

Der Test auf signifikanten Bias nach AIAG Core Tool MSA

Stephan Conrad, TEQ® Training & Consulting GmbH

SigBi oder SickBi?

Nachdem in der letzten PIQ® dargelegt wurde, welche signifikanten Schwachpunkte der Kennwert „ndc“ (number of distinct categories) aufzeigt, wurde von Lesern der Wunsch geäußert, doch auch noch mal Sinn und Unsinn des Tests auf signifikanten Bias in anschaulicher Weise zu beschreiben. Während seitens des ndc von großen Automotive-Zulieferkonzernen schon beklagt wurde, dass der hier angerichtete Schaden in die Millionen gehe, ist die Gefahr beim Test auf signifikanten Bias, hier kurz SigBi genannt, wohl nicht so groß. Denn noch immer haben viele Verantwortliche nicht erkannt, dass das überall verwendete „Verfahren 1“ (C_g/C_{gk}) gar nicht in dem AIAG Core Tool MSA enthalten ist. Statt dessen müsste zum Überprüfen der systematischen Abweichungen der Test auf SigBi durchgeführt werden. Das Verfahren 1 ist weder in der aktuellen 4. Auflage noch in einer vorherigen Auflage enthalten, obwohl es in den Richtlinien von GM und Ford explizit erwähnt und gefordert ist. Im Grund dürften sogar diese beiden Firmen zu den „Erfindern“ des Verfahren 1 gehören, denn schon vor der ersten Auflage der MSA wurde dieses Verfahren 1 (C_g/C_{gk}) in den 80er Jahren in den internen Richtlinien beschrieben.

Verlangt man nun also, eine „MSA“ im allgemeinen Sinne zu machen, ist man gut beraten, das Verfahren 1 als erstes Verfahren durchzuführen. Wird allerdings eine MSA explizit nach „AIAG Core Tool MSA 4. Auflage“ gefordert, gehört dieses Verfahren 1 nicht dazu. Stattdessen ist auf Basis eines sehr ähnlichen Versuches ein Test auf SigBi durchzuführen.

Klären wir zuerst, worum es in den Verfahren geht:

Die Verfahren möchten klären, ob der Messwert unseres zu untersuchenden Messsystems nicht zu weit von „richtigen Wert“ entfernt ist. Der „richtige Wert“ wird durch ein kalibriertes Werkstück oder ein käufliches Normal verkörpert, sofern dieses Normal die Messaufgabe repräsentiert. Das Referenzteil wird nun mehrfach gemessen. Für Verfahren 1 sind 50 Messungen üblich, manchmal werden auch 25 Messungen zugelassen, mal werden 100 Messungen empfohlen. Nach AIAG MSA 4. Auflage sind mindestens 30 Messungen durchzuführen, das Beispiel wird in der 4. Auflage jedoch noch immer mit 10 Messwerten vorgeführt. Aber damit sind die Gemeinsamkeiten auch schon beschrieben.

Im Verfahren 1 werden zwei Aspekte berücksichtigt:

- der Bias, d. h. die mittlere systematische Abweichung, sozusagen die „Fehlanzeige“ oder „Missweisung“, und

- die zufällige Streuung, die dazu führt, dass der angezeigte Wert mal über und mal unter den richtigen Wert liegen würde.

Allgemein bekannte Praxisbeispiele sind z. B.

- für den Bias: Der „gesetzlich voreilende Tacho“ im PKW, der nie eine zu geringe Geschwindigkeit anzeigen darf und deshalb lieber immer etwas voreilt.
- für die zufällige Streuung: Die Geschwindigkeitsanzeige des Navi, die je nach Situation mal etwas über oder unter der tatsächlichen Geschwindigkeit liegt.

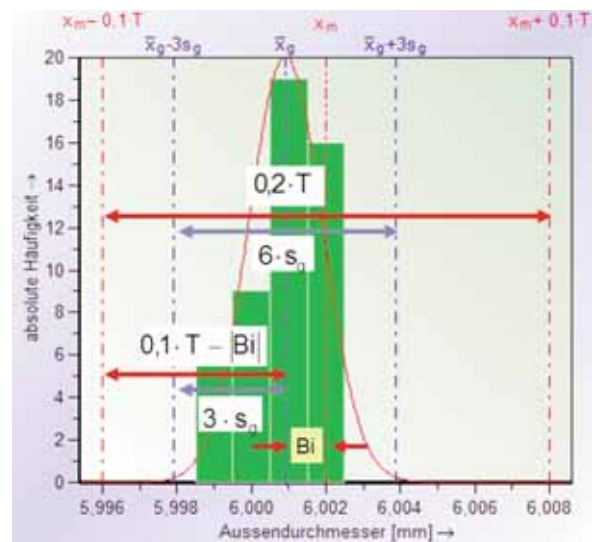


Abbildung 1: \bar{x}_m = Richtiger Wert (Referenzwert);
 s_g = Standardabweichung, x_g = Mittelwert,
 Bi = Bias, T = Merkmalstoleranz

Kurz gesagt führt die Forderung $C_g \geq 1,33$ und $C_{gk} \geq 1,33$ dazu, dass der Bias und die Streuung gemeinsam ausreichend klein sein müssen. Ist die Streuung des Messsystems klein, so darf der Bias etwas größer sein, ist die Streuung größer, dann muss eben der Bias kleiner sein.

Die AIAG MSA 4. Auflage bewertet allerdings nur den Bias, nicht die Streuung. Darüber hinaus geht sie sogar davon aus, dass ein Messsystem gar keinen Bias haben darf. Zwar heißt es in Kapitel 2 (Section D) noch, dass der Bias sich nicht signifikant von Null unterscheiden bzw. nicht den in der Kalibriervorschrift festgelegten Wert überschreiten darf. In der Umsetzung in Kapitel 3 geht es allerdings nur noch um die signifikante Abweichung von Null. Das heißt im Klartext: **Sobald man nachweisen kann, dass ein Messsystem einen Bias hat, gilt dieses Messsystem als nicht geeignet.**

Wo liegt das Problem bei diesem Nachweis? Ein jedes Experiment mit „ein paar Messwerten“ wird einen Bias errechnen. Es wäre wahrlich ein besonderer Zufall, wenn die Messwerte so gezielt ober- und unterhalb des Referenzwertes liegen würden, dass der Mittelwert der Messungen genau auf dem Referenzwert zu liegen kommt. Der Mittelwert liegt im Allgemeinen „immer etwas daneben“, auch bei dem besten Messgerät. Würden wir statt 50 eher 50.000 Messungen machen, dann würde der Mittelwert bei optimalen Messsystemen sicher auch eher auf dem Referenzwert liegen, nicht aber wenn wir auf 15 Messungen reduzieren. Dann spielt der Zufall eine viel zu große Rolle.

Genau hier greift der Test auf SigBi. Ob die errechnete Abweichung noch zufällig oder schon ein deutlicher Hinweis auf einen Bias ist, das ist eine typische Domäne der Statistiker. Üblicherweise führt man dann einen sogenannten t-Test durch, mit dem man z. B. auch prüfen kann, ob Maschinen den gleichen Istwert einhalten, ob der Glasfasergehalt in den Granulat-Chargen gleich ist, ob ein Prozess stabil oder ob eine Optimierung des Werkzeuges gelungen ist.

Nun treffen uns aber die Schwächen eines jeden statistischen Tests:

- Streuen die Messwerte stark, dann lassen sich natürlich kleinere mittlere Abweichungen nur schwer nachweisen. Das heißt umgekehrt, wenn das Messsystem nur wenig streut, dann lässt sich auch ein minimaler Bias schon leicht belegen. Da es in AIAG MSA 4. Auflage aber gar nicht darum geht, wie groß der Bias ist, sondern nur ob der Bias nachgewiesen werden kann, ist ein gering streuendes Messsystem mit kleinem Bias schon nicht mehr geeignet. Sollte Ihr Messsystem glücklicherweise stärker streuen, könnte der Nachweis des Bias nicht möglich sein, und das Messsystem wäre geeignet. Die AIAG MSA hat zwar

in der 4. Auflage erkannt, dass hier ein Problem vorliegt, und warnt davor, dass der Test bei zu großer Streuung nicht durchführbar ist, nennt dafür aber keine Grenzwerte. Das gegenteilige Problem, die „Überempfindlichkeit“ bei zu geringer Streuung, hat allerdings noch keinen Eingang in das Papier gefunden.

- Wie schon oben beschrieben, steigt die Qualität des t-Tests auch mit der Stichprobengröße. Das heißt, mit den in Verfahren 1 üblichen 50 Messungen zeigen nahezu alle Beispiele aus bekannten Firmenrichtlinien, dass der Bias nachweisbar ist. Damit sind die Messsysteme nicht geeignet, trotz ausreichendem C_g/C_{gk} . Wird der Test, wie im Beispiel der AIAG MSA, mit 10 Messwerten durchgeführt, ist das nicht ausreichend, um den Bias nachzuweisen, das Messsystem gilt also als geeignet.

Sie sehen also, besser Finger weg von dieser Methode. Bei einem internationalen Automobilzulieferer wurden die nach Verfahren 1 fähigen Messmittel exemplarisch nochmals mit SigBi bewertet. Der Erfolg war, dass 80% der fähigen Messmittel nicht mehr fähig waren. Nicht weil der Bias zu groß war, sondern weil der kleine Bias signifikant war. Halten wir es mit Ford und GM, die selbst in den eigenen Unterlagen diese Methode ihres Interessenverbandes AIAG ignorieren.

Natürlich liegt mir am Herzen, dass die Anwender auch verstehen, dass das Problem nicht am t-Test sondern an der fehlerhaften Interpretation der Ergebnisse liegt. Schlägt ein statistischer Test an, dann hat er etwas gefunden, das deutlich aus dem Zufallsrauschen hervorsticht. Schlägt er nicht an, dann heißt das nicht, dass da nichts ist, sondern nur, dass er es nicht „mit ausreichender Sicherheit“ nachweisen konnte.

Das eigentliche Problem liegt aber in der Tatsache begründet, dass nur bewertet wird, ob ein Bias vorhan-

Zeichnungswerte		Gemessene Werte		Statistische Werte	
X_m	= 1,99600			\bar{x}_g	= 2,00303
USG	= 1,800	$X_{min,g}$	= 1,983	s_g	= 0,00886
OSG	= 2,200	$X_{max,g}$	= 2,020	$ B = \bar{x}_g - X_m $	= 0,00703
T	= 0,400	R_g	= 0,037	n_{ges}	= 50
		n_{eff}	= 35		
Minimale Bezugsgröße für fähiges Prüfsystem					
$C_g = \frac{0,2 \cdot T}{4 \cdot s_g}$	= 2,26			$T_{min}(C_g)$	= 0,235
$C_{gk} = \frac{0,1 \cdot T - \bar{x}_g - X_m }{2 \cdot s_g}$	= 1,86			$T_{min}(C_{gk})$	= 0,306
%RE	= 0,25%			$T_{min}(RE)$	= 0,0200
Test auf systematische Messabweichung (Bias) Nullhypothese wird zum Niveau $\alpha \leq 0,1\%$ verworfen					
Prüfsystem fähig (RE,U,Cg,Cgk)					
(Q-DAS Messunsicherheit (03/2005)[S]): (Verfahren 1)					

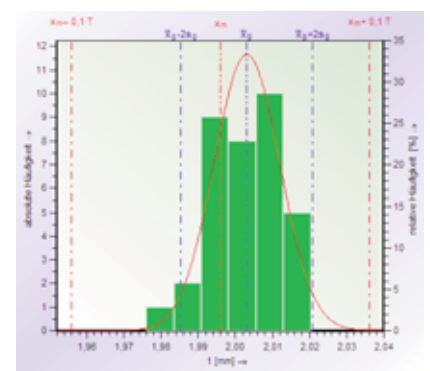


Abbildung 2: $B_i = 7,03 \mu m$ laut signifikantem Bias „n.i.O.“ (Test zu sensibel wegen kleiner Streuung)

Zeichnungswerte		Gemessene Werte		Statistische Werte	
X _m	= 1,99500			\bar{x}_g	= 2,00394
USG	= 1,800	X _{min g}	= 1,966	s _g	= 0,0196
OSG	= 2,200	X _{max g}	= 2,034	Bi = $\bar{x}_g - x_m$	= 0,00894
T	= 0,400	R _g	= 0,068	n _{eff}	= 18
		n _{ges}	= 50		
Minimale Bezugsgröße für fähiges Prüfsystem					
$C_g = \frac{0,2 \cdot T}{4 \cdot s_g}$	= 1,02			T _{min} (C _g)	= 0,522
$C_{gk} = \frac{0,1 \cdot T - \bar{x}_g - x_m }{2 \cdot s_g}$	= 0,79			T _{min} (C _{gk})	= 0,611
%RE	= 0,25%			T _{min} (RE)	= 0,0200
Test auf systematische Messabweichung (Bias)			Nullhypothese wird nicht widerlegt		
Die Anforderungen sind nicht erfüllt (RE, U, C _g , C _{gk})					
(Q-DAS Messunsicherheit (03/2005)[S]): (Verfahren 1)					

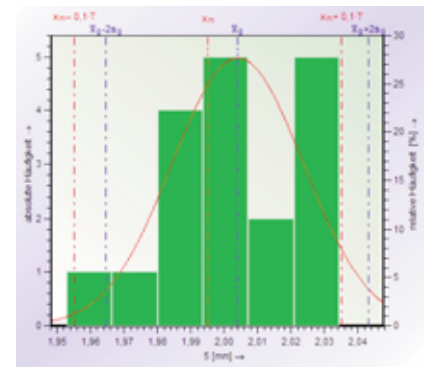


Abbildung 3: Bi= 8,94 µm laut signifikantem Bias „i.O.“ (Test versagt wegen großer Streuung)

den nicht ist, und nicht, wie groß der Bias ist. Hinter dem Glauben, ein Bias wäre an sich schon schlecht, verstecken sich zwei grundlegende Fehlinterpretationen:

- Sobald Sie versuchen, per Normalübertrag ein Messsystem zu justieren, werden Sie höchstwahrscheinlich einen kleinen Bias einjustieren. Bei mehreren Messen des Normals haben wir Streuung gesehen, und beim Justieren des Systems werden wir irgendeinen dieser Messwerte „abnullen“ und somit einen Bias herstellen. Der ist zwar klein und meist unerheblich, aber er ist da.
- Der Referenzwert ist nicht der „wahre Wert“. Er ist nur der „richtige Wert“, die beste Annäherung an den „waren Wert“ im Rahmen der Kalibrierung. Weil ein gutes Kalibrierlabor weiß, dass auch eine Kalibrierung Messunsicherheiten enthält, kann es Messunsicherheiten nach GUM (ISO/IEC Guide 98) angeben. Sollte sich der Bias also im Rahmen der Messunsicherheit der Kalibrierung des Referenzteils bewegen, ist er kein Mangel des Messsystems. Dahinter liegen Unmengen weiterer Risiken zur Frage der korrekten Kalibrierung bezogen auf die konkrete Messaufgabe und der Auswahl des Referenzteiles.

Hier ein Beispiel für den Fehlentscheid des Tests auf signifikanten Bias. Im ersten Fall (Abb. 2) ist der relativ kleine Bias auf Grund recht geringer Streuung nachweisbar. Obwohl C_g und C_{gk} auf „i.O.“ entscheiden, führt die Nachweisbarkeit des Bias zur Ablehnung.

Im zweiten Fall (Abb. 3) ist sowohl Bias als auch Streuung größer. Damit ist das Messsystem nach C_g/C_{gk} als „n.i.O.“ zu bewerten, der SigBi kann aber den Bias nicht mehr nachweisen, wodurch das Messsystem fälschlicherweise als „i.O.“ bewertet wird.

Schlussfolgerung:

Auch wenn Sie sich MSA-konform verhalten wollen, lassen Sie die Finger von diesem Test. Er mag zusätzliche Informationen liefern, ist aber nicht geeignet, an Stelle des Verfahren 1 über die Eignung eines Messsystems zu entscheiden. Bleiben Sie beim allseits anerkannten und üblichen Verfahren 1, auch wenn es nicht in der AIAG MSA erwähnt wird. Interpretieren Sie eine MSA-konformität dahingehend, dass Sie das Verfahren 1 mit „6s“ berechnen, weil ein Vergleich von Streubreiten nach der Einleitung der MSA immer mit „6s“ erfolgen sollte. Ob das „richtiger“ ist als die „4s“ vieler anderer Firmenrichtlinien ist, wollen wir hier nicht diskutieren. Es ist einfach so nach AIAG MSA 4. Auflage.

P.S.: Diese Anwendung des t-Tests steckt gleich zweimal in der Linearitätsstudie der AIAG MSA, mit exakt den gleichen Fehlinterpretationen. Aber das ist eine weitere Geschichte ...

Literatur

- [1] AIAG - Chrysler Corp., Ford Motor Co., General Motors Corp. Measurement Systems Analysis, Reference Manual 4. Auflage, Michigan, USA, 2010.
- [2] Robert Bosch GmbH Schriftenreihe „Qualitätsmanagement in der Bosch-Gruppe“ Heft 10, Fähigkeit von Mess- und Prüfprozessen, Ausgabe 5, Stuttgart, 2010.
- [3] General Motors Co. - GM Powertrain SP-Q-MSS - Measurement Systems Specification Version: G2.0, Detroit, 2010.
- [4] Ford Motor Co. - Ford Powertrain PTP 07-116 Initial Process Study for Machine Acceptance, Neuauflage 2009.