Prof. Dr. Ludwig Paditz, 17.10.2016

Statistik für Produktionstechniker

================================

1. Eine Befragung von Studenten eines Studienganges ergab die folgende Urliste:

seq(x,x,1,51,1)⇒ListNr

{1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20

seq(20,x,1,51,1)

{20,20,20,20,20,20,20,20,20,20,20,20,20,20,20,20,20

per Hand anpassen:

{23,22,22,25,24,20,24,23,23,22,20,21,21,25,25,23,23

{23,22,22,25,24,20,24,23,23,22,20,21,21,25,25,23,23

seq(4,x,1,51,1)

{4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,

per Hand anpassen:

{4,4,2,5,5,2,4,4,3,3,4,3,3,1,3,4,3,2,4,4,1,4,1,4,4,

{4,4,2,5,5,2,4,4,3,3,4,3,3,1,3,4,3,2,4,4,1,4,1,4,4,

Urdaten als Matrix:

trn(augment(augment(listToMat(ListNr),listToMat(Alter)),listT

$$\begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 & 16 & 17 \\ 23 & 22 & 22 & 25 & 24 & 20 & 24 & 23 & 23 & 22 & 20 & 21 & 21 & 25 & 25 & 23 & 23 \\ 4 & 4 & 2 & 5 & 5 & 2 & 4 & 4 & 3 & 3 & 4 & 3 & 3 & 1 & 3 & 4 & 3 \end{bmatrix}$$

Für jedes einzelne Merkmal (Alter, Semester) gebe man an bzw. berechne man:

(a) die Strichliste,

(b) die Tabelle der absoluten, relativen, kumulierten und relativen kumulierten Häufgkeiten,

(c) das Stabdiagramm,

(d) die Summenkurve,

(e) das arithmetische Mittel,

(f) alle Mediane,

(g) den Modalwert, falls er existiert,

(h) die Spannweite,

(i) den Quartilsabstand,

(j) die Varianz,

(k) die Standardabweichung,

(l) den Variationskoeffizienten.

Stat–Editor

Alter (Strichliste aus dem Histogramm ablesen) bzw. Variationsreihe:

```
sortA(Alter)
    {18,20,20,20,20,20,20,20,21,21,21,21,21,21,21,21,21▶
{1,0,7,11,9,13,4,3,0,2,0,0,0,0,1}→hAlter
                        {1,0,7,11,9,13,4,3,0,2,0,0,0,0,1}
sum(ans)
                                                       51
seq(x,x,18,32,1)→NAlter
          {18,19,20,21,22,23,24,25,26,27,28,29,30,31,32}
```

primäre Häufigkeitstabelle:

(absolute/relative Häufigkeiten)

trn(augment(listToMat(NAlter),listToMat(hAlter)))

$$\begin{bmatrix} 18 & 19 & 20 & 21 & 22 & 23 & 24 & 25 & 26 & 27 & 28 & 29 & 30 & 31 & 32 \\ 1 & 0 & 7 & 11 & 9 & 13 & 4 & 3 & 0 & 2 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

bereinigt um nichtauftretende Merkmale:

$$\begin{bmatrix} 18 & 20 & 21 & 22 & 23 & 24 & 25 & 27 & 32 \\ 1 & 7 & 11 & 9 & 13 & 4 & 3 & 2 & 1 \end{bmatrix}$$

[18 20 21 22 23 24 25 27 32]⇒NAlter

$$\begin{bmatrix} 18 & 20 & 21 & 22 & 23 & 24 & 25 & 27 & 32 \end{bmatrix}$$

[1 7 11 9 13 4 3 2 1]⇒hAlter

$$\begin{bmatrix} 1 & 7 & 11 & 9 & 13 & 4 & 3 & 2 & 1 \end{bmatrix}$$

trn(augment(trn(NAlter),trn(hAlter/51)))

$$\begin{bmatrix} 18 & 20 & 21 & 22 & 23 & 24 & 25 & 27 & 32 \\ \frac{1}{51} & \frac{7}{51} & \frac{11}{51} & \frac{3}{17} & \frac{13}{51} & \frac{4}{51} & \frac{1}{17} & \frac{2}{51} & \frac{1}{51} \end{bmatrix}$$

approx(ans)

$$\begin{bmatrix} 18.00 & 20.00 & 21.00 & 22.00 & 23.00 & 24.00 & 25.00 & 27.00 & 32. \\ 0.02 & 0.14 & 0.22 & 0.18 & 0.25 & 0.08 & 0.06 & 0.04 & 0.(\end{bmatrix}$$

in Prozent:

approx(trn(augment(trn(NAlter),trn(100*hAlter/51))))

$$\begin{bmatrix} 18.00 & 20.00 & 21.00 & 22.00 & 23.00 & 24.00 & 25.00 & 27.00 & 32. \\ 1.96 & 13.73 & 21.57 & 17.65 & 25.49 & 7.84 & 5.88 & 3.92 & 1.(\end{bmatrix}$$

kumuliert:

matToList(trn(NAlter),1)⇒NAlter

{18,20,21,22,23,24,25,27,32}

matToList(trn(hAlter),1)⇒hAlter

{1,7,11,9,13,4,3,2,1}

trn(augment(listToMat(NAlter),listToMat(cuml(hAlter))))

$$\begin{bmatrix} 18 & 20 & 21 & 22 & 23 & 24 & 25 & 27 & 32 \\ 1 & 8 & 19 & 28 & 41 & 45 & 48 & 50 & 51 \end{bmatrix}$$

trn(augment(listToMat(NAlter),listToMat(cuml(hAlter/51))))

$$\begin{bmatrix} 18 & 20 & 21 & 22 & 23 & 24 & 25 & 27 & 32 \\ \frac{1}{51} & \frac{8}{51} & \frac{19}{51} & \frac{28}{51} & \frac{41}{51} & \frac{15}{17} & \frac{16}{17} & \frac{50}{51} & 1 \end{bmatrix}$$

approx(trn(augment(listToMat(NAlter),listToMat(100*cuml(hAlt

$$\begin{bmatrix} 18.00 & 20.00 & 21.00 & 22.00 & 23.00 & 24.00 & 25.00 & 27.00 & 32. \\ 1.96 & 15.69 & 37.25 & 54.90 & 80.39 & 88.24 & 94.12 & 98.04 & 10(\end{bmatrix}$$

Semester (Strichliste aus dem Histogramm ablesen)

bzw. Variationsreihe:

sortA(Semester)

{1,1,1,2,2,2,2,2,3,3,3,3,3,3,3,3,3,4,4,4,4,4,4,4,4,

{3,5,9,31,3}→hSemest

{3,5,9,31,3}

sum(ans)

51

seq(x,x,1,5,1)→NSemest

{1,2,3,4,5}

primäre Häufigkeitstabelle:

(absolute/relative Häufigkeiten)

trn(augment(listToMat(NSemest),listToMat(hSemest)))

$$\begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 5 & 9 & 31 & 3 \end{bmatrix}$$

trn(augment(listToMat(NSemest),listToMat(hSemest/51)))

$$\begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ \frac{1}{17} & \frac{5}{51} & \frac{3}{17} & \frac{31}{51} & \frac{1}{17} \end{bmatrix}$$

approx(ans)

$$\begin{bmatrix} 1.00 & 2.00 & 3.00 & 4.00 & 5.00 \\ 0.06 & 0.10 & 0.18 & 0.61 & 0.06 \end{bmatrix}$$

in Prozent:

approx(trn(augment(listToMat(NSemest),listToMat(100*hSemes

$$\begin{bmatrix} 1.00 & 2.00 & 3.00 & 4.00 & 5.00 \\ 5.88 & 9.80 & 17.65 & 60.78 & 5.88 \end{bmatrix}$$

kumuliert:

trn(augment(listToMat(NSemest),listToMat(cuml(hSemest))))

$$\begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 8 & 17 & 48 & 51 \end{bmatrix}$$

trn(augment(listToMat(NSemest),listToMat(cuml(hSemest/51))

$$\begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ \frac{1}{17} & \frac{8}{51} & \frac{1}{3} & \frac{16}{17} & 1 \end{bmatrix}$$

approx(trn(augment(listToMat(NSemest),listToMat(100*cuml(h

$$\begin{bmatrix} 1.00 & 2.00 & 3.00 & 4.00 & 5.00 \\ 5.88 & 15.69 & 33.33 & 94.12 & 100.00 \end{bmatrix}$$

100*cuml(hAlter/51)⇒list5

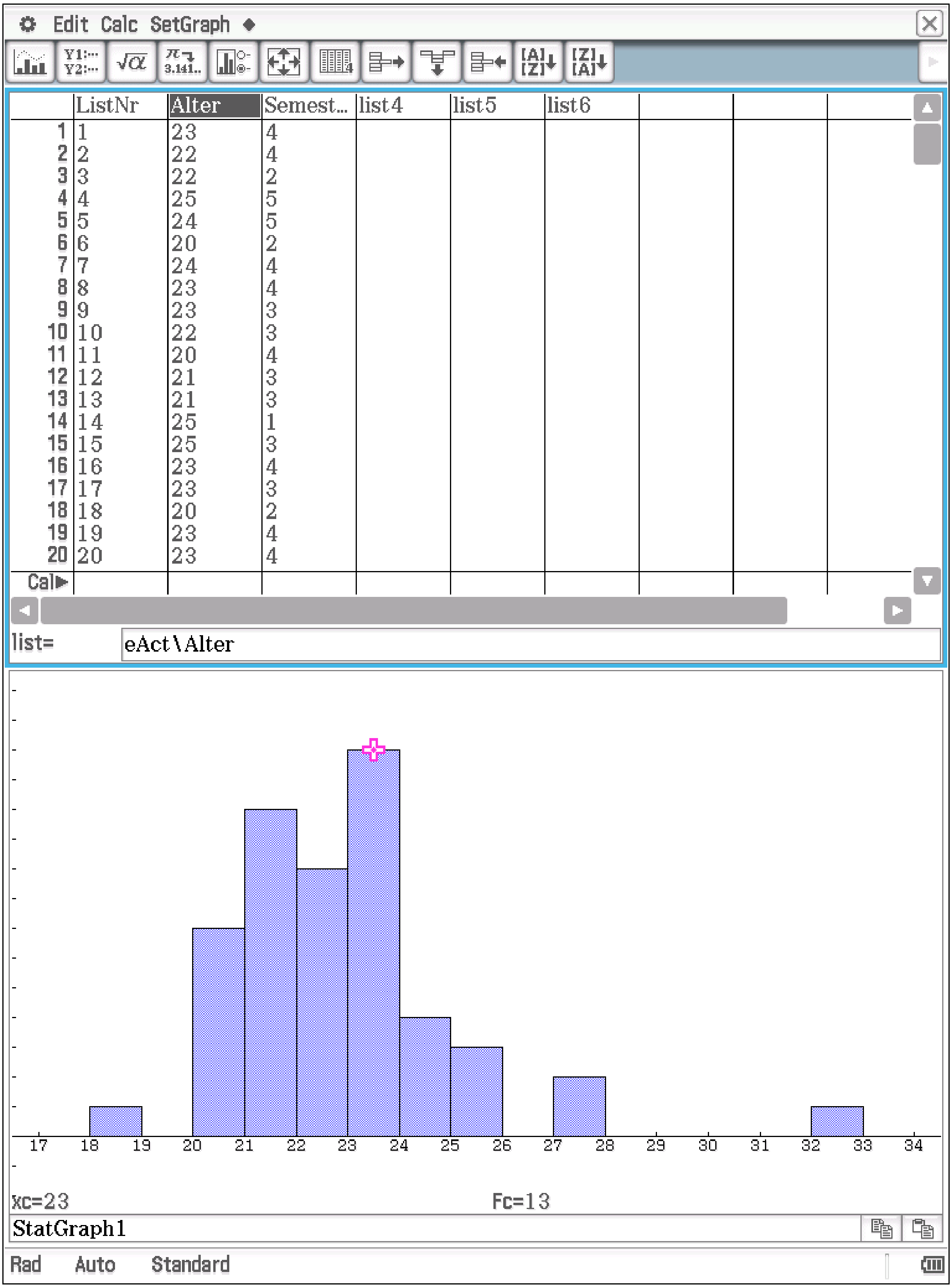
$$\left\{\frac{100}{51}, \frac{800}{51}, \frac{1900}{51}, \frac{2800}{51}, \frac{4100}{51}, \frac{1500}{17}, \frac{1600}{17}, \frac{5000}{51}, 100\right\}$$

100*cuml(hSemest/51)⇒list7

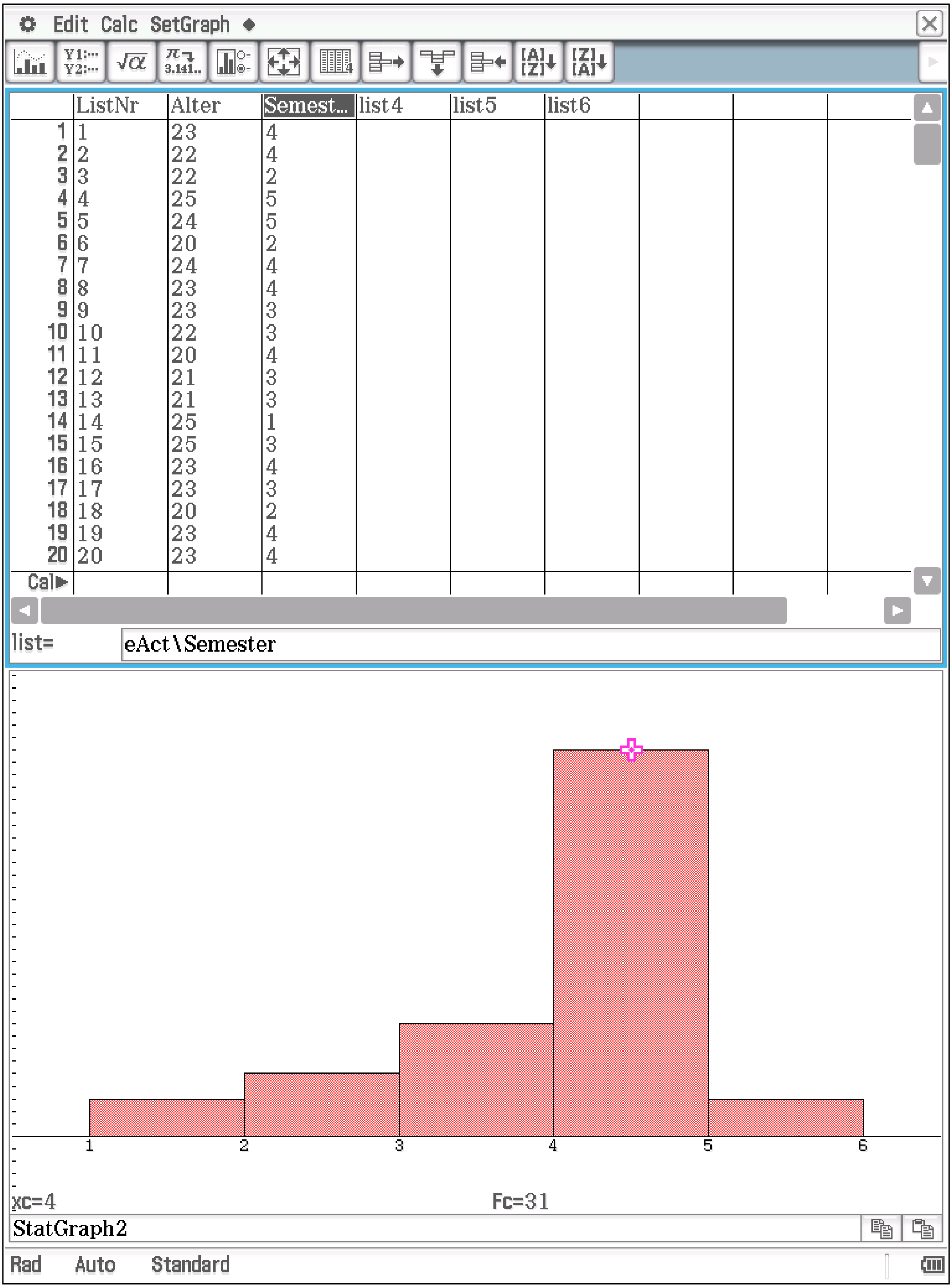
$$\left\{\frac{100}{17}, \frac{800}{51}, \frac{100}{3}, \frac{1600}{17}, 100\right\}$$

Stat-editor

Histogramm Alter:



Histogramm Semester:

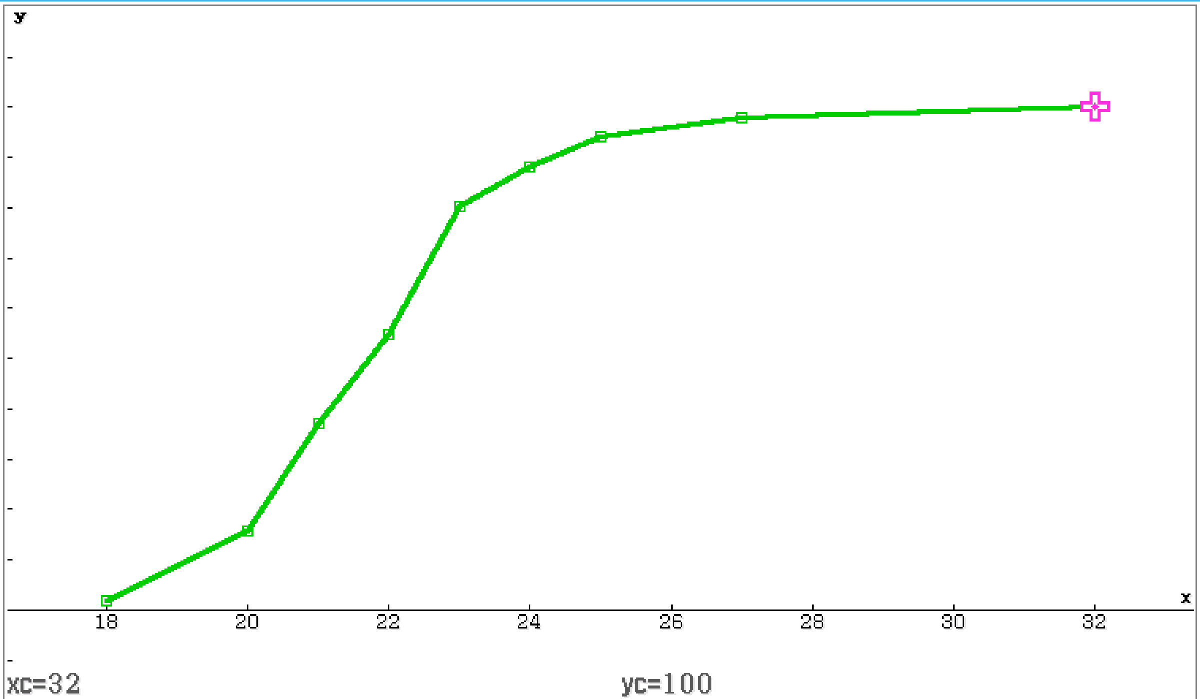


Summenkurve (Alter, bis 100%):

Edit Calc SetGraph

	ListNr	Alter	Semest...	NAlter	list5	NSemest	list7
1	1	23	4	18	1.96	1	5.88
2	2	22	4	20	15.69	2	15.69
3	3	22	2	21	37.26	3	33.33
4	4	25	5	22	54.9	4	94.12
5	5	24	5	23	80.39	5	100
6	6	20	2	24	88.24		
7	7	24	4	25	94.12		
8	8	23	4	27	98.04		
9	9	23	3	32	100		
10	10	22	3				
11	11	20	4				
12	12	21	3				
13	13	21	3				
14	14	25	1				
15	15	25	3				
16	16	23	4				
17	17	23	3				
18	18	20	2				
19	19	23	4				
20	20	23	4				
Cal▶							

list= list5



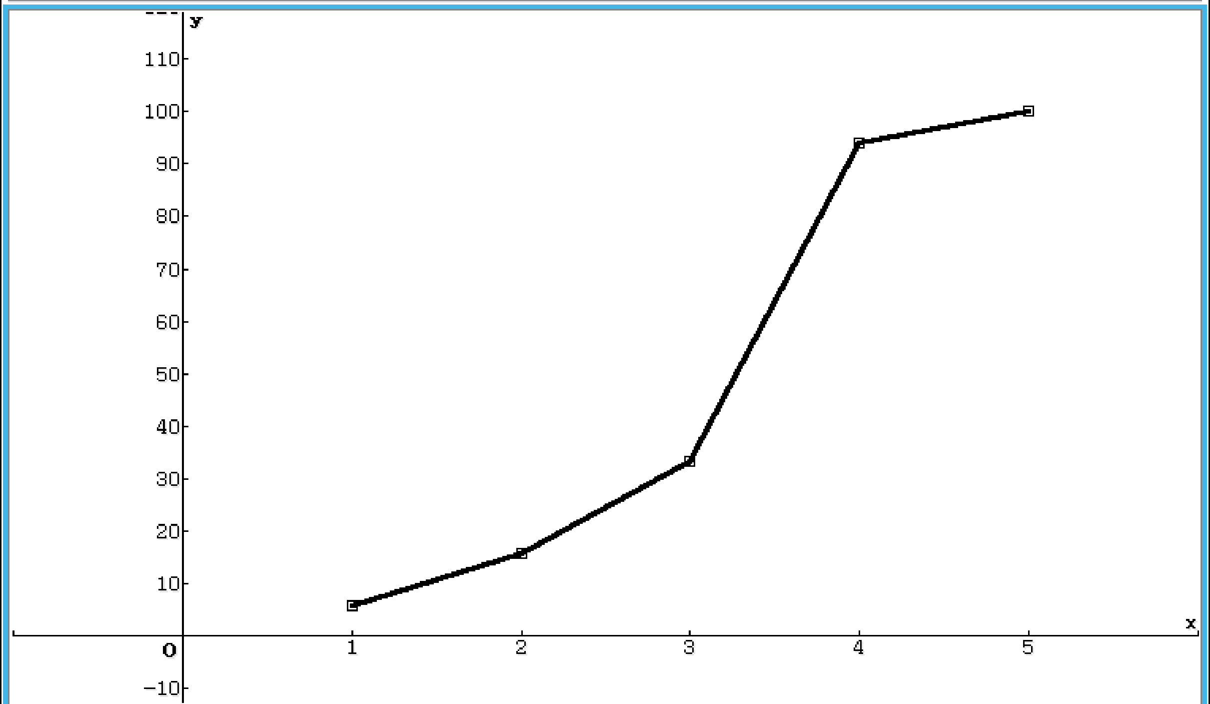
StatGraph3

Rad Auto Standard

Summenkurve (Semester, bis 100%):

	ListNr	Alter	Semest...	NAlter	list5	NSemest	list7
1	1	23	4	18	1.96	1	5.88
2	2	22	4	20	15.69	2	15.69
3	3	22	2	21	37.26	3	33.33
4	4	25	5	22	54.9	4	94.12
5	5	24	5	23	80.39	5	100
6	6	20	2	24	88.24		
7	7	24	4	25	94.12		
8	8	23	4	27	98.04		
9	9	23	3	32	100		
10	10	22	3				
11	11	20	4				
12	12	21	3				
13	13	21	3				
14	14	25	1				
15	15	25	3				
16	16	23	4				
17	17	23	3				
18	18	20	2				
19	19	23	4				
20	20	23	4				

[17]=



Rad

Mittelwerte:

mean(Alter)

22.41

mean(Semester)

3.51

Mediane:

median(Alter)

22

median(Semester)

4

Modalwerte:

mode(Alter)

23

mode(Semester)

4

Spannweite:

max(Alter)−min(Alter)

14

max(Semester)−Min(Semester)

4

Quartilabstand:

Q_3(Alter)−Q_1(Alter)

2

Q_3(Semester)−Q_1(Semester)

1

Varianz:

variance(Alter)

5.05

$\sqrt{ans}$

2.25

variance(Semester)

0.93

$\sqrt{ans}$

0.97

Standardabweichung:

stdDev(Alter)

2.25

stdDev(Semester)

0.97

Variationskoeffizienten:

$$\frac{\text{mean(Alter)}}{\text{stdDev(Alter)}} * 100\%$$

9.98

$$\frac{\text{mean(Semester)}}{\text{stdDev(Semester)}} * 100\%$$

3.63

Zusammenfassung:

================

OneVariable Alter

done

DispStat

done

=====================

One Variable

$\bar{x}$ = 22.41

Σx = 1143

$\Sigma x^2 = 25869$

$\sigma_x = 2.22$

$s_x = 2.25$

$n = 51$

$\text{minX} = 18$

$Q_1 = 21$

$\text{Med} = 22$

$Q_3 = 23$

$\text{maxX} = 32$

$\text{Mode} = 23$

$\text{ModeN} = 1$

$\text{ModeF} = 13$

======================

OneVariable Semester

done

DispStat

done

======================

One Variable

$\bar{x} = 3.51$

$\Sigma x = 179$

$\Sigma x^2 = 675$

$\sigma_x = 0.96$

$s_x = 0.97$

$n = 51$

$\text{minX} = 1$

$Q_1 = 3$

$$Med = 4$$

$$Q_3 = 4$$

$$maxX = 5$$

$$Mode = 4$$

$$ModeN = 1$$

$$ModeF = 31$$

======================

Prof. Dr. Ludwig Paditz, 17.10.2016

Statistik für Produktionstechniker

==============================

2. Die Größe (in cm) von 200 achtjährigen Kindern einer Stadt wurde gemessen.

Die Ergebnisse sind aus der beiliegenden Tabelle ersichtlich:

seq(x,x,121,152,1)⇒Größe

{121,122,123,124,125,126,127,128,129,130,131,132,13▶

Spannweite:

152−121

31

{2,2,0,0,4,0,6,4,6,12,4,8,6,4,14,4,6,8,6,12,8,12,10▶

{2,2,0,0,4,0,6,4,6,12,4,8,6,4,14,4,6,8,6,12,8,12,10▶

sum(ans)

200

trn(augment(listToMat(Größe),listToMat(Anzahl)))

121	122	123	124	125	126	127	128	129	130	131	132	133	▶
2	2	0	0	4	0	6	4	6	12	4	8	6	

bereinigt um nichtauftretende Größen:

121	122	125	127	128	129	130	131	132	133	134	135	136	▶
2	2	4	6	4	6	12	4	8	6	4	14	4	

121	122	125	127	128	129	130	131	132	133	134	135	136	▶
2	2	4	6	4	6	12	4	8	6	4	14	4	

(a) Man fasse die Daten jeweils nach folgenden Klasseneinteilungen zusammen:

(i) Reduktionslage 119.5, Klassenbreite 5, Klassenanzahl 7

[119.5; 124.5); [124.5; 129.5); [129.5; 134.5); [134.5; 139.5); [139.5; 144.5); [144.5; 149.5); [149.5; 154.5)

(ii) Reduktionslage 120.5, Klassenbreite 5, Klassenanzahl 7

[120.5; 125.5); [125.5; 130.5); [130.5; 135.5); [135.5; 140.5); [140.5; 145.5); [145.5; 150.5); [150.5; 155.5)

(b) Man stelle die Klassen in einem Säulendiagramm dar. (Histogramm)

(c) Man bestimme jeweils die Klassenmitten und berechne das gewichtete Mittel der Klassenmitten.

```
matToList(trn([121 122 125 127 128 129 130 131 132 133▶
    {121,122,125,127,128,129,130,131,132,133,134,135,13▶
matToList(trn([2 2 4 6 4 6 12 4 8 6 4 14 4 6 8 6 12 8 1▶
    {2,2,4,6,4,6,12,4,8,6,4,14,4,6,8,6,12,8,12,10,4,16,▶
```

Stat-Editor

Klassenmitten und Häufigkeiten aus dem Histogramm

(i)

```
seq(x,x,119.5+2.5,149.5+2.5,5)➔iMitte
                              {122,127,132,137,142,147,152}
{4,20,34,38,46,52,6}➔hlisti
                                       {4,20,34,38,46,52,6}
```

mean(iMitte, hlisti)

139.05

(ii)

seq(x, x, 120.5+2.5, 150.5+2.5, 5)⇒iiMitte

{123, 128, 133, 138, 143, 148, 153}

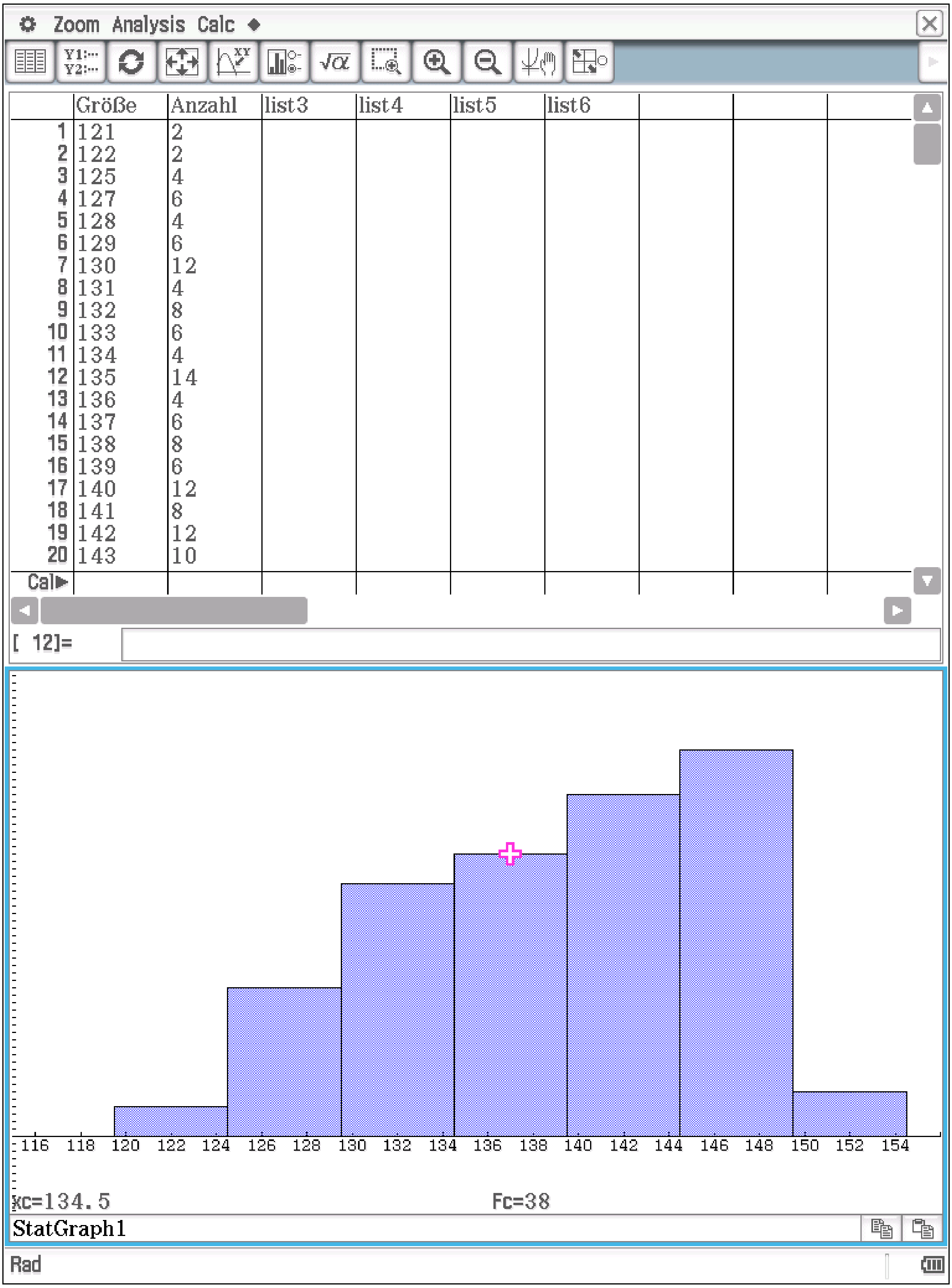
{8, 28, 36, 36, 50, 40, 2}⇒hlistii

{8, 28, 36, 36, 50, 40, 2}

mean(iiMitte, hlistii)

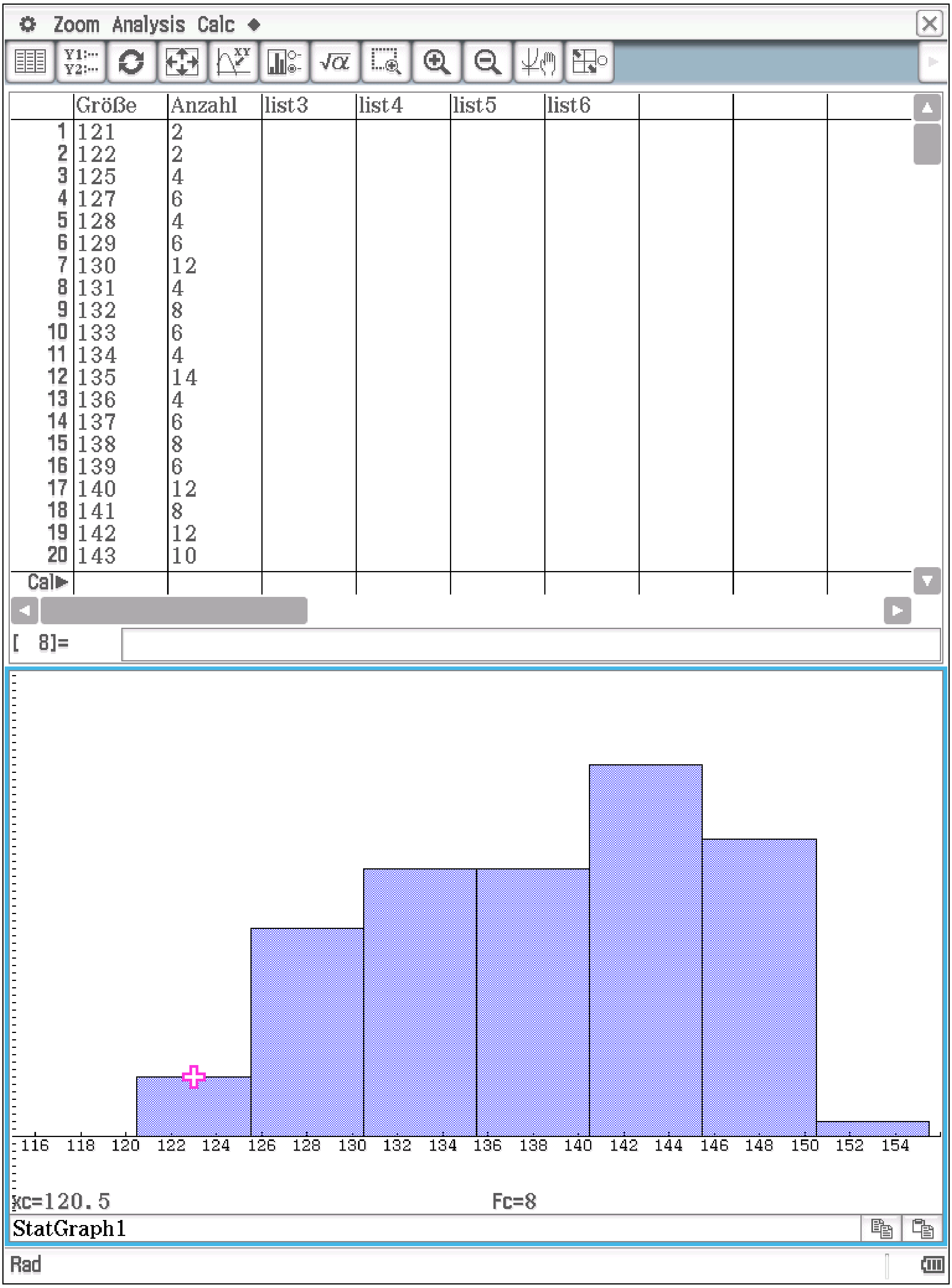
138.5

Histogramm (i):



	Größe	Anzahl	list3	list4	list5	list6			
1	121	2							
2	122	2							
3	125	4							
4	127	6							
5	128	4							
6	129	6							
7	130	12							
8	131	4							
9	132	8							
10	133	6							
11	134	4							
12	135	14							
13	136	4							
14	137	6							
15	138	8							
16	139	6							
17	140	12							
18	141	8							
19	142	12							
20	143	10							
Cal▶									

Histogramm (ii):



	Größe	Anzahl	list3	list4	list5	list6
1	121	2				
2	122	2				
3	125	4				
4	127	6				
5	128	4				
6	129	6				
7	130	12				
8	131	4				
9	132	8				
10	133	6				
11	134	4				
12	135	14				
13	136	4				
14	137	6				
15	138	8				
16	139	6				
17	140	12				
18	141	8				
19	142	12				
20	143	10				
Cal▶						

Prof. Dr. Ludwig Paditz, 17.10.2016

Statistik für Produktionstechniker

==============================

3. Es sei $\bar{x}$ das arithmetische Mittel der reellen Zahlen x_1, ..., x_n, $n \in \mathbf{N}$.

Man beweise, dass $\bar{x}$ die folgende Funktion minimiert:

$$f(z) = \sum_{i=1}^{n} \left((x_i - z)^2 \right)$$

Lösung:

1. Ableitung von f(z)

$$f'(z) = \sum_{i=1}^{n} (2 * (x_i - z) * (-1)) = 0$$

ergibt:

$$\sum_{i=1}^{n} (x_i - z) = 0 \Rightarrow n * \bar{x} = n * z$$

Somit $z = \bar{x}$

2. Ableitung von f(z)

$$\sum_{i=1}^{n} \left(2 * x_i * (-1)^2 \right) = 2n * \bar{x} > 0 \text{ für } \bar{x} > 0, \text{ d.h. min!}$$

4. Welches Skalenniveau können Daten zu folgenden Merkmalen maximal aufweisen?

(a) Klausurpunktzahl – metrisch/Absolut (Noten dagegen ordinal)

(b) Kinderzahl – metrisch/Absolut

(c) Intelligenz – ordinal

(d) Einkommen – metrisch/Absolut

(e) Preis einer Ware – metrisch/Absolut

(f) Geburtsjahr – metrisch/Absolut

(g) Kraftstoffverbrauch eines PKW auf 100km – metrisch/Absolut

(h) Erlernter Beruf – nominal

(i) Familienstand – nominal

(j) Güteklasse – ordinal

(k) Seitenzahl eines Buches – metrisch/Absolut

(l) Fahrpreise – metrisch/Absolut

(m) Freizeitbeschäftigung – nominal

(n) Nationalität – nominal

(o) Handelsklasse bei Obst – ordinal

(p) Körpergröße – metrisch/Absolut

Bem.:

metrisch/Absolut bedeutet: "metrische Daten mit absolutem Nullpunkt" (Verhältnisskala)

abgeschwächt: "metrische Daten ohne absoluten Nullpunkt" (Intervallskala)

Prof. Dr. Ludwig Paditz, 17.10.2016

Statistik für Produktionstechniker
==============================

5. In einem Unternehmen entwickelten sich die jährlichen Umsätze über sechs Jahre
wie folgt: Nach einem Umsatzeinbruch um 50% im ersten Jahr stiegen die Umsätze
in den nächsten fünf Jahren um jährlich 10%.

(a) Berechnen Sie die durchschnittliche jährliche prozentuale Umsatzentwicklung
des Unternehmens für die beschriebenen sechs Jahre.

(b) Im Jahr vor der beschriebenen Entwicklung hatte das Unternehmen einen
Umsatz von 1 Mio. €. Wie groß war der Umsatz im sechsten Jahr danach?
Vergleichen Sie das Ergebnis mit dem von (a).

Lösung:

seq(x, x, 1, 6, 1)→Jahr

$$\{1,2,3,4,5,6\}$$

$\mathrm{seq}(\frac{U}{2}, x, 1, 6, 1)$

$$\left\{\frac{U}{2},\frac{U}{2},\frac{U}{2},\frac{U}{2},\frac{U}{2},\frac{U}{2}\right\}$$

$\left\{\frac{U}{2}, \frac{U}{2}*1.1, \frac{U}{2}1.1^2, \frac{U}{2}1.1^3, \frac{U}{2}1.1^4, \frac{U}{2}1.1^5\right\}$ ⇒Umsatz

$$\left\{\frac{U}{2}, \frac{11\cdot U}{20}, \frac{121\cdot U}{200}, \frac{1331\cdot U}{2000}, \frac{14641\cdot U}{20000}, \frac{161051\cdot U}{200000}\right\}$$

(a) geometrisches Mittel:

$\sqrt[6]{\frac{1}{2}1.1^5}$

0.964544432

seq(U*0.964544432^x, x, 1, 6, 1)

{0.964544432·U, 0.9303459613·U, 0.8973600168·U, 0.8655▶

Die Umsätze sanken durchschnittlich um jährlich 3,55%:

solve(0.964544432·U*x=0.9303459613·U, x)

{x=0.964544432}

1−0.964544432

0.035455568

seq(U*0.964544432^x, x, 1, 6, 1)|U=10^6

{964544.432, 930345.9613, 897360.0168, 865543.6077, 834▶

Nach 6 Jahren betrug der Umsatz nur noch 805255€.

6. Ein Eisenbahnzug fährt ohne Zwischenhalte mit folgenden Geschwindigkeiten:
200km mit 200km/h, 150km mit 120km/h, 60km mit 80km/h und 10km
mit 30km/h. Berechnen Sie die Durchschnittsgeschwindigkeit des Zuges auf dieser
Strecke. Phasen der Beschleunigung und des Bremsens sollen der Einfachheit halber
vernachlässigt werden.

Lösung:

Gesamtstrecke:

sum({200,150,60,10})

420

Physik: $v=\frac{s}{t} \Rightarrow t=\frac{s}{v}$

Zeiten:

$sum(\{\frac{200}{200},\frac{150}{120},\frac{60}{80},\frac{10}{30}\})$

3.333333333

$v:=\frac{420}{ans}$

126

allgemein: (harmonisches Mittel)

$$v=\frac{\text{Gesamtstrecke}}{\text{Gesamtzeit}}=\frac{sum(s_i)}{sum(\frac{s_i}{v_i})}$$

7. Die Messungen der Körperlänge bei 17- bis 18-jährigen Mädchen und bei 6-jährigen Mädchen ergaben folgende Werte:

$$\begin{bmatrix} & n & \bar{x} & s \\ \text{6-jährige Mädchen} & 77 & 112.6 & 4.64 \\ \text{17- bis 18-jährige Mädchen} & 51 & 162.6 & 5.12 \end{bmatrix}$$

Berechnen Sie für beide Messreihen den Variationskoeffzienten und vergleichen Sie

die Variabilität der Körperlänge in den beiden Gruppen.

Lösung:

Variationskoeffizient $v=\frac{s}{\bar{x}}*100\%$

6-jährige Mädchen:

$v:=\frac{4.64}{112.6}*100$

4.120781528

17- bis 18-jährige Mädchen:

s größer aber v kleiner!

$v:=\frac{5.12}{162.6}*100$

3.148831488

Prof. Dr. Ludwig Paditz, 17.10.2016

Statistik für Produktionstechniker

==============================

8. Bei 150 Akten zu Steuerhinterziehungen fand man die folgende gemeinsame Häufgkeitsverteilung der Variablen X(Schulabschluss) und Y(Entdeckungszeitraum).

X/Y				
	1	2	3	Σ
1	36	30	□	120
2	□	□	16	□
Σ	40	□	□	150

Kodierung:

X=1 Volksschule, X=2 höhere Schule

Y=1 bereits beim ersten Versuch entdeckt

Y=2 innerhalb eines Jahres entdeckt

Y=3 nach einem Jahr entdeckt

(a) Vervollständigen Sie die Tabelle.

(b) Ist der Anteil der erst nach einem Jahr entdeckten Steuerhinterziehungen bei den Tätern mit höherem Schulabschluss größer als bei den Volksschülern?

(c) Wie groß ist der Anteil der spätestens bis Ende eines Jahres ertappten Steuersünder?

(d) Haben die Volksschüler unter denen, die bereits beim Versuch ertappt wurden, einen größeren Anteil als unter allen Steuersündern?

(e) Berechnen Sie Prozentsätze in den Zeilen und stellen Sie diese in Form gestapelter Balkendiagramme für die Kategorien von X dar! Kann man aus dieser Darstellung auf eine Abhängigkeit des Entdeckungszeitraumes vom Schulabschluss schließen?

(f) Berechnen Sie ausgehend von den marginalen (Rand-) Verteilungen die Anzahl in den einzelnen Zellen, die sich im Falle der empirischen Unabhängigkeit ergeben würde (Indifferenztabelle).

(g) Berechnen Sie den Wert (Chi-Quadrat)

$$\chi^2 = \Sigma_{i,j} \frac{(n_{ij} - \hat{n}_{ij})^2}{\hat{n}_{ij}}$$

als Maß für die Abweichung der beobachteten Tabelle von der Tabelle bei Unabhängigkeit (Indifferenztabelle).

Lösung:

Kreuztabelle (Kontingenztafel)

$$\text{(a)} \quad \begin{bmatrix} X/Y & & & & \\ & 1 & 2 & 3 & \Sigma \\ 1 & 36 & 30 & 54 & 120 \\ 2 & 4 & \square & 16 & 30 \\ \Sigma & 40 & \square & \square & 150 \end{bmatrix} \Rightarrow \begin{bmatrix} X/Y & & & & \\ & 1 & 2 & 3 & \Sigma \\ 1 & 36 & 30 & 54 & 120 \\ 2 & 4 & 10 & 16 & 30 \\ \Sigma & 40 & 40 & 70 & 150 \end{bmatrix}$$

(b) $\frac{54}{120}=0.45$ zu $\frac{16}{30}=\frac{64}{120}=0.533$ Antwort: Ja!

(c) $\frac{40+40}{150}=\frac{80}{150}=53{,}33\%$

$\frac{80}{150}$

0.5333333333

(d) $\frac{36}{40}=0.9$ zu $\frac{120}{150}=0.8$ Antwort: Ja!

(e) Prozentsätze, gestapeltes Balkendiagramm, gruppierte Balken

$$\begin{bmatrix} 36*\frac{100}{120} & 30*\frac{100}{120} & 54*\frac{100}{120} & 120*\frac{100}{120} \\ 4*\frac{100}{30} & 10*\frac{100}{30} & 16*\frac{100}{30} & 30*\frac{100}{30} \\ 40*\frac{100}{150} & 40*\frac{100}{150} & 70*\frac{100}{150} & 150*\frac{100}{150} \end{bmatrix}$$

$$\begin{bmatrix} 30 & 25 & 45 & 100 \\ 13.33333333 & 33.33333333 & 53.33333333 & 100 \\ 26.66666667 & 26.66666667 & 46.66666667 & 100 \end{bmatrix}$$

fRound(ans, 2)

$$\begin{bmatrix} 30 & 25 & 45 & 100 \\ 13.33 & 33.33 & 53.33 & 100 \\ 26.67 & 26.67 & 46.67 & 100 \end{bmatrix}$$

$$\begin{bmatrix} X/Y & & & & \\ & 1 & 2 & 3 & \Sigma \\ 1 & 36 & 30 & 54 & 120 \\ 2 & 4 & 10 & 16 & 30 \\ \Sigma & 40 & 40 & 70 & 150 \end{bmatrix} \Rightarrow \begin{bmatrix} X/Y & & & & \\ & Y=1 & Y=2 & Y=3 & \Sigma \\ X=1 & 36 & 30 & 54 & 120 \\ & 30 & 25 & 45 & 100 \\ X=2 & 4 & 10 & 16 & 30 \\ & 13.33 & 33.33 & 53.33 & 100 \\ \Sigma & 40 & 40 & 70 & 150 \\ & 26.67 & 26.67 & 46.67 & 100 \end{bmatrix}$$

TK = Tabellenkalkulation

Grafik im TK-Menü

(f) Indifferenztabelle mit $n\hat{\ }_{ij}=\frac{n_{i.}*n_{.j}}{n_{..}}$

$$\begin{bmatrix} X/Y & & & & \\ & 1 & 2 & 3 & \Sigma \\ 1 & 32 & 32 & 56 & 120 \\ 2 & 8 & 8 & 14 & 30 \\ \Sigma & 40 & 40 & 70 & 150 \end{bmatrix}$$

(g) Formel und Berechnung

$$\begin{bmatrix} 36 & 30 & 54 \\ 4 & 10 & 16 \end{bmatrix} \Rightarrow A$$

$$\begin{bmatrix} 36 & 30 & 54 \\ 4 & 10 & 16 \end{bmatrix}$$

Stat-Editor (Chi^2-Test)

ChiTest A

done

DispStat

done

==================

χ^2 Test

χ^2 = 3.4821429

prob = 0.1753324

df = 2

==================

A

$$\begin{bmatrix} 36 & 30 & 54 \\ 4 & 10 & 16 \end{bmatrix}$$

Expected

$$\begin{bmatrix} 32 & 32 & 56 \\ 8 & 8 & 14 \end{bmatrix}$$

Matrix der erwarteten Häufigkeiten im Fall der Unabhängigkeit der Merkmale X, Y.

A−Expected

$$\begin{bmatrix} 4 & -2 & -2 \\ -4 & 2 & 2 \end{bmatrix}$$

χ^2value

3.482142857

$$\chi^2 := \frac{16}{32}+\frac{4}{32}+\frac{4}{56}+\frac{16}{8}+\frac{4}{8}+\frac{4}{14}$$

3.482142857

Kontingenzkoeffizient: nach Pearson

$C=\sqrt{\frac{\chi^2}{\chi^2+n}}$ mit $n=150$

$$C=\sqrt{\frac{\chi^2}{\chi^2+150}}$$

C=0.1506240654

Kontingenzkoeffizient: nach Cramer (genauer als C)

s berücksichtigt die Anzahl der Merkmalsausprägungen für X, Y

$K=\sqrt{\frac{\chi^2}{n*(s-1)}}$ mit n=150, s=min(2,3)=2

$K=\sqrt{\frac{\chi^2}{150*1}}$

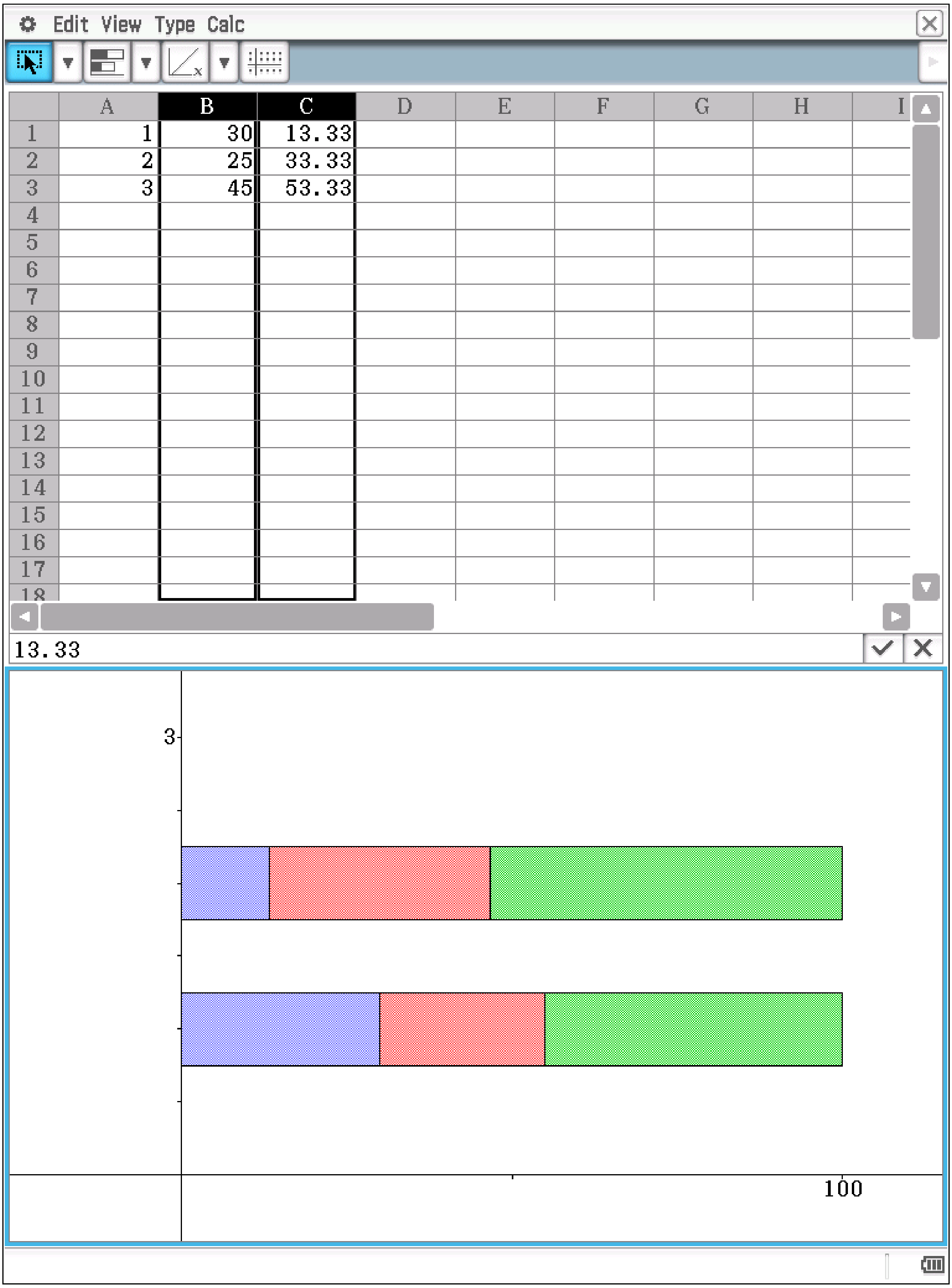
K=0.1523623501

https://de.wikipedia.org/wiki/Kontingenzkoeffizient

Balkendiagramm (gestapelt):

A: Merkmale für Y=1,2,3,

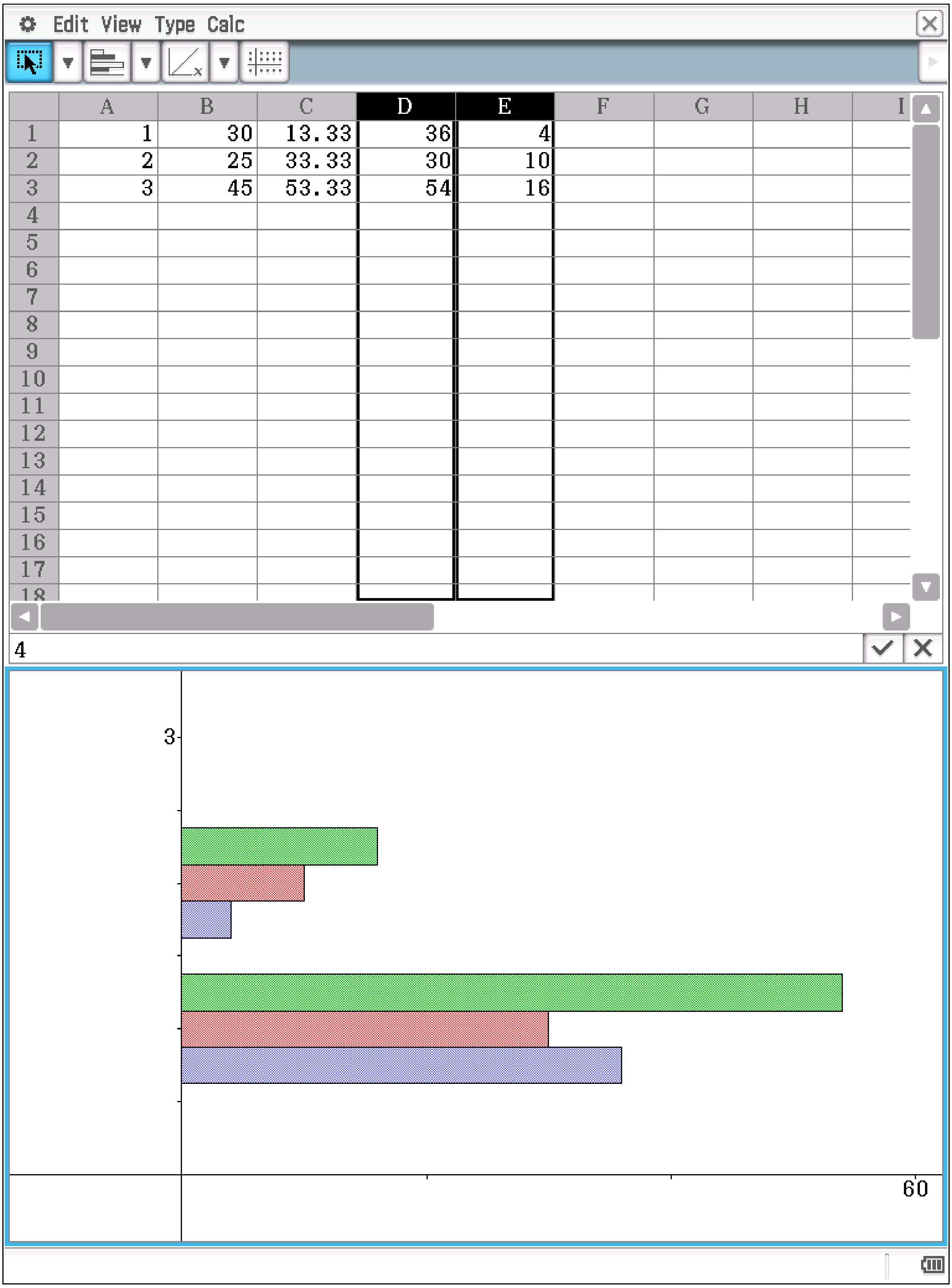
B: Prozente für X=1 (unterer Balken), C: Prozente für X=2 (oberer Balken)



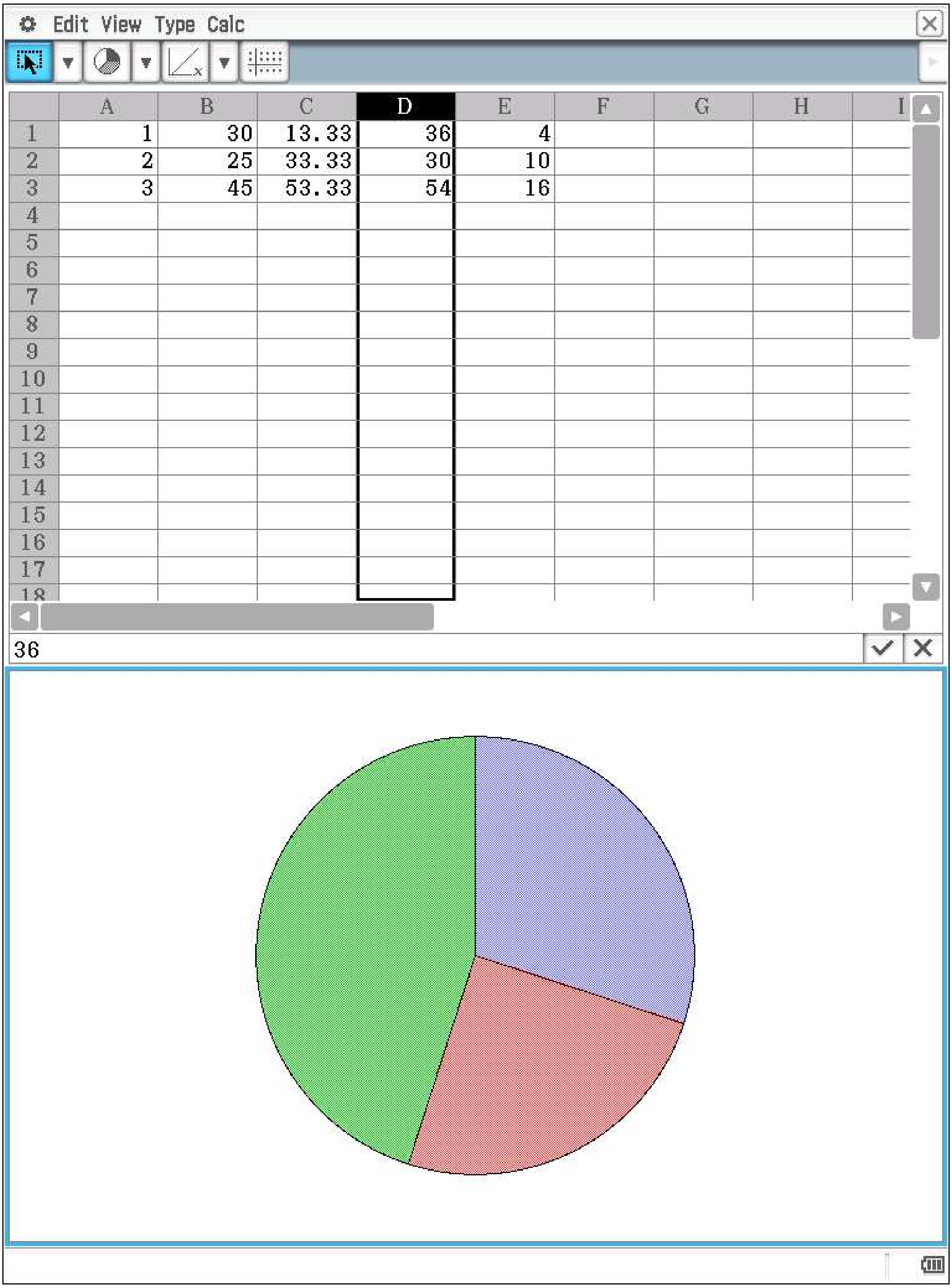
Balkendiagramm (gruppiert):

A: Merkmale für Y=1,2,3,

D: Anzahlen für X=1 (untere Balken), E: Anzahlen für X=2 (obere Balken)

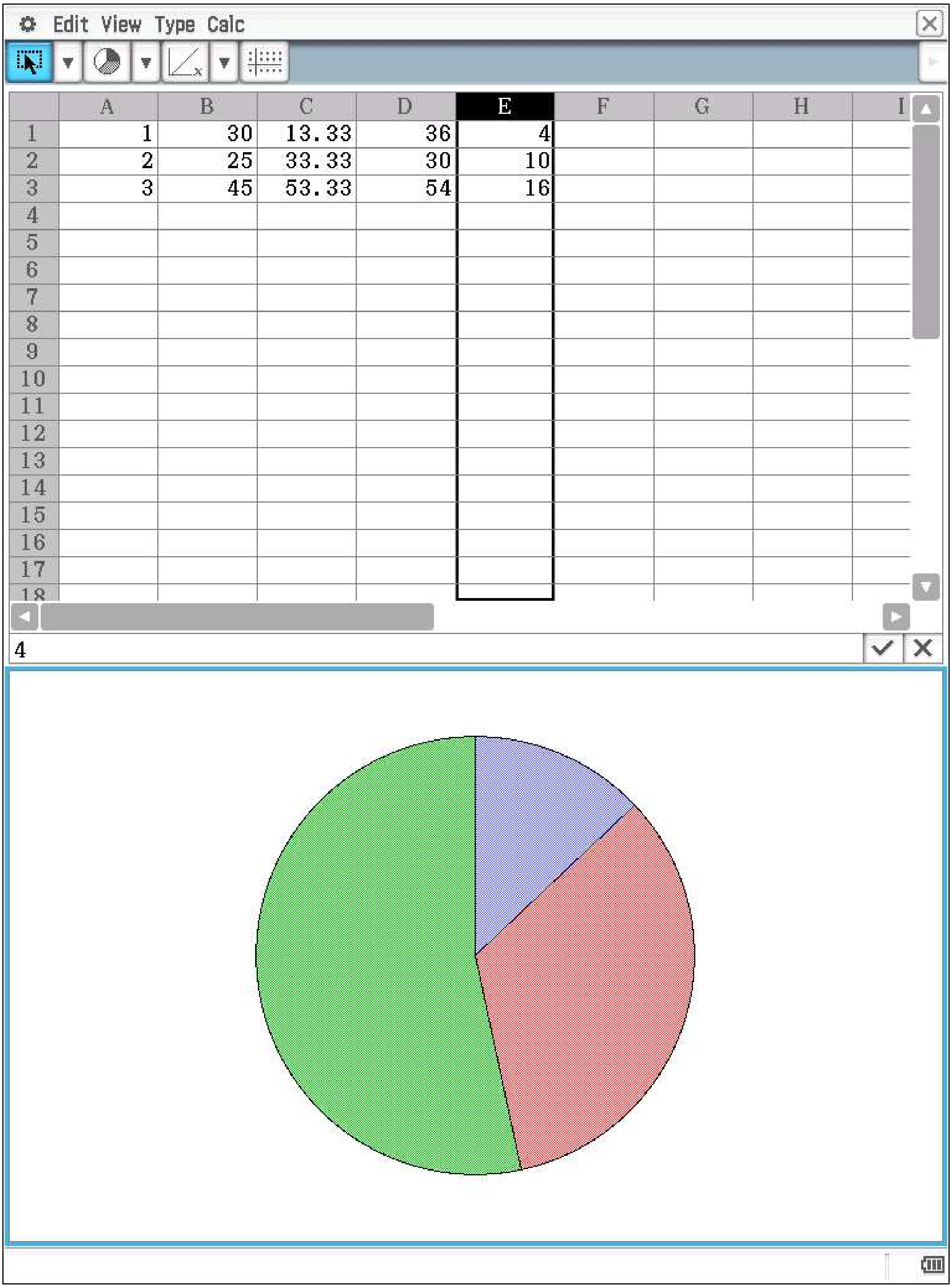


Kreisdiagramm für X=1



	A	B	C	D	E
1	1	30	13.33	36	4
2	2	25	33.33	30	10
3	3	45	53.33	54	16

Kreisdiagramm für X=2



	A	B	C	D	E
1	1	30	13.33	36	4
2	2	25	33.33	30	10
3	3	45	53.33	54	16