

Statistik: Wirklichkeit und Wahrscheinlichkeit

Statistiken. Statistiken dienen dazu, etwas zu beweisen, zu widerlegen, zu rechtfertigen oder zu fordern. Sei es die Wirksamkeit neuer Medikamente oder die demografische Entwicklung der Gesellschaft. Statistik ist die Aufzeichnung vergangener Entwicklungen und Ereignisse. Jede Extrapolation einer Statistik in die Zukunft ist reine Spekulation.

Manipulation. Wenn das Resultat einer Statistik durch einen Auftraggeber bereits vorgegeben ist, wird es in der Regel auch entsprechend ausfallen. Dasselbe gilt für Studien, Gutachten und Expertisen.

Es gibt verschiedene Möglichkeiten, Statistiken zu manipulieren. Die gebräuchlichste Methode ist die Datenauswahl. Es werden nur Daten verwendet, die das gewünschte Resultat unterstützen. Andere Daten werden nicht berücksichtigt.

Eine weitere Methode ist die Darstellung der Statistik. Auf der einen Achse werden die Werte linear, auf der anderen Achse einfach oder doppelt logarithmisch dargestellt. Das verändert den Kurvenverlauf in der gewünschten Weise.

Signifikanz. Signifikanz ist die Wahrscheinlichkeit, dass ein Resultat kein Zufallsergebnis ist, ausgedrückt durch den p-Wert (probability). Je kleiner der p-Wert, desto sicherer ist das Resultat kein Zufallsprodukt. Eine statistische Signifikanz liegt vor, wenn p kleiner als 0.05 ist. Das bedeutet, dass die zufallsbedingte Unsicherheit unter 5% liegt. Die statistische Signifikanz ist von der zur Verfügung stehenden Datenmenge abhängig. Je mehr Daten, desto grösser die Wahrscheinlichkeit, dass das Resultat kein Zufallsergebnis ist. Eine Umfrage ist nicht repräsentativ, wenn nur wenige Personen eines Kollektivs befragt werden. Das Ergebnis ist nicht signifikant.

Korrelationen. Oft werden Zusammenhänge zwischen Beobachtungen und Ereignissen hergestellt, die nichts miteinander zu tun haben, aber positiv oder negativ korrelieren können. Beispielsweise der Milchpreis mit dem Zigarettenkonsum oder Störche mit den Geburtenraten.

Relationen. Man kann die Anzahl der an Grippe erkrankten Personen in einem bestimmten Land in absoluten Zahlen wiedergeben. Man kann diese Zahlen mit dem Vorjahr vergleichen. Das ergibt keinen Sinn. Der Bezug ist beispielsweise die Gesamtbevölkerungszahl: 500 Grippekranke auf 8 Millionen Menschen. Im laufenden Jahr sind es 600 Grippekranke. Das sind 20% mehr als im vergangenen Jahr. Eine Epidemie. Die Bevölkerung ist in diesem Zeitraum auf 9 Millionen Menschen angewachsen, also ein Plus von 12.5%. Rechnen wir es anders: 500 Grippekranke auf 8 Millionen ergibt 0.00625%. 600 Grippekranke auf 9 Millionen ergibt gerundet 0.00667%. Also doch keine Epidemie. Zudem werden nur Grippefälle erfasst, die auf Grund eines Arztbesuchs festgestellt werden. Eine häufige Fehlerquelle entsteht dadurch, dass absolute Zahlen und Prozentwerte vermischt werden. Oder es werden in manipulatorischer Ansicht gezielt absolute Zahlen oder Prozentwerte ausgewählt.

Prozentrechnen. Beim Prozentrechnen haben manche Menschen ein Defizit, aber trotzdem Erfolg, wie die folgende frei erfundene Geschichte zeigt: Ein vermögender Geschäftsmann wurde nach seinem Erfolgsrezept gefragt. Er antwortete: Ich handle mit Krimskram. Ich kaufe sie für einen Euro ein und verkaufe sie für drei Euro. Von diesen 2 Prozent lebe ich.

Durchschnitt. Der Durchschnitt ist das arithmetische Mittel, die Summe aller Werte geteilt durch die Anzahl der Werte. Beim Durchschnittseinkommen eines Unternehmens oder eines Landes verschieben Spitzenverdiener den Durchschnitt nach oben. Plötzlich verdient statistisch jeder mehr, als er wirklich verdient. Dies wiederum ist die Rechtfertigung für hohe Preise.

Median. Es gibt einen anderen Durchschnittswert, den Median. Das ist der Wert, der in der Mitte liegt. Bei der Ermittlung des Median werden die einzelnen Werte oder Gruppen von Werten sortiert. Die Einkommen werden beispielsweise in Einkommensgruppen eingeteilt und aufsteigend sortiert. Die Einkommensgruppe die in der Mitte liegt, ist der Median oder Zentralwert. Links und rechts der Mitte gibt es genau so viele Einkommensgruppen, die mehr und die weniger verdienen. Der Median hat den Vorteil, dass er resistent gegenüber extremen Einzeldaten ist.

Normalverteilung. Bei der Normalverteilungskurve (Gauss-Verteilung) gibt es ein mehr oder weniger breites Mittelfeld, das sich um einen Mittelwert konzentriert. Links und rechts des Mittelwerts fällt die Kurve stetig in Richtung der Extremwerte ab. Es gibt die grosse Masse des Mittelstands, einige Reiche und noch weniger Superreiche auf der einen Seite und arme und bitterarme Menschen auf der anderen Seite.

Die Verteilungskurve ist glockenförmig.

Pareto-Verteilung. Vilfredo Pareto (Italienischer Ökonom 1848-1923) hat nach seinen Beobachtungen festgestellt, dass zu seiner Zeit 20 Prozent der italienischen Familien 80 Prozent des Volksvermögens besaßen. Daraus entwarf er die 20/80 Prozent-Formel. Sie besagt, dass mit 20 Prozent Aufwand (Finanzaufwand, Energieaufwand, Zeitaufwand, Arbeitsaufwand) 80 Prozent der angestrebten Ergebnisse erreicht werden. Für die restlichen 20 Prozent an Ergebnissen beträgt der zusätzliche Aufwand 80 Prozent. Für die letzten zwei Prozent an Perfektion steigt der Aufwand weiter. Zwei Prozent mehr Perfektion kosten zusätzliche 98 Prozent Aufwand.

Es gibt Bereiche, in denen absolute Perfektion und Spitzenleistungen unabdingbar sind. Und es gibt Bereiche, in denen 80 Prozent Perfektion ausreichen.

Gemessen an der Bevölkerungszahl gibt es wenige Reiche und Superreiche, dann fällt die Kurve steil nach unten. Das ist der fehlende Mittelstand, der es nicht nach oben geschafft hat und nach unten durchgefallen ist. Am langen Ende der Verteilungskurve („long tail“) dümpelt die Kurve auf tiefem Niveau asymptotisch gegen Null (L-Verteilung).

Die Verteilungskurve ist asymmetrisch.