DATENWAHL BEIM «*GESETZ ZUR GRUPPENBILDUNG»

***ZU VERSTEHEN ALS MODELLANWENDUNG ZUM «GESETZ DER GROSSEN ZAHLEN» (GGZ)**

WORUM ES GEHT

Basierend auf der «Modellanwendung zum GGZ» soll für eine beliebige Abfolge von Daten ausgesagt werden, inwieweit (mit welchem prozentualen Anteil) sich eben diese Daten aus reinem Zufall – also gegenseitig völlig unbeeinflusst – gefolgt sind.

Entsprechende Berechnungen basieren dabei stets auf einem individulellen Referenzprofil, welches die Voraussetzung für absolute Zufälligkeit im Einzelfall abbildet. Bei der Entwicklung dieses Referenzprofils ist dabei folgendes zu beachten:

Es gibt einerseits die sogenannten «Lieferdaten», also das, was die erfasste Zahlenfolge (oder Stimmabgabe, oder...) tatsächlich quantifiziert, bzw. benennt.

Daneben gibt es die «Eingangsdaten für die Berechnung», wobei die «Lieferdaten» fallweise auch die «Eingangsdaten für die Berechnung» sein können.

Wie ist das zu verstehen?

Wenn es darum geht, die Unabhängigkeit der Stimmabgabe bei Befragungen, oder die Zufälligkeit der Anzahl von Schiffspassagen innerhalb eines definierten Zeitabschnitts, oder die Zufälligkeit von Messwerten bei der Materialprüfung, etc. zu quantifizieren, können die «Lieferdaten» zugleich als «Eingangsdaten für die Berechnung» verwendet werden. Man spricht dann vom «Datentyp Absolutwerte». Es kann damit vorbehaltlos und direkt auf die Zufälligkeit des Geschehens an sich geschlossen werden.

Würde man hingegen die realen Aktienkurse eines Wertpapiers direkt als «Eingangsdaten für die Berechnung» verwenden, ergäbe sich daraus eine unsinnige Information. Ähnliches gilt bspw. auch beim Wetter für den zeitlichen Verlauf der Aussenlufttemperatur:

Werden verschiedene Temperaturen aus zeitlichem Ablauf registriert und als Zahlen in eine Urne gelegt, so wäre es bei reinem Zufall nach der Ziehung der «Temperatur 10°C» gleichwahrscheinlich, dass unmittelbar danach die «Temperatur 11°C» oder die «Temperatur 18°C» folgen würde.

So spielt es sich aber in der Realität nicht ab, was belegt, dass sich Aussenlufttemperaturen (oder Aktienkurse, etc.) nicht «gegenseitig unbeeinflusst», also nicht rein zufällig folgen.

Hiergegen erscheint es zumindest als wahrscheinlich, dass sich die «Differenzen» oder die «logarithmischen Renditen» von nachfolgenden Temperaturen oder Kurswerten, etc. rein zufällig, d.h. unabhängig von der vorliegenden Differenz oder Rendite einstellen. Ob hierbei die Häufigkeiten der unterschiedlichen Grössen eine Normalverteilung bilden, was als Regelfall vermutet wird, ist dabei für den «Zufälligkeitsnachweis» ohne jede Bedeutung.

Wird somit – wie am praktischen Beispiel «Herzfrequenz – Variabilität» mittels der Pulsabstände = «Differenzen», oder am Beipiel «Aktienkurse» mittels deren «logarithmischen Renditen» als «Eingangsdaten» gerechnet, beziffert der resultierende PLAUSIBILITÄTSGRAD (P*ZF) % daraus das zufällige Zustandekommen eben DIESER Differenzen oder Renditen – UND NICHT UNMITTELBAR DIE ZUFÄLLIGKEIT DER «LIEFERDATEN»!

Die für eine sinnvolle Berechnung zu verwendenden «Eingangsdaten» sind hier also nicht die «Lieferdaten» selbst, sondern fallweise deren «Differenzen» oder « logarithmischen Renditen».

Fazit: Die Wahl des Datentyps als «Eingangsgrösse für die Berechnung» sowie für die daraus abzuleitende Bewertung des Geschehens ist stets vom ANWENDER des Verfahrens, d.h. mit dessen Sachverstand zum Thema vorzunehmen!

REGELN ZUR DATENWAHL

Im Sinn einer Orientierungshilfe für den Anwender des Verfahrens werden die folgenden Hinweise und EMPFEHLUNGEN zur Wahl des «Datentyps» abgegeben:

Methodisch stehen drei Datentypen als «Eingangsdaten zur Berechnung» zur Auswahl:

«Datentyp Absolutwerte»; die Lieferdaten bilden direkt die «Eingangsdaten»

«Datentyp Differenzen der Absolutwerte»; diese sind individuell herzuleiten

«Datentyp logarithmische Renditen der Absolutwerte»; sind ebenfalls herzuleiten

Merke: Jeder je nach «Datentyp» resultierende PLAUSIBILITÄTSGRAD hat seine mathematische Richtigkeit! Die Frage ist jedoch, aus welchem dieser Resultate eine naheliegende Schlussfolgerung für das interessierende «Geschehen per se» gezogen werden kann!

Zur Wahl des «Datentyps» wird empfohlen (Spezialfälle vorbehalten):

Alle Geschehen, deren Variablen augenscheinlich keinen direkten Bezug zueinander haben – oder von denen individuelle Unabhängigkeit vermutet oder erwartet wird – sind dem «Datentyp Absolutwerte» zuzuordnen. Der resultierende PLAUSIBILITÄTSGRAD liefert unmittelbar die «Zufälligkeit des Geschehens» in seinem Ablauf.

Jene Geschehen resp. Abläufe, deren Erkennungsmerkmal die «Differenzbildung per se» ist (z.B. Herzfrequenz – Variabilität HFV; allgemeiner: zufällige zeitliche Abstände für ein sich ad Infinitum wiederholendes Ereignis) sind logischerweise dem «Datentyp Differenzen» zuzuordnen. Hier stellen die «Lieferdaten» ebenfalls zugleich die «Eingangsdaten zur Berechnung» dar. Konsequenz wie oben!

Jene Geschehen, deren Variablen im zeitlichen Ablauf offensichtlich keine gegenseitige Unabhängigkeit aufweisen (Börsenkurse, Aussentemperaturen, Bevölkerungsentwicklung , etc., etc.) können mittels des «Datentyps Differenzen (D)» oder des «Datentyps logarithmische Renditen (R)» auf «Unregelmässigkeiten» untersucht werden. Es liegt dann an der Erfahrung und am Geschick des Anwender des Verfahrens, aus dem resultierenden Grad an Zufälligkeit (für den fallweise gewählten Datentyp (D) oder (R) bzw. für dessen komplementäre Nicht – Zufälligkeit) die richtigen Schlüsse mit **Blick auf das fragliche «Geschehen per se»** zu ziehen.

02.08.2026 / Ba.