HTW Dresden
Fak. Informatik/Mathematik
Prof. Dr. L. Paditz 07.01.2012

Arbeitsblatt zur Statistik - 18. HA im Fernstudium (10/023/21)

Warum die Statistik nicht zur absoluten Wahrheit führt, jedoch die Kritikfähigkeit fördert:

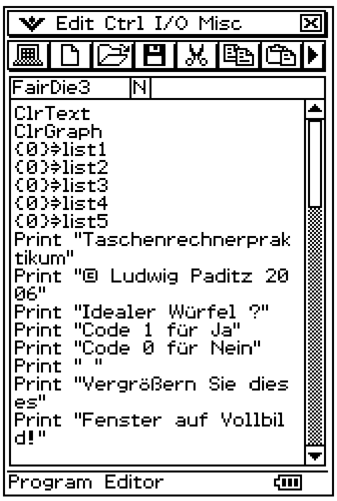
(Quelle/Programmidee: **Edwards, C.C.: Does a TI-8x Cast a Fair Die?** in: Eightysomething!, Vol. 6, No. 3, p. 9-10 (1997), siehe Internet: http://www.ti.com/calc/docs/act83stat.htm,
Aulenbacher, **Paditz,** Wabel-Frenk: Lehr- und Übungsbuch Mathematik Bd.3
(**Teil Stochastik: Beispiel 6.1 und Aufgabe 19.1**),
Fachbuchverlag Leipzig 1996 (ISBN 3-446-18715-4)), 2.Aufl. 2001: ISBN 978- 3-446-21682-2.

Übung:
Interaktives Arbeiten mit dem ClassPad 330 (TI-89) unter Nutzung von Programmen für gewisse Teilschritte und Ausführung von Zwischenschritten durch Direkteingabe von TR-Befehlen. Alle Rechnerergebnisse/ Grafiken sind schriftl. zu protokollieren.

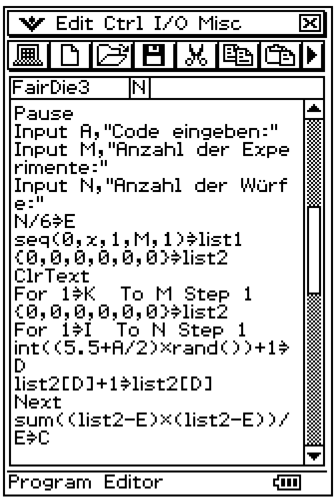
Inhalt:
Auf der Grundlage von M (z.B. M=300) Würfelexperimenten (Chi-Quadrat-verteilten Testgrößen), wobei in jedem Experiment N (z.B. N=100) Augenzahlen simuliert werden sollen, errechnet man gemäß dem Chi-Quadrat-Anpassungstest M Chi-Quadrat-verteilte Testgrößen. Die Auswertung dieser Testgrößen im Histogramm wird mit der Chi-Quadrat-Dichte-Funktion als statistische Prüfverteilung des Chi-Quadrat-Anpassungstests verglichen.
Schließlich werden weitere statistische Untersuchungen mit dem simulierten Datenmaterial durchgeführt.

Simulation der M (z.B. M=300) Chi-Quadrat-verteilten Testgrößen zum Würfelexperiment:
Starten Sie das CP-Programm „FairDie3“ und folgen Sie den Eingabefenstern!
Hinweis: Um Rechenzeit und Speicherplatz zu sparen, ist es ausreichend, z.B. nur M=250 oder M=200 zu wählen.

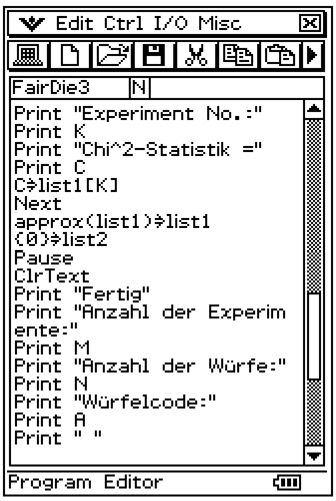
Eine Testgröße wird hierbei aus N (z.B. N=100) Würfeldaten gebildet, die wie folgt simuliert werden:
int((5.5+A/2) * rand( )) +1, hierbei ist A der Code (A=1 fair die, A=0 unfair die),
int(…) ist der Ganzteil**, rand(„ohne Argument“)** ist eine Zufallszahl aus (0,1), gleichmäßig stetig verteilt.
Der unfaire Würfel hat die simulierte Wahrscheinlichkeitsverteilung
P(X=k) = 2/11 für k=1,2,3,4,5 und P(X=k) = 1/11 für k=6.



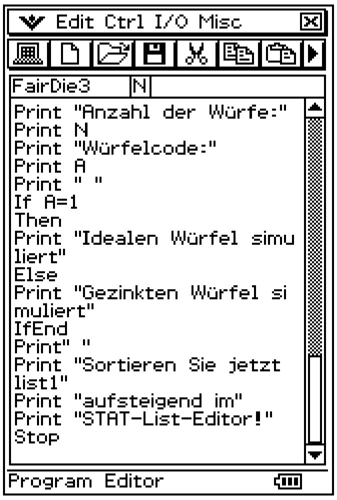

```
ClrText
ClrGraph
{0}⇒list1
{0}⇒list2
{0}⇒list3
{0}⇒list4
{0}⇒list5
Print "Taschenrechnerpraktikum"
Print "© Ludwig Paditz 2006"
Print "Idealer Würfel ?"
Print "Code 1 für Ja"
Print "Code 0 für Nein"
Print " "
Print "Vergrößern Sie dieses"
Print "Fenster auf Vollbild!"
```

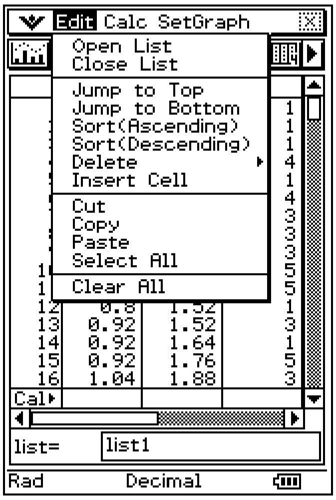


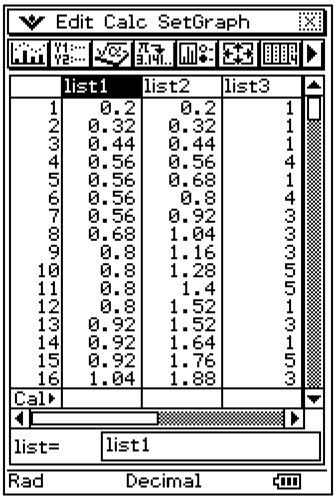

```
Pause
Input A,"Code eingeben:"
Input M,"Anzahl der Experimente:"
Input N,"Anzahl der Würfe:"
N/6⇒E
seq(0,x,1,M,1)⇒list1
{0,0,0,0,0,0}⇒list2
ClrText
For 1⇒K  To M Step 1
{0,0,0,0,0,0}⇒list2
For 1⇒I  To N Step 1
int((5.5+A/2)×rand())+1⇒D
list2[D]+1⇒list2[D]
Next
sum((list2-E)×(list2-E))/E⇒C
```




```
Print "Experiment No.:"
Print K
Print "Chi^2-Statistik ="
Print C
C⇒list1[K]
Next
approx(list1)⇒list1
{0}⇒list2
Pause
ClrText
Print "Fertig"
Print "Anzahl der Experimente:"
Print M
Print "Anzahl der Würfe:"
Print N
Print "Würfelcode:"
Print A
Print " "
```





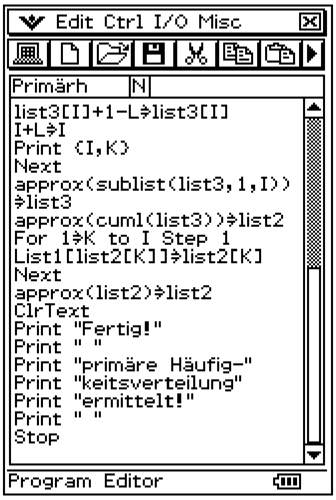


Aufgaben:

1a) Nutzen Sie das Programm **FairDie3** zur **Erzeugung von M=250 Chi-Quadrat-verteilten Zufallszahlen** (Parameter der Chi-Quadrat-Verteilung: 5 Freiheitsgrade) auf der Grundlage von jeweils **N=100** Würfen eines Würfels (vgl. Bsp. 6.1 u. Statistik-Aufg. 19.10 im LÜB Bd. 3). Am Ende der Datensimulation mögen die M simulierten Daten in der **Urliste list1** abgespeichert sein.
Hinweis: Rechenzeit ca. 15 min. (TI-89: 10MHz Prozessor-Chip), mit dem **RandSeed**-Befehl kann der Zufallszahlengenerator an einer festen Stelle gestartet werden, womit eine bestimmte Simulation reproduziert werden kann.

b) Erzeugen Sie aus der Urliste list1 die zugehörige **Variationsreihe** und speichern Sie diese wieder in list1 ab!
Hinweis: Sortierungsbefehl für die aufsteigende Reihenfolge: **SortA list1**)

In den letzten beiden Bildern erfolgte die Sortierung im STAT-Menü (Listeneditor) aufsteigend.
Hinweis: die Sortierung im Main-Menü ist unvorteilhaft (im Main-Menü wird mit CAS gerechnet, was die Sortierung verzögert). Im STAT-Menü wird rein numerisch sortiert, ohne CAS-Hintergrund, was die Sortierung sehr beschleunigt.



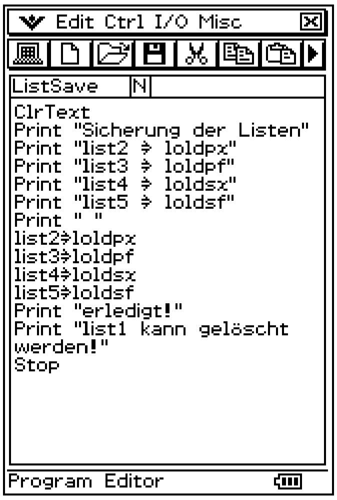
Primäre Häufigkeitstabelle

2a) Erzeugen Sie nun die **primäre Häufigkeitstabelle**, wobei list2 die geordneten Messwerte und list3 die zugehörigen absoluten Häufigkeiten enthalten soll. Nutzen Sie dazu das im vcp-file angegeben Programm **Primärh**. Sichern Sie list2 und list3 in **loldpx** und **loldpf**.

b) Speichern Sie nun die (absoluten) **Summenhäufigkeiten** in die Datenliste **loldps** ab!
Hinweis: **CumSum(list3)** /STO/ loldps . Nutzen Sie das Programm **ListSave** nach Ausführung von **Sekundh**, s. Aufg. 3a)

c) Stellen Sie die **empirische Verteilungsfunktion y21** als rechtsseitig stetige Treppenfunktion und die theoretische χ^2_5-Verteilungsfunktion (**y22**) im TR-Display grafisch dar. Nutzen Sie zur Definition dieser Funktionen die Funktionen y21(x)=**dotP(pf,px)/500** (500=2M) und y22(x)=**ChiCDf(0,x,5), s.u.**

Sekundäre Häufigkeitstabelle



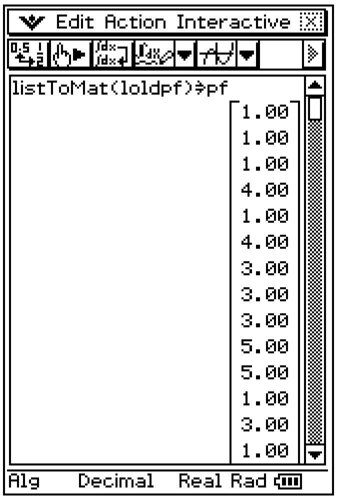
Sicherung der Daten vor dem Start einer neuen Simulation.

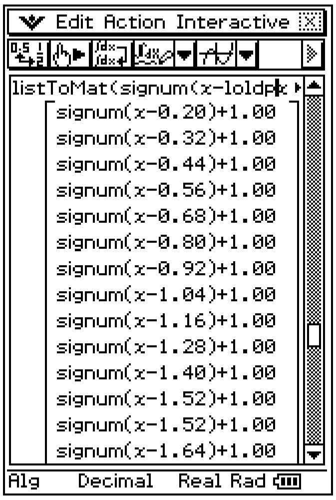
Step function without vertical line possible!

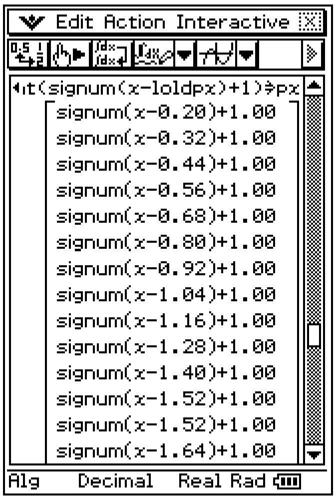
Use “Draw Plot” to draw the empirical distribution function:
list “loldpf” contains the frequencies (absolute values) sum(loldpf)=250.
and loldpx contains the sample data x.
(We have simulated chi-squared sample data with 5 degrees of freedom)
We want to draw an empirical distribution function with a primary frequency table:
We use the formulae y = dotP(pf,px)/500 , where pf and px are vectors with the primary frequency table. The step function we create with the signum-function:

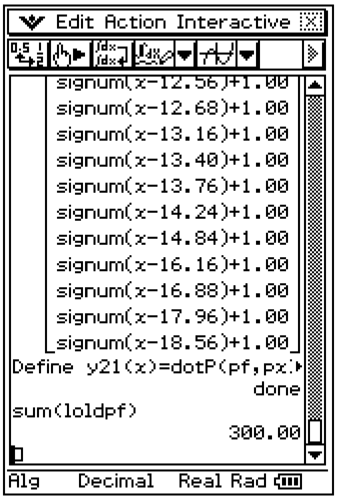
3a) Erzeugen Sie nun eine **sekundäre Häufigkeitstabelle** (Klassenbreite 1, Reduktionslage 0) mit dem Programm **Sekundh**. Die Klassenmitten sollen in Liste list4 und die zugehörigen Klassenhäufigkeiten in Liste list 5 abgespeichert werden. Anschließend werden diese Datenlisten in **loldsx** und **loldsf** gesichert.
Hinweis: list4 enthält gemäß dem Programm **Sekundh** als erstes Element zusätzlich die Koordinate 0 für den Anfangspunkt des Häufigkeitspolygons, dessen „Knickpunkte“ dann in list4, list5 stehen, wobei list5[1] = list5[2] festgelegt wird.

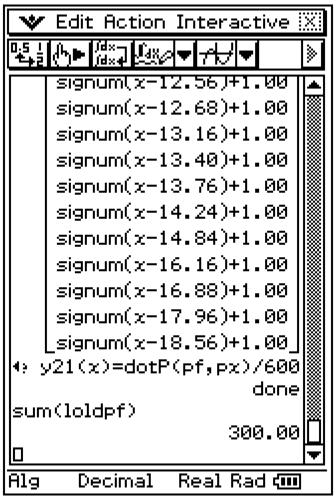
b) Stellen Sie das **Histogramm** zu loldpx und loldpf (**Plot1**) und das zugehörige **Häufigkeitspolygon** (**Plot2** mittels loldsx und loldsf) grafisch dar, einschließlich der χ^2_5-Dichtefunktion (**y3**), wobei die χ^2_5-Dichtefunktion über das Programm **defchidf( )** definiert werden kann.

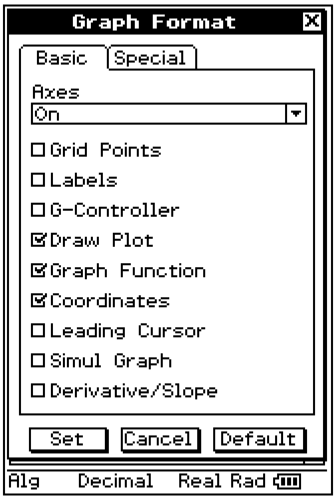


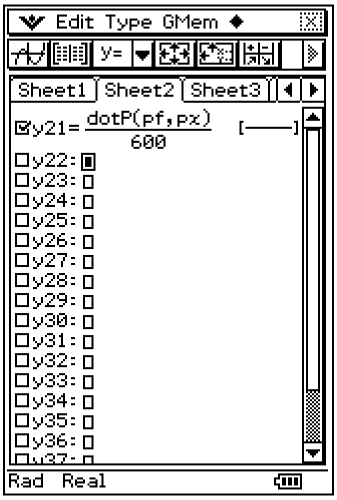


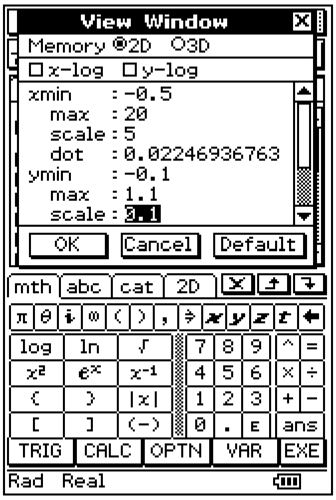


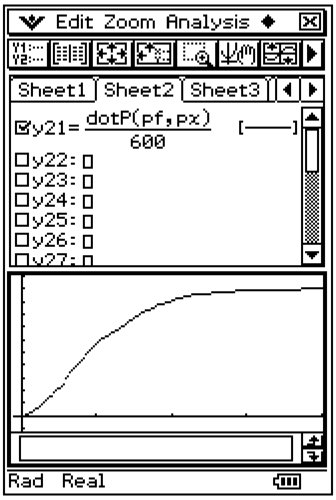








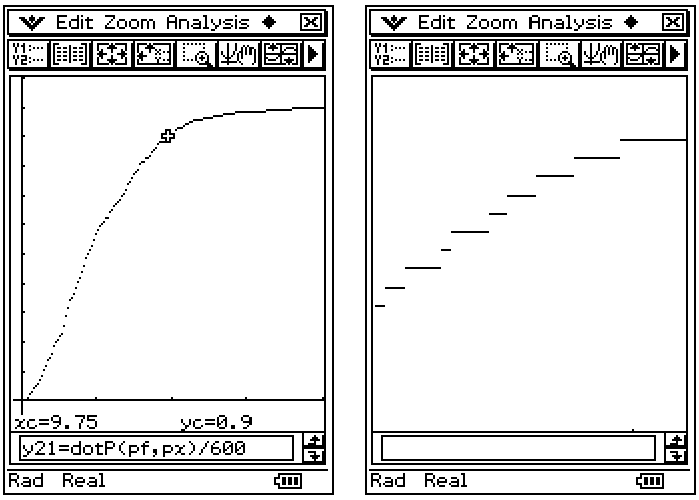




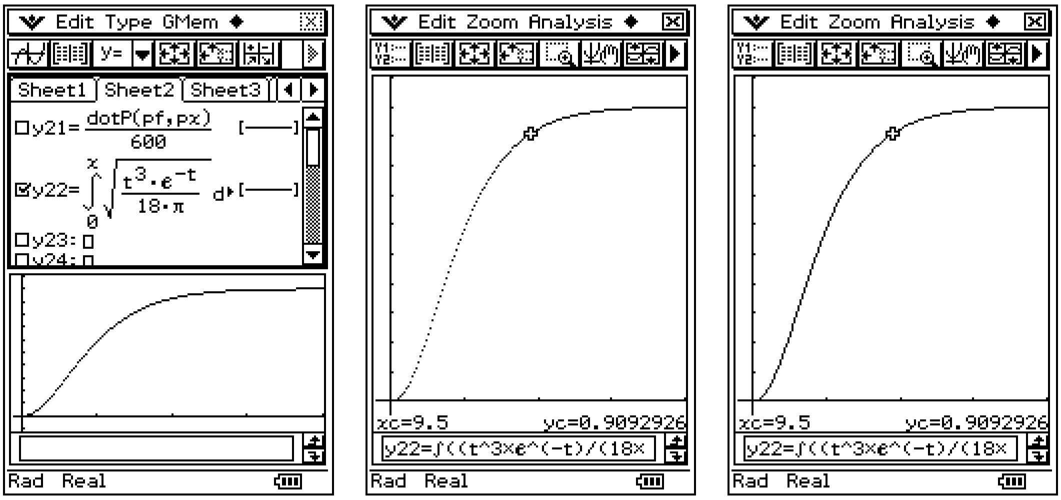
4) Stellen Sie die **Treppenkurve der relativen Summenhäufigkeiten** (y4) (Grafikmodus „Dot“) und erneut die theoretische Verteilungsfunktion (y22) graphisch dar. Die (absoluten) Summenhäufigkeiten sollen in der Liste loldss abgespeichert sein.

Hinweis: **CumSum(list5) - list5[1]** /STO/ **loldss**

Die Treppenkurve der rel. Summenhäufigkeiten wird durch die Funktionen y24(x)=**dotP(sf,sx)/500** (500=2M) definiert (dann Parameter-Modus, s.o.):

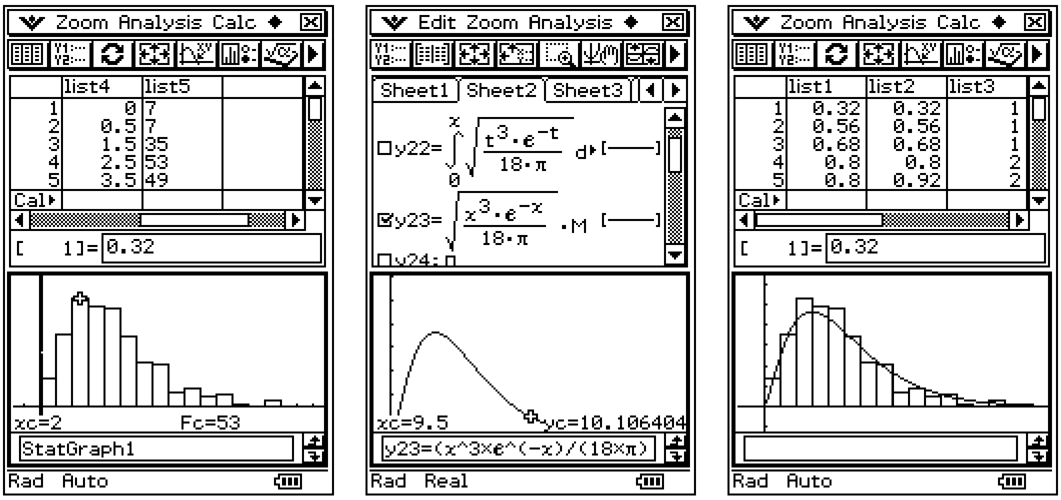


empirische Verteilungsfunktion

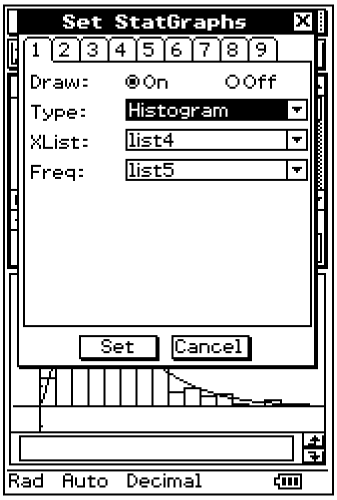


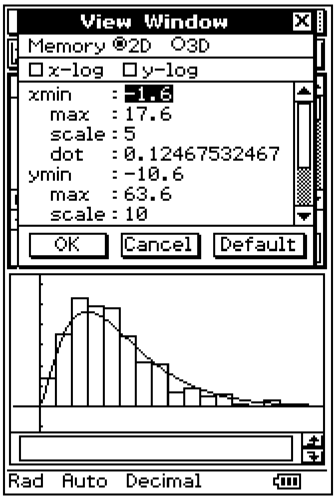
theoretische Verteilungsfunktion (pixelweise geplottet bzw. connected style)

Hinweis: das Zeichnen der theoretischen Verteilungsfunktion nimmt etwas Zeit in Anspruch, da jeder Pixelpunkt der Kurve als numerisches Integral berechnet wird.



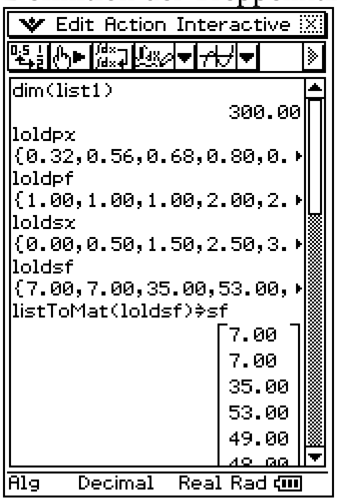
Histogramm und Dichtefunktion (idealer Würfel)

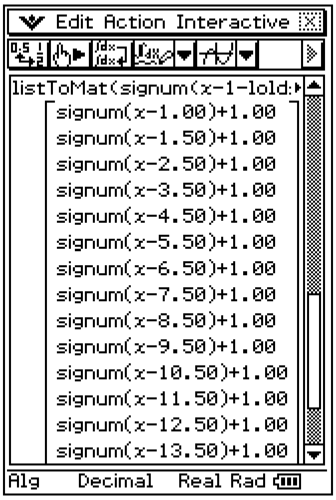


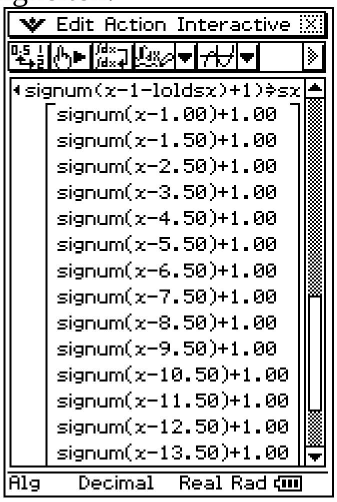


Automatische Fenstereinstellung

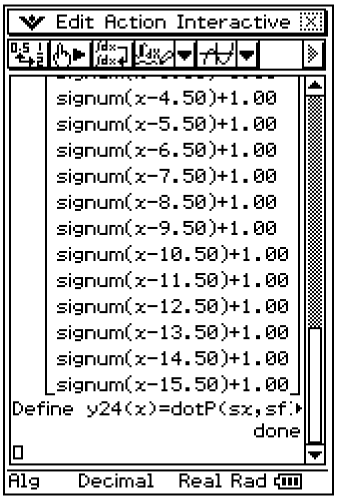
Definition der Treppenkurve der relativen Summenhäufigkeiten:

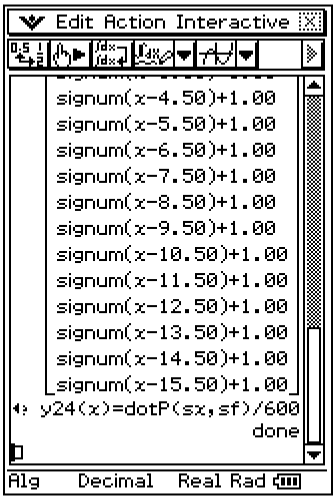


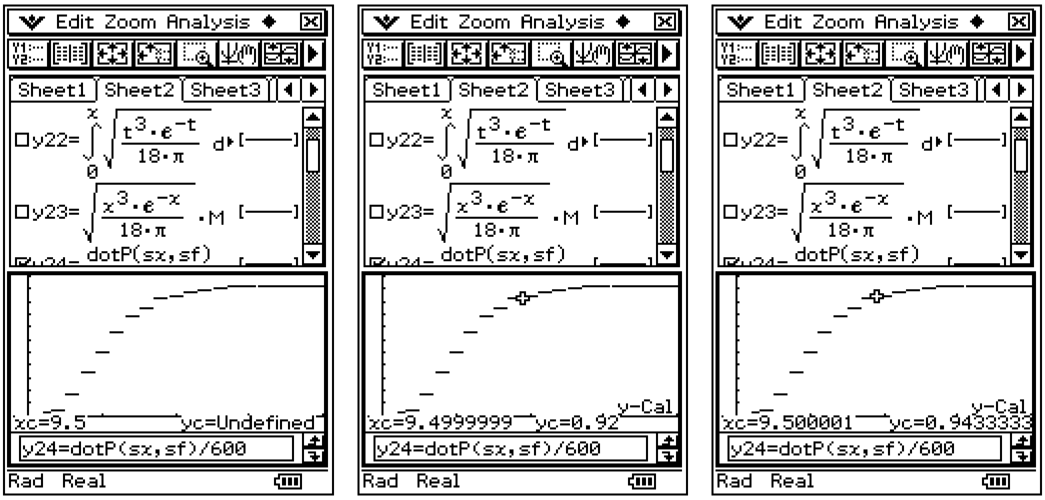




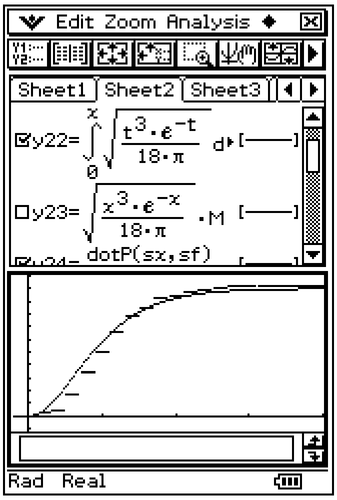
Verschiebung der Sprungstelle jeweils an das rechte Klassenende: signum(x-**1**-loldsx)





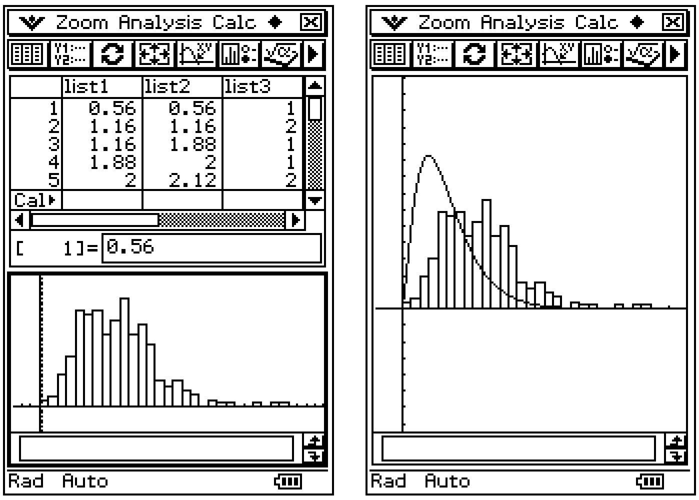


Treppenkurve der relativen Summenhäufigkeiten an der Unstetigkeitsstelle
(Die Treppenkurve ist per Def. rechtsseitig stetig, d.h. für x=9.5 gilt der Wert für x=9.5+0)



Treppenkurve und theoretische Verteilungsfunktion

Gezinkte Würfel:



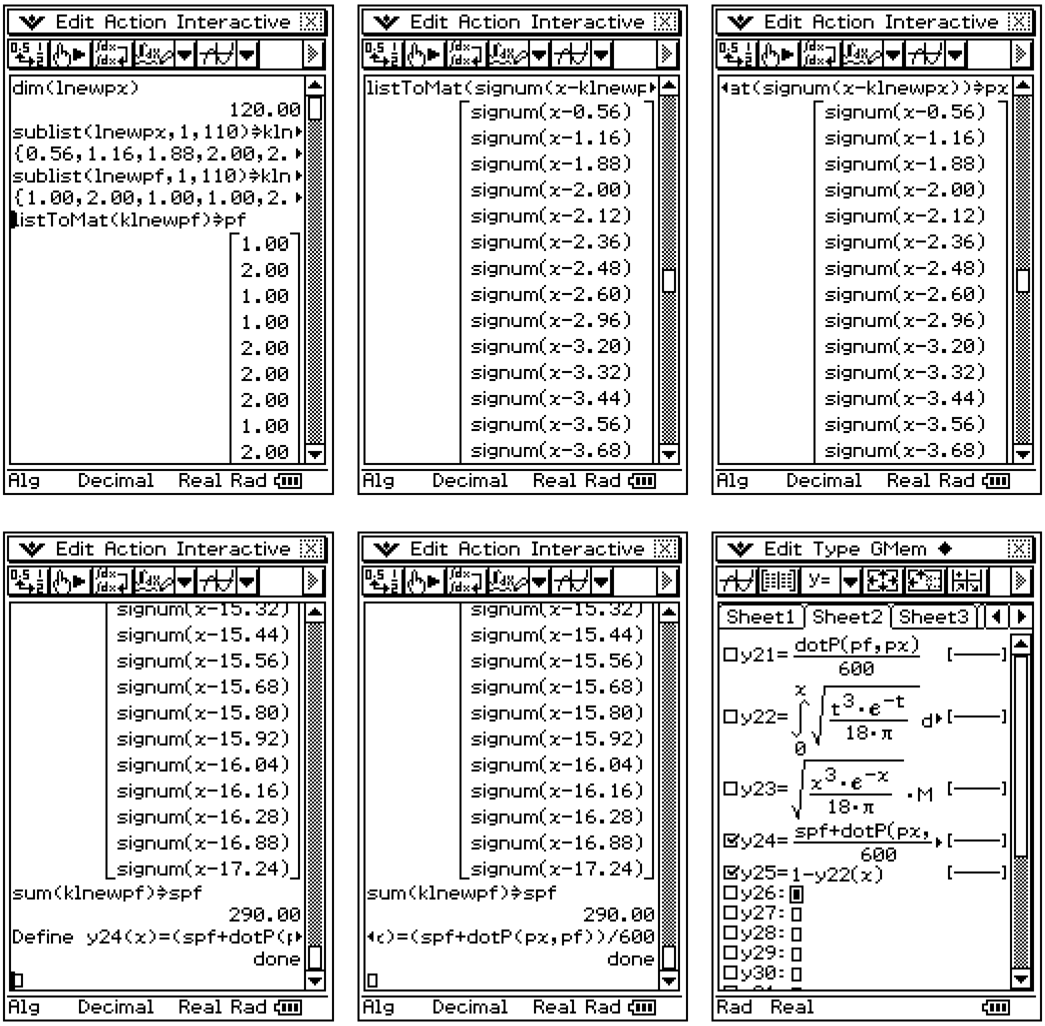
Histogramm und Chi^2-Dichte (andere Skalierung beachten)

Empirische Verteilungsfunktion (gezinkter Würfel):

5) Simulation des „gezinkten“ Würfels mit $p_k = 2/11$, k=1(1)5, $p_6 = 1/11$ (Die „6“ ist gegenüber den anderen Augenzahlen benachteiligt!)
Nutzen Sie dazu das bereits oben erwähnte Programm faidie9(code,m,n) mit dem code=0 .
Im Programm fairdie9(code,m,n) wird lediglich die Zeile **rand(6) /STO/ d** ersetzt durch **int((rand(11)+1)/2) /STO/ d** ersetzt.

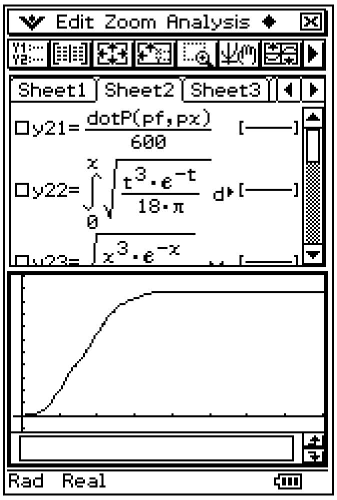
Die Aufgabenschritte 1) bis 4) sind zu wiederholen, wobei vorher die Listen list2 bis list5 in den Listen **loldpx, loldpf** bzw. **loldsx, loldsf**
gerettet werden. Entsprechend werden die Listen jetzt in **lnewpx, lnewpf, lnewps** bzw. **lnewsx, lnewsf, lnewss** abgespeichert.

Wegen Speicherüberlauf wird die primäre Häufigkeitstabelle nach 110 Datenpaaren abgeschnitten:

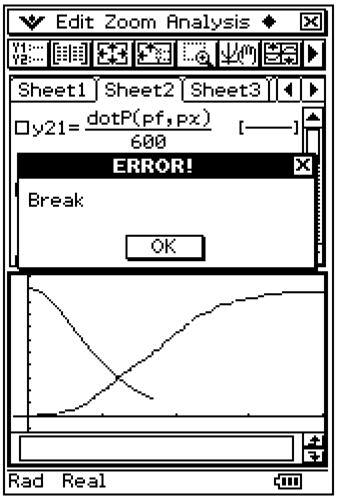


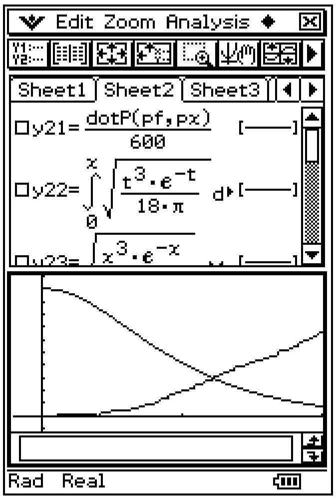
6) Untersuchen Sie die **Wahrscheinlichkeit ß für den Fehler 2. Art** im Experiment fairdie9(code,m,n), code=1, (Irrtumswahrscheinlichkeit α=5%, dann 10%), wenn die Alternative im Würfel (code=0), vgl. 5) besteht. Für welches α gilt hier α = ß ? Zugehöriges Quantil $\chi^2_{5,1-\alpha}$ angeben. Ermitteln Sie dazu grafisch den Schnittpunkt von y2 und **y7 = 1-y5** . Sichern Sie alle Funktionen in der Datenbasis GDB1.

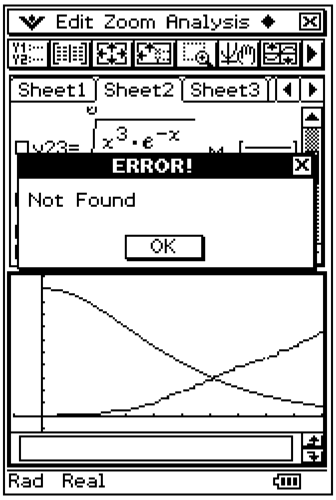
7) Schätzen Sie aus den Daten (lold..) empir. Mittelwert und empir. Streuung und vergleichen Sie mit den entsprechenden statistischen Kennzahlen einer theoretischen χ^2_5-Verteilung. Stellen Sie schließlich **Histogramm** (Plot1) und **Boxplot** (Plot5) für die alte Simulation (loldpx, loldpf) bzw. für die neue Simulation (Histogramm Plot3 und Boxplot Plot6 mit lnewpx, lnewpf) jeweils in einer Grafik dar.



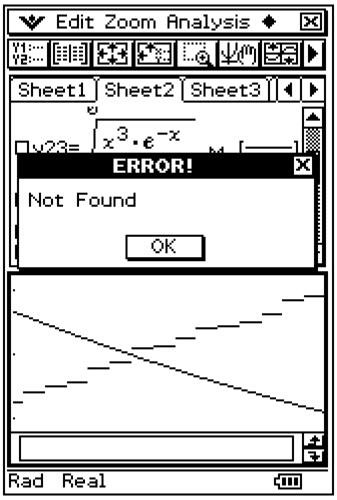
„abgeschnittene“ Treppenkurve erreicht nur „fast“ 1

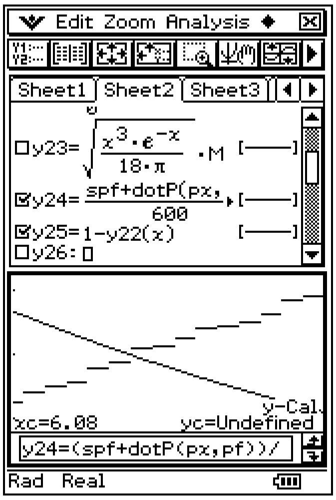


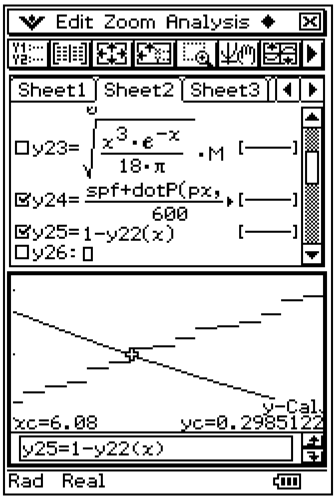




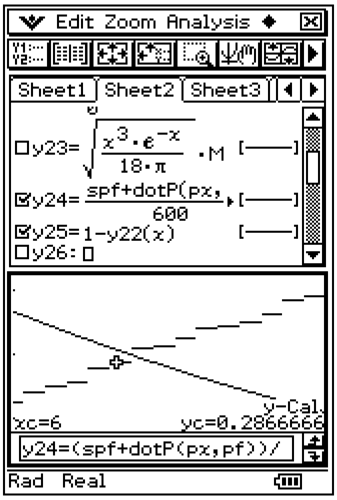
Abbruch nach erfolgtem Schnitt der empirischen Verteilungsfunktion mit 1-F(x), F(x) theoretische Chi^2-Verteilungsfunktion. Nun: Verkleinerung des Betrachtungsfensters.

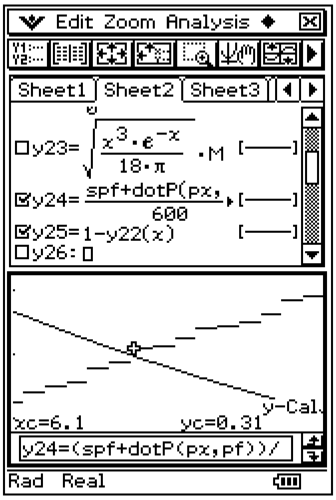


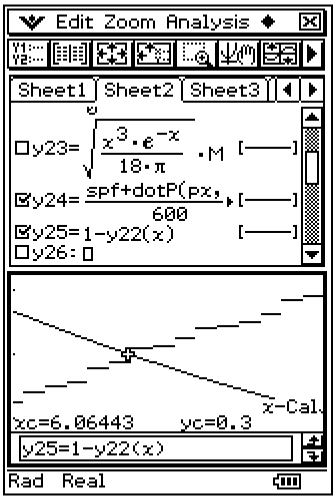




Es existiert tatsächlich kein Schnittpunkt.







Damit erfolgt der Sprung bei (exakt) 6,08 und es gilt für diese Simulation:
Mit einer Wahrscheinlichkeit von 0,2985 werden sowohl der Fehler 1. Art wie auch 2. Art begangen, wenn man als Chi^2-Quantil den Wert 6,08 zur Entscheidung verwendet, ob Ho abzulehnen oder nicht abzulehnen sei.

8) Beantworten Sie abschließend die Frage: Ist ein Würfel zu beanstanden, wenn er in 100 Würfen die Augenzahlen 1 bis 6 mit den Häufigkeiten 15, 16, 18, 17, 16, 18 (vgl. LÜB Bd.3, Beisp. 6.1 bzw. Aufg. 19.10) zeigt?

Die früher genutzten TI-89-Programme sind aus dem Internet abrufbar:

fairdie9(code,m,n)	unter	http://www.informatik.htw-dresden.de/~paditz/fairdie9.89p	bzw.	fairdie9.txt
primfreq(list1)	unter	http://www.informatik.htw-dresden.de/~paditz/primfreq.89p	bzw.	primfreq.txt
secufreq( )	unter	http://www.informatik.htw-dresden.de/~paditz/secufreq.89p	bzw.	secufreq.txt
defempvf( )	unter	http://www.informatik.htw-dresden.de/~paditz/defempvf.89p	bzw.	defempvf.txt
defchivf( )	unter	http://www.informatik.htw-dresden.de/~paditz/defchivf.89p	bzw.	defchivf.txt
defchidf( )	unter	http://www.informatik.htw-dresden.de/~paditz/defchidf.89p	bzw.	defchidf.txt
deftrels( )	unter	http://www.informatik.htw-dresden.de/~paditz/deftrels.89p	bzw.	deftrels.txt
plot1234( )	unter	http://www.informatik.htw-dresden.de/~paditz/plot1234.89p	bzw.	plot1234.txt

Ein Programmvariante zum TI-83 mit weiteren Hinweisen (z.B. PC-Variante unter Windows 3.1 bzw. Windows 95) findet man unter der Internetadresse: http://www.informatik.htw-dresden.de/~paditz/fairdti2.html

Für Classpad siehe: http://www.informatik.htw-dresden.de/~paditz/fairdie_simulation_CP300.pdf

Link zum vcp-file: http://www.informatik.htw-dresden.de/~paditz/FairDie2012.vcp

(Hinweis zum engl. Text: **NCTM** bedeutet: **National Council of Teachers of Mathematics**)