ARM, ANOVA und der ganze Rest …

Dipl.-Ing. Stephan Conrad, TEQ® Training & Consulting GmbH

In unseren Seminaren sehen wir in letzter Zeit eine zunehmende Verunsicherung bezüglich der Vorgehensweisen und Interpretationen der Ergebnisse bei dem allseits bekannten „Verfahren 2“, „R&R-“ oder „GRR-Verfahren“.

Vorbei sind die Zeiten, als man die Berechnungsmethoden noch mit einer kurzen Skizze und ein paar Spannweiten erklären konnte, und sich der Zauber der Wahrheit in zwei K-Faktoren versteckte. Spätestens seit der vierten Auflage der AIAG-MSA ist die ANOVA-Methode durchweg die empfohlene Methode und ARM nur noch im „Notfall“ erlaubt.

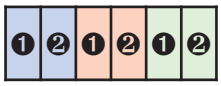
Begründet wird dies fast durchweg mit der Tatsache, dass ARM nur die Komponenten EV und AV (Wiederhol- und Vergleichspräzision) ermitteln kann, während ANOVA auch eine Wechselwirkung IA (Interaction) erkennen kann. Das reicht, um in die Diskussion einzusteigen, eventuell auch, um sie zu beenden, aber was versteckt sich wirklich dahinter?

Vorab ein kurzes Schema zu Verfahren 2: Typischerweise messen 2-3 Prüfer zehn unterschiedliche Werkstücke je 2-3 mal. Das ist ein ganz normaler vollfaktorieller Versuchsplan, was wir in den Seminaren zu MSA aber nur ganz leise sagen.

Ein Beispiel: 3 Prüfer messen 10 Teile je 2 mal:

Prüfer A		Prüfer B		Prüfer C	
Trial 1	Trial 2	Trial 1	Trial 2	Trial 1	Trial 2
1-A1	1-A2	1-B1	1-B2	1-C1	1-C2
2-C1	2-C2	2-B1	2-B2	2-C1	2-C2
3-C1	3-C2	3-B1	3-B2	3-C1	3-C2
4-C1	4-C2	4-B1	4-B2	4-C1	4-C2
5-C1	5-C2	5-B1	5-B2	5-C1	5-C2
6-C1	6-C2	6-B1	6-B2	6-C1	6-C2
7-C1	7-C2	7-B1	7-B2	7-C1	7-C2
8-C1	8-C2	8-B1	8-B2	8-C1	8-C2
9-C1	9-C2	9-B1	9-B2	9-C1	9-C2
10-C1	10-C2	10-B1	10-B2	10-C1	10-C2

Dieses Schema stellen wir verkürzt so dar:

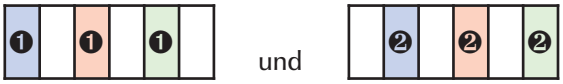


Was macht ARM?

Mit dieser Methode versucht man, die Fragen zu beantworten:

- Wie gut kann ein Prüfer seine Messung wiederholen?

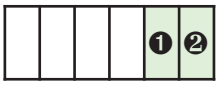
  Dazu vergleicht man die mittleren Spannweiten zwischen

  und nennt das Ergebnis „Wiederholpräzision“
- „Wie vergleichbar sind die Messungen der Prüfer untereinander?“

  Dazu vergleicht man die Spannweite zwischen

  und

  und nennt das Ergebnis „Vergleichspräzision“.

Weitere Einflüsse bleiben unberücksichtigt, weil einfach niemand danach fragt.

In einem nächsten Schritt werden die Spannweiten in Standardabweichungen umgerechnet. Da jeder weiß, dass bei gleicher Standardabweichung verschiedene Stichproben zufallsbedingt unterschiedliche Spannweiten haben können, nimmt man aus einer sogenannten w-Verteilung einfach die Standardabweichung, bei der die hier erhaltene Spannweite am wahrscheinlichsten ist. Das ist eine allseits akzeptierte Methode, die dennoch etwas Mut verlangt. Oder würden Sie grundsätzlich davon ausgehen, dass Sie bei einer Lotterie immer den wahrscheinlichsten Wert ziehen? Das wäre ja schade, denn Sie würden nie etwas gewinnen.

Was macht ANOVA?

ANOVA geht einen etwas anderen Weg. Zuerst wird „alle Streuung“ der 60 Werte in einen Topf geworfen. Dann beginnt das Fischen:

- Aus den Mittelwerten der 10 Teile kann man die Teilestreuung berechnen und abziehen
- Aus den Mittelwerten der Prüfer kann man die Vergleichbarkeit berechnen und abziehen
- Die Wechselwirkung lassen wir an dieser Stelle noch außen vor …
- Was übrig bleibt, das muss wohl die Wiederholbarkeit sein.

Die Wiederholstreuung ist also „der Rest“, der auch die Einflüsse enthält, die bei ARM unter den Tisch fallen würden. Das ist ein prinzipieller Unterschied ANOVA zu ARM.

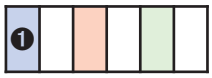
Was ist EV und AV?

Das sind sehr unglückliche Benennungen

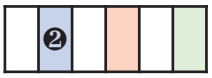
- EV = Equipment Variation = Wiederholpräzision
- AV = Appraiser Variation = Vergleichspräzision

Spätestens mit der Übersetzung der Begriffe beginnen wir zu straucheln. EV heißt „Equipment Variation“ und wäre mit „Gerätestreuung“ zu übersetzen. Leider weisen wir damit dieser Streukomponente eine Ursache zu, ohne dass es dafür einen Grund gäbe. Ähnlich ist es mit AV, der „Appraiser Variation“, der Bedienerstreuung. Dass die Messwerte der Bediener nicht vergleichbar sind, muss nicht am Bediener liegen.

Ein Beispiel: In einem unserer Seminare hatten die Teilnehmer die Aufgabe, die Höhe einer Zentrierbuchse aus Aluminium mit einem Messschieber zu messen. Die Teilnehmer dürfen dabei gerne Fehler machen, bekanntlich lernt man daraus. Also durften sie auch entscheiden, dass die Bediener ihre beiden Durchgänge nacheinander abarbeiteten. Der erste Bediener schnäbelte das Aluminiumteil mit den Schneiden der Messschenkel an und konnte deshalb seine Messungen nicht wiederholen. Definitiv ein Bedienerfehler, aber



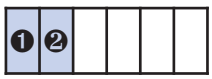
stimmte nicht mit



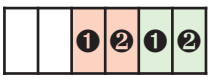
überein, somit eine zu große Equipment Variation EV im Sinne Wiederholpräzision. Diese vom Bediener verursachte Streuung „Gerätestreuung“ zu nennen, wäre definitiv falsch.

Nachdem der erste Bediener den beiden anderen zugesehen hatte, die beide den Messschieber völlig korrekt benutzten und weitaus bessere Ergebnisse erzielten, wurden die Ergebnisse diskutiert. Daraufhin wollte der erste Bediener seine Messungen wiederholen. Ein Griff zum Messschieber und er legte los - ohne den Messschieber zu nullen, der sich ja in der Zwischenzeit dejustiert hatte.

Nun konnte er seine Ergebnisse wunderbar wiederholen, er hat also klar dazugelernt, aber seine Werte waren auf Grund der Gerätedrift nicht mehr mit denen der anderen vergleichbar. Das heißt



stimmte nicht mit



überein, somit ein zu großes AV im Sinne von Vergleichspräzision. Diese von der Drift des Gerätes verursachte Streuung „Bedienerstreuung“ zu nennen, wäre ebenso falsch.

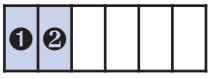
Dennoch muss man sich mit den Kürzeln wohl arrangieren, denn auch der VDA Band 5 hat diese Abkürzungen übernommen.

Was ist die Wechselwirkung IA (InterAction)?

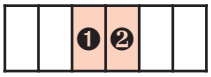
Auch hier muss man mit der Benennung vorsichtig umgehen. Dieser hier sehr allgemein verstandene Begriff darf nicht mit der Wechselwirkung verwechselt werden, die in der klassischen statistischen Versuchsplanung (Design of Experiment) benutzt wird. Dort unterscheidet man zweifach, dreifach oder andere Mehrfachwechselwirkungen. Eine Zweifach-WW ist zum Beispiel gegeben, wenn man eine definierte Menge Luft bei unterschiedlichen Temperaturen beigibt. Die Volumenveränderung hängt dabei von der Temperatur ab, das Ganze lässt sich sogar mit einer Formel beschreiben. Diese Wechselwirkung ist systematisch und kann produktiv genutzt werden, um ein Prozessoptimum zu finden.

In der Messsystemanalyse nimmt man diesen Begriff aber, wenn sich die Streuung eines oder mehrere der zehn Teile von der zu erwartenden Vergleichspräzision signifikant unterscheidet. Die Gründe dafür können vielfältig sein, oftmals zufällig, aber fast immer störend. Die Wechselwirkung ist somit keine Eigenschaft oder „hinzunehmendes Streumaß“ des Messsystems, wie z. B. die Wiederholpräzision, sondern eher ein Hinweis auf einen noch unerkannten Störeinfluss im System, dessen tatsächliche Auswirkungen möglicherweise noch gar nicht vollständig erkannt wurden.

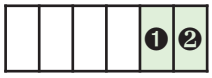
Ein Beispiel: In unserem Versuch haben die drei Bediener die zehn Teile je zweimal gemessen, also sechs Messungen pro Teil. Daraus ergibt sich eine Vergleichspräzision aus den mittleren Abweichungen von



,



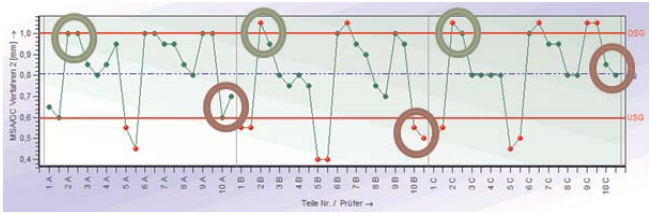
und



Nun prüft man jedes einzelne der zehn Teile, ob die Streuung auch dem erwarteten mittleren Abweichungen entsprechen. Zeigt sich dabei, dass einzelne Teile völlig anders streuen, unterscheidet sich also die „Prüferstreuung“ bei diesen einzelnen Teilen von AV, dann ist vermutlich bei diesen Prüfer-Teil-Kombinationen ein weiterer Einfluss zu sehen, der vereinfacht Wechselwirkung genannt wird.

Ein Beispiel:

Drei Prüfer A, B, und C messen 10 Teile je zweimal.



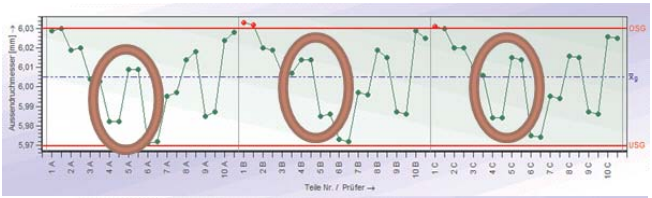
Es ist gut zu erkennen, dass die Unterschiede beim zehnten Teil enorm sind, wesentlich größer als bei den anderen Teilen (exemplarisch ist das 2. Teil noch hervorgehoben). Im Abweichungsdiagramm, die dem die Abweichungen der einzelnen Messungen zum Mittelwert des jeweiligen Teils dargestellt ist, kann man das noch deutlicher erkennen.

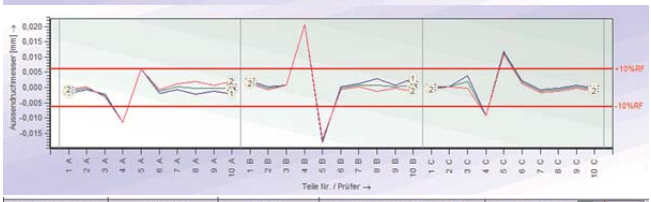
Das Ergebnis zeigt eine hohe Wechselwirkung IA.

Wiederholpräzision	0,00129	0,0359	EV	=	0,148 ≤ **0,185** ≤ 0,247	%EV	= 46,29%
Vergleichspräzision	0,000912	0,0302	AV	=	0,000 ≤ **0,156** ≤ 1,117	%AV	= 38,89%
Wechselwirkung	0,00223	0,0473	IA	=	0,131 ≤ **0,243** ≤ 0,412	%IA	= 60,87%
Prüfsystemstreuung	0,00444	0,0666	R&R	=	0,307 ≤ **0,343** ≤ 1,159	%R&R	= 85,79%

Ursache dafür könnten Handhabungsfehler oder auch Probleme mit der Spannvorrichtung sein.

Ähnliches passiert, wenn ein Prüfer zwei Teile vertauscht.

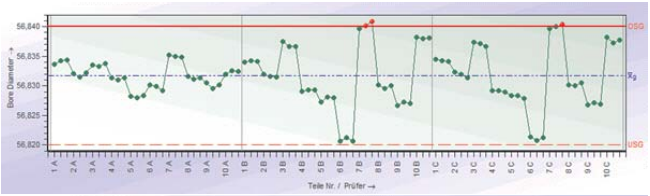


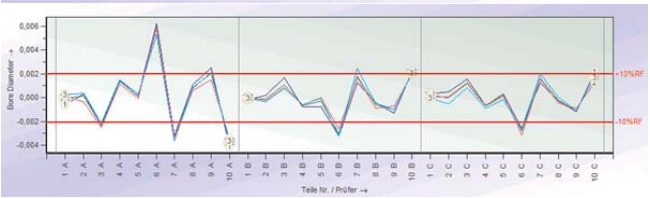


Wiederholpräzision	0,00000175	0,00132	EV	=	0,00545 ≤ **0,00682** ≤ 0,00911	%EV	= 11,36%
Vergleichspräzision	0,000	0,000	AV	=	0,000 ≤ **0,000** ≤ 0,0294	%AV	= 0,00%
Wechselwirkung	0,0000615	0,00784	IA	=	0,0268 ≤ **0,0404** ≤ 0,0634	%IA	= 67,33%
Prüfsystemstreuung	0,0000833	0,00795	R&R	=	0,0390 ≤ **0,0410** ≤ 0,0505	%R&R	= 68,29%

Das ist zwar ein trivialer Fehler im Ablauf und kein Problem des Messsystems, sollte aber auf jeden Fall erkannt werden.

Auch wenn ein Prüfer (hier A) durch Handhabungsfehler, Justierfehler oder Verschleiß eine grundlegend andere Teilestreuung misst, schlägt sich das in der Wechselwirkung nieder.





Wiederholpräzision	0,000000114	0,000337	EV	=	0,00147 ≤ **0,00174** ≤ 0,00212	%EV	= 8,69%
Vergleichspräzision	0,000	0,000	AV	=	0,000 ≤ **0,000** ≤ 0,000	%AV	= 0,00%
Wechselwirkung	0,00000596	0,00244	IA	=	0,00890 ≤ **0,0126** ≤ 0,0192	%IA	= 62,89%
Prüfsystemstreuung	0,00000608	0,00246	R&R	=	0,0121 ≤ **0,0127** ≤ 0,0123	%R&R	= 63,49%

Es wird also deutlich, dass die Wechselwirkung nicht ein einfach zu definierendes Phänomen beschreibt, sondern sehr deutlich darauf hinweist, dass in den Daten Störeinflüsse zu sehen sind, die nicht einer mittleren Wiederhol- oder Vergleichpräzision zuzuordnen sind. Weitere Untersuchungen sind notwendig, um die Ursachen zu erkennen. Es wäre auch falsch, diese Wechselwirkung als „übliche Streuung des Messsystems“ einfach hinzunehmen. Vielmehr ist ein Hinweis darauf, dass weitere Störungen vorliegen, die vor Freigabe des Messprozesses untersucht werden müssen.

Aber, um das nochmals zu betonen, auch EV und AV dürfen keinesfalls ungesehen dem Gerät oder dem Bediener zugewiesen werde, nur weil sie mit Equipment und Appraiser Variation übersetzt werden!

Also doch besser ANOVA?

Mit Sicherheit ja. ARM mag erlaubt sein, wenn man weit weg jeglicher Zivilisation im Dschungel mit Bleistift, Papier und einem Solartaschenrechner auf die Schnelle eine Messsystemanalyse durchführen muss. (Der Autor bittet um Verzeihung für diese polemisierenden Äußerungen, es macht aber so viel Spaß). Aber selbst dafür gäbe es sogenannte ANOVA-Tabellen, mit denen die Qualitätsingenieure schon vor Erscheinen des PCs gearbeitet haben. Die einzige vertretbare Entschuldigung ist, dass man das eigene Unternehmen ähnlich einem Ozeandampfer nur langsam auf neuen Kurs bringen kann. Das bedeutet aber, den Kurswechsel jetzt einzuleiten, damit das Unternehmen nicht in absehbarer Zeit das Schicksal der Titanic teilt.

Um die vorne geschilderte Vorgehensweise von ANOVA nun noch zu ergänzen:

- Aus den Mittelwerten der 10 Teile kann man die Teilestreuung berechnen und abziehen
- Aus den Mittelwerten der Prüfer kann man die Vergleichpräzision berechnen und abziehen
- Falls die Streuung mindestens eines Teils von der mittleren Vergleichpräzision signifikant abweicht, dann diese Wechselwirkung berechnen und abziehen
- Was übrig bleibt, das muss wohl die Wiederholpräzision sein.

Dieses „Falls … signifikant …“ weist auch darauf hin, dass die errechnete Wechselwirkung mit Hilfe eines F-Tests bewertet wird. Ist das Ergebnis so gering, dass es zufällig sein könnte, dann weist man den Wert nicht aus, sondern packt diese zufälligen Reste in die Wiederholpräzision.

Ein weiterer Vorteil der ANOVA Methode ist die Erweiterbarkeit. Nach VDA Band 5 kann der Versuch zusätzliche Einflussgrößen enthalten, wie z. B.

- Unterschiedliche Messeinrichtungen
- Unterschiedliche Messorte am Objekt
- Unterschiedliche Messmethoden

Diese können per ANOVA aufgeschlüsselt und getrennt ausgewiesen werden. Eine jede untersuchte Einflussgröße erhöht natürlich die Anzahl der benötigten Messungen. Damit kommen wir direkt in die Methoden der Versuchsplanung, aber das ist eine andere Geschichte, die wir an einem anderen Tag erzählen wollen …