**Prognose- und Schätzbereiche**

**Genauigkeit von Simulationen – Eine Prognose wagen**

In der GeoGebra-Datei wurde nur mit gearbeitet.

Der gefundene Zusammenhang lässt sich aber auch für alle anderen Werte für verallgemeinern:

|  |  |
| --- | --- |
| |  | | --- | | **Vorhersage von relativen Häufigkeiten bei bekannter Wahrscheinlichkeitund Simulationen ( - Gesetz)**  Die relative Häufigkeit nach n Simulationen, , liegt mit 95 % Wahrscheinlichkeit vom bekannten maximal entfernt.  Wir können folgende Aussage treffen:  liegt im Intervall .  Dieser Schluss kann mit 95 % Sicherheit gezogen werden. | |

**Faustregeln**

|  |  |  |
| --- | --- | --- |
| **Stichprobenumfang** | **95 %-Prognosebereich**  **(teilweise gerundet)** | **Breite des Prognosebereichs** |
| 50 |  |  |
| 100 |  |  |
| 1.000 |  |  |
| 10.000 |  |  |

**Beispiel**

Bei 1.000 Simulationen und einem bekannten hat der 95 %-Prognosebereich die Grenzen   
0,4 ± 0,03.

**Unter- bzw. Überschreiten von 0 und 1**

Der Term kann den Wert 0 unterschreiten, der Term den Wert 1 überschreiten. ist aber immer durch 0 und 1 nach unten bzw. nach oben begrenzt. Prognosebereiche, deren Grenzen 0 unter- bzw. 1 überschreiten, werden stattdessen von 0 nach unten bzw. von 1 nach oben begrenzt.

In unserem Beispiel mit p = 0,5 würde bei einer Stichprobe von weniger als 5 noch keine nützliche Aussage entstehen, da einen Prognosebereich von 0 bis 1 liefert, und erst das Intervall verkleinert.

**Genauigkeit von Simulationen – Schätzen von Wahrscheinlichkeiten aus Simulationen**

Bisher konnten wir mit 95 % Wahrscheinlichkeit etwas über die Lage eines Simulationsergebnisses vorhersagen. Man kann aber den gefundenen Zusammenhang auch umkehren. Das folgende Beispiel verdeutlicht das:

|  |  |
| --- | --- |
| **A**  *Mithilfe einer Simulation werden aus einer Urne mit 10.000 Kugeln 100 gezogen (Ziehen mit Zurücklegen). Sie wissen, dass in der Urne 4.000 blaue und 6.000 rote Kugeln sind. Der Anteil p ist also bekannt. Wie groß schätzen Sie den Anteil h100 an roten Kugeln in der Stichprobe, bevor Sie die Simulation durchführen?* | **B**  *Mithilfe einer Simulation werden aus einer Urne mit 10.000 Kugeln 100 Kugeln gezogen (Ziehen mit Zurücklegen). Sie wissen, dass von den Kugeln in der Urne manche blau und die anderen rot sind. Sie simulieren eine Zufallsstichprobe von 100 Kugeln. Darin befinden sich 52 % rote Kugeln. Wie groß schätzen Sie den Anteil p an roten Kugeln in der Urne?* |

Angelehnt an: Lergenmüller, A., Schmidt, G., & Krüger, K. (Hrsg.) (2012): *Mathematik Neue Wege. Arbeitsbuch für Gymnasien. Stochastik*. Braunschweig: Schroedel, Seite 153.

In A sind die Anteile in der Gesamtheit bekannt. Mithilfe des Anteils an roten Kugeln und Vorgabe einer Sicherheitswahrscheinlichkeit von 95 % kann man den 95 %-Prognosebereich für berechnen. Nach den Faustregeln beträgt die Abweichung von 0,6 bei gerade 0,1. Der Prognosebereich beginnt also bei 0,5 und endet bei 0,7. Er entspricht dem Intervall .

In B möchte man von einem Simulationsergebnis auf eine unbekannte Wahrscheinlichkeit schließen. Mit 95 % Wahrscheinlichkeit liegt 0,52 maximal von p entfernt. Wenn wir sagen: liegt im Bereich 0,42 bis 0,62 (kurz , dann können wir zu 95 % sicher sein, dass das stimmt.

|  |
| --- |
| **Schätzen von Wahrscheinlichkeiten aus relativen Häufigkeiten**  **nach n Simulationen ( - Gesetz)**  Aus n Simulationen mit unbekanntem *p* ermittelt man eine relative Häufigkeit . Mit 95 % Wahrscheinlichkeit liegt vom unbekannten *p* maximal entfernt.  Wir können folgende Aussage treffen:  *p* liegt im Intervall .  Dieser Schluss kann mit 95 % Sicherheit gezogen werden.  Man nennt diesen Bereich auch 95 %-Schätzbereich für die unbekannte Wahrscheinlichkeit. |

**Faustregeln**

|  |  |  |
| --- | --- | --- |
| **Stichprobenumfang *n*** | **95 %-Schätzbereich**  **(teilweise gerundet)** | **Breite des Schätzbereichs** |
|  |  |  |
| 100 |  |  |
| 1.000 |  |  |
| 10.000 |  |  |

**Unter- bzw. Überschreiten von 0 und 1**

Der Term kann den Wert 0 unterschreiten, der Term den Wert 1 überschreiten. ist aber immer durch 0 und 1 nach unten bzw. nach oben begrenzt. 95 %-Schätzbereiche, deren Grenzen 0 unter- bzw. 1 überschreiten, werden stattdessen von 0 nach unten bzw. von 1 nach oben begrenzt.

**Ausblick**

Den im Kasten „Schätzen von Wahrscheinlichkeiten aus Simulationen“ ermittelten 95 %-Schätzbereich nennt man in der Statistik auch 95 %-*Konfidenzintervall*.

Man könnte bei 95 % noch einen genaueren Bereich angeben, die Faustformel mit liefert eher einen größeren Bereich. Für den Einstieg reicht die Genauigkeit aber aus.

Die 95 %-Sicherheit muss man so interpretieren: Wenn man eine Simulation durchführt und dann einen Schätzbereich aus der beobachteten relativen Häufigkeit ermittelt, so enthält dieser das unbekannte p oder auch nicht. Wenn man eine solche Simulation oft durchführt liefert das Verfahren in 95 % der Fälle ein Konfidenzintervall, das das unbekannte p enthält. Der Betreiber eines Statistikbüros, das am laufenden Band Konfidenzintervalle bestimmt, liefert seinen Auftraggeber in 95 % der Fälle Intervalle, die das unbekannte p enthalten und in 5 % solche, die das nicht tun. Man weiß in keinem Einzelfall, ob wirklich in dem Intervall enthalten ist.

Das ist so, wenn man aus einer Urne mit 100 Kugeln, von denen 95 weiß und 5 schwarz sind, eine Kugel zieht, deren Farbe man aber nicht ansehen kann. Man ist sich dann zu 95 % sicher, eine weiße Kugel gezogen zu haben.