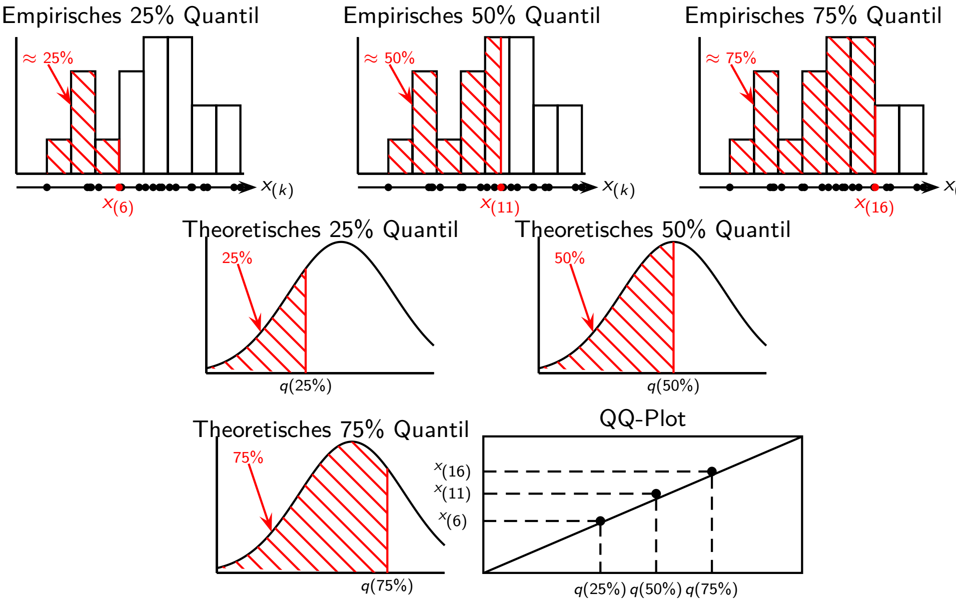
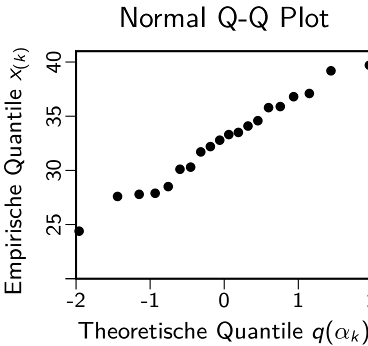
QQ-Plot Betondruckfestigkeit



Normal-Plot

- Plotten theoretischen Quantile $q(\alpha_k) = \Phi^{-1}(\alpha_k)$ der Standardnormaverteilung gegen die empirischen Quantile $x_{(k)}$

1. Datensatz: $x_1, x_2, \ldots, x_n$
2. $\alpha_{(k)} = \frac{k-0.5}{n}, \; k = 1, 2, \ldots, n$
3. Theoretische Quantile: $q(\alpha_k) = \Phi^{-1}(\alpha_k)$
4. Empirische Quantile: $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$
5. $\big(q(\alpha_k), x_{(k)}\big)$ plotten

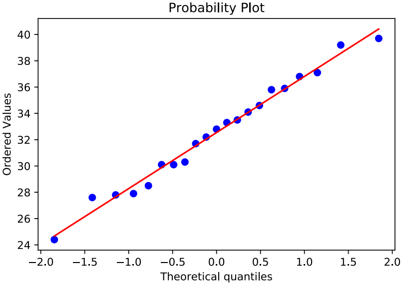


Normal-Plot mit **Python**

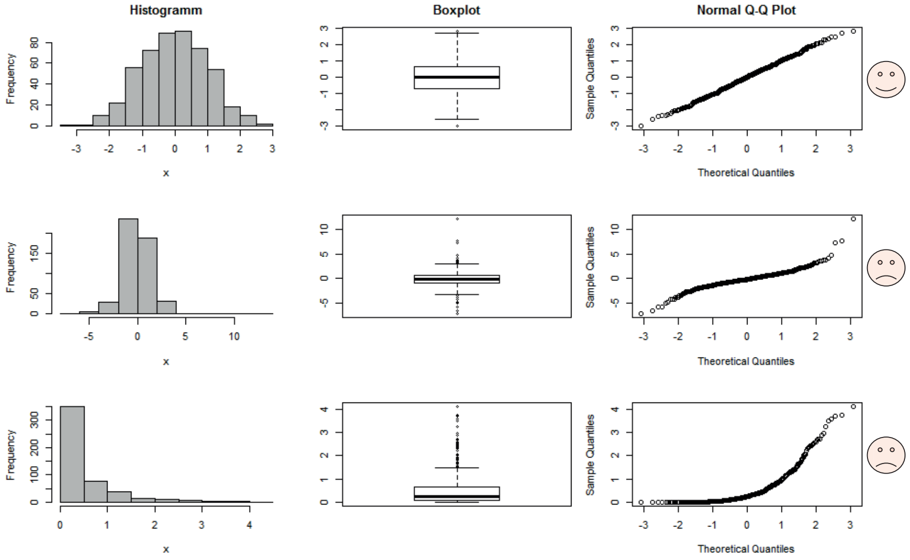
```
import numpy as np
import scipy.stats as st
import matplotlib.pyplot as plt

x = np.array([24.4, 27.6, 27.8, 27.9, 28.5, 30.1, 30.1, 30.3,
31.7, 32.2, 32.8, 33.3, 33.5, 34.1, 34.6, 35.8, 35.9, 36.8, 37.1,
39.2, 39.7])

st.probplot(x, plot=plt)
```



Beispiele Normalplots für 3 Datensätze mit $n = 500$



Parameterschätzung - Maximum-Likelihood

- $X \sim \mathrm{Pois}(\lambda)$ mit unbekanntem Parameter λ, welchen wir mit der Maximum-Likelihood Methode schätzen wollen.

- Datensatz : $x_1, \ldots, x_n$

- **Likelihood-Funktion**: Wahrscheinlichkeit, dass $X_1 = x_1$ **und** $X_2 = x_2$ etc. **und** $X_3 = x_3$ beobachtet werden, falls $X_i \sim$Pois(λ) und alle Beobachtungen x_i unabhängig voneinander sind:

$$
\begin{aligned}
P(X_1 = x_1 \cap \ldots \cap X = x_n) &= P(X_1 = x_1) \cdot P(X_2 = x_2) \cdots P(X = x_n) \\
&= e^{-\lambda}\frac{\lambda^{x_1}}{x_1!} \cdot e^{-\lambda}\frac{\lambda^{x_2}}{x_2!} \cdots e^{-\lambda}\frac{\lambda^{x_n}}{x_n!} \\
&= \prod_{i=1}^{n} e^{-\lambda}\frac{\lambda^{x_i}}{x_i!} \\
&= L(\lambda)
\end{aligned}
$$

- **log-Likelihood-Funktion** ist:

$$\begin{aligned} l(\lambda) &= \log(L(\lambda)) \\ &= \log\left(\prod_{i=1}^{n} e^{-\lambda}\frac{\lambda^{x_i}}{x_i!}\right) \\ &= \sum_{i=1}^{n} \log\left(e^{-\lambda}\frac{\lambda^{x_i}}{x_i!}\right) \end{aligned}$$

- Leitet man $l(\lambda)$ nach λ ab und setzt $l'(\lambda) = 0$, so erhält man die Gleichung

$$\frac{1}{\lambda}\sum_{i=1}^{n} x_i - n = 0\,.$$

- Maximum-Likelihood Schätzer : $\hat{\lambda} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

Maximum-Likelihood-Schätzer für Poisson-Verteilung

Maximum-Likelihood Schätzer

Sind die Datenpunkte x_i unabhängige Realisierungen der Zufallsvariablen $X_i \sim \text{Poisson}(\lambda)$, so ergibt der Maximum-Likelihood-Schätzer:

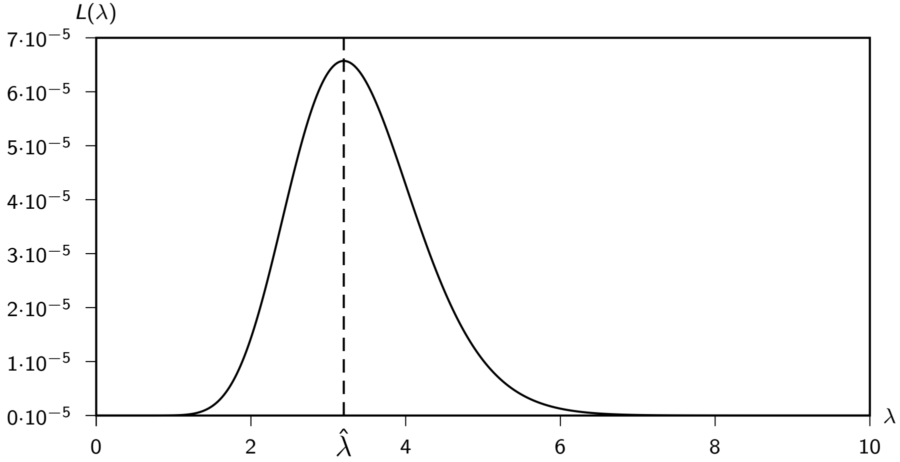
$$\hat{\lambda} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

- **Beispiel:** Hersteller für Isolationsmaterialen misst die Anzahl kanzerogener Fasern pro mm^2 in 5 Proben. Gemessenen Werte $x_1, \ldots, x_5$ seien:

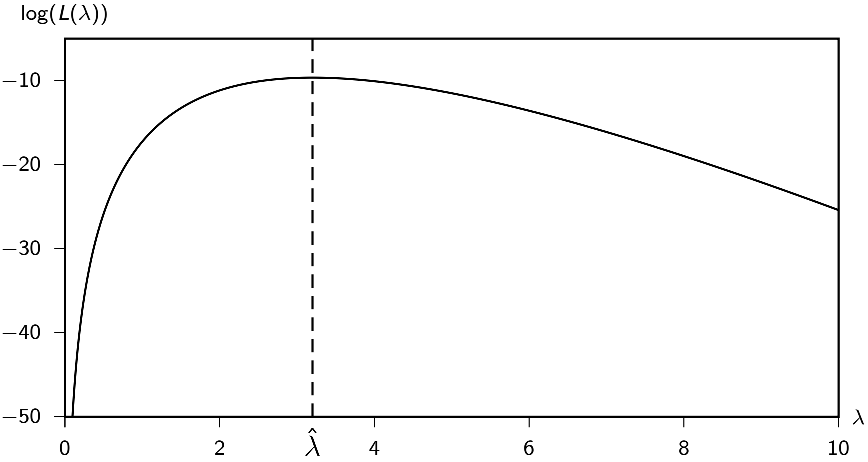
$$4, 1, 2, 3, 6$$

- Anzahl kanzerogener Fasern folgt Poisson-Verteilung: Beobachtungen sind $x_1, \ldots, x_5$ Realisierungen von $X_1, \ldots, X_5 \sim \text{Poisson}(\lambda)$ sind. Maximum-Likelihood-Schätzer: $\hat{\lambda} = \bar{x} = 3.2$

Likelihood-Funktion für Datensatz kanzerogener Fasern



Log-Likelihood Funktion für Datensatz kanzerogener Fasern



Momentenmethode

- Maximum-Likelihood Schätzer bei Poisson-Verteilung:

$$\hat{\lambda} = \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \,,$$

d.h. $\hat{\lambda} = \widehat{E(X)} = \bar{x}$. Geschätzter Erwartungswert ist arithmetisches Mittel.

- Maximum-Likelihood Schätzer bei Binomial-Verteilung, z.B. x_1 Gewinne von n_1 Losen beim 1. Versuch, etc. (siehe Serie 5):

$$\hat{\pi} = \frac{x_1 + \ldots + x_n}{n_1 + \ldots + n_2} = \frac{x}{n} \,.$$

Da $\mathrm{E}(X) = n\pi$ und $\hat{\pi} = \frac{\widehat{E(X)}}{n}$, ist der geschätzte Erwartungswert die beobachtete Anzahl Gewinne $\widehat{E(X)} = x$

Momentenmethode

- X Binomial-verteilt: $\mathrm{E}(X) = n\pi$
- Also $\pi = \mathrm{E}(X)/n$
- n (Anzahl unabhängiger Versuche) wird als bekannt vorausgesetzt
- Pragmatisch motivierte Schätzung ist dann:

$$\widehat{\mathrm{E}(X)} = x = \text{ beobachtete Anzahl Gewinne}$$

- Man schätzt den Erwartungswert also durch die Beobachtung.

Momentenmethode

Somit ergibt sich aufgrund der **Momentenmethode** die relative Häufigkeit

$$\hat{\pi} = x/n$$

Zusammenfassung Parameterschätzer

- Für unsere beobachteten Daten nehmen wir ein **Modell** an (z.B. Binomialverteilung)
- Das Modell enthält unbekannte Parameter (z.B. π)
- Basierend auf den beobachteten Daten versuchen wir, die Parameter zu schätzen (z.B. mit Momentenmethode, Maximum-Likelihood Methode).
- **Das heisst, dass wir basierend auf den beobachteten Daten versuchen, Rückschlüsse über den datengenerierenden Mechanismus zu ziehen!**