Statistik I

Prof. Dr. Simone Abendschön

Letzte Sitzung am 8. Februar 2024

Plan heute

Grundlagen der Inferenzstatistik

- Zentrales Grenzwerttheorem
- (Standardfehler → SoSe 24)
- Wiederholung Inhalte mit Übungen

Lernziele heute

JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN

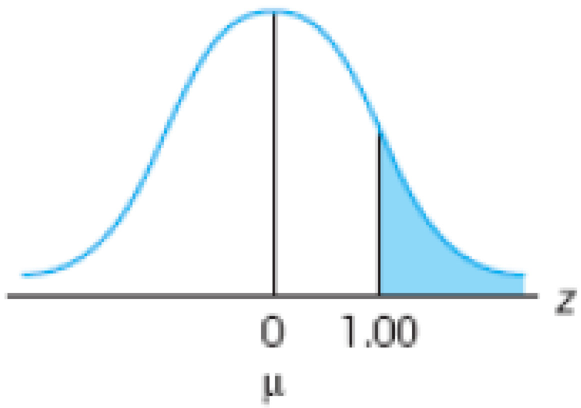
- Kennen und Verstehen des Zentralen Grenzwerttheorems

Einführung

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

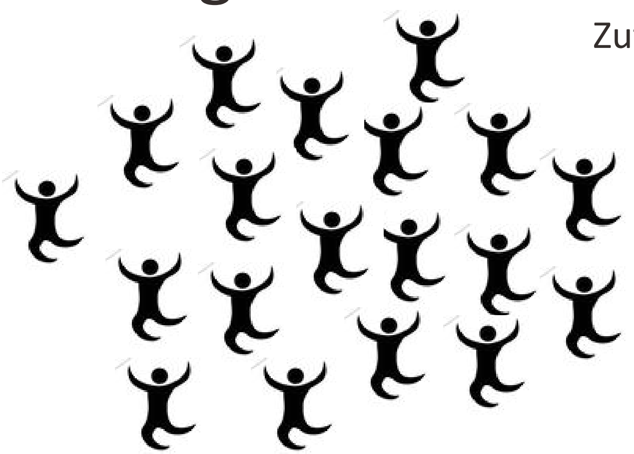
Wie groß ist die Wahrscheinlichkeit per Zufallsauswahl bei gegebenem Mittelwert und Standardabweichung einen Fall in einem bestimmten Werteintervall auszuwählen ?

Wdh.



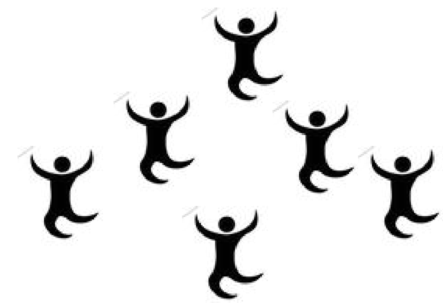
Welcher Flächenanteil der Normalverteilung entspricht z-Werten >1?
Wie groß ist die Wahrscheinlichkeit, für normalverteilte Werte einen z-Wert > 1.0 zu erhalten?

Grundgesamtheit



Zufallsauswahl/Wahrscheinlichkeit

Stichprobe



μ

(Erwartungswert – „Durchschnitt der Grundgesamtheit“)

Inferenz

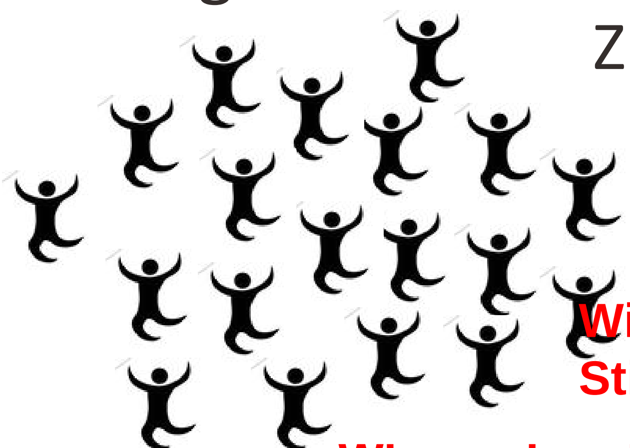
Schätzung

$\bar{x}$ Statistik

(Arithmetisches Mittel Stichprobe)

Inferenzstatistik

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“)

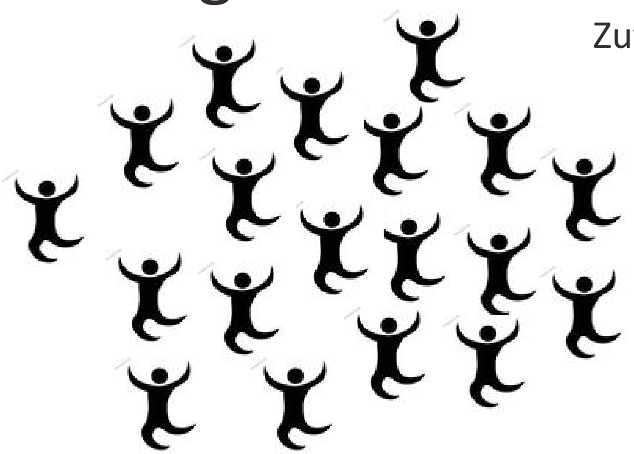
Inferenz

$\bar{x}$ Statistik

(Arithmetisches Mittel Stichprobe)

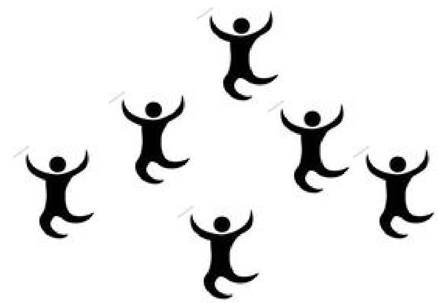
Inferenzstatistik

Grundgesamtheit



Zufallsauswahl/Wahrscheinlichkeit

Stichprobe



Grundannahmen über die Verteilung von Stichprobenkennwerten

μ

(Erwartungswert – „Durchschnitt der Grundgesamtheit“)

Inferenz

Schätzung

$\bar{x}$ Statistik

(Arithmetisches Mittel Stichprobe)

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-Werte transformiert und für Wahrscheinlichkeitsaussagen genutzt werden

Stichproben und Grundgesamtheit

JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN

- 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

- 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:

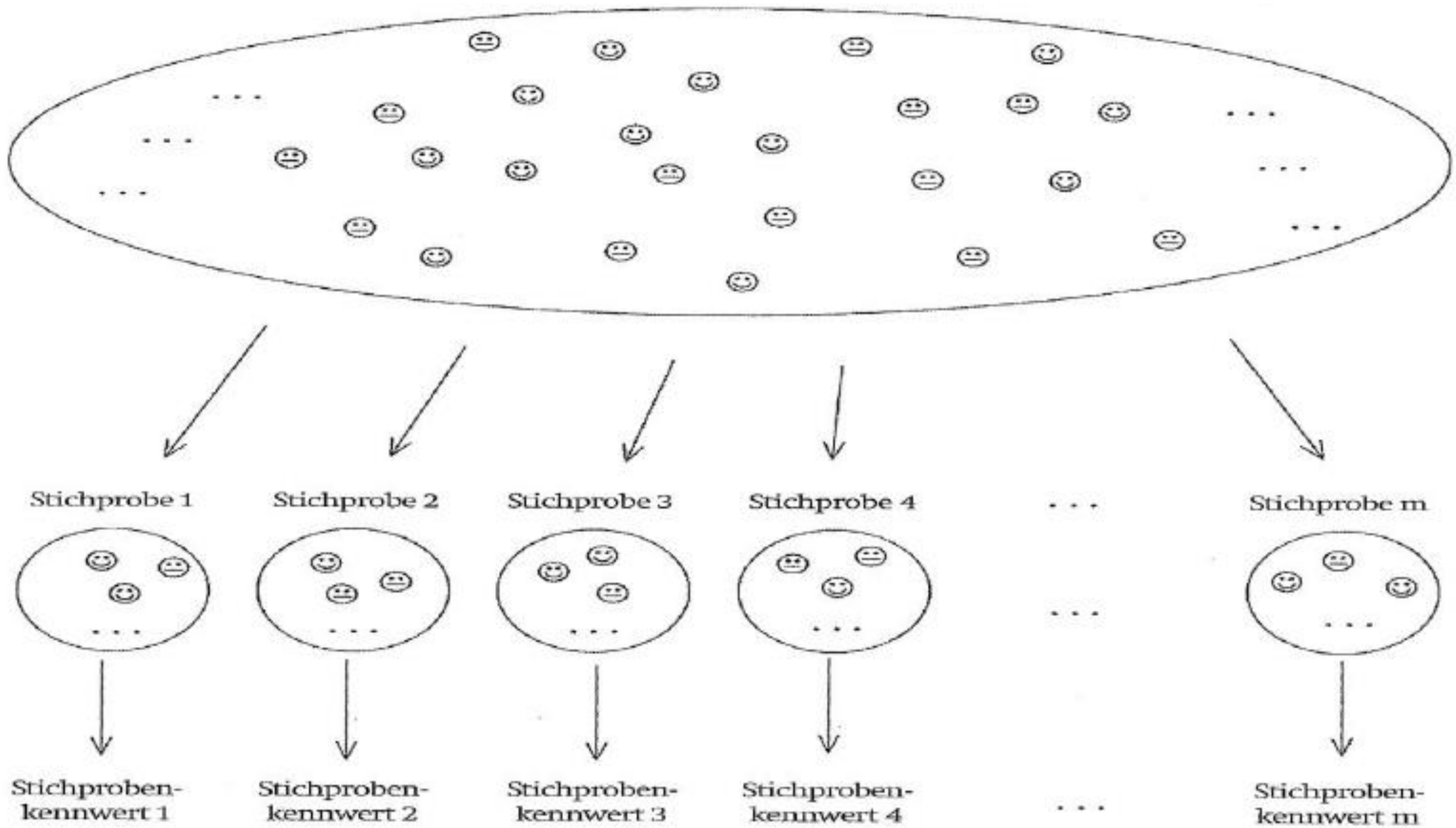
- Eine Stichprobenkennwerteverteilung für unendlich viele Stichproben von Mittelwerten nähert sich der Normalverteilung an, falls die Stichprobe ausreichend groß ist (*n*>= 30) oder die Werte in der GG normalverteilt sind
- Der Erwartungswert E der Stichprobenmittelwerte entspricht dem „wahren" Mittelwert der GG
- μ: $E(\bar{x}) = \mu$

Stichprobenkennwerteverteilung

„Nachweis" über Simulation

- 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



Simulationsbeispiel

- 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

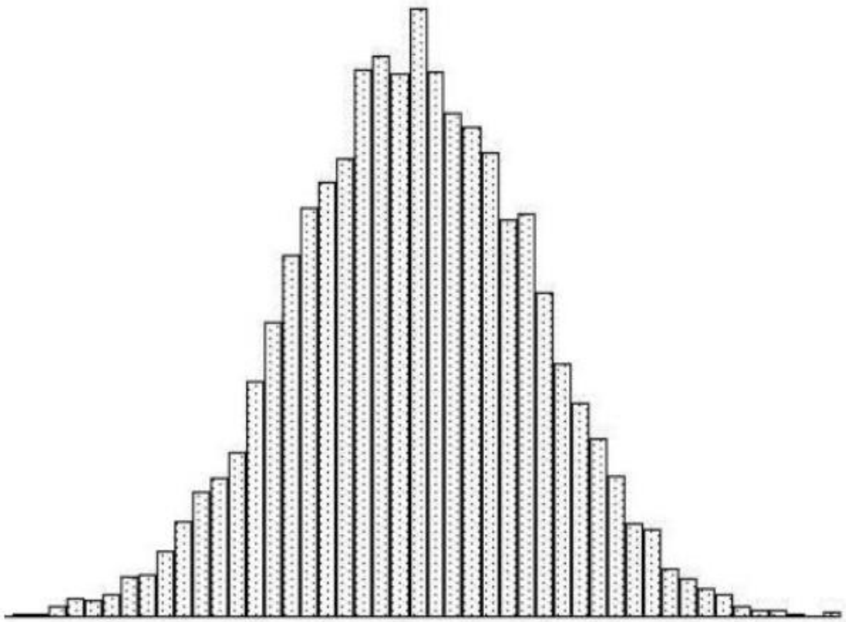
Simulationsbeispiel

JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN

- 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:

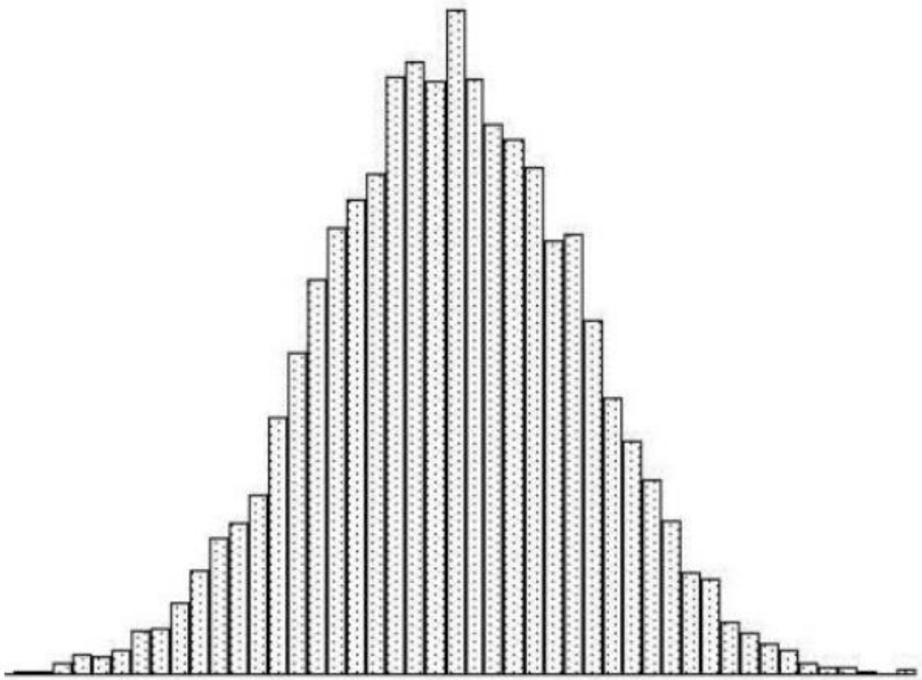


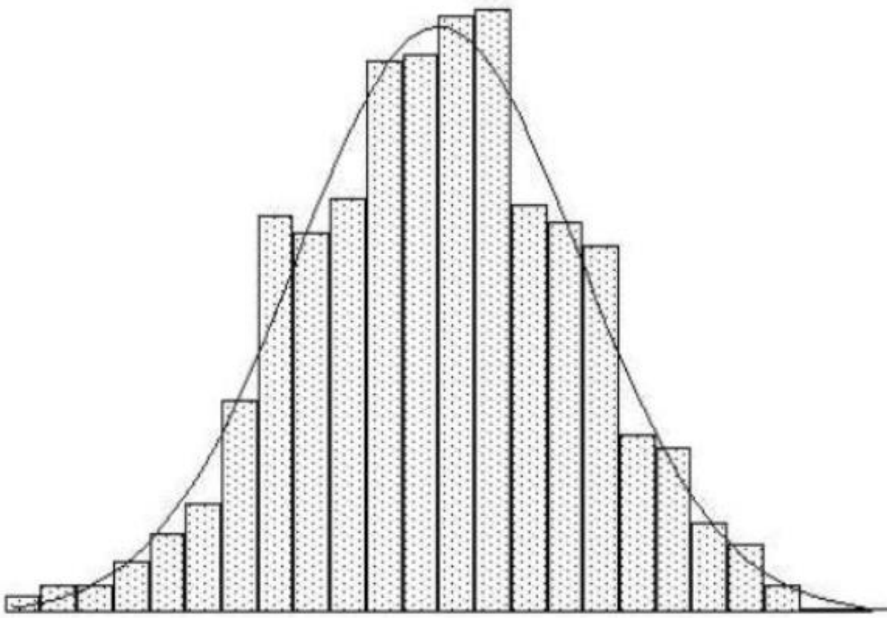
Normalverteilung

Simulationsbeispiel

JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN

Population:	Verteilung der Stichprobenmittelwerte:

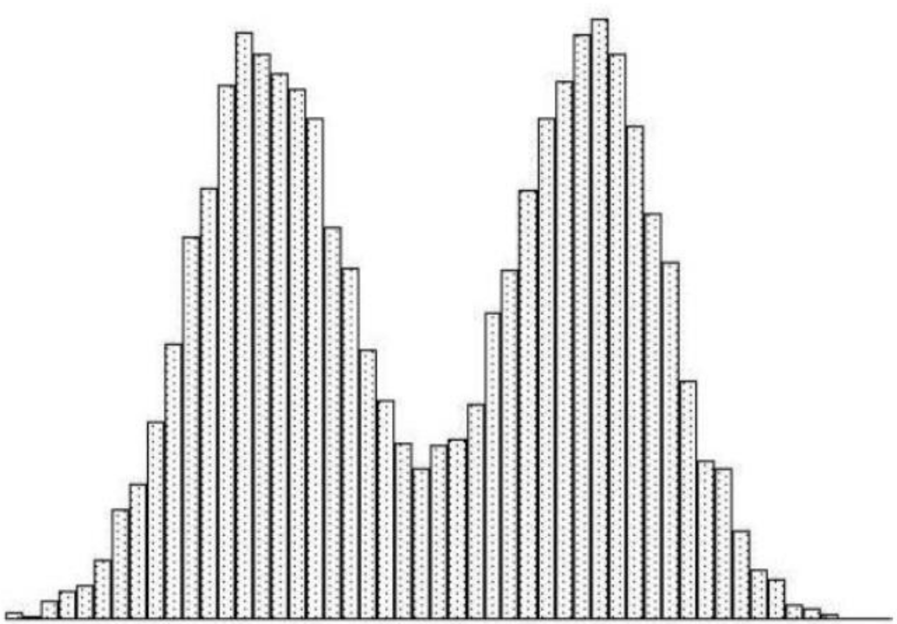




Normalverteilung

Simulationsbeispiel

Population:	Verteilung der Stichprobenmittelwerte:

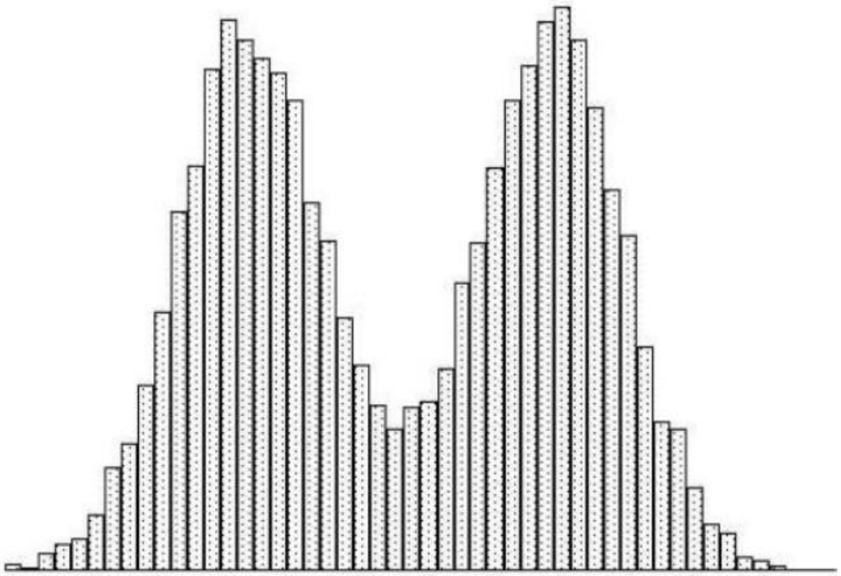


Bimodale Verteilung

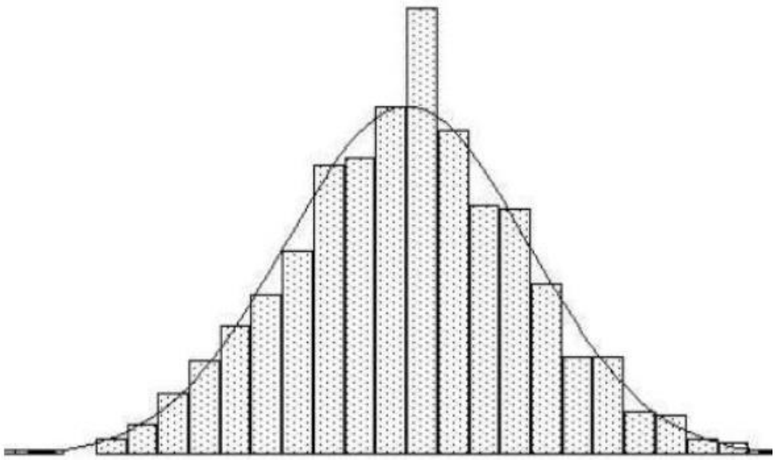
Simulationsbeispiel

JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN

Population:	Verteilung der Stichprobenmittelwerte:

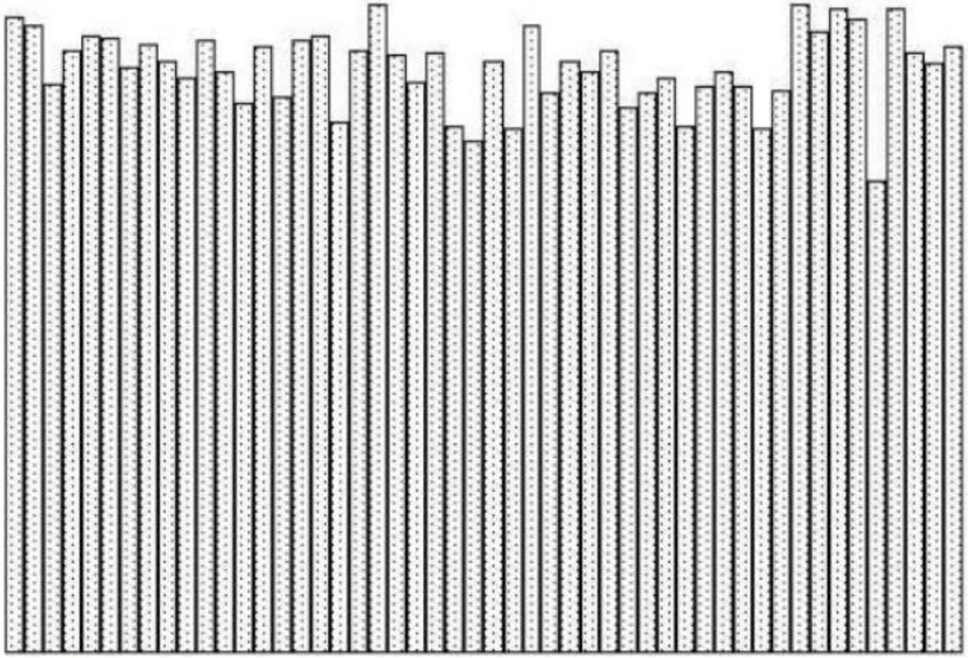


Bimodale Verteilung



Simulationsbeispiel

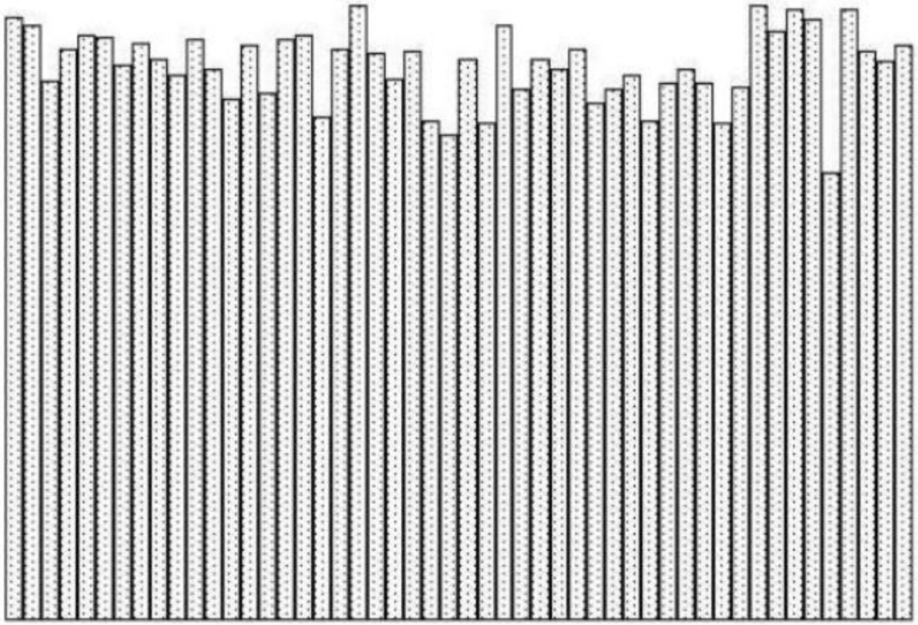
Population:	Verteilung der Stichprobenmittelwerte:



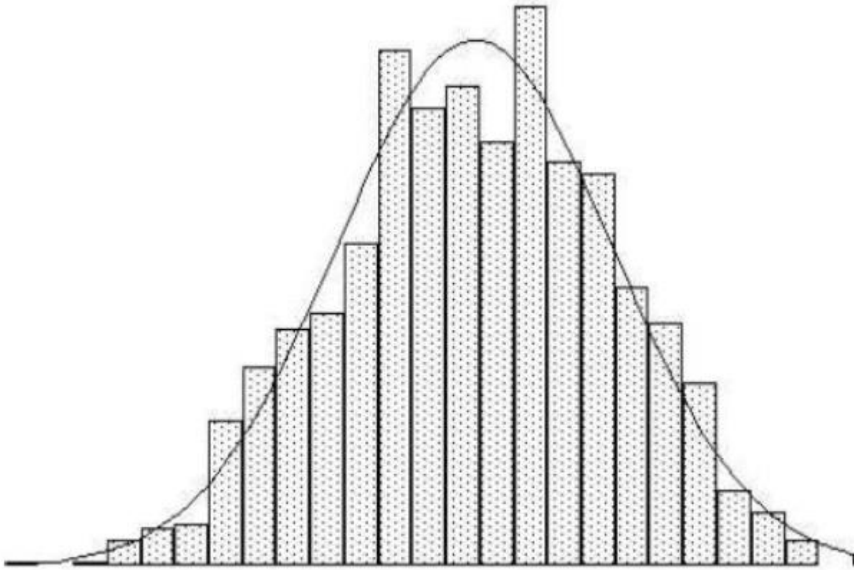
Gleichverteilung

Simulationsbeispiel

Population:	Verteilung der Stichprobenmittelwerte:



Gleichverteilung



Zentrales Grenzwerttheorem

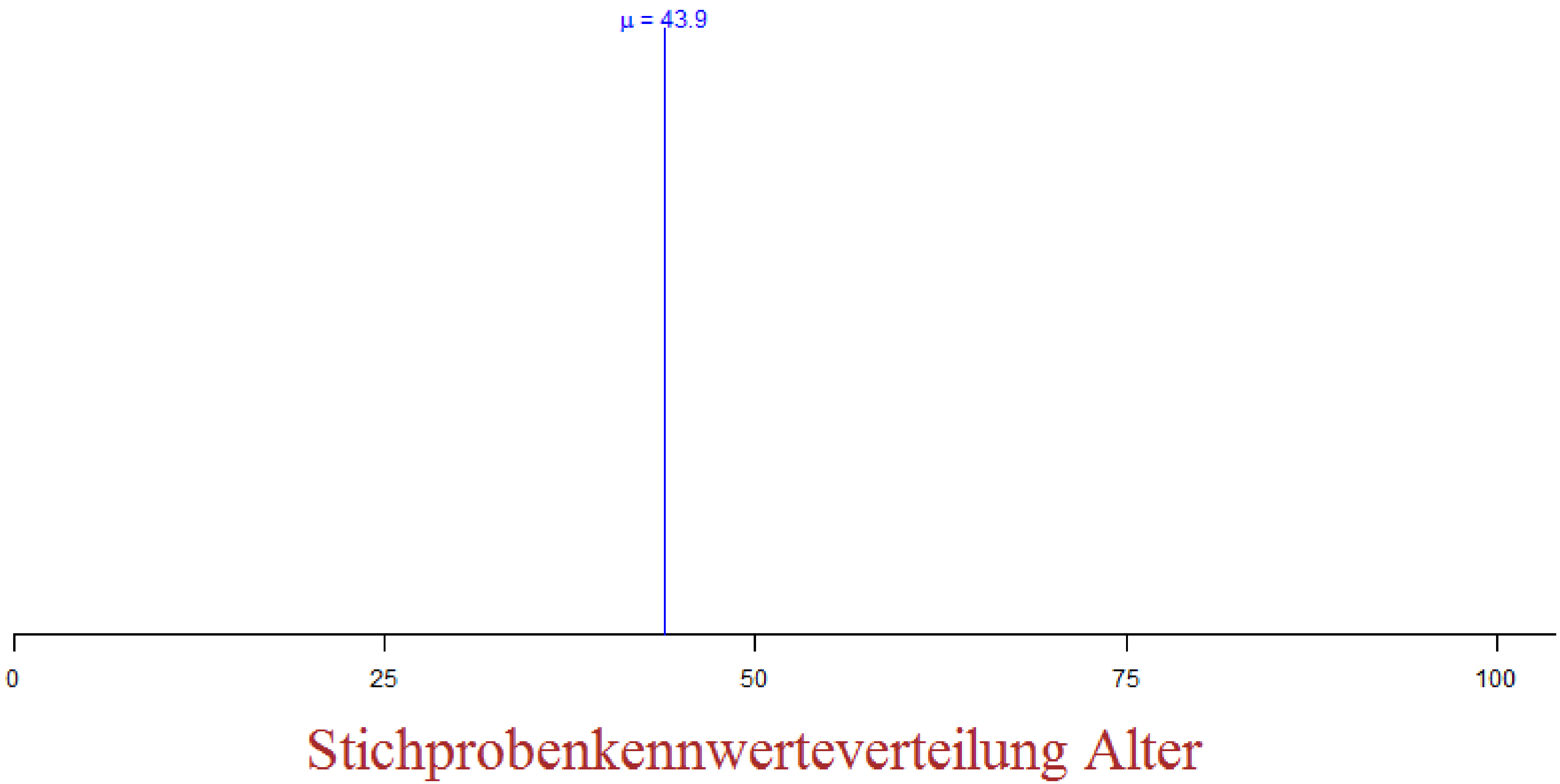
JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN

- 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

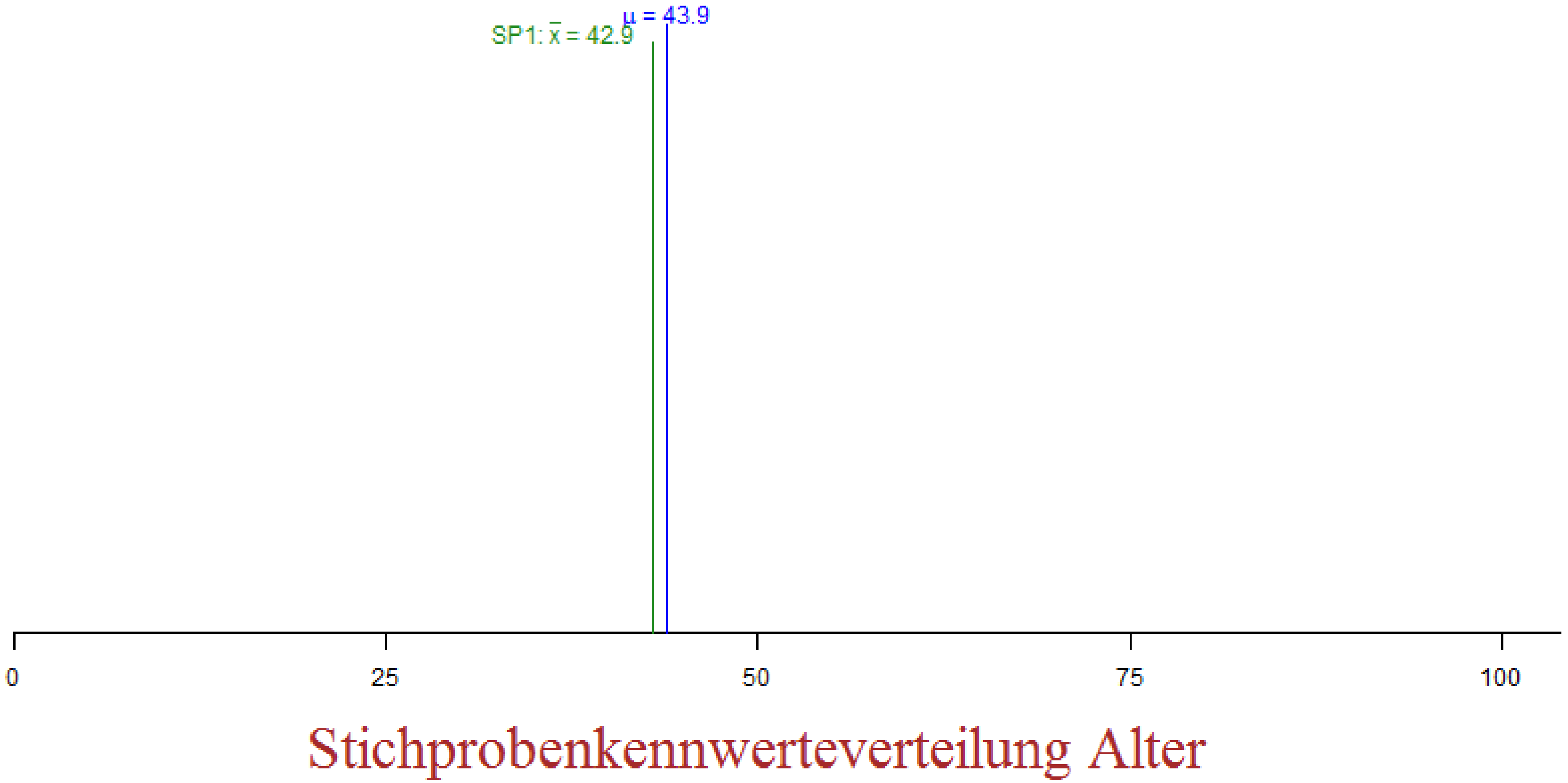
JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN

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)



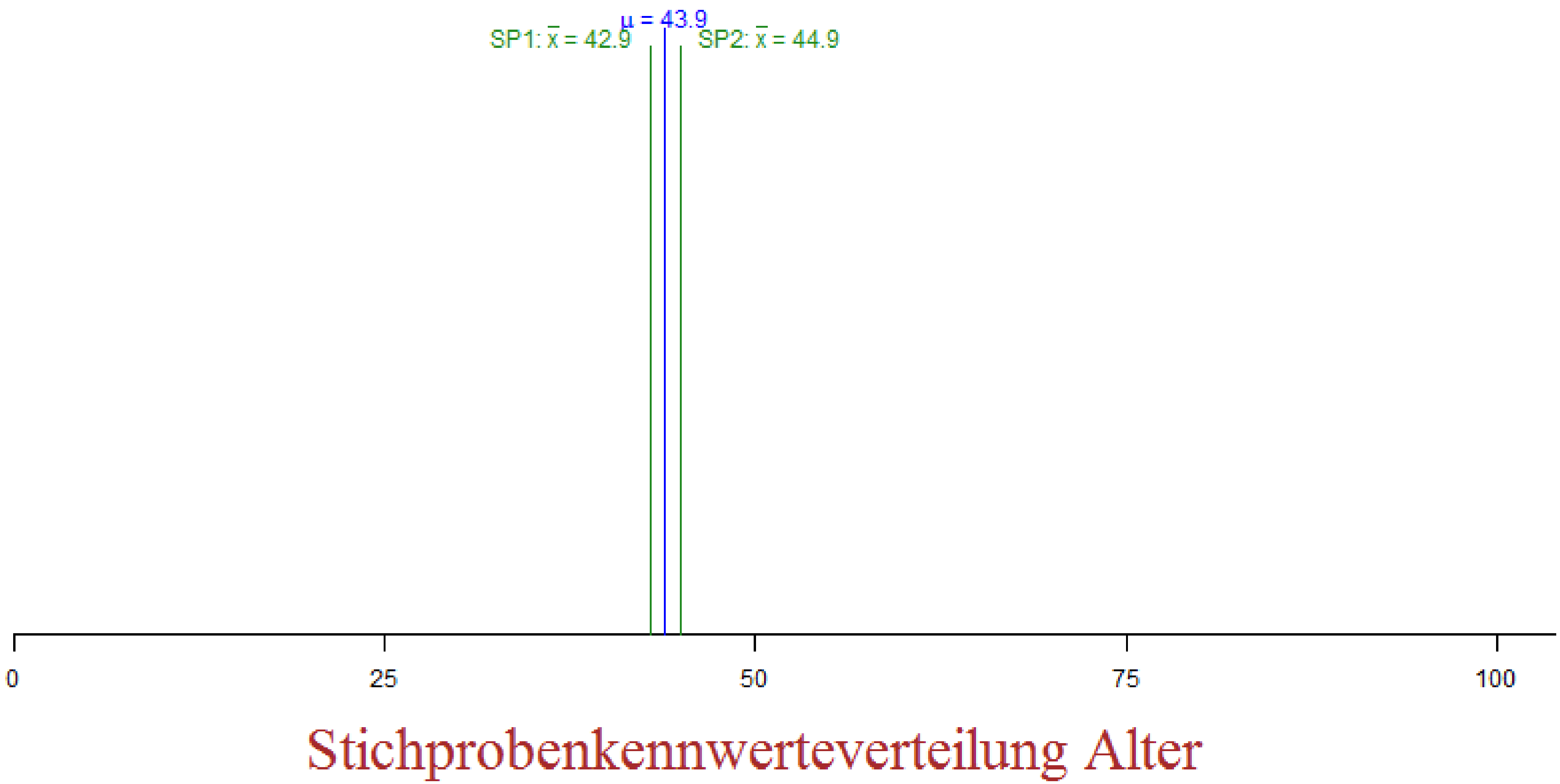
Beispiel: Zentrales Grenzwerttheorem

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)



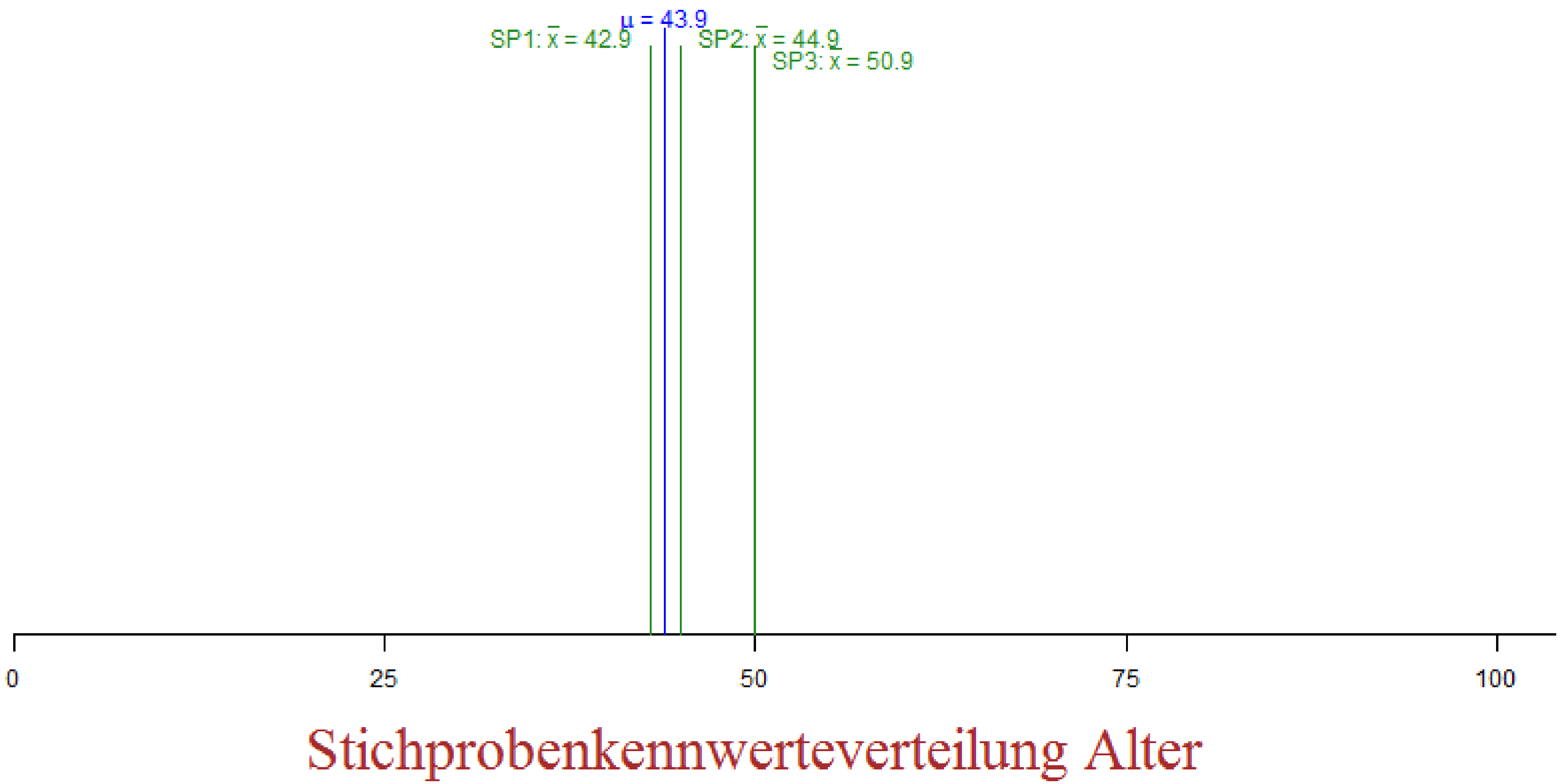
Beispiel: Zentrales Grenzwerttheorem

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)



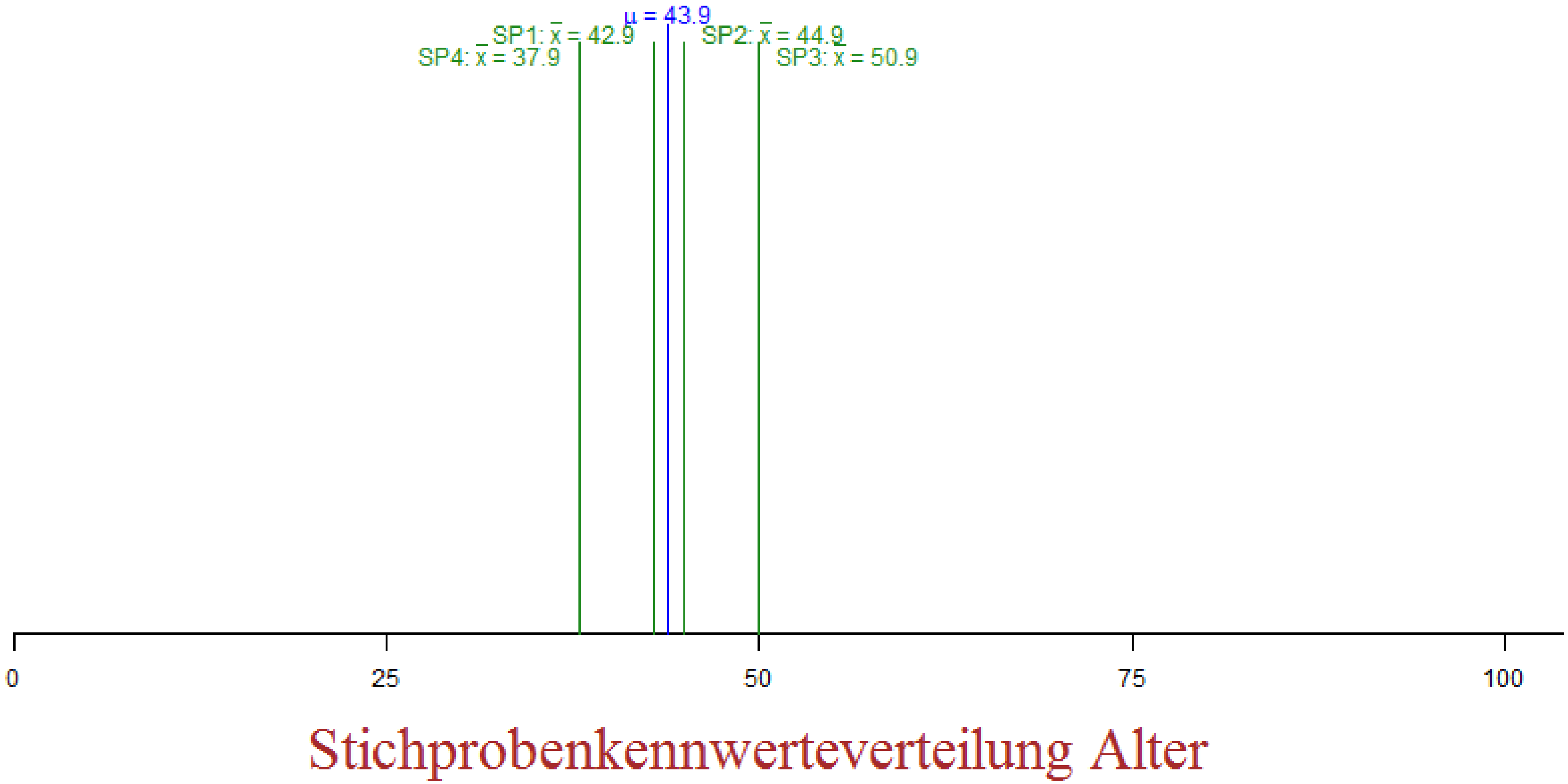
Beispiel: Zentrales Grenzwerttheorem

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)



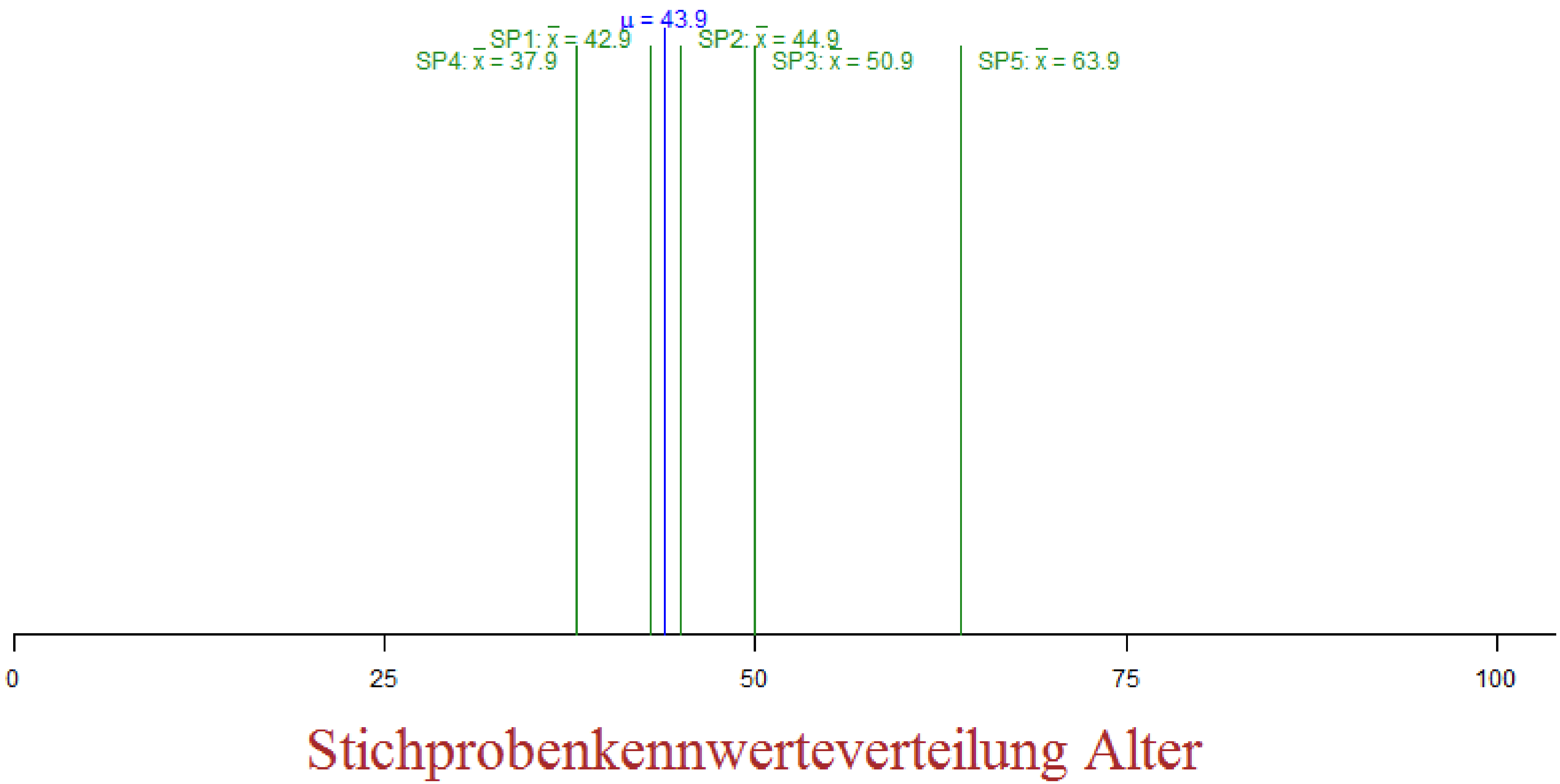
Beispiel: Zentrales Grenzwerttheorem

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)



Beispiel: Zentrales Grenzwerttheorem

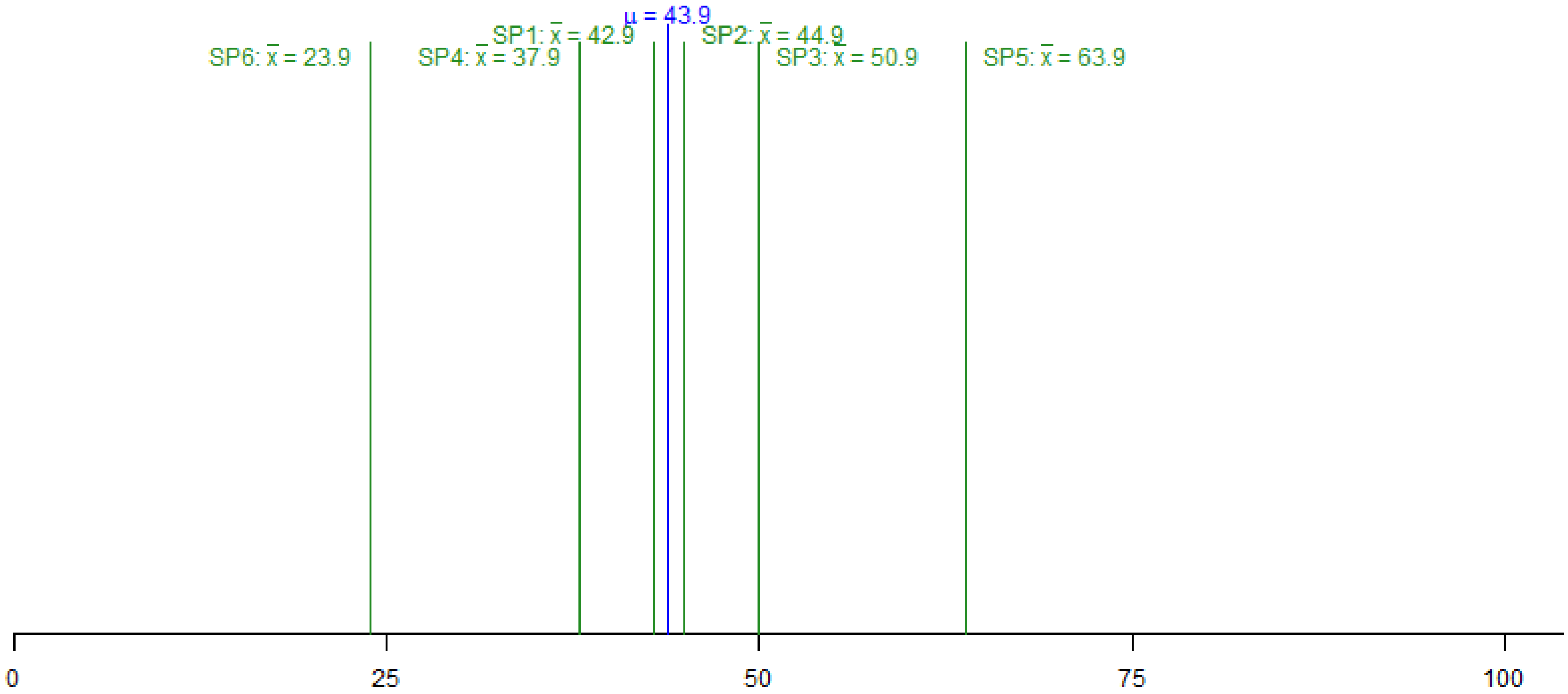
Arithmetisches Mittel des Alters der dt. Bevölkerung (μ) ist 43,9 Jahre (vgl. (Destatis Zensus 2011: https://www.destatis.de/DE/ZahlenFakten/ImFokus/Bevoelkerung/ AltersstrukturZensus.html)



Beispiel: Zentrales Grenzwerttheorem

JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN

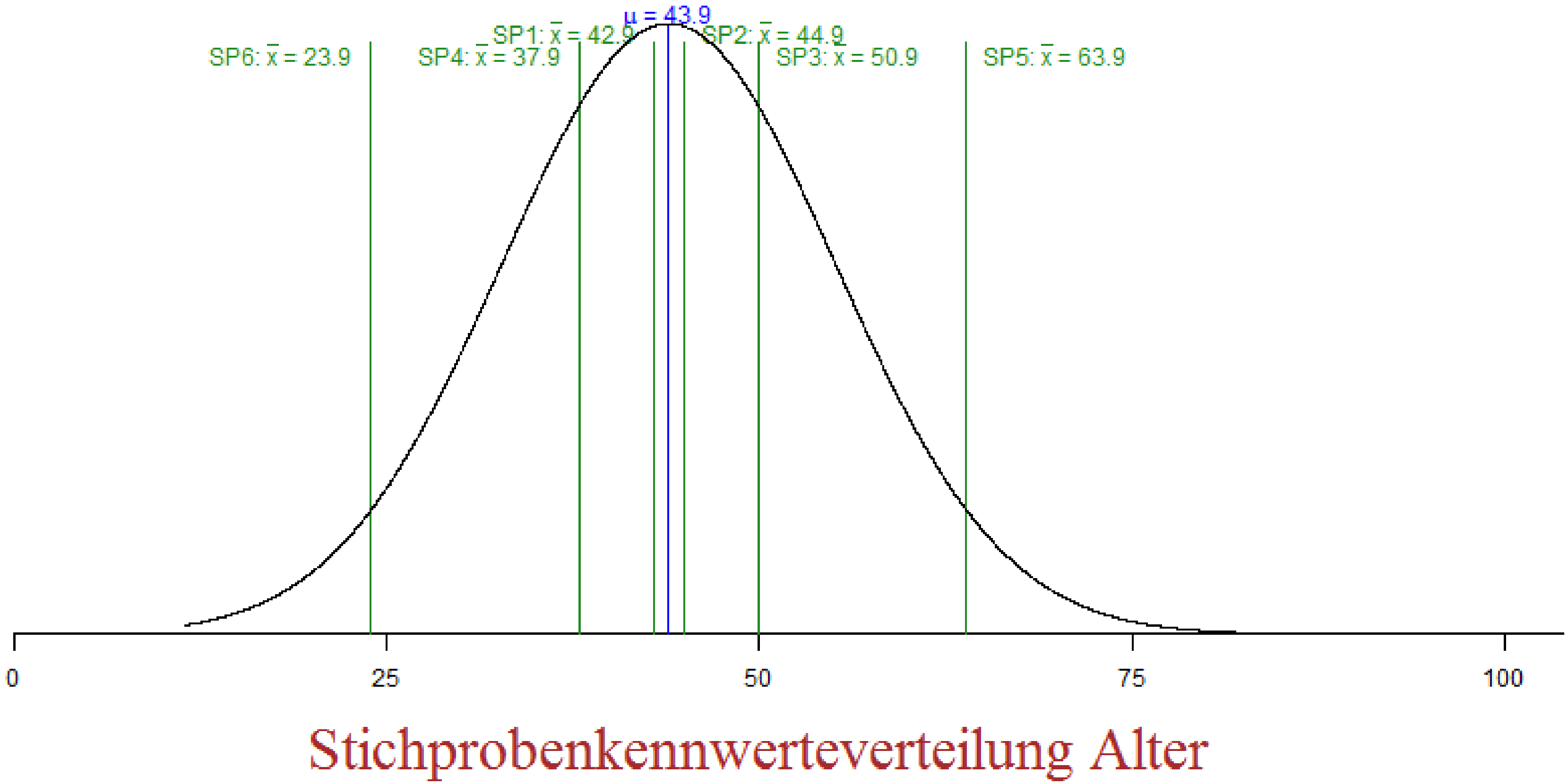
Arithmetisches Mittel des Alters der dt. Bevölkerung (μ) ist 43,9 Jahre (vgl. (Destatis Zensus 2011: https://www.destatis.de/DE/ZahlenFakten/ImFokus/Bevoelkerung/ AltersstrukturZensus.html)



Stichprobenkennwerteverteilung Alter

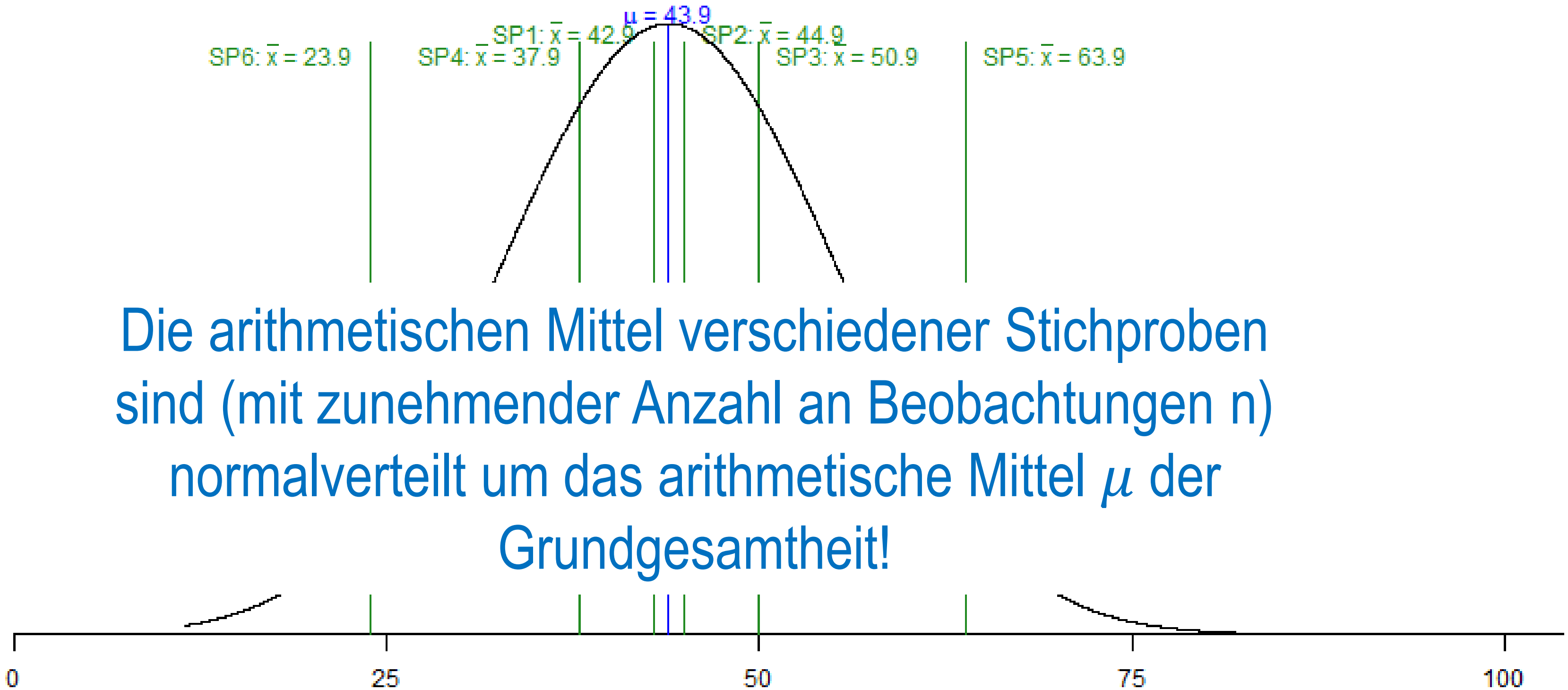
Beispiel: Zentrales Grenzwerttheorem

- 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)



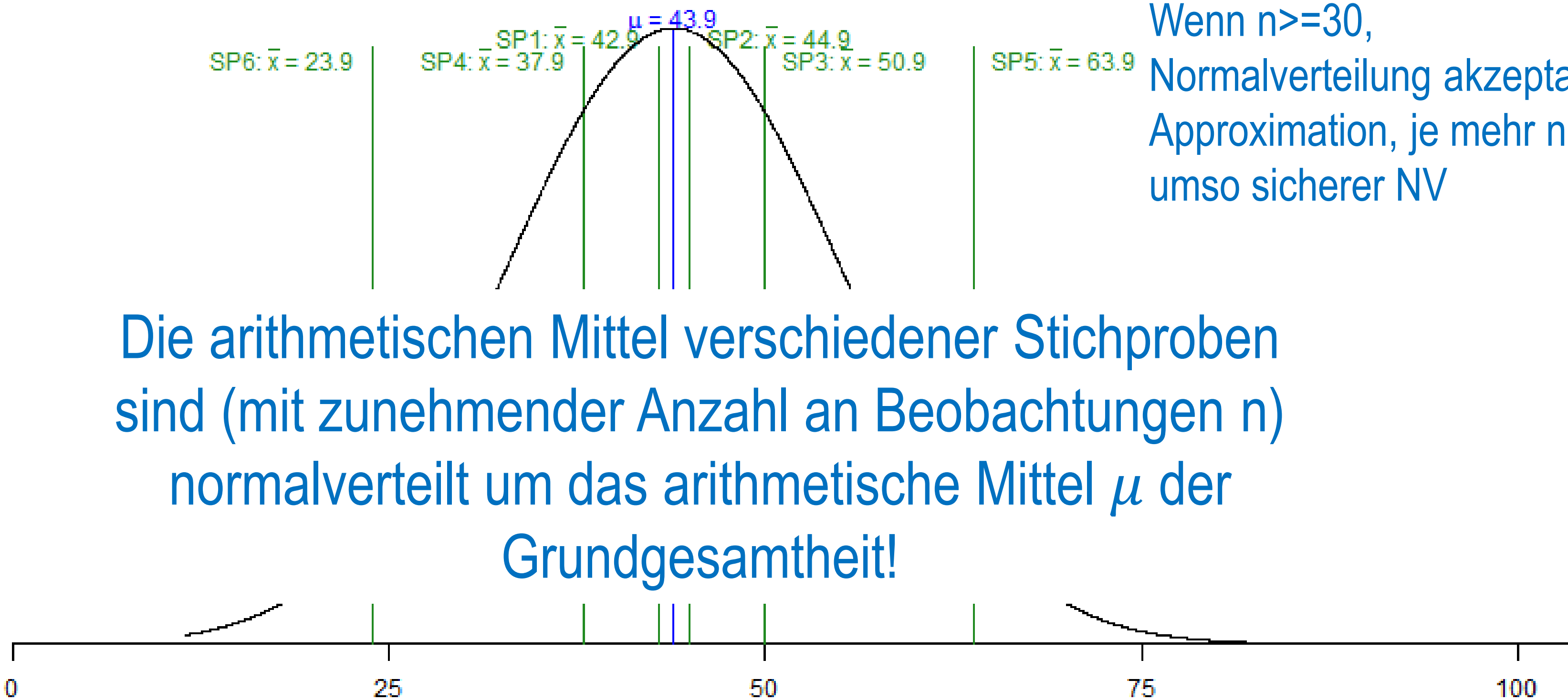
Die arithmetischen Mittel verschiedener Stichproben sind (mit zunehmender Anzahl an Beobachtungen n) normalverteilt um das arithmetische Mittel μ der Grundgesamtheit!

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)



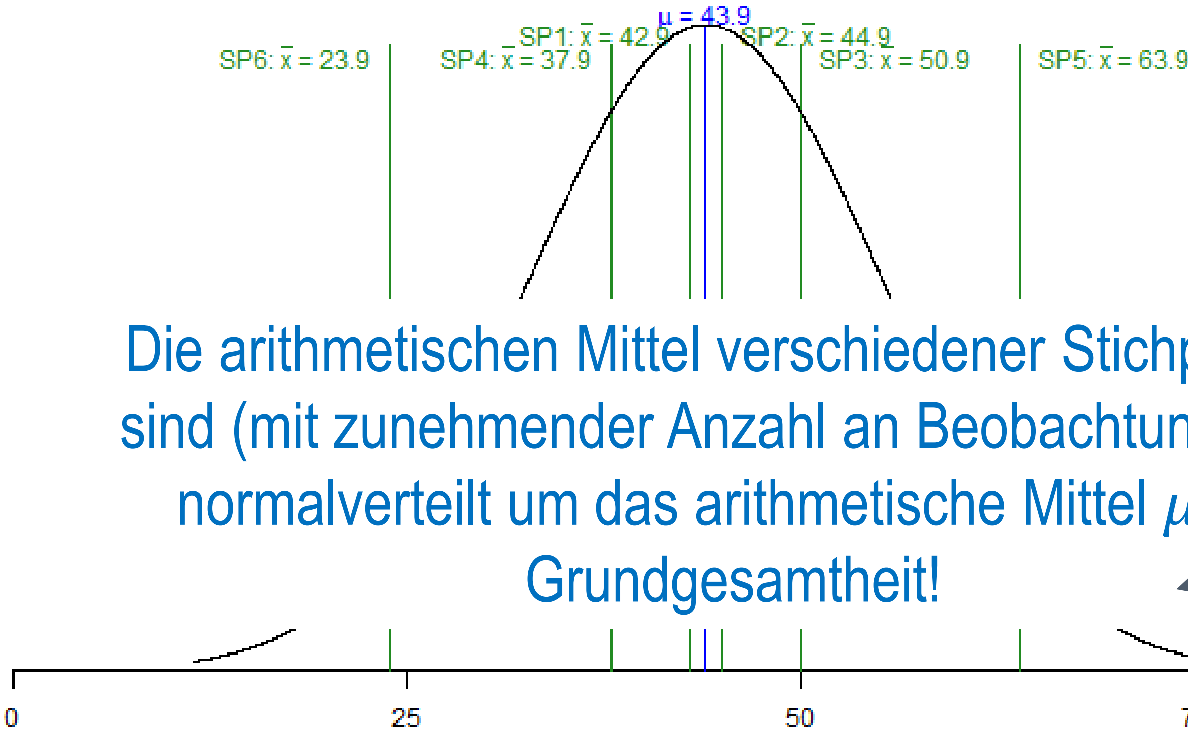
Wenn n>=30, Normalverteilung akzeptable Approximation, je mehr n umso sicherer NV

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

- 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

- 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
- Fortsetzung folgt…

Plan heute

Grundlagen der Inferenzstatistik

- Zentrales Grenzwerttheorem
- (Standardfehler → SoSe 24)
- Wiederholung Inhalte mit Übungen

Wiederholungsquiz

Quiz über Pingo

https://pingo.coactum.de

Zugang: **017613**

QR Code #017613

pingo.coactum.de → 017613