3,120 Methoden: Statistik

Übungsblatt 4: Stichprobenverteilungen und Schätztheorie

Mathis Mörke
Michael Schürle
und Alois Weigand

Universität St.Gallen (HSG)

Herbstsemester 2024

Aufgabe 1

In einem grossen Produktionsunternehmen beträgt die **durchschnittliche Anzahl geleisteter Überstunden 62** pro Jahr mit einer zugehörigen **Standardabweichung von 15** Stunden.
Wie gross ist die Wahrscheinlichkeit, dass in einer zufällig gezogenen Stichprobe mit der Grösse $n = 36$

1. die **durchschnittliche Überstundenzahl** kleiner 65 ist?
2. die durchschnittliche Überstundenzahl zwischen 55 und 65 liegt?
3. die durchschnittliche Überstundenzahl kleiner als 60 ist?

Refresher Aufgabe 1

- Gesucht: $P(x_1 \leq \overline{X} \leq x_2)$
  - Schritt 1: Standardisieren: $Z = \frac{\overline{X} - E(\overline{X})}{\sigma_{\overline{X}}}$
  - Schritt 2: Berechne z-Werte
  - Schritt 3: Finde den entsprechenden Wert in der Tabelle

- Warum die Standardnormalverteilung?

  $\Rightarrow$ Zentraler Grenzwertsatz

Refresher Aufgabe 1

- **Zentraler Grenzwertsatz (Intuition)**
  - Verteilung des arithmetischen Mittelwerts von n unabhängig identisch (wie die Grundgesamtheit) verteilten Zufallsvariablen X_i strebt mit wachsendem Stichprobenumfang n gegen eine Normalverteilung mit dem Erwartungswert μ und der Varianz σ^2/n
  - für jede Verteilung der Grundgesamtheit kann die Verteilung des Mittelwertes durch die Normalverteilung angenähert werden $\Rightarrow$ gilt wenn: Stichprobengrösse ausreichend gross ($n \geq 30$).

Refresher Aufgabe 1

- **Zentraler Grenzwertsatz**
  - X_1, X_2,..., X_n seien stochastisch unabängige Zufallsvariablen
    $\Rightarrow$ gleiche Verteilung (identisch)
    $\Rightarrow$ gleicher Erwartungswert μ, gleiche Varianz σ^2
    $\Rightarrow$ **Idee:** Zufallsstichprobe mit Zurücklegen
  - Dann: Zufallsvariable $Z = \frac{\frac{1}{n}\sum_{i=1}^{n} X_i - \mu}{\frac{\sigma}{\sqrt{n}}}$

    ist näherungsweise standardnormalverteilt (für $n \geq 30$)
  - Achtung: $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ ist auch eine Zufallsvariable mit der Realisation $\overline{x}$ aus den Stichprobenwerten x_1, x_2, x_3,...,x_n

Refresher Aufgabe 1

- **Zentraler Grenzwertsatz**
- Warum Zufallsvariable $Z = \frac{\frac{1}{n}\sum_{i=1}^{n} X_i - \mu}{\frac{\sigma}{\sqrt{n}}}$? $\rightarrow Z = \frac{\overline{X} - E(\overline{X})}{\sigma_{\overline{X}}}$
  - Erwartungswert $E(\overline{X})$
    $\Rightarrow E(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1}^{n} \mu = \frac{1}{n} \cdot n\mu = \mu$
  - Varianz $\sigma^2_{\overline{X}}$
    $\Rightarrow Var(\overline{X}) = Var(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n^2}\sum_{i=1}^{n} \sigma^2 = \frac{1}{n^2} \cdot n\sigma^2 = \frac{\sigma^2}{n}$
  - Standardabweichung von $\overline{X}$ ist dann $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$

Aufgabe 1

1. **durchschnittliche Überstundenzahl kleiner 65**

   $\Rightarrow$ Mittelwert $\mu = 62$, Standardabweichung $\sigma = 15$, Stichprobengrösse $n = 36$

- $P(\overline{X} < 65) = P(Z < \frac{65-62}{2.5}) = P(Z < 1.20)$
- Was brauchen wir?
  - $Z = \frac{\overline{X}-\mu}{\sigma_{\overline{X}}} = \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \Rightarrow z = \frac{65-\mu}{\frac{\sigma}{\sqrt{n}}}$
  - $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{36}} = \frac{15}{6} = 2.5$

Aufgabe 1

1. **durchschnittliche Überstundenzahl kleiner 65**

   $\Rightarrow$ Mittelwert $\mu = 62$, Standardabweichung $\sigma = 15$, Stichprobengrösse $n = 36$

- $P(\overline{X} < 65) = P(Z < \frac{65-62}{2.5}) = P(Z < 1.20)$
- für $z = 1.20 \Rightarrow 0.8849$

Aufgabe 1

2. **durchschnittliche Überstundenzahl zwischen 55 und 65**

   $\Rightarrow$ Mittelwert $\mu = 62$, Standardabweichung $\sigma = 15$, Stichprobengrösse $n = 36$

- $P(55 < \overline{X} < 65) = P(\frac{55-62}{2.5} < Z < \frac{65-62}{2.5})$
  $\Rightarrow P(-2.80 < Z < 1.20)$

- Was brauchen wir?
  - $Z = \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$
  - $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{36}} = \frac{15}{6} = 2.5$

Aufgabe 1

2. **durchschnittliche Überstundenzahl zwischen 55 und 65**

   $\Rightarrow$ Mittelwert $\mu = 62$, Standardabweichung $\sigma = 15$, Stichprobengrösse $n = 36$

- $P(55 < \overline{X} < 65) = P(\frac{55-62}{2.5} < Z < \frac{65-62}{2.5})$
  $\Rightarrow P(-2.80 < Z < 1.20)$

  Tabelle: für $z = 2.80 \Rightarrow 0.9974$

  Tabelle: für $z = 1.20 \Rightarrow 0.8849$

  $\Rightarrow P(-2.80 < Z < 1.20) = 0.8849 - (1 - 0.9974) = 0.8849 - 0.0026 = 0.8823$

- Fläche unter Dichtefunktion ist immer 1!

Aufgabe 1

3. **durchschnittliche Überstundenzahl kleiner 60**

   $\Rightarrow$ Mittelwert $\mu = 62$, Standardabweichung $\sigma = 15$, Stichprobengrösse $n = 36$

- $P(\overline{X} < 60) = P(Z < \frac{60-62}{2.5}) = P(Z < -0.80)$
- Was brauchen wir?
  - $Z = \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$
  - $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{36}} = \frac{15}{6} = 2.5$

Aufgabe 1

3. **durchschnittliche Überstundenzahl kleiner 60**

   $\Rightarrow$ Mittelwert $\mu = 62$, Standardabweichung $\sigma = 15$, Stichprobengrösse $n = 36$

- $P(\overline{X} < 60) = P(Z < \frac{60-62}{2.5}) = P(Z < -0.80)$
- aus Tabelle: für $z = 0.80 \Rightarrow 0.7881$
- $P(\overline{X} < 60) = P(Z < \frac{60-62}{2.5}) = P(Z < -0.80) = 1 - 0.7881 = 0.2119$

Aufgabe 2

Die **durchschnittliche Verweildauer für einen Spitalaufenthalt betrage 4.9 Tage** bei einer zugehörigen **Standardabweichung von 3.5** Tagen. Angenommen es werde eine **Stichprobe von 50** Patienten gezogen, wie gross ist dann die **Wahrscheinlichkeit, dass die durchschnittliche Verweildauer für diese Gruppe nicht mehr als 5.7 Tage beträgt?**

Aufgabe 2

- **durchschnittliche Verweildauer nicht mehr als 5.7 Tage**

  $\Rightarrow$ Mittelwert $\mu = 4.9$, Standardabweichung $\sigma = 3.5$, Stichprobengrösse $n = 50$

  $n = 50 \Rightarrow$ Stichprobenmittelwert $\overline{x}$ kann als normalverteilt angenommen werden (Zentraler Grenzwertsatz)

- $P(\overline{X} < 5.7) = P(Z < \frac{5.7-4.9}{0.495}) = P(Z < 1.62)$

- Was brauchen wir?
  - $Z = \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$
  - $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{3.5}{\sqrt{50}} = 0.495$

Aufgabe 2

- **durchschnittliche Verweildauer nicht mehr als 5.7 Tage**
  $\Rightarrow$ Mittelwert $\mu = 4.9$, Standardabweichung $\sigma = 3.5$, Stichprobengrösse $n = 50$

- $P(\overline{X} < 5.7) = P(Z < \frac{5.7-4.9}{0.495}) = P(Z < 1.62) = 0.9474$

Recap: Verteilung des Stichprobenmittelwerts

- **Verteilung des Stichprobenmittelwerts $\overline{X}$**
  - bei grossem Stichprobenumfang n:
    $E(\overline{X}) = \mu$
    $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$
    $\overline{X}$ annähernd normalverteilt
  - $\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$ ist standardnormalverteilt:

    $$Prob(-z < \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < +z) = F_{Sn}(z) - F_{Sn}(-z) = D(z)$$

Aufgabe 3

Es wird geschätzt, dass 17% der Besitzer von Aktienfonds-Anteilen Rentner seien ($\pi = 0.17$). Angenommen es wurde eine zufällige Stichprobe von $n = 400$ Aktienfonds-Besitzern gezogen:

1. Geben Sie den **Erwartungswert für den Anteil** p der Rentner in der Stichprobe an. Geben Sie ferner die **Standardabweichung für den Anteil** p an.
2. Wie gross ist die Wahrscheinlichkeit, **dass mind. 20% der in der Stichprobe enthaltenen Personen Rentner sind?**
3. Wie gross ist die **Wahrscheinlichkeit, dass sich der Anteil Rentner zwischen 15% und 25% bewegt**?

Refresher Aufgabe 3

- **Stichprobenanteilswert** $P = \frac{X}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i$
  - Was wissen wir:
    - X_1, X_2,..., X_n sind stochastisch unabängige Bernoulli-Experimente
    - Anzahl der Erfolge X durch Addition
    - jedes X_i: 1 für Erfolg und 0 für Misserfolg
    - mit Zurücklegen!!
    - gleicher Erwartungswert $E(X_i) = \pi$
    - gleiche Varianz $Var(X_i) = \pi(1 - \pi)$
  - Für die Zufallsvariable $X = \sum_{i=1}^{n} X_i$:
    $\Rightarrow E(X) = n \cdot \pi$
    $\Rightarrow Var(X) = n \cdot \pi \cdot (1 - \pi)$
    $\Rightarrow$ siehe Binomialverteilung

Refresher Aufgabe 3

- **Zentraler Grenzwertsatz**
  - für Zufallsvariable $Z = \frac{P - E(P)}{\sigma_P}$
  - mit Anteilswert $P = \frac{X}{n} = \frac{1}{n} \sum_{i=1}^{n} X_i$

    $\Rightarrow E(P) = \frac{1}{n} E(X) = \frac{1}{n} \cdot n \cdot \pi = \pi$

    $\Rightarrow$ Varianz $\sigma_P^2 = Var(P) = Var(\frac{X}{n}) = \frac{1}{n^2} Var(X) = \frac{1}{n^2} \cdot n \cdot \pi \cdot (1 - \pi) = \frac{\pi(1-\pi)}{n}$

    $\Rightarrow$ Standardabweichung $\sigma_P = \sqrt{\frac{\pi(1-\pi)}{n}}$

Refresher Aufgabe 3

- **Zentraler Grenzwertsatz**
  - X_1, X_2,..., X_n seien stochastisch unabängige Zufallsvariablen
    $\Rightarrow$ gleiche Verteilung
    $\Rightarrow$ gleicher Erwartungswert und gleiche Varianz
  - Dann ist die Zufallsvariable $Z = \frac{P - E(P)}{\sigma_P} = \frac{P - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}}$

    näherungsweise standardnormalverteilt
  - wenn gilt: $n\pi \geq 5$ und $n(1 - \pi) \geq 5$ (Weiers)

Aufgabe 3

Es wurde geschätzt, dass 17% der Besitzer von Aktienfonds-Anteilen Rentner seien ($\pi = 0.17$). Angenommen es wurde eine zufällige Stichprobe von $n = 400$ Aktienfonds-Besitzern gezogen:

1. Geben Sie den Erwartungswert für den Anteil p der Rentner in der Stichprobe an. Geben Sie ferner die Standardabweichung für den Anteil p an.
2. Wie gross ist die Wahrscheinlichkeit, dass mind. 20% der in der Stichprobe enthaltenen Personen Rentner sind?
3. Wie gross ist die Wahrscheinlichkeit, dass sich der Anteil Rentner zwischen 15% und 25% bewegt?

Aufgabe 3

1. **Erwartungswert und Standardabweichung für Anteil p?**

- Erwartungswert $E(P) = \pi = 0.17$
- Standardabweichung $\sigma_P = \sqrt{\frac{\pi(1-\pi)}{n}} = \sqrt{\frac{0.17(1-0.17)}{400}} = 0.0188$

Aufgabe 3

2. **Wahrscheinlichkeit, dass mind. 20% der in der Stichprobe enthaltenen Personen Rentner sind**

- Es gilt $n\pi = 68 \geq 5$ und $n(1-\pi) = 332 \geq 5$
  $\Rightarrow Z = \frac{P-E(P)}{\sigma_P} = \frac{P-\pi}{\sqrt{\frac{\pi(1-\pi)}{n}}}$ als normalverteilt angenommen
- $P(p \geq 0.20) = P(z \geq \frac{0.20-0.17}{0.0188}) = P(z \geq 1.60) = 1 - 0.9452 = 0.0548$
- aus Tabelle: für $z = 1.60 \Rightarrow 0.9452$

Aufgabe 3

3. **Wahrscheinlichkeit, dass sich der Anteil Rentner zwischen 15% und 25% bewegt**

- $P(0.15 \leq p \leq 0.25) = P(\frac{0.15-0.17}{0.0188} \leq z \leq \frac{0.25-0.17}{0.0188}) =$
  $P(0.15 \leq p \leq 0.25) = P(-1.06 \leq z \leq 4.26) =$
  $1-(1-0.8554) = 0.8554$

Recap: Grenzwertsätze

- Annahme: Zufallsvariable X_i stochastisch unabhängig $\Rightarrow$ Wahrscheinlichkeiten bleiben gleich

  ... oder Mit Zurücklegen!!

- Ohne Zurücklegen bei endlicher Grundgesamtheit
  $\Rightarrow$ **Korrekturfaktor (siehe Weiers)**
  - $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} \times \sqrt{\frac{N-n}{N-1}}$
  - $\sigma_P = \sqrt{\frac{\pi(1-\pi)}{n}} \times \sqrt{\frac{N-n}{N-1}}$

- für Stichprobengrösse n und N Grösse der Grundgesamtheit
  $\Rightarrow$ vernachlässigbar wenn $n/N < 0.05$ (siehe Weiers)

Recap: Verteilung des Stichprobenanteilswerts

- **Verteilung des Stichprobenanteilswerts** P
  - wenn: $n\pi \geq 5$ und $n(1-\pi) \geq 5$ (Weiers):
    $E(P) = \pi$
    $\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}}$
    p annähernd normalverteilt
  - $Z = \frac{P-E(P)}{\sigma_P} = \frac{P-\pi}{\sqrt{\frac{\pi(1-\pi)}{n}}}$ ist standardnormalverteilt:
    $$Prob\left(-z < \frac{P-\pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} < +z\right) = F_{Sn}(z) - F_{Sn}(-z) = D(z)$$

Aufgabe 4

Eine Bierbrauerei füllt ihre 0.5 Liter Bügelflaschen maschinell ab. Wegen des näherungsweise normalverteilten Abfüllfehlers nach $\mathrm{N}(\mu = 0, \sigma^2 = 3.84\mathrm{cl}^2)$ ist die Maschine auf 50.2cl eingestellt. Von den verwendeten Flaschen seien (aufgrund einer im Laugengehalt falsch eingestellten Waschmaschine) 4% der Gummidichtungen der Verschlüsse porös. Ein Kunde lässt sich 1 Palette (40 Kasten zu je 15 Flaschen) schicken.

1. Geben Sie ein 95%-**Prognoseintervall** (resp. Wahrscheinlichkeitsintervall) **für den durchschnittlichen Flascheninhalt und den Gesamtinhalt** der Sendung an,
2. sowie **entsprechende Intervalle für den Anteil und die Gesamtzahl der Flaschen mit porösen Gummidichtungen**.
3. Wie gross ist die Wahrscheinlichkeit, dass der durchschnittliche Flascheninhalt weniger als 0.5 Liter ist?

Aufgabe 4

1) **95%-Wahrscheinlichkeitsintervall für durchschnittl.Flascheninhalt (und Gesamtinhalt)**

- Was wissen wir?
  - $1 - \alpha = 95\% \Rightarrow \alpha = 0.05$, aber zweiseitiger Test $\Rightarrow$ für $\alpha/2 = 0.025$ bekommen wir z-Wert von 1.96 (1-0.025=0.975)
  - Stichprobe $n = 600$ (40 Kisten zu je 15 Flaschen)
  - Mittelwert $\mu = 50.2cl$ und Varianz $\sigma^2 = 3.84\mathrm{cl}^2$
  - Standardabweichung $\sigma_{\overline{x}} = \sqrt{\frac{3.84}{600}} = 0.08$

Aufgabe 4

1) **95%-Wahrscheinlichkeitsintervall für durchschnittl.Flascheninhalt (und Gesamtinhalt)**

- Direkter Schluss von Grundgesamtheit auf Stichprobe

$$P(\mu - z \cdot \sqrt{\frac{\sigma^2}{n}} < \overline{X} < \mu + z \cdot \sqrt{\frac{\sigma^2}{n}}) = 0.95$$

- einsetzen:

$$P(50.2 - 1.96 \cdot \sqrt{\frac{3.84}{600}} < \overline{X} < 50.2 + 1.96 \cdot \sqrt{\frac{3.84}{600}}) = 0.95$$

$\Rightarrow$ 95%-Wahrscheinlichkeitsintervall zwischen 50.04 und 50.36

- Mit 95% Wahrscheinlichkeit liegt der durchschnittliche Flascheninhalt der Stichprobe zwischen 50.04 und 50.36.

Aufgabe 4

1) **95%-Wahrscheinlichkeitsintervall für durchschnittl.Flascheninhalt (und Gesamtinhalt)**

- Mit Wahrscheinlichkeit von 95% liegt der durchschnittliche Flascheninhalt der Stichprobe zwischen 50.04 und 50.36.
- Für Gesamtinhalt ($n\overline{X}$): Intervallgrenzen $50.04 \cdot 600$ und $50.36 \cdot 600$

  $\Rightarrow$ mit 95% ist der Gesamtinhalt zwischen 30024 und 30216 cl

  ... oder zwischen 300.24 und 302.16 liter ($1cl = 0.01l$)

Aufgabe 4

2) **... Intervall für Anteil (und Gesamtzahl) der Flaschen mit porösen Gummidichtungen**

- Was wissen wir?
  - $1 - \alpha = 95\% \Rightarrow \alpha = 0.05$, aber zweiseitiger Test
    $\Rightarrow$ für $\alpha/2 = 0.025$ bekommen wir z-Wert von 1.96 (1-0.025=0.975)
  - Stichprobe $n = 600$ (40 Kisten zu je 15 Flaschen)
  - **jetzt: Anteilswert** $\pi = 0.04$ in Grundgesamtheit
    $\Rightarrow$ selbe Logik

Aufgabe 4

2) **... Intervall für Anteil (und Gesamtzahl) der Flaschen mit porösen Gummidichtungen**

- Direkter Schluss von Grundgesamtheit auf Stichprobe

$$P(\pi - z \cdot \sqrt{\frac{\pi(1-\pi)}{n}} < P < \pi + z \cdot \sqrt{\frac{\pi(1-\pi)}{n}}) = 0.95$$

- $P(0.04 - 1.96 \cdot \sqrt{\frac{0.04 \cdot 0.96}{600}} < P < 0.04 + 1.96 \cdot \sqrt{\frac{0.04 \cdot 0.96}{600}}) = 0.95$

- 95%-Wahrscheinlichkeitsintervall zwischen 0.02432 und 0.05568 (also zwischen 2.4% und 5.6%)
- für nP: zwischen $600 \cdot 0.02432 = 14.59$ und $600 \cdot 0.5568 = 33.41$ Flaschen (also zwischen 14 und 34 Flaschen)

Aufgabe 4

3. **Wahrscheinlichkeit, dass durchschnittlicher Flascheninhalt** $< 50cl$**?**

- $P(\overline{X} < 50) = P(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} < \frac{50-50.2}{0.08})$
  $\Rightarrow z = -0.2/0.08 = -2.5$
  $\Rightarrow$ für $z = 2.5 \Rightarrow 0.9938$ aus Tabelle
  $\Rightarrow$ aber: $P(Z > 2.5) = 1 - 0.9938 = 0.0062$

- Kontrolle: Wir wissen dass nur ein Anteil von 5% der Durchschnitte ausserhalb des Intervalls $[50.04; 50.36]$ liegt

Aufgabe 5

In einer einfachen Stichprobe von 15 HSG-Studenten wurde die Anzahl ausgelassener Vorlesungstage im letzten Monat ermittelt. Die Stichprobe ergab folgendes Ergebnis:
$\{0, 3, 8, 1, 3, 0, 1, 2, 1, 0, 4, 5, 2, 1, 3\}$

1. Ermitteln Sie eine **Schätzung für μ, dem Mittelwert** der ausgelassenen Vorlesungstage eines HSG-Studenten.
2. Geben Sie eine **Schätzung für σ^2 (Varianz)** an.

Aufgabe 5

- **Gesucht: Mittelwert μ der Grundgesamtheit**
- Zufallsvariable X: ausgelassene Vorlesungstage
  $\Rightarrow$ Zufallsstichprobe von $n = 15$ Studenten
  $\Rightarrow$ beobachteten Ausprägungen x_1, x_2,..., x_{15}
- Schätzung $\hat{\mu} = \overline{x} = \frac{1}{n}\sum_{i=1}^{15} x_i$
- $\overline{x} = \frac{\sum_{i=1}^{15} x_i}{n} = \frac{34}{15} = 2.267$

Aufgabe 5

- **Warum $\hat{\mu} = \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ für μ?**

1. Merkmalswerte x_1, x_2,...x_n als Realisationen von Zufallsvariablen $X_1, X_2, ..., X_n \Rightarrow$ Zufallsstichprobe als Zufallsexperiment

2. Stichprobenmittelwert $\overline{x}$ als Realisation der Zufallsvariable $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

3. alle X_i sind identisch und unabhängig mit $E(X_i) = \mu$ und $Var(X_i) = \sigma^2$ (mit Zurücklegen)

4. $E(\hat{\mu}) = E(\overline{X}) = E(\frac{1}{n}\sum_i^n X_i) = \frac{1}{n}\sum_i^n E(X_i) = \frac{1}{n}n\mu = \mu$

Aufgabe 5

- **Warum** $\hat{\mu} = \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ **für** μ**?**
  - $E(\hat{\mu}) = E(\overline{X}) = E(\frac{1}{n}\sum_i^n X_i) = \frac{1}{n}\sum_i^n E(X_i) = \frac{1}{n} n\mu = \mu$
  - Erwartungstreuer Schätzer $\Rightarrow$ im Durchschnitt trifft der Schätzer den gesuchten Wert

Aufgabe 5

- **Gesucht: Varianz σ^2 der Grundgesamtheit**
- Stichprobenvarianz $S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}$

  ... oder umgeformt: $S^2 = \frac{n\sum_{i=1}^{n} X_i^2 - (\sum_{i=1}^{n} X_i)^2}{n(n-1)}$
- $s^2 = \frac{n\sum x_i^2 - (\sum x_i)^2}{n(n-1)} = \frac{15(144) - 34^2}{15(15-1)} = 4.781$

Aufgabe 5

- **Warum** $\hat{\sigma}^2 = S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ **für** σ^2**?**

  $\Rightarrow \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ als Schätzfunktion nicht erwartungstreu

- Intuition:

  $$E(\tfrac{1}{n}\sum(X_i - \overline{X})^2) = E(\tfrac{1}{n}\sum X_i^2 - \tfrac{2}{n}\overline{X}\sum X_i + \tfrac{1}{n}\sum \overline{X}^2)$$
  $$= E(\tfrac{1}{n}\sum X_i^2 - \tfrac{2}{n}\overline{X}n\overline{X} + \tfrac{1}{n}n\overline{X}^2) = \tfrac{1}{n}(\sum E(X_i^2) - nE(\overline{X}^2))$$
  $$= \tfrac{1}{n}(n(\sigma^2 + \mu^2) - n(\sigma_{\overline{X}}^2 + \mu^2)) = \tfrac{1}{n}(n\sigma^2 + n\mu^2 - n\tfrac{\sigma^2}{n} - n\mu^2)$$
  $$= \tfrac{1}{n}(n-1)\sigma^2 = \tfrac{n-1}{n}\sigma^2$$

  - Es gilt: $\sigma^2 = E(X_i^2) - \mu^2 \Rightarrow E(X_i^2) = \sigma^2 + \mu^2$

Aufgabe 5

- **Warum** $\hat{\sigma}^2 = S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ **für** σ^2**?**
- $\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ als Schätzfunktion nicht erwartungstreu, da $E(\frac{1}{n}\sum(X_i - \overline{X})^2) = \frac{n-1}{n}\sigma^2$
- Für Erwartungstreue: $\hat{\sigma}^2 = \frac{n}{(n-1)}\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$
  $\Rightarrow$ Korrekturfaktor
- Erwartungstreuer Schätzer: $\hat{\sigma}^2 = S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}$
  $\Rightarrow E(S^2) = \sigma^2$
  $\Rightarrow$ im Durchschnitt trifft der Schätzer den gesuchten Wert

Recap: Punktschätzer

- Stichprobenmittelwert, -anteilswert, und -varianz als Schätzfunktionen für den Mittelwert, Anteilswert, und die Varianz in der Grundgesamtheit
  - $\hat{\mu} = \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ für μ
  - $\hat{\sigma}^2 = S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ für σ^2
  - Anteilswert p für π

- Werte der Schätzfunktionen hängen von den zufällig gezogenen Stichprobenelementen ab

  $\Rightarrow$ **Stichprobenverteilung** mit Erwartungswert und Varianz.

- Im Durchschnitt entspricht der Schätzer dem (unbekannten) wahren Wert in der Grundgesamtheit (über den wir eine Aussage treffen wollen)

Refresher: Intervallschätzung (zweiseitig)

- **Bsp: Stichprobenmittelwert**

  $$Prob(-z < \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < +z) = F_{Sn}(z) - F_{Sn}(-z) = D(z)$$

  - durch umformen: Intervall um Stichprobenmittelwert

    $$Prob(\mu - z \cdot \frac{\sigma}{\sqrt{n}} < \overline{X} < \mu + z \cdot \frac{\sigma}{\sqrt{n}}) = F_{Sn}(z) - F_{Sn}(-z) = D(z)$$

  - mit welcher Wahrscheinlichkeit ein Stichprobenmittelwert in ein vorher bestimmtes Intervall fällt
  - Annahme: μ, σ, n, $D(z)$ bekannt $\Rightarrow$ berechne z

Refresher: Intervallschätzung (zweiseitig)

- **Schluss von Stichprobe auf Grundgesamtheit**
  - $\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$ standardnormalverteilt

    $$Prob(-z < \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < +z) = F_{Sn}(z) - F_{Sn}(-z)$$
  - durch umformen: Intervall um μ

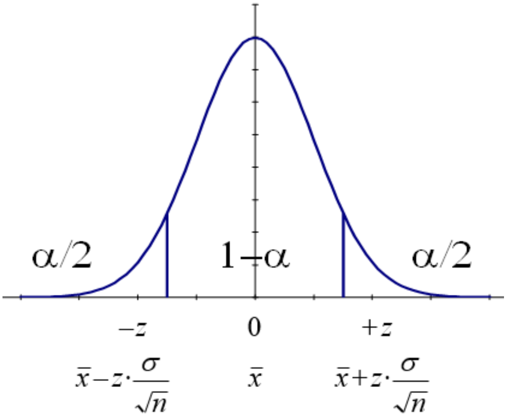
    $$Prob(\overline{X} - z \cdot \frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z \cdot \frac{\sigma}{\sqrt{n}}) = F_{Sn}(z) - F_{Sn}(-z) = 1 - \alpha$$
  - Konfidenzwahrscheinlichkeit $1 - \alpha$
  - Irrtumswahrscheinlichkeit α

Refresher: Intervallschätzung (zweiseitig)

$$Prob(\overline{X} - z \cdot \frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z \cdot \frac{\sigma}{\sqrt{n}}) = 1 - \alpha$$

- Konfidenzwahrscheinlichkeit $1 - \alpha$: wie sehr man darauf vertraut, dass der unbekannte Wert μ im Konfidenzintervall liegt
- Irrtumswahrscheinlichkeit α: wie oft man sich im Mittel irrt
- α und $1 - \alpha$ vorgegeben
  $\Rightarrow$ Anspruch an Genauigkeit der Intervallschätzung
- bekannt: $\overline{x}$, σ (oder S aus der Stichprobe), n, α $\Rightarrow$ berechne z

Refresher: Intervallschätzung (zweiseitig)



Refresher: Intervallschätzung (zweiseitig)

- **Fall 1: Grundgesamtheit normalverteilt und σ bekannt:**

  $\Rightarrow z = \frac{\overline{x}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$

  $\Rightarrow \overline{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$

- **Fall 2: Grundgesamtheit beliebig, σ bekannt, $n \geq 30$:**

  $\Rightarrow z = \frac{\overline{x}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$

  $\Rightarrow \overline{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$

Refresher: Intervallschätzung (zweiseitig)

- **Fall 3: Grundgesamtheit beliebig, σ unbekannt, $n \geq 30$:**

  $\Rightarrow z = \frac{\overline{x}-\mu}{s/\sqrt{n}} \sim N(0,1)$

  $\Rightarrow \overline{x} \pm z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$

  $\Rightarrow s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2$

- **Fall 4: Grundgesamtheit normalverteilt, σ unbekannt, $n < 30$:**

  $\Rightarrow t = \frac{\overline{x}-\mu}{s/\sqrt{n}} \sim t_{n-1}$ mit $n-1$ Freiheitsgraden

  $\Rightarrow \overline{x} \pm t_{\alpha/2}^{n-1} \cdot \frac{s}{\sqrt{n}}$

  $\Rightarrow s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2$

Refresher: Intervallschätzung (zweiseitig)

- Kleine Anmerkung:
  - falls Grundgesamtheit N endlich und $n/N \geq 0.05$:
    $\Rightarrow$ Ohne Zurücklegen angenommen
    $\Rightarrow$ Korrekturfaktor $\frac{N-n}{N-1}$ bei Standardfehlern $\sigma_{\overline{X}}$ oder σ_P

    $\Rightarrow$ In Vorlesung nicht weiter behandelt.

Aufgabe 6

Von 1000 Bachelor-Studenten werden 37 zufällig ausgewählte Studenten nach ihrem Alter befragt. Es ergibt sich ein Durchschnittsalter von $\overline{x} = 23$ Jahren mit einer Standardabweichung von 0.6 Jahren. Bestimmen Sie ein 95%-Konfidenzintervall für das Durchschnittsalter.
Anmerkung: Wir interessieren uns für das Durchschnittsalter aller Bachelor Studenten der Grundgesamtheit (mit $N = 1000$)

Aufgabe 6

- **Gesucht: 95%-Konfidenzintervall für das Durchschnittsalter**

- Was wissen wir?
  - $N = 1000$ mit Stichprobe $n = 37$
  - aus Stichprobe: $\overline{x} = 23$ und $s = 0.6$
  - $n/N = 37/1000 = 0.037 < 5\%$
    $\Rightarrow$ auf Korrekturfaktor kann verzichtet werden
  - Varianz σ unbekannt: $\hat{\sigma}_{\overline{X}} = \frac{s}{\sqrt{n}}$
  - Verteilung der Grundgesamtheit unbekannt, aber $n > 30$

Aufgabe 6

- **Gesucht: 95%-Konfidenzintervall für das Durchschnittsalter**

- Konfidenzintervall für unbekanntes Durchschnittsalter μ
  $\Rightarrow \overline{x} \pm z_{\alpha/2} \cdot \frac{\hat{\sigma}}{\sqrt{n}}$
  $Prob(\overline{X} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}) = 1 - \alpha$

- $\hat{\sigma}_{\overline{X}} = \frac{s}{\sqrt{n}} = \frac{0.6}{\sqrt{37}} = 0.099$

- $z = 1.96$ für $\alpha/2 = 0.025$ (zweiseitig)

- einsetzen:
- $P(23 - 1.96 \cdot 0.099 < \mu < 23 + 1.96 \cdot 0.099) = 0.95$
  $\Rightarrow$ Konfidenzintervall zwischen 22.806 und 23.194

Aufgabe 7

Aus den (angenommenen) 20000 Studierenden der Zürcher Universität werden 600 für eine Befragung zufällig ausgewählt. In der Stichprobe gaben 360 Studierende an, ein festes Natel-Abo zu besitzen.

1. Geben Sie einen erwartungstreuen Schätzwert für den Anteil der Natel-Telefonierer (mit Abo) der ZH-Uni an.
2. Ermitteln Sie für den Anteil und die Anzahl der Natel-Abo-Besitzer in der Grundgesamtheit ein 95%-Konfidenzintervall und interpretieren Sie dieses.

Aufgabe 7

- **Erwartungstreuer Schätzer für Anteilswert**

- Stichprobenanteilswert: Anteil des Erfolgs x in einer Stichprobe von n Versuchen
- $\hat{\pi} = p = \frac{360}{600} = 0.6$

- Warum Stichprobenanteilswert? $\Rightarrow$ Erwartungstreue
  - p als Realisation von $P = \frac{X}{n} = \frac{1}{n} \sum_{i=1}^{n} X_i$
  - Bernoulli-Zufallsvariable X_i mit $E(X_i) = \pi$
  - Für $X = \sum_{i=1}^{n} X_i$ gilt $E(X) = n \times \pi$
    $\Rightarrow E(P) = \frac{1}{n} E(X) = \frac{1}{n} \times n \times \pi = \pi$

Aufgabe 7

- 95%-**Konfidenzintervall für Anteilswert und Anzahl der Abo-Besitzer**

- $x = np = 360 \geq 5$ und $n(1-p) = 240 \geq 5$
  $\Rightarrow$ Verteilung durch Normalverteilung approximiert
- $n/N = 600/20000 = 0.03 < 5\%$
  $\Rightarrow$ Korrekturfaktor vernachlässigbar

- $\sigma_P = \sqrt{\frac{\pi(1-\pi)}{n}} \Rightarrow s_P = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.6(0.4)}{600}} = 0.02$

- Konfidenzwahrscheinlichkeit von 95% $\Rightarrow \alpha = 5\%$
- zweiseitig z-Wert für $\alpha/2 = 0.025$ entspricht 1.96

Aufgabe 7

- 95%-**Konfidenzintervall für Anteilswert und Anzahl der Abo-Besitzer**

- Konfidenzwahrscheinlichkeit von 95% $\Rightarrow \alpha = 5\%$
- zweiseitig z-Wert für $\alpha/2 = 0.025$ entspricht 1.96

- $Prob(p - z_{\alpha/2} \cdot \sigma_P \leq \pi \leq p + z_{\alpha/2} \cdot \sigma_P) = 0.95$
- $Prob(0.6 - 1.96 \cdot 0.02 \leq \pi \leq 0.6 + 1.96 \cdot 0.02) = 0.95$
- $Prob(0.56 \leq \pi \leq 0.64) = 0.95$

Aufgabe 7

- 95%-**Konfidenzintervall für Anteilswert und Anzahl der Abo-Besitzer**
- Konfidenzintervall $Prob(0.56 \leq \pi \leq 0.64) = 0.95$
- $N = 20000$ Studenten (davon $\pi = 0.6$ (oder 60%))
  $\Rightarrow$ Konfidenzintervall für Anzahl der Abo-Besitzer $N\pi$
  $\Rightarrow Prob(0.56 \cdot 20000 \leq N\pi \leq 0.64 \cdot 20000) = 0.95$
  $\Rightarrow Prob(11200 \leq N\pi \leq 12800) = 0.95$