PERSONAL*notes*

IMST^3: Testen von e-Learning Sequenzen zur Normalverteilung

1 *Projekttagebuch, 5bk, November 2004 - Feber 2005*

Print Open / Close

Das Ziel dieses Projekttagebuches ist es, uns, der Klasse, und dem Leser die wichtigsten Themen vom Mathematikunterricht zu protokollieren.

Das Projekttagebuch wird in der Stunde geführt, daher Verzeihung, wenn die Formulierung nicht exquisit ist.

Miriam Dick, 5bk, HAK Grazbachgasse, 19.11.2004. Dick.Miriam@hak-graz.at

1) Warum Normalverteilung? Forellenbeispiel 19.11.2004

Open / Close

Software: Normal Distribution Notebook und NormalDistribution Palette von *M@th Desktop*:

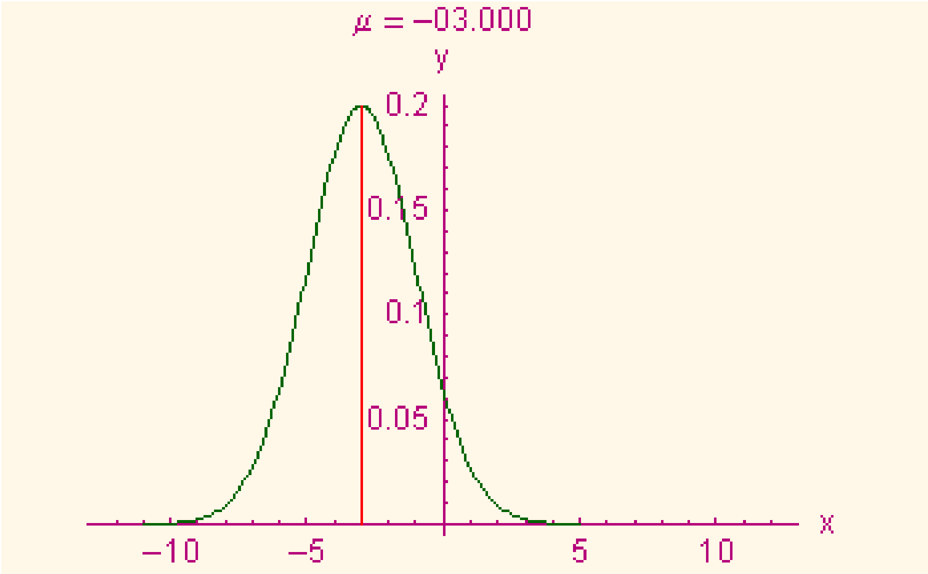
Wiederholung zur Binomialverteilung: Was ist ...μ, σ? Bedeutung.

Heute Einführung der Normalverteilung, und wir werden ca. 3 - 4 Stunden für dieses Thema brauchen! Mathematiker unterscheidet zwischen diskreter und stetiger Verteilung.

Im SÜ-Heft:
Die Binomialverteilung ist eine diskrete Verteilung,zB 4,5,6 Fragen in einem Test anhaken.Die Normalverteilung ist eine stetige Verteilung,zB die Körpergröße kann jede Zahl annehmen.Bei jeder Verteilung braucht man Mittelwert und Abweichung (μ,σ).Sigma ist ein Streuungsmaß.

Am PC:
Normalverteilung-Movies zu besserem Verständnis.
Bedeutung von μ und σ erklärt.
Welchen Einfluss hat μ in der Normalverteilung?

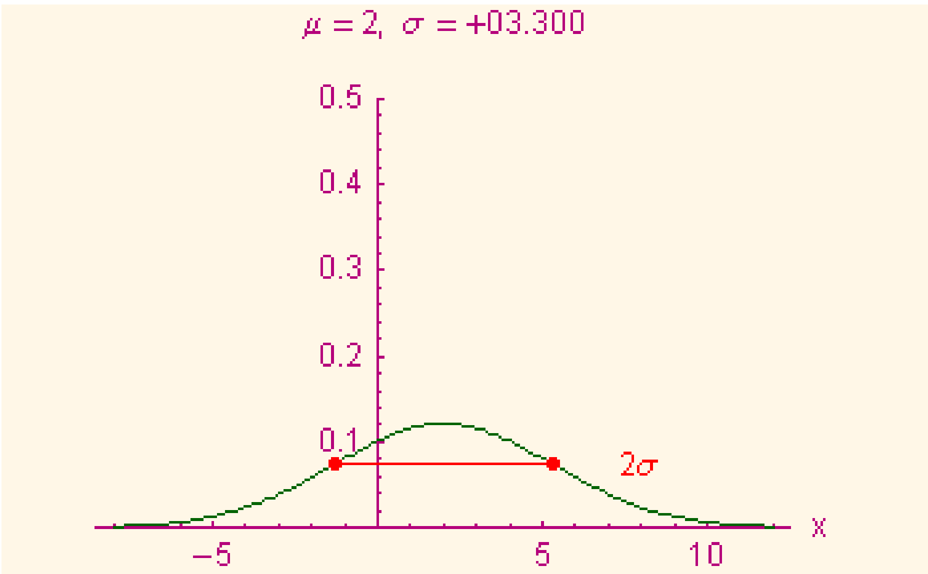
Mittelwert μ verschiebt die Verteilung.



Am PC:

Welchen Einfluss hat σ in der Normalverteilung?

Je kleiner σ ist, umso spitzer ist die Verteilung, automatisch auch höher!
Wenn σ groß ist (sehr breit), legt sie sich "schlafen".
Die Distanz zwischen zwei bestimmten Punkten in der Normalverteilung, beträgt immer 2σ, unabhängig von der Form.



Im SÜ-Heft:
Die Dichtefunktion von der standardisierten Normalverteilung lautet:

$$f[x] = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

Gaußsche Glockenkurve (Bellkurve)
Dichtefunktion (PDF=Probability Density Function)
Wir zeichnen die Kurve!

Wir rechnen die Wertetabelle mit der Hand aus, damit wir ein Gefühl für die Formel und die Kurve bekommen.
Bei -3 kommt das Gleiche heraus wie bei +3, weil durch das Quadrat die Funktionswerte assymetrisch werden. (links und rechts gleich groß).

Das bestimmte Integral bedeutet Fläche.(Fläche der Kurve=1)
Bestimmt man jedoch das Z,muss man das Z durch das "Unendlich" ersetzen!

$$P(z_1 \le Z \le z_2) = \frac{1}{\sqrt{2\pi}} \int_{z_1}^{z_2} e^{-\frac{x^2}{2}} \, dx = \phi[z_2] - \phi[z_1]$$

Das Integral ist tabelliert,oder mit dem PC ausrechenbar!!!
Werte von ϕ (z) erhalten (Tabelle, kopierter Zettel vom Herrn Professor)
ϕ (1,5)=0.93319
ϕ (3,2)=0.99874
ϕ (20)=0.9999999999
ϕ (20)=0.

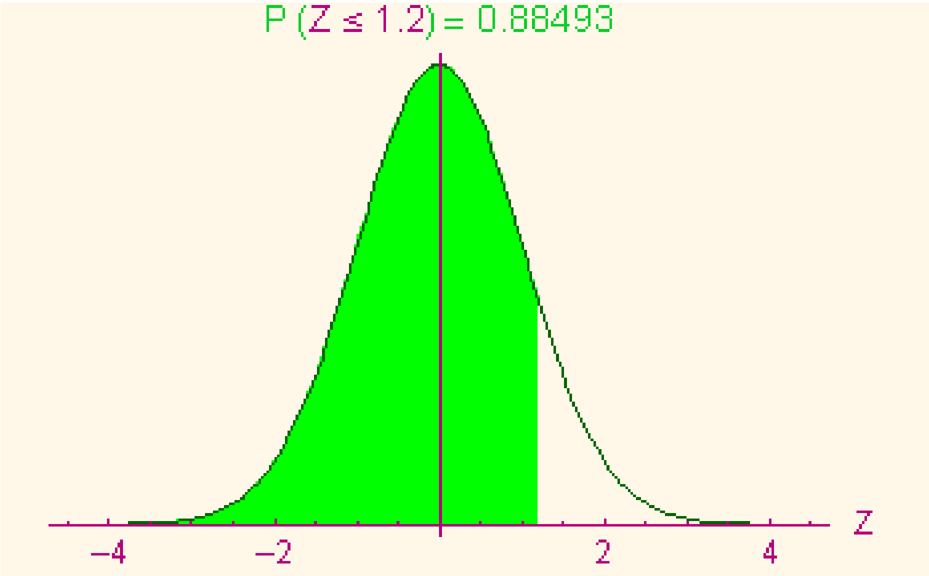
Am PC:
Wir arbeiten mit dem ϕ[z] button und P(x1 ≤ X ≤ x2) button von der Normal Distribution Basics Palette.

Wir berechnen und zeichnen mit P(x1≤X≤x2) die ϕ[z]-Werte!
Die Fläche von -∞ und -2, ist dieselbe Fläche von ∞ bis 2!

ϕ[-2] = 1 - ϕ[2] --> sehr wichtige Formel!

Diesesmal keine Hausübung!!!

Input ▹
```
z = 1.2;
probability = MDSϕ[z];

MDSNormalDistributionPμσ[ Z ≤ z,  {0, 1}]
(* P ( Z ≤ z ) = ? *)
```



Normal Distribution

z (−∞)	z (1.2)	P (Z ≤ 1.2)
−∞	1.2	0.88493
Dev (z σ)		
0. + 1.2		

2) φ[z] Formeln, Forellenbeispiel, Würfelbeispiele 26.11.2004

Open / Close

Software: Normal Distribution Notebook und NormalDistribution Palette von *M@th Desktop*:

Wiederholen der Dichtefunktion!

$$f[x] = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

Wann ist eine Normalverteilung standardisiert?
Wenn μ =0 und σ =1
Wenn σ groß ist, ist die Verteilung breit!

Wir wiederholen ϕ[z]!

$$P(z_1 \le Z \le z_2) = \frac{1}{\sqrt{2\pi}} \int_{z_1}^{z_2} e^{-\frac{x^2}{2}} dx$$
$$= \phi[z_2] - \phi[z_1]$$

Beispiel: händische Überlegung wie groß ϕ[-1] ist!
gleich groß wie von ϕ[1], weil es quadratisch ist.

Jede Exponentialfunktion kann nur positiv oder 0 sein im Unendlich!!!

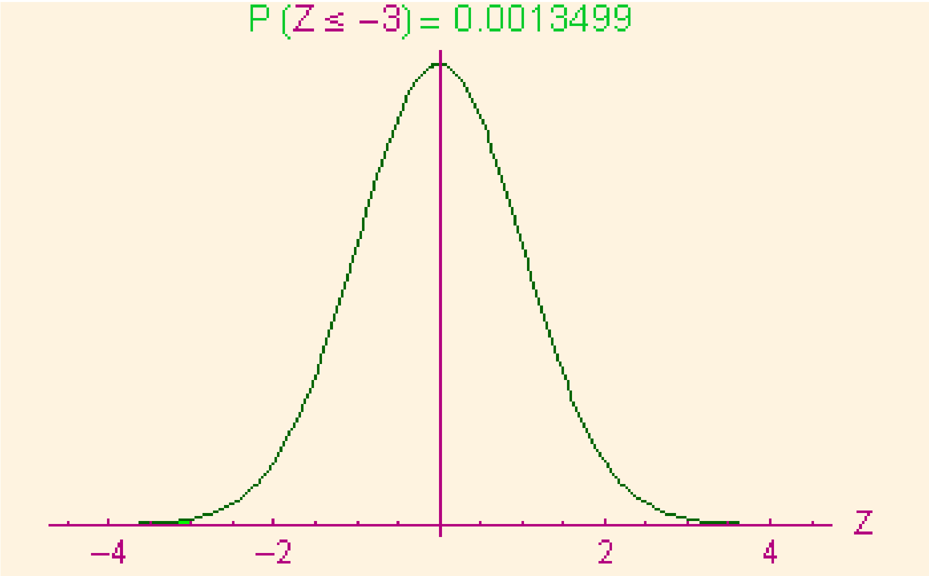
Werte zwischen -99 und -∞ sind praktisch unwahrscheinlich!!!
Alles spielt sich in der Gegend vom Mittelwert ab (99.7 %).
Exponentialfunktion ist sehr stark und drückt Wahrscheinlichkeit sehr weit hinunter.

Am PC: Wir erstellen ein Notebook mit dem Titel "Z-Wert und Z-Formeln"

Im SÜ-Heft:
Berechne ϕ[2.2] mit der Tabelle und am PC!
Berechne ϕ[-1.5] mit der Tabelle und am PC1
Berechne ϕ[-3] mit der Tabelle und am PC!

Unsere Antwort bei ϕ[-3]:

Input ▹
```
z = -3;
probability = MDSϕ[z];

MDSNormalDistributionPμσ[ Z ≤ z,  {0, 1}]
(* P ( Z ≤ z ) = ? *)
```



Normal Distribution

z (−∞)	z (−3.)	P (Z ≤ −3.)
−∞	−3.	0.0013499
Dev (z σ)		
0. −3.		

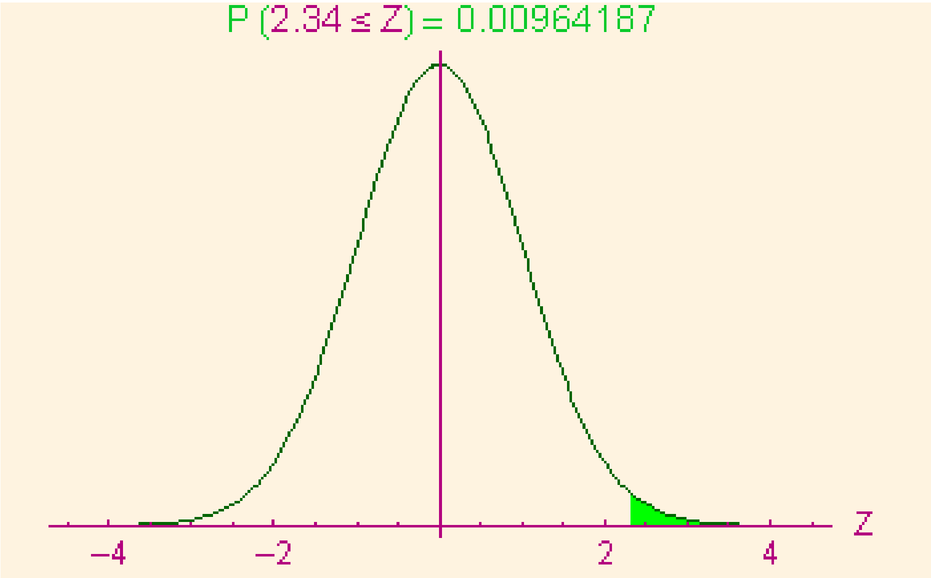
Antwort: Das Integral geht von -∞ bis ca. -3!
Die Wahrscheinlichkeit ist bereits sehr gering
Die Wahrscheinlichkeit, dass die Zufallsvariable Z Werte zwischen -∞ und -3 annimmt, beträgt 0.00135!!!

Am PC: Z-Beispiele und Z-Formeln mit der Hand!
Wir rechnen von den Examples 3.1!

Im SÜ-Heft:
P(Z ≤ 1.2) is equivalent to ϕ[1.2]. It gives the probability that the random variable Z lies in the interval -∞ to 1.2

Im SÜ-Heft:
P(2.34 ≤ Z) ist gleich 1- ϕ[2.34] = 0.00964

Input ▹
```
Switch;
z = 2.34;
probability = 1 - MDSϕ[z];

MDSNormalDistributionPμσ[ z ≤ Z,  {0, 1}]
(* P ( z ≤ Z ) = ? *)
```



Normal Distribution

z (2.34)	z (+∞)	P (2.34 ≤ Z)
2.34	+∞	0.00964187
Dev (z σ)		
0. + 2.34		

Antwort: Die W, dass die Zufallsvariable Z Werte ab 2.34 annimmt beträgt knapp 1%.

Im SÜ-Heft
P(-1.3≤Z≤1.3)

Wir erstellen eine Formelsammlung im SÜ-Heft!
Formelsammlung mit Zeichnung erklärt.

$$P(Z \le z) = \phi[z]$$

$$P(z \le Z) = 1 - \phi[z]$$

$$P(z \le Z \le z) = 2\,\phi[z] - 1$$

$$P(z1 \le Z \le z2) = \phi[z2] - \phi[z1]$$

Im SÜ-Heft

P(|Z| ≥0.54)

Input ▷
```
2 - 2 MDSϕ[0.54, {0, 1}]
```

```
0.589197
```

Antwort: Mit dem MDSϕ-Button bekommt man zwar das richtige Ergebnis, jedoch muss man beim Visualisieren 1-(Ergebnis) rechnen - dann gleiches Ergebnis!

Beispiel f mit der Hand und am PC ausgerechnet!

3) Trouts-Beispiel, Standardisierung, Forellenbeispiel 3.12.2004

Open / Close

Allgemeines:

Aufgabe und vorhande Mappe kontrolliert
kleine Überprüfung von der Normalverteilung + theoretische Fragen
weitere Benotungsart wird eine kleine Präsentation unseres Projektes sein (Aufstellen von Statistiken, Erkennen der Qualität von Daten etc.)

Am PC:

Wir gehen von den Examples 3.3 Trouts aus. X = Zufallsvariable, die die Forellenlänge angibt

Trout Example 3 year old trouts in a basin have the length of $\mu = 45.2$ cm and $\sigma = 3$ cm.

(a) What is the probability that a randomly caught trout is smaller than 46 cm, $P(X \leq 46) = ?$

(b) What is the probability that a randomly caught trout is longer than 44.5 cm, $P(44.5 \leq X) = ?$

(c) What is the probability that a randomly caught trout has a length between 44.2 cm and 46.2 cm, $P(44.2 \leq X \leq 46.2) = ?$ (symmetrical interval)

(d) What is the probability that a randomly caught trout has a length between 45 cm and 47 cm, $P(45 \leq X \leq 47) = ?$

Notiz: Der Herr Professor erklärt uns mit welcher Formel man von X auf z komm und was standardisieren heißt!

Formel:

$$\mathbf{X} = \mu + \mathbf{z}\sigma$$

Im SÜ-Heft:

Wir errechnen zuerst z, und dann die Wahrscheinlichkeit von dem Trouts-Beispiel!

Es gibt auch einen einfacheren Weg, die Wahrscheinlichkeit auszurechnen - der PC macht es automatisch mit der Formel P(X ≤ x).

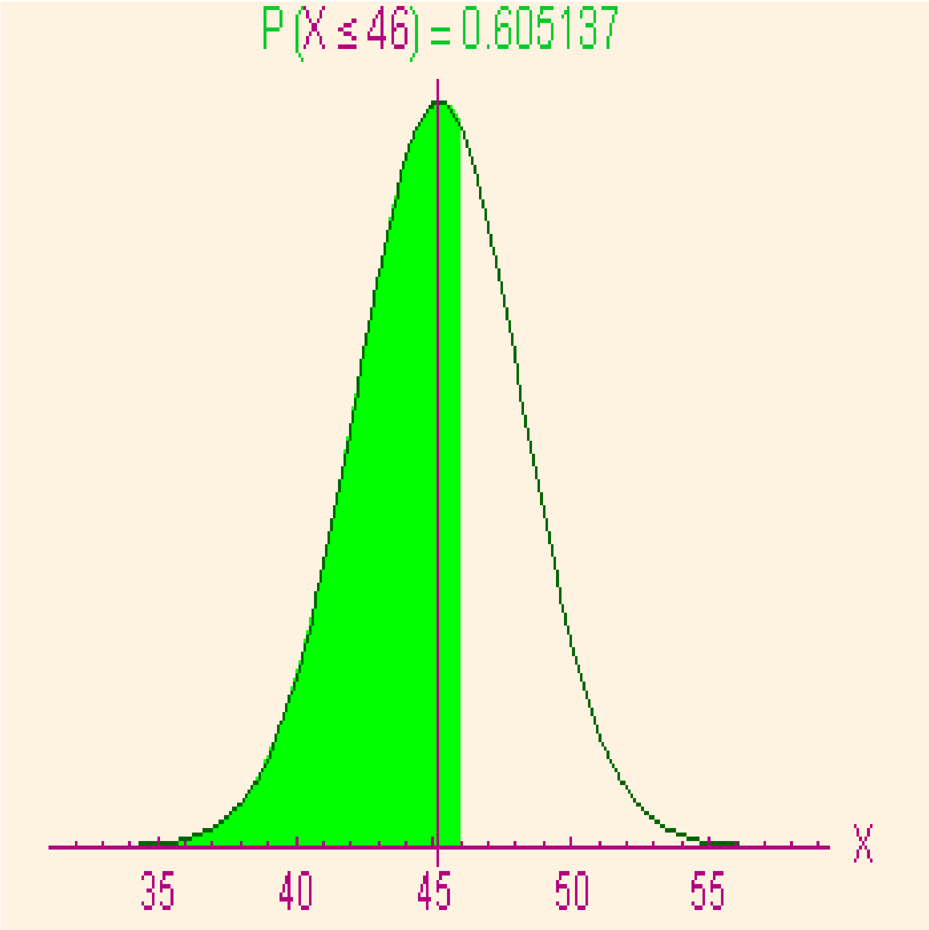
Warum haben wir dann diese Formel gelernt?
Für jede Aufgabe gibt es ein eigenes Integral, deswegen "Trick" mit der z-Tabelle und Formel zum Umrechnen von X auf z!
--> es ist eine rein historische Sache, heute braucht es keiner mehr, aber es dient zum Vereinfachen!

Warum standardisieren?
Mit Hilfe der Standardisierung wird das Beispiel auf die 0,1-Verteilung (Standard-Normalverteilung) reduziert und mit der z-Tabelle ausgewertet.
--> Heutzutage nicht mehr notwendig, weil man jede μ ,σ -Aufgabe am PC ausrechnen kann.

Wir errechnen einen Wert von z=0.2666
Anhand mit der Formel für die Wahrscheinlichkeit errechnen wir diese aus!

Am PC:

Input ▹
```
Switch ;
μ = 45.2 ; σ = 3 ; x = 46 ;
probability = MDSϕ[x, {μ, σ}];

MDSNormalDistributionPμσ[ X ≤ x ,  {μ, σ}]
(* P ( X ≤ value ) = ? *)
```



Normal Distribution

z (−∞)	z (46.)	P (X ≤ 46.)
−∞	0.266667	0.605137
Dev (z σ)		
45.2 ± 0.8		

Antwort: Die Wahrscheinlichkeit, dass die zufällig gefischte Forelle ≤ 46 cm ist, beträgt ca. 61 %.

Notiz: Das bestimmte Integral berechnet auch die Wahrscheinlichkeiten mit einer negativen Länge von -8 bis 46 cm.

Notiz: Die Normalverteilung ist jedoch nur ein Modell um ein Wahrscheinlichkeiten-Movie zu beschreiben - der ganze wichtige Wert liegt in der Sigmaumgebung.

Die weiteren Beispiele haben wir immer mit den gleichen Schritten gerechnet:
1. Umrechnung von X auf z mit der kennen gelernten Formel
2. ϕ[z] aus unserer Tabelle heraussuchen
2. Überlegen welche Wahrscheinlichkeitsformel anzuwenden ist

3. Ausrechnen der Wahrscheinlichkeit

Erklärung anhand des Beispiels TroutExample (c)
Wir haben μ=45.2 , σ=3 und X=46.2 gegeben! Es wird gefragt, wie hoch die W ist, dass die zufällig gefischte Forelle eine Größe zwischen 44,2 und 46,2 cm hat.
Wir rechnen den X-Wert in den z-Wert um mit Hilfe unserer Formel.
Ergebnis z = 0.333
Wir schauen in der Tabelle nach: ϕ[0.333] ist gleich z=0.63
Mit Hilfe der Formel 2ϕ[z] - 1 errechnen wir P (die Wahrscheinlichkeit)
Wir bekommen ein Ergebnis von 26 % heraus!

Notiz: Man muss immer gut überlegen, welche Formel man für die Wahrscheinlichkeit wählt, da der PC die Wahrscheinlichkeit sofort mit dem gegebenen X-Wert ausrechnet, wir die W aber mit dem z-Wert errechnent.

Notiz: Bei dem letzten Beispiel tauchte zum ersten Mal die Formel 2 ϕ[z]-1 auf!
Warum?
Weil der Abstand von den beiden Anhaltsgrößen (44,2 und 46,2 cm) zum Mittelwert (45,2) jeweils der Gleiche ist.

Im SÜ-Heft:

Wir lernen auch das Reverse-Beispiel kennen - die umgekehrte Formel!
Man hat die Wahrscheinlichkeit gegeben und muss den X-Wert errechnen.

Gleich wie bei den obigen Beispielen, errechnet der Computer X auf einmal.
Mit der Hand errechnen wir den z-Wert und müssen diesen mit der Standardisierungsformel (siehe oben) umrechnen.

Im SÜ-Heft und am PC:

Reverse Trout Examples 3 year old trouts in a basin have the length of $\mu = 45.2$ cm and $\sigma = 3$ cm.

(a) 33 % of small trouts in the basin must be put in a special basin with more food and oxygen in water.
What is the max length x of a small trout? $P(X \le x) = 0.33, \quad x = ?$

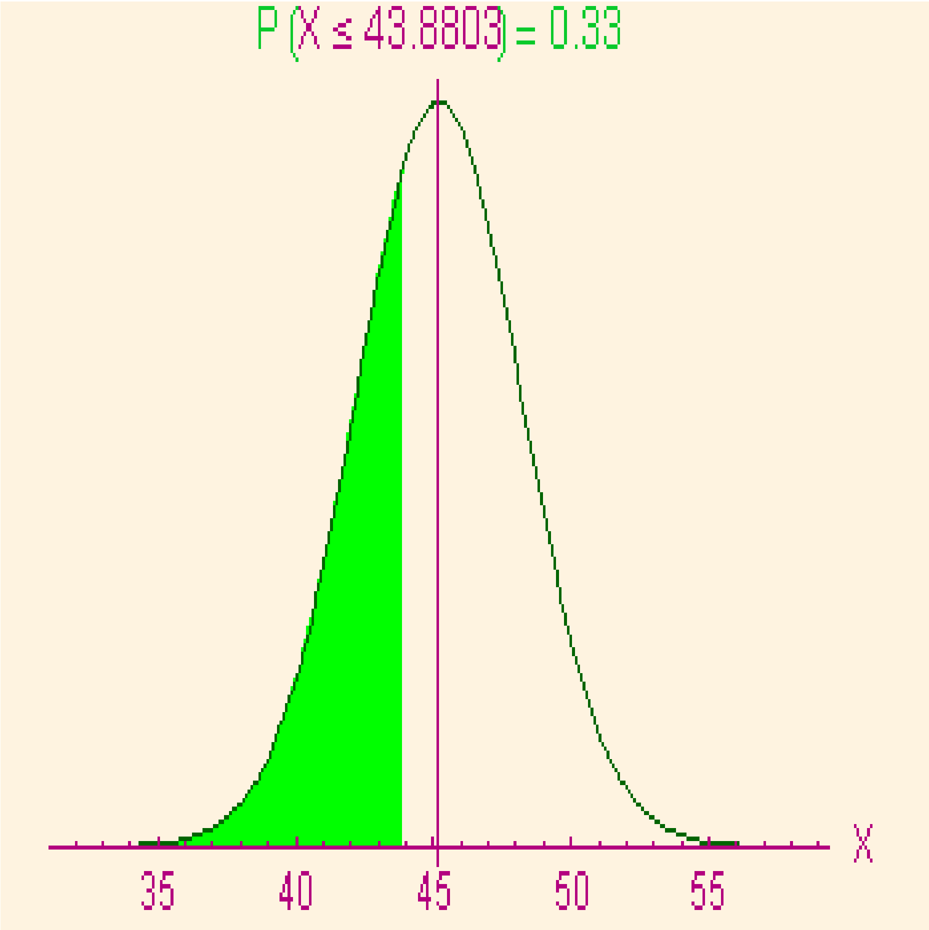
(b) 40% of the largest trouts in the basin are to be sold to a restaurant.
What is the minimum length x of a trout, $P(x \le X) = 0.40, \quad x = ?$

(c) 90 % of the trouts in the basin are how long?
$P(\mu - x \le X \le \mu + x) = 0.90, \quad \mu - x, \ \mu + x = ?$

Notiz: Das wichtigste an dieser Rechnung ist, dass man zwar die gleichen Formeln verwendet aber vor dem Ausrechnen den SWITCH-Button drückt.

a) 33 % von den gefischten Forellen haben eine Länge von???

Input ▹
```
Switch;
μ = 45.2; σ = 3;
probability = 0.33;
Clear[z];
 z = z /. FindRoot[MDSϕ[μ + z σ, {μ, σ}] == probability, {z, 0}];

MDSNormalDistributionPμσ[ X ≤ μ + z σ, {μ, σ}]
(* P ( X ≤ ? ) = probab *)
```



Normal Distribution

z (-∞)	z (43.8803)	P (X ≤ 43.8803)
-∞	-0.439912	0.33
Dev (z σ)		
45.2 - 1.31974		

Antwort: 33 % der Forellen im Becken haben eine Länge bis rund 43,9 cm!

Der Herr Professor hat uns vor dem Berechnen eines neuen Beispiels gefragt, was zB P(μ- x ≤ X ≤ μ + x) = 0.90 heißt - also nach unserer Interpretation von dieser Formel gefragt!

Antwort: 90 % der Fische haben eine Länge von 45.2 - einer unbekannten Zahl und 45.2 und + dieser unbekannten Zahl!!!

4) Mitarbeit über Normalverteilung, Tossing a Die Example 17.12.2004

Öffnen / Schließen

Kleine Mitarbeit über Normalverteilung:

Aufgabe:
--> alles mit der Hand und am PC (+Ausdruck)

GROUP PROJECT Finding the First Employment
In an European country, the age at which people find their first employment can be assumed to be roughly normal distributed for men and women. The mean value for men and women is given below. The standard deviation for both sexes is 2.3 at all times.

Age at first employment

Year	Women (Mean)	Men (Mean)
1970	16.4	17.5
1980	16.8	17.8
1990	17.1	18.0

1.

(a) Consider a woman randomly selected out of all women finding their first employment in 1990. What is the probability that she was over 21 when finding her first employment? What is the probability that she was under the age of 18? What is the probability that she was between the age of 16 and 18?

(b) Consider a man randomly selected out of all men finding their first employment in 1990. What is the probability that he was over 23 when finding his first employment? What is the probability that he was under the age of 19? What is the probability that he was between the age of 17 and 19?

(c) Consider a man randomly selected out of all men finding their first employment in 1990. How old did you have to be, to belong to the oldest 15% of all men who hadn't found an employment yet? How old did you have to be, in order to be among the oldest 7% of men not having been employed yet?

Ergebnis:

13 Schüler waren anwesend, 10 davon hatten alles richtig, 3 hatten 0 richtig und eine Schülerin hatte 1/4 richtig!

neues Thema

Da die Zeit knapp war, sind wir das Beispiel (a) nur handschriftlich durchgegangen; wir haben den Unterschied zwischen der absoluten und der relativen Häufigkeit kennen gelernt; am PC haben wir lediglich einen Teil gemacht.

Tossing a Die Example (1)

(a) A die is tossed 600 times. X is the relative frequency that the face 3 turns up. What is the probability that X differs from 1/6 by 0.02 at most?
0.02 corresponds to how many tosses? (Wurf)

P(1/6 - rel deviation ≤ X ≤ 1/6 + rel deviation) = ?

(b) A die is tossed 480 times. X is the relative frequency that the face 1 turns up. With a probability of 95% X will lie in which interval around 1/6?
What is the corresponding interval for the tosses?

P(1/6 - rel deviation ≤ X ≤ 1/6 + rel deviation) = 0.95

rel deviation = ? abs deviation = ?

(c) X is the relative frequency that the face 2 turns up. How often has the die to be tossed that X will lie in the interval 1/6 ± 0.015 with 99% ?

P(1/6 - rel deviation ≤ X ≤ 1/6 + rel deviation) = 0.99,

n = ?, interval of tosses = ?

Visualize your result in (a) - (d).

Im SÜ-Heft:

--> Wir gehen das Beispiel Schritt für Schritt durch (was mit dem PC gerechnet, einfach ausgelassen wird), um für dieses Beispiel und die dazu gebrauchten/entstandenen Formel ein "Gefühl" bekommen.

a) Würfel, 600 Mal gewürfelt, X ist die relative Häufigkeit, dass die Zahl 3 kommt. Wie groß ist die Wahrscheinlichkeit, dass die relative Häufigkeit X von 1/6 um 0,02 abweicht?

Input ▹
```
n = 600;
p = 1/6;
{μ = n p} // N
```

```
{100.}
```

Antwort: Wir erwarten, dass bei 600 Würfelversuchen die Zahl 3, 100 Mal kommt.

Input ▷ `{ p = 1 / 6 } // N`

`{0.166667}`

Antwort: Der exakte Wert beträgt 0.1666667.

Notiz: Der Herr Professor zeigt uns ein kleines "Experiment" um uns die Relative Häufigkeit verständlicher zu machen.

Input ▷ `{h3 = 107 / 600} // N`

`{0.178333}`

Input ▷ `h3 - p // N`

`0.0116667`

Antwort: Die relative Häufigkeit h3 beträgt 0.178333 und die Schwankung zwischen der relativen Häufigkeit und dem exakten Wert beträgt knapp 1%.

Wir kommen wieder zurück zu unserem eigentlichen Beispiel:
--> Wir zählen zu unserem p=1/6 einmal 0.02 dazu (absoluter Wert) und ziehen einmal 0.02 (Abweichung) ab.

Input ▷ `{rW = 1 / 6 - 0.02} // N`

`{0.146667}`

Input ▷ `rW * 600`

`88.`

Antwort: Der relative Wert gibt uns 88 Mal die Zahl drei aus bei 600 Würfen.

Input ▷ `{aW = 1 / 6 + 0.02} // N`

`{0.186667}`

Input ▷ `aW * 600`

`112.`

Antwort: Der absolute Wert gibt uns 112 Mal die Zahl drei aus bei 600 Würfen.

Am PC:
Wir rechnen die Wahrscheinlichkeit aus, dass bei 100 Versuchen, die Häufigkeit zwischen diesen Schwankungen (88 bis 112) liegt.

Input ▹

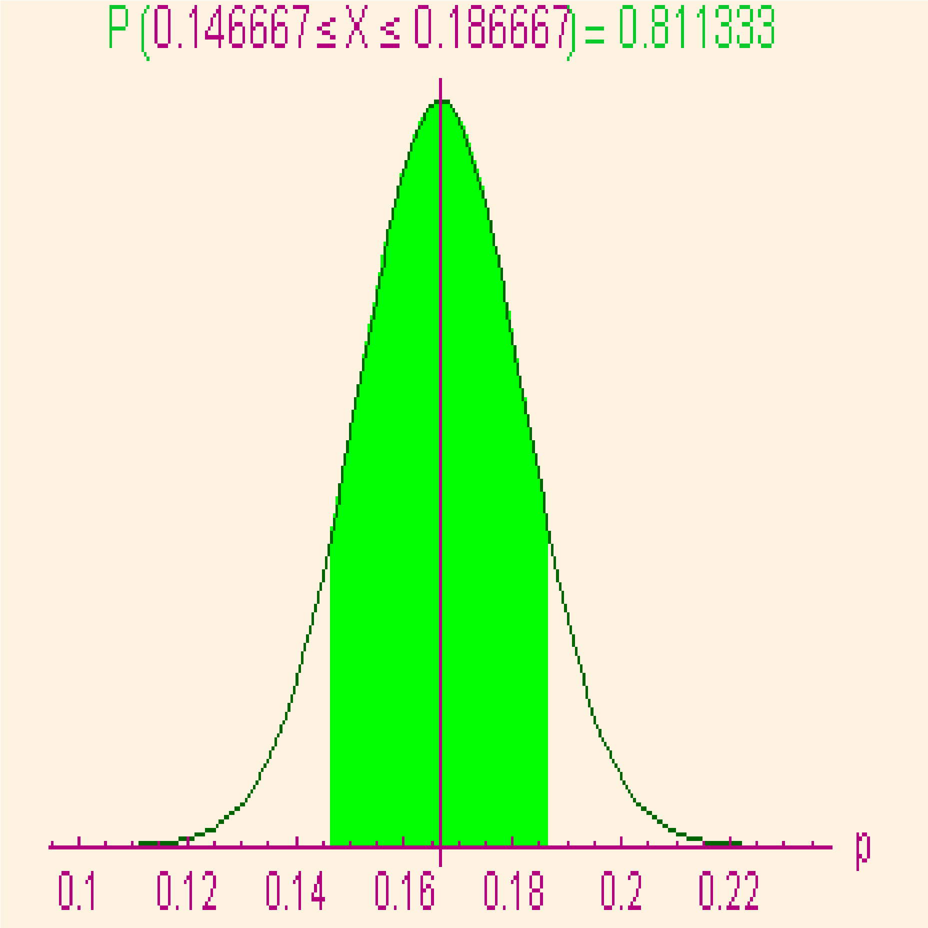
```
Switch to Abs Deviation ;
n = 600;
p = 1 / 6;
z = □;
reldeviation = 0.02;

var = MDEnterVar[z][[1]]; (*n,p,z,reldeviation*)
{μ = n p, σ = √(n p (1 - p)) , deviation = reldeviation n};

z = var /. NSolve[ z σ / n == reldeviation , var ][[1]]
```

```
1.31453
```

Antwort: Am PC gerechnet haben wir nur den z-Wert errechnet. Dieser beträgt 1.31453.

Input ▹

```
Switch to μ, σ ;
reldeviation = Abs[reldeviation];

MDSNormalDistributionPnp[
    p - reldeviation ≤ X ≤ p + reldeviation, {n, p} ]
```



Normal Distribution

z (0.146667)	z (0.186667)	P (0.146667 ≤ X ≤ 0.186667)
-1.31453	1.31453	0.811333
Dev (z σ)	Rel Dev (z σ / n)	
100. ± 12.	0.166667 ± 0.02	

Antwort: Die Wahrscheinlichkeit, dass die relative Häufigkeit von 1/6 um 0.02 abweicht, liegt bei 81 %.

Wir erstellen eine weitere Formelsammlung:

absolute Abweichung (Deviation $z\sigma$):

$$P(\mu - z\sigma \leq x \leq \mu + z\sigma)$$

relativer Wert: (absolut Wert / n):

$$\frac{P\left(np - z\sqrt{np\,(1-p)} \le x \le np + z\sqrt{np\,(1-p)}\right)}{n}$$

Zerlegen wir die Formel erhält man die einzelnen relative Werte:

relative Häufigkeitsvariable

$$\frac{x}{n}$$

relative Abweichung

$$z\,\frac{\sqrt{np\,\sqrt{(1-p)}}}{n}$$

5) Wiederholung des Tossing a Die Examples 14.01.2005

Öffnen / Schließen

Da wir vor den Weihnachten erstmals das "Tossing a Die" Example gemacht haben, und wir unseren Professor darum gebeten haben es noch einmal durchzugehen, haben wir in unseren letzten Stunde vor der Schularbeit am 21.01.2004 das Beispiel noch einmal wiederholt.

Dieses Mal haben wir uns besonders auf das Handschriftliche konzentriert.

- SA-Stoff:
  Normalverteilung
  Tossing a Die Example

Im SÜ-Heft:
Wir sind von folgender Formel ausgegangen: (relative Häufigkeit)

$$P\left(\frac{1}{6} - 0.02 \le x \le \frac{1}{6} + 0.02\right) = ?$$

--> 600 x wird gewürfelt und man zählt die Anzahl der 3er! In diesem Fall stehen die $\frac{1}{6}$ für den Mittelwert.

Wir führen zum Verständnis ein kleines Experiment durch und gehen willkürlich davon aus, dass bei 600 Würfen 107 3er dabei sind.

Input ▹
```
h3 = 107/600 // N
```

```
0.178333
```

Notiz: Unsere relative Häufigkeit liegt bei 0.178333.

Wenn man die obige Formel ausrechnet, kommt man auf folgendes Ergebnis

$$P(0.1466 \le x \le 0.1866) = ?$$

Antwort: Unser Experiment zeigt uns, dass das Ergebnis der relativen Häufigkeit (0.178333) im Bereich der Formel liegen würde!!!

...zurück zu unserem eigentlichen Beispiel

Im SÜ-Heft:
Jetzt errechnen wir die relative Häufigkeit anhand der oben genannten Formel
P (μ -zσ $\le$ x $\le$ μ +zσ).
Dividiert man diesen ganzen Term durch n (Anzahl der gesamten Würfe), bekommt man die absolute Häufigkeit.

Wir kommen zu folgenden Formeln:

$$z\sigma \;=\; \text{absolute Abweichung}$$

$$z\sqrt{\frac{p\,(1-p)}{n}} = \text{relative Abweichung}$$

Da wir eine Abweichung von 0,02 gegeben haben, können wir anhand dieser Formeln das z errechnen, um dann mit der Formel

$$2\,\sigma\,[z] - 1 = ?$$

die Wahrscheinlichkeit zu errechnen.

Woher wissen wir, dass diese Formel anzuwenden ist?
--> Anhand der symmetrischen Abweichung von 0.02. (gegeben in der Angabe)

b.) Dieses Mal ist eine Wahrscheinlichkeit von 95 % gegeben und wir müssen die relative Abweichung errechnen! Also gehen wir wieder von folgender Formel aus:

$$P\left(\frac{1}{6} - \text{reldeviation} \leq x \leq \frac{1}{6} + \text{reldeviation}\right) = 0.95$$

--> In diesem Fall rechnet man wieder Schritt für Schritt um zu den Ergebnis (rel.Häufigkeit) zu gelangen, jedoch in umgekehrter Reihenfolge!

Wir wissen, dass es sich um eine symmetrische Abweichung handelt, also wenden wir zu Anfang wieder diese Formel an:

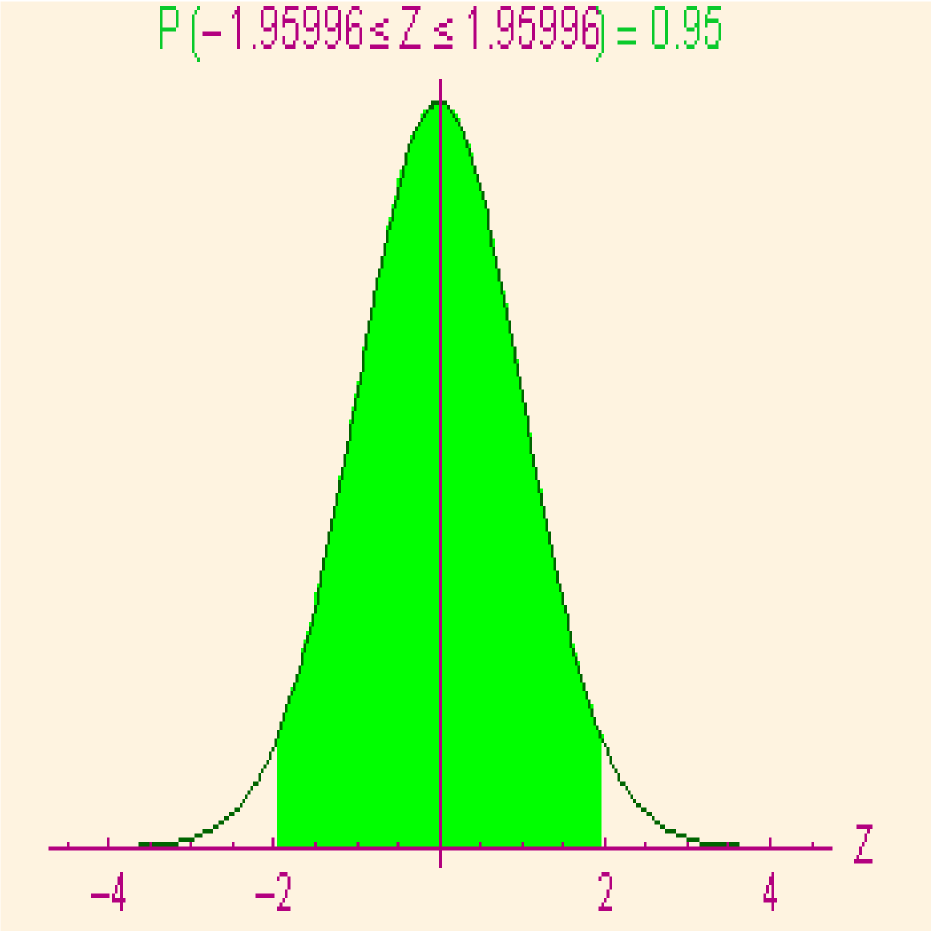
$$2\,\sigma\,[z - 1] = 0.95$$

<u>Am PC:</u>
Mit *Mathematica* sieht die Rechnung wiefolgt aus:

```
Switch;
probability = 0.95;
Clear[z];
```

Input ▹

```
z = z /. FindRoot[2 MDSϕ[z] - 1 == probability, {z, 0}];

MDSNormalDistributionPμσ[ -z ≤ Z ≤ z,  {0, 1}]
(* P ( -? ≤ Z ≤ ?) = probab *)
```



Normal Distribution

z (-1.95996)	z (1.95996)	P (-1.95996 ≤ Z ≤ 1.95996)
-1.95996	1.95996	0.95
Dev (z σ)		
0. ± 1.95996		

Antwort: Der z-Wert beträgt 1.95996.

Um nun die relative Abweichung zu errechnen, nimmt man das errechnete z her und und setzt es in die passende Formel ein:

$$z\sqrt{\frac{p(1-p)}{n}} = \text{reldeviation}$$

Input ▷

```
Switch to Abs Deviation ;
n = 600;
p = 1/6;
z = 1.95996;
reldeviation = r;

var = MDEnterVar[r][[1]];(*n,p,z,reldeviation*)
{μ = n p, σ = √(n p (1 - p)), deviation = reldeviation n};

r = var /. NSolve[z σ / n == reldeviation, var][[1]]
```

```
0.0298198
```

Antwort: Die relative Abweichung beträgt 0.0298198.

c.) Aufgabenstellung: Wie oft muss man Würfeln, dass mit 99 %iger Wahrscheinlichkeit der 3er in dem Intervall von $\frac{1}{6}$-0.015 und $\frac{1}{6}$+0.015 liegt?

--> Nun haben wir alles bis auf n gegeben und müssen auf die gleichen Formel wie in Punkt b.) zurückgreifen, jedoch müssen wir die Formel der relativen Abweichung umformen, um auf n zu gelangen!

Tossing a Die 2:
P (50 ≤ x) = 0.9 --> ist gegeben
Jetzt müssen wir das z in unserer Tabell nachschauen.
Da x größer als X ist, wendet man diese Formel an:

$$1 - \sigma[z] = -1.285$$

50 bzw. das X steht für $\mu + z\sigma$ = Häufigkeit!!!

$$n\,\frac{1}{6} \;-\; 1.2816\;\sqrt{n\,\frac{1}{6}\,\frac{5}{6}} = 50$$

μ $\quad$ **z** $\quad$ σ

Im SÜ-Heft:
die $\sqrt{n}$ haben wir einfach zu **u** gemacht und haben somit eine sogenannte Substitution (Vereinfachung) erstellt, die wir für eine quadratische Gleichung nutzen konnten.

6) Mathematik-SA, 21.01.2005

Öffnen / Schließen

- Das Stoffgebiet:
  Normalverteilung
  Tossing a Die Example

Es war alles am PC und mit der Hand zu machen!

3. (1) z-values:

Solve the problems by hand and on the PC.
Write an precise answer to each problem on the PC.

(a) Find the following probabilities for the standard normal random variable Z:

$P(Z \geq 0.37) = ?$

(b) Find a value of the standard normal random variable Z, call it z, such that

$P(Z \geq z) = 0.41$

4. (2) Ambulance:

Solve the problems by hand and on the PC.
Write an precise answer to each problem on the PC.

The time between an emergency call and the first ambulance arriving at the place of accident has been measured over a long time. This time is approximately normal distributed with a mean of 8.6 minutes and a standard deviation of 3.7 minutes.

Consider a randomly chosen accident. What is the probability that the first ambulance arrives

(a) after more than 13 minutes?
(b) after less than 7 minutes?

(c) In which time intervall will the ambulance arrive with 95%?

5. (3) Arnold Schwarzenegger stadium:

Solve the problems by hand and on the PC.
Write an precise answer to each problem on the PC.

The Arnold Schwarzenegger stadium contains 15400 seats. It is known, that 13% of all people reserving seats never turn up at their seat.

How many reservations can the manager accept for a game, such that he can still be 99% confident, that he'll not run out of seats?

(a) Find the z value
(b) Determine n. Visualize your result.

Hint to (b) : After clicking on Plot Deviation, click Switch to μ,σ and edit the input to MDSNormalDistributionP$\mu\sigma$[X ≤ 15400, {μ,σ}]

Ergebnis: Die Schularbeit ist allgemein sehr gut ausgefallen.

Sehr gut: 8
Gut: 4
Befriedigend: 3

7) Assessing Normality Random Data, 28.01.2005

Öffnen / Schließen

Zuerst hat uns der Herr Professor noch ein paar Beispiele zu "Tossing a Die" rechnen lassen, wie zB Ausrechnen von **p**.
--> mit Hand und am PC
--> *MD Stat* - Normalverteilung - Grundlagen - Exercise Pool - 2. n,p Deviation - Calculating n,p deviation

Im SÜ-Heft und am PC:

6. Tossing a Coin A fair coin is tossed n = 600 times. Find the probability that the number of heads will not differ from 1/2 by more than 0.02

Solution: P(1/2 - 0.02 ≤ X ≤ 1/2 + 0.02) = 0.67

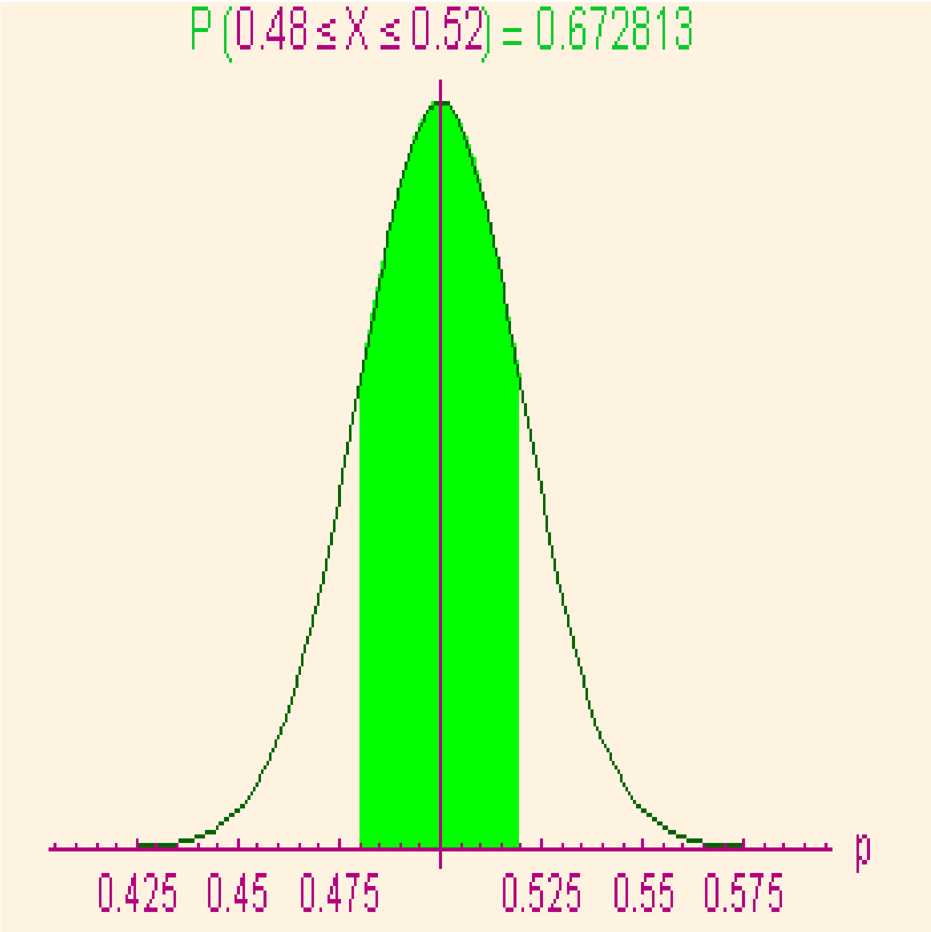
Input ▹

```
Switch to Abs Deviation ;
n = 600;
p = 1 / 2;
z = z;
reldeviation = 0.02;

var = MDEnterVar[z] [[1]]; (*n,p,z,reldeviation*)
{μ = n p, σ = √(n p (1 - p)) , deviation = reldeviation n};

z = var /. NSolve[ z σ / n == reldeviation , var ] [[1]]
```

0.979796

Input ▹

```
Switch to μ, σ ;
reldeviation = Abs[reldeviation];

MDSNormalDistributionPnp[
    p - reldeviation ≤ X ≤ p + reldeviation, {n, p} ]
```



Normal Distribution

z (0.48)	z (0.52)	P (0.48 ≤ X ≤ 0.52)
-0.979796	0.979796	0.672813
Dev (z σ)	Rel Dev (z σ / n)	
300. ± 12.	0.5 ± 0.02	

Antwort: Die Wahrscheinlichkeit, dass unter 600 Würfen von einer Münze, der Kopf 300 mal kommt unter Berücksichtigung einer Abweichung von 12 mal, beträgt 67 %.

7. Number of Heads A fair coin is tossed n = 600 times. Find the probability that the number of heads will not differ from 300 by more than 13.

Solution: P(300 - 13 ≤ X ≤ 300 + 13) = 0.71

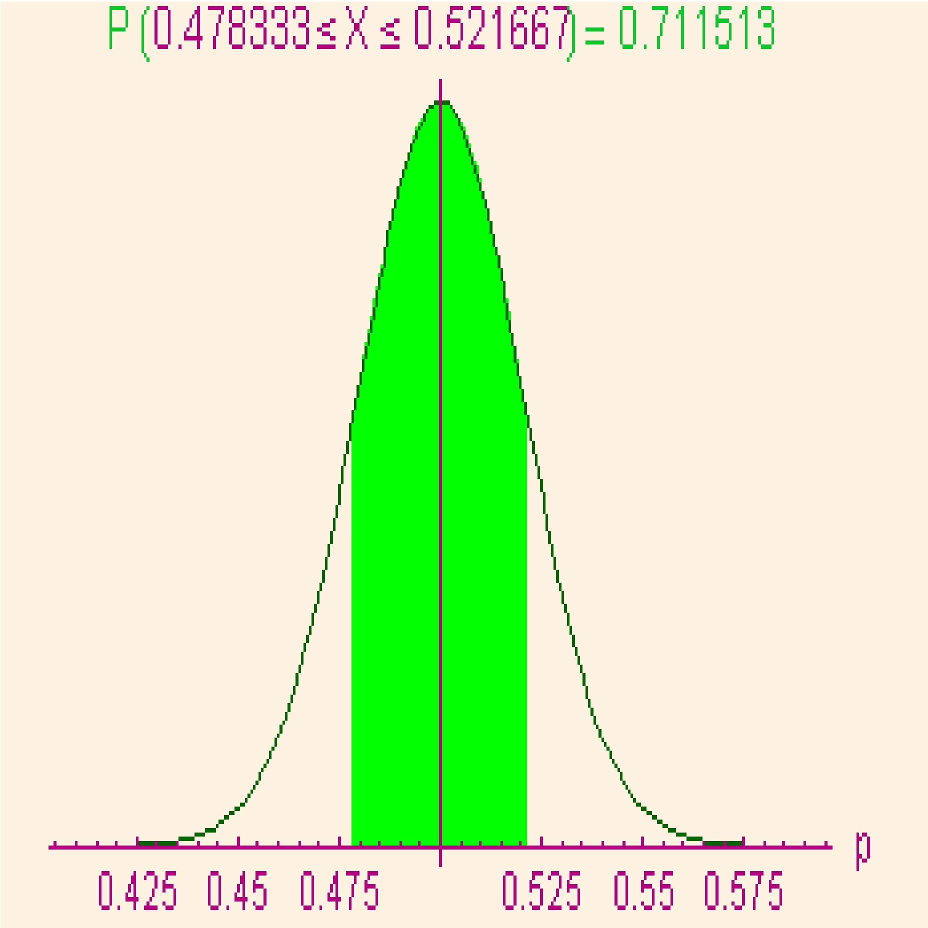
Input ▹

```
Switch to Abs Deviation ;
n = 600;
p = 300 / 600;
z = z;
reldeviation = 13 / 600;

var = MDEnterVar[z][[1]]; (*n,p,z,reldeviation*)
{μ = n p, σ = √(n p (1 - p)), deviation = reldeviation n};

z = var /. NSolve[ z σ / n == reldeviation , var ][[1]]
```

1.06145

Input ▹

```
Switch to μ, σ ;
reldeviation = Abs[reldeviation];

MDSNormalDistributionPnp[
    p - reldeviation ≤ X ≤ p + reldeviation, {n, p} ]
```



Normal Distribution

z (0.478333)	z (0.521667)	P (0.478333 ≤ X ≤ 0.521667)
-1.06145	1.06145	0.711513
Dev (z σ)	Rel Dev (z σ / n)	
300. ± 13.	0.5 ± 0.0216667	

Antwort: Die Wahrscheinlichkeit, dass unter 600 Würfen von einer Münze, der Kopf 300 mal kommt unter Berücksichtigung einer Abweichung von 13 mal, beträgt 71 %.

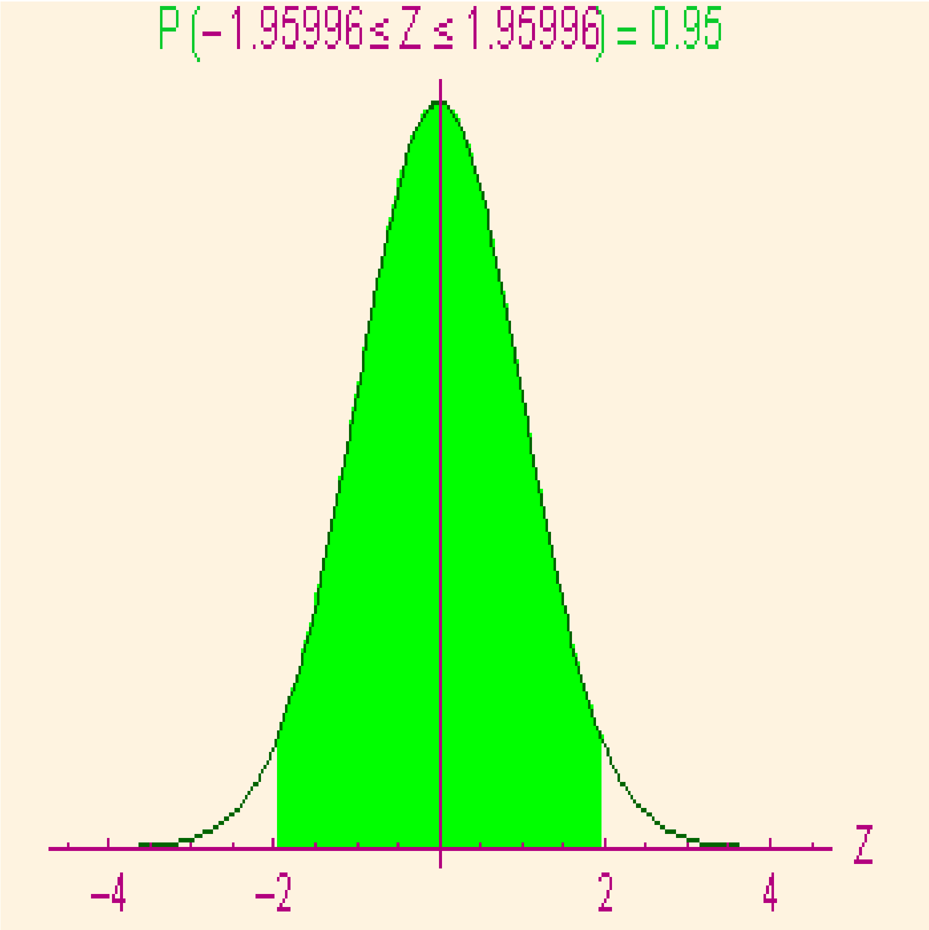
8. Relative Deviation A fair coin is tossed n = 600 times. With 95% probability the deviation will not differ from 1/2 by more than ± which value?

Solution: relative deviaiton = 0.04

```
Switch;
probability = 0.95;
Clear[z];
```

Input ▹

```
z = z /. FindRoot[2 MDSϕ[z] - 1 == probability, {z, 0}];

MDSNormalDistributionPμσ[ -z ≤ Z ≤ z,  {0, 1}]
(* P ( -? ≤ Z ≤ ?) = probab *)
```



Normal Distribution

z (-1.95996)	z (1.95996)	P (-1.95996 ≤ Z ≤ 1.95996)
-1.95996	1.95996	0.95
Dev (z σ)		
0. ± 1.95996		

```
Switch to Abs Deviation ;
n = 600;
p = 1 / 2;
z = z;
reldeviation = r;
```

Input ▹

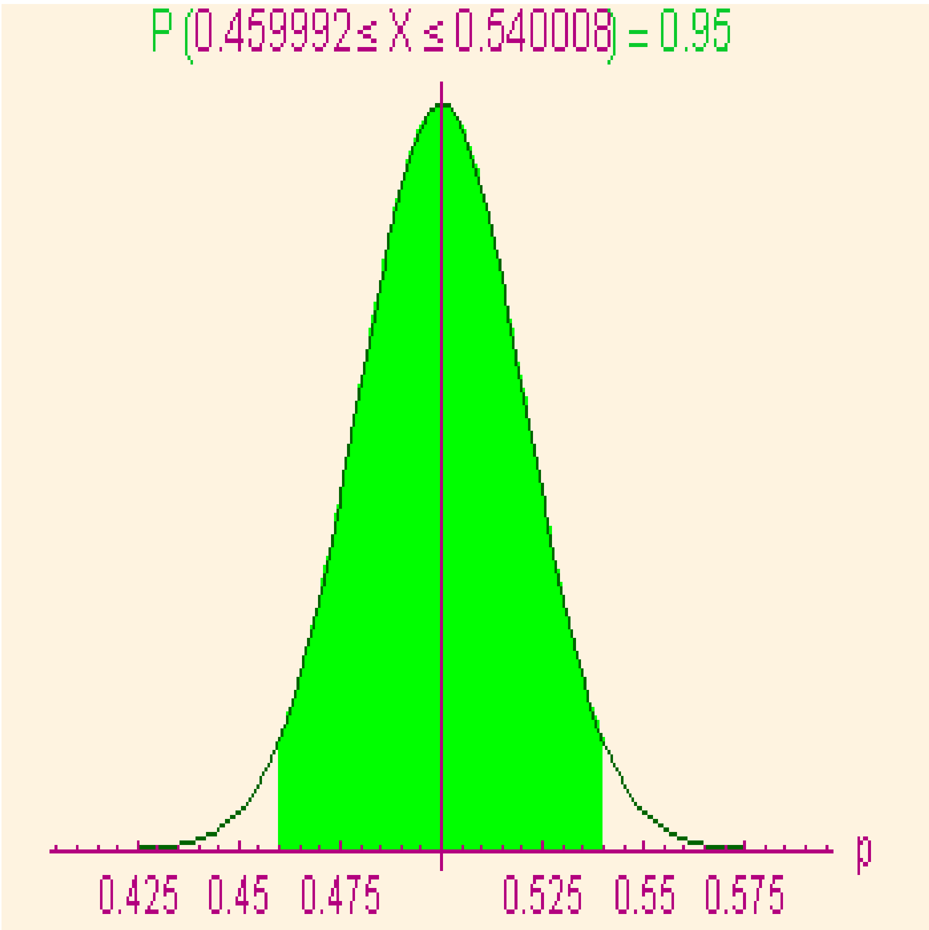
```
var = MDEnterVar[r][[1]];(*n,p,z,reldeviation*)
{μ = n p, σ = √(n p (1 - p)), deviation = reldeviation n};

r = var /. NSolve[ z σ / n == reldeviation, var ][[1]]
```

0.0400076

```
Switch to μ, σ ;
reldeviation = Abs[reldeviation];
```

Input ▹

```
MDSNormalDistributionPnp[
    p - reldeviation ≤ X ≤ p + reldeviation, {n, p} ]
```



Normal Distribution

z (0.459992)	z (0.540008)	P (0.459992 ≤ X ≤ 0.540008)
-1.95996	1.95996	0.95
Dev (z σ)	Rel Dev (z σ / n)	
300. ± 24.0046	0.5 ± 0.0400076	

Antwort: Bei einer Wahrscheinlichkeit von 95 %, dass der Kopf (einer Münze) 300 mal unter 600 Würfen kommt, liegt die relative Abweichung bei 0.04.

9.

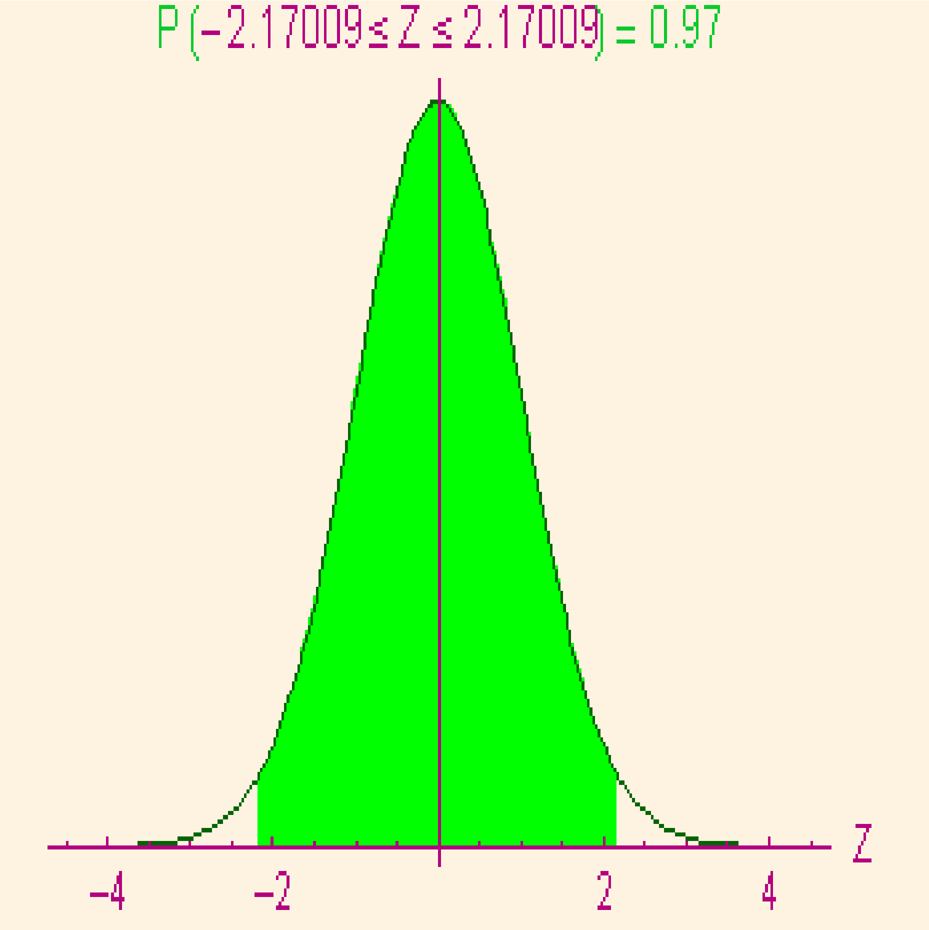
Difficult * Finding p   p is the percentage of broken bulbs. 400 bulbs are delivered. For which p is
the deviation from p at most 0.02 with a probability of 97%?

P(p - reldeviation ≤ X ≤ p + reldeviation) = 0.97 , p = ?

Hint: Note there are **two** p solutions to the quadratic equation.
....NSolve[z σ /n == reldeviation ,var][[1]] or
....NSolve[z σ /n == reldeviation ,var][[2]]
Take that one with the lower p.

Solution: p = 0.035

Input ▹
```
Switch;
probability = 0.97;
Clear[z];
z = z /. FindRoot[2 MDSϕ[z] - 1 == probability, {z, 0}];

MDSNormalDistributionPμσ[ -z ≤ Z ≤ z,  {0, 1}]
(* P ( -? ≤ Z ≤ ?) = probab *)
```



Normal Distribution

z (−2.17009)	z (2.17009)	P (−2.17009 ≤ Z ≤ 2.17009)
−2.17009	2.17009	0.97
Dev (z σ)		
0. ± 2.17009		

Input ▹

```
Switch to Abs Deviation ;
n = 400;
p = p;
z = z;
reldeviation = 0.02;

var = MDEnterVar[p][[1]];(*n,p,z,reldeviation*)
{μ = n p, σ = √(n p (1 − p)), deviation = reldeviation n};

p = var /. NSolve[ z σ / n == reldeviation , var ][[2]]
```

0.0352155

Antwort: Der Prozentanteil kaputter Glühbirnen mit einer relativen Abweichung von 0.02 unter 400 gelieferten Glühbirnen beträgt in 97 % der Fälle nur 3,5 %.

10. Probability n = 650, p = 0.75, relative deviation = 0.02

P(p - relative deviation ≤ X ≤ p + relative deviation) = ?

Solution: 0.76

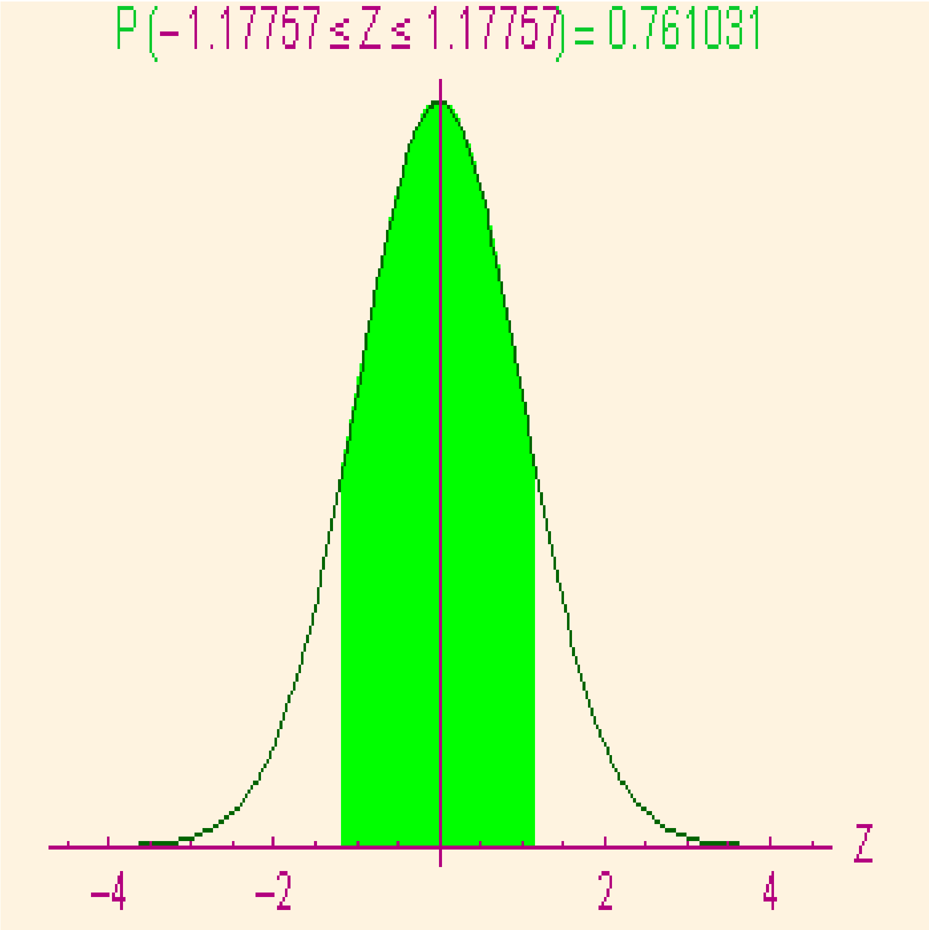
Input ▹

```
Switch to Abs Deviation ;
n = 650;
p = 0.75;
z = z;
reldeviation = 0.02;

var = MDEnterVar[z][[1]];(*n,p,z,reldeviation*)
{μ = n p, σ = √(n p (1 - p)) , deviation = reldeviation n};

z = var /. NSolve[ z σ / n == reldeviation , var ][[1]]
```

```
1.17757
```

Input ▹

```
Switch ;
z = z ; (* z > 0 *)
probability = 2 MDSϕ[z] - 1;

MDSNormalDistributionPμσ[ -z ≤ Z ≤ z,  {0, 1}]
(* P ( -z ≤ Z ≤ z) = ? *)
```



Normal Distribution

z (−1.17757)	z (1.17757)	P (−1.17757 ≤ Z ≤ 1.17757)
−1.17757	1.17757	0.761031
Dev (z σ)		
0. ± 1.17757		

Antwort: Bei einer Anzahl von n=650, einem Mittelwert von p= 0.75 und einer relativen Abweichung von 0.02, beträgt die Wahrscheinlichkeit 76 %.

Wir bekommen das aktuelle *Math Desktop* und beginnen mit dem Kapitel:

Accessing Normality-Random Data

Für besseres Verständnis schreiben wir uns Merksätze für das neue Kapitel auf, machen ein paar Probebeispiele am PC, rechnen jedoch nichts mit der Hand.

In der Wirtschaft wird nicht normalverteilt (sondern komplexer - rechnen wir aber nicht), in der Biologie schon, was wir auch lernen.
Random Data Generation produziert Zufallsdaten - wir simulieren eine Stichprobe!
Warum???

Antwort: Die Stichprobe wird deshalb mittels Zufallszahlen simuliert, damit wir uns eine echte (und vor allem zeitaufwändige) Stichprobe ersparen.

Wie schon gesagt, lernen wir nur normalverteilte Zufallszahlen, was man bei unseren bisherigen Beispielen sehen kann.

zB Simulieren von Würfelresultaten:
Die Zufallszahlen müssen natürlich gleich verteilt sein (jede Zahl von 1-6 kommt nur einmal vor)
oder
zB Stichprobe von Körpergrößen:
Die Zufallszahlen müssen normal verteilt sein, um auch eine normal-verteilte Stichprobe zu erhalten

Notiz: Wir benutzen die Assessity Normality - Palette und üben mit "Random Data".

Am PC:
Mit dem μ,σ **Random Sample** -Button werden Zufallsdaten erzeugt.

Gas Mileage
(a) A car journal performs extensive tests on new car models to determine their mileage ratings. The mean was approximately 36 mpg and the standard deviation 2.7 mpg.
(a) Simulate testresults of 40 new cars.
(b) Show all testresults.
(c) Calculate that $\bar{x}$ (Xbar) and s of the testresults. The values of the testa are to be close to μ and σ.

Solution:

(a) The μ,σ Rand Sample button creates a sample of a population which is μ,σ normal distributed with an arbitrary (beliebig) sample size. Here you simulate a test on 40 cars.

Click the μ,σ Rand Sample button with μ = 36 and σ = 2.7. Change sampleSize to 40.

```
μ = 36 ; σ = 2.7 ;
sampleSize = 40;
```

```
data = Table[Random[NormalDistribution[μ, σ]],
        {sampleSize}];
Short[data, 40](* displays only some lines *)
```

```
{37.0487, 38.5175, 29.7668, 28.6475, 36.9618, 38.0133, 41.3504,
 37.5878, 32.2691, 36.967, 36.883, 36.6894, 35.6408, 37.199,
 38.0156, 33.7531, 38.9242, 31.9882, 37.2889, 33.7949, 40.2317,
 40.9545, 32.8534, 38.9539, 39.8039, 35.1642, 37.9018,
 33.8115, 38.3593, 29.8223, 34.4234, 38.3262, 33.1669,
 37.2711, 37.3045, 35.793, 37.5322, 37.5389, 39.3726, 32.1134}
```

(b) The Short command prints as a short form of data. If you want to see all simulated data, change 4 to 400. Note. Because of the random generator this data are different from the data of (a).

> Click the μ,σ Rand Sample button with μ = 36 and σ = 2.7. Change sampleSize to 40.
> Change 4 in Short to 400 hundred.

```
μ = 36 ; σ = 2.7 ;
sampleSize = 200;
```

Input ▹

```
data = Table[Random[NormalDistribution[μ, σ]],
        {sampleSize}];
Short[data, 200](* displays only some lines *)
```

```
{38.4463, 34.8481, 38.0697, 35.5688, 38.1466, 35.5267, 32.8751,
 35.2397, 32.5196, 33.9839, 37.4441, 36.5721, 40.0973, 32.3781,
 39.1776, 34.3853, 36.6686, 37.6252, 34.5003, 32.759, 36.3488,
 33.4284, 36.5971, 37.3223, 38.3245, 30.5495, 34.967, 36.9669,
 38.2175, 34.2587, 33.7656, 28.2357, 34.4703, 38.4105, 35.4946,
 39.2695, 36.6907, 30.3164, 35.4487, 35.6467, 37.0583, 37.4088,
 35.213, 34.875, 33.9597, 36.0567, 37.324, 35.3436, 35.2612,
 35.3935, 32.8614, 35.8563, 35.5364, 38.5761, 39.4252,
 34.4835, 37.2001, 36.7723, 34.5039, 34.2076, 37.702, 37.4961,
 39.1288, 33.3039, 37.7337, 32.2541, 34.3145, 37.0366, 30.373,
 34.7, 37.6673, 38.5297, 40.2465, 36.2764, 38.5152, 40.7091,
 35.0262, 37.7853, 33.1586, 34.0156, 33.9578, 36.2791, 29.3608,
 37.6991, 36.7262, 33.5675, 40.771, 33.4513, 33.3536, 37.5133,
 40.8266, 39.0177, 39.1131, 37.5889, 32.9753, 34.5455, 35.5441,
 33.17, 37.7756, 36.3902, 38.697, 35.6043, 35.1813, 33.9384,
 33.527, 34.5965, 39.1684, 34.2565, 36.1168, 35.5025, 33.0316,
 35.1853, 34.5591, 37.557, 37.7758, 40.2267, 38.3637, 34.1166,
```

```
37.3502, 35.9777, 39.8054, 37.7103, 37.4366, 37.1508, 38.0457,
29.1735, 32.7708, 39.5913, 38.3259, 35.3434, 37.9914, 35.9265,
39.2057, 34.1438, 38.0417, 36.8621, 34.0267, 33.4009, 36.2585,
31.5563, 35.6974, 34.5959, 37.0725, 37.6448, 33.7222, 36.7821,
38.9204, 36.4328, 36.1948, 38.7433, 35.4762, 35.188, 36.7283,
35.8537, 34.3867, 38.7077, 30.7888, 41.3733, 32.5623, 34.4547,
33.3109, 40.234, 37.4569, 38.9081, 39.9417, 39.1141, 33.4411,
37.7669, 31.5115, 36.0762, 31.1471, 39.2828, 34.7704, 38.5516,
37.0377, 36.1998, 36.9098, 40.7119, 39.2817, 39.3033,
35.1416, 33.7061, 37.7253, 33.9441, 34.6693, 35.2351,
34.3755, 33.4993, 37.7674, 36.1934, 34.55, 38.4761, 41.9129,
38.8703, 31.6724, 36.1899, 33.7172, 35.9053, 35.8982, 37.58}
```

Am PC:
Mit dem **$\bar{x}$, s** – Button wird die gerade gemachte Stichprobe analysiert.

(c) The mean and standard deviation are close to μ=36 and σ=2.7.

Use the $\bar{x}$,s button.

Input ▹
```
x̄ = Mean[data]; s = StandardDeviation[data];
n = Length[data];

MDSElementaryStatistics[ data];
```

Elementary Statistics

n	$\bar{x}$	s	SE Mean $s/\sqrt{n}$	s^2
200	36.0364	2.52174	0.178314	6.35915

Antwort: Normalerweise ist der $\bar{x}$, s – Button in der Nähe von μ und σ in der Population.

Notiz: Nebenbei bemerkt kann natürlich bei solchen Rechnungen keiner das Gleiche herausbekommen, da es sich immer nur um Zufallszahlen handelt, die der PC zufällig auswählt.

Am PC:
ein weiteres Beispiel

Downhill Race
44 competitors were in the race. The mean time was 113.4 seconds with a standard

deviation of 3.24 seconds.
(a) Simulate the elapsed time of the race. Put the data in a table.
(b) Calculate that $\overline{x}$ and s of the testresults. The values of the simulation are to be close to μ and σ.

Solution:

(a) The notebook to provide the normal distributed data with the table is called NormalDCreatingFilesTables.nb.

It can be opened by the menu command
MD Stat > Normal Distribution > Create Random Numbers.

Click on the menu *MD Stat* > Normal Distribution > Create Random Numbers
Open the section Creating data tables for simulations and create the table.

The simulation of the race might look like

Seconds	Seconds	Seconds	Seconds
113.976	108.659	115.542	114.589
114.688	113.756	111.099	112.848
115.041	115.991	110.169	115.342
114.081	111.696	109.938	111.519
117.913	111.83	111.927	108.88
113.308	113.29	116.842	109.236
122.961	110.935	113.543	108.749
108.84	110.756	118.454	110.592
111.843	115.605	117.463	108.374
113.247	118.79	110.491	113.478
114.467	117.043	114.49	116.166

Now create one of your own.

Wir haben uns an die oben genannte Anweisung gehalten und haben unsere eigene Tabelle erzeugt mit "*MD Stat* - Normalverteilung - Zufallsdaten erzeugen"

8) WH Assessing Normality, Random Data, 04.02.2005

Öffnen / Schließen

Am PC:
Simulation der Abfahrtszeiten -> wir gehen von jeder einzelnen Sigmaumgebung aus und vergleichen die Ergebnisse.

μ = 113,4
σ = 3,24

Notiz: Mit 68 %-iger Wahrscheinlichkeit liegen die Resultate in der Umgebung 113,4 ± 3,24 sec. (110 ≤ X ≤ 116)

Notiz: Mit 95 %-iger Wahrscheinlichkeit erwarten wir Resultate im Zeitintervall 113 ± 9 sec. (107 ≤ X ≤ 119)

Notiz: Wir erwarten, dass 99,7 % der Resultate im Bereich von 113,4 ± 9 sec. liegen (104 ≤ X ≤ 122)

--> Short [data ...
Die letzte Zahl im Short-Befehl gibt nur an, wieviele Zeilen ca. ausgegeben werden sollen.

--> Wir erzeugen ein File "HeightsWomenDick.dat", den wir mittels Befehl in den Tempordner speichern.

Input ▹

```
filename = "C:\\Temp\\HeightsWomenDick.dat";
μ = 64.7; σ = 2.4;
sampleSize = 40;

zleft = -4; (* randomrange delimiter *)
zright = 4; (* randomrange delimiter *)

text = "Heights";
columns = 4;

data := RandomArray[NormalDistribution[μ, σ], {sampleSize}];
data = Select[data, μ + zleft σ <= # <= μ + zright σ &];
If[MemberQ[data, _?(# < 0 &)],
  Print[ "Note: some elements in the file are <
     0. Check if your problem allows data < 0."]];
StylePrint[ToString@filename <> "  written with the data:",
  "Text"];

MDShowTable[Partition[data, columns], Table[text, {columns}],
  ColumnAlignments -> Center, CellStyle -> "Text"];

datafile = If[DirectoryName@filename == "",
   OpenWrite[ ToFileName[{MathDesktopStatistics`$MDSInstallDir,
      "Distributions", "NormalDistribution",
```

```
        "NormalDSampleFiles"}, filename]], OpenWrite[filename]];
    Write[datafile, data];
    Close[datafile];
```

C:\Temp\HeightsWomenDick.dat written with the data:

Heights	Heights	Heights	Heights
65.505	63.1648	66.7151	65.7872
59.6155	64.5197	68.0605	63.3092
64.4922	61.7167	65.868	63.883
64.1238	67.5379	63.3492	68.3608
66.3556	63.9291	70.0236	64.998
68.6081	62.3598	66.8364	64.4518
67.7558	66.4786	65.8113	63.4194
65.1984	67.6977	63.1003	58.5385
63.5127	66.6531	69.7706	64.7087
65.4409	61.9123	61.8949	64.7561

(**MD*ShowTable ...*) --> mit diesem Befehl kann man die Anzeige einer Tabelle ein- und ausschalten - bei vielen Daten sehr praktisch.

(b) Read the file HeightsOfYoungWome.dat and determine $\bar{x}$ and s of the data.

Click on the Import button to import the file HeightsOfYoungWomen.dat. Evaluate the cell.
Then use the $\bar{x}$,s button.

Input ▹
```
Short[data = ReadList[
    "C:\\Temp\\HeightsWomenDick.dat" , Expression] [[1]], 2]
(* Fileformat y or {x,y}:
          {1,-2.2,3,...} or { {1,2}, {-2.3,3},...} *)
```

```
{64.4695, 59.0229, ≪246≫, 64.0905, 61.7163}
```

Input ▹
```
x̄ = Mean[data]; s = StandardDeviation[data];
n = Length[data];

MDSElementaryStatistics[ data];
```

Elementary Statistics

n	$\bar{x}$	s	SE Mean $s/\sqrt{n}$	s^2
250	64.5928	2.33016	0.147372	5.42963

Antwort: Nach dem Befehl "xReadTable" oder "Import" werden alle Daten in der Variable "data" abgespeichert werden.

Im SÜ-Heft:

Simple Methods of Accessing Normality

Wir lernen 3 Methoden kennen, um die Normalität von Daten zu überprüfen.

1. Histogramm (lernen aber keine Hypothesentests)
2. Die 68-95-99.7 Regeln (3 Sigmaumgebungen)
3. Der Normal-Wahrscheinlichkeitsplot

Wichtig: Jede Stichprobe einer normalverteilten Population bringt eine normalverteilte Stichprobe mit sich.

Outliner: = Ausreißer, was bedeutet, dass eine Anhäufung von Daten außerhalb der üblichen Datenbreite liegt.

Am PC:
... haben wir zu jeder Methode ein Beispiel gemacht.

Click the μ,σ Rand Sample button with μ = 80 and σ = 3.2, then the 68-95-99.7 Rule button and the Normal Prob Plot button.
Highlight the cells and merge them with the menu command Cell > Merge Cells.

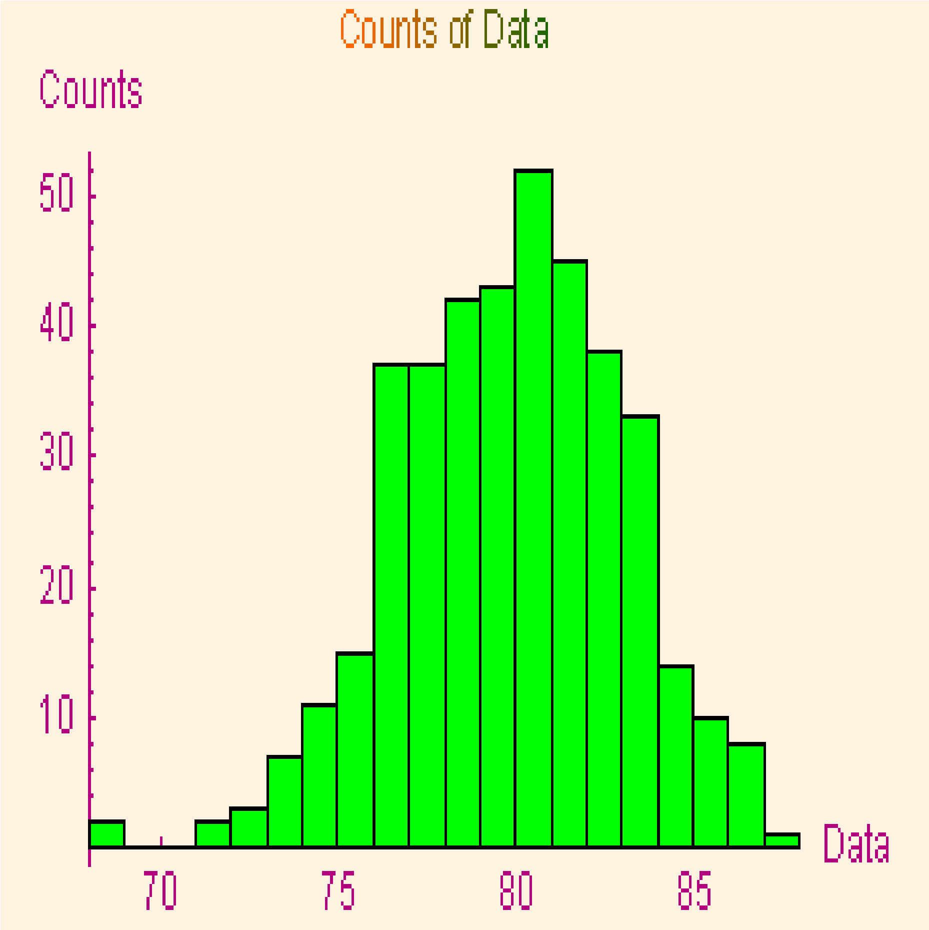
Then click the Answer button in the Tools section to describe your results.

```
μ = 80 ; σ = 3.2 ;
sampleSize = 400;
```

Input ▷
```
data = Table[Random[NormalDistribution[μ, σ]],
        {sampleSize}];
Short[data, 4](* displays only some lines *)
```

```
{80.1845, 75.2467, 83.9482, 79.1989, 76.7062, 76.0595,
 ≪389≫, 74.983, 79.1131, 82.5424, 79.4635, 77.2243}
```

Input ▷
```
MDSHistogram[ data ] ;  (* data format like {3,-2.2,1/2,...} *)
```



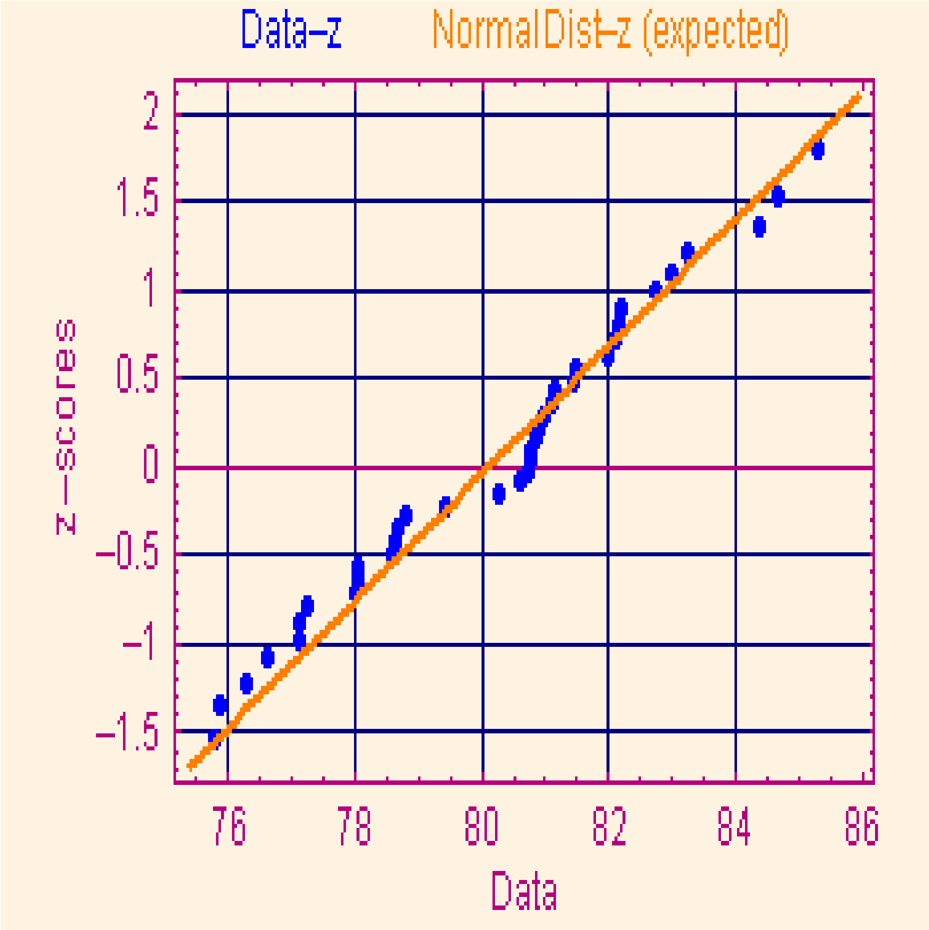
Antwort:
Mit einer erhöhten sampleSize von zB 400 oder sogar 4000 kann man ein genaueres und aussagekräftigeres Histogramm erstellen.
Hier zB haben wir einen Outliner.

Input ▹ `MDSCheckStandardDeviation[ data ];`

s – Interval Report

Interval	Percentage of Measurement	Expected
x̄ ± s	62.5 %	68 %
x̄ ± 2 s	97.5 %	95 %
x̄ ± 3 s	100. %	99.7 %

Input ▹ `MDSNormalProbabilityPlot[ data ];`



9) Besprechung des Projekts u.der HÜ 11.02.2005

Open / Close

Am Anfang der Stunde haben wir unser bevorstehendes Projekt anhand folgender Angabe von Herrn Prof. Simonovits besprochen:

Was ist ein Projekt
"Tell me and I will forget; show me and I may remember; involve me and I will understand"
-- Chinesisches Sprichwort
Der Projektinhalt: Das Projekt soll eine oder mehrere der statistischen Methoden, die in Vorlesung und Proseminar behandelt werden, an Hand von aktuellen Frage-stellungen anwenden. Die Wahl des Themas wird der Gruppe selbst überlassen. Die gewählten Themen sollen interessant sein und zum Verständnis von aktuellen, realistischen Fragen beitragen. Die Planung von Stichproben zur Beantwortung von interessanten Fragestellungen und die Analyse der erhobenen Daten können ein Projekt bilden. Aber auch Projekte, die bereits erhobene Daten auf ihre Zusam-menhänge analysieren, sind von Interesse. Öffentliche Daten (zum Beispiel vom Statistischen Zentralamt) können dazu verwendet werden. Daten können auch über das Internet erhoben werden. Projekte, die

zur Verbesserung von Prozessen inner-halb der HAK beitragen, sind besonders aktuell. Auch Projekte, die zur Verbesse-rung von Produktionsprozessen und/oder Dienstleistungsprozessen (sowohl der öffentlichen wie auch der privaten Hand) beitragen, sind von Interesse.

Beispiele für Projektthemen: Ausmaß der Nebenjobtätigkeiten von Schülern der HAK; über welches Budget verfügen die Schüler der HAK; Erhebung von Kundenzufriedenheit (z.B. des Buffets); Erfolg an der Schule und ihre Abhängigkeit von persönlichen Eigenschaften und Vorleistungen; Verdienst von Managern und deren Abhängigkeit von ihrer Ausbildung; usw. In all diesen Fällen muß zuerst überlegt werden, wie man die interessierenden Variablen am besten messen kann.
Euer Projekt muss zwei Aufgabenstellungen umfassen:
Die erste Aufgabenstellung wird mit den Methoden der Linearen Regression gelöst.
Die zweite Aufgabenstellung mit Hilfe der Beurteilung der Normalität von Daten.

Der Ablauf: Das Projekt ist ein wichtiger Teil des MNI Projketes. Das Projekt wird zu zweit durchgeführt. Die Projekte werden am 18.3.2005 präsentiert. Jedes Gruppenmitglied bekommt die gleiche Projektbeurteilung.
Die Arbeit ist in einem M@th Desktop notebook zu schreiben.
Der Projektbericht von ca. 7 Seiten muß färbig ausgedruckt bis 18.3.2005 mit dem M@th Desktop.nb vorgelegt werden. Der Bericht soll die Beschreibung der Problemstellung, die Erhebung der Daten, die zur Beantwortung der Fragestellung benötigt werden, die statistische Analyse, und die Interpretation der Ergebnisse beinhalten. Die Arbeit soll mit einer kurzen Beschreibung und Interpretation der Resultate beginnen (Executive Summary; nicht mehr als eine Seite). Die Arbeit soll auch auf etwaige Schwach-stellen der Analyse hinweisen und mögliche Weiterführungen aufzeigen. Detail-lierte statistische Analysen, die über das Seitenlimit hinausgehen, sollen im Anhang der Arbeit dargestellt werden. Eine Auflistung des verwendeten Daten-materials soll ebenfalls Teil des Anhangs sein.

Projektbewertung mit Beitrag zur Gesamtnote für das Projekt:
Problemstellung (Originalität der Fragestellung, Realitätsbezug) 20%
Qualität von Daten und Aussage (Bezug zur Fragestellung, Plausibilität der Argumentation) 30%
Präsentation (Qualität, verwendete Medien, Diskussion) 20%
Bericht (Qualität, Umfang) 30%

Hier ist ein Link von einer (komplizierteren) Aufgabe an der WU Wien
http://eeyore.wu-wien.ac.at/stat4/hackl/ss02/0207StudRef.htm

Notiz: Da wir 15 Schülerinnen und Schüler in der Klasse sind und 3 davon eine Prüfung machen, ergeben sich 6 Gruppen zu je 2 Personen
Fasching & Lambauer

Ponsold & Dieregger
Wolf & Bloder
Reiss & Urwalek
Kiegerl & Hable
Dick & Kloiber

Im SÜ-Heft:

Wir haben uns das Beispiel aus dem Punkt 2.Simple Methods of Asseccing Normality die Angabe von "Wie ein Normal Probality Plot funktioniert" herausgenommen und haben dieses Beispiel im SÜ-Heft errechnet - wieder mit Hilfe der ϕ[z]-Tabelle - und nachher auch gezeichnet.
Anschließend haben wir unseren gezeichneten Plot interpretiert.

How does a normal probability plot work?

Let's try an example with the small data set: {1, 0.9, 1.1, 1.8, 2.3 }, n = 5.

Step **1**: DIvide 100% in 5 equal intervals. If the 5 values are normally distributed, you might expect them to occur at, say, 10th, 30th, <50th, 70th and 90th percentiles.
The percentiles are the z-scores with comulative probabilties of 0.1, 0.3, 0.5, 0.7 and 0.9.

Use the Normal Distribution palette Basics, the ϕ^{-1}button, to look up the percentiles, the z-scores. The result is {-1.28, -0.52, 0, 0.52, 1.28}

The 0.1, 0.3, 0.5, 0.7 and 0.9 probabilities are the middle of each of the 5 intervals, determined by the formula (i - 0.5) / n, i ϵ {1,...,5}.

Step **2**: Sort the data. Plot to each data point on the x-axis {0.9,1,1.1,1.8,2.3}, the blue points, the percentiles on the y-axis {-1.28, -0.52, 0, 0.52, 1.28}.
The orange line is constructed by plotting the z value to each point, e.g. (0.9 - $\overline{x}$) /s, (1.1 - $\overline{x}$) /s, etc.

Notiz: {1, 0.9, 1.1, 1.8, 2.3 }, n = 5
Da n=5 gegeben ist, sind wir von 100 % ausgegangen und haben diese in 5 gleiche Intervalle dividiert.
1. 0 - 20 % = 10 %
2. 20 - 40 % = 30 %
3. 40 - 60 % = 50 %
4. 60 - 80 % = 70 %
5. 80 - 100 % = 90 %
Wir haben willkürlich die Mitte der einzelnen Intervalle gewählt und diese in z-Werte mit

Hilfe unserer ϕ[z]-Tabelle umgewandelt.
ϕ[z1] = 0.1
ϕ[z2] = 0.3
ϕ[z3] = 0.5
ϕ[z4] = 0.7
ϕ[z5] = 0.9

Input ▹
```
Switch;
probability = 0.1;
Clear[z]; z = z /. FindRoot[MDSϕ[z] == probability, {z, 0}];

MDSNormalDistributionPμσ[ Z ≤ z,  {0, 1}]
(* P ( Z ≤ ? ) = probab *)
```

Normal Distribution

z (−∞)	z (−1.28155)	P (Z ≤ −1.28155)
−∞	−1.28155	0.1
Dev (z σ)		
0. − 1.28155		

Antwort: Da 0.1 nicht "nachschlagbar" ist sozusagen, müssen wir von 0.9 ausgehen und beim Ergebnis ein Minus vorsetzen.
z1 = - 1.28

Input ▹
```
Switch;
probability = 0.3;
Clear[z]; z = z /. FindRoot[MDSϕ[z] == probability, {z, 0}];

MDSNormalDistributionPμσ[ Z ≤ z,  {0, 1}]
(* P ( Z ≤ ? ) = probab *)
```

Normal Distribution

z (−∞)	z (−0.524401)	P (Z ≤ −0.524401)
−∞	−0.524401	0.3
Dev (z σ)		
0. − 0.524401		

Input ▹
```
Switch;
probability = 0.5;
Clear[z]; z = z /. FindRoot[MDSϕ[z] == probability, {z, 0}];

MDSNormalDistributionPμσ[ Z ≤ z,  {0, 1}]
(* P ( Z ≤ ? ) = probab *)
```

Normal Distribution

z (−∞)	z (0.)	P (Z ≤ 0.)
−∞	0.	0.5
Dev (z σ)		
0. + 0.		

Input ▹
```
Switch ;
probability = 0.7 ;
Clear[z]; z = z /. FindRoot[MDSϕ[z] == probability, {z, 0}];

MDSNormalDistributionPμσ[ Z ≤ z,  {0, 1}]
(* P ( Z ≤ ? ) = probab *)
```

Normal Distribution

z (−∞)	z (0.524401)	P (Z ≤ 0.524401)
−∞	0.524401	0.7
Dev (z σ)		
0. + 0.524401		

Input ▹
```
Switch ;
probability = 0.9 ;
Clear[z]; z = z /. FindRoot[MDSϕ[z] == probability, {z, 0}];

MDSNormalDistributionPμσ[ Z ≤ z,  {0, 1}]
(* P ( Z ≤ ? ) = probab *)
```

Normal Distribution

z (−∞)	z (1.28155)	P (Z ≤ 1.28155)
−∞	1.28155	0.9
Dev (z σ)		
0. + 1.28155		

Notiz: Mit dem data und dem $\bar{x}$,s-Button werten wir die Daten aus, und errechnen $\bar{x}$.

Input ▹
```
data = { 0.9 , 1, 1.1 , 1.8, 2.3 } ;  Add ;
data
```

{0.9, 1, 1.1, 1.8, 2.3}

Input ▹
```
x̄ = Mean[data]; s = StandardDeviation[data];
n = 1 / 5[data];

MDSElementaryStatistics[ data];
```

Elementary Statistics

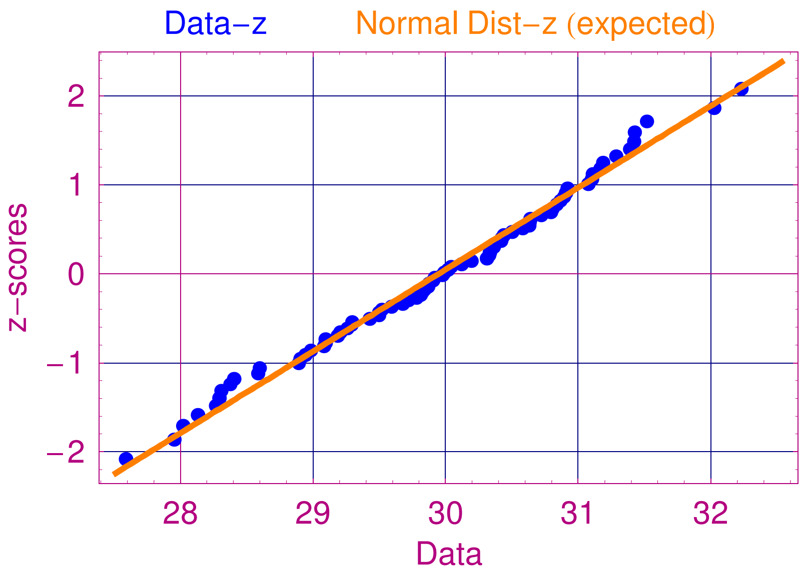
n	$\bar{x}$	s	SE Mean $s/\sqrt{n}$	s^2
5	1.42	0.605805	0.270924	0.367

Notiz: Nun haben wir die $\overline{x}$ und s-Werte und gehen von folgender Formel aus:
$X = \mu + z.\sigma$ und wandeln sie in $x = \overline{x} + z.s$ um!!!
Es ergeben sich folgende Koordinatenpunkte für die theoretische Normalverteilung:
{0.9, -0.86}
{1, -0.69}
{1.1, -0.53}
{1.8, 0.63}
{2.3, 1.45}
Diese Punkte ergeben die orange Linie im Plot

praktische Normalverteilung:
{0.9, -1.28}
{1, -0.53}
{1.1, 0.0}
{1.8, 0.53}
{2.3, 1.28}
Diese Punkte ergeben die blauen Werte im Plot.

Antwort:
Der Normalitätsplot rechnet alle Datenpunkte z-Werte um (über die Intervalle - Schritt 1); gleichzeitig werden die Punkte ausgerechnet, die eine ideale Normalverteilung wiedergeben (theoretische Normalverteilung) --> Formel $z = \frac{X - \overline{x}}{s}$ (Wir schätzen, dass μ durch das $\overline{x}$ von der Stichprobe und das σ durch das s ersetzt wird.

<u>Am PC:</u>
Beispiel für einen Normalitätsplot:



Antwort:
Wenn die Datenpunkte nicht stark von der orangen Linie (theoretische Verteilung) abweichen, können wir sagen, dass die Stichprobe normalverteilt ist und damit auch die Population.

10) Random Data, Poissonverteilung
Projektvorlage, 18.02.2005

Open / Close

Wir machen den Punkt 3.1 von den Examples und vergleichen die Ergebnisse bei Veränderung der sampleSize!

Am PC:

Impact of Sample Size
Create random number with $\mu = 80$ and $\sigma = 3.2$

(a) Simulate a sample with sampleSize = 40. Draw a histogram, check the 68-95-99.7 rule, perform a normal distribution plot.
Repeat this 5 times.
Describe the histogram forms, the greatest deviation from the expected percentages, the departure from straightness in the normal plot.

(b) Simulate a sample with sampleSize = 400. Draw a histogram, check the 68-95-99.7 rule, perform a normal distribution plot.
Repeat this 5 times.
Describe the histogram forms, the greatest deviation from the expected percentages, the departure from straightness in the normal plot.

(c) Simulate a sample with sampleSize = 4000. Draw a histogram, check the 68-95-99.7 rule, perform a normal distribution plot.
Repeat this 5 times.
Describe the histogram forms, the greatest deviation from the expected percentages, the departure from straightness in the normal plot.

Solution:

(a) - (c) The μ,σ Rand Sample button creates a sample of a population which is μ = 80 and σ = 3.2 distributed.
With the Histogram button, the 68-95-99.7 Rule button and the Normal Prob Plot button you can analyse the data.

Click the μ,σ Rand Sample button with μ = 80 and σ = 3.2, then the 68-95-99.7 Rule button and the Normal Prob Plot button.
Highlight the cells and merge them with the menu command Cell > Merge Cells.

Then click the Answer button in the Tools section to describe your results.

```
μ = 80 ; σ = 3.2 ;
sampleSize = 40;
```

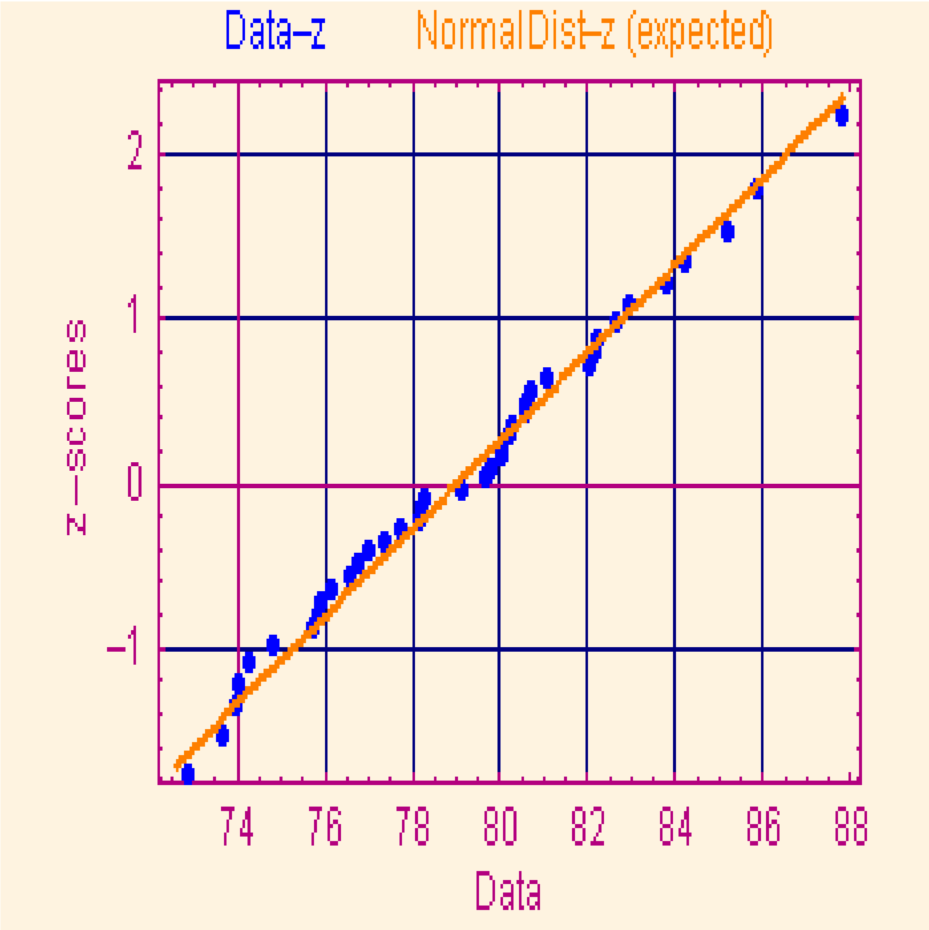
Input ▹

```
data = Table[Random[NormalDistribution[μ, σ]],
        {sampleSize}];
Short[data, 4](* displays only some lines *)
```

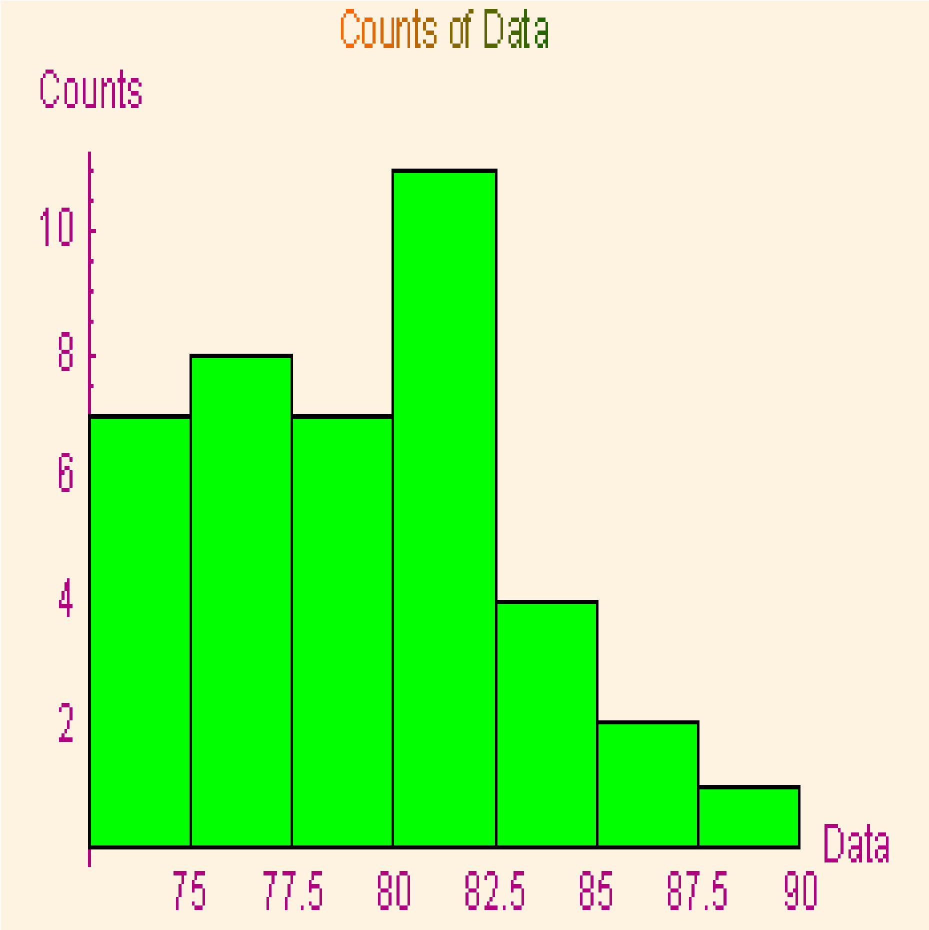
```
{77.7196, 75.8442, 80.5918, 74.0018, 80.29,
 ≪30≫, 81.0752, 72.8893, 82.673, 72.5979, 74.2895}
```

Input ▹

```
MDSNormalProbabilityPlot[ data ];
```



Input ▹ `MDSHistogram[ data ] ;  (* data format like {3,-2.2,1/2,...} *)`



Input ▹ `MDSCheckStandardDeviation[ data ];`

s – Interval Report

Interval	Percentage of Measurement	Expected
$\bar{x} \pm s$	67.5 %	68 %
$\bar{x} \pm 2\,s$	97.5 %	95 %
$\bar{x} \pm 3\,s$	100. %	99.7 %

Antwort:
Wir überlegen, ob die Outliner bzw. Schwankungen, welche man in den Auswertungen erkennen kann zu der Statistik gehören oder Fehler sind. Es handelt sich hierbei um typische Schwankungen.

```
μ = 80 ; σ = 3.2 ;
sampleSize = 4000;
```

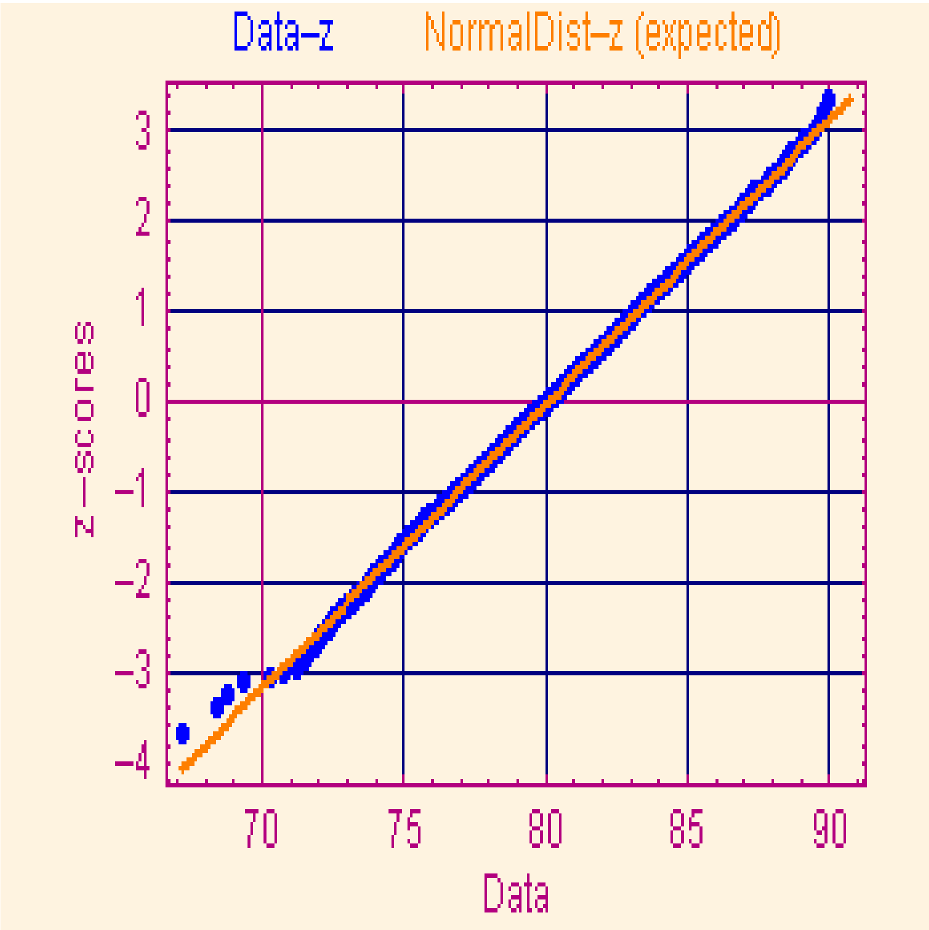
Input ▹

```
data = Table[Random[NormalDistribution[μ, σ]],
        {sampleSize}];
Short[data, 4](* displays only some lines *)
```

```
{82.1664, 79.559, 84.3378, 77.8965, 81.2039,
 ≪3990≫, 74.2425, 79.104, 82.1261, 85.6147, 75.4196}
```

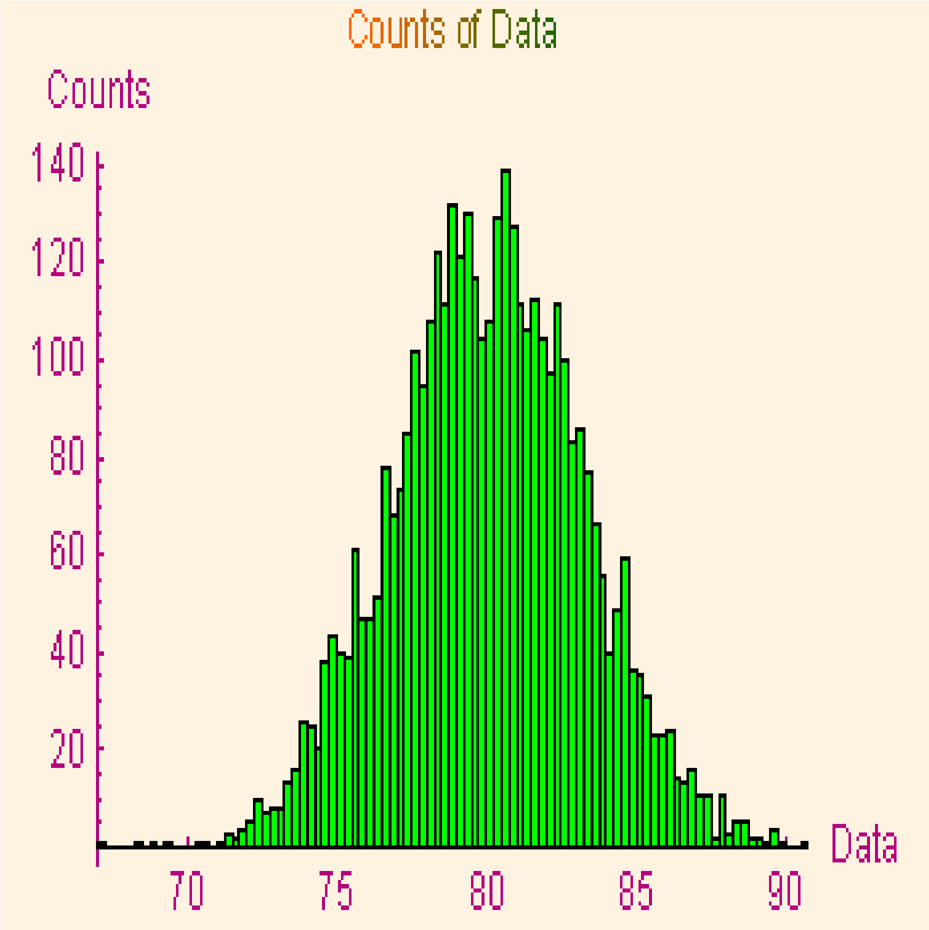
Input ▹

```
MDSNormalProbabilityPlot[ data ];
```



Input ▹

```
MDSHistogram[ data ] ;  (* data format like {3,-2.2,1/2,...} *)
```



Input ▹ `MDSCheckStandardDeviation[ data ];`

s – Interval Report

Interval	Percentage of Measurement	Expected
$\bar{x} \pm s$	68.475 %	68 %
$\bar{x} \pm 2\,s$	95.65 %	95 %
$\bar{x} \pm 3\,s$	99.75 %	99.7 %

Antwort:
Bei einer sampleSize von 4000 wird die Auswertung natürlich viel genauer, jedoch kann man zB beim Normalitätsplot durchaus die Outliner erkennen.

Am PC:
Jetzt machen wir das gleiche nur mit nicht normal verteilten Daten (Poisson)

Die Poissonverteilung ist sozusagen die "kleine Schwester" der Normalverteilung, da sie bei einer geringeren Wahrscheinlichkeit auftritt, jedoch aber immer mit der Normalverteilung zusammenhängt.

Man kann statistisch gesehen, ein Fehlurteil sprechen, weil die Ergebnisse einmal gut und einmal schlecht sein können.

Non Normal Distributions

Generate data from non normal distributions. Apply the three methods on these data. Describe how the methods "react" on non normal data.

(a) Simulate a Poisson distribution sample with sampleSize = 50 and $\lambda = 5$. Draw a histogram, check the 68-95-99.7 rule, perform a normal distribution plot.
Repeat this also with a sampleSize = 500.
Describe the histogram forms, the greatest deviation from the expected percentages, the departure from straightness in the normal plot.
Do you think the Poisson data can be described sufficiently by a normal distribution?

(b) Simulate a uniform distribution sample with sampleSize = 50. Draw a histogram, check the 68-95-99.7 rule, perform a normal distribution plot.
Repeat this also with a sampleSize = 500.
Describe the histogram forms, the greatest deviation from the expected percentages, the departure from straightness in the normal plot.
Do you think the uniform data can be described sufficiently by a normal distribution?

Solution:

(a) The μ,σ Rand Sample button creates a sample of a population which is μ,σ distributed. Change the input in order to generate Poisson distributed data with the parameter $\lambda = 5$.

Click the μ,σ Rand Sample button. Change NormalDistribution[μ,σ] to PoissonDistribution[5] in the input, sampleSize to 50 resp 500.

```
μ = □; σ = □;
sampleSize = 50;
```

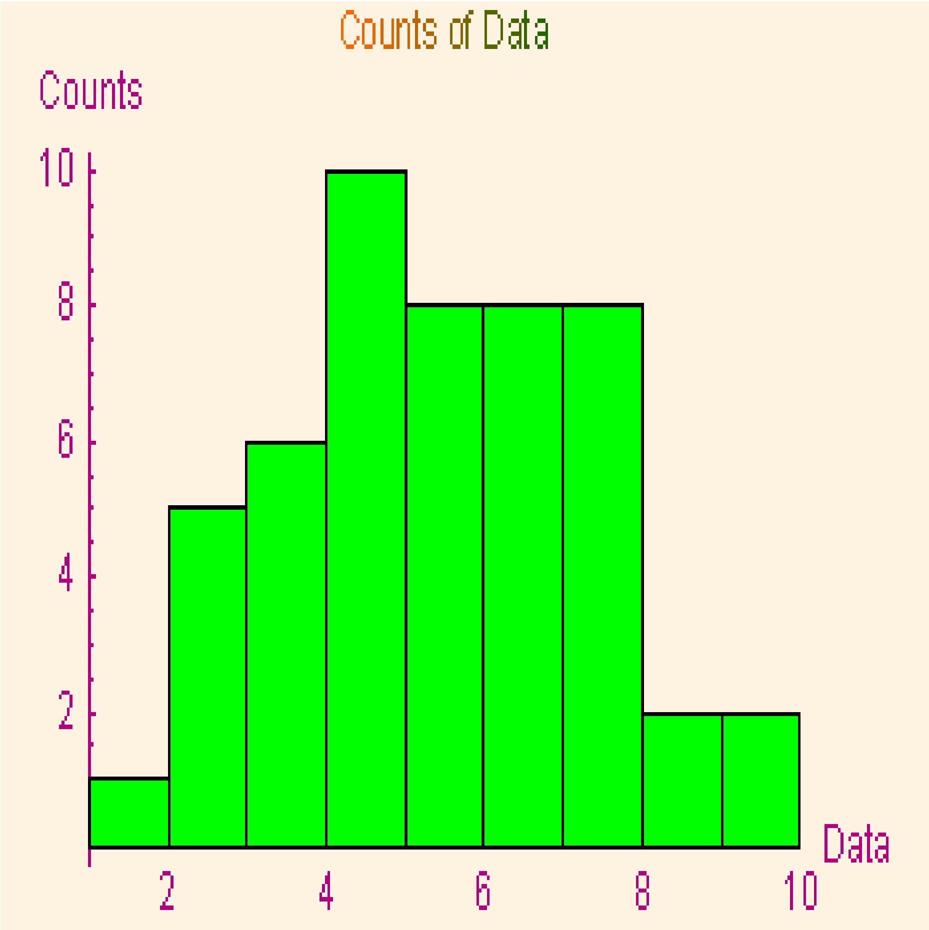
Input ▹

```
data = Table[Random[PoissonDistribution[5]],
        {sampleSize}];
Short[data, 4](* displays only some lines *)
```

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

Input ▹

```
MDSHistogram[ data ] ;  (* data format like {3,-2.2,1/2,...} *)
```

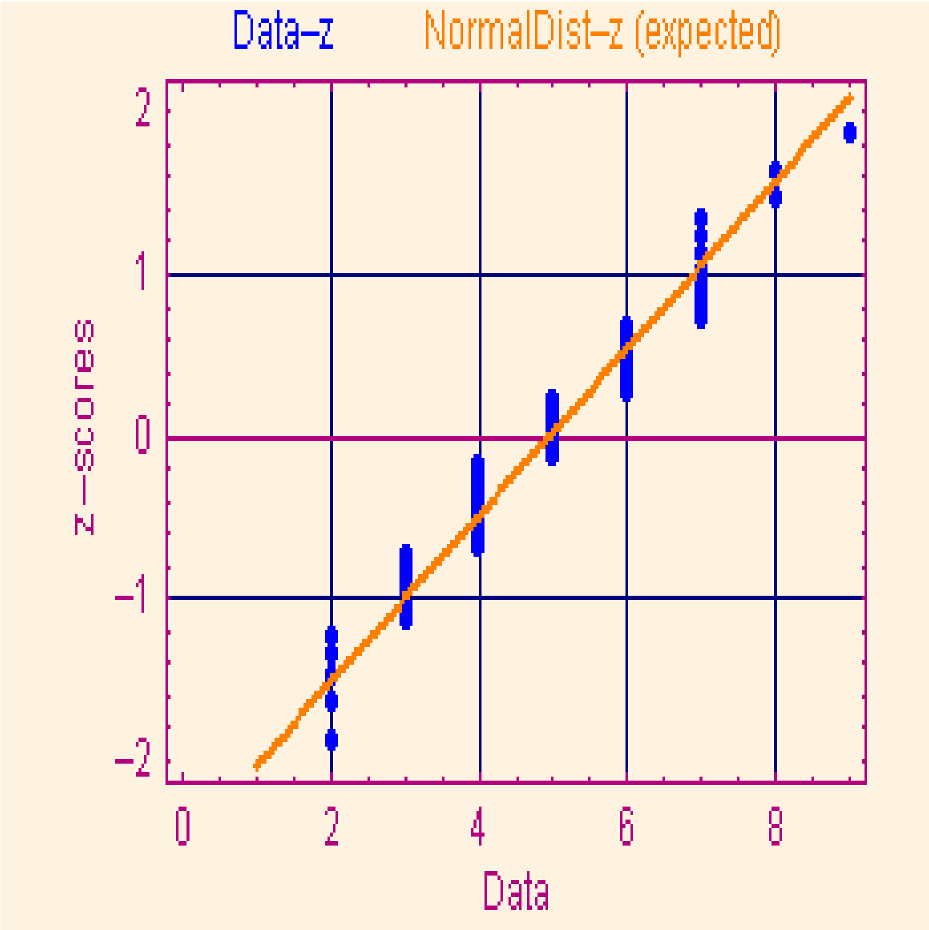


Input ▹ `MDSCheckStandardDeviation[ data ];`

s – Interval Report

Interval	Percentage of Measurement	Expected
$\bar{x} \pm s$	64. %	68 %
$\bar{x} \pm 2\,s$	94. %	95 %
$\bar{x} \pm 3\,s$	100. %	99.7 %

Input ▹ `MDSNormalProbabilityPlot[ data ];`




```
μ = □; σ = □;
sampleSize = 500;
```

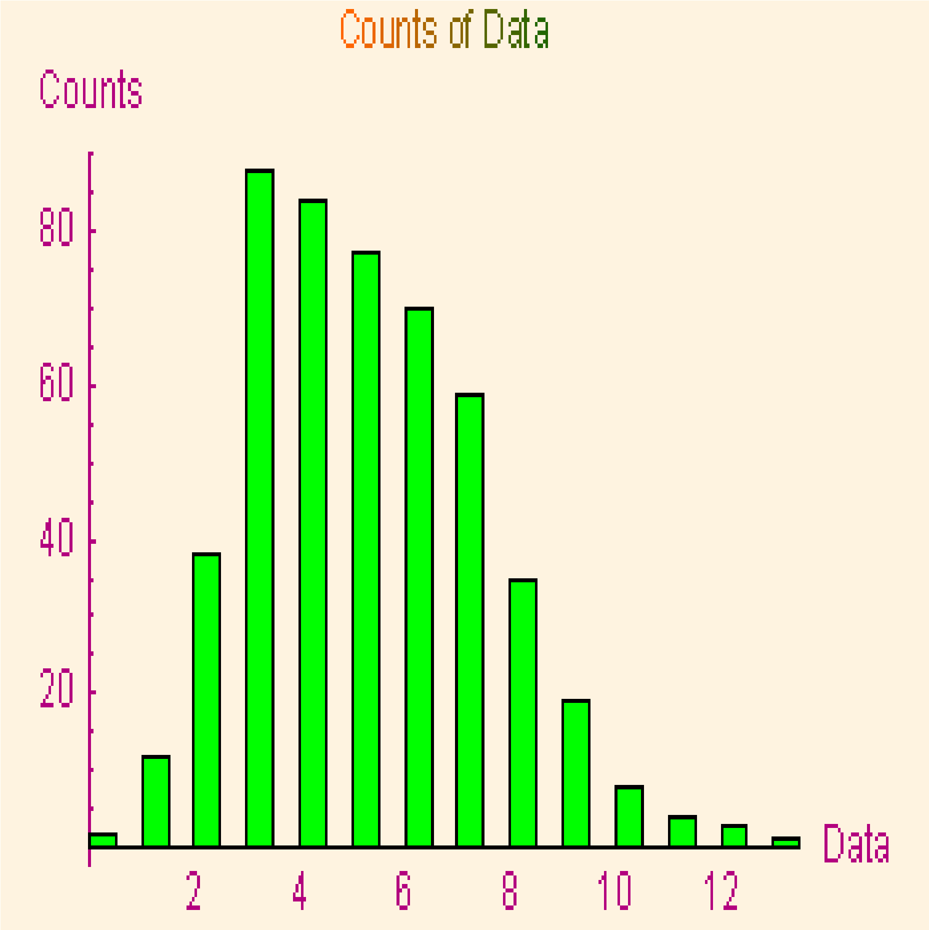
Input ▹

```
data = Table[Random[PoissonDistribution[5]],
        {sampleSize}];
Short[data, 4](* displays only some lines *)
```

```
{2, 2, 5, 6, 4, 4, 8, 6, 8, 4, 4, 1, 4, 9, 3, 3, 5, 8, 2, 5, 9,
 5, 3, 5, 4, 4, 5, 6, 3, 8, 1, 9, 6, 5, 3, 8, 3, 5, 3, 6, 5, 4,
 3, 1, 5, 6, 12, 2, 6, 3, ≪400≫, 7, 5, 4, 9, 3, 3, 5, 8, 8,
 3, 7, 6, 4, 5, 3, 10, 7, 4, 8, 2, 3, 3, 3, 7, 3, 5, 5, 7, 3,
 5, 3, 9, 4, 6, 2, 6, 6, 3, 9, 9, 6, 7, 5, 3, 10, 3, 5, 6, 5, 5}
```

Input ▹

```
MDSHistogram[ data ] ;  (* data format like {3,-2.2,1/2,...} *)
```

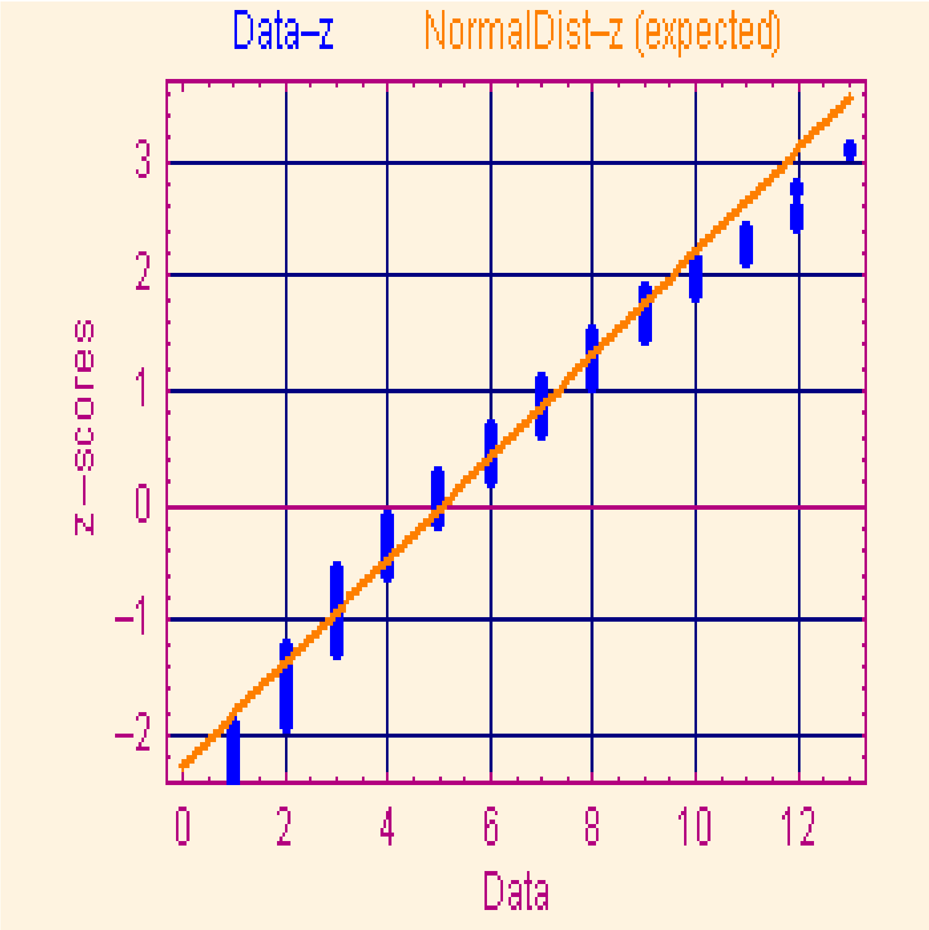


Input ▹ `MDSCheckStandardDeviation[ data ];`

s – Interval Report

Interval	Percentage of Measurement	Expected
$\bar{x} \pm s$	75.6 %	68 %
$\bar{x} \pm 2\,s$	96.4 %	95 %
$\bar{x} \pm 3\,s$	99.2 %	99.7 %

Input ▹ `MDSNormalProbabilityPlot[ data ];`



Antwort:
Um eine Poissonverteilung durchzuführen haben wir nichts anderes gemacht als den Teil des Inputs "NormalDistribution[μ,σ]" in PoissonDistribution[5] umgewandelt.

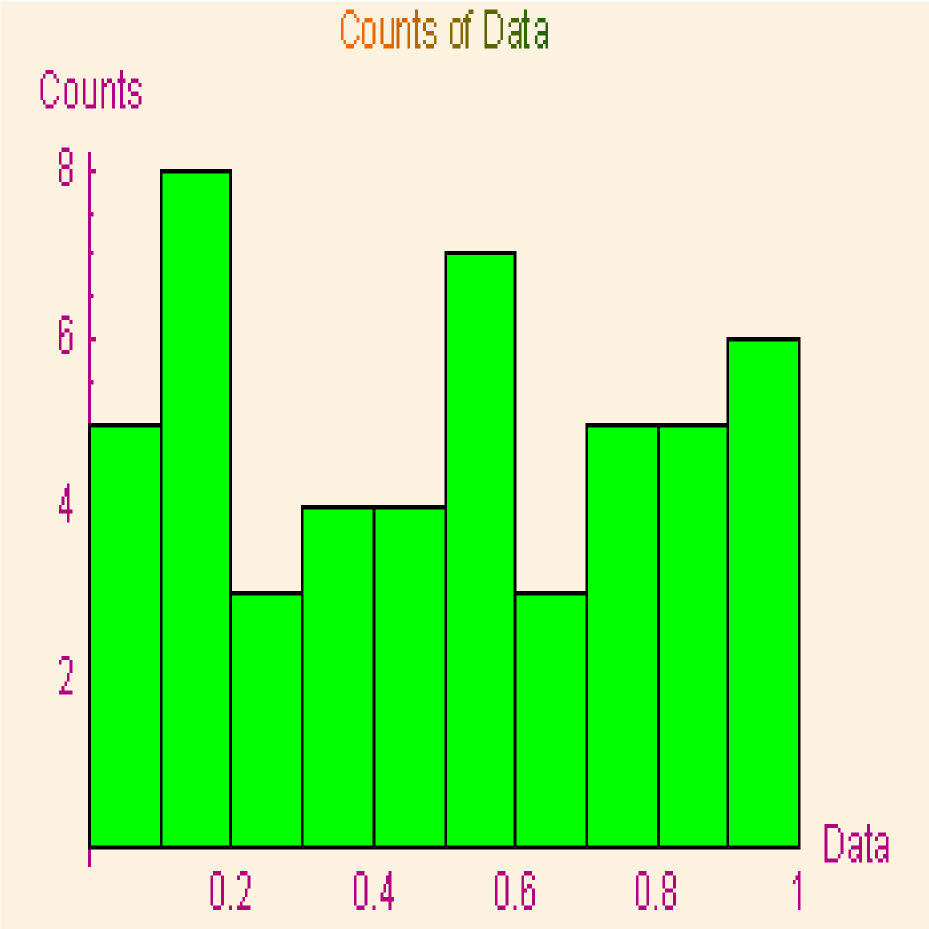
(b) Change the input of the μ,σ Rand Sample button in order to generate uniform distributed data in the interval in the range 0 to 1.

Click the μ,σ Rand Sample button. Simply delete NormalDistribution[μ,σ] in the input.
Change sampleSize to 50 resp 500.

Input ▹
```
μ = □; σ = □;
sampleSize = 50;

data = Table[Random[],
        {sampleSize}];
Short[data, 1](* displays only some lines *)
```

```
{0.779683, 0.998309, ≪47≫, 0.58299}
```

Input ▷ `MDSHistogram[ data ] ;   (* data format like {3,-2.2,1/2,...} *)`

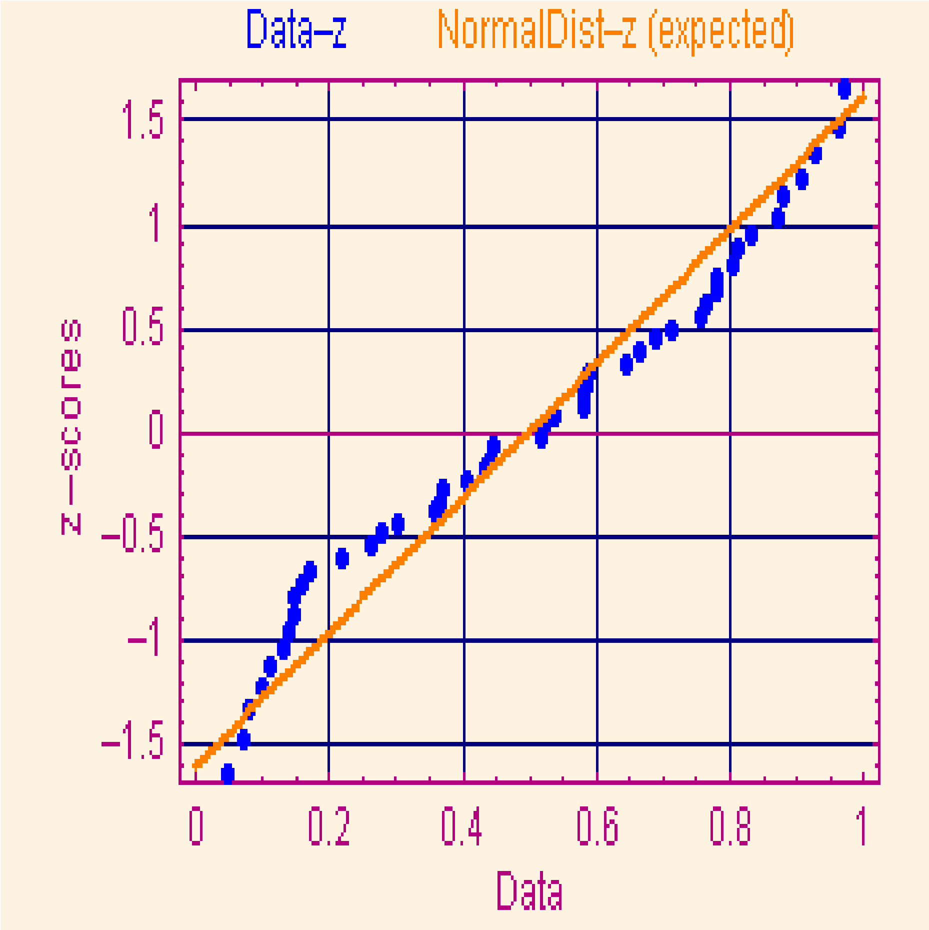


Input ▷ `MDSCheckStandardDeviation[ data ];`

s – Interval Report

Interval	Percentage of Measurement	Expected
$\bar{x} \pm s$	54. %	68 %
$\bar{x} \pm 2\,s$	100. %	95 %
$\bar{x} \pm 3\,s$	100. %	99.7 %

Input ▷ `MDSNormalProbabilityPlot[ data ];`




```
μ = □ ; σ = □ ;
sampleSize = 500;
```

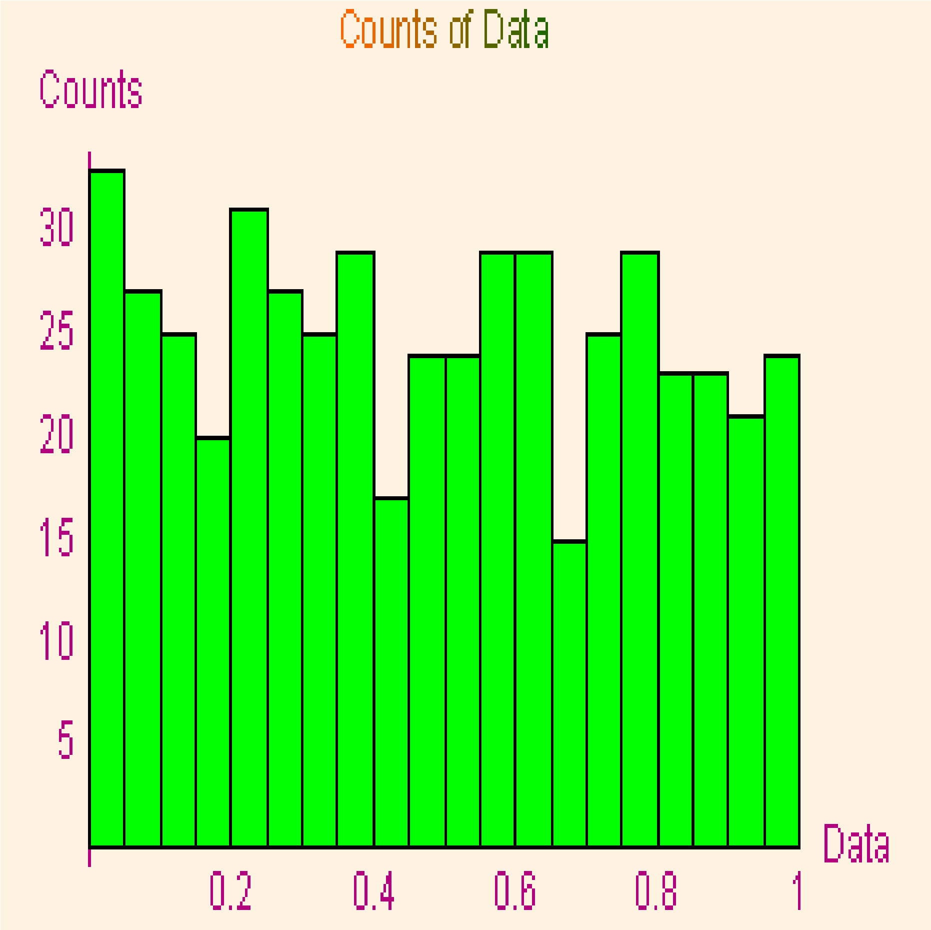
Input ▹

```
data = Table[Random[],
        {sampleSize}];
Short[data, 1] (* displays only some lines *)
```

```
{0.785887, ≪498≫, 0.0690215}
```

Input ▹

```
MDSHistogram[ data ] ;  (* data format like {3,-2.2,1/2,...} *)
```

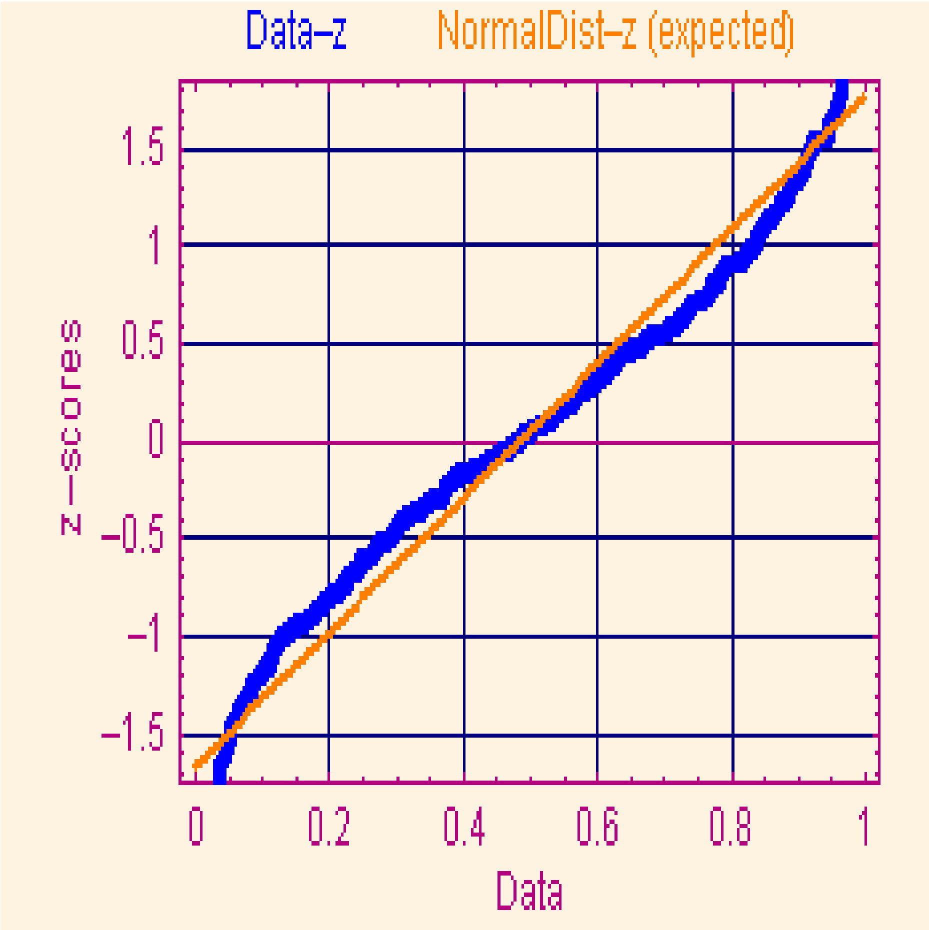


Input ▹ `MDSCheckStandardDeviation[ data ];`

s – Interval Report

Interval	Percentage of Measurement	Expected
x̄ ± s	58.6 %	68 %
x̄ ± 2 s	100. %	95 %
x̄ ± 3 s	100. %	99.7 %

Input ▹ `MDSNormalProbabilityPlot[ data ];`



Antwort:
Wir können erkennen, egal ob die sampleSize 50 oder 500 beträgt, die Daten nicht normalverteilt sind, da sie sehr stark abweichen.

Den 2. Teil der Stunde haben wir noch genutzt um unser Projekt zu besprechen. Wir haben eine eigene Projektvorlage erstellt, in die die Projektberichte, Auswertungen, Zusammenfassungen etc. hineinkommen.

Was ist wichtig für das Projekt?

- auch Standardbücher vom Gymnasium verwenden (Theorie usw.)
- Math-Desktop verwenden
- Summary --> reiner Text (Gelerntes beschreiben, etwaige Erweiterungen des Projektes, Seite der WU anschauen)
- Projekt mit Schiene + Hülle versehen, gutes Format wählen

- bei der Präsentation nur Punkt 1 - 3 von der Projektvorlage vorstellen

INTERNET Print **Open / Close**