Statistik II

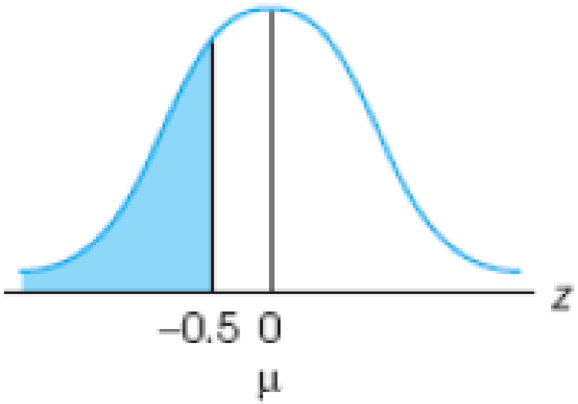
Prof. Dr. Simone Abendschön

Vorlesung am 15.5.23

Plan heute

- Kurze Wiederholung und Auflösung der Übungen vom letzten Mal
- Grundlagen der Inferenzstatistik
  - Zentrales Grenzwerttheorem
  - Standardfehler

Übungsbeispiel 3): Hausaufgabe bzw. Tutorium

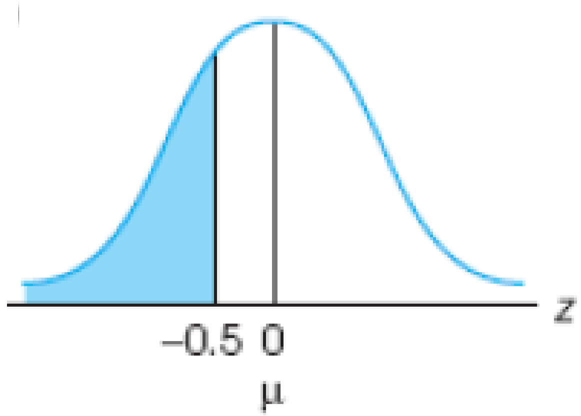


Welcher Flächenanteil der Normalverteilung entspricht z-Werten <-0,5?

Wie groß ist die Wahrscheinlichkeit, für normalverteilte Werte einen z-Wert <-0,5 zu erhalten?

Übungsbeispiel 3)

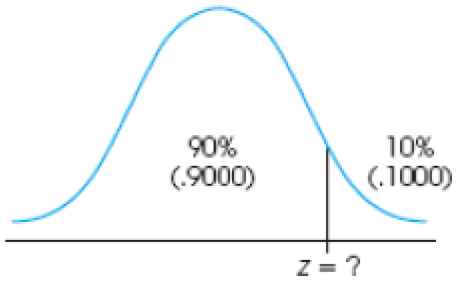
Bsp. c)



Welcher Flächenanteil der Normalverteilung entspricht z-Werten <-0,5?
Wie groß ist die Wahrscheinlichkeit, für normalverteilte Werte einen z-Wert <-0,5 zu erhalten?

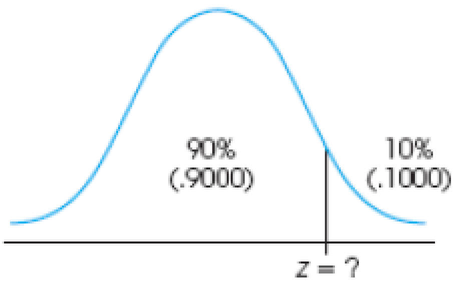
Vorgehen:
Skizzieren NV und gesuchte Fläche
Bestimme z = -0,5 in der z-Werte Tabelle:
P(z < -0.5) = 0.3085 =30.85%

Übungsbeispiel 4)



Welcher z-Wert separiert die obersten 10% aller Werte von den restlichen 90% der Verteilung?

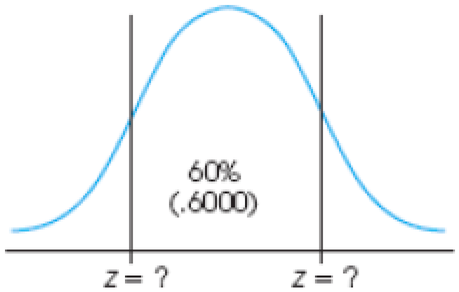
Übungsbeispiel 4)



Welcher z-Wert separiert die obersten 10% aller Werte von den restlichen 90% der Verteilung?

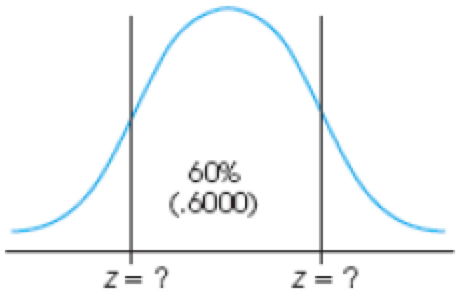
- Skizzieren der Normalverteilung und der gesuchten Fläche
- Bestimme P = 0.90 in der z-Werte Tabelle
- Bestimme korrespondierenden z-Wert: z = 1.28

Übungsbeispiel 5): Hausaufgabe bzw. Tutorium



Welche z-Werte separieren die mittleren 60% aller Werte von den restlichen 40% der Verteilung?

Übungsbeispiel 5)



Welche z-Werte separieren die mittleren 60% aller Werte von den restlichen 40% der Verteilung?

- Skizzieren der Normalverteilung und der gesuchten Fläche
- Bestimme P = 0.20 in der z-Werte Tabelle
- Bestimme korrespondierende z-Werte:

z = -0.84; z = 0.84

Flächenanteile & Wahrscheinlichkeiten für z-Werte

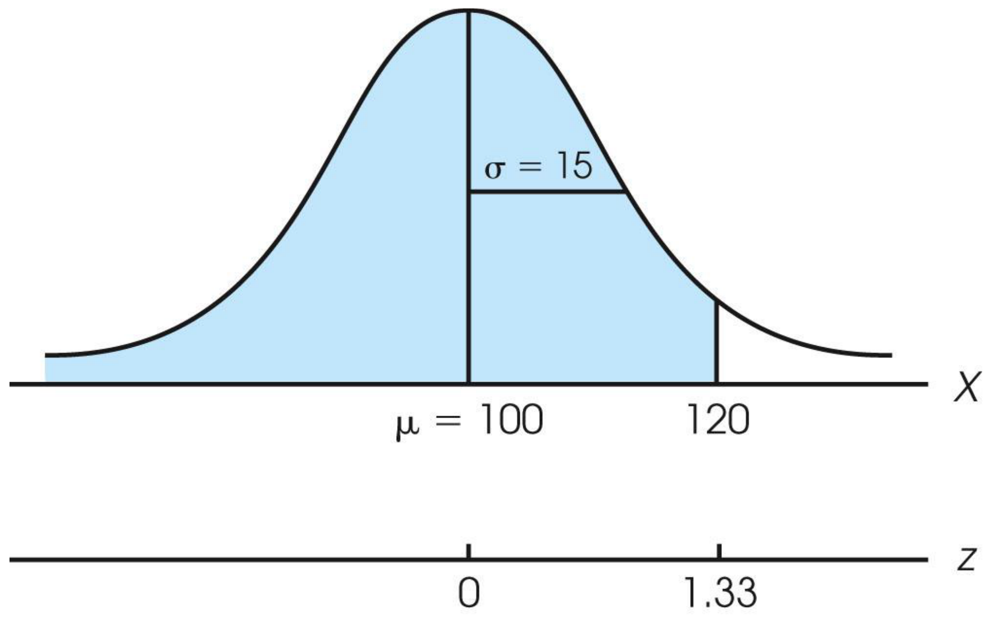
Anwendungsbeispiel A:

- Gegeben sei eine Verteilung von IQ-Werten mit $\mu = 100$ und $\sigma = 15$. Wie groß ist die Wahrscheinlichkeit, zufällig eine Person mit einem IQ < 120 auszuwählen?

Flächenanteile & Wahrscheinlichkeiten für z-Werte

Anwendungsbeispiel A:

- Gegeben sei eine Verteilung von IQ-Werten mit μ= 100 und σ= 15. Wie groß ist die Wahrscheinlichkeit, zufällig eine Person mit einem IQ < 120 auszuwählen?

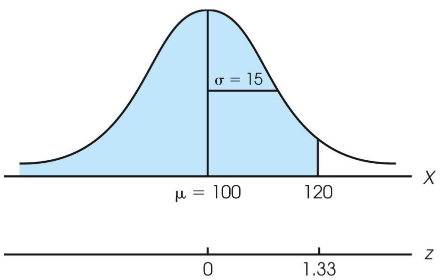


Flächenanteile & Wahrscheinlichkeiten für z-Werte

- Anwendungsbeispiel A:
- Gegeben sei eine Verteilung von IQ-Werten mit μ= 100 und σ= 15. Wie groß ist die Wahrscheinlichkeit, zufällig eine Person mit einem IQ < 120 auszuwählen?

1) Transformieren Rohwerte in z-Werte

$$z = \frac{X-\mu}{\sigma} = \frac{120-100}{15} = \frac{20}{15} = 1.33$$



IQ-Wert von 120 entspricht einem z-Wert von 1.33
IQ-Werte kleiner als 120 entsprechen z-Werten kleiner als 1.33
2) Korrespondierenden z-Wert in Tabelle auswählen:
P = 0.9082
P (X < 120) = P (z < 1.33) = 0.9082 = 90.82%

Flächenanteile & Wahrscheinlichkeiten für z-Werte

Anwendungsbeispiel B:

- Wahrscheinlichkeiten bzw. Anteile zwischen zwei (normalverteilten) X-Werten bestimmen
- In der Gießener Innenstadt werden Geschwindigkeitsmessungen für Autofahrer durchgeführt. Bei der letzten Überprüfung sei für Autofahrer eine Durchschnitts-Geschwindigkeit von μ= 58km/h mit einer Standardabweichung von σ= 10 festgestellt worden. Die Messwerte seien (näherungsweise) normalverteilt.
- Wie hoch ist der Anteil der Autofahrer, die zwischen 55km/h und 65km/h in der Gießener Innenstadt fahren?

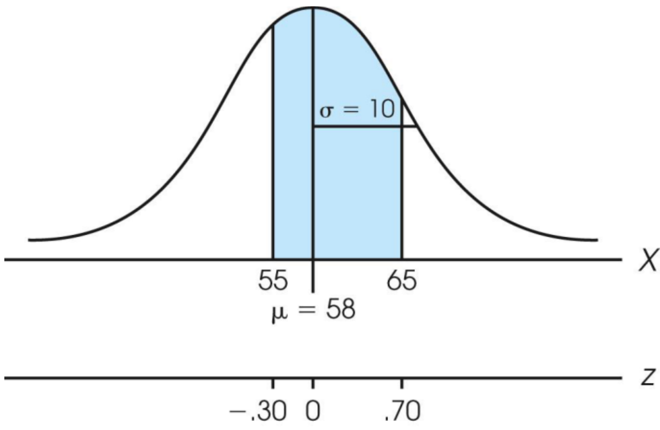
JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN

Anwendungsbeispiel B:

1) Transformieren der Rohwerte in z-Werte

Für X = 55km/h: $z = \frac{X-\mu}{\sigma} = \frac{55-58}{10} = -\frac{3}{10} = -0.3$

Für X = 65km/h: $z = \frac{X-\mu}{\sigma} = \frac{65-58}{10} = \frac{7}{10} = 0.7$



2. Verteilung mit gesuchtem Intervall skizzieren

3a. Bestimmen der Fläche links von X = 65

Für z = .70, P = 0.76

3b. Bestimmen der Fläche links von X = 55

Für z = -.30, P= 0.38

4. Subtrahieren: 0.76 –0.38 = 0.38

Flächenanteile & Wahrscheinlichkeiten für z-Werte

JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN

Anwendungsbeispiel B:

- Wahrscheinlichkeiten/Anteile zwischen zwei (normalverteilten) X-Werten bestimmen
- In der Gießener Innenstadt werden Geschwindigkeitsmessungen für Autofahrer durchgeführt. Bei der letzten Überprüfung sei für Autofahrer eine Durchschnitts-Geschwindigkeit von μ= 58km/h mit einer Standardabweichung von σ= 10 festgestellt worden. Die Messwerte seien (näherungsweise) normalverteilt.
- Wie hoch ist der Anteil der Autofahrer, die zwischen 55km/h und 65km/h in der Gießener Innenstadt fahren? → 38%

Anwendungsbeispiel C:

X-Werte für Wahrscheinlichkeiten/Anteile bestimmen

- Der Asta der JLU finanziert eine sozialwissenschaftliche Untersuchung zur Belastung durch Pendeln unter Studierenden. Die Ergebnisse zeigen, dass von den Studierenden im Durchschnitt $\mu = 24.3$ Minuten pro Studientag für An-und Abreise verbraucht werden; die Standardabweichung sei $\sigma = 10$.
- Wieviel Minuten müssten Sie mindestens pendeln, um zu den 10% Studis mit der höchsten Pendeldauer für An-und Abreise zum Studienort zu gehören?

- Anwendungsbeispiel C: X-Werte für Wahrscheinlichkeiten/Anteile bestimmen

1. Bestimme 90% bzw. 0.90 in der z-Werte Tabelle und den dazugehörigen z-Wert: z = 1.282
2. Bestimme das Vorzeichen des gesuchten z-Wertes: positiv
3. Transformiere den z-Wert in den Rohwert:

$X = \mu + z\sigma$

$= 24.3 + 1.282 \cdot 10$

$= 24.3 + 12.82$

$= 37.1$

Anwendungsbeispiel C:

X-Werte für Wahrscheinlichkeiten/Anteile bestimmen

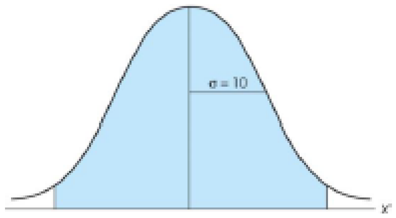
- Der Asta der JLU finanziert eine sozialwissenschaftliche Untersuchung zur Belastung durch Pendeln unter Studierenden. Die Ergebnisse zeigen, dass von den Studierenden im Durchschnitt $\mu = 24.3$ Minuten pro Studientag für An-und Abreise verbraucht werden; die Standardabweichung sei $\sigma = 10$.
- Wieviel Minuten müssten Sie mindestens pendeln, um zu den 10% Studis mit der höchsten Pendeldauer für An-und Abreise zum Studienort zu gehören?

→ ca. 37 Minuten

Hausaufgabe / Tutorium!

Anwendungsbeispiel D (gleiche Population wie eben):

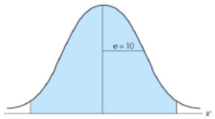
- X-Werte zwischen zwei Wahrscheinlichkeiten/Anteilswerten bestimmen
- Wie lautet die Spannweite für die mittleren 90% der Verteilung?



Flächenanteile & Wahrscheinlichkeiten für z-Werte

Anwendungsbeispiel D (gleiche Population):

- X-Werte zwischen zwei Wahrscheinlichkeiten/Anteilswerten bestimmen
- Wie lautet die Spannweite für die mittleren 90% der Verteilung?

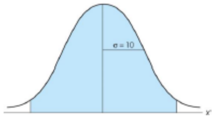


- 1) 90% = jeweils 5% auf beiden Seiten der symmetrischen Normalverteilung
- 2) Bestimmung der gesuchten z-Werte: …

Flächenanteile & Wahrscheinlichkeiten für z-Werte

Anwendungsbeispiel D (gleiche Population):

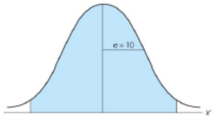
- X-Werte zwischen zwei Wahrscheinlichkeiten/Anteilswerten bestimmen
- Wie lautet die Spannweite für die mittleren 90% der Verteilung?



- 1) 90% = jeweils 5% auf beiden Seiten der symmetrischen Normalverteilung
- 2) Bestimmung der gesuchten z-Werte (z-Tabelle!): z = +1.65 und z = -1.65 trennen jeweils 5% von der Gesamtfläche
- 3) Bestimmung der X-Werte:
- $X = \mu + z\sigma = 24.3 + 1.65 \cdot 10 = 40.8$
- $X = \mu + z\sigma = 24.3 + (-1.65) \cdot 10 = 7.8$

Flächenanteile & Wahrscheinlichkeiten für z-Werte

- Anwendungsbeispiel D:
- X-Werte zwischen zwei Wahrscheinlichkeiten/Anteilswerten bestimmen
- Wie lautet die Spannweite für die mittleren 90% der Verteilung?



- 1)90% = jeweils 5% auf beiden Seiten der symmetrischen Normalverteilung
- 2)Bestimmung der gesuchten z-Werte: z = +1.65 und z = -1.65 trennen jeweils 5% von der Gesamtfläche
- 3) Bestimmung der X-Werte:
- $X=\mu+z\sigma=24.3+1.65\cdot 10=40.8$
- $X=\mu+z\sigma=24.3+(-1.65)\cdot 10=7.8$

90% aller Gießener Studierenden pendeln zwischen 7.8 und 40.8 Minuten zum Studienort, was einer Spannweite von 33 entspricht

Zusammenfassung

- Dichtefunktion der Normalverteilung als Hilfsmittel, um Häufigkeiten bzw. Wahrscheinlichkeiten für kontinuierliche Variablen zu ermitteln
- Wahrscheinlichkeiten können als (Flächen-)Anteile interpretiert werden
- Für normalverteilte Daten liegen tabellarische Darstellungen für interessierende Anteilwerte/Wahrscheinlichkeiten vor, die mit den jeweiligen z-Werten korrespondieren
  - Anhand der Formel zur z-Transformation können X-Werte in z-Werte und z-Werte in X-Werte transformiert werden
  - Für z-Werte können die zugehörigen Wahrscheinlichkeiten/Anteile aus der z-Tabelle entnommen werden

Lernziele

- Sie erweitern Ihre Kenntnisse über die sog. „Normalverteilung“ und wissen wozu sie in der Statistik dient
- Sie können Flächenanteile und damit Wahrscheinlichkeiten innerhalb der Normalverteilung berechnen

Plan heute

Grundlagen der Inferenzstatistik

- Zentrales Grenzwerttheorem
- Standardfehler

Lernziele heute und nächste Woche

JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN

- Kennen und Verstehen des Zentralen Grenzwerttheorems
- Kennen und Bestimmen des Standardfehlers

Einführung

JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN

- Bislang haben wir die Konzepte der Wahrscheinlichkeit, z-Wert-Transformation und Normalverteilung nur für Stichproben mit der Größe $n = 1$ angewendet, d.h.

Wie groß ist die Wahrscheinlichkeit per Zufallsauswahl bei gegebenem Mittelwert und Standardabweichung einen Fall in einem bestimmten Werteintervall auszuwählen ?

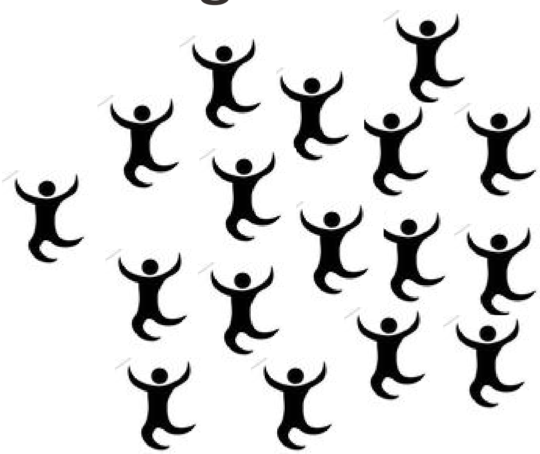
Stichproben und Grundgesamtheit

- Aber sozialwissenschaftliche Forschungspraxis: Stichproben sind typischerweise (sehr) viel größer
  - Z.B. ALLBUS: > 3000 Befragte; European Social Survey: ca. 35.000 Befragte
- Schätzungen auf Basis von Stichprobenkennwerten (z.B. Mittelwerte oder Anteilswerte)
- Diese Kennwerte können ebenfalls in z- (bzw. t-) Werte transformiert und für Wahrscheinlichkeitsaussagen genutzt werden

Inferenzstatistik

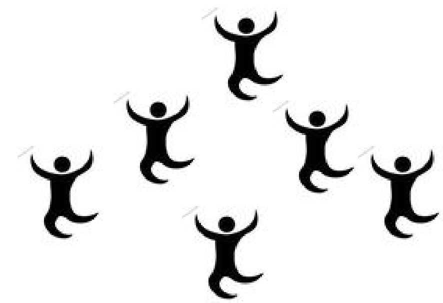
JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN

Grundgesamtheit



Zufallsauswahl/Wahrscheinlichkeit

Stichprobe



μ

(Erwartungswert – „Durchschnitt der Grundgesamtheit“)

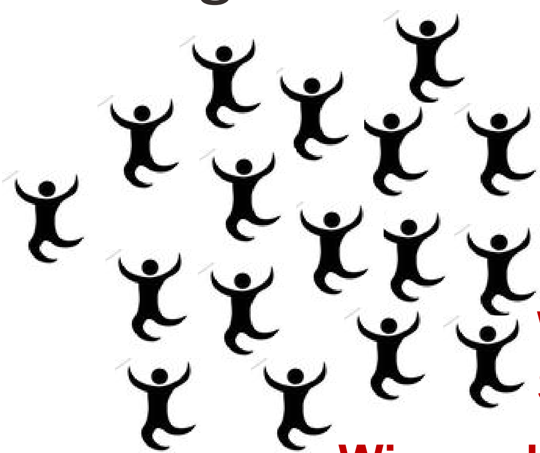
Inferenz

Schätzung

$\bar{x}$ Statistik

(Arithmetisches Mittel Stichprobe)

Grundgesamtheit



Zufallsauswahl

Stichprobe

Wie unsicher ist die Schätzung auf Basis der Stichprobe?

Wie „wahrscheinlich" ist das Stichprobenergebnis auch für die Grundgesamtheit gültig?

Stichprobenfehler

μ

(Erwartungswert – „Durchschnitt der Grundgesamtheit")

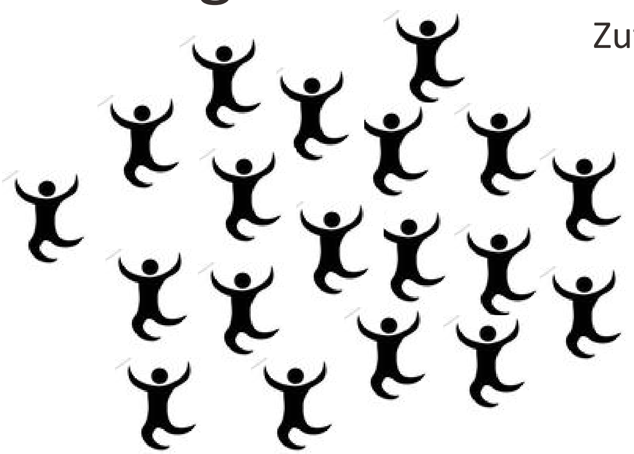
Inferenz

$\bar{x}$ Statistik

(Arithmetisches Mittel Stichprobe)

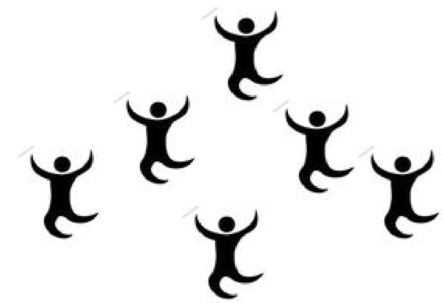
Inferenzstatistik

Grundgesamtheit



Stichprobe

Zufallsauswahl/Wahrscheinlichkeit



Grundannahmen über die Verteilung von Stichprobenkennwerten

μ

(Erwartungswert – „Durchschnitt der Grundgesamtheit“)

Inferenz

Schätzung

$\bar{x}$

Statistik

(Arithmetisches Mittel Stichprobe)

Stichproben und Grundgesamtheit

- Stichprobenfehler (Stichprobenschwankung/Sampling Error):
  - Empirische Ergebnisse einer Zufallsstichprobe weichen immer (mehr oder weniger) vom tatsächlichen Wert in Grundgesamtheit ab
  - → Diskrepanz zwischen Stichprobenkennwert $\bar{x}$ und Populationskennwert μ
  - Berechnung eines Standardfehlers

Stichproben und Grundgesamtheit

- Da wir den „wahren“ Wert in der GG nicht kennen, wissen wir nicht ob unser Stichprobenfehler groß oder klein ist
  - Stichprobenergebnisse variieren – wir können eine „gute“ oder „schlechte“ Stichprobe erwischen
  - Zufällige Einflüsse: Unterschiedliche Stichproben = unterschiedliche Beobachtungseinheiten
- Aber: Grundannahmen über die Verteilung von Stichprobenkennwerten!

Zentrales Grenzwerttheorem

Auch: zentraler Grenzwertsatz

Definition:

- Eine Stichprobenkennwerteverteilung für unendlich viele Stichproben von Mittelwerten nähert sich der Normalverteilung an, falls die Stichproben ausreichend groß sind (n>= 30) oder die Werte in der GG normalverteilt sind
- Der Erwartungswert E der Stichprobenmittelwerte entspricht dem „wahren" Mittelwert der GG

$$\mu: E(\bar{x}) = \mu$$

Stichprobenkennwerteverteilung

JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN

Wie können wir das wissen?

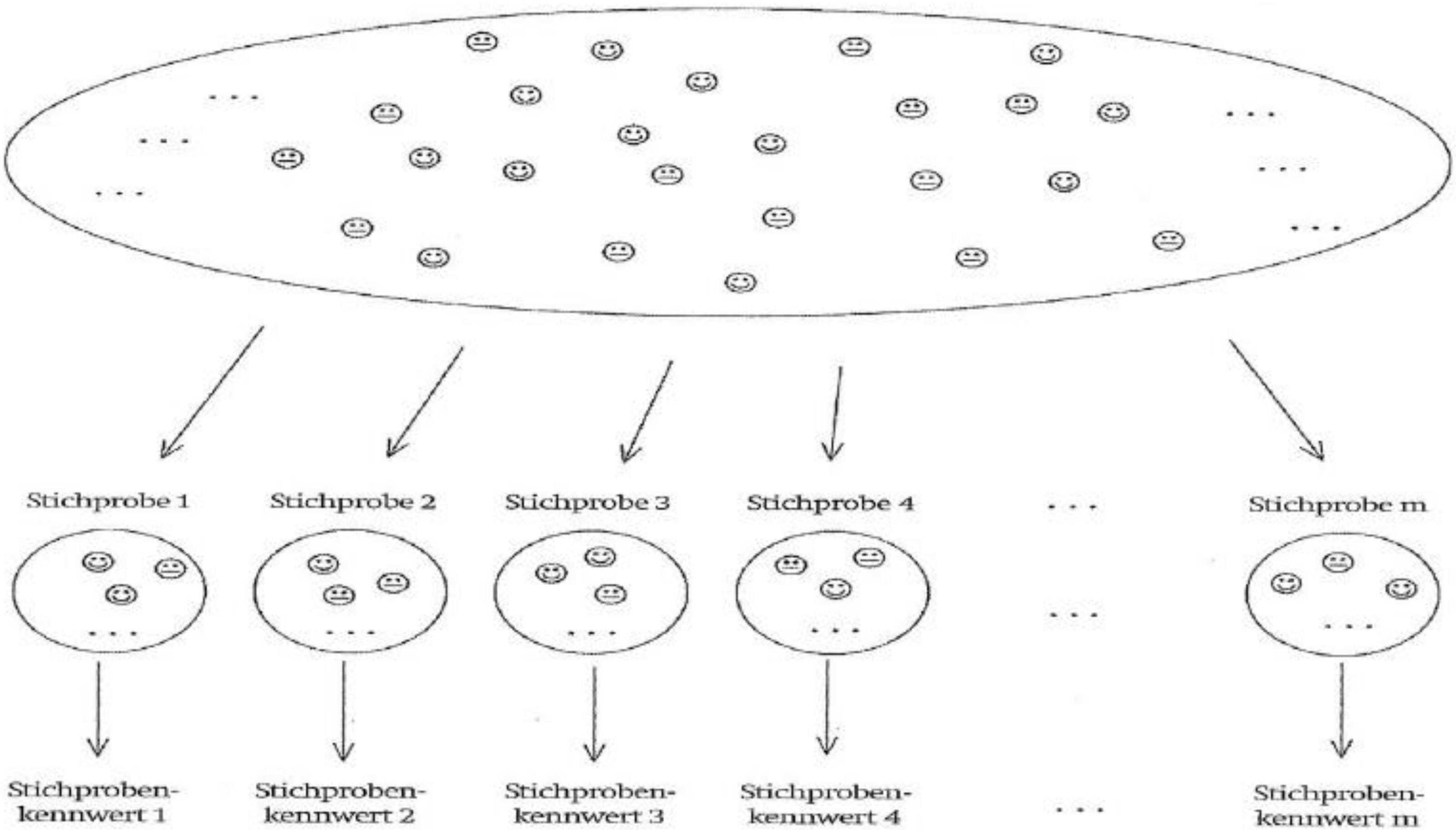
- Es werden theoretisch unendlich viele Stichproben vom jeweils gleichen Umfang *n* aus derselben Grundgesamtheit gezogen.
- Für jede einzelne Stichprobe wird der interessierende Kennwert (hier arithmetisches Mittel) berechnet
- → **Stichprobenmittelwerteverteilung** (Stichprobenkennwerteverteilung), „theoretische“ Verteilung

Simulationsbeispiel

JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN

- Es werden theoretisch unendlich viele Stichproben vom jeweils gleichen Umfang *n* aus derselben Population gezogen (Simulationsbeispiel n=100.000)
- Für jede einzelne Stichprobe wird der interessierende Kennwert (hier arithmetisches Mittel, funktioniert aber auch mit Anteilswert) berechnet

Simulationsbeispiel

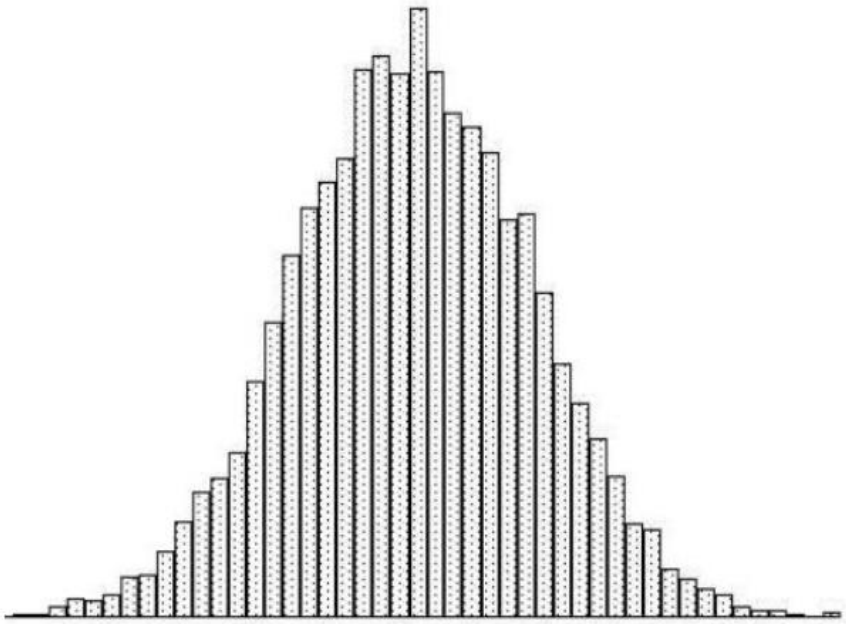


Simulationsbeispiel

- Simulierte Daten, Modellpopulation N= 100.000,
- Unterschiedliche Verteilungsformen
- Für jede Verteilungsform: jeweils 1.000 Zufallsstichproben vom Umfang *n*= 500; Berechnung $\bar{x}$ für jede einzelne Stichprobe
- Berechnung des arithmetischen Mittels aus diesen 1000 Mittelwerten
- Wie sieht die Verteilung der Mittelwerte aus? Was passiert? (Siehe auch Abbildung 22 im Lehrbrief)

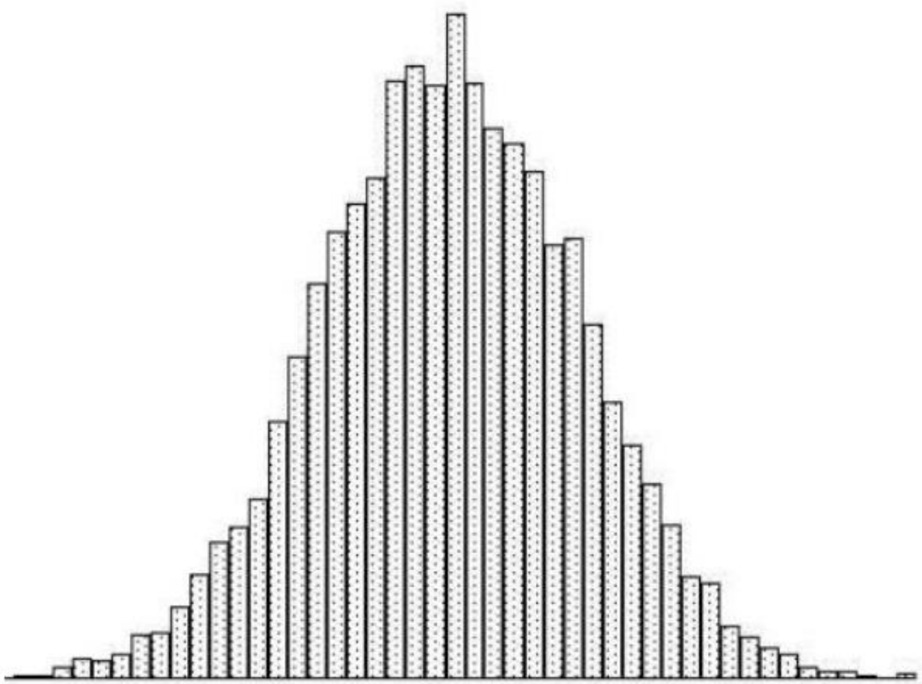
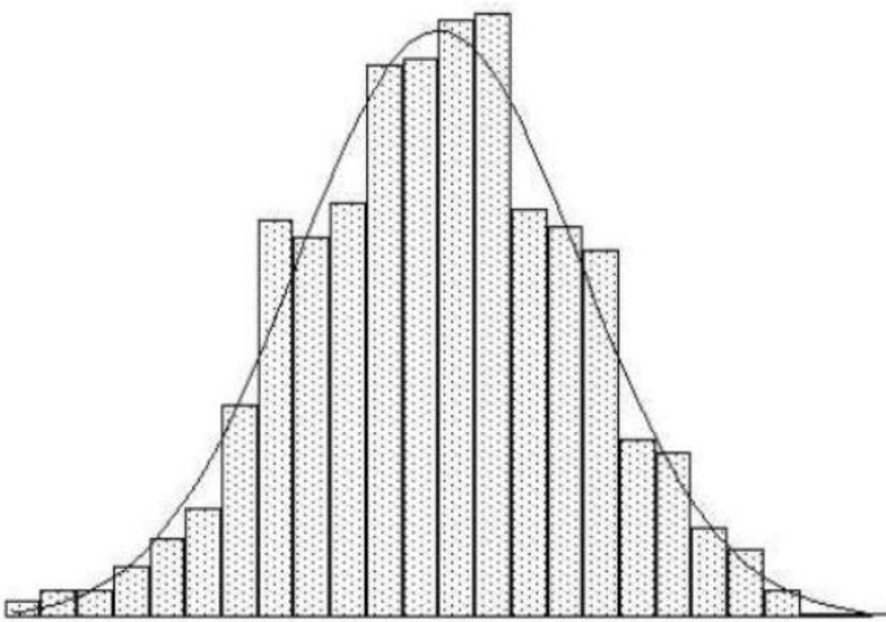
Simulationsbeispiel

Population:	Verteilung der Stichprobenmittelwerte:



Normalverteilung

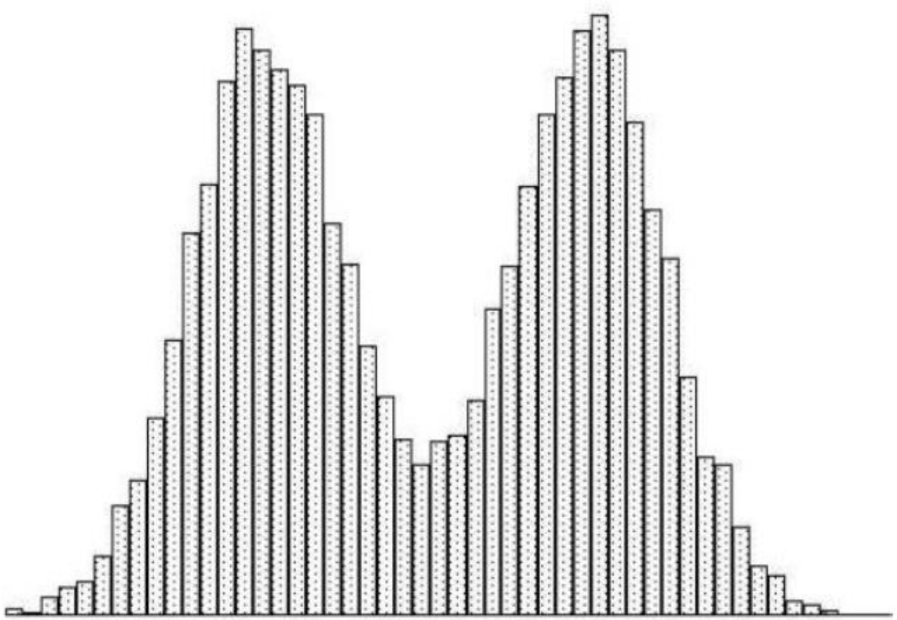
Simulationsbeispiel

Population:	Verteilung der Stichprobenmittelwerte:
	

Normalverteilung

Simulationsbeispiel

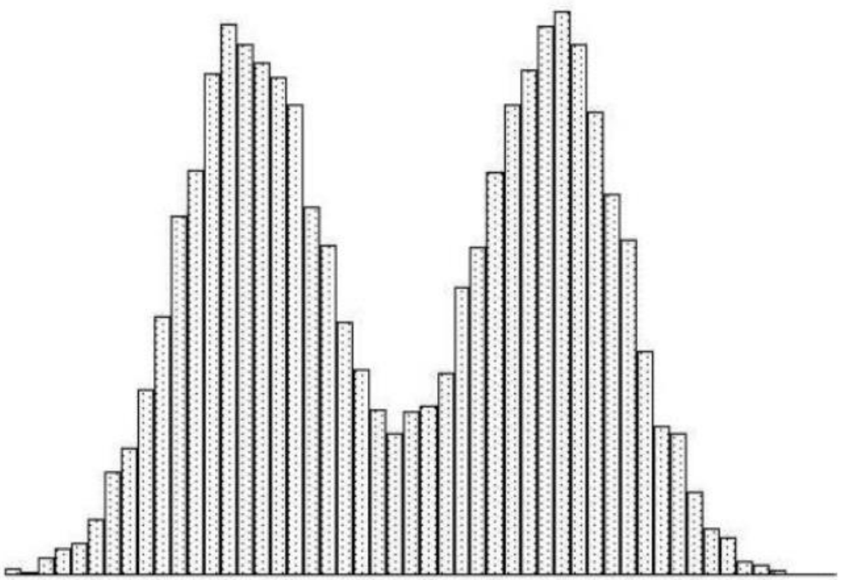
Population:	Verteilung der Stichprobenmittelwerte:



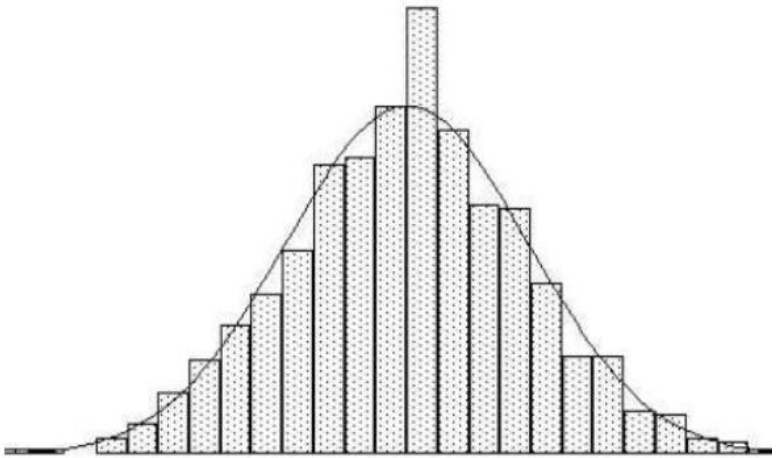
Bimodale Verteilung

Simulationsbeispiel

Population:	**Verteilung der Stichprobenmittelwerte:**

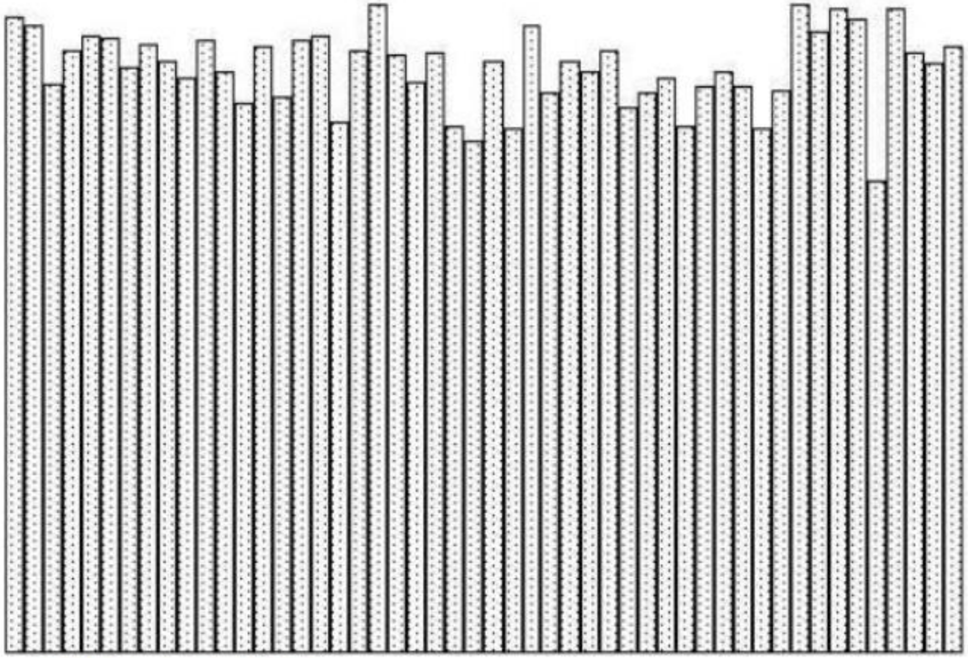


Bimodale Verteilung



Simulationsbeispiel

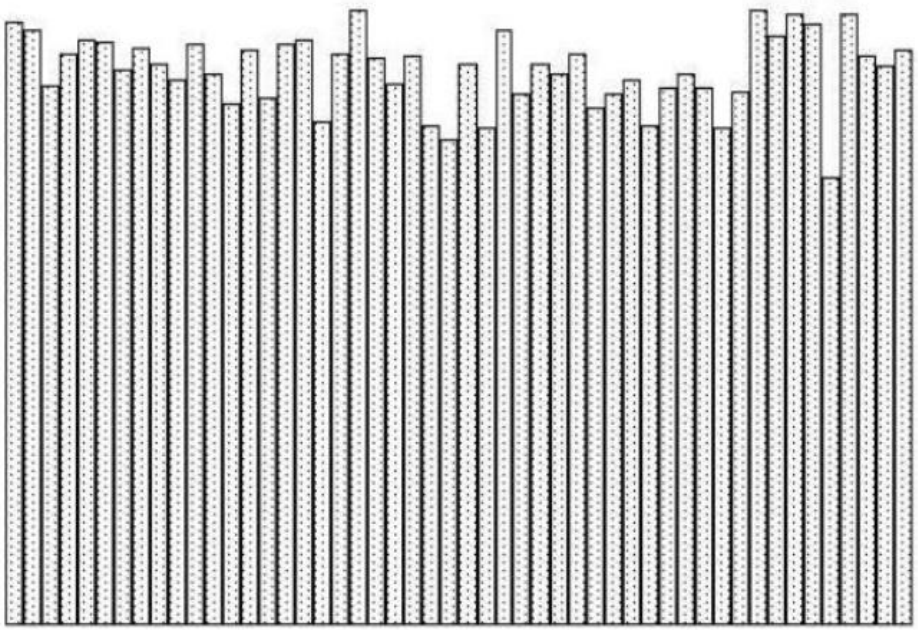
Population:	Verteilung der Stichprobenmittelwerte:



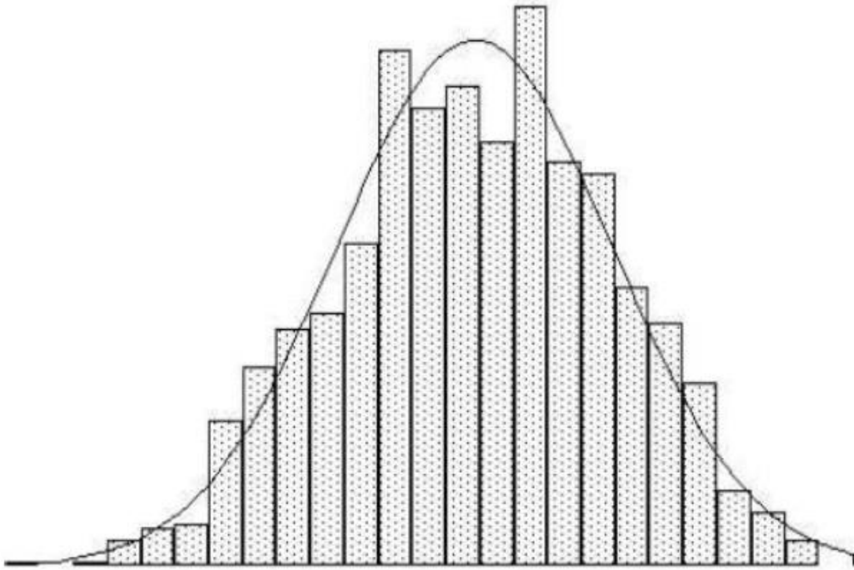
Gleichverteilung

Simulationsbeispiel

Population:	Verteilung der Stichprobenmittelwerte:



Gleichverteilung



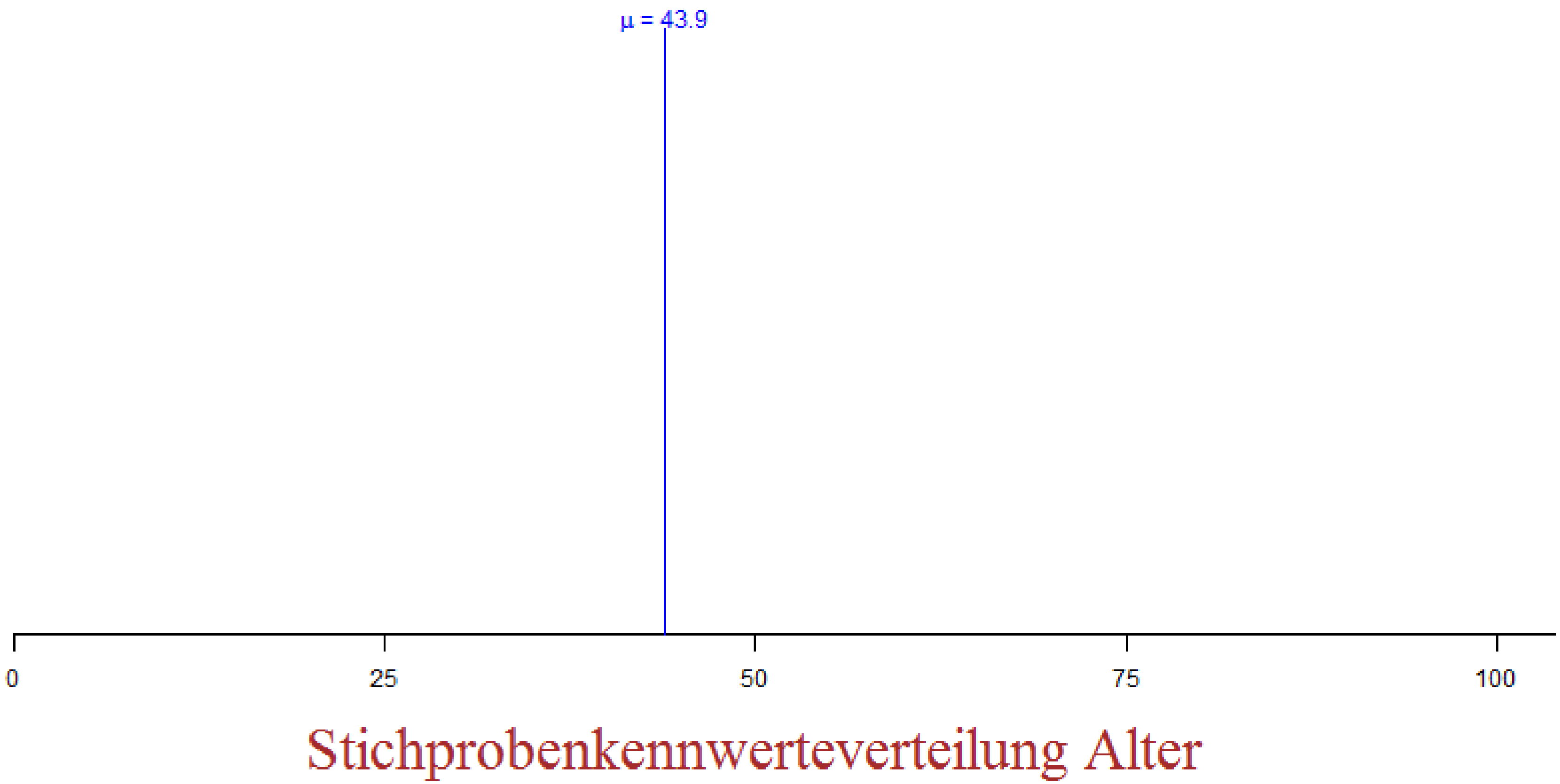
Zentrales Grenzwerttheorem

JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN

- Zentrale Tendenz der Verteilung von Stichprobenkennwerten (Mittelwerte, aber auch Anteilswerte)
- Unabhängig von der Verteilung eines interessierenden Merkmals in der Population wird die Verteilung der Stichprobenmittelwerte (und Anteilswerte) normalverteilt um μ sein
  - falls die Stichprobe ausreichend groß ist (n>= 30)
  - oder die Werte in der Population normalverteilt sind

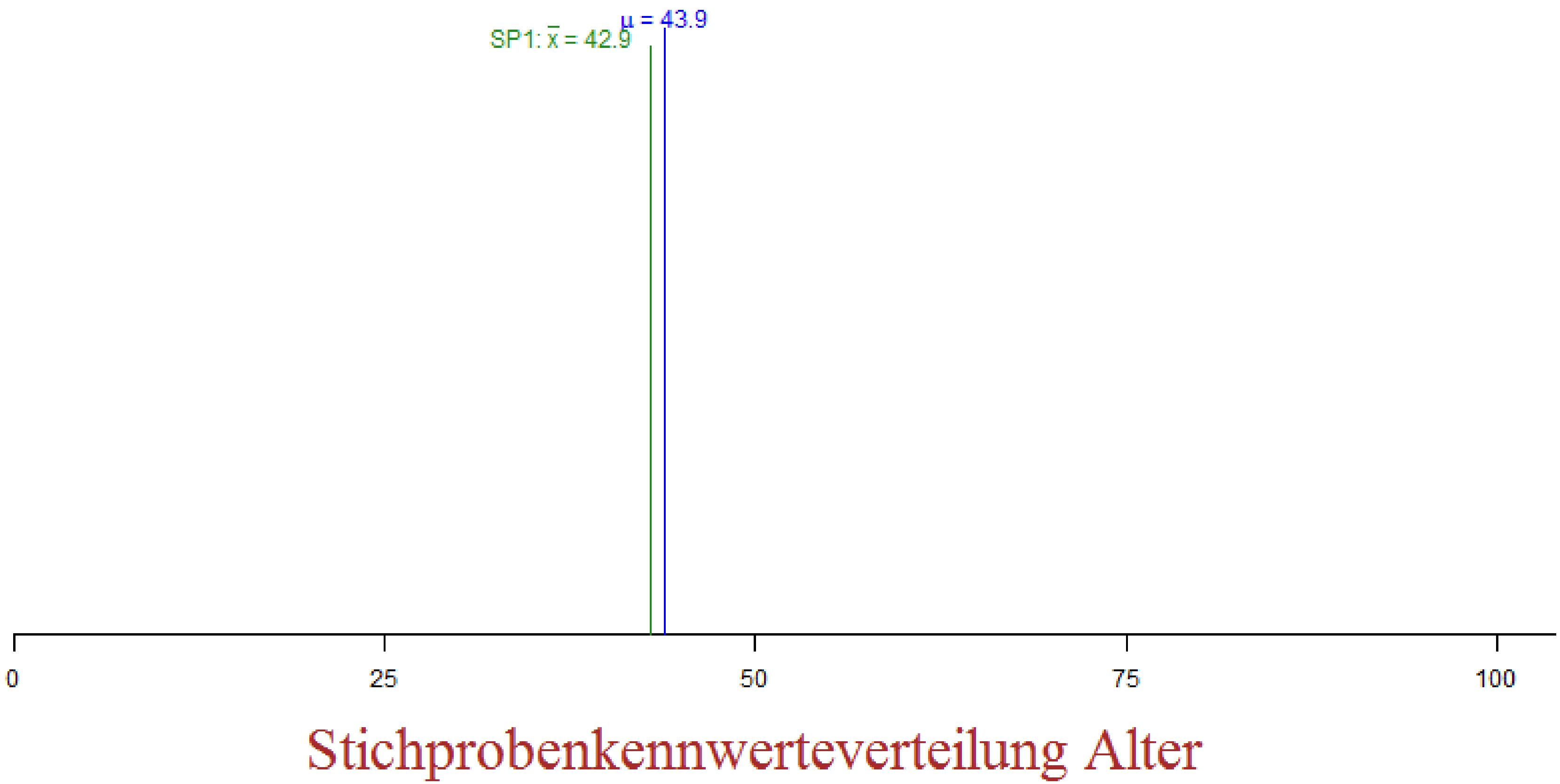
Beispiel: Zentrales Grenzwerttheorem

Arithmetisches Mittel des Alters der dt. Bevölkerung (μ) ist 43,9 Jahre (vgl. (Destatis Zensus 2011: https://www.destatis.de/DE/ZahlenFakten/ImFokus/Bevoelkerung/AltersstrukturZensus.html)



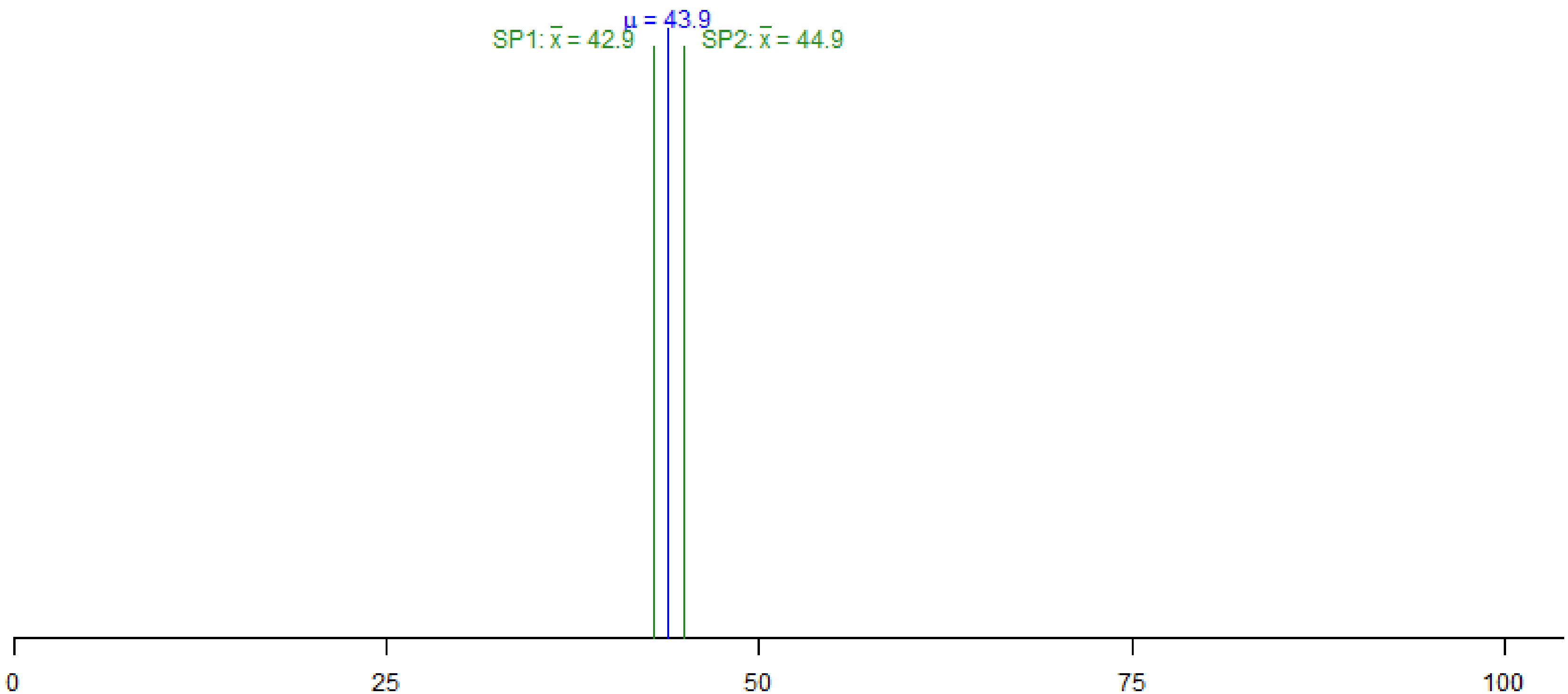
Beispiel: Zentrales Grenzwerttheorem

Arithmetisches Mittel des Alters der dt. Bevölkerung (μ) ist 43,9 Jahre (vgl. (Destatis Zensus 2011: https://www.destatis.de/DE/ZahlenFakten/ImFokus/Bevoelkerung/AltersstrukturZensus.html)



Beispiel: Zentrales Grenzwerttheorem

Arithmetisches Mittel des Alters der dt. Bevölkerung (μ) ist 43,9 Jahre (vgl. (Destatis Zensus 2011: https://www.destatis.de/DE/ZahlenFakten/ImFokus/Bevoelkerung/AltersstrukturZensus.html)

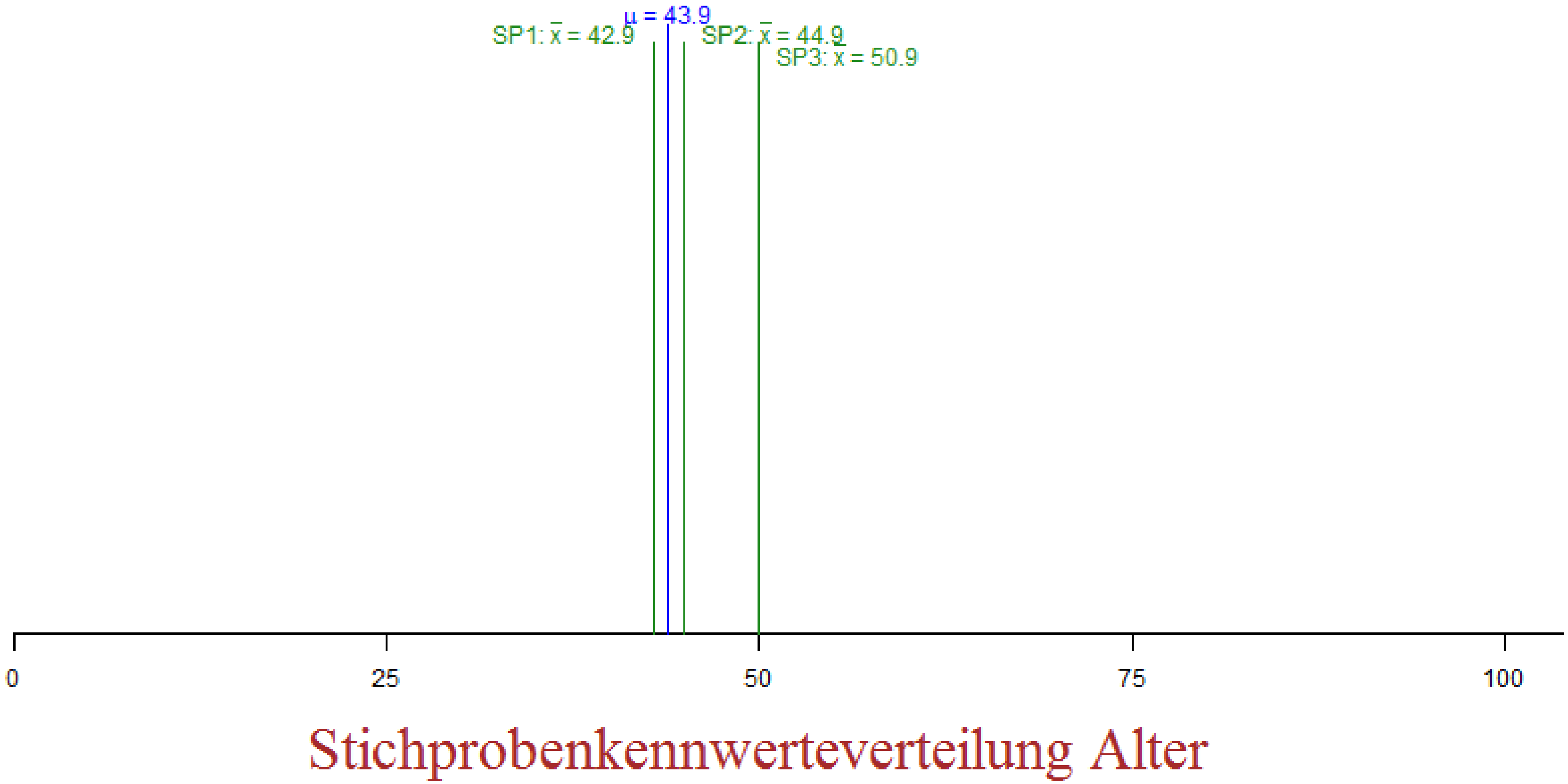


Stichprobenkennwerteverteilung Alter

Beispiel: Zentrales Grenzwerttheorem

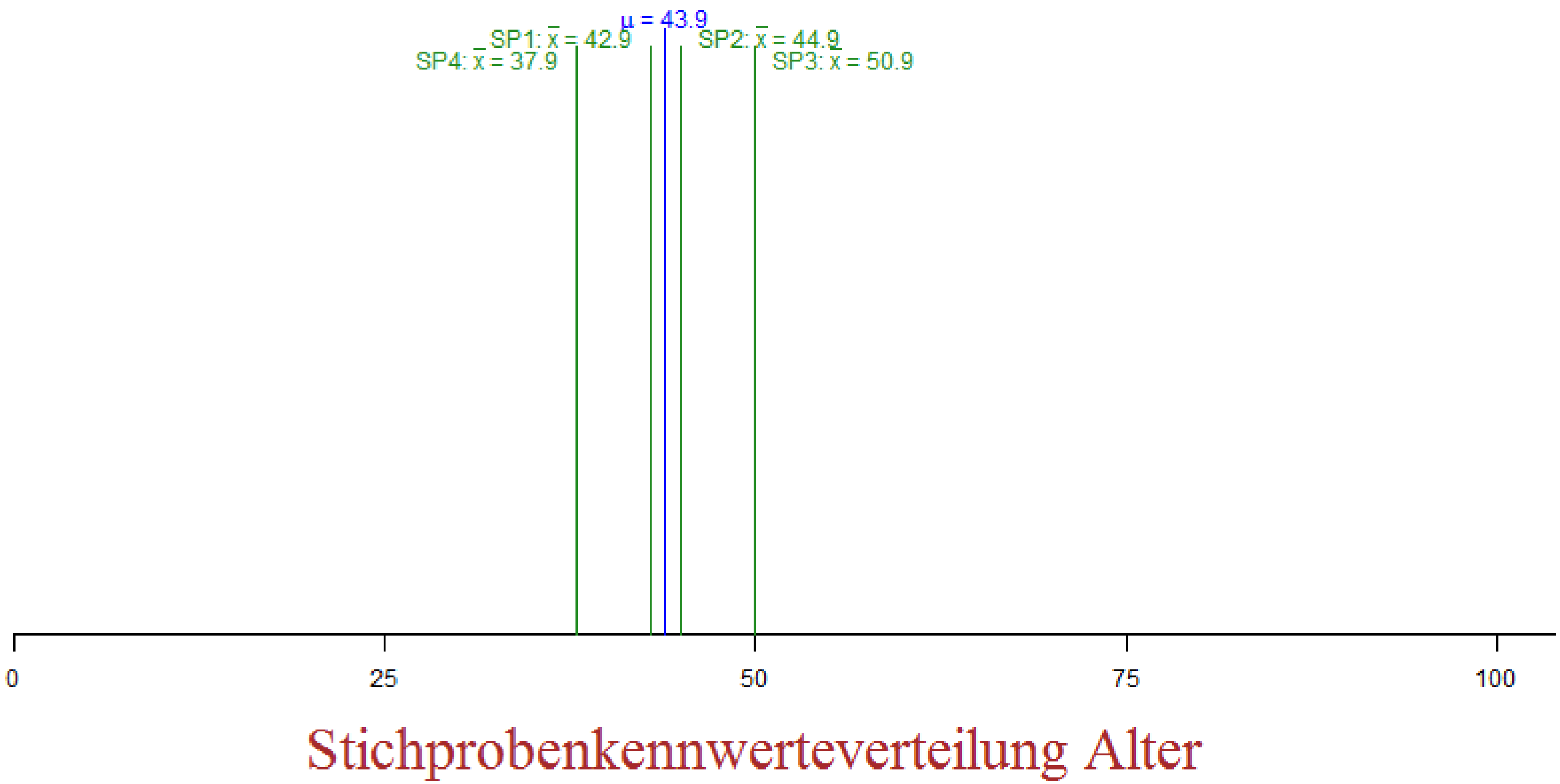
JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN

Arithmetisches Mittel des Alters der dt. Bevölkerung (μ) ist 43,9 Jahre (vgl. (Destatis Zensus 2011: https://www.destatis.de/DE/ZahlenFakten/ImFokus/Bevoelkerung/ AltersstrukturZensus.html)



Beispiel: Zentrales Grenzwerttheorem

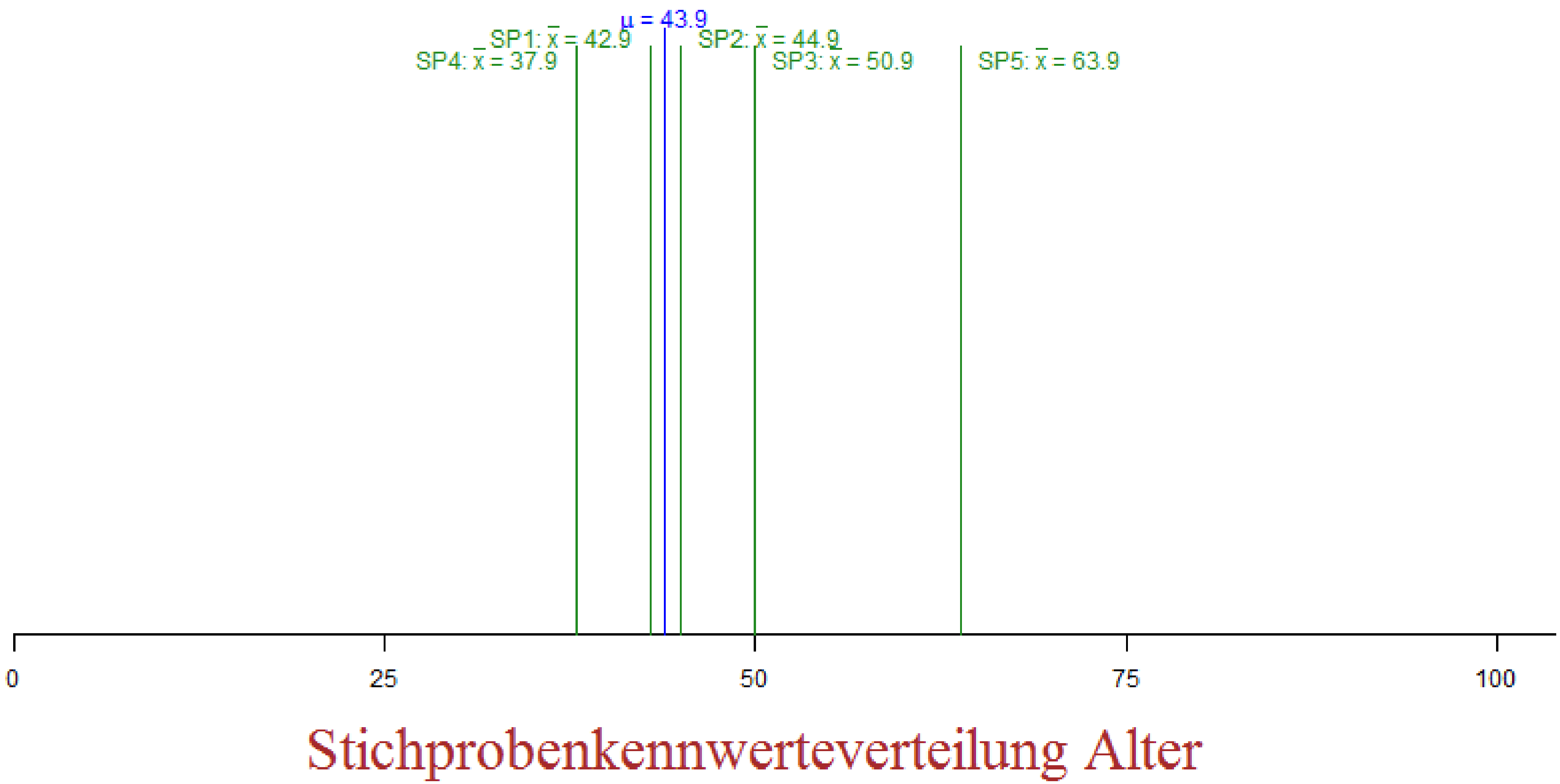
Arithmetisches Mittel des Alters der dt. Bevölkerung (μ) ist 43,9 Jahre (vgl. (Destatis Zensus 2011: https://www.destatis.de/DE/ZahlenFakten/ImFokus/Bevoelkerung/ AltersstrukturZensus.html)



Beispiel: Zentrales Grenzwerttheorem

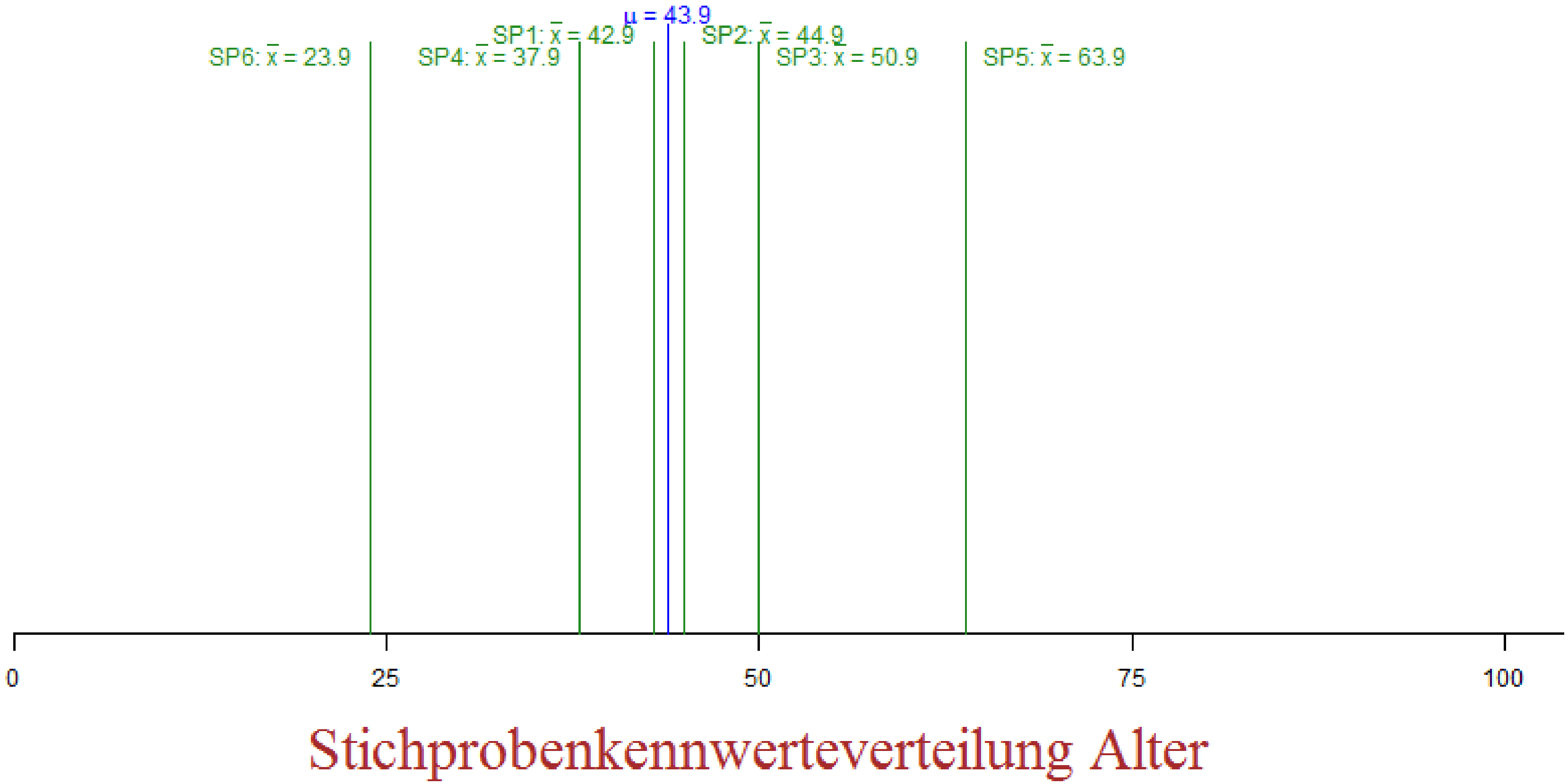
JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN

Arithmetisches Mittel des Alters der dt. Bevölkerung (μ) ist 43,9 Jahre (vgl. (Destatis Zensus 2011: https://www.destatis.de/DE/ZahlenFakten/ImFokus/Bevoelkerung/ AltersstrukturZensus.html)



Beispiel: Zentrales Grenzwerttheorem

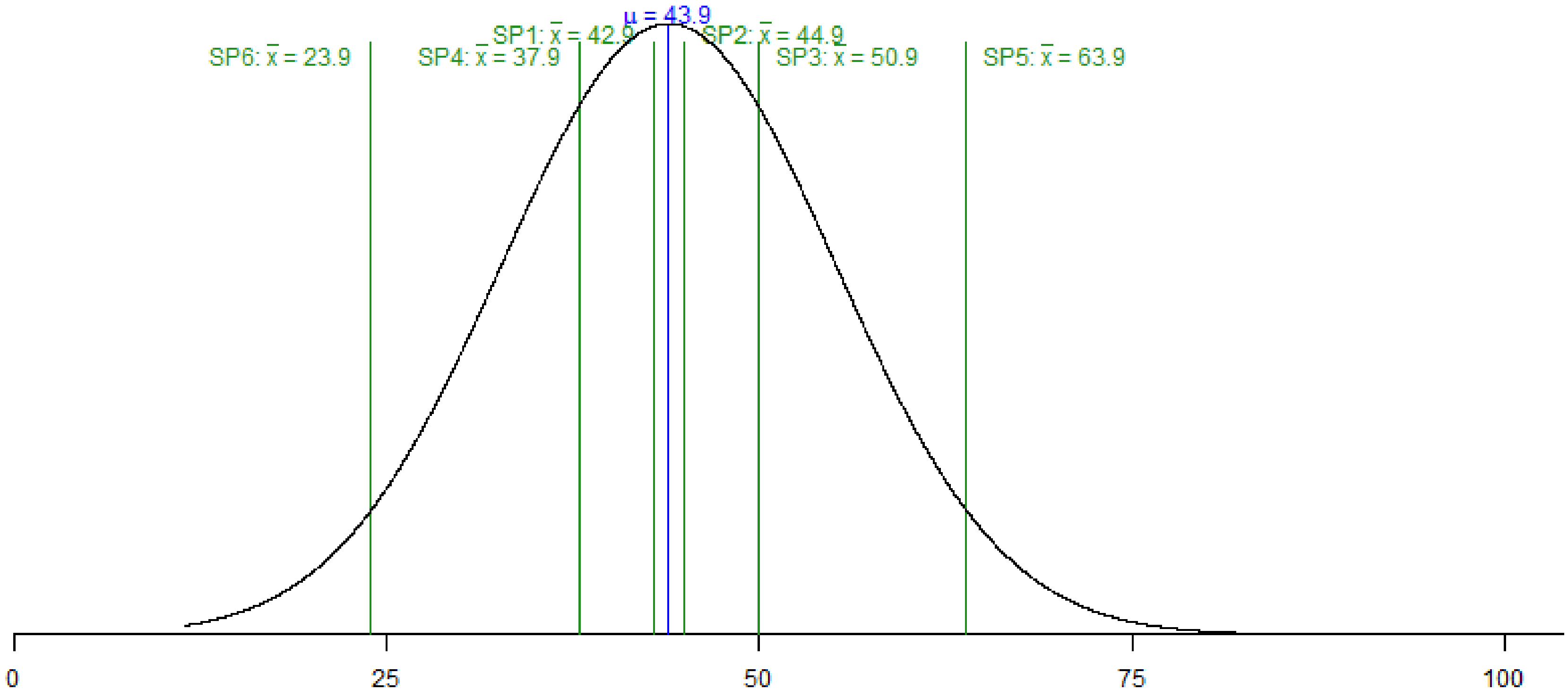
Arithmetisches Mittel des Alters der dt. Bevölkerung (μ) ist 43,9 Jahre (vgl. (Destatis Zensus 2011: https://www.destatis.de/DE/ZahlenFakten/ImFokus/Bevoelkerung/ AltersstrukturZensus.html)



Beispiel: Zentrales Grenzwerttheorem

JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN

- Arithmetisches Mittel des Alters der dt. Bevölkerung (μ) ist 43,9 Jahre (vgl. (Destatis Zensus 2011: https://www.destatis.de/DE/ZahlenFakten/ImFokus/Bevoelkerung/AltersstrukturZensus.html)

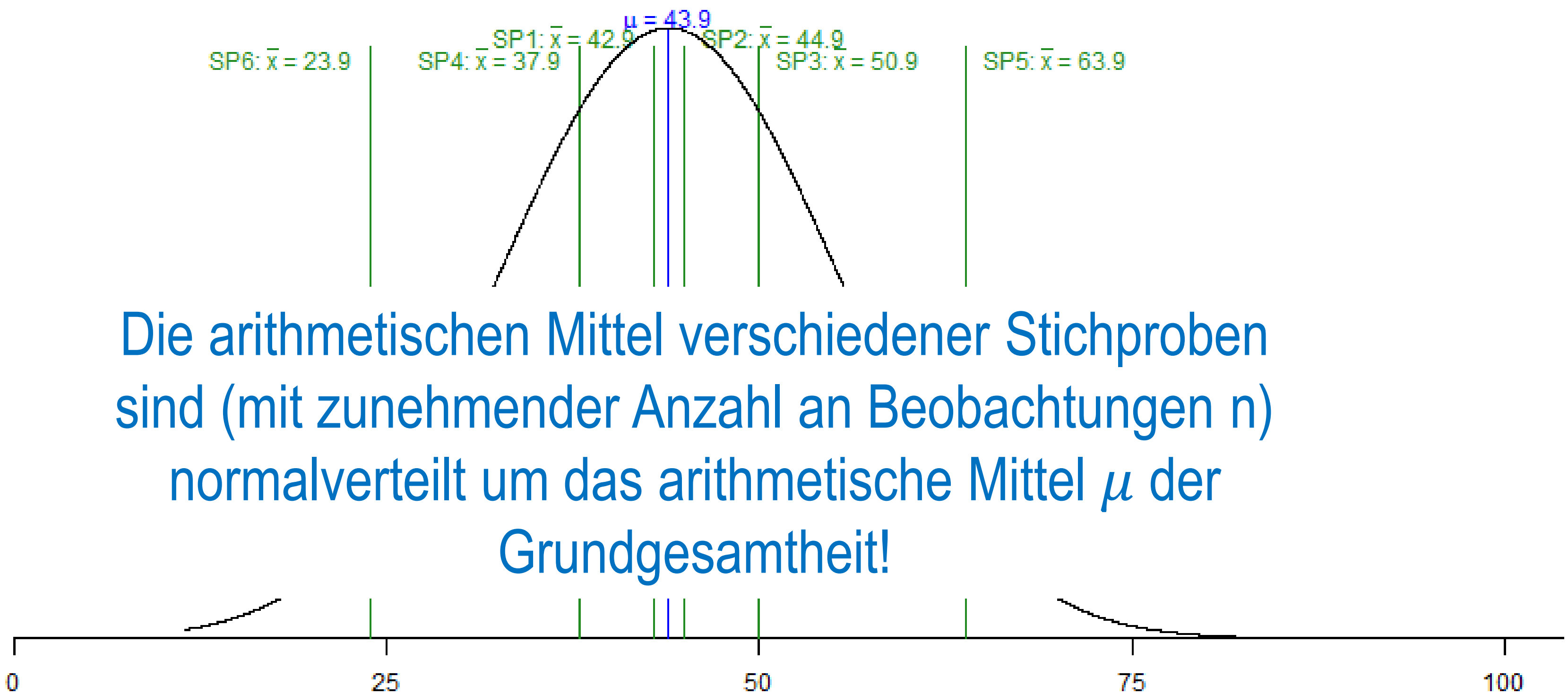


Stichprobenkennwerteverteilung Alter

Beispiel: Zentrales Grenzwerttheorem

JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN

- Arithmetisches Mittel des Alters der dt. Bevölkerung (μ) ist 43,9 Jahre (vgl. (Destatis Zensus 2011: https://www.destatis.de/DE/ZahlenFakten/ImFokus/Bevoelkerung/AltersstrukturZensus.html)



Die arithmetischen Mittel verschiedener Stichproben sind (mit zunehmender Anzahl an Beobachtungen n) normalverteilt um das arithmetische Mittel μ der Grundgesamtheit!

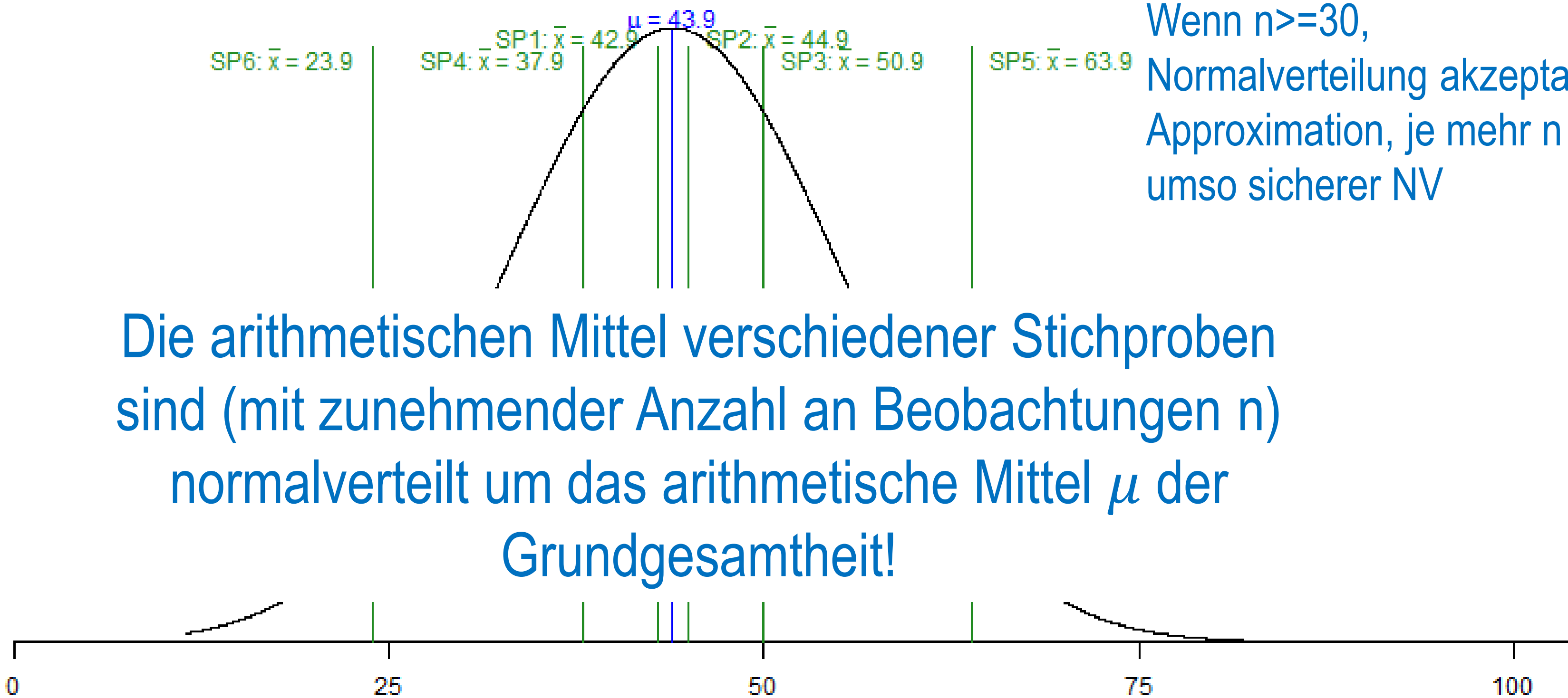
Stichprobenkennwerteverteilung Alter

Beispiel: Zentrales Grenzwerttheorem

JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN

- Arithmetisches Mittel des Alters der dt. Bevölkerung (μ) ist 43,9 Jahre (vgl. (Destatis Zensus 2011: https://www.destatis.de/DE/ZahlenFakten/ImFokus/Bevoelkerung/AltersstrukturZensus.html)

Wenn n>=30, Normalverteilung akzeptable Approximation, je mehr n umso sicherer NV



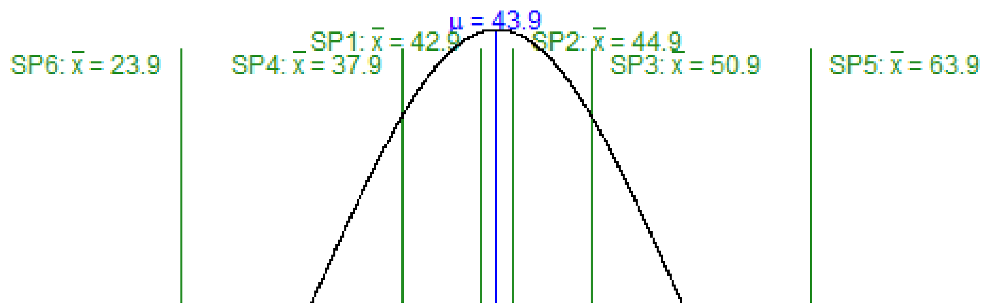
Die arithmetischen Mittel verschiedener Stichproben sind (mit zunehmender Anzahl an Beobachtungen n) normalverteilt um das arithmetische Mittel μ der Grundgesamtheit!

Stichprobenkennwerteverteilung Alter

Beispiel: Zentrales Grenzwerttheorem

JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN

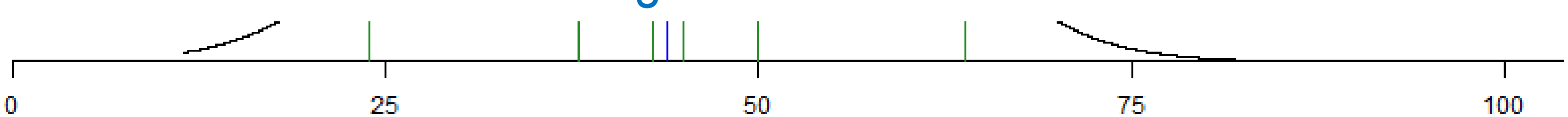
- Arithmetisches Mittel des Alters der dt. Bevölkerung (μ) ist 43,9 Jahre (vgl. (Destatis Zensus 2011: https://www.destatis.de/DE/ZahlenFakten/ImFokus/Bevoelkerung/AltersstrukturZensus.html)



Wenn n>30, Normalverteilung akzeptable Approximation, je mehr n umso sicherer NV approximativ.

Die arithmetischen Mittel verschiedener Stichproben sind (mit zunehmender Anzahl an Beobachtungen n) normalverteilt um das arithmetische Mittel μ der Grundgesamtheit!

Zentrales Grenzwerttheorem!



Stichprobenkennwerteverteilung Alter

Standardfehler des Mittels

JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN

- Standardabweichung der Stichprobenmittelwerte als **Standardfehler der Stichprobenmittelwerte** oder Standardfehler des Mittels (kurz: **Standardfehler, $\boldsymbol{\sigma_{\bar{X}}}$**)
- Durchschnittliche Streuung der arithmetischen Mittel
- Informiert darüber, wie präzise ein Stichprobenmittelwert den Populationsmittelwert schätzt
- Informiert über die Größe der Diskrepanz zwischen einem Stichprobenmittelwert $\bar{x}$ und dem Populationsmittelwert μ
- Englische Bezeichnung: Standard Error (S.E.)

Standardfehler des Mittels

JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN

- Ein relativ kleiner Standardfehler bedeutet, dass die Stichprobenmittelwerte alle relativ ähnlich sind, d.h. grafisch wenig streuen
- Ein relativ großer Standardfehler bedeutet, dass die Stichprobenmittelwerte alle relativ unähnlich sind, d.h. stärker streuen
- Je größer der Standardfehler desto unsicherer die Schätzung
- Formal: $\boldsymbol{\sigma}_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$

Beispiel: Standardfehler

- Gegeben sei für ein interessierendes Merkmal eine Population mit der Standardabweichung σ= 10.
- Wie groß ist die durchschnittliche Abweichung zwischen dem Mittelwert einer Stichprobe für n= 4 zufällig aus dieser Population ausgewählten Beobachtungseinheiten und dem Populationsmittelwert?

Beispiel: Standardfehler

- Gegeben sei für ein interessierendes Merkmal eine Population mit der Standardabweichung σ= 10.
- Wie groß ist die durchschnittliche Abweichung zwischen dem Mittelwert einer Stichprobe für n= 4 zufällig aus dieser Population ausgewählten Beobachtungseinheiten und dem Populationsmittelwert?

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

Beispiel: Standardfehler

JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

- Gegeben sei für ein interessierendes Merkmal eine Population mit der Standardabweichung σ= 10.
- Wie groß ist die durchschnittliche Abweichung zwischen dem Mittelwert einer Stichprobe für n= 4 zufällig aus dieser Population ausgewählten Beobachtungseinheiten und dem Populationsmittelwert? → S.E. =5

Übung: Standardfehler

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

- Gegeben sei für ein interessierendes Merkmal eine Population mit der Standardabweichung σ= 10.
- Wie groß ist die durchschnittliche Abweichung zwischen dem Mittelwert einer Stichprobe für n= 25 zufällig aus dieser Population ausgewählten Beobachtungseinheiten und dem Populationsmittelwert?

Standardfehler

- Durch welche Faktoren wird der Standardfehler beeinflusst?

1. Varianz des Merkmals in der Grundgesamtheit

  - Je größer die Varianz des Merkmales in der Population, desto größer ist der Standardfehler der Stichprobenmittelwerte.

Standardfehler

JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN

- Durch welche Faktoren wird der Standardfehler noch beeinflusst?

Stichprobengröße (n)	Standardfehler	
1	$\sigma_{\bar{X}} = \frac{10}{\sqrt{1}}$	= 10
9	$\sigma_{\bar{X}} = \frac{10}{\sqrt{9}}$	= 3.33
25	$\sigma_{\bar{X}} = \frac{10}{\sqrt{25}}$	= 2
100	$\sigma_{\bar{X}} = \frac{10}{\sqrt{100}}$	= 1

Standardfehler

JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN

- Durch welche Faktoren wird der Standardfehler beeinflusst?

2. Stichprobenumfang

- **„Gesetz der großen Zahl“:** Je größer der Stichprobenumfang, desto kleiner ist der Standardfehler, denn:
  - mit steigendem Stichprobenumfang wird die Informationsunsicherheit über die Grundgesamtheit reduziert

Standardfehler

JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN

- Zusammenhang ist negativ: Je größer die Stichprobe, desto kleiner der Standardfehler
- Zusammenhang ist monoton, aber nicht-linear

Stichprobengröße (*n*)	Standardfehler	
1	$\sigma_{\bar{X}} = \frac{10}{\sqrt{1}}$	= 10
9	$\sigma_{\bar{X}} = \frac{10}{\sqrt{9}}$	= 3.33
25	$\sigma_{\bar{X}} = \frac{10}{\sqrt{25}}$	= 2
100	$\sigma_{\bar{X}} = \frac{10}{\sqrt{100}}$	= 1

Fazit Standardfehler

- Durchschnittliche Streuung aller Stichprobenmittelwerte
- Maß für die Genauigkeit des Stichprobenmittelwerts
- Interpretation?
  - Je kleiner desto besser (da präziser)
  - (Standard-)Normalverteilung als „Hilfe“ für Berechnung von Wahrscheinlichkeiten

Übungen

Gegeben sei eine Grundgesamtheit mit **μ= 50 und σ= 12.**

a) Wie lautet der Standardfehler für die Verteilung der Stichprobenmittelwerte für eine Stichprobengröße von n= 4?

b) Angenommen die Verteilung in der Population sei nicht normalverteilt; welche Form wäre dann bei n = 4 für die Verteilung der Stichprobenmittelwerte zu erwarten?

c) Wie lautet der Mittelwert und der Standardfehler für die Verteilung der Stichprobenmittelwerte für eine Stichprobengröße von n= 36?

d) Angenommen die Verteilung in der Population sei nicht normalverteilt; welche Form wäre dann bei n = 36 für die Verteilung der Stichprobenmittelwerte zu erwarten?

Übungen

JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN

Gegeben sei eine Population mit **μ= 50 und σ= 12.**

a) Wie lautet der Erwartungswert und der Standardfehler für die Verteilung der Stichprobenmittelwerte für eine Stichprobengröße von *n*= 4?

$$\mu = 50; \sigma_{\bar{X}} = \frac{12}{\sqrt{4}} = 6$$

b) Angenommen die Verteilung in der Population sei nicht normalverteilt; welche Form wäre dann bei *n* = 4 für die Verteilung der Stichprobenmittelwerte zu erwarten? Keine Normalverteilung.

c) Wie lautet der Mittelwert und der Standardfehler für die Verteilung der Stichprobenmittelwerte für eine Stichprobengröße von *n*= 36?

d) Angenommen die Verteilung in der Population sei nicht normalverteilt; welche Form wäre dann bei *n* = 36 für die Verteilung der Stichprobenmittelwerte zu erwarten?

Übungen

Gegeben sei eine Population mit **μ= 50 und σ= 12.**

a) Wie lautet der Erwartungswert und der Standardfehler für die Verteilung der Stichprobenmittelwerte für eine Stichprobengröße von *n*= 4?

$$\mu = 50; \sigma_{\bar{X}} = \frac{12}{\sqrt{4}} = 6$$

b) Angenommen die Verteilung in der Population sei nicht normalverteilt; welche Form wäre dann bei *n* = 4 für die Verteilung der Stichprobenmittelwerte zu erwarten? Keine Normalverteilung.

c) Wie lautet der (erwartete) Mittelwert und der Standardfehler für die Verteilung der Stichprobenmittelwerte für eine Stichprobengröße von *n*= 36?

$$\mu = 50; \sigma_{\bar{X}} = \frac{12}{\sqrt{36}} = 2$$

d) Angenommen die Verteilung in der Population sei nicht normalverteilt; welche Form wäre dann bei *n* = 36 für die Verteilung der Stichprobenmittelwerte zu erwarten?

Übungen

Gegeben sei eine Population mit **μ= 50 und σ= 12.**

a) Wie lautet der Erwartungswert und der Standardfehler für die Verteilung der Stichprobenmittelwerte für eine Stichprobengröße von *n*= 4?

$$\mu = 50; \sigma_{\bar{X}} = \frac{12}{\sqrt{4}} = 6$$

b) Angenommen die Verteilung in der Population sei nicht normalverteilt; welche Form wäre dann bei *n* = 4 für die Verteilung der Stichprobenmittelwerte zu erwarten? Keine Normalverteilung.

c) Wie lautet der Mittelwert und der Standardfehler für die Verteilung der Stichprobenmittelwerte für eine Stichprobengröße von *n*= 36?

$$\mu = 50; \sigma_{\bar{X}} = \frac{12}{\sqrt{36}} = 2$$

d) Angenommen die Verteilung in der Population sei nicht normalverteilt; welche Form wäre dann bei *n* = 36 für die Verteilung der Stichprobenmittelwerte zu erwarten? Normalverteilung

Stichprobenmittelwerte, Standardfehler und Wahrscheinlichkeit

Beispiel

- Arithmetisches Mittel des Alters der Bevölkerung (μ) ist 43,9 Jahre
- Wie hoch ist die Wahrscheinlichkeit einen Stichprobenmittelwert $\bar{x} = 32$ Jahre zufällig zu ziehen?

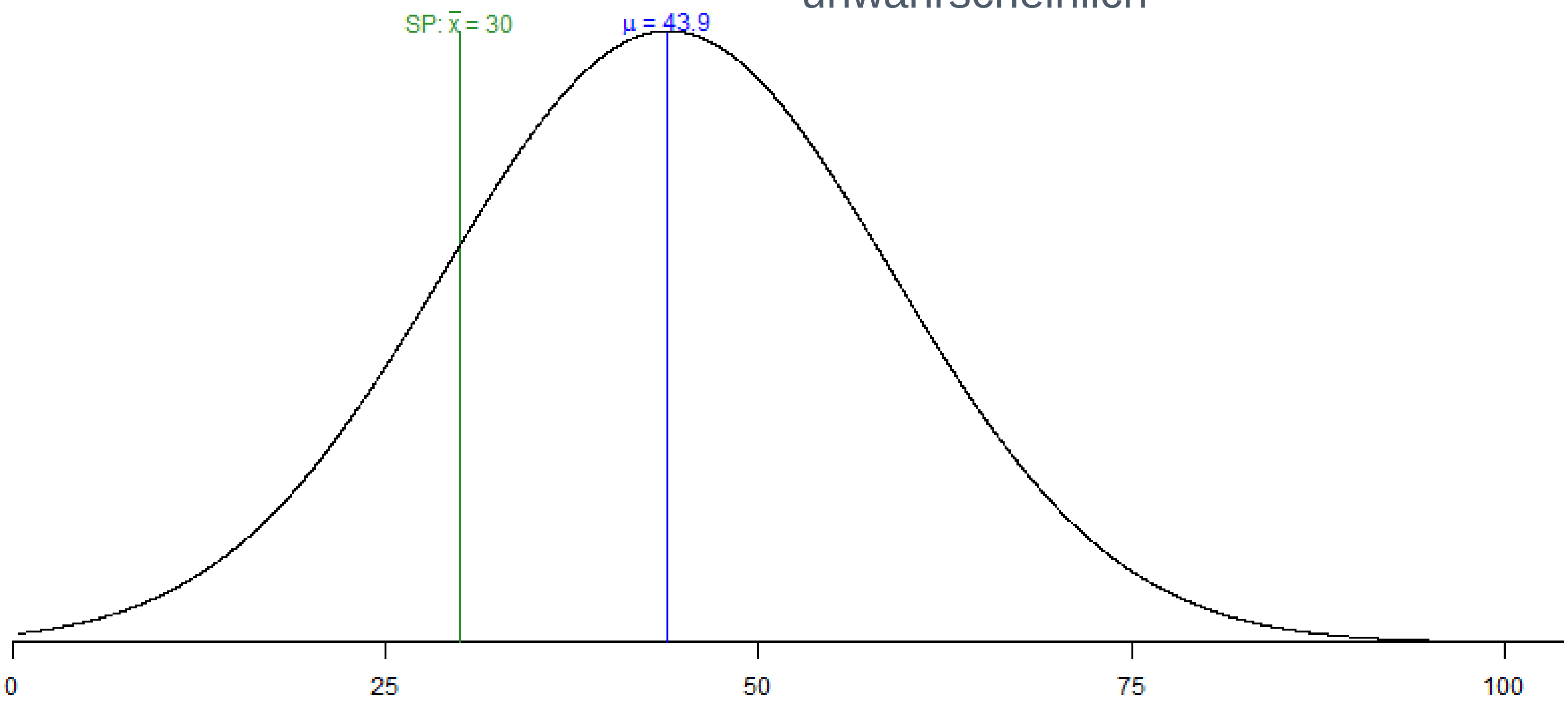
Stichprobenmittelwerte, Standardfehler und Wahrscheinlichkeit

- Hängt vom Standardfehler $\sigma_{\bar{x}}$ ab
- Arithmetische Mittel des Alters der Bevölkerung (μ) ist 43,9 Jahre
- Wie hoch ist die Wahrscheinlichkeit einen Stichprobenmittelwert $\bar{x} = 32$ Jahre zufällig zu ziehen?
- 3 verschiedene Streuungsbeispiele:
  - $\sigma_{\bar{x}} = 15$ Jahre, $\bar{x} = 32$ Jahre
  - $\sigma_{\bar{x}} = 10$ Jahre, $\bar{x} = 32$ Jahre
  - $\sigma_{\bar{x}} = 5$ Jahre, $\bar{x} = 32$ Jahre

Stichprobenkennwerteverteilung

- $\mu = 43{,}9$ Jahre und $\sigma_{\bar{x}} = 15$

$\bar{x} = 32$ Jahre nicht besonders unwahrscheinlich



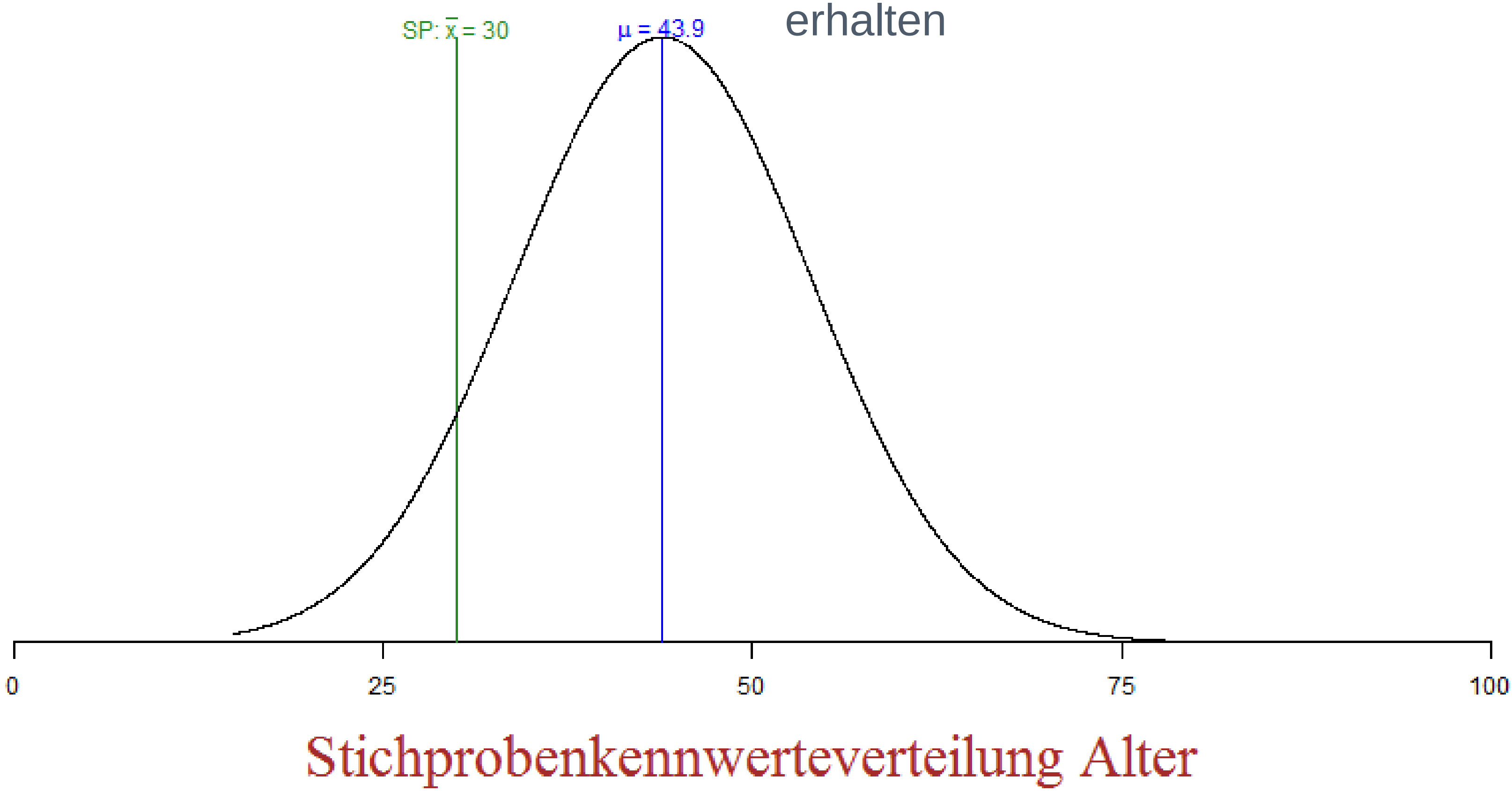
Stichprobenkennwerteverteilung Alter

Stichprobenkennwerteverteilung

JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN

- $\mu = 43{,}9$ Jahre und $\sigma_{\bar{x}} = 10$

$\bar{x} = 32$ Jahre schon „besonderer" zu erhalten



Stichprobenkennwerteverteilung

- $\mu = 43{,}9$ Jahre und $\sigma_{\bar{x}} = 5$

$\bar{x} = 32$ Jahre sehr unwahrscheinlich