Prof. Dr. Ludwig Paditz 03.11.2014

1.Ü STAT 1.1

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

vgl. LÜB Bd.3, Kapitel 17 S.230ff

Dateneingabe:

seq(x,x,7,19,1)⇒xliste

{7,8,9,10,11,12,13,14,15,16,17,18,19}

{1,1,2,2,6,10,7,11,7,7,3,2,1}⇒hliste

{1,1,2,2,6,10,7,11,7,7,3,2,1}

Datenauswertung in der primären Häufigkeitstabelle:

OneVariable xliste,hliste

done

DispStat

done

Hinw.: im Grundformat Q_1 und Q_3 aktivieren

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

Eindim. Variable

$\bar{x}$ = 13.45

Σx = 807

Σx^2 = 11223

σ_X = 2.4794153 (Normierung mit n)

s_X = 2.500339 (Normierung mit n−1)

n = 60

minX = 7

Q_1 = 12 (1.Quartil)

Med = 14 (Median, Zentralwert)

Q_3 = 15 (3. Quartil)

maxX = 19

Mode = 14 (der wahrscheinlichste Wert,
Datenwert mit maximaler Häufigkeit)

ModeN = 1 (Anzahl an Modalwerten)

ModeF = 11 (Häufigkeit des Modalwertes)

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

OneSampleTInt 0.95, xliste, hliste

done

DispStat

done

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

1-Stichprob. t-Int.

Daten=Liste

Lower = 12.804094

Upper = 14.095906

$\bar{x}$ = 13.45

s_x = 2.500339

n = 60

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

Berechnung des t-Wertes mit dem Befehl invTCDf(1-q, M)

mit $1-q=\frac{1-S}{2}$ und M=n-1

z.B. S=0.95, dann 1-q=0.025, n=60 ergibt M=59

invTCDf(0.025, 59)

2.000995378

Damit erhält man die Grenzen für das Vertrauensintervall,
vgl. **Taschenbuch** der Wirtschaftsmathematik S. 325

$13.45+\frac{2.500339}{\sqrt{60}}*2.000995378$

14.09590605

$13.45-\frac{2.500339}{\sqrt{60}}*2.000995378$

12.80409395

Alternativ kann man den t-Wert in der Tabelle im Taschenbuch der Wirtschaftsmathematik S. 377 ablesen bzw. durch Interpolation ermitteln:

M=60 hat t=2,000
M=50 hat t=2,009
Damit hat M=59 etwas mehr als 2.000 und zwar

$\frac{2.009-2.000}{10}=0.0009,$

d.h. $t=2.0009\approx 2.00$

Die t-Werte gehen auf **Gosset** zurück, der diese 1908 unter dem Pseudonym **"Student"** veröffentlichte, vgl.
http://de.wikipedia.org/wiki/William_Sealy_Gosset
Student: **"The Probable Error of a Mean"**
In: Biometrika. Band 6 Heft 1. 1908. S. 1-25

Datenauswertung in der sekundären Häufigkeitstabelle:

seq(x, x, 7.5, 19.5, 2)→xlistes

{7.5, 9.5, 11.5, 13.5, 15.5, 17.5, 19.5}

{2, 4, 16, 18, 14, 5, 1}→hlistes

{2, 4, 16, 18, 14, 5, 1}

OneVariable xlistes, hlistes

done

DispStat

done

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

Eindim. Variable

$\bar{x}$ = 13.4

Σx = 804

Σx^2 = 11145

σ_x = 2.4879711 (Normierung mit n)

s_x = 2.508967 (Normierung mit n-1)

n = 60

minX = 7.5

Q_1 = 11.5 (1.Quartil)

Med = 13.5 (Median, Zentralwert)

Q_3 = 15.5 (3.Quartil)

maxX = 19.5

Mode = 13.5 (der wahrscheinlichste Wert,
Datenwert mit maximaler Häufigkeit)

ModeN = 1 (Anzahl an Modalwerten)

ModeF = 18 (Häufigkeit des Modalwertes)

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

Damit erhält man die Grenzen für das Vertrauensintervall,

$$13.40+\frac{2.508967}{\sqrt{60}}*2.000995378$$

14.0481349

$$13.40-\frac{2.508967}{\sqrt{60}}*2.000995378$$

12.7518651

Prof. Dr. Ludwig Paditz 28.10.2014

1.HA STAT 1.2

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

vgl. LÜB Bd.3, Aufg.17.3 S.240 und S.334ff

Lösung als eActivity generiert:

Dateneingabe per Hand: **Urdatenliste**

{499,484,493,487,500,493,504,507,485,501,495,497,49▶

{499,484,493,487,500,493,504,507,485,501,495,497,49▶

dim(xliste)

40

Aufsteigende Sortierung der Daten: **Variationsreihe**

sortA(xliste)⇒xlistea

{484,485,486,487,487,488,488,490,491,492,493,493,49▶

Statistik–Editor

Wird die Zeile "Statistik–Editor" angeklickt, öffnet sich ein Hintergrundfenster mit den Datenlisten als Spalten (analog zu SPSS) und einem Untermenü zur Definition statistischer Grafiken.

Je kleiner die Schrittweite, desto unruhiger wirkt das Histogramm (kleine Säulen in der Mitte, vgl. 1.Übung).

Schrittweite 1 nicht sinnvoll.

Definition der Treppenkurve der rel. Summenhäufigkeiten (Sprung am Klassenende):

$$\text{Define } y1(x)=\begin{cases} 0, & -\infty<x<485 \\ 1/40, & 485\le x<490 \\ 7/40, & 490\le x<495 \\ 17/40, & 495\le x<500 \\ 25/40, & 500\le x<505 \\ 34/40, & 505\le x<510 \\ 38/40, & 510\le x<515 \\ 1, & 515\le x<\infty \end{cases}$$

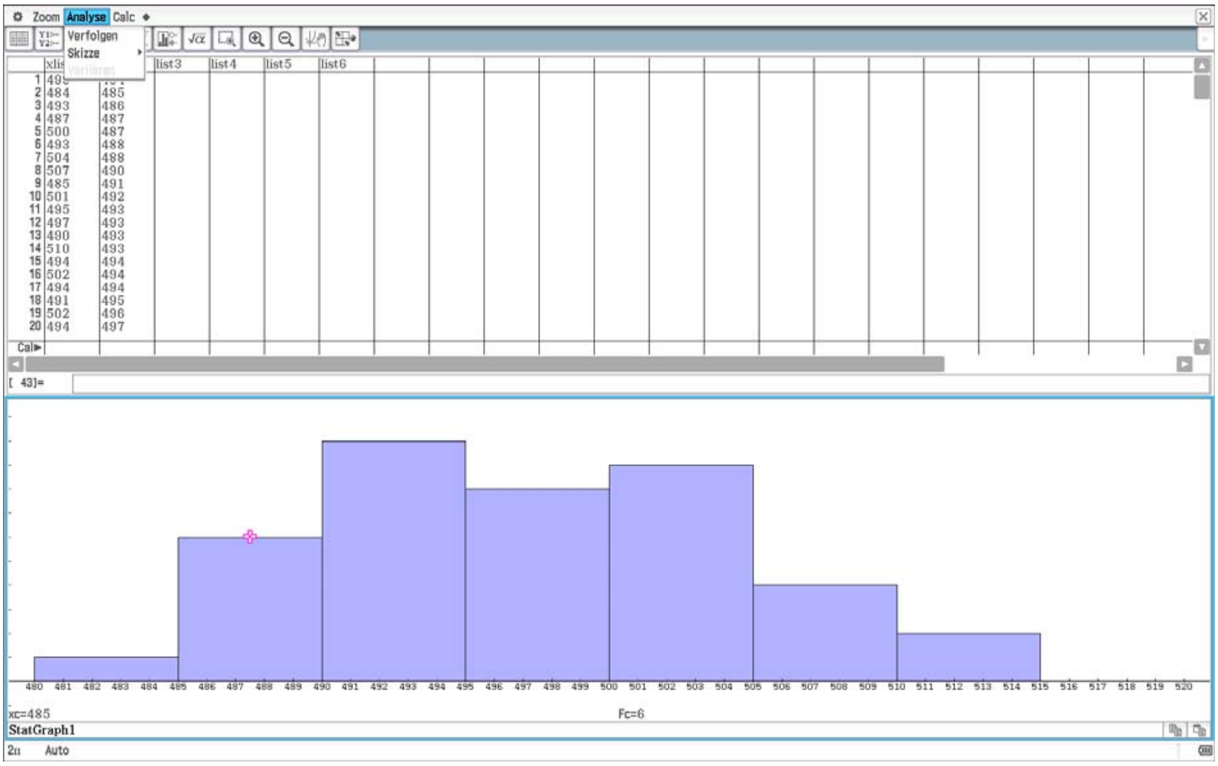
done

Bei Klassenbreite 1 ist die Treppenfunktion die (rechtsseitig stetige) **empirische Verteilungsfunktion.**

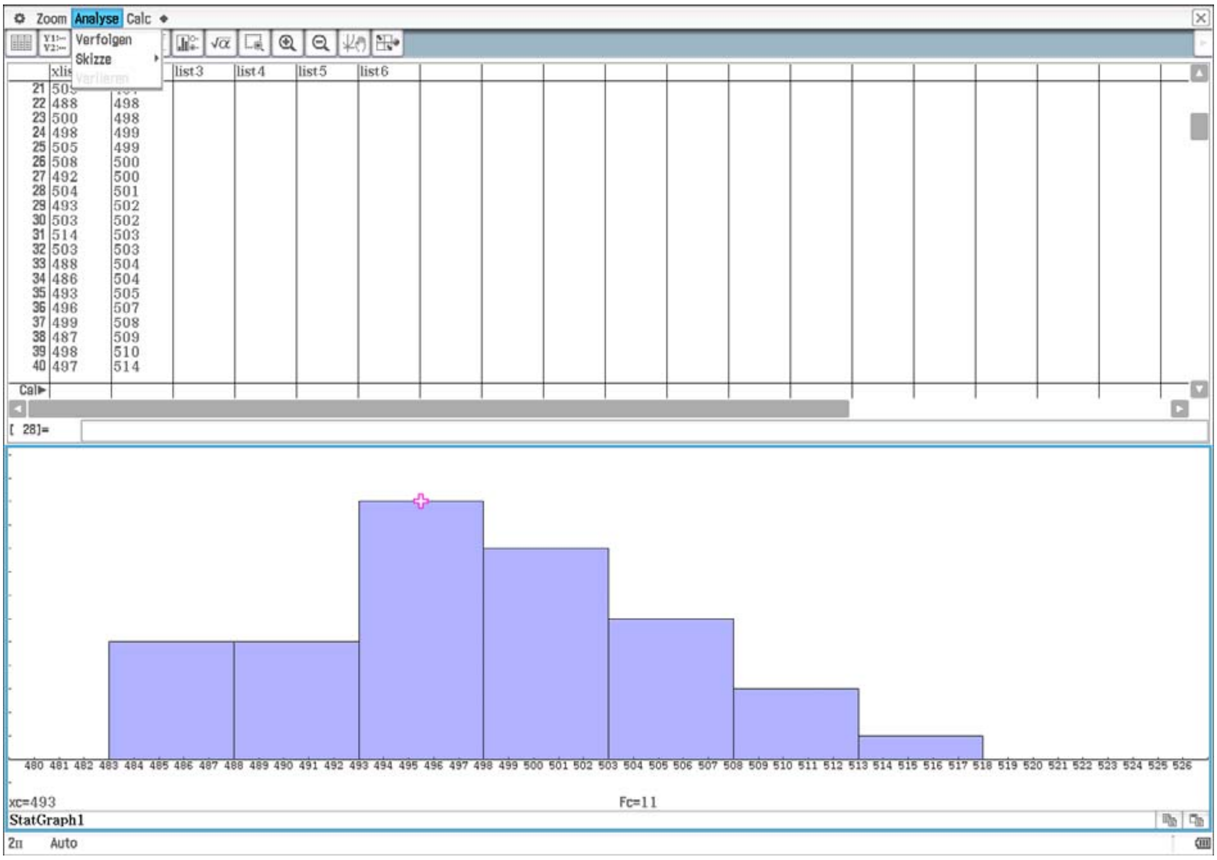
2D-Grafik einer Treppenfunktion Y1:… Y2:…

Statistische Grafiken zu STAT 1.2

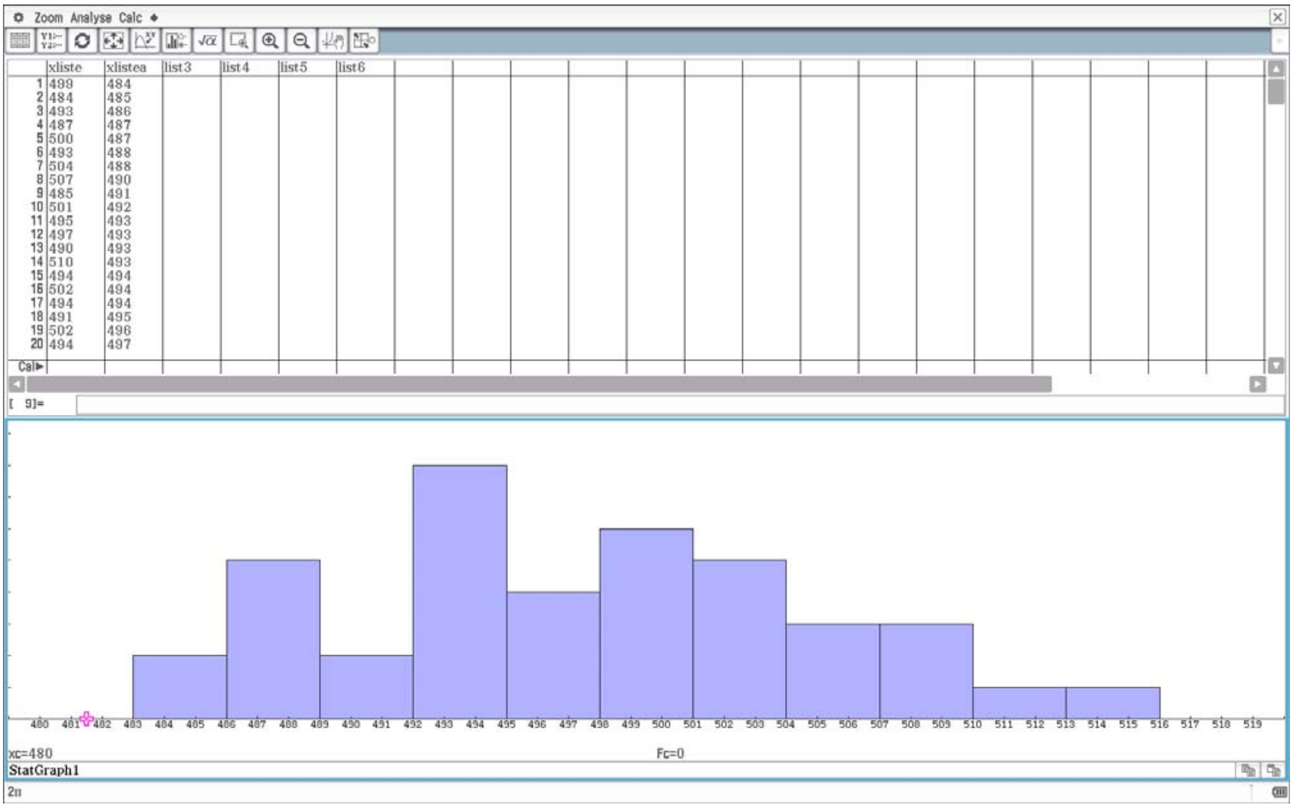
H-Start: 480 (Reduktionslage), H-Schr.: 5 (Klassenbreite) einstellen



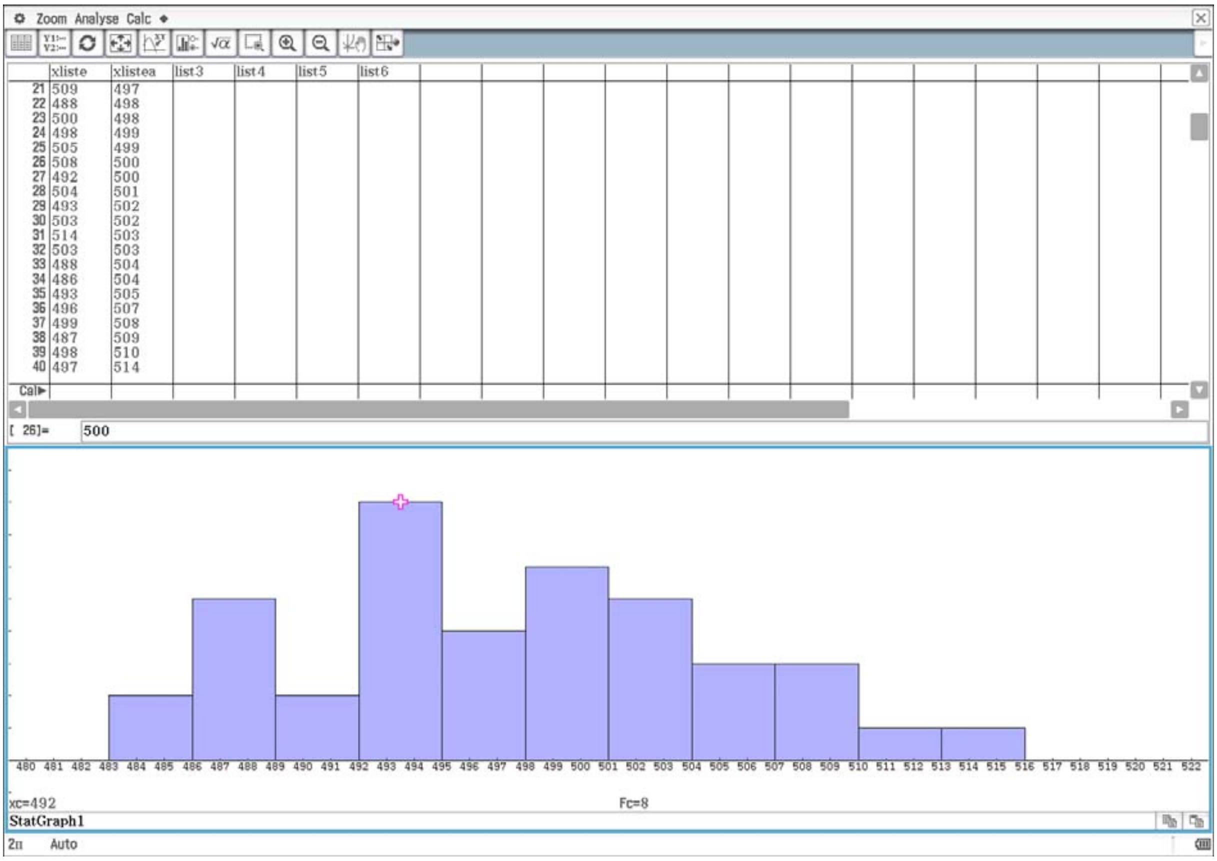
H-Start: 483 (Reduktionslage), H-Schr.: 5 (Klassenbreite) einstellen



H-Start: 480 (Reduktionslage), H-Schr.: 3 (Klassenbreite) einstellen

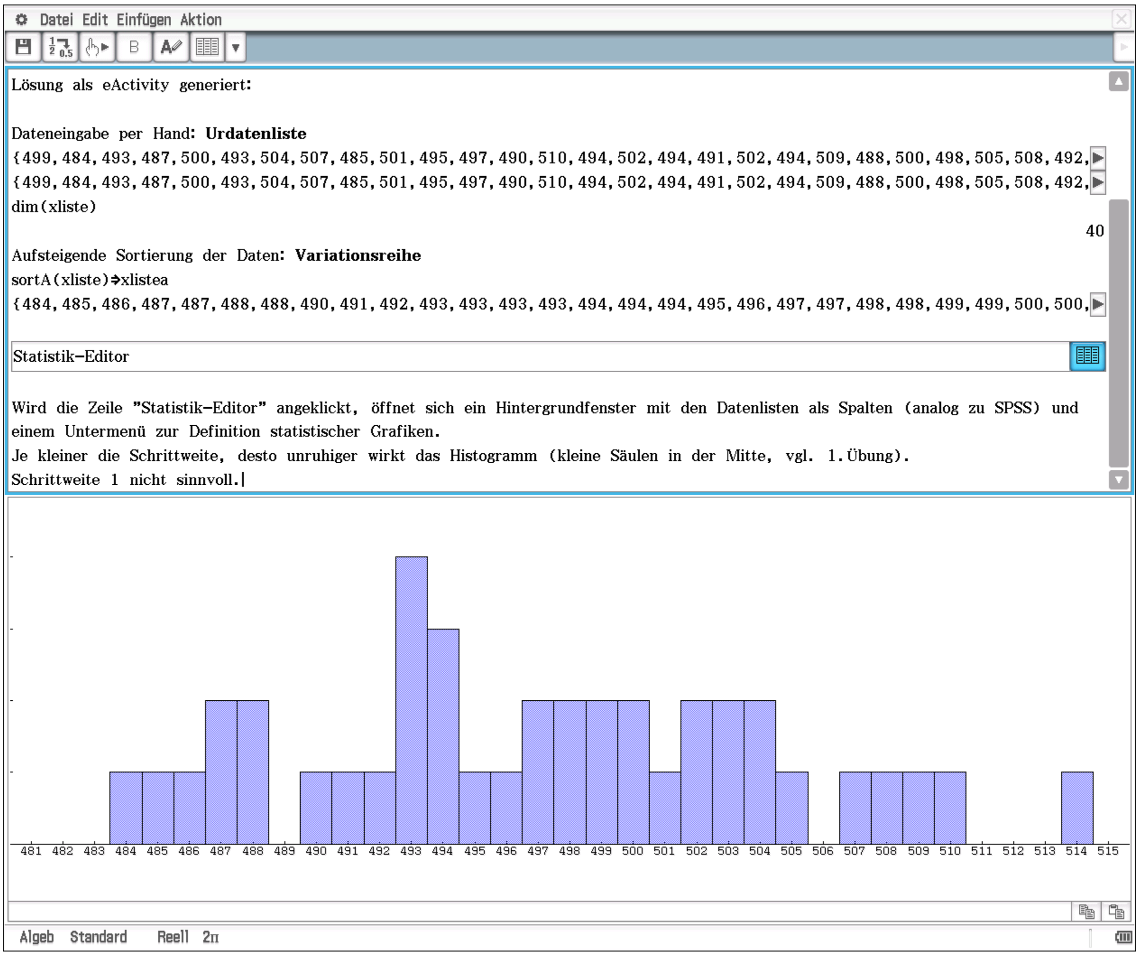


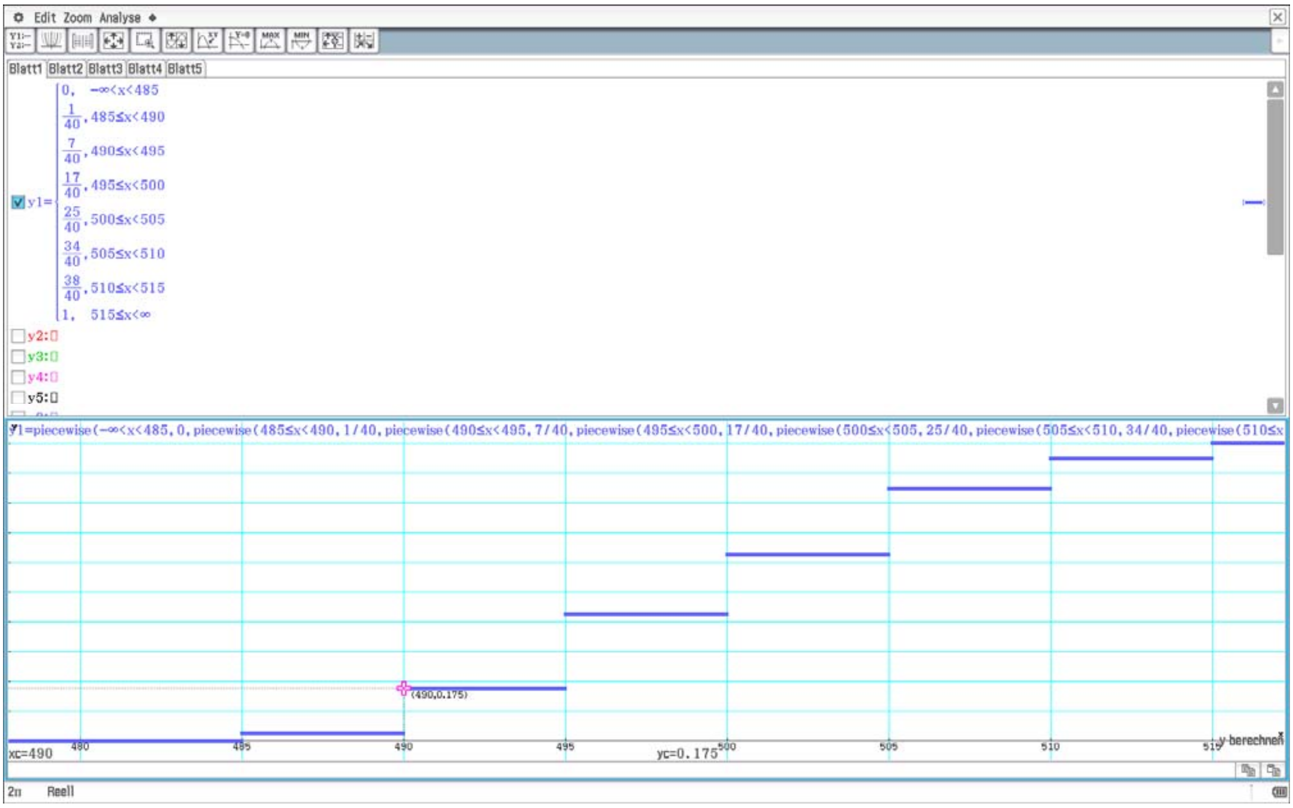
H-Start: 483 (Reduktionslage), H-Schr.: 3 (Klassenbreite) einstellen



H-Start: 480.5 (Reduktionslage), H-Schr.: 1 (Klassenbreite) einstellen

Säulendiagramm zur primären Häufigkeitstabelle:





Prof. Dr. Ludwig Paditz 28.10.2014

1.HA STAT 1.4

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

vgl. LÜB Bd.3, Aufg.16.7 (lies EUR statt DM) S.224ff

Lösung als eActivity generiert:

Dateneingabe per Hand: **Urdatenlisten** als **verbundene Listen** für Zahlenpaare

{12,15,18,18,20,21,24,25,36,37}⇒xliste

{12,15,18,18,20,21,24,25,36,37}

{11,12,16,17,18,18,20,21,26,31}⇒yliste

{11,12,16,17,18,18,20,21,26,31}

Ausführen der linearen Regression und Anzeige der Ergebnisse:

LinearReg xliste,yliste,1,y1,On

done

DispStat

done

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

Lineare Regression

y=a+b•x

a = 2.9409474

b = 0.7105775

r = 0.9770887 (Korrelationskoeffizient)

r^2 = 0.9547023 (Bestimmtheitsmaß B)

MSe= 1.8458793 (MeanSquareError)

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

y1(x)

0.710577547·x+2.940947437

Regressionsgerade

Statistik-Editor

Berechnung von Kennzahlen zur zweidimensionalen Stichprobe für den Zufallsvektor (X, Y) und Anzeige der Ergebnisse:

TwoVariable xliste, yliste, 1

done

DispStat

done

Stop

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

Zweidim. Variable (X, Y)

$\bar{x}$ = 22.6

Σx = 226

Σx^2 = 5724

σ_x = 7.8511146 (Normierung mit n)

s_x = 8.2758014 (Normierung mit n−1)

n = 10

$\bar{y}$ = 19

Σy = 190

Σy^2 = 3936

σ_y = 5.709641 (Normierung mit n)

s_y = 6.01849 (Normierung mit n−1)

Σxy = 4732

minX = 12
maxX = 37
minY = 11
maxY = 31
================================

individueller Aufruf der soeben berechneten Kennzahlen (Systemvariablen):

$\bar{x}$

22.6

$\bar{y}$

19

s_x

8.275801405

s_y

6.018490028

s_x^2

68.48888889

Berechnung von s_{xy} über den Korrelationskoeffizienten r_{xy} (Systemvariable dazu ist rCorr):

LinearReg xliste,yliste,1,y1,On

done

rCorr

0.9770887107

r²Corr

0.9547023485

rCorr×8.275801405×6.018490028

48.66666667

Hinweis: Mit jedem neuen Aufruf eines Statistik-Befehls werden die vorherigen Systemvariablen gelöscht und neu beschrieben.

Schrittweise Berechnung des $MSe=\frac{1}{n-2}\sum_{i=1}^{n}((yi-y(xi))^2)$

n:=10

10

residual

{-0.4678780013,-1.599610642,0.2686567164,1.26865671

listToMat(ans)

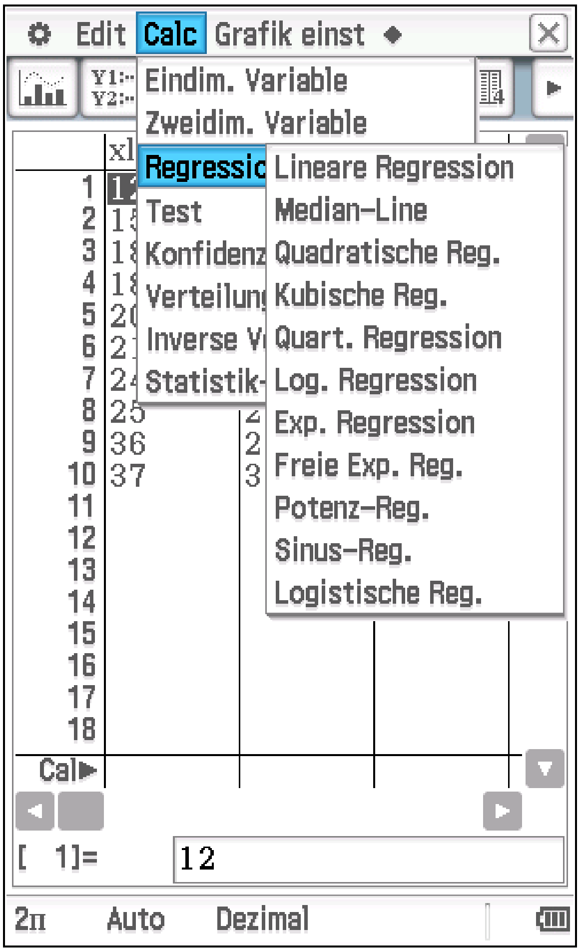
$$\begin{bmatrix} -0.4678780013 \\ -1.599610642 \\ 0.2686567164 \\ 1.268656716 \\ 0.8475016223 \\ 0.1369240753 \\ 5.191434134\text{E-}3 \\ 0.2946138871 \\ -2.52173913 \\ 1.767683323 \end{bmatrix}$$

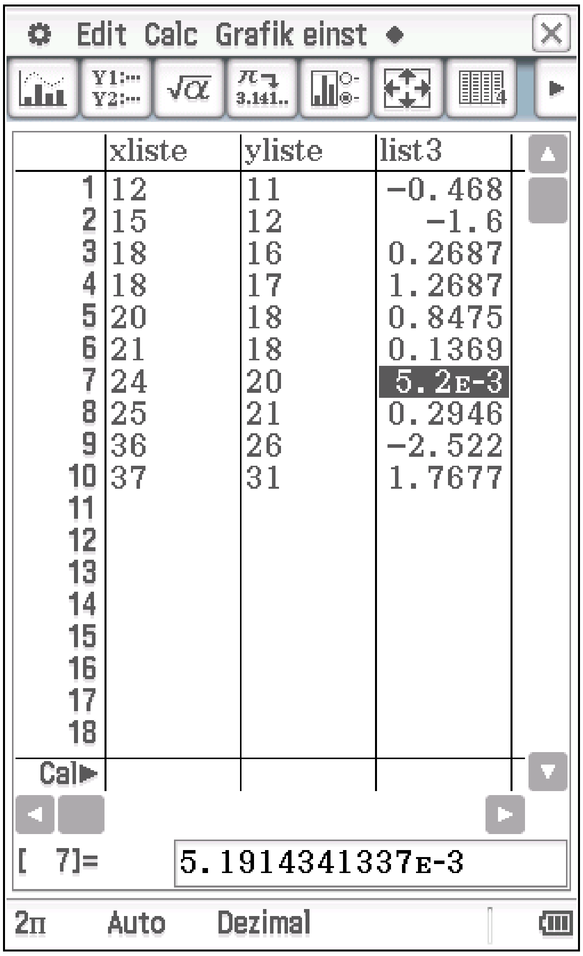
Listenarithmetik:

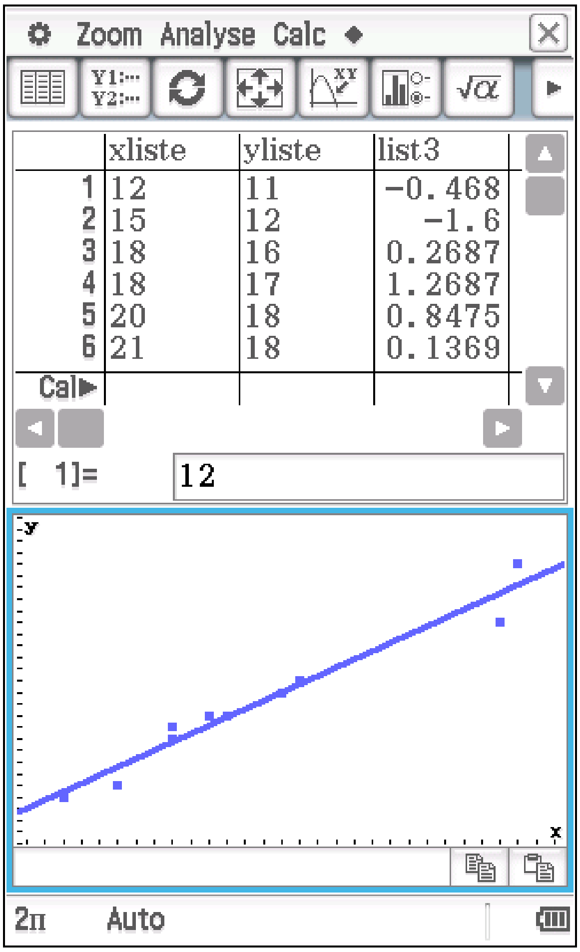
sum(residual^2)/(n-2)

1.845879299

vgl. auch Taschenbuch der Wirtschaftsmathematik S.293ff







list3 enthält die Residuen

(Reste als Differenz yi-y(xi))

Prof. Dr. Ludwig Paditz 28.10.2014

1.HA STAT 1.7
==============

vgl. LÜB Bd.3, Aufg.17.7 S.243 und S.340f

Dateneingabe per Hand mit anschließender Kommaverschiebung:
(Urdatenliste bereits als Variationsreihe gegeben)

{5,6,6,8,11,13,15,16,20,22,24,25,25,28,31,34,37,42,▶
{5,6,6,8,11,13,15,16,20,22,24,25,25,28,31,34,37,42,▶

dim(xliste)

87

xliste/100➔xliste

{0.05,0.06,0.06,0.08,0.11,0.13,0.15,0.16,0.2,0.22,0▶

Die Darstellung der Daten in der Anzeige wird durch die Hintergrundeinstellung im "Grundformat" bestimmt (hier: Zahlenformat "Normal 2")

a) sekundäre Häufigkeitstabelle als verbundene Datenlisten:

linke Klassengrenzen:

seq(x,x,0,9,1)➔Glinks

{0,1,2,3,4,5,6,7,8,9}

Klassenrepräsentanten:

Glinks+0.5➔KRepräs

{0.5,1.5,2.5,3.5,4.5,5.5,6.5,7.5,8.5,9.5}

Häufigkeiten (per Hand in der Variationsreihe auszählen) –

Listeneditor nutzen (list2 erstellen):

Statistik−Listeneditor

list2⇒hliste

{37, 19, 15, 6, 3, 2, 3, 1, 0, 1}

sum(ans)

87

hliste/87⇒rhliste

{0.4252873563, 0.2183908046, 0.1724137931, 0.068965517▶

listToMat(ans)

$$\begin{bmatrix} 0.4252873563 \\ 0.2183908046 \\ 0.1724137931 \\ 0.06896551724 \\ 0.03448275862 \\ 0.02298850575 \\ 0.03448275862 \\ 0.01149425287 \\ 0 \\ 0.01149425287 \end{bmatrix}$$

b) Grafiken: Histogramm und Häufigkeitspolygon

statistische Grafiken

Treppenfunktion und Summenpolygon:

Erzeugung der Liste der Summenhäufigkeiten und relativen Summenhäufigkeiten

n:=87

87

cuml(hliste)⇒shliste

{37,56,71,77,80,82,85,86,86,87}

shliste/n⇒srhliste

{0.4252873563,0.6436781609,0.816091954,0.8850574713▶

listToMat(ans)

$$\begin{bmatrix} 0.4252873563 \\ 0.6436781609 \\ 0.816091954 \\ 0.8850574713 \\ 0.9195402299 \\ 0.9425287356 \\ 0.9770114943 \\ 0.9885057471 \\ 0.9885057471 \\ 1 \end{bmatrix}$$

$$\text{Define } y1(x)=\begin{cases} 0, & x<1 \\ 37/n, & 1\le x<2 \\ 56/n, & 2\le x<3 \\ 71/n, & 3\le x<4 \\ 77/n, & 4\le x<5 \\ 80/n, & 5\le x<6 \\ 82/n, & 6\le x<7 \\ 85/n, & 7\le x<8 \\ 86/n, & 8\le x<9 \\ 86/n, & 9\le x<10 \\ 1, & 10\le x<\infty \end{cases}$$

done

Grafikformat "pixelweise" einstellen, "Auto-Stat-Fenster" aus

2D-Grafik

Statistik-Editor

OneVariable xliste

done

DispStat

done

Hinw.: im Grundformat Q_1 und Q_3 aktivieren

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

Eindim. Variable

$\bar{x}$ = 1.8914943

Σx = 164.56

Σx^2 = 622.1474

σ_x = 1.8903357 (Normierung mit n)

s_x = 1.9012943 (Normierung mit n-1)

n = 87

minX = 0.05

Q_1 = 0.51

Med = 1.38

Q_3 = 2.49

maxX = 9.93

Mode = 0.06

Mode = 0.25

Mode = 0.51

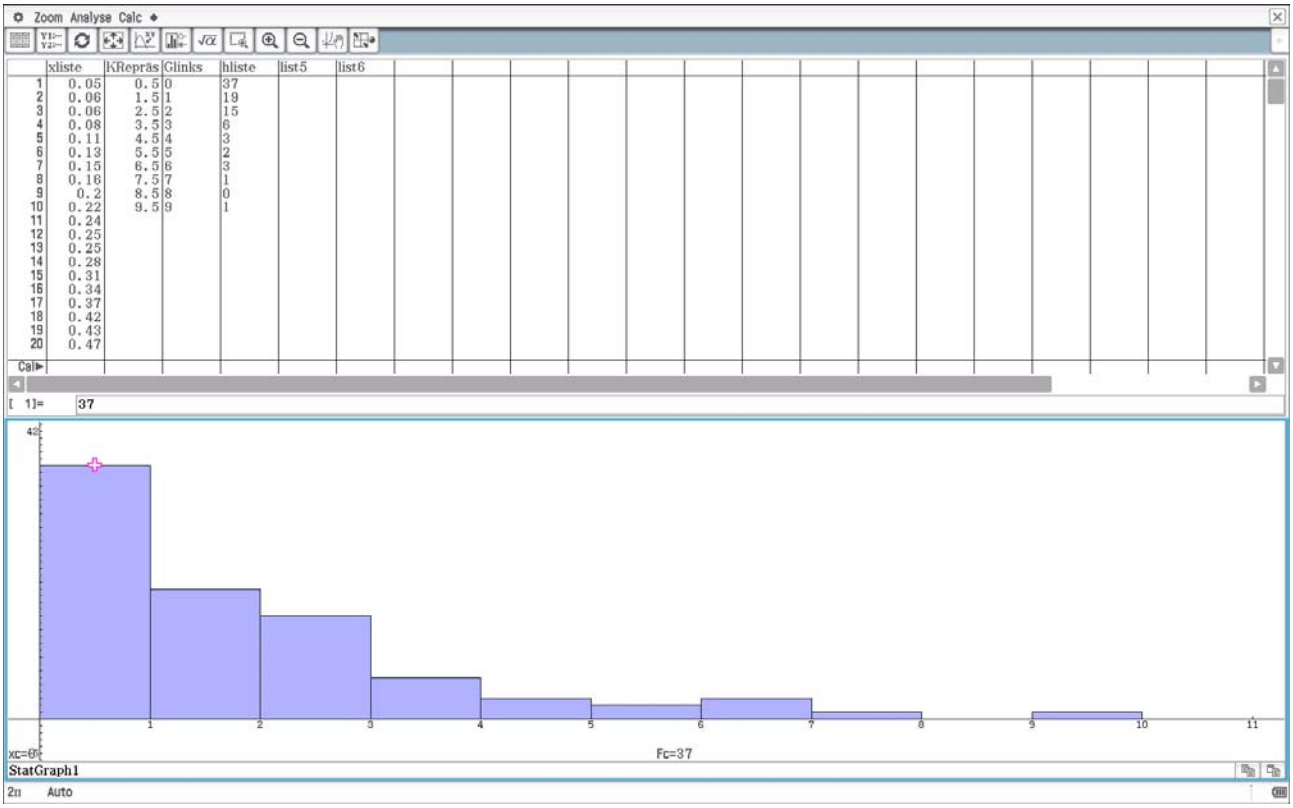
Mode = 0.76

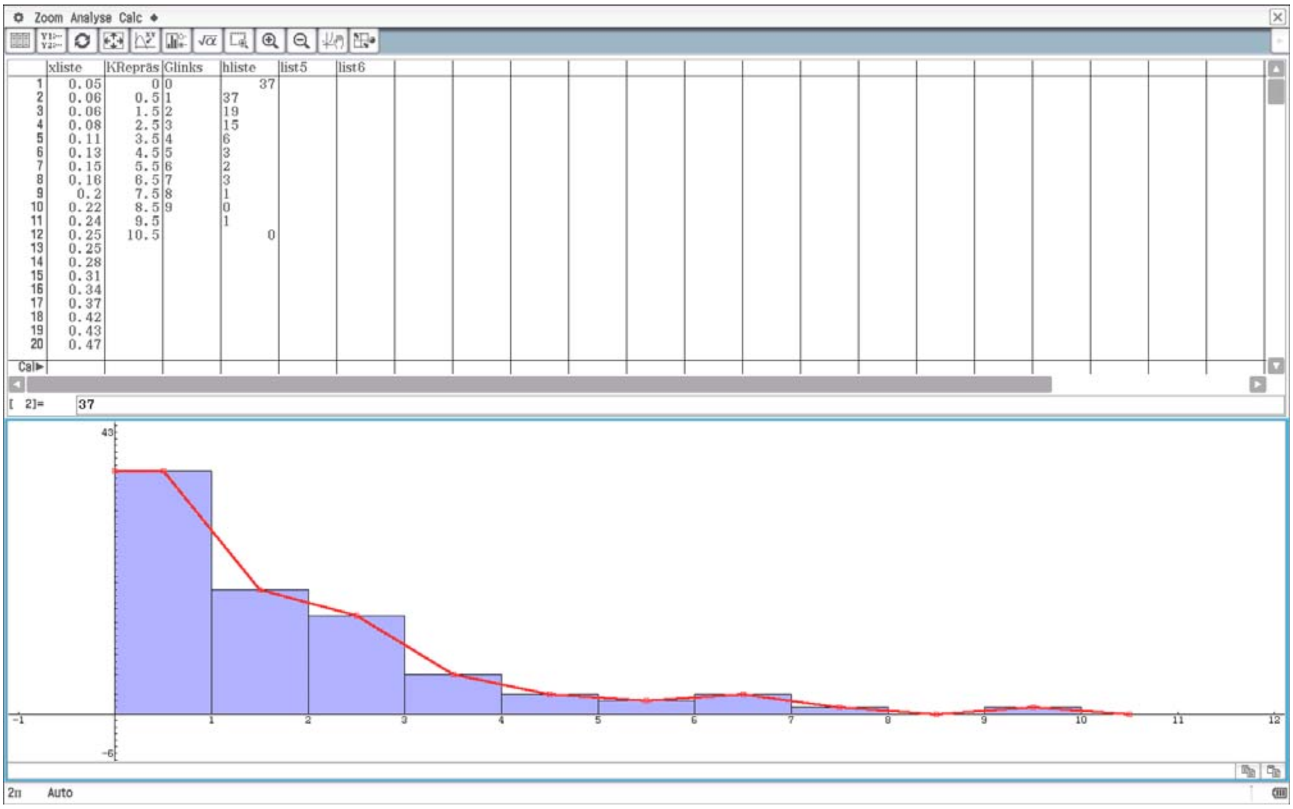
Mode = 0.88

ModeN = 5

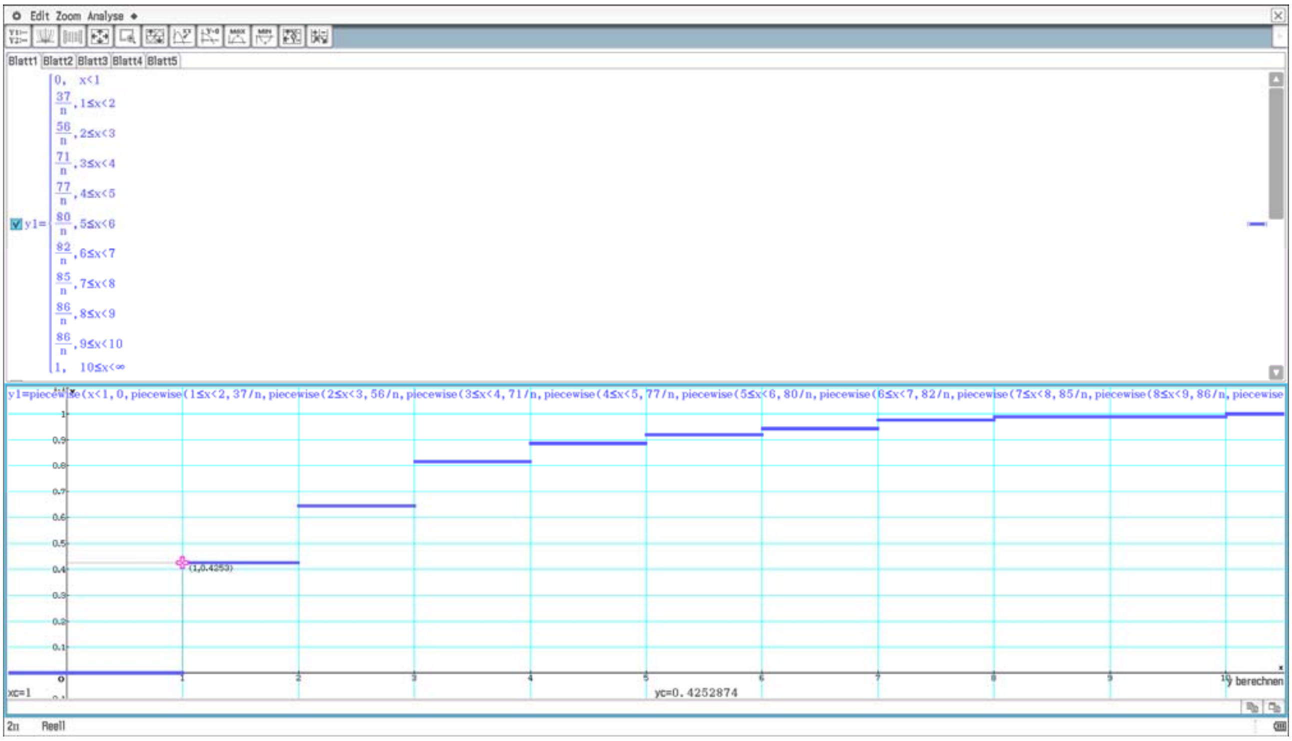
ModeF = 2

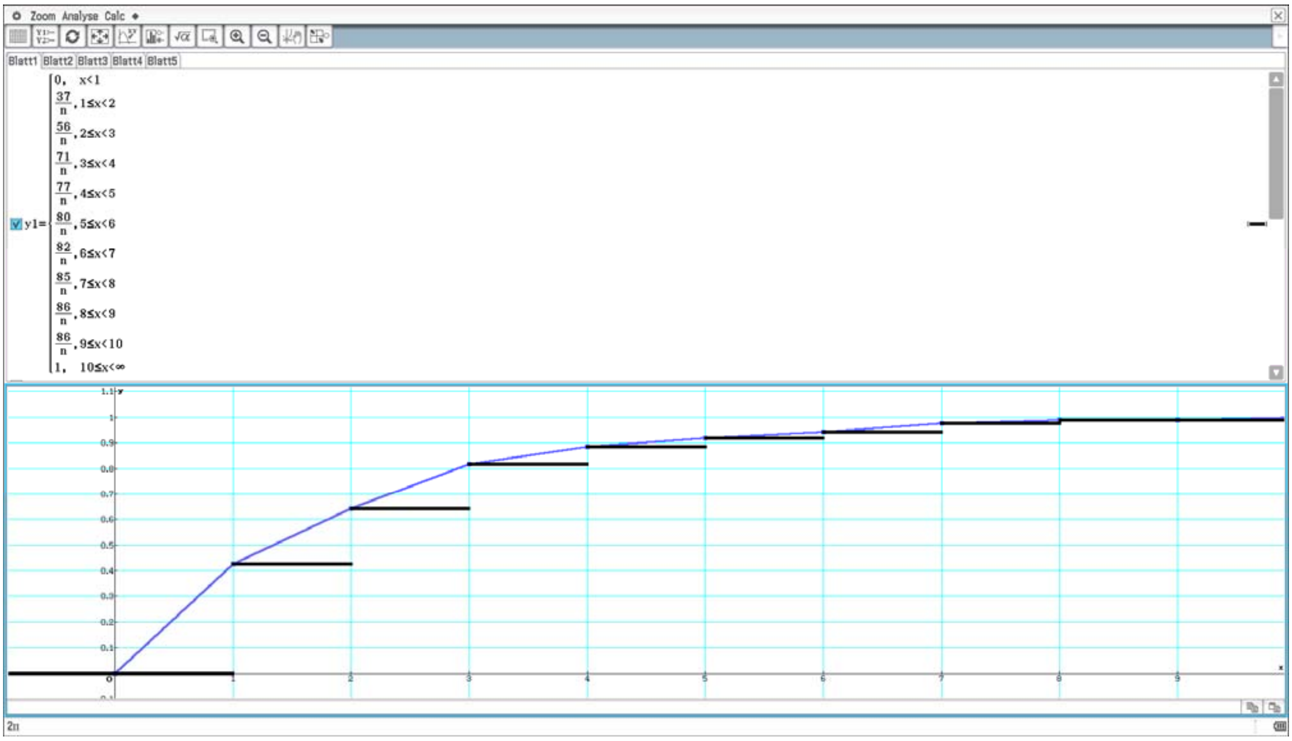
================================





Die KRepräs und hliste wurden um die Datenpaare (0,37) und (10.5,0) erweitert, um ein vollständiges Häufigkeitspolygon zu erhalten. Da es sich um Lebensdauern handelt, wurde links nicht über die Reduktionslage 0 hinausgegangen.





Treppenkurve und Summenpolygon

c) Hypothese zur Theoretischen Verteilung:

Häufigkeitspolygon erinnert an eine fallende e-Funktion: Exponentialverteilung.