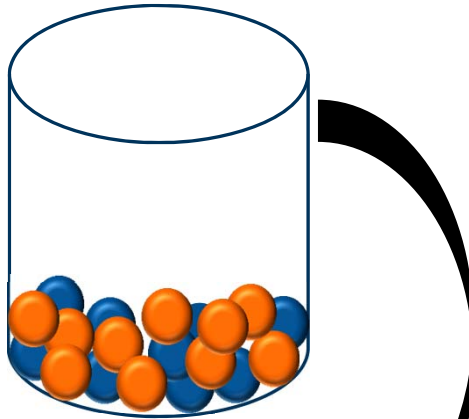


Wahrscheinlichkeitsrechnung vs. Statistik

Wahrscheinlichkeitsrechnung

Modell

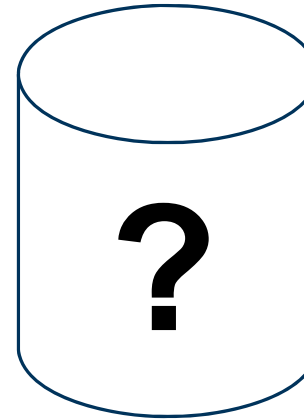


Daten



Gegeben die Informationen über die Urne:
Was werden wir in den Händen haben, und
mit welcher Wahrscheinlichkeit ?

Statistik



Gegeben die Informationen in unserer Hand:
Was ist in der Urne enthalten, und wie sicher
sind wir darüber?

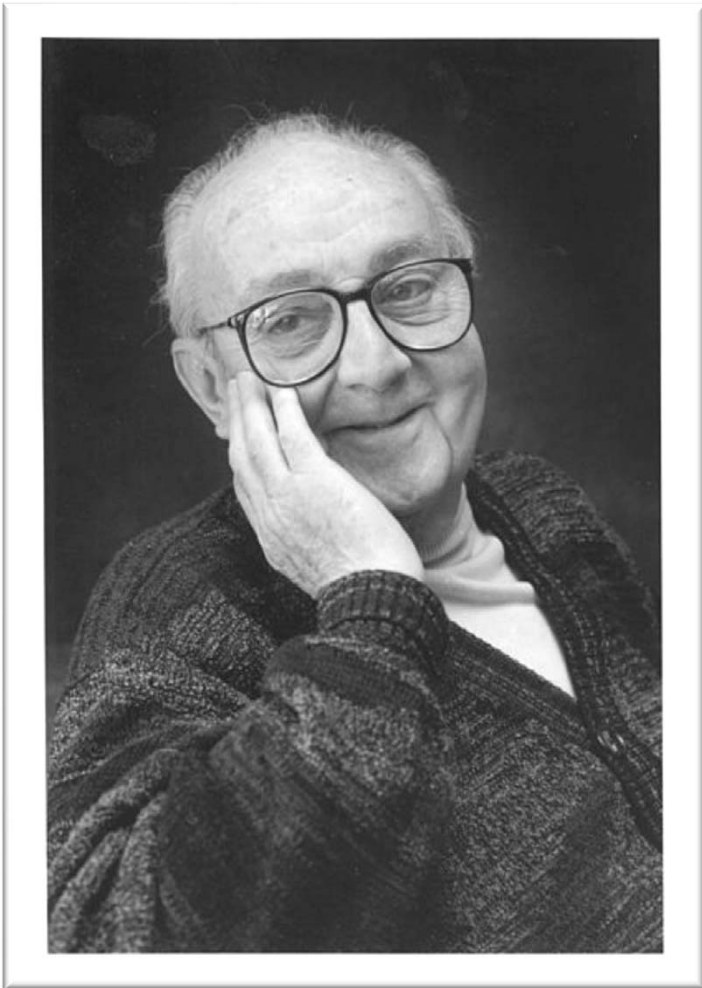
Grundidee

- Beobachtete Daten x_1, \dots, x_n werden betrachtet als Realisierungen von i.i.d. Zufallsvariablen $X_1, \dots, X_n \sim F(\theta)$
- Man nimmt an, dass die Modellfamilie $F(\cdot)$ (z.B. normalverteilt) bekannt ist, aber der Parameter θ (z.B. $\theta = (\mu, \sigma^2)$) nicht.
- Ziel: anhand von den Daten Rückschlüsse auf das Modell (Parameter) ziehen
- Beispiele für $F(\theta)$:
 - $F(\theta) = \text{Bern}(p)$, also $\theta = p$
 - $F(\theta) = \text{Poisson}(\lambda)$, also $\theta = \lambda$
 - $F(\theta) = N(\mu, \sigma^2)$, also $\theta = (\mu, \sigma^2)$

Drei Hauptfragen in der schliessenden Statistik

- Welches ist der plausibelste Wert für den unbekanntem Parameter θ ?
→ Punktschätzung (best guess)
- Ist ein bestimmter vorgegebener Parameterwert θ_0 mit den Daten verträglich?
→ Test
- Was ist der Bereich von plausiblen Parameterwerten?
→ Vertrauensintervall

Zitat George Box



"Essentially, all models are wrong, but some are useful"

Wahl der Verteilungsfamilie

Man kann die Verteilungsfamilie wählen aufgrund von

- Erfahrung (“was sich bisher bewährt hat”)
- physikalischen Argumenten
(Summen von vielen Effekten sind z.B. gemäss ZGS normalverteilt)
- grafischen Methoden:
 - Vergleiche empirische Verteilungsfunktion und theoretische Verteilungsfunktion
 - Vergleiche empirisches Histogramm und theoretische Dichte
 - Oft am besten: vergleiche empirische Quantile und theoretische Quantile (QQ-Plot)

Beispiel QQ-Plot: Druckfestigkeit

k	$x_{(k)}$
1	24.4
2	27.6
3	27.8
4	27.9
5	28.5
6	30.1
7	30.3
8	31.7
9	32.2
10	32.8
11	33.3
12	33.5
13	34.1
14	34.6
15	35.8
16	35.9
17	36.8
18	37.1
19	39.2
20	39.7

- Messe Druckfestigkeit an $n = 20$ verschiedenen Prüfkörpern.
- Wir wollen untersuchen, wie gut die Daten mit einer **Normalverteilung** beschrieben werden können.

Beispiel QQ-Plot: Datensatz Druckfestigkeit

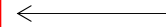
k	$\alpha_k = (k - 0.5)/n$	$q_{\alpha_k} = x_{(k)}$
1	0.025	24.4
2	0.075	27.6
3	0.125	27.8
4	0.175	27.9
5	0.225	28.5
6	0.275	30.1
7	0.325	30.3
8	0.375	31.7
9	0.425	32.2
10	0.475	32.8
11	0.525	33.3
12	0.575	33.5
13	0.625	34.1
14	0.675	34.6
15	0.725	35.8
16	0.775	35.9
17	0.825	36.8
18	0.875	37.1
19	0.925	39.2
20	0.975	39.7

Empirische
($\alpha_k \times 100\%$)
Quantile

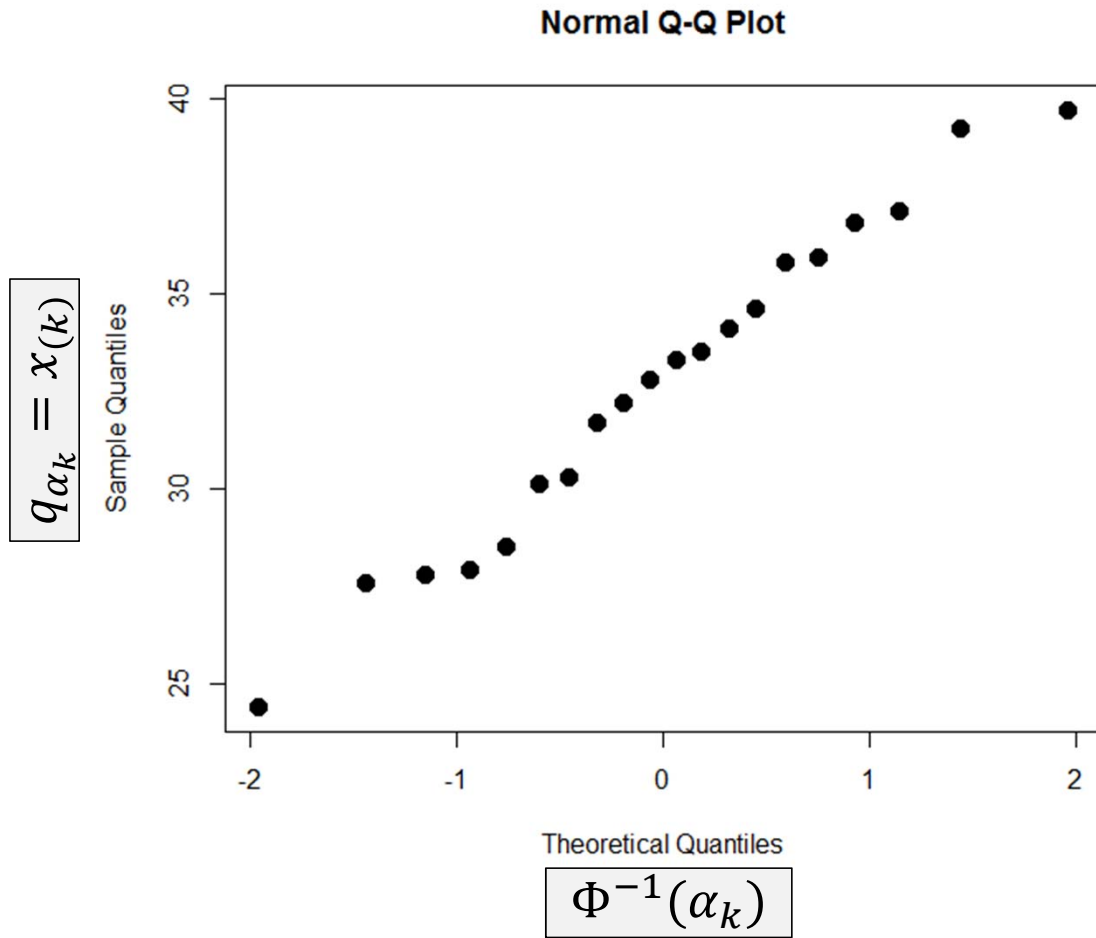


k	$\alpha_k = (k - 0.5)/n$	$\Phi^{-1}(\alpha_k)$
1	0.025	-1.9600
2	0.075	-1.4395
3	0.125	-1.1503
4	0.175	-0.9346
5	0.225	-0.7554
6	0.275	-0.5978
7	0.325	-0.4538
8	0.375	-0.3186
9	0.425	-0.1891
10	0.475	-0.0627
11	0.525	0.0627
12	0.575	0.1891
13	0.625	0.3186
14	0.675	0.4538
15	0.725	0.5978
16	0.775	0.7554
17	0.825	0.9346
18	0.875	1.1503
19	0.925	1.4395
20	0.975	1.9600

Theoretische
($\alpha_k \times 100\%$)
Quantile einer
 $N(0,1)$ Verteilung



Beispieldatensatz Druckfestigkeit



Bemerkungen zu QQ-Plots (allgemein)

- Der Trick mit einer **Standard-Referenzverteilung** (wie beim Normalplot) geht auch bei anderen Verteilungen. Manchmal muss man hierzu die Achsen noch umskalieren, z.B. mit log.