

2020-04-06 - SOCOS

Qualitätsmanagement in der Bosch-Gruppe | Technische Statistik

1. Grundlagen der Technischen Statistik

Kontinuierliche Merkmale



BOSCH
Technik fürs Leben



Qualitätsmanagement in der Bosch-Gruppe
Technische Statistik

Heft 1 – Grundlagen der technischen Statistik:
Kontinuierliche Merkmale

Ausgabe 01.2016

2020-04-06 - SOCCOS





Ausgabe 01.2016

Vorwort

Bevor der Computer erfunden und komfortable Statistikprogramme entwickelt wurden, mussten die im vorliegenden Heft beschriebenen grafischen Darstellungen und Auswertungen ausnahmslos unter Verwendung spezieller Formblätter und statistischer Tabellen von Hand erstellt werden.

Im Interesse eines tiefgehenden Verständnisses ist es auch heute noch sinnvoll, z.B. im Rahmen des Selbststudiums oder von Schulungen zur Statistik die entsprechenden Vorgehensweisen im Detail nachzuvollziehen. Aus diesem Grund sind die Vorgehensweisen in einer Weise beschrieben, welche die handschriftliche Auswertung ohne Rechnerhilfe ermöglicht.



Inhaltsverzeichnis

1	Einleitung	5
2	Begriffe	7
2.1	Merkmal	7
2.2	Grundgesamtheit	7
2.3	Stichprobe.....	8
2.4	Zufallsvariable.....	8
2.5	Wahrscheinlichkeit.....	10
3	Statistische Kenngrößen	13
3.1	Median.....	13
3.2	Arithmetischer Mittelwert	15
3.3	Gleitender Mittelwert	16
3.4	Geometrischer Mittelwert	17
3.5	Harmonischer Mittelwert.....	18
3.6	Standardabweichung.....	19
3.7	Variationskoeffizient.....	21
3.8	Spannweite	22
3.9	Range-Methode zur Bestimmung der Standardabweichung.....	22
4	Statistische Berechnungen in EXCEL	24
5	Grafische Darstellung von Daten	27
5.1	Urwertdiagramm.....	27
5.2	Strichliste, Punktdiagramm	28
5.3	Klasseneinteilung, Histogramm	28
5.4	Summenkurve	33
6	Statistische Verteilungen.....	34
6.1	Gaußsche Normalverteilung.....	34
6.1.1	Eigenschaften und Kenngrößen der Normalverteilung.....	35
6.1.2	Verteilungsfunktion	37
6.1.3	Standardnormalverteilung.....	39
6.2	Das Wahrscheinlichkeitsnetz der Normalverteilung.....	41
6.3	Logarithmische Normalverteilung	48
6.3.1	Wahrscheinlichkeitsnetz der Lognormalverteilung	50
6.3.2	Zusammenhang zwischen Normalverteilung und Lognormalverteilung	53
6.4	Mischverteilungen	54

2020-04-06 - SOCOS



7	Qualitätsregelkarten	56
7.1	Regelkarten der Lage	56
7.1.1	Mittelwertkarte	57
7.1.2	Urwertkarte (x-Karte)	59
7.2	Regelkarten der Streuung	60
7.2.1	s-Karte	61
7.2.2	R-Karte	62
8	Beurteilung von Häufigkeitsverteilungen im Zusammenhang mit einer Toleranz	63
9	Genauigkeit der Schätzung von Mittelwert und Standardabweichung.....	66
10	Standardnormalverteilung	68
11	Literaturverzeichnis	72
12	Formelzeichen und Begriffe	73
	Index.....	75

2020-04-06 - SOCOS



1 Einleitung

Die mathematische Statistik oder kurz „Statistik“ hat sich ursprünglich aus den Volkszählungen entwickelt, die dazu dienen sollten, den Zustand (lat. „status“) eines Staates festzustellen und z.B. hinsichtlich volkswirtschaftlicher Merkmale zu beschreiben. Laut Duden ist sie die „Wissenschaft von der zahlenmäßigen Erfassung, Untersuchung und Auswertung von Massenerscheinungen“.

Diese Definition beinhaltet zwei wesentliche Aspekte der Statistik: das Erfassen, Ordnen und Darstellen statistischer Daten sind Ziele der deskriptiven (beschreibenden) Statistik, während die Auswertung (Analyse, Deutung) der Daten Aufgabe der induktiven (schließenden) Statistik ist. In den Medien begegnen uns immer wieder Beispiele für entsprechende Anwendungen aus beiden Teilbereichen.

Beispiele für Anwendungen der deskriptiven (beschreibenden) Statistik sind etwa Darstellungen der

- zeitlichen Kursentwicklung von Devisen oder von Aktienindices (Urwertdiagramme)
- Sitzverteilungen in Landtagen (Tortendiagramme)
- Anteile verschiedener Automarken an der Gesamtzahl der in Deutschland neu zugelassenen Pkw in einem Jahr (Histogramme)

oder die Angabe des pro-Kopf-Verbrauchs von Milcherzeugnissen in den EU-Ländern in einem Jahr (Mittelwerte).

Folgende Anwendungsbeispiele für die Vorgehensweise der induktiven Statistik haben noch unverkennbar eine gewisse Verwandtschaft mit Volkszählungen:

- Prognose des Wahlergebnisses am Wahlsonntag auf Grundlage repräsentativer Umfragen,
- Hochrechnungen der Zuschauerzahlen bei Fernsehsendungen auf Grundlage der Sehbeteiligung ausgewählter „Testseher“,
- Schätzung der Besucherzahlen bei (offenen) Großveranstaltungen,
- Schätzung des Bestands einer bestimmten Tierart (Gesamtpopulation) in einem Gebiet bekannter Größe,
- Analyse der Auswirkung von Werbekampagnen auf das Kaufverhalten eines Marktes anhand des Verhaltens ausgewählter Testkunden.

In allen letztgenannten Fällen wird auf Grundlage des Wissens über eine begrenzte Teilmenge von Individuen („Stichprobe“) eine Aussage über ein größeres Kollektiv („Gesamtheit“) abgeleitet.

Dabei wird ausgenutzt, dass bei vielen Massenerscheinungen zwar das Ergebnis einer Einzelbeobachtung (jährliche Anzahl von Blitzeinschlägen pro Quadratkilometer im Gebiet xy) zufällig ist (und somit nicht mit Sicherheit vorhergesagt werden kann), aber dennoch eine mathematisch beschreibbare Gesetzmäßigkeit vorliegt.

Häufig hat die schließende Statistik aber auch zum Ziel, von einem augenblicklichen Zustand („Trend“) auf künftiges Verhalten zu schließen, d.h. in gewisser Weise die Zukunft vorherzusagen.

Die Statistik arbeitet zu diesem Zweck mit mathematischen Modellen (Verteilungsfunktionen) welche die Eigenschaften sogenannter Zufallsvariablen beschreiben.



Missverständnisse bei der Anwendung statistischer Methoden resultieren fast immer aus der Vernachlässigung der Modellbezogenheit und der im Zusammenhang mit den Methoden genannten Voraussetzungen.

Das Verständnis statistischer Aussagen und Methoden wird dem Einsteiger in die Materie durch mehrere, auf der Folgeseite aufgeführte, problematische Gesichtspunkte erschwert.

1. Begriffliche Schwierigkeiten

In der Umgangssprache wird der Begriff „wahrscheinlich“ häufig ersetzt durch andere Begriffe, wie z.B. „unmöglich“, „vielleicht“, „vermutlich“, „mit ziemlicher Sicherheit“, oder „todsicher“, die unserer Erfahrung nach stets ein Maß für das Vertrauen in die Richtigkeit einer Aussage darstellen sollen. Je nach Person, die einen solchen Begriff benutzt, deren Stimmungslage (Euphorie, Depression) und der jeweiligen Situation kann aber jeder einzelne ganz unterschiedliche Bedeutung haben.

Demgegenüber benutzt die Statistik eine mathematisch definierte „Wahrscheinlichkeit“, eine Zahl zwischen null (unmögliches Ereignis) und eins (sicheres Ereignis) als Maß für das voraussichtliche „Eintreten“ oder „Nicht-Eintreten“ eines Ereignisses. Die Schwierigkeiten, den Begriff Wahrscheinlichkeit losgelöst vom „statistischen Rahmen“ mit einfachen Worten zu erläutern, sind nur allzu offensichtlich.

2. Logische Schwierigkeiten

Es besteht grundsätzlich die Gefahr, dass der Anwender der induktiven Statistik den Eindruck einer objektiven „Sicherheit“ erhält, die tatsächlich nicht gegeben ist. Dieses Missverständnis spiegelt sich wider in der Begriffsfolge „unbekannt“, „zufällig“ („Zufallsvariable“), „wahrscheinlich“ („Wahrscheinlichkeit“) die im üblichen Sprachgebrauch häufig im Begriff „sicher“ („Aussagesicherheit“) gipfelt.

Es sollte einleuchten, dass es keine Möglichkeit gibt, in der Realität zwischen den Zuständen „unbekannt“ und „sicher“ eine Brücke zu schaffen.

3. Übertragbarkeit

Die in Statistiklehrbüchern betrachteten Beispiele zeigen, dass man kaum umhin kommt, Naturphänomene und diesbezügliche Messgrößen oder Beispiele aus der Spieltheorie mit bekannten Randbedingungen zur Veranschaulichung zufälligen (chaotischen?) Verhaltens heranzuziehen:

- Zahl der Blitze pro Quadratkilometer Erdoberfläche und Jahr,
- Jährliche Niederschlagsmenge pro Quadratmeter,
- Bewegung von Gasmolekülen (Brownsche Molekularbewegung)
- Radioaktiver Zerfall,
- Gewinnaussichten bei Glücksspielen (Würfeln, Roulette, Lotto).

Im Vergleich zu solchen Beispielen erscheinen die in der industriellen Praxis untersuchten Phänomene kaum mit den Begriffen „zufällig“ oder gar „chaotisch“ vereinbar zu sein.

Trotz dieser grundsätzlichen Probleme haben sich statistische Methoden in der industriellen Praxis einen festen Platz erobert.

Das vorliegende Heft bildet den „Einstieg“ in die Schriftenreihe „Qualitätssicherung in der Bosch-Gruppe, Technische Statistik“, die eine Vielzahl von Spezialthemen abdeckt.



2 Begriffe

2.1 Merkmal

Gegenstand statistischer Überlegungen und Berechnungen in der industriellen Praxis sind in der Regel kontinuierlich veränderliche, messbare Merkmale und diskrete, abzählbare Merkmale von Betrachtungseinheiten. Entsprechend diesen Untergruppen tragen die ersten beiden Hefte der Bosch-Schriftenreihe „Qualitätssicherung in der Bosch-Gruppe“ die Untertitel „Kontinuierliche Merkmale“ (Heft Nr. 1) und „Diskrete Merkmale“ (Heft Nr. 2).

Die im vorliegenden Heft 1 behandelten kontinuierlichen Merkmale sind messbare oder beobachtbare Eigenschaften (Länge, Gewicht, Temperatur) von Objekten oder Ereignissen (Lebensdauer, Berstdruck).

Obwohl eine physikalische Messgröße stets in Form einer Maßzahl (z.B. 48) und einer Maßeinheit (z.B. mm) angegeben wird, sind für statistische Betrachtungen die Maßeinheiten von untergeordneter Bedeutung. Wir können uns daher in den Beispielen meist auf die reinen Maßzahlen konzentrieren.

Statistische Analysen der Eigenschaften kontinuierlicher Merkmale finden in vielen Bereichen der industriellen Praxis Anwendung, z.B. bei der

- Untersuchung der Fähigkeit von Messgeräten und Maschinen,
- Beurteilung von Fertigungsprozessen,
- Statistischen Prozessregelung (SPC),
- Auswertung von Versuchsdaten.

Die Verfahren der induktiven Statistik sind insbesondere im Zusammenhang mit Risikoanalysen von Interesse und natürlich dort, wo aus Gründen der Wirtschaftlichkeit mit relativ kleinen Stichprobenumfängen gearbeitet werden muss, z.B. im Bereich aufwendiger (teurer) Qualitätsprüfungen (zerstörende Prüfung, Lebensdaueruntersuchungen).

2.2 Grundgesamtheit

Unter einer Grundgesamtheit verstehen wir eine endliche oder unendliche Menge von Betrachtungseinheiten, die bzgl. einer vorliegenden statistischen Fragestellung als gleichartig zu betrachten sind. Solche Betrachtungseinheiten können beispielsweise auch „Beobachtungen“ oder Ergebnisse von unter gleichen Bedingungen durchgeführten „Versuchen“ sein.

Beispiele für endliche Grundgesamtheiten sind die Menge der

- Schüler einer Schule,
- wahlberechtigten Personen eines Bundeslandes,
- Fernsehzuschauer, die das Endspiel der letzten Fußball-WM verfolgt haben,
- Teile einer Warenlieferung,
- innerhalb einer Schicht gefertigten Erzeugnisse des Werks XY.



Beispiele für (theoretisch) unendliche Grundgesamtheiten sind die Menge der

- bei Würfeln mit einem Würfel beobachteten „Augenzahlen“,
- bei Wiederholmessungen an einem Längennormal ermittelten Messergebnisse,
- Teile, die eine Maschine herstellen wird, unter der Annahme, dass sie für alle Zeiten in dem augenblicklich vorliegenden Zustand bleibt.

Insbesondere die letzten Beispiele zeigen, dass eine Grundgesamtheit durchaus nicht immer real sein muss sondern auch fiktiv sein kann. Darüber hinaus ist erkennbar, dass eine statistische Fragestellung mitunter auf eine Prognose (Vorhersage) zukünftiger Ereignisse abzielen kann.

2.3 Stichprobe

Eine Stichprobe ist dagegen stets eine reale und damit endliche Menge von „Dingen“ oder Ereignissen. Beispiele hierfür sind die Menge der

- Fahrzeuge, die den Engelbergtunnel (bei Leonberg) am 1.5.2015 passiert haben,
- beobachteten Wurfresultate bei 10 Würfeln mit einem Würfel,
- bei 25 Wiederholmessungen an einem Normal ermittelten Längenmaße,
- im Rahmen einer Maschinenfähigkeitsuntersuchung hergestellte 50 Teile.

Der Begriff Stichprobe kommt übrigens von dem in früheren Zeiten üblichen „Anstechen“ der Getreidesäcke und Baumwollballen zum Zwecke der Qualitätsprüfung. Eine Stichprobe besteht aus einer Einheit oder mehreren Einheiten, die einer realen oder fiktiven Grundgesamtheit nach dem Zufallsprinzip „entnommen“ werden. Die Anzahl dieser Elemente wird Stichprobenumfang genannt. Die Eigenschaften der Stichprobe sollen repräsentativ für die Grundgesamtheit sein. Eine zufällige Entnahme setzt voraus, dass jedes Element der Grundgesamtheit die gleiche Chance (gleiche Wahrscheinlichkeit) hat, in die Stichprobe zu gelangen. In der Regel ist es nur in wenigen Fällen möglich, das Zufallsprinzip in nahezu idealer Weise zu verwirklichen (Münzwurf, Roulettespiel, Ziehung der Lottozahlen). Insbesondere versagt die Anschauung meist in Zusammenhang mit fiktiven Grundgesamtheiten; das „Entnehmen“ ist hier nur im übertragenen Sinn zu verstehen.

2.4 Zufallsvariable

In der Statistik umgeht man diese Schwierigkeit durch Einführung der Begriffe „Zufallsexperiment“ und „Zufallsvariable“. Mit „Zufallsexperiment“ bezeichnet man einen Vorgang, der sich prinzipiell beliebig oft wiederholen lässt, und dessen (Einzel-) Ergebnis nicht vorhersagbar ist (z.B. Wurf mit einem Würfel). Die „Zufallsvariable“ repräsentiert die möglichen Ergebnisse eines Zufallsexperiments (z.B. die Zahlen 1, 2, ..., 6). Mathematisch gesehen ist sie eine Funktion, der durch ein „Zufallsexperiment“ eine reelle Zahl (z.B. gewürfelte Augenzahl) zugeordnet wird. Die „Einheiten“ oder „Elemente“, die als Ergebnisse eines Zufallsexperiments „beobachtet“ werden (der Grundgesamtheit als Stichprobe „entnommen“ werden), sind sogenannte „Realisationen“ dieser Zufallsvariablen.

Vergleicht man diese Definitionen mit den Erläuterungen aus 2.2 und 2.3, so ist erkennbar, dass die dem üblichen Sprachgebrauch und der Anschauung entsprechenden Begriffe Grundgesamtheit und Stichprobe durch die mathematischen Größen „Zufallsvariable“ und „Realisationen der Zufallsgröße“ ersetzt worden sind, mit der Einschränkung, dass es sich bei den letztgenannten beiden stets um reelle Zahlen handelt.



Beispielsweise sind die Ergebnisse x_1, x_2, \dots, x_{10} einer Folge von 10 Wiederholmessungen an einem Längennormal Realisationen einer Zufallsgröße X , welche die Gesamtheit aller möglichen (unendlich vielen) Messergebnisse an diesem Normal repräsentiert.

Im üblichen Sprachgebrauch werden häufig auch die an einer Stichprobe realer Teile gemessenen Merkmalswerte (Menge von Messergebnissen) als Stichprobe bezeichnet.

BEISPIEL 2.1:

Die Bilder 2.1 und 2.2 zeigen jeweils eine Grundgesamtheit von 4941 Kugeln. Die Grundgesamtheit nach Bild 2.1 besteht aus 4465 weißen und 476 schwarzen Kugeln. Letztere wollen wir als Stellvertreter für fehlerbehaftete Teile betrachten. Der Anteil p' der schwarzen Kugeln ist also $p' = \frac{476}{4941} \approx 0,096 = 9,6\%$.

Zur Schätzung des Fehleranteils (er ist in diesem Fall die interessierende Zufallsvariable) wird eine Stichprobe von 490 Kugeln entnommen, hier veranschaulicht durch den rechteckig begrenzten Bereich (49 · 10 Kugeln). Die Stichprobe enthält 48 schwarze Kugeln, deren Anteil in der Stichprobe ist also $p = \frac{48}{490} \approx 0,098 = 9,8\%$.

Somit liefert in diesem Fall die Stichprobe eine relativ gute Schätzung des Fehleranteils der Grundgesamtheit.

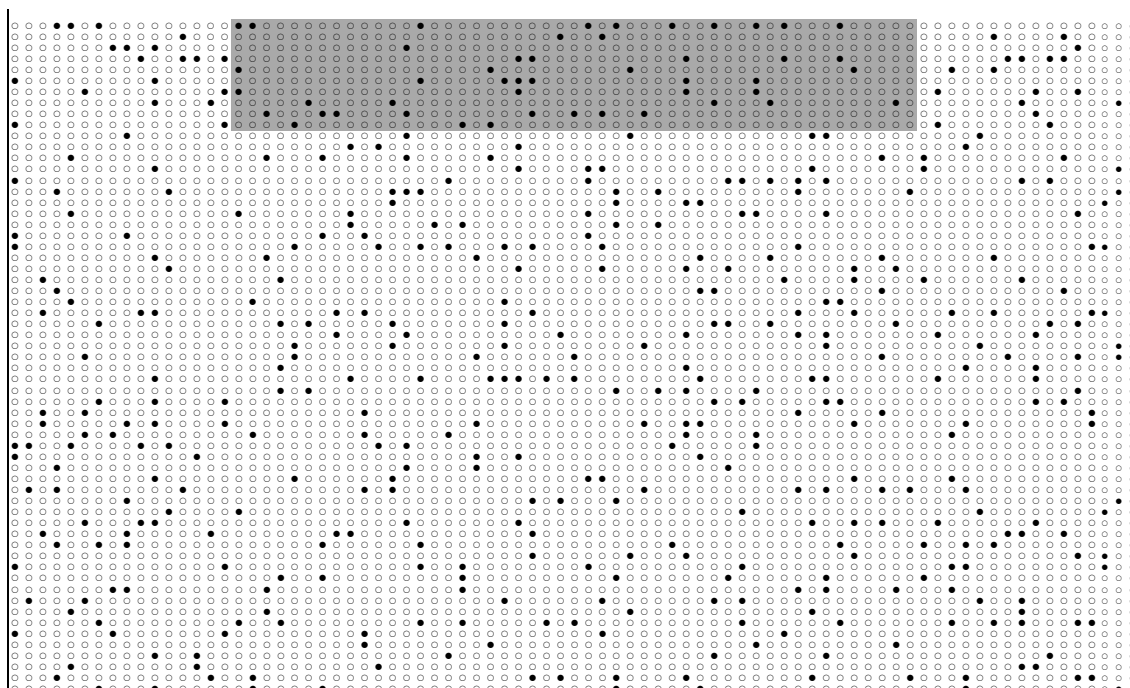


Bild 2.1: Gleichmäßig durchmischte Grundgesamtheit mit etwa 10 % Fehleranteil. Der Fehleranteil wird anhand der Stichprobe (Rechteck) recht gut geschätzt.

Dass eine solche Schätzung auch zu Fehlurteilen führen kann, zeigt das Beispiel nach Bild 2.2. Hier ist die Grundgesamtheit nicht homogen durchmischte, der Fehleranteil nimmt von unten nach oben ab. Diese Situation kann beispielsweise eintreten, wenn der Fehleranteil einer Fertigungsline innerhalb eines begrenzten Zeitraums stetig abnimmt, und die produzierten Teile der Fertigungsreihenfolge entsprechend in einen Container gelegt werden. Im vorliegenden Beispiel sind 499 Kugeln der Grundgesamtheit schwarz, und 4442 weiß.



Es liegt also ein Fehleranteil $p' = \frac{499}{4941} \approx 0,101 = 10,1\%$ vor, der sich kaum von dem in Bild 2.1 unterscheidet. Es ist aber offensichtlich, dass der Fehleranteil in der Stichprobe $p = \frac{8}{490} \approx 0,0163 = 1,63\%$ zu einem Fehlschluss führt.

Der Unterschied in den beiden Stichprobenergebnissen und den daraus resultierenden Anteilsschätzungen ist offenbar darauf zurückzuführen, dass das Zufallsprinzip (in beiden Fällen) verletzt wurde. Es hatte eben nicht jedes Teil der Grundgesamtheit die gleiche Chance, in die Stichprobe zu gelangen.

In der Realität sind wir natürlich in der Situation, dass wir den Fehleranteil in der Grundgesamtheit nicht kennen und uns ausschließlich auf die Aussagefähigkeit der Stichprobe verlassen müssen.

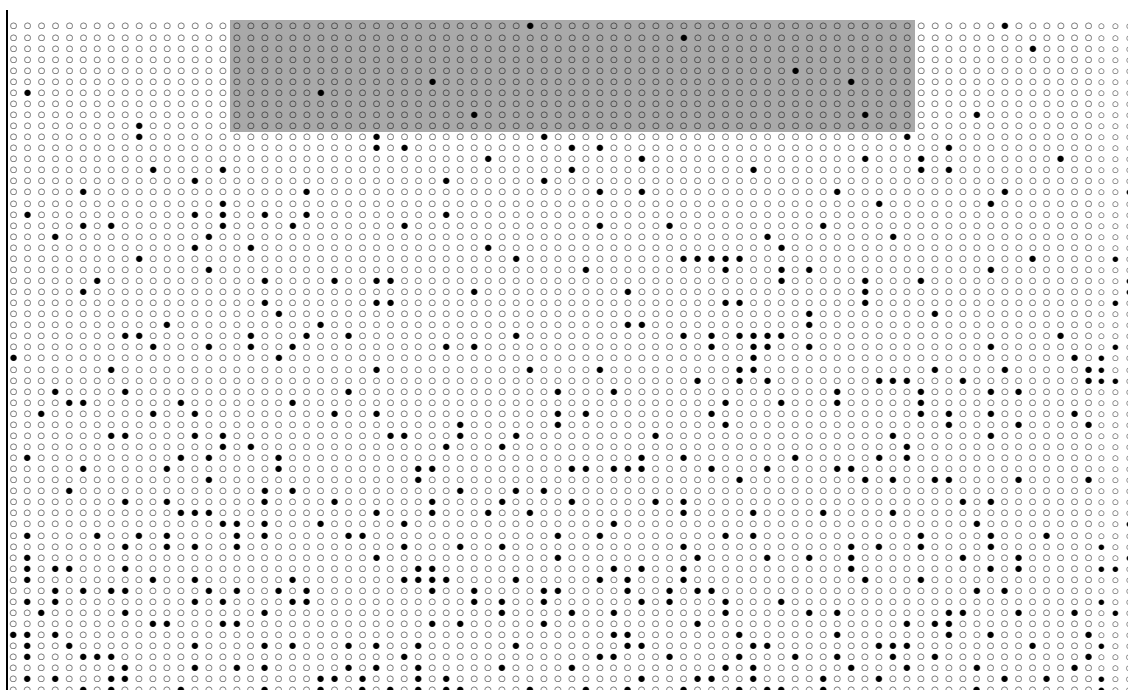


Bild 2.2: Inhomogene Grundgesamtheit mit etwa 10 % Fehleranteil. Der Fehleranteil wird anhand der Stichprobe (Rechteck) falsch geschätzt.

2.5 Wahrscheinlichkeit

Die mathematische Wahrscheinlichkeit ist eine Zahl, die eng mit dem Ergebnis eines Zufallsexperiments verknüpft ist.

Ihre klassische Definition leitet sich aus der Spieltheorie ab, aus der sich die Wahrscheinlichkeitsrechnung und Statistik ursprünglich entwickelt haben.

Ein in Statistiklehrbüchern häufig betrachtetes Zufallsexperiment ist das Werfen einer Münze. Es wird allgemein akzeptiert, dass das Ergebnis eines Münzwurfs nicht vorhersagbar ist, und aufgrund der (hinreichenden) Symmetrie der Münze die Ereignisse „Kopf“ und „Zahl“ gleich wahrscheinlich sind.

Nach der klassischen Definition ist die mathematische Wahrscheinlichkeit $P(A)$ eines Ereignisses A bei einem Zufallsexperiment gegeben durch

$$P(A) = \frac{g}{m} .$$



Darin bezeichnet

g die Anzahl der (günstigen) Fälle, bei denen A eintritt,

und

m die Anzahl aller möglichen Fälle

bei dem betrachteten Experiment. Bezogen auf den Münzwurf bedeutet dies:

die Wahrscheinlichkeit für das Ereignis „Kopf“ ist $P(\text{„Kopf“}) = \frac{1}{2} = 0,5 = 50\%$.

Die Anzahl g der Fälle, in denen „Zahl“ zutrifft (günstiges Ergebnis für den Spieler, der „Zahl“ als Ergebnis vorhergesagt hat) ist gleich 1, die Anzahl aller möglichen Ausgänge des (einmaligen) Münzwurfs ist gleich 2 („Kopf“ und „Zahl“). Offenbar folgt aus der Symmetrie der Münze die „Symmetrie der Wahrscheinlichkeiten“. Beide Ergebnisse sind gleich wahrscheinlich: $P(\text{„Kopf“}) = P(\text{„Zahl“}) = \frac{1}{2}$.

Betrachtet man ein Zufallsexperiment mit einer endlichen Zahl möglicher Ereignisse, deren Auftretenswahrscheinlichkeiten nicht unmittelbar aus Symmetriebetrachtungen abgeleitet werden können, so besteht (zumindest theoretisch) die Möglichkeit, das Experiment sehr oft zu wiederholen und die relativen Häufigkeiten (vgl. Abschnitt 5.3) zu bestimmen, mit der jedes Ereignis auftritt. Man kann dann die Wahrscheinlichkeit für ein bestimmtes Ereignis als Grenzwert definieren, dem sich die relative Häufigkeit (eine Zahl zwischen null und eins) des betrachteten Ereignisses bei einer großen (gegen unendlich strebenden) Zahl von Wiederholungen des Zufallsexperiments nähert.

Beim wiederholten Münzwurf z.B. wird man feststellen, dass die relativen Häufigkeiten (z.B. Anzahl der Würfe mit Ergebnis „Kopf“ dividiert durch die Gesamtzahl der Würfe), mit denen beide Ergebnisse auftreten, sich mit wachsender Zahl von Würfeln immer mehr dem Wert 0,5 nähern (Bild 2.3). Allgemein wird dieses Phänomen durch das „Gesetz der großen Zahlen“ beschrieben.

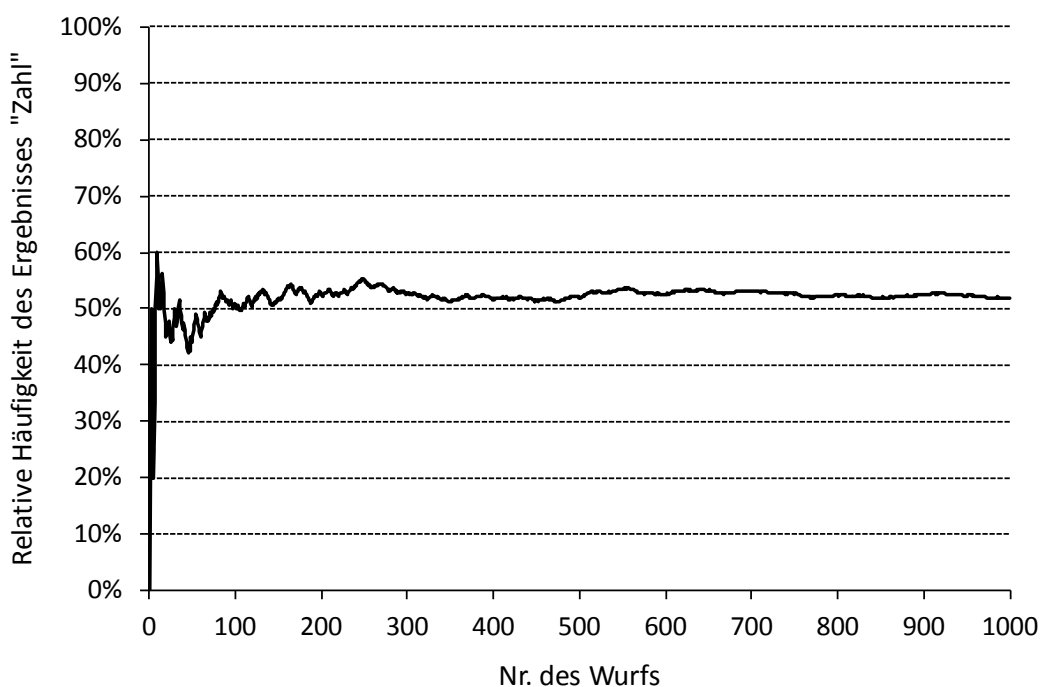


Bild 2.3: Veranschaulichung des Gesetzes großer Zahlen. Die relative Häufigkeit des beim wiederholten Wurf einer Münze erzielten Ergebnisses „Zahl“ nähert sich nach hinreichend vielen Würfeln dem theoretischen Wert 0,5.



Aus nachstehender Tabelle geht hervor, wie die in Bild 2.3 dargestellten relativen Häufigkeiten berechnet worden sind.

Wurf Nr.	Wurf- ergebnis	rel. Häufigkeit des Er- gebnisses „Zahl“	Wurf Nr.	Wurf- ergebnis	rel. Häufigkeit des Er- gebnisses „Zahl“
1	K	$0/1 = 0,00$	991	Z	0,504
2	Z	$1/2 = 0,50$	992	Z	0,504
3	K	$1/3 = 0,33$	993	Z	0,505
4	K	$1/4 = 0,25$	994	Z	0,505
5	K	$1/5 = 0,20$	995	Z	0,506
6	Z	$2/6 = 0,33$	996	K	0,505
7	K	$2/7 = 0,29$	997	K	0,505
8	Z	$3/8 = 0,38$	998	K	0,504
9	Z	$4/9 = 0,44$	999	Z	0,505
10	K	$4/10 = 0,40$	1000	Z	0,505

Tabelle 2.1: Berechnung der relativen Häufigkeiten für Bild 2.3



3 Statistische Kenngrößen

Wesentliche Eigenschaften eines Datensatzes sind die „mittlere Lage“ der Einzelwerte sowie deren „Streuung“ auf der von minus unendlich bis plus unendlich reichenden Zahlengeraden. In diesem Abschnitt werden Kenngrößen erläutert, die zur Beschreibung dieser Eigenschaften geeignet sind.

3.1 Median

Bereits bei der Sammlung von Einzelwerten muss berücksichtigt werden, dass deren Reihenfolge i.a. eine wesentliche Information enthält, nämlich die zeitliche Abfolge ihrer Entstehung (z.B. Temperaturverlauf, Versuchsreihenfolge). Die Werte werden also in der Reihenfolge aufgeschrieben, in der sie gerade anfallen. Die so entstandene Liste heißt Urliste.

Werden einer Serienproduktion nacheinander mehrere Teile entnommen, deren Merkmalswerte anschließend vermessen und analysiert werden sollen, so ist es zweckmäßig, die Teile entsprechend ihrer Produktionsreihenfolge zu nummerieren.

Dies ist insbesondere dann wichtig, wenn etwa der Messvorgang an einem anderen Ort als dem Produktionsort stattfindet und die Gefahr besteht, dass die Reihenfolge sonst verlorenght. Selbstverständlich werden die an den Teilen ermittelten Merkmalswerte (Messwerte) entsprechend der Nummerierung der Teile aufgeschrieben.

BEISPIEL 3.1:

Es wurden folgende neun Messwerte ermittelt: 5, 6, 6, 3, 5, 8, 6, 7, 4.

Man kann annehmen, dass es sich dabei um Abweichungen von einem bestimmten Sollwert (Mittelpunkt des Toleranzbereichs) z.B. in 1/100mm oder mV handelt.

Allgemein bezeichnet man einen Merkmalswert mit dem Buchstaben x und die Anzahl der Werte mit n . Dem Symbol x wird ein Laufindex i angefügt:

$$x_i; i = 1, 2, 3, \dots, n,$$

d.h., die Werte werden bezeichnet mit

$$x_1, x_2, x_3, \dots, x_n.$$

Werden die Werte ihrer Größe nach geordnet und mit dem kleinsten Wert beginnend aufgeschrieben, so spricht man von einer geordneten Liste.

Nach Beispiel 3.1: 3, 4, 5, 5, 6, 6, 6, 7, 8

Mathematisch formuliert:

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$$

Zur Unterscheidung von den Werten der Urliste sind die Indizes hier in Klammern gesetzt.

Bei einer geordneten Liste entspricht also die erste Größe dem minimalen, die letzte dem maximalen Merkmalswert:

$$x_{(1)} = x_{\min} \quad x_{(n)} = x_{\max}$$

Werden die Merkmalswerte über einer Merkmalsachse aufgetragen, so erhält man die Häufigkeitsverteilung der Merkmalswerte (Bild 3.1).



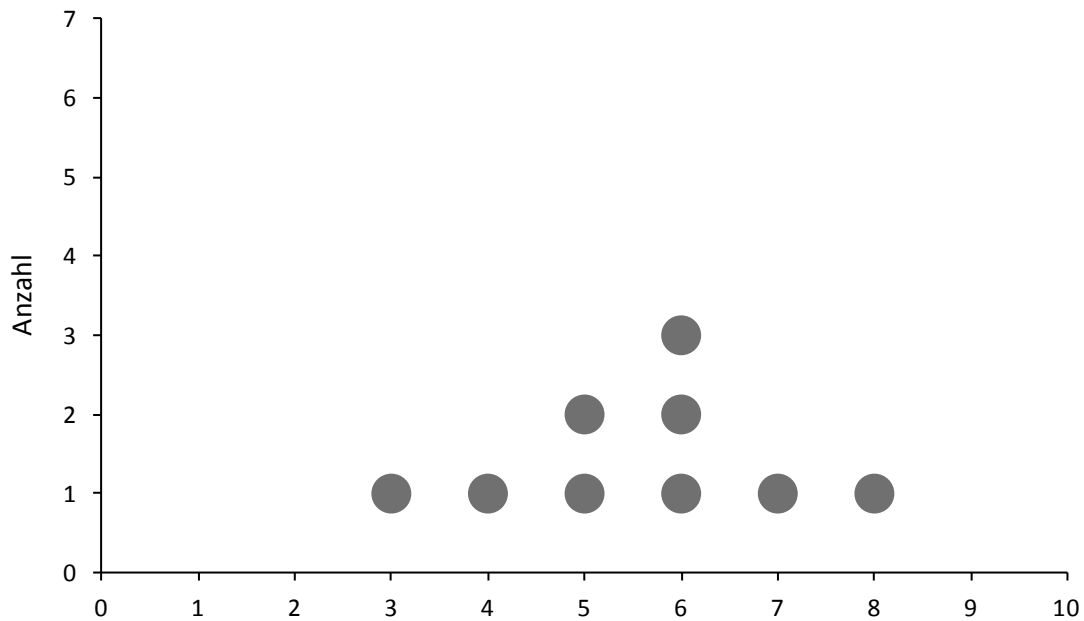


Bild 3.1: Häufigkeitsschaubild (Punktdiagramm)

Eine einfach zu bestimmende Kenngröße der Lage ist der Median, auch Zentralwert genannt. Er teilt die Stichprobe in zwei Hälften gleicher Anzahl. Der Median, bezeichnet mit \tilde{x} (sprich: „x Schlange“), wird durch Abzählen der Werte einer geordneten Liste ermittelt:

$$\tilde{x} = x_{\left(\frac{n+1}{2}\right)} \quad \text{falls } n \text{ ungerade,}$$

$$\tilde{x} = \frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2} \quad \text{falls } n \text{ gerade.}$$

Der Zentralwert kommt in der Wertemenge also nur im Falle einer ungeraden Anzahl von Messwerten vor; für eine gerade Anzahl wird er als Mittelwert der benachbarten Werte $x_{\left(\frac{n}{2}\right)}$ und $x_{\left(\frac{n}{2}+1\right)}$ definiert.

Die bereits oben angegebene geordnete Liste:

$$x_{(1)} = 3, \quad x_{(2)} = 4, \quad x_{(3)} = 5, \quad x_{(4)} = 5, \quad x_{(5)} = 6, \quad x_{(6)} = 6, \quad x_{(7)} = 6, \quad x_{(8)} = 7, \quad x_{(9)} = 8$$

hat also den Median: $\tilde{x} = x_{(5)} = 6$.

Der wesentliche Vorteil des Medians besteht in seiner Unabhängigkeit von den Extremwerten des Datensatzes.



3.2 Arithmetischer Mittelwert

Der arithmetische Mittelwert (das arithmetische Mittel) ist definiert als Quotient aus der Summe aller Einzelwerte und der Anzahl der Einzelwerte:

$$\bar{x} = \frac{\text{Summe aller Einzelwerte}}{\text{Anzahl aller Einzelwerte}} \quad (\bar{x} : \text{sprich „x quer“})$$

oder mathematisch formuliert:

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^n x_i \quad (\text{Arithmetisches Mittel}).$$

Die Summe aller Einzelwerte wird vereinfacht durch ein Summenzeichen (großer griechischer Buchstabe Sigma) dargestellt. Diese Schreibweise bedeutet, dass alle n Merkmalswerte x , angefangen vom ersten gemessenen Wert x_1 (entsprechend $i = 1$) bis zum letzten gemessenen Wert x_n (entsprechend $i = n$), aufsummiert werden.

Ausgeschrieben sieht das folgendermaßen aus:

$$\sum_{i=1}^n x_i = x_1 + x_2 + \dots + x_n.$$

Für die neun Messwerte ($n = 9$) aus Beispiel 3.1 ergibt sich als Summe:

$$\sum_{i=1}^9 x_i = 5 + 6 + 6 + 3 + 5 + 8 + 6 + 7 + 4 = 50$$

und damit als arithmetischer Mittelwert

$$\bar{x} = \frac{50}{9} \approx 5,6.$$

Wenn \bar{x} nicht indiziert ist, bezeichnet es in dieser Unterlage stets den arithmetischen Mittelwert.

Wie folgendes Beispiel zeigt, stellt der Mittelwert lediglich einen Anhaltswert für die „mittlere Lage der Werte“ auf dem Zahlenstrahl dar. Ohne zusätzliche Informationen kann er nahezu wertlos sein.

BEISPIEL 3.2:

Die 20 Schüler einer Klasse haben eine durchschnittliche Körpergröße von 1,70 m.

Kann daraus ein Schluss auf die tatsächliche Verteilung der Körpergrößen gezogen werden?

Denkbar ist es, dass die Körpergrößen 1,50 m, 1,60 m, 1,80 m, und 1,90 m jeweils etwa gleich häufig vertreten sind, und die übrigen Schüler etwa 1,70 m groß sind.

Möglich ist es aber auch, dass eine Hälfte der Schüler etwa 1,60 m, die andere Hälfte etwa 1,80 m groß ist.

Es kann auch nicht ausgeschlossen werden, dass 19 Schüler etwa 1,66 m groß sind und ein extrem großer 2,46 m.

Das Beispiel verdeutlicht, dass die Aussagekraft des arithmetischen Mittelwerts stets in Zusammenhang mit einem zugehörigen Verteilungsmodell (eingipflige, mehrgipflige, symmetrische, schiefe Verteilung) steht. Insbesondere wird deutlich, dass Extremwerte den arithmetischen Mittelwert stark beeinflussen können.



3.3 Gleitender Mittelwert

Ein gleitender Mittelwert wird aus einer Folge von Merkmalswerten gebildet, indem jeweils n Werte dieser Folge formal zu einer Gruppe zusammengefasst werden und der Mittelwert dieser n Werte berechnet wird.

Für jeden neuen Merkmalswert, welcher der Wertefolge hinzugefügt wird, lässt man jeweils den ersten Wert der letzten Wertegruppe weg, so dass sich eine neue Wertegruppe vom Umfang n ergibt, aus der der neue gleitende Mittelwert berechnet wird usw.

Beispiel für $n = 5$:

3 7 4 9 1	$\bar{x}_1 = 4,8$
3 7 4 9 1 8	$\bar{x}_2 = 5,8$
3 7 4 9 1 8 5	$\bar{x}_3 = 5,4$
3 7 4 9 1 8 5 2	$\bar{x}_4 = 5,0$

Selbstverständlich sind die so ermittelten gleitenden Mittelwerte nicht mehr unabhängig voneinander. Diese Kenngröße spricht deshalb auf plötzlich auftretende Veränderungen nur mit Verzögerung an, was durchaus auch beabsichtigt ist.

Beispielsweise lässt sich ein langfristiger Trend in einer Darstellung der Zahl der monatlichen Neuzulassungen von Pkw und Kombi über der Zeit leichter erkennen, wenn aus den Zahlen der vorangegangenen 6 oder 12 Monate ein gleitender Mittelwert gebildet wird (vgl. Bild 3.2). Kurzfristige Schwankungen wirken sich auf den gleitenden Mittelwert kaum aus.

Im Rahmen der Statistischen Prozessregelung besteht die Möglichkeit, eine Qualitätsregelkarte mit gleitendem Mittelwert zur Prozessregelung zu verwenden. In diesem Fall kann das verzögerte „Ansprechen“ des gleitenden Mittelwerts auf plötzliche unerwünschte Prozesszustände dagegen von Nachteil sein.

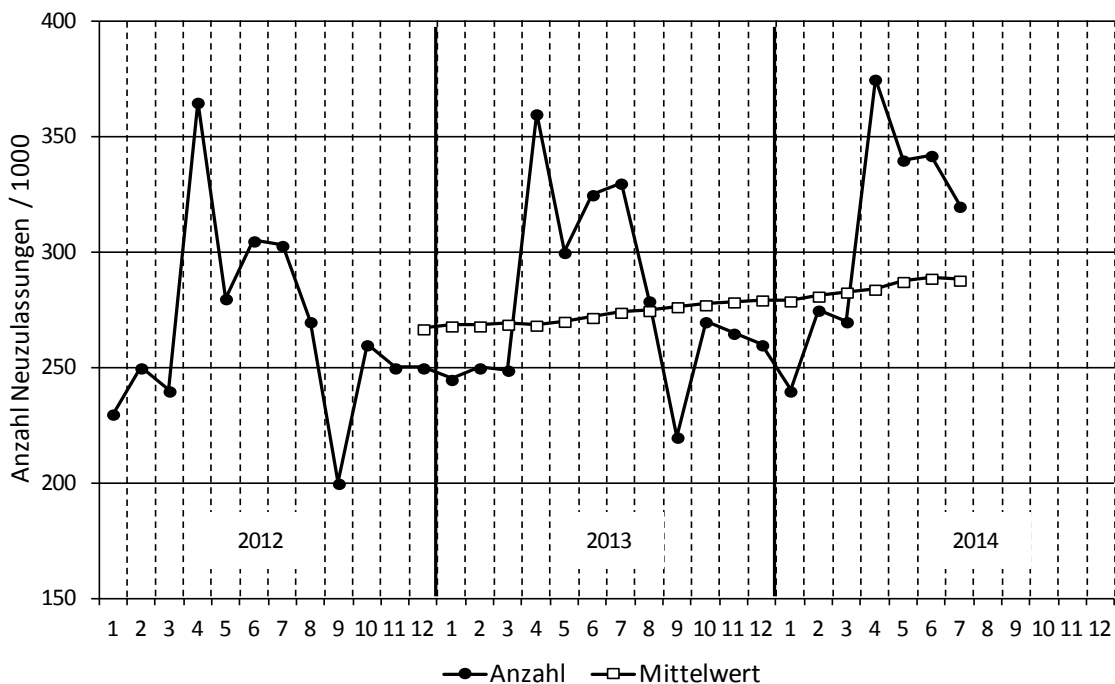


Bild 3.2: Gleitender 12-Monats-Mittelwert



3.4 Geometrischer Mittelwert

Der geometrische Mittelwert entspricht der n-ten Wurzel aus dem Produkt aller n Werte einer Zahlenreihe:

$$\bar{x}_g = \sqrt[n]{\text{Produkt aller } n \text{ Einzelwerte}}$$

oder in Symbolschreibweise:

$$\bar{x}_g = \sqrt[n]{\prod_{i=1}^n x_i} \quad (\text{Geometrisches Mittel}).$$

Das Produkt aller Einzelwerte wird vereinfacht durch den großen griechischen Buchstaben Pi dargestellt. Diese Schreibweise bedeutet, dass alle n Merkmalswerte x, angefangen vom ersten Wert x₁ (entsprechend i = 1) bis zum letzten Wert x_n (entsprechend i = n), miteinander multipliziert werden:

$$\prod_{i=1}^n x_i = x_1 \cdot x_2 \cdot \dots \cdot x_n.$$

Als Beispiel sollen nochmals die neun Messwerte aus Beispiel 3.1 dienen. Ihr Produkt ist:

$$\prod_{i=1}^9 x_i = 5 \cdot 6 \cdot 6 \cdot 3 \cdot 5 \cdot 8 \cdot 6 \cdot 7 \cdot 4 = 3628800,$$

und damit ergibt sich als geometrischer Mittelwert:

$$\bar{x}_g = \sqrt[9]{3628800} \approx 5,4.$$

Der geometrische Mittelwert (auch Geometrisches Mittel genannt) findet Anwendung in Zusammenhang mit Wachstumsprozessen.

BEISPIEL 3.3:

Nehmen wir an, dass die Einwohnerzahl einer Stadt exponentiell wächst. Plötzliche (unstetige) Veränderungen durch massenhafte Zuzüge oder Katastrophen wollen wir ausschließen.

Zeit in Jahren	Jahr	Einwohnerzahl
$t_1 = 0$	1990	$N_1 = 100000$
$t_2 = 10$	2000	$N_2 = 141000$
$t_3 = 20$	2010	$N_3 = 200000$

Wir nehmen an, die Einwohnerzahl sei für die Jahre 1990 und 2010 bekannt, und der Stand im Jahre 2000 solle aus diesen beiden Angaben ermittelt werden.

Das Arithmetische Mittel würde als Ergebnis für das Jahr 2000 eine Einwohnerzahl von 150000 liefern. Bei dieser Art der Berechnung würde man jedoch das exponentielle Wachstum außer Acht lassen und fälschlicherweise den Wert berechnen, der sich bei linearem Wachstum ergeben würde.

Das Geometrische Mittel liefert in diesem Beispiel die korrekte Schätzung:

$$\bar{x}_g = \sqrt[3]{100000 \cdot 200000} \approx 141000.$$



Der Grund für diesen Zusammenhang wird klar, wenn man sich überlegt, was exponentielles Wachstum bedeutet. Die Einwohnerzahl nimmt entsprechend der Funktion

$$N = N_0 \cdot e^{a \cdot t}$$

mit der Zeit t zu. Mit Hilfe dieses Wachstumsgesetzes und der Angaben für die Jahre 1990 und 2010 lässt sich der Wachstumsparameter a berechnen:

$$a = \frac{1}{t} \cdot \ln\left(\frac{N_3}{N_1}\right) = \frac{1}{20} \cdot \ln\left(\frac{200000}{100000}\right) = 0,03466.$$

Die Einwohnerzahl im Jahr 2000 lässt sich dann bestimmen, indem man für t die mittlere Zeit $t_2 = \frac{t_1 + t_3}{2}$ einsetzt:

$$N_2 = N_1 \cdot e^{a \cdot \frac{t_1 + t_3}{2}} = 100000 \cdot e^{0,03466 \cdot 10} \approx 141000.$$

Dies entspricht dem Ergebnis, das man über das geometrische Mittel erhält.

3.5 Harmonischer Mittelwert

Sofern es sich bei den Messwerten x_i um Verhältnisse (bzw. Kehrwerte) handelt, führt die Berechnung des arithmetischen Mittels zu einem falschen Ergebnis.

BEISPIEL 3.4:

Ein Autofahrer legt auf der Autobahn eine Strecke von 200 km zurück. Die erste Hälfte der Strecke $s_1 = 100$ km fährt er mit $v_1 = 80$ km/h, die zweite Hälfte $s_2 = 100$ km mit $v_2 = 160$ km/h. Wie ist die mittlere Geschwindigkeit?

Die naheliegende Antwort $\bar{v} = \frac{80 + 160}{2} \text{ km/h} = 120 \text{ km/h}$ ist falsch!

Das korrekte Ergebnis erhält man, wenn man die Gesamtstrecke durch die benötigte Gesamtzeit dividiert:

$$\bar{v} = \frac{s_1 + s_2}{t_1 + t_2} = \frac{s_1 + s_2}{\frac{s_1}{v_1} + \frac{s_2}{v_2}}.$$

Da die beiden Teilstrecken gleich lang sind ($s_1 = s_2$), ergibt sich schließlich:

$$\bar{v} = \frac{2}{\frac{1}{80 \text{ km/h}} + \frac{1}{160 \text{ km/h}}} \approx 107 \text{ km/h}$$

Die im allgemeinen Fall zu betrachtende Größe

$$\bar{x}_H = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}}$$

heißt Harmonischer Mittelwert (Harmonisches Mittel).



3.6 Standardabweichung

Die in Bild 3.3 dargestellten, aus je 7 Messwerten bestehenden Datensätze sind alle durch den gleichen arithmetischen Mittelwert $\bar{x} = 5$ gekennzeichnet.

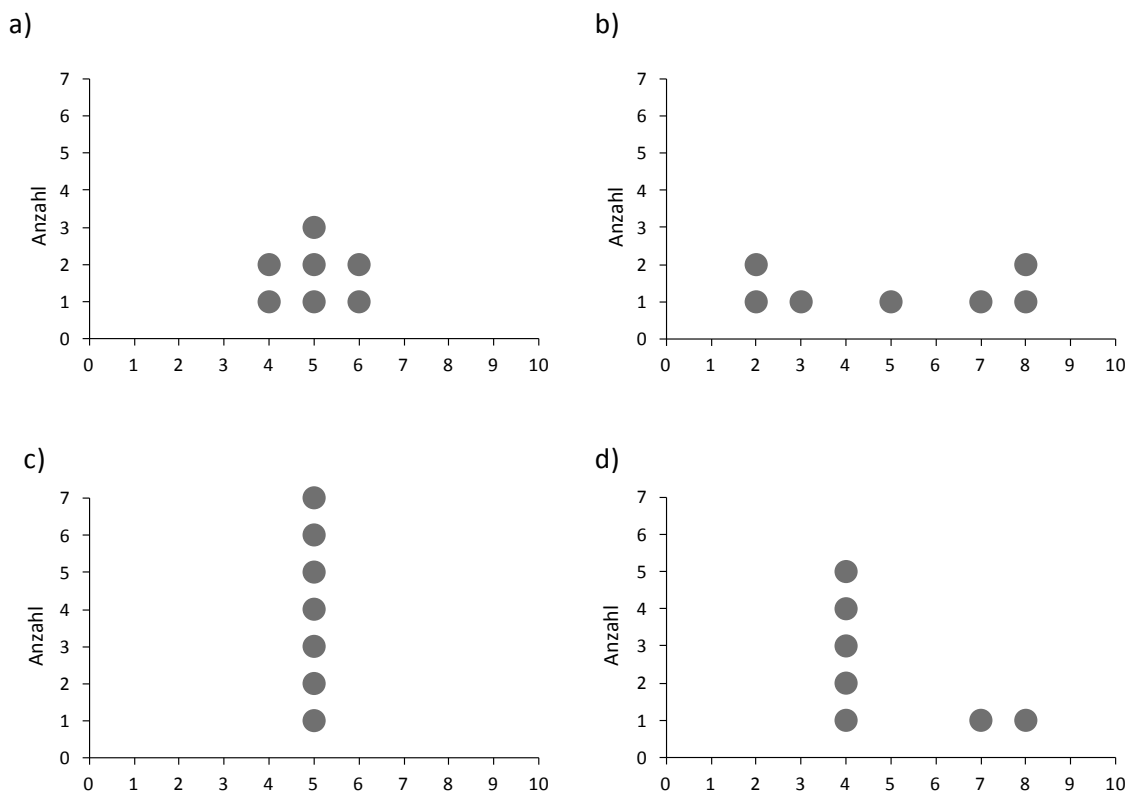


Bild 3.3: Punktdiagramm zu Datensätzen mit gleichem Mittelwert ($\bar{x} = 5$)

Obwohl bei allen Verteilungen der gleiche Mittelwert vorliegt, streuen die Einzelwerte offensichtlich unterschiedlich stark um den Mittelwert. D.h., es besteht eine mehr oder minder große Abweichung der Einzelwerte vom Mittelwert. In Bild 3.3c ist diese Abweichung am geringsten, in 3.3b am größten.

Daher erscheint es zweckmäßig, eine mittlere Abweichung vom Mittelwert in der Weise zu berechnen, dass man die Summe aller Einzelabweichungen $\sum_{i=1}^n (x_i - \bar{x})$ durch die Anzahl

der Einzelwerte n dividiert:
$$\text{Mittlere Abweichung} = \frac{1}{n} \cdot \sum_{i=1}^n (x_i - \bar{x}).$$

Dabei tritt jedoch die Schwierigkeit auf, dass die Summe aller Einzelabweichungen gleich null wird:

$$\begin{aligned} \sum_{i=1}^n (x_i - \bar{x}) &= (x_1 - \bar{x}) + (x_2 - \bar{x}) + (x_3 - \bar{x}) + \dots + (x_n - \bar{x}) \\ &= (x_1 + x_2 + x_3 + \dots + x_n) - n \cdot \bar{x} \\ &= \left(\sum_{i=1}^n x_i \right) - n \cdot \bar{x}. \end{aligned}$$



Wegen der Beziehung $\sum_{i=1}^n x_i = n \cdot \bar{x}$ (Definition des Mittelwerts) folgt schließlich

$$\sum_{i=1}^n (x_i - \bar{x}) = 0.$$

Offenbar ist die Summe der Abweichungen der Einzelwerte vom Mittelwert kein brauchbares Streuungsmaß.

Eine Alternative besteht darin, die Summe der Absolutbeträge der Einzelabweichungen vom Mittelwert zu bilden und durch den Stichprobenumfang zu dividieren. Das so definierte Streuungsmaß heißt mittlere lineare Abweichung:

$$D = \frac{1}{n} \cdot \sum_{i=1}^n |x_i - \bar{x}|.$$

Diese Maßzahl ist jedoch nicht gebräuchlich.

Eine sehr wichtige und häufig benutzte Maßzahl für die Streuung erhält man dagegen, wenn nicht die absoluten Beträge, sondern die quadrierten Einzelabweichungen aufsummiert und der Maßzahl zugrundegelegt werden:

$$\sum_{i=1}^n (x_i - \bar{x})^2.$$

Durch die Quadrierung werden einerseits die Einzelbeiträge zur Gesamtabweichung positiv, und andererseits werden die vom Mittelwert weiter entfernt liegenden Einzelwerte stärker berücksichtigt. Ein geeignetes Streuungsmaß erhält man schließlich, indem die vorstehende Summe der quadratischen Abweichungen noch durch den um eins verminderten Stichprobenumfang dividiert wird.

$$s^2 = \frac{1}{n-1} \cdot \sum_{i=1}^n (x_i - \bar{x})^2$$

Diese Kenngröße s^2 für die Abweichung bezeichnet man als „Varianz“.

Es fällt auf, dass nicht durch die Anzahl der Einzelwerte n , sondern durch die um eins verminderte Anzahl $n-1$ dividiert wird. Der Grund dafür ist, dass die so definierte Varianz der Stichprobe eine „gute“ (mathematisch ausgedrückt: erwartungstreue) Schätzung der unbekanntem Varianz der untersuchten Grundgesamtheit ist.

Die Größe, die sich durch Radizieren (Wurzelziehen) der Varianz s^2 ergibt, heißt (empirische) Standardabweichung s :

$$s = \sqrt{\frac{1}{n-1} \cdot \sum_{i=1}^n (x_i - \bar{x})^2}.$$

Da sie aus einer Summe von quadratischen Ausdrücken berechnet wird, ist sie stets eine positive Zahl (größer oder gleich null).

In Zusammenhang mit Wiederholmessungen zum Zwecke der Einschätzung der Messunsicherheit findet man gelegentlich Angaben von Maßen in der Form $(23 \pm 0,2)$ mm.

Dabei entspricht die Zahl 23 dem aus den einzelnen Messwerten berechneten Mittelwert \bar{x} und die Zahl 0,2 z.B. dem Dreifachen der Standardabweichung dieser Einzelwerte.



Die Angabe $\bar{x} \pm 3 \cdot s$ beinhaltet also neben dem gemessenen Mittelwert auch eine Information über die „Streubreite“ der Messwerte.

Da jeder Messwert (z.B. 23 mm) aus einer Maßzahl (die Zahl 23) und einer Maßeinheit (mm) besteht, ist offensichtlich, dass die Varianz zur Angabe der Messunsicherheit ungeeignet ist (eine Angabe der Form $23 \text{ mm} \pm 0,04 \text{ mm}^2$ wäre sinnlos).

Die manuelle Berechnung der Standardabweichung s soll anhand der Werte aus Beispiel 3.1 erläutert werden. Dazu ist folgende Tabelle zweckmäßig:

Laufindex	Einzelwerte	Abweichung	Abweichungsquadrat
i	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1	5	-0,55	0,303
2	6	0,45	0,203
3	6	0,45	0,203
4	3	-2,55	6,503
5	5	-0,55	0,303
6	8	2,45	6,003
7	6	0,45	0,203
8	7	1,45	2,103
9	4	-1,55	2,403
Summe	50		18,227

Tabelle 3.1: Auswertung der Daten aus Beispiel 3.1

$$\bar{x} = \frac{50}{9} = 5,5 \quad s = \sqrt{\frac{1}{9-1} \cdot 18,227} = 1,51$$

3.7 Variationskoeffizient

Eine zur Beurteilung von Grundgesamtheiten nicht minder wichtige Kenngröße stellt der Variationskoeffizient v dar. Beim Variationskoeffizient wird das Ausmaß der Streuung der Einzelwerte auf die Größe des arithmetischen Mittelwertes bezogen:

$$v = \frac{s}{\bar{x}} \cdot 100 \%$$

Die Verwendung dieser Größe ist stets dann von Vorteil, wenn sich die Mittelwerte zweier zu vergleichender Datensätze, die dem gleichen Verteilungstyp unterliegen, stark voneinander unterscheiden.



3.8 Spannweite

Ein weiteres gebräuchliches Streuungsmaß ist die Spannweite R (engl.: Range). Die Spannweite ist die Differenz des letzten und des ersten Wertes einer geordneten Wertefolge:

$$R = x_{(n)} - x_{(1)}$$

bzw., bezogen auf eine beliebige, ungeordnete Wertemenge, die Differenz des größten und des kleinsten Werts:

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

Die Spannweite ist stets eine positive Zahl (größer oder gleich null).

BEISPIEL 3.5:

Für die Wertemenge (2, 3, 7, 5, 3, 2, -2, 0, 4, 3) ergibt sich:

$$x_{\max} = 7 \quad x_{\min} = -2 \quad R = 7 - (-2) = 9.$$

3.9 Range-Methode zur Bestimmung der Standardabweichung

Die Range-Methode (Spannweiten-Verfahren) ist ein vereinfachtes Rechenverfahren zur schnellen Ermittlung einer Standardabweichung s_R . Dieses Streuungsmaß s_R stimmt in guter Näherung mit s überein und ist für viele in der Praxis auftretende Fälle hinreichend genau. Voraussetzung für die Anwendung dieses einfachen Verfahrens ist, dass dem Datensatz eine Normalverteilung zugrundeliegt und er insbesondere keine Ausreißer enthält.

Die Werte der Messreihe werden in m Gruppen (Stichproben) zu je n Einzelwerten aufgeteilt. Der Datensatz besteht also aus insgesamt $m \cdot n$ Einzelwerten. Dieses Verfahren wird i.a. dann angewendet, wenn die Messwerte ohnehin gruppenweise anfallen, z.B. bei der Median-R-Karte im Rahmen der SPC in Form von 5er-Stichproben.

Der jeweilige Gruppenmittelwert \bar{x}_j ist dann

$$\bar{x}_j = \frac{1}{n} \cdot \sum_{i=1}^n x_{i,j}$$

mit i : Laufindex innerhalb einer Gruppe,

j : Laufindex für die Gruppen ($j = 1, 2, \dots, m$).

Die Spannweite jeder Gruppe ist $R_j = R_{j,\max} - R_{j,\min}$.

Der Mittelwert \bar{R} der Spannweiten aller Gruppen ist

$$\bar{R} = \frac{1}{m} \cdot \sum_{j=1}^m R_j.$$

Aus \bar{R} und einer tabellierten Hilfsgröße d_2^* wird schließlich die Standardabweichung s_R berechnet:

$$s_R = \frac{\bar{R}}{d_2^*}.$$



d_2^* ist von der Anzahl n der Einzelwerte je Gruppe sowie der Anzahl m der Gruppen abhängig (s. Tabelle 3.2).

		Stichprobenumfang n								
		2	3	4	5	6	7	8	9	10
Anzahl m der Gruppen	1	1,414	1,912	2,239	2,481	2,673	2,830	2,963	3,078	3,179
	2	1,279	1,805	2,151	2,405	2,604	2,768	2,906	3,024	3,129
	3	1,231	1,769	2,120	2,379	2,581	2,747	2,886	3,006	3,112
	4	1,206	1,750	2,105	2,366	2,570	2,736	2,877	2,997	3,103
	5	1,191	1,739	2,096	2,358	2,563	2,730	2,871	2,992	3,098
	6	1,181	1,731	2,090	2,353	2,558	2,726	2,867	2,988	3,095
	7	1,173	1,726	2,085	2,349	2,555	2,723	2,864	2,986	3,092
	8	1,168	1,721	2,082	2,346	2,552	2,720	2,862	2,984	3,090
	9	1,164	1,718	2,080	2,344	2,550	2,719	2,860	2,982	3,089
	10	1,160	1,716	2,077	2,342	2,549	2,717	2,859	2,981	3,088
...	
d_2		1,128	1,693	2,059	2,326	2,534	2,704	2,847	2,970	3,078

Tabelle 3.2: Werte von d_2^* in Abhängigkeit von n und m . d_2 ist der Grenzwert von d_2^* für unendlich viele Gruppen (d.h. $m \rightarrow \infty$).

BEISPIEL 3.6:

i	Gruppe Nr.					
	1	2	3	4	5	6
1	70	71	68	72	72	72
2	68	67	72	76	66	69
3	69	66	69	67	63	63
4	69	64	67	68	73	68
5	75	72	69	69	72	68
\tilde{x}_j	$\tilde{x}_1 = 69$	$\tilde{x}_2 = 67$	$\tilde{x}_3 = 69$	$\tilde{x}_4 = 69$	$\tilde{x}_5 = 72$	$\tilde{x}_6 = 68$
R_j	$R_1 = 7$	$R_2 = 8$	$R_3 = 5$	$R_4 = 9$	$R_5 = 10$	$R_6 = 9$

$$\bar{R} = \frac{1}{6} \cdot \sum_{j=1}^6 R_j = \frac{7+8+5+9+10+9}{6} = 8$$

$$s_R = \frac{\bar{R}}{d_2^*} = \frac{8}{2,353} \approx 3,4$$



4 Statistische Berechnungen in EXCEL

Messdaten werden häufig in EXCEL-Tabellen erfasst. Meist besteht dann auch die Notwendigkeit, einfache statistische Kenngrößen unmittelbar in dieser Tabelle zu ermitteln.

Die Berechnung eines Mittelwerts kann man so realisieren, dass man z.B. die in Spalte A untereinander stehenden Einzelwerte mit Hilfe des Befehls "=SUMME(A1:An)" addiert und das Ergebnis dann durch die Anzahl n dividiert. Das entspräche dem in 3.2 erklärten Vorgehen.

Zur Ermittlung der Varianz kann man mühsam die in Kap. 3.6 beschriebenen Schritte nachvollziehen, indem man den berechneten Mittelwert von jedem Wert in Spalte A subtrahiert und das Ergebnis in die entsprechende Zeile der Spalte B einträgt, anschließend jedes Feld in Spalte B mit sich selbst multipliziert und das Ergebnis in die entsprechende Zeile in Spalte C einträgt und schließlich die Spalte C summiert und das Ergebnis durch $n-1$ dividiert. Auf diese Weise können Kenngrößen durch Anwendung der Grundrechenarten Addition, Subtraktion, Multiplikation und Division ermittelt werden. Zur Berechnung der Standardabweichung wird zusätzlich noch das Wurzelziehen benötigt.

Solche „Programmierungsübungen“ sind dem Verständnis der Formeln und dem Umgang mit einfachen mathematischen EXCEL-Funktionen sicher dienlich. Sobald es aber um die schnelle Ermittlung der Kenngrößen geht, wird der geübte Anwender die in EXCEL verfügbaren Statistikfunktionen nutzen.

Beispielsweise kann der Mittelwert der in den Zeilen 1 bis 10 der Spalte A stehenden Werte durch Eingabe des Befehls "=MITTELWERT(A1:A10)" z.B. in die Zelle A11 berechnet werden, ähnlich die Standardabweichung durch Eingabe von "=STABW(A1:10)" z.B. in Feld A12. Alternativ können entsprechende Funktionalitäten der grafischen Bedienungsfläche genutzt werden.

Im Zusammenhang mit der Standardabweichung sei auf folgende Besonderheit hingewiesen. Die EXCEL-Funktion "STABW" geht davon aus, dass es sich bei den eingegebenen Werten um Stichprobenergebnisse handelt, mit deren Hilfe die Standardabweichung σ einer größeren Grundgesamtheit geschätzt werden soll. Die Berechnung von s (als Schätzwert für σ) erfolgt dann entsprechend der Formel:

$$s = \sqrt{\frac{1}{n-1} \cdot \sum_{i=1}^n (x_i - \bar{x})^2} .$$

In diesem Fall wird die Summe der Abweichungsquadrate also durch den um eins verringerten Stichprobenumfang geteilt.

Die EXCEL-Funktion "STABWN" geht davon aus, dass die eingegebenen Zahlen bereits der Grundgesamtheit entsprechen (z.B. Körpergrößen der Schüler einer Klasse), und die Summe der Abweichungsquadrate wird durch den Umfang der Grundgesamtheit (Anzahl der eingegebenen Werte) dividiert. Die Berechnung erfolgt also nach der Formel:

$$s_n = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^n (x_i - \bar{x})^2} .$$

Bei größeren Datenmengen, etwa ab $n = 50$, wird der Unterschied der Faktoren $\frac{1}{n-1}$

und $\frac{1}{n}$ bedeutungslos (der relative „Fehler“ beträgt im Falle $n = 50$ ungefähr 2 %).



Vorsicht mit Kenngrößen!

Das folgende Beispiel soll zeigen, dass eine Angabe statistischer Kenngrößen allein noch keinen eindeutigen Schluss hinsichtlich des Überschreitungsanteils bzgl. eines tolerierten Grenzwerts (oberer Grenzwert OGW) erlaubt.

BEISPIEL 4.1:

Messreihe 1, Messwerte x_i				
22	25	25	26	26
26	22	22	23	25
24	23	24	23	23
23	24	22	26	25
26	25	24	22	24

Mittelwert: $\bar{x} = 24,0$ Standardabweichung: $s_x = 1,4434$

Messreihe 2, Messwerte y_i				
24	23	22	25	23
23	25	21	24	24
24	23	22	26	24
23	26	24	24	25
24	25	28	24	24

Mittelwert: $\bar{y} = 24,0$ Standardabweichung: $s_y = 1,4434$

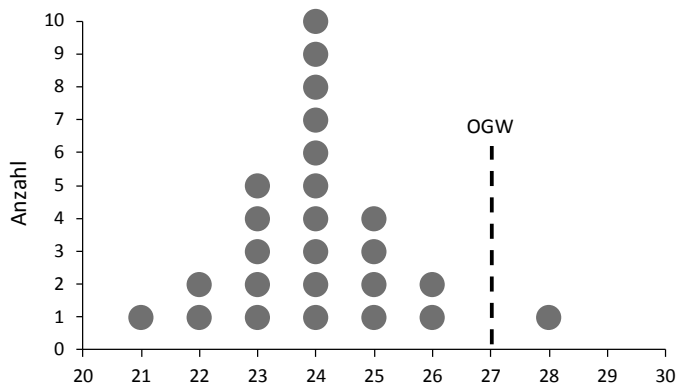
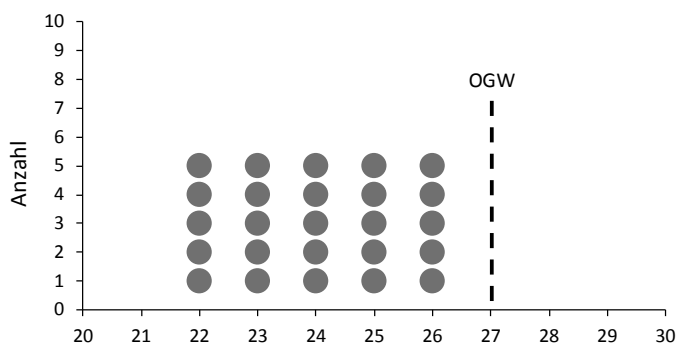


Bild 4.1: Punktdiagramme zu Messreihe 1 (oben) und Messreihe 2 (unten)



Obwohl Mittelwerte und Standardabweichungen beider Messreihen jeweils gleich sind, liegt im ersten Fall kein einziger Wert oberhalb des oberen Grenzwerts OGW („obere Toleranzgrenze“), während im zweiten Fall ein Messwert OGW überschreitet.

Das Beispiel lässt erkennen, dass es bei der Auswertung von Messreihen unumgänglich ist, sich eine möglichst umfassende, ganzheitliche Betrachtungsweise zu eigen zu machen, und Schlüsse nicht allein aufgrund weniger Einzelinformationen zu ziehen. Offensichtlich darf im vorliegenden Beispiel die „Verteilung“ der Messwerte nicht unberücksichtigt bleiben (vgl. Abschnitt 6.).

HINWEIS:

Häufig wird angenommen, dass die Begriffe Mittelwert und Standardabweichung sich stets auf die Normalverteilung (siehe Abschnitt 6.) beziehen. Diese Annahme ist nicht richtig.

2020-04-06 - SOCOS



5 Grafische Darstellung von Daten

Eine grafische Darstellung von Daten ermöglicht dem Betrachter, wesentliche Eigenschaften eines Datensatzes schnell zu erfassen. Sie unterstützt und erleichtert damit die Auswertung von Messreihen.

Es ist zum Beispiel recht leicht, anhand eines Urwertdiagramms besondere Merkmale wie Anfangspunkt, Endpunkt, Trend, Periodizität, Anhäufungen von Punkten oder einzelne Punkte, die von der Mehrheit der übrigen Punkte deutlich entfernt liegen, sogenannte Ausreißer, zu erkennen.

Darüber hinaus dient aber die Mehrzahl grafischer Darstellungen dazu, Daten im Hinblick auf ihre statistischen Eigenschaften bewerten und statistische Kenngrößen auch ohne Rechnerhilfe bestimmen zu können.

5.1 Urwertdiagramm

Das Urwertdiagramm ist eine Auftragung von Messwerten in der Reihenfolge ihrer Entstehung, die Abszisse (x-Achse) entspricht daher häufig der Zeit. Im Rahmen von Prozessuntersuchungen werden Daten beispielsweise in Abständen von Minuten, Stunden, Schichten oder Tagen erfasst. Die Zeit kann also in diesen Fällen durch Angabe des Datums und der Uhrzeit angegeben werden. Wird eine Untersuchung an Teilen durchgeführt, die einem Fertigungsprozess als Stichprobe entnommen worden sind, so können Entnahmezeitpunkt und der Zeitpunkt der Vermessung der Teilemerkmale stark differieren. Dies spielt unter Umständen eine Rolle bei Produkten, die sich zeitlich verändern können (z.B. Kunststoffteile, Klebeverbindungen).

Bei experimentellen Untersuchungen kann dagegen ein Vorgang von Interesse sein, dessen Beginn als Nullpunkt der Zeitskala definiert wird (z.B. Einschwingverhalten einer Regleinrichtung). Das folgende Beispiel zeigt den Temperaturverlauf in einem Trockenofen mit einfacher Zweipunktregelung.

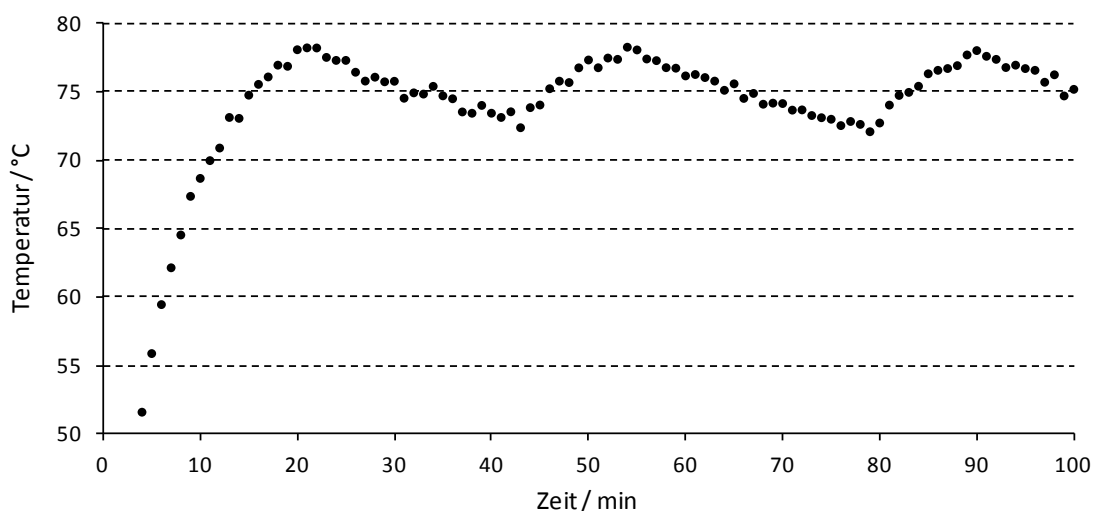


Bild 5.1: Beispiel eines Urwertdiagramms



5.2 Strichliste, Punktdiagramm

Beispiele für Punktdiagramme sind in den vorangegangenen Abschnitten 3.1, 3.6 und 4. zu finden. Diese Punktdiagramme sowie Häufigkeitsdiagramme oder Histogramme werden im Folgenden auch zur Erläuterung einer Klasseneinteilung herangezogen.

Ein Punktdiagramm geht aus einer Strichliste hervor, wenn wie im nachfolgenden Beispiel jede Fünfergruppe (natürlich sind auch andere Gruppierungen möglich) durch einen Punkt repräsentiert wird. Man vergleiche Bild 5.2 mit Bild 4.1. (unten). In beiden Fällen liegt sozusagen in natürlicher Weise eine Klasseneinteilung (s. Abschnitt 5.3) vor, da nur ganzzahlige Merkmalswerte auftreten. Die Höhe einer „Säule“ ist ein Maß für die absolute Häufigkeit des zugehörigen Merkmalswerts.

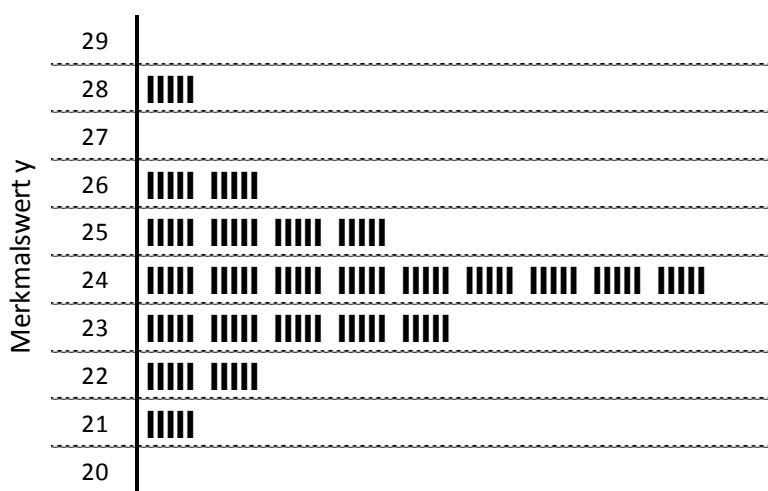


Bild 5.2: Beispiel einer Strichliste

5.3 Klasseneinteilung, Histogramm

Ist die Anzahl der einer Stichprobe zugrundeliegenden Messwerte größer als etwa 25, so wird zweckmäßigerweise eine Klasseneinteilung vorgenommen.

Die Vorgehensweise bei einer Klasseneinteilung soll an einem Beispiel mit nachfolgend aufgeführten Messwerten erläutert werden. Die Messwerte, in Millimetern (mm) angegeben, sind angenommen; sie könnten aber z.B. aus einem Fertigungsprozess stammen, wie etwa dem Absägen von Stangenmaterial.

BEISPIEL 5.1:

Gegeben sei folgende Urliste mit 50 Werten:

8,0	7,0	7,4	8,0	7,0
7,4	7,8	7,5	7,7	6,9
6,5	7,5	7,6	7,3	8,0
7,0	7,5	7,1	7,4	8,6
6,0	8,0	7,0	8,0	6,9
7,5	8,4	6,8	8,3	8,0
8,3	7,3	7,0	7,5	7,9
8,0	7,5	7,0	6,5	7,8
5,8	7,8	6,3	7,5	7,9
9,0	8,0	7,1	7,0	7,4



Trägt man unter Berücksichtigung einer Klasseneinteilung die Messwerte in ein Häufigkeitsschaubild ein, so ergibt sich unter Zugrundelegung von 7 Klassen folgende Darstellung:

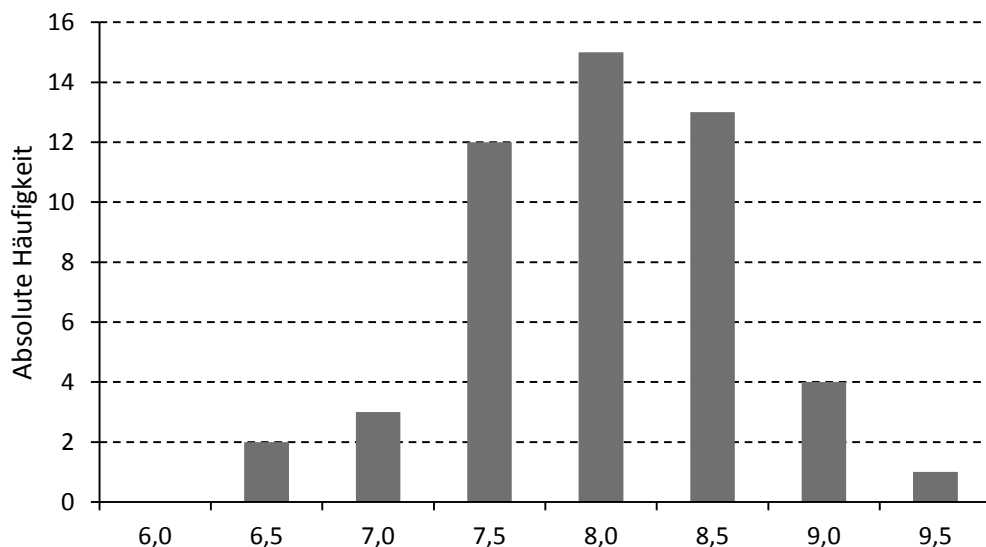


Bild 5.3: Häufigkeitsschaubild zu Beispiel 5.1; $k = 7$, $w = 0,5$

Eine Klasseneinteilung mit 22 Klassen ergibt dagegen folgendes Häufigkeitsschaubild:

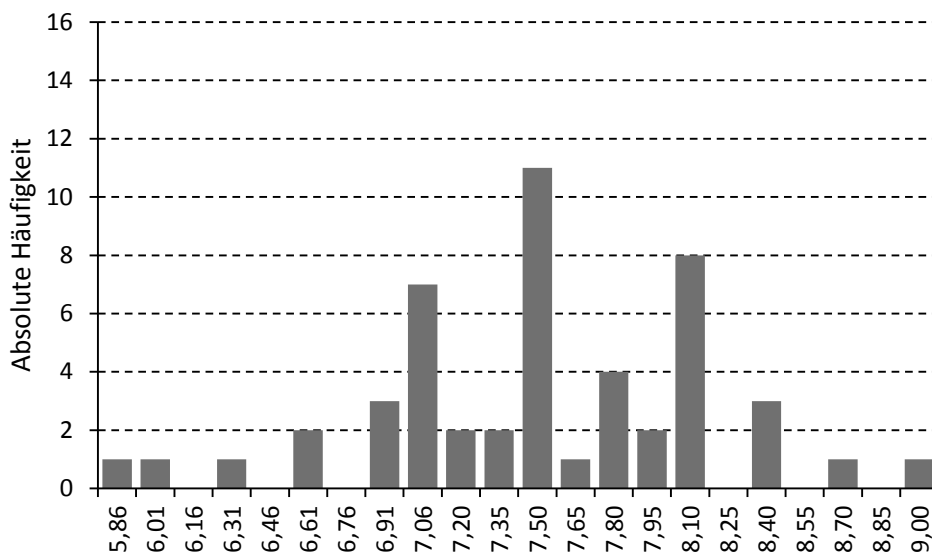


Bild 5.4: Häufigkeitsschaubild zu Beispiel 5.1; $k = 22$, $w = 0,15$

Offensichtlich haben die gewählte Klassenzahl und die Klassenbreite einen großen Einfluss auf das „Aussehen“ des Häufigkeitsschaubildes.

Einen Anhaltswert für eine geeignete Wahl der Anzahl k der Klassen gibt die Formel (Faustregel):

$$k \approx \sqrt{n} \quad \text{für } 25 \leq n \leq 100.$$

Die Einschränkung bzgl. des Stichprobenumfangs lässt erkennen, dass diese Formel nur für Messreihen bis zu 100 Werten gilt. Bei weniger als 25 Werten ist es meist nicht mehr sinnvoll, ein Häufigkeitsschaubild zu erstellen.



Liegen dagegen mehr als 100 Messwerte vor, so empfiehlt sich die Wahl der Klassenzahl entsprechend der Formel:

$$k \approx 5 \cdot \log(n) \quad \text{für } n > 100.$$

Unter Berücksichtigung dieser Regeln ergeben sich also etwa folgende Klassenzahlen:

Anzahl der Messwerte n	Anzahl der Klassen k
ab 25	5
30	6
45	7
60	8
75	9
100	10
200	12

Tabelle 5.1: Anzahl der Klassen k in Abhängigkeit von der Anzahl der Messwerte n

Wie man den Häufigkeitsschaubildern entnehmen kann, umfasst jede Klasse einen bestimmten Wertebereich. Die Grenzen eines jeden Bereichs (Intervalls) heißen untere und obere Klassengrenze, die Länge eines solchen Intervalls bezeichnet man als Klassenbreite w.

Ein möglicher Ansatz zur Wahl dieser Klassenbreite besteht darin, sie anhand der Formel $w = \frac{x_{max} - x_{min}}{k - 1}$ aus der Spannweite zu berechnen. Dies hat aber meist Klassengrenzen mit mehreren Nachkommastellen zur Folge und kann zu leeren Klassen führen.

Das Häufigkeitsschaubild nach Bild 5.3 wurde anhand der n = 50 Werte aus Beispiel 5.1 erstellt. Entsprechend obiger Regeln für die Wahl der Größen k und w ergaben sich folgende Werte:

$$k \approx \sqrt{50} \approx 7 \quad \text{und} \quad w = \frac{x_{max} - x_{min}}{k - 1} = \frac{9,0 - 5,8}{7 - 1} \approx 0,5.$$

In den Bildern 5.3 und 5.4 wurden die Klassengrenzen durch entsprechende Wahl der zweiten Nachkommastelle so festgelegt, dass jeder Merkmalswert eindeutig einer Klasse zugeordnet werden kann. Eine andere Möglichkeit einer eindeutigen Zuordnung besteht darin, jeweils die rechte Klassengrenze mit zum Intervall zu zählen oder Werte, die mit einer Klassengrenze identisch sind, jeweils zur Hälfte auf die beiden benachbarten Klassen aufzuteilen.

HINWEIS:

Es sind Situationen denkbar, in denen es vorteilhaft sein kann, die Klassenbreiten unterschiedlich zu wählen. Beispielsweise versagt obige Regel zur Berechnung der Klassenbreite meist, wenn der Datensatz einen Ausreißer enthält.

Die strikte Befolgung dieser Regel könnte dann dazu führen, dass nur die beiden äußeren Klassen besetzt sind (eine dieser Klassen enthält lediglich den Ausreißer) und alle übrigen leer bleiben. Man kann dies vermeiden, indem man beispielsweise bei der Klasseneinteilung einzelne Extremwerte zunächst unberücksichtigt lässt und sie nach Festlegung einer für die Situation sinnvollen Klasseneinteilung jeweils der entsprechenden äußeren Klasse (erste bzw. letzte Klas-



se) zuordnet. Dies bedeutet, dass die äußere rechte Klasse nach oben, die äußere linke Klasse nach unten offen ist.

Es ist nicht auszuschließen, dass auch Statistikprogramme, die (beispielsweise zur Histogramm-erstellung) Klasseneinteilungen nach wenigen einfachen Regeln vornehmen, je nach „Exotik“ des Datensatzes aus genannten Gründen unbrauchbare Darstellungen liefern. Daher bieten sie dem Benutzer meist die Möglichkeit, die Klasseneinteilung nach eigenem Ermessen zu korrigieren.

Bevor auf die Berechnung des Mittelwertes \bar{x} und der Standardabweichung s aus einer Klasseneinteilung eingegangen wird, sollen noch einige wichtige Begriffe erläutert werden.

Klassengrenzen:

Jede Klasse einer Klasseneinteilung wird begrenzt durch eine untere Klassengrenze x'_{j-1} und eine obere Klassengrenze x'_j .

Klassenmitte x_j :

Die Klassenmitte entspricht dem arithmetischen Mittel von unterer und oberer Klassengrenze:

$$x_j = \frac{x'_{j-1} + x'_j}{2}.$$

Klassenbreite w_j :

Die Klassenbreite entspricht dem Abstand zwischen unterer und oberer Klassengrenze:

$$w_j = x'_j - x'_{j-1}.$$

Im Allgemeinen haben alle Klassen gleiche Klassenbreite, d.h. $w_j = w$ für alle Klassen.

Absolute Häufigkeit n_j :

Anzahl der Werte, die auf die betreffende (j -te) Klasse entfallen (man spricht auch von „absoluter Klassenhäufigkeit“).

Relative Häufigkeit h_j :

Absolute Häufigkeit dividiert durch die Gesamtzahl n der Werte des Datensatzes:

$$h_j = \frac{n_j}{n} \quad \text{mit } n = \sum_{j=1}^k n_j = n_1 + n_2 + n_3 + \dots + n_k.$$

Absolute Summenhäufigkeit G_j :

Summe der absoluten Häufigkeiten n_j von der ersten bis zur j -ten Klasse (einschließlich).

$$G_j = \sum_{i=1}^j n_i = n_1 + n_2 + n_3 + \dots + n_j$$

Relative Summenhäufigkeit H_j :

Relativer Anteil aller Werte unterhalb der oberen Klassengrenze der j -ten Klasse:

$$H_j = \sum_{i=1}^j h_i = h_1 + h_2 + h_3 + \dots + h_j$$

oder, was bei Rechnung von Hand einfacher ist: $H_j = \frac{G_j}{n}$.



Bei einem Häufigkeitsschaubild wird die Verteilung der Messwerte, also der Zusammenhang zwischen einer Variablen x und der Häufigkeit ihres Auftretens, dargestellt. Werden über der Merkmalsachse die absoluten Häufigkeiten aufgetragen, so erhält man ein Häufigkeitsdiagramm (vgl. Bilder 5.3 und 5.4).

Trägt man dagegen die relativen Häufigkeiten auf, so erhält man ein sogenanntes Histogramm (auch Balkendiagramm genannt).

Aus den Werten von Beispiel 5.1 ergibt sich z.B. folgende Darstellung:

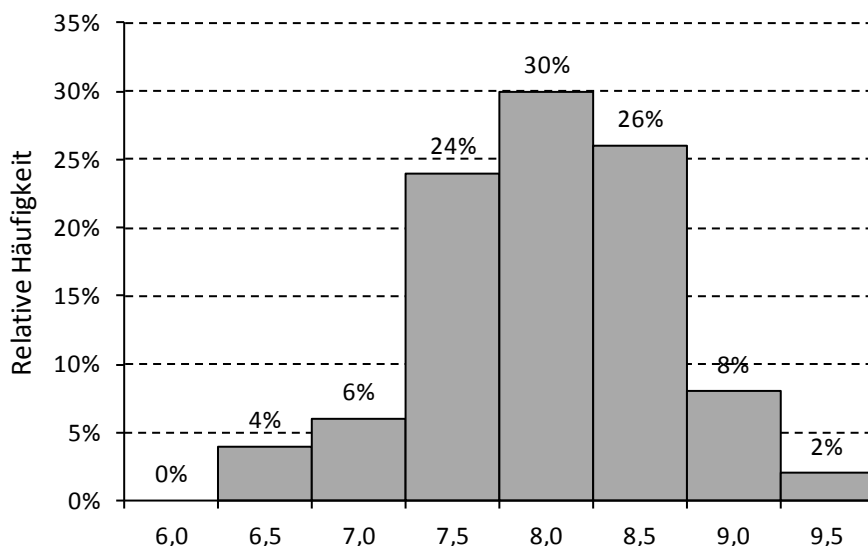


Bild 5.5: Histogramm zu den Werten aus Beispiel 5.1

Dabei sind über den Merkmalsklassen Rechtecke gezeichnet, deren Höhen den Häufigkeiten h_j entsprechen (bei konstanter Klassenbreite). Die folgende Tabelle fasst alle wesentlichen Kenngrößen zusammen:

Klasse Nr.	untere Klassen-grenze x'_{j-1}	obere Klassen-grenze x'_j	Absolute Häufigkeit n_j	Relative Häufigkeit h_j	Relative Summen-häufigkeit H_j	Hilfsgröße (s. Text) $n_j \cdot x_j$	Hilfsgröße (s. Text) $n_j \cdot x_j^2$
1	5,75	6,25	2	4 %	4 %	12,0	72,0
2	6,25	6,75	3	6 %	10 %	19,5	126,8
3	6,75	7,25	12	24 %	34 %	84,0	588,0
4	7,25	7,75	15	30 %	64 %	112,5	843,8
5	7,75	8,25	13	26 %	90 %	104,0	832,0
6	8,25	8,75	4	8 %	98 %	34,0	289,0
7	8,75	9,25	1	2 %	100 %	9,0	81,0
Σ			$n = 50$	100 %		375,0	2832,6

Tabelle 5.2



5.4 Summenkurve

Werden die relativen Summenhäufigkeiten über den oberen Klassengrenzen aufgetragen, so erhält man einen s-förmigen Kurvenzug, die sogenannte Summenkurve.

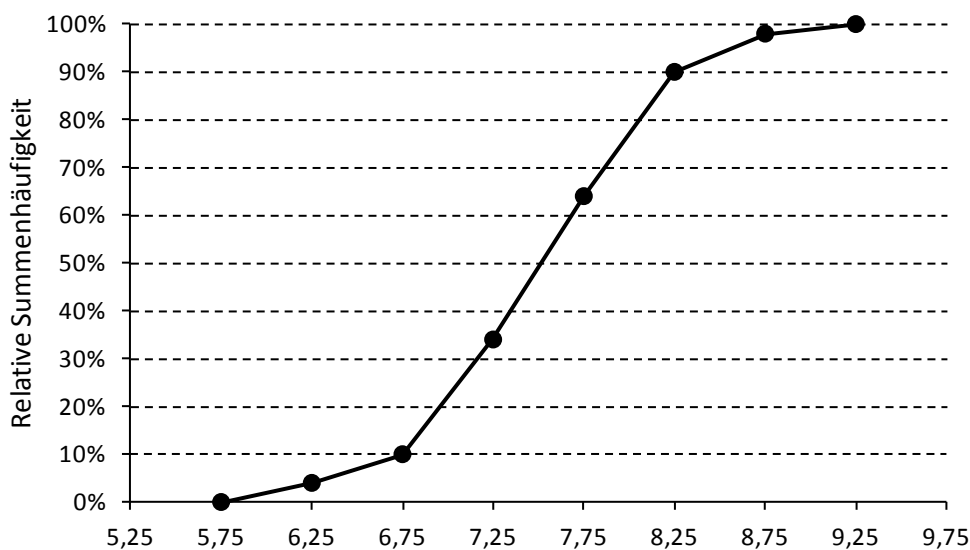


Bild 5.6: Summenkurve zu den Werten aus Beispiel 5.1

Der Vorteil der Summenkurve gegenüber dem Häufigkeitsschaubild ist leicht erkennbar. Es kann ohne große Mühe abgelesen werden, wie viel Prozent der Messwerte unterhalb eines bestimmten Wertes auf der x-Achse liegen (z.B. zur Abschätzung eines Ausschussanteils). Im gezeigten Beispiel liegen 90% der Daten unterhalb des Wertes 8,25, bzw. 10% der Daten oberhalb dieses Wertes. Sofern die Urwerte des Datensatzes nicht bekannt sind, so können folgende Formeln nützlich sein, die es erlauben, Mittelwert und Standardabweichung anhand der Informationen zu berechnen, die im Histogramm enthalten sind (zur Erinnerung: die x_j bezeichnen hier die Klassenmitten).

$$\text{Mittelwert: } \bar{x} = \frac{1}{n} \cdot \sum_{j=1}^k (n_j \cdot x_j) = \frac{n_1 \cdot x_1 + n_2 \cdot x_2 + n_3 \cdot x_3 + \dots + n_k \cdot x_k}{n}$$

$$\text{Varianz: } s^2 = \frac{1}{n-1} \cdot \sum_{j=1}^k n_j \cdot (x_j - \bar{x})^2 = \frac{1}{n-1} \cdot \left(\sum_{j=1}^k (n_j \cdot x_j^2) - n \cdot \bar{x}^2 \right)$$

$$\text{Standardabweichung: } s = \sqrt{s^2}$$

Die absoluten Häufigkeiten n_j können anhand der relativen Häufigkeiten h_j berechnet werden: $n_j = n \cdot h_j$.

Im obigen Beispiel (vgl. Tabelle 5.2) findet man:

$$\bar{x} = \frac{375,0}{50} = 7,5 \quad \text{und} \quad s^2 = \frac{1}{50-1} \cdot (2832,6 - 50 \cdot 7,5^2) = 0,41 \quad \text{und schließlich} \quad s = 0,64.$$

Bei Berechnung dieser Größen anhand der Urwerte ergibt sich: $\bar{x} = 7,454$ und $s = 0,6399$.



6 Statistische Verteilungen

6.1 Gaußsche Normalverteilung

Vergrößert man (bei konstanten Messbedingungen) den Umfang n einer Messreihe fortlaufend (d.h. die Anzahl der Messwerte wird theoretisch unendlich groß), und verringert man gleichzeitig die Klassenbreite (gegen null), so nähert sich die Summenkurve (vgl. 5.4) einer Grenzkurve, die der Verteilung (Verteilungsfunktion) der (unendlichen) Grundgesamtheit entspricht. Analog dazu nähert sich die gestufte Linie, die den oberen Rand des Histogramms darstellt, einer Grenzkurve, welche die grafische Darstellung einer Wahrscheinlichkeitsdichtefunktion ist (vgl. Bild 6.1, $n \rightarrow \infty$, Klassenbreite $w \rightarrow 0$).

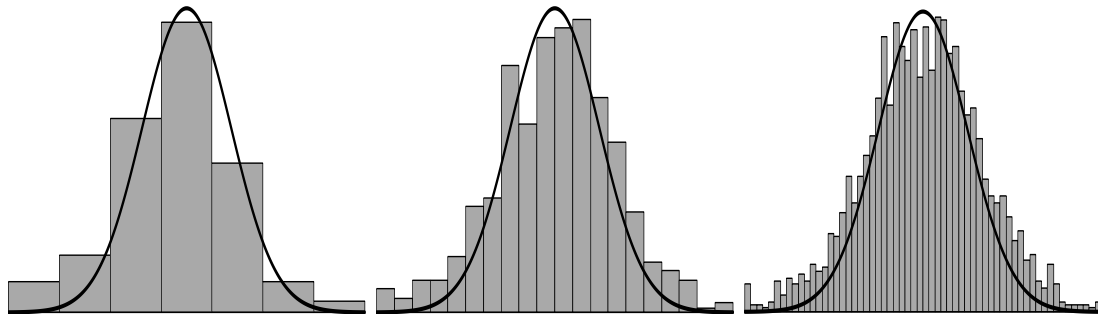


Bild 6.1: Schema zur Veranschaulichung des Übergangs vom Histogramm zur Dichtefunktion am Beispiel Normalverteilung

Jedem Punkt auf der Merkmalsachse entspricht eine Zahl mit theoretisch unendlich vielen Nachkommastellen, also z.B. $x = 73,26452718 \dots$. Die Wahrscheinlichkeit, dass ein Merkmal (mit bekannter Verteilung) exakt diesen Wert annimmt, ist null. Dagegen ist die Wahrscheinlichkeit, dass ein Merkmalswert im Bereich (Intervall) zwischen 73,2 und 73,4 liegt, eine endliche Zahl größer null. Man erhält eine solche Wahrscheinlichkeit, indem man einen Wert der Wahrscheinlichkeitsdichtefunktion mit der Intervallbreite multipliziert. Die Wahrscheinlichkeitsdichtefunktion ist sozusagen die Verallgemeinerung der relativen Häufigkeit bei gegen null schrumpfender Klassenbreite.

Der Begriff „Dichte“ beruht auf einer Analogie zwischen der Wahrscheinlichkeitsrechnung und der Mechanik starrer Körper (vgl. z.B. [3]).

Die Fläche, die von der Kurve der Dichtefunktion und einem bestimmten Intervall der Merkmalsachse begrenzt wird, entspricht der Wahrscheinlichkeit, mit der Merkmalswerte der Grundgesamtheit auf dieses Intervall entfallen. Diese Fläche ist also ein grafisches Analogon zur Wahrscheinlichkeit. Die Gesamtfläche, die von einer beliebigen Wahrscheinlichkeitsdichtefunktion und der Merkmalsachse (zwischen minus unendlich und plus unendlich) begrenzt wird, entspricht stets dem Wert 1 (= 100 %).

In der Vergangenheit hat sich gezeigt, dass man bei experimentellen Untersuchungen und statistischen Beobachtungen häufig Merkmalsverteilungen findet, die Histogramme mit ähnlichem Aussehen ergeben. Der Mathematiker C. F. Gauß hat dieses Phänomen an Landvermessungsdaten untersucht. Dieser Verteilungstyp hat den Namen „Normalverteilung“ und dient häufig als Verteilungsmodell für technisch-statistische Phänomene. Wegen ihrer charakteristischen Form wird die Darstellung der Dichtefunktion dieser Verteilung auch „Gaußsche Glockenkurve“ genannt.



6.1.1 Eigenschaften und Kenngrößen der Normalverteilung

Die Glockenkurve ist die grafische Darstellung der Dichtefunktion der Normalverteilung, welche durch die mathematische Beziehung

$$f(x) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{-\frac{1}{2} \cdot \left(\frac{x-\mu}{\sigma}\right)^2}$$

beschrieben wird.

Die Funktion (und damit die Kurve) ist durch die Parameter μ und σ eindeutig bestimmt. Dabei stellt μ den Mittelwert der Verteilung dar und σ deren Standardabweichung.

Anhand der Funktionsgleichung bzw. ihrer grafischen Darstellung lassen sich weiterhin einige Besonderheiten erkennen:

- Die Kurve ist symmetrisch zum Mittelwert μ .
- Die Kurve besitzt an den Stellen $\mu - \sigma$ und $\mu + \sigma$ jeweils einen Wendepunkt. Dies bedeutet, dass z.B. an der Stelle $\mu - \sigma$ die von der x -Achse weg gerichtete Krümmung der Kurve in eine zur x -Achse hin gerichtete Krümmung übergeht.
- Die Kurve verläuft von $x = -\infty$ bis $x = +\infty$. Das ist jedoch nur für eine theoretische Betrachtung interessant. Praktisch bedeutend ist die Kurve meistens nur im Abstand von jeweils etwa drei bis vier Standardabweichungen vom Mittelwert μ nach links und nach rechts in Bild 6.2. Dort nähert sich die Kurve schon der x -Achse.

Wie bereits angedeutet entspricht die Fläche unter der Gaußkurve einer unendlich großen Anzahl von Messwerten aus einer normalverteilten Grundgesamtheit. Wird diese Fläche gleich 1 (entsprechend 100 %) gesetzt, so kann ein zwischen zwei Punkten liegender Anteil (in %) ermittelt werden.

Werden auf der Merkmalsachse, ausgehend vom Mittelwert μ nach links und rechts entsprechende Strecken in Vielfachen der Standardabweichung abgetragen, so kann in Abhängigkeit von σ der Anteil der Verteilung angegeben werden. Im Bild 6.2 sind diese Anteile für die Bereiche $\mu \pm 1 \cdot \sigma$, $\mu \pm 2 \cdot \sigma$ und $\mu \pm 3 \cdot \sigma$ als graue Flächen hervorgehoben.

Dementsprechend ergibt sich für den Bereich

$\mu \pm 1 \cdot \sigma$ ein Anteil von 68,3 %,

$\mu \pm 2 \cdot \sigma$ ein Anteil von 95,4 %,

$\mu \pm 3 \cdot \sigma$ ein Anteil von 99,7 %.

Man sieht, dass außerhalb von $\mu \pm 3 \cdot \sigma$ nur noch ein verschwindend geringer Anteil, nämlich nur 0,3 % (= 100 % - 99,7 %) der Verteilung liegt (siehe Bild 6.2).



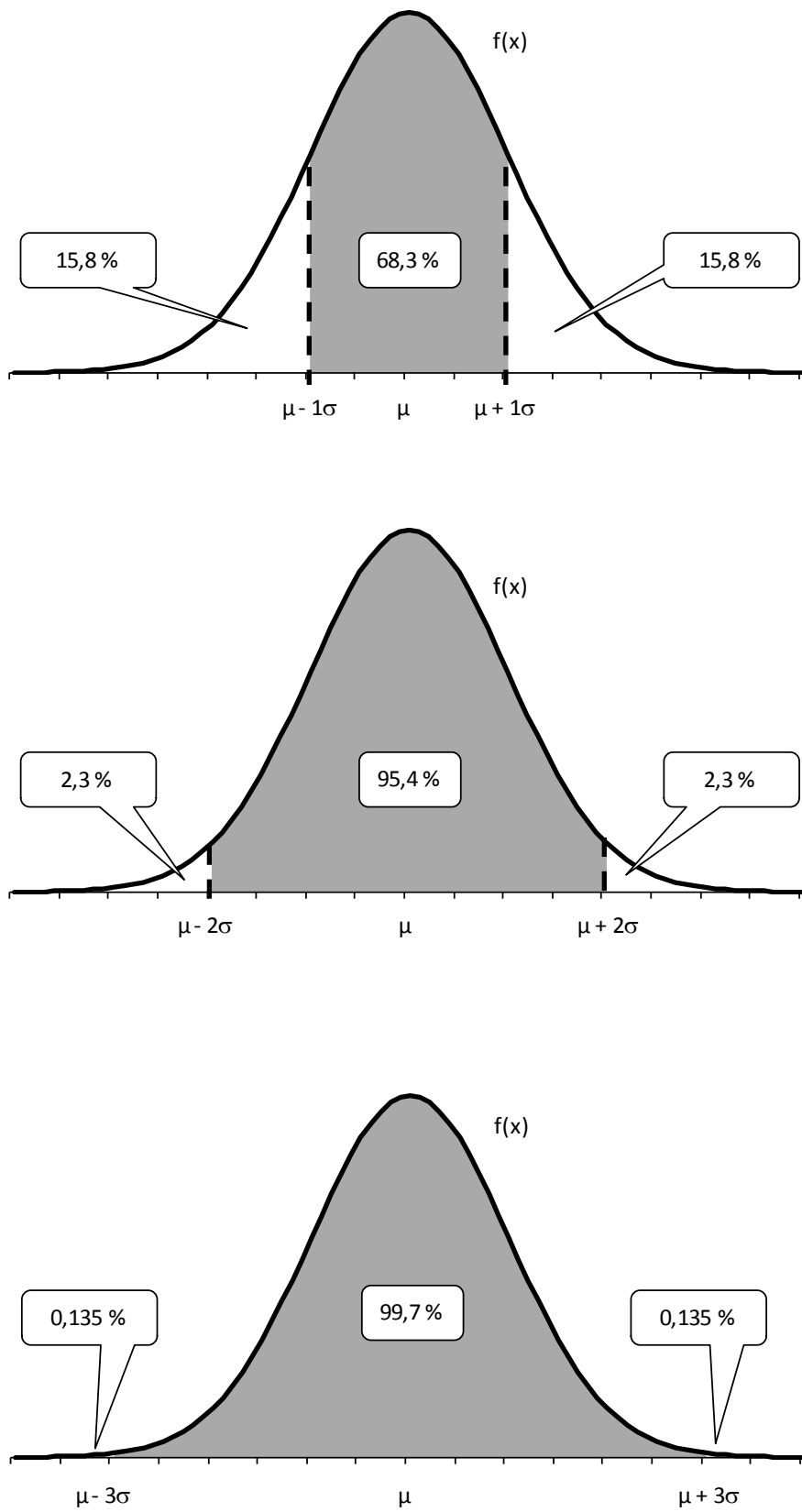


Bild 6.2: Flächeninhalte unter der Glockenkurve



6.1.2 Verteilungsfunktion

In Abschnitt 5.4 ist beschrieben, wie die Summenkurve einer Häufigkeitsverteilung ermittelt wird. Durch Addition der einzelnen relativen Klassenhäufigkeiten werden relative Summenhäufigkeiten bestimmt, die schließlich über den oberen Klassengrenzen aufgetragen werden. Die zugehörigen Punkte werden stückweise durch Geraden verbunden. An der letzten oberen Klassengrenze sind dann alle Werte (= 100 % der Verteilung) berücksichtigt.

Die Summenkurve der Gaußverteilung, wird im Prinzip ebenso ermittelt. Dabei muss die in Abhängigkeit vom Merkmalswert x gebildete Häufigkeitssumme, die ja einem bestimmten Flächenanteil unter der Gaußkurve entspricht, durch ein besonderes mathematisches Verfahren, die Integration, berechnet werden.

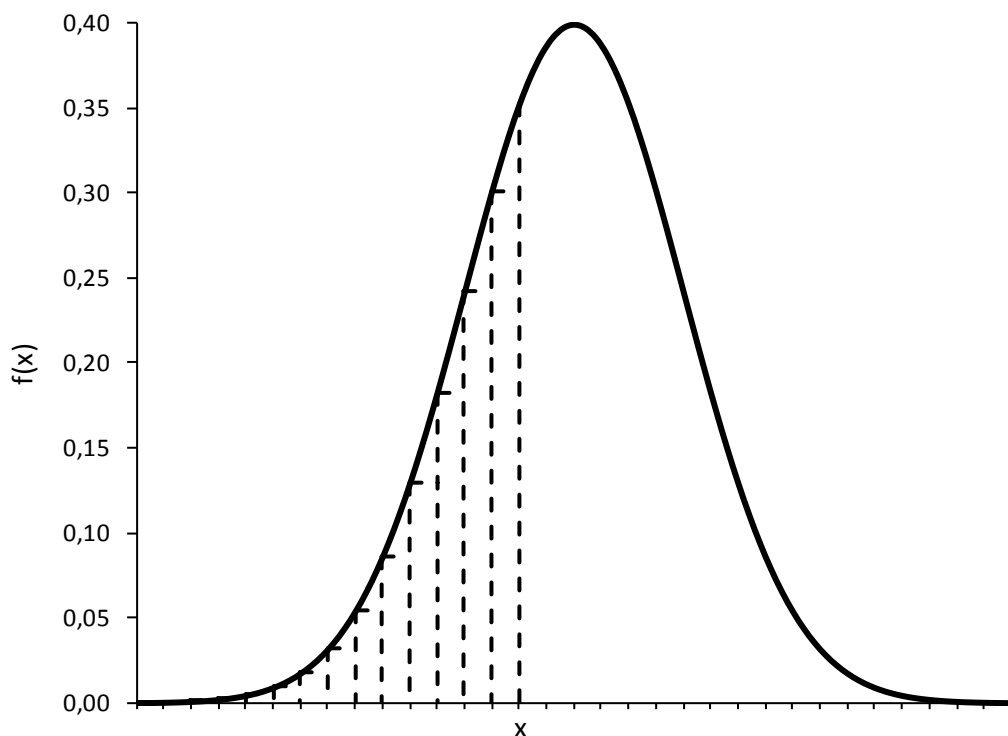


Bild 6.3: Veranschaulichung der Integration

Die Funktion, welche die Summenkurve einer Wahrscheinlichkeitsverteilung beschreibt, nennt man Verteilungsfunktion $F(x)$. Sie gibt für jedes x die Wahrscheinlichkeit an, dass ein zufällig gemessener Wert kleiner oder gleich x ist.

Mathematisch formuliert ergibt sich damit die Summenwahrscheinlichkeit bis zur Stelle x durch die Verteilungsfunktion:

$$F(x) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot \int_{-\infty}^x e^{-\frac{1}{2} \cdot \left(\frac{v-\mu}{\sigma}\right)^2} dv.$$

$F(x)$ entspricht der Fläche unter der Gaußschen Glockenkurve bis zum Wert x .

Bild 6.3 veranschaulicht die Bedeutung der Integration. Die Fläche unter der Kurve bis zur Stelle x wird näherungsweise berechnet, indem man jeweils die Flächen schmaler Rechtecke (Breite Δx) bestimmt und aufsummiert. Das Ergebnis wird umso genauer, je schmaler die Rechtecke werden (Grenzübergang $\Delta x \rightarrow 0$).



In Bild 6.4 (unten) ist die Verteilungsfunktion als Kurve dargestellt.

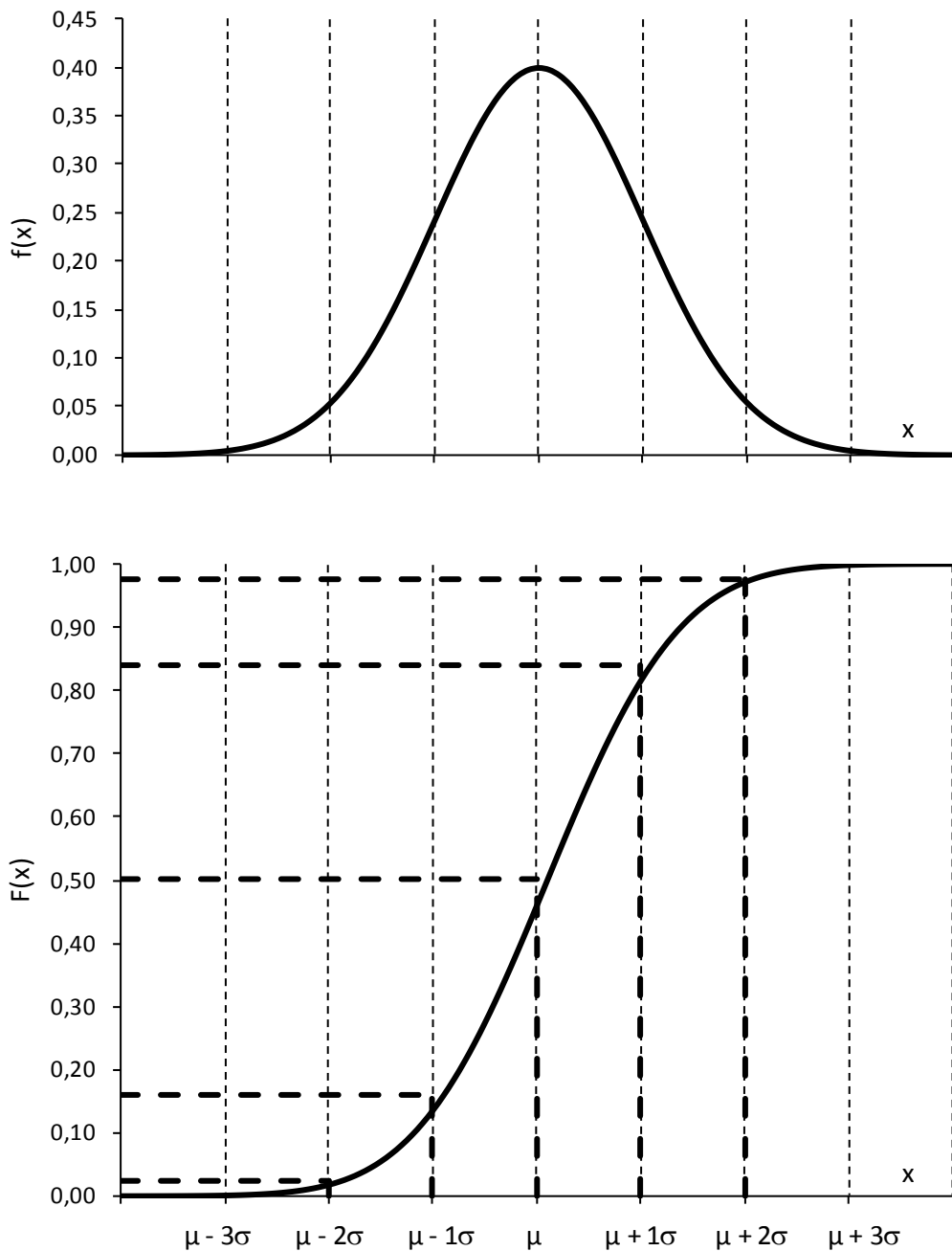


Bild 6.4: Gegenüberstellung der Wahrscheinlichkeitsdichtefunktion (oben) und der Verteilungsfunktion (unten) der Normalverteilung

Aus dieser Darstellung geht hervor, dass beim Mittelwert μ die Summenhäufigkeit 50 % beträgt. Die 0%-Linie, sowie die 100%-Linie werden von der Kurve theoretisch erst im Unendlichen berührt, bei $\mu - 3 \cdot \sigma$, bzw. $\mu + 3 \cdot \sigma$ sind jedoch die entsprechenden Linien nahezu schon erreicht. Bei $\mu - 3 \cdot \sigma$ beträgt die Summenwahrscheinlichkeit 0,135 %, bei $\mu + 3 \cdot \sigma$ hat sie den Wert 99,865 %.

Es lässt sich nun sehr leicht erkennen, dass der Anteil der Verteilung zwischen $\mu - 3 \cdot \sigma$ und $\mu + 3 \cdot \sigma$ ungefähr 99,73 % beträgt, nämlich $99,863 \% - 0,135 \%$.



6.1.3 Standardnormalverteilung

Ihre große praktische Bedeutung erhalten die Gaußkurven erst durch einen Standardisierungsprozess. Das wird verständlich, wenn man bedenkt, dass für jede beliebige Normalverteilung die dazugehörige Gaußkurve gezeichnet werden kann.

Die Standardisierung bewirkt, dass alle Gaußkurven in eine Standardkurve mit Mittelwert $\mu = 0$ und Standardabweichung $\sigma = 1$ überführt werden. Dies wird durch folgende Transformation erreicht:

$$u = \frac{x - \mu}{\sigma}$$

Durch die Subtraktion $x - \mu$ wird der Mittelwert zum Nullpunkt hin verschoben. Die Division durch die Standardabweichung entspricht einer Stauchung (bzw. Dehnung) der Abszisse, so dass sich die Standardabweichung 1 ergibt. Bild 6.6 veranschaulicht diese Skalentransformation.

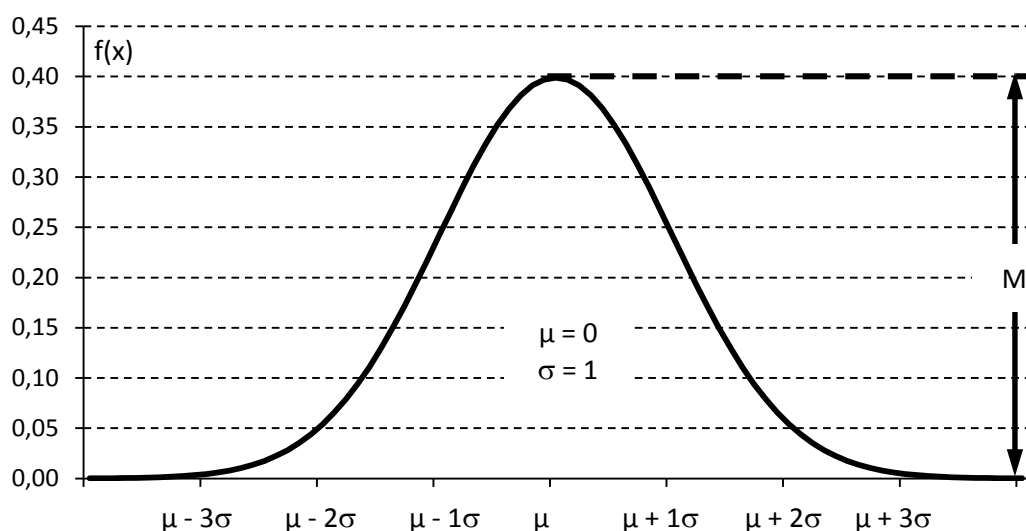


Bild 6.5: Glockenkurve der Standardnormalverteilung

Zum Zeichnen der Glockenkurve kann man folgende Näherungswerte benutzen, bei einer beliebigen Wahl des Maximums M in mm (vgl. Bild 6.5):

Abszisse	$\mu \pm 0,5 \cdot \sigma$	$\mu \pm 1 \cdot \sigma$	$\mu \pm 1,5 \cdot \sigma$	$\mu \pm 2 \cdot \sigma$	$\mu \pm 3 \cdot \sigma$
Ordinate	$\frac{7}{8} \cdot M$	$\frac{5}{8} \cdot M$	$\frac{2,5}{8} \cdot M$	$\frac{1}{8} \cdot M$	$\frac{0,1}{8} \cdot M$
	$\approx 0,88 \cdot M$	$\approx 0,63 \cdot M$	$\approx 0,33 \cdot M$	$\approx 0,13 \cdot M$	$\approx 0,01 \cdot M$

Tabelle 6.1: Näherungswerte zur grafischen Darstellung der Gaußkurve

Der Vorteil der Standardisierung liegt darin, dass für Gaußverteilungen mit beliebigen μ und σ die Wahrscheinlichkeitsdichte und damit auch die Summenfunktion (Summenhäufigkeit) nur vom Wert der Variablen u abhängt.



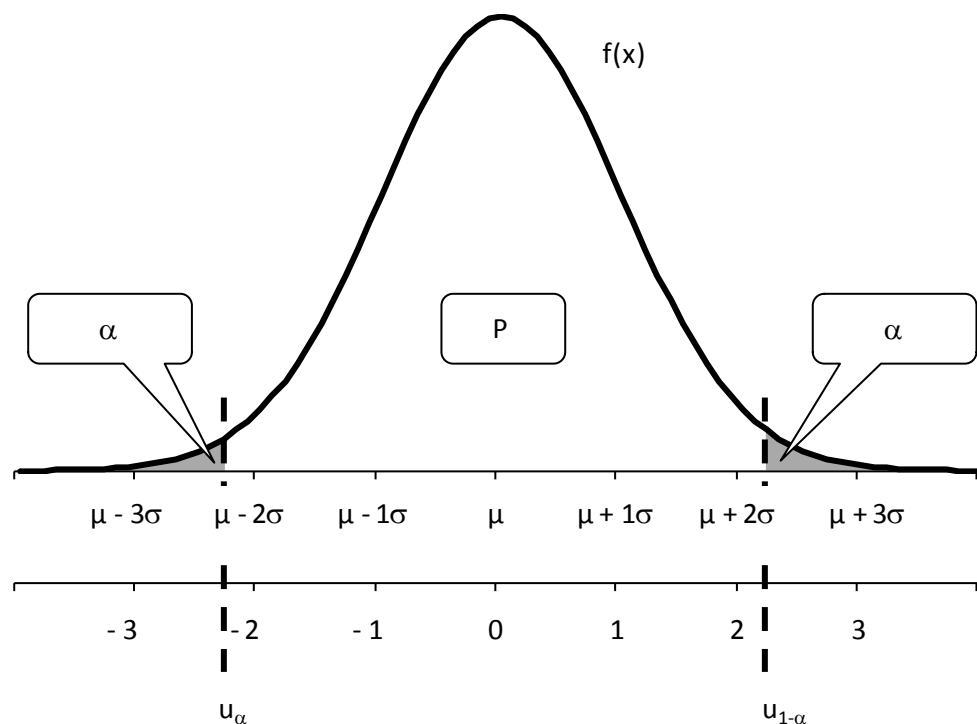


Bild 6.6: Darstellung der Überschreitungsteile

Zwei neue Begriffe tauchen in Zusammenhang mit dieser Darstellung auf:

- Die Größe P , welche die Wahrscheinlichkeit dafür angibt, dass ein zufällig gemessener Wert zwischen x_α und $x_{1-\alpha}$ liegt (engl.: Probability = Wahrscheinlichkeit) und
- der Überschreitungsteil α . Diese Größe entspricht der Wahrscheinlichkeit, mit welcher der gemessene Wert kleiner als x_α ist. Wegen der Symmetrie ist α auch gleich der Wahrscheinlichkeit, dass der gemessene Wert größer als $x_{1-\alpha}$ ist.

Die Fläche unter der Gaußkurve entspricht $P + 2 \cdot \alpha = 1 = 100\%$.

Für den Bereich $\mu \pm 3 \cdot \sigma$ ergibt sich die Wahrscheinlichkeit $P = 99,73\%$ und daher für den einseitigen Überschreitungsteil der Wert $\alpha = 0,135\%$ (wegen $2 \cdot \alpha = 0,27\%$).

Dem Überschreitungsteil α sind auf der Merkmalsachse sogenannte Streugrenzen zugeordnet und zwar eine untere, die mit x_α , und eine obere, die mit $x_{1-\alpha}$ bezeichnet wird.

Durch Umstellung der Normierungsgleichung

$$u = \frac{x - \mu}{\sigma}$$

lassen sich diese Streugrenzen leicht berechnen. Es gilt: $x_{1-\alpha} = \mu + u_{1-\alpha} \cdot \sigma$.

Da die Normalverteilung symmetrisch ist, gilt außerdem:

$$x_\alpha = \mu + u_\alpha \cdot \sigma \text{ mit } u_\alpha = -u_{1-\alpha} \Rightarrow x_\alpha = \mu - u_{1-\alpha} \cdot \sigma.$$

Nun wird der praktische Nutzen der Standardisierung verständlich: Für jede Gaußverteilung mit beliebigen μ und σ liegt zwischen den Streugrenzen x_α , und $x_{1-\alpha}$ der Anteil P der Verteilung.



Dieser allgemeingültige Zusammenhang zwischen u und P ist aus folgender Tabelle ersichtlich:

u	1,0	1,28	1,64	1,96	2,0	2,33	2,58	3,0	3,2
$P(\pm u)$ %	68,3	80,0	90,0	95,0	95,4	98,0	99,0	99,7	99,9
α %	15,9	10,0	5,0	2,5	2,28	1,0	0,5	0,135	0,05

Tabelle 6.2: Zusammenhang zwischen u und P

In Abschnitt 10 ist eine detailliertere Tabelle der Standardnormalverteilung wiedergegeben. Allerdings werden dort etwas andere Bezeichnungen verwendet. $D(u)$ entspricht der hier P genannten Größe und $\Phi(-u)$ ist mit α identisch.

6.2 Das Wahrscheinlichkeitsnetz der Normalverteilung

Das Wahrscheinlichkeitsnetz der Normalverteilung ist ein Hilfsmittel zur grafischen Bestimmung statistischer Kenngrößen eines Datensatzes. Zudem bietet es die Möglichkeit zu prüfen, ob die Werte des Datensatzes aus einer Normalverteilung stammen können oder nicht.

Die Ordinaten-Teilung des Wahrscheinlichkeitsnetzes ist derart konstruiert, dass aus Normalverteilungen stammende Datensätze Punktefolgen liefern, die auf einer Geraden liegen. Bildlich gesprochen wird die Summenkurve aus Bild 6.4 durch entsprechende Verzerrung der Ordinatenachse „gerade gebogen“. Dies wird erreicht, indem die zu ganzzahligen Vielfachen der (dimensionslosen) Größe u gehörenden Summenhäufigkeiten auf der Ordinate gleiche Abstände zueinander erhalten. Die Abszisse, also die Merkmalsachse (hier für die u dargestellt, vgl. Abschnitt 6.1.3), hat eine lineare Teilung.

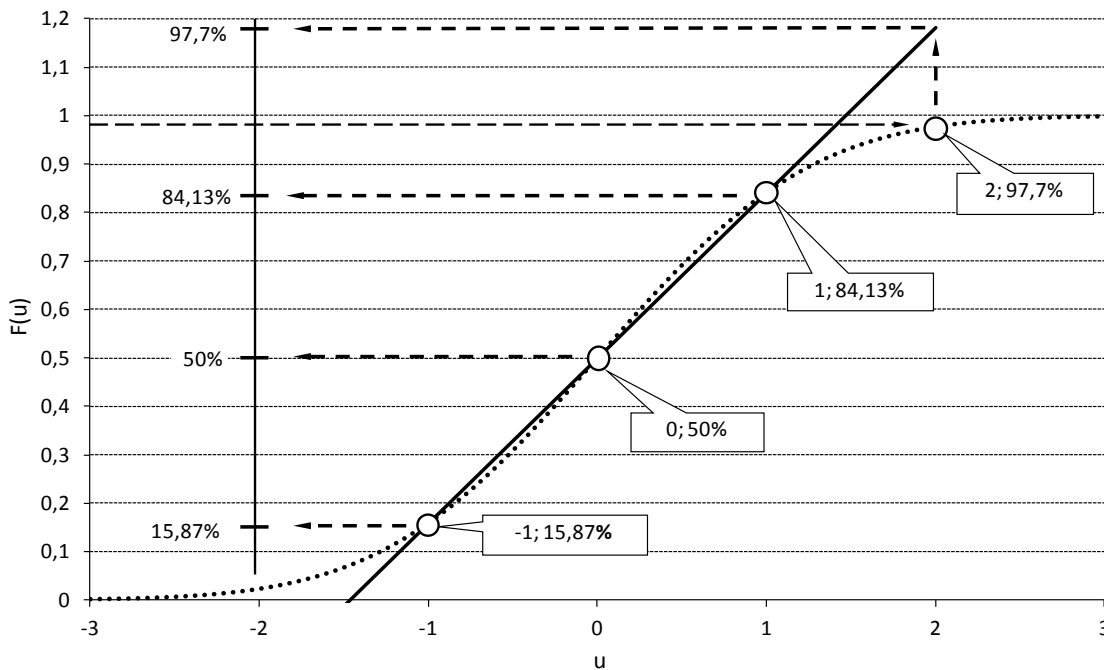


Bild 6.7: Schema zur Entstehung des Wahrscheinlichkeitsnetzes der Normalverteilung



Da normalerweise die aus einer Stichprobe ermittelten Werte der relativen Summenhäufigkeit nur angenähert der theoretischen Normalverteilung folgen, sind die entsprechenden Punkte im Wahrscheinlichkeitsnetz durch eine Ausgleichsgerade zu approximieren.

Die Anwendung und der praktische Nutzen des Wahrscheinlichkeitsnetzes werden im folgenden Abschnitt erläutert.

Häufigkeitsschaubild mit Wahrscheinlichkeitsnetz

Mit Hilfe des Wahrscheinlichkeitsnetzes der Normalverteilung kann man prüfen, ob für den betrachteten Datensatz die Annahme dieses Verteilungsmodells gerechtfertigt ist oder nicht. Darüber hinaus können statistische Kenngrößen des Datensatzes auf einfache Weise ermittelt/abgelesen werden.

Der relativen Summenhäufigkeit	entspricht
50 %	der Mittelwert \bar{x}
99,865 %	der Wert $\bar{x} + 3 \cdot s$
0,135 %	der Wert $\bar{x} - 3 \cdot s$

Diese Zusammenhänge begründen die Handhabung des Wahrscheinlichkeitsnetzes.

Bei Messreihen, die aus mehr als 25 Messwerten bestehen, und die mit Hilfe einer Klasseneinteilung ausgewertet werden, trägt man die Werte der relativen Summenhäufigkeit an den oberen Klassengrenzen in das Wahrscheinlichkeitsnetz ein.

Bei kleineren Messreihen erstellt man zunächst eine geordnete Liste (Messwerte der Größe nach ordnen) und ordnet diesen Werten Summenhäufigkeiten $H_i(n)$ zu, die mit Hilfe

der Näherungsformel $H_i(n) = \frac{i - 0,3}{n + 0,4}$ berechnet werden können ($i = 1, 2, \dots, n$).

$$x_{(1)}, x_{(2)}, \dots, x_{(n)}$$

$$H_1(n), H_2(n), \dots, H_n(n).$$

Anschließend werden die Punkte $(x_{(i)}, H_i(n))$ ins Wahrscheinlichkeitsnetz eingezeichnet.

Durch die sich ergebende Punktefolge wird eine Ausgleichsgerade in der Weise gezeichnet, dass die Abstände der Punkte zur Geraden möglichst klein werden und etwa gleich viele Punkte oberhalb und unterhalb der Geraden liegen. Je besser die Punktefolge durch die Gerade approximierbar ist, um so eher ist das Modell der Normalverteilung zur Beschreibung des ausgewerteten Datensatzes geeignet.



Die grafische Bestimmung statistischer Kenngrößen sei anhand des Beispiels 6.1 erläutert.

Grafische Bestimmung des Mittelwerts

Den Mittelwert \bar{x} findet man, indem man den Schnittpunkt der horizontalen Linie bei 50 % Summenhäufigkeit mit der Ausgleichsgeraden aufsucht und den zugehörigen Wert auf der x-Achse abliest.

Man findet $\bar{x} = 346$.

Grafische Bestimmung der Standardabweichung

Die Standardabweichung s lässt sich ermitteln, indem man die zu den Summenhäufigkeiten 99,865 % (entspricht $\bar{x} + 3 \cdot s$) und 0,135 % (entspricht $\bar{x} - 3 \cdot s$) gehörenden x -Werte abliest — im vorliegenden Beispiel sind dies die Werte $x = 366$ und $x = 324$ — und deren Differenz durch 6 teilt.

$$\text{Man findet also: } s = \frac{366 - 324}{6} = \frac{42}{6} = 7.$$

Grafische Bestimmung eines Überschreitungsanteils

Im vorliegenden Fall ist der obere Grenzwert $OGW = 360$ („obere Toleranzgrenze“). Den theoretischen Anteil der Grundgesamtheit, der OGW überschreitet, findet man, indem man an dieser Stelle eine senkrechte Linie zeichnet und deren Schnittpunkt mit der Ausgleichsgeraden bestimmt. Die zu diesem Schnittpunkt gehörende Summenhäufigkeit entspricht dem Anteil der Grundgesamtheit, der unterhalb OGW liegt, hier also etwa 97,5 %.

Der gesuchte Überschreitungsanteil (einseitig rechts) ist dann die Differenz zu 100 % also 2,5 %.

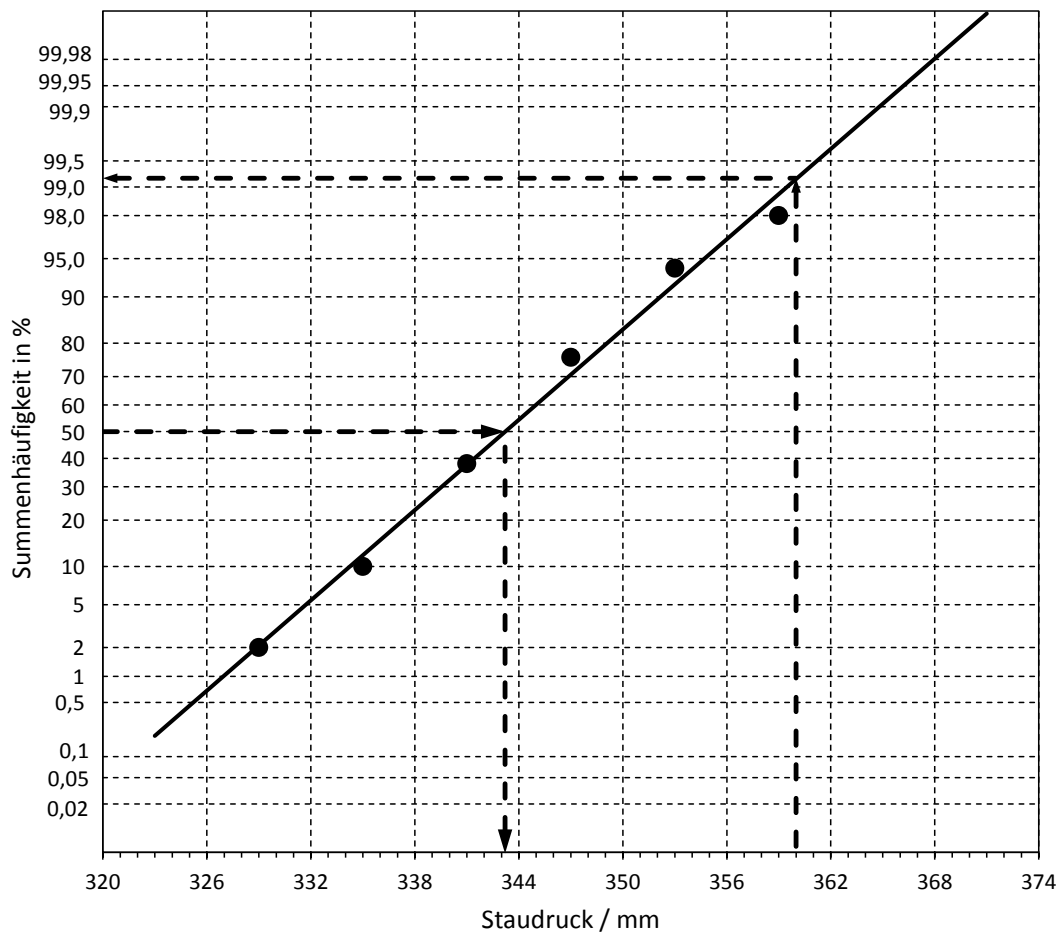
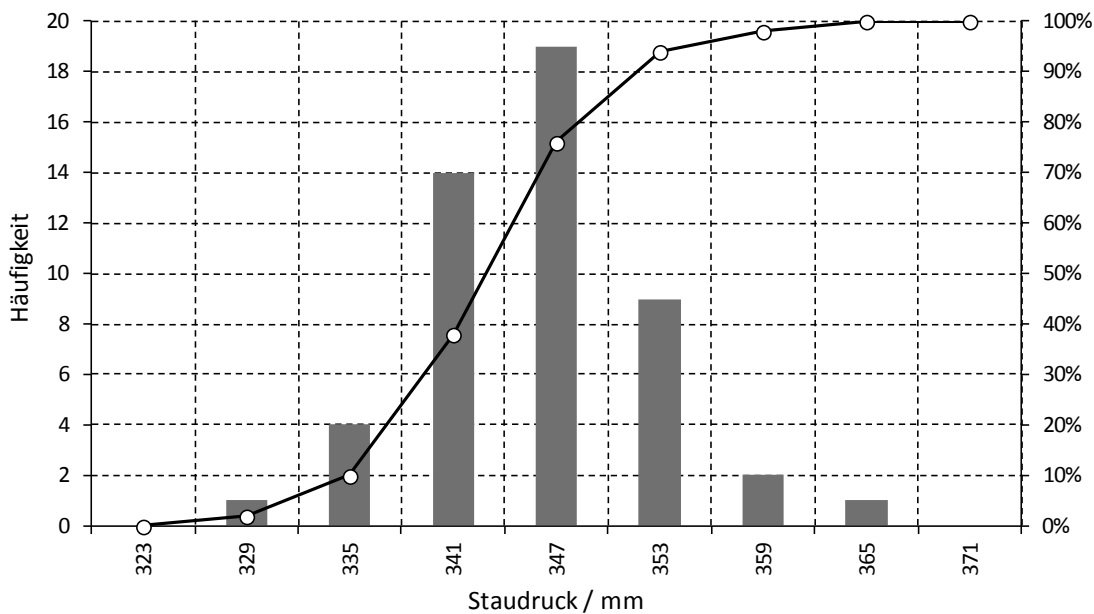
Der Anteil, der einen bestimmten unteren Grenzwert UGW („untere Toleranzgrenze“) unterschreitet, ergibt sich analog dazu. Man muss also bei UGW eine Senkrechte zeichnen und die zum Schnittpunkt mit der Ausgleichsgeraden gehörige Summenhäufigkeit ablesen. Dieser Wert entspricht unmittelbar dem gesuchten Überschreitungsanteil (einseitig links).

Der sogenannte zweiseitige Überschreitungsanteil ist die Summe aus dem unteren (einseitig linken) und dem oberen (einseitig rechten) Überschreitungsanteil. Er gibt den Anteil der Einheiten der Grundgesamtheit an, deren Merkmalswerte außerhalb des Toleranzbereichs liegen.



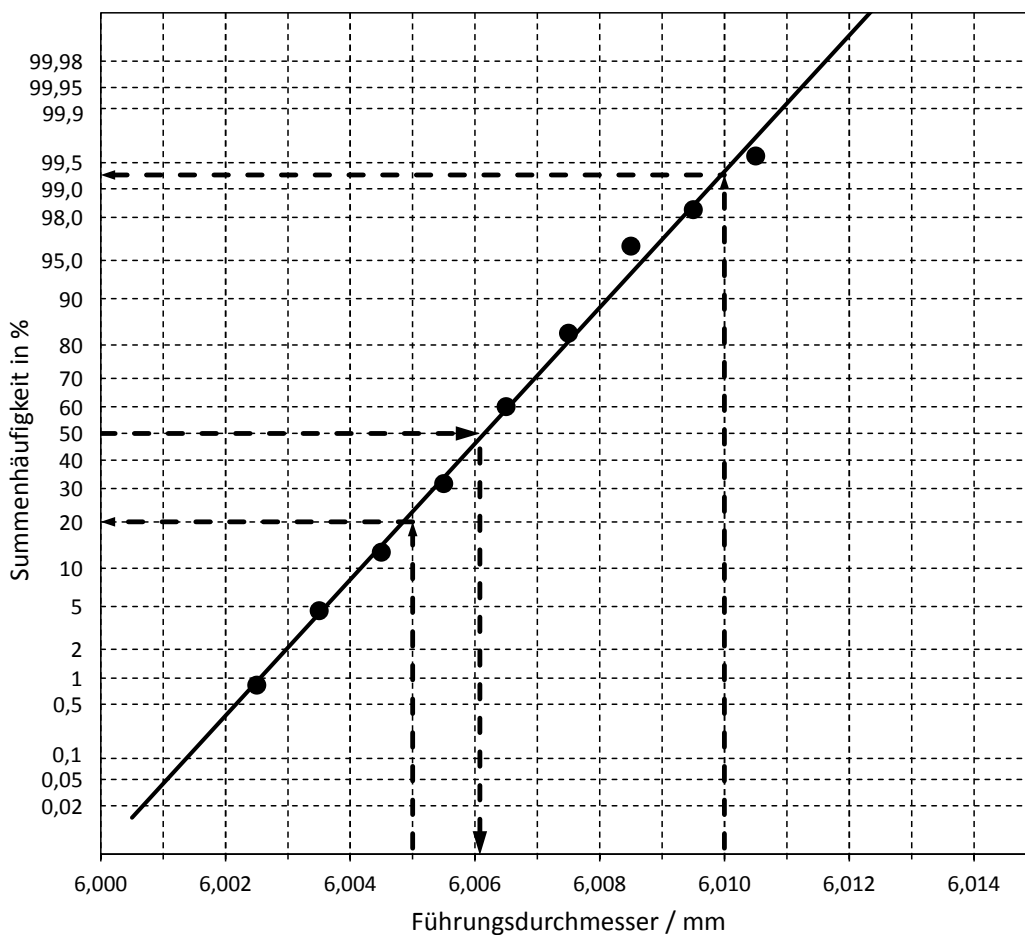
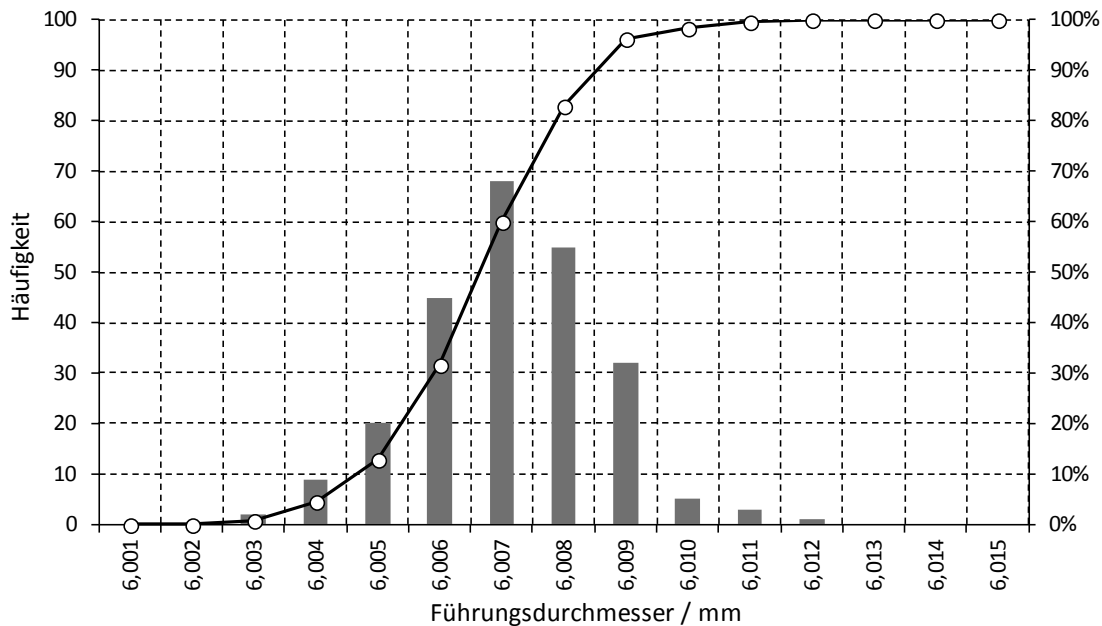
Beispiel 6.1: Staudruckwerte von Ventilkugeln

Hier ist die Häufigkeitsverteilung einer Stichprobe von 50 Messwerten dargestellt.



Beispiel 6.2: Führungsdurchmesser von Düsennadeln

Diese „Stichprobe“ umfasst $n = 240$ Messwerte.



BEISPIEL 6.3: Staudruckwerte an Ventilkegeln

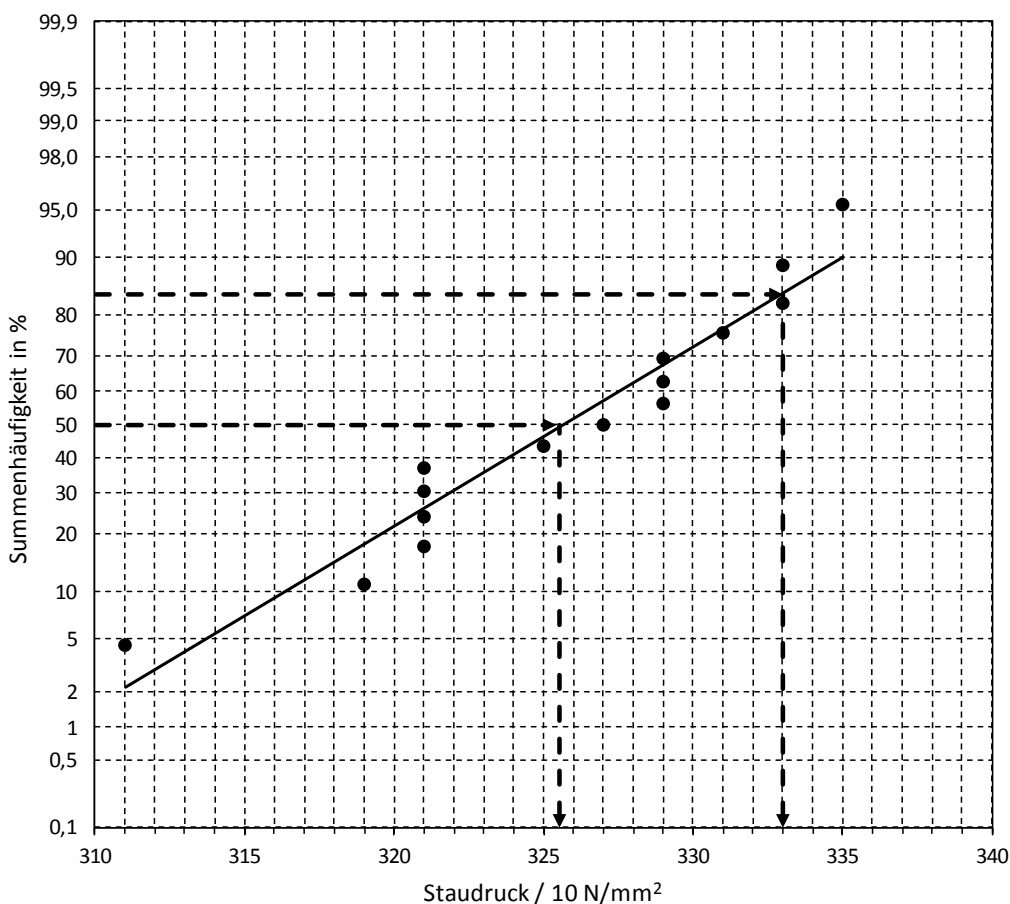
Dieses Beispiel zeigt die Auswertung einer „kleinen Messreihe“ ($n = 15$ Messwerte) im Wahrscheinlichkeitsnetz. In folgender Tabelle sind die Messwerte x_i (10 N/mm^2) und die zugehörigen relativen Summenhäufigkeiten $H_i(n)$ zur Eintragung der Punkte $(x_{(i)}, H_i(n))$ ins Wahrscheinlichkeitsnetz (in aufsteigender Reihenfolge) angegeben.

Letztere wurden mit Hilfe der Näherungsformel $H_i(n) = \frac{i - 0,3}{n + 0,4}$ für $i = 1, 2, \dots, 15$ berechnet.

Beispiel: $H_2(15) = \frac{2 - 0,3}{15 + 0,4} \approx 0,11 = 11\%$.

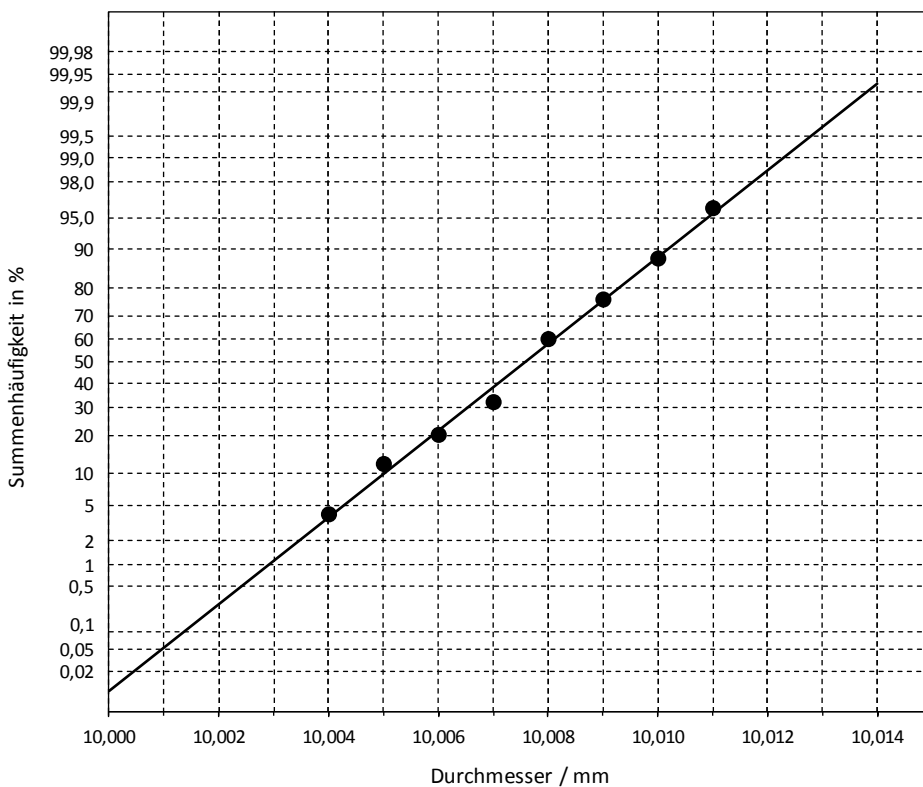
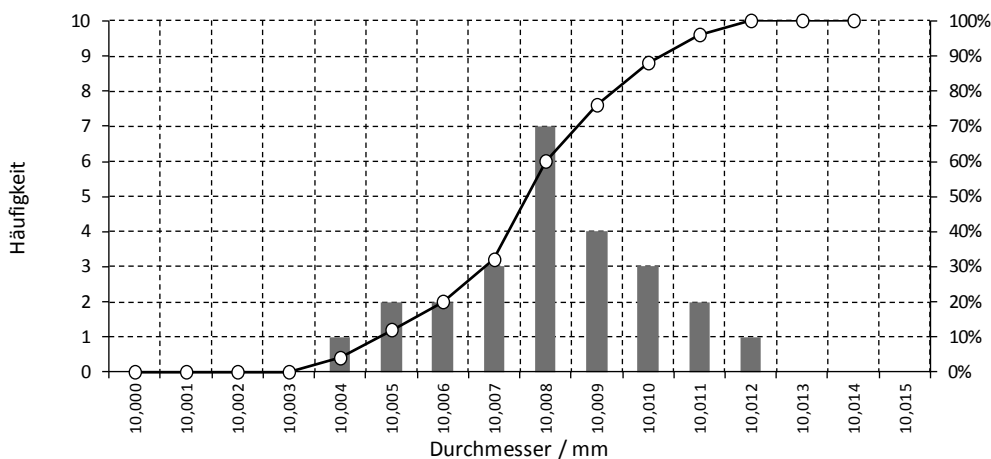
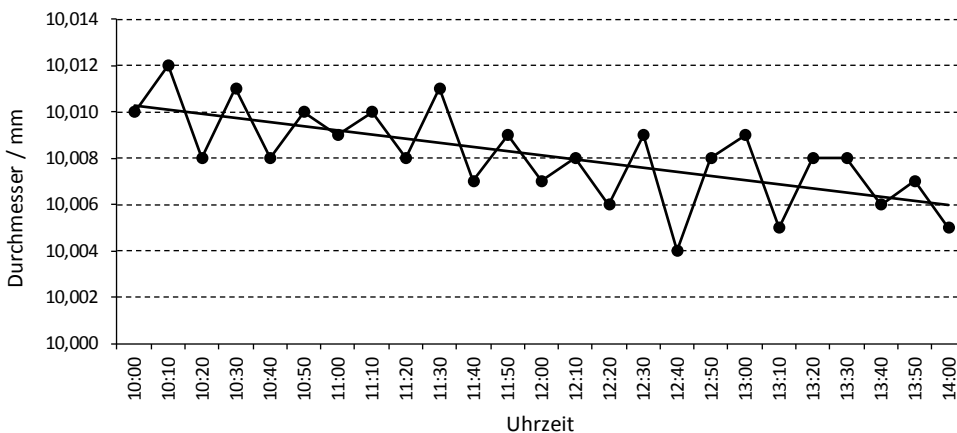
Nr.	1	2	3	4	5	6	7	8
x_i	311	319	321	321	321	321	325	327
$H_i(n)$	4,5	11,0	17,5	24,0	30,5	37,0	43,5	50,0
Nr.	9	10	11	12	13	14	15	
x_i	329	329	329	331	333	333	335	
$H_i(n)$	56,5	63,0	69,5	76,0	82,5	89,0	95,5	

Der Maßstab auf der Merkmalsachse muss mit etwas Überlegung gewählt werden. Einerseits sollte die Ausgleichsgerade noch vollständig gezeichnet werden können, damit man möglichst die Schnittpunkte am oberen und unteren Rand des Netzes ablesen kann. Andererseits sollte die Gerade aber im Interesse der Ablesegenauigkeit nicht zu steil verlaufen.



BEISPIEL 6.4: Untersuchung des Prozesses „Buchsen schleifen“

In diesem Fall wurden die Messwerte ähnlich wie bei einer Urwertkarte in zeitlicher Reihenfolge aufgetragen. Aus dieser Darstellung ergibt sich in einfacher Weise das Häufigkeitsschaubild.



6.3 Logarithmische Normalverteilung

Wenn ein Merkmal aus physikalischen Gründen eine bestimmte Grenze nicht über- bzw. unterschreiten kann, ergibt sich häufig eine sogenannte schiefe (asymmetrische) Verteilung der Merkmalswerte. Das ist z.B. der Fall bei den Merkmalen Durchbiegung und Rundheit, bei denen der untere Grenzwert „Null“ ist oder bei der Rockwell-Härteprüfung von Stahl, bei der eine bestimmte Mindesthärte nicht unterschritten wird. Die Auswertung des Häufigkeitsschaubildes einer solchen schiefen Verteilung im Wahrscheinlichkeitsnetz der Normalverteilung ergibt eine gekrümmte Linie (Bild 6.8).

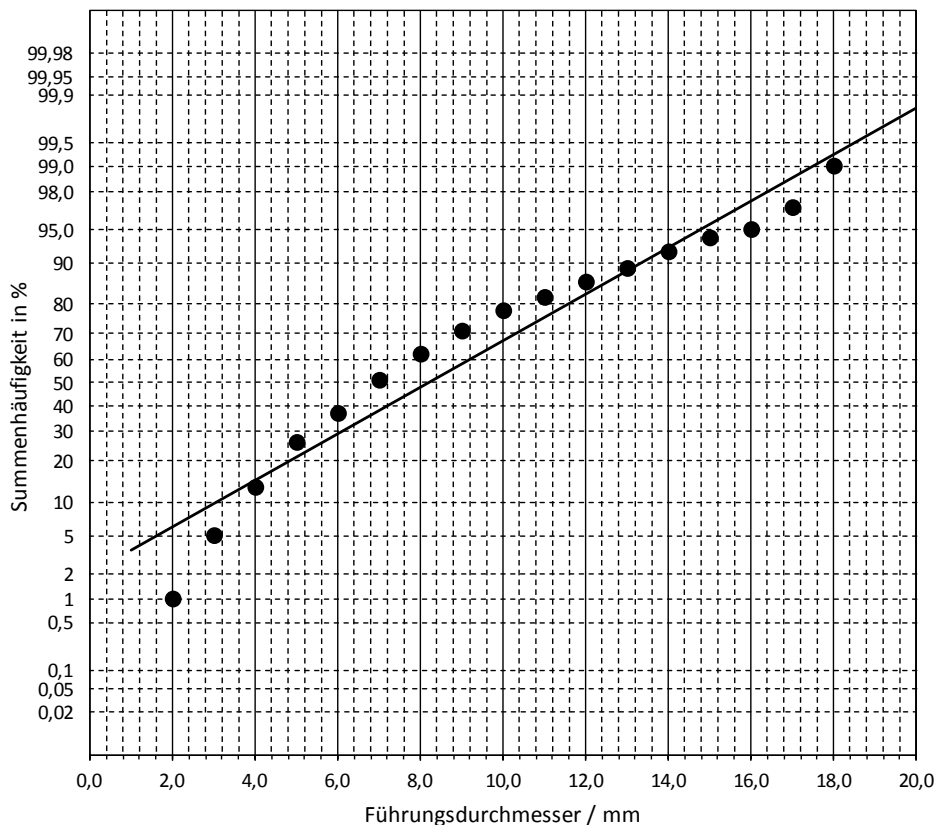
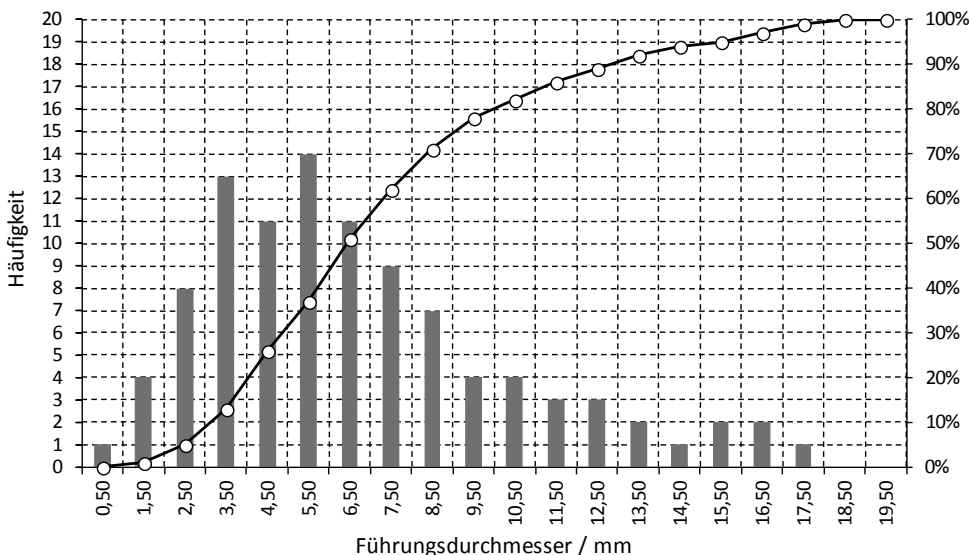


Bild 6.8: Auswertung einer schiefen Verteilung im Wahrscheinlichkeitsnetz der Normalverteilung



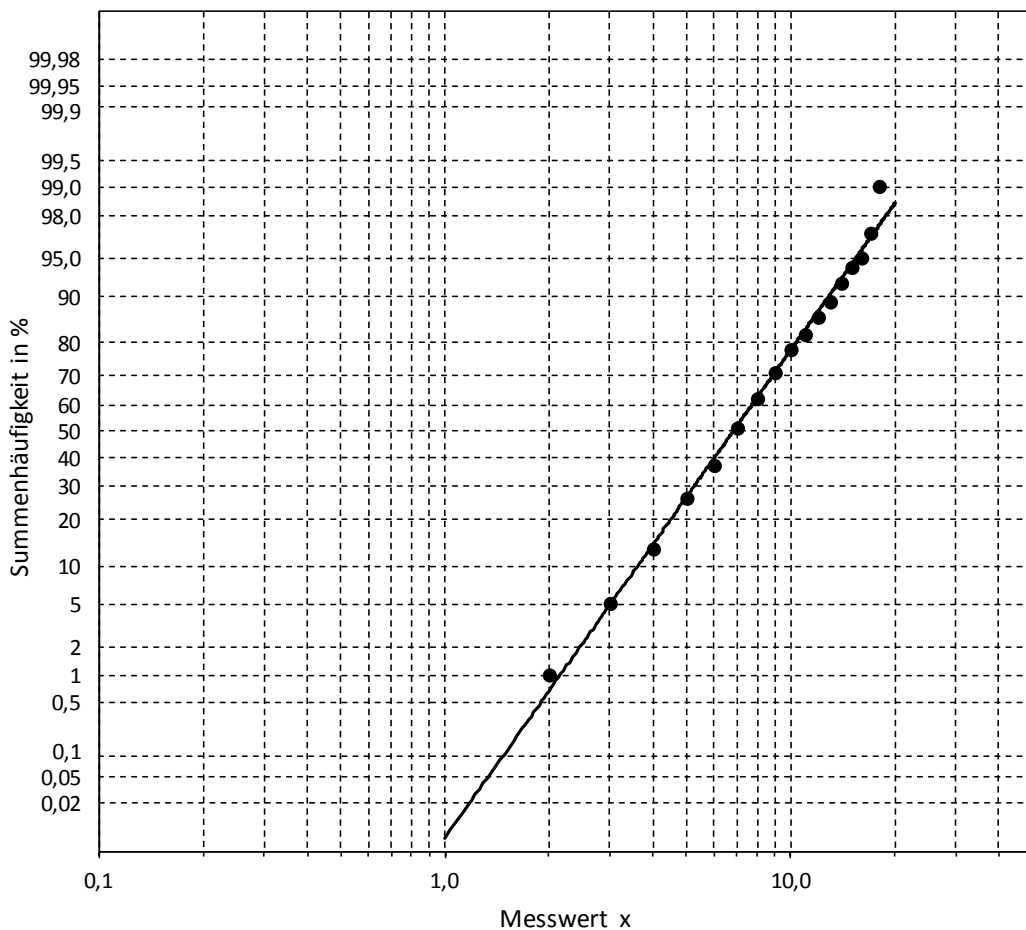
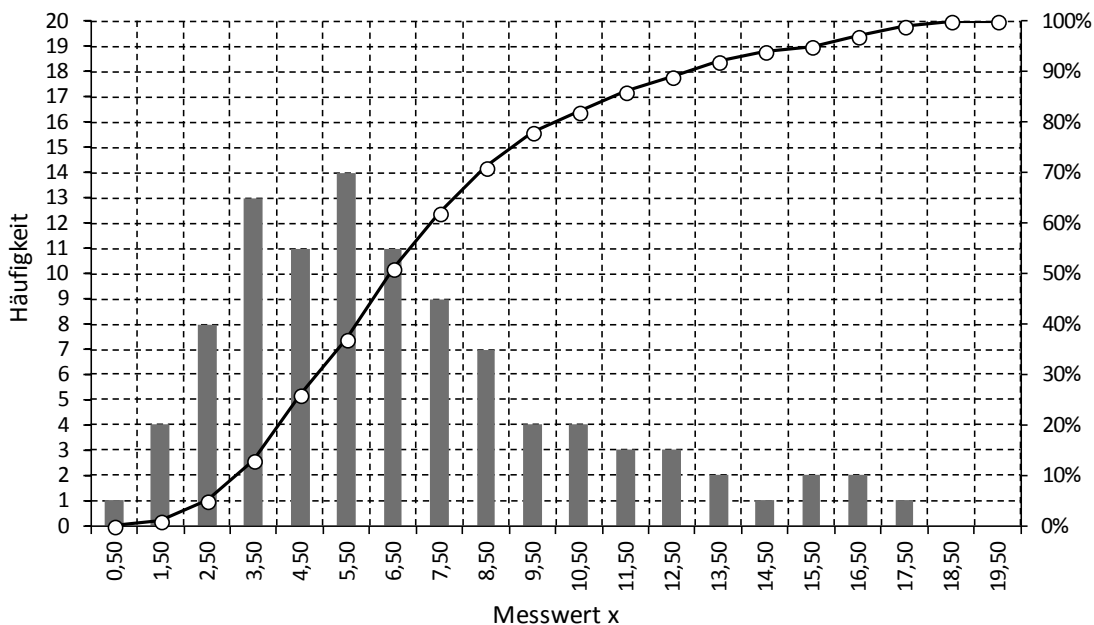


Bild 6.9: Gleicher Datensatz im Wahrscheinlichkeitsnetz der logarithmischen Normalverteilung

Stellt man dagegen den gleichen Datensatz im Wahrscheinlichkeitsnetz der Lognormalverteilung dar, so ergibt sich näherungsweise ein geradliniger Punkteverlauf (vgl. Bild 6.9).



6.3.1 Wahrscheinlichkeitsnetz der Lognormalverteilung

Die Ordinatenskala des Wahrscheinlichkeitsnetzes der Lognormalverteilung ist mit der Skala des Wahrscheinlichkeitsnetzes der Normalverteilung identisch. Die beiden Netze unterscheiden sich lediglich in der Skalenteilung der Abszisse (x-Achse). Beim logarithmischen Wahrscheinlichkeitsnetz ist diese Skalenteilung logarithmisch. Wegen der in folgender Tabelle angegebenen Beziehungen (die ersten beiden Spalten enthalten unterschiedliche Schreibweisen für die gleiche Zahl x) haben auf einer logarithmisch geteilten Achse z.B. die den Werten 10, 100, 1000 entsprechenden Stellen gleiche Abstände.

x	x	log(x)
$\frac{1}{100} = 0,01$	10^{-2}	-2
$\frac{1}{10} = 0,1$	10^{-1}	-1
1	10^0	0
10	10^1	1
100	10^2	2
1000	10^3	3

Es ist somit auf einfache Weise möglich, Datensätze darzustellen, die sich über zwei oder mehr Zehnerpotenzen erstrecken (vgl. Bild 6.9). Es ist zu beachten, dass bei einer logarithmischen Skala niemals der Wert Null auftauchen kann, da die Position, an der die Null stehen müsste, bei „minus unendlich“ liegt.

Die Eintragung der Punkte erfolgt wie im „normalen“ Wahrscheinlichkeitsnetz (vgl. Abschnitt 6.2).

Kenngrößen der Lognormalverteilung sind der geometrische Mittelwert \bar{x}_g und der Streufaktor ε („epsilon“). \bar{x}_g ist der Median, d.h. links und rechts von dieser Zahl liegen jeweils 50 % der Werte der Verteilung. Er kann also grafisch leicht ermittelt werden, indem der Schnittpunkt der eingezeichneten Ausgleichsgeraden mit der horizontalen Linie bei 50 % aufgesucht und der zugehörige x-Wert abgelesen wird.

Der Mittelwert \bar{x} ist wegen der Asymmetrie der Verteilung nicht mit dem häufigsten Wert (Modalwert) identisch, sondern liegt zwischen diesem und dem Median \bar{x}_g .

Zur Ermittlung des Streufaktors ε bestimmt man zunächst im Wahrscheinlichkeitsnetz den Schnittpunkt der Ausgleichsgeraden mit der horizontalen Linie bei $\bar{x}_g \cdot \varepsilon$ (entsprechend 84,13 % Summenhäufigkeit), liest den zugehörigen Wert auf der x-Achse ab und teilt diese Zahl schließlich durch \bar{x}_g .

Ähnlich wie bei der Normalverteilung können mit Hilfe von \bar{x}_g und ε Flächenanteile bestimmt werden, die innerhalb bestimmter Intervalle von der Kurve der Dichtefunktion und der x-Achse begrenzt werden.



Zwischen $\frac{\bar{x}_g}{\varepsilon}$ und $\bar{x}_g \cdot \varepsilon$ liegen 68,3 % aller Werte einer Lognormalverteilung.

Zwischen $\frac{\bar{x}_g}{\varepsilon^2}$ und $\bar{x}_g \cdot \varepsilon^2$ liegen 95,4 % aller Werte einer Lognormalverteilung.

Zwischen $\frac{\bar{x}_g}{\varepsilon^3}$ und $\bar{x}_g \cdot \varepsilon^3$ liegen 99,7 % aller Werte einer Lognormalverteilung.

In der Regel ist im Zusammenhang mit nullbegrenzten Merkmalen nur ein oberer Grenzwert OGW von Interesse. In diesem Fall kann im logarithmischen Wahrscheinlichkeitsnetz der zugehörige Überschreitungsanteil bestimmt werden. Dazu zeichnet man an der Stelle OGW (auf der x -Achse) eine senkrechte Linie und bestimmt den zum Schnittpunkt mit der Ausgleichsgeraden gehörenden Summenhäufigkeitswert $H(\text{OGW})$ in Prozent. Der Wert $100 - H(\text{OGW})$ entspricht dem theoretischen Anteil der Verteilung (in Prozent), der den Grenzwert überschreitet.

Bild 6.10 zeigt ein Beispiel für eine solche Auswertung. Man beachte, dass das Häufigkeitsdiagramm (oben) und das Wahrscheinlichkeitsnetz (unten) unterschiedliche Abszissenskalen besitzen (lineare und logarithmische Teilung) und die Punkte im Wahrscheinlichkeitsnetz daher nicht mehr senkrecht unter den Positionen der oberen Klassengrenzen des Häufigkeitsdiagramms zu liegen kommen.

Geht man auf der horizontalen Linie bei 50 % Summenhäufigkeit nach rechts bis zur Ausgleichsgeraden und an diesem Schnittpunkt dann senkrecht nach unten, so findet man den geometrischen Mittelwert $\bar{x}_g = 0,31$. Entsprechend findet man zum Schnittpunkt der 84,13%-Linie mit der Ausgleichsgeraden den Wert $\bar{x}_g \cdot \varepsilon = 0,51$. Division dieser beiden

Zahlen liefert den Streufaktor $\varepsilon = \frac{0,51}{0,31} = 1,65$. Verfolgt man die Senkrechte beim oberen

Grenzwert $\text{OGW} = 1,0$ so findet man den Schnittpunkt mit der Ausgleichsgeraden bei 99 % Summenhäufigkeit. Der Überschreitungsanteil bzgl. dieser Grenze ist also 1 %.



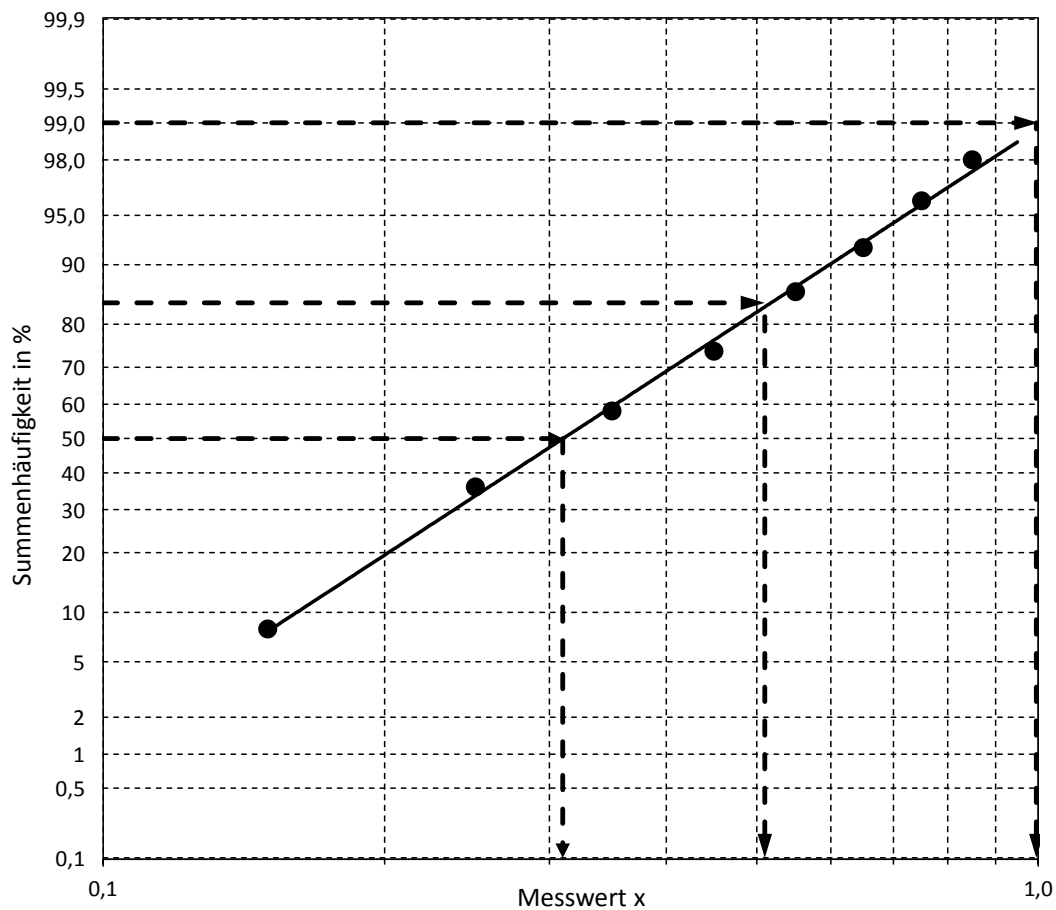
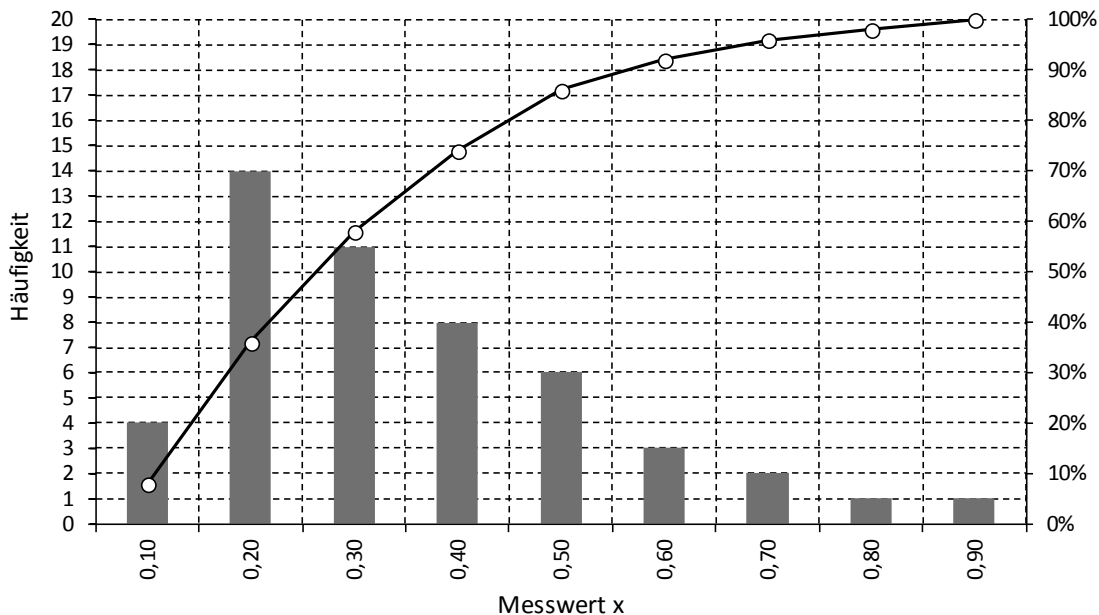


Bild 6.10: Grafische Auswertung im logarithmischen Wahrscheinlichkeitsnetz



6.3.2 Zusammenhang zwischen Normalverteilung und Lognormalverteilung

Durch Logarithmieren wird ein lognormalverteiltes Merkmal x in ein normalverteiltes Merkmal z überführt. Durch das Logarithmieren wird der obere Teil der Lognormalverteilung stark gestaucht und der Bereich zwischen dem geometrischen Mittelwert \bar{x}_g und dem Nullpunkt stark gedehnt. Dies entspricht einer „Spiegelung“ an der Kurve der Funktion $z = \ln(x)$. Bild 6.11 veranschaulicht, dass \bar{x}_g auf \bar{z} abgebildet wird und die Stelle $\bar{x}_g \cdot \varepsilon$ dem Wert $\bar{z} + s_z$ entspricht. Es ist:

$$\bar{z} = \ln(\bar{x}_g) \quad \Leftrightarrow \quad \bar{x}_g = e^{\bar{z}} \quad \text{und} \quad s_z = \ln(\varepsilon) \quad \Leftrightarrow \quad \varepsilon = e^{s_z}.$$

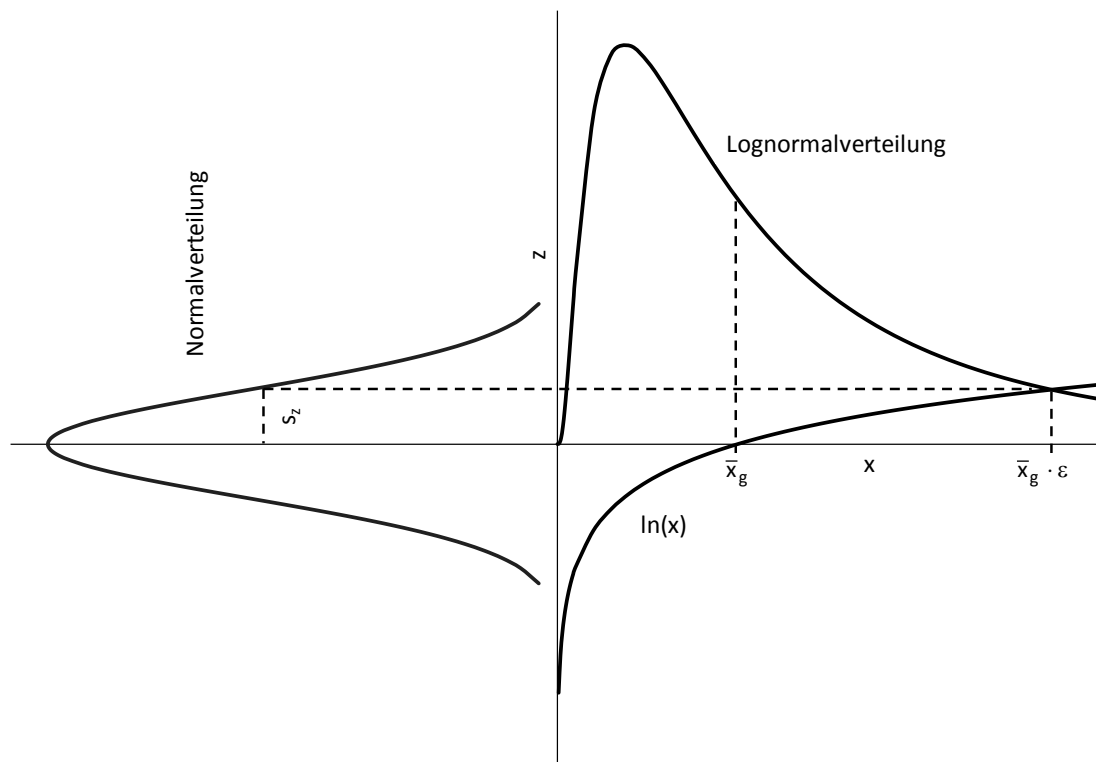


Bild 6.11: Darstellung zur Veranschaulichung des Zusammenhangs zwischen der Normalverteilung und der Lognormalverteilung

HINWEIS:

Es mag etwas verwirrend erscheinen, dass die vorstehenden Erläuterungen sich auf den natürlichen Logarithmus $\ln(x)$ (zur Basis e) beziehen, das Wahrscheinlichkeitsnetz aber eine dem dekadischen Logarithmus $\log(x)$ (zur Basis 10) entsprechende Teilung aufweist. Diese Tatsache ist jedoch hinsichtlich der Vergleichbarkeit der Auswertergebnisse ohne Bedeutung, da bei der Berechnung von \bar{x}_g und ε stets eine Rücktransformation in das ursprüngliche Koordinatensystem erfolgt, und Berechnungen von Überschreitungsanteilen oder Grenzwerten nur mit Hilfe dieser beiden Größen durchgeführt werden. Bei Verwendung von $\log(x)$ sowie dessen Umkehrfunktion 10^x werden obige Ausdrücke durch $\bar{x}_g = 10^{\bar{z}}$ und $\varepsilon = 10^{s_z}$ ersetzt.



6.4 Mischverteilungen

Es kann vorkommen, dass die Maße von Teilen, die auf zwei verschiedenen Maschinen oder Fertigungslinien hergestellt wurden, verschiedene Verteilungen aufweisen. In der Regel wird man beobachten, dass bei gleichem Verteilungstyp Mittellage und/oder Streuung voneinander abweichen. Werden die Teile nicht getrennt, so äußert sich dies statistisch gesehen durch Entstehung einer Mischverteilung. Mischverteilungen können sich aber auch bilden, wenn sich in der laufenden Produktion wesentliche Einflussgrößen praktisch sprunghaft ändern (Werkzeugwechsel, Wechsel der Materialcharge).

Das Histogramm einer Mischverteilung zeigt in der Regel zwei oder mehr Maxima. Man spricht auch von einer bimodalen bzw. einer multimodalen Verteilung.

Sofern die Mischverteilung durch Überlagerung zweier Normalverteilungen entstanden ist, deren Mittelwerte stark differieren, so zeigt eine Auftragung (der Summenhäufigkeiten bzw. der Einzelwerte) im Wahrscheinlichkeitsnetz, dass die entstehende Punktefolge sich stückweise durch zwei verschiedene Geraden annähern lässt.



2020-04-06 - SOCOS

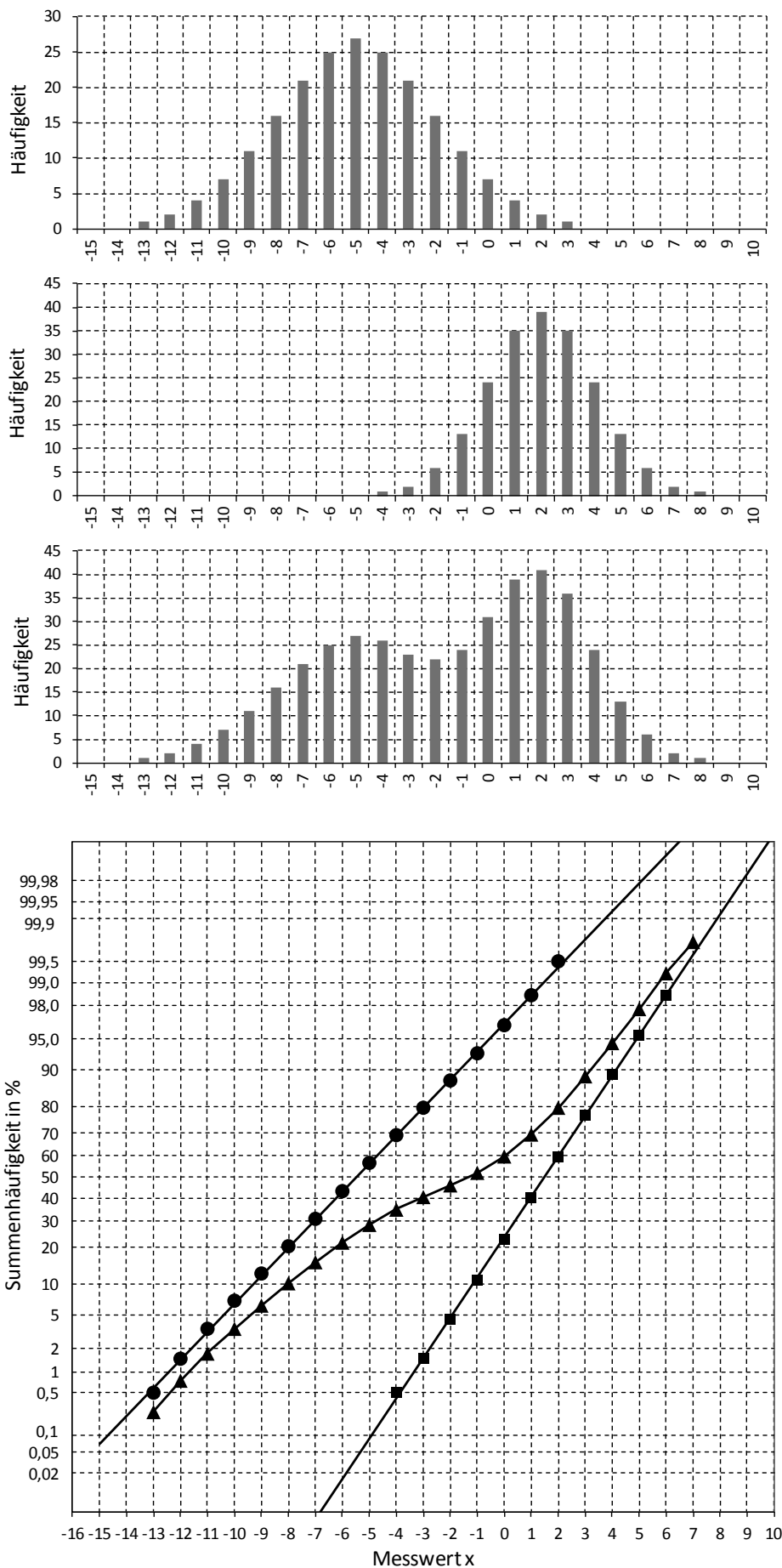


Bild 6.12: Darstellung einer aus zwei Kollektiven entstandenen Mischverteilung



7 Qualitätsregelkarten

HINWEIS:

In diesem Kapitel werden lediglich statistische Grundlagen der Regelkartentechnik dargestellt. Spezielle Vorgehensweisen und aktuelle Regelungen bzgl. deren praktischer Anwendung sind im Heft 7 „Statistische Prozessregelung (SPC)“ der Bosch-Schriftenreihe „Qualitätssicherung in der Bosch-Gruppe, Technische Statistik“ beschrieben.

Die Statistische Prozessregelung (**Statistical Process Control**) ist ein bewährtes Verfahren zur Regelung eines Fertigungsprozesses auf der Grundlage statistischer Methoden.

Dazu werden dem Prozess nach prozessspezifischen Entnahmeregeln Stichproben von Teilen entnommen, deren Merkmalswerte werden gemessen und in die sogenannte Qualitätsregelkarte eingetragen. Aus den Merkmalswerten berechnete statistische Kenngrößen werden sodann zur Beurteilung des aktuellen Prozesszustands herangezogen. Gegebenenfalls wird der Prozesszustand durch geeignete Maßnahmen korrigiert.

Die Regelkartentechnik wurde von Walter Andrew Shewhart (1891-1967) in den zwanziger Jahren entwickelt und 1931 in seinem Werk „Economic Control of Quality of Manufactured Product“ ausführlich beschrieben.

SPC ist eine Anwendung aus dem Bereich der induktiven (schließenden) Statistik. Es liegen nicht alle Messwerte vor, so wie es bei einer 100%-Prüfung der Fall wäre. Von einem kleinen Datensatz, den Stichprobenwerten, wird auf die Grundgesamtheit geschlossen.

Das mathematische Modell für variable Größen basiert auf der Vorstellung, dass auf einen Prozess viele Einflussgrößen einwirken. Die „5 M“ — Mensch, Maschine, Material, Mitwelt, Methode — bilden die Hauptgliederungspunkte der Einflussgrößen.

Jedes „M“ lässt sich weiter untergliedern, z.B. Mitwelt (= Umwelt, Milieu) in Temperatur, Luftfeuchtigkeit, Erschütterung, Schmutz, Beleuchtung,

Das unkontrollierbare, zufällige Einwirken vieler Einflussgrößen führt trotz sorgfältigen Vorgehens zu Abweichungen der realen Merkmalswerte vom Zielwert (i.a. Mittenwert des Toleranzbereichs).

Aus dem zufälligen Zusammenwirken vieler Einflussgrößen ergibt sich für das betrachtete Teilemerkmal i.a. näherungsweise eine Gaußsche Normalverteilung. Dieser Sachverhalt wird durch den zentralen Grenzwertsatz der Statistik beschrieben. Die Normalverteilung ist darum von grundlegender Bedeutung für die SPC.

7.1 Regelkarten der Lage

Einen ersten Hinweis auf die Prozesslage erhält man bereits, indem man ein gefertigtes Teil als Zufallsstichprobe entnimmt, das interessierende Merkmal misst und den gefundenen Merkmalswert mit der Schätzung des Mittelwerts $\hat{\mu}$ vergleicht. Liegt der Einzelwert innerhalb eines Bereiches von beispielsweise $\mu \pm 3\sigma$, so wird dieses Ergebnis nicht weiter verwundern, da unter den genannten Voraussetzungen bei rein zufälligem Prozessverhalten etwa 99,73 % aller Merkmalswerte in diesem Bereich liegen sollten. Man könnte nun bereits aus dem einen gefundenen Ergebnis den Schluss ziehen, dass die momentane Prozessmittellage mit derjenigen des Vorlaufs (Schätzwert $\hat{\mu}$) übereinstimmt, bzw. dass es keinen Hinweis auf eine Verschiebung der Prozesslage gibt.



Das Vertrauen in eine solche Aussage wäre allerdings deutlich größer, wenn sie nicht nur durch einen einzelnen Stichprobenwert begründet wäre, sondern durch beispielsweise $n = 5$ Werte. Die vorstehenden Überlegungen bilden die Grundzüge der Funktionsweise einer Qualitätsregelkarte. Aus dem Ergebnis einer aktuellen 5er-Stichprobe wird ein Schluss auf die momentane Prozesslage gezogen. Offen ist jedoch noch die Frage, welche Kenngröße(n) als Träger der Information über die Prozesslage betrachtet werden soll(en).

Neben der Möglichkeit, alle fünf Einzelwerte getrennt zu betrachten, bietet sich an, eine Beurteilung aufgrund der Lage des Mittelwerts \bar{x} bzw. des Medians (Zentralwert) \tilde{x} der fünf Werte vorzunehmen. Diesen drei Möglichkeiten entsprechen drei unterschiedliche Regelkarten der Lage, die Urwertkarte, die Mittelwertkarte und die Mediankarte. Wir wollen uns in dieser Unterlage auf eine Darstellung der Mittelwertkarte und der Urwertkarte beschränken.

7.1.1 Mittelwertkarte

Die Mittelwertkarte ist die wichtigste und in der Praxis am häufigsten anzutreffende Qualitätsregelkarte. Ihre Funktionsweise soll im Folgenden erklärt werden.

In gleichbleibenden zeitlichen Abständen werden einem Fertigungsprozess Stichproben vom Umfang $n = 5$ entnommen, das zu überwachende Merkmal wird vermessen, und die fünf einzelnen Messergebnisse sowie deren Standardabweichung s und Mittelwert \bar{x} werden jeweils in die Qualitätsregelkarte eingetragen.

Betrachten wir zunächst nur die in zeitlicher Reihenfolge gefundenen Stichprobenmittelwerte \bar{x}_i . Auf die Standardabweichungen s_i wird im Zusammenhang mit der s -Karte (Abschnitt 7.2.1) eingegangen.

Die berechneten Mittelwerte werden in einem Diagramm grafisch aufgetragen und durch einen Linienzug verbunden.

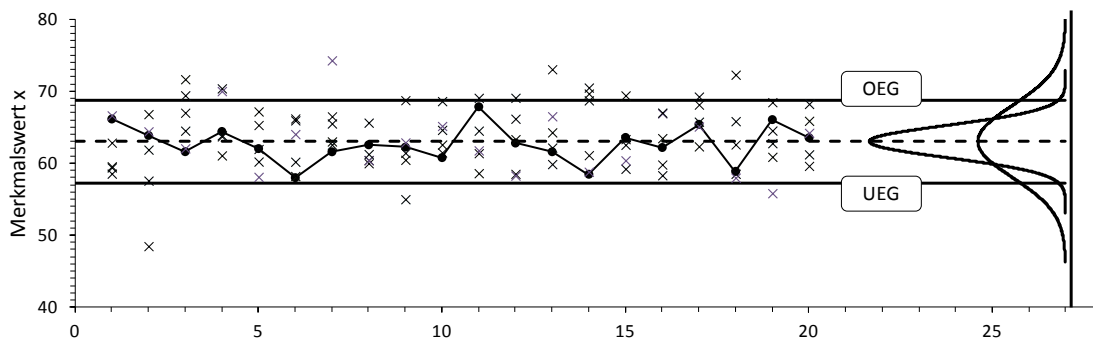


Bild 7.1: Schematische Darstellung zur Funktionsweise einer Mittelwertkarte. Zur Verdeutlichung der Zusammenhänge sind zusätzlich zu den Mittelwerten (Punkte) auch die Einzelwerte (Kreuze) jeder Stichprobe dargestellt.

Die Mittelwerte zeigen wie die Einzelwerte eine Streuung um die Prozessmittellage μ , jedoch ist ihre Streuung um den Faktor $\frac{1}{\sqrt{5}}$ kleiner als die der Einzelwerte. Diesen Zusammenhang verdeutlicht das folgende Bild.



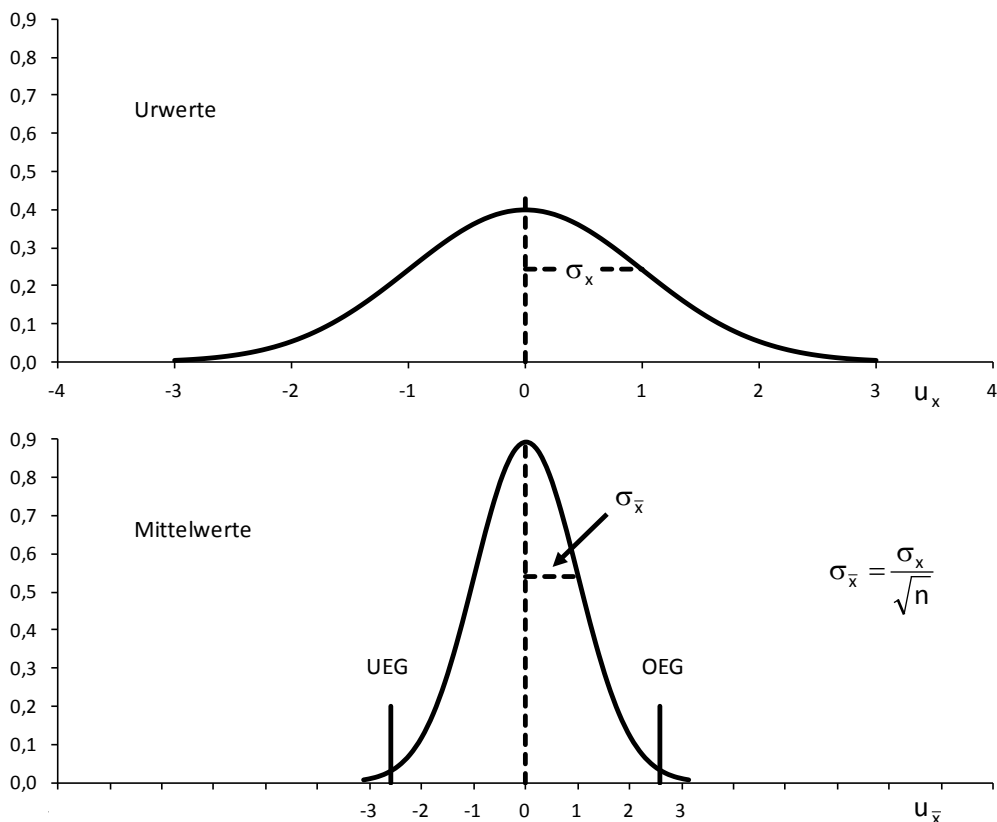


Bild 7.2: Zusammenhang zwischen der Streuung der Einzelwerte (Urwerte) und der Streuung der Mittelwerte.

Allgemein gilt: Streuen die Einzelwerte eines Prozessmerkmals mit der Standardabweichung σ um den Mittelwert μ , so ist die Standardabweichung der Mittelwerte \bar{x} von n Werten gleich $\frac{\sigma}{\sqrt{n}}$.

Wir gehen hier auf die Betrachtung des allgemeinen Falls einer Stichprobe vom Umfang n über.

Den Zufallsstrebereich der Mittelwerte findet man nun sehr leicht, wenn man den Sachverhalt ausnutzt, dass die Transformation $u = \frac{x - \mu}{\sigma}$ eine normalverteilte Variable x in eine standardnormalverteilte Variable u überführt (vgl. Abschnitt 6.1.3) und statt x mit der Standardabweichung σ die Mittelwerte \bar{x} mit der Standardabweichung $\frac{\sigma}{\sqrt{n}}$ einsetzt:

$$u = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Aus der Tabelle der Standardnormalverteilung entnimmt man die Grenzen $u_{\text{unten}} = -2,58$ und $u_{\text{oben}} = +2,58$ für den (zweiseitigen) 99%-Zufallsstrebereich der Größe u . Mit einer Wahrscheinlichkeit von 99 % liegt u also innerhalb der Grenzen $-2,58$ und $+2,58$:

$$-2,58 \leq u \leq +2,58$$

Durch Einsetzen von u ergibt sich daraus der 99%-Zufallsstrebereich des Mittelwerts \bar{x} einer Stichprobe vom Umfang n :



$$-2,58 \leq \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq +2,58 \quad \Leftrightarrow \quad \mu - 2,58 \cdot \frac{\sigma}{\sqrt{n}} \leq \bar{x} \leq \mu + 2,58 \cdot \frac{\sigma}{\sqrt{n}}$$

Ersetzt man noch die unbekannte Größe μ durch den Sollwert C und das ebenfalls unbekannte σ der Grundgesamtheit durch den entsprechenden Schätzwert $\hat{\sigma}$ (er wird im Rahmen einer Prozessfähigkeitsuntersuchung anhand von mindestens 20 Stichproben zu je 5 Teilen ermittelt), so enthält man mit

$$C - 2,58 \cdot \frac{\hat{\sigma}}{\sqrt{n}} \leq \bar{x} \leq C + 2,58 \cdot \frac{\hat{\sigma}}{\sqrt{n}}$$

eine zentrale Beziehung im Zusammenhang mit allen Regelkarten der Lage.

Die Größen

$$\text{OEG} = C + 2,58 \cdot \frac{\hat{\sigma}}{\sqrt{n}} \quad \text{und} \quad \text{UEG} = C - 2,58 \cdot \frac{\hat{\sigma}}{\sqrt{n}}$$

heißen obere und untere Eingriffsgrenze für den Mittelwert \bar{x} . Da sie lediglich von der Prozessstreuung σ abhängig und unabhängig von der Merkmalstoleranz sind, spricht man von natürlichen oder prozessbezogenen Eingriffsgrenzen.

Sie begrenzen den Bereich, in dem 99 % aller Mittelwerte von jeweils n Einzelwerten liegen. Diese Eingriffsgrenzen sind in der Regelkarte (vgl. Bilder 7.1 und 7.2) als horizontale durchgezogene Linien eingezeichnet. Alle bisherigen Überlegungen haben zur Voraussetzung, dass die Prozessmittellage stabil ist. Überschreitet ein beobachteter Mittelwert \bar{x}_i die obere oder unterschreitet er die untere Eingriffsgrenze, so wird daraus geschlossen, dass die Voraussetzung einer stabilen Prozessmittellage nicht mehr gegeben ist, der Prozess sich also signifikant verschoben hat und ein korrigierender Eingriff durch den Maschinenbediener erfolgen muss.

7.1.2 Urwertkarte (x-Karte)

Wie praktische Erfahrungen zeigen, ist es in einigen Fällen wünschenswert, die Prozesslage direkt anhand der Stichprobeneinzelwerte zu beurteilen. Diesem Wunsch trägt die Urwertkarte Rechnung.

Die Herleitung ihrer natürlichen Eingriffsgrenzen erfolgt über eine Betrachtung der Wahrscheinlichkeit, dass alle Einzelwerte einer Stichprobe vom Umfang n innerhalb dieser Grenzen liegen.

Im Abschnitt über die Mittelwertkarte wurde gezeigt, dass unter bestimmten Voraussetzungen (normalverteiltes Merkmal, stabiler Prozess) ein einzelner Merkmalswert mit einer Wahrscheinlichkeit von 99 % im Bereich zwischen

$$C - 2,58 \cdot \sigma \quad \text{und} \quad C + 2,58 \cdot \sigma$$

liegt. Da beispielsweise 5 Stichprobenwerte voneinander unabhängige Zufallsergebnisse sind, ist die Wahrscheinlichkeit, dass alle 5 Werte in diesem Bereich liegen, gleich dem Produkt aus den Einzelwahrscheinlichkeiten, also

$$P_{\text{ges}} = 0,99 \cdot 0,99 \cdot 0,99 \cdot 0,99 \cdot 0,99$$

$$P_{\text{ges}} = 0,99^5 \approx 0,95.$$

Zur Berechnung der natürlichen Eingriffsgrenzen für die Urwertkarte mit $n = 5$ muss man nun lediglich die Gesamtwahrscheinlichkeit $P_{\text{ges}} = 0,99$ setzen und die Gleichung



$$P_{\text{ges}} = (P_{\text{einzel}})^5 = 0,99 \quad \text{umformen in}$$

$$P_{\text{einzel}} = \sqrt[5]{0,99} = 0,998.$$

Sie sind also aus der Tabelle der Standardnormalverteilung als (zweiseitige) 99,8%-Zufallsstregrenzen zu entnehmen. Für das Beispiel $n = 5$ ist dies der Wert $u = 3,09$, und die Eingriffsgrenzen sind demnach

$$\text{OEG} = C + 3,09 \cdot \hat{\sigma} \quad (n = 5)$$

$$\text{UEG} = C - 3,09 \cdot \hat{\sigma}.$$

Verallgemeinert auf Stichproben vom Umfang n erfolgt die Berechnung der natürlichen Eingriffsgrenzen für die Urwertkarte nach

$$\text{OEG} = C + u_{\sqrt{n}, 0,99} \cdot \hat{\sigma}$$

$$\text{UEG} = C - u_{\sqrt{n}, 0,99} \cdot \hat{\sigma}.$$

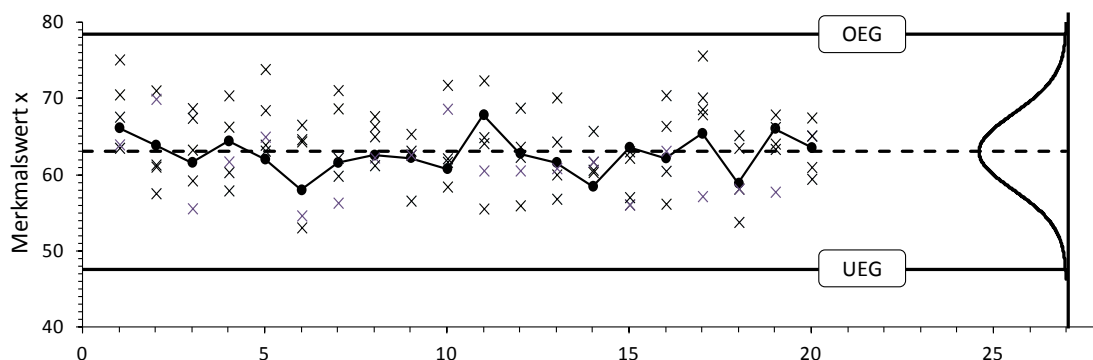


Bild 7.3: Schematische Darstellung zur Funktionsweise einer Urwertkarte. Zur Verdeutlichung sind jeweils der größte und der kleinste Einzelwert einer Stichprobe durch eine Linie verbunden (gleiche Daten wie in Bild 7.1).

7.2 Regelkarten der Streuung

Ebenso wie die Prozessmittellage ist die Prozessstreuung, die i.a. mit der Standardabweichung eines Teilemerkmals gleichgesetzt wird, eine zentrale Kenngröße zur Beurteilung der Fertigungsqualität. Dabei ist das Erkennen von Streuungszunahmen genauso von Interesse, wie das von Streuungsverringerungen. Der letztere Fall ist als Chance zu sehen, die Ursache für die kurzfristig aufgetretene Verbesserung herauszufinden und den günstigen Prozesszustand dauerhaft beizubehalten.

Qualitätsregelkarten für die Streuung sind geeignete Hilfsmittel, um solche Änderungen festzustellen. Zu diesem Zweck werden die in Zusammenhang mit der Mittelwertkarte betrachteten Stichproben vom Umfang n genutzt. Neben der Information über die aktuelle Prozesslage enthält jede Gruppe von Stichprobeneinzelwerten auch eine Information über die momentane Prozessstreuung.

Als Kenngröße dieser Information kann man die Standardabweichung s oder die Spannweite R jeder Stichprobe betrachten.



7.2.1 s-Karte

Wie bereits angedeutet bietet es sich an, die s-Karte nicht als separate Karte zu führen, sondern als zweites Diagramm auf dem Formblatt für die Mittelwertkarte darzustellen. Man nennt diese Kombination \bar{x} -s-Karte.

Parallel zur Darstellung des Mittelwerts jeder i-ten 5-er Stichprobe (bzw. allgemein Stichprobe vom Umfang n) wird also hier deren Standardabweichung s_i aufgetragen.

Die Größe des aktuellen Werts dieser Standardabweichung s_i hängt natürlich von der durch $\hat{\sigma}$ geschätzten mittleren Prozessstreuung ab und kann im speziellen Fall zufällig etwas größer oder kleiner als ein langfristiger mittlerer Wert s sein. Um entscheiden zu können, ab wann solche Schwankungen von s nicht mehr als Zufallsergebnis sondern als Hinweis auf eine tatsächlich vorliegende Veränderung der Prozessstreuung zu deuten sind, braucht man Grenzwerte für s, die beispielsweise den 99%-Zufallsstrebereich von s umfassen. Ist der aktuelle Wert von s größer als der obere bzw. kleiner als der untere Grenzwert, so wird dies als Hinweis auf eine signifikante Veränderung der Prozessstreuung interpretiert. Solche Grenzwerte lassen sich bestimmen, wenn man den Sachverhalt berücksichtigt, dass die Größe $f \cdot \frac{s^2}{\sigma^2}$ einer χ^2 -Verteilung (gesprochen: chi-Quadrat) unterliegt.

Aus $f \cdot \frac{s^2}{\sigma^2} = \chi^2$ ergeben sich mit

$$s_{ob} = \sqrt{\frac{\chi_{f, 1-\alpha/2}^2}{f}} \cdot \sigma \quad \text{und} \quad s_{un} = \sqrt{\frac{\chi_{f, \alpha/2}^2}{f}} \cdot \sigma$$

die gesuchten Eingriffsgrenzen für s.

Die χ^2 -Verteilung zum Freiheitsgrad f (f entspricht der um 1 verminderten Stichprobengröße, also n – 1) ist wie die Standardnormalverteilung tabelliert. Allerdings ist die χ^2 -Verteilung nicht symmetrisch.

Deshalb ergeben sich unterschiedliche Faktoren zur Berechnung der oberen und unteren Eingriffsgrenzen. Die Wurzelterme in obigen Ausdrücken werden der Einfachheit halber mit B'_{Eob} und B'_{Eun} bezeichnet und direkt mit der zugehörigen Stichprobengröße n tabelliert angegeben. Die Gleichungen für die Eingriffsgrenzen der s-Karte lauten somit:

$$OEG = B'_{Eob} \cdot \hat{\sigma}$$

$$UEG = B'_{Eun} \cdot \hat{\sigma}$$

n	2	3	4	5	6
B'_{Eob}	2,807	2,302	2,069	1,927	1,830
B'_{Eun}	0,006	0,071	0,155	0,227	0,287

Tabelle 7.1: Faktoren zur Berechnung der Eingriffsgrenzen der s-Karte



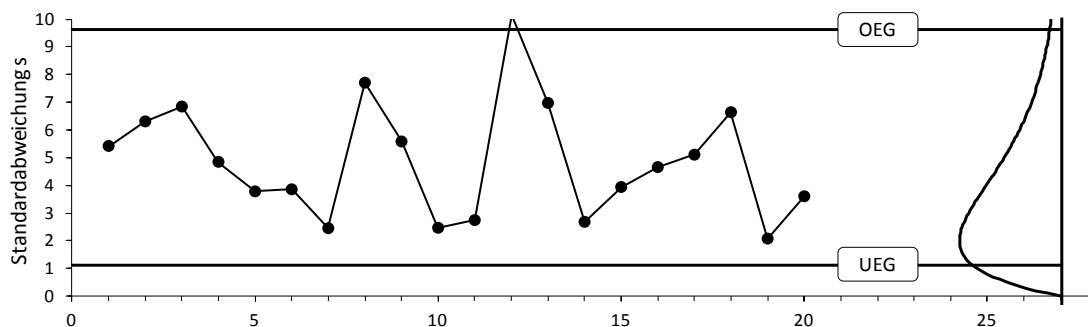


Bild 7.4: Schematische Darstellung einer s-Karte (gleiche Daten wie in den Bildern 7.1 und 7.3). Die Standardabweichungen unterliegen einer schiefen Verteilung. Da die Berechnung der Eingriffsgrenzen für s über diejenigen für s^2 erfolgt, werden die Flächenanteile oberhalb OEG bzw. unterhalb UEG in der Darstellung zu groß wiedergegeben.

7.2.2 R-Karte

Im Falle der R-Karte benutzt man die Spannweite R , also die Differenz des größten und kleinsten Einzelwerts einer 5-er-Gruppe (bzw. einer Stichprobe vom Umfang n) als Maß für die momentane Prozessstreuung.

Die Grundlage zur Berechnung der Streugrenzen von R bildet die Verteilung der standardisierten Spannweite w_n .

Diese Verteilung lässt sich beispielsweise mit Hilfe eines Rechners simulieren, indem wiederholt aus einer Grundgesamtheit standardnormalverteilter Werte jeweils eine Zufallsstichprobe vom Umfang n „gezogen“ und deren Spannweite R bestimmt wird. Der 99%-Zufallsstrebereich dieser Größe R ist dann gegeben durch die obere Grenze $w_{n;0,995}$ und die untere Grenze $w_{n;0,005}$.

Da bei Verwendung der Spannweitenkarte die Prozessstandardabweichung σ zweckmäßigerweise über die mittlere Spannweite \bar{R} geschätzt wird (vgl. Abschnitt 3.9)

$$\hat{\sigma} = \frac{\bar{R}}{d_n},$$

ist es vorteilhaft, die Faktoren $w_{n;0,005}$ bzw. $w_{n;0,995}$ und $\frac{1}{d_n}$ zu den Faktoren D_{Eun} bzw. D_{Eob} zusammenzufassen. Diese Größen sind tabelliert und erlauben eine Berechnung der Eingriffsgrenzen für die Spannweitenkarte nach

$$OEG = D_{Eob} \cdot \bar{R}$$

$$UEG = D_{Eun} \cdot \bar{R}.$$

n	2	3	4	5	6
D_{Eob}	3,518	2,614	2,280	2,100	1,986
D_{Eun}	0,008	0,080	0,166	0,239	0,296

Tabelle 7.2: Faktoren zur Berechnung der Eingriffsgrenzen der R-Karte



8 Beurteilung von Häufigkeitsverteilungen im Zusammenhang mit einer Toleranz

In nachfolgenden Abbildungen werden schematisch einige Beispiele möglicher Häufigkeitsverteilungen dargestellt und stichwortartig erläutert, die bei der Untersuchung eines Fertigungsprozesses auftreten können.

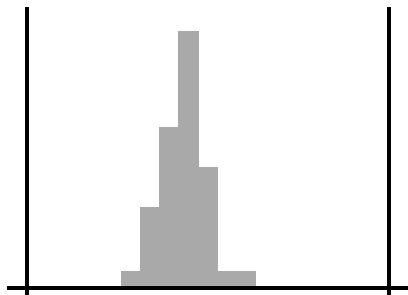


Bild 8.1:

Streubereich wesentlich kleiner als Toleranz. Mittelwert stimmt gut mit der Toleranzmitte überein.

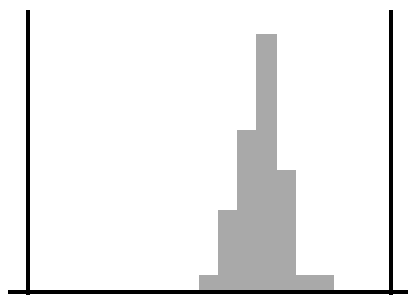


Bild 8.2:

Streubereich wesentlich kleiner als Toleranz. Mittelwert liegt jedoch außerhalb Toleranzmitte. Mit der Entstehung von Ausschuss ist zu rechnen.

Prozess zentrieren!

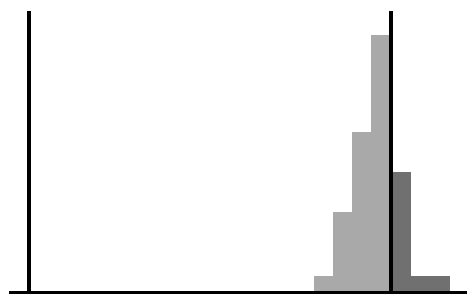


Bild 8.3:

Streubereich wesentlich kleiner als Toleranz. Mittelwert der Verteilung jedoch weit von der Toleranzmitte entfernt.

Prozess zentrieren!



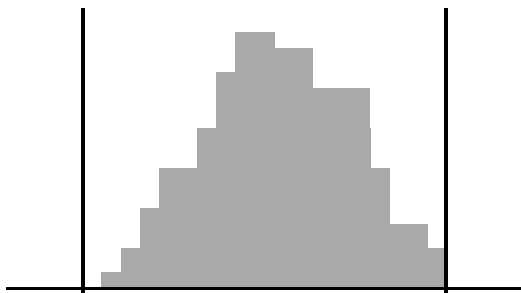


Bild 8.4:

Streubreite ungefähr gleich der Toleranz.

Mittelwert der Verteilung stimmt gut mit Toleranzmitte überein. Eine systematische Mittelwertsverschiebung kann jedoch Ausschuss zur Folge haben.

Streuung reduzieren!

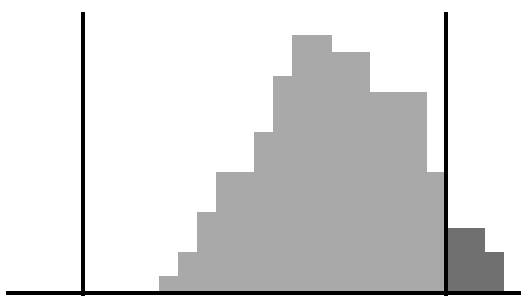


Bild 8.5:

Streubreite ungefähr gleich der Toleranz.

Jedoch fällt Mittelwert der Verteilung nicht mit Toleranzmitte zusammen. Ausschuss an der oberen Toleranzgrenze.

Prozess zentrieren, Streuung reduzieren!

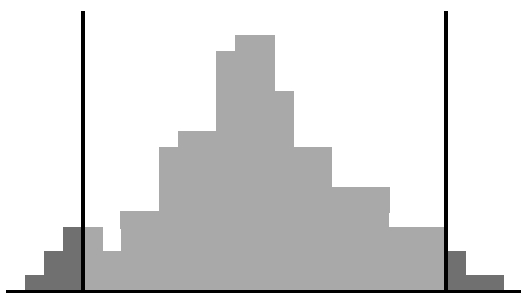


Bild 8.6:

Mittelwert der Verteilung stimmt gut mit der Toleranzmitte überein. Streubreite ist zu groß. Überschreitung an beiden Toleranzgrenzen.

Streuung reduzieren!

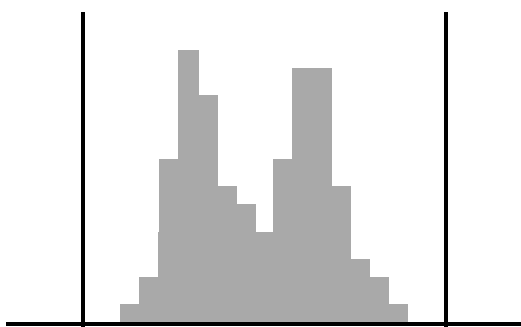


Bild 8.7:

Überlagerung zweier Verteilungen.

Möglicherweise durch systematische Prozessveränderung (z.B. Werkzeug, Material) verursacht.

Nach Beseitigung der Ursache kann die Toleranz leicht eingehalten werden, weil die Streubreite beider Verteilungen vergleichsweise klein ist.



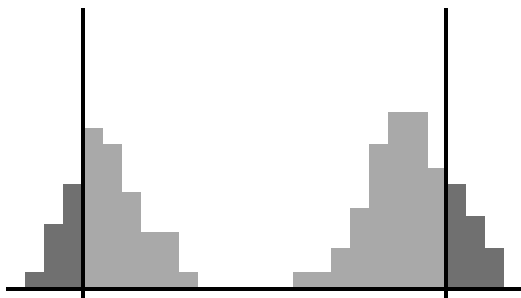


Bild 8.8:

Ähnliche Situation wie in Bild 8.7, jedoch liegen die Mittelwerte der beiden Verteilungen so weit auseinander, dass an beiden Toleranzgrenzen Ausschuss entsteht.

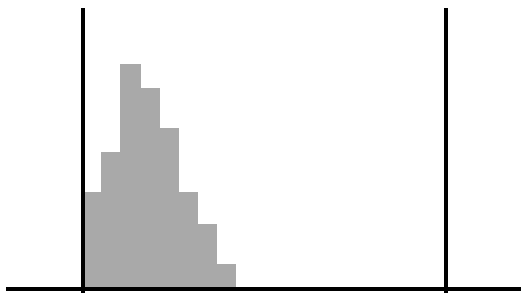


Bild 8.9:

Mittelwert der Verteilung ist zur unteren Toleranzgrenze hin verschoben. Das Los wurde offenbar 100 % verlesen. Wenn der Prozess zentriert werden kann, wird das Verlesen überflüssig.

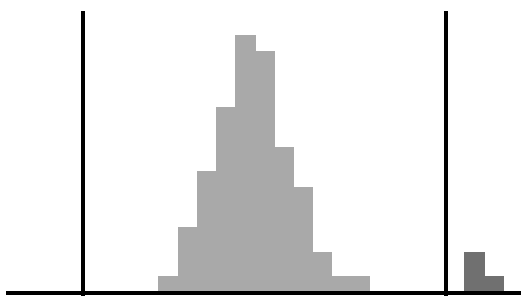


Bild 8.10:

Die Hauptverteilung hat eine geringe Streubreite. Der Mittelwert stimmt gut mit der Toleranzmitte überein. Ein kleinerer Anteil liegt jenseits der oberen Toleranzgrenze. Es könnte sich um Ausschuss handeln, der während der Einrichtung der Maschine entstanden und nicht aussortiert worden ist.



9 Genauigkeit der Schätzung von Mittelwert und Standardabweichung

Von Stichproben abgeleitete Aussagen über eine Grundgesamtheit sind stets mit einer statistischen Unsicherheit behaftet. Diese Unsicherheit ist i.a. umso größer, je kleiner die zugrundeliegende Datenbasis, d.h. der Umfang n der Stichprobe ist.

Wir setzen im Folgenden voraus, dass es sich bei der Grundgesamtheit um eine Normalverteilung handelt, deren Mittelwert μ und Standardabweichung σ nicht bekannt sind.

Die aus einer Stichprobe berechneten (empirischen) Größen \bar{x} und s sind Schätzungen der unbekanntenen Größen μ bzw. σ . Dies wird durch die Schreibweise

$$\hat{\mu} = \bar{x} \quad (\bar{x} \text{ ist eine Schätzung für } \mu)$$

$$\hat{\sigma} = s \quad (s \text{ ist eine Schätzung für } \sigma)$$

zum Ausdruck gebracht.

Man kann einen Bereich um \bar{x} bzw. um s angeben, in denen die unbekanntenen Größen μ und σ jeweils mit großer Wahrscheinlichkeit liegen. Die Breite dieser sogenannten Vertrauensbereiche ist einerseits vom Stichprobenumfang n , andererseits von einer vorgegebenen Aussagewahrscheinlichkeit P_A abhängig. Die Größe $1 - P_A$ ist die zugehörige Irrtumswahrscheinlichkeit.

In den Bildern 9.1 und 9.2 sind die Vertrauensbereiche von μ und σ in Abhängigkeit von der Stichprobengröße dargestellt. Die Kurven gelten für eine Aussagewahrscheinlichkeit von 95 %. Das bedeutet, dass im Durchschnitt in 95 von 100 Fällen, in denen μ durch \bar{x} bzw. σ durch s geschätzt wird, μ bzw. σ innerhalb der den Kurven entnommenen Vertrauensgrenzen liegt.

a) Vertrauensbereich für σ :

$$\frac{s_R}{D_{ob}} \leq \sigma \leq \frac{s_R}{D_{un}}$$

s_R ist die mit der Range-Methode (vgl. Abschnitt 3.9) bestimmte Standardabweichung der Stichprobe. Die Faktoren $\frac{1}{D_{un}}$ und $\frac{1}{D_{ob}}$ findet man, indem man auf der Abszisse (diese ist logarithmisch geteilt) die Stichprobengröße n aufsucht, senkrecht nach oben geht bis auf die beiden eingezeichneten Kurven und von hier waagrecht nach links auf die Ordinate. Dort liest man die entsprechenden Werte für $\frac{1}{D_{un}}$ bzw. $\frac{1}{D_{ob}}$ ab (Bild 9.1).

b) Vertrauensbereich für μ :

$$\bar{x} - \frac{t'}{\sqrt{n}} \cdot s_R \leq \mu \leq \bar{x} + \frac{t'}{\sqrt{n}} \cdot s_R$$

Den Faktor $\frac{t'}{\sqrt{n}}$ findet man analog zu der in a) beschriebenen Vorgehensweise (Bild 9.2).

Da der Vertrauensbereich für μ symmetrisch ist, genügt es, $\frac{t'}{\sqrt{n}}$ an der oberen Kurve abzulesen.



Je größer die Stichprobe ist, umso mehr nähern sich die Faktoren $\frac{1}{D_{un}}$, $\frac{1}{D_{ob}}$ dem Wert 1 bzw. $\frac{t'}{\sqrt{n}}$ dem Wert 0.

Die Kurvenverläufe lassen aber auch erkennen, dass bei Stichprobenumfängen oberhalb etwa $n = 50$ (trotz steigenden Versuchsaufwands) die Vertrauensbereiche nur noch unwesentlich kleiner werden. Dagegen sind Aussagen über Mittelwert und Standardabweichung bei Stichprobenumfängen $n < 25$ mit Vorsicht zu interpretieren, da die zugehörigen Vertrauensbereiche sehr groß werden.

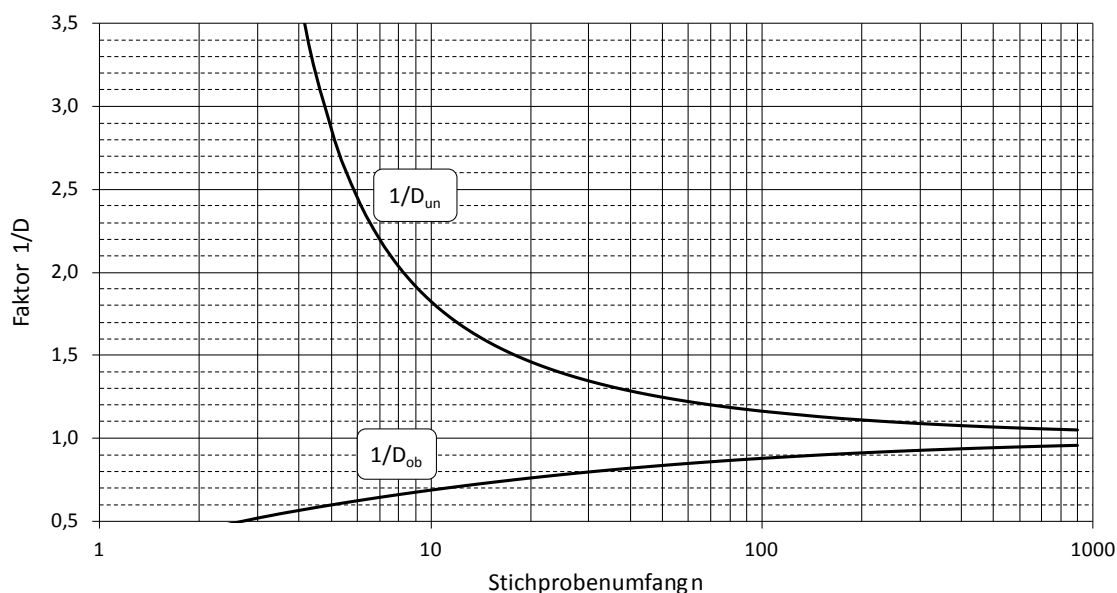


Bild 9.1: Diagramm zur Bestimmung des Vertrauensbereichs für σ ($P_A = 95\%$)

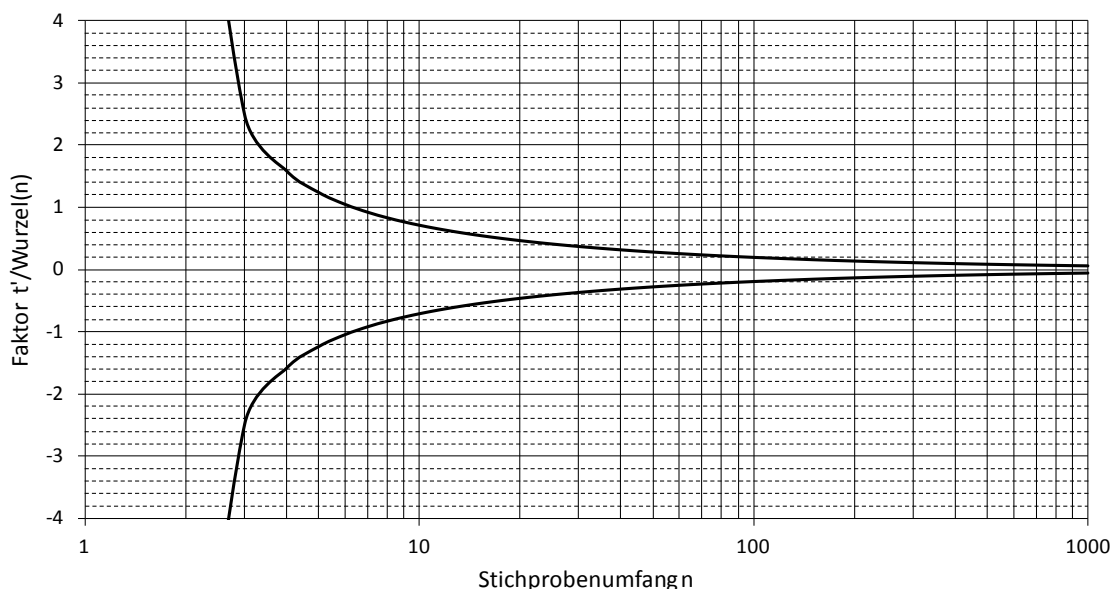


Bild 9.2: Diagramm zur Bestimmung des Vertrauensbereichs für μ ($P_A = 95\%$)

2020-04-06 - SOCOS



10 Standardnormalverteilung $\Phi(-u) = 1 - \Phi(u)$ $D(u) = \Phi(u) - \Phi(-u)$

u	$\Phi(-u)$	$\Phi(u)$	D(u)
0,01	0,496011	0,503989	0,007979
0,02	0,492022	0,507978	0,015957
0,03	0,488033	0,511967	0,023933
0,04	0,484047	0,515953	0,031907
0,05	0,480061	0,519939	0,039878
0,06	0,476078	0,523922	0,047845
0,07	0,472097	0,527903	0,055806
0,08	0,468119	0,531881	0,063763
0,09	0,464144	0,535856	0,071713
0,10	0,460172	0,539828	0,079656
0,11	0,456205	0,543795	0,087591
0,12	0,452242	0,547758	0,095517
0,13	0,448283	0,551717	0,103434
0,14	0,444330	0,555670	0,111340
0,15	0,440382	0,559618	0,119235
0,16	0,436441	0,563559	0,127119
0,17	0,432505	0,567495	0,134990
0,18	0,428576	0,571424	0,142847
0,19	0,424655	0,575345	0,150691
0,20	0,420740	0,579260	0,158519
0,21	0,416834	0,583166	0,166332
0,22	0,412936	0,587064	0,174129
0,23	0,409046	0,590954	0,181908
0,24	0,405165	0,594835	0,189670
0,25	0,401294	0,598706	0,197413
0,26	0,397432	0,602568	0,205136
0,27	0,393580	0,606420	0,212840
0,28	0,389739	0,610261	0,220522
0,29	0,385908	0,614092	0,228184
0,30	0,382089	0,617911	0,235823
0,31	0,378281	0,621719	0,243439
0,32	0,374484	0,625516	0,251032
0,33	0,370700	0,629300	0,258600
0,34	0,366928	0,633072	0,266143
0,35	0,363169	0,636831	0,273661
0,36	0,359424	0,640576	0,281153
0,37	0,355691	0,644309	0,288617
0,38	0,351973	0,648027	0,296054
0,39	0,348268	0,651732	0,303463
0,40	0,344578	0,655422	0,310843
0,41	0,340903	0,659097	0,318194
0,42	0,337243	0,662757	0,325514
0,43	0,333598	0,666402	0,332804
0,44	0,329969	0,670031	0,340063
0,45	0,326355	0,673645	0,347290
0,46	0,322758	0,677242	0,354484
0,47	0,319178	0,680822	0,361645
0,48	0,315614	0,684386	0,368773
0,49	0,312067	0,687933	0,375866
0,50	0,308538	0,691462	0,382925

u	$\Phi(-u)$	$\Phi(u)$	D(u)
0,51	0,305026	0,694974	0,389949
0,52	0,301532	0,698468	0,396936
0,53	0,298056	0,701944	0,403888
0,54	0,294598	0,705402	0,410803
0,55	0,291160	0,708840	0,417681
0,56	0,287740	0,712260	0,424521
0,57	0,284339	0,715661	0,431322
0,58	0,280957	0,719043	0,438085
0,59	0,277595	0,722405	0,444809
0,60	0,274253	0,725747	0,451494
0,61	0,270931	0,729069	0,458138
0,62	0,267629	0,732371	0,464742
0,63	0,264347	0,735653	0,471306
0,64	0,261086	0,738914	0,477828
0,65	0,257846	0,742154	0,484308
0,66	0,254627	0,745373	0,490746
0,67	0,251429	0,748571	0,497142
0,68	0,248252	0,751748	0,503496
0,69	0,245097	0,754903	0,509806
0,70	0,241964	0,758036	0,516073
0,71	0,238852	0,761148	0,522296
0,72	0,235762	0,764238	0,528475
0,73	0,232695	0,767305	0,534610
0,74	0,229650	0,770350	0,540700
0,75	0,226627	0,773373	0,546745
0,76	0,223627	0,776373	0,552746
0,77	0,220650	0,779350	0,558700
0,78	0,217695	0,782305	0,564609
0,79	0,214764	0,785236	0,570472
0,80	0,211855	0,788145	0,576289
0,81	0,208970	0,791030	0,582060
0,82	0,206108	0,793892	0,587784
0,83	0,203269	0,796731	0,593461
0,84	0,200454	0,799546	0,599092
0,85	0,197662	0,802338	0,604675
0,86	0,194894	0,805106	0,610211
0,87	0,192150	0,807850	0,615700
0,88	0,189430	0,810570	0,621141
0,89	0,186733	0,813267	0,626534
0,90	0,184060	0,815940	0,631880
0,91	0,181411	0,818589	0,637178
0,92	0,178786	0,821214	0,642427
0,93	0,176186	0,823814	0,647629
0,94	0,173609	0,826391	0,652782
0,95	0,171056	0,828944	0,657888
0,96	0,168528	0,831472	0,662945
0,97	0,166023	0,833977	0,667954
0,98	0,163543	0,836457	0,672914
0,99	0,161087	0,838913	0,677826
1,00	0,158655	0,841345	0,682689

2020-04-06 - SOCOS



u	$\Phi(-u)$	$\Phi(u)$	D(u)
1,01	0,156248	0,843752	0,687505
1,02	0,153864	0,846136	0,692272
1,03	0,151505	0,848495	0,696990
1,04	0,149170	0,850830	0,701660
1,05	0,146859	0,853141	0,706282
1,06	0,144572	0,855428	0,710855
1,07	0,142310	0,857690	0,715381
1,08	0,140071	0,859929	0,719858
1,09	0,137857	0,862143	0,724287
1,10	0,135666	0,864334	0,728668
1,11	0,133500	0,866500	0,733001
1,12	0,131357	0,868643	0,737286
1,13	0,129238	0,870762	0,741524
1,14	0,127143	0,872857	0,745714
1,15	0,125072	0,874928	0,749856
1,16	0,123024	0,876976	0,753951
1,17	0,121001	0,878999	0,757999
1,18	0,119000	0,881000	0,762000
1,19	0,117023	0,882977	0,765953
1,20	0,115070	0,884930	0,769861
1,21	0,113140	0,886860	0,773721
1,22	0,111233	0,888767	0,777535
1,23	0,109349	0,890651	0,781303
1,24	0,107488	0,892512	0,785024
1,25	0,105650	0,894350	0,788700
1,26	0,103835	0,896165	0,792331
1,27	0,102042	0,897958	0,795915
1,28	0,100273	0,899727	0,799455
1,29	0,098525	0,901475	0,802949
1,30	0,096801	0,903199	0,806399
1,31	0,095098	0,904902	0,809804
1,32	0,093418	0,906582	0,813165
1,33	0,091759	0,908241	0,816482
1,34	0,090123	0,909877	0,819755
1,35	0,088508	0,911492	0,822984
1,36	0,086915	0,913085	0,826170
1,37	0,085344	0,914656	0,829313
1,38	0,083793	0,916207	0,832413
1,39	0,082264	0,917736	0,835471
1,40	0,080757	0,919243	0,838487
1,41	0,079270	0,920730	0,841460
1,42	0,077804	0,922196	0,844392
1,43	0,076359	0,923641	0,847283
1,44	0,074934	0,925066	0,850133
1,45	0,073529	0,926471	0,852941
1,46	0,072145	0,927855	0,855710
1,47	0,070781	0,929219	0,858438
1,48	0,069437	0,930563	0,861127
1,49	0,068112	0,931888	0,863776
1,50	0,066807	0,933193	0,866386

u	$\Phi(-u)$	$\Phi(u)$	D(u)
1,51	0,065522	0,934478	0,868957
1,52	0,064256	0,935744	0,871489
1,53	0,063008	0,936992	0,873983
1,54	0,061780	0,938220	0,876440
1,55	0,060571	0,939429	0,878858
1,56	0,059380	0,940620	0,881240
1,57	0,058208	0,941792	0,883585
1,58	0,057053	0,942947	0,885893
1,59	0,055917	0,944083	0,888165
1,60	0,054799	0,945201	0,890401
1,61	0,053699	0,946301	0,892602
1,62	0,052616	0,947384	0,894768
1,63	0,051551	0,948449	0,896899
1,64	0,050503	0,949497	0,898995
1,65	0,049471	0,950529	0,901057
1,66	0,048457	0,951543	0,903086
1,67	0,047460	0,952540	0,905081
1,68	0,046479	0,953521	0,907043
1,69	0,045514	0,954486	0,908972
1,70	0,044565	0,955435	0,910869
1,71	0,043633	0,956367	0,912734
1,72	0,042716	0,957284	0,914568
1,73	0,041815	0,958185	0,916370
1,74	0,040929	0,959071	0,918141
1,75	0,040059	0,959941	0,919882
1,76	0,039204	0,960796	0,921592
1,77	0,038364	0,961636	0,923273
1,78	0,037538	0,962462	0,924924
1,79	0,036727	0,963273	0,926546
1,80	0,035930	0,964070	0,928139
1,81	0,035148	0,964852	0,929704
1,82	0,034379	0,965621	0,931241
1,83	0,033625	0,966375	0,932750
1,84	0,032884	0,967116	0,934232
1,85	0,032157	0,967843	0,935687
1,86	0,031443	0,968557	0,937115
1,87	0,030742	0,969258	0,938516
1,88	0,030054	0,969946	0,939892
1,89	0,029379	0,970621	0,941242
1,90	0,028716	0,971284	0,942567
1,91	0,028067	0,971933	0,943867
1,92	0,027429	0,972571	0,945142
1,93	0,026803	0,973197	0,946393
1,94	0,026190	0,973810	0,947620
1,95	0,025588	0,974412	0,948824
1,96	0,024998	0,975002	0,950004
1,97	0,024419	0,975581	0,951162
1,98	0,023852	0,976148	0,952297
1,99	0,023295	0,976705	0,953409
2,00	0,022750	0,977250	0,954500



u	$\Phi(-u)$	$\Phi(u)$	D(u)
2,01	0,022216	0,977784	0,955569
2,02	0,021692	0,978308	0,956617
2,03	0,021178	0,978822	0,957644
2,04	0,020675	0,979325	0,958650
2,05	0,020182	0,979818	0,959636
2,06	0,019699	0,980301	0,960602
2,07	0,019226	0,980774	0,961548
2,08	0,018763	0,981237	0,962475
2,09	0,018309	0,981691	0,963382
2,10	0,017864	0,982136	0,964271
2,11	0,017429	0,982571	0,965142
2,12	0,017003	0,982997	0,965994
2,13	0,016586	0,983414	0,966829
2,14	0,016177	0,983823	0,967645
2,15	0,015778	0,984222	0,968445
2,16	0,015386	0,984614	0,969227
2,17	0,015003	0,984997	0,969993
2,18	0,014629	0,985371	0,970743
2,19	0,014262	0,985738	0,971476
2,20	0,013903	0,986097	0,972193
2,21	0,013553	0,986447	0,972895
2,22	0,013209	0,986791	0,973581
2,23	0,012874	0,987126	0,974253
2,24	0,012545	0,987455	0,974909
2,25	0,012224	0,987776	0,975551
2,26	0,011911	0,988089	0,976179
2,27	0,011604	0,988396	0,976792
2,28	0,011304	0,988696	0,977392
2,29	0,011011	0,988989	0,977979
2,30	0,010724	0,989276	0,978552
2,31	0,010444	0,989556	0,979112
2,32	0,010170	0,989830	0,979659
2,33	0,009903	0,990097	0,980194
2,34	0,009642	0,990358	0,980716
2,35	0,009387	0,990613	0,981227
2,36	0,009137	0,990863	0,981725
2,37	0,008894	0,991106	0,982212
2,38	0,008656	0,991344	0,982687
2,39	0,008424	0,991576	0,983152
2,40	0,008198	0,991802	0,983605
2,41	0,007976	0,992024	0,984047
2,42	0,007760	0,992240	0,984479
2,43	0,007549	0,992451	0,984901
2,44	0,007344	0,992656	0,985313
2,45	0,007143	0,992857	0,985714
2,46	0,006947	0,993053	0,986106
2,47	0,006756	0,993244	0,986489
2,48	0,006569	0,993431	0,986862
2,49	0,006387	0,993613	0,987226
2,50	0,006210	0,993790	0,987581

u	$\Phi(-u)$	$\Phi(u)$	D(u)
2,51	0,006037	0,993963	0,987927
2,52	0,005868	0,994132	0,988264
2,53	0,005703	0,994297	0,988594
2,54	0,005543	0,994457	0,988915
2,55	0,005386	0,994614	0,989228
2,56	0,005234	0,994766	0,989533
2,57	0,005085	0,994915	0,989830
2,58	0,004940	0,995060	0,990120
2,59	0,004799	0,995201	0,990402
2,60	0,004661	0,995339	0,990678
2,61	0,004527	0,995473	0,990946
2,62	0,004397	0,995603	0,991207
2,63	0,004269	0,995731	0,991461
2,64	0,004145	0,995855	0,991709
2,65	0,004025	0,995975	0,991951
2,66	0,003907	0,996093	0,992186
2,67	0,003793	0,996207	0,992415
2,68	0,003681	0,996319	0,992638
2,69	0,003573	0,996427	0,992855
2,70	0,003467	0,996533	0,993066
2,71	0,003364	0,996636	0,993272
2,72	0,003264	0,996736	0,993472
2,73	0,003167	0,996833	0,993666
2,74	0,003072	0,996928	0,993856
2,75	0,002980	0,997020	0,994040
2,76	0,002890	0,997110	0,994220
2,77	0,002803	0,997197	0,994394
2,78	0,002718	0,997282	0,994564
2,79	0,002635	0,997365	0,994729
2,80	0,002555	0,997445	0,994890
2,81	0,002477	0,997523	0,995046
2,82	0,002401	0,997599	0,995198
2,83	0,002327	0,997673	0,995345
2,84	0,002256	0,997744	0,995489
2,85	0,002186	0,997814	0,995628
2,86	0,002118	0,997882	0,995763
2,87	0,002052	0,997948	0,995895
2,88	0,001988	0,998012	0,996023
2,89	0,001926	0,998074	0,996147
2,90	0,001866	0,998134	0,996268
2,91	0,001807	0,998193	0,996386
2,92	0,001750	0,998250	0,996500
2,93	0,001695	0,998305	0,996610
2,94	0,001641	0,998359	0,996718
2,95	0,001589	0,998411	0,996822
2,96	0,001538	0,998462	0,996923
2,97	0,001489	0,998511	0,997022
2,98	0,001441	0,998559	0,997117
2,99	0,001395	0,998605	0,997210
3,00	0,001350	0,998650	0,997300

2020-04-06 - SOCCOS



u	$\Phi(-u)$	$\Phi(u)$	D(u)
3,01	0,001306	0,998694	0,997387
3,02	0,001264	0,998736	0,997472
3,03	0,001223	0,998777	0,997554
3,04	0,001183	0,998817	0,997634
3,05	0,001144	0,998856	0,997711
3,06	0,001107	0,998893	0,997786
3,07	0,001070	0,998930	0,997859
3,08	0,001035	0,998965	0,997930
3,09	0,001001	0,998999	0,997998
3,10	0,000968	0,999032	0,998065
3,11	0,000936	0,999064	0,998129
3,12	0,000904	0,999096	0,998191
3,13	0,000874	0,999126	0,998252
3,14	0,000845	0,999155	0,998310
3,15	0,000816	0,999184	0,998367
3,16	0,000789	0,999211	0,998422
3,17	0,000762	0,999238	0,998475
3,18	0,000736	0,999264	0,998527
3,19	0,000711	0,999289	0,998577
3,20	0,000687	0,999313	0,998626
3,21	0,000664	0,999336	0,998673
3,22	0,000641	0,999359	0,998718
3,23	0,000619	0,999381	0,998762
3,24	0,000598	0,999402	0,998805
3,25	0,000577	0,999423	0,998846
3,26	0,000557	0,999443	0,998886
3,27	0,000538	0,999462	0,998924
3,28	0,000519	0,999481	0,998962
3,29	0,000501	0,999499	0,998998
3,30	0,000483	0,999517	0,999033
3,31	0,000467	0,999533	0,999067
3,32	0,000450	0,999550	0,999100
3,33	0,000434	0,999566	0,999131
3,34	0,000419	0,999581	0,999162
3,35	0,000404	0,999596	0,999192
3,36	0,000390	0,999610	0,999220
3,37	0,000376	0,999624	0,999248
3,38	0,000362	0,999638	0,999275
3,39	0,000350	0,999650	0,999301
3,40	0,000337	0,999663	0,999326
3,41	0,000325	0,999675	0,999350
3,42	0,000313	0,999687	0,999374
3,43	0,000302	0,999698	0,999396
3,44	0,000291	0,999709	0,999418
3,45	0,000280	0,999720	0,999439
3,46	0,000270	0,999730	0,999460
3,47	0,000260	0,999740	0,999479
3,48	0,000251	0,999749	0,999498
3,49	0,000242	0,999758	0,999517
3,50	0,000233	0,999767	0,999535

u	$\Phi(-u)$	$\Phi(u)$	D(u)
3,51	0,000224	0,999776	0,999552
3,52	0,000216	0,999784	0,999568
3,53	0,000208	0,999792	0,999584
3,54	0,000200	0,999800	0,999600
3,55	0,000193	0,999807	0,999615
3,56	0,000185	0,999815	0,999629
3,57	0,000179	0,999821	0,999643
3,58	0,000172	0,999828	0,999656
3,59	0,000165	0,999835	0,999669
3,60	0,000159	0,999841	0,999682
3,61	0,000153	0,999847	0,999694
3,62	0,000147	0,999853	0,999705
3,63	0,000142	0,999858	0,999717
3,64	0,000136	0,999864	0,999727
3,65	0,000131	0,999869	0,999738
3,66	0,000126	0,999874	0,999748
3,67	0,000121	0,999879	0,999757
3,68	0,000117	0,999883	0,999767
3,69	0,000112	0,999888	0,999776
3,70	0,000108	0,999892	0,999784
3,71	0,000104	0,999896	0,999793
3,72	0,000100	0,999900	0,999801
3,73	0,000096	0,999904	0,999808
3,74	0,000092	0,999908	0,999816
3,75	0,000088	0,999912	0,999823
3,76	0,000085	0,999915	0,999830
3,77	0,000082	0,999918	0,999837
3,78	0,000078	0,999922	0,999843
3,79	0,000075	0,999925	0,999849
3,80	0,000072	0,999928	0,999855
3,81	0,000070	0,999930	0,999861
3,82	0,000067	0,999933	0,999867
3,83	0,000064	0,999936	0,999872
3,84	0,000062	0,999938	0,999877
3,85	0,000059	0,999941	0,999882
3,86	0,000057	0,999943	0,999887
3,87	0,000054	0,999946	0,999891
3,88	0,000052	0,999948	0,999896
3,89	0,000050	0,999950	0,999900
3,90	0,000048	0,999952	0,999904
3,91	0,000046	0,999954	0,999908
3,92	0,000044	0,999956	0,999911
3,93	0,000042	0,999958	0,999915
3,94	0,000041	0,999959	0,999918
3,95	0,000039	0,999961	0,999922
3,96	0,000037	0,999963	0,999925
3,97	0,000036	0,999964	0,999928
3,98	0,000034	0,999966	0,999931
3,99	0,000033	0,999967	0,999934
4,00	0,000032	0,999968	0,999937



11 Literaturverzeichnis

- [1] Broschüre: Elementare Werkzeuge der Qualitätstechnik (C/QMM)
- [2] M. Sadowy: Industrielle Statistik, Vogel-Verlag, Würzburg, 1970
- [3] Erwin Kreyszig: Statistische Methoden und ihre Anwendungen, Vandenhoeck u. Ruprecht, Göttingen, 1988
- [4] H. Weber: Einführung in die Wahrscheinlichkeitsrechnung und Statistik für Ingenieure, Teubner, Stuttgart, 1988
- [5] Karl Bosch: Elementare Einführung in die Wahrscheinlichkeitsrechnung, Vieweg, Braunschweig, 1989
- [6] Karl Bosch: Elementare Einführung in die angewandte Statistik, Vieweg, Braunschweig, 1989
- [7] F. Barth, H. Berghold, R. Haller: Stochastik (Grundkurs, Leistungskurs), Ehrenwirth, München, 1992
- [8] G. Wagner und R. Lang: Statistische Auswertung von Meß- und Prüfergebnissen, Hrsg.: Deutsche Gesellschaft für Qualität e. V., DGQ 14, 1976
- [9] Graf, Henning, Stange, Wilrich: Formeln und Tabellen der angewandten mathematischen Statistik, Springer-Verlag, Berlin, 1987
- [10] Lothar Sachs: Angewandte Statistik, Springer-Verlag, Berlin, 1992
- [11] Hartung: Statistik, Oldenbourg, München, 1989



12 Formelzeichen und Begriffe

$\int_{-\infty}^{+\infty}$	= Integral von minus unendlich bis plus unendlich
$\sqrt{\quad}$	= Wurzelzeichen
\sum	= Summenzeichen
\prod	= Produktzeichen
\leq	= kleiner als oder gleich
\geq	= größer als oder gleich
\neq	= ungleich
$ x $	= Betrag von x (positiver Wert von x)
B_{un}, B_{ob}	= Faktoren zur Berechnung der Streugrenzen von s
C	= Mittenwert des Toleranzbereichs oder Sollwert
D_{un}, D_{ob}	= Faktoren zur Berechnung des Vertrauensbereichs für σ
e	= Basis des natürlichen Logarithmus ($\approx 2,71828$)
f	= Zahl der Freiheitsgrade
f(x)	= Wahrscheinlichkeitsdichtefunktion
G_j	= aufsummierte Besetzungszahl
h_j	= relative Häufigkeit
H_j	= relative Summenhäufigkeit
i, j	= Zählindizes
k	= Klassenzahl
ln	= natürlicher Logarithmus
m	= Anzahl von Wertegruppen
n	= Stichprobenumfang
n_j	= Absolute Häufigkeit in der j-ten Klasse
P	= Wahrscheinlichkeit
P_A	= Aussagewahrscheinlichkeit
R	= Range, Spannweite
s	= Standardabweichung der Stichprobe
s_R	= mit der Range-Methode bestimmte Standardabweichung
t'	= Faktor zur Berechnung des Vertrauensbereiches für μ bei unbekannter Standardabweichung der Grundgesamtheit

2020-04-06 - SOCCOS



u	= standardisierter Parameter der Normalverteilung $N(\mu = 0; \sigma^2 = 1)$
v	= Variationskoeffizient
w	= Klassenbreite
x	= stetige Merkmalswerte
x_i, y_i	= Werte einer Messreihe
$x_{(1)}, \dots, x_{(n)}$	= nach der Größe geordnete Werte einer Messreihe
\bar{x}_g	= geometrischer Mittelwert einer Stichprobe
\bar{x}_H	= Harmonischer Mittelwert
x_{\max}	= größter Wert einer Stichprobe
x_{\min}	= kleinster Wert einer Stichprobe
\tilde{x}	= Median (Zentralwert) einer Stichprobe
\bar{x}	= arithmetischer Mittelwert einer Stichprobe
Δ	= Intervallbreite
ε	= Streufaktor der Lognormalverteilung
μ	= Mittelwert einer Grundgesamtheit
μ_g	= geometrischer Mittelwert einer Grundgesamtheit
σ	= Standardabweichung einer Grundgesamtheit
σ^2	= Varianz einer Grundgesamtheit
π	= Zahl Pi ($\approx 3,1416$)

2020-04-06 - SOCCOS



Index

- Balkendiagramm 32
- deskriptive Statistik 5
- Dichtefunktion 34, 38
- Einflussgrößen ("5M") 56
- Gauß 34
- geordnete Liste 14
- Gesetz großer Zahlen 12
- Glockenkurve 35, 36, 37
- Grundgesamtheit 7
- Häufigkeit
 - absolute 31, 32
 - relative 31, 32
- Häufigkeitsschaubild 28, 42
- Histogramm 28
- induktive Statistik 5
- Integration 37
- Klassen-
 - breite 31
 - einteilung 28
 - grenze 30, 31, 32
 - mitte 31
- Logarithmus 50, 53
- Lognormalverteilung 48, 53
- Median 13, 50
- Merkmal 7
- Mischverteilung 54
- Mittelwert 35, 42
 - arithmetischer 15
 - geometrischer 17, 50, 53
 - gleitender 16
 - harmonischer 18
- Mittelwertkarte 57
- mittlere Abweichung 19
- Normalverteilung 34, 35, 56
- Punktdiagramm 28
- Qualitätsregelkarten 56
- Range-Methode 22
- Realisation einer Zufallsgröße 9
- Regelkarten
 - der Lage 56
 - für die Prozessstreuung 60
- R-Karte 62
- schiefe Verteilung 48
- Shewhart 56
- s-Karte 60, 61
- Spannweite 22, 62
- Standardabweichung 19, 35, 42
- Standardnormalverteilung 39
- Statistik 5
- Stichprobe 8
- Streufaktor 50
- Streuung
 - der Einzelwerte 58
 - der Mittelwerte 58
- Strichliste 28
- Summenhäufigkeit
 - absolute 31
 - relative 31, 32
- Summenkurve 33, 37
- Überschreitungsanteil 40, 43
- Urwertdiagramm 27
- Urwertkarte (x-Karte) 59
- Varianz 20
- Variationskoeffizient 21
- Verteilungsfunktion 34, 37, 38
- Vertrauensbereich
 - der Standardabweichung 66
 - des Mittelwerts 66
- Wahrscheinlichkeit 10, 34
- Wahrscheinlichkeitsnetz
 - der Lognormalverteilung 49, 50
 - der Normalverteilung 41, 42
- Zentralwert 14
- Zufalls-
 - experiment 8, 10
 - streubereich
 - der Mittelwerte 58
 - variable 8



Robert Bosch GmbH
C/QMM
Postfach 30 02 20
D-70442 Stuttgart
Germany
Phone +49 711 811 - 0
www.bosch.com



Robert Bosch GmbH

C/QMM

Postfach 30 02 20

D-70442 Stuttgart

Germany

Phone +49 711 811-0

www.bosch.com

