Auch A/B-Tests mit guten Testkonzepten führen mit falschen Statistikansätzen schnell und in jeder Testing-Phase zu nicht aussagekräftigen Ergebnissen und Fehlinterpretationen.
Wir zeigen Dir die 10 wichtigsten Statistik-Fallen, die in den einzelnen Phasen des Optimierungsprozesses auftreten können und geben praktische Tipps, wie Du sie vermeidest und valide Ergebnisse erhältst.

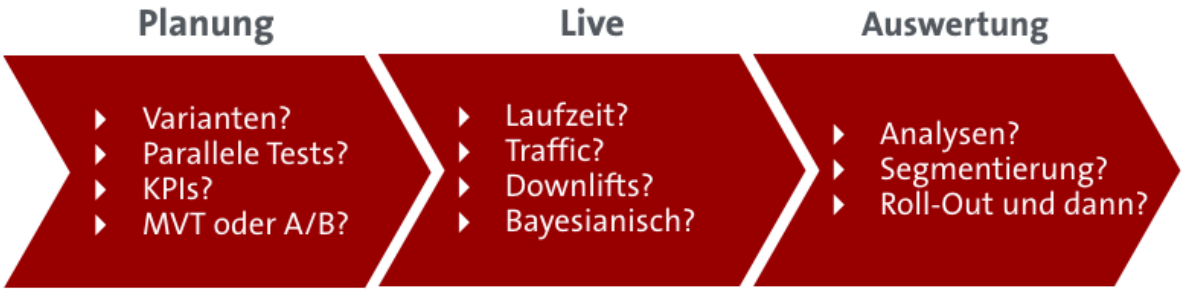


Abbildung 1: In jeder Phase eines Tests lauern Statistik-Stolperfallen.

Phase 0: Vor dem Test

Statistik-Falle #1:
„Lasst uns **möglichst viele Varianten** testen, irgendeine wird schon funktionieren."

Keine besonders gute Idee. Generell sollte man im Optimierungsprozess immer hypothesengetrieben arbeiten. Wahllos drauf los testen bringt nichts. Es gibt aber noch ein Problem und das heißt: Validität. Je mehr Testvarianten Du laufen lässt, desto höher ist auch die Wahrscheinlichkeit, eine davon als Gewinner zu deklarieren, obwohl sie gar keiner ist.
Vielleicht ist dem einen oder anderen der Begriff „Alpha-Fehler Kumulierung" bereits begegnet. Bei jedem Test akzeptiert man, dass es eine gewisse Irrtumswahrscheinlichkeit über das Testergebnis gibt. Da man beim A/B-Testing nur einen Ausschnitt aus der kompletten Kundenbasis betrachtet, gibt es immer Unsicherheiten über die tatsächlichen Effekte. Ziel des Tests ist es aber, die gemessenen Effekte im Test auf die gesamte Kundenbasis zu übertragen. Dabei spielen Konfidenzen eine wichtige Rolle.

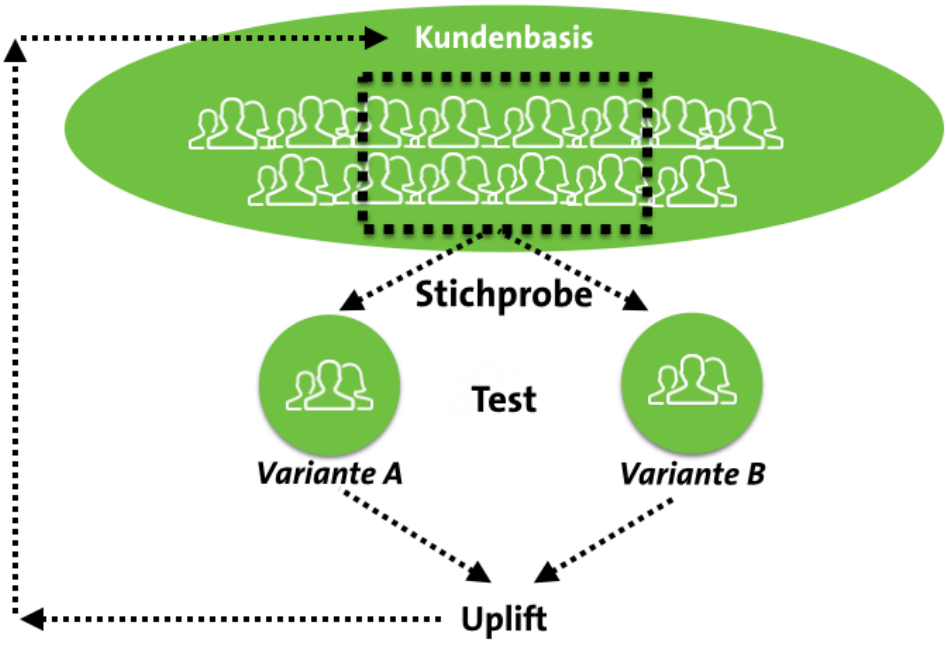


Abbildung 2: Testergebnisse sollen auf alle Kunden übertragen werden, um allgemein gültige Aussagen machen zu können.

Mit Hilfe von Konfidenzen kann man festlegen, wie sicher man sein möchte, dass gemessene Uplifts tatsächlich existieren. In der Regel arbeitet man mit einem Konfidenzniveau von 95%. Im Umkehrschluss heißt dies, dass man eine 5-prozentige Wahrscheinlichkeit für den Fehler 1. Art (Alpha-Fehler) akzeptiert. Das heißt, dass man in 5% aller Fälle auf einen signifikanten Effekt der

Testvariante schließt, obwohl es in Wirklichkeit gar keinen gibt. Jedoch gilt diese geringe Fehlerwahrscheinlichkeit nur für den Fall, dass man EINE Testvariante ins Rennen schickt. Bei mehreren Varianten erhöht sich dieser Fehler exponentiell.
Bei einer Testvariante beträgt die Irrtumswahrscheinlichkeit 5%. Der kumulierte Alpha-Fehler tritt ein, wenn mehrere Gruppen mit einer anderen Gruppe verglichen werden. In diesem Fall hat jeder einzelne Gruppenvergleich einen einfachen Alpha-Fehler, wobei dieser über alle Gruppen kumuliert zu betrachten ist (dies folgt aus den Axiomen der Wahrscheinlichkeitstheorie). Bei drei Testvarianten, die jeweils mit der Control verglichen werden, liegt die Irrtumswahrscheinlichkeit schon bei 14 %; mit 8 Testvarianten bei über 34 %. Das bedeutet, dass es bei einem solchen Test in mehr als jedem dritten Fall zu einem signifikanten Uplift kommt, der aber nur rein zufällig entstanden ist.

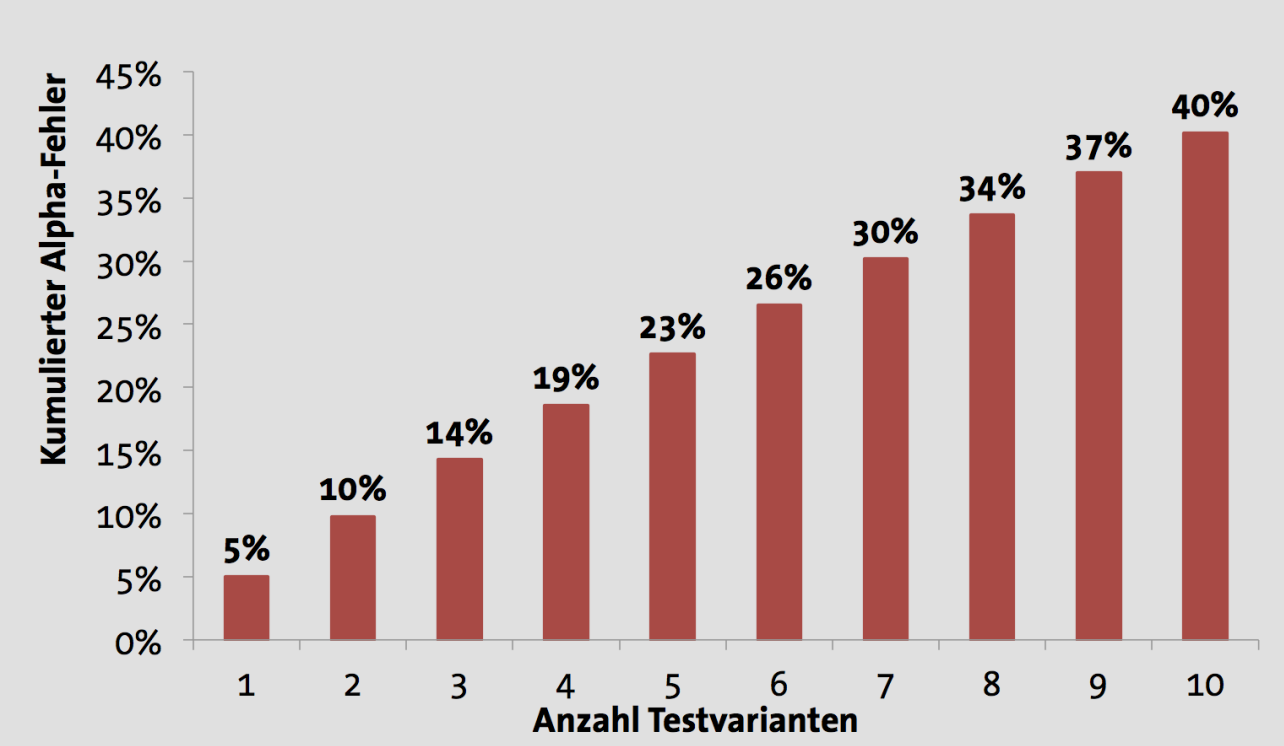


Abbildung 3: Je mehr Varianten man testet, desto höher ist das Risiko einer falschen Entscheidung.

Falls Du es genau wissen willst, so kannst Du den kumulierten Alpha-Fehler für deinen Test berechnen:

$$Alpha_{kumuliert} = 1-(1-Alpha)^n$$

Alpha = gewähltes Signifikanzniveau, in der Regel 0,05
n = Anzahl der Testvarianten in deinem Test (ohne Control)

Die gute Nachricht: Verwendest Du die neue Smart Stat von VWO oder Optimizely's Stats Engine, werden automatisch Verfahren eingesetzt, die für eine steigende Fehlerwahrscheinlichkeit bei mehreren Testvarianten korrigieren.
Hast Du ein Testing Tool, welches keine Korrekturverfahren verwendet, gibt es zwei Möglichkeiten:

1. Alpha-Fehler selber korrigieren
   Mit Hilfe der statistischen Varianzanalyse (ANOVA) lässt sich zunächst testen, ob es bei mehreren Testvarianten überhaupt einen signifikanten Unterschied im Gesamtvergleich gibt (der sogenannte F-Test). Dann weiß man aber noch nicht, bei welcher Testvariante das der Fall ist. Durch eine einfaktorielle Varianzanalyse wird geprüft, ob sich die Varianzen der einzelnen Stichproben unterscheiden und welche Testvariante eine signifikant höhere Conversion Rate als die Control aufweist.
2. Anzahl Testvarianten begrenzen
   Wem das zu kompliziert ist, der kann ganz einfach die Anzahl der Testvarianten von vorne herein begrenzen. Aus Erfahrung sollte man nicht mehr als drei Varianten plus die Control in einem Test laufen lassen. Die Varianten sollten dabei immer hypothesengetrieben konzipiert werden.

Übrigens: Das Problem der steigenden Fehlerwahrscheinlichkeit tritt auch bei multivariaten Tests (MVT) auf! Praktische Tipps zu einem erfolgreichen MVT findest Du unter „Statistik-Falle #4".

Statistik-Falle #2:

„Wir können auf **keinen Fall zwei Tests gleichzeitig** laufen lassen. Die Ergebnisse beeinflussen sich gegenseitig und **verzerren**."

In der Tat gibt es unterschiedliche Meinungen dazu, ob man zulassen sollte, dass ein Nutzer in mehreren Tests gleichzeitig sein kann. Während die einen argumentieren, dass das unproblematisch sei, da sich die Effekte gegenseitig ausgleichen, sehen andere Verschmutzungseffekte durch paralleles Testen; Interaktionen zwischen Tests und Varianten können zu nicht-linearen Effekten, höheren Varianzen und sinkender Validität führen.
Tatsächlich hängt das konkrete Vorgehen davon ab, wie hoch man das Risiko für solche asymmetrischen Effekte einschätzt. Hier bieten sich verschiedene Lösungen an:

1. Geringes Risiko von Interdependenzen

Ist das Risiko gering und der Traffic-Overlap zwischen den Tests überschaubar, kann man zwei Tests durchaus parallel laufen lassen, so dass ein und derselbe Nutzer auch in mehreren Tests sein kann.
Das Vorgehen führt zu validen Ergebnissen, wenn die Testkonzepte nicht oder wenig miteinander konkurrieren. Zum Beispiel könnte man ein Konzept, welches UVPs auf der Startseite testet, parallel zum Test laufen lassen, welcher

die Darstellung der Kundenmeinungen auf der Detailseite testet. Beide Tests kann man rein methodisch und technisch mit 100% Traffic laufen lassen, d.h. dass ein Nutzer in beiden Tests sein kann. Die Wahrscheinlichkeit, dass die Tests miteinander konkurrieren, ist hier eher gering.

Warum das geht? Ohne zu tief in die Mathematik des A/B-Testings einzusteigen: Es liegt an dem großartigen Zufallsprinzip, welches dem A/B-Testing zugrunde liegt.

Nutzer werden zufällig der Test- bzw. Kontrollvariante zugeordnet. Das gilt bei jedem Test und auch bei solchen, die parallel laufen, also den Nutzer mehreren Tests gleichzeitig zuordnen. Durch dieses Zufallsprinzip in Kombination mit dem Gesetz der großen Zahlen wird sichergestellt, dass die Anteile von Nutzern, die auch an einem anderen Test teilnehmen, in Test- und Kontrollvariante proportional sind (siehe Abbildung 4). So gleichen sich die Effekte auf die gemessene Conversion Rate in den Gruppen aus und man kann den Uplift verlässlich messen.

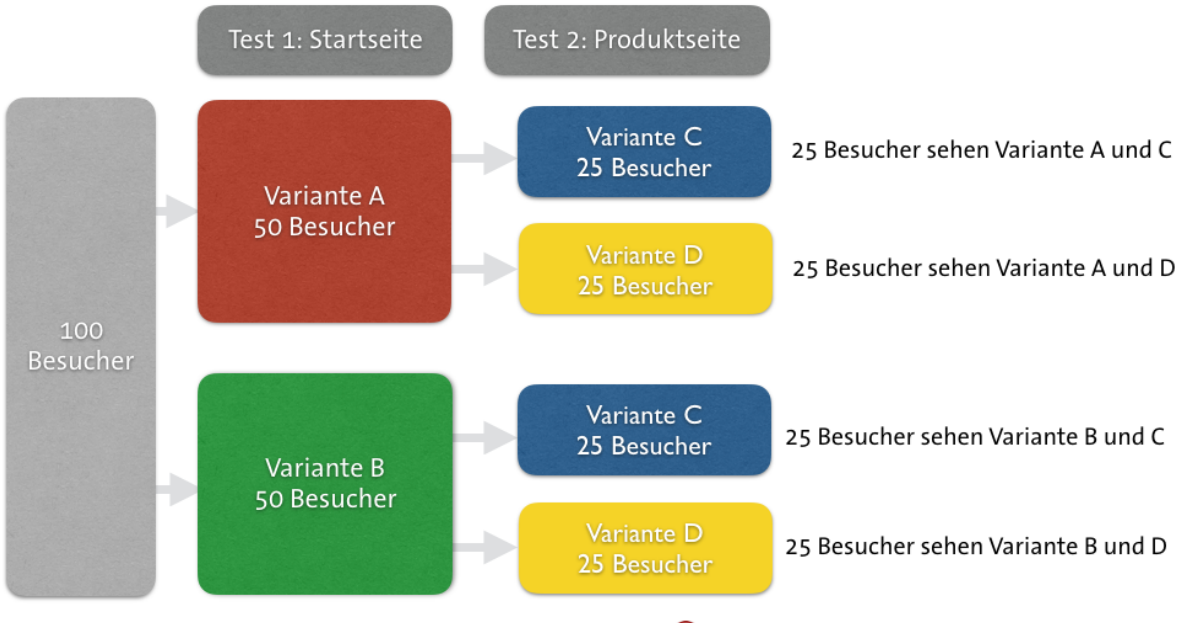


Abbildung 4: Der Anteil der Nutzer in zwei Tests ist bei einem sauberen Zufallsprinzip gleich hoch (Quelle: Optimizely).

2. Hohes Risiko von Interdependenzen

Es kann aber auch Tests geben, die besser nicht parallel laufen sollten. Vor allem bei Tests auf der gleichen Seite, oder schlimmer noch, auf den selben Elementen einer Seite läufst Du Gefahr, keine validen Daten zu bekommen.

Ein (Extrem-)Beispiel: In Test 1 möchtest Du die Platzierung der Produkt-Recommendations auf der Detailseite testen. In Test 2 veränderst Du den

Algorithmus der Reco-Engine, indem Du hier vor allem Sale-Artikel anzeigst. Kommt ein Nutzer in beiden Tests in die Testvariante, so kann die Kombination der Änderungen zu ganz anderen Effekten führen, als wenn der Nutzer nur eine der Änderungen sieht.

Bei stark interagierenden Tests bieten sich zwei Lösungen an:

Lösung A) Auf Nummer sicher gehen: Tests getrennt voneinander laufen lassen

Zum einen kann man in allen gängigen Testing Tools den Traffic auf mehrere Tests aufteilen. So könnte man zum Beispiel den Test zur Platzierung der Recommendations mit 50% Traffic testen und die anderen 50% für die Sale-Artikel in den Recommendations verwenden. Bei diesem Vorgehen wird sichergestellt, dass die Nutzer nur eines der laufenden Testkonzepte sehen. Beachte dabei, dass sich die Testlaufzeit für jeden Test entsprechend verlängert.
Ist ein 50/50 Split nicht möglich, kannst Du die Tests immer noch nacheinander laufen lassen. Dabei solltest Du stets priorisieren, von welchem Test Du mehr erwartest und dementsprechend die Tests in die Roadmap eintakten.

Lösung B) Multivariat Testen

Eine andere Möglichkeit besteht darin, die Testkonzepte in einen multivariaten Test zu kombinieren. Das ist sinnvoll, wenn die Testkonzepte auf das gleiche Goal optimieren und sie auf gleichen Seiten laufen, wie im Beispiel mit den Produkt-Recommendations. Das heißt, die Varianten müssen sinnvoll kombinierbar sein. Zielt der eine Test z.B. auf die Änderung der Main-Navi ab und der andere auf UVPs im Checkout, so macht das wenig Sinn, da es kaum Interaktionen zwischen den Konzepten gibt.

Statistik-Falle #3:

„ Wenn die **Klickraten steigen**, dann sehen wir das auch in der Conversion Rate.“

Wenn das nur so einfach wäre. Nur weil die Besuche auf den Produktseiten steigen und die Besucher öfter einen Artikel in den Warenkorb legen, heißt das nicht, dass sie am Ende auch kaufen.

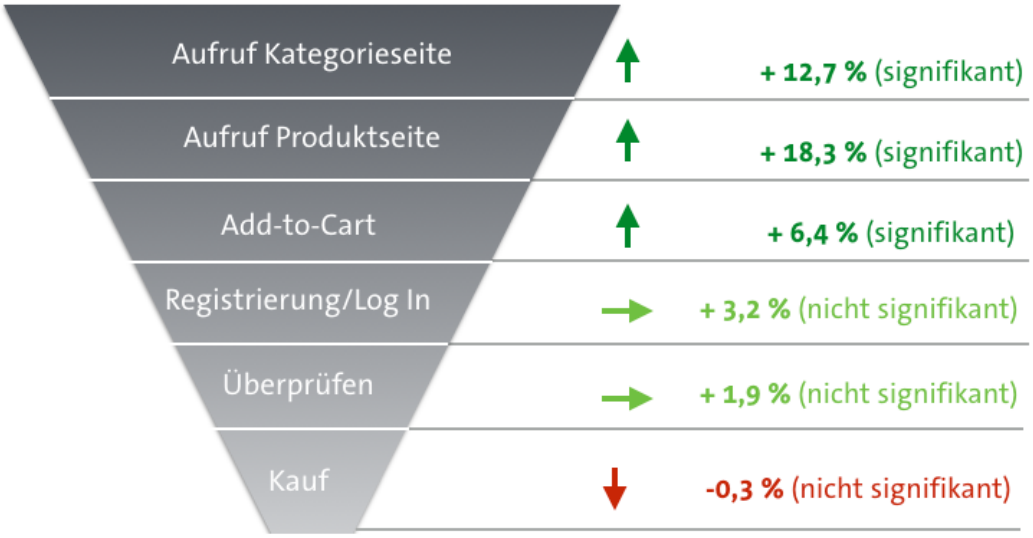


Abbildung 5: Im Laufe des klassischen Funnels können Uplifts verloren gehen.

Bei der Wahl der Key-Performance-Indicator (KPI) solltest Du folgendes beachten:

- **Die richtige Haupt-KPI wählen und priorisieren.**

In jedem Test sollte man immer die Macro-Conversions messen, anhand derer der Erfolg eines Testkonzepts beurteilt wird. Zum Beispiel die Conversion Rate oder Umsatzwerte. Welche die richtige KPI ist, hängt von der konkreten Testhypothese ab:

Conversion Rate:
In der Regel ist das die Haupt-KPI eines Onlineshops. Geht es also im Test darum, Nutzer in erster Linie in Käufer zu konvertieren, hat die Conversion Rate die höchste Priorität.

Warenkorbwerte/Revenue pro Besucher:
Jedoch kann es auch Konzepte geben, die in erster Linie darauf ausgelegt sind, den Warenkorbwert zu erhöhen. Eine Optimierung der Produktempfehlungen auf der Produktseite (Design, Platzierung, Algorithmus) führt nicht zwingend dazu, dass Leute überhaupt kaufen, sondern vermutlich eher dazu, dass sie mehr kaufen, wenn sie denn kaufen. Durch Empfehlungen von Komplementärprodukten (X-Sell) entsteht zum Beispiel ein Anreiz, die passende Jeans zum neuen Oberteil gleich dazu zu kaufen. Dann wäre der Umsatzwert pro Conversion oder pro Besucher relevant.

Retouren:
Testet man zum Beispiel, ob Angaben zur Passform („Fällt kleiner aus“) auf der Produktseite dabei helfen, die passende Größe zu finden, sollte man noch einen Schritt weiter gehen. Letztendlich hilft es nicht, den Kunden mit tollen Versprechungen zum Artikel zum Kauf zu bringen, wenn er dann unzufrieden

ist und alles retourniert. Der Umsatz nach Retoure oder die Retourenquote geben als Haupt-KPI Aufschluss darüber, ob man mit einem Konzept letztendlich auch wirklich mehr Geld verdient. Die Daten dazu lassen sich durch eine Verknüpfung zwischen Testing Tool und Data Warehouse (DWH) analysieren (siehe Statistik-Falle #8).

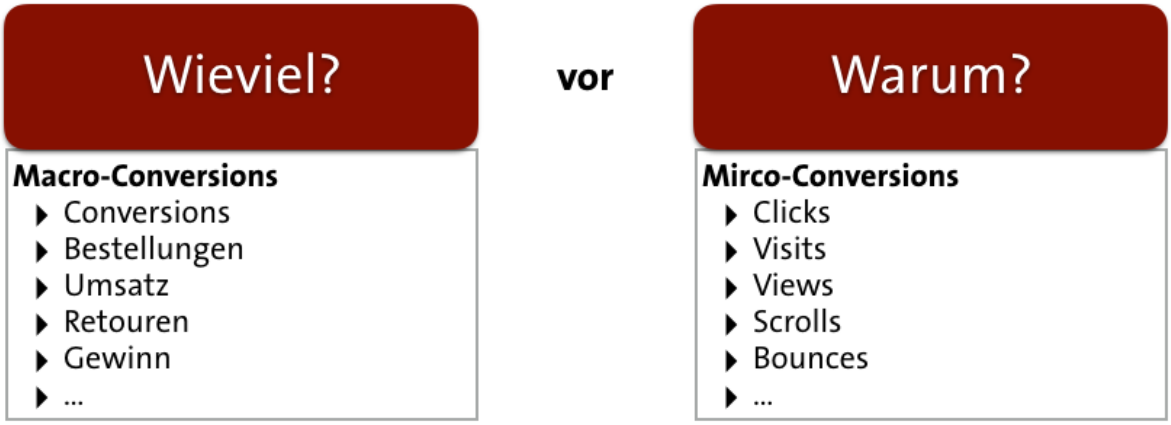


Abbildung 6: Macro-Conversions haben eine höhere Priorität als Micro-Conversions.

• Micro-Conversions nur als Diagnose-Instrumente

Das heißt natürlich nicht, dass Du Micro-Conversions wie Klickraten oder Seitenaufrufe nicht anschauen solltest. Sie sollten jedoch nicht dazu dienen, den Erfolg eines Konzepts letztendlich zu beurteilen sondern maximal als Diagnoseinstrumente.

Mal angenommen, Du testest eine optimierte Variante der Meta-Navigation und möchtest herausfinden ob die Kunden mehr kaufen. Überraschenderweise zeigen die Ergebnisse keinerlei Änderungen in der Conversion Rate oder dem Revenue. Verwenden die Kunden die neue Navigation genauso wie die alte? Oder andere Einstiegspunkte, wie die Side-Navi, Suche oder Kateogrie-Teaser?

In einem solchen Fall kann es hilfreich sein, wenn Du dir die Klickraten auf Navigationselemente oder die Nutzung der internen Suche in den Varianten anschaust. So kommst Du der Antwort nach dem „Warum“ ein Stück näher und kannst das Verhalten Deiner Kunden besser verstehen.

• Testgröße auf die KPI abstimmen

Oft beobachtet man in laufenden Tests, dass es nach relativ kurzer Zeit bereits signifikante Uplifts in den Micro-Conversions, wie Add-to-Basket oder Klicks gibt. In der Conversion Rate ist aber nichts zu sehen.

Tatsächlich gibt es einen Zusammenhang zwischen der Art der KPI und der benötigten Testgröße. Das liegt an den Schwankungen und der Unsicherheit, die je nach KPI variieren: Je höher diese Unsicherheiten sind, desto länger muss der Test laufen, um signifikante Effekte nachzuweisen.

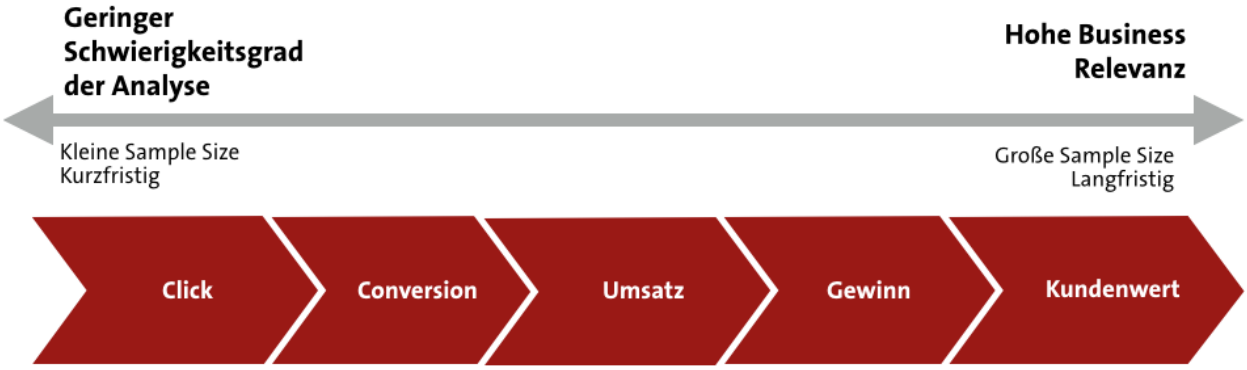


Abbildung 7: Je stärker eine KPI schwankt, desto länger muss der Test laufen.

Klickraten oder Seitenaufrufe sind Zähldaten und haben eine relativ geringe Schwankungsbreite: Klick oder kein Klick.
Bei der **Conversion Rate** sieht das schon anders aus. Ob jemand tatsächlich kauft oder nicht, wird von weiteren Faktoren beeinflusst, die dazu führen, dass die Unsicherheit über diese KPI steigt.
Um Effekte in den **Umsatz-KPIs**, wie dem Warenkorbwert nachzuweisen, benötigt man in der Regel (je nach Testkonzept) eine noch größere Sample Size als bei der Conversion Rate. Bei diesen sogenannten metrischen KPIs besteht zusätzlich Unsicherheit über die Höhe des Kaufs, wenn ein Nutzer denn überhaupt kauft.
Möchte man auf die **Umsatzwerte nach Retoure** optimieren, sollte man die Testgröße weiter erhöhen. Bei dieser KPI ist es mit am schwierigsten, verlässliche Aussagen zu treffen. Wenn jemand kauft, wieviel kauft er dann und wenn er retourniert, wieviel retourniert er dann? Die Unsicherheit steigt weiter was eine längere Testlaufzeit erforderlich macht.
Um sicherzustellen, dass Du tatsächlich verlässliche Ergebnisse bekommst, solltest Du Dir zuerst Gedanken darüber machen, welche KPI Du als Haupt-KPI des Tests siehst und dann die Testgröße entsprechend darauf abstimmen.

- Nicht zu viele Metriken messen

Je mehr KPIs man hat, desto schwieriger mag am Ende die Entscheidung sein. Dies lässt sich zwar durch eine vernünftige Priorisierung vereinfachen, jedoch solltest Du dir bereits vor dem Test Gedanken darüber machen, welche Macro- und Micro-Conversions Du wirklich benötigst. Misst man wahllos alles, was man im Tool messen kann, verliert man am Ende das Ziel aus den Augen. Eine sehr hohe Anzahl von Metriken führt außerdem dazu, dass die Wahrscheinlichkeit steigt, in einer Metrik einen Effekt zu finden, der rein zufällig entstanden ist.

Statistik-Falle #4:
„Auf keinen Fall als multivariaten Test. **Viel zu aufwendig** und die Ergebnisse sind sowieso **nicht valide**."

Kommt drauf an, wie man´s macht. Richtig aufgesetzt sind multivariate Tests ein tolles Werkzeug, um Faktoren und Kombinationen zu testen. Dass man mit einem MVT mehrere Faktoren auf einer Seite gleichzeitig testen kann, klingt kompliziert. Muss es aber gar nicht sein. Es gibt nur ein paar Regeln, die Du beachten solltest:

- Teste nicht zu viele Varianten

Auch bei einem MVT mit mehreren Varianten tritt das Problem der Alpha-Fehler-Kumulierung auf. Um zu verhindern, dass Du eine Variante zu einem Gewinner erklärst, obwohl sie gar keiner ist, solltest Du auch hier die Anzahl der Kombinationen so gut es geht begrenzen. Auf Basis einer guten Hypothesenbildung sollten Faktoren und deren Kombinationen mit Bedacht gewählt werden. Zudem kannst Du mit einem höheren Konfidenzniveau arbeiten (zum Beispiel 99,5 %), um sicher zu stellen, dass die Ergebnisse valide sind.

Variations	Revenue per Visitor	Percentage Improvement	Chance to Beat Original
Control		-	-
[redacted]		+32.44%	94%
[redacted]		+30.88%	96%
[redacted]		+27.31%	91%
[redacted]		+21.57%	89%
[redacted]		+20.70%	76%

Abbildung 8: Den Ergebnissen eines MVT sollte man nicht blind vertrauen.

- Werte die Testergebnisse richtig aus

Zuerst schaut man bei den Ergebnissen eines MVT natürlich darauf, welche Variante den höchsten (und signifikanten) Uplift erzielt hat. Dabei erhält man aber nur Informationen darüber, welche Kombination von Faktoren welchen Uplift erzielt. Darüber hinaus ist es aber wichtig, zu analysieren, wie hoch der Einfluss einzelner Faktoren auf die Conversion Rate ist. Das kann man mit Hilfe einer sogenannten Varianzanalyse machen. Diese isoliert den Effekt der

einzelnen Faktoren (im Beispiel sind das Farbe und Layout) auf die Conversion Rate.

- Validiere die Ergebnisse im Folgetest

Um Dein Vertrauen in die Ergebnisse Deines MVT zu erhöhen, kannst Du den ermittelten Testsieger auch mit einem anschließenden A/B-Test validieren. Die Gewinnerkombination lässt Du dann einfach gegen die derzeitige Control laufen. Die Fehlerwahrscheinlichkeit, einen falschen Gewinner zu deklarieren, reduziert sich so auf die üblichen 5%.

Section Report

Section	Impact	Variation	Conversion Rate Range	Percentage Improvement	Chance to Beat Original
Farbe	-	Control		-	-
		Farbe 1		-	-
		Farbe 2		-2.98%	5%
		Farbe 3		+0.21%	55%
Layout	29%	Control		-	-
		Layout 1		-	-
		Layout 2		+3.96%	98%
		Layout 3		+4.73%	99%

Abbildung 9: Mit Hilfe einer Varianzanalyse lassen sich Effekte einzelner Faktoren, z.B. Farbe und Layout, isoliert berechnen.

Phase 1: Der Test ist live

Statistik-Falle #5:
„Der Test **läuft schon drei Tage** und die Variante perfomt schlechter. Wir sollten den Test stoppen."

Klar, es mag gute Gründe geben, einen Test vorzeitig zu beenden. Zum Beispiel wenn Umsatz-Ziele durch eine signifikant schlecht performende Testvariante bedroht sind.

Um aber mit Hilfe eines Tests über den Erfolg oder Misserfolg eines Konzepts zu entscheiden, reichen drei Tage nicht aus. Jeder Test benötigt eine minimale Testgröße, um überhaupt etwas darüber sagen zu können, ob ein Konzept funktioniert oder nicht.
Gerade in den ersten Tagen nach dem Teststart beobachtet man im Tool oft ein wildes Hin und Her der Linien. Es kann sogar sein, dass die Testvariante schlechter gestartet ist. Vielleicht hat ein Kunde rein zufällig in der Control soviel bestellt, dass sie künstlich nach oben gezogen wird und der Test einen temporären Downlift aufweist (siehe Statistik-Falle #8).

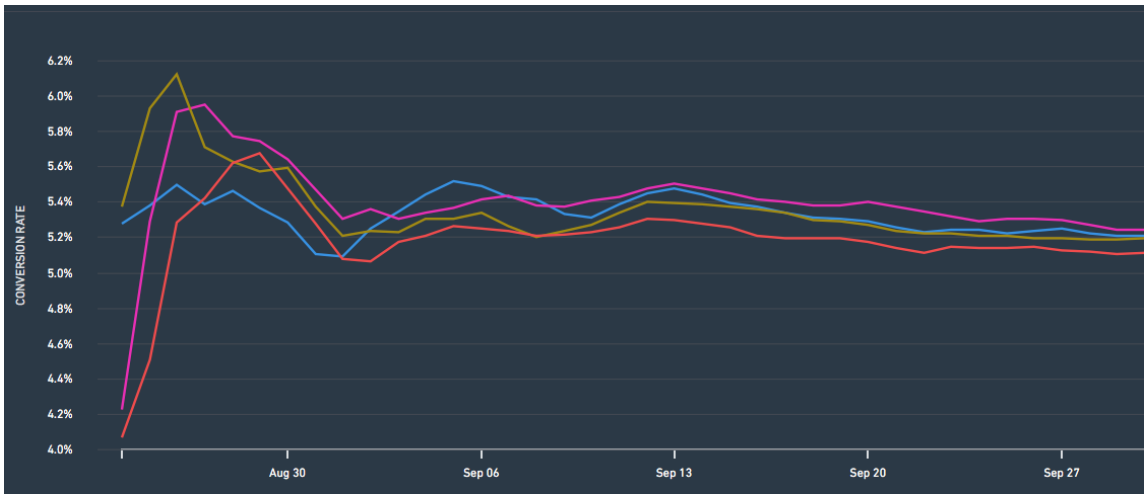


Abbildung 10: Gerade zu Beginn eines Tests schwanken die Testergebnisse sehr stark. Diesen Zahlen solltest Du nicht trauen.

Schon oft habe ich bei Tests beobachtet, dass sich anfangs negative Effekte mit der Zeit umdrehen und die Testvariante nach drei Wochen einen signifikant positiven Effekt zeigt. Schon vor dem Teststart kannst Du mit Hilfe unseres Tools zur Testdauerberechnung schätzen, wie lange ein Test laufen muss (die Excel-Datei kannst Du am Ende des Artikels herunterladen).
Zeigt ein Test in der Anfangszeit keine signifikanten Downlifts, heißt es bis dahin schlichtweg: Keep calm and continue testing.

A/B Test Mindestlaufzeit / Mindestconversion-Anzahl

Im Sheet "Guideline" findest du mehr Informationen dazu, welche Informationen du hier eingeben musst und woher du diese bekommst.

Dein Input		Dein Output	
Aktuelle Conversion Rate der Testseite:	5%	Benötigte Anzahl Visitors pro Variante	29.418
Erwarteter Uplift in %:	9%	Benötigte Anzahl Visitors gesamt	58.836
Anzahl aller Varianten (inkl. Control):	2	Benötigte Anzahl Conversions pro Variante	1.471
Conversions/Monat:	4.000	Benötigte Anzahl Conversions gesamt	2.942

Test-Settings:

Power	80%
Konfidenz	95%

Mindestlaufzeit deines Tests

22 Tage

Abbildung 11: Vor jedem Test solltest Du eine Sample Size Schätzung durchführen.

Statistik-Falle #6:

„Kein Problem. Wenn die eine **Variante** nicht funktioniert, **ändern** wir sie oder **schalten sie einfach ab.**“

Ich bin sicher, das passiert ständig und manchmal gibt es auch sehr gute Gründe dafür. Performt eine Testvariante über mehrere Tage so schlecht, dass relevante Business Ziele bedroht sind, ist es sinnvoll diese Variante aus dem Rennen zu nehmen.
Oft werden Varianten auch mit geringem Traffic-Anteil gestartet, welcher dann über die Zeit hinweg erhöht wird (Ramp-Up). Oder Varianten werden mitten im Test verändert.
Alle drei Aspekte können die Ergebnisse verzerren: Beim A/B-Testing sind die Testergebnisse repräsentativ für den gesamten Testzeitraum, und zwar in gleichen Teilen. Schraubt man zum Beispiel am Traffic, führt das dazu, dass gewisse Zeiträume über- oder unterrepräsentiert sind. Den gleichen Effekt hat das Abschalten einer Variante mitten im Test. Verändert man den Content einer Variante, ist das Testergebnis eine skurrile Mischung aus unterschiedlichen Hypothesen und Konzepten, die am Ende gar keine Aussage mehr liefert.
Hier muss man eine gesunde Mischung aus wissenschaftlichem Anspruch an Validität und der praktischen Business Relevanz finden.

Folgende Tipps kannst Du mitnehmen:

- Wird eine Variante bereits in den ersten Tagen nach Teststart abgeschaltet, empfiehlt es sich, den kompletten Test einfach ohne die Variante neu starten. Hier hält sich der Zeitverlust in Grenzen.
- Bevor Du Varianten während des Tests veränderst, starte lieber einen neuen Test und teste die angepasste Variante gegen die Control.
- Schraubst Du den Traffic von Tag zu Tag höher, dann stelle sicher, dass Du einen ausreichenden Testzeitraum mit der von dir gewünschten maximalen Traffic-Menge abdeckst, in dem der Traffic stabil über die Zeit bleibt.

Statistik-Falle #7:

„Mit den neuen **bayesianischen Testverfahren** haben wir viel **schneller** Ergebnisse."

Die neuen Statistik-Verfahren, die bereits von VWO und Optimizely eingesetzt werden, bieten vor allem den Vorteil, dass man zu jeder Zeit seine Testergebnisse interpretieren kann und valide Ergebnisse hat, auch wenn man die benötigte Testgröße noch nicht erreicht hat.

Jedoch schützen die bayesianischen Verfahren auch vor falscher Methodik nicht:

1. Das Problem mit den Erwartungen

Bayesianische Verfahren setzen voraus, dass man vor dem Teststart eine sogenannte a priori Annahme treffen muss, mit welcher Wahrscheinlichkeit das Testkonzept denn erfolgreich sein wird. Liegt man in seinen Annahmen falsch, riskiert man lange Testlaufzeiten und invalide Ergebnisse.

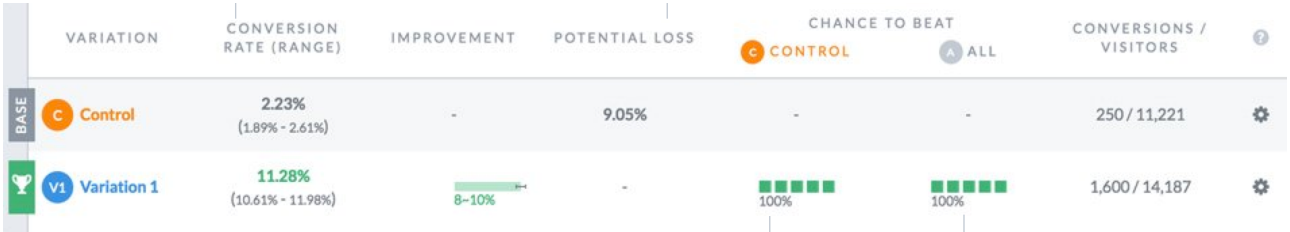


Abbildung 12: Das VWO Reporting Interface.

2. Das Problem mit der geringen Testgröße

Die Ergebnisse sind zwar zu jedem Zeitpunkt während des Tests interpretierbar. Aber gerade zu Anfang, wo die Schwankungen noch sehr hoch sind, ist Vorsicht geboten. Die Anzahl der Visitors im Test ist noch sehr klein und wird von außergewöhnlichen Beobachtungen (z.B. sehr hohen Bestellwerten) stark beeinflusst. Denn trotz der Vorteile, die der bayesianische Ansatz bietet, gilt auch hier, dass die Methode bei geringen Datensätzen keine ausreichende

Validität bietet. Die neueren Verfahren schützen also auch vor falscher Vorgehensweise beim A/B-Testing nicht (das Gleiche gilt im übrigen auch für die traditionellen Verfahren!).

Phase 2 - Nach dem Test: Auswertung

Statistik-Falle #8:

„Die **Ergebnisse des Testing Tools reichen völlig** aus, um eine Entscheidung zu treffen."

Klar. Mit dem Testing Tool findet man heraus, ob ein getestetes Konzept zu einer höheren Conversion Rate führt oder nicht. Allerdings lohnt sich oft auch ein detaillierter Blick in die Testergebnisse um mehr Insights zu generieren und Ergebnisse zu validieren. Ergebnisse aus dem Testing Tool sollten nicht in einem Datensilo gelagert, sondern mit anderen Datenbanken, wie dem Webanalytics System oder dem DWH kombiniert werden.
Damit kann man eine Reihe wichtiger Fragen beantworten:

1. Warum?

Gerade wenn die Ergebnisse eines Tests nicht so sind wie erwartet, kommt sie, die Frage nach dem Warum. Wie in dem oben beschriebenen Beispiel zum Test der Meta-Navi (siehe Statistik-Falle #3) kann es sinnvoll sein, sich Micro-Conversions anzuschauen um Testergebnisse besser zu verstehen. Hat man vor dem Teststart versäumt, Klicks auf Navigationselemente oder die Nutzung der internen Suche als Goal einzurichten, lassen sich solche Fragen im Nachhinein nicht mit dem Testing Tool beantworten. Ist Dein Testing Tool hingegen mit dem Webanalytics Tool verknüpft, kannst Du dir diese Metriken anschauen.

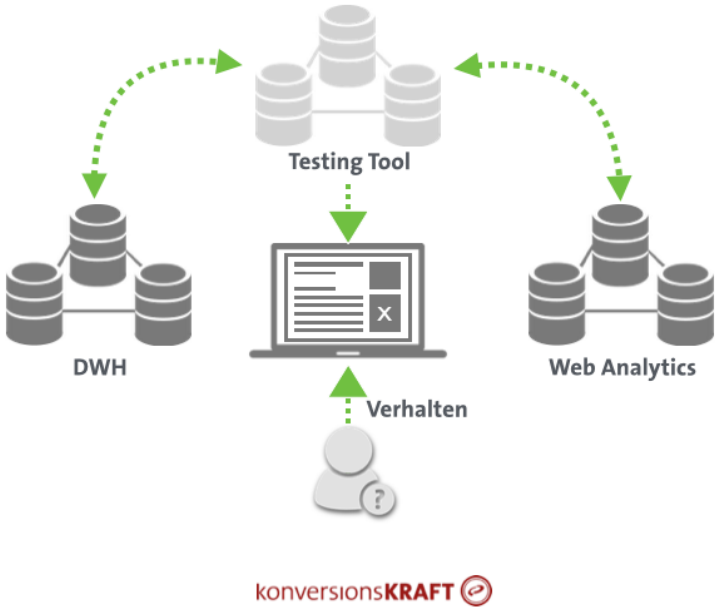


Abbildung 13: Für ein umfassendes Kundenverständnis hilft eine gute Verbindung von Datenquellen.

2. Führt der Conversion Rate Uplift auch zu mehr Gewinn?

Wenn Kunden aufgrund eines Optimierungskonzepts mehr bestellen, ist das gut. Wenn sie jedoch genauso viel oder noch mehr retournieren, kann das am Ende sogar einen Verlust bedeuten. Eine Verbindung zwischen Testing Tool und DWH ist also auch aus dem Grund sinnvoll, um das Ergebnis nach Retouren zu analysieren. Hier werden die Varianten IDs ins Backend übergeben und Du kannst dir die Retourenquoten oder die Netto-Umsätze als zusätzliche KPI in der Test- und Controlvariante anschauen.

3. Berücksichtigung von Vielbestellern

Fast jeder Online Shop hat sie, und sie verursachen in der Regel Probleme bei der validen Auswertung eines Tests: Die Vielbesteller. Dazu können Call-Center, Kunden aus dem B2B-Bereich oder auch exzessive Online-Shopper gehören. Diese sogenannten Ausreißer verzerren die Ergebnisse. Warum ist das so? Nun ja, bei der Auswertung von Testergebnissen schaut man in der Regel auf Mittelwerte:

- Die Conversion Rate ist nichts anderes, als die **durchschnittliche Bestellrate** einer Gruppe von Besuchern.
- Der Revenue per Visitor ist der **durchschnittliche Bestellwert** pro Besucher.

Außergewöhnlich hohe Werte bei der Anzahl der Bestellungen oder den Warenkorbwerten ziehen diese Mittelwerte nach oben. Schon wenn die Control rein zufällig einige Kunden mit extrem hohen Werten aufweist, kann das bei relativ geringer Testgröße dazu führen, dass Ergebnisse nicht mehr positiv, sondern vielleicht sogar signifikant negativ werden.

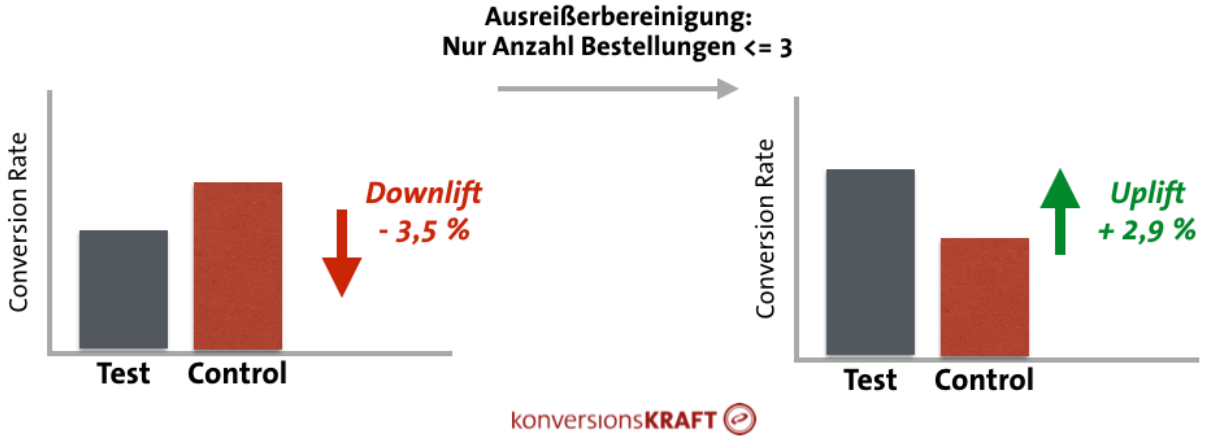


Abbildung 14: Durch eine Ausreißerbereinigung können sich Testergebnisse ändern.

Was tun? Bei allen gängigen Testing Tools kann man Dateien mit den Rohdaten erhalten. Hier sind alle Bestellungen der Kunden aufgelistet. Eine andere

Möglichkeit besteht darin, entsprechende Reports im angeknüpften Webanalytics Tool anzulegen und abnormal hohe Bestellungen durch entsprechende Filtersetzung auszuschließen.

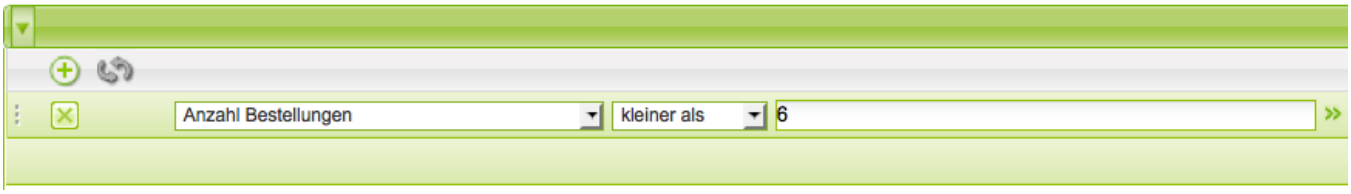


Abbildung 15: Filter für Vielbesteller im Analytics Tool einrichten.

Auch wenn das mit ein wenig Aufwand verbunden ist, lohnt sich der Blick in die Rohdaten, da man so oft signifikante Effekte entdeckt, die Durch die Ausreißer „versteckt" waren.

4. Für die Daten-Freaks: Valide Berechnung von Konfidenzintervallen

Klassische Methoden zur Berechnung von Konfidenzintervallen haben ein Problem: Sie unterstellen, dass die zugrunde liegenden Daten einer bestimmten Verteilung, nämlich der Normalverteilung, folgen. Die linke Grafik zeigt eine perfekte (theoretische) Normalverteilung. Die Anzahl der Bestellungen schwankt um einen positiven Mittelwert. Der Großteil der Kunden bestellt im Beispiel 5 mal. Weniger oder mehr Bestellungen kommen weniger oft vor. Die rechte Grafik zeigt die Realität.

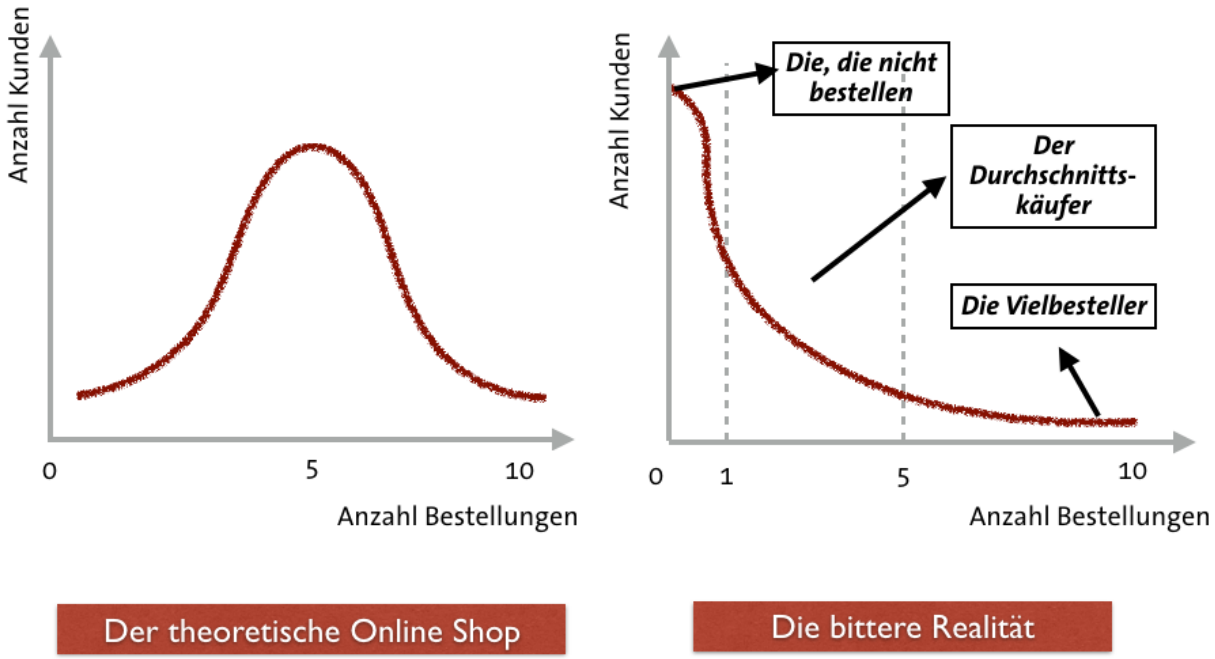


Abbildung 16: Theorie und Praxis.

Bei einer durchschnittlichen Conversion Rate von 5% bleiben 95% Kunden, die nicht kaufen. Der Großteil der Käufer hat wahrscheinlich eine oder zwei Bestellungen getätigt und dann gibt es noch die wenigen Kunden, die extrem viel bestellen.
Solche Verteilungen nennt man „rechtsschief“ (right skewed) und sie haben vor allem bei geringer Testgröße einen Einfluss auf die Validität der Konfidenzintervalle. Diese sind dann nicht mehr verlässlich zu berechnen. Der zentrale Grenzwertsatz führt bei sehr großen Stichproben dazu, dass diese Verzerrungen weniger ins Gewicht fallen. Aber wie „falsch“ die Konfidenzintervalle tätsächlich sind, hängt auch davon ab, wie stark die Daten von einer perfekten Normalverteilung abweichen. Da in einem durchschnittlichen Onlineshop mindestens 90% der Kunden nicht kaufen, ist der Anteil der „Nullen“ im Datensatz extrem und die Abweichungen in aller Regel enorm (mit dem Shapiro-Wilk-Test kannst Du deine Daten auf Normalverteilung testen).
Aufgrund dieser Probleme lohnt sich ein Blick abseits des klassischen t-Tests. Es gibt weitere Methoden, die bei nicht-normalverteilten Daten verlässliche Ergebnisse liefern:

- U-Test
  Der Mann-Whitney U-Test (=Wilcoxon-Rangsummentest) ist eine Alternative zum t-Test, wenn die Daten stark von der Normalverteilung abweichen.
- Robuste Statistik
  Methoden aus der robusten Statistik werden verwendet, wenn Daten nicht normalverteilt sind oder durch Ausreißer verzerrt werden. Hier werden Mittelwerte und Varianzen so berechnet, dass sie nicht von außergewöhnlich hohen oder niedrigen Werten beeinflusst werden.
- Bootstrapping
  Dieses sogenannte nicht-parametrische Verfahren arbeitet unabhängig von jeder Verteilungsannahme und liefert verlässliche Schätzungen für Konfidenzniveau und Intervalle. Im Kern gehört es zu den Resampling-Methoden, welche auf Basis der beobachteten Daten durch Stichprobenziehungen die Verteilungen von Variablen verlässlich schätzen.

Statistik-Falle #9:

„Ich schaue mir jetzt mal die **Ergebnisse in den Segmenten** an. Irgendwo finde ich bestimmt einen Uplift.“

Generell ist das schon eine gute Idee. Die Ergebnisse eines A/B-Tests werden in der Regel über alle Besucher reportet. Dabei kann es aber Kundengruppen geben, die besonders gut oder schlecht auf ein bestimmtes Testkonzept reagieren.

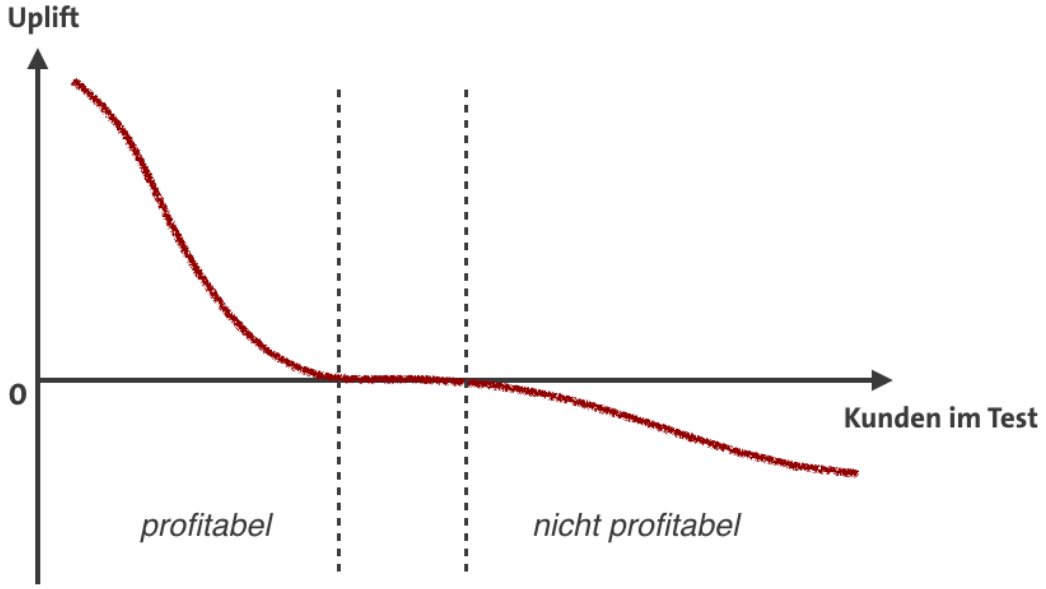


Abbildung 17: Segmentierung in profitable und nicht profitable Kundengruppen.

Ein beliebtes Beispiel sind Unique Value Propositions (UVP), also Merkmale, die einen Online Shop einzigartig machen. Oft beobachtete man, dass es hier unterschiedliche Effekte bei Neu- und Bestandskunden gibt. Während Bestandskunden die Vorzüge eines Kaufs schon kennen, helfen UVPs dabei, Neukunden vom Kauf zu überzeugen und Vertrauen aufzubauen. Diese unterschiedlichen Reaktionen werden mit den aggregierten Reports in der Regel aber nicht aufgedeckt und man ist enttäuscht ,weil sich overall kein signifikanter Uplift nachweisen lässt.

Durch eine Verbindung von Testing Tool mit anderen Datenquellen (Web Analytics, DWH) kannst Du die Testergebnisse für eine Reihe von Kundeneigenschaften analysieren, die in deinem Testing Tool nicht verfügbar sind, zum Beispiel:

- Geschlecht
- Alter
- Kategorien, in denen jemand kauft
- Besuchte Seiten
- Daten zum Klickverhalten
- Retourenquote
- Geolocation
- ...

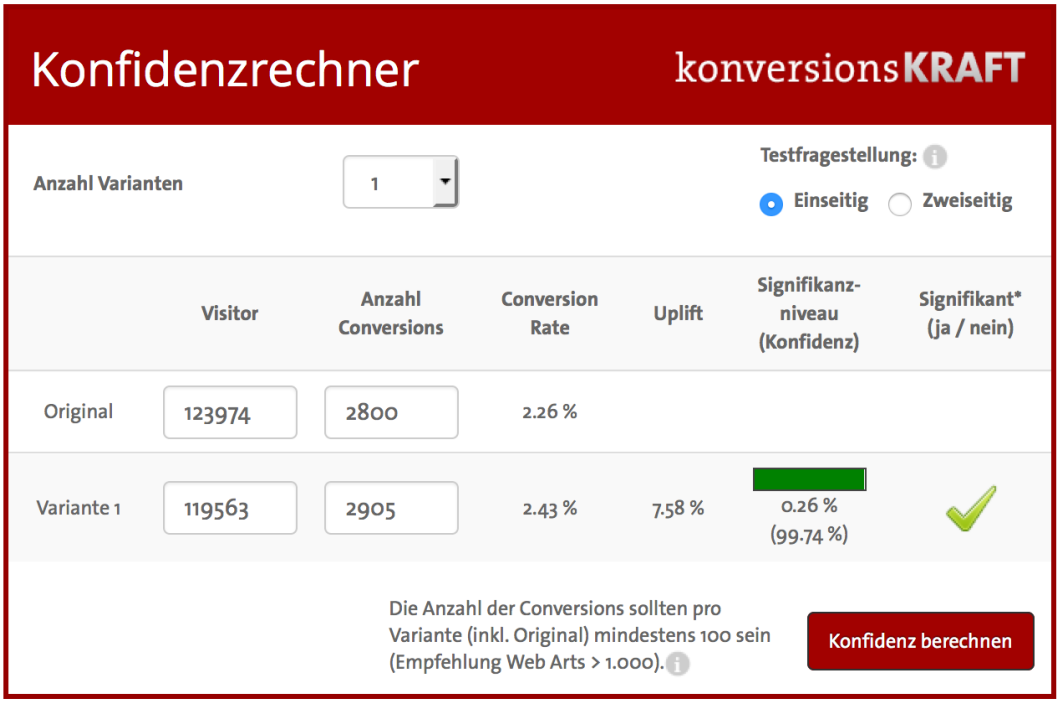


Abbildung 18: Auch bei Ergebnissen in den Segmenten ist eine Konfidenzanalyse Pflicht.

Allerdings ist hier Vorsicht geboten. Die Fehlerwahrscheinlichkeit, Stichwort Alpha-Fehler Kumulierung) erhöht sich exponentiell, je mehr Segmente man miteinander vergleicht. Folglich solltest Du vermeiden, wahllos die Segmente zu durchforsten. Sie sollten immer interpretierbar bleiben und auch relevant für das Testkonzept sein.
Außerdem gilt es, darauf zu achten, dass die Segmente ausreichend groß sind. Je spezieller man Segmente definiert, desto geringer wird die Sample Size und die Verlässlichkeit der Ergebnisse. Schaut man sich zum Beispiel nur Besucher an, die über ihr Tablet auf einer Kategorieseite einsteigen, danach auf eine Produktseite gehen, weiblich sind und vor allem am Wochenende kaufen, führt das dazu, dass man nur einen Bruchteil aller Besucher anschaut. Achte also stets darauf, dass die Testgröße auch in den Segmenten ausreichend ist. Weißt Du schon vor dem Test, dass Du eine Segmentierung vornehmen wirst, empfiehlt es sich, die Testdauer von vorne herein mindestens zu verdoppeln. Mit dem konversionsKRAFT Signifkanzrechner (www.konversionskraft.de/tools) kannst Du prüfen, wie sicher die Ergebnisse in dem jeweiligen Segment sind.

Statistik-Falle #10:
„Der Test war **wohl doch nicht so erfolgreich**. Wir haben das Konzept ausgerollt, **die Conversion Rate steigt aber nicht**."

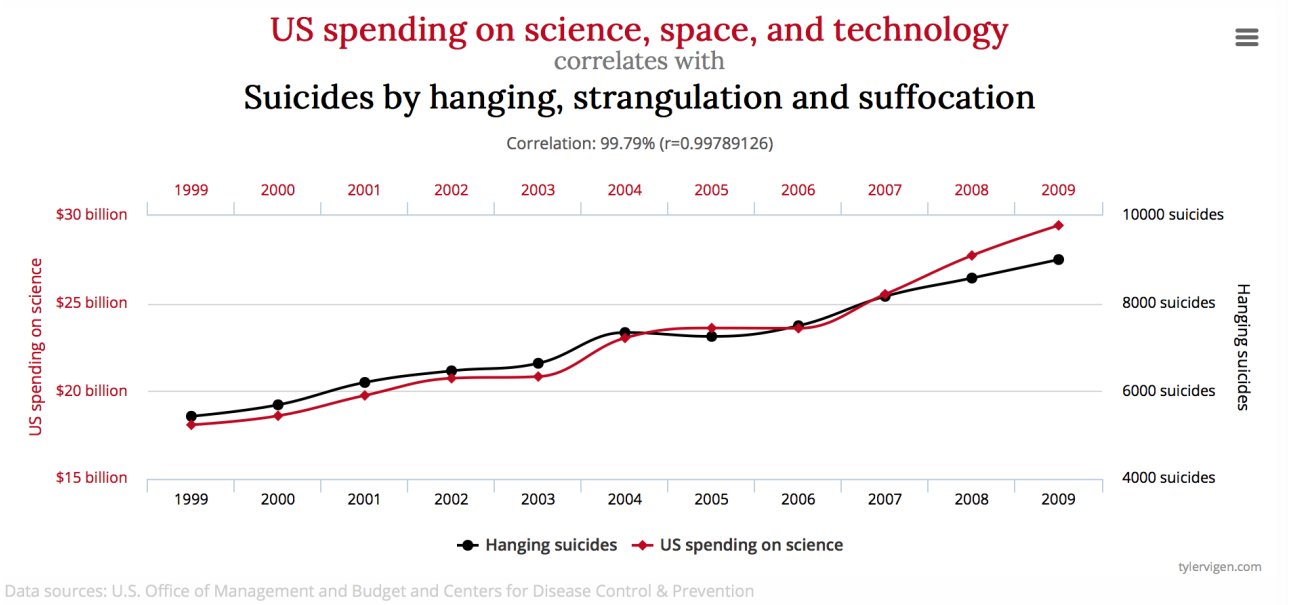
Oft werden auf Basis von Testergebnissen Hochrechnungen erstellt, wie viel zusätzlicher Umsatz in der Zukunft damit erzeugt wird. Heraus kommen schöne Folien, auf denen suggeriert wird, dass durch das eine Testkonzept mit einem Umsatzplus von 40% in den nächsten zwei Jahren zu rechnen ist. Klingt toll, aber ist das wirklich so?
Sorry für die harten Worte, aber: Bitte tu es nicht! Das Ergebnis einer solchen Hochrechnung ist so verlässlich, als würde ich sagen, dass es heute in drei Monaten regnen wird. Das würdest Du auch nicht glauben oder? Das Ergebnis von Tests einfach linear in die Zukunft zu prognostizieren funktioniert aus folgenden Gründen nicht:

Grund 1: Kurz - vs. langfristige Effekte
In der Regel misst man beim A/B-Testing, ob das Konzept zu kurzfristigen Änderungen im Kundenverhalten führt. Der Test läuft drei Wochen und man sieht, dass sich dadurch die Conversion Rate um x% erhöht. Soweit so gut. Jedoch sagt diese Änderung nichts über den Effekt auf das langfristige Kundenverhalten und KPIs wie Kundenzufriedenheit und Kundenbindung aus. Dafür müsste der Test weitaus länger laufen, um zum Beispiel Gewöhnungseffekte oder Änderungen im Kundenstamm und den Nutzerbedürfnissen zu erfassen. In einem solchen Test konnte ich beobachten, dass der Effekt kurzfristig sehr hoch war, sich aber nach mehreren Monaten auf ein deutlich niedrigeres, aber dennoch signifikantes Niveau eingependelt hat. Nutzer gewöhnen sich an Neuheiten: Was heute noch begeistert hat, ist morgen schon Gewohnheit und wird nicht mehr so stark wahrgenommen. Folglich ist es schlichtweg falsch, wenn man den gemessenen kurzfristigen Effekt als konstant annimmt und in die Zukunft fortschreibt.

Grund 2: Kausalität vs. Korrelation
Angenommen, der Test hat einen signifikanten Uplift und das Testkonzept wird fest implementiert. Was dann oft passiert sind sogenannte Vorher-Nachher-Vergleiche mit dem Ziel, den gemessenen Effekt auch in der Conversion Rate nachzuweisen. Hier wird die Conversion Rate im Zeitraum vor der Implementierung mit einem Zeitraum nach der Implementierung verglichen. Man erwartet, dass die Differenz der beiden Raten ja genau dem gemessenen Effekt aus dem Test entsprechen muss.
Klar, livegestellte Änderungen können durchaus zu Steigerungen in der Conversion Rate führen. Jedoch gibt es parallel noch hunderte andere Einflussfaktoren, die ebenfalls die Conversion Rate bestimmen, z.B. Saison, Sale Aktionen, Lieferschwierigkeiten, neue Produkte, andere Kunden oder Bugs.

Es ist wichtig, zwischen Korrelation und Kausalität zu unterscheiden. Korrelation gibt lediglich an, inwiefern zwei Kennzahlen einem gemeinsamen Trend folgen. Jedoch sagt Korrelation nichts darüber aus, ob die eine Kennzahl kausal, also ursächlich für die Änderung der anderen Variable ist.
Dieser Trugschluss wird in Abbildung 18 deutlich: Die US-Ausgaben für Science, Space and Technology steigen und gleichzeitig bringen sich deshalb mehr Leute um. Macht irgendwie keinen Sinn oder?
Abbildung 18: Hier findest Du weitere skurrile Beispiele die den Unterschied zwischen Korrelation und Kausalität deutlich machen: http://tylervigen.com/spurious-correlations

Beobachtet man also nach einem Livegang Änderungen in der Conversion Rate, so lässt sich dieser Effekt nicht isolieren. Man weiß nicht genau, wie stark welche Faktoren ursächlich für die Conversion Rate sind. Angenommen, das livegestellte Testkonzept hat einen positiven Effekt, dann kann es darüber hinaus aber andere negative Einflussfaktoren geben, so dass der Effekt insgesamt nicht mehr nachweisbar ist.
Die einzig saubere Möglichkeit, kausale Zusammenhänge und den Effekt eines



Testkonzepts langfristig zu messen, ist folgende: Nach dem erfolgreichen Test rollst Du das Konzept für 95% der Kunden aus (zum Beispiel über Dein Testing Tool). Die restlichen 5% belässt Du als Kontrollgruppe. Durch einen ständigen Vergleich dieser beiden Gruppen kannst Du den langfristigen Effekt Deines Konzepts messen. Wenn diese Methode für Dich nicht praktikabel ist, findest Du hier weitere Tipps, um zwischen Korrelation und Kausalität zu unterscheiden.

Grund 3: Missinterpretation von Konfidenz

Ein weiteres Problem ist, dass es leider immer noch zu Missverständnissen bei der Interpretation von Konfidenzen kommt. Angenommen, der Test zeigt einen Uplift von 4,5 % bei einem Konfidenzniveau von 98%. Das heißt nicht, dass der Effekt mit einer Wahrscheinlichkeit von 98% genau 4,5% beträgt! Jede Konfidenzanalyse liefert als Output ein Intervall, in dem der zu erwartende Uplift mit einer bestimmten Wahrscheinlichkeit (Konfidenz) liegt. Im Beispiel könnte das bedeuten, dass der Effekt auf Basis der gemessenen Werte mit einer Wahrscheinlichkeit von 98% zwischen 2% und 7% liegt. Es ist daher ein Trugschluss, wenn man davon ausgeht, dass der tatsächliche Effekt genau dem aus dem Test entspricht. Dieses Intervall wird zwar umso kleiner, je länger der Test läuft, es wird jedoch nie auf eine genaue Punktschätzung hinauslaufen. Das Konfidenzniveau sagt viel mehr etwas darüber aus, wie stabil das Ergebnis ist.

Übrigens: Es gibt einen kleinen aber feinen Unterschied zwischen der Konfidenz und der Chance-to-Beat-Original, welche die meisten Testing Tools reporten: Die Konfidenz gibt lediglich an, mit welcher Wahrscheinlichkeit eine Variable innerhalb eines bestimmten Intervalls (dem Konfidenzintervall) liegt und sagt erstmal nichts darüber, ob ein Konzept erfolgreich ist oder nicht. Die CTBO hingegen misst, ob und wie stark sich Konfidenzintervalle überlappen und wie wahrscheinlich es ist, dass die Testvariante irgendwie besser ist, als die Control. Diesen Unterschied solltest Du kennen. Mehr darüber erfährst Du hier.

Fazit

Wenn man sich bei der Auswertung eines Tests rein auf die Zahlen aus dem Testing Tool verlässt, kann das zu Fehlern führen. Diese wiederum gefährden die Validität deiner Ergebnisse. Um das zu vermeiden, gibt es für jede Phase des Tests ein paar Grundregeln, die man einhalten sollte.
Beim Thema A/B-Testing gibt es dabei aber auch einen regelmäßigen Konflikt zwischen dem wissenschaftlichen Anspruch an die Sauberkeit eines Tests und den Business Bedürfnissen. Du solltest versuchen, einen guten Mittelweg zu finden, so dass Deine Tests immer noch praktisch relevante Ergebnisse liefern und das Vertrauen in die Testergebnisse trotzdem bestehen bleibt.

37 Statistik-Tipps für Optimierer

#1 Anzahl der Testvarianten

1. Eine 100%ige Sicherheit gibt es nicht. Mit dem Konfidenzniveau bestimmst Du, welche Fehlerwahrscheinlichkeit Du akzeptieren willst.
2. Bei mehr als 1 Testvariante kommt es zur Alpha-Fehler-Kumulierung.
3. Es gibt Testing Tools, die dafür korrigieren.
4. Mit einer Varianzanalyse kannst Du den Fehler selbst korrigieren.
5. Einfacher geht es mit einer Reduzierung der Testvarianten (max. 3 + Control).

#2 Paralleles Testen

6. Sind keine hohen Interdependenzen zwischen Tests zu erwarten, können Tests durchaus gleichzeitig laufen.
7. Bei einem hohen Risiko von asymmetrischen Effekten hingegen können Ergebnisse verzerrt werden. In diesem Fall empfehlen wir einen Split oder zeitlich versetzte Tests.
8. Haben Tests die gleichen Goals und sind sinnvoll miteinander kombinierbar, kann multivariates Testen eine gute Lösung sein.

#3 Die richtigen KPIs

9. Mache dir VOR jedem Test Gedanken über die richtige KPI.
10. Der Erfolg sollte immer anhand von Macro-KPIs beurteilt werden (Conversion oder (Netto-)Umsatz).
11. Bei mehreren KPIs ist eine Priorisierung notwendig.
12. Wähle nicht zu viele KPIs.
13. Nutze Micro-Conversions als Diagnoseinstrumente.
14. Je höher die Varianz von KPIs, desto länger muss der Test laufen.

#4 Multivariates vs. A/B-Testing

15. Multivariate Tests sind super, wenn sie richtig gemacht werden.
16. Begrenze die Anzahl an Varianten und wähle die Kombinationen in Abstimmung auf dein Testkonzept.
17. Durch eine Varianzanalyse kannst du den Effekt einzelner Faktoren auswerten.
18. Validiere die Ergebnisse mit einem Folgetest, um die Fehlerwahrscheinlichkeit zu reduzieren.

#5 Testlaufzeit

19. Jeder Test benötigt eine minimale Testgröße, um valide Ergebnisse zu erzielen.
20. Berechne die Testlaufzeit vorab, z.B. mit unserem Kalkulator.
21. Gib dem Test genügend Zeit und schalte ihn nicht vorher einfach ab.

#6 Änderungen während des Tests

22. Änderst Du Varianten oder schaltest diese ab, so macht es Sinn, den Test neu zu starten, bevor die Ergebnisse verzerrt werden.
23. Schraube nicht zu viel am Traffic und wenn doch, erweitere den Testzeitraum, um eine ausreichende Testgröße zu bekommen.

#7 Bayesianische Verfahren

24. Treffe vor Teststart realistische „a priori“ Annahmen zum Erfolg des Testkonzepts.
25. Liegst Du mit den Annahmen daneben, riskierst Du lange Testlaufzeiten und invalide Ergebnisse
26. Vertraue einer geringen Testgröße nie, denn sie sind auch bei bayesianischen Verfahren häufig nicht valide.

#8 Testauswertung

27. Kombiniere Testergebnisse mit anderen Datenbanken.
28. Analytics-Tools können im Nachhinein Auskunft über Metriken geben, die im Testing Tool vergessen wurden.
29. Im DWH kannst Du Retouren analysieren und und sehen, ob der Uplift auch zu mehr Gewinn führt.
30. Mit Hilfe von Rohdaten kannst du Ausreißer identifizieren und z.B. abnormal hohe Bestellungen ausschließen.
31. Daten entsprechen häufig nicht der Normalverteilung, sondern sind „rechtsschief“ und verfälschen Konfidenzintervalle. Daten abseits der Normalverteilung kannst Du mit verschiedenen Methoden auswerten.

#9 Segmentierung

32. Die Verbindung mit anderen Datenquellen ermöglicht eine Analyse der Testergebnisse nach Kundeneigenschaften.
33. Achte beim Vergleich der Segmente auf eine ausreichende Stichprobengröße, um Fehler zu vermeiden.
34. Wähle die Kundeneigenschaften mit Bedacht und stimme sie auf Deine Hypothese ab.

#10 Implementierung

35. Projiziere das Testergebnis nicht einfach linear auf die Zukunft, denn A/B-Tests messen in der Regel kurzfristige Verhaltensänderungen. Gewöhnungseffekte, Änderungen in Kundenstamm und Nutzerbedürfnissen werden nicht berücksichtigt.
36. Den isolierten Effekt auf die Conversion Rate langfristig kannst du mit einer dauerhaften 95%/5% - Ausspielung messen.
37. Interpretiere die Konfidenz nicht falsch und verwechsle sie nicht mit der CTBO.

// konversionsKRAFT Trainings

Hier geht es zu den konversionsKRAFT Trainings 2016 >>
http://www.konversionskraft.de/ausbildung-conversion-manager

// konversionsKRAFT Webinare

Aktuelle Webinare hier >>
http://www.konversionskraft.de/webinare

// Autor

Dr. Julia Freese ist Head of Data Analytics bei der Web Arts AG. Nach der Promotion im Bereich Makroökonomie war Julia Freese bei der Zalando SE in Berlin im Bereich Data Intelligence und Conversion Optimierung tätig.

Bei konversionsKRAFT (Web Arts AG) kümmert sie sich neben der strategischen Kundenberatung um die Themen Webanalytics, Big Data und Personalisierung.

// Über konversionsKRAFT

konversionsKRAFT (Web Arts AG) ist führender Dienstleister im Bereich der Conversion- und Landing Page Optimierung. Die 1996 gegründete Agentur mit 50 Mitarbeitern ist Veranstalter des ConversionSUMMIT, Herausgeber des Blogs www.konversionsKRAFT.de und Gründer der Global Conversion Alliance.

Mit maßgeschneiderten Konzepten optimiert konversionsKRAFT Webseiten und Portale, um **mehr** Besucher zum Kauf oder zur Anfrage zu führen. Mithilfe von Methoden aus der Konsumpsychologie und dem Neuromarketing garantiert konversionsKRAFT den Erfolg der durchgeführten Optimierungsmaßnahmen.

Das Ziel der Arbeit ist es, die Effizienz des eingesetzten Online Marketing Budgets von Unternehmen zu erhöhen, positive Markenerlebnisse zu schaffen und damit **mehr Erfolg** im digitalen Wettbewerb zu erzielen.