Empirische Quantile

Seien $x_1 \leq x_2 \leq ... \leq x_n$ der Größe nach aufsteigend geordnete Beobachtungswerte mit den Rängen 1, 2, ..., n, so heißt der kleinste Wert, der von einem Anteil p aller Werte nicht überschritten wird, das p - Quantil x_p ($0<p<1$). Beispielsweise ist $x_{0,5}$ das 0,5 - Quantil (Median). 50% der Beobachtungen liegen unterhalb $x_{0,5}$.
Spezielle Quantile sind:

Median:	p = 0,5
Quartile:	p = 0,25 ; 0,5 ; 0,75
Dezile:	p = k / 10 (k = 1, ..., 9)
Perzentile	p = k / 100 (k = 1, ..., 99)

Es gibt mehrere Berechnungsmöglichkeiten bei gegebenen n Beobachtungen.

1. Es sind nur Beobachtungswerte zugelassen

$x_p = x_k$ mit $k = [n \cdot p+1]_G$, wobei $[\,z\,]_G$ die **Gauß-Klammer-Funktion** darstellt. Sie bezeichnet die größte ganze Zahl, die kleiner oder gleich z ist, z.B. $[4{,}5]_G = 4$.

2. Es sind auch „berechnete“ Quantile zugelassen

2.1. Übliche Berechnung) ([1])

$x_p = x_k$ mit $k = [n \cdot p+1]_G$ wenn $n \cdot p$ nicht ganzzahlig,
$x_p = (x_{np} + x_{np+1}) / 2$ wenn $n \cdot p$ eine ganze Zahl ist.

2.2. Interpolation über die Ränge der aufsteigend geordneten Beobachtungen ([2])

Die Rangzahl für das i-te Quantil erhält man aus $i = (n+1) \cdot p$, sie liegt zwischen den Rängen $k = [(n+1) \cdot p]_G$ und (k+1) Man interpoliert:

Rang	Wert
k	x_k
$(n+1) \cdot p$	x_p
k+1	x_{k+1}

$$\frac{(k+1) - k}{x_{k+1} - x_k} = \frac{(n+1) \cdot p - k}{x_p - x_k}$$

das ergibt: $x_p = x_k + ((n+1) \cdot p - k) \cdot (x_{k+1} - x_k)$ für $k < (n+1) \cdot p < (k+1)$

2.3. Bestimmung aus der empirischen Verteilungsfunktion F(X) (kumulative rel. Häufigkeit in bactcount.sav)

Für das p - Quantil x_p gilt $F(x_p) \geq p$
Für den Median gilt bei gerader Anzahl von Beobachtungen ($n \cdot p$ ist eine ganze Zahl):
$x_{0,5} = (x_{n/2} + x_{n/2+1}) / 2$

Beispiel zur Berechnung der Quartile Q_1 , Q_2 und Q_3 mit dem Datensatz bactcount.sav (SPSS)
Quelle: J.S.Kim, R.J. Dailey: Biostatistics for Oral Healthcare. Tab. 2.3.4 Blackwell Munksgaard 2008
$Q_1 = x_{0,25}$; $Q_2 = x_{0,5}$; $Q_3 = x_{0,75}$

1. Nur Beobachtungswerte

Q_1 : n = 90 ; p = 0,25 ; $n \cdot p$ = 22,5 ; $k = [22{,}5 +1]_G = 23$; Q_1 = **220** (der 23. Wert der aufsteigend geordnete Beobachtungswerte).
Q_2 (Median): n = 90 ; p = 0,5 ; $n \cdot p$ = 45 ; $k = [45 +1]_G = 46$; Q_2 = **304**
Q_3 : n = 90 ; p = 0,75 ; $n \cdot p$ = 67,5 ; $k = [67{,}5 +1]_G = 68$; Q_2 = **366**

2.1. Übliche Berechnung) ([1])

Q1: aus $n \cdot \alpha$ = 22,5 folgt Q1 = „23. Wert in der Reihe“ = **220**
Q2: aus $n \cdot \alpha$ = 45 folgt Q2 = „1/2·(45.+46.Wert)“ = **303**
Q3: aus $n \cdot \alpha$ = 67,5 folgt Q3 = „68. Wert in der Reihe“ = **366**

2.2. Interpolation über die Ränge ([2])

Q_1 : n = 90 ; p = 0,25 ; $(n+1) \cdot p$ = 22,75 ; $k = [22{,}75]_G = 22$; $x_{22} = 213$; $x_{23} = 220$ das ergibt
$Q_1 = 213 + (22{,}75 - 22) \cdot (220 - 213) =$ **218,25**
Q_2 (Median): n = 90 ; p = 0,5 ; $(n+1) \cdot p$ = 45,5 ; $k = [45{,}5]_G = 45$; $x_{45} = 302$; $x_{46} = 304$ das ergibt
$Q_2 = 302 + (45{,}5 - 45) \cdot (304 - 302) =$ **303**
Q_3 : n = 90 ; p = 0,75 ; $(n+1) \cdot p$ = 68,25 ; k = 68 ; $Q_3 = 366 + 0{,}25 \cdot 14 =$ **369,5**

Ausdruck aus SPSS:

COUNT

N	Gültig	90
	Fehlend	0
Perzentile	25	218,25
	50	303,00
	75	369,50

SPSS verwendet die Methode der Interpolation über die Ränge.

[1] Hartung: Statistik. Oldenbourg Verlag
[2] Bland: An Introduction to Medical Statistics. Oxford University Press