

Statistik 2 - Musterlösung 9

Markus Kalisch

Thursday, October 16, 2014

Aufgabenstellung

Ich bitte im Folgenden meine landwirtschaftliche Ignoranz zu entschuldigen! Der Kontext ist extrem vereinfacht und soll lediglich helfen, sich die Situation möglichst einfach vorstellen zu können.

In dieser Serie machen wir ein Gedankenexperiment. Stellen Sie sich vor, Sie haben in Ihrem neuen Job einen Beratungsauftrag erhalten: Bauer Meier hat zwei Sorten Rüben (A, B) und zwei Sorten Dünger (1, 2). Gesät, gedüngt und geerntet wird immer zu einer vorgegebenen Zeit. Bauer Meier fragt sich nun, welche Kombination von Rüben und Dünger wohl den grössten Ertrag bringt.

Helfen Sie Bauer Meier diese Frage zu beantworten. Sie haben dazu 4 Ackerstücke (Parzellen zu je 4m x 4m) an unterschiedlichen Standorten zur Verfügung. Jede Pflanze benötigt 1m x 1m Fläche (Mutantenrüebli nahe AKW ;).

- 1) Beschreiben Sie ein Experiment, das die Frage von Bauer Meier möglichst gut beantworten kann. Beschreiben Sie so viele Details, dass Bauer Meier das Experiment komplett durchführen könnte (evtl. mit Tabellen). Dabei soll auch ein Skizze enthalten sein, die anzeigt, auf welchem Teil vom Ackerstück was angepflanzt und gedüngt werden soll. Z.B. könnte die Skizze so aussehen, wenn auf jedem der vier Felder in jedem Quadratmeter Pflanze A mit Dünger 1 angepflanzt wird (Acker 1 | Acker 2 | Acker 3 | Acker 4):

A1	A1	A1	A1		A1	A1	A1	A1		A1	A1	A1	A1		A1	A1	A1	A1
A1	A1	A1	A1		A1	A1	A1	A1		A1	A1	A1	A1		A1	A1	A1	A1
A1	A1	A1	A1		A1	A1	A1	A1		A1	A1	A1	A1		A1	A1	A1	A1
A1	A1	A1	A1		A1	A1	A1	A1		A1	A1	A1	A1		A1	A1	A1	A1

- 2) Schreiben Sie nun einen Kommentar zu dem Design, sodass ein Experte (z.B. kontrollierender Statistiker) überzeugt ist, dass Sie alle wichtigen Prinzipien des experimentellen Designs berücksichtigt haben. Falls in gewissen Teilen des Designs mehrere Varianten möglich sind, beschreiben Sie Vor- und Nachteile der Varianten.
- 3) Beschreiben Sie, mit welcher Methode die Daten angemessen ausgewertet werden können.

Lösung zu Teilaufgabe 1: Anleitung für Bauer Meier

Nehmen Sie 64 Pflanzensamen und von beiden Düngersorten genug Dünger um je 32 Pflanzen düngen zu können.

Wir verwenden alle 4 Ackerstücke (jeweils 4m x 4m) und teilen jedes Ackerstück in 16 kleine Sub-Felder auf (jeweils 1m x 1m). Insgesamt gibt es also 64 Felder. Auf jedes Feld wird eine Pflanze (Typ A oder Typ B) gesetzt und mit einem vorgegebenem Dünger (1 oder 2) bedüngt.

Die genaue Bepflanzung und Bedüngung je nach Sub-Feld ist in folgender Skizze schematisch gezeigt (die vier Acker sind mit vertikalen Strichen getrennt; Begründung für dieses Design steht in Lösung zu Teilaufgabe 2):

A1	A1	B1	B1		A2	A1	B2	B1		B2	B2	A2	A2		A1	B2	B2	B2
B2	A1	A2	B2		B1	B2	B2	B1		B2	B1	B1	A2		B1	A1	A2	A1
B2	B2	B1	A2		B1	A1	A1	A2		B1	A1	A2	A1		A2	B2	B1	B1
A1	A2	A2	B1		A1	A2	B2	A2		B2	B1	A1	A1		A2	B1	A2	A1

Lösung: Diskussion des Designs

Im Folgenden werden diese Punkte illustriert: Es kann eine klare, messbare Hypothese aufgestellt werden. Zudem werden alle besprochenen Prinzipien für gutes experimentelles Design berücksichtigt.

Hypothese

Es gibt signifikante Unterschiede bzgl. Ernteertrag (in kg) am Stichtag zwischen den verschiedenen Kombinationen von Pflanze und Dünger.

Experiment

Wir könnten Bauern suchen, die in der Vergangenheit gewisse Kombinationen von Pflanze und Dünger verwendet haben und dann den Ertrag pro Fläche zwischen den Kombinationen vergleichen. Dieses Vorgehen hat den Nachteil, dass der Effekt von Kombination und Bauer vermischt wird: Evtl. hat der unerfahrene Bauer Huber nur A1 und der erfahrene Bauer Schmitt nur A2 angepflanzt. Bauer Huber hat viele Fehler gemacht und deswegen ist die Ernte von A1 gering, obwohl die Kombination A1 bei richtiger Behandlung eigentlich ertragreich wäre.

Um solche Vermischungen möglichst ausschliessen zu können, machen wir daher am besten ein Experiment. In diesem Beispiel ist das auch mit vertretbarem Aufwand und ohne ethische Bedenken durchführbar.

Kontrolle

Wir haben mehrere Kombinationen, die wir miteinander vergleichen. Die Kombinationen dienen sich gegenseitig also als Kontrolle.

Replikate

Wir können insgesamt 64 Kombinationen anbauen. Am günstigsten ist ein balanciertes Design mit 16 Replikaten pro Kombination.

Zufall / Randomisierung

Es gibt 4 Felder mit jeweils 4 Spalten und 4 Zeilen. D.h., es können insgesamt 64 Pflanzen angepflanzt werden. Wir stellen nun verschiedene mögliche Verteilungen der Pflanzen (mit Dünger) vor. Dabei schreiben wir aus Platzgründen die vier Blöcke nebeneinander und trennen mit vertikalen Strichen (auch wenn die Blöcke in Wirklichkeit nicht nebeneinander liegen):

Block1|Block2|Block3|Block4

Block als Confounder

A1	A1	A1	A1		A2	A2	A2	A2		B1	B1	B1	B1		B2	B2	B2	B2
A1	A1	A1	A1		A2	A2	A2	A2		B1	B1	B1	B1		B2	B2	B2	B2
A1	A1	A1	A1		A2	A2	A2	A2		B1	B1	B1	B1		B2	B2	B2	B2
A1	A1	A1	A1		A2	A2	A2	A2		B1	B1	B1	B1		B2	B2	B2	B2

Diese Aufteilung ist sehr ungünstig. Angenommen, A2 hat einen sehr niedrigen Ertrag. Lag das daran, das A2 wirklich eine ungünstige Kombination ist oder war Feld 2 einfach ungünstig gelegen (z.B. schlechter Boden). Der Effekt von Block und Behandlung ist vermischt und daher ist keine Aussage über Behandlung mehr möglich. Das Experiment war umsonst!

Pro Spalte eine Kombination

A1	A2	B1	B2		A1	A2	B1	B2		A1	A2	B1	B2		A1	A2	B1	B2
A1	A2	B1	B2		A1	A2	B1	B2		A1	A2	B1	B2		A1	A2	B1	B2
A1	A2	B1	B2		A1	A2	B1	B2		A1	A2	B1	B2		A1	A2	B1	B2
A1	A2	B1	B2		A1	A2	B1	B2		A1	A2	B1	B2		A1	A2	B1	B2

Diese Aufteilung hat den Nachteil, dass die Kombinationen immer in den gleichen Positionen im Feld vorkommen. Folgende Probleme könnten z.B. auftreten:

- Kombinationen am Rand vom Feld könnte einen Nachteil haben, z.B. weil Schädlinge leichter an diese Pflanzen kommen können. Somit wären bei uns A1 und B2 systematisch im Nachteil gegenüber A2 und B1.
- A2 ist immer rechts neben A1. Angenommen, es gibt in allen Parzellen eine Hangneigung, sodass das Wasser und die Nährstoffe in Richtung A2 abfließen. Dann besteht die Gefahr, dass A1 wegen der fehlenden Nährstoffe schlechter gedeiht, obwohl A1 eigentlich eine gute Kombination ist.

Ein Vorteil dieser Methode ist jedoch, dass es evtl. bedeutend einfacher ist eine Spalte mit derselben Saat zu bepflanzen und mit demselben Dünger zu düngen (z.B. weil man mit dem Traktor fährt).

Strip-plot design

Eine Verbesserung wäre es, wenn man in jedem Block die Spalten permutiert. Z.B. so:

A1	A2	B1	B2		A2	B2	B1	A1		B1	A1	A2	B2		A2	A1	B2	B1
A1	A2	B1	B2		A2	B2	B1	A1		B1	A1	A2	B2		A2	A1	B2	B1
A1	A2	B1	B2		A2	B2	B1	A1		B1	A1	A2	B2		A2	A1	B2	B1
A1	A2	B1	B2		A2	B2	B1	A1		B1	A1	A2	B2		A2	A1	B2	B1

So kann man immer noch spaltenweise säen und düngen, aber es sind jetzt nicht mehr immer die gleichen Kombinationen Nachbarn bzw. am Rand. Das ist ein klarer Vorteil.

Bei der Auswertung einer solchen Anordnung gibt es (evtl. für Sie überraschend) einige technische Details zu

beachten, die wir in der VL nicht besprochen haben. In der fortgeschrittenen Literatur findet man mehr zu diesem Thema unter dem Stichwort *strip-plot design*.

Randomisierung pro Block

Bei dieser Anordnung werden pro Block alle 4 Kombinationen genau viermal völlig zufällig (aber innerhalb von einem Block) platziert. Z.B. so:

A1	A1	B1	B1		A2	A1	B2	B1		B2	B2	A2	A2		A1	B2	B2	B2
B2	A1	A2	B2		B1	B2	B2	B1		B2	B1	B1	A2		B1	A1	A2	A1
B2	B2	B1	A2		B1	A1	A1	A2		B1	A1	A2	A1		A2	B2	B1	B1
A1	A2	A2	B1		A1	A2	B2	A2		B2	B1	A1	A1		A2	B1	A2	A1

Falls die Ausbringung und Düngung für jede Pflanze einzeln gemacht werden kann (und nicht z.B. mit dem Traktor streifenweise erfolgen muss), ist diese Anordnung sehr geschickt. Wir haben pro Block jede Kombination 4 mal (also ein balanciertes Design pro Block und auch insgesamt). Es gibt keine systematischen Nachbarschafts- oder Randeffekte, weil die Position jeder Kombination randomisiert ist.

Völlige Randomisierung

In dieser Anordnung werden 16 Replikate jeder Kombination völlig zufällig über alle Blöcke verteilt. Z.B. so:

A1	A1	B2	A2		B2	A1	A1	A1		A2	B1	A1	B2		A1	B2	A2	B1
B2	A1	A1	B2		A2	B1	B2	A2		B1	A2	A1	B2		A1	A1	A1	B1
B2	B2	A2	A1		A2	B2	B2	B1		A2	B1	B1	B2		B1	B1	B1	A2
A2	B1	B1	B2		A1	A2	A1	B1		A2	B2	B1	B2		A2	B1	A2	A2

Dieses Vorgehen ist zwar nicht falsch, aber weniger effizient als die Randomisierung pro Block (Ackerfläche). Weil die vier Ackerflächen an unterschiedlichen Orten liegen, wird es wohl systematische Unterschiede geben zwischen ihnen geben, die wir in der Auswertung mit einem Blockfaktor berücksichtigen können. Wir wenden hier folgenden Grundsatz an: "Block what you can, randomize the rest."

Es gibt noch viele weitere Möglichkeiten: Split-Plot Design, Latin Square Design, ...

Lösung: Statistische Auswertung

Die Auswertung erfolgt mit einer zwei-Weg ANOVA (Pflanze P und Dünger D) mit Blockfaktor ($Block$):

$$Y \sim P * D + Block$$

Abschliessend können wir mit *TukeyHSD* paarweise Vergleiche aller Kombinationen untersuchen.