusammenhang der beiden Größen vermutbar, da die Punkte recht weit um die Trendlinie herum streuen. Naheliegender scheint ein nicht-linearer Zusammenhang, der durch eine trigonometrische Funktion beschrieben werden kann. Dabei handelt es sich jedoch nur um eine Vermutung und soll hier lediglich als Idee aufgeführt sein.

2.1.2 Mathematische Analyse des linearen Zusammenhangs

Der Zusammenhang soll nun mithilfe des Korrelationskoeffizienten genauer untersucht werden.
Dafür wird in einem ersten Schritt unter Zuhilfenahme einer Arbeitstabelle[2] die **Kovarianz** beider Werte (), sowie die individuelle Standardabweichung der Merkmale () ermittelt.

$$s_{xy} = \frac{37}{36} \cdot \left(\frac{5.499.666{,}9}{37} - 144{,}7 \cdot 1.029{,}7297\right) = -326{,}5426\,\text{ct}\cdot\text{l}^{-1} \cdot \text{Mrd€} \quad (18)$$

Die Kovarianz gewährt bereits erste Einblicke und weist auf einen negativen linearen Zusammenhang hin. Weitere Aussagen sind an dieser Stelle nicht möglich,

[2]siehe Anhang Arbeitstabelle

da es sich bei der Kovarianz um eine nicht standardisierte Größe handelt.[3]

$$s_x^2 = \frac{37}{36} \cdot (\frac{39.711.148}{37} - 1.029{,}7297^2) = 13.290{,}1471\ [\text{Mrd€}]^2 \tag{19}$$

$$s_x = \sqrt{13.290{,}1471\ [\text{Mrd€}]^2} = 115{,}2829\ \text{Mrd€} \tag{20}$$

Für die Varianz $s_y{}^2$ und die Standardabweichung s_y (des Benzinpreises) können wir die Berechnungen (15) und (16) aus dem vorangegangenen Kapitel übernehmen.
Mit diesen Werten können wir nun den Korrelationskoeffizienten ermitteln:

$$r = \frac{s_{xy}}{s_x \cdot s_y} = \frac{-326{,}5426\ \text{ct}\cdot\text{l}^{-1} \cdot \text{Mrd€}}{115{,}2829\ \text{Mrd€} \cdot 14{,}926\ \text{ct}\cdot\text{l}^{-1}} \tag{21}$$

$$= -0{,}1898 \tag{22}$$

Der Korrelationskoeffizienten bestätigt die anfängliche Vermutung und zeigt, dass zwischen den untersuchten Merkmalen ein negativer, schwacher linearer Zusammenhang besteht. Der Benzinpreis fällt also tendentiell höher aus, je niedriger das Bruttoinlandsprodukt.

2.1.3 Test auf lineare Abhängigkeit

Es besteht jedoch die Möglichkeit, dass der zuvor ermittelte Zusammenhang lediglich zufällig im Rahmen der Stichprobe zustande kam. Aus diesem Grund muss geprüft werden, ob der lineare Zusammenhang auch für die (unbekannte) Grundgesamtheit gültig ist.
Dafür wird zunächst die Nullhypothese H_0 formuliert: ”Die Merkmale X (BIP_D) und Y (Benzinpreis) sind linear unabhängig.”

Um die Gültigkeit der Nullhypothese zu testen, wird eine Testgröße T benötigt, die mithilfe des Korrelationskoeffizienten berechnet werden kann:

$$T = \frac{r}{\sqrt{1-r^2}} \cdot \sqrt{n-2} \quad = \frac{-0{,}1898}{\sqrt{1-(-0{,}1898)^2}} \cdot \sqrt{37-2} = -1{,}1435 \tag{23}$$

Im Folgenden wird die Signifikanzwahrscheinlichkeit einer t-verteilten Zufallsvariablen mit df = 37-2 = 35 Freiheitsgraden $X_{t;35}$ mithilfe der Excelfunktion T.VERT bestimmt. Etwas trivialer formuliert könnte man sagen: Es wird geprüft, wie hoch die Wahrscheinlichkeit dafür ist, dass eine Zufallsvariable (t-verteilt, mit df - Freiheitsgraden) den in der Stichprobe ermittelten Wert der Testgröße mindestens

[3] vgl. Statistikskript 2022 S. 116

erreicht. Dafür wird die ermittlete Wahrscheinlichkeit mit dem festgelegten Signifikanzniveau von $\frac{\alpha}{2}$ = 0,025 bzw. 2,5% verglichen.

$$P(X_{t;35} \leq T) = 1 - \text{T.VERT}(-1{,}1435; 35; \text{wahr}) \tag{24}$$
$$= 0{,}86971 \tag{25}$$

Da die ermittelte Signifikanzwahrscheinlichkeit mit 86,971% weit über dem Signifikanznivea liegt, kann die Nullhypothese nicht abgelehnt werden. Demnach besteht außerhalb der Stichprobe kein signifikanter linearer Zusammenhang zwischen dem Bruttoinlandsprodukt und dem Benzinpreis.

2.2 Einflussgröße II: Wechselkurs Euro $\rightarrow$ USD

EURUSD steht für den **Wechselkurs von EURO in USD**. Zwar wir im vorliegenden Datenmaterial keine eindeutige Aussage über die Einheit getroffen, rein logisch müssten die Werte jedoch in USD · €$^{-1}$ vorliegen. Aufgrund der Erfassung wird zudem von einem durchschnittlichen Wechselkurs für den betrachteten Monat ausgegangen.
Der Wechselkurs stellt das Verhältnis der Werte von einer Währungseinheit zu einer anderen dar. Da ein Verhältnis nicht negativ bzw. kleiner 0 werden kann und die Abstände zwischen den Werten gleichmäßig und messbar sind, liegt ein **metrisch verhältnisskaliertes Merkmal** vor. Die Analyse des Zusammenhangs erfolgt deshalb mithilfe des Korrelationskoeffizienten.

2.2.1 Einleitende Analyse

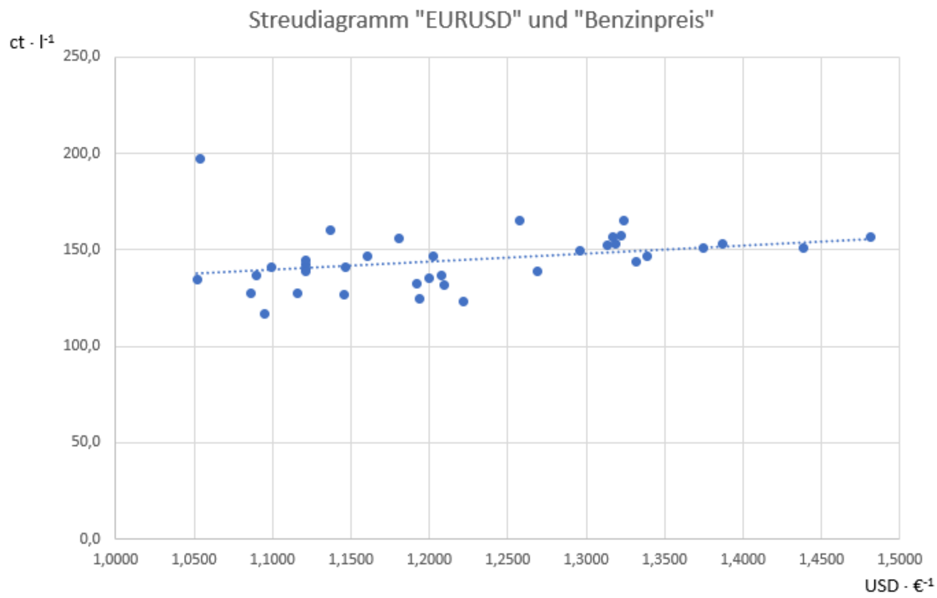

Abbildung 7: Streudiagramm ËURUSD” und ”Benzinpreis”

Anhand des Streudiagramms lässt sich bereits ein positiver linearer Zusammenhang vermuten, der aufgrund der Streuung um die Trendlinie tendentiell schwach bis mittelmäßig ausfallen wird.

2.2.2 Mathematische Analyse des linearen Zusammenhangs

Wie im Kapitel zuvor, soll nun der Zusammenhang mathematisch analysiert werden. Dafür werden die Kovarianz[4] und - ergänzend zu dem Wert der Standardabweichung für Merkmal Y - die Standardabweichung für Merkmal X[5] berechnet. Die ermittelten Werte dienen dann der Berechnung des Korrelationskoeffizienten[6].

$$s_{xy} = 0{,}5084\,\text{ct·l}^{-1} \cdot \text{USD·€}^{-1} \tag{26}$$
$$s_x^2 = 0{,}0125\,[\text{USD·€}^{-1}]^2 \tag{27}$$
$$s_x = 0{,}1116\,\text{USD·€}^{-1} \tag{28}$$
$$r = 0{,}3052 \tag{29}$$

[4]vgl. Berechnung 18 Kapitel 2.1.2
[5]vgl. Berechnung 19/20 Kapitel 2.1.2
[6]vgl. Berechnung 21/22 Kapitel 2.1.2

Mit einem Korrelationskoeffizienten von 0,3052 ist von einem mittelmäßigen, positiven linearen Zusammenhang zwischen den Merkmalen X (EURUSD) und Merkmal Y (Benzinpreis) auszugehen. Tendentiell steigt der Benzinpreis also mit steigendem Wechselkurs.

2.2.3 Test auf lineare Abhängigkeit

Die Nullhypothese H_0 für den vorliegenden Sachverhalt lautet: "Die Merkmale X (EURUSD) und Y (Benzinpreis) sind linear unabhängig."
Für den Test der Signifikanz des Korrelationskoeffizienten, wird nun die Testgröße ermittelt und die Signifikanzwahrscheinlichkeit bestimmt:

$$T = 1{,}8957 \tag{30}$$
$$P(X_{t;35} \geq T) = 1 - \text{T.VERT}(1{,}8957; 35; \text{wahr}) \tag{31}$$
$$= 0{,}0331 \tag{32}$$

Die Signifikanzwahrscheinlichkeit liegt damit knapp über dem halbierten Signifikanzniveau von 0,025. Die Nullhypothese kann damit nicht abgelehnt werden; es besteht kein signifikanter linearer Zusammenhang zwischen dem Wechselkurs EURUSD und dem Benzinpreis.

2.3 Einflussgröße III: Rohölpreis

Der vorliegende **Rohölpreis** beschränkt sich auf die Ölsorte Brent. Das ist wenig verwunderlich, denn Brent ist inzwischen die Referenzsorte für den Weltmarkt und damit repräsentativer Indikator für globale Ölpreise.[7] Sie werden in USD $\cdot$ bbl^{-1} angegeben.
Da die Preise für Öl den Benzinpreisen in ihrer Art sehr ähnlich sind, lassen sie sich als **metrisch verhältnisskalierten Merkmal** definieren; der Korrelationskoeffizient dient auch hier als Instrument zur Analyse des Zusammenhangs. Leider sind keine Aussagen dazu getroffen, ob es sich um einen Durchschnittswert oder konkreten Wert zum Ende des betrachteten Monats handelt. Da der Benzinpreis ebenfalls als durchschnittlicher Preis erfasst wurde, wird der Vollständigkeit halber an dieser Stelle ebenfalls von einem durchschnittlichen Rohölpreis ausgegangen.

[7] vgl. Kimani, o.S.

2.3.1 Einleitende Analyse

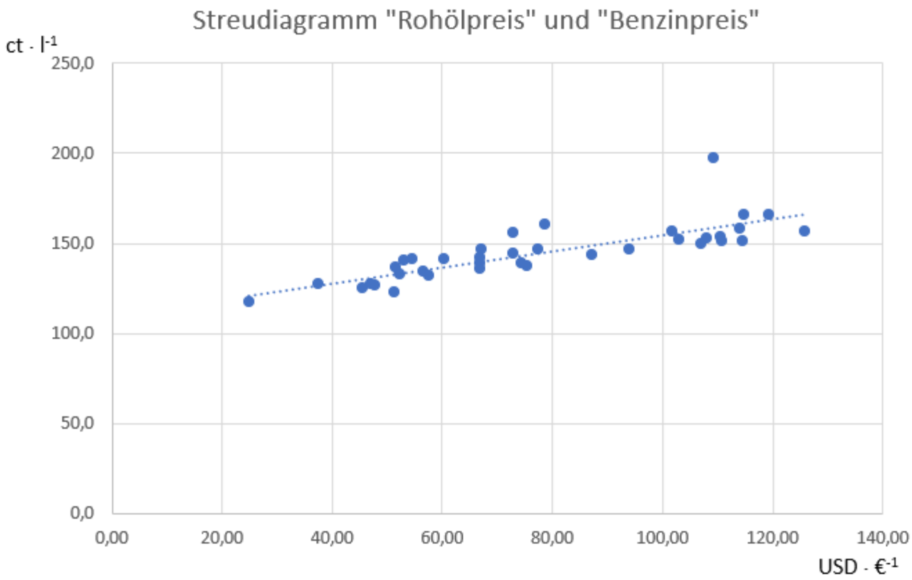

Abbildung 8: Streudiagramm ”Rohölpreis” und ”Benzinpreis”

Im Kontrast zu den vorangegangenen Merkmalen lässt sich hier ein mittelmäßiger bis starker positiver linarer Zusammenhang vermuten. Die Punkte liegen nah bei der eingezeichneten Trendlinie und streuen bis auf wenige Ausnahmen nur schwach um diese herum.

2.3.2 Mathematische Analyse des linearen Zusammenhangs

Wie bei den Merkmalsanalysen zuvor, soll auch hier die Vermutung mathematisch untersucht werden. Dafür werden wieder Kovarianz und Standardabweichung des neuen Merkmals X (Rohölpreis) ermittelt und aus ihnen sowie den bereits vorhandenen Werten für Merkmal Y (Benzinpreis) der Korrelationskoeffizient berechnet.

$$s_{xy} = 327{,}8951 \,\text{ct·l}^{-1} \cdot \text{USD·bbl}^{-1} \tag{33}$$

$$s_x^2 = 743{,}2488 \,[\text{USD·bbl}^{-1}]^2 \tag{34}$$

$$s_x = 27{,}2626 \,\text{USD·bbl}^{-1} \tag{35}$$

$$r = 0{,}8058 \tag{36}$$

Der Korrelationskoeffizient beträgt in diesem Fall 0,806 und weist damit auf einen starken linearen positiven Zusammenhang zwischen den beiden betrachteten Merkmalen hin. Steigt der Preis für Rohöl, wird höchst wahrscheinlich auch der Preis für Benzin steigen.

2.3.3 Test auf lineare Abhängigkeit

Im Zuge des Tests auf lineare Abhängigkeit wird folgende Nullhypothese H_0 formuliert: "Die Merkmale X (Rohölpreis) und Y (Benzinpreis) sind linear unabhängig."
Für den Test der Signifikanz des Korrelationskoeffizienten wird nun die Testgröße ermittelt und die Signifikanzwahrscheinlichkeit bestimmt:

$$T = 8{,}0504 \tag{37}$$
$$P(X_{t;35} \geq T) = 1 - \text{T.VERT}(8{,}0504; 35; \text{wahr}) \tag{38}$$
$$= 0{,}00000... \tag{39}$$

Die Signifikanzwahrscheinlichkeit fällt so gering aus, dass es aus Gründen der Verständlichkeit anschaulicher erscheint, das Ergebnis der T.VERT-Funktion zu zeigen. Mit 0,99999... fällt dieses erstaunlich hoch aus; die Signifikanzwahrscheinlichkeit liegt damit weit unter dem Signifikanzniveau und führt damit zur Ablehnung der Nullhypothese.
Der positive lineare Zusammenhang zwischen Rohölpreis und Benzinpreis ist damit signifikant.

2.4 Einflussgröße IV: Stimmungsindikator

Der **Stimmungsindikator** wird im Rahmen des vorliegenden Materials als Stimmungsindikator am Markt für Rohöl definiert. Stimmungsindikatoren sind - vereinfacht gesagt - Verhältniszahlen, die Aufschluss über Trends, Vermögenswerte und die Wirtschaft aus einer entsprechenden Perspektive geben.[8] Für den Rohölmarkt wird dieser Indikator mit einem Wert = 1 als neutral, einem Wert < 1 als negativ und einem Wert > 1 als positiv beschrieben.
Per se funktionieren Stimmungsindikatoren wie Barometer, wobei sie nie einen Wert unter 0 erreichen, der Wertebereich jedoch nach oben offen bleibt. Dabei ist die Aussagekraft der Werte unterschiedlich und die Bedeutung ihrer Abstände nicht eindeutig bestimmbar. Aus diesem Grund handelt es sich um ein quasi-stetig **ordinal skaliertes Merkmal**, für die folgende Analyse wird der **Rangkorrelationskoeffizient** verwendet.

[8] vgl. bienngoccruise.com

2.4.1 Einleitende Analyse

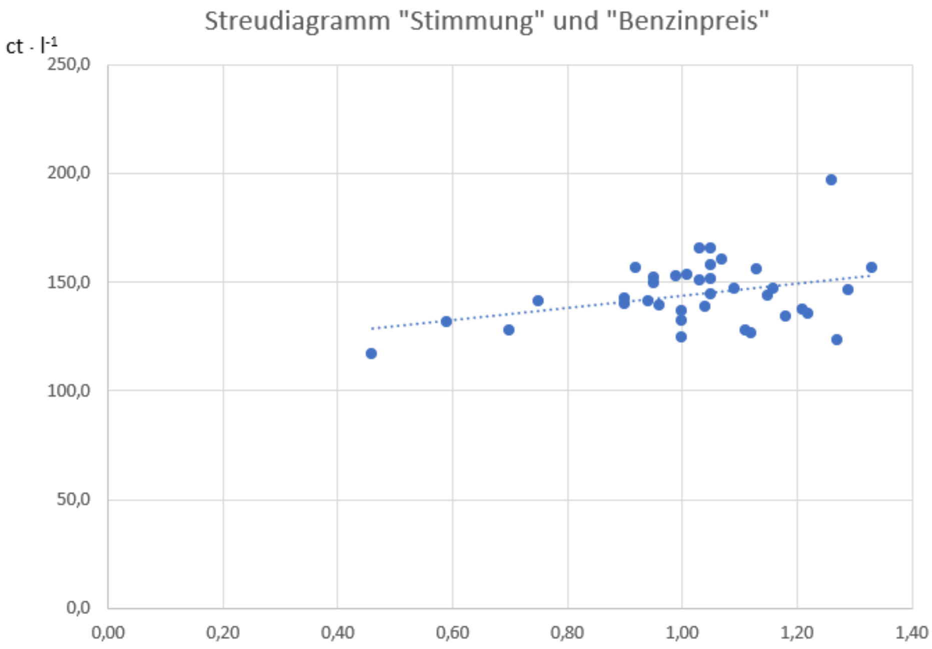

Abbildung 9: Streudiagramm SStimmung” und ”Benzinpreis”

Das Streudiagramm lässt bereits einen schwachen positiven linearen Zusammenhang vermuten. Obgleich die Werte zum den niedrigen Stimmungsindikatorwert nahe bei der Trendlinie liegen, streuen sie doch immer extremer, je positiver der Wert wird. Der konkrete Zusammenhang soll im Weiteren wieder mathematisch analysiert werden.

2.4.2 Mathematische Analyse des monotonen Zusammenhangs

Aufgrund der unterschiedlichen Skalenniveaus ist die Berechnung von Kovarianz und Korrelationskoeffizienten hier nicht ohne weiteres möglich.
Stattdessen werden die Werte der Stimmung und des Benzinpreises mithilfe einer Arbeitstabelle[9] ihren Werten entsprechend aufsteigend mit Rängen versehen. Die Ränge können nun wie ein metrisches Merkmal behandelt und verrechnet werden.
Analog zu den vorherigen Kapiteln wird aus der Summe der Rangprodukte die Kovarianz errechnet, sowie die Standardabweichungen der individuellen Merkmalsränge. Die vorliegenden Werte können dann zur Berechnung des Rangkorrelationskoeffizienten genutzt werden:

[9] siehe Anhang Arbeitstabelle

$$s_{R(x)R(y)} = \frac{37}{36} \cdot (\frac{14.082{,}5}{37} - 19 \cdot 19) = 20{,}1528 \tag{40}$$

$$s^2_{R(x)} = \frac{37}{36} \cdot (\frac{17.566{,}5}{37} - 19^2) = 116{,}9306 \Rightarrow s_{R(x)} \quad = 10{,}8134 \tag{41}$$

$$s^2_{R(y)} = \frac{37}{36} \cdot (\frac{17.574}{37} - 19^2) = 117{,}1389 \Rightarrow s_{R(y)} \quad = 10{,}8231 \tag{42}$$

$$R = \frac{s_{R(x)R(y)}}{s_{R(x)} \cdot s_{R(y)}} = \frac{20{,}1528}{10{,}8134 \cdot 10{,}8231} = 0{,}1722 \tag{43}$$

Mit einem Wert von 0,1722 deutet der Rangkorrelationskoeffizient auf einen schwachen positiven monotonen Zusammenhang der gebildeten Rangzahlen hin. Steigt die Rangzahl von X, erhöht sich tendentiell auch die Rangzahl von Y. Berücksichtigt man die Transformation und damit die ursprünglichen Merkmale, bedeutet dies, dass tendentiell der Benzinpreis steigt, wenn die Stimmung am Rohölmarkt steigt bzw. positiver wird.
Um eine so allgemeine Aussage tatsächlich treffen zu können, ist es jedoch unabdingbar, das Ergebnis auf Signifikanz zu testen.

2.4.3 Test auf monotonen Zusammenhang

Da der vorliegende Rangkorrelationskoeffizient in der angewandten Form dem Korrelationskoeffizienten entspricht, kann auch das gleiche Testverfahren angewandt werden.
Die Nullhypothese H_0 lautet: "Die Rangzahlen der Merkmale X (Stimmungsindikator) und Y (Benzinpreis) sind monoton unabhängig."Dafür wird die Testgröße mithilfe des Rangkorrelationskoeffizienten bestimmt und die Signifikanzwahrscheinlichkeit berechnet.

$$T = \frac{R}{\sqrt{1-R^2}} \cdot \sqrt{n-2} = \frac{0{,}1722}{\sqrt{1-0{,}1722^2}} \cdot \sqrt{37-2} = 1{,}0342 \tag{44}$$

$$P(X_{t;35} \geq T) = 1 - \text{T.VERT}(1{,}0342; 35; \text{wahr}) \tag{45}$$

$$= 0{,}1541 \tag{46}$$

Die Signifikanzwahrscheinlichkeit von 0,1541 liegt über dem Signifikanzniveau, die Nullhypothese kann nicht abgelehnt werden. Damit liegt kein signifikanter monotoner Zusammenhang vor.

2.5 Einflussgröße V: Entscheidungsverhalten der OPEC

OPEC ist eine dauerhafte, zwischenstaatliche, internationale Organisation von 13 erdölexportierenden Mitgliedsstaaten mit Sitz in Wien, deren Ziel die Stabilisierung des Ölpreises auf dem Weltmarkt zum Wohle aller beteiligten ist. Dafür wird

die Erdölförderung künstlich verknappt oder gesteigert, um so den Preis zu stabilisieren, zu drücken oder anzuheben.[10]
Die hier erfassten Daten beziehen sich auf die Verhandlungen der OPEC über konzertierte Aktionen im jeweiligen Zeitraum und ob eine Entscheidung getroffen wurde oder nicht. Die Ausprägungen des Merkmals sind dabei nicht nach Wert sortierbar und können lediglich als gleich oder ungleich unterschieden werden. Damit liegt ein diskretes **nominal skaliertes Merkmal** vor, für dessen Analyse sich die Maßzahl von Cramèr eignet.

2.5.1 Einleitende Analyse

Da für die folgende Analyse mindestens ein nominal skaliertes Merkmal vorliegt, erübrigt sich die Erstellung eines Streudiagramms aus Gründen der Unmöglichkeit. Stattdessen kann eine zweidimensionale Häufigkeitstabelle mit den klassierten Daten des Benzinpreises als Merkmal Y, und den Ausprägungen der OPEC als Merkmal X angefertigt werden.

Merkmal Y	**[110;130)**	**[130;140)**	**[140;150)**	**[150;160)**	**[160;200)**	SUMME
Merkmal X						
E	3	4	5	4	2	18
kE	3	4	5	5	2	19
SUMME	6	8	10	9	4	**37**

Abbildung 10: zweidimensionale Häufigkeitstabelle der Merkmale OPEC und Benzinpreis

Für die Zellen der Häufigkeitstabelle, welche als Grundlage zur Berechnung der Maßzahl von Cramèr genutzt wird, gilt die Bedingung:

$$\frac{h_{j.} \cdot h_{.k}}{n} = e_{jk} \geq 5 \tag{47}$$

Diese Bedigung wird aber nur von einigen wenigen Zellen der Häufigkeitstabelle erfüllt. Um die Aussagekraft der Maßzahl von Cramèr nicht einzuschränken, wird die Häufigkeitstabelle noch einmal verdichtet.

[10]vgl. opec.org

Merkmal Y	**[110; 140)**	**[140;150)**	**[150;200)**	SUMME
Merkmal X				
E	7	5	6	18
kE	7	5	7	19
SUMME	14	10	13	**37**

Abbildung 11: verdichtete zweidimensionale Häufigkeitstabelle

Um einer ersten Einschätzung über die statistische Abhängigkeit näher zu kommen, werden die bedingten relativen Häufigkeiten wie folgt berechnet:

$$f(y_k|x_j) = \frac{h_{jk}}{h_{j.}} \tag{48}$$

$$f(y_1|x_1) = \frac{7}{18} = 0{,}3889 \tag{49}$$

Merkmal Y	**[110; 140)**	**[140;150)**	**[150;200)**	SUMME
Merkmal X				
E	0,3889	0,2778	0,3333	1,0000
kE	0,3684	0,2632	0,3684	1,0000
SUMME	0,3784	0,2703	0,3514	**1,0000**
DIFFERENZ	*0,0205*	*0,0146*	*-0,0351*	

Abbildung 12: zweidimensionale Häufigkeitstabelle mit bedingten relativen Häufigkeiten

Die so ermittelten, bedingten relativen Häufigkeiten können spaltenweise verglichen werden.
Da sie nicht übereinstimmen, liegt auf alle Fälle eine statistische Abhängigkeit vor. Die maximale Differenz von 0,0351 (Spalte 3) lässt dabei eine sehr schwache Abhängigkeit vermuten. Für eine genauere Aussage genügt jedoch nicht nur die Betrachtung einer Spalte der Häufigkeitstabelle, sondern das Einbeziehen aller Werte oder genauer: die Maßzahl von Cramèr.

2.5.2 Mathematische Analyse der statistischen Abhängigkeit

Um die statistische Abhängigkeit zu prüfen, muss zunächst χ^2 berechnet werden.

$$\chi^2 = n \cdot [(\sum_{j=1}^{m}\sum_{k=1}^{l} \frac{h_{jk}^2}{h_{j.} \cdot h_{.k}}) - 1] \tag{50}$$

$$= 37 \cdot [(\frac{7^2}{14 \cdot 18} + \frac{5^2}{10 \cdot 18} + ... + \frac{7^2}{13 \cdot 19}) - 1] \tag{51}$$

$$= 37 \cdot (1{,}0013495 - 1) = 0{,}0499 \tag{52}$$

Die Maßzahl von Cramèr lässt sich mit dem vorliegenden Wert wie folgt ermitteln:

$$C = \sqrt{\frac{1}{n} \cdot \frac{\chi^2}{\text{min}((m-1);(l-1))}} \tag{53}$$

$$= \sqrt{\frac{1}{37} \cdot \frac{0{,}0499}{\text{min}(2-1)(3-1)}} \tag{54}$$

$$= \sqrt{\frac{0{,}0499}{37}} = 0{,}0367 \tag{55}$$

Mit einem Wert von 0,0367 bestätigt die Maßzahl von Cramèr damit die eingangs aufgestellte Vermutung: Es besteht eine schwache statistische Abhängigkeit zwischen dem Entscheidungsverhalten der OPEC und den Benzinpreisen. Der Benzinpreis wird vom Entscheidungsverhalten der OPEC kaum beeinflusst.

2.5.3 Test auf statistische Abhängigkeit

Für den Test auf statistische Abhängigkeit existiert auch für diese Maßzahl ein Verfahren mithilfe einer Testgröße. Im Gegensatz zu den vorangegangenen Berechnungen wird hier χ^2 selbst als Testgröße verwendet und in die Excel-Funktion CHIQU.VERT eingesetzt.

$$df = (m-1) \cdot (l-1) = (2-1) \cdot (3-1) = 1 \cdot 2 = 2 \tag{56}$$

$$P({X^2}_2 \geq T) = 1 - \text{CHIQU.VERT}(0{,}0499; 2; \text{wahr}) \tag{57}$$

$$= 0{,}9753 \tag{58}$$

Der sich daraus ergebende Wert beträgt 0,9753 und führt zur Ablehnung der Nullhypothese. Von einer signifikanten Abhängigkeit zwischen den Merkmalen Y (Benzinpreis) und X (OPEC) kann daher nicht ausgegangen werden.

2.6 Zwischenergebnis der bivariaten Datenanalyse