Box-Plots

Ziel: übersichtliche Darstellung der Daten.
Boxplot zu dem Eingangsbeispiel mit n=5:

`Descr_Boxplot0.sas`

Prozeduren: UNIVARIATE, GPLOT, BOXPLOT

Box-Plots

Ziel: übersichtliche Darstellung der Daten.
Boxplot zu dem Eingangsbeispiel mit n=5:

`Descr_Boxplot0.sas`

Prozeduren: UNIVARIATE, GPLOT, BOXPLOT

```
PROC UNIVARIATE PLOT;
```

Box-Plots

Ziel: übersichtliche Darstellung der Daten.
Boxplot zu dem Eingangsbeispiel mit n=5:

`Descr_Boxplot0.sas`

Prozeduren: UNIVARIATE, GPLOT, BOXPLOT

```
PROC UNIVARIATE PLOT;

SYMBOL1 INTERPOL=BOXT10;
PROC GPLOT;
   PLOT y*x=1;
```

Box-Plots

Ziel: übersichtliche Darstellung der Daten.
Boxplot zu dem Eingangsbeispiel mit n=5:

`Descr_Boxplot0.sas`

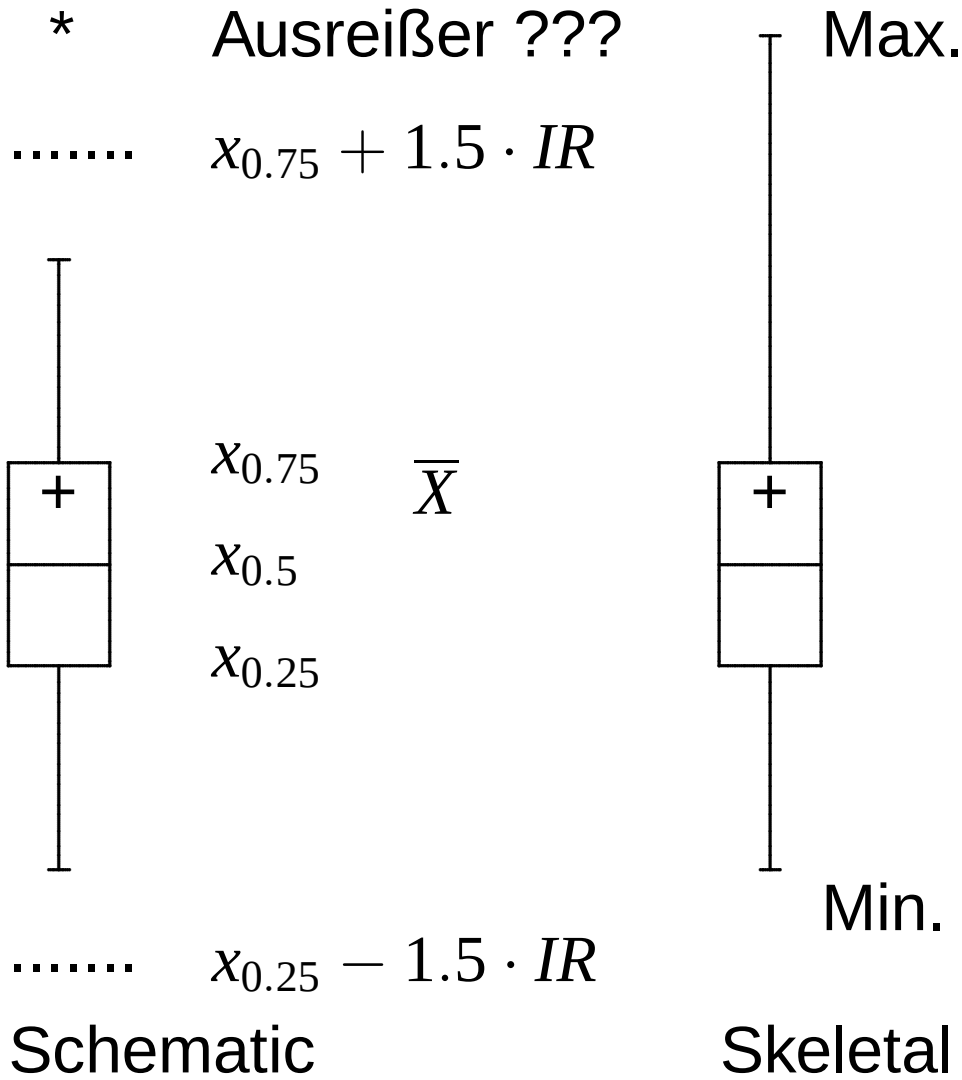
Prozeduren: UNIVARIATE, GPLOT, BOXPLOT

```
PROC UNIVARIATE PLOT;

SYMBOL1 INTERPOL=BOXT10;
PROC GPLOT;
   PLOT y*x=1;

PROC BOXPLOT;
   PLOT y*x /BOXSTYLE=SCHEMATIC;
            /BOXSTYLE=SKELETAL;
```

Prozedur BOXPLOT



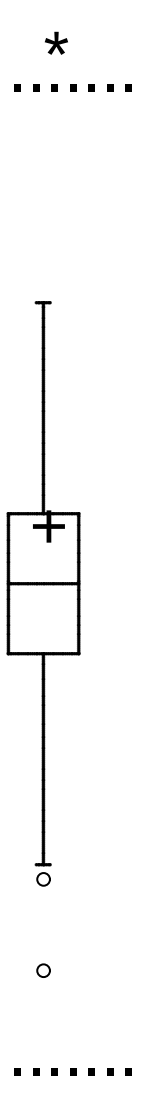
Erläuterung zu BOXSTYLE=Schematic

$X \sim \mathcal{N}(\mu, \sigma^2)$

etwa 99% der Daten liegen zwischen den “fences”.

$$
\begin{aligned}
0.99 &= 0.995 - 0.005 \\
&= \Phi(2.575) - \Phi(-2.575) \\
&= P(\mu - 2.575\sigma < X < \mu + 2.575\sigma) \\
&\approx P(x_{0.5} - 2.575 \cdot \underbrace{0.7434 \cdot IR}_{} < X < \\
&\qquad x_{0.5} + 2.575 \cdot \underbrace{0.7434 \cdot IR}_{}) \\
&= P(x_{0.5} - 1.914 \cdot IR < X < x_{0.5} + 1.914 \cdot IR) \\
&\approx P(x_{0.5} - 2 \cdot IR < X < x_{0.5} + 2 \cdot IR) \\
&= P(x_{0.25} - 1.5 \cdot IR < X < x_{0.75} + 1.5 \cdot IR)
\end{aligned}
$$

Prozedur UNIVARIATE, Option PLOT



Box-Plots in SAS

Ein Merkmal, eine Gruppe (Merkmal gr)

```
gr = 1;
PROC BOXPLOT;
   PLOT zeit*gr; RUN;
```

Ein Merkmal (zeit), mehrere Gruppen (gr)

```
PROC BOXPLOT;
   PLOT zeit*gr; RUN;
```

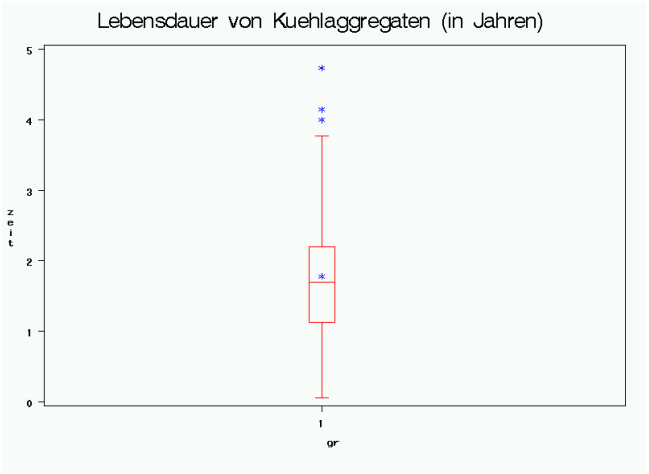
Ein Merkmal (X), mehrere Gruppen (gr)

```
SYMBOL INTERPOL=BOXT10;
PROC GPLOT; PLOT X*gr; RUN;
```

`Descr_Boxplot.sas` `Descr_Boxplot1.sas`

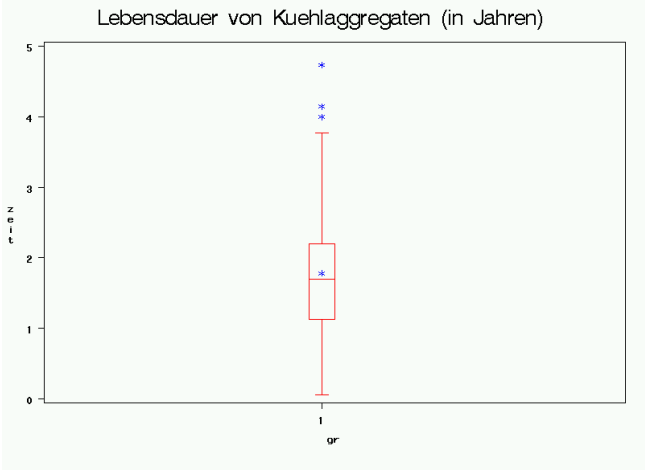
Boxplots - Beispiele

Lebensdauern von 100 Kühlaggregaten

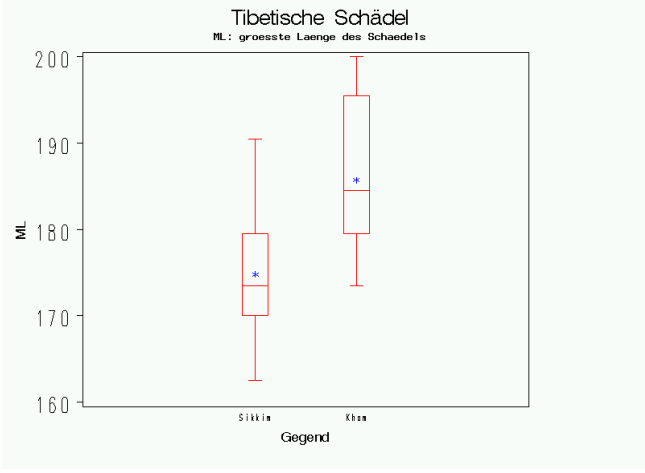


Boxplots - Beispiele

Lebensdauern von 100 Kühlaggregaten



Schädelmaße in zwei Regionen Tibets



Box-Plots in SAS (2)

Box-Plots von mehreren Variablen

`Descr_Boxplot2.sas`

1. Data-Step:
   Definition von neuen Variablen, die konstant gesetzt werden.
2. Symbol-Anweisungen für die einzelnen darzustellenden Variablen definieren.
3. Achsenbeschriftung entsprechend den Variablen definieren.
4. Prozedur GPLOT;

Probability Plots
Erinnerung: Normalverteilung

(i) Dichte der Standard-Normalverteilung

$$\phi(x) = \frac{1}{\sqrt{2 \cdot \pi}} \cdot e^{-\frac{x^2}{2}}, \qquad -\infty < x < \infty$$

(ii) Verteilungsfunktion der Standard-Normal

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2 \cdot \pi}} \cdot e^{-\frac{t^2}{2}} dt, \qquad -\infty < x < \infty$$

(iii) Dichte der Normalverteilung

$$\frac{1}{\sigma}\phi(\frac{x-\mu}{\sigma}) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x-\mu)^2}{\sigma^2}},$$

mit Erwartungswert μ und Varianz σ^2.

Probability Plots

Erinnerung: Normalverteilung, Quantile

Der Wert $\Phi^{-1}(u)$ heißt u-Quantil

der Standard-Normalverteilung.

Die Funktion $\Phi^{-1}(u), u \in (0,1)$, heißt Quantilfunktion

der Standard-Normalverteilung.

$\alpha = 0.05$

$$\begin{aligned} \Phi^{-1}(1-\alpha) &= \Phi^{-1}(0.95) = 1.645 \\ \Phi^{-1}(1-\frac{\alpha}{2}) &= \Phi^{-1}(0.975) = 1.96 \end{aligned}$$

$\Phi^{-1}(\alpha)$: α-Quantil, theoretisch

$x_\alpha = x_{(\lfloor \alpha n \rfloor)}$: α-Quantil, empirisch

Q-Q-Plot
Variante 1

Wenn Normalverteilung zutrifft, so müssen die Punkte

$$(\Phi^{-1}(\alpha), x_\alpha)$$

etwa auf einer Geraden liegen,

$$\Phi^{-1}(\alpha) \approx \frac{x_\alpha - \mu}{\sigma} = \frac{x_{(\lfloor \alpha n \rfloor)} - \mu}{\sigma}$$

```
PROC UNIVARIATE PLOT; RUN;
```

Die theoretischen Werte (+) werden durch die empirischen Werte (*) überschrieben.
Je weniger “+”-Zeichen zu sehen sind, desto näher sind wir an der NV.

`Descr_QQPlot.sas`

Q-Q-Plot
Variante 2


```
PROC UNIVARIATE;
   QQPLOT var /Optionen;
RUN;
```

wie oben, bessere Grafik, aber keine Linie.
Es werden die Punkte

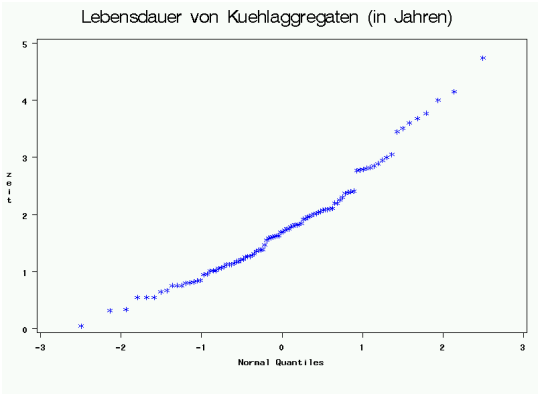
$$\left(\Phi^{-1}\left(\frac{i-0.375}{n+0.25}\right), x_{(i)}\right)$$

geplottet. $i = 1, \ldots, n$.

Bem.: $\Phi^{-1}\left(\frac{i-0.375}{n+0.25}\right)$ ist eine Approximation von $\mathbf{E}X_{(i)}$ bei Standard-Normalverteilung.

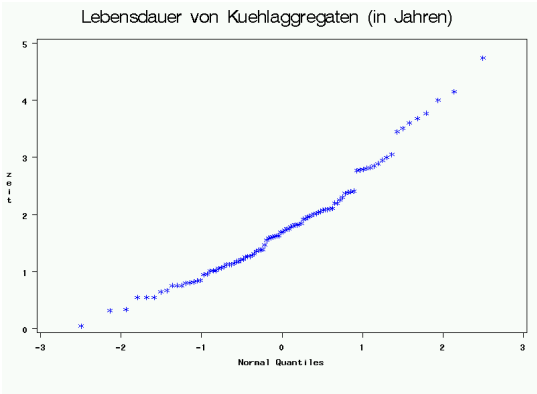
Q-Q Plots - Beispiele

Lebensdauern von 100 Kühlaggregaten

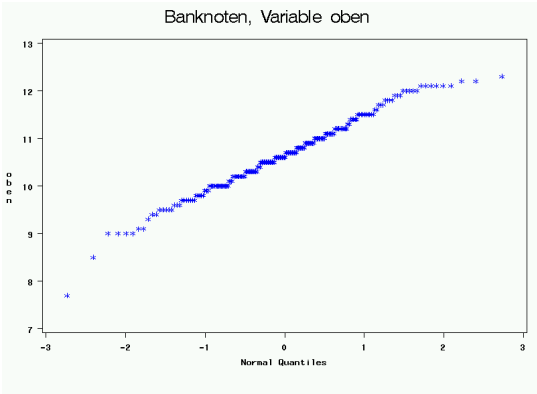


Q-Q Plots - Beispiele

Lebensdauern von 100 Kühlaggregaten

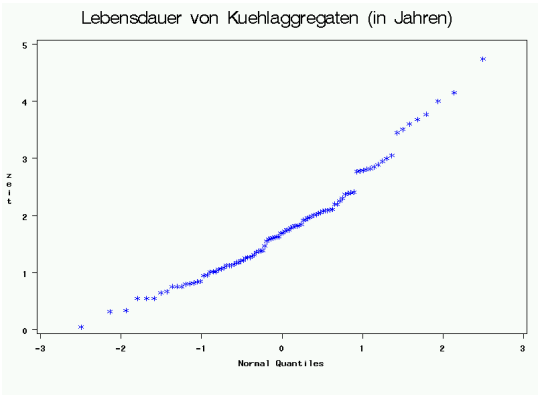


Abmessungen von Banknoten

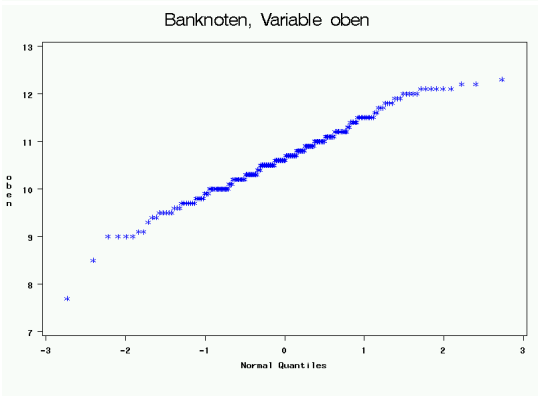


Q-Q Plots - Beispiele

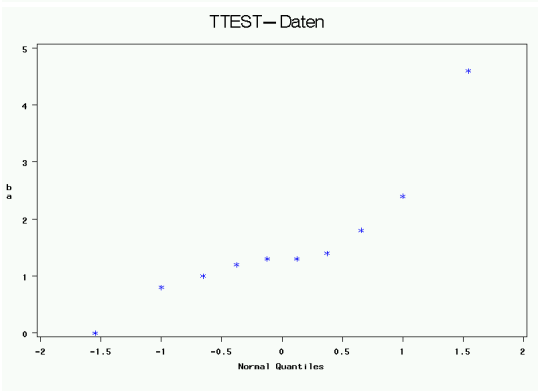
Lebensdauern von 100 Kühlaggregaten



Abmessungen von Banknoten



Verlängerung der Schlafdauer



Probability Plot


```
PROC UNIVARIATE;
   PROBPLOT var /Optionen;
RUN;
```

wie oben, x-Achse hat die selbe Skala, aber eine andere Beschriftung, statt x_α steht α, also

$$\left(\alpha, x_{(i)}\right) = \left(\frac{i - 0.375}{n + 0.25}, x_{(i)}\right)$$

Bem.: Es können auch einige andere Verteilungen verwendet werden.

Q-Q Plot
Übersicht

wenige Punkte weg von der Geraden	Ausreißer
linkes Ende unter der Linie rechtes Ende über der Linie	lange Tails
linkes Ende über der Linie rechtes Ende unter der Linie	kurze Tails
gebogene Kurve, steigender Anstieg	rechtsschief
gebogene Kurve, fallender Anstieg	linksschief
Plateaus und Sprünge	diskrete Daten gerundete Dat.

Gerade $y = x$	empirische $\approx$ theoretische Verteil.
Gerade $y = ax + b$	nur Lage- oder Skalenunterschied