

## Der erste Schritt in die digitale Zukunft

### Erkennen Sie Zusammenhänge in Ihren Daten

Schnellere Computer und die fortschreitende Digitalisierung aller Prozesse haben dafür gesorgt, dass die Datenmengen in Industrie und Verwaltung schneller wachsen als erwartet. Aber kaum sind diese Datenmengen in der Welt, entstehen auch die Imperative ihrer Nutzung: „Daten sind das Rohöl des einundzwanzigsten Jahrhunderts“, „Algorithmen werden 30 % aller Arbeitsplätze ersetzen“, „Unternehmen müssen ihre Produkte oder Dienstleistungen digitalisieren“.

Leider weist keiner den Weg, der vom Status quo zur digitalen Zukunft führt. Was also ist dran an der digitalen Revolution, und wie kann man möglichst unmittelbaren Nutzen ziehen und sich für die Zukunft wappnen?



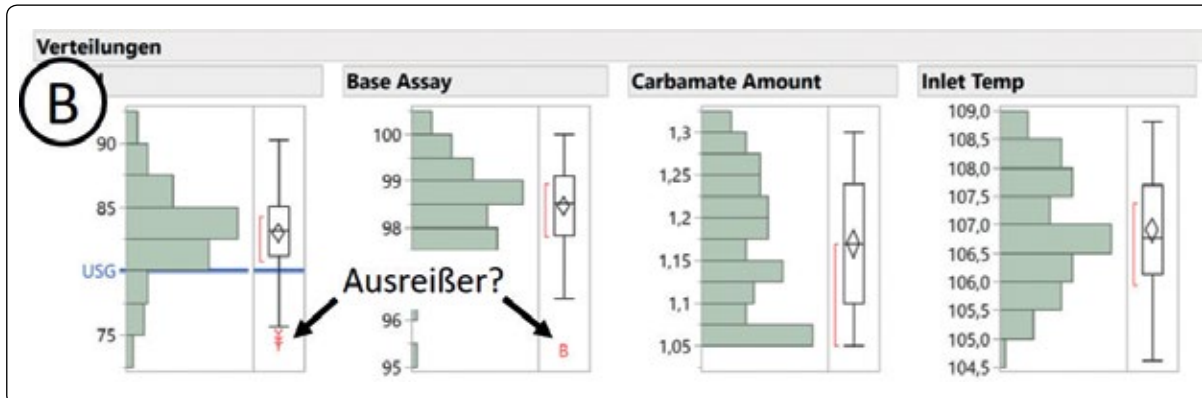
Bernd Heinen,  
SAS Institute

Wie immer, wenn neue Technologien zur Verfügung stehen, gibt es eine Menge vollkommen innovativer Ideen, die vorher nicht zu verwirklichen waren. Autonom fahrende Autos, eine annähernd perfekte Gesichtserkennung, selbstlernende Roboter und vieles mehr zeigen das Potenzial einer konsequenten Nutzung von Daten und digitaler Technik. So interessant und verheißungsvoll (oder bedrohlich, je nach Anwendung und Auffassung) diese einzelnen Projekte sind, so wenig taugen sie als Richtlinie für tägliches Handeln im Rahmen der Prozesse, die heute aktuell sind oder als Wegweiser für eine mittelfristige Entwicklung bestehender Produkte oder Dienstleistungen. Der Wandel vollzieht sich meistens schrittweise, und mit dem nötigen Rüstzeug kann man sowohl unmittelbare Erfolge erzielen als auch die Basis für eine nachhaltige Entwicklung legen.

Unter Rüstzeug ist zunächst die Offenheit und Bereitschaft zu verstehen, sich regelmäßig und konkret mit den Daten auseinanderzusetzen, die im Umfeld der eigenen beruflichen Tätigkeit anfallen oder beschafft werden können. Dieser Appell richtet sich vor allem an Menschen, deren Ausbildung nicht in erster Linie der Datenanalyse galt – also an Laboranten, Naturwissenschaftler oder Ingenieure. Es ist sicher gut, wenn einzelne Personen sich diese Sicht auf die Arbeit aneignen; mittelfristig sollte es aber für eine Mehrheit im Unternehmen selbstverständlich werden, Datenaspekte zu berücksichtigen. Nicht zuletzt ist auch die Unterstützung der Vorgesetzten gefragt, denn wo Digitalisierung eine Rolle spielt, ist Software nicht fern. Und obwohl die zu analysierenden Daten in Tabel-

lenform überführt werden müssen und Grafiken eine entscheidende Rolle spielen, sind die gängigen Tabellenverarbeitungs- und Grafikprogramme nicht Teil der Lösung, sondern Teil des Problems. Nur Statistikprogramme bieten die Hilfe, die den Einstieg in diese Methodik erleichtert und eine zuverlässige Basis für zukünftige Entwicklung bildet. Man sollte sich von dem Begriff Statistik allerdings nicht abschrecken lassen. Die möglicherweise schlechten Erfahrungen der eigenen Ausbildung haben keinen Bestand angesichts der vielversprechenden Lösungen aktueller Anwendungen.

Angenommen, die erste Hürde ist geschafft, Software ist vorhanden, Daten sind importiert. Worin kann dann der unmittelbare Gewinn liegen? Meistens steht man ja vor spezifischen Aufgaben oder Fragestellungen. Man möchte den Zusammenhang zwischen verschiedenen Tabellenspalten – in der Statistik spricht man von Variablen – so gut verstehen, dass man dieses Wissen für Verbesserungen einsetzen kann. Dazu ist es schon einmal gut zu wissen, über welche Bereiche oder Stufen sich die beteiligten Größen überhaupt bewegen. Bewegt sich alles in demselben Bereich oder gibt es extreme Beobachtungen? Sind diese häufig oder selten? Treten sie nur bei einem Messwert auf oder bei mehreren? Wenn bei mehreren, dann gleichzeitig oder unabhängig voneinander (Grafik 1)? Informative Grafiken, die gezielt für den Einsatz in einem statistischen Umfeld geschaffen wurden, geben Einblick in diese Zusammenhänge und helfen, solche und ähnliche Fragen zu beantworten. Mitunter werfen sie aber auch Fragen auf, auf die man sonst nicht gekommen wäre, die aber



Grafik 1: Was sind die Verteilungen der Messgrößen? Sind sie symmetrisch? Unsymmetrisch? Gibt es Ausreißer? Wo liegen die Messwerte im Vergleich zu Referenzgrößen (USG = Untere Spezifikationsgrenze).

durchaus von fachlicher Relevanz sind. Somit liegt der erste Vorteil klar auf der Hand: Man kann nur auf Situationen, Trends oder Abweichungen reagieren, wenn man sie kennt.

Sind die wichtigsten Variablen mit Ihren Ausprägungen bekannt, kann man sich einen Eindruck von wechselseitigen Abhängigkeiten und Einflüssen verschaffen, oder ihn, soweit man ihn aus den erstellten Grafiken schon gewonnen hat, vertiefen. Intuitiv nimmt man dabei schon eine Einteilung der Variablen vor: in Zielgrößen, die die eigentlich interessierenden Ergebnisse beschreiben und in Einflussgrößen oder Faktoren, die alles umfassen, was die Zielgrößen beeinflusst. So ist man eigentlich an den Komponenten einer Kunststoffmischung oder den Betriebsparametern eines Extruders primär gar nicht interessiert, wohl aber an den Qualitäten des produzierten Werkstücks, das in mehrfacher Hinsicht beschrieben werden kann. Bruchfestigkeit, Oberflächenstruktur, Maß-

treue, Homogenität, Flexibilität und andere Qualitäten beschreiben das gewünschte Ergebnis. Die verwendeten Ausgangsstoffe und Prozessparameter sind „nur“ insofern von Bedeutung, als man ohne ihre Kenntnis das gewünschte Ergebnis nicht erzielen kann. Einfache, paarweise Vergleiche zwischen je einer Zielgröße (interessant!) und einem Faktor (Mittel zum Zweck) helfen, die Art und das Gewicht der jeweiligen Abhängigkeit einzuschätzen (Grafik 2). Auch hier ergibt sich ein unmittelbarer Vorteil, denn in Einzelfällen mag diese Kenntnis schon ausreichen, um korrigierend in Rezepturen oder Prozesse eingreifen zu können. Auch für eine Diskussion unter Kollegen oder für den Bericht an Entscheider wird so eine sachliche, rationale Basis geschaffen.

Ein wirklich mächtiges Werkzeug kann man sich schaffen, indem man die einzelnen Aspekte in einem statistischen Modell vereint. Dies geschieht hinzubekommen erfordert etwas Erfahrung, aber die hat man

in den beiden vorherigen Schritten ja schon gesammelt. Indem man nun den simultanen Einfluss der wichtigsten Faktoren beschreibt, kann man mit einer einzigen Methode mehrere Fragen untersuchen (Grafik 3):

- Wieviel Einfluss haben die gewählten Faktoren?
- Gibt es Hinweise darauf, dass Faktoren in dem Modell fehlen?
- Wirken die einzelnen Faktoren über ihren ganzen Wertebereich gleichmäßig auf die Zielgröße, oder gibt es Bereiche mit stärkeren oder schwächeren Wirkungen?
- Wirken die Faktoren unabhängig voneinander auf das Zielergebnis, oder gibt es Synergismen oder Antagonismen zwischen mehreren Faktoren?

Somit hilft das Modell, ein besseres Verständnis eines Ablaufs oder von Materialeigenschaften zu erlangen. Im besten Fall bestätigt es das schon vorhandene Fachwissen auf eine unabhängige, sachliche Art und Weise; im zweitbesten Fall weist es auf Verständnislücken hin, deren Folgen sich eventuell sogar schnell beheben lassen, wenn man sich ihrer erst einmal bewusst ist. Das wichtigste Ziel der Modellbildung besteht aber darin, dass es ermöglicht, abhängig von unterschiedlichen Kombinationen der Faktoren Vorhersagen über die zu erwartenden Werte der Zielgrößen zu erhalten. Somit kann man einfach und schnell alternative Szenarien durchspielen, optimale Kombinationen ermitteln und Ergebnisse simulieren, die auch Variationen im

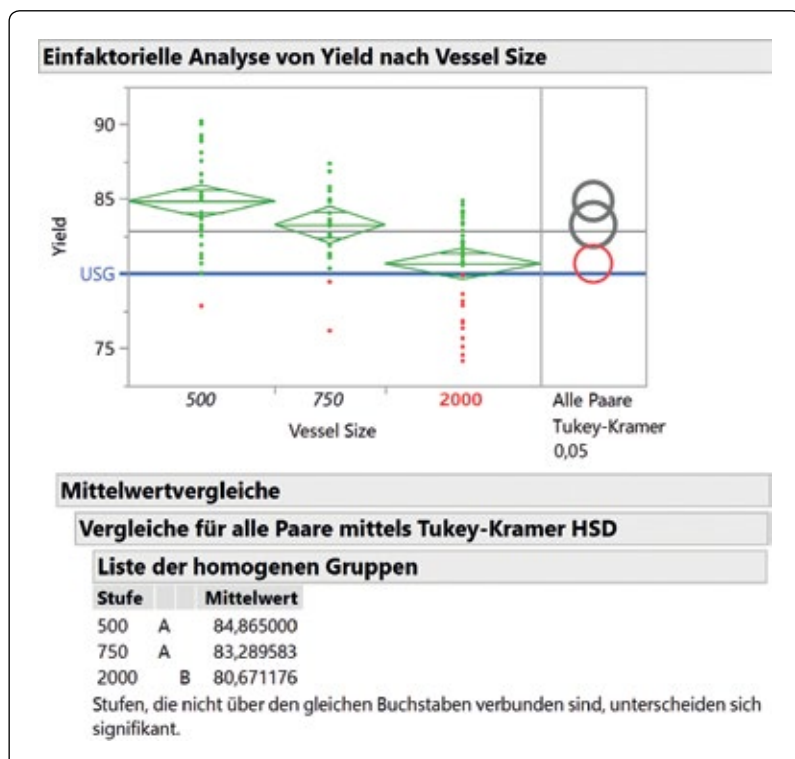
Ausgangsmaterial, Hysterese bei Schaltungen oder andere Schwankungsursachen berücksichtigen. In den meisten Fällen wird man so berechnete Konstellationen experimentell überprüfen, aber auf jeden Fall erspart man sich den ganzen mühseligen, aufwändigen und teuren Weg, diese Ergebnisse durch Experimente zu erarbeiten. Dabei können vielfältige Ziele erreicht werden:

- eine Spezifikation möglichst genau einhalten
- einen Zielwert maximieren oder minimieren
- möglichst robust gegenüber Schwankungen des Ausgangsmaterials oder einzelner Prozessschritte sein

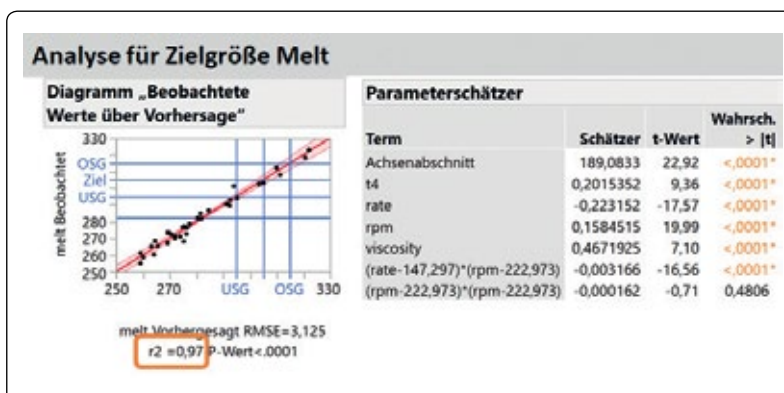
Den Wert von Daten und deren Analyse kann man nicht überschätzen. Mit der Analyse bestehender Daten kann man auch keinen Fehler machen, wenn man mit kleinen Schritten beginnt und sich an immer komplexere Fragestellungen herantastet. Tut man dies in einem integrierten Umfeld, wie es z.B. die Software JMP von SAS Institute bietet, sind die Lernzyklen kurz und Erfolge stellen sich schnell ein. Ein Erfolg wird Sie auf dem ganzen Weg begleiten: Sie werden in der Lage sein, bessere Entscheidungen schneller zu treffen. Und die auf diesem Weg gesammelten Erfahrungen bilden dann auch die solide Basis, von der aus Sie für sich und Ihr Unternehmen die digitale Zukunft gestalten können.

Bernd Heinen, JMP Systems Engineer, SAS Institute GmbH, Heidelberg

- bernd.heinen@jmp.com
- www.jmp.com



Grafik 2: Die statistische Analyse zeigt, dass sich die Beobachtungen mit Vessel Size 500 hinsichtlich der Zielgröße Yield signifikant von den beiden anderen Gruppen unterscheiden.



Grafik 3: Das Modell für die Zielgröße „melt“ erklärt insgesamt 97 % der Variation in den Daten, nur 3 % bleiben zufällige Unsicherheit.