



Planungen eines Experiments

2

Vor einem Experiment müssen mehrere Überlegungen getätigt und Entscheidungen getroffen werden. Welchen Effekt möchte ich untersuchen? Welche Einflussvariablen muss ich dafür manipulieren? Wie genau soll mein Experiment aussehen? Wie viele Messwiederholungen benötige ich?

2.1 Anatomie eines Experiments

Bevor wir über die konkrete Planung eines Experiments sprechen, sollten wir uns einen Überblick über die Grundbegriffe und die Bestandteile eines herkömmlichen computergestützten Experiments der Experimentalpsychologie verschaffen. Natürlich können sich diese Bestandteile je nach ihrer inhaltlichen Orientierung auch ändern oder weitere Phasen hinzukommen. Hier wird der Fokus jedoch primär auf Experimente gelegt, wie sie in der visuellen Aufmerksamkeitsforschung verwendet werden. Nichtsdestotrotz sind die Begrifflichkeiten und Konzepte auch über diesen spezifischen Forschungsbereich hinaus gebräuchlich. Es ist daher sehr wichtig, sich über die Bedeutung dieser Begriffe im Klaren zu sein, da dies nicht nur das Verständnis von Experimenten, welche vorgestellt werden, erhöht, sondern auch den Austausch mit Kollegen und Kolleginnen bei der Planung von Experimenten wesentlich erleichtert. Also, was sind Bildschirme, Durchgänge, gemischte Blöcke und reine Blöcke?

Bildschirme

Wenn Sie einen Blick auf Abb. 2.1 werfen, sehen Sie, dass ein Bildschirm die kleinste „Einheit“ eines Experiments darstellt. Auch wenn schnell klar ist, was

„Bildschirm“ im gegenwärtigen Kontext bedeutet, mag der Begriff zunächst etwas verwirrend sein. Logischerweise hieven wir während eines Experiments nicht eine Reihe von Computermonitoren vor die Versuchspersonen. Bildschirm im Kontext eines Experiments bezieht sich auf jenen Reiz bzw. jene Konfiguration von Reizen, die einer Versuchsperson zu einem gegebenen Zeitpunkt präsentiert werden.

Durchgänge

Ein oder (meist) mehrere Bildschirme konstituieren einen Durchgang. Genereller gesagt, sind Durchgänge jene Aufgaben, die Versuchspersonen im Verlauf eines Experiments erledigen sollen. Ein Durchgang ist beendet, wenn Versuchspersonen die von ihnen geforderte Aufgabe erledigt haben, woraufhin ein neuer Durchgang folgt. Welche Aufgabe die Versuchspersonen erledigen sollen, hängt natürlich vom jeweiligen Experiment ab. Oft sollen Versuchspersonen eine manuelle Antwort geben, eine schnelle Augenbewegung auf einen Reiz hin oder von einem Reiz weg ausführen. Das bedeutet, dass es in einem Verhaltensexperiment pro Durchgang meist einen einzelnen Datenpunkt gibt. Verwendet man andere Methoden, wie etwa Elektroenzephalografie (EEG), werden Daten über die Hirnaktivität über einen gesamten Durchgang hinweg gesammelt. Das Prinzip bleibt aber auch bei der Verwendung eines EEGs gleich: Man vergleicht Datenpunkte zu einem bestimmten Zeitpunkt innerhalb eines Durchgangs unter verschiedenen Experimentalbedingungen.

Blöcke

Gleich wie Durchgänge eine Sammlung an Bildschirmen darstellen, sind Blöcke eine Sammlung von sukzessiven aufeinanderfolgenden Durchgängen. Werfen Sie wieder einen Blick in Abb. 2.1: In dieser Abbildung sind zwei Blöcke mit den Bezeichnungen „Block 1“ und „Block 2“ dargestellt. Natürlich wäre es auch möglich, die Übungsphase als „Übungsblock“ zu bezeichnen, da Versuchspersonen auch in dieser Phase meist mehrere aufeinanderfolgende Durchgänge absolvieren und nicht lediglich einen. Es gibt jedoch verschiedene Möglichkeiten, experimentelle Bedingungen innerhalb eines Blockes zu präsentieren: gemischt oder rein (oft auch „geblockt“ genannt) (siehe Abb. 2.2). Kurz gesagt: in *gemischten Blöcken* kommen alle Bedingungen, die Sie in Ihrem Experiment testen, randomisiert vor. Das bedeutet, innerhalb eines Blocks kann ein Durchgang aus der Bedingung

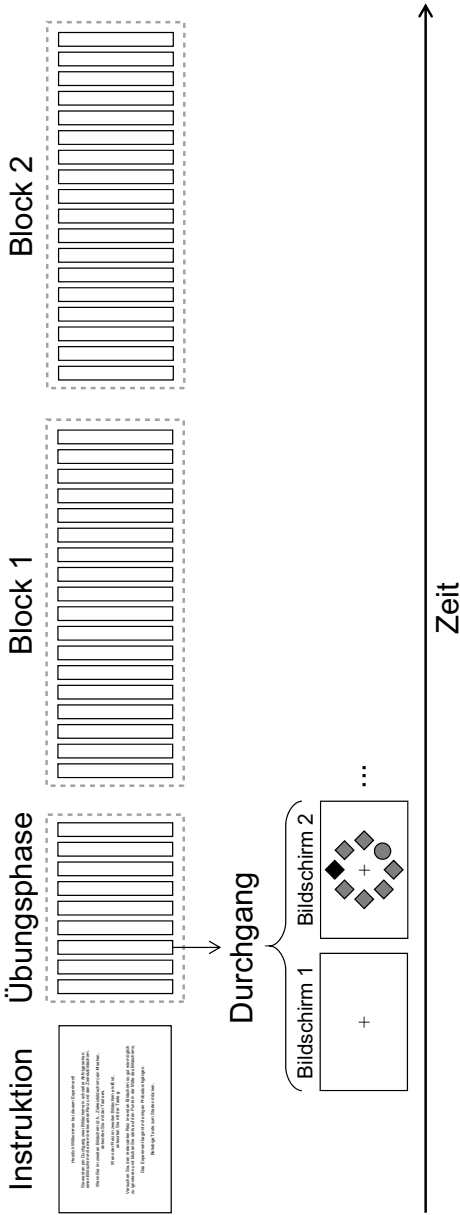


Abb. 2.1 Ein typischer Aufbau eines herkömmlichen Experiments: Nach der Instruktion für die Aufgabe erledigen die Versuchspersonen einige Durchgänge zur Übung (Übungsphase), woraufhin das Hauptexperiment mit der Datenerhebung beginnt. Dieses Hauptexperiment ist wiederum meist in mehrere Abschnitte (Block 1 & Block 2) unterteilt

A stammen und der nächste Durchgang zufällig aus Bedingung A oder Bedingung B. In *reinen Blöcken*, bzw. *geblockten Designs*, erledigen die Versuchspersonen über eine längere Zeit hinweg Durchgänge aus stets der gleichen experimentellen Bedingung. Für beide Arten von Blöcken kann es gute Gründe geben. Wollen Sie etwa Lerneffekte oder den Einfluss von Erwartungen und Vorbereitung über mehrere Durchgänge hinweg untersuchen, kann es sinnvoll sein, stets die gleiche Bedingung zu präsentieren. Wollen Sie aber Lerneffekte oder damit verwandte Alternativverklärungen möglichst ausschließen, dann sind gemischte Blöcke das Mittel der Wahl. Warum diese beiden Arten von Blöcken einen relevanten Einfluss auf die Studienergebnisse und deren Interpretation haben können, wird in Abschn. 2.6 noch zusätzlich näher beleuchtet.

Reines und gemischtes Design am Beispiel von Treisman und Gelade (1980)

Die Rolle der Aufmerksamkeit in der visuellen Suche ist eine traditionelle und seit Langem erforschte Forschungsfrage in der Kognitionspsychologie. Unter welchen Umständen können wir effizient relevante Zielreize von irrelevanten Distraktoren unterscheiden? Welche Suche ist schwieriger und welche ist leichter? Wann benötigen wir mehr oder weniger Aufmerksamkeit? Anne

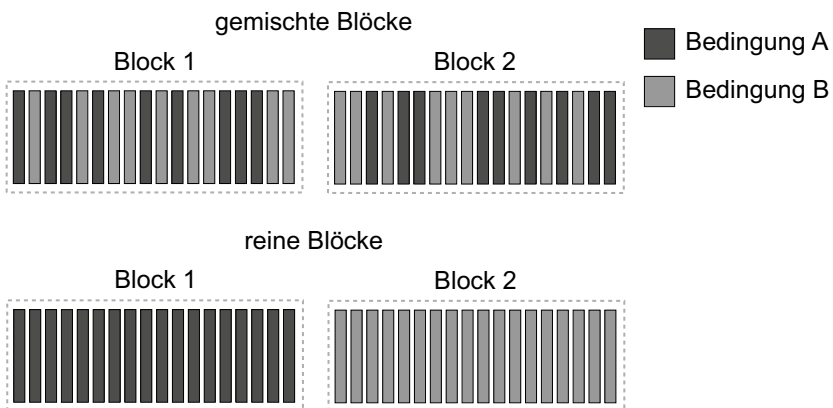
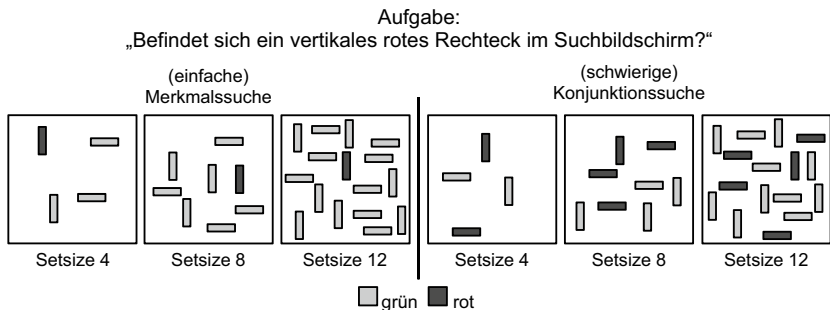


Abb. 2.2 In einem Experiment mit Messwiederholungen wird zumeist nicht nur eine Bedingung untersucht, sondern mehrere, die anschließend verglichen werden. Diese Bedingungen können entweder (pseudo-)randomisiert (gemischte Blöcke) oder in zwei verschiedenen Blöcken (reine Blöcke) dargeboten werden

Treisman und ihre Kolleginnen und Kollegen leisteten einen so großen Beitrag zur Beantwortung dieser Fragen, dass Treisman 2011 sogar von Präsident Obama mit der National Medal of Science ausgezeichnet wurde.

In ihrer *Feature-Integration Theory* stellten Treisman und Gelade (1980) ein Modell auf, das eine Erklärung für den Einfluss der Aufmerksamkeit unter schwierigen und einfachen Suchbedingungen erklären soll. Dazu ließen Treisman und Gelade ihre Versuchspersonen einfache und schwierige Suchaufgaben erledigen. Während des gesamten Experiments sollten die Versuchspersonen mittels Tastendruck angeben, ob sich ein Zielreiz im Suchbildschirm befindet oder nicht. In der *einfachen Suchbedingung* unterschied sich der Zielreiz stark von den Distraktoren (für eine sinngemäße Darstellung siehe Abbildung unten links). Genauer gesagt unterschied sich der Zielreiz in der einfachen Suchbedingung in einem Merkmal (z. B. Farbe) von den umgebenden Distraktoren.

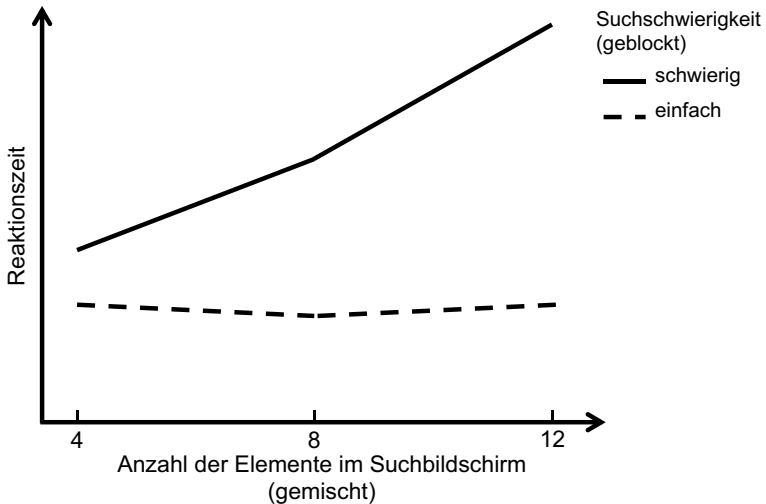
In der schwierigen Suchbedingung unterschied sich der Zielreiz von den Distraktoren jedoch nicht nur in einer Farbe oder einer Form von den Distraktoren, sondern in einer Kombination dieser beiden Merkmale. Das bedeutet, dass eine Merkmalsuche nicht mehr ausreichte, um den Zielreiz aufzuspüren, sondern eine *Merkmalskombinationssuche* (kürzer: Konjunktionssuche) nötig war, um den Zielreiz zu identifizieren. Versuchspersonen erledigten abwechselnd mehrere schwierige und mehrere einfache Suchblöcke.



Nun gut, wir sehen, wie Treisman und Gelade (1980) den Faktor der Suchschwierigkeit geblockt manipulierten. Wie wiesen sie aber nach, ob die Versuchspersonen effizient nach dem Zielreiz suchten oder nicht? Hier kommt ein zweiter Faktor ins Spiel: die Anzahl der Elemente im Suchbildschirm (bzw. kürzer im Englischen: *setsize*). In der Abbildung sehen Sie, dass 4, 8 oder 12 Reize im Suchbildschirm sein konnten. Warum?

Nehmen wir an, wir können die Anwesenheit eines Zielreizes ohne Umschweife sofort erkennen, weil sich der Zielreiz so markant von seiner Umgebung abhebt. Wie wirkt sich Ihrer Meinung nach die Anzahl der Distraktoren im Suchbildschirm aus? Genau, so gut wie gar nicht. Wir können den Suchbildschirm schnell scannen, die visuellen Reize praktisch *parallel* verarbeiten und entsprechende Abweichungen effizient und schnell erkennen. Wenn die Zielreize und Distraktoren einander allerdings hinreichend ähnlich sind, dann genügt diese parallele Suche nicht mehr. Stattdessen müssen wir die einzelnen Reize *seriell* absuchen und nach jeder Selektion eines Reizes entscheiden, ob es sich dabei um den Zielreiz handelt oder nicht. Entsprechend länger dauert die Suche.

Diese Annahme klingt vernünftig, doch wollen wir sie mit Daten stützen. Genau hier kommt die *Setsize*-Manipulation ins Spiel: Wenn die Suche effizient und parallel vonstatten geht, sollte sich die Zeit, die man zur Identifikation des Zielreizes benötigt, zwischen den verschiedenen *Setsize*-Bedingungen nicht wesentlich unterscheiden. Anders verhält es sich bei einer ineffizienten und seriellen Suche. Unter diesen Suchbedingungen hat man mit mehr ähnlichen Distraktoren noch mehr Reize, die man absuchen und klassifizieren muss. Daraus folgt, dass die Suchzeit als eine Funktion der Anzahl der Elemente im Suchbildschirm ansteigen sollte. Um diese Hypothese untersuchen zu können, variierten Treisman und Gelade (1980) die Anzahl der Elemente im Suchbildschirm zufällig von Durchgang zu Durchgang. Die *Setsize*-Bedingung wurde also innerhalb der Blöcke (= Suchschwierigkeit) gemischt dargeboten. Stellt man die Reaktionszeiten der Versuchspersonen im Verhältnis zur *Setsize* als eine Suchfunktion dar, konnten Treisman und Gelade ganz genau das finden:



In ihrer *Feature-Integration Theory* argumentieren Treisman und Gelade (1980), dass man Unterschiede in einzelnen Merkmalsdimensionen bereits ohne Aufmerksamkeitszuwendung („prä-attentiv“) identifizieren kann (ein hinreichend großer Merkmalskontrast vorausgesetzt, vgl. Duncan & Humphreys, 1989). Muss jedoch nach einer Kombination mehrerer Merkmale gesucht werden, benötigt es Aufmerksamkeit, um diese unterschiedlichen Merkmale zu kombinieren.

Wir kennen das auch aus dem realen Leben: Stellen Sie sich vor, sie fahren nachts mit dem Auto auf einer wenig befahrenen Landstraße. Sie passieren Bäume, Sträucher, Felder und plötzlich sehen sie etwas: einen gesperrten Bahnübergang. Für Sie ist es in einer solchen Situation vermutlich belanglos, wie viele Sträucher Sie neben der Straße sehen. Sofern die Sträucher oder Bäume die Ampel am Bahnübergang nicht verdecken, nehmen Sie dieses Warnsignal sofort wahr.

Fahren Sie mit Ihrem Auto jedoch in einer größeren Stadt (why? ... why???) und wollen herausfinden, ob sie Vorfahrt geben müssen, dann ist ein Vorfahrt-achten-Schild nicht auffällig genug, dass Sie sich locker zurücklehnen können und das Verkehrsschild schon Ihre Aufmerksamkeit einfängt. Stattdessen müssen Sie den Schilderwald gezielt und aufmerksam nach der Information absuchen, die für Sie relevant ist. Dreieckig sind nämlich mehrere Warnzeichen, ebenso wie rot. Sie müssen also gezielt nach einem auf dem Kopf stehenden, rot umrahmten Dreieck suchen. ◀

2.2 Max-Kon-Min Prinzip

Wir versuchen für gewöhnlich, in unseren Experimenten bestimmte Effekte nachzuweisen. Stellen wir uns ein Experiment daher kurz anhand eines sehr zeitgemäßen Beispiels vor: Wir möchten manuell zwischen Radiosendern wechseln (OK, Boomer...). Dabei ist die gewünschte Radiostation der Effekt, den wir erreichen wollen. Das Rauschen zwischen den einzelnen Stationen ist etwas, das wir minimieren wollen. Nur durch ein sensibles Herumdrehen des Knopfes können wir dabei unser gewünschtes Ergebnis erreichen: ein optimales Verhältnis des Signals (der Musik des Radiosenders) zum Hintergrundrauschen (in der englischsprachigen Literatur werden Sie hierzu oft „signal-to-noise ratio“ lesen können). Um unsere grauen Zellen noch zusätzlich zu fordern, stellen wir uns vor, dass der gewünschte Radiosender in Wien eine Frequenz von 92 MHz hat und in Innsbruck 87,6 MHz. Wir müssen also (1) das Signal **maximieren**, (2) die für die Örtlichkeit korrekte Frequenz wählen, also für den Ort **kontrollieren** und (3) das Rauschen **minimieren**.

Was hat das jetzt mit der Experimentalpsychologie zu tun? Um einen spezifischen Effekt zu finden, sollten wir uns zunächst bewusst sein, was dieser Effekt denn an und für sich ist: das Variieren der abhängigen Variable in Abhängigkeit der jeweiligen Bedingung (der Experimental- oder Kontrollbedingung). Wir untersuchen also, ob die Varianz unserer Daten durch die von uns gewählten Bedingungen erklärt werden kann. Das klingt nun vielleicht weniger trivial, als es eigentlich ist. Die Varianz, die wir in den Daten beobachten (Gesamtvarianz) kann nämlich durch drei Quellen zustande kommen (siehe Abb. 2.3): der Primärvarianz, der Sekundärvarianz und des Zufallsfehlers (Kerlinger, 1973).

Primärvarianz

Unter der Primärvarianz versteht man den Anteil der systematischen Varianz der durch die systematische Variation der Experimentalbedingungen (UV) zustande kommt. In einem guten Experiment gilt es, diese Primärvarianz zu maximieren. Dies wird durch die Wahl von optimalen Faktoren und Faktorstufen erreicht, die miteinander verglichen werden sollen. Ideal ist es hier, Extremstufen von Faktoren zu wählen, welche die Unterschiede zwischen den Bedingungen maximiert. Wenn Sie zum Beispiel demonstrieren wollen, dass kongruente Bedingungen in der Stroop-Aufgabe besonders hilfreich sind (beispielsweise das Wort „Rot“ in roter Farbe), dann vergleichen Sie diese Bedingung für gewöhnlich

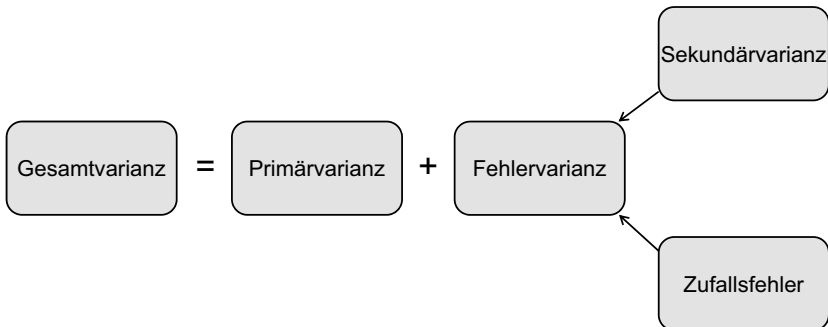


Abb. 2.3 Die Gesamtvarianz der abhängigen Variable in einem Experiment setzt sich aus der Primär- und Fehlervarianz zusammen. Die Fehlervarianz setzt sich ihrerseits wiederum aus der Sekundärvarianz und dem Zufallsfehler zusammen

nicht mit einer neutralen Bedingung (beispielsweise das Wort „Rindfleischetikettierungsüberwachungsaufgabenübertragungsgesetz“ in roter Farbe), sondern mit einer inkongruenten Bedingung (bspw. das Wort „Blau“ in roter Farbe). Zu Faktoren werden Sie im gleichnamigen Abschnitt noch Näheres lernen.

Sekundärvarianz

Unter der Sekundärvarianz versteht man den Anteil systematischer Varianz, der durch nicht berücksichtigte und unkontrollierte Faktoren zustande kommt. Die Sekundärvarianz kann die Interpretation der gefundenen Ergebnisse erschweren bzw. im schlimmsten Falle sogar verunmöglichen. Stellen Sie sich folgendes „Experiment“ vor: Wir wollen Geschlechtsunterschiede in den arithmetischen Fähigkeiten in einem Computerexperiment überprüfen, in dem die Versuchspersonen randomisiert Additionen, Subtraktionen, Multiplikationen und Divisionen durchführen sollen. Dafür rekrutieren Sie junge Frauen aus einem humanistisch ausgerichteten Gymnasium und junge Männer aus einer Höheren Technischen Lehranstalt (HTL). Ihre Ergebnisse suggerieren große Unterschiede in den mathematischen Kompetenzen zwischen den Geschlechtern: Männer erreichten signifikant mehr Punkte als Frauen. Kann man die Ergebnisse jedoch dahingehend interpretieren, dass Männer generell besser in Mathematik sind als Frauen? Mitnichten!

Eine Vielzahl an nicht berücksichtigten Variablen könnten diese Unterschiede erklären. Hier nur zwei wahrscheinlich höchst relevante Konfundierungen:

1. *Selbstselektion*: Es ist anzunehmen, dass sich technisch und mathematisch interessierte und begabte Personen eher für eine technische Schule entscheiden als Personen, die sich eher für Sprachen begeistern.
2. *Unterrichtsfächer*: Schüler:innen an einer HTL haben wesentlich mehr Unterrichtseinheiten, die sich mit Mathematik und verwandten Fächern beschäftigen, als Schüler:innen an einem humanistischen Gymnasium. Das bedeutet klarerweise, dass Schüler:innen an der HTL wesentlich mehr Übung in Arithmetik haben als jene, die ein humanistisches Gymnasium besuchen.

Sie sehen, die Interpretation des eben beschriebenen (und erfundenen!) Ergebnisses ist, streng genommen, gar nicht möglich. Die gefundenen Unterschiede könnten genauso gut durch die Konfundierungen erklärt werden.

Kollinearität

Kollinearität beschreibt in der Statistik das Ausmaß eines Zusammenhanges zwischen zwei Variablen (spezifischer: UVs). Kollinearität sollte in Verfahren wie etwa einer Regressionsanalyse tunlichst vermieden werden. Die Problematik sollte durch das oben genannte Beispiel ersichtlich sein: Korrelieren zwei Prädiktoren (etwa das Geschlecht und die Anzahl an Mathematikstunden) zu hoch miteinander, dann ist eine getrennte Interpretation der einzelnen Prädiktorvariablen in einem Regressionsmodell nicht möglich.

Zufallsfehler

Den Zufallsfehler kann man mit dem Messfehler aus der klassischen Testtheorie vergleichen, und tatsächlich gibt es zwei hervorstechende Gemeinsamkeiten mit dem Messfehler der klassischen Testtheorie:

1. Der Erwartungswert, also jener Wert, den eine Variable im Mittel nach unendlich vielen Messwiederholungen annimmt, ist beim Messfehler 0. Ähnlich verhält es sich mit der Varianz. Wir werden die Sekundärvarianz realistisch gesehen wohl nie auf 0 bekommen, werfen Sie aber einen Blick auf die Berechnung der Varianz:

$$var = \frac{\sum_{i=0}^n (x_i - \bar{x})}{n}$$

Anhand der Formel ist gut ersichtlich, dass die Größe der Varianz als eine Funktion der Stichprobengröße (n) abnimmt – je größer der Nenner wird, desto kleiner wird das Resultat (siehe Abb. 2.4).

2. Fast wichtiger als der erste Punkt: Der Zufallsfehler korreliert nicht mit dem gemessenen Effekt. Das bedeutet, dass der Zufallsfehler die Primärvarianz nicht systematisch beeinflusst, sondern lediglich Rauschen in den Daten ist, das es zu minimieren gilt, da eine Kontrolle des Zufalls ein gleichermaßen ermüdendes wie hoffnungsloses Unterfangen ist.

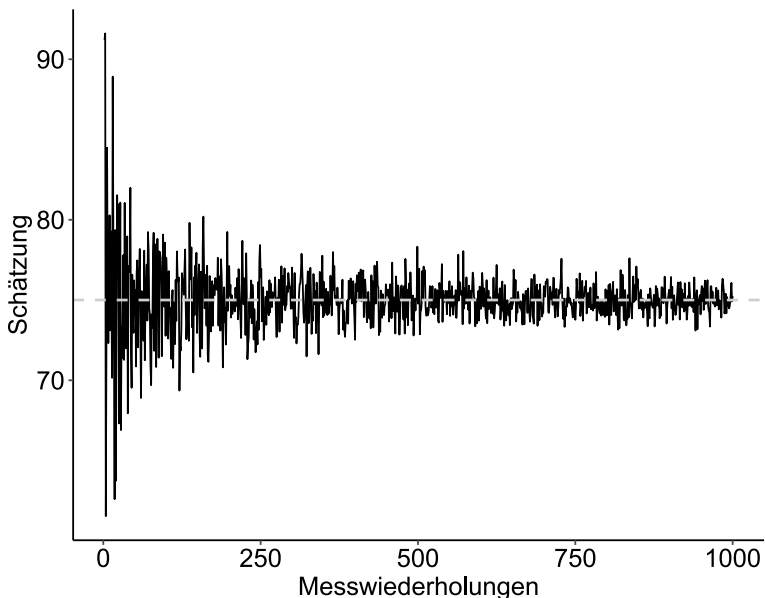


Abb. 2.4 Simulation zur Zunahme der Messgenauigkeit. Zwei bis 1000 Werte werden zufällig aus einer Normalverteilung (N[75; 25]) gezogen und gemittelt. Je mehr Werte gezogen werden, desto präziser wird die Schätzung des tatsächlichen Mittelwerts einer Variable (hier: die gestrichelte graue Linie)