5 SCHÄTZEN VON ANTEILEN

W 5.01 Es sei p der bekannte relative Anteil eines Merkmals in der Gesamtbevölkerung. Es soll die unbekannte relative Häufigkeit h in einer großen Stichprobe vom Umfang n geschätzt werden. Was versteht man unter dem γ-Streubereich von p? Gib eine Formel für den γ-Streubereich von p an!

W 5.02 Es sei h die bekannte relative Häufigkeit eines Merkmals in einer großen Stichprobe vom Umfang n. Es soll der unbekannte relative Anteil p des Merkmals in der Grundgesamtheit geschätzt werden. Was versteht man unter einem γ-Konfidenzintervall für p? Gib eine Formel für die Berechnung des γ-Konfidenzintervalls für p an!

W 5.03 Wie kann ein 95 %-Konfidenzintervall (frequentistisch) gedeutet werden?

W 5.04 Wie ändert sich ein γ-Konfidenzintervall, wenn man eine größere Sicherheit γ wählt (aber n und h gleich bleiben)?

W 5.05 Wie ändert sich ein γ-Konfidenzintervall, wenn man einen größeren Stichprobenumfang n wählt (aber γ und h gleich bleiben)?

W 5.06 Zum Ermitteln des unbekannten relativen Anteils p eines Merkmals in der Gesamtbevölkerung wird eine Stichprobe von großem Umfang n erhoben. In dieser ergibt sich der Wert h für die relative Häufigkeit des Merkmals. Es soll ein Konfidenzintervall der Länge d für p angegeben werden. Wie kann man die Sicherheit γ dieses Konfidenzintervalls berechnen?

W 5.07 Für einen unbekannten relativen Anteil p in einer Bevölkerung soll aufgrund einer Stichprobe ein 95 %-Konfidenzintervall von vorgegebener Länge d angegeben werden. Beschreibe, wie man den nötigen Mindestumfang der Stichprobe ermitteln kann, falls
- eine Voruntersuchung die relative Häufigkeit h ergeben hat,
- keine Voruntersuchung vorliegt!

W 5.01 Der γ-Streubereich von h ist jenes Intervall symmetrisch um p, welches die unbekannte relative Häufigkeit in einer Stichprobe vom vorgegebenen Umfang n mit der Wahrscheinlichkeit γ enthält.

Ist n groß, so gilt: γ-Streubereich von $h \approx \left[p - z \cdot \sqrt{\frac{p \cdot (1-p)}{n}}; p + z \cdot \sqrt{\frac{p \cdot (1-p)}{n}}\right]$, wobei z so zu ermitteln ist, dass $\Phi(z) = \frac{1+\gamma}{2}$ gilt.

W 5.02 Das γ-Konfidenzintervall für p ist die Menge aller Schätzwerte von p, deren zugehörige γ-Streubereiche den bekannten Wert h überdecken.

Es gilt: γ-Konfidenzintervall für $p \approx \left[h - z \cdot \sqrt{\frac{h \cdot (1-h)}{n}}; h + z \cdot \sqrt{\frac{h \cdot (1-h)}{n}}\right]$, wobei z so zu ermitteln ist, dass $\Phi(z) = \frac{1+\gamma}{2}$ gilt.

W 5.03 Würde man sehr oft Stichproben vom Umfang n erheben und jedesmal durch $\left[h - z \cdot \sqrt{\frac{h \cdot (1-h)}{n}}; h + z \cdot \sqrt{\frac{h \cdot (1-h)}{n}}\right]$ mit $\Phi(z) = \frac{1+\gamma}{2}$ ein γ-Konfidenzintervall für den unbekannten relativen Anteil p eines Merkmals in einer Grundgesamtheit ermitteln, dann würden diese Intervalle bei ca. 95 % aller Stichproben den unbekannten relativen Anteil p überdecken.

W 5.04 Je größer die Sicherheit γ ist, desto größer ist die Länge des zugehörigen γ-Konfidenzintervalls, dh. desto ungenauer ist die Schätzung von p durch das Konfidenzintervall.

W 5.05 Je größer der Stichprobenumfang n ist, desto kleiner ist die Länge des zugehörigen γ-Konfidenzintervalls, dh. desto genauer ist die Schätzung von p durch das Konfidenzintervall.

W 5.06 $\gamma \approx 2 \cdot \Phi(z) - 1$, wobei $z = \frac{d}{2 \cdot \sqrt{\frac{h \cdot (1-h)}{n}}}$

W 5.07 Für die halbe Länge des 95 %-Konfidenzintervalls soll gelten:

$z \cdot \sqrt{\frac{h \cdot (1-h)}{n}} = \frac{d}{2}$, wobei z so ermitteln ist, dass $\Phi(z) = \frac{1+0{,}95}{2}$. Damit ist $z \approx 1{,}96$.

- Ist eine Schätzung für h durch eine Voruntersuchung vorhanden, so erhält man für den nötigen minimalen Stichprobenumfang $n \approx \frac{4 \cdot 1{,}96^2 \cdot h \cdot (1-h)}{d^2}$.
- Ist keine Schätzung für h aus einer Voruntersuchung bekannt, überlegt man, dass der Term $\frac{4 \cdot 1{,}96^2 \cdot h \cdot (1-h)}{d^2}$ bei konstantem d für $h = 0{,}5$ seinen größten Wert annimmt. Damit erhält man für den nötigen minimalen Stichprobenumfang $n \approx \left(\frac{1{,}96}{d}\right)^2$.