

## 8. Planung und Auswertung von Tierversuchen

### 8.1. Wissenschaftlicher Zweck

Tierversuche werden durchgeführt, um biologische oder biomedizinische Fragestellungen zu beantworten. Hierbei muss die wissenschaftliche Fragestellung vor Versuchsbeginn eindeutig festgelegt sein. Es ist unwissenschaftlich und zudem tierschutzrechtlich unzulässig, Tierversuche ohne präzise Fragestellung aufzunehmen in der vagen Hoffnung, „verwertbare“ Ergebnisse zu erhalten. Der Weg kann nicht das Ziel sein. Es obliegt primär dem Leiter tierexperimenteller Studien, durch intensive wissenschaftliche Recherchen (Medline, Kongressbesuche, Kommunikation mit anderen Arbeitsgruppen) die Relevanz einer tierexperimentellen Fragestellung zu gewährleisten. Aber auch die Tierschutzbeauftragten, die zu jedem Versuchstierantrag eine Stellungnahme abgeben müssen, die Mitglieder der §15 Kommissionen, von denen nach Deutschem Tierschutzgesetz 1/3 aus anerkannter Tierschutzorganisationen rekrutiert werden müssen und die die Behörden bei der Genehmigung von Tierversuchsanträgen beraten, sowie letztendlich auch die entscheidenden Behördenvertreter sind in die tierschutzrechtliche Pflicht genommen, die Relevanz der jeweiligen wissenschaftlichen Fragestellung kritisch zu hinterfragen.

### 8.2. Eigenschaften tierexperimenteller Studien

In der Humanmedizin durchgeführte Studien werden nach folgenden Kriterien eingeteilt:

-Beobachtung oder Experiment:

-In Beobachtungsstudien solche Merkmale von Individuen erhoben, die sowieso vorhanden sind, also nicht erst durch gezielte Manipulation induziert werden müssen. Der Wissenschaftler greift bei Beobachtungsstudien nicht aktiv in das Geschehen ein sondern spielt eine völlig passive Rolle. Die meisten epidemiologischen Studien stellen Beobachtungsstudien dar (z.B. die Erfassung der Körpergewichte und der Cholesterinspiegel bei Diabetikern). Die Ergebnisse von Beobachtungsstudien sind beschreibend, Rückschlüsse auf kausale (ursächliche) Zusammenhänge sind in der Regel nicht möglich.

-In einem Experiment werden solche Merkmale bei Individuen erfasst, denen im Rahmen der Studie bestimmte Charakteristika zugewiesen wurden. In einem Experiment spielt der Wissenschaftler eine aktive Rolle, er führt aktiv Manipulationen an Individuen durch und versucht dadurch Einfluss auf Merkmale zu nehmen. Klinische Studien, bei denen an freiwilligen Probanden oder Patienten neue Therapieformen erprobt werden, stellen Experimentalstudien dar.

-Prospektive oder retrospektive Studie:

- Bei retrospektiven Studien werden Daten erfasst und ausgewertet, die in der Vergangenheit entstanden sind und somit ausnahmslos bereits zu Beginn der Studie vorliegen. Somit können bei retrospektiven Studien nur diejenigen Daten bearbeitet werden, die auch vorhanden sind. Interessierende Parameter, die eben nicht erfasst wurden, können nicht in die Studie einfließen. Ein Beispiel für eine retrospektive Studie ist die Untersuchung der Eigenschaften von Tumoren, die mit dem Einverständnis der Patienten nach der chirurgischen Entfernung aufbewahrt wurden.
- Bei prospektiven Studien werden die Merkmale erst nach Beginn der Studie erfasst. Bei prospektiven Studien kann der Studienleiter im Rahmen der Kooperationsbereitschaft der Studienteilnehmer im voraus planen, welche Daten zu gewinnen und zu bearbeiten sind. Alle klinischen Studien zur Prüfung neuer Therapien bei Probanden / Patienten stellen prospektive Studien dar.
- Longitudinal- oder Transversalstudie (Längsschnitt oder Querschnitt)
  - Bei Transversal- oder Querschnittstudien werden bei den Studienteilnehmern nur zu einem einzigen Zeitpunkt Merkmale erfasst. Ein Beispiel für eine Transversalstudie stellt z.B. die einmalige Erhebung des Body-Mass-Index bei den Mitarbeitern und den Studenten einer Universität dar.
  - Bei Longitudinal- oder Längsstudien werden bei den Studienteilnehmern zu mehreren Zeitpunkten Merkmale erfasst. Es findet also eine Beobachtung der Merkmale statt.
  - Bei sogenannten Pseudolongitudinalstudien werden bei den Studienteilnehmern zwar wie bei Transversalstudien nur jeweils zu einem Zeitpunkt Merkmale erhoben. Da die Merkmale jedoch bei Individuen unterschiedlicher Altersklassen bestimmt werden, kann durch die altersspezifische Datenauswertung trotzdem eine Aussage über die zeitliche Entwicklung des Merkmals getroffen werden. Werden beispielsweise die Cholesterinwerte einmalig bei allen Bürgern der Stadt Mainz bestimmt und anschließend die Mittelwerte jeweils für alle Bürger mit demselben Geburtsjahr errechnet, so kann durch diese pseudolongitudinale Studie durchaus eine Aussage über die Altersabhängigkeit des Cholesterinspiegels getroffen werden.

### **Merke:**

-Tierversuche sind immer **Experimentalstudien**, in denen der Wissenschaftler aktiv Einfluss auf die Versuchstiere nicht (z.B. Applikation von Substanzen, Diätverabreichung, Vergleich gentechnisch veränderter Tiere mit Kontrolltieren). Somit erlauben Tierversuche die Klärung kausaler wissenschaftlicher Zusammenhänge.

-Tierversuche sind immer **prospektive Studien**, d.h. die Studie kann und muss detailliert in voraus geplant werden.

-Tierversuche können **Transversalstudien, Longitudinalstudien oder Pseudolongitudinalstudien** sein. So werden bei vielen Tierversuchen der Grundlagenforschung Tiere getötet und dann bestimmte Organe entnommen und histologisch untersucht. Solche Experimente stellen prinzipiell Transversalstudien dar. Da häufig eine Altersabhängigkeit der histologischen Veränderungen dargestellt werden soll, werden Tiere unterschiedlicher Altersgruppen histologisch untersucht. Hierdurch ergeben sich typische Pseudolongitudinalstudien. Tierexperimentelle Longitudinalstudien, bei denen bestimmte Merkmale (z.B. Gewicht, Blutwerte) zu mehreren Zeitpunkten bei den Versuchstieren erhoben werden, werden ebenfalls häufig realisiert.

### **8.3. Planung tierexperimenteller Studien**

#### **8.3.1 Auswahl des Tiermodells (Tierart, Tierstamm, gentechnische Veränderungen, Geschlecht, Alter)**

Auch die bestmögliche Auswahl eines geeigneten Tiermodells obliegt in erster Linie dem verantwortlichen Tierversuchsleiter. Die Modellauswahl impliziert nicht nur die Auswahl der Tierart, sondern auch die des Tierstamms sowie des Geschlechts und des Alters der Versuchstiere. Darüber hinaus muss der Versuchsleiter auch entscheiden, ob der Einsatz gentechnisch veränderter Tiere sinnvoll ist. In diesem Zusammenhang ist anzumerken, dass in den letzten Jahrzehnten eine breite Vielzahl unterschiedlicher gentechnisch veränderter Mausstämme entwickelt wurden, die nun von der wissenschaftlichen Gemeinde genutzt werden können. Die Auswahl des Tiermodells muss wissenschaftlichen, ethisch / tierschutzrechtlichen und praktischen Kriterien genügen. In erster Linie steht die fachliche Prüfung des Tiermodells, d.h. die Auswahl desjenigen Modells, das die besten Erfolgsaussichten bietet, eine spezifische wissenschaftliche Fragestellung auch beantworten zu können. Bei der Modellauswahl ist aber auch zu berücksichtigen, dass den tierschutzrechtlichen Forderungen der Durchführung von Tierversuchen bei phylogenetisch möglichst niedrig stehenden Spezies sowie der ethischen Vertretbarkeit der Belastung der Tiere im Hinblick auf den zu erwartenden Erkenntnisgewinn genüge getan wird. Schließlich müssen in die Modellauswahl aber auch praktische Erwägungen einfließen (z.B. gesetzliche Registrierungen oder Genehmigungen vorhanden; Arbeitsschutz gewährleistet; erforderliches Hygieneniveau gewährleistet; bauliche, personelle und organisatorische Voraussetzungen gegeben). Obwohl die letztgenannten Kriterien von Seiten der Wissenschaftler häufig als „banal“ abgetan werden, sind sie nicht selten ausschlaggebend für Erfolg oder Misserfolg der Etablierung eines versuchstierkundlichen Modells. Der Tierversuchsleiter sollte bei der Modellauswahl den Rat von Experten (Biologen, Tierhaltungsmanagement) suchen. Er hat durch intensive Recherchen

zu gewährleisten, dass ein für die gegebenen Verhältnisse optimales Tiermodell zur Auswahl kommt.

### **8.3.2 Definition von Abbruchkriterien**

Bereits vor Versuchsbeginn sind diejenigen Kriterien festzulegen, bei denen belastete Tieren zu töten sind. Für spezifische Versuchszwecke existieren zum Teil bewährte Empfehlungen (z.B. „Kriterien zur vorzeitigen Tötung von tumortragenden Mäusen und Ratten im Versuch“, Erstellung durch den „Ausschuss für Tierschutzbeauftragte in der GV-SOLAS“ sowie den „Arbeitskreis 4 der Tierärztlichen Vereinigung für Tierschutz – TVT“). In diesem Zusammenhang ist auch darauf hinzuweisen, dass in §9 (2) 6 des Deutschen Tierschutzgesetzes gefordert wird, dass „bei Tierversuchen zur Ermittlung der tödlichen Dosis oder tödlichen Konzentration eines Stoffes das Tier schmerzlos zu töten ist, sobald erkennbar ist, dass es infolge der Wirkung des Stoffes stirbt.“ Dies bedeutet, dass bei Experimenten, bei denen der Tod der Versuchstiere ein Messparameter ist, eine enge Gesundheitsüberwachung zu organisieren ist, so dass die frühzeitige Erkennung und unverzügliche schmerzfreie Tötung moribunder Tiere gewährleistet werden kann.

### **8.3.3 Bestimmung der Gruppengröße**

Für die Bestimmung der Gruppengröße eines Tierexperiments stehen folgende Methoden zur Verfügung:

- Gesetzliche Vorschriften: Für bestimmte staatlich vorgeschriebene Tierversuche wie z.B. Toxizitätstests ist die Anzahl einzusetzender Tiere vom Gesetzgeber vorgegeben.
- Empirie: Die Gruppengröße wird aufgrund eigener Erfahrungswerte oder anhand von Literaturrecherchen (Erfahrungswerte anderer Wissenschaftler) festgelegt.
- Mathematische Gleichungen: Hierbei kann aufgrund bestimmter vorher festzulegender oder abzuschätzender Parameter (Art des zur Datenanalyse anzuwendenden statistischen Tests, gewünschtes Signifikanzniveau des statistischen Tests, gewünschte Trennschärfe des Tests, Standardabweichung der Messwerte) die erforderliche Gruppengröße errechnet werden. Hierzu ist eine gewisse Expertise erforderlich, über die wohl nur wenige Leiter tierexperimenteller Studien verfügen. Bei Bedarf sollte der Rat von Experten auf dem Gebiet der Biometrie gesucht werden.
- Ressource-Gleichungsmethode: Mead (1988) hat ein vereinfachtes rechnerisches Verfahren zur Abschätzung der Gruppengröße entwickelt. An dieser Stelle kann nicht weiter auf dieses Verfahren eingegangen werden. Es wird auf die Originalpublikation verwiesen (Mead, 1988).

### 8.3.4 Gewährleistung der Strukturgleichheit

**Randomisierung:** In vielen Tierexperimenten wird eine Prüftiergruppe mit einer Kontrolltiergruppe verglichen. Bei der Planung solcher Versuche ist es wichtig, die Tiere auf völlig zufällige Art und Weise auf Test- und Kontrollgruppe zu verteilen. Dieser Vorgang, der als Randomisierung bezeichnet wird, kann am besten durch Zufallszahlentabellen oder durch die Verwendung von geeigneten computergestützten Zufallszahlengeneratoren erreicht werden. Es hat sich herausgestellt, dass die interindividuelle Variation durch die Randomisierung auf bestmögliche Weise auf Tiergruppen verteilt wird. Es sei jedoch angemerkt, dass die Randomisierung eben ein zufälliger Vorgang ist, d.h. in Einzelfällen auch extreme Imbalancen (z.B. bei der Gruppengröße) schaffen kann. In solchen Fällen ist es legitim, die randomisierte Tiergruppenbildung zu wiederholen.

**Stratifizierte Randomisierung:** Wird nicht eine gesamte Tiergruppe in Prüf- und Kontrollgruppe randomisiert, sondern wird die Gruppe zunächst nach bestimmten Kriterien, insbesondere solcher Kriterien, bei denen ein Studieneinfluss vermutet wird, unterteilt, so wird dies als stratifizierte Randomisierung oder Blockbildung bezeichnet. Die Blockbildung begründet sich dabei auf offensichtliche Merkmale der Versuchstiere (z. B. starke Unterschiede bezüglich des Körpergewichts oder bezüglich der Herkunft der Tiere) oder auf die Art, wie die Daten erhoben werden (z.B. unterschiedliche Zeitpunkte der Probengewinnung).

**Cross-over Versuche:** Es gibt Versuchsansätze, bei denen die Tiere nacheinander unterschiedlichen Behandlungen ausgesetzt werden. Für das Beispiel eines Tierversuchs mit einer Prüf- und einer Kontrollgruppe ergibt sich beim Cross-over Versuch folgendes experimentelles Design. Zunächst werden die Tiere in beide Gruppen (Prüfung und Kontrolle) randomisiert und die erste Phase des Versuchs wird komplett durchgeführt. Nach Abschluss der ersten Experimentalphase und Einhaltung einer adäquaten Karenzzeit (Erholung, „wash-out“ der Testsubstanz) wird die Gruppenzugehörigkeit gerade umgekehrt, d.h. diejenigen Tiere, die vorher zur Kontrollgruppe gehörten, werden nunmehr der Prüfgruppe zugeordnet und umgekehrt. Nach dem Tausch der Gruppenzugehörigkeit wird der Versuch mit veränderten Rollen der Versuchstiere wiederholt.

**Faktorielle Tierversuchspläne:** Bei faktoriellen Versuchsplanungen wird im Rahmen des Tierversuchs nicht nur ein Faktor untersucht (wie z.B. Prüfbehandlung versus Kontrollbehandlung) sondern es werden gleich mehrere Faktoren abgearbeitet. So können beispielweise mehrere Prüfsubstanzen mit einer Kontrollsubstanz verglichen werden oder Prüf- und Kontrollbehandlung können bei Männchen und Weibchen oder bei unterschiedlichen Tiergenotypen getestet werden.

### 8.3.5 Gewährleistung der Beobachtungsgleichheit

Es hat sich in der Vergangenheit gezeigt, dass bei der Durchführung humanmedizinischer klinischer Studien sowohl Probanden bzw. Patienten (Studienteilnehmer) als auch die durchführenden Ärzte (Studienleiter) Erwartungshaltungen entwickeln können. Zudem ist es durch zahlreiche Beispiele belegt, dass als psychologische Folge von Erwartungshaltungen Messergebnisse oder deren Interpretation beeinflusst werden können. So besteht beispielsweise im Rahmen einer Medikamentenstudie, bei der das allgemeine Wohlbefinden eines Probanden/Patienten beurteilt werden soll, bei beiden Parteien ein beträchtlicher Ermessungsspielraum der Beurteilung des Merkmals „Wohlbefinden“. Selbstverständlich muss aber bei einer klinischen Studie die gleichwertige Beurteilung von Behandlungs- und Kontrollgruppe gewährleistet werden können (Beobachtungsgleichheit). Der möglichen Beeinflussung humanmedizinischer klinischer Studien durch Erwartungshaltungen der beteiligten Parteien wird durch doppelblinde Studiendurchführung entgegen gewirkt. Beim Vergleich einer neuen Therapieform mit einer bereits bewährten wissen also weder Proband bzw. Patient noch behandelnder Arzt, welche Therapieform beim jeweiligen Studienteilnehmer zur Anwendung kommt. Auch die Auswertung der Testergebnisse erfolgt blind, d.h. die Zuordnung zur Gruppenzugehörigkeit ist unbekannt.

Bei Tierexperimenten braucht nicht befürchtet werden, dass die Versuchstiere Erwartungshaltungen entwickeln und dies die Studie beeinflusst. Sehr wohl besteht jedoch die Möglichkeit, dass der Experimentator bei der experimentellen Auswertung keinen neutralen Standpunkt einnimmt, sondern aufgrund einer Erwartungshaltung befangen ist. Folgende Maßnahmen sind geeignet, die Beobachtungsgleichheit tierexperimenteller Studien zu gewährleisten:

- Blinde Auswertung: Der Experimentator ist bei der Auswertung der Versuchsdaten nicht informiert, welcher Versuchsgruppe das auszuwertende Material zugehört. Dies ist insbesondere bei Auswertungsmethoden, die einen großen Ermessensspielraum bieten (z.B. histologische Untersuchungen, morphometrische Analysen), essentiell.
- Festlegung der Interpretationskriterien von Messwerten vor Studienbeginn: Auch bei der Verwendung von Laborwerten als Zielparame-ter können Probleme bei der Beobachtungsgleichheit auftauchen; zwar nicht bei der Bestimmung der Werte selbst (dies übernimmt ein Automat), sehr wohl aber bei der Interpretation der Messergebnisse. Es ist wichtig, dass bereits bei der Planung des Tierversuchs, also vor Studienbeginn, eindeutig festgelegt wird, wie die zu erhebenden Messwerte zu interpretieren sind, d.h. ab welchen Grenzwerten die Messwerte als normal bzw. als erhöht oder als erniedrigt anzusehen sind.

### **8.3.6 Festlegung des Zielkriteriums**

Bei der Planung einer tierexperimentellen Studie ist neben der Gewährleistung der Struktur- und Beobachtungsgleichheit die eindeutige Festlegung des Zielkriteriums wichtig. Soll beispielsweise der Einfluss einer Testsubstanz auf den Cholesterinspiegel von Tieren untersucht werden, so kann der Cholesterinwert einer einzigen Messung oder bei mehreren Bestimmungen der Verlauf des Cholesterinspiegels oder aber auch die Differenz zwischen zuerst und zuletzt erhobenem Wert als Zielkriterium dienen. Es ist sinnvoll, das Zielkriterium vor Studienbeginn zu definieren. Dabei sollten vorzugsweise solche Zielkriterien ausgewählt werden, bei denen eine geringe Standardabweichung erwartet werden kann.

## **8.4. Auswertung tierexperimenteller Studien:**

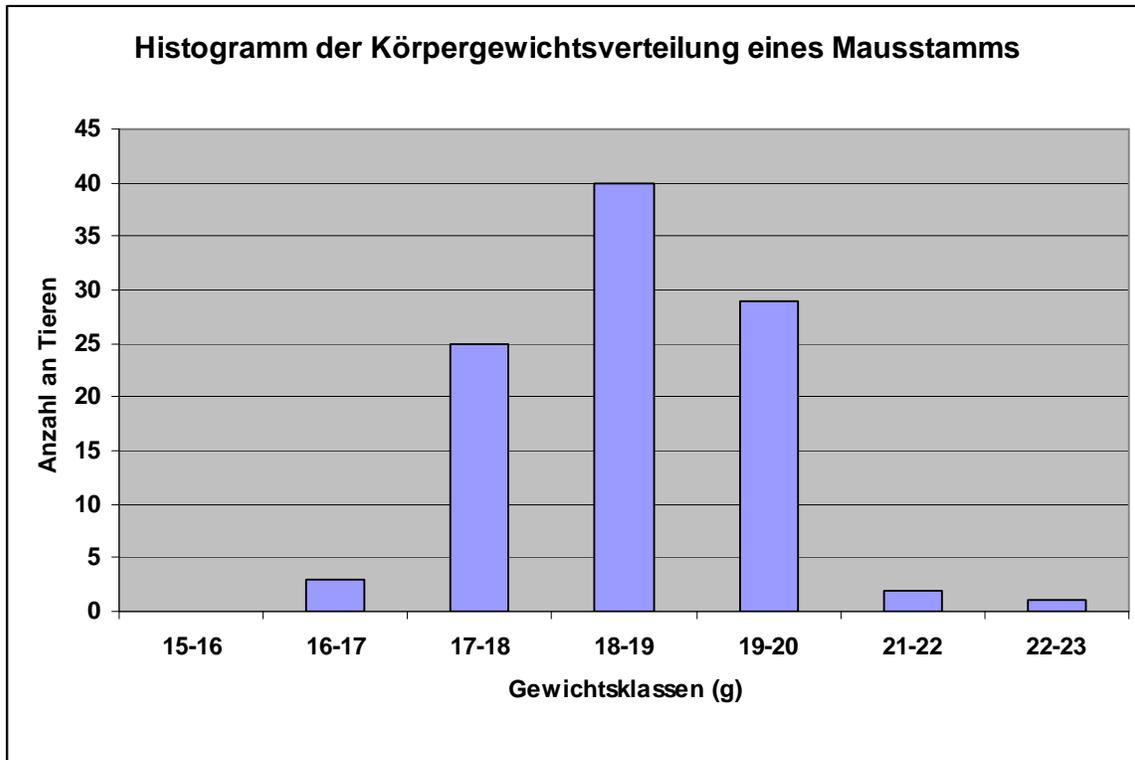
### **8.4.1 Typen von Merkmalen**

Qualitative Merkmale sind Merkmale, deren Ausprägung keine zahlenmäßige Ordnung haben (wie z.B. Augenfarbe, Geschlecht oder Genotyp).

Quantitative Merkmale hingegen können durch Zahlen ausgedrückt werden. Hierbei sind diskrete und stetige Merkmale zu unterscheiden. Bei diskreten Merkmalen handelt es sich um Anzahlen: sie werden durch natürliche Zahlen wiedergegeben (z.B. Anzahl an Jungtieren eines Weibchens, Anzahl an Leukozyten). Stetige Merkmale liegen auf einer kontinuierlichen Skala: sie werden also durch die reellen Zahlen wiedergegeben (z.B. Größe, Gewicht, Alter, Triglyceridspiegel).

### **8.4.2 Normalverteilung**

Für die Darstellung der Ergebnisse tierexperimenteller Studien und auch für deren statistische Analyse ist es ganz entscheidend, ob die erhobenen Ergebnisdaten einer Normalverteilung entsprechen. Die Beurteilung, ob eine spezifische Datensammlung eine Normalverteilung darstellt, ist zumeist schwierig. Hilfreich ist die Erstellung eines Daten-Histogramms. Hierbei werden die Ergebnisdaten einer Versuchsgruppe in Klassen zusammengefasst, wobei die Anzahl der Klassen sinnvoll zu wählen ist. Im Daten-Histogramm werden nun die Häufigkeiten dargestellt, mit der Daten in eine bestimmte Klasse fallen. Die nachfolgende Abbildung zeigt das Histogramm für die Körpergewichte von 100 Inzuchtmäusen eines Alters von 6-8 Wochen.



**Abbildung 8.1: Histogramm der Körpergewichtsklassen von 100 Inzucht-Mäusen im Alter von 6-8 Wochen**

Bei normalverteilten Datensätzen ist das Histogramm symmetrisch und glockenförmig. Liegt ein asymmetrisches (=schiefes) Histogramm vor, so kann eventuell durch einfaches oder zweifaches Logarithmieren der Daten eine symmetrische Verteilung erreicht werden. In solchen Fällen wird von einer log- oder log-log-Normalverteilung gesprochen. Bei log- und log-log-Normalverteilungen ist zu berücksichtigen, dass die Retransformierung bestimmter statistischer Parameter, die mit den logarithmierten Daten bestimmt werden können, zum Teil unzulässig ist. Es stehen mathematische Tests (z.B. Shapiro-Wilk) zur Verfügung, mit denen für vorgegebene Datensammlungen eine Normalverteilung ausgeschlossen werden kann.

#### **8.4.3. Darstellung stetiger Daten mit Normalverteilung – arithmetischer Mittelwert mit Standardabweichung**

Stetige Daten, bei denen eine Normalverteilung unterstellt werden kann, können am günstigsten durch den arithmetischen Mittelwert dargestellt werden. Der Mittelwert ist die Summe aller Messergebnisse dividiert durch deren Anzahl. Es ist zu beachten, dass der Mittelwert in starker Weise auf Ausreißer reagiert, d. h. extreme Messpunkte werden den Mittelwert sehr stark beeinflussen.

**Rechenbeispiele:**

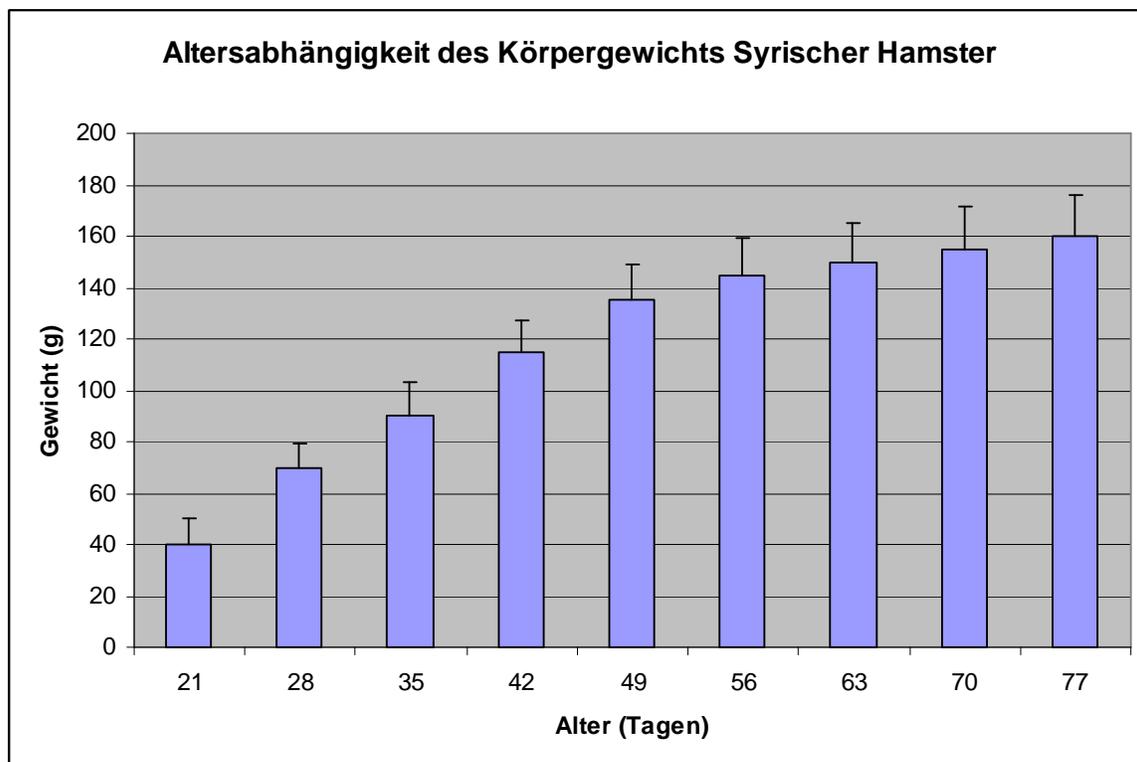
Die Merkmalswerte 3, 7, 70, 75 und 110 weisen den Mittelwert 53,0 auf.

$$53,0 = (3+7+70+75+110)/5$$

Die Merkmalswerte 3, 7, 70, 75 und 10.000 weisen den Mittelwert 2031,0 auf.

Die Beispiele belegen die starke Empfindlichkeit des Mittelwerts gegenüber Ausreißern (hier der Wert 10.000 im 2. Beispiel).

Bei normalverteilten Daten wird die Streuung der Daten durch die Varianz  $s^2$  bzw. deren Wurzel, die Standardabweichung  $s$ , angegeben. Hierbei ist die Varianz definiert als die Summe aller Abstandsquadrate zwischen den Messwerten und dem Mittelwert dividiert durch  $(n-1)$ . Der Platzhalter  $n$  bezeichnet dabei die Anzahl aller Merkmale. In sehr vielen wissenschaftlichen Publikationen wird der Mittelwert von Messdaten als Balkendiagramm dargestellt und die Standardabweichung als Linie aufgesetzt. Im Intervall Mittelwert  $\pm s$  werden sich ca. 68 % aller Merkmale befinden; im Intervall Mittelwert  $\pm 2s$  werden ca. 95 % aller Merkmale vorzufinden sein.



**Abbildung 8.2: Darstellung stetiger Daten als Mittelwert mit Standardabweichung. Die Balkenhöhe zeigt den Mittelwert, der aufgesetzte Strich die Standardabweichung auf.**

#### 8.4.4. Darstellung stetiger Daten ohne Normalverteilung– empirischer Median mit empirischen Quantilen

Stetige Daten, bei denen keine Normalverteilung unterstellt werden kann, sind durch den empirischen Median wiederzugeben. Der empirische Median beschreibt die Mitte eines nach Datengröße sortierten Datensatzes. Der Median unterteilt den Datensatz in 2 Hälften: bei der einen Datenhälfte sind die Daten größer als der Median, bei der anderen Hälfte kleiner. Für einen sortierten Datensatz mit ungerader Datenanzahl ist der empirische Median gleich dem  $[(n+1)/2]$  Wert: dies ist genau derjenige Datenwert, der in der Mitte des sortierten Datensatzes lokalisiert ist. Bei gerader Datenanzahl ist der empirische Median definiert als der Mittelwert des  $n/2$  und  $(n/2+1)$  Wertes. Dies ist der Mittelwert genau der beiden Datenwerte, die in der Mitte des sortierten Datensatzes liegen.

##### Rechenbeispiele:

Die Merkmalswerte 3, 7, 70, 75 und 110 weisen den Median 70 auf.

$$\text{Median} = \text{mittlerer Wert} = 70$$

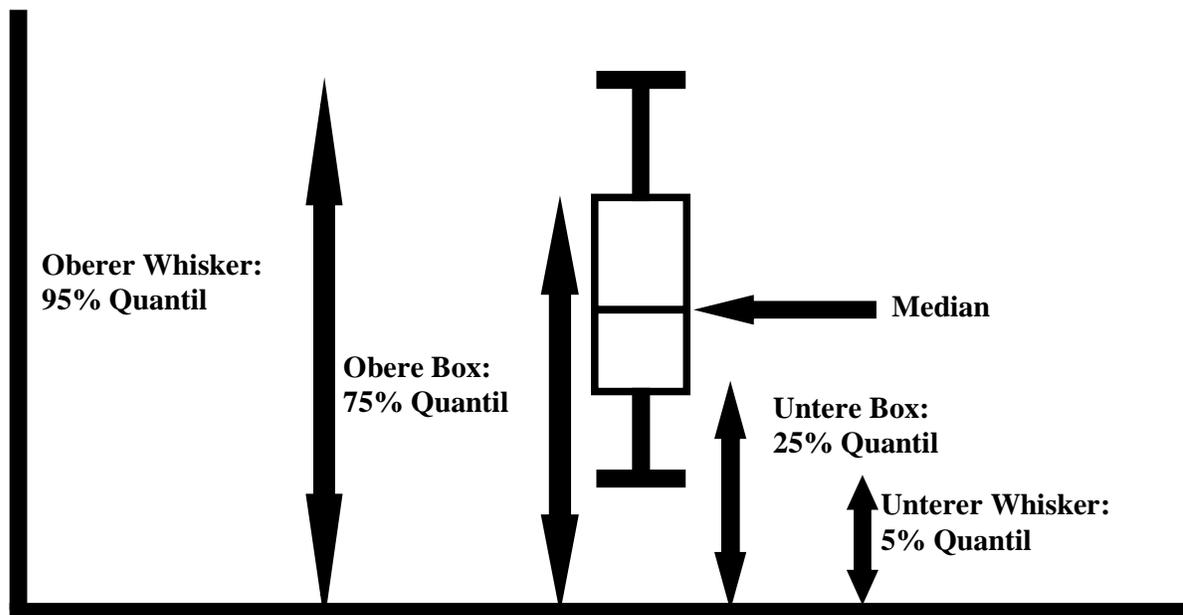
Die Merkmalswerte 3, 7, 70, 75 und 10.000 weisen ebenfalls den Median 70 auf.

Die Merkmalswerte 4, 6, 72, 80, 111 und 325 weisen den Median 76 auf

$$\text{Median} = \text{Mittelwert aus } 72 \text{ und } 80 = 76$$

Die Beispiele belegen die geringe Empfindlichkeit des Medians gegenüber Ausreißern (hier der Wert 10.000 im 2. Beispiel).

Bei nicht normalverteilten Daten wird die Streuung der Daten durch Quantile angegeben. Hierbei bezeichnet man als  $x\%$ -Quantil dasjenige Intervall, in dem sich  $x\%$  der Merkmalswerte einer Datensammlung befinden. Eine häufig benutzte Darstellung des Medians und der zugehörigen Quantile ist der „Box and Whiskers Plot“. Dieser Plot besteht –wie Abbildung 3 zeigt- aus zwei Boxen und zwei „Whiskers“ (=Schnurrhaare). Die Trennlinie zwischen beiden Boxen markiert den Median. Beide Boxen enthalten jeweils 25% der Merkmalswerte. Jede Box samt Whisker enthält jeweils 45% der Merkmalswerte. Zwischen X-Achse (Annahme: nur positive Datenwerte!) und unterem Rand des unteren Whiskers liegen 5% der Merkmalswerte (5% Quantil), d.h. unterhalb des „Box and Whiskers Plots“ liegen noch 5% der Merkmale. Zwischen X-Achse und unterem Rand der unteren Box liegen 25% der Merkmalswerte (25% Quantil). Zwischen X-Achse und oberem Rand der oberen Box liegen 75% der Merkmalswerte (75% Quantil). Und schließlich liegen zwischen X-Achse und oberem Rand des oberen Whiskers 95% der Merkmalswerte (95% Quantil).



**Abbildung 8.3:** „Box and Whiskers Plot“

Der “Box and Whiskers Plot” liefert sehr vielfältige Informationen über die Symmetrie und die Verteilung der Merkmalswerte.

#### 8.4.5. Statistische Tests

In der überwiegenden Mehrzahl tierexperimenteller Studien wird ein Testverfahren (Wirksamkeit einer Prüfsubstanz oder einer Prüftechnik) mit einem Kontrollverfahren (Wirksamkeit einer Kontrollsubstanz oder einer Kontrolltechnik) verglichen. Dabei wird die hypothetische Annahme, dass das Prüfverfahren sich nicht vom Kontrollverfahren unterscheidet, als Nullhypothese bezeichnet. Durch den Tierversuch soll somit die Nullhypothese überprüft werden. Hierzu wird eine bestimmte Anzahl von Tieren dem Prüfverfahren und eine andere Tiergruppe dem Kontrollverfahren ausgesetzt. Es werden zur Überprüfung der Nullhypothese also bei Prüf- und Kontrollverfahren Stichproben gezogen. Mit den Ergebnissen der Stichproben wird ein statistischer Test (Signifikanztest) durchgeführt und diese Datenanalyse soll Auskunft über die Bestätigung oder Nicht-Bestätigung der Nullhypothese liefern.

		„Wahrheit“	
		Nullhypothese gilt nicht	Nullhypothese gilt
Test- ergeb- nisse	Nullhypothese wird abgelehnt		<b>Fehler 1. Art (Wahrscheinlichkeit: <math>\alpha</math>)</b>
	Nullhypothese wird nicht abgelehnt	<b>Fehler 2. Art (Wahrscheinlichkeit: <math>\beta</math>)</b>	

**Abbildung 8.4: Mögliche Ausgänge statistischer Tests und Fehler-Wahrscheinlichkeiten.**

Wie Abbildung 4 zeigt, können bei einem statistischen Test prinzipiell folgende „Komplikationen“ auftreten:

**Fehler 1. Art:** Ein Fehler 1. Art tritt dann auf, wenn die Nullhypothese zwar tatsächlich wahr ist (Testverfahren ist dem Kontrollverfahren nicht überlegen), aber dies vom statistischen Test nicht erkannt wird. Das Testverfahren wird folglich zu Unrecht als dem Kontrollverfahren überlegen angesehen, das Ergebnis ist falsch positiv. Die Folge eines Fehlers 1. Art kann beispielsweise sein, dass kranke Haustiere mit einer Substanz behandelt werden, die überhaupt keine Heilungswirkung hat. Diesen Tieren würde folglich eine effiziente Therapie vorenthalten und darüber hinaus würden die Tiere unnötigerweise den Nebenwirkungen einer therapeutisch unwirksamen Substanz ausgesetzt. Die Wahrscheinlichkeit eines Fehlers 1. Art wird als  $\alpha$  bezeichnet und muss gering gehalten werden. Der  $\alpha$ -Wert ist vor der Durchführung des statistischen Tests festzulegen. In Abhängigkeit von den Konsequenzen des Fehlers 1. Art wird als Obergrenze für die Wahrscheinlichkeit  $\alpha$  üblicherweise ein Niveau von 5 % oder von 1 % akzeptiert. Dieses Niveau wird als Signifikanzniveau bezeichnet.

**Fehler 2. Art:** Ein Fehler 2. Art tritt dann auf, wenn die Nullhypothese tatsächlich nicht zutrifft (Testverfahren ist dem Kontrollverfahren überlegen!), aber dies vom statistischen Test nicht erkannt wird. Die Überlegenheit des Prüfverfahrens wird folglich vom statistischen Test nicht wahrgenommen; das Testergebnis ist falsch negativ. Die Folgen eines Fehlers 2. Art sind weniger gravierend als die eines Fehlers 1. Art. Eine mögliche Konsequenz ist, dass ein effizientes Therapieverfahren verworfen wird und somit nicht zur klinischen Anwendung kommt. Die Wahrscheinlichkeit für Fehler 2. Art wird als  $\beta$  bezeichnet. Die zu  $\beta$  komplementäre Wahrscheinlichkeit ( $1-\beta$ ) wird als Trennschärfe oder „Power“ eines statistischen Tests bezeichnet. Die Trennschärfe ( $1-\beta$ ) charakterisiert dabei die Wahr-

scheinlichkeit eines statistischen Tests, richtigerweise das Nicht-Zutreffen der Nullhypothese zu erkennen. Bei statistischen Test ist ein geringer  $\beta$ -Fehler und -unmittelbar daraus folgend- eine hohe Trennschärfe gewünscht. Folgende Punkte sind geeignet, um den  $\beta$ -Fehler zu verringern und damit die Trennschärfe zu vergrößern:

-Erhöhung des Stichprobenumfangs

-Verringerung der Standardabweichung der Stichprobe: Dies kann teilweise durch geschickte Auswahl des Zielparameters erreicht werden: soll beispielsweise der Einfluss eines Lipidsenkers auf den Cholesterinspiegel untersucht werden, so ist es ungünstig, den absoluten Cholesterinspiegel als Zielparameter zu wählen, da dieser sehr stark variiert. Stattdessen sollte die Differenz der Cholesterinwerte vor Studienbeginn und nach Therapie als Zielparameter eingesetzt werden, da dieser Wert eine geringere Standardabweichung aufweist.

#### **8.4.6. Auswahl eines geeigneten statistischen Tests**

Für die Auswahl eines geeigneten Signifikanztests ist entscheidend, ob die Daten der Stichprobe einer Normalverteilung unterliegen.

Ist anzunehmen, dass die Daten in der Prüf- und Kontrollstichprobe einer Normalverteilung unterliegen, so sollte der T-Test durchgeführt werden. Der T-Test stellt einen parametrischen Test dar, d.h. es fließen die originären Messwerte der Stichproben in den Test ein. Es ist zu entscheiden, ob ein einseitiger oder ein zweiseitiger T-Test eingesetzt wird. Der einseitige T-Test darf nur dann Anwendung finden, wenn sich die Daten der Prüfstichprobe prinzipiell nur in einer Richtung in Bezug auf die Daten der Kontrollstichprobe verhalten können (z.B. Prüfverfahren kann prinzipiell nicht ungünstiger als Kontrollverfahren sein).

Muss angenommen werden, dass die Daten in der Prüf- und Kontrollstichprobe keiner Normalverteilung unterliegen, so müssen nicht-parametrische statistische Tests zum Einsatz kommen. Als Beispiel sei hier der Wilcoxon-Test aufgeführt, der auf Rängen basiert. Hier fließen nicht die Absolutwerte der Messwerte in den statistischen Test ein, sondern deren Reihenfolge.

#### **8.5. Literatur**

- Mead R, The design of experiments. Cambridge University Press, Cambridge, New York, 1988  
 Van Zutphen LFM, Baumans V, Beynen AC (eds) (1995) Grundlagen der Versuchstierkunde, Gustav Fischer, Stuttgart, Jena, New York  
 Trampisch HJ, Windeler J, (eds) Medizinische Statistik, Springer-Verlag, Berlin, Heidelberg, New York, 1997

Festing MFW and Weigler BJ, Experimental Design and Statistical Analysis, in: Handbook of Laboratory Animal Science, Volume 1, Essential Principles and Practices, Hau J and VanHoosier GL (eds), CRC Press, Boca Raton, 2003