Das Mittelwertproblem

Dipl.- Ing. Björnstjerne Zindler, M.Sc.

www.Zenithpoint.de

Erstellt: 24. Dezember 2009 – Letzte Revision: 22. Februar 2021

Inhaltsverzeichnis

Literatur

[001] Keine für vorliegenden Text.

.

Die Kuh ist ersoffen, obwohl der Graben im Durchschnitt 30cm tief ist.

Statistiker

1 Mittelwertproblem I

[001]ff.

1.1 Vorbetrachtung

In der Statistik gibt es die vergleichende Möglichkeit der Ereignishäufigkeit η (Inzidenz) aus einer gegebenen festen Grundmenge P, konvertiert auf eine (sinnvolle) fixe Teilgrundmenge B bei bekannter Anzahl von beobachteten Fällen F.

$$\frac{F}{P} = \frac{\eta}{B}$$

$\Rightarrow$

$$\eta = \frac{F}{P} \cdot B$$

Um eine Verfeinerung „in der Fläche“ zu erreichen, wird die Grundmenge aufgeteilt in Teilgrundmengen $P_1, P_2, P_3 \cdots P_n$. Die einzelnen Ereignishäufigkeiten $\eta_1, \eta_2, \eta_3 \cdots \eta_n$ dienen dann als Vergleich zur Beurteilung des Ereignisses.

Die Frage ist, ob das überhaupt ohne weitergehende kritische Betrachtung möglich ist. Es ist evident, dass gelten muss:

$$P = P_1 + P_2 + P_3 + \ldots + P_n \qquad F = F_1 + F_2 + F_3 + \ldots + F_n$$

Dann muss aber auch:

$$\eta = \frac{\eta_1 + \eta_2 + \eta_3 + \ldots + \eta_n}{n}$$

Diese Forderung ist unbedingte Voraussetzung für eine Nutzung als eine vergleichende Ereignishäufigkeit.

Weiterhin:

$$\frac{F_1}{P_1} = \frac{\eta_1}{B} \qquad \frac{F_2}{P_2} = \frac{\eta_2}{B} \qquad \frac{F_3}{P_3} = \frac{\eta_3}{B} \qquad \cdots \qquad \frac{F_n}{P_n} = \frac{\eta_n}{B}$$

$\Rightarrow$

$$P_1 \cdot \frac{\eta_1}{F_1} = B \qquad P_2 \cdot \frac{\eta_2}{F_2} = B \qquad P_3 \cdot \frac{\eta_3}{F_3} = B \qquad \cdots \qquad P_n \cdot \frac{\eta_n}{F_n} = B$$

$\Rightarrow$

$$P_1 \cdot \frac{\eta_1}{F_1} = P_2 \cdot \frac{\eta_2}{F_2} = P_3 \cdot \frac{\eta_3}{F_3} = \cdots = P_n \cdot \frac{\eta_n}{F_n} = B$$

Sowie:

$$P = \frac{F_1}{\eta_1} \cdot B + \frac{F_2}{\eta_2} \cdot B + \frac{F_3}{\eta_3} \cdot B + \ldots + \frac{F_n}{\eta_n} \cdot B$$

$\Rightarrow$

$$\frac{F}{\eta} = \frac{F_1}{\eta_1} + \frac{F_2}{\eta_2} + \frac{F_3}{\eta_3} + \ldots + \frac{F_n}{\eta_n}$$

$\Rightarrow$

$$F = F_1 \cdot \frac{\eta}{\eta_1} + F_2 \cdot \frac{\eta}{\eta_2} + F_3 \cdot \frac{\eta}{\eta_3} + \ldots + F_n \cdot \frac{\eta}{\eta_n}$$

Das erfordert unbedingt:

$$\frac{\eta}{\eta_1} = \frac{\eta}{\eta_2} = \frac{\eta}{\eta_3} = \ldots = \frac{\eta}{\eta_n} = 1$$

Bedeutet, man vergleicht und sortiert gleich großes nach ihrer Größe. Unterschiede sind vornherein ein Fehler in der Aufteilung in den Grundmengen. Für diese muss grundsätzlich gelten:

$$P_1 = P_2 = P_3 = \ldots = P_n = \frac{P}{n}$$

$\Rightarrow$

$$\frac{F_n}{P_n} = \frac{\eta_n}{B} = n \cdot \frac{F_n}{P}$$

$\Rightarrow$

$$Q_n = n \cdot \frac{B}{P} \cdot \frac{F_n}{\eta_n} = 1$$

Letzteres ist ein nutzbarer Gütefaktor Q für alle η_n wobei gelten muss:

$$Q_1 + Q_2 + Q_3 + \ldots + Q_n = n$$

1.2 Schlussfolgerungen

• Bei einer möglichen Gleichverteilung von F ergibt sich somit insgesamt ein erhöhtes η_n für kleines P_n, was eindeutig dem Modell zuzuordnen ist und nicht einer Fallanhäufung von F.

•Eine vergleichende Beurteilung ist deshalb nur für $P_1 = P_2 = P_3 = \ldots = P_n$ zulässig.

2 Beispiele

2.1 Beispiel 1

In einem Landkreis leben $P = 200.000$ Einwohner. Davon gewinnen $F = 2$ Personen im Monat August einen Lottogewinn höher 1.000.000€. Das statistische Büro vermeldet dem Landrat eine Ereignishäufigkeit $\eta = 1$:

$$\frac{2}{200.000} = \frac{1}{100.000} = \frac{\eta}{100.000}$$

Der Landrat möchte ab einem $\eta_{\text{Grenz}} = 20$ eine Sonderabgabe für seine dann reichen Einwohner generieren, um den Haushalt des Kreises zu konsolidieren. Deshalb verlangt er eine feinere Untersuchung bis hinab zu einer Kommunalaufteilung. Diese ergibt Grunddaten $n = 2$ für die zentrale Stadt mit $P_S = 199.500$ Einwohner bei $F_S = 1$ und für das einzige Dorf im Landkreis bei $P_D = 500$ mit $F_D = 1$:

Stadt

$$\frac{1}{199.500} = \frac{0,501}{100.000} = \frac{\eta_S}{100.000}$$

$\Rightarrow$

$$\eta_S = 0,501$$

Dorf

$$\frac{1}{500} = \frac{200}{100.000} = \frac{\eta_D}{100.000}$$

$\Rightarrow$

$$\eta_D = 200$$

Damit ergibt sich für den gesamten Landkreis:

$$\eta_{\text{vorh}} = \frac{0,501 + 200}{2} = 100,25 > 20 = \eta_{\text{Grenz}}$$

Was den Landrat dazu veranlasst, sein Vorhaben durchzusetzen.

!Bitte jetzt keine rechtlichen Spitzfindigkeiten zum Beispiel!

Die Güten dazu hätten jedoch die Verzerrung der Ergebnisse an das Licht gebracht. Der Landrat hörte jedoch nicht auf mahnende Stimmen aus dem statistischen Büro. Die beiden Lottogewinner zogen in den Nachbarkreis. Die sonstig Steuerpflichtigen verrringerten sich somit um mindest 2. Ebenso eine Spende an den dörflichen Kindergarten und das städtische Schwimmbad flossen vorbei.

$$Q_S = 2 \cdot \frac{100.000}{200.000} \cdot \frac{1}{0,501} = 1,995 \neq 1 \qquad Q_D = 2 \cdot \frac{100.000}{200.000} \cdot \frac{1}{200} = 0,005 \neq 1$$

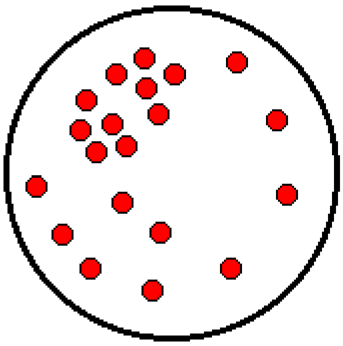
$\Rightarrow$

$$Q = Q_S + Q_D = 1,995 + 0,005 = 2 = n$$

2.2 Beispiel 2

Alle Fälle mit $B = 100.000$ und einem $\eta_{\text{Grenz}} = 2,1$

2.2.1 Fall 1



Mit:

$$P = 1.000.000 \qquad F = 20$$

⇒

$$\frac{20}{1.000.000} = \frac{2}{100.000} = \frac{\eta}{100.000}$$

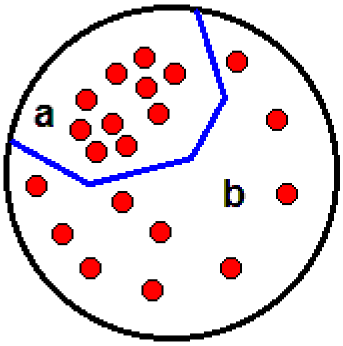
⇒

$$\eta = \mathbf{2} < \mathbf{2,1} = \eta_{\text{Grenz}}$$

Güte:

$$Q = 1 \cdot \frac{100.000}{1.000.000} \cdot \frac{20}{2} = 1$$

2.2.2 Fall 2



• $\mathbf{P}_a$

Mit:

$$P_a = 500.000 \qquad F_a = 10$$

$\Rightarrow$

$$\frac{10}{500.000} = \frac{2}{100.000} = \frac{\eta_a}{100.000}$$

$\Rightarrow$

$$\boldsymbol{\eta_a = 2 < 2,1 = \eta_{\text{Grenz}}}$$

Güte:

$$Q_a = 2 \cdot \frac{100.000}{1.000.000} \cdot \frac{10}{2} = 1$$

• $\mathbf{P}_b$

Mit:

$$P_b = 500.000 \qquad F_b = 10$$

$\Rightarrow$

$$\frac{10}{500.000} = \frac{2}{100.000} = \frac{\eta_b}{100.000}$$

$\Rightarrow$

$$\boldsymbol{\eta_b = 2 < 2,1 = \eta_{\text{Grenz}}}$$

Güte:

$$Q_b = 2 \cdot \frac{100.000}{1.000.000} \cdot \frac{10}{2} = 1$$

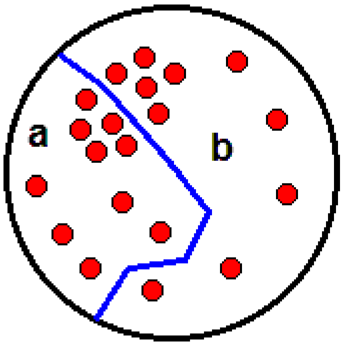
Zusammenfassen von η**:**

$$\eta = \frac{2+2}{2} = 2$$

Zusammenfassen von Q:

$$Q_a + Q_b = 1 + 1 = n = 2$$

2.2.3 Fall 3



• **P**$_a$

Mit:

$$P_a = 550.000 \qquad F_a = 10$$

$\Rightarrow$

$$\frac{10}{550.000} = \frac{1,818}{100.000} = \frac{\eta_a}{100.000}$$

$\Rightarrow$

$$\boldsymbol{\eta_a = 1,818 < 2,1 = \eta_{\text{Grenz}}}$$

Güte:

$$Q_a = 2 \cdot \frac{100.000}{1.000.000} \cdot \frac{10}{1,818} = 1,100 \approx 1$$

• **P**$_b$

Mit:

$$P_b = 450.000 \qquad F_b = 10$$

$\Rightarrow$

$$\frac{10}{450.000} = \frac{2,222}{100.000} = \frac{\eta_b}{100.000}$$

$\Rightarrow$

$$\eta_b = 2,222 > 2,1 = \eta_{\text{Grenz}}$$

Güte:

$$Q_b = 2 \cdot \frac{100.000}{1.000.000} \cdot \frac{10}{2,222} = 0,900 \approx 1$$

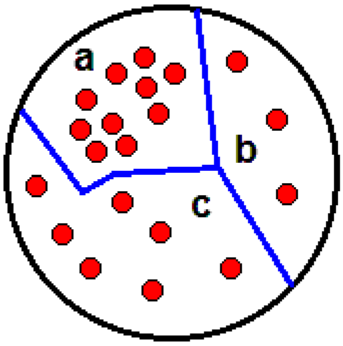
Zusammenfassen von η:

$$\eta = \frac{1,818 + 2,222}{2} = 2,020 \approx 2$$

Zusammenfassen von Q:

$$Q_a + Q_b = 1,100 + 0,900 = n = 2$$

2.2.4 Fall 4



• $\mathbf{P}_a$

Mit:

$$P_a = 500.000 \qquad F_a = 10$$

$\Rightarrow$

$$\frac{10}{500.000} = \frac{2,000}{100.000} = \frac{\eta_a}{100.000}$$

$\Rightarrow$

$$\boldsymbol{\eta_a = 2,000 < 2,1 = \eta_{\text{Grenz}}}$$

Güte:

$$Q_a = 3 \cdot \frac{100.000}{1.000.000} \cdot \frac{10}{2,000} = 1,500 \neq 1$$

• $\mathbf{P}_b$

Mit:

$$P_b = 200.000 \qquad F_b = 3$$

$\Rightarrow$

$$\frac{3}{200.000} = \frac{1,500}{100.000} = \frac{\eta_b}{100.000}$$

$\Rightarrow$

$$\boldsymbol{\eta_b = 1,500 < 2,1 = \eta_{\text{Grenz}}}$$

Güte:

$$Q_b = 3 \cdot \frac{100.000}{1.000.000} \cdot \frac{3}{1,5} = 0,600 \neq 1$$

• $\mathbf{P}_c$

Mit:

$$P_c = 300.000 \qquad F_c = 7$$

$\Rightarrow$

$$\frac{7}{300.000} = \frac{2,333}{100.000} = \frac{\eta_c}{100.000}$$

$\Rightarrow$

$$\eta_c = 2,333 > 2,1 = \eta_{\text{Grenz}}$$

Güte:

$$Q_c = 3 \cdot \frac{100.000}{1.000.000} \cdot \frac{7}{2,333} = 0,900 \approx 1$$

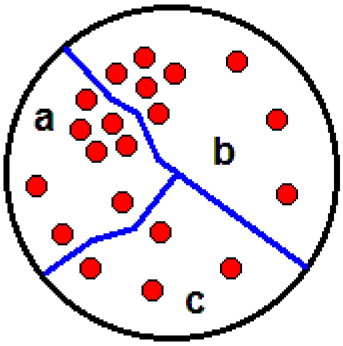
Zusammenfassen von η**:**

$$\eta = \frac{2,000 + 1,500 + 2,333}{3} = 1,944 \approx 2$$

Zusammenfassen von Q:

$$Q_a + Q_b + Q_c = 1,500 + 0,600 + 0,900 = n = 3$$

2.2.5 Fall 5



• **P**$_a$

Mit:

$$P_a = 400.000 \qquad F_a = 10$$

$\Rightarrow$

$$\frac{10}{400.000} = \frac{2,500}{100.000} = \frac{\eta_a}{100.000}$$

$\Rightarrow$

$$\eta_a = 2,500 > 2,1 = \eta_{\text{Grenz}}$$

Güte:

$$Q_a = 3 \cdot \frac{100.000}{1.000.000} \cdot \frac{10}{2,500} = 1,200 \neq 1$$

• **P**$_b$

Mit:

$$P_b = 450.000 \qquad F_b = 6$$

$\Rightarrow$

$$\frac{6}{450.000} = \frac{1,333}{100.000} = \frac{\eta_b}{100.000}$$

$\Rightarrow$

$$\boldsymbol{\eta_b = 1,333 < 2,1 = \eta_{\text{Grenz}}}$$

Güte:

$$Q_b = 3 \cdot \frac{100.000}{1.000.000} \cdot \frac{6}{1,333} = 1,350 \neq 1$$

• **P**$_c$

Mit:

$$P_c = 150.000 \qquad F_c = 4$$

$\Rightarrow$

$$\frac{4}{150.000} = \frac{2,666}{100.000} = \frac{\eta_c}{100.000}$$

$\Rightarrow$

$$\eta_c = 2,666 > 2,1 = \eta_{\text{Grenz}}$$

Güte:

$$Q_c = 3 \cdot \frac{100.000}{1.000.000} \cdot \frac{4}{2,666} = 0,450 \approx 1$$

Zusammenfassen von η**:**

$$\eta = \frac{2,500 + 1,333 + 2,666}{3} = 2,166 \neq 2$$

Zusammenfassen von Q:

$$Q_a + Q_b + Q_c = 1,200 + 1,350 + 0,450 = n = 3$$

3 Mittelwertproblem II

[001]ff.

Der wahre Mittelwert ist immer unbekannt. Der Mittelwert großer Stichproben nähert sich dem wahren Mittelwert nur an. Der Mittelwert kleinerer Teile dieser Stichproben sind selbst nur ein Teilmittelwert und Repräsentant des Mittelwertes großer Stichproben. Problem ist die Aussagekraft von Abweichungen in Bezug des Mittelwertes und Teilmittelwertes. Kann es dadurch sein, dass die Aussage *zu klein*, *zu warm* oder *zu wenig* in Bezug Mittelwert sich umkehrt auf *zu groß*, *zu kalt* oder *zu viel* in Bezug auf den Teilmittelwert? Kann folgende Bedingung erfüllt sein?

$$\triangle\overline{\mathbf{x}}_{\mathbf{4}} > n \cdot |\triangle\overline{\mathbf{x}}_{\mathbf{1}00}| \qquad \text{mit} \qquad n > 1 \qquad n \in R$$

3.1 Voraussetzungen

Gegeben sei eine Zufallsfolge von 100 Messungen. Der Mittelwert beträgt:

$$\overline{x}_{100} = \frac{1}{100} \sum_{i=1}^{100} x_i$$

Daraus werden 96 Messungen extrahiert und dessen Mittelwert gebildet:

$$\overline{x}_{96} = \frac{1}{96} \sum_{i=1}^{96} x_i$$

Die restlichen 4 Messungen ergeben einen gesonderten Mittelwert:

$$\overline{x}_{4} = \frac{1}{4} \sum_{i=1}^{4} x_i$$

Die Abweichung des 100. Wertes vom eigenen Mittelwert ist gefragt:

$$\triangle\overline{x}_{100} = x_{100} - \frac{1}{100} \sum_{i=1}^{100} x_i$$

Für eine Annahme, dass gilt:

$$x_{100} < \frac{1}{100} \sum_{i=1}^{100} x_i$$

Ergibt sich so für $\Delta\overline{x}_{100}$:

$$\triangle\overline{x}_{100} < 0$$

Die Abweichung des 100. Wertes vom 4- Messungs- Mittelwert ist gesucht:

$$\triangle\overline{x}_{4} = x_{100} - \frac{1}{4} \sum_{i=1}^{4} x_i$$

Für eine Annahme, dass gilt:

$$x_{100} > \frac{1}{4} \sum_{i=1}^{4} x_i$$

Ergibt sich so für $\Delta\overline{x}_{4}$:

$$\triangle\overline{x}_{4} > 0$$

3.2 Behauptung

Für die Beschreibung des Mittelwert- Problems soll überprüft werden, ob gelten könnte:

$$\triangle\overline{x}_4 > n \cdot |\triangle\overline{x}_{100}|$$

Wobei für die Konstante n gelten soll:

$$n > 1 \qquad\qquad n \in R$$

3.3 Beweis

Für:

$$\triangle\overline{x}_4 > n \cdot |\triangle\overline{x}_{100}|$$

Es wird eingesetzt:

$$x_{100} - \frac{1}{4}\sum_{i=1}^{4} x_i > n \cdot \left| x_{100} - \frac{1}{100}\sum_{i=1}^{100} x_i \right|$$

⇒

$$x_{100} - \frac{1}{4}\sum_{i=1}^{4} x_i > n \cdot \left(\frac{1}{100}\sum_{i=1}^{100} x_i - x_{100} \right)$$

Es wird weiter umgeformt:

$$x_{100} - \frac{1}{4}\sum_{i=1}^{4} x_i > \frac{n}{100}\sum_{i=1}^{100} x_i - n \cdot x_{100}$$

⇒

$$100 \cdot (1+n) \cdot x_{100} > n\sum_{i=1}^{100} x_i + 25\sum_{i=1}^{4} x_i$$

⇒

$$100 \cdot (1+n) \cdot x_{100} > 100 \cdot n \cdot \overline{x}_{100} + 100 \cdot \overline{x}_4$$

Vereinfacht:

$$(1+n) \cdot x_{100} > n \cdot \overline{x}_{100} + \overline{x}_4$$

⇒

$$n > \frac{\overline{x}_4 - x_{100}}{x_{100} - \overline{x}_{100}}$$

Die Konstante n kann aus der Behauptung substituiert werden:

$$n > 1$$

⇒

$$0 > \frac{\overline{x}_4 - x_{100}}{x_{100} - \overline{x}_{100}} - 1$$

⇒

$$x_{100} > \frac{\overline{x}_4 + \overline{x}_{100}}{2}$$

4 Beispiele

4.1 Beispiel 1

Die Wetterdaten speziell der Temperatur für den Monat November der letzten 100 Jahre ergaben einen Monatsdurchschnitt von:

$$\overline{x}_{100} = 5,0°C$$

Die letzten vier Messungen ergaben folgende Einzelwerte:

$$\begin{array}{ll} x_{97} = 4,0°C & x_{98} = 4,1°C \\ x_{99} = 3,9°C & x_{100} = 4,8°C \end{array}$$

Das ergibt einen Vierjahres- Mittelwert von:

$$\overline{x}_4 = 4,2°C$$

Man kann je nach benutztem Durchschnitt zwei folgende richtige Aussagen aufstellen:

1. Dieses Jahr (x_{100}) war mit $+0,6K$ der Monat November zu warm. ($\overline{x}_4$)
2. Dieses Jahr (x_{100}) war mit $-0,2K$ der Monat November zu kalt. ($\overline{x}_{100}$)

Diese Aussage sieht deshalb auf der Zu-warm-Seite dramatisch aus, weil gilt:

$$x_{100} > \frac{\overline{x}_4 + \overline{x}_{100}}{2}$$

$\Rightarrow$

$$4,8°C > 4,6°C$$

Der dazugehörige n-Wert:

$$n > \frac{\overline{x}_4 - x_{100}}{x_{100} - \overline{x}_{100}}$$

$\Rightarrow$

$$n > 3$$

Die Erwärmung fällt demnach $n = 3x$ dramatischer aus als die Abkühlung. Welche der beiden Aussagen ist aber nun aussagekräftiger? Das erklärt nur die Kenntnis der restlichen x_1 bis x_{96}.

4.2 Beispiel 2

Die Wetterdaten speziell der Temperatur für den Monat November der letzten 100 Jahre ergaben einen Monatsdurchschnitt von:

$$\overline{x}_{100} = 5,0°C$$

Die letzten fünf Messungen ergaben folgende Einzelwerte:

$$x_{96} = 5,2°C \quad x_{97} = 4,0°C$$
$$x_{98} = 4,1°C \quad x_{99} = 3,9°C \quad x_{100} = 4,8°C$$

Das ergibt einen Vierjahres- Mittelwert von (ohne x_{100}):

$$\overline{x}_4 = 4,3°C$$

Man kann je nach benutztem Durchschnitt zwei folgende richtige Aussagen aufstellen:

1. Dieses Jahr (x_{100}) war mit $+0,5K$ der Monat November zu warm. ($\overline{x}_4$)
2. Dieses Jahr (x_{100}) war mit $-0,2K$ der Monat November zu kalt. ($\overline{x}_{100}$)

Diese Aussage sieht deshalb auf der Zu-warm-Seite (immer noch) dramatisch aus, weil gilt:

$$x_{100} > \frac{\overline{x}_4 + \overline{x}_{100}}{2}$$

$\Rightarrow$

$$4,8°C > 4,65°C$$

Der dazugehörige n-Wert:

$$n > \frac{\overline{x}_4 - x_{100}}{x_{100} - \overline{x}_{100}}$$

$\Rightarrow$

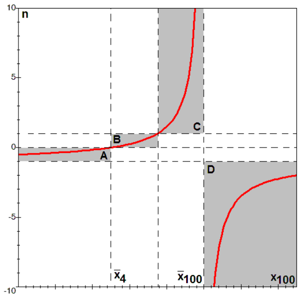
$$n > 2,5$$

Die Erwärmung fällt demnach (immer noch) $n = 2,5x$ dramatischer aus als die Abkühlung obwohl ein Kälteloch die letzten vier Jahre mit anschließender 100-Jahre-Mittelwert-Angleichung beobachtbar ist.

5 Zusammenfassung und Erweiterung auf $n \in R$

5.1 Für $\overline{x}_4 < \overline{x}_{100}$

Die obigen Erkenntnisse bildlich dargestellt.



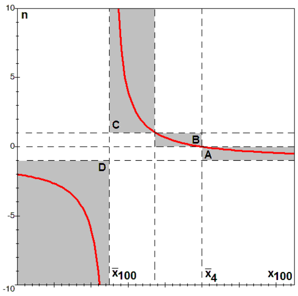
Darstellung der verschiedenen n-Wertbereiche
in Abhängigkeit von Mittel- und Teilmittelwert
für $\overline{x}_4 < \overline{x}_{100}$.

Wobei für die einzelnen Bereiche A bis D letztendlich gilt:

	D	A	B	C
Mittelwert	x_{100} zu warm	x_{100} zu kalt	x_{100} zu kalt	x_{100} zu kalt
Teilmittelwert	x_{100} zu warm	x_{100} zu kalt	x_{100} zu warm	x_{100} zu warm
Wert n	$n < -1$	$-1 < n < 0$	$0 < n < +1$	$n > +1$
Übereinstimmung	$(+)$	$(+)$	$(-)$	$(-)$

5.2 Für $\overline{x}_4 > \overline{x}_{100}$

Die obigen Erkenntnisse bildlich dargestellt.



Darstellung der verschiedenen n-Wertbereiche
in Abhängigkeit von Mittel- und Teilmittelwert
für $\overline{x}_4 > \overline{x}_{100}$.

Wobei für die einzelnen Bereiche A bis D letztendlich gilt:

	D	A	B	C
Mittelwert	x_{100} zu kalt	x_{100} zu warm	x_{100} zu warm	x_{100} zu warm
Teilmittelwert	x_{100} zu kalt	x_{100} zu warm	x_{100} zu kalt	x_{100} zu kalt
Wert n	$n < -1$	$-1 < n < 0$	$0 < n < +1$	$n > +1$
Übereinstimmung	$(+)$	$(+)$	$(-)$	$(-)$

Resümee: Bei unkommentierter Nutzung eines Teilmittelwerts können Temperaturwerte schnell zu falschen Aussagen entarten. Zu jedem Mittelwert gehört unbedingt die Angabe und Nutzung der (Standard)Abweichung.

LaTeX 2_ε