

appinio

Interactive webinars

The future of research

Research Academy

Free certification

01 Intro

02 Recap

03 Modul 2

04 Key Takeaways

05 Q&A



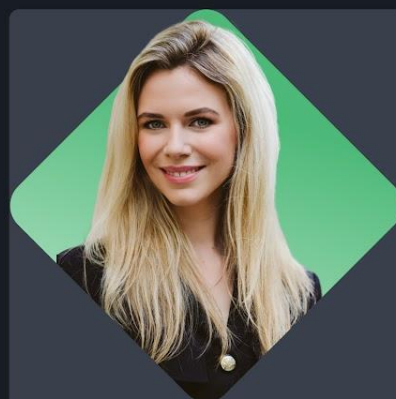
Dein Expert:innen Team



Dr. Jan Walsken
Research Consultant



Dr. Julia Vollstedt
Senior Research
Consultant



Anna Breiting
Customer Success
Manager

Dein Weg zum MaFo Zertifikat

Wir bescheinigen dir grundlegende MaFo Kenntnisse, die du mit der Teilnahme an allen 4 Modulen erwerben kannst.

- Nimm an allen Terminen live teil ODER schau dir die Recordings an
- Dafür musst du dich einmal mit deiner E-Mail Adresse registrieren und diese beibehalten
- Am Ende gibt es einen kurzen Wissenstest, den du bestehen musst.

Nutze das Zertifikat

- Bei deinem Arbeitgeber als Weiterbildung
- Bei Bewerbungen
- Auf LinkedIn

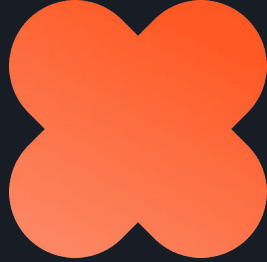


Übersicht: 4 Module



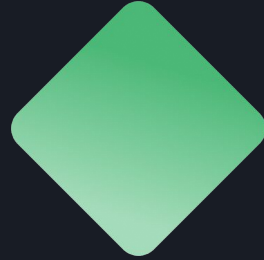
Forschungsfrage & Zielgruppe

Die richtigen Fragen stellen
Zielgruppe definieren
Quotas festlegen



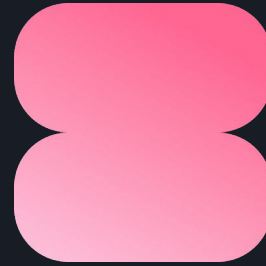
Fragebogen- Design

Logischer Aufbau
Bias Vermeidung
Fragetypen



Myth Busting

Demografie-Dilemma
Brand Tracking
Signifikanz



Vom Datenpunkt zur Story

Zahlen richtig lesen
Storytelling
Stakeholder Management

Modul 1 - RECAP

Forschungsfrage & Zielgruppe

Studien von Anfang an richtig aufsetzen



Repräsentativität

Take-Aways

Repräsentativität an sich ist nicht unbedingt ein Gütesiegel für die Qualität einer Umfrage.



Umfragen können nicht per se „repräsentativ“ sein

Wir müssen uns immer fragen, WEN wir mit unserer Studie repräsentieren wollen: unsere Kund:innen? Ganz Deutschland? Alle Hundebesitzer:innen?



Perfekte Repräsentativität ist ein unerreichbares Ideal

Quota Sampling kann sie nur annähern: Wir kennen nie alle Merkmale und deren Verteilung in der Grundgesamtheit – und könnten nie alle Menschen erreichen.



Eine große Stichprobe allein macht eine Studie nicht repräsentativ

Spanische Männer unter 25 können deutsche Frauen über 65 nicht repräsentieren – egal ob wir 100, 1.000 oder 10.000 von ihnen befragen.



Dennoch: Je größer die (repräsentative) Stichprobe, desto genauer die Messung

Mehr Teilnehmende senken die Fehlermarge und erlauben genauere, zuverlässige Aussagen über die Grundgesamtheit.

Drei Fragen – drei Antworten:



Konkret und messbar formulieren – sonst misst du nichts.

Erst die Zielgruppe definieren, dann die Stichprobe ziehen.

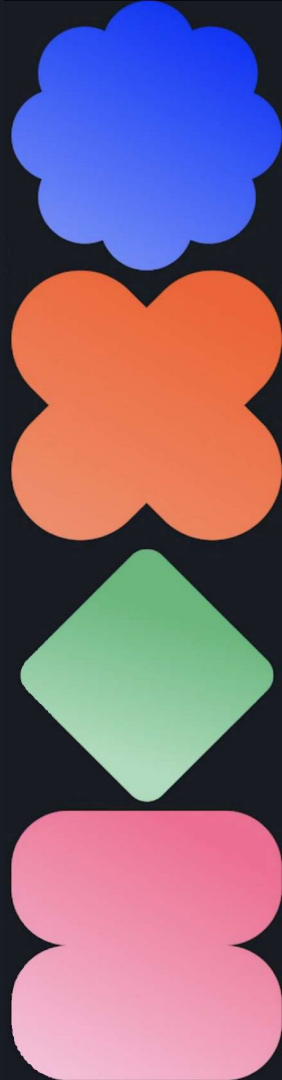


Repräsentativ heißt: repräsentativ für wen?

Modul 2

Fragebogen-Design

Wie aus einer Forschungsfrage eine gute Frage wird



Product & Service Development

Range Optimization

User Experience (UX)

Product Testing

Concept / Idea Screening

Packaging Testing

Claim Testing

Naming Testing

Pricing

Market Understanding

Shopper Understanding

Usage & Attitude

Customer Experience (CX)

Consumer Understanding

Brand & Communication

Ad Test (Pre-Post)

Brand / Health Performance

Logo / Slogan Test

Sinngleiche Frage, ganz andere Ergebnisse

Rugg, D. (1941). Experiments in wording questions: II. *Public Opinion Quarterly* , 5, 91–92

"Should we **allow** public speeches
against democracy?"

75 %

„Nein“ – nicht erlauben

"Should we **forbid** public speeches
against democracy?"

54 %

„Ja“ – verbieten

💡 Beide Antworten gehen in dieselbe Richtung, trotzdem 21 Prozentpunkte Unterschied.

Welche dieser drei Fragen geht ins Feld?

Stimme jetzt im Umfrage-Tool ab!

A Anmelden & empfehlen

„Wie wahrscheinlich würdest du dich für unsere hervorragende neue ‚Appinio Research Academy‘ anmelden und sie deinen Kolleg:innen empfehlen?“

- a) Sehr wahrscheinlich
- b) Wahrscheinlich
- c) Eher nicht

B Direkter Vergleich

„Welcher der beiden Namen spricht dich mehr an – ‚Appinio Research Academy‘ oder ‚Next Gen Insights Program‘?“

- a) Appinio Research Academy
- b) Next Gen Insights Program

C Split, nur ein Name

Jede Person sieht nur einen Namen. Kurzer Infotext, dann: „Angenommen, das Angebot startet in den nächsten Monaten – wie wahrscheinlich meldest du dich an?“

- a) Sehr wahrscheinlich
- b) Wahrscheinlich
- c) Weder noch
- d) Unwahrscheinlich
- e) Sehr unwahrscheinlich

 Bitte wechsele jetzt in das Umfrage-Tool und gib deine Stimme ab!

Auflösung

Zwei Fragen fallen durch – aber aus sehr unterschiedlichen Gründen

A



Suggestivfrage

- „hervorragende“, „unsere“ → suggestiv
- anmelden und empfehlen → zwei Fragen
- unbalancierte Skala, kein Zeitbezug

B



Sauber formuliert, aber falsche Messung

- Präferenz ist nicht Anmeldeabsicht
- keine Skala, also keine Stärke
- erzwingt einen Vergleich, den niemand hat

C



Monadisch – KPI aus Modul 1

- jede Zelle sieht nur einen Namen
- Skala 1–5, voll verbalisiert
- direkt vergleichbar zwischen A und B

 Die häufigste kaputte Frage ist nicht die peinliche – sondern die, die gut aussieht und am Messziel vorbeischießt.

01 Reihenfolge & Aufbau

Wo eine Frage steht, verändert ihre und folgende Antworten

Erst ungestützt, dann gestützt fragen

Erst „Welche Marken fallen dir ein?“, dann die Markenliste zeigen. Andersherum hast du die Antwort vorgesagt.

Filter setzen statt Pseudo-Meinungen

Ohne Filterfrage geben 20–40 % eine Meinung zu einem frei erfundenen Gesetz ab. Mit Filter rund 10 %.

Vorherige Fragen färben die nächste

Lebenszufriedenheit vor Dating-Häufigkeit gefragt: $r = .16$. Dieselben zwei Fragen, Reihenfolge getauscht: $r = .55$.

Generelle vor konkreter Preisfrage

Erst „Würdest du zu einem fairen Preis kaufen?“, dann Preispunkte. Ein früher Preis ankert alles danach.

Kurz halten – Länge kostet Qualität

Über 25 Minuten brechen dreimal so viele ab wie unter 5. Und die letzten Fragen werden nur noch geklickt.

 Der Fragebogen ist eine Dramaturgie: erst offen, dann gestützt; bzw. erst allgemein, dann konkret.

02 Die Frage selbst

Jedes Wort ist eine Messentscheidung

Pro Konstrukt nur eine Frage

„... anmelden und weiterempfehlen?“ – bei „Nein“ weißt du nicht, welche der beiden Hälften gemeint war.

Wertungen und Färbungen raus

„unsere hervorragende neue Academy“ → „die Academy“. Niemand kritisiert gern seinen Gastgeber.

Spezifisch statt „oft“ und „regelmäßig“

„Oft“ heißt für die einen täglich, für die anderen monatlich. Besser: „Wie oft in den letzten 7 Tagen?“

Kurze, klar begrenzte Zeiträume

„In den letzten 4 Wochen“ statt „üblicherweise“. Je länger der Zeitraum, desto mehr wird hineingerechnet.

Keine Agree/Disagree-Batterien

14 % mehr Zustimmung als bei Forced Choice – und 22 % stimmen einer Aussage und ihrer Umkehrung zu.



Wenn zwei Menschen eine Frage unterschiedlich verstehen, messen wir zwei verschiedene Dinge.

03 Die Antwortoptionen

Die Skala ist die halbe Frage

Symmetrisch und voll verbalisiert

„sehr wahrscheinlich / wahrscheinlich / eher nicht“ → mindestens 5 Stufen

Skalenzahlen färben die Antwort

Gleiche Worte, andere Zahlen: bei 0 bis 10 wählten 34 % die untere Hälfte, bei -5 bis +5 nur 13 %.

Wettbewerber und „keine davon“ mitgeben

Ohne Distraktoren misst du Zustimmung, nicht Bekanntheit. Erfundene Marken entlarven Straightliner.

Antwortoptionen randomisieren

Am Bildschirm gewinnt, was oben steht. Ausnahme: „keine davon“ bleibt unten und exklusiv.

Matrixfragen sparsam einsetzen

Grids erzeugen mehr Abbrüche, mehr fehlende Werte und mehr Straightlining als Einzelfragen.

 Antwortoptionen sind kein Formular-Detail – sie sind Teil des Messinstruments.

Manche Fragen kann man nicht direkt stellen

Direkt gefragt

„Wie wichtig ist Ihnen ...?“

Alles ist wichtig. Alles über 4,2.

„Was würden Sie dafür zahlen?“

Hypothetisch genannt: +35 % bis +200 %.

Warum

**Menschen sind schlechte
Introspektionisten –
aber gute Entscheider.**



Also nicht nach Wichtigkeit fragen – Entscheidungen beobachten und die Wichtigkeit ausrechnen.

→ Genau dafür gibt es Conjoint, MaxDiff & Co.



Methoden Tool Box

Fortgeschrittene Research Methoden bei Appinio



Das Introspektions-Paradox

Die Fehlannahme

„Menschen wissen genau, was sie wollen, wie viel sie zahlen würden und warum sie kaufen.“

- **Alles ist wichtig:** Auf 5er-Skalen bewerten Befragte im Schnitt 85 % aller Features mit 4 oder 5.
-
- **Hypothetischer Bias:** Unverbindlich genannte Zahlungsbereitschaften liegen 35 % bis 200 % über dem Kassentyp.
-
- **Soziale Erwünschtheit:** Befragte rationalisieren Motive post-hoc (z. B. Nachhaltigkeit statt Preis).

Der Ausweg

„Menschen sind schlecht in Introspektion – aber exzellente Entscheider.“

- **Trade-offs erzwingen:** Reale Entscheidungen erfordern immer Kompromisse zwischen Nutzen und Kosten.
- **Implizite Messung:** Wichtigkeit nicht erfragen, sondern aus Auswahlentscheidungen statistisch ableiten.
- **Monadische Isolation:** Varianten getrennt testen, um künstliche Vergleiche zu verhindern.

Daumenregel: Wenn eine Forschungsfrage unbewusste Motive oder Zielkonflikte berührt, gehört sie in der Regel in eine Methode oder ein Research Design - nicht ein Textfeld.

Ein Blick in unsere Method Tool Box

Fortgeschrittene Research Methoden bei Appinio

Nicht jede Fragestellung lässt sich direkt durch eine Frage beantworten



Priorisierungen

Maximum Difference
Scaling
KANO



Pricing

Van Westendorp
Gabor-Granger
Feature-Value-Matrix
Conjoint



Design- & Konzepttests

A/B Testing



Marktverständnis

Treiberanalyse

Rankings: trügerisch simpel, aber kognitiv komplex

Use Case: Claim Testing
& Priorisierungen

— Wie wir es NICHT machen sollten

„Bringen Sie die folgenden 12 Produktmerkmale in eine Rangfolge von 1 (am wichtigsten) bis 12 (am unwichtigsten).“

- Kognitiver Overload
- Keine Distanzmessung
- Likert-Skalen sind keine gute Rating Alternative

Ergebnis:

Pseudo-Differenzierung ohne handlungsleitende Aussagekraft

Rankings: trügerisch simpel, aber kognitiv komplex

Use Case: Claim Testing
& Priorisierungen

❌ Wie wir es NICHT machen sollten

„Bringen Sie die folgenden 12 Produktmerkmale in eine Rangfolge von 1 (am wichtigsten) bis 12 (am unwichtigsten).“

- Kognitiver Overload
- Keine Distanzmessung
- Likert-Skalen sind keine gute Rating Alternative

Ergebnis:

Pseudo-Differenzierung ohne handlungsleitende Aussagekraft

✅ So funktioniert Maximum-Difference Scaling

Teilnehmende durchlaufen mehrere rotierende Sets mit je 4 Merkmalen. Das mathematische Modell berechnet die relative Bedeutsamkeit über alle Kombinationen hinweg:

- Einfache Aufgabe
- Metrische Skala
- Kein Skalen-Bias

Ergebnis:

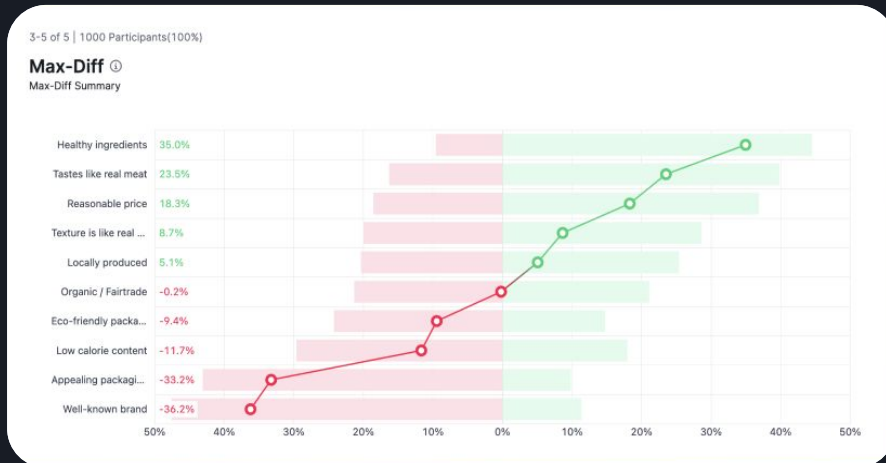
Granulare, normierte & segmentierungsfähige Ergebnisse

💡 **Daumenregel: Sobald mehr als 5-6 Items priorisiert werden müssen, sollten wir von einem Ranking Abstand nehmen!**

MaxDiff: Eine Alternative zu Rankings

Use Case: Claim Testing
& Priorisierungen

Typischer Output / Beispiel



Einsatzgebiet:

Ranking von USPs, Claims, Value Propositions und sogar Designs (Picture MaxDiff)



Praxis-Tipp:

Halte die Items in einer MaxDiff auf derselben Abstraktionsebene und die Formulierungen ähnlich lang.

Limitation:

Eindimensionalität. Viele Entscheidungen erfordern aber Kompromisse (Trade-Offs).



Daumenregel:

Bei mehreren Attributen, Features, die miteinander kombiniert werden sollen, oder Produktpaketen → Conjoint

„Der Preis ist heiß“ funktioniert leider nur im Fernsehen

Use Case: Pricing

— Wie wir es NICHT machen sollten

„Was wärest du maximal bereit, für diesen neuen Snack im Supermarkt zu bezahlen? [Offenes Zahlenfeld]“

- Kein Bezugsrahmen
- Strategisches Antworten
- Anker-Verzerrung

Ergebnis:

Unbrauchbare Fantasiaezahlen, die katastrophal für Business Cases sind

„Der Preis ist heiß“ funktioniert leider nur im Fernsehen

Use Case: Pricing

❌ Wie wir es NICHT machen sollten

„Was wärest du maximal bereit, für diesen neuen Snack im Supermarkt zu bezahlen? [Offenes Zahlenfeld]“

- Kein Bezugsrahmen
- Strategisches Antworten
- Anker-Verzerrung

Ergebnis:

Unbrauchbare Fantasiazahlen, die katastrophal für Business Cases sind

✅ Unser Verständnis von Preisbereitschaft

Preisbereitschaft ist keine einzelne Zahl, sondern ein psychologischer Korridor.

- Qualitätsgrenze
- Schmerzgrenze (Pain-of-Paying)
- Zwei erprobte Methoden: Van Westendorp (Korridor) & Gabor-Granger (Elastizität)

Empfehlung:

Ersetze ein Freitextfeld durch strukturierte Pricing-Methoden



Merke: Frage niemals ungestützt nach „Was würdest du zahlen?“
Frage stattdessen nach der Wahrnehmung konkreter Preisschwellen für ein konkretes Produkt.

VWD & GG: Zwei geläufige Pricing Methoden

Use Case: Pricing

Van Westendorp (VWD)

Ziel: Akzeptablen Preiskorridor für Innovationen ermitteln.



Ergebnis:

Optimaler Preispunkt (OPP) & Indifferenzpreispunkt (IDP).

VWD & GG: Zwei geläufige Pricing Methoden

Use Case: Pricing

Van Westendorp (VWD)

Ziel: Akzeptablen Preiskorridor für Innovationen ermitteln.



Ergebnis:

Optimaler Preispunkt (OPP) & Indifferenzpreispunkt (IDP).

Gabor-Granger (GG)

Ziel: Konkrete Preiselastizität & Umsatzoptimum testen.



Ergebnis:

Ermittelt den Preispunkt mit dem maximalen Gesamtumsatz.

 **Merke:** VWD für Markterkundung & Preiskorridor (Range) • Gabor-Granger für konkrete Preiselastizitätskurven

— Wie wir es NICHT machen sollten

„Hier sind Design A und Design B. Welches gefällt dir besser und warum?“ [Beide nebeneinander]

- Macht Laien zu Art Direction
- Aktiviert System 2
- Keine Kaufabsicht

Artefakt:

Gegenseitige Beeinflussung. Subtilere oder neuere Designs verlieren häufig gegen bestehende Designs.

❌ Wie wir es NICHT machen sollten

„Hier sind Design A und Design B. Welches gefällt dir besser und warum?“ [Beide nebeneinander]

- Macht Laien zu Art Direction
- Aktiviert System 2
- Keine Kaufabsicht

Artefakt:

Gegenseitige Beeinflussung. Subtilere oder neuere Designs verlieren häufig gegen bestehende Designs.

✅ Das Monadische Prinzip

Reality Check: Niemand sieht im Supermarkt zwei Designs nebeneinander.

- Monadischer Split
- Gleiche Kernfragen
- Echte Kausalität

Ergebnis:

Wir messen die tatsächliche Wirkung eines Designs oder Konzepts auf eine Ergebnisvariable.

💡 Praxis Tipp: Immer monadisch testen, wenn das Ziel eine Vorhersage des realen Markterfolgs ist.

💎 Merke: Wir erzeugen einen künstlichen Vergleich zwischen Alternativen, den Konsument:innen in der Realität niemals vorfinden würden.

Was sie sagen, ist nicht immer, was sie antreibt

Use Case: Marktverständnis

— Wie wir es NICHT machen sollten

„Wie wichtig ist Ihnen bei Ihrer Bank

- A) ... die Datensicherheit,
- B) ... der Kundenservice
- C) ... ein modernes App-Design?“ [1–5 Skala]

- Angegebene Wichtigkeit differenziert nicht
- Unbewusste Hebel werden ignoriert
- Risiko von Fehlinvestitionen

Problem:

Angegebene Wichtigkeit misst Pflichtanforderungen,
nicht Differenzierungstreiber

Was sie sagen, ist nicht immer, was sie antreibt

Use Case: Marktverständnis

❌ Wie wir es NICHT machen sollten

„Wie wichtig ist Ihnen bei Ihrer Bank

- A) ... die Datensicherheit,
- B) ... der Kundenservice
- C) ... ein modernes App-Design?“ [1–5 Skala]

- Angegebene Wichtigkeit differenziert nicht
- Unbewusste Hebel werden ignoriert
- Risiko von Fehlinvestitionen

Problem:

Angegebene Wichtigkeit misst Pflichtanforderungen, nicht Differenzierungstreiber

✅ Die Lösung: Treiberanalyse (TA)

„Welchen **Einfluss** haben

- A) ... die Datensicherheit,
- B) ... der Kundenservice
- C) ... ein modernes App-Design

... auf die erklärbare Kundenzufriedenheit bei der Bank x?“

- Regressionsanalyse
- Statistisch abgeleitete Wichtigkeit
- Aufdeckung impliziter Hebel

Ergebnis:

TA trennt echte Differenzierungstreiber von reinen Hygiene-Faktoren



Merke: Angegebene Wichtigkeit = Was Befragte behaupten, was wichtig ist
Abgeleitete Wichtigkeit = Was die Befragten wirklich bewegt

Treiberanalyse: Angegebene vs. abgeleitete Wichtigkeit

Use Case: Marktverständnis

Die 4-Felder Matrix

Indem Performance (Zufriedenheit) gegen den Treiber-Einfluss (Beta-Koeffizient) geplottet wird, entsteht eine glasklare Roadmap:

1. Hebel (Hohe Relevanz, schlechte Note)

Hier muss sofort investiert werden!

2. Stärken (Hohe Relevanz gute Note)

USP & Basis für Marketing-Story.

3. Low Priority (Geringe Relevanz, schlechte Note)

Keine Panik - bewegt die Kunden kaum.

4. Basics (Geringe Relevanz, gute Note)

Kosten reduzieren, Effizienz sichern.

Typische Fragestellungen

- Welche Markenattribute (z. B. Innovation, Nachhaltigkeit, Verlässlichkeit) treiben die Kaufabsicht?
- Welche Service-Touchpoints haben den größten Einfluss auf die Gesamtzufriedenheit der Kunden?
- Welche Features einer App/Software tragen am stärksten zur dauerhaften Nutzerzufriedenheit bei?
- Welche Aspekte (z. B. Gehalt, Führungskultur, Work-Life-Balance, Entwicklungschancen) haben den stärksten Einfluss auf das Mitarbeiter-Engagement?

 **Merke:** Eine TA ist immer dann geeignet, wenn wir wissen wollen, welche Faktoren den stärksten Einfluss auf eine Zielgröße haben, um zu erfahren, wo sich Optimierungen am meisten lohnen.

Überblick Method Tool Box

Fortgeschrittene Research Methoden bei Appinio

Nicht jede Fragestellung lässt sich direkt durch eine Frage beantworten



Maximum Difference Scaling

Anstelle von Rankings für echte Differenzierung und handlungsleitende Aussagekraft.



Van Westendorp Gabor-Granger

Anstelle von offenen Preisanfragen, um Preisspannen (VWD) und Preiselastizität Kurven zu ermitteln. Preisbereitschaft ist keine Zahl, sondern ein psychologischer Korridor.



A/B Testing

Anstelle von direkten Vergleichen, um den realen Markterfolg eines Designs-/Konzeptes unter Bedingungen zu testen.

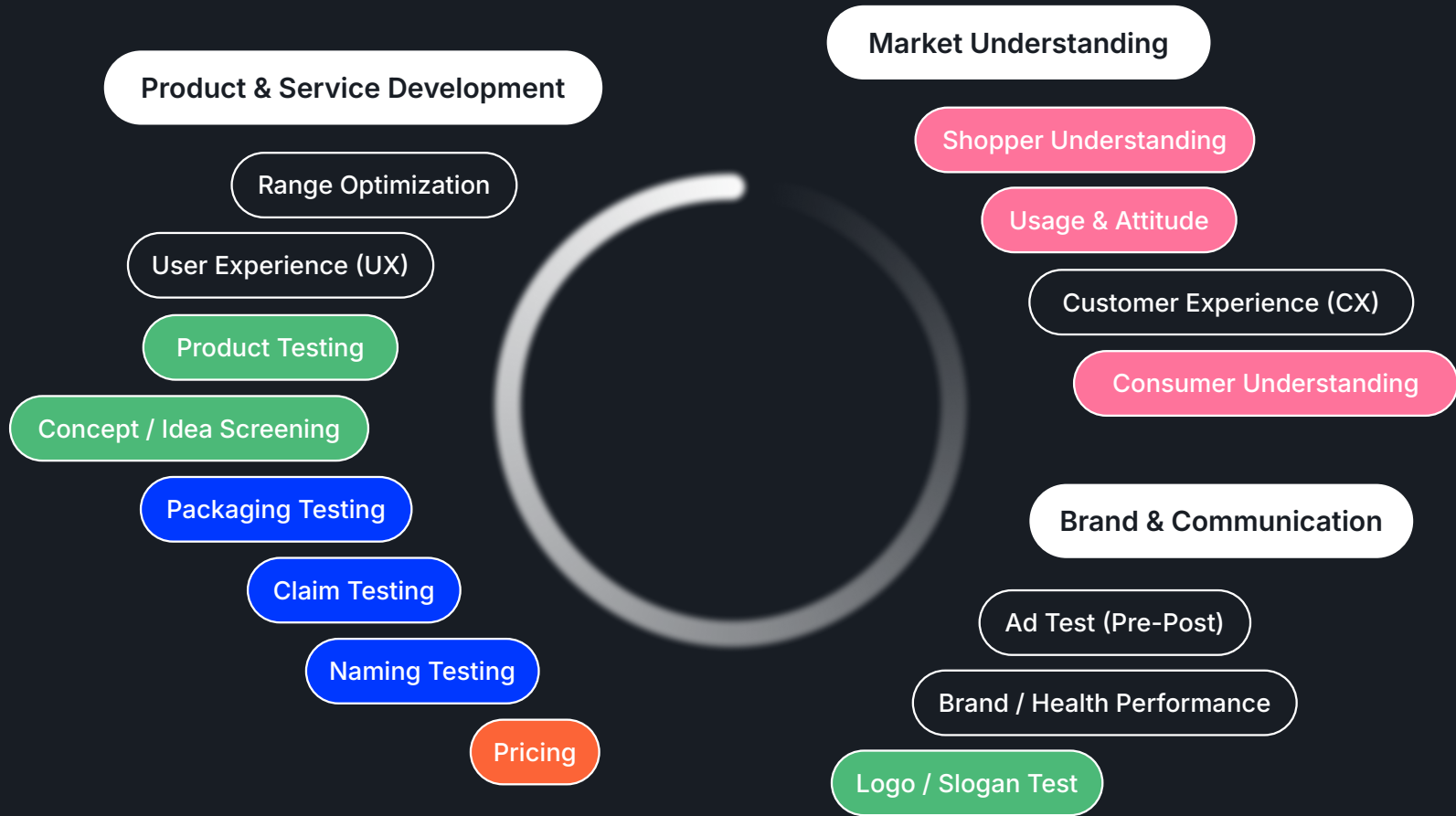


Treiberanalyse

Anstelle von deskriptiven Wichtigkeits-Statements, um Hygienefaktoren von wahren Erfolgstreibern zu unterscheiden.

Key Takeaways ganze Session





Nächstes Modul:

29. Oktober



12:00



Myth Busting

Demografie Dilemma
Brand Tracking
Signifikanz

Zum Schluss

Q&A

Stellt eure Fragen im Chat –
wir gehen sie gemeinsam durch.



Von der Idee zu den Ergebnissen in drei Fragen:



Was wollen wir eigentlich herausfinden?

"Wen befragen wir?"



"Wie befragen wir die richtigen Leute?"

Schritt 1:

Die Forschungsfrage

Was wollen wir eigentlich herausfinden?



Welche dieser drei Forschungsfragen ist gut?

Stimme jetzt in deinem Umfrage-Tool für den besten Ansatz ab!



Option A

**„Ist unsere neue
Verpackung gut?“**



Option B

**„Wie zufrieden
(Skala 1–5) sind unsere
Kund:innen mit unserem
Kundenservice?“**



Option C

**„Wie zeigen wir, dass
unsere Marke die
beliebteste ist?“**



Bitte wechsle jetzt in das Umfrage-Tool und gib deine Stimme ab!

Formulierung von konkreten, quantifizierbaren Forschungsfragen herausfinden?



Gute Forschungsfragen

„Welches von drei Verpackungsdesigns erzielt die höchste Kaufwahrscheinlichkeit (Skala 1–5) bei Biertrinker:innen?“

konkret, messbar, klare Zielgruppe

„Wie zufrieden (Skala 1–5) sind unsere Kund:innen mit unserem Kundenservice?“

eine Frage, klare Skala

„Wie hoch ist die gestützte Bekanntheit unserer Marke in Deutschland (18–65)?“

klar definierter KPI



Schlechte Forschungsfragen

„Ist unsere neue Verpackung gut?“

nicht quantifizierbar

„Wie zufrieden sind unsere Kund:innen mit Preis und Kundenservice?“

double-barrelled: zwei Fragen in einer

„Wie zeigen wir, dass unsere Marke die beliebteste ist?“

post hoc reasoning: „Es muss herauskommen, dass ...“

Teste dich selbst: Welche der drei Fragen ist eine gute Forschungsfrage?

“Welcher Name – 'Appinio Research Academy' (Option A) vs. 'Next Gen Insights Program' (Option B) – erzielt eine höhere Anmeldeabsicht (Skala 1–5)?”

→ Konkret, messbar, vergleicht zwei reale Optionen



“Welcher Name – 'Appinio Research Academy' (Option A) vs. 'Next Gen Insights Program' (Option B) – erzielt eine höhere Anmeldeabsicht (Skala 1–5)?”

→ Konkret, messbar, vergleicht zwei reale Optionen

Die Forschungsfrage steht – aber da fehlt doch noch etwas...



Schritt 2:

Die Zielgruppe

Wen befragen wir?



Zielgruppendefinition

Wen befragen wir?

Unsere Forschungsfrage:

Welcher Name – 'Appinio Research Academy' (Option A) vs. 'Next Gen Insights Program' (Option B) – erzielt eine höhere Anmeldeabsicht (Skala 1–5) bei ...?

Mögliche Zielgruppen:

- Mitarbeitende bei Appinio?
- Studierende aus den Bereichen Psychologie und BWL?
- Fachkräfte aus den Bereichen Insights, Brand Management und Marketing?



“Welcher Name – 'Appinio Research Academy' (Option A) vs. 'Next Gen Insights Program' (Option B) – erzielt eine höhere Anmeldeabsicht (Skala 1–5) bei Fachkräften aus den Bereichen Insights, Brand Management und Marketing?”

→ Konkret, messbar, vergleicht zwei reale Optionen, klare Zielgruppe

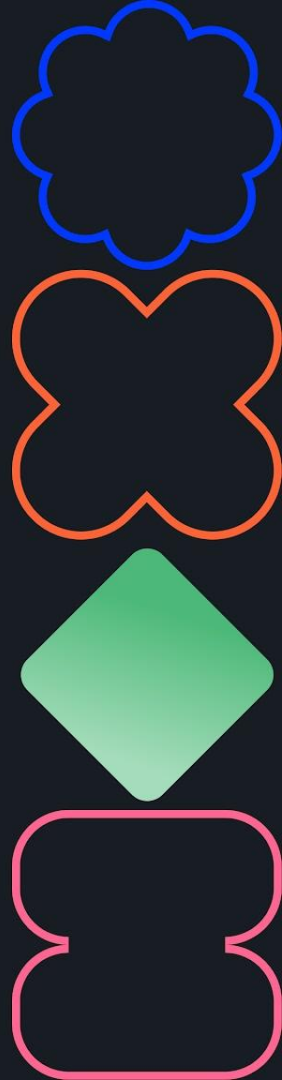
Die Forschungsfrage steht – aber wie kommen wir an genau diese Leute?



Schritt 3:

Die Stichprobe

Wie befragen wir die richtigen Leute?



Repräsentativität

Der Goldstandard



Definition

Eine Stichprobe ist repräsentativ, wenn sie die Grundgesamtheit (z. B. alle Menschen in Deutschland) in allen relevanten Merkmalen – etwa Alter, Geschlecht oder Region – im Kleinen widerspiegelt. Nur dann lassen sich die Ergebnisse der Stichprobe auf die Grundgesamtheit übertragen.

Wie stellen wir also sicher, dass die Ergebnisse der Umfrage wirklich die Meinung der Grundgesamtheit (= Fachkräfte aus den Bereichen Insights, Brand Management und Marketing) widerspiegeln?



Wie befragen wir die richtigen Leute? – Option A

Grundgesamtheit



Fachkräfte aus den Bereichen Insights, Brand Management und Marketing in Deutschland

Stichprobenziehung (das unerreichbare Ideal)



Erstelle eine Liste mit allen Fachkräften aus den Bereichen Insights, Brand Management und Marketing in Deutschland und wir befragen die Leute komplett zufällig



Zufall wird beispielsweise durch Würfeln oder statistische Verfahren hergestellt



Befrage **nur** die zufällig Ausgewählten — aber **alle** von ihnen

= Perfekt repräsentativ – aber in der Praxis unmöglich umzusetzen



Wie befragen wir die richtigen Leute? – Option B

Grundgesamtheit



Fachkräfte aus den Bereichen Insights, Brand Management und Marketing in Deutschland

Willkürliche Auswahl / Convenience Sample



Wir befragen irgendwelche Leute auf der Straße, im Supermarkt, aus unserem Bekanntenkreis, beim Sport...



Und/oder: wir teilen die Befragung in sozialen Medien (z.B. LinkedIn)



Wir befragen nur die Leute, die sich bereiterklären, befragt zu werden

= Schnell und günstig – aber systematisch verzerrt



Wie befragen wir die richtigen Leute? – Option C

Grundgesamtheit



Fachkräfte aus den Bereichen Insights, Brand Management und Marketing in Deutschland

Quotenstichprobe



Stichprobe vermessen, z. B. per Zensus: Nach welchen Kriterien unterscheiden sich die Personen (Alter, Geschlecht, Einkommen)?



„Quoten- bzw. Stichprobenplan“ erstellen, z. B. gen. pop / nat. rep. Hier: 50 % je Geschlecht; 45 % blau; 25 % orange; 17 % grün; 12 % pink...



Fachkräfte aus den Bereichen Insights, Brand Management und Marketing in Deutschland auf Basis des Quotenplans befragen

= Der pragmatische Kompromiss: bildet die Zielgruppe in den Merkmalen ab, die bekannt sind



Repräsentativität Quiz



Der Fall

Ein regionales Nachrichtenportal startet auf seiner Startseite eine offene Online-Umfrage zur neuen Fußgängerzone der Stadt.



Wer wurde befragt?

- **Stichprobe:** N = 3.500 Website-Besucher:innen, die spontan auf das Umfrage-Widget geklickt haben.
- **Methode:** Freiwillige Online-Teilnahme ohne Quoten oder Screening (Self-Selection)



Welche Schlagzeile ist methodisch korrekt?

- **Option A:** „82 % der Bürger:innen lehnen die neue Fußgängerzone ab.“
- **Option B:** „Die Mehrheit unserer Leser:innen lehnt sie ab.“
- **Option C:** „82 % der Teilnehmenden unserer Umfrage lehnen sie ab.“

Repräsentativität Quiz



Welche Schlagzeile ist methodisch korrekt?

- **Option A** „82 % der Bürger:innen lehnen die neue Fußgängerzone ab.“
✗ Zu weit gefasst – befragt wurden nur Website-Besucher:innen, die freiwillig geklickt haben.
- **Option B** „Die Mehrheit unserer Leser:innen lehnt sie ab.“
✗ Immer noch zu weit – die 3.500 Klickenden sind ein selbst selektierter Bruchteil der Leserschaft.
- **Option C** „82 % der Teilnehmenden unserer Umfrage lehnen sie ab.“
✓ Richtig – beschreibt genau die Gruppe, die befragt wurde.



Auflösung & Learning

Achtung vor Self-Selection Bias! Bei offenen Online-Links nehmen vor allem Menschen teil, die wütend oder stark emotionalisiert sind — Zufriedene und Neutrale klicken solche Widgets nicht an. Ein offener Link misst Lautstärke, nicht Mehrheit.

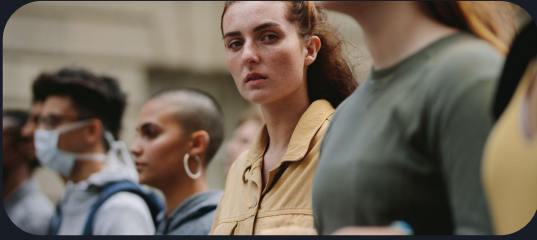
Und: Ein „nicht repräsentativ“-Hinweis unter der Schlagzeile repariert eine zu weit gefasste Aussage nicht. Die Reichweite der Aussage muss in der Schlagzeile selbst stimmen.

Repräsentativität - Die Realität

Muss eine Umfrage immer repräsentativ sein, damit sie uns Insights bringen kann?

Nein – aber wir müssen bei unseren Ableitungen aufpassen, für wen sie gelten.

Für welche Fragestellungen muss eine Umfrage repräsentativ sein?



Repräsentatives Sample unverzichtbar

- Aussagen über die Gesamtbevölkerung („die Deutschen ...“)
- Marktgrößen- und Potenzialschätzungen
- Politische Meinungs- und Wahlforschung
- Ergebnisse für Presse und Öffentlichkeit



Nicht unbedingt notwendig

- Explorative Studien und Ideen-Screening
- A/B-Vergleiche, z. B. Konzept- oder Verpackungstests
- Befragungen spezifischer Zielgruppen (z. B. eigene Kund:innen)
- UX- und Usability-Tests

**„Appinio Research Academy“ gewinnt: Ø
4,69 vs. 4,11 Punkte Teilnahmeabsicht auf
einer Skala von 1–5.**

→ Deshalb heißt dieses Webinar so, wie es heißt.

n = 500 Professionals · Insights, Brand Management & Marketing



Nächstes Modul:

24. September



12:00



Fragebogen-Design

Logischer Aufbau
Bias Vermeidung
Fragetypen

Zum Schluss

Q&A

Stellt eure Fragen im Chat –
wir gehen sie gemeinsam durch.

