§2 Daten graphisch darstellen pfad

▼

```
SetDirectory[
   "C:\Dokumente und Einstellungen\Weiss\Desktop\Stochastik mit Mathematica 7.0\Statistik\Datenordner"];

<< StatisticalPlots`;
<< MultivariateStatistics`;

RawDataPlot[stich_, options___] := Module[{a, b, c, d, e, plotl, plot2, plot3},
    a = Min[stich];
    b = Max[stich];
    c = N[Median[stich], 4];
    d = N[Mean[stich], 4];
    e = (b − a) / 5;
    plot1 = ListPlot[stich + (b − a) + e, PlotStyle → {Red, PointSize[0.015]}];
    plot2 = ListPlot[Sort[stich], PlotStyle → PointSize[0.015]];
    plot3 = Graphics[{Gray, Line[{{1, b + e / 2}, {Length[stich], b + e / 2}}]}];
    Show[{plot3, plot1, plot2}, AspectRatio → 0.6, PlotRange → All, Axes → False, Frame → True,
      FrameLabel → {None, "   sortiert            nicht  sortiert"},
      FrameTicks → {Automatic, {a, c, b, {b + e, a}, {(b − a) + d + e, d}, {2 b − a + e, b}}, None, None}, options]]

MovingHistogram[stich_, b_, bereich_, options___] := Module[{x, xmin, xmax, table, plot1, plot2},
    x = bereich[[1]];
    xmin = bereich[[2]];
    xmax = bereich[[3]];
    table = Table[{x, BinCounts[stich, {{x − b / 2, x + b / 2}}][[1]]}, {x, xmin, xmax, b / 50}];
    plot1 = ListPlot[table, Joined → True, Axes → True, Ticks → {Automatic, None}, options];
    plot2 = Graphics[{Thickness[0.02], Line[{{xmin, 0}, {xmin + b, 0}}]}];
    Show[plot1, plot2]]

ScatterPlot[matrix_, options___] := Module[{ax, ay, bx, plot1, plot2},
      ax = Min[Part[matrix, All, 1]];
      ay = Min[Part[matrix, All, 2]];
      bx = Max[Part[matrix, All, 1]];
      plot1 = ListPlot[matrix, options];
      plot2 = Plot[Evaluate[Fit[matrix, {1, x}, {x}]], {x, ax, bx}, PlotStyle → {Thickness[0.01], Red}];
      Show[{plot1, plot2}]];

Kontingenztafel[matrix_, options___] := Module[{li1, li2, l12, kon},
    li1 = Union[Part[matrix, All, 1]];
    li2 = Union[Part[matrix, All, 2]];
    l12 = Flatten[Outer[List, li1, li2], 1];
    kon = Partition[Table[Count[matrix, l12[[i]]], {i, 1, Length[l12]}], Length[li2]];
    TableForm[kon, TableHeadings → {li1, li2}, options]]

Balkendiagramm3D[matrix_, options___] := Module[{li1, li2, l12, kon},
    li1 = Union[Part[matrix, All, 1]];
    li2 = Union[Part[matrix, All, 2]];
    l12 = Flatten[Outer[List, li1, li2], 1];
    kon = Partition[Table[Count[matrix, l12[[i]]], {i, 1, Length[l12]}], Length[li2]];
    BarChart[Transpose[kon], ChartLabels → {li2, li1}, options]]
```

Wenn statistische Daten sehr umfangreich sind, erweist es sich als hilfreich, diese Daten graphisch zu *veranschaulichen. Man bekommt dadurch einen wesentlich besseren Eindruck von diesen Daten. In vielen Fällen erkennt man dabei auch Merkwürdigkeiten, die sonst verborgen bleiben würden - und genau diese Merkwürdigkeiten geben in vielen Fällen Anlass für weitere interessante Fragestellungen.*

Wir werden in diesem Abschnitt einige der vielen Befehle und Möglichkeiten besprechen, welche Mathematica zur Verfügung stellt, um Daten graphisch zu veranschaulichen. Man vergleiche dazu die ausführliche Beschreibung aller Möglichkeiten, welche sich in ChartingAndInformationVisualization findet.

2.1 Graphische Darstellung univariater Daten

Wir gehen davon aus, dass unser statistisches Datenmaterial *daten* bereits in Form einer Datenmatrix, welche im Datenfile *datenfile* unseres Datenordners abgelegt ist, vorliegt und befassen uns in diesem Abschnitt mit der Frage, wie sich eine Stichprobe dieses Datenmaterials (also eine Spalte der Datenmatrix ohne dem zugehörigen Namen) graphisch veranschaulichen lässt. Dabei haben wir zu unterscheiden, ob diese Stichprobe zu einem quantitativen oder einem qualitativen Merkmal gehört.

2.1.1 Bemerkung: Ein *quantitatives* Merkmal lässt sich durch ein Rohdaten-Plot, ein Histogramm bzw ein gleitendes Histogramm sowie durch ein Box-Plot graphisch veranschaulichen; ein *qualitatives* Merkmal kann durch ein Balkendiagramm bzw ein Tortendiagramm graphisch dargestellt werden.

■ Rohdaten-Plot

Um einen ersten Eindruck der zu analysierenden Stichprobe *stich* eines *quantitativen* Merkmals zu erhalten, verwendet man häufig ein sogenanntes Rohdaten-Plots. Ein derartiges Plot besteht aus zwei Hälfte. In der oberen Hälfte werden die einzelnen Werte der darzustellenden Stichprobe sequentiell von links nach rechts aufgetragen, in der unteren Hälfte werden die Werte dieser Stichprobe zuerst der Größe nach geordnet und dann ebenfalls sequentiell von links nach rechts aufgetragen. Die Ordinaten der Datenpunkte entsprechen dabei den jeweiligen Werten der darzustellenden Stichprobe. Mit Hilfe eines derartigen Plots lässt sich festellen, ob Ausreißer vorhanden sind bzw ob in der Stichprobe ein Trend zu bemerken ist.

Zur graphischen Darstellung der Stichprobe *stich* eines *quantitativen* Merkmals durch ein Rohdaten-Plot dient der Befehl RawDataPlot (hinsichtlich der Begriffe Median und Mittelwert vergleiche man Definition 3.1.1):

■ RawDataPlot[*stich*]

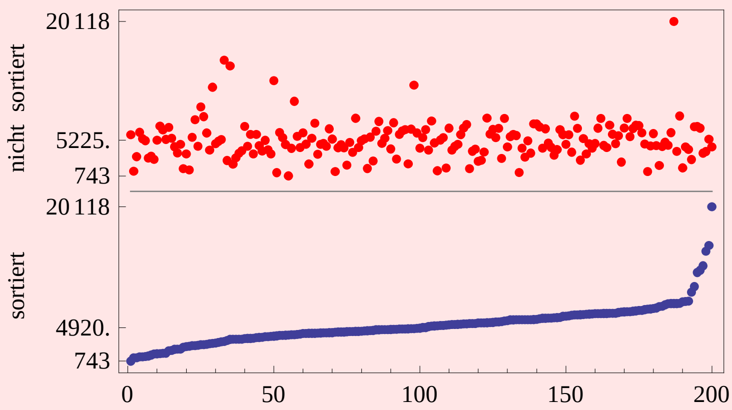
erzeugt für die Stichprobe *stich* ein Rohdaten-Plot. Neben dem Minimum und dem Maximum sind am linken Rand bei den nicht sortierten Daten der Mittelwert und bei den sortierten Daten der Median der Stichprobe *stich* angeführt.

2.1.2 Beispiel: Vom Datenmaterial *einkommen* erstelle man ein Rohdaten-Plot für das monatliche Brutto-Einkommen der Männer.

▼

Lösung: Mit Hilfe von Cases und Part wählen wir von dem im Datenordner abgelegten Datenfile *einkommenfile* das monatliche Brutto-Einkommen der Männer aus, nennen diese Stichprobe "männer" und zeichnen von dieser Stichprobe ein Rohdaten-Plot:

```
männer = Part[Cases[<< einkommenfile, {m, x_}], All, 2];
RawDataPlot[männer, ImageSize → {300, 150}]
Clear[männer]
```



An Hand dieses Plots lässt sich ablesen: Das minimale Einkommen der befragten Männer beträgt 743 €; ihr maximales Einkommen beträgt 20118 €; durchschnittlich verdienen die befragten Männer 5225 €. Außerdem zeigt dieses Plot, dass es einige wenige Männer gibt, welche ein deutlich höheres Einkommen beziehen als der Rest. Diese Asymmetrie wird auch dadurch deutlich, dass der Median mit 4920 € deutlich niedriger ist als der Mittelwert 5225 €. Die Punkte in der oberen Hälfte des Plots weisen keinen Trend auf. Dies lässt darauf schließen, dass bei der Datenaufnahme die Männer in zufälliger Reihenfolge befragt wurden.

Histogramm

Zur graphischen Darstellung einer oder mehrerer Stichproben eines *quantitativen* Merkmals durch ein Histogram dient der Befehl Histogram.

■ Histogram[{*stich*1, *stich*2, …}]

zeichnet Histogramme der Stichproben *stich*1, *stich*2, … in eine gemeinsame Zeichnung, wobei der Wertebereich der Stichproben automatisch in gleich lange Intervalle eingeteilt wird.

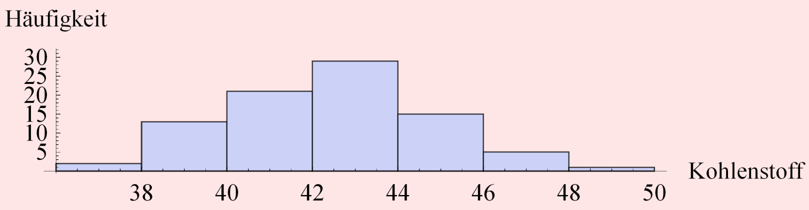
Mit der Option n bzw $\{\{b_1, b_2, \ldots\}\}$ wird der Wertebereich der Stichproben in n gleich lange Intervalle eingeteilt bzw werden nur die Häufigkeiten der Intervalle $[b_i, b_{i+1}[$ graphisch dargestellt.

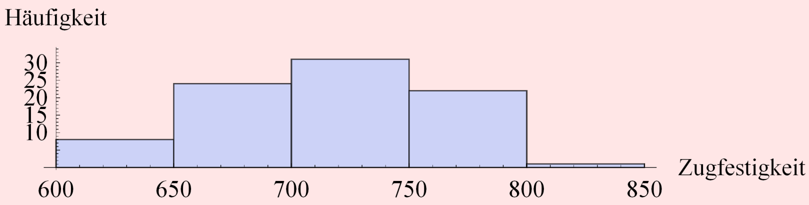
2.1.3 Beispiel: Vom Datenmaterial *stahl* erstelle man Histogramme für die beiden quantitativen Merkmale Kohlenstoff (zweite Spalte) bzw Zugfestigkeit (dritte Spalte).

▼

Lösung: Wir lesen dazu das im Datenordner abgelegte Datenfile *stahlfile* ein, rufen mit Hilfe von Part die zweite bzw dritte Spalte dieser Datenmatrix auf und bezeichnen die zugehörigen Stichproben mit "kohlenstoff" bzw "zugfestigkeit". Von diesen Stichproben zeichnen wir Histogramme:

```
kohlenstoff = Rest[Part[<< stahlfile, All, 2]];
zugfestigkeit = Rest[Part[<< stahlfile, All, 3]];
Histogram[kohlenstoff, AxesLabel → {"Kohlenstoff", "Häufigkeit"}, AspectRatio → 0.2, ImageSize → {300, 80}]
Histogram[zugfestigkeit, AxesLabel → {"Zugfestigkeit", "Häufigkeit"}, AspectRatio → 0.2,
  ImageSize → {300, 80}]
Clear[kohlenstoff, zugfestigkeit]
```



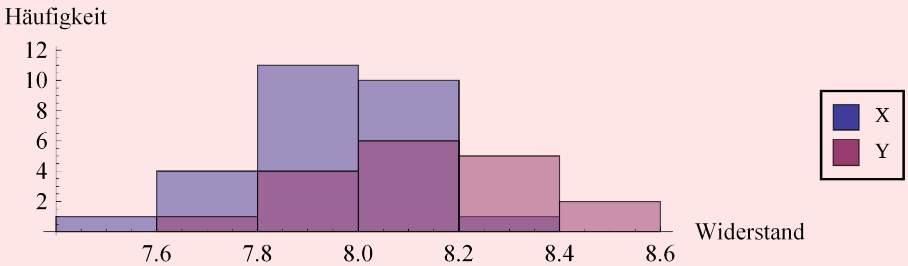


2.1.4 Beispiel: Vom Datenmaterial *widerstand* erstelle man ein Histogramm für das quantitative Merkmal Widerstand (erste Spalte), wobei wir nur jene Fälle beachten, in denen der Widerstand auf der Fertigungsstraße X bzw auf der Fertigungsstraße Y hergestellt wurde.

▼

Lösung: Wir lesen dazu das im Datenordner abgelegte Datenfile *widerstandfile* ein, erzeugen unter Verwendung von Cases und Part Stichproben der auf den Fertigungsstraßen X bzw Y erzeugten Widerstände und bezeichnen diese Stichproben mit "straßeX" bzw "straßeY". Von diesen beiden Stichproben zeichnen wir Histogramme.

```
straßeX = Part[Cases[<< "widerstandfile", {x_, X}], All, 1];
straßeY = Part[Cases[<< "widerstandfile", {y_, Y}], All, 1];
Histogram[{straßeX, straßeY}, AxesLabel → {"Widerstand", "Häufigkeit"}, AxesOrigin → {7.4, 0},
  ChartLegends → {"X", "Y"}, AspectRatio → 0.3, ImageSize → {300, 100}]
Clear[straßeX, straßeY]
```



Wir erkennen an Hand dieses Histogramms, dass die auf der Fertigungsstraße Y erzeugten Widerstände tendenziell einen etwas höheren Widerstandswert aufweisen, als die auf der Fertigungsstraße X erzeugten Widerstände. Ob diese Tatsache "statistisch signifikant" ist oder allein durch zufällige Schwankungen erklärbar ist, wird Gegenstand weiterer Untersuchungen sein.

■ Gleitendes Histogramm

Beim klassischen Histogramm werden die Werte der Stichprobe in starre Intervalle eingeteilt. Dadurch kommt es in Abhängigkeit von der jeweiligen Intervalleinteilung zu unterschiedlichen Histogrammen. Außerdem treten oft "unschöne" Sprünge zwischen den einzelnen Balken auf.

Beim gleitenden Histogramm werden - wie im klassischen Histogramm - ebenfalls die Anzahl der Werte innerhalb eines Intervalls der Breite b ermittelt, allerdings gleitet dieses Intervall "kontinuierlich" über den gesamten Datensatz. Die zum Wert x gehörende Ordinate entspricht dabei der Häufigkeit des Intervalls $[x - b/2, x + b/2[$.

Je breiter dieses gleitende Intervall gewählt wird, um so "glatter" wird die Kurve. Allerdings wird ihre Aussagekraft mit zunehmender Intervallbreite immer geringer. Die Kunst, gleitende Histogramme richtig einzusetzen, liegt somit in der Wahl einer geeigneten Intervallbreite.

Zur graphischen Darstellung der Stichprobe *stich* eines *quantitativen* Merkmals durch ein gleitendes Histo-gramm dient der Befehl MovingHistogram:

■ MovingHistogram[*stich*, *b*, {*x*, *xmin*, *xmax*}]

zeichnet ein gleitendes Histogramm der Stichprobe *stich*, wobei das über die Stichprobe "gleitende" Intervall die Breite b besitzt (diese Breite wird durch den schwarzen Balken links unten symbolisiert) und der Mittelpunkt dieses Intervalls mit einer Schrittweite von $b/50$ über das Intervall [*xmin*, *xmax*] gleitet.

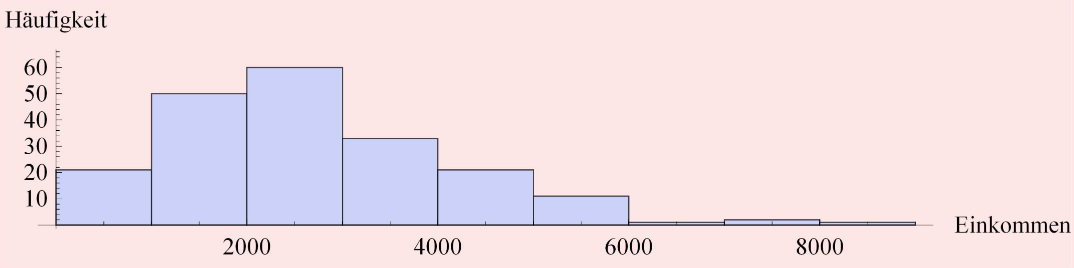
Man beachte, dass beim gleitenden Histogramm nur die Form der Kurve von Bedeutung ist - daher findet man auf der y-Achse auch keine Skaleneinteilung. Mit den üblichen Plot-Optionen kann die Zeichnung den eigenen Bedürfnissen angepasst werden.

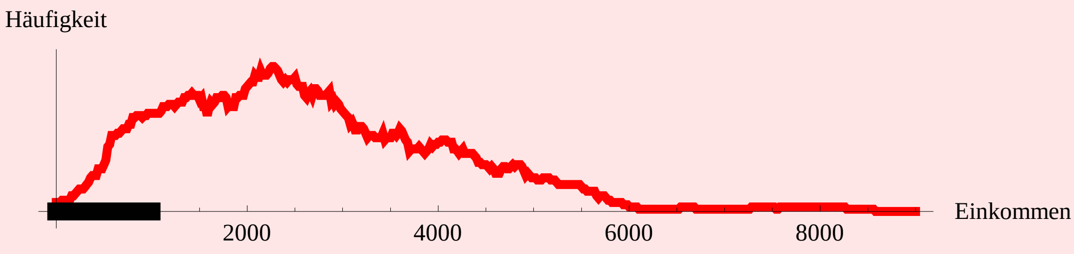
2.1.5 Beispiel: Vom Datenmaterial *einkommen* erstelle man ein Histogramm sowie ein gleitendes Histo-gramm für das monatliche Brutto-Einkommen der Frauen.

▼

Lösung: Mit Hilfe von Cases und Part wählen wir von dem im Datenordner abgelegten Datenfile *einkommenfile* das monatliche Brutto-Einkommen der Frauen aus, bezeichnen diese Stichprobe mit "frauen" und veranschaulichen diese Stichprobe sowohl durch ein Histogramm als auch durch ein gleitendes Histogramm:

```
frauen = Part[Cases[<< "einkommenfile", {w, y_}], All, 2];
Histogram[frauen, AspectRatio → 0.2, AxesLabel → {"Einkommen", "Häufigkeit"}, ImageSize → {400, 100}]
MovingHistogram[frauen, 1000, {x, 0, 9000}, PlotStyle → {Thickness[0.01], Red}, AspectRatio → 0.2,
  AxesLabel → {"Einkommen", "Häufigkeit"}, ImageSize → {400, 100}]
Clear[frauen]
```





■ Box-Plot

Zur graphischen Darstellung einer oder mehrerer Stichproben eines *quantitativen* Merkmals durch ein Boxplot dient der Befehl BoxWhiskerPlot (dieses Plot erinnert an die Barthaare einer Katze, was den etwas eigenartigen Namen erklärt). Hinsichtlich der dabei verwendeten Begriffe Median und Quantil bzw Quartil vergleiche man Definition 3.1.1 und Definition 3.1.5.

■ BoxWhiskerPlot[*stich*1, *stich*2, …] bzw BoxWhiskerPlot[*matrix*]

zeichnet Box-Plots der Stichproben *stich*1, *stich*2, … in eine gemeinsame Zeichnung. Der mittlere Strich gibt den Median, der untere Strich das Minimum und der obere Strich das Maximum der jeweiligen Stichprobe an. Der mittlere Balken entspricht dem sogenannten Interquartile-Range (also dem Bereich zwischen dem 25%-Quartil und dem 75%-Quartil) der Stichprobe. Die alternative Form dieses Befehls wird dann verwendet, wenn die Stichproben *stich*1, *stich*2, … die Spalten der Matrix *matrix* bilden.

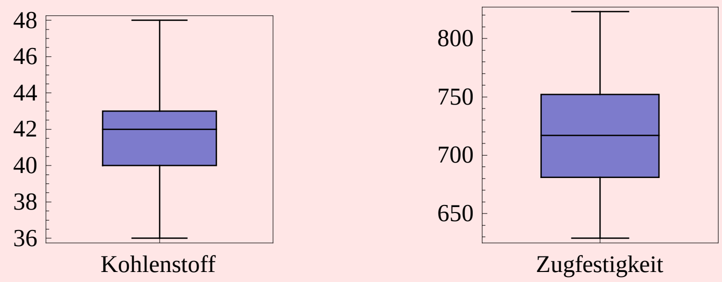
Mit der Option BoxOutliers → All werden Ausreißer markiert; mit der Option BoxOutliers → Automatic in Verbindung mit der Option BoxOutlierMarkers lassen sich gewöhnliche Ausreißer und extreme Ausreißer in verschiedener Weise markieren. Unter einem Ausreißer versteht man dabei einen Wert, welcher außerhalb des 3/2-fachen des Interquartile-Range liegt; unter einem extremen Ausreißer versteht man einen Wert, der außerhalb des 3-fachen des Interquartile-Range liegt. Ausreißer, welche keine extremen Ausreißer sind, nennt man gewöhnliche Ausreißer. Mit der Option BoxLabels können die einzelnen Box-Plots benannt werden.

2.1.6 Beispiel: Vom Datenmaterial *stahl* erstelle man Box-Plots für die quantitativen Merkmale Kohlenstoff (zweite Spalte) bzw Zugfestigkeit (dritte Spalte).

▼

Lösung: Wir lesen dazu das im Datenordner abgelegte Datenfile *stahlfile* ein, rufen mit Hilfe von Part die zweite bzw dritte Spalte dieser Datenmatrix auf und bezeichnen die zugehörigen Stichproben mit "kohlenstoff" und "zugfestigkeit". Von diesen Stichproben zeichnen wir Box-Plots:

```
kohlenstoff = Rest[Part[<< "stahlfile", All, 2]];
zugfestigkeit = Rest[Part[<< "stahlfile", All, 3]];
plot1 = BoxWhiskerPlot[kohlenstoff, BoxLabels → "Kohlenstoff", ImageSize → {150, 100}];
plot2 = BoxWhiskerPlot[zugfestigkeit, BoxLabels → "Zugfestigkeit", ImageSize → {150, 100}];
GraphicsGrid[{{plot1, plot2}}]
Clear[kohlenstoff, zugfestigkeit, plot1, plot2]
```

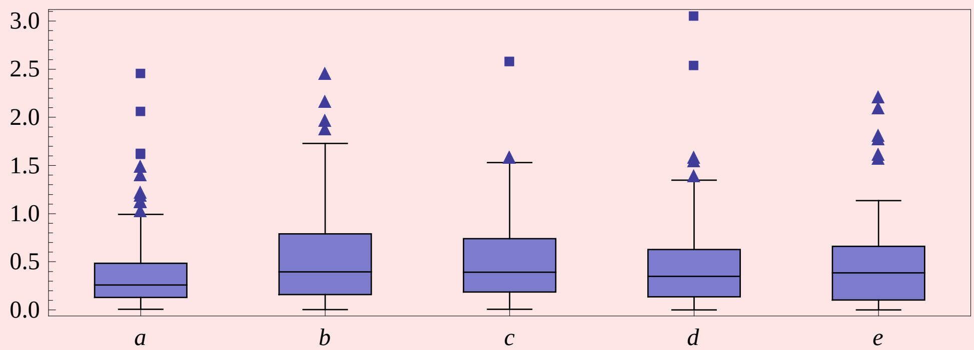


2.1.7 Beispiel: Durch Simulation erzeuge man $k = 5$ Stichproben von jeweils $n = 100$ mit dem Parameter $\lambda = 2$ exponentialverteilten Zufallszahlen, zeichne die Box-Plots dieser Stichproben in eine gemeinsame Zeichnung und versehe die einzelnen Box-Plots mit den Namen a, b, c, d, e.

▼

Lösung: Wir erzeugen zuerst in der bekannten Weise das darzustellende Datenmaterial "daten". Es handelt sich dabei um eine Matrix mit n Zeilen und k Spalten von mit dem Parameter λ exponentialverteilten Zufallszahlen. Anschließend zeichnen wir die Box-Plots der k Spalten dieser Matrix "daten" in eine gemeinsame Zeichnung, versehen die einzelnen Boxen mit den Namen a, b, c, d, e und achten besonders auf Ausreißer:

```
k = 5; n = 100; λ = 2;
daten = RandomReal[ExponentialDistribution[λ], {n, k}];
BoxWhiskerPlot[daten, BoxLabels → {a, b, c, d, e}, BoxOutliers → Automatic, BoxOutlierMarkers → {▲, ■}]
Clear[k, n, λ, daten]
```



■ Balkendiagramm und Tortendiagramm

Zur graphischen Darstellung der Stichprobe *stich* eines *qualitativen* Merkmals durch ein Balkendiagramm bzw Tortendiagramm müssen zuerst die absoluten Häufigkeiten der einzelnen Werte der darzustellenden Stichprobe ermittelt werden. Dazu dient der bereits früher besprochene Befehl Tally. Nachdem damit eine Liste der absoluten Häufigkeiten der Stichprobe *stich* ermittelt wurde, lassen sich diese Häufigkeiten mit den Befehlen BarChart bzw PieChart durch ein Balkendiagramm bzw Tortendiagramm graphisch veranschaulichen:

■ BarChart[Part[Tally[*stich*], *All*, *2*]] zeichnet ein Balkendiagramm der absoluten Häufigkeiten der Stichprobe *stich*.

■ PieChart[Part[Tally[*stich*], *All*, *2*]]

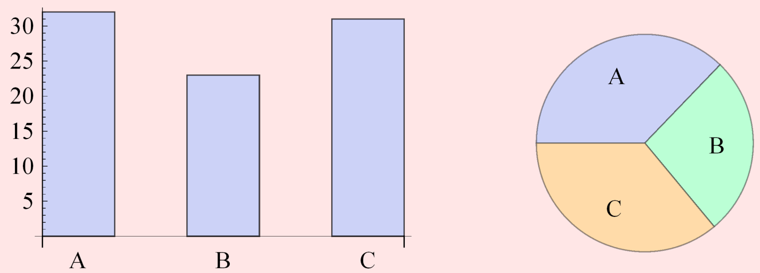
zeichnet ein Tortendiagramm der absoluten Häufigkeiten der Stichprobe *stich*.

2.1.8 Beispiel: Vom Datenmaterial *stahl* zeichne man sowohl ein Balkendiagramm als auch ein Tortendiagramm des qualitativen Merkmals Sorte (erste Spalte).

▼

Lösung: Wir lesen dazu das im Datenordner abgelegte Datenfile *stahlfile* ein, rufen die erste Spalte dieser Datenmatrix auf und bezeichnen die zugehörige Stichprobe mit "sorte". Von dieser Stichprobe ermitteln wir mit Hilfe von Part und Tally die Häufigkeiten der einzelnen Sorten und zeichnen von diesen Häufigkeiten ein Balkendiagramm und ein Tortendiagramm:

```
sorte = Rest[Part[<< stahlfile, All, 1]];
plot1 = BarChart[Part[Tally[sorte], All, 2], ChartLabels → {"A", "B", "C"}, BarSpacing → 1, ImageSize → {150, 10
plot2 = PieChart[Part[Tally[sorte], All, 2], ChartLabels → {"A", "B", "C"}, ImageSize → {150, 100}];
GraphicsGrid[{{plot1, plot2}}]
Clear[sorte, plot1, plot2]
```



Zeigt man mit der Maus auf einen Balken bzw ein Tortenstück, so wird die Häufigkeit des entsprechenden Merkmals angezeigt.

2.2 Graphische Darstellung bivariater Daten

Wir gehen wieder davon aus, dass unser statistisches Datenmaterial *daten* bereits in Form einer Datenmatrix, welche im Datenfile *datenfile* unseres Datenordners abgelegt ist, vorliegt und befassen uns in diesem Abschnitt mit der Frage, wie die gegenseitige Abhängigkeit zweier Merkmale graphisch sichtbar gemacht werden kann. Dabei haben wir zu unterscheiden, ob beide Merkmale quantitativ bzw beide Merkmale qualitativ sind oder ob eines dieser beiden Merkmale quantitativ und das andere Merkmal qualitativ ist.

2.2.1 Bemerkung: Die Abhängigkeit zweier *quantitativer* Merkmale lässt sich durch ein Scatter-Plot bzw ein dreidimensionales Histogramm graphisch veranschaulichen; die Beziehungen zwischen einem *qualitativen* und einem *quantitativen* Merkmal kann durch ein multiples Box-Plot graphisch dargestellt werden; die Abhängigkeit zweier *qualitativer* Merkmale lässt sich durch eine Kontingenztafel in übersichtlicher Weise angeben und durch ein dreidimensionales Balkendiagramm graphisch veranschaulichen.

■ Scatter-Plot

Bei dem darzustellenden Datenmaterial *matrix* handelt es sich um eine Matrix mit n Zeilen und zwei Spalten (den Stichproben der beiden zur Diskussion stehenden *quantitativen* Merkmale). Zur graphischen Veranschaulichung der Abhängigkeiten dieser beiden *quantitativen* Merkmale durch ein Scatter-Plot dient der Befehl ScatterPlot:

■ ScatterPlot[*matrix*]

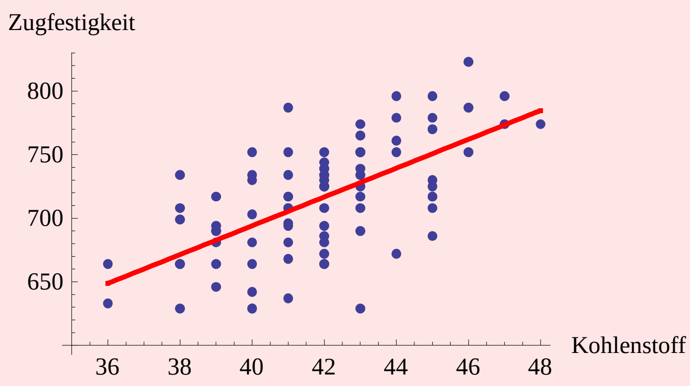
die in der Matrix *matrix* der Form $\{\{x_1, y_1\}, \{x_2, y_2\}, \ldots, \{x_n, y_n\}\}$ angegebenen Listen $\{x_i, y_i\}$ werden als Koordinaten von Punkten in der xy-Ebene gedeutet. Diese Punkte werden zusammen mit der durch diese Punkte gehenden Regressionsgeraden gezeichnet (dabei handelt es sich um jene Gerade, bei der die Summe der Quadrate der (in y-Richtung gemessenen) Abstände der Punkte von dieser Geraden minimal ist). Es stehen die üblichen Optionen des Befehls ListPlot zur Verfügung.

2.2.2 Beispiel: Für das Datenmaterial *stahl* veranschauliche man die Abhängigkeit des quantitativen Merkmals Zugfestigkeit (zweite Spalte) vom quantitativen Merkmal Kohlenstoff (erste Spalte) durch ein Scatter-Plot.

▼

Lösung: Wir lesen das im Datenordner abgelegte Datenfile *stahlfile* ein, erzeugen mit Hilfe von Part und Rest eine Matrix, deren zwei Spalten aus den Stichproben der beiden Merkmale Kohlenstoff und Zugfestigkeit bestehen, nennen diese Matrix "stahlKoZu" und zeichnen von dieser Matrix ein Scatter-Plot:

```
stahlKoZu = Rest[Part[<< stahlfile, All, {2, 3}]];
ScatterPlot[stahlKoZu, PlotStyle → PointSize[0.02], AxesOrigin → {35, 600},
  AxesLabel → {"Kohlenstoff", "Zugfestigkeit"}]
Clear[stahlKoZu]
```



Man erkennt an diesem Scatter-Plot deutlich, dass die Zugfestigkeit von Stahlblechen mit zunehmendem Kohlenstoffgehalt zunimmt. Da die einzelnen Punkte aber noch relativ weit von der Regressionsgeraden entfernt sind, dürfte die Zugfestigkeit von Stahlblechen noch von weiteren Einflussgrößen (die in unserem Datenmaterial aber nicht zur Verfügung stehen) abhängen.

■ Dreidimensionales Histogramm

Bei dem darzustellenden Datenmaterial *matrix*1, *matrix*2, … handelt es sich um Matrizen mit $n_1, n_2, \ldots$ Zeilen und zwei Spalten (den Stichproben der beiden zur Diskussion stehenden *quantitativen* Merkmale). Zur graphischen Veranschaulichung der Abhängigkeiten dieser beiden *quantitativen* Merkmale durch ein dreidimensionales Histogramm dient der Befehl Histogram3D:

■ Histogram3D[{*matrix*1, *matrix*2, …}]

die in den Matrizen *matrix*1, *matrix*2, … der Form $\{\{x_1, y_1\}, \{x_2, y_2\}, \ldots, \{x_n, y_n\}\}$ angegebenen Listen $\{x_i, y_i\}$ werden als Koordinaten von Punkten in der *xy*-Ebene gedeutet. Von diesen Punkten wird ein dreidimensionales Histogramm gezeichnet, wobei die Intervalleinteilung automatisch vorgenommen wird.

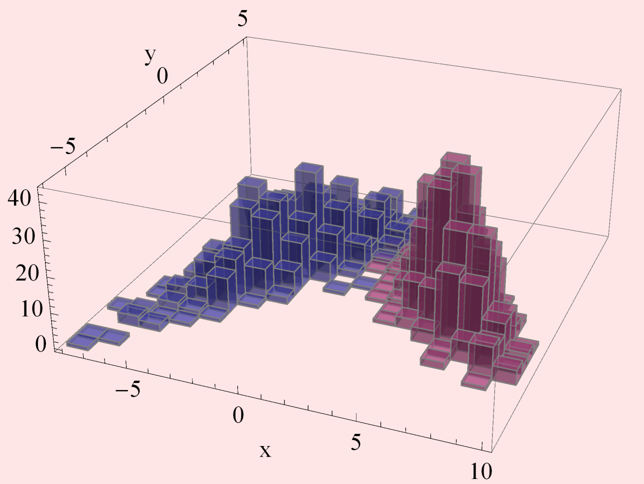
Mit der Option $\{n, m\}$ bzw $\{\{\{b_1, b_2, \ldots\}\}, \{\{c_1, c_2, \ldots\}\}$ wird der gesamte Wertebereich der Stichproben in $n \times m$ gleich große Rechtecke eingeteilt bzw werden nur die Häufigkeiten der Rechtecke $[b_i, b_{i+1}[\times [c_k, c_{k+1}[$ graphisch dargestellt.

2.2.3 Beispiel: Man erzeuge $n_1 = 600$ mit den Parametern $\vec{\mu}_1 = \{-2, 0\}$ und $\mathbf{\Sigma}_1 = \{\{4, 3\}, \{3, 3\}\}$ sowie $n_2 = 600$ mit den Paramtern $\vec{\mu}_2 = \{6, -1\}$ und $\mathbf{\Sigma}_2 = \{\{2, -1\}, \{-1, 1\}\}$ multinormal verteilte Zufallspunkte in der *xy*-Ebene und veranschauliche dieses Datenmaterial durch ein dreidimensionales Histogramm.

▼

Lösung: Wir erzeugen zuerst in der bekannten Weise das darzustellende Datenmaterial und geben diesem Datenmaterial den Namen *daten*1 bzw *daten*2. Anschließend stellen wir diese Punkte durch ein dreidimensionales Histogramm graphisch dar:

```
n1 = 600; μ1 = {−2, 0}; Σ1 = {{4, 3}, {3, 3}}; n2 = 600; μ2 = {6, −1}; Σ2 = {{2, −1}, {−1, 1}};
daten1 = RandomReal[MultinormalDistribution[μ1, Σ1], n1];
daten2 = RandomReal[MultinormalDistribution[μ2, Σ2], n2];
Histogram3D[{daten1, daten2}, AxesLabel → {"x", "y"}]
Clear[n1, μ1, Σ1, n2, μ2, Σ2, daten1, daten2]
```



■ Multiples Box-Plot

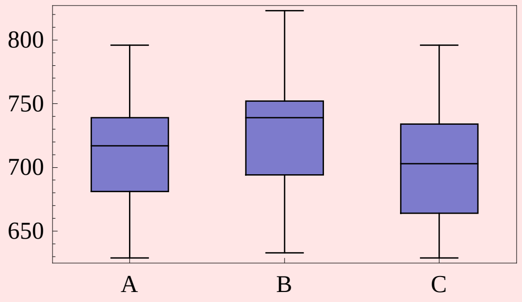
Bei dem darzustellenden Datenmaterial handelt es sich um eine Matrix mit *n* Zeilen und zwei Spalten (den Stichproben der beiden zur Diskussion stehenden Merkmale), wobei die erste Spalte dem *qualitativen* Merkmal und die zweite Spalte dem *quantitativen* Merkmal entspricht. Um die Abhängigkeit des *quantitativen* Merkmals vom *qualitativen* Merkmal graphisch darzustellen, wählen wir zuerst mit Hilfe von Cases jeweils jene Fälle aus, die den gleichen Eintrag im *qualitativen* Merkmal besitzen und erzeugen Listen mit den zugehörigen Einträgen des *quantitativen* Merkmals. Von diesen Listen zeichnen wir mit Hilfe des Befehls BoxWhiskerPlot ein multiples Box-Plot.

2.2.4 Beispiel: Für das Datenmaterial *stahl* veranschauliche man die Abhängigkeit des quantitativen Merkmals Zugfestigkeit (dritte Spalte) vom qualitativen Merkmal Sorte (erste Spalte) durch ein multiples Box-Plot.

▼

Lösung: Wir lesen dazu das im Datenordner abgelegte Datenfile *stahlfile* ein, wählen mit Hilfe von Cases die Sorten A bzw B bzw C aus, rufen davon mittels Part jeweils die dritten Spalten auf und bezeichnen diese so erzeugten Stichproben mit "sorteA" bzw "sorteB" bzw "sorteC". Von diesen drei Stichproben zeichnen wir mit Hilfe von BoxWhiskerPlot das gewünschten Plot:

```
sorteA = Part[Cases[<< stahlfile, {A, x_, y_}], All, 3];
sorteB = Part[Cases[<< stahlfile, {B, x_, y_}], All, 3];
sorteC = Part[Cases[<< stahlfile, {C, x_, y_}], All, 3];
BoxWhiskerPlot[sorteA, sorteB, sorteC, BoxLabels → {"A", "B", "C"}, ImageSize → {200, 110}]
Clear[sorteA, sorteB, sorteC]
```



Wir erkennen an dieser Zeichnung, dass die Zugfestigkeit von Stahlblechen der Sorte *B* insgesamt etwas höher sein dürfte, als die Zugfestigkeit von Stahlblechen der beiden anderen Sorten. Ob diese Tatsache aber "statistisch signifikant" ist oder durch zufällige Schwankungen allein erklärt werden kann, ist Gegenstand weiterer statistischer Untersuchungen.

■ Kontingenztafel und dreidimensionales Balkendiagramm

Bei dem darzustellenden Datenmaterial handelt es sich um eine Matrix mit *n* Zeilen und zwei Spalten (den Stichproben der beiden zur Diskussion stehenden *qualitativen* Merkmale). Die Beziehungen dieser beiden qualitativen Merkmale lässt sich in Form einer Kontingenztafel übersichtlich angeben und durch ein dreidimensionales Balkendiagramm graphisch darstellen. Dazu dienen die beiden Befehle Kontingenztafel bzw Balkendiagramm3D:

■ Kontingenztafel[*matrix*]

gibt die Kontingenztafel der Matrix *matrix* an. Darunter versteht man eine Tabelle mit den absoluten Häufigkeiten aller Paare {*x*, *y*}, wobei *x* alle möglichen Werte des ersten qualitativen Merkmals und *y* alle möglichen Werte des zweiten qualitativen Merkmals durchläuft.

■ Balkendiagramm3D[*matrix*]

stellt die in der Kontingenztafel angeführten absoluten Häufigkeiten in Form eines Bar-Charts graphisch dar.

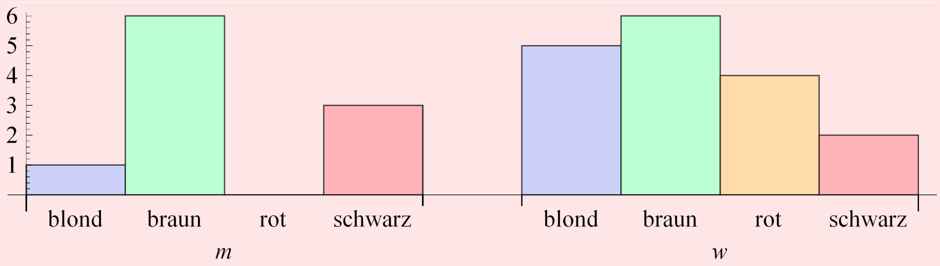
2.2.5 Beispiel: Für das Datenmaterial *studenten* veranschauliche man die Abhängigkeit des qualitativen Merkmals Haarfarbe (dritte Spalte) vom qualitativen Merkmal Geschlecht (erste Spalte) durch eine Kontingenztafel sowie durch ein 3D-Balkendiagramm.

▼

Lösung: Wir lesen dazu das im Datenordner abgelegte Datenfile *studentenfile* ein, rufen mit Hilfe von Part die dritte Spalte (Haarfarbe) und erste Spalte (Geschlecht) auf, streichen die erste Zeile (mit den Namen der Variablen) und bezeichnen diese Matrix mit "studentenHaarGeschlecht". Von dieser Matrix ermitteln wir die Kontingenztafel sowie das zugehörige dreidimensionale Balkendiagramm:

```
studentenHaarGeschlecht = Rest[Part[<< studentenfile, All, {3, 1}]];
Kontingenztafel[studentenHaarGeschlecht, TableSpacing → {1, 5}]
Balkendiagramm3D[studentenHaarGeschlecht, BarSpacing → {0, 1}, AspectRatio → 0.2,
  ImageSize → {350, 100}]
Clear[studentenHaarGeschlecht]
```

	m	w
blond	1	5
braun	6	6
rot	0	4
schwarz	3	2



Man erkennt, dass die Haarfarbe und das Geschlecht der untersuchten Studenten eine gewisse Abhängigkeit aufweisen. So kommt beispielsweise die Haarfarbe "rot" nur bei Frauen vor. Außerdem trit bei Männern die Haarfarbe "blond" deutlich seltener auf als bei Frauen. Die Überprüfung, ob diese Abhängigkeiten tatsächlich "statistisch signifikant" sind, ist Aufgabe weiterer Untersuchungen.