8. Prüfung auf Normalverteilung

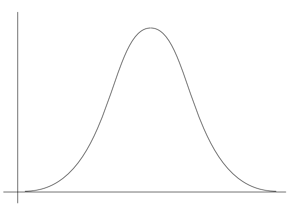
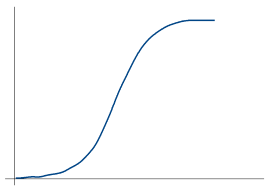
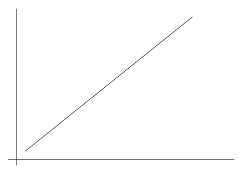
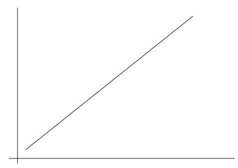
8.1 Warum Prüfung aus Normalverteilung?

Eine häufige Aufgabe bei naturwissenschaftlichen Untersuchungen besteht in der Feststellung, ob sich zwei (oder mehr) Datengruppen mehr als es durch den Einfluß des Zufalls zu erwarten wäre, unterscheiden. Bei einem Analgesietest reagierten die Mäuse der Kontrollgruppe auf einen Schmerzreiz nach durchschnittlich 18,7 s. Die mit einem Präparat behandelten Tiere zeigen 1 Stunde nach Applikation eine mittlere Reaktionszeit von 38,6 s. Das weist auf eine analgetische Wirkung hin, da die behandelten Tiere den Schmerzreiz länger tolerierten als die der Kontrollgruppe. Die Zeitdifferenz ist aber kein Beweis für eine Wirkung des Präparates, denn bei zwei Stichproben können wir durchaus zufallsbedingt unterschiedliche Mittelwerte erwarten. Die Frage ist nun ob die Zeitdifferenz zufällig oder auf die Präparatewirkung zurückzuführen ist. Ist die Differenz klein, dann schreiben wir das eher dem Zufall zu. Ist sie dagegen groß, dann neigen wir eher zu der Annahme, sie sei nicht durch den Zufall zu erklären sondern beruhe auf der Wirkung des Präparates. Wo ist aber die Grenze zwischen einer großen Differenz und einer kleinen? Zur Beantwortung dieser Frage stellen wir folgende Hypothese auf: „Die Differenz ist zufällig. Das Präparat hat keine analgetische Wirkung". Zur Prüfung der Hypothese können wir die Daten dem so genannten t-Test unterziehen. Dieser Test setzt aber voraus, dass die zu untersuchenden Daten normalverteilt sind, zumindest approximiert. Vor Anwendung des Tests müssen wir daher prüfen, ob diese Voraussetzung gegeben ist. Es gibt also einen guten Grund, sich mit der Prüfung auf Normalverteilung zu beschäftigen. [Genaueres zum Thema Hypothesenprüfung und t-Test folgt in späteren Kapiteln.]

8.2 Prüfung auf Normalverteilung

Hierzu gibt es verschiedene Verfahren. Wir wollen hier ein graphisches Schätzverfahren kennen lernen, mit dem wir relativ schnell und einfach prüfen können, ob empirische Datensätze approximiert normalverteilt sind. Dazu werden für die zu prüfenden Daten die Summenprozenthäufigkeiten berechnet und gegen die oberen Klassengrenzen in ein spezielles Koordinatensystem (Wahrscheinlichkeitsskala oder Probitskala auf der Abszisse) eingetragen. Zeigt die Punktfolge eine Tendenz zur Linearität, lässt sich ihr also eine Gerade anpassen, die möglichst nahe an allen Punkten liegt, dann können wir davon ausgehen, dass die Daten zumindest approximiert normalverteilt sind. Dies gilt um so eher, je besser die Anpassung ist.

Vor der eigentliche Prüfung (Punkt 8.6 und 8.7) müssen wir mit den folgenden Schritten einige Vorarbeiten leisten, die zum Verständnis des Verfahrens führen. Dazu werden wir die Daten von Beispiel 1 so

Schritt 1 Erstellung einer absoluten und einer relativen Häufigkeitskurve	Schritt 2 Erstellung einer sigmoiden Summenprozent-häufigkeitskurve	Schritt 3 Erstellung einer Geraden im Wahrscheinlichkeits-netz	Schritt 4 Erstellung einer Geraden im Probitnetz
 Abb.1	 Abb.2	 Abb.3	 Abb.4

transformieren, dass aus der Glockenkurve eine sigmoide Kurve (Ogive) wird und aus dieser zwei Geraden entstehen. Mit einer dieser beiden Geraden können wir dann die eigentliche Schätzung durchführen. Der Begriff sigmoid ist abgeleitet von dem kleinen griechischen Buchstaben sigma (σ) welcher am Wortende ähnlich dem kleinen lateinischen s geschrieben wird.

Beispiel 1 Im mikroskopischen Praktikum wurden die Durchmesser von 369 Exemplaren des marinen Dinoflagellaten Noctiluca miliaris in µm bestimmt. [Inkrement 1 µm]. Der range betrug 200 µm bis 640 µm. Die Daten wurden zur weiteren Bearbeitung mit einer Klassenbreite von 20 µm gruppiert. Die Ergebnisse finden Sie in den Spalten 1 bis 3 der Tabelle 1, $\bar{x}$ = 460,6 µm, s_x = 73,4 µm. Frage: Sind die Durchmesser normalverteilt? Die Antwort könnte dann von Bedeutung sein, wenn wir z.B. die Durchmesser einer anderen Population von Noctiluca m. mit den Werten unserer Population vergleichen möchten.

1	2	3	4	1	2	3	4
Klasse	Klassen-mitte	absolute Häufigkeit H	relative Häufigkeit h (%H)	Klasse	Klassen-mitte	absolute Häufigkeit H	relative Häufigkeit h (%H)
1	200	1	0,271	14	460	46	12,470
2	220	1	0,271	15	480	39	10,570
3	240	2	0,542	16	500	37	10,030
4	260	0	0,000	17	520	29	7,860
5	280	4	1,084	18	540	21	5,690
6	300	3	0,813	19	560	14	3,790
7	320	5	1,350	20	580	11	2,980
8	340	10	2,710	21	600	7	1,900
9	360	11	2,980	22	620	5	1,360
10	380	18	4,880	23	640	2	0,542
11	400	23	6,230	24	660	0	0
12	420	37	10,030	25	680	0	0
13	440	43	11,650	26	700	0	0

Tabelle 1

8.3 Schritt 1
Erstellung der Glockenkurve (absolute Häufigkeiten gegen die Klassenmitten)

Wir berechnen mit $\bar{x}$ = 460,6 µm, s_x = 73,4 µm die Wahrscheinlichkeitsdichte für die Daten. Die große Stichprobe gestattet uns, an Stelle der Parameter µ und σ deren Schätzwerte einzusetzen. So entsteht die berechnete Häufigkeitsverteilungskurve (Abb.5, grün), in die wir die absoluten Häufigkeiten (orange) eingezeichnet haben.

Die linke Ordinate trägt die absolute Häufigkeit und die rechte die f(x)-Werte. Da die Punkteschar der alsoluten Häufigkeiten der Form der Glockenkurve recht gut folgt, ist dies schon ein erster Hinweis auf normalverteilte Daten. Da es aber einfacher ist, einer Punktfolge mit linearer Tendenz eine Gerade anzupassen, fahren wir fort auf dem Wege zu eben dieser Geraden und verwenden dazu die f(x)-Werte.

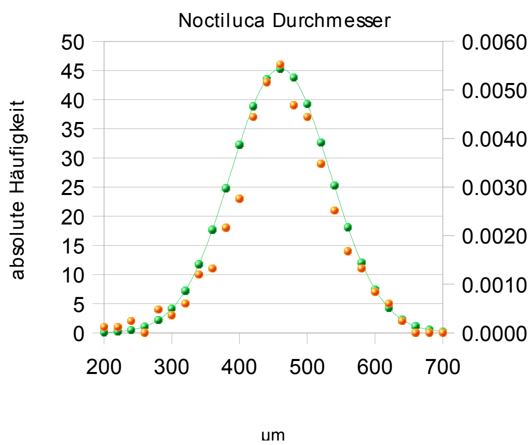


Abb. 5

8.4 Erstellung der Glockenkurve (relative Häufigkeiten gegen Klassenmitten)

Unser Ziel ist es, die Daten so zu bearbeiten, dass eine Punktfolge mit linearer Tendenz resultiert. Im ersten Schritt zu dieser Transformation der Daten werden aus den absoluten Häufigkeiten die relativen (prozentualen) Häufigkeiten berechnet.

In Spalte 4 der Tabelle 1 sind die nach

$$h = \frac{H}{n}$$

h = relative Häufigkeiten
H = absolute Häufigkeiten
n = ∑H

für jede Klassenmitte berechneten relativen Häufigkeiten h rot eingetragen.

Wenn wir die relativen Häufigkeiten gegen die Klassenmitten graphisch darstellen, dann resultiert die gleiche Verteilungskurve wie die grüne Kurve in Abb.5, nur dass die Ordinate in Abb.6 nun die relativen statt die absoluten Häufigkeiten trägt. Wir benötigen die relativen Häufigkeiten für die Berechnungen der Summenprozenthäufigkeiten in Anschnitt 8.5.

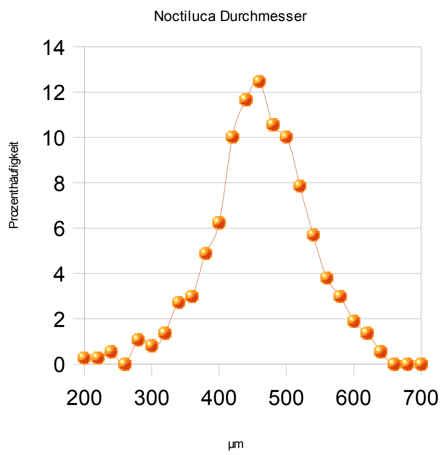


Abb.6

Die relative Häufigkeit beim Vergleich von Datengruppen

Wenn zum Vergleich mehrerer Datengruppen die absoluten Häufigkeiten gegenüber gestellt werden, dann erschweren die unterschiedlichen Ordinatenteilungen die Vergleichbarkeit. Günstiger ist die Gegenüberstellung der relativen Häufigkeitskurven, da hier bei allen Gruppen die Ordinaten gleich sind.

Bei Prozenthäufigkeiten immer n angeben!

Die Angabe von relativen Häufigkeiten birgt eine Gefahr der Fehlinterpretation auf Grund der Nichtbeachtung der Stichprobengröße. Die Aussagekraft einer Stichprobe hängt u.a. von deren Umfang ab. Relative Häufigkeiten enthalten aber keinerlei Information dazu, wie groß die Stichprobe war. Es ist deswegen zwingend notwendig, bei der Angabe von Prozentwerten immer anzugeben, welchem Stichprobenumfang sie entsprechen.

8.5 Schritt 2
Erstellung der sigmoiden Kurve (∑%H gegen die oberen Klassengrenzen)

Der nächste Schritt zur Geraden führt zu einer sigmoiden Kurve, die dann entsteht, wenn wir auf der Ordinate statt der relativen Häufigkeiten die Summenprozenthäufigkeiten (∑%H = kumulierte relative Häufigkeiten) auftragen. Dazu müssen wir für jede Klasse die ∑%H ermitteln. Dies geschieht nach folgender

Überlegung. In die 1. Klasse gehören 0,271% aller Messwerte. Dies ist auch die ∑%H der ersten Klasse. Unter der ∑%H der 2. Klasse verstehen wir die Summe aller Prozenthäufigkeiten die in die 1. und in die 2. Klasse bis zu deren oberer Grenze gehören. Das sind 0,271% + 0,271% = 0,542%.

Die ∑%H der Klasse i entspricht der Summe aller relativer Häufigkeiten, von Klasse i=1 bis zur oberen Grenze der Klasse i.

$$\sum_{i=1}^{n} \%H = \%H_1 + \%H_2 + \%H_3 + \ldots + \%H_n$$

Alle so berechneten Summenprozenthäufigkeiten sind in Spalte 3 der Tabelle 2 eingetragen.

1	2	3	1	2	3
obere Klassengrenze	relative Häufigkeit h (%H)	Summenprozent-häufigkeit	obere Klassengrenze	relative Häufigkeit h (%H)	Summenprozent-häufigkeit
<210	0,271	0,271	<470	12,470	55,281
<230	0,271	0,542	<490	10,570	65,851
<250	0,542	1,084	<510	10,030	75,881
<270	0,000	1,084	<530	7,860	83,741
<290	1,084	2,168	<550	5,690	89,431
<310	0,813	2,981	<570	3,790	93,221
<330	1,350	4,331	<590	2,980	96,201
<350	2,710	7,041	<610	1,900	98,101
<370	2,980	10,021	<630	1,360	99,461
<390	4,880	14,901	<650	0,542	100
<410	6,230	21,131	<670	0	100
<430	10,030	31,161	<690	0	100
<450	11,650	42,811	<710	0	100

Tabelle 2

Bei der graphischen Darstellung der Summenprozenthäufigkeitskurve trägt die Ordinate die kumulierten Prozentzahlen von 0 bis 100%. Die Abszisse trägt die Klassengrenzen. Da die untere Grenze einer Klasse mit der oberen Grenze der vorhergehenden Klasse identisch ist, entfällt bei der Skalenbeschriftung an den oberen Klassengrenzen das Zeichen „kleiner als“.

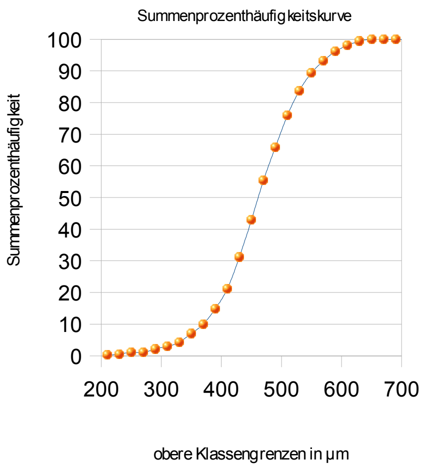


Abb.7

Während bei den Graphiken zur absoluten und relativen Häufigkeit die Funktionswerte immer über den Klassenmitten aufgetragen werden, müssen wir die Summenprozentwerte immer über der oberen Klassengrenze auftragen, weil die Summenprozentwerte ja allen Messwerten der Klassen bis zur oberen Grenze entsprechen und nicht nur bis zu der Klassenmitte. Es entsteht die S-förmige Kurve (Sigmoid, Ogive) in Abb. 7.

Je symmetrischer die Glockenkurve war, um so symmetrischer ist auch die sigmoide Kurve. Unsere Kurve weist damit auch schon auf eine Normalverteilung hin. Nur im Bereich zwischen 200 µm und 300 µm erkennen wir kleine Unregelmäßigkeiten, die sich in Abb.6 schon andeuteten.

Schätzung des Mittelwertes und des Streubereichs an der Ogive

Innerhalb des Bereiches $\bar{x} \pm s_x$ liegen bei einer Normalverteilung ca. 68% aller Werte. An der Ogive können wir – Normalverteilung vorausgesetzt - durch Interpolation den Mittelwert bei 50%, die untere einfache Streugrenze bei 16% und die obere einfache Streugrenze bei 84% ablesen. Da die Ogive bei 16% und 84% nicht linear ist, erhalten wir nur Schätzwerte.

8.6 Schritt 3
Erstellung der Geraden (∑%H an der Wahrscheinlichkeitsskala gegen die oberen Klassengrenzen)

Die Wahrscheinlichkeitsskala

Um mit den gleichen Daten, die die sigmoide Kurve lieferten, eine Gerade zu erhalten, müssen wir die Ordinatenskalierung ändern. Anstelle der metrisch geteilten Prozentskala benötigen wir die so genannte Wahrscheinlichkeitsskala, die nach dem Gaußschen Integral $\int_{-\infty}^{\infty} f(x)dx$ geteilt ist. Die Berechnung der Skalenteilung über die Integralrechnung ist aufwendig und wird daher in der Praxis nicht durchgeführt. Üblicherweise verwenden wir für die Darstellung so genanntes Wahrscheinlichkeitspapier welches der Handel anbietet (Schleicher und Schüll).

Die Ordinate (Abb.8) zeigt die Wahrscheinlichkeitsskala. Sie ist nicht äquidistant geteilt. Sie sehen, dass die Differenzen zweier benachbarter Zahlen nicht gleich sind (99,9 – 97,8 = 2,1; 97,8 – 84 = 13,8; 84 – 50 = 34), die den Zahlen entsprechenden Strecken aber wohl. Auch die Zwischenräume sind wie Abb.9 zeigt, wieder inäquidistant geteilt. Der Mittelpunkt entspricht dem Summenprozentwert 50%. Nach oben und unten wird die Skala symmetrisch immer stärker gespreizt, so dass 0% und 100% im Unendlichen liegen. Der Grund dafür liegt letztlich an dem Integral $\int_{-\infty}^{\infty} f(x)dx$ und damit am asymptotischen Verlauf der Glockenkurve.

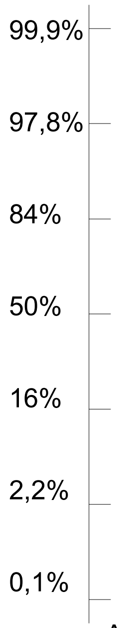


Abb.8

Die Abb.9 trägt auf der Ordinate die Wahrscheinlichkeitsskala, die Abszisse trägt die oberen Klassengrenzen, zu denen wir die entsprechenden ∑%H-Werte eingetragen haben. Die Punktfolge zeigt eine Tendenz zur Linearität, was auf normalverteilte Daten hinweist.

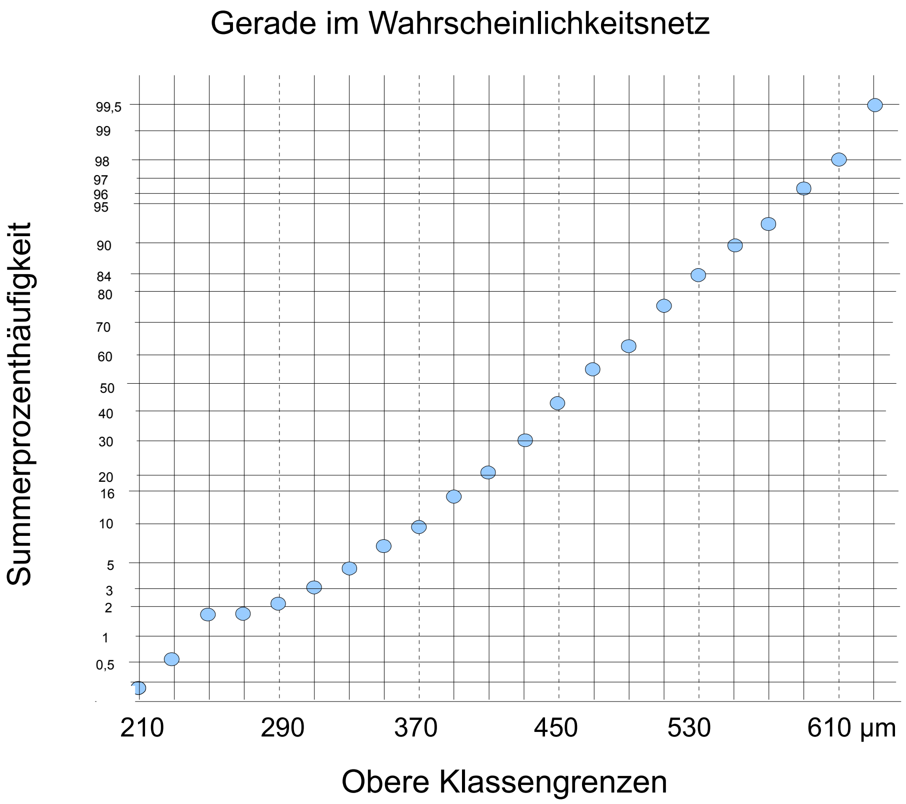


Abb.9

Die eigentliche Prüfung auf Normalverteilung
Um der Punktfolge in Abb.9 eine Gerade anzupassen legen wir zunächst eine Gerade nach Augenmaß. Dabei ist zu beachten, dass die Summe der parallel zur Ordinate gemessenen Abweichungen der Punkte oberhalb der Geraden gleich der Summe unterhalb der Geraden ist und dass die Summen minimiert sind. Die folgende Abb.10 zeigt das Prinzip.

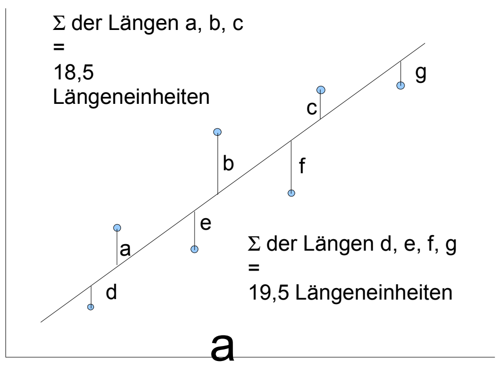


Abb.10

Optimal liegt die Gerade in Abb.10 offensichtlich noch nicht, da die untere Summe größer als die obere ist. Die Anpassung der Geraden muss u.U. mehrmals nachgebessert werden. Dazu könnten wir die eingezeichnete Gerade in geeignet scheinender Weise etwas verschieben. Und dies so oft, bis die oben genannte Forderung erfüllt ist. Beim Thema Regression lernen wir die Lage der Geraden zu berechnen.

Unsere Punktfolge in Abb.9 weicht nur in unteren Bereich deutlich von der Linearität ab. Mit Hilfe einer angepassten Geraden können wir relativ schnell erkennen, dass zumindest eine approximierte Normalverteilung der Daten vorliegt. Je besser die Anpassung ist, um so eher gehen wir von einer Normalverteilung aus. Das beschriebene Verfahren ist daran gebunden, dass das handelsübliche Wahrscheinlichkeitspapier vorliegt. Wie wir selber ein Koordinatensystem erstellen können, mit dem wir diese Gerade erhalten, das zeigt Schritt 4.

8.7 Schritt 4
Erstellung einer Geraden (Probits an der Probitskala gegen die oberen Klassengrenzen)

Die Probiskala

Den gleichen Effekt, nämlich die Linearisierung der Ogive erreichen wir, wenn wir anstelle der Wahrscheinlichkeitsskala die so genannte Probitskala anwenden. Diese hat den Vorteil, dass wir sie im Gegensatz zur Wahrscheinlichkeitsskala selber erstellen können. Es ist eine Skala, die in Wahrscheinlichkeitseinheiten (<u>prob</u>ability un<u>its</u> = Probits) äquidistant geteilt ist. Sie beginnt in der Regel mit Probit 2 und endet mit Probit 8 wobei diese Grenzen abhängig von den Versuchsdaten sind. In der Mitte der Ordinate liegt immer Probit 5. An der Probitskala, die ja keine Prozentskala ist, können wir allerdings keine ∑ %H eintragen. Daher müssen wir diese mit Hilfe der so genannten Probittabelle in Probits transformieren. Wir zeigen mit Tabelle 3 einen Ausschnitt aus der umfangreichen Probittabelle, die in einschlägigen Tabellenbüchern (z.B. Wissenschaftliche Tabellen Geigy) zu finden ist.

Summenprozentwerte (Zehntel)

Ausschnitt aus der Probittabelle
(Aus F.Keller, Statistik für naturwissenschaftliche Berufe, pmi-Verlag, siehe Literaturverzeichnis)

	0	0.1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
0		*1,91*	*2,12*	***2,25***	*2,35*	*2,42*	*2,49*	*2,54*	*2,59*	*2,63*
1	*2,67*	***2,71***	*2,74*	*2,77*	*2,80*	*2,83*	*2,86*	*2,88*	*2,90*	*2,93*
2	*2,95*	*2,97*	*2,99*	*3,00*	*3,02*	*3,04*	*3,06*	*3,07*	*3,09*	*3,10*
3	*3,12*	*3,13*	*3,15*	*3,16*	*3,18*	*3,19*	*3,20*	*3,21*	*3,23*	*3,24*
4	*3,25*	*3,26*	*3,27*	*3,28*	*3,29*	*3,30*	*3,32*	*3,33*	*3,34*	*3,35*
5	*3,36*	*3,36*	*3,37*	*3,38*	*3,39*	*3,41*	*3,41*	*3,42*	*3,43*	*3,44*

Tabelle 3

Probits (kursiv)

Summenprozentwerte (ganze Zahlen)

Probit für Summenprozentwert 0,271% runden zu 0,3% ---> Probit 2,25
0,542% runden zu 0,5% ---> Probit 2,42
1,084% runden zu 1,1% ---> Probit 2,71

Weitere Werte müssen einer Originaltabelle entnommen werden, die wir hier nicht darstellen können. Mit Hilfe der Abb.11, die der Probittabelle entspricht, können wir die Werte auch durch graphische Interpolation gewinnen. Errichten Sie die Vertikale über z.B. 40% (∑%H = 40) und projizieren Sie deren Schnittpunkt mit der Kurve horizontal auf die Ordinate. Dort finden Sie den Probitwert für 40% nämlich 4,75 Probis. Beachten Sie, dass graphische Interpolationen je nach Größe der Graphik oft nur Näherungswerte ergeben.

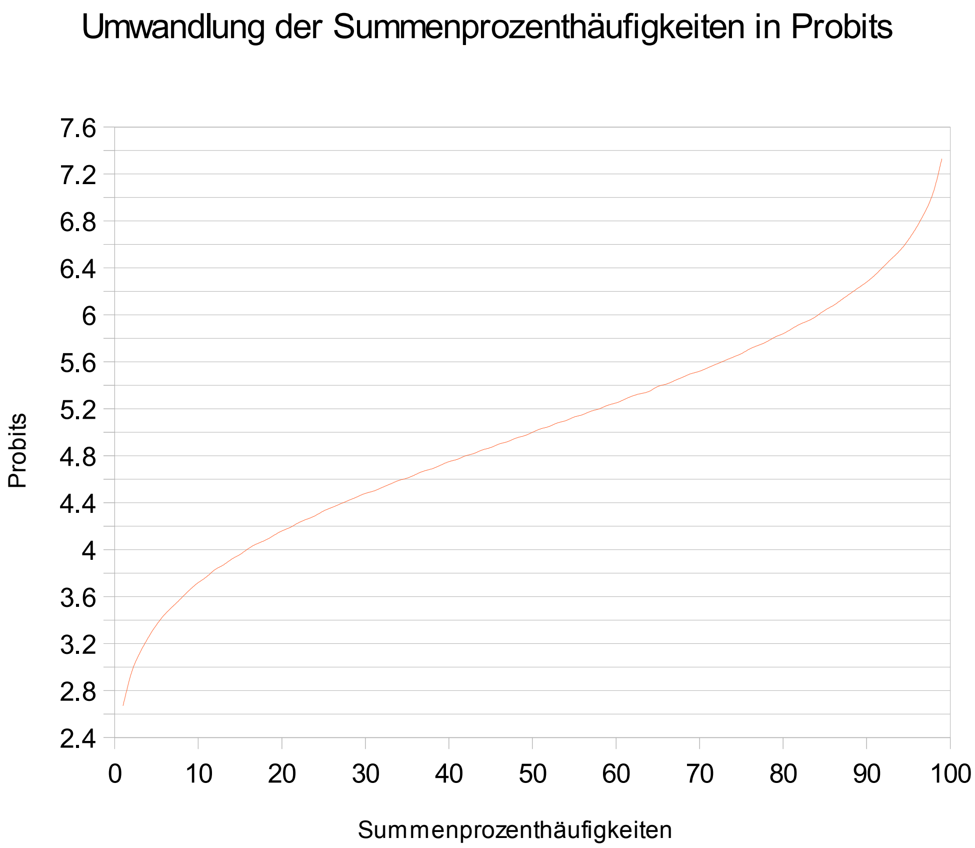


Abb.11

Die Tabelle 4 enthält bereits alle über eine Tabelle ermittelten Probits.

Klasse	obere Klassen-grenze	Summenprozent-häufigkeit	Probits	Klasse	obere Klassen-grenze	Summenprozent-häufigkeit	Probits
1	<210	0,271	-	14	<470	55,281	5,13
2	<230	0,542	-	15	<490	65,851	5,41
3	<250	1,084	2,7	16	<510	75,881	5,70
4	<270	1,084	2,7	17	<530	83,741	5,98
5	<290	2,168	2,9	18	<550	89,431	6,25
6	<310	2,981	3,1	19	<570	93,221	6,49
7	<330	4,331	3,3	20	<590	96,201	6,77
8	<350	7,041	3,5	21	<610	98,101	7,04
9	<370	10,021	3,72	22	<630	99,461	7,58
10	<390	14,901	3,96	23	<650	100	
11	<410	21,131	4,2	24	<670	100	
12	<430	31,161	4,5	25	<690	100	
13	<450	42,811	4,82	26	<710	100	

Tabelle 4

Wenn alle Probits über den oberen Klassengrenzen aufgetragen sind, dann erhalten wir die gleiche Punkteschar wie bei der Wahrscheinlichkeitsskala. Die Abb.12 wurde mit einem Tabellenkalkulationssystem erstellt, welches die Lage der Geraden bereits eingefügt hat. Die Gerade ist den Punkten so gut angepasst, dass wir davon ausgehen, dass die Durchmesser von Noctiluca m. zumindest approximiert normalverteilt sind.

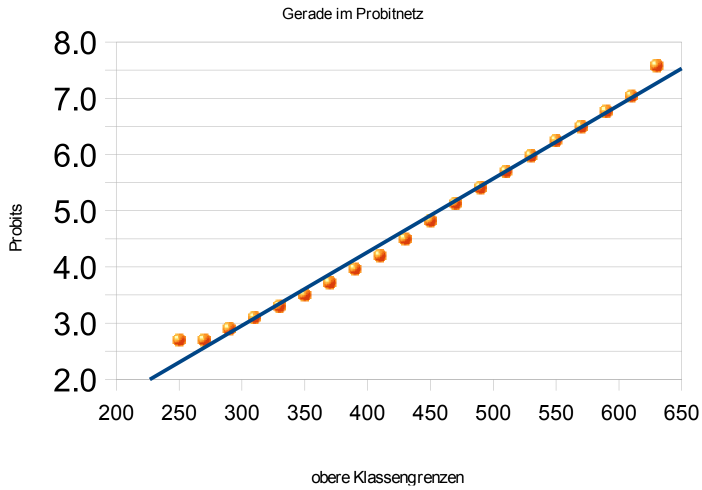


Abb.12

Ob wir die Prüfung auf Normalverteilung mit der Wahrscheinlichkeitsskala oder mit der Probitskala durchführen ist einerlei. Die Punktfolgen sind identisch. Wenn wir Daten auf Normalverteilung prüfen wollen, stellen wir nicht, wie oben beschrieben, die Glockenkurve und dann die Ogive dar, sondern erstellen nach Berechnung der relativen Häufigkeiten und der Summenprozenthäufigkeiten gleich die Gerade an der Wahrscheinlichkeits- oder Probitskala und entscheiden dann. Der Weg über die Glockenkurve war nur eine didaktische Hilfskonstruktion.

8.8 Weitere Anwendungen der Geraden

Geraden haben neben ihrer Funktion bei der Prüfung auf Normalverteilung noch andere nützliche Eigenschaften.

8.8.1 Graphische Ermittlung von $\bar{x}$ und s_x.

Mit Hilfe der Geraden können wir ohne weitere Berechnung graphisch das arithmetische Mittel und die Standardabweichung schätzen. Nach der Probittabelle gilt:

Σ%H 16% --> Probit 4
Σ%H 50% --> Probit 5
Σ%H 84% --> Probit 6

Projizieren wir Probit 5 horizontal auf die Gerade und den Schnittpunkt vertikal auf die Abszisse, so wird dort $\bar{x}$ angezeigt. Projizieren wir Probit 4 auf die Gerade, so zeigt der Schnittpunkt auf der Abszisse die untere Grenze des einfachen Streubereichs an. Mit Probit 6 erhalten wir dessen obere Grenze.

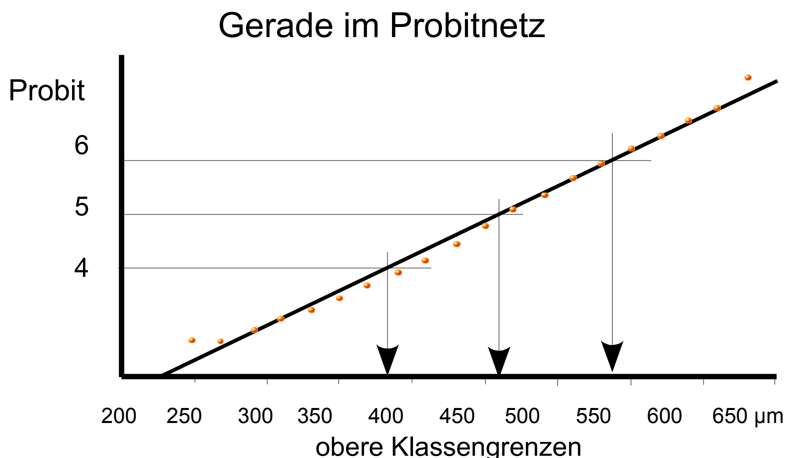


Abb.13

In der Abb.13 finden wir für $\bar{x}$ ~ 458 µm (berechnet: 460,6 µm) und für $\bar{x}+s_x$ ~ 535 µm , für $\bar{x}-s_x$ ~ 383 µm. Das entspricht einer Standardabweichung von ~76,5 µm (berechnet: 73,4 µm). Das sind schon bessere Werte als die an der Ogive ermittelten.

8.8.2 Graphischer Vergleich der Streuungen mehrerer Datensätze.

Beispiel 2 Nehmen wir an, uns lägen die Kornmassen (in g/L) zweier Ernten eines Saatgutes vor und wir wollten wissen, ob die Werte in den beiden Gruppen unterschiedlich stark streuen.
[Rechnerisch ist das über den Variationskoeffizienten (Kapitel 5) prüfbar.]

Eine vergleichende Visualisierung der Streuung erreichen wir, wenn die Summenprozenthäufigkeitskurven mehrerer Gruppen in ein Probitnetz eingezeichnet werden. Der Zusammenhang zwischen der Steilheit der Geraden und der Streuung der Daten ist sofort zu erkennen. Wie Abb.14 zeigt, ist die Gerade um so steiler, je geringer die Messwerte streuen. In Gruppe 1 (blau) ist s_x = 30 g/L, in Gruppe 2 (rot) ist s_x = 13 g/L. Im Beispiel sind die Mittelwerte gleich. Das ist aber für einen Vergleich aber nicht notwendig. Allerdings muss die Abszissenspreizung für alle Gruppen gleich sein. Durch graphische Interpolation finden wir auch schnell das jeweilige arithmetische Mittel.

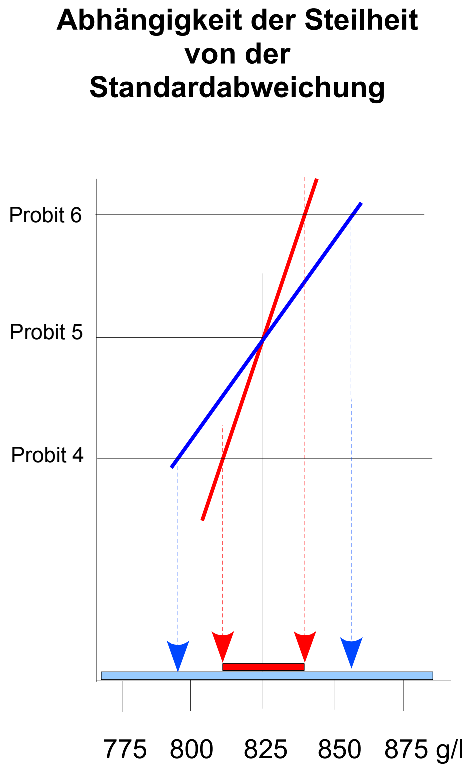


Abb.14

Beispiel 3: Im parasitologischen Praktikum haben wir bei 36 männlichen Spulwürmern (Ascaris suum) die Länge der Tiere gemessen. Die Ergebnisse in mm:

213	192	168	211	246	189	201	194	201	166	129	196
176	192	199	192	141	152	217	161	221	222	182	222
194	187	179	197	212	215	232	176	198	192	214	181

Wir wollen über die Gerade im Probitnetz feststellen, ob die Längen normalverteilt sind. Dazu haben wir die Daten in fünf Gruppen mit der Breite 26 mm klassiert und nach den bekannten Verfahren die Daten in Tabelle 5 errechnet.

Gruppe	Mittelwert	Obere Klassengrenze	absolute Häufigkeit	Prozenthäufigkeit	Summenprozent häufigkeit	Probits
1	133	<146	2	5,56	5,56	3,4
2	159	<172	4	11,11	16,67	4,03
3	185	<198	15	41,67	58,33	5,2
4	211	<224	13	36,11	94,44	6,59
5	237	<250	2	5,56	100	

Tabelle 5

Versuchen Sie, den Punkten in Abb.15 eine Gerade optimal anzupassen und bilden Sie sich dann eine Meinung zur Normalverteilung der Daten.

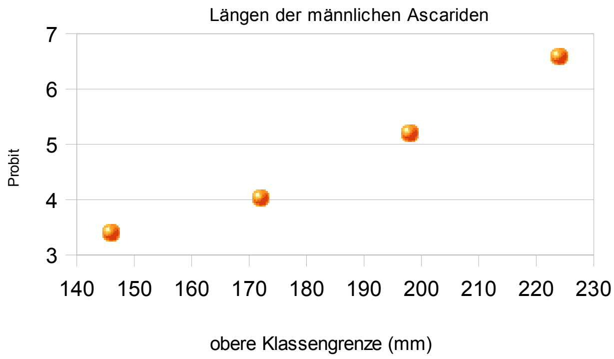


Abb.15

Übungen

Übung 1 Im Zusammenhang mit der Untersuchung in Beispiel 3 haben wir bei 46 weiblichen Ascariden die Längen in mm gemessen. Die Messwerte liegen in der folgenden Liste vor. Prüfen Sie mit einer Geraden im Probitnetz auf Hinweise zur Normalverteilung.

275	263	248	298	283	262	269	254	266	289
262	265	269	257	309	284	262	191	316	298
262	269	266	244	303	283	291	249	274	339
237	300	299	303	284	242	231	262	248	262
292	292	271	307	286	266				

Lösung:
Die Abbildung weist auf eine deutliche lineare Tendenz hin. Wir betrachten die Längen als zumindest approximiert normalverteilt.

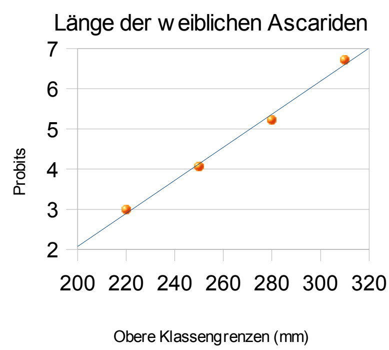


Abb.16