

Werner Fricke

Statistik in der Arbeitsorganisation

5. Auflage



HANSER

REFA-Fachbuchreihe Unternehmensentwicklung

Dr.-Ing. Werner Fricke

Statistik in der Arbeitsorganisation

REFA-Fachbuchreihe Unternehmensentwicklung

Dr.-Ing. Werner Fricke

Statistik in der Arbeitsorganisation

5. Auflage

HANSER

Die Deutsche Bibliothek -CIP Einheitsaufnahme
Ein Titelsatz für diese Publikation ist bei der Deutschen Bibliothek erhältlich.

PRINT ISBN 978-3-446-47270-9
E-PDF ISBN 978-3-446-47271-6

5. Auflage

© Copyright 2021 by REFA Bundesverband e.V., Darmstadt.

Nachdruck oder fotomechanische Wiedergabe – auch auszugsweise – verboten.

Printed in Germany.

Druck: Beltz Bad Langensalza GmbH, Bad Langensalza

Vorwort des Herausgebers

Statistische Daten und Angaben sind in der heutigen Arbeitswelt nicht nur weit verbreitet und vielseitig verwendbar, sie sind und bleiben offenkundig auch notwendig. Ein spezielles Anwendungsgebiet stellt über Jahrzehnte die Arbeitswirtschaft dar. Dies hat sich mit deren Weiterentwicklung zur prozessorientierten Arbeitsorganisation keineswegs verringert. Im Gegenteil – es treten weitere Aufgaben und neue Nutzer auf, z.B. in Verbindung mit selbstverantwortlichen Unternehmenseinheiten wie Segmenten, bei sich durch ihre Arbeit selbststeuernde Gruppen und Teams oder bei der Verwendung von Zielvereinbarungen.

Für Gewinnung und Nutzung gerade arbeitsbezogener Daten und Informationen haben statistische Methoden, Regeln und Kriterien nach wie vor eine spezielle Bedeutung. Einerseits geht es stets um die Aussagekraft und Zuverlässigkeit arbeitsbezogener Daten, andererseits um die Beeinflussung des mit ihrer Ermittlung und Verarbeitung verbundenen, nicht immer geringen Aufwandes.

Bedenkt man in diesem Zusammenhang auch, welches Risiko damit verbunden sein kann, aus einer relativ geringen Datenmenge – der Stichprobe – Ableitungen für weitere Entwicklungen oder für andere Bedingungen vorzunehmen, so wird die Notwendigkeit der Kenntnis sowie des sachgerechten Gebrauchs statistischer Methoden erkennbar.

Hierbei ist eigentlich unerheblich, ob es sich im einzelnen um folgendes handelt:

- Ermittlung der tatsächlichen Auslastung von Anlagen oder Arbeitssystemen
- Erfassung der Häufigkeit und Dauer von Störungen und Ausfällen sowie die Bestimmung ihrer Auswirkungen für das Unternehmen
- Bestimmung von Risiken, etwa von Unfallrisiken oder Expositionen
- Fehleranalyse und -verfolgung
- Erfassung von Arbeitsaufwandsdaten.

Statistische Kenntnisse sind auch zum Vorbereiten, Durchführen und Auswerten von Befragungen der Kunden oder der Beschäftigten sowie von Messungen erforderlich. Schließlich geht es nicht nur darum, Sachverhalte und Zusammenhänge darzustellen, sondern auch darum, die Ursachen solcher Symptome möglichst treffsicher aufzudecken.

In diesem Band werden die Grundlagen aller statistischen Anwendungen in den REFA-Ausbildungen vermittelt. Die Aktualität dieser Ausbildung wird nicht zuletzt dank der Verwendung einschlägiger Software demonstriert.

Das Buch ist Teil einer dreibändigen Reihe, zu der weiter die Titel „Anwendung statistischer Qualitätsmethoden“ und „Führungsorientiertes Qualitätsmanagement“ gehören. Diese neuen Produkte sind Ergebnis angestrenzter Arbeit erfahrener Fachleute im Rahmen eines REFA-Entwicklungsprojektes. Sie werden dazu beitragen, die Attraktivität der REFA – Aus- und Weiterbildung in einem für die Wirtschaft wichtigen Arbeitsfeld deutlich zu steigern.

Prof. Dr.- Ing. habil. E. Kruppe
Vorsitzer REFA-Entwicklungsausschuss

Vorwort

In allen Bereichen der betrieblichen Praxis werden Daten durch Messungen, Erhebungen oder andere Experimente erfasst. Zur anschaulichen Darstellung, Auswertung und Interpretation dieser Daten sind statistische Methoden unerlässlich. Die Vergangenheit hat gezeigt, dass theoretische Vermittlung und praktische Anwendung dieses Stoffes häufig noch nicht ausreichend auf die Bedürfnisse der Lernenden ausgerichtet sind. Hier soll der Versuch unternommen werden, durch verständliche Inhalte, in Verbindung mit professionellen Softwareprodukten und modernen Lehrmethoden, diesen Nachteil zu vermeiden. Deshalb werden alle in diesem Buch verwendeten **Beispiele** und **Abbildungen** mit Hilfe einer **Demonstrationssoftware**, die Sie über untenstehenden Link erreichen, detailliert dargestellt und begreifbar gemacht. Wenn zu einem Beispiel eine entsprechende Datei vorhanden ist, dann steht der Name der Datei in der Beispielüberschrift. Außerdem enthält der Link eine Powerpoint-Demo mit entsprechenden Abbildungen.

Download:

<https://www.refa-verband.de/downloads/SidA-Download.zip>

Dieses Buch soll dazu dienen, Ihnen ein Grundverständnis für statistische Methoden zu vermitteln. Die beschriebenen Methoden können in der Praxis u.a. auf folgenden Gebieten zum Einsatz kommen:

- Untersuchung der Eigenschaften von Zulieferteilen
- Überwachung von Qualitätsstandards und Einhaltung von Gütevorschriften
- Prozessorientiertes Qualitätsmanagement
- Zeitwirtschaft und Planzeitbildung zur Zeit- und Kostenkalkulation
- Nutzung von Betriebsmitteln
- Personal- und Materialbedarfsplanung
- Ermittlung von Verteilzeitanteilen
- Wirkungen bzw. Nebenwirkungen von Behandlungsmethoden
- Erstellung von Unfallstatistiken
- Ermittlung von Schadenshäufigkeiten
- Analyse des Krankenstands eines Unternehmens.

An dieser Stelle möchte ich allen danken, die mit ihrer Hilfe und ihren Anregungen konstruktiv am Zustandekommen dieses Buches beteiligt waren. Besonders möchte ich Herrn Prof. Dr. Eberhard Kruppe für die stets konstruktive Zusammenarbeit bei der Durchsicht der Entwürfe und seine Anregungen danken. Weiterhin gilt mein Dank Herrn Konrad Siegel für die effektive Zusammenarbeit beim Aufbau der GEONExT-Beispiele und meiner Frau Bettina für die fachlich kompetente Durchsicht und Korrektur der Manuskripte.

Inhaltsverzeichnis

1	Einleitung.....	13
1.1	Was ist Statistik.....	13
1.2	Statistik im REFA-Bereich.....	13
1.3	Erhebung von Stichproben.....	13
1.3.1	Bewusste Auswahl.....	14
1.3.2	Zufallsauswahl	16
1.4	Merkmalsausprägungen.....	18
2	Grundbegriffe der Wahrscheinlichkeitsrechnung.....	20
2.1	Grundlagen	20
2.2	Binomial- und hypergeometrische Verteilung.....	23
2.2.1	Binomialverteilung.....	23
2.2.2	Hypergeometrische Verteilung	28
3	Analyse qualitativer Daten (Zählstichproben)	30
3.1	Häufigkeit von Merkmalsausprägungen	31
3.2	Vertrauensbereich für den Anteilswert	32
3.3	Vertrauensbereich für die Differenz zweier Anteilswerte.....	36
4	Beschreibende (deskriptive) Statistik	37
4.1	Grundlegende Begriffe	37
4.2	Darstellung statistischer Daten mit Häufigkeiten	37
4.3	Lagemaße von Häufigkeitsverteilungen	45
4.3.1	Modalwert	45
4.3.2	Zentralwert oder Median	45
4.3.3	α -Quantil	47
4.3.4	Arithmetischer Mittelwert.....	49
4.3.5	Geometrischer Mittelwert	56
4.3.6	Gegenüberstellung Zentralwert – arithmetischer Mittelwert	57
4.4	Streuemaße von Häufigkeitsverteilungen	59
4.4.1	Spannweite, mittlere Spannweite und Streuzahl.....	59
4.4.2	Quartilsabstand QA.....	62
4.4.3	Durchschnittliche absolute Abweichung.....	63
4.4.4	Varianz, Standardabweichung und Variationszahl	63

5	Verteilungen und ihre Eigenschaften	68
5.1	Gaußsche Normalverteilung	68
5.2	t-Verteilung	72
5.3	χ^2 -Verteilung	74
5.4	F-Verteilung	75
5.5	Weitere Verteilungen von Messwertreihen	76
5.5.1	Allgemeines zu Verteilungsmodellen	76
5.5.2	Verschiebung und Transformation von Messwertreihen	78
5.5.3	Logarithmische Normalverteilung	82
5.5.4	Betragsverteilung 1. Art	83
5.5.5	Betragsverteilung 2. Art (Rayleigh-Verteilung)	84
5.5.6	Weibull-Verteilung	87
5.5.7	Pearson-Funktionen	89
5.5.8	Johnson-Transformationen	89
5.5.9	Mischverteilungen	90
5.5.10	Zweidimensionale Normalverteilung	91
5.6	Auswahl von Verteilungsmodellen	92
6	Schließende Statistik	94
6.1	Stichprobe und Grundgesamtheit	94
6.2	Vertrauensbereich für den arithmetischen Mittelwert	95
6.3	Statistische Testverfahren	103
6.3.1	Allgemeines zu Testverfahren	103
6.3.2	Zweiseitige und einseitige Tests	105
6.3.3	Allgemeine Vorgehensweise bei der Testdurchführung	107
6.3.4	Test des arithmetischen Mittelwertes	107
6.3.5	Test der Standardabweichung	108
6.3.6	Test für die Differenz zweier arithmetischer Mittelwerte	109
6.3.7	Test für die Differenz zweier Varianzen (F-Test)	110
6.3.8	Test auf Zufälligkeit (nach Swed und Eisenhart)	111
6.3.9	Test auf Normalverteilung	116
6.3.9.1	Anpassungstest nach Kolmogoroff und Smirnov ($n \leq 50$)	116
6.3.9.2	Chi-Quadrat-Anpassungstest ($n > 50$)	119
6.3.10	Test auf Ausreißer	123
6.3.10.1	Test auf Ausreißer nach David, Hartley und Pearson	123
6.3.10.2	Test auf Ausreißer (nach Grubbs)	126
6.3.10.3	Kritische Diskussion des Ausreißerproblems	127

Inhaltsverzeichnis

7	Regressionsanalyse	129
7.1	Einflussgrößen	130
7.2	Lineare Regression mit einer Einflussgröße	132
7.3	Korrelation und Bestimmtheitsmaß	136
7.3.1	Korrelation.....	136
7.3.2	Bestimmtheitsmaß	142
7.4	Nichtlineare Regression mit einer Einflussgröße	146
7.5	Nichtlineare Regression mit mehreren Einflussgrößen	149
7.6	Beurteilung von Regressionsformeln, Testverfahren	151
7.6.1	Residualanalyse.....	152
7.6.2	Test auf Zufälligkeit bei der Regressionsrechnung	154
7.6.3	Test auf Normalverteilung bei der Regressionsrechnung	157
7.6.4	Test auf Ausreißer bei der Regressionsrechnung	158
7.6.5	Weitere Testverfahren bei der Regressionsrechnung	158
7.6.6	Vorgehensweise bei der Beurteilung	159
7.6.7	Berücksichtigung der Anzahl Messpunkte	160
7.6.8	Vermeidbare Fehler beim Einsatz der Regressionsanalyse.....	163
7.6.9	Beispiele zur Beurteilung von Regressionsfunktionen	164
7.7	Trendrechnung, Zeitreihenanalyse	183
7.7.1	Methode der gleitenden Durchschnitte	183
7.7.2	Trendextrapolation	186
8	Anhang.....	190
8.1	Tabellen	190
8.2	Formelsammlung	201
8.3	Statistische Methoden in der Arbeitsorganisation	203
8.4	Griechisches Alphabet	209
9	Literaturverzeichnis	210
10	Bilderverzeichnis	211
11	Stichwortverzeichnis	214

1 Einleitung

1.1 Was ist Statistik

Der Begriff „Statistik“ wird in vielen Bereichen verwendet, z.B. Wirtschaft, Medizin, Verkehr, Ingenieurwesen, Physik, Chemie usw. Mit diesem Begriff werden häufig Zahlenkolonnen in Tabellen oder auch Grafiken in Verbindung gebracht. Man kennt die Arbeitslosenstatistik, die Unfallstatistik oder die Bevölkerungsstatistik. In diesem Sinne wurde auch der Begriff „Statistik“ gegen Ende des 17. Jahrhunderts geprägt und bedeutete allgemein „die Lehre von der Zustandsbeschreibung des Staates“ (vgl. auch Hartung / 1 / und REFA / 2 /).

Nach REFA / 2 / geht die moderne Statistik über diese rein beschreibende Aufgabe hinaus und stellt Methoden zur Verfügung, die zur Datensammlung, zur Aufbereitung und Analyse der Daten dienen. Das Ziel der Verwendung von statistischen Methoden besteht darin, Behauptungen und Hypothesen zu überprüfen und zu einer rationalen Entscheidungsfindung beizutragen. Grundlage für eine statistische Analyse ist stets eine inhaltlich möglich genau abgegrenzte Fragestellung und die Beschreibung des Untersuchungsziels. Hierbei spielen auch wirtschaftliche Überlegungen eine Rolle, z.B. wie groß der Aufwand für die Datenerhebung ist, um statistisch gesicherte Aussagen mit einer bestimmten vorgegebenen Genauigkeit zu erhalten.

1.2 Statistik im REFA-Bereich

Im Bereich der Arbeitsorganisation spielen statistische Methoden in vielerlei Hinsicht eine wichtige Rolle. Dieses Buch soll deshalb auch für REFA-Fachkräfte eine Grundlage für die korrekte Anwendung und Umsetzung dieser Methoden bilden. An geeigneter Stelle wird jeweils auf den Bezug zu REFA hingewiesen. Außerdem werden die für REFA relevanten Methoden und Verfahren in einem eigenen Abschnitt im Anhang 8.3 des Buches zusammengefasst.

1.3 Erhebung von Stichproben

Will man Informationen über bestimmte Problemstellungen erhalten, so sind zunächst Daten zu erheben. Es stellt sich hierbei häufig die Frage, ob es notwendig und sinnvoll ist alle Daten zu erfassen oder ob ein Teil der Daten ausreicht. Sicherlich liefert die Totalerhebung die umfassendsten Informationen über die Problemstellung. In den meisten Fällen ist es jedoch unmöglich, alle Informationen einer Problemstellung zusammenzutragen, weil die Datenmenge einfach zu groß oder unendlich ist.

Selbst wenn alle Daten ermittelt werden könnten, ist häufig der Zeit- und damit der Kostenaufwand so groß, dass dies nicht sinnvoll erscheint. Aus diesem Grund wird man sich in der Regel mit einer Untermenge der gesamten Daten zufrieden geben und aus dieser **Stichprobe** Rückschlüsse auf die Gesamtmenge der Daten, die **Grundgesamtheit**, ziehen.

Bei der Erhebung von Daten ist es notwendig, dass die Stichprobe möglichst alle Eigenschaften der Grundgesamtheit aufweist. Will man z.B. wissen, wie hoch der Frauenanteil an der Gesamtzahl der Studierenden in einem bestimmten Fach ist, so werden natürlich nur diejenigen in die Stichprobe mit einbezogen, die das entsprechende Fach studieren. Auch bei der Auswahl der Elemente der Stichprobe werden verschiedene Verfahren unterschieden.

1.3.1 Bewusste Auswahl

Bei der bewussten Auswahl werden die Einheiten der Grundgesamtheit nach den Kenntnissen über ihre Zusammensetzung, d.h. aufgrund gezielter Überlegungen, in der Stichprobe mit dem Ziel einer möglichst guten Repräsentativität berücksichtigt / 2 /. Die Verfahren der bewussten Auswahl werden häufig in der Marktforschung eingesetzt und zwar immer dann, wenn eine Auswahl nach dem Zufallsprinzip nicht möglich oder zu aufwendig ist. Nach Hartung / 1 / werden diese Verfahren Beurteilungsstichproben genannt. Sie werden wie folgt von den Zufallsstichproben abgegrenzt (vgl. auch Menges/Skala / 3 /, Hauser / 4 /).

Zufallsstichproben	Beurteilungsstichproben
erfordern eingehende Planung	sind einfach zu planen
eine bestimmte Erhebungseinheit muss in die Stichprobe gelangen (Substitutionen sind nicht zugelassen)	Substitutionen der Erhebungseinheiten sind zugelassen
sind relativ teuer	sind relativ billig
da sie auf der Wahrscheinlichkeitsrechnung beruhen, ist eine genaue Fehlerrechnung möglich	da sie nicht auf der Wahrscheinlichkeitsrechnung beruhen, ist keine Fehlerrechnung möglich (der Fehler muss ohne statistische Mittel „beurteilt“ werden)
besitzen eine theoretische Fundierung	sind nicht theoretisch zu fundieren
sind eindeutig besser	sind eindeutig schlechter

Bild 1: Gegenüberstellung von Zufallsstichproben und Beurteilungsstichproben / 1, 3, 4 /

Bei den Beurteilungsstichproben werden folgende Verfahren unterschieden:

- Gezielte Auswahl typischer Fälle (typische Auswahl)
- Auswahl nach dem Konzentrationsprinzip („cut-off“-Verfahren)
- Quotenauswahl

Bei der **typischen Auswahl** (Hartung / 1 /) werden Einheiten bestimmt, die für die Grundgesamtheit als typisch gelten. Dieses „als typisch gelten“ wird vom Bearbeiter anhand geeigneter Auswahlmerkmale bestimmt, die man aus einer hinreichenden Kenntnis über die Grundgesamtheit gewinnt. Soll z.B. die Investitionsbereitschaft von Betrieben einer Branche ermittelt werden, so könnte man diese so auswählen, dass alle ausgewählten Betriebe die gleiche durchschnittliche Größe aufweisen wie alle Betriebe der Branche.

Bei dem **Konzentrationsprinzip** („cut-off“-Verfahren) werden nur die wesentlichen Einheiten in die Stichprobe einbezogen. Unwesentliche, unbedeutende Einheiten werden nicht berücksichtigt, sie werden abgeschnitten (cut-off). Das Verfahren hat den Sinn, die Grundgesamtheit einzuengen, ohne dass dies eine große Auswirkung auf die Aussagen hat. Soll z.B. wieder die Investitionsbereitschaft von Betrieben einer Branche ermittelt werden (vgl. auch Hartung / 1 / und REFA / 2 /) und wird diese hauptsächlich durch einige wenige Großbetriebe bestimmt, so ist es einleuchtend, dass nur diese Betriebe in der Stichprobe berücksichtigt werden.

Bei der **Quotenauswahl** wird die Stichprobe nach bestimmten, bekannten Quoten, also dem Verhältnis, nach dem bestimmte Ausprägungen eines Merkmals in der Stichprobe enthalten sein sollen, ausgewählt. Soll z.B. die Nachfrage nach einem bestimmten Konsumgut ermittelt werden, so kann die Stichprobe geschlechtsspezifisch, nach Altersgruppen oder nach Erwerbstätigkeit quotiert werden (vgl. auch Hartung / 1 / und v. Mises / 5 /).

Da die Beurteilungsstichproben hauptsächlich für Marktforschungsanalysen zum Einsatz kommen und die angegebenen Nachteile besitzen, werden diese in den folgenden Ausführungen nicht weiter berücksichtigt. Für alle weiteren Betrachtungen fordern wir also die Auswahl der Stichprobenelemente nach dem Zufallsprinzip.

1.3.2 Zufallsauswahl

Das Prinzip der Zufallsauswahl bedient sich der Wahrscheinlichkeitsrechnung und erlaubt es, Vertrauensbereiche für die unbekannten Kenngrößen der Grundgesamtheit anzugeben. Es werden folgende Verfahren unterschieden:

- einfache Zufallsstichprobe
- geschichtete Stichprobe
- Klumpenstichprobe

Einfache Zufallsstichprobe: Hierbei wird die Stichprobe so ausgewählt, dass jedes Element die gleiche Chance hat, in die Auswahl zu gelangen. Liegen z.B. 1000 Werkstücke als endliche Grundgesamtheit vor und wir wollen 100 davon zufällig auswählen, so können wir alle Werkstücke von 1 bis 1000 nummerieren. Wir erzeugen nun Zufallszahlen ($0 < Z_i < 1$), multiplizieren diese mit 1000 und runden die Ergebnisse auf. Es ergeben sich Zufallszahlen zwischen 1 und 1000, z.B. wie folgt: 707, 832, 179, 772, 472, 178, 680, 968, 76, ...

Nun werden die Werkstücke mit den entsprechenden Nummern entnommen. Sollte eine Zufallszahl mehrfach erscheinen, wird einfach die Nächste verwendet. In der Literatur werden manchmal noch Tabellen mit Zufallszahlen abgedruckt, in der heutigen Zeit wird man jedoch EDV-Programme verwenden, die entsprechende Zahlen liefern. Das Microsoft Produkt Excel® verfügt z.B. über die Funktion „Zufallszahl()“. Natürlich sind in den beiliegenden Demoversionen ebenfalls Zufallsgeneratoren enthalten.

Geschichtete Zufallsauswahl: Bei einer inhomogenen Stichprobe mit großen Differenzen wird bei einer einfachen Zufallsauswahl die Streuung und damit die Anzahl der notwendigen Messpunkte für eine abgesicherte Aussage sehr groß. In diesem Fall kann die Grundgesamtheit in Schichten aufgeteilt werden, die in sich möglichst homogen, untereinander jedoch heterogen sein sollten.

Bei einer geschichteten Zufallsauswahl wird die Grundgesamtheit vom Umfang N in L Teilmengen (Schichten) vom Umfang N_h zerlegt. Aus jeder Teilmenge wird nun eine Stichprobe von Umfang n_h gezogen und ausgewertet. Das folgende Beispiel soll dies näher erläutern:

1 Einleitung

In einem Lager liegen Artikel mit stark unterschiedlichen Preisen. Ziel ist es, den mittleren Preis je Produkt zu ermitteln. Würde man eine einfache Zufallsstichprobe ziehen, so wäre der Stichprobenumfang verhältnismäßig groß. Folgende Informationen liegen über das Lager vor:

Anzahl der Artikel		$N = 70.000$
Anzahl Artikel im Preissegment	50 bis 100	$N_1 = 37.000$ (ca. 53 %)
Anzahl Artikel im Preissegment	100 bis 200	$N_2 = 23.000$ (ca. 33 %)
Anzahl Artikel im Preissegment	200 bis 500	$N_3 = 10.000$ (ca. 14 %)

Für jede Schicht wurde eine Stichprobe von Umfang $n_1=n_2=n_3=50$ gezogen. Der Gesamtumfang der Stichprobe beträgt also $n=n_1+n_2+n_3=150$. Je Schicht werden z.B. folgende Mittelwerte und Varianzen berechnen:

$$\begin{array}{ll} \bar{x}_1 = 75,62 & s_1^2 = 214,2 \\ \bar{x}_2 = 145,86 & s_2^2 = 788,6 \\ \bar{x}_3 = 356,8 & s_3^2 = 8201,2 \end{array}$$

Man ermittelt nun den mittleren Artikelpreis wie folgt:

$$\bar{x}_{ges} = \frac{N_1}{N} \cdot \bar{x}_1 + \frac{N_2}{N} \cdot \bar{x}_2 + \frac{N_3}{N} \cdot \bar{x}_3 = 0,53 \cdot 75,62 + 0,33 \cdot 145,85 + 0,14 \cdot 356,8 = 138,2$$

Klumpenstichprobe: Bei der Durchführung dieses Verfahrens wird die Grundgesamtheit in Klumpen (engl. = cluster) zerlegt. Danach werden nach dem Zufallsprinzip einige dieser Klumpen ausgewählt. Erhebt man nun alle Merkmale der ausgewählten Klumpen, so spricht man von einem einstufigen Verfahren, wählt man jedoch innerhalb der Klumpen nochmal nach dem Zufallsprinzip aus, so spricht man von einem mehrstufigen Verfahren. Bei diesem Verfahren sollte jeder Klumpen die Grundgesamtheit im Bezug auf das interessierende Merkmal möglichst gut widerspiegeln.

Mehr über die Erhebung von Stichproben erfahren Sie in Hartung / 1 / Kapitel V „Aspekte der Datengewinnung“.

1.4 Merkmalsausprägungen

Bei der Ermittlung von Merkmalen einer Stichprobe können unterschiedliche Methoden zur Anwendung kommen, z.B.:

- Befragung
- Beobachtung von Merkmalen (z.B. Multimomentaufnahme)
- Selbstaufschreibung
- Zählen, Messen oder Schätzen von quantitativen Merkmalen

Je nach angewandter Methode sind natürlich unterschiedliche Messgenauigkeiten und -fehler zu erwarten. Die Auswahl der richtigen Methode sollte sich dabei am späteren Verwendungszweck der Daten orientieren. Will man z.B. die Abmessungen von hochgenauen Einbauteilen messen, wird es nicht ausreichen, ein Lineal zu verwenden, sondern man wird sicherlich Messsysteme zum Einsatz bringen, die eine höhere Messgenauigkeit aufweisen.

Die Objekte (Personen oder Systeme), die befragt oder beobachtet oder an denen Messungen vorgenommen werden, nennt man auch **Untersuchungseinheiten** / 1 /. Bei Multimomentaufnahmen wird auch noch der Begriff **Beobachtungssystem** verwendet. Die Größen oder Eigenschaften, auf die sich die Fragen oder Messungen beziehen, **heißen Merkmale**. Es werden die in Bild 2 dargestellten Merkmalsarten unterschieden.

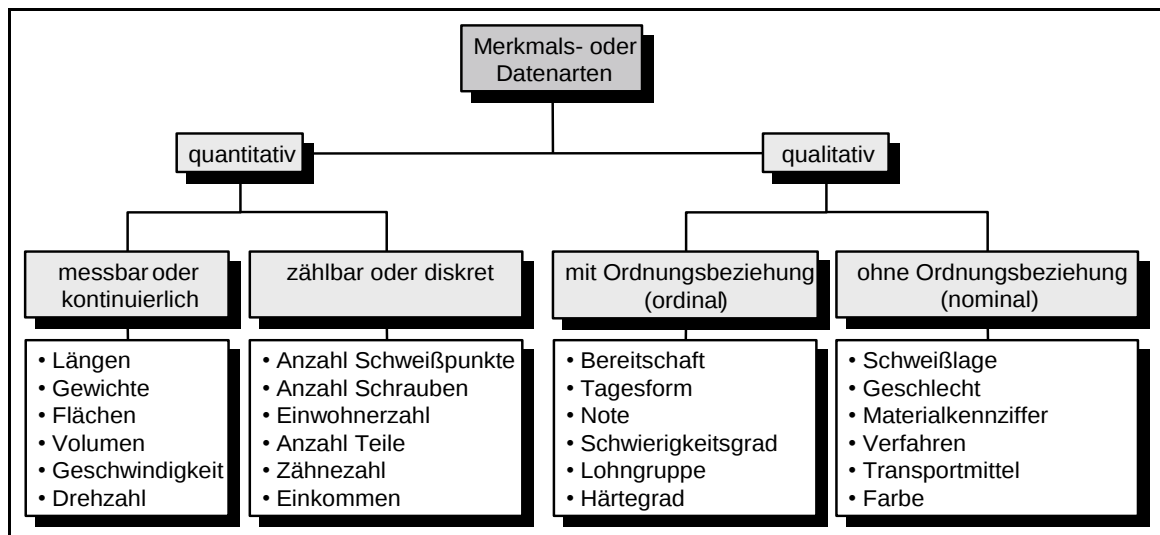


Bild 2: Einteilung der Merkmals- oder Datenarten (quantitativ / qualitativ)

1 Einleitung

Messbare, kontinuierliche oder auch stetige Merkmale können auf einer Skala jeden beliebigen Wert annehmen. Zwischen zwei Werten kontinuierlicher Merkmale liegen theoretisch unendlich viele weitere Werte.

Zählbare, diskrete Merkmale können auf einer Skala nur ganzzahlige Werte annehmen. Zwischen zwei aufeinanderfolgenden Werten ist kein weiterer Beobachtungspunkt möglich. Die Gesamtzahl diskreter Merkmale kann maximal abzählbar unendlich sein.

Qualitative Merkmale mit Ordnungsbeziehung (ordinal) unterliegen einer gewissen Rangfolge. Die Abstände zwischen den Werten sind dabei unterschiedlich und nicht interpretierbar. Bei Schulnoten etwa kann man sagen, dass ein „gut“ besser ist als ein „befriedigend“, der Notenabstand kann jedoch nicht interpretiert werden, weil die Merkmale nicht eindeutig bestimmten Leistungen zugeordnet werden können.

Qualitative Merkmale ohne Ordnungsbeziehung (nominal) unterliegen keiner Rangfolge. Es ist nicht möglich, zu sagen, dass die Farbe grün größer oder kleiner ist als die Farbe gelb.

In der Literatur werden die quantitativen Merkmale manchmal auch metrisch skalierte Merkmale genannt. Diese werden teilweise auch noch nach intervall-skalierten (ohne absoluten Nullpunkt) und verhältnisskalierten (mit absolutem Nullpunkt) Merkmalen unterschieden.

2 Grundbegriffe der Wahrscheinlichkeitsrechnung

2.1 Grundlagen

Die Wahrscheinlichkeitsrechnung beschäftigt sich mit dem Auftreten zufälliger Ereignisse und ist nach *Laplace* (1749 bis 1827) wie folgt definiert:

Liegt eine Grundmenge Ω (Grundraum, Ereignisraum) von unbeschränkt wiederholbaren Versuchen oder Stichproben vor und sind die Teilmengen von Ω zufällige Ereignisse E , so ist die Wahrscheinlichkeit P (P engl. = probability) für das Eintreten des Ereignisses E wie folgt definiert:

$$P_{(E)} = \frac{g}{m}$$

g : Anzahl der günstigen Ereignisse

m : Anzahl der möglichen Ereignisse

Ein Ereignis, das nur aus einem Element besteht, nennt man Elementarereignis.

Beispiel 1: einfacher Würfelwurf $E = \{2, 4, 6\}$

Ereignisraum $\Omega = \{1, 2, 3, 4, 5, 6\}$

Elementarereignisse $\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}$.

Gesucht: Wahrscheinlichkeit für das Auftreten einer geraden Zahl: $E = \{2, 4, 6\}$

$$g = 3, m = 6 \Rightarrow P_{(E)} = 0,5 = 50\%$$

Nach v. Mises / 5 / wird die Wahrscheinlichkeit auch als Grenzwert der relativen Häufigkeit wie folgt definiert:

$$P_{(E)} = \lim_{n \rightarrow \infty} \frac{n}{m}$$

n : Anzahl der Beobachtungen des Ereignisses E

m : Anzahl der Beobachtungen insgesamt

Dieser Grenzwert wird auch als **statistische Wahrscheinlichkeit** bezeichnet.

2 Grundbegriffe der Wahrscheinlichkeitsrechnung

Beispiel 2: einfacher Würfelwurf $E = \{1\}$

Ereignisraum $\Omega = \{1, 2, 3, 4, 5, 6\}$

Elementarereignisse $\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}$,

Gesucht: Wahrscheinlichkeit für das Auftreten einer 1: $E = \{1\}$

Folgende Würfe wurden durchgeführt:

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
P %	100	50	67	50	60	50	57	63	67	60	55	50	46	43	40	38	35	39	37	35
Wurf	1	5	1	5	1	3	1	1	1	3	4	5	3	2	4	4	3	1	5	2
E=1	1		1		1		1	1	1									1		
n	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
P %	33	32	30	33	36	35	33	32	31	30	29	28	27	26	26	25	24	26	26	25
Wurf	6	4	4	1	1	2	2	5	3	6	2	3	5	6	6	5	6	1	2	6
E=1				1	1													1		
n	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
P %	24	24	23	25	24	24	23	23	22	22	22	23	25	26	25	25	25	24	24	23
Wurf	2	3	3	1	4	3	3	2	2	3	4	1	1	1	6	2	6	4	2	3
E=1				1									1	1	1					
n	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80
P %	23	23	22	22	22	21	21	21	20	20	20	19	19	19	20	20	19	19	20	20
Wurf	3	3	2	6	6	2	5	6	2	2	2	2	4	4	1	4	4	6	1	6
E=1															1				1	
n	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
P %	20	20	19	19	19	19	18	18	18	18	18	17	17	17	17	17	16	16	16	16
Wurf	6	5	4	3	4	3	2	5	5	5	5	2	4	5	3	4	3	6	3	2
E=1																				

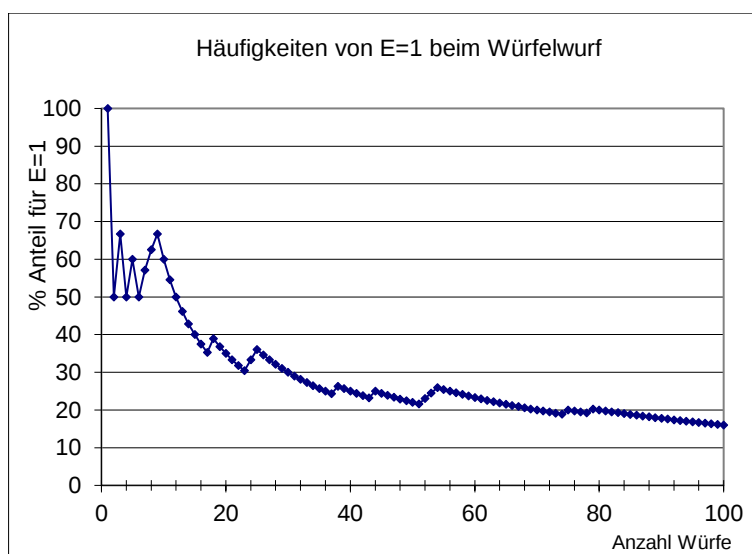


Bild 3: Verlauf der Häufigkeit für das Ereignis $E=\{1\}$ beim Würfelwurf