Hypothesenprüfung "für Anfänger" - ein schulisches Minimalprogramm mit Anwendungsbeispielen aus Medizin, Pädagogik und Soziologie

Roman Laußermayer, Imst

1. Vorbemerkung

Es gibt kaum Lehrstoffe, die nicht unter irgendeinem Gesichtspunkt wichtig und - bei entsprechendem zeitlichen und vor allem didaktischen Aufwand - auch für Schüler erklärbar sind. Die eigentliche Voraussetzung für die Behandlung neuer zusätzlicher Lehrstoffe in der Schule kann also nicht die Frage sein, ob die Behandlung möglich ist, sondern nur, ob folgende vier "Extremalprinzipien" erfüllbar sind:

1) möglichst geringer Zeitbedarf;
2) möglichst geringer begrifflicher Aufwand (Vor- und Zusatzkenntnisse),
3) mit möglichst geringem Wissensaufwand möglichst viel erklären, keine reinen "ad-hoc-Hilfsmittel": d.h. neue Lehrstoffe (wie Hypothesenprüfung) sollten nur dann in den Unterricht aufgenommen werden, wenn die dafür notwendigen Hilfsmittel (Begriffe, Formeln, Lehrsätze) solcher Art sind, daß ihre Behandlung auch dann noch nützlich bleibt (z.B. als Wiederholung bzw. Vertiefung des übrigen Stoffes), wenn das eigentliche Ziel z.B. aus Zeitgründen (Erkrankung, Stundenausfall) nicht mehr behandelt werden kann;
4) möglichst praxisechte Problem- und Aufgabenstellungen (Motivation für den Schüler!).

2. Gebrauchte Begriffe (Definitionen)

Grundgesamtheit (Population) = Die größte Menge von Dingen, an denen wir gerade interessiert sind (Bsp.: wir messen eine Variable an jedem Element einer Menge; Ergebnis: eine Grundgesamtheit von Werten dieser Variablen).

Stichprobe = Teil einer Grundgesamtheit.

Zufallsstichprobe = Eine Stichprobe der Größe n aus einer Grundgesamtheit der Größe N, die so gezogen wird, daß für jede mögliche Stichprobe der Größe n die Wahrscheinlichkeit, gezogen zu werden, gleich groß ist.

N.B.: Bei allen unseren Beispielen wird angenommen, daß es sich um solche Zufallsstichproben handelt. Allerdings ist man in der Praxis hinsichtlich der Erfülltheit dieser Definition sehr großzügig!

Bsp.: freiwillige Versuchspersonen, etwa Studenten, sind keine Zufallsstichproben!

Statistik = Beschreibende Größe, die aus den Daten einer Stichprobe errechnet wird (lateinische Buchstaben, z.B. $\bar{x}$, s).

Parameter = Beschreibende Größe, die aus den Daten einer Grundgesamtheit errechnet wird (griechische Buchstaben, z.B. μ, σ).

3. Statistische Inferenz

(beurteilende, inferentielle Statistik)

Grundgedanke: Auf Grund von Ergebnissen über eine Stichprobe aus einer Grundgesamtheit werden Schlüsse über eine Grundgesamtheit gezogen.

Wichtigste Arten:

a) Schätzung

b) Hypothensenprüfung

Beide hängen sehr eng zusammen. Dieser enge Zusammenhang liefert die didaktische Grundidee für die Behandlung der Hypothesenprüfung in der Schule.

3.1. Notwendige Vor- bzw. Zusatzkenntnisse:

Gebraucht werden: Begriffe und Eigenschaften der Normalverteilung (wird hier vorausgesetzt), Begriff und Eigenschaften der Stichprobenverteilung und der zentrale Grenzwertsatz.

Stichprobenverteilung (sampling distribution):

Anschauliche Beschreibung: für diskrete, endliche Grundgesamtheiten kann man eine Stichprobenverteilung wie folgt herstellen:

a) aus einer endlichen Grundgesamtheit vom Umfang N werden alle möglichen Zufallsstichproben vom Umfang n gezogen;

b) für jede dieser Zufallsstichproben wird die interessierende Statistik berechnet (in der Regel: $\bar{x}$,s);

c) es wird eine Häufigkeitstabelle erstellt:

berechneter Wert der Statistik	Häufigkeit dieses Wertes

d) wenn die Grundgesamtheit groß (oder unendlich) ist, wird die Häufigkeitsverteilung durch eine möglichst große Anzahl von Stichproben approximiert.

Verallgemeinerung (Definition): Stichprobenverteilung = Verteilung aller möglichen Werte einer Statistik (z.B. $\bar{x}$), die für gleich große Zufallsstichproben aus ein und derselben Grundgesamtheit errechnet wird.

Die Eigenschaften der Stichprobenverteilung und der Inhalt des für das Weitere so wichtigen zentralen Grenzwertsatzes werden nun dem Schüler an einem einfachen Rechenbeispiel anschaulich erklärt. Ein Beweis des zentralen Grenzwertsatzes für Schüler ist natürlich nicht möglich. Aber auch an medizinischen, naturwissenschaftlichen und wirtschaftswissenschaftlichen Fakultäten amerikanischer Universitäten ist der hier eingeschlagene Weg der "Überzeugung durch Beispiele" anstelle von Beweisen unangefochtene Praxis!

<u>Beispiel 1:</u>

Das Alter von 5 Kindern, die in der Kinderstation einer Klinik zu einem bestimmten Zweck untersucht werden (= Grundgesamtheit), beträgt 6, 8, 10, 12 und 14 Jahre. Bestimme alle möglichen Stichproben der Größe 2 sowie die Mittelwerte dieser Stichproben. Zeichne das Histogramm für die Verteilung der Grundgesamtheit und für die Verteilung der Stichprobe. (Stichprobe mit Zurücklegen).

LÖSUNG:

Die Parameter der Grundgesamtheit sind: $\mu = 10$, $\sigma = \sqrt{8}$

Die Bestimmung aller möglichen Stichproben der Größe 2 und ihrer Mittelwerte ergibt folgende Tabelle ($\bar{x}$ wird jeweils in Klammer angegeben:

		zweiter Zug				
		6	8	10	12	14
erster Zug	6	6,6(6)	6,8(7)	6,10(8)	6,12(9)	6,14(10)
	8	8,6(7)	8,8(8)	8,10(9)	8,12(10)	8,14(11)
	10	10,6(8)	10,8(9)	10,10(10)	10,12(11)	10,14(12)
	12	12,6(9)	12,8(10)	12,10(11)	12,12(12)	12,14(13)
	14	14,6(10)	14,8(11)	14,10(12)	14,12(13)	14,14(14)

Die Stichproben ohne Zurücklegen befinden sich oberhalb oder unterhalb der Hauptdiagonalen.

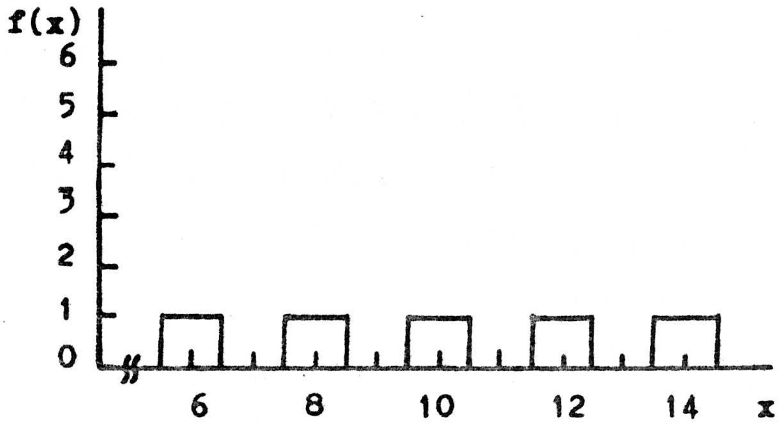
Als nächstes wird nun die Häufigkeitstabelle für $\bar{x}$ erstellt:

$\bar{x}$	Häufigkeit	relative Häufigkeit
6	1	1/25
7	2	2/25
8	3	3/25
usw.		
Summe	25	1

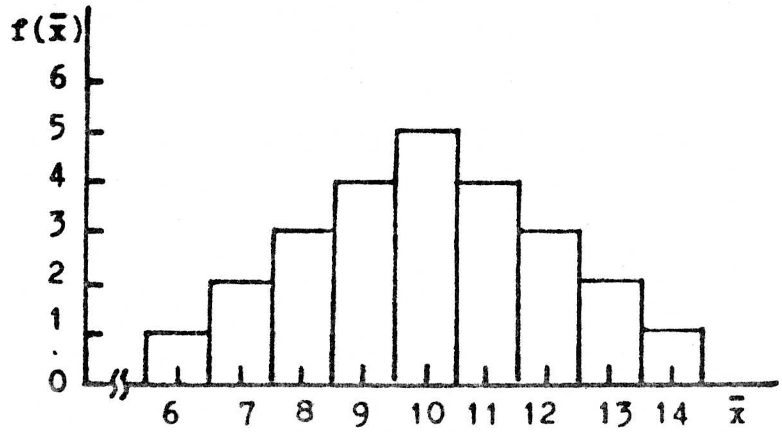
Zu beachten: Diese Stichprobenverteilung erfüllt die Bedingungen einer Wahrscheinlichkeitsverteilung: Die Einzelwahrscheinlichkeiten sind alle positiv, ihre Summe ergibt 1.

Wir interessieren uns natürlich vor allem für die funktionelle Form der Verteilung der $\bar{x}$-Werte:

Verteilung der Grundgesamtheit:



Verteilung der Stichprobe:



Besonders zu beachten ist dabei (im Hinblick auf später: zentraler Grenzwertsatz!) der radikale Unterschied zwischen den beiden

Verteilungen: "rechteck-förmige" Verteilung der Grundgesamtheit - "normal-förmige" Verteilung der zugehörigen Stichprobe!

Die Rechnung für dieses Beispiel liefert folgendes wichtige Ergebnis:

$$\text{I)} \quad \mu_{\bar{x}} = \mu \left[\frac{\Sigma \bar{x}_i}{N^n} = 10 = \frac{\Sigma x_i}{N} \right]$$

$$\text{II)} \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \quad \left[\frac{\Sigma(\bar{x}_i - \mu_{\bar{x}})^2}{N^n} = 4 = \frac{1}{n} \cdot \frac{\Sigma(x_i - \mu)^2}{N} \right]$$

Analog für die Stichproben ohne Zurücklegen:

$$\text{I)} \quad \mu_{\bar{x}} = \mu \left[\frac{\Sigma \bar{x}_i}{\binom{N}{n}} = 10 = \frac{\Sigma x_i}{N} \right]$$

$$\text{II)} \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}} \quad \left[\frac{\Sigma(\bar{x}_i - \mu_{\bar{x}})^2}{\binom{N}{n}} = 3 = \frac{\sigma^2}{n} \cdot \frac{N-n}{N-1} \right]$$

NB: Der Faktor $\sqrt{\frac{N-n}{N-1}}$ heißt "Endlichkeitskorrektur" (finite population correction) und wird in der Praxis immer dann vernachlässigt, wenn $n/N \leq 0{,}05$ ist.

Verallgemeinerung:

Die erhaltenen Ergebnisse sind kein Zufall, sondern typisch für Stichprobenverteilungen:

Es sei N der Umfang einer Grundgesamtheit mit dem Mittelwert μ und der Standardabweichung σ. n sei der Stichprobenumfang. Dann gibt es $\binom{N}{n}$ mögliche Stichproben ohne Zurücklegen bzw. N^n Stichproben mit Zurücklegen und ebensoviele Stichprobenmittelwerte $\bar{x}$. Diese Stichprobenmittelwerte ergeben eine Verteilung mit den folgenden Eigenschaften:

I) Ist die Grundgesamtheit normalverteilt, dann sind auch die Stichprobenmittelwerte $\bar{x}$ normalverteilt.

II) Ist die Grundgesamtheit groß, aber nicht normalverteilt, so nähert sich die Verteilung der Stichprobenmittelwerte $\bar{x}$ einer Normalverteilung, wenn der Stichprobenumfang groß (≥ 30) ist.

Satz II ist der wichtige "zentrale Grenzwertsatz" !

In beiden Fällen (I und II) gilt:

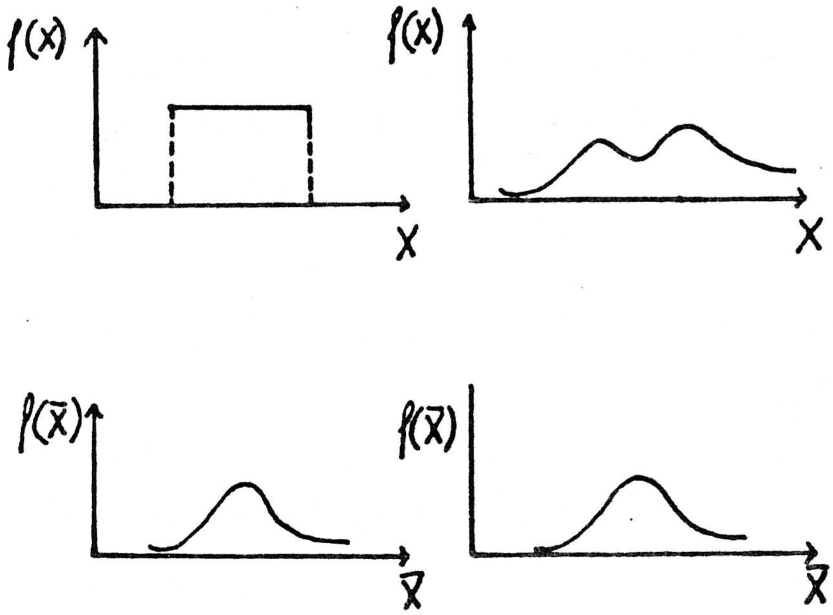
a) Der Mittelwert der Verteilung der Stichprobenmittelwerte $\mu_{\bar{x}}$ ist gleich dem Mittelwert μ der Grundgesamtheit.

b) Die Standardabweichung der Verteilung der Stichprobenmittelwerte beträgt:

für Stichproben mit Zurücklegen bzw. für $n/N \leq 0{,}05$: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$,

für Stichproben ohne Zurücklegen: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$.

Der für die angewandte Statistik so grundlegende zentrale Grenzwertsatz wird im Lehrbuch von Yamane (Statistik, Band 1, Fischer Taschenbuch, 1976, S.147) durch folgende Abbildung sehr eindrucksvoll veranschaulicht:



Didaktische Anmerkung:

Es muß betont werden, daß das bisher Behandelte für die Statistik so wichtig ist, daß es sich auch dann lohnt, wenn man aus Zeitgründen im Stoff nicht mehr weiter kommt (siehe oben, 3."Extremalprinzip"). Der zentrale Grenzwertsatz ist ja einer der Hauptgründe für die große Bedeutung der Normalverteilung in der Statistik.

Er ist daher so wichtig, daß man - etwas überspitzt - fragen könnte: Wenn man die Eigenschaften der Stichprobenverteilung und den zentralen Grenzwertsatz nicht (wenigstens in dieser einfachen Art) behandeln kann, sollte man dann nicht auch auf die Behandlung der Normalverteilung verzichten? Die "zentrale" Bedeutung des zentralen Grenzwertsatzes für die Statistik wird daher im Schulbuch von Laub (Lehrbuch der Mathematik für die Oberstufe der AHS, IV.Teil für die 8.Klasse (erstes Heft), Wien 1974, S.52 f) sehr zu Recht besonders betont.

Die beiden folgenden Beispiele über Stichprobenverteilungen zeigen, daß man nun dem Grundgedanken der Hypothesenprüfung schon sehr nahe ist:

Beispiel 2:

Es ist bekannt, daß in einer großen menschlichen Population die Schädellänge ungefähr normal verteilt ist mit $\mu = 185{,}6$ mm und $\sigma = 12{,}7$ mm. Wie groß ist die Wahrscheinlichkeit dafür, daß für eine Stichprobe von $n = 10$ aus dieser Population $\bar{x} > 190$ ist?

LÖSUNG:

Für die Stichprobe der $\bar{x}$ gilt: $\bar{x}$ ist normalverteilt mit $\mu_{\bar{x}} = \mu = 185{,}6$ und $\sigma_{\bar{x}} = 12{,}7/\sqrt{10}$.

Somit (Transformation auf die Standard-Normalverteilung):

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma/\sqrt{n}} = 1{,}09$$

Aus der Tabelle der z-Werte (Normalverteilung) folgt:

Fläche rechts von 1,09 = 0,1379. Somit $P(\bar{x} > 190) = 0{,}1379$.

Didaktische Anmerkungen:

- Im Unterricht wird man diese Fläche 0,1379 selbstverständlich auch graphisch darstellen.
- Als Vorbereitung der Schüler auf die Hypothesenprüfung lohnt sich die Zusatzüberlegung, was diese rund 14 % Wahrscheinlichkeit bedeuten (ist diese Wahrscheinlichkeit klein oder groß?). Man

wird darauf hinweisen, daß manzur Beantwortung dieser Fragd einen Vergleichsmaßstab braucht und daß dafür üblicherweise 5 % genommen werden.

Beispiel 3 :

Mittelwert und Standardabweichung der Eisenwerte des Blutserums von gesunden Männern sind 120 µg pro 100 ml und 15 µg pro 100 ml. Die Form der Verteilung der Grundgesamtheit ist nicht gegeben. Wie groß ist die Wahrscheinlichkeit dafür, daß eine Zufallsstichprobe von 50 normalen Männern einen Mittelwert zwischen 115 und 125 µg pro 100 ml ergibt?

LÖSUNG:

Wegen $n = 50$ (> 30) gilt der zentrale Grenzwertsatz. Standardisierung der annähernd normalen Stichprobenverteilung von $\bar{x}$ ($\mu_{\bar{x}} = 120$; $\sigma_{\bar{x}} = 15/\sqrt{50} = 2{,}12$) ergibt:

$$P(115 \leq \bar{x} \leq 125) = P\left(\frac{115-120}{2{,}12} \leq z \leq \frac{125-120}{2{,}12}\right) = P(-2{,}36 \leq z \leq 2{,}36) =$$

$$2.0{,}491 \approx 98\ \%$$

Didaktische Anmerkungen: analog Beispiel 2

3.2. Schätzung

Grundgedanke:

Man berechnet mit Hilfe einer Stichprobe eine Statistik, die als Näherungswert (Approximation) für den betreffenden Parameter der Population dient, aus der die Stichprobe genommen wurde.

Beispiele:

- Der Verwalter eines großen Krankenhauses braucht das Durchschnittsalter der in einem bestimmten Jahr aufgenommenen Patienten. Es wäre zu aufwendig (zu teuer), alle Akten durchzugehen.
- Eine Klinik möchte wissen, auf welchen Bruchteil einer bestimmten

Art von Patienten ein bestimmtes Medikament unerwünschte Nebenwirkungen hat. Man kann natürlich nicht warten, bis die Gesamtheit aller potentiellen Patienten beobachtet ist (Gefahr für die Patienten, Schaden für den Ruf der Klinik).

Grundbegriffe: Punktschätzung - Intervallschätzung

Problemstellung:

Ein Forscher möchte den Mittelwert einer normalverteilten Population schätzen. Er zieht eine Zufallsstichprobe der Größe n und nimmt den daraus berechneten Mittelwert $\bar{x}$ als Schätzwert für μ (= Punktschätzung!). Man kann zwar beweisen, daß $\bar{x}$ ein guter Schätzwert für μ ist, aber man kann natürlich nicht erwarten, daß $\bar{x} = \mu$ (Wirkungen des Zufalls bei der Auswahl der Stichprobe!). Daher: Schätzung von μ mit Hilfe eines Intervalls (Intervallschätzung!)

Didaktische Aufbereitung:

Es folgt nun die für die Hypothesenprüfung wichtige Grundidee, auf die alles ankommt und die daher möglichst einfach und anschaulich erklärt werden muß.

Annahmen: Der Mittelwert μ der Grundgesamtheit sei nicht bekannt, die Standardabweichung σ der Grundgesamtheit aber sei gegeben (z.B. aus früheren Untersuchungen übernommen).

N.B.: Wenn σ unbekannt ist, wird in der Praxis meistens die Standardabweichung der Stichprobe s an Stelle von σ verwendet. Im Unterricht wird man selbstverständlich ebenso vorgehen!

1.Schritt:

Der Forscher weiß

a) daß die Stichprobenmittelwerte $\bar{x}$ normalverteilt sind mit $\mu_{\bar{x}} = \mu$ und $\sigma_{\bar{x}} = \sigma/\sqrt{n}$ (zentraler Grenzwertsatz!).

b) daß 95 % aller möglichen Werte der Verteilung von $\bar{x}$ zwischen $\mu-1{,}96\ \sigma_{\bar{x}}$ und $\mu+1{,}96\ \sigma_{\bar{x}}$ liegen, m.a.W., daß das Intervall $\mu\pm1{,}96\ \sigma_{\bar{x}}$ 95 % aller möglichen Werte von $\bar{x}$ enthält (<u>Eigenschaft der Normalverteilung!</u>)

<u>2.Schritt:</u>

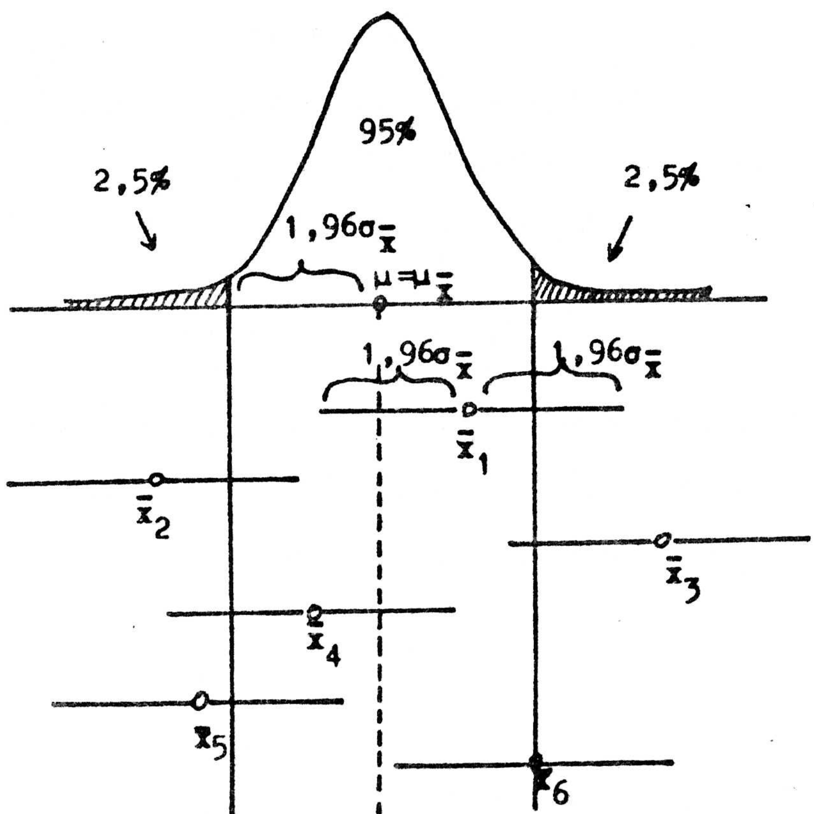
Der Forscher weiß aber nicht, wie groß $\mu=\mu_{\bar{x}}$ ist. Er kann also zwar die Form der Kurve der Verteilung von $\bar{x}$ zeichnen (<u>Eigenschaft der Normalverteilung!</u>), weiß aber nicht, wo auf der $\bar{x}$-Achse er diese Kurve anbringen soll (weiter rechts oder weiter links).

<u>3.Schritt:</u>

Der Forscher überlegt sich das Folgende (Grundidee!):

a) Angenommen, der Wert $\mu = \mu_{\bar{x}}$ wäre bekannt, so könnte man mit $\mu = \mu_{\bar{x}}$ und $\sigma_{\bar{x}}$ die zugehörige Stichprobenverteilung von $\bar{x}$ zeichnen und zu jedem möglichen $\bar{x}$ ein Intervall der Größe $(\bar{x} \pm 1{,}96\ \sigma_{\bar{x}})$ konstruieren (siehe folgende Abbildung).

b) Von 95 % dieser Intervalle würden die Mittelpunkte im Intervall $(\mu\pm1{,}96\ \sigma_{\bar{x}})$ liegen und jedes dieser Intervalle würde somit μ enthalten (siehe Abbildung).



Verallgemeinerung (Merksätze):

(I) Bei sehr oft wiederholtem ("in the long run") Ziehen einer Stichprobe aus einer normalverteilten Grundgesamtheit bzw. bei einer entsprechend großen Stichprobe ($n \geq 30$) aus einer beliebig verteilten Grundgesamtheit würden 95 % aller Intervalle der Form ($\bar{x} \pm 1{,}96 . \sigma_{\bar{x}}$) den Mittelwert μ der Grundgesamtheit enthalten (= statistische Wahrscheinlichkeitsinterpretation für $\bar{x} \pm 1{,}96 . \sigma_{\bar{x}}$!)

(II) In der Praxis der Hypothesenprüfung werden natürlich nicht beliebig viele $\bar{x}$, sondern nur ein Wert $\bar{x}$ ermittelt (1 Stichprobe). Für die Praxis der Hypothesenprüfung ist daher folgende "praktische Interpretation" des Konfidenzintervalls grundlegend: "Wir sind zu 95 % überzeugt ("confident"), daß das eine berechnete Intervall ($\bar{x} \pm 1{,}96 . \sigma_{\bar{x}}$) den Mittelwert der Grundgesamtheit enthält (= subjektive Wahrscheinlichkeitsinterpretation für($\bar{x} \pm 1{,}96 . \sigma_{\bar{x}}$!)

(III) Das Intervall ($\bar{x} \pm 1{,}96 . \sigma_{\bar{x}}$) heißt 95 % - Konfidenzintervall für μ (analog:($\bar{x} \pm 1{,}65 . \sigma_{\bar{x}}$)= 90 % - Konfidenzintervall und ($\bar{x} \pm 2{,}58 . \sigma_{\bar{x}}$)= 99 % - Konfidenzintervall)

Anmerkung:

Zwischen (I) und (II) liegt ein wichtiger "begrifflicher Sprung": der Wechsel vom statistischen zum subjektiven Wahrscheinlichkeitsbegriff. Diesem Umstand wird leider auch in der Fachliteratur kaum Beachtung geschenkt.

Beispiel 4:

In einem Experiment soll für eine bestimmte Personen-Grundgesamtheit die mittlere Anzahl der Herzschläge pro Minute ermittelt werden. Eine Stichprobe von 49 Personen ergab den Wert 90 Schläge pro Minute.

Auf Grund früherer Untersuchungen sei bekannt, daß σ der Grundgesamtheit 10 Schläge pro Minute beträgt. Bestimme das 95 % - Konfidenzintervall für μ.

LÖSUNG: 95 % - Konfidenzintervall $= \left(90 \pm 1{,}96 \cdot \frac{10}{\sqrt{49}}\right) = (87;93)$

Interpretation: (etwas anders)

5 % der Intervalle $(\bar{x} \pm 1{,}96 \cdot \sigma_{\bar{x}})$ enthalten μ nicht. Das Risiko, ein solches Intervall zu erwischen, beträgt also 5 %. Man nimmt dieses Risiko in Kauf, d.h. wenn das Konfidenzintervall $(\bar{x} \pm 1{,}96 \cdot \sigma_{\bar{x}})$ den Wert μ nicht enthält, ist man mit 95 % Wahrscheinlichkeit (subjektive Wahrscheinlichkeit!) überzeugt, daß der Meßwert $\bar{x}$ durch andere Ursachen zu erklären ist, als durch Zufall.

Beispiel 5:

Bei einer immunologischen Untersuchung wurde für eine Stichprobe von 49 Erwachsenen u.a. der Durchmesser eines Erythems (Hautrötung) als Reaktion auf ein Antigen bestimmt. Das Erythem der Stichprobe hatte einen Mittelwert von 21 mm und eine Standardabweichung von 11 mm. Kann aus diesen Daten geschlossen werden, daß der Mittelwert der Grundgesamtheit kleiner als 30 mm ist?

LÖSUNG: 95 % - Konfidenzintervall für μ: $\left(21 \pm 1{,}96 \cdot \frac{11}{\sqrt{49}}\right) = (18;24)$

INTERPRETATION: Wir können also mit 95 % Wahrscheinlichkeit annehmen, daß der Mittelwert der Grundgesamtheit in diesem Intervall liegt und somit kleiner ist als 30 mm. Im Jargon der Hypothesenprüfung: "Die Hypothese $\mu \geq 30$ (H_0) ist auf dem Signifikanzniveau 5 % zu verwerfen".

Didaktische Anmerkung:

Bei sehr starkem Zeitdruck könnte man an dieser Stelle bereits abschließen. Die Grundidee der Hypothesenprüfung ist damit im wesentlichen behandelt. Falls es die Zeit noch erlaubt, empfiehlt sich als Abrundung (Vertiefung) die Behandlung des im folgenden angeführten Stoffes.

4. Hypothesenprüfung (Randstoff!)

Sinn und Zweck der Hypothesenprüfung:

Hypothesenprüfung soll dem Forscher (Mediziner, Biologen, Soziologen, Pädagogen, Psychologen) oder dem Verwaltungsexperten helfen, durch Untersuchung von Stichproben aus einer Population Entscheidungen über diese Population selbst zu treffen.

Grundbegriffe:

Hypothese = Aussage über eine oder mehrere Grundgesamtheiten.

Forschungshypothese = Eine durch Forschung entstandene Vermutung. Eine Forschungshypothese führt zu einer statistischen Hypothese.

Statistische Hypothese = Formulierung einer Forschungshypothese, die geeignet ist, mit statistischen Methoden geprüft zu werden.

Nullhypothese (H_0) = Eine zu testende Annahme (Hypothese, Theorie), die in der Regel aufgestellt wird, um verworfe oder nicht verworfen zu werden.

Alternativhypothese (H_1) = Das Gegenteil der Nullhypothese. In der Regel die Hypothese, die der Forscher bestätigt sehen möchte.

Beispiel 6:

Ein Lehrer gibt einer Klasse einen Test, von dem er aus jahrelanger Erfahrung weiß, daß er normalerweise einen Mittelwert von $\mu = 78$ Punkten und eine Standardabweichung von $\sigma = 7$ Punkten liefert. Eine gegenwärtige Klasse mit 30 Schülern erreicht einen Mittelwert von 82 Punkten. Berechtigt dieses Ergebnis zur Annahme (Hypothese H_1), daß es sich um eine überdurchschnittlich gute Klasse handelt, m.a.W., daß diese Klasse nicht der Grundgesamtheit mit

$\mu = 78$, sondern einer Grundgesamtheit mit $\mu > 78$ entstammt? (Signifikanzniveau 0,05).

LÖSUNG: 95 % - Konfidenzintervall für μ: $\left(82 \pm 1{,}96 \cdot \frac{7}{\sqrt{30}}\right) = (80;85)$

INTERPRETATION: Da man mit 95 % Sicherheit annehmen muß, daß das Konfidenzintervall den Wert μ enthält, der Wert 78 aber nicht in diesem Konfidenzintervall liegt, ist die Hypothese H_0 ("die Klasse entstammt einer Grundgesamtheit mit $\mu = 78$") mit 95 % Sicherheit zu verwerfen ("auf dem Signifikanzniveau 0,05" oder "auf dem 5 % - Niveau der Verläßlichkeit" zu verwerfen). M.a.W.: die Klasse muß mit 95 % Sicherheit aus einer Grundgesamtheit mit $\mu > 78$ stammen, also überdurchschnittlich sein.

Verallgemeinerung:

Wir verwerfen H_0 (akzeptieren H_1) auf dem Signifikanzniveau 0,05, wenn der von der Hypothese H_0 angenommene Parameter nicht im 95 % - Konfidenzintervall liegt. Wenn er aber in diesem Intervall liegt, wird die Hypothese H_0 "auf dem Signifikanzniveau 0,05" akzeptiert.

5. Abschließende Bemerkungen für den Lehrer:

5.1. Die dargestellte Form der Hypothesenprüfung setzt nur den Begriff des Konfidenzintervalls (also eine Eigenschaft der Normalverteilung und den zentralen Grenzwertsatz) voraus, ist aber mathematisch äquivalent der Hypothesenprüfung mittels Teststatistik (z-Wert als Entscheidungszahl).

Beispiel 6: (diesmal mit Test-Statistik zu lösen)

Kritischer z-Wert: $z_k = 1{,}96$

Test-Statistik: $z_T = \frac{82-78}{7/\sqrt{30}} = 3{,}13$

Da $z = 3{,}13 > z = 1{,}96$, ist H_0 zu verwerfen.

Die Äquivalenz der beiden Methoden der Hypothesenprüfung (Konfidenzintervall bzw. Test-Statistik) ergibt sich durch Umformung der Ungleichung

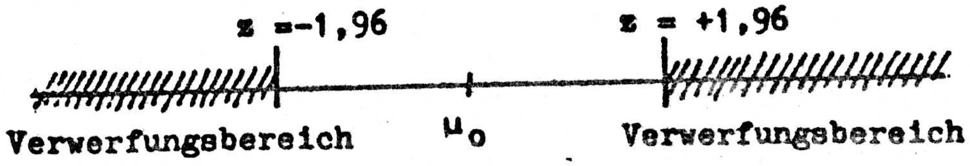
$$z_T = \frac{82-78}{7/\sqrt{30}} > z_k = 1{,}96 \Leftrightarrow 78 < 82-1{,}96 \cdot \frac{7}{\sqrt{30}}$$

d.h. 78 liegt nicht im Konfidenzintervall!

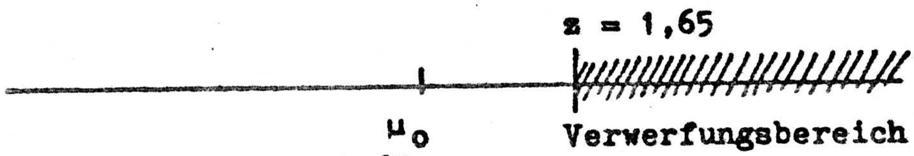
5.2. Die Methode der Hypothesenprüfung mittels Konfidenzintervall hat gegenüber den anderen Methoden (z-Wert, p-Wert) den Vorteil, daß sie zusätzlich auch eine Vorstellung von der wahrscheinlichen Größe des unbekannten Parameters gibt. Sie ist also für den Schüler anschaulicher.

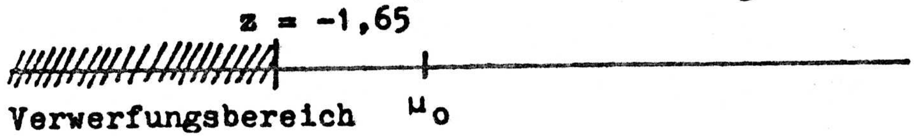
5.3. Die wohl häufigste Art, in der die Bekanntgabe der Ergebnisse von statistischen Überprüfungen von Forschungshypothesen erfolgt, ist die Angabe des Signifikanzniveaus. Die hier vorgeschlagene Methode ermöglicht eine einfache Erklärung dieses Begriffes.

5.4. Die Methode der Hypothesenprüfung mittels Konfidenzintervall ist sehr universell. Sie ist in gleicher Weise auch für die Student-t-Verteilung anwendbar und funktioniert ebenso problemlos auch bei "einseitiger Fragestellung". Allerdings wird vor "einseitigen Tests" und damit vor "einseitigen Signifikanzniveaus" dringend abgeraten!



Beim "einseitigen Test" besteht der Verwerfungsbereich aus nur einem Teil und man hat nur einen (und zwar einen kleineren!) kritischen z-Wert (z.B. für 5 %: $z = 1,65$ oder $z = -1,65$):





Bei der Methode der Hypothesenprüfung mittels Konfidenzintervall ist daher bei "einseitigen Tests" das Konfidenzintervall unsymmetrisch (z.B. für 5%: $(-\infty|\bar{x}+1,65\ \sigma_{\bar{x}})$ oder $(\bar{x}-1,65\ \sigma_{\bar{x}}|\infty)$. Ansonst ist die Vorgangsweise nicht wesentlich anders: "Für 95% der $\bar{x}$-Werte gilt: Entweder enthält das Konfidenzintervall $(-\infty|\bar{x}+1,65\ \sigma_{\bar{x}})$ bzw. $(\bar{x}-1,65\ \sigma_{\bar{x}}|\infty)$ den Wert μ oder der Wert μ ist größer (kleiner) als die rechte (linke) Grenze des Konfidenzintervalls $(-\infty|\bar{x}+1,65\ \sigma_{\bar{x}})$ (bzw. $(\bar{x}-1,65\ \sigma_{\bar{x}}|+\infty)$)".

Obwohl also auch "einseitige Tests" mit der Methode des Konfidenzintervalls problemlos lösbar sind, wird vor "einseitigen Tests" in der Schule sowohl aus didaktischen wie auch aus sachlichen Gründen abgeraten:

a) didaktische Gründe:

Der Wechsel des Konfidenzintervalls (z.B. für 5 %: $(\bar{x} \pm 1,96 . \sigma_{\bar{x}})$ (zweiseitig), $(\bar{x} \pm 1,65 . \sigma_{\bar{x}})$ (einseitig)) ist für den Schüler verwirrend und eine unnötige Quelle für Fehler.

b) sachliche Gründe:

Einseitige Tests haben eine Änderung der Signifikanz zur

Folge, und zwar so, daß sich eine Vergrößerung des Risikos (Fehlers) erster Art ergibt: Die Hypothese H_0 wird eher verworfen, obwohl sie richtig ist.

Da diese Änderung der Signifikanz durch den (vielleicht nur erwarteten oder erhofften?) Ausgang des Experiments verursacht wurde, bedeutet dies die Einführung eines subjektiven Faktors in die gerade zum Zwecke größerer Objektivität eingeführte Hypothesenprüfung. Überzeugtheit bezüglich der Richtung des Ergebnisses ist etwas Subjektives. D.h., einen einseitigen Test zuzulassen, wenn jemand "fühlt" oder "glaubt", daß die Richtung des Ergebnisses einigermaßen sicher ist, hätte zur Folge, daß ein Experiment für Herrn Maier (der sich hinsichtlich der Richtung des Ergebnisses sicher fühlt) signifikant und für Herrn Müller (der sich hinsichtlich der Richtung nicht so sicher fühlt) nicht signifikant wäre.

<u>Beispiel 7:</u>

Nach bisheriger Erfahrung beträgt das Durchschnittsergebnis von Studenten im Fach Ökonomie 65 Punkte (Standardabweichung 16 Punkte). Eine neue Lehrmethode wird eingeführt und eine Stichprobe von 64 Studenten ausgewählt. Ihr Mittelwert beträgt 69 Punkte. Ist die neue Lehrmethode a) signifikant verschieden? b) signifikant besser? (α = 5 % in beiden Fällen)

LÖSUNG:

a) zweiseitiger Test:

95 % - Konfidenzintervall: $\left(69 \pm 1{,}96 \cdot \frac{16}{\sqrt{64}}\right) = (65;\ 73)$

Da der "alte" Mittelwert $\mu = 65$ nicht außerhalb dieses Konfidenzintervalls liegt, kann H_0 ("$\mu = 65$") nicht verworfen werden.

Die neue Lehrmethode ist nicht signifikant verschieden von der alten Lehrmethode (und somit natürlich auch nicht signifikant besser als die alte Methode!)

b) einseitiger Test:

95 % - Konfidenzintervall: $\left(69 \pm 1{,}65 \cdot \frac{16}{\sqrt{64}}\right) = (66;72)$

Da der "alte" Testwert μ = 65 nicht in diesem Konfidenzintervall liegt, wird H_0 verworfen. Die neue Lehrmethode ist signifikant besser!

Wie das folgende Beispiel zeigt, unterscheiden sich einseitige und zweiseitige Tests im Ergebnis meistens überhaupt nicht, so daß die Unterscheidung auch vom Standpunkt der Praxis vom Schüler mit Recht erspart werden kann.

Beispiel 8:

Ein Sozialarbeiter vermutet (!), daß Erwachsene, die auf Sozialfürsorge angewiesen sind, im Mittel weniger als 5 Jahre Schulbildung besitzen (einseitige Fragestellung!). Eine Zufallsstichprobe von 169 solcher Erwachsenen ergab einen Mittelwert von $\bar{x}$ = 4,6 Jahren Schulbildung und eine Standardabweichung von s = 3,9 Jahren. Wird die Vermutung des Sozialarbeiters durch diese Daten gestützt? (α = 5 %)

LÖSUNG:

a) zweiseitiger Test:

95 % - Konfidenzintervall: $\left(4{,}6 \pm 1{,}96 \cdot \frac{3{,}9}{\sqrt{169}}\right) = (4{,}0;5{,}2)$

H_0("$\mu \geq 5$") ist nicht zu verwerfen. Die Vermutung des Sozialarbeiters wird nicht bestätigt.

b) einseitiger Test:

95 % - Konfidenzintervall: $\left(4{,}6 \pm 1{,}65 \cdot \frac{3{,}9}{\sqrt{169}}\right) = (4{,}1;\ 5{,}1)$

Gleiches Ergebnis wie bei a)!

5.6. <u>Dem Schüler zu vermittelndes "kritisches Wissen"</u>:

Im Hinblick auf die Grundaufgabe der österreichischen Schule (Erziehung der Schüler zu selbständigem Urteil, zu Kritikfähigkeit, zum Erkennen von Manipulation, zu sachgerechter Verwertung von Information usw.) wäre es wichtig, den Schülern folgendes Grundwissen über Hypothesenprüfung zu vermitteln:

5.6.1. Die Festlegung des Signifikanzniveaus ist weitgehend subjektiv. Sie wird bestimmt von der Stärke des Wunsches zur Vermeidung von sogenannten Fehlern vom Typ I, d.h. vom Wunsch, das Risiko, eine richtige Hypothese H_0 zu verwerfen, möglichst klein zu machen. Je stärker dieser Wunsch, desto kleiner wird das Signifikanzniveau gewählt.
Die Trennlinie zwischen wahrscheinlich und unwahrscheinlich und damit zwischen Ablehnung und Annahme von H_0 hängt also davon ab, wie gewiß (subjektiv!) der Forscher oder politische Entscheidungsträger eine inkorrekte Entscheidung (Verwerfung von H_0, obwohl richtig) vermeiden will.

5.6.2. Das Signifikanzniveau muß im Interesse der Objektivität <u>vor</u> Ermittlung der Daten (Durchführung des Experiments) und zwar möglichst "durch andere Personen als den Statistiker festgelegt werden; es handelt sich um eine politische Entscheidung" (Yamane, Statistik I, S.204). Außerdem darf das vorher festgelegte Signifikanzniveau während einer Untersuchung nicht mehr geändert werden.

5.6.3. Es kann sein, daß der durch das Experiment ermittelte Wert $\bar{x}$ durch einen ungünstigen Zufall so weit vom wahren Wert μ abweicht, daß das Konfidenzintervall diesen Wert μ nicht mehr enthält, obwohl dieser Wert μ der wahre Mittelwert der Grundgesamtheit ist: H_0 wird dann verworfen, obwohl es richtig ist (Fehler vom Typ I). Das Risiko, einen solchen Fehler zu begehen, kann nur dadurch verringert werden, daß man das Konfidenzintervall vergrößert (90 % → 95 % → 99 %) und damit das Signifikanzniveau verkleinert (0,1 → 0,05 → 0,01). Jedoch (Warnung!): Gleichzeitig mit der Verkleinerung des Signifikanzniveaus wächst die Wahrscheinlichkeit für eine andere Fehlerart: Die Nullhypothese H_0 wird beibehalten, obwohl sie falsch ist (Fehler vom Typ II).

5.6.4. Es haben sich in Medizin, Biologie, Wirtschaftswissenschaft, Soziologie, Psychologie und Pädagogik drei verschieden strenge Maßstäbe der Verläßlichkeit (Signifikanzniveaus) eingebürgert: 0,05/0,01/0,001.

Als Auswahlregeln werden angegeben (z.B. von Lienert, Verteilungsfreie Methoden der Biostatistik, 1962):

0,05: Es soll eine Hypothese bestätigt oder in einem Spezialfall nachgewiesen werden, die allgemein anerkannt ist oder zwanglos in den Rahmen einer anerkannten Theorie paßt oder die nur von geringer theoretischer oder praktischer Bedeutung ist.

0,01: H_1 widerspricht einer anerkannten Theorie oder enthält einen theoretisch oder praktisch bedeutsamen Befund.

0,001: Bei Untersuchungen von ganz besonderer Wichtigkeit ("hochsignifikant").

5.6.5. Die Formulierung "H_0 wir beibehalten, abgelehnt, zurückgewiesen, verworfen, akzeptiert u.ä." drückt aus, daß bei der Hypothesenprüfung grundsätzlich keine Sätze als falsch oder richtig erwiesen (falsifiziert oder verifiziert), sondern nur als falsch oder richtig bezeichnet (subjektiv dafür gehalten!) werden. Es geht hier also um subjektive Wahrscheinlichkeiten! "Keinen Grund zur Ablehnung der Nullhypothese zu haben, ist noch kein Beweis für die Richtigkeit ... Annahme der Nullhypothese ist gewissermaßen ein Freispruch aus Mangel an Beweisen. Was man jetzt braucht, ist aber der Beweis der Unschuld" (Adam, Einführung in die medizinische Statistik, 1963).

5.6.6. Im Sinne der vom Lehrplan geforderten Erziehung der Schüler zu Kritikfähigkeit wäre es natürlich sehr wertvoll, auch noch kurz (aber eindringlich!) auf den wichtigen Unterschied zwischen statistischer und praktischer Signifikanz hinzuweisen. Wenn die statistische Analyse von Forschungsergebnissen statistisch signifikante Ergebnisse liefert, so darf man daraus nicht schließen, daß diese Ergebnisse auch notwendigerweise praktisch signifikant sind. Nur persönliches Fachwissen und große Erfahrung können dies entscheiden! Univ.Prof. Dr.Wohlzogen (Universität Wien) hat bei einem Vortrag über "Die Statistik in der Medizin" anläßlich der 150-Jahr-Feier des Österreichischen Statistischen Zentralamtes folgende sehr bemerkemsnwerte Feststellung gemacht: "Nun zu einem anderen Gesichtspunkt noch, dem Begriff der statistischen Signifikanz und der medizinischen Relevanz. Es werden oft Studien einfach nicht zur Publikation angenommen, weil sie nicht mit dem Epitheton ornans P = 0,05 versehen sind. Andererseits werden Studien angenommen, wenn diese ominöse Zahl, diese heilige Kuh, die endlich einmal geschlachtet werden sollte, für Ergebnisse zutrifft, die aber medizinisch keine Bedeutung haben. Wir müssen hier unterscheiden, es kann eine Studie zu medizinisch relevanten Ergebnissen geführt haben, die aber nicht statistisch signifikant sind ... Es ist ein Unfug, wenn wir im Verlauf eines Experimentes sehen, daß medizinisch nichts herauskommt, daß die Studie weitergeführt wird, nur damit wir am Schluß endlich dann dieses $P = 0{,}05$ hinter die Ergebnisse, die medizinisch keine Bedeutung haben, anhängen können." (Mitteilungsblatt der Österr.Gesellschaft für Statistik und Informatik, Nr.36, 1979, S.164).

Mag.Roman Laußermayer,
Gymnasium Imst, Tirol