Mess-System-Analyse diskret

Voraussetzung und verwandte Themen

Für diese Beschreibungen sind Grundlagen der Statistik und insbesondere der Hypothesentests vorteilhaft. Weiterführende Themen sind

www.versuchsmethoden.de/Mess-System-Analyse.pdf

www.versuchsmethoden.de/Hypothesentests.pdf

Einführung

Als Mess-System-Analyse, kurz MSA, bezeichnet man die Analyse der Fähigkeit eines Messmittels und des Mess-Prozesses. Unter einer diskreten Messung versteht man ein Ergebnis auf Basis gut/schlecht, 0 / 1 usw. Wenn möglich ist die Messung durch stetige Messwerte vorzuziehen. Nur wenn es keine andere Möglichkeit gibt, ist die diskrete Messung anzuwenden. Das ist z.B. in den meisten Fällen eine subjektive Beurteilung. Die nächste Stufe ist eine ordinal skalierte Messung ab 3 Stufen, z.B. die Bewertung nach Schulnoten.

Ziel und Nutzen

In einer diskreten MSA soll für Untersuchungen und Prozesse aufzeigen, dass das Mess-System oder der Prüfer besser ist, als ein zufällig zustande gekommenes Ergebnis.

Grundlagen

Im Folgenden werden diese Verfahren beschrieben:

- **Gage R&R für diskrete Merkmale**
- **Kappa Methode (Fleiss & Cohens)**
- **Bowker-Verfahren**
- **Kendalls Konkordanz (ordinale Daten)**

Gage R&R für diskrete Merkmale

Im Verfahren Gage R&R für diskrete Merkmale lässt man mehrere Prüfer jeweils zweimal verschiedene Teile beurteilen. Das könnten z.B. Prüfungen an Teilen sein, die entweder intakt, oder fehlerhaft sind. Die Prüfer dürfen nicht wissen, welches Teil sie vor sich haben und die Reihenfolge muss zufällig sein. Zu beachten sind die Hinweise zu den Prüfobjekten in einem späteren Kapitel.
Gibt es innerhalb eines Prüfers, oder zwischen verschiedenen Prüfern Abweichungen, so werden diese gezählt.

Im Verfahren **mit Referenzwert** sind den Prüfern zufällige Muster vorzulegen, von denen sie die Einstufung ebenfalls nicht kennen. Es werden mindestens 30 zufällige Muster empfohlen, die jeder Prüfer zweimal zu bewerten hat. Die Ergebnisse werden tabellarisch aufgetragen:

Mess-System-Analyse diskret

	Referenz	Prüfer A		Prüfer B		Prüfer C	
1	gut	gut	gut	gut	gut	gut	gut
2	schlecht	schlecht	schlecht	schlecht	schlecht	schlecht	gut
3	gut	gut	gut	gut	gut	gut	gut
4	gut	gut	gut	gut	gut	gut	gut
5	gut	schlecht	schlecht	schlecht	schlecht	schlecht	schlecht
6	gut	gut	gut	gut	gut	gut	gut
7	gut	gut	gut	schlecht	schlecht	gut	gut
8	schlecht	schlecht	schlecht	schlecht	schlecht	schlecht	schlecht
9	schlecht	schlecht	schlecht	schlecht	schlecht	schlecht	schlecht

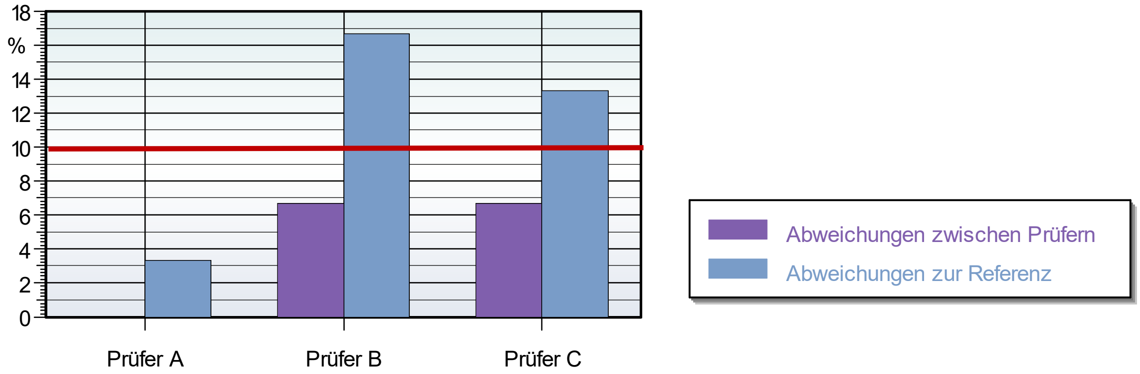
Abweichungen:

innerhalb des Prüfers

zur Referenz

zwischen den Prüfern

Eine Anforderung kann sein, dass das Verhältnis unterschiedlicher Ergebnisse zu der Anzahl Teile soll nicht größer als 5% oder 10% sein soll.
In diesem Beispiel haben Prüfer B und C mehr als 10% Abweichungen zur Referenz:



Kappa Methode

In der sogenannten **Kappa Methode** werden nicht nur, wie in der vorherigen Darstellung, die Abweichungen gezählt. Es stellt sich die Frage welcher Anteil nur zufällig möglich gewesen wäre. Die folgende Kenngröße Kappa berücksichtigt dies:

$$\kappa = \frac{p_o - p_c}{1 - p_c}$$

p_o : beobachtete Übereinstimmung
pc : zufällig mögliche Übereinstimmung

Kappa stellt die Übereinstimmung zwischen verschiedenen Prüfern, oder zu einer Referenz dar, abzüglich einer zufälligen Übereinstimmung. Die Anforderung an Kappa ist in der Regel wie folgt gestaffelt:

$\kappa < 0{,}7$ ⇨ nicht fähig
$0{,}7 \leq \kappa < 0{,}9$ ⇨ bedingt fähig
$0{,}9 \leq \kappa < 1{,}0$ ⇨ fähig

Man kann die Übereinstimmung jedes Prüfers mit sich selbst ermitteln, also ob jedes Objekt/Teil immer gleich bewertet wird (**Wiederholbarkeit**). Dabei dürfen die Prüfer nicht wissen, welches Teil gerade bewertet wird. Das nächste ist die Überprüfung, wie sich die Prüfer untereinander verhalten (**Reproduzierbarkeit**). Für diese beiden Fragestellungen wendet man das sogenannte **Fleiss-Kappa** an. Für die Fragen nach der Übereinstimmung jedes Prüfers mit einer Referenz ist der **Cohen's Kappa** relevant.
Die auszuwertende Tabelle ist im Grunde genauso aufgebaut, wie bei der Methode Gage R&R diskret. Hier wird jedoch anstelle der Kennzeichnung gut/schlecht 0 und 1 verwendet. Weiterhin können hier auch 3 Wiederholungen gemacht werden.

In folgendem Beispiel wurden 30 Teile von 3 Prüfern bewertet:

untere Vertrauensgrenze — Standardabweichung — analog Normalv. z = κ/SE — Irrtumswahrsch.

	korrekt	%	Fleiss k	95% CI	SE	z	p-val
Prüfer A	29	96,7	0,955	0,643	0,10541	9,06	0,000
Prüfer B	28	93,3	0,910	0,597	0,10541	8,63	0,000
Prüfer C	29	96,7	0,954	0,642	0,10541	9,05	0,000

Am Ende der Tabelle wird die Nullhypothese getestet, dass sich die Übereinstimmungen signifikant von einem zufälligen Ergebnis unterscheiden (p-value < 0,05). Zu beachten ist hier, dass diese Statistik nicht mit den dargestellten Grenzen von Kappa zu vergleichen ist! Nur der eigentliche Kappa-Wert liefert die Entscheidung, ob die Messung fähig ist.

Eine ähnliche Berechnung ist in der AIAG-Richtlinie Measurement System Analysis (4th Edition) beschrieben. Im Gegensatz zu der hier gezeigten Variante können dort nur die Beziehungen jeweils zwischen 2 Prüfern paarweise angewendet werden.

Bowker-Verfahren

Im sogenannten Bowker-Verfahren gibt es ebenfalls die Beurteilungen gut/schlecht. Dabei können die Prüfer gleiche Ergebnisse haben, unterschiedliche, oder uneinheitlich sein. Mindestens 40 verschiedene Prüfobjekte werden von 2 Prüfern je 3mal geprüft. Jedes der 40 Ergebnisse wird in drei Klassen aufgeteilt:

Klasse 1 : alle 3 Wiederholungen ergaben das Ergebnis gut
Klasse 2 : innerhalb der 3 Wiederholungen abweichende Ergebnisse
Klasse 3 : alle 3 Wiederholungen ergaben das Ergebnis schlecht.

Das Ergebnis wird in Form einer Kreuztabelle dargestellt:

		Prüfer B		
		gut	gemischt	schlecht
Prüfer A	gut	7	3	0
	gemischt	10	1	7
	schlecht	2	5	5

Für die Prüfung wird mit Hilfe der χ^2-Verteilung die Symmetrie getestet. Die Nullhypothese ist, dass die zu erwartenden Häufigkeiten $n_{j,i}$ in der Tabelle symmetrisch zur Hauptdiagonalen liegen. Diese Nullhypothese wird abgelehnt, wenn die folgende Testgröße mit 3 Freiheitsgraden größer ist, als der Wert aus der χ^2-Verteilung

$$\chi^2 = \sum_{i>j} \frac{(n_{i,j} - n_{j,i})^2}{n_{i,j} - n_{j,i}}$$

Weitere Informationen sind dem VDA-Band 5 /31/ zu entnehmen.

Kendalls Konkordanz (ordinale Daten)

Mit Hilfe der folgenden Prüfung nach Kendall ist eine MSA für ordinal skalierte Daten möglich. Das könnte z.B. eine subjektive Prüfung mit einem definierten aufsteigenden Bewertungsindex sein, wie etwa Schulnoten.

Hierfür werden mindestens 30 Teile, bzw. Prüfobjekte benötigt. Dabei ist eine ideale Verteilung der Eigenschaften wichtig (realistische Streubreite). Alle Bewertungsnoten sollten, soweit möglich, gleichhäufig vorkommen. Auch hier gilt natürlich, dass die Prüfer zufällige Teile in zufälliger Reihenfolge bekommen, deren Eigenschaft sie nicht kennen.
Die Auswertung erfolgt über die Rangfolge der Bewertung. Der nichtparametrische Test mit Hilfe des sogenannte Kendall's Konkordanzkoeffizient W ist dabei wie folgt definiert:

$$W = \frac{12 \sum_{i=1}^{n} R_i^2 \; - \; 3\, m^2 n\, (n+1)^2}{m^2 n\, (n^2+1) - m \sum_{j=1}^{m} T_j}$$

mit

$$T_j = \sum_{k=1}^{g_j} (t_{jk}^3 - t_{jk})$$

R_i Rang des bewerteten Teils

t_{jk} Anzahl Teile mit gleichem Rang (Rangbindungslänge)

g_j Zahl der Rangbindungen bei Prüfer j

n Anzahl Teile

m Anzahl Prüfer

k Anzahl Rangplätze

Der Konkordanzkoeffizient W ist χ^2-verteilt mit dem Freiheitsgrad f = n – 1.

$$\chi^2 = \mathrm{m\,(n-1)\,W}$$

worüber sich der p-Value bestimmen lässt. Hiermit wird die Nullhypothese geprüft, dass es keine Übereinstimmung der Prüfer gibt. Ist der p-value < α (0,05), so ist die Übereinstimmung ausreichend.

Das Ranking wird entsprechend einer aufsteigenden Zahlenbewertung gebildet. Folgendes Beispiel zeigt eine sortierte Bewertung. Da 3mal die Bewertung 3 vorkommt, ist hier 3mal die mittlere Rangzahl 4 zu vergeben. Würde es die 3 nur 2mal geben, so wären die Rangzahlen … 2; 3,5; 3,5; 5; …
Die Anzahl der Rangbindungen ist hier g_j = 1 und die Rangbindungslänge t_{jk} = 3. Somit ergibt sich

Nr.	Bewer-tungszahl	Ranking
1	1	1
2	2	2
3	3	4
4	3	4
5	3	4
6	4	6

$$T_j = \sum_{k=1}^{g_j} (t_{jk}^3 - t_{jk}) \;=\; \sum_{k=1}^{1} (3_1^3 - 3) \;=\; 24$$

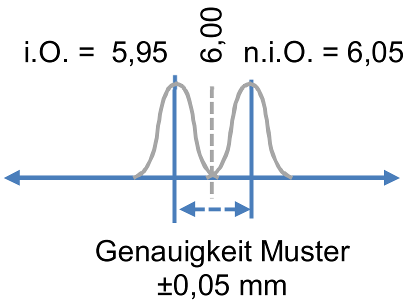
Dieser Test lässt relativ große Abweichungen der Prüfer zu, bevor der p-value die 0,05 erreicht. Es macht Sinn eine zusätzliche Anforderung auf eine konkrete Anzahl Übereinstimmungen zu definieren, z.B. mindestens 10% der Teile müssen alle Prüfer gleich bewertet haben.
Die gleiche Rechnung kann zwischen den Prüfern, aber auch gegenüber Wiederholungen jedes Prüfers angewendet werden. Weiterhin ist ein Test gegenüber einer Referenzbewertung möglich. Referenzwerte liegen allerdings selten bei subjektiven Bewertungen vor.

Anmerkungen zu den Prüfobjekten

Für die genannten Verfahren werden Prüfobjekte, Muster oder Teile benötigt, bei denen die Frage besteht, wie deren Beschaffenheit sein soll.

Es macht keinen Sinn ein Muster so schlecht zu machen, dass es unter allen Umständen immer als schlecht erkannt wird. Ebenso ist es bei guten Mustern. Alle Muster sollten knapp an der Grenze zwischen gut und schlecht liegen, damit die Möglichkeit einer Falschbeurteilung auch „beobachtet“ werden kann.

Dabei ist die physikalische Eigenschaft so zu wählen, dass trotzdem eine eindeutige Zuordnung noch möglich ist. Beispiel: Es besteht die Aufgabe mit Hilfe einer selbst hergestellten Lehre zu testen, ob eine Bohrung in einem Blech (Muster) einen Maximaldurchmesser von 6,0 mm hat. Geht die Lehre in die Bohrung, so gilt diese als zu groß und somit als n.i.O. (schlecht). Die Frage ist, mit welcher Genauigkeit die Bohrung der Muster hergestellt werden kann. Wird als Messgerät z.B. ein Messschieber verwendet, so ist die Genauigkeit hier 0,05mm. Zur „sicheren“ Unterscheidung müsste somit eine i.O. Bohrung 5,95mm und eine n.i.O. Bohrung 6,05mm haben.

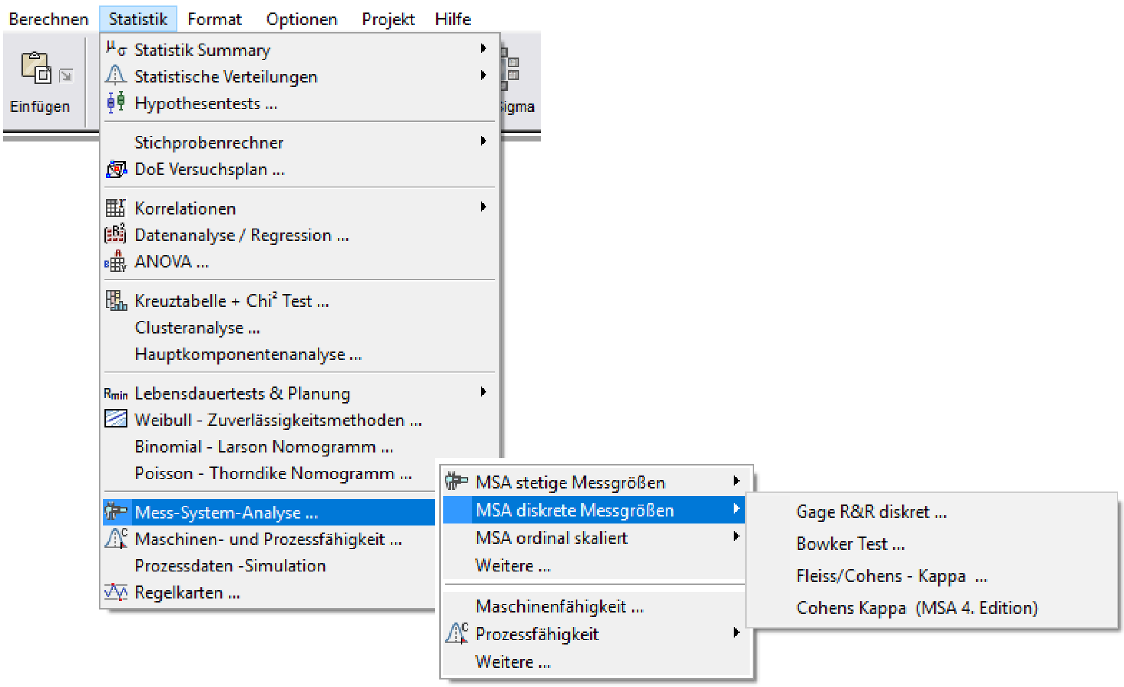


Legt man nun die n.i.O. Muster auf 6,05, so ist die Wahrscheinlichkeit ungewollt 6,00 oder weniger zu erhalten 2,5% (95% entsprechen ±0,05 mm). Dafür gibt es aber auch auf der i.O. Seite 2,5%, die unbeabsichtigt n.i.O. sein könnten.

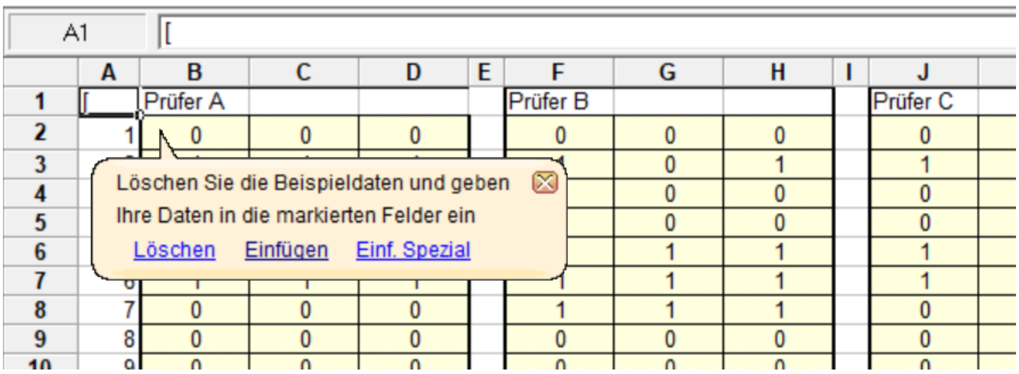
Weiterhin geht die Genauigkeit der Lehre mit ein. Während sich die Fehler der verschiedenen Muster über deren Anzahl verteilen, gibt es nur eine Lehre, die immer den gleichen Fehler hat. Wenn nun die Aufteilung der Muster zu 50% i.O. und 50% n.i.O. ist, verschiebt der Fehler der Lehre den Anteil beobachteter i.O.- und n.i.O.-Muster einseitig. Die Genauigkeit der Lehre sollte deshalb deutlich besser sein.

Anwendung in Visual-XSel 15

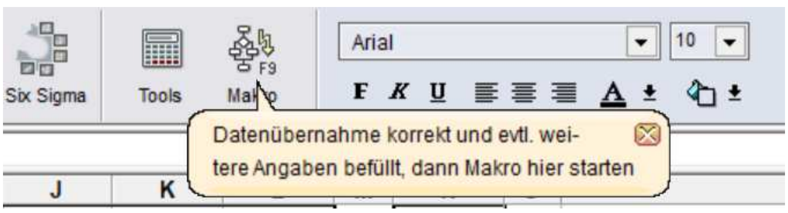
www.crgraph.de



Nach Auswahl des entsprechend gewünschten Templates sind über den Link Einfügen Daten aus der Zwischenablage einzufügen, oder per Hand einzugeben:



Evtl. sind weitere Felder (gelb hinterlegt) zu befüllen. Danach ist das Makro mit F9, zu starten, oder über die Ikone Makro.



Literatur

/27/ Measurement System Analysis MSA 4
Fourth Edition 7/2010
ISBN# 978-1-60-534211-5

/28/ Leitfaden zum “Fähigkeitnachweis von Messsystemen”
Stand September 2002,
Version 2.1 D/E

/29/ Statistische Methoden der Qualitätssicherung
Horst Rinne, Hans-Joachim Mittag
Hanser, München/Wien 2002, ISBN 3-446-15503-1.

/30/ Handbuch Qualitätsmanangement
Masing
Hanser, München, ISBN 978-3-446-40752-7

/31/ VDA Band 5 Prüfprozesseignung
Verband der Automobilindustrie e.V. VDA – QMC
2. Auflage, Frankfurt 2010, ISSN 0943-9412

Literatur

Taschenbuch der statistischen Qualitäts- und Zuverlässigkeitsmethoden

Die wichtigsten Methoden und Verfahren für die Praxis.

Beinhaltet statistische Methoden für Versuchsplanung & Datenanalyse, sowie Zuverlässigkeit & Weibull.

- Statistische Verteilungen und Tests & Mischverteilungen
- Six Sigma Einführung und Zyklen
- Systemanalysen Wirkdiagramm, FMEA, FTA, Matrizen-Methoden
- Shainin- und Taguchi-Methoden
- Versuchsplanung DoE, D-Optimal
- Korrelations- und Regressionsverfahren
- Multivariate Datenauswertungen
- Prozessfähigkeit – Messmittelfähigkeit MSA 4 und VDA 5
- Regelkarten
- Toleranzrechnung und Monte-Carlo-Simulation
- Statistische Hypothesentests
- Weibull und Lebensdaueranalysen
- Stichprobengröße

190 Seiten, Ringbuch

ISBN: 978-3-00-043678-9