

Von Sporen, Histogrammen und Ellipsen Was man tun und lassen soll

Heinz Clémenton
Faculté de Biologie et de Médecine
Université de Lausanne

Zusammenfassung: Es wird versucht, grundlegende Konzepte statistischer Analysen ohne Mathematik zu erläutern. Einige grafische Beispiele sollen dabei helfen.

Stichwörter: Mykologie, Sporenstatistik, Methoden, Histogramm, QQ-Diagramm, Streudiagramm, Toleranzellipse, grafische Darstellung.

Einleitung

Im letzten Abschnitt ihrer Arbeit über Dimensionen in der Pilzmorphologie erwähnen Halbwachs und Karasch (2019) bei der Besprechung von Sporenmessungen kurz die Normalverteilung, Streudiagramme und Ellipsen, ohne weiter auf diese Begriffe einzugehen. Dieser Abschnitt weckte in mir alte Erinnerungen an allerhand Probleme, die manche Mykologen plagten, wenn es sich um statistische Aussagen zu ihren Sporenmessungen handelt, und die manchmal zu fragwürdigen Aussagen führen. Und diese Probleme veranlassten mich bereits in der zweiten Hälfte der 1980er Jahre einen Vortrag zum Thema «Sinn und Unsinn bei der Sporenmessung» auszuarbeiten. Die vorliegende Arbeit stützt sich auf meine alten Erinnerungen und den alten Vortrag. Ich zitiere keine fragwürdige Veröffentlichung, denn es geht mir nicht um direkte Stellungnahmen zu diesen Publikationen, sondern vielmehr um eine allgemeine und neutrale Diskussion zum Thema Sporenstatistik.

Das Messen der Sporen (mit oder ohne die Ornamentation, die Auswahl der zu messenden Sporen, in KOH oder Wasser gemessen, gerundete Messwerte, usw.) wird nicht besprochen, da nicht zum Thema gehörend. Lediglich die Normalverteilung, die grafische Darstellung der Resultate (Histogramme, QQ-Diagramme, Streudiagramme, Ellipsen) und, im Text verstreut, die Anzahl der nötigen Messungen kommen hier zur Sprache.

Im Internet kann man zum Thema Statistik alles und jedes nachlesen, meist mit mathematischen Formeln und Definitionen. Weshalb denn hier alles nochmal ausbreiten? Weil mathematische Formeln und Definitionen vermieden werden sollen, und weil die Statistik auch human dargeboten werden kann

Im Folgenden wird das gesamte Sporenpulver mit dem statistischen Ausdruck **Grundgesamtheit** bezeichnet, und die tatsächlich gemessenen Sporen bilden die **Stichprobe**. Es ist wichtig, diese beiden Begriffe klar auseinander zu halten. Aufgabe des Mykologen ist es nun, aus der Stichprobe auf die Grundgesamtheit zu schließen, was immer eine Schätzung, nie eine Gewissheit ist.

Material und Methoden

Die Beispiele beruhen auf meinen Sporenmessungen von *Conocybe velutipes*, *Gymnopus confluens* und *Pseudobaeospora calcarea*, auf den veröffentlichten Streudiagrammen von *Flammulina fennae*, *F. velutipes* und *F. elastica* (Abb. 15 im Anhang), auf den von Möls in Parmasto (1987) aufgelisteten Messungen von Sporen der *Auricularia auricula-judae* sensu lato, auf den von Fannechère (2005) gemessenen *Inocybe*-Sporen, und auf einem theoretischen "Modell 5000": Virtuelle, mit dem Computer erzeugte "Sporen" mit normalverteilten Längen und Breiten, ein "Sporenpulver", aus dem verschieden große Stichproben entnommen werden.

Pilz	n	Länge 95% µm	Breite 95% µm	Q 95%
Modell 5000	5000	9,9 - 10,6	4,1 - 5,0	2,11 - 2,40 *
<i>Conocybe velutipes</i>	1630	11,7 - 13,4	6,4 - 7,7 *	1,83 - 1,92 *
<i>Gymnopus confluens</i>	77	6,5 - 8,6	3,2 - 4,5	1,56 - 2,34 *
<i>Pseudobaeospora calcarea</i>	60	3,1 - 4,8 *	2,5 - 3,4	1,15 - 1,55
<i>Flammulina fennae</i>	14	6,1 - 7,9	3,6 - 4,8	1,53 - 1,77
<i>Flammulina velutipes</i>	19	6,6 - 9,5	3,2 - 4,4	1,84 - 2,39
<i>Flammulina elastica</i>	14	8,2 - 10,6	3,2 - 3,8	2,41 - 2,99
<i>Inocybe spec.</i>	44	9,8 - 13,0	3,9 - 5,7	2,0 - 2,8 *
<i>Auricularia auricula-judae</i> s.l.	33	12,8 - 20,9 **	–	–

* von der Normalverteilung (unbedeutend) abweichend
 ** von der Normalverteilung stark abweichend, da zweigipfelig,
 aus *Auricularia americana* und *A. auricula-judae* bestehend

Benützte Programme (Macintosh Computer)

Adobe Photoshop CC 2015 zur Fertigstellung der Abbildungen.

Free Statistics Software, Wessa, P. Office for Research Development and Education, version 1.2.1, URL <https://www.wessa.net/>

https://www.wessa.net/rwasp_fitdistrnorm.wasp - output

Für die Histogramme und QQ-Diagramme.

JMP Ein professionelles Statistikprogramm der SAS Gruppe. SAS Institute GmbH, MP Office Germany, Austria, Switzerland and Italy. Otto-Lilienthal-Straße 36, D-71034 Böblingen.

Für die Abb. 9.

Microsoft Excel für die Säulendiagramme und für die Berechnung der Toleranzellipsen nach den Documenta Geigy.

MjoGraph-4.6.1 von Makoto Tanahashi, seit 2006. <http://www.ochiailab.dnj.ynu.ac.jp/mjograph/>
 Für das Zeichnen der mit Excel berechneten Toleranzellipsen.

Statdisk v12, Release Date: January 2013. Copyright 1986-2004, Pearson Education, Inc. Copyright 2004-2016 Triola Statistics. www.statdisk.org

Für die Berechnung des Ryan-Joiner Testes und des Korrelationskoeffizienten.

Notizen zur Normalverteilung

Wenn die Messresultate grafisch dargestellt werden sollen, so stehen dem Mykologen mancherlei Möglichkeiten zur Verfügung, die aber nicht immer richtig eingesetzt werden. Und immer wieder stößt man auf den Begriff der Normalverteilung, die all diesen Messungen und Grafiken zugrunde liegen soll. Nur, was ist die Normalverteilung und wieso ist sie "normal"? Zu dieser Frage möchte ich einige Bemerkungen machen bevor von Histogrammen, QQ-Diagrammen und Ellipsen die Rede ist.

Wenn eine Messreihe aus vielen (am besten mehr als 1000) Messungen besteht, so findet man oft, dass die Messwerte mit guter Annäherung symmetrisch um einen Mittelwert verteilt sind. Es gibt etwa gleich viele kleinere und kleinste, wie auch größere und größte Werte. Das wurde und wird heute noch von vielen Biologen als völlig normal empfunden, und die Messwerte wurden als «normal verteilt» verstanden. Die große Überraschung kam 1844, als der belgische Statistiker Quetelet fand, dass sich seine 5738 Messungen zum Brustumfang schottischer Soldaten, die er als durchaus «normal» ansah, in auffallend guter Weise an die vom Mathematiker Gauß anno 1809 auf ganz anderer Basis rein theoretisch erarbeitete «Glockenkurve» anpassen ließen (wieso ausgerechnet ein Belgier schottische Soldaten vermaß ist mir rätselhaft ... fide "https://de.wikipedia.org/wiki/Adolphe_Quetelet"). Daraus ergab sich der Begriff der Normalverteilung, erst im biologischen, dann im mathematischen Sinn. Tatsächlich lassen sich viele biologische (und andere) Messungen gut an eine Glockenkurve anpassen, z.B. die Länge der Fichtennadeln, das Gewicht der Kartoffeln, der Chlorophyllgehalt des Spinates. Und die in der Natur festgestellte weite Verbreitung dieser Übereinstimmungen führte dazu, dass die Biologen (und auch andere Wissenschaftler) die Glockenkurve als ein fundamentales Grundgesetz, ein Dogma empfanden, das unbedingte Gehorsamkeit forderte. Dabei wird aber oft übertrieben.

Es muss verstanden werden, dass es nicht nur eine, sondern unendlich viele (im mathematischen Sinn) verschiedene Glockenkurven gibt, flachere und höhere, schmalere und breitere; so wie es auch unendlich viele verschiedene Kreise, Rechtecke und Ellipsen gibt. Dazu einige Bemerkungen.

Streuung und Standardabweichung werden von manchen Biologen für synonym gehalten; aber dem ist nicht so. Natürlich streuen die Messungen einer Stichprobe um einen Mittelwert, und die Weite dieser Streuung reicht einfach vom kleinsten bis zum größten Messwert der Stichprobe. Wenn die Sporenlänge eines Pilzes mit 8-12 μm angegeben wird, so beträgt die Streuung der Messwerte 4 μm mit einem vermuteten aber nicht bestätigten Mittelwert von 10 μm . Die Extremwerte sind stark vom Zufall abhängig; was wenn eine einzige zusätzliche Messung zufälligerweise den bisher größten Messwert übertrifft? Dann ändert natürlich auch die Streuung, vielleicht zufälligerweise sogar sehr stark. Die Streuung hat keine mathematische Basis, und sie ist meist nicht symmetrisch (es gibt in der Regel nicht gleich viele Messwerte unter dem Mittelwert wie über dem Mittelwert, besonders wenn nur wenige Messungen vorliegen).

Um etwas Ordnung in diese Sache zu bringen führte der britische Statistiker Pearson anno 1892 die **Standardabweichung** ein, Symbol «s». Diese ist weit weniger vom Zufall abhängig, beruht auf der mathematisch begründeten, «glockenförmigen» Normalverteilung und ist symmetrisch. Anstelle der (mäßig komplexen) mathematischen Formel stelle ich die Standardabweichung grafisch dar. Sie ist gleich dem Wert vom Mittelwert bis zum Wendepunkt der Glockenkurve (Abb. 1); und sie kann

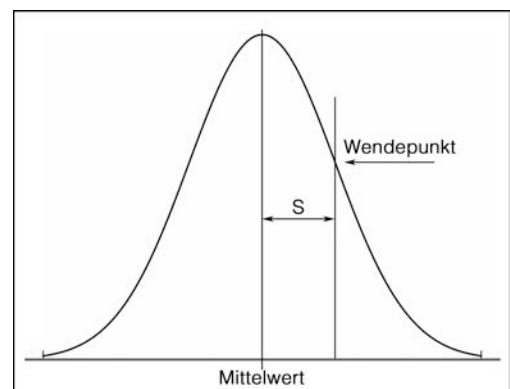


Abb. 1: Eine Normalverteilung und ihre Glockenkurve mit eingezeichneter Standardabweichung s . Der Wendepunkt der Kurve befindet sich im flachsten Teil und ist von Auge nicht erkennbar; er liegt nicht in halber Höhe des Mittelwertes.

sehr leicht mit einem Computerprogramm wie etwa Excel oder mit einem beliebigen Statistikprogramm gefunden werden. Und nun kann ein mathematisch begründeter **Toleranzbereich** der Grundgesamtheit als ungefähr 4 Standardabweichungen geschätzt werden. “Ungefähr” deshalb, weil der **Toleranz-Faktor** von der Anzahl Messungen abhängt. Für 50 und mehr Messungen darf der Toleranz-Faktor 4 mit gutem Gewissen gebraucht werden; für weniger Messungen weicht er jedoch stark ab. Und so werden die untere und obere Grenze mit $\langle m - 2s \rangle$ und $\langle m + 2s \rangle$ berechnet. Dies entspricht etwa den berühmten 95% Grenzen des Toleranzbereiches. Natürlich stimmt das nicht 100%ig, aber die Abweichungen von der mathematischen Genauigkeit sind, biologisch gesehen, ganz und gar bedeutungslos und können vernachlässigt werden (für 50 Messungen beträgt der Faktor 4,06, für 100 Messungen 3,99, für 500 Messungen 3,95; alles zufriedenstellende Annäherungen an den Faktor 4. Kritisch wird es bei 30 Messungen, denn da beträgt der Faktor 4,18; und er wird umso größer, je kleiner die Stichprobe ist). Warum $\langle m \pm 2s \rangle$ und nicht einfach $\langle m \pm s \rangle$? Weil diese engeren Grenzen nur rund zwei Drittel der Grundgesamtheit einschließen (etwa 68,3%); und dem Pilz aufzuzwingen, dass ein Drittel seiner Sporen anormal sein sollen, wäre doch empörend. Schon 5% anormale Sporen sind eine Zumutung (s. Abb. 9).

Prüfung auf Normalverteilung – Eine Notwendigkeit?

Die Forderung nach Normalität der Grundgesamtheit darf nicht allzu streng genommen werden. Abweichungen sind an der Tagesordnung, aber sie sind in der Regel so klein, dass die Grundgesamtheit als normalverteilt angenommen werden darf. Parmasto et al. (1987: 19) hielten fest dass “According to our observations the spore sample of one specimen usually reveals the normal distribution of measurements (Gemäß unserer Beobachtungen zeigt eine Sporen-Stichproben eines Individuums eine Normalverteilung der Messungen)”. Dies kann ich anhand von 16 eigenen Untersuchungen bestätigen (von insgesamt 263 geprüften Stichproben sprachen deren 229 (= 87%) nicht gegen eine Normalverteilung der Grundgesamtheit). Es darf in der Regel angenommen werden, dass die Sporen eines Sporenpulvers, genauer gesagt deren Längen und Breiten, etwa normal (“genügend normal”) verteilt sind. Überraschenderweise sind die Q-Werte (Länge/Breite) bisweilen nicht normalverteilt, trotz der Normalität der Längen und Breiten. Das spielt aber bei den zweidimensionalen Längen-Breiten-Streudiagrammen keine Rolle und darf in diesem Zusammenhang ohne weiteres vergessen werden.

Angesichts der Tatsache, dass mit großer Wahrscheinlichkeit (etwa 85-90%) ein Sporenpulver normalverteilt ist, geht man kein großes Risiko ein, wenn auf eine Prüfung auf Normalverteilung verzichtet wird. Und wenn, was immerhin möglich ist, die Grundgesamtheit der Sporen von der Normalverteilung abweicht, so ist die Abweichung meist unbedeutend und beruht auf anormal kleinen oder großen Sporen, die ausgeschlossen oder einfach ignoriert werden können. Solche Stichproben dürfen wie gewohnt ausgewertet werden. Und wenn damit etwas nicht recht in Ordnung ist, so weichen die Schlussresultate so wenig von der Realität ab, dass die Abweichungen keine biologische Bedeutung haben. Ich weiß, dass diese Aussage den mathematisch puristischen Statistikern zuwider ist, aber die Biologie ist nicht so starr und kann eine etwas lockerere Einstellung gut verkraften.

Es bleibt die Frage, wann ist eine Abweichung geringfügig? Die Größe der Abweichung kann mit mathematischen Prüfverfahren geschätzt werden, aber mit so aufwändigen Berechnungen, dass hier nicht darauf eingegangen wird. Man findet sie in den besseren Lehrbüchern der Statistik, und natürlich auch im Internet. Wer es genau wissen will, kommt um die mathematischen Prüfverfahren nicht herum; aber im mykologischen Alltag ist das kaum je nötig. Die Abweichungen können grafisch mit den weiter unten besprochenen QQ-Diagrammen sichtbar gemacht werden, was die Beurteilung der Abweichungen erleichtert.

Mathematische “Prüfung” auf Normalverteilung der Grundgesamtheit

Wie oben schon angedeutet wird die Glockenkurve vielerorts als ein fundamentales Dogma empfunden, das unbedingte Gehorsamkeit fordert. Deshalb wird oft verlangt, dass die Messwerte einer *Stichprobe* normal verteilt sein müssen. Das ist aber unangebracht, ja geradezu falsch. In der Tat sind die Messwerte einer kleinen Stichprobe (nur 30 bis 50 Sporen) kaum je normal verteilt.

Die Stichproben werden einer großen, im Falle eines Sporenpulvers oft mehrere Millionen Sporen umfassenden *Grundgesamtheit* entnommen, und *diese, also die Gesamtheit aller Sporen eines Sporenpulvers*, sollte normalverteilt sein. Es ist gewagt, von nur 30 oder 50 Sporen auf das ganze Sporenpulver zu extrapolieren, aber es gibt mehrere mathematische Prüfverfahren (z.B. der D’Agostino-Pearson K^2 Test oder der Ryan-Joiner Test), die erlauben zu schätzen ob die Stichprobe aus einer normalverteilten Grundgesamtheit stammen *könnte*; aber diese Verfahren sagen nicht, ob die Grundgesamtheit wirklich normalverteilt ist. Alle mathematischen Prüfverfahren sind nur annähernde Schätzungen, denen nicht zu viel Glauben geschenkt werden darf. Je kleiner die Stichprobe, desto gröber die Schätzung (und deshalb vorsichtigerweise auch ein größerer Toleranz-Faktor); und die Schätzungen verschiedener Prüfverfahren können sich zudem widersprechen. Die Prüfung auf Normalverteilung ist eine mathematisch begründete Schätzung, eine gewagte Extrapolation von der Stichprobe auf die Grundgesamtheit, aber nie eine exakte Erfassung der Grundgesamtheit.

Eine solche Prüfung erlaubt nicht, eine Normalität der Grundgesamtheit zu beweisen, sondern nur auszusagen, dass die Stichprobe nicht gegen eine Normalverteilung der Grundgesamtheit spricht. Es sind dies Methoden um zu raten ob die Stichprobe aus einer normalverteilten Grundgesamtheit kommen kann.

Notizen zu den QQ-Diagrammen

Quantil-Quantil-Diagramme vergleichen die Messungen der Stichprobe mit der Normalverteilung; sie beurteilen also die Stichprobe, nicht die Grundgesamtheit und unterscheiden sich damit grundsätzlich von den oben erwähnten mathematischen Schätzungen der Grundgesamtheit. Um das zu tun, werden die Messungen vom kleinsten bis zum größten Wert zunehmend geordnet. Dann wird die Normalverteilung in ebenso viele Abschnitte («Quantile») aufgeteilt, wie die Stichprobe Messungen enthält. Somit hat man gleichviele beobachtete Werte (die Messungen) wie erwartete (theoretische) Werte (aus der Normalverteilung), die nun paarweise miteinander verglichen werden. Das kann grafisch als QQ-Diagramm dargestellt werden. Die x-Achse enthält die theoretischen Werte, auf der y-Achse werden die beobachteten Werte aufgetragen. Sind alle Messwerte mit den erwarteten Werten identisch, so sind die Messwerte offensichtlich normal verteilt, da sie sich genau an die Normalverteilung halten. Grafisch dargestellt ergibt sich eine von links unten nach rechts oben aufsteigende Gerade.

In der Regel weichen die beobachteten Werte der Stichprobe von den theoretischen Werten der Normalverteilung ab, und im QQ-Diagramm liegen die Punkte über oder unter der theoretischen Gerade. Je enger sich die Punkte um die Gerade scharen, desto besser sind die Messwerte der Normalverteilung angeglichen. Man sieht die Übereinstimmungen und Abweichungen von der Normalität auf einen Blick. In den QQ-Diagrammen sind die Gerade und deren Vertrauensgrenzen eingezeichnet. Punkte außerhalb dieser Vertrauensgrenzen sind zu weit von der Normalen entfernt. Sind das nur wenige Punkte, so dürfen sie vernachlässigt werden. Vielfach liegen die Außenseiter ganz links unten und (oder) ganz rechts oben, was bedeutet, dass die kleinsten und (oder) die größten Sporen die Normalität verlassen haben; der Großteil der Sporen jedoch folgt der Normalverteilung, und die Stichprobe darf normal ausgewertet werden.

Die beiliegenden Beispiele sollen die Wirksamkeit der QQ-Diagramme belegen (Abb. 2, 12, 13, 14).

Bemerkung: In unseren Beispielen werden die Messwerte mit der Normalverteilung verglichen. Solche QQ-Diagramme werden deshalb oft NQ-Diagramme, Normal-Quantil-Diagramme, genannt.

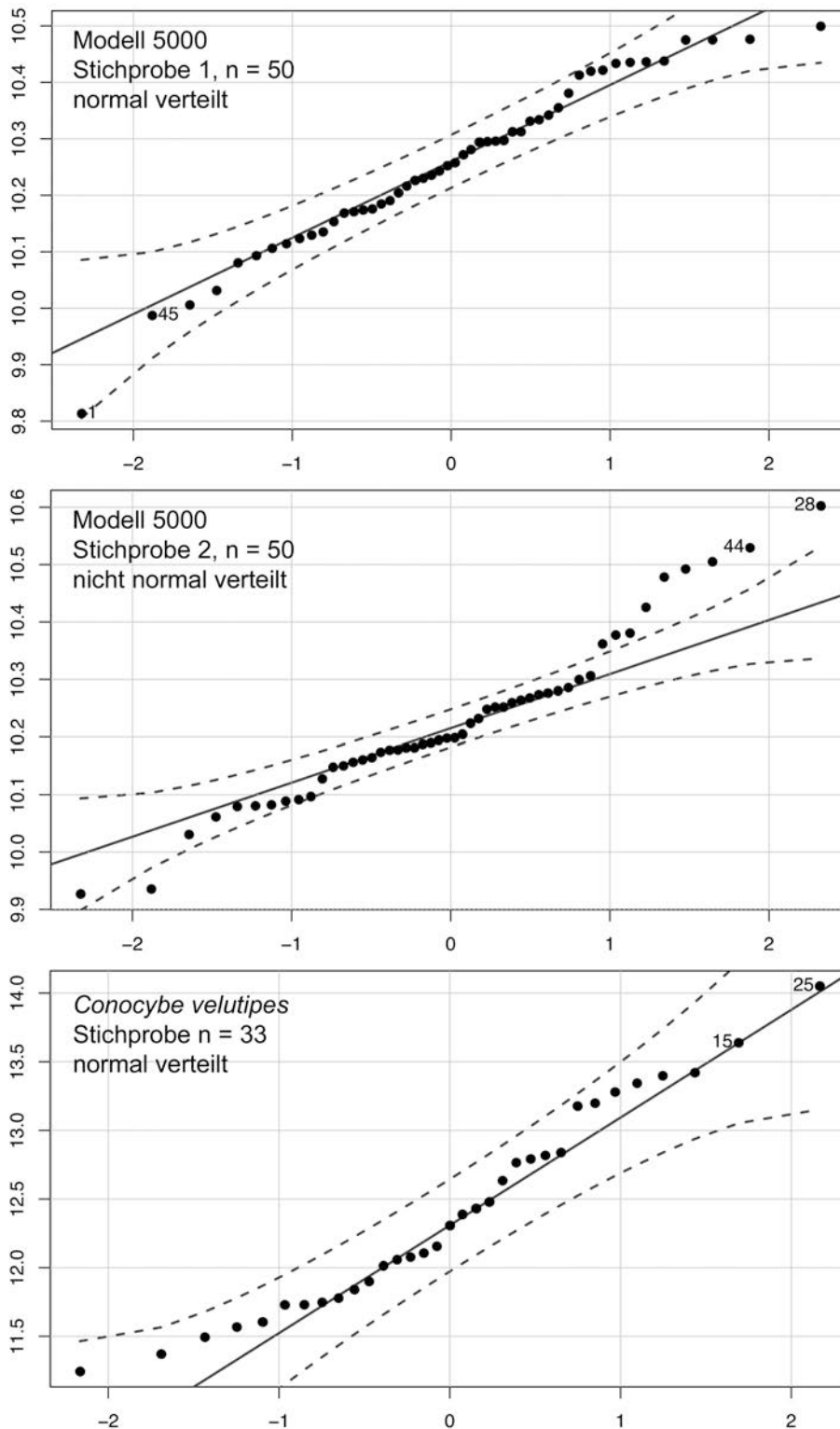


Abb. 2: Drei QQ-Diagramme.

Stichprobe 1 (n = 50, Modell 5000): Alle Messungen liegen in der Vertrauenszone und zeigen so, dass die Stichprobe normalverteilt ist.

Stichprobe 2 (n = 50, Modell 5000): 10 Messwerte befinden sich außerhalb der Vertrauenszone und zeigen so, dass die Stichprobe nicht normal verteilt ist, aber der Ryan-Joiner Test spricht nicht gegen die Hypothese, dass diese Stichprobe aus einer normalverteilten Grundgesamtheit kommen kann.

Conocybe velutipes, diese kleine Stichprobe aus den 1630 vermessenen Sporen ist normalverteilt.

Notizen zu den Histogrammen

Es ist vielleicht angebracht, das Wesen eines Histogramms kurz zu charakterisieren und es von einem Säulendiagramm zu unterscheiden (Abb. 3), denn leider werden Säulendiagramme bisweilen Histogramme genannt, was krass falsch ist.

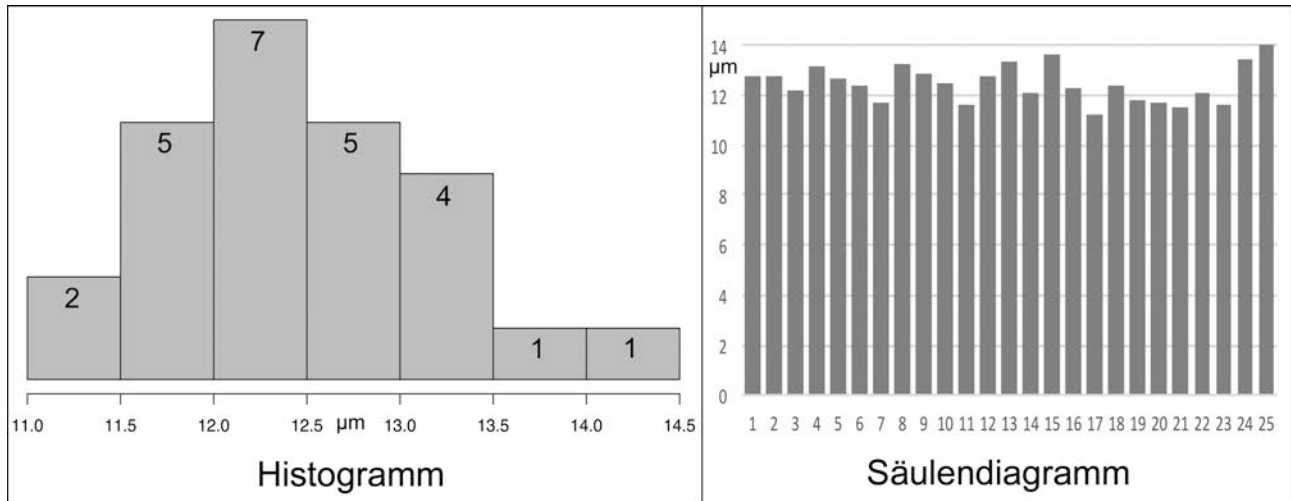


Abb. 3: Die Längen der ersten 25 Sporen aus den 1630 vermessenen Sporen der *Conocybe velutipes* als Histogramm und als Säulendiagramm dargestellt.

Im eingebauten Wörterbuch meines Macintosh Computers findet man für das Histogramm folgende Definition: «*Grafische Darstellung einer Häufigkeitsverteilung in Form von Säulen, die den Häufigkeiten der Messwerte entsprechen*». Also doch ein Säulendiagramm? Nein! Wenn das Wort Säulen durch «Rechtecke» ersetzt wird, so erinnert die genannte Definition weniger an ein Säulendiagramm. Man beachte, dass zwischen den Rechtecken eines Histogramms keine Lücken bestehen. In einem Säulendiagramm hingegen werden die Säulen der besseren Lesbarkeit halber meist mit mehr oder weniger breiten Zwischenräumen voneinander getrennt. Diese sind statistisch bedeutungslos.

Für ein Histogramm werden 1) die Messwerte in Klassen eingeteilt, (für das Beispiel der Abb. 3 wurden 7 Klassen von je 0,5 µm Breite gebildet). 2) jede Messung wird in die ihrem Wert entsprechende Klasse eingewiesen, und 3) wird gezählt wie viele Messwerte jede Klasse erhalten hat, im Beispiel sind diese Anzahlen oben in den Rechtecken angegeben.

Das Säulendiagramm der Abbildung 3 listet einfach die Sporen willkürlich in beliebiger Reihenfolge nacheinander auf, wobei die Höhe der Säulen der Länge der Spore entspricht. Es handelt sich um die gleichen Sporen, die auch für das Histogramm gebraucht wurden.

Säulendiagramme können aber auch brauchbare Information liefern. Wenn die Messungen mit zunehmender Größe geordnet werden, so können zweigipfelige Verteilungen oft mühelos erkannt und dann von Hand getrennt werden (s. Beispiel *Auricularia*, Abb. 13).

Histogramm und Normalverteilung in der gleichen Grafik

Jedem Histogramm kann die «**angepasste Normalverteilung**» überlagert werden. Diese wird aus dem Mittelwert, der Standardabweichung und der Klassenbreite des Histogramms berechnet und grafisch dem Histogramm überlagert. Das erlaubt einen besseren optisch-grafischen Vergleich der Stichprobe mit der Normalverteilung, ist aber kein wirksamer Test für die Normalverteilung der Grundgesamtheit und bringt keinen wirklichen Vorteil. Dazu kommt die Tatsache, dass verschiedene Histogramme einer gegebenen kleinen Stichprobe je nach der Anzahl ihrer Klassen verschieden aussehen können (Abb. 4,5). Die Berechnung und die grafische Darstellung einer

angepassten Normalverteilung muss man einem Computer überlassen. Ich empfehle das kostenlose online-Programm **Free Statistics Software**.

Es wird oft vorgeschlagen (und auch praktiziert), das **Histogramm einer kleinen Stichprobe** mit der Glockenkurve zu vergleichen, rein optisch, gefühlsmäßig, “von Auge”. Das sagt zwar etwas über die Messreihe aus, aber nicht über die Grundgesamtheit. Zudem ist der Vergleich “von Auge” eines Histogramms mit der Normalverteilung trügerisch. Es ist deshalb unangebracht, rein optisch aus einem Histogramm von nur 30-50 Messungen auf die Normalität der Grundgesamtheit schließen zu wollen (Abb. 4,5); außer in krassen Fällen (s. Beispiel *Auricularia*, Abb. 13).

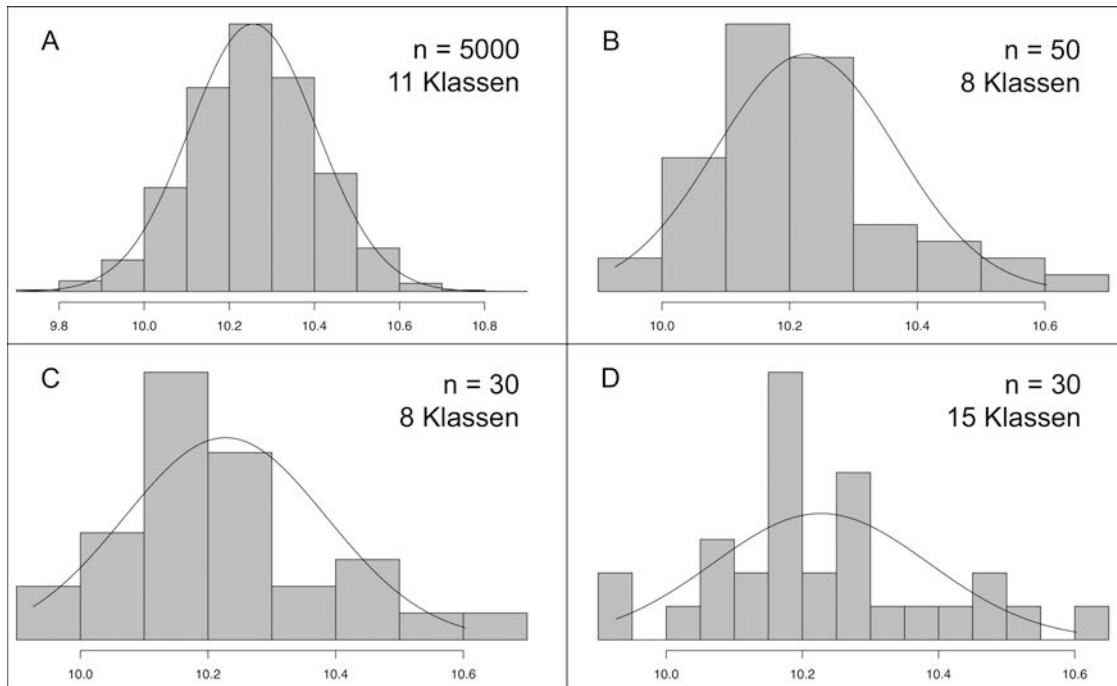


Abb. 4: Histogramme mit angepassten Normalverteilungen zum “Modell 5000” der virtuellen, vom Computer erzeugten, normal verteilten “Sporen”. **A:** Alle 5000 Sporen sind erfasst, das Histogramm ist symmetrisch und sehr gut an die Normalverteilung angepasst. **B:** Werden nur 50 Messwerte dargestellt, so weicht das Histogramm deutlich von der Normalverteilung ab. Die Stichprobe ist nicht normalverteilt. **C,D:** Die gleiche, kleine, nicht normalverteilte Stichprobe wird mit zwei verschiedenen Histogrammen dargestellt. Der Einfluss der Anzahl der Klassen auf das Aussehen der Histogramme ist augenscheinlich. Die optisch-grafischen Abweichungen von der Normalverteilung in B,C und D beziehen sich nur auf die Stichproben, die nicht gegen die Normalverteilung der Grundgesamtheit sprechen (Ryan-Joiner Test, und die Tatsache, dass das Modell 5000 normalverteilt erzeugt wurde).

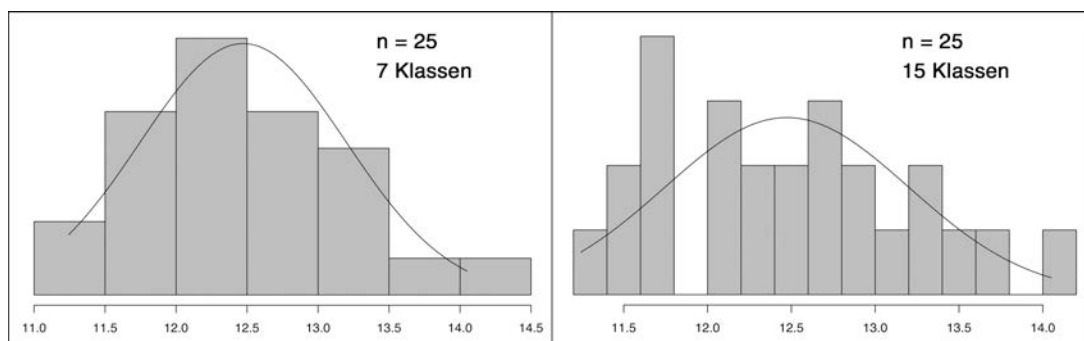


Abb. 5: Zwei Histogramme der Längen der ersten 25 Sporen aus dem Sporenpulver von *Conocybe velutipes*. Beide Histogramme zeigen die gleiche Stichprobe; ihr stark verschiedenes Aussehen beruht lediglich auf der verschiedenen Anzahl Klassen. Die Längen der Sporen sind normal verteilt, was aus dem linken Histogramm vermutet werden kann, wohingegen das rechte Histogramm eine starke Abweichung von der Normalverteilung anzuzeigen scheint. Das ist ein Beispiel der Unzulässigkeit, aus kleinen Stichproben auf die Grundgesamtheit schließen zu wollen.

Aber wie steht es mit dem **Histogramm einer großen Stichprobe**? Wann ist eine Stichprobe groß? Damit ein Histogramm statistisch aussagekräftig ist, sollte es mindestens 7 Klassen umfassen (3 vor dem Mittelwert, 1 um den Mittelwert und 3 nach dem Mittelwert), besser sind deren 9. Und es müssen genügend Messungen vorliegen, dass in jeder Klasse mindestens 1 Zählung erwartet werden darf. Das ist annähernd der Fall, wenn die durchschnittliche Anzahl Zählungen pro Klasse etwa 50 beträgt. Daraus ergibt sich eine untere Grenze für eine große Stichprobe von rund $7 \cdot 50 = 350$ Messungen, besser sind $9 \cdot 50 = 450$ Messungen. Zugegeben, das ist viel, besonders wenn von 450 Sporen sowohl die Länge, als auch die Breite gemessen werden müssen. So viele Messungen erhöhen zwar die Genauigkeit der 95%-Grenzen sowohl der Länge, als auch der Breite der Sporen, aber die Histogramme solcher Stichproben können im rein optischen Vergleich immer noch trügen (Abb. 6).

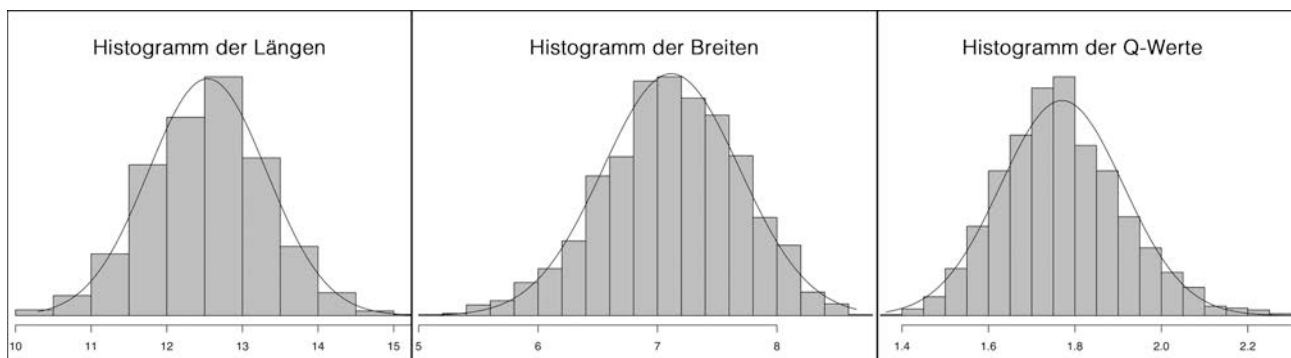


Abb. 6: Histogramme mit angepasster Normalverteilung der Längen, Breiten und Q-Werte der gesamten Messreihe von 1630 Sporen der *Conocybe velutipes*. Alle drei scheinen optisch-grafisch normalverteilt; aber die Breite und die Q-Werte sind es nicht. Die Längen sind normalverteilt, trotz der leichten optischen Abweichung von der Glockenkurve (Ryan-Joiner Test).

Zweigipfelige Histogramme treten oft bei Sporenpulvern auf, die von Fruchtkörpern mit zwei- und viersporigen Basidien stammen. Solche Sporenpulver sind aus zwei verschiedenen, meist ungleich großen Grundgesamtheiten zusammengesetzt und zeigen im besten Fall eine ungleichmäßige zweigipfelige Verteilung (Abb. 13), allenfalls eine verbreiterte Glockenkurve. Ob zweierlei Basidien vorkommen, kann anhand eines Lamellen-Quetschpräparates festgestellt werden, aber um die beiden Grundgesamtheiten zu trennen ist man auf den Computer angewiesen. Hier kann ein Histogramm manchmal helfen, aber diese Hilfe ist unsicher. Es gibt auch verschiedene statistische Tests, z.B. der Möls Test in Parmasto (1987), die erlauben die beiden Grundgesamtheiten auseinanderzuhalten, und schließlich können auch QQ-Diagramme und geordnete Säulendiagramme aufschlussreiche Hinweise geben.

Notizen zu den Ellipsen

Die Normalverteilung und die Histogramme sind eindimensional, univariat, wie auch die Histogramme mit angepasster Normalverteilung. Sie behandeln ausschließlich nur die Länge, nur die Breite, oder nur den Q-Wert und widmen den jeweils anderen Dimensionen keine Aufmerksamkeit. **Streudiagramme** und **Ellipsen** hingegen sind zweidimensionale, **bivariate** Darstellungen zweier korrelierter, bis zu einem gewissen Grad voneinander abhängiger Größen; etwa wie «je länger eine Spore, desto breiter ist sie auch».

Die **Grafische Darstellung der Resultate mit Streudiagrammen** stellt die gemessenen Sporen mit kleinen Symbolen in einem Koordinatensystem dar, auf dessen Achsen die Längen und die Breiten

aufgetragen sind. In meinen Streudiagrammen entspricht die Abszisse der Länge, die Ordinate der Breite der Sporen.

Das einfachste Streudiagramm zeigt nur gerade die Symbole der Sporen (Abb. 7A). Mit etwas mehr Aufwand kann das Zentrum der Sporenwolke mit einem besonderen Symbol dort eingezeichnet werden, dessen Abszissenwert den Mittelwert der Länge, der Ordinatenwert den Mittelwert der Breite enthält. In einem weiteren Schritt werden die 90% oder 95% Grenzen der Länge und der Breite als horizontale und vertikale Strecken eingezeichnet, deren Schnittpunkt im Zentrum der Sporenwolke liegt (Abb. 7B). Soweit so recht, wenn auch nur beschränkt informativ.

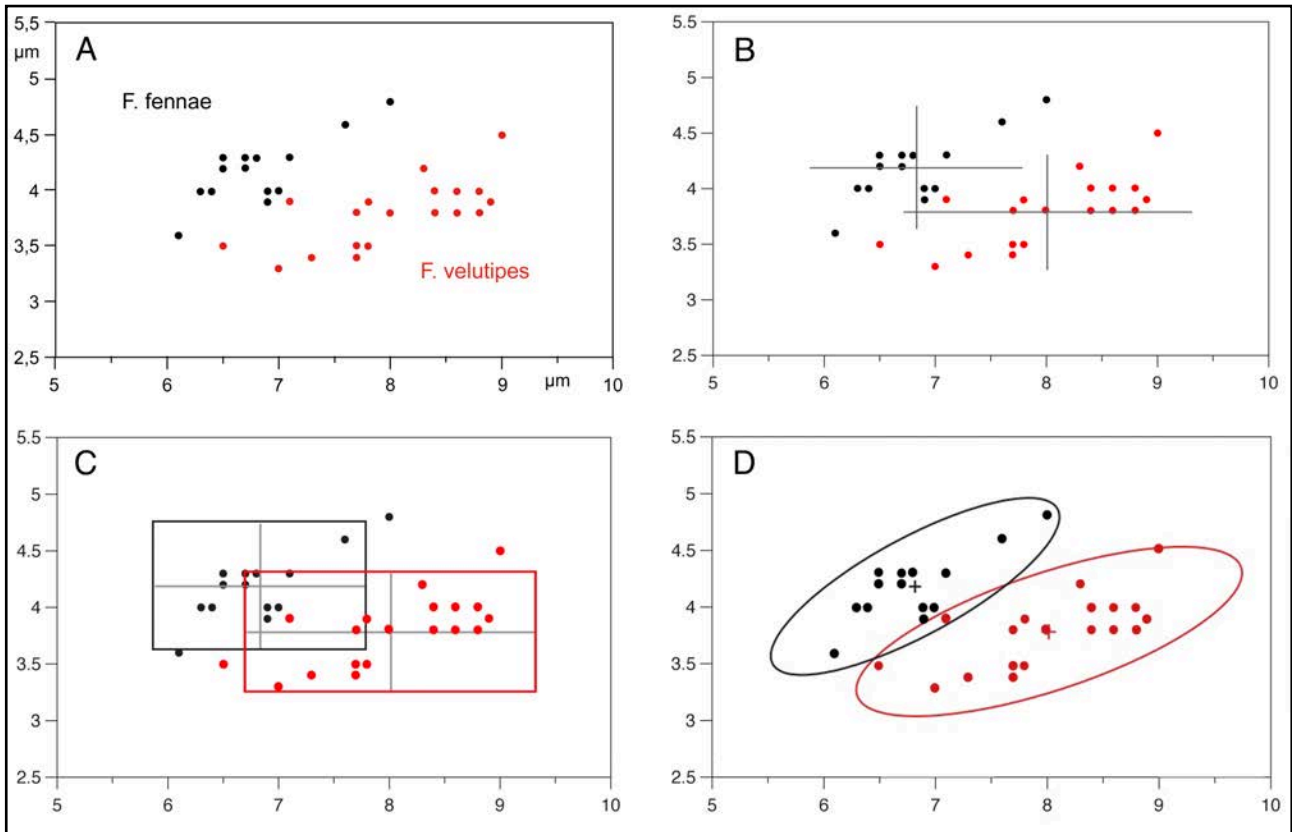


Abb. 7: Vier Streudiagramme der Sporenmasse von *Flammulina fennae* und *Flammulina velutipes*. In A,B und C scheinen die beiden Arten stark zu überlappen, aber die 95% Toleranzellipsen in D trennen die Arten recht gut. Das Diagramm C ist falsch, da der Neigung der Symbolwolken nicht Rechnung getragen wird. Die Koordinaten der Messwerte wurden den Streudiagrammen von Bas (1983) entnommen.

Falsch wird die Sache, wenn die horizontalen und vertikalen Strecken dazu benützt werden um Rechtecke zu zeichnen, wie in der Abb. 7C gezeigt. Denn Längen und Breiten der Sporen sind nicht unabhängige Größen, sondern sie sind meist bis zu einem gewissen Grad miteinander korreliert. Je länger die Spore, desto breiter ist sie in der Regel auch. Deshalb liegt die Sporenwolke im Streudiagramm meist schief; in meinen Darstellungen von links unten nach rechts oben. Dieser schiefen Lage werden die Darstellungen mit Rechtecken nicht gerecht. Die mathematisch richtige Darstellung ist eine Ellipse (Abb. 7D), die weiter unten besprochen wird.

Gelegentlich findet man in der Literatur auch Streudiagramme mit von Hand eingezeichneten, vielgestaltigen und völlig willkürlichen Umgrenzungen der Symbolwolken, die nur psychologischen Wert haben, so wie das in der Abb. 8 gezeigt ist.

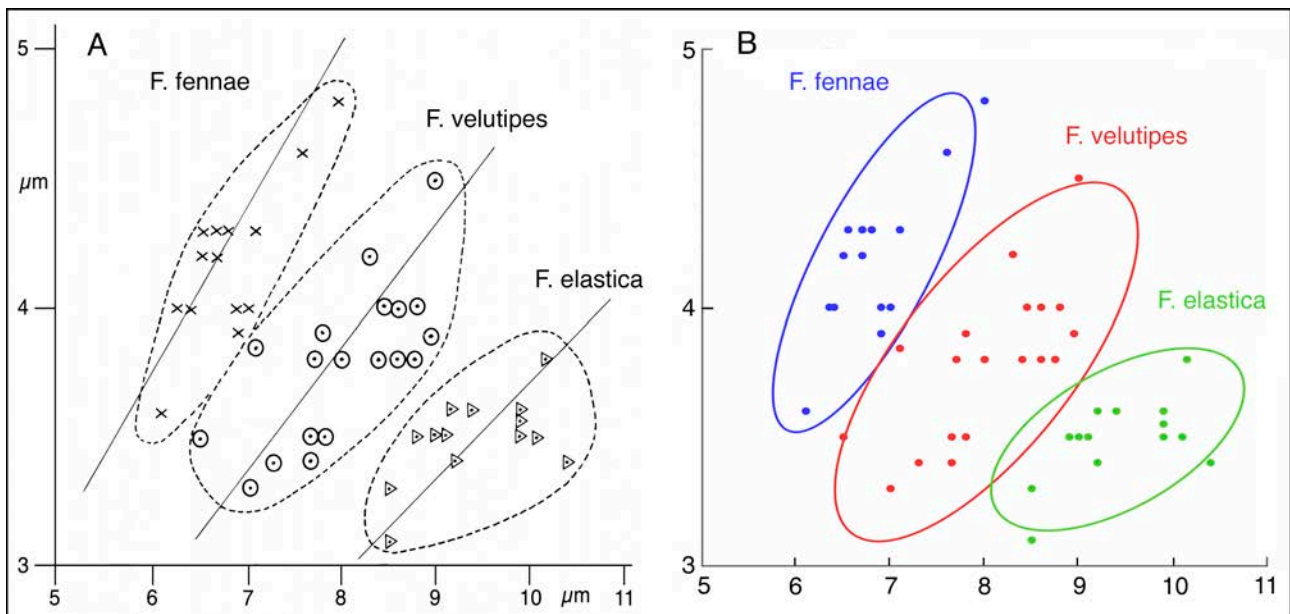


Abb. 8: Die gut gemeinten, von Auge geschätzten und von Hand gezeichneten Abgrenzungen der Sporenbezirke der drei *Flammulina*-Arten sind gefühlsmäßig ansprechend aber statistisch unhaltbar. Die mathematisch-statistische Erfassung durch die 90% Toleranzellipsen sind objektiv und daher besser, wenn auch eine kleine Überlappung vorkommt. Die Koordinaten der Messwerte wurden den Streudiagrammen von Bas (1983) entnommen. A nach Bas, gedreht um der hier bevorzugten Darstellung gerecht zu werden. *Flammulina elastica* heisst bei Bas *F. velutipes* f. *longispora*.

Die **Toleranzellipse**, auch Populationsellipse genannt, ist die beste Umgrenzung einer Sporenwolke im zweidimensionalen Streudiagramm. Aber sie darf natürlich nicht "von Auge" so ungefähr und von Hand eingezeichnet werden; sie muss berechnet werden.

Es gibt für jedes Streudiagramm theoretisch unendlich viele Toleranzellipsen, jede mit ihrer eigenen Toleranzgrenze. In der Praxis werden die Toleranzgrenzen auf 90%, 95% oder 99% gesetzt (Abb. 9); in der Mykologie kommen vor allem die 90% Ellipse und die 95% Ellipse zum Einsatz (Abb. 10, 11).

Welche Toleranzgrenze ist angebracht? Ich habe die Erfahrung gemacht, dass für kleine Stichproben die 90% Toleranzellipse besser angebracht ist, für große hingegen eignet sich die 95% Ellipse besser. Kleine Stichproben sind solche mit weniger als etwa 100 Messungen. Aber am besten berechnet man beide Ellipsen und schaut sich das Resultat an; die Wahl ist dann eher psychologisch als statistisch begründet.

Die Abbildung 10 zeigt schon auf den ersten Blick, dass eine Ellipse (hier die 95%-Ellipse) befriedigend oder unbefriedigend sein kann, je nach der Größe des Stichprobenumfangs. Bei mehr als etwa 500 Messungen empfiehlt sich die 99%-Ellipse (Abb. 9).

Die **Berechnung der Toleranzellipse** ist komplex und aufwändig. Vielen on-line Programmen oder kostenlos erhältlichen Programmen fehlt die Ellipsen-Funktion. Eine Ausnahme macht das Programm "R", aber dessen Anwendung bleibt dem Unerfahrenen kryptisch. Bequemer sind professionelle Statistikprogramme, aber die sind sündhaft teuer ...

Ich bin für jeden Hinweis auf erschwingliche Programme dankbar, die erlauben, Toleranzellipsen zu berechnen und zu zeichnen.

In den Documenta Geigy (1960) wird eine detaillierte Anleitung mit vielen Formeln gegeben, die es dem Erfahrenen erlaubt, ein Computerprogramm zu schreiben. Das Resultat ist eine Liste von x-y Koordinaten beliebig vieler Punkte auf dem Umfang der Ellipse. Diese müssen dann zusammen mit den Koordinaten der Messwerte (die Längen und Breiten der Sporen) in ein Grafikprogramm kopiert werden.

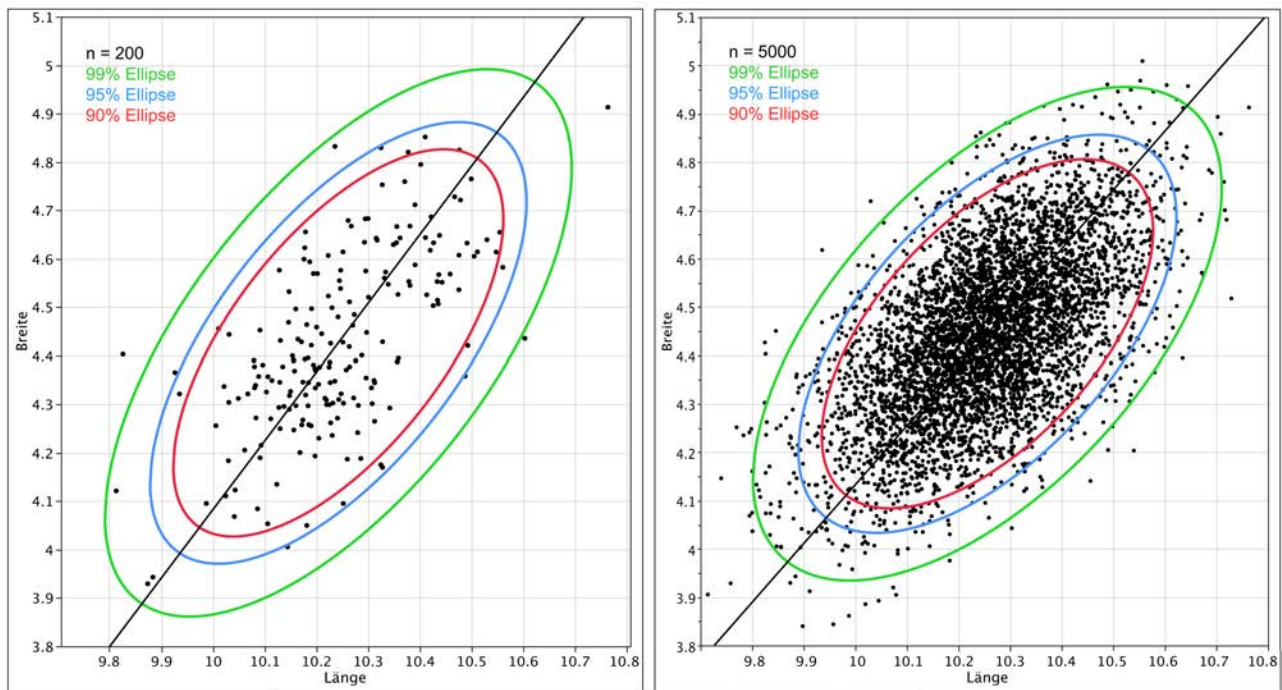


Abb. 9: Die drei meist gebrauchten Toleranzellipsen mit eingezeichneten Hauptachsen; «Modell 5000». Die auf 200 Messungen beruhenden Ellipsen sind deutlich größer als die der Grundgesamtheit ($n = 5000$), da die Unsicherheit (Ungenauigkeit) der Grenzen einer Toleranzellipse mit abnehmendem Stichprobenumfang zunimmt. Sie sind auch etwas steiler als die der Grundgesamtheit, da auch deren Neigung nur eine annähernde Schätzung ist.

Optisch-grafische Auswertung der Toleranzellipse

Außer der banalen Tatsache, dass eine Toleranzellipse einen berechneten (aber trotzdem nur geschätzten) Bereich anzeigt, innerhalb dessen die allermeisten Sporen der Grundgesamtheit liegen, erlaubt die Ellipse noch weitere Aussagen.

Die vertikalen und horizontalen **Tangenten** an die Ellipse können als **bivariat bestimmte Toleranzgrenzen** der Länge und der Breite angesehen werden. Diese aus korrelierten Messwerten zweidimensional geschätzten Toleranzen sind aber deutlich größer als die eindimensional berechneten Toleranzbereiche («Zusammenarbeit» gegen «jeder für sich»). Ungefähr vergleichbare Grenzwerte zeigen die Tangenten einer 90% Toleranzellipse und die eindimensionalen 95% Toleranzbereiche ($0,95 \cdot 0,95 = 0,9025$); aber auch das sind immer nur Annäherungen, die stark vom Stichprobenumfang n abhängig sind. **Beispiel:** Die Sporen des *Gymnopus confluens* ($n = 77$) messen $6,5-8,6 \times 3,2-4,5 \mu\text{m}$ (95%, Länge und Breite unabhängig voneinander berechnet) und $6,4-8,7 \times 3,2-4,6 \mu\text{m}$ (Tangenten an die 90% Ellipse); gerundete Werte (Abb. 11).

Individuelle univariate Toleranzbereiche der Länge und der Breite lassen sich an der Toleranzellipse ablesen. Legt man eine vertikale Gerade durch eine ausgewählte Stelle der Abszisse (Länge der Sporen), so zeigt die Ellipse den Bereich der dazu passenden Breiten an. Das gilt auch für das Auffinden der wahrscheinlichen Längen zu einer bestimmten Breite, wenn horizontale Geraden gebraucht werden. Dies ist möglich, weil die Toleranzellipse eine korrelierte, bivariate Analyse darstellt.

Die **Korrelation** zwischen Länge und Breite (“je länger desto breiter”) kann mit dem Korrelationskoeffizienten « r » angegeben werden, dessen Wert in unserer Darstellungsweise zwischen 0 (keine Korrelation) und 1 (maximale Korrelation) liegt; aber dessen Berechnung liegt außerhalb des Themas. Man benutze die Programme Statdisk oder Free Statistics Software.

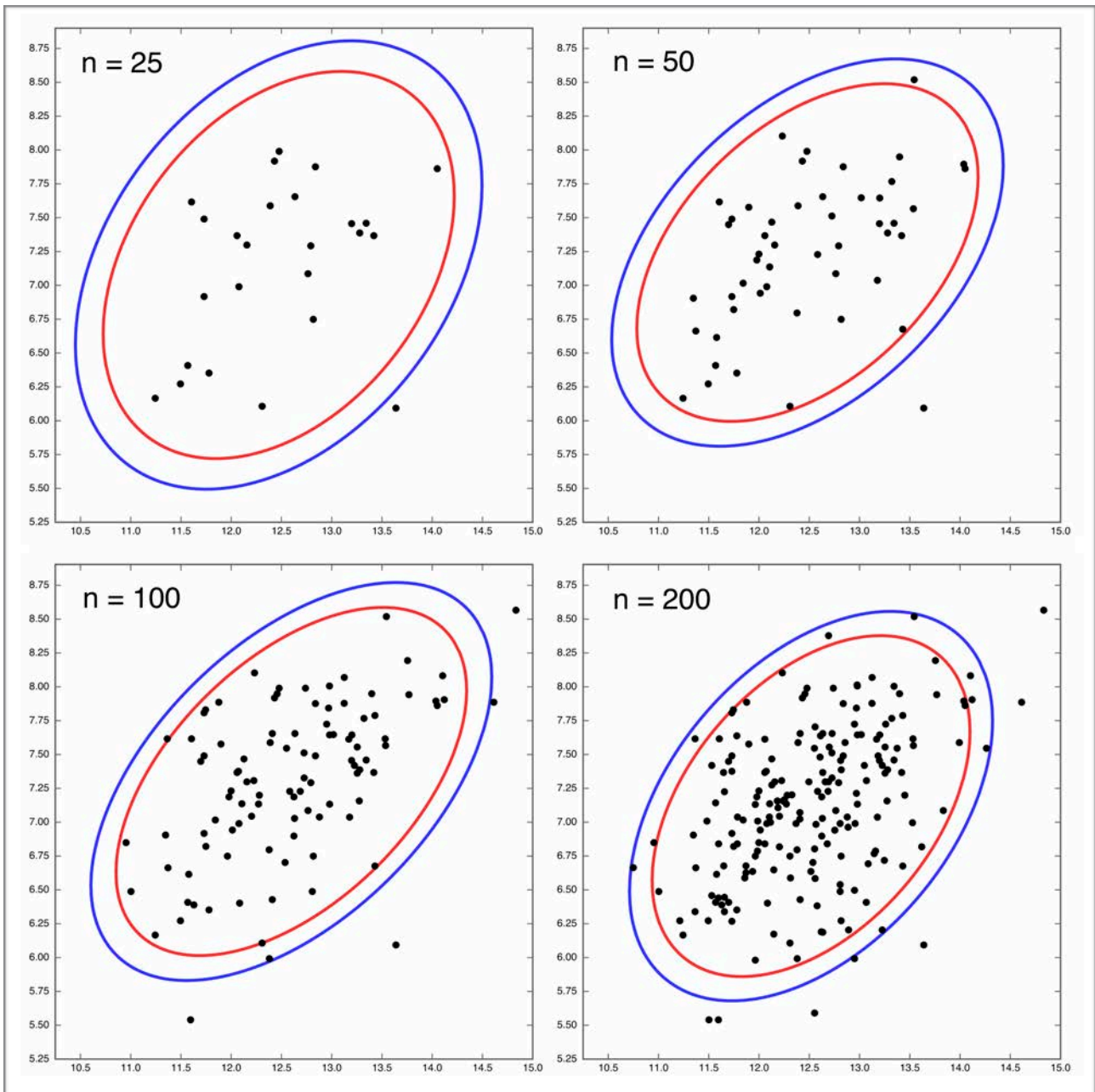


Abb. 10: Vier verschieden große Stichproben aus den 1630 Messungen der Sporen von *Conocybe velutipes*. Je größer die Stichprobe, desto enger umschließen die Ellipsen die Messpunkte, da der Toleranzfaktor mit zunehmendem n kleiner wird. Rot = 90% Ellipse, blau = 95% Ellipse.

Die Korrelation drückt sich auch in der Toleranzellipse aus. Die Form und die Neigung der Ellipse geben einen optisch-grafischen Eindruck der Stärke der Korrelation. Grob gesagt, je mehr sich eine Ellipse der Kreisform nähert, desto geringer ist die Korrelation, vorausgesetzt, die Neigung der Ellipse ist deutlich von der Horizontalen verschieden (Abb. 11).

Banal aber wichtig: Die Form der Toleranzellipse ist in keiner Weise mit der Form der Sporen verbunden. Die Sporen der *Pseudobaeospora calcarea* sind kurz ellipsoidisch, die Toleranzellipse hingegen lang elliptisch; die Sporen des *Gymnopus confluens* sind lang ellipsoidisch, die Toleranzellipse aber ist kurz elliptisch.

Auf eine Besprechung weiterer Eigenschaften der Toleranzellipsen muss verzichtet werden, denn diese führen rasch zu komplexeren Analysen, die außerhalb des gestellten Themas liegen.

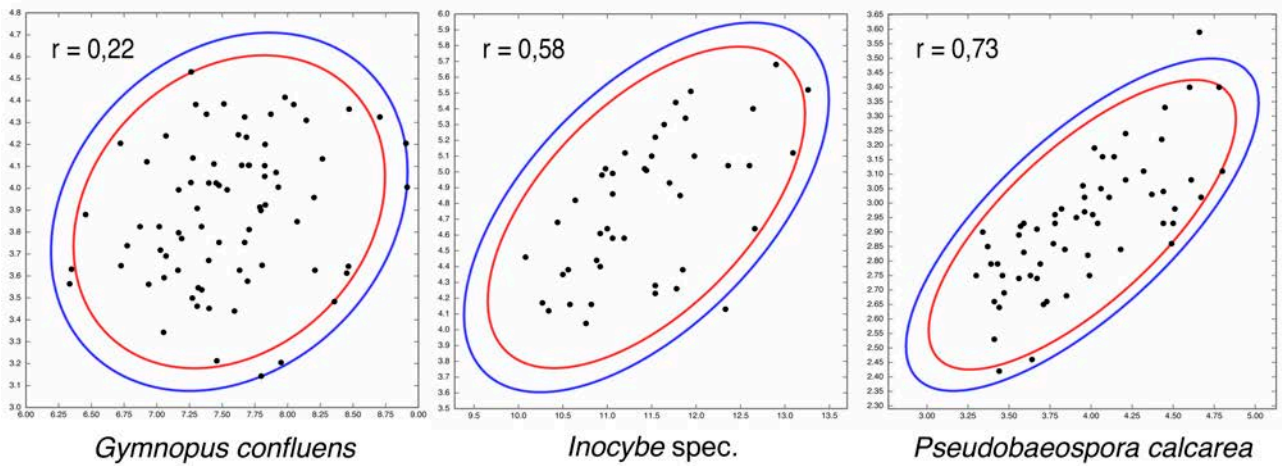


Abb. 11: Die Form der Ellipsen spiegelt den Korrelationskoeffizienten r zwischen der Länge und der Breite der Sporen wider.

Zwei weitere Beispiele aus der Literatur

Inocybe spec. – Fannechère (2005) fand, dass kleine Stichproben ($n < 50$) nicht normalverteilt sind und rät deshalb von der Verwendung der Toleranzbereiche und der Toleranzellipsen ab, da diese auf der Normalverbreitung beruhen. Wie oben schon festgehalten wurde, sind kleine Stichproben fast notgedrungen nicht normalverteilt, aber das bedeutet nicht, dass auch die Grundgesamtheit (das gesamte Sporenpulver) nicht normalverteilt ist.

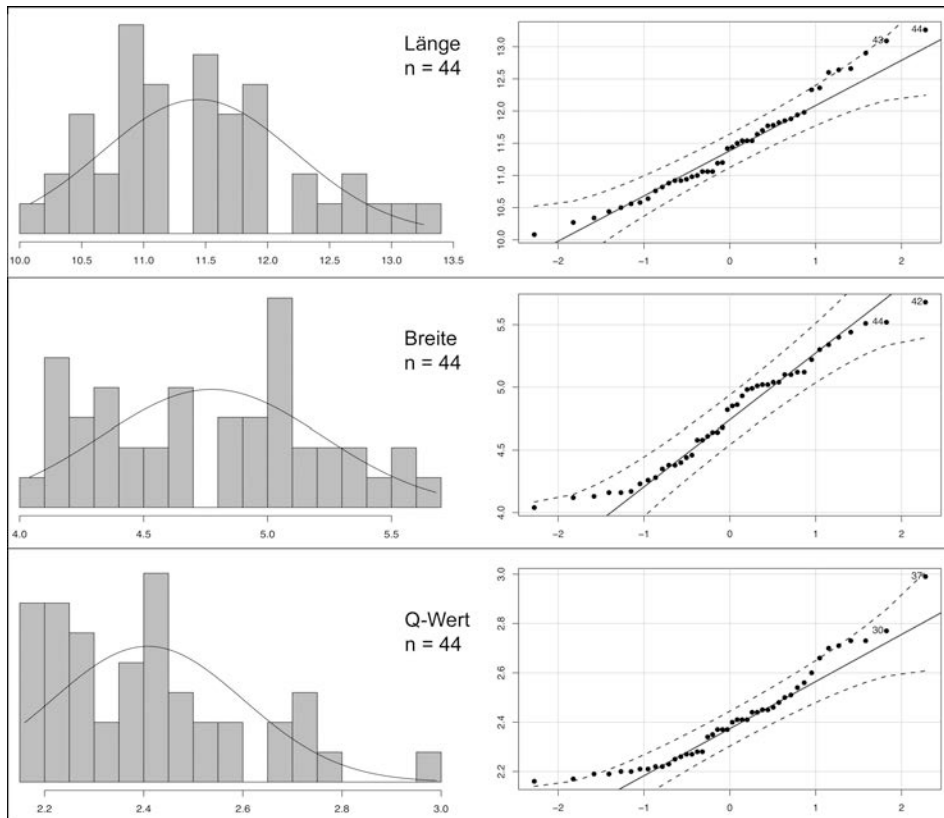


Abb. 12: Grafische Darstellungen der Längen, Breiten und Q-Werte eines nicht näher bestimmten Risspilzes (*Inocybe spec.*), nach den Messungen von Fannechère (persönliche Mitteilung). Die Stichprobe ist nicht normalverteilt, aber die Längen und Breiten der Grundgesamtheit sind normalverteilt, die Q-Werte sind es nicht (Ryan-Joiner Test). Dieses Beispiel zeigt klar, dass aus den unregelmäßigen Histogrammen kleiner Stichproben nicht auf eine "Nichtnormalität" der Grundgesamtheit geschlossen werden darf.

Herr Fannechère hat mir freundlicherweise seine Messungen zur Verfügung gestellt. Meine Analysen seiner Stichproben ($n = 44$) sind in den Abbildungen 11 und 12 dargestellt. Die 90% Toleranzellipse zeigt nur zwei Außenseiter und umfasst die Stichprobe recht gut; aber die Verteilung der Sporen im Innern der Ellipse erscheint unausgeglichen, wie das oft bei kleinen Stichproben der Fall ist. Die Histogramme der Länge, der Breite und der Q-Werte weichen optisch-grafisch bedeutend von der Normalverteilung ab, wie dies auch Herr Fannechère fand. Der Ryan-Joiner Test der Längen und der Breiten der Stichprobe hingegen erlaubt nicht, die Normalverteilung der Längen und der Breiten in der Grundgesamtheit (im Sporenpulver) auszuschließen. Dies gilt nicht für die Q-Werte, die im Sporenpulver mit großer Wahrscheinlichkeit nicht normalverteilt sind.

***Auricularia auricula-judae* s.l.** – Möls (1987) benutzte die mittleren Sporenlängen von 33 Pilz-Individuen der (damaligen) Sammelart *Auricularia auricula-judae* (genannt *A. auricula*) um mit seinem Algorithmus zu zeigen, dass diese heterogene Population in zwei homogene Populationen aufgeteilt werden kann. Die kurzsporige Population entspricht der *Auricularia americana*, die langsporige der *Auricularia auricula-judae*.

In der Abbildung 13 wird die Mischpopulation als zweigipfelige, aus 33 Messungen bestehende Grundgesamtheit grafisch dargestellt, und die Abbildung 14 zeigt die beiden homogenen Grundpopulationen, die den beiden oben genannten Arten entsprechen.

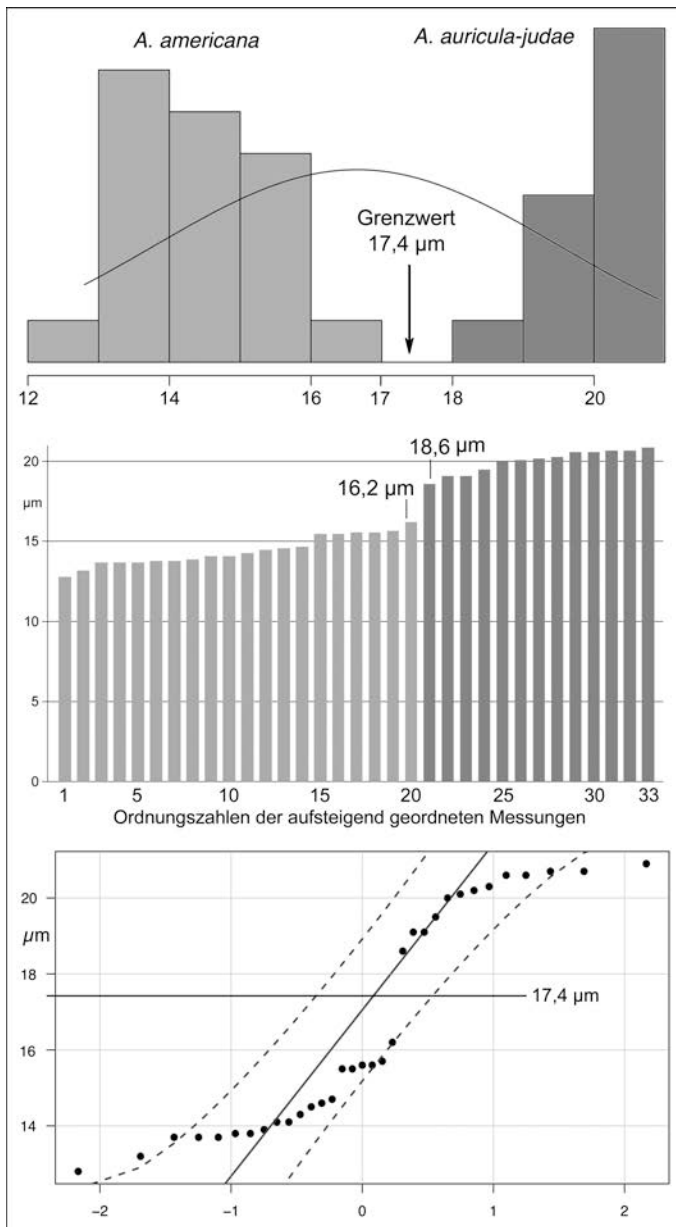


Abb. 13: Die aus zwei Grundgesamtheiten zusammengesetzte Messreihe resultiert in einer stark nichtnormalen Stichprobe, die sowohl *Auricularia americana*, als auch *Auricularia auricula-judae* umfasst. Diese beiden Arten können in allen drei Grafiken erkannt werden. Daten nach Möls (1987), als *Auricularia auricula* sensu lato.
 $(16,2 + 18,6)/2 = 17,4$

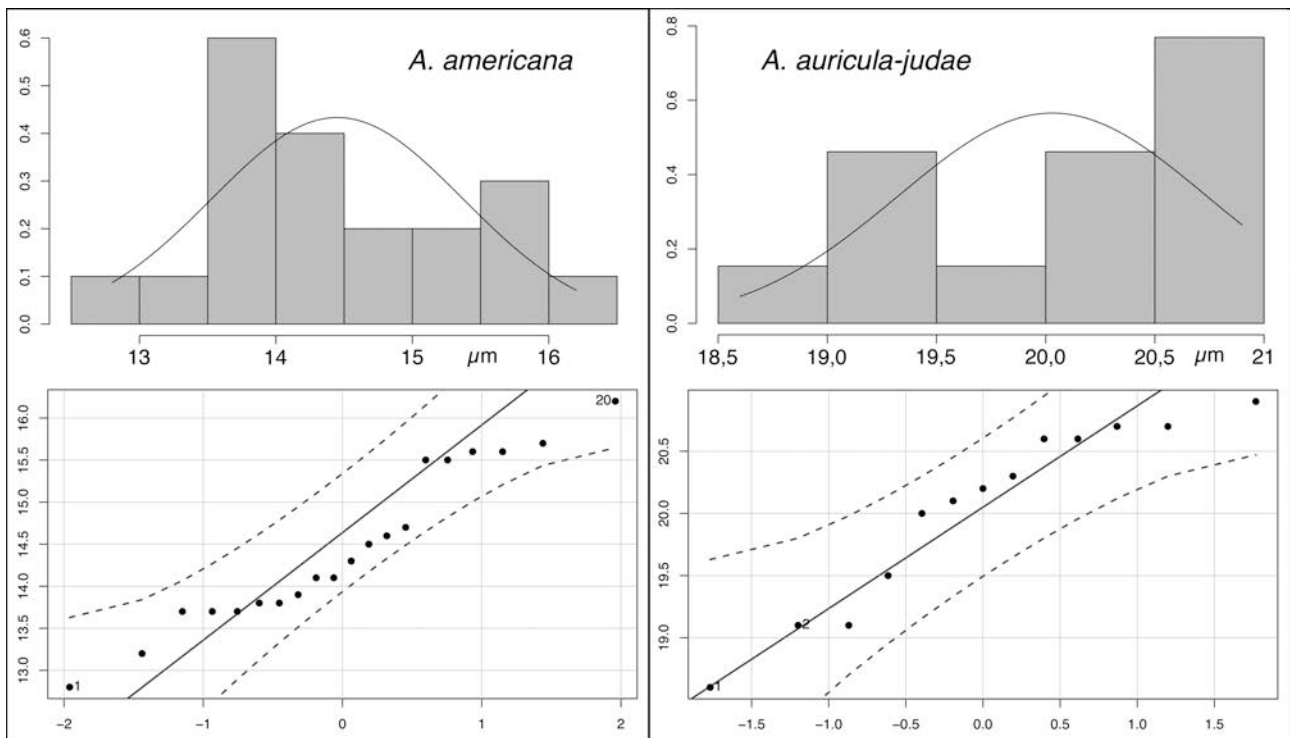


Abb. 14: Die Stichproben der beiden *Auricularia*-Arten der Abbildung 13 getrennt analysiert zeigen je eine Normalverbreitung ihrer Sporenlängen, trotz stark von der Glockenkurve verschiedenen Histogrammen.

Diskussion

Die besprochenen Methoden und Bemerkungen mögen dem Mathematiker unvollständig und ungenau, manchem Mykologen hingegen nebensächlich und verwirrend erscheinen. Das ist die unvermeidliche Folge des bewussten Verzichtes auf (oft recht komplexe) mathematische Formeln, die man jedoch im Internet finden und zu verstehen versuchen kann. Die hier geduldete mathematische Ungenauigkeit hat in den allermeisten Fällen einen so kleinen Einfluss auf das Endresultat, dass die biologische Bedeutung der Resultate in keiner nennenswerten Weise beeinflusst wird.

Das Wichtigste in Kürze

- Es ist zum Verständnis der Methoden fundamental wichtig, Stichprobe und Grundgesamtheit zu unterscheiden. Die Stichprobe besteht aus den Messungen einzelner Sporen, die Grundgesamtheit ist das gesamte Sporenpulver (oder im Falle eines Lamellen-Quetschpräparates die Gesamtheit aller reifen Sporen des Pilzes – eine nicht empfehlenswerte Arbeitsweise).
- Die Streuung der Messungen einer Stichprobe ist zufällig und hat keine mathematische Basis; sie bezieht sich nur auf die Stichprobe. Die Standardabweichung ist mathematisch begründet und kann sich sowohl auf die Stichprobe, als auch auf die Grundgesamtheit beziehen. Die Standardabweichung der Grundgesamtheit ist immer eine annähernde Schätzung, nie eine genaue Messung.
- Kleine Stichproben (weniger als etwa 50 Messungen) sind fast immer nicht normalverteilt. Dies bedeutet aber nicht, dass auch die Grundgesamtheit nicht normal verteilt ist. Kleine Stichproben dürfen statistisch analysiert werden, aber die Genauigkeit ist geringer, als Analysen, die auf großen Stichproben beruhen.
- Die Grundgesamtheit ist mit großer Wahrscheinlichkeit (rund 85-90%) normal verteilt. Ist sie es nicht, so sind die Abweichungen meist banal und dürfen vernachlässigt werden.

- Die mathematische Prüfung der Normalverteilung der Grundgesamtheit ist immer ein annäherndes Rätselraten und darf unterlassen werden. Eine solche Prüfung erlaubt nicht, eine Normalität der Grundgesamtheit zu beweisen, sondern nur auszusagen, dass die Stichprobe nicht gegen eine Normalverteilung der Grundgesamtheit spricht.
- Eine optisch-grafische Prüfung der Normalverteilung der Stichprobe (nicht der Grundgesamtheit) ist das QQ-Diagramm. Dieses zeigt auch deutlich, ob die Stichprobe homogen oder aus 2 (oder mehreren) unterschiedlichen Populationen besteht. Das QQ-Diagramm sagt nichts über die Grundgesamtheit aus.
- Histogramme sind keine Säulendiagramme. Sie sind nur dann aussagekräftig, wenn sie auf wenigstens ungefähr 350 Messungen beruhen.
- Das Aussehen eines Histogramms kleiner Stichproben hängt stark von der Anzahl Klassen ab und ist deshalb nur von beschränkter Aussagekraft.
- Toleranzellipsen müssen berechnet werden; sie dürfen nicht von Auge so ungefähr gezeichnet werden. Sie sind umso größer, je kleiner die Stichprobe, weil kleine Stichproben die Grundgesamtheit schlechter abschätzen können. – Leider sind Computerprogramme zur Berechnung der Toleranzellipsen sehr teuer.
- Je weniger Messungen die Stichprobe enthält, desto mehr weicht die Toleranzellipse von der theoretisch wahren Ellipse der Grundgesamtheit ab. Um eine brauchbare Annäherung zu erhalten, sollte die Ellipse auf wenigstens 100 (besser 200) Messungen beruhen.
- Die Form der Ellipse sagt etwas über die Bindung der Länge mit der Breite der Spore aus (“je länger desto breiter”). Je schmaler die Ellipse, desto stärker die Bindung (Korrelation).
- Die Regel der «10 muss, 20 soll, 30 darf man messen» ist brauchbar zur Pilzbestimmung, aber unbrauchbar für die wissenschaftliche Beschreibung einer Art.

Literatur

- Bas C (1983) *Flammulina* in Western Europe. *Persoonia* 12: 51-66.
- Documenta Geigy (1960) Wissenschaftliche Tabellen. J.R. Geigy A.G. Basel. Pharmazeutische Abteilung.
- Fannechère G (2005) Statistiques et notation des dimensions des spores. *Bulletin de la Société mycologique de France* 121: 255-292, 2005 (erhalten Dezember 2006).
- Halbwachs H, Karasch P (2019) Dimensionen: Ein Blick in die unsichere Welt der Pilzmorphologie. *Zeitschrift für Mykologie* 85: 93-108.
- Möls T (1987) Testing of the Population Homogeneity. In: Parmasto et al. (1987).
- Parmasto E et al. (1987) Variation of basidiospores in the Hymenomycetes and its significance to their taxonomy. *Bibliotheca Mycologica* Band 115. J. Cramer, Stuttgart.

Anhang

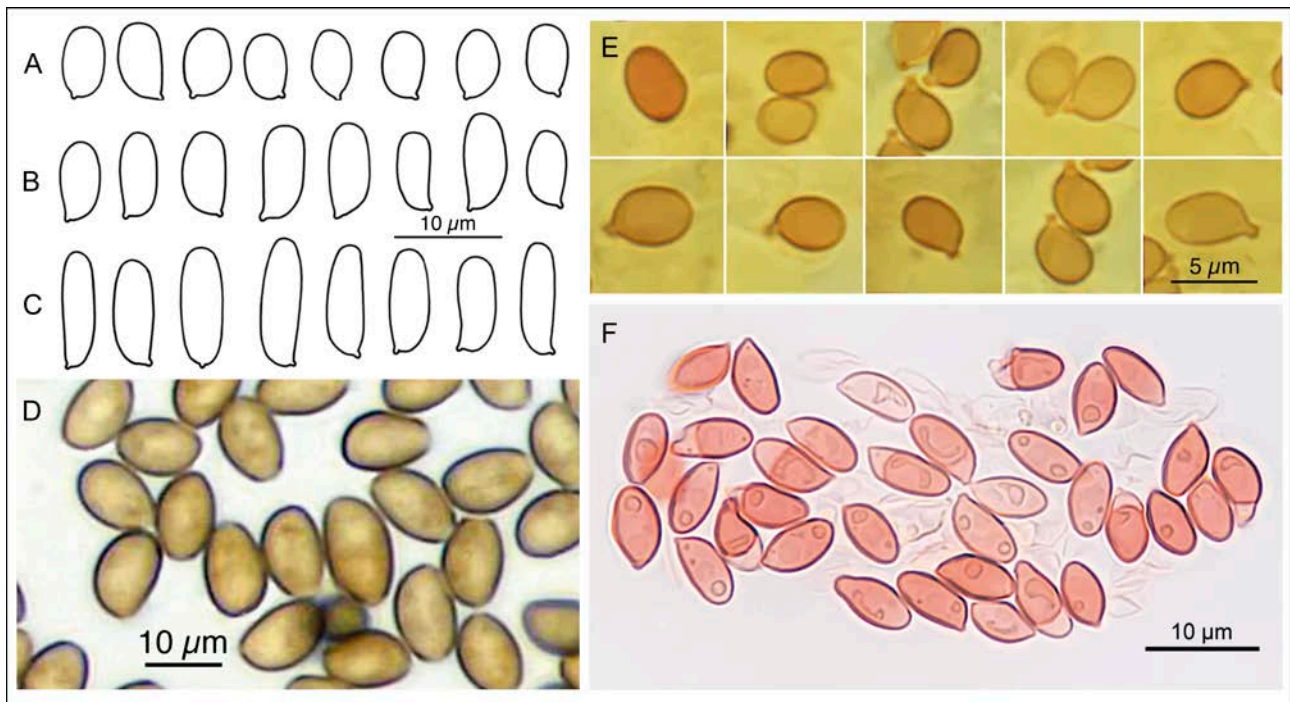


Abb. 15: Einige repräsentative Sporen. **A** *Flammulina fennae*; **B** *Flammulina velutipes*; **C** *Flammulina elastica*; **D** *Conocybe velutipes* in KOH; **E** *Pseudobaeospora calcarea* in Melzer's Jodlösung, **F** *Gymnopus confluens* in SDS-Kongorot, Sporenpulver 24 h feucht gehalten, die meisten Sporen entwickelten eine sekundäre, dicke, innere Wandschicht. – A,B,C nach Bas (1983); D,E,F originale Fotografien.