Einführung in die deskriptive Statistik

Übersicht:

Literatur:

Bortz, Jürgen (1999): Lehrbuch der Statistik. Für Sozialwissenschaftler, 3., 4. oder 5. Aufl. Berlin, Kapitel 1, 3 und 6 (Auszüge)

Hoppe, Siegfried / Detlef Liepmann (1974): Einführung in die Statistik für die Verhaltenswissenschaften. Teil I: Deskriptive Statistik, Stuttgart, Kapitel 1, 2 und 3.

Einführung in die deskriptive Statistik (Teil 1)

1. Grundlagen: Messen, Skalieren, eindimensionale Häufigkeitsverteilungen

Ziel der **deskriptiven Statistik** (beschreibenden Statistik) ist es, die erhobenen Daten zu strukturieren und zu beschreiben. Erst die schließende Statistik (Inferenzstatistik) beschäftigt sich mit der Frage, inwieweit sich die an einer Stichprobe erhobenen Daten auch darüber hinaus aussagekräftig und gültig sind.
Nachdem die Messwerte in einer Datenmatrix angeordnet wurden, können die Häufigkeiten der bestimmter Messwerte ermittelt und graphisch darstellt werden (Abschnitt 1). Sodann werden Zusammenhänge innerhalb der Daten mit verschiedenen statistischen Kennwerten beschrieben.
Die eindimensionale oder univariate Datenbeschreibung bezieht sich auf die Messung jeweils einer einzelnen Variablen, wobei Maßzahlen zur Kennzeichnung der zentralen Tendenz und zur Streuung der Messwerte um die zentrale Tendenz gebildet werden (Abschnitt 2).
Zweidimensionale oder bivariate Datenbeschreibung versucht die Beziehung zweier Variablen zueinander mit bestimmten Kennwerten zu charakterisieren (Abschnitt 3).

1.1. Messen

Wiederholung: Die Fragestellung einer empirischen Untersuchung bezieht sich auf ein oder mehrere **Merkmale** in einer Gruppe von Untersuchungseinheiten (z.B. Personen). Die Gruppe aller Untersuchungseinheiten, in denen Merkmale von Interessen sind, nennt man **Grundgesamtheit** oder **Population**. Da man nicht *alle* Elemente der Population untersuchen kann, befasst man sich meistens mit einer *Auswahl* aus der Grundgesamtheit, einer **Stichprobe**.

Die Elemente der untersuchten Stichprobe (z.B. Objekten, Personen oder Ereignissen) werden auch **Merkmalsträger** genannt. Die Merkmale, die verschiedene **Ausprägungen** annehmen, nennt man **Variablen**; die Merkmalsausprägungen selbst werden Variablenwerte oder Kategorien genannt.

Beim **Messen** wird jedem Merkmalsträger der Stichprobe genau ein Messwert zugeordnet und zwar dergestalt, dass diese Zuordnung die Verhältnisse der Merkmalsträger bezüglich der Merkmalsausprägungen untereinander wiedergibt. Messen heißt also: Zuordnen von Zahlen zu Merkmalsträgern nach bestimmten Regeln.

Der Messvorgang beinhaltet die Abbildung eines empirischen Relativs (Merkmalsausprägungen bei Merkmalsträgern) in ein numerisches Relativ (Zahlen).
Ein **Relativ** ist eine Menge von Elementen (z.B. Merkmalsträger oder Zahlen), innerhalb derer es Beziehungen zwischen den Elementen bezüglich eines Merkmals (z.B. Geschlecht oder Körpergröße) gibt (z.B. die Beziehungen: „gleich“ „ungleich“, „kleiner als“).

Bsp.: Wir haben zwei Relative:

Relativ 1 = Stichprobe von Studenten, Gleichheit bzw. Ungleichheit des Geschlechts

Relativ 2 = Menge zweier Zahlen (1, 2), Gleichheit bzw. Ungleichheit dieser Zahlen.

Relativ 1 ist ein empirisches Relativ, Relativ 2 ein numerisches Relativ.

Messen ist eine Zuordnung von Zahlen zu Merkmalsträgern, bei der die Relationen der Merkmalsträger bezüglich ihrer Merkmalsausprägungen durch die ihnen zugeordneten Zahlen wiedergegeben werden.

Bei *quantitativen Variablen* sind die Merkmalsausprägungen (Variablenwerte, Kategorien) abgestuft und können durch Zahlen wiedergegeben werden. Bei *qualitativen Variablen* lassen sich die Merkmalsausprägungen (Variablenwerte, Kategorien) nicht in eine abgestufte Reihenfolge bringen. Die Zuordnung von Zahlen ist daher willkürlich (Bsp. Geschlecht). *Diskontinuierliche / diskrete* Variablen besitzen eine begrenzte Anzahl von Variablenwerten. In jedem Messwertintervall liegen nur endlich viele Messwerte. *Kontinuierliche / stetige* Variablen haben eine unbegrenzte Anzahl von Variablenwerten (z.B. physikalische Maße: Dauern, Längen usw.) In jedem Intervall zwischen zwei Messwerten liegen unendlich viele Messwerte.

1.2. Skalenniveaus

Je nach der Qualität der Beziehungen zwischen den Merkmalsträgern bezüglich ihrer Merkmalsausprägungen kann die Messung der Variablen auf unterschiedliche Weise erfolgen:

(1) Feststellung der Gleichheit oder Ungleichheit: $A = B$ oder $A \neq B$ (Nominalskala).
(2) Beziehung einer Rangordnung: A < B < C (Ordinalskala).
(3) Rangordnungsbeziehung, bei der auch die Größen der Unterschiede zwischen den einzelnen Rängen miteinander verglichen werden können: B - A = C – B (Intervallskala).

Unter einer **Skala** verstehen wir die Menge von Messwerten einer Variablen, die durch den Messvorgang den Merkmalsträgern zugeordnet werden können. Dabei wird jedem Merkmalsausprägung ein **Skalenwert** zugewiesen.

Anm.: Natürlich muss nicht bei jedem Skalenwert / jeder Merkmalsausprägung in der gegebenen Stichprobe auch ein oder mehrere Messwerte auftreten.

Typologie der **Skalenniveaus:**
(1) Bei **Nominalskalen** können nur Aussagen darüber getroffen werden, ob die Messwerte eines Merkmals gleich oder ungleich sind. Die Skalenwerte müssten nicht unbedingt durch Zahlen, sondern könnten genauso gut andere durch andere Symbole bezeichnet werden.
Dichotome Skalen besitzen zwei Skalenwerte.
Trichotome Skala besitzen drei Skalenwerte;
Mehrkategoriale Skala besitzen mehr als drei Skalenwerte.
(2) Bei **Ordinalskalen** (oder: Rangskalen) kann darüber hinaus ausgesagt werden, ob ein

Messwert größer, kleiner oder gleich einem anderen ist. Alle Messwerte können in eine Rangordnung gebracht werden.
(3) Bei **Intervallskalen** kann darüber hinaus ausgesagt werden, ob der *Unterschied* zweier Messwerte größer, gleich oder kleiner als der Unterschied zweier anderer Messwerte ist.
Erst auf dem Skalenniveau der Intervallskalen können Skalenwerte bezüglich ihres Abstandes zueinander, also bezüglich ihrer Differenz (und ihrer Summe), miteinander verglichen werden („gleichabständige" Skalenwerte); erst hier ist die Addition oder Subtraktion von Messwerten zulässig.
In einer Intervallskala können der Nullpunkt und die absolute Größe der Abstände oder Intervalle frei gewählt werden. Die auf diese Weise entstandenen Skalen lassen sich durch einfache mathematische Transformationen ineinander überführen. Daraus folgt jedoch:
- Es können *keine* Aussagen über die absoluten Beträge eines Skalenintervalls bzw. des Unterschieds zweier Messwerte getroffen werden.
- Es können keine Aussagen über das Verhältnis zweier Messwerte (= Prozente, Anteilwerte) getroffen werden, weil hierzu ein Nullpunkt als Ausgangspunkt notwendig wäre.

(4) **Ratioskalen** (Verhältnisskalen) besitzen zudem einen natürlichen Nullpunkt. Diese erlaubt ein Multiplizieren und Dividieren von Skalenwerten miteinander. Ratioskalen sind in der Psychologie und Soziologie jedoch ausgesprochen selten.
Bsp.: physikalische Maße: Länge, Gewicht usw.

Beachte:
1. Die Skalenniveaus sind *hierarchisch* angeordnet. D.h. Skalen auf einem höheren Skalenniveau besitzen alle Eigenschaften des niedereren Skalenniveaus.
2. Das Skalenniveau einer Messung wird durch Plausibilitätsschluss ermittelt.
Während physikalischen Größen fast immer Intervallskalen (oder sogar Ratioskalen) sind, ist dieses Skalenniveau bei psychologischen oder soziologischen Größen nur selten eindeutig nachweisbar (Ausnahme: Psychophysik).
3. In der Forschung geht man bisweilen dennoch von Intervallskalen aus, obwohl dieses Skalenniveau nicht nachweisbar ist.
- Annahme eines „regelmäßigen" Aufbaus der Welt(sicht) der meisten Menschen mit gleichabständigen Skalenintervallen. Bei Befragungen sollen die Versuchspersonen (Merkmalsträger) diesen „inneren Maßstab" gleichabständig auf ein numerisches Relativ (z.B. Ratingskala) übertragen.
4. Das Skalenniveau einer operationalisierten, manifesten Variable (z.B. Ausbildungszeit in Jahren) entspricht nicht automatisch demjenigen der dahinter stehenden latenten Variable (theoretisches Konzept: „Bildungsstand"). Auch wenn der Indikator offensichtlich intervall- oder ratioskaliert ist, bleibt die latente Variable ordinalskaliert.

1.3. Messwertklassen

Bei stetigen Variablen (insbesondere bei physikalischen Messungen) sowie bei Variablen mit einer großen Anzahl von Ausprägungen werden aus Gründen der Anschaulichkeit und Übersichtlichkeit Messwerte vielfach zu **Messwertklassen** (Intervalle, Kategorien) zusammengefasst. Man spricht dann von „gruppierten Daten". Die Anzahl k der Klassen / Kategorien sollte unter 20 liegen. Prinzipiell kann jede Skala von Messwerten durch Klassenbildung verändert werden. Damit das Skalenniveau erhalten bleibt, sind folgende Regeln zu beachten:

1. Kategorien von Nominalskalen können beliebig zusammengefasst werden, insofern die Zuordnung der Messwerte eindeutig bleibt. Die Klassenbildung sollte inhaltlich begründet sein.
2. Ordinalskalen können durch Zusammenlegen benachbarter Ränge vergröbert werden. Je größer die Variationsbreite der Messwerte ist, desto größer können die Klassen sein.
3. Bei der Vergröberung von Intervallskalen ist darauf zu achten, dass die Klassenbreiten (kb) der Klassen / Intervalle gleich sind. Die neu gebildete Klasse wird durch die Klassenmitte bezeichnet, das ist das arithmetische Mittel der unteren und der oberen Intervallgrenze.
4. Insbesondere bei Skalen von physikalischen Messungen, die nur selten über den gesamten theoretisch möglichen Bereich Messwerte liefern, kann es sinnvoll sein, am oberen und unteren Rand der Skala **offene Messwertklassen** zu bilden, in die alle Messwert bis bzw. ab einem bestimmten Messwert fallen. Bei einer statistischen Beschreibung dieser Skalen muss jedoch berücksichtigt werden, dass bei Einbeziehung der offenen Klassen kein Intervallskalenniveau mehr gegeben ist.

1.4. Univariate Häufigkeitsverteilungen

Nach einer Datenerhebung liegen die Messwerte eines Merkmals zunächst in Form einer sog. **Urliste** vor. In dieser Zahlenliste werden die Messwerte der einzelnen Untersuchungseinheiten hintereinander aufgelistet. In einer **Datenmatrix** werden die Messwerte von mehreren Variablen in Spalten nebeneinander zusammengefasst. Eine Untersuchungseinheit (Person) bildet mit ihren Messwerten jeweils eine Zeile der Matrix. Alle Messwerte einer bestimmte Variable stehen untereinander in einer Spalte.
Vielfach ist es sinnvoll, vor der statistischen Auswertung der Daten
- durch Indexbildung mehrere Variablen zu neuen Variablen (Indices) zusammenzufassen;
- durch Klassenbildung mehrere Kategorien in eine Klasse zusammenzufassen;
- durch Klassenbildung eine stetige in eine diskrete Variable (mit einer endlichen Anzahl von Klassen/Messwertintervallen) zu transformieren.

Einige konventionelle Abkürzungen und Bezeichnungen in der deskriptiven Statistik:
Vp = „Versuchsperson" (heute sagt man eher: Versuchs- oder Untersuchungsteilnehmer);
n = die Anzahl der Vps bei Stichprobenuntersuchungen, Umfang der Messung/Untersuchung. Die Größe der Grundgesamtheit wird mit **N** bezeichnet.

Die Variable, die wir bei den Versuchspersonen gemessen haben, wird **x** genannt. x ist der Messwert-Vektor der verschiedenen Messwerte x_i.
Die einzelnen Untersuchungsteilnehmer werden durchnummeriert: 1, 2, 3, ... i .... bis n. („i“ bedeutet irgendeine beliebige Versuchsperson).
Mit $\mathbf{x_i}$ wird der Messwert bezeichnet, der bei der Versuchsperson Vp_i bezüglich einer Variablen x gemessen wurde.
Die Urliste ist die Auflistung aller Messwerte x_i aller Vps in der Reihenfolge 1, 2, 3, ... i .... bis n.
Die einzelnen Skalenwerte der Messung nennen wir x_1, x_2, x_3 ... x_l; die Anzahl der Skalenwerte m.
Skalenklassen werden mit k_1, k_2, k_3 usw. bezeichnet.

Die Verteilung der Messwerte (x_i) einer Variable (x) auf die Skalenwerte (x_l) werden **Häufigkeitsverteilung** (oder einfach: Verteilung) genannt

Mit $\mathbf{f(x_l)}$ wird die Häufigkeit bezeichnet, mit der ein Skalenwert x_l innerhalb der Stichprobe als Messwert auftritt (f steht hier für englisch *frequency*, Häufigkeit).

- Vielfach wird in Statistik-Lehrbüchern zwischen Skalenwert x_l und dem Messwert einer Vp x_i *nicht* unterschieden.
- Mitunter steht statt $f(x_l)$ auch $f(x)$ oder f_i (bei Klassen: f_k).
- Bei Nominal- oder Ordinalskalen müssen die Skalenwerte nicht unbedingt durch Zahlen, sondern können auch durch die Kategoriennamen bezeichnet werden.

In einer **Häufigkeitstabelle** werden jedem Skalenwert x_l bzw. jeder Klasse k_l (1. Spalte) in der zweiten Spalte ein entsprechender Häufigkeitswert f(x) bzw. f(k) (genauer: $f(x_l)$ bzw. $f(k_l)$) zugeordnet.
Den Häufigkeitswert kann man auch als Prozentwert ausdrücken %(x) bzw. %(k).

$$\%(x_l) = \frac{f(x_l)}{n} \cdot 100\%.$$

Manchmal möchte man wissen, wie viele Messwerte gerade bis zu einem bestimmten Skalenwert vorliegen. Die Verteilung dieser bis zu einem Skalenwert angehäuften Messwerte nennt man **kumulierte Häufigkeitsverteilung $\mathbf{f_{kum}(x_l)}$** bzw. $\mathbf{f_{kum}(k_l)}$ (auch: $f_c(x)$). Vielfach wird die kumulierte Häufigkeitsverteilung der Häufigkeitstabelle als eine dritte Spalte angefügt. Auch die kumulierte Häufigkeitsverteilung lässt sich in Prozentzahlen ausdrücken (sog. Prozentränge, PR):

$$PR = \%_{kum}(x_l) = \frac{f_{kum}(x_l)}{n} \cdot 100\%.$$

1.5. Graphische Darstellung von univariaten Häufigkeitsverteilungen

Häufigkeitstabellen können auf verschiedene Weise anschaulich graphisch dargestellt werden.

(1) **Diagramme**: Streckendiagramm, Säulendiagramm. Die Häufigkeiten der einzelnen Skalenwerte bzw. Klassen werden in Flächen oder Längen übersetzt, sodass die Flächen- oder Längenverhältnisse die Verhältnisse der Häufigkeiten wiedergeben. Diese Darstellung ist auch bei Daten aus Nominalskalen sinnvoll. Bei Ordinalskalen sind nur solche Diagramme sinnvoll, in denen die Flächen / Längen waagrecht oder senkrecht nebeneinander angeordnet sind, sodass eine Rangfolge graphisch abgebildet wird.
Kreisdiagramme („Tortendiagramm") sind eher bei Nominalskalen angemessen.

(2) Die Häufigkeiten von kontinuierlichen Intervallskalen werden durch ein Histogramm oder Polygon dargestellt.
Bei einem **Histogramm** werden über eine Intervallskala Rechtecke errichtet, wobei deren Breite der Klassenbreite und deren Höhe der jeweiligen Häufigkeit der Messwerte innerhalb der Klasse entspricht.
Ein **Polygon** (oder Polygonzug) entsteht, wenn man die Intervall- oder Klassenmitten der Histogrammrechtecke durch gerade Linien miteinander verbindet. (Oft erstellt man ein Polygon auch direkt, ohne Umweg über ein Histogramm.)

Histogramme werden bei diskreten Verteilungen benutzt, Polygone dagegen auch bei stetigen Verteilungen (bzw. bei Messungen, die prinzipiell stetige Messwerte liefern könnten, wären die Messinstrumente nur genau genug).

Beachte:
- Bei der Anfertigung eines Histogramms oder Polygons sollte man den Maßstab auf der Ordinate (senkrechte Achse) und der Abszisse (waagrechte Achse) so wählen, dass keine verzerrten oder falschen Eindrücke der Verteilungsform entstehen. Vielmehr sollte die Wahl des Maßstabes von Abzissen- und Ordinatenachse eine unverzerrte, anschauliche Interpretation der Ergebnisse ermöglichen.
Bei Ratioskalen sollten sowohl Abszisse als auch Ordinate möglichst im Nullpunkt ansetzen.
Wenn aus Platzgründen eine Abszissen- oder Ordinatenachse unterbrochen wird, sollte dies durch doppelte Trennlinien an der Achse vermerkt werden.

1.6. Typen von Häufigkeitsverteilungen in der empirischen Forschung

Aus der graphischen Darstellung einer Häufigkeitsverteilung lassen sich Aussagen über den jeweils vorliegenden Typ der Verteilung ablesen:

(1) **Modalität** der Verteilung:
- ohne Gipfel = Gleichverteilung
- eingipflig = **unimodal**
- zweigipflig = **bimodal**

(Mehrgipflige Verteilungen weisen in der Psychologie mitunter auf Fehler bei der Stichprobenauswahl hin. Man hat u.U. aus zwei Gruppen von Merkmalsträgern – z.B. einer 5. und einer 12. Schulklasse eine gemeinsame hergestellt, obwohl sich beide Gruppen bezüglich einer Variable stark unterscheiden.)

(2) Lage der Verteilung: **Symmetrie** und **Schiefe**
- symmetrisch (der Gipfel liegt in der Mitte; auf beiden Seiten des Gipfels gleich steil)
- linksschief / rechtssteil (Gipfel rechts von der Mitte)
- rechtsschief / linkssteil (Gipfel links von der Mitte)
- J-Verteilung, abfallend (Gipfel sehr stark nach einer Seite verschoben)
- U-Verteilung (entsteht durch Polarisierung)

(3) Normalverteilung
Die wahrscheinlich häufigste Verteilung in der empirischen Wissenschaft ist die **Normalverteilung** (vgl. Interferenzstatistik): Die Normalverteilung ist das mathematische Modell einer symmetrischen, unimodalen Verteilung, wobei die Häufigkeiten zum Gipfel hin stärker zu- bzw. abnehmen als bei den Rändern (Glockenform).

2. Kennwerte der eindimensionalen Häufigkeitsverteilung

Charakteristische Eigenschaften der Messwerteverteilung in einer eindimensionalen (univariaten) Häufigkeitsverteilung können mit Hilfe von statistischen Kennwerten oder Maßzahlen beschrieben werden. Besonders wichtig sind die Maßzahlen der zentralen Tendenz einer Verteilung (Lagemaße) sowie Maßzahlen, die über die Unterschiedlichkeit (Variabilität) der Merkmalsausprägungen informieren (Dispersions- oder Streuungsmaße).
Viele dieser Kennwerte / Maßzahlen sind speziell für bestimmte Skalenniveaus entwickelt worden; einige wichtige Kennwerte (z.B. Mittelwert und Standardabweichung) sind nur ab dem Skalenniveau von Intervallskalen zulässig und nur bei stetigen Messwerten sinnvoll!

2.1. Kennzahlen der zentralen Tendenz

Maße der zentralen Tendenz (Lagemaße) sind Skalenwerte, durch welche die zentrale Tendenz von Häufigkeitsverteilungen am besten ausgedrückt werden kann bzw. die Skalenwerte, welche die Messwerte eines Merkmals am besten „repräsentieren".
Abhängig vom Skalenniveau gibt es dabei verschiedene Möglichkeiten:

(1) Modalwert oder Modus

Der **Modus** (Mod) einer Verteilung ist derjenige Skalenwert, der am häufigsten auftritt.

Der Modus ist das einzig sinnvolle Maß der zentralen Tendenz bei Nominalskalen.

Es kann auch zwei oder *mehrere* Modi geben (mehrere Gipfel im Polygonzug): bimodale und polymodale Verteilungen.
Wenn mehrere Skalenwerte zu Klassen zusammengefasst wurden, gilt die Klassenmitte der häufigsten Klasse als Modalwert der Verteilung. Bei breitgipfligen Verteilungen (auf Ordinal- oder Intervallskalenniveau), bei denen mehrere Skalenwerte mit gleicher Häufigkeit nebeneinander liegen, entspricht der Modus der Mitte zwischen diesen Intervallen.

(2) Median

Der Median ist der Wert, von dem alle anderen Werte im Durchschnitt am wenigsten abweichen.

Der **Median** (Md, Mdn) einer Verteilung ist derjenige Wert einer Skala, welcher die der Größe nach geordneten Messwerte einer Stichprobe in genau zwei gleich große Hälften teilt.

Der Median ist ein Maß der zentralen Tendenz auf Ordinalskalenniveau, d.h. wenn alle Messwerte können in eine Rangfolge gebracht werden können.

Bei einer ungeraden Anzahl von Untersuchungseinheiten ist der Median genau der Messwert der Untersuchungseinheit, die in der Mitte liegt, wenn man alle Untersuchungseinheiten der Größe nach anordnet. Bei einer geraden Anzahl von Untersuchungseinheiten ist es die Mitte zwischen den beiden mittleren Fällen. Der Median kann auch mithilfe der Prozentränge ermittelt werden; er ist der Wert x_l, für den gilt:

$$\%_{kum}(x_l) = \frac{f_{kum}(x_l)}{n} \cdot 100\% = 50\ \%$$

Bei Intervallskalen mit extremen Messwerten auf einer Seite der Skala (Ausreißer) ist der Median meist eine bessere Maßzahl zur Kennzeichnung der zentralen Tendenz als das arithmetische Mittel.

(3) Das **arithmetische Mittel** ($\bar{x}$, AM) ist das gebräuchlichste Maß zur Kennzeichnung der zentralen Tendenz einer Verteilung auf Intervall- oder Rationalskalenniveau. Es wird zumeist einfach **Mittelwert** genannt.
Das arithmetische Mittel ist die Summe aller Messwerte x_i (also vom ersten Messwert i = 1 bis zum letzten Messwert i = n, vgl. das Summenzeichen) geteilt durch die Anzahl der Messwerte bzw. den Umfang der Stichprobe n.

$$AM = \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Falls bereits eine Häufigkeit $f(x_l)$ für die einzelnen Skalenwerte vorliegt, können auch die mit ihrer Häufigkeit $f(x_l)$ multiplizierten Skalenwerte x_l addiert und anschließend durch die Anzahl der Merkmalsträger n geteilt werden.

$$AM = \bar{x} = \frac{\sum_{l=1}^{m} x_l \cdot f(x_l)}{n}$$

(Vergleichbare Berechnungsformeln gelten für in Klassen gruppierte Skalenwerte.)

Das arithmetische Mittel („Mittelwert") ist die Summe der Messwerte bzw. der mit ihren Häufigkeiten multiplizierten Skalenwerte geteilt durch den Umfang der Stichprobe.

Die Position von arithmetischem Mittel (AM), Median (Mdn) und Modus (Mod) gibt Auskunft darüber, ob eine Verteilung rechtssteil, linkssteil oder symmetrisch ist:
rechtssteil/linksschief: AM < Md < Mod
linkssteil/rechtsschief: AM > Md > Mod
symmetrisch: AM = Md = Mod
Der Unterschied von Modus, Median und AM ist also ein Kennwert für die Schiefe der Verteilung.

Bei Verteilungen mit extremen Messwerten („Ausreißer") werden mitunter nur die mittleren der Messwerte zur Bildung des Mittelwertes verwendet (z.B. nur die mittleren 80 %). Die restlichen Messwerte werden entweder nicht in die Berechnung einbezogen oder aber durch den Skalenwert, der den Ausreißern am nächsten liegt, ersetzt ($\bar{x}_{10\%}$, 10%-Winsorisiertes AM).

Weitere Kennzahlen der zentralen Tendenz:

(4) Die arithmetischen Mittel verschiedener Stichprobenmessungen können zu einem **gewichteten arithmetischen Mittel** verbunden werden, wenn die einzelnen Stichproben *gleich groß* sind. Man addiert die Mittelwerte $\bar{x}_j$ und teilt die Summe durch die Anzahl der Stichproben m.

$$\bar{x}_{gesamt} = \frac{\sum_{j=1}^{m} \bar{x}_j}{m}$$

Das **gewichtete Gesamtmittel** (GAM) wird verwendet, wenn die Kennzahlen von m *verschieden großer* Stichproben zusammengefasst werden sollen. Das gewichtete arithmetische Mittel ist die Summe der mit dem jeweilgen Stichprobenumfang n_j multiplizierten Mittelwerte $\bar{x}_j$, dividiert durch die Summe aller Stichprobenumfänge.

$$GAM = \frac{\sum_{j=1}^{m} n_j \bar{x}_j}{\sum_{j=1}^{m} n_j}$$

Anmerkung: Dies ist allerdings nur möglich, wenn die Varianzen (s. Abschnitt 2.2) der einzelnen Stichproben gleich sind.

(5) Das **geometrische Mittel** (GM) kann bei der Messung von Empfindungsstärken aufgrund von psychophysischen Gesetzmäßigkeiten verwendet werden.

$$GM = \sqrt[n]{x_1 \cdot x \quad \cdot x \quad \cdot \ldots \cdot x_n}$$

2.2. Kennzahlen der Dispersion / Streuung

Auch wenn sich die zentralen Tendenzen zweier Häufigkeitsverteilungen ähneln, können die relativen Häufigkeiten der jeweiligen Skalenwerte stark voneinander abweichen.
Die Maßzahlen der Dispersion, Variation oder Streuung beschreiben diese Unterschiedlichkeit der Häufigkeitsverteilungen von Stichproben, auch wenn diese Stichproben dieselbe zentrale Tendenz (und dieselbe Größe) besitzen.

(1) Spannweite / range

Der **range** (Spannweite, Variationsbreite) ist die Differenz zwischen größten und kleinstem Messwert und gibt somit die Größe des Skalenbereichs an, in dem alle Messwerte liegen.
Range = $x_{max} - x_{min}$

Nachteil: Einzelne extreme Messwerte („Ausreißer“) erhalten unverhältnismäßig großes Gewicht. Deshalb bestimmt man die Spannweite mitunter abzüglich der Extremwerte an beiden Enden der Verteilung - z.B. nur der mittelern 90 % der Werte betrachten, abzüglich der ersten fünf (0-5%) und der letzen fünf Perzentile (95-100%).

(2) Mittlerer Quartilabstand
Mit Hilfe der kumulierten Häufigkeiten bzw. Prozentränge lässt sich eine Verteilung leicht in vier Quartile (25%-Gruppen) aufteilen. Man bestimmte den mittleren Quartilabstand als die Hälfte der *Differenz* zwischen dem Skalenwert nach 25% (P_{25}) und nach 75% (P_{75}) der Merkmalsträger.

Da sich die folgenden Dispersionsmaße auf die *Abstände* von Messwerten zum arithmetischen Mittel beziehen, sind sie nur bei Intervallskalen sinnvoll, denn nur bei Intervallskalen lassen sich die Abstände zwischen Messwerten überhaupt bestimmen.

(3) Durchschnittliche Abweichung (AD) / mittlere Abweichung (average/mean deviation)

Die mittlere/durchschnittliche Abweichung (AD) ist die Summe der Abstände aller Messwerte zum arithmetischen Mittel $\bar{x}$ geteilt durch die Anzahl der Messwerte n.

Die Abstände von Messwerte x_i und Mittelwert werden als Betrag, d.h. mit positivem Vorzeichen, geschrieben. Jeder einzelne Messwert geht in das Streuungsmaß ein!

$$AD = \frac{\sum_{i=1}^{n} \left| x_i - \bar{x} \right|}{n}$$

Liegen die Häufigkeiten der einzelnen Messwerte x_l vor, so kann man auch rechnen (m bezieht sich auf die Anzahl der Skalenwerte x_l):

$$AD = \frac{\sum_{l=1}^{m} f(x_l) \cdot \left| x_l - \bar{x} \right|}{n}$$

(4) Varianz (quadratische Abweichung) und Standardabweichung

Da es mitunter wünschenswert ist, besonders große Abweichungen der Messwerte vom Mittelwert besonders stark gewichtet in das Streuungsmaß einzubringen, kann man auf die Quadrate der Abweichungen zurückgreifen. Durch die Quadrierung bekommen größere Abweichungen ein größeres Gewicht. Außerdem werden durch die Quadrierung automatisch alle Abstände positiv, d.h. man muss nicht – wie beim AD – die Beträge der Abstände addieren.

Die **Varianz s^2** (quadratische Abweichung) ist die Summe der Quadrate aller Abstände geteilt durch die Anzahl der Messwerte, bzw.: das arithmetische Mittel der Summe aller quadrierten Abstände zwischen Messwerten und arithmetischen Mittel.

$$s^2 = \frac{\sum_{i=1}^{n} \left(x_i - \bar{x} \right)^2}{n}$$

Auch hier kann natürlich auf die Häufigkeiten der Skalenwerte x_l zurückgegriffen werden:

$$s^2 = \frac{\sum_{l=1}^{m} \left(x_l - \bar{x} \right)^2 \cdot f(x_l)}{n}$$

Die **Standardabweichung s** (SD, „Streuung“ im engeren Sinne) ist die Wurzel aus der Varianz.

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^{n} \left(x_i - \bar{x} \right)^2}{n}}$$

Auch hier kann natürlich auf die Häufigkeiten der Skalenwerte x_l zurückgegriffen werden:

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{l=1}^{m}(x_l - \bar{x})^2 \cdot f(x_l)}{n}}$$

Die Varianz und die Standardabweichung sind die gebräuchlichsten Kennwerte zur Beschreibung für die Dispersion einer Häufigkeitsverteilung in der empirischen Forschung; wobei die Standardabweichung s handlicher ist, weil sie auf derselben Einheit wie die Messwerte - und nicht auf deren Quadrate - beruht. Allerdings sind Varianz und Standardabweichung nur bei unimodalen und (zumindest annährend) symmetrischen Verteilungen wirklich aussagekräftig.

Bei den obigen Formeln wird die Standardabweichung in zwei Rechenschritten bestimmt werden: Zunächst muss für jeden Messwert die Abweichung zum Mittelwert berechnet und erst anschließend kann der SD-Wert selbst bestimmt werden. Dies lässt sich durch folgende Formel vereinfachen (zur Herleitung vgl. Bortz 1999, S. 44).

$$s = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n} - \bar{x}^2}$$

Nebenbemerkungen:
1. In den Formeln für Varianz und Standardabweichung, wie sie sich in vielen Statistik-Lehrbüchern findet, wird nicht durch n sondern durch (n-1) geteilt:

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} \quad \text{und} \quad s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

Dies wird jedoch erst in der Inferenzstatistik notwendig, um „erwartungstreue" Kennwerte für die Schätzung der Streuung in unbekannten Grundgesamtheiten zu erhalten.

2. Bedeutung der Standardabweichung
Für eine Normalverteilung gilt, dass zwischen den Skalenwerten $\bar{x}$ + s und $\bar{x}$ - s genau 68,26 % (also ca. zwei Drittel) aller Merkmalsträger liegen. Im Bereich von $\bar{x} \pm 2s$ befinden sich 95,44 % der Merkmalsträger. Diesen Umstand macht man sich in der Inferenzstatistik zunutze.

(5) **Variationskoeffizient**
Gelegentlich wird als ein weiteres Dispersionsmaß der Varitionskoeffizient (V, auch: Variabilitätskoeffizient) verwendet. Er erlaubt es, Streuungen von Verteilungen mit unterschiedlichen Mittelwerten miteinander zu vergleichen. Der Varitionskoeffizient relativiert die

Standardabweichung an Mittelwert bzw. drückt die Standardabweichung in Mittelwerteinheiten aus (nur bei $\bar{x}$ >0):

$$V = \frac{s}{\bar{x}}$$

Der Variationskoeffizient ist allerdings nur bei Ratio-Skalen mit einem natürlichen Nullpunkt sinnvoll, weil nur dann der absolute Betrag des Mittelwerts feststeht und sich die Verteilung (mitsamt des Mittelwertes) nicht kontinuierlich verschieben lässt.

2.3. Vergleich von Messwerten aus verschiedenen Messungen: z-Werte

Wenn die Merkmalsausprägungen von Personen, die aus verschiedenen Stichproben stammen (z.B. Testnoten von Schülern aus verschiedenen Schulklassen), miteinander bezüglich der Position ihrer Merkmalsausprägung innerhalb der jeweiligen Stichprobe verglichen werden sollen, ist es sinnvoll, den Messwert jeder Person auf alle Werte der betreffenden Messung (im Beispiel: der jeweiligen Schulklasse, aus der ein Schüler stammt) zu beziehen. Individuelle Ausprägungen (z.B. schulische Leistungen) werden also vor dem direkten Vergleich am jeweiligen Kollektiv relativiert. Hierzu bestehen zwei Möglichkeiten:
(1) Durch einen Vergleich der Prozentränge der Merkmalsträger: Auf welchem Platz in der Rangordnung befindet sich die Person innerhalb des Kollektivs / der Stichprobe?
(2) Oder durch eine Relativierung der Messwerte an der Standardabweichung im jeweiligen Kollektiv. Die Abweichung des jeweilige Messwerts zum Mittelwert $\bar{x}$ wird durch die Standardabweichung der betreffenden Stichprobe dividiert. Führt man diese Transformation an allen Messwerten durch, so erhält man die sog. **z-Standardwerte** einer Verteilung.

$$z_i = \frac{x_i - \bar{x}}{s}$$

> Um die Abweichungen der Messwerte vom Mittelwert bei unterschiedlichen Stichproben besser miteinander vergleichen zu können, müssen sie zuvor an der Streuung aller Werte in der jeweiligen Stichprobe relativiert werden. Dies geschieht, indem die Abweichungen der Messwerte zum arithmetischen Mittel der Stichprobe durch die Standardabweichung der Stichprobe dividiert wird (**z-Transformation**).

Die z-Standardwerte sind nichts anderes als in Standardabweichungen gemessene Abweichungseinheiten unter bzw. über dem Mittelwert. Z-Standardardwerte informieren über die relative Position eines Individuums in einem Kollektiv.

Werden alle Messwerte einer normalverteilten Stichporbe z-transformiert, so erhalten wir eine sog. Standard-Normalverteilung, bei der gilt:

Eine z-transformierte Verteilung hat einen Mittelwert $\bar{x}$ = 0 und eine Standardabweichung von s = 1.

Beachte:
1. Die Varianz bzw. Standardabweichung ist nur bei normalverteilten Stichproben aussagekräftig. Diese Einschränkung gilt deshalb auch für die Vergleichbarkeit von z-Standard-Werten aus verschiedenen Stichproben!
2. Messungen, die sich aus inhaltlichen Gründen nicht vergleichen lassen, sind natürlich auch nach einer z-Transformation nicht miteinander vergleichbar!