

Theorie Aufgabe 14

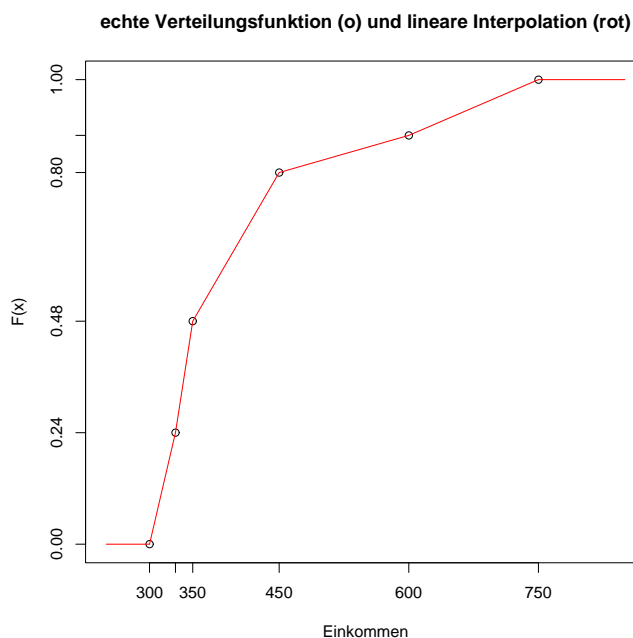
Theorie: Lineare Interpolation

- Absolute Häufigkeiten für klassierte Daten eines Merkmals  $X$  mit  $k$  Klassen sind folgendermaßen definiert

$$h_j = h(c_{j-1} \leq X < c_j), \quad j = 1, \dots, k,$$

mit Klassengrenzen  $c_0 < c_1 < \dots < c_k$ .

- Die kumulierte relative Häufigkeitsverteilung lässt sich graphisch folgendermaßen veranschaulichen:



- Idee: Approximation durch Polygonzug (d.h. eine stückweise lineare Funktion)
- Annahme: Merkmalswerte sind innerhalb jeder der Klassen gleichverteilt.
- Dazu wird die lineare Interpolation verwendet: Gegeben sind jeweils zwei Punkte  $(c_{j-1}, F(c_{j-1}))$  und  $(c_j, F(c_j))$ , wobei  $F$  die echte Verteilungsfunktion bezeichnet, die nur an einzelnen Punkten bekannt ist. Diese Punkte sollen durch eine lineare Funktion  $y = ax + b$  verbunden werden. Problem:  $a$  und  $b$  sind unbekannt. Idee: Löse folgendes lineares Gleichungssystem:

$$\begin{array}{rcl} (I) & F(c_{j-1}) & = ac_{j-1} + b \\ (II) & F(c_j) & = ac_j + b \\ \hline (II) - (I) & F(c_j) - F(c_{j-1}) & = a(c_j - c_{j-1}) \end{array}$$

Wir erhalten

$$a = \frac{F(c_j) - F(c_{j-1})}{c_j - c_{j-1}}$$

und durch Einsetzen in (I)

$$b = F(c_{j-1}) - \frac{F(c_j) - F(c_{j-1})}{c_j - c_{j-1}}c_{j-1}$$

und damit für  $x \in [c_{j-1}, c_j)$ :

$$\begin{aligned} F(x) &= \frac{F(c_j) - F(c_{j-1})}{c_j - c_{j-1}}x + F(c_{j-1}) - \frac{F(c_j) - F(c_{j-1})}{c_j - c_{j-1}}c_{j-1} \\ &= F(c_{j-1}) + \frac{x - c_{j-1}}{c_j - c_{j-1}}(F(c_j) - F(c_{j-1})) \end{aligned}$$