

Applied Statistics for Data Science

Serie 8

Aufgabe 8.1 $\bar{X}_{36} \approx \mathcal{N}\left(\mu, \frac{\sigma_x^2}{36}\right) = \mathcal{N}\left(8.2, \frac{36}{36}\right) = \mathcal{N}(8.2, 1)$

Die Zeit, die ein Passagier an einem Flughafen Check-in Schalter verbringt ist eine Zufallsvariable mit Mittelwert 8.2 Minuten und Standardabweichung 6 Minuten. Wir beobachten zufällig 36 Passgiere.

- Berechnen Sie die Wahrscheinlichkeit, dass die durchschnittliche Wartezeit dieser Passagiere weniger als 10 Minuten beträgt.
- Berechnen Sie die Wahrscheinlichkeit, dass die durchschnittliche Wartezeit dieser Passagiere zwischen 5 und 10 Minuten beträgt.
- Berechnen Sie die Wahrscheinlichkeit, dass die durchschnittliche Wartezeit dieser Passagiere mehr als 20 Minuten beträgt.
- Alle haben wohl schon die Erfahrung gemacht, dass man länger beim Check-in gewartet hat. Warum ist die Wahrscheinlichkeit von c) dann so klein?
durchschnittliche Wartezeit wird betrachtet. Unwahrscheinlich, dass diese mehr als 20 Minuten ist
- Gilt hier die i.i.d.-Annahme überhaupt?
kommt darauf an, wo und wann Messungen gemacht werden

Aufgabe 8.2

Ein Zigarettenhersteller gibt an, dass der Nikotingehalt in einer Zigarette durchschnittlich 2.2 mg mit Standardabweichung 0.3 mg hat. Bei einer Stichprobe von 100 zufällig ausgewählten Zigaretten ist das Stichprobenmittel allerdings 3.1 mg.

Wie hoch ist die Wahrscheinlichkeit, dass das Stichprobenmittel einen Wert von 3.1 mg oder mehr erreicht, wenn die Aussage des Zigarettenherstellers wahr ist?

Interpretieren Sie dieses Resultat.

$$\mathcal{N}\left(2.2, \frac{0.3^2}{100}\right) = \mathcal{N}(2.2, 0.0009)$$

Aussage vom Zigarettenherstellers ziemlich sicher nicht wahr.

Aufgabe 8.3

Ein Dozent weiss aus Erfahrung, dass bei einer Prüfung die Punktezahlen durchschnittlich bei 77 Punkte mit einer Standardabweichung von 15 Punkten liegen. In diesem Semester unterrichtet der Dozent zwei Kurse; der eine hat 25, der andere 64 TeilnehmerInnen.

a) Wie gross ist näherungsweise die Wahrscheinlichkeit, dass das durchschnittliche Prüfungsergebnis im dem Kurs mit 25 TeilnehmerInnen zwischen 72 und 82 Punkten liegt?

b) Wiederholen Sie die Rechnung aus Teil a) für den Kurs mit 64 TeilnehmerInnen.

Vergleichen Sie die beiden Werte und erklären Sie den Unterschied.

$$a) \mathcal{N}\left(77, \frac{15^2}{25}\right) = \mathcal{N}(77, 9)$$

$$\text{pnorm}(\dots, 77, \underline{\text{sqrt}(9)})$$

$$b) \mathcal{N}\left(77, \frac{15^2}{64}\right)$$

unterschied entsteht, da die sd kleiner wird bei mehr Kameraden in der Klasse

Applied Statistics for Data Science

Musterlösungen zu Serie 8

Lösung 8.1

X_i ist die Zufallsvariable der Wartezeit für den i -ten Passagier. Es gilt $\mu = 8.2$ und $\sigma_X = 6$.

Wir betrachten die durchschnittliche Wartezeit \bar{X}_{36} , die annähernd wie folgt verteilt ist:

$$\bar{X}_{36} \sim \mathcal{N}\left(\mu, \frac{\sigma_X^2}{n}\right) = \mathcal{N}\left(8.2, \frac{6^2}{36}\right) = \mathcal{N}(8.2, 1)$$

a) Gesucht ist

$$P(\bar{X}_{36} \leq 10) = 0.9640697$$

```
pnorm(q = 10, mean = 8.2, sd = 1)
## [1] 0.9640697
```

b) Gesucht ist

$$P(5 \leq \bar{X}_{36} \leq 10) = 0.9633825$$

```
pnorm(q = 10, mean = 8.2, sd = 1) - pnorm(q = 5, 8.2, 1)
## [1] 0.9633825
```

Der Unterschied zu b) ist sehr klein. Das heisst aber auch, dass die Wahrscheinlichkeit, dass der Durchschnitt unter 5 Minuten liegt sehr klein ist.

c) Gesucht ist

$$P(\bar{X}_{36} \geq 20) \approx 0$$

```
1 - pnorm(q = 20, mean = 8.2, sd = 1)
## [1] 0
```

Hier wird die W'keit 0 angegeben, aber dies ist sie nicht. Sie ist nur so klein, dass sie mit 0 dargestellt wird.

- d) Die Wahrscheinlichkeit, dass man selber mehr als 20 Minuten warten kann, ist natürlich viel grösser.

Die Wahrscheinlichkeit in c) beschreibt die Wahrscheinlichkeit, dass 36 zufällig beobachtete Personen durchschnittlich mehr 20 warteten und diese ist fast 0.

Dass viele Personen durchschnittlich mehr als 20 Minuten warten müssen, ist kleiner als die Wahrscheinlichkeit dass *eine* Person mehr als 20 Minuten warten muss.

- e) Dies ist eine heikle Angelegenheit. Wenn Sie einen grossen Flughafen nehmen und zu jeder Tageszeit zufällig Passagiere an einem beliebigen Abfertigungsschalter auswählen, dann scheint die i.i.d.-Annahme gerechtfertigt zu sein.

Wenn Sie einen Abfertigungsschalter nach dem Zufallsprinzip auswählen und 36 Passagiere von diesem Schalter auswählen, ist die Annahme nicht gerechtfertigt. Wenn es bereits eine lange Schlange gibt, dann warten alle Passagiere überdurchschnittlich lange.

Dasselbe gilt, wenn Sie eine Tageszeit wählen, zu der der Flughafen ungewöhnlich stark ausgelastet ist. Die Wartezeit für alle Passagiere wäre länger.

Lösung 8.2

X_i ist die Zufallsvariable des Nikotingehalts in der i -ten Zigarette. Es gilt $\mu = 2.2$ und $\sigma_X = 0.3$.

Wir betrachten den durchschnittlichen Nikotingehalt \bar{X}_{100} , die annähernd wie folgt verteilt ist:

$$\bar{X}_{100} \sim \mathcal{N}\left(\mu, \frac{\sigma_X^2}{n}\right) = \mathcal{N}\left(2.2, \frac{0.3^2}{100}\right) = \mathcal{N}(2.2, 0.0009)$$

Gesucht ist

$$P(\bar{X}_{100} \geq 3.1) \approx 0$$

Diese Wahrscheinlichkeit, dass das Stichprobenmittel 3.1 mg oder grösser ist, ist praktisch 0. Das heisst aber auch, dass unter der Annahme von 2.2 mg ein Stichprobenmittel von 3.1 mg (oder mehr) *sehr* unwahrscheinlich ist. So können wir vermuten, dass etwas mit der Angabe von 2.2 mg nicht stimmt.

Lösung 8.3

X_i ist die Zufallsvariable für die Punktezah der i -ten StudentIn. Es gilt $\mu = 77$ und $\sigma_X = 15$.

- a) Wir betrachten die durchschnittliche Punktezah \bar{X}_{25} , die annähernd wie folgt verteilt ist:

$$\bar{X}_{25} \sim \mathcal{N}\left(\mu, \frac{\sigma_X^2}{n}\right) = \mathcal{N}\left(77, \frac{15^2}{25}\right) = \mathcal{N}(77, 9)$$

Gesucht ist

$$P(72 \leq \bar{X}_{25} \leq 82) = 0.9044193$$

```
pnorm(q = 82, mean = 77, sd = 3) - pnorm(q = 72, 77, 3)
## [1] 0.9044193
```

- b) Wir betrachten die durchschnittliche Punktezahl \bar{X}_{64} , die annähernd wie folgt verteilt ist:

$$\bar{X}_{64} \sim \mathcal{N}\left(\mu, \frac{\sigma_X^2}{n}\right) = \mathcal{N}\left(77, \frac{15^2}{64}\right) = \mathcal{N}(77, 225/64)$$

Gesucht ist

$$P(72 \leq \bar{X}_{64} \leq 82) = 0.9923392$$

```
pnorm(q = 82, 77, sqrt(225/64)) - pnorm(q = 72, 77, sqrt(225/64))
## [1] 0.9923392
```

Der 2. Wert ist grösser als der erste. Der Grund ist, dass im 2. Fall die Standardabweichung kleiner ist und sich somit mehr Werte in diesem Bereich befinden.