Prof. Dr. Ludwig Paditz, 25.10.2016

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

9. Zwei Studenten führen unabhängig voneinander, jeder in seinem Heimatort, eine Befragung durch, wobei sie neben dem Geschlecht der Probanden u.a. die Antwort auf eine Frage mit den Antwortmöglichkeiten: "ja", "egal" und "nein" registrieren.
Beide Studenten befragen je 120 Personen.
Student 1 befragt 48 Männer und 72 Frauen;
Antworten: 10 mal "ja", 20 mal "egal" und 90 mal "nein".
Student 2 befragt 84 Männer und 36 Frauen;
Antworten: 100 mal "ja", 10 mal "egal" und 10 mal "nein".

I) Stellen Sie beide Befragungsergebnisse in je einer Kontingenztafel dar, wobei Sie in die Zellen diejenigen Häufgkeiten eintragen, die zu erwarten gewesen wären, wenn das Antwortverhalten auf die betreffende Frage unabhängig vom Geschlecht wäre. **(Indifferenztabelle)**

II) Wir wollen annehmen, die unter I) bestimmten Tafeln hätten sich tatsächlich bei den Befragungen ergeben.
a) Wie wäre ein solches Ergebnis zu interpretieren?
b) Die Studenten beschließen, ihre Daten zu vereinigen. Welche Kontingenztafel für die nunmehr 240 Fälle ergibt sich?
(Indifferenztabelle)
c) Bestimmen Sie den χ^2-Wert für die unter b) erhaltene Tafel

und stellen Sie das Ergebnis grafsch durch gestapelte Balken dar! Interpretieren Sie das Ergebnis!

Lösung:

I)

X ... (zufälliges) Geschlecht

Kodierung: X=1 männlich, X=2 weiblich

Y ... (zufällige) Antwort

Kodierung: Y=1 ja, Y=2 egal, Y=3 nein

Kontingenztafel für Student 1:

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

$$\begin{bmatrix} X/Y & & & & \\ & 1 & 2 & 3 & \Sigma \\ 1 & \square & \square & \square & 48 \\ 2 & \square & \square & \square & 72 \\ \Sigma & 10 & 20 & 90 & 120 \end{bmatrix}$$

Ansatz für Chi-Test mit einer fingierten Tabelle:

$$A:=\begin{bmatrix} 5 & 10 & 33 \\ 5 & 10 & 57 \end{bmatrix}$$

$$\begin{bmatrix} 5 & 10 & 33 \\ 5 & 10 & 57 \end{bmatrix}$$

ChiTest A

done

Expected

$$\begin{bmatrix} 4 & 8 & 36 \\ 6 & 12 & 54 \end{bmatrix}$$

Indifferenztabelle:

$$\begin{bmatrix} X/Y & & & & \\ & 1 & 2 & 3 & \Sigma \\ 1 & 4 & 8 & 36 & 48 \\ 2 & 6 & 12 & 54 & 72 \\ \Sigma & 10 & 20 & 90 & 120 \end{bmatrix}$$

Hinweis: direkte Berechnung der Indifferenztabelle

mit $n\hat{}_{ij}=\dfrac{n_{i.}*n_{.j}}{n_{..}}$

Berechnung mittels Matrizenrechnung:

$$\begin{bmatrix} 48 \\ 72 \end{bmatrix}*[10\ \ 20\ \ 90]*\frac{1}{120}$$

$$\begin{bmatrix} 4 & 8 & 36 \\ 6 & 12 & 54 \end{bmatrix}$$

Kontingenztafel für Student 2:

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

$$\begin{bmatrix} X/Y & & & & \\ & 1 & 2 & 3 & \Sigma \\ 1 & \square & \square & \square & 84 \\ 2 & \square & \square & \square & 36 \\ \Sigma & 100 & 10 & 10 & 120 \end{bmatrix}$$

Ansatz für Chi-Test mit einer fingierten Tabelle:

$$A:=\begin{bmatrix} 74 & 5 & 5 \\ 26 & 5 & 5 \end{bmatrix}$$

$$\begin{bmatrix} 74 & 5 & 5 \\ 26 & 5 & 5 \end{bmatrix}$$

ChiTest A

done

Expected

$$\begin{bmatrix} 70 & 7 & 7 \\ 30 & 3 & 3 \end{bmatrix}$$

Indifferenztabelle:

$$\begin{bmatrix} X/Y & & & & & \\ & & 1 & 2 & 3 & \Sigma \\ & 1 & 70 & 7 & 7 & 84 \\ & 2 & 30 & 3 & 3 & 36 \\ & \Sigma & 100 & 10 & 10 & 120 \end{bmatrix}$$

Hinweis: direkte Berechnung der Indifferenztabelle

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

Lösung:

II)

a)

Student 1: hoher Anteil für "nein", geringer Anteil für "ja"

Prozentanteile für X=1 bzw. X=2 nicht unterscheidbar

(Geschlechterunabhängig):

$$\begin{bmatrix} 4/48 & 8/48 & 36/48 \\ 6/72 & 12/72 & 54/72 \end{bmatrix}*100$$

$$\begin{bmatrix} 8.33 & 16.67 & 75.00 \\ 8.33 & 16.67 & 75.00 \end{bmatrix}$$

Student 2: hoher Anteil für "ja", geringer Anteil für "nein",

d.h. umgekehrtes Antwortverhalten.

Prozentanteile für X=1 bzw. X=2 nicht unterscheidbar

(Geschlechterunabhängig):

$$\begin{bmatrix} 70/84 & 7/84 & 7/84 \\ 30/36 & 3/36 & 3/36 \end{bmatrix} * 100$$

$$\begin{bmatrix} 83.33 & 8.33 & 8.33 \\ 83.33 & 8.33 & 8.33 \end{bmatrix}$$

b)

$$\begin{bmatrix} X/Y & & & & \\ & 1 & 2 & 3 & \Sigma \\ 1 & \square & \square & \square & 132 \\ 2 & \square & \square & \square & 108 \\ \Sigma & 110 & 30 & 100 & 240 \end{bmatrix}$$

Ansatz für Chi-Test mit einer fingierten Tabelle:

$$A := \begin{bmatrix} 74 & 15 & 43 \\ 36 & 15 & 57 \end{bmatrix}$$

$$\begin{bmatrix} 74 & 15 & 43 \\ 36 & 15 & 57 \end{bmatrix}$$

anderer Zugang:

Kreuztabelle als Zusammenfassung der separaten

Indifferenztabellen aus **I)**:

$$\begin{bmatrix} X/Y & & & & \\ & 1 & 2 & 3 & \Sigma \\ 1 & 4 & 8 & 36 & 48 \\ 2 & 6 & 12 & 54 & 72 \\ \Sigma & 10 & 20 & 90 & 120 \end{bmatrix} + \begin{bmatrix} X/Y & & & & \\ & 1 & 2 & 3 & \Sigma \\ 1 & 70 & 7 & 7 & 84 \\ 2 & 30 & 3 & 3 & 36 \\ \Sigma & 100 & 10 & 10 & 120 \end{bmatrix}$$

$$\begin{bmatrix} 4 & 8 & 36 & 48 \\ 6 & 12 & 54 & 72 \\ 10 & 20 & 90 & 120 \end{bmatrix} + \begin{bmatrix} 70 & 7 & 7 & 84 \\ 30 & 3 & 3 & 36 \\ 100 & 10 & 10 & 120 \end{bmatrix}$$

$$\begin{bmatrix} 74 & 15 & 43 & 132 \\ 36 & 15 & 57 & 108 \\ 110 & 30 & 100 & 240 \end{bmatrix}$$

ChiTest A

done

approx(Expected)

$$\begin{bmatrix} 60.5 & 16.5 & 55.0 \\ 49.5 & 13.5 & 45.0 \end{bmatrix}$$

approx(χ^2value)

12.8

Indifferenztabelle:

(mit gebrochenen Anteilen, nur theoretisch möglich)

$$\begin{bmatrix} X/Y & & & & \\ & 1 & 2 & 3 & \Sigma \\ 1 & 60.5 & 16.5 & 55 & 132 \\ 2 & 49.5 & 13.5 & 45 & 108 \\ \Sigma & 110 & 30 & 100 & 240 \end{bmatrix}$$

Hinweis: direkte Berechnung der Indifferenztabelle

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

Prozentanteile für X=1 bzw. X=2 nicht unterscheidbar

(Geschlechterunabhängig):

$$\begin{bmatrix} 60.5/132 & 16.5/132 & 55/132 \\ 49.5/108 & 13.5/108 & 45/108 \end{bmatrix}*100$$

$$\begin{bmatrix} 45.8 & 12.5 & 41.7 \\ 45.8 & 12.5 & 41.7 \end{bmatrix}$$

hohe Anteile für "ja" und "nein" im Gegensatz zur Einzelbefragung Student1/Student2

c) Z=1 (Student1), Z=2(Student2)

$$\begin{bmatrix} Z/Y & & & & \\ & 1 & 2 & 3 & \Sigma \\ 1 & 10 & 20 & 90 & 120 \\ 2 & 100 & 10 & 10 & 120 \\ \Sigma & 110 & 30 & 100 & 240 \end{bmatrix}$$

$$B:=\begin{bmatrix} 10 & 20 & 90 \\ 100 & 10 & 10 \end{bmatrix}$$

$$\begin{bmatrix} 10 & 20 & 90 \\ 100 & 10 & 10 \end{bmatrix}$$

ChiTest B

done

approx(χ^2value)

141.0

Expected

$$\begin{bmatrix} 55 & 15 & 50 \\ 55 & 15 & 50 \end{bmatrix}$$

Der hohe χ^2-Wert erscheint kritisch, d.h. Befragungsergebnisse stark davon abhängig, wer die Befragung vorgenommen hat (Student1 bzw. Student2).

anderer Zugang:

Kreuztabelle als Zusammenfassung der separaten Indifferenztabellen:

$$\begin{bmatrix} X/Y & & & & \\ & 1 & 2 & 3 & \Sigma \\ 1 & 4 & 8 & 36 & 48 \\ 2 & 6 & 12 & 54 & 72 \\ \Sigma & 10 & 20 & 90 & 120 \end{bmatrix} + \begin{bmatrix} X/Y & & & & \\ & 1 & 2 & 3 & \Sigma \\ 1 & 70 & 7 & 7 & 84 \\ 2 & 30 & 3 & 3 & 36 \\ \Sigma & 100 & 10 & 10 & 120 \end{bmatrix}$$

$$\begin{bmatrix} 4 & 8 & 36 & 48 \\ 6 & 12 & 54 & 72 \\ 10 & 20 & 90 & 120 \end{bmatrix} + \begin{bmatrix} 70 & 7 & 7 & 84 \\ 30 & 3 & 3 & 36 \\ 100 & 10 & 10 & 120 \end{bmatrix}$$

$$\begin{bmatrix} 74 & 15 & 43 & 132 \\ 36 & 15 & 57 & 108 \\ 110 & 30 & 100 & 240 \end{bmatrix}$$

$$A := \begin{bmatrix} 74 & 15 & 43 \\ 36 & 15 & 57 \end{bmatrix}$$

$$\begin{bmatrix} 74 & 15 & 43 \\ 36 & 15 & 57 \end{bmatrix}$$

ChiTest A

done

approx(Expected)

$$\begin{bmatrix} 60.5 & 16.5 & 55.0 \\ 49.5 & 13.5 & 45.0 \end{bmatrix}$$

approx(χ^2value)

12.8

$$\begin{bmatrix} X/Y & & & & \\ & 1 & 2 & 3 & \Sigma \\ 1 & 74 & 15 & 43 & 132 \\ 2 & 36 & 15 & 57 & 108 \\ \Sigma & 110 & 30 & 100 & 240 \end{bmatrix}$$

Auch hier erscheint der hohe χ^2–Wert kritisch, d.h. das

Antwortverhalten ist hier geschlechterabhängig:

X=1(männl.) eher Y=1(ja) und X=2(weibl.) eher Y=3(nein)

$$\begin{bmatrix} 74/132 & 15/132 & 43/132 & 132/132 \\ 36/108 & 15/108 & 57/108 & 108/108 \\ 110/240 & 30/240 & 100/240 & 240/240 \end{bmatrix} * 100$$

$$\begin{bmatrix} 56.1 & 11.4 & 32.6 & 100.0 \\ 33.3 & 13.9 & 52.8 & 100.0 \\ 45.8 & 12.5 & 41.7 & 100.0 \end{bmatrix}$$

Prof. Dr. Ludwig Paditz, 25.10.2016

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

10. Vierzehn zufällig ausgewählteMänner gaben folgende Schuhgrößen und Körperhöhen (in cm) an:

$$\begin{bmatrix} \text{Schuhgröße} & 42 & 45 & 42.5 & 45.5 & 43 & 39 & 42 & 41 & 41.5 & 42 \\ \text{Körperhöhe} & 175 & 188 & 178 & 189 & 182 & 169 & 182 & 171 & 175 & 17 \end{bmatrix}$$

(a) Besteht zwischen Schuhgröße und Körperhöhe ein linearer Zusammenhang?
(a1) Skizzieren Sie die Datenpaare in der (x,y){Ebene.
(a2) Berechnen Sie getrennt für beide Merkmale arithmetisches Mittel und empirische Standardabweichung.
(a3) Berechnen und interpretieren Sie den Pearsonschen Korrelationskoeffzienten.

(b) Berechnen Sie nach der Methode der kleinsten Quadrate eine lineare Funktion aus der sich bei bekannter Körperhöhe(xliste) eines Mannes dessen Schuhgröße(yliste) schätzen lässt.
(c) Ist die bei (b) berechnete Beziehung dazu geeignet, für einen Mann mit Schuhgröße 45 die Körpergröße vorherzusagen?

Lösung:
[42 45 42.5 45.5 43 39 42 41 41.5 42.5 42 40

$$\begin{bmatrix} 42 & 45 & \frac{85}{2} & \frac{91}{2} & 43 & 39 & 42 & 41 & \frac{83}{2} & \frac{85}{2} & 42 & 40 & 42 & 45 \end{bmatrix}$$

[175 188 178 189 182 169 182 171 175 179 173 174 176▶

[175 188 178 189 182 169 182 171 175 179 173 174 176▶

matToList(trn(Schuhgr),1)⇒yliste

$$\left\{42, 45, \frac{85}{2}, \frac{91}{2}, 43, 39, 42, 41, \frac{83}{2}, \frac{85}{2}, 42, 40, 42, 45\right\}$$

matToList(trn(Körperh),1)⇒xliste

{175,188,178,189,182,169,182,171,175,179,173,174,17▶

(a1) (x,y)-Plot

(a2)

OneVariable xliste

done

DispStat

done

mean(xliste)

178.21

stdDev(xliste)

6.09

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

OneVariable yliste

done

DispStat

done

mean(yliste)

42.36

stdDev(yliste)

1.84

(a3) Korr.-Koeff. als Teilergebnis der LinearReg.

LinearReg xliste, yliste

done

DispStat

done

rCorr

0.91

(b)

LinearReg xliste, yliste, 1, y1, On

done

DispStat

done

aCoef

0.28

bCoef

-6.94

y1(x)

0.28·x-6.94

MKQ:

Define y(x)=a*x+b

done

$$\text{Define } F(a,b)=\sum_{k=1}^{14}\left((\text{yliste}[k]-y(\text{xliste}[k]))^2\right)$$

done

$$\frac{d}{da}(F(a,b))=0 \Rightarrow Gl1$$

338·(169·a+b-39)+348·(174·a+b-40)+342·(171·a+b-41)+350▶

$\frac{d}{db}(F(a,b))=0 \Rightarrow Gl2$

$$2\cdot(169\cdot a+b-39)+2\cdot(174\cdot a+b-40)+2\cdot(171\cdot a+b-41)+2\cdot\Big(175\cdot a+b$$

$\begin{cases} Gl1 \\ Gl2 \end{cases}\Big|_{a,b}$

$$\left\{a=\frac{1868}{6753}, b=-\frac{93733}{13506}\right\}$$

approx(ans)

$$\{a=0.28, b=-6.94\}$$

(c) nein, besser Regression für Schuhgr. in Abhängigkeit von Körperh. durchführen.

solve(y1(x)=45, x)

$$\{x=187.77\}$$

LinearReg yliste, xliste, 1, y2, On

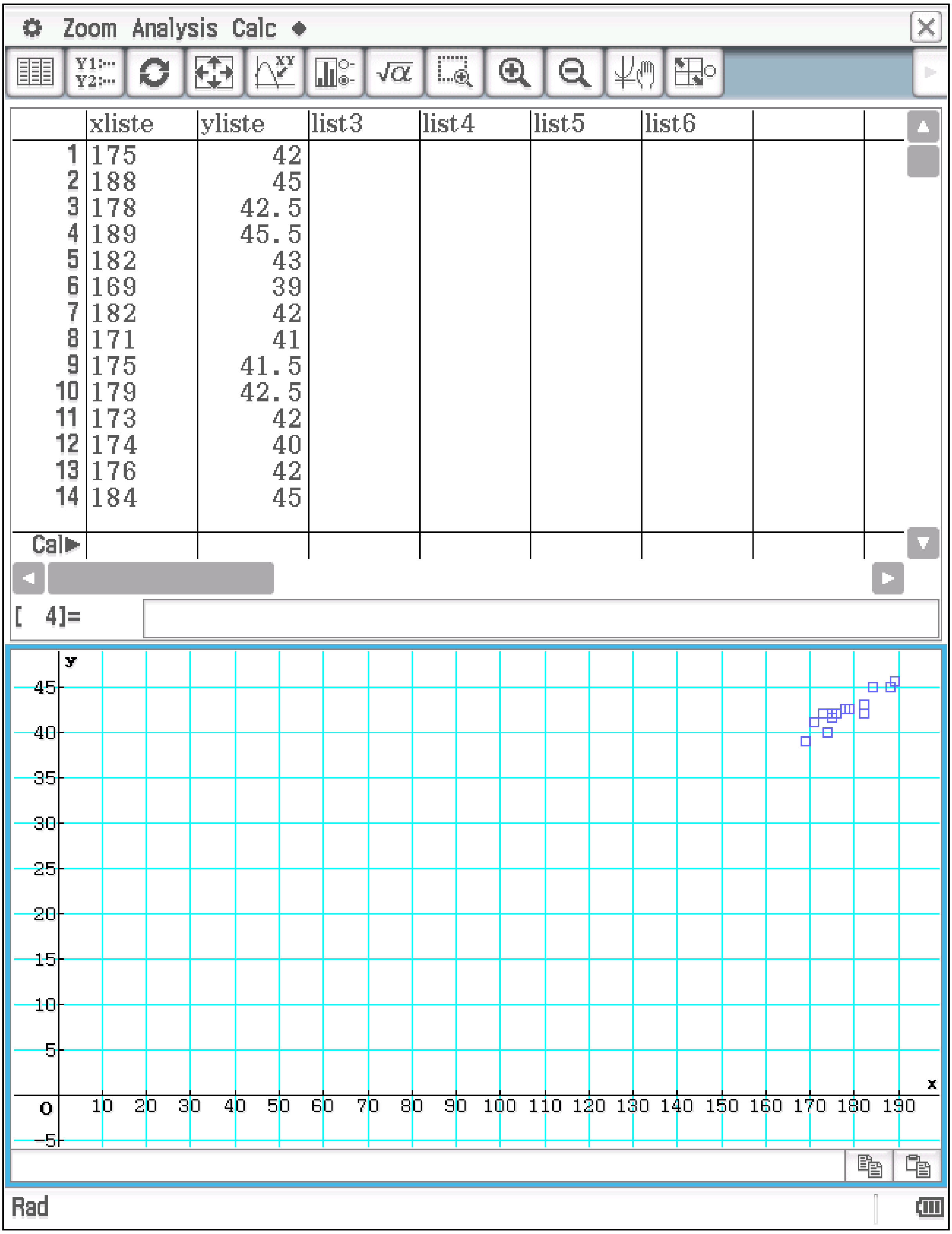
done

y2(45)

186.19

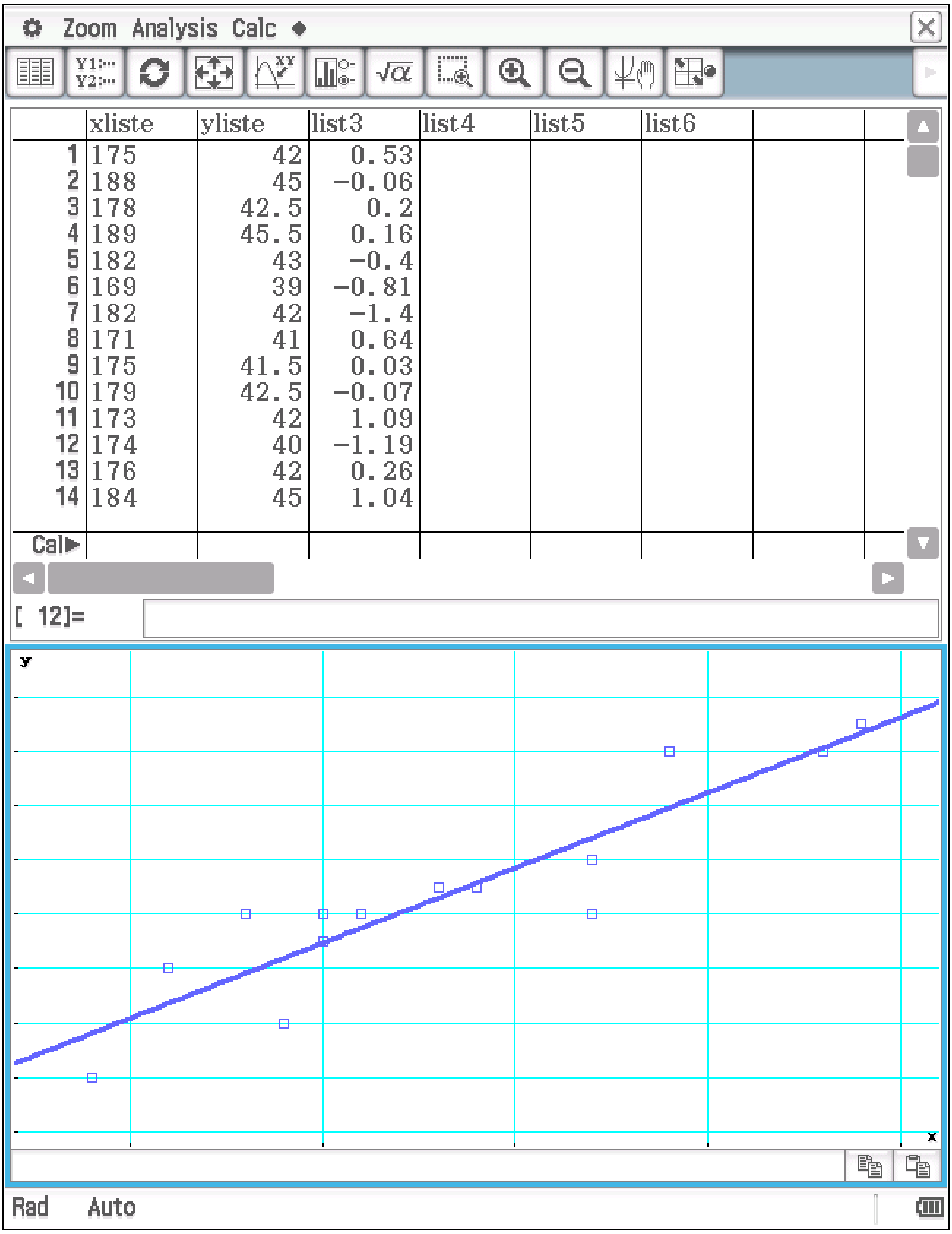
Besseres Ergebnis 186.19 im Vergleich zu 187.77

(x,y)-Plot (Scatter-Plot)



	xliste	yliste	list3	list4	list5	list6
1	175	42				
2	188	45				
3	178	42.5				
4	189	45.5				
5	182	43				
6	169	39				
7	182	42				
8	171	41				
9	175	41.5				
10	179	42.5				
11	173	42				
12	174	40				
13	176	42				
14	184	45				

vergrößert:



	xliste	yliste	list3	list4	list5	list6
1	175	42	0.53			
2	188	45	−0.06			
3	178	42.5	0.2			
4	189	45.5	0.16			
5	182	43	−0.4			
6	169	39	−0.81			
7	182	42	−1.4			
8	171	41	0.64			
9	175	41.5	0.03			
10	179	42.5	−0.07			
11	173	42	1.09			
12	174	40	−1.19			
13	176	42	0.26			
14	184	45	1.04			

Prof. Dr. Ludwig Paditz, 01.11.2016

Statistik für Produktionstechniker

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

11. Die folgenden Daten wurden in den 60-er Jahren von Dereck Whiteside (UK Building Research Station) gesammelt. Er zeichnete den wöchentlichen Gasverbrauch (in 1 000 cu feet, 1 cubic foot = 0,028 m^3) und die durchschnittliche wöchentliche Außentemperatur (in °C) für sein Haus in Südost-England auf, und zwar in einem Zeitraum von 26 Wochen vor bis 30 Wochen nach einer durchgeführten Hausisolierung. Das Hausthermostat war die ganze Zeit über auf 20°C eingestellt. Von Interesse ist der Zusammenhang zwischen Gasverbrauch und Außentemperatur. Die folgende Tabelle enthält die Messwerte vor der Isolierung.

X ... Außentemperatur
Y ... Gasverbrauch

X	Y	X	Y	X	Y
-0.8	7.2	4.3	5.2	7.4	4.2
-0.7	6.9	5.4	4.9	7.5	4.0
0.4	6.4	6.0	4.9	7.5	3.9
2.5	6.0	6.0	4.3	7.6	3.5
2.9	5.8	6.0	4.4	8.0	4.0
3.2	5.8	6.2	4.5	8.5	3.6
3.6	5.6	6.3	4.6	9.1	3.1
3.9	4.7	6.9	3.7	10.2	2.6
4.2	5.8	7.0	3.9		

(a) Stellen Sie diese Datenpaare in einem kartesischen Koordinatensystem dar, indem Sie für das abhängige Merkmal die Ordinatenachse verwenden.

(b) Beschreiben Sie den Zusammenhang zwischen den beiden Merkmalen, und beurteilen Sie seine Stärke durch eine geeignete Kennzahl.

(c) Bestimmen Sie die Regressionsgerade nach der Methode der kleinsten Quadrate, und zeichnen Sie diese in das Streudiagramm ein. Beurteilen Sie die Güte der Anpassung anhand des Bestimmtheitsmaßes.

(d) Interpretieren Sie die beiden Koeffzienten der in (c) berechneten Geradengleichung. Mit welchem wöchentlichen Gasverbrauch in m^3 musste Herr Whiteside vor der Isolierung seines Hauses bei einer Außentemperatur von 5,0°C rechnen?

(e) Wie werden sich die statistischen Kennzahlen, die den Zusammenhang zwischen Außentemperatur und wöchentlichem Gasverbrauch kennzeichnen, nach der Isolierung ändern?

Lösung:

$$\text{matToList}\left(\begin{bmatrix}-0.8\\-0.7\\0.4\\2.5\\2.9\\3.2\\3.6\\3.9\\4.2\end{bmatrix},1\right)\Rightarrow\text{list11}$$

{-0.8,-0.7,0.4,2.5,2.9,3.2,3.6,3.9,4.2}

$$\text{matToList}\left(\begin{bmatrix}4.3\\5.4\\6.0\\6.0\\6.0\\6.2\\6.3\\6.9\\7.0\end{bmatrix},1\right)\Rightarrow\text{list12}$$

{4.3,5.4,6.0,6.0,6.0,6.2,6.3,6.9,7.0}

$$\text{matToList}\left(\begin{bmatrix}7.4\\7.5\\7.5\\7.6\\8.0\\8.5\\9.1\\10.2\end{bmatrix},1\right)\Rightarrow\text{list13}$$

{7.4,7.5,7.5,7.6,8.0,8.5,9.1,10.2}

augment(augment(list11,list12),list13)⇒listx

{-0.8,-0.7,0.4,2.5,2.9,3.2,3.6,3.9,4.2,4.3,5.4,6.0, ▶

dim(ans)

26.0

$$\mathrm{matToList}\left(\begin{bmatrix}7.2\\6.9\\6.4\\6.0\\5.8\\5.8\\5.6\\4.7\\5.8\end{bmatrix},1\right)\Rightarrow \mathrm{list21}$$

{7.2, 6.9, 6.4, 6.0, 5.8, 5.8, 5.6, 4.7, 5.8}

$$\mathrm{matToList}\left(\begin{bmatrix}5.2\\4.9\\4.9\\4.3\\4.4\\4.5\\4.6\\3.7\\3.9\end{bmatrix},1\right)\Rightarrow \mathrm{list22}$$

{5.2, 4.9, 4.9, 4.3, 4.4, 4.5, 4.6, 3.7, 3.9}

$$\mathrm{matToList}\left(\begin{bmatrix}4.2\\4.0\\3.9\\3.5\\4.0\\3.6\\3.1\\2.6\end{bmatrix},1\right)\Rightarrow \mathrm{list23}$$

{4.2, 4.0, 3.9, 3.5, 4.0, 3.6, 3.1, 2.6}

augment(augment(list21, list22), list23)⇒listy

{7.2, 6.9, 6.4, 6.0, 5.8, 5.8, 5.6, 4.7, 5.8, 5.2, 4.9, 4.9, 4. ▶

dim(ans)

26.0

(a)

Stat–Editor

(b) Korrelationskoeffizient

LinearReg listx, listy, 1, y1, On

done

DispStat

done

y1(x)

$-0.3932\cdot x+6.8538$

Residual

{0.0316, -0.2291, -0.2965, 0.1293, 0.0866, 0.2045, 0.1618 ▶

r^2Corr

0.9438

rCorr

-0.9715

(c) $y=b_0+b_1x$

mit $b_0=6.853827699$ und $b_1=-0.3932388222$

Gasverbrauch = 6.8538-0.3932*Temperatur

Bestimmtheitsmaß = r^2Corr = 0.9438, d.h.
94,38% des Zusammenhanges sind über die lineare Regression bestimmt.

(d)

$b_0=6.8538$ [1000 cu feet] ... Gasverbrauch pro Woche bei einer durchschnittlichen Außentemperatur von $x=0$[°C]

$|b_1|=0.3932$ [1000 cu feet/grad] **...** Anstieg des wöchentlichen Gasverbrauches bei Absinken der durchschnittlichen Außentemperatur um 1[°C]

$x=5$[°C] ⇒ 4887,6 [cu feet] ≈ 137m^3

y1(5)*1000

4887.6336

ans*0.028

136.8537

(e) Nach der Hausisolierung:

b_0 wird kleiner (Gasverbrauch pro Woche bei einer durchschnittlichen Außentemperatur von $x=0$[°C])

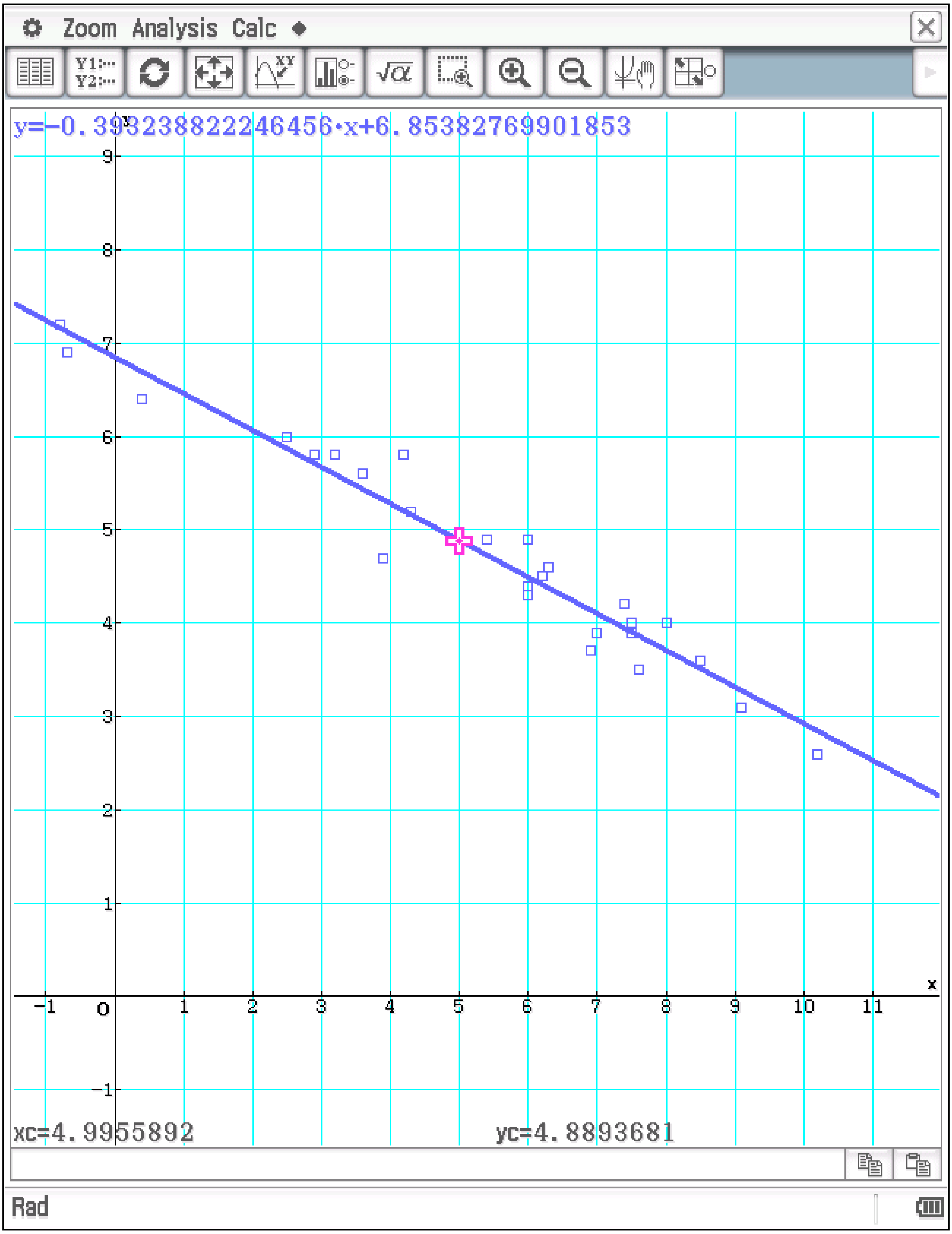
$b_1<0$ wird größer (der Anstieg der Regressionsgeraden wird "flacher")

Edit Calc SetGraph

	listx	listy	residual	list4	list5	list6
1	−0.8	7.2	0.0316			
2	−0.7	6.9	−0.229			
3	0.4	6.4	−0.297			
4	2.5	6	0.1293			
5	2.9	5.8	0.0866			
6	3.2	5.8	0.2045			
7	3.6	5.6	0.1618			
8	3.9	4.7	−0.62			
9	4.2	5.8	0.5978			
10	4.3	5.2	0.0371			
11	5.4	4.9	0.1697			
12	6	4.9	0.4056			
13	6	4.3	−0.194			
14	6	4.4	−0.094			
15	6.2	4.5	0.0843			
16	6.3	4.6	0.2236			
17	6.9	3.7	−0.44			
18	7	3.9	−0.201			
19	7.4	4.2	0.2561			
20	7.5	4	0.0955			
21	7.5	3.9	−5E−3			
22	7.6	3.5	−0.365			
23	8	4	0.2921			
24	8.5	3.6	0.0887			
25	9.1	3.1	−0.175			
26	10.2	2.6	−0.243			
27						
28						
29						
30						
31						
32						
33						
34						
35						
Cal▶						

list= residual

Rad Decimal



Zoom Analysis Calc
y=−0.393238822246456·x+6.85382769901853
xc=4.9955892
yc=4.8893681
Rad

Prof. Dr. Ludwig Paditz, 01.11.2016

Statistik für Produktionstechniker

12. Die folgende Tabelle enthält die Klausurergebnisse einer Gruppe von Studierenden in einer Mathematik- und einer Statistikklausur.

X ... Punkte Mathematik

Y ... Punkte Statistik

Student	X	Rang	Y	Rang
1	38	□	39	□
2	47	□	34	□
3	44	□	31	□
4	51	□	48	□
5	35	□	46	□
6	29	□	23	□
7	22	□	17	□
8	14	□	12	□
9	12	□	16	□
10	19	□	28	□
11	9	□	10	□

(a) Welches Skaleniveau weisen die Variablen **Punkte Mathematik** bzw. **Punkte Statistik** auf?

(b) Stellen Sie den Datensatz mit Hilfe eines Scatterplots grafisch dar!

(c) Berechnen Sie den Spearmanschen

Rangkorrelationskoeffzienten für die Variablen **Punkte Mathematik** und **Punkte Statistik**!

(d) Wie ist die Abhängigkeit zwischen den beiden Datenreihen einzuschätzen (stark, mittel, schwach, nicht vorhanden)?

(e) Berechnen Sie den Pearsonschen Korrelationskoeffzienten.

(f) Wie ist das Vorzeichen des Pearsonschen Korrelationskoeffzienten zu interpretieren?

Lösung:

(a) Skalenniveau: ordinal (metrisch)

(b)

$$\text{matToList}\left(\begin{bmatrix}38\\47\\44\\51\\35\\29\\22\\14\\12\\19\\9\end{bmatrix},1\right)\Rightarrow \text{listx}$$

$$\{38,47,44,51,35,29,22,14,12,19,9\}$$

$$\text{matToList}\left(\begin{bmatrix}39\\34\\31\\48\\46\\23\\17\\12\\16\\28\\10\end{bmatrix},1\right)\Rightarrow\text{listy}$$

{39,34,31,48,46,23,17,12,16,28,10}

Stat−Editor

(c) Spearmanscher Rangkorrelationskoeffzient

sortD(listx)

{51,47,44,38,35,29,22,19,14,12,9}

sortD(listy)

{48,46,39,34,31,28,23,17,16,12,10}

$$\begin{bmatrix} \text{Student} & X & \text{Rang} & Y & \text{Rang} \\ 1 & 38 & 4 & 39 & 3 \\ 2 & 47 & 2 & 34 & 4 \\ 3 & 44 & 3 & 31 & 5 \\ 4 & 51 & 1 & 48 & 1 \\ 5 & 35 & 5 & 46 & 2 \\ 6 & 29 & 6 & 23 & 7 \\ 7 & 22 & 7 & 17 & 8 \\ 8 & 14 & 9 & 12 & 10 \\ 9 & 12 & 10 & 16 & 9 \\ 10 & 19 & 8 & 28 & 6 \\ 11 & 9 & 11 & 10 & 11 \end{bmatrix}$$

$$\begin{bmatrix} 4 \\ 2 \\ 3 \\ 1 \\ 5 \\ 6 \\ 7 \\ 9 \\ 10 \\ 8 \\ 11 \end{bmatrix} - \begin{bmatrix} 3 \\ 4 \\ 5 \\ 1 \\ 2 \\ 7 \\ 8 \\ 10 \\ 9 \\ 6 \\ 11 \end{bmatrix}$$

$$\begin{bmatrix} 1 \\ -2 \\ -2 \\ 0 \\ 3 \\ -1 \\ -1 \\ -1 \\ 1 \\ 2 \\ 0 \end{bmatrix}$$

$(\text{norm}(\text{ans}))^2$

26

$R:=1-\dfrac{6*\text{ans}}{11*(11^2-1)}$

$\dfrac{97}{110}$

approx(ans)

0.8818181818

(d) wegen R=0.88 ist die Abhängigkeit (Punkte Mathematik bzw. Statistik) als stark einzuschätzen:
Die Ergebnisse in Mathematik entsprechen etwa denen in Statistik und umgekehrt.

(e)

LinearReg listx, listy, 1, y1, On

done

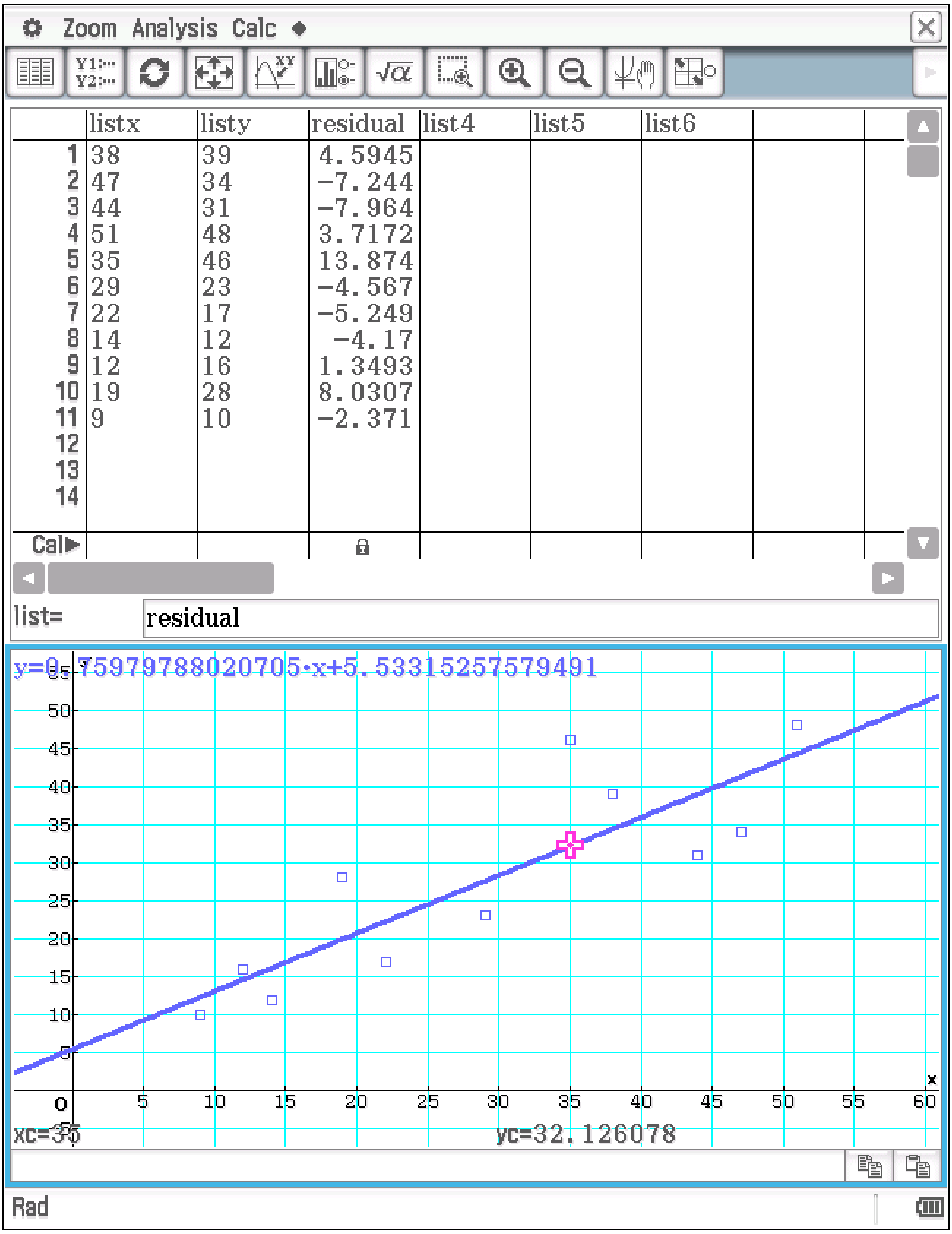
DispStat

done

rCorr

0.8523207564

(f) r=0.8523 bedeutet starke pos. Korrelation:
Höhere Punktzahlen in Mathematik ziehen auch höhere Punktzahlen in Statistik nach sich (positiver Trend)



	listx	listy	residual	list4	list5	list6
1	38	39	4.5945			
2	47	34	−7.244			
3	44	31	−7.964			
4	51	48	3.7172			
5	35	46	13.874			
6	29	23	−4.567			
7	22	17	−5.249			
8	14	12	−4.17			
9	12	16	1.3493			
10	19	28	8.0307			
11	9	10	−2.371			
12						
13						
14						
Cal▶						

Prof. Dr. Ludwig Paditz, 01.11.2016

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

13. Man lässt 5 Personen, die sich für eine ausgeschriebene Stelle beworben haben, jeweils die beiden Intelligenztests I und II durchlaufen und erhält folgende Punktzahlen:

X ... Punktzahl der Person i bei Test I
Y ... Punktzahl der Person i bei Test II

Person Nr. i	X	Y
1	95	109
2	110	126
3	121	117
4	87	89
5	105	81

(a) Skizzieren Sie die zugehörige Punktwolke in der (x, y)-Ebene.
(b) Berechnen Sie den Spearmanschen Rangkorrelationskoeffzienten.

Lösung:
(a)

$$\mathrm{matToList}\left(\begin{bmatrix} 95 \\ 110 \\ 121 \\ 87 \\ 105 \end{bmatrix}, 1\right) \Rightarrow \mathrm{listx}$$

$$\{95, 110, 121, 87, 105\}$$

$$\mathrm{matToList}\left(\begin{bmatrix} 109 \\ 126 \\ 117 \\ 89 \\ 81 \end{bmatrix}, 1\right) \Rightarrow \mathrm{listy}$$

$$\{109, 126, 117, 89, 81\}$$

Stat−Editor

(b) Spearmanscher Rangkorrelationskoeffzient

sortD(listx)

$$\{121, 110, 105, 95, 87\}$$

sortD(listy)

$$\{126, 117, 109, 89, 81\}$$

$$\begin{bmatrix} 4 \\ 2 \\ 1 \\ 5 \\ 3 \end{bmatrix} - \begin{bmatrix} 3 \\ 1 \\ 2 \\ 4 \\ 5 \end{bmatrix}$$

$$\begin{bmatrix} 1 \\ 1 \\ -1 \\ 1 \\ -2 \end{bmatrix}$$

$$(\mathrm{norm}(\mathrm{ans}))^2$$

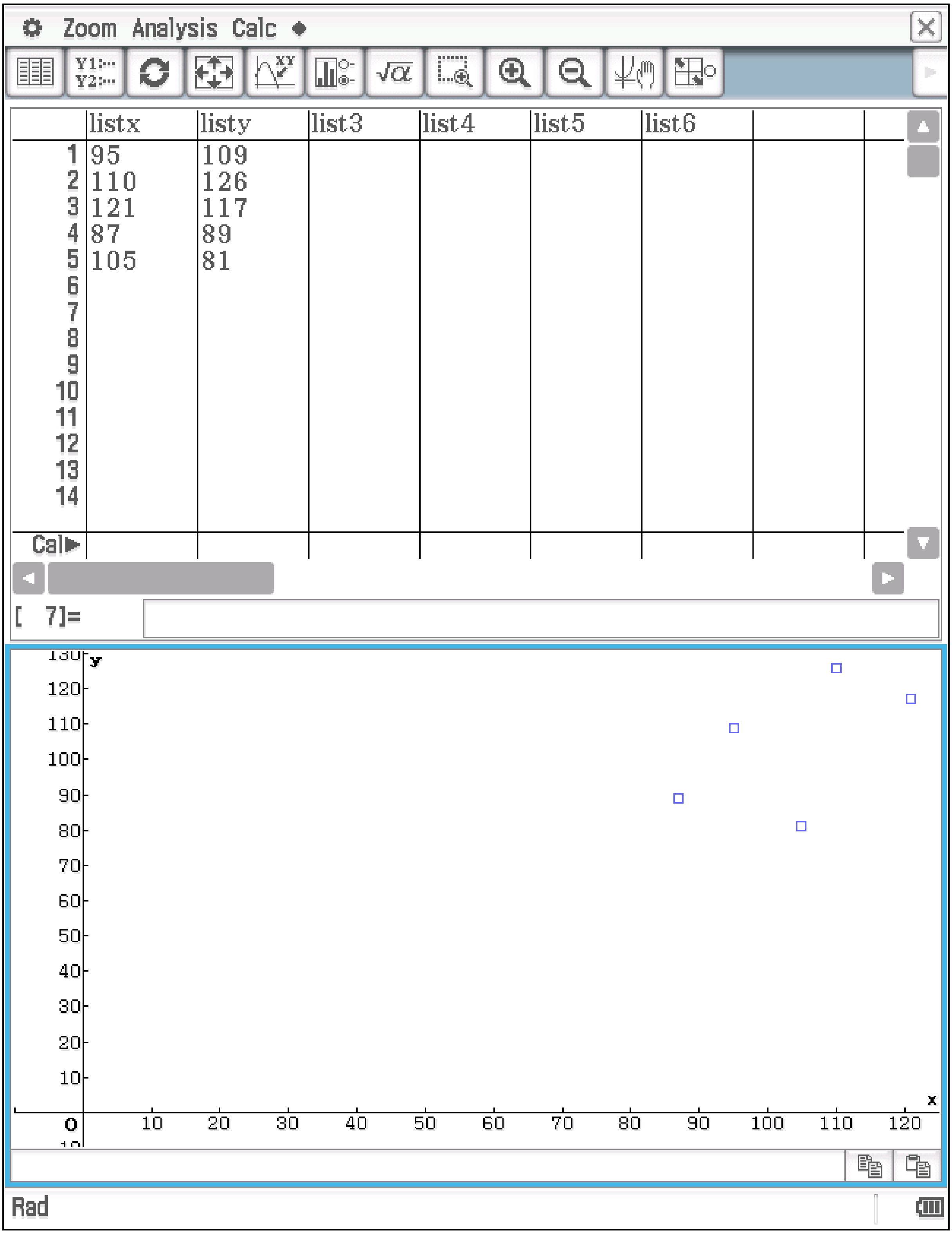
$$8$$

$$R:=1-\frac{6*ans}{5*(5^2-1)}$$

$$\frac{3}{5}$$

approx(ans)

$$0.6$$



Zoom Analysis Calc
listx listy list3 list4 list5 list6
1 95 109
2 110 126
3 121 117
4 87 89
5 105 81
Cal▶
[7]=
y
x
Rad

Prof. Dr. Ludwig Paditz, 01.11.2016

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

14. Für die 23 Teilnehmerinnen im Siebenkampf der Damen bei denWeltmeisterschaften der Leichtathletik im Sommer 2009 in Berlin, die für jede Teildisziplin ein Ergebnis haben, beschreibe das Merkmal XH das Ergebnis im 100-m-Hürdenlauf in s, XW das Ergebnis im Weitsprung in m und XK das Ergebnis im Kugelstoßen in m.

(a) Der Pearsonsche Korrelationskoeffzient für die beiden Merkmale XH und XW ist -0,432. Interpretieren Sie diesen Wert, insbesondere das Vorzeichen!

(b) Der Pearsonsche Korrelationskoeffzient für die beiden Merkmale XH und XK ist 0,025. Interpretieren Sie diesen Wert!

(c) Für eine Zufallsstichprobe im Umfang 10 aus dem Starterinnenfeld ergab sich die untenstehende Tabelle. Berechnen Sie den Rangkorrelationskoeffzienten für die beiden Teildisziplinen 100-m-Hürdenlauf und Weitsprung!

XH ... Ergebnis 100-m-Hürden in s
XW ... Ergebnis Weitsprung in m

$$\begin{bmatrix} \text{Nr.} & \text{XH} & \text{XW} \\ === & ===== & ===== \\ 1 & 13.62 & 6.42 \\ 2 & 13.44 & 6.32 \\ 3 & 13.58 & 6.50 \\ 4 & 13.50 & 6.24 \\ 5 & 13.78 & 5.97 \\ 6 & 13.60 & 6.13 \\ 7 & 13.69 & 5.82 \\ 8 & 13.96 & 5.90 \\ 9 & 13.99 & 5.99 \\ 10 & 14.23 & 6.23 \end{bmatrix}$$

Lösung:

(a) negative Korralation, d.h. etwa: je schneller im Hürdenlauf, desto weniger weit im Weitsprung

(b) fast keine Korrelation, d.h. etwa: zwischen Ergebnis im Hürdenlauf und im Kugelstoßen besteht kein Zusammenhang.

(c)

$$\text{matToList}\left(\begin{bmatrix} 13.62 \\ 13.44 \\ 13.58 \\ 13.50 \\ 13.78 \\ 13.60 \\ 13.69 \\ 13.96 \\ 13.99 \\ 14.23 \end{bmatrix}, 1\right) \Rightarrow \text{listxh}$$

{13.62, 13.44, 13.58, 13.5, 13.78, 13.6, 13.69, 13.96, 13.9▶

$$\text{matToList}\left(\begin{bmatrix} 6.42 \\ 6.32 \\ 6.50 \\ 6.24 \\ 5.97 \\ 6.13 \\ 5.82 \\ 5.90 \\ 5.99 \\ 6.23 \end{bmatrix}, 1\right) \Rightarrow \text{listxw}$$

{6.42, 6.32, 6.5, 6.24, 5.97, 6.13, 5.82, 5.9, 5.99, 6.23}

sortD(listxh)

{14.23, 13.99, 13.96, 13.78, 13.69, 13.62, 13.6, 13.58, 13.

sortD(listxw)

{6.5, 6.42, 6.32, 6.24, 6.23, 6.13, 5.99, 5.97, 5.9, 5.82}

$$\begin{bmatrix} \text{XH} & \text{Rang} \\ 13.62 & 6 \\ 13.44 & 10 \\ 13.58 & 8 \\ 13.50 & 9 \\ 13.78 & 4 \\ 13.60 & 7 \\ 13.69 & 5 \\ 13.96 & 3 \\ 13.99 & 2 \\ 14.23 & 1 \end{bmatrix}, \begin{bmatrix} \text{XW} & \text{Rang} \\ 6.42 & 2 \\ 6.32 & 3 \\ 6.50 & 1 \\ 6.24 & 4 \\ 5.97 & 8 \\ 6.13 & 6 \\ 5.82 & 10 \\ 5.90 & 9 \\ 5.99 & 7 \\ 6.23 & 5 \end{bmatrix}$$

$$\begin{bmatrix} 6 \\ 10 \\ 8 \\ 9 \\ 4 \\ 7 \\ 5 \\ 3 \\ 2 \\ 1 \end{bmatrix} - \begin{bmatrix} 2 \\ 3 \\ 1 \\ 4 \\ 8 \\ 6 \\ 10 \\ 9 \\ 7 \\ 5 \end{bmatrix}$$

$$\begin{bmatrix} 4 \\ 7 \\ 7 \\ 5 \\ -4 \\ 1 \\ -5 \\ -6 \\ -5 \\ -4 \end{bmatrix}$$

$(\text{norm}(\text{ans}))^2$

258

$R:=1-\frac{6*\text{ans}}{10*(100-1)}$

-0.5636363636

Hinweis: mit aufsteigender Sortierung

fill(11, 10, 1)

$$\begin{bmatrix} 11 \\ 11 \\ 11 \\ 11 \\ 11 \\ 11 \\ 11 \\ 11 \\ 11 \\ 11 \end{bmatrix}$$

$$\text{ans}= \begin{bmatrix} 6 \\ 10 \\ 8 \\ 9 \\ 4 \\ 7 \\ 5 \\ 3 \\ 2 \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} 5 \\ 1 \\ 3 \\ 2 \\ 7 \\ 4 \\ 6 \\ 8 \\ 9 \\ 10 \end{bmatrix}$$

$$\mathrm{fill}(11,10,1)-\begin{bmatrix} 2 \\ 3 \\ 1 \\ 4 \\ 8 \\ 6 \\ 10 \\ 9 \\ 7 \\ 5 \end{bmatrix}$$

$$\begin{bmatrix} 9 \\ 8 \\ 10 \\ 7 \\ 3 \\ 5 \\ 1 \\ 2 \\ 4 \\ 6 \end{bmatrix}$$

$$\begin{bmatrix} 5 \\ 1 \\ 3 \\ 2 \\ 7 \\ 4 \\ 6 \\ 8 \\ 9 \\ 10 \end{bmatrix}-\begin{bmatrix} 9 \\ 8 \\ 10 \\ 7 \\ 3 \\ 5 \\ 1 \\ 2 \\ 4 \\ 6 \end{bmatrix}$$

$$\begin{bmatrix} -4 \\ -7 \\ -7 \\ -5 \\ 4 \\ -1 \\ 5 \\ 6 \\ 5 \\ 4 \end{bmatrix}$$

$(\mathrm{norm}(\mathrm{ans}))^2$

$$258$$

$R:=1-\frac{6*\mathrm{ans}}{10*(100-1)}$

$$-0.5636363636$$

Ergebnis aus (a) wird bestätigt.

Prof. Dr. Ludwig Paditz, 01.11.2016

Statistik für Produktionstechniker

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

15. Sei X eine N(7,4)-verteilte Zufallsvariable. Bestimmen Sie:

(a) $P(X \geq 9)$

(b) $P(X < 8)$

(c) $P(6 \leq X \leq 11)$

(d) $P(|X - 7| > 3.5)$

Lösung: Wiederholung aus Mathematik2

Syntax: normCDf(a, b, σ, μ)=

normCDf$\left(\frac{a-\mu}{\sigma}, \frac{b-\mu}{\sigma}, 1, 0\right)$=normCDf$\left(\frac{a-\mu}{\sigma}, \frac{b-\mu}{\sigma}\right)$

(zentrieren mit μ, normieren mit σ)

Standardnormalverteilung $\Phi(x)$ tabelliert

(a) $P(X \geq 9)$

normCDf(9,∞,2,7)

0.1586552539

normCDf(1,∞)

0.1586552539

(b) $P(X < 8)$

normCDf(−∞,8,2,7)

0.6914624613

normCDf(−∞,0.5)

0.6914624613

(c) $P(6 \leq X \leq 11)$

normCDf(6, 11, 2, 7)

0.6687123293

normCDf(−0.5, 2)

0.6687123293

(d) $P(|X - 7| > 3.5)$

$P(|X - 7| > 3.5)=$

$P(X - 7 > 3.5)+P(X - 7 < -3.5)=$

$P(X > 10.5)+P(X < 3.5)$

normCDf(−∞, 3.5, 2, 7)+normCDf(10.5, ∞, 2, 7)

0.08011831373

alternativ:

$P(|X - 7| > 3.5)=1-P(|X - 7| \leq 3.5)=$

$1-P(-3.5 \leq X - 7 \leq 3.5)=$

1−normCDf(3.5, 10.5, 2, 7)

0.08011831373

1−normCDf$\left(\frac{-3.5}{2}, \frac{3.5}{2}\right)$

0.08011831373

1−2*normCDf$\left(0, \frac{3.5}{2}\right)$

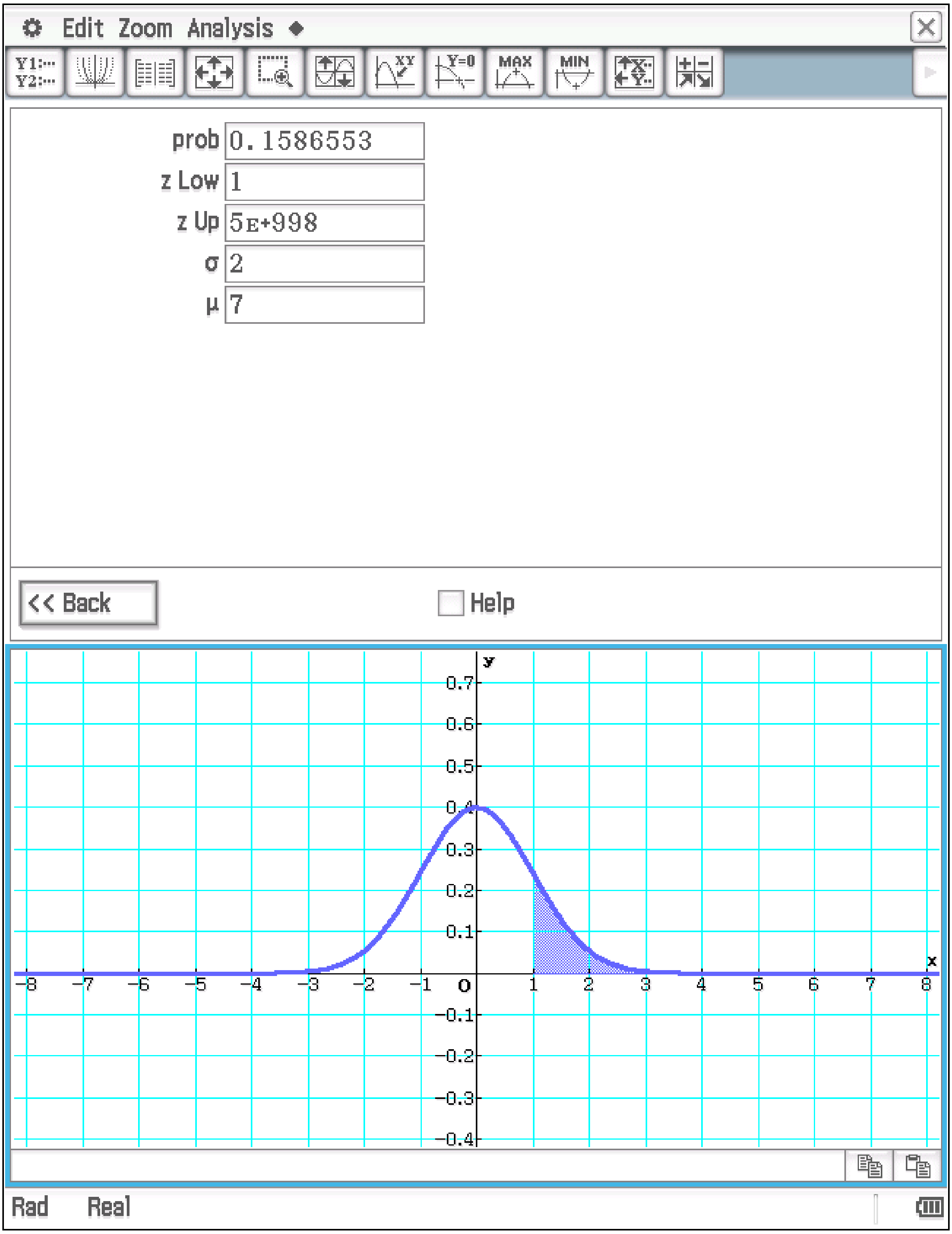
0.08011831373

1−2*normCDf(0, 1.75)

0.08011831373

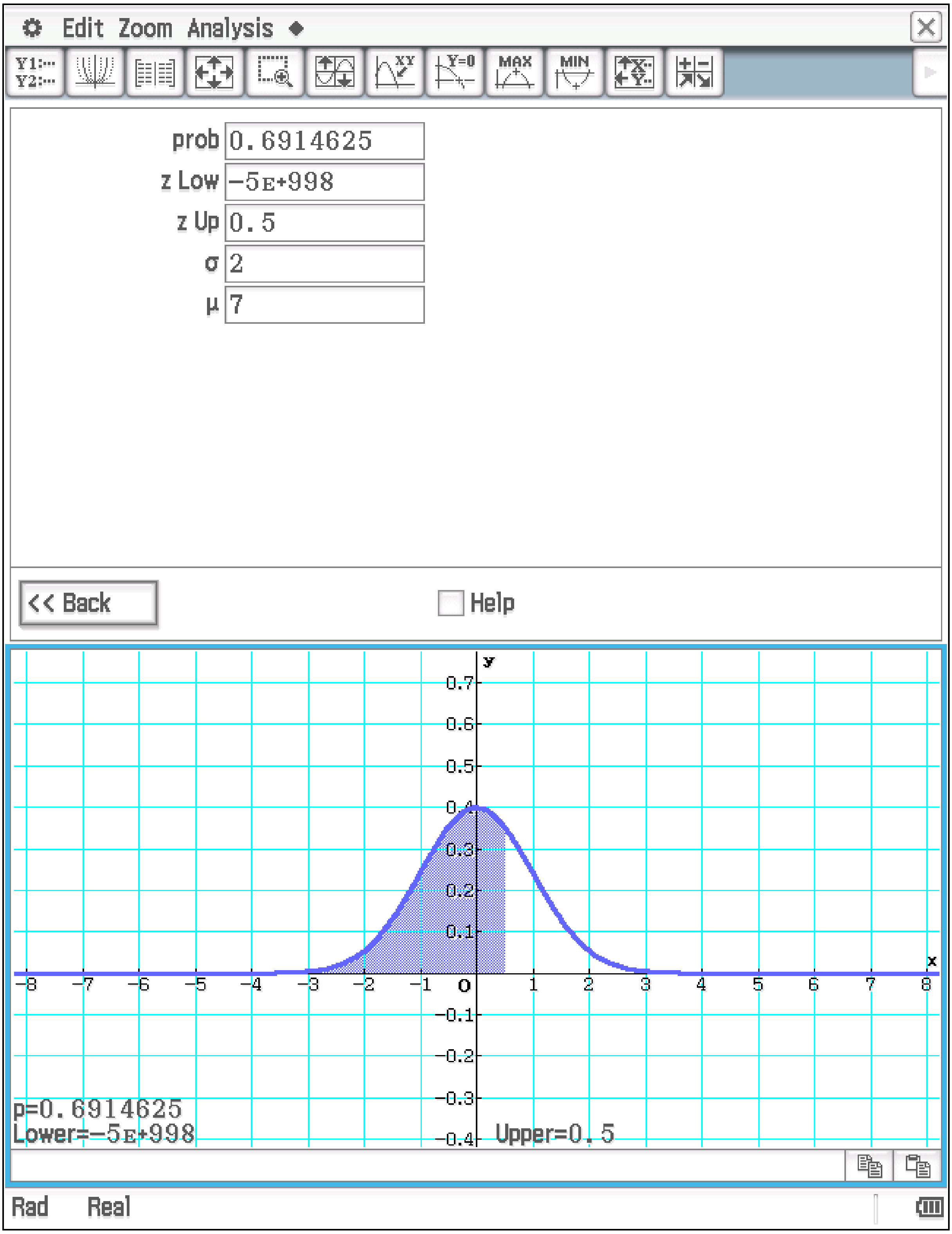
Stat−Editor

(a) P(X ≥ 9)



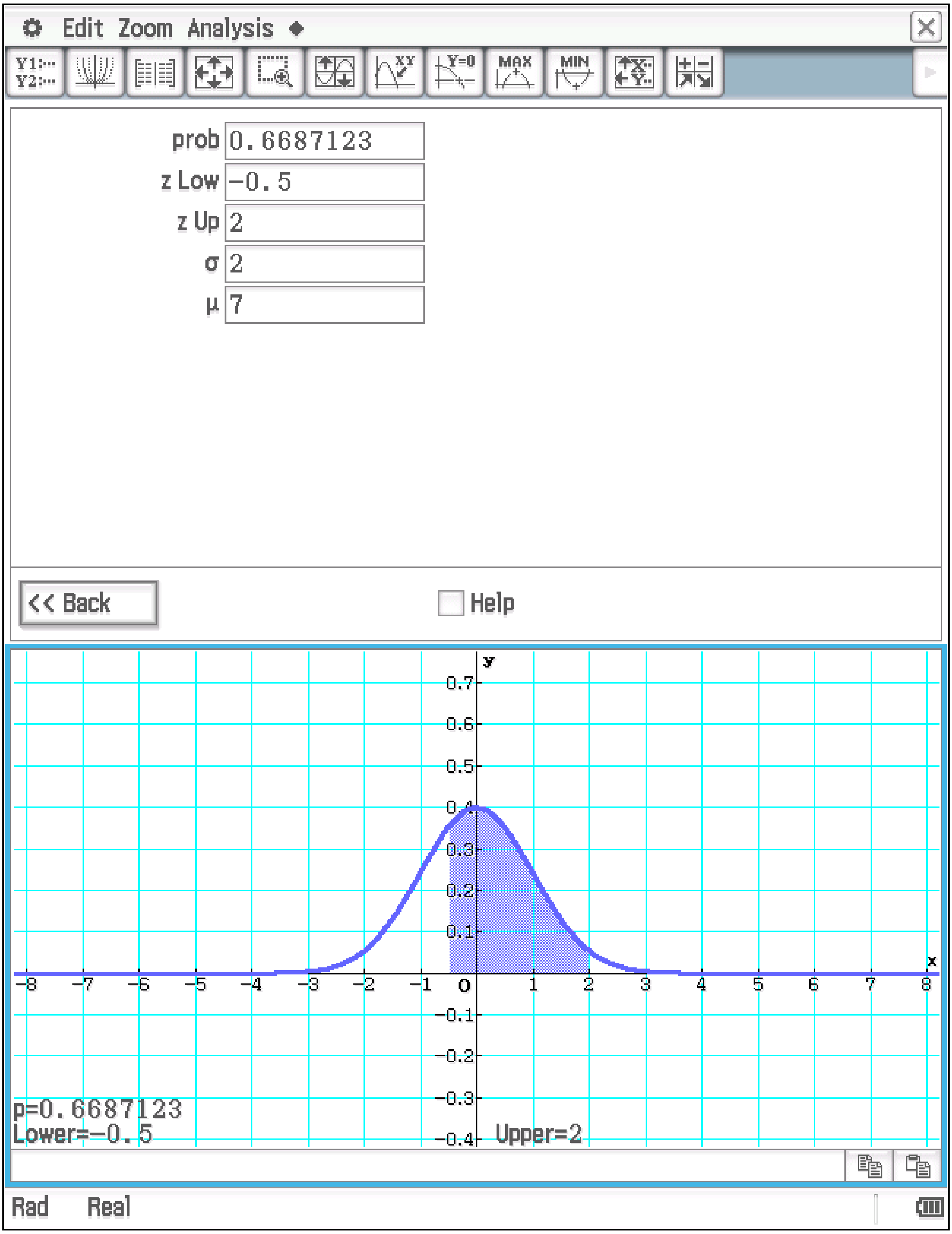
Edit Zoom Analysis
prob 0.1586553
z Low 1
z Up 5E+998
σ 2
μ 7
<< Back
Help
Rad Real

(b) P(X < 8)



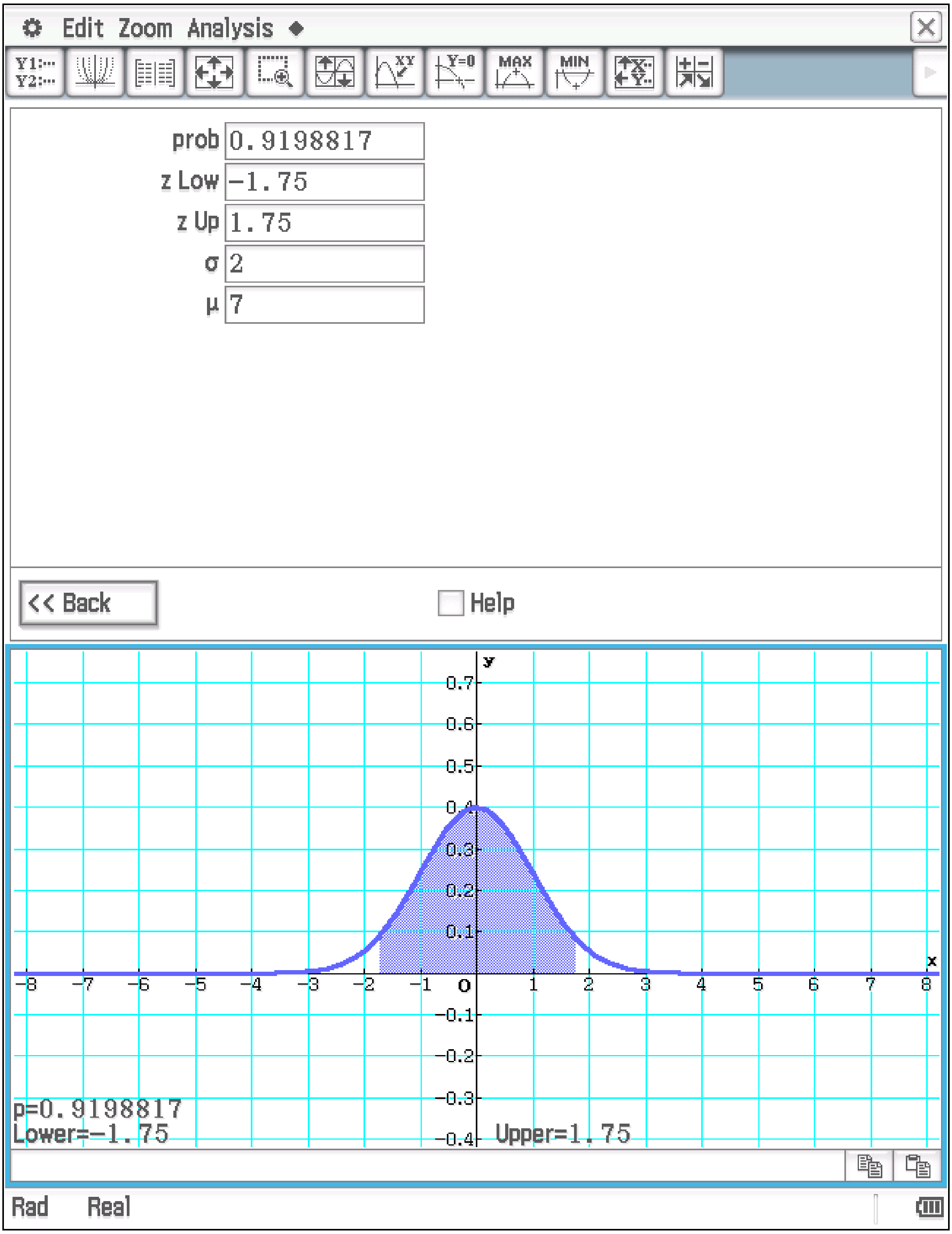
Edit Zoom Analysis
prob 0.6914625
z Low −5E+998
z Up 0.5
σ 2
μ 7
<< Back
Help
p=0.6914625
Lower=−5E+998
Upper=0.5
Rad Real

(c) P(6 ≤ X ≤ 11)



Edit Zoom Analysis
prob 0.6687123
z Low −0.5
z Up 2
σ 2
μ 7
<< Back
Help
p=0.6687123
Lower=−0.5
Upper=2
Rad Real

(d) P(|X-7|>3.5) = 1 - P(|X-7| ≤ 3.5) = 1 – 0.9198817 = 0.0801183



Prof. Dr. Ludwig Paditz, 01.11.2016

Statistik für Produktionstechniker

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

16. Sei X eine N(10,0.8)-verteilte Zufallsvariable. Bestimmen Sie jeweils $c\in\mathbf{R}$ so, dass gilt:

(a) $P(X \le c) = 0.85$

(b) $P(X > c) = 0.7$

(c) $P(|X - 10| \le c) = 0.9$

Lösung:

(a) $P(X \le c) = 0.85$

solve(normCDf(−∞, c, $\sqrt{0.8}$, 10)=0.85, c)

{c=10.92701421}

alternativ: Quantilfunktion

c=invNormCDf(0.85, $\sqrt{0.8}$, 10)

c=10.92701421

(b) $P(X > c) = 0.7$

solve(normCDf(c, ∞, $\sqrt{0.8}$, 10)=0.7, c)

{c=9.530961922}

alternativ: Quantilfunktion

c=invNormCDf(0.3, $\sqrt{0.8}$, 10)

c=9.530961922

(c) $P(|X - 10| \le c) = 0.9$

solve(normCDf(−c, c, $\sqrt{0.8}$, 0)=0.9, c)

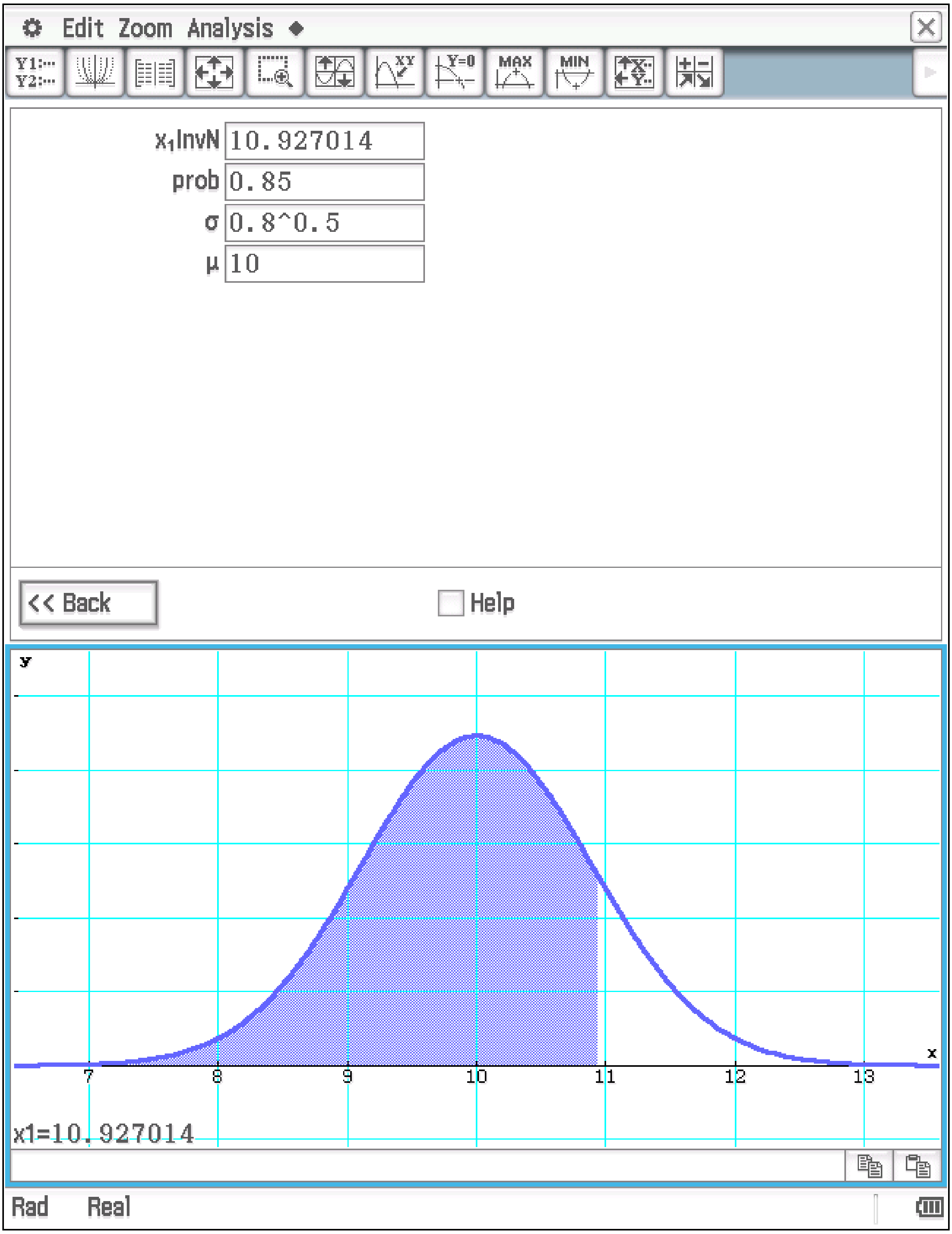
{c=1.471201809}

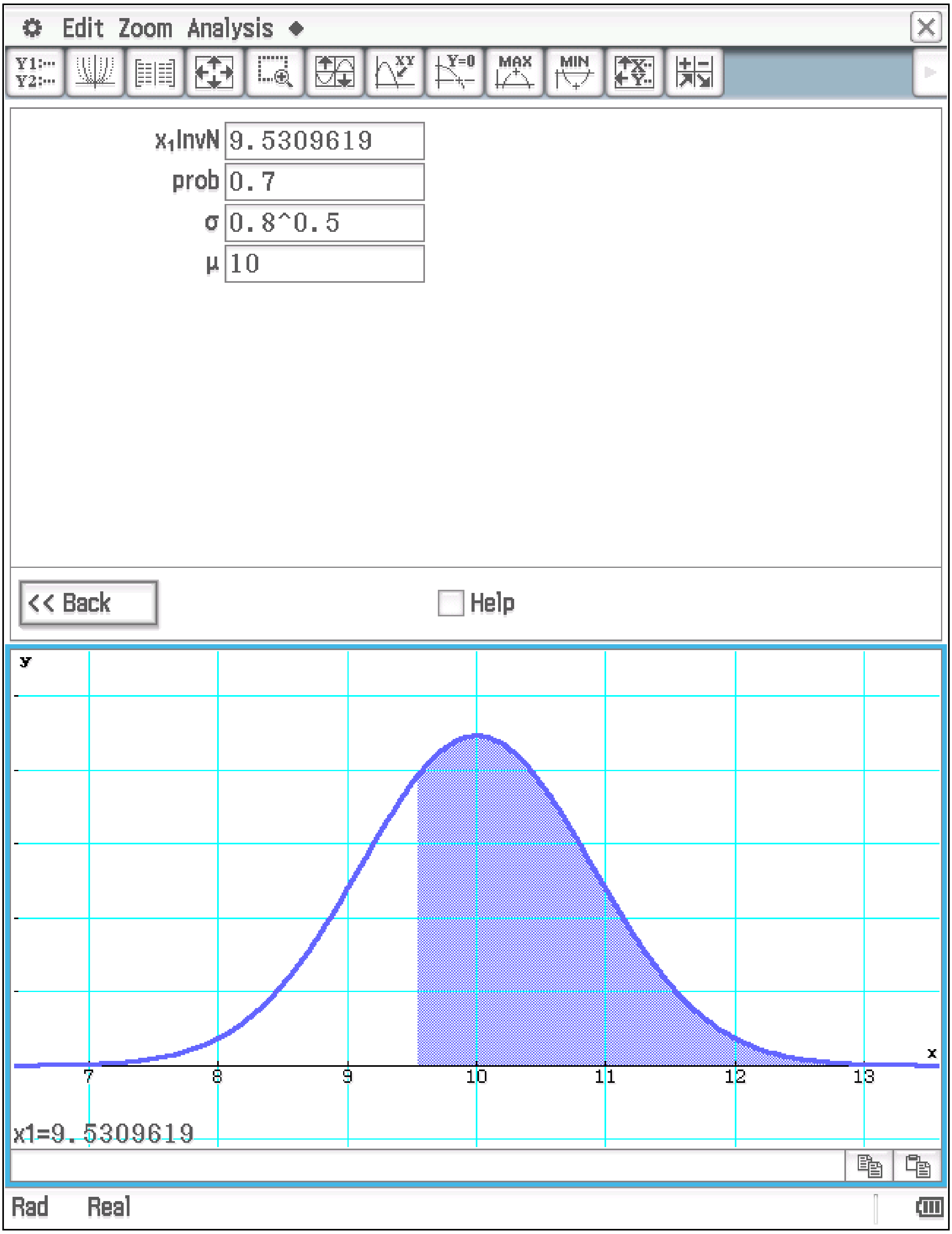
alternativ: Quantilfunktion

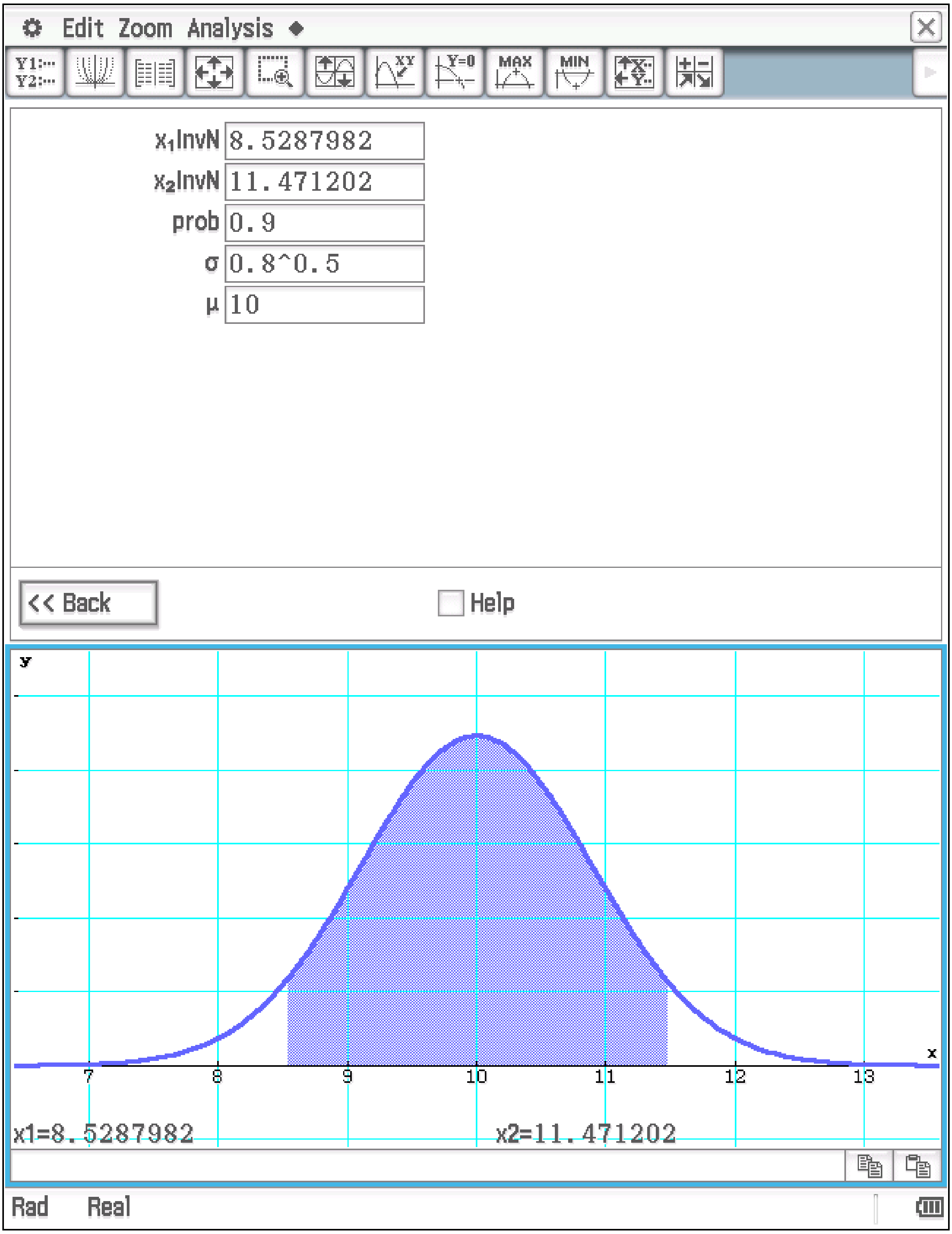
c=invNormCDf(0.95,$\sqrt{0.8}$,0)

c=1.471201809

Stat-Editor


Edit Zoom Analysis
x1InvN 10.927014
prob 0.85
σ 0.8^0.5
μ 10
<< Back
Help
y
x
7
8
9
10
11
12
13
x1=10.927014
Rad Real


Edit Zoom Analysis
x1InvN 9.5309619
prob 0.7
σ 0.8^0.5
μ 10
<< Back
Help
y
x
7
8
9
10
11
12
13
x1=9.5309619
Rad Real


Edit Zoom Analysis
x1InvN 8.5287982
x2InvN 11.471202
prob 0.9
σ 0.8^0.5
μ 10
<< Back
Help
y
x
7
8
9
10
11
12
13
x1=8.5287982
x2=11.471202
Rad Real

Prof. Dr. Ludwig Paditz, 01.11.2016

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

17. Die Lebensdauer der Computerchips eines bestimmten Produzenten werde durch eine normalverteilte Zufallsvariable mit dem Erwartungswert $4,4 \cdot 10^6$ Stunden und der Standardabweichung von $3 \cdot 10^5$ Stunden beschrieben.

(a) Ein Computerhersteller stellt an die von ihm verwendeten Chips die Anforderung, dass wenigstens 90% der Chips aus einer großen Lieferung eine Lebensdauer von mindestens $4 \cdot 10^6$ Stunden aufweisen. Sollte er mit dem Chipproduzenten ins Geschäft kommen?

(b) Wie groß ist die Wahrscheinlichkeit, dass in einer Lieferung von 100 Chips des Produzenten wenigstens vier Chips enthalten sind, deren Lebensdauer unter $3.8 \cdot 10^6$ Stunden liegt?

Lösung:
(a) X ... zufällige Lebensdauer,

$N(4.4 \cdot 10^6, (3 \cdot 10^5)^2)$-verteilt

Ansatz: $P(X \geq 4 \cdot 10^6) \geq 0.90$

normCDf$(4 \cdot 10^6, \infty, 3 \cdot 10^5, 4.4 \cdot 10^6)$

0.9087887803

Ja, die Forderung ist erfüllt!

(b) n=100 (Y ... zufällige Anzahl der fehlerhaften Chips)
Binomialverteilung mit $p=P(X<3.8\cdot10^6)$ (Forderung nicht erfüllt)

$p:=\text{normCDf}(-\infty,3.8\cdot10^6,3\cdot10^5,4.4\cdot10^6)$

0.02275013195

binomialCDf(4,100,100,p)

0.193889386

1−binomialCDf(0,3,100,p)

0.193889386

100p*(1−p)

2.223256344

Faustregel für ZGWS (Zentraler Grenzwertsatz) nicht erfüllt:

(Rechnung mit Stetigkeitskorrektur 0.5)

$\text{normCDf}(3.5,\infty,\sqrt{100p*(1-p)},100p)$

0.205665106

$1-\text{normCDf}(-\infty,3.5,\sqrt{100p*(1-p)},100p)$

0.205665106