

Probenahme

Statistische Grundlagen der Probenahme
bei einfacher Aufteilung**DIN**
53 803
Teil 1Sampling; statistical basis; one-way layout
Échantillonnage; base statistique; classification simple

Ersatz für Ausgabe 03.79

Inhalt

	Seite		Seite
1 Anwendungsbereich und Zweck	2	7 Rundungsempfehlungen	13
2 Begriffe	2	7.1 Stichprobenaufbau bei vorgegebener Weite des Vertrauensbereiches für den Mittelwert der Grundgesamtheit	13
3 Anordnungsschema für die Beobachtungswerte bei einfacher Aufteilung	3	7.2 Stichprobenaufbau bei vorgegebener Weite des Vertrauensbereiches für den Mittelwert der Grundgesamtheit und minimalen Gesamtkosten	13
4 Statistische Auswertung bei einfacher Aufteilung (inhomogene Grundgesamtheit)	3	7.3 Stichprobenaufbau bei vorgegebenen Gesamtkosten	13
4.1 Allgemeines	3	7.4 Stichprobenaufbau bei vorgegebenen Gesamtkosten und kleinster erreichbarer Weite des Vertrauensbereiches für den Mittelwert der Grundgesamtheit	13
4.2 Modellvorstellungen	4	8 Programmablaufpläne für die Berechnung des Stichprobenaufbaues	13
4.2.1 Modell mit Zufallskomponenten	4	Anhang A Beispiele aus der Textiltechnik	19
4.2.2 Modell mit systematischen Komponenten	4	A.1 Aufteilung nach einem Kriterium; Modell mit Zufallskomponenten; Durchführung der Einfachen Varianzanalyse; Bestimmung der Schätzwerte für die Varianzen; Testen von Hypothesen über die Modellparameter; Vertrauensbereich für den Mittelwert der Grundgesamtheit; Höchstzugkraft von Spinnfasergarn	19
4.3 Auswertung	5	A.2 Modell mit Zufallskomponenten; Stichprobenaufbau bei vorgegebener Weite des Vertrauensbereiches für den Mittelwert der Grundgesamtheit; varianzbestimmte Stichprobe; Höchstzugkraft von Spinnfasergarn	20
4.4 Weitere Einzelheiten zum Modell mit Zufallskomponenten	5	A.3 Modell mit Zufallskomponenten; Stichprobenaufbau bei vorgegebener Weite des Vertrauensbereiches für den Mittelwert der Grundgesamtheit und minimalen Gesamtkosten; varianzbestimmte kostenminimale Stichprobe; Höchstzugkraft von Spinnfasergarn	20
4.4.1 Schätzwerte für die Modellparameter	5	A.4 Aufteilung nach einem Kriterium; Modell mit Zufallskomponenten; Durchführung der Einfachen Varianzanalyse; Bestimmung der Schätzwerte für die Varianzen; Testen von Hypothesen über die Modellparameter; Vertrauensbereich für den Mittelwert der Grundgesamtheit; Spinn Drehung von Spinnfasergarn	21
4.4.2 Testen von Hypothesen über die Modellparameter	6	A.5 Modell mit Zufallskomponenten; Stichprobenaufbau bei vorgegebener Weite des Vertrauensbereiches für den Mittelwert der Grundgesamtheit bei homogener Grundgesamtheit; varianzbestimmte Stichprobe; Spinn Drehung von Spinnfasergarn	22
4.4.3 Vertrauensbereiche für die Modellparameter ...	6	A.6 Aufteilung nach einem Kriterium; Modell mit systematischen Komponenten; Durchführung der Einfachen Varianzanalyse; Bestimmung der Schätzwerte für die Gruppenmittelwerte und die Varianz; Testen von Hypothesen über die Modellparameter; Vertrauensbereich für die Mittelwerte der Gruppen; Kett schrumpfung bei einer Gewebelieferung	23
4.5 Weitere Einzelheiten zum Modell mit systematischen Komponenten	6		
4.5.1 Schätzwerte für die Modellparameter	6		
4.5.2 Testen von Hypothesen über die Modellparameter	7		
4.5.3 Vertrauensbereiche für die Modellparameter ..	7		
4.6 Sonderfall: Homogene Grundgesamtheit	7		
5 Stichprobenaufbau bei inhomogener Grundgesamtheit	7		
5.1 Modell mit Zufallskomponenten	7		
5.1.1 Stichprobenaufbau bei vorgegebener Weite des Vertrauensbereiches für den Mittelwert der Grundgesamtheit; varianzbestimmte Stichprobe	8		
5.1.2 Stichprobenaufbau bei vorgegebener Weite des Vertrauensbereiches für den Mittelwert der Grundgesamtheit und minimalen Gesamtkosten; varianzbestimmte kostenminimale Stichprobe	9		
5.1.3 Stichprobenaufbau bei vorgegebenen Gesamtkosten; kostenbestimmte Stichprobe	9		
5.1.4 Stichprobenaufbau bei vorgegebenen Gesamtkosten und kleinster erreichbarer Weite des Vertrauensbereiches für den Mittelwert der Grundgesamtheit; kostenbestimmte varianzminimale Stichprobe	11		
5.2 Modell mit systematischen Komponenten	11		
5.2.1 Aussagen über den Vertrauensbereich für den Mittelwert der Grundgesamtheit oder für die Mittelwerte der Gruppen	12		
5.2.1.1 Stichprobenaufbau bei vorgegebener Weite des Vertrauensbereiches für den Mittelwert der Grundgesamtheit oder für die Mittelwerte der Gruppen; varianzbestimmte Stichprobe	12		
5.2.1.2 Stichprobenaufbau bei vorgegebenen Gesamtkosten; kostenbestimmte Stichprobe	12		
6 Stichprobenaufbau bei homogener Grundgesamtheit	12		

Fortsetzung Seite 2 bis 30

Ausschuß Qualitätssicherung und angewandte Statistik (AQS) im DIN Deutsches Institut für Normung e.V.
Normenausschuß Materialprüfung (NMP) im DIN

Documents
Document Sharing Hub

	Seite
A.7 Aufteilung nach einem Kriterium; Modell mit systematischen Komponenten; Durchführung der Einfachen Varianzanalyse; Bestimmung der Schätzwerte für die Gruppenmittelwerte und die Varianz; Testen von Hypothesen über die Modellparameter; Vertrauensbereich für die Mittelwerte der Gruppen; Trocknungsverlust einer Kammgarnlieferung	24
A.8 Modell mit Zufallskomponenten; Stichprobenaufbau bei vorgegebener Weite des Vertrauensbereiches für den Mittelwert der Grundgesamtheit; varianzbestimmte Stichprobe; Höchstzugkraft von Spinnfasergarn	25
A.9 Modell mit Zufallskomponenten; Stichprobenaufbau bei vorgegebener Weite des Vertrauensbereiches für den Mittelwert der Grundgesamtheit; varianzbestimmte kostenminimale Stichprobe; Gehalt an gewaschener Wolle	25
A.10 Modell mit systematischen Komponenten; Stichprobenaufbau bei vorgegebener Weite des Ver-	

trauensbereiches für die Gruppenmittelwerte; varianzbestimmte Stichprobe; Kettshrumpfung einer Gewebelieferung	26
A.11 Bestimmung der Varianz in der Grundgesamtheit; Stichprobenaufbau bei vorgegebener Weite des Vertrauensbereiches für den Mittelwert der Grundgesamtheit; varianzbestimmte Stichprobe; Einzelfaser-Naßhöchstzugkraft	26
A.12 Stichprobenaufbau bei vorgegebener Weite des Vertrauensbereiches für den Mittelwert der Grundgesamtheit; varianzbestimmte Stichprobe; Einzelfaser-Naßhöchstzugkraft	26
Anhang B Formelzeichen	27
Zitierte Normen und andere Unterlagen	29
Weitere Normen und andere Unterlagen	29
Änderungen	30
Erläuterungen	30

1 Anwendungsbereich und Zweck

Bei den meisten Untersuchungen werden aus der vorliegenden Grundgesamtheit (dem Prüfling) Proben entnommen. Die daran für ein bestimmtes Merkmal ermittelten Kenngrößen müssen den Schluß auf die entsprechenden Kenngrößen in der Grundgesamtheit (dem Prüfling) zulassen. Die Probe muß deshalb so entnommen werden, daß die Differenz zwischen der an der Probe gefundenen Kenngröße und der Kenngröße der Grundgesamtheit zufällig ist, d. h., die Probe muß für die Grundgesamtheit repräsentativ sein (repräsentative Probe).

Weiterhin müssen die zufälligen Abweichungen quantitativ abgeschätzt und Vertrauensbereiche für die unbekannten Kenngrößen der Grundgesamtheit bei vorgegebenem Vertrauensniveau angegeben werden können.

Schließlich muß damit gerechnet werden, daß möglicherweise mehrere Streuungsursachen vorhanden sind. Die Probenahme hat so zu erfolgen, daß die verschiedenen Streuungsursachen berücksichtigt und ihre Auswirkungen getrennt erfaßt werden können.

Für alle Untersuchungen, bei denen von der Stichprobe auf die Grundgesamtheit geschlossen werden soll, müssen die Proben nach dieser Norm entnommen werden. Falls Proben untersucht werden, die nicht nach dieser Norm entnommen wurden, gelten die Ergebnisse nur für die Probe selbst, nicht für die Grundgesamtheit.

Diese Norm bezieht sich auf alle Bereiche der Technik, Wissenschaft, Medizin, Wirtschaft usw.

In DIN 53 803 Teil 3 werden die statistischen Grundlagen der Probenahme bei **zweifacher** Aufteilung nach zwei **gleichberechtigten** Kriterien und in DIN 53 803 Teil 4 die statistischen Grundlagen der Probenahme bei **zweifacher** Aufteilung nach zwei einander **nachgeordneten** Kriterien behandelt.

In dieser Norm werden die statistischen Grundlagen der Probenahme bei **einfacher** Aufteilung behandelt, d. h. beim Vorliegen von zwei Streuungsursachen.

Die folgenden beiden Möglichkeiten sind zu unterscheiden:

- Die Existenz zweier Streuungsursachen ist zwar bekannt, über das Ausmaß ihrer Auswirkungen liegen jedoch keine Angaben vor. Da in diesem Falle die Einfache Varianzanalyse durchzuführen ist, muß bereits

die Probenahme darauf abgestellt werden (siehe Abschnitte 2 bis 4).

- Liegen dagegen quantitative Erfahrungswerte für die beiden Streuungen vor, dann können damit konkrete Vorschriften für die Probenahme gemacht werden, um zu einem Ergebnis mit vorgegebener Weite des Vertrauensbereiches zu gelangen (siehe Abschnitte 5 bis 8).

Wie sich praktisch gezeigt hat, ist jede für ein bestimmtes Merkmal repräsentative Stichprobe nicht ohne weiteres auch für ein anderes Merkmal repräsentativ.

2 Begriffe

Die in dieser Norm verwendeten Begriffe sind in DIN 55 350 Teil 11, Teil 12, Teil 14, Teil 21, Teil 22, Teil 23, Teil 24 und Teil 31 festgelegt.

Bei der Gruppeneinteilung einer Grundgesamtheit nach **einem** Kriterium ist die **Einfache Varianzanalyse** ein statistisches Auswertungsverfahren, das auf der Aufteilung der gesamten Summe der quadrierten Abweichungen (S.d.q.A.) in zwei Anteile beruht:

- die S.d.q.A. innerhalb der Gruppen und
- die S.d.q.A. zwischen den Gruppen.

Sie dient unter anderem zur Entscheidung der Frage, ob die Grundgesamtheit bezüglich eines bestimmten Merkmals bei einer bestimmten Gruppeneinteilung als homogen anzusehen ist, und zur Abschätzung der Anteile, mit denen die beiden Streuungsursachen zur Gesamtstreuung beitragen.

Zur Berechnung des Vertrauensbereiches für eine Kenngröße der Grundgesamtheit muß das Vertrauensniveau $1-\alpha$ vorgegeben werden. Zum Testen von Hypothesen muß das Signifikanzniveau α vorgegeben werden. Das Vertrauensniveau und das Signifikanzniveau sind nach technischen und wirtschaftlichen Kriterien **vor Beginn der Untersuchung** festzulegen.

Ausdrücklich sei darauf hingewiesen, daß das Signifikanzniveau für den Test und das Vertrauensniveau für die Abgrenzung der Vertrauensbereiche unabhängig voneinander nach den Folgen von Fehlentscheidungen festgelegt werden können.

Der Begriff „Stichprobenumfang“ ist nach DIN 55 350 Teil 14 definiert als „Anzahl der Auswahleinheiten in der Stichprobe“. In dieser Norm wird dieser Begriff auf die „Anzahl der Beobachtungswerte einer Stichprobe“ übertragen.

Tabelle 1. Anordnungsschema für die Beobachtungswerte; Zwischensummen und Hilfsgrößen bei einfacher Aufteilung

Beobach- tungs- wert Gruppe	1	2	bis	j	bis	n_2	Zwischensummen		
							A_i	B_i	A_i^2
1	x_{11}	x_{12}	bis	x_{1j}	bis	x_{1n_2}	A_1	B_1	A_1^2
2	x_{21}	x_{22}	bis	x_{2j}	bis	x_{2n_2}	A_2	B_2	A_2^2
bis	bis	bis		bis		bis	bis	bis	bis
i	x_{i1}	x_{i2}	bis	x_{ij}	bis	x_{in_2}	A_i	B_i	A_i^2
bis	bis	bis		bis		bis	bis	bis	bis
n_1	$x_{n_1 1}$	$x_{n_1 2}$	bis	$x_{n_1 j}$	bis	$x_{n_1 n_2}$	A_{n_1}	B_{n_1}	$A_{n_1}^2$
Hilfsgrößen							$A = \sum A_i$	$B = \sum B_i$	$C = \sum A_i^2$

3 Anordnungsschema für die Beobachtungswerte bei einfacher Aufteilung

Die Beobachtungswerte werden nach dem Schema in Tabelle 1 (stark umrandeter linker Teil) zusammengestellt. Das Schema enthält zu jeder der n_1 Gruppen ($i = 1, 2, 3, \dots, n_1$) die zugehörigen n_2 Beobachtungswerte ($j = 1, 2, 3, \dots, n_2$). Dabei wird in dieser Norm die Anzahl der Beobachtungswerte für jede Gruppe als gleich groß vorausgesetzt. Bei unterschiedlichen Anzahlen von Beobachtungswerten in den Gruppen siehe [1].

Man bildet weiter für jede Zeile i die folgenden Zwischensummen:

Summe aller Beobachtungswerte innerhalb der Zeile i :

$$A_i = \sum_{j=1}^{n_2} x_{ij} \quad (1)$$

Summe aller quadrierten Beobachtungswerte innerhalb der Zeile i :

$$B_i = \sum_{j=1}^{n_2} x_{ij}^2 \quad (2)$$

Quadrat der Summe aller Beobachtungswerte innerhalb der Zeile i :

$$A_i^2 = \left(\sum_{j=1}^{n_2} x_{ij} \right)^2 \quad (3)$$

Diese für jede Zeile i vorliegenden drei Zwischensummen werden jeweils über alle n_1 Zeilen zu den folgenden Hilfsgrößen (Spaltensummen der Zwischensummen) summiert:

Summe aller Beobachtungswerte:

$$A = \sum_{i=1}^{n_1} A_i = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} x_{ij} \quad (4)$$

Summe aller quadrierten Beobachtungswerte:

$$B = \sum_{i=1}^{n_1} B_i = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} x_{ij}^2 \quad (5)$$

Summe der Quadrate der Summe aller Beobachtungswerte innerhalb der Zeilen i :

$$C = \sum_{i=1}^{n_1} A_i^2 = \sum_{i=1}^{n_1} \left(\sum_{j=1}^{n_2} x_{ij} \right)^2 \quad (6)$$

Die Zwischensummen und Hilfsgrößen dürfen nicht gerundet werden.

Bis hierher ist der Gang der Rechnung unabhängig davon, ob später das Modell mit Zufallskomponenten oder das Modell mit systematischen Komponenten benutzt wird.

4 Statistische Auswertung bei einfacher Aufteilung (inhomogene Grundgesamtheit)

4.1 Allgemeines

Bei der Untersuchung muß man das Prüflös aus sachlichen Gründen häufig in Gruppen unterteilen. Dabei ist insbesondere das zu untersuchende Merkmal zu berücksichtigen, siehe Tabelle 2.

Tabelle 2. Beispiele für die Unterteilung des Prüflöses

Grund- gesamtheit Prüflös	Merkmal	Gruppe	Einheiten (innerhalb einer Gruppe)
mehrere Ballen Schweiß- wolle	Gehalt an rein ge- waschener Wolle	Ballen	Teile eines Ballens (z. B. entnehmbare Kernbohr- proben)
ein Ballen Chemie- fasern	Feuchtig- keitsgehalt	Schicht oder Schale	aus einer Schicht oder Schale entnehmbare Teile bestimmter gleicher Größe
Garn- lieferung	Höchstzug- kraft	Kops	Teile eines Kops (z. B. entnehm- bare Abschnitte von 50 cm Länge)
Gewebe- lieferung	Flächen- gewicht	Gewebe- stück	aus einem Gewebestück entnehmbare Teilabschnitte bestimmter gleicher Größe

Die Gruppe ist ein Teil der Grundgesamtheit, der sich bei einer sachlich begründeten Unterteilung der Grundgesamtheit ergibt. Die Definitionen für die Kenngrößen der Grundgesamtheit gelten in entsprechender Form auch für die Kenngrößen der Gruppen. Es gibt also Kenngrößen für die

einzelne Gruppe i als Teilgesamtheit $(\mu_i, \sigma_i^2, \sigma_i, y_i)$ und Kenngrößen der aus der Gruppe entnommenen Stichprobe $(\bar{x}_i, s_i^2, s_i, v_i)$.

Homogen (in statistischem Sinne) bezüglich eines bestimmten Merkmals bei einer bestimmten Gruppeneinteilung ist eine Grundgesamtheit dann, wenn für dieses Merkmal die Kenngrößen $\mu_i, \sigma_i^2, \sigma_i, y_i$ aller Gruppen untereinander, also auch mit denen der Grundgesamtheit, übereinstimmen, bzw. die Kenngrößen $\bar{x}_i, s_i^2, s_i, v_i$ nur zufällig voneinander abweichen.

Wenn nichts anderes bekannt ist, muß man davon ausgehen, daß das Prüflös inhomogen ist, d. h., die Streuung der Beobachtungswerte um ihren Mittelwert – den Gesamtmittelwert – besteht aus zwei Komponenten: der Streuung innerhalb jeder Gruppe (Streuung der Beobachtungswerte um den jeweiligen Gruppenmittelwert) und der Streuung zwischen den Gruppen (Streuung der Gruppenmittelwerte um den Gesamtmittelwert).

Dann hat die Probenahme so zu erfolgen, daß folgende Aussagen gemacht werden können:

- 1a) Für den Gesamtmittelwert muß ein Schätzwert angegeben und ein Vertrauensbereich abgegrenzt werden können. Das gleiche gilt für die Gruppenmittelwerte beim Modell mit systematischen Komponenten.
- 1b) Beide Streuungen (zwischen den Gruppen und innerhalb der Gruppen) müssen getrennt und quantitativ erfaßt werden können.
- 1c) Insbesondere muß statistisch entschieden werden können, ob die Streuung des untersuchten Merkmals zwischen den Gruppen existiert, d. h. ob das Material bezüglich des untersuchten Merkmals bei der gewählten Unterteilung der Grundgesamtheit in Gruppen als inhomogen anzusehen ist.

Damit die genannten Forderungen erfüllt werden können, muß folgendes beachtet werden:

- 2a) Mehr als eine Gruppe ist zur Probenahme heranzuziehen; ebenso ist aus jeder untersuchten Gruppe mehr als eine Einzelprobe zu entnehmen. Die Anzahl der Einzelproben je Gruppe und, falls für die Gruppen getrennte Untersuchungen durchgeführt werden müssen, die Anzahl der Endproben je Gruppe, muß für alle untersuchten Gruppen gleich sein, damit die nachstehenden statistischen Auswertungsverfahren anwendbar sind.

Beim Fehlen jeglicher Vorkenntnisse wählt man die Anzahl der zur Probenahme heranzuziehenden Gruppen mindestens so groß wie die Anzahl der Endproben je Gruppe; fast immer ist es günstiger, die Anzahl der Gruppen erheblich größer zu wählen als die Anzahl der Endproben je Gruppe.

- 2b) Falls nicht alle Gruppen zur Probenahme herangezogen werden, müssen die Gruppen, denen Proben entnommen werden sollen, zufällig aus der Grundgesamtheit ausgewählt werden. Bei der Entnahme der Endproben aus den Gruppen sind die Richtlinien von DIN 53 803 Teil 2 zu beachten.
- 2c) Zu jeder Einzelprobe bzw. Endprobe ist die Gruppenzugehörigkeit festzuhalten.

4.2 Modellvorstellungen

Um Aussagen über die Grundgesamtheit machen zu können, müssen einige Modellannahmen über die Grundgesamtheit gemacht werden. Entsprechend den praktischen Gegebenheiten sind zwei Modelle zu unterscheiden,

- das Modell mit Zufallskomponenten und
- das Modell mit systematischen Komponenten.

Bei beiden Modellen wird unterstellt, daß sich jeder Merk-

malswert a_{ij} der Grundgesamtheit aus mehreren Komponenten gemäß folgender Gleichung zusammensetzt:

$$a_{ij} = \mu + (\mu_i - \mu) + \varepsilon_{ij} = \mu + \xi_i + \varepsilon_{ij} \quad (7)$$

Dabei ist

- μ_i der Mittelwert der Gruppe i ($i = 1, 2, 3, \dots, M$)
- $\xi_i = \mu_i - \mu$ die Abweichung des Mittelwertes der Gruppe i vom Mittelwert μ der Grundgesamtheit
- ε_{ij} eine Zufallsabweichung innerhalb der Gruppe i .

In den folgenden Ausführungen wird vorausgesetzt, daß die ε_{ij} unabhängig voneinander und normalverteilt sind; sie haben für jede Gruppe den Mittelwert Null und die Varianz σ_ε^2 ; σ_ε^2 wird Varianz innerhalb der Gruppen genannt.

Diese Voraussetzungen, wonach die Beobachtungswerte innerhalb einer Gruppe normalverteilt sind und die Varianz σ_ε^2 für alle Gruppen gleich ist, sind gegebenenfalls mit Hilfe geeigneter statistischer Testverfahren nachzuprüfen (siehe [1], z. B. Cochran-Test).

Die beiden Modelle unterscheiden sich hinsichtlich der Art der Abweichungen $\xi_i = \mu_i - \mu$ der Gruppenmittelwerte vom Mittelwert der Grundgesamtheit.

4.2.1 Modell mit Zufallskomponenten

Kann oder soll die einzelne der insgesamt vorhandenen M Gruppen nicht als Individuum aufgefaßt werden und will man Aussagen über den Mittelwert μ der Grundgesamtheit und die Streuung der Gruppenmittelwerte μ_i um den Mittelwert μ der Grundgesamtheit machen, dann ist das Modell mit Zufallskomponenten anzuwenden. Bei diesem Modell ist der einzelne Gruppenmittelwert ohne Bedeutung. Man faßt die Abweichungen $\xi_i = \mu_i - \mu$ als Zufallsabweichungen mit dem Mittelwert Null und der Varianz σ_ξ^2 auf; σ_ξ^2 wird Varianz zwischen den Gruppen genannt.

Von den M Gruppen werden n_1 Gruppen zufällig ausgewählt und zur Probenahme herangezogen. (Im Grenzfall kann $n_1 = M$ werden.)

Das Ziel der Auswertung beim Modell mit Zufallskomponenten ist also, Aussagen über den Mittelwert μ der Grundgesamtheit und die beiden Varianzen σ_ξ^2 und σ_ε^2 zu machen, aus denen sich die Varianz σ^2 der Beobachtungswerte additiv zusammensetzt:

$$\sigma^2 = \sigma_\xi^2 + \sigma_\varepsilon^2 \quad (8)$$

Aus diesem Grunde werden die Varianzen σ_ξ^2 und σ_ε^2 häufig Varianzkomponenten, die Standardabweichungen σ_ξ und σ_ε Streuungskomponenten zwischen den Gruppen und innerhalb der Gruppen genannt.

Bei einfacher Gruppenunterteilung müssen damit neben μ noch zwei Varianzen (σ_ξ^2 und σ_ε^2) berücksichtigt werden.

Das Modell mit Zufallskomponenten wird bei Abnahmeprüfungen sehr häufig angewendet; dabei kommt es nur auf den Mittelwert μ der Grundgesamtheit und die beiden Varianzen σ_ξ^2 und σ_ε^2 an, die Gruppen werden in der Regel nicht als Individuen behandelt.

4.2.2 Modell mit systematischen Komponenten

Kann oder muß jede der M vorhandenen Gruppen als ein Individuum aufgefaßt werden, über das eine Aussage sinnvoll ist, dann ist das Modell mit systematischen Komponenten zugrunde zu legen.

Bei diesem Modell ist jeder der einzelnen Gruppenmittelwerte von Bedeutung, und man faßt die Abweichungen $\xi_i = \mu_i - \mu$ der Gruppenmittelwerte vom Mittelwert

$$\mu = \frac{1}{M} \sum_{i=1}^M \mu_i \quad (9)$$

der Grundgesamtheit als systematische Abweichungen auf. Damit jeder Gruppenmittelwert μ_i einzeln erfaßt werden

kann, müssen alle M Gruppen zur Probenahme herangezogen werden ($n_1 = M$).

Insgesamt ist das Ziel der Auswertung, Aussagen über den Mittelwert μ der Grundgesamtheit und/oder die M Gruppenmittelwerte μ_i , in jedem Falle aber über die Varianz σ_ε^2 innerhalb der Gruppen zu machen.

Das Modell mit systematischen Komponenten wird bei innerbetrieblichen Untersuchungen sehr häufig angewendet; dabei kommt es z. B. auf die Untersuchung von Arbeitspositionen, Maschinen, Laboranten, Meßgeräten, Meßverfahren an, die Gruppen werden also als Individuen behandelt.

4.3 Auswertung

Jede Einzelprobe bzw. Endprobe liefert für das untersuchte Merkmal einen Beobachtungswert; zu jedem dieser Beobachtungswerte muß die Gruppenzugehörigkeit festgehalten werden.

Bezeichnet man die Anzahl der zur Probenahme herangezogenen Gruppen mit n_1 und die Anzahl der auf jede dieser Gruppen entfallenden Beobachtungswerte mit n_2 , dann fallen insgesamt $n = n_1 \cdot n_2$ Beobachtungswerte x_{ij} an. Dabei ist i die Nummer der Gruppe und j die Nummer des Beobachtungswertes in dieser Gruppe.

Aus den entsprechend dem Schema in Tabelle 1 aufgeschriebenen Beobachtungswerten werden die Mittelwerte $x_{i.}$ und die Varianzen s_i^2 der n_1 Gruppen berechnet:

$$x_{i.} = \frac{1}{n_2} \sum_{j=1}^{n_2} x_{ij} \quad (10)$$

($x_{i.}$ wird auch als $\bar{x}_i$ bezeichnet).

$$s_i^2 = \frac{1}{n_2 - 1} \sum_{j=1}^{n_2} (x_{ij} - x_{i.})^2 \quad (11)$$

Der Gesamtmittelwert der $n = n_1 \cdot n_2$ Beobachtungswerte ist:

$$x_{..} = \frac{1}{n_1 \cdot n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} x_{ij} = \frac{1}{n_1} \sum_{i=1}^{n_1} x_{i.} \quad (12)$$

($x_{..}$ wird auch als $\bar{x}$ bezeichnet.)

Zur Durchführung der Varianzanalyse berechnet man die S.d.q.A. innerhalb der Gruppen

$$Q_w = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} (x_{ij} - x_{i.})^2 = (n_2 - 1) \sum_{i=1}^{n_1} s_i^2 \quad (13)$$

und die zugehörige Zahl der Freiheitsgrade

$$f_w = n_1 (n_2 - 1) \quad (14)$$

sowie die S.d.q.A. zwischen den Gruppen

$$Q_b = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} (x_{i.} - x_{..})^2 = n_2 \sum_{i=1}^{n_1} (x_{i.} - x_{..})^2 \quad (15)$$

und die zugehörige Zahl der Freiheitsgrade

$$f_b = n_1 - 1. \quad (16)$$

Die Summe dieser beiden S.d.q.A. ergibt die gesamte S.d.q.A.

$$Q_{\text{total}} = Q_w + Q_b = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} (x_{ij} - x_{..})^2 \quad (17)$$

$$\text{mit } f_{\text{total}} = f_w + f_b = n_1 n_2 - 1 \quad (18)$$

Freiheitsgraden.

Die Indizes w und b leiten sich von den englischen Benennungen „within“ für „innerhalb“ und „between“ für „zwischen“ ab.

Rechentchnisch ist es ungünstig, die S.d.q.A. aus den Definitionsgleichungen (13) und (15) zu gewinnen. Zweckmäßiger ist folgender Weg:

Die gesuchten Summen der quadrierten Abweichungen werden aus den in Tabelle 1 enthaltenen Hilfsgrößen A, B und C ermittelt. Dabei ist:

$$Q_b = \frac{C}{n_2} - \frac{A^2}{n_1 \cdot n_2} \quad (19)$$

$$Q_w = B - \frac{C}{n_2} \quad (20)$$

$$Q_{\text{total}} = B - \frac{A^2}{n_1 \cdot n_2} \quad (21)$$

Weiter ist der Gesamtmittelwert der $n_1 \cdot n_2$ Beobachtungswerte:

$$x_{..} = \frac{A}{n_1 \cdot n_2} \quad (22)$$

Zusammen mit der jeweils zugehörigen Zahl der Freiheitsgrade f wird die Zerlegungstafel aufgestellt, siehe Tabelle 3 und Tabelle 4.

Bei der dargestellten Rechentechnik erhält man die S.d.q.A. immer als Differenz nahezu gleichgroßer Zahlen. Deshalb muß die Stellenzahl der Rechenmaschine mindestens doppelt so groß sein wie die Stellenzahl der Beobachtungswerte. Die während der Rechnung anfallenden Zwischenergebnisse dürfen keinesfalls gerundet werden. Erst die S.d.q.A. selbst darf gegebenenfalls gerundet werden.

Man kann statt mit den Beobachtungswerten selbst auch mit transformierten Beobachtungswerten rechnen, wenn durch Subtraktion eines Hilfswertes bei der Mehrzahl der Beobachtungswerte Stellen wegfallen. Werden alle Beobachtungswerte um einen Hilfswert vermindert, dann vermindern sich alle Mittelwerte um den gleichen Hilfswert, während alle Varianzen unverändert bleiben.

4.4 Weitere Auswertung beim Modell mit Zufallskomponenten

4.4.1 Schätzwerte für die Modellparameter

Mit Hilfe der S.d.q.A. und der zugehörigen Zahl von Freiheitsgraden wird die Zerlegungstafel aufgestellt, siehe Tabelle 3.

Damit die einfache Aufteilung sinnvoll bleibt, müssen die Freiheitsgrade f_w und f_b mindestens gleich 1 sein, die unteren Schranken für n_1 und n_2 sind also jeweils gleich 2.

Aus (22) und der Zerlegungstafel erhält man als Ergebnis:

$$\text{Schätzwert für } \mu \text{ ist } x_{..} \quad (23)$$

$$\text{Schätzwert für } \sigma_\varepsilon^2 \text{ ist } \hat{\sigma}_\varepsilon^2 = s_w^2 \quad (24)$$

$$\text{Schätzwert für } \sigma_\xi^2 \text{ ist } \hat{\sigma}_\xi^2 = \frac{s_b^2 - s_w^2}{n_2} \quad (25)$$

Falls $\hat{\sigma}_\xi^2$ negativ ist, wird σ_ξ^2 durch Null geschätzt.

Neben den Schätzwerten $\hat{\sigma}_\xi^2$ und $\hat{\sigma}_\varepsilon^2$ werden häufig die Schätzwerte $\hat{\sigma}_\xi$ und $\hat{\sigma}_\varepsilon$ für die Streuungskomponenten benutzt:

$$\text{Schätzwert für } \sigma_\varepsilon \text{ ist } \hat{\sigma}_\varepsilon = \sqrt{s_w^2} \quad (26)$$

$$\text{Schätzwert für } \sigma_\xi \text{ ist } \hat{\sigma}_\xi = \sqrt{\frac{s_b^2 - s_w^2}{n_2}} \quad (27)$$

Weiter ist für die Variationskoeffizienten y_ε und y_ξ :

$$\text{Schätzwert für } y_\varepsilon = \frac{\sigma_\varepsilon}{\mu} \text{ ist } v_\varepsilon = \frac{\hat{\sigma}_\varepsilon}{x_{..}} \quad (28)$$

$$\text{Schätzwert für } y_\xi = \frac{\sigma_\xi}{\mu} \text{ ist } v_\xi = \frac{\hat{\sigma}_\xi}{x_{..}} \quad (29)$$

Tabelle 3. Zerlegungstafel für das Modell mit Zufallskomponenten

Variabilität	S.d.q.A. Q	Zahl der Freiheitsgrade f	Quotient $s^2 = \frac{Q}{f}$	Der Quotient ist Schätzwert für
zwischen den Gruppen	$Q_b = \frac{C}{n_2} - \frac{A^2}{n_1 \cdot n_2}$	$f_b = n_1 - 1$	$s_b^2 = \frac{Q_b}{f_b}$	$\sigma_\varepsilon^2 + n_2 \sigma_\xi^2$
innerhalb der Gruppen	$Q_w = B - \frac{C}{n_2}$	$f_w = n_1 (n_2 - 1)$	$s_w^2 = \frac{Q_w}{f_w}$	σ_ε^2
gesamt	$Q_{\text{total}} = B - \frac{A^2}{n_1 \cdot n_2}$	$f_{\text{total}} = n_1 n_2 - 1 = n - 1$	$s^2 = \frac{Q_{\text{total}}}{f_{\text{total}}}$	σ_ε^2 (falls $\sigma_\xi^2 = 0$)

4.4.2 Testen von Hypothesen über die Modellparameter

Mit dem F-Test läßt sich statistisch prüfen, ob die Varianz σ_ξ^2 zwischen den Gruppen existiert.

Dazu stellt man die Nullhypothese auf:

$$H_0: \sigma_\xi^2 = 0,$$

d.h., die Varianz zwischen den Gruppen ist nicht von Null verschieden; das untersuchte Material ist homogen.

Die zugehörige Alternativhypothese lautet:

$$H_1: \sigma_\xi^2 > 0,$$

d.h., die Varianz zwischen den Gruppen ist von Null verschieden; das untersuchte Material ist inhomogen.

Die Nullhypothese wird mit dem Prüfwert $F_B = s_b^2/s_w^2$ geprüft, der mit dem Tabellenwert $F_{f_b, f_w; 1-\alpha}$ der F-Verteilung zum Signifikanzniveau α verglichen wird. Die Tabellenwerte $F_{f_b, f_w; 1-\alpha}$ erhält man beispielsweise aus [1].

Ist $F_B \leq F_{f_b, f_w; 1-\alpha}$, wird die Nullhypothese nicht verworfen und das Material als homogen angesehen.

Ist $F_B > F_{f_b, f_w; 1-\alpha}$, wird die Nullhypothese zugunsten der Alternativhypothese verworfen und das Material als inhomogen angesehen.

4.4.3 Vertrauensbereiche für die Modellparameter

Wird die Nullhypothese über die Varianz zwischen den Gruppen nicht verworfen, wird die Grundgesamtheit durch μ und σ_ε^2 gekennzeichnet. Der Vertrauensbereich für den Gesamtmittelwert μ wird mit $n, x_{..}$ und s nach DIN 53 804 Teil 1 berechnet.

Wird die Nullhypothese über die Varianz zwischen den Gruppen zugunsten der Alternativhypothese verworfen, ergibt sich der Vertrauensbereich für den Mittelwert μ der Grundgesamtheit zum Vertrauensniveau $1 - \alpha$ zu:

$$x_{..} - t_{f_b; 1-\alpha/2} \cdot \sqrt{\left(1 - \frac{n_1}{M}\right) \frac{\hat{\sigma}_\xi^2}{n_1} + \frac{\hat{\sigma}_\varepsilon^2}{n_1 n_2}} \leq \mu \leq x_{..} + t_{f_b; 1-\alpha/2} \cdot \sqrt{\left(1 - \frac{n_1}{M}\right) \frac{\hat{\sigma}_\xi^2}{n_1} + \frac{\hat{\sigma}_\varepsilon^2}{n_1 n_2}} \quad (30)$$

wobei $t_{f_b; 1-\alpha/2}$ der Tabellenwert der t -Verteilung bei zweiseitiger Abgrenzung ist, siehe [1].

In dem häufig auftretenden Fall $M \gg n_1$, d.h., die Anzahl M der vorhandenen Gruppen ist groß gegen die Anzahl n_1 der zur Untersuchung herangezogenen Gruppen, ist der in (30)

auftretende Faktor $\left(1 - \frac{n_1}{M}\right)$ praktisch gleich Eins und kann vernachlässigt werden; der Vertrauensbereich hat dann die Form

$$x_{..} - t_{f_b; 1-\alpha/2} \cdot \sqrt{\frac{\hat{\sigma}_\xi^2}{n_1} + \frac{\hat{\sigma}_\varepsilon^2}{n_1 n_2}} \leq \mu \leq x_{..} + t_{f_b; 1-\alpha/2} \cdot \sqrt{\frac{\hat{\sigma}_\xi^2}{n_1} + \frac{\hat{\sigma}_\varepsilon^2}{n_1 n_2}} \quad (31)$$

Für σ_ξ^2 und σ_ε^2 lassen sich ebenfalls Vertrauensbereiche angeben (siehe [1]).

4.5 Weitere Auswertung beim Modell mit systematischen Komponenten

4.5.1 Schätzwerte für die Modellparameter

Mit Hilfe der S.d.q.A. und der zugehörigen Zahl von Freiheitsgraden wird die Zerlegungstafel aufgestellt, siehe Tabelle 4.

Tabelle 4. Zerlegungstafel für das Modell mit systematischen Komponenten

Variabilität	S.d.q.A. Q	Zahl der Freiheitsgrade f	Quotient $s^2 = \frac{Q}{f}$	Der Quotient ist Schätzwert für
zwischen den Gruppen	$Q_b = \frac{C}{n_2} - \frac{A^2}{n_1 \cdot n_2}$	$f_b = n_1 - 1$	$s_b^2 = \frac{Q_b}{f_b}$	$\sigma_\varepsilon^2 + \frac{n_2}{n_1 - 1} \sum_{i=1}^{n_1} (\mu_i - \mu)^2$
innerhalb der Gruppen	$Q_w = B - \frac{C}{n_2}$	$f_w = n_1 (n_2 - 1)$	$s_w^2 = \frac{Q_w}{f_w}$	σ_ε^2
gesamt	$Q_{\text{total}} = B - \frac{A^2}{n_1 \cdot n_2}$	$f_{\text{total}} = n_1 n_2 - 1 = n - 1$	$s^2 = \frac{Q_{\text{total}}}{f_{\text{total}}}$	$\sigma_\varepsilon^2 \left(\text{falls } \sum_{i=1}^{n_1} (\mu_i - \mu)^2 = 0 \right)$

Aus (1) und (22) und der Zerlegungstafel erhält man als Ergebnis:

$$\text{Schätzwert für } \mu_i \text{ ist } x_{i.} = \frac{A_i}{n_2} \quad (32)$$

mit $i = 1, 2, \dots, n_1$

$$\text{Schätzwert für } \mu \text{ ist } x_{..} = \frac{A}{n_1 \cdot n_2} \quad (33)$$

$$\text{Schätzwert für } \sigma_e^2 \text{ ist } \hat{\sigma}_e^2 = s_w^2 \quad (34)$$

4.5.2 Testen von Hypothesen über die Modellparameter

Mit dem F-Test läßt sich statistisch prüfen, ob Unterschiede zwischen den Mittelwerten μ_i bestehen.

Dazu stellt man die Nullhypothese auf:

$$H_0: \mu_1 = \mu_2 = \mu_3 = \dots = \mu_M$$

$$\text{bzw. } \sum_{i=1}^M (\mu_i - \mu)^2 = 0,$$

d. h., die Unterschiede zwischen den Mittelwerten μ_i sind nicht von Null verschieden; das untersuchte Material ist homogen.

Die zugehörige Alternativhypothese lautet:

$$H_1: \sum_{i=1}^M (\mu_i - \mu)^2 \neq 0,$$

d. h., die Unterschiede zwischen den Mittelwerten μ_i sind von Null verschieden; das untersuchte Material ist inhomogen.

Die Nullhypothese wird mit dem Prüfwert $F_B = s_b^2/s_w^2$ geprüft, der mit dem Tabellenwert $F_{f_b, f_w; 1-\alpha}$ der F-Verteilung zum Signifikanzniveau α verglichen wird. Die Tabellenwerte $F_{f_b, f_w; 1-\alpha}$ erhält man beispielsweise aus [1].

Ist $F_B \leq F_{f_b, f_w; 1-\alpha}$, wird die Nullhypothese nicht verworfen und das Material als homogen angesehen.

Ist $F_B > F_{f_b, f_w; 1-\alpha}$, wird die Nullhypothese zugunsten der Alternativhypothese verworfen und das Material als inhomogen angesehen.

4.5.3 Vertrauensbereiche für die Modellparameter

Wird die Nullhypothese über die Unterschiede zwischen den Mittelwerten μ_i nicht verworfen, wird die Grundgesamtheit bereits durch μ und σ_e^2 gekennzeichnet. Der Vertrauensbereich für μ wird mit $n, x_{..}$ und s nach DIN 53 804 Teil 1 berechnet.

Wird die Nullhypothese über die Unterschiede zwischen den Mittelwerten μ_i zugunsten der Alternativhypothese verworfen, sind die Vertrauensbereiche für die Gruppenmittelwerte μ_i zum Vertrauensniveau $1 - \alpha$:

$$\begin{aligned} x_{i.} - t_{f_w; 1-\alpha/2} \frac{\hat{\sigma}_e}{\sqrt{n_2}} &\leq \mu_i \\ &\leq x_{i.} + t_{f_w; 1-\alpha/2} \frac{\hat{\sigma}_e}{\sqrt{n_2}} \end{aligned} \quad (35)$$

Der Vertrauensbereich für den Mittelwert μ der Grundgesamtheit zum Vertrauensniveau $1 - \alpha$ ist

$$x_{..} - t_{f_w; 1-\alpha/2} \frac{\hat{\sigma}_e}{\sqrt{n_1 n_2}} \leq \mu \leq x_{..} + t_{f_w; 1-\alpha/2} \frac{\hat{\sigma}_e}{\sqrt{n_1 n_2}} \quad (36)$$

Dabei ist $t_{f_w; 1-\alpha/2}$ der Tabellenwert der t-Verteilung bei zweiseitiger Abgrenzung, siehe [1].

Für σ_e^2 läßt sich ebenfalls ein Vertrauensbereich angeben, siehe [1].

4.6 Sonderfall: Homogene Grundgesamtheit

Ist aus Erfahrung die Homogenität einer Grundgesamtheit hinsichtlich eines bestimmten Merkmals bei einer bestimmten Gruppeneinteilung bekannt, ist bereits eine dieser Gruppen repräsentativ für die Grundgesamtheit. Es ist dann ohne Einfluß auf das Ergebnis, ob eine, mehrere und welche Gruppen zur Probenahme herangezogen werden und wie die Gesamtanzahl n der zu untersuchenden Endproben auf die zur Untersuchung bereitgestellten Gruppen aufgeteilt wird. Bei homogener Grundgesamtheit kommt man also ohne Gruppeneinteilung aus.

Unerlässlich ist es aber, bei der Entnahme der Endproben aus den Gruppen Richtlinien zu beachten, die in DIN 53 803 Teil 2 behandelt werden.

5 Stichprobenaufbau bei inhomogener Grundgesamtheit

5.1 Modell mit Zufallskomponenten

Für das weitere wird angenommen, daß die Varianzen σ_ξ^2 und σ_e^2 aufgrund bereits vorliegender Erfahrungen (z. B. Auswertungen nach Abschnitt 4.4.1 bzw. Tabelle 3 anhand genügend großer Anzahlen von Beobachtungswerten) hinreichend genau bekannt sind.

Ziel der Auswertung ist, Aussagen über den Mittelwert μ der Grundgesamtheit zu machen. μ wird geschätzt durch den Mittelwert $x_{..}$ einer Stichprobe, wobei $x_{..}$ die Varianz σ_G^2 hat:

$$\sigma_G^2 = \left(1 - \frac{n_1}{M}\right) \frac{\sigma_\xi^2}{n_1} + \frac{\sigma_e^2}{n_1 \cdot n_2} \quad (37)$$

In dem häufig auftretenden Fall $M \gg n_1$, d. h., die Zahl M der vorhandenen Gruppen ist groß gegen die Zahl n_1 der zur Untersuchung herangezogenen Gruppen, ist der Faktor

$1 - \frac{n_1}{M}$ praktisch gleich Eins; $x_{..}$ hat dann die Varianz:

$$\sigma_G^2 = \frac{\sigma_\xi^2}{n_1} + \frac{\sigma_e^2}{n_1 \cdot n_2} \quad (38)$$

Der mit $x_{..}$ gebildete Vertrauensbereich zum vorgegebenen Vertrauensniveau $1 - \alpha$ hat in beiden Fällen die Gestalt:

$$x_{..} - u_{1-\alpha/2} \cdot \sigma_G \leq \mu \leq x_{..} + u_{1-\alpha/2} \cdot \sigma_G \quad (39)$$

wobei $u_{1-\alpha/2}$ der Tabellenwert der standardisierten Normalverteilung bei zweiseitiger Abgrenzung ist. Der Vertrauensbereich nach (39) hat also die halbe absolute Weite $W = u_{1-\alpha/2} \cdot \sigma_G$; W charakterisiert die Genauigkeit der Schätzung von μ durch $x_{..}$.

Falls die relative Weite vorgeschrieben ist, wird sie durch Multiplizieren mit $x_{..}$ in die absolute Weite umgerechnet.

Seien c_1 die Kosten für die Bereitstellung einer Gruppe und c_2 die Kosten für die Entnahme und Herstellung einer Endprobe und deren Messung, dann entstehen beim Stichprobenaufbau ($n_1; n_2$) folgende Gesamtkosten:

$$K_G = n_1 \cdot c_1 + n_1 \cdot n_2 \cdot c_2 \quad (40)$$

Die halbe Weite W des Vertrauensbereiches und die Gesamtkosten sind also vom Stichprobenaufbau ($n_1; n_2$) abhängig. Bezüglich des Stichprobenaufbaues werden (jeweils bei fest vorgegebenem Vertrauensniveau) folgende Fragestellungen behandelt:

- Varianzbestimmte Stichprobe (siehe Abschnitt 5.1.1)
- Varianzbestimmte kostenminimale Stichprobe (siehe Abschnitt 5.1.2)
- Kostenbestimmte Stichprobe (siehe Abschnitt 5.1.3)
- Kostenbestimmte varianzminimale Stichprobe (siehe Abschnitt 5.1.4)

5.1.1 Stichprobenaufbau bei vorgegebener Weite des Vertrauensbereiches für den Mittelwert der Grundgesamtheit; varianzbestimmte Stichprobe

n_1 und n_2 sind so zu bestimmen, daß für W die Forderung $W \leq W_0$ bei vorgegebenem Wert W_0 erfüllt wird. Berechnet man aus W_0 die Größe σ_0^2 gemäß

$$\sigma_0^2 = \left(\frac{W_0}{u_{1-\alpha/2}} \right)^2 \quad (41)$$

wobei $u_{1-\alpha/2}$ der Tabellenwert der standardisierten Normalverteilung bei zweiseitiger Abgrenzung ist, so läßt sich die Forderung $W \leq W_0$ auch in der Form $\sigma_G^2 \leq \sigma_0^2$ schreiben.

Die Varianz σ_G^2 für den Mittelwert der Stichprobe $\bar{x}_n$ nach (37)

$$\sigma_G^2 = \left(1 - \frac{n_1}{M} \right) \frac{\sigma_\xi^2}{n_1} + \frac{\sigma_\varepsilon^2}{n_1 \cdot n_2}$$

soll den vorgegebenen Wert σ_0^2 nicht überschreiten, also $\sigma_G^2 \leq \sigma_0^2$.

Sind die Varianzen σ_ξ^2 und σ_ε^2 bekannt und werden die Anzahlen n_1 und n_2 zunächst als stetig betrachtet, kann man diese aus der Forderung $\sigma_G^2 \leq \sigma_0^2$ berechnen.

Aus der Forderung $\sigma_G^2 \leq \sigma_0^2$ erhält man zunächst folgenden im allgemeinen nicht-ganzzahligen Wert für die untere Schranke für n_1 :

$$n'_1 = \frac{1}{\frac{\sigma_0^2}{\sigma_\xi^2} + \frac{1}{M}} = \frac{\sigma_\xi^2}{\sigma_0^2 + \frac{\sigma_\xi^2}{M}} \quad (42)$$

n'_1 ist für die weitere Rechnung eine Hilfsgröße, die **ohne** Rundung in (43), (44), (46) und (47) eingesetzt wird.

Zu dieser unteren Schranke für n_1 gehört folgender, im allgemeinen nicht-ganzzahliger Wert für die Schranke für n_2 :

$$n'_2 = \frac{n'_1}{n_{1+} - n'_1} \cdot \frac{\sigma_\varepsilon^2}{\sigma_\xi^2} \quad (43)$$

Dabei ist n_{1+} in (43) der auf die nächstgrößere ganze Zahl **aufgerundete** Wert n'_1 ; falls n'_1 ganzzahlig ist, wird dieser Wert um Eins vergrößert.

Die untere Schranke für n_2 ist $n'_2 = 2$; hierzu gehört der im allgemeinen nicht-ganzzahlige Wert für die obere Schranke für n_1 :

$$n''_1 = n'_1 \left(1 + \frac{\sigma_\varepsilon^2}{2\sigma_\xi^2} \right) \quad (44)$$

Diese oberen Schranken n''_1 und n'_2 werden jeweils auf die nächstgrößeren ganzen Zahlen n_{1++} und n_{2++} **aufgerundet**; sie bleiben unverändert, falls sie bereits ganzzahlig sind.

Weitere Ausführungen zu den Rundungsempfehlungen sind in Abschnitt 7 zusammengestellt.

Die Bereiche für die zulässigen ganzzahligen Werte n_1 und n_2 sind damit und unter Beachtung der Bedingung $n_1 \geq 2$:

$$2 \leq n_{1+} \leq n_1 \leq n_{1++} \quad 2 \leq n_2 \leq n_{2++} \quad (45)$$

Mit einer Anzahl von Gruppen unterhalb n_{1+} kann man die Forderung $\sigma_G^2 \leq \sigma_0^2$ nicht mehr einhalten. Mit einer Anzahl von Gruppen oberhalb n_{1++} bzw. mit einer Anzahl von Endproben pro Gruppe oberhalb n_{2++} wird σ_G^2 zunehmend kleiner gegenüber σ_0^2 , da im ersten Fall n_2 nicht kleiner als 2 bzw. im zweiten Fall n_1 nicht kleiner als n_{1+} werden kann. Die Anzahlen n_{1++} und n_{2++} brauchen also nicht überschritten zu werden.

Die zulässigen Wertepaare innerhalb der Bereiche nach (45) berechnet man wie folgt:

Ist ein nach (45) zulässiger ganzzahliger Wert n_1 vorgegeben, berechnet man:

$$\tilde{n}_2 = \frac{\sigma_\varepsilon^2}{n_1 \cdot \sigma_0^2 - \left(1 - \frac{n_1}{M} \right) \sigma_\xi^2} = \frac{\sigma_\varepsilon^2}{\sigma_\xi^2} \cdot \frac{n'_1}{n_1 - n'_1} \quad (46)$$

Der Wert $\tilde{n}_2$ wird auf die nächstgrößere ganze Zahl n_2 **aufgerundet** oder bleibt unverändert, falls er ganzzahlig ist.

Ist ein nach (45) zulässiger ganzzahliger Wert n_2 vorgegeben, berechnet man:

$$\tilde{n}_1 = \frac{\sigma_\xi^2 + \frac{\sigma_\varepsilon^2}{n_2}}{\sigma_0^2 + \frac{\sigma_\xi^2}{M}} = n'_1 \left(1 + \frac{1}{n_2} \cdot \frac{\sigma_\varepsilon^2}{\sigma_\xi^2} \right) \quad (47)$$

Der Wert $\tilde{n}_1$ wird auf die nächstgrößere ganze Zahl n_1 **aufgerundet** oder bleibt unverändert, falls er ganzzahlig ist.

Von den zulässigen Wertepaaren werden diejenigen als „günstige Wertepaare“ bezeichnet, bei denen die Varianz σ_G^2 nur wenig kleiner als σ_0^2 ist.

Nach fallweise verschiedenen Kriterien wählt man aus den zulässigen Wertepaaren ein Paar aus.

Aus den oberen und unteren Schranken der Bereiche nach (45) berechnet man die Spannweiten

$$R_1 = n_{1++} - n_{1+} \quad \text{und} \quad R_2 = n_{2++} - n_{2+}$$

Zur Berechnung der günstigen Wertepaare ist es zweckmäßig, einen zulässigen ganzzahligen Wert n_1 vorzugeben, falls $R_1 < R_2$ ist, und den zugehörigen Wert n_2 nach (46) zu berechnen, oder einen zulässigen ganzzahligen Wert n_2 vorzugeben, falls $R_2 < R_1$ ist, und den zugehörigen Wert n_1 nach (47) zu berechnen. Damit wird der Rechenaufwand am geringsten.

Beispiel 1:

Aus früheren Untersuchungen sind folgende Werte für die Varianzen bekannt:

$$\sigma_\xi^2 = 14 \quad \sigma_\varepsilon^2 = 47$$

Die Varianz σ_G^2 soll den Wert $\sigma_0^2 = 1,2$ nicht überschreiten. (Für dieses und die folgenden grundsätzlichen Beispiele bleibt die (Maß-)Einheit der Varianzen außer Betracht.) Die Grundgesamtheit besteht aus $M = 50$ Gruppen.

Nach den in Abschnitt 5.1.1 angegebenen Beziehungen (42), (43) und (44) berechnet man:

$$n'_1 = 9,46 \quad \text{aufgerundet auf } n_{1+} = 10$$

$$n'_1 = 25,34 \quad \text{aufgerundet auf } n_{1++} = 26$$

$$n'_2 = 58,75 \quad \text{aufgerundet auf } n_{2++} = 59$$

Die Bereiche, in denen n_1 und n_2 variiert werden können, sind damit in diesem Fall:

$$10 \leq n_1 \leq 26 \quad 2 \leq n_2 \leq 59$$

In der Tabelle 5 sind sämtliche günstigen Wertepaare ($n_1; n_2$) für dieses Beispiel zusammengestellt. Außerdem ist der Stichprobenumfang $n_1 \cdot n_2$ und die zugehörige Varianz σ_G^2 angegeben. Für die günstigen Wertepaare ist σ_G^2 nur wenig kleiner als $\sigma_0^2 = 1,2$. In diesem Fall wurden, beginnend mit $n_1 = 10$, ganzzahlige Werte für n_1 vorgegeben; mit (46) wurden die zugehörigen Werte n_2 berechnet.

Für die (in Tabelle 5 nicht aufgeführten) Wertepaare (25;2) bis (22;2), (20;3) und (19;3) sowie (17;4) ist $\sigma_G^2 > \sigma_0^2 = 1,2$; diese Wertepaare sind deshalb nicht zulässig.

Tabelle 5. Wertepaare $(n_1; n_2)$ und Varianz σ_G^2 für den Mittelwert der Stichprobe für Beispiel 1

n_1	n_2	$n_1 \cdot n_2$	σ_G^2
10	59	590	1,1997
11	21	231	1,1962
12	13	156	1,1879
13	9	117	1,1986
14	7	98	1,1996
15	6	90	1,1756
16	5	80	1,1825
18	4	72	1,1506
21	3	63	1,1327
26	2	52	1,1623

Für die (in Tabelle 5 nicht aufgeführten) Wertepaare (25;3) bis (22;3), (20;4) und (19;4) sowie (17;5) ist σ_G^2 merklich kleiner als $\sigma_0^2 = 1,2$; diese Wertepaare sind deshalb weniger günstig. Bei den in Tabelle 5 angegebenen „günstigen Wertepaaren“ nimmt n_1 um mindestens 1 zu und gleichzeitig n_2 um mindestens 1 ab im Vergleich zum vorhergehenden günstigen Wertepaar.

Da man $n_2 = 59$ nicht zu überschreiten braucht und $n_1 = 10$ nicht unterschreiten darf, um $\sigma_G^2 \leq \sigma_0^2 = 1,2$ einzuhalten, ist das Paar (10;59) das erste günstige Wertepaar.

Wie Beispiel 1 zeigt, kann man ausgehend von dem günstigen Wertepaar (10;59) mit dem „maximalen“ Umfang $n_1 \cdot n_2 = 590$ der Stichprobe (der nicht überschritten zu werden braucht) durch mäßiges Vergrößern von n_1 über n_{1+} hinaus die Werte für n_2 verkleinern, wobei sich der Stichprobenumfang $n_1 \cdot n_2$ drastisch reduzieren läßt bis hinab zum „minimalen“ Stichprobenumfang $n_1 \cdot n_2 = 52$, der nicht unterschritten werden darf.

5.1.2 Stichprobenaufbau bei vorgegebener Weite des Vertrauensbereiches für den Mittelwert der Grundgesamtheit und minimalen Gesamtkosten; varianzbestimmte kostenminimale Stichprobe

n_1 und n_2 sind so zu bestimmen, daß einerseits die Forderung $W \leq W_0$ bei vorgegebenem Wert W_0 eingehalten wird, andererseits die Gesamtkosten K_G minimal werden.

Die Varianz für den Mittelwert der Stichprobe nach (37)

$$\sigma_G^2 = \left(1 - \frac{n_1}{M}\right) \frac{\sigma_\xi^2}{n_1} + \frac{\sigma_\epsilon^2}{n_1 \cdot n_2}$$

soll den vorgegebenen Wert σ_0^2 nicht überschreiten, also $\sigma_G^2 \leq \sigma_0^2$. Die Gesamtkosten der Untersuchung nach (40)

$$K_G = n_1 \cdot c_1 + n_1 \cdot n_2 \cdot c_2$$

sollen minimiert werden.

Sind die Varianzen σ_ξ^2 und σ_ϵ^2 bekannt und werden die Anzahlen n_1 und n_2 zunächst als stetig betrachtet, ergibt sich mit Hilfe der Differentialrechnung theoretisch folgendes Paar im allgemeinen nicht-ganzzahliger Werte für die varianzbestimmte kostenminimale Stichprobe:

$$n_2^* = \sqrt{\frac{c_1}{c_2} \cdot \frac{\sigma_\epsilon}{\sigma_\xi}} \quad (48)$$

Der zugehörige Wert n_1^* ist nach (47):

$$n_1^* = \frac{\sigma_\xi^2 + \frac{\sigma_\epsilon^2}{n_2^*}}{\sigma_0^2 + \frac{\sigma_\xi^2}{M}} \quad (49)$$

Zu diesem Wertepaar $(n_1^*; n_2^*)$ gehören die theoretischen Minimalkosten

$$K_{G \min} = \frac{(\sigma_\xi \sqrt{c_1} + \sigma_\epsilon \sqrt{c_2})^2}{\sigma_0^2 + \frac{\sigma_\xi^2}{M}} \quad (50)$$

Die Werte n_1^* und n_2^* werden auf die nächstgrößeren ganzen Zahlen n_{1+}^* und n_{2+}^* aufgerundet.

Beginnend mit $(n_{1+}^*; n_{2+}^*)$ werden die Anzahlen n_1 und n_2 in der Umgebung von $(n_1^*; n_2^*)$ variiert. Dazu wird n_2 in dem Bereich $n_{2+}^* - d \leq n_2 \leq n_{2+}^* + d$ variiert, wobei $d = 4$ oder 3 gesetzt wird. Zu jedem dieser n_2 -Werte wird der zugehörige Wert n_1 nach (47) berechnet. Dabei werden der Stichprobenumfang $n_1 \cdot n_2$ sowie alle Wertepaare notiert, für die σ_G^2 nur wenig kleiner als σ_0^2 und K_G nur wenig größer als $K_{G \min}$ ist.

Nach fallweise verschiedenen Kriterien wählt man aus den in der Umgebung von $(n_1^*; n_2^*)$ liegenden Wertepaaren ein Paar aus.

Beispiel 2:

Für das Beispiel 1 mit den dort angegebenen Varianzen σ_ξ^2 und σ_ϵ^2 sowie der Forderung $\sigma_G^2 \leq \sigma_0^2 = 1,2$ mögen die Kosten (in frei wählbaren (Maß-)Einheiten) sein:

$$c_1 = 4 \quad c_2 = 1$$

Mit (48), (49) und (50) wird jetzt:

$$n_1^* = 18,13 \text{ aufgerundet auf } n_{1+}^* = 19$$

$$n_2^* = 3,66 \text{ aufgerundet auf } n_{2+}^* = 4$$

$$K_{G \min} = 138,92$$

In der Tabelle 6 sind für einige Wertepaare in der Umgebung von $(n_1^*; n_2^*)$ die zugehörige Varianz und die Kosten zusammengestellt.

Bei n_1^* ist die Rundung relativ groß. Damit entfernt man sich mit dem Wertepaar (19;4) von dem theoretischen Kostenminimum.

In solchen Fällen kann man oft andere Wertepaare finden, die in der Umgebung des Paares $(n_1^*; n_2^*)$ liegen, für die σ_G^2 nur wenig kleiner als σ_0^2 ist, und bei denen die Kosten K_G näher am theoretischen Minimum liegen als die Kosten beim Paar $(n_{1+}^*; n_{2+}^*)$.

Wie Tabelle 6 zeigt, sind die Wertepaare (16;5) und (18;4) kostengünstiger als das zunächst aus (48) und (49) berechnete und gerundete Paar (19;4).

Tabelle 6. Wertepaare $(n_1; n_2)$, Varianzen σ_G^2 und Gesamtkosten K_G für Beispiel 2

n_1	n_2	$n_1 \cdot n_2$	σ_G^2	K_G
18,13	3,665	66,45	1,199	138,97
19	4	76	1,075	152
14	7	98	1,199	196
15	6	90	1,176	150
16	5	80	1,183	144
18	4	72	1,151	144
21	3	63	1,133	147
26	2	52	1,162	156

5.1.3 Stichprobenaufbau bei vorgegebenen Gesamtkosten; kostenbestimmte Stichprobe

n_1 und n_2 sind so zu bestimmen, daß die Gesamtkosten K_G den vorgegebenen Betrag K_0 nicht überschreiten.