Praktische Hinweise
zur Anwendung statistischer Methoden
bei einfach strukturierten bewegungs- und trainingswissenschaft-lichen Experimenten

> **Statistik** ist eine Zusammenfassung von Methoden, die es uns erlauben, im Falle der Ungewißheit vernünftige Entscheidungen zu treffen. Um solche Entscheidungen treffen zu können, ist es erforderlich, Daten zu sammeln und zu analysieren. Statistik ist somit eine Zusammenfassung von Methoden zur Sammlung und Analyse von Daten.

Die folgenden “praktischen Hinweise” sollen keine abgeschlossene Theorie des wissenschaftlichen Testens darstellen, sondern lediglich Studenten, die an der Durchführung eines einfach strukturierten Experimentes in den Bewegungs- und Trainingswissenschaften interessiert sind (und die die Grundrechenarten noch nicht ganz vergessen haben sowie sich langsam daran gewöhnen, sich bei der Erledigung ihrer Arbeiten in Zusammenhang mit dem Studium eines Personalcomputers zu bedienen), die wichtigsten Definitionen zur Statistik und die grundlegenden Maßnahmen zur Planung, Durchführung und Auswertung eines Experimentes liefern. Die “praktischen Hinweise” orientieren sich nicht an Aufbau und Inhalt von Lehrbüchern der Statistik, sondern an den Problemen, Wünschen und Bedürfnissen von Studenten, so wie sich diese seit Jahren in der Betreuung und Beurteilung von schriftlichen Hausarbeiten herauskristallisiert haben.

Die “praktischen Hinweise” können auch bei Arbeiten in anderen Teilbereichen der Sportwissenschaften (Sportpsychologie, -soziologie, -pädagogik u.a.) von Nutzen sein und wenn es darum geht, in der Sportfachliteratur Forschungsberichte zu lesen und zu beurteilen. Ganz nebenbei sollen sie auch zu einer kritischen Haltung gegenüber allzu voreiligen, euphorisch vorgetragenen oder doktrinär vertretenen Wahrheiten (oder “Halbwahrheiten”) erziehen und vor ideologisierenden Rattenfängern und mystifizierenden Wundertätern schützen helfen.

Die “praktischen Hinweise” sind wie folgt gegliedert:

1. Die vier Stufen einer wissenschaftlichen Untersuchung

- Beobachtung; Entstehen einer Ungewißheit:
Ein Wissenschaftler beobachtet Vorgänge, Geschehnisse oder Verhaltensweisen und stößt dabei auf ein ihm anfänglich unerklärliches Phänomen (= Fragestellung).

- Vorhersage:
Der Wissenschaftler beschäftigt sich mit den bestehenden Theorien bzw. mit dem derzeitigen Wissensstand zum vorliegenden Problem und versucht, die offene Frage aus der Sicht der bisherigen Kenntnisse hypothetisch zu beantworten
(= Hypothesenbildung). Er trifft somit eine Vorhersage, wobei weiterhin ungewiß ist, ob und unter welchen Bedingungen die Vorhersage tatsächlich zutrifft.

- Durchführung einer Untersuchung:
Um die Gültigkeit der Vorhersage zu prüfen bzw. um die Ungewißheit über den hypothetischen Zusammenhang zu beseitigen, plant der Wissenschaftler eine Untersuchung, führt diese durch und wertet sie aus.

- Bestätigung:
Auf dem Wege über die Auswertung und Analyse der in der Untersuchung gewonnenen Daten versucht der Wissenschaftler herauszufinden, ob er die Vorhersage mit ausreichender Sicherheit (Gewißheit) als bestätigt annehmen kann. Dabei ist das Ergebnis mit der bestehenden Theorie zu vergleichen bzw. in sie zu integrieren. Ist die Wahrscheinlichkeit, daß das Ergebnis der Untersuchung nur zufällig zustandegekommen ist, zu groß, darf er die Vorhersage nicht als Tatsache oder Wahrheit verkünden bzw. er darf die Hypothese nicht annehmen.

2. Die Hypothese

Am Anfang steht die Hypothese! Das soll heißen: Ohne eine begründete, auf der Basis der bestehenden Theorie abgeleitete Hypothese macht die Planung einer Erhebung, einer Untersuchung oder eines Experimentes keinen Sinn.

> Eine **Hypothese** ist eine Vorhersage über den Ablauf eines Ereignisses oder einer Verhaltensänderung, die als Antwort auf eine Fragestellung aus dem derzeitigen Kenntnisstand der Wissenschaft resultiert.

Eine Hypothese muß drei Forderungen erfüllen:

1. Eine Hypothese muß eine konkrete Ereignis- bzw. Verhaltens**vorhersage** liefern.

   Beispiel: Gültige Vorhersage: "Krafttraining der ischiokruralen Muskeln verbessert die Sprintgeschwindigkeit".
   Keine gültige Vorhersage: "Spiel ist Bewegung".

2. Eine Hypothese muß **überprüfbar** sein, d.h. sie muß so formuliert werden, daß durch experimentelle Methoden mit Hilfe gewonnener Daten geprüft werden kann, ob die Vorhersage zutrifft oder nicht. Sie muß so formuliert sein, daß - unabhängig

vom Urheber der Voraussage - **jeder**, der ein Interesse daran hat, die Überprüfung vornehmen kann.

Beispiel: Überprüfbar: "Dehnungstraining vergrößert die Gelenkreichweite." oder "Turnen stärkt die Schultermuskulatur".
Schwer überprüfbar: "Laufen macht Laune".
Nicht überprüfbar: "Angenehme Gedanken mindern innere Verkrampfungen" oder "Turnen stärkt den Charakter".

3. Die in der Hypothese verwendeten Begriffe müsssen **definierbar** und **operationalisierbar** sein.

*Beispiel: "Gelenkreichweite" ist definiert als der maximale in einem Gelenk mögliche Bewegungswinkel. "Gelenkreichweite" ist operationalisierbar durch Einsatz entsprechender Winkelmesser, und zwar durch **jede Person,** die dazu bereit und daran interessiert ist.*
"Innere Verkrampfungen" sind weder definierbar noch meßbar (es sei denn, durch einen Wunderheiler und nur durch diesen).

Hypothesen können so formuliert werden, daß sie für eine **einseitige** Fragestellung oder für eine **zweiseitige** Fragestellung eine Vorhersage treffen.
Die einseitige Fragestellung liefert stets die konkretere Vorhersage:

Beispiel: "Dehnungstraining führt zu einer Abnahme der Ruhespannung". (Es wird in der Hypothese nicht erwartet, daß Dehnungstraining auch zu einer Zunahme der Ruhespannung führen könnte).

Die zweiseitige Fragestellung erwartet einen Effekt, läßt die Richtung des Effektes aber offen.

Beispiel: "Regelmäßiges Krafttraining verändert die Körperhaltung". (Es bleibt offen, in welcher Weise und in welcher Richtung Krafttraining die Körperhaltung beeinflussen wird)

Es gibt generell zwei Arten von Hypothesen:

1. Unterschiedshypothesen:

Unterschiedshypothesen sagen einen Unterschied zwischen zwei Personengruppen bzw. Stichproben oder aber einen Unterschied innerhalb einer Stichprobe vor und nach einer Behandlung vorher.

a) Unterschied **zwischen** zwei Personengruppen:

Beispiel: "Sprinter sind zu einer höheren Schrittfrequenz fähig als Langstreckler".

b) Unterschiede **innerhalb** einer Personengruppe vor und nach einer Behandlung.

Beispiel: "Nach einem Schnellkrafttraining zeigen Langstreckler höhere Schrittfrequenzen als vorher."

2. Zusammenhangshypothesen:

In Zusammenhangshypothesen wird eine einseitige oder wechselseitige Abhängigkeit zwischen zwei oder mehreren Variablen prognostiziert.

Beispiel: "Es besteht ein Zusammenhang zwischen Körpergröße und Schrittfrequenz".

(Diese Hypothese ist zwar als zweiseitige Fragestellung formuliert, zielt aber stillschweigend auf einen einseitigen Effekt hin, da man erwarten kann, daß niemand mit zunehmender Körpergröße eine Zunahme der Schrittfrequenz vorhersehen wird.)

Die Art der Hypothese bestimmt sowohl die Art des Experimentes und dessen Planung als auch die statistische Auswertung der Daten.

Solange ein Experiment die Richtigkeit einer hypothetischen Vorhersage nicht bestätigt hat, sollte der ernsthafte, kritische Wissenschaftler mit ihr nicht argumentieren oder gar auf ihr weitergehende Maßnahmen oder Theorien aufbauen.

Beispiel: Die Vorhersage “Krafttraining führt zu Muskelverkürzungen” klingt in mancher Hinsicht plausibel, ist aber eine bisher unbestätigte Hypothese. Zudem ist der Begriff “Muskelverkürzungen” weder ausreichend definiert noch operationalisierbar. Trotzdem werden auf der Basis dieser Vorhersage Stretchingtheorien aufgebaut, Therapiemaßnahmen gestaltet und fleißig praktiziert und Theorien über muskuläre Dysbalancen entwickelt.

Um einer solchen Gefahr zu begegnen, sollte man zu der Hypothese (auch: **Alternativhypothese, Arbeitshypothese**) stets die **Nullhypothese** bilden.

Beispiel: “Sprinter und Langstreckler unterscheiden sich nicht in der Schrittfrequenz”. “Regelmäßiges Krafttraining beeinflußt nicht die Körperhaltung”.

Die Nullhypothese ist so lange beizubehalten, bis ein Experiment sie als falsch (nicht haltbar) identifiziert bzw. die Gültigkeit der Alternativhypothese bestätigt hat.

Diese Empfehlung widerspricht der Einstellung naiver Menschen, die nur zu gerne bereit sind, ihnen angenehme Vorhersagen zu glauben oder gar als Wahrheiten zu behandeln. So begegnet man häufig Aussagen wie (im täglichen Leben) “Vom tüchtigen Essen wird man groß und stark!” oder (in pädagogischen Schriften) “Bewegungsspiele verbessern die Orientierungsfähigkeit!”. Solche Aussagen klingen oft äußerst plausibel, sind aber in den meisten Fällen nicht empirisch auf ihre Gültigkeit hin überprüft oder entbehren gar jeglicher realen Grundlage. Statt dessen entspringen sie einem Wunschdenken zur Befriedigung egoistischer Bedürfnisse oder stellen erstrebenswerte Ziele dar. Die kritiklose Annahme solcher “Hypothesen” als Wahrheiten kann unterschiedliche Konsequenzen haben: Im ersten Fall (“Vom tüchtigen Essen....”) können diese höchst nachteilig sein, weil zu “tüchtiges” Essen in der Regel zu Fettsucht und Kreislauferkrankungen führt. Im zweiten Fall (“Bewegungsspiele....”) können auf den ersten Blick keine negativen Konsequenzen gesehen werden; denn Bewegungsspiele sind aus den verschiedensten Gründen empfehlenswert und werden irgendeinen positiven Effekt schon haben - was aber noch nicht dazu berechtigt, die obige Aussage als Wahrheit zu behandeln.

Kritische Wissenschafter sollten in solchen Fällen wie folgt formulieren: “Es ist (aus diesen und jenen Gründen) anzunehmen, daß ....!” Dadurch wird deutlich gemacht, daß es sich um eine (wenn auch gut begründete) Vermutung, Vorhersage oder Hypothese handelt, die auf ihre Verifizierung wartet.

3. Zu den Daten

3.1 Definitionen

Das, was in bewegungstheoretischen Untersuchungen beobachtet, beurteilt oder gemessen wird, sind **Merkmale** von Bewegungen oder Personen. Bewegungen oder Personen werden deshalb auch **Merkmalsträger** genannt. In der Regel treten die Merkmale (z.B. Kraft, Geschicklichkeit, Fitness) bei einzelnen Personen oder unter bestimmten Bedingungen (vor oder nach einem Training...) in verschiedenen Merkmals**ausprägungen** auf. Sie sind veränderlich, variabel und werden deshalb auch **Variablen** genannt.

Generell unterscheidet man zwischen **qualitativen** und **quantitativen** Merkmalen. Die qualitativen Merkmale (qualitativen Variablen, Nominalvariablen) unterscheiden sich bezüglich der Ausprägungs**art**. Die quantitativen Variablen unterscheiden sich hinsichtlich des Ausprägungs**grades**.

3.2 Vier Skalentypen

Um den Ausprägungsgrad festzulegen, wird die Ausprägung des Merkmals bewertet oder gemessen (quantifiziert) und ihr eine **Maßzahl** zugeordnet. Das Zuordnen bzw. Einordnen von Maßzahlen erfolgt innerhalb einer **Skala**. Je nach Art des Merkmals sind **4 Skalentypen** zu unterscheiden:

1. Nominalskala:
Die Ausprägungen eines Merkmals unterscheiden sich nur **qualitativ** (siehe vorn).

2. Ordinalskala (Rangskala):
Die Ausprägungen eines Merkmals unterscheiden sich graduell, wobei eine größer-kleiner-Relation (besser-schlechter-Relation) festgestellt werden kann. Die Abstände zwischen den einzelnen Stufen der Ordinalskala bleiben jedoch ungewiß. (Es ist nicht klar, ob der Abstand zwischen den Noten gut und sehr gut genauso groß ist wie zwischen den Noten befriedigend und gut). Somit wird lediglich **die Reihenfolge von Rängen (bzw. von Rangplätzen)** festgestellt.

3. Intervallskala
Die Ausprägungen der Merkmale unterscheiden sich quantitativ-graduell, wobei die Intervalle zwischen den benachbarten Skalenwerten gleich groß sind. Hat die Skala einen absoluten Nullpunkt, stellt sie eine

4. Verhältnisskala (metrische Skala)
dar. Nur innerhalb dieses Skalentyps sind mathematische Operationen (Addition, Subtraktion, Multiplikation, Division ...) möglich.

Während die nominal- und ordinalskalierten Variablen **diskret**, d.h. in einzelnen *Stufen* ausgeprägt sind, können die intervall- und metrisch skalierten Variablen **kontinuierlich** ausgeprägt sein. In der Regel werden sie jedoch diskret gemessen **-** in Abhängigkeit von den Erfordernissen oder der Meßgenauigkeit des Meßinstrumentes.

Beispiele zu den Skalentypen.

Merkmalsträger	***Merkmal***	***Merkmalsausprägung***
1. Nominalskalierung		
Sportstudenten des 1. Semesters	*Geschlecht*	*weiblich; männlich*
Sportstudenten des 3. Semesters	*bevorzugte Sportart*	*Schwimmen; Volley-ball; Fußball; Gymnastik*
2. Ordinalskalierung		
Teilnehmer des Methodikkurses	*Leistung in der Methodikklausur*	*Noten: 2; 2,3; 2,7; 3,0; 3,7; 4,0; 5..*
Teilnehmer der Turnprüfung	*Leistung bei der Bodenkür*	*Punkte: 8,3; 8,7; 9,0; 9,5;.....*
3. Intervallskalierung		
Mitglieder einer Trainingsgruppe	*Körpertemperatur nach 10.000m-Lauf*	*Grad Clesius: 37,1; 37,5; 37,6; 37,9.....*
4. Verhältnisskalierung		
Personen einer Stichprobe	*Leistung in einem Maximalkrafttest*	*Kraft in Newton: 114,5; 115; 132,8; 140,1; ....*
Teilnehmer der Leichtathletikprüfung	*Kugelstoßleistung*	*Weite in m: 7,51; 7,67; 8,01; 9,67; 9,68;.....*

Häufig werden in Datenerhebungen **zwei** (oder mehrere) unterschiedliche Merkmale auf ihre Ausprägung hin untersucht bzw. festgestellt, ob sich Merkmale gegenseitig beeinflussen.

Beispiel: Bei der Erhebung, ob Sportstudentinnen dehnfähiger als Sportstudenten sind, werden das Merkmal (die Variable) “Geschlecht” und das Merkmal (die Variable) “Dehnfähigkeit” untersucht.

Dabei erwartet man, daß das ***eine*** Merkmal (hier: Geschlecht) einen Einfluß auf das ***andere*** Merkmal (Dehnfähigkeit) ausübt, d.h. eines der beiden Merkmale (hier Dehnfähigkeit) scheint in der Ausprägung von dem anderen Merkmal (hier: Geschlecht) ***abhängig*** zu sein. Man nennt es **abhängige Variable**, das andere Merkmal **unabhängige Variable** (da zweifellos das Merkmal “Geschlecht” nicht von der Dehnfähigkeit bestimmt werden kann).

Beispiele:

unabhängiges Merkmal	***abhängiges Merkmal***
Alter	*Ausdauerleistung*
Körpergröße	*Hochsprungleistung*
Zugehörigkeit zur Kraft-trainingsgruppe (ja; nein)	*Maximalkraft*
Vortestleistung	*Nachtestleistung*

4. Methoden der Datensammlung

In der Regel werden Daten durch **Beobachtung, Befragung, Tests** und **Experimente** gesammelt. (Im folgenden wird nur das *Experiment* besprochen, da die Anforderungen an das Planen, Durchführen und Auswerten von Experimenten auch für die übrigen Methoden gelten.)

4.1 Das Experiment (Definition)

Ein **Experiment** ist ein Verfahren, bei dem unter kontrollierten und gezielt variierten Bedingungen Vorgänge, Reaktionen oder Verhaltensänderungen herbeigeführt werden, um Hypothesen zu verifizieren oder zu falsifizieren.

An ein Experiment werden 4 Bedingungen gestellt:

1. Willkürlichkeit:
Der Experimentator muß den Vorgang bzw. das Verhalten, das er zu beobachten beabsichtigt, willkürlich jederzeit herbeiführen können.

2. Variierbarkeit:
Die zu untersuchende Variable muß unter Konstanthaltung aller übrigen Bedingungen variierbar sein, um die Variation der Ergebnisse auf die Variation der Variablen zurückführen zu können.

3. Wiederholbarkeit:
Experimente müssen sich unter möglichst gleichen Bedingungen wiederholen lassen, um zu gewährleisten, daß das Ergebnis eines Experimentes (unabhängig vom früheren Experimentator) überprüft werden kann.

4. Kontrollierbarkeit
Anordnung und Ablauf des Experimentes muß kontrollierbar (und somit auch protokollierbar) sein, um die Bedingungen einer identischen Wiederholung des Experimentes erfüllen zu können.

4.2 Arten von Experimenten

1. Querschnittsuntersuchung
Querschnittsuntersuchungen vergleichen zwei oder mehrere Personen oder Gruppen im Hinblick auf die Ausprägung eines oder mehrerer Merkmale.

Beispiel: Die Hypothese “Turner haben dehnfähigere Hüftstreckmuskeln als Leichtathleten” überprüft man, indem man die momentane Ausprägung des Merkmals Dehnfähigkeit der Hüftstreckmuskeln einer Gruppe Turner mit der einer Gruppe Leichtathleten vergleicht.

2. Längsschnittsuntersuchung
Längsschnittsuntersuchungen prüfen zeitliche Entwicklungsprozesse einer Person oder einer Gruppe von Personen, indem die Veränderung der Ausprägung eines

Merkmals im Laufe der Zeit oder unter der Wirkung einer Behandlung festgestellt wird. Längsschnittuntersuchungen verlangen somit, daß das Merkmal (mindestens) zweimal geprüft wird, nämlich einmal vor Beginn der Behandlung (= Vortest) und einmal nach Abschluß der Behandlung (= Nachtest). Um sicher zu gehen, daß die Merkmalsänderung auf die Behandlung zurückgeführt werden kann, sollte stets zusätzlich eine Kontrollgruppe (Kap. 4.4) eingerichtet werden, die in der Zeit zwischen Vortest und Nachtest nicht behandelt wird.

*Beispiel: Die Hypothese “Ein 10wöchiges Dehnungstraining verbessert die Dehnfähigkeit der Hüftstreckmuskeln” prüft man anhand einer entsprechenden Längsschnittuntersuchung, indem man zuerst mit der Versuchsgruppe (und der Kontrollgruppe) einen **Vortest** zur Dehnfähigkeit der Hüftstreckmuskeln ausführt, dann die Versuchsgruppe mit einem Dehnungstraining behandelt und danach mit der Versuchsgruppe (und der Kontrollgruppe) einen **Nachtest** zur Dehnfähigkeit der Hüftstreckmuskeln absolviert. Vortest und Nachtest müssen unter identischen Bedingungen realisiert werden.*

Längsschnittuntersuchungen erfordern einen hohen Zeitaufwand. Um die damit verbundenen Kosten niedrig zu halten, werden häufig Einzelfallstudien (in Form von Zeitreihenanalysen) durchgeführt. Diese unterliegen abweichenden Regeln, die hier nicht behandelt werden sollen.

A. Erkundungsexperiment, Pilot-Studie

Wenn zu einer Fragestellung nur wenige Forschungsergebnisse vorliegen, sich die hypothetischen Vorhersagen nur vage formulieren lassen, kann man Erkundungsexperimente durchführen, um einen ersten Eindruck von den Wirkungszusammenhängen zu bekommen. Läßt dieses Erkundungsexperiment Gesetzmäßigkeiten erkennen, müssen diese durch ein Entscheidungsexperiment überprüft werden.

B. Entscheidungsexperiment

Mit Hilfe eines Entscheidungsexperimentes werden Daten gesammelt, die die Grundlage für die Entscheidung, ob eine Hypothese als gültig angesehen werden kann, liefern.

a) Laborexperiment

Das Laborexperiment stellt das typische Experiment dar, für das die vorn gelieferte Definition gilt. Die Bezeichnung “Labor...” bezieht sich auf den Umstand, daß man die Bedingungen von Versuchsort, Versuchsablauf u.ä. konstant hält und nur die zu prüfende Variable unter Kontrolle variiert.

Beispiel: Zur Prüfung der Hypothese “Die Weitsprungleistung ist von der Dämpfungseigenschaft des Absprungbalkens abhängig” variiert man die Materialeigenschaft des Absprungbalkens und versucht, alle übrigen Bedingungen, die die Weitsprungleistung beeinflussen könnten (Tageszeit, Jahreszeit, Wind und Wetter, Anlauflänge ....) konstant zu halten.

b) Das Feldexperiment (quasi-experimentelle Untersuchung):

Häufig ist es schwierig, die zu prüfende Einflußgröße gezielt zu variieren. Deshalb hofft man auf zwangsläufig (natürlich) auftretende Schwankungen der Einflußgröße “im Feld” und mißt, wie sich diese Schwankungen auf das zu messende Merkmal auswirken.

Beispiel: Zur Püfung der Hypothese “Die Weitsprungleistung ist von der Richtung der Absprungkraft abhängig” läßt man eine oder mehrere Personen mehrfach (unter sonst konstanten Bedingungen) Weitsprünge absolvieren und untersucht den Einfluß der zwangsläufig auftretenden Variationen in der Richtung der Absprungkraft auf die Sprungleistung.

(Der Unterschied zwischen Labor- und Feldexperiment liegt somit nur darin, daß im ersten Fall die Bedingungen (die unabhängigen Variablen) **gezielt variiert** werden, im zweiten Fall **zufällig von selbst variieren**.)

4.3 Gütekriterien von Meßmethoden

Je nach Art des Merkmals, das im Experiment quantifiziert werden soll, ist ein passendes Meßinstrument (Fragebogen, Schätzskala, Wertungsrichtlinien, Waage, Bandmaß....) zu wählen. Dabei sind die Kriterien der Validität, Reliabilität und Objektivität zu beachten.

1. Validität: Darunter versteht man den Grad der *Gültigkeit* einer Meßmethode im Hinblick auf das Meßziel. Im Rahmen naturwissenschaftlicher Meßvorgänge ergibt sich die Validität in der Regel von selbst. Zeit mißt man mit Zeitmeßgeräten (Stoppuhr, elektronische Lichtschranken, selbst eine Eieruhr ist ein valides Zeitmeßinstrument), Längen mißt man mit Längenmeßgeräten, Kraft .... Schwieriger ist es, die Validität von Fragebögen für Einstellungstests, Testbatterien zur Quantifizierung der Koordinationsfähigkeit usw. zu bestimmen. Hier sollte man auf validierte, in der Literatur beschriebene oder im Handel erhältliche Instrumentarien zurückgreifen.

2. Reliabilität ist der Grad der *Genauigkeit* bzw. *Zuverlässigkeit*, mit der ein Meßinstrument ein Merkmal quantifiziert.

Beispiel: Die Reliabilität einer elektronischen Lichtschranke, die Zeiten in 1/100 sec mißt, liegt selbstverständlich deutlich höher als die einer Stoppuhr mit Handbetrieb. Die Reliabilität eines Stahlmaßbandes ist höher als die eines in sich etwas nachgiebigen Leinenmaßbandes.

Bei selbst konstruierten Meßinstrumenten (auch wenn sie auf elektronischer Basis funktionieren) ist die Reliabilität stets zu prüfen. Dies wird in der Regel mit Wiederholungsmessungen (siehe Kap. 5.8), Parallelmessungen, oder mit der Testhalbierungsmethode (siehe einschlägige Literatur) praktiziert.

3. Objektivität ist der Grad der *Unabhängigkeit* einer Meßmethode vom messenden *Subjekt*.

Beispiel: Bei einer elektronischen Lichtschranke, die die Maßzahl digital liefert und zum Datenspeicher weiterleitet, gibt es keine Verfälschungsmöglichkeit durch den Versuchsleiter.
Das Messen der Schulterbreite von Versuchspersonen mit Hilfe eines Maßbandes ist dagegen davon abhängig, wie genau der Versuchsleiter die Meßpunkte bestimmt, wie er von Person zu Person das Maßband anlegt, wie er beim Ablesen auf- oder abrundet usw.

In solchen Fällen, in denen die Objektivität nicht eindeutig ist, sollte man diese prüfen, indem man eine Versuchsserie von zwei oder mehreren Versuchsleitern unabhängig voneinander messen läßt und den Korrelationskoeffizienten (siehe Kap. 5.8) der Meßreihen berechnet.

Kennzeichnet sich ein Meßinstrument durch eine geringere Objektivität, leidet darunter zwangsläufig auch die Reliabilität. Entsprechend muß bei geringer Reliabilität auch die Validität eines Meßverfahrens leiden.

Zum Bereich der Objektivität von empirischen Methoden gehören auch Effekte, die sich aus der Erwartung der am Experiment beteiligten Personen über den Ausgang des Experimentes ergeben. Sowohl Versuchsleiter als auch Versuchspersonen verhalten sich in Experimenten in der Regel (unbewußt) im Sinne der Hypothesen, sofern sie diese kennen.

a) Versuchsleiter-Erwartungseffekt (ROSENTHAL-Effekt):
Läßt ein Meßverfahren einen Entscheidungsspielraum des Versuchsleiters zu (etwa bei einem Abrunden oder Aufrunden), entscheidet der Versuchsleiter unbewußt und unbeabsichtigt meistens **im** Sinne der Hypothesen und dies um so eher, je weniger distanziert er an den Versuch herangeht bzw. je leidenschaftlicher er von seiner Vorhersage überzeugt ist.

Das gilt beispielsweise auch bei mündlichen Instruktionen der Versuchspersonen, bei denen durch unterschiedliche Betonung oder unterschiedliches Engagement Versuchspersonen unterschiedlich zur Lösung von Aufgaben motiviert werden können, je nachdem, ob sie zur Versuchs- oder Kontrollgruppe gehören.

Um solche Effekte zu vermeiden, sollte man alle Maßnahmen möglichst standardisieren (schriftliche Instruktionen, mechanisierte Auswerteverfahren usw.).

b) Versuchspersonen-Erwartungseffekt
Auch Versuchspersonen verhalten sich häufig unbewußt so, wie sie glauben, daß der Versuchsleiter dies von ihnen erwarten könnte, und dies um so eher, je mehr sie den Versuchsleiter respektieren, anerkennen oder schätzen.

Beispiel: Eine Vp mit Muskelverspannung erfährt vom Versuchsleiter eine Behandlung zur Entspannung. Nach der Behandlung glaubt die Vp (im Vertrauen auf die Wirksamkeit der Behandlung), die Verspannung habe nachgelassen, und gibt dies auf Anfrage dem Versuchsleiter bekannt, auch wenn objektiv keine Besserung eingetreten ist (Plazebo-Effekt). Unter Umständen scheut sich auch die Vp, ihre wahre Befindlichkeit bekanntzugeben, um den Versuchsleiter nicht über die Effektivität seiner Behandlung zu enttäuschen.

Auch sollte man Versuchspersonen vor einem Experiment nur über allgemeine bzw. über die dringend notwendigen Belange informieren, niemals jedoch über die hypothetischen Vorhersagen.

Versuchsleiter- und Versuchspersoneneffekte lassen sich am besten durch Doppel-Blindversuche vermeiden, bei denen weder der Versuchsleiter noch die Versuchspersonen über den erwarteten Ausgang des Experimentes informiert sind.

Die Einhaltung der Gütekriterien sollte stets **vor Beginn** des Experimentes (mit Hilfe entsprechender Vorversuche) abgesichert sein.

4.4 Versuchsgruppe, Kontrollgruppe

Hypothesen machen in der Regel Vorhersagen über Merkmale bzw. Merkmalsänderungen einer bestimmten Personengruppe (Sportstudenten eines Jahrganges, Sprinter einer Leistungsklasse...). Da es in der Regel zu aufwendig ist, **alle** Personen zu testen, die sich durch ein bestimmtes Merkmal oder eine Merkmalskombination auszeichnen (=**Grundgesamtheit**), wählt man aus ihr eine kleine Gruppe, **eine Stichprobe**, aus. Dies darf nicht willkürlich geschehen, sondern muß nach dem Zufallsprinzip (= randomisiert) erfolgen, so daß alle Personen der Grundgesamtheit gleiche Chancen haben, in die Stichprobe aufgenommen zu werden.

Diese Bedingung kann bei der Abfassung von studentischen Hausarbeiten meistens nicht erfüllt werden, da der studentische “Versuchsleiter” in der Regel - notgedrungen - seine besten Freunde bittet, als Versuchsperson zu dienen. Auf diese Weise sind die Vpn schon nach bestimmten Persönlichkeitsmerkmalen ausgewählt (Verfügungsbereitschaft, positive Einstellung zum Versuch und zum Versuchsleiter...), die nicht allen Personen der Grundgesamtheit eigen sind.

Gleiches gilt für die Aufteilung der Teilnehmer am Experiment auf die **Versuchsgruppe** bzw. die **Kontrollgruppe.** Die Versuchsgruppe ist derjenige Teil der Versuchspersonen, an denen die Wirkungen der Versuchbedingungen geprüft werden. Die Kontrollgruppe dient als “unbehandelte” Gruppe der Vergleichsmöglichkeit.

Bei Leistungs- oder Lernexperimenten sollte man außerdem dafür sorgen, daß das Vortestniveau in Versuchs- und Kontrollgruppen gleich hoch ist, weil Personen mit einem niedrigen (Ausgangs-) Leistungsstand ihre Leistung leichter und schneller verbessern können als Personen mit einem hohen Leistungsstand. Aus diesem Grunde teilt man die Gruppen erst nach dem Vortest auf, indem die Versuchsperson mit dem höchsten Leistungsstand in Gruppe 1, die zweit- und drittbeste in Gruppe 2, die viert- und fünftbeste in Gruppe 1 usw. aufgenommen werden (= parallelisieren).

Versuchsgruppen zur Prüfung von Unterschiedshypothesen im Bereich der Bewegungs- und Trainingswissenschaften sollten bei einfach strukturierten Experimenten 12-16 Personen (aber nicht weniger als 8 Personen) umfassen (siehe auch Kap. 5.10). Für Versuchsgruppen zur Prüfung von Zusammenhangshypothesen gibt es nicht solch eine einfache Empfehlung. Ihr Umfang hängt von unterschiedlichen Bedingungen ab (siehe Literatur).

Grundssätzlich sollte man die Vpn stets vor Beginn des Experimentes (ohne die Hypothesen zu verraten) über die körperlichen und psychischen Belastungen informieren und ihnen freistellen, jederzeit die Teilnahme zu kündigen bzw. den laufenden Versuch abzubrechen.

4.5 Versuchsplan

Vor der Durchführung eines Experimentes sollte man einen Versuchsplan erstellen, aus dem die Gruppenzugehörigkeit (unabhängige Variable), die Behandlung (treatment) der Versuchsgruppe, das Leistungsmerkmal (abhängigie Variable) sowie der Zeitpunkt der Datenerhebung hervorgeht.

Beispiel: Grundplanung:

unabhängige Variable	*Trainingsgruppe*	*Kontrollgruppe*
abhängige Variable	*Maximalkraft* *Schnellkraft*	

Versuchspläne lassen sich stufenweise erweitern, bis sie alle den Versuchsablauf betreffenden Maßnahmen enthalten:

Beispiel: Erweiterte Versuchsplanung:

	Vps-Kontingent: 24 Vpn	
Test	*Vortest*	
unabhängige Variable 1	*Maximalkraft, Schnellkraft*	
unabhängige Variable 2	*(Zugehörigkeit zur)* *Trainingsgruppe: 12 Vpn*	*(Zugehörigkeit zur)* *Kontrollgruppe: 12 Vpn*
Behandlung	*Querschnittstraining* *3 Wochen; 2mal/Woche;* *5 Sätze á 10 Wiederholungen mit 80% Fmax* *und 1 min Satzpause*	
Test	*Nachtest*	
abhängige Variable	*Maximalkraft, Schnellkraft*	
Auswertung	*Datenanalyse*	

4.6 Versuchsprotokoll

Der endgültige Versuchsplan kann gleichzeitig die Basis für das Protokollieren der Versuchsdurchführung sein, indem man z.B. die einzelnen Schritte abhakt. Zusätzlich sollten alle Vorkommnisse, Abweichungen und Auffälligkeiten protokollarisch festgehalten werden, um die Ursachen später (bei Datenanalyse) auftretender Ungereimtheiten aufdecken zu können.

5. Beschreibende Statistik

Ein Experiment liefert in der Regel eine Reihe von Beobachtungswerten (Meßwerten, Daten), die entweder in der Reihe des zeitlichen Ablaufes des Experimentes oder in alphabetischer Reihenfolge der Versuchspersonen gesammelt werden (= Urliste). Diese gilt es, gemäß den Erfordernissen der Fragestellung (Hypothese) zu ordnen, zu gliedern, zu analysieren und zu prüfen. Dazu dient
a) die beschreibende Statistik (deskriptive Statistik),
b) die Prüfstatistik (schließende Statistik, Inferenzstatistik).

Die im folgenden besprochenen Methoden und Verfahren beschränken sich auf ein Minimum. Für Vertiefung wird jeweils Literatur angegeben. Da derzeit die wichtigsten Berechnungen (z.B. des Mittelwertes, der Standardabweichung, des Korrelationskoeffizienten) mit Hilfe von Taschenrechnern und PC-Programmen möglich sind, werden die zutreffenden Formeln nicht im einzelnen abgeleitet und erläutert, sondern für den daran interessierten Studenten in Kleindruck mitgeliefert. Statt dessen wird der entsprechende Eingabe-Befehl in dem derzeit gebräuchlichsten Kalkulationsprogramm (MICROSOFT-EXCEL Version 5.0) wiedergegeben, erkennbar an der grauen Unterlegung in den Tabellen - verbunden mit dem Hinweis “exc”. Der mit Hilfe der Eingabe-Befehle berechnete Wert wird ***fett - kursiv*** kenntlich gemacht. Nur Formeln, für die im EXCEL-Programm keine zufriedenstellende Fassung vorliegt, werden ausführlich in der Anwendung besprochen. Bei der Wiedergabe von Formeln wird das Multiplikationssymbol (*) stetz eingesetzt. Die hier verwendeten Symbole und Abkürzungen der statistischen Kenngrößen entsprechen nicht in allen Fällen der (in dieser Hinsicht ohnehin nicht einheitlichen) Literatur, sondern sind mit der Maßgabe ausgewählt, Verwechselungen mit Symbolen und Abkürzungen der Physik zu vermeiden (siehe dazu auch das **Verzeichnis benutzter Abkürzungen** unter **9.**).

5.1 Allgemeine Erläuterungen, Beispiel

Die Erläuterung der beschreibenden (und Prüf-)Statistik soll anhand praktischer Beispiele erfolgen. Um den Leser nicht bei jedem Problemkomplex mit einem neuen Beispiel zu überfallen, wird **ein** Fall “konstruiert”, der einerseits für möglichst alle zu klärenden Fragen eingesetzt werden kann, andererseits aber derart einfach strukturiert ist, daß er als Einführung in die Fragestellungen, wie sie in schriftlichen Hausarbeiten gelöst werden können, dienen kann (Nur in wenigen Ausnahmefällen werden zusätzliche Fallbeispiele gebracht.):

Beispiel: Ein Student, der an der Hochschule X die Praxisprüfung im Gerätturnen beobachtet, kommt zu der Vermutung, daß Personen, die schnell laufen können, besser über das Pferd springen können als langsam laufende Personen, und daß aus diesem Grunde diejenigen Studenten, die sich auf die Sportart Leichtathletik spezialisiert haben, (neben den Turnern) Vorteile beim Pferdsprung haben könnten.

Nachdem er zu diesem Thema in der Literatur keine Aussagen gefunden hat, die ihm bei der Beantwortung der Fragen hätten helfen können, beschließt er, seine Annahme durch eine Untersuchung zu erhärten. Er plant eine ***Pilotstudie*** *(Kap. 4.2). Da er aus organisatorischen Gründen nicht alle Sportstudenten der Hochschule X (= Grundgesamtheit) überprüfen kann, sieht er vor, bei einer Stichprobe von 20 Vpn (Sportstudenten) die Ausprägung folgender Variablen zu bestimmen und zu vergleichen:*

Variable 1: bevorzugte Sportart, ermittelt durch Befragung,
Variable 2: Zensur des Pferdsprunges in der Turn-Praxisprüfung,
Variable 3: 100m-Sprintzeit der Leichtathletik-Praxisprüfung in sec.

Die Prüfung der Gütekriterien (siehe Kap. 4.3) unterbleibt vorerst (obwohl man sich zur Objektivität der Benotung im Turnen und über die Objektivität und Re-

liabilität von handgestoppten 100m-Zeiten so seine Gedanken machen könnte; siehe auch Kap. 5.8). Die gesammelten Daten werden in einer Liste in alphabetischer Reihenfolge der Vp-Namen zusammengefaßt (Tab. I).
An dieser Stelle beginnt die beschreibende Statistik.

Tab.I: Sportartspezialisierung (Disz.), Pferdsprung-Zensur (Note) und 100m-Sprintzeit (Zeit) von 20 Sportstudenten (Vp)

Vp.	Disz.	Note	Zeit [sec]
A	Fußb.	3,0	13,0
B	L.athlet.	4,0	11,4
C	Fußb.	2,0	12,7
D	L.athlet.	3,3	12,6
E	Fußb.	4,3	12,4
F	Fußb.	4,0	12,9
G	Fußb.	3,7	13,1
H	L.athlet.	5,0	12,1
I	Fußb.	2,7	12,3
J	L.athlet.	3,3	11,8
K	Turnen	1,7	12,9
L	L.athlet.	2,0	12,8
M	Fußb.	4,3	12,6
N	L.athlet.	3,7	12,3
O	Turnen	2,3	13,0
P	Fußb.	4,0	13,1
Q	Fußb.	3,7	12,2
R	Turnen	1,3	13,5
S	L.athlet.	3,3	12,0
T	Fußb.	3,7	11,9

Datensammlungen lassen sich durch 3 unterschiedliche Möglichkeiten aufbereiten:

1. durch Tabellen,
2. durch Diagramme,
3. durch statistische Kenngrößen.

Während Tabellen und Diagramme die Funktion haben, die Daten übersichtlich und anschaulich darzustellen, z.B. um einen ersten Eindruck von Unterschieden und Zusammenhängen zu bekommen, erfüllen die **statistischen Kenngrößen** den Zweck, die Stichprobe zu charakterisieren und Grundlagen für weitere Berechnungen und Interpretationen zu liefern.

Tabellen werden mit römischen Ziffern numeriert und erhalten eine Überschrift mit der Angabe aller zum Verständnis der Tabelle notwendigen Informationen, insbesondere
- die Ausprägung welcher Variablen dargestellt ist,
- um welche Kenngrößen (Häufigkeit, Mittelwert.....) es sich handelt,
- welche Abkürzungen benutzt werden,
- an welcher Stichprobe die Daten erhoben wurden,
- woher die Daten stammen (Quelle, Literatur .....).

Ersatzweise oder zur Ergänzung können **Diagramme und Abbildungen** angefertigt werden, die mit arabischen Ziffern zu numerieren sind und eine Unterschrift (Legende) mit entsprechend vollständigen Informationen erhalten. Weitere Regeln zur Anfertigung von Abbildungen und Tabellen sind der Literatur und dem Skript “Empfehlungen zur Anfertigung sportwissenschaftlicher (empirischer) Arbeiten” zu entnehmen.

Beispiel: Tabelle I zeigt die Urliste des hier zugrundegelegten Beispiels einer Pilotstudie.

5.2 Die Häufigkeitsverteilung

Die Häufigkeitsverteilung stellt die grundlegenste Art der Stichproben-Charakterisierung dar. Sie beschreibt, wie häufig die einzelnen Merkmalsausprägungen auftreten. Die Häufigkeitsverteilung kann die **absoluten Häufigkeiten** oder die **relativen Häufigkeiten** (in % der Gesamtstichprobe) wiedergeben.

Beispiel: Das erste Merkmal des Erhebungsbeispiels, die Sportartspezialisierung, zeigt 3 mögliche Ausprägungen. Um die Häufigkeiten in den Ausprägungen des Merkmals “Sportartspezialisierung” bestimmen zu können, sind die Einzelbeobachtungen der Urliste (Tab. I) alphabetisch nach den Sportartspezialisierungen (Spalte B) zu sortieren (Tab. II). Jetzt zeigt sich, daß die Ausprägung “Fußball” die absolute Häufigkeit 10, die Ausprägung “Leichtathletik” die absolute Häufigkeit 7 und die Ausprägung “Turnen” die absolute Häufigkeit 3 aufweisen.

Gelegentlich kann es von Interesse sein, die relativen Häufigkeiten (in Prozent der Anzahl der Gesamtstichprobe) zu ermitteln. Diese sind in Tabelle II mit aufgeführt.

Zur besseren Überschaubarkeit der Häufigkeitsverteilung lassen sich Häufigkeitsdiagramme benutzen (Abb. 1-3).

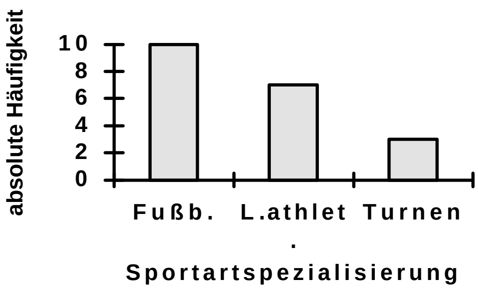

Abb. 1: Häufigkeitsverteilung in der Sportartspezialisierung von Sportstudenten (n = 20); Quelle: Urliste Tab. I

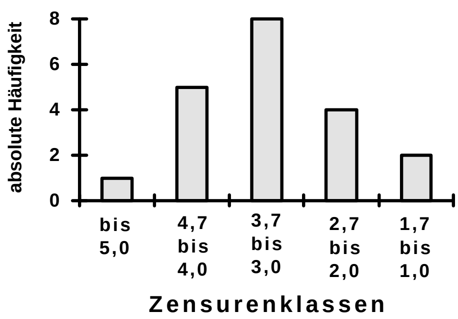

Abb. 2: Häufigkeitsverteilung der Pferdsprung-Zensuren von Sportstudenten (n = 20); Quelle: Urliste Tab. I

Besitzt ein Merkmal viele unterschiedliche Ausprägungen, empfiehlt es sich, diese zu **Merkmalsklassen** zusammenzufassen.

Beispiel: Die Pferdsprungzensuren kommen in 11 verschiedenen Ausprägungen vor, *wobei 6 Ausprägungen nur je 1mal auftreten. Zur besseren Überschaubar-*

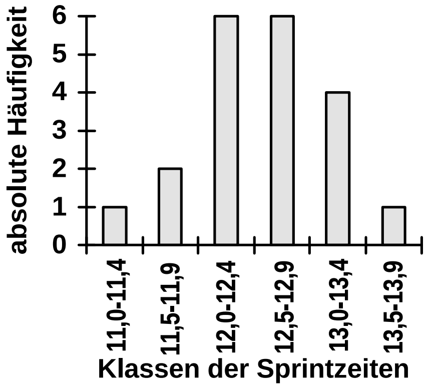

Abb. 3: Häufigkeitsverteilung der 100m-Sprintzeiten von Sportstudenten (n = 20); Quelle: Urliste Tab. **I**

Tab.II: Sportartspezialisierung (Disz.), Pferdsprung-Zensur (Note), 100m-Sprintzeit (Zeit) in sec, absolute und relative Häufigkeit der Sportartspezialisierung von 20 Sportstudenten(Vp) (Quelle: Urliste Tab. I)

	A	B	C	D
1	Vp.	Disz.	Note	Zeit
2	C	Fußb.	2,0	12,7
3	I	Fußb.	2,7	12,3
4	A	Fußb.	3,0	13,0
5	G	Fußb.	3,7	13,1
6	Q	Fußb.	3,7	12,2
7	T	Fußb.	3,7	11,9
8	F	Fußb.	4,0	12,9
9	P	Fußb.	4,0	13,1
10	E	Fußb.	4,3	12,4
11	M	Fußb.	4,3	12,6
12	L	L.athlet.	2,0	12,8
13	D	L.athlet.	3,3	12,6
14	J	L.athlet.	3,3	11,8
15	S	L.athlet.	3,3	12,0
16	N	L.athlet.	3,7	12,3
17	B	L.athlet.	4,0	11,4
18	H	L.athlet.	5,0	12,1
19	R	Turnen	1,3	13,5
20	K	Turnen	1,7	12,9
21	O	Turnen	2,3	13,0

absolute Häufigkeit	Fußb.	***10***	exc	=anzahl2(B2:B11)
	L.athlet.	***7***	exc	=anzahl2(B12:B18)
	Turnen	***3***	exc	=anzahl2(B19:B21)
	Summe	***20***		
relative Häufigkeit	Fußb.	***0,50***		
	L.athlet.	***0,35***		
	Turnen	***0,15***		

keit können 5 Merkmalsklassen gebildet werden, nämlich die Merkmalsklassen "bis 5,0", "4,7 bis 4,0", "3,7 bis 3,0", "2,7 bis 2,0" und "1,7 bis 1,0". Wie sich jetzt die einzelnen Merkmalsausprägungen auf die Merkmalsklassen verteilen, zeigt das Häufigkeitsdiagramm Abb. 2. Das Merkmal "100m-Sprintzeit" tritt sogar in 15 verschiedenen Ausprägungen auf, wobei nur die Ausprägungen 12.3, 12.6, 12.9, 13.0 und 13.1 je zweimal erscheinen. Bildet man nun 6 Merkmalsklassen, ergibt sich das in Abb. 3 dargestellte Häufigkeitsdiagramm.

Bei psychologischen, soziologischen und biologischen Merkmalen zeigen die Häufigkeitsverteilungen in der Regel eine typische symmetrische "Glockenform", d.h. die mittleren Ausprägungen treten relativ häufig auf, nach den Extremwerten hin nehmen die Häufigkeiten immer mehr ab (Abb. 2, 3 und 4): Diese Art der Verteilung heißt **Normalverteilung** (**GAUSS**sche Normalverteilung).

Beispiel: Sowohl das Häufigkeitsdiagramm der Pferdsprungnoten (Abb. 2) als auch das der 100m-Sprintzeiten (Abb. 3) zeigen eine annähernd symmetrische Glockenform. Aber nur die 100m-Sprintzeiten dürfen als normalverteilt gelten, nicht aber die Pferdsprungnoten. Das liegt daran, daß die Feststellung der Normalverteilung voraussetzt, daß die Abstände zwischen den Merkmalsklassen gleich sind. Das trifft aber nur für intervallskalierte Daten (Kap. 3.2), also hier nur für die Sprintzeiten, zu.

5.3 Statistische Kenngrößen, Gliederung

Statistische Kenngrößen sind (meistens berechnete) Werte zur Kennzeichnung bzw. Charakterisierung von Stichproben, z.B. um diese von anderen Stichproben unterscheiden zu können.

Statistische Kenngrößen sollen Stichproben nach zwei Kriterien beschreiben:

a) Wie ist die untersuchte Variable in einer Stichprobe im Durchschnitt (in der "**zentralen Tendenz**") ausgeprägt? Entsprechende Kenngrößen sind der **Mittelwert**, der **Median** und der **Modalwert**.

b) Über welchen Bereich **streuen** die Ausprägungen der untersuchten Variablen in der Stichprobe. Entsprechende Kenngrößen sind u.a. die **Spannweite** und die **Standardabweichung.**

(Kenngrößen zur Beschreibung der ***Form*** einer Verteilung sollen hier nicht behandelt werden; siehe dazu die Literatur.)

5.4 Kenngrößen für die zentralen Tendenzen

Die geläufigste Kenngröße zur Charakterisierung von Verteilungen ist der **Mittelwert** (das arithmetische Mittel), der bei allen möglichen (und unmöglichen, sprich: nicht zulässigen) Gelegenheiten Anwendung findet. Da er nach einer mathematischen Formel **berechnet** wird (Summe der Ausprägungen aller Meßwerte, dividiert durch die Anzahl der Meßwerte), darf er nur bei intervallskalierten Daten (Kap. 3.2) ermittelt werden.

Beispiel: Im vorliegenden Beispiel der Pilotstudie tritt nur die Variable "Sprintzeit" intervallskaliert auf. Aus diesem Grunde ist es, strenggenommen, nur hier er

Tab.III: Sportartspezialisierung (Disz.), Pferdsprung-Zensur (Note), 100m-Sprintzeit (Zeit) in sec und Mittelwerte(MW) der Sprintzeiten bei insgesamt 20 Sportstudenten (Vp) (Quelle: Urliste Tab. I)

	A	B	C	D		
1	Vp.	Disz.	Note	Zeit		
2	C	Fußb.	2,0	12,7		
3	I	Fußb.	2,7	12,3		
4	A	Fußb.	3,0	13,0		
5	G	Fußb.	3,7	13,1		
6	Q	Fußb.	3,7	12,2		
7	T	Fußb.	3,7	11,9		
8	F	Fußb.	4,0	12,9		
9	P	Fußb.	4,0	13,1		
10	E	Fußb.	4,3	12,4		
11	M	Fußb.	4,3	12,6		
12	L	L.athlet.	2,0	12,8		
13	D	L.athlet.	3,3	12,6		
14	J	L.athlet.	3,3	11,8		
15	S	L.athlet.	3,3	12,0		
16	N	L.athlet.	3,7	12,3		
17	B	L.athlet.	4,0	11,4		
18	H	L.athlet.	5,0	12,1		
19	R	Turnen	1,3	13,5		
20	K	Turnen	1,7	12,9		
21	O	Turnen	2,3	13,0		
		Gesamt	MW:	***12,5***	exc:	=mittelwert(D2:D21)
		Fußb.	MW:	***12,6***	exc:	=mittelwert(D2:D11)
		L.athlet.	MW:	***12,1***	exc:	=mittelwert(D12:D18)
		Turnen	MW:	***13,1***	exc:	=mittelwert(D19:D21)

laubt, Mittelwerte zu berechnen (Tab. III). Sortiert man die Tabelle alphabetisch nach der Sportartenspezialisierung, lassen sich mit Hilfe des in Tab. III wiedergegebenen EXCEL-Befehls leicht die Mittelwerte der Sprintzeiten innerhalb der Sportartenklassen berechnen (Tab. III).

Formel zur Berechnung des "arithmetischen Mittels" (MW):

$$MW = \frac{1}{N} * \sum x_i$$

Hier sind N die Anzahl der Meßwerte und x_i die einzelnen Meßwerte von 1 bis N.

Die Meßwerte für die Ausprägung der Variablen “Pferdsprung-Zensur” stellen zwar auch Zahlen dar, aber die Bedingungen für intervallskalierte Daten (Kap. 3.2) sind nicht erfüllt. Auch wenn es in der Praxis üblich ist, z.B. “Durchschnittsnoten” zu berechnen, hat eine solche Kenngröße statistisch keinen Aussagewert. Bei ordinalskalierten Daten wird die zentrale Tendenz durch den **Median** bestimmt. Das ist derjenige (mittlere) Meßwert, oberhalb und unterhalb dessen gleich viele Beobachtungswerte liegen.

Beispiel: Der Median (MD) der Pferdsprung-Zensuren beträgt 3,5 (EXCEL-Befehl siehe Tab. IV). Um das Zustandekommen dieses Wertes einzusehen, sollte man die Daten der Tabelle I nach aufsteigenden Pferdsprung-Zensuren sortieren, wie in Tab. IV dargestellt. Da bei einer geraden Anzahl von Meßwerten (hier: 20) kein “mittlerer” Meßwert existiert, wird als Median das Mittel aus den beiden mittleren Meßwerten (hier: aus dem 10. und 11. Meßwert, nämlich 3,3 und 3,7) genommen.

(Zur Anwendung des EXCEL-Befehls zur Berechnung des Medians ist ein Sortieren der Datenreihe nicht erforderlich! Dies wurde in Abb. IV nur deshalb vollzogen, um das Zustandekommen des Medians zu verdeutlichen. Zusätzlich ist in Tabelle IV zum Vergleich der Mittelwert der Pferdsprung-Zensuren angegeben. Er beträgt 3,27, hat aber aus den oben genannten Gründen keine statistische Berechtigung. Der Median der Sprintzeiten beträgt im übrigen 12,6.)

Tab.IV: Sportartspezialisierung (Disz.), Pferdsprung-Zensur (Note), 100m-Sprintzeit (Zeit) in sec und Median (MD) der Pferdsprung-Zensur von 20 Sportstudenten (Quelle: Urliste Tab. I)

	A	B	C	D		
1	Vp.	Disz.	Note	Zeit		
2	R	Turnen	1,3	13,5		
3	K	Turnen	1,7	12,9		
4	C	Fußb.	2,0	12,7		
5	L	L.athlet.	2,0	12,8		
6	O	Turnen	2,3	13,0		
7	I	Fußb.	2,7	12,3		
8	A	Fußb.	3,0	13,0		
9	D	L.athlet.	3,3	12,6		
10	J	L.athlet.	3,3	11,8		
11	S	L.athlet.	3,3	12,0		
12	G	Fußb.	3,7	13,1		
13	Q	Fußb.	3,7	12,2		
14	T	Fußb.	3,7	11,9		
15	N	L.athlet.	3,7	12,3		
16	F	Fußb.	4,0	12,9		
17	P	Fußb.	4,0	13,1		
18	B	L.athlet.	4,0	11,4		
19	E	Fußb.	4,3	12,4		
20	M	Fußb.	4,3	12,6		
21	H	L.athlet.	5,0	12,1		
		MD	***3,5***		exc:	= median(C2:C21)
		[MW	***3,27***		exc:	=mittelwert(C2:C21)]

Um den Median der Pferdsprung-Zensuren innerhalb der Sportartklassen zu bestimmen, muß die Tabelle nach den Sportarten und (aufsteigend) nach den Pferdsprung-Zensuren sortiert werden. Dies ist in Tab. III geschehen. Hier läßt sich auch ohne Rechenoperationen der Median der Pferdsprung-Zensuren der Sportartklasse "Fußball" leicht erkennen. Er beträgt 3,7. Der Median der Zensuren der Leichtathleten beträgt 3,3 und derjenige der Turner 1,7.

Die Berechnung des Medians setzt eine aufsteigend (oder absteigende) Reihe von Meßwerten (also eine Ordinalskala) voraus. Aus diesem Grunde ist es nicht möglich, für die Beschreibung der zentralen Tendenzen in der Ausprägung der Variablen "Sportartspezialisierung" den Median (oder gar den Mittelwert) zu benutzen. Hier - bei der Beschreibung der zentralen Tendenzen von Nominalvariablen (Kap. 3.2) - bleibt nur die Benutzung des **Modalwertes** (des **Modus**), der demjenigen Meßwert innerhalb einer Verteilung entspricht, der am häufigsten vorkommt.

Beispiel: Der Modus der Variablen "Sportartspezialisierung" ist die Merkmalsausprägung "Fußball", wie Tab. II, die nach der Sportartspezialisierung sortiert ist, zu entnehmen ist. D.h., durch die Berechnung der absoluten Häufigkeiten der Meßwert-Klassen "Fußball", "Leichtathletik" und "Turnen" ergibt sich der Modalwert als größter Wert der absoluten Häufigkeit von selbst.

5.5 Kenngrößen für die Streuung

Zusätzlich zur Charakterisierung der "zentralen Tendenzen" der Häufigkeitsverteilung innerhalb einer Stichprobe durch Mittelwert, Median und Modus kann es sinnvoll sein, eine Stichprobe auch dadurch zu beschreiben, zwischen welcher maximalen bzw. minimalen Ausprägung die einzelnen Meßwerte angeordnet sind, d.h. **zwischen welchen Extremwerten die Meßwerte** "**streuen"**. Die zutreffende Kenngröße (Streuungsmaß) ist die **Spannweite** (Streubereich, Variationsbreite, Range). Diese ergibt sich, wenn der kleinste in der Stichprobe auftretende Meßwert vom größten abgezogen wird.

Beispiel: Da in Tab. IV die Pferdsprung-Zensuren nach der Größe sortiert sind, lassen sich die beiden extremen Meßwerte (1,3 und 5,0) leicht erkennen. Die Spannweite berechnet sich dann gemäß der Formel:
Spannweite = 5,0 - 1,3 = 3,7.

Die Spannweite der Sprintzeiten läßt sich nicht auf den ersten Blick erkennen. Wenn man sich ein Sortieren der Sprintzeiten ersparen will, lassen sich mit entsprechenden EXCEL-Befehlen der größte und kleinste Meßwert heraussuchen und mit ihrer Hilfe die Spannweite bestimmen (Tab. V).

Für intervallskalierte Daten ist die Spannweite kein ergiebiger Kennwert zur Charakterisierung der Streuung der Meßwerte, weil nicht berücksichtigt werden kann, wie die Meßwerte zwischen den Extremwerten angeordnet sind. Sie könnten z.B. zwischen den Extremwerten relativ gleichmäßig verteilt sein (Abb. 4, links), sie könnten sich aber auch um den Mittelwert drängen (Abb. 4, rechts). Dies verdeutlicht Abb. 4 mit zwei Normalverteilungen, bei denen sich trotz gleicher Spannweite (16-4 = 12), gleicher Gesamtzahl der Meßwerte (= 104) und gleichem Mittelwert (= 10) die Meßwerte unterschiedlich dicht um den Mittelwert (gestrichelte Senkrechte) scharen und dadurch dem Kurvenverlauf eine “schlankere” bzw. “gedrungenere” Form geben.

Tab.V: Sportartspezialisierung (Disz.), Pferdsprung-Zensur (Note), 100m-Sprintzeit (Zeit) in sec, Berechnung der Spannweite und Standardbweichung (s) der Sprintzeit von 20 Sportstudenten (Vp) (Quelle: Urliste Tab. I)

	A	B	C	D		
1	Vp.	Disz.	Note	Zeit		
2	C	Fußb.	2,0	12,7		
3	I	Fußb.	2,7	12,3		
4	A	Fußb.	3,0	13,0		
5	G	Fußb.	3,7	13,1		
6	Q	Fußb.	3,7	12,2		
7	T	Fußb.	3,7	11,9		
8	F	Fußb.	4,0	12,9		
9	P	Fußb.	4,0	13,1		
10	E	Fußb.	4,3	12,4		
11	M	Fußb.	4,3	12,6		
12	L	L.athlet.	2,0	12,8		
13	D	L.athlet.	3,3	12,6		
14	J	L.athlet.	3,3	11,8		
15	S	L.athlet.	3,3	12,0		
16	N	L.athlet.	3,7	12,3		
17	B	L.athlet.	4,0	11,4		
18	H	L.athlet.	5,0	12,1		
19	R	Turnen	1,3	13,5		
20	K	Turnen	1,7	12,9		
21	O	Turnen	2,3	13,0		
22						
23	Maximale Sprintzeit:			***13,5***	exc	=max(D2:D21)
24	Minimale Sprintzeit:			***11,4***	exc	=min(D2:D21)
25	Spannweite:			***2,1***	exc	=D23-D24
26	Standardabweichung **s =**			***0,53***	exc	=stabw(D2:D21)

Ein Kennwert für die unterschiedliche Form der Normalverteilungskurve, d.h. für die unterschiedliche Dichte der Meßwerte um den Mittelpunkt (also ein Streuungsmaß) ist die **Standardabweichung (s).** Sie bestimmt, bis zu welchem Skalenwert 34,13 % aller Beobachtungsfälle vom Mittelwert aus nach unten (MW minus s) bzw. nach oben (MW plus s) verteilt sind (Abb. 4). Somit liegen 68,26 Prozent aller Beobachtungsfälle zwischen MW minus s und MW plus s.

Bei “gedrungenen” Normalverteilungen ist die Standardabweichung größer (in Abb. 4 beträgt sie für die linke Kurve s_1 = 2,27), bei “schlankeren” Normalverteilungen ist die Standardabweichung kleiner (in Abb. 4 beträgt sie für die rechte Kurve s_2 = 1,84).

Beispiel: Die Standardabweichung der 100m-Sprintzeiten der untersuchten 20 Sportstudenten beträgt 0,53 sec (Tab. V). Das bedeutet: 68,26 Prozent *aller Meßwerte der Sprintzeit liegen zwischen MW minus s (nämlich 12,5 sec - 0,53 sec = **11,97** sec) und MW plus s (nämlich 12,5 sec + 0,53 sec = **13,03** sec). Innerhalb der Gruppe der Fußballer beträgt die Standardabweichung s = 0,413 sec, bei den Leichtathleten s = 0,476 sec und bei den Turnern s = 0,321 sec.*

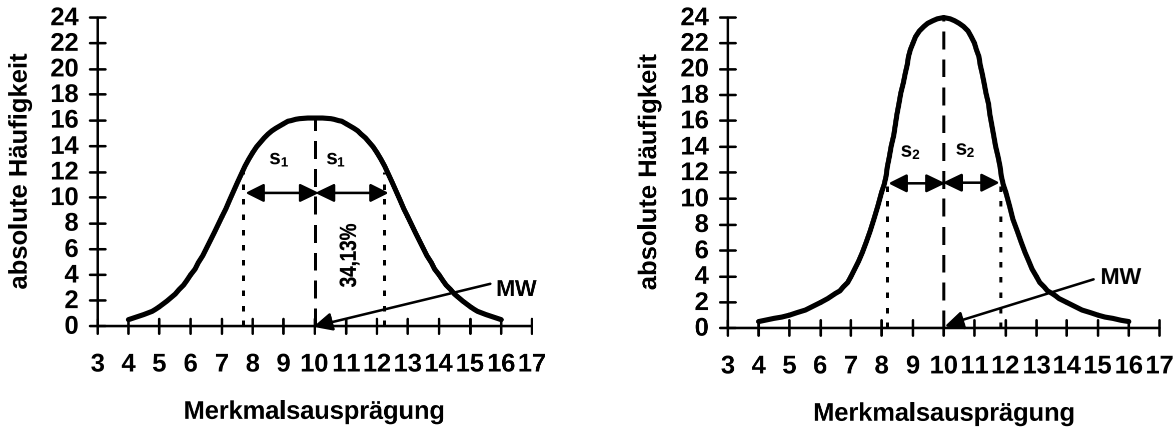

Abb 4: Beispiel zweier angenäherter Normalverteilungen mit gleicher Spannweite, gleichem Mittelwert (MW) und gleicher Anzahl von Meßwerten, aber unterschiedlicher Standardabweichung (s1 und s2)

Formel zur Berechnung der Standardabweichung (s):

$$s = \sqrt{\frac{1}{N-1} * \sum (x_i - MW)^2}$$

Hier sind N die Anzahl der Meßwerte, xi die einzelnen Meßwerte und MW der Mittelwert.

5.6 Nutzung der Kennwerte zur Beschreibung der Unterschiede von Stichproben

Die Kennwerte zur Charakterisierung der zentralen Tendenzen (Mittelwert, Median, Modus) und zur Charakterisierung der Streuung (Spannweite, Standardabweichung) werden zwar allgemein dazu benutzt, die Ausprägung eines Merkmals in einer Stichprobe (Personengruppe, Verteilung) zu beschreiben, häufiger aber werden sie eingesetzt, um deutlich zu machen, wie sich zwei (oder mehrere) Stichproben im Hinblick auf ein Merkmal (eine Variable) **unterscheiden,** z.B. um die Richtigkeit von Unterschiedshypothesen (Kap. 2) nachzuprüfen. Dazu gehört auch die Notwendigkeit, zu zeigen, wie sich die Ausprägung eines Merkmals in einer Stichprobe durch eine Behandlung (z.B. ein Training) ändert, d.h. welcher Unterschied sich in der Ausprägung des Merkmals vor bzw. nach der Behandlung ergibt.

Beispiel: Innerhalb der untersuchten 20 Sportstudenten erreichen die Turner bessere Pferdsprungzensuren (Median 1,7) als die Fußballer (Median 3,7). Dem-

gegenüber zeigen die Leichtathleten bessere Sprintzeiten (Mittelwert 12,1 sec) als die Turner (Mittelwert 13,1 sec).

Die Spannweite der Pferdsprungzensuren beträgt bei den Fußballern 2,3 Notenstufen, bei den Turnern nur 1,0 Notenstufen. Die Standardabweichung der Sprintzeiten der Leichtathleten ist größer (0,48 sec) als die der Turner (0,32 sec).

5.7 Eine Kenngröße zur Beschreibung von Zusammenhängen

Häufig ist es notwendig, zu zeigen, wie in einer Stichprobe die Ausprägungen von zwei Merkmalen **zusammenhängen** (= “korrelieren”).

Beispiel: Jedem ist bekannt, daß die Merkmale “Körpergröße” und “Körpergewicht” zusammenhängen, d.h. je größer Menschen sind, desto mehr wiegen sie. In gleicher Weise hängen die Merkmale “Körpergröße” und “Hochsprungleistung” zusammen, also: die Hochsprungleistung steigt mit der Körpergröße.
Im vorliegenden Untersuchungsbeispiel wurde vermutet, daß das Merkmal “Pferdsprung-Zensur” von dem Merkmal “Sprintleistung” abhängt.

Wie aus den Beispielen deutlich wird, müssen bei der Prüfung von Zusammenhängen von jeder Versuchsperson zwei Meßwerte vorliegen, nämlich ein Meßwert über die Ausprägung der Variablen a und ein Meßwert über die Ausprägung der Variablen b. (Man spricht auch von einem Meßwert**paar**). Man erwartet beim Vorliegen eines Zusammenhanges, daß zusammen mit der Änderung (z.B. Vergrößerung) der Ausprägung von Merkmal a sich entsprechend auch die Ausprägung von Merkmal b ändert (vergrößert). Wenn das stets der Fall ist, spricht man von einem **perfekten** Zusammenhang. Aber Zusammenhänge brauchen nicht immer “hundertprozentig” zu sein.

Beispiel: Untersucht man eine Gruppe Personen im Hinblick auf Körpergröße und Hochsprungleistung wird in der Regel der Größere auch höher springen können. Hin und wieder wird es jedoch vorkommen, daß eine kleinere Person höher springen kann als eine andere größere Person oder daß zwei gleich große Personen unterschiedlich hoch springen. In einem solchen Fall ist der Zusammenhang nicht so “eng” wie bei einer “hundertprozentigen” bzw. “perfekten” Korrelation.

Um den Grad des Zusammenhanges (der Korrelation) von zwei Merkmalen bestimmen zu können, benötigt man also einen Kennwert. Dies ist der **Korrelationskoeffizient r** (bzw. R, siehe unten).

Bei einem perfekten **“positiven”** Zusammenhang nimmt der Korrelationskoeffizient den Wert r = 1 an (je größer die Ausprägung der Variablen a, desto größer die Ausprägung der Variablen b) bzw. bei einer **“negativen”** Korrelation (je größer a, desto kleiner b) den Wert r = -1. Fehlt jeglicher Zusammenhang zwischen den beiden untersuchten Variablen, geht der Korrelationskoeffizient r gegen Null. Aus der Höhe des Wertes zwischen r = 0 und r = 1 (bzw. zwischen r = 0 und r = -1) läßt sich auf die Stärke des Zusammenhanges schließen.

Die passende bildliche Darstellung für den Zusammenhang zweier Variablen (x und y oder a und b) ist das x-y-Diagramm, bei dem man die Ausprägung des einen Merkmals (x oder a) auf der Abszisse (der x-Achxe), die Ausprägung des anderen Merkmals (y oder b) auf der Ordinate (der y-Achse) abträgt (Abb. 5), so daß jedes Merkmalspaar einen Punkt im Diagramm zugeordnet bekommt. Bei einem perfekten Zusammenhang (r = 1) bilden die Punkte eine aufsteigende oder (bei r = -1) eine absteigende Gerade, bei einem weniger engen Zusammenhang eine schräg auf- oder absteigende Punkt**wolke** (Abb. 5).

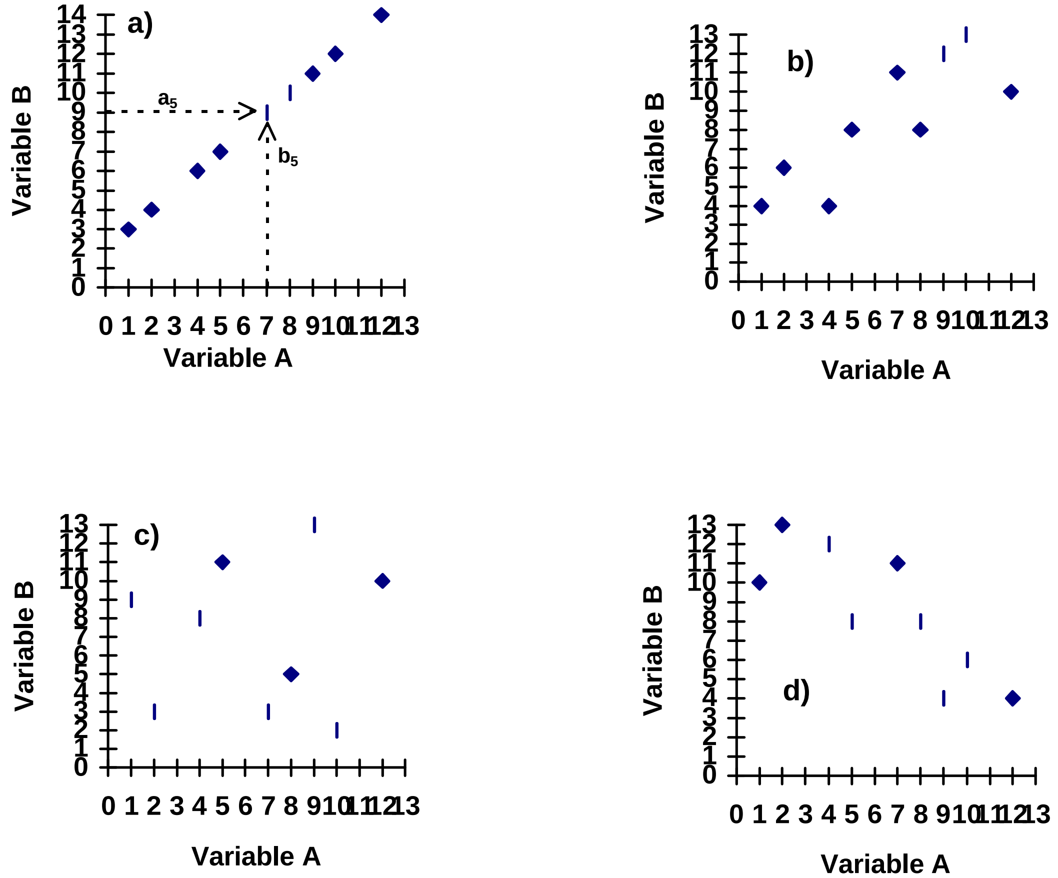

Abb. 5: x-y-Diagramme von Merkmalszusammenhängen.
a) nahezu perfekter Zusammenhang: r = 0,976. b) hoher positiver Zusammenhang: r = 0,824
c) nahezu kein Zusammenhang: r = 0,064. d) hoher negativer Zusammenhang r = -0,819

Wie eng nun der Zusammenhang zwischen Variablen einer Stichprobe ist, d.h. welchen Wert der Korrelationskoeffizient annimmt, muß mit Hilfe der einzelnen Meßwertpaare berechnet werden. Die beiden bedeutendsten Korrelationsberechnungen stellen die **“Produkt-Moment-Korrelation”** und die **“Rangkorrelation”** dar. Erstere darf nur bei intervallskalierten Daten angewendet werden, letztere - wie der Begriff ankündigt - bei ordinalskalierten Daten.

Beispiel: Die Erhebung über die Pferdsprung-Zensuren und 100m-Sprintzeiten wurde aufgrund der Vermutung geplant, daß Personen, die schneller laufen können, besser über das Pferd springen können und deshalb bessere Pferd-

sprung-Zensuren bekommen. Es wird somit ein Zusammenhang zwischen der Variablen Pferdsprung-Zensur und der Variablen 100m-Sprintzeit vermutet.

Da eine der Variablen (Note) ordinalskaliert, die andere (Zeit) intervallskaliert vorliegt, darf die Produkt-Moment-Korrelation zwischen beiden Variablen nicht berechnet werden (obwohl dies - allerdings nur zum Zweck des Vergleiches und der Verdeutlichung - in Tab. VI geschehen ist), sondern es ist der Rangkorrelationskoeffizient zu berechnen. Er stellt mit ***R = -0,344*** *einen eher mäßigen Zusammenhang zwischen der Pferdsprungzensur und der 100m-Sprintzeit fest und hat darüber hinaus ein negatives Vorzeichen. Letzteres deutet auf eine negative Korrelation hin, die besagt, daß mit* ***zu****nehmenden Pferdsprung-Zensuren die 100m-Sprintzeit* ***ab****nimmt, erkennbar an der von links nach rechts abfallenden Punktwolke in Abb. 6. Der Wert des Korrelationskoeffizienten von R = -0,344 liegt näher dem Wert Null als dem Wert -1, so daß man eher von einem* ***schwachen*** *negativen Zusammenhang der beiden untersuchten Parameter sprechen muß.*

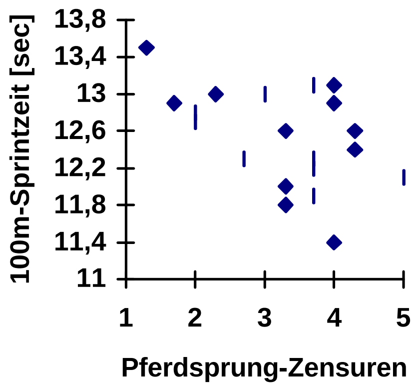

Abb. 6: Zusammenhang zwischen Pferdsprung-Zensuren und 100m-Sprintzeit bei 20 Sportstudenten. R = - 0,367

Zur Berechnung des Rangkorrelationskoeffizienten bildet man aus den beiden Meßreihen je eine **Rangreihe**: Die höchste Zensur bekommt den Rangplatz 1, die zweithöchste den Rangplatz 2 usw. (Tab. VI, Spalte E). Treten 2 oder mehrere gleiche Noten auf, bekommen sie als Rang den Mittelwert derjenigen Rangplätze, die sie besetzen. Die Vpn C und L , die beide die Zensur 2,0 im Pferdsprung erhalten, teilen sich den 3. und 4. Rangplatz und erhalten deshalb den Rangplatz 3,5. Die Personen T, Q, N und G, die alle die Zensur 3,7 besitzen, teilen sich die Rangplätze 11 bis 14. Deshalb erhalten sie den mittleren Rangplatz 12,5. In gleicher Weise wird mit der zweiten Meßreihe (Zeiten) verfahren (Tab. VI, Spalte F). Danach wird die Differenz der beiden Rangplätze (Tab. VI, Spalte G) quadriert (Tab. VI, Spalte H) und die Summe der quadrierten Rangplatzdifferenzen gebildet. Der Rangkorrelationskoeffizient berechnet sich nach folgender Formel:

$$R = 1 - \frac{6 * \Sigma D^2}{N * (N^2 - 1)}$$

wobei ΣD^2 die Summe der quadrierten Rangplatzdifferenzen, und N die Anzahl der untersuchten Personen darstellt.

Diese Berechnung kann man sich dadurch erleichtern, daß man die Rangplätze durch EXCEL-Befehle (Tab. VI) errechnen läßt. Man braucht dazu nicht einmal die Datenreihe zu sortieren. Doch **Achtung!** Beim EXCEL-Befehl bekommen gleiche Meßwerte nicht den mittleren Rangplatz, sondern den niedrigsten gemeinsamen Rangplatz (Tab. VI unten). Das hat zur Folge, daß dann, wenn in einer Datenreihe viele Meßwerte auf nur wenige Rangplätze verteilt werden, der Rangkorrelationskoeffizient niedriger ausfällt.

Tab.VI: Sportartspezialisierung (Disz.), Pferdsprung-Zensur (Note) und 100m-Sprintzeit (Zeit) in sec von 20 Sportstudenten (Vp). Berechnung des Rangkorrelationskoeffizienten (R) zwischen "Note" und "Zeit". R1: Rangreihe der Noten. R2: Rangreihe der Zeiten. Untere Tabelle: Ermittlung der Rangplätze mit Hilfe des EXCEL-Befehls. (Quelle: Urliste Tab. I)

	A	B	C	D	E	F	G	H
1	Vp.	Disz.	Note	Zeit	R_1	R_2	R_1-R_2	$(R_1-R_2)^2$
2	R	Turnen	1,3	13,5	1	20	-19	361
3	K	Turnen	1,7	12,9	2	14,5	-12,5	156,25
4	C	Fußb.	2,0	12,7	3,5	12	-8,5	72,25
5	L	L.athlet.	2,0	12,8	3,5	13	-9,5	90,25
6	O	Turnen	2,3	13,0	5	16,5	-11,5	132,25
7	I	Fußb.	2,7	12,3	6	7,5	-1,5	2,25
8	A	Fußb.	3,0	13,0	7	16,5	-9,5	90,25
9	J	L.athlet.	3,3	11,8	9	2	7	49
10	S	L.athlet.	3,3	12,0	9	4	5	25
11	D	L.athlet.	3,3	12,6	9	10,5	-1,5	2,25
12	T	Fußb.	3,7	11,9	12,5	3	9,5	90,25
13	Q	Fußb.	3,7	12,2	12,5	6	6,5	42,25
14	N	L.athlet.	3,7	12,3	12,5	7,5	5	25
15	G	Fußb.	3,7	13,1	12,5	18,5	-6	36
16	B	L.athlet.	4,0	11,4	16	1	15	225
17	F	Fußb.	4,0	12,9	16	14,5	1,5	2,25
18	P	Fußb.	4,0	13,1	16	18,5	-2,5	6,25
19	E	Fußb.	4,3	12,4	18,5	9	9,5	90,25
20	M	Fußb.	4,3	12,6	18,5	10,5	8	64
21	H	L.athlet.	5,0	12,1	20	5	15	225
22							Summe:	1787

Rangkorrelationskoeffizient: R = ***- 0,344***

exc = 1-6*H22/20/(20^2-1)

Produkt-Moment-Korrelationskoeffizient: r = ***- 0,480***

exc =korrel(C2:C21;D2:D21)

exc = rang(C2;C$2:C$21;1)

1	Vp.	Disz.	Note	Zeit	R_1	R_2	$(R_1-R_2)^2$
2	R	Turnen	1,3	13,5	***1***	20	361
3	K	Turnen	1,7	12,9	***2***	14	144
4	C	Fußb.	2,0	12,7	***3***	12	81
5	L	L.athlet.	2,0	12,8	***3***	12	81
6	O	Turnen	2,3	13,0	***5***	16	121
7							

Zur Interpretation der Stärke von Zusammenhängen ist der **Wert** des Korrelationskoeffizienten zu Rate zu ziehen, wie oben schon mehrfach durch die Bezeichnungen “perfekter” oder “schwacher” Zusammenhang deutlich gemacht wurde. Zur generellen Bewertung kann folgende Auflistung gemäß BÖS (1986) dienen:

absoluter Wert von r (bzw. R)	“Höhe” des Zusammenhanges
0,00	kein Zusammenhang
bis 0,39	schwacher Zusammenhang
0,40 bis 0,69	mittlerer Zusammenhang
0,70 bis 0,99	hoher Zusammenhang
1,00	perfekter Zusammenhang

Formel zur Berechnung des Koeffizienten der Produkt-Moment-Korrelation:

$$r = \frac{N * \sum x_i * y_i - \sum x_i * \sum y_i}{\sqrt{\left(N * \sum x_i^2 - \left(\sum x_i\right)^2\right) * \left(N * \sum y_i^2 - \left(\sum y_i\right)^2\right)}}$$

Hier sind N die Anzahl aller Merkmalspaare, x_i die einzelnen Meßwerte der Variablen x und y_i die einzelnen Meßwerte der Variablen y.

5.8 Der Korrelationskoeffizient zur Prüfung der Gütekriterien

Die Korrelationsberechnungen lassen sich auch einsetzen, um die Erfüllung der Gütekriterien empirischer Tests, speziell von Objektivität und Reliabilität (Kap. 4.3), zu überprüfen.

Beispiel: Nehmen wir an, die Pferdsprünge der 20 Versuchspersonen sind während der Prüfung per Videokamera aufgezeichnet worden. Nun will der Versuchsleiter, der die Untersuchung zur Abhängigkeit der Pferdsprung-Zensur von der Sprintfähigkeit durchführt, wissen, ob die Pferdsprung-Zensuren hinreichend objektiv vergeben wurden. Dazu führt er das Videoband einem zweiten Gutachter vor , der - unabhängig vom ersten Gutachter - die Sprünge erneut bewertet. Auf diese Weise erhält der Versuchsleiter für alle Sprünge Meßwertpaare, die er vergleichen kann:

A	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*
3	*4*	*2*	*3,3*	*4,3*	*4*	*3,7*	*5*	*2,7*	*3,3*
3,3	*5*	*2,7*	*2,3*	*3,7*	*3,7*	*3*	*4,3*	*2*	*3*

K	*L*	*M*	*N*	*O*	*P*	*Q*	*R*	*S*	*T*
1,7	*2*	*4,3*	*3,7*	*2,3*	*4*	*3,7*	*1,3*	*3,3*	*3,7*
2,3	*1,3*	*5*	*2,7*	*3,3*	*5*	*3*	*1*	*2,3*	*3*

Für die beiden Reihen von Pferdsprung-Zensuren wird jetzt als Koeffizient für die Objektivität der Rangkorrelationskoeffizient berechnet, was den Wert

R = 0,81 ergibt. Dies deutet - wie nicht anders zu erwarten war - eine hohe Korrelation an (Kap. 5.7). Trotzdem liegt der Verdacht nahe, daß die Person des Prüfers das Prüfergebnis stark beeinflußt; denn die Pferdsprung-Zensuren differieren teilweise um eine ganze Notenstufe. Es ist also verständlich, daß man für eine annehmbare Objektivität hohe Zusammenhänge von Vergleichsmeßreihen verlangen muß.

Für die Bewertung von Korrelationskoeffizienten innerhalb der Prüfung der Gütekriterien gibt es keine einheitlichen Richtlinien, weil Forscher - je nach Art des Forschungsgebietes, des Meßinstrumentariums, der Problemstellung u.a.m. - unterschiedlich hohe Anforderungen an die Güte ihrer Methoden stellen. In Anlehnung an WILLIMCZIK (1977) soll hier die folgende Übersicht als Bewertungsgrundlage dienen:

Koeffizient r bzw R	**Bewertung**
kleiner 0,70	fraglich
0,70 bis 0,79	gering
0,80 bis 0,89	annehmbar
0,90 bis 0,94	gut
ab 0,95	sehr gut

Korrelationkoeffizienten zur Überprüfung der Gütekriterien sollten mindestens im annehmbaren Bereich liegen, also mindestens den Wert 0,80 erreichen. Der im obigen Beispiel produzierte Objektivitätskoeffizient liegt mit R = 0,81 so gerade auf einem annehmbaren Niveau, so daß Zweifel an der Objektivität (und - zwangsläufig damit verbunden - auch an der Reliabilität) des hier praktizierten Bewertungsverfahrens nach wie vor berechtigt sind. Als Konsequenz wäre zu überlegen, wie sich der Grad der Objektivität bei Notengebungen verbessern ließe. Einige von mehreren Möglichkeiten wären das Erstellen von Bewertungskriterien, die Schulung der Gutachter in der Anwendung der Bewertungskriterien sowie die Optimierung und Standardisierung der Beobachtungsbedingungen.

Beim Einsatz physikalischer Meßsysteme (Waagen mit elektronischer Anzeige, Blutdruckmeßgeräte, elektronische Lichtschranken) ist die Frage nach der Objektivität zweitrangig, da die Person des Versuchsleiters bzw. Diagnostikers keinen (oder nur einen geringen) Einfluß auf das Meßergebnis nehmen kann. Hier steht statt dessen die Frage nach der Reliabilität im Vordergrund.

Beispiel: Einem Versuchsleiter, der die Sprintfähigkeit von Studenten überprüfen will, stehen zwar elektronische Lichtschranken zur Verfügung, aber - da er seine Tests in einer Sporthalle durchführen muß - nur eine Laufstrecke von 20m, wobei die letzten 15m als Meßstrecke dienen. Die Frage ist, ob in dieser Situation das Ergebnis die zu messende Variable (Sprintfähigkeit) hinreichend genau abbildet.

Das einfachste Verfahren zur Bestimmung der Reliabilität ist die Testwiederholung (ein Verfahren, das aber nicht in jeder Situation angewendet werden darf, siehe Literatur), wobei die beiden Meßreihen miteinander korreliert werden.

Beispiel: Im vorliegenden Fall läßt der Versuchsleiter - nach einer entsprechenden Aufwärmphase - die Strecke aus einer aufrechten Startposition einmal durchlaufen und wiederholt den Test nach einer Erholungspause von 5 min. Die folgende Zusammenstellung zeigt die Sprintergebnisse in Sekunden:

A	***B***	***C***	***D***	***E***	***F***	***G***	***H***	***I***	***J***
1,972	*1,952*	*2,086*	*1,878*	*1,924*	*1,960*	*1,970*	*1,889*	*1,925*	*1,877*
1,941	*1,918*	*2,055*	*1,886*	*1,957*	*1,886*	*1,941*	*1,927*	*1,992*	*1,903*

K	***L***	***M***	***N***	***O***	***P***	***Q***	***R***	***S***	***T***
2,222	*2,031*	*1,847*	*1,992*	*1,938*	*2,157*	*2,006*	*1,945*	*2,001*	*1,897*
2,249	*2,034*	*1,808*	*1,954*	*1,998*	*2,220*	*1,994*	*1,904*	*2,003*	*1,964*

Die gewonnenen Meßreihen produzieren einen Reliabilitätskoeffizienten von r = 0,914, natürlich errechnet mit Hilfe der Produkt-Moment-Korrelation, da intervallskalierte Daten vorliegen. Dieses Ergebnis zeigt, daß trotz der hohen Meßgenauigkeit der elektronischen Lichtschranken die Art der Durchführung des Tests nur eine schwach gute Reliabilität erreicht.

5.9 Meßgenauigkeit und Variabilitätskoeffizient

Der Koeffizient für die Reliabilität läßt zwar eine Bewertung der Meßgenauigkeit zu, kann aber nicht verdeutlichen, wie weit die in einem Meßvorgang ermittelte Ausprägung mit der (unbekannten) tatsächlichen Ausprägung übereinstimmt. Das läßt sich mit Hilfe des **Variabilitätskoeffizienten** errechnen, der festlegt, um wieviel Prozent die einzelnen Messungen durchschnittlich vom Mittelwert abweichen.

Beispiel: Ein Versuchsleiter will mit Hilfe des oben erwähnten Pferdsprungvideos die Kriterien zur Benotung von Pferdsprüngen überarbeiten und kommt zu dem Schluß, daß nur derjenige Prüfling die Zensur “sehr gut” erhalten sollte, dessen Körper im Augenblick des Aufstützens mit der Horizontalen einen bis zu den Zehen aufsteigenden Winkel von 10° bildet. Um dies am Beispiel zu überprüfen, sucht er sich aus dem Videofilm einen geeigneten Sprung heraus, den er in Standbildprojektion auf den Bildschirm bringt. Mit Hilfe eines Winkelmessers, den er auf den Bildschirm legt - den Nullpunkt auf das Schultergelenk, die Basis in der Horizontalen -, ermittelt er einen Winkel von 11°. Eine spontane Wiederholungsmessung ergibt 12,5°. Um sich zu vergewissern, führt der Versuchsleiter den Meßvorgang weitere 8 Mal durch, *indem er jeweils den Winkelmesser neu anlegt und den Wert abliest. Dadurch erhält er folgende Datenreihe:*

11°   12,5°   12°   10,5°   12°   10°   10,5°   11°   10,5°   11°

Der Mittelwert (MW) der Messungen beträgt 11,1°, die Standardabweichung (s) 0,77°.

Der Variabilitätskoeffizient (VK) errechnet sich nach der Formel:

$$VK = \frac{s}{MW} * 100$$

Im obigen Beispiel ergibt sich ein Variabilitätskoeffizient von 6,94%. Dieser Wert kann als Meßfehler interpretiert werden und zwar - da nur **eine** Person die einzelnen Messungen durchführte - als **intrapersoneller Meßfehler**. Es ist anzunehmen, daß bei einer Wiederholungsmessung unterschiedlicher Versuchsleiter der Meßfehler (der **inter**personelle Meßfehler) höher ist, vor allem dann, wenn die einzelnen Auswerter die Ergebnisse der anderen nicht kennen.

Beispiel: Ließe der Versuchsleiter den Winkel des im Videofilm ausgewählten Pferdsprunges durch 10 Personen unabhängig voneinander bestimmen, könnte folgende Datenreihe entstehen:

12°   11,5°   12,5°   10°   11°   12,5°   9,5°   11,5°   10°   11°

Der Mittelwert der 10 Messungen weicht mit 11,15° nicht wesentlich von dem der intrapersonellen Meßreihe ab, die Standardabweichung ist mit 1,00° jedoch deutlich größer. Damit vergrößert sich der Variabilitätskoeffizient auf VK = 8,98%.

In der Regel fällt aus naheliegenden Gründen der interpersonelle Meßfehler größer aus als der intrapersonelle Meßfehler. Ersterer läßt sich durch genaue Anweisungen zur Handhabung des Meßinstrumentes und durch Definition des zu messenden Merkmals reduzieren. In bewegungs- und trainingswissenschaftlichen Untersuchungen sollte der Meßfehler nicht mehr als 5% betragen.

5.10 Standardfehler und Vertrauensgrenzen

Eine Stichprobe sollte zwar eine Grundgesamtheit möglichst genau widerspiegeln, trotzdem wird der Mittelwert des in der Stichprobe untersuchten Merkmals nicht völlig mit dem Mittelwert der Grundgesamtheit (den man natürlich nicht kennt; denn sonst hätte man ja keine Stichprobe zu untersuchen brauchen) übereinstimmen, und es bleibt vorerst offen, inwieweit man dem Ergebnis der Stichprobe trauen kann.

Beispiel: Wenn wir wissen wollen, wie schnell die Sportstudenten der Hochschule X im Durchschnitt über 100 m sprinten, könnte man sich mit der in Tab. I aufgeführten Stichprobe von 20 Sportstudenten begnügen, sofern man der Meinung ist, daß sie eine repräsentative Stichprobe darstellt. Trotzdem muß man damit rechnen, daß der errechnete Mittelwert von 12,5 sec (Tab. III) nicht völlig mit dem Mittelwert übereinstimmt, den man gewonnen hätte, wenn man alle Personen der Grundgesamtheit vermessen hätte.

Mit Hilfe der statistischen Kenngröße “**Standardfehler des Mittelwertes, s_{MW}**” (oder kurz “Standardfehler”) läßt sich nun **abschätzen**, wie weit der Stichprobenmittelwert durchschnittlich vom (unbekannten) Mittelwert der Grundgesamtheit abweicht.

Der Standardfehler berechnet sich nach der Formel

$$s_{MW} = \frac{s}{\sqrt{N}} \ ,$$

wobei s die Standardabweichung der Stichprobe und N der Umfang der Stichprobe bedeuten.

Mit Hilfe des Standardfehlers lassen sich nun die **Vertrauensgrenzen** einer Stichprobe festlegen. Diese bestimmen dasjenige Intervall (**Vertrauensintervall**), in dem mit einer vorbestimmten statistischen Sicherheit der Mittelwert der Grundgesamtheit erwartet werden kann. Da man sich in bewegungs- und trainingswissenschaftlichen Untersuchungen mit einer 95%igen Sicherheit zufriedengeben kann, berechnet man das Vertrauensintervall nach der Formel

$$\text{Vertrauensintervall} = MW \pm 1{,}96 * s_{MW} \ .$$

Hier bedeuten MW den Mittelwert der Stichprobe und s_{MW} den Standardfehler des Stichprobenmittelwertes (1,96 ist der Faktor für eine 95%ige Sicherheit.).

*Beispiel: Der Standardfehler s_{MW} des Mittelwertes der 100m-Sprintzeiten von Tab. I ergibt mit Hilfe der Standardabweichung von Tab. V den Wert $0{,}53\,\text{sec} / \sqrt{20} = 0{,}12\,\text{sec}$. Daraus lassen sich die Vertauensgrenzen berechnen, nämlich die untere Vertrauensgrenze mit 12,5 sec - 1,96*0,12 sec und die obere Vertrauensgrenze mit 12,5 sec + 1,96*0,12 sec. Das bedeutet, wir können aufgrund des Ergebnisses der Stichprobe von 20 Sportstudenten mit 95%iger Sicherheit darauf vertrauen, daß der Mittelwert der 100m-Sprintzeiten aller Sportstudenten der Hochschule X in dem Vertrauensintervall zwischen 12,26 sec und 12,74 sec liegt.*

Die Berechnung des Standardfehlers kann helfen, die für eine Stichprobe notwendige Anzahl von Versuchspersonen abzuschätzen (siehe dazu die Literatur!). Um in einfach strukturierten bewegungs- und trainingswissenschaftlichen Experimenten (bei denen häufig ökonomische und terminliche Zwänge zu berücksichtigen sind) keine zu großen Standardfehler zu riskieren, sollten Stichproben zur Untersuchung von Unterschiedshypothesen etwa 14 Vpn (möglichst nicht weniger als 8 Vpn) umfassen.

6. Prüfstatistik

Mit Hilfe der beschreibenden Statistik (Kap. 5) werden die Kennwerte von Stichproben festgestellt, um eine Strichprobe charakterisieren und zwei oder mehrere Stichproben vergleichen zu können. Mit diesem Vergleich ist aber noch nichts darüber ausgesagt, ob die festgestellten Unterschiede oder Zusammenhänge deutlich genug sind, um die Ungewißheit über die Gültigkeit der Hypothese beseitigen zu können. Um den Forscher dazu in die Lage zu versetzen, bietet die **Prüfstatistik** Rechenverfahren an, deren Ergebnisse darüber aufklären, mit welcher **Wahrscheinlichkeit** das Ergebnis einer Untersuchung durch den **Zufall** zustande gekommen ist.

In der Regel gibt man sich in den Bewegungs- und Trainingswissenschaften mit einem Anteil des Zufalls am Zustandekommen eines Testergebnisses (mit einer **“Zufallswahrscheinlichkeit”** bsw. **“Irrtumswahrscheinlichkeit”, p**) von 5% und weniger zufrieden. Bei besonders schwerwiegenden Entscheidungen wird nur eine Irrtumswahrscheinlichkeit von weniger als 1% akzeptiert. Überschreitet die Zufälligkeit des Ergebnisses den festgelegten Wert von 5 % (oder 1%) nicht, nennt man das Ergebnis “**signifikant**”. Dieser Begriff darf in der Statistik nicht im umgangssprachlichen Sinn (= bedeutsam, wichtig) verwendet werden, sondern bedeutet stets **“überzufällig”.** Hat ein statistisches Prüfverfahren die **Signifikanz** eines Ergebnisses herausgestellt, ist es erlaubt, die Nullhypothese zu verwerfen und die Alternativhypothese anzunehmen. Die anfängliche Ungewißheit über das Eintreffen der in der Hypothese formulierten Vorhersage ist damit reduziert, aber - wohl gemerkt - nicht völlig beseitigt. Diesen Rest an Ungewißheit (mit der Wahrscheinlichkeit von 5% bzw. 1% trifft die Nullhypothese möglicherweise doch zu) sollte der ernsthafte Wissenschaftler niemals außer acht lassen und sich vor Pauschalaussagen hüten wie etwa “Es ist wissenschaftlich bewiesen, daß......” (zum “Fehler erster oder zweiter Art” siehe einschlägige Literatur).

Welches Rechenverfahren man zur Signifikanzprüfung anwendet, hängt u.a. davon ab,

1. ob eine Unterschieds- oder eine Zusammenhangshypothese vorliegt (Kap. 2),
2. welches Skalenniveau die geprüften Variablen besitzen (Kap. 3.2)
3. ob es sich um abhängige oder unabhängige Stichproben handelt (Kap. 6.2 und 6.3),
4. welche Anzahl von Variablen und
5. welche Anzahl von Stichproben verglichen werden,
6. wie umfangreich die Stichproben sind.

Da in einfach strukturierten Experimenten der Bewegungslehre in der Regel kleine Stichproben (N < 30) untersucht, und meistens nur ein bis zwei Stichproben (oder ein bis zwei Variablen) verglichen werden, sollen im folgenden nur die Bedingungen 1-3 beachtet werden.

Wendet man das ausgewählte Prüfverfahren (= den ausgewählten statistischen Test) an dem vorliegenden Datenmaterial an, wird als Ergebnis der Wert einer **Prüfgröße** errechnet. Dieser im Experiment erzeugte Wert (= ”tatsächlicher” Wert, **“beobachteter” Wert**) wird in einer (für jeden statistischen Test speziell angefertigten) Tabelle mit einem (“theoretischen”) **Mindestwert** für signifikante Ergebnisse, der vom Stichprobenumfang und vom gewählten Signifikanzniveau abhängt, verglichen. Ist der beobachtete Wert gleich oder größer als der Mindestwert der Tabelle, kann das Ergebnis des Experimentes signifikant genannt werden. Im anderen Fall ist die Nullhypothese beizubehalten. Im folgenden wird diese Vorgehensweise nicht immer praktiziert, weil manche EXCEL-Befehle nach Eingabe des Datenmaterials automatisch die errechnete Prüfgröße (ohne diese anzuzeigen) mit der Mindestgröße vergleichen und dann als Ergebnis der Prüfrechnung und des Vergleiches direkt die Irrtumswahrscheinlichkeit ausgeben.

6.1 Signifikanzprüfung bei Zusammenhangshypothesen

Die Stärke des Zusammenhanges zweier Variablen wird durch den Korrelationskoeffizienten bestimmt (Kap. 5.7). Ob jedoch eine signifikante Korrelation vorliegt, hängt außer von der **Höhe des Korrelationskoeffizienten** von **der Anzahl**

der untersuchten Versuchspersonen ab. Welchen Mindestwert ein Korrelationskoeffizient bei einer gegebenen Anzahl von Vpn annehmen muß, um das gewünschte Signifikanzniveau von p = 5% (oder p = 1%) zu erfüllen, kann aus Tabelle VII abgelesen werden.

Beispiel: Bei den untersuchten 20 Personen korreliert die Variable "Note" mit der Variablen "Zeit" nur mäßig: R = -0,344 (s. Tab. VI). Als Mindestwert für eine signifikante Korrelation findet man in Tab. VII (rechte Wertetabelle) bei 20 Vpn den Koeffizienten 0,377 für eine 5%-Irrtumswahrscheinlichkeit bei Rangkorrelation. Dieser Tabellenwert wird von dem in der Stichprobe errechneten Wert (dem beobachteten Wert) nicht erreicht, was bedeutet,

Tab. VII: Mindestgrößen von Korrelationskoeffizienten, die signifikant sind

Produkt-Moment-Korrelation

Anzahl der Vpn.	Mindestgröße von **r** bei	
	p = 5%	p = 1%
3	0 , 950	0 , 990
4	0 , 878	0 , 959
5	0 , 811	0 , 917
6	0 , 754	0 , 874
7	0 , 707	0 , 834
8	0 , 666	0 , 798
9	0 , 632	0 , 765
10	0 , 602	0 , 735
11	0 , 576	0 , 708
12	0 , 553	0 , 684
13	0 , 532	0 , 661
14	0 , 514	0 , 641
15	0 , 497	0 , 623
16	0 , 482	0 , 606
17	0 , 468	0 , 590
18	0 , 456	0 , 575
19	0 , 444	0 , 561
20	0 , 433	0 , 549
21	0 , 423	0 , 537
22	0 , 413	0 , 526
23	0 , 404	0 , 515
24	0 , 396	0 , 505
25	0 , 388	0 , 496
26	0 , 381	0 , 487
27	0 , 374	0 , 478
28	0 , 367	0 , 470
29	0 , 361	0 , 463
30	0 , 355	0 , 456
40	0 , 304	0 , 393
50	0 , 273	0 , 354
60	0 , 250	0 , 332
80	0 , 217	0 , 283
100	0 , 195	0 , 254

Rangkorrelation

Anzahl der Vpn.	Mindestgröße von **R** bei	
	p = 5%	p = 1%
5	0 , 900	
6	0 , 829	0 , 943
7	0 , 714	0 , 893
8	0 , 643	0 , 833
9	0 , 600	0 , 783
10	0 , 564	0 , 746
12	0 , 506	0 , 712
14	0 , 456	0 , 645
16	0 , 425	0 , 601
18	0 , 399	0 , 564
20	0 , 377	0 , 534
22	0 , 359	0 , 508
24	0 , 343	0 , 485
26	0 , 329	0 , 465
28	0 , 317	0 , 448
30	0 , 306	0 , 432

daß der Einfluß des Zufalls am Zustandekommen des Ergebnisses größer ist, als man akzeptieren kann. Somit muß die Korrelation als ***nicht signifikant*** *gelten. Die Hypothese, daß die Pferdsprungnote mit der Sprintfähigkeit zusammenhängt, kann nicht als gültig angenommen werden, es ist die Nullhypothese beizubehalten.*

Die besprochenen Bedingungen gelten uneingeschränkt auch bei der Verwendung der Produkt-Momentkorrelation im Falle intervallskalierter Daten. Zu beachten ist, daß die Mindestwerte des Rangkorrelationskoeffizienten (Tab. VII, rechts) und des Produkt-Moment-Korrelationskoeffizienten (Tab. VII, links) nicht übereinstimmen.

6.2 Signifikante Unterschiede zwischen “unabhängigen Stichproben”

Unabhängig sind Stichproben dann, wenn das Ergebnis der einen das Ergebnis der anderen nicht beeinflußt. Bevor dazu ein Beispiel behandelt wird, soll zuerst die Auswahl statistischer Prüftests bei Unterschiedshypothesen besprochen werden:

Welche statistischen Testverfahren zur Überprüfung von Unterschiedshypothesen angewendet werden, hängt von folgenden Bedingungen ab:

Tab. VIII: Schema zur Auswahl statistischer Tests bei Unterschiedshypothesen

Bedingung A (Skalenniveau)	**Bedingung B** (Abhängigkeit)	**Bedingung C** (Varianz)	**anzuwendender** **Test**
Nominalskala	unabhängige Stichproben		CHI^2-Test (siehe Literatur)
	abhängige Stichproben		Vorzeichen-Test (siehe Literatur)
Ordinalskala	unabhängige Stichproben		WILCOXON-WHITE-Test (S.36)
	abhängige Stichproben		WILCOXON-Test (S. 39)
Intervallskala	unabhängige Stichproben	homogene Varianz	t-Test bei homogener Varianz (S. 36)
		heterogene Varianz	t-Test bei heterogener Varianz (S. 36)
	abhängige Stichproben		t-Test für abhängige Stichproben (S. 40)

Beispiel: Aus Tabelle V geht hervor, daß die 7 Leichtathleten unter den 20 Sportstudenten mit 12,1 sec im Mittel niedrigere Sprintzeiten erreichen als die Fußballer mit 12,6 sec. Um zu prüfen, ob diese Differenz nur zufällig zustande gekommen ist oder signifikant ist, steht bei intervallskalierten Daten wie der Sprintzeit der t-Test für unabhängige Stichproben zur Verfügung.

Bevor der **t-Test** zur Prüfung der Mittelwertsunterschiede intervallskalierter Daten bei unabhängigen Stichproben angewendet wird, ist festzustellen, ob die Streuung der Merkmalsausprägungen beider Stichproben ähnlich ist oder nicht; denn in beiden Fällen muß eine andere Form des t-Tests angewendet werden (Begründung siehe Literatur).

Das Maß für die Streuung bei intervallskalierten Daten ist die Standardabweichung (Kap. 5.5). Mit Hilfe des **F-Test** läßt sich nun prüfen, ob sich die **Varianzen** (= die quadrierten Standardabweichungen) beider Stichproben gleichen (d.h. **homogen** sind) oder nicht (d.h. **heterogen** sind). Die Standardabweichungen der Sprintzeiten der Fußballer und Leichtathleten unserer Erhebung in Kap. 5.5 betragen beispielsweise 0,413 sec bzw. 0,476 sec. Es scheinen demnach auf den ersten Blick keine allzu großen Differenzen in der Varianz der Sprintzeiten von Fußballern und Leichtathleten zu bestehen. Dies bestätigt der F-Test.

*Beispiel: In Tabelle IX ist zuerst mit Hilfe des F-Tests geprüft, ob es Unterschiede in der Varianz der Sprintzeiten von Fußballern und Leichtathleten gibt. Als Ergebnis liefert der EXCEL-Befehl die Irrtumswahrscheinlichkeit für die Annahme eines Varianz**unterschiedes**. Diese beträgt 67,5%, ist demnach so hoch, daß die Nullhypothese ("kein Varianzunterschied") bestehen bleiben kann. Somit können die Mittelwertsunterschiede der Sprintzeiten beider Sportartgruppen mit dem **t-Test für Stichproben mit homogener Varianz** berechnet werden. Der entsprechende EXCEL-Befehl liefert die Irrtumswahrscheinlichkeit dieses Unterschiedes. Sie ist kleiner als 5%, man schreibt auch: p < 0,05. Somit kann die Nullhypothese verworfen und die Alternativhypothese, die einen Unterschied in den Sprintzeiten der Fußballer und Leichtathleten unseres Experiments erwartet, als bestätigt angenommen werden.*

*Würde man ein Interesse haben, den Unterschied der Sprintzeiten der Turner und Leichtathleten unserer Erhebung zu vergleichen, könnte man aus guten Gründen von vornherein annehmen, daß die Turner langsamer sprinten als die Leichtathleten. Hier läge eine Hypothese mit **einseitiger Fragestellung (**Kap. 2) vor. In diesem Fall könnte man den **t-Test für einseitige Fragestellungen** anwenden, dessen EXCEL-Befehl in Tab. IX wiedergegeben ist. Dieser führt leichter zu einem signifikanten Ergebnis als der Test für zweiseitige Fragestellung, darf aber nur angewendet werden, wenn eine begründete einseitig formulierte Hypothese vorliegt.*

Tab.IX: Sportartspezialisierung (Disz.), Pferdsprung-Zensur (Note), 100m-Sprintzeit (Zeit) in sec und Irrtumswahrscheinlichkeit des F-Tests und des t-Tests der Sprintzeit-Unterschiede von Fußballern und Leichtathleten (Quelle: Urliste Tab. I)

	A	B	C	D
1	Vp.	Disz.	Note	Zeit
2	C	Fußb.	2,0	12,7
3	I	Fußb.	2,7	12,3
4	A	Fußb.	3,0	13,0
5	G	Fußb.	3,7	13,1
6	Q	Fußb.	3,7	12,2
7	T	Fußb.	3,7	11,9
8	F	Fußb.	4,0	12,9
9	P	Fußb.	4,0	13,1
10	E	Fußb.	4,3	12,4
11	M	Fußb.	4,3	12,6
12	L	L.athlet.	2,0	12,8
13	D	L.athlet.	3,3	12,6
14	J	L.athlet.	3,3	11,8
15	S	L.athlet.	3,3	12,0
16	N	L.athlet.	3,7	12,3
17	B	L.athlet.	4,0	11,4
18	H	L.athlet.	5,0	12,1
19	R	Turnen	1,3	13,5
20	K	Turnen	1,7	12,9
21	O	Turnen	2,3	13,0

F-Test Fußballer/Leichtathleten: **p =** ***0,675*** exc =ftest(D2:D11;D12:D18)

t-Test Fußballer/Leichtathleten: **p =** ***0,044*** exc =ttest(D2:D11;D12:D18;2;2)

t-Test bei ungleicher Varianz: exc =ttest(D2:D11;D12:D18;2;**3**)

t-Test bei einseitiger Fragestellung: exc =ttest(D2:D11;D12:D18;**1**;2)

Formel für den F-Test:

$$F = \frac{s_1^2}{s_2^2}$$

Formel für den t-Test für unabhängige Stichproben mit homogener Varianz:

$$t = \frac{|MW_1 - MW_2|}{\sqrt{\frac{(N_1 - 1) * s_1^2 + (N_2 - 1) * s_2^2}{N_1 + N_2 - 1} * \left(\frac{1}{N_1} + \frac{1}{N_2}\right)}}$$

Formel für den t-Test für unabhängige Stichproben mit heterogener Varianz:

$$t = \frac{|MW_1 - MW_2|}{\sqrt{\frac{s_1^2}{N_1 - 1} + \frac{s_2^2}{N_2 - 1}}}$$

Hier bedeuten MW_1 und MW_2 die Mittelwerte der beiden Stichproben, N_1 und N_2 die Anzahl der Meßwerte der jeweiligen Stichproben und s_1 und s_2 die Standardabweichungen der beiden Stichproben. **Beachte**: Beim F-Test muß die größere Standardabweichung stets im Zähler stehen. Der Wert der Prüfgröße t gibt keine direkte Auskunft über die Signifikanz des Befundes. Dazu muß der errechnete t-Wert mit dem Mindestwert einer Tabelle für signifikante t-Werte verglichen werden (siehe Literatur).

Die t-Tests sind, wie aus Tab. VIII hervorgeht, nur bei intervallskalierten Variablen anzuwenden. Liegt ein Ordinalniveau vor, müssen andere Prüfverfahren zum Vergleich der zentralen Tendenzen von 2 Stichproben herangezogen werden, etwa der **WILCOXON-WHITE-Test.**

Beispiel: Bei Überprüfung der hypothetischen Vermutung, daß die Leichtathleten unserer Erhebung bessere Pferdsprung-Zensuren (Median = 3,3) bekommen als die Fußballer (Median = 3,7), weil sie schneller sprinten und deshalb besser über das Pferd springen können, ist ein Prüfverfahren für unabhängige Stichproben mit Rangvariablen anzuwenden, z.B. der WILCOXON-WHITE-Test. Im EXCEL-Programm gibt es keine brauchbare Formel zur Berechnung eines Vergleiches von Stichproben mit Rangvariablen. Aus diesem Grunde muß der WILCOXON-WHITE-Test schrittweise selbst berechnet werden (Tab. X). Dazu werden - nach entsprechender Sortierung der Datei nach aufsteigenden

Notenwerten (Tab. X, Spalte C) - zuerst Noten-Rangplätze an die beteiligten Versuchspersonen verteilt. (Zum Bilden von Rangreihen siehe Kap. 5.7. ***Achtung!*** *Hier nicht den EXCEL-Befehl zur Bildung von Rangreihen benutzen! Dieser liefert für den WILCOXON-WHITE-Test unbrauchbare Werte!) Dann wird die Prüfgröße (T_{WW}) nach folgender Formel bestimmt:*

$$T_{WW} = \left| T_1 - \frac{N_1 * (N+1)}{2} \right|$$

Dabei ist Tww die Prüfgröße des WILCOXON-WHITE-Tests, T_1 die Rangteilsumme und N_1 die Vpn-Anzahl in der kleineren der beiden Stichproben (in Tab. X: Leichtathleten) und N die Gesamtzahl der Vpn in beiden Stichproben.
Das Ergebnis dieser Berechnung, der Betrag der Prüfgröße T_{WW} , stellt nicht, wie bei der Berechnung des t-Tests, die Irrtumswahrscheinlichkeit dar. Statt dessen muß mit Hilfe dieser Prüfgröße anhand einer Tabelle (Tab. XI) die zugehörige Irrtumswahrscheinlichkeit noch ermittelt werden: Der Betrag der Prüfgröße ($T_{WW} = 4$) ist ***kleiner*** *als der Mindestwert (= 21) für signifikante Unterschiede zwischen Stichproben der vorliegenden Umfänge ($N_1 = 10$ und $N_2 = 7$; Tab. XI). D.h., die Unterschiede in den Noten von Leichtathleten und Fußballern sind nicht signifikant. Aus diesem Grunde muß die Nullhypothese, die besagt, daß die beiden Gruppen keinen Unterschied in den Pferdsprung-Zensuren zeigen, beibehalten werden.*

Tab. X: Sportartspezialisierung (Disz.), Pferdsprung-Zensur (Note), 100m-Sprintzeit (Zeit) in sec und Signifikanzprüfung der Notenunterschiede von Fußballern und Leichtathleten mit Hilfe des WILCOXON-WHITE-Tests. R_{FB}: Notenrangreihe der Fußballer. R_{LA}: Notenrangreihe der Leichtathleten. (Quelle: Urliste Tab. I)

	A	B	C	D	E	F	
1	Vp.	Disz.	Note	Zeit	RLA	RFB	
2	C	Fußb.	2,0	12,7		1,5	
3	L	L.athlet.	2,0	12,8	1,5		
4	I	Fußb.	2,7	12,3		3	
5	A	Fußb.	3,0	13,0		4	
6	J	L.athlet.	3,3	11,8	6		
7	S	L.athlet.	3,3	12,0	6		
8	D	L.athlet.	3,3	12,6	6		
9	T	Fußb.	3,7	11,9		9,5	
10	Q	Fußb.	3,7	12,2		9,5	
11	N	L.athlet.	3,7	12,3	9,5		
12	G	Fußb.	3,7	13,1		9,5	
13	B	L.athlet.	4,0	11,4	13		
14	F	Fußb.	4,0	12,9		13	
15	P	Fußb.	4,0	13,1		13,0	
16	E	Fußb.	4,3	12,4		15,5	
17	M	Fußb.	4,3	12,6		15,5	
18	H	L.athlet.	5,0	12,1	17		
19				Teilsumme T_1:	***59***	exc	=summe(E2:E18)
20				Anzahl N_1:	***7***	exc	=anzahl(E2:E18)
21				Gesamtanzahl N:	***17***	exc	=anzahl(E2:E18;F2:F18)
22				Prüfgröße T_{WW}:	***4***	exc	=abs(E19-E20*(E21+1)/2

Tab. XI: Mindestwerte für die Signifikanzprüfung der Testgröße T_{WW} des WILCOXON-WHITE-Tests für unabhängige Stichproben mit Rangvariablen (Quelle: nach MITTENECKER 1970)

	4	5	6	7	8	9	10	11	12	13	14	N2 / N1
5%	-	-	-	-	8,0	9,0	10,0	10,0	11,0	12,0	13,0	2
	-	7,5	8,0	9,5	10,0	11,5	12,0	13,5	14,0	15,5	16,0	3
	8,0	9,0	10,0	11,0	12,0	13,0	15,0	16,0	17,0	18,0	19,0	4
	9,0	10,5	12,0	12,5	14,0	15,5	17,0	18,5	19,0	20,5	22,0	5
14	56,0		13,0	15,0	16,0	17,0	19,0	20,0	22,0	23,0	25,0	6
13	53,0	50,5		16,5	18,0	19,5	21,0	22,5	24,0	25,5	27,0	7
12	50,0	47,0	44,0		19,0	21,0	23,0	25,0	26,0	28,0	29,0	8
11	47,0	44,5	42,0	39,5		22,5	25,0	26,5	28,0	30,5	32,0	9
10	44,0	41,0	39,0	36,0	34,0		27,0	29,0	30,0	32,0	34,0	10
9	41,0	38,5	36,0	33,5	32,0	29,5		30,5	33,0	34,5	37,0	11
8	38,0	35,0	33,0	31,0	29,0	27,0	25,0		35,0	37,0	39,0	12
7	34,0	32,5	30,0	28,5	26,0	24,5	22,0	20,5		38,5	41,0	13
6	31,0	29,0	27,0	26,0	24,0	22,0	20,0	18,0	16,0		43,0	14
5	28,0	25,5	24,0	22,5	21,0	19,5	18,0	15,5	14,0	12,5	-	
4	24,0	22,0	21,0	20,0	18,0	17,0	15,0	14,0	12,0	-	-	
3	20,0	18,5	17,0	16,5	15,0	13,5	-	-	-	-	-	1%
N_1 / N_2	14	13	12	11	10	9	8	7	6	5	4	

6.3 Signifikante Unterschiede bei “abhängigen Stichproben”

Abhängige Stichproben liegen immer dann vor, wenn eine Personengruppe mindestens **zweimal** (meistens vor und nach einer Behandlung) im Hinblick auf **ein** Merkmal vermessen (oder beurteilt) wurde.

Beispiel: Im vorliegenden Erhebungsbeispiel könnte man vorschlagen, die Pferdsprung-Zensuren von Sportstudenten dadurch zu verbessern, daß man mit ihnen ein Sprinttraining absolviert. Tab XII liefert das Ergebnis dieses Tests. Dazu wurden die Personen zuerst, d.h. vor dem Sprinttraining, beim Pferdsprung benotet (= Vortest; Tab. XII Spalte B). Dann bildete man zwei parallelisierte Gruppen (Tab. XII, Spalte C bis H). Die eine Gruppe (Vg) absolvierte ein 3-wöchiges Sprinttraining, während die andere als Kontrollgruppe (Kg) unbehandelt blieb. Am Ende der Behandlungszeit wurden beide Gruppen wieder beim Pferdsprung benotet (Nachtest; Tab. XII Spalte I).

Um die zentralen Tendenzen der Pferdsprung-Zensuren innerhalb der Gruppen (Vg bzw. Kg) und von Vortest und Nachtest zu charakterisieren, dient - bei ordinalskalierten Daten - der Median (Kap. 3.2). Sowohl die Versuchsgruppe, als auch die Kontrollgruppe zeigen im Nachtest einen geringfügig niedrigeren Median als im Vortest (Tab. XII).

Tab.XII: Pferdsprung-Zensuren (Note) von 20 Sportstudenten vor (Vortest) und nach (Nachtest) einem 3-wöchigen Sprinttraining der Versuchsgruppe (Vg). Anwendung des WILCOXON-Tests. D: Differenz der Noten von Vortest und Nachtest in der Versuchsgruppe. Kg: Kontrollgruppe. T1 und T2: Rang-Teilsummen

	A	B	C	D	E	F	G	H	I	J	K	L	M
1			Gruppenbildung					Vor-	Nach-		Rang	Rang	Rang
2	Vortest							test	test		von	bei	bei
3	Vp.	Note	Vp.	Note	Gruppe		Vp.	Note	Note	**D**	|D|	+D	-D
4	A	3,0	R	1,3	Vg	**Vg**	D	3,3	2,7	0,6	4	4	
5	B	4,0	K	1,7	Kg		F	4,0	4,0	0,0			
6	C	2,0	C	2,0	Kg		H	5,0	4,0	1,0	1	1	
7	D	3,3	L	2,0	Vg		J	3,3	3,3	0,0			
8	E	4,3	O	2,3	Vg		L	2,0	1,7	0,3	6	6	
9	F	4,0	I	2,7	Kg		N	3,7	4,0	-0,3	6		6
10	G	3,7	A	3,0	Kg		O	2,3	3,0	-0,7	2,5		2,5
11	H	5,0	D	3,3	Vg		P	4,0	3,3	0,7	2,5	2,5	
12	I	2,7	J	3,3	Vg		Q	3,7	4,0	-0,3	6		6
13	J	3,3	S	3,3	Kg		R	1,3	1,3	0,0			
14	K	1,7	G	3,7	Kg	**Kg**	A	3,0	2,7	N=	***7***		
15	L	2,0	N	3,7	Vg		B	4,0	5,0	exc	=anzahl(K4:K13)		
16	M	4,3	Q	3,7	Vg		C	2,0	2,3				
17	N	3,7	T	3,7	Kg		E	4,3	4,0				
18	O	2,3	B	4,0	Kg		G	3,7	3,3		T_1=	**13,5**	
19	P	4,0	F	4,0	Vg		I	2,7	2,0			T_2=	**14,5**
20	Q	3,7	P	4,0	Vg		K	1,7	2,0				
21	R	1,3	E	4,3	Kg		M	4,3	3,7	PrüfgrößeT_W:			***0,50***
22	S	3,3	M	4,3	Kg		S	3,3	4,0	exc			
23	T	3,7	H	5,0	Vg		T	3,7	3,5	=abs(L18-K14*(K14+1)/4)			
					Median	Vg:		***3,5***	***3,3***				
						Kg:		***3,5***	***3,4***				

Um jetzt zu prüfen, ob sich Vpn vom Vortest zum Nachtest hinsichtlich der Pferdsprungzensuren verbessert haben, vergleicht man die Unterschiede in den zentralen Tendenzen beider Meßreihen (Stichproben) durch entsprechende statistische Prüfverfahren für abhängige Stichproben. Bei intervallskalierten Daten benutzt man den **t-Test für abhängige Stichproben** (“gepaarter t-Test”), bei ordinalskalierten Daten den **WILCOXON-Test**.

Beispiel: Zur Prüfung der Veränderung der Pferdsprungzensuren vom Vortest zum Nachtest muß man selbstverständlich den WILCOXON Test einsetzen, da die zu prüfenden Daten (Zensuren) ordinalskaliert sind. Auch für diesen gilt, was schon für den WILCOXON-WHITE-Test festgestellt werden mußte: Auf Grund des fehlens einer geeigneten Formel im EXCEL-Programm müssen die einzelnen Rechenschritte selbst ausgeführt werden (Tab. XII). Zur Anwendung des WILCOXON-Tests berechnet man zuerst die Differenzen der Vor- und Nachtest-Noten. Danach werden für die absoluten Werte der Differenzen, die *ungleich Null sind, Rangplätze vergeben (Tab. XII, Spalte K) und dann zwei Untergruppen gebildet, in denen die Rangplätze derjenigen Vpn stehen, die sich verbessert (Tab. XII, Spalte L) bzw. die sich verschlechtert haben (Tab. XII, Spalte M). Jetzt bildet man die Rang-Teilsummen T_1 und T_2 und benutzt die kleinere Rangteilsumme (hier: T_1) zur Berechnung der Prüfgröße T_W, und zwar nach der Formel:*

$$T_W = \left| T_1 - \frac{N * (N+1)}{4} \right|$$

Hier ist N die Anzahl derjenigen Vpn, die einen Rangplatz zugeordnet bekommen haben.

*Auch hier sagt das Ergebnis (der Betrag der Prüfgröße T_W) noch nichts über die Irrtumswahrscheinlichkeit aus. Dazu ist wieder eine Prüftabelle (Tab. XIII) zu Rate zu ziehen. Der Wert der Prüfgröße T_W , **0,5**, ist gemäß Tabelle XIII kleiner als die Mindestgröße von 12 bei 7 Versuchspersonen und bei einer Irrtumswahrscheinlichkeit von p = 5%. Aus diesem Grunde liegt keine signifikante Veränderung der Pferdsprungzensur vor, d.h. das Anheben der Pferdsprungzensur von 3,5 im Vortest auf 3,3 im Nachtest (jeweils Median) muß als zufallsbedingt angesehen werden. Die Alternativhypothese (Sprinttraining verbessert die Pferdsprungnote) kann nicht angenommen werden. Statt dessen ist die Nullhypothese beizubehalten. Dieses Ergebnis korrespondiert durchaus mit dem in Kap. 6.1 gewonnenen Befund, der keinen signifikanten Zusammenhang zwischen Sprintleistung und Pferdsprung-Zensur aufdecken konnte.*

Tab.XIII: Mindestwerte für die Signifikanzprüfung der TestgrößeT_W des WILCOXON-Tests für abhängige Stichproben mit Rangvariablen (Ouelle: nach MITTENECKER 1970)

p \ N	6	7	8	9	10	11	12	13	14	15
5%	10,5	12,0	14,0	16,5	19,5	22,0	25,0	28,5	31,5	35,0
1%			18,0	20,5	24,5	28,0	32,0	35,5	39,5	44,0

p \ N	16	17	18	19	20	21	22	23	24	25
5%	38,0	41,5	45,5	49,0	53,0	56,5	60,5	65,0	69,0	73,5
1%	48,0	53,5	57,5	63,0	67,0	72,5	77,5	83,0	89,0	94,5

In gleicher Weise kann mit Hilfe des WILCOXON-Tests nachgepüft werden, ob die Änderung der Pferdsprungnote in der Kontrollgruppe vom Vortest zum Nachtest signifikant ist. Auch dies kann - wie nicht anders zu erwarten - nicht bestätigt werden.

Bei der Prüfung von Behandlungseffekten bei intervallskalierten Variablen ist der **t-Test für abhängige Stichproben** anzuwenden. Im EXCEL-Programm liegt dazu der entsprechende Befehl vor (Tab. XIV). Als Ergebnis wird die Irrtumswahrscheinlichkeit angezeigt, so daß es sich hier erübrigt, die entsprechende Formel herzuleiten bzw. eine Tabelle mit Mindestwerten für die Signifikanzprüfung parat zu haben.

Beispiel: Bei der Prüfung, ob sich die Pferdsprungzensuren der Versuchsgruppe im vorigen Beispiel vom Vortest zum Nachtest verbessert haben, wurde kein signifikantes Ergebnis erzielt. Das könnte möglicherweise daran liegen, daß - in unserem fiktiven Experiment - das 3-wöchige Sprinttraining der Versuchsgruppe gar nicht zu einer Verbesserung der Sprintfähigkeit geführt hat und somit die Fähigkeit, über das Pferd zu springen, auch nicht verbessert werden konnte. Nehmen wir an, wir hätten von den 20 Vpn nicht nur im Vortest, sondern auch im Nachtest zusätzlich die 100m-Sprintzeiten getestet, dann läßt sich das obige Problem klären. Tabelle XIV liefert die Sprintzeiten vom Vortest und Nachtest und zeigt in der Versuchsgruppe eine mittlere Verbesserung der Sprintzeit um 0,14 sec von 12,63 sec auf 12,49 sec. Die Anwendung des t-Tests für abhängige Stichproben liefert für das Zustandekommen dieser Verbesserung eine Irrtumswahrscheinlichkeit von 2,95 %, was deutlich unter der 5%-Grenze liegt. Somit kann eine signifikante Verbesserung der Sprintleistung in der Versuchsgruppe festgestellt werden.

Die Anwendung des t-Tests für abhängige Stichproben zur Prüfung und Veränderung der Sprintzeit in der Kontrollgruppe liefert eine Irrtumswahrscheinlichkeit von 89%, d.h. die Sprintleistungen ändern sich in der Kontrollgruppe nicht signifikant.

Formel für den t-Test für abhängige Stichproben:

$$t = \frac{D}{\sqrt{\frac{\sum d_i^2 - N * D^2}{N * (N-1)}}}$$

Hier bedeuten d_i die einzelnen Differenzen der abhängigen Meßwerte, D die Summe der Differenzen und N die Anzahl der Versuchspersonen bzw. Meßwertpaare.

Tab.XIV: 100m-Sprintzeiten von 20 Sportstudenten im Vortest und Nachtest und Mittelwerte. Behandlung der Versuchsgruppe (Vg): 3-wöchiges Sprinttraining. t-Test für abhängige Stichproben. (Quelle: Vortest-Zeiten Urliste Tab.I)

	A	B	C	D	E	F	G
1			Zeit [sec]			Mittelwert	
2		Vp.	Vortest	Nachtest	Diff.	Vortest	Nachtest
3	**Vg**	D	12,6	12,4	0,2	**12,63**	**12,49**
4		F	12,9	12,8	0,1		
5		H	12,1	12,2	-0,1		
6		J	11,8	11,7	0,1		
7		L	12,8	12,4	0,4		
8		N	12,3	12,3	0,0		
9		O	13,0	12,7	0,3		
10		P	13,1	13,2	-0,1		
11		Q	12,2	12,0	0,2		
12		R	13,5	13,2	0,3		
13	**Kg**	A	13,0	13,2	-0,2	**12,43**	**12,53**
14		B	11,4	11,8	-0,4		
15		C	12,7	12,5	0,2		
16		E	12,4	12,4	0,0		
17		G	13,1	13,3	-0,2		
18		I	12,3	12,1	0,2		
19		K	12,9	13,0	-0,1		
20		M	12,6	12,8	-0,2		
21		S	12,0	12,2	-0,2		
22		T	11,9	12,0	-0,1		

t-Test für abhängige Stichproben, Vg: **p =** ***0,0295***

exc =ttest(C3:C12;D3:D12;2;1)

t-Test für abhängige Stichproben, Kg: **p =** ***0,8971***

exc =ttest(C13:C22;D13:D22;2;1)

7. Diskussion der Befunde

Die beschreibende und prüfende Statistik hat ausschließlich die Aufgabe, die in einem Experiment gewonnenen Daten rechnerisch auszuwerten und festzustellen, ob die **Nullhypothese verworfen** und eine **Hypothese angenommen** werden kann oder ob die **Nullhypothese beizubehalten** ist. Das dabei erreichte Prüfergebnis ist nun zu interpretieren. Dabei sollten unter anderem auf folgende Gesichtspunkte eingegangen werden:

- Vergleich mit ähnlichen Ergebnissen aus der Literatur,
- Diskussion möglicher Fehlerquellen,
- Erläuterung der Faktoren, die zum Ergebnis geführt haben,
- Einordnung der Bedeutung der Befunde in die bestehende Theorie,
- Konsequenzen der Befunde für die Praxis,
- Aufzeigen weiterer offener Fragen und weiterführender Hypothesen,
- Konsequenzen für anschließende Forschungen.

(siehe auch Skript "Empfehlungen zur Anfertigung sportwissenschaftlicher (empirischer) Arbeiten")

8. Literatur

In der folgenden Literaturliste sind Lehrbücher, die sich zur Einführung in die Statistik eignen, durch ✷ und Lehrbücher mit vorwiegend sportwissenschaftlicher Thematik durch ▼ gekennzeichnet.

BORTZ, J.: Lehrbuch der Statistik. Berlin 1979

✷▼ BÖS, K.: Statistikkurs I. Ahrensburg 1986

EHRENBERG, A.S.C.: Statistik oder der Umgang mit Daten. Weinheim 1986

CLAUSS, G., und EBNER, H.: Grundlagen der Statistik für Psychologen, Pädagogen und Soziologen. Frankfurt 1971

LIENERT, G.A.: Testaufbau und Testanalyse. Meisenheim 1969

✷ MITTENECKER, E.: Planung und statistische Auswertung von Experimenten. Wien 1970

✷▼ WILLIMCZIK, K.: Grundkurs Statistik. Frankfurt 1975

WIEMANN, K.: Empfehlungen zur Anfertigung sportwissenschaftlicher (empirischer) Arbeiten. Uni Wuppertal

9. Verzeichnis der benutzten Abkürzungen

exc	Hinweis auf einen Formelbefehl im EXCEL-Programm
F	Prüfgröße für den F-Test
Kg	Kontrollgruppe
MD	Median
MW	arithmetisches Mittel, Mittelwert
N, n	Anzahl von Versuchspersonen, Stichprobenumfang
p	Irrtumswahrscheinlichkeit
r	Koeffizient der Produkt-Moment-Korrelation
R	Rangkorrelationskoeffizient
R_{FB}, R_1...	Rang, Rangplätze
s	Standardabweichung
s_{MW}	Standardfehler des Mittelwertes
sec	Sekunden
t	Prüfgröße für den t-Test
T, T_1,T_2..	Rangsumme, Rangteilsumme
Tw	Prüfgröße für den WILCOXON-Test
Tww	Prüfgröße für den WILCOXON-WHITE-Test
Vg	Versuchsgruppe
Vl	Versuchsleiter
Vp, Vpn	Versuchsperson, Versuchspersonen

10. Verzeichnis von Formeln und Nachschlage-Tabellen

11. **Register** (die Seitenangaben beziehen sich auf die Stelle im Text, in der die Begriffe definiert werden)