7. Wahrscheinlichkeitsverteilungen

7.1 Häufigkeitsverteilungen – Wahrscheinlichkeitsverteilungen

Häufigkeitsverteilungen (empirische Verteilungen, Kapitel 3) werden in der deskriptiven Statistik angewendet um die absolute und/oder relative Häufigkeit von Daten tabellarisch und/oder graphisch darzustellen. Dieser Verteilungstyp ist an empirische Daten gebunden. Liegen keine Daten vor, dann gibt es auch keine Häufigkeitsverteilung.

Wahrscheinlichkeitsverteilungen (theoretische Verteilungen) sind von Mathematikern theoretisch entwickelte Funktionen, die es erlauben, die Verteilung der Wahrscheinlichkeiten, mit denen Experimentalergebnisse auftreten, zu berechnen. Selbst wenn keine Daten vorliegen, sind solche Verteilung als mathematische Formulierung vorhanden. Für die Statistik wichtige Wahrscheinlichkeitsverteilungen sind u.a. die Normalverteilung, die Binominalverteilung, die Poissonverteilung, die F-Verteilung, die t-Verteilung und die Chi²-Verteilung. Bei der Normalverteilung haben wir in Kapitel 3 gesehen, wie empirische Daten angenähert durch eine Normalverteilung beschrieben werden können. Erfahrungsgemäß lassen sich viele Experimentalergebnisse bei hinreichend großen Stichproben durch die eine oder andere theoretische Verteilung approximiert beschreiben.

Wir unterscheiden zwei Formen der Wahrscheinlichkeitsverteilungen.

1. Diskrete Verteilungen z.B. Binominalverteilung
Poissonverteilung

Hier kann für jedes denkbare Versuchsergebnis berechnet werden, mit welcher Wahrscheinlichkeit es unter bekannten Bedingungen eintreten wird. Beispiel: Wir werfen fünf mal eine Münze und können berechnen, wie wahrscheinlich es ist, dass bei den fünf Würfen 3 mal die Zahl oben liegen wird.

2. Stetige Verteilungen z.B. Normalverteilung
Chi²-Verteilung
t-Verteilung
F-Verteilung

Hier können wir nicht für ein einzelnes Versuchsergebnis berechnen, mit welcher Wahrscheinlichkeit es eintreten wird. Weil es bei stetigen Daten (Reelle Zahlen) unendlich viele Ergebnisse geben kann ist für das Einzelergebnis die Wahrscheinlichkeit seines Auftretens gleich Null. Beispiel: Sie wiegen ein Hühnerei. Wie wahrscheinlich ist es, dass das Ei exakt 55,0000 ... 0 g wiegt. Es sind theoretisch unendlich viele Ergebnisse denkbar, da die Stellenzahl unbegrenzt ist. Selbst zwischen 55,0 g und 55,1 g liegen prinzipiell unendlich viele Werte, z.B. 55,0065492743456789 ... g. Die Wahrscheinlichkeit dafür, dass irgendein Wert auftritt ist so gering, dass wir dafür Null angeben. Wir können nur die Wahrscheinlichkeit dafür berechnen, dass ein beliebiges Ei zwischen 55,0 g und z.B. 60,0 g liegt. Siehe dazu bei Normalverteilung.

Ziel dieses Kapitels ist es, die Binominalverteilung und die Normalverteilung vorzustellen. Auf Chi²-Verteiung, t-Verteilung und F-Verteilung gehen wir später ein.

Theoretische, also mathematische Verteilungen werden durch Funktionsgleichungen beschrieben wie die folgende

Dichtefunktion der Binominalverteilung $f(k) = P_{n;k} = \binom{n}{k} * p^k * (1-p)^{n-k}$

Der Begriff Dichte bezeichnet das Auftreten von Wahrscheinlichkeiten von Zufallsvariablen.

Mit Hilfe dieser Gleichung, auf die wir weiter unten eingehen, können wir berechnen, wie groß die Wahrscheinlichkeit dafür ist, dass bei Zufallsexperimenten, bei denen nur zwei Ausgänge möglich sind, ein bestimmtes Ereignis eintritt. Betrachten wir in diesem Zusammenhang mal nicht Würfelexperimente sondern eine etwas komplexere biologische Situation, nämlich eine Maus, die sechs Junge wirft. Das Werfen eines Neugeborenen könnten wir als einen Versuch ansehen, ein Männchen zu gebären. Manchmal tritt bei einem solchen Versuch der Erfolg (Männchen) ein und manchmal der Misserfolg (Weibchen). [Natürlich können wir bei diesem Gedankenexperiment die Geschlechter wechseln.] Es hängt vom Zufall ab, wie viele Männchen und wie viele Weibchen in dem Sechserwurf sind. Wir werden mit dieser Gleichung berechnen, wie groß die Wahrscheinlichkeit dafür ist, dass in einem Sechserwurf genau 2 Männchen sind.

7.2 Die Binominalverteilung B(n;p)

(Nach Jacob Bernoulli auch Bernoulli-Verteilung)

Die Binominalverteilung ist eine bedeutsame Verteilung von Eintrittswahrscheinlichkeiten diskreter Daten. Im Vorfeld müssen wir zwei Begriffe erläutern.

Bernoulli-Experiment
Bernoulli-Experimente sind Zufallsexperimente, bei denen nur zwei Ausgänge, nämlich Erfolg und Misserfolg möglich sind. (Münzwurf-Experiment mit dem Ausgang „Kopf oder Zahl“, Würfelexperiment mit dem Ausgang „drei oder nicht drei“, Tierversuche mit dem Ausgang „tot oder lebendig“, „geheilt oder nicht geheilt“ oder bei der Überprüfung von Geräten die Möglichkeit „funktionstüchtig oder defekt“.) Die Wahrscheinlichkeit für den Erfolg wird mit p (Erfolgswahrscheinlichkeit) und die für den Misserfolg mit q bezeichnet. Da nur zwei Ausgänge möglich sind, muss auf jeden Fall einer der beiden Ausgänge eintreten. Nach einem Münzwurf liegt mit der Wahrscheinlichkeit 1 Kopf oder Zahl oben. Daher ist die Wahrscheinlichkeit dafür, dass Erfolg oder Misserfolg also p oder q eintritt gleich 1. Für die folgenden Berechnungen werden wir das Zeichen q durch den Term 1 – p ersetzen, (p + q = 1).

Bernoulli-Kette
Wenn wir ein Bernoulli-Experiment unter gleichen Bedingungen n mal durchführen und wenn die Ergebnisse der Experimente voneinander unabhängig sind, dann haben wir eine so genannte Bernoulli-Kette der Länge n und der Erfolgswahrscheinlichkeit p. Nach diesen beiden Parametern schreiben wir für eine Binominalverteilung auch B(n;p). Die Häufigkeit des Erfolgs bei n Experimenten bezeichnen wir mit k. k ist die Zufallsvariable deren Auftreten durch den Zufall beeinflusst wird.

Mit der Werten n = Anzahl der Versuche,
k = Anzahl der erfolgreichen Ausgänge bei n Versuchen und
p = Erfolgswahrscheinlichkeit

ist die Vorbedingung gegeben, die zur Berechnung der Binominalverteilung erforderlich ist. Mit diesen Werten können wir berechnen, wie groß die Wahrscheinlichkeit $P_{n;k}$ dafür ist, dass bei n Versuchen k mal der Erfolg eintritt. Wir wollen das an einem fiktiven Beispiel aus der Biologie untersuchen.

7.2.1 Wie groß ist die Wahrscheinlichkeit, dass bei n Versuchen k mal der Erfolg eintritt?

Beispiel 1 Wir züchten Mäuse des Stammes M. Eine Eigenschaft dieses Stammes sei, dass seine Nachkommen 55% ♂ und 45% ♀ sind. Eine Maus wirft n = 6 Junge. Die Anzahl der

Männchen und Weibchen im Sechserwurf ist vom Zufall abhängig. Auch die Reihenfolge, in der Männchen und Weibchen bei der Geburt erscheinen. Wie groß ist die Wahrscheinlichkeit $P_{n;k}$ dafür, dass bei dem Sechserwurf (n = 6) genau k = 2 ♂ vorhanden sind.

Es bedeuten
k = 2, gefragt ist nach 2 Männchen.
n = 6, es sind 6 „Versuche" der Maus, ein Männchen zu gebären (Erfolg).
p = 0,55 ist die Wahrscheinlichkeit dafür, ein Männchen zu gebären.
1 – p = 0,45 ist die Wahrscheinlichkeit dafür, kein Männchen, also ein Weibchen zu gebären.
$P_{n;k}$ = Wahrscheinlichkeit dafür, dass bei n Versuchen k mal der Erfolg eintritt. Das hier also unter n = 6 Jungtieren k = 2 Männchen sind.

1. Berechnung der Wahrscheinlichkeit für das Auftreten von 2 ♂ bei der Reihenfolge ♂ ♂ ♀ ♀ ♀ ♀.
Die im Beispiel gestellte Frage bezieht sich nicht auf eine bestimmte Reihenfolge der Geschlechter bei der Geburt. Wir wollen aber trotzdem die Wahrscheinlichkeit für das Auftreten dieser Reihenfolge berechnen. Betrachten wir zunächst die Reihenfolge ♂ ♂ ♀ ♀ ♀ ♀, bei der die beiden zuerst geborenen Tiere Männchen sind. Die Wahrscheinlichkeit P_k dafür, dass diese (♂ ♂ ♀ ♀ ♀ ♀) Reihenfolge eintritt ist,

P_k(♂∩♂∩♀∩♀∩♀∩♀) = p(♂) * p(♂) * p(♀) * p(♀) * p(♀) * p(♀) (Und-Verknüpfung, disjunkte Ereignisse)
P_k(♂∩♂∩♀∩♀∩♀∩♀) = 0,55 * 0,55 * 0,45 * 0,45 * 0,45 * 0,45
P_k(♂∩♂∩♀∩♀∩♀∩♀) = $0{,}55^2 * 0{,}45^4 = p^2 * (1-p)^4$
P_2(♂∩♂∩♀∩♀∩♀∩♀) = $p^2 * (1-p)^4$ ==> $\boxed{\mathbf{P_k = p^k\ (1-p)^{n-k}}}$
P_k(♂∩♂∩♀∩♀∩♀∩♀) = $p^k * (1-p)^{n-k}$
P_k(♂∩♂∩♀∩♀∩♀∩♀) = $0{,}55^2 * 0{,}45^4$
P_2(♂∩♂∩♀∩♀∩♀∩♀) = 0,3025 * 0,0410
P_2 = 0,01240

Die Wahrscheinlichkeit dafür, dass die Reihenfolge bei der Geburt diese ♂ ♂ ♀ ♀ ♀ ♀ ist, ist also P_2 = 1,24%.

2. Andere Reihenfolgen als ♂ ♂ ♀ ♀ ♀ ♀
Nun könnte die Geschlechterfolge bei der Geburt auch eine andere als die genannte (♂ ♂ ♀ ♀ ♀ ♀) sein, z.B. ♂ ♀ ♀ ♂ ♀ ♀. Wie viele Kombinationsmöglichkeiten es hier gibt, das können Sie durch kombinieren der Symbole ja mal versuchen festzustellen. Sie werden merken, dass das schnell unübersichtlich wird. Daher wollen wir die Anzahl der Kombinationen lieber über eine Gleichung aus der Kombinatorik berechnen.

Berechnung der Anzahl der Kombinationsmöglichkeiten von 2 ♂ und 4 ♀ bei einem Sechserwurf

Gleichung für die Kombination ($C_{n;k}$) von k = 2 Elementen aus n = 6 Elementen ohne Berücksichtigung der Anordnung.

$$C_{n;k} = \binom{n}{k} = \frac{n!}{k! * (n-k)!}$$

Es bedeuten
$C_{n;k}$ = Anzahl der Kombinationen von k Elementen aus n Elementen.
$\binom{n}{k}$ = gesprochen „n über k"
n! = gesprochen „n Fakultät" = n*(n-1)*(n-2)* ... *1
n = 6
k = 2

$$C_{n;k} = \binom{n}{k} = \frac{6!}{2! * 4!} = \frac{6*5*4*3*2*1}{2*1 * 4*3*2*1} = \frac{720}{48} = 15$$

Danach sind $C_{6;2}$ = 15 verschiedene Kombinationen möglich. Für ♂∩♂∩♀∩♀∩♀∩♀ gilt P_2= 1,24%. Welche Wahrscheinlichkeiten gelten für das Auftreten der 14 anderen Kombinationen?

3. Die Wahrscheinlichkeiten für alle 15 Kombinationsmöglichkeiten.
Die ursprünglich gestellte Frage bezog sich auf 2 Männchen, unabhängig davon, an welcher Stelle der Geburtenfolge sie stehen. Wir wissen nun, dass es 15 verschiedene Kombinationen für die Reihenfolge gibt und es stellt sich die Frage, wie groß P_k für die verschiedenen Kombinationen ist. Wir kennen diesen Wert für ♂ ♂ ♀ ♀ ♀ ♀ (1,24%). Würden Sie P_k für alle Kombinationen berechnen, dann würden Sie feststellen, dass für alle 15 Kombinationen mit 2♂ und 4♀ die Wahrscheinlichkeit mit 0,01240 identisch ist.

Bei einem Sechslingswurf kann aber nur <u>eine</u> der 15 möglichen Kombinationen (C) eintreten, also C1 ∪ C2 ∪ C3 ∪ C4 ∪ ... ∪ C15. Für Oder-Verknüpfungen [die Wahrscheinlichkeit dafür, dass eine der 15 Möglichkeiten eintritt] folgt die Addition der Einzelwahrscheinlichkeiten, also P(C1) + P(C2) + P(C3) +... P (C15) = 0,0124 + 0,0124 + 0,0124 + ... 0,0124 = 15 * 0,0124

Wir haben jetzt folgende Werte: $C_{n;k} = 15$
$P_k = 0,01240$

unsd danach folgt
$P_{n;k} = C_{n;k} * P_k$
$P_{6;2} = C_{6,2} * P_2$
$P_{n;k} = 15 * 0,01240$
$P_{n;k} = 0,186$

[Vielleicht lesen Sie nochmal in Kapitel 6 nach.]

Damit ist die Frage zu Beispiel 1 beantwortet:
Die Wahrscheinlichkeit für das Auftreten von 2 Männchen an beliegiger Stelle in der Reihenfolge bei einem Sechserwurf beträgt 18,6%. Bei 18,6% aller Sechserwürfe können wir erwarten, dass 2 ♂ dabei sind.

Erweiterte Frage zu Beispiel 1.

1. Andere Verteilungen der Anzahl der Geschlechter bei einem Sechserwurf
Wie ist die Verteilung der Wahrscheinlichkeiten, wenn k ($k \leq n$) variiert?

Diese Überlegungen folgen nicht mehr der Aufgabenstellung von Beispiel 1. Wir können Beispiel 1 aber nutzen, um eine weitere Betrachtung anzustellen: Neben der Realisation 2 ♂ und 4 ♀ (in beliebiger Folge) sind folgende Realisationen denkbar, nämlich

0 ♂ und 6 ♀
1 ♂ und 5 ♀
3 ♂ und 3 ♀
4 ♂ und 2 ♀
5 ♂ und 1 ♀
6 ♂ und 0 ♀

Wie sind die Wahrscheinlichkeiten für jede der sieben Realisationsmöglichkeiten? Wir variieren bei den Experimenten mit konstanten n = 6 und p = 0,55 den Wert für k von 0 bis 6 [Inkrement = 1] und fragen: Wie groß ist die Wahrscheinlichkeit, dass bei einem Sechserwurf 0,1,2,3,4,5,6 Männchen sind? Also sieben Fragen. Da wir die Reihenfolge der Geschlechter außer acht lassen können (die Wahrscheinlichkeit ist ja bei allen Kombinationsmöglichkeiten gleich), genügt die Berechnung der Wahrscheinlichkeiten für jeweils eine Reihenfolge der Realisationen, z.B. 0♂ und 6♀**.**

2. Berechnung der Wahrscheinlichkeiten für alle Variationen von k.
Beispielhaft hier die Berechnung der Anzahl der Kombinationen bei 0♂ und 6♀

0♂ und 6♀**:** P_k(♀∩♀∩♀∩♀∩♀∩♀) = p(♀) * p(♀) * p(♀) * p(♀) * p(♀) * p(♀)

$P_k = P_0 = 0{,}45^6 = \underline{0{,}008038}$ [n = 6; k = 0]

$$C_{n;k} = C_{6,0} = \binom{n}{k} = \frac{n!}{k! * (n-k)!} = \frac{6*5*4*3*2*1}{0! * 6!} = \frac{6*5*4*3*2*1}{1 * 6*5*4*3*2*1} = \frac{720}{720} = 1$$ [0! = 1]

Daraus folgt: $P_{6;0} = C_{6;0} * P_0 = 1 * 0{,}008038 = 0{,}008038$
Die Wahrscheinlichkeit dafür, dass 0 Männchen geboren werden ist 0,8%.

Und für alle Kombinationen gilt

k = 0 $P_{n;k}$ (0 ♂ und 6 ♀) = $C_{n;k} * p^k * (1-p)^{n-k} = 1 * 0{,}55^0 * 0{,}45^6 = 0{,}0080$
k = 1 $P_{n;k(}$(1 ♂ und 5 ♀) = $C_{n;k} * p^k * (1-p)^{n-k} = 6 * 0{,}55^1 * 0{,}45^5 = 0{,}0609$
k = 2 $P_{n;k(}$(2 ♂ und 4 ♀) = $C_{n;k} * p^k * (1-p)^{n-k} = 15 * 0{,}55^2 * 0{,}45^4 = 0{,}1860$
k = 3 $P_{n;k(}$(3 ♂ und 3 ♀) = $C_{n;k} * p^k * (1-p)^{n-k} = 20 * 0{,}55^3 * 0{,}45^3 = 0{,}3032$
k = 4 $P_{n;k(}$(4 ♂ und 2 ♀) = $C_{n;k} * p^k * (1-p)^{n-k} = 15 * 0{,}55^4 * 0{,}45^2 = 0{,}2780$
k = 5 $P_{n;k(}$(5 ♂ und 1 ♀) = $C_{n;k} * p^k * (1-p)^{n-k} = 6 * 0{,}55^5 * 0{,}45^1 = 0{,}1359$
k = 6 $P_{n;k(}$(6 ♂ und 0 ♀) = $C_{n;k} * p^k * (1-p)^{n-k} = 1 * 0{,}55^6 * 0{,}45^0 = 0{,}0277$

Die zusammengefassten Ergebnisse zeigen in Abb.1, dass es am wahrscheinlichsten ist, dass gleich viele Männchen und Weibchen in dem Wurf sind. ($P_{6;3} = 0{,}3032$). Extrema sind am unwahrscheinlichsten. Die Verteilung der Wahrscheinlichkeiten für alle möglichen k ist schwach asymmetrisch. Symmetrisch wäre Sie nur bei p = 0,5.

k	$P_{n;k}$
0	0,0080
1	0,0609
2	0,1816
3	0,3032
4	0,2780
5	0,1359
6	0,0277

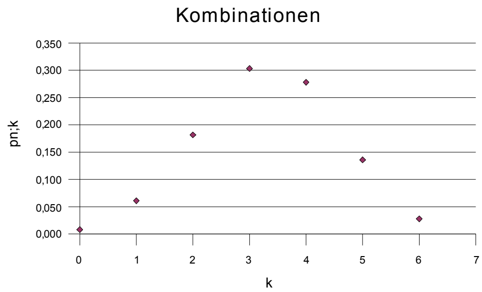

Abb.1

Beispiel 2 Wie groß ist die Wahrscheinlichkeit dafür, dass bei n Versuchen k mal der Erfolg eintritt? Eine Urne enthält 20 rote Kugel (r) und 30 nichtrote Kugeln (n). Es soll gewettet werden. Spieler A sagt, von den vier Kugeln, die er bei vier Zügen erhält, würden zwei rote sein (und natürlich zwei nichtrote). Spieler B sagt, er würde fünf Züge machen und drei mal rot ziehen (und natürlich zwei mal nichtrot). Nach jedem Zug wird die gezogene Kugel zurückgelegt damit die Wahrscheinlichkeiten gleich bleiben. Welcher Spieler hat die größere Chance, das Spiel zu gewinnen?
Zu berechnen ist die Wahrscheinlichkeit, mit der das Erfolgsergebnis = rote Kugeln bei n Versuchen k mal eintritt.

Betrachten wir die Situation von Spieler A.
Wie groß ist die Wahrscheinlichkeit, dass bei vier maligem Ziehen je einer Kugel zwei rote Kugeln gezogen werden?

Zur Berechnung über die Gleichung $P_{n;k} = C_{n;k} * p^k * (1-p)^{n-k}$ benötigen wir
$C_{n;k} = \binom{n}{k}$

n = Anzahl der Versuche, hier 4 (Ziehungen)
k = Anzahl der Erfolgsergebnisse, hier 2 rote Kugeln (bei 4 Versuchen)
p = Wahrscheinlichkeit für das Auftreten des Erfolgsergebnisses bei <u>einem</u> Versuch, hier p(r) = 20/50 = 0,4
1-p = 0,6

Dann gilt

$$P_{n;k} = \frac{n!}{k!\,(n-k)!} * p^k * (1-p)^{n-k}$$

$$P_{4;2} = \frac{4!}{2! * (4-2)!} * 0{,}4^2 * 0{,}6^{4-2}$$

$$P_{4;2} = \frac{4*3*2*1}{2*1 * 2*1} * 0{,}4^2 * 0{,}6^{4-2}$$

$$P_{4;2} = \frac{24}{4} * 0{,}16 * 0{,}36$$

$$P_{4;2} = 6 * 0{,}0576$$

$$\underline{P_{4;2} = 0{,}3456}$$

Die Wahrscheinlichkeit, dass eine der 6 Kombinationen (r r n n, r n r n, r n n r, n r r n, n r n r, n n r r) auftritt ist = 34,65%.

Nun die Situation für Spieler B

n = Anzahl der Versuche, hier 5 (Ziehungen)
k = Anzahl der Erfolgsergebnisse, hier 3 rote Kugeln bei 5 Versuchen
p = 0,4

$$P_{n;k} = \frac{n!}{k!\,(n-k)!} * p^k * (1-p)^{n-k}$$

$$P_{5;3} = \frac{120}{12} * 0{,}064 * 0{,}36 = 10 * 0{,}064 * 0{,}36 = 0{,}2304$$

Da $P_{4;2} = 0{,}3456 > P_{5;3} = 0{,}2304$, hat Spieler A hat also größere Chance, das Spiel zu gewinnen.

7.2.2 Wie ändert sich P, wenn die Anzahl n der Versuche gegen unendlich strebt?

Beispiel 3 Bei einem Münzwurf ist die Erfolgswahrscheinlichkeit p für „Zahl liegt oben" = 0,5.
Wie groß ist die Wahrscheinlichkeit dafür, dass <u>ein mal</u> „Zahl oben" liegt bei jeweils n = 1, 2, 3, 4, 5,10, 20, 50, 100 Würfen?

Beispielrechnung für n = 3

n = 3, k = 1; p = 0,5

$$P_{n;k} = \binom{n}{k} * p^k * (1-p)^{n-k}$$

$$P_{3,1} = \frac{3!}{1! * 2!} * 0{,}5^1 * 0{,}5^2$$

$$P_{3,1} = 3 * 0{,}5 * 0{,}25$$
$$P_{3,1} = 3 * 0{,}125 * 0{,}25$$
$$P_{3,1} = \underline{0{,}375}$$

Über die Excelfunktion BINOMVERT erhalten wir folgende Ergebnisse
[In die Eingabemaske für BINOMVERT muss bei „Kumuliert“ eine 0 eingegeben werden]

k=1	n	1	2	3	4	5	10	20	50	100
	$P_{n;k}$	0,5	0,5	0,375	0,2500	0,1562	0,00976	$1{,}907 \cdot 10^{-5}$	$4{,}44 \cdot 10^{-14}$	$7{,}89 \cdot 10^{-29}$

Tabelle 1

Aus Abb.2 ist zu entnehmen, dass die Wahrscheinlichkeit, das gewünschte Ziel (ein mal „Zahl liegt oben“) zu erreichen mit steigendem n immer geringer wird. Letztendlich strebt die Wahrscheinlichkeit gegen Null, wenn n gegen unendlich strebt. [Wenn $n \Rightarrow \infty$, dann $p \Rightarrow 0$]. Wir empfinden auch intuitiv, dass es sehr, sehr unwahrscheinlich ist, dass bei sehr vielen Würfen insgesamt nur ein mal die Zahl oben liegen wird.

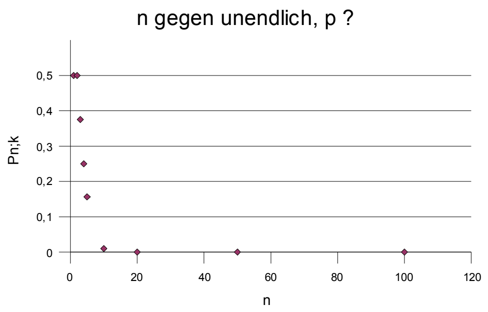

Abb.2

7.2.3 Die Symmetrie einer Binominalverteilung ist abhängig von p.

Beispiel 4 Um dies zu zeigen berechnen wir die Wahrscheinlichkeit ($P_{n;k}$) dafür, dass bei n = 4 der Erfolg mit k = 0; 1; 2; 3; 4 mal eintritt. Wir führen den Versuch mit folgenden Anordnungen durch:

1. n = 4; k = 0; k = 1; k = 2; k = 3; k= 4; p = 0,3
2. n = 4; k = 0; k = 1; k = 2; k = 3; k= 4; p = 0,5
3. n = 4; k = 0; k = 1; k = 2; k = 3; k= 4; p = 0,7

Beispielhaft die Berechnung für p = 0,3 und k = 0

$$P_{n;k} = \binom{n}{k} p^k * (1-p)^{n-k}$$

$$P_{4;0} = \binom{4}{0} p^0 * (1-p)^4$$

$$P_{4;0} = \binom{4}{0} 0{,}3^0 * 0{,}7^4$$

$P_{4;0} = 1 * 1 * 0{,}2401$
$\underline{P_{4;0} = 0{,}2401}$

Die Ergebnisse über BINOMVERT berechnet:

p = 0,3 Graphik grün		p = 0,5 Graphik rot		p = 0,7 Graphik schwarz	
k	**P**	**k**	**P**	**k**	**P**
0	0,2401	0	0,0625	0	0,0081
1	0,4116	1	0,25	1	0,0756
2	0,2646	2	0,375	2	0,2646
3	0,0756	3	0,25	3	0,4116
4	0,0081	4	0,0625	4	0,2401

Tabelle 2

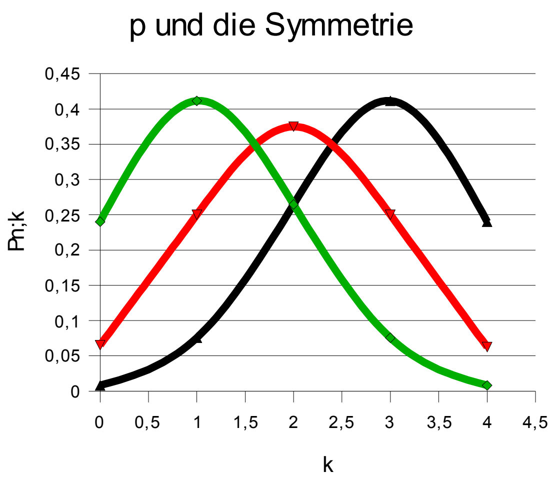

Abb.3

Wir entnehmen Abb.3 dass die Binominalverteilung bei p = 0,5 symmetrisch ist. Dies gilt grundsätzlich bei p = 0,5 und nur dann.

7.3 Normalverteilung N(μ,σ)

Die Normalverteilung ist eine der wichtigsten Wahrscheinlichkeitsverteilungen stetiger Daten. Die ihr zugrunde liegenden mathematischen Zusammenhänge wurden unabhängig voneinander gleich drei mal gefunden. Zunächst erkannte Abraham de Moivre diese Verteilung im 18. Jh. im Zusammenhang mit Glücksspielen. Später wurde sie von Pierre Simon Laplace ebenfalls bei der Beschäftigung mit Glücksspielen „wieder entdeckt“ und Carl Friedrich Gauß entwickelte sie anlässlich von Untersuchungen zu Messfehlern bei astronomischen Messungen. Die Bezeichnung „normal distribution“ wurde erst 1893 von Karl Pearson eingeführt. Wir verwenden heute meist die Bezeichnungen Normalverteilung, seltener de-Moivre-Verteilung oder Gaußverteilung. Die Normalverteilung steht in engem Zusammenhang mit Zufallsvariablen.

7.3.1 Zufallsvariable

Viele empirische Daten sind Zufallsvariable. Sie entstehen, wenn z.B. bei Messvorgängen die Messwerte durch viele unabhängige Faktoren zufällig beeinflusst werden. Wir füllen mit einer Vollpipette jeweils 100 µL einer Lösung in 20 Reaktionsgefäße. Ob wir exakt 100 µL abmessen oder etwas weniger/mehr, das hängt - grobe Fehler ausgeschlossen - vom Zufall ab. Den Messwert 100 µL beeinflussen während des Messvorgangs viele Faktoren wie z.B. Schwankungen der Umgebungstemperatur, der Aufmerksamkeit des Messenden sowie dessen Ermüdungszustand. Dies führt über unterschiedliche Meniskuseinstellungen und

Abtropfzeiten zu Abweichungen vom Sollwert 100 µL. Solche Einflüsse wirken sich in ihrer Summe so auf die Messwerte aus, dass bei vielen Messungen im Mittel 100 µL abgemessen werden (Kapitel 2 Zufallsfehler). Aber es kommen Abweichungen nach oben und unten vor, beide gleichermaßen stark. Die Abweichungen werden um so seltener, je extremer sie sind. So wie der Volumenmesswert sind viele natürliche Daten als Zufallsvariable zu betrachten, die durch viele voneinander unabhängige Einflussgrößen zustande gekommen sind. Dies ist der Hintergrund dafür, dass zumindest approximiert normalverteilte Daten in der Natur relativ weit verbreitet sind. Der Zusammenhang zwischen Zufallsvariablen und der Normalverteilung wird durch einen wichtigen Satz der Statistik erklärt, durch den Zentralen Grenzwertsatz. Dieser Satz, auf den wir nicht näher eingehen werden, besagt, dass die Summe bzw. die arithmetischen Mittel unabhängiger Zufallsvariabler zumindest approximiert normalverteilt sind.

7.3.2 Berechnung der Normalverteilung

Die Graphik einer Normalverteilung ist die aus Kapiteil drei bekannte Glockenkurve (Abb. 4). Wir haben dort empirische Daten (Gewichte von Hühnereiern) mit der Glockenkurve verglichen. Wir wollen hier an einem Beispiel zeigen, wie die Kurve zustande kommt. In ihrer idealen, symmetrischen Form werden wir sie bei empirische Daten nicht vorfinden, je mehr Daten vorliegen um so besser nähert sich die Kurve vieler empirischer Daten aber der idealen Form.

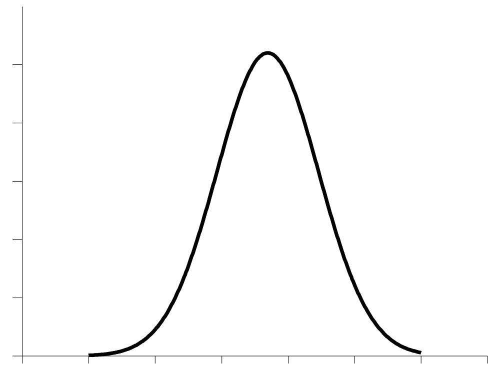

Abb.4

Beispiel 5 Durchmesser menschlicher Erythrozyten

An einem gefärbten Blutausstrich haben wir bei 300 Erythrozyten den Durchmesser in µm gemessen. Durch systematische Auswahl der vermessenen Zellen wurden Doppelmessungen vermieden. Wir gehen davon aus, dass der Erythrozytendurchmesser eine Zufallsvariable ist, deren Messergebnisse zur Protokollierung auf 0,25 µm gekörnt wurden.

Aus den Messergebnissen

µm	5,00	5,25	5,50	5,75	6,00	6,25	6,50	6,75	7,00	7,25	7,50	7,75	8,00	8,25	8,50	8,75	9,00	9,25	9,50	9,75	10,00
H	1	0	2	2	3	4	10	15	25	35	40	45	40	25	20	16	10	0	4	2	1

H = absolute Häufigkeit.

haben wir die folgenden Kennwerte berechnet, diese sind gerundet: $\bar{x}$ = 7,69 µm und s_x = 0,77 µm.

Die Messwerte der Stichprobe ergeben das Bild von Abb.5. Danach liegt das Maximum etwa in der Mitte des ranges. Extremwerte kommen um so seltener vor, je extremer sie sind. Die Kurve kommt also der Glockenkurve recht nahe. Wir folgern daraus, dass die Daten approximiert einer Normalverteilung entsprechen.

Abb. 5

Über die Wahrscheinlichkeitsdichte [f(x)] können wir berechnen, wie die ideale Glockenkurve für die Erythrozytendurchmesser der Stichprobe aussieht. Was wir bei der empirischen Häufigkeitsverteilung die absolute Häufigkeit genannt haben, das nennen wir hier die Wahrscheinlichkeitsdichte.

$$f(x) = y = \frac{1}{\sigma * \sqrt{(2\pi)}} * e^{-0,5(\{x-\mu\} / \sigma)^2}$$

Notationen $f(x) = y =$ Wahrscheinlichkeitsdichte
σ = Parameter der Standardabweichung
μ = Parameter des arithmetischen Mittels
π = 3,1416
e = Eulersche Zahl 2,7182

Die in die Gleichung einzusetzenden Parameter μ und σ stehen nicht zur Verfügung, da wir ja nur eine Stichprobe untersucht haben. Wir betrachten den Stichprobenumfang von n = 300 aber als so groß, dass die Kennwerte $\bar{x}$ und s_x hinreichend gute Repräsentanten für deren Parameter sind und somit an deren Stelle in der Rechnung verwendet werden dürfen. π und e sind Konstanten und daher ist der Verlauf der Verteilung bzw. die Form der Kurve ausschließlich von den beiden Parametern μ und σ abhängig. Die Normalverteilung wir daher mit dem Ausdruck **N(μ,σ)** gekennzeichnet.

Wir wollen nun zeigen, wie die Normalverteilungskurve für die Erythrozytendurchmesser mit den Werten $\bar{x}$ (μ) = 7,69 µm und s_x (σ) = 0,77 µm berechnet wird.
Dazu werden wir für jeden der 21 Messwerte berechnen, wie groß die Wahrscheinlichkeitsdichte für ihr auftreten ist, also jeweils deren f(x)-Wert. Beispielhaft folgt hier die Berechnung von f(x) für x = 6,00 µm.

$$f(x) = y = \frac{1}{\sigma * \sqrt{(2\pi)}} * e^{-0,5(\{x-\mu\} / \sigma)^2}$$

$$f(x) = \frac{1}{0,77 * \sqrt{6,2832}} * e^{-0,5(\{6 - 7,69\} / 0,77)^2}$$

$$f(x) = \frac{1}{1,9301} * e^{-0,5(-1,69 / 0,77)^2}$$

$$f(x) = 0,5181 * e^{-0,5(-2,1948)^2}$$

$$f(x) = 0,5181 * e^{-0,5 * 4,8172}$$

$$f(x) = 0,5181 * e^{-2,4086}$$

$$f(x) = 0,5181 * 2,7182^{-2,4086}$$

$$f(x) = 0,5181 * 0,08995$$

$$\underline{f(x) = 0,0466}$$

Da die Rechnung relativ zeitaufwendig ist, haben wir für alle x im geschlossenen Intervall [5;10] über die Excelfunktion NORMVERT die f(x)-Werte berechnet. Wer möchte, der kann die Berechnung nach Eingabe der obigen Formel über die Tastatur ausführen. Vorsicht, die Formeleingabe ist fehlerträchtig.

Die mit NORMVERT berechneten Werte sind (auf fünf Nachkommastellen gerundet)
[In die Eingabemaske von NORMVERT muss bei „Kumuliert“ eine 0 eingegeben werden.]

µm	5,00	5,25	5,50	5,75	6,00	6,25	6,50	6,75	7,00	7,25	7,50	7,75	8,00	8,25	8,50	8,75	9,00	9,25	9,50	9,75	10,00
f(x)	0,00116	0,00342	0,00908	0,02168	0,04660	0,09015	0,15696	0.24593	0,34678	0,44006	0,50257	0,51654	0,47777	0,39771	0,29794	0,20087	0,12187	0,06655	0,03270	0,01446	0,00576

Tabelle 3

Die Graphik der Wahrscheinlichkeitsdichte ist die folgende Glockenkurve in Abb.6.

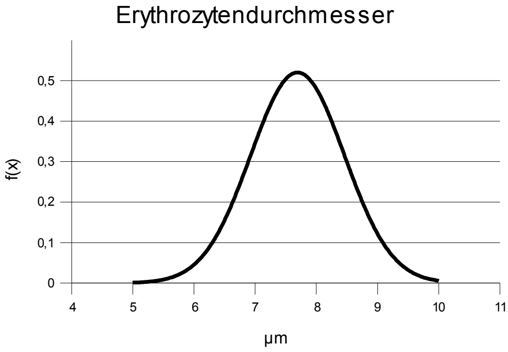

Abb.6

Typische Eigenschaften der Glockenkurve
Die Abb.6 zeigt die berechnete Wahrscheinlichkeitsverteilung für die Parameter $\mu = 7{,}69$ und $\sigma = 0{,}77$. Wir können der Kurve die typischen Eigenschaften der Normalverteilung entnehmen.(Siehe auch Kapitel 3.)

1. Die Graphik ist symmetrisch und glockenförmig
2. Sie ist monomodal
3. Die Funktionswerte f(x) nähern sich der Abszisse beidseitig asymptotisch. Das bedeutet, Extremwerte kommen um so seltener vor, je extremer sie sind.
4. Der Erwartungswert = µ liegt unter $f(x)_{max}$ in der Symmetrieachse. Das Lot von $f(x)_{max}$ der Kurve zeigt auf der Abszisse das arithmetisches Mittel, den Modalwert und den Medianwert an. Diese drei Werte sind bei einer Normalverteilung identisch.
5. Die Grenzen der Standardabweichung liegen auf der Abszisse bei $\sigma \pm s_x$. Diese Punkte entsprechen den beiden Wendepunkten (~$0{,}6 * f(x)_{max}$).
6. Normalverteilte Variable sind charakterisiert durch $N(\mu;\sigma)$.

7.4 Die Variation der Messwerte in der Normalverteilung

Zu den speziellen Eigenschaften der Normalverteilung gehört der Zusammenhang zwischen den Streubereichen ($\bar{x} \pm s_x$) und dem Flächenanteil unter der Kurve (Abb.7). Wir wollen daher die Grenzen des einfachen, doppelten und dreifachen Streubereichs, die wir auch deren Schranken nennen, berechnen.

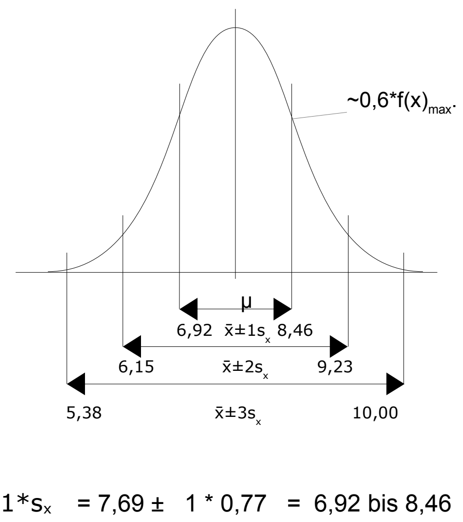

Abb.7

Einfacher Streubereich : $\bar{x} \pm 1*s_x = 7{,}69 \pm 1 * 0{,}77 = 6{,}92$ bis $8{,}46$
Doppelter Streubereich : $\bar{x} \pm 2*s_x = 7{,}69 \pm 2 * 0{,}77 = 6{,}15$ bis $9{,}23$
Dreifacher Streubereich: $\bar{x} \pm 3*s_x = 7{,}69 \pm 3 * 0{,}77 = 5{,}38$ bis $10{,}00$

Wenn wir auf der Abszisse einer Normalverteilungskurve μ und die Streubereichsgrenzen eintragen und dann über diesen Punkten die Lote errichten, dann entstehen abgegrenzte Bereiche unter der Kurve. Würden wir die Flächen dieser Bereiche ausmessen oder berechnen, dann würden wir folgendes finden.

Die Flächen links und rechts von der Symmetrieachse entsprechen jeweils 50% der Gesamtfläche unter der Kurve (Abb.8).

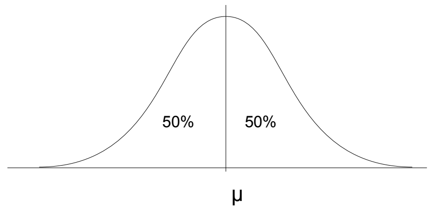

Abb.8

Innerhalb der Schranken des einfachen Streubereichs, also im Bereich $\bar{x} \pm 1*s_x$ liegen rund 68,3% der Gesamtfläche unter der Kurve (Abb.9).

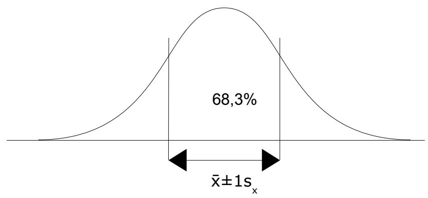

Abb.9

Innerhalb der Schranken von $\bar{x} \pm 2*s_x$ liegen rund 95,5% der Gesamtfläche unter der Kurve (Abb.10).

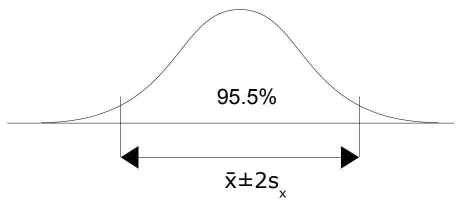

Abb.10

Innerhalb der Schranken von $\bar{x} \pm 3*s_x$ liegen rund 99,7% der Gesamtfläche unter der Kurve (Abb.11).

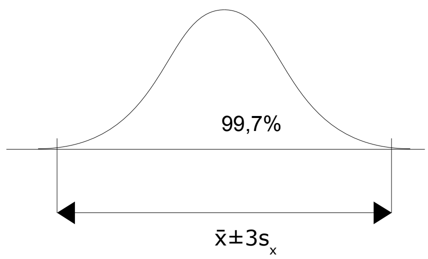

Abb.11

Die genannten Flächenanteile der Streubereiche gelten für alle N(μ;s), unabhängig davon, welche Werte μ und s annehmen.

Die Gesamtfläche unter der Kurve (100%) entspricht der Wahrscheinlichkeit dafür, dass ein beliebiger Messwert irgendwo auf der Abszisse unter der Kurve liegt, also der Wahrscheinlichkeit 1 (100%). Die Fläche unter der Kurve links von der Symmetrieachse entspricht 50% der Gesamtfläche und somit beträgt die Wahrscheinlichkeit, dass ein beliebiger Messwert links der Symmetrieachse liegt, 50%. Die prozentualen Flächenanteile werden also als Wahrscheinlichkeiten interpretiert. Diese entsprechen auf der Abszisse einer Strecke, nicht einem Punkt. $\bar{x} \pm 1*s_x$ ist eine Strecke, ein Intervall. Der Faktor für s_x mag noch so klein sein, es bleibt immer eine Strecke wie z.B. bei $\bar{x} \pm 0{,}001*s_x$. Da ein Flächenanteil aber stets über einem Intervall auf der Abszisse liegt, kann auch nur für ein Intervall eine Wahrscheinlichkeit angegeben werden, nicht für einen einzelnen Wert auf der Abszisse. Durch unsere Kenntnis über die Streubereiche können wir nun z.B folgende Aussagen machen:

Ein beliebiger Wert der Population liegt mit ca.
68,3%iger Wahrscheinlichkeit in $\bar{x} \pm 1*s_x$
95,5%iger Wahrscheinlichkeit in $\bar{x} \pm 2*s_x$
99,7%iger Wahrscheinlichkeit in $\bar{x} \pm 3*s_x$
31,7%iger Wahrscheinlichkeit ausserhalb $\bar{x} \pm 1*s_x$
4,5%iger Wahrscheinlichkeit ausserhalb $\bar{x} \pm 2*s_x$
0,3%iger Wahrscheinlichkeit ausserhalb $\bar{x} \pm 3*s_x$
15,8%iger Wahrscheinlichkeit unterhalb $\bar{x} - 1*s_x$
15,8%iger Wahrscheinlichkeit oberhalb $\bar{x} + 1*s_x$

Prüfen Sie mal wie viel % der 300 Messwerte innerhalb der Streugrenzen liegen. Die dazu nötige Häufigkeitsliste finden Sie weiter vorne.

Sie sollten zu folgendem Ergebnis kommen: Im einfachen Streubereich liegen 70%, im doppelten Streubereich liegen 95%, im dreifachen Streubereich liegen 99,3%. Die Werte liegen nahe bei den theoretischen Werten.

Alle zu den genannten drei Streubereichen gehörenden Prozentwerte sind „krumm“. Mit „krumm“ meine ich Zahlen wie 95,5, mit „glatt“ solche wie 95,0. Wir könnten nun fragen, wie wir zu Streubereichen kommen können, die „glatten“ Prozentwerten entsprechen. Hier schon mal zwei Werte:

„krumm“	95,5% --> $\sigma + 2*s_x$	99,7% --> $\sigma + 3*s_x$
„glatt“	95,0% --> $\sigma + 1,96*s_x$	99,9% --> $\sigma + 3,29*s_x$

Bei der folgenden Standardisierung der Normalverteilung werden wir sehen, wie diese Werte errechnet werden. Fast alle Werte (99,7%) liegen bei einer Normalverteilung innerhalb der dreifachen Streugrenzen. Nach der so genannten „3-sigma-Regel“ zählen manche Statistiker die 0,3% die außerhalb dieses Bereiches an den Enden der Verteilung liegen, nicht mehr zur Grundgesamtheit sondern werten sie als Ausreißer. Solche Werte können fälschlich z.B. durch Messfehler in die Stichprobe geraten sein. Sie können aber auch zur Grundgesamtheit gehören. Ob ein solcher Wert, der außerhalb der dreifachen Streuung liegt, wirklich ein Ausreißer ist, dass prüfen wir in der Inferenzstatistik mit einem speziellen Ausreißertest.

7.4.1 Wie beeinflusst die Standardabweichung die Form der Kurve?

Da in die Berechnung der Kurve nur die beiden Parameter μ und s eingehen, kann die Form der Kurve nur von diesen beiden Werten abhängen. Eine Vergrößerung von μ bewirkt eine Verschiebung der Kurve auf der Abszisse nach rechts. Bei Verkleinerung von μ rutscht sie nach links. (Abb.12). σ wirkt sich auf die Breite der Kurve aus. Steigt σ, so wird die Kurve flacher (grün σ = 1,8), sinkt σ, so wird sie schmaler und höher. In Abb. 13 gilt schwarz: σ = 0,25, rot: σ = 0,77.

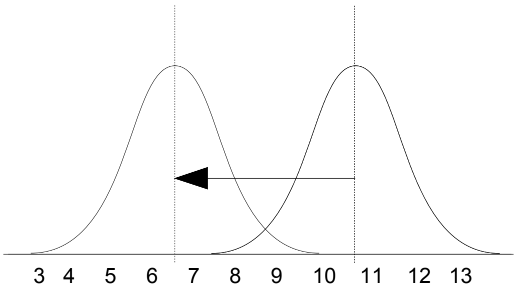

Abb.12

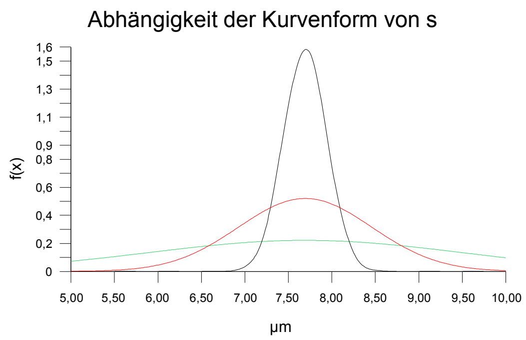

Abb.13

7.5 Standardisierung der Normalverteilung

Transformation der Normalverteilung (N(μ,σ) in die Standardnormalverteilung N(0,1)

Die Normierung hat das Ziel, Berechnungen zur Normalverteilung zu vereinfachen.

Beispiel 6 Wir wollen berechnen, mit welcher Wahrscheinlichkeit ein bestimmter X-Wert einer Verteilung oberhalb oder unterhalb einer festgelegten Grenze oder zwischen zwei Grenzen auf der Abszisse liegt.

Folgende Daten aus Beispiel 5 sind bekannt: range 5,0 µm bis 10,0 µm
$\bar{x}$ = 7,69 µm
s_x = 0,77 µm
$\bar{x} \pm 1^* s_x$ = 6,92 bis 8,46
$\bar{x} \pm 2 * s_x$ = 6,15 bis 9,23
$\bar{x} \pm 3 * s_x$ = 5,38 bis 10,00

Frage: Wie groß ist die Wahrscheinlichkeit dafür, dass ein beliebiger Erythrozyt aus der Population einen Durchmesser von >9,0 µm hat? Wie wir wissen, entspricht die gesuchte Wahrscheinlichkeit dem Flächenanteil unter der Kurve rechts von 9,0 µm. Stünde statt der 9 µm die 9,23 µm in der Frage, dann wäre die Lösung einfach. 9,23 µm entsprechen nämlich der oberen Schranke des doppelten Streubereichs. Im doppelten Streubereich liegen 95,5% aller Werte. Oberhalb der oberen Schranke liegt dann die Hälfte der Differenz zu 100%, das sind rund 2,25%. Die Ermittlung der Fläche rechts von 9,0 µm über eine Integralrechnung ist sehr aufwendig aber machbar. Wollten wir die Fläche rechts von 7,4 µm wissen, dann wäre eine neue Berechnung notwendig. Je nach Fragestellung wäre für jeden x-Wert-Spanne in [5;10] eine Neuberechnung des Flächenanteils notwendig. Auch das wäre machbar und die erhaltenen Werte könnten tabelliert werden. Da die Flächenanteile für die Abszissenwerte jeder anderen Normalverteilung (andere Datensätze) ebenfalls neu zu berechnen wären, müsste für jeden denkbaren Datensatz eine solche Tabelle erstellt werden. Da das kaum praktikabel wäre wird die Normalverteilung standardisiert. Damit erreichen wir, dass für beliebige Datensätze die Flächen mit nur <u>einer</u> Tabelle, der z-Tabelle, sehr einfach berechnet werden können. Bei der Standardisierung gehen wir im Prinzip folgendermaßen vor. Wir verschieben die Kurve auf der Abszisse soweit nach links, dass $(f(x)_{max}$ über dem Koordinatenmittelpunkt (Null) steht (Abb.14). Da, wo die Streugrenzen auf der Abszisse liegen, kennzeichnen wie diese Punkte wie folgt:

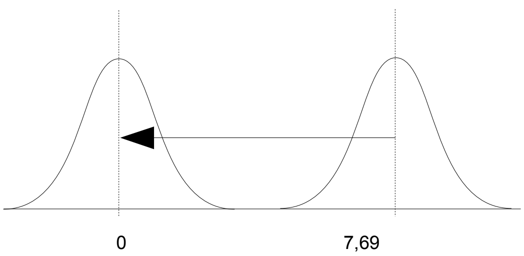

Abb. 14

Alte Abszisse	Neue Abszisse
5,38 µm - 3 s	- 3 z
6,15 µm - 2 s	- 2 z
6,93 µm - 3 s	- 1 z
7,69 µm μ	0
8,46 µm 1 s	1 z
9,23 µm 2 s	2 z
10,00 µm 3 s	3 z

Tabelle 4

1s wird zu 1z, 2s zu 2z und 3s zu 3z, das Entsprechende bei den negativen Streugrenzen. Die so erhaltene Skala nennen wir z-Skala und die Werte z-Werte. Unsere Kurve (Abb.14 und 15) hat sich nicht verändert, sie ist nur auf der Abszisse so verschoben worden, dass $f(x)_{max}$. über x = 0 liegt. Und auf der Abszisse finden wir nun nicht mehr die Messwerte, sondern die Grenzen der Streubereiche in z-Werten. Siehe Abb.16.

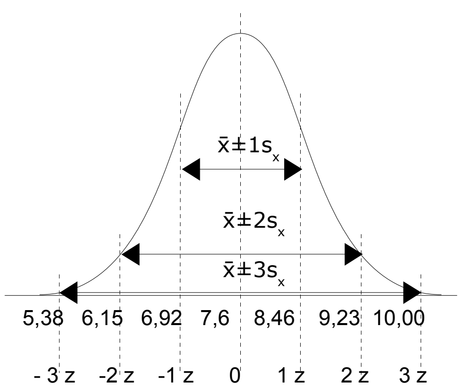

Abb. 15

Der einfache Streubereich geht von - 1 z bis 1 z
Der zweifache Streubereich geht von - 2 z bis 2 z
Der dreifache Streubereich geht von - 3 z bis 3 z

Nun zurück zu unserer Frage:
Mit welcher Wahrscheinlichkeit ist der Durchmesser eines Erythrozyten >9 µm? Auf der z-Abszisse gibt es den Wert 9 nicht mehr, da stehe nur z-Werte. Jetzt kommt der wesentliche Punkt: Wir müssen den Wert 9 in einen z-Wert umrechnen.

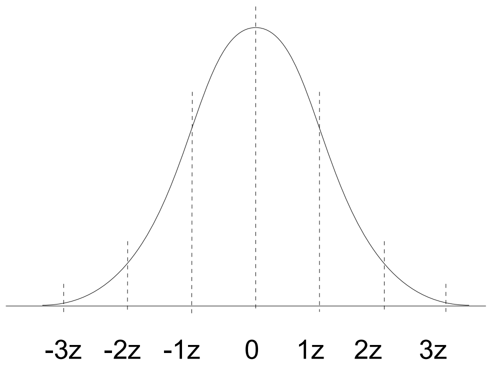

Abb.16

Die geschieht nach $\boxed{z = (x - \mu)/\sigma}$

Dabei ist $x = 9$, $\mu = 7{,}69$ und $\sigma = 0{,}77$

Es folgt also $z = (9 - 7{,}69)/0{,}77$
$z = 1{,}7$

Nur benötigen wir die schon erwähnte z-Tabelle, die z.B. über das Internet zu erreichen ist. Auf die Frage „Wie groß ist der Flächenanteil rechts von 9 µm?“ ist nun die Frage geworden „Wie groß ist der Flächenanteil rechts von 1,7 z?“ Aus der Tabelle können wir entnehmen, welcher Prozentwert für z = 1,7 gilt.

Die folgende verkürzte z-Tabelle enthält nur positive z-Werte mit einer Nachkommastelle. Erhalten wir bei Berechnungen z-Werte mit mehr Nachkommastellen, so runden runden wir den Tabellenwerten entsprechend oder wir finden den P-Wert durch graphische Interpolation in der, der Tabelle entsprechenden Kurve Abb. 18. Erhalten wir bei der Berechnung negative z-Werte, so ignorieren wir das -Zeichen. Die Tabelle entspricht dem rechten Teil der Kurve. Da diese symmetrisch ist, entsprechen die negativen z-Werte ihren positiven Pendant.

Flächen (Wahrscheinlichkeiten), die zwischen z und 0 liegen

z	p	z	p	z	p	z	p	z	p	z	p
0,0	0,0000	0,6	0,2257	1,2	0,3849	1,8	0,4641	2,4	0,4918	3,0	0,4987
0,1	0,0398	0,7	0,2580	1,3	0,4032	1,9	0,4713	2,5	0,4938		
0,2	0,0793	0,8	0,2881	1,4	0,4192	2,0	0,4773	2,6	0,4953		
0,3	0,1179	0,9	0,3159	1,5	0,4332	2,1	0,4821	2,7	0,4965		
0,4	0,1554	1,0	0,3413	1,6	0,4452	2,2	0,48,61	2,8	0,4974		
0,5	0,1915	1,1	0,3643	1,7	0,4554	2,3	0,4893	2,9	0,4981		

Tabelle 5

Es gibt verschiedene Tabellendarstellungen, die unterschiedliche Zahlen enthalten aber bei richtiger Anwendung alle zum gleichen Ergebnis führen. Wichtig ist, den Tabellenaufbau zu kennen und die Tabelle dann richtig anzuwenden.

Die vorliegende Tabelle gibt den prozentualen Anteil der Flächen an, der zwischen 0 und einem z-Wert (zwischen 0 und 3 z) liegt. Für z = 1,7 finden wir 0,4554 = 45,54%. Nun müssen wir überlegen: Im negativen Bereich, links von 0 liegen 50% der Fläche. Von 0 bis 1,7 z liegen 45,54%. Das

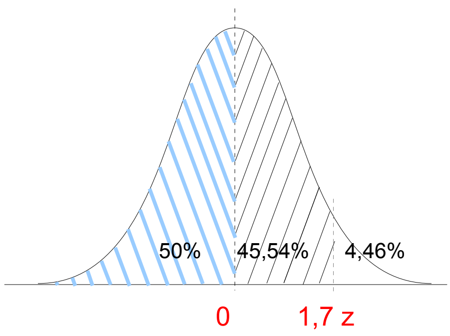

Abb.17

bedeutet: links von z = 1,7 liegen insgesamt 50% + 45,54% = 95,54%der Gesamtfläche (Abb.17). Rechts von z = 1 liegen demnach 100% - 95,54% = 4,46%. Ein kürzerer Weg: Die Fläche unter der rechten Kurvenhälfte entspricht 50% der Gesamtfläche. 45,54% liegen zwischen 0 und 1,7 z. Die Differenz ist 4,46%. Das bedeutet: Ein beliebiger Erythrozyt hat mit 4,46%iger Wahrscheinlichkeit einen Durchmesser von > 9 µm.

Beispiel 7 Mit welcher Wahrscheinlichkeit ist der Durchmesser eines beliebigen Erythrozyten <8,5 µm?

1. Umrechnen von 8,5 µm in einen z-Wert. $z = (x - \mu)/\sigma$ --> z = (8,5 - 7,69)/0,77 = 1,05
2. Ermittlung des p-Wertes für z = 1,05: Interpoliert über Kurve erhalten wir p = 0,354
3. Die Fläche links von 0 entspricht 50%
4. Dazu kommen von 0 bis z = 1,05: 35,4%
5. Links von z = 1,05 (8,5 µm) liegen also 85,4% aller Erythrozytendurchmesser. Das bedeutet: Ein beliebiger Erythrozyt hat also mit 85,4%iger Wahrscheinlichkeit einen Durchmesser < 8,5 µm.

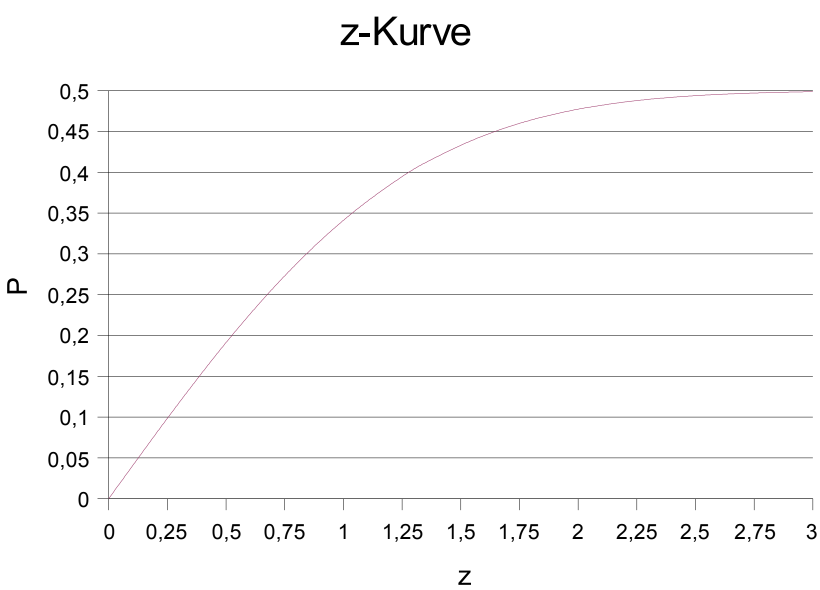

Abb.18

Beispiel 8 Mit welcher Wahrscheinlichkeit hat ein beliebiger Erythrozyt einen Durchmesser zwischen 6,5 µm und 9,0 µm?

1. Die gesuchte Wahrscheinlichkeit entspricht der Fläche unter der Kurve zwischen 6,5 und 9,0 µm.
2. Ermittlung von z für 6,5: z = (x – µ)/s --> z = (6,5 - 7,69)/0,77 = -1,55 (minus ignorieren)
3. Ermittlung von z für 9,0: z = (x – µ)/s --> z = (9,0 - 7,69)/0,77 = 1,70
4. Fläche von 0 bis - 1,55 : p = 0,44.
5. Fläche von 0 bis 1,70: p = 0,4554
6. Die Gesamtfläche setzt sich zusammen aus den beiden Teilflächen links und rechts von 0. Summe der beiden p-Werte: 0,895 (Abb.19)
7. Mit 89,5%iger Wahrscheinlichkeit hat ein beliebiger Erythrozyt einen Durchmesser zwischen 6,5 und 9,0 µm.

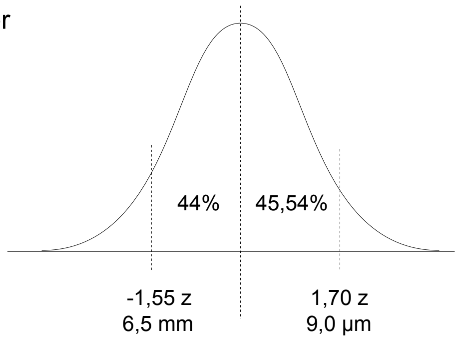

Abb.19

7.6 Übungen

Übung 1
Bei Mäusen ist eine Infektion mit Trypanosoma brucei unbehandelt tödlich. Nach Behandlung mit der Substanz X in der Dosierung 75 mg/kg s.c. ist die übliche Überlebensrate 98%. Wie wahrscheinlich ist es, dass von 10 behandelten Tieren alle 10 überleben?

Übung 2
Die Gewichte der Eier einer Hühnerpopulation sind mit N(64,8;5,9) normalverteilt. (Kapitel 3)
Wie groß ist die Wahrscheinlichkeit dafür, dass ein beliebiges Ei aus der Population >70 g wiegt? Gesucht ist P(x > 70 g)

Übung 3
Wie groß ist die Wahrscheinlichkeit dafür, dass ein Ei zwischen 55 g und 60 g wiegt?
N(64,8;5,9) : P (55 < x < 60)

Übung 4
Von einem Diagnoseverfahren ist bekannt, dass in 5% aller Fälle ein falsch positives Ergebnis auftritt.
Berechnen Sie die Wahrscheinlichkeit dafür, dass bei 20 Diagnosen zwei falsche Entscheidungen auftreten.

Übung 5
Über Jahre gemittelt ist der Männeranteil der Auszubildenden in einem Beruf im 1. Ausbildungsjahr 12%.
Wie groß ist die Wahrscheinlichkeit dafür, dass bei der nächsten Einstellung von 26 Personen 3 Männer dabei sind?

Lösung Übung 1
Zur Berechnung über die Binominalverteilung-Gleichung $P_{n;k} = C_{n;k} * p^k * (1-p)^{n-k}$ benötigen wir

$$C_{n;k} = \binom{n}{k}$$

n = Anzahl der Versuche, hier 10
k = Anzahl der Erfolgsergebnisse, hier 10 Tiere überleben
p = Wahrscheinlichkeit für das Auftreten des Erfolgs bei einem Versuch, p = 0,98

Dann gilt

$$P_{n;k} = \frac{n!}{k!\,(n-k)!} * p^k * (1-p)^{n-k}$$

$$P_{n;k} = \frac{10!}{10! * (10-10)!} * 0{,}098^{10} * 0{,}02^{0}$$

$$P_{n;k} = 1 * 0{,}8171 * 1$$

$$P_{n;k} = 0{,}8171$$

Mit 81,7% iger Wahrscheinlichkeit ist zu erwarten, dass alle 10 Tiere überleben.

Lösung Übung 2: 1. Wie groß ist der Flächenanteil unter der Kurve rechts von 70? 2. z = (70 - 64,8)/5,9 = 0,8814. Da die z-Tabelle nur Zahlen mit einer Nachkommastelle enthält, runden wir auf 0,88 und interpolieren. 3. Nach der Kurve entspricht z = 0,88 dem Wert 0,31. 4. Links von 0 liegt 50%, rechts von 0 bis z = 088 liegen 31%. 5. Rechts von 70 g liegen also 100% - 81% = 19%. Lösung Übung 3: 1. Berechnung von z für 55 g => (55-64,8)/5,9 = -1,66 z => runden zu -1,7 z => Tabelle => P = 0,4554. Das ninus vor 1,66 z ignorieren wir, da die Kurve symmetrisch ist und die Tabelle nur positive Werte enthält. 2. Zwischen 55 g (1,77 z) und 0 liegen also ca. 45% der Fläche. 3. Berechnen von z für 60 g führt zu 0,81 z => ca. 29% der Fläche. 4. Die Differenz zwischen 55 g (1,77 z) und 60 g (0,81 z) entspricht demnach 45% - 29% = 16% (gerundet). 5. Ein beliebiges Ein liegt also mi 15%iger Wahrscheinlichkeit zwischen 55 und 60 g.

Lösung Übung 4
n = 20; k = 2; p = 0,05

$$P_{n;k} = \binom{n}{k} * p^k * (1-p)^{n-k}$$

$$P_{n;k} = \frac{20!}{2! * 18!} * 0{,}05^2 * 0{,}95^{18}$$

$$P_{n;k} = 190 * 0{,}0025 * 0{,}3972 = \underline{0{,}1886}$$

Lösung Übung 5
n= 26; k = 3; p = 0,12

$$P_{26,3} = \frac{26!}{3! * 23!} * 0{,}12^3 * 0{,}88^{23} = 2600 * 0{,}001728 * 0{,}05286 = \underline{0{,}2374}$$