

## 2. Datengewinnung

2	Datengewinnung
2.1	Erhebungen
2.1.1	Primärerhebung
2.1.2	Sekundärerhebung
2.2	Experiment
2.3	Grundgesamtheit und Stichprobe
2.3.1	Stichprobenauswahl
2.3.1.1	Technik der Zufallsauswahl
2.3.1.1.1	Randomisieren mit Zufallszahlen
2.3.1.1.2	Randomisieren mit Schlussziffern
2.3.1.1.3	Randomisieren mit Losen
2.3.1.1.4	Geschichtete Stichproben
2.3.1.1.5	Clusterstichproben
2.3.1.1.6	Mehrstufige Stichproben
2.4	Stichprobenfehler
2.5	Fehler bei der Datengewinnung
2.5.1	Zufallsfehler
2.5.2	Systematische Fehler, Bias
2.5.2.1	Konstante systematische Fehler
2.5.2.2	Variable systematische Fehler
2.5.3	Grobe Fehler
2.5.4	Erkennen von Fehlern
2.6	Biologische Variabilität
2.7	Ünungen

Im Kapitel „Datenformen“ haben wir diverse Arten von Daten kennengelernt. Wir wollen uns jetzt mit Methoden beschäftigen, mit denen wir empirische Daten gewinnen können. Untersuchungen wie etwa zu der Frage, welches von zwei Arzneimitteln die bessere therapeutische Wirkung hat, bedürfen einer sehr genauen Planung. Und zu dieser Planung gehören auch die Methoden der Datengewinnung.

Sorgfältiges Vorgehen ist bei der Datengewinnung von großer Bedeutung, da Fehler die hier – gleich zu Beginn einer Untersuchung – gemacht werden, später nicht mehr korrigiert werden können.

### Vorbemerkung

Wenn wir Daten erfassen, so ist der erste Schritt, mal abgesehen von der Datenerfassung durch EDV-Systeme, die Protokollierung in der Reihenfolge der Kenntnisnahme. Diese Protokollierung, die wir als Urdatenliste oder Rohdatenliste bezeichnen, ist häufig unübersichtlich, weswegen wir sie in „Reinform“ übertragen und/oder eine Zusammenfassung (aggregierte = zusammengefasste Daten, Tabelle oder Graphik) erstellen. Bei dieser Übertragung können Fehler auftreten z.B. Werte vertauscht werden. Aus diesem Grunde ist es unabdingbar, die ursprünglichen Protokollierungen aufzuheben um bei Unklarheiten gegebenenfalls darauf zurückgreifen zu können. Für die elektronische Datenerfassung gilt das Entsprechende.

In der Literatur werden die Verfahren zur Datengewinnung je nach Arbeitsgebiet (Medizin, Demoskopie, Naturwissenschaften) unterschiedlich strukturiert dargestellt. Wir wollen uns mit den Grundlagen der beiden Verfahren „Erhebung“ und „Experiment“ beschäftigen.

Unter Erhebungen verstehen wir Verfahren zur Datengewinnung bei denen wir auf die Faktoren, die die Daten beeinflussen können, keinen Einfluss haben.

Unter Experimenten verstehen wir Verfahren zur Datengewinnung, bei denen wir die Faktoren, die einen Einfluss auf die Daten haben, beeinflussen können.

## 2.1 Erhebungen

### 2.1.1 Primärerhebung

Bei einer Primärerhebung (prospektive Erhebung) sammeln wir Daten während diese beobachtbar sind. Dabei können wir die Faktoren, die eine Wirkung auf die Daten haben, nicht beeinflussen.

#### Beispiel 1

Ein Zoologe beschäftigt sich mit einer Untersuchung zur Individualentwicklung von **Zebras** in der Gefangenschaft. Er reist zu den Zebrageburten in die Zoos und misst die Geburtsgewichte um letztlich das mittlere Geburtsgewicht bei Zebras kennenzulernen. Dabei hat er keinen Einfluss auf Bedingungen, die das Geburtsgewicht beeinflussen könnten, wie z.B. Alter und Gesundheitszustand der Mutter, Jahreszeit der Geburt. Diese sind Bedingungen, die er nur zur Kenntnis nimmt und protokolliert. Nur auf die Messung selber hat er Einfluss.

#### Beispiel 2

Ich interessiere mich für den **Temperaturverlauf** über einen bestimmten Zeitraum und messe dazu täglich zu einer bestimmten Uhrzeit die Temperatur an einem bestimmten Ort. Auf die Entstehung der Daten habe ich keinen Einfluss aber ich protokolliere sie nach Ablesen des Thermometers.

Tag	3.1.09	4.1.09	5.1.09	6.1.09	7.1.09	8.1.09	9.1.09	10.1.09	11.1.09
° C	-7	-3	-2	-9	-13	-9	-7	-6	-6

Zu solchen Primärerhebungen zählen z.B. Biotopuntersuchungen zur Populationsdichte von Tieren und Pflanzen, deren Ergebnisse oft zu den Grundlagen für Bebauungspläne gehören.

#### Beispiel 3

Ein Biologe hat die Aufgabe in einem Hochmoor die Häufigkeit der typischen Heidekrautpflanzen Moosbeere, Rosmarinheide, Besenheide und Glockenheide, sowie der carnivoren Pflanze rundblättriger Sonnentau zu ermitteln. Er wird durch Begehung der zu untersuchenden Fläche des Moores das Vorkommen der Pflanzen protokollieren und so die gewünschte Biotopübersicht erstellen. Auf die Wachstumbedingungen der Pflanzen hat er dabei keinen Einfluss.

### 2.1.2 Sekundärerhebung

Wenn wir uns mit einer Fragestellung z.B. auf dem Gebiet der Entwicklung von Insektiziden befassen, dann werden wir, bevor es ans Experimentieren geht, zunächst in der Literatur recherchieren und Datenmaterial, welches andere zu dem Thema schon erarbeitet haben, analysieren. Die hier angewendete Art der Datengewinnung bezeichnen wir als Sekundärerhebung, retrospektive Erhebung oder Sekundärstatistik. Bei einer solchen älteren Datensammlung war das Ziel der Untersuchung möglicherweise ein ganz anderes als unseres. Wir können aus ihr aber das herausuchen, was uns interessiert. Zur Entstehung und Protokollierung der Daten liegen uns u.U. keine oder nur unvollständige Informationen vor. Wir haben keinen Einfluss auf Entstehung und Protokollierung und sind auf die Sorgfalt anderer Personen angewiesen. Dies ist ein Problem bei allen Sekundärerhebungen.

#### Beispiel 4

Wenn der Hochmoorbiologe von Beispiel 3 in seine Arbeit auch einen Vergleich mit dem Zustand des Moores von vor 30 Jahren aufnehmen möchte, so hätte er diesen Teil der Daten retrospektiv aus alten Aufzeichnungen erheben müssen. Ob die damaligen Forscher die gleiche Sorgfalt bei der Bestandsaufnahme hatten walten lassen, wie er heute, das ist ihm nicht bekannt. Daher die Unsicherheit bei Sekundärerhebungen.

#### Beispiel 5

Ich interessiere mich aufgrund aktueller Fragen für die **Diurese (Harnausscheidung) bei Ratten**. Bevor ich einen Versuch plane, recherchiere ich in alten Protokollen und finde Aufzeichnungen zur Diurese bei Ratten nach Applikation von Hypophysin. Die Aufzeichnungen sind vom 14.3.1978 und enthalten neben den Messwerten und der Dosierung die folgenden Angaben, die an 180 g – 200 g schweren Tieren des Stammes Wistar gewonnen wurden.

Nach min	15	30	45	60	75	90	105	120	135	150	165	180	195	210	225	240
mL Urin	0	0	0	1	2	4	5	6	7	8	8	8	10	11	12	13

Leider enthält das Protokoll keine Angaben über das Geschlecht der Tiere und die Uhrzeit zu der die Applikation erfolgte. Das wurde in der Originalaufzeichnung offensichtlich vergessen.

Dies ist ein häufiges Problem bei retrospektiven Erhebungen. Sie sind manchmal unvollständig und damit mehr oder weniger unsicher. Gelegentlich ist eine retrospektive Erhebung aber die einzige Möglichkeit zur Beantwortung einer Frage. Etwa, weil aus ökonomischen, ethischen oder anderen Gründen Untersuchungen nicht wiederholt werden können oder weil gezielt Geschehnisse aus der Vergangenheit zu analysieren sind. Wenn bei der Recherche nicht die Rohdaten zur Verfügung stehen also die Urliste, die bei der Datenerfassung niedergeschrieben wurde, sondern z.B. nur der daraus berechnete Mittelwert und die Streuung (aufbereitete, aggregierte Daten), so sprechen wir von einer Tertiärerhebung. Aggregierte Daten enthalten stets weniger Information als die Urliste, die ja eine „Kurzfassung“ der Urliste ist. Andererseits kann man aus den aggregierten Daten die Urliste nicht mehr rekonstruieren.

## 2.2 Experimente

Bei Arbeiten in Laboratorien ist in der Regel das Experiment die wichtigste Datenquelle. Das Typische für Experimente ist, dass wir die Bedingungen, die die Datenentstehung beeinflussen, planmäßig variieren können.

### Beispiel 6

Wenn wir die Abhängigkeit des Wachstums von **Fibroblasten** von Zusatzstoffen im Kulturmedium untersuchen wollen, dann können wir in parallelen Ansätzen dem oft verwendeten Medium DMEM unterschiedliche Mengen L-Glutamin, Natriumpyruvat oder gegen Mycoplasmeninfektionen Gentamicin zusetzen um die Zusammensetzung zu finden, in der unsere Zellen am besten wachsen.

So können wir den L-Glutaminzusatz variieren:

1. Ansatz 250 mg/L
2. Ansatz 300 mg/L
3. Ansatz 350 mg/L

Nach der Kultivierung würden wir die Zellzahl/mL Medium bestimmen und könnten dann eine Aussage darüber machen, welche Konzentration an L-Glutamin das beste Wachstum zur Folge hat.

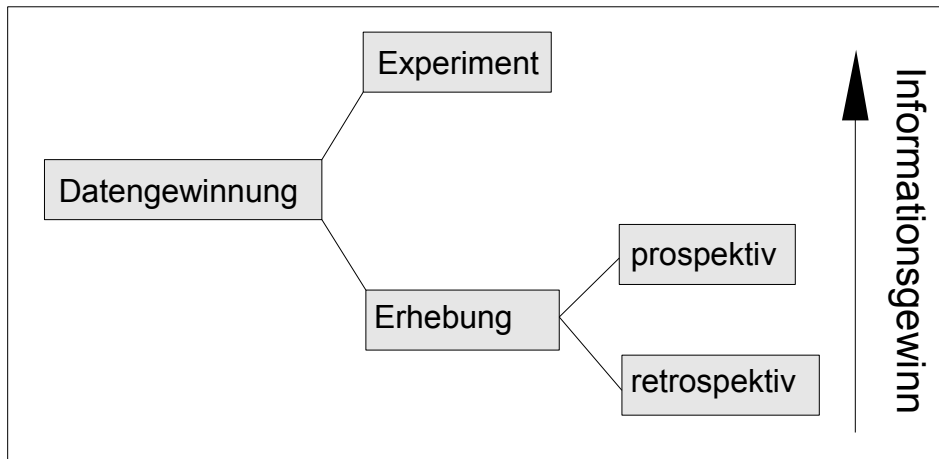
### Beispiel 7

Wollen wir das Wachstum von **Pflanzen** in Abhängigkeit vom Lehmgehalt des Kultursubstrats, von Niederschlag, Boden- und Lufttemperatur untersuchen, so können wir die Pflanzen in Substraten mit definierten Lehmgehalten kultivieren. Wir können in Klimakammern Niederschlag, Boden- und Lufttemperatur variieren und so Schritt für Schritt den Einfluss dieser variierten Bedingungen auf das Wachstum prüfen. Bei Wachstumsbeobachtungen im Freiland (Primärerhebung) wäre das nicht möglich, da müssten wir das Biotop mit seinen Eigenschaften so akzeptieren, wie wir es vorfinden.

### Beispiel 8

Wenn wir die therapeutische Wirkung eines Präparates gegen Hautparasiten an **Schafen** prüfen wollen, dann können wir die Tiere eingeteilt in Gruppen unterschiedlich stark infizieren. Wir können zu bestimmten, wählbaren Zeiten nach der Infektion mit verschiedenen Präparaten therapieren, wir können die Tiere nach Körpermasse, nach Geschlecht usw. Behandlungsgruppen zuordnen.

Dadurch, dass bei Experimenten nahezu alle Einflussmöglichkeiten auf die Daten geplant variiert werden können, ist das Experiment die Form der Datengewinnung mit dem höchsten Informationsgewinn.



## 2.3 Grundgesamtheit und Stichprobe

Wir wollen diese beiden grundlegenden Begriffe der Statistik, die im vorhergehenden Kapitel schon kurz ohne Erklärung erwähnt wurden, anhand einiger Beispiele genauer kennenlernen.

### Beispiel 9

Für ein Versuchsvorhaben benötigen wir 50 kg Saatgut der **Ackererbse** (*Pisum arvense*) und wollen das Material von einem Landwirt kaufen. Für unseren Versuch ist der Rohproteingehalt der Erbsen von Bedeutung. Wir müssen diesen Wert für die gesamte Ernte, von der wir ja 50 kg erhalten, kennen. Zur Ermittlung dieses Wertes stellt uns der Landwirt 50 g Erbsen zur Verfügung. Nach Aufbereitung des Materials erhalten wir durch eine Kjeldahlbestimmung den Proteingehalt  $w = 0,229$ . Die Frage ist nun, ob dieser, an 50 g Erbsen gewonnene Wert ein guter Schätzwert für den Rohproteingehalt der gesamten Erntemasse ist. Hätten wir vielleicht besser 500 g (oder gar die ganze Ernte) aufarbeiten sollen, um zu einem aussagekräftigeren Wert zu kommen?

Dieses Problem tritt bei der Planung von Versuchsvorhaben immer wieder auf. Wir benötigen Informationen über die Ausprägung eines Merkmals in einer bestimmten Menge von uns interessierenden Objekten (Merkmalsträgern). Aus verschiedenen Gründen können wir aber nicht alle Objekte der uns interessierenden Menge untersuchen, sondern nur einen begrenzten Teil davon. Hätten wir die gesamte Ernte, also alle Objekte, untersucht, so hätten wir zwar einen hervorragenden Wert für  $w(\text{Rohprotein})$  aber es wären keine Erbsen für unseren Versuch mehr da.

In diesem Zusammenhang wollen wir für die weiteren Überlegungen die Menge über die wir eine Aussage (wie „ $w(\text{Rohprotein der gesamten Ernte} = x$ )“ machen wollen, künftig als **Grundgesamtheit** (Population) bezeichnen und den Teil der Menge, den wir untersucht haben oder untersuchen wollen, als **Stichprobe**.

Bei der Auswahl von Stichproben aus Grundgesamtheiten tritt oft eine Problematik auf, die ich an dem folgenden Beispiel erklären möchte. Die vom Gesetzgeber geforderte Bestimmung der  $DL_{50}$  einer Substanz an Mäusen ist die Bestimmung der Dosis, die bei 50% der behandelten Tiere zum Tode führt. Das Ergebnis einer solchen Untersuchung, welches an z.B. 30 Tieren gewonnen wurde, verallgemeinern wir (Hochrechnung) auf „Mäuse“ eines bestimmten Stammes, Geschlechts usw. Die Grundgesamtheit, auf die wir das Ergebnis verallgemeinern, sind also alle Mäuse mit den genannten Eigenschaften. Dies ist eine fiktive – gedachte – Grundgesamtheit, deren Umfang unbekannt ist und aus der wir im Labor natürlich keinen Teil als Stichprobe auswählen können. Der Züchter wird uns die gewünschte Anzahl Tiere liefern, die er unseren Anforderungen, was die Eigenschaften der Tiere angeht, entsprechend aus den ihm zur Verfügung stehenden Tieren ausgewählt hat. Die 100 Tiere, die wir erhalten haben, sind für uns nun die „Grundgesamtheit“ oder ein Teil derselben, aus der wir unsere Stichprobe entnehmen werden. Wir verallgemeinern dann unser Stichprobenergebnis auf Mäuse, hatten aber nicht die Möglichkeit, die Stichprobe aus allen Mäusen der Grundgesamtheit auszuwählen. Wir können jedoch davon ausgehen, dass die uns zur Verfügung gestellten Mäuse in ihren Eigenschaften denen entsprechen, auf die wir verallgemeinern. Wenn wir also sagen, „wir wählen aus einer Grundgesamtheit von 100 Mäusen eine Stichprobe aus“, dann ist der Begriff Grundgesamtheit in dem erklärten Sinne zu verstehen. Diese Überlegungen gelten für viele Stichprobenauswahlen.

Wenn, falls es möglich ist, eine Grundgesamtheit vollständig untersucht wird, dann sprechen wir von einer Vollerhebung. Bei der Untersuchung einer Stichprobe von Teilerhebung. Weitere Gründe für Teilerhebungen zeigen die folgenden Beispiele.

### Beispiel 10

Wir benötigen **200 Mäuse** des Stammes NMRI, ♀, 15 – 17 Wochen alt und wollen vom Züchter das mittlere Gewicht der Tiere wissen. Die Grundgesamtheit ist in diesem Falle die Gesamtheit aller Tiere mit den gewünschten Merkmalen, die ihm zu diesem Zeitpunkt zur Verfügung stehen. Der Züchter kann die Tiere aus ökonomischen Gründen nicht alle wiegen und auch nicht die bestellten 200 Tiere. Er wird eine Stichprobe seiner Tiere wiegen und uns dem Mittelwert der Gewichte angeben.

### Beispiel 11

Wir interessieren uns für die Länge des Protozoons **Spirostomum ambiguum** in einem Tümpel. Die Grundgesamtheit, über die wir eine Aussage wie „Spirostomum hat eine mittlere Länge von  $x \mu\text{m}$ “ machen wollen, ist in ihrem Umfang unbekannt. Es ist praktisch unmöglich alle Tiere des Tümpels auszumessen. Wir werden uns mit einer Wasserprobe als Stichprobe begnügen und deren Ergebnis auf die Grundgesamtheit verallgemeinern.

### Beispiel 12

Der Hersteller von **Dragees** muss prüfen, ob die Dragees einer Charge den geforderten Wirkstoffgehalt haben. Wollte er die gesamte Charge untersuchen, dann könnte er keine Dragees mehr verkaufen, hätte aber einen sehr verlässlichen Wert des Wirkstoffgehaltes. Er wird also auch nur eine Stichprobe untersuchen. Das gleiche Argument gilt für alle Untersuchungen, bei denen das zu untersuchende Material zerstört wird, z.B. Brenndauer von Glühbirnen, Haltbarkeit von Autoreifen und Tabletten.

Wenn wir nun Untersuchungen an Stichproben vornehmen, dann stellt sich die Frage, wie wir vorgehen müssen, um die Elemente der Stichprobe aus den Elementen der Grundgesamtheit auszuwählen. Da wir die an einer Stichprobe gewonnene Information auf die Grundgesamtheit verallgemeinern wollen, muss die Information eine gute Schätzung für die Grundgesamtheit sein. Das bedeutet, dass die Stichprobe in ihrer Zusammensetzung repräsentativ für die Grundgesamtheit sein muss. Ist das nicht der Fall, dann kann das Stichprobenergebnis ein verfälschtes Bild der Grundgesamtheit geben. Das bedeutet dann, dass das Ergebnis der Untersuchung zu der ursprünglichen Frage wertlos ist.

### Beispiel 13

Eine **Erbsensuppe** besteht aus mehreren Komponenten, z.B. Erbsen, Fleischstückchen, Kartoffeln und Wasser. Die einzelnen Komponenten sind nicht gleichmäßig verteilt, da die Festbestandteile nach kurzer Zeit sedimentieren. Wollten wir den Geschmack der Suppe prüfen, so würden wir natürlich nicht die für die ganze Familie hergestellten 2 L essen, sondern eine Stichprobe davon. Bei der Auswahl derselben müssen wir beachten, dass die Stichprobe alle Teile der Suppe in der Zusammensetzung enthält, wie sie auch in der Grundgesamtheit vorliegen. Würden wir nur die Kartoffel probieren, so würden wir gar nicht merken, dass das Fleisch etwas salzig ist. Wir müssen also gut umrühren und dann eine "typische" Probe entnehmen.

Die Grundgesamtheit ist die Menge von Elementen, die bezüglich bestimmter Merkmale für eine konkrete Frage interessant ist.

Die Stichprobe ist eine Teilmenge der Elemente der Grundgesamtheit, die bezüglich der interessierenden Merkmale für die Grundgesamtheit repräsentativ ist.

Wir wollen jetzt lernen, was wir bei der Auswahl einer Stichprobe aus eine Grundgesamtheit beachten müssen.

## 2.3.1 Auswahl der Stichprobe

Damit eine Stichprobe die Grundgesamtheit hinreichend repräsentiert sind bei ihrer Auswahl folgende Aspekte zu beachten.

Repräsentanz der Stichprobe  
 Umfang der Stichprobe  
 Zufälligkeit bei der Auswahl der Stichprobe  
 Unabhängigkeit der Elemente der Stichprobe

Repräsentanz Die Stichprobe soll die charakteristischen Eigenschaften der Grundgesamtheit möglichst exakt abbilden. Nach dem Erbsensuppenbeispiel hier noch ein anderes zu diesem Aspekt.

### Beispiel 14

Es soll in der Bevölkerung einer Stadt eine **Meinungsumfrage** zu einem die Ethik betreffenden Thema durchgeführt werden. Als Grundgesamtheit gelten alle Personen > 18 Jahre in dieser Stadt. Da aus ökonomischen Gründen keine Vollerhebung durchgeführt werden kann, wird eine Stichprobe untersucht. Bei der Planung der Stichprobenauswahl muss darauf geachtet werden, dass in der Stichprobe Eigenschaften wie Geschlecht, Religionszugehörigkeit, Alter, Schulbildung und Einkommen, die die Meinung zum Thema beeinflussen können, annähernd mit dem gleichen Prozentsatz vorkommen wie in der Grundgesamtheit. Das zu realisieren ist nicht einfach. Eine simple Telefonumfrage (jede hundertste Telefonnummer im Telefonbuch) ist hier sicher nicht der richtige Weg, da dieses Auswahlverfahren die genannten Eigenschaften nicht berücksichtigt.

Umfang Es ist direkt einleuchtend, dass eine Stichprobe die Grundgesamtheit um so besser repräsentiert, je näher ihr Umfang ( $n$  = Anzahl der Elemente der Stichprobe) dem Umfang der Grundgesamtheit ( $N$  = Anzahl der Elemente in der Grundgesamtheit) kommt. Wie wir leicht einsehen können, verbietet sich eine "zu große" Stichprobe u.a. weil mit dem Stichprobenumfang die Kosten der Untersuchung wachsen. Aber auch die ethischen und gesetzlichen Auflagen z.B. bei Tierversuchen spielen hier eine Rolle. Im Übrigen führt aber einer Vergrößerung von  $n$  ab einem gewissen Stichprobenumfang nicht mehr zu einer nennenswerten Verbesserung des repräsentativen Charakters der Stichprobe. Wie groß der Stichprobenumfang in einem konkreten Fall sein muss, lässt sich näherungsweise berechnen, worauf wir hier aber nicht eingehen können. In vielen Fällen stützen wir uns auf Erfahrungswerte.

Zufälligkeit Um der Forderung nach Repräsentanz nachzukommen, müssen die Elemente der Grundgesamtheit, die in die Stichprobe kommen, streng nach den Regeln des Zufalls ausgewählt werden. Diese Notwendigkeit wird im folgenden Beispiel deutlich. Es muss gewährleistet sein, dass für alle Elemente der Grundgesamtheit die gleiche Chance besteht, in die Stichprobe gelangen zu können.

### Beispiel 15

Wir interessieren uns dafür, ob Mäuse durch Coffein in ihrer motorischen Aktivität beeinflusst werden. Dazu wollen wir eine Gruppe von 20 Mäusen auf vier Gruppen zu je 5 Tieren verteilen und die Gruppen wie folgt behandeln

Gruppe 1	Coffein 0,16 mg/kg
Gruppe 2	Coffein 0,32 mg/kg
Gruppe 3	Coffein 0,64 mg/kg
Gruppe 4	Kontrolle ohne Koffein

Wir wollten im Vorfeld wissen, ob die 20 ausgewählten Tiere eine gleichermaßen starke Aktivität hatten und entnahmen dem Käfig ein Tier nach dem anderen, maßen deren Aktivität und teilten sie der Reihe nach den Gruppen 1, 2, 3 und 4 zu. Erwartet hatten wir eine annähernd gleiche Aktivität der 20 Tiere. Wir erhielten vor der Behandlung aber folgende Werte:

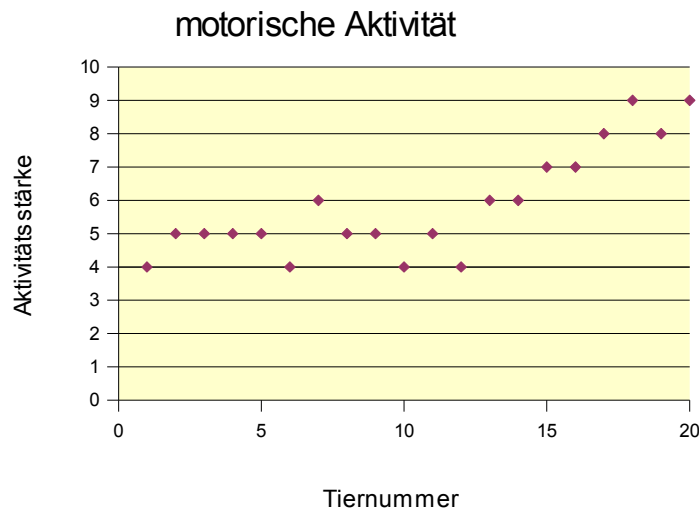


Abb.1

Der Graphik ist zu entnehmen, dass die zuletzt gefangenen Tiere deutlich höhere Aktivitäten aufweisen als die übrigen. In der Reihenfolge der Entnahme können wir die Tiere demnach nicht auf die Gruppen verteilen, denn dann kämen in die Kontrollgruppe die aktivsten Tiere. Was haben wir bei der Auswahl falsch gemacht? Die 20 Tiere wurden nicht zufällig den vier Gruppen zugeordnet. Die Tiere mit relativ geringer Aktivität waren leichter zu fangen und die zuletzt gefangenen Tiere hatten aufgrund ihrer stärkeren Aktivität der fangenden Hand immer wieder entkommen können. Jeder, der schon mal Mäuse auch einem Vorratskäfig gefangen hat, kennt das. Hätten wir den Fehler nicht erkannt, dann wäre das Versuchsergebnis sicher verfälscht worden, denn die Kontrolltiere hatten eine stärkere Grundaktivität als die übrigen Tiere. Nachdem der Fehler erkannt war, haben wir die Tiere durchnummeriert und sie dann durch Zufallsauswahl (Randomisieren, siehe weiter unten) auf die vier Gruppen verteilt.

Nur durch eine Zufallsauswahl kann man sicherstellen, dass keine gerichtete Auswahl (hier nach der Aktivität) vorkommt, die das Versuchsergebnis verfälscht.

### Beispiel 16

Ein anderes Beispiel zum gleichen Problem. Wir haben in einem Behälter 20 **Blutegel** (*Hirudo medicinalis*) und wollen für ein Experiment fünf auswählen. Manche Tiere sitzen träge an der Gefäßwand festgesaugt und andere schwimmen agil herum. Letztere sind schwerer zu fangen. Würden wir die fünf Tiere entnehmen, die am einfachsten zu fangen sind (die festsitzenden) so hätten die agilen eine schlechtere Chance gehabt als die trägen in die Stichprobe zu kommen. Nun sind Blutegel schlecht zu kennzeichnen. Wir könnten uns aber helfen, indem wir die Tiere einzeln in kleine Gefäße setzen. Dann könnten wir wie wir später noch sehen werden, durch eine Zufallsauswahl (randomisieren) fünf Gefäße für die Stichprobe auswählen.

### Unabhängigkeit

Die Elemente, die in die Stichprobe gelangen, dürfen nicht voneinander abhängig sein. Damit soll vermieden werden, dass die Ausprägung des einen Elements von der eines anderen beeinflusst wird.

### Beispiel 17

Dies ist etwa wichtig bei **Meinungsumfragen**. Wenn alle Mitglieder einer Familie nach ihrem Wahlverhalten am nächsten Wahltag befragt werden, dann wird u.U. bei verschiedenen Personen die gleiche Meinung geäußert, da innerhalb der Familie eine Angleichung der Meinungen stattgefunden haben könnte. Sinnvoller wäre es, Personen zu befragen, die sich nicht kennen und sich nicht in der Meinung zu dem Problem ausgetauscht haben.

**Beispiel 18**

Bei Mäusen wollten wir die Dauer der Vollnarkose nach der Gabe von Nembutal-Na in der Dosierung 50 mg/kg bestimmen. Die 20 narkotisierten Mäuse lagen in Abständen von ca. 7 cm auf einem Tisch nebeneinander auf dem Rücken. Unsere Aufgabe bestand darin, den Zeitpunkt zu protokollieren, zum dem die Tiere sich von sich aus, ohne Störung durch andere Mäuse aus der Rückenlage in die Bauchlage umdrehten. Der Zeitwert jeder Maus musste also unanhängig vom „Wachwerden“ anderer Tiere festgestellt werden. Dazu war es notwendig, eine Maus (Nr. 7), die sich umgedreht hat, sofort vom Tisch zu entfernen. Im anderen Fall könnte sie herumlaufen und eine andere Maus (Nr.5) früher, als diese es von sich aus getan hätte, zum Umdrehen bewegen. Der Zeitwert von Maus 5 wäre dann nicht unabhängig von Maus 7, denn diese hätte Maus 5 beeinflusst. Würden wir einen solchen Fall nicht bemerken, so wären die Versuchsergebnisse fehlerbehaftet.

Es gibt in der später zu besprechenden schließenden Statistik allerdings Tests, die abhängige Stichproben erfordern.

**2.3.1.1 Technik der Zufallsauswahl**

Nachdem wir nun wissen, welche Aspekte bei der Stichprobenauswahl zu beachten sind, wollen wir sehen, **wie** man die Auswahl aus der Grundgesamtheit konkret durchführt. Bei vielen Forschungsvorhaben wird die Auswahl durch ein PC-Programm realisiert. Wir wollen hier aber die klassischen Verfahren kennenlernen. Nach dem englischen Wort für „aufs Gratewohl“ at random bezeichnen wir Verfahren bei denen der Zufall die Auswahl steuert, als **Randomisieren**. Wir wollen uns hier mit einige einfache Techniken des Randomisierens beschäftigen.

**2.3.1.2 Randomisieren mit Zufallszahlen.**

Zunächst einige Bemerkungen zu Zufallszahlen. Dies sind Zahlen, die in zufälliger Folge in einer Liste aufgeführt sind. Wir finden solche Listen in der Literatur als Zufallszahlentabellen, können sie aber auch selber mit einem zehneitigen Würfel mit den Ziffern 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 erstellen. Wir würden dazu die Ergebnisse der Würfe als Ziffernfolge fortlaufend notieren. In einer Zufallszahlentabelle hinreichender Größe müsste jede der 10 Ziffern mit annähernd gleicher Häufigkeit vorkommen. Vorsicht! Ein willkürliches Aufschreiben von Ziffern, die uns gerade einfallen führt, wie wir leicht prüfen könnten, **nicht** zu einer Zufallsfolge. Das bedeutet, es gibt Häufungen oder Minderungen im Auftreten bestimmter Ziffern, die wir offensichtlich bevorzugt oder unterdrückt haben. Zufallszahlen können wir leicht mit einem Tabellenkalkulationssystem erzeugen. Beim Randomisieren mit Zufallszahlen müssen wir alle Elemente der Menge, aus der wir die Stichprobe ziehen wollen, nummerieren. Das Verfahren ist nur durchführbar, wenn diese Menge (Grundgesamtheit) überschaubar klein genug ist, um den Aufwand der Kennzeichnung ökonomisch vertretbar zu halten. In vielen Fällen ist die Größe der Grundgesamtheit unbekannt, weswegen eine Durchnummerierung dann gar nicht möglich ist.

**Beispiel 19**

Wir wollen wissen, ob Wirkstoff A bei **Mäusen** eine bessere Wirkung gegen eine Trypanosomeninfektion hat als Wirkstoff B. Die in diesem Fall vom Tierlieferanten vorausgewählte „Grundgesamtheit“ bestehe aus 100 Mäusen mit den gewünschten Eigenschaften bezüglich Stamm, Geschlecht und Gewicht. Für den Versuch wird eine Stichprobe von 20 Tieren benötigt. Wir kennzeichnen alle 100 Tiere in fortlaufender Folge mit den Zahlen 00, 01, 02 bis 99. Jedes Tier bekommt also eine zweiziffrige Codezahl. Wären 300 Tiere zu kennzeichnen, so bekäme jedes Tier eine dreiziffrige Codezahl von 000 bis 299. Nun suchen wir mit Hilfe von Zufallszahlen 20 Tiere für die Stichprobe aus. In der Tabelle sind aus Gründen der Übersichtlichkeit jeweils fünf Ziffern zu einem Block zusammen gefasst. Wo immer wir beginnen diese Tabelle zu lesen und ob wir nach rechts, links, oben oder unten lesen, immer bekommen wir eine Folge zufällig angeordneter Zahlen. Es ist auch unerheblich, ob wir einziffrige Zahlen, zweiziffrige Zahlen usw. entnehmen. Bei der Suche nach 20 Tieren aus den 100 Tieren verfahren wir wie folgt: Wir legen willkürlich fest, dass wir in der 4. Zeile von oben und in der 4. Spalte von links beginnen wollen und nach rechts lesen (siehe Tabelle). Ist die Zeile zu Ende, fahren wir mit der nächsten Zeile darunter, links beginnend, fort. Jede andere Vereinbarung wäre auch möglich. Sie muss aber vorher festgelegt werden, um zu verhindern, dass wir mit einer willkürlich ausgesuchten Startzahl beginnen. Wir fangen mit der zweiziffrigen Zahl 74 an und teilen Maus 74 dem Käfig A zu. Dann Maus 46 zu Käfig B, Maus 47 zu Käfig A usw. Die Zahlen schreiben wir auf. Nach Maus 99 (Käfig B) finden wir die Zahl 06. Diese Zahl, wie auch nach Maus 72 die Zahl 58 überschlagen wir, da die Zahlen 06 und 58 bereits vergeben sind.



Das Ergebnis sind die Zufallszahlen:

74, 46, 47, 40, 84, 06, 12, 61, 68, 99, 06, 86, 18, 20, 95, 30, 58, 31, 72, 58, 42, 16
--

Die 20 Mäuse mit den obigen Nummern sind nun rein zufällig in die Stichprobe gelangt wobei die folgende Zuordnung resultiert.

Käfig A:	74 47 84 12 68 86 20 30 31 42
Käfig B:	46 40 06 61 99 18 95 58 72 16

Wir haben auf diesem Wege also aus 100 Mäusen rein zufällig 20 ausgewählt und diese wiederum zufällig in zwei Gruppen aufgeteilt.

### 2.3.1.1.2 Randomisieren mit Schlussziffern

#### Beispiel 20

Wir wollen aus einer Gruppe von 60 **Kanarienvögeln**, deren Ringe gekennzeichnet sind, einen bestimmten Prozentsatz für eine Stichprobe auswählen. Alle Tiere der Gruppe sind fortlaufend nummeriert. Sollen 20% der Gruppe ausgewählt werden, so legen wir in freier Wahl zwei Ziffern fest, z.B. 4 und 9. Alle Tiere mit den Endziffern 4 und 9 kommen nun in die Stichprobe. Bei einer Gruppe von 60 Tieren wären das die Tiere 4,9,14,19,24,29,34,39,44,49,54,59. Das sind 12 Tiere und somit 20% der Gruppe. Benötigen wir 30% der Gruppe, so werden drei Endziffern festgelegt.

Die Ringe der Tiere üblicherweise nicht von 00 bis 59 nummeriert, sie tragen wahrscheinlich beliebige Zahlen- oder Buchstabenkombinationen der Züchter, denen wir aber eine von 00 bis 59 fortlaufende Codierung zuordnen können..

Etwa so:

Ring AB 7312	erhält Nummer 00
Ring CD 14	erhält Nummer 01
Ring CD 234	erhält Nummer 02
	usw.
Ring FT 33	erhält Nummer 99

### 2.3.1.1.3 Randomisieren mit Losen

#### Beispiel 21

Wir interessieren uns für die Wirkung des Herbizids Glyphosate auf die Anthocyanbiosynthese bei **Buchweizensämlingen**. Zur Untersuchung sollen 15 Töpfe mit 7 Tage alten vorkultivierten Pflanzen zufällig auf 3 Gruppen verteilt werden, die mit unterschiedlichen Dosen des Wirkstoffs zu behandeln sind. Die Töpfe werden von 1 – 15 nummeriert. Damit die Zuteilung zu den Gruppen A, B und C zufällig geschieht, arbeiten wir hier mit Loskarten, die mit den Zahlen 1 bis 15 gekennzeichnet sind. Die Karten werden gemischt und alternierend den drei Gruppen zugeordnet wie die folgende Darstellung zeigt.

Gruppe A	Gruppe B	Gruppe C
Losnummer	Losnummer	Losnummer
12, 3, 8, 10, 9	4,6,11,1,7	2, 5,13,14,15

Nun gelangt Topf 1 in Gruppe B, Topf 2 in Gruppe C, Topf 3 in Gruppe A usw. Dieses Verfahren ist sehr einfach, wir benötigen keine Zufallszahlentabelle und die Loskarten kann man rasch selber erstellen.

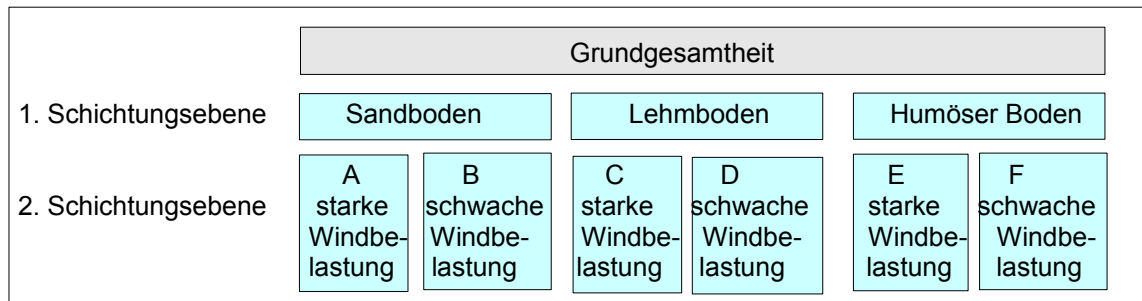
### 2.3.1.1.4 Geschichtete Stichproben

Es gibt Situationen, in denen die Grundgesamtheit sehr heterogen ist. Das kann bedeuten, dass in ihr Gruppen mit interessierenden Eigenschaften vorhanden sind, die jede für sich, sehr klein sind. Würden wir eine repräsentative Stichprobe entnehmen, so würden bei einer vertretbar kleinen Stichprobe, aus diesen kleinen Gruppen kaum Elemente in die Stichprobe gelangen. In solchen Fällen bilden wir geschichtete Stichproben. Hier hat nicht jedes Element der Grundgesamtheit die gleiche Chance in die Stichprobe zu kommen. Das Verfahren beruht darauf, dass wir die Grundgesamtheit in mehrere Teile, eben die Schichten, zerlegen, die jeweils in sich homogener sind. Aus jeder Schicht

wird dann eine eigene Stichprobe entnommen.

### Beispiel 22

Wir untersuchen im Zusammenhang mit einem **Schädlingsbefall** bei einer Wildpflanzenart den Zustand der Schädlinge im Freiland. Der Befall in der Grundgesamtheit variiert stark in Abhängigkeit von den Biotopbedingungen Bodenart und Windbelastung. Auf Lehmf Flächen finden wir sehr wenige Schädlinge. Wir schichten daher das Biotop nach der Bodenart, auf der die Pflanzen wachsen und nach der Windbelastung der Pflanzen. Die Schichtung ergibt folgendes Bild:



Aus den sechs Schichten wir nun per Zufall - wie wir es kennen - die geforderte Anzahl an Arealen für die Stichprobe ausgesucht. Alle Pflanzen der Stichproben werden dann begutachtet. Mit diesem Verfahren ist eher sichergestellt, dass alle Standortvariablen gleichermaßen in der Stichprobe vertreten sind. Auch bei den Schichten C und D (Lehmböden) werden wir genügend Schädlinge für deren Untersuchung finden. Bei einer repräsentativen Stichprobe hätten wir wahrscheinlich in den beiden Gruppen C und D kaum einen Schädling gefunden.

#### 2.3.1.1.5 Clusterauswahl

Ein typisches Beispiel dieses Verfahrens ist demographischer Art.

### Beispiel 23

In einer Kleinstadt soll das **Kaufverhalten** der über 18 jährigen Bewohner untersucht werden, die regelmäßig beim Discounter einkaufen. Da keine Informationen vorliegen, die diese Gruppe kennzeichnen, können wir aus ihr auch keine Stichprobe entnehmen. In einem solchen Fall teilen wir die bewohnte Fläche der Stadt in Raster, z.B. in Planquadrate geeigneter Größe ein. Wir bilden auf diesem Wege Klumpen (Cluster) von Bewohnern, in denen sich wahrscheinlich auch Mitglieder der gesuchten Gruppe befinden. Die Cluster werden durchnummeriert und es wird eine Stichprobe je Cluster gewählt. In den Stichprobenclustern werden alle Bewohner nach ihrem Kaufverhalten befragt. Durch dieses Verfahren hatten nicht alle Bewohner der Stadt die gleichen Chancen in die Stichprobe zu gelangen, sie wurden der Stichprobe indirekt durch Clusterauswahl zugeführt.

#### 2.3.1.1.6 Mehrstufige Auswahl

Hierunter verstehen wir eine Stichprobenauswahl aus einem schon selektierten Teil der Grundgesamtheit.

### Beispiel 24

Wir bestellen beim Züchter 200 **Goldhamster**, männlich, Stamm Hoe:SYHK, 150 g  $\pm$  10 g. Der Züchter hat zum Zeitpunkt, an dem er den Versandkarton für uns fertig macht, 2000 solcher Tiere vorrätig. Er teilt sie für uns und andere Kunden in einer ersten Selektierungsstufe in Käfige zu je 50 auf. Wir erhalten vier Käfige. Im der zweiten Selektierungsstufe wählen wir aus den vier Käfigen einen aus und in einer dritten Stufe aus diesem eine Stichprobe von n = 20. Diese 20 Tiere sind aus der Grundgesamtheit mit N = 2000 über ein mehrstufiges Verfahren ausgewählt worden. Dieses Vorgehen ist in der Praxis sehr häufig.

## 2.4 Der Stichprobenfehler e

Stichprobenergebnisse repräsentieren immer nur mehr oder weniger gut die Merkmalsausprägungen in der Grundgesamtheit. Bei einer Untersuchung variieren üblicherweise die Messwerte und wir geben häufig als Ergebnis der Untersuchung stellvertretend an Stelle aller Messwerte das arithmetische Mittel  $\bar{x}$  (gesprochen x quer) der Messungen an. Dabei müssen wir den Stichprobenmittelwert  $\bar{x}$  als Schätzwert für  $\mu$ , den Mittelwert der Grundgesamtheit, den sogenannten "Erwartungswert" ansehen.

Der Idealfall  $\bar{x} = \mu$  wird wegen des begrenzten Stichprobenumfangs in der Regel nicht erreicht und wir formulieren die Abweichung

$$\bar{x} - \mu = e$$

als Stichprobenfehler  $e$ . Je größer der Stichprobenumfang  $n$ , umso besser repräsentiert die Stichprobe die Grundgesamtheit, umso kleiner wird  $e$ . In der Regel ist  $e$  nicht bekannt. Wir können  $e$  aber aus Kenntnissen über die Variation der Mittelwerte vieler vergleichbarer Stichproben aus einer Grundgesamtheit schätzen. Auf das Verfahren können wir hier nicht eingehen. Die Schätzung von  $\mu$  durch  $\bar{x}$  wird mit zunehmendem  $n$  so gut, dass eine Vergrößerung des Stichprobenumfangs dann oft nicht mehr sinnvoll ist.

## 2.5 Fehler bei der Datengewinnung

Bei der Gewinnung von Daten und deren Aufbereitung zur späteren Analyse können Fehler auftreten. Diese Fehler können das Ergebnis einer Untersuchung im günstigsten Falle fragwürdig erscheinen lassen, etwa wenn uns die Werte unwahrscheinlich vorkommen. Dann können wir nach eventuellen Fehlern suchen und diese gegebenenfalls eliminieren. Im schlimmsten Fall können wir an den Daten nicht erkennen, dass sie fehlerhaft sind und wir ziehen dann Schlüsse aus falschen Ergebnissen. Wir wollen uns hier mit den drei wesentlichen Fehlertypen beschäftigen.

Dies sind

Zufallsfehler  
Systematischer Fehler  
Grober Fehler

### 2.5.1 Zufallsfehler (statistischer Fehler, persönlicher Fehler)

#### Beispiel 25

Bei einer Gesundheitskontrolle soll die  $K^+$ -Konzentration im Blut festgestellt werden. Ein Laborant analysiert die Blutprobe. Er führt fünf Parallelbestimmungen der gleichen Blutprobe durch und erhält folgende Ergebnisse der Stoffmengenkonzentration  $c(K^+)$  in mmol/L:

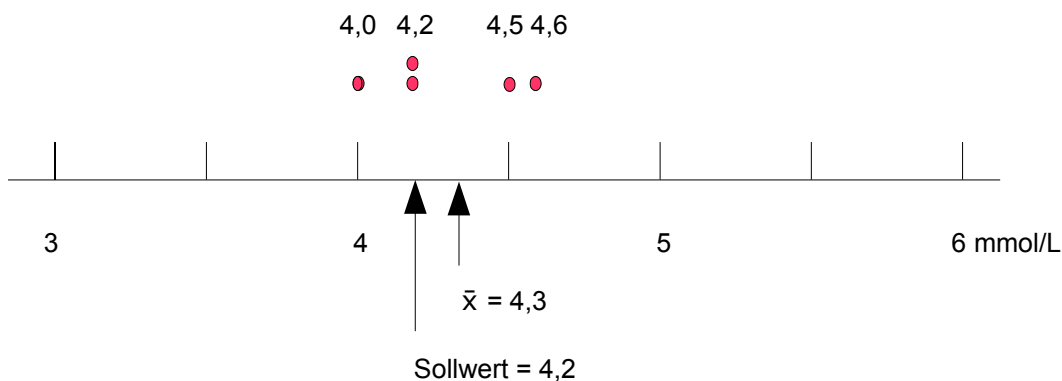


Abb.2

Der Mittelwert der fünf Messungen beträgt 4,3 mmol/L. Der „richtige“ von einem Analysenautomaten gefundene Wert, der Sollwert, ist 4,2 mmol/L. Wir ziehen daraus folgende Schlüsse:

1. Die fünf Einzelwerte streuen um ihren Mittelwert herum.
2. Der Mittelwert liegt nahe beim Sollwert.

Dieses Bild ist typisch für den Zufallsfehler. Für ihn ist nämlich charakteristisch, dass die Messwerte mal zu hoch und mal zu niedrig liegen. Die Gründe für diese Streuung liegen in der Arbeitsweise des Analysierenden (daher persönlicher Fehler).

**Gründe für Zufallsfehler**

Unaufmerksamkeit, Ablenkung, Ermüdung  
 Empfindlichkeit der Augen lässt nach (Analogskala falsch ablesen)  
 Pipettierfehler (mal zu viel, mal zu wenig abgemessen)  
 Allgemeine Arbeitsweise

Zufallsfehler können wir nicht vermeiden, wir streben aber an, sie zu minimieren.

Der Zufallsfehler ist ein Maß dafür, wie weit die Werte um ihren Mittelwert streuen. Er ist ein Maß für die Präzision der Arbeit.

## 2.5.2 Systematischer Fehler (methodischer Fehler, Bias)

**Beispiel 26**

Das Ergebnis des vorherigen Beispiels hätte auch so ausfallen können:

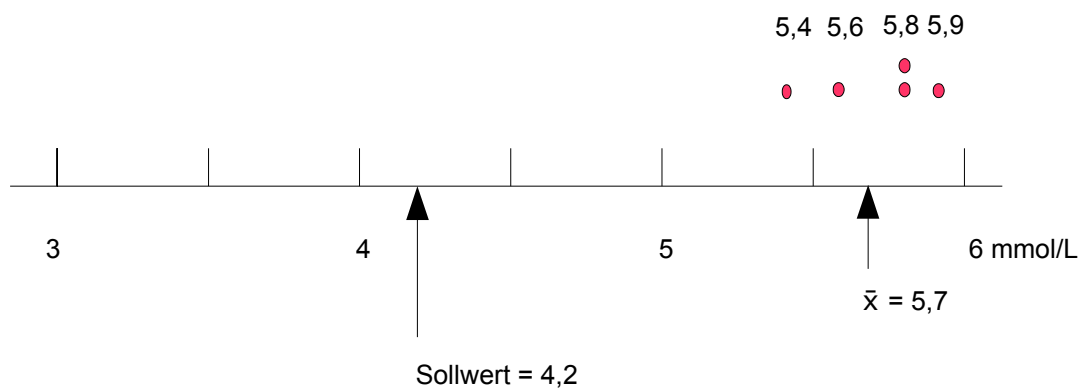


Abb.3

Wie im obigen Beispiel streuen die fünf Ergebnisse etwa gleich stark um ihren Mittelwert 5,7 mmol/L. Dafür ist wieder der Zufallsfehler verantwortlich. Der Mittelwert 5,7 mmol/L weicht aber auffallend weit vom Sollwert 4,2 mmol/L ab. Dieser liegt sogar außerhalb der Spanne der fünf Messwerte. Alle Werte sind auf der Skala nach rechts verschoben. Das kann nicht mehr durch den Zufall erklärt werden, denn dann würden die Werte in der Nähe des Sollwertes streuen. Hier liegt ein systematischer Fehler vor, der zu falschen Ergebnissen führt. Bei einem systematischen Fehler entstehen Werte, die immer zu niedrig oder immer zu hoch bezogen auf den Sollwert liegen. Der systematische Fehler wird nach dem englischen Begriff für „befangen“ auch als „bias“ bezeichnet.

**Gründe für systematische Fehler**

Die falsche Pipette benutzt.  
 Die Waage war nicht richtig nivelliert.  
 Zum Ablesen wurde die falsche Kalibrationskurve benutzt.  
 Am Photometer wurde die falsche Wellenlänge eingestellt.  
 Konzentration der Standardlösung war fehlerhaft.  
 Probenmaterial und/oder Reagenzien falsch gelagert.

Der systematische Fehler ist ein Maß dafür, wie weit die Werte in einer Richtung vom Sollwert abweichen. Er ist ein Maß für die **Richtigkeit** der Ergebnisse.

Durch sachlich richtiges Arbeiten können systematische Fehler vermieden werden.

### 2.5.2.1 Konstanter systematischer Fehler

#### Beispiel 27

Im physikalischen Praktikum wird ein **Experimental-Flüssigkeitsthermometer** benutzt, bei dem wir die Skala längs der Kapillare verschieben können. Bei der richtigen Vorgehensweise wird vor einer Messung die Lage der Skala mit einer geeichten Thermometer abgeglichen, so dass bei beiden Thermometern der gleich Wert, z.B. 0 °C angezeigt wird. Hierbei trat ein Einstellfehler auf, in dessen Folge jeder Messwert um 2 °C zu hoch liegt. Die folgenden Zahlen zeigen, wie sich dieser Fehler im Vergleich mit richtigen Messwerten bei verschiedenen hohen Messwerten auswirkt.

Thermometer richtig kalibriert zeigt bei 0 °C 0 °C an	Thermometer falsch kalibriert zeigt bei 0 °C 2 °C an	Absolute Abweichung	Relative Prozentuale Abweichung
A	B	C	D
		$ A-B $	$100 \cdot  A-B /A$
10	12	2	20%
20	22	2	10%
30	32	2	7%
40	42	2	5%

Beispielrechnung für A = 20 °C

Absolute Abweichung  $|20 - 22| = |2|$

Relative Abweichung Bei 20 °C ist die absolute Abweichung 2 °C  
Bei 100 °C sind dies 10 °C, das entspricht 10%

**Typische Eigenschaft des konstanten systematischen Fehlers:**  
Der absolute Fehler ist bei allen Messwerten gleich hoch.  
Der relative Fehler (prozentuale Fehler) sinkt mit zunehmender Messwerthöhe.

### 2.5.2.2 Variabler systematischer Fehler

#### Beispiel 28

Hier geht es um einen Fehler bei einer **maßanalytischen Titration**. Bei der Faktorisierung der Maßlösung ist ein Fehler aufgetreten. Während der richtige Faktor 1,02 ist, wurde der Faktor auf der Flasche mit 1,04 angegeben. Das Zahlenbeispiel zeigt, wie sich dieser Fehler auf die Titrationsergebnisse auswirkt.

Verbrauch	falscher Titer	falsches Ergebnis	richtiger Titer	richtiges Ergebnis	absolute Abweichung	Relative prozentuale Abweichung
A	B	C	D	E	F	G
					$ E-C $	$100 \cdot  (E-C) /E$
10,00 mL	1,04	10,40 mL	1,02	10,20 mL	0,20 mL	1,96%
12,00 mL	1,04	12,48 mL	1,02	12,24 mL	0,24 mL	1,96%
14,00 mL	1,04	14,56 mL	1,02	14,28 mL	0,28 mL	1,96%
16,00 mL	1,04	16,64 mL	1,02	16,32 mL	0,32 mL	1,96%

Wie wir sehen, hängt die absolute Abweichung von der Höhe des Messwertes ab, da der falsche Titer ja multiplikativ (wahrer Verbrauch = praktischer Verbrauch \* Faktor) in das Ergebnis eingeht. Der prozentuale Fehler dagegen ist unabhängig von der Höhe des Verbrauchs immer gleich. Bei der Datengewinnung müssen wir uns überlegen, wo bei der Arbeit systematische Fehler auftreten können und diese Störquellen dann möglichst ausschalten.

### 2.5.3 Grobe Fehler

Grobe Fehler sind von systematischen Fehlern u.U. schwer abzugrenzen. Sie entstehen durch grundsätzlich fehlerhaftes Arbeiten.

#### Beispiel 29

Bei der Untersuchung des Einflusses von **Erythropoetin** auf die Erythrozytenzahl/L Blut wurden Probanden mit unterschiedlichen Mengen Erythropoetin behandelt. Vor der Blutanalyse wurden die Proben von Proband C und Proband D verwechselt. Dies führte, bevor wir den Fehler später nachweisen konnten, zunächst zu fragwürdig scheinenden Ergebnissen.

#### Gründe für grobe Fehler

Fehler in der Zuordnung von Proben  
Benutzen falschen Auswertetabellen  
Benutzen einer falschen Analysenmethode  
Abschreibfehler (Originaldaten → Sekundärlisten)

Grobe Fehler sind ein Hinweis auf schlechte handwerkliche Arbeit.

### 2.5.4 Erkennung von Fehlern

Das Erkennen von Fehlern ist von großer Bedeutung. Erkennt man sie, so können u.U. teure Untersuchungen, bei denen gleich zu Beginn Fehler gemacht wurden, rechtzeitig abgebrochen werden. Eine Langzeittoxizitätsstudie zu einem medizinisch interessanten Wirkstoff an Ratten dauert ca. zwei Jahre. Fehler bei der Stichprobenauswahl, die nicht erkannt werden, binden über zwei Jahre Arbeitskräfte und das Ergebnis ist wertlos. Fehler bei klinisch chemischen Analysen können zu gravierenden gesundheitlichen Problemen führen.

#### Grobe Fehler und systematische Fehler

Prüfung auf Plausibilität (plausibilis – beifallwürdig)

Jeder, der sich mit der Datengewinnung beschäftigt, sollte genügend Fachkenntnisse auf dem Gebiet zu dem die Daten gehören haben, um erkennen zu können, ob die gewonnenen Daten aus sachlicher Sicht als überhaupt möglich, einleuchtend, annehmbar, vernünftig d.h. plausibel erscheinen.

#### Beispiel 30

Bei der Suche nach einem Wirkstoff, der bei **Mäusen** eine Infektion mit *Trypanosoma brucei* erfolgreich therapieren kann, wurde eine Kontrollgruppe infizierter Tiere mit einem Placebo behandelt. Bei dieser Kontrollgruppe überlebten alle Tiere die nächsten sieben Tage. Das war nicht zu erwarten, da die Infektion erfahrungsgemäß nach vier Tagen tödlich verläuft. Dieses unerwartete Überleben der Tiere war uns aufgefallen. Eine Recherche ergab, dass die Tiere dieser Gruppe versehentlich gar nicht infiziert worden waren. So wurde durch eine Plausibilitätsprüfung der Ergebnisse ein grober Fehler aufgedeckt. Hätte der der die Daten zu beurteilen hatte, keine Kenntnis über den Infektionsverlauf gehabt, dann wäre der Fehler wahrscheinlich nicht aufgefallen.

Er kommt u.U. vor, dass in einer Messwertreihe ein einziger Wert weit weg vom Feld der anderen liegt. Wenn wir sicher sind, dass dieser Wert durch einen offensichtlichen Fehler entstanden ist, dann können wir ihn als „Ausreißer“ von der weiteren Bearbeitung ausschließen. Dies muss dann aber protokolliert werden. Aus der Urliste dürfen wir den Ausreißer nicht entfernen. Wenn nur der Verdacht auf einen Fehler vorliegt, dann können wir mit einem Ausreissertest (z.B. dem NalimovTest, mit dem wir uns später beschäftigen werden.) prüfen, ob es sich wirklich um einen Ausreißer handelt oder nicht. Kann der Test, die Ausreisservermutung nicht bestätigen, dann muss der Wert bei den folgenden Berechnungen mit einbezogen werden.

### Zufallsfehler und systematische Fehler überlagern sich oft.

Die Größe zufälliger Fehler kann durch Berechnung der Streuung (machen wir später) mit anderen Datenreihen verglichen werden. Ist das Streuungsmaß größer als „üblich“ (das kann man mit einem Test feststellen, siehe später), dann sollten wir aufhorchen.

Systematische Fehler sind nicht so leicht aufzudecken, da ja der Sollwert in der Regel nicht bekannt ist. Werden systematische Fehler vermutet, so können wir das u.U. prüfen, indem wir die Analyse mit einem anderen Verfahren wiederholen.

**Ziel:** Grobe und systematische Fehler ausschalten!  
Zufallsfehler minimieren!

## 2.6 Biologische Variabilität

### Beispiel 31

Ich habe bei einer **Kaliumhydroxidlösung** mit  $w(\text{KOH}) = 7,00 \text{ g}/100 \text{ g}$  an fünf aufeinanderfolgenden Tagen titrimetrisch  $w$  bestimmt und erhielt folgende Werte in  $\text{g}/100\text{g}$ : 6,90; 6,99; 7,05; 7,01; 7,06. Die Streuung um den Mittelwert 7,00 (genau 7,002) betrachte ich als auf dem Zufallsfehler beruhend. Einen systematischen Fehler schließe ich aus, da mein Mittelwert ja dem Sollwert entspricht.

Dann habe ich an fünf aufeinanderfolgenden Tagen bei **einem Kaninchen** Blut entnommen und jeweils die Leukozytenzahl bestimmt. Dabei erhielt ich die folgenden Werte: 6,8; 7,0; 7,1; 7,3; 7,6 \*  $10^9/\text{Leucos}/\text{L}$ . Für die Streuung mache ich auch hier die den Zufallsfehler verantwortlich. Da mein Mittelwert mit  $7,16 * 10^9/\text{Leucos}/\text{L}$  dem Erwartungen eines gesunden Kaninchens entspricht, schließe ich auch hier einen systematischen Fehler aus.

Nun eine dritte Messgruppe. Ich untersuche bei **fünf Kaninchen** an einem Tage die Zahl der Leukozyten/L und erhalte folgende Ergebnisse in  $\text{Leucos}/\text{L}$ : 5,6; 6,4; 7,3; 8,0; 8,3 \*  $10^9/\text{L}$ . Der Mittelwert beträgt  $7,12 * 10^9/\text{Leucos}/\text{L}$  und repräsentiert den Erwartungswert gut. Also liegt auch hier ein Zufallsfehler aber kein systematischer Fehler vor.

Die Unterschiede in den Streuungen bei den drei Versuchen mögen zunächst nicht auffallen, sie erscheinen aber deutlich, wenn die Daten an einer Skala dargestellt werden.

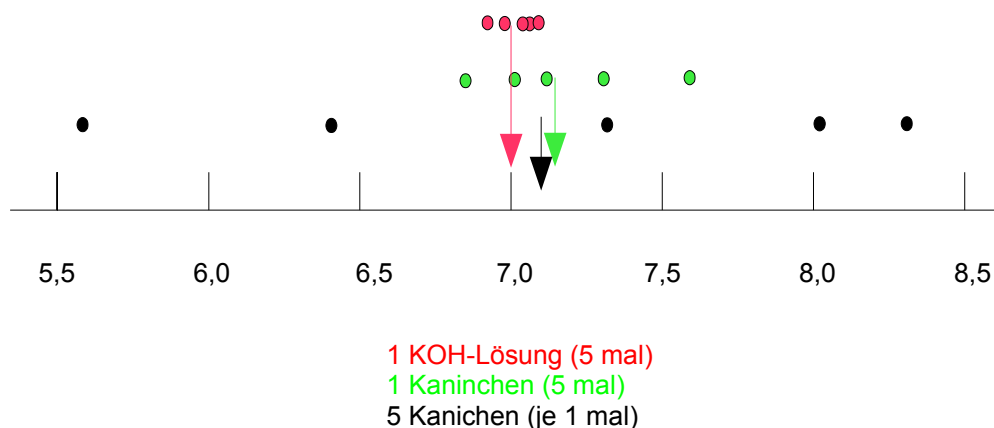


Abb.4

Da wir systematische Fehler (und auch grobe Fehler) ausgeschlossen haben, stellt sich die Frage nach der Ursache für die unterschiedlich starken Streuungen.

Die verschlossene KOH-Lösung hat sicher an allen fünf Tagen den gleichen Wert für  $w$ . Bei einem Tier ist das anders. Bei unserem Kaninchen variieren die Zahlen der Leukozyten/L physiologisch von Tag zu Tag. Das ist normal. Untersuchen wir mehrere Kaninchen, so sind die Schwankungen von Tier zu Tier noch größer als bei einem Kaninchen von Tag zu Tag. Das ist auch normal. Im Gegensatz zur



KOH-Lösung habe wir es bei den Kaninchen mit Lebewesen zu tun. Und hier ist es physiologisch, d.h. natürlich, dass die Blutwerte wie viele andere Werte zu verschiedenen Zeiten mehr oder weniger großen Schwankungen unterliegen. Diese Erscheinung, die auf der Individualität von Lebewesen beruht, nennen wir **Biologische Variabilität**. Diese Eigenschaft, die zu den üblichen, die Messwerte/Zählwerte beeinflussenden Fehler hinzukommt, macht die Interpretation biologischer Versuchsergebnisse problematischer als die Ergebnisse chemischer und physikalischer Versuche.

Wie aus dem Beispiel hervorging müssen wir unterscheiden zwischen:

<u>Intraindividueller Variabilität</u> ,	die zu unterschiedlichen Werten eines Individuums in Abhängigkeit von z.B. der Tageszeit führt.
und	
<u>Interindividueller Variabilität</u> ,	die zu unterschiedlichen Werten bei verschiedenen Individuen führt.

Wir könnten nun denken, dass wir die Stärke der interindividuellen Variabilität durch Vergrößern der Stichprobe minimieren können. Das ist aber nicht der Fall. Wenn zusätzliche Elemente der Grundgesamtheit in die Stichprobe aufgenommen werden, so haben diese Elemente die übliche Bandbreite an Ausprägungen wie in der Grundgesamtheit. Eine Minimierung der interindividuellen Variabilität erreichen wir dadurch, dass wir durch Wahl der Grundgesamtheit dafür sorgen, dass diese in Bezug auf das zu untersuchende Merkmal eine geringe Streuung hat. Das bedeutet bei Tieren etwa, dass man sich auf ein Geschlecht beschränkt, auf eine bestimmte Alters- und Gewichtsklasse oder auf die Wahl eines Stammes, der genetisch bedingt eine geringe Streuung der Merkmalsausprägung hat.

## 2.7 Übungen

### Übung 1

1. Von welchem Fehler wird die Richtigkeit der Daten beeinflusst und was ist das Typische für diese Fehlerart?
2. Von welchem Fehler wird die Präzision der Daten beeinflusst und was ist das Typische für diese Fehlerart?
3. Welchen Fehler kann man besser erkennen und warum, den systematisch Fehler oder den Zufallsfehler ?
4. Erklären den Begriff Plausibilität.
5. Welche Auswirkung hat die biologische Variabilität auf die Auswertung von Versuchen?

Antwort 1:

Systematischer Fehler, die Werte liegen, bezogen auf den Sollwert immer zu hoch oder zu niedrig.

Antwort 2:

Zufallsfehler, die Werte streuen mehr oder weniger stark in der Nähe des Sollwertes, wenn kein systematischer oder grober Fehler vorliegt.

Antwort 3:

Den systematischen Fehler, da die Werte vom Erwartungswert deutlich abweichen.

Antwort 4:

Das bedeutet soviel wie Stimmigkeit eines Wertes. Der Wert erscheint uns einleuchtend.

Antwort 5:

Durch die Biologische Variabilität werden die Streuungen, die Messwerte an sich schon anhaften, noch vergrößert. Das macht die Auswertung von Versuchen oft schwieriger.

### Übung 2

- a. Erstellen Sie in Excel eine **Zufallszahlentabelle** mit mindestens 300 Ziffern (0 – 9) und suchen Sie aus 100 Tieren 25 für eine Stichprobe aus.
- b. Erstellen Sie eine ebenso große Tabelle mit Zahlen, die Ihnen gerade einfallen. Prüfen Sie in beiden Tabellen durch Abzählen, wie häufig die Ziffern 0 bis 9 vorkommen.  
Keine Lösung



### Übung 3

Erklären Sie an Beispielen was eine **retrospektive Erhebung** von einem **Experiment** unterscheidet.

Antwort:

Eine retrospektive Erhebung ist ein Verfahren zur Datengewinnung bei der auf Daten zugegriffen wird, deren Entstehung und Aufzeichnung in der Vergangenheit liegt. Beispiel: Ein Meteorologe muss eine Zusammenstellung schreiben über den Temperaturverlauf auf dem Feldberg/Schwarzwald. Ihn interessieren die täglichen Temperaturwerte von 12:00 bis 15:00 in den Jahren 1962 bis 1965.

Ein Experiment ist ein Verfahren zur Datengewinnung, bei dem die Bedingungen im Umfeld vom Experimentator mehr oder weniger frei gewählt werden können. Beispiel: Ich möchte den Sauerstoffverbrauch von Fischen (Goldorfen) in Abhängigkeit von der Fließgeschwindigkeit des Wassers ( $v$ ), der Temperatur des Wassers ( $T$ ) und vom Gewicht ( $m$ ) der Tiere ermitteln. Dazu bilde ich mehrere Gruppen von Fischen und variiere die Bedingungen wie in der Tabelle angegeben.

Gruppe	V in mL/min	T in °C	M in g
A	50	18	20
B	50	18	25
C	50	24	20
D	50	24	25
E	30	18	20
F	30	18	25
G	30	24	20
H	30	24	25

Die beiden Verfahren zur Datengewinnung unterscheiden sich darin, dass ich bei den Erhebungen darauf angewiesen bin, dass die Aufzeichnungen von vor ca. 40 Jahren ordentlich zu den gewünschten Zeitpunkten durchgeführt wurden. Über Sorgfalt und Richtigkeit der Daten liegen keine weiteren Angaben vor. Insofern sind diese Daten mit einer Unsicherheit belastet. Im Experiment kann ich die gewünschten Bedingungen wie ich möchte einstellen und auf die korrekte Protokollierung der Daten habe ich Einfluss, da ich sie selber vornehme. Die Daten des Experimentes sind also sicherer und somit aussagekräftiger als die der Erhebung.

### Übung 4 Fehler bei der Datengewinnung

Es wurde die Hydrolyse des Fettes in der Milch durch das Enzym Lipase untersucht. Gemessen wurden über 6 h die Veränderungen der pH-Werte der Proben. Kolben 1 und 2 enthielten Milch und das aktive Enzym, Kolben 3 enthielt Milch und inaktiviertes Enzym. Die Tabelle zeigt die Messwerte (pH-Werte).

	0-Wert	1 h	2 h	3 h	4 h	5 h	6 h
Kolben 1	6,45	6,35	6,23	6,34	6,29	6,25	6,25
Kolben 2	6,62	6,62	6,52	6,68	6,65	6,64	6,64
Kolben 3	6,63	6,63	6,54	6,70	6,67	6,65	6,65

Welcher Versuchsfehler ist zu erkennen?

Lösung:

An einer Graphik erkennt man schnell, dass Kolben 2 inaktives Enzym enthielt. Es liegt ein grober Fehler vor.

### Übung 5 Fehler bei der Datengewinnung

Untersucht wurden die Erythrozyten-Zahlen und die Leukozyten-Verteilung bei **Mäusen**, die alle gleich stark mit *Trypanosoma brucei* infiziert waren. Jeder Untersucher untersuchte 2 Mäuse. Die Tiere wurden den Untersuchern durch Randomisieren zugeteilt.

Methode: Zählkammer, Blutausschlag, Mikroskop  
Die Ergebnisse waren

Untersucher Nr.	Maus Nr.	Erys/ $\mu\text{L} \cdot 10^9$	Neutr.%	Eos.%	Lymph.%	Mono.%
1	1	8,56	9,0	0	89,5	1,5
	2	8,52	9,0	0,5	89,5	1,0
2	1	8,24	23,0	0,5	71,5	5,0
	2	9,04	18,5	1,5	79,5	4,5
3	1	10,90	22,5	11,0	55,5	11,0
	2	10,20	23,0	11,0	60,5	5,5
4	1	3,89	64,0	1,0	21,0	14,0
	2	4,60	62,0	3,0	29,0	6,0
5	1	9,68	60,0	5,0	24,0	11,0
	2	6,29	58,0	6,0	27,0	9,0
6	1	8,44	29,0	2,0	65,0	4,0
	2	7,52	35,0	0	59,0	6,0
7	1	1,54	6,4	7,9	78,8	6,9
	2	1,31	5,0	9,0	81,0	5,5
8	1	8,92	34,0	2,0	58,5	5,5
	2	6,91	30,5	2,0	61,0	6,5
9	1	10,24	35,0	1,0	60,0	4,0
	2	9,01	34,0	1,5	9,5	5,0
10	1	10,60	21,0	6,0	70,0	3,0
	2	10,51	18,0	4,0	74,0	4,0

- Was ist an den Ergebnissen auffällig?
- Wo vermuten Sie welche Fehler?
- Begründen Sie Ihre Antworten.

#### Lösungen

Untersucher 1: Neutrophile und Monozyten beide zu niedrig und damit nicht plausibel, systematischer Fehler

Untersucher 3: Eosinophile beide zu hoch, nicht plausibel, systematischer Fehler Monozyten 1 zu hoch, vermutlich grober Fehler, bei systematischem Fehler wäre eher beide Werte zu hoch.

Untersucher 4: Monozyten 1 zu hoch, vermutlich grober Fehler, bei systematischem Fehler wären eher beide Werte zu hoch.

Untersucher 7: Erys beide zu tief, systematischer Fehler

Untersucher 9: Lymphocyten 2 zu tief, vermutlich grober Fehler, bei systematischem Fehler wären eher beide Werte zu tief.