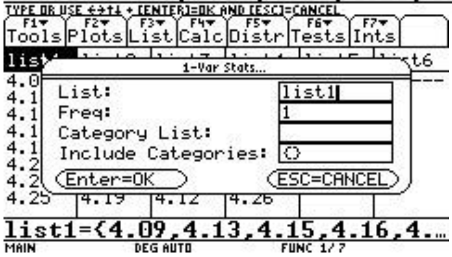
1. Auswertung von Daten

Gegeben sind die folgenden Daten aus dem Weitsprungwettbewerb der Klasse 8a:

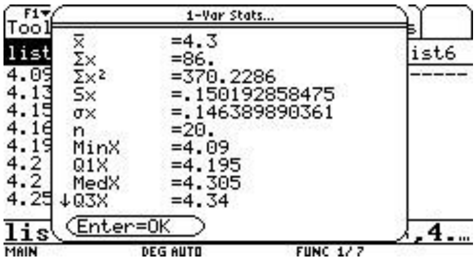
4,32m 4,19m, 4,30m, 4,44m, 4,31m, 4,09m, 4,13m, 4,72m, 4,40m, 4,20m, 4,15m, 4,35m, 4,20m, 4,56m, 4,31m, 4,27m, 4,32m, 4,16m, 4,25m, 4,33m

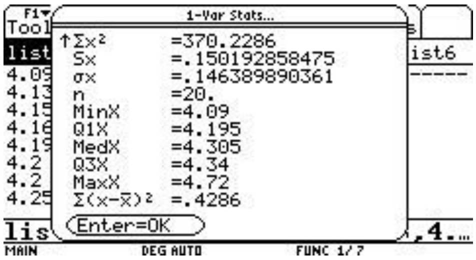
Werten Sie die Daten aus und stellen Sie das Ergebnis graphisch dar.
Ein fertiges Modul findet sich im Stats/List-Editor unter F4 Calc 1: 1-Var Stats. Die Daten werden deshalb in List1 eingetragen.
(Im Data/Matrix Editor gibt es das gleiche Programm unter F5 Calc 1: OneVar. Allerdings müssen die Daten dann als Typ DATA eingetragen werden, nicht als LIST. Deshalb lieber Stats/List-Editor verwenden!).
Im Menü für die 1-Var Statistik ist die Liste anzugeben.



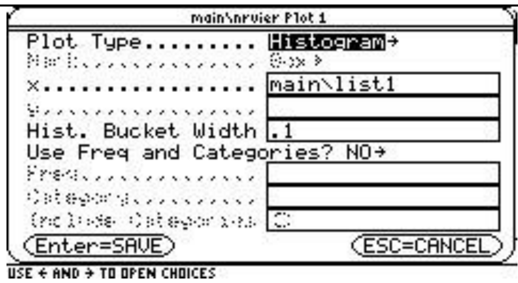
Auf dem Ergebnisbildschirm bedeutet:

- $\bar{x}$ den Mittelwert der x-Werte (entspricht dem Erwartungswert)
- Sx die empirsche Varianz der x-Werte *(nicht zu verwechseln mit der Varianz! Bei der Varianz dividieren wir die Summe der Quadrate der Abweichungen durch n, bei der empirischen Varianz durch n-1.)*
- σx die Standardabweichung der x-Werte
- MinX Minimum der x-Werte
- Q1X erstes Quartil der x-Werte (das kleinste Viertel der Werte geht bis Q1X)
- MedX Median der x-Werte (Median ist der mittlere Wert, z.B. bei 19 Werten der Größe nach geordnet der 10.Wert)
- Q3X drittes Quartil der x-Werte
- MaxX Maximum der x-Werte

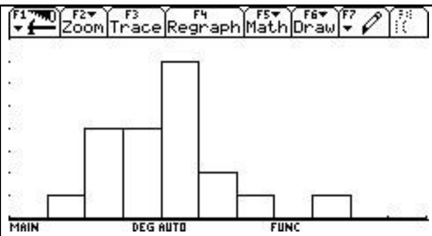




Graphische Darstellung von Daten
a) Histogramm
Mit Histogramm werden Ein-Variablendaten als Histogramm geplottet. Die x-Achse wird in gleichbreite Balken oder Stäbe unterteilt. Die Höhe eines Stabes gibt an, wie viele Datenpunkte in den Bereich des jeweiligen Stabes fallen.
Unter Hist. Bucket Width lässt sich die Stabbreite einstellen. Je größer die Stabbreite, desto weniger Balken entstehen. Bei unter 50

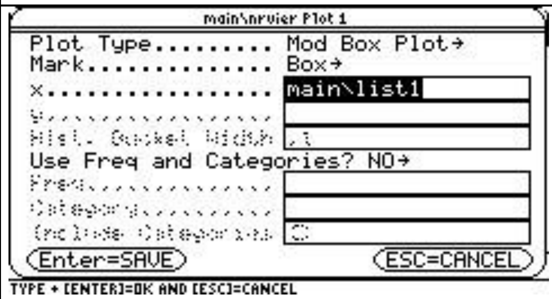


Messungen verwendet man in der Regel 5 bis 7 Balken, bis 100 Messungen 6 bis 10 Balken.

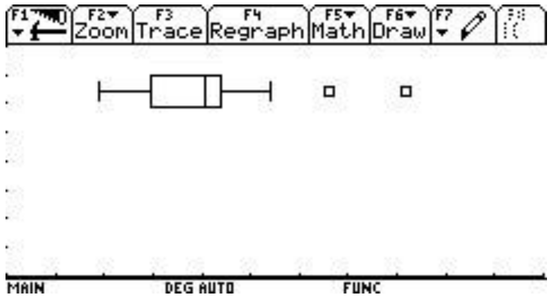


b) Boxplot

Der Boxplot ist ein Diagramm zur graphischen Darstellung einer Reihe numerischer Daten. Er vermittelt einen Eindruck davon, ob die Verteilung symmetrisch oder „schief“ ist. Als „Box“ bezeichnet man das mittlere Rechteck. Es enthält den Bereich, der die mittlere Hälfte der Daten, das 2. und das 3. Quartil, enthält. Der senkrechte Strich in der Box kennzeichnet den Median (s. oben). Die beiden „Whisker“ oder Fühler kennzeichnen den Bereich, in dem das untere Viertel (1. Quartil) bzw. das obere Viertel (4. Quartil) der Daten liegen (ohne die Ausreißer). Werte, die mehr als das 1,5fach der Boxlänge vom Median entfernt sind, werden als „Ausreißer“ bezeichnet und gesondert dargestellt, wenn als Plot Type „Mod Box Plot“ gewählt wurde. In der Darstellungsform „Box Plot“ fallen die Ausreißer weg.



Im Window ist lediglich der x-Bereich einzustellen



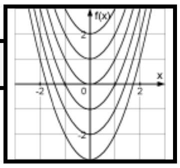
Die Funktion mean berechnet den Mittelwert einer Liste. Dabei kann man die Liste von Hand definieren, indem man die Werte durch Komma getrennt in geschweifte Klammern einschließt. Alternativ lassen sich die Werte im List-Editor eintragen, dann kann auf den Namen der Liste (list1 usw.) zurückgegriffen werden.

Auf mean kann im Hauptbildschirm zugegriffen werden. Im Stat/List-Editor findet sich mean unter F2 List: 3Math: 3 mean.

Die Varianz berechnet man mit VarPop (Varianz der Population). Die Funktion steht im Hauptbildschirm zur Verfügung und kann auf selbst definiert Liste wie auch Liste aus dem Stats/List-Editor angewandt werden.

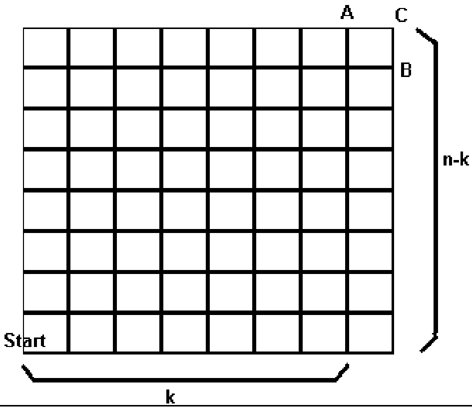
Im Stat/List-Editor findet sich unter F2 List: 3Math: A varPop.

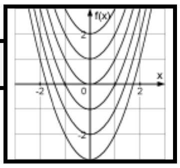
Achtung: Der Befehl variance liefert die Varianz normiert mit n-1 statt mit n!

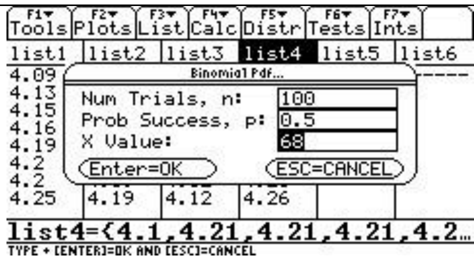
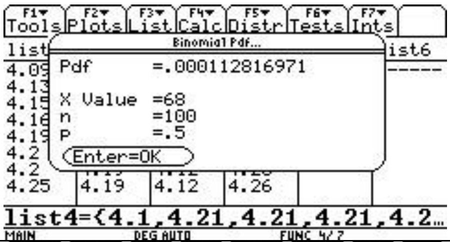
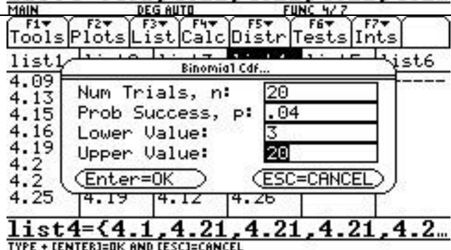
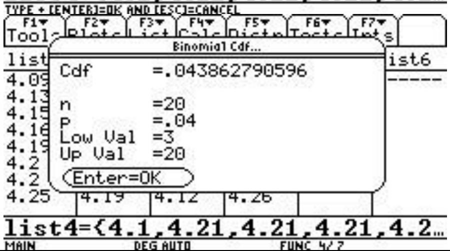
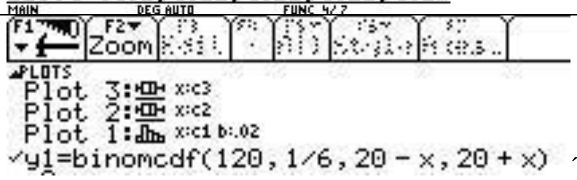


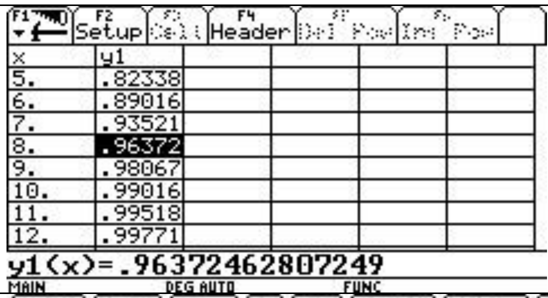
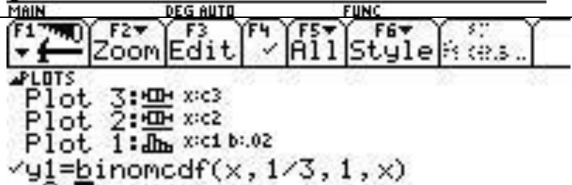
StDevPop liefert die Standardabweichung der Population, normiert mit Wurzel aus n. Auf stDevPop kann im Hauptbildschirm zugegriffen werden. Im Stat/List-Editor findet sich mean unter F2 List: 3Math: 9 stDevPop. Achtung: StdDev liefert die empirische Standardabweichung, normiert mit Wurzel aus n-1!	

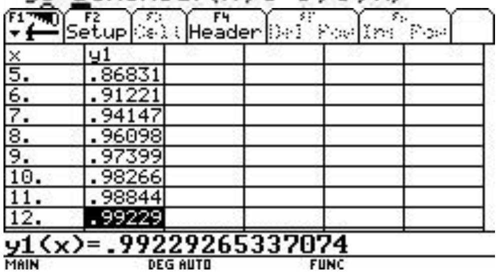
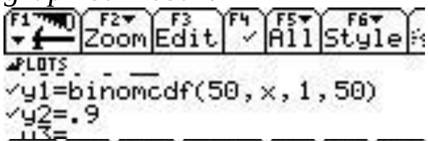
2. Binomialverteilung

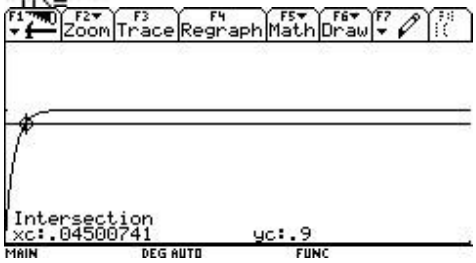
Die Binomialkoeffizienten beschreiben die Anzahl der Möglichkeiten, aus n Elementen k auszuwählen. Mit dem Voyage: $\binom{n}{k}$ = nCr (n, k) Es gibt nur eine Möglichkeit, aus n Elementen 0 Elemente oder n Elemente auszuwählen. Die Binomialkoeffizienten sind symmetrisch. Es spielt keine Rolle, ob man k Elemente auswählt oder k Elemente weglässt.	Für die Binomialkoeffizienten $\binom{n}{k}$ gilt: $\binom{n}{k} = \frac{n!}{k! \cdot (n-k)!}$ $\binom{n}{0} = \binom{n}{n} = 1$ $\binom{n}{k} = \binom{n}{n-k}$ Symmetrie
Die Rekursionsformel kann man sich am besten im Stadtplan von Manhattan klarmachen: Wenn man insgesamt n+1 Blocks (davon k+1 nach Osten) weit von Start nach C geht, dann kommt man auf dem kürzesten Weg nach n Blocks entweder über Punkt A (n Blocks weit von Start, davon k nach Osten) oder über Punkt B (n Blocks weit, davon k+1 nach Osten). Im Pascal'schen Dreieck erhält man einen Koefffizienten, indem man die beiden Zahlen links oben und rechts oben addiert!	Rekursionsformel: $\binom{n}{k} + \binom{n}{k+1} = \binom{n+1}{k+1}$ 
Die abgekürzte Schreibweise binomPdf und binomCdf darf nur verwendet werden, wenn sie entsprechend vereinbart worden ist! Zu Beginn der schriftlichen Arbeit oder der Aufgabe muss also stehen: Die Binomialverteilung findet man im Stats/List-Editor unter F5 DISTR. In den Formeln bedeutet n die Anzahl der Vesuche, p die Erfolgswahrscheinlichkeit, k die Anzahl der Erfolge, u die untere Grenze und o die obere Grenze für die Anzahl der	Im Folgenden gilt: $binompdf(n,p,k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$ $binomcdf(n,p,u,o) = \sum_{i=u}^{o} \binom{n}{i} \cdot p^i \cdot (1-p)^{n-i}$ Die rechte Seite der beiden Formeln stehen in der Formelsammlung S. 44 und 45 oben mit der einzigen Änderung, dass die Summe statt von 0 bis k von u bis o laufen muss.



Erfolge.	
Voraussetzungen für die Anwendung der Binomialverteilung sind: - die Erfolgswahrscheinlichkeit p für jeden einzelnen Bernoulli-Versuch ist konstant - die Anzahl der Wiederholungen n ist konstant, es gibt keinen vorzeitigen Abbruch - gesucht wird nur die Anzahl der Erfolge, nicht bei welchem Versuch ein Erfolg erzielt wird	Gegenindikationen sind: - durch Trainings- oder Ermüdungseffekt oder durch Ziehen ohne Zurücklegen aus kleinen Stichproben schwankt die Erfolgswahrscheinlichkeit - es gibt einen vorzeitigen Abbruch (z.B. „würfle, bis eine 6 kommt“) - es wird ein Erfolg in einem bestimmten Versuch erwartet
Typ 1: Einzelereignis Wie groß ist die Wahrscheinlichkeit, dass bei 100 Münzwürfen genau 68 Wappen auftreten? X Anzahl der Wappen n=100 p = 0,5 k=68 P(X=68) gesucht binomPdf(100,1/2,68)=0,000113 Die Wahrscheinlichkeit für genau 68 Wappen bei 100 Münzwürfen beträgt 0,0113%.	
Typ 2: Bereich von Ereignissen Auf einer Leitung werden 20 Bits übertragen. Wie groß ist die Wahrscheinlichkeit, dass mehr als 2 Bits falsch übertragen werden, wenn die Wahrscheinlichkeit für eine falsche Übertragung bei jedem Bit bei 4% liegt? X Anzahl der fehlerhaften Bits n = 20 p=0,04 k>2 P(X>2) gesucht binomCdf(20,0.04,3,20)=0,0439 Die Wahrscheinlichkeit liegt bei 4,4%.	
Typ 3: k gesucht (Anzahl der Erfolge) Eine Würfel wird 120 mal geworfen. Bestimme einen Bereich um den Erwartungswert herum, in dem mit 95% Sicherheit die Anzahl der Sechsen liegt. X Anzahl der Sechsen n=120 p= 1/6 k gesucht P(μ-k≤X≤μ+k) = 0,95 E(X)=120 * 1/6 = 20 = μ BinomCdf(120,1/6,20-x,20+x) wird x=8 erstmals größer als 0,95. Die Zahl der Sechsen liegt mit 95% Wahr-	

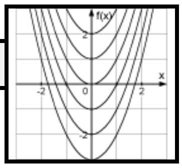
<table>
<tr>
<td>scheinlichkeit zwischen 12 und 28 (jeweils einschließlich).</td>
<td>
</td>
</tr>
<tr>
<td>Typ 4: n gesucht (Zahl der Wiederholungen)
Ein Schütze trifft mit einer Wahrscheinlichkeit von 1/3 das Ziel. Wie oft muss er schießen, damit er mit einer Sicherheit von 99% mindestens einmal das Ziel trifft?
X Anzahl der Treffer
n gesucht   p = 1/3   k≥1
P(X≥1) > 0,99
BinomCdf (x, 1/3, 1, x) wird für x=12 erstmals größer als 0,99.
Der Schütze muss zwölf mal schießen, damit er mit 99% Sicherheit mindestens einmal das Ziel trifft.</td>
<td>

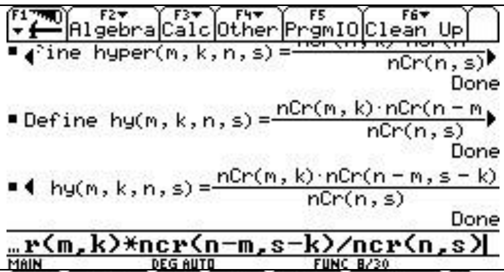
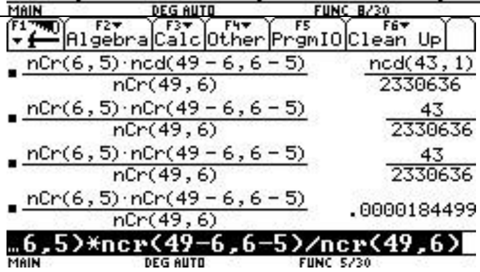

</td>
</tr>
<tr>
<td>Typ 5: p gesucht (Erfolgswahrscheinlichkeit)
p sei die Wahrscheinlichkeit, an einer Infektion zu erkranken. Für welches p ist die Wahrscheinlichkeit, dass unter 50 Versuchspersonen mindestens eine erkrankt ist, mindestens 90%?
X Anzahl Erkrankungen
n = 50   p gesucht   k ≥1   P(X≥1)≥0,9
y1=binomCdf(50,x,1,50)
y2=0.9
intersection (y1,y2,0,0.2) | x=0,04500741
| y = 0,9
Bei einer Ansteckungswahrscheinlichkeit von mindestens 4,5% ist mit 90% Sicherheit unter 50 Versuchspersonen mindestens eine erkrankt.</td>
<td>Hinweis:
Die Gleichung binomCdf(50,x,1,50)=0,9 kann das CAS nicht algebraisch, sondern nur graphisch lösen!



</td>
</tr>
</table>

3. Hypergeometrische Verteilung

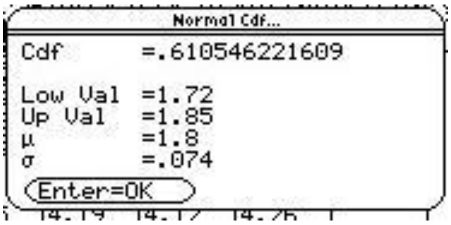
Während man die Binomialverteilung mit dem Modell „Ziehen mit Zurücklegen“ beschreiben kann, steht die hypergeometrische Verteilung für „Ziehen ohne Zurücklegen“, wobei die Urne in beiden Fällen zwei Sorten Kugeln enthält. Das klassische Beispiel für eine hypergeometrische Verteilung ist die Wahrscheinlichkeit für 6, 5, 4, usw. Richtige im Zahlenlotto.	Bei Bedarf kann die hypergeometrische Verteilung definiert werden. Dabei ist zu beachten, dass der Voyage nicht zwischen Groß- und Kleinschreibung unterscheidet. Für n wird deshalb s (Stichprobenumfang) eingesetzt:



Die Formel für die hypergeometrische Verteilung findet man in der blauen Formelsammlung auf Seite 42.	
Wie groß ist die Wahrscheinlichkeit für 5 Richtige bei Lotto „6 aus 49“? M = 6 (Anzahl Richtige), k = 5 (Richtige in der Stichprobe), N = 49 (Anzahl Kugeln), n = 6 (Stichprobenumfang) Die Wahrscheinlichkeit für 5 Richtige beträgt 0,001845%.	
Für große n nähert sich die Hypergeometrische Verteilung der Binomialverteilung an, weil bei sehr vielen Kugeln sich die Erfolgswahrscheinlichkeit beim Ziehen einzelner Kugeln kaum noch ändert.	

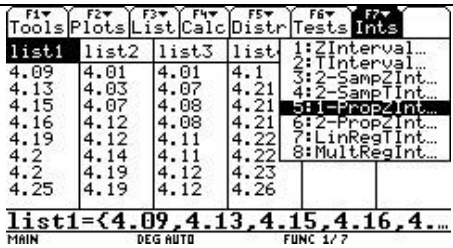
4. Normalverteilung

Eine Funktion f heißt Dichtefunktion einer stetigen Zufallsgröße X, wenn nebenstehende Eigenschaften erfüllt sind: Eine Zufallsgröße und deren Verteilung heißen stetig, falls es eine geeignete Dichtefunktion mit den Eigenschaften (1) bis (3) gibt.	(1) Für alle $x \in \Re$ gilt: $f(x) \geq 0$ (2) $P(a \leq X \leq b) = \int_a^b f(x)dx$ (3) $\int_{-\infty}^{+\infty} f(x)dx = 1$
Eine Zufallsgröße X ist normalverteilt mit dem Erwartungswert μ und der Standardabweichung σ, wenn $f(x) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{-\frac{1}{2}\cdot\left(\frac{x-\mu}{\sigma}\right)^2}$ Dichtefunktion ist. Für normalverteilte Zufallsgrößen gilt: $P(a \leq X \leq b) = \int_a^b \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{-\frac{1}{2}\cdot\left(\frac{x-\mu}{\sigma}\right)^2} dx$	Zur Berechnung der Normalverteilung steht im Stats/List-Editor der Menüpunkt NormalCdf zur Verfügung: 

Die abgekürzte Schreibweise **normcdf(U,O,μ,σ)** darf nur verwendet werden, wenn zu Beginn der Arbeit/Aufgabe folgendes vereinbart wurde: Die Funktion unter dem Integral finden Sie in der Formelsammlung S. 42, vorletzter Block.	Im Folgenden gilt: $normcdf(U,O,\mu,\sigma)=\int_U^O \frac{1}{\sigma\cdot\sqrt{2\pi}}\cdot e^{-\frac{1}{2}\cdot\left(\frac{x-\mu}{\sigma}\right)^2}\,dx$
Beim Mikrozensus 1999 ergab sich für die Körpergröße von 18-20jährigen Männern ein Mittelwert von 1,80 m bei einer Standardabweichung von 7,4 cm. Die Körpergröße ist annähernd normalverteilt. Mit welcher Wahrscheinlichkeit ist ein zufällig ausgewählter Mann dieser Altersgruppe zwischen 1,72 und 1,85 m groß? Die Wahrscheinlichkeit beträgt 61,05%. Im Hauptbildschirm lautet die Syntax Normcdf(untere Grenze, obere Grenze, Erwartungswert, Standardabweichung). Wird die Reihenfolge der Parameter vertauscht, so erhält man die Fehlermeldung „Domain Error“.	 
In Tabellenwerken (auch im „Großen Tafelwerk“) ist häufig nur eine Tabelle für die Standardnormalverteilung vorgegeben. In der Vor-CAS-Zeit musste deshalb zur Lösung eine gegebenen Normalverteilung standardisiert werden mit Hilfe der Formal $P(a\le X\le b)=\Phi\left(\frac{b-\mu}{\sigma}\right)$ - $\Phi\left(\frac{a-\mu}{\sigma}\right)$	Wird eine Binomialverteilung durch eine Normalverteilung angenähert, so erhält man eine bessere Näherung, wenn man bei der unteren Grenze 0,5 subtrahiert und bei der oberen Grenze 0,5 addiert. ACHTUNG: Nur im Fall der Näherung darf ein Korrekturfaktor eingebaut werden! Sonst ist immer von a bis b zu integrieren, d.h. die in der Aufgabenstellung vorgegebenen Grenzen sind zu übernehmen!
Bestimmung von μ und σ bei symmetrischen Randbereichen mit Rechner graphisch	

5. Konfidenzintervalle

Bei der Bestimmung eines Konfidenzintervalls schätzen wir die Erfolgswahrscheinlichkeit p, die einem Zufallsversuch zu Grunde liegt. Dabei schließen wir von einem Stichprobenergebnis X auf diejenigen Werte von p, mit denen das Stichprobenergebnis vereinbar ist. Soweit nicht anders vereinbar, beträgt die Sicherheitswahrscheinlichkeit 95%. Wir beantworten also die Frage: Für welche Werte von p liegt das Stichprobenergebnis X in der 1,96-σ-Umgebung von	Für binomial verteilte Zufallsgrößen mit μ=np und $\sigma=\sqrt{n\cdot p\cdot(1-p)}$ gilt im Extremfall: $n\cdot p-\sqrt{n\cdot p\cdot(1-p)}=X$ am unteren bzw. $n\cdot p+\sqrt{n\cdot p\cdot(1-p)}=X$ am oberen Rand. Zusammengefasst und umgestellt ergibt sich die Gleichung $1{,}96\cdot\sqrt{n\cdot p\cdot(1-p)}=\lvert X-n\cdot p\rvert$

μ?	
120 Schüler wurden befragt, davon spielen 36 DotA. Wie groß ist mit 95% Sicherheit der Anteil der DotA-Spieler? Die Lösung kann mit Hilfe der o.a. Gleichung berechnet werden.. Der Anteil der DotA-Spieler liegt mit 95% Sicherheit zwischen 21,8 und 38,2%.	
Die im Stats/List-Editor unter F7 Ints 5:1-PropZInt implementierte Funktion ist – anders als im Handbuch angegeben! - nicht geeignet, unbekannte Wahrscheinlichkeiten zu schätzen. Hierbei wird die Wahrscheinlichkeit p nicht als Variable eingesetzt, sondern durch X/n abgeschätzt. Das folgende Beispiel mit einer sehr kleinen Wahrscheinlichkeit zeigt den Fehler: In einer Stichprobe von n=500 tritt kein Mal das Ereignis auf. 1-PropZInt liefert mit X=0 und n=500 das Konfidenzintervall [0;0]. $solve\left(1{,}96\cdot\sqrt{500\cdot p\cdot(1-p)}=\lvert 0-500p\rvert, p\right)$ liefert das richtige Ergebnis. Die Wahrscheinlichkeit für das Ergebnis liegt also mit 95% Sicherheit zwischen 0% und 0,76%.