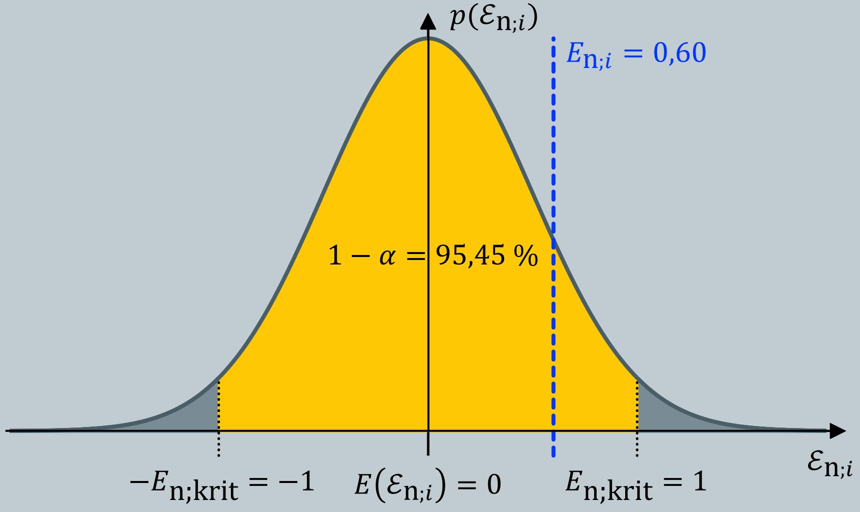
esz

calibration & metrology



Der E_n-Wert

Kriterium zur Leistungsbewertung von Eignungsprüfungen

06. Mai 2026 I Patricia Mayer
333. PTB-Seminar

Datensatz

- Teilnehmerwert x_i des Teilnehmers i mit $i \in \{1, \dots, N\}$ (N: Gesamtzahl Teilnehmer)
- Erweiterte Messunsicherheit $U_i = k_i \cdot u_i$ des Teilnehmers i → typischerweise*): $k_i = 2$
- Referenzwert x_{ref} und erweiterte Referenzunsicherheit $U_{\mathrm{ref}} = k_{\mathrm{ref}} \cdot u_{\mathrm{ref}}$ → typischerweise*): $k_{\mathrm{ref}} = 2$
  - messtechnisch bestimmt durch Referenzlabor
  - statistisch berechnet basierend auf Teilnehmerdaten (Konsenswert)

White Paper
„Eignungsprüfungen zwischen Kalibrierlaboratorien – Statistische Auswertung von Vergleichsmessungen“
https://www.esz-ag.de/downloads/

*) Basierend auf der Annahme von Normalverteilungen und einer Überdeckungswahrscheinlichkeit von ca. 95,45 %.

Definition

- **Leistungskenngröße** E_n-Wert (normalized Error) für den Teilnehmer i

$$E_{\mathrm{n};i} = \frac{d_i}{k_{1-\alpha} \cdot u(d_i)} = \frac{x_i - x_\mathrm{ref}}{k_{1-\alpha} \cdot \sqrt{u_i^2 + u_\mathrm{ref}^2 - 2u(x_i; x_\mathrm{ref})}}$$

  - $k_{1-\alpha}$ entsprechend Faltung der Wahrscheinlichkeitsdichten von Teilnehmer und Referenz und Überdeckungswahrscheinlichkeit $1 - \alpha$.
  - Berücksichtigung von Korrelationen anhand der Kovarianz $u(x_i; x_\mathrm{ref})$.

- **Leistungskriterium** zur Leistungsbewertung des Teilnehmers i

Positive Leistung (= bestanden) $\Leftrightarrow$ $-1 < E_{\mathrm{n};i} < 1$

Definition

- **Leistungskenngröße** E_{n}-Wert (normalized Error) für den Teilnehmer i

$$E_{\mathrm{n};i} = \frac{d_i}{k_{1-\alpha} \cdot u(d_i)} = \frac{x_i - x_{\mathrm{ref}}}{2 \cdot \sqrt{u_i^2 + u_{\mathrm{ref}}^2}}$$

  - $k_{1-\alpha} = 2$ entsprechend Normalverteilung für Überdeckungswahrscheinlichkeit von ca. 95,45 % ($\alpha = 4{,}55$ %).
  - In Abwesenheit von Korrelationen entfällt $u(x_i; x_{\mathrm{ref}})$.

- **Leistungskriterium** zur Leistungsbewertung des Teilnehmers i

Positive Leistung (= bestanden) $\Leftrightarrow$ $-1 < E_{\mathrm{n};i} < 1$

Graphische Eigenschaften

$$E_{\mathrm{n};i} = \frac{x_i - x_{\mathrm{ref}}}{2 \cdot \sqrt{u_i^2 + u_{\mathrm{ref}}^2}}$$

positive Leistung $\Leftrightarrow$ $|E_{\mathrm{n};i}| < 1$

(Annahmen: keine Korrelation, Normalverteilungen, $\alpha = 4{,}55\ \%$)

Graphische Eigenschaften

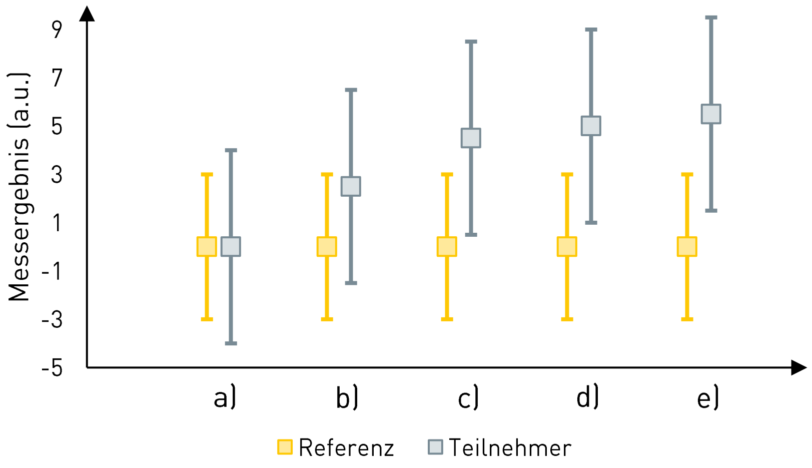
$$E_{n;i} = \frac{x_i - x_{ref}}{2 \cdot \sqrt{u_i^2 + u_{ref}^2}}$$

positive Leistung $\Leftrightarrow$ $|E_{n;i}| < 1$

(Annahmen: keine Korrelation, Normalverteilungen, $\alpha = 4{,}55\ \%$)

Formulierung

„Das Teilnehmerergebnis $x_i \pm U_i$ stimmt im Rahmen der Messunsicherheit mit dem Referenzergebnis $x_{ref} \pm U_{ref}$ überein."



Graphische Eigenschaften

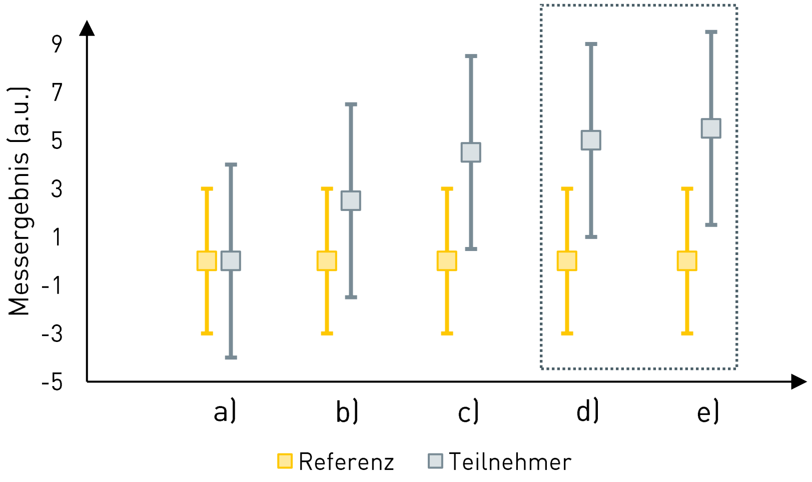
$$E_{\mathrm{n};i} = \frac{x_i - x_\mathrm{ref}}{2 \cdot \sqrt{u_i^2 + u_\mathrm{ref}^2}}$$

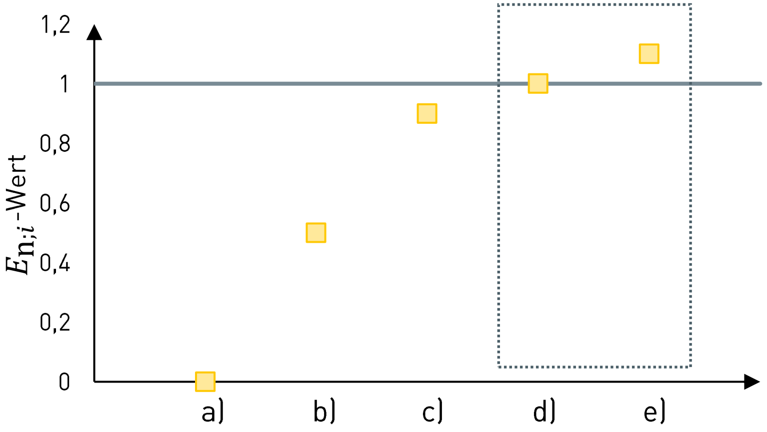
positive Leistung $\Leftrightarrow$ $|E_{\mathrm{n};i}| < 1$

(Annahmen: keine Korrelation, Normalverteilungen, $\alpha = 4{,}55\ \%$)

Formulierung

„Das Teilnehmerergebnis $x_i \pm U_i$ stimmt im Rahmen der Messunsicherheit mit dem Referenzergebnis $x_\mathrm{ref} \pm U_\mathrm{ref}$ überein.“





Trotz Überlappung der MU-Balken kann eine negative Leistungsbewertung, d.h. $|E_{\mathrm{n};i}| \geq 1$ vorliegen.

Graphische Eigenschaften

$$E_{\mathrm{n};i} = \frac{x_i - x_{\mathrm{ref}}}{2 \cdot \sqrt{u_i^2 + u_{\mathrm{ref}}^2}}$$

positive Leistung $\Leftrightarrow$ $|E_{\mathrm{n};i}| < 1$

(Annahmen: keine Korrelation, Normalverteilungen, $\alpha = 4{,}55\ \%$)

- **Wichtig**: Trotz der Überlappung der Messunsicherheitsbalken kann eine negative Leistungsbewertung, d.h. $|E_{\mathrm{n};i}| \geq 1$ erfolgen.

- **Grund**: Überlappung der Messunsicherheitsbalken $\Leftrightarrow$ $|x_i - x_{\mathrm{ref}}| < U_i + U_{\mathrm{ref}}$

  $|E_{\mathrm{n};i}| < 1 \Leftrightarrow |x_i - x_{\mathrm{ref}}| < \sqrt{U_i^2 + U_{\mathrm{ref}}^2} < U_i + U_{\mathrm{ref}}$

- **Ergebnis**: Die Überlappung der Messunsicherheitsbalken ist ein notwendiges, aber kein hinreichendes Kriterium für eine positive Leistungsbewertung über den E_{n}-Wert.

  $|E_{\mathrm{n};i}| < 1 \Rightarrow$ die Messunsicherheitsbalken überlappen

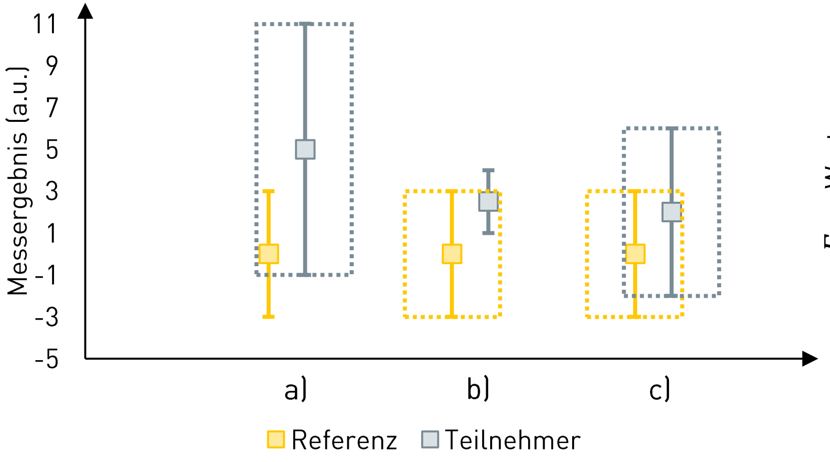
Graphische Eigenschaften

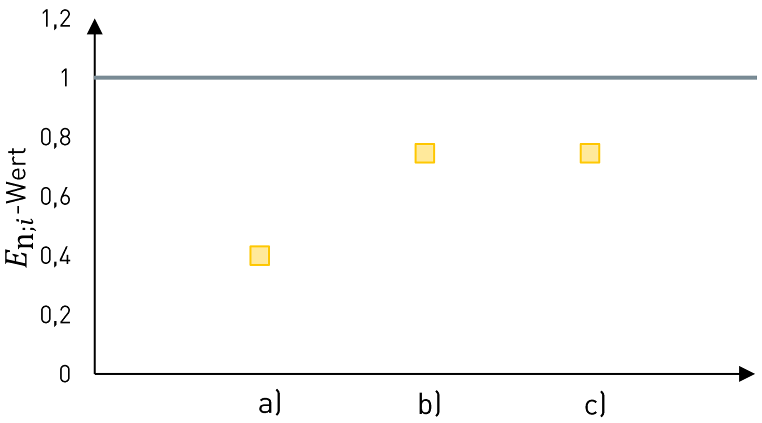
$$E_{n;i} = \frac{x_i - x_{ref}}{2 \cdot \sqrt{u_i^2 + u_{ref}^2}}$$

positive Leistung $\Leftrightarrow$ $|E_{n;i}| < 1$

(Annahmen: keine Korrelation, Normalverteilungen, $\alpha = 4{,}55$ %)

Kriterium 1[*)] Wenn der Messwert des Teilnehmers innerhalb des Messergebnisses der Referenz liegt und/oder, wenn der Referenzwert innerhalb des Messergebnisses des Teilnehmers liegt, dann folgt daraus eine positive Leistungsbewertung, d.h. $|E_{n;i}| < 1$.





Denn dann gilt:

$$|x_i - x_{ref}| < U_i < \sqrt{U_i^2 + U_{ref}^2}$$

bzw.

$$|x_i - x_{ref}| < U_{ref} < \sqrt{U_i^2 + U_{ref}^2}$$

*) hinreichendes Kriterium

Graphische Eigenschaften

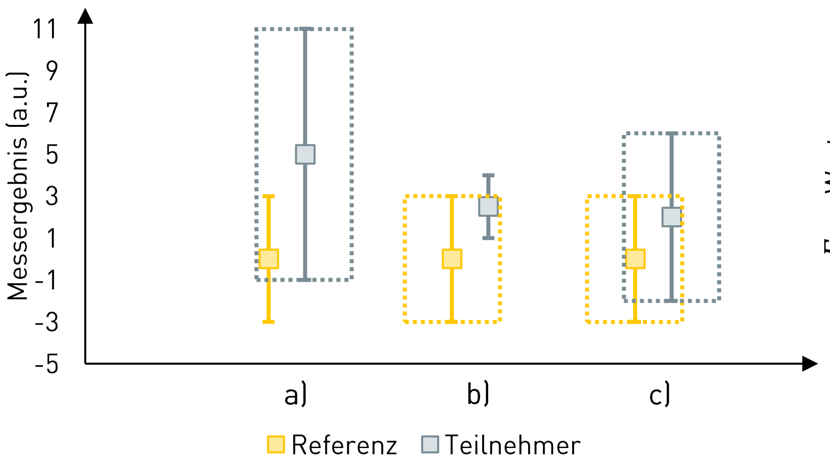
$$E_{\mathrm{n};i} = \frac{x_i - x_{\mathrm{ref}}}{2 \cdot \sqrt{u_i^2 + u_{\mathrm{ref}}^2}}$$

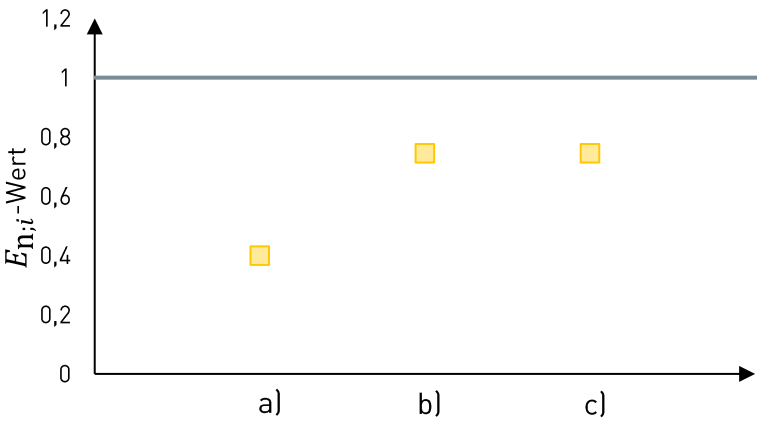
positive Leistung $\Leftrightarrow$ $|E_{\mathrm{n};i}| < 1$

(Annahmen: keine Korrelation, Normalverteilungen, $\alpha = 4{,}55\ \%$)

Kriterium 1*) Abschätzung über das Maximum der Messunsicherheiten

$$|x_i - x_{\mathrm{ref}}| \leq \max(U_i; U_{\mathrm{ref}}) \;\Rightarrow\; |E_{\mathrm{n};i}| < 1 \quad \text{(positive Leistungsbewertung)}$$





Denn dann gilt:

$$|x_i - x_{\mathrm{ref}}| < U_i < \sqrt{U_i^2 + U_{\mathrm{ref}}^2}$$

bzw.

$$|x_i - x_{\mathrm{ref}}| < U_{\mathrm{ref}} < \sqrt{U_i^2 + U_{\mathrm{ref}}^2}$$

*) hinreichendes Kriterium

Statistische Eigenschaften

$$E_{\mathrm{n};i} = \frac{d_i}{k_{1-\alpha} \cdot u(d_i)}$$

positive Leistung $\Leftrightarrow$ $|E_{\mathrm{n};i}| < 1$

- **Zufallsgröße**

$$\mathcal{E}_{\mathrm{n};i} = \frac{\mathcal{D}_i}{k_{1-\alpha} \cdot u(\mathcal{D}_i)}$$

  - beschrieben durch Wahrscheinlichkeitsdichte $p(\mathcal{E}_{\mathrm{n};i})$
  - Standardunsicherheit $u(\mathcal{E}_{\mathrm{n};i}) = 1/k_{1-\alpha}$

- **Stichprobe**

$$E_{\mathrm{n};i} = \frac{d_i}{k_{1-\alpha} \cdot u(d_i)}$$

  - konkreter Zahlenwert

Hypothesentest (zweiseitiger Signifikanztest)

Statistische Eigenschaften

$$E_{\mathrm{n};i} = \frac{d_i}{k_{1-\alpha} \cdot u(d_i)} \qquad \text{positive Leistung} \quad \Leftrightarrow \quad |E_{\mathrm{n};i}| < 1$$

- **Nullhypothese $\mathbb{H}_0$** $E(\mathcal{E}_{\mathrm{n};i}) = 0$ **Alternativhypothese $\mathbb{H}_1$** $E(\mathcal{E}_{\mathrm{n};i}) \neq 0$
  - Die Nullhypothese wird angenommen.
  - Anhand der Stichprobe $E_{\mathrm{n};i}$ wird entschieden, ob die Nullhypothese verworfen werden muss.

- **Zufallsgröße**

$$\mathcal{E}_{\mathrm{n};i} = \frac{\mathcal{D}_i}{k_{1-\alpha} \cdot u(\mathcal{D}_i)}$$

  - beschrieben durch Wahrscheinlichkeitsdichte $p(\mathcal{E}_{\mathrm{n};i})$
  - Standardunsicherheit $u(\mathcal{E}_{\mathrm{n};i}) = 1/k_{1-\alpha}$
  - Erwartungswert $E(\mathcal{E}_{\mathrm{n};i}) = 0$

Statistische Eigenschaften

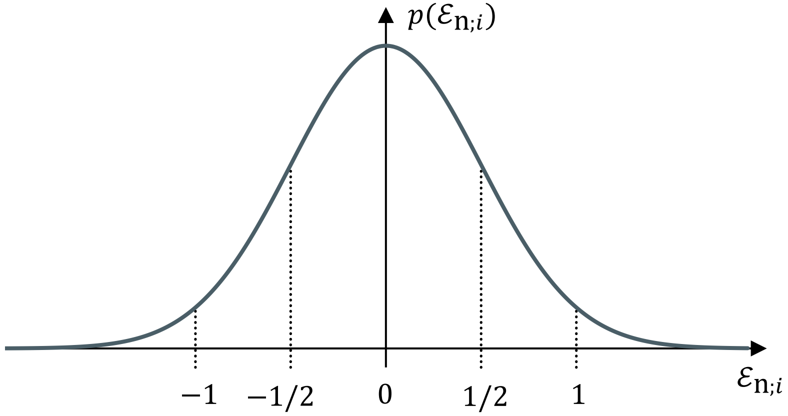
$$E_{\mathrm{n};i} = \frac{d_i}{k_{1-\alpha} \cdot u(d_i)} \qquad \text{positive Leistung} \Leftrightarrow \left|E_{\mathrm{n};i}\right| < 1$$

- **Zufallsgröße**

$$\mathcal{E}_{\mathrm{n};i} = \frac{\mathcal{D}_i}{k_{1-\alpha} \cdot u(\mathcal{D}_i)}$$

- beschrieben durch Wahrscheinlichkeitsdichte $p\left(\mathcal{E}_{\mathrm{n};i}\right)$
- Standardunsicherheit $u\left(\mathcal{E}_{\mathrm{n};i}\right) = 1/k_{1-\alpha}$
- Erwartungswert $E\left(\mathcal{E}_{\mathrm{n};i}\right) = 0$

- **Beispiel**



- Normalverteilung
- Überdeckungswahrscheinlichkeit $1 - \alpha = 95{,}45\ \%$ $(\alpha = 4{,}55\ \%)$
- $k_{1-\alpha} = 2$

Statistische Eigenschaften

$$E_{\mathrm{n};i} = \frac{d_i}{k_{1-\alpha} \cdot u(d_i)}$$

positive Leistung $\Leftrightarrow$ $|E_{\mathrm{n};i}| < 1$

- **Nullhypothese $\mathbb{H}_0$** $E(\mathcal{E}_{\mathrm{n};i}) = 0$ **Alternativhypothese $\mathbb{H}_1$** $E(\mathcal{E}_{\mathrm{n};i}) \neq 0$
- **Signifikanzniveau** α Typischerweise $\alpha = 4{,}55\ \%$ (Normalverteilung)
- **Kritischer Wert** $\int_{-E_{\mathrm{n;krit}}}^{E_{\mathrm{n;krit}}} p(\mathcal{E}_{\mathrm{n};i})\, \mathrm{d}\mathcal{E}_{\mathrm{n};i} = 1 - \alpha$

$E_{\mathrm{n;krit}} = 1$ unabhängig von der Wahrscheinlichkeitsdichte

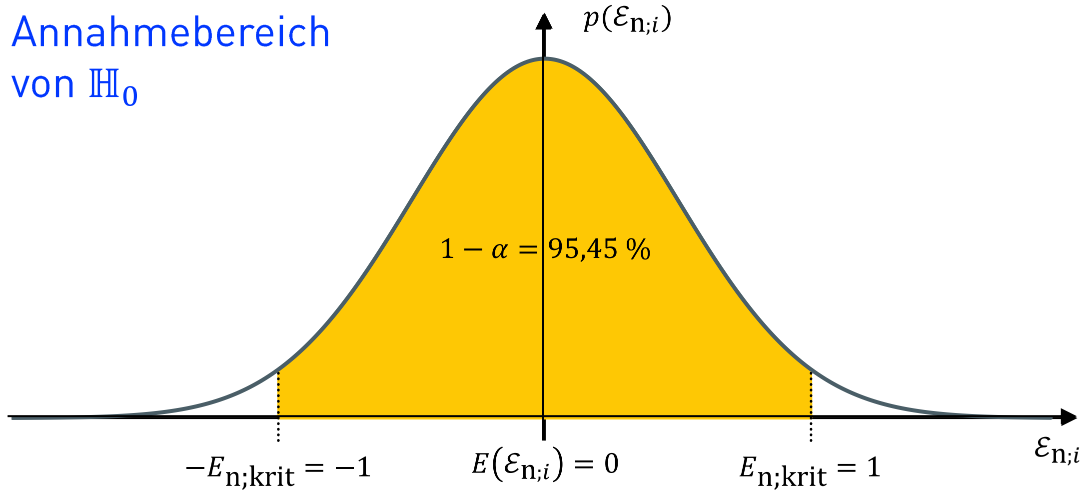
Statistische Eigenschaften

$$E_{\mathrm{n};i} = \frac{d_i}{k_{1-\alpha} \cdot u(d_i)}$$

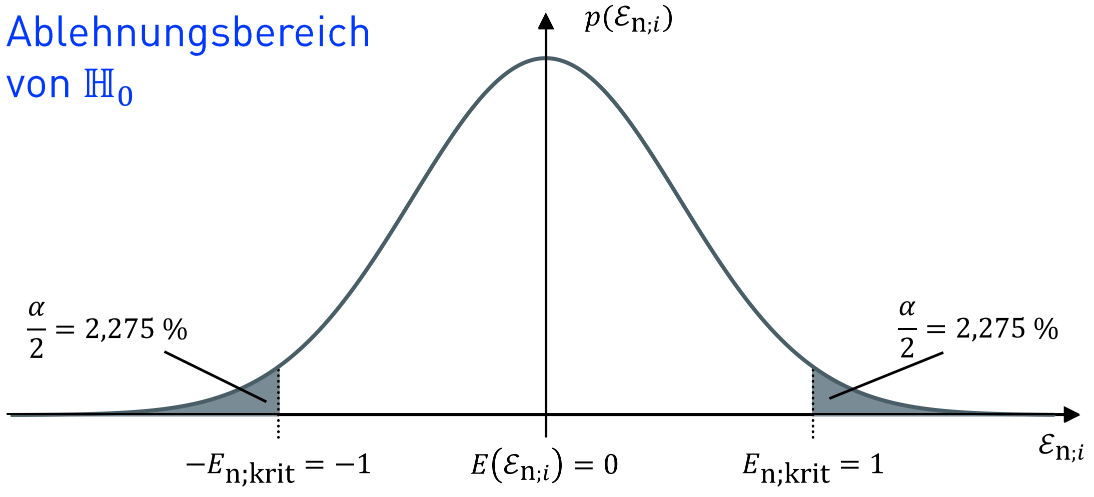
positive Leistung $\Leftrightarrow$ $|E_{\mathrm{n};i}| < 1$

- **Nullhypothese $\mathbb{H}_0$** $E(\mathcal{E}_{\mathrm{n};i}) = 0$
- **Signifikanzniveau** $\alpha = 4{,}55\ \%$
- **Kritischer Wert** $E_{\mathrm{n;krit}} = 1$

Annahmebereich von $\mathbb{H}_0$



Ablehnungsbereich von $\mathbb{H}_0$

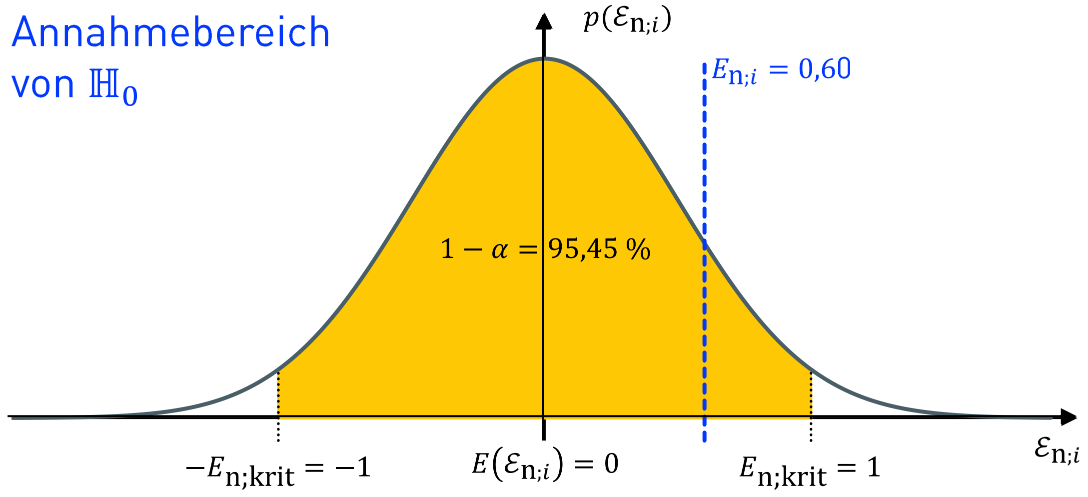


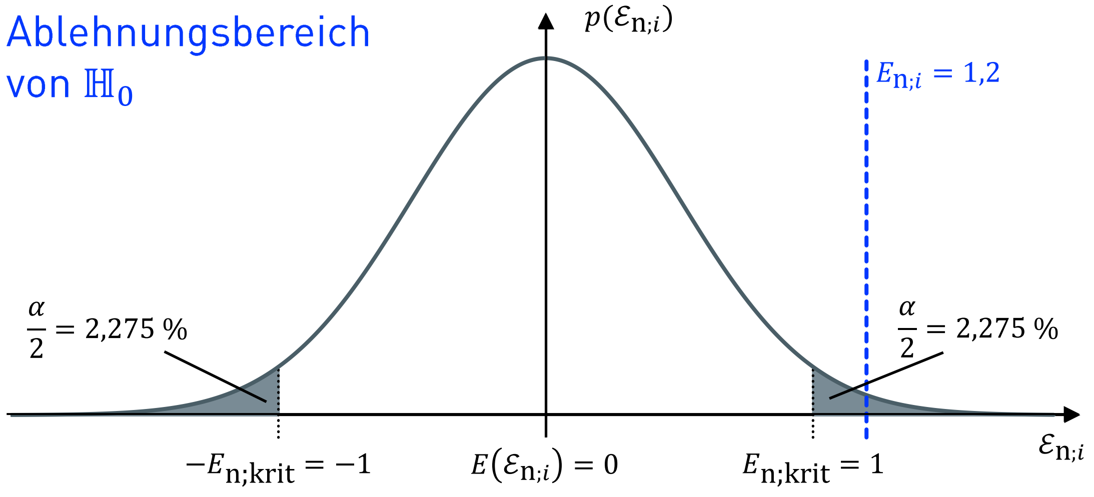
Statistische Eigenschaften

$$E_{\mathrm{n};i} = \frac{d_i}{k_{1-\alpha} \cdot u(d_i)} \qquad \text{positive Leistung} \Leftrightarrow |E_{\mathrm{n};i}| < 1$$

- **Nullhypothese** $\mathbb{H}_0$ $\quad E(\mathcal{E}_{\mathrm{n};i}) = 0$
- **Signifikanzniveau** $\quad \alpha = 4{,}55\ \%$
- **Kritischer Wert** $\quad E_{\mathrm{n;krit}} = 1$

Der konkret erzielte E_{n}-Wert ist eine Stichprobe aus der Wahrscheinlichkeitsdichte $p(\mathcal{E}_{\mathrm{n};i})$.





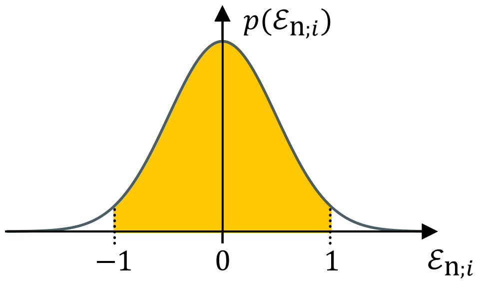
Statistische Eigenschaften

$$E_{\mathrm{n};i} = \frac{d_i}{k_{1-\alpha} \cdot u(d_i)}$$

positive Leistung $\Leftrightarrow$ $|E_{\mathrm{n};i}| < 1$

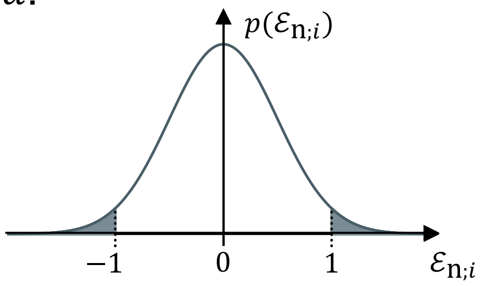
Annahme von $\mathbb{H}_0$

- Annahme, dass alle systematischen Fehler im Teilnehmerwert x_i und im Referenzwert x_ref korrigiert werden.
- Die verbleibende Differenz zwischen x_i und x_ref ist dann zufällig (vgl. Wahrscheinlichkeitsdichte).
- Die Annahme von $\mathbb{H}_0$ ist möglich, auch wenn $E_{\mathrm{n};i} \neq E(\mathcal{E}_{\mathrm{n};i}) = 0$.



Ablehnung von $\mathbb{H}_0$

- E_n-Werte, mit $|E_\mathrm{n}| > E_{\mathrm{n;krit}}$ werden unter $\mathbb{H}_0$ als möglich aber unwahrscheinlich erachtet und daher abgelehnt.
- Die Verwerfung von $\mathbb{H}_0$ bedeutet die Annahme von $\mathbb{H}_1$ und führt zu einer negativen Leistungsbewertung.
- Das Risiko einer Fehlentscheidung entspricht dem Signifikanzniveau α.



Statistische Eigenschaften

$$E_{\mathrm{n};i} = \frac{d_i}{k_{1-\alpha} \cdot u(d_i)}$$

positive Leistung $\Leftrightarrow$ $|E_{\mathrm{n};i}| < 1$

Stichprobe \ Grundgesamtheit	$\mathbb{H}_0$ **ist wahr**	$\mathbb{H}_0$ **ist falsch** ($\mathbb{H}_1$ **ist wahr**)
$\lvert E_\mathrm{n}\rvert < 1$ $\mathbb{H}_0$ **wird angenommen**	positive Leistungsbewertung korrekt	falsch-positive Leistungsbewertung Risiko kann i.A. nicht näher bestimmt werden, da $\mathbb{H}_1$ zu unkonkret.
$\lvert E_\mathrm{n}\rvert \geq 1$ $\mathbb{H}_0$ **wird abgelehnt** ($\mathbb{H}_1$ **wird angenommen**)	falsch-negative Leistungsbewertung Risiko ist gegeben durch Signifikanzniveau α	negative Leistungsbewertung korrekt

Statistische Eigenschaften

$$E_{\mathrm{n};i} = \frac{d_i}{k_{1-\alpha} \cdot u(d_i)}$$

positive Leistung $\Leftrightarrow$ $|E_{\mathrm{n};i}| < 1$

- **Nullhypothese** $\mathbb{H}_0$ $E(\mathcal{E}_{\mathrm{n};i}) = 0$
- **Signifikanzniveau** α
- **Kritischer Wert** $E_{\mathrm{n;krit}} = 1$

- Risiko der falsch-negativen Leistungsbewertung (α-Fehler): α (bspw. $\alpha = 4{,}55\ \%$)
- Risiko der falsch-positiven Leistungsbewertung (β-Fehler): $< 1 - \alpha = 95{,}45\ \%$ (nicht konkret angebbar)

- **Ergebnis:**
  - Die Leistungsbewertung in Eignungsprüfungen ist risikobehaftet.
  - Das Risiko hängt von der Wahl des Signifikanzniveaus α ab.
  - Die Verwerfung der Nullhypothese ist gesicherter als deren Annahme.
  - Positive Leistungsbewertung: Die Nullhypothese kann nicht widerlegt werden – sie wird aber auch nicht bestätigt. Daher ist eine regelmäßige Teilnahme an Eignungsprüfungen vorgesehen.

Messpunktübergreifende Leistungsbewertung

- **Bisher** Leistungsbewertung an einem Messpunkt
- **Praxis** In einer Eignungsprüfung häufig Vielzahl an Messpunkte
  → Leistungsbewertung erfolgt standardmäßig an jedem Messpunkt
- **Zusätzlich** Z.T. wird eine messpunktübergreifende Leistungsbewertung durchgeführt
  → „Wenn $|E_{\text{n};i}| < 1$ an mehr als 80 % der Messpunkte, dann ist die Eignungsprüfung insgesamt bestanden.“

Ist das sinnvoll?

Messpunktübergreifende Leistungsbewertung

- **Bisher** — Leistungsbewertung an einem Messpunkt
  In einer Eignungsprüfung häufig Vielzahl an Messpunkten
- **Statistische Eigenschaften** — negative Leistungsbewertung für Anteil α der Messpunkte akzeptabel
  (bspw. 4,55 % der Messpunkte)
- **Aber** — Nicht alle Messpunkte gehören zur gleichen statistischen Grundgesamtheit
  (bspw. verschiedene Messgrößen innerhalb einer Eignungsprüfung)

Eine messpunktübergreifende Leistungsbewertung kann irreführend sein.

Messpunktübergreifende ~~Leistungsbewertung~~ Leistungsanalyse

- **Statistischen Eigenschaften** Streuung der E_n-Werte wird erwartet (bspw. gemäß einer Normalverteilung)

- **Vorzeichen des E_n-Wertes**
  - Messpunkte weisen einheitliches Vorzeichen auf

    → ggf. Hinweis auf einen systematischen Messfehler

- **Größenordnung des E_n-Wertes**
  - Messpunkte haben einheitlich betragsmäßig einen sehr kleinen E_n-Wert

    → ggf. Hinweis auf Überschätzung der Messunsicherheit

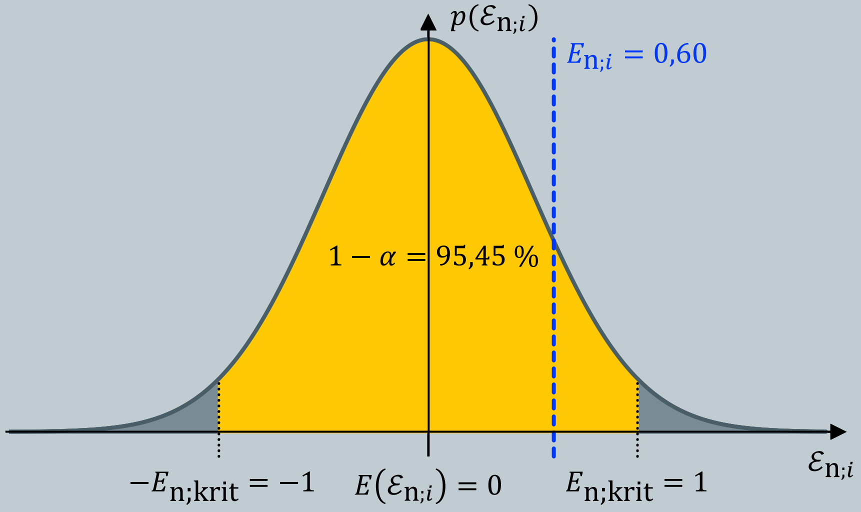
    → ggf. Hinweis auf nicht berücksichtigte Korrelationen

Zusammenfassung

DIN ISO 13528:2025, Anmerkung zu Absatz 9.1.1

„Leistungskenngrößen sind am nützlichsten, wenn die Kenngrößen und ihre Ableitung von den Teilnehmern und anderen interessierten Parteien verstanden werden."

- Definition des E_{n}-Wertes
- Graphische Eigenschaften
- Grenzen des E_{n}-Wertes bei $U_i < U_{\mathrm{ref}}$
- Statistische Eigenschaften
- Messpunktübergreifende Leistungsbewertung



Referenzen

[1] P. Mayer, „Der E_n-Wert – Kriterium zur Leistungsbewertung von Eignungsprüfungen“, 2025, VDI-Berichte 2464, „12. VDI-Fachtagung Messunsicherheit praxisgerecht bestimmen – Prüfprozesse in der industriellen Praxis 2025“, https://www.vdi-nachrichten.com/shop/messunsicherheit-praxisgerecht-bestimmen-pruefprozesse-in-der-industriellen-praxis-2025/

[2] Deutsches Institut für Normung / International Organization for Standardization, DIN EN ISO/IEC 17025:2018-03, „Allgemeine Anforderungen an die Kompetenz von Prüf- und Kalibrierlaboratorien“, 2018, https://www.dinmedia.de/de/norm/din-en-iso-iec-17025/278030106

[3] Deutsches Institut für Normung / International Organization for Standardization, DIN ISO 13528:2025-07, „Statistische Verfahren für Eignungsprüfungen durch Ringversuche“, 2025, https://www.dinmedia.de/de/norm/din-iso-13528/384736378

[4] Deutsches Institut für Normung / International Organization for Standardization, DIN EN ISO/IEC 17043:2023-10, „Konformitätsbewertung – Allgemeine Anforderungen an Eignungsprüfungen“, 2023, https://www.dinmedia.de/de/norm/din-en-iso-iec-17043/369196889

[5] P. Mayer, „Eignungsprüfungen zwischen Kalibrierlaboratorien – Statistische Auswertung von Vergleichsmessungen“, 2024, White Paper esz AG, https://www.esz-ag.de/downloads/

[6] BIPM, JCGM 100:2008, “Evaluation of measurement data – Guide to the expression of uncertainty in measurement”, 2008, https://www.bipm.org/en/committees/jc/jcgm/publications

[7] UKAS, M3003, „The expression of uncertainty and confidence in measurement”, 2024, Appendix P, https://www.ukas.com/wp-content/uploads/2023/05/M3003-The-expression-of-uncertainty-and-confidence-in-measurement.pdf

[8] BIPM, CIPM-MRA, „ Mutual recognition of national measurement standards and of calibration and measurement certificates issued by national metrology institutes”, 1999/2003 https://www.bipm.org/en/cipm-mra/cipm-mra-documents

Sicherheit

braucht **Vertrauen**. Wir machen

Vertrauen **messbar**.

Für **Entscheidungen** von heute

und **Produkte**

der **Zukunft**.