3.1 Was ist der t-Test?

Das folgende Unterkapitel erklärt schrittweise die Fragestellung und Funktion eines t-Tests und die benötigten theoretischen Grundlagen: die Nullhypothese, die Stichprobenkennwerteverteilung und deren Streuung (Kap. 3.1.1 und 3.1.2). Dies führt zur Entwicklung der Formel für den t-Wert und der t-Verteilung. Die Form dieser Verteilung wird durch ihre Freiheitsgrade bestimmt (Kap. 3.1.3). Die weiteren Abschnitte befassen sich mit der Auswertung eines empirisch ermittelten t-Werts, mit dem Einfluss der Stichprobengröße sowie mit den Voraussetzungen, die für die Durchführung eines t-Tests gegeben sein müssen (Kap. 3.1.4 bis 3.1.9).

3.1.1 Die Fragestellung des t-Tests

Der t-Test kann nur bei intervallskalierten Daten angewendet werden. Er gehört zur Gruppe der parametrischen Verfahren.

Der t-Test ist eine Entscheidungsregel auf einer mathematischen Grundlage, mit deren Hilfe ein Unterschied zwischen den empirisch gefundenen Mittelwerten zweier Gruppen näher analysiert werden kann. Er liefert nur für intervallskalierte Daten zuverlässige Informationen. Deshalb gehört er zur Gruppe der parametrischen Verfahren.

Parametrische Verfahren schätzen Populationsparameter mittels statistischer Kennwerte wie dem arithmetischen Mittel oder der Varianz, für deren Berechnung die Intervallskaliertheit der Daten Voraussetzung ist.

Der t-Test untersucht, ob sich die Mittelwerte zweier Gruppen systematisch unterscheiden.

Der t-Test arbeitet mit den Populationsparametern der Streuung und des arithmetischen Mittels, die mit Hilfe der Stichprobe geschätzt werden. Er liefert eine Entscheidungshilfe dafür, ob ein gefundener Mittelwertsunterschied rein zufällig entstanden ist, oder ob es wirklich bedeutsame Unterschiede zwischen den zwei untersuchten Gruppen gibt. Mathematisch gesprochen beurteilt dieses Verfahren, ob sich zwei untersuchte Gruppen systematisch in ihren Mittelwerten unterscheiden oder nicht.

Der Stichprobenkennwert des t-Tests ist die Differenz der Mittelwerte.

Der wichtigste Wert für die Durchführung eines t-Tests ist die Differenz der Gruppenmittelwerte. Diese Differenz bildet den Stichprobenkennwert des t-Tests:

$$\bar{x}_1 - \bar{x}_2$$

Die zentrale Frage des t-Tests lautet: Wie wahrscheinlich ist die empirisch gefundene oder eine größere Mittelwertsdifferenz unter allen möglichen rein theoretisch denkbaren Differenzen (Abb. 3.1)?

Der t-Test dient wie viele andere statistische Verfahren zur Überprüfung aufgestellter Hypothesen. Dabei ist es wichtig, vor der Durchführung eines t-Tests die zu untersuchende Hypothese inhaltlich zu präzisieren. Die inhaltliche Hypothese muss dann in eine mathematische Schreibweise gebracht und somit in eine statistische Hypothese überführt werden. Der t-Test prüft damit, ob diese statistische Hypothese zutrifft.

Betrachten wir die Entwicklung der Fragestellung für einen t-Test anhand des in der Einführung beschriebenen Gedächtnisexperiments. In der Einleitung dieses Buches wurden die inhaltlichen Hypothesen für die verschiedenen Verarbeitungsbedingungen vorgestellt. Bei einer strukturellen Verarbeitung sollten weniger Wörter als bei bildhafter bzw. emotionaler Verarbeitung erinnert werden. Zwischen bildhafter und emotionaler Verarbeitung sollte kein Unterschied in der Erinnerungsleistung auftreten. Die empirisch gefundenen Mittelwerte der einzelnen Bedingungen sind:

Verarbeitungsbedingung:	Anzahl erinnerter Wörter
strukturell	$\bar{x}_{\text{strukturell}} = 7{,}2$
bildhaft	$\bar{x}_{\text{bildhaft}} = 11$
emotional	$\bar{x}_{\text{emotional}} = 12$

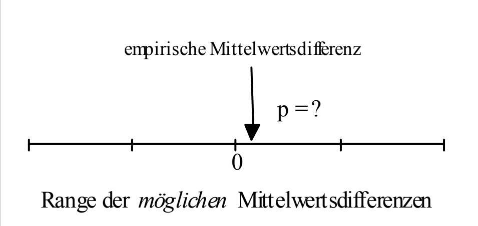
Da der t-Test jeweils nur zwei Gruppen betrachten kann, greifen wir den Mittelwertsvergleich zwischen bildhafter und struktureller Verarbeitung heraus und wandeln die inhaltliche Hypothese in eine statistische um. Auf dieses Beispiel werden sich die meisten Rechnungen in den folgenden Abschnitten beziehen:

$$\bar{x}_{\text{bildhaft}} - \bar{x}_{\text{strukturell}} > 0$$

$\bar{x}_{\text{bildhaft}}$: Mittelwert der erinnerten Wörter unter bildhafter Verarbeitung

$\bar{x}_{\text{strukurell}}$: Mittelwert der erinnerten Wörter unter struktureller Verarbeitung

Abb. 3.1. Fragestellung des t-Tests



Die inhaltliche Hypothese muss in eine statistische Hypothese umgewandelt werden.

Download der Daten unter:

http://www.quantitative-methoden.de

Die Formulierung der statistischen Hypothese bestimmt die Bildung der Differenz der Mittelwerte.

Die erhobenen Daten erlauben nun die Bestimmung des zu prüfenden Stichprobenkennwerts. Die Bildung der Differenz wird entscheidend durch die Formulierung der statistischen Hypothese mitbestimmt: Sie legt fest, welcher Wert vom anderen abgezogen wird.

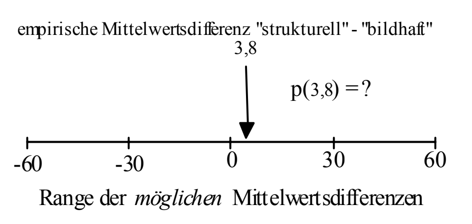
$$\bar{x}_{\text{bildhaft}} - \bar{x}_{\text{strukturell}} = 11 - 7{,}2 = 3{,}8 > 0$$

Es wäre auch möglich, dieselbe inhaltliche Vorhersage umgekehrt in die statistische Hypothese zu übersetzen: Bei struktureller Verarbeitung werden weniger Wörter erinnert als bei bildhafter Verarbeitung.

$$\bar{x}_{\text{strukturell}} - \bar{x}_{\text{bildhaft}} < 0 \quad => \quad 7{,}2 - 11 = -3{,}8 < 0$$

Der Wert 3,8 ist größer als Null und bestätigt zumindest in der Tendenz die inhaltliche Vorhersage. Doch stellt sich noch immer die Frage, ob es systematische Unterschiede in der Erinnerungsleistung bei unterschiedlicher Verarbeitung gibt, oder ob der gefundene Unterschied zufällig aufgetreten ist.

Abb. 3.2. Einordnung der empirischen Mittelwertsdifferenz



Die maximal möglichen Differenzen liegen in dem Gedächtnisexperiment bei –60 und +60, denn es wurden insgesamt 60 Wörter präsentiert. Eine solche Maximaldifferenz träte auf, wenn die Gruppe mit bildhafter (struktureller) Erinnerung im Durchschnitt kein einziges, die Gruppe mit struktureller (bildhafter) Verarbeitung dagegen alle Wörter erinnert hätte. An dieser Stelle kommt der t-Test ins Spiel: Er gibt Auskunft darüber, wie wahrscheinlich ein Auftreten der Differenz von 3,8 oder einer größeren unter allen möglichen Differenzen ist (Abb. 3.2).

3.1.2 Die Nullhypothese

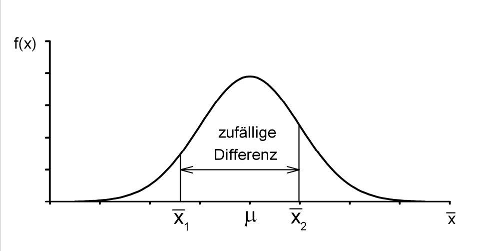
Die Nullhypothese (H_0) nimmt an, dass die Mittelwertsdifferenz zufällig entstanden ist.

Für die Erklärung der Mittelwertsdifferenz gibt es neben der Annahme eines systematischen Unterschieds zwischen den beiden Gruppen eine weitere Möglichkeit: Die Differenz zwischen den Mittelwerten ist zufällig zustande gekommen und es gibt keinen echten Unterschied zwischen den beiden untersuchten Gruppen. Die beiden Gruppen stammen im Grunde aus zwei Populationen mit demselben Mittelwert. Die Differenz zwischen den Gruppen sollte demzufolge Null betragen. Diese Annahme heißt deshalb Nullhypothese oder H_0.

Wieso kann es überhaupt zu einer Differenz der Stichprobenmittelwerte kommen, wenn diese Stichproben aus Populationen mit einem identischen Populationsmittelwert stammen? Ein solcher Unterschied auf Stichprobenebene ist deshalb möglich, weil die Stichprobenmittelwerte aufgrund der begrenzten Anzahl von Werten in einer Stichprobe fast nie genau dem Populationsmittelwert entsprechen, sondern mit einem Stichprobenfehler behaftet sind. Dieser ist in der Regel nicht besonders groß, denn Stichprobenmittelwerte sind erwartungstreue Schätzer des Populationsmittelwerts (Kap. 2.3). Allerdings ist es durchaus möglich, dass die Mittelwerte verschiedene Punkte auf einer Stichprobenkennwerteverteilung repräsentieren und so eine Differenz zwischen den Mittelwerten zustande kommt. Der Unterschied zwischen den beiden empirisch gefundenen Mittelwerten ist also noch kein Beweis dafür, dass die Stichproben aus zwei unterschiedlichen Populationen stammen.

Unter der Annahme der Nullhypothese beruht die Variation der Stichprobenmittelwerte also auf Zufall, oder anders gesagt, auf einem Stichprobenfehler (Abb. 3.3). Noch einmal: Die Nullhypothese postuliert, dass die Populationsmittelwerte der beiden Gruppen identisch sind und deshalb eine Mittelwertsdifferenz von Null zu erwarten ist.

Abb. 3.3. Stichprobenkennwerteverteilung von Mittelwerten mit zwei zufällig entstandenen Stichprobenmittelwerten sowie dem gemeinsamen Populationsmittelwert μ



Stichprobenkennwerteverteilung unter der Nullhypothese

Unter Annahme der Nullhypothese kann eine Stichprobenkennwerteverteilung von Mittelwertsdifferenzen konstruiert werden. In Kapitel 2 wurde die Stichprobenkennwerteverteilung von Mittelwerten bereits ausführlich behandelt. Zu jener besteht aber ein entscheidender Unterschied: In Kapitel 2 ist der interessierende Kennwert der Mittelwert einer Stichprobe, im jetzigen Fall betrachten wir die Differenz zweier Mittelwerte. Das bedeutet also, dass auf der Abszisse der Verteilung jetzt Mittelwertsdifferenzen abgetragen sind. Alle möglichen zwei Stichprobenmittelwerte, aus denen die Differenzen gebildet werden, stammen unter Annahme der Nullhypothese aus zwei Populationen mit identischem Populationsmittelwert.

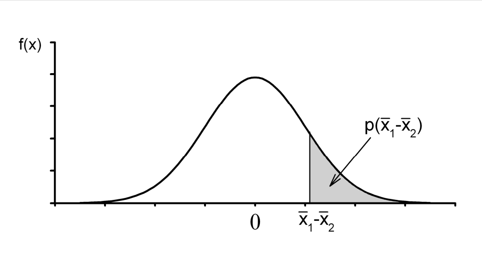
Wird aus zwei Populationen mit identischem Populationsmittelwert jeweils eine Stichprobe gezogen, so kann die Differenz der beiden

Stichprobenmittelwerte theoretisch jeden beliebigen Wert annehmen. Die zu erwartende Differenz aber ist gleich Null, denn die Stichprobenmittelwerte sind normalverteilt um ihren jeweiligen Erwartungswert, den Populationsmittelwert.

Standardfehler der Mittelwertsdifferenz

Da die Populationsmittelwerte identisch sind, wird sich die Mehrzahl der gefundenen Differenzen folglich in der Nähe von Null befinden. Aus diesen Überlegungen resultiert nach unendlich vielen Ziehungen von Stichproben eine Normalverteilung der Mittelwertsdifferenzen mit dem arithmetischen Mittel Null und einer von der Populationsstreuung und den Stichprobenumfängen abhängigen Streuung (vgl. Kap. 2). Diese Verteilung heißt Stichprobenkennwerteverteilung von Mittelwertsdifferenzen unter der Nullhypothese, ihre Streuung nennt sich Standardfehler von Mittelwertsdifferenzen. Diese Verteilung erlaubt die Bestimmung der Auftretenswahrscheinlichkeit des Bereichs einer empirisch gefundenen oder größeren Differenz (Abb. 3.4). Dadurch wird eine Bewertung der gefundenen Differenz möglich (Kapitel 2 beinhaltet dieselbe Argumentation für den Mittelwert, jetzt geht es um Mittelwertsdifferenzen.).

Abb. 3.4. Zuordnung einer Wahrscheinlichkeit zu einer empirischen Mittelwertsdifferenz



Hinweis: Wahrscheinlichkeiten lassen sich bei kontinuierlichen Verteilungen nur für Bereiche bestimmen (Kap. 2.1.2). Wenn in diesem Buch einem einzelnen Wert eine Wahrscheinlichkeit zugeordnet wird, ist dieser als Grenze eines Bereichs zu verstehen.

Die Abb. 3.5 zeigt eine Stichprobenkennwerteverteilung von Mittelwertsdifferenzen, die durch 26.000-maliges Ziehen von jeweils zwei Stichproben mit der Größe n = 40 entstanden ist. Die Stichproben entstammen identischen Populationen. Die zwei kleinen Graphiken zeigen die Populationsverteilungen. Nach der Berechnung der Stichprobenmittelwerte werden diese voneinander subtrahiert. Die entstandenen Differenzen bilden den Stichprobenkennwert. In einem Koordinatensystem wird dann abgetragen, wie häufig eine bestimmte Differenz auftritt. Diese große Graphik zeigt die resultierende Stichprobenkennwerteverteilung.

In der Computersimulation ist die Form der Normalverteilung deutlich erkennbar, der Mittelwert liegt mit 0,004 sehr nahe an dem von der Nullhypothese erwarteten Mittelwert von Null. Sehr kleine Differenzen um Null treten also am häufigsten auf, Differenzen größer als 8 oder –8 kommen so gut wie gar nicht vor.

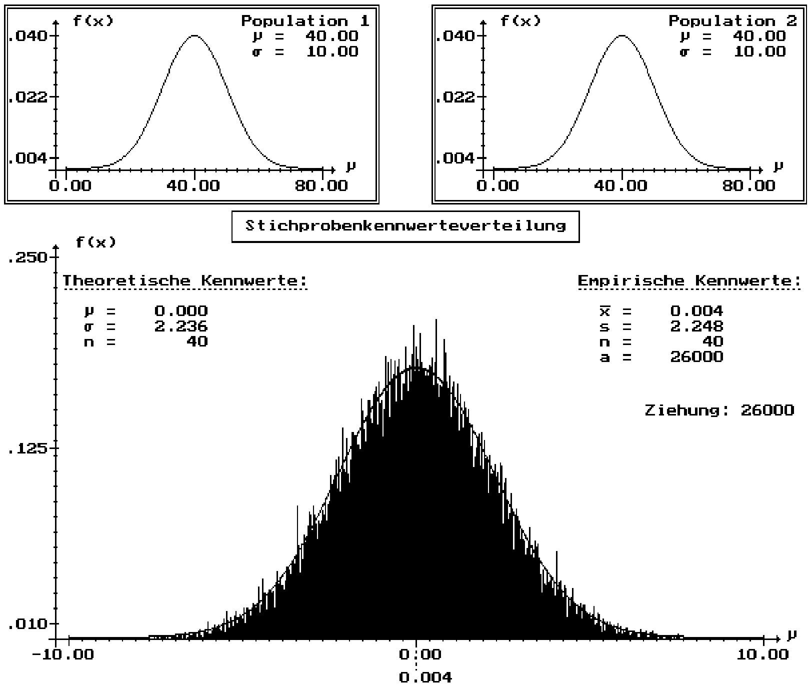


Abb. 3.5. Computersimulation einer Stichprobenkennwerteverteilung von Mittelwertsdifferenzen. Die Stichproben entstammen identischen Populationen.

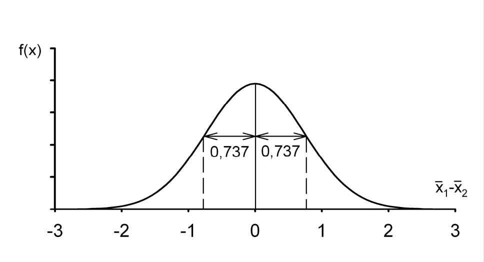
Für die Bestimmung der Stichprobenkennwerteverteilung ohne Simulation muss ihre Streuung (Standardfehler der Mittelwertsdifferenz) mit Hilfe der Stichprobe geschätzt werden. In die Formel gehen die Stichprobenumfänge der betrachteten Gruppen und die geschätzten Streuungen der zugehörigen Populationen ein. Die Formel lautet:

Formel für die Schätzung des Standardfehlers der Mittelwertsdifferenz

$$\hat{\sigma}_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{\hat{\sigma}_1^2}{n_1} + \frac{\hat{\sigma}_2^2}{n_2}}$$

$\hat{\sigma}_{\bar{x}_1-\bar{x}_2}$	:	geschätzter Standardfehler der Mittelwertsdifferenz
n_1	:	Anzahl der Vpn bzw. Beobachtungen in Stichprobe 1
$\hat{\sigma}_1^2$	:	geschätzte Varianz der Population 1
n_2	:	Anzahl der Vpn bzw. Beobachtungen in Stichprobe 2
$\hat{\sigma}_2^2$	:	geschätzte Varianz der Population 2

Abb. 3.6 . Stichprobenkennwerteverteilung von Mittelwertsdifferenzen unter der Annahme der H_0 mit $\hat{\sigma}_{\bar{x}_1-\bar{x}_2} = 0{,}737$



Ein Beispiel aus dem Gedächtnisexperiment: Für die Mittelwertsdifferenzen der Gruppen „bildhaft“ und „strukturell“ soll der Standardfehler der Mittelwertsdifferenz geschätzt werden. Die Stichprobengröße ist in beiden Verarbeitungsgruppen $n_1 = n_2 = 50$, die geschätzte Populationsstreuung der bildhaft enkodierenden Gruppe beträgt 4,14, die der strukturell verarbeitenden Gruppe 3,16.

$$\hat{\sigma}_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{4{,}14^2}{50} + \frac{3{,}162^2}{50}} = 0{,}737$$

Der Standardfehler der Mittelwertsdifferenz beträgt also unter der Annahme der Nullhypothese 0,737. Zusammen mit dem angenommenen Mittelwert von Null legt die Streuung die Form der Verteilung fest (Abb. 3.6). Bei größeren Stichproben verkleinert sich der Standardfehler der Mittelwertsdifferenz (Kap. 2.3).

3.1.3 Die t-Verteilung

t-Werte sind die standardisierten Differenzen der Stichprobenmittelwerte. Die Wahrscheinlichkeit eines t-Werts ist über die t-Verteilung bestimmbar.

Für die Bewertung der Auftretenswahrscheinlichkeit einer empirisch gefundenen Differenz ist ein standardisiertes Maß für eine Mittelwertsdifferenz sehr hilfreich (analog zu den in Kap. 1.4 besprochenen z-Werten). Die Standardisierung der Stichprobenkennwerteverteilung erfolgt ähnlich wie bei den z-Werten an ihrer geschätzten Streuung. Die standardisierten Stichprobenkennwerte heißen t-Werte, die standardisierten Verteilungen sind die t-Verteilungen (im Englischen auch „Student t“ genannt). Sie entsprechen nicht ganz der Standardnormalverteilung, sondern sind schmalgipfliger. Das liegt daran, dass die Form einer t-Verteilung von den Stichprobengrößen bzw. den Freiheitsgraden der Verteilung abhängig ist (siehe die folgenden Abschnitte 3.1.4 und 3.1.5). In einer t-Verteilung sind die Wahrscheinlichkeiten für die einzelnen t-Werte genau ablesbar. Die allgemeine Definition des t-Werts lautet:

Allgemeine Definition des t-Werts

$$t_{df} = \frac{\text{empirische Mittelwertsdifferenz - theoretische Mittelwertsdifferenz}}{\text{geschätzter Standardfehler der Mittelwertsdifferenz}}$$

formal: $$t_{df} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\hat{\sigma}_{\bar{x}_1-\bar{x}_2}}$$

Der t-Test findet in den meisten Fällen als Nullhypothesensignifikanztest Anwendung. Diesem t-Test liegt die Annahme zu Grunde, dass die Populationsmittelwerte der beiden zu vergleichenden Gruppen identisch sind. Die theoretische Mittelwertsdifferenz unter der Nullhypothese ist $\mu_1 - \mu_2 = 0$ und kann bei der Berechnung weggelassen werden. Die vereinfachte Formel lautet:

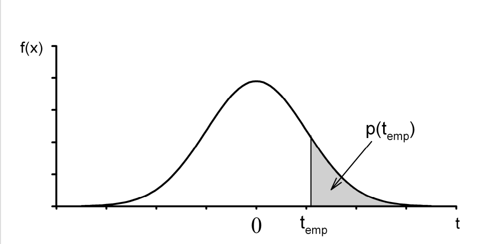
Definition des t-Werts unter der Nullhypothese

$$t_{df} = \frac{\bar{x}_1 - \bar{x}_2}{\hat{\sigma}_{\bar{x}_1 - \bar{x}_2}}$$

Der t-Test kann auch zur Testung anderer Hypothesen als der Nullhypothese dienen, in denen von einem in den Populationen vorhandenen Unterschied in den Mittelwerten ausgegangen wird. In einem solchen Fall wäre die theoretische Mittelwertsdifferenz größer Null. Allerdings wird der t-Test nur sehr selten in dieser Form verwendet. Deshalb findet sie hier keine weitere Beachtung.

Die obige Formel ermöglicht unter Kenntnis der entsprechenden Streuung die Umrechnung einer empirischen Mittelwertsdifferenz in einen t-Wert. Anhand der t-Verteilung kann einem empirischen t-Wert eine Wahrscheinlichkeit zugeordnet werden, mit der exakt dieser oder ein größerer t-Wert unter der Annahme der Nullhypothese auftritt. Die Auftretenswahrscheinlichkeit eines positiven t-Werts entspricht dem Anteil der Fläche unter der Kurve, den der t-Wert nach rechts abschneidet (Abb. 3.7). Die Wahrscheinlichkeiten der verschiedenen t-Werte sind in Tabelle B im Anhang aufgelistet. Dort findet sich auch eine ausführliche Beschreibung für die Benutzung aller Tabellen sowie für die Bestimmung von Wahrscheinlichkeiten negativer t-Werte.

Abb. 3.7. Wahrscheinlichkeit eines t-Werts in einer t-Verteilung



3.1.4 Die Freiheitsgrade einer t-Verteilung

Die exakte Form der t-Verteilung ist trotz der Standardisierung weiterhin vom Stichprobenumfang abhängig und deckt sich aus diesem Grunde nicht exakt mit der z-Verteilung. Der Unterschied zwischen diesen Verteilungen ist dadurch zu erklären, dass in die Berechnung des t-Werts nicht einer, sondern zwei erwartungstreue Schätzer für Populationsparameter eingehen: die empirische Mittelwertsdifferenz und der geschätzte Standardfehler der Mittelwertsdifferenz (In der Formel zur Berechnung der z-Werte ist die Streuung kein Schätzer der Populationsstreuung, sondern bezieht sich direkt auf die Population, siehe Kap. 1.4.). Leider liefert aber

Die Form der t-Verteilung ist von ihren Freiheitsgraden abhängig.

Abb. 3.8. Formen von t-Verteilungen in Abhängigkeit von ihren Freiheitsgraden

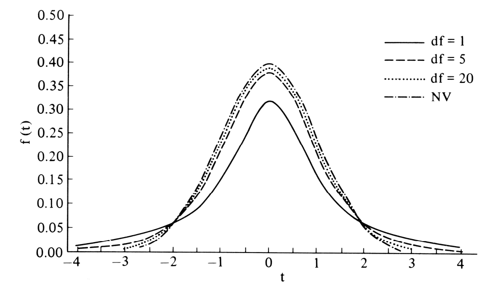
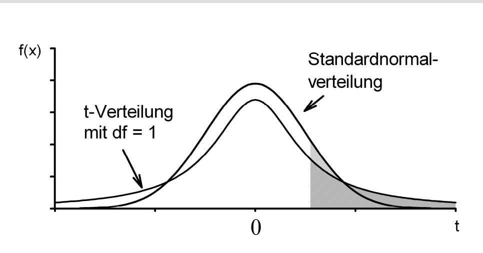


Abb. 3.9. Wahrscheinlichkeit eines t-Werts in Abhängigkeit der Freiheitsgrade



Bei einer geringen Anzahl von Freiheitsgraden sind große t-Werte unter der Nullhypothese wahrscheinlicher.

eine Formel mit zwei erwartungstreuen Schätzern kein erwartungstreues Ergebnis mehr. Mathematisch geht die t-Verteilung erst bei $n = \infty$ in eine z-Verteilung über, bei $n < \infty$ ist die t-Verteilung schmalgipfliger und flacher als eine z-Verteilung und nähert sich im Vergleich zur Standardnormalverteilung langsamer asymptotisch der x-Achse an. Die Freiheitsgrade der gefundenen Mittelwertsdifferenz erlauben eine genaue Beschreibung der zu verwendenden t-Verteilung. Sie werden in dem hier besprochenen t-Test durch folgende Formel berechnet:

$$df = n_1 + n_2 - 2$$

In Abbildung 3.8 sind t-Verteilungen mit verschiedenen Freiheitsgraden eingetragen (Abbildung mit freundlicher Genehmigung entnommen aus Bortz, 2005, S. 81). Bei df = 20 schmiegt sich diese Verteilung schon sehr nahe an die Standardnormalverteilung an. Bei df = 120 sind die beiden Verteilungen so gut wie identisch. Es ist deutlich zu sehen, dass die t-Verteilung umso schmalgipfliger und flacher verläuft und sich umso zögerlicher an die x-Achse annähert, je kleiner die Zahl der Freiheitsgrade ist.

Die Form der t-Verteilung nimmt Einfluss auf die Wahrscheinlichkeit, die einem bestimmten t-Wert zugeordnet wird. In einer t-Verteilung mit wenigen Freiheitsgraden schneidet ein positiver t-Wert einen größeren Teil der Fläche unter der Kurve nach rechts ab als bei einer Verteilung mit vielen Freiheitsgraden (siehe Abb. 3.9). Je flacher die t-Verteilung, desto größer wird also die Auftretenswahrscheinlichkeit eines bestimmten t-Werts. Auf der praktischen Ebene ist deshalb unter Annahme der Nullhypothese eine bestimmte empirische Mittelwertsdifferenz bei großen Stichproben unwahrscheinlicher als bei kleinen Stichproben. Für die Bestimmung der Wahrscheinlichkeit eines gefundenen t-Werts müssen also erst die Freiheitsgrade berechnet werden, um damit die richtige Fläche unter der Kurve zu erhalten.