

## Gefühlte Wahrheiten, Risiken und statistische Tests

Roman Wenig, Frank Stockhaus,  
TEQ® Training & Consulting GmbH

Unsere moderne Welt ist ohne elektronische Datenverarbeitung nicht mehr vorstellbar: Aus Daten werden Informationen generiert. Diese dienen wiederum dazu, Entscheidungen zu treffen. Der Prozess der Entscheidungsfindung kommt praktisch überall und ständig vor. Entscheidungen sind es, nach denen wir unsere Prozesse lenken, unsere Dienstleistungen und materiellen Produkte entstehen lassen. Dabei geht es nicht nur darum, dass Entscheidungen überhaupt getroffen werden, sondern auch, wie gut diese sind. Entscheidungen können aufgrund von Erfahrungen und aus der persönlichen Einschätzung einer vorliegenden Situation heraus getroffen werden, aber auch unabhängig von den handelnden Personen methodisch strukturiert auf der Basis harter Zahlen, Daten und Fakten.

Insbesondere in der industriellen Produktion werden Prozesse aus wirtschaftlichen Gründen in der Regel anhand von Stichprobenprüfungen gelenkt. Das zwingt uns dazu, Risiken in Kauf zu nehmen: Eine Stichprobenprüfung birgt immer das Risiko in sich, eine Fehlentscheidung zu treffen. Um dieses Risiko abwägen zu können, ist das Beherrschen von statistischen Methoden unerlässlich. Zu den grundlegenden Methoden gehören die statistischen Tests.

Momentan gibt es neben den statistischen Tests keine einfache Methode, die eleganter und gleichzeitig simpler mit der Variation unserer Prozesse umgeht: Es wird ein binäres Ergebnis geliefert, welches mit einem kalkulierten Risiko versehen ist. Statistische Tests trennen den Zufall von systematisch wirkenden Änderungen, das Signal vom Rauschen, die Spreu vom Weizen. Sie sind aus keinem Schritt des Produktlebenszyklus wegzudenken und sollten deshalb ein Standardwerkzeug eines jeden sein, der seine Entscheidungen mit Hilfe von Zahlen, Daten und Fakten fällt.

Damit der statistische Test das gesuchte Ergebnis zuverlässig liefert, muss der Anwender zum einen die Methode an sich beherrschen, zum anderen aber auch gewisse Bedingungen für den Prozess des Vorbereitens, Durchführens und Nachbereitens einhalten. Diesen Prozess haben wir im Folgenden für die Parametertests, die Tests von Parametern der Grundgesamtheit auf Signifikanz, schrittweise beschrieben. In jedem Schritt sind die Fragen und Aufgaben aufgelistet, die Sie vor der weiteren Bearbeitung beantworten bzw. lösen sollten. Im Anschluss wird das Vorgehen an einem Beispiel veranschaulicht.

### 1. Welche Wahrscheinlichkeitsverteilung beschreibt die vorliegende Situation am besten?

- Klären Sie:
  - o Haben Sie eine Stichprobe? Beschreibt Ihre Stichprobe die Situation „verlässlich“?
  - o Liegen Ihnen diskrete oder kontinuierliche Daten vor?
- Bei diskreten Daten: Welches Modell liegt vor? (Danach richten sich die Wahrscheinlichkeitsverteilungen und deren Parameter.)
  - o Anzahl der fehlerhaften Einheiten: Binomialverteilung
  - o Anzahl der Fehler je Einheit: Poissonverteilung
- Bei kontinuierlichen Daten: Testen Sie, ob die Daten mit dem Modell der Normalverteilung beschrieben werden können.

### 2. Welcher Parameter der Verteilung soll getestet werden?

- Was wollen Sie wissen? Wozu wollen Sie die Stichprobe verwenden? Wie gut kennen Sie die Grundgesamtheit – ist Ihnen eventuell deren Streuung bekannt?
- Klären Sie:
  - o Soll ein Lageparameter oder soll
  - o ein Streuungsparameter oder sollen
  - o beide oder soll
  - o eine Fehlerrate getestet werden?
- Führen Sie eine Datenanalyse durch: Ermitteln Sie die statistischen Kennwerte für die Lage- und Streuungsparameter bzw. für die Fehlerrate aus Ihrer Stichprobe.

### 3. Was für einen Vergleich wollen Sie durchführen?

- Legen Sie fest, ob es sich um
  - o einen Soll-Ist-Vergleich (Stichprobe gegen Kundenforderung),
  - o einen Ist-Ist-Vergleich (zwei Stichproben) oder um
  - o den Vergleich mehrerer Stichproben handelt.

### 4. Welche Hypothesen sind zu formulieren?

- Erstellen Sie eine Null- und eine Alternativhypothese.
  - o Die Nullhypothese  $H_0$  beschreibt, dass sich nichts geändert hat: Das zu erwartende Signal verschwindet im Rauschen (d. h. in der Streu-

ung bzw. im Vertrauensbereich des zu testenden Parameters).

- o Die Alternativhypothese  $H_1$  ist eine Aussage, die der  $H_0$  entgegensteht. ( $H_1$  soll die zu prüfende Vermutung enthalten!)
- o Entweder
  - weicht der Parameter signifikant ab (d. h. er ist ungleich) oder
  - er ist signifikant kleiner oder
  - er ist signifikant größer.

**5. Welcher Test ist zu wählen?**

- Wählen Sie den richtigen statistischen Test aus. Dabei assistiert Ihnen statistische Software, bei der Entscheidungsbäume hinterlegt sind.
- Legen Sie die Irrtumswahrscheinlichkeit  $\alpha$  fest. Übliche Werte für  $\alpha$  sind 5 %, 1 % und 0,1 %.
- Führen Sie den statistischen Test durch.

**6. Wie lautet das Testergebnis?**

- Lesen Sie den P-Wert aus dem Testergebnis ab. Vergleichen Sie diesen mit der Irrtumswahrscheinlichkeit  $\alpha$ .
- Alternativ: Lesen Sie Prüfgröße ab. Vergleichen Sie diese mit dem kritischen Quantil.
- Schlussfolgern Sie aus diesem Vergleich die „statistisch formale“ Entscheidung bezüglich der Nullhypothese:
  - o  $P \geq \alpha$ : Die Nullhypothese wird nicht verworfen.
  - o  $P < \alpha$ : Die Nullhypothese wird verworfen, damit ist die Alternativhypothese als zutreffend anzusehen.

**7. Wie ist das Testergebnis in Bezug auf die reale Situation zu interpretieren?**

- Formulieren Sie einen Antwortsatz, der das Testergebnis in folgender Form beschreibt: „Das heißt, dass... (der getestete Parameter signifikant/nicht signifikant ist bzw. das Ergebnis der Datenanalyse zufällig/nicht zufällig ist).“
- Übertragen Sie das Testergebnis auf Ihre reale Situation.

**8. Wie „gut“ ist der Test?**

- Stellen Sie fest, ob das Testergebnis in Abhängigkeit
  - o vom Stichprobenumfang,
  - o von der Irrtumswahrscheinlichkeit  $\alpha$ ,
  - o von der Wahrscheinlichkeit für einen Fehler zweiter Art bzw. von der Trennschärfe
  - o und von einer nachzuweisenden „technologisch sinnvollen“ Lageverschiebung oder von einer „technologisch sinnvollen“ Streuungsänderung oder von einer „technologisch sinnvollen“ Änderung der Fehlerrate

plausibel ist.

Diese acht Schritte stellen gewissermaßen einen Wegweiser bei der Durchführung von statistischen Tests dar und sollen Ihnen helfen, sich dabei besser zurechtzufinden. Sie sollen nun an einem Beispiel veranschaulicht werden. Die notwendigen statistischen Auswertungen werden hier mit der Software qs-STAT® gezeigt.

Bei der Prozessfähigkeitsuntersuchung für die Breite einer Passnut wurden für  $C_p$  und  $C_{pk}$  nur Werte um 1 ermittelt. Die Forderungen von mindestens 1,33 sind somit nicht erfüllt. Um bessere, d. h. größere Werte zu erreichen, ist es also erforderlich, die Streuung zu reduzieren. Der verantwortliche Prozessingenieur hat auch sofort eine Idee zur Verbesserung und setzt diese um. Die nächsten 20 Teile werden unter den geänderten Prozessbedingungen gefertigt und danach gemessen.

Die Verteilung von kontinuierlichen Merkmalen wird häufig durch Grafiken wie Werteverlauf, Histogramm oder Box-Plot visualisiert. Im Beispiel waren sich Prozessingenieur, Anlagenbediener und Fertigungsleiter schnell einig: Aus dem Box-Plot in Bild 1 ist eindeutig zu erkennen, dass sich die Streuung unseres Merkmales durch die Verbesserungsmaßnahme deutlich reduziert hat!

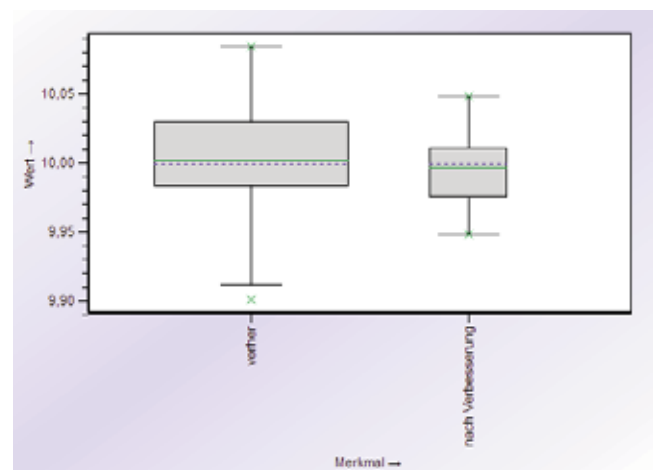


Bild 1: Box-Plot für das Merkmal vor und nach der vermeintlichen Verbesserungsmaßnahme

Auch die numerischen Werte der Standardabweichungen aus beiden Stichproben belegen eine kleinere Prozessstreuung, siehe Bild 2. Daraus wird eine Verringerung der Standardabweichung auf 76 % gegenüber dem Ausgangswert berechnet.

Teilnr.	Teilebez.	Drehteil
Merkm.Nr.	Merkm.Bez.	$\bar{x}$ s
1	vorher	10,002222      0,0334
2	nach Verbesserung	9,995945      0,0254

Bild 2: Mittelwert  $\bar{x}$  und Standardabweichung s vor und nach der Verbesserungsmaßnahme

Mit einem guten Gefühl beauftragt der Prozessingenieur eine neue Prozessfähigkeitsstudie auf der Basis von insgesamt 125 Teilen. Nach mehreren Schichten liegt das Ergebnis vor und alle Beteiligten sind etwas irritiert, als sie die ermittelten Werte erfahren:  $C_p = 0,99$  und  $C_{pk} = 0,95$ .

**Was war geschehen?**

Bei den oben beschriebenen „gefühlten Wahrheiten“ wurden zwei wichtige Punkte übersehen:

- Die aus den beiden Stichproben berechneten Standardabweichungen  $s_{vorher}$  und  $s_{nachher}$  sind nur Schätzungen der wahren, aber unbekanntes Standardabweichungen  $\sigma$  der beiden Grundgesamtheiten. Allein aus der Tatsache, dass die Standardabweichung  $s_{nachher}$  kleiner ist als  $s_{vorher}$ , kann nicht geschlossen werden, dass dies in der Grundgesamtheit genauso ist. Die festgestellte Abweichung kann auch Zufall sein!
- Der **Stichprobenumfang** nach der Verbesserungsmaßnahme ist mit 20 Teilen relativ **klein**. Allgemein gilt: Je kleiner der Stichprobenumfang, umso größer ist die Unsicherheit für den Bereich, in dem die wahre Standardabweichung der Grundgesamtheit erwartet wird.

Eine objektive Bewertung der Frage, ob die Verbesserungsmaßnahme zu einer signifikanten Verringerung der Standardabweichung  $\sigma$  der Grundgesamtheit geführt hat, kann nur ein statistischer Test liefern.

Halten wir uns bei der Testdurchführung an die oben eingeführten 8 Schritte.

**1. Welche Wahrscheinlichkeitsverteilung beschreibt die vorliegende Situation am besten?**

Das kontinuierliche Merkmal „Breite der Passnut“ kann mit der Normalverteilung beschrieben werden. Die Software qs-STAT® findet für beide Stichproben eine Normalverteilung als geeignetes Modell, das bestätigen die Histogramme in Bild 3.

**2. Welcher Parameter der Verteilung soll getestet werden?**

Es soll getestet werden, ob sich der Parameter Standardabweichung durch die Maßnahme des Prozessingenieurs verringert hat.

**3. Was für einen Vergleich wollen Sie durchführen?**

Es sind zwei Grundgesamtheiten bezüglich ihrer Standardabweichung zu vergleichen. Dieser Test wird als Ist-Ist-Vergleich oder auch als 2-Stichproben-Test bezeichnet.

**4. Welche Hypothesen sind zu formulieren?**

Mathematisch formuliert lauten die Null- ( $H_0$ ) und Alternativhypothese ( $H_1$ ):

$$H_0 : \sigma_{vorher}^2 \leq \sigma_{nachher}^2$$

$$H_1 : \sigma_{vorher}^2 > \sigma_{nachher}^2$$

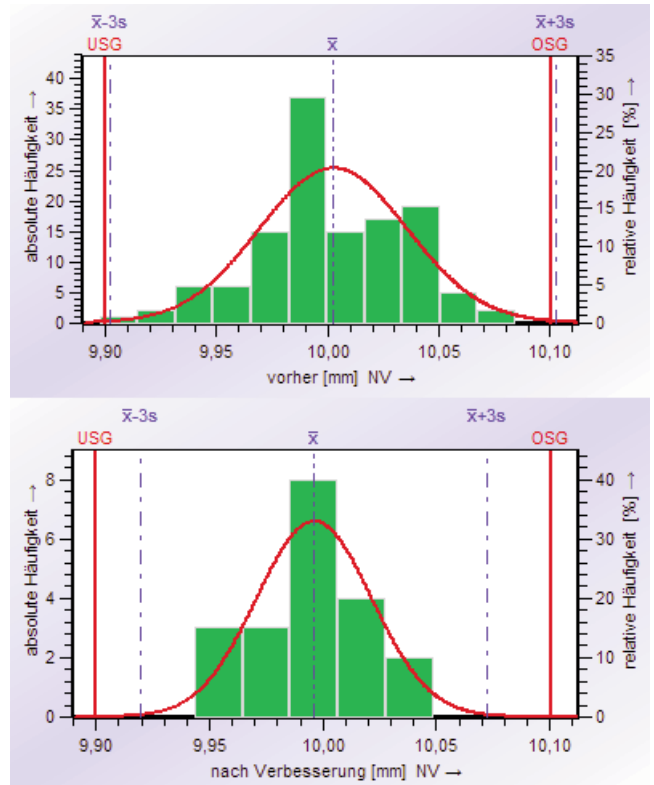


Bild 3: Histogramm vor und nach der Verbesserungsmaßnahme

Inhaltlich formuliert soll mit dem Test bestätigt werden, dass die Streuung nach der Verbesserung  $\sigma_{nachher}^2$  kleiner ist als die Streuung vor der Verbesserung  $\sigma_{vorher}^2$ . Die Streuung wird häufig durch die Bezeichnung Varianz ausgedrückt.

**5. Welcher Test ist zu wählen?**

Für Statistikexperten steht bereits nach den Ausführungen der Schritte 1. bis 4. fest: Es ist der „Zweistichproben F-Test“ zu wählen.

Für alle anderen Anwender ist die Nutzung des Test-Assistenten im Modul Stichprobenanalyse der Software qs-STAT® zu empfehlen. Die Auswahl des geeigneten Tests wird durch die Darstellung in einer

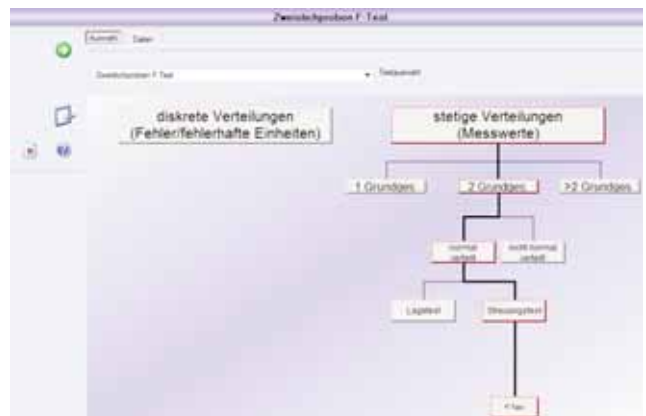


Bild 4: Auswahl des Zweistichproben F-Tests

Baumstruktur erleichtert. Durch einen Mausklick in das gewünschte Rechteck öffnet sich die nächste untergeordnete Ebene. Auf diese Weise erreicht man sukzessive den geeigneten Test. Für das Beispiel ist das Ergebnis in Bild 4 dargestellt.

**6. Wie lautet das Testergebnis?**

Im Test-Assistenten von qs-STAT® ist das Testergebnis auf der Registerkarte „Testen“ dargestellt (Bild 5). Achtung: Die Auswahl der gewünschten Hypothesen (siehe Schritt 4) erfolgt im linken unteren Bereich des Fensters!



Bild 5: Auswahl der Hypothesen und Testergebnisse

Die eleganteste Art, das Testergebnis zu bewerten, ist der Vergleich des P-Wertes mit der vorgegebenen Irrtumswahrscheinlichkeit  $\alpha$ . Im Beispiel soll die Irrtumswahrscheinlichkeit  $\alpha = 5\%$  verwendet werden. Dem Bild 6 kann entnommen werden, dass der P-Wert mit  $8,5\%$  größer ist als die Irrtumswahrscheinlichkeit  $\alpha = 5\%$ . Damit wird die Nullhypothese nicht widerlegt. Dieses Ergebnis wird auch verbal im Feld „Testergebnis“ ausgegeben.

Vergleich der Varianzen von zwei Normalverteilungen					
F-Test 2					
vorher			nach Verbesserung		
n	s	s <sup>2</sup>	n	s	s <sup>2</sup>
128	0,0234	0,00112	20	0,0254	0,000648
H <sub>0</sub> : Die Varianz der 1. Grundgesamtheit ist kleiner/gleich der Varianz der 2. Grundgesamtheit					
H <sub>1</sub> : Die Varianz der 1. Grundgesamtheit ist größer der Varianz der 2. Grundgesamtheit					
Testniveau	kritische Werte		Prüfgröße		
	unten	oben			
$\alpha = 5\%$	—	1,93	$\frac{s_1^2}{s_2^2} = 1,73800$		
$\alpha = 1\%$	—	2,58			
$\alpha = 0,1\%$	—	3,67			
Testergebnis					
Nullhypothese wird nicht widerlegt					
P-Wert %	8,457 %				

Bild 6: Testergebnisse (Auszug aus Bild 5)

Es ist zu beachten, dass in diesem Schritt nur eine „statistisch formale“ Entscheidung bezogen auf die Nullhypothese getroffen wird. Es gibt dabei nur zwei Möglichkeiten:

- Die Nullhypothese wird widerlegt oder
- die Nullhypothese wird nicht widerlegt.

**7. Wie ist das Testergebnis in Bezug auf die reale Situation zu interpretieren?**

Das Ergebnis aus Schritt 6 bedeutet inhaltlich, dass keine signifikante Verringerung der Streuung festgestellt werden konnte.

Anders formuliert: Der Wert der Standardabweichung in der Stichprobe nach der Verbesserungsmaßnahme ( $s_{nachher} = 0,0254$  mm) ist nur zufällig kleiner als die Standardabweichung aus der Stichprobe vor der Verbesserung ( $s_{vorher} = 0,0334$  mm).

Damit konnte die „gefühlte Wahrheit“ der vermeintlichen Verbesserung, die die beteiligten Personen aus den Stichprobendaten abgeleitet hatten, durch den objektiven statistischen Test nicht bestätigt werden!

**8. Wie „gut“ ist der Test?**

Die „Güte“ eines statistischen Tests hängt mit den möglichen Fehlentscheidungen zusammen, die man treffen könnte. Die Zusammenhänge sind kompliziert und hier nicht vollständig zu erörtern. Deshalb nur soviel, dass man aus statistischer Sicht zwei Fehlentscheidungen treffen kann:

- Die Nullhypothese wird abgelehnt, obwohl sie richtig ist (Fehler 1. Art oder  $\alpha$ -Fehler).
- Die Nullhypothese wird nicht abgelehnt, obwohl sie falsch ist (Fehler 2. Art oder  $\beta$ -Fehler).

In der Software qs-STAT® wird in der Registerkarte „Planung“ die sogenannte Operationscharakteristik angezeigt. Das ist  $\beta$ , die Wahrscheinlichkeit für den Fehler 2. Art. Sie hängt in unserem Fall vom Verhältnis

$$\text{der Streuungen in den Grundgesamtheiten } \frac{\sigma_{vorher}}{\sigma_{nachher}}$$

ab und wird als Funktion dieses Quotienten dargestellt. Man muss jetzt einen Wert für den Quotienten und die Wahrscheinlichkeit  $\beta$  bzw. die Trennschärfe  $1 - \beta$  vorgeben, dann wird der Mindest-Stichprobenumfang für den Test berechnet.

Im Beispiel wird  $\frac{\sigma_{vorher}}{\sigma_{nachher}} = 1,5$

und  $\beta = 10\%$  bzw.  $1 - \beta = 90\%$  vorgegeben. Das bedeutet, dass der Test

eine Verringerung der Streuung auf  $\frac{2}{3} \approx 67\%$

mit einer Wahrscheinlichkeit von  $90\%$  erkennen soll. Bild 7 zeigt die gewünschte Operationscharakteristik als rote Linie, der dafür notwendige Stichprobenumfang beträgt mindestens 54 Teile.

Im geschilderten Beispiel hat der Prozessingenieur allerdings nur eine Stichprobe von 20 Teilen gezogen. Welche Konsequenzen hat das? Die Operationscharakteristik für  $n = 20$  wird als grüne Linie im Bild 7 angezeigt.



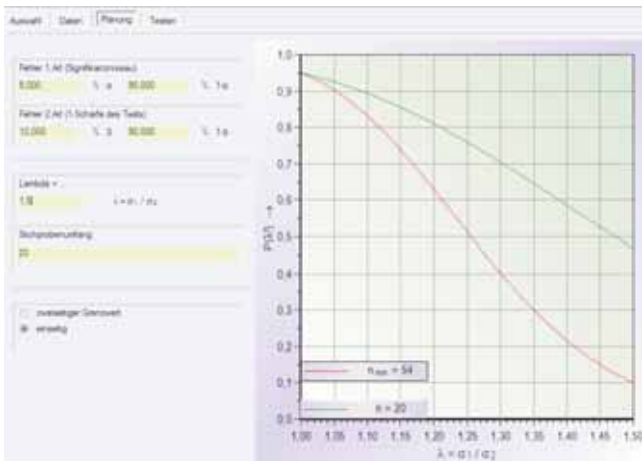


Bild 7: Operationscharakteristik für den Zweistichproben F-Test

Bei einem Abszissenwert  $\frac{\sigma_{\text{vorher}}}{\sigma_{\text{nachher}}} = 1,5$

liest man eine Wahrscheinlichkeit von ca. 47 % auf der Ordinate ab. Das heißt, eine Verringerung der Standardabweichung auf ca. 67 % führt mit einer Wahrscheinlichkeit von 47 % ( $\beta$ ) zur Beibehaltung der Nullhypothese. Das bedeutet, dass keine Verringerung der Standardabweichung erkannt wurde! Anders formuliert: Eine Verringerung der Standardabweichung auf ca. 67 % wird nur mit einer Wahrscheinlichkeit von 53 % ( $1-\beta$ ) vom Test erkannt!

Wir müssen also feststellen, dass der vom Prozessingenieur gewählte Stichprobenumfang von  $n = 20$  zu klein war, um die gewünschte Reduzierung der Standardabweichung statistisch gesichert zu erkennen.

## Fazit

Bei den statistischen Tests handelt es sich um ein grundlegendes Werkzeug der schließenden Statistik. Sie sind ein wesentlicher Bestandteil vieler angewandter statistischer Verfahren, wie beispielsweise der Messsystemanalyse, der statistischen Prozessregelung SPC oder auch der statistischen Versuchsplanung DoE. Auch zur Entscheidungsfindung bei Prozessverbesserungsmaßnahmen und bei Six Sigma-Projekten sind sie unverzichtbar.

Dieser Beitrag soll einen kleinen Einblick in die Möglichkeiten der Anwendung statistischer Tests geben. Keinesfalls war es beabsichtigt, alle Aspekte bei der Anwendung detailliert zu diskutieren.

Wir hoffen, Ihr Interesse an dieser Thematik geweckt zu haben. In diesem Fall empfehlen wir Ihnen den Besuch unserer Seminare „Statistische Verfahren – Teil 1 bis 3 (020-STM, 021-STM und 039 STM)“. Diese Seminare legen den Schwerpunkt auf das Verständnis der Verfahren, einschließlich der mathematischen Zusammenhänge und auf eine gewisse Breite in der Darstellung verfügbarer Tests. Rechenintensive Methoden werden mit der Standardsoftware Microsoft® Excel realisiert. Die nächsten Termine für diese Seminare entnehmen Sie bitte der Terminübersicht auf Seite 78 dieser PIQ-Ausgabe.

Bei In-House-Trainings ist alternativ auch die Nutzung der kommerziellen Statistikpakete von qs-STAT® oder MINITAB® möglich.

Haben Sie Fragen? Schreiben Sie einfach an die Autoren [frank.stockhaus@teq.de](mailto:frank.stockhaus@teq.de) oder [roman.wenig@teq.de](mailto:roman.wenig@teq.de).

# Best of Class – Worldleader DIE Lasermikrometer

Ihr kompetenter Partner für präzise, berührungsfreie Dimensionsmessungen im fertigungsbegleitenden Einsatz!

## Eigenschaften

- Kein Nachkalibrieren
- Absolute Präzision rückführbar auf nationale Normale
- Meßobjekt muß nicht mittig im Meßfeld sein
- Oberflächenbeschaffenheit ohne Einfluß: matt, glänzend, transparent, hell, dunkel, farbig
- Messung ab  $\varnothing 0,025$  mm
- Wiederholpräzision ab  $\pm 0,1\mu\text{m}$ !
- Meßunsicherheit  $\leq \pm 0,4\mu\text{m}$

## Anwendungen

- PKD-Schneidwerkzeuge
- Hartmetall-Fräser und -Bohrer
- Rundstäbe
- Lehrdorne, Zylinderstifte
- Achsen, Wellen, Kolben
- Einspritzdüsen, Ventile
- Nocken- und Kurbelwellen
- Kopierer- und Druckerwalzen

Z-Mike Lasermeßtechnik GmbH  
Steinschönauer Str. 4c Tel.: 06078-935711  
D-64823 Groß-Umstadt Fax.: 06078-935749  
E-Mail: [info@z-mike.de](mailto:info@z-mike.de) [www.zmike.de](http://www.zmike.de)

