

Planen und Auswerten mehrfaktorieller Versuche ausformuliertes Skript

Dr. Karin Hartung

Inhaltsverzeichnis

Einführung.....	3
Lineare Modelle.....	3
Varianzanalyse.....	5
Zusammenhang von p-Wert und F-Wert bzw. t-Wert.....	8
Voraussetzungen für eine Varianzanalyse.....	8
Transformation	9
Korrelierte Daten.....	12
Regression und Kovarianzanalyse.....	12
Signifikanz und Relevanz.....	13
Prinzipien der Versuchsplanung:	
Randomisation, Wiederholung und Blockbildung.....	13
Wiederholung: das Ermitteln der Varianz.....	14
Blockbildung: das Ausschließen von Störgrößen.....	15
Wiederholungsanzahl.....	15
Versuchsdesigns mit vollständigen Blöcken.....	15
Vollständig randomisierte Anlage	
(engl: Completely Randomized Design = CRD).....	16
Randomisierte vollständige Blockanlage	
(engl: Randomized Complete Block Design = RCBD).....	17
Lateinisches Quadrat (engl: Latin Square Design = LSD)	17
Spaltanlage (engl: Split plot).....	18
Streifenanlage (engl: Strip plot).....	19
Unvollständige Blöcke.....	20
Feste und zufällige Effekte.....	22
Fehlende Werte.....	23
Modellbildung.....	23
Literatur.....	24

Einführung

Bei Versuchen und Erhebungen werden viele Einzeldaten gewonnen. Diese lassen sich ohne Aufbereitung nicht auf einmal erfassen oder beurteilen. Eine Möglichkeit diese Daten in eine interpretierbare Form zu bringen, ist die grafische Darstellung, beschreibende Zahlen eine andere. Letzteres ist beschreibende Statistik hierzu gehören Mittelwert, Varianz und Standardabweichung der Rohdaten. Einen Schritt weiter geht die schließende Statistik. Diese will die Daten nicht nur beschreiben sondern Gruppen miteinander vergleichen. Ihr Ziel ist es den Einfluss einzelner Faktoren ohne Störgrößen darzustellen und aufzuzeigen welche Gruppen sich wirklich unterscheiden. Allerdings kann schließende Statistik keine Aussage darüberemachen, ob die gefundene Signifikanz auch relevant ist.

Lineare Modelle

Viele Versuche können mittels Varianzanalyse ausgewertet werden. Dafür muss man eine Varianz berechnen können. Das bedeutet, dass man mindestens zwei Werte je Behandlung braucht. Besser ist es drei oder mehr Wiederholungen zu haben. Die abhängige Variable (y) muss für eine Varianzanalyse quantitativ sein, die beeinflussenden Variablen können quantitativ oder qualitativ sein oder wie z.B. im Fall der Kovarianzanalyse eine Kombination aus beidem sein. Sind alle beeinflussenden Variablen quantitativ spricht man von 'Regression'. Der Varianzanalyse liegt in der Regel ein lineares Modell zugrunde. Diese Modelle setzen voraus, dass es einen linearen, d.h. rein additiven Zusammenhang zwischen den einzelnen Effekten im Modell gibt oder dass dieser durch Transformation der Daten erzielt werden kann.

Für die Versuchsfrage:

„Wie verhalten sich 3 Sorten bei unterschiedlicher Düngung (2 Düngerarten)?“ oder

„Wie verhalten sich 3 Rassen bei unterschiedlicher Fütterung (2 Futterarten)?“

Bedeutet dies, dass das Modell für den Behandlungsanteil wie folgt lautet:

$$y_{ik} = \tau_{ik} + e_{ik}$$

mit τ_{ik} : Behandlungskombination der i-te Sorte bei k-ter Düngung

Von dieser Behandlungskombination kann man einen allgemeinen Mittelwert (μ) abtrennen, so dass sich folgendes Modell ergibt:

$$y_{ik} = \mu + \tau_{ik} + e_{ik}$$

Dieser Mittelwert ergibt sich aus der Summe aller Beobachtungen geteilt durch die Anzahl aller Beobachtungen. Der Wert von τ_{ik} ist nun die Abweichung der einzelnen Beobachtung vom allgemeinen Mittelwert.

Will man diesen Versuch nicht einfaktoriell auswerten, sondern möchte die Faktoren Sorte und Düngung separat betrachten, kann jede einzelne Behandlung τ_{ik} in einen Sorteneffekt und einen Düngereffekt aufgeteilt werden.

$$\tau_{ik} = \alpha_i + \delta_k$$

mit α_i : i-te Sorte und δ_k : k-te Düngung

Dies führt zu folgendem Modell

$$y_{ik} = \mu + \alpha_i + \delta_k + e_{ik}$$

Dieses Modell beschreibt den Fall, dass sich alle Sorten bei beiden Düngerarten gleich verhalten, also Sorte A bei beiden Düngungen immer auf die gleiche Art besser ist als Sorte C und immer schlechter als Sorte B. Das Modell sagt aber auch, dass sich beide Düngerarten bei allen Sorten gleich verhalten, d.h. Dünger 2 immer gleich viel besser ist als Dünger 1 (Abb. 1 links). Nun kann es aber sein, dass sich zumindest eine Sorte (hier C) bei einem Dünger anders verhält als die beiden anderen (Abb. 1 rechts).

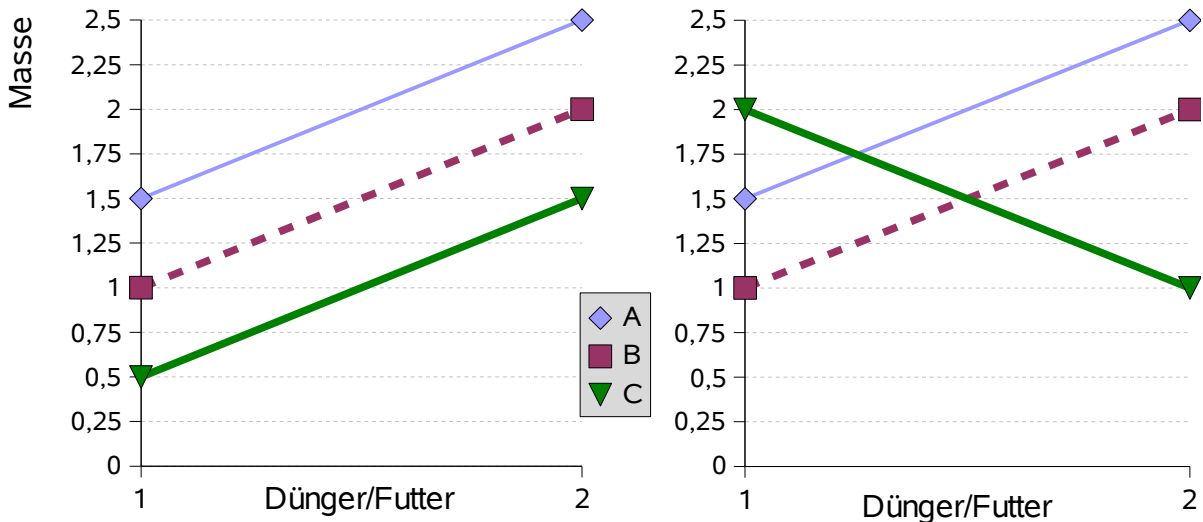


Abbildung 1: Massezunahme bei 3 Sorten bzw. Rassen (A, B, C) bei unterschiedlicher Dünger- bzw. Futtergabe; links haben nur die Haupteffekte einen Einfluss, rechts gibt es eine Interaktion, d.h. C verhält sich anders als A und B.

Jetzt ist es nicht mehr möglich etwas allgemeingültiges über das Verhalten der drei Sorten bei unterschiedlicher Düngung bzw. der beiden Dünger bei unterschiedlichen Sorten zu sagen. Es liegt eine Wechselwirkung zwischen Sorte und Dünger vor. Diese Wechselwirkung, oder auch Interaktion, muss im Modell berücksichtigt werden. Hierzu muss ein Interaktionsterm ($\alpha\delta_{ij}$) ins Modell aufgenommen werden. Das Modell sieht dann wie folgt aus:

$$y_{ik} = \mu + \alpha_i + \delta_k + \alpha\delta_{ik} + e_{ik}$$

Neben den Behandlungen gibt es bei einem Versuch auch immer ein Versuchsdesign, d.h. die Anordnung der Behandlungen in einem Versuch folgt einem bestimmten Muster/Design. Ordnet man die Wiederholungen jeder Sorte-Düngerkombination völlig zufällig den einzelnen Versuchsparzelle zu, nennt sich dieses Versuchsdesign „vollständig randomisierte Versuchsanlage“. Man kann aber auch bestimmte Einheiten zusammenfassen, z.B. vollständige Wiederholungen. Dieses Versuchsdesign ist im einfachen Falle eine randomisierte vollständige Blockanlage. Das für den Versuch gewählte Design muss im Modell berücksichtigt werden. Im Fall der vollständig randomisierte Versuchsanlage gibt es keine Randomisationseinheit, die Untersuchungseinheiten gruppiert. Daher ist auch kein Designanteil ins Modell aufzunehmen. Bei der Blockanlage gibt es eine solche Untereinheit: die Blöcke. Wenn das Sorten-Dünger-Beispiel als Blockanlage angelegt wurde, wird das bisherige Modell $y_{ik} = \mu + \alpha_i + \delta_k + \alpha\delta_{ik} + e_{ik}$ um einen Blockeffekt (β_j) erweitert, sodass das Modell nun wie folgt lautet:

$$y_{ijk} = \mu + \beta_j + \alpha_i + \delta_k + \alpha\delta_{ik} + e_{ijk}$$

Man hat nun ein Modell in dem eine abhängige Variable (y_{ijk}) von mehreren beeinflussenden Variablen (β_j , α_i , δ_k) und der Interaktionen der Behandlungseffekte ($\alpha\delta_{ik}$) erklärt wird. Dieses

Modell setzt sich wie jedes Modell aus einem Designanteil (β_j) und einem Behandlungsanteil ($\alpha_i + \delta_k + \alpha\delta_{ik}$) zusammen. Zwischen Design- und Behandlungsanteil gibt es in der Regel keine Interaktion. Hinzu kommen noch der allgemeine Mittelwert (μ) und der Restfehler (e_{ijk}).

In einem Experiment versucht man abgesehen von den Behandlungen für alle Prüfglieder alle weiteren beeinflussenden Faktoren konstant zu halten. Geht das nicht versucht man Prüfglieder, die dem gleichen Einfluss ausgesetzt sind, in Gruppen zusammenzufassen. In der anschließenden Analyse kann der Einflussfaktor dann berücksichtigt werden. Beeinflussenden Faktoren, die nicht kontrolliert werden können beruhen häufig auf Umwelteinflüssen. Diese können die Form von Gradienten haben, z.B. Feuchtigkeit oder Nährstoffversorgung am Hang, Licht im Gewächshaus, Temperatur im Klimaschrank, oder völlig heterogen sein, z.B. Bodenstruktur, Begleitvegetation, Regenmengen. Bei Tierversuchen könnte es sich um Anfangsgewicht, Alter, Nährstoff- und Nahrungsaufnahme, Keimkonzentration, Luftzirkulation oder Temperatur im Stall handeln. All diese Faktoren können für sich einen Einfluss auf die Prüfglieder haben oder mit anderen Faktoren, also auch den Behandlungen, interagieren. Auch die räumliche-zeitliche Position innerhalb des Versuchsdesigns kann einen Einfluss auf das Versuchsglied haben, z.B. Randeffekte im Feldversuch, Zeiteffekte bei Laboranalysen. Diese Faktoren können häufig entweder durch geschickte Wahl des Versuchsdesigns oder durch reines Erfassen später in der Auswertung berücksichtigt werden oder schlagen sich andernfalls im Versuchsfehler nieder und vergrößern diesen.

Varianzanalyse

Betrachtet man den die Beobachtungswerte (y) vom Standpunkt der linearen Modelle kommend, so setzt sich, wie schon gezeigt, jeder y -Wert aus einzelnen Effekten zusammen. Bei einem Versuch in dem drei Sorten in zwei Blöcken getestet werden, sieht das Modell wie folgt aus:

$$y_{ij} = \mu + \beta_j + \alpha_i + e_{ij}.$$

Die Werte könnten wie folgt aussehen:

	SorteA	SorteB	SorteC	Block-Mittelwert	allgemeiner Mittelwert
Block I	35,2	46,2	56,6	46	
Block II	37,8	46,8	56,4	47	
Sorten-Mittelwert	36,5	46,5	56,5		46,5

Um μ zu erhalten bildet man den Mittelwert: $(35,2 + 46,2 + 56,6 + 37,8 + 46,8 + 56,4)/6 = 46,5$. Diesen kann man nun von jedem Wert abziehen und erhält so die um den Mittelwert bereinigten Werte für die Sorten in den Blöcken.

	SorteA	SorteB	SorteC	Block-Mittelwert
Block I	-11,3	-0,3	10,1	-0,5
Block II	- 8,7	0,3	9,9	0,5
Sorte-Mittelwert	-10,0	0,0	10,0	

Nun kann man die Mittelwerte der Blöcke bilden und dies subtrahieren, um den Sorteneffekt zu

erhalten der nur noch vom Restfehler beeinflusst ist.

	SorteA	SorteB	SorteC
Block I	-10,8	0,2	10,6
Block II	- 9,2	-0,2	9,4
Sorte-Mittelwert	-10,0	0,0	10,0

Subtrahiert man den Sortenmittelwert ergibt das den Restfehler.

	SorteA	SorteB	SorteC
Block I	-0,8	0,2	0,6
Block II	0,8	-0,2	-0,6

Die erhaltenen Mittelwerte jeder Sorte und jedes Blockes sind nun die Effekte der Sorte bzw. des Blocks im linearen Modell, dass man dem linearen Modell $y_{ijl} = \mu + \beta_j + \alpha_i + e_l$ entsprechend wie folgt darstellen kann:

$$\begin{bmatrix} 35,2 & 46,2 & 56,6 \\ 37,8 & 46,8 & 56,4 \end{bmatrix} = \begin{bmatrix} 46,5 & 46,5 & 46,5 \\ 46,5 & 46,5 & 46,5 \end{bmatrix} + \begin{bmatrix} -0,5 & -0,5 & -0,5 \\ 0,5 & 0,5 & 0,5 \end{bmatrix} + \begin{bmatrix} -10 & 0 & 10 \\ -10 & 0 & 10 \end{bmatrix} + \begin{bmatrix} -0,8 & 0,2 & 0,6 \\ 0,8 & -0,2 & -0,6 \end{bmatrix}$$

Nun kann man die Quadratsummen (SQ) von Blockeffekt (SQ_{Block}), Sorteneffekt (SQ_{Sorte}) und Restfehler (SQ_{Fehler}) errechnen.

$$SQ_{\text{Block}} = (-0,5)^2 + (-0,5)^2 + (-0,5)^2 + (0,5)^2 + (0,5)^2 + (0,5)^2 = 1,5$$

$$SQ_{\text{Sorte}} = (-10)^2 + (0)^2 + (10)^2 + (-10)^2 + (0)^2 + (10)^2 = 400$$

$$SQ_{\text{Fehler}} = (-0,8)^2 + (0,2)^2 + (0,6)^2 + (0,8)^2 + (-0,2)^2 + (-0,6)^2 = 2,08$$

Diese Werte erhält man auch, wenn man den üblicheren Weg geht und die Formeln nutzt, die in der Literatur zu finden sind:

$$SQ_{\text{Sorte}} = r \sum_{i=1}^t (\bar{y}_{.i} - \bar{y}_{..})^2 = \sum_{i=1}^t y_{.i}^2 / r - y_{..}^2 / (rt)$$

$$SQ_{\text{Block}} = t \sum_{j=1}^r (\bar{y}_{.j} - \bar{y}_{..})^2 = \sum_{j=1}^r y_{.j}^2 / t - y_{..}^2 / (rt)$$

$$SQ_{\text{Fehler}} = t \sum_{j=1}^r (\bar{y}_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2 = \sum_{i=1}^t \sum_{j=1}^r y_{ij}^2 - \sum_{j=1}^r y_{.j}^2 / t - \sum_{i=1}^t y_{i.}^2 / r + y_{..}^2 / (rt)$$

Um die Werte für das Beispiel zu berechnen, helfen die folgenden Terme:

$$I = \sum_{j=1}^r y_{.j}^2 / t = (138^2 + 141^2) / 3 = 38925 / 3 = 12975$$

$$II = \sum_{i=1}^t y_{i.}^2 / r = (73^2 + 93^2 + 113^2) / 2 = 13373,5$$

$$III = y_{..}^2 / (rt) = (35,2 + \dots + 56,4)^2 / 6 = 279^2 / 6 = 12973,5$$

$$IV = \sum_{i=1}^t \sum_{j=1}^r y_{ij}^2 = (35,2^2 + \dots + 56,4^2) = 13377,08$$

$$SQ_{\text{Sorte}} = II - III = 13373,5 - 12973,5 = 400$$

$$SQ_{\text{Block}} = I - III = 12975 - 12973,5 = 1,5$$

$$SQ_{\text{Fehler}} = IV - I - II + III = 13377,08 - 12975 - 13373,5 + 12973,5 = 2,08$$

Es ergeben sich die gleichen Quadratsummen. Die Werte können nun wie folgt in einer Varianzanalysetabelle angegeben werden. Wobei die Freiheitsgrade (DF) der einzelnen Effekte benötigt werden, um das Mittelquadrat (MQ) zu berechnen. Teilt man nun das MQ des interessierenden Effekts durch den MQ des Restfehlers, so erhält man den F-Wert des Versuches, denn man dann mit dem zugehörigen F-Wert aus der Tabelle vergleicht. Statt des F-Werts geben die meisten Statistikprogramme heute den p-Wert an, der eine Größe darstellt die direkt mit dem jeweils gewählten Wert des Fehlers erster Art (Alpha) verglichen werden kann.

Im allgemeinen berechnen sich die Freiheitsgrade aus der Differenz der Anzahl der Beobachtungen (n) abzüglich eines Freiheitsgrads für den Mittelwert und der benötigten Freiheitsgrade je Effekt im Modell.

Für die hier gezeigte Blockanlage ergeben sich daher mit n=6, r=2, und t=3:

$$DF = n-1-(r-1)-(t-1)=6-1-(2-1)-(3-1)= 2$$

<i>Effekt</i>	<i>DF</i>	<i>SQ</i>	<i>MQ</i>	<i>F – Wert</i>
<i>Block</i>	$(r-1)$	SQ_{Sorte}	$MQ_{Block} = \frac{SQ_{Block}}{(r-1)}$	$F_{Block} = \frac{MQ_{Block}}{MQ_{Fehler}}$
<i>Sorte</i>	$(t-1)$	SQ_{Block}	$MQ_{Sorte} = \frac{SQ_{Sorte}}{(t-1)}$	$F_{Sorte} = \frac{MQ_{Sorte}}{MQ_{Fehler}}$
<i>Fehler</i>	$(t-1)*(r-1)$	SQ_{Fehler}	$MQ_{Fehler} = \frac{SQ_{Fehler}}{((t-1)*(r-1))}$	

Effekt	DF	SQ	MQ	F-Wert	p-Wert
Block	1	1,50	1,50	1,44	0,3527
Sorte	2	400,00	200,00	192,31	0,0052
Fehler	2	2,08	1,04		

Zusammenhang von p-Wert und F-Wert bzw. t-Wert

Als Ergebnis einer Varianzanalyse erhält man den F-Wert des jeweils betrachteten Effekts. Hat man einen t-Test durchgeführt, erhält man den zugehörigen t-Wert. Wobei der t-Test, der Vergleich von genau zwei Gruppen, als Spezialfall des F-Tests angesehen werden kann, da beim F-Test beliebig viele Gruppen verglichen werden können. Dieser Zusammenhang zeigt sich mathematisch insofern, als der quadrierte t-Wert gleich dem entsprechenden F-Wert ist. p-Wert und t-Wert – und ebenso der F-Wert – hängen in soweit zusammen, als der p-Wert die zum t-Wert zugehörige Fläche unter der Wahrscheinlichkeits-Kurve angibt (Abb. 2). Flächeberechnung ist rechenintensiv. Daher war es bevor rechenstarke Computer zur Verfügung standen, notwendig für bestimmte Alphaniveaus und verschiedene Freiheitsgradkombinationen Tabellen zu erstellen. Aus diesen konnte der zugehörige t-Wert abgelesen werden. Heute geben Statistikprogramme den p-Wert an.

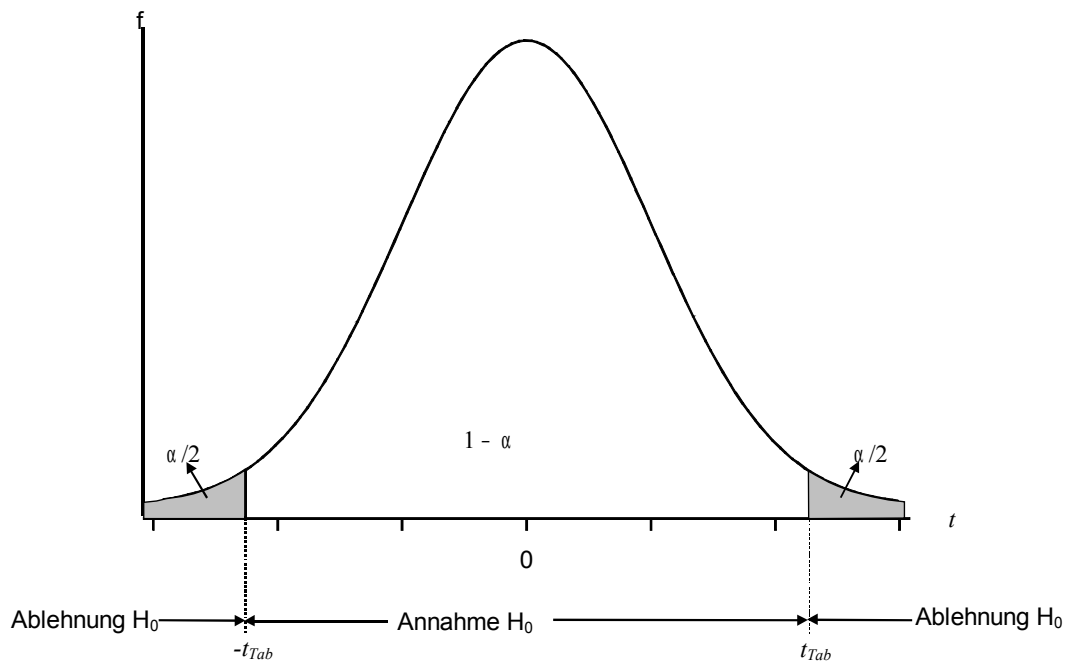


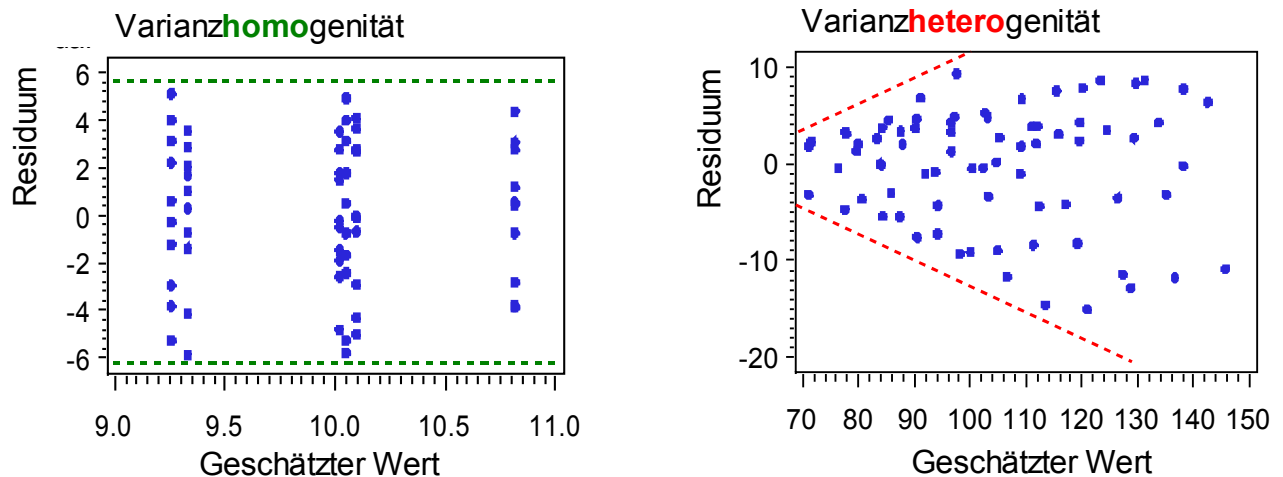
Abbildung 2: Dichteverteilung einer t-Verteilung mit eingetragenem Alphaniveau und zugehörigen Werten der t-Tabelle. (Aus Piepho 2009: Statistik-Skript)

Voraussetzungen für eine Varianzanalyse

Wie schon weiter oben erwähnt, gibt es notwendige Voraussetzungen, die erfüllt sein müssen, um eine Varianzanalyse durchführen zu dürfen. Diese Voraussetzungen sind neben der Additivität *Varianzhomogenität* und *Normalverteilung* der Residuen (Restfehler). Varianzhomogenität und Normalverteilung kann man sowohl numerisch als auch grafisch prüfen. Für die visuelle Überprüfung der Varianzhomogenität trägt man grafisch die Residuen gegen den vorhergesagten Wert auf. Die Schwankung (Varianz) der Werte der Residuen sollte dabei über den gesamten Bereich der geschätzten Werte gleich sein (Abb. 3 oben links). Falsch wäre z.B., wenn bei kleinen geschätzten Werten die Varianz klein und bei großen geschätzten Werten groß wäre (Abb. 3 oben rechts). Die Gruppierungen in Abbildung 3 oben links ergeben sich dadurch, dass alle Faktoren im Modell qualitativ sind. Sind die Faktoren quantitativ oder haben die qualitativen Faktoren sehr viele Stufen, streuen die Daten stärker im gesamten Bereich der geschätzten Werte (Abb. 3 oben rechts). Das heißt, dass die grundsätzliche Form (gruppiert oder streuend) von den Daten abhängt und Varianzhomogenität und Normalverteilung unabhängig davon erzielt sein müssen. Die Varianzhomogenität kann numerisch mit dem Levene-Test getestet werden. Um die Normalverteilung zu prüfen, kann

numerisch der Shapiro-Wilk-Test verwendet werden. Der Wert sollte möglichst nahe bei 1 liegen und eine Überschreitungswahrscheinlichkeit von 10% aufweisen. Grafisch kann man die erhaltenen Residuen als Histogramm darstellen und mit der zugehörigen Normalverteilung vergleichen (Abb.3 unten links). Eine weitere Möglichkeit ist die Darstellung der Quantile der Residuen gegen die Quantile der zugehörigen Normalverteilung. Diese Darstellung ist unter dem Namen q-q-Plot bekannt (Abb.3 unten rechts). Ist Normalverteilung oder Varianzhomogenität nicht gegeben, kann man versuchen diese Voraussetzungen mittels Transformation zu erhalten. Nicht immer wird beides gleichzeitig zu voller Zufriedenheit zu erzielen sein. Dann ist es wichtiger eine gute Varianzhomogenität bei brauchbarer Normalverteilung zu erhalten als anders herum.

Varianzhomogenität der Residuen



Normalverteilung der Residuen

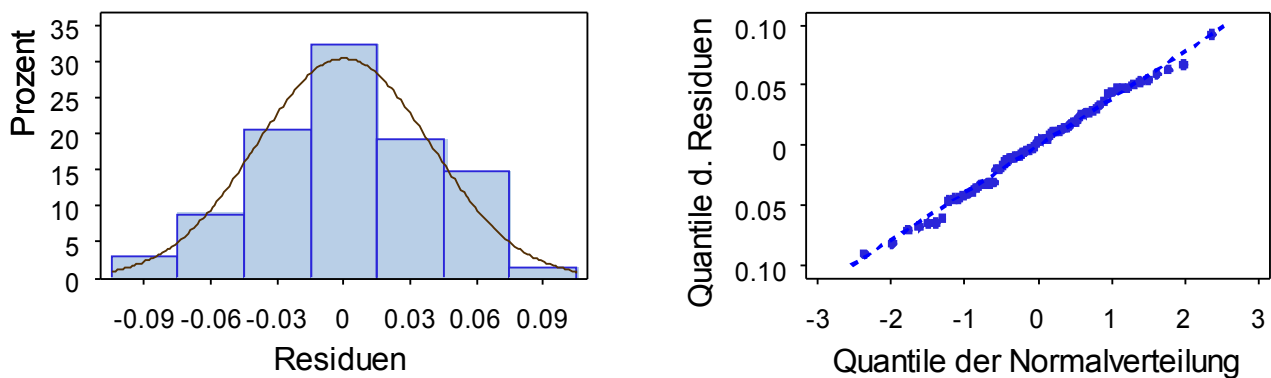


Abbildung 3: Grafische Darstellung der Varianzhomogenität bei rein qualitativen Effekten (oben links) und Varianzheterogenität bei quantitativen Effekten (oben rechts) und der Normalverteilung als Histogramm (unten links) und q-q-Plot (unten rechts).

Transformation

Es gibt sehr viele mögliche Transformationen, mit denen man versuchen kann Varianzhomogenität und Normalverteilung zu erlangen. Wurzel- oder log-Transformation kann bei schiefen Verteilungen hilfreich sein, die bei hohen Schätzwerten größere Varianz als bei kleinen aufweist. Im umgekehrten Fall helfen Potenz- oder Exponentialtransformationen. Bei Prozentzahlen sollte man eine Logit- oder Winkeltransformation verwenden. Eine sehr allgemeine Transformation ist die Box-Cox-Power-Transformation. Bei dieser erhält man über die flexible Wahl des Transformationsparameters „Phi“ eine ganze Familie von unterschiedlichen Transformationen.

Einige mögliche Transformationen sind:

	SAS-Umsetzung
logarithmische Transformation ¹	$\ln_y = \log(y);$
Wurzeltransformation	$\text{sqrt_y} = \text{sqrt}(y);$ $\text{sqrt2_y} = \text{sqrt}(y) + \text{sqrt}(y+1);$
Kehrwert / reziproker Wert	$\text{Kehr} = 1/y;$
Potenzieren	$y_2 = y^{**2};$ aber auch andere Werte sind möglich
Exponentielle Transformation	$\text{exp_y} = \exp(y);$
Logit-Transformation ^{1,2}	$\text{logit_y} = \log(y / (1-y));$
optimierte Logit-Transformation ^{2,3}	$\text{logit1_y} = \log((y + (1/(2*m))) / (1-(y+(1/(2*m)))));$
Arcus-Sinus-Winkeltransformation	$\text{Wink} = \arcsin(\text{sqrt}(y));$
optimierte Winkeltransformation ³	$\text{Wink1} = \arcsin(\text{sqrt}((z+3/8)/(m+3/4)));$
Box-Cox-Transformation	$\text{box_y} = ((y^{**\text{phi}})-1)/\text{phi};$
Folded-Exponential	$\text{folded_y} = (\exp(\text{phi}*y)-1)/\text{phi}/2 - (\exp(\text{phi}*(1-y))-1)/\text{phi}/2.$

¹ Logarithmus von Null ist nicht definiert, ² Daten müssen Werte zwischen Null und Eins haben, ³ optimiert für „gezählte“ Prozentwerte z.B. z = 5 von m = 20 waren erkrankt. Erwähnt sei hier kurz, dass bei SAS der Befehl LOG den natürlichen Logarithmus anfordert. LOG2 und LOG10 fordern den Logarithmus zur Basis 2 bzw. 10 an.

Bei der Log- und Logit-Transformation treten Probleme bei der Logarithmierung von Null auf, da der Logarithmus von Null nicht definiert ist. In einem solchen Falle kann es sinnvoll sein einen kleinen Wert (c) zu allen y-Werte dazu zu addieren:

$\ln_y = \log(y + c);$
 $\text{logit_y} = \log(y + c / (1 - y + c)).$

Zur Box-Cox und zur Folded-Exponential Transformation finden sich Makros für deren Umsetzung in SAS auf der Homepage des Fachgebiets Bioinformatik (<https://www.uni-hohenheim.de/bioinformatik/beratung/index.htm>). Diese Makros helfen den optimalen Wert von Phi für die beiden Transformationen zu finden. Hinweise zur Erstellung und Verwendung von Makros finden sich unter <http://www.urz.uni-heidelberg.de/statistik/sas-ah/04.01/makro/1-Einfuehrung.html>

Die endgültige Varianzanalyse der Daten erfolgt dann mit den transformierten Werten, für die die Voraussetzungen für eine Varianzanalyse erfüllt sind. Die Signifikanztests für die transformierten Daten sind auch für die Originaldaten valide. Zur besseren Verständlichkeit der Ergebnisse ist eine Rücktransformation der Mittelwerte möglich. Transformiert man die Mittelwerte der transformierten Daten zurück, so entsprechen die erhaltenen Werte dem Median der Originaldaten.

Zur besseren Übersicht über den Ablauf einer Varianzanalyse ist der Vorgang in einem Flussdiagramm dargestellt (Abb. 4). Einige Begriffe wie z.B. zufällige Effekte und AIC-Wert werden später eingeführt, sind hier aber der Vollständigkeit halber schon berücksichtigt.

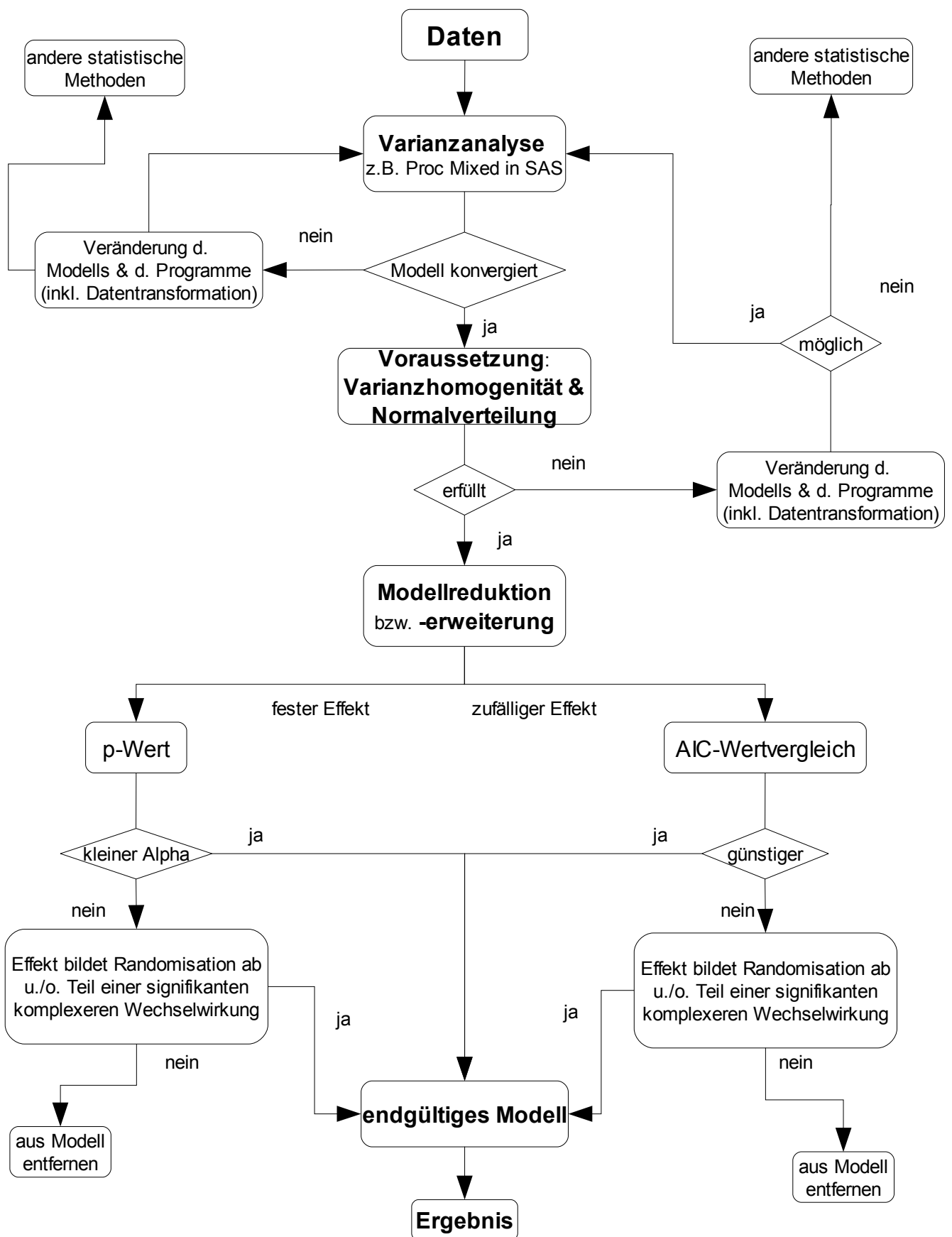


Abbildung 4: Theoretischer Ablauf einer Varianzanalyse dargestellt als Flussdiagramm

Korrelierte Daten

Korrelation beschreibt die Beziehung zwischen zwei oder mehreren Variablen. Besteht eine Korrelation zwischen zwei Variablen, ist nicht gesagt, ob eine Größe die andere kausal beeinflusst, ob beide von einer dritten Größe kausal abhängen oder ob es überhaupt ein Kausalzusammenhang gibt. Gibt es einen ungerichteten Kausalzusammenhang wird dieser in der Regel durch Korrelationsmaße, wie das Bestimmtheitsmaß (R^2) dargestellt. Ist der Zusammenhang gerichtet kommen Methoden wie Regressionsanalyse in Frage. Wichtig werden Korrelationen in Versuchen auch auf eine weitere Art: Werden mehrere Messungen zeitlicher und oder räumlicher Art auf derselben Untersuchungseinheit durchgeführt, kommt es dadurch zur Korrelation, da dieselbe Versuchseinheit Gegenstand mehrerer nicht-randomisierter Messungen (unechte Wiederholungen) ist (Abb. 5).

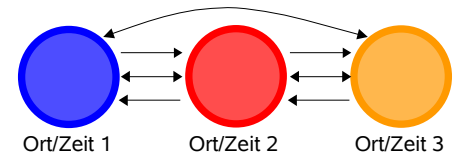


Abbildung 5: Korrelierte Daten

Diese serielle Korrelation sollte in der Varianzanalyse

berücksichtigt werden, da andernfalls die Voraussetzung der statistischen Berechnungen, die Randomisation, verletzt ist. Noch eine Form der Korrelation ist für die Anlage von Versuchen und Analysen wichtig: vermengte Daten (engl. Confounded Data; Abb. 6).

Hierbei beeinflusst eine Variable, beobachtet oder nicht beobachtet, eine oder mehr als eine erhobene beeinflussende Variable. Dadurch kommt es zu Scheinkorrelationen (der Storch bringt die Kinder) und zu Fehleinschätzungen der vermengten Variablen. Hat man im Extremfall zwei Variablen, die nahezu das gleiche erklären und jede für sich in einem varianzanalytischen Modell signifikant sind, und nimmt beide ins Modell auf, so ist keine von beiden signifikant.

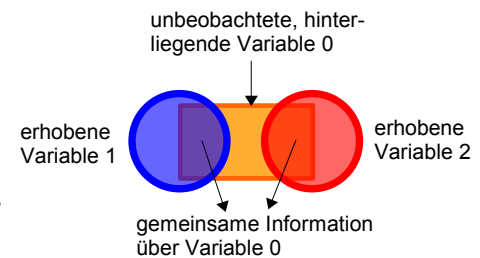


Abbildung 6: Konfundierte Daten

Ein Beispiel für eine Fehleinschätzung ist, dass Körpergröße und Gewicht von Schulkinderen einen Einfluss auf die Englischkenntnisse haben. Viel mehr ist es so, dass das beide verbindende Element „Alter“ sowohl die zunehmende Größe und Gewicht als auch die Englischkenntnisse erklärt. Als Beispiel für Multikollinearität kann die Schätzung des Stammvolumens herangezogen werden. Es zeigt sich, dass wenn man die Stammhöhe so wie den Baumumfang in 1 und 1,5 Meter Höhe nutzt beide Umfänge nicht signifikant sind, obwohl jeder für sich signifikant ist.

Regression und Kovarianzanalyse

Liegt eine „gerichtete Korrelation“ zwischen abhängiger und beeinflussender Variable vor, so kann dieser Zusammenhang in Form einer Regressionsanalyse untersucht werden. Die Regression beschreibt in Form einer Funktion die Abhängigkeit der Zielvariable (y) von einer quantitativen beeinflussenden Variable (x). Die y-Variable muss dabei einer Normalverteilung folgen, wohingegen die x-Variable im Versuch systematisch variiert werden kann. Wird die x-Variable systematisch variiert kommt eine Berechnung der Korrelation nicht mehr infrage. Die Kovarianzanalyse kombiniert Varianzanalyse von qualitativen Daten mit Regression von quantitativen Störgrößen. Die Kovarianzanalyse beantwortet die Frage, wie die Ergebnisse ausgesehen hätten, wenn der Einfluss der quantitativen Störgröße (Kovariablen) einheitlich gewesen wäre. Der Regressionsanalyse-Anteil „entfernt“ die Varianz der Kovariablen aus der abhängigen Variablen durch eine Regression der abhängigen Variablen auf die Kovariable. Die Regressionsresiduen beschreiben dann den Anteil der abhängigen Variablen, der nicht durch die Kovariable erklärt werden kann. Mit diesen Residuen wird in der Varianzanalyse weiter gearbeitet. Daraus wird klar, dass die Kovarianzanalyse nur bei quantitativen Merkmalen einsetzbar ist, und auch nur dann, wenn der Effekt der Kovariablen in allen Gruppen gleichmäßig wirkt, also keine Interaktion vorliegt (d.h. die Regres-

sionslinien parallel laufen). In Abbildung 7 sind Ergebnisse eines Fütterungsversuchs bei Schweinen dargestellt. Zwei Futtermittel (1 und 2) wurden jeweils einer Gruppe Schweine gefüttert. Aufgrund schlechter Randomisation waren alle leichten Tiere in der Gruppe, die Futter 1 erhielten, und alle schweren Tiere erhielten Futter 2. Bildet man nun die einfachen Mittelwerte je Gruppe, berücksichtigt also das Anfangsgewicht der Tiere nicht so ergibt sich das Futtermittel 2 zu einer größeren Gewichtszunahme führt als Futter 1 (Abb. 7 links). Dieses Ergebnis kann jedoch daran liegen, dass die absolute Gewichtszunahme bei schweren Tieren höher ist als bei leichten. Die relative Gewichtszunahme, also die Zunahme je Kilogramm Körpergewicht kann sich anders verhalten. Daher möchte man wissen, wie ein Tier gleichen Gewichtes sich in jeder Gruppe verhalten hätte. Man passt daher einen Effekt für die quantitative Variable „Anfangsgewicht“ (Kovariable) an. Haben die angepassten Geraden die gleiche Steigung, unterscheiden sich also nur im Achsenabschnitt, liegt eine Kovarianzanalyse vor. Dann kann für das mittlere Gewicht der Gesamtgruppe je Futter ein Wert ermittelt werden. Diese kann sich deutlich von dem der einfachen Gruppenmittelwerte unterscheiden (Abb. 7 rechts). Bei diesem konstruierten Beispiel wird extrapoliert, d.h. angenommen das die ermittelte Funktion auch für schwerere bzw. leichtere Tiere gilt. Solche Extrapolationen sind nicht verlässlich, da man keine Informationen über die Bereiche außerhalb der gewonnenen Daten hat. Sie sollten daher mit Vorsicht behandelt werden.

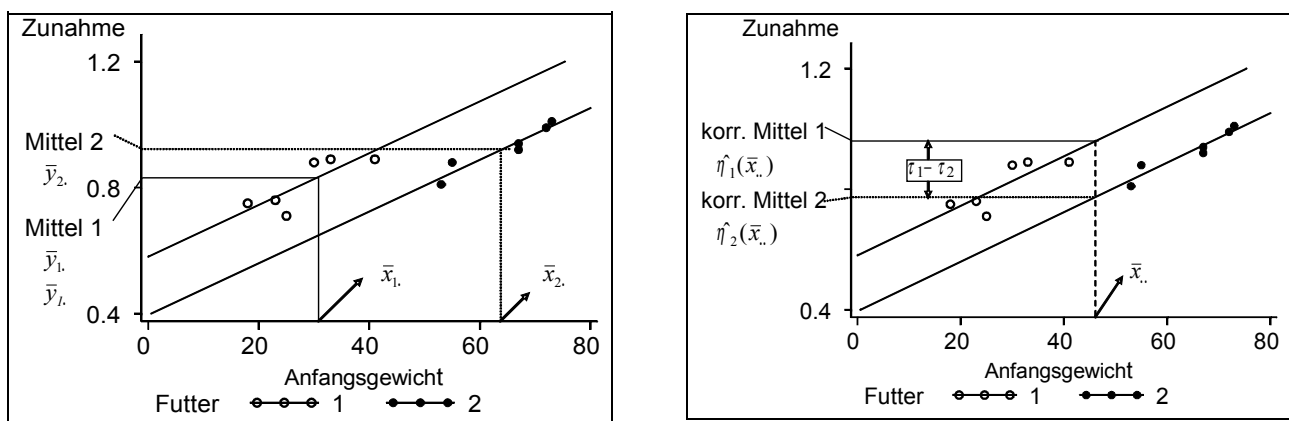


Abbildung 7: Anfangsgewicht (Kovariable) und Zunahme (beeinflusste Variable) für zwei Futtermittel; links: einfache Mittelwerte auf der y-Achse abgetragen, rechts adjustierte (korrigierte) Mittelwerte (aus Piepho 2010: Biometrie-Skript).

Signifikanz und Relevanz

Signifikanz ist nicht gleichbedeutend mit Relevanz. Ein signifikanter Unterschied kann so klein sein, dass er in der Praxis keine Bedeutung hat. Umgekehrt kann ein relevanter Unterschied, zumindest mit dem durchgeführten Versuch, nicht nachweisbar sein. Dies ist leicht möglich, wenn die Grenzdifferenz, also die mit dem Versuch erzielte kleinstmögliche Differenz, größer ist als der relevante Unterschied. Oder die Power des Tests ist nicht stark genug, also der Fehler 2. Art (Alternativhypothese wird fälschlicher Weise abgelehnt) tritt zu häufig auf.

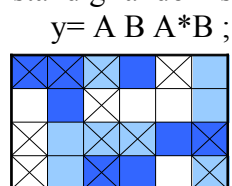
Prinzipien der Versuchsplanung: Randomisation, Wiederholung und Blockbildung

Für die Planung von Versuchen gilt, dass Randomisation, Wiederholung und meist auch Blockbildung nötig ist. Randomisieren bedeutet die Prüfglieder zufällig den Versuchseinheiten (Parzellen, Reagenzgläser, Geräte, etc.) zuzuordnen. Dadurch versucht man den Einfluss von Störgrößen

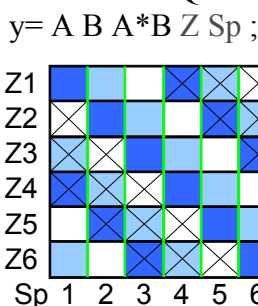
zufällig auf alle Prüfglieder zu verteilen, so dass kein Faktor durch einen unbekannten Einfluss generell bevorzugt oder benachteiligt ist. Die Wiederholungen ermöglichen, dass eine Varianz geschätzt werden kann. Und durch die Wiederholungen gleichen sich positive und negative Einflüsse aus. Mittels Blockbildung versucht man systematische Einflüsse zu erfassen und diese dadurch vom Restfehler abtrennen zu können. Blockbildung kann bedeuten, räumlich nahe gelegenes zusammenzufassen, da sich dieses in der Regel ähnlicher ist als weiter entferntes. Es kann aber auch sinnvoll sein weiter entferntes zusammenzufassen, wenn diese Teile sich ähnlicher sind, z.B. zwei Tonlinsen in einem Schlag (stark Ton-haltiger Bereich in einem Feld) oder genetisch verwandte Tiere.

Diese Prinzipien finden ihren Ausdruck in den verschiedenen Versuchsdesigns. Fünf einfache Versuchsanlagen sollen vorgestellt werden: die vollständig randomisierte Anlage, die randomisierte vollständige Blockanlage, das lateinische Quadrat, die Spaltanlage und die Streifenanlage. Später wird neben der randomisierten vollständigen Blockanlage auch noch die randomisierte unvollständige Blockanlage vorgestellt werden. Diese gehört zu den Gitteranlagen.

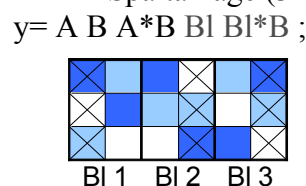
Vollständig randomisierte Anlage (6 Bl)



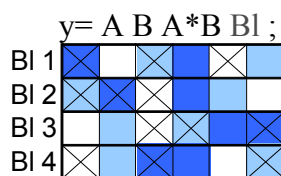
Lateinisches Quadrat (6 Bl)



Spaltanlage (3 Bl)



Randomisierte Blockanlage (4 Bl)



Streifenanlage (3 Bl)

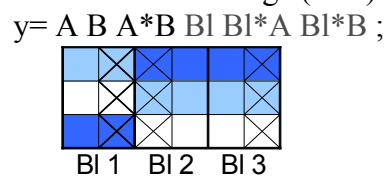


Abbildung 8: Struktur und Varianzanalysemodell (in SAS Schreibweise) von fünf einfachen Versuchsdesigns für einen Versuch mit zwei Faktoren A und B mit drei bzw. zwei Stufen je Faktor bei dem alle Faktorkombinationen in vier, sechs bzw. drei Wiederholungen (Bl) geprüft werden. Bis auf das Lateinische Quadrat sind alle abgebildeten Versuchsdesigns randomisiert.

All diese Anlagen lassen sich beliebig kombinieren und können als lineares Modell dargestellt werden. Die daraus gewonnenen Daten lassen sich mittels Varianzanalyse auswerten, wenn die Voraussetzungen der Varianzanalyse erfüllt sind. Bevor diese Anlagen besprochen werden, soll noch auf Wiederholungen und Blockbildung eingegangen werden.

Wiederholung: das Ermitteln der Varianz

Wie oben schon angesprochen, ermöglicht erst die Wiederholung die Schätzung der zufallsbedingten Varianz. Neben echten Wiederholungen, das heißt Wiederholungen, die randomisiert wurden, gibt es noch sogenannte unechte Wiederholungen. Dies sind Messwiederholungen, die auf einer Randomisationseinheit ausgeführt werden. Unechte Wiederholungen liefern keine unabhängigen Informationen sondern haben Information gemeinsam. So wird die Stickstoffversorgung verschiedener Pflanzen innerhalb einer Parzelle eines Düngerversuches ähnlicher sein, als die Stickstoffversorgung zweier Pflanzen aus zwei verschiedenen Parzellen. Blutwerte eines Tieres werden an verschiedenen Tagen ähnliche Informationen liefern, Blutwerte unterschiedlicher Tiere

werden dagegen Informationen über die Spannweite dieser Werte liefern. Bei unechten Wiederholungen arbeitet man mit deren Mittelwerten. Ausnahme davon bilden wiederholte Messungen innerhalb einer Randomisationseinheit über die Zeit oder innerhalb des Raumes (z.B. Tiefe) hier ist man an der Veränderung interessiert, kann diesen Effekt jedoch nicht randomisieren. In diesen Fällen modelliert man diesen Zusammenhang, d.h. man gibt die für diese Randomisationseinheit vermutete Korrelationsstruktur im Varianzanalysemodell mit an.

Blockbildung: das Ausschließen von Störgrößen

Blockbildung ist vom Grundgedanken her einfach zu verstehen: gleiches wird zusammengefasst. Das heißt, dass Unterschiede in der Gruppe, die durch die Zusammenfassung entsteht, klein(er) sind als der Unterschied zwischen den gebildeten Gruppen. In der Umsetzung scheint dieses Gruppieren mitunter schwerer zu sein. Die Teile einer Gruppe müssen sich nicht nebeneinander befinden, z.B. können im Feldversuch zwei feuchte Senken zusammengefasst werden oder Morgen- und Abendstunden aufgrund der Lichtintensität eine Einheit bilden. Blockung entsteht aber auch durch Personeneffekte. Bearbeiten oder bewerten mehrere Personen einen Versuch, können Unterschiede zwischen Untersuchungseinheiten an der Art der Bearbeitung bzw. der Bewertung des jeweiligen Bearbeiters liegen. Dieser Personeneffekt kann erfasst und durch geschickte Blockbildung herausgerechnet werden.

Es gibt aber auch Störgrößen z.B. Randeffekte, die mit Blockbildung nicht ausgeschlossen werden können. Als Beispiel hierfür seien Stirnrand- und Nachbarschaftseffekte im Feldversuch erwähnt. Stirnrandeffekte führen in der Sortenprüfung zur Überschätzung der Sortenleistung. Trifft dies gleichermaßen auf alle Sorten zu, handelt es sich also um einen allgemeinen Effekt, stellt dies kein Problem dar. Dieser Effekt kann durch Blockung abgefangen werden. Gibt es allerdings Unterschiede zwischen den Sorten, ist also der jeweilige Effekt vom Genotyp abhängig, verzerrt dieser Effekt die Sortenbewertung. Eine Möglichkeit dies zu verhindern ist keine oder nur schmale Wege anzulegen, eine andere die Wege zu begrünen, einen weiteren Schutzstreifen an der Stirnseite anzulegen. Nachbarschaftseffekte führen zu einer Unterschätzung der konkurrenzschwächeren Sorte. Um dies zu verhindern können Parzellenkerne beerntet werden. Hierdurch wird der Effekt am stärksten reduziert aber der Flächenaufwand steigt. Der größere Flächenbedarf kann den Bodenunterschied und die damit verbundenen Effekte vergrößern. Die Kernbeerntung selbst kann technisch sehr aufwendig und von daher schwer umsetzbar sein. Statt einer Kernbeerntung kann man die Parzelle vergrößern und durch die veränderte Relation von Rand zu Kern den Einfluss der Randeffekte auf das Gesamtergebnis verringern. Das Zusammenfassen von Prüfgliedern mit ähnlicher Nachbarschaftswirkung zu Gruppen hingegen greift wieder die Methode der Blockbildung auf.

Wiederholungsanzahl

Die Anzahl an Wiederholungen, die benötigt werden, ist Abhängig von der gewünschten Grenzdifferenz. Die Grenzdifferenz, also die kleinsten Differenz zwischen zwei Gruppen, die man nachweisen können möchte, hängt von der Varianz der untersuchten Größe ab. Als Faustzahl kann gelten, dass 30 echte Wiederholungen ein großer Probenumfang ist, alles darunter kann als kleiner Probenumfang angesehen werden.

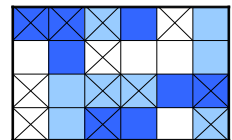
Versuchsdesigns mit vollständigen Blöcken

Mehrere Aspekte haben einen Einfluss auf die Wahl der Versuchsanlage. Dies sind zum einen bekannte oder vermutete Umwelteinflüsse, zum anderen sind es technische Rahmenbedingungen, wie Arbeitsbreite oder Wurfgröße, aber auch die Zahl der Prüfglieder in Relation zur Zahl der Versuchseinheiten. Umwelteinflüsse führen häufig zur Wahl von Blockanlagen sowie lateinischen

Quadraten und lateinischen Rechtecken (unvollständige lat. Quadrate). An dieser Stelle sollen nun die einfachen Versuchsdesigns vorgestellt werden: die vollständig randomisierte Anlage, die randomisierte vollständige Blockanlage, das lateinische Quadrat, die Spaltanlage und die Streifenanlage. Für die letzten vier Designs können die Begriffe Block und Wiederholung synonym verwendet werden, da jeder Block einer vollständigen Wiederholung entspricht.

Zu jedem Design wird im folgenden das zugehörige Modell, die Umsetzung in SAS mit nur festen Effekten (in MIXED und GLM) und die Varianzanalysetabelle gegeben. Die Versuchsdesigns können kombiniert werden. Darüber hinaus gibt es weitere Möglichkeiten, Versuche anzulegen, die mit unvollständigen Blöcken agieren. Von unvollständigen Blöcken spricht man dann, wenn nicht alle Prüfglieder bzw. alle untersuchten Faktorkombinationen in jedem Block geprüft werden. Solche Versuchsdesigns werden später vorgestellt.

Vollständig randomisierte Anlage (engl: Completely Randomized Design = CRD)



Vorteil der vollständig randomisierte Anlage ist, dass die maximale Zahl an Freiheitsgraden genutzt werden kann. Voraussetzung ist, dass die Position des Prüfglieds im Versuch keinen Einfluss hat.

Mit p Ausprägungen des Faktor Alpha (a ; z.B. 3 Sorte oder Rassen), q Ausprägungen des Faktor Delta (d ; z.B. 2 Dünger oder Futter) und $N=p*q*r$ Prüfgliedern lautet das Modell:

$$y_{ik} = \mu + a_i + d_k + ad_{ik} + e_{ik}$$

SAS-Code für ein Modell lautet:

```
Proc mixed data=...;
class A D ;
model y= A D A*D /s outp=pred2;
lsmeans A D A*D ; run;

Proc glm data=...;
class A D ;
model y= A D A*D /ss3;
means A D A*D / t;run;
```

Die Varianzanalysetabelle setzt sich wie folgt zusammen:

Effekt	DF	SQ	MQ	F _{Vers}
Faktor A	$p-1$	SQ_A	$MQ_A = SQ_A / (p-1)$	MQ_A / MQ_{Fehler}
Faktor D	$q-1$	SQ_D	$MQ_D = SQ_D / (q-1)$	MQ_D / MQ_{Fehler}
AxD	$(p-1)*(q-1)$	SQ_{AxD}	$MQ_{AxD} = SQ_{AxD} / ((p-1)(q-1))$	MQ_{AxD} / MQ_{Fehler}
Fehler	$q*(p-1)*(r-1)$	SQ_{Fehler}	$MQ_{Fehler} = SQ_{Fehler} / (q(p-1)(r-1))$	

Randomisierte vollständige Blockanlage (engl: Randomized Complete Block Design = RCBD)

Vorteil der vollständigen Blockanlage ist, dass Unterschiede zwischen den Blöcken vom Versuchsfehler abgetrennt werden können. Nur noch die Unterschiede innerhalb der Blöcke gehen in den Versuchsfehler ein. Daher sollten Blöcke nicht zu groß werden, um möglichst einheitlich zu sein. Die Gruppierung in Blöcke koste jedoch $(r-1)$ Freiheitsgrade.

Bl 1	
Bl 2	
Bl 3	
Bl 4	

Mit p Ausprägungen des Faktor Alpha (z.B. 3 Sorte), q Ausprägungen des Faktor Delta (z.B. 2 Dünger), r Blöcken (Beta) und $N=p*q*r$ Prüfgliedern lautet das Modell:

$$y_{ijk} = \mu + a_i + b_j + d_k + ad_{ik} + e_{ijk}$$

Der SAS-Code für ein Modell lautet:

```
Proc mixed data=...;
class A D Bl;
model y= A D A*D Bl /s outp=pred2;
lsmeans A D A*D Bl ; run;

Proc glm data=...;
class A D R Bl;
model y= A D A*D Bl /ss3;
means A D A*D Bl / t;run;
```

Die Varianzanalysetabelle setzt sich wie folgt zusammen:

Quelle	DF	SQ	MQ	F _{Vers}
Block	$r-1$	SQ_{Bl}	$MQ_{Bl} = SQ_{Bl} / (r-1)$	MQ_{Bl} / MQ_{Fehler}
Faktor A	$p-1$	SQ_A	$MQ_A = SQ_A / (p-1)$	MQ_A / MQ_{Fehler}
Faktor D	$q-1$	SQ_D	$MQ_D = SQ_D / (q-1)$	MQ_D / MQ_{Fehler}
AxD	$(p-1)*(q-1)$	SQ_{AxD}	$MQ_{AxD} = SQ_{AxD} / ((p-1)(q-1))$	MQ_{AxD} / MQ_{Fehler}
Fehler	$q*(p-1)*(r-1)$	SQ_{Fehler}	$MQ_{Fehler} = SQ_{Fehler} / (q*(p-1)*(r-1))$	

Lateinisches Quadrat (engl: Latin Square Design = LSD)

Vorteil des Lateinisches Quadrats ist, dass zwei orthogonale (voneinander unabhängige) Effekte, z.B. Bodeneffekte die etwa in einem 90° Winkel zueinander einen Einfluss auf den Versuch haben, berücksichtigt werden können. Diese Effekte, die in Zeilen (Z) und Spalten (Sp) erfasst werden können, gehen nicht mehr in den Versuchsfehler ein. Hierdurch gehen aber $(Z-1)+(Sp-1)$ Freiheitsgrade verloren. Da die Anzahl der Prüfglieder hier gleich der Anzahl der Wiederholungen ist eignet sich das Design meist nur für Versuche mit geringer Prüfgliedzahl.

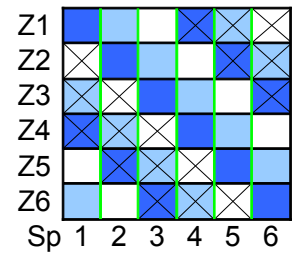
Mit p Ausprägungen des Faktor Alpha (z.B. 3 Sorten), q Ausprägungen des Faktor Delta (z.B. 2 Dünger), r Blöcken (Beta) und $N=p*q*r$ Prüfgliedern lautet das Modell:

$$y_{ijkl} = \mu + Z_j + Sp_i + a_i + d_k + ad_{ik} + e_{ijkl}$$

Der SAS-Code für ein Modell lautet:

```
Proc mixed data=...;
class A D R Sp;
model y= A D A*D R Sp /s outp=pred2;
lsmeans A D A*D R Sp;
run;

Proc glm data=...;
class A D R Sp;
model y= A D A*D R Sp /ss3;
means A D A*D R Sp / t;
run;
```



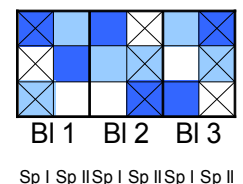
Die Varianzanalysetabelle setzt sich wie folgt zusammen:

Effekt	DF	SQ	MQ	F _{Vers}
Zeile	(p*q)-1	SQ_Z	$MQ_Z = SQ_Z / (r-1)$	MQ_Z / MQ_{Fehler}
Spalte	(p*q)-1	SQ_{Sp}	$MQ_{Sp} = SQ_{Sp} / (r-1)$	MQ_{Sp} / MQ_{Fehler}
Faktor A	p-1	SQ_A	$MQ_A = SQ_A / (p-1)$	MQ_A / MQ_{Fehler}
Faktor D	q-1	SQ_D	$MQ_D = SQ_D / (q-1)$	MQ_D / MQ_{Fehler}
AxD	(p-1)*(q-1)	SQ_{AxD}	$MQ_{AxD} = SQ_{AxD} / ((p-1)(q-1))$	MQ_{AxD} / MQ_{Fehler}
Fehler	n-1-3*(p*q-1)	SQ_{Fehler}	$MQ_{Fehler} = SQ_{Fehler} / (n-1-3*(p*q-1))$	

Faktor A und D können auch als Faktor C betrachtet werden, der (p*q) Stufen hat und mithin (p*q)-1 Freiheitsgrade

Spaltanlage (engl: Split plot)

Vorteil der Spaltanlage ist, dass innerhalb einer Wiederholung die Stufen eines Effekts zu einem Großteilstück (Gt) zusammengefasst werden können. Innerhalb dieser Großteilstücke werden dann in die Kleinteilstücke (Kt) die Stufen der weiteren Effekte angelegt. Das führt zu technischer Erleichterung, da jedes Gt einheitlich bearbeitet wird. Nachteil ist die Reduktion der Freiheitsgrade, was zu einem größeren Restfehler führt.



Mit p Ausprägungen des Faktor Alpha (z.B. 3 Sorten), q Ausprägungen des Faktor Delta (z.B. 2 Dünger), r Blöcken (Beta) und $N=p*q*r$ Prüfgliedern lautet das Modell:

$$y_{ijk} = \mu + b_j + bd_{jk} + a_i + d_k + ad_{ik} + e_{ijk}$$

$bd_{jk} = BI * D = BI * Spalte = \text{Großteilstück (Gt) innerhalb einer Wiederholung}$

Der SAS-Code für ein Modell lautet:

```
Proc mixed data=...;
class A D Bl;
model y= A D A*D Bl Bl*D/s outp=pred2;
lsmeans A D A*D Bl Bl*D; run;
```

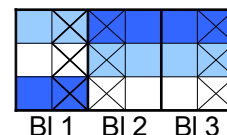
```
Proc glm data=...;
class A D R Bl;
model y= A D A*D Bl Bl*D/ss3;
means A D A*D Bl Bl*D / t;run;
```

Die Varianzanalysetabelle setzt sich wie folgt zusammen:

Quelle	DF	SQ	MQ	F _{Vers}
Block	r-1	SQ_{Bl}	$MQ_{Bl} = SQ_{Bl} / (r-1)$	MQ_{Bl} / MQ_{Fehler}
Faktor A	p-1	SQ_A	$MQ_A = SQ_A / (p-1)$	MQ_A / MQ_{Fehler}
Faktor D	q-1	SQ_D	$MQ_D = SQ_D / (q-1)$	MQ_D / MQ_{Gt}
Gt-Fehler	$(q-1)*(r-1)$	SQ_{Gt}	$MQ_{Gt} = SQ_{Gt} / ((q-1)*(r-1))$	
AxD	$(p-1)*(q-1)$	SQ_{AxD}	$MQ_{AxD} = SQ_{AxD} / ((p-1)*(q-1))$	MQ_{AxD} / MQ_{Fehler}
Fehler	$q*(p-1)*(r-1)$	SQ_{Fehler}	$MQ_{Fehler} = SQ_{Fehler} / (q*(p-1)*(r-1))$	

Streifenanlage (engl: Strip plot)

Vorteil der Streifenanlage ist, dass innerhalb einer Wiederholung je Stufe des einen Faktors zusammenhängend Bearbeitet werden kann. Nachteil ist die Reduktion der Freiheitsgrade, was zu einem größeren Restfehler führt.



Mit p Ausprägungen des Faktor Alpha (z.B. 3 Sorten), q Ausprägungen des Faktor Delta (z.B. 2 Dünger), r Blöcken (Beta) und $N=p*q*r$ Prüfgliedern lautet das Modell:

$$y_{ijk} = \mu + b_j + bd_{jk} + ba_{ij} + a_i + d_k + ad_{ik} + e_{ijk}$$

$bd_{jk} = Bl * D = Bl * Spalte = Spalte$ innerhalb einer Wiederholung und
 $ba_{jk} = Bl * A = Bl * Zeile = Zeile$ innerhalb einer Wiederholung

Der SAS-Code für ein Modell lautet:

```
Proc mixed data=...;
class A D Bl;
model y= A D A*D Bl Bl*A Bl*D/s outp=pred2;
lsmeans A D A*D Bl Bl*A Bl*D; run;
```

```
Proc glm data=...;
class A D R Bl;
model y= A D A*D Bl Bl*A Bl*D/ss3;
means A D A*D Bl Bl*A Bl*D / t;run;
```

Die Varianzanalysetabelle setzt sich wie folgt zusammen:

Quelle	FG	SQ	MQ	F _{Vers}
Block	r-1	SQ_{Bl}	$MQ_{Bl} = SQ_{Bl} / (r-1)$	MQ_{Bl} / MQ_{Fehler}
Faktor A	p-1	SQ_A	$MQ_A = SQ_A / (p-1)$	$MQ_A / MQ_{A \times Bl}$
Fehler A	$(p-1) \cdot (r-1)$	$SQ_{A \times Bl}$	$MQ_{A \times Bl} = SQ_{A \times Bl} / ((p-1)(r-1))$	
Faktor B	q-1	SQ_B	$MQ_B = SQ_B / (q-1)$	$MQ_B / MQ_{B \times Bl}$
Fehler B	$(q-1) \cdot (r-1)$	$SQ_{B \times Bl}$	$MQ_{B \times Bl} = SQ_{B \times Bl} / ((q-1)(r-1))$	
AxB	$(p-1) \cdot (q-1)$	$SQ_{A \times B}$	$MQ_{A \times B} = SQ_{A \times B} / ((p-1)(q-1))$	$MQ_{A \times B} / MQ_{Fehler}$
Fehler	$(p-1) \cdot (q-1) \cdot (r-1)$	SQ_{Fehler}	$MQ_{Fehler} = SQ_{Fehler} / ((p-1)(q-1)(r-1))$	

Unvollständige Blöcke

Nicht immer ist es bei Versuchen möglich alle Varianten in jedem Block zu testen, sei es, weil die Zahl der Varianten sehr groß ist, z.B. Zahl der Genotypen in der Pflanzenzüchtung oder weil die Zahl der Untersuchungseinheiten innerhalb eines Blocks stark eingeschränkt ist, z.B. bei fistulierten Tieren oder Geschwistern eines Wurfs. Für diese Problematik sind verschiedene Lösungen entwickelt worden. So gibt es verschiedene Varianten, unvollständige Blöcke zu vollständigen Wiederholungen zusammenzufassen. Es gibt aber auch unvollständige Wiederholungen. Daneben können unvollständige Blockanlagen balanciert und teilbalanciert sein. Balanciert sind unvollständige Blockanlagen dann, wenn jedes Prüfglied mit jedem anderen Prüfglied gleich häufig geprüft wird. Das heißt, dass alle Prüfgliedpaare gleich häufig vorkommen. Von teilbalancierten unvollständigen Blockanlagen spricht man dann, wenn nicht alle Prüfgliedpaare gleichhäufig vorkommen. Je mehr unterschiedliche Werte für gemeinsames Auftreten vorkommen (z.B. manche Paare gar nicht andere 1, 2 oder 3 mal), um so schwieriger wird die Auswertung, da daraus unterschiedliche Standardfehler entstehen und mithin mehrere Grenzdifferenzen (LSD) vorliegen.

Das Alphadesign ist unter den Versuchsanlagen mit unvollständigen Blöcken in vollständigen Wiederholungen, die Form mit der höchsten Flexibilität. Gitteranlagen stellen bestimmte Anforderungen an die Zahl der Prüfglieder und Anzahl der Blöcke, so dass es nur bestimmte Formen gibt. Der Begriff Zwei- bzw. Dreisatz Gitter bedeutet, dass der Versuch aus zwei bzw. drei vollständigen wiederholten Gittern besteht. Weitere Information dazu findet sich unter anderem in Cochran und Cox (1992).

Das Einbinden von wiederholten Standards ist eine Möglichkeit unwiederholte Prüfglieder schätzen zu können. So entstehen unvollständige Blöcke, diese sind über die Wiederholten Standards miteinander verbunden. Ein Beispiel dafür sind so genannte Augmented Designs. Hier werden bestimmte Varianten (Standards) in alle Blöcke geprüft und die einzelnen Blöcke um unwiederholte Varianten (Prüfglieder) erweitert. Die Standards, die in allen Blöcken geprüft werden, werden dabei entsprechend eines Versuchsdesigns (z.B. eines der fünf oben gezeigten) randomisiert. Die Prüfglieder werden dann zufällig auf die freien Positionen im Versuchsdesign verteilt.

1	2	3	4	5	6	7	8	9
10	Std 3	11	12	Std 1	13	14	Std 2	15
16	17	18	19	20	21	22	23	24
25	26	27	28	29	30	31	32	33
34	Std 2	35	36	Std 1	37	38	Std 3	39
40	41	42	43	44	45	46	47	48
49	50	51	52	53	54	55	56	57
58	Std 1	59	60	Std 3	61	62	Std 2	63
64	65	66	67	68	69	70	71	72

Abbildung 9: Augmented Design mit drei Standards (Std 1-3) in vollständiger Blockanlage und 72 Prüfgliedern

Unvollständige Blöcke ergeben sich auch dann, wenn es bei einer Versuchsanlage mit vollständigen Blöcken zu Datenverlust kommt. Sei es, dass einzelne Werte nicht erhoben werden können, weil Parzellen zerstört, Reagenzgläser runtergefallen oder Tiere erkrankt sind oder beim Datentransfer einzelne Werte verloren gegangen sind. Gerade in den letztgenannten Fällen ergibt sich intuitiv die Nutzung adjustierter Mittelwerte. Ein theoretisches Beispiel soll die Nutzung adjustierter Mittelwerte erläutern. Drei Behandlungen wurden in drei Wiederholungen (drei vollständige Blöcke) untersucht und die folgenden Ergebnisse erhalten:

Block	Behandlung		
	1	2	3
1	1	2	3
2	4	5	6
3	7	8	9
Mittel:	4	5	6

Beim Datentransfer ging der Wert der dritten Behandlung im dritten Block verloren:

Block	Behandlung		
	1	2	3
1	1	2	3
2	4	5	6
3	7	8	.
	4	5	?

Würde man nun für Behandlung 3 den einfachen Mittelwert aus zwei Werten bilden, würde Behandlung 3 mit einem Mittelwert von 4,5 zwischen Behandlung 1 und 2 liegen, da das Ergebnis aus Block 3 mit den höchsten Werten fehlt. Verglichen mit dem Ergebnis des Gesamtdatensatzes ist dies eine Verfälschung des Ergebnisses.

Es liegt nahe die Unterschiede von Behandlung 3 zu Behandlung 1 und 2 innerhalb der Blöcke 1 und 2 zu ermitteln und diese auf den dritten Block anzuwenden:

	Block 1	Block 2
Behandlung 3 – 1	2	2
Behandlung 3 – 2	1	1

In beiden Blöcken ist Behandlung 3 um 2 Einheiten größer als Behandlung 1 bzw. um 1 Einheit größer als Behandlung 2. Wendet man diese Information auf Block 3 an so erhält man in beiden Fällen 9 als Wert und damit einen adjustierten Mittelwert von 6. Die Nutzung der Intra-Block-Information hat die Schätzung der Behandlung also deutlich verbessert. Im Normalfall werden sich die Differenzen von zwei Behandlungen in verschiedenen Blöcken unterscheiden und man arbeitet mit Mittelwerten dieser Differenzen.

Bei unvollständigen Blöcken gibt es neben der Möglichkeit der direkten Intra-Block-Information auch noch die indirekten Intra-Block-Information. Nimmt man Block 1 und 2 aus dem Beispiel von oben und löscht eine weitere Beobachtung, z.B. Behandlung 2 in Block 2, so erhält man zwei unvollständige Blöcke:

Block	Behandlung		
	1	2	3
2	4	.	6
3	7	8	.

Wie vorher ist man an den Vergleichen der Behandlungen interessiert: 1 vs. 2, 1 vs. 3 und 2 vs. 3. Die ersten beiden Vergleiche beruhen auf dem direkten Vergleich aus einem Block. Block 3: Beh1 – Beh2 = 7-8=-1. Block 2: Beh1 – Beh3 = 4-6=-2. Der Vergleich Behandlung 2 mit Behandlung 3 ist indirekt möglich:

$$\begin{aligned} \text{Beh1} - \text{Beh3 (Block 2)} - (\text{Beh1} - \text{Beh2 (Block3)}) &= \\ \text{Beh1} - \text{Beh3} - \text{Beh1} + \text{Beh2} &= \\ \text{Beh2} - \text{Beh3} &= -2 - (-1) = -1 \end{aligned}$$

Es gibt noch eine weitere Möglichkeit die Information der Blöcke zu nutzen, um Informationen über den Vergleich Behandlung 2 und 3 zu bekommen: Man kann die Differenz der Summen der Behandlungen je Block miteinander vergleichen:

$$\begin{aligned} \text{Beh1} + \text{Beh3 (Block 2)} &= 4 + 6 = 10 \\ \text{Beh1} + \text{Beh2 (Block 3)} &= 7 + 8 = 15 \\ \text{Beh1} + \text{Beh3} - (\text{Beh1} + \text{Beh2}) &= \text{Beh3} - \text{Beh2} = 10 - 15 = -5 \end{aligned}$$

Diese Information wird Inter-Block-Information genannt. Sie ist nicht so genau wie die Intra-Blockinformation, da sie neben der Restvarianz auch die Blockvarianz enthält.

Feste und zufällige Effekte

Bei der einfachen Varianzanalyse erhält man für jeden Effekt, den man prüft, Schätzwerte (feste Effekte) und für den Restfehler eine Varianz (zufälliger Effekt). Im Zusammenhang mit komplexeren Versuchsanlagen, in denen es mehr als einen Fehlerterm gibt, wie bei Spalt- und Streifenanlage, sowie bei unvollständigen Blöcken, werden zufällige Effekte interessant. Bei zufälligen Effekten liegt das Interesse auf der, durch den Effekt hervorgerufenen, Varianz. Hinzu kommt, dass erst bei zufälligen Effekten die Inter-Blockinformation genutzt werden kann. Ob ein Effekt als fester oder als zufälliger Effekt genutzt werden sollte, ist von mehreren Aspekten abhängig, die in der folgenden Tabelle gegeneinander gestellt werden

Fester Effekt (fix)

- Es werden wenig Faktorstufen geprüft (<10, z.B. 4 Sorten, 3 Jahre).
- Die Faktorstufen sind nicht repräsentativ für den Faktor.
- Man ist nur an geprüften Faktorstufen interessiert bzw. die geprüften Stufen stehen im Vordergrund.
- Modellvergleich erfolgt über p-Werte.

Zufällige Effekte (random)

- Es werden viele Faktorstufen einer (theoretisch) unendlich großen Grundgesamtheit geprüft (z.B. Orte, Jahre).
- Die Faktorstufen sind repräsentativ für den Faktor (Zufallsstichprobe aus Grundgesamtheit).
- Man ist an Faktor allgemein (Grundgesamtheit) interessiert (z.B. Schätzung der Varianz der zufälligen Effekte).
- Wird ein fester Effekt mit einem zufälligen gekreuzt (Interaktion) so wird diese Interaktion ein zufälliger Effekt.
- Der Modellvergleich erfolgt über den AIC-Wert (Akaike's Informations-kriterium) damit können nur Modelle, die dieselben festen Effekte besitzen, miteinander verglichen werden.
- Randomisationseinheiten.

Hilfreiche Fragen, um zu entscheiden, ob ein Effekt als zufälliger oder fester Effekt ins Modell aufgenommen werden soll, sind:

- Sind die betrachteten Faktorstufen repräsentativ für den Faktor?

- Hat der Faktor viele Stufen und betrachtet man viele Stufen des Faktors?
- Interessiert der Faktor oder nur genau die geprüften Faktorstufen?

Am Beispiel Block soll die Entscheidung fester oder zufälliger Effekt verdeutlicht werden. Nimmt man Blöcke fix nutzt man die Intra-Block-Information, nimmt man Blöcke zufällig kann man die Intra- und die Inter-Block-Information nutzen. In der Regel ist man nicht an Schätzwerten für die Blöcke interessiert sondern hat Blöcke angelegt, um eine Störgröße auszuschließen. Das spricht dafür, die Blöcke zufällig zu nehmen. Was dagegen sprechen kann, ist die geringe Zahl der Blöcke. Als Faustzahl kann hier die Zahl 10 gelten. Häufig hat man nur drei oder vier, vor allem, wenn man mit vollständigen Blöcken arbeitet, und wird die Blöcke daher eher als festen Effekt betrachten. Dazu kommt, dass bei vollständigen Blöcken die Inter-Blockinformation nicht genutzt werden kann. Von daher gibt es im Fall vollständiger Blöcke keinen Unterschied, wenn man diesen Effekt als festen oder als zufälligen Effekt ins Modell auf nimmt. Wenn ausreichend unvollständige Blöcke zur Schätzung der Varianz der Blockeffekte zur Verfügung stehen, dann ist es jedoch besser diese als zufälligen Effekt ins Modell aufzunehmen.

Allgemein gilt, dass Randomisationseinheiten (Block, Zeile, Spalte etc.), die nicht alle Prüfglieder enthalten, als zufällige Effekte ins Modell aufgenommen werden sollten. Randomisationseinheiten, die alle Prüfglieder enthalten können fixer oder zufälliger Effekt im Modell sein, denn für die Schätzung von Prüfglieddifferezenzen ist es in diesem Falle gleichgültig ob der Effekt fix oder zufällig betrachtet wird. Für Prüfgliedmittelwerte und Mittelwertdifferezenzen im Falle unvollständiger Randomisationseinheiten ist dies jedoch NICHT egal, da die Inter-Block-Information genutzt werden kann. In diesem Fall ist die Konsequenz: Veränderte Adjustierung der Prüfgliedmittelwerte. Bei ausreichend vielen Wiederholungen des Effekts werden in der Regel die Standardfehler (bzw. der mittlere Standardfehler) kleiner.

Fehlende Werte

Wie oben schon erwähnt kommt es immer wieder zu fehlenden Werten. Diese können wie bei unvollständigen Blöcken geplant sein oder im Zuge der Versuchsdurchführung entstehen. Für eine valide statistische Auswertung ist dabei wichtig, dass die Werte zufällig fehlen und nicht aufgrund eines bestimmten Prozesses. Liegt dem Fehlen ein bestimmter Prozess zugrunde, beinhaltet das Fehlen eine Information. Dies ist zum Beispiel bei einer mehrstufigen Selektion der Fall. Eine Auswertung der Daten ist dann mit Vorsicht zu betrachten. Näheres dazu findet sich zum Beispiel in Piepho & Möhring (2006).

Modellbildung

Mit Hilfe der Grundstrukturen, die bis jetzt vorgestellt wurden, kann das Modell zur Analyse des Versuchs hergeleitet werden. Dabei gilt: Analysed as randomized. Jeder Randomisationsschritt muß abgebildet werden. Dabei gilt Ockhams Sparsamkeitsprinzip. Das Modell sollte so einfach als möglich sein, d.h. aus möglichst wenig Effekten bestehen, bei Regressionen sollten die Polynome möglichst niedrig sein, d.h. eher x^2 als x^4 . Um ein passendes Modell zu finden, gibt es keine festen Weg, aber Regeln. Dabei ist es hilfreich zu wissen, dass es das richtige Modell nicht gibt. In der Regel gibt es eine Gruppe angemessener Modelle.

Zum Ablauf der Modellbildung

Als erstes sammelt man die Effekte des Modells. Hierbei erweist es sich als hilfreich, die Behand-

lungsanteile und die Anteile des Versuchsdesign (Randomisationseinheiten) getrennt zu betrachten. Dies bedeutet, separate Bezeichnungen für die Behandlungen und die Randomisationseinheiten zu vergeben. Erst zum Schluss kann man die Bezeichnungen der Randomisationseinheiten durch die möglicherweise zugehörigen Behandlungsbezeichnungen (z.B. bei Spalt- und Streifenanlagen) ersetzen.

Die einzelnen Schritte der Modellbildung sind:

- Behandlungsanteil aufstellen.
- Designanteil aufstellen, mit eigenen Bezeichnungen für jede Randomisationseinheit.
- Nicht randomisierbare Faktoren des Designs notieren, diese sind häufig Messwiederholungen in Zeit oder Raum.
- Prüfen, ob es Interaktionen zwischen Faktoren des Designs und Effekten des Behandlungsanteils gibt und ob diese geprüft werden können. In der Regel gibt es keine Interaktionen. Ausnahmen sind zum Beispiel der Einfluss von Jahren oder Orten auf Behandlungen (Sorten/Tiere).
- Entscheiden welche Effekte als zufällige Effekte ins Modell eingehen sollen.
- Bei den Effekten, die nicht-randomisierte Faktoren (Messwiederholungen) enthalten, überlegen, ob und welche Korrelationsstruktur genutzt werden soll.

Das so erhaltene Modell kann nun in einem Statistikprogramm, wie z.B. SAS umgesetzt werden. Um nun zu entscheiden, ob das Modell angemessen ist, muss erst geprüft werden, ob die Voraussetzung für eine Varianzanalyse, Varianzhomogenität und Normalverteilung, erfüllt sind. Andernfalls muss durch Transformation der abhängigen Variable oder durch Modellmodifikation so lange probiert werden bis diese erreicht sind oder zu anderen Analyseverfahren gegriffen werden. Hat man die Voraussetzungen erfüllt können die festen Effekte entsprechend ihres p-Wertes ausgewählt werden. Die zufälligen Effekte können bei gleichen Effekten im festen Teil des Modells (identischen festen Effekten) z.B. über den Wert des Aikaike-Informationskriteriums (AIC) verglichen werden. Der AIC-Wert setzt sich zusammen aus der Loglikelihood so wie einem Strafterm für die Zahl der verwendeten zufälligen Effekte. Hierdurch wird dem Sparsamkeitsprinzip Rechnung getragen: Das gewählte Modell soll so einfach als möglichst sein. Detailliertere Informationen hierzu finden sich in Piepho et al. (2003) und Piepho et al. (2004).

Literatur

- Cochran, W.G., and Cox, G.M.(1992) Experimental Designs. Wiley, 2 edition, 640 p.
- Piepho, H.P., Büchse, A., and Emrich, K. (2003): A hitchhiker's guide to the mixed model analysis of randomized experiments. Journal of Agronomy and Crop Science 189, 310-322
- Piepho, H.P., Büchse, A., and Richter, C. (2004): A mixed modelling approach to randomized experiments with repeated measures. Journal of Agronomy and Crop Science 190, 230-247
- Piepho, H.P., and J. Möhring. (2006). Selection in cultivar trials - Is It Ignorable? Crop Science 46, 192-201.

Mein Dank gilt Prof. Dr. H.-P. Piepho auf dessen Skripten dieses Skript beruht und dessen wertvolle Kritik zusammen mit den Vorschlägen und Korrekturen meiner Kolleginnen Frau Dr. Kuzyakowa und Frau Kleinknecht dieses Skript seine Form erhielt.