Manfred Borovcnik, Klagenfurt

EIN DIREKTER ZUGANG ZUR BEURTEILENDEN STATISTIK

1. Einleitung - Zur Gestaltung des Unterrichts in Wahrscheinlichkeitsrechnung und Statistik

Bei der Rechtfertigung von Stochastik-Unterricht tauchen immer wieder zwei Argumente auf:

> Es handelt sich dabei um eine eigene Denkweise, an die Probleme heranzugehen - in Abhebung vom noch immer deterministisch geprägten Weltbild.
>
> Die Durchdringung der Arbeitswelt, vieler Wissenschaften und des täglichen Lebens - die enorme Anwendungsbreite von Stochastik - macht es erforderlich, sich damit auseinanderzusetzen.

Man beschränkt sich allerdings zu sehr auf den Bereich Wahrscheinlichkeitsrechnung und damit auf die innermathematischen Beziehungen.
Beschreibende Statistik erscheint vom mathematischen Problemgehalt zu leicht - beurteilende Statistik als zu schwierig und wird häufig nicht unterrichtet. Zu einem verlieren dadurch die angeführten Rechtfertigungsargumente sehr viel an Überzeugungskraft, zum anderen ist durch die Erarbeitung der mathematischen Beziehungen ein ausgewogenes Verständnis der erarbeiteten Begriffe und Methoden (i.a. noch) nicht gewährleistet.

Wir möchten im folgenden

- zentrale Ideen der beschreibenden Statistik herausarbeiten, an denen deutlich wird, daß dieses Teilgebiet interessant für den Unterricht gestaltet werden kann,
- einen direkten Zugang zur beurteilenden Statistik aufzeigen. Wie können statistische Fragestellungen erarbeitet werden, ohne daß man den Wahrscheinlichkeitskalkül aufbaut? Wahrscheinlichkeitsrechnung erscheint dabei nur als (mögliches) Anhängsel zur Formalisierung (wie soll mit Wahrscheinlichkeiten kohärent umgegangen werden?) und damit zur "mathematischen" Rechtfertigung der "empirisch" vorweggenommenen Zusammenhänge.

2. Häufigkeitsverteilungen

Zur Erfassung bestimmter Sachverhalte, zur Beantwortung bestimmter

Fragestellungen müssen oft Daten erhoben werden. Das Ergebnis einer Urliste mit mehreren Variablen bietet jedoch eine Fülle von Informationen, die es erst zu ordnen gilt.

Tabelle: Daten von 25 befragten Personen

Person	Alter	Geschlecht	Familienstand	Kinderzahl	abgeschlossene Schulbildung der Ehegattin	abgeschlossene Schulbildung des Ehegatten
Mandl	70	männl.	verh.	3	7 Kl.AHS	6 Kl. AHS
Binder	28	weibl.	getr.	3	Matura AHS	Matura HAK
Kassl	47	männl.	verh.	1	3 Kl.HAK	Matura AHS
Fellner	48	männl.	verh.	3	Matura HAK	7 Kl. AHS
Polzer	23	männl.	verh.	0	HS	Matura HTL
Mack	69	weibl.	gesch.	3	-	3 Kl. HTL
Draxl	31	weibl.	verh.	1	1.Dipl.	Magisterium
Peuker	70	weibl.	ledig	0	-	HS
Hütter	80	weibl.	verw.	0	-	HS
Strobl	37	weibl.	verh.	3	1.Dipl.	7 Kl. AHS
Bauer	65	weibl.	ledig	0	-	HS
Gigler	71	weibl.	verw.	2	-	HS
Steiner	41	weibl.	getr.	3	Matura	Matura AHS
Isola	70	männl.	verh.	5	HS	VS
Zwick	56	weibl.	verh.	1	Matura	3 Kl. HTL
Walzl	34	weibl.	verh.	4	HS	HS
Krenn	48	weibl.	verh.	6	Matura	3 Kl. HAK
Zöttl	43	weibl.	gesch.	3	-	3 Kl. HAK
Vieler	50	weibl.	verh.	1	Matura	7 Kl. AHS
Mitterer	24	weibl.	verh.	2	Matura	HS
Hütl	23	männl.	verh.	1	1.Dipl.	Matura AHS
Angerer	47	weibl.	verh.	4	Doktorat	Doktorat
Kaus	63	männl.	ledig	0	-	Matura HTL
Gasser	31	weibl.	verh.	4	7 Kl. AHS	Matura AHS
Spitzer	21	weibl.	getr.	3	HS	HS

Zunehmende Raffung kann Klarheit darüber verschaffen, was für die untersuchte "Personengruppe" (die untersuchten Objekte) typisch ist bzw., wo sich darüber hinaus Besonderheiten (Ausreißer etwa) ergeben. Zuerst werden die Daten von den Personen losgelöst, des weiteren wird zunächst jede einzelne Variable für sich isoliert betrachtet. Häufigkeitstabellen geben einen raschen Überblick darüber, wie häufig die verschiedenen Variablenwerte aufgetreten sind.

Geschlecht	Anzahl	relative Häufigkeit in %
männlich	7	28 %
weiblich	18	72 %
	25	100 %

Kinderzahl	relative Häufigkeit in %
0	20
1	20
2	8
3	32
4	12
mehr als 4	8

Alter	relative Häufigkeit in %
20 - 30	20
30 - 40	16
40 - 50	24
50 - 60	8
60 - 70	12
70 - 80	16
≥ 80	4

Tabellen: Verteilung des Geschlechts, der Kinderzahl bzw. des Alters in obiger Erhebung

Die graphische Darstellung erhöht die Lesbarkeit sehr:

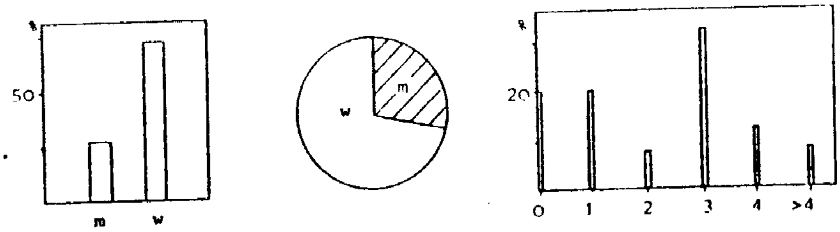


Bild 1: Möglichkeiten der Darstellung: Stabdiagramm, Kreisdiagramm bzw. Histogramm

Bei quantitativen Variablen führt oft erst geeignete Klassifizierung der Daten zu mehr Übersicht und zu erkennbaren "Mustern" in der graphischen Darstellung. Dabei ergibt sich das Problem der Wahl einer geeigneten Klasseneinteilung: der Gewinn an Übersicht ist gegen den Verlust an ursprünglich vorhandener Information abzuwägen.

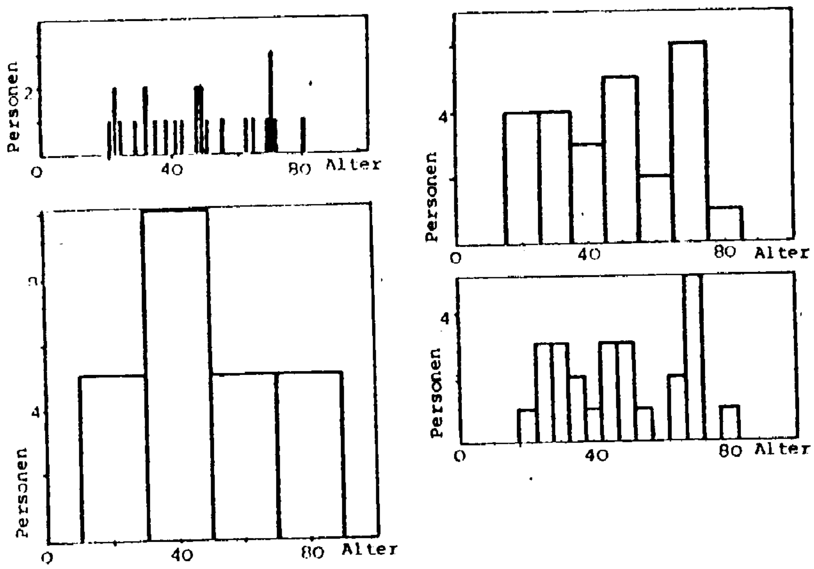


Bild 2: Verteilung des Alters - ohne bzw. mit verschiedenen Klasseneinteilungen. Wahl der Klasseneinteilung: GEWINN an ÜBERSICHT - VERLUST an INFORMATION

Werden ungleich breite Klassen gewählt, so kann sich durch die Wahl der Darstellung ein optisch ungünstiger Eindruck ergeben:

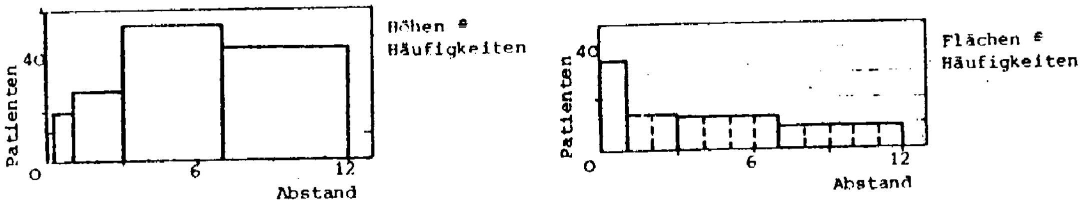


Bild 3: Histogramm bei Klasseneinteilung mit unterschiedlichen Klassenbreiten - Höhen ≙ Häufigkeiten gegen Flächen ≙ Häufigkeiten; Abstand zwischen ersten Symptomen und Herzinfarkt.

In der Regel erweist es sich als günstig, Histogramme so zu zeichnen, daß die Flächen der einzelnen Blöcke proportional zu den entsprechenden Häufigkeiten der zugehörigen Intervallklassen sind.

Lernziele:

- Elementare Kulturtechniken, wie Lesen und Erstellen von Tabellen und graphischen Darstellungen.
- Leitidee: Raffung von Information:
  - Gewinn von Übersicht,
  - Vermittlung von Information,
  - Erkennung von "Mustern" und Besonderheiten (etwa Ausreißern).
- Metaziele:
  - Durch Raffung der Daten geht der Bezug zu Personen verloren,*) ebenso (zunächst) die Information über Zusammenhänge zwischen verschiedenen Variablen.
  - Methoden der Darstellung hängen von der Art der Fragestellung und vom Zweck der Untersuchung ab (Manipulation!).

3. Mittelwerte

Stabdiagramme bieten eine gute Übersicht über Ergebnisse, stellen jedoch eine simultan dargebotene vielfältige Information dar, die oft nur einen qualitativen Eindruck hinterläßt. Wir sind es gewohnt, Sachverhalte mit Zahlen zu erfassen und mittels dieser "Kennziffern" Vergleiche anzustellen. Dies können wir leicht, wenn die Variablen quantitativ sind.

Beispiel: Bernd und Alfred spielen Golf. Ihr Ergebnis als Stabdiagramm:

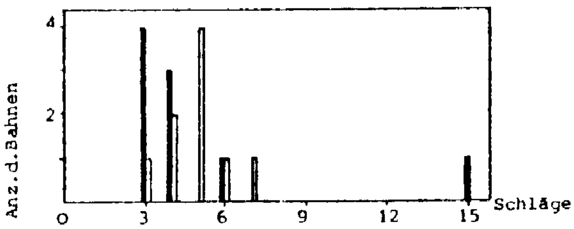
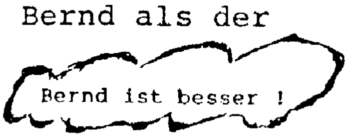


Bild 4: Stabdiagramm des Ergebnisses der Golfpartie Alfred (▮) gegen Bernd (▯).

Gesucht ist eine für den jeweiligen Golfer typische Zahl, die die Häufigkeitsverteilung repräsentiert, die das Ergebnis jedes Golfers in dieser einen Kennziffer zusammenfaßt. Diese Zahl soll dann die Fragestellung beantworten helfen. Nimmt man den Mittelwert als Basis für die Entscheidung, so gilt Bernd als der bessere Golfer.



----

*) Dies ist eine Besonderheit statistischer (im Gegensatz zu historischer) Betrachtungsweise, die allzu leicht übersehen wird, sie bildet jedoch eine der Ursachen, warum Ergebnisse statistischer Untersuchungen für die Einzelperson für nicht relevant erachtet werden.

Als grundlegende Eigenheiten des Mittelwertes sollen erarbeitet werden:

- Aus dem Mittelwert läßt sich die Summe der Variablenwerte rekonstruieren.
- Die Summe der Variablenwerte muß eine sinnvolle Größe geben, dazu ein Beispiel:
  Wohnungsstatistik einer Gemeinde
  1960 500 Wohnungen,
  1970 1500 Wohnungen,
  daher gab es zwischen 1960 und 1970 durchschnittlich 1000 Wohnungen ??
- Ob der Mittelwert eine sinnvolle Raffung der Information ist, hängt vom Zweck, vom Standpunkt des Betrachters ab.

  Beispiel: Gehälter einzelner Angestellter in einem kleinen Betrieb:
  6000, 6000, 8000, 24000; $\bar{x} = 11000$?

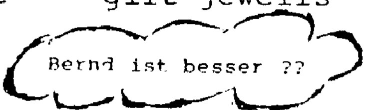
  - Für den Buchhalter, der am Monatsende die Gehälter auszuzahlen hat: $n\bar{x}$ = auszuzahlender Betrag.
  - Für jemanden, der sich um einen Posten in dem Betrieb bewirbt ??
- Der Mittelwert ist bezüglich Ausreißern empfindlich.

Heuristische Vorstellungen für den Mittelwert helfen bei der Deutung von Ergebnissen sowie beim Verständnis technischer Details*), wie der

- Berechnung von Mittelwerten aus Häufigkeitsverteilungen, auch mit Klasseneinteilung,
- Methode des vorläufigen Mittels zur Rechenerleichterung.

- Mittelwert als Gleichgewichtswert - Hebelgesetz.
- Mittelwert als Ausgleichswert - geometrische Veranschaulichung.**)

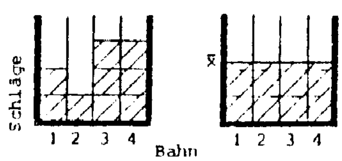
Der Mittelwert ist jedoch nicht die einzig mögliche Kennziffer, mittels der die Frage nach dem besseren Golfer beantwortet werden kann: Auf der Basis von Modus bzw. Zentralwert***) gilt jeweils Alfred als der bessere Golfer.



----

*) Hier bietet die rezeptartige Statistik eine Fülle von Formeln, die sich allerdings nur als Resultat vernünftiger Rechenvereinfachungen herausstellen. Es erscheint vordringlich, den Prozeß der Rechenvereinfachung aus einer konkreten Situation heraus in den Vordergrund zu stellen und die Formeln allenfalls auch wegzulassen bzw. sie in dynamische Rechenschemata aufzulösen.

**) Es gilt, die Schlagzahl der Golfer gleichmäßig auf die einzelnen Bahnen zu verteilen. Wir stellen uns einen Sandbehälter mit Fächern für die einzelnen Bahnen vor, der in jedem Fach entsprechend der Schlagzahl mit Sand gefüllt ist. Entfernen wir nun die Trennwände und ebnen den Sand ein, so erhalten wir auf allen Bahnen gleich viel Sand, nämlich gerade dem Mittelwert entsprechend.



***) Häufigster Wert einer Urliste bzw. derjenige Wert, der die Urliste halbiert: die Hälfte der Daten liegt über bzw. unter diesem Wert.

Die eingangs beinahe definitiv getroffene Feststellung, daß Bernd der bessere Golfer ist, muß zurückgenommen werden: Die Beurteilung des Sachverhalts hängt von der Kennziffer ab, die dazu herangezogen wird.

Welche Kennziffer sollen wir nun heranziehen ?

Diese Frage kann nicht allgemein beantwortet werden !

Statistische Ergebnisse hängen von der gewählten Methode ab, sie verlieren ohne den Kontext viel an Aussagekraft. Statistische Problemstellungen sind offene Problemstellungen. Festlegungen und Bewertungen sind bei der mathematischen Modellbildung notwendig. Etwa soll durch die Heranziehung bestimmter Kennziffern ein Phänomen der Realität erfaßt werden. Man kann nicht innermathematisch entscheiden, welches Maß das richtige ist.

Es handelt sich um nicht eindeutig beantwortbare Fragen.

Diese Willkür soll auch im Rahmen der beschreibenden Statistik herausgestellt werden. Es gibt wohl Abgrenzungen, unter welchen Bedingungen gewisse Kennziffern nicht anwendbar sind ("metrische Eigenschaft" etwa), sie können jedoch nur Orientierungen geben, immer bleiben Fragen offen.*)

Lernziele:

- Operatives Verständnis vom Mittelwert: Berechnen von Mittelwerten in verschiedensten Zusammenhängen.
- Heuristische Vorstellungen zum Mittelwert: Gleichgewichtswert - Ausgleichswert.
- Leitidee: Quantifizierung von Sachverhalten mittels Kennziffern.
- Metaziele:
  - Aussagekraft des Mittelwerts hängt vom Zweck ab.
  - Statistische Fragestellungen sind prinzipiell offen.

----

*) Hier bietet sich etwa die Diskussion der Sinnhaftigkeit der Bildung des Notendurchschnitts in der Schule an. Noten können allenfalls der Größe nach angeordnet werden, keinesfalls kann man aber sagen, daß die Distanz (der Leistungsunterschied) zwischen den Noten 1 und 2 gleich groß ist wie die (der) zwischen den Noten 4 und 5. Dies müßte jedoch eigentlich erfüllt sein, wenn man der Durchschnittsbildung eine sinnvolle Deutung abgewinnen möchte. Notengebung ist überhaupt ein sehr heikles Kapitel bezüglich Quantifizierung von Sachverhalten. Daran anschließend kann man auch die Grenzen der Quantifizierbarkeit und alternative Herangehensweisen diskutieren. IQ-Tests und Leistungstests bei der Überprüfung von Bewerbern um eine Stelle sind weit verbreitet und ähnlicher Kritik ausgesetzt (Politische Bildung).

4. Streuungsmaße

Die Repräsentation eines Sachverhalts durch eine Kennzahl ist für viele Zwecke unzureichend. Der Mittelwert kennzeichnet nur das Zentrum einer Verteilung, wesentlich ist aber auch, wie die Werte um dieses Zentrum verteilt sind. Die Raffung durch eine Kennziffer ist zu extrem. Wir würden ja nicht Statistik treiben, wenn zwischen den zu untersuchenden Objekten keinerlei Variabilität bezüglich der interessierenden Variablen besteht.*) Variabilität stellt daher eine zentrale Idee der Statistik dar. Auch zu ihrer mathematischen Erfassung und Beschreibung gibt es wieder verschiedenste Kennzahlen.

- Variationsbreite,
- mittlere lineare Abweichung,
- mittlere quadratische Abweichung,
- Standardabweichung.

Wichtig erscheint dabei, nicht durch die Vielfalt an Formeln Verwirrung zu stiften, sondern prozeßartig aufzuzeigen, wie eine intuitive Idee mathematisch erfaßt wird, und, daß es alternative Möglichkeiten gibt.**)

Die "Brauchbarkeit" der Standardabweichung zur Erfassung kann durch das folgende Beispiel demonstriert werden.

Hans hat bei drei Schularbeiten sieben Punkte erreicht.

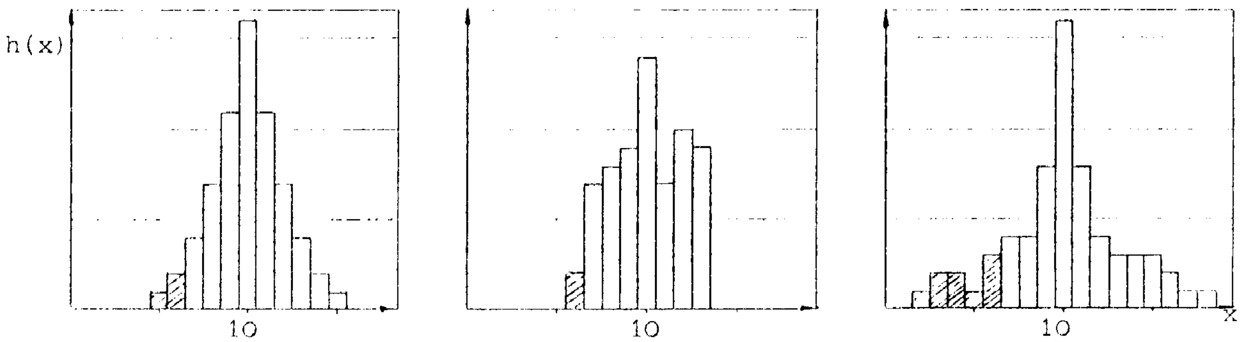


Bild 5: Punkteverteilung bei drei verschiedenen Schularbeiten - schraffiert jeweils der Bereich bis einschließlich sieben Punkte.

Ist er daher bei allen Schularbeiten gleich gut gewesen ?

----

*) Die weite Verbreitung von Mittelwerten paßt ganz gut in ein deterministisches Weltbild: der Zufall ist im gängigen Verständnis von Mittelwerten weitgehend "eliminiert".

**) Wiederum kann man nicht mit Mitteln der Mathematik die Wahl einer bestimmten Kennzahl rechtfertigen.

Die absoluten (Punkt-)Werte x_i können zur Beurteilung der Frage offenbar nicht herangezogen werden, das zeigt die Betrachtung der Histogramme: Die Ergebnisse der Schularbeiten weisen ein unterschiedliches Profil auf. Es bietet sich an:

- anzugeben, wieviele Schüler jeweils schlechter als Hans abgeschnitten haben,
- die Kennzahlen $\bar{x}$ und s zu verwenden, etwa um anzugeben, wieviele Standardabweichungen Hans unter dem Mittelwert liegt: $z_i = \frac{x_i - \bar{x}}{s}$, diese normierten Werte differenzieren das Ergebnis und sollen wohl ähnliches leisten wie in der ersten Variante: aus normierten z-Werten soll man auch ungefähr Rangplätze ablesen können.

  Die sogenannten s-Regeln lauten:

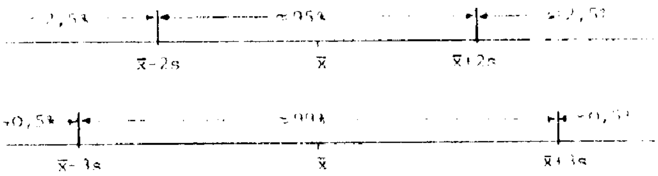


Bild 6: 2 s- (3 s-) Regel für symmetrische Verteilungen: Werte außerhalb des Intervalls ($\bar{x}$-2s, $\bar{x}$+2s) (bzw. ($\bar{x}$-3 s, $\bar{x}$+3s) zählen zu den 5% (0,5%) extremsten (atypischen) Werte für die Häufigkeitsverteilung, je 2,5% (%.) liegen darunter bzw. darüber.

Diese s-Regeln gilt es, empirisch bei einigen Häufigkeitsverteilungen zu überprüfen. Dabei zeigt sich, daß diese Regeln für verschiedene Typen von Verteilungen unterschiedlich gut brauchbar sind. Setzt man sie voraus, so stellen sie ein weitreichendes Instrument dar, sie dienen:

- zum Vergleich von Werten aus verschiedenen Häufigkeitsverteilungen,
- zur Einstufung eines Wertes innerhalb der bestehenden Verteilung zur Beurteilung, ob dieser Wert innerhalb dieser Verteilung, als extrem gelten soll oder nicht.*)

Lernziele:

- Berechnen der empirischen Standardabweichung (bei Klasseneinteilung, mit Rechenerleichterung) sowie der z-Werte.
- Leitziele:
  - Variabilität und ihre mathematische Erfassung,
  - Normierung von Variablenwerten: z-Werte geben ungefähr Auskunft über Rangplätze (zumindest für symmetrische Verteilungen).
- Metaziel: Wissen um Probleme mit der mathematischen Erfassung der intuitiven Vorstellung Variabilität.

----

*) Liegt ein Wert x_i etwa unterhalb von $\bar{x}$-2s, so sind nur ca. 2,5% der Werte kleiner als er selbst, er wird daher als extrem eingestuft.

5. Empirische Verteilungen

Die Untersuchung verschiedenster empirischer Verteilungen ergibt, daß die s-Regeln relativ gut brauchbar sind, falls die Verteilung symmetrisch glockenförmig (vom Typ) ist, daß sie aber bei schiefen Verteilungen "versagen".

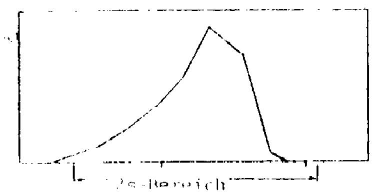


Bild 7: Für schiefe Verteilungen lassen sich aus z-Werten keine Rangplätze ablesen, der 2s-Bereich geht in der Regel "weit" über die tatsächlich beobachteten Werte hinaus.

Bei dieser Typisierung empirischer Verteilungen kann und soll auch ein Übergang von Histogrammen als Darstellung des Ergebnisses konkreter Erhebungen zu sogenannten idealisierten Häufigkeitsverteilungen vollzogen werden.

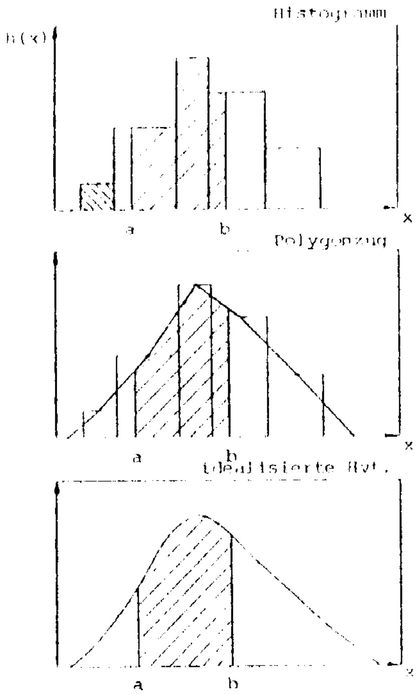


Bild 8: Nur wenn Intervallgrenzen mit Klassengrenzen ident sind, gilt im Histogramm exakt:
Fläche = relative Häufigkeit der Werte, die in das betreffende Intervall hineinfallen. Angestrebt wird eine Ersetzung des Histogramms durch einen Polygonzug und in weiterer Folge durch den Graphen einer stetigen Funktion - der idealisierten Häufigkeitsverteilung, für die auch gelten soll:

Fläche $\approx$ relative Häufigkeit.

Die idealisierte Häufigkeitsverteilung dient

- zur Beschreibung einer Erhebung und zwar ihrer allgemeinen Züge, ihrer "Muster",
- zur Einsicht über Zusammenhänge - Interpretation des Musters sowie von Abweichungen davon,
- als Basis für zukünftige Erhebungen: Prognosen

Sie ist ferner der weiteren mathematischen Behandlung leichter zugänglich.
Es gibt viele statistische Fragestellungen, bei denen die Information aus Häufigkeitsverteilungen nicht weiter hilft; dazu ein Beispiel:
Das Sterberisiko verschiedener Altersklassen ist zu beurteilen.

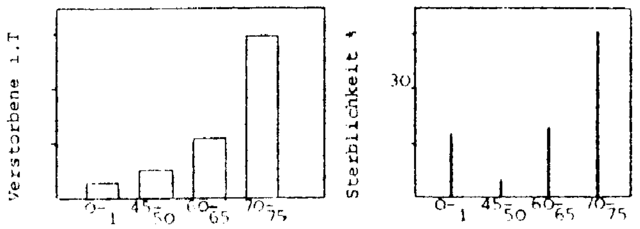


Bild 9:
a) Anzahl der Verstorbenen verschiedener Altersklassen.
b) Beziehungszahlen: $\frac{\text{Verstorbene/Altersklasse}}{\text{Lebende/Altersklasse}}$

Erst die Berechnung von Beziehungszahlen (b) hilft, das Sterberisiko verschiedener Altersklassen miteinander zu vergleichen.

Lernziele:

- Verfestigung der elementaren Techniken.
- Abgrenzung der s-Regeln: symmetrische, schiefe Verteilungen.
- Leitidee: "Muster" - Typisierung von statistischen Sachverhalten zur Beschreibung und zur Prognose.
- Metaziele: - Anwendungsvielfalt
  - Interpretation von Daten in Zusammenhängen.

6. Wo stehen wir ?

Wir haben versucht, in zentrale Problemstellungen der (beschreibenden) STATISTIK einzuführen. Es wurde dabei nicht eigens problematisiert, daß es sich jeweils nur um eine Teilmenge, eine Stichprobe der Grundgesamtheit handelt. Es ging um Erfassung und Vermittlung von Information, Bewertung von Sachverhalten, Erkennen von Regularitäten, im einzelnen waren folgende Konzepte von leitender Bedeutung:

- Raffung von Daten,
- Quantifizierung von Sachverhalten,
- Idee der Variabilität,
- Normierung und Bewertung von Einzelwerten innerhalb und zwischen verschiedenen Verteilungen,
- Idee der statistischen "Muster".

Wir sind bislang ohne Wahrscheinlichkeiten und ohne Wahrscheinlichkeitsrechnung ausgekommen. Die folgenden zwei Abschnitte sind von entscheidender Bedeutung. Der Standpunktwechsel: "Untersuchte Gesamtheit ist nur eine Stichprobe einer umfassenderen Grundgesamtheit" wirft eine Reihe von Problemen auf. Kann mann, und wenn man kann, wie kann man die Ergebnisse der Stichprobe auf die größere Gesamtheit übertragen? Welche Fehler sind dabei möglich? Wir werden uns dabei an der Idee der repräsentativen Stichprobe, einem Minimodell der Gesamtheit, und den Auswahlverfahren, die dazu führen sollen, orientieren. Die Rolle, die zufällige Auswahl dabei spielt, ist zentral für ein Verständnis der beurteilenden Statistik, an deren Schwelle wir im Moment stehen.

7. Stichproben

Beispiel:

Eine Bundespräsidentenwahl steht bevor. Ein Meinungsforschungsinstitut wird beauftragt, eine Aussage über den derzeitigen Wählerwillen zu treffen. Bei der Wahl selbst werden ja alle Wähler "befragt". Deshalb und auch aus finanziellen Gründen wäre es unsinnig, jetzt schon - vorgreifend - alle zu befragen. So wählt dieses Meinungsforschungsinstitut eine gewisse Anzahl von Wählern als stellvertretende Gruppe aus, befragt diese und rechnet das Ergebnis ihrer Umfrage auf die Gesamtanzahl aller Wähler um, um so eine Prognose stellen zu können.

Trifft die Generalannahme zu, daß die Verhältnisse in der Stichprobe, bezüglich "relevanter" Merkmale strukturgleich mit denen in der Grundgesamtheit sind, so ist eine Übertragung bezüglich des in Frage stehenden Merkmals zulässig. Die zentrale Frage ist daher: Wie zieht man eine repräsentative Stichprobe, eine Stichprobe, die ein Minimodell der Grundgesamtheit darstellt?

Dazu gibt es zwei grundverschiedene Methoden:

a) Bewußte Auswahl - Quotenauswahl

Das Wahlverhalten sei (allein) durch die Merkmale Parteizugehörigkeit, Konfession und sozialer Status bestimmt. Wir denken uns die Bevölkerung als Würfel:

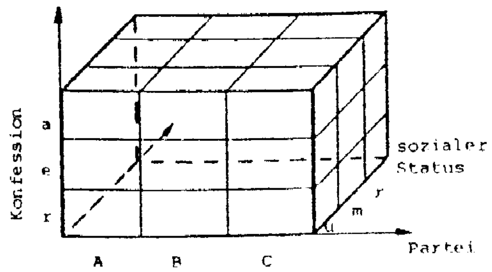


Bild 10: Schichtung der Bevölkerung nach mehreren Variablen - symbolische Darstellung als Würfel.

Durch die Werte der einzelnen Merkmale wird die Bevölkerung in Zellen zerlegt, der Anteil der Bevölkerung in bestimmten Zellen (etwa (B,m,r)) heißt Quote. Das erklärte Ziel der Quotenauswahl besteht darin, für die Stichprobe einen Miniwürfel gleicher Struktur, d.h. mit gleichen Besetzungszahlen (Quoten) bezüglich der bekannten Merkmale herzustellen, dies soll dann eine Übertragung auch der Verhältnisse bezüglich unbekannter Merkmale (im Beispiel des Wahlverhaltens) von der Stichprobe auf die Grundgesamtheit ermöglichen.

Das Problem bei bewußter Auswahl besteht darin, daß die Auswahl der für relevant erachteten (das Wahlverhalten "determinierenden") Merkmale schwierig ist und daß die Information über die Quoten allenfalls fehlt oder zumindest unsicher ist*).

b) Zufällige Auswahl

Information ist unsicher, man versucht daher, jegliche ((un)bewußte) Information auszuschalten, eine Trennung der Eigenschaften der auszuwählenden Elemente und dem Auswahlvorgang zu erreichen. Im Gegensatz zur bewußten Auswahl versucht man durch sogenanntes "zufälliges" Ziehen einer Stichprobe allen denkbaren Auswahlen gleiche "Chancen" einzuräumen.

Beispiel: Ein Reporter einer Rundfunkstation wird auf die Straße geschickt, um die Meinung der Bevölkerung dieser Stadt zu einer geplanten Gesetzesänderung - Fristenlösung etwa - zu erheben.

Können wir dies als zufällige Auswahl ansehen ?

Unsere Einschätzung, ob ein Auswahlvorgang als "zufällig" gilt, hängt vom Stand der Information und der Art der Fragestellung ab: Gibt es innerhalb von Gruppierungen relativ homogene Ansichten zum Befragungsthema, zwischen den einzelnen Gruppierungen jedoch starke Unterschiede? Sind verschiedene Gruppierungen über-(unter)repräsentiert?

Denkbar im Beispiel: Hausfrauen sind überrepräsentiert und reagieren überwiegend ablehnend der Einführung der Fristenlösung gegenüber. Wenn dagegen die Augenfarbe erhoben wird, so mag die Überrepräsentierung der Hausfrauen keine Rolle spielen.

Man kann nicht erklären, was zufällige Auswahl ist, sondern nur, was sie nicht bewirken soll: Einseitige Bevorzugung gewisser Stichproben. Die Urnenmethode zielt darauf ab, zu verhindern, daß Eigenschaften der auszuwählenden Objekte den Auswahlvorgang beeinflussen.

Urnenmethode:**) Alle N Elemente der Grundgesamtheit sind namentlich bekannt. N Kugeln, beschriftet mit den Namen der Elemente, ansonsten identisch, werden in eine Urne gelegt und sorgfältig gemischt, n Kugeln davon werden gezogen und stellen die Stichprobe dar.

Bei zufälliger Auswahl sind wir mit unserer Vorstellung am anderen Ende einer "Skala" angelangt als bei bewußter Auswahl (Quotenverfahren):

- maximale Ausnutzung von Information zur Bewertung der Anteile (Quoten) relevanter Gruppierungen,
- Ausschaltung jeglicher möglicher ((un)bewußter) Information, weitestgehende Trennung von auszuwählenden Elementen und dem Auswahlvorgang.

Paradox erscheint zufällige Auswahl zunächst insofern, als eine normale Stichprobe dieselbe Chance hat, gezogen zu werden wie eine extreme Stichprobe. Wie ist das mit der Idee des Minimodells zu vereinbaren?

Lernziele:

- Verschiedene Methoden der Auswahl und deren Vor- und Nachteile: Ausnutzung von Information - Ausschaltung von Information.
- Leitidee: Minimodell zur Übertragung der Verhältnisse von der Stichprobe auf die Grundgesamtheit.
- Metaziel: Die Bewertung, ob zufällige Auswahl vorliegt, hängt vom Informationsstand ab.

----

*) Man behilft sich z.B. so, daß man nicht verlangt, daß die Quoten aller Zellen für die Stichprobe erfüllt werden, sondern daß lediglich die sogenannten Randquoten (Prozentsatz der A-,B-,C-Wähler, der Konfession r,e,a usf.) eingehalten werden.

**) Auch diese Methode ist nicht frei von Problemen - ihre Brauchbarkeit hängt wesentlich vom "sorgfältigen Mischen" ab. Letzlich bleibt es eine Sache der Bewertung, ob man etwas als zufällige Auswahl einstuft.

8. Zufällige Auswahl und Verteilung der Kennzahl Stichprobenmittelwert

Beispiel: In einer Kleinstadt gibt es sechs Geschäfte, die Sporthemden einer bestimmten Qualität anbieten.

Geschäft	A	B	C	D	E	F
Preis	140,-	155,-	166,-	149,-	130,-	144,-

Wir erheben den Preis lediglich in drei Geschäften. Ist es sinnvoll, den mittleren Hemdenpreis in der Stichprobe auf die Grundgesamtheit zu übertragen ? *)

Die Fragestellung ist natürlich wirklichkeitsfremd: Wozu führen wir eine Erhebung durch, wenn wir die Grundgesamtheit schon kennen? Der Vorteil einer Stichprobe ist auch nicht unmittelbar einzusehen, wir könnten den Preis doch gleich in allen Geschäften erfragen. Wir opfern Wirklichkeitsnähe der (hoffentlich) besseren Verständlichkeit.

Hat sich in der konkreten Stichprobe A,C,D ergeben, so erhält man einen mittleren Preis $\bar{x} = 154,33$. Wie groß ist nun der Mittelwert in der Grundgesamtheit? Naiv würde man schätzen: $\mu \approx \bar{x} = 154,33$. Um die Genauigkeit dieser Schätzung zu beurteilen, wollen wir untersuchen, welche Werte der Stichprobenmittelwert überhaupt annehmen kann. Wir fassen diese Werte zu einer neuen Grundgesamtheit, der der Variablen Stichprobenmittelwert zusammen.

Der Vergleich Verteilung in der Grundgesamtheit - Verteilung der Variablen Stichprobenmittelwert zeigt:

Bild 11: Gegenüberstellung: Verteilung in der Grundgesamtheit - Verteilung der Variablen Stichprobenmittelwert.

- o der maximale Fehler passiert bei der Auswahl AEF ($\mu \approx 140,\dot{6}$) bzw. bei BCD ($\mu \approx 156,\dot{6}$), Fehler dabei ca. 8,- ($\mu = 148,\dot{6}$), was angesichts der großen Variabilität der Einzelpreise (130,- bis 166,-) akzeptabel erscheint.

  es gibt viele "normale", nur wenige "extreme" - von einem Minimodell abweichende Stichproben.

----

*) Man kann die Fragestellung etwa durch allgemeine Überlegungen zum Verbraucherpreisindex motivieren.

Im Unterschied zur Verteilung in der Grundgesamtheit bzw. in einer speziellen Stichprobe bezieht man sich nun nicht mehr auf eine Gesamtheit von Geschäften (in anderen Fällen: von Personen), sondern auf eine Gesamtheit von möglichen Stichproben. Eine Grundgesamtheit ist in der Statistik auch eine Menge von denkbaren Beobachtungen.

Die Verteilung der Variablen Stichprobenmittelwert kann nun in zweierlei Hinsicht gedeutet werden:

- statisch: Überblick über alle Möglichkeiten
- dynamisch: Anteile in dieser Verteilung können als Chance gedeutet werden. Zufällige Auswahl soll ja bewirken, daß jeder einzelne mögliche Wert gleichberechtigt ist: Ein Fehler von mehr als 4 S passiert bei nur 8 von 20 möglichen Stichproben - das Risiko, einen solchen Fehler zu machen bei der Schätzung $\mu \approx \bar{x}$, wird daher mit 8/20 bewertet.

Zufällige Auswahl ist mit der Idee des Minimodells dahingehend vereinbar, als es eben viele "normale" und nur wenige "extreme" Stichproben gibt. Das Risiko, eine extreme Stichprobe zu erhalten ist klein. Das Auswahlverfahren greift indirekt und dient gerade dazu, dieses Risiko berechenbar zu machen.

Wenn wir den Stichprobenumfang variieren, so zeigt sich überdies:

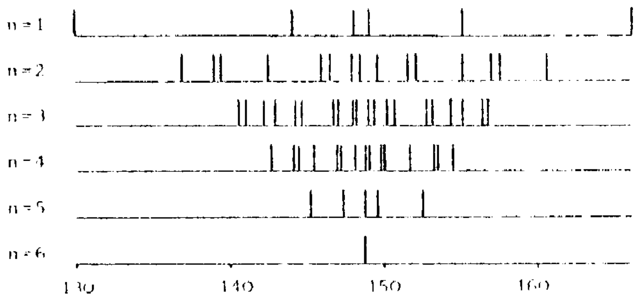


Bild 12: Verteilung der Variablen Stichprobenmittelwert bei variierendem Stichprobenumfang.

- Mit zunehmendem Stichprobenumfang zieht sich die Verteilung der Variablen Stichprobenmittelwert zusammen, die "Genauigkeit" der Schätzung $\mu \approx \bar{x}$ wird besser.

Auch bei zufälliger Auswahl kann man Information über die Grundgesamtheit positiv verwerten:

Wir nehmen an, daß wir wissen, daß B und C exklusive, A,D,E und F normale Geschäfte sind. Wir wählen aus den zwei "Schichten" von Geschäften getrennt aber zufällig aus, und zwar ein Geschäft aus {B,C}, zwei aus {A,D,E,F}. Damit wird etwa B C D als Stichprobenergebnis ausgeschlossen.

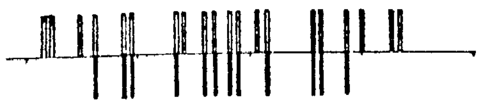


Bild 13: Verteilung der Variablen Stichprobenmittelwert mit und ohne Schichtung.

- Je homogener die Schichten, aus denen zufällig ausgewählt wird, umso besser. Die Abbildung zeigt, wie durch Einbringen von Information in Kombination mit zufälliger Auswahl die "Genauigkeit" der Schätzung verbessert wird.

Für das weitere Vorgehen ist das Studium der gedachten Verteilung aller Stichprobenmittelwerte notwendig:

Variable Stichprobenmittelwert $\overline{X}$		Grundgesamtheit X
$\mu_{\overline{X}}$	=	μ
$\sigma_{\overline{X}}$	=	$f(\sigma,n) \approx \frac{\sigma}{\sqrt{n}}$

Je eher X vom Typ ⏊,

desto eher $\overline{X}$ vom Typ ⏊.

Grundgesamtheit

Verteilung des Stichprobenmittelwerts

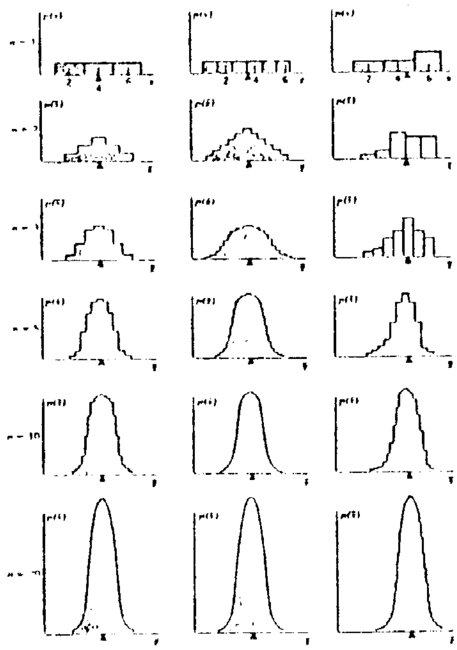


Bild 14:
Verteilung der Variablen Stichprobenmittelwert bei verschiedenen Verteilungen in der Grundgesamtheit.

Die Zusammenhänge können empirisch anhand einiger Beispiele, aber auch durch Simulation von Stichproben aus bestimmten Grundgesamtheiten erarbeitet werden.

Die angesprochenen Zusammenhänge zwischen Grundgesamtheit und Verteilung der Variablen $\overline{X}$ sind Kern der Wahrscheinlichkeitstheorie.*) Damit können wir die s-Regeln auch für die Verteilung von $\overline{X}$ anwenden:

----

*) Einschließlich des zentralen Grenzverteilungssatzes.

- $[\mu - 2\sigma_{\bar{X}}, \mu + 2\sigma_{\bar{X}}]$ ist dann ein Intervall, das ca. 95% der normalen Stichprobenmittelwerte enthält, ein sogenanntes 95%-Prognoseintervall für $\bar{X}$.
- $2\sigma_{\bar{X}}$ kann als Maß für den Fehler der Schätzung $\mu \approx \bar{x}$ gedeutet werden.
- $\mu - 2\sigma_{\bar{X}}$, $\mu + 2\sigma_{\bar{X}}$ zeigen an, welche Stichproben als extrem einzustufen sind.

Etwas suggestiver ist die Sprechweise: Die Wahrscheinlichkeit - das Risiko, daß der Wert von $\bar{X}$ aus dem Prognoseintervall $[\mu - 2\sigma_{\bar{X}}, \mu + 2\sigma_{\bar{X}}]$ *) herausfällt, wird mit 5% bewertet. Diese Bewertung ist natürlich darauf bezogen, daß zufällig ausgewählt wird und soll "prognosefähig" sein, d.h., wenn man tatsächlich "zufällig" unter denselben Bedingungen auswählt, soll in ca. 5% der Stichproben tatsächlich ein so extremer Stichprobenmittelwert realisiert werden, daß er aus diesem Prognoseintervall herausfällt.

Lernziele:

- Zusammenhänge zwischen Grundgesamtheit und Variablen Stichprobenmittelwert ($\bar{X}$).
- Leitziel: Menge von denkbaren Beobachtungen für $\bar{X}$ als neue (hypothetische) Grundgesamtheit - stellt den Vergleichsmaßstab für die Beurteilung eines einzelnen, isolierten Stichprobenmittelwerts $\bar{x}$ dar.
- Metaziele:
  - Bedeutung zufälliger Auswahl für die Deutung der Verteilung von $\bar{X}$.
  - Zusammenhänge zwischen der Idee "zufällige Auswahl" und der Zielvorstellung "Minimodell".

9. Statistische Beurteilung

Das Sporthemdenbeispiel ist insoferne von tatsächlichen Entscheidungssituationen unter Unsicherheit verschieden, als die Grundgesamtheit fiktiv bekannt ist, es dient nur als Aufhänger, die Zusammenhänge zwischen der Verteilung in der Grundgesamtheit und der Variablen Stichprobenmittelwert zu untersuchen.

---

*) Verwendet man die 3 s-Regel, so hat man das entsprechende Risiko mit ca. 1% zu bewerten.

Beispiel: Herr Mahlfreund möchte überprüfen, ob die Abfüllmaschine für die 1kg-Mehlpackungen richtig eingestellt ist. Gewisse Abweichungen, vom Sollgewicht sind einfach nicht zu vermeiden. Er muß jedoch darauf achten, daß er die untere Grenze von 985 g nur in sehr wenigen Fällen unterschreitet. Deshalb ist es notwendig, den Abfüllvorgang von Zeit zu Zeit zu überprüfen. In einem solchen Probelauf (Umfang n=30) ergibt sich: $\bar{x}$ = 995,49 g, s = 4,73 g.

Soll er die Maschine neu justieren oder stellt das Stichprobenergebnis keinen Anlaß dar, in den Abfüllvorgang einzugreifen und die Maschine neu zu justieren?

Das Beispiel illustriert die grundsätzliche Problematik der statistischen Beurteilung:

Man soll aufgrund der verfügbaren Information aus einer Stichprobe

- Entscheidungen fällen,
- gewisse Hypothesen über die Grundgesamtheit prüfen: die Maschine ist gut justiert, d.h. die gefüllten Mehlpackungen haben einen Mittelwert von μ = 1000 g.

Es ist klar, daß dabei Fehler möglich sind. Stoppt man den Abfüllvorgang, so kann dies zu Unrecht erfolgen und es entgeht einem "teure" Maschinenzeit. Andererseits kann es passieren, daß der Abfüllvorgang nicht in Ordnung ist und daß dies nicht erkannt wird. Dann bekommt Herr Mahlfreund eventuell Schwierigkeiten mit den Kontrollbehörden bzw. mit den Abnehmern.

Folgendes Entscheidungskriterium ist naheliegend:

$\bar{x} \notin (\mu_{\bar{X}} - 2\sigma_{\bar{X}}, \ \mu_{\bar{X}} + 2\sigma_{\bar{X}}) \rightarrow$ Hypothese μ = 1000 g verwerfen $\rightarrow$ neu justieren.

Von den Möglichkeiten:

(a) die Abweichungen sind auf "Fehler" bei der zufälligen Auswahl zurückzuführen,

(b) es ist etwas passiert, was nur geringe Wahrscheinlichkeit (nämlich 5%) hat,

(c) die Hypothese über den Mittelwert μ trifft nicht zu,

wird nur die letzte in Betracht gezogen.

$\bar{x} \in (\mu_{\bar{X}} - 2\sigma_{\bar{X}}, \ \mu_{\bar{X}} + 2\sigma_{\bar{X}}) \rightarrow$ kein Einwand gegen Hypothese über den Mittelwert μ $\rightarrow$ Abfüllvorgang geht weiter.

Man sagt: Das Risiko dieses Entscheidungsverfahrens - Test - wird mit 5% bewertet.*)

----

*) Für die 3s-Regel: Risiko = 1%.

Für die Auswertung des Entscheidungskriteriums braucht man noch die Werte $\mu_{\bar{X}}$ und $\sigma_{\bar{X}}$.

Die Hypothese $\mu = 1000$ g liefert $\mu_{\bar{X}} = \mu = 1000$ g,
für $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$ *) schätzt man $\sigma \approx s$, $s = 4{,}73$ g, die Standardabweichung aus der Stichprobe.

Der Nachteil der bisherigen Vorgangsweise liegt darin, daß man eine einzelne Hypothese über μ isoliert "testet".
Die Frage: "Welche Hypothesen über den Mittelwert können bei einem solchen "Test" nicht ausgegliedert werden" führt auf den Begriff der sogenannten Vertrauensintervalle.

Die Stichprobe hat $\bar{x}$ ergeben. Eine Hypothese μ kann nicht abgelehnt werden, wenn $\bar{x} \in (\mu - 2\sigma_{\bar{X}}, \mu + 2\sigma_{\bar{X}})$, d.h., wenn der realisierte Stichprobenmittelwert im Vergleich zur gedachten Verteilung von $\bar{X}$ unter der Annahme μ als nicht "extrem" ausgewiesen wird. Es gilt: $\bar{x} \in (\mu - 2\sigma_{\bar{X}}, \mu + 2\sigma_{\bar{X}})$ genau dann, wenn $\mu \in (\bar{x} - 2\sigma_{\bar{X}}, \bar{x} + 2\sigma_{\bar{X}})$; man sagt: $(\bar{x} - 2\sigma_{\bar{X}}, \bar{x} + 2\sigma_{\bar{X}})$ ist eine Realisierung eines Vertrauensintervalls zur "Sicherheit" 95%.**) Der Begriff Vertrauensintervall präzisiert gleichzeitig, was unter Genauigkeit der Schätzung $\mu \approx \bar{x}$ zu verstehen ist.

Von den Einschränkungen der Risiko-Werte mit 5% (1%) bzw. der Sicherheits-Werte 95% (99%) kann man sich durch die Einführung der Normalverteilung befreien. Über Deutung und mögliche Fehldeutungen der "statistischen Standardaussagen" könnte man noch viel anfügen. Es sei hier nur daran erinnert, daß die Einschätzungen bezüglich Risiko und Sicherheit entscheidend von der Voraussetzung der zufälligen Auswahl abhängen.

Lernziele:

- Grundprobleme statistischer Beurteilung.
- Standardmethoden: Test und Vertrauensintervalle für den Mittelwert.
- Metaziele:
  - Deutung von grundlegenden statistischen Aussagen wie "das Risiko eines Tests wird mit 5% bewertet", "die Sicherheit der herangezogenen Vertrauensintervallmethode wird mit 95% bewertet".
  - Einfache Fehldeutungen von Grundaussagen der statistischen Beurteilung.
  - Rolle der zufälligen Auswahl für die Bewertung von Risiko und Sicherheit.

----

*) Auf die Unterscheidung von Stichproben aus endlichen Grundgesamtheiten und Wiederholungen desselben Experiments wollen wir hier nicht eingehen.

**) Der 3s-Regel entspricht ein 99%-Vertrauensintervall.

10. Zur Rolle von Wahrscheinlichkeiten und innermathematischen Zusammenhängen

Wahrscheinlichkeiten haben nur die Funktion der Bewertung von Risiken, sie treten - im Gegensatz zu üblichen Darstellungen der Stochastik - als Hilfsgrößen auf. Inhaltlich verstehen wir Wahrscheinlichkeit in zweierlei Hinsicht:

- Idealisierte relative Häufigkeit, aus der Prognosewerte für tatsächlich auftretende, relative Häufigkeiten abgeleitet werden sollen.
- Ausdruck der subjektiven Bewertung einer Situation. Die Forderung, daß auch daraus Prognosen für relative Häufigkeiten ableitbar sein sollen, stellt für Testsituationen allerdings eine Fiktion dar.

Im Zentrum der Wahrscheinlichkeitstheorie sind die Zusammenhänge zwischen der Verteilung in der Grundgesamtheit (X) und der Variablen Stichprobenmittelwert ($\overline{X}$) anzusiedeln. Im Falle der Herleitung der Verteilung von $\overline{X}$ müßte man den Begriffsapparat mehrdimensionaler Wahrscheinlichkeitsdichten, Randdichten sowie Integraltransformationen aufbauen. Wir plädieren dafür, diese innermathematisch anspruchsvollen Beziehungen empirisch vorwegzunehmen, und das schon im Hinblick auf die für die Praxis bedeutsame Normalapproximation der Variablen $\overline{X}$ (implizit unsere Verteilung vom Typ , zentraler Grenzverteilungssatz).

Wenn man diese innermathematischen Beziehungen nicht aufgreift, so besteht keine Notwendigkeit, den Kalkül der Wahrscheinlichkeitsrechnung explizit aufzubauen. Den dadurch gewonnenen Freiraum möchten wir dadurch nützen, daß wir direkt zur statistischen Beurteilung vordringen, und ferner, daß den Fragen "Welche Konzepte stehen im Hintergrund?", "Was bedeuten statistische Aussagen?" mehr Raum gewidmet werden kann.

Literatur

BOROVCNIK, Manfred, 1981: Statistik im Selbststudium - Von der beschreibenden zur beurteilenden Statistik, Klagenfurt.

BOROVCNIK, Manfred, 1982: Internationale Tendenzen in der Statistik-Ausbildung; in: Mitteilungsblatt der Österreichischen Gesellschaft für Statistik und Informatik 48, 1982, S. 193-207 (Wien).

SCHRAGE, Georg, 1980: Schwierigkeiten mit stochastischer Modellbildung - Zwei Beispiele aus der Praxis; in: Journal für Mathematik-Didaktik 1, 1980, S. 86-101 (Schöningh/Paderborn).

SCHRAGE, Georg, 1981: Entscheiden und Begründen - Leitlinien für den Statistikunterricht; in: Schriftenreihe Didaktik der Mathematik 3, 1981, S. 179-208 (Hölder-Pichler-Tempsky / Wien).

WINTER, Heinrich, 1982: Why teach descriptive statistics in secondary schools (7th to 10th grade)?; in: Grey/Holmes/Barnett/Constable (eds.): Proceedings of the First International Conference on Teaching Statistics, Sheffield (Teaching Statistics Trust).