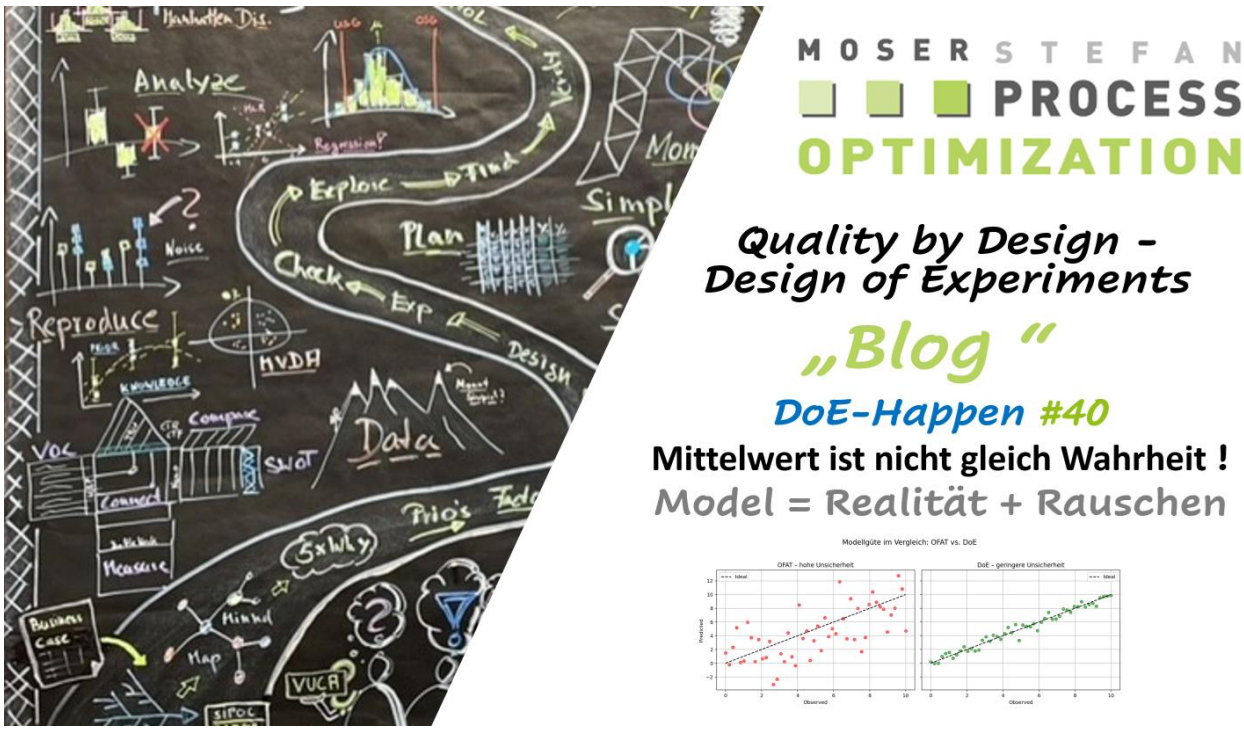


🧪 DoE-Happen #040: Deskriptive vs. Inferenzstatistik – und warum OFAT uns oft in die Irre führt

🗓️ 06. April. 2025 / ✍️Stefan Moser

Willkommen zurück im Reich der Möglichkeiten, DoE-Fans!

"Wir haben doch schon so viele Versuche gemacht – warum kommen wir trotzdem nicht weiter?"

Ein Satz, den ich in Workshops leider oft höre. Meist dann, wenn viele Daten gesammelt wurden, aber keine echten Erkenntnisse ableitbar sind. Das liegt selten an der Menge der Daten – sondern daran, wie wir sie betrachten und was eigentlich dahintersteckt. Zeit also, einen Blick auf zwei Grundpfeiler der Statistik zu werfen: **Deskriptiv vs. Inferenz**. Und auf das, was DoE anders (und besser) macht.

Unser Einstig in die Thematik anhand eines Beispiels

Gehen wir davon aus, dass wir einen Prozess verbessern wollen, zu dem bereits einige Daten vorliegen. Vielleicht haben wir im Vorfeld schon ein paar Variablen verändert – mal einzeln, mal parallel – je nach Gusto, Bauchgefühl oder erstem Teilerfolg.

📊 Deskriptive Statistik – die Lupe auf unsere Daten

Wenn wir uns mit **deskriptiver Statistik** beschäftigen, tun wir erstmal etwas sehr Menschliches: Wir **beschreiben**, was wir sehen.

- Mittelwert, Median, Standardabweichung
- Boxplots, Histogramme, Tabellen
- Häufigkeiten und Verteilungen innerhalb *einer* Stichprobe

Schon George Box warnte davor, bei Prozessverbesserungen auf OFAT zu setzen: ‚To understand interactions, you have to change more than one factor at a time.' (Box et al., 2005) – genau hier setzt DoE an.

Ein Beispiel: Der pH-Wert von 10 Proben liegt im Mittel bei 6,5. Klingt erst mal gut – aber reicht das wirklich für eine verlässliche Aussage? // **„Was bedeutet das für Ihre Daten?**

Nein. Denn diese Aussage gilt nur für die gemessene Gruppe. Was sie uns nicht verrät: **Ist das Ergebnis zufällig? Stabil? Übertragbar?** Genau hier kommt die **inferenzstatistische Sichtweise** ins Spiel.

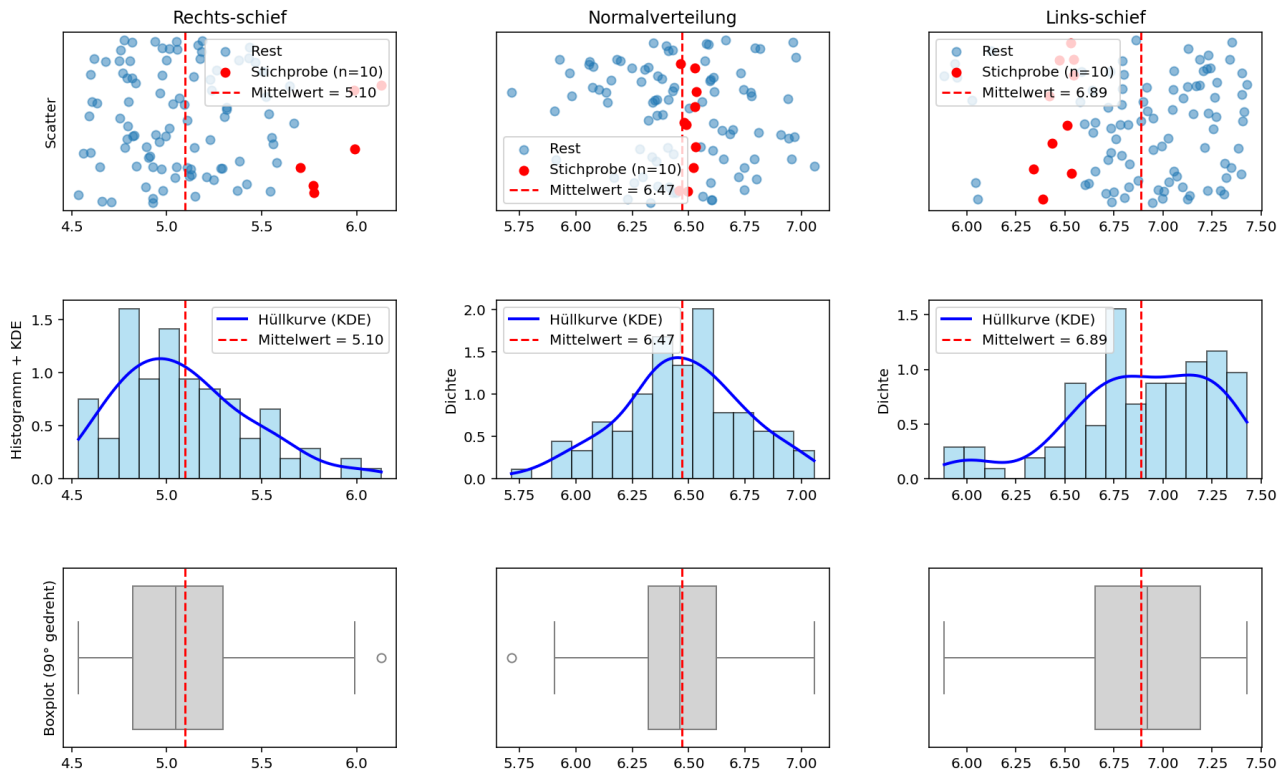


Abbildung 1: Warum eine Stichprobe (fast) nie die ganze Wahrheit erzählt

Im voranstehenden Beispiel habe ich fiktive Daten verwendet, um drei verschiedene Verteilungen zu visualisieren. (*Rechts schief, normal und linksschiefe Verteilung*)
Walter A. Shewhart, Pionier der statistischen Qualitätskontrolle, wies bereits 1939 darauf hin, dass Daten nur dann sinnvoll interpretiert werden können, wenn sie im Kontext des zugrundeliegenden Systems stehen – eine Momentaufnahme genügt nicht.

Denn auch wenn ein Mittelwert scheinbar eindeutig ist:
👉 Wenn die Faktoren im gesamten möglichen Raum variiert werden – also innerhalb der für den Prozess definierten Spezifikationen –, können instabile Prozesse bei anderen Faktorkonstellationen ganz andere Ergebnisse liefern.

🧠 Inferenzstatistik – aus Daten lernen (und extrapolieren)

Die **inferenzstatistische Statistik** will mehr. Sie fragt:

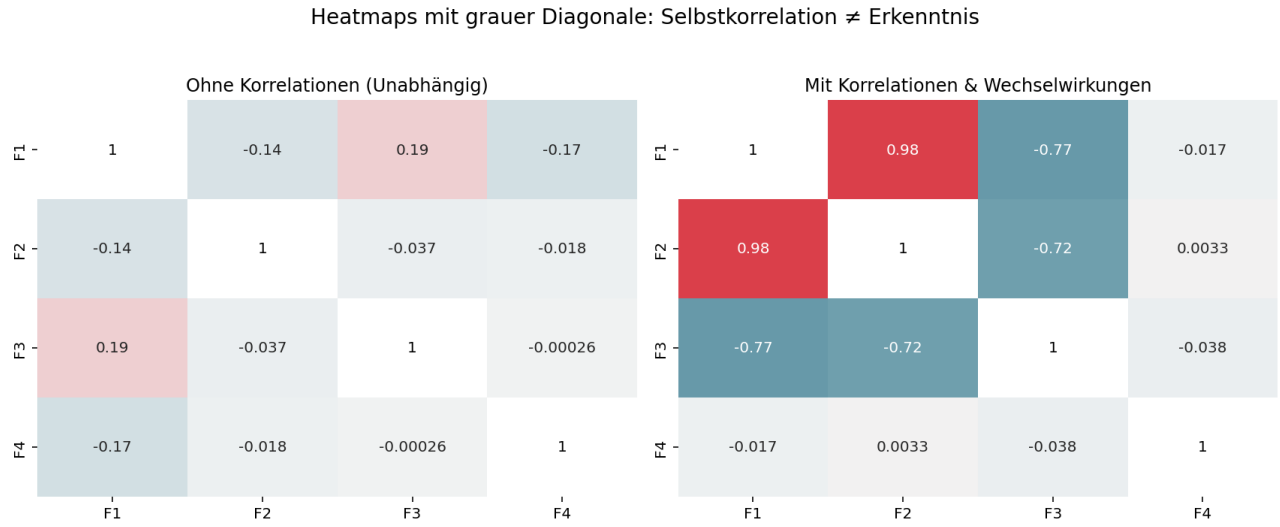
- Ist dieser Effekt **reproduzierbar**?
- Können wir auf die **gesamte Population** schließen?
- Wie **verlässlich** sind unsere Aussagen?
- Typische Werkzeuge: Konfidenzintervalle, Hypothesentests, p-Werte.

In der Inferenzstatistik dreht sich alles um Unsicherheit. Wie Larry Wasserman es formuliert: ‚Statistical inference is about drawing conclusions beyond the data at hand – and acknowledging the uncertainty that comes with it.' (Wasserman, 2004)

Ein klassisches Beispiel: Wir testen zwei Chargen mit und ohne Additiv. Die eine zeigt 5% weniger Ausschuss. Ist das ein echter Effekt oder Zufall?

Von Inferenz zur strukturierten Wirkungserkennung

In der Inferenzstatistik geht es darum, aus einer begrenzten Datenbasis **verlässliche Rückschlüsse** auf den Gesamtprozess zu ziehen. Doch genau hier stellt sich die zentrale Frage: **Sind unsere Einflussgrößen unabhängig voneinander – oder wirken sie im Verbund?**



Im voranstehenden Plot habe ich eine Heatmap genutzt, um die **Korrelationen zwischen den Faktoren** darzustellen.
Je **unabhängiger** die Faktoren voneinander variiert wurden, desto besser lassen sich ihre **Einzeleffekte** später analysieren und interpretieren.

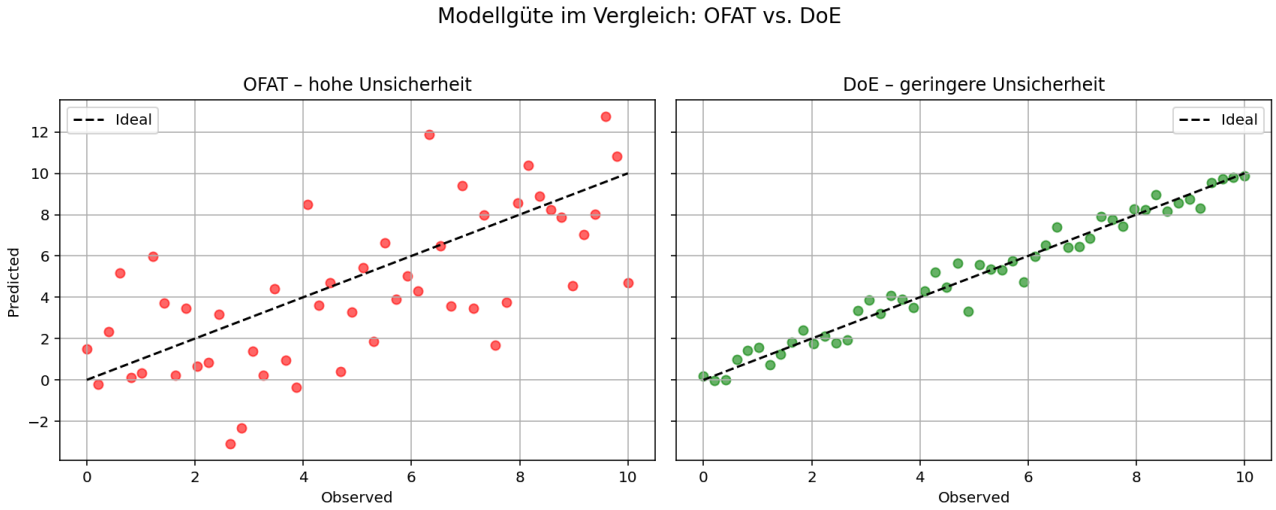
- Die Werte auf der Diagonalen – die „1“ – repräsentieren die Selbstkorrelation eines Faktors mit sich selbst. Das ist wenig überraschend (100 %) und für unsere Analyse irrelevant, da es hier ausschließlich um die Beziehungen zwischen verschiedenen Faktoren geht.
- Die übrigen Korrelationswerte reichen von 0 (keine Korrelation) bis ±1 (perfekte Korrelation) – und genau hier zeigt sich, ob wir die Effekte später überhaupt sauber auseinanderhalten können.
- Ob wir Variablen **unabhängig oder in Wechselwirkung** betrachten können, hängt direkt davon ab, **wie** wir sie untersucht haben. Wenn Faktoren *nicht systematisch variiert* wurden, sondern zufällig oder isoliert – dann können wir Effekte nicht sauber voneinander trennen.

Und genau hier setzt **Design of Experiments (DoE)** an: Ein gutes Versuchsdesign ermöglicht es, mit **wenigen, aber gezielt gewählten Versuchen** die Effekte der Einflussgrößen **sichtbar und interpretierbar** zu machen.

Der DoE-Anwender spricht dabei vom **Ursache-Wirkung-Modell**:
Eine strukturierte Annäherung an die Frage: *Welche Faktoren beeinflussen welche Zielgrößen – und wie sicher ist diese Aussage?*

Oder, ganz einfach ausgedrückt: **Realität = Modell + Rauschen** – so, wie es Douglas Montgomery in seinen Arbeiten zu DoE beschreibt: Jede gemessene Antwort ist eine Kombination aus erklärbaren Effekten und

einem zufälligen Fehleranteil.



- Das **Modell** erklärt, was wir durch unsere Versuche und Messungen systematisch abbilden können. Das **Rauschen** steht für den Anteil der Variation, der (noch) außerhalb unserer Kontrolle oder unseres Verständnisses liegt – oft bedingt durch unbekannte oder schwer kontrollierbare Umgebungsbedingungen.
- Je schlechter die Qualität der Datenbasis – insbesondere, wenn die untersuchten Faktoren **nicht unabhängig voneinander variiert** wurden –, desto größer ist die Abweichung zwischen Modellvorhersage und Realität.
- Ein Teil dieser Abweichung wird als ‚**Lack of Fit**' bezeichnet – also der **systematische Anteil, den unser Modell nicht erklären kann**, **weil wichtige Einflussgrößen fehlen oder falsch modelliert wurden.“**

Wieviel hat nun DoE damit zu tun bzw. wo greift dessen Hebel methodisch an?

Sehr viel! Denn Versuchsplanung ist letztlich ein clever konstruiertes Design, das genau diese Unsicherheiten strukturiert adressiert. Übrigens: Das „D“ in DoE steht nicht ohne Grund für „Design“ – also für die gezielte Konstruktion des Versuchsplans. Schon **George Box warnte davor, bei Prozessverbesserungen auf OFAT** zu setzen: ‚To understand interactions, you have to change more than one factor at a time.' (Box et al., 2005) – genau hier setzt DoE an.“

Was macht DoE dabei konkret besser?

🔍 Der untersuchte Raum wird gezielt abgesteckt – mit Blick auf sinnvolle und robuste Einstellungen der Faktoren.

🔍 Effekte werden gezielt geprüft, nicht bloß beobachtet.

🔍 Wechselwirkungen und Nichtlinearitäten werden sichtbar – und lassen sich modellieren.

👉 **Deshalb liefert DoE nicht nur hübsche Mittelwerte, sondern belastbare Aussagen mit Substanz.**

Was passiert ohne inferenzstatistische Grundlage?

- Einzelne Versuche → führen zu überinterpretierten Resultaten
- **Viele Versuche** → bedeuten nicht automatisch Erkenntnisgewinn
- **Blindes Variieren / unstrukturiertes Probieren** → endet in Trial-and-Error-Spiralen – wie unser Student Heinz, der an einem Rezept tüftelt und ständig nur eine Zutat ändert, ohne zu wissen, warum es mal klappt und mal nicht. In DoE-Projekten sehen wir regelmäßig, dass:

1. **Einzelversuche** nicht ausreichen, um holistische Aussagen zu treffen.

2. **Mehr Versuche** ohne Fokus nicht zielführend sind.
3. Nur die **richtigen Versuche im richtigen Bereich** mit den richtigen Faktoren zu Erkenntnis führen.

Deshalb gilt: Kein Erkenntnisgewinn ohne Plan – und keine sinnvolle Planung ohne klares Problemverständnis. Gerade in strukturierten Methoden wie **DFSS (Design for Six Sigma)**, **DMAIC** oder dem **PADC-Zyklus** wird dieser Zusammenhang besonders betont: Nur wer das **Problem präzise definiert**, kann auch den **Untersuchungsrahmen sinnvoll abstecken** – und daraus belastbare, nutzbare Ergebnisse ableiten. Schon Shewhart betonte, dass jede Datenanalyse ohne Verständnis des Systems in die Irre führen kann."
(vgl. Shewhart, 1939: Statistical Method from the Viewpoint of Quality Control)

Die Rolle der Fachkompetenz

DoE ist kein Selbstläufer. Ohne **fachliche Expertise** bleibt auch der beste Versuchsplan blind. Warum?

- Nur wer den Prozess versteht, erkennt sinnvolle Grenzen.
- Nur wer die Chemie kennt, weiß, wann ein Additiv wirkt – und wann es zum Weichmacher verkommt.
- Nur wer die Zielgrößen kennt, kann sagen, was Erfolg wirklich ist.

💬 **Fazit – Statistik ist kein Zahlenfriedhof**

Deskriptive Statistik ist der Einstieg. Inferenzstatistik bringt Substanz. Doch **erst mit DoE** machen wir beides gezielt nutzbar – wir **erheben**, **verstehen** und **modellieren** unsere Daten.

Die Kunst liegt nicht im Sammeln, sondern im **gezielten Fragen**.

🧩 Und nur mit einer klaren Problemformulierung (siehe Blog #009 zum A3-Report!) können wir diesen Fragen auch den nötigen Rahmen geben.

🧩 Vielleicht der wichtigste Punkt: Wie John Tukey schon 1962 betonte: ‚The greatest value of a picture is when it forces us to notice what we never expected to see.' Eine klare Problemformulierung schafft genau diesen Blick für das Wesentliche – bevor wir uns im Datendschungel verlieren.

👉 Kein Plan = kein Schutz.
👉 Klarer Plan = ~~keine~~, deutlich weniger Diskussion.

Oder wie man im Projektjargon sagt: **„Cover your Ask."** 😉

Bleiben Sie experimentierfreudig und neugierig!



💬 Bleiben Sie am Ball! In den kommenden #DoE-Happen vertiefen wir diese und andere Fragestellungen. Bitte teilen Sie Ihre Erfahrungen z.B. bei LinkedIn in den #Kommentaren. Sie können Sie die Themenrichtung mitgestalten. Ich freue mich darauf, von Ihnen zu hören!

🎁 Für regelmäßige Updates besuchen Sie meine Webseite: www.stefan-moser.com, wo Sie eine Übersicht und die Chronologie der Blog-Reihe finden.

Ihr DFSS & DoE Trainer,

Stefan Moser

DFSS-Proj.-mgmt. Trainer, DoE & MVDA Lecturer, Trainer, Facilitator, Specialist SIMCA, MODDE, Impulse-Geber

Mein Angebot zur Begleitung und Weiterentwicklung:

Ich biete DoE-Kurse an, die vom Einsteiger- bis zum Masterkurs reichen. Dabei decke ich alle relevanten Bereiche ab: von Fokus-Kursen zu Themen wie Screening, Charakterisierung, Optimierung und Robustheit bis hin zu Spezialkursen zu Mischungs- und Formulierungsdesigns, Stabilität oder spezifischen hierarchischen Designs wie Red Mup.

Neben diesen Kursen unterstütze ich meine Kunden bei der Versuchsplanung – sei es durch gezielte Beratung oder in Form von Troubleshooting oder Workshops. Ich begleite Sie in allen Phasen: von der Problemformulierung und Machbarkeitsstudie über die Optimierung bis zur robusten Absicherung Ihrer Prozesse.

Neben meinem Lieblingsthema DoE biete ich auch Kurse in den Bereichen MVDA, DFSS und QFD an. Hier unterstütze ich unter anderem die Ausbildung zum DFSS-Manager in den Stufen Yellow, Green und Black Belt. Diese Kurse realisiere ich in Zusammenarbeit mit meinen Partnern.

Wenn Sie Ihre Prozesse und Methoden auf das nächste Level bringen möchten, finden wir gemeinsam die passende Lösung!



Wenn Sie dazu mehr erfahren möchten, So finden Sie diese Hinweise auf meiner Webseite. www-stefan-moser.com

Gerne können Sie mich auch direkt anschreiben unter info@stefan-moser.com

Literaturhinweise:

- Montgomery, D.C. (2020): *Design and Analysis of Experiments*, Wiley.
- Box, G.E.P., Hunter, W.G., Hunter, J.S. (2005): *Statistics for Experimenters*, Wiley.
- Shewhart, W.A. (1939): *Statistical Method from the Viewpoint of Quality Control*, Dover.
- Wasserman, L. (2004): *All of Statistics*, Springer.
- Tukey, J.W. (1962): *The Future of Data Analysis*, in: *Annals of Mathematical Statistics*.