Zentraler Grenzwertsatz/Konfidenzintervalle

Statistik I

Sommersemester 2009

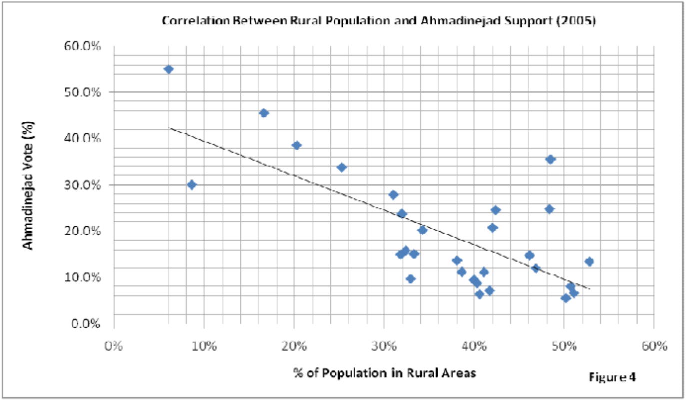
Kann Ahmadinejad die Wahl gewonnen haben?

- Im wesentlichen Dreiteilung der polit. Elite
- 2005: 17.3 Millionen Stimmen (Stichwahl), Wahlbeteiligung 48%
- 2009: 24.5 Millionen Stimmen, Wahlbeteiligung ca. 83%
- Anders als 2005 kaum Unterschiede in Wahlbeteiligung zwischen den 30 Provinzen
- Woher kommt der Zuwachs von 7.2 Millionen Stimmen?

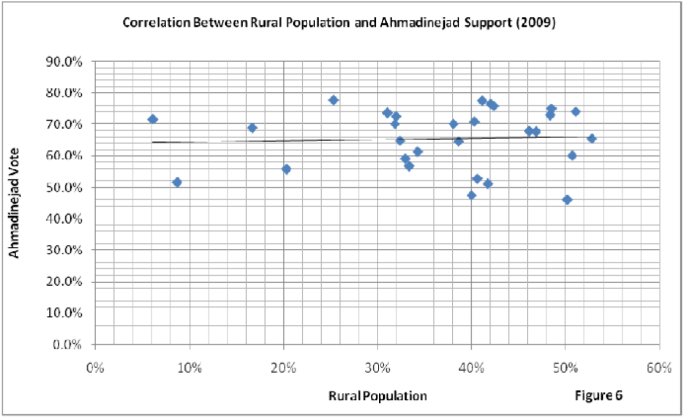
Was ist problematisch?

- Etwas über 100% Wahlbeteiligung in zwei konservativen Provinzen
- In zehn Provinzen müßte Ahmadinejad nach dem offiziellen Ergebnis
  - Alle konservativen Wähler von 2005
  - Alle Erst-/früheren Nichtwähler
  - Alle früheren Zentrumswähler
  - Und fast die Hälfte der reformistischen Wähler für sich gewonnen haben
- 1997, 2001 und 2005 konservative Kandidaten in ländlichen Provinzen sehr unpopulär
- 2009
  - Kein Zusammenhang zwischen Zunahme Wahlbeteiligung/Zunahme Ahmadinejad
  - Negativer Zusammenhang zwischen ländlich/Ahmadinejad verschwindet

2005: Land/Ahmadinejad

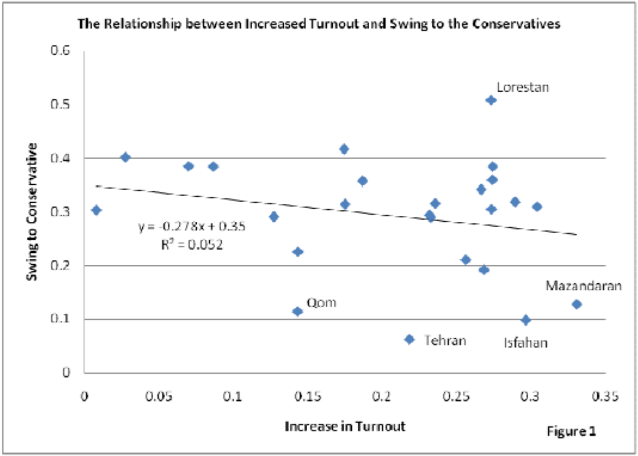


2009: Land/Ahmadinejad



Mehr: `http://www.chathamhouse.org.uk/files/14234_iranelection0609.pdf`

Wahlbeteiligung/Zuwachs Ahmadinejad



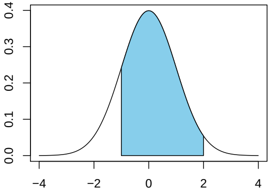
Iran

Wiederholung

Inferenzstatistik

- Punktschätzungen
- Zentraler Grenzwertsatz
- Konfidenzintervalle

Zusammenfassung



Zum Nachlesen

- Agresti/Finlay Kapitel 5
- Gehring/Weins Kapitel 11
- Schumann Kapitel 6

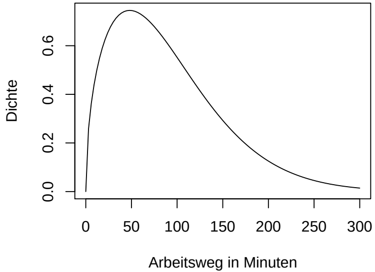
Zufallsexperiment/Wahrscheinlichkeiten

- Zufallsexperiment:
  - Ausgang hängt nur vom Zufall ab
  - (Theoretisch) beliebig oft wiederholbar
  - Relative Häufigkeiten $\rightarrow$ Wahrscheinlichkeiten
- Gedankenexperiment; Standardbeispiele: Münz- und Würfelwurf
- Nützliches Modell für Sozialwissenschaftler
- Z. B. für Stichprobenziehung

Wahrscheinlichkeitsverteilungen

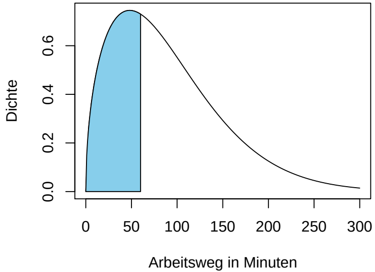
- Wahrscheinlichkeitsverteilungen fassen mögliche Ergebnisse von Zufallsexperimenten und deren Wahrscheinlichkeiten zusammen
- Summe der Balken über alle Ereignisse $= 1$
- Ok für abzählbare Ereignisse (diskrete Zufallsvariablen)
- Was tun mit kontinuierlichen Zufallsvariablen?
- Histogramm $\rightarrow$ Dichteschätzung
- Fläche unter Dichtefunktion $= 1$ (Gesamtwahrscheinlichkeit)
- Wahrscheinlichkeit für Wert aus einem bestimmten Bereich $=$ Fläche über diesem Intervall

Dauer: Fahrt zur Arbeit



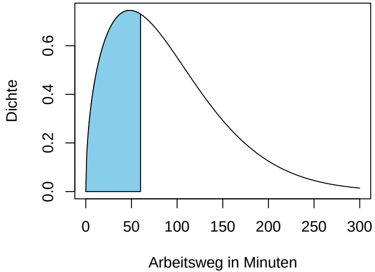
- Wieviel % der Bürger brauchen 60 Minuten oder weniger für den Weg zur Arbeit?
- Potentielle Unterstützer für Infrastrukturmaßnahmen
- Fläche links von 60?

Dauer: Fahrt zur Arbeit



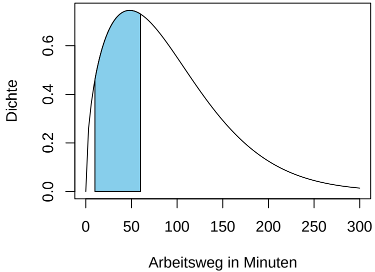
- Wieviel % der Bürger brauchen 60 Minuten oder weniger für den Weg zur Arbeit?
- Potentielle Unterstützer für Infrastrukturmaßnahmen
- Fläche links von 60? → 37%

Dauer: Fahrt zur Arbeit



- Wieviel % der Bürger brauchen 60 Minuten oder weniger für den Weg zur Arbeit?
- Potentielle Unterstützer für Infrastrukturmaßnahmen
- Fläche links von 60? → 37%
- Wieviele % brauchen mehr als 10 und weniger als 60 Minuten?

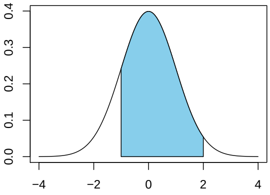
Dauer: Fahrt zur Arbeit



- Wieviel % der Bürger brauchen 60 Minuten oder weniger für den Weg zur Arbeit?
- Potentielle Unterstützer für Infrastrukturmaßnahmen
- Fläche links von 60? → 37%
- Wieviele % brauchen mehr als 10 und weniger als 60 Minuten?
- Wichtige Verteilungen tabelliert

Normalverteilung

- Bekannteste und wichtigste aller kontinuierlichen Zufallsverteilungen
- Wahrscheinlichkeitsdichte für Werte zwischen $-\infty$ und $+\infty$
- Fläche unter der Kurve $=1$
- Über dem Intervall [-1;2] liegen rund 82% der Fläche

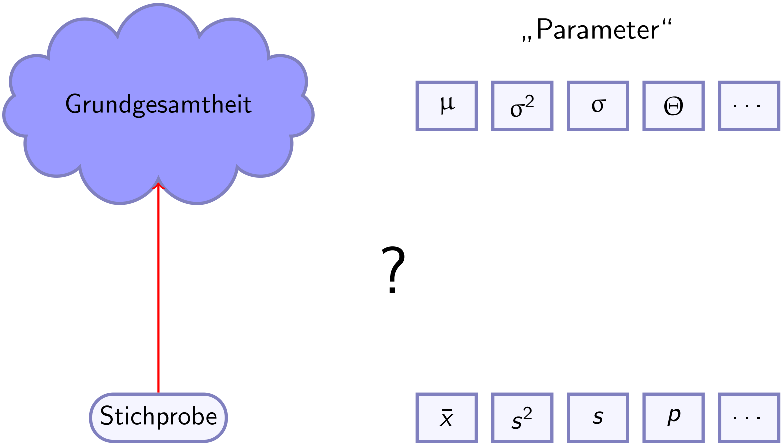


- D. h. in $\approx$ 4 von 5 Fällen zieht man einen Wert aus diesem Intervall

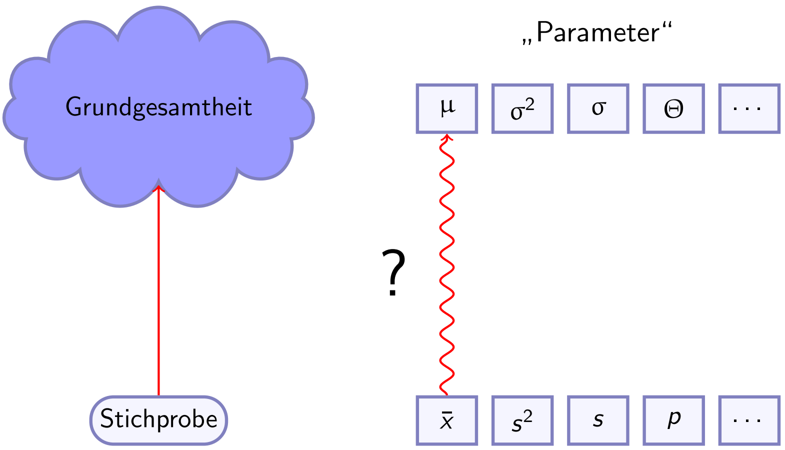
Worum geht es bei der Inferenzstatistik?

- Wie kommt man von einer Stichprobe zu einem (gültigen) Schluß auf die Grundgesamtheit?
- Bzw. wie groß ist der Fehler, den man dabei wahrscheinlich macht?
- Parameter in der Stichprobe vs. Parameter in der Grundgesamtheit
- Parameter z. B.:
  - Mittelwert
  - Varianz/Standardabweichung
  - Anteilswerte
  - Zusammenhangsmaße
  - Regressionskoeffizienten
  - ...

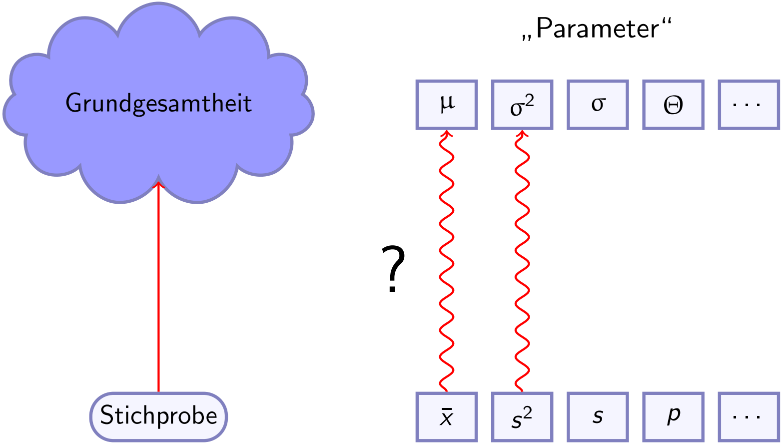
Grundgesamtheit/Stichprobe



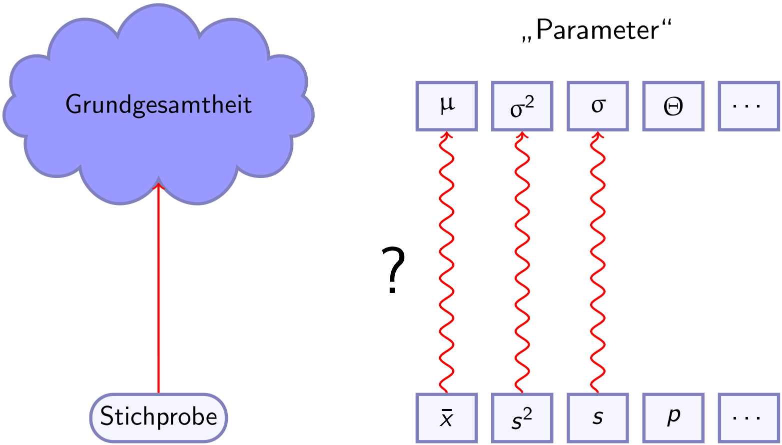
Grundgesamtheit/Stichprobe



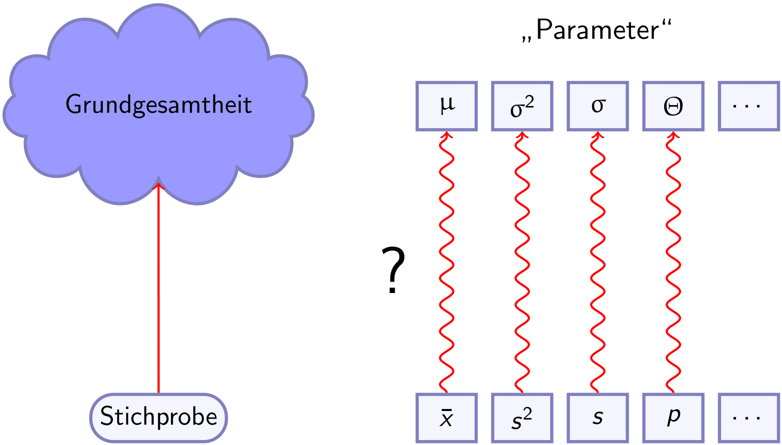
Grundgesamtheit/Stichprobe



Grundgesamtheit/Stichprobe



Grundgesamtheit/Stichprobe



Wie funktionieren Punktschätzungen?

- *Ein* Wert als Schätzung für Parameter in der Grundgesamtheit
- Punktschätzungen sollen präzise sein (wenig Variationen über Stichproben)
- Punktschätzungen sollen unverzerrt / „erwartungstreu“ sein (kein bias)
- Stichprobenmittelwert: unverzerrte Schätzung für Mittelwert in Grundgesamtheit
- Anteilswert in Stichprobe: unverzerrte Schätzung für Anteilswert in Grundgesamtheit

Wie funktionieren Punktschätzungen?

- *Ein* Wert als Schätzung für Parameter in der Grundgesamtheit
- Punktschätzungen sollen präzise sein (wenig Variationen über Stichproben)
- Punktschätzungen sollen unverzerrt / „erwartungstreu“ sein (kein bias)
- Stichprobenmittelwert: unverzerrte Schätzung für Mittelwert in Grundgesamtheit
- Anteilswert in Stichprobe: unverzerrte Schätzung für Anteilswert in Grundgesamtheit
- Varianz (Standardabweichung) in Stichprobe:

Wie funktionieren Punktschätzungen?

- *Ein* Wert als Schätzung für Parameter in der Grundgesamtheit
- Punktschätzungen sollen präzise sein (wenig Variationen über Stichproben)
- Punktschätzungen sollen unverzerrt / „erwartungstreu“ sein (kein bias)
- Stichprobenmittelwert: unverzerrte Schätzung für Mittelwert in Grundgesamtheit
- Anteilswert in Stichprobe: unverzerrte Schätzung für Anteilswert in Grundgesamtheit
- Varianz (Standardabweichung) in Stichprobe: unterschätzt Varianz in Grundgesamtheit

Was ist los mit der Varianz?

- Varianz s^2 in Stichprobe unterschätzt Varianz σ^2 um Faktor $\frac{n-1}{n}$
- Mit $\frac{n}{n-1}$ multiplizieren
- Berechnung von Stichprobenvarianz *als Schätzung für Grundgesamtheit* läßt sich vereinfachen:

$$\begin{aligned}\hat{\sigma}^2 &= s^2 \times \frac{n}{n-1} \\ &= \frac{\Sigma(x_i - \bar{x})^2}{n} \times \frac{n}{n-1} \\ &= \frac{\Sigma(x_i - \bar{x})^2}{n-1}\end{aligned}$$

- Irrelevant in großen Stichproben

Warum reichen Punktschätzungen nicht?

$$\hat{\sigma} = \sqrt{\hat{\sigma}^2}$$

- Schätzungen basieren auf Zufallsstichproben
- Parameter in Stichprobe sind Zufallsvariablen – kontinuierliche Verteilung um wahren Wert (Grundgesamtheit)
- Wahrscheinlichkeit wahren Wert exakt zu treffen $= 0$
- Wahrscheinlichkeit, Stichprobenwert aus einem bestimmten Intervall zu erhalten?
- Analyse der Zufallsverteilung

Wiederholung: Repräsentative Umfrage

„Die Repräsentativitätslüge“

„Meinungsumfragen nennen sich oft , ‚repräsentativ‘ tatsächlich aber werden die Befragten meist nach dem Zufallsprinzip ausgewählt (aus „Vorwärts“, 10/1994, S. 23)

Def. repräsentativ:

- Jedes Element der GG hat gleiche
- oder angebbare
- Wahrscheinlichkeit in die Stichprobe zu gelangen

Wiederholung: Repräsentative Umfrage

„Die Repräsentativitätslüge"

„Meinungsumfragen nennen sich oft , ‚repräsentativ' tatsächlich aber werden die Befragten meist nach dem Zufallsprinzip ausgewählt (aus „Vorwärts", 10/1994, S. 23)

Def. repräsentativ:

- Jedes Element der GG hat gleiche
- oder angebbare
- Wahrscheinlichkeit in die Stichprobe zu gelangen

- *Zufällige Auswahl* der Untersuchungsobjekte

Wiederholung: Repräsentative Umfrage

„Die Repräsentativitätslüge“

„Meinungsumfragen nennen sich oft , ‚repräsentativ‘ tatsächlich aber werden die Befragten meist nach dem Zufallsprinzip ausgewählt (aus „Vorwärts“, 10/1994, S. 23)

Def. repräsentativ:

- Jedes Element der GG hat gleiche
- oder angebbare
- Wahrscheinlichkeit in die Stichprobe zu gelangen

- *Zufällige Auswahl* der Untersuchungsobjekte
- Voraussetzung für Inferenzstatistik

Stichprobenziehung als komplexes Zufallsexperiment

- Zufällige Auswahl von n Befragten
- Berechnung eines Stichprobenparameters (z. B. Mittelwert LRS) über n Befragte
- Stichprobenparameter als *Ergebnis* des komplexen Zufallsexperimentes
- Experiment (prinzipiell) sehr oft wiederholbar
- *Wahrscheinlichkeit*, ein bestimmtes Ergebnis/Ergebnis innerhalb eines bestimmten Wertebereiches zu erhalten
- Berechenbar, wenn Verteilung der Zufallsvariable bekannt

Stichprobenziehung als komplexes Zufallsexperiment

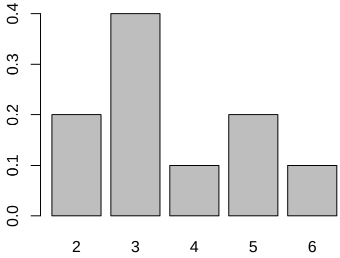
- Zufällige Auswahl von n Befragten
- Berechnung eines Stichprobenparameters (z. B. Mittelwert LRS) über n Befragte
- Stichprobenparameter als *Ergebnis* des komplexen Zufallsexperimentes
- Experiment (prinzipiell) sehr oft wiederholbar
- *Wahrscheinlichkeit*, ein bestimmtes Ergebnis/Ergebnis innerhalb eines bestimmten Wertebereiches zu erhalten
- Berechenbar, wenn Verteilung der Zufallsvariable bekannt
- Wie sieht die Verteilung des Stichprobenparameters (z. B. Mittelwert) aus? $\rightarrow$ „Grenzwertsatz“

Stichprobenziehung als komplexes Zufallsexperiment

- Zufällige Auswahl von n Befragten
- Berechnung eines Stichprobenparameters (z. B. Mittelwert LRS) über n Befragte
- Stichprobenparameter als *Ergebnis* des komplexen Zufallsexperimentes
- Experiment (prinzipiell) sehr oft wiederholbar
- *Wahrscheinlichkeit*, ein bestimmtes Ergebnis/Ergebnis innerhalb eines bestimmten Wertebereiches zu erhalten
- Berechenbar, wenn Verteilung der Zufallsvariable bekannt
- Wie sieht die Verteilung des Stichprobenparameters (z. B. Mittelwert) aus? $\longrightarrow$ „Grenzwertsatz“
- Bei Konstruktion der Zufallsvariable Mittelwert wirken viele (n) einfache Zufallsvariablen zusammen

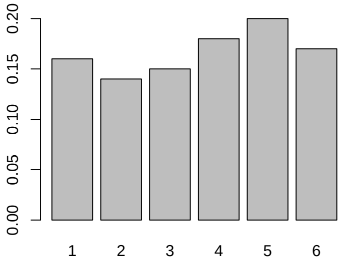
Einfaches Zufallsexperiment: $n = 1$

- Würfel einmal werfen ($n = 1$), Ergebnis notieren
- Experiment sehr oft wiederholen
- Häufigkeit der Werte $1, 2, \cdots 6$ nähert sich an Verteilung in GG (Gleichverteilung) an („Gesetz der großen Zahl“)
- 10 Wiederholungen



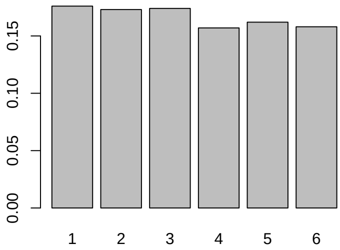
Einfaches Zufallsexperiment: $n = 1$

- Würfel einmal werfen ($n = 1$), Ergebnis notieren
- Experiment sehr oft wiederholen
- Häufigkeit der Werte $1, 2, \cdots 6$ nähert sich an Verteilung in GG (Gleichverteilung) an („Gesetz der großen Zahl“)
- 100 Wiederholungen



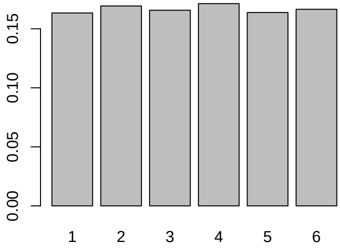
Einfaches Zufallsexperiment: $n = 1$

- Würfel einmal werfen ($n = 1$), Ergebnis notieren
- Experiment sehr oft wiederholen
- Häufigkeit der Werte $1, 2, \cdots 6$ nähert sich an Verteilung in GG (Gleichverteilung) an („Gesetz der großen Zahl“)
- 1 000 Wiederholungen



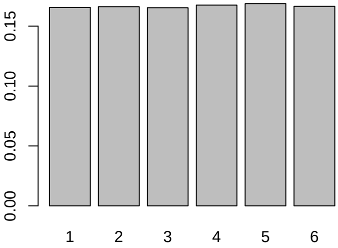
Einfaches Zufallsexperiment: $n = 1$

- Würfel einmal werfen ($n = 1$), Ergebnis notieren
- Experiment sehr oft wiederholen
- Häufigkeit der Werte $1, 2, \cdots 6$ nähert sich an Verteilung in GG (Gleichverteilung) an („Gesetz der großen Zahl“)
- 10 000 Wiederholungen

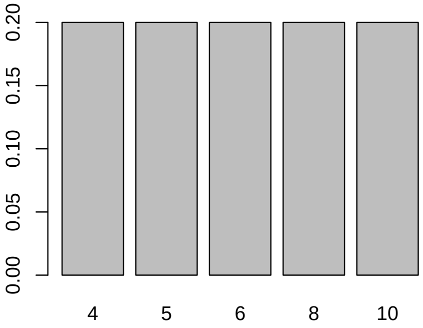


Einfaches Zufallsexperiment: $n = 1$

- Würfel einmal werfen ($n = 1$), Ergebnis notieren
- Experiment sehr oft wiederholen
- Häufigkeit der Werte $1, 2, \cdots 6$ nähert sich an Verteilung in GG (Gleichverteilung) an („Gesetz der großen Zahl“)
- 100 000 Wiederholungen

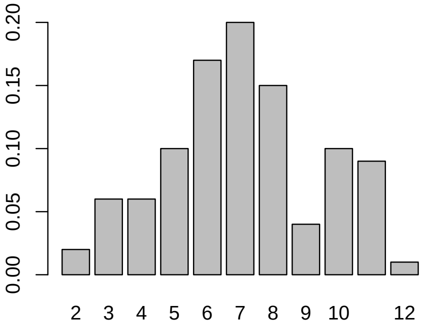


Komplexeres Zufallsexperiment: n=2



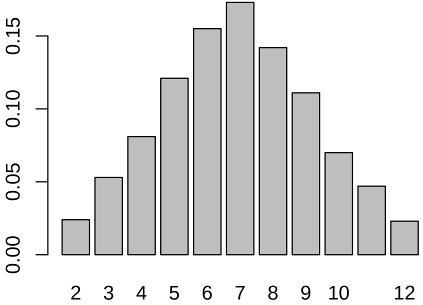
10 Wiederholungen

Komplexeres Zufallsexperiment: n=2



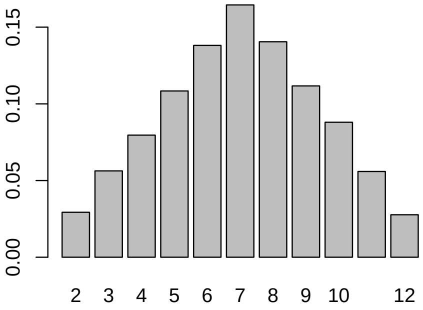
100 Wiederholungen

Komplexeres Zufallsexperiment: n=2



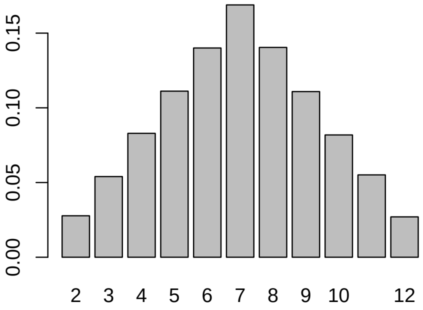
1 000 Wiederholungen

Komplexeres Zufallsexperiment: n=2



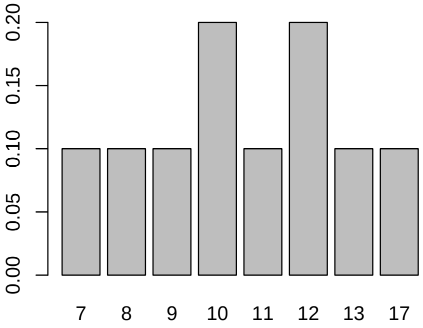
10 000 Wiederholungen

Komplexeres Zufallsexperiment: n=2



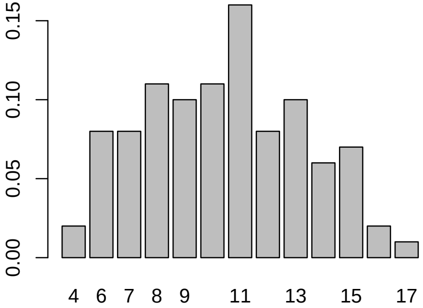
100 000 Wiederholungen

Komplexeres Zufallsexperiment: n=3



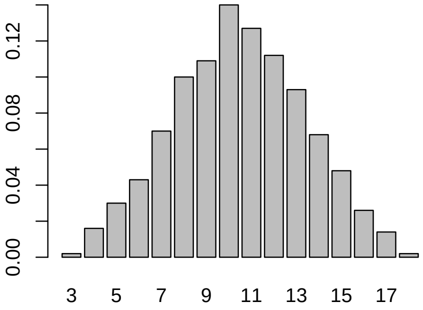
10 Wiederholungen

Komplexeres Zufallsexperiment: n=3



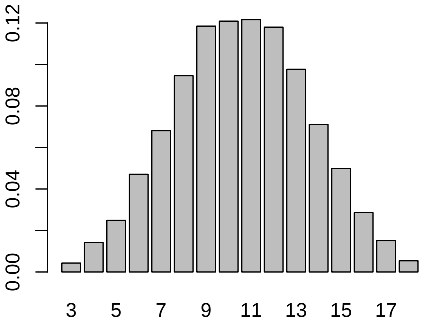
100 Wiederholungen

Komplexeres Zufallsexperiment: n=3



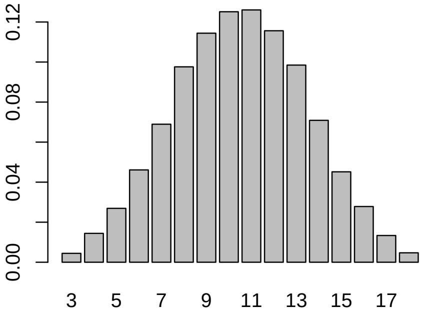
1 000 Wiederholungen

Komplexeres Zufallsexperiment: n=3



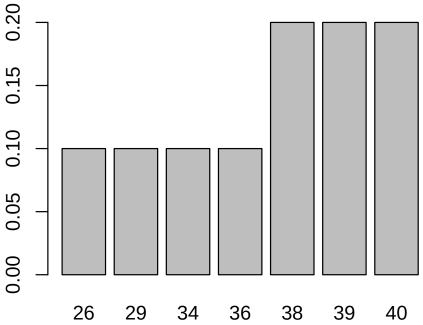
10 000 Wiederholungen

Komplexeres Zufallsexperiment: n=3



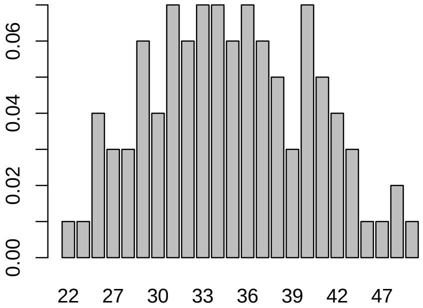
100 000 Wiederholungen

Komplexes Zufallsexperiment: n=10



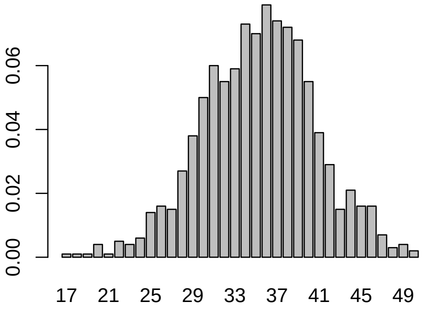
10 Wiederholungen

Komplexes Zufallsexperiment: n=10



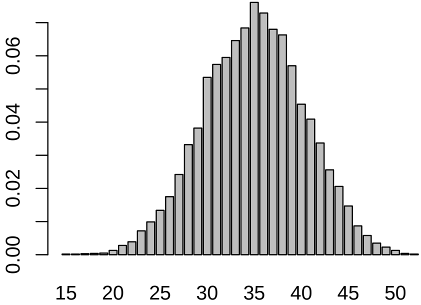
100 Wiederholungen

Komplexes Zufallsexperiment: n=10



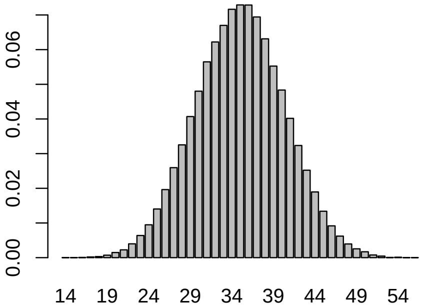
1 000 Wiederholungen

Komplexes Zufallsexperiment: n=10



10 000 Wiederholungen

Komplexes Zufallsexperiment: n=10



100 000 Wiederholungen

Zentraler Grenzwertsatz

- Stichprobenmittelwert: komplexe Zufallsvariable, additiv zusammengesetzt aus n einfachen Zufallsvariablen
- Verteilung der *Stichprobenmittelwerte* aus (theoretisch) unendlich vielen Stichproben nähert sich mit zunehmendem Stichprobenumfang an Normalverteilung an
- Mittelwert der Normalverteilung entspricht Mittelwert in GG (kein bias)
- Wenn $n > 30$, Normalverteilung brauchbare Approximation
- Unabhängig von Verteilung in GG/Stichproben

Zentraler Grenzwertsatz II

Variable	Verteilung
Ausgangsvariable in GG, z. B. Alter	z. B. linkssteil
Alter *eines* Befragten: einfache Zufallsvariable	wie in GG (linkssteil)
Arithmetisches Mittel des Alters: komplexe Zufallsvariable	normalverteilt *über viele mögliche Stichproben*

Berechnung des Standardfehlers

Standardfehler

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \sqrt{\frac{\sigma^2}{n}}$$

- Je größer Streuung in GG, desto größer Standardfehler
- Je kleiner Stichprobenumfang, desto größer Standardfehler
- Umgekehrt quadratischer Zusammenhang: Um Standardfehler zu halbieren muß Stichprobenumfang vervierfacht werden

Eigenschaften der Normalverteilung

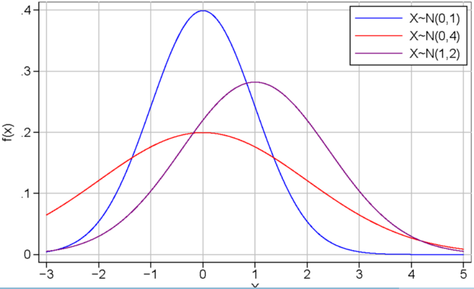
- Familie; gleiche Form, aber unterschiedliche Lage/Breite (Mittelwert/Streuung)
- Für alle Normalverteilungen
  - Ca. 68% der Fläche über Intervall $\pm$ 1 Standardabweichung um Mittelwert
  - Ca. 95% der Fläche über Intervall $\pm$ 1.96 Standardabweichungen um Mittelwert
- „Standardnormalverteilung" (z-Verteilung): Mittelwert $=0$, Standardabweichung/Varianz$=1$ $\rightarrow$ Fläche unter Kurve tabelliert
- Tabelle für alle Normalverteilungen anwendbar (z-Transformation)
- Viele andere Zufallsverteilungen gehen für große n in Normalverteilung über

Eigenschaften der Normalverteilung II

Dichtefunktion

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

- e Eulersche Zahl: $2.718\cdots$
- $\pi : 3.142\cdots$
- μ: Mittelwert der Verteilung
- σ: Standardabweichung der Verteilung

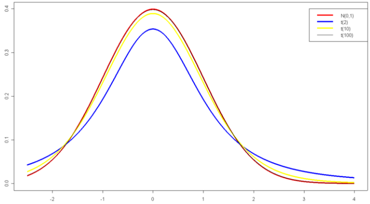


Konfidenzintervalle

- In 95% aller Stichproben ist Stichprobenmittelwert nicht weiter als 1.96 Standardfehler vom wahren Mittelwert entfernt
- Intervall von 1.96 Standardfehlern um Stichprobenmittelwert schließt in 95% der Fälle wahren Mittelwert in GG mit ein
- Intervallschätzung

Konfidenzintervalle II

- Stichprobenmittelwerte nicht standardnormal-, sondern normalverteilt (z-Transformation umkehren)
- Streuung in GG muß nach bekannter Formel aus Streuung in Stichprobe geschätzt werden $\rightarrow$ doppelte Unsicherheit
- t-Verteilung (größere Wahrscheinlichkeit für extreme Werte, Freiheitsgrade)
- Geht bei großen Fallzahlen in Normalverteilung über



Konfidenzintervalle III

- Wahrscheinlichkeit, daß Intervall wahren Wert nicht einschließt: Irrtumswahrscheinlichkeit, α
- Wahrscheinlichkeit, daß Intervall wahren Wert einschließt: Vertrauenswahrscheinlichkeit $1-\alpha$, manchmal γ genannt
- In der Forschungspraxis meistens Vertrauenswahrscheinlichkeiten von 95% oder 99%
  - Symmetrisches Intervall aus (Standard)Normalverteilung
  - „Kritische Werte“ ± 1.96; ± 2.58

Beispiel: LRS in Schweden

- $\bar{x} = 5.822$; $\sigma = \hat{2}.393$; $n = 1891$; gesucht: 90% Konfidenzintervall

Beispiel: LRS in Schweden

- $\bar{x} = 5.822$; $\sigma = \hat{2}.393$; $n = 1891$; gesucht: 90% Konfidenzintervall
- Standardfehler berechnen: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{2.393}{\sqrt{1891}} = 0.055$

Beispiel: LRS in Schweden

- $\bar{x} = 5.822$; $\sigma = \hat{2}.393$; $n = 1891$; gesucht: 90% Konfidenzintervall
- Standardfehler berechnen: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{2.393}{\sqrt{1891}} = 0.055$
- „Kritischen Wert“ suchen

Beispiel: LRS in Schweden

- $\bar{x} = 5.822$; $\sigma = \hat{2}.393$; $n = 1891$; gesucht: 90% Konfidenzintervall
- Standardfehler berechnen: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{2.393}{\sqrt{1891}} = 0.055$
- „Kritischen Wert“ suchen
  - $n > 1000 \rightarrow$ Normalverteilung als Approximation für t-Verteilung

Beispiel: LRS in Schweden

- $\bar{x} = 5.822$; $\sigma = \hat{2}.393$; $n = 1891$; gesucht: 90% Konfidenzintervall
- Standardfehler berechnen: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{2.393}{\sqrt{1891}} = 0.055$
- „Kritischen Wert" suchen
  - $n > 1000 \rightarrow$ Normalverteilung als Approximation für t-Verteilung
  - In Normalverteilungstabelle Wert finden, der links und rechts 5% der Fläche abtrennt

Beispiel: LRS in Schweden

- $\bar{x} = 5.822$; $\sigma = \hat{2}.393$; $n = 1891$; gesucht: 90% Konfidenzintervall
- Standardfehler berechnen: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{2.393}{\sqrt{1891}} = 0.055$
- „Kritischen Wert" suchen
  - $n > 1000 \rightarrow$ Normalverteilung als Approximation für t-Verteilung
  - In Normalverteilungstabelle Wert finden, der links und rechts 5% der Fläche abtrennt
  - $\pm 1.644 \rightarrow$ 5% der Fläche über Intervall $[-\infty; -1.644]$ + 5% der Fläche über Intervall $[1.644; +\infty]$

Beispiel: LRS in Schweden

- $\bar{x} = 5.822$; $\sigma = \hat{2}.393$; $n = 1891$; gesucht: 90% Konfidenzintervall
- Standardfehler berechnen: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{2.393}{\sqrt{1891}} = 0.055$
- „Kritischen Wert" suchen
  - $n > 1000 \rightarrow$ Normalverteilung als Approximation für t-Verteilung
  - In Normalverteilungstabelle Wert finden, der links und rechts 5% der Fläche abtrennt
  - $\pm 1.644 \rightarrow$ 5% der Fläche über Intervall $[-\infty; -1.644]$ + 5% der Fläche über Intervall $[1.644; +\infty]$
- Konfidenzintervall:
  $5.822 \pm 1.644 \times 0.055 = 5.822 \pm 0.09 = [5.732; 5.912]$

Konfidenzintervalle für Anteilswerte

- Vorgehensweise im Grunde identisch
- Verteilung von Anteilswerten binomial
- Kann durch Normalverteilung approximiert werden wenn $n \times p \times (p - 1) \geqslant 9$
- Varianz von Anteilswerten: $p \times (p - 1)$
  - Zwei Ausprägungen gleich häufig ($p = 0.5$) $\rightarrow$ extreme Heterogentität
  - Je geringer Anteil eines Merkmales, desto homogener
  - Keine Varianz wenn $p = 1$
- Konfidenzintervalle können für relative Häufigkeiten oder Prozente berechnet werden

Beispiel: Wahre Wahlabsicht für FDP

- Umfrage mit $n = 1221$: 7.3% (=0.073) würden FDP wählen
- 95% Konfidenzintervall?
- $1221 \times 0.073 \times (1 - 0.073) = 82.63$ (> 9)
- Standardfehler: $\sigma_p = \sqrt{\frac{92.7\% \times 7.3\%}{1221}} = 0.744\%$
- „Kritischer Wert“: 1.96
- Konfidenzintervall: $7.3\% \pm 1.96 \times 0.744\% = [5.84\%; 8.76\%]$
- Alternativ: mit relativen Häufigkeiten rechnen

Logik Konfidenzintervall

- Über sehr viele Stichprobenziehungen hinweg Normalverteilung der Stichprobenkennwerte um wahren Wert in GG mit Streuung = Standardfehler
- 95% aller Stichprobenkennwerte nicht weiter als $\pm$ 1.96 Standardfehler vom wahren Mittelwert entfernt
- Internvall von $\pm$ 1.96 Standardfehlern um Stichprobenmittelwert wird in 95% aller Fälle wahren Mittelwert einschließen
- Habe ich persönlich eine der vielen (relativ) guten Stichproben?

Logik Standardnormalverteilung

- Standardnormalverteilung (z-Verteilung): Mittelwert: 0; Varianz/Standardabweichung: 1
- 95% aller Werte im Intervall $\pm$ 1.96
- Alle Normalverteilungen haben gleiche Form
- 95% der Werte innerhalb von $\pm$ 1.96 Standardfehler ($>$ oder $<$ 1) um jeweiligen Mittelwert

Zusammenfassung

- Inferenzstatistik – Schluß von Stichproben auf GG
- Funktioniert nur für Zufallsstichproben
- Modell für zufällige Abweichungen der Stichprobenkennwerte vom wahren Mittelwert in GG
- Konfidenzintervall: Bereich, der mit gewünschter Sicherheit wahren Wert einschließt
- Nächste Woche: Hypothesentests:
  - Agresti/Finlay: Kapitel 6+7
  - Gehring/Weins: Kapitel 12
  - (Schumann: Kapitel 7.3)