

STATISTIK I – Übung 05**Varianz und Standardabweichung****1 Kurze Wiederholung**Varianz und Standardabweichung

Bei der Varianz handelt es sich um den wichtigsten und meistgenutzten Dispersionsparameter. Sie berechnet sich als die Summe der quadrierten Abweichungen der Einzelwerte einer Verteilung vom arithmetischen Mittel eben dieser Verteilung, abschließend geteilt durch die Gesamtzahl der Werte.

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

An dieser Stelle lohnt es sich, die Formel nicht nur stur anzuwenden, sondern näher zu betrachten – was wird hier eigentlich gerechnet? Zunächst einmal wird die Differenz der einzelnen Werte der Verteilung ($x_1, x_2, x_3 \dots x_n$) vom arithmetischen Mittel gebildet, die Summe aller Differenzen wird anschließend durch die Anzahl der Werte der Verteilung geteilt. Man errechnet somit quasi das arithmetische Mittel der Differenzen aller Werte vom arithmetischen Mittel – oder anders formuliert die durchschnittliche Abweichung der Werte vom Zentrum der Verteilung. Diese durchaus naheliegende und intuitive Vorgehensweise für die Bestimmung der Streuung überrascht nur in einem Detail: Die zu addierenden Differenzen werden vorab noch quadriert um zu verhindern, dass sich die positiven und die negativen Abweichungen gegenseitig neutralisieren. In der Formelsammlung findet sich neben der bekannten Formel noch diese leicht vereinfachte Berechnungsvorschrift.

$$s^2 = \left\{ \frac{1}{n} \sum_{i=1}^n x_i^2 \right\} - \bar{x}^2$$

Aufgrund des unumgänglichen Quadrierens ist allerdings auch das Ergebnis eine quadrierte Zahl, d.h. die Varianz einer Gehaltsverteilung steht nicht in der Einheit Euro, sondern in Euro². Dadurch lässt sie sich schlecht interpretieren – oder aber es besteht sogar die Gefahr einer inhaltlichen Fehlinterpretation, wenn die quadrierte Einheit „sinnvoll“ zu sein scheint (Quadratmeter). Aus diesem Grund wird bei der Berechnung der Varianz in der Regel auch noch die Standardabweichung als positive Wurzel aus der Varianz und damit „korrekt dimensionierter“ Dispersionsparameter angegeben.

$$s = +\sqrt{s^2}$$

Variationskoeffizient

Wie wir in der letzten Übungseinheit zu den Dispersionsparametern gelernt haben, spielen die Streuungsmaße unter anderem beim Vergleich von Verteilungen eine wichtige Rolle. Hierbei kann es erforderlich werden, die Standardabweichung in eine dimensionslose Größe umzuwandeln, um einen Vergleich der Streuung von Verteilungen zu ermöglichen, die in verschiedenen Dimensionen erfasst werden. Als Beispiel sei auf den Vergleich von Preisverteilungen in unterschiedlichen Währungen – etwa in Euro und mexikanische Pesos – verwiesen. Da ein Euro etwa 20 Pesos entspricht, besteht bei einem direkten Vergleich der Standardabweichungen beider Verteilungen das Risiko, die Streuung der Preisangaben in Pesos im Verhältnis zu den Preisangaben in Euro aufgrund der größeren absoluten Differenzwerte erheblich zu überschätzen. Aus diesem Grund dividiert man die Standardabweichungen beider Verteilungen vor dem Vergleich noch durch die jeweiligen arithmetischen Mittel und generiert somit den dimensionslosen Variationskoeffizienten, mit dem ein Vergleich möglich ist. Die Berechnung des Variationskoeffizienten ist nur möglich, wenn das arithmetische Mittel positiv ist.

$$v = \frac{s}{x}$$

Empirische Varianz und Stichprobenvarianz

Im Rahmen dieser Vorlesung arbeiten wir mit der sogenannten empirischen Varianz:

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

Bei der Berechnung von Varianzen mit SPSS, PSPP und anderen statistischen Analyseprogrammen wird dagegen gelegentlich die Stichprobenvarianz berechnet – allerdings nicht als solche, sondern als Varianz deklariert, was zu erheblicher Verwirrung bei Studierenden führen kann, die ihre Lösungen für Übungsaufgaben mit Hilfe von Software überprüfen möchten.

$$s_v^2 = \frac{1}{(n-1)} \sum_{i=1}^n (x_i - \bar{x})^2$$

Die Stichprobenvarianz unterscheidet sich von der empirischen Varianz darin, dass anstatt durch die Anzahl der Werte der Verteilung durch die Anzahl an Freiheitsgraden (n-1) dividiert wird – ein Begriff, auf den in späteren Übungseinheiten noch einmal näher eingegangen werden wird.

2 Beispielrechnungen

Für die Beispielrechnung greifen wir auf die Daten aus der letzten Übungseinheit zurück – Angaben zum Körpergewicht von 30 Probandinnen und Probanden.

Nr.	Körpergewicht	Nr.	Körpergewicht
1	61,2 kg	16	62,2 kg
2	72,8 kg	17	79,5 kg
3	62,3 kg	18	61,0 kg
4	85,7 kg	19	55,5 kg
5	91,8 kg	20	98,9 kg
6	64,9 kg	21	94,4 kg
7	74,3 kg	22	85,3 kg
8	95,2 kg	23	93,0 kg
9	87,3 kg	24	72,8 kg
10	82,4 kg	25	88,2 kg
11	78,4 kg	26	87,4 kg
12	91,2 kg	27	68,6 kg
13	89,3 kg	28	90,5 kg
14	76,3 kg	29	85,0 kg
15	68,2 kg	30	71,9 kg

Berechnung der Varianz

Für die Berechnung der Varianz empfiehlt sich die Anlage einer Hilfstabelle, über die sich die beiden benötigten Größen – das arithmetische Mittel sowie die Summe der quadrierten Abstände der Werte vom arithmetischen Mittel – auf einfache Art und Weise ermitteln lassen.

Nr.	Körpergewicht	$(x_i - \bar{x})^2$
1	61,2	323,28
2	72,8	40,70
3	62,3	284,93
4	85,7	42,51
5	91,8	159,26
6	64,9	203,92
7	74,3	23,81
8	95,2	256,64
9	87,3	65,93
10	82,4	10,37
11	78,4	0,61
12	91,2	144,48
13	89,3	102,41
14	76,3	8,29
15	68,2	120,56
16	62,2	288,32
17	79,5	0,10
18	61,0	330,51
19	55,5	560,74
20	98,9	388,88
21	94,4	231,65
22	85,3	37,45
23	93,0	190,99
24	72,8	40,70
25	88,2	81,36
26	87,4	67,57
27	68,6	111,94
28	90,5	128,14
29	85,0	33,87
30	71,9	53,00
Σ	2375,5	4332,96
\emptyset	79,18	144,43

Die Varianz dieser Verteilung liegt bei 144,43 kg².

(Für alle Softwarenutzer: Die Stichprobenvarianz liegt bei 149,41 kg².)

Berechnung der Standardabweichung

Die Standardabweichung berechnet sich als positive Wurzel der Varianz.

$$s = +\sqrt{s^2}$$

$$s = 12,02$$

Die Standardabweichung dieser Verteilung liegt bei 12,02 kg.

(Für alle Softwarenutzer: Die Wurzel der Stichprobenvarianz beträgt 12,22 kg.)

Da hier keine unterschiedlich dimensionierten Verteilungen miteinander verglichen werden sollen (zum Beispiel eine Gewichtsverteilung in kg und eine Gewichtsverteilung in g) erübrigt sich an dieser Stelle die Bestimmung des Variationskoeffizienten.

3 Übungsaufgaben (Lösungen folgen in der kommenden Woche)

Auch bei der Übungsaufgabe bleiben wir bei den Beispieldaten aus der vergangenen Übungseinheit – den Altersangaben der 30 schon nach ihrem Körpergewicht befragten Probandinnen und Probanden.

Nr.	Alter	Nr.	Alter
1	17 Jahre	16	33 Jahre
2	36 Jahre	17	22 Jahre
3	41 Jahre	18	62 Jahre
4	34 Jahre	19	44 Jahre
5	32 Jahre	20	41 Jahre
6	54 Jahre	21	56 Jahre
7	45 Jahre	22	62 Jahre
8	22 Jahre	23	23 Jahre
9	71 Jahre	24	86 Jahre
10	14 Jahre	25	41 Jahre
11	86 Jahre	26	65 Jahre
12	44 Jahre	27	53 Jahre
13	34 Jahre	28	35 Jahre
14	21 Jahre	29	21 Jahre
15	54 Jahre	30	19 Jahre

a) Bestimmen Sie die Varianz.

b) Bestimmen Sie die Standardabweichung.