Schließende Statistik

Die schließende Statistik befaßt sich mit dem Rückschluß von einer Stichprobe auf die Grundgesamtheit (Population). Es muß eine repräsentative (d.h. nur zufallsbeeinflusste) Stichprobe aus der Grundgesamtheit gezogen werden. Grundlage der schließenden Statistik ist die Wahrscheinlichkeitsrechnung.

Typische Fragestellungen sind:

Mit Hilfe der schließenden Statistik lassen sich also Fragen beantworten und Behauptungen überprüfen:

- Welche Zahnpasta ist für die Kariesprophylaxe zu empfehlen?
- Kann Mukoviszidose mit einem Schnelltest frühzeitig diagnostiziert werden?
- Welche Therapie wirkt bei Kindern mit Asthma am besten?
- Welche Faktoren beeinflussen die Heilungschancen von Karzinompatienten?
- Treten Mißbildungen bei Neugeborenen nach Tschernobyl häufiger auf?
- Die neue Therapie wirkt bei 85% aller Patienten.
- Warzen verschwinden auch ohne Behandlung bei 30% aller Personen.
- In Japan tritt Dickdarmkrebs seltener auf als Magenkrebs; in den USA ist es genau umgekehrt.

Typische Aufgabenstellungen sind:

(a) das Schätzen von Parametern, Angabe von Konfidenzintervallen

(b) das Testen von Hypothesen

Konfidenzintervalle dienen dem Zweck, die Genauigkeit von Zählungen und Messungen zu bestimmen. Testverfahren werden angewandt, um vermutete Sachverhalte (Hypothesen) anhand von Versuchen gegenüber täuschenden Zufallseffekten abzusichern.

Wir benötigen:

- Grundbegriffe der Wahrscheinlichkeitstheorie
- Studiendesign

Grundbegriffe der Wahrscheinlichkeitsrechnung

1. Zufällige Ereignisse und Wahrscheinlichkeit

Kann das Ergebnis eines Versuches nicht sicher vorausgesagt werden, so spricht man von einem **zufälligen Ereignis** (Bsp.: Ergebnis eines Münzwurfes, Geschlecht eines ungeborenen

Kindes, Wirkung eines Medikaments, Verlauf einer Erkrankung). Ursachen hierfür sind etwa die natürliche Variabilität der Versuchsobjekte, zufallsbedingte Ungenauigkeiten bei der Versuchsausführung und zufällige Meßfehler.

Das Eintreffen solcher Zufallsereignisse soll quantifiziert werden, d.h. es ist eine Zahl – seine **Wahrscheinlichkeit** – anzugeben, mit der das Ereignis eintritt.

Tritt ein Ereignis E stets ein, so nennen wir es das sichere Ereignis und ordnen ihm die Wahrscheinlichkeit P(E) = 1 zu. Ein Ereignis, welches niemals eintritt, ist das unmögliche Ereignis und wir ordnen ihm die Wahrscheinlichkeit P(E) = 0 zu. Damit ist zunächst der Bereich der Wahrscheinlichkeit des Eintreffens eines zufälligen Ereignisses E mit $0 \leq P(E) \leq 1$ festgelegt.

Das Bestimmen der Auftrittswahrscheinlichkeit eines beliebigen Ereignisses:

- theoretische Überlegungen: alle Elementarereignisse (nicht weiter aufteilbare Ereignisse: z.B. Würfeln einer 1) sind gleichwahrscheinlich – Würfel, Kartenspiel

Eine mögliche Definition für den Begriff der Wahrscheinlichkeit ist daher

$$P = \frac{Anzahl der günstigen Fälle}{Anzahl der möglichen Fälle}$$

Beispiel:

Lotto: Wie groß ist die Wahrscheinlichkeit einen 6er zu setzen?

Lösung:
Beim Lotto gibt es 45 Zahlen. Die Wahrscheinlichkeit einer richtigen Zahl ist 6/45, einer zweiten richtigen Zahl 5/44 usw. Es gilt also

$$\text{P(6 richtige)} = \frac{6}{45} \cdot \frac{5}{44} \cdot \frac{4}{43} \cdot \frac{3}{42} \cdot \frac{2}{41} \cdot \frac{1}{40} = \frac{1}{8145060}$$

oder mit Hilfe des Binomialkoeffizienten $\binom{45}{6} = \frac{45!}{6! \cdot (45-6)!}$

Es gibt $\binom{6}{4} \cdot \binom{39}{2} = 11115$ Möglichkeiten, vier richtige (und damit zwei falsche) Zahlen zu setzen. Damit ergeben sich folgende Wahrscheinlichkeiten:

P(6 richtige) =	1/8145060 =	0,000000123
P(5 richtige) =	234/8145060 =	0,000028729
P(4 richtige) =	11115/8145060 =	0,001364631
P(3 richtige) =	182780/8145060 =	0,022440596
P(2 richtige) =	1233765/8145060 =	0,151474022
P(1 richtige) =	3454542/8145060 =	0,424127262
P(0 richtige) =	3262623/8145060 =	0,400564637

- Empirie – relative Häufigkeiten
  mit wachsender Anzahl von Versuchen d.h. einer langen Folge von unabhängigen Durchführungen des zugrundeliegenden Experiments nähert sich die relative Häufigkeit einem bestimmten Zahlenwert – der Wahrscheinlichkeit.

Statistiker	*Münzwürfe (n)*	*Wappen (k)*	*k/n*
Buffon	4000	2048	0,5080
Pearson	12000	6019	0,5016
Pearson	24000	12012	0,5005

2. Zufallsvariablen und Wahrscheinlichkeitsverteilungen

Eine **Zufallsvariable** X ist eine Größe, die jedem zufälligen Ereignis eine Zahl zuordnet. Die Zahl (der Wert) x, den die Zufallsvariable X als Ergebnis des Zufallsexperimentes annimmt, nennt man Realisation x von X. Eine Zufallsvariable nimmt ihre möglichen Werte in Abhängigkeit vom Zufall an, d.h. gemäß ganz bestimmten Wahrscheinlichkeiten.

Zufallsexperiment	Zufälliges Ereignis	Zufallsvariable X	Werte der ZV X
Münzwurf	Wappen oder Zahl	Wurfergebnis	0 oder 1
Med. Therapie	Patient geheilt, unverändert, schlechter	Therapieerfolg	1,2 oder 3
Herstellung eines Produkts	Ausschuß, kein Ausschuß	Produktqualität	0 oder 1
Messung des Blutdrucks	erfaßter Wert	Blutdruckwert	$80 \leq X \leq 300$

Man nennt die Menge dieser Wahrscheinlichkeiten die **Wahrscheinlichkeitsverteilung** der zufälligen Variablen.

Eine **diskrete Zufallsvariable** X ist durch die Angabe ihres Wertebereiches

$$x_1, x_2, ..., x_n$$

und den Einzelwahrscheinlichkeiten für das Auftreten jeden Wertes x_i

$$P(X=x_1) = p_1, P(X=x_2) = p_2, ..., P(X=x_n) = p_n$$

vollständig definiert (**Wahrscheinlichkeitsfunktion**)

z.B. Würfel: $x_i = i$ und $p_i = 1/6$ für $i = 1,2,...,6$ (diskrete Gleichverteilung)

Mit der Wahrscheinlichkeit

$$F(x) = P(X \leq x_i) = p_1+p_2+...+p_i$$

wird die **Verteilungsfunktion** der diskreten Zufallsvariablen definiert. Sie ergibt sich als Summe der Einzelwahrscheinlichkeiten bis zur Stelle x, d.h. sie gibt nicht über die Wahrscheinlichkeit für die Annahme eines bestimmten Wertes Auskunft, sondern man kann durch sie erfahren, mit welcher Wahrscheinlichkeit Werte im ganzen Intervall unterhalb der Stelle x angenommen werden.

z. B. Würfel: $P(X \leq 3) = 1/6+1/6+1/6 = 1/2$

Stetige Zufallsgrößen werden durch die Dichtefunktion bzw. Verteilungsfunktion charakterisiert.

Die Wahrscheinlichkeitsfunktion ist das theoretische Gegenstück zur empirischen Häufigkeitsverteilung, die Verteilungsfunktion zur kumulativen Häufigkeit.

Wie bei Häufigkeitsverteilungen kann die in einer Wahrscheinlichkeitsverteilung enthaltene Information durch Kenngrößen (**Parameter**) beschrieben werden. Die Parameter der Grundgesamtheit werden meist mit griechischen Buchstaben bezeichnet: z.B. Populationsmittelwert (Erwartungswert) μ und Varianz σ^2.

Die wichtigsten Wahrscheinlichkeitsverteilungen unter den diskreten sind die **Binomialverteilung** und die **Poissonverteilung**, unter den stetigen Verteilungen ist es die **Normalverteilung**.
Weitere stetige Verteilungen sind die t-Verteilung, die χ^2-Verteilung und die F-Verteilung. Sie sind Prüfverteilungen von statistischen Tests.

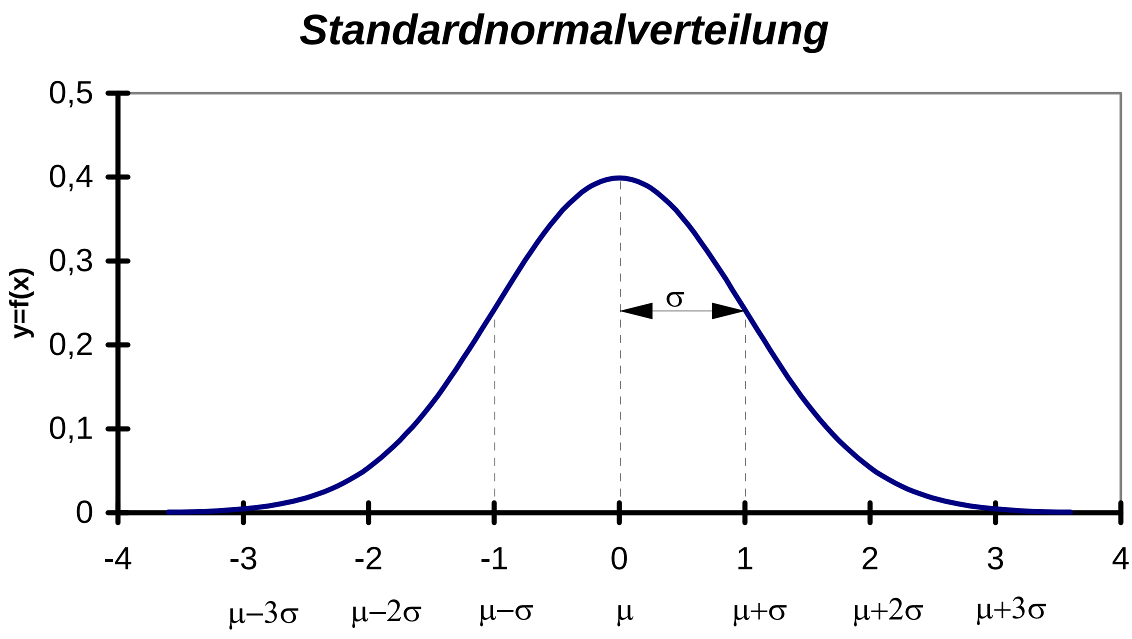
NORMALVERTEILUNG

Diese Verteilung hat in der Statistik eine zentrale Bedeutung: Eine Summe von vielen unabhängigen, beliebigen Zufallsvariablen ist angenähert normalverteilt; das bedeutet in der Praxis, daß viele Probleme unter Verwendung der Normalverteilungsannahme gelöst werden können - vorausgesetzt, die Stichprobe ist groß genug. Die Wahrscheinlichkeitsdichte der Normalverteilung (auch Gauß-Verteilung oder Glockenkurve) beträgt

$$y = f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \cdot \frac{(x-\mu)^2}{\sigma}}$$

$$\left(-\infty < x < \infty, -\infty < \mu < \infty, \sigma > 0\right)$$

Die *Standardnormalverteilung* hat einen Mittelwert von μ=0 und eine Standardabweichung von σ=1.



Wichtig sind folgende Bereiche:

$\mu \pm 1\sigma$	$\Rightarrow$	68,2% aller Werte liegen zwischen $\mu \pm \sigma$
$\mu \pm 2\sigma$	$\Rightarrow$	95,4% aller Werte liegen zwischen $\mu \pm 2\sigma$
$\mu \pm 3\sigma$	$\Rightarrow$	99,7% aller Werte liegen zwischen $\mu \pm 3\sigma$
$\mu \pm 1{,}96\sigma$	$\Rightarrow$	95% aller Werte liegen zwischen $\mu \pm 1{,}96\sigma$
$\mu \pm 2{,}58\sigma$	$\Rightarrow$	99% aller Werte liegen zwischen $\mu \pm 2{,}58\sigma$

Schätzen von Parametern, Konfidenzintervalle

Da man nicht die gesamte Population erfaßt, sondern so gut wie immer auf Stichproben von begrenztem Umfang angewiesen ist, muß man sogenannte Schätzungen für die Populationsparameter angeben. Die Kennzahlen, die wir in der deskriptiven Statistik kennengelernt haben, stellen solche Schätzungen für die Populationsparameter dar.
Im Falle der Normalverteilung (oder zumindest eingipfeligen, symmetrischen Verteilung) sind das arithmetische Mittel $\bar{x}$ und die Stichprobenvarianz s^2 „gute“ Schätzer für Erwartungswert μ und Varianz σ^2 der Population.

Stichprobenverteilung

Hat man viele Zufalls-Stichproben des Umfanges n aus der gleichen Population gezogen, so werden sich die Mittelwerte $\bar{x}_1, \bar{x}_2, ..., \bar{x}_k$ im allgemeinen unterscheiden.
Der Stichprobenmittelwert ist eine Zufallsvariable.

Folgende Aussagen lassen sich treffen:

- Der **Erwartungswert** des Stichprobenmittelwerts ist der Erwartungswert μ der Population
- **Standardfehler des Mittelwertes (SE of mean)**: Der Standardfehler des Mittelwertes ist die Standardabweichung der Mittelwerte-Verteilung $s/\sqrt{n}$.
- **Zentraler Grenzwertsatz**: Unabhängig von der Verteilung in der Population folgt die Verteilung des Stichprobenmittelwertes asymptotisch, d.h. approximativ für „großes“ n einer Normalverteilung mit Mittelwert μ und Standardabweichung $\sigma/\sqrt{n}$.

Konfidenzintervalle

Die Punktschätzung liefert einen einzelnen Wert für den unbekannten Parameter. Mehr Information bietet ein Schätzintervall (Konfidenzintervall), in dem der unbekannte (wahre) Parameter mit entsprechend hoher Wahrscheinlichkeit (z.B. 95%) enthalten ist. Ein solches Schätzintervall ist deshalb von besonderer Bedeutung, weil seine Breite die Genauigkeit oder Ungenauigkeit der Schätzung repräsentiert. Die Grenzen werden aus der Stichprobe bestimmt.

Schätzt man beispielsweise den Erwartungswert μ bei vorausgesetzter Normalverteilung durch den Mittelwert $\bar{x}$ und die Standardabweichung σ durch s, so bestimmen sich die Grenzen eines (1-α) Konfidenzintervalls für μ durch

$$\bar{x} - t_{1-\alpha/2,n-1} s/\sqrt{n} \le \mu \le \bar{x} + t_{1-\alpha/2,n-1} s/\sqrt{n}$$

$t_{1-\alpha/2,n-1}$ ist das (1-α/2)-Quantil der t-Verteilung mit n-1 Freiheitsgraden. α wird als **Irrtumswahrscheinlichkeit** bezeichnet, da α die Wahrscheinlichkeit angibt, daß der wahre Parameter nicht in diesem Bereich liegt.

Testen von Hypothesen

Prinzip statistischer Tests

Statistische Testverfahren dienen zur Überprüfung von Hypothesen.

<u>Einführungsbeispiel:</u>

Der Parameter PTT (Partielle Thromboplastinzeit) streut bei einem bestimmten Reagenz bei Erwachsenen mit einer Standardabweichung von σ=4,0 um ein Mittel μ_0=36 sec. In einem klinisch-chemischen Labor erhält man in einer Stichprobe von 25 Bestimmungen der PTT bei Kindern mit demselben Reagenz das arithmetische Mittel von $\bar{x}$ = 42 sec.

Frage: Weichen die PTT-Werte der Kinder „echt“ von den Werten der Erwachsenen ab oder sind diese Unterschiede nur zufallsbedingt?

Eine Beantwortung der Sachfrage kann durch eine Entscheidung zwischen folgenden beiden Aussagen (Hypothesen) vorgenommen werden:

Nullhypothese (H_0): die PTT-Werte bei Kindern unterscheiden sich nicht von den PTT-Werten bei Erwachsenen
$H_0 : \mu = \mu_0$ (= 36)

Alternativhypothese (H_1): die PTT-Werte unterscheiden sich

$H_1: \mu \neq \mu_0$

Da die Verteilung, der die Stichprobe entstammt nicht, vollständig bekannt ist, kann die Entscheidung zwischen H_0 und H_1 nicht mit Bestimmtheit, sondern nur unter Einschluß von Fehlerrisiken getroffen werden. Offensichtlich gibt es zwei Möglichkeiten einer Fehlentscheidung:

Ablehnung von H_0, obwohl H_0 zutrifft	Fehler 1.Art
Beibehaltung von H_0, obwohl H_1 gilt	Fehler 2.Art

Beide Fehlerarten können grundsätzlich nicht vermieden werden, man kann höchstens die Wahrscheinlichkeit für ihr Auftreten kontrollieren, indem man einen Maximalwert vorgibt, der nicht überschritten werden darf. Die Wahrscheinlichkeit für einen Fehler 1.Art bezeichnet man dabei als Irrtumswahrscheinlichkeit α oder **Signifikanzniveau** des Tests. Beide Fehler können nicht beliebig klein gehalten werden.
Die Wahrscheinlichkeit, einen Fehler 2.Art zu begehen, hängt davon ab, welche Alternative konkret vorliegt und wie groß die Stichprobe ist, aufgrund derer entschieden werden soll.
Bei der Festlegung der Wahrscheinlichkeiten für einen Fehler 1.Art bzw. 2.Art wird folgendermaßen vorgegangen:

- Die Irrtumswahrscheinlichkeit α wird vorgegeben (üblich α=0,05, α=0,01)

Bei geplanten Studien:
- Die Wahrscheinlichkeit <u>keinen</u> Fehler 2. Art zu begehen, die **Power** des Tests, mit der eine relevante Abweichung von der Nullhypothese aufgedeckt werden soll, wird festgelegt und entsprechend der dafür erforderliche Stichprobenumfang berechnet.

Entscheidungsregel („Test“) für das Beispiel:

Irrtumswahrscheinlichkeit α=0.05

<u>Modellannahmen</u>: wir gehen davon aus, daß die PTT-Werte der Kinder normalverteilt sind, mit unbekannten Erwartungswert μ und bekannter Standardabweichung σ (σ=4,0)

Einstichproben-Gaußtest:

$$T = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} \sim N(0,1) \qquad T = \frac{42 - 36}{4 / \sqrt{25}} = 7.5$$

T heißt Teststatistik, deren Verteilung unter der Nullhypothese bestimmt werden kann. Gegen die Nullhypothese – und für die Alternative - sprechen „sehr große" oder „sehr kleine" Werte von T, sodass man die Nullhypothese immer dann verwerfen wird, wenn solche extremen Werte auftreten (**kritischer Bereich, Ablehnungsbereich**).
Extreme Werte können auch unter der Nullhypothese vorkommen und würden zu ihrer Ablehnung führen; das wäre gerade ein Fehler 1.Art. Die Wahrscheinlichkeit dafür sollte insgesamt höchstens α=0.05 betragen und sich zu gleichen Anteilen auf zu große wie zu kleine Werte verteilen.
Unter diesen Bedingungen wird man die Nullhypothese also genau dann zurückweisen, wenn $T > z_{0.975} = 1.96$ oder $T < z_{0.025} = -1.96$ ist, wobei die z_q die entsprechenden Quantile der Standardnormalverteilung sind.

In SPSS wird sowohl die Teststatistik als auch der p-Wert ausgegeben. Der **p-Wert (Überschreitungswahrscheinlichkeit)** ist die Wahrscheinlichkeit, mit der bei Zutreffen der Nullhypothese der beobachtete Wert t der Teststatistik oder ein noch extremerer Wert realisiert würde: $p = P(T \geq t \mid H_0)$. **Man lehnt H_0 ab, falls $p \leq \alpha$!**

Das behandelte Testproblem heißt 2-seitig, da es in der Alternative Abweichungen von der Nullhypothese nach beiden Seiten zuläßt. Entsprechend gibt es unter demselben Modell zwei 1-seitige Testprobleme mit entsprechend zu modifizierenden Entscheidungsregeln.
(H_0: $\mu \leq \mu_0$ H_1: $\mu > \mu_0$ bzw. H_0: $\mu \geq \mu_0$ H_1: $\mu < \mu_0$)

<u>Testablauf zusammengefaßt:</u>

(1) Formulierung der Hypothesen
Nullhypothese - Alternativhypothese
einseitige, zweiseitige Fragestellung

(2) Wahl des Signifikanzniveaus α

(3) Bestimmung der Teststatistik

(4) Ausführung des Tests und Entscheidung

Häufig verwendeter Test auf Lageunterschiede

Der t-Test für 2 unabhängige Stichproben:

Hypothesen: H_0: $\mu_1 = \mu_2$ H_1: $\mu_1 \neq \mu_2$

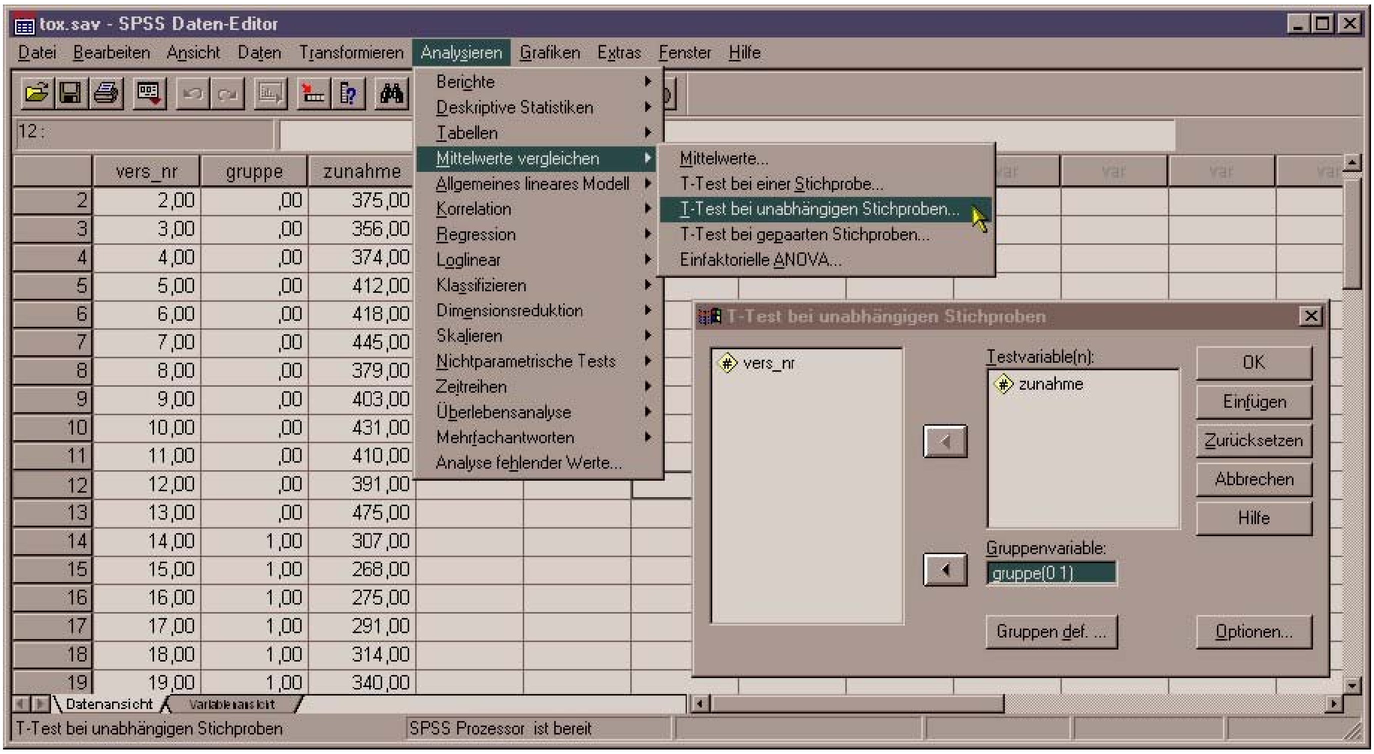
Voraussetzungen:
Die Beobachtungen der beiden Gruppen stammen aus unabhängigen normalverteilten Beobachtungen mit Mittelwerten μ_1 und μ_2 und die Standardabweichungen sind gleich $\sigma_1 = \sigma_2$, aber unbekannt.

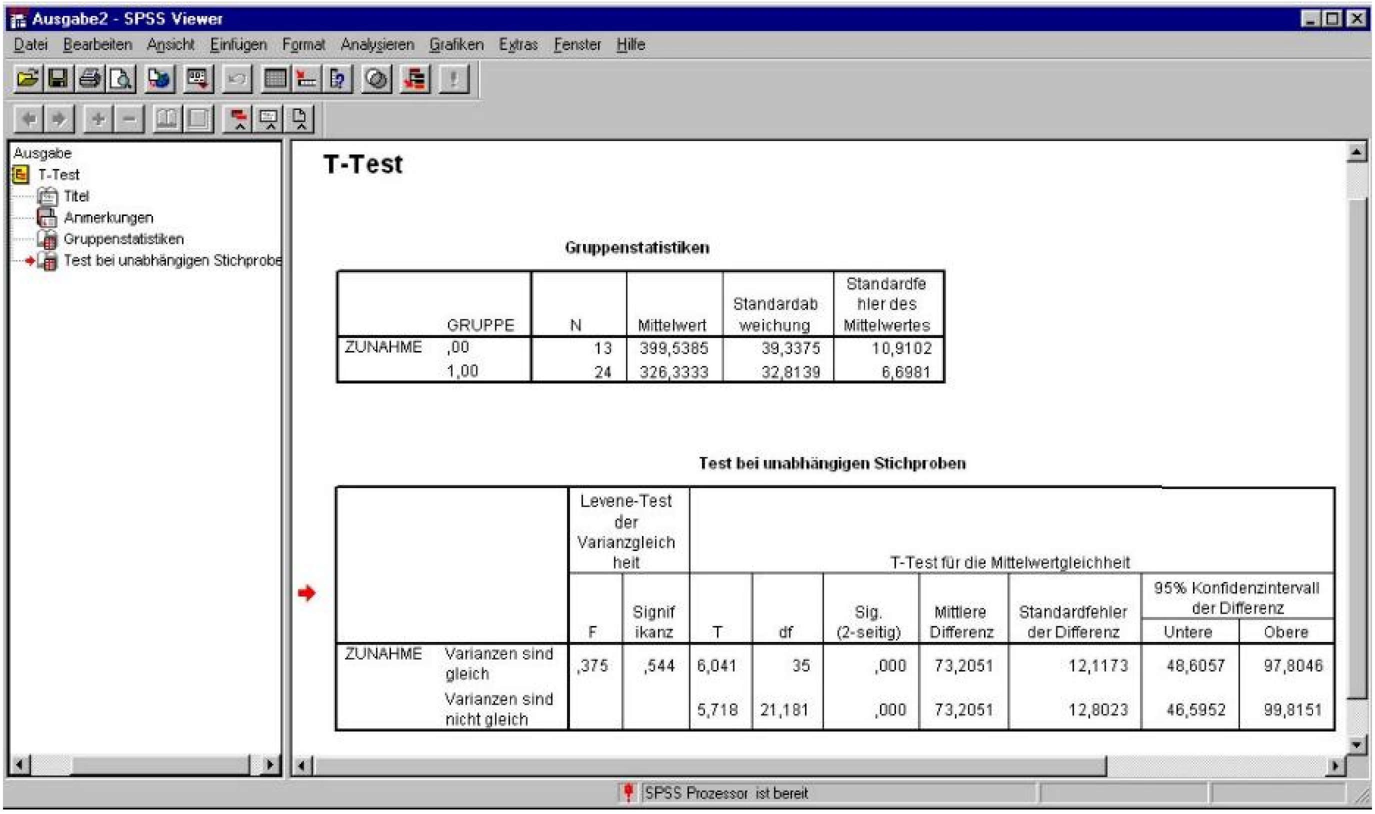
Teststatistik: $T = \dfrac{\bar{x}_1 - \bar{x}_2}{s\sqrt{1/n_1 + 1/n_2}}$ ist t-verteilt mit n_1+n_2-2 Freiheitsgraden, mit

$$s = \sqrt{\left[(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2\right]/(n_1 - n_2 - 2)}$$

SPSS Ausgabe (Menü Statistik, Mittelwerte vergleichen, Unabhängige-Stichproben T-Test)

- Beschreibende Statistik der beiden Gruppen durch Anzahl, Mittelwert, Standardabweichung, Standardfehler und Differenz der Mittelwerte
- Test auf Gleichheit der Varianzen nach Levene H_0: $s_1 = s_2$
- Ergebnis des t-Tests: Teststatistik, Freiheitsgrade, p-Wert, Konfidenzintervall





T-Test

Gruppenstatistiken

	GRUPPE	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes
ZUNAHME	,00	13	399,5385	39,3375	10,9102
	1,00	24	326,3333	32,8139	6,6981

Test bei unabhängigen Stichproben

		Levene-Test der Varianzgleichheit		T-Test für die Mittelwertgleichheit						
									95% Konfidenzintervall der Differenz	
		F	Signifikanz	T	df	Sig. (2-seitig)	Mittlere Differenz	Standardfehler der Differenz	Untere	Obere
ZUNAHME	Varianzen sind gleich	,375	,544	6,041	35	,000	73,2051	12,1173	48,6057	97,8046
	Varianzen sind nicht gleich			5,718	21,181	,000	73,2051	12,8023	46,5952	99,8151

Häufig verwendete Tests auf Lageunterschiede:

Anzahl und Art der Stichproben	quantitative Zielgröße		qualitative Zielgröße
	parametrische Testverfahren (Normalverteilung)	nichtparametrische Testverfahren	
eine Stichprobe	Einstichproben t-Test	Wilcoxon-Vorzeichen Rangsummentest	Binomialtest
2 verbundene Stichproben	t-Test für verbundene Stichproben	Wilcoxon-Vorzeichen-Rangsummentest	Mc Nemar Test
2 unabhängige Stichproben	t-Test für unabhängige Stichproben (Gleichheit der Varianzen), Welch-Test	Wilcoxon Rangsummentest (U-Test von Mann und Whitney)	Chi-Quadrat Test Fishers Exakter Test für 2x2 Tafel
> 2 verbundene Stichproben	Varianzanalyse (randomisierte Blockanlage)	Friedman Test	
>2 unabhängige Stichproben	Varianzanalyse	Kruskal-Wallis Test	