

# Inhaltsverzeichnis

Inhaltsverzeichnis.....	1
Grundsätzliches zur Statistik in den Biowissenschaften .....	2
Stichproben und Fehler .....	2
Verteilungen .....	4
Beschreibung von Stichproben mit Einzelwerten und Streuung.....	5
Übersichtstabelle Formeln Mittelwerte und Streuung .....	8
Beispiel zu Mittelwerten und Streuung .....	8
Fehlerrechnung.....	9
Anhang: Herleitung der Fehlerfortpflanzungsregeln .....	11
Darstellung von Ergebnissen mit Streuung .....	11
Angabe von Werten im Text .....	11
<b>Graphische Darstellung von Stichproben (kommt noch)</b> .....	12
Die Geschichte mit der Signifikanz.....	12
<b>t-Tests (kommt noch)</b> .....	14
<b>Chi-Quadrat-Test (kommt noch)</b> .....	14
<b>Wie man mit Statistik keinen Unsinn macht... (kommt noch)</b> .....	14
<b>Tipps zum Weiterlesen und Surfen</b> .....	14
Register.....	15

# Grundsätzliches zur Statistik in den Biowissenschaften

Schon der Begriff „Statistik“ hat oft einen schlechten Ruf („Glaube keiner Statistik, die Du nicht selbst gefälscht hast“) und ist auch beim Auswerten von Daten oft der unbeliebteste Teil. Richtig angewandt ist Statistik aber einfach die mathematische Disziplin, die sich mit quantitativen Daten beschäftigt, also im Laboralltag mit Messwerten. Im Gegensatz zu absoluten Zahlen, mit denen wir das Rechnen gewohnt sind, sind experimentell erhobene Daten immer eine Stichprobe aller theoretisch möglichen Daten und damit entweder ein Teil der Gesamtmenge (z.B. die Körpergrößen von 100 Individuen einer Art), eine Annäherung an den „echten“ Wert (z.B. 100 Messungen eines Km-Werts, die um den Mittelwert schwanken) oder eine Mischung aus beidem. Die Statistik dient jetzt dazu, solche schwer greifbaren Daten objektiv und nachvollziehbar zu handhaben. Das bedeutet zum einen, solche Daten sinnvoll darzustellen, aber auch Daten weiterzuverarbeiten, um zum Beispiel Hypothesen zu testen und aus den gemessenen Werten weitere Werte herzuleiten.

Eine korrekte Anwendung statistischer Methoden berücksichtigt also gerade die Unsicherheiten von Messwerten, sichert so den Auswerter möglichst gut gegen Fehlschlüsse ab und ist essentiell für wissenschaftliche Arbeiten. Schlechte statistische Arbeit hat nur Mittelwerte vor Augen, vergleicht diese ohne Rücksicht auf die Unsicherheiten und findet sich erschreckend häufig in der Presse oder Talkshows zu politischen Themen...

Die folgenden Kapitel sollen eine kurze Einführung in eine Reihe statistischer Grundideen und ein paar häufige Anwendungen in den Biowissenschaften geben. Sie ersetzen definitiv keinen Statistikkurs, können aber hoffentlich den Einstieg für den nicht allzu Mathe-affinen Biologen erleichtern.

## ***Stichproben und Fehler***

Alle experimentellen Daten sind **Stichproben** aus der Gesamtheit aller theoretisch möglichen Werte. Im Idealfall spiegelt solch eine Stichprobe die echte Verteilung der Werte realistisch wieder, allerdings können verschiedene **Fehler** dem Experimentator hier einen Strich durch die Rechnung machen:

- **Systematische Fehler** sind Fehler, die jede Messung gleich beeinflussen, zum Beispiel ein falsch geeichtes Messgerät, falsche Annahmen bei der Messung, ein stets vorhandener störender Einfluss etc. Systematische Fehler lassen sich durch gute Planung und Kontrollen minimieren, aber nie vollständig vermeiden. Statistisch sind sie normalerweise nicht greifbar, das heißt, bei der Auswertung fallen sie unter den Tisch, allerdings können bekannte systematische Fehler teilweise aus den Messdaten

herausgerechnet werden (Wobei man natürlich die korrekte Fehlerrechnung beachten muss, auf die wir später eingehen). Vergleicht man Daten innerhalb einer Messung miteinander, kann es sein, dass systematische Fehler keine Rolle mehr spielen, da alle Werte gleich betroffen sind, unterscheiden sich die Ergebnisse verschiedener Messreihen an verschiedenen Tagen, von unterschiedlichen Experimentatoren, in anderen Laboren oder mit verschiedenen Messmethoden, dann sind unterschiedliche systematische Fehler häufig die Ursache.

- **Statistische Fehler** oder **zufällige Fehler** sind Fehler, die bei jedem Messwert unterschiedlich groß und in unterschiedliche Richtungen ausfallen können und sind damit die klassischen „Messfehler“, die mit statistischen Methoden beschrieben werden können. Statistische Fehler können teilweise durch eine optimale Versuchsdurchführung minimiert werden, allerdings sind nie alle „statistisch störenden“ Einflüsse völlig ausschaltbar. Will man durch Messungen, die mit statistischen Fehlern behaftet sind einen echten Wert möglichst genau bestimmen, ist ein möglichst großer **Stichprobenumfang** ( $n$ ) entscheidend, da sich statistische Fehler bei größeren Stichproben besser gegenseitig aufheben.

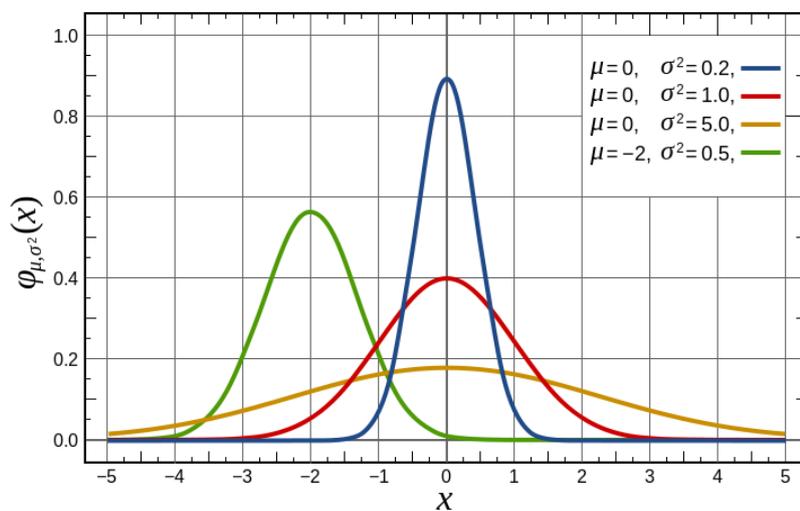
Neben echten Messfehlern gibt es gerade in biologischen Systemen häufig auch eine natürliche Varianz, die man auch als eine statistische Abweichung vom idealen Mittelwert betrachten und statistisch behandeln kann. Natürliche Varianz und Messfehler lassen sich allerdings in den Datenreihen am Ende nicht mehr unterscheiden und werden dann als ein „Fehler“ mitgeführt.

- **Grobe Fehler** sind solche, die zu vermeiden der Experimentator in der Hand hat. Hierzu zählen zum Beispiel eine falsche Planung des Experiments, die Anwendung eines falschen Geräts oder falscher Einstellungen, das Vertauschen von Proben, bemerkbare Störungen des Experiments von außen. Grobe Fehler können zu systematischen oder statistischen Fehlern führen (So führt eine falsch tarierte Waage zu einem systematischen Fehler, da sie immer den gleichen falschen Nullwert subtrahiert, zu viel Kaffee vor dem Messen führt über das Zittern zu statistischen Fehlern, die in alle Richtungen vom Soll abweichen).

Grobe Fehler sollten nicht mit Hilfe von „statistischen Tricks“ ausgeglichen, sondern vermieden werden und mit groben Fehlern belastete Messwerte oder Experimente aus einer Auswertung weggelassen werden. Zum nachträglichen Aufspüren grober Fehler ist ein möglichst gut geführtes Laborbuch natürlich äußerst nützlich!

## Verteilungen

Die in einer Stichprobe erhobenen Daten können auf verschiedene Arten verteilt sein. Sehr häufig sind zufällig streuende Daten **normalverteilt**, können also gut durch eine **Normalverteilung** (auch **Gauß-Verteilung** oder **Glockenkurve**, Abbildung 1) beschrieben werden, was die Sache sehr einfach macht, da hierzu nur zwei Werte notwendig sind: der Mittelwert  $\mu$  als Angabe, wo das Maximum der Verteilung liegt und die Standardabweichung  $\sigma$  als Maß für die Breite der Verteilung, wobei etwa 68% aller Werte in einem Bereich von  $\mu \pm \sigma$  liegen, 95% in einem Bereich von  $\mu \pm 2\sigma$  und 99,7% in einem Bereich von  $\mu \pm 3\sigma$ . Auf die Berechnung von Mittelwert und Standardabweichung gehen wir später genauer ein.



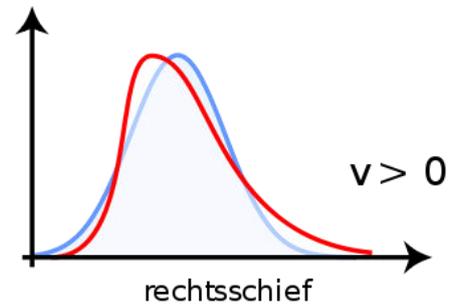
**Abbildung 1:** Vier Beispiele für Normalverteilungen mit unterschiedlichen Mittelwerten ( $\mu$ ) und Standardabweichungen ( $\sigma$ ).

(Quelle: Wikimedia Commons, Inductiveload, Datei in public domain)

Häufig wird daher jede Messreihe im Labor als normalverteilt angenommen, was allerdings aus verschiedenen Gründen problematisch sein kann:

- Die Normalverteilung **stetig**, hat also für jeden Wert  $x$  auch eine Häufigkeit, während manche Verteilungen **diskret** sind, also nur bestimmte  $x$ -Werte überhaupt annehmen können (häufig ganze Zahlen z.B. die Tentakelzahl bei *Hydra* oder die Zahl der Samen in einer *Arabidopsis*-Schote). Hier mit einer Normalverteilung zu arbeiten ist also im besten Fall eine grobe Näherung!
- Die Normalverteilung streckt sich mathematisch betrachtet unendlich weit in beide Richtungen, was in der Realität eigentlich nie der Fall ist. Das ist selten ein Problem, da die Normalverteilung oft immer noch eine sehr gute Näherung darstellt, wer aber die Körpergröße von Siebtklässlern als normalverteilt annimmt und daraus dann die Wahrscheinlichkeit für 15 Meter oder -1,7 Meter große Kinder ableitet, hat etwas nicht richtig verstanden...

- Viele Werte sind schlicht und einfach nicht normalverteilt, so kann eine Verteilung z.B. **schief** sein, also nicht symmetrisch zum Mittelwert (z.B. Zellgrößen, die nicht unter eine lebensfähige Größe fallen können aber nach oben hin weniger streng beschränkt sind). Verteilungen können auch komplizierter sein, wenn sie zum Beispiel mehr als ein Maximum aufweisen (z.B. Körpergröße bei Arten mit zwei Geschlechtern, wo die Verteilung eventuell die Summe zweier Normalverteilungen ist) oder schlicht und einfach eine ganz andere Form annehmen (z.B. Treten beim Werfen eines idealen Würfels alle Zahlen von 1 bis 6 gleich häufig auf, alle anderen aber nie!).



**Abbildung 2:** Beispiel für eine rechtsschiefe Verteilung (rot) *Quelle: Wikimedia Commons, Chrischi, Datei in public domain*

Ob die eigenen Daten einer angenommenen Verteilung entsprechen, lässt sich natürlich auch statistisch testen, dazu später mehr beim  $\chi^2$ -Test.

## ***Beschreibung von Stichproben mit Einzelwerten und Streuung***

Die vollständigste, aber unhandliche Beschreibung einer Stichprobe ist natürlich die Angabe aller Einzelwerte, zum Beispiel in Form einer Tabelle. Einfacher und daher häufiger verwendet ist die Angabe als **Mittelwert** mit einer Angabe zur Streuung der Messwerte, häufig in der Form  $\bar{x} \pm \Delta x$  (mit  $\bar{x}$  als Mittelwert und  $\Delta x$  als „Fehler“) dazu gibt es mehrere Möglichkeiten, wobei so eine Beschreibung z.B. bei asymmetrischen Verteilungen oder solchen mit mehreren Maxima die Realität nur sehr unzureichend wiedergibt. (Zum Beispiel könnten bei einer Operation etwa die Hälfte der Patienten direkt sterben und der Rest noch 10 Jahre leben, eine mittlere Lebenserwartung von 5 Jahren anzugeben, wäre hier also eher irreführend!) Die Formeln für alle im Folgenden beschriebenen Werte finden sich am Ende dieses Kapitels.

Die am häufigsten verwendeten Mittelwerte sind:

- Das **arithmetische Mittel** oder der **Durchschnitt** ( $\bar{x}$  englisch: *arithmetic mean*) wird salopp auch oft einfach nur als „Mittelwert“ bezeichnet – so auch von den gebräuchlichsten Tabellenkalkulationsprogrammen. Es berechnet sich aus der Summe aller Einzelwerte geteilt durch die Gesamtzahl der Werte. Das arithmetische Mittel eignet sich gut als Abschätzung des „echten Werts“, wenn die Daten zufällig und symmetrisch um diesen streuen (z.B. wenn der Km-Wert eines Enzyms in mehreren Wiederholungen gemessen wird). Auch ist es ideal, wenn man die Summe größerer

Mengen von Einzelmessungen abschätzen will (Wenn eine Familie im Durchschnitt 2,1 Kinder hat, muss man bei 100 Familien im Ort mit etwa 210 Kindern rechnen). Nachteile des arithmetischen Mittels sind, dass es besonders bei diskreten Verteilungen oft Werte ergibt, die kein Einzelwert annehmen kann (z.B. hat keine Familie genau 2,1 Kinder...) und dass es z.B. bei schiefen Verteilungen mehr oder weniger weit vom einem „typischen Wert“ entfernt liegen kann.

- Der **Median** oder **Zentralwert** (englisch *median*) ist derjenige Wert über und unter dem gleich viele Einzelwerte liegen. Er wird ermittelt, indem alle Werte der Größe nach aufgereiht werden und der Wert genau in der Mitte ausgewählt (bei ungerader Zahl der Werte) oder der Durchschnitt aus den beiden Werten in der Mitte (bei ungerader Zahl der Werte, bei großen Messreihen meist der gleiche Wert) gebildet wird. Der Median eignet sich gut, um den „typischen“ Messwert darzustellen (z.B. ist der Median des Einkommens ein besserer Wert für das, was „Otto Normalverbraucher“ an Geld zur Verfügung hat als das Durchschnittseinkommen, da die Verteilung rechtsschief ist), allerdings lässt sich mit dem Median schlecht auf größere Mengen schließen (z.B. Trinken über die Hälfte der Gäste eines Wirts nur ein Bier, ein Teil feiert aber kräftig und trinkt viel mehr. Wenn der Wirt nun vom Median=1 schließt, dass er für 200 Gäste etwa 200 Biere benötigt wird er den echten Wert deutlich unterschätzen, was mit einem Durchschnittswert von 1,3 Bier pro Gast nicht passiert wäre. Kommt aber ein Gast in die Bar, liegt der Wirt mit der Vermutung er müsse wahrscheinlich ein Bier zapfen meist richtig). Bei symmetrischen Verteilungen sind Median und arithmetisches Mittel gleich.

Um die Streuung der Einzelwerte um den Mittelwert anzugeben, sind die folgenden Angaben am gebräuchlichsten:

- Die **Varianz** (englisch *variance*) ist eine Summe der Abweichungen der Einzelwerte vom Mittelwert, geteilt durch die Zahl der Messungen. Um zu verhindern, dass sich Abweichungen zu den beiden Seiten des Mittelwerts gegenseitig aufheben, werden nicht die absoluten Differenzen, sondern deren Quadrate addiert. Da bei der Varianz durch die Zahl der aufaddierten Werte geteilt wird, ist sie grundsätzlich unabhängig von der Zahl der Messwerte – die Varianz sinkt also nicht mit einer steigenden Zahl der Messungen, der bestimmte Wert spiegelt aber immer genauer die wirkliche Varianz der betrachteten Verteilung wieder! Als quadratischer Wert ist die Varianz meist allerdings kein besonders anschauliches Maß.

- Die **Standardabweichung**, auch **mittlerer Fehler des Einzelwerts** (englisch *standard deviation*, abgekürzt SD) ist die Wurzel aus der Varianz und ist das gebräuchlichste Maß für die Streuung der Messwerte um den Mittelwert. Sie hat die gleiche Einheit wie die ermittelten Messwerte und ist wie die Varianz prinzipiell unabhängig von der Zahl der Messwerte. Bei einer Normalverteilung lässt sich über die Standardabweichung auch abschätzen, wie viele Messwerte in welchem Bereich um den Mittelwert liegen, bzw. umgekehrt, wie wahrscheinlich es ist, bei einer solchen Verteilung einen Wert außerhalb dieser Bereiche zu finden (siehe oben). Die Standardabweichung wird in Tabellenkalkulationsprogrammen mit der Formel STABW.N oder STABWN (Open Office) berechnet.

Da bei Experimenten immer nur ein Teil der potentiell möglichen Werte gemessen werden kann und die so ermittelte Standardabweichung vom echten Wert der betrachteten Verteilung abweicht, wird hier die **empirische Standardabweichung** verwendet, bei der statt durch die Zahl der Messwerte  $n$  durch  $n-1$  geteilt wird. Die empirische Standardabweichung ist von der Zahl der Messungen abhängig und nähert sich bei großen  $n$  der tatsächlichen Standardabweichung an, ist bei kleineren  $n$  aber größer, was die experimentelle Unsicherheit reflektiert. Die empirische Standardabweichung wird in Tabellenkalkulationsprogrammen mit der Formel STABW (Open Office, Excel) oder STABW.S (Excel) berechnet.

- Während Varianz und Standardabweichung ein gutes Maß dafür sind, wie die Werte einer Verteilung um den Mittelwert streuen, ist bei vielen Messungen die Frage interessanter, wie verlässlich eigentlich der bestimmte Mittelwert ist. Der **Standardfehler** oder **mittlere Fehler des Mittelwerts** (englisch *standard error of the mean*, abgekürzt SEM) ist ein Maß für die Verlässlichkeit des erhaltenen Mittelwerts und ist definiert als die Standardabweichung, geteilt durch die Wurzel der Zahl der Messwerte – der Standardfehler ist damit abhängig von der Zahl der Messwerte und wird mit mehr Messungen kleiner. Um den Standardfehler zu halbieren benötigt man viermal so viele, um ihn auf ein Zehntel herunterzubringen einhundertmal so viele Messungen! Neben der Angabe als absolutem Fehler (also Zahlenwert) ist häufig auch eine Angabe als relativer Fehler (in % des Mittelwerts sinnvoll).

Da auch dem Standardfehler mathematisch eine Gaußfunktion zu Grunde liegt, erlaubt er auch abzuschätzen wie wahrscheinlich es ist, dass der echte Mittelwert in einem Bereich um den experimentell bestimmten Wert liegt, man kann mit seiner Hilfe also

einen **Vertrauens-**oder **Erwartungsbereich** definieren und als **Konfidenzintervall** ausdrücken, so liegt der echte Wert im Bereich

- $\bar{x} \pm \bar{s}$  mit einer Wahrscheinlichkeit von ca. 68%
- $\bar{x} \pm 2\bar{s}$  mit einer Wahrscheinlichkeit von ca. 95%
- $\bar{x} \pm 3\bar{s}$  mit einer Wahrscheinlichkeit von ca. 99,7%

Anders herum ausgedrückt, liegt dann die **Irrtumswahrscheinlichkeit** bei der Annahme der echte Wert läge in den oben genannten Bereichen bei etwa 32%, 5% und 0,3%.

## Übersichtstabelle Formeln Mittelwerte und Streuung

Wert	Formel/Zeichen
Einzelner Messwert	$x_1, x_2, x_3$ etc., allgemein $x_i$
Zahl der Messwerte (Stichprobengröße)	$n$
Arithmetisches Mittel (Durchschnitt)	$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$
Varianz	$S^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$
Standardabweichung (Grundgesamtheit)	$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$
Standardabweichung (empirisch)	$S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$
Standardfehler des Mittelwerts, absolut	$\sigma(\bar{x}) = \Delta\bar{x} = \frac{\sigma}{\sqrt{n}}$
Standardfehler des Mittelwerts, relativ	$\partial\bar{x} = \frac{\Delta\bar{x}}{\bar{x}}$

### Beispiel zu Mittelwerten und Streuung

Die beiden Biologen E.Ager und L.Azy haben eine neue Fliegenart entdeckt und messen nun für eine Veröffentlichung die Körpergrößen der gefangenen Tiere. Beide greifen auf die gleichen Tiere zurück, L.Azy ist aber deutlich weniger fleißig als sein Kollege und misst nur jede vierte Fliege und so kommen die beiden zu folgenden Messwerten:

E.Ager: 7,7 6,9 5,8 7,1 7,8 8,7 9,0 5,9 7,1 7,6 6,3  
 7,4 8,6 8,3 7,8 7,7 (n = 16)  
 L.Azy: 7,1 5,9 7,4 7,7 (n = 4)

Daraus ergibt sich folgende Auswertung:

	Mittelwert	Standard- abweichung	Standardfehler, absolut	Standardfehler, relativ
E.Ager	7,48	0,94	0,24	3,2%
L.Azy	7,03	0,79	0,39	5,5%

Vergleicht man die beiden Messungen, dann stellt man fest, dass die Mittelwerte und Standardabweichungen relativ nah beieinander liegen – was ja auch zu erwarten ist, wenn beide die gleiche Verteilung vermessen haben. Interpretiert man die Standardabweichung fälschlich als „Fehler“, dann scheint die Messung von L.Azy sogar „besser“ zu sein. Der wirkliche Unterschied liegt aber im Standardfehler des Mittelwerts, der bei E.Ager mit seiner vierfachen Zahl an Messungen nur etwa halb so groß ist. Und das wird spätestens bei einem Signifikanztest ganz entscheidend sein...

## Fehlerrechnung

Häufig will man mit experimentell bestimmten Daten weiterarbeiten und aus diesen zum Beispiel abgeleitete Werte bestimmen. Rechnet man mit fehlerbehafteten Werten, dann muss man allerdings beachten, dass sich die Fehler beim Rechnen meist auch verändern. Für eine Abschätzung des resultierenden Fehlers wird dabei gewöhnlich die sogenannte Gauß'sche Fehlerfortpflanzung verwendet. Allerdings ist diese streng genommen nur dann sinnvoll, wenn die Werte normalverteilt sind, die Fehler rein statistisch sind (und keine Mischung aus statistischem Fehler und abgeschätztem systematischem Fehler) und wenn die Fehler im Vergleich zum Messwert relativ klein sind – anders ausgedrückt: Bei schlechten Messwerten wird alles bei der Fehlerrechnung noch einmal schlechter...

Für die häufigsten Rechenarten folgt jetzt eine Übersicht über die entsprechende Gauß'sche Fehlerfortpflanzung – dabei ist anzumerken, dass je nach Rechenart die Arbeit mit absoluten oder relativen Fehlern einfacher sein kann. Bei komplizierteren Rechnungen empfiehlt es sich, die resultierenden Fehler für jeden Zwischenschritt auszurechnen und nicht den deutlich komplizierteren Weg zu wählen, alles in einem Schritt auszurechnen! Für die matheaffineren

Leser gibt es am Ende dieses Kapitels noch eine kurze Übersicht dazu, wie sich die Angaben in der Tabelle herleiten lassen.

In der folgenden Tabelle sind  $\bar{x}$  und  $\bar{y}$  Werte mit den Fehlern  $\Delta\bar{x}$  und  $\Delta\bar{y}$  während  $a$  ein Wert ohne Fehler ist. Beachte, dass beim **Normieren** von Daten ein Referenzwert (normalerweise die Kontrolle) als 1 oder 100% gesetzt wird, also alle Werte durch diesen geteilt werden. Das ist aber ein Teilen durch den absoluten Mittelwert des und nicht durch den Wert mit Fehler – und natürlich muss dabei der Fehler des Referenzwerts mitgeführt werden (Beispiel: In einem Experiment werden die Werte Kontrolle:  $4,0 \pm 0,4$  Experiment:  $8,0 \pm 0,8$ . Um auf den Kontrollwert zu normieren würde man alle Werte und Fehler durch 4 teilen und erhielte die folgenden relativen Werte: Kontrolle  $1,0 \pm 0,1$  Experiment:  $2,0 \pm 0,2$ )

### Beispiele mit einer fehlerbehafteten Variable

Funktion	Absoluter Fehler	Relativer Fehler
$f(\bar{x})$	$\Delta f(\bar{x})$	$\frac{\Delta f(\bar{x})}{f(\bar{x})}$
$\bar{x} + a$	$\Delta\bar{x}$	$ \frac{\Delta\bar{x}}{\bar{x} + a} $
$\bar{x} - a$	$\Delta\bar{x}$	$ \frac{\Delta\bar{x}}{\bar{x} - a} $
$\bar{x} * a$	$ a * \Delta\bar{x} $	$ a * \frac{\Delta\bar{x}}{\bar{x}} $
$\frac{\bar{x}}{a}$	$ \frac{\Delta\bar{x}}{a} $	$ \frac{1}{a} * \frac{\Delta\bar{x}}{\bar{x}} $
$\bar{x}^a$	$ a\bar{x}^{a-1} * \Delta\bar{x} $	$ a * \frac{\Delta\bar{x}}{\bar{x}} $

### Beispiele mit zwei fehlerbehafteten Variablen

Funktion	Absoluter Fehler	Relativer Fehler
$f(\bar{x}, \bar{y})$	$\Delta f(\bar{x}, \bar{y})$	$\frac{\Delta f(\bar{x}, \bar{y})}{f(\bar{x}, \bar{y})}$
$\bar{x} + \bar{y}$	$\sqrt{(\Delta\bar{x})^2 + (\Delta\bar{y})^2}$	$\frac{\sqrt{(\Delta\bar{x})^2 + (\Delta\bar{y})^2}}{\bar{x} + \bar{y}}$
$\bar{x} - \bar{y}$	$\sqrt{(\Delta\bar{x})^2 + (\Delta\bar{y})^2}$	$\frac{\sqrt{(\Delta\bar{x})^2 + (\Delta\bar{y})^2}}{\bar{x} - \bar{y}}$
$\bar{x} * \bar{y}$	$\sqrt{(\bar{y} * \Delta\bar{x})^2 + (\bar{x} * \Delta\bar{y})^2}$	$\sqrt{(\frac{\Delta\bar{x}}{\bar{x}})^2 + (\frac{\Delta\bar{y}}{\bar{y}})^2}$

$$\frac{\bar{x}}{\bar{y}} \quad \sqrt{\left(\frac{\Delta\bar{x}}{\bar{y}}\right)^2 + \left(\frac{\bar{x} * \Delta\bar{y}}{\bar{y}^2}\right)^2} \quad \sqrt{\left(\frac{\Delta\bar{x}}{\bar{x}}\right)^2 + \left(\frac{\Delta\bar{y}}{\bar{y}}\right)^2}$$


---

## Anhang: Herleitung der Fehlerfortpflanzungsregeln

Für eine Funktion mit mehreren Variablen  $f(x_i)$  gilt nach der Gauß'schen Fehlerfortpflanzung, dass der Fehler der Funktion die Wurzel aus der Summe der Quadrate der Fehler der einzelnen Variablen multipliziert mit der Ableitung der Funktion nach eben diesen Variablen ist, also:

$$\Delta f(\bar{x}_i) = \sqrt{\sum_{i=1}^n \left(\frac{\partial \bar{x}_i}{\partial f} \Delta \bar{x}_i\right)^2}$$

oder für zwei Werte x und y:

$$\Delta f(\bar{x}, \bar{y}) = \sqrt{\left(\frac{\partial \bar{x}}{\partial f} \Delta \bar{x}\right)^2 + \left(\frac{\partial \bar{y}}{\partial f} \Delta \bar{y}\right)^2}$$

Funktion	Ableitung nach x	Ableitung nach y	Fehler der Funktion
$f(\bar{x}, \bar{y})$	$\frac{\partial \bar{x}}{\partial f}$	$\frac{\partial \bar{y}}{\partial f}$	$\Delta f(\bar{x}, \bar{y})$
$\bar{x}^a$	$a * \bar{x}^{a-1}$	–	$ a\bar{x}^{a-1} * \Delta\bar{x} $
$\bar{x} + \bar{y}$	1	1	$\sqrt{(\Delta\bar{x})^2 + (\Delta\bar{y})^2}$
$\bar{x} - \bar{y}$	1	–1	$\sqrt{(\Delta\bar{x})^2 + (\Delta\bar{y})^2}$
$\bar{x} * \bar{y}$	y	x	$\sqrt{(y * \Delta\bar{x})^2 + (\bar{x} * \Delta\bar{y})^2}$
$\frac{\bar{x}}{\bar{y}}$	$\frac{1}{\bar{y}}$	$-\frac{\bar{x}}{\bar{y}^2}$	$\sqrt{\left(\frac{\Delta\bar{x}}{\bar{y}}\right)^2 + \left(\frac{\bar{x} * \Delta\bar{y}}{\bar{y}^2}\right)^2}$

---

## Darstellung von Ergebnissen mit Streuung

### Angabe von Werten im Text

Ein mit Messungenauigkeit behafteter Wert wird meist in der Form  $\bar{x} \pm \Delta\bar{x}$  angegeben. Wichtig ist dabei allerdings, den Stichprobenumfang (n) anzugeben und im Zweifelsfall klar zu machen, was  $\Delta x$  ist (Standardabweichung, Standardfehler). Bei tabellarischen Übersichten oder längeren Arbeiten können diese Angaben natürlich auch in der Tabellenunterschrift oder zentral in einem Material- und Methodenteil gemacht werden.

Sinnvoll ist es auch, den Fehler mit ein oder zwei Stellen anzugeben und den Mittelwert dann auf die letzte Stelle des Fehlers zu runden, also z.B.  $4,56 \pm 0,12$  oder  $4,6 \pm 0,1$  und nicht  $4,563456 \pm 0,1$  oder  $4,563456 \pm 0,125476$  (hier täuschen die vielen Stellen eine offensichtlich nicht vorhandene Genauigkeit vor) und auch nicht  $4,5 \pm 0,0001$  (hier ist der Mittelwert viel ungenauer angegeben als die Messung es eigentlich erlauben würde!). Eine sinnvolle Genauigkeit sollte man auch bei der Angabe als relativem Fehler beachten, wobei es sinnvoll ist, sich zu erinnern, dass 100% etwa einer, 10% zwei, 1% drei und so weiter Stellen Genauigkeit entspricht. Zum Beispiel wären hier wieder  $4,56 \pm 2,6\%$  oder  $4,5678 \pm 0,02\%$  sinnvolle Angaben während  $4,56784 \pm 5,3\%$  wieder eine zu hohe Genauigkeit des Mittelwerts vortäuscht und  $4,6 \pm 0,002\%$  sinnvolle Genauigkeit unterschlägt.

### **Graphische Darstellung von Stichproben (kommt noch)**

- \* Komplette,
- \* Balken mit Fehlerbalken
- \* Box-whisker plot

## **Die Geschichte mit der Signifikanz**

Ein Problem bei experimentell erhobenen Daten ist die Frage, ob zwei erhaltene Datensätze der gleichen Verteilung oder zwei unterschiedlichen entsprechen – anders ausgedrückt, ob die aus verschiedenen Versuchsreihen tatsächlich unterschiedlich sind oder vielleicht nur auf Grund der fehlerbehafteten Messungen unterschiedlich aussehen. Die Annahme, dass ein Messwert sich nicht von der Kontrolle unterscheidet wird dabei als **Nullhypothese  $H_0$**  bezeichnet und das Ziel ist die Antwort auf die Frage, ob  $H_0$  verworfen werden und stattdessen die Alternativhypothese „Da ist etwas anders“ angenommen werden darf.

Als Maß hierfür dient die **Signifikanz** – dabei wird ein Unterschied dann als **signifikant** bezeichnet, wenn die Wahrscheinlichkeit, dass er zufällig zustande gekommen ist unter einer vorher festgelegten Schwelle liegt. Diese Schwelle wird auch als **Signifikanzniveau  $\alpha$**  bezeichnet, die Wahrscheinlichkeit, dass zwei Messwerte in Wirklichkeit gleich sind als **Irrtumswahrscheinlichkeit  $p$**  – ist ein Unterschied signifikant, dann ist die Wahrscheinlichkeit, sich zu irren wenn man einen echten Unterschied annimmt, also  $H_0$  verwirft, kleiner als das Signifikanzniveau. Ein paar Signifikanzniveaus sind dabei besonders gebräuchlich, wobei die sprachlichen Umschreibungen nicht scharf definiert sind:

- Ist  $p < 0,05$  (5%) spricht man von einem „signifikanten Unterschied“, mit  $p < 0,05$  von einem kontrollwert abweichende Wert werden häufig mit \* markiert
- Ist  $p < 0,01$  (1%) spricht man von einem „sehr signifikanten Unterschied“, markiert wird mit \*\*

- Ist  $p < 0,001$  (0,1%) spricht man von einem „hoch signifikanten Unterschied“, markiert wird mit \*\*\*

Signifikanztests sind ein unverzichtbares Mittel, um experimentell erhobene Daten objektiv zu vergleichen und so „echte Unterschiede“ von zufälligen Schwankungen durch Messungen zu unterscheiden. Mehr können sie allerdings nicht. Insbesondere sollte man die folgenden Fallstricke beachten:

- **Ein signifikanter Unterschied ist nicht absolut**, sondern nur mit einer als vertretbar angesehenen Wahrscheinlichkeit echt. Bei  $p < 0,05$  sind immer noch bis zu 5% der signifikanten Unterschiede zufällig entstanden, also 1 von 20 (Der Irrtum wäre damit auf Medikamentenbeipackzetteln eine häufige Nebenwirkung!). Bei 10 mit  $p < 0,05$  signifikanten Unterschieden in einer Veröffentlichung wäre damit die Wahrscheinlichkeit, dass alle Unterschiede echt sind bei nur noch etwa 60%!
- **Signifikanztests sehen nur statistische Fehler!** Alle anderen Fehler, die zum Beispiel in der Experimentplanung und –durchführung anfallen werden hier übersehen – gerade bei nicht-verblindeten Tests ist das Risiko, dass der Experimentator z.B. Kontrolle und Probe nicht exakt gleich behandelt schwer abzuschätzen. Die echte Irrtumswahrscheinlichkeit wird also immer schwer abschätzbar höher sein, als im Signifikanztest berechnet...
- **Signifikanz testet nur die Hypothese, dass sich zwei Messreihen unterscheiden!** Das ist zwar für die Überlegung ob die Theorie dazu, warum sich die Reihen unterscheiden äußerst wichtig, genau genommen ist aber die Aussage „Stoff X senkt Wert Y in Zelltyp Z signifikant“ falsch, denn getestet hat man nur „Bei Zelltyp Z haben wir nach Behandlung mit Stoff X einen signifikant niedrigeren Wert Y gefunden“ – dass dieser Effekt auf Stoff X zurückgeht oder gar auf dem angenommenen Wirkprinzip beruht kann der Signifikanztest nicht abschätzen (z.B. könnte die ganze Stoffklasse zu der X gehört wirksam sein, X mit Stoff V verunreinigt gewesen sein, X die Wirkung auf ganz anderem Weg als gedacht bewirken usw.)
- **Signifikanztests prüfen nur Fehler erster Art!** Die Irrtumswahrscheinlichkeit  $\alpha$  gibt nur an, wie wahrscheinlich es ist, dass die Annahme es bestände ein unterschied zwischen zwei Werten falsch ist (Also falsches ablehnen der Nullhypothese). Fehler zweiter Art, also solche bei denen die Nullhypothese nicht verworfen wird, obwohl sie falsch ist, lassen sich so nicht abschätzen. So ließe sich mit einem Signifikanztest also die Hypothese  $H_1$  „Medikament A wirkt besser als Medikament B“ untermauern, die Hypothese  $H_0$  „Medikament A und B wirken gleichgut“ aber nicht. Dass das nicht möglich ist liegt daran, dass zwischen den beiden Bereichen „ $H_1$  stimmt ziemlich sicher“ (durch Signifikanztest testbar) und „ $H_0$  stimmt ziemlich sicher“ eben noch der große Bereich „Sicher ist hier nichts“ liegt.

- **Signifikanztests testen nur vorher aufgestellte Hypothesen sinnvoll.** Ob eine vor dem Experiment aufgestellte Hypothese durch das Experiment untermauert wird lässt sich mit einem Signifikanztest abschätzen. Rein rechnerisch kann man natürlich auch einfach drauflosmessen, dann Signifikanztests machen und zu den auffälligen Werten dann Hypothesen aufstellen. Hier gibt der Signifikanztest aber dann keinen Aufschluss mehr über die Irrtumswahrscheinlichkeit, da die Werte ja aus „bestandenen Tests“ ausgewählt wurden – zur Erinnerung bei  $p < 0,05$  weist eins von 20 zufälligen Ergebnissen einen signifikanten Unterschied auf! Dieses „Suchen signifikanter Unterschiede“ kann zwar wissenschaftlich extrem nützlich sein, da es Hypothesen generiert, getestet werden müssen diese dann aber in einem weiteren, unabhängigen Experiment! Mehr dazu im Kapitel „Wie man mit Statistik keinen Unsinn macht...“

### **t-Tests (kommt noch)**

Was testet man?

Wie funktioniert es?

Worauf achten?

### **Chi-Quadrat-Test (kommt noch)**

Was testet man?

Wie funktioniert es?

Worauf achten?

### **Wie man mit Statistik keinen Unsinn macht... (kommt noch)**

Sinnvolle Angabe von Werten

Irrtumswahrscheinlichkeiten senken

Hypothesen testen und generieren

### **Tipps zum Weiterlesen und Surfen**

# Register

<b>D</b>		<b>M</b>	
Durchschnitt .....	5	Median .....	6
<b>E</b>		Mittel, arithmetisches .....	5, 8
Erwartungsbereich .....	8	Mittelwert .....	5
<b>F</b>		<b>N</b>	
Fehler		Normieren .....	10
absoluter .....	7	Nullhypothese .....	12
erster Art .....	13	<b>S</b>	
grobe .....	3	Signifikanz .....	12
mittlerer der Einzelwerte .....	7	Signifikanzniveau .....	12
mittlerer des Mittelwerts .....	7	Standardabweichung .....	7
relativer .....	7	empirische .....	7, 8
statistische .....	3	Grundgesamtheit .....	8
systematische .....	2	Standardfehler .....	7, 8
zufällige .....	3	Stichprobe .....	2
zweiter Art .....	13	Stichprobenumfang .....	3
Fehlerrechnung .....	9	<b>V</b>	
<b>G</b>		Varianz .....	6
Glockenkurve .....	4	Verteilung	
<b>I</b>		diskrete .....	4
Irrtumswahrscheinlichkeit .....	8, 12	Gauß-Verteilung .....	4
<b>K</b>		Normalverteilung .....	4
Konfidenzintervall .....	8	schiefe .....	5
		stetige .....	4
		Vertrauensbereich .....	8
		<b>Z</b>	
		Zentralwert .....	6