Varianzanalyse S 2006 Version 1.0

Problemstellung

Mit der Varianzanalyse können Probleme von der folgenden Art behandelt werden:

Problem 1: Man hat r verschiedene Futtermittel für Schweine und möchte wissen, ob sie sich wirklich in unterschiedlichem Maße auf die Gewichtszunahme von Schweinen auswirken. Dazu bildet man r Gruppen von Schweinen, d.h. man teilt eine Menge von n Schweinen in r Gruppen ein. Die Schweine der i-ten Gruppe werden mit dem i-ten Futtermittel gefüttert, und man misst bei jedem Schwein die Gewichtszunahme. Die Gewichtszunahme des j-ten Schweines der i-ten Gruppe sei x_{ij} . Wenn also n_i die Anzahl der Schweine der i-ten Gruppe ist, so erhält man Zahlen $x_{ij}, \quad j = 1, \ldots, n_i, \quad i = 1, \ldots, r$.

Die verschiedenen Gruppen sollen sich möglichst nur bezüglich der Fütterung voneinander unterscheiden, deshalb ist es zur Vermeidung systematischer Unterschiede sinnvoll, die Aufteilung in r Gruppen zufallsmäßig zu machen. Ferner sollte die Menge von n Schweinen eine Zufallsstichprobe sein, denn man will ja allgemeingültige Aussagen gewinnen (wenngleich natürlich mit den üblichen Fehlerwahrscheinlichkeiten) und nicht etwa Aussagen, die sich nur auf die betrachteten n Schweine beziehen.

Man kann nicht erwarten, dass alle Tiere einer Gruppe genau die gleiche Gewichtszunahme haben. Denn dem Einfluss (falls es überhaupt einen Einfluss gibt) der unterschiedlichen Fütterung überlagern sich Zufallseinflüsse, die man nicht genau kennt und nicht kontrollieren kann. Um zu klären, ob die unterschiedlichen Futtermittel einen Einfluss auf die Gewichtszunahme haben, muss man versuchen, die Zufallseinflüsse zu trennen vom (etwaigen) Einfluss des Futtermittels. Anders formuliert: in den Zahlen x_{ij} steckt eine Variabilität (falls nicht gerade alle x_{ij} denselben Wert haben), die man zerlegen kann in eine Variabilität, die durch die unterschiedliche Fütterung bewirkt wird und in eine Zufallsvariabilität. Es wird später deutlich, wieso man hier von einer Streuungszerlegung reden kann.

Eine solche Problemstellung ist eine Problemstellung der *einfachen Varianzanalyse* (einfach,weil es außer dem Zufallseinfluss einen Faktor gibt,der von Einfluss ist beziehungsweise sein kann, nämlich das Futtermittel).

Problem 2: Wenn man neben der Wirkung von r verschiedenen Futtermitteln auch die Wirkung von p verschiedenen Stallarten (mit z.B. unterschiedlich viel Platz, unterschiedlicher Temperatur o.ä.) auf die Gewichtszunahme untersuchen möchte, so teilt man die Tiere in r · p Gruppen ein und erhält Zahlen $x_{ijk}\ , \quad k = 1, \ldots, n_{ij}\ , i = 1, \ldots, r\ , \quad j = 1, \ldots, p$.

Hierbei ist n_{ij} die Anzahl der Tiere in der Gruppe (i,j), deren Tiere mit dem i-ten Futtermittel gefüttert werden und in einem Stall vom j-ten Typ gehalten werden, und x_{ijk} ist die Gewichtszunahme des k-ten Tieres in der Gruppe. (Der Vollständigkeit halber sei aber bemerkt, dass man hier die Gleichheit aller n_{ij} voraussetzt.)

Eine solche Problemstellung ist eine Problemstellung der zweifachen Varianzanalyse.

Problem 3: Man will untersuchen, ob die Ernteerträge bei r verschiedenen Düngemitteln sich im Mittel voneinander unterscheiden, anders ausgedrückt: ob das Düngemittel einen Einfluss auf den Ertrag hat.

Problem 4: Man will untersuchen, ob die mittleren Studienleistungen von Absolventen verschiedener Schultypen (z.B. Gymnasien verschiedener Richtungen, Fachoberschulen verschiedener Richtungen) sich voneinander unterscheiden.

Problem 5: Man will untersuchen, ob die mittlere Lebensdauer von Batterien fünf verschiedener Hersteller unterschiedlich ist.

Alle Problemstellungen kann man in etwas abstrahierter Form so formulieren:

Man hat verschiedene Gruppen von Einheiten (z.B. Schweine, Ackerflächen, Studenten, Batterien), die Einheiten in verschiedenen Gruppen werden verschiedenen Behandlungen (z.B. Fütterung mit verschiedenen Futtermitteln, Düngung mit verschiedenen Düngemitteln) unterzogen,im weiteren Sinne kann man auch den Besuch eines bestimmten Schultyps, die Herstellung bei einem bestimmten Hersteller als Behandlung ansehen. Es interessiert die Frage, ob die verschiedenen Behandlungen im Mittel dieselbe Wirkung haben. Dabei ist natürlich jeweils das Mittel über die jeweilige Grundgesamtheit gemeint, aus der die Gruppen als Stichprobe ausgewählt sind. [1] Die verschiedenen Grundgesamtheiten beim Batteriebeispiel sind alle von den verschiedenen Batterieherstellern erzeugten Batterien. Im Beispiel mit der Schweinefütterung müsste man sich vorstellen, dass die i-te Grundgesamtheit beispielsweise die Menge aller Schweine der BRD ist, und zwar der Menge aller Schweine der BRD, wie sie wären, wenn sie alle mit dem i-ten Futtermittel gefüttert würden. (Da ja gar nicht alle Schweine der BRD mit diesem Futtermittel gefüttert werden, ist diese Grundgesamtheit natürlich nur fiktiv.)

Einerseits ist die einfache Varianzanalyse eine Verallgemeinerung des Vergleichs zweier Mittelwerte aufgrund unabhängiger Stichproben (genaueres dazu in einem späteren Kapitel). Andererseits hat die Varianzanalyse etwas Ähnlichkeit mit der Regressionsanalyse. Worin der entscheidende Unterschied besteht, sieht man am besten, wenn man das obige Futtermittelbeispiel in ein Problem der Regressionsanalyse umwandelt: Nehmen wir an, man will wissen, wie bzw. ob überhaupt die Gewichtszunahme bei Schweinen von der Menge des gegebenen Futtermittels abhängt. Um es einfach zu machen, d.h. auf eine einfache lineare Regression zu kommen, sei angenommen, dass nur ein Futtermittel untersucht wird. Man variiert bei dem Versuch die Dosis des Futtermittels, d.h. gibt verschiedenen Schweinen verschiedene Mengen. Hier ist jetzt das Futtermittel ein sogenannter quantitativer Faktor. Bei dem obigen Problem der Varianzanalyse waren die Futtermittel jedoch sogenannte qualitative Faktoren: ein Futtermittel einer bestimmten Marke wird entweder gegeben oder nicht, und wenn es gegeben wird, dann in einer ganz bestimmten, festen Dosis. Man sieht, dass man zu dem Beispiel mit den Batterien nicht so leicht ein ähnliches Regressionsanalyse-Problem finden kann: denn eine Batterie wird (i.a.) bei einem bestimmten Hersteller hergestellt oder nicht, und daneben gibt es keine quantitativen Abstufungen.

Das mathematische Modell der einfachen Varianzanalyse

Gegeben sind $n = n_1 + n_2 + \ldots + n_r$ Realisierungen $x_{11}, \ldots, x_{1n_1}$, $x_{21}, \ldots, x_{2n_2}$, $\ldots, x_{r1}, \ldots, x_{rn_r}$ von stochastisch unabhängigen Zufallsgrößen $X_{11}, \ldots, X_{1n_1}$, $X_{21}, \ldots,$ X_{2n_2} , $X_{r1}, \ldots, X_{rn_r}$.

Die Verteilung von X_{ij} ist eine Normalverteilung mit Mittelwert μ_i und Varianz σ^2 (d.h. die Varianz hängt weder von i noch von j ab, sie ist also für alle r Gruppen dieselbe, der Mittelwert hängt nur von i, genauer: höchstens von i ab).

Es wird die Fragestellung untersucht, ob $\mu_1 = \mu_2 = \ldots = \mu_r$ gilt, d.h. die Mittelwerte der Verteilung für alle r Gruppen dieselben sind oder nicht. "Nicht" bedeutet dabei natürlich: es gibt mindestens ein i_1 und i_2 mit $\mu_{i_1} \neq \mu_{i_2}$. Wie schon vorher in einer Fußnote bemerkt wurde, geht es selbstverständlich nicht darum, ob die r Stichprobenmittelwerte gleich sind. Man sucht also einen Test zum Vergleich (genauer: zum simultanen Vergleich) der Mittelwerte mehrerer Normalverteilungen. Es wird sich zeigen, dass der unten hergeleitete Test für den Spezialfall r=2 äquivalent zum früher behandelten Test zum Vergleich der Mittelwerte zweier Normalverteilungen aufgrund zweier unabhängiger Stichproben ist. ("Äquivalent" bedeutet, dass beide Tests in jeder Situation dasselbe Ergebnis liefern, natürlich nicht, dass sie formal identisch sind).

Der Vollständigkeit halber sei bemerkt, dass es daneben noch die -ziemlich triviale- Aufgabenstellung gibt, die Mittelwerte μ_i zu schätzen. Selbstverständlich nimmt man als Schätzung für den Mittelwert μ_i bei der i-ten Gruppe

$\hat{\mu_i} = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}$

also das Stichprobenmittel der i-ten Gruppe.

Es ist bei der Varianzanalyse (und nicht nur dabei) äußerst zweckmäßig, die folgende Schreibweise für Summen und Durchschnitte zu verwenden, die am Beispiel erläutert wird:

$x_{i.} := \sum_{j=1}^{n_i} x_{ij}$

[1] Ob die aus den Gruppen errechneten Mittelwerte gleich sind oder nicht, ist eine triviale Rechenaufgabe, die mit *schließender* Statistik nichts zu tun hat und natürlich ein klares Ja oder Nein ohne irgendwelche Fehlermöglichkeiten liefert.

$\overline{x_{i.}} := \frac{1}{n_i} \cdot x_{i.}$
$x_{..} := \sum_{i=1}^{r} x_{i.} = \sum_{i=1}^{r} \sum_{j=1}^{n_i} x_{ij}$
$\overline{x_{..}} := \frac{1}{n_1+n_2+\ldots+n_r} \cdot x_{..} = \frac{1}{n} x_{..}$
Mit dieser Schreibweise wäre bei einer Matrix $A = (a_{ij})$ also $a_{i.}$ die Summe über die Elemente der i-ten Zeile, $a_{.j}$ die Summe über die Elemente der j-ten Spalte).

Test zum Vergleich der Mittelwerte mehrerer Normalverteilungen (F-Test der einfachen Varianzanalyse)

Es treffe das oben dargestellte Modell der einfachen Varianzanalyse zu. Gesucht ist ein Test für

H: $\mu_1 = \mu_2 = \ldots = \mu_r$ gegen K : es gibt mindestens ein i und k mit $\mu_i \neq \mu_k$

Die Gegenhypothese K bedeutet keineswegs, dass alle Mittelwerte μ_i voneinander verschieden sein müssen, sondern nur, dass mindestens zwei voneinander verschieden sein müssen. Das ist gerade die Verneinung der Hypothese H.

Die folgenden Überlegungen sollen den Test plausibel machen.

Zur gesamten Streuung der x_{ij}-Werte gehört die Quadratsumme
$q = \sum_{i=1}^{r} \sum_{j=1}^{n_i} (x_{ij} - \overline{x_{..}})^2$
Diese Quadratsumme kann man entsprechend dem folgenden Satz zerlegen:

Satz: Es gilt
$\sum_{i=1}^{r} \sum_{j=1}^{n_i} (x_{ij} - \overline{x_{..}})^2 = \sum_{i=1}^{r} \sum_{j=1}^{n_i} (x_{ij} - \overline{x_{i.}})^2 + \sum_{i=1}^{r} n_i \cdot (\overline{x_{i.}} - \overline{x_{..}})^2$
Wenn im folgenden auf diese Gleichung Bezug genommen wird, wird der Term auf der linken Seite mit q bezeichnet, und die Terme auf der rechten Seite werden der Reihe nach mit q_1 und q_2 bezeichnet.

Beweis: Man formt $x_{ij} - \overline{x_{..}}$ um, indem man $\overline{x_{i.}}$ dazwischenschiebt

$x_{ij} - \overline{x_{..}} = x_{ij} - \overline{x_{i.}} + \overline{x_{i.}} - \overline{x_{..}} = (x_{ij} - \overline{x_{i.}}) + (\overline{x_{i.}} - \overline{x_{..}})$

Quadrieren dieser Gleichung liefert nach dem binomischen Lehrsatz

$(x_{ij} - \overline{x_{..}})^2 = (x_{ij} - \overline{x_{i.}})^2 + 2 \cdot (x_{ij} - \overline{x_{i.}}) \cdot (\overline{x_{i.}} - \overline{x_{..}}) + (\overline{x_{i.}} - \overline{x_{..}})^2$

Also gilt für $i = 1, \ldots, r$
$\sum_{j=1}^{n_i} (x_{ij} - \overline{x_{..}})^2 = \sum_{j=1}^{n_i} (x_{ij} - \overline{x_{i.}})^2 + 2 \cdot \sum_{j=1}^{n_i} (x_{ij} - \overline{x_{i.}}) \cdot (\overline{x_{i.}} - \overline{x_{..}}) + n_i \cdot (\overline{x_{i.}} - \overline{x_{..}})^2 =$
$\sum_{j=1}^{n_i} (x_{ij} - \overline{x_{i.}})^2 + n_i \cdot (\overline{x_{i.}} - \overline{x_{..}})^2$
Dabei erklärt sich das letzte Gleichheitszeichen daraus, dass $\sum_{j=1}^{n_i} (x_{ij} - \overline{x_{i.}}) = 0$ gilt und der Faktor $\overline{x_{i.}} - \overline{x_{..}}$ nicht von j abhängt.

Wenn man nun noch die hergeleitete Gleichung über $i = 1, \ldots, r$ aufsummiert, dann erhält man die zu beweisende Gleichung $\boxed{-}$

Bemerkung: $q_1 = \sum_{i=1}^{r} \sum_{j=1}^{n_i} (x_{ij} - \overline{x_{i.}})^2$ kann man als Quadratsumme, die zur Streuung *innerhalb* der Gruppen gehört, auffassen, und $q_2 = \sum_{i=1}^{r} n_i \cdot (\overline{x_{i.}} - \overline{x_{..}})^2$ kann man als Quadratsumme, die zur Streuung *zwischen* den Gruppen gehört, betrachten. Wenn jeweils alle x-Werte in einer Gruppe gleich wären, so wäre q_1 gleich 0. Wenn aber alle Gruppen denselben Mittelwert der x-Werte hätten (gemeint ist hier der Stichprobenmittelwert $\overline{x_{i.}}$, nicht der Mittelwert μ_i der Verteilung von X_{ij}), so wäre q_2 gleich 0.

Die folgenden Überlegungen dienen zur Motivierung der Prüfgröße, die man beim Test zum Vergleich der Mittelwerte mehrerer Normalverteilungen verwendet.

Wenn die Hypothese H zutrifft, so ist
$\frac{1}{\sigma^2} \cdot q_2(X_{11}, \ldots, X_{rn_r}) = \frac{1}{\sigma^2} \cdot \sum_{i=1}^{r} n_i \cdot (\overline{X_{i.}} - \overline{X_{..}})^2$
chiquadratverteilt mit $r - 1$ Freiheitsgraden, wie nun begründet wird. Wenn die Hypothese zutrifft, kann man

$\mu := \mu_1 = \mu_2 = \ldots = \mu_r$ setzen. Für $i = 1, \ldots, r$ ist dann

$\frac{\overline{X_{i.}} - \mu}{\frac{\sigma}{\sqrt{n_i}}}$

standardnormalverteilt, und diese r Zufallsgrößen sind voneinander unabhängig. Folglich ist die Summe der Quadrate dieser Zufallsgrößen, also

$\sum_{i=1}^{r} (\frac{\overline{X_{i.}} - \mu}{\frac{\sigma}{\sqrt{n_i}}})^2 = \frac{1}{\sigma^2} \sum_{i=1}^{r} n_i \cdot (\overline{X_{i.}} - \mu)^2$

bekanntlich chiquadratverteilt mit r Freiheitsgraden. Ersetzt man hier μ durch $\overline{X_{..}}$, so verringert sich die Zahl der Freiheitsgrade um 1. Die so gebildete Größe

$\frac{1}{\sigma^2} \cdot \sum_{i=1}^{r} n_i \cdot (\overline{X_{i.}} - \overline{X_{..}})^2$

ist also χ^2_{r-1}-verteilt.

Wenn nun σ^2 bekannt wäre, so wäre es naheliegend, diese Größe als Prüfgröße beim Test zu verwenden, und zwar so, dass man sich bei kleinen Werten der Prüfgröße für H, bei großen Werten der Prüfgröße für K entscheidet. Denn wenn die wahren Mittelwerte μ_i alle gleich sind, so ist zu erwarten, dass die Stichprobenmittelwerte $\overline{X_{i.}}$ der verschiedenen Gruppen Werte annehmen, die sich nicht stark voneinander und somit auch nicht stark von $\overline{X_{..}}$ unterscheiden und somit die obige Größe einen kleinen Wert annimmt.

Weil aber σ^2 nicht bekannt ist, wird es durch einen Schätzer ersetzt, nämlich durch

$\frac{1}{n-r} \cdot q_1(X_{11}, \ldots, X_{rn_r}) = \frac{1}{n-r} \sum_{i=1}^{r} \sum_{j=1}^{n_i} (X_{ij} - \overline{X_{i.}})^2$

Dieser Schätzer ist übrigens immer (nicht nur unter der Hypothese) erwartungstreu, denn es gilt für $i = 1, \ldots, r$

$E \sum_{j=1}^{n_i} (X_{ij} - \overline{X_{i.}})^2 = (n_i - 1) \cdot E \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (X_{ij} - \overline{X_{i.}})^2 = (n_i - 1) \cdot \sigma^2$

und somit

$E \frac{1}{n-r} \sum_{i=1}^{r} \sum_{j=1}^{n_i} (X_{ij} - \overline{X_{i.}})^2 = \frac{1}{n-r} \sum_{i=1}^{r} E \sum_{j=1}^{n_i} (X_{ij} - \overline{X_{i.}})^2 = \frac{1}{n-r} \sum_{i=1}^{r} (n_i - 1) \cdot \sigma^2 =$
$\frac{1}{n-r} \cdot (n_1 + \ldots + n_r - r) \cdot \sigma^2 = \frac{n-r}{n-r} \sigma^2 = \sigma^2$

Nun könnte man also plausiblerweise den Bruch

$\frac{\sum_{i=1}^{r} n_i \cdot (\overline{X_{i.}} - \overline{X_{..}})^2}{\frac{1}{n-r} \sum_{i=1}^{r} \sum_{j=1}^{n_i} (X_{ij} - \overline{X_{i.}})^2}$

als Prüfgröße verwenden, aber stattdessen dividiert man noch den Zähler durch $r - 1$. Dann hat nämlich die Prüfgröße eine bekannte Verteilung, wie aus dem folgenden Satz hervorgeht, und der Test ändert sich nicht [2] , denn es ist egal, ob man $T(x_{11}, \ldots, x_{rn_r})$ mit einer Vergleichsgröße c oder das $\frac{1}{r}$-fache mit dem $\frac{1}{r}$-fachen von c vergleicht.

[2] Es ändert sich wohl die Form, in der man den Test angibt, aber die Abbildung d, die dem Argument $x_{11}, \ldots, x_{r n_r}$ die Entscheidung d_h bzw. d_K zuordnet, ändert sich nicht.