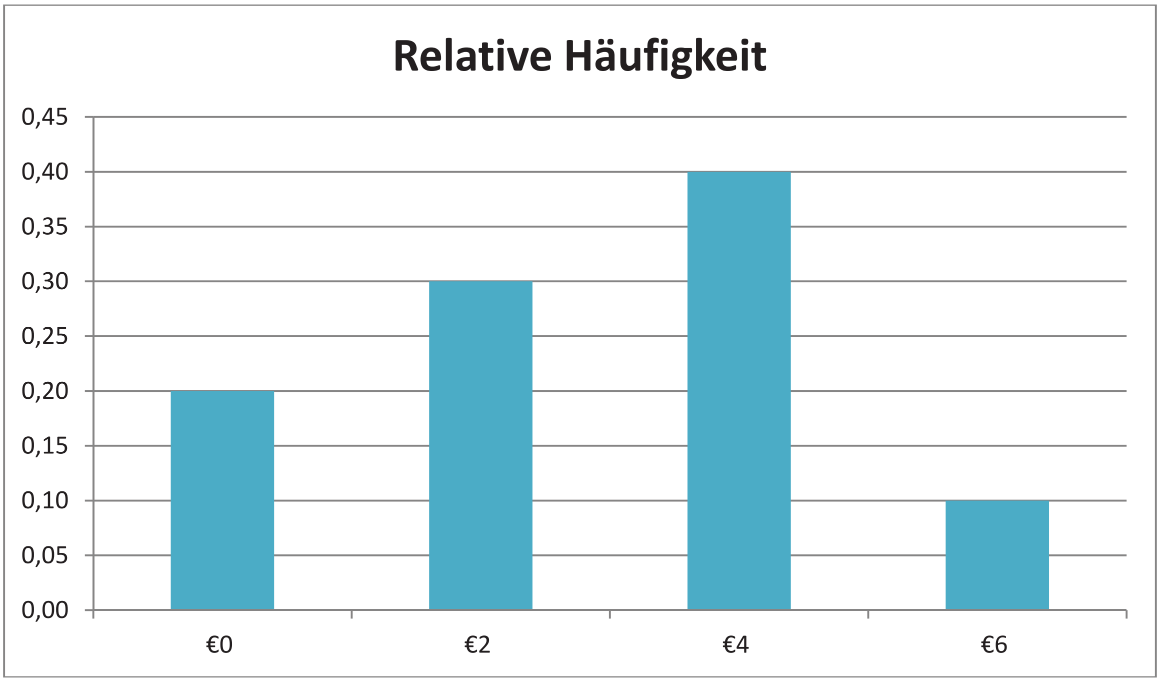
4 Stichproben und deren Kenngrößen

Beispiel: 0€, 2€, 2€, 6€, 2€, 4€, 0€, 4€, 4€, 4€ ← Stichprobe

a_i	absolute Häufigkeit $a(a_i)$	relative Häufigkeit $r(a_i)$
0€	2	$\frac{2}{10}$
2€	3	$\frac{3}{10}$
4€	4	$\frac{4}{10}$
6€	1	$\frac{1}{10}$

Bemerkung zu relative Häufigkeit:
Würde man z.B. $n = 10$ mal spielen und sich jeden Gewinn notieren, so könnte man die absolute Häufigkeit für jeden Wert angeben, als auch die relative Häufigkeit.



Wenn man den Stichprobenumfang vergrößern würde, dann liegen die relativen Häufigkeiten „in etwa“ bei der Wahrscheinlichkeit. Genau genommen konvergiert die relative Häufigkeit (bei einem Grenzwert nach Wahrscheinlichkeit und unter gewissen Voraussetzungen) gegen die entsprechende Wahrscheinlichkeit.

Von dieser Stichprobe können wir nun auch den Mittelwert bzw. das arithmetische Mittel berechnen:

$$\bar{x} = \frac{1}{n} \cdot (x_1 + \cdots + x_n) = a_1 \cdot r(a_1) + \cdots + a_m \cdot r(a_m)$$

a_m sind die möglichen Ausprägungen.

Die empirische Varianz (Stichprobenvarianz) s^2 berechnet sich wie folgt:

$$s^2 = \frac{1}{n-1} \cdot ((x_1 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2)$$

Im **Beispiel**:

$$\begin{aligned} \bar{x} &= \frac{1}{10} \cdot (0€ + 2€ + 2€ + 6€ + 2€ + 4€ + 0€ + 4€ + 4€ + 4€) \\ &= \tfrac{1}{10} \cdot (2 \cdot 0€ + 3 \cdot 2€ + 4 \cdot 4€ + 6€) = \tfrac{14}{5}€ \end{aligned}$$

$$s^2 = \frac{1}{9} \cdot \left(2 \cdot \left(0€ - \frac{14}{5}€\right)^2 + 3 \cdot \left(2€ - \frac{14}{5}€\right)^2 + 4 \cdot \left(4€ - \frac{14}{5}€\right)^2 + \left(6€ - \frac{14}{5}€\right)^2\right)$$

$$= \tfrac{56}{15}€^2 \approx 3.73€^2$$

Bemerkung zum Faktor $\frac{1}{n-1}$:
Möchte man über eine Stichprobe die Varianz der „Gesamtpopulation“ schätzen, dann wird der Faktor $\frac{1}{n-1}$ anstatt $\frac{1}{n}$ verwendet. Der Grund ist, dass (unter gewissen Voraussetzungen) damit s^2 ein Erwartungstreuer Schätzer für die theoretisch unbekannte Varianz σ^2 ist, d.h. es gilt $E(S^2) = \sigma^2$. Geht man davon aus, dass keine Stichprobe (als Teilmenge einer größeren Gesamtpopulation) vorliegt, sondern die Gesamtpopulation, dann verwendet man den Faktor 1/n. In diesem Fall wäre die relative Häufigkeit gleich der Wahrscheinlichkeit. **Achtung:** Dies machen offensichtlich die meisten Schulbücher. Die Varianz der (Gesamt)-Population ist dann:

$$s^2 = \frac{1}{n} \cdot ((x_1 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2) = r(a_1)(a_1 - \bar{x})^2 + \cdots + r(a_m)(a_m - \bar{x})^2$$

An einem Beispiel wollen wir noch mal verdeutlichen, was es mit der Varianz auf sich hat. Nehmen wir mal an, dass von 10 Personen jede 100.000€ auf dem Konto hat. Der Mittelwert der Kontostände ist dann $\bar{x}$ =100.000€. Die Streuung bzw. die Varianz ist gleich 0€².

Derselbe Mittelwert ergibt sich aber auch, wenn 9 Personen 0€ besitzen und eine Person 1.000.000€ besitzt. Nun ist aber die Streuung deutlich größer (wir betrachten die 10 Personen nun, wie die meisten Schulbücher dies tun, als Gesamtpopulation und verwenden den Faktor 1/n):

$$s^2 = \frac{1}{10} \cdot (9 \cdot (0€ - 100000€)^2 + (1000000€ - 100000€)^2)$$

$$= 90000000000\, €^2$$