

Gesetz zur Gruppenbildung – Beispielsammlung

DER PLAUSIBILITÄTSGRAD – BEDEUTUNG UND BERECHNUNG

VORBEMERKUNG

Mit der Wortschöpfung «Plausibilitätsgrad» sollte ursprünglich beschrieben werden, in welchem Umfang (bis zu welchem Grad) eine Abfolge von aus einer Urne gezogenen, einstelligen Ziffern je nach Häufigkeitsverteilung dem «Gesetz zur Gruppenbildung» folgt. Dies in Anlehnung an das «Gesetz von Benford», bei welchem sich ebenso die Frage stellt, als «wie plausibel» eine gelieferte Zahlenreihe erscheint, welche der «Erstzifferregel nach Benford» entsprechen soll.

Obwohl sich der «Plausibilitätsgrad» - was dessen Bedeutung als auch dessen Berechnung anbelangt – im Zusammenhang mit dem «Gesetz zur Gruppenbildung» auf sehr unterschiedliche Fragestellungen anwenden lässt, und sich dazu fallweise eine passendere Bezeichnung als wie «Plausibilitätsgrad» aufdrängen würde, soll die gewählte Wortschöpfung prinzipiell für alle Anwendungen beibehalten werden.

BEDEUTUNG

Der «Plausibilitätsgrad» ist eine Verhältniszahl zwischen einer Abfolge von Zahlen, welche dem «Gesetz zur Gruppenbildung» entspricht, und einer REALEN Abfolge von gemessenen, gelieferten oder sonst wie zustande gekommenen Zahlenreihe. Je nach Prozentsatz (Deckungsgrad), ist damit die «Güte der Zahlenlieferung» (oder eben deren Plausibilität) quantifiziert.

Als Bezugsgrösse dient ein sogenanntes **Referenzprofil** für die aus «geordnet – zufälliger Ziehung» in der Abfolge entstehenden ***Zahlenpakete (siehe unten)**. Dabei kommen zwei mögliche Referenzprofile in Betracht:

a) «Sicheres Referenzprofil»

In diesem Fall wird vorausgesetzt, dass die Zahlen in der Urne unter sich einer bekannten, bzw. gesicherten Häufigkeitsverteilung entsprechen. Damit ist die ***Paketbildung**, welche sich aus der Abfolge der aus der Urne gezogenen Zahlen (Variablen) ergibt, vorbestimmt. Dies zum Beispiel beim fairen Münzwurf, je 50%; beim Würfeln, je 16.666%; bei normalverteilten Zahlen zwischen 0 und 9; bei der Erstziffer – Verteilung nach Benford; etc.

Für die hier mittels Plausibilitätsgrad zu quantifizierende Übereinstimmung einer REALEN Stichproben (-Abfolge) mit ihrer Bezugsgrösse, wird zunächst in gleicher Weise ein «**reales (Stichproben -) Referenzprofil**» erstellt und dieses nach definierten Regeln (siehe Abschnitt **BERECHNUNG**) mit dem «**sicheren Referenzprofil**» verglichen.

***Zahlenpakete, *Paketbildung:**

Unter der Voraussetzung von «geordnet – zufälliger Ziehung» von Zahlen aus einer Urne, bilden sich – je nach Anzahl unterschiedlicher Ziffern, und je nach deren Häufigkeitsverteilung – laut dem (empirischen) «Gesetz zur Gruppenbildung» in der Abfolge der Ziehungen ganz bestimmte, nach Anzahl und Grösse strukturierte Pakete mit jeweils identischer Zahl.

Folgt aus der Untersuchung ein niedriger Plausibilitätsgrad bedeutet dies, dass die Abfolge der Stichprobedaten NICHT «geordnet – zufällig» erfolgt ist. Zum Beispiel: Der Münzwurf war «nicht – zufällig», also unfair; gleiches beim Würfeln; oder: Die Häufigkeitsverteilung der Erstziffern entspricht nicht dem Benford – Gesetz; usw.

Unterschiede zwischen «sicherem (oder logischem, siehe unten) Referenzprofil» und «realem Stichprobenprofil» ergeben **ungleiche Paketbildungen**, aus denen der Plausibilitätsgrad für die Datenlieferung berechnet wird. Der Unterschied zwischen SOLL und IST kann sich aus **drei Gründen* einstellen:

**Als Folge von zu kleiner Stichprobenzahl (Datenmenge), mit entsprechend grosser relativer Streuung.*

**Als Folge von Häufigkeitsverschiebungen unter den Urnenzahlen (Variablen) WÄHREND der Erhebungsphase. Klassischer Fall: Meinungsumschwung innerhalb einer laufenden Umfragekampagne.*

**Ungleichförmiges Ziehungsprozedere der Stichprobe, z.B. hervorgerufen durch ungenaues Mess – und/oder Anzeigegerät bei einer laufenden Materialprüfung.*

b) «Logisches Referenzprofil»

In diesem (häufigeren) Fall existiert keine Vorgabe über die Zusammensetzung der Zahlengruppe in der Urne, woraus ein «sicheres Referenzprofil» generiert werden könnte. Stattdessen wird hier ein «logisches Referenzprofil» erstellt. Grundlage dafür bildet die Häufigkeitsverteilung aus der Stichprobe selbst, z.B. die nach Abfolge festgehaltenen Stimmenanteile/Voten zu einer Umfrage. Für diese Häufigkeiten nach Voten werden auch hier aus 20x1000 Ziehungen die laut «Gesetz zur Gruppenbildung» entstehenden Zahlenpakete (siehe oben) generiert. Das weitere Prozedere ist sodann identisch zu jenem für/mit dem «sicheren Referenzprofil»

Bei der Überprüfung auf Plausibilität geht es darum, festzustellen, ob die Paketbildung aus der realen Stichproben – Abfolge jener Paketbildung möglichst nahe kommt, welche aus dem erzeugten «logischen Referenzprofil» hervorgegangen ist. Auch hier ist ein tiefer Plausibilitätsgrad (BERECHNUNG siehe unten) auf eine der drei obgenannten Gründe zurückzuführen.

BERECHNUNG

Die Berechnung des Plausibilitätsgrads (besser: der Plausibilitätsgrade) wird anhand des praktischen Beispiels «Ist dem Messgerät zu trauen?» beschrieben. Dazu wird auf dessen 4 Tabellen Bezug genommen (**ANHANG**)..

Tabelle 1 beziffert «Stück» und «Grösse» der Paketbildungen für das «logische Referenzprofil», welches aus den 122 Messdifferenzen zur Druckfestigkeitsermittlung eines Baumaterials generiert wurden. Unter «Stück» ist die Anzahl Zahlenpakete, bezogen auf jeweils 1000 Ziehungen, zu verstehen. Mit «Grösse» ist die Anzahl identischer Zahlen in einem Paket gemeint. Aus **Tabelle 1** geht hervor, dass sich Stückzahl und Grösse der Pakete unter den 10 verschiedenen Differenzbereichen (Bandbereiche von Differenzen) deutlich unterscheiden.

Tabelle 2 ist als Ergänzung zu **Tabelle 1** zu verstehen. Aus den 20 x 1000 Ziehungen folgen 20 minimal voneinander abweichende Paketzahlen. Während **Tabelle 1** den Mittelwert daraus angibt, liefert **Tabelle 2** die aus «Mittelwert, Standardabweichung und Wiederholung (20x) resultierende ($\pm$) Fehlerabweichung von **Tabelle 1**.

Tabelle 3 liefert die Paketbildungen analog **Tabelle 1**, wobei diese hier aus dem oben beschriebenen «Stichprobenprofil» – ausgehend von der ermittelten Häufigkeitsverteilung der Stichprobe und der tatsächlichen Abfolge der 123 Druckmessungen, resp. 122 Differenzen – gewonnen werden.

Mit diesen drei Tabellen sind die Vorbereitungen für die Berechnung des / der Plausibilitätsgrade abgeschlossen. **Tabelle 4** liefert die gesuchten Plausibilitätsgrade.

Berechnung der EINZELWERTE

Diese sind über die ganze **Tabelle 4** verteilt

Es sind vier Fälle zu unterscheiden:

Wenn $SS > (SR+FB)$, dann $P = ((SR+FB)/SS)*100$

Wenn $SS \leq (SR - FB)$, dann $P = (SS/(SR-FB))*100$

Wenn $(SR - FB) < SS < (SF+FB)$, dann $P = 100$

Wenn $SS = 0$ und $(SR = 0)$, dann $P = 0$

P = Plausibilitätsgrad; SR = Stückzahl Referenzpakete; SS = Stückzahl Stichprobenpakete;

FB = (±) Fehlerbandbreite

Berechnung der gewogenen ZAHLENBANDWERTE

Diese finden sich «spaltenweise» am Fuss von **Tabelle 4** verteilt

Es erfolgt «spaltenweise» eine Umrechnung der einzelnen Plausibilitätsgrade in einen – entsprechend der unterschiedlichen Inhaltsgrößen der Pakete – gewogenen Mittelwert (je Spalte resp. Zahlenband).

Berechnung der gewogenen GESAMT – PLAUSIBILITÄT

Diese findet sich am rechten unteren Rand von **Tabelle 4**.

Zur Ermittlung werden die «spaltenweise» gewogenen Mittelwerte laut unterster (roter) Tabellenzeile entsprechend der je Spalte/Zahlenband enthaltenen Anzahl Ziffern (zweitunterste blaue Tabellenzeile) gewichtet und daraus das gewogene Gesamtmittel = Gesamt – Plausibilität berechnet.

Gruppenbildung "m"	Promille Zahlenband 1	Promille Zahlenband 2	Promille Zahlenband 3	Promille Zahlenband 4	Promille Zahlenband 5	Promille Zahlenband 6	Promille Zahlenband 7	Promille Zahlenband 8	Promille Zahlenband 9	Promille Zahlenband 10	Promille ALLE Zahlenbänder 1 bis 10
m = 1	32	23	61	101	102	126	127	107	38	24	740
m = 2	1.1	0.7	3.7	14.5	15.8	22.3	24.3	15.2	1.8	0.6	100.5
m = 3	0	0.1	0.2	2.4	1.9	4.5	5.4	1.9	0.2	0	14
m = 4	0	0	0	0.2	0.3	1	1	0.2	0	0	2.4
m = 5	0	0	0	0.1	0	0.2	0.3	0	0	0	0.3
m = 6	0	0	0	0	0	0	0.1	0.1	0	0	0.1
m = 7	0	0	0	0	0	0.1	0	0	0	0	0.1
m = 8	0	0	0	0	0	0	0	0	0	0	0
m = 9	0	0	0	0	0	0	0	0	0	0	0
m = 10	0	0	0	0	0	0	0	0	0	0	0
NICHT ERFASSTE ANZAHL DATEN (alle für m > 10; gemittelt)	Bei extrem unterschiedlichen Gewichtungen der Urmenzahlen kommt es vor, dass bis zur maximal erfassten Gruppenbildung (bis m = 10) nicht alle 1000 Daten erfasst werden. Diese (gemittelte!) ANZAHL - bestehend aus Gruppen mit m > 10 - ist rot eingetragen.										0
ERFASSTE ANZAHL DATEN (alle bis m = 10; gemittelt)	34.2	24.7	69	138.5	140.5	189.8	197.9	144.5	42.2	25.2	1006.5 (GEMITTELTE ANZAHL AUS 20 x 1000)

Hinweis:
Da das Total (blau + rot) aus je 20 (mal 1000) getrennt gemittelten Ergebnissen besteht, muss dieses nicht zwingend = 1000 betragen!

Tabelle 1

Gruppenbildung "m"	Zahlenband 1	Zahlenband 2	Zahlenband 3	Zahlenband 4	Zahlenband 5	Zahlenband 6	Zahlenband 7	Zahlenband 8	Zahlenband 9	Zahlenband 10	Promille ALLE Zahlenbänder 1 bis 10
m = 1	6.75	4.36	6.43	8.99	7.31	10.04	10.95	7.8	5.31	6.32	14.9
m = 2	1.17	0.96	2.24	4.05	5.22	4.26	3.94	3.62	1.49	0.85	7.57
m = 3	0	0.21	0.53	1.39	1.49	2.13	2.24	1.6	0.43	0	4.05
m = 4	0	0	0	0.43	0.43	0.96	0.96	0.43	0	0	1.39
m = 5	0	0	0	0.21	0	0.43	0.53	0	0	0	0.53
m = 6	0	0	0	0	0	0	0.21	0.21	0	0	0.32
m = 7	0	0	0	0	0	0.21	0	0	0	0	0.21
m = 8	0	0	0	0	0	0	0	0	0	0	0
m = 9	0	0	0	0	0	0	0	0	0	0	0
m = 10	0	0	0	0	0	0	0	0	0	0	0
m = 10+	Bei extrem unterschiedlichen Gewichtungen der Urmenzahlen kommt es vor, dass bis zur maximal erfassten Gruppenbildung (bis m = 10) nicht alle 1000 Daten erfasst werden. Diese (gemittelte!) ANZAHL - bestehend aus Gruppen mit m > 10 - ist rot eingetragen.										0

Tabelle 2 (Fehlerbandbreite zu Tabelle 1)

	PROMILLEANTEILE	PROMILLEANTEILE	PROMILLEANTEILE	PROMILLEANTEILE	PROMILLEANTEILE	PROMILLEANTEILE	PROMILLEANTEILE	PROMILLEANTEILE	PROMILLEANTEILE	PROMILLEANTEILE	PROMILLEANTEILE
Gruppenbildung "m"	Zahl1	Zahl 2	Zahl 3	Zahl 4	Zahl 5	Zahl 6	Zahl 7	Zahl 8	Zahl 9	Zahl 10	PROMILLEANTEILE aller Zahlen 1 bis 10
m = 1	58	17.9	53.6	93.8	98.2	156.3	129.5	98.2	35.7	0	741.1
m = 2	0	4.5	4.5	13.4	22.3	8.9	40.2	4.5	0	0	98.2
m = 3	0	0	0	0	0	0	4.5	0	0	0	4.5
m = 4	0	0	0	0	4.5	0	0	0	0	0	4.5
m = 5	0	0	0	0	4.5	0	0	0	0	0	4.5
m = 6	0	0	0	0	0	0	0	0	0	0	0
m = 7	0	0	0	0	0	0	0	0	0	0	0
m = 8	0	0	0	0	0	0	0	0	0	0	0
m = 9	0	0	0	0	0	0	0	0	0	0	0
m = 10	0	0	0	0	0	0	0	0	0	0	0
NICHT ERFASSTE ANZAHL DATEN (alle für m > 10)	Bei extrem unterschiedlichen Gewichtungen der Urnenzahlen kommt es vor, dass die maximal erfasste Gruppenbildung (bis m = 10) nicht alle GELIEFERTEN Daten umfasst. Diese ANZAHL - bestehend aus Gruppen mit m > 10 - ist in "Zelle rechts unten" rot angegeben.										2
ERFASSTE ANZAHL DATEN (alle bis m = 1)	13	6	14	27	41	39	50.0416	24	8	0	222
GEWICHTUNG DER GELIEFERTEN DATEN (m)	58.56	27.03	63.06	121.62	184.88	175.68	225.41	108.11	36.04	0	1000

Tabelle 3

(Hilfsspalten überdeckt oder Zahlen weisst)	PLAUSIBILITÄTS - PROZENTE	PLAUSIBILITÄTS - PROZENTE	PLAUSIBILITÄTS - PROZENTE	PLAUSIBILITÄTS - PROZENTE	PLAUSIBILITÄTS - PROZENTE	PLAUSIBILITÄTS - PROZENTE	PLAUSIBILITÄTS - PROZENTE	PLAUSIBILITÄTS - PROZENTE	PLAUSIBILITÄTS - PROZENTE	PLAUSIBILITÄTS - PROZENTE	
Gruppenbildung "m"	Zahl1	Zahl 2	Zahl 3	Zahl 4	Zahl 5	Zahl 6	Zahl 7	Zahl 8	Zahl 9	Zahl 10	
m = 1	66.8	96	98.2	100	100	87	100	99	100	0	
m = 2	100	36.9	100	100	94.3	49.3	70.2	38.9	0	100	
m = 3	0	100	100	0	0	0	100	0	100	0	
m = 4	0	0	0	100	16.2	0	0	100	0	0	
m = 5	0	0	0	100	0	100	100	0	0	0	
m = 6	0	0	0	0	0	0	100	100	0	0	
m = 7	0	0	0	0	0	0	0	0	0	0	
m = 8	0	0	0	0	0	0	0	0	0	0	
m = 9	0	0	0	0	0	0	0	0	0	0	
m = 10	0	0	0	0	0	0	0	0	0	0	
ADDIERTE PROZENTE PRO ZAHL	3874.4	2050.5	6163.5	12060	14317.4	14475.6	19944.1	10071.9	3570	0	PLAUSIBILITÄT GESAMT [%]
ERFASSTE ANZAHL DATEN (alle bis m = 10)	13	6	14	27	41	39	50	24	8	0	222
GEMITTELTE PLAUSIBILITÄT [%] PRO ZAHL	66.8	76.2	98.5	100	78.1	83.1	89.3	94	100		87.2
HILFSZEILE	66.8	76.2	98.5	100	78.1	83.1	89.3	94	100	0	

23.06.2025 /Ba.// GESAMT - PLAUSIBILITÄT VERBESSERT 27.07.2025/Ba.

Tabelle 4