Tanzende Grössen
oder
Die Geschichte mit dem n-1 im Nenner der Varianz

Peter Gallin, Kantonsschule Zürcher Oberland in Wetzikon

Warum berechnet man die Varianz einer Stichprobe mit der Formel

$$\frac{1}{n-1}\sum_{i=1}^{n}(x_i-m)^2 \quad ?$$

Warum dieser Nenner $n-1$ und nicht etwa n? Solche Fragen stellen unsere Schülerinnen und Schüler meist ausgerechnet dann, wenn wir auch selber nicht mehr genau wissen, wie alles zusammenhängt. Irgendwelche mehr oder weniger plausible Gründe fallen uns dann schon ein; selten werden sie aber Klarheit schaffen. Die nun folgende Herleitung ist eine Zusammenstellung der einschlägigen Schritte, die man in Lehrbüchern über Statistik meist ziemlich verstreut vorfindet. Hier liegt das Lehrbuch "Statistische Methoden und ihre Anwendungen" von Erwin Kreyszig (Vandenhoeck & Ruprecht, 1972) zugrunde. Ziel ist es, mit möglichst kleinem Aufwand sichtbar zu machen, dass die zitierte Formel eine Zufallsgrösse darstellt, deren Erwartungswert gerade die unbekannte Varianz σ^2 der Zufallsgrösse x_i ausmacht. (Wir machen in der Notation keinen Unterschied zwischen der Zufallsgrösse x_i und einer ihrer Realisierungen x_i.)

Ist man in der seltenen Lage, alle N Objekte einer Grundgesamtheit messen zu können, so erhalten wir aus den N Messwerten x_i mittels der Formeln

$$\mu=\frac{1}{N}\sum_{i=1}^{N}x_i \quad \text{und} \quad \sigma^2=\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2$$

den wahren Mittelwert μ und die wahre Varianz σ^2 (Mittelwert der Abweichungsquadrate bezüglich des Mittelwerts μ) der Grundgesamtheit. Sobald man aber nur noch eine Stichprobe vom Umfang $n < N$ erhebt, ändert sich die Situation drastisch. Jetzt tanzen nicht nur die Messwerte x_i auf der Messskala wild umher, sondern auch der Ausdruck

$$m=\frac{1}{n}\sum_{i=1}^{n}x_i \quad \text{(Mittelwert)}$$

wird zu einer – wenn auch weniger wild – tanzenden Grösse, einer Zufallsgrösse. In diesem Moment muss man vom Erwartungswert zu sprechen beginnen. In Anlehnung an den ersten Fall, bei dem man alle Messwerte kannte, ist es hilfreich, sich den Erwartungswert der Messwerte x_i vorübergehend als

$$\mu=\frac{1}{N}\sum_{i=1}^{N}x_i$$

vorzustellen, wobei N die riesige Zahl aller zu messenden Objekte der Grundgesamtheit bedeutet. Man braucht diese Formel allerdings nicht und begnügt sich mit der Schreibweise $\mu = E(x_i)$. Da mit E tatsächlich gerechnet werden muss, sind die Gesetze, denen E genügt, unumgänglich. Dass E linear ist, dass also die Gesetze

$$E(X+Y) = E(X) + E(Y),$$
$$E(aX) = a\mathrm{E}(X)$$

gelten, kann man mit ein paar Beispielen untermauern. Dass E bei *unabhängigen* Zufallsgrössen X und Y auch der Gleichung

$$E(X\cdot Y) = E(X)\cdot E(Y)$$

genügt, muss vielleicht etwas näher untersucht werden. Ein sehr einleuchtendes Beispiel liefert der Zufallsgenerator eines Taschenrechners: In einem Programm, welches das Würfeln mit dem Befehl RND simuliert und den Erwartungswert der Quadrate der Augenzahlen berechnet, erlebt man die Überraschung, dass die Befehle $(\mathrm{RND})^2$ und $(\mathrm{RND})\cdot(\mathrm{RND})$ verschiedene Erwartungswerte (nämlich 15.167 und 12.25) erzeugen. Denkt man sich die erste Version als Produkt zweier Zufallsvariablen X und Y, so sind diese extrem abhängig, weil sie ja gleich sind. In der zweiten Version, die bereits als Produkt erscheint, sind sie unabhängig, weil man zweimal auf den Zufallsgenerator zugreift. Das hat zur Folge, dass die Erwartungswerte von X und Y einfach miteinander multipliziert werden können: $3.5\cdot 3.5 = 12.25$.

Mit Hilfe der Linearität von E zeigt man nun direkt, dass der Erwartungswert der Zufallsgrösse m auch μ ist:

$$E(m)=E(\frac{1}{n}\sum x_i)=\frac{1}{n}\sum E(x_i)$$
$$=\frac{1}{n}\sum \mu=\frac{1}{n}\cdot n\cdot\mu=\mu$$

(Summation für i von 1 bis n)

Das bedeutet, dass bei häufigen Messreihen zu je n Messungen $x_1,\ldots,x_n$ erwartet werden kann, dass m um μ tanzt. Man sagt dafür auch, m sei eine Schätzung von μ, dessen Wert man jetzt ja nicht kennt.

Nun befassen wir uns damit, wie heftig der Tanz der Zufallsgrösse x_i ist. Das Mass dafür ist die Varianz σ^2. Sie ist festgelegt als Erwartungswert der Abweichungsquadrate der Messwerte bezüglich des Erwartungswerts μ.

Definition:
$\sigma^2 = E((x_i - \mu)^2)$

Auch hier ist – wie im Fall von μ – die Vorstellung von

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2$$

vorübergehend hilfreich. Da wir nun aber μ nicht kennen, kommt man auf die Vermutung, dass man mit der Grösse

$$\bar{s}^2 = \frac{1}{n}\sum(x_i - m)^2$$

eine gute Schätzung für das unbekannte σ^2 aufstellen könnte. Man beachte, dass $\bar{s}^2$ durch diese Festlegung auch zu einer tanzenden Grösse geworden ist. Man hofft also, dass $\bar{s}^2$ um das feste σ^2 tanzt. Dass dies ein Irrtum ist, soll jetzt gezeigt werden. Damit die Rechnung nicht zu unübersichtlich wird, schalten wir zwei Nebenrechnungen (I. und II.) ein, die auch für sich alleine genommen interessante Aussagen liefern.

I. Wir berechnen, wie heftig der Tanz von m um das feste aber unbekannte μ ist, das heisst, wir untersuchen den Erwartungswert des Abweichungsquadrats von m bezüglich μ:

$$\begin{aligned}
&E((m-\mu)^2) = \\
&E\left(\left(\frac{1}{n}\sum x_i - \frac{1}{n}n\mu\right)^2\right) = E\left(\left(\frac{1}{n}\sum(x_i-\mu)\right)^2\right) \\
&= \frac{1}{n^2}E\Big(\sum(x_i-\mu)^2 + 2\sum_{i<j}(x_i-\mu)(x_j-\mu)\Big) \\
&= \frac{1}{n^2}\Big(\sum E((x_i-\mu)^2) + 2\sum_{i<j}E((x_i-\mu)(x_j-\mu))\Big)
\end{aligned}$$

Da nun die Messungen x_i und x_j unabhängig voneinander sind (was bei einer Stichprobe im Gegensatz zur einer Vollunter-suchung aller N Objektre der Grundgesamtheit der Fall ist), kann man $E((x_i - \mu)\cdot(x_j - \mu))$ in $E(x_i - \mu)\cdot E(x_j - \mu)$ umformen. Wegen $\mu = E(x_i) = E(x_j)$ sind beide Faktoren Null. Damit vereinfacht sich die Rechnung schlagartig:

$$\begin{aligned}
&E((m-\mu)^2) = \frac{1}{n^2}\sum E(x_i-\mu)^2 \\
&= \frac{1}{n^2}\sum\sigma^2 = \frac{1}{n^2}(n\cdot\sigma^2) = \frac{\sigma^2}{n} \qquad \text{(I)}
\end{aligned}$$

Das ist der rechnerische Nachweis dafür, dass der Mittelwert m nur $1/n$ so stark um μ tanzt wie die Messwerte x_i.

II. So wie der Mittelwert m einer Stichprobe mit Hilfe eines beliebigen Schätzwerts M mit der Formel $m = (1/n)\cdot\sum x_i = (1/n)\cdot\sum(x_i - M) + M$ manchmal bequemer auszurechnen ist, verfügt man auch für die Berechnung einer Varianz $\bar{s}^2$ über eine entsprechende Formel (vgl. Steinerscher Satz zur Berechnung von Trägheitsmomenten), die sich auf den Schätzwert M anstatt auf m bezieht und die für $M = 0$ in der Statistik der Taschenrechner normalerweise verwendet wird:

$$\begin{aligned}
\bar{s}^2 &= \frac{1}{n}\sum(x_i-m)^2 = \frac{1}{n}\sum((x_i-M)-(m-M))^2 = \\
&= \frac{1}{n}\Big(\sum(x_i-M)^2 - 2\sum(x_i-M)(m-M) + \sum(m-M)^2\Big) \\
&= \frac{1}{n}\Big(\sum(x_i-M)^2 - 2(m-M)\sum(x_i-M) + n\cdot(m-M)^2\Big) \\
&= \frac{1}{n}\Big(\sum(x_i-M)^2 - 2(m-M)(nm-nM) + n\cdot(m-M)^2\Big) \\
&= \frac{1}{n}\Big(\sum(x_i-M)^2 - n\cdot(m-M)^2\Big) \\
&= \frac{1}{n}\sum(x_i-M)^2 - (m-M)^2 \qquad \text{(II)}
\end{aligned}$$

Man kann also den Mittelwert der Abweichungsquadrate in Bezug auf einen beliebigen Schätzwert M berechnen, sofern man ihn anschliessend um das Quadrat des mit dem Schätzwert begangenen Fehlers verkleinert.

Nun können wir uns der eigentlichen Aufgabe zuwenden und untersuchen, ob die oben definierte Grösse $\bar{s}^2$ tatsächlich um σ^2 tanzt oder nicht. Wir berechnen also den Erwartungswert von $\bar{s}^2$. Da $\bar{s}^2$ definitionsgemäss bezüglich m berechnet wird, σ^2 dagegen bezüglich μ, wenden wir gleich zu Beginn die Formel (II) an, um auf den Bezugswert μ umzuschalten:

$$\begin{aligned}
E(\bar{s}^2) &= E\Big(\frac{1}{n}\sum(x_i-m)^2\Big) \\
&= E\Big(\frac{1}{n}\sum(x_i-\mu)^2 - (m-\mu)^2\Big) \\
&= \frac{1}{n}\sum E((x_i-\mu)^2) - E((m-\mu)^2)
\end{aligned}$$

Mit Hilfe von (I) und der Definition von σ^2 ergibt sich

$$E(\bar{s}^2) = \frac{1}{n}\cdot n\cdot\sigma^2 - \frac{\sigma^2}{n} = \sigma^2\cdot\Big(\frac{n-1}{n}\Big).$$

Unsere Vermutung hat sich nicht ganz bestätigt: Der Erwartungswert von unserem $\bar{s}^2$ ist nicht σ^2, sondern ein wenig kleiner. Wenn wir also anstelle von $\bar{s}^2$ die Zahl

$$s^2 = \frac{n}{n-1}\cdot\bar{s}^2$$

als Schätzung für das gesuchte σ^2 wählen, so ist garantiert, dass wenigstens deren Erwartungswert genau σ^2 ist. Man berechnet also mittels der Formel

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

den besten Schätzwert für σ^2.