



Sommersemester 2021

Fakultät Wirtschaftswissenschaften
Volkswirtschaftslehre & Statistik
International Graduate Center
☎: (0421) 5905-4691

Prof. Dr. Peter Schmidt

Peter@Schmidt-bremen.de
www.schmidt-bremen.de → mkm

Modul 8 Unit 6:

Statistik im Kulturbetrieb

Grundlagen der beschreibende Statistik

mkm

Masterstudiengang **Kulturmanagement**



»Sollen wir das arithmetische Mittel als durchschnittliche Körpergröße nehmen und den Gegner erschrecken, oder wollen wir ihn einlullen und nehmen den Median?«

Zielsetzungen

- Angewandte Fragestellungen des Kulturbetriebs bearbeiten
 - Praxisorientierte Kennzahlen ermitteln
 - Wissenschaftliche Fragestellungen und Methoden
 - Vermittlung statistisch-methodischen Wissens
 - Gewinnung praktischer Erfahrung in der EDV-Umsetzung
 - Außendarstellung durch Dokumentation der verwendeten Methoden
- ⇒ Hilfe bei der täglichen Arbeit.

Inhalt

1	Quantitative Methoden; Statistik.....	3
1.1	Das „Forum“	4
1.2	Definitionen – Was ist Statistik?	4
1.3	Häufigkeiten und grafische Darstellungen	7
1.3.1	Eindimensionale und mehrdimensionale Häufigkeitsdarstellungen	7
1.3.2	Besondere Häufigkeitskonzepte	10
1.4	Lagemaße (Mittelwerte) und Streuungsmaße	11
1.5	Zusammenhänge zwischen mehreren Merkmalen.....	13
1.5.1	Zusammenhangsmaße	14
1.5.2	Regressionsanalyse.....	16
1.6	Zeitreihen und Indexzahlen.....	19
1.6.1	Zeitreihen.....	19
1.6.2	Indexzahlen.....	20
1.6.3	Praxiswissen.....	23
2	Literaturhinweise und Weitere Informationen.....	24
3	Schlagwortindex.....	25

Abbildungen

Abbildung 1	Häufigkeitsdiagramm	6
Abbildung 2	Säulendiagramm für Mittelwerte	6
Abbildung 3	Skalierung von Merkmalen	7
Abbildung 4	Absolute Häufigkeiten und zweidimensionales Säulendiagramm	9
Abbildung 5	Kreisdiagramme zur Darstellung von relativen Häufigkeiten	9
Abbildung 6	Einfache Häufigkeiten und Summenhäufigkeiten.....	10
Abbildung 7	Verteilungen mit unterschiedlicher Streuung.....	12
Abbildung 8	Abweichungen der Einzelbeobachtungen vom Mittelwert	13
Abbildung 9	xy-Diagramm für Zusammenhang Werbeausgaben und Besucherzahl	14
Abbildung 10	Stärke von Zusammenhängen und Werte des Korrelationskoeffizienten	15
Abbildung 11	Regressionsgrade Werbeausgaben und Besucherzahl	16
Abbildung 12	Regressionsanalyse	17
Abbildung 13	Regressionsanalyse mit multiplen Einflussfaktoren	18
Abbildung 14	Umsatzentwicklung im Zeitablauf.....	20

Tabellen

Tabelle 1	Rohdaten in Excel	5
Tabelle 2	Häufigkeitsauswertungen: Univariat und als Kreuztabelle.....	5
Tabelle 3	Tabellarische Auswertungen: Mittelwerte nach Alter und Geschlecht.....	6
Tabelle 4	Besucherdaten (Beispiel für Ursprungsdaten).....	8
Tabelle 5	Eindimensionale Häufigkeitstabelle – Anzahl BesucherInnen nach Genre	8
Tabelle 6	Zweidimensionale Häufigkeitstabelle – Befragte nach Alter und Geschlecht	8
Tabelle 7	Prozentuale Häufigkeiten – Befragte nach Alter und Geschlecht	9
Tabelle 8	Mittelwerte.....	11
Tabelle 9	Beispiel für Lage- und Streuungsmaße	13
Tabelle 10	Wertetabelle für Zusammenhang Werbeausgaben und Besucherzahlen.....	14
Tabelle 11	Zusammenhangsmaße für unterschiedliche Skalenniveaus	15
Tabelle 12	Zeitreihe der Umsatzentwicklung.....	19

1 QUANTITATIVE METHODEN; STATISTIK

Statistik ist ein Gebiet, das mit vielen Vorbehalten und Vorurteilen behaftet ist. Sie dies die Sorge vor zu viel Mathematik, Formeln und anderem schwer Verständlichem. Oder seien es Redensarten, die ein unbedarftes Herangehen an dieses Gebiet erschweren, wie der berühmte Ausspruch (Winston Churchill fälschlich zugeschrieben) „Ich traue keiner Statistik, die ich nicht selbst gefälscht habe“ oder der beliebten Steigerung „Lüge – gemeine Lüge – Statistik“.

Trotzdem begegnet uns Statistik an vielen Stellen des täglichen (kultur- / betrieblichen) Lebens und es ist wichtig, damit umgehen zu können. Es ist *nicht* sinnvoll, aufgrund kompliziert klingender Begriffe davon auszugehen, dass das Gegenüber „schon Recht haben wird“. Wenn Sie selbst in der Lage sind, fachkundig nachzufragen können Sie Aussagen kritisch hinterfragen. Nicht alle statistischen Modelle und Kennzahlen sind in allen Zusammenhängen und für alle Arten von Daten anwendbar.

Aber nicht nur die Situation, vorgelegte statistische Auswertungen verstehen und (kritisch) interpretieren zu müssen, kann in der täglichen Praxis auftauchen, sondern auch der Wunsch, vorhandene Daten selbst auszuwerten und anschaulich darzustellen und daraus zu lernen. Dies kann die (grafische) Aufbereitung zur Präsentation der Daten sein, aber auch die Analyse von statistischen Zusammenhängen bzw. Unterschieden von Daten oder Sachverhalten.

Ja, Statistik kann und soll Spaß machen!

Fun with data

Wir begleiten in diesem Skript drei Studierende, Adam, Beate und Chris, die ein studentisches Unternehmen gegründet haben, das „**Forum**“ – eine Café-Bar mit einer kleinen Bühne für Auftritte lokaler Künstler.

*Begleitende
Excel-Datei*

In diesem Skript werden wichtige betriebsstatistische Methoden nicht nur vorgestellt, sondern können mit dem PC selbst durchgeführt werden, da sich auf der Webseite <http://schmidt-bremen.de> eine Excel-Datei befindet, mit der die Beispiele aus dem Text nachvollzogen werden können.

Es wird hier beispielhaft das Tabellenkalkulationsprogramm MS Excel verwendet, da dieses eine sehr große Verbreitung hat. In anderen Tabellenkalkulationen können die dargestellten Methoden ebenso verwendet werden. Darüber hinaus gibt es spezielle Statistik-Programme, die die Verarbeitung von Daten zwar erleichtern, für den täglichen Gebrauch jedoch i.d.R. nicht notwendig sind, z.B. SPSS, SAS, R, EViews u.v.m. Auf diese wird hier nicht eingegangen.

Diese Unterlagen wollen mehr bieten als nur die Aufzählung verschiedener Methoden und deren kurze verbale Beschreibung. Ziel ist eine lesbare und alltagstaugliche Übersicht über gängige Methoden und nicht eine mathematisch umfassende Darstellung der Statistik. Auf Formeln wird weitgehend verzichtet; zur Vergleichbarkeit mit bzw. Orientierung in Nachschlagewerken werden die üblichen Symbole (Buchstaben, Abkürzungen) verwendet. Am Ende des Skripts findet sich ein Schlagwortindex, der das Auffinden einzelner Begriffe erleichtern soll. Für tiefergehende Fragen sind am Schluss einige Literaturhinweise zusammengestellt.

Ziel dieses Skriptes ist es, zu zeigen, dass quantitative Methoden in der täglichen Arbeit – v.a. durch den Einsatz von PC-Programmen – einfach zu erstellen und dadurch praktisch und nutzbringend einsetzbar sind.

1.1 Das „Forum“

Pauken statt Kreativität – immer wieder saßen Adam, Beate und Chris abends in der geschlossenen Mensa und fragten sich, ob es noch etwas Anderes gibt als für die nächste Prüfung zu lernen. Und dass diese ehemalige Kellerkneipe auf dem Campus leer steht, ist wirklich ein Jammer. So entstand die Idee von drei Studierenden, Adam, Beate und Chris, das „Forum“ zu gründen – eine Café-Bar mit einer kleinen Bühne für Auftritte lokaler Künstler.

Nach einigen Diskussionen entschieden sie sich, das tatsächlich zu tun. Das Erarbeiten des Business-Plans passte sogar ins BWL-Modul, die Dozentin sagte Hilfe zu, die Fakultät war bereit, den leeren Kellerraum zur Verfügung zu stellen und so wurde die Planung konkret.

Im „Forum“ – treffen sich jetzt Studierende und Gäste und genießen Kaffee, Kultur und kulinarisches.

Die UnternehmerInnen des „Forum“:

Adam studiert Tourismusmanagement im Schwerpunkt Marketing, ist 28 Jahre und arbeitet neben dem Studium in einer Szene-Bar. Er bringt das gastronomische Know-how sowie Kontakte aus der Gastro-Szene mit, hat aber wenig Affinität zu Zahlen, dafür mehr zur Kleinkunst.

Ganz anders die 23-Jährige **Beate**. Sie studiert Finanzen und Rechnungslegung und soll dafür sorgen, dass die Aktivseite der Bilanz stets wächst und auch alles richtig dokumentiert wird.

Chris studiert Kulturmanagement. Der gesellschaftskritischen 19 Jährigen ist es wichtig, dass nur biologische und ethisch unbedenkliche Produkte angeboten werden.

1.2 Definitionen – Was ist Statistik?

Statistik ist ein Hilfsmittel, ein Werkzeug zur systematischen Darstellung und Auswertung von Zahlenmaterial, meist kurz als „Daten“ bezeichnet. Mit statistischen Methoden werden Kennzahlen gebildet, die dabei helfen, vorliegendes Datenmaterial - vor allem aber die entsprechenden Sachverhalte - möglichst objektiv zu bewerten.

Es gibt zwei grundlegende Ziele statistischer Analysen:

- Beschreibung vorhandener Daten: **Beschreibende oder Deskriptive Statistik**
 Es liegen Daten (Zahlen) vor, die ausgewertet werden sollen: z.B. Alter und Einkommen von BesucherInnen, Zuschauerzahlen nach Wochentag oder Umsatzzahlen in Quartalen, ... usw.
- Ableiten allgemeiner Aussagen **Schließende oder Induktive Statistik**
 aus einer kleinen Auswahl von Daten
 Es liegt nur eine (kleine) **Stichprobe** von Daten vor, aus diesen sollen allgemeingültige Schlüsse über die Grundgesamtheit aller Daten gezogen werden: z.B.: Aus den Angaben über Alter, Geschlecht und Vorlieben von 50 zufällig ausgesuchten Besuchern soll auf alle potentiellen Besucher, Ihre Zusammensetzung und Vorlieben geschlossen werden, um so das Programm des nächsten Halbjahres optimal zu planen.

Dieses Skript behandelt die beschreibende Statistik.

Verdichtung von Informationen – abhängig von der Skalierung der Daten

Eine Hauptaufgabe statistischer Methoden ist es, die oft sehr große Fülle von Informationen auf wenige (Kenn-) Zahlen zu *verdichten*. Beispiel: Von 150 Besuchern einer Ausstellung mögen z.B. für jede Person Alter, Geschlecht, Dauer des Aufenthalts und die besuchte Veranstaltung vorliegen, dies sind 600 einzelne Datenpunkte. Was können wir aus diesen lernen?

beschreibende
Statistik

schließende
Statistik

Verdichtung von
Datenmaterial

Statistisch sprechen wir von **Merkmalen** (z.B. Alter, Geschlecht) und deren **Ausprägungen** (z.B. 20 Jahre, 44 Jahre bzw. männlich, weiblich, divers).

Durch Auszählung von Häufigkeiten oder Angabe eines Mittelwertes können diese auf wenige „Kennzahlen“ reduziert werden, die für betriebliche Entscheidung hilfreich sind.

Fallstudie

Das „**Forum**“ ist eingerichtet; Geräte, Geschirr und Sitzmöbel sind da. Nun soll es losgehen. Zunächst einmal müssen die Öffnungszeiten geplant und alles bestellt werden, was benötigt wird: Vom Kaffeepulver, Tee über Milch bis hin zu Servietten, Snacks ... Wie viel wird benötigt?

Auf der Webseite finden Sie einen Fragebogen, der in der Hochschule verwendet wurde sowie eine Excel-Datei mit Auswertungen. Aus diesen stammen die folgenden Beispiele.

Von der „Urliste“ zur Auszählung, Auswertung und Grafik

Nach der Eingabe der Daten (in ein Tabellenkalkulationsprogramm, z.B. Excel, Calc, Sheets, ...), schauen wir zunächst einmal auch eine große Menge Daten und mögen uns fragen: „Und nun?“ Hier ein Beispiel für einen kleinen Ausschnitt: die ersten 8 Personen und die ersten 5 Fragen. Für drei Merkmale wurden bereits Rechnungen durchgeführt bzw. Beschriftungen eingefügt:

Person	Kaffee			Alter		Geschlecht		Berufsjahre
	Menge/Tag	Tage/Woche	Menge/Woche	Alter	Altersgruppe	(Kürzel)	Geschlecht	
1	0	0	0	19	unter 20	m	männlich	0
2	0	0	0	22	22++	m	männlich	0
3	1	5	5	20	20-21	w	weiblich	0
4	0	0	0		-	m	männlich	0
5	1	3	3	20	20-21	m	männlich	1
6	2,5	5	12,5	27	22++	m	männlich	12
7	1	7	7	19	unter 20	m	männlich	3
8	1	2	2	29	22++	m	männlich	2

Tabelle 1 Rohdaten in Excel

Um aus diesen Zahlen Informationen ziehen zu können, müssen sie aufbereitet werden. Ein paar Beispiele:

1. Auszählung der Antworten, z.B. nach Altersgruppen – oder Alter und Geschlecht gleichzeitig

Altersgruppe	Anzahl
-	2
unter 20	14
20-21	18
22++	9
Gesamt	43

Altersgruppe	männlich	weiblich	Gesamt
-	2		2
unter 20	2	12	14
20-21	10	8	18
22++	4	5	9
Gesamt	18	25	43

Tabelle 2 Häufigkeitsauswertungen: Univariat und als Kreuztabelle

Diese können wir auch graphisch darstellen:

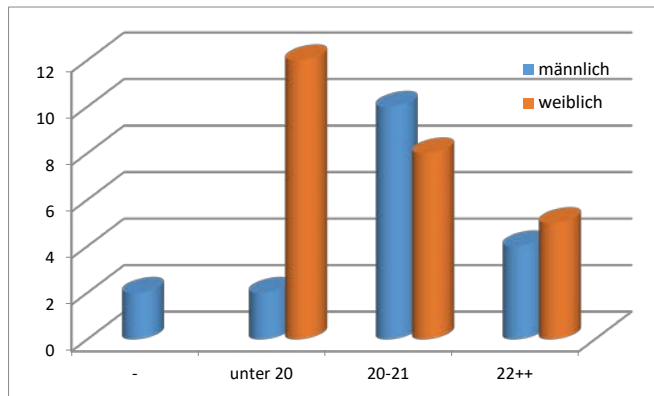


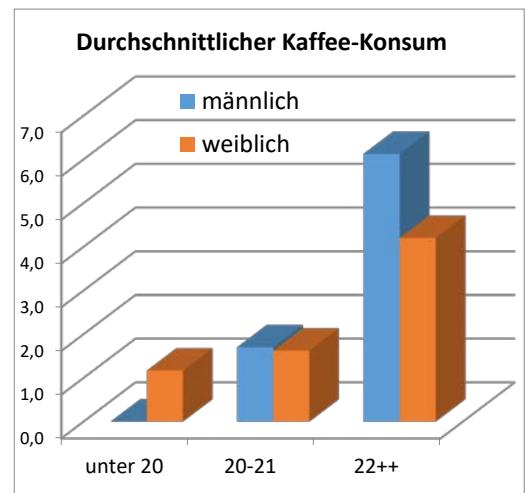
Abbildung 1
Häufigkeitsdiagramm

2. Auswertung der Antworten, z.B. durchschnittlicher Kaffeekonsum – nach Alter und Geschlecht

Durchschnittlicher Kaffee-Konsum/Woche nach Alter und Geschlecht

Altersgruppen	männlich	weiblich	Gesamt
unter 20	0,0	1,2	1,0
20-21	1,7	1,6	1,7
22++	6,1	4,2	5,1
Gesamt	2,6	1,9	2,2

Abbildung 2
Säulendiagramm für Mittelwerte



Wichtigkeit alkoholischer Getränke nach Alter und Geschlecht

Altersgruppen	männlich	weiblich	Gesamt
unter 20	3,0	1,3	1,5
20-21	3,3	1,6	2,6
22++	3,5	1,0	2,1
Gesamt	3,3	1,3	2,1

Tabelle 3 Tabellarische Auswertungen: Mittelwerte nach Alter und Geschlecht

Solche und viele weitere Auswertungen werden wir in den kommenden Kapiteln kennen lernen. Freuen wir uns auf spannende Ergebnisse ☺

Adam, Beate und Chris sitzen zusammen und besprechen die Ergebnisse einer Kunden-Befragung. Es fällt ihnen auf, dass nicht alle Variablen gleich behandelt werden können. Die einen sind „echte“ Zahlen, andere stellen Kategorien oder ja-nein Antworten dar.
 Welche Arten von Daten gibt es? Wie können diese dargestellt und interpretiert werden?

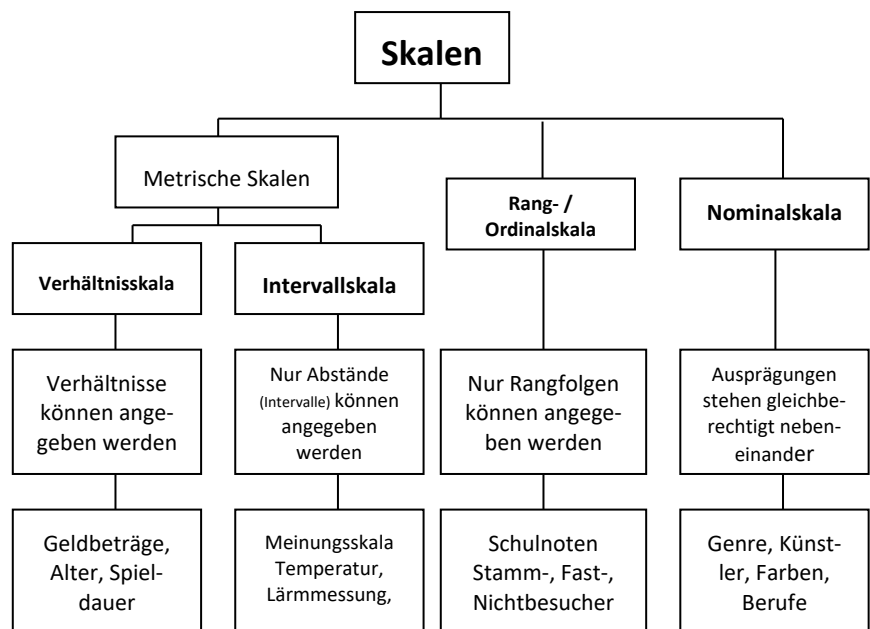
Während in den Ursprungsdaten also alle Personen mit allen Eigenschaften enthalten sind, beinhalten Verdichtungen nur einzelne ausgewertete Kennzahlen.

Es zeigt sich jedoch, dass nicht alle Maßzahlen für alle Merkmale möglich sind, so wäre z.B. ein Mittelwert beim Genre nicht sinnvoll. Genauer gesagt ist die Auswahl der statistischen Maßzahlen von der **Skalierung des Merkmals** abhängig. Abbildung 3 zeigt die vier Skalen, die üblicherweise unterschieden werden.

Abbildung 3 Skalierung von Merkmalen

Skalierung von Daten als Basis für die Anwendbarkeit statistischer Methoden

Verhältnisskalierte Daten beinhalten die meiste Information, nominal skalierte die wenigste. Entsprechend stehen mehr oder weniger statistische Methoden zur Auswertung der Daten zur Verfügung. Merkmale können in diskreter oder stetiger Form vorliegen. Diskrete Merkmale können nur abzählbar viele Ausprägungen annehmen, wie z.B. oben Genre, Geschlecht oder Farben. Stetige Merkmale hingegen können beliebig viele Ausprägungen annehmen, oft werden sie in Dezimalzahlen gemessen, z.B. Geldbeträge, Gewichte oder Mengen.



Die Unterscheidungen von Typen und Skalen werden im Folgenden wichtig sein, wenn die Methoden zur Auswertung beschrieben werden.

1.3 Häufigkeiten und grafische Darstellungen

Wie oben gesehen, ist auch die Auszählung von Häufigkeiten ein Mittel zur Verdichtung von Daten, gerne wird diese grafisch dargestellt.

1.3.1 Eindimensionale und mehrdimensionale Häufigkeitsdarstellungen

Die Ursprungsdaten (oder Rohdaten - vgl. Tabelle 1 Rohdaten in Excel) werden oft in Tabellen dargestellt, die aus Zeilen und Spalten bestehen. Dabei stellt jede Zeile eine statistische Einheit (Person, Werkstück, Summe, ...) und jede Spalte ein bestimmtes Merkmal dar.

Diese Darstellung wird auch in Tabellenkalkulationsprogrammen verwendet. Hier kann dann jede Zelle (z.B. Zeile 3, Spalte 4) einzeln angesteuert bzw. berechnet werden. Beispiele hierzu finden sich in den Excel-Dateien. Die laufende Nummer wird auch als (Lauf-) Index bezeichnet, daher die übliche Abkürzung i . Es können dann alle Angaben anhand dieses Index angegeben werden. In Tabelle 4 kam beispielsweise die 3. Person zweimal ins Theater: $B_3 = 2$.

Tabelle 4 Besucherdaten (Beispiel für Ursprungsdaten)

Ursprungsdaten

i	Geschlecht	Besuchshäufigkeit	↓ Spalte	
			Besuchtes Genre	Zufriedenheit
i	G_i	B_i	S_i	N_i
1	w	10	Cabaret	2
2	w	5	Musik	1
3	m	2	Theater	3
4	w	9	Cabaret	8
5	m	11	Theater	1
6	m	4	Cabaret	9
7	m	6	Musik	2

Zeile →

Zelle

Die einfachste Verdichtung von Daten ist die Angabe von Häufigkeiten, oft ebenfalls in tabellarischer Form, wie Tabelle 5 für den Fall einer einfachen Häufigkeitstabelle für das Merkmal „Berufstatus“ zeigt.

Tabelle 5 Eindimensionale Häufigkeitstabelle – Anzahl BesucherInnen nach Genre

Häufigkeitstabelle

Kürzel	Status	Anzahl (Häufigkeit n_i)	(Bezeichnung)
T	Theater	100	= n_1
C	Cabaret	250	= n_2
M	Musik	50	= n_3
Summe:		400	= n BesucherInnen

Der Buchstabe „ n “ als Symbol für „Anzahl der Beobachtungen“ wird in der Statistik sehr häufig verwendet. Wenn es sich auf die Grundgesamtheit aller statistischen Einheiten bezieht, wird auch ein großes „ N “ verwendet.

Interessanter ist die Aufbereitung mehrerer Dimensionen, etwa die Auszählung der Anzahl der Besucher, diesmal nach Alter *und* Geschlecht, wie sie in Tabelle 6 vorgenommen wird.

Tabelle 6 Zweidimensionale Häufigkeitstabelle – Befragte nach Alter und Geschlecht

zweidimensionale Häufigkeiten

Besuchshäufigkeit	Geschlecht		alle Personen	Rand-
	weiblich	männlich		
bis 2	80	40	120	summen
2 – 5	100	80	180	
5 und mehr	120	80	200	
alle Personen	300	200	500	Gesamtsumme

Randsummen

In dieser Tabelle 6 sind zum einen die Einzelhäufigkeiten für die Kombinationen bestimmter Eigenschaften (z.B. kamen 40 Männer bis zu zweimal), aber auch - in den „Randsummen“ die Häufigkeitsauszählungen für die einzelnen Merkmale (z.B. insgesamt gibt es 180 Personen mit einer Besuchshäufigkeit zwischen 2 und 5 Vorstellungen). Für das Merkmal Besuchshäufigkeit wurden *Klassen* (von ... bis ...) gebildet. Dies ist sinnvoll, wenn so viele Ausprägungen vorhanden sind, dass diese nicht mehr übersichtlich in einer Tabelle oder Grafik dargestellt werden können. Üblich ist auch die Darstellung von *relativen* oder *prozentualen* Häufigkeiten.

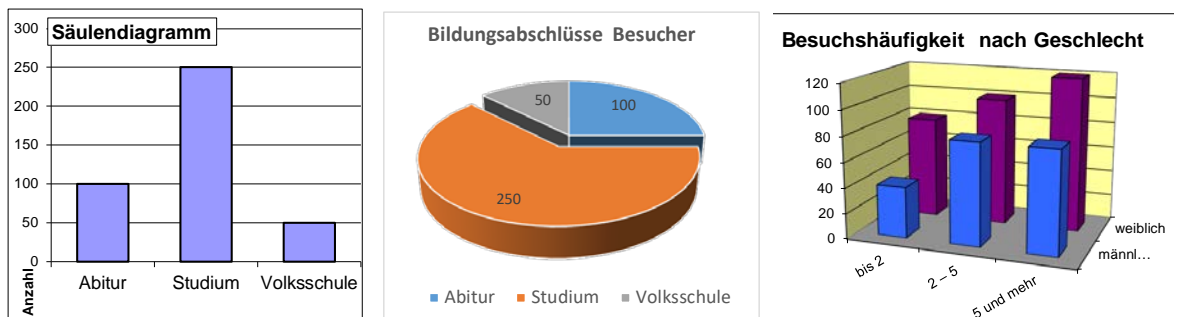
Tabelle 7 Prozentuale Häufigkeiten – Befragte nach Alter und Geschlecht

relative Häufigkeiten

Besuchshäufigkeit	Zeilenprozent			Spaltenprozent		
	weiblich	männlich	alle Personen	weiblich	männlich	alle Personen
bis 2	66,7%	33,3%	100%	26,7%	20,0%	24,0%
2 – 5	55,6%	44,4%	100%	33,3%	40,0%	36,0%
5 und mehr	60,0%	40,0%	100%	40,0%	40,0%	40,0%
alle Personen	60,0%	40,0%	100%	100%	100%	100%

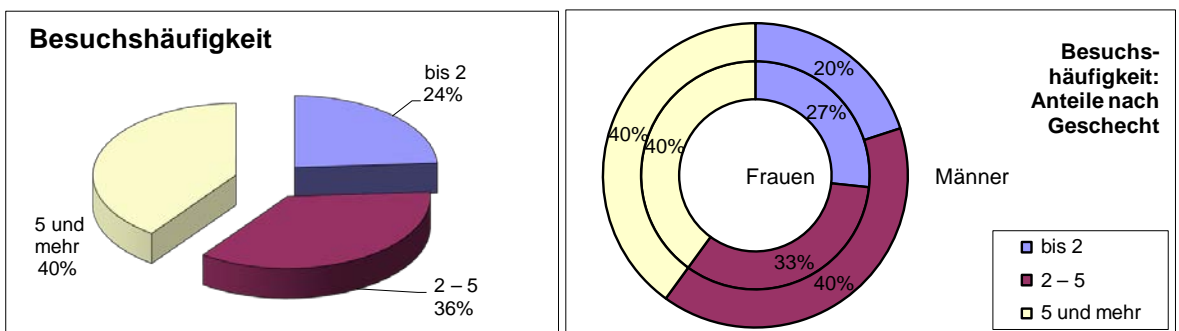
Diese Häufigkeitsdarstellungen, ob in absoluten Zahlen oder relativen Anteilen gemessen, werden oft grafisch dargestellt. So lassen sich die Zahlen aus Tabelle 5 z.B. in einem Säulen- oder Tortendiagramm darstellen, wie in Abbildung 4 zu sehen.

Abbildung 4 Absolute Häufigkeiten und zweidimensionales Säulendiagramm



Auch die zweidimensionalen Häufigkeiten aus Tabelle 6 lassen sich grafisch veranschaulichen (z.B. wie in Abbildung 4 recht oder Abbildung 5).

Abbildung 5 Kreisdiagramme zur Darstellung von relativen Häufigkeiten



Grafiken wie die hier beispielhaft vorgestellten lassen sich mit Hilfe von Computerprogrammen relativ einfach erzeugen. Es gibt eine sehr große Anzahl von Darstellungsmöglichkeiten und es

sollte jeweils aus dem konkreten Zusammenhang entschieden werden, welche Darstellung hilfreich „für den Transport der Botschaft“ ist.

Die Daten und die hier dargestellten Beispiele finden sich in der begleitenden Excel-Datei.

Die verschiedenen Skalen und Typen helfen dabei, die Daten mit den korrekten statistischen Maßen zu analysieren. Wir haben kennengelernt:

- Metrische Skalen (Verhältnisskala und Intervallskala)
- Ordinalskala
- Nominalskala

Außerdem gibt es diskrete und stetige Merkmale als Typen von Variablen.

Für verschiedene Typen werden die Häufigkeiten auf unterschiedliche Weise ermittelt.

Auch in den folgenden Kapiteln werden die Typen und Skalen der betrachteten Merkmale wichtig sein.

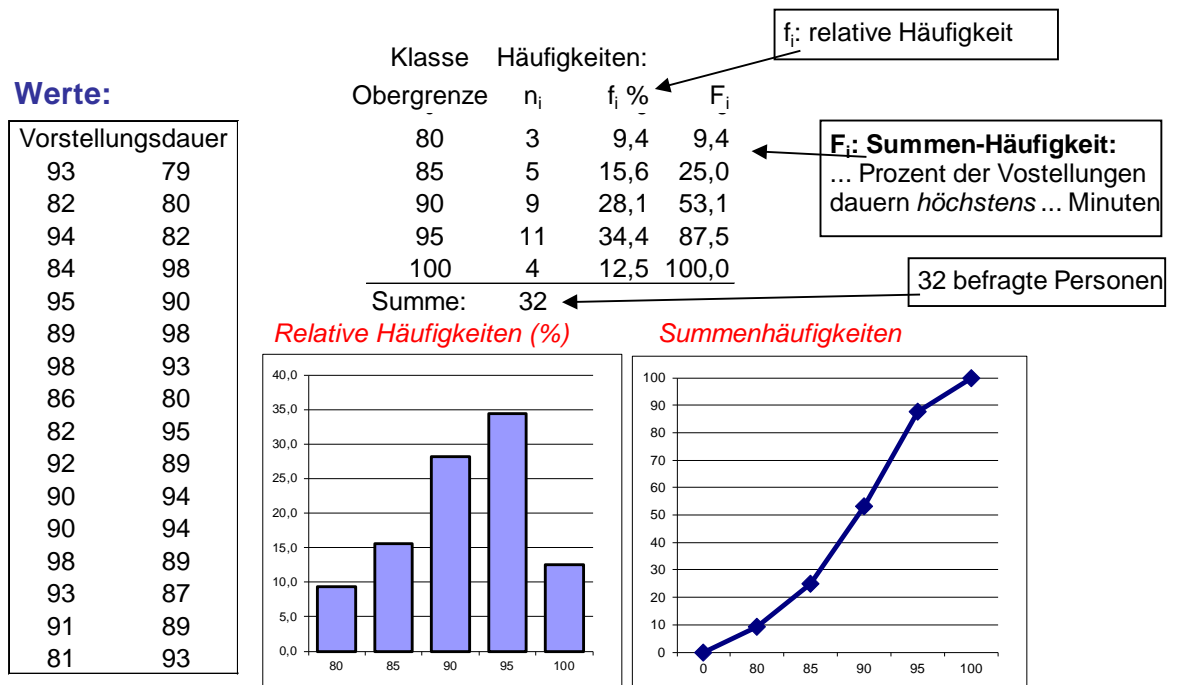
1.3.2 Besondere Häufigkeitskonzepte

Histogramme werden verwendet, wenn die Ausprägungen wie oben in Klassen eingeteilt werden und diese unterschiedlich breit sind. Einfache Säulendiagramme würden in diesem Fall falsche Häufigkeiten vermuten lassen, so dass die Häufigkeiten als Fläche dargestellt werden.

Summenhäufigkeitsfunktionen zeigen, wie viel (Prozent der) Ausprägungen *höchstens* einem bestimmten Wert annehmen (*bis zu ...*). Abbildung 6 zeigt dieses Häufigkeitskonzept neben einem einfachen Säulendiagramm.

Histogramme und Summenhäufigkeiten

Abbildung 6 Einfache Häufigkeiten und Summenhäufigkeiten



1.4 Lagemaße (Mittelwerte) und Streuungsmaße

Beate ist unzufrieden. Das **Forum-Café** brummt nicht so wie erwartet und sie weiß nicht genau warum. Sie vermutet, dass die Lage des **Forum-Cafés** nicht ideal gewählt worden ist. Deswegen möchte sie die Entfernung (in m) zu den verschiedenen Seminarräumen vom **Forum-Café** aus berechnen. Die Ergebnisse sind: 100, 130, 900, 20, 80, 20, 150 Meter.

Was ist die mittlere Entfernung der Seminarräume zum Forum-Café?

Als Antwort sind verschiedene Mittelwert-Maße möglich – warum? Wie interpretieren wir diese?

Die bisherige Beschreibung zeigte Möglichkeiten der Darstellung, die deutlich anschaulicher sind als die Betrachtung von Ursprungsdaten, aber die Tabellen und Abbildungen in Punkt 1.3 müssen jeweils ihrerseits interpretiert werden: „Wie unterscheiden sich zwei Grafiken?“; „Was ist das wichtige an diese Tabelle?“ ... mögen die Fragen lauten. Daraus ergibt sich der Wunsch, nach noch knapperen statistischen (bzw. betrieblichen) *Kennzahlen*.

Die bekanntesten statistischen Maße sind **Mittelwerte**, unter Ihnen der „prominenteste“ das arithmetische Mittel, oft einfach als „Mittelwert“ bezeichnet. Aber nicht für alle Daten (jeder Skalierung) kann ein arithmetisches Mittel errechnet werden. In der betrieblichen Praxis sind die in Tabelle 8 angegebenen Mittelwerte relevant.

Mittelwerte werden auch als Lagemaße bezeichnet, da sie die Lage einer Verteilung (auf der waagrechten Achse) angeben. So haben die Verteilungen von Abbildung 4 oder Abbildung 6 ihren jeweiligen Schwerpunkt in ihren Mittelwerten.

Mittel-
werte:
auf die
Skala
achten

Tabelle 8 Mittelwerte

Mittelwert	für Skalen	Definition	Beispiel
Modus (Modalwert)	alle Skalen	häufigster Wert	In Tabelle 5 ist C der Modalwert, da Cabaret (mit 250 Personen) die größte Einzelhäufigkeit aufweist
Median oder Zentralwert	Ordinalskalen und metrische Skalen	Mitte aller geordneten Ausprägungen	Alter: 44, 19, 24, 60, 21, 42, 11 geordnet: 11,19,21, 24 ,42,44,60 Median = 24, da mittlerer Wert, es stehen rechts und links davon je drei Zahlen
arithmetisches Mittel („Mittelwert“)	metrische Skalen	Summe aller Werte geteilt durch die Anzahl der Beobachtungen (n)	Alter: 44, 19, 24, 60, 21, 42, 11 $\frac{44 + 19 + 24 + 60 + 21 + 42 + 11}{7} =$ <i>Durchschnittsalter $\bar{x} = 31,6$ Jahre</i>
geometrisches Mittel	Steigerungsraten von Wachstumsdaten (nur Verhältnisskalen)	n-te Wurzel aus dem Produkt aller Werte	Steigerung Zuschauerzahlen in 4 Jahren: 10 %, 20 %, 5 %, 1 % $\sqrt[4]{1,10 \cdot 1,20 \cdot 1,05 \cdot 1,01}$ <i>GM = 1,0877 oder rund 8,77 Prozent</i>

Die einzelnen Berechnungen können in der Excel-Datei nachvollzogen werden.

Wichtig ist hier, dass falsche Verwendung der Mittelwerte eben auch zu falschen (oder verfälschten) Ergebnissen führt. So ist das arithmetische Mittel i.d.R. größer als das geometrische Mittel. Würde man im letzten Beispiel unten fälschlicherweise ein arithmetisches Mittel errechnen, so hätte dies einen Wert von 9 Prozent ($10+20+5+1=36 / 4$). Die tatsächliche Steigerung der letzten vier Jahre würde also höher angegeben als sie tatsächlich war.

Ziel war es, die durchschnittliche Entfernung zur *StudierBar* zu ermitteln.

Wir haben gesehen, dass es verschiedene Mittelwerte gibt, die auch verschiedene Werte ergeben:

- arithmetisches Mittel: 200 Meter
- Median: 100 Meter
- Modus 20 Meter

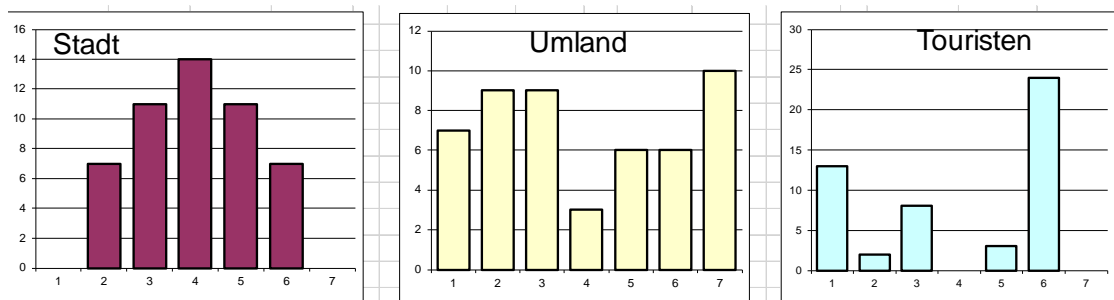
Diskutieren Sie: Welches Maß „trifft es“ am besten? Warum?

Streuungsmaße

Streuungsmaße

Neben der Lage einer Verteilung ist diese durch ihr Aussehen, etwa ihre „Breite“ bezeichnet: Wie verteilen sich die Ausprägungen des Merkmals um den Mittelwert? Statistisch wird hier von *Schwankung* oder *Streuung* der Werte gesprochen, so dass die entsprechenden Kennzahlen als **Streuungsmaße** bezeichnet werden. Abbildung 7 zeigt ein Beispiel für 3 verschiedene Verteilungen (Häufigkeitsauszählungen als Säulendiagramm). In allen drei Fällen wurden 50 BesucherInnen (aus drei Herkunftsregionen) befragt: „Wie hat Ihnen die heutige Vorstellung gefallen?“. Die Antworten aus allen drei Regionen ergaben denselben arithmetischen Mittelwert von 4,0 – die Verteilungen sehen aber unterschiedlich aus.

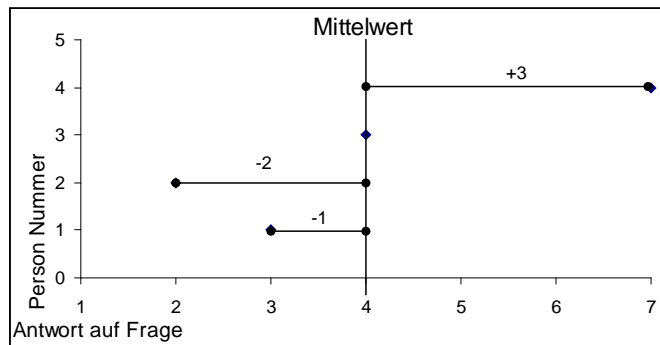
Abbildung 7 Verteilungen mit unterschiedlicher Streuung



Mit Streuungsmaßen kann das unterschiedliche Aussehen dieser drei Verteilungen, statistisch gesagt die unterschiedliche Schwankung gemessen werden: Das einfachste Streuungsmaß ist die **Spannweite**. Sie wird ermittelt, indem die kleinste Ausprägung von der größten abgezogen wird (Stadt: $6-2=4$; Umland: $7-1=6$; Touristen: $6-1=5$) und spiegelt so die Breite der Verteilung wider.

Üblichere Streuungsmaße messen die Abweichung der einzelnen Werte vom Mittelwert, was in Abbildung 8 veranschaulicht wird (am Beispiel von vier Personen aus der Stadt). Als statistische Kennzahl dient wiederum ein Mittelwert dieser Abweichungen. Die „Durchschnittliche Absolute Abweichung (DAA)“ ist ein mögliches Maß, weitaus bekannter jedoch ist die **Standardabweichung**. Sie wird ermittelt aus der Hilfsgröße „Varianz“ (dem Durchschnitt aller quadrierten Abweichungen). Die Formeln zu deren Ermittlung finden Sie in der Excel-Datei.

Abbildung 8 Abweichungen der Einzelbeobachtungen vom Mittelwert



Antworten auf einer Ratingskala (Werte von 1 bis 7 konnten angegeben werden)

In Tabelle 9 sind die vorgestellten Maßzahlen für das obige Beispiel (Befragung von je 50 Personen aus drei Herkunftsregionen) zusammengefasst.

Tabelle 9 Beispiel für Lage- und Streuungsmaße

Herkunft	Median	arithm. Mittel	Spannweite	Standardabweichung
Stadt	4	4,0	4 (von 2 bis 6)	1,26
Umland	3	4,0	6 (von 1 bis 7)	2,16
Touristen	5	4,0	5 (von 1 bis 6)	2,18

Es mag sich angesichts dieses einfachen Beispiels die Frage stellen, welchen Sinn solche recht aufwendigen Maßzahlen haben. Dieser liegt vor allem in der *Verarbeitung großer Datenmengen*. Sind nicht nur drei Herkunftsregionen, sondern z.B. 20 PLZ-Bereiche und nicht nur eine Frage, sondern z.B. 35 zu bewerten und so zu verdichten, dass eine Orientierung „auf einen Blick“ (oder zumindest wenige Blicke) nicht mehr möglich ist, so geht dies nur mit Hilfe von Kennzahlen. Nicht alle sind an jeder Stelle geeignet. So zeigt hier das arithmetische Mittel eine Übereinstimmung der drei Regionen an, was aber angesichts Abbildung 7 nicht zu überzeugen vermag. Schon der Median, vor allem aber die Streuungsmaße zeigen - auch ohne den Blick auf die Grafiken - dass die Antworten aus der Stadt recht einheitlich verteilt sind, wogegen diejenigen in der Region und von Touristen größere Schwankungen aufweisen. Dort gehen also die Meinungen weiter auseinander, in unserem Beispiel könnte hiermit ein Anhaltspunkt dafür gegeben sein, dass die Außenkommunikation für weiter anreisende BesucherInnen verbessert werden kann.

1.5 Zusammenhänge zwischen mehreren Merkmalen

Beate und Adam diskutieren wie sie die Umsätze im *Forum-Café* erhöhen können. Adam schlägt vor, mehr Werbung zu machen, z.B. Flugblätter oder Inserate im Szene-Magazin. Beate glaubt nicht, dass diese Ausgaben gerechtfertigt sind. Wie können wir messen, ob es einen Zusammenhang zwischen Werbeausgaben und Umsatz gibt? Und wenn ja, ist es möglich, für bestimmte Werbeausgaben erwartete Besucherzahlen abzuschätzen?

Oft ist das Ziel statistischer Analysen nicht nur, *ein einzelnes* Merkmal zu beschreiben, sondern es interessiert das Verhältnis verschiedener Merkmale zueinander. So wurde bereits in Tabelle 7 und Abbildung 4 eine *zweidimensionale* Betrachtung angestellt.

Es können zwei unterschiedliche Fragestellungen unterschieden werden:

- Wird ein **zufälliger** Zusammenhang zwischen zwei Merkmalen untersucht, in dem Sinne, dass die Merkmale sich gegenseitig beeinflussen; oder
- wird ein **ursächlicher Zusammenhang** vermutet, in dem Sinne, dass bestimmte Merkmale ein anderes beeinflussen bzw. steuern (Kausalität)?

Für die erste Frage eignen sich Zusammenhangmaße, die zweite kann mit Regressionsmodellen untersucht werden.

1.5.1 Zusammenhangmaße

Messen der
Stärke
eines Zu-
sammen-
hangs

Zusammenhangmaße beschreiben die *Stärke* eines Zusammenhangs. Typischerweise wird versucht, die Besucherzahl durch Werbemaßnahmen zu verbessern – die natürlich Geld kosten. Daher gibt es regelmäßig Diskussionen zwischen der Marketing-Abteilung (hier Adam, der mehr Werbung machen möchte) und der Geschäftsführung (hier Beate, die die Kosten im Auge hat). „Dann lasst und doch mal schauen, ob Werbung wirkt“ - Es werden sechs Vorstellungen ($i = 1, \dots, 6$), entsprechend sechs Wertepaare (x_i, y_i) miteinander verglichen. Diese sind in Tabelle 10 angegeben.

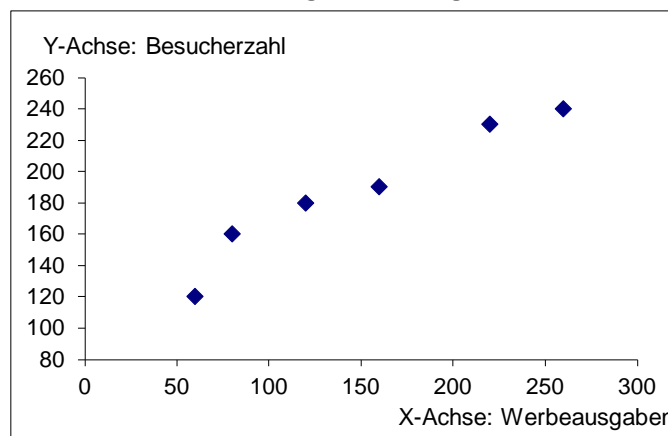
Tabelle 10 Wertetabelle für Zusammenhang Werbeausgaben und Besucherzahlen

Woche	Werbeausgaben	BesucherInnen	Woche	Werbeausgaben	BesucherInnen
i	X_i	Y_i	i	X_i	Y_i
1	60	120	4	160	190
2	80	160	5	220	230
3	120	180	6	260	240

Da es sich um zwei metrisch skalierte Merkmale handelt, können sie - in Abbildung 9 - als xy-Diagramm (Streu- oder Punktdiagramm) dargestellt werden.

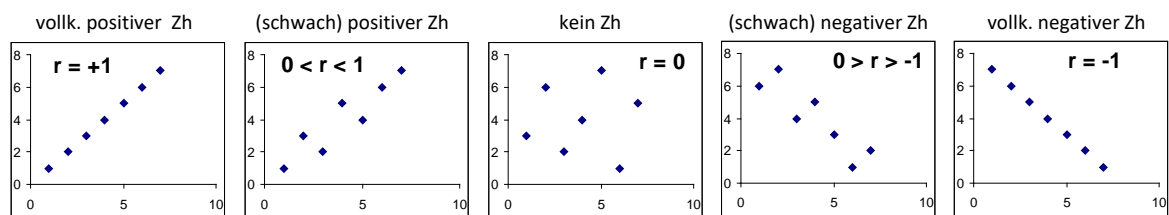
In der folgenden Grafik ist ein Zusammenhang zu erkennen: Je höher die Werbeausgaben, desto mehr BesucherInnen kommen. Das statistische Maß, welches die Stärke eines solchen Zusammenhanges metrischer Merkmale misst, heißt **Korrelationskoeffizient** (nach Bravais-Pearson), üblicherweise mit dem Buchstaben r (oder dem griechischen ρ) bezeichnet. r gibt sowohl die Richtung des Zusammenhanges als auch dessen Stärke an, denn er **kann Werte zwischen -1 und 1 annehmen**. Im vorliegenden Falle ergibt sich ein Wert von $r=0,97$ und damit ein starker positiver Zusammenhang zwischen Werbeausgaben (X) und Besucherzahl (Y).

Abbildung 9 xy-Diagramm für Zusammenhang Werbeausgaben und Besucherzahl



Ein positiver Zusammenhang (r größer als 0) heißt, dass je größer die Ausprägung des einen Merkmals (X), desto *größer* auch die des anderen Merkmals (Y); Ein negativer Zusammenhang (r kleiner als 0) heißt, dass je größer die Ausprägung des einen Merkmals (X), desto *kleiner* die des anderen Merkmals (Y), wobei der Zusammenhang desto stärker ist, je näher r an 1 bzw. -1 liegt. Ein Korrelationskoeffizient nahe oder gleich Null bedeutet, dass es *keinen* Zusammenhang zwischen X und Y gibt. Der Korrelationskoeffizient r ist also eine Kennzahl, die eine große Menge an Informationen verdichten kann, indem das Verhältnis beliebig vieler Wertepaare in einer Maßzahl r zusammengefasst wird.

Abbildung 10 Stärke von Zusammenhängen und Werte des Korrelationskoeffizienten



Auch hier gilt wieder, dass ein solches Maß besonders dann nützlich ist, wenn große Mengen von Daten betrachtet werden und nicht für jedes Merkmalpaar ein solches xy-Diagramm erstellt werden kann. Es können dann mittels des Korrelationskoeffizienten schnell diejenigen Merkmale herausgefunden werden, die einen starken Zusammenhang aufweisen und diese näher untersucht werden. Korrelationskoeffizienten sind ein in der betrieblichen Praxis sehr gebräuchliches Maß.

Allerdings kann der Korrelationskoeffizient r nach Bravais-Pearson nur für *metrische* Merkmale ermittelt werden. Bei *ordinal* skalierten Merkmalen muss auf den Rangkorrelationskoeffizienten r_s zurückgegriffen werden; bei *nominal* skalierten Daten steht nur der Kontingenzkoeffizient zur Verfügung. Tabelle 11 zeigt die Zusammenhangsmaße für die verschiedenen Skalenniveaus.

Je größer der Informationsgehalt der Skala (vgl. Abbildung 3), desto höher ist auch die Aussagekraft des Zusammenhangsmaßes. Der Rangkorrelationskoeffizient r_s kann nur Sortierungen vergleichen, aber keine Zahlenwerte, der Kontingenzkoeffizient C beinhaltet keine Richtung des Zusammenhangs. Bezüglich des Wertes von C gilt ebenfalls, dass ein Wert von 0 keinen Zusammenhang bedeutet und je näher C sich dem Wert 1 nähert, desto stärker ist der untersuchte Zusammenhang zwischen den Merkmalen X und Y.

Tabelle 11 Zusammenhangsmaße für unterschiedliche Skalenniveaus

Zusammenhangsmaß	für Skalen	Wertebereich	Beispiele
r Korrelationskoeffizient nach Bravais-Pearson	metrische Skalen	$-1 \leq r \leq 1$	- Werbung und Besucherzahl - Alter und Einkommen
r_s Korrelationskoeffizient nach Spearman	Ordinalskalen	$-1 \leq r_s \leq 1$	- Schulnote und Altersklasse - Schulabschluss und Besuchertypen (Stamm-, Fast-, Nichtbesucher)
C Kontingenzkoeffizient	Nominalskalen	$0 \leq C \leq 1$	- Geschlecht und Lieblings-Genre - Beruf und Herkunft

1.5.2 Regressionsanalyse

Bei metrisch skalierten Merkmalen wurden in Abbildung 9 und Abbildung 10 „Punktewolken“ betrachtet, also die Verteilung der xy-Wertepaare in einem Koordinatensystem. Als Referenz für die Messung der Stärke eines Zusammenhanges dient dabei eine gedachte Linie durch die Punktewolke und die Betrachtung, wie die Beobachtungspunkte zu dieser Linie liegen. Bei Korrelationskoeffizienten $r = 1$ und $r = -1$ liegen die Punkte *auf* dieser gedachten Gerade bzw. bilden diese Gerade.

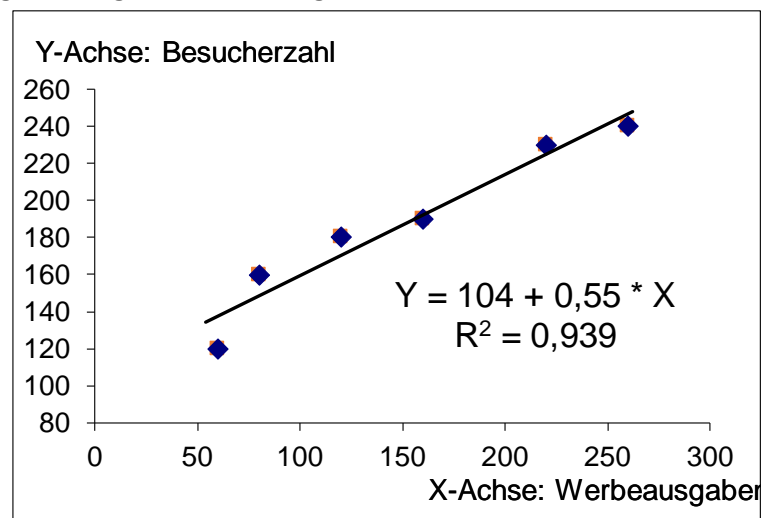
Regressionsanalyse misst den kausalen Zusammenhang als Funktion

Im Beispiel der Tabelle 10 wurde ein statistischer Zusammenhang zwischen den Werbeausgaben und der Besucherzahl ermittelt. Es stellt sich im nächsten Schritt die Frage, *wie* die beiden Merkmale zusammenhängen, welcher Art ihre Beziehung ist. Um dies statistisch zu untersuchen, muss zunächst eine Annahme aufgestellt werden, diese sei:

- inhaltlich: **Die Besucherzahl hängt von den Werbeausgaben ab**
- mathematisch: Die Besucherzahl Y ist eine Funktion der Werbeausgaben X
- funktional: $Y = f(X)$ und im linearen Fall: $Y = a + b \cdot X$.

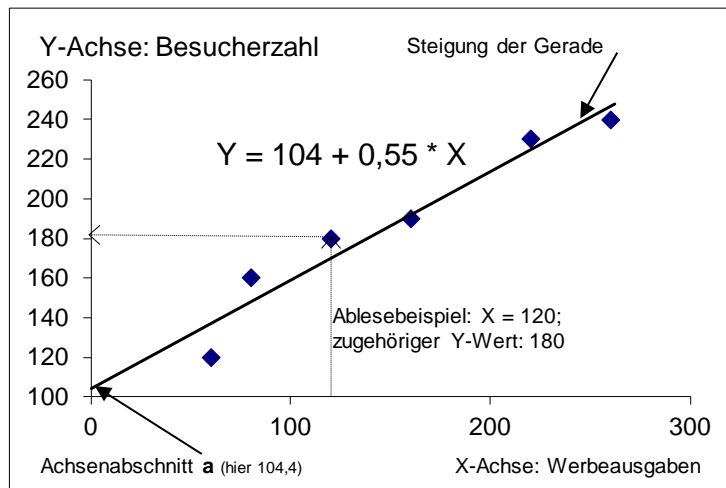
In Worten bedeutet dies, dass eine Gerade gesucht wird, die durch das Zentrum der „Punktewolke“ geht, wie Abbildung 11 zeigt.

Abbildung 11 Regressionsgrade Werbeausgaben und Besucherzahl



Die „Regressionsgrade“ in Abbildung 11 wird bestimmt durch ihren Schnittpunkt mit der Y-Achse (hier $a = 104$) und ihre Steigung (hier $b = 0,55$). Mit dieser Grade bzw. der Formel $Y = a + b \cdot x$; hier $Y = 104 + 0,55 \cdot x$ kann für jede denkbare Höhe von Werbeausgaben ein *erwarteter* Wert für die Besucherzahl errechnet werden. Daher hat das Modell seinen Namen, denn „re-gressere“ kommt aus dem Lateinischen und bedeutet „zurückführen“; hier wird also die Besucherzahl auf die Werbeausgaben zurückgeführt. Dies kann zum einen geschehen durch Einsetzen von X-Werten in die Formel: so ergibt sich für 100 Euro Werbeausgaben eine erwartete Besucherzahl von $Y = 104 + 0,55 \cdot 100 = 159$ BesucherInnen. X wird auch als das *erklärende (unabhängige)* und Y als das *erklärte (abhängige)* Merkmal bezeichnet. Dies ist in Abbildung 12 verdeutlicht.

Abbildung 12 Regressionsanalyse



Eine solche Regressionsanalyse kann in Computerprogrammen sehr einfach erzeugt werden. In der Excel-Datei in der beiliegenden Datei finden Sie diese Grafiken, die zugrundeliegenden Zahlen und Hinweise zur Erstellung der Analysen.

Eine Regressionsanalyse bietet somit zwei praktische Möglichkeiten:

- Ein Zusammenhang kann formal beschrieben werden (wie hängen X und Y zusammen?)
- Es kann für gegebene X-Werte ausgerechnet werden, welche Y-Werte zu erwarten sind.

Anwendung finden Regressionsanalysen in verschiedensten Bereichen der betrieblichen Praxis und sind sehr verbreitet.

Gütemaß

Die Güte einer Regressionsanalyse bemisst sich daran, wie gut die Regressionsgrade den tatsächlichen Zusammenhang beschreibt bzw. vorhersagt. Dies wird darin gemessen, wie stark die einzelnen Beobachtungspunkte um die Gerade schwanken. Liegen alle Punkte auf der Gerade, so ist die Regressionsschätzung perfekt. Liegen sie nahe neben der Grade, so ist die Vorhersage, wie im obigen Beispiel, gut – je weiter die Werte von der Grade entfernt liegen, desto „schlechter“ ist die Regression. Diese Darstellung erinnert an die des Korrelationskoeffizienten und tatsächlich ist im bisher besprochen Fall der linearen Einfachregression $R^2 = r^2$ (also das Quadrat des Korrelationskoeffizienten) ein **Gütemaß** für die Regressionsanalyse. (Es wird allgemein als „R-Quadrat“ ausgesprochen, wobei dies als Eigenname zu verstehen ist. Es gibt keine Zahl R, die dann quadriert wird, sondern das Gütemaß heißt R^2 , bei machen Autoren aber auch B für Bestimmtheitsmaß). In Worten sagt R^2 aus, wie viel Prozent der Schwankungen der Y-Werte durch die X-Werte vorhergesagt werden. R^2 liegt also zwischen 0 und 1 ($0 \leq R^2 \leq 1$). Im obigen Beispiel ist in Abbildung 11 das Gütemaß mit 93,9 Prozent angegeben, diese Beispielregression beschreibt die Daten also gut, was auch grafisch erkennbar ist.

In der Praxis sind allerdings die wenigsten zu untersuchenden Zusammenhänge so einfacher Natur wie das obige Beispiel:

- Der Zusammenhang kann nicht-linear sein, d.h. dass die Punktwolke nicht durch eine Gerade beschrieben werden kann, sondern nur durch eine Kurve bzw. nichtlinear Funktion;
- Y hängt nicht genau von einer Erklärungsgröße X ab, sondern von mehreren.

Anders ausgedrückt: Die Realität ist in der Regel weder einfach noch linear.

Beide Erweiterungen des Regressionsmodells sind in der Praxis sehr gebräuchlich.

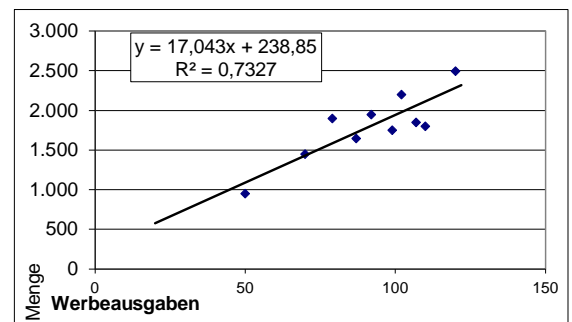
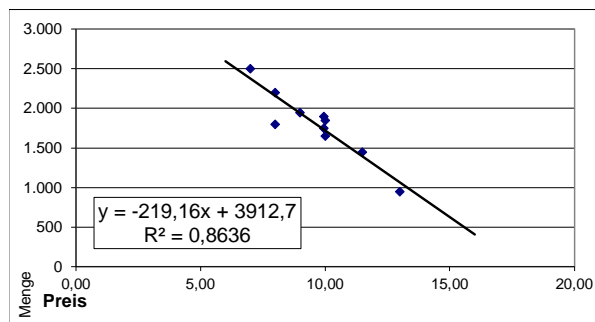
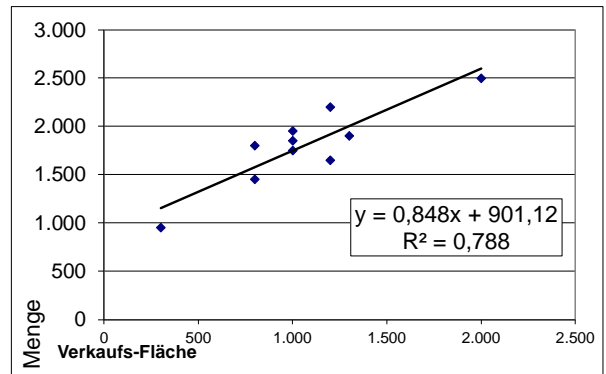
Das „Forum“ möchte eine Ausstellung einer regionalen Künstlerin organisieren und abschätzen, wie viele BesucherInnen sie erwarten können. Sie sprechen mit 10 vergleichbaren regionalen Kulturzentren. Abbildung 13 zeigt dieses Beispiel für einen Zusammenhang zwischen einem abhängigen Merkmal, der Besucherzahl, und drei Einflussfaktoren: der Ausstellungsfläche, den Werbeausgaben und des Eintrittspreises.

Abbildung 13 Regressionsanalyse mit multiplen Einflussfaktoren

Multivariate Zusammenhänge

Besucherzahlen von 10 Ausstellungen

	Besucher / Monat	Ausstellungs-Fläche	Werbe-Ausgaben	Eintritts-Preis
	Stück	qm	€	€
i	y_i	$x1_i$	$x2_i$	$x3_i$
Nr	Absatz	Fläche	Werbung	Preis
1	2.500	2.000	120	7,00
2	1.850	1.000	107	10,00
3	1.750	1.000	99	9,95
4	1.450	800	70	11,50
5	950	300	50	13,00
6	2.200	1.200	102	8,00
7	1.800	800	110	8,00
8	1.950	1.000	92	9,00
9	1.650	1.200	87	10,00
10	1.900	1.300	79	9,95



In Abbildung 13 sind zunächst die Ursprungsdaten und die drei einzelnen Regressionen dargestellt. Anhand der Gütemaße ist zu erkennen, dass der Preis (mit einem R^2 von 86,4 %) den höchsten Erklärungsgrad aufweist, die Ausstellungsfläche ($R^2 = 78,8 \%$) den zweithöchsten und auch die Werbeausgaben ($R^2 = 73,3 \%$) einen messbaren Einfluss auf die Absatzmenge haben.

Die inhaltliche Aussage kann der Steigung der Regressionsgraden bzw. dem Vorzeichen des jeweiligen b entnommen werden: Während die Ausstellungsfläche ebenso wie die Werbeausgaben positiv auf die Absatzmenge wirken, hat der Preis einen negativen Einfluss. Das heißt: je größer die Fläche einer Ausstellung und je höher die dortigen Werbeausgaben, desto höher die Besucherzahl. Je höher jedoch der Preis, desto weniger BesucherInnen werden kommen.

Dies kann auch in einer einzigen, *multiplen* Regression errechnet werden. Die Bestimmungsgleichung für den Absatz Y lautet dann:

*Lineare
Mehrfach-
regression*

$$Y = a + b_1 \cdot X_1 + b_2 \cdot X_2 + b_3 \cdot X_3$$

oder hier: Absatz = a + b_1 * Fläche + b_2 * Werbung + b_3 * Preis

Die Durchführung der multiplen Regression in Excel ergibt das folgende Ergebnis:

$$\text{Absatz} = 2398 + 0,4 \cdot \text{Fläche} + 1,7 \cdot \text{Werbung} - 123,8 \cdot \text{Preis}$$

Damit wird in einer Gleichung das oben dargestellte Ergebnis beschrieben. Eine größere Ausstellungsfläche erhöht den Absatz um das 0,4-fache, also z.B. 100 qm mehr Ausstellungsfläche bringen im Durchschnitt 40 BesucherInnen mehr. Die Erhöhung der Werbeausgaben um 10 Euro erhöht die erwartete Besucherzahl um 17 und eine Senkung des Preises um 1 Euro würde zu einer Erhöhung des Absatzes um knapp 124 BesucherInnen führen. Damit ist auch eine Rangfolge geeigneter Maßnahmen zur Absatzerhöhung erkennbar, die Preissenkung hat in diesem Beispiel die stärkste Wirkung.

Tatsächlich ist die multiple Regression besser als die Einzelregressionen, den hier werden mehr mögliche Einflussfaktoren verwendet und damit ein realistischeres „Modell“ der Realität entworfen. Die bessere Qualität zeigt sich auch im R^2 , das mit 92% höher ist als die einzelnen.

Regressionsanalysen finden in verschiedenen Varianten Anwendung. Die hier besprochene lineare Regression wird oft auch als „KQ-Regression“ als Abkürzung für Kleinste-Quadrate-Regression (weil mathematisch die Abstände zwischen den Beobachtungspunkten und der Regressionsgrade quadriert werden und die Gerade dann so gewählt, dass die Summe dieser quadrierten Abstände möglichst klein wird) oder OLS-Regression von der englischen Bezeichnung „Ordinary Least Squares“-Regression.

Hinweise zur Durchführung von Regressionsrechnungen finden sich in der Excel-Datei.

Bei der Beschreibung mehrerer Merkmale und ihres Zusammenhangs muss zunächst beachtet werden, dass es für verschiedene Skalen auch verschiedene Zusammenhangsmaße gibt. Für metrische und ordinale Merkmale sprechen wir von Korrelation, für nominal skalierte von Kontingenz. Der Korrelationskoeffizient nach Bravais-Pearson beschreibt den linearen Zusammenhang zwischen metrischen Merkmalen, er kann Werte zwischen -1 und 1 annehmen. Beate erkennt, dass es einen starken Zusammenhang zwischen den Werbeausgaben und Besucherzahlen gibt. Die geschätzte Regressionsgerade ist: $Y = 104 + 0,55 * x$, d.h. beispielsweise bei 100 Euro Werbeausgaben eine erwartete Besucherzahl von $Y = 104 + 0,55 * 100$ (Stunden) = 159 BesucherInnen

1.6 Zeitreihen und Indexzahlen

1.6.1 Zeitreihen

Ein weiteres Anwendungsgebiet von Zusammenhangsmaßen ergibt sich, wenn die Entwicklung eines Merkmals im Zeitablauf betrachtet werden soll, also quasi der Zusammenhang zwischen diesem Merkmal und der Zeit.

Eine **Zeitreihe** wird dabei (künstlich) in mehrere Komponenten zerlegt:

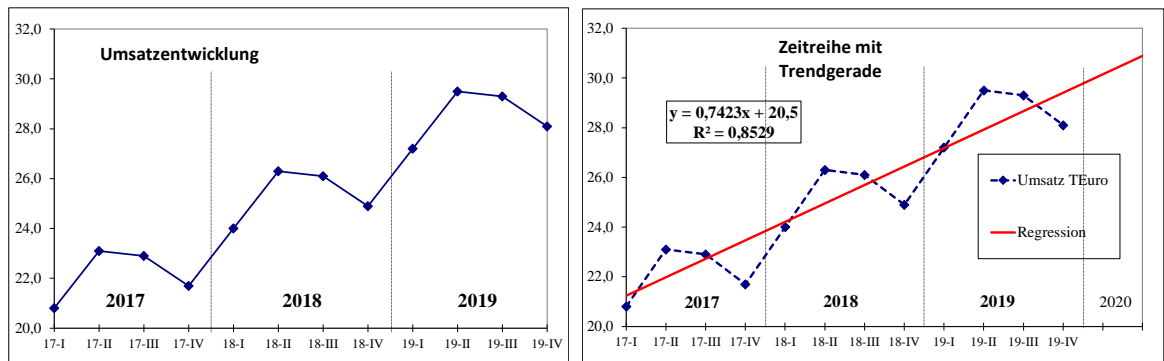
$$Y = \text{Trend-Komponente (+ Konjunktur-Komponente)} + \text{Saison-Komponente} + \text{Rest-Komponente}$$

Beispielsweise seien Umsatzzahlen für die Quartale von 2017 bis 2019 (Tabelle 12) betrachtet, die in Abbildung 14 als Zeitreihe darstellt sind.

Tabelle 12 Zeitreihe der Umsatzentwicklung

Jahr	2017				2018				2019			
Quartal	17-I	17-II	17-III	17-IV	18-I	18-II	18-III	18-IV	19-I	19-II	19-III	19-IV
Zeitpunkt t	1	2	3	4	5	6	7	8	9	10	11	12
y_t	20,8	23,1	22,9	21,7	24,0	26,3	26,1	24,9	27,2	29,5	29,3	28,1

Abbildung 14 Umsatzentwicklung im Zeitablauf



Die linke Grafik in Abbildung 14 zeigt die Entwicklung des Umsatzes in den drei Jahren, wobei erkennbar ist, dass sich in jedem Jahr eine recht regelmäßige Entwicklung wiederholt. Die *saisonale* Komponente zeigt ein Ansteigen des Umsatzes im 1. und 2. Quartal sowie Rückgänge im 3. und 4. Quartal. Um diese Saisoneinflüsse zu bereinigen und den Entwicklungstrend betrachten zu können, wird auch hier eine lineare Regression durchgeführt, deren Ergebnis auf der rechten Seite von Abbildung 14 zu erkennen ist. Die Steigung der Regressionsgerade ist positiv, d.h. der Umsatztrend geht über die drei Jahre nach oben, die Rückgänge in der zweiten Jahreshälfte sind nur *saisonbedingt*.

Glättung
von
Zeitreihen-
werten

Es ist erkennbar, dass die Trendkomponente eine **Glättung** der schwankenden Zeitreihe darstellt und damit eine Referenzgröße für die Ermittlung der Saisoneinflüsse darstellen kann. Als weitere Methode zur Glättung von Zeitreihen sind *Gleitende Durchschnitte* (Moving Average) üblich, bei der aus jeweils vier Quartalswerten ein Mittelwert gebildet wird. Die hier verwendete Methode der linearen Trendfunktion hat dabei den Vorteil, dass der Trend für alle Beobachtungszeitpunkte gebildet werden kann und auch Prognosen über diesen Zeitraum hinaus vorgenommen werden können.

Diese Möglichkeit sowie die Saisonbereinigung sind in der Excel-Datei dargestellt.

1.6.2 Indexzahlen

Index-
zahlen

Indexzahlen (oder Indizes) sind gewichtete arithmetische Mittelwerte aus Messzahlen. Bekannt ist etwa der Preisindex der Lebenshaltung, der durch das Statistische Bundesamt veröffentlicht wird. Hier werden die Preisentwicklungen aller Güter und Dienstleistungen, die Haushalte im Durchschnitt verbrauchen, zu einer mittleren Preissteigerung zusammengefasst. Dabei werden die (relativen) Mengen und daraus folgend die Ausgabenanteile für diese Produkte berücksichtigt.

Bei Zeitreihenanalysen werden anstelle der absoluten Werte oft **Reihen von Indexzahlen** verwendet. Diese werden dadurch gebildet, dass ein Basiszeitraum = 100 (Prozent) gesetzt wird und alle anderen Werte in Bezug auf dieses Basisjahr umgerechnet werden. Entwicklungen von Preisen, Umsätzen, Marktanteilen können damit für verschiedene Merkmale verglichen werden, die eine unterschiedliche absolute Höhe haben und deshalb (z.B. in einer Grafik) nicht „zusammenpassen“.

Besonders beliebt bei den Studierenden sind die original von der englischen Partneruni importierten Scones (mit Lemon Curd). Um deren Preise zu kalkulieren, muss der Wechselkurs des Euro zum britischen Pfundes berücksichtigt werden.

Um die Entwicklung im Verhältnis zum Basisjahr 2020 besser verfolgen zu können, möchte Chris einen Index zur Basis 2020 verwenden. Aber auch die Steigerungen zwischen den Jahren und die durchschnittlichen Steigerungen sind interessant.

Jahr:	2020	2021	2022	2023	2024	2025
Außenwert € gegenüber dem Pfund	0,64	0,76	0,80	0,76	0,88	0,96

Messziffern oder Messzahlen:

X = Reihe von Werten x_t mit $t = 0, \dots, T$.

0 = **Basisperiode** ; $t =$ **Berichtsperiode (laufend)**

Messzahl für die „Periode t zur Basis 0“:

$$M_0^t = \frac{x_t}{x_0} \cdot 100$$

zum obigen Problem:

Jahr:	2020	2021	2022	2023	2024	2025
Außenwert € gegenüber Pfund	0,64	0,76	0,80	0,76	0,88	0,96
a) „2020 = 100“						
b)						
c)						
d)						

a) Ermitteln Sie einen Index zur Basis 2020 (in Tabelle eintragen -- zwei Nachkommastellen genügen)

Rechenweg:

b) Ermitteln Sie die jährlichen Zuwachsraten ... in Prozenten (in Tabelle eintragen -- zwei Nachkommastellen)

c) ... in Prozentpunkten (in Tabelle eintragen -- zwei Nachkommastellen genügen)

d) Gesucht Sie die durchschnittliche Steigerung für den gesamten Zeitraum 2020 bis 2025. Mit welchem Maß wird diese ermittelt - warum?

Umbasierung und Verkettung von Messziffern:

($A =$ altes Basisjahr; $N =$ neues Basisjahr)

$$M_N^t = \frac{M_A^t}{M_A^N} \cdot 100 [\%] \quad \text{oder (wenn } M_A^N \text{ nicht bekannt) } \quad M_N^t = M_A^t \cdot M_N^A / 100$$

oder doch lieber gleich Dreisatz ... ;-)

Für die Kammerstatistik soll folgende Zeitreihe als Indexzahlen zur Basis 2022 angegeben werden:

	2020	2021	2022	2023	2024	2025
alter Index (Basis 2017)	110	112,5	118	115	119	121,5
neuer Index (Basis 2022)						

mit welchem Basisjahr würde die o.a. Entwicklung am besten erkennbar sein?

neuer Index (Basis)

Adam nimmt sich die geschichtsträchtige Aufgabe vor, die Geschäftsentwicklung über 10 Jahre darzustellen. Er entstaubt den Geschäftsbericht 2016 und legt den aktuellen von 2020 daneben. Stellen Sie die Geschäftsentwicklung als eine durchgehende Indexreihe dar:

Geschäftsbericht 2016		Geschäftsbericht 2020		Basis:	
Jahr	Umsatzentwicklung (2020=100)	Jahr	Umsatzentwicklung (2026=100)	Jahr	
2010	100,0	2016	100,0		
2011	102,9	2017	100,5		
2012	110,3	2018	101,5		
2013	114,9	2019	104,2		
2014	114,0	2020	108,0		
2015	112,9				
2016	114,0				

Dies war unser Einstieg zu **Fun with Data.** 😊

Sollten Sie Lust auf mehr oder/und tiefere Einblicke bekommen, sei empfohlen:

Schmidt, Peter, 2019: „Statistik schrittweise verstehen“; Lehr- und Arbeitsbuch für Wirtschafts- und Sozialwissenschaften, UVK Lucius

Ein Index bezieht sich immer auf ein Basisjahr, wogegen eine Zuwachsrate (z.B. Preissteigerung, s.u.) sich jeweils auf das Vorjahr bezieht.

Im untersuchten Fall hat der Index für 2025 zur Basis 2020 den Wert 150.

Die prozentualen Steigerungen (Zuwachsraten) liegen zwischen -5,0% und 18.8% mit einer durchschnittlichen Steigerung über den gesamten Zeitraum von 8,45%.

Durch Umbasierung kann ein anderes Basisjahr verwendet werden, z.B. um mehrere Indexreihen zusammenzufassen.

1.6.3 Praxiswissen

Preisindizes sind ein sehr wichtiges Thema in der Praxis. Wir können dort einerseits die Preisentwicklung für uns selbst verfolgen (Harmonisierter Verbraucher Preis Index, HVPI), sondern auch die Entwicklung von Preisen, die für Unternehmen und die Gesamtwirtschaft wichtig sind, z.B. den „BIP-Deflator“. Beispiele:

- [Preise im Überblick](https://www.destatis.de/DE/Themen/Wirtschaft/Preise/_inhalt.html) (Statistisches Bundesamt)
https://www.destatis.de/DE/Themen/Wirtschaft/Preise/_inhalt.html
- [Preisindex der Lebenshaltung](https://www.destatis.de/DE/Themen/Wirtschaft/Preise/Verbraucherpreisindex/_inhalt.html) des Statistischen Bundesamtes, kann auch auf Inflationsraten umgeschaltet werden
https://www.destatis.de/DE/Themen/Wirtschaft/Preise/Verbraucherpreisindex/_inhalt.html
- [Verbraucherpreisindex: Warenkorb und Wägungsschema](https://www.destatis.de/DE/Themen/Wirtschaft/Preise/Verbraucherpreisindex/inflation.html) (Statistisches Bundesamt)
https://www.destatis.de/DE/Themen/Wirtschaft/Preise/Verbraucherpreisindex/inflation.html
- [Persönlicher Inflationsrechner](https://www.destatis.de/DE/Service/Statistik-Visualisiert/persoenerlicher-inflationsrechner-uebersicht.html) (Statistisches Bundesamt)
https://www.destatis.de/DE/Service/Statistik-Visualisiert/persoenerlicher-inflationsrechner-uebersicht.html
- Preisentwicklung für einzelne Gütergruppen sehr anschaulich dargestellt: [Preis-Kaleidoskop](https://www.destatis.de/DE/Themen/Wirtschaft/Preise/Verbraucherpreisindex/PreisKaleidoskopUebersicht.html). (Statistisches Bundesamt) *https://www.destatis.de/DE/Themen/Wirtschaft/Preise/Verbraucherpreisindex/PreisKaleidoskopUebersicht.html*
- Die Preisentwicklung häufig gekaufter Produkte können Sie auch im [Preismonitor](http://www.destatis.de/DE/Themen/Wirtschaft/Konjunkturindikatoren/Preismonitor/Preismonitor.html) beobachten (Statistisches Bundesamt)
www.destatis.de/DE/Themen/Wirtschaft/Konjunkturindikatoren/Preismonitor/Preismonitor.html
- [Verbraucherpreisindex und Inflationsrate](http://www.destatis.de/DE/ZahlenFakten/GesamtwirtschaftUmwelt/Preise/Verbraucherpreisindizes/WarenkorbWaegungsschema/VPIVideo_.html) (Video, Statistisches Bundesamt)
www.destatis.de/DE/ZahlenFakten/GesamtwirtschaftUmwelt/Preise/Verbraucherpreisindizes/WarenkorbWaegungsschema/VPIVideo_.html

Ebenfalls sehr wichtig im Wirtschaftsbereich sind **Aktienindizes**:

Ein Aktienindex ist eine Kennziffer, mit der die Kursentwicklung dargestellt werden. Es gibt diverse Aktienindizes, die jeweils eine vorgegebene Anzahl an Aktien enthalten und ein gewichtetes arithmetisches Mittel (vgl. Abschnitt **Fehler! Verweisquelle konnte nicht gefunden werden.**) der Entwicklung der Einzelwerte darstellen.

In Deutschland ist der DAX (Deutscher Aktienindex) der wichtigste Index, international der Dow Jones Index aus den USA, der FTSE usw. Im Netz finden sich viele Auflistungen mit den jeweils aktuellen Kursen, wie z.B. bei www.finanzen.net/indizes, www.onvista.de/aktienkurse, www.finanznachrichten.de/aktienkurse/uebersicht.htm, www.boersennews.de u.v.m.

Aktuelles zum DAX z.B. bei der [Deutsche Börse Group](http://www.deutsche-boerse.com) (www.deutsche-boerse.com)

2 LITERATURHINWEISE UND WEITERE INFORMATIONEN

Aus der großen Menge guter Literatur zu Statistik bzw. Evaluation (in der Kultur) seien einige genannt, die jeweils praktische Einführungen in die betriebliche Anwendung darstellen:

Birnkrant, Gesa, 2011: Evaluation im Kulturbetrieb. VS Verlag, Wiesbaden.

Bourier, Günther, 2018: „Beschreibende Statistik“ und „Wahrscheinlichkeitsrechnung und Schließende Statistik“

Glogner-Pilz, Patrick und Föhl, Patrick (Hrsg.), 2016: Handbuch Kulturpublikum - Forschungsfragen und -befunde, Springer Fachmedien, Wiesbaden

Hennefeld, Vera und Reinhard Stockmann (Hrsg.), 2013: Evaluation in Kultur und Kulturpolitik, Waxmann

Hoffmann, Andreas, Astrid Kurzeja, Jutta Schmidt und Peter Schmidt (2017): „Mit interdisziplinären Veranstaltungsprogrammen neue und junge Besucher für Museen und Ausstellungshäuser gewinnen“, Zeitschrift für Kulturmanagement, Vol 3, Nr. 2 2017

Kirchhoff, Sabine u.a. 2010: Der Fragebogen. 5. Überarbeitete Auflage, VS Verlag, Wiesbaden.

Matthäus, Wolf-Gert und Jörg Schulze, 2005: "Statistik mit Excel - Beschreibende Statistik für jedermann"

Munro, Patricia u.a. 2009: Wegweiser Evaluation. Oekom, München.

Puhani, Josef, 2008: "Statistik - Einführung mit praktischen Beispielen"

Renz, Thomas, 2016: Nichtbesucherforschung. Transcript Verlag, Bielefeld.

Schmidt, Peter, 2014: „Statistik Formeln“; Formelsammlung der wichtigsten Formeln der deskriptiven und induktiven Statistik, UVK Lucius

Schmidt, Peter, 2019: „Statistik schrittweise verstehen“; Lehr- und Arbeitsbuch für Wirtschafts- und Sozialwissenschaften, UVK Lucius

Stockmann, Reinhard (Hrsg.), 2007: Handbuch zur Evaluation. Waxmann Verlag, Münster .

Spaß an der Statistik und trotzdem - oder eben deshalb - viele interessante Informationen rund um das Thema und seine Anwendungen finden sich in:

Krämer, Walter: „So lügt man mit Statistik“ und: „Statistik verstehen“

Praktische Arbeit am PC wird mit diesem sehr empfehlenswerten Buch erleichtert, da es neben dem theoretischen Hintergrund auch die praktische Umsetzung in Excel zeigt, Text und Hinweise auf CD-ROM mitliefert und vor allem durch den erfrischenden Schreibstil die statistische Arbeit zur Freude macht:

Monka, Michael und **Voß**, Werner: "Statistik am PC - Lösungen mit Excel"

Weitere Informationen zur (amtlichen) Statistik sowie interessante Datengrundlagen können im Internet gefunden werden. Wichtige Web-Adressen mit (betriebs-) wirtschaftlich relevanten Informationen finden sich z.B. auf meiner Webseite: schmidt-bremen.de.

3 SCHLAGWORTINDEX

Deskriptive Statistik.....	4	Korrelationskoeffizient	14	Stichprobe	4
Diskrete Merkmale	7	Merkmale	5	Streuung	12
Glättung.....	20	Mittelwerte	11, 15	Streuungsmaße	
Gütemaß	17	arithmetisches Mittel	11	DAA	12
Häufigkeiten		geometrisches Mittel.....	11	Spannweite	12
Histogramm	10	Median.....	11	Standardabweichung.....	12
prozentuale Häufigkeiten .	9	Modus.....	11	Varianz	12
relative Häufigkeiten	9	Regression		<u>Umbasierung</u>	21
Summenhäufigkeit	10	multiple Regression	18	Zeitreihe	19
Indexzahlen	20	Regressionsanalyse.....	17	Trend(gerade)	20
Induktive Statistik.....	4	Regressionsgrade.....	16	Zusammenhangmaße	15
Klassenbildung	9	Schwankung	12	Zuwachsrate	21
Kontingenzkoeffizient	15	Skalierung	7		
		Stetige Merkmale	7		