t-test:

Wenn man untersucht, ob sich zwei Messwertreihen unterscheiden (z.B. Formantwerte bei Männern und Frauen des Vokals [a]), dann wird man sich die Verteilung der Daten in einem Histogramm ansehen und die Mittelwerte (um zu sehen, wo die Daten im Durchschnitt liegen) und Standardabweichungen (um zu sehen, wie stark die Messungen schwanken; bzw., wie dicht sie am Mittelwert liegen) berechnen. Man weiß dann aber nicht, ob ein Unterschied nur **zufällig** zwischen den Mittelwerten besteht, oder ob man diesem Unterschied eine gewisse **Signifikanz** zuschreiben darf. Man hat mit Mittelwert und Standardabweichung zwar die Daten **beschrieben**, aber man kann nicht **testen**, ob der Unterschied signifikant ist. Hier helfen statistische **Testverfahren**, wie der sogenannte ***t*-test**, der die Wahrscheinlichkeit berechnet, dass ein Unterschied statistisch signifikant ist, oder nur zufällig besteht. Der folgende Text versucht zu motivieren, wie der *t*-test das macht, ohne dabei durch die Mathematik zu gehen, die dahinter steht. (Achtung: der *t*-test, wie alle statistischen Tests, gibt nur eine Wahrscheinlichkeit an, und kann nicht sagen, ob ein Unterschied wirklich besteht.)

1. Überlegung:

Stellen Sie sich vor, Sie kennen alle Daten einer Verteilung (z.B. alle [a] aller deutschen Sprecher – das geht natürlich nicht; stellen Sie sich aber einen hypothetischen Überwachungsstaat vor, der alles was jede/r jemals gesagt hat aufgezeichnet hätte). Diese Originaldaten hätten eine gewisse Verteilung und man könnte den Mittelwert und die Standardabweichung davon berechnen. In Wirklichkeit kennen Sie diesen Mittelwert und die Standardabweichung aber nicht, obwohl er besteht. Stellen Sie sich jetzt vor, Sie könnten völlig zufällig 10 Werte aus dieser Grundgesamtheit ziehen. Vermutlich würden Sie einige Werte oberhalb und andere Werte unterhalb des originalen, 'wirklichen' Mittelwertes ziehen; es könnte auch sein, dass Sie gerade die 10 kleinsten oder 10 größten Werte erwischen würden, aber das ist äußerst unwahrscheinlich (aber nicht ausgeschlossen). Der Mittelwert dieser 10 Werte wird irgendwo zwischen den kleinsten und größten Werten der Originaldaten liegen; wahrscheinlich aber eher in der Gegend des wirklichen Mittelwertes liegen (weil vermutlich kleine und große Werte zufällig gezogen werden – wie beim Lotto, wo die gezogenen Zahlen häufig unterhalb und oberhalb des Mittelwertes '25' liegen). Das heißt, obwohl sie nichts über den wirklichen Mittelwert wissen, können Sie annehmen, dass der Mittelwert der 10 zufällig gezogenen Zahlen irgendwo in der Gegend des wirklichen Mittelwertes liegt.

2. Überlegung

Stellen Sie sich weiter vor, sie könnten sehr sehr oft 10 solcher Werte ziehen (und immer wieder die gezogenen Zahlen zurücklegen – so, wie im Lotto zweimal pro Woche 6 Zahlen aufs Neue gezogen werden). Da die gezogenen Zahlen wahrscheinlicher dichter am wirklichen Mittelwert liegen, und nur selten alle 10 Zahlen am unteren oder oberen Rand der (wirklichen) Verteilung liegen, folgen die Mittelwerte der zufällig 'gezogenen' Zahlen einer Normalverteilung um den wirklichen Mittelwert herum (das lässt sich mathematisch beweisen; aber dass die Werte um den wirklichen Mittelwert herumliegen ist hoffentlich 'einsichtig'). Das ist etwas sehr Überraschendes: obwohl die wirkliche Verteilung 'irgendwie' aussehen kann, liegen die Mittelwerte der zufällig gezogenen Werte in einer Normalverteilung um den wirklichen Mittelwert herum! Dabei hängt es von zwei Faktoren ab, 'wie dicht' die Mittelwerte der zufällig gezogenen Zahlen beim wirklichen Mittelwert liegen: wenn die Originaldaten sehr stark streuen (d.h., eine große Standardabweichung haben), dann werden die zufällig gezogenen Daten auch stärker streuen und die aus ihr berechneten Mittelwerte stärker streuen, als wenn die Originaldaten alle dicht beim wirklichen Mittelwert liegen (denn dann können die zufällig gezogenen Daten auch nicht weit abweichen). Der zweite Faktor ist die Anzahl der zufällig gezogenen Zahlen: wenn es sehr viele sind (z.B. 100 oder 1000) dann wird deren Mittelwert dichter am 'wahren' Mittelwert der Originaldaten liegen, weil es unwahrscheinlicher ist, dass alle 100 (oder 1000) Werte unterhalb oder oberhalb des wirklichen

Mittelwertes liegen. Wenn man dagegen nur zufällig drei Werte zieht, dann wird es häufiger vorkommen, dass der Mittelwert der gezogenen Zahlen weiter vom wirklichen Mittelwert abweicht. Die Abbildungen 1a-d stellen das graphisch dar.

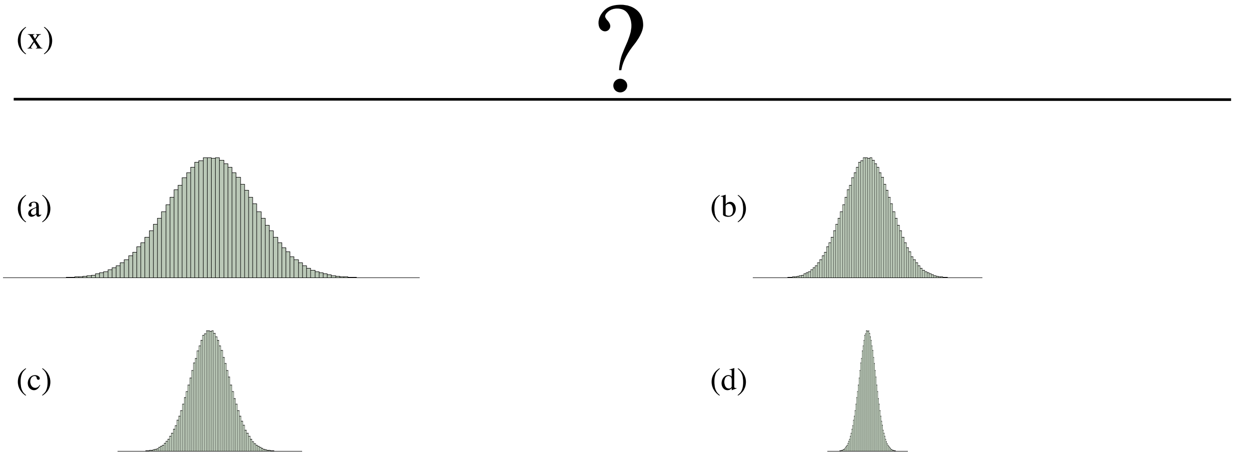

Abb. 1: (x) Originale (unbekannte) Verteilung, die einen (unbekannten) Mittelwert und eine (unbekannte) Standardabweichung hat. (a) Verteilung der Mittelwerte von jeweils 10 zufällig gezogenen Werten aus einer Originalverteilung mit einer großen Standardabweichung/Varianz. (b) Wie (a), aber die Originaldaten haben eine kleine Varianz. (c) Wie (a), aber es wurden 100 Daten zufällig gezogen. (d) Wie (c), aber wiederum mit wenig streuenden Originaldaten.

3. Überlegung

Wir wissen zwar nichts über die Verteilung der Originaldaten (1x), aber wir wissen, dass die zufällig gezogenen Werte in einer Glockenkurve (= (gauß'chen) Normalverteilung) um den wirklichen Mittelwert liegen (1a-d). Wenn man jetzt wiederum zufällig aus dieser Normalverteilung zwei Werte zieht, dann wird die Differenz dieser beiden Werte auch wieder einer Normalverteilung folgen (nach denselben Überlegungen wie Überlegung 1 und 2), allerdings mit dem Mittelwert '0' (weil die beiden zufällig gezogenen Mittelwerte der aus den Verteilungen der zufällig gezogenen Werte (1a-d) oberhalb und unterhalb des wirklichen Mittelwertes (von 1x) liegen, und ihre Differenz mal positiv, mal negativ und auch häufig mal dicht bei oder wirklich '0' sein können). Dies ist in Abbildung 2a-d dargestellt: auch hier werden die Differenzen der gezogenen Mittelwerte um so weniger streuen (d.h., um so dichter bei '0' liegen), wenn die Originaldaten wenig streuten (2b, d) und die Zufallsstichproben groß waren (2c, d).

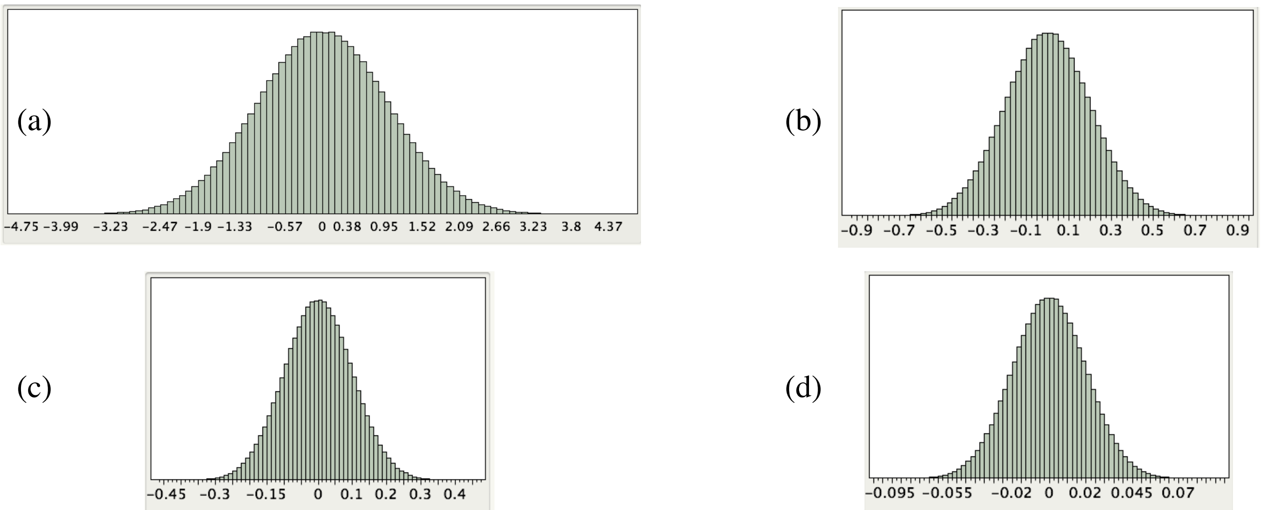

Abb. 2: Verteilungen der Differenzen von zwei zufällig gezogene Mittelwerten aus 1a-d (beachten Sie bitte die unterschiedlichen Skalen der vier Verteilungen!).

4. Der *t*-test

Diese Überlegungen lassen sich mathematisch exakt formulieren, und man kann deswegen Wahrscheinlichkeiten berechnen, wie weit zwei Mittelwerte einer Zufallsstichprobe voneinander abweichen, obwohl man die Originaldaten nicht kennt. Nur aufgrund (i) der Streuung der (zufälligen) Messwerte, die man als Zufallsstichprobe aus den Originaldaten zieht, und (ii) der Größe der Zufallsstichprobe kann der *t*-test berechnen, wie zufällig der Unterschied zwischen zwei so gefundenen Mittelwerten ist, da diese Mittelwerte normalverteilt sein *müssen* (obwohl die Originaldaten es nicht überhaupt nicht sein müssen!). Die Mittelwerte von zwei Gruppen von zufällig gezogenen Werte werden nur eine geringe Wahrscheinlichkeit haben, weit voneinander zu liegen – es hat sich in der Linguistik (wie in vielen Sozialwissenschaften) eingebürgert, dass man sagt, wenn es nur in 5% oder weniger aller Fälle zufällig vorkommen kann, dass sie weit auseinanderliegen, dass der Unterschied **auf dem 5%-Niveau signifikant** ist (wobei die Berechnung dieses 'Abstands' von der der Streuung in den Originaldaten und der Stichprobengröße abhängt). Wenn man jetzt aber seine beiden Mittelwerte nicht zufällig 'gezogen' hat, sondern z.B. ganz bewusst einmal Werte von Männern und einmal von Frauen genommen hat, dann möchte man, dass es recht unwahrscheinlich ist, dass ein beobachteter Unterschied nur zufällig ist – man möchte dass die beiden Werte möglichst weit auseinanderliegen, also eine sehr geringe (kleiner 5%) Wahrscheinlichkeit haben, nur durch Zufall so unterschiedlich zu sein. Man will also möglichst kleine Prozentwerte für die Wahrscheinlichkeit eines signifikanten Mittelwertunterschieds haben (z.B. 1% oder noch kleiner).

Wahrscheinlichkeiten & Prozente:

Wahrscheinlichkeiten werden i.d.R. mit 'p' abgekürzt und die Werte liegen zwischen 0 und 1, was 0% bis 100% entspricht: man multipliziert einfach die Wahrscheinlichkeiten mit '100' und hat dann die Wahrscheinlichkeiten in Prozent. Zum Beispiel: $p = 0.37$ entspricht einer Wahrscheinlichkeit von 37%; $p = 0.002$ heißt 0.2%; $p <= 0.0001$ sind weniger als 0.01%.

Typen von *t*-tests:

Es gibt drei Arten von *t*-tests:

1- vs. 2-seitiger (*one vs. two tailed*) *t*-test: Wenn man aus theoretischen Überlegungen oder aus vorherigen Untersuchungen weiß, dass einer der beiden Mittelwerte größer oder kleiner als der andere Mittelwert sein *muss*, dann benutzt den 1-seitigen *t*-test; wenn man nicht sicher weiß, dass der Unterschied in eine Richtung gehen muss, dann benutzt man den 2-seitigen *t*-test.

Mit gleicher oder ungleicher Varianz (*equal vs. unequal variance*): Wenn aus irgendwelchen Gründen davon ausgegangen werden muss, dass die beiden Mittelwerte aus Verteilungen mit unterschiedlichen Streuungen stammen (z.B. Messwerte von Gesunden und Leuten mit einer Pathologie, die nahelegt, dass es große Streuungen in deren Messwerten geben wird), dann benutzt man den *t*-test für unterschiedliche Varianzen, sonst den mit gleicher Varianz.

Gepaarter/gebundener vs. ungepaarter/ungebundener (*paired vs. unpaired*) *t*-test: Wenn zwei Messungen 'verbunden' (= 'gepaart') sind (z.B. ein Patient vor und nach einer Behandlung), dann benutzt man den gepaarten *t*-test, sonst immer den ungepaarten.

In der Regel benutzt man in der Phonetik den 2-seitigen, ungepaarten *t*-test für gleiche Varianzen.

t-Werte:

Der *t*-test benutzt intern einen '*t*-Wert', der in Abhängigkeit von der Stichprobengröße in *p*-Werte (Wahrscheinlichkeiten) oder Prozente umgerechnet wird. Gewisse Programme (z.B. Excel) geben diesen *t*-wert gar nicht an.

Wichtig:

Der *t*-test berechnet zwar die Wahrscheinlichkeit, dass ein Unterschied der Mittelwerte signifikant ist. Man darf aber **nicht** nur die beiden Mittelwerte in die Berechnung einfließen lassen, sondern muss **alle** Werte angeben (z.B. in Excel), weil die Programme den Umfang der Stichprobe (z.B. 3, 10, 100 oder so) und deren Verteilung bewerten muss.

Vorgehen für eine statistische Auswertung:

Für eine Untersuchung wird man meist in Excel (oder einem anderen Tabellenkalkulations-programm/*spreadsheet*) die Daten in ein oder zwei Spalten haben, daraus in ein oder zwei Spalten die Häufigkeiten für Daten in einzelnen Intervallen/Klassen zählen, von denen man dann wiederum die Histogramme zeichnen kann. Aus den Daten wird man auch die Mittelwerte und Standardabweichungen berechnen, damit man sie selber sehen kann, und man wird alle Daten in den *t*-test geben, der einen Wert ausgibt, der hoffentlich kleiner als 0.05 ist (weil dann die Wahrscheinlichkeit, dass beide Mittelwerte zufällig so unterschiedlich sind, unter 5% liegt; man also sagen kann, die beiden Mittelwerte sind auf dem 5% Niveau signifikant unterschiedlich.