



Eignungsnachweise bei attributiven Prüfprozessen

Teil 2: Eignungsnachweise für diskrete Merkmale

Dieses Dokument ist Teil 2 einer 3-teiligen Whitepaper-Reihe zu attributiven Prüfprozessen. Die Reihe stellt die verschiedenen Methoden zum Eignungsnachweis attributiver Prüfprozesse in einer kompakten Übersicht vor. Dabei wird auf aktuelle Verbands- und Firmenrichtlinien Bezug genommen und es werden Voraussetzungen, typische Anwendungsbereiche sowie methodische Grundlagen erläutert.

Das erste Whitepaper der Reihe hat sich mit den Grundlagen und Begrifflichkeiten der attributiven Prüfprozesse beschäftigt.

Dieses Whitepaper stellt ausgewählte Methoden vor, die für den Eignungsnachweis von Prüfprozessen nominaler Merkmale verwendet werden können:

- Short-Method
- Kappa Cohen/Fleiss
- Effektivität/Miss Rate/False Alarm Rate
- Bowker-Test

Der Konkordanz-Koeffizient nimmt eine Sonderrolle ein, da er nur für ordinale Merkmale anwendbar ist.

Short-Method

Die Short-Method wurde in der AIAG MSA in der ersten und zweiten Auflage beschrieben, ist aber seit der dritten Auflage nicht mehr enthalten, weil sie für die Durchführung eines Eignungsnachweises als nicht geeignet angesehen wird. Dennoch ist diese Methode noch in vielen Firmenrichtlinien vorhanden und kann laut VDA Band 5 für ein grobes Screening auf kapitale Fehlentscheidungen verwendet werden. Das Problem der Short-Method besteht darin, dass alle Prüferentscheide je Teil identisch sein müssen, um die Short-Method zu bestehen. Das heißt, der Unsicherheitsbereich darf nicht erkannt werden, was natürlich dem Bestreben entgegensteht, die Größe der Unsicherheit zu bewerten.

Die Short-Method wird im VDA Band 5 allerdings als geeignet angesehen, um an einem sinnvoll ausgewählten Prüflös die Stabilität des Prüfprozesses zu überwachen oder zu bewerten, ob neue Mitarbeiter in der Lage sind, alle Entscheidungen richtig zu treffen.

n	$X_{A;1}$	$X_{A;2}$	$X_{B;1}$	$X_{B;2}$	Beurteilung
1	-	-	+	+	Rot
2	-	-	-	-	Grün
3	+	+	+	+	Grün
4	+	-	-	-	Rot
5	+	+	+	+	Grün
6	-	-	-	-	Grün
7	-	-	-	-	Grün
8	+	+	+	+	Grün
9	-	-	-	+	Rot
10	+	+	+	+	Grün
11	+	+	+	+	Grün
12	+	-	-	-	Rot
13	+	+	+	+	Grün
14	+	+	-	+	Rot
15	+	+	+	+	Grün
16	+	+	+	+	Grün
17	-	-	-	-	Grün
18	+	+	+	+	Grün
19	-	-	-	-	Grün
20	+	+	+	+	Grün

Abbildung 1: Auswertung Short-Method

Voraussetzung

Die Short-Method benötigt ein Prüflös von 20 repräsentativen i. o. - und n. i. o.-Teilen. Optional können für die Teile Referenzbewertungen vorliegen, wodurch auch eine Bewertung der Prüfer gegen eine Referenz durchgeführt werden kann.

Durchführung

Bei der Short Method prüfen mindestens 2 Prüfer ein kleines Prüflös von mindestens 20 Teilen mindestens 2 mal in zufälliger Reihenfolge.

Anforderungen

Alle Entscheidungen der Prüfer müssen gleich ausfallen. Falls auch gegen eine Referenzbewertung geprüft wird, muss die Entscheidung der Prüfer mit der Referenzbewertung übereinstimmen.

Anmerkung

Bei der Nutzung als Stabilitätsprüfung ist der Vergleich mit einem Referenzentscheid obligatorisch.

Kappa (Cohen/Fleiss)

Die Kappa Methode bewertet die Übereinstimmung zwischen Prüfern, eines Prüfers mit sich selbst und von Prüfern mit der Referenz. Dabei werden die tatsächlichen Übereinstimmungen denen gegenübergestellt, die sich durch „reinen Zufall“ erklären lassen. Kappa beschreibt also, wie viele der möglichen Übereinstimmungen gefunden werden, die über rein zufällige Übereinstimmungen hinausgehen. Üblicherweise sollten Kappa Werte zwischen 0 und 1 liegen, wobei $\kappa = 0$ bedeutet, dass die Prüfentscheidungen alle durch rein zufällige Übereinstimmungen erklärbar sind, während $\kappa = 1$ zeigt, dass die Prüfer alle möglichen Übereinstimmungen gefunden haben. Ist $\kappa < 0$, so ist die Übereinstimmung sogar schlechter als durch reinen Zufall zu erwarten. Daraus kann man ableiten, dass möglicherweise Missverständnisse oder Fehlinterpretationen durch die Prüfer vorliegen.

Voraussetzung

Das Prüflös sollte laut VDA Band 5 und Bosch Heft 10 mindestens 100 - 200 Teile enthalten.

Durchführung

Ist Bedienerinfluss zu erwarten sollten mindestens 3 Prüfer jedes Teil mindestens 2 mal in zufälliger Reihenfolge prüfen. Ist kein Bedienerinfluss zu erwarten, sollte jedes Teil mindestens 6 mal geprüft werden.

Anforderungen

Laut AIAG MSA weist ein $\kappa < 0,4$ auf einen nicht geeigneten Prüfprozess hin, bei einem $\kappa \geq 0,75$ ist von einem geeigneten Prüfprozess auszugehen. Nach Bosch Heft 10 muss $\kappa \geq 0,7$ sein, um eine bedingte Fähigkeit darzustellen, ab $\kappa \geq 0,9$ gilt der Prüfprozess als fähig. Letzteres ist auch die Anforderung nach VDA Band 5.

Die Durchführung der Kappa-Verfahren nach Cohen und Fleiss ist formal identisch, allerdings werden unterschiedliche Vergleiche durchgeführt.

- Cohen's Kappa bewertet die paarweise Übereinstimmung zwischen je zwei Prüfern oder dem Prüfer und der Referenz. Mit 3 Prüfern A, B, und C, mit Einbeziehung der Referenz R, erhält man sechs Kappa Werte. – paarweise
 - Übereinstimmungen der Prüfer A x B, B x C, C x A, auch Interrater Agreement genannt.
 - Übereinstimmungen der Prüfer gegen die Referenz A x R, B x R, C x R

A*B Kreuztabelle					
Vergleich Prüfer A mit B					
		B*C Kreuztabelle			
Vergleich Prüfer A mit B		A*C Kreuztabelle			
A	0	C		Gesamt	
		0,00	1,00		
B	0	Zählung	43	7	50
		Erwartete Zählung	17,0	33,0	50,0
Gesamt	1	Zählung	8	92	100
		Erwartete Zählung	34,0	66,0	100,0
Gesamt	1,00	Zählung	51	99	150
		Erwartete Zählung	51,0	99,0	150,0
Vergleich	A x B	B x C		A x C	
Kappa	0,86	0,79		0,78	

Abbildung 2: Cohen's Kappa gegen Prüfer
Beispieldaten aus AIAG MSA 4

A*Referenz Kreuztabelle							
B*Referenz Kreuztabelle							
C*Referenz Kreuztabelle							
A	0	C			Gesamt		
	1	0		0,00	1,00		
	Gesamt	1	0,00	Zählung	42	9	51
			Erwartete Zählung	17,3	34,7	51,0	
	Gesamt	1,00	Zählung	6	92	99	
			Erwartete Zählung	31,7	67,3	99,0	
Gesamt	1,00	Zählung	48	102	150		
		Erwartete Zählung	48,0	102,0	150,0		
Vergleich		A x Referenz		B x Referenz		A x Referenz	
Kappa		0,88		0,92		0,77	

Abbildung 3: Cohen's Kappa gegen Referenz
Beispieldaten aus AIAG MSA 4

- Fleiss' Kappa bewertet die Übereinstimmung der Prüfer mit sich selbst und mit der Referenz, sowie die Übereinstimmung aller Prüfer miteinander und gemeinsam mit dem Referenzentscheid. Mit 3 Prüfern A, B und C, mit Einbeziehung der Referenz R, erhält man acht Kappa Werte.
 - Übereinstimmung der Prüfer mit sich selbst A x A, B x B, C x C, auch Intrarater Agreement genannt
 - Übereinstimmung aller Prüfer untereinander (A x B x C)
 - Übereinstimmungen der Prüfer gegen die Referenz A x R, B x R, C x R
 - Übereinstimmung aller Prüfer gemeinsam mit der Referenz (A x B x C x R)

Vielfach wird nur der schlechteste Kappa-Wert zur Beurteilung herangezogen.

Anzahl paarweise gleicher Kombinationen je Prüfobjekt i = 1, ... N ₀																				
Kategorien: 0 - Nicht in Ordnung 1 - In Ordnung																				
Prüfobjekt	Referenz	Prüfer-Durchlauf						A x A	B x B	A x B x C	A x B x C	A - 1 x Ref	A - 2 x Ref	A - 3 x Ref	B - 1 x Ref	B - 2 x Ref	B - 3 x Ref	C - 1 x Ref	C - 2 x Ref	C - 3 x Ref
		A-1	A-2	A-3	B-1	B-2	B-3													
		A-1	A-2	A-3	B-1	B-2	B-3	C-1	C-2	C-3										
Kappa: $\frac{P_{\text{Übs}} - P_{\text{Exp}}}{1 - P_{\text{Exp}}} = K$																				
Kappa: Jeder Prüfer gegen Referenz (Mittelwerte)																				
Kappa: Jeder Prüfer gegen Referenz (Mittelwerte)																				

Abbildung 4: Fleiss' Kappa nach Bosch Heft 10

Effektivität/Miss Rate/False Alarm Rate

Die Effektivität bewertet die Anzahl korrekter Entscheidungen in Bezug auf die Anzahl der möglichen Entscheidungen.

$$\text{Effektivität} = \frac{\text{Anzahl korrekter Entscheidungen}}{\text{Anzahl möglicher Entscheidungen}}$$

Neben der Effektivität werden in der Regel auch die Miss Rate (Fehlerrate) und die False Alarm Rate (Pseudofehler) ermittelt. Dabei bewertet die Miss Rate den Anteil fehlerhafter i. o.-Entscheidungen und die False Alarm Rate den Anteil fehlerhafter n. i. o.-Entscheidungen.

$$\text{Miss Rate} = \frac{\text{Anzahl fehlerhafter i. o.-Entscheidungen}}{\text{Anzahl möglicher richtiger n.i. o.-Entscheidungen}}$$

$$\text{False Alarm Rate} = \frac{\text{Anzahl fehlerhafter n. i. o.-Entscheidungen}}{\text{Anzahl möglicher richtiger i. o.-Entscheidungen}}$$

Voraussetzung und Durchführung

Analog des Fleiss'Kappa-Methode

Anforderung

Die Tabelle zeigt eine Empfehlung für die Bewertung der Ergebnisse nach AIAG MSA (2010).

	Effektivität	Miss Rate	False Alarm Rate
Akzeptabel	≥ 90 %	≤ 2 %	≤ 5 %
Bedingt Akzeptabel	≥ 90 %	≤ 5 %	≤ 10 %
Nicht Akzeptabel	> 80 %	> 5 %	> 10 %

Abbildung 5: Bewertungsgrenzen Effektivität, Miss-, und False Alarm Rate

Hinweis: Die Zusammensetzung des Prüfloses beeinflusst signifikant das Ergebnis des Eignungsnachweises. Soll z. B. im Wesentlichen eine Aussage über Miss Rate getroffen werden, tragen i. o.-Teile nichts zur Beurteilung bei. Das Prüflos muss dann gegebenenfalls anders aufgebaut sein als bei der Beurteilung der Effektivität oder eines Kappa-Wertes. Deshalb sollte auch hier dringend Expertenrat eingeholt werden.

Bowker-Test

Der Bowker-Test wird nur in VDA Band 5 genannt und ist ein Test auf symmetrische Entscheidungen der Prüfer. Er wird im Wesentlichen aus historischen Gründen aufgeführt, aber heute nicht mehr als Stand der Technik erachtet.

Voraussetzung

Der Bowker-Test benötigt ein Prüflos von 40 repräsentativen i. o.- und n. i. o.-Teilen. Eine Referenzbewertung wird nicht benötigt und kann auch in der Auswertung nicht berücksichtigt werden.

Durchführung

Die 40 Prüfbjekte werden von 2 Prüfern jeweils 3 mal geprüft. Die Richtigkeit der Entscheidungen wird nicht bewertet. Abhängig von den Beurteilungen wird das Teil je Prüfer jeweils einer von drei Klassen zugeordnet. Das Teil gehört zu

- Klasse 1, wenn alle Entscheidungen des Prüfers i. o. ergeben
- Klasse 2, wenn die Entscheidungen des Prüfers nicht eindeutig sind
- Klasse 3, wenn alle Entscheidungen des Prüfers n. i. o. ergeben

Die Ergebnisse werden für beide Prüfer in einer Kreuztabelle aufgetragen:

Häufigkeiten n_{ij}		Prüfer B		
		Klasse 1 Ergebnisse +++	Klasse 2 gemischte Ergebnisse	Klasse 3 Ergebnisse ---
Prüfer A	Klasse 1 Ergebnisse +++	1/1 7	1/2 6	1/3 7
	Klasse 2 gemischte Ergebnisse	2/1 2	2/2 12	2/3 3
	Klasse 3 Ergebnisse ---	3/1 0	3/2 1	3/3 2

Abbildung 6: Beispiel Bowker-Test nach VDA 5

In den Feldern (1/1), (2/2) und (3/3) werden von beiden Prüfern einheitliche Entscheidungen getroffen. Allen anderen Felder zählen Teile mit nicht einheitlichen Entscheidungen.

Anforderung: Die Bewertung wird nun über die Symmetrie der Matrix bezüglich der Diagonale der Übereinstimmungen durchgeführt. Es wird also mit Hilfe eines x-Tests verglichen:

- Feld (2/1) mit (1/2)
- Feld (3/1) mit (1/3)
- Feld (3/2) mit (2/3)

Die Nullhypothese des χ^2 -Tests lautet: „Die Prüfer weisen eine Symmetrie in ihren Entscheidungen auf“, bzw. „Die Matrix ist symmetrisch“. Diese Hypothese wird verworfen, wenn die Teststatistik größer ist als das 5 %-Quantil der χ^2 -Verteilung.

Spezielle Methode für originale Merkmale

Konkordanz-Koeffizient (Kendall's W)

Der Konkordanz-Koeffizient nach Kendall (Kendall's W) kann ausschließlich für ordinale Merkmale berechnet werden. Die Durchführung entspricht der Kappa-Methode, wobei die Prüfer den Prüfobjekten ordinale Kategorien (z.B. Schulnoten) zuweisen müssen. Zur Beurteilung der Prüferübereinstimmung werden die Abweichungen der Prüfer untereinander und gegen die Referenz analog der Kappa-Methoden bewertet. Allerdings werden die Abweichungen umso höher gewichtet, je mehr Kategorien im Vergleich zur Referenzbewertung übersprungen werden. Eine Abweichung um eine Note wird also geringer gewichtet als eine Abweichung um zwei oder mehr Noten. Je höher der ermittelte Konkordanz-Koeffizient, desto besser die Übereinstimmung zwischen den Prüfern mit der Referenz. Auch hier bedeutet ein Konkordanz-Koeffizient von 0, dass die Bewertungen durch reinen Zufall erklärt werden können und ein Konkordanz-Koeffizient von 1, dass alle Prüfobjekte korrekt eingestuft wurden.

Fazit

Die Darstellungen zeigen, dass es dem Anwender obliegt, das „richtige“ Verfahren auszuwählen, falls kundenseitig keine Vorgaben zu erfüllen sind. Es sollte also zuerst festgelegt werden, welcher Aspekt des Prüfprozesses bewertet werden soll (Vergleichbarkeit der Prüfer, Miss-Rate ...) und dann das dazu passende Verfahren durchgeführt werden. Möglicherweise sollte auch in Betracht gezogen werden, auf Basis einer umfassenden Untersuchung parallel mehrere Kennwerte zu ermitteln.

Wie einleitend erwähnt, stellen ordinale Merkmale eine Besonderheit dar und können ergänzend mit dem Konkordanz-Koeffizient nach Kendall bewertet werden.

Diskretisierte Merkmale bieten die Möglichkeit, über die genannten Methoden hinaus auch eine Messunsicherheit U_{ATTR} und somit einen Eignungsindex $Q_{MP,ATTR}$ zu ermitteln. Dazu mehr im dritten Teil dieser Reihe.

Über Testo Industrial Services

Die Testo Industrial Services GmbH ist einer der führenden Anbieter von messtechnischen Dienstleistungen für Kalibrierung und GxP-Services in Deutschland. Als Tochter des Messgeräteherstellers Testo SE & Co. KGaA verfügen wir über mehr als 60 Jahre Erfahrung in der Messtechnik. Zum Dienstleistungsportfolio zählen neben ISO-/Werkskalibrierung ebenso DAkkS-akkreditierte Kalibrierungen nach DIN EN ISO/IEC 17025:2018 und ganzheitliche Prüfmittelmanagementlösungen für eine zuverlässige Prüfmittelüberwachung. Darüber hinaus bieten wir auch Services in den Bereichen Qualitätssicherung, Qualifizierung und Validierung an. Ein breites Weiterbildungsangebot mit Seminaren und kostenfreien Webinaren rundet das Serviceportfolio ab.

Consulting Quality Assurance

Wir bieten Schulung und Beratung zur Anlaufbegleitung nach APQP und Qualitätssicherung sowie Akkreditierungsvorbereitungen nach ISO/IEC 17025 in den Branchen:

- Automotive (OEMs und Zulieferer)
- Aerospace
- Medical und
- Defence

Unsere Beratungsdienstleistungen im Detail

- Qualifizierung von Mess- und Prüfprozessen
- Messsystemanalyse, MSA
- Eignungsnachweise nach VDA5.X
- Qualifizierung von Fertigungsprozessen
- Maschinen- und Prozessfähigkeit, MFU/PFU
- Statistische Prozesslenkung, SPC
- Prüfplanung
- Identifikation der zu messenden Merkmale
- Technologiescreening für Planung und Aufbau eines neuen Prüfstands

- Prüfmittelmanagement
- Digitale Prüfmittelverwaltung (PRIMAS)
- Optimiertes Prüfmittelmanagement z.B.: Kalibrierintervall/Fehlergrenze/Entscheidungsregeln
- Akkreditierungsvorbereitung nach ISO/IEC17025
- Aufnahme Ist-Situation/Gap-Analyse
- Handlungsempfehlung anhand der identifizierten GAPs und Ausarbeitung einer Projektskizze für die Akkreditierungsvorbereitung

Autoren

**Frank Platte**

Junior Consultant Quality Assurance
fplatte@testotis.de

Frank Platte ist Juniorberater im Bereich Consulting Quality Assurance bei der Testo Industrial Services GmbH. Seit 2023 ist er Teil des Consulting Quality Assurance-Teams. Zuvor studierte er Produktion und Logistik an der TH Köln. Bei Testo Industrials Services beschäftigt er sich mit

Eignungsnachweis nach VDA 5, MSA nach AIAG und Akkreditierungsvorbereitungen für Prüf und Kalibrierlaboratorien nach ISO/IEC 17025:2017.

**Dipl.-Ing. Stephan Conrad**

Senior Consultant Quality Assurance
sconrad@testotis.de

Dipl.-Ing. Stephan Conrad ist Senior Berater im Team Consulting Quality Assurance bei Testo Industrial Services GmbH. Seit 2000 beschäftigt er sich mit den Anforderungen und Vorgehensweisen zur Qualifikation von Maschinen und Prozessen, Statistical Process Control (SPC), Prozessoptimierung sowie Messsystemanalysen

und dem Eignungsnachweis von Mess- und Prüfprozessen. Mit seiner langjährigen Expertise unterstützt und berät er Unternehmen aus der Automobilindustrie, der Medizintechnik und dem Maschinenbau bei der Entwicklung von Qualitätsstrategien, der Prozess- und Richtliniengestaltung sowie deren methodischer Umsetzung.

