

UNIVERSITÄT LEIPZIG

Medizinische Fakultät

QSB 1: Epidemiologie, Medizinische Biometrie und Medizinische Informatik

Übung 3
Deskriptive Statistik
Verteilungen

Wintersemester 2026/27

imise.

Lernziele

Die Studierenden sollen

- die Anwendung der wichtigsten Grafiken zur Darstellung von Verteilungen kennen: Histogramm, empirische Verteilungsfunktion, Boxplot, Balkendiagramm,
- den Begriff Referenzbereich kennen,
- aus Verteilungskurven Wahrscheinlichkeiten ablesen und
- die standardisierte Normalverteilung kennen und anwenden können.

Probandencharakteristik (1)

am Beispiel der Herzsportstudie

Beschreibung der Probandeneigenschaften vor der Intervention („Baseline – Characteristics“)

- getrennt nach Studienarm:
  - Alter, Geschlecht, Schwangerschaft, sportliche Betätigung, Raucherstatus, Erkrankungen, Medikamenteneinnahme
- absolute / relative Häufigkeiten
- arithmetisches Mittel und Standardabweichung
- Median / Minimum / Maximum
- geeignete tabellarische oder grafische Darstellung

Probandencharakteristik (2)

Kategoriale Daten: z. B. Armverteilung

Studienarm

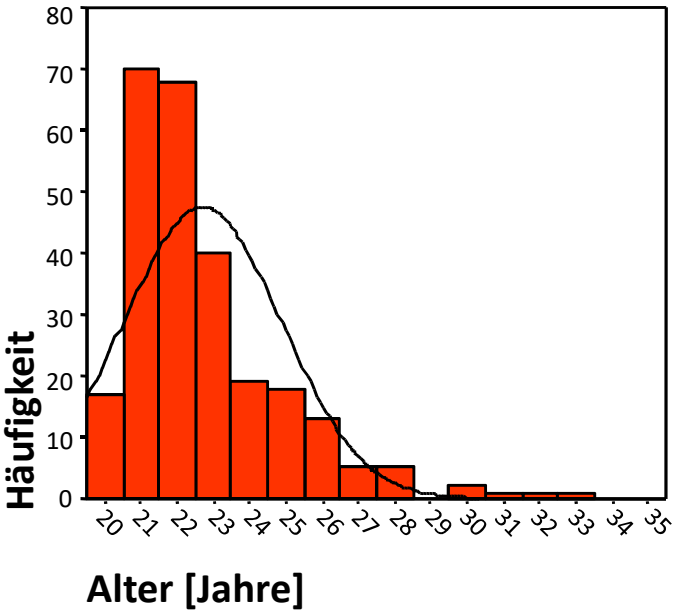
	Häufigkeit	Prozent
Arm A (5 Kniebeugen)	129	49,4
Arm B (30 Kniebeugen)	132	50,6
Gesamt	261	100,0

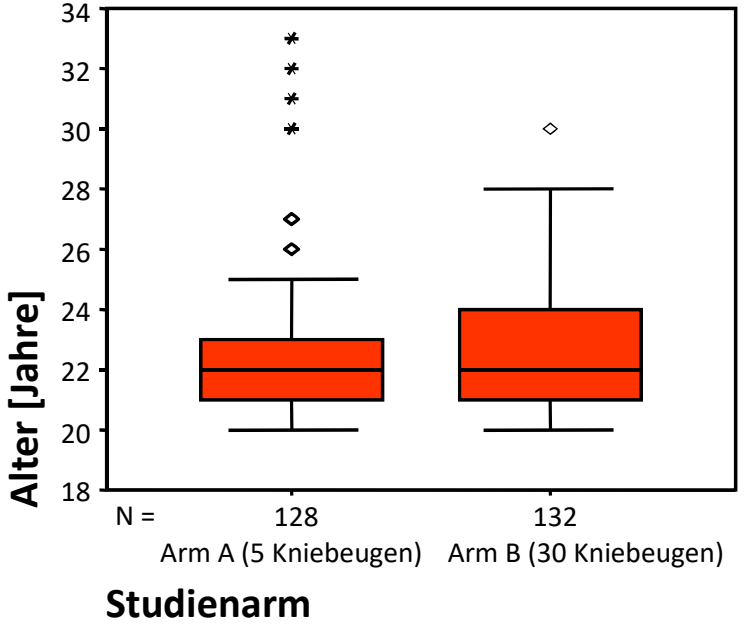
Standard:

- absolute Häufigkeiten
- relative Häufigkeiten (Prozent)

Probandencharakteristik (3)

Metrische Daten: z. B. Alter

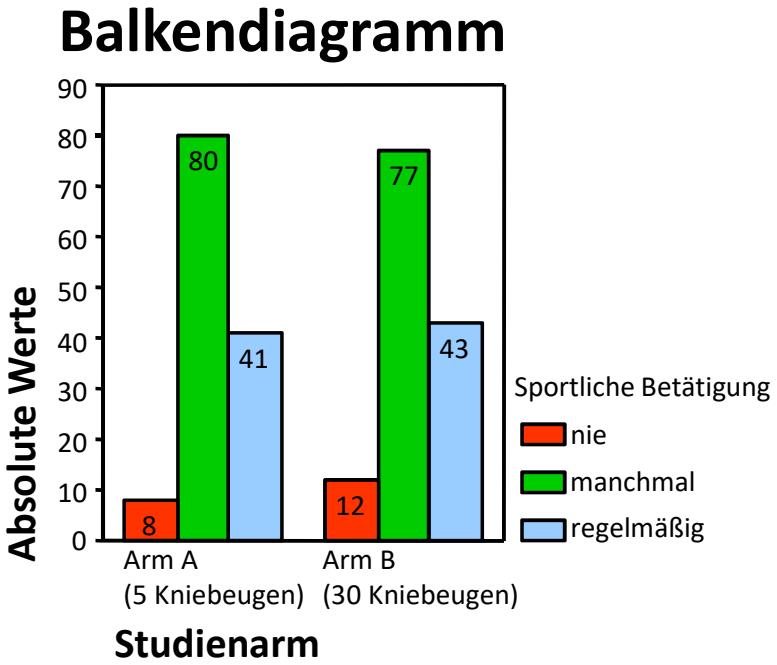


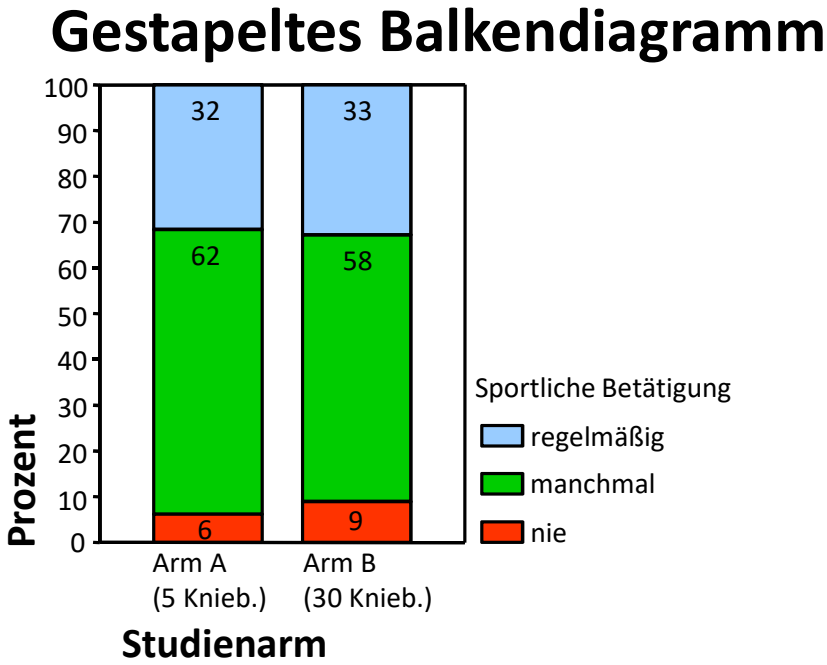


		Studienarm		
		Arm A (5 Knieb.)	Arm B (30 Knieb.)	Gesamt
Alter [Jahre]	Mittelwert	23	23	23
	Standardabweichung	2	2	2
	Median	22	22	22
	Minimum	20	20	20
	Maximum	33	30	33

Probandencharakteristik (5)

Ordinale Daten: Sportliche Betätigung





		Studienarm			
		Arm A (5 Knieb.)		Arm B (30 Knieb.)	
		Anzahl	Spalten%	Anzahl	Spalten%
Sportliche Betätigung	nie	8	6,2%	12	9,1%
	manchmal	80	62,0%	77	58,3%
	regelmäßig	41	31,8%	43	32,6%
Total		129	100,0%	132	100,0%

Probandencharakteristik (6)

Tabellarische Zusammenfassung

Charakteristik	Arm A (5 Kniebeugen) n=129	Arm B (30 Kniebeugen) n=132	Gesamt n=261
Alter (Jahre)			
Median (min – max)	22 (20 – 33)	22 (20 – 30)	22 (20 – 33)
Geschlecht			
männlich	46 (35,7 %)	48 (36,4 %)	94 (36,0 %)
weiblich	83 (64,3 %)	84 (63,6 %)	167 (64,0 %)
Schwangerschaft			
unbekannt	1 (1,2 %)	0	1 (0,6 %)
ja	0	0	0
Raucher	31 (24,0 %)	37 (28,0 %)	68 (26,1 %)
Sportliche Betätigung			
nie	8 (6,2 %)	12 (9,1 %)	20 (7,7 %)
manchmal	80 (62,0 %)	77 (58,3 %)	157 (60,2 %)
regelmäßig	41 (31,8 %)	43 (32,6 %)	84 (32,2 %)
Erkrankung Knie			
unbekannt	5 (3,9 %)	1 (0,8 %)	6 (2,3 %)
ja	4 (3,1 %)	4 (3,0 %)	8 (3,1 %)
Erkrankung Herz			
unbekannt	5 (3,9 %)	4 (3,0 %)	9 (3,4 %)
ja	2 (1,6 %)	3 (2,7 %)	5 (1,9 %)
Medikamente			
ja	0	1 (0,8 %)	1 (0,4 %)

Probandencharakteristik (7)

Literaturbeispiel: Perkovic et al. 2019

Tabelle 1 fast jeder medizinischen Publikation beschreibt die Kohorte.

Table 1. Demographic and Clinical Characteristics of the Patients at Baseline.*

Characteristic	Canagliflozin (N=2202)	Placebo (N=2199)	All Patients (N=4401)
Age — yr	62.9±9.2	63.2±9.2	63.0±9.2
Female sex — no. (%)	762 (34.6)	732 (33.3)	1494 (33.9)
Race or ethnic group — no. (%)†			
White	1487 (67.5)	1444 (65.7)	2931 (66.6)
Black	112 (5.1)	112 (5.1)	224 (5.1)
Asian	425 (19.3)	452 (20.6)	877 (19.9)
Other	178 (8.1)	191 (8.7)	369 (8.4)
Current smoker — no. (%)	341 (15.5)	298 (13.6)	639 (14.5)
Median urinary albumin-to-creatinine ratio (IQR)¶	923 (459–1794)	931 (473–1868)	927 (463–1833)

* Plus–minus values are means ±SD. Percentages may not total 100 because of rounding. IQR denotes interquartile range.

- Welche deskriptiven statistischen Kenngrößen erkennen Sie?
- Welche sind Lage-, welche Streuungsmaße?
- Zu welchen Datentypen passen sie?

Referenzbereiche (1)

Story

Mutter zum Kinderarzt:

„Mein Paul (8) war schon im Kindergarten der Kleinste, in der Schule ist er es wieder, und die Nachbarsjungen sind auch größer. Kann es sein, dass er eine Wachstumsstörung hat?“

Referenzbereiche (2)

Story - Fortsetzung

- Frage

  Wie viele gleichaltrige Jungen waren in der Kindergartengruppe von Paul, in seiner Schulklasse und in der Nachbarschaft?

- Antwort

Kindergartengruppe:	insgesamt 5 Jungen
Schulklasse:	insgesamt 7 Jungen
Nachbarschaft:	2 gleichaltrige Jungen

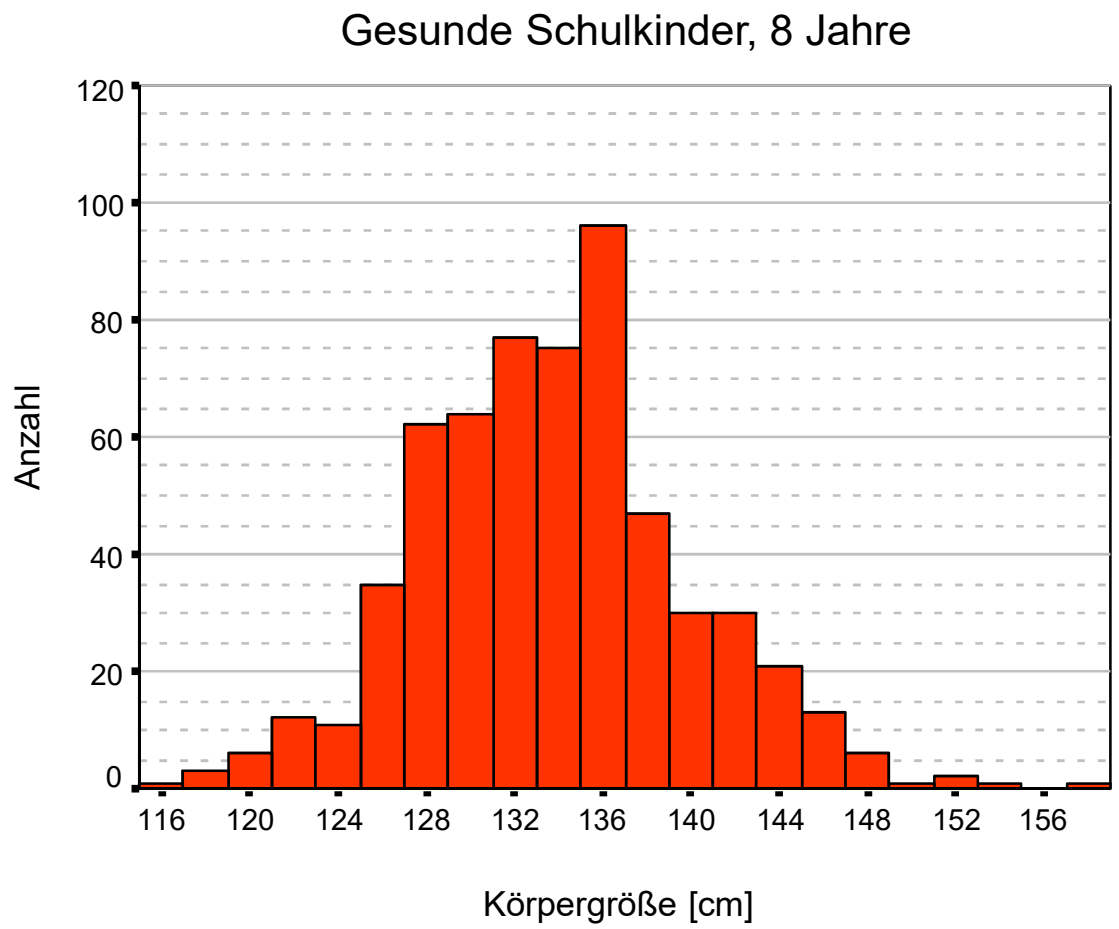
Referenzbereiche (3)

Überlegung

- Paul hat $(5-1)+(7-1)+2=12$ Jungen zum Vergleich.
- Die Wahrscheinlichkeit, unter 13 Jungen der Kleinste zu sein, beträgt … .

Referenzbereiche (5)

Körpergrößen von 594 Kindern



Paul ist
128 cm groß.

Wie viele
Kinder sind
kleiner als er?

Referenzbereiche (6)

Einordnung von Paul

- Die Histogramm-Säule bei 128 umfasst den Bereich von >127 bis ≤129.
- Die sechs Säulen links daneben repräsentieren

  _____,

  _____,

  _____,

  _____,

  _____ und

  _____ Kinder.

Referenzbereiche (8)

Fazit

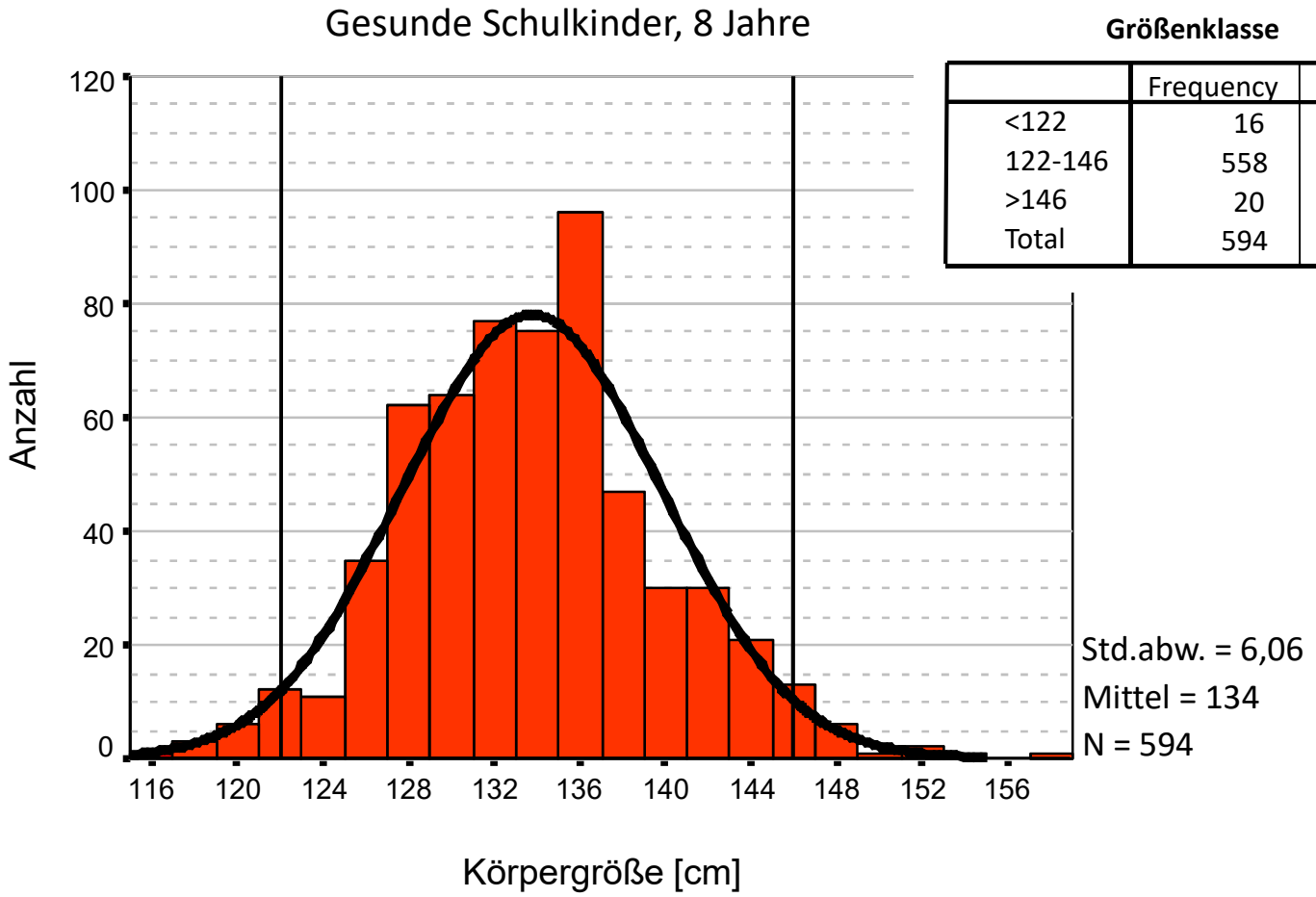
- „Normalität“ wird durch den Vergleich mit der gesamten Population definiert.
- Die „Mitte“ der Population sowie kleine Abweichungen davon gelten als normal.
- Je weiter ein Individuum „am Rande“ liegt, desto eher gilt es als auffällig.

Frage

Welche „Randlage“ begründet einen starken Verdacht? → Festlegung von Referenzbereichen

Referenzbereiche (9)

Körpergrößen

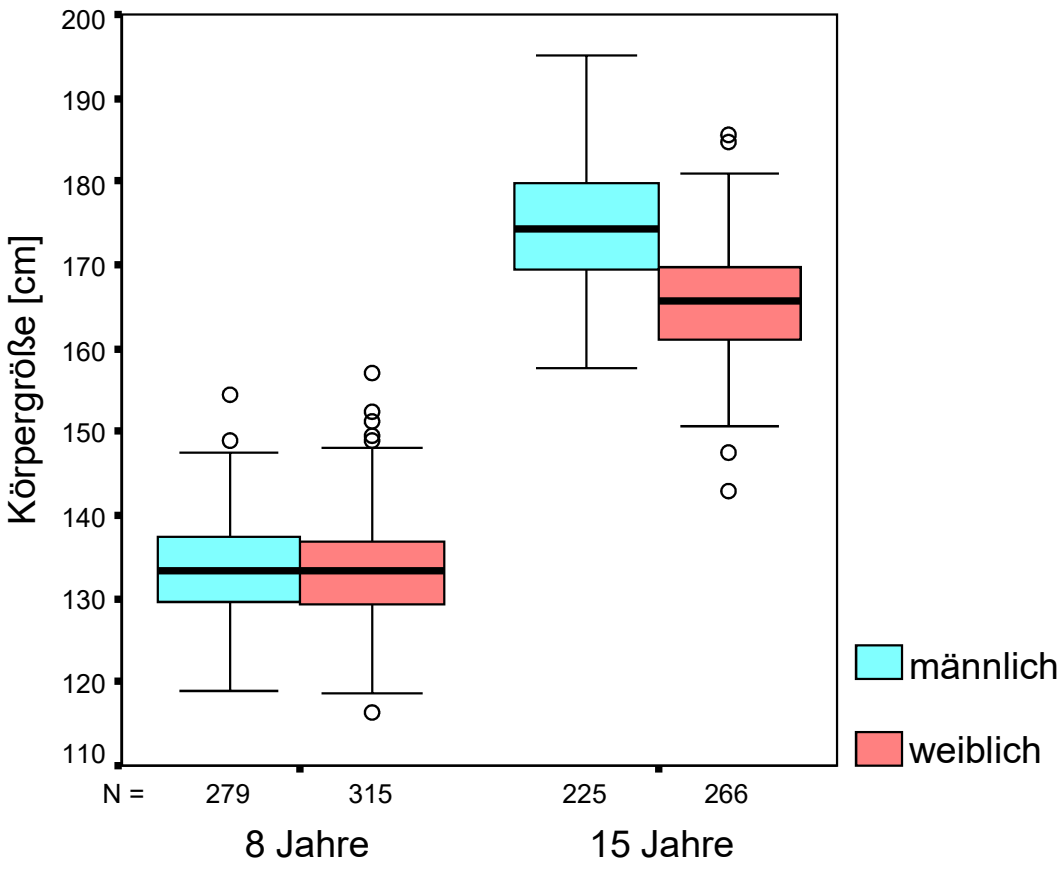


Größenklasse

	Frequency	Percent
<122	16	2,7
122-146	558	93,9
>146	20	3,4
Total	594	100,0

Referenzbereiche (10)

Alters- / Geschlechtsunterschiede



Boxplot:

- Box: I. / III. Quartil
- dicke Medianlinie
- „Antennen“ und Ausreißer:

Werte, die > 1,5 Boxhöhen von der Box entfernt liegen, sind Ausreißer und werden einzeln als o dargestellt

„Antennen“ bis zum letzten Wert zeichnen, der noch kein Ausreißer ist

vereinfacht: „Antennen“ bis Min./Max. zeichnen

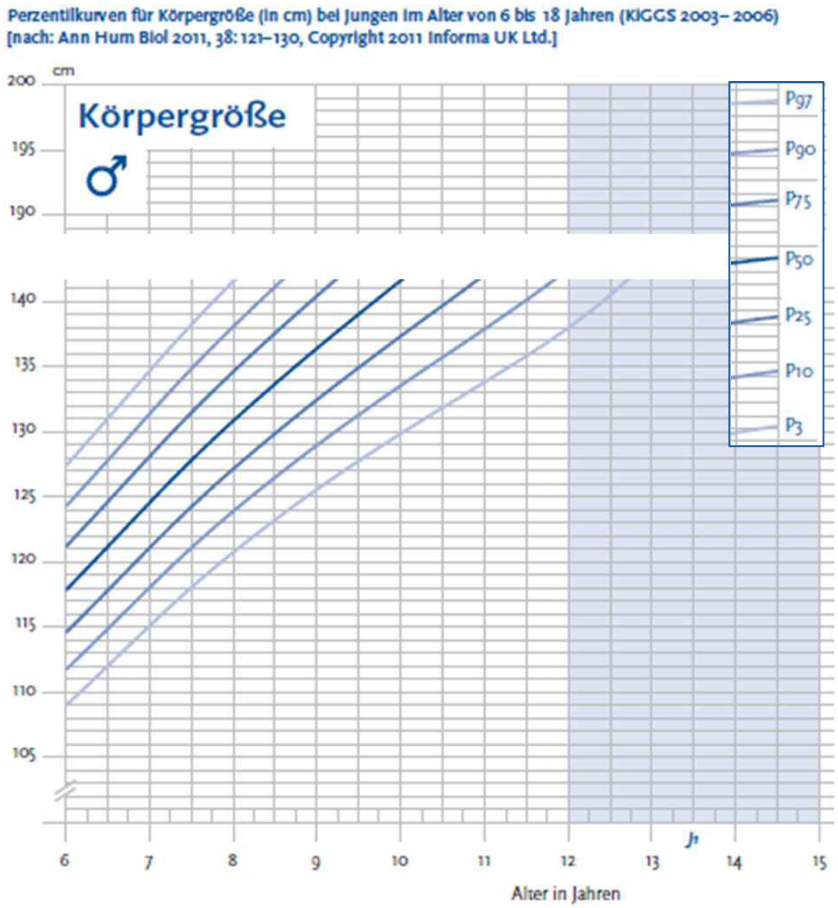
Referenzbereiche (12)

- Referenzbereich: innerhalb = unauffällig, außerhalb = auffällig (aber noch nicht: krank)
- Wähle für den Referenzbereich 2 Grenzen (untere und obere), sodass ein bestimmter Prozentsatz der Referenzstichprobe < untere Grenze und ein bestimmter Prozentsatz > obere Grenze
- Beispiel: untere Grenze = 3. Perzentil
  obere Grenze = 97. Perzentil

(häufig verwendet z.B. in der Anthropometrie)

Referenzbereiche (13)

Perzentilkurven



Die Linien bedeuten das 3., 10., 25., 50., 75., 90. und 97. Perzentil.

Schätzen Sie das Perzentil für Paul ab (128 cm bei 8 Jahren)!

Referenzbereiche (15)

- Für halbwegs normalverteilte (Gauß-verteilte) Größen (Histogramm-Oberkanten beschreiben etwa eine Gaußsche Glockenkurve) gilt:

  95% aller Werte liegen im Bereich

  $$Mittelwert\ \bar{x} \pm 1{,}96\ Standardabweichungen\ s$$

- Daher definiert man den Referenzbereich häufig:

  $$untere\ Grenze = \bar{x} - 1{,}96\ s$$
  $$obere\ Grenze = \bar{x} + 1{,}96\ s$$

  (z.B. legt das Zentrallabor am Uniklinikum Leipzig auf diese Art Referenzbereiche auf der Grundlage von Blutspenderdaten fest)

Referenzbereiche (16)
Bemerkungen

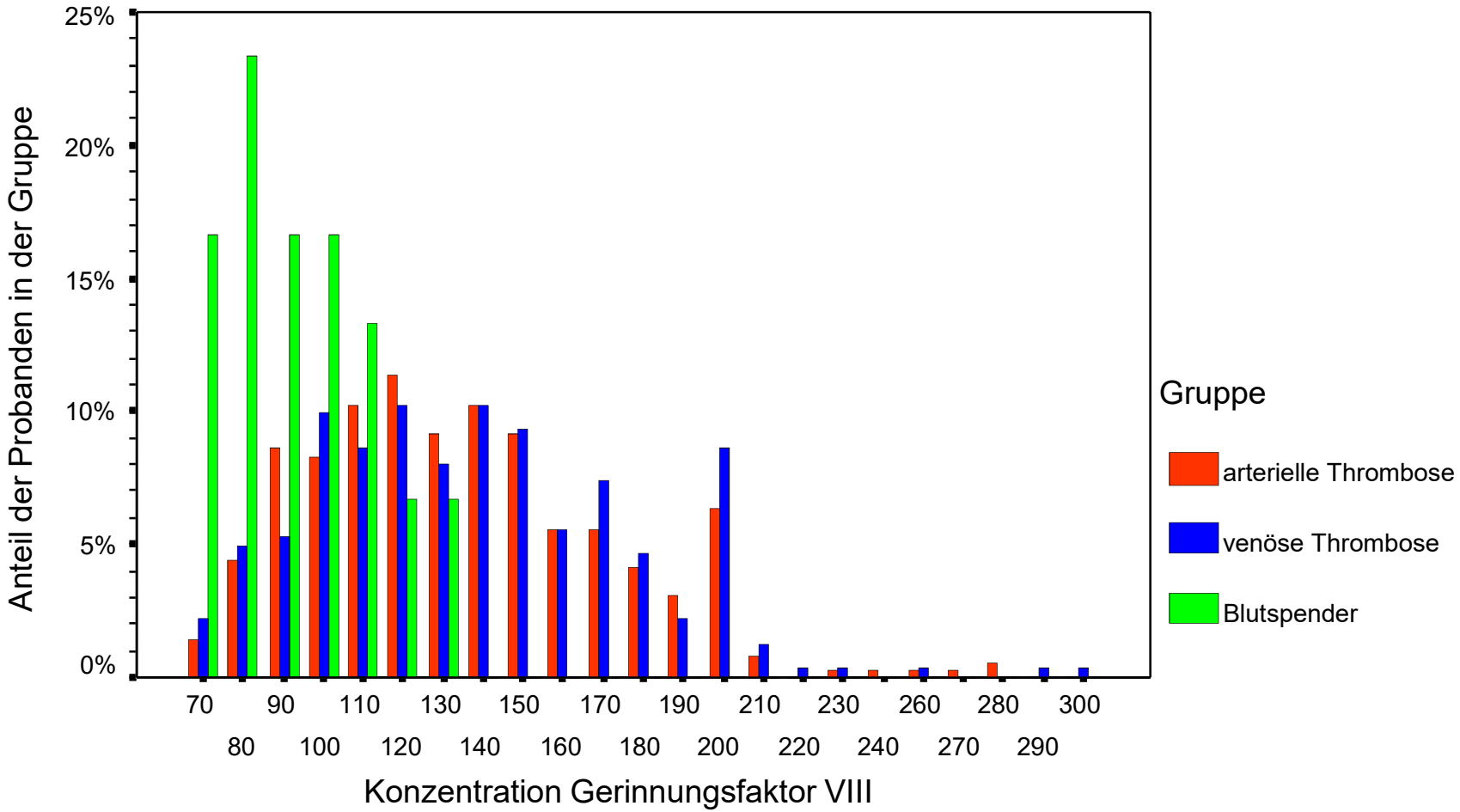
- **ggf. nur einseitige Referenzbereiche**

  Bsp.: Bei Myokardinfarkt ist die Myoglobin-Fraktion der Creatinkinase (CK-MB) erhöht. Für CK-MB wird oft nur die obere Grenze des Referenzbereiches angegeben, i.d.R. 6% von CK.

- **Zielbereich ≠ Referenzbereich**

  Bsp.: Als Referenzbereich für HbA_{1c} gilt < 6%. Als Therapieziel für die Einstellung von Diabetikern wird HbA_{1c} < 6,5% angegeben (Richtlinie des Disease Management Programm).

Gerinnungsfaktoren (1)



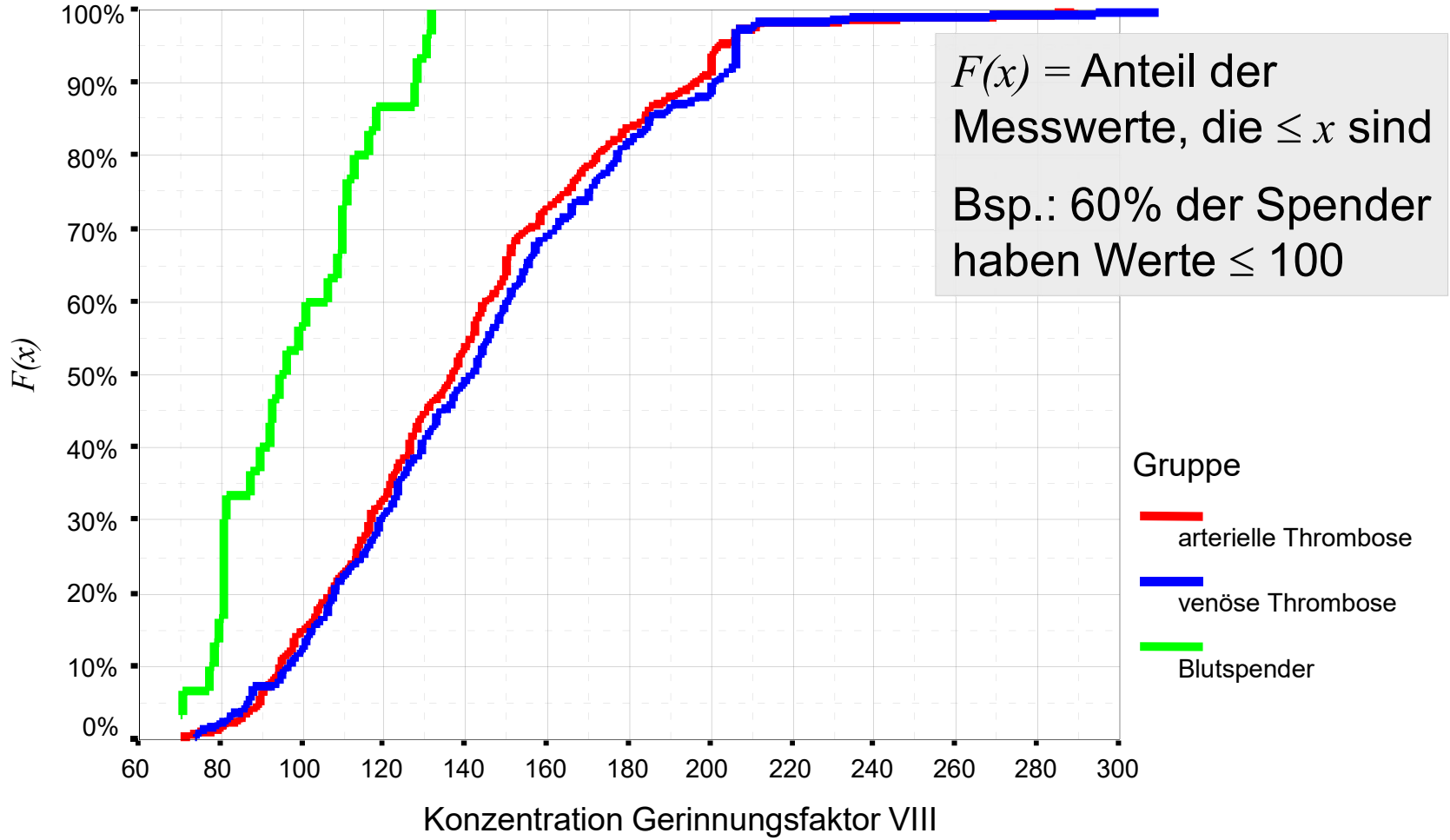
Gerinnungsfaktoren (2)

Diskussion

- Inwiefern ist Gerinnungsfaktor VIII für Diagnose/Ausschluss von Thrombose geeignet?
- Trennung venös/arteriell?
- Wie entsteht der Peak auf der rechten Flanke der Verteilung?

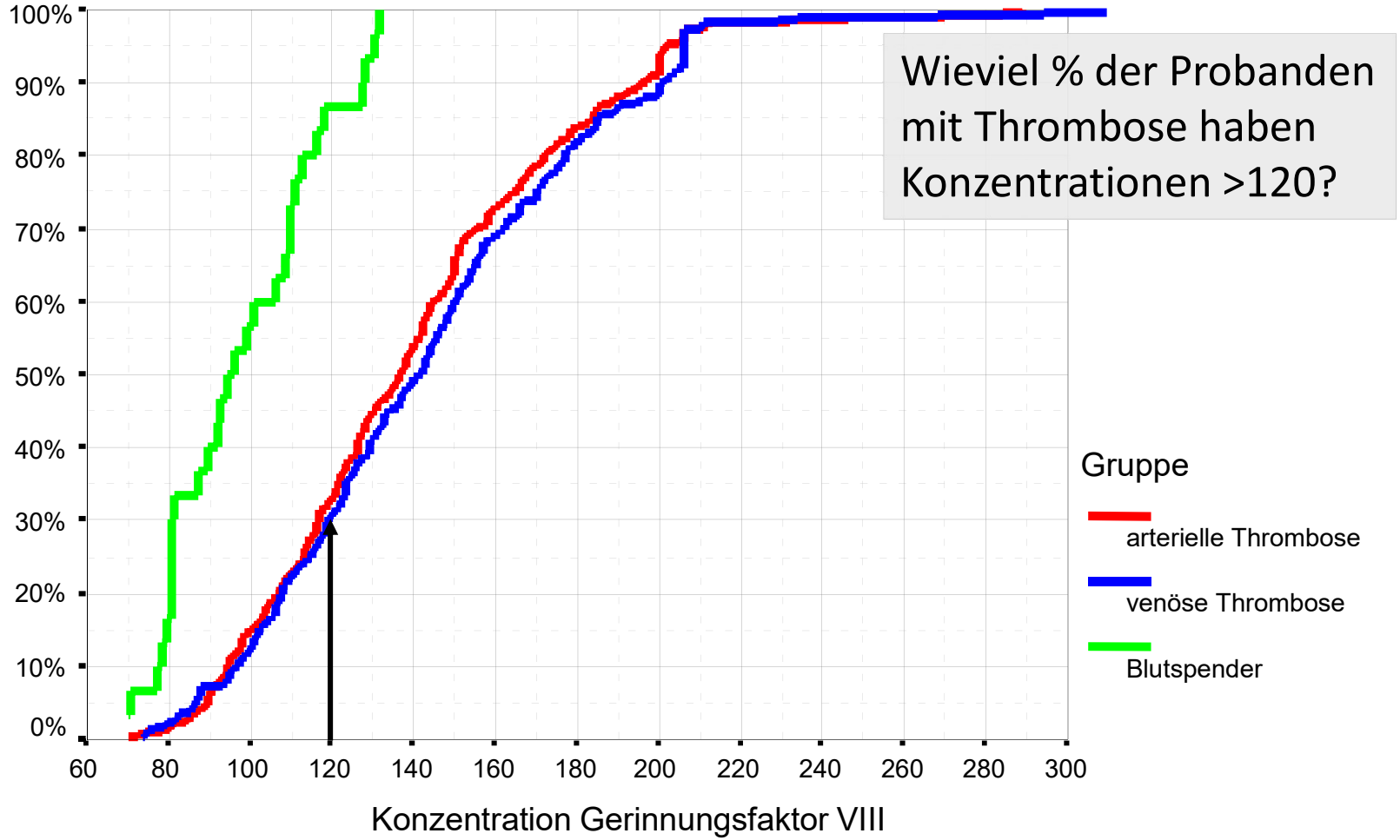
Gerinnungsfaktoren (4)

Empirische Verteilungsfunktion $F(x)$



Gerinnungsfaktoren (5)

Empirische Verteilungsfunktion $F(x)$



Gerinnungsfaktoren (7)

Sensitiviät & Spezifität

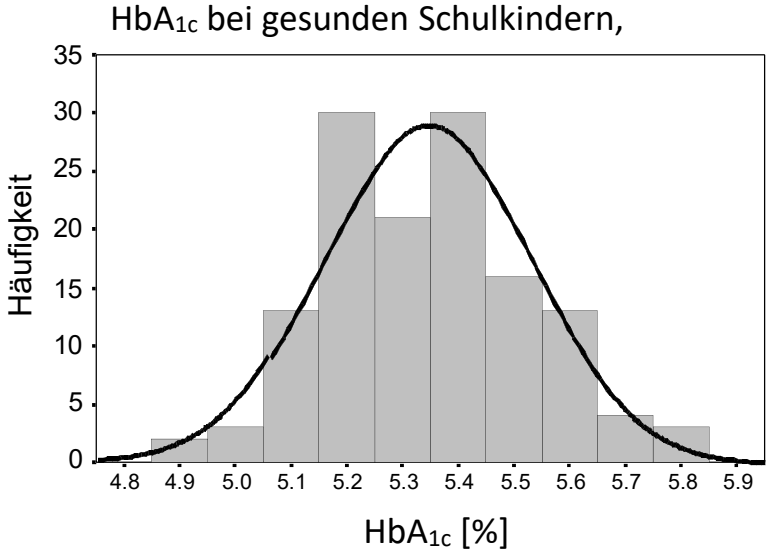
Diagnose von Thrombose

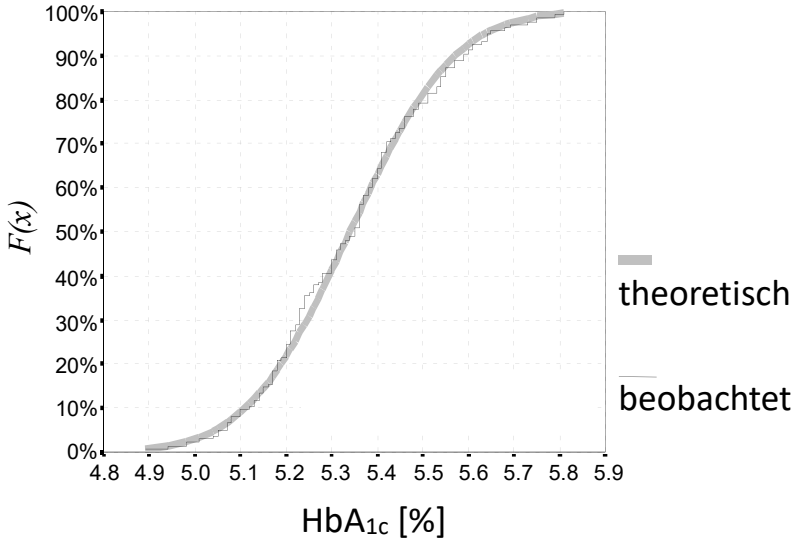
Ermitteln Sie anhand der empirischen Verteilungsfunktionen die fehlenden Werte!

Schwellwert	Sensitivität	Spezifität
120		
140		
	90%	

Normalverteilung (1)

HbA_{1c}





Bis auf kleine Schwankungen, die sich aus der Zufälligkeit der Stichprobenauswahl ergeben, sind die HbA_{1c}-Werte bei Gesunden normalverteilt ($\bar{x} = 5{,}34; s = 0{,}18$ in dieser Stichprobe).

Damit liegen 95% aller Werte im Bereich: $5{,}34 \pm 1{,}96 \cdot 0{,}18$

(4,99 - 5,69)

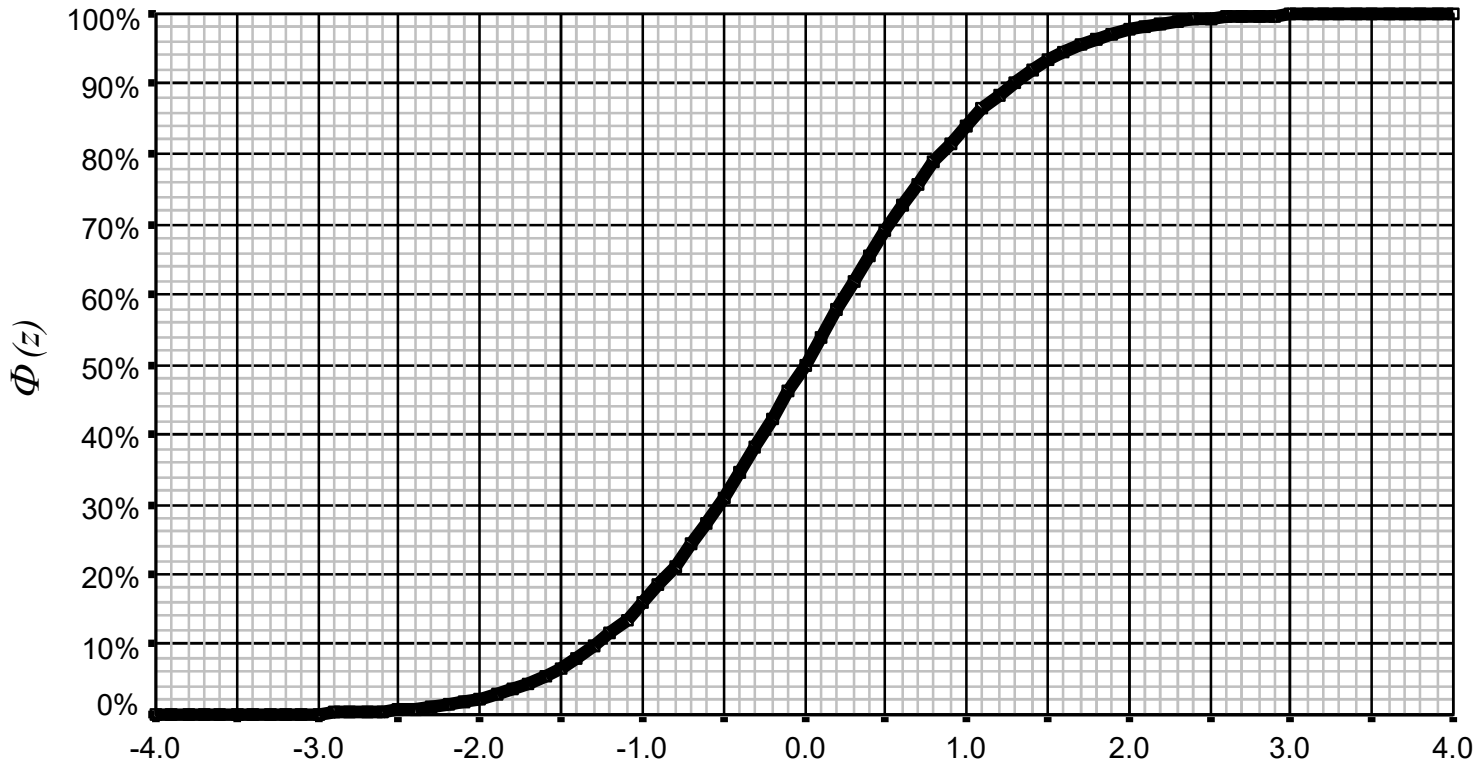
Normalverteilung (2)

HbA_{1c} - Aufgabe 1

1. Wie groß ist der Anteil der Gesunden mit einem $HbA_{1c} \leq 5{,}8\%$?

Normalverteilung (3)

Verteilungsfunktion der standardisierten Normalverteilung



Modell: $Z = \frac{X-\mu}{\sigma}$ Konkrete Stichprobe: $z = \frac{x-\bar{x}}{s}$

Hausaufgabe (1)

2. Wie groß ist der Anteil der Gesunden mit einem $HbA_{1c} > 5{,}8\%$?

3. Wie groß ist der Anteil der Gesunden mit einem $HbA_{1c} \leq 5{,}1\%$?

4. Wie groß ist der Anteil der Gesunden mit einem HbA_{1c} zwischen 5,1 und 5,8%?

5. Oberhalb welches HbA_{1c}-Wertes hat nur etwa jeder tausendste Gesunde seinen HbA_{1c}-Wert?