STOCHASTIK

Die Normalverteilung

Hartmut Meyer
https://mathemeyer.com

Inhalt

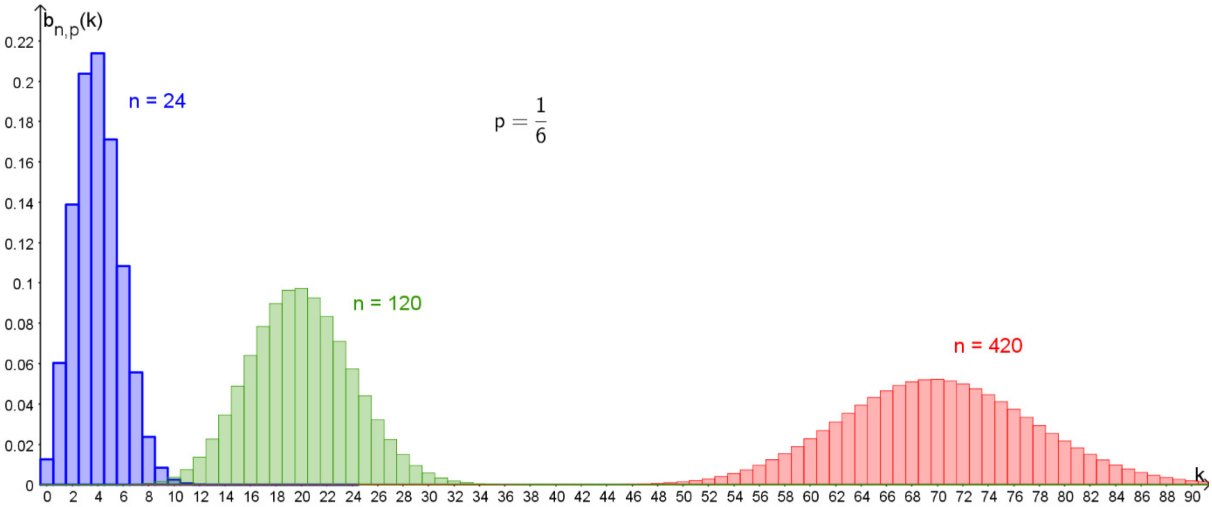
Von der Binomialverteilung zur Normalverteilung

Eigenschaften der Binomialverteilung

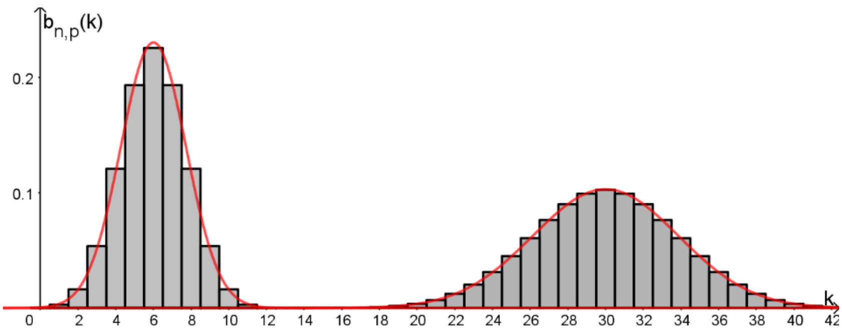
Betrachtet man die zu einer Binomialverteilung mit der festen Treffer-wahrscheinlichkeit p gehörigen Histogramme, so fällt auf, dass sie mit zunehmenden Stichprobenumfang n

- breiter
- flacher
- symmetrischer zum Erwartungswert μ

werden:

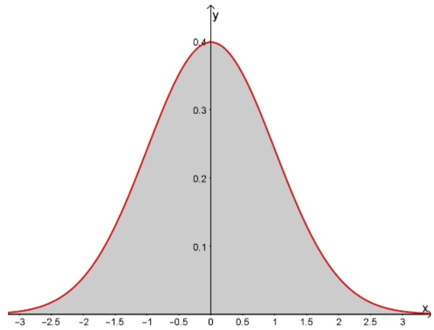


Gemeinsam aber haben alle diese Histogramme, dass eine sogenannte **Glockenkurve** entsteht, wenn man die Rechteckmitten verbindet:



Diese Beobachtungen gehen ursprünglich auf die beiden französischen Mathematiker Abraham de Moivre (1667 – 1754) und Pierre Simon Laplace (1749 – 1827) zurück.

Die Gauss'sche Dichtefunktion



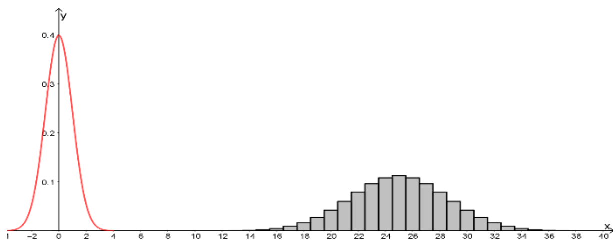
Im Jahre 1794 fand der deutsche Mathematiker

Carl Friedrich Gauss (1777 – 1855) in einem anderen Zusammenhang die Funktionsgleichung einer solchen Glockenkurve: $\varphi(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}$.

Ihr Graph ist achsensymmetrisch zur y-Achse, außerdem gilt, dass die Fläche zwischen dem Graphen und der x-Achse den Inhalt 1 hat: $\int_{-\infty}^{+\infty} \varphi(x)\, dx = 1$

Approximation der Binomialverteilung durch die Gauss-Kurve

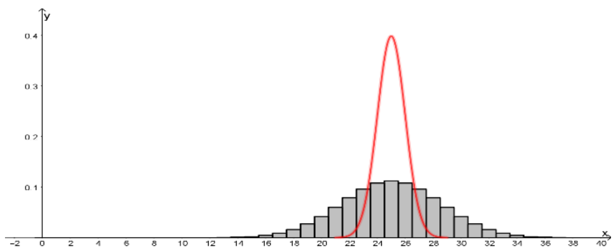
Zeichnet man das Histogramm einer Binomialverteilung mit den Parametern n und p sowie die Gauss-Kurve mit der Gleichung $\varphi(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}$ in ein Koordinatensystem, so ergibt sich zunächst folgendes Bild:



Die Kurve ist symmetrisch zu $x = 0$, das Histogramm hingegen zum Erwartungswert μ, also zur Geraden $x = \mu$. Daher muss der Graph von φ um μ Einheiten nach rechts verschoben werden.

Die Kurve hat dann die Funktionsgleichung

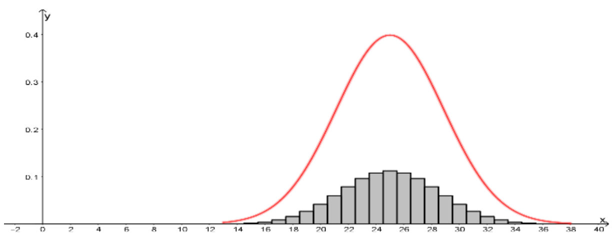
$$\varphi(x - \mu) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2}}$$



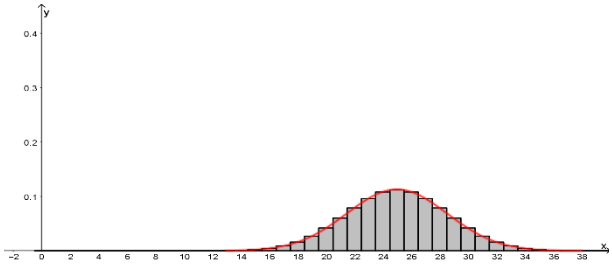
Die Kurve ist noch zu schmal. Damit sie die richtige Breite bekommt, muss sie mit dem Faktor σ (die Standardabweichung) in x-Richtung gestreckt werden.

Die Kurve hat dann die Funktionsgleichung

$$\varphi\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



Lage und Breite des Histogramms und der Gauss-Kurve sind nun ähnlich. Staucht man nun noch die Höhe der Kurve mit dem Faktor σ, so passt sie genau.



Die Kurve hat schließlich die Funktionsgleichung

$$\frac{1}{\sigma} \cdot \varphi\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Berechnung von Wahrscheinlichkeiten mit der Normalverteilung

Die auf diese Weise modifizierte Gauss-Funktion kann somit – unter der Bedingung, dass der Stichprobenumfang n hinreichend groß ist – als gute Näherung für die Binomialverteilung genutzt werden.

Statt $\frac{1}{\sigma} \cdot \varphi\left(\frac{x-\mu}{\sigma}\right)$ wird üblicherweise die Abkürzung $\boldsymbol{\varphi_{\mu,\sigma}(x)}$ verwendet.

Es gilt also:

Eine Binomialverteilung mit dem Erwartungswert μ und der Standardabweichung σ lässt sich für große n näherungsweise durch eine Glockenkurve beschreiben:

$$\boldsymbol{b_{n;p}(k) \approx \varphi_{\mu,\sigma}(k) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}}$$

Als Voraussetzung muss die sogenannte **Laplace-Bedingung**: $\boldsymbol{\sigma > 3}$ erfüllt sein.

Beispiel 1: *Die Zufallsgröße X mit den Parametern $n = 200$ und $p = 0{,}5$ hat den Erwartungswert $\mu = 200 \cdot 0{,}5 = 100$ sowie die Standardabweichung $\sigma = \sqrt{200 \cdot 0{,}5 \cdot (1 - 0{,}5)} = \sqrt{50}$.*
Die Laplace-Bedingung ist erfüllt: $\sigma = \sqrt{50} > 3$, somit kann die Normalverteilung als Näherung der Binomialverteilung genutzt werden:

Dann beträgt die Wahrscheinlichkeit für z.B. genau 95 Treffer:

$$P(X = 95) \approx \varphi_{100;\sqrt{50}}(95) = \frac{1}{\sqrt{50} \cdot \sqrt{2\pi}} \cdot e^{-\frac{(95-100)^2}{2\cdot(\sqrt{50})^2}} \approx 0{,}0439$$

Das Beispiel 1 zeigt, wie gut die Näherung durch die Normalverteilung ist. Berechnet man die Wahrscheinlichkeit $P(X = 95)$ exakt mit der Binomialverteilung, so gilt: $b_{200;0,5}(95) \approx 0{,}0439$.

Ein Unterschied wird erst erkennbar, wenn man die gesamte Taschenrechner-anzeige betrachtet:

$\varphi_{100;\,\sqrt{50}}(95) \approx 0{,}04393912895$ und $b_{200;0,5}(95) \approx 0{,}04393443286$.

Die Normalverteilung ist um so exakter je größer der Stichprobenumfang n ist und je näher die Trefferwahrscheinlichkeit p an dem Wert 0,5 ist.

Die Funktion $\varphi_{\mu;\sigma}$ ist unter dem Namen *„Normal-Dichte"* in unserem Taschenrechner gespeichert. Man erreicht die Funktion über die Tastenfolge „MENU" „7" „1".

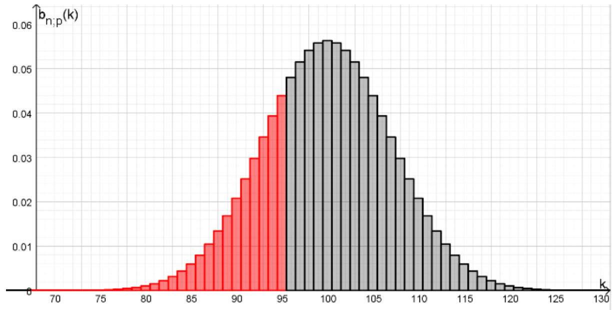
Nach Eingabe von z.B. $X : 95$, $\sigma : \sqrt{50}$, $\mu : 100$ lautet das Ergebnis $0{,}04393912895$.
(vergl. Beispiel 1)

Die Berechnung von Wahrscheinlichkeiten mit der $\varphi_{\mu;\sigma}$ – Funktion nennt man auch die **Methode von Moivre und Laplace**.

Die kumulierte Normalverteilung

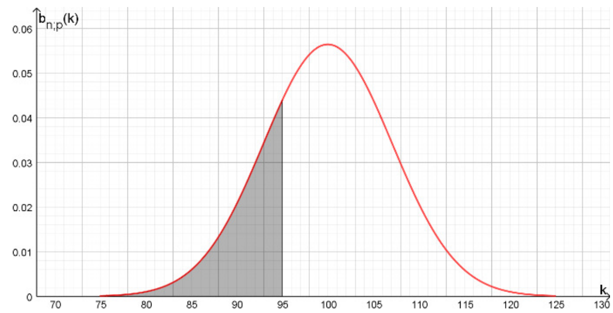
Wie auch bei der Binomialverteilung ist es bei der Praxis der Normalverteilung relevanter, sogenannte Intervallwahrscheinlichkeiten zu berechnen.
Für Beispiel 1 könnte z.B. die Frage lauten:
„Mit welcher Wahrscheinlichkeit wird man bei $n = 200$ Versuchen und der Trefferwahrscheinlichkeit $p = 0{,}5$ ***höchstens 95 Treffer*** *erzielen."*

Um solch eine Intervallwahrscheinlichkeit $(k \leq 95)$ zu berechnen, müsste man entweder mit der Formel von Bernoulli oder der Methode von Moivre und Laplace die Summe aller Einzelwahrscheinlichkeiten $P(X = 0) + P(X = 1) + \cdots + P(X = 95)$ berechnen.



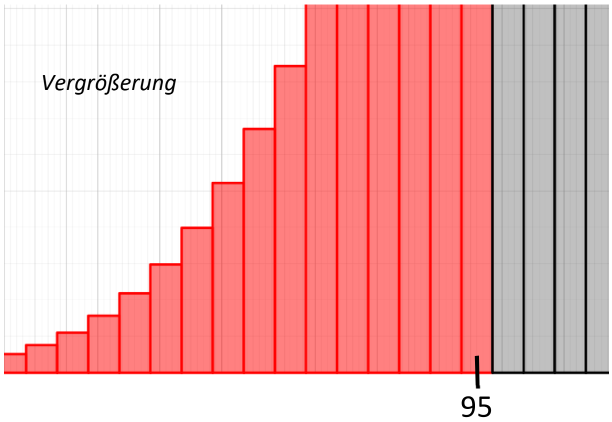
Beide Verfahren sind doch sehr aufwendig und wenig praktikabel.

Einfacher ist es hingegen, die entsprechende Fläche unter der zugehörigen Glockenkurve mit Hilfe der Integralrechnung zu berechnen. Da die φ-Funktion nicht elementar zu integrieren ist, liegen die Werte in Form einer Tabelle (auf dem Taschenrechner) vor.
Die Stammfunktion von φ wird mit Φ bezeichnet.



$$\Phi(k) = \int_{-\infty}^{k} \varphi_{\mu,\sigma}(x)\, dx$$

Zu beachten ist hierbei, dass die Rechtecke der Histogramme bei der Binomialverteilung symmetrisch zur Stelle k eingezeichnet sind. Um auch den letzten Balken mitzurechnen, muss also bis zu dem x-Wert $k + 0{,}5 = 95{,}5$ integriert werden.



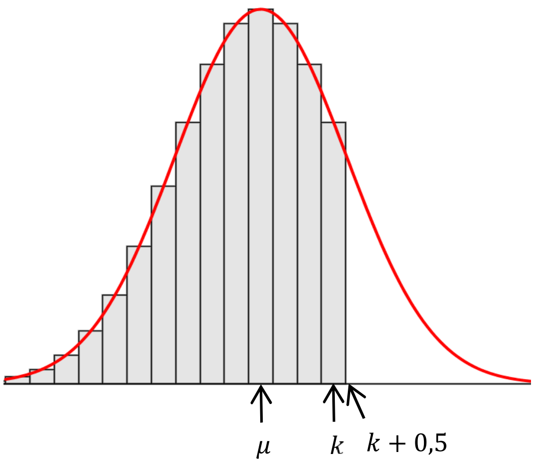
Es ist hier:

$$P(X \leq 95) \approx \Phi\left(\frac{95 + 0{,}5 - 100}{\sqrt{50}}\right) = \Phi_{100;\,\sqrt{50}}(95{,}5) \approx 0{,}26226$$

(Zum Vergleich: Die exakte Berechnung ergibt $F_{200;0,5}(95) \approx 0{,}26231$)

Allgemein gilt:

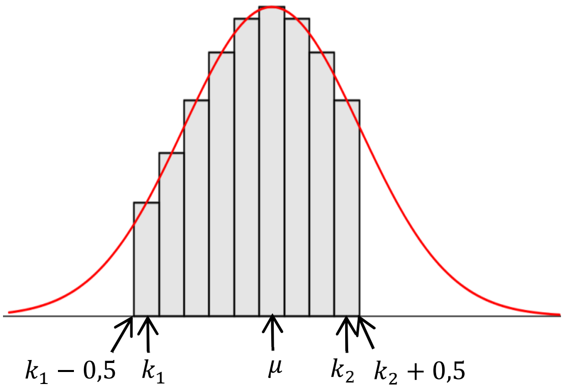
Für eine binomialverteilte Zufallsgröße X mit den Parametern n und p und dem Erwartungswert μ und der Standardabweichung σ, lässt sich die Wahrscheinlichkeit für **höchstens k Treffer** näherungsweise durch die **Formel von MOIVRE-LAPLACE** berechnen:



$$\boldsymbol{P(X \leq k) \approx \Phi\left(\frac{k+0,5-\mu}{\sigma}\right) \approx \Phi_{\mu;\,\sigma}(k+0,5)}$$

Diese Formel lässt sich leicht auf allgemeine Intervalle erweitern:

Die Wahrscheinlichkeit für **mindestens k_1 aber höchstens k_2 Treffer** lässt sich näherungsweise berechnen durch :



$$\boldsymbol{P(k_1 \leq X \leq k_2)}$$
$$\boldsymbol{\approx \Phi\left(\frac{k_2+0,5-\mu}{\sigma}\right) - \Phi\left(\frac{k_1-0,5-\mu}{\sigma}\right)}$$
$$\boldsymbol{\approx \Phi_{\mu;\,\sigma}(k_2+0,5) - \Phi_{\mu;\,\sigma}(k_1-0,5)}$$

Lautet bei Beispiel 1 z.B. die Frage:

*„Mit welcher Wahrscheinlichkeit wird man bei $n = 200$ Versuchen und der Trefferwahrscheinlichkeit $p = 0{,}5$ **mindestens 80 aber höchstens 90 Treffer** erzielen.“*

So ergibt die Formel von MOIVRE-LAPLACE:

$$P(80 \leq X \leq 90) \approx \Phi\left(\frac{90+0{,}5-100}{\sqrt{50}}\right) - \Phi\left(\frac{80-0{,}5-100}{\sqrt{50}}\right)$$
$$\approx \Phi_{100;\,\sqrt{50}}(90{,}5) - \Phi_{100;\,\sqrt{50}}(79{,}5) \approx 0{,}08768$$

(Zum Vergleich: $F_{200;0{,}5}(90) - F_{200;0{,}5}(79) \approx 0{,}08766$)

Auch die Funktion $\Phi_{\mu;\sigma}$ ist unter dem Namen *„Kumul.-Normal-V.“* in unserem Taschenrechner gespeichert. Man erreicht die Funktion über die Tastenfolge „MENU“ „7“ „2“.

Nach Eingabe von z.B.

$$\begin{array}{ll} Untere & : 79{,}5 \\ Obere & : 90{,}5 \\ \sigma & : \sqrt{50} \\ \mu & : 100 \end{array}$$

lautet das Ergebnis $0{,}08768364428$ *(vergl. Seite 7)*

Beispiel 2: *In Deutschland haben ca. 11% der Menschen die Blutgruppe B. Es wird eine repräsentative Stichprobe von 500 Personen betrachtet.*

a) *Mit welcher Wahrscheinlichkeit befinden sich in der Stichprobe genau 60 Personen mit der Blutgruppe B?*

b) *Mit welcher Wahrscheinlichkeit befinden sich in der Stichprobe zwischen 50 und 60 Personen mit der Blutgruppe B?*

c) *Mit welcher Wahrscheinlichkeit befinden sich in der Stichprobe mehr als 65 Personen mit der Blutgruppe B?*

Lösungen:

Die Zufallsgröße X steht für die Menschen mit der Blutgruppe B. X ist binomialverteilt mit $n = 500$ und $p = 0{,}11 \Rightarrow \mu = 55$ und $\sigma = \sqrt{48{,}95}$.

Da die Laplace-Bedingung [$\sigma > 3$] erfüllt ist, kann die Normalverteilung als Näherung der Binomialverteilung genutzt werden.

a) $P(X = 60) \approx \varphi_{55;\sqrt{48{,}95}}(60) \approx 0{,}0442 \approx 4{,}4\%$

b) $$P(50 \leq X \leq 60) \approx \Phi_{55;\sqrt{48{,}95}}(60{,}5) - \Phi_{55;\sqrt{48{,}95}}(49{,}5) \approx 0{,}5682 \approx 56{,}8\%$$

c) $$P(X > 65) = 1 - P(X \leq 65) \approx 1 - \Phi_{55;\sqrt{48{,}95}}(65{,}5) \approx 1 - 0{,}9333 \approx 0{,}0667 \approx 6{,}7\%$$

Die Umkehrung der Normalverteilung

Im Gegensatz zur kumulierten Binomialverteilung (also zur Funktion $F_{n;p}$) besitzt die Normalverteilung (also die Integralfunktion $\Phi_{\mu;\,\sigma}$) eine Umkehrfunktion.

So ergibt z.B. $\Phi^{-1}{}_{100;\,\sqrt{50}}(0{,}26226) \approx 95{,}500$ *(vergl. Seite 6)*

Auch die Funktion $\Phi^{-1}{}_{\mu;\sigma}$ ist unter dem Namen *„Inv.-Normal-V."* in unserem Taschenrechner gespeichert. Man erreicht die Funktion über die Tastenfolge „MENU" „7" „3".

Nach Eingabe von z.B.	$Fläche : 0{,}26226$ $\sigma : \sqrt{50}$ $\mu : 100$	lautet das Ergebnis 95,50088774

Beispiel 3: *Ein Laplace-Würfel wird 300-mal geworfen. Hierbei wird gezählt, wie oft in einer solchen Serie eine Sechs gewürfelt wird.*

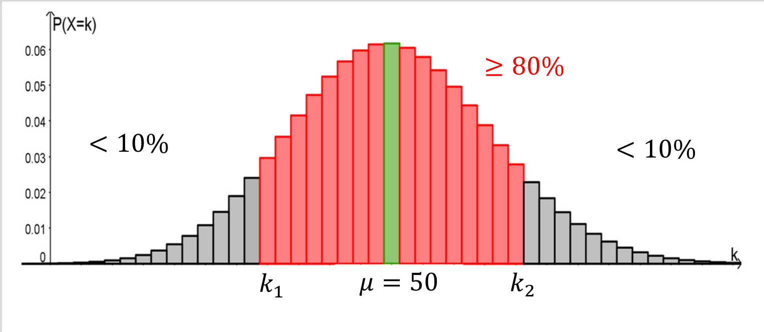
Gesucht ist der Bereich für die Anzahl der Sechsen um den Erwartungswert $\mu = 50$ *für den die Wahrscheinlichkeit mindestens 80% beträgt.*

Lösung:

Die Zufallsgröße X steht für die Anzahl der Sechsen in einer 300er-Serie.

X ist binomialverteilt mit $n = 300$ und $p = \frac{1}{6} \Rightarrow \mu = 50$ und $\sigma = \sqrt{\frac{125}{3}}$.

Da die Laplace-Bedingung $[\sigma > 3]$ erfüllt ist, kann die Normalverteilung als Näherung der Binomialverteilung genutzt werden.



Es soll gelten: $P(k_1 \leq X \leq k_2) \geq 0{,}8$ *also:*

$P(X \leq k_1 - 1) < 0{,}1$

$\Leftrightarrow \Phi_{50;\sqrt{\frac{125}{3}}}(k_1 - 1 + 0{,}5) < 0{,}1$

$\Leftrightarrow k_1 - 1 + 0{,}5 < \Phi^{-1}_{50;\sqrt{\frac{125}{3}}}(0{,}1)$

$\Leftrightarrow k_1 - 0{,}5 < 41{,}73$

$\Leftrightarrow k_1 < 42{,}23$

$\Rightarrow k_1 = 42$

$P(X \leq k_2) \geq 0{,}9$

$\Leftrightarrow \Phi_{50;\sqrt{\frac{125}{3}}}(k_2 + 0{,}5) \geq 0{,}9$

$\Leftrightarrow k_2 + 0{,}5 \geq \Phi^{-1}_{50;\sqrt{\frac{125}{3}}}(0{,}9)$

$\Leftrightarrow k_2 + 0{,}5 \geq 58{,}27$

$\Leftrightarrow k_2 \geq 57{,}77$

$\Rightarrow k_2 = 58$

Mit einer Wahrscheinlichkeit von mindestens 80% werden bei einer Serie von 300 Würfen mit dem Würfel zwischen 42 und 58 Sechsen gewürfelt.

Dies bedeutet im Umkehrschluss, dass die Wahrscheinlichkeit für höchstens 41 und mindestens 59 Sechsen bei einer solchen Serie zusammengenommen weniger als 20% beträgt.