Gesetz zur Gruppenbildung - Beispielsammlung

ZUR GLEICHWERTIGKEIT VON PLAUSIBILITÄTSGRADEN

VERSTÄNDIGUNG

Die Herleitung der jeweiligen Plausibilitätsgrade (im Speziellen jene des Gesamt – Plausibilitätsgrades) – so wie in den Querformat – Tabellen üblicherweise dargestellt – basiert stets auf einem gesicherten Referenzprofil **oder** auf einem logischen Referenzprofil.

Ein **logisches Referenzprofil** muss erstellt werden, wenn keine gesicherte Häufigkeitsverteilung für die im Datensatz enthaltenden Variablen (umgeformt in maximal zehn unterschiedliche einstellige Zahlen) vorliegt. Das ist bei vielen praktischen Anwendungen der Fall.

Damit aber hängt der resultierende Plausibilitätsgrad **auch** von der Grösse der Datenlieferung ab.

Qualitativ: Weist ein Datensatz von bspw. 100 Messwerten (gut / schlecht → n =2) in der Summe eine Verteilung von z.B. 50% / 50% auf und führt dabei zu einem Plausibilitätsgrad $P_{(n = 2, Z = 50)}$ von bspw.75%, so ergibt ein **identischer** Datensatz – jedoch im Umfang von bspw. 500 Messungen und gleicher Abfolge – einen **merklich höheren**, d.h. günstigeren Wert.

Der Grund liegt darin, dass dann, wenn ein logisches Referenzprofil aus nur 100 Messwerten gebildet wird, die ermittelte Aufteilung der Einzelwerte nach Häufigkeit – und damit deren Gewichtung in der Urne – deutlich unzuverlässiger ist als bei einer Serie mit 1000 Einzelwerten von gleicher Häufigkeitsverteilung. Anders gesagt: Würde man 100 gewichtete Einzelwerte aus der realen Datenlieferung 10-mal in gleicher Zusammensetzung in die Urne geben, das Ganze gut durchschütten und dann wiederum 1000 Ziehungen (mit jeweiligem Zurücklegen) vornehmen, ergäbe sich ein systematisch höherer Plausibilitätsgrad für die zwar vergrösserte, aber strukturmässig unveränderte Datenreihe.

Zur Eliminierung dieses Unterschieds im Ergebnis (P) für ein und dieselbe Datencharakteristik, muss eine (ideelle) Angleichung auf einen einheitlichen Datenumfang von 1000 Einzelwerten vorgenommen werden.

Betrachtet man nach umfangreichen Vergleichsrechnungen die Plausibilitätsgrade für Datensätze aus ungestörter Zufallsentnahme je nach Grösse (Z) und Anzahl (n), so ergibt sich ein approximierter Zahlenraster (P%) gemäss **Tabelle 1.**

Das bedeutet: Bestünde der Datensatz nicht bloss aus bspw. Z = 50 Werten (bei n = 2), sondern aus Z = 500 Werten, beliefe sich P nicht auf 67%, sondern auf 88%. Unter Berücksichtigung des unterschiedlichen Datenumfangs sind demnach die beiden Prozentzahlen gleichwertig.

APPROXIMIERUNG

Z	n = 2	n = 3	n = 4	n = 5	n = 6	n = 7	n = 8	n = 9	n =10
50	67	69	71	73	75	77	79	81	83
100	74	75.5	77	78.5	80	81.5	83	82.5	86
200	81	82	83	84	85	86	87	88	89
500	88	88.5	89	89.5	90	90.5	91	91.5	92
1000	95	95	95	95	95	95	95	95	95

Tabelle 1

Es werden demnach in **Tabelle 2** entsprechende Quotienten (Zuschlagsfaktoren ZF) aus gegebenem $P_{(n = ...; Z = ...)}$, bezogen auf den jeweiligen Wert $ZF_{(n = ...; \mathbf{Z = 1000})}$ gebildet. Dieser Quotient hebt somit den naturgemäss tiefen Prozentsatz (infolge kleinerer Datenmenge) auf das Niveau für einen identischen Datensatz aus 1000 Werten der Abfolge.

Z	n = 2	n = 3	n = 4	n = 5	n = 6	n = 7	n = 8	n = 9	n =10
50	1.417910448	1.376811594	1.3380282	1.3013699	1.26667	1.2337662	1.202532	1.1728395	1.144578313
100	1.283783784	1.258278146	1.2337662	1.2101911	1.1875	1.1656442	1.144578	1.1515152	1.104651163
200	1.172839506	1.158536585	1.1445783	1.1309524	1.11765	1.1046512	1.091954	1.0795455	1.06741573
500	1.079545455	1.073446328	1.0674157	1.0614525	1.05556	1.0497238	1.043956	1.0382514	1.032608696
1000	1	1	1	1	1	1	1	1	1
r- generell > 0.996	FUNKTION generell = A*X^B								
A =	2.18669	2.048834	1.922957	1.807753	1.70021	1.604913	1.515422	1.432826	1.356457
B =	-0.114225	-0.104641	-0.095322	-0.086255	-0.07743	-0.068823	-0.060436	-0.052254	-0.044268

Tabelle 2

Aus Regressionsanalysen je nach Datenumfang (Z) folgt daraus für «jedes n»
eine Gesetzmässigkeit der Form: ZF = A*Z^B, mit den jeweils gültigen Werten A und B.

Mittels Regressionsanalyse je nach Anzahl der Variablen (n) schliesslich, folgt für den «Term A», wie auch für den «Term B» die Gesetzmässigkeit der Form A = a(A)+b(A)*n, bzw. B = a(B)+b(B)*n. Siehe **Tabelle 3 unten.**

	TERM A	TERM B
n = 2	2.18669	-0.114225
3	2.048834	-0.104641
4	1.922957	-0.095322
5	1.807753	-0.086255
6	1.700207	-0.077426
7	1.604913	-0.068823
8	1.515422	-0.060436
9	1.432826	-0.052254
10	1.356457	-0.044268
r > 0.9985	FUNKTION = a(A) + b(A)*n	FUNKTION = a(B) + b(B)*n
a(A) =	2.349357	A ~ 2.35 - 0.13*n
b (A) =	-0.103114	
a(B) =	-0.130605	B ~ (-) 0.13 +0.00874*n
b (B) =	0.008737	

ZUSCHLAG "ZF" ~ (2.35-0.1*n)*Z^(0.009*n - 0.13)

Tabelle 3

Der approximierte Zuschlagsfaktor (ZF) auf den jeweiligen Plausibilitätsgrad laut «Tabellen – Querformat» kann auch für einen auf (25 <= Z <=1000; 2<=n<=10) erweiterten Bereich angewendet werden. Eine Datenreihe mit weniger als 25 Einzelwerten sollte hingegen nicht mehr untersucht werden, da bei dieser geringen Anzahl die relativen Streuungen bei der benötigten «Paketbildung» zu gross werden.

Im Übrigen wird davon ausgegangen, dass die für gleiche Häufigkeitsverteilungen der «n» Variablen ermittelten Zuschlagsfaktoren (ZF) **auch für ungleiche Häufigkeiten** «hinreichende Gültigkeit» haben.

ERGÄNZENDE BETRACHTUNGEN

Mit der geschaffenen Gleichwertigkeit unter verschiedenen Plausibilitätsgraden je nach Datenumfang ist es möglich, eine eindeutige Aussage über «das Mass an Abweichung» einer Zahlenreihe von der empirischen Gesetzmässigkeit zur Gruppenbildung zu machen. Eine vollständige Übereinstimmung ($P_{100\%)}$ würde bedeuten, dass die nach «fairem Zufallsprinzip» anfallenden Daten/Voten/Messungen/Ereignisse/etc. in der Urne – d.h. innerhalb der Messperiode – absolut **homogen verteilt** sind.

Anschauliches Beispiel: Würde eine Sieblinie mit bestimmten Kornanteilen (z.B. für Betonkies) durch zufällige Kornentnahmen stichprobenartig überprüft, könnte so der Grad an (erwünschter) Homogenität/Durchmischung) quantifiziert werden.

Ähnliches gilt für alle Mischungen (und auch für Lösungen), die sich möglichst homogen darstellen sollen – soweit sich diese überhaupt messen lassen.

Unterhalb welchem Prozentsatz für (P) von chaotisch – turbulentem Verlauf in der Abfolge der Daten/Messwerte gesprochen werden muss, kann so allerdings nicht gesagt werden. Dies auch deshalb nicht, weil viele als chaotisch bezeichnete Abläufe sich nicht «Knall auf Fall» einstellen. Klassisches Beispiel: Der Übergang von laminarer zu turbulenter Strömung in der Hydromechanik. Zwar beziffert hier die sog. Reynoldszahl den Übertritt von laminar (heisst geordnet) zu turbulent (heisst chaotisch), doch die Reynoldszahl weist eine erhebliche Bandbreite von «Beginn» bis «Erfüllung» auf. Ab wann genau also ist die Strömung chaotisch?

Anderseits gibt es auch Reihenentwicklungen /Zahlenfolgen, welche offenbar keiner geordneten Gesetzmässigkeit folgen und demzufolge per se als chaotisch gelten. Ein klassisches Beispiel: Die Entwicklung der Collatz – Reihe. Es wird interessant sein, in einem weiteren Beitrag deren Plausibilitätsgrad (je nach gewählter Erstzahl) zu ermitteln.

Januar 2026

DIESER SCHLUSSATZ IST FALSCH(!):

Die Collatz – Reihen sind **grundsätzlich nicht** in der Lage, «Pakete» aus sich mehrfach folgenden identischen Zahlen (hier gerade oder ungerade) zu bilden, da die Bildungsregel dieser «(3n + 1) – Vermutung» zwingend dazu führt, dass sich **niemals zwei ungerade Zahlen unmittelbar folgen**. Die Reihen mögen zwar je nach Anfangszahl «unterschiedlich – chaotisch» (nicht vorhersehbar) verlaufen, dürfen diesbezüglich aber **nicht** mit dem «Gesetz zur Gruppenbildung» in Verbindung gebracht werden!

27.01.2026/Ba