

stark verlangsamt wird. Teilweise ist eine solche Eigenschaft von der Trockenresistenz abhängig, teilweise ist sie aber ein Ausdruck der Frühzeitigkeit und der speziellen Periodizität gewisser Pflanzenarten. Nach mehrjährigen Erfahrungen, auch in feuchten Jahren, konnte in den Svalöfer Versuchen festgestellt werden, daß das Wiesenrispengras und besonders der Weidetimothee (*Phleum nodosum*) im Hochsommer und Herbst mehr produzieren als z. B. der Rot-schwingel, der bei den ersten Schnitten des

Jahres einen hohen Ertrag gibt, später aber sehr wenig Grünmasse produziert.

Es wurden einige Grundzüge der jetzigen Prüfungsmethoden kurz geschildert, die in der Weidepflanzenzüchtung des Schwedischen Saatzuchtvereins benutzt werden. Da aber dieser Züchtungszweig noch verhältnismäßig jung ist und während der Arbeit noch immer neue Probleme hinzukommen, werden in Zukunft voraussichtlich wesentliche Änderungen und Erweiterungen der Pläne hinzukommen.

Einige Gesichtspunkte zur Benutzung der Statistik im Versuchswesen.

Von **Olof Tedin**, Svalöf.

Während der zwei letzten Jahrzehnte hat eine zahlenmäßige Behandlung von Versuchsergebnissen, die auf theoretisch-statistischen Überlegungen fußt, in allen Ländern Eingang gefunden. In den Versuchen des Schwedischen Saatzuchtvereins, wie auch sonst in Schweden, wurden die von dänischen und norwegischen Versuchsanstellern (vor allem LINDHARD, KRISTENSEN und VIK) ausgearbeiteten Methoden ziemlich früh benutzt. In den letzten Jahren wurde besonders den Methoden von FISHER und dessen Schülern Interesse entgegengebracht. Wie gewöhnlich, wenn eine neue Forschungsrichtung hervortritt, ist auch betreffs der Versuchsstatistik eine gewisse Überschätzung ihrer Bedeutung eingetreten, und eine gewisse Zurückhaltung dürfte am Platze sein, da Übertreibungen nur schädlich wirken, unter anderem dadurch, daß sie das Vertrauen für das wirklich Wertvolle in der neuen Richtung untergraben. Bei dem Schwedischen Saatzuchtverein in Svalöf mit seinen Filialstationen wie auch bei verschiedenen anderen schwedischen Versuchsorganisationen, mit welchen wir zusammenarbeiten, hat man im Laufe der Jahre umfassende Erfahrungen mit der mathematischen Behandlung von Versuchsergebnissen gemacht. Der Verfasser hat sich seit mehr als zehn Jahren besonders für diese Fragen interessiert und sie mit mehreren Kollegen eingehend besprochen. Es ist ihm dadurch möglich geworden, einen Überblick zu gewinnen, der es ihm ermöglicht, sich eine bestimmte persönliche Auffassung über diese Fragen zu bilden. Da das Zahlenmaterial, welches er behandelt hat, größtenteils aus Svalöf stammt, hat er diese Gelegenheit benutzt, gewisse Gesichtspunkte hervorzuheben, die, obgleich sie weder erschöpfend noch unbestritten sind, trotzdem als ein Beitrag zu einer aktuellen Frage ein gewisses Interesse haben können.

Wenn man seine Versuchsergebnisse mathematisch behandeln will, muß man zuerst darüber klar sein, was durch eine solche Behandlung erreicht werden kann. Vor allem muß dabei nachdrücklich betont werden, daß die Statistik nur ein Mittel und nicht das Ziel ist, und daß man, wieviel man auch rechnet, einen schlechten Versuch nicht gut machen kann. Dies hervorzuheben könnte überflüssig sein, wenn die Erfahrung nicht gezeigt hätte, daß diese Tatsache öfters nicht beachtet wird. In vielen Fällen kann indessen eine korrekte Statistik es möglich machen, beschränkte Versuchsmöglichkeiten wirklich auszunützen. Inzwischen können Tatsachen, die bei einem einfachen Studium der Resultate nicht hervortreten, durch eine statistische Behandlung kenntlich gemacht werden, und vor allem kann die Statistik in vielen Fällen Anleitung für eine zweckmäßige Planlegung der Versuche geben. In erster Linie ist ihre Bedeutung indessen eine aufklärende, sie ist ein Mittel zur Bestimmung des Wertes eines Versuches.

Das praktisch eingestellte Versuchswesen auf dem Gebiete der Landwirtschaft — von dem mehr theoretisch-wissenschaftlichen ist in diesem Zusammenhang nicht die Rede — hat die Aufgabe, kontrollierte Erfahrungen zu machen, auf welche eine Beratung für die Praxis gegründet werden kann. Diese Beratung betrifft häufig wirtschaftlich bedeutungsvolle Fragen, und es ist daher sehr wichtig, daß sie zuverlässig ist. Der mit statistischen Methoden berechnete mittlere Fehler eines Versuches oder einer Versuchsreihe gibt uns eine Möglichkeit, die Zuverlässigkeit der Resultate zu berechnen, und gibt also Auskunft darüber, ob sie als Grundlage für eine Beratung dienen können oder nicht. Ein weiterer Gesichtspunkt muß in diesem Zusammenhang klar hervorgehoben werden. Eine auf

Versuchsergebnisse begründete praktische Beratung bedeutet immer eine gewisse Verallgemeinerung. Wenn eine solche Verallgemeinerung berechtigt sein soll, müssen nicht nur die Resultate zuverlässig sein, sondern eben so wichtig ist es, daß die Versuche repräsentativ sind, d. h. die Verhältnisse, unter welchen die Versuche ausgeführt worden sind, müssen im Durchschnitt denen entsprechen, die in der praktischen Landwirtschaft herrschen und für welche die Beratung beabsichtigt ist. Dieser Gesichtspunkt hat eine interessante Konsequenz. Der Wunsch nach hochgradiger Zuverlässigkeit eines Versuches, also nach einem niedrigen mittleren Fehler, führt dazu, daß die Bodenverhältnisse, wie auch andere in dem Versuch nicht geprüfte Faktoren so gleichartig wie möglich sein müssen. Hierbei wird indessen die Repräsentativität des Versuches vermindert, und in diesem Konflikt muß der Versuchsleiter mit Takt und Sorgfalt zwischen zwei teilweise antagonistischen Prinzipien wählen. Vielleicht muß zur Zeit die Bedeutung der Repräsentativität besonders hervorgehoben werden. Verf. möchte hier die scheinbar paradoxe Behauptung aufstellen, daß, wenn viele gleichsinnige Versuche nicht angestellt werden können, man die Versuche so ausführen muß, daß diejenigen Außenfaktoren, welche nicht durch eine geeignete Versuchstechnik kontrolliert werden können, die größtmöglichen Versuchsfehler verursachen.

Eine Frage, die häufig diskutiert wird, ist die des Wertes des mittleren Fehlers eines Einzelversuches. Bestimmte Behauptungen und praktische Beratung können nur selten auf das Resultat eines Einzelversuches abgestellt werden. Ausnahmen kommen zwar vor, aber im allgemeinen muß man eine Reihe von mehr oder weniger gleichsinnigen Versuchen haben. Es ist nun bekannt, daß, wenn man den mittleren Fehler des Durchschnittes einer Versuchsreihe bestimmt, dieser in sich auch die Unsicherheit der Einzelversuche enthält, denn die Variation zwischen den Einzelresultaten beruht teilweise auf ihrer Unsicherheit, teilweise auf wirklichen Verschiedenheiten. Hieraus wird häufig der Schluß gezogen, daß der mittlere Fehler des Einzelversuches nur wenig oder keinen Wert habe. Es muß jedoch betont werden, daß er erstens eine gewisse Bedeutung für den Versuchsleiter hat, wenn dieser Kenntnis von der gewöhnlichen Größe der mittleren Fehler seiner Versuche hat, und wenn er dann in einem Versuch einen viel größeren Fehler bekommt, so liegt darin eine Warnung, entweder, daß seine Versuchsanstellung in diesem Falle schlecht gewesen ist,

oder daß er ungewöhnlich variable Versuchsverhältnisse gehabt hat. Zweitens hat der mittlere Fehler des Einzelversuches eine andere große Bedeutung, die unten erwähnt wird.

Wenn man aus einer Versuchsreihe den Durchschnittswert berechnet und in der Beratung ausnützen will, muß man sich zuerst darüber klar sein, ob dieser Wert tatsächlich eine Zusammenfassung der Resultate gibt. Ein Beispiel mag dies verdeutlichen. Die zwei Svalöfer Sorten von vierzeiliger Gerste für Nordschweden, Vega und Dore, stellen ganz verschiedene Ansprüche an den Kalkgehalt des Bodens. Auf kalkhaltigem Boden in der Provinz Jämtland gibt Dore den höchsten Ertrag, auf schwach saurem dagegen sind die beiden Sorten ungefähr gleichwertig, und auf stark saurem liefert Dore kaum die halbe Ernte derjenigen von Vega. Wenn man nun einen einfachen Mittelwert von Versuchen auf verschiedenen Böden berechnet, so würde dieser offenbar keinen richtigen Ausdruck für den Anbauwert der Sorten geben. Eine Reihe von solchen Versuchen bildet in der statischen Terminologie keine einheitliche Population, sondern ist eine Mischung von zwei, drei oder mehreren Populationen, jede mit ihrem eigenen Mittelwert, und beim Zusammenstellen der Ergebnisse muß man versuchen, diese Teilpopulationen auseinanderzuhalten und die Mittelwerte jeder einzelnen zu berechnen. In diesem Beispiel ist dies ganz offenbar und kann ohne statistische Methoden festgestellt werden. Öfters sind indessen die Verschiedenheiten nicht so auffallend, aber nichtsdestoweniger von großer Bedeutung. Man hat zuweilen Ursache, anzunehmen, daß verschiedene Sorten sich verschiedenartig zueinander verhalten, z. B. auf Lehm-, Sand- oder Moorböden, in verschiedenen Klimagebieten usw. Solche Unterschiede können im allgemeinen nur mit Hilfe statistischer Methoden festgestellt werden. Dabei müssen erstens die Versuchsergebnisse nach verschiedenen Gründen gruppiert werden, und dann sollen die mittleren Fehler der Gruppen berechnet und nach den Methoden der Varianzanalyse (Analysis of variance) miteinander und mit anderen Größen verglichen werden. Der gewissenhafte Versuchsleiter sollte überhaupt nicht Durchschnittswerte angeben, ohne zuerst untersucht zu haben, ob ein solcher Wert mit Berechtigung als der wahre Ausdruck der Einzelresultate betrachtet werden kann. Dabei muß die Streuung in der Reihe der einzelnen Versuchsergebnisse berechnet und mit den mittleren Fehlern der Einzelversuche mit Hilfe geeigneter Methoden und Tabellen verglichen werden. Wenn die beiden Größen unge-

fähr gleich sind, so ist es berechtigt, den Mittelwert anzugeben. Ist die Streuung dagegen beträchtlich größer als die Fehler, dann hat man keine einheitliche Population, und der Mittelwert hat keinen oder wenigstens einen stark beschränkten Wert.

Wie aus den obigen Darlegungen hervorgeht, hat der mittlere Fehler des Einzelversuches große Bedeutung für den Versuchsleiter und vor allem, wenn er prüfen will, ob die Berechnung von einfachen Mittelwerten berechtigt ist. Wenn dies festgestellt ist, kann eine andere Frage, die inzwischen diskutiert wird, erörtert werden, nämlich: Soll der Versuchsplan so gemacht werden, daß eine absolut korrekte Berechnung des mittleren Fehlers möglich ist, oder sind andere Versuchspläne vorzuziehen. Ein „mittlerer Fehler“ hat an und für sich keinen Wert, wenn aber das Material „normal“ verteilt ist, kann der Quotient: $\frac{\text{Differenz}}{\text{mittlerer Fehler}}$ mit von Mathematikern aufgestellten Tabellen verglichen werden und somit Auskunft geben über die Zuverlässigkeit der Resultate. Die Annahme einer „Normalverteilung“ enthält gewöhnlicherweise eine Approximation, die aber in sehr vielen Fällen berechtigt zu sein scheint. Man muß aber immer danach trachten, daß korrekte Tabellen benutzt werden. Bei einer geringen Anzahl von Einzelbeobachtungen sind die gewöhnlichen Integralen der Normalkurve nämlich nicht verwendbar, sondern müssen durch Spezialtabellen (STUDENT, FISHER) ersetzt werden. Daß der mittlere Fehler erst beim Vergleich mit gewissen Tabellen von Wert ist, folgt hieraus, daß dieser Fehler auch so berechnet sein muß, daß er den theoretischen Voraussetzungen der Tabellen entspricht. Hieraus folgen gewisse Ideenkonflikte, die kurz behandelt werden sollen.

Die korrekte Fehlerberechnung macht es bei den gewöhnlich benutzten Versuchsplänen nötig, eine gewisse Zufälligkeit in der Verteilung der Parzellen zu haben. Diese innerhalb gewisser Grenzen durchgeführte Zufälligkeit hat indessen einige Unannehmlichkeiten. Erstens muß der Plan von Versuch zu Versuch geändert werden, was Schwierigkeiten machen kann. Zweitens gibt der zufällig aufgestellte Versuchsplan manchmal größere wirkliche Fehler als die ganz systematische Anordnung der Parzellen. Als Beispiel kann der quadratische Plan mit 5 Versuchsgliedern und 5 Wiederholungen erwähnt werden. Die Einzelheiten können hier

nicht erörtert werden, sondern Verf. muß auf seine Untersuchung im J. *agricult. Sci.* 1931, S. 191—208 hinweisen. Laut dieser bekommt man beim Benutzen des sogenannten Rennerzuges durchschnittlich einen etwas niedrigeren wirklichen Fehler als bei der zufälligen Verteilung innerhalb des Rahmens des sogenannten „Latin square“. Im ersten Falle wird indessen der berechnete Fehler zu hoch und beim Gebrauch der gewöhnlichen Tabellen bekommt man also eine zu niedrige Einschätzung der Zuverlässigkeit der Resultate. Der Unterschied ist jedoch ziemlich klein und die Abweichung von der korrekten Einschätzung des Fehlers also von keiner größeren Bedeutung. Andererseits ist auch die Erhöhung des wirklichen Fehlers bei Benutzung des zufälligen Planes gering, was die Vorkämpfer der systematischen Pläne beachten sollten. Nach der Meinung des Verfassers muß man prinzipiell solche Pläne benutzen, die eine korrekte Fehlerberechnung gestatten; will man gewisse systematische Pläne benutzen, muß man zuerst durch besondere Untersuchungen darüber klar sein, ob die Abweichungen in der Schätzung der Zuverlässigkeit von größerer Bedeutung sind oder nicht.

Endlich mögen einige Worte über die komplizierten (sog. „confounded“) Versuchspläne gesagt werden, die in den letzten Jahren von FISHER und seinen Schülern vorgelegt worden sind. Oben wurde hervorgehoben, daß die Repräsentativität nicht zum Vorteil des niedrigen mittleren Fehlers geopfert werden sollte. Gewöhnlich hat man in den Versuchen nur die Möglichkeit, wenige Faktoren zu variieren, während in der Praxis eine Menge von solchen gleichzeitig variieren. Ein Versuch hat daher, im Vergleich zu der Praxis, eine geringe Repräsentativität. Je mehr Faktoren in ein und demselben Versuch variiert werden können, desto bessere Auskunft erhält man über den Wert der einen oder anderen Maßregel, Sortenwahl, Düngung u. dgl. Die gleichzeitige Prüfung mehrerer Faktoren ist indessen bisher unmöglich gewesen, da sie allzu verwinkelte Versuchspläne forderte. Diese Schwierigkeit scheint jetzt durch die statistische Theorie teilweise gelöst zu sein. Wir haben bis jetzt aber noch keine nennenswerten Erfahrungen mit diesen neuen Plänen sammeln können. Ich muß mich deshalb darauf beschränken, die Überzeugung auszusprechen, daß diese neuen Pläne wertvolle Beiträge zu der Vervollkommnung des Versuchswesens geben werden.