Dr. Alfred Eisler

Skriptum zur Wahrscheinlichkeitsrechnung

Themenbereich	
Skriptum zur Wahrscheinlichkeitsrechnung	
Inhalte	**Ziele**
• Einführung der Binomialverteilung mit Beispielen • Kenngrößen wie Erwartungswert und Varianz mit ihren speziellen Eigenschaften. • Übergang zur stetigen Verteilung; Vergleich der diskreten mit der stetigen Verteilung, wobei sich die Notwendigkeit des Übergangs aus der langen Rechenzeit des TI 92 ergibt. • Beispiele zur Normalverteilung	• Anhand von Beispielen und Analogieschlüssen soll die Binomialverteilung eingeführt werden. • Der Schüler soll erkennen, daß bei großen Zahlen der Rechner “zu lange” braucht und daraus auf die Notwendigkeit eines anderen Modells schließen. • Es soll die Fähigkeit gestärkt werden, eigene Formeln zu erarbeiten. • Die Tabelle soll als Lösungswerkzeug eingesetzt werden.
[Kurzbeschreibung ...]siehe Inhalte	

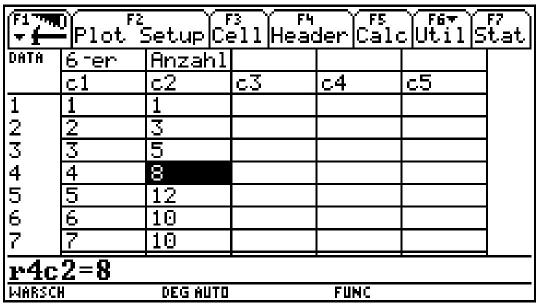
Wahrscheinlichkeitsrechnung

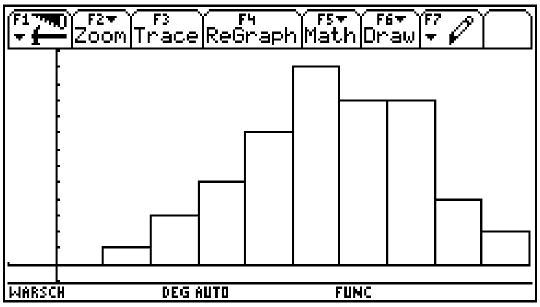
BINOMIALVERTEILUNG

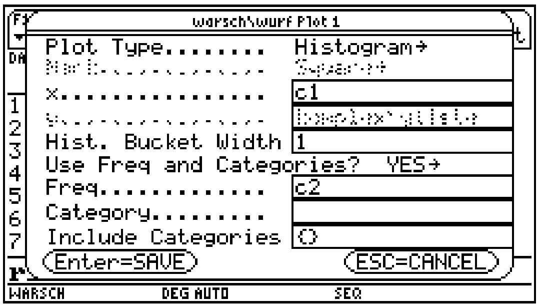
Einstiegsbeispiele :

Jeder Schüler würfelt zu Hause 30 mal; gezählt wird die Anzahl der 6er. Die Daten werden zusammengefaßt und mit dem TI-92 in einem Histogramm dargestellt.

Eine typische Tabelle und ein typisches Histogramm könnten etwa so aussehen :







Für die Tabelle verwendet man den *Data/Matrix Editor*, als *Plot Type* definiert man *Histogramm*, für die *x-Werte* nimmt man die Spalte *c1*, als *Freq* verwendet man *c2*.

Für die **Binomialverteilung** gilt die bekannte Formel:

$$P(X = k) = \binom{n}{k} \cdot p^{k} \cdot (1 - p)^{n-k}$$

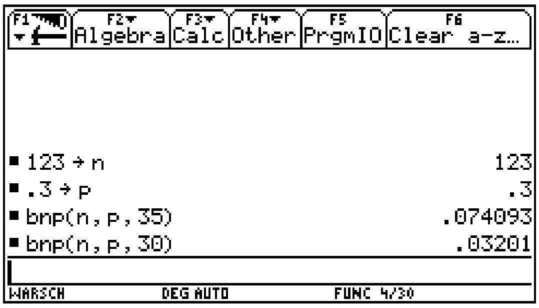
Statt P(X=k) schreibt man auch $b_{n,p}(k)$.
Es soll nun die allgemeine Formel auf dem TI 92 eingegeben werden. Sie wird mit **bnp(n,p,k)** bezeichnet.

nCr(n,k).p^k.(1-p)^(n-k) --> bnp(n,p,k)

Bsp: Ein Glücksrad wird 2x gedreht. Dabei ist die Wahrscheinlichkeit zu gewinnen p.
Stelle die Wahrscheinlichkeitsverteilung der Trefferhäufigkeit k für k = 0, 1, 2 durch eine Tabelle dar.

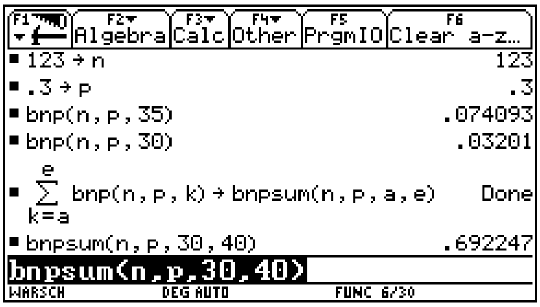
Bsp :
1) In einem Autobus befinden sich 30 Personen. Im Durchschnitt sind aus der Sicht der Zöllner 10% Schmuggler. Wie groß ist die Wahrscheinlichkeit, bei zufälliger Auswahl von 3 Personen keinen, genau einen, genau zwei, genau drei Schmuggler zu erwischen?
n = 3, p = 0.1, k = 0,1,2,3
Für die Berechnung soll die Formel (Funktion) bnp(n,p,k) verwendet werden.

2) In einer Schule befinden sich 750 Schüler. 30% sind fehlsichtig. Der Schularzt untersucht die ersten Klassen (123 Schüler).
a) Wie groß ist die W, genau 35 fehlsichtige Schüler zu erhalten
b) genau 30



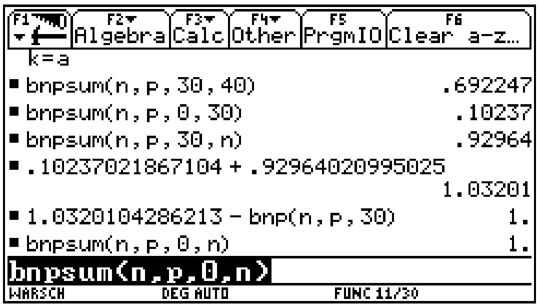
genau 35
genau 30

c) mindestens 30 und höchstens 40
1. Stufe : 11 Berechnungen, 11 Summanden
2. Stufe : neue Formel **bnpsum(n,p,a,e)**



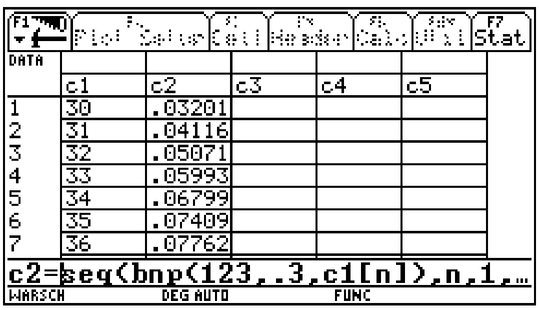
Mindestens 30, höchstens 40 fehlsichtige.

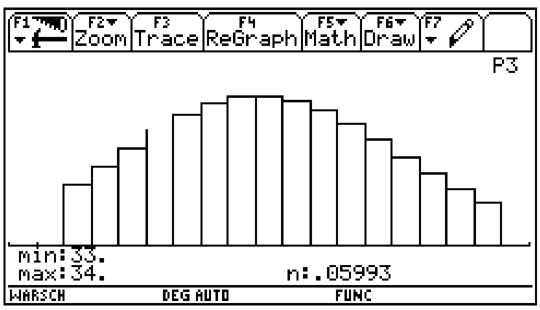
d) höchstens 30 fehlsichtige
e) mindestens 30 fehlsichtige
Die Summe aus d) und e) ergibt mehr als 1 - warum? P(X=30) abziehen.
Wie muß man die Frage anders stellen, damit 1 herauskommt?



höchstens 30
mindestens 30
Summe !!
P für k=30 abziehen.
Allgemeines Ergebnis !

f) Erstelle ein Histogramm der Wahrscheinlichkeiten für k = 30 bis 45.





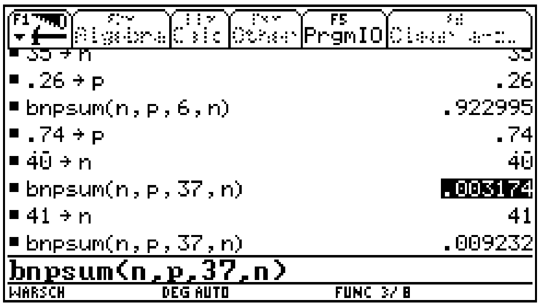
Weitere Beispiele

Der Anteil der Linkshänder wird mit 1% der Bevölkerung angenommen. Berechne die P dafür, daß in einer Klasse mit 29 Schülern genau 2 Linkshänder, mindestens 3 Linkshänder sitzen.
Zeichne ein Histogramm für die Wahrscheinlichkeiten, in dieser Klasse 0 bis 5 Linkshänder zu finden.

Beim Pfeilwerfen rechnet man bei 100 Würfen mit 9 Volltreffern. Wie groß ist die P, daß ein Schütze mit 25 Würfen mindestens 3 Volltreffer erzielt. Schätze das Ergebnis vorher ab!
Erstelle ein Histogramm für k = 2 bis 5 Volltreffer!

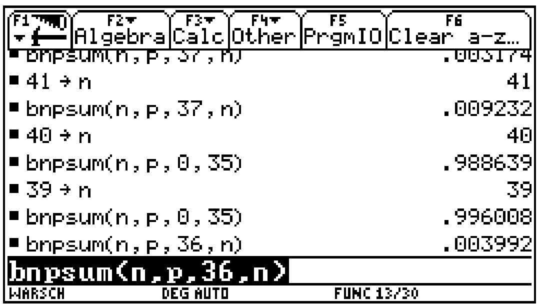
Ein Industriebetrieb kann seinen Energiebedarf an 80% aller Arbeitstage durch Eigenproduktion decken. Berechne die P, daß er während der kommenden Arbeitswoche (5 Arbeitstage) mindestens 3ma zusätzliche Energielieferungen benötigt! Wie sieht die Verteilung (grafisch) für k = 0 bis 5 aus?

Bsp : In einem Hotel werden erfahrungsgemäß nur 74% der Reservierungen auch wirklich in Anspruch genommen. Wie groß ist die Wahrscheinlichkeit, daß bei voller Reservierung von den 35 Betten mehr als 5 Betten frei bleiben. Um wieviele Personen darf höchstens überbucht werden, damit mit einer P von 1% mehr als 1 Person kein Bett mehr bekommt?



erster Versuch mit 40
zu klein
zweiter Versuch mit 41
OK !!

Zusatz : Wieviele Personen dürfen gebucht werden, damit mit 99,5%iger Wahrscheinlichkeit alle ein Bett bekommen?



erster Versuch mit 40.
Zu viele!
Versuch mit 39; OK!!
Andere Variante (es kommen mehr als 35)

Einführung des Erwartungswertes

Für den Erwartungswert gilt die Formel :

$$E(X) = \sum_{k=0}^{n} P(X = k) * k$$

Im TI 92 : Die Formel als *ew(n,p)* in den Rechner eingeben.
∑(bnp(n,p,k),k,0,n) --> *ew(n,p)*

Für obige Beispiele soll der Erwartungswert ausgerechnet werden. Wir kontrollieren (vergleichen) mit den Histogrammen.

Wir berechnen :
ew(3,0.5) ergibt *1.5*
ew(m,t) ergibt kein vernünftiges Ergebnis (Grenzen des Rechners)
ew(10,t) ergibt *10.t*
ew(43,t) ergibt *43.t* **Dauert!!**

Was kann man aus den letzten beiden Beispielen für e(n,p) schließen

eb(n,p) = n.p Erwartungswert bei der Binomialverteilung (iin Rechner eingeben!)

Übungsbeispiele

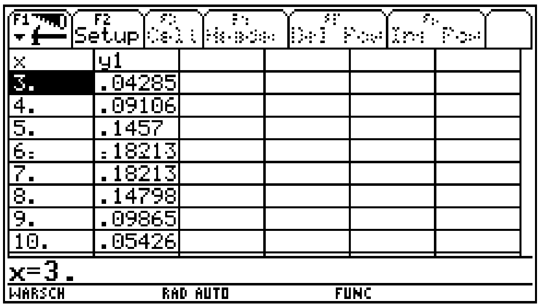
1. Bei einer Scheibe beträgt die Trefferwahrscheinlichkeit 2/3. Wie groß ist bei
a) 2 mal schießen die P für 0 Treffer, 1 Treffer, 2 Treffer ?
b) 3 mal schießen die P für 0 Treffer, 1 Treffer, 2 Treffer, 3 Treffer?
2. Franz spielt gegen Grete Tennis. Aus der Erfahrung weiß man, daß Grete mit der Wahrscheinlichkeit 0,52 gewinnt.
Wie groß ist die P, daß Franz von 10 Spielen 2,3 oder 8 gewinnt?
3. Eine Maschine erzeugt ein Produkt, wobei bei der Produktion im Mittel 2% schadhafte Teile anfallen.
a) Wie groß ist die P, daß von einer Tagesproduktion von 400 Stück, mehr als 10 Stück schadhaft sind
b) Weniger als 1% schadhaft sind?
c) genau 5 Stück schadhaft sind
d) Mehr als 395 Stück fehlerfrei sind?
e) Wieviele schadhafte Stücke sind pro Tag zu erwarten
f) Auf wieviel % muß der Auaschuß reduziert werden, damit pro Tag nur 6 defekte Stücke zu erwarten sind?

Bsp : Ein homogener Würfel wird 20 mal geworfen. X gebe die Anzahl der 5er oder 6er an.
Ermittle die Wahrscheinlichkeitsverteilung rechnerisch und grafisch.
Für die Lösung dieser Aufgabe stehen die folgenden Möglichkeiten zur Verfügung :

1. Eingabe im Data/Matrix Editor und Darstellung mittels Plot.
2. Verwenden einer Tabelle
3. Eingabe der Werte im Sequence Mode und zeichnen der Funktionen
4. Ein eigenes Programm für die grafische Darstellung.

1) Wurde bereits oben behandelt.

2) Arbeiten mit der Tabelle
Wir setzen im Home-Screen n=20 und p=1/3 und definieren im y-Editor als Funktion y1(x)=bnp(n,p,x).



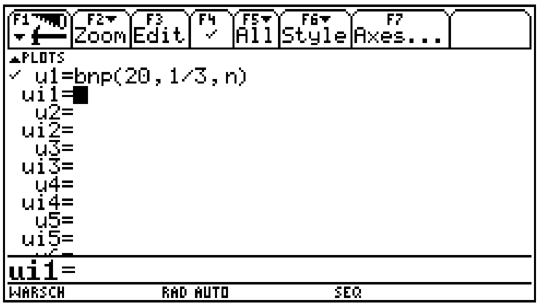
x	y1
3.	.04285
4.	.09106
5.	.1457
6.	.18213
7.	.18213
8.	.14798
9.	.09865
10.	.05426

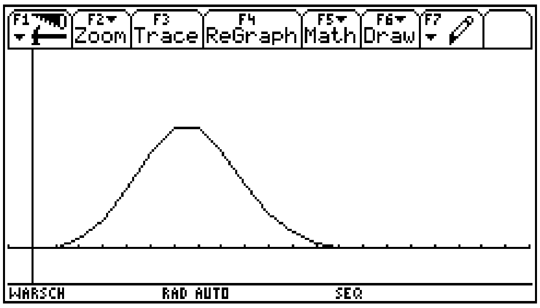
x=3.

Man erhält dann die folgende Tabelle für die Wahrscheinlichkeiten. Eine grafische Darstellung ist hier nicht möglich, weil für beliebige reele Werte kein Funktionswert existeirt.

3) Arbeiten mit dem Sequence-Mode
Der Funktionsmodus wird umgestellt (*MODE/Function/Sequence*).

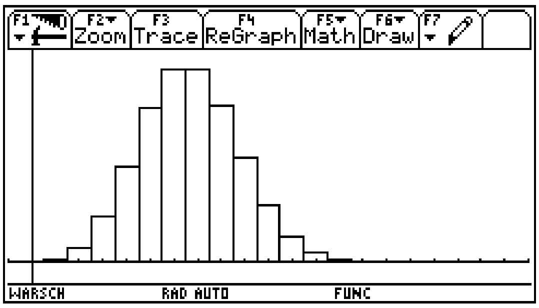
Jetzt kann für u1 die Funktion bnp eingeben werden. Als Variable ist hier immer ein n zu verwenden. Mit ◆GRAPH erhält man die folgende Grafik.





4) Eigenes Programm

Eingabe *Binplot(n,p)*



Das Programm stellt die Bildschirmkoordinaten selbstständig ein.
Man erhält ein Balkendiagramm.

Begriff der VARIANZ, STANDARDABWEICHUNG

Für die Varianz gilt die bekannte Formel : $Var(X) = \sum_{i=0}^{n} (x_i - E(X))^2.P(X=x_i)$

wobei man bei einer Binomialverteilung schreiben kann : $Var(X) = \sum_{k=0}^{n} (k - E(X))^2.P(X=k)$

Außerdem gilt für die Varianz : Var(X) = E(X²) - E(X)²

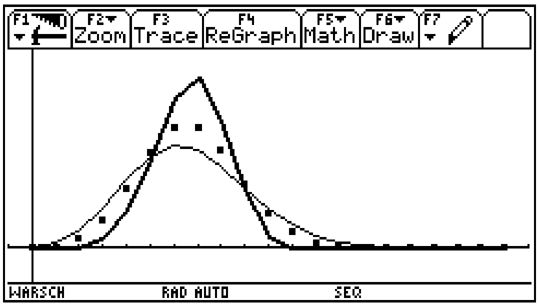
und für die Standardabweichung : $\sigma = \sqrt{Var(X)}$

Für alle Berechnungen wird nun eine Formel für die Varianz im Rechner definiert; wir speichern als **var(n,p)**.
Für größere Werte von n sind die Rechnungen sehr aufwändig, daher verwendet man besser den Verschiebungssatz und definiert speziell für die Binomialverteilung

$$\sum_{k=0}^{n} (k^2.bnp(n,p,k)) - eb\ (n,p)^2 \quad \rightarrow \quad var1(n,p)$$

Damit erhält man die Ergebnisse etwas schneller.
So erhält man etwa var1(20,q) = -20.q.(q-1) oder var1(27,q) = -27.q.(q-1).
Ein allgemeines Ergebnis var1(n,p) kann der Rechner nicht liefern.
Wir schließen trotzdem für die Binomialverteilung auf folgende Formel : varb(n,p) = n*p*(1-p) und geben diese in den Rechner ein.

Wir wollen nun den Begriff der Varianz genauer untersuchen. Dazu verwenden wir die Verteilungen mit
n = 200, p = 1/30, Line
n = 20, p = 1/3,Sqare
n = 10, p = 2/3, Thick und stellen Sie gleichzeitig grafisch dar.

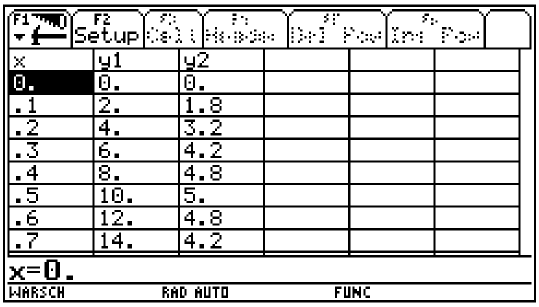


Der Erwartungswert ist für alle drei Verteilungen gleich.

Allerdings wird für größeres p das Maximum stärker ausgeprägt, die Ränder werden steiler, die Verteilung wird schärfer.

Berechnet man jeweils die Varianz, so erhält man die Werte 6,44 4,44 und 2,22. Schärfere Verteilungen haben eine kleinere Varianz (Streuung). Die Varianz stellt somit ein Maß für die Abweichung vom Erwartungswert, für die Breite der Verteilung, für die Schärfe der Verteilung dar.

Vergleicht man verschiedne Verteilungen (n=20, p=0,1 bis p=0,9) und stellt man den Erwartungsert und die Varianz in einer Tabelle dar, so erhält man :



x	y1	y2
0.	0.	0.
.1	2.	1.8
.2	4.	3.2
.3	6.	4.2
.4	8.	4.8
.5	10.	5.
.6	12.	4.8
.7	14.	4.2

y1(x) = eb(n,x)
y2(x) = varb(n,p)

Bei p = 0,5 ist die Varianz am Größten, für alle anderen Werte nimmt Sie symmetrisch ab.

Für die **Standardabweichung** geben wir folgende Formel ein :

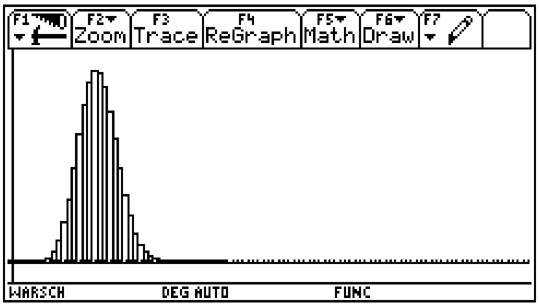
$$\sqrt{var(n,p)} \rightarrow sa(n,p) \quad und$$
$$\sqrt{n.p.(1-p)} \rightarrow sb(n,p)$$

NORMALVERTEILUNG

Bsp : Wie groß ist die Wahrscheinlichkeit, daß beim 1000 maligen Würfeln mindestens 120 und höchstens 155 6er vorkommen?
Wir rechnen *bnpsum(1000,1/6,120,155)*; der Rechner benötigt mehr als 2 Minuten, das Ergebnis laute 0,171892.
Wir untersuchen das gleiche Beispiel mit n = 200 und dem Intervall [30,45]. Das Ergebnis von 0,75 benötigt etwa **20s**.

Für Rechnungen mit großen Zahlen ist die Summenbildung eine sehr zeitaufwendige Methode. Man such daher eine bessere Variante.

Zur Veranschaulichung rechnen wir nochmals für n = 100, I = [10,18]
bnpsum(n,p,10,18) -> 0.675
Zur grafischen Veranschaulichung stellen wir die Verteilung mit *binplot(n,p)* dar.



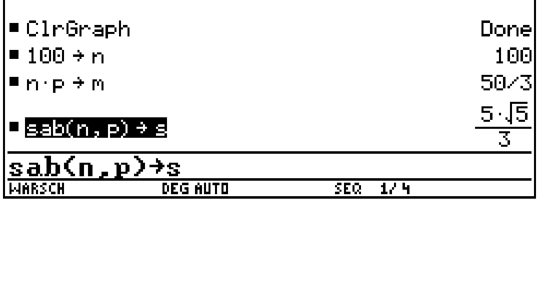
Durch die Balkenbreite 1 läßt sich die Wahrscheinlichkeit als Fläche darstellen, man verwendet die bekannte Formel von Gauss.

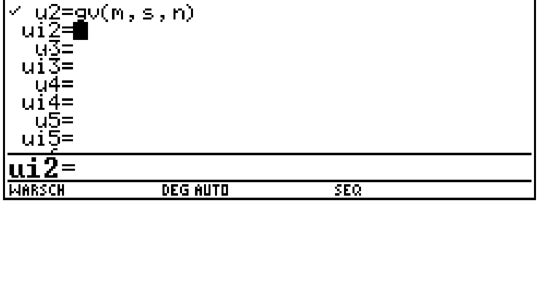
$$W(X = x) = \frac{1}{\sigma.\sqrt{2.\pi}} \; . \; e^{-\frac{1}{2} \cdot \left(\frac{x - my}{\sigma}\right)^2}$$

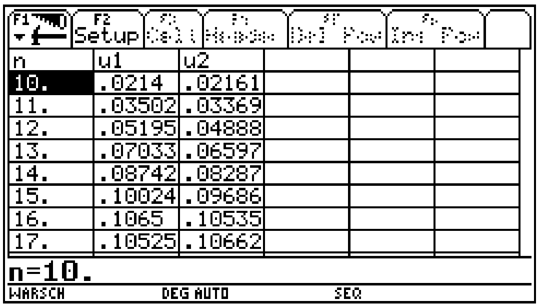
Diese Formel wird in den Rechner eingespeichert und mit ***gv(x,μ,σ)*** (Gauß-Verteilung) bezeichnet.

Wir wollen nun die Güte dieser Formel untersuchen. Dazu schalten wir den Funktionsmodus aus Sequence um und verwenden wieder die Tabelle.

Zuerst legen wir im Home-Screen einige Parameter fest, um die Eingabe später zu vereinfachen. Es sei **n = 100 und p = 1/6.** Dann geben die Funktionen ein, erstellen eine Tabelle und lassen die Funktionen grafisch darstellen.





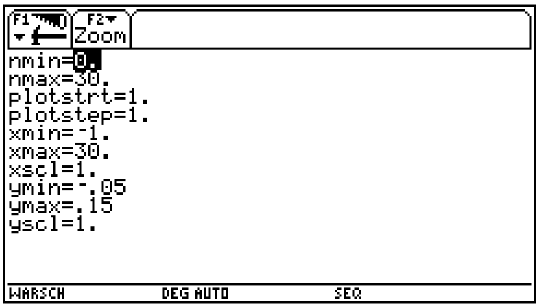


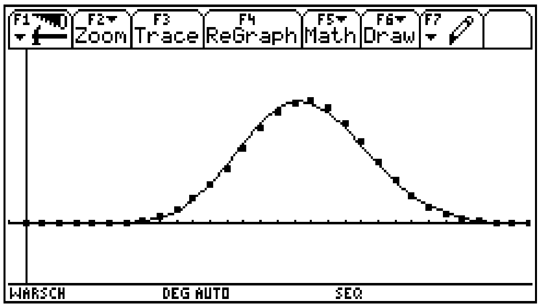
n	u1	u2
10.	.0214	.02161
11.	.03502	.03369
12.	.05195	.04888
13.	.07033	.06597
14.	.08742	.08287
15.	.10024	.09686
16.	.1065	.10535
17.	.10525	.10662

Start : 10
Δtbl : 1

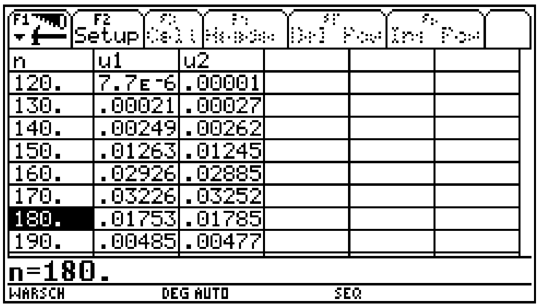
u1 : Line
u2 : Sqare

Dabei stellt die Funktion *u1* die Werte der Binomialverteilung und *u2* die Werte der Normalverteilung dar.

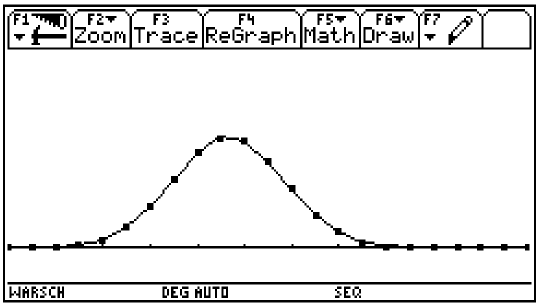




Für **n = 1000** ergibt sich mit dem **Startwert 100 und Δtbl : 10** (Dauert sehr lange !!!!!)



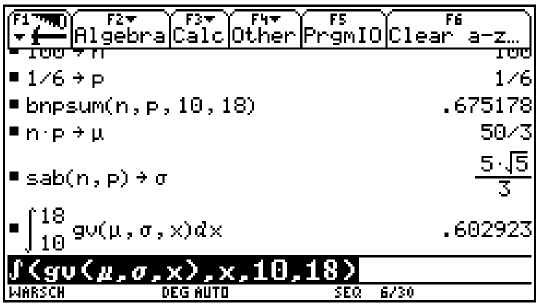
n	u1	u2
120.	7.7E-6	.00001
130.	.00021	.00027
140.	.00249	.00262
150.	.01263	.01245
160.	.02926	.02885
170.	.03226	.03252
180.	.01753	.01785
190.	.00485	.00477



dh. Die Treppenfunktion der Binomialverteilung kann durch eine stetige Funktion angenähert werden. Diese Annäherung ist umso besser, je größer n ist.
Aus der Summe der Einzelwahrscheinlichkeiten wird die Fläche unter der Kurve.

Es sollen nun einige Beispiele berechnet werden.

Bsp : Wie groß ist die Wahrscheinlichkeit, beim 100 maligen Würfeln mindestens 10 und höchstens 18 6er zu erhalten?
Zuerst rechnen wir als Binomialverteilung, dann als Normalverteilung.



Die Differenz der beiden Ergebniswerte läßt sich durch Verwendung der Stetigkeitskorrektur, dh Integration von 9,5 bis 18,5 verkleinern.

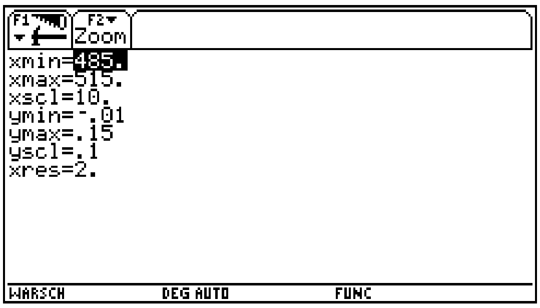
Zur Vereinfachung der Berechnung definiert man die Funktion

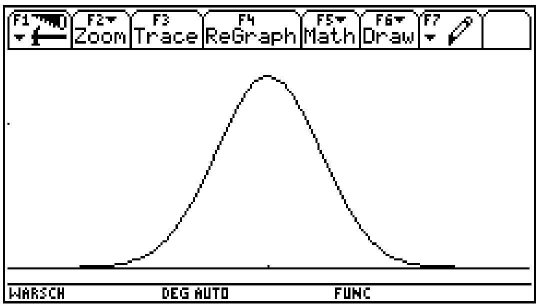
$\int (gv(x,\mu,\sigma),x,a,e)$ -> $gfl(\mu,\sigma,a,e)$ **Gauß-Fläche**

Damit ergibt sich für das obige Beispiel : gfl(μ,σ,10,18) --> 0,6029
bzw gfl(μ,σ,9.5,18.5) --> 0,66

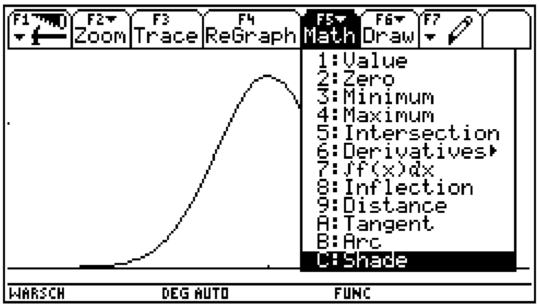
Bsp : Eine Abfülleinrichtung ist auf eine Abfüllmenge von 500g eingestellt. Die Standardabweichung beträgt 3g. Stelle die Verteilungsfunktion grafisch dar. Berechne die W, daß die Abfüllmenge
a) höchstens 506g beträgt
b) nicht unter 497g sinkt.
c) um nicht mehr als 3g vom MW abweicht. Stelle den Bereich grafisch dar !
d) Wie groß ist die Toleranz d zu wählen, damit 90% der abgefüllten Pakete in diesem Bereich liegen.

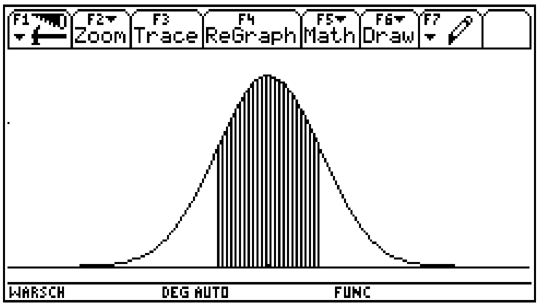
Für die grafische Darstellung definiert man zuerst die *Variablen m und s im Homescreen*. Für *y1* setzt man *gv(x,m,s)* und läßt die Funktion mit den angegebenen Fensterwerten zeichnen.





a) $\mu = 500, \sigma = 3$
gfl(μ, σ, - ∞, 506) = 0,97725 oder gfl(μ, σ, 0, 506) = 0,97725
b) gfl(μ, σ, 497, ∞) = 0,84135 ≙ 84,14%
c) gfl(μ, σ,497, 503) = 0,97725
Für die Grafik wählt man *F5/Shade*,





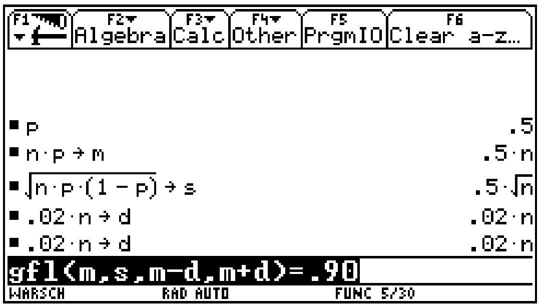
bestätigt *Above X axis* und gibt die untere und obere Grenze ein.

d) Ansatz : gfl(μ, σ, 500 - d, 500 + d)
nsolve(.....= 0.9,d) ergibt für d den Wert d = 4,93g
Diese Rechnung braucht relativ lange (ca 4 min). Eine andere Möglichkeit wäre die Verwendung einer Tabelle und das Herantasten an die Lösung.

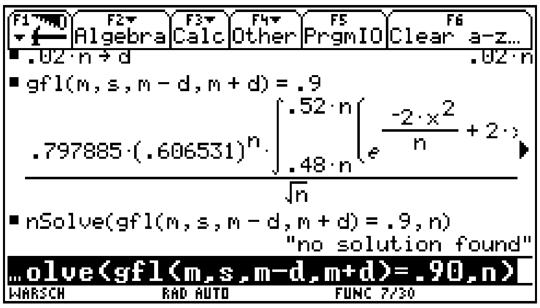
Bsp : Durch eine Stichprobe soll der Anteil p der Ausschußstücke einer Maschine mit 90% iger Sicherheit auf 0,02 genau geschätzt werden. Wie groß ist die Stichprobe zu wählen

n sei die Größe der Stichprobe und p der Anteil der Ausschußstücke.
Es soll dann **W(μ-d ≤ X ≤ μ+d) = 0,90** gelten, wobei d = 0,02*n.

Wir können davon ausgehen, daß gilt p≤ 0,5, und setzen p = 0,5 für den Grenzfall.
Definiert man nun m = n*p und s = √(n*p*(1-p)), und übersetzt man die obige Gleichung in die Sprache des TI-92, so erhalten wir : *gfl(m,s,m-d,m+d) = 0.90*



Bestätigt man die letzte Zeile mit *Enter*, so erhält man einen wenig estetischen Ausdruck. Dies soll uns weiter nich stören, es muß nur nach n gelöst werden, und die gesuchte Stichprobengröße wurde erhalten.

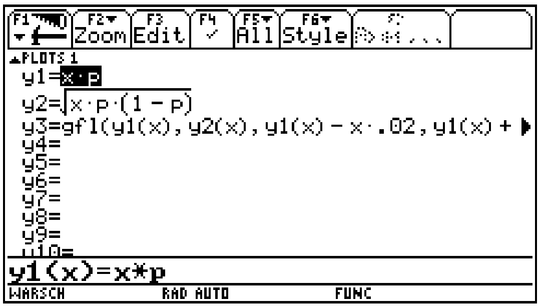


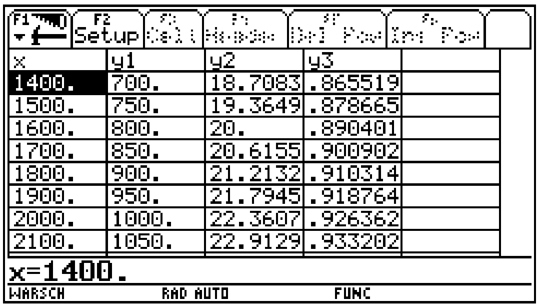
Leider macht uns hier der Rechner einen Strich durch die Rechnung. Wartet man lange genug (einige Minuten) so erhält man die Meldung: *No solution found*.

Diesen Lösungsweg können wir vergessen.

Versuch mit der Tabelle

Nimmt man als x die Stichprobengröße an, und definiert man die notwendigen Parameter sowie die Funktion *gfl* wie angegeben, so erhält an diese Tabelle :





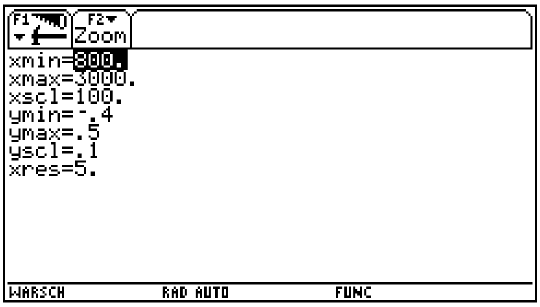
x	y1	y2	y3
1400.	700.	18.7083	.865519
1500.	750.	19.3649	.878665
1600.	800.	20.	.890401
1700.	850.	20.6155	.900902
1800.	900.	21.2132	.910314
1900.	950.	21.7945	.918764
2000.	1000.	22.3607	.926362
2100.	1050.	22.9129	.933202

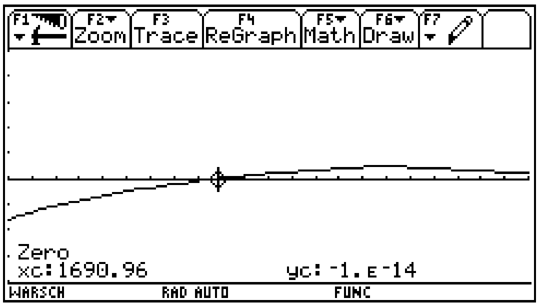
Daraus läßt sich eine **Stichprobengröße von etwa 1700** ablesen.

Führt man dieselbe Aufgabe für p = 0.2 durch, so erhält man eine Stichprobengröße von n = 1100. (Übung)

Grafische Lösung

Wir definieren die Funktion y4(x) = y3(x) - 0,90, und lassen diese Funktion mit geeigneten Fensterparametern darstellen. Von dieser Funktion suchen wir die Nullstellen.



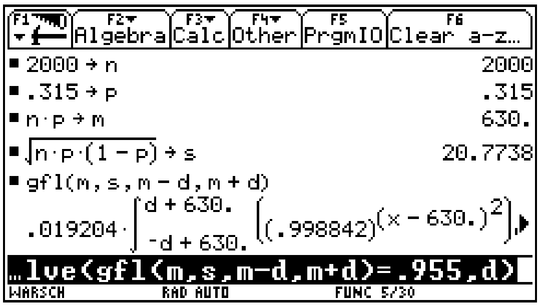


Wir erhalten das “gleiche Ergebnis” wie oben.

Probleme : Der Rechner zeichnet sehr lange. Wenn man den richtigen Bereich nicht erraten kann, dann is diese Variante doch sehr zeitaufwendig.

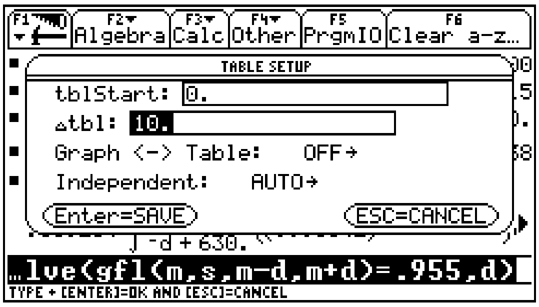
Beispiel :
Ein Meinungsforschungsinstitutbefragt 2000 zufällig ausgewählte Personen und stellt fest, daß 31,5% davon für die Partei A sind. Mit 95,5%-iger Sicherheit soll eine Aussage über den Anteil p der A-Sympatisanten in der wahlberechtigten Bevölkerung gemacht werden.

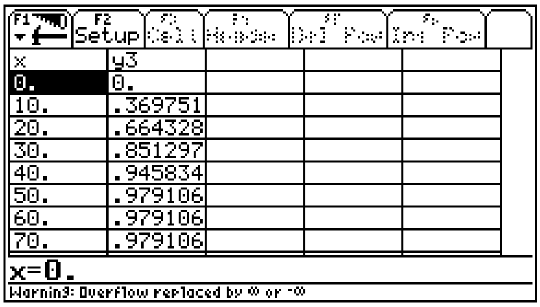
X sei der Anteil der A-Freunde unter den 2000 befragten. Dann ist X binomialverteilt mit den Parametern n=2000 und p = 0,315.



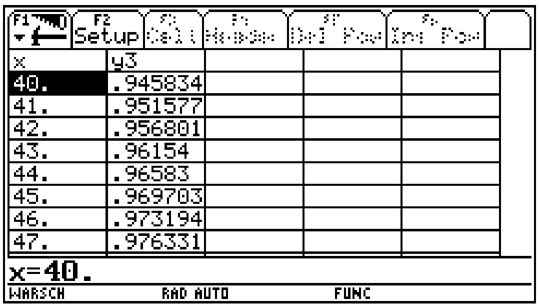
Wir berechnen m und s und jenes symmetrische Interva [m-d,m+d] für das gilt *W(m-d≤ X≤ m+d) = 0,955.*
Die Lösung der Integralgleichung ist sehr aufwendig, innerhalb einer vernünftigen Zeit erhält man kein Ergebnis.

Wir versuchen einen Ansatz über die Tabelle. Da wir nicht wissen, was herauskommtsetzen wir für die Parameter vorerst einmal grobe Werte an und erhalten :





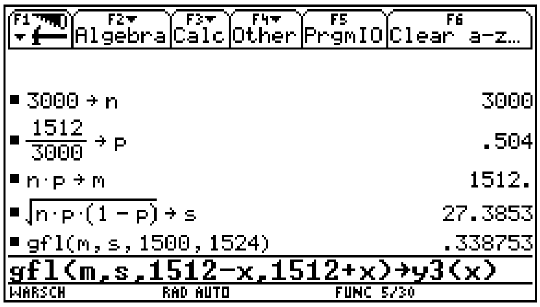
Der gesuchte Wert liegt somit zwischen 40 und 50. Damit läßt sich die Tabelle verfeinern.



Man erkennt : d = 42.

Somit gilt W(630-42≤ X ≤ 630+42) = 0,955 bzw **W(588≤ X ≤ 672) = 0,955**.
Dh **die Wahrscheinlichkeit liegt zwischen 29,4% und 33,6%**.

Beispiel :
Vor einer Bundespräsidentenwahl, bei der nur zwei Personen kandidieren, wird unter 3000 Personen eine Meinungsbefragung durchgeführt, von denen 1512 den Kandidaten A bevorzugen. Es soll mit 99%-iger Sicherheit prognostiziert werden, ob Kandidat A die absolute Mehrheit erhalten wird. Ist diese Befragung für eine solche Prognose ausreichend?



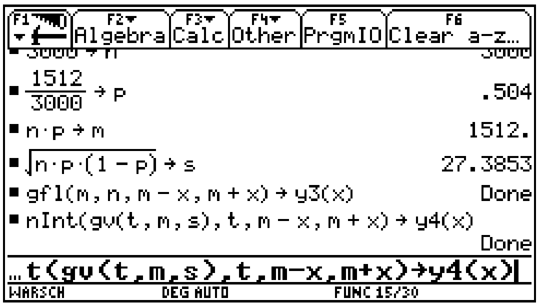
Nach Eingabe der Parameter berechnen wir das Integral :

Man versucht diesen Ansatz :
Dieser Wert ist zu klein; wir verwenden wieder die Tabelle.

Läßt man die Tabelle so wie in obigem Beispiel berechnen, so braucht der Rechner sehr lang. Zusätzlich kommen für x>60 offensichtlich falsche Werte heraus.
Hier dürften die Grenzen des Rechners erreicht worden sein.

Wir versuchen einen anderen Ansatz mit Hilfe der **numerischen Integration**:

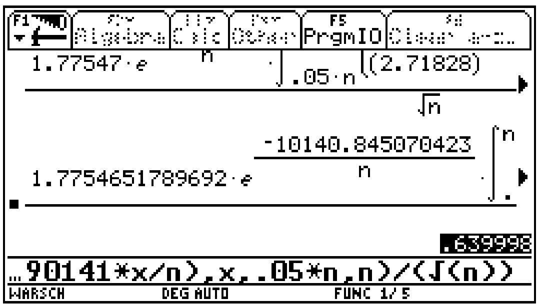
nInt(gv(t,m,s) ,t, m-x, m+x) -> y_4(x)



Mit diesem Ansatz kann die Tabelle erstellt werden. Man lies das Ergebnis 71 ab.

Die Anzahl der A-Wähler liegt somit mit 99%-iger Sicherheit zwischen 1441 und 1583 dh zwischen 48,03% und 52,76%. Es kann mit 99%iger Sicherheit keine absolute Mehrheit vorhergesagt werden.

Beispiel :
In einem Wahlkreis werden 600 Wahlberechtigte hinsichtlich ihres beabsichtigten Wahlverhaltens befragt, von denen sich 32 für die Partei Z entscheiden. Mit welcher Wahrscheinlichkeit kann prognostiziert werden, ob diese Partei die 5% Hürde überspringen wird?
Berchnet man m und s und setzt man für n=600, so erhält man :



Die Partei Z überspringt die 5% Hürde zu 64%.