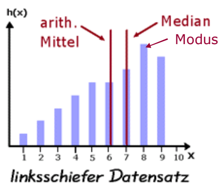
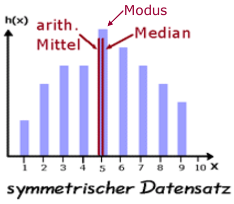
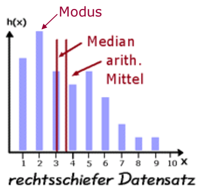
Maße zur Kennzeichnung der Form einer Verteilung (1)

- Schiefe (*skewness*): Definition I -

Eine Verteilung von Messwerten wird als schief bezeichnet, wenn sie in der Weise asymmetrisch ist, dass links oder rechts des Durchschnitts eine Häufung der Messwerte auftritt. Bei linksschiefen Verteilungen sind Modus und Median größer als das arithmetische Mittel, bei rechtsschiefen kleiner.







AM liegt links vom Md
AM<Md<Mo

AM ≈ Md ≈ Mo*

AM liegt rechts vom Md
Mo<Md<AM

Begriffssynonyme:

- rechtsgipflig
- rechtssteil
- Verteilung mit negativer Schiefe („die Differenz AM-Md ist negativ")

* Es gibt symmetrische Verteilungen in denen Mo≠AM bzw. Md

- linksgipflig
- linkssteil
- Verteilung mit positiver Schiefe („die Differenz AM-Md ist positiv")

Maße zur Kennzeichnung der Form einer Verteilung (2)

- Schiefe (*skewness*): Definition II -

- Der Kennwert der Schiefe (*skewness*) ist definiert als der Durchschnittswert der dritten Potenz z-transformierter Messwerte (drittes Moment der Verteilung).

$$Skewness = AM(z_i^3) = \frac{\sum_{i=1}^{n} z_i^3}{n}$$

- Bei symmetrischen Verteilungen ist die Schiefe = 0.
- Bei linearen Transformationen (y=bx+a) kommt es zu Rechts- bzw. Links-Verschiebungen auf der X-Achse um die Konstante a und zu Stauchungen (0<b<1) bzw. Verbreiterungen (b>1) der Verteilung, nicht aber zu einer Veränderung der Schiefe.
- Bei negativer Konstante b (b<0), spricht man von einer Umpolung.
  In diesem Fall ändert sich lediglich das Vorzeichen der Schiefe.

Maße zur Kennzeichnung der Form einer Verteilung (3)

- Wölbung (*kurtosis*): Definition -

- Die Wölbung (auch Exzess genannt) ist ein Maß für die Steilgipfligkeit einer Verteilung.
- Es gibt zwei Formeln für die Wölbung. Am häufigsten wird die unten angegebene Formel benutzt, bei der vom Durchschnitt der vierten Potenz der z-transformierten Werte (viertes Moment der Verteilung) der Wert 3 subtrahiert wird (dies entfällt bei der anderen Formel):

$$Kurtosis = AM(z_i^4) - 3 = \frac{\sum_{i=1}^{n} z_i^4}{n} - 3$$

Dann gilt:

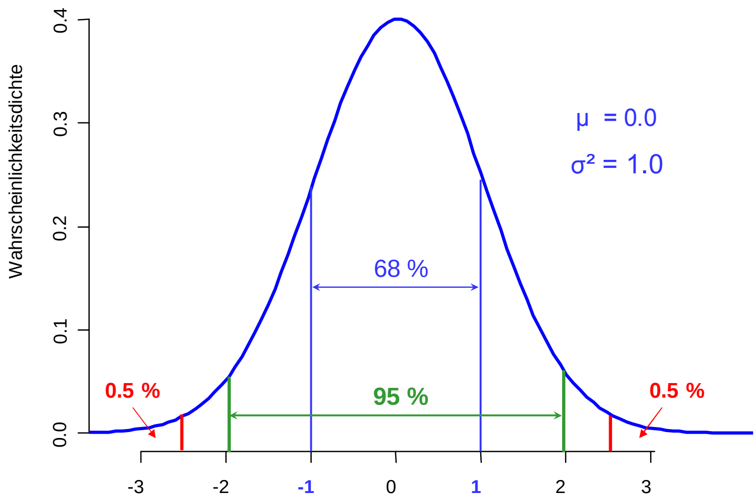
- Bei einer positiven Wölbung ist die Verteilung schmalgipfliger als eine Normalverteilung.
- Bei einer negativen Wölbung ist die Verteilung breitgipfliger (plateauartiger) als die Normalverteilung.

Für beide Formeln gilt:

- Gleichverteilungen haben eine niedrige Wölbung.
- Bei linearen Skalentransformationen ändert sich die Wölbung nicht.

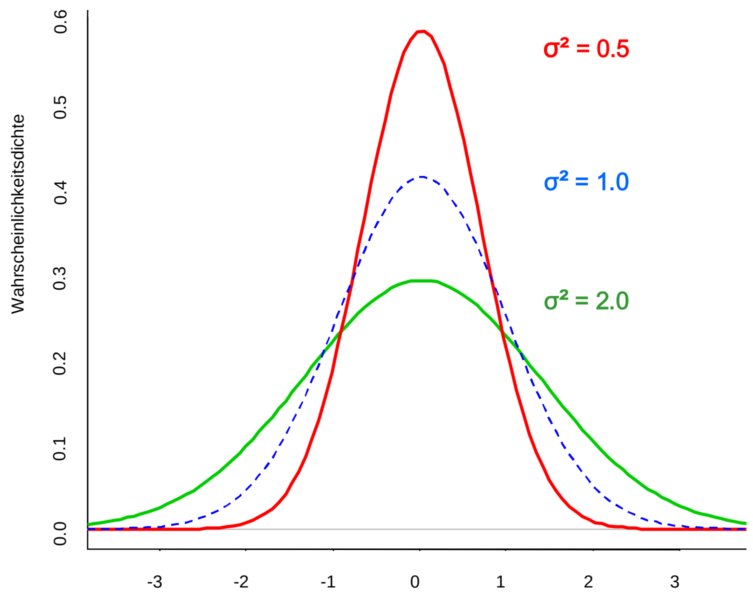
Unterschiedliche Verteilungsformen (1)

- Die Standardnormalverteilung und ihre Flächenanteile-



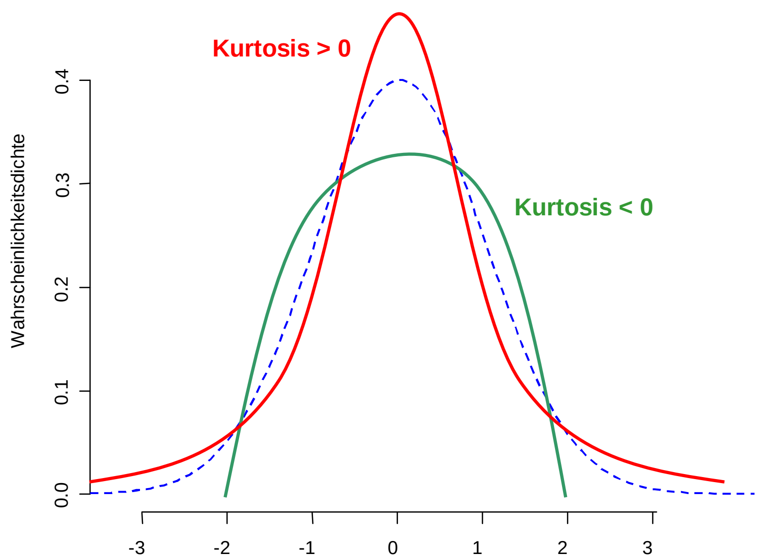
Unterschiedliche Verteilungsformen (2)

- Drei Normalverteilungen mit unterschiedlichen Varianzen -



Unterschiedliche Verteilungsformen (3)

- Illustration positiver und negativer Kurtosis (Wölbung) -



Zusammenfassung Verteilungsformen (1)

Anmerkung: Im Folgenden wird immer von der oben angegebenen Formel (mit -3) für die *kurtosis* ausgegangen.

- Kennziffern der Verteilungsform sind Schiefe (*skewness*) und Wölbung (*kurtosis*).
- Die Maße der *skewness* und der *kurtosis* ändern sich bei Lineartransformationen der Werte nicht.
- Ist das Maß der ***skewness* positiv**, wird die Verteilung als **rechtsschief** (oder linkssteil) bezeichnet. Dabei gilt: Modus < Median < arithmetisches Mittel.
- Ist das Maß der ***skewness* negativ**, wird die Verteilung als **linksschief** (oder rechtssteil) bezeichnet. Dabei gilt: arithmetisches Mittel < Median < Modus.
- Die Flächenanteile unter der Kurve der **Standardormalverteilung** (μ = 0.0, σ = 1.0) sind:

```
[-1.00 σ, +1.00 σ] = 68.3%
[-1.96 σ, +1.96 σ] = 95.0%
[-2.58 σ, +2.58 σ] = 99.0%
```

- Verteilungskurven normalverteilter Werte mit **Standardabweichungen <> 1.0** schneiden die Kurve der Standardnormalverteilung auf jeder Seite nur an **einer** Stelle. Durch z-Standardisierung der Werte lässt sich jede Normalverteilung in die Standardnormalverteilung überführen.

Zusammenfassung Verteilungsformen (2)

- *Kurtosis* und Varianz sind unabhängig von einander.
- Ist das Maß der ***kurtosis* <> 0**, schneidet die Verteilungskurve die Kurve der Standardnormalverteilung auf jeder Seite (links und rechts) an **zwei** Stellen.
- Ist das Maß der ***kurtosis* positiv**, verläuft die Verteilungskurve steiler als die Kurve der Standardnormalverteilung, dabei hat die Verteilungskurve „schwächere Schultern“: Es liegt mehr „Masse“ im Zentrum und an den Enden der Verteilung.
- Ist das Maß der ***kurtosis* negativ**, verläuft die Verteilungskurve flacher als die Kurve der Standardnormalverteilung, dabei hat die Verteilungskurve „stärkere Schultern“: Es liegt weniger „Masse“ im Zentrum und an den Enden der Verteilung.
- **Normalverteilte** Werte haben eine ***kurtosis* von 0.0**, **gleichverteilte** eine ***kurtosis* von -1.2**; **zweigipflig verteilte** eine ***kurtosis* < -1.2**.
- Bei nicht normalverteilten Werten (*skewness* <> 0 und/oder *kurtosis* <> 0) ist die gängige Interpretation der Flächenanteile anhand der Standardabweichung und des arithmetischen Mittels als Maß der zentralen Tendenz möglicherweise unangemessen!

Nachtrag (1)

- Stichprobenstatistiken und geschätzte Populationsparameter -

- Die bisher dargestellten Formeln der Schiefe (*skewness*) und Wölbung (*kurtosis*) beschreiben die Verteilungsform der Stichprobe, und die in ihnen benutzten *z*-Werte werden anhand der Standardabweichung der Stichprobenwerte berechnet, wobei gilt:

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

- Gebräuchlicher (und in den meisten Statistikprogrammen verwendet) sind allerdings Formeln zur Berechnung der geschätzten Populationsparameter der Schiefe und der Wölbung (die Schätzer werden auch als unverzerrt bzw. *unbiased* bezeichnet), wobei die Populationsparameter aus den Stichprobendaten geschätzt werden. Der wesentliche Unterschied besteht in Korrekturfaktoren, in denen die Zahl für die Stichprobengröße „korrigiert" wird. Deshalb werden auch die in den Formeln benutzten *z*-Werte anhand der geschätzten Populationsstandardabweichung berechnet, wobei gilt:

$$z_i = \frac{x_i - \bar{x}}{\hat{\sigma}} \quad \text{mit} \quad \hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{(n-1)}}$$

Nachtrag (2)

- Stichprobenstatistiken und geschätzte Populationsparameter -

- Die Formel für den aus den Stichprobendaten geschätzten Populationsparameter der Schiefe (*skewness, unbiased*) unter Verwendung der anhand der geschätzten Populationsstandardabweichung berechneten *z*-Werte ist:

$$Skewness_{unbiased} = \frac{n}{(n-1)\cdot(n-2)} \cdot \sum_{i=1}^{n} z_i^3$$

- Die Formel für den aus den Stichprobendaten geschätzten Populationsparameter der Wölbung (*kurtosis, unbiased*) unter Verwendung der anhand der geschätzten Populationsstandardabweichung berechneten *z*-Werte ist:

$$Kurtosis_{unbiased} = \frac{n\cdot(n+1)}{(n-1)\cdot(n-2)\cdot(n-3)} \cdot \sum_{i=1}^{n} z_i^4 - 3 \cdot \frac{(n-1)^2}{(n-2)\cdot(n-3)}$$

- Je größer die Stichprobe ist, um so weniger unterscheiden sich die Werte der Stichprobenstatistiken und der geschätzten Populationsparameter.