

Aufgabe 10

Das Ergebnis der Untersuchung eines kardinalskalierten Merkmals X sei in folgender Tabelle wiedergegeben:

Ausprägung	1	2	3	4	7
Anzahl	4	4	6	4	2

- Bestimmen Sie das arithmetische Mittel, den Modus und den Median.
- Berechnen Sie die mittlere quadratische Abweichung.
- Obige Daten werden nun mittels der Intervalle $[0; 3)$, $[3; 4)$ und $[4; 7]$ klassiert. Bestimmen Sie die Rechteckhöhen des Histogramms.

Lösungshinweis:

```
x <- rep(c(1, 2, 3, 4, 7), times = c(4, 4, 6, 4, 2))
median(x)
mean(x)
SP <- max(x) - min(x) # Spannweite
MQA <- mean((x - mean(x))^2) # mittlere quadratische Abweichung
s <- sqrt(MQA) # Standardabweichung
V <- s/mean(x) # Variationskoeffizient
```

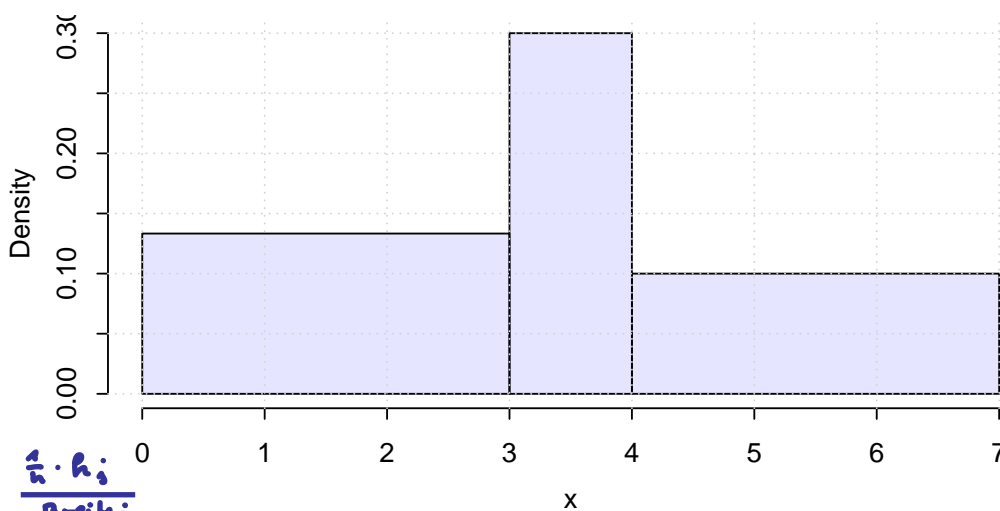
$$x_{\text{med}} = 3, \quad \bar{x} = 3,$$

$$SP = 6, \quad s^2 = 2.8,$$

$$s \approx 1.6733201, \quad V \approx 0.5577734.$$

Handwritten note: $x_{\text{med}} = \frac{1}{2}(x_{10} + x_{11}) = \frac{1}{2}(3+3) = 3$

```
require(MASS)
hist(x, breaks = c(0, 3, 4, 7), right = FALSE, col = rgb(0,
  0, 1, 0.1), main = "")
grid()
```



$$\text{Höhe}_j = \frac{\frac{1}{n} \cdot h_j}{\text{Breite}_j}$$

j	Klasse	Breite	h_j	Höhe ρ_j
1	$[0, 3)$	3	8	$\frac{2}{15}$
2	$[3, 4)$	1	6	0.3
3	$[4, 7]$	3	6	0.1

Σ 20

Aufgabe 11

Eine Umfrage über den Bierkonsum Münchner Bürger ergibt bei 10 Personen folgende Zahlenreihe (Liter pro Woche):

3	10	1	2	3	0	2	1	0	3
---	----	---	---	---	---	---	---	---	---

Berechnen Sie den Modalwert, den Median, das arithmetische Mittel, die Spannweite, die mittlere quadratische Abweichung, die Standardabweichung und den Variationskoeffizienten.

Lösungshinweis:

```
x <- c(3, 10, 1, 2, 3, 0, 2, 1, 0, 3)
table(x) # Kontingenztabelle
```

Ausprägung	0	1	2	3	10
Häufigkeit	2	2	2	3	1

$$\frac{1}{2}(x_5 + x_6) = 2 = x_{med}$$

also: $x_{mod} = 3$

0	0	1	1	2	2	3	3	3	10
1	2	3	4	5	6	7	8	9	10

```
median(x)
mean(x)
SP <- max(x) - min(x) # Spannweite
MQA <- mean((x - mean(x))^2) # mittlere quadratische Abweichung
s <- sqrt(MQA) # Standardabweichung
V <- s/mean(x) # Variationskoeffizient
```

$$x_{med} = 2,$$

$$\bar{x} = 2.5,$$

$$SP = 10,$$

$$s^2 = 7.45,$$

$$s \approx 2.7294688,$$

$$V \approx 1.0917875.$$

Aufgabe 12

Es wurde eine Untersuchung bei 600 zufällig ausgewählten Personen zum Thema „Selbstsicherheit und Autostärke“ durchgeführt. Dabei wurden folgende Beobachtungskategorien gebildet:

Selbstsicherheit $x \in \{\text{sehr stark, stark, mittel, schwach, sehr schwach}\}$

Autostärke $y \in \{[0, 40]; (40, 70]; (70, 100]; (100, 200]\}$ [in KW gemessen]

Es ergab sich die folgende gemeinsame Häufigkeitstabelle:

bei Unabhängigkeit würde man hier $\frac{90}{600} \cdot 130 = 19,5$ erwarten

		Selbstsicherheit					Σ	
		sehr stark	stark	mittel	schwach	sehr schwach		
Mittel	Autostärke							
	20	[0, 40]	60	25	0	0	5	90
	55	(40, 70]	30	40	60	10	15	
	85	(70, 100]	40	25	70	70	30	
150	(100, 200]	0	20	30	20	50	600	
			130					

- Wie viele Personen sind jeweils insgesamt in jeder Klasse der Autostärken? Zeichnen Sie ein Histogramm der Autostärken *aller* Personen.
- Warum läßt sich kein Histogramm der Selbstsicherheit zeichnen?
- Sind in der gegebenen Häufigkeitstabelle die beiden Merkmale Autostärke und Selbstsicherheit unabhängig? Es wird der Beweis für die richtige Antwort erwartet, nicht nur eine Behauptung.

Berechnen Sie für die folgenden beiden Teilaufgaben die Klassenmittelwerte der Autostärken (als Näherung für die exakten Werte).

Ferner wird die Gesamtheit aller 600 Personen aufgeteilt in 5 Gruppen aufgrund der Zugehörigkeit zu den einzelnen Selbstsicherheitskategorien. Betrachten Sie bei d) und e) jede dieser Gruppen separat.

- Wie groß sind (näherungsweise) Mittelwerte und Spannweiten der Autostärken in jeder der 5 Gruppen?
- Wie würden Sie den Vorschlag bewerten, die Standardabweichung der Autostärken durch die Klassenmitten verbunden mit den jeweiligen Häufigkeiten zu berechnen? Ist das Ergebnis einer solchen Rechnung korrekt, zu klein oder zu groß? Bitte begründen Sie jeweils Ihre Antwort.

Lösungshinweis:

```
a) # definiere Werte jeweils in Mitte der Klasse
Klassenmitten <- c(20, 55, 85, 150)
Haeufigkeiten <- c(90, 155, 235, 120)
x <- rep(Klassenmitten, Haeufigkeiten)
table(x)

## x
## 20 55 85 150
## 90 155 235 120

# Anzahl der Objekte insgesamt
n <- length(x)
n
## [1] 600
```

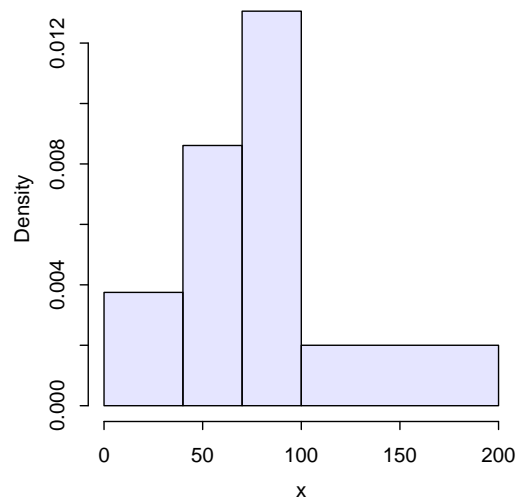
Berechne Höhe_{*i*} = $\frac{c \cdot h(a_i)}{\text{Breite}_i}$ mit $c = \frac{1}{n}$ (hier: $n = 600$):

Intervall	Breite	Häufigkeit	Höhe
[0; 40]	40	90	0.00375
(40; 70]	30	155	0.0086111
(70; 100]	30	235	0.0130556
(100; 200]	100	120	0.002

```
hist(x, breaks = c(0, 40, 70, 100, 200),
     right = TRUE, col = rgb(0, 0, 1, 0.1),
     main = "")
```

[..], (...], (...]
[..), [..), [..]

right=T
right=F



- b) Das geht nicht, weil *Selbstsicherheit* ein ordinales Merkmal ist.
- c) Betrachte die Gruppe der Leute mit *Autostärke* $\in [0; 40]$ und *Selbstsicherheit* gleich „sehr stark“: Die Häufigkeiten sind dort $h_{[0;40]} = 90$ und $h_{\text{sehr stark}} = 130$; Gesamtzahl der Leute $n = 600$. bei gleichmäßiger Verteilung würde man in der Gruppe insgesamt $90 \cdot 130/600 = 19,5$

Leute erwarten; es sind aber in dieser Gruppe 60 Leute. Die beiden Merkmale sind somit nicht unabhängig.

- d) Da man die genaue Lage der Werte in den Klassen nicht kennt, kann man jeweils mit dem Klassenmittel als Schätzwert rechnen. Es ergibt sich:

$$\begin{aligned}\bar{x}_{\text{sehr stark}} &= \frac{1}{130} \cdot (20 \cdot 60 + 55 \cdot 30 + 85 \cdot 40 + 150 \cdot 0) \\ &= 48,08\end{aligned}$$

$$\bar{x}_{\text{stark}} = 71.1363636$$

$$\bar{x}_{\text{mittel}} = 85.9375$$

$$\bar{x}_{\text{schwach}} = 95$$

$$\bar{x}_{\text{sehr schwach}} = 109.75$$

Für die Spannweite kann man von den maximal möglichen Werten ausgehen:

$$SP_{\text{sehr stark}} = 100 - 0 = 100$$

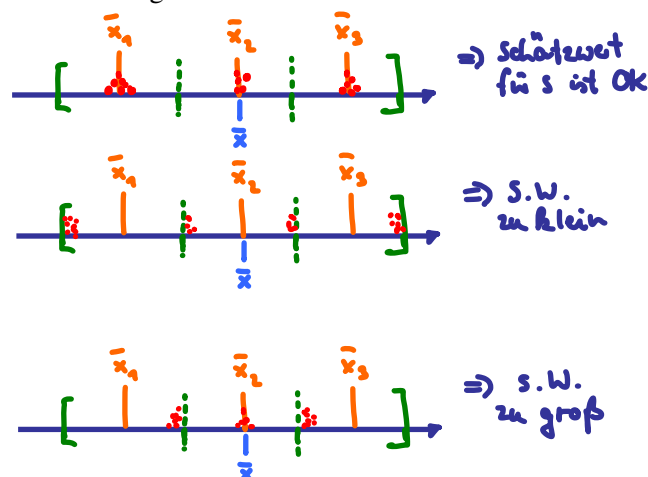
$$SP_{\text{stark}} = 200 - 0 = 200$$

$$SP_{\text{mittel}} = 200 - 40 = 160$$

$$SP_{\text{schwach}} = 200 - 40 = 160$$

$$SP_{\text{sehr schwach}} = 200 - 0 = 200$$

- e) Die Standardabweichung könnte so berechnet den wahren Wert über- oder unterschätzen: Angenommen alle tatsächlichen Werte würde immer möglichst nahe am arithmetischen Mittel aller Werte liegen, also jeweils am „dem arithmetischen Mittel zugewandten“ Rand der Intervalle, würde die Standardabweichung so überschätzt; im Fall der in Wirklichkeit maximal weit vom arithmetischen Mittel liegenden Werte würde die tatsächliche Standardabweichung so unterschätzt.



Aufgabe 13

Deskriptiv: Lage Streuung

In der deutschen Fußballnationalmannschaft wird die Körpergröße (in cm) einiger (aktiver wie ehemaliger) Spieler erhoben. Es ergeben sich folgende Daten:

Körpergröße	[150, 160)	[160, 170)	[170, 175)	[175, 180)	[180, 190)	[190, 210)
Anzahl	5	35	47	38	101	27

- Zeichnen Sie ein Histogramm der Daten.
- Schätzen Sie das arithmetische Mittel und den Median der Körpergröße.
- Schätzen Sie die Standardabweichung der Körpergröße, indem Sie davon ausgehen, dass alle Nennungen pro Klasse jeweils in der Klassenmitte liegen. Ist dieser Wert größer oder kleiner als die tatsächliche Standardabweichung in der Grundgesamtheit?

Lösungshinweis:

```
a) # definiere Werte jeweils in Mitte der Klasse
Klassenmitten <- c(155, 165, 172.5, 177.5, 185, 200)
Haeufigkeiten <- c(5, 35, 47, 38, 101, 27)
x <- rep(Klassenmitten, Haeufigkeiten)
table(x)

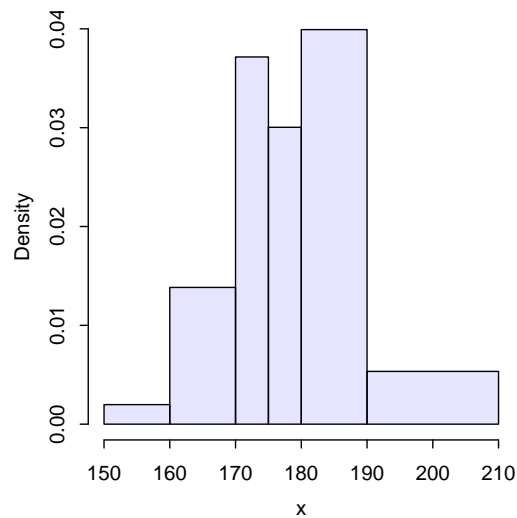
## x
##  155  165 172.5 177.5  185  200
##    5   35   47   38  101   27

# Anzahl der Objekte insgesamt
n <- length(x)
n
## [1] 253
```

Berechne $H\ddot{o}he_i = \frac{c \cdot h(a_i)}{Breite_i}$ mit $c = \frac{1}{n}$ (hier: $n = 253$):

	n_i	h_i	155	165	172,5	177,5	185	200
Körpergröße	[150, 160)	[160, 170)	[170, 175)	[175, 180)	[180, 190)	[190, 210)		
Häufigkeit		5	35	47	38	101	27	
Breite		10	10	5	5	10	20	
Höhe		0.002	0.0138	0.0372	0.03	0.0399	0.0053	

```
hist(x, breaks = c(150, 160, 170, 175, 180, 190, 210),
     right = TRUE, col = rgb(0, 0, 1, 0.1),
     main = "")
```



b) Rechnet man jeweils (näherungsweise) mit den Klassenmitten, erhält man:

$$\bar{x} = 179.7924901 \quad \text{und} \quad x_{\text{med}} = 185.$$

c) Für die Standardabweichung ergibt sich somit als Näherungswert:

$$s = 10.3678119$$

Die Standardabweichung könnte so berechnet den wahren Wert über- oder unterschätzen: Angenommen alle tatsächlichen Werte würde immer möglichst nahe am arithmetischen Mittel aller Werte liegen, also jeweils am „dem arithmetischen Mittel zugewandten“ Rand der Intervalle, würde die Standardabweichung so überschätzt; im Fall der in Wirklichkeit maximal weit vom arithmetischen Mittel liegenden Werte würde die tatsächliche Standardabweichung so unterschätzt.

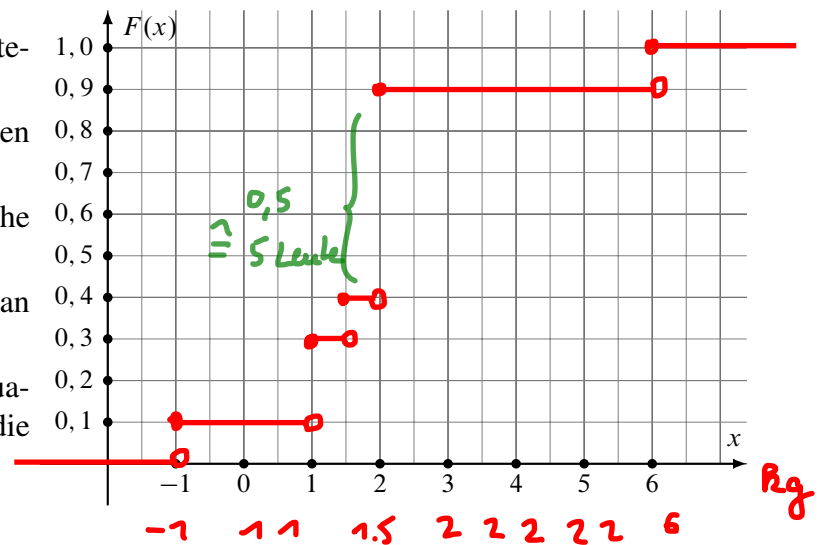
Aufgabe 14

Deskriptiv: Lage Streuung Vtgl.fkt.

10 Personen werden befragt, wieviel Sie in den Weihnachtsferien zugenommen haben. Für die empirische Verteilungsfunktion des abgefragten Merkmals $X \hat{=} „Gewichtszunahme vom 23. Dezember bis zum darauffolgenden 7. Januar“$ ergibt sich:

$$F(x) = \begin{cases} 0,0 & \text{für } x < -1,0 \\ 0,1 & \text{für } -1,0 \leq x < 1,0 \\ 0,3 & \text{für } 1,0 \leq x < 1,5 \\ 0,4 & \text{für } 1,5 \leq x < 2,0 \\ 0,9 & \text{für } 2,0 \leq x < 6,0 \\ 1,0 & \text{für } x \geq 6,0 \end{cases}$$

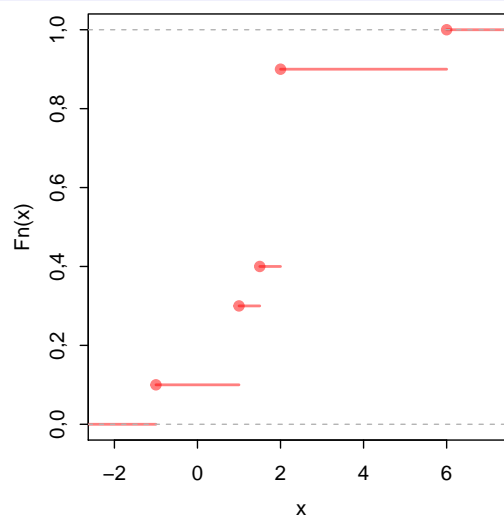
- Zeichnen Sie $F(x)$ in nebenstehendes Koordinatensystem ein.
- Schreiben Sie die ursprünglichen Daten als Urliste von X auf.
- Berechnen Sie das arithmetische Mittel von X .
- Bestimmen Sie den Median von X .
- Berechnen Sie die mittlere quadratische Abweichung sowie die Standardabweichung von X .



Lösungshinweis:

- b) und a) Da die 10%-Sprünge jeweils einer Person entsprechen (bei 10 Leuten), kann man die Urliste eindeutig aus F ablesen:

```
x <- c(-1, 1, 1, 1.5, 2, 2, 2, 2, 2, 6)
plot(ecdf(x), lwd = 2, col = rgb(1, 0, 0, 0.5), main = "")
```



- Für das arithmetische Mittel ergibt sich $\bar{x} = 1,85$.
- Der Median ist $x_{\text{med}} = 2$.
- $s^2 = 2,703$ und $s = 1,644$.

Aufgabe 15

Deskriptiv: Lageparameter Konzentration

Ein bestimmtes Gut wird von genau 7 Firmen produziert. Folgende Tabelle gibt an, wie viele tausend Stück jede Firma herstellt:

Firma:	A	B	C	D	E	F	G
prod. Stückzahl:	3	2	3	5	6	15	6

(tausend Stück)

- Bestimmen Sie den Medianwert der produzierten Stückzahlen.
- Skizzieren Sie für x -Werte aus dem Intervall $[0;20]$ den Verlauf der Funktion $F(x) =$ Anteil der Firmen, die höchstens $1000 \cdot x$ Stück produzieren.
- Errechnen Sie die Knickpunkte der zugehörigen Lorenzkurve.
- Errechnen Sie den normierten Gini-Koeffizienten.
- Bestimmen Sie den Konzentrationskoeffizienten CR_2 .

Lösungshinweis:

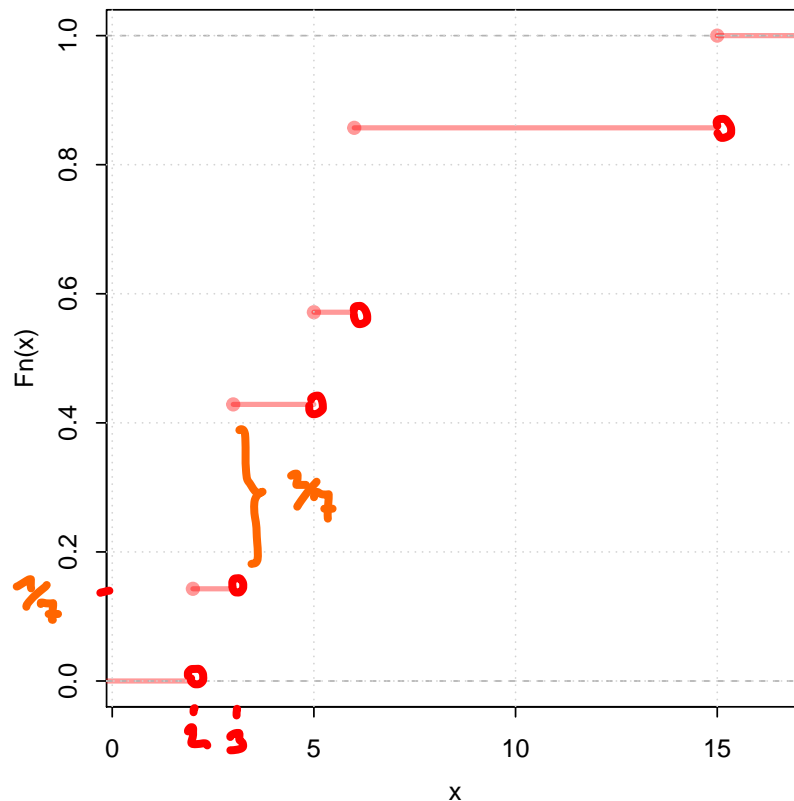
`x <- c(3, 2, 3, 5, 6, 15, 6)`

a_i	h_i	F_i
3	1	$\frac{1}{7}$
2	2	$\frac{2}{7}$
3	1	$\frac{3}{7}$
5	2	$\frac{4}{7}$
6	1	$\frac{6}{7}$
15	1	1
	$\Sigma 7$	

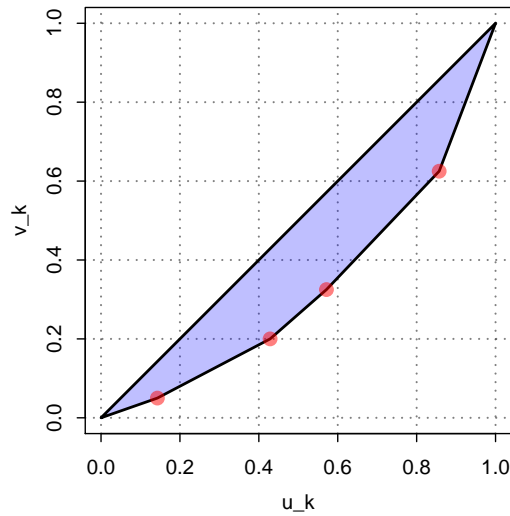
$F(x)$	f_i	Intervall
0	f_i	$x < 2$
$\frac{1}{7}$	f_i	$2 \leq x < 3$
$\frac{2}{7}$	f_i	$3 \leq x < 5$
$\frac{4}{7}$	f_i	$5 \leq x < 6$
$\frac{6}{7}$	f_i	$6 \leq x < 15$
1	f_i	$x \geq 15$

- a) Median: $x_{med} = 5$

- b) `plot(ecdf(x), lwd = 3, col = rgb(1, 0, 0, 0.4), main = "")`



```
c) library(ineq)
plot(Lc(x), las = 2, col = rgb(1, 0, 0, 0.5)) #
```



Für die Knicke ergibt sich:

k	u_k	v_k	u_k
1	0.143	0.050	$\frac{1}{7}$
3	0.429	0.200	$\frac{3}{7}$
4	0.571	0.325	$\frac{4}{7}$
6	0.857	0.625	$\frac{6}{7}$

d) $G = 0.3428571$ und damit $G^* = \frac{n}{n-1} \cdot G = \frac{7}{6} \cdot 0.3428571 = 0.4$.

e) $CR_2 = 0.525$.

P_i
2
3
3
3
6
6
6
15
$\Sigma \frac{15}{40}$

$$\frac{2 \sum_{i=1}^n i p_i - (n+1)}{n}$$

$$G = \frac{1}{7} \left[2 \cdot \frac{1}{40} (1 \cdot 2 + 2 \cdot 3 + 3 \cdot 3 + 4 \cdot 5 + 5 \cdot 6 + 6 \cdot 6 + 7 \cdot 15) - (7+1) \right]$$

$$= \dots \approx 0,3428571$$

$$CR_2 = \frac{6+15}{40} = \frac{21}{40} = 0,525$$