



Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich



Auswertung einer Brauchbarkeitsuntersuchung von alternativen Tools für Tangible Interaction

Semesterarbeit
von Raphael Wirz

Chalmers University of Technology, Göteborg;
Human-Computer Interaction (HCI)
&
ETH, Zürich;
Zentrum für Produkt-Entwicklung (ZPE)

Betreuung durch
Dr. Morten Fjeld (Chalmers)
Dr. Andreas Kunz (ETH)

August 2005

Vorwort

Diese Semesterarbeit wurde an der Chalmers University of Technology in Göteborg, Schweden, durchgeführt.

Ich möchte mich bei Dr. Morten Fjeld, meinem Betreuer in Chalmers, bedanken, dass er mir ermöglicht hat, nach Schweden zu gehen und dort meine Arbeit durchzuführen. Er war stets offen bei Fragen und Unklarheiten und hat mir wichtige Anregungen und Unterstützung bei Problemen geboten.

Des weiteren möchte ich mich bei Dr. Andreas Kunz bedanken, der an der ETH die nötigen Vorbereitungen für meinen Austausch getroffen hat. Er hat mich dann vor allem auch bei der schriftlichen Ausarbeitung des Berichtes unterstützt.

Für den Rat in Fragen zu statistischen Problemstellungen möchte ich mich bei Kristina Bötschi und Kerstin Wiklander bedanken.

Zürich, August 2005

Raphael Wirz

Zusammenfassung

Das an der ETH Zürich entwickelte kollaborative Planungssystem BUILD-IT wird in einer Untersuchung mit zwei alternativen Tools verglichen und auf seine Brauchbarkeit hin geprüft. Aus der Untersuchung mit 60 Versuchsteilnehmern resultiert eine grössere Menge an Daten.

Diese Daten werden in der vorliegenden Arbeit strukturiert und ausgewertet. Es wird eine statistische Auswertungsmethode gesucht, mit der die Daten analysiert werden können. Die Methode wird den spezifischen Gegebenheiten angepasst und ein detailliertes Vorgehen wird festgelegt und dann umgesetzt. Abschliessend werden die Resultate der statistischen Auswertung in einer strukturierten Weise dargestellt und diskutiert.

Inhaltsverzeichnis

Vorwort	ii
Zusammenfassung	iii
Inhaltsverzeichnis	iv
Abbildungsverzeichnis	vii
Tabellenverzeichnis	viii
1. Einleitung	1
1.1. Aufgabenstellung	1
1.2. Vorgehensweise	1
2. Zielsuche	2
2.1. Situationsanalyse	2
2.1.1. Untersuchungsdurchführung	2
2.1.1.1. Aufgabendesign	2
2.1.1.2. Tools	4
2.1.1.3. Versuchsdesign	8
2.1.2. Software	9
2.1.3. Sichtung der Daten	10
2.1.4. Strukturierung der Daten	11
2.1.5. Methode für die Auswertung	12
2.2. Zielformulierung	14
3. Theoretische Grundlagen	15
3.1. Vorgehensweise	15
3.2. Mathematische Grundlagen	19
3.2.1. Boxplot	19
3.2.2. Standardabweichung	20
3.2.3. Normalverteilung	20
3.2.4. Arithmetisches Mittel	21
3.2.5. Transformationen	21
3.2.6. Schiefe	22
3.2.7. Kurtosis	23
3.2.8. Quadratwurzel	24
3.2.9. Natürlicher Logarithmus	24
3.2.10. Dyadischer Logarithmus	24
3.2.11. Levene-Test	24
3.2.12. Kolmogorov-Smirnov-Test	25
3.2.13. Histogramm	25
3.2.14. ANOVA	26
3.2.15. Post-Hoc-Tests	27
4. Auswertung der Messreihen	28
4.1. Prüfen der Voraussetzungen	28
4.1.1. Fehlerhafte Daten	28

4.1.1.1.	Untransformierte Daten	28
4.1.1.2.	Arithmetisches Mittel	29
4.1.1.3.	Transformation – sqrt	30
4.1.2.	Korrekte Daten	31
4.1.2.1.	Untransformierte Daten	31
4.1.2.2.	Arithmetisches Mittel	32
4.1.2.3.	Transformationen	33
4.2.	Beurteilung der Transformationen	35
4.2.1.	Boxplot	35
4.2.2.	Standardabweichung.....	37
4.2.3.	Histogramm	38
4.3.	Entscheidung über Transformationen.....	41
5.	Resultate der Messreihen	43
5.1.	Kognitive Unterstützung.....	43
5.2.	Hypothesen	44
5.3.	Kriterien zur Messung der kognitiven Unterstützung	44
5.4.	Darstellung der Resultate	45
5.4.1.	Abhängige Variablen der ANOVA	45
5.4.2.	ANOVA.....	47
5.4.3.	Post-Hoc-Test	48
5.4.4.	Zusammenfassung der Resultate.....	49
5.4.5.	Zusätzliche Auswertungen	50
5.4.5.1.	Qualitative Auswertung der Lerneffekte	50
5.4.5.2.	Qualitative Kontrolle sekundärer Daten.....	53
5.5.	Diskussion der Resultate	54
A.	Anhang	58
A.1.	Literaturverzeichnis	58
A.2.	Abkürzungsverzeichnis.....	59
A.3.	Glossar	60
A.4.	Verwendete SPSS Funktionen	61
A.5.	Excel Datentabellen.....	62
A.5.1.	Fehlerhafte Daten.....	62
A.5.2.	Korrekte Daten	62
A.5.2.1.	Untransformierte Daten	62
A.5.2.2.	Transformierte Daten	79
A.6.	SPSS Ausgabetafeln	81
A.6.1.	Fehlerhafte Daten.....	81
A.6.1.1.	Untransformierte Daten	81
A.6.1.2.	Arithmetisches Mittel	81
A.6.1.3.	Transformierte Daten – sqrt	82
A.6.2.	Korrekte Daten	82
A.6.2.1.	Untransformierte Daten	82
A.6.2.2.	Arithmetisches Mittel	83
A.6.2.3.	Transformierte Daten	84
A.6.3.	Resultate	86
A.6.3.1.	Zeit & Anzahl Blöcke	86

A.6.3.2.	Benutzerzufriedenheit	87
A.6.3.3.	Versuchserfolg	90

Abbildungsverzeichnis

Abbildung 1: Positionierungsproblem.....	3
Abbildung 2: BUILD-IT System mit Spiegel, Tisch und Leinwand.....	4
Abbildung 3: Draufsicht auf dem Tisch und Perspektive auf der Leinwand	5
Abbildung 4: Bricks	5
Abbildung 5: BUILD-IT Testumgebung	6
Abbildung 6: Modeler Testumgebung	7
Abbildung 7: Physische Testumgebung.....	8
Abbildung 8: Entscheidungsbaum für Unterschiedshypothesen	13
Abbildung 9: Berechnung des arithmetischen Mittels	16
Abbildung 10: Boxplot.....	20
Abbildung 11: Normalverteilungskurve	21
Abbildung 12: Beispiele von Verteilungen mit Rechtsschiefe (links), Linksschiefe (Mitte) und keiner Schiefe (symmetrische Verteilung) (rechts)	22
Abbildung 13: Beispiele von Verteilungen mit positiver (rechts) und negativer (links) Kurtosis.....	23
Abbildung 14: Histogramm mit Normalverteilungskurve	25
Abbildung 15: Boxplots Zeit und Anzahl Blöcke; fehlerhafte untransformierte Daten.....	28
Abbildung 16: Boxplots Zeit und Anzahl Blöcke; fehlerhafte Daten, arithmetisches Mittel.....	29
Abbildung 17: Boxplots Zeit und Anzahl Blöcke; fehlerhafte transformierte Daten	30
Abbildung 18: Boxplots Zeit und Anzahl Blöcke; korrekte untransformierte Daten	32
Abbildung 19: Boxplots Zeit und Anzahl Blöcke; korrekte Daten, arithmetisches Mittel.....	33
Abbildung 20: Boxplots Zeit und Anzahl Blöcke; korrekte transformierte Daten .	36
Abbildung 21: Histogramme Zeit; korrekte transformierte Daten	39
Abbildung 22: Histogramme Anzahl Blöcke; korrekte transformierte Daten.....	40
Abbildung 23: Erwartete kognitive Unterstützung (in Pfeilrichtung zunehmend) für jedes Tool und die Hypothesen H1-H3 zwischen den Tools.	44
Abbildung 24: Boxplot von Zeit (ln) und Anzahl Blöcke (sqrt)	46
Abbildung 25: Lerneffekte; Boxplots für Zeit und Anzahl Blöcke.....	52

Tabellenverzeichnis

Tabelle 1: Variablen	9
Tabelle 2: Standardabweichungen; fehlerhafte untransformierte Daten	29
Tabelle 3: Standardabweichungen; fehlerhafte Daten, arithmetisches Mittel.....	30
Tabelle 4: Falsche- und dazugehörige richtige Zeitwerte.....	31
Tabelle 5: Standardabweichungen; korrekte untransformierte Daten	32
Tabelle 6: Standardabweichungen; korrekte Daten, arithmetisches Mittel.....	33
Tabelle 7: p-Wert und Signifikanzentscheidung für Kolmogorov-Smirnov- und Levene-Test	34
Tabelle 8: Bewertung der Boxplots	37
Tabelle 9: Standardabweichungen Zeit; korrekte transformierte Daten	37
Tabelle 10: Standardabweichungen Anzahl Blöcke; korrekte transformierte Daten	37
Tabelle 11: Bewertung Standardabweichung	38
Tabelle 12: Bewertung der Histogramme.....	41
Tabelle 13: Gesamtbewertung.....	41
Tabelle 14: Erwartete kognitive Unterstützung	43
Tabelle 15: Arithmetisches Mittel von Zeit und Anzahl Blöcke.....	45
Tabelle 16: Klarheit, Schwierigkeit und Tauglichkeit ergeben die Benutzerzufriedenheit	46
Tabelle 17: Versuchserfolg	47
Tabelle 18: Resultate der einfaktoriellen ANOVA	48
Tabelle 19: Resultate von Tukey HSD, Post-Hoc-Tests	49
Tabelle 20: Zusammenfassung der Resultate	49
Tabelle 21: Lerneffekte?	51
Tabelle 22: Antwortbild des Fragebogens.....	53
Tabelle 54: Verwendete SPSS Funktionen.....	61
Tabelle 23: Fehlerhafte Daten	62
Tabelle 24: Primäre, untransformierte, korrekte Daten	75
Tabelle 25: Sekundäre Daten	77
Tabelle 26: Fragebogen.....	78
Tabelle 27: Transformierte Daten	80
Tabelle 28: SPSS Ausgabe; Standardabweichungen; fehlerhafte untransformierte Daten.....	81
Tabelle 29: SPSS Ausgabe; Standardabweichungen; fehlerhafte Daten, arithmetisches Mittel.....	81
Tabelle 30: SPSS Ausgabe; Standardabweichungen; fehlerhafte transformierte (sqrt) Daten	82
Tabelle 31: SPSS Ausgabe; Standardabweichungen; korrekte untransformierte Daten.....	82
Tabelle 32: SPSS Ausgabe; Standardabweichungen; korrekte Daten, arithmetisches Mittel.....	83
Tabelle 33: SPSS Ausgabe; Standardabweichungen; korrekte transformierte Daten.....	84

Tabelle 34: SPSS Ausgabe; Levene-Test.....	84
Tabelle 35: SPSS Ausgabe; Kolmogorov-Smirnov-Test für BUILD-IT	85
Tabelle 36: SPSS Ausgabe; Kolmogorov-Smirnov-Test für Modeler	85
Tabelle 37: SPSS Ausgabe; Kolmogorov-Smirnov-Test für Physical.....	86
Tabelle 38: SPSS Ausgabe; ANOVA	86
Tabelle 39: SPSS Ausgabe; Tukey HSD Post-Hoc-Test zur ANOVA	87
Tabelle 40: SPSS Ausgabe; Benutzerzufriedenheit; Arithmetisches Mittel.....	87
Tabelle 41: SPSS Ausgabe; Benutzerzufriedenheit; Levene-Test.....	87
Tabelle 42: SPSS Ausgabe; Benutzerzufriedenheit; Kolmogorov-Smirnov-Test für BUILD-IT	88
Tabelle 43: SPSS Ausgabe; Benutzerzufriedenheit; Kolmogorov-Smirnov-Test für Modeler	88
Tabelle 44: SPSS Ausgabe; Benutzerzufriedenheit; Kolmogorov-Smirnov-Test für Physical.....	89
Tabelle 45: SPSS Ausgabe; Benutzerzufriedenheit; ANOVA	89
Tabelle 46: SPSS Ausgabe; Benutzerzufriedenheit; Tukey HSD Post-Hoc-Test für ANOVA	89
Tabelle 47: SPSS Ausgabe; Versuchserfolg; Standardabweichung	90
Tabelle 48: SPSS Ausgabe; Versuchserfolg; Levene-Test.....	90
Tabelle 49: SPSS Ausgabe; Versuchserfolg; Kolmogorov-Smirnov-Test für BUILD-IT	90
Tabelle 50: SPSS Ausgabe; Versuchserfolg; Kolmogorov-Smirnov-Test für Modeler	91
Tabelle 51: SPSS Ausgabe; Versuchserfolg; Kolmogorov-Smirnov-Test für Physical.....	91
Tabelle 52: SPSS Ausgabe; Versuchserfolg; ANOVA	91
Tabelle 53: SPSS Ausgabe; Versuchserfolg; Tukey HSD Post-Hoc-Test für ANOVA.....	92

1. Einleitung

1.1. Aufgabenstellung

In der Semesterarbeit sollen die folgenden Arbeitspakete behandelt werden:

- Einarbeitung in die Thematik der Tangible User Interfaces (TUI)
- Festlegen einer geeigneten Vorgehensweise und Erarbeitung des Wissens für die statistische Auswertung
- Statistische Auswertung vorgängig durchgeführter Testserien
- Darstellung und Diskussion der Resultate
- Schriftliche Dokumentation der Arbeit

1.2. Vorgehensweise

Die Vorgehensweise basiert auf den vom Systems Engineering her bekannten Problemlösungszyklus [8], der für komplexe Projekte in mehreren Phasen durchschritten wird. Für die vorliegende Arbeit wurde der Problemlösungszyklus als roter Faden verwendet, jedoch den spezifischen Anforderungen einer eher statistisch geprägten Auswertung angepasst.

Mit der Situationsanalyse werden die vorgängig durchgeführten Testserien und die daraus hervorgegangenen Daten beschrieben. Es findet eine Evaluation von statistischer Software statt. Ausgehend von der Situationsanalyse werden Ziele für das weitere Vorgehen formuliert.

Im Kapitel Theoretische Grundlagen wird die grundsätzliche Vorgehensweise für die statistische Auswertung erarbeitet, mit den dazugehörigen mathematischen Grundlagen.

Danach werden die Daten basierend auf der Vorgehensweise ausgewertet. Es werden die Voraussetzungen für die Auswertung geprüft, verschiedene Vorgehensvarianten werden beurteilt und schliesslich wird eine bestimmte Variante des Vorgehens ausgewählt.

Abschliessend wird die Auswertung mit dieser Variante durchgeführt und die Resultate werden in einer strukturierten Weise dargestellt und diskutiert.

2. Zielsuche

In der Situationsanalyse wird die vorgängige Untersuchungsdurchführung im Vorfeld dieser Arbeit und die daraus hervorgegangenen Daten beschrieben. Ausgehend von der Situationsanalyse werden die Ziele formuliert, die das weitere Vorgehen bestimmen.

2.1. Situationsanalyse

In diesem Kapitel wird die bisherige Untersuchungsdurchführung beschrieben, statistische Software wird evaluiert, die aus der Untersuchung hervorgegangenen Daten werden gesichtet und strukturiert und eine grundsätzliche Methode zur Datenauswertung wird gesucht.

2.1.1. Untersuchungsdurchführung

Im Vorfeld zu dieser Semesterarbeit wurde eine empirische Untersuchung durchgeführt und die Untersuchungsdaten dazu wurden gesammelt.

Bei der Untersuchung wird ein TUI (Tangible User Interface) mit zwei alternativen Tools verglichen und auf seine Brauchbarkeit zur Problemlösung durch Einzelbenutzer hin untersucht. Um die Brauchbarkeit zu messen, wird die kognitive Unterstützung (s. Kapitel 5.1. Kognitive Unterstützung) gemessen, die ein Tool dem Benutzer bietet. Das TUI heisst BUILD-IT und ist ein Tool, das an der ETH Zürich zur kollaborativen, räumlichen Planung entwickelt wurde. Bei den alternativen Tools handelt es sich um einen Oberflächen-Modeler, in der Folge Modeler Tool genannt und um reale Objekte, in der Folge Physical Tool genannt.

Die Motivation für die Brauchbarkeitsuntersuchung ist die Beurteilung der Qualität von BUILD-IT als Planungstool.

2.1.1.1. Aufgabendesign

Um die alternativen Tools für Planungsaktivitäten miteinander vergleichen zu können, muss die zu konzipierende Aufgabe für alle drei in gleichem Masse geeignet sein. Bei Planungsaktivitäten ist beispielsweise eine typische Aufgabe, 3D Modelle in eine bestimmte räumliche Anordnung zu bringen. Normalerweise finden solche Aufgaben unter Zeitdruck statt.

Man entschied sich deshalb für ein Laser-Positionierungsproblem, bei dem Versuchszeit, Anzahl Benutzeroperationen und Benutzerzufriedenheit gemessen werden. Es sollte einer Situation aus der realen Arbeitswelt ähnlich sein.

Die Versuchsteilnehmer haben für alle drei Tools die selbe Aufgabe zu lösen. Der Versuchsaufbau besteht aus den drei Elementen Blöcke (B), Laserpointer (L) und Ziel (T) (s. Abbildung 1). Neun unterschiedlich hohe Blöcke sind auf einer Grundplatte zu einem quadratischen Grundriss angeordnet. Das Ziel hat immer den gleichen Abstand von den Blöcken. Nur der Laserpointer ist frei verschiebbar und wird auf die einzelnen Blöcke gestellt, um zu testen, ob das Ziel getroffen wird. Nur von einem Block aus kann die richtige Lösung gefunden werden. Für die Aufgabe gibt es nur eine Lösung.

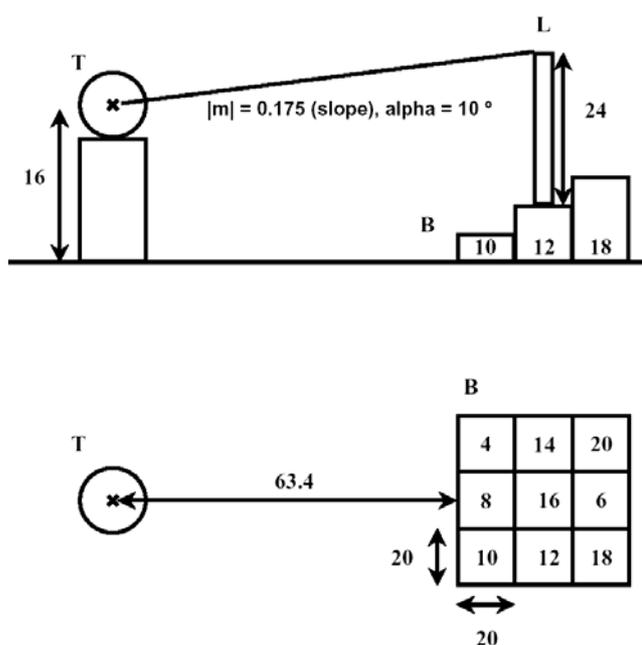


Abbildung 1: Positionierungsproblem

2.1.1.2. Tools

Es gibt die drei Tools BUILD-IT, Modeler und Physical. Diese werden anschliessend kurz beschrieben.

BUILD-IT

BUILD-IT ist ein System zur Unterstützung der räumlichen Planung für Projekte, beispielsweise in den Bereichen Innenarchitektur, Stadtplanung oder Produktionsplanung. Es ist vor allem für die Teamnutzung konzipiert. Die Benutzer können Projekte zusammen an einem Tisch unter der Benutzung von BUILD-IT besprechen. Im Experiment wird die Einzelnutzung und nicht die Teamnutzung von BUILD-IT untersucht.

Das System besteht aus einem Rack mit zwei Beamern und einem Computer (s. Abbildung 2). Ein Beamer projiziert die Draufsicht des virtuellen Modells über einen Spiegel auf den Tisch, der andere Beamer die Perspektive auf eine Leinwand (s. Abbildung 3). Über reale Klötze (im folgenden Bricks genannt) (s. Abbildung 4) kann man in der Draufsicht durch den virtuellen Raum navigieren und die virtuellen Objekte bearbeiten und verschieben. Eine Kamera erfasst die Bewegungen der Bricks auf dem Tisch. Aufgrund dessen werden die Veränderungen des Raumes über Computer und Beamer ständig angepasst.



Abbildung 2: BUILD-IT System mit Spiegel, Tisch und Leinwand



Abbildung 3: Draufsicht auf dem Tisch und Perspektive auf der Leinwand

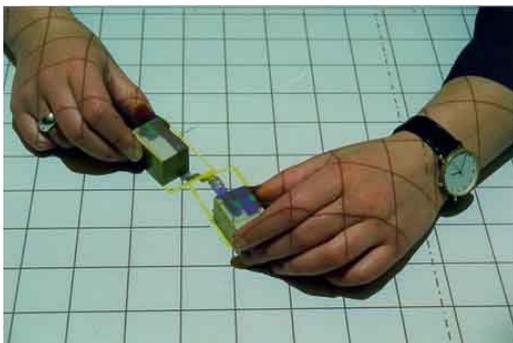


Abbildung 4: Bricks

Blöcke, Laserpointer und Ziel, die Elemente der Aufgabe, sind im BUILD-IT Tool virtuell modelliert [3]. Es ist keine Navigation der Ansichten erlaubt und die Ansichten bleiben in ihrer Position fixiert (s. Abbildung 5). Die Versuchspersonen benutzen einen Brick, um den virtuellen Laserpointer auf einen Block zu verschieben. Draufsicht und Perspektive zeigen, ob der Laserstrahl das Ziel trifft. In Abbildung 5 sind einige typische Vorgehensschritt a) bis d) zum Testen der Blöcke dargestellt. Die linke Spalte zeigt die Draufsicht und die rechte die Perspektive:

- In a) ist eine mögliche Startsituation dargestellt.
- Dann wird in b) ein erster Block getestet. Der Laserstrahl schießt über das Ziel hinaus.
- In c) wird ein zweiter Block getestet. Der Laserstrahl schießt unter dem Ziel hindurch.
- In d) wird schliesslich ein dritter Block getestet. Der Laserstrahl trifft das Ziel.

Die Lösung ist gefunden.

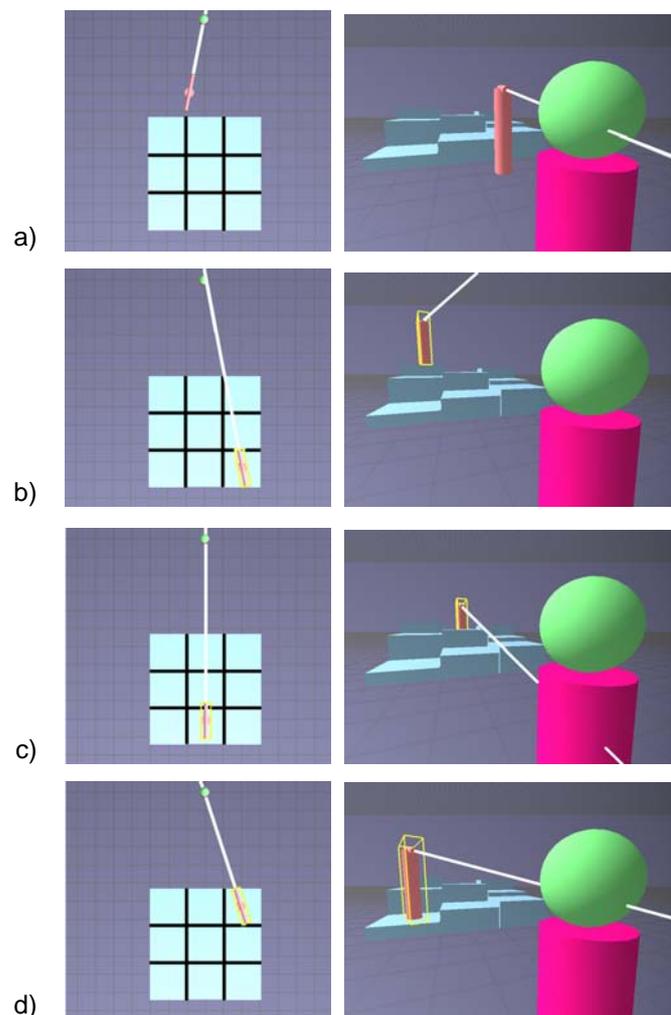


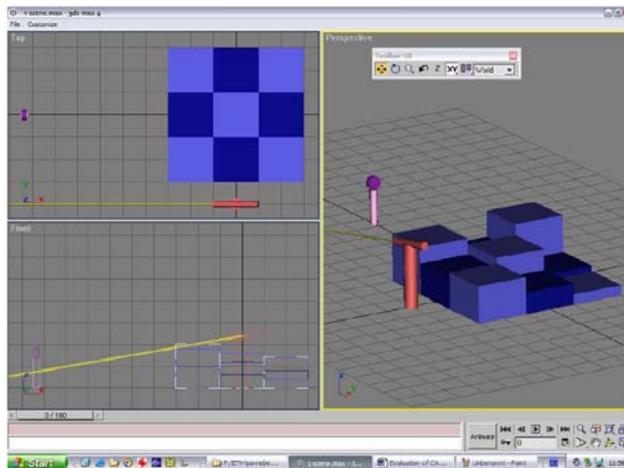
Abbildung 5: BUILD-IT Testumgebung

Modeler

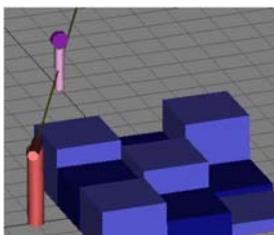
Das Modeler Tool zeigt das Positionierungsproblem im virtuellen Raum (s. Abbildung 6) [4]. Die Modellierung erfolgte mit dem Modeler- und Animationsprogramm „3D Studio Max“. Am Bildschirm werden die drei Ansichten Draufsicht, Seitenansicht und Perspektive gezeigt. Die Interaktion im virtuellen Raum geschieht über die Maus. Der Versuchsperson stehen verschiedene Funktionen zum Navigieren und Positionieren in einer Toolbox zur Verfügung. In Abbildung 6 sind einige typische Situationen und Tools a) bis d) dargestellt:

- a) Bildschirmansicht von 3D Studio Max mit Draufsicht (oben links), Seitenansicht (unten links) und Perspektive (rechts)
- b) und c) sind zwei Testszenarien in der Perspektive

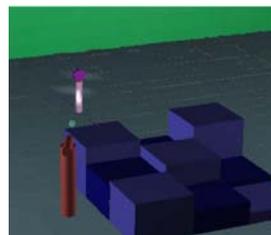
- d) die Toolbox umfasst verschiedene Funktionen zum Navigieren und Positionieren: Verschieben, Rotieren, Umschalten der Ansichtgröße, Undo, Restriktion der Bewegung.



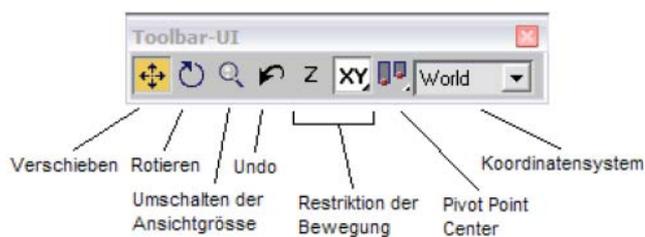
a)



b)



c)



d)

Abbildung 6: Modeler Testumgebung

Physical

Das Physical Tool ist das reale, physische Modell der Aufgabe (s. Abbildung 7) [3]. Die Masse sind gegenüber den Angaben in Abbildung 1 mit einem Faktor vier vergrößert. Der Aufbau besteht aus neun Metallblöcken, die von einer Grundplatte zusammengehalten werden, einer Laserquelle (maximaler Durchmesser: 15 mm, Länge: 50mm) und einem Ziel aus einem Metallstift mit einem 5mm mal 15mm grossen Querstift aus Metall.

Die Versuchsperson greift den Laserpointer mit der Hand, positioniert und rotiert ihn auf verschiedenen Blöcken. Solange die Lösung nicht gefunden ist, deutet

ein Lichtpunkt auf der weissen Wand hinter dem Ziel die Richtung des Laserstrahls an.

In Abbildung 7 sind einige typische Vorgehensschritte a) bis d) zum Testen der Blöcke dargestellt. Der Laserpunkt auf der Wand ist rot hervorgehoben:

- a) zeigt eine mögliche Startsituation.
- Dann wird in b) ein erster Block getestet. Der Laserstrahl schiesst über das Ziel hinaus und trifft die Wand.
- In c) wird ein zweiter Block getestet. Der Laserstrahl schiesst unter dem Ziel hindurch und trifft die Wand.
- In d) wird schliesslich ein dritter Block getestet. Der Laserstrahl trifft das Ziel.

Die Lösung ist gefunden.

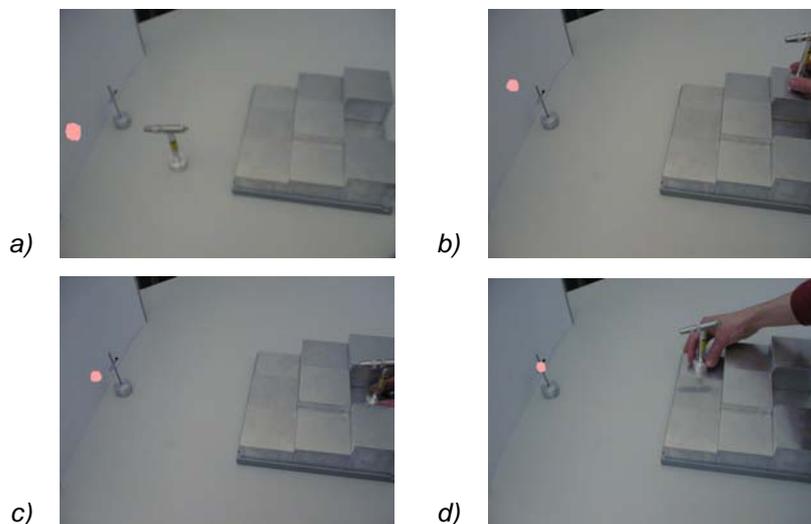


Abbildung 7: Physische Testumgebung

2.1.1.3. Versuchsdesign

Der Versuch wird nach einem „between-subject“ Design durchgeführt. Das bedeutet, dass jedem Tool eine andere Gruppe von Versuchspersonen zugeordnet wird. Damit können Lerneffekte zwischen den Tools ausgeschlossen werden. Jedes Tool wird von 20 Versuchspersonen benutzt, das sind insgesamt 60 Teilnehmer.

Es werden 12 verschiedene Varianten benutzt, um die Blöcke auf der Grundplatte zu positionieren. Eine Variante (immer dieselbe) wird für einen Demonstrationsversuch benutzt. Eine andere Variante (immer dieselbe) wird für einen Versuch mit Hilfe benutzt. Die restlichen zehn Aufgabenvarianten werden für das eigentliche Experiment verwendet, bei dem nicht geholfen wird. Diese zehn Varianten so permutiert, dass nur zwei Versuchspersonen eines Tools die gleiche Reihenfolge haben.

Für jeden Versuch werden die Werte für die Versuchszeit und die Anzahl der getesteten Blöcke notiert. Ebenfalls wird festgehalten, ob die richtige Lösung

gefunden wurde. Nach den Versuchen füllt der Teilnehmer einen Fragebogen (s. Anhang A.5.2.1, Tabelle 27) aus, der für die Evaluation der Benutzerzufriedenheit dient. In Tabelle 1 sind die Bezeichnungen dieser aufgenommenen Werte zusammengefasst. Die Werte werden als Variablen ausgedrückt. In der Folge werden sie zur Vereinfachung nur noch in gekürzter Form verwendet.

Abkürzung der Variable (in der Folge verwendet)	Beschreibung der Variable
Zeit	Zeit pro Versuch
Anzahl Blöcke	Anzahl der getesteten Blöcke pro Versuch
Benutzerzufriedenheit	Benutzerzufriedenheit (bleibt gleich)
Versuchserfolg	prozentualer Anteil der Versuche mit richtiger Lösung pro Tool

Tabelle 1: Variablen

2.1.2. Software

Es ist wichtig, die Software für die statistische Auswertung in einem frühen Stadium zu definieren, da sie für die an das Programm angepasste Strukturierung der Daten und das Kennen der zur Verfügung stehenden Funktionen wichtig ist.

Zur Auswahl als Statistikprogramme stehen Systat, SPSS und Excel.

- *Excel*: Das Programm ist eher ein Aussenseiter und nicht gerade bekannt für Stärken im statistischen Bereich. Es bietet aber trotzdem einige statistische Funktionen, wie z.B. die ANOVA. Es ist jedoch nicht möglich Post-Hoc-Tests durchzuführen. Aus diesem Grund kam das Programm nicht für den Gebrauch in Frage.
- *Systat*: Es ist ein umfangreiches Universal-Statistik-Paket, das einen grossen Teil der bekannten statistischen und graphischen Methoden abdeckt. Fast alle Möglichkeiten des Programms sind durch die Menüsteuerung zugänglich, so dass die Bedienung relativ komfortabel ist.

- *SPSS*: Das Programm gehört zu den meist verwendeten und umfassendsten Statistikprogrammen. Es hat alle wichtigen Funktionalitäten zu den meisten Anwendungsgebieten der Statistik. *SPSS* funktioniert menügesteuert und ist daher nicht schwierig zu bedienen. Ursprünglich wurde *SPSS* vor allem für Anwendungen in den Sozialwissenschaften entwickelt, besitzt heute jedoch Prozeduren zu den meisten Anwendungsgebieten der Statistik.

Da Excel für den Gebrauch nicht in Frage kommt, muss die Entscheidung zwischen den ziemlich ähnlichen Programmen *SPSS* und *Systat* gefällt werden. Da beide Programme an der ETH über *IDES* für Studenten verfügbar sind und alle wichtigen statistischen Funktionen umfassen, die für Auswertung im Rahmen dieser Semesterarbeit notwendig sind, ist es nicht so entscheidend, welches der beiden letztlich gewählt wird. Die Wahl fällt schliesslich, aufgrund der grossen Verbreitung und Bekanntheit, auf *SPSS*. Es wird die *SPSS* Version 13.0.1 verwendet. Excel wird weiterhin für die Verwaltung und den Unterhalt der Datentabellen verwendet, d.h. für Veränderungen im Layout und für Transformationen der Daten.

Im Anhang A.6 befinden sich die Tabellen mit der Originalausgabe von *SPSS*. Die relevanten Daten werden in den Bericht übertragen. Sie sind in die Kapitel 4.1, 4.2 und 5.4 eingeflossen. Ebenfalls im Anhang, und zwar unter A.4, befindet sich eine Zusammenstellung mit den Menübefehlen der verwendeten *SPSS*-Funktionen. Es handelt sich um die Funktionen, mit denen die Ausgabetablen generiert wurden.

2.1.3. Sichtung der Daten

Während dem Versuch wurden alle Informationen in ein Papierformular eingetragen. Daraus ist eine grosse Menge von Daten hervorgegangen. Die Daten von jedem Tool wurden in eine andere Exceltabelle eingetragen, jedoch nicht in einer einheitlichen Form. Für jedes Tool besteht eine Exceltabelle mit einer anderen Form. Beispiele dafür sind:

- Die Personen sind unterschiedlich nummeriert. Die Numerierung beginnt für jedes Tool wieder von vorne.
- Die Versuchsvarianten sind für das Modeler Tool anders bezeichnet als für die anderen beiden.
- Die Zeit ist sowohl in der Form Minuten und Sekunden als auch nur in Sekunden vorhanden.
- Der Versuchserfolg ist sowohl in Englisch als auch in Deutsch angegeben.

2.1.4. Strukturierung der Daten

Die Daten der drei unterschiedlichen Exceltabellen werden in einer Tabelle zusammengeführt und in eine einheitliche Form gebracht. Die Antworten des Fragebogens sind nur handgeschrieben auf den Papierformularen vorhanden. Diese müssen in die Exceltabelle übertragen werden.

Die Daten werden in primäre-, mit dem Versuch zusammenhängende Daten, und sekundäre-, mit dem Fragebogen und der Person zusammenhängende Daten eingeteilt. Die Zeilen sind so strukturiert, dass für jeden Versuch eine neue Spalte verwendet wird, d.h. pro Person gibt es zehn Zeilen für die zehn Versuche.

Die Spalten der *primären* Daten werden folgendermassen bezeichnet (s. Anhang A.5.2.1, Tabelle 25):

<i>Person:</i>	Die Nummerierung für die 60 Versuchspersonen beginnt bei 11 und endet bei 70. Damit sollte verhindert werden, dass später im Statistikprogramm Probleme daraus entstehen, dass die Personen eins bis neun nur eine Ziffer als Bezeichnung haben.
<i>Tool:</i>	Die drei Tools BUILD-IT, Modeler und Physical erhalten die Nummern 1-3: BUILD-IT = 1, Modeler = 2, Physical = 3.
<i>Versuchsvariante:</i>	Die zehn Versuchsvarianten tragen die Nummern 1 bis 10, wobei sie wie schon erwähnt mit lateinischen Quadraten permutiert sind, d.h. pro Tool gibt es nur zwei Personen mit der gleichen Kombination von Versuchsvarianten. Wie sich die Varianten unterscheiden, soll nicht weiter dokumentiert werden.
<i>Versuchsnummer:</i>	Die Versuchsnummern sind innerhalb einer Person chronologisch von T1 bis T10 durchnummeriert.
<i>Anzahl getestete Blöcke:</i>	Es wird die effektive Anzahl der getesteten Blöcke in einem Versuch notiert.
<i>Zeit pro Versuch (sec):</i>	Es wird die Zeit in Sekunden notiert, die für die Lösung einer Versuchsvariante gebraucht wird.
<i>Versuchserfolg (erfüllt?):</i>	Es wird notiert, ob der Versuchsteilnehmer die richtige Lösung gefunden hat oder nicht. Es werden die Nummern 8 und 9 verwendet, für „Ja“ Versuch erfüllt oder „Nein“ Versuch nicht erfüllt: Ja = 8, Nein = 9.

Die Spalten der *sekundären* Daten (Kontrolldaten) werden folgendermassen bezeichnet (s. Anhang A.5.2.1, Tabelle 26):

<i>Geschlecht:</i>	Das Geschlecht der Personen wird mit 6 = männlich und 7 = weiblich bezeichnet.
<i>Alter:</i>	Das Alter ist mit der effektiven Jahreszahl bezeichnet.
<i>Fragen:</i>	Es gibt acht Fragen 1.1–1.3, 2.1–2.3, 3.1 und 3.2 wobei die Frage 1.2 mehrere Antworten zulässt. Jede Frage bildet eine Spalte, wobei Frage 1.2 fünf Spalten für die möglichen Mehrfachnennungen umfasst. Die Antworten sind als Zahlenwerte in die Tabelle eingetragen Die Zahlenwerte umfassen die Skala [-2,-1,0,1,2]. Die Bedeutung der Skala ist für jede Frage anders (Fragebogen s. Anhang A.5.2.1, Tabelle 27).

2.1.5. Methode für die Auswertung

Eine zentrale Bedeutung kommt der Wahl der richtigen Analysemethode zu. Je nachdem welche Eigenschaften die Versuchsdaten aufweisen, kann aus einer Vielzahl von verschiedenen Methoden die richtige gewählt werden (s. Abbildung 8). Die entscheidenden Eigenschaften sind:

- Normalverteilung
- Anzahl der Gruppen
- Abhängige Daten
- Homogene Varianz

Die Daten der Untersuchung umfassen drei Gruppen, nämlich die drei Tools, und, wie später noch gezeigt wird, sind sie nicht abhängig. Bezogen auf Abbildung 8, grenzt dies den Fokus auf zwei mögliche Methoden ein, unabhängig davon, ob die Daten normalverteilt und die Varianzen homogen sind oder nicht. Die beiden Methoden sind die „Varianzanalyse für unabhängige Stichproben“ (im Folgenden ANOVA (analysis of variance) genannt) und die „Kruskal-Wallis Rangvarianzanalyse“. Später wird auch noch gezeigt, dass die ANOVA verwendet werden kann, da die Voraussetzungen dafür erfüllt sind.

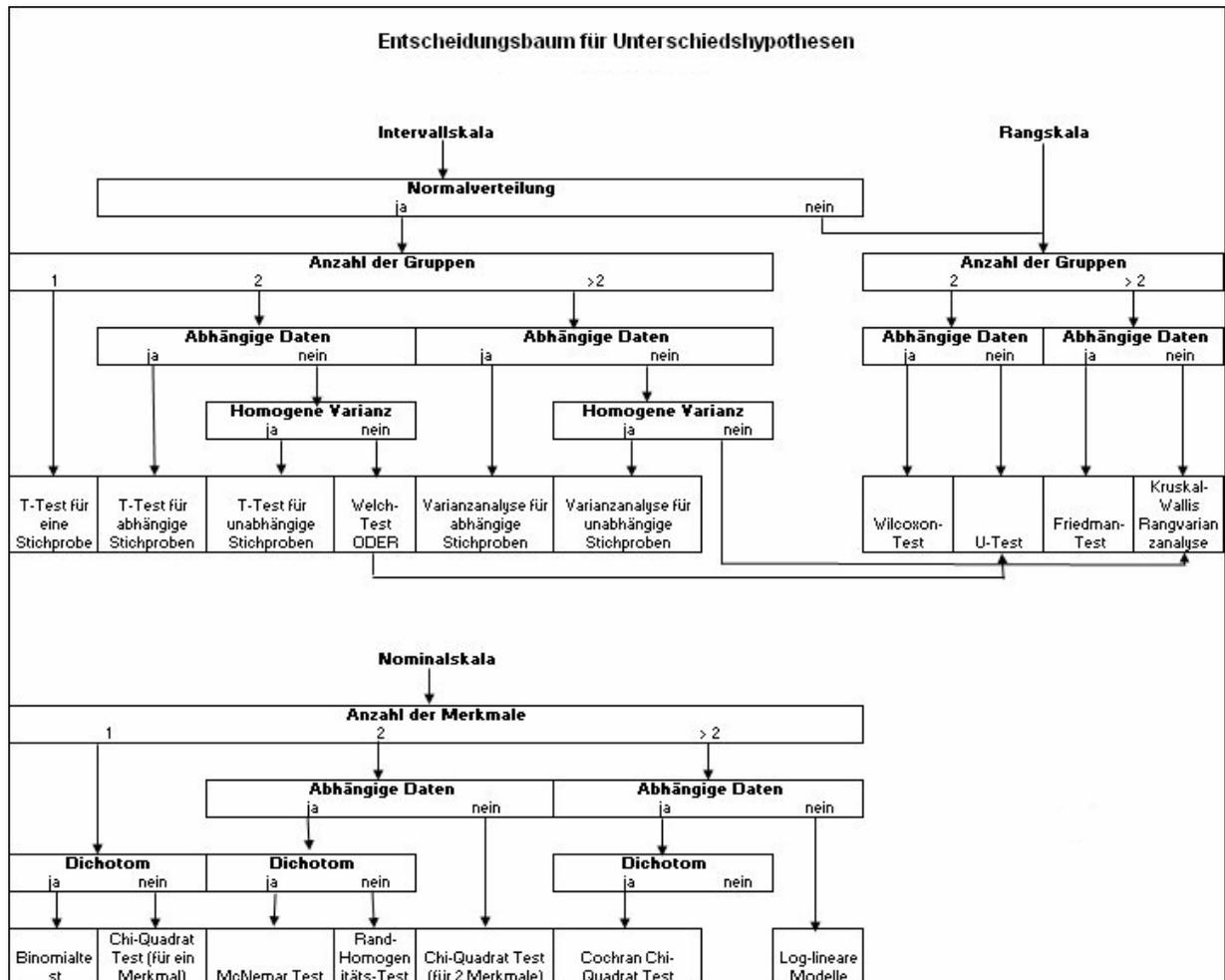


Abbildung 8: Entscheidungsbaum für Unterschiedshypothesen

Die ANOVA ist ein statistisches Verfahren der Datenanalyse, das versucht, den Einfluss einer unabhängigen Variablen auf die abhängige Variable zu überprüfen. Das Verfahren untersucht, ob (und gegebenenfalls wie) sich der Mittelwert einer abhängigen Variablen in verschiedenen Gruppen unterscheidet. In Prüfgrößen des Verfahrens wird getestet, ob die Varianz zwischen den Gruppen größer ist als die Varianz innerhalb der Gruppen. Dadurch kann ermittelt werden, ob sich die Gruppen signifikant unterscheiden oder nicht [1].

Die Daten der Untersuchung umfassen unabhängige-, abhängige- und Kontrollvariablen:

unabhängige Variable:

- Person
- Tool
- Versuchsvariante
- Versuchsnummer

abhängige Variablen: - Zeit

- Anzahl Blöcke
- Benutzerzufriedenheit (wird später eingeführt)
- Versuchserfolg

Kontrollvariablen:

- Geschlecht
- Alter
- Fragen (fließen teilweise in Benutzerzufriedenheit ein)

2.2. Zielformulierung

Aus der Situationsanalyse ergeben sich Ziele, die das weitere Vorgehen bestimmen sollen:

- Eine statistische Methode für die weitere Auswertung festlegen.
- Das detaillierte Vorgehenskonzept der statistischen Methode erarbeiten.
- Die Voraussetzungen für die statistische Methode überprüfen.
- Eine Vorgehensvariante auswählen.
- Die ausgewählte Variante mit der statistischen Methode analysieren.
- Die abhängigen Variablen sinnvoll mit der kognitiven Unterstützung verbinden.
- Die Unterschiede der Tools in der Performance (Zeit und Anzahl Blöcke) und in der Benutzerzufriedenheit auf Signifikanz überprüfen.
- Den Einfluss von Lerneffekten ausarbeiten.
- Den Einfluss der Kontrollvariablen (sekundäre Daten) aufzeigen.

3. Theoretische Grundlagen

Wie schon im Kapitel 2.1.5 beschrieben, kommen für die Daten grundsätzlich die zwei Analysemethoden ANOVA und Kruskal-Wallis Test in Frage, wobei die ANOVA dem Kruskal-Wallis Test vorzuziehen ist. Da vorweggenommen werden kann, dass die Daten die Voraussetzungen für die ANOVA erfüllen, wird der Kruskal-Wallis Test im weiteren Verlauf nicht behandelt.

In Kapitel 3.1 wird die grundsätzliche Vorgehensweise für die Auswertung der Daten mit ANOVA beschrieben. In Kapitel 3.2 werden gewisse mathematische Ausdrücke genauer erläutert, die in der Beschreibung des Vorgehens ohne tiefere Erklärung benutzt werden.

Die Vorgehensweise wird in den Kapiteln 4.1, 4.2, 4.3 und 5.4 umgesetzt.

3.1. Vorgehensweise

Für die ANOVA wird in ganz bestimmten Schritten vorgegangen. In diesem Kapitel soll das genaue Vorgehen beschrieben werden. Eine eingehendere Erläuterung der in diesem Kapitel verwendeten mathematischen- und statistischen Ausdrücke findet sich dann im folgenden Kapitel 3.2.

Für die Anwendung der ANOVA ist es zentral, dass drei Voraussetzungen erfüllt sind [1]:

- Unabhängigkeit der Probenwerte
- Normalverteilung der Probenwerte innerhalb der Gruppen
- Homogenität der Varianzen der Probenwerte

Durch die Schritte 1-6 werden die Daten auf diese Voraussetzungen hin überprüft und gegebenenfalls durch Transformationen angepasst. Diese können bewirken, dass sich Verteilungen der Normalverteilung annähern und sich die Homogenität der Varianzen verbessert. Als Transformationen werden die Quadratwurzel, der natürliche Logarithmus und der dyadische Logarithmus verwendet. Die Schritte 7 und 8 umfassen die eigentliche ANOVA. In Klammern wird auf die Kapitel verwiesen, wo der jeweilige Schritt umgesetzt wird:

1. Untransformierte Daten (Kapitel 4.1.1.1 und 4.1.2.1)

In einem ersten Schritt werden die ursprünglichen Daten unverändert anhand von Boxplot (s. Kapitel 3.2.1), eine Methode zur graphischen Darstellung von

statistischen Verteilungen, und Standardabweichung (s. Kapitel 3.2.2) untersucht, obwohl die Unabhängigkeitsvoraussetzung für die ANOVA nicht erfüllt ist. Es besteht Abhängigkeit innerhalb der Versuchswiederholungen einer Person, weil die Person zehn Versuche mit demselben Tool zur grundsätzlich selben Aufgabe wiederholt.

Dieser unnötig erscheinende Schritt wird ausgeführt, um einen ersten Eindruck von den Daten zu bekommen und spätere Verbesserungen durch Transformation der Daten besser erkennen zu können.

Der Boxplot dient der Überprüfung auf Normalverteilung (s. Kapitel 3.2.3). Mit der Standardabweichung wird die Homogenität der Varianzen geprüft [Für die Untersuchung der Homogenität der Varianzen wird die Standardabweichung benutzt. Dies ist problemlos möglich, da die Standardabweichung als die Quadratwurzel aus der Varianz definiert ist.]

2. Arithmetisches Mittel (Kapitel 4.1.1.2 und 4.1.2.2)

Im nächsten Schritt werden die Daten pro Person mit Hilfe des arithmetischen Mittels (s. Kapitel 3.2.4) zu einem Wert zusammengefasst.

Es wird das arithmetische Mittel für die Zeit und die Anzahl Blöcke aus den Werten von Versuchsnummer T1-T10 jeder Person berechnet (s. Abbildung 9). Pro Tool gibt es danach noch je 20 Werte für Zeit und Anzahl Blöcke, anstatt 60 zuvor; pro Person noch einen anstatt 10.

Person	Tool	Versuchsnummer	Anzahl getestete Blöcke	Zeit pro Versuch (sec)
11	1	T1	4	14
11	1	T2	1	11
11	1	T3	1	8
11	1	T4	1	7
11	1	T5	5	28
11	1	T6	6	32
11	1	T7	1	9
11	1	T8	5	28
11	1	T9	4	25
11	1	T10	11	61

Person	Tool	Arithmetisches Mittel Zeit	Arithmetisches Mittel Anzahl Blöcke
11	1	22.3	3.9

Abbildung 9: Berechnung des arithmetischen Mittels

Durch die Zusammenfassung der Werte sind die zugrundeliegenden Daten einer einzelnen Person zwar immer noch abhängig. Für jede Person besteht aber nur noch ein Wert und diese Werte der Personen untereinander sind unabhängig.

Gleichzeitig kann durch das arithmetische Mittel die weitgestreute Verteilung der Werte eingedämmt und geglättet werden. Variiert beispielsweise die Zeit einer Person von sehr lange bis sehr kurz, dann gleichen sich diese beiden Extremwerte aus und führen zu einem moderateren Durchschnittswert. Hat eine Person jedoch durchwegs lange für die Lösung der Aufgabe gebraucht, dann hilft auch das arithmetische Mittel zur Ausgleichung der Extremwerte nichts.

Wiederum dient der Boxplot der Überprüfung auf Normalverteilung und die Standardabweichung der Überprüfung auf Homogenität der Varianzen.

3. Transformationen (Kapitel 4.1.1.3 und 4.1.2.3)

Im Fall, dass die Daten die Anforderungen an die ANOVA nach der Berechnung des arithmetischen Mittels noch nicht erfüllen, sind Transformationen (s. Kapitel 3.2.5) zu benutzen. Durch nichtlineare Transformationen können Verteilungen verändert werden. Beispielsweise lassen sich Schiefe (s. Kapitel 3.2.6) und Kurtosis (s. Kapitel 3.2.7) verringern. Die beiden Masse charakterisieren Symmetrie und Form einer Verteilung.

Die abhängigen Variablen Zeit und Anzahl Blöcke werden ausgehend von den Werten des arithmetischen Mittels nichtlinear transformiert. Es werden drei verschiedene Transformationen angewendet: die Quadratwurzel ($\sqrt{}$) (s. Kapitel 3.2.8), der natürliche Logarithmus (\ln) (s. Kapitel 3.2.9) und der dyadische Logarithmus ($2\log$) (s. Kapitel 3.2.10). Es geht um die Frage, welche der Transformationen für welche abhängige Variable am geeignetsten ist. Für die zwei Variablen können zwei unterschiedliche Transformationen resultieren.

4. Levene- und Kolmogorov-Smirnov-Test (Kapitel 4.1.2.3)

Nachdem die Daten transformiert sind, wird mit dem Levene- (s. Kapitel 3.2.11) und dem Kolmogorov-Smirnov-Test (s. Kapitel 3.2.12) geprüft, ob sie die Homogenität der Varianzen und die Annahme der Normalverteilung erfüllen [6].

5. Beurteilung der Transformationen (Kapitel 4.2)

Daraufhin werden die Transformationen anhand von drei Kriterien miteinander verglichen. Für den Vergleich dienen die drei Kriterien Boxplot, Standardabweichung und Histogramm (s. Kapitel 3.2.13). Für jedes Kriterium werden die Transformationen bewertet und eine Rangliste wird erstellt, von Rang eins für die am besten geeignete Transformation bis Rang drei für die am schlechtesten geeignete Transformation.

5.1 Boxplot

Mit dem Boxplot wird beurteilt, ob die Verteilung symmetrisch ist und damit letztlich auch, ob Normalverteilung angenommen werden kann.

5.2 Standardabweichung

Mit dem Vergleich der Standardabweichung wird beurteilt, ob die Standardabweichungen der Tools untereinander homogen sind.

5.3 Histogramm

Mit dem Histogramm wird noch einmal, dieses mal aber direkt beurteilt, ob die Verteilung mit der Normalverteilungskurve näherungsweise übereinstimmt.

6. Entscheidung über Transformationen (Kapitel 4.3)

Aus den drei beurteilten Transformationen sollen nun, in einer Gesamtbewertung, diejenigen ausgewählt werden, welche die Kriterien am besten erfüllen. Zu diesem Zweck werden die bisherigen Ränge der Kriterien Boxplot, Standardabweichung und Histogramm zu einem neuen Punktestotal addiert. Aus diesem Punktestand resultiert der Gesamtrang für jede Transformation.

Die geeignetste Transformation für jede der beiden abhängigen Variablen wird ausgewählt und für die weitere Analyse mit ANOVA verwendet.

7. ANOVA (Kapitel 5.4.2)

Nun werden die abhängigen Variablen in transformierter Form (nur die zwei ausgewählten Transformationen) mit ANOVA (s. Kapitel 3.2.14) untersucht.

8. Post-Hoc-Test (Kapitel 5.4.3)

Anschliessend wird ein Post-Hoc-Test (s. Kapitel 3.2.15) für die Bestimmung der signifikanten Unterschiede zwischen den Tools durchgeführt.

3.2. Mathematische Grundlagen

Dieses Kapitel liefert die mathematischen Grundlagen zum vorhergehenden Kapitel 0 über die Vorgehensweise. Es werden mathematische Definitionen und Erklärungen für bisher verwendete Ausdrücke geliefert, deren Bedeutung noch nicht ausreichend klar ist. Jedes Unterkapitel behandelt einen Ausdruck. Die Unterkapitel sind nach der Reihenfolge des Auftretens der Ausdrücke in Kapitel 0 geordnet.

3.2.1. Boxplot

Der Boxplot (s. Abbildung 10) ist eine graphische Methode, um statistische Verteilungen von Daten darzustellen und untereinander zu vergleichen. Es ist es möglich, einen ersten Gesamteindruck über die Daten und Hinweise über die wichtigsten Eigenschaften einer Verteilung wie Lage, Median, Streuung (s. Kapitel 3.2.2 zur Standardabweichung und Kapitel 3.2.7 zur Kurtosis), Symmetrie, Schiefe (s. Kapitel 3.2.6) zu erhalten. Es kann sehr schnell ein visueller Eindruck gewonnen werden, ob die Verteilung symmetrisch ist oder ob Extremwerte vorliegen. Die Abstände zwischen Median und oberer bzw. unterer Begrenzung der Box geben Aufschluss darüber, ob die Verteilung symmetrisch oder schief ist. Die Form eines Boxplots gibt auch Hinweise auf die Art der Verteilung in den gegenübergestellten Teilstichproben, unter anderem auf Abweichungen von der Normalverteilung [10].

Ein Box-Plot besteht aus einem Kasten (Box) und zwei Strecken oben und unten von der Box. Die Darstellung basiert auf den fünf Kenngrößen Median, unteres Quartil, oberes Quartil, unteres Percentil und oberes Percentil.

Der Median wird als Querstrich innerhalb der Box dargestellt. Es ist der Wert, über und unter dem jeweils die Hälfte der Fälle liegt. Analog dazu sind 25% der Fälle kleiner als das untere Quartil und 25% der Fälle sind grösser als das obere Quartil. 10% der Fälle sind kleiner als das untere Percentil und 10% der Fälle sind grösser als das obere Percentil.

Die Länge der Box umfasst die Werte zwischen dem unteren- und dem oberen Quartil, auch Interquartilabstand (IQR) genannt. Demnach beinhaltet die Box 50% aller Werte.

Ausreisser und Extremfälle sind klar erkennbar, sie liegen ausserhalb der kleinen Querstriche (Whiskers). Ausreisser sind zwischen $1,5$ und $3 \cdot \text{IQR}$ vom oberen oder unteren Rand der Box entfernt und werden als Kreise dargestellt. Extremfälle sind mehr als $3 \cdot \text{IQR}$ vom oberen oder unteren Rand entfernt und werden als Sterne dargestellt [6] [10].

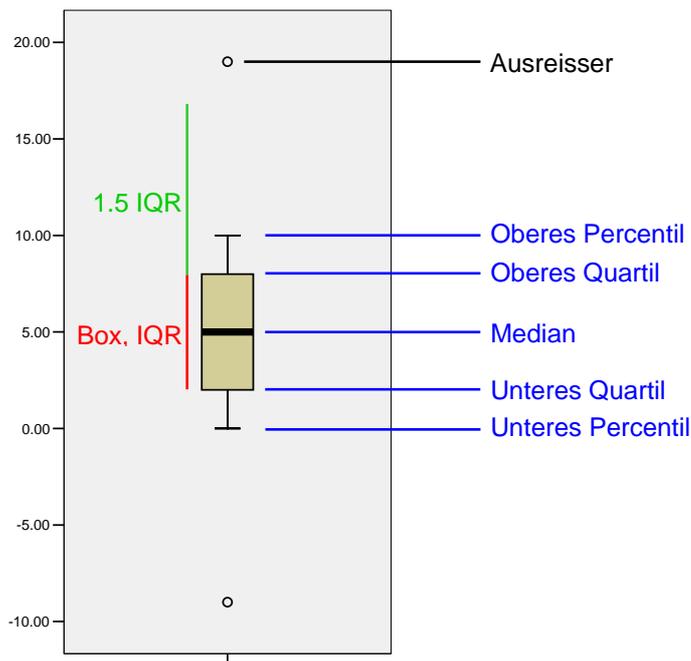


Abbildung 10: Boxplot

3.2.2. Standardabweichung

Die Standardabweichung ist ein Mass für die Streuung der Werte einer Variablen um das arithmetische Mittel (s. Kapitel 3.2.4). Sie ist definiert als die Quadratwurzel aus der Varianz [9].

Die Standardabweichung ist mathematisch definiert als [7]:

$$s := \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$$

- n Anzahl der Werte
- x_i das i -te Element in der Menge der Werte
- \bar{x} das arithmetische Mittel (für Formel s. Kapitel 3.2.4)

3.2.3. Normalverteilung

Die Normalverteilung ist eine symmetrische, stetige Verteilung um das arithmetische Mittel (s. Kapitel 3.2.4) herum. Sie ist durch das arithmetische Mittel und die Standardabweichung vollständig definiert. Die Kurve (s. Abbildung 11) ihrer Dichtefunktion ist glockenförmig [12].

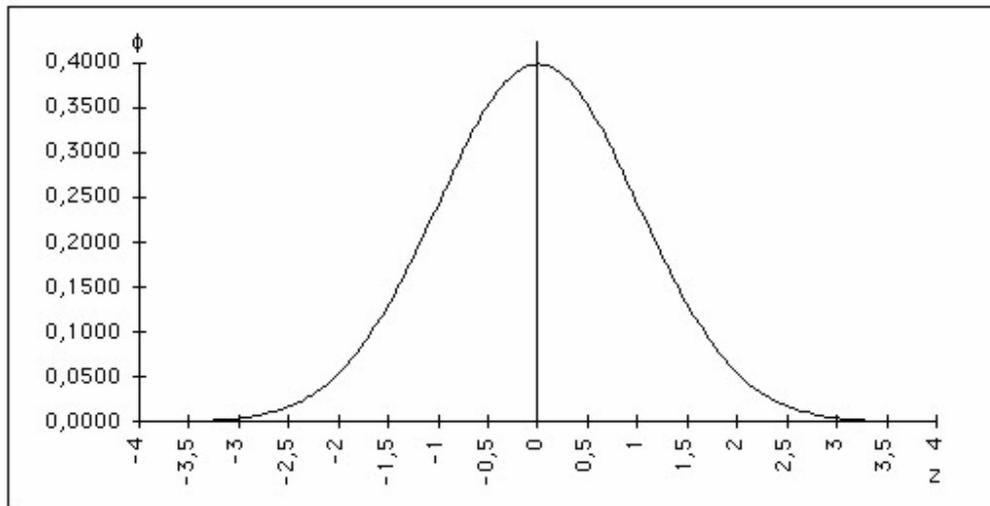


Abbildung 11: Normalverteilungskurve

3.2.4. Arithmetisches Mittel

Das Arithmetische Mittel wird folgendermassen berechnet [7]:

$$\bar{x}_{arithm} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{x_1 + x_2 + \dots + x_n}{n}$$

- n Anzahl der Werte
- x_i das i -te Element in der Menge der Werte

Das bedeutet, dass die Summe der Einzelwerte der Daten durch die Anzahl der Einzelwerte dividiert wird.

3.2.5. Transformationen

Viele statistische Verfahren stellen gewisse Anforderungen an die Verteilung von Zufallsvariablen, wie die Normalität der Daten und die Homogenität der Varianzen. Geringfügige Abweichungen von der Normalverteilung, in der Form einer leichten Links- oder Rechtsschiefe (s. Kapitel 3.2.6) oder einer Kurtosis (s. Kapitel 3.2.7), lassen sich mittels geeigneter nichtlinearer Transformationen beseitigen. Gleichzeitig verbessert sich häufig auch die Homogenität der Varianzen.

Einige nichtlineare Transformation (wobei x der zu transformierende Wert ist) und ihre Wirkung auf die Schiefe sind:

- $x^3 \rightarrow$ reduziert extreme Linksschiefe
- $x^2 \rightarrow$ reduziert Linksschiefe
- $\ln(x) \rightarrow$ reduziert Rechtsschiefe

Durch die Transformation der Daten gehen Informationen verloren. Dies muss gezwungenermassen in Kauf genommen werden, da die Daten sonst nicht ausgewertet werden könnten. Weil die Daten jedes Tools gleich transformiert werden, können sie immer noch relativ zueinander verglichen werden. Die absoluten Zahlenwerte sind aber nicht mehr aussagekräftig.

In dieser Arbeit werden für die statistische Auswertung drei verschiedene Transformationen angewendet: die Quadratwurzel (s. Kapitel 3.2.8), der natürliche Logarithmus (s. Kapitel 3.2.9) und der dyadische Logarithmus (s. Kapitel 3.2.10).

3.2.6. Schiefe

Die Schiefe ist ein Mass für die Asymmetrie einer Verteilung. Die Normalverteilung ist symmetrisch und hat keine Schiefe, somit ist der Wert der Schiefe gleich null (s. Abbildung 12, rechts). Eine Verteilung mit einer deutlich positiven Schiefe (Rechtsschiefe) läuft nach rechts lang aus (langer rechter Schwanz) (s. Abbildung 12, links), eine Verteilung mit einer deutlich negativen Schiefe (Linksschiefe) läuft nach links lang aus (langer linker Schwanz) (s. Abbildung 12, Mitte). Als Faustregel kann man verwenden, dass ein Schiefe-Wert, der mehr als doppelt so gross ist wie sein Standardfehler für eine Abweichung von der Symmetrie spricht [12].

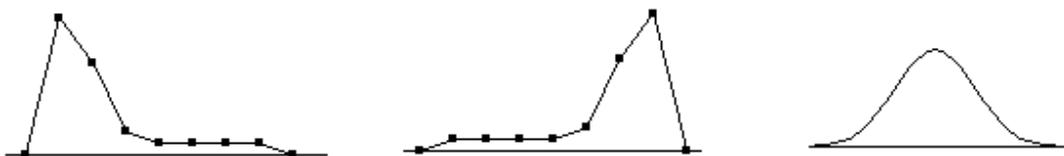


Abbildung 12: Beispiele von Verteilungen mit Rechtsschiefe (links), Linksschiefe (Mitte) und keiner Schiefe (symmetrische Verteilung) (rechts)

Die Formel für die Berechnung der Schiefe lautet [11]:

$$g_1 = \sum_{i=1}^n \frac{(x_i - \bar{x})^3}{s^3}$$

- n Anzahl der Werte
- x_i das i -te Element in der Menge der Werte
- \bar{x} das arithmetische Mittel (für Formel s. Kapitel 3.2.4)
- s die Standardabweichung

3.2.7. Kurtosis

Die Kurtosis ist ein Mass dafür, wie sich die Beobachtungen um einen zentralen Punkt gruppieren, man könnte auch sagen, ein Mass für die Steilheit oder Wölbung einer Verteilung. Die Normalverteilung hat definitionsgemäss keine Kurtosis, somit ist der Wert der Kurtosis gleich null. Bei positiver Kurtosis gruppieren sich die Beobachtungen dichter als bei der Normalverteilung und haben längere Flügel (s. Abbildung 13, links). Bei negativer Kurtosis gruppieren sich die Beobachtungen weniger dicht zusammen und haben kürzere Flügel (s. Abbildung 13, rechts) [12].



Abbildung 13: Beispiele von Verteilungen mit positiver (rechts) und negativer (links) Kurtosis

Die Formel für die Berechnung der Kurtosis lautet [11]:

$$g_2 = \sum_{i=1}^n \frac{(x_i - \bar{x})^4}{s^4 - 3}$$

- n Anzahl der Werte
- x_i das i -te Element in der Menge der Werte
- \bar{x} das arithmetische Mittel (für Formel s. Kapitel 3.2.4)
- s die Standardabweichung

3.2.8. Quadratwurzel

Das mathematische Symbol für die Quadratwurzel aus x lautet [7]:

$$\sqrt{x} = x^{\frac{1}{2}}$$

Unter der Quadratwurzel einer Zahl x versteht man eine Zahl, deren Quadrat gleich der gegebenen Zahl x ist. Im Folgenden wird die Quadratwurzel mit *sqrt* (square root) abgekürzt.

3.2.9. Natürlicher Logarithmus

Das mathematische Symbol für den natürlichen Logarithmus aus x lautet [7]:

$$\ln(x)$$

Es ist der Logarithmus von x zur Basis e . Im Folgenden wird der natürliche Logarithmus mit *ln* abgekürzt.

3.2.10. Dyadischer Logarithmus

Das mathematische Symbol für den dyadischen Logarithmus, oder auch Zweier-Logarithmus, lautet [7]:

$$ld(x) = \log_2(x)$$

Es ist der Logarithmus von x zur Basis 2. Im Folgenden wird der dyadische Logarithmus mit *2log* abgekürzt.

Die Formel zur Berechnung des Logarithmus von x zu einer beliebigen Basis b lautet [7]:

$$\log_b(x) = \frac{\log_a(x)}{\log_a(b)}$$

3.2.11. Levene-Test

Der Levene-Test prüft zwischen allen abhängigen Variablen, ob die Homogenität der Varianzen erfüllt ist. Dieser Test setzt keine Normalverteilung voraus.

Ein p -Wert (Irrtumswahrscheinlichkeit) < 0.05 zeigt eine signifikante Abweichung an und bedeutet, dass die Homogenität der Varianzen abgelehnt werden muss.

ANOVA ist robust gegen diese Verletzung der Voraussetzung der Homogenität der Varianzen, wenn die Gruppen gleiche oder fast gleiche Grösse aufweisen. Trotzdem ist es besser die Daten in diesem Fall zu transformieren [6].

3.2.12. Kolmogorov-Smirnov-Test

Der Kolmogorov-Smirnov-Test prüft, ob die Daten einer Gruppe normalverteilt sind. Es wird die beobachtete kumulative Verteilungsfunktion einer Variable mit einer festgelegten theoretischen Verteilung verglichen, in diesem Fall einer Normalverteilung.

Ein p-Wert (Irrtumswahrscheinlichkeit) < 0.05 zeigt eine signifikante Abweichung an und bedeutet, dass die Normalverteilung abgelehnt werden muss.

Mit diesem Test für die Güte der Anpassung wird getestet, ob die Beobachtung wahrscheinlich aus der Normalverteilung stammt [6].

3.2.13. Histogramm

Ein Histogramm (s. Abbildung 14) ist die graphische Darstellung der Verteilung einer Variablen. Die Daten werden der Größe nach geordnet und in Intervalle aufgeteilt. Über jedem Intervall wird ein Rechteck errichtet. Die Höhe des Rechtecks gibt die Anzahl der Werte (Häufigkeit) der Variablen wieder, die innerhalb des Intervalls liegen [9].

Eine zusätzlich über das Histogramm gelegte Normalverteilungskurve (wie in Abbildung 14) erleichtert die Beurteilung, ob die Daten normalverteilt sind. Die Normalverteilungskurve weist denselben Mittelwert und dieselbe Standardabweichung auf wie die Daten.

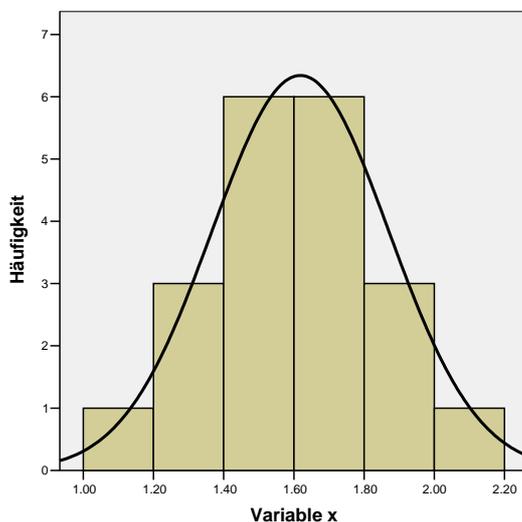


Abbildung 14: Histogramm mit Normalverteilungskurve

3.2.14. ANOVA

Mit der ANOVA wird die Hypothese überprüft, dass die Mittelwerte *mehrerer* Gruppen gleich sind. Ist der p-Wert (Irrtumswahrscheinlichkeit) kleiner als ein zuvor festgelegtes Signifikanzniveau (häufig $\alpha = 0.05$), dann wird die Hypothese verworfen [6]. Ein Signifikanzniveau $\alpha = 0.05$ bedeutet, dass man sich bei dem gefundenen Ergebnis mit einer Wahrscheinlichkeit von 95% nicht geirrt hat. Dieses Verfahren ist eine Erweiterung des t-Tests. Mit dem t-Test wird die Hypothese überprüft, dass die Mittelwerte *zweier* Gruppen gleich sind [2].

Gibt es mehr als zwei Gruppen zu vergleichen, dann ist es auch möglich t-Tests zu benutzen, um jede Gruppe mit allen anderen zu vergleichen. Es kann jedoch durch die mehrfache Verwendung von t-Tests zu einer Zunahme der Fehlerrate 1. Art (d.h. irrtümliches Verwerfen einer gültigen Hypothese) kommen. ANOVA hingegen kann die Unterschiede zwischen mehreren Gruppen auf Signifikanz testen, ohne die Fehlerrate 1. Art zu erhöhen [12].

Für die Anwendung von ANOVA müssen gewisse Voraussetzungen erfüllt sein [1]:

- Unabhängigkeit der Probenwerte
- Hinreichende Normalverteilung der Probenwerte innerhalb der Gruppen (→ Kolmogorov-Smirnov-Test)
- Homogenität der Varianzen der Probenwerte (→ Levene-Test)

Die Homogenität der Varianzen der Probenwerte kann mit dem Levene-Test (s. Kapitel 3.2.11) überprüft werden. Die Normalverteilung der Probenwerte kann mit dem Kolmogorov-Smirnov-Test (s. Kapitel 3.2.12) überprüft werden. Zur Überprüfung der Verteilungsform werden auch graphische Darstellungen der Verteilung in Form eines Histogrammes oder eines Boxplots hinzugezogen.

Die Varianzanalyse ist grundsätzlich unempfindlich gegenüber Abweichungen von der Normalverteilung. Die Gruppen müssen jedoch eine symmetrische Verteilungen und gleiche Varianzen aufweisen. Die Homogenität der Varianzen wird mit Hilfe des Levene-Tests überprüft [6].

Zusätzlich zur Feststellung, dass Differenzen zwischen den Mittelwerten von Gruppen vorhanden sind, kann auch bestimmt werden, welche Mittelwerte abweichen. Für den Vergleich von Mittelwerten gibt es Post-Hoc-Tests (s. Kapitel 3.2.15). Diese werden nach der ANOVA ausgeführt.

3.2.15. Post-Hoc-Tests

Um herauszufinden, welche Mittelwerte voneinander verschieden sind, bzw. welche Gruppen sich von den anderen signifikant unterscheiden, muss im Anschluss an die ANOVA ein Post-Hoc-Test durchgeführt werden. Mit paarweisen Mehrfachvergleichen wird die Differenz zwischen gepaarten Mittelwerten getestet. Wenn die Mittelwerte auf einem α -Niveau von weniger als 0.05 voneinander abweichen, dann weisen sie einen signifikanten Unterschied auf [5].

Je nachdem, wie sicher, bzw. konservativ, dieser Test sein soll (z.B.: Ausschluss von Fehlern 1. Art, d.h. irrtümliches Verwerfen einer gültigen Hypothese), gibt es verschiedene Post-Hoc-Tests: unter anderen Tukey HSD, Newman-Keuls, Duncan oder Scheffé Test. Konservativ sind Tests dann, wenn erst grosse Effekte signifikant werden, radikalere (d.h. weniger konservative) Tests weisen bereits kleinere Effekte als signifikant aus.

Der Newman-Keuls und der Duncan Test sind weniger konservativ, akzeptieren also mehr Fehlermöglichkeiten als beispielsweise der Tukey HSD Test. Der Scheffé Test findet vorzugsweise bei komplexen Vergleichen Anwendung [6].

In dieser Arbeit wird als Post-Hoc-Test der *Tukey HSD (honestly significant difference) Test* verwendet, da er einen Mittelweg zwischen sehr konservativen und radikalen Tests bietet.

4. Auswertung der Messreihen

In diesem Kapitel werden die Vorgehensschritte 1 bis 6 aus Kapitel 3.1 durchgeführt.

4.1. Prüfen der Voraussetzungen

In diesem Kapitel werden die Daten auf das Erfüllen der Voraussetzungen der ANOVA hin überprüft und bei Nichterfüllen mit Transformationen angepasst. Das Vorgehen orientiert sich an den in Kapitel 3.1 beschriebenen Vorgehensschritten 1 bis 4. Das Kapitel ist in zwei Abschnitte unterteilt. In einem Teil wird mit fehlerhaften Daten gearbeitet. Die Daten werden dann kontrolliert und korrigiert. Im anderen Teil werden die korrekten Daten verwendet. Es wird gezeigt, wie die fehlerhaften Daten durch die Analyse entdeckt wurden.

4.1.1. Fehlerhafte Daten

4.1.1.1. Untransformierte Daten

Boxplots

Die Boxplots sind noch weit von der Idealform entfernt (s. Abbildung 15). In der Idealform ist der Boxplot symmetrisch. Beim Boxplot für die Zeit hat es noch überdurchschnittlich viele Extremfälle und Ausreisser. Die Werte für die Anzahl Blöcke sind im Vergleich zur Zeit schon weit besser verteilt, wobei auch noch kein Idealzustand erreicht ist.

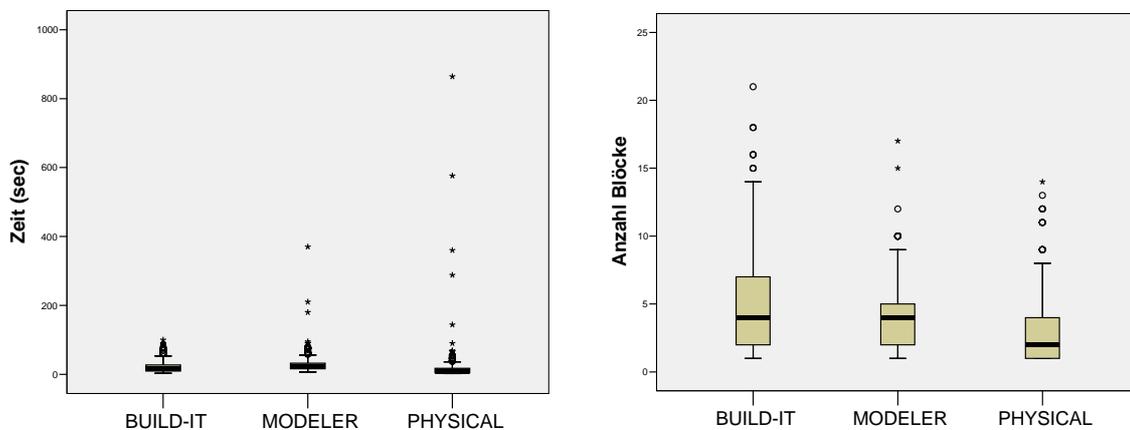


Abbildung 15: Boxplots Zeit und Anzahl Blöcke; fehlerhafte untransformierte Daten

Standardabweichung

Die Standardabweichung zwischen den einzelnen Tools variiert vor allem für die Zeit noch sehr stark (s. Tabelle 2).

Tool	Zeit (sec)	Anzahl Blöcke
BUILD-IT	18.09	4.09
Modeler	34.00	2.54
Physical	79.72	2.66

Tabelle 2: Standardabweichungen; fehlerhafte untransformierte Daten

4.1.1.2. Arithmetisches Mittel

Boxplots

Im Vergleich zu vorher hat sich vor allem die Zahl der Ausreisser massiv verkleinert (s. Abbildung 16). Die Boxplots für die Anzahl Blöcke sehen in Bezug auf die Symmetrie schon relativ ansprechend aus. Vor allem der Boxplot für BUILD-IT kommt der Idealform schon nahe. Alle Ausreisser bis auf einen sind verschwunden.

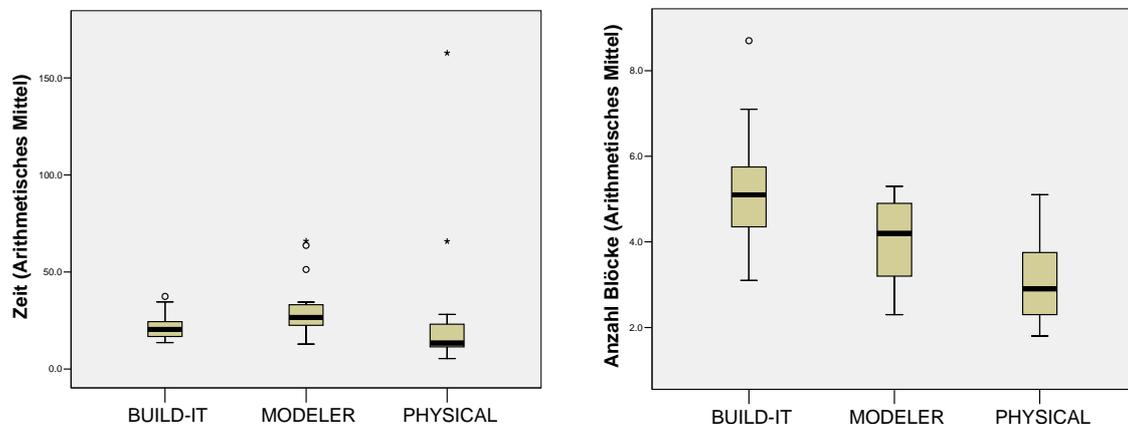


Abbildung 16: Boxplots Zeit und Anzahl Blöcke; fehlerhafte Daten, arithmetisches Mittel

Standardabweichung

Die Unterschiede in der Standardabweichung haben sich verbessert. Bei der Zeit bestehen aber immer noch beträchtliche Differenzen (s. Tabelle 3). Die Standardabweichung ist daher noch nicht in zufriedenstellendem Masse einheitlich.

Tool	Arithmetisches Mittel Zeit	Arithmetisches Mittel Anzahl Blöcke
BUILD-IT	6.71	1.34
Modeler	14.04	0.98
Physical	35.04	0.88

Tabelle 3: Standardabweichungen; fehlerhafte Daten, arithmetisches Mittel

4.1.1.3. Transformation – sqrt

In diesem Abschnitt wird nur eine nichtlineare Transformation, für die Quadratwurzel, beispielhaft durchgeführt. Die anderen nichtlinearen Transformationen, der natürliche Logarithmus und der dyadische Logarithmus, werden an dieser Stelle nicht dargestellt, da jeweils nur geringfügige Veränderungen gegenüber der Quadratwurzel resultieren.

Aus den einzelnen arithmetischen Mittelwerten für jede Person wird die Quadratwurzel gezogen.

Boxplot

Diese Transformation genügt um zu zeigen, dass immer noch Ausreisser die Symmetrie der Boxplots negativ beeinflussen und zwar vor allem im Boxplot zur Zeit (s. Abbildung 17).

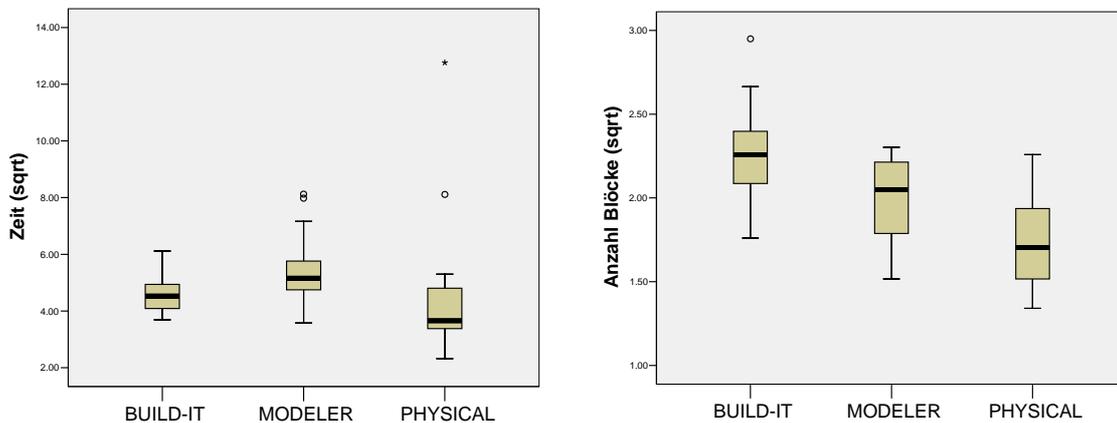


Abbildung 17: Boxplots Zeit und Anzahl Blöcke; fehlerhafte transformierte Daten

Die Daten werden auf Grund der Ausreisser einer genaueren Untersuchung unterzogen. Es sollte herausgefunden werden, ob es einzelne Personen gibt, die eine Häufung von Ausreissern aufweisen oder ob sie über alle Personen verteilt sind.

Es hat sich dann gezeigt, dass einzelne Versuchspersonen extrem lange (Maximum 864 sec!) für das Lösen der Aufgabe gebraucht haben. Diese nicht realistisch erscheinenden Werte sind der Anlass, alle Einträge für die Zeit und die Anzahl Blöcke in der Exceltabelle mit den originalen Werten in den Versuchsprotokollen zu vergleichen.

Bei diesem Vergleich wird entdeckt, dass einzelne Einträge zu einem früheren Zeitpunkt, im Vorfeld dieser Arbeit, falsch übertragen wurden und zwar mit massiven Differenzen, was natürlich die Verteilung der Daten stark verzerrt hat. Falsche Einträge werden bei der Zeit gefunden, nicht aber bei der Anzahl Blöcke. Die Tabelle 4 zeigt alle falschen Zeitwerte mit den dazugehörigen richtigen Werten. Die Fehler betreffen vor allem das Physical Tool.

Person	Versuch	Tool	Falsche Zeit (sec)	Richtige Zeit (sec)
12	T3	BUILD-IT	40	41
45	T2	Modeler	90	110
53	T9	Physical	576	40
56	T1	Physical	288	20
56	T3	Physical	360	25
56	T7	Physical	864	6
61	T4	Physical	144	10
63	T1	Physical	19	17

Tabelle 4: Falsche- und dazugehörige richtige Zeitwerte

4.1.2. Korrekte Daten

Der Boxplot und die Standardabweichung für die Anzahl Blöcke bleiben unverändert, da für diese keine falschen Werte gefunden wurden. Sie sind in der Folge nur der Vollständigkeit halber aufgeführt und werden nicht mehr explizit besprochen.

4.1.2.1. Untransformierte Daten

Boxplots

Mit den korrigierten Zeitwerten sind die Ausreisser und Extremfälle bei weitem nicht mehr so gross wie bei den untransformierten falschen Daten (s. Kapitel 4.1.1.1). Im Boxplot zur Zeit weist das Modeler Tool die gravierendsten Ausreisser und Extremfälle aus (s. Abbildung 18).

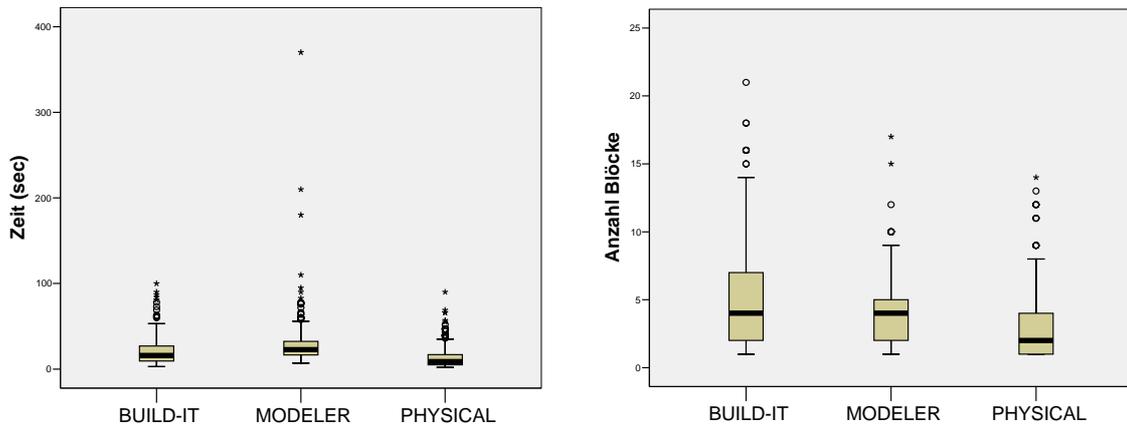


Abbildung 18: Boxplots Zeit und Anzahl Blöcke; korrekte untransformierte Daten

Standardabweichung

Die Standardabweichung zwischen den einzelnen Tools variiert für die Zeit noch sehr stark (s. Tabelle 5).

Tool	Zeit (sec)	Anzahl Blöcke
BUILD-IT	18.10	4.09
Modeler	34.17	2.54
Physical	13.98	2.66

Tabelle 5: Standardabweichungen; korrekte untransformierte Daten

4.1.2.2. Arithmetisches Mittel

Boxplots

Im Vergleich zu vorher hat sich vor allem die Zahl der Ausreisser bei der Zeit massiv verkleinert. Die Idealform in Bezug auf die Symmetrie ist jedoch noch nicht erreicht (s. Abbildung 19). Zur Verbesserung sind deshalb noch nichtlineare Transformationen nötig.

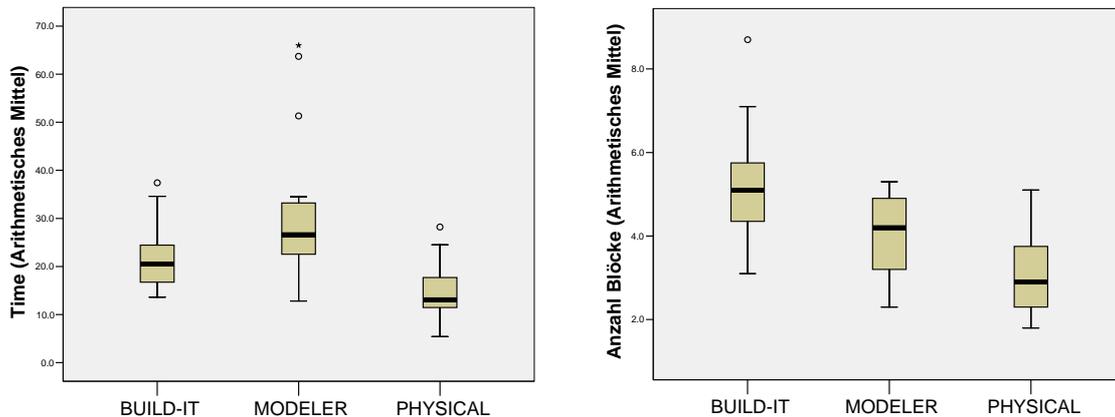


Abbildung 19: Boxplots Zeit und Anzahl Blöcke; korrekte Daten, arithmetisches Mittel

Standardabweichung

Die Unterschiede in der Standardabweichung haben sich verbessert. Bei der Zeit bestehen aber immer noch beträchtliche Differenzen (s. Tabelle 6). Die Standardabweichung ist daher noch nicht in zufriedenstellendem Masse einheitlich.

Tool	Arithmetisches Mittel Zeit	Arithmetisches Mittel Anzahl Blöcke
BUILD-IT	6.72	1.34
Modeler	13.98	0.98
Physical	5.91	0.88

Tabelle 6: Standardabweichungen; korrekte Daten, arithmetisches Mittel

4.1.2.3. Transformationen

Es wird darauf verzichtet in diesem Kapitel die Boxplots und Standardabweichungen der drei Transformationen für die Zeit und die Anzahl Blöcke darzustellen. Dies hätte sechs Boxplots zur Folge und wäre nicht mehr übersichtlich. Boxplots und Standardabweichung werden im nächsten Kapitel 4.2 noch dargestellt und genauer untersucht. Stattdessen sollen der Levene- und der Kolmogorov-Smirnov-Test zur Prüfung der Homogenität der Varianzen und der Annahme der Normalverteilung der transformierten Daten herangezogen werden.

In der Tabelle 7 sind die Resultate des Levene-Tests, zur Prüfung der Homogenität der Varianzen, und des Kolmogorov-Smirnov-Tests, zur Prüfung der Normalverteilung für jede Transformation zusammengefasst. Die Zahlenwerte geben den p-Wert (Irrtumswahrscheinlichkeit) der Tests wieder. Das

Signifikanzniveau ist $\alpha = 0.05$ und bedeutet, dass man sich bei dem gefundenen Ergebnis mit einer Wahrscheinlichkeit von 95% nicht geirrt hat.

Für den Levene-Test bedeutet ein p-Wert (Irrtumswahrscheinlichkeit) < 0.05 eine signifikante Abweichung von der Homogenität der Varianzen. Für den Kolmogorov-Smirnov-Test bedeutet ein p-Wert (Irrtumswahrscheinlichkeit) < 0.05 eine signifikante Abweichung von der Normalverteilung.

Die Signifikanzentscheidung lautet dann „ja“, wenn der Test nicht verworfen werden kann und somit keine signifikante Abweichung vorliegt. Ist die Abweichung signifikant lautet die Entscheidung „nein“. Ist für einen der beiden Tests die Entscheidung negativ, dann sind die Voraussetzungen insgesamt nicht erfüllt.

Dies ist nur beim dyadischen Logarithmus für die Variable Zeit der Fall. Diese Transformation wird aber der Vollständigkeit halber im nächsten Kapitel 4 trotzdem weiterhin aufgeführt.

		Zeit			Anzahl Blöcke		
		sqrt	ln	2log	sqrt	ln	2log
Kolmogorov-Smirnov-Test; p-Wert	BUILD-IT	0.43	0.63	0.63	0.98	0.63	0.63
	Modeler	0.35	0.61	0.66	0.64	0.49	0.49
	Physical	0.54	0.58	0.51	0.97	0.97	0.97
Kolmogorov-Smirnov-Test erfüllt?		ja	ja	ja	ja	ja	ja
Levene-Test; p-Wert		0.54	0.74	0.01	0.85	0.26	0.74
Levene-Test erfüllt?		ja	ja	nein	ja	ja	ja
Voraussetzungen insgesamt erfüllt?		ja	ja	nein	ja	ja	ja

Tabelle 7: p-Wert und Signifikanzentscheidung für Kolmogorov-Smirnov- und Levene-Test

4.2. Beurteilung der Transformationen

Es wird der Schritt 5 gemäss Kapitel 3.1 durchgeführt. Die drei Transformationen sollen anhand der drei Kriterien Boxplot, Standardabweichung und Histogramm miteinander verglichen und bewertet werden.

4.2.1. Boxplot

In Abbildung 20 sind die Boxplots der Zeit und der Anzahl Blöcke für die drei Transformationen sqrt, ln und 2log tabellarisch dargestellt, damit die einzelnen Diagramme direkt auf ihre Symmetrie hin verglichen werden können. Der Boxplot, welcher der Symmetrie am nächsten kommt, ist am ehesten für die ANOVA geeignet.

Aus den Boxplots für die Zeit wird die Box eines Tools innerhalb des Boxplots einer Transformation betrachtet und mit den Boxen desselben Tools in den Boxplots der anderen Transformation mit der gleichen Variable verglichen. Also z.B. BUILD-IT mit der sqrt-Transformation für die Variable Zeit, wird mit BUILD-IT für die ln- und 2log-Transformation bei gleichbleibender Variable betrachtet (mit roten Kästchen markiert).

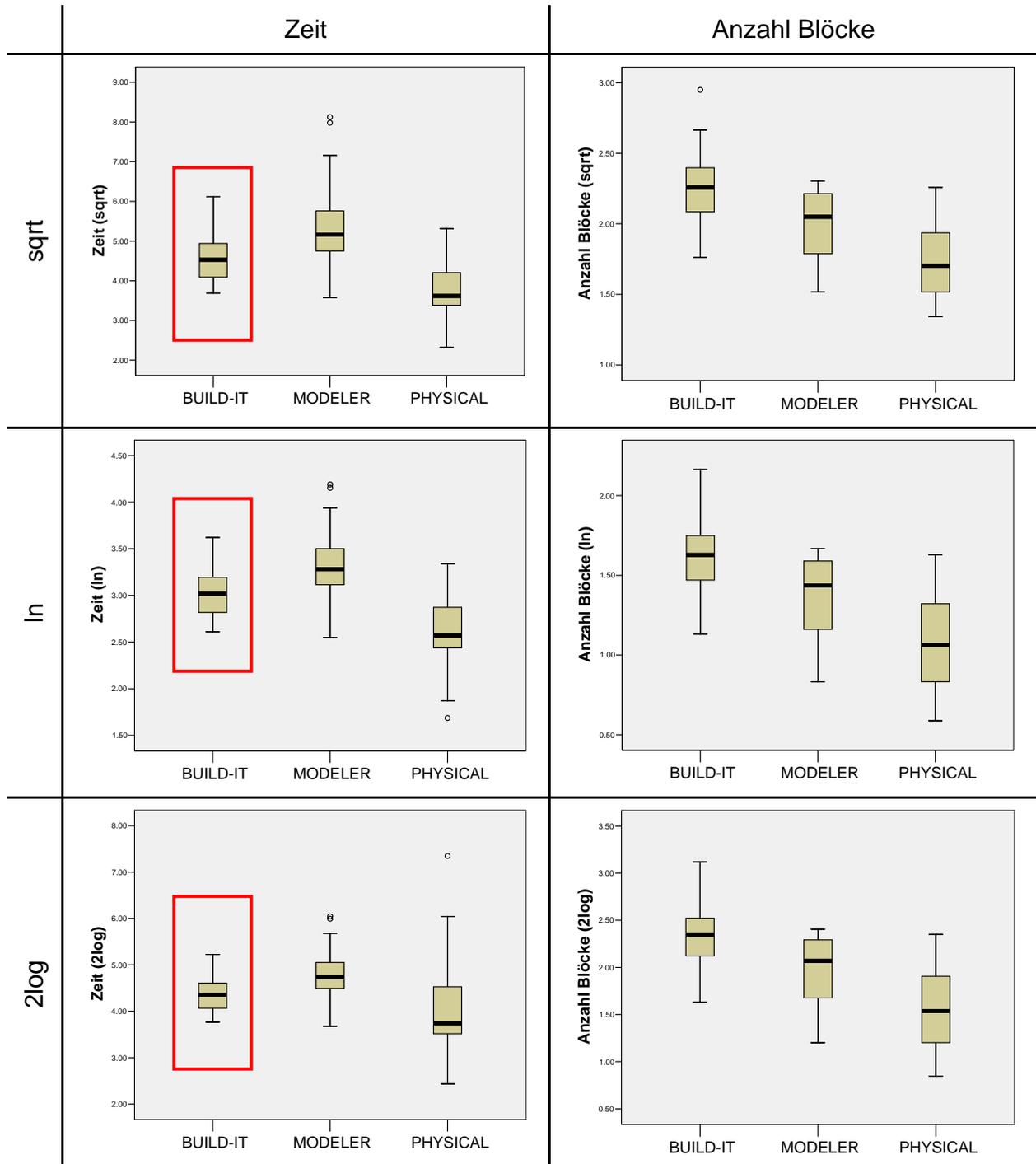


Abbildung 20: Boxplots Zeit und Anzahl Blöcke; korrekte transformierte Daten

Die drei Transformationen werden in die Ränge eins bis drei eingeteilt (s. Tabelle 8). 2log erhält den ersten Rang, ln und sqrt erhalten die folgenden Ränge. Dasselbe Vorgehen wird erst für die anderen Tools und dann auch noch für die Variable Anzahl Blöcke angewandt. Die Ränge pro Transformation werden addiert (Zeile „Totale Punktzahl“) und für jede Variable wiederum die Ränge eins bis drei vergeben (Zeile „Rang Boxplot“).

		Zeit			Anzahl Blöcke		
		sqrt	ln	2log	sqrt	ln	2log
Rang	BUILD-IT	3	2	1	1	2	3
	Modeler	3	2	1	1	2	3
	Physical	1	2	3	3	2	1
Totale Punktzahl		7	6	5	5	6	7
Rang Boxplot		3	2	1	1	2	3

Tabelle 8: Bewertung der Boxplots

4.2.2. Standardabweichung

Tabelle 9 und Tabelle 10 zeigen die Werte der Standardabweichungen der Variablen Zeit und Anzahl Blöcke für die drei Transformationen. Innerhalb einer Transformation wird die Homogenität der Werte überprüft.

Tool	sqrt	ln	2log
BUILD-IT	0.68	0.28	0.41
Modeler	1.15	0.40	0.58
Physical	0.78	0.43	1.14

Tabelle 9: Standardabweichungen Zeit; korrekte transformierte Daten

Tool	sqrt	ln	2log
BUILD-IT	0.29	0.25	0.36
Modeler	0.26	0.26	0.39
Physical	0.25	0.28	0.41

Tabelle 10: Standardabweichungen Anzahl Blöcke; korrekte transformierte Daten

Dazu wird das Verhältnis der kleinsten zur grössten Standardabweichung einer Transformation berechnet (s. Tabelle 11). Das Ergebnis ist der gleichnamigen Zeile der Tabelle zu entnehmen. Das Verhältnis mit dem grössten Endwert zeigt die Transformation mit den homogensten Standardabweichungen an. Für die Zeit weist ln die homogenste Standardabweichung auf und erhält den Rang eins. Sqrt und 2log erhalten die folgenden Ränge. Dasselbe Vorgehen wird für die Variable Anzahl Blöcke angewandt.

	Zeit			Anzahl Blöcke		
	sqrt	ln	2log	sqrt	ln	2log
Verhältnis kleinste/grösste Standardabweichung	0.68/ 1.15	0.28/ 0.43	0.41/ 1.14	0.25/ 0.29	0.25/ 0.28	0.36/ 0.41
Ergebnis	0.59	0.65	0.36	0.86	0.89	0.88
Rang Standardabweichung	2	1	3	3	1	2

Tabelle 11: Bewertung Standardabweichung

4.2.3. Histogramm

In Abbildung 21 und Abbildung 22 sind die Histogramme der Zeit und der Anzahl Blöcke dargestellt. Sie beinhalten die imaginäre Kurve der Normalverteilung sowie das eigentliche Histogramm der realen Daten. Je näher die Balken bei der Normalverteilungskurve liegen, desto eher erfüllen die zugrundeliegenden Daten die Annahme der Normalverteilung. Da die Zahl der Proben mit 20 ziemlich gering ist, weisen die Histogramme markante Lücken auf.

Wiederum werden die drei Histogramme der transformierten Daten des selben Tools miteinander verglichen. Dabei ist das Kriterium, dass die Balken in ihrer Gesamtheit möglichst nahe bei der Normalverteilungskurve liegen. Das Histogramm, das dieses Kriterium am besten erfüllt, erhält den ersten Rang.

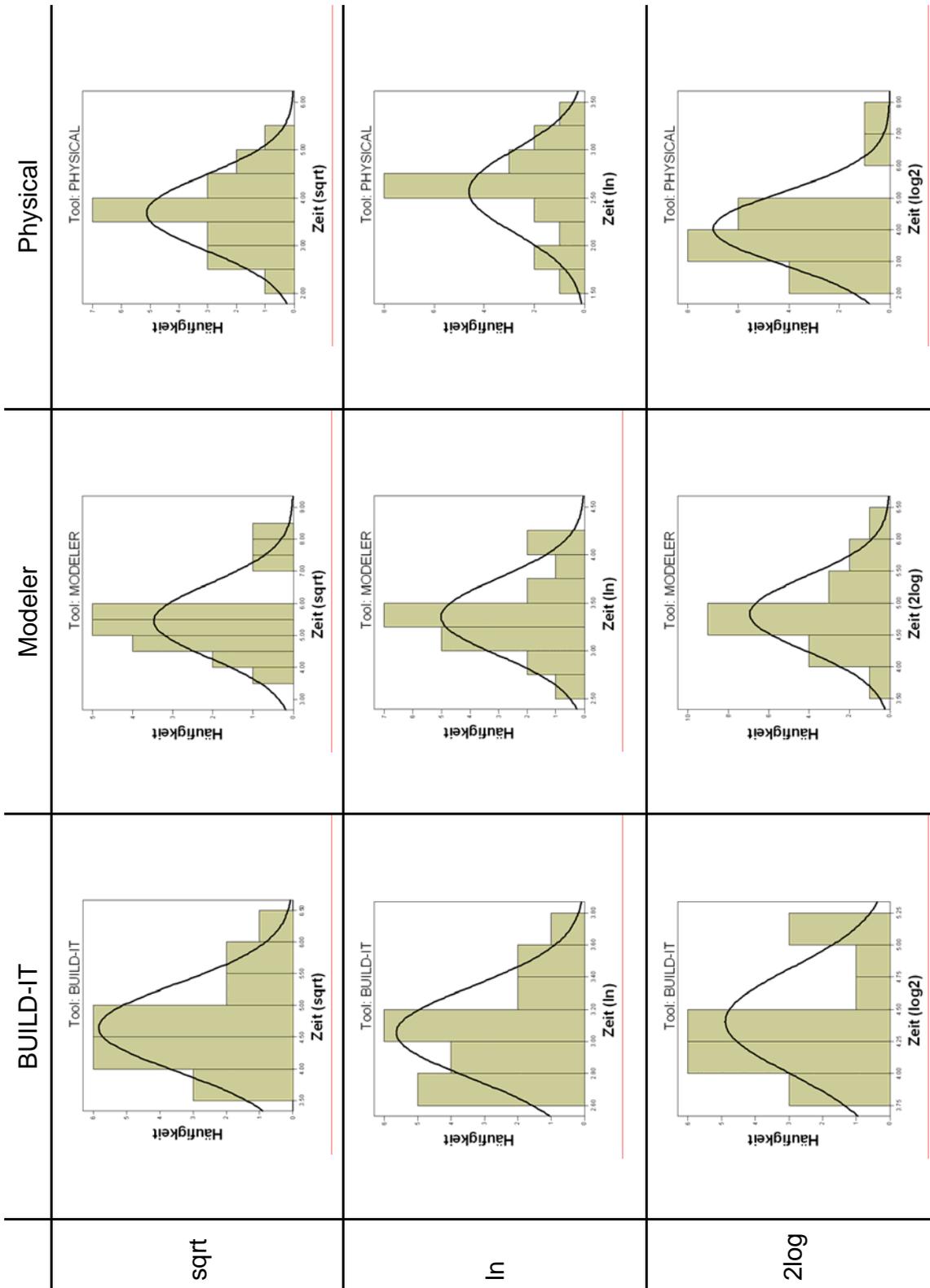


Abbildung 21: Histogramme Zeit; korrekte transformierte Daten

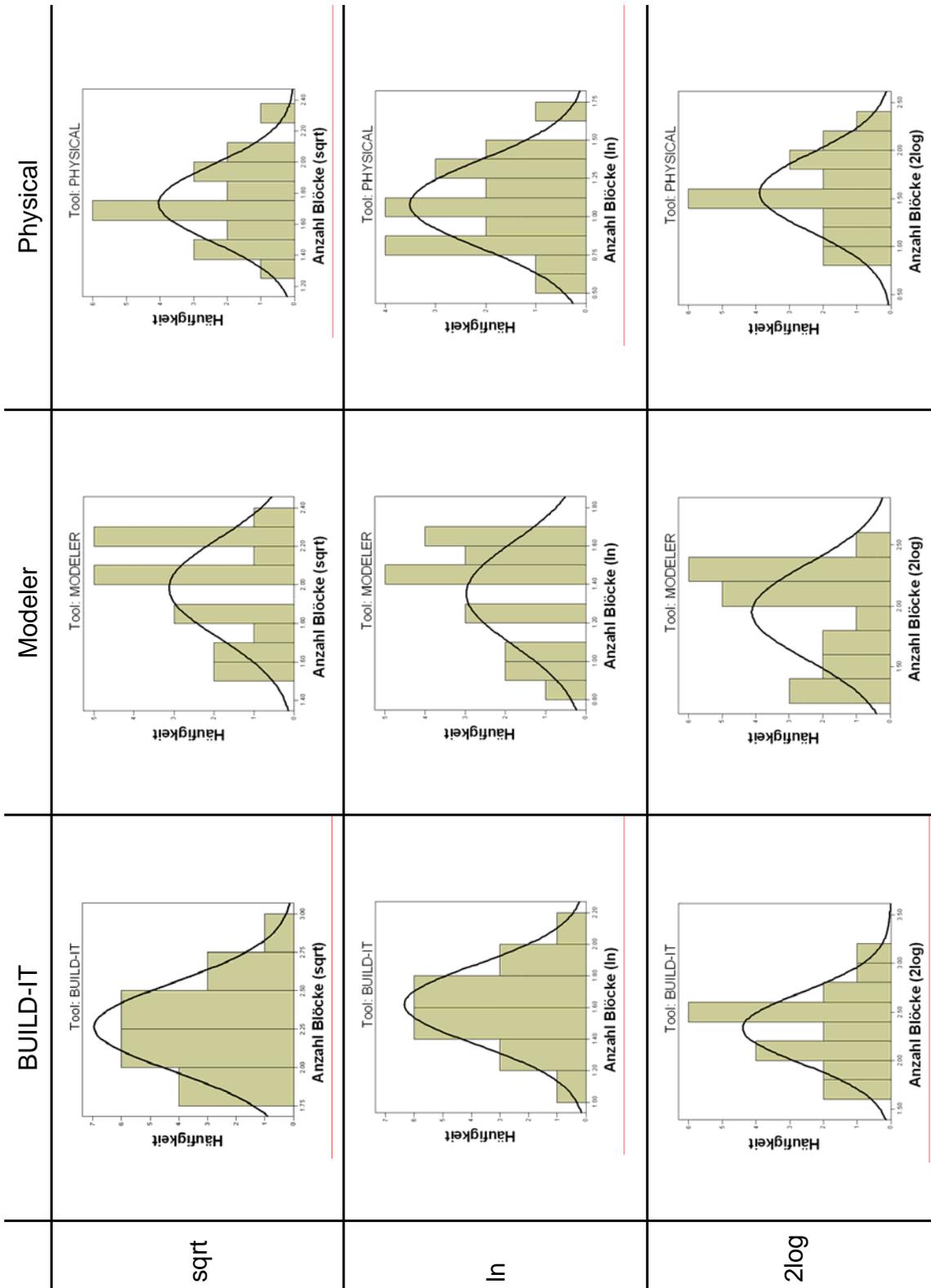


Abbildung 22: Histogramme Anzahl Blöcke; korrekte transformierte Daten

Der Vergleich der Ränge ist in Tabelle 12 zusammengefasst. Die Ränge der einzelnen Tools einer Transformation werden danach zu einer totalen Punktzahl addiert. Die Transformation mit der geringsten Punktzahl erhält Rang eins und die anderen entsprechend die folgenden Ränge. Bei der Variable Zeit haben zufällig zwei Transformationen die geringste Punktzahl von 5 und erhalten deshalb beide den ersten Rang.

		Zeit			Anzahl Blöcke		
		sqrt	ln	2log	sqrt	ln	2log
Rang	BUILD-IT	1	1	3	1	2	3
	Modeler	3	2	1	2	2	1
	Physical	1	2	3	1	3	2
Totale Punktzahl		5	5	7	4	7	6
Rang Histogramm		1	1	3	1	3	2

Tabelle 12: Bewertung der Histogramme

4.3. Entscheidung über Transformationen

Nach Kapitel 3.1 wird der Schritt 6 durchgeführt.

Für die Gesamtbewertung werden die bisherigen Ränge der Kriterien Boxplot, Standardabweichung und Histogramm zu einem neuen Punkttotal addiert (s. Tabelle 13). Aus diesem Punktstand resultiert der Gesamtrang für jede Transformation.

Für die Variable Zeit ist die Transformation mit dem natürlichen Logarithmus am besten geeignet. Für die Variable Anzahl Blöcke ist die Transformation mit der Quadratwurzel am besten geeignet.

		Zeit			Anzahl Blöcke		
		sqrt	ln	2log	sqrt	ln	2log
Rang	Boxplot	3	2	1	1	2	3
	Standardabweichung	2	1	3	3	1	2
	Histogramm	1	1	3	1	3	2
Totale Punktzahl		6	4	7	5	6	7
Rang Gesamt		2	1	3	1	2	3

Tabelle 13: Gesamtbewertung

Die abhängigen Variablen Zeit und Anzahl Blöcke erfüllen somit durch diese Transformationen die Voraussetzungen für die ANOVA: Homogenität der Standardabweichungen und Normalverteilung der Daten. Im folgenden Kapitel werden die transformierten Daten für die ANOVA benutzt.

5. Resultate der Messreihen

Nachdem im vorhergehenden Kapitel 4 die Voraussetzungen für die ANOVA geprüft worden sind, sollen nun sinnvolle Hypothesen für die Analyse festgelegt werden. Diese werden anschliessend mit der ANOVA auf ihre Gültigkeit hin überprüft. In der Diskussion werden die Resultate zusammengefasst und besprochen.

5.1. Kognitive Unterstützung

Gegenstände wie die Tools in dieser Untersuchung können das Denken unterstützen. Kognitive Unterstützung kann in diesem Zusammenhang dann definiert werden, als die von einem Gegenstand, also hier von einem Tool, dem Benutzer gebotene Unterstützung für das Denken und das Lösen von Problemen.

Da das Physical Tool dem Benutzer erlaubt, die Elemente mit den eigenen Fingern zu berühren, von verschiedenen Seiten zu betrachten und intuitiv zu lösen, wie man das aus dem Alltag gewohnt ist, wird erwartet, dass dieses Tool die höchste kognitive Unterstützung gibt. Von BUILD-IT wird die zweitbeste kognitive Unterstützung erwartet, da das Problem durch das virtuelle 3D-Modell nahe der realen Situation ist und der Benutzer den Brick, der den Laserpointer repräsentiert, auch mit den eigenen Fingern greifen kann. Vom Modeler Tool wird am wenigsten kognitive Unterstützung erwartet, da das Navigieren und Manipulieren menügesteuert ist und es daher mehr abstraktes Denken vom Benutzer verlangt.

Die Tabelle 14 fasst diese Einteilung zusammen.

Tool	Erwartete kognitive Unterstützung
BUILD-IT	mittel
Modeler	tief-mittel
Physical	hoch

Tabelle 14: Erwartete kognitive Unterstützung

5.2. Hypothesen

Aufgrund der erwarteten kognitiven Unterstützung (s. Kapitel 5.1) werden Vermutungen, oder Hypothesen, darüber aufgestellt, welche kognitive Unterstützung die Tools effektiv geben. Die Daten sollen dann daraufhin überprüft werden, ob diese Hypothesen zutreffen oder nicht.

Hypothesen:

- **H1:** Modeler gibt weniger kognitive Unterstützung als Physical
- **H2:** Modeler gibt weniger kognitive Unterstützung als BUILD-IT
- **H3:** BUILD-IT gibt weniger kognitive Unterstützung als Physical

Die Verschmelzung von erwarteter kognitiver Unterstützung und den Hypothesen führt zu folgender Abbildung 23.

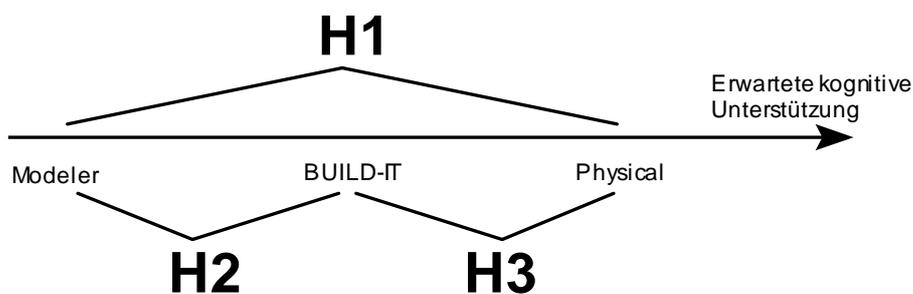


Abbildung 23: Erwartete kognitive Unterstützung (in Pfeilrichtung zunehmend) für jedes Tool und die Hypothesen H1-H3 zwischen den Tools.

5.3. Kriterien zur Messung der kognitiven Unterstützung

Die Hypothesen H1-H3 können nicht unmittelbar mit einem Test geprüft werden, da die kognitive Unterstützung nicht direkt mit einer Variablen messbar ist. Die kognitive Unterstützung wird aufgeteilt und durch einzelne, messbare, abhängige Variablen repräsentiert. Bei den Variablen handelt es sich um die Zeit, die Anzahl Blöcke, die Benutzerzufriedenheit und den Versuchserfolg. Aus diesen Variablen werden die Kriterien K1 bis K4 gebildet, mit denen die kognitive Unterstützung gemessen wird:

- K1:** tiefere Versuchszeit
- K2:** höhere Anzahl getesteter Blöcke pro Versuch (das Ausprobieren von mehr Blöcken reduziert die kognitive Belastung)

- K3:** höhere Benutzerzufriedenheit (wahrgenommene Klarheit der Aufgabenformulierung, Aufgabenschwierigkeit und Tooltauglichkeit)
- K4:** höhere Rate an erfolgreichen Versuchen (am Ende eines Versuches wurde die richtige Lösung gefunden)

Ein Kriterium wird immer bezogen auf eine Hypothese angewendet. Beispielsweise ist das Kriterium K1 unter Hypothese H1 erfüllt wenn die Zeit von Physical tiefer ist als von Modeler.

5.4. Darstellung der Resultate

5.4.1. Abhängige Variablen der ANOVA

Zeit & Anzahl Blöcke

Die arithmetischen Mittel von Zeit, Anzahl Blöcke und Versuchserfolg sind in Tabelle 15 dargestellt. Es wurden die ursprünglichen Daten verwendet, d.h. die Zeit hat die Einheit Sekunden und die Anzahl Blöcke besteht aus der effektiven Anzahl Blöcke

In Abbildung 24 sind die Boxplots von Zeit (ln) und Anzahl Blöcke (sqrt) zu sehen. Es wurden die transformierten Daten verwendet. Die Abkürzung in Klammern zeigt die verwendete Transformation an.

Boxplots und arithmetische Mittel geben einen ersten Eindruck von Lage und Streuung der Daten. Das Bild der Lage der Tools in den Boxplots ist für Zeit und Anzahl Blöcke erstaunlicherweise nicht übereinstimmend. Gemäss Kriterien ist die Abfolge für die Zeit (von tief nach hoch) Physical, BUILD-IT, Modeler; und für die Anzahl Blöcke (von hoch nach tief) BUILD-IT, Modeler, Physical.

Tool	Arithmetisches Mittel	
	Zeit (sec)	Anzahl Blöcke
BUILD-IT	22.08	5.2
Modeler	31.02	4.0
Physical	14.18	3.1

Tabelle 15: Arithmetisches Mittel von Zeit und Anzahl Blöcke

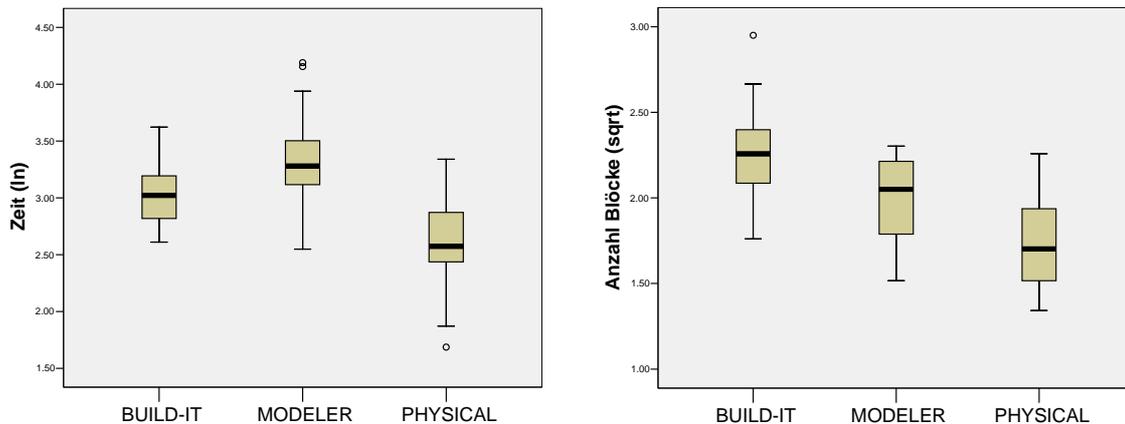


Abbildung 24: Boxplot von Zeit (ln) und Anzahl Blöcke (sqrt)

Benutzerzufriedenheit

Mit einem Fragebogen (s. Anhang A.5.2.1, Tabelle 27) wurden unter anderem drei Fragen zur Benutzerzufriedenheit gestellt. Es handelt sich um die folgenden Fragen:

- Frage 2.3:** War die Aufgabenstellung verständlich formuliert? → *Klarheit*
- Frage 2.1:** Wie empfanden Sie den Schwierigkeitsgrad der Aufgabenstellung? → *Schwierigkeit*
- Frage 3.1:** Wie schätzen Sie das Werkzeug zur Lösung dieser Aufgabe ein? → *Tauglichkeit*

Die Fragen können mit den Begriffen *Klarheit*, *Schwierigkeit* und *Tauglichkeit* charakterisiert werden. Die Antwortskala stellt den Befragten eine Skala [-2,-1,0,1,2] von tiefer bis grosser Zufriedenheit zur Auswahl.

Die Resultate der Antworten sind mit den neuen Begriffen in Tabelle 16 zusammengefasst. Die Summe der gemittelten Werte der drei Fragen ergibt die Benutzerzufriedenheit, die im Folgenden als abhängige Variable weiterverwendet wird.

Tool	Arithmetisches Mittel			Benutzerzufriedenheit
	<i>Klarheit</i>	<i>Schwierigkeit</i>	<i>Tauglichkeit</i>	
BUILD-IT	1.7	0.5	0.6	0.9
Modeler	0.6	0.6	0.7	0.6
Physical	1.8	0.6	0.6	1.0

Tabelle 16: Klarheit, Schwierigkeit und Tauglichkeit ergeben die Benutzerzufriedenheit

Versuchserfolg

Der Versuchserfolg drückt den prozentualen Anteil der Versuche mit richtiger Lösung pro Tool aus. In Tabelle 17 sind die Resultate zusammengestellt.

Tool	Versuchserfolg (in Prozent)
BUILD-IT	96,5%
Modeler	79%
Physical	100%

Tabelle 17: Versuchserfolg

Die Variable Versuchserfolg erfüllt die Voraussetzungen für die ANOVA nicht. Sowohl Homogenität der Varianz als auch Normalverteilung sind nicht gegeben. Auch für einen nichtparametrischen Test wie den Kruskal-Wallis-Test sind die Voraussetzungen nicht erfüllt. Die Verteilungen der drei Tools müssten dazu mindestens ähnliche Form aufweisen. Dies ist jedoch nicht gegeben, da zumindest das Physical Tool mit einem Wert von 100% nur einen einzigen Wert annimmt und man daher nicht von einer Verteilung sprechen kann.

Da kein Test durchgeführt werden kann, werden allein die Zahlenwerte miteinander verglichen. Die Benutzer des Modeler Tools zeigen einen klar tieferen Anteil an richtig ausgeführten Versuchen als die Benutzer der anderen zwei Tools. Aus diesem Grund ist das Kriterium K.4 für H1 und H2 erfüllt. Obwohl auch BUILD-IT keinen 100% Anteil aufweist, kann zwischen BUILD-IT und Physical kein klarer Unterschied im Versuchserfolg ausgemacht werden. K.4 ist für H3 nicht erfüllt. Diese Resultate werden in die Tabelle 20 (s. Kapitel 5.4.4) übertragen.

5.4.2. ANOVA

Zur statistischen Analyse der verbleibenden drei Variablen wird gemäss Kapitel 3.1 (Schritt 7) eine ANOVA durchgeführt. Die unabhängige Variable ist Tool. Die abhängigen Variablen sind Zeit (ln), Anzahl Blöcke (sqrt) und Benutzerzufriedenheit (s. Tabelle 18). Das Signifikanzniveau ist $\alpha = 0.05$.

Die ANOVA ergibt für alle abhängigen Variablen signifikante Unterschiede zwischen den Tools, ausgewiesen durch einen p-Wert (Irrtumswahrscheinlichkeit) < 0.05 . Man weiss somit erst, dass signifikante Unterschiede vorhanden sind. Jedoch ist noch nicht bekannt, zwischen welchen Tools diese bestehen. Aus diesem Grund muss noch ein Post-Hoc-Test durchgeführt werden, um herauszufinden, zwischen welchen Tools der Unterschied effektiv signifikant ist und der Wert welches Tools grösser ist.

Abhängige Variable	p-Wert	signifikant?
Zeit (ln)	< 0.001	ja
Anzahl Blöcke (sqrt)	< 0.001	ja
Benutzerzufriedenheit	0.024	ja

Tabelle 18: Resultate der einfaktoriellen ANOVA

5.4.3. Post-Hoc-Test

Für den Post-Hoc-Test wird gemäss Kapitel 3.1 (Schritt 8) der Tukey HSD (honestly significant difference) Test verwendet. Die unabhängige Variable ist Tool. Die abhängigen Variablen sind Zeit (ln), Anzahl Blöcke (sqrt) und Benutzerzufriedenheit (s. Tabelle 19). Das Signifikanzniveau ist $\alpha = 0.05$.

Signifikante Unterschiede treten daher für einen p-Wert (Irrtumswahrscheinlichkeit) < 0.05 auf. Es bestehen nicht für alle abhängigen Variablen signifikante Unterschiede zwischen den Tools (Spalte „Signifikant?“). Alle nichtsignifikanten Resultate werden ausgeschieden. Von den signifikanten Resultaten wird die paarweise Differenz zwischen den Tools, unter dem Gesichtspunkt der Hypothesen und Kriterien, genauer betrachtet. Beispielsweise gilt für die abhängige Variable Zeit (ln) das Kriterium K1. Jede der drei Toolpaar-Kombinationen, welche eine der Hypothesen H1-H3 repräsentiert, wird mit K1 verglichen, ob dieses erfüllt ist. Gleich wird für K2 und K3 vorgegangen. Zu beachten ist, dass ein Resultat zwar signifikant sein kann, aber in die falsche Richtung, so dass das Kriterium nicht erfüllt ist. So z.B. bei K2 unter H1. Bei signifikanten Resultaten ist deshalb Vorsicht vor schnellen Rückschlüssen auf ein erfülltes Kriterium geboten.

Die Resultate werden in Tabelle 20 (s. Kapitel 5.4.4) übertragen.

Abhängige Variable	Toolpaar I&J (Hypothese)	Paarweise Differenz (I-J)	p-Wert	Signifikant?
Zeit (ln)	Modeler & Physical (H1)	0.79	< 0.001	ja
"	BUILD-IT & Modeler (H2)	-0.30	0.038	ja
"	BUILD-IT & Physical (H3)	0.49	< 0.001	ja
Anzahl Blöcke (sqrt)	Modeler & Physical (H1)	0.25	0.012	ja
"	BUILD-IT & Modeler (H2)	0.28	0.004	ja
"	BUILD-IT & Physical (H3)	0.53	< 0.001	ja
Benutzer-zufriedenheit	Modeler & Physical (H1)	-0.38	0.028	ja
"	BUILD-IT & Modeler (H2)	0.32	0.083	nein
"	BUILD-IT & Physical (H3)	-0.07	0.890	nein

Tabelle 19: Resultate von Tukey HSD, Post-Hoc-Tests

5.4.4. Zusammenfassung der Resultate

Die bisherigen Resultate der Kriterien K1 bis K4 zu den Hypothesen H1-H3 können nun zusammengefasst werden (s. Tabelle 20). Für jedes Kriterium wird die Entscheidung getroffen, ob es erfüllt ist oder nicht und entsprechend wird dies ausgedrückt mit ja oder nein. Drei von vier Kriterien sind für H1 und H2 erfüllt. Für H3 ist nur ein Kriterium erfüllt.

Hypothese	K1	K2	K3	K4	Anzahl erfüllter Kriterien
H1	ja	nein	ja	ja	3 (von 4)
H2	ja	ja	nein	ja	3 (von 4)
H3	ja	nein	nein	nein	1 (von 4)

Tabelle 20: Zusammenfassung der Resultate

Als Voraussetzung für die Bestätigung einer Hypothese wird die Erfüllung von mehr als der Hälfte der Kriterien genommen, d.h. mehr als zwei. Daher können H1 und H2 übers Ganze gesehen bestätigt werden. Nicht genügend bestätigt hat sich H3.

Die Ergebnisse für die Hypothesen sind wie folgt:

- H1: Es wird weniger Zeit gebraucht, die Benutzerzufriedenheit ist höher und der Versuchserfolg ist grösser mit Physical als mit Modeler.
- H2: Es wird weniger Zeit gebraucht, mehr Blöcke werden getestet und der Versuchserfolg ist grösser mit BUILD-IT als mit Modeler.
- H3: Es wird weniger Zeit gebraucht mit Physical als mit BUILD-IT.

5.4.5. Zusätzliche Auswertungen

Zusätzlich zur Auswertung mit ANOVA werden Lerneffekte und sekundäre Daten nur anhand qualitativer Kriterien ausgewertet. Aus Gründen, die noch genannt werden, kann keine statistische Methode zur Auswertung angewandt werden.

5.4.5.1. Qualitative Auswertung der Lerneffekte

Zusätzlich zu K1-K4 werden zwei weitere Kriterien definiert:

- Lerneffekt in der Versuchszeit, d.h. tendenziell weniger Versuchszeit
- Lerneffekt in der Anzahl getestete Blöcke, d.h. tendenziell weniger Blöcke getestet

Die Lerneffekte können ebenfalls als eine Art der Messung der kognitiven Unterstützung gedeutet werden. Sie stehen in Beziehung mit der Leichtigkeit der Erlernung und des Gebrauchs eines Tools.

Jedoch kann mit den Kriterien kein Vergleich zwischen zwei Tools angestellt werden, da der Unterschied im Lerneffekt zwischen zwei Tools nicht quantifiziert werden kann. Dies ist der Fall, weil wegen der Verletzung der Voraussetzung der Unabhängigkeit der einzelnen Versuche kein statistischer Test durchgeführt werden kann. Die Kriterien werden deshalb nur qualitativ anhand von Boxplots ausgewertet.

Die Unabhängigkeit ist aus folgendem Grund nicht erfüllt: Um Lerneffekte eines Tools zu untersuchen, werden für beide Variablen die Werte von Versuch eins bis zehn miteinander verglichen. Da diese zehn Versuche aber von derselben Person ausgeführt wurden, sind diese Werte voneinander abhängig.

Um viele Ausreisser und Extremfälle in den Diagrammen zu verhindern, sind die ursprünglichen Daten für die Zeit ln- und für die Anzahl Blöcke sqrt-transformiert worden.

Abbildung 25 zeigt die Boxplots von Zeit und Anzahl Blöcke der drei Tools von Versuch 1 bis 10. Allfällige Lerneffekte werden direkt aus den Graphiken herausgelesen. Dazu werden die Trends von Median, Box oder Whiskers verglichen werden. Ein wichtiges Kriterium ist auch, ob der letzte Boxplot tiefer liegt als der erste. Dazwischen können sich schon, in beschränktem Masse, Schwankungen ergeben.

Grundsätzlich sind bei der Zeit die Daten homogener. Sie variieren nicht so stark von Versuch zu Versuch. Bei der Anzahl Blöcke hingegen sind die Daten inhomogener zwischen den Versuchen und breiter gestreut innerhalb eines Versuches.

In Tabelle 21 wird zusammengefasst, ob es Lerneffekte gibt:

- Für BUILD-IT sind Lerneffekte in der Zeit zu erkennen, aber nicht in der Anzahl Blöcke.
- Für Modeler sind Lerneffekte sowohl in der Zeit als auch in der Anzahl Blöcke zu erkennen.
- Für Physical sind keine Lerneffekte in der Zeit zu erkennen, jedoch kommt es in der Anzahl Blöcke in den letzten Versuchen zu Ermüdungserscheinungen.

Tools	Zeit	Anzahl Blöcke
BUILD-IT	ja	nein
Modeler	ja	ja
Physical	nein	nein/Ermüdung

Tabelle 21: Lerneffekte?

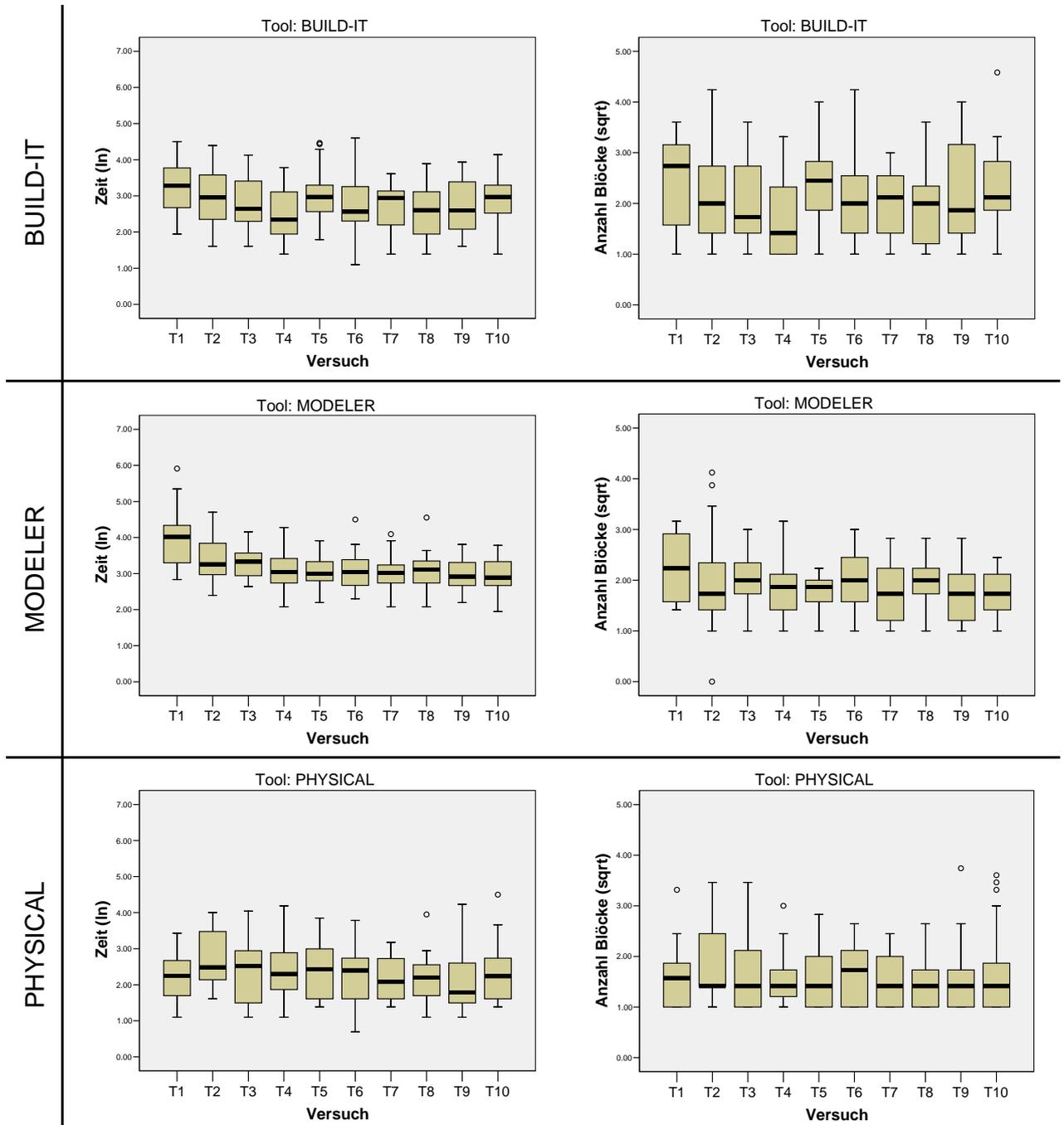


Abbildung 25: Lerneffekte; Boxplots für Zeit und Anzahl Blöcke

5.4.5.2. Qualitative Kontrolle sekundärer Daten

Von den Teilnehmern ist ein Fragebogen ausgefüllt worden. Dieser hat zum Ziel, sogenannte Kontrollvariablen, in diesem Fall Informationen zu den Versuchspersonen wie Alter, Geschlecht, Art und Häufigkeit des Computergebrauchs, Vorkenntnisse mit Tools und Aufgabe, zu erfassen. Damit wird der eventuelle Einfluss der Kontrollvariablen auf die abhängigen Variablen untersucht.

Der komplette Fragebogen (s. Tabelle 27) mit allen Antworten (s. Tabelle 26) kann im Anhang A.5.2.1 eingesehen werden.

Tabelle 22 fasst die Antworten des Fragebogens zusammen. Die Fragen 2.1, 2.3 und 3.1 sind hier nicht erwähnt, da sie schon im Kapitel 5.4.1 zur Benutzerzufriedenheit behandelt werden.

Das Antwortbild birgt keine grossen Überraschungen. Alle Personen benutzen den Computer zwischen einmal pro Woche und einmal pro Tag. Einzig bei der Benutzung verteilen sich die Antworten über alle fünf Möglichkeiten. Vorkenntnisse mit den Tools und der Aufgabe hat praktisch niemand, abgesehen von ein paar Ausnahmen.

Die Antworten sind im Allgemeinen ziemlich einheitlich. Deshalb sind auch keine grösseren Auswirkungen auf die abhängigen Variablen zu erwarten

Frage	Antworten
1.1: Wie oft benutzen Sie einen Rechner (PC, Workstation)?	[0] Minimum 1 Stunde Wöchentlich [1] Minimum 1 Stunde täglich
1.2: Wofür benutzen Sie einen Rechner? (Mehrfachnennungen möglich)	[-2] Freizeit (Spiele usw.) [-1] Kommunikation (E-Mail, WWW...) [0] Planung und Konstruktion [1] Code implementieren, Textverarbeitung, Simulation [2] Andere Zwecke
1.3: Wie gut sind Ihre Vorkenntnisse in BUILD-IT?	Hauptsächliche Antwort: [-2] Kenne BUILD-IT nicht
2.2: Haben Sie früher ähnliche Aufgaben gelöst?	Hauptsächliche Antwort: [-2] noch nie
3.2: Wie oft haben Sie dieses Werkzeug eingesetzt?	Hauptsächliche Antwort: [-2] noch nie

Tabelle 22: Antwortbild des Fragebogens

Die weiblichen Versuchspersonen sind mit einer Anzahl von fünf bei BUILD-IT im Vergleich zu den anderen zwei Tools klar übervertreten. Dort gibt es nämlich nur noch eine weitere weibliche Person, nämlich beim Physical.

Das Durchschnittsalter der Teilnehmer beträgt etwa 23 Jahre. Das Alter ist unter den Tools ausgewogen verteilt, so dass keine Beeinflussung der abhängigen Variablen erwartet wird.

5.5. Diskussion der Resultate

Es hat sich gezeigt, dass zwei Hypothesen angenommen werden können und eine Hypothese verworfen werden muss:

Hypothese H1 hat gezeigt, dass das Physical Tool mehr kognitive Unterstützung gibt als das Modeler Tool.

Einzig bezüglich der Anzahl der getesteten Blöcke kann das Physical- nicht mit dem Modeler Tool mithalten.

Die Hypothese H2 hat gezeigt, dass BUILD-IT mehr kognitive Unterstützung gibt als das Modeler Tool.

Einzig bezüglich der Benutzerzufriedenheit kann BUILD-IT nicht mit dem Modeler Tool mithalten.

Die Hypothese H3 hat gezeigt, dass das Physical Tool nicht mehr kognitive Unterstützung gibt als BUILD-IT.

Einzig bezüglich der Versuchszeit hat das Physical Tool BUILD-IT übertroffen.

BUILD-IT

BUILD-IT kommt der realen Umgebung des Physical Tools bezüglich Benutzerzufriedenheit und Versuchserfolg ziemlich nahe (keine signifikanten Unterschiede) und übertrifft es gar (signifikant) in der Anzahl getesteter Blöcke. Die Versuchszeit lässt zu wünschen übrig (signifikant schlechter).

Man kann folgern, dass die Benutzer nicht ganz so schnell sind im Arbeiten mit BUILD-IT, probieren aber mehr aus, was das Denken entlastet. Man kann vermuten, dass das vermehrte Ausprobieren mehr Zeit als nötig verbraucht. Es kann nur darüber spekuliert werden, dass der Benutzer vielleicht schneller gewesen wäre, hätte er vermehrt überlegt, wo die Lösung liegen könnte und nicht durch Ausprobieren dieses Defizit kompensiert. Das Ausprobieren hätte sich vielleicht aber ausbezahlt, wenn die Blöcke schneller und präziser hätten verschoben werden können. BUILD-IT könnte deshalb noch hinsichtlich Geschwindigkeit und Präzision, mit der man Objekte verschieben kann, verbessert werden.

Die hohe Quote an erfolgreichen Versuchen deutet darauf hin, dass die Benutzer sehen, ob sie die richtige Lösung gefunden haben und das TUI somit übersichtlich gestaltet ist. Die Benutzer sind zufrieden mit dem Tool.

Modeler

Das Modeler Tool ist hinter BUILD-IT (signifikantes) Mittelmaß in der Anzahl der getesteten Blöcke, am schlechtesten in der Versuchszeit und in der Benutzerzufriedenheit und abgeschlagen im Versuchserfolg (alles signifikant). Das sehr schlechte Abschneiden im Versuchserfolg könnte mit der Übersichtlichkeit des Tools in Beziehung stehen. Es könnte so interpretiert werden, dass die Versuchspersonen Mühe hatten zu erkennen, ob der Laserstrahl das Ziel trifft oder nicht.

Es wurde von Anfang an erwartet, dass das Modeler Tool die schlechtesten Ergebnisse und somit die tiefste kognitive Unterstützung hervorbringen wird. Diese Erwartungshaltung wird durch die Ergebnisse bestätigt.

Physical

Das Physical Tool ist am besten (signifikant) in der Versuchszeit, in der Benutzerzufriedenheit und im Versuchserfolg (beide nicht signifikant). Am schlechtesten (signifikant) schneidet es in der Anzahl getesteter Blöcke ab. Das schlechte Abschneiden in der Anzahl getesteter Blöcke erstaunt sehr. Die Benutzer haben weniger ausprobiert und die Lösung vermehrt durch analytisches Denken errahnt.

Es wurde erwartet, dass das Physical Tool die höchste kognitive Unterstützung zeigt. Diese Erwartung wurde nicht erfüllt. Es war nur wenig besser als BUILD-IT, wobei teilweise nur geringfügige Unterschiede zwischen den beiden bestanden.

Daraus ergibt sich die logische Schlussfolgerung, dass die Modellierung von virtuellen 3D Modellen für BUILD-IT schneller und billiger ist, als die Konstruktion von physischen 3D Modellen wie den in dieser Arbeit präsentierten physischen Blöcken, da, wenn überhaupt, dann nur ein geringer Vorteil daraus resultiert.

Benutzerzufriedenheit

Gewisse Ergebnisse aus der Befragung zur Benutzerzufriedenheit sind ein wenig erstaunlich und deshalb zu kommentieren.

Es entspricht der Erwartung, dass die Klarheit der Aufgabenformulierung für das Physical Tool am besten eingeschätzt wurde. Es erstaunt, aber dass BUILD-IT fast denselben Wert erreicht. Bei der Schwierigkeit der Aufgabenstellung und der Tooltauglichkeit ergeben sich keine wesentlichen Unterschiede. Die Werte unterscheiden sich jeweils um höchstens 0.1.

Bedenkt man, dass die Versuchsteilnehmer ihr Urteil für ein Tool völlig subjektiv auf einer Antwortskala mit fünf Abstufungen gefällt haben, ohne einen Vergleich mit den anderen Tools zu haben, dann werden diese Resultate ziemlich stark relativiert. Der Unterschied zwischen Modeler und den anderen beiden in der

Klarheit der Aufgabenstellung ist sicher deutlich und beeinflusst auch die Benutzerzufriedenheit insgesamt. Alle anderen Unterschiede sind aber sehr undeutlich.

Lerneffekte

Die beiden Kriterien für Lerneffekte stehen in Beziehung mit der Leichtigkeit der Erlernung und des Gebrauchs eines Tools. Sie stellen ebenfalls eine Art der Messung der kognitiven Unterstützung dar.

Keine Lerneffekte

Keine Lerneffekte bedeuten entweder, dass ein Tool von Anfang an intuitiv verständlich ist, also schon perfekt benutzt wird und somit keine Verbesserung in der Performance mehr möglich ist. Oder es zeigt, dass das Tool zu komplex ist und keine oder nur geringfügige Verbesserungen zu erzielen sind.

Lerneffekte vorhanden

Sind Lerneffekte vorhanden, dann zeigt dies, dass die perfekte Benutzung des Tools nicht von Anfang an klar war. Es hat aber eine begrenzte Komplexität, die es dem Benutzer erlaubt, mit jedem Versuch seine Effizienz zu steigern und somit schneller zu werden und weniger Blöcke zu testen.

Die Lerneffekte von BUILD-IT in der Versuchszeit werden so interpretiert, dass die Benutzung des Tools eine gewisse Komplexität aufweist und der Benutzer erst nach einigen Versuchen sicherer wird. In der Anzahl der getesteten Blöcke zeigen sich keine Verbesserungen. Dem Benutzer ist von Anfang an klar, wie er zum Finden der Lösung vorgehen muss.

Beim Modeler Tool verbesserte sich die Versuchszeit. Analog zu BUILD-IT wurde die Versuchsperson erst mit einigen Versuchen sicherer in der Benutzung des Tools. Die Verbesserung in der Anzahl getesteter Blöcke zeigt, dass auch das Vorgehen zur Lösung der Aufgabe zunächst nicht klar war.

Das Nichtvorhandensein von Lerneffekten beim Physical Tool bestätigt, dass das Tool intuitiv klar ist, sowohl bei der Benutzung des Tools, als auch im Vorgehen zu Lösung der Aufgabe. Die Ermüdungserscheinungen bei der Anzahl der getesteten Blöcke in den letzten Versuchen kann entweder gedeutet werden als Konzentrationsverlust bei den Versuchspersonen oder als Veränderung des Vorgehens hin zum vermehrten Ausprobieren.

Zu Bemerkungen gilt es noch, dass die Lerneffekte beeinflusst sein könnten durch die wechselnde Abfolge der Versuchsvarianten und deren unterschiedliche Schwierigkeit. Innerhalb eines Tools haben nur zwei Personen die gleiche Abfolge.

Eignung der Aufgabe

Zum Schluss sollten noch einige Gedanken zur Angemessenheit der Aufgabe festgehalten werden:

- Ist das Positionierungsproblem eine geeignete Aufgabe zur Untersuchung von BUILD-IT, oder sind die alternativen Tools durch die Art der Aufgabe im Vorteil? Für das Physical Tool könnte es ein Vorteil gewesen sein, dass es direkt aus dem Positionierungsproblem hervorgegangen ist.
- BUILD-IT wurde für eine intuitive Interaktion konzipiert, während die Aufgabe eher ein analytisches Vorgehen nahe legte. Möglicherweise war die Aufgabe zu analytisch ausgelegt für das ursprüngliche Untersuchungsziel. Geeigneter wäre eine Aufgabe, die weniger Nachdenken erfordert, z.B. eine Suchaufgabe.
- Die Untersuchung wurde für Einzelbenutzer durchgeführt. Da aber BUILD-IT vor allem mit dem Ziel der Teamnutzung entwickelt wurde, hätte vielleicht auch der Aspekt der kollaborativen Nutzung noch verstärkt in die Aufgabe mit einfließen sollen oder sollte es zumindest in vergleichbaren, zukünftigen Untersuchungen.

A. Anhang

A.1. Literaturverzeichnis

- [1] Bortz, J.: *Statistik für Human- und Sozialwissenschaftler*. 6. vollständig überarbeitete und aktualisierte Auflage, Springer, Heidelberg, 2005.
- [2] Bühlmann, P.: *Stochastik*. Vorlesungsskript WS 2002/03 (D-MAVT) ETH Zürich, Zürich, 2002.
- [3] Fjeld, M., Guttormsen Schär, S., Signorello, D., Krueger, H.: Alternative Tools for Tangible Interaction: A Usability Evaluation. In *Proceedings of the IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR 2002)*, pp. 157-166, 2002.
- [4] Gähwiler, M., Fjeld, M.: *Evaluation of CAD-systems as alternative tools to a Tangible User Interface (TUI)*. Semesterarbeit am Institut für Hygiene und Arbeitsphysiologie (IHA) der ETH Zürich, Zürich, 2002, http://e-collection.ethbib.ethz.ch/ecol-pool/semarb/semarb_44.pdf (Zugriff: 10.8.2005)
- [5] Landau, S., Everitt, B.S.: *A Handbook of Statistical Analyses using SPSS*. Chapman & Hall/CRC, London 2004.
- [6] SPSS Inc.: *SPSS Hilfsmenü*. SPSS Version 13.0.1, 2004.
- [7] Verein Schweizer Mathematik- und Physiklehrer: *Formeln und Tafeln, Mathematik – Physik*. 9. durchgesehene Auflage, Orell Füssli Verlag AG, Zürich, 2001.
- [8] Züst, R.: *Einstieg ins Systems Engineering: systematisch denken, handeln und umsetzen*. 2. überarbeitete Auflage, Verlag Industrielle Organisation, Zürich, 2000.

Internet (Zugriff: 10.8.2005):

- [9] <http://de.wikipedia.org/wiki/Hauptseite>
- [10] <http://www.physics.csbsju.edu/stats/>
- [11] <http://www.ma.uni-heidelberg.de/inst/biom/lexikon/>
- [12] <http://davidmlane.com/hyperstat/index.html>

A.2. Abkürzungsverzeichnis

ANOVA	ANalysis Of Variance (deutsch: Varianzanalyse)
bzw.	beziehungsweise
d.h.	das heisst
IQR	Interquartile Range (deutsch: Interquartilabstand)
ln	Natürlicher Logarithmus
s.	siehe
sec	Sekunden
sqrt	Square root (deutsch: Quadratwurzel)
TUI	Tangible User Interface
z.B.	zum Beispiel
2log	Dyadischer Logarithmus oder auch Zweier- Logarithmus

A.3. Glossar

Bricks	Klötze in BUILD-IT zum manipulieren und navigieren
Beamer	Ein Projektor, der Daten aus einem visuellen Ausgabegerät für ein Publikum in vergrößerter Form an eine Projektionsfläche wirft
Laserpointer permutieren	Eine Laserquelle die einen Laserstrahl erzeugt verändern der Reihenfolge von Elementen einer Menge
Whiskers	Querstriche ausserhalb der Box eines Boxplots, die das untere- und obere Percentil bezeichnen

A.4. Verwendete SPSS Funktionen

Funktion	SPSS Menübefehl
Einfaktorielle ANOVA	Analysieren → Mittelwerte vergleichen → Einfaktorielle ANOVA...
Post Hoc Tukey	Analysieren → Mittelwerte vergleichen → Einfaktorielle ANOVA... → Post Hoc... → bei „Tukey“ Häkchen setzen
Standardabweichung	Analysieren → Mittelwerte vergleichen → Einfaktorielle ANOVA... → Optionen... → bei „Deskriptive Statistik“ Häkchen setzen
Levene-Test	Analysieren → Mittelwerte vergleichen → Einfaktorielle ANOVA... → Optionen... → bei Test auf „Homogenität der Varianzen“ Häkchen setzen
Kolmogorov-Smirnov-Test	Analysieren → Nichtparametrische Test → K-S bei einer Stichprobe... → bei „Normal Häkchen“ setzen
Boxplot	Graphiken → Boxplot... → Einfach...
Histogramm mit Normalverteilungskurve	Graphiken → Histogramm... → bei „Normalverteilungskurve“ Häkchen setzen

Tabelle 23: Verwendete SPSS Funktionen

A.5. Excel Datentabellen

A.5.1. Fehlerhafte Daten

Person	Tool	Versuchs- variante	Versuchs- nummer	Anzahl getestete Blöcke	Zeit pro Versuch (sec)	Versuchs- erfolg (erfüllt?)
	1=BUILD-IT 2=MODELER 3=PHYSICAL					8=Ja 9=Nein
12	1	9	T3	7	40	8
45	2	2	T2	15	90	8
53	3	1	T9	7	576	8
56	3	2	T1	4	288	8
56	3	4	T3	5	360	8
56	3	5	T7	1	864	8
61	3	9	T4	2	144	8
63	3	3	T1	6	19	8

Tabelle 24: Fehlerhafte Daten

A.5.2. Korrekte Daten

A.5.2.1. Untransformierte Daten

Primäre Daten

Person	Tool	Versuchs- variante	Versuch- nummer	Anzahl getestete Blöcke	Zeit pro Versuch (sec)	Versuchs- erfolg (erfüllt?)
	1=BUILD-IT 2=MODELER 3=PHYSICAL					8=Ja 9=Nein
11	1	1	T1	4	14	8
11	1	5	T2	1	11	8
11	1	3	T3	1	8	8
11	1	9	T4	1	7	8
11	1	7	T5	5	28	8
11	1	2	T6	6	32	8
11	1	6	T7	1	9	8
11	1	4	T8	5	28	8
11	1	10	T9	4	25	8
11	1	8	T10	11	61	8
12	1	5	T1	8	60	8
12	1	3	T2	4	25	8
12	1	9	T3	7	41	8

12	1	7	T4	2	9	8
12	1	2	T5	15	69	8
12	1	6	T6	3	10	8
12	1	4	T7	5	22	8
12	1	10	T8	3	17	8
12	1	8	T9	3	11	8
12	1	1	T10	4	23	8
13	1	3	T1	4	17	8
13	1	9	T2	1	5	8
13	1	7	T3	2	8	8
13	1	2	T4	4	16	8
13	1	6	T5	2	11	8
13	1	4	T6	4	13	8
13	1	10	T7	1	4	8
13	1	8	T8	1	4	8
13	1	1	T9	13	51	9
13	1	5	T10	2	7	8
14	1	9	T1	1	11	8
14	1	7	T2	6	40	8
14	1	2	T3	3	16	8
14	1	6	T4	2	12	8
14	1	4	T5	4	26	8
14	1	10	T6	1	13	8
14	1	8	T7	4	20	8
14	1	1	T8	5	29	8
14	1	5	T9	1	12	8
14	1	3	T10	4	27	8
15	1	7	T1	9	26	8
15	1	2	T2	18	81	8
15	1	6	T3	8	34	8
15	1	4	T4	7	25	8
15	1	10	T5	1	7	8
15	1	8	T6	18	100	8
15	1	1	T7	6	26	8
15	1	5	T8	6	21	8
15	1	3	T9	4	15	8
15	1	9	T10	10	39	8
16	1	2	T1	13	46	8
16	1	6	T2	9	23	8
16	1	4	T3	4	11	8
16	1	10	T4	1	4	8
16	1	8	T5	7	19	8
16	1	1	T6	3	15	8
16	1	5	T7	3	12	8
16	1	3	T8	4	14	8
16	1	9	T9	1	6	8

16	1	7	T10	4	13	8
17	1	6	T1	3	15	8
17	1	4	T2	4	16	8
17	1	10	T3	1	6	8
17	1	8	T4	1	6	8
17	1	1	T5	6	25	8
17	1	5	T6	2	10	8
17	1	3	T7	8	24	8
17	1	9	T8	1	7	8
17	1	7	T9	2	8	8
17	1	2	T10	21	63	8
18	1	4	T1	13	51	8
18	1	10	T2	1	8	8
18	1	8	T3	5	14	8
18	1	1	T4	7	26	8
18	1	5	T5	7	14	8
18	1	3	T6	5	12	8
18	1	9	T7	2	8	8
18	1	7	T8	8	20	8
18	1	2	T9	6	12	8
18	1	6	T10	2	7	8
19	1	10	T1	1	7	8
19	1	8	T2	2	7	8
19	1	1	T3	11	39	8
19	1	5	T4	2	5	8
19	1	3	T5	8	20	8
19	1	9	T6	1	3	8
19	1	7	T7	2	5	8
19	1	2	T8	6	18	8
19	1	6	T9	16	43	8
19	1	4	T10	8	11	8
20	1	8	T1	7	33	8
20	1	1	T2	5	13	8
20	1	5	T3	8	27	8
20	1	3	T4	11	43	8
20	1	9	T5	4	12	8
20	1	7	T6	3	6	8
20	1	2	T7	9	26	8
20	1	6	T8	7	24	8
20	1	4	T9	2	8	8
20	1	10	T10	2	12	9
21	1	1	T1	8	38	8
21	1	5	T2	3	14	8
21	1	3	T3	2	15	8
21	1	9	T4	1	5	8
21	1	7	T5	3	9	8

21	1	2	T6	16	78	9
21	1	6	T7	2	15	8
21	1	4	T8	5	28	8
21	1	10	T9	1	8	8
21	1	8	T10	3	14	8
22	1	5	T1	8	27	8
22	1	3	T2	4	25	8
22	1	9	T3	1	9	8
22	1	7	T4	1	9	8
22	1	2	T5	4	20	8
22	1	6	T6	8	34	8
22	1	4	T7	5	22	8
22	1	10	T8	1	5	8
22	1	8	T9	3	12	8
22	1	1	T10	4	24	9
23	1	3	T1	13	90	8
23	1	9	T2	2	15	8
23	1	7	T3	1	12	8
23	1	2	T4	7	35	8
23	1	6	T5	15	84	8
23	1	4	T6	6	28	8
23	1	10	T7	4	17	8
23	1	8	T8	1	7	8
23	1	1	T9	10	39	8
23	1	5	T10	5	19	8
24	1	9	T1	1	9	8
24	1	7	T2	8	32	8
24	1	2	T3	3	14	8
24	1	6	T4	2	9	8
24	1	4	T5	3	17	8
24	1	10	T6	5	15	8
24	1	8	T7	5	19	8
24	1	1	T8	4	13	8
24	1	5	T9	10	21	8
24	1	3	T10	4	13	8
25	1	7	T1	4	20	8
25	1	2	T2	10	49	8
25	1	6	T3	2	13	8
25	1	4	T4	4	18	8
25	1	10	T5	1	6	8
25	1	8	T6	7	24	8
25	1	1	T7	7	26	8
25	1	5	T8	4	13	8
25	1	3	T9	5	17	8
25	1	9	T10	1	4	8
26	1	2	T1	11	41	8

26	1	6	T2	2	10	9
26	1	4	T3	3	12	8
26	1	10	T4	1	7	8
26	1	8	T5	8	18	8
26	1	1	T6	8	13	8
26	1	5	T7	8	19	8
26	1	3	T8	3	7	8
26	1	9	T9	1	5	8
26	1	7	T10	8	27	8
27	1	6	T1	2	22	8
27	1	4	T2	7	51	8
27	1	10	T3	2	17	8
27	1	8	T4	1	8	8
27	1	1	T5	14	87	8
27	1	5	T6	2	9	8
27	1	3	T7	7	37	8
27	1	9	T8	1	5	8
27	1	7	T9	2	8	8
27	1	2	T10	5	20	8
28	1	4	T1	8	33	8
28	1	10	T2	13	43	8
28	1	8	T3	2	5	8
28	1	1	T4	4	20	8
28	1	5	T5	6	16	8
28	1	3	T6	4	15	9
28	1	9	T7	2	9	8
28	1	7	T8	4	13	8
28	1	2	T9	13	41	9
28	1	6	T10	10	30	8
29	1	10	T1	1	7	8
29	1	8	T2	2	9	8
29	1	1	T3	13	62	8
29	1	5	T4	3	12	8
29	1	3	T5	6	21	8
29	1	9	T6	1	6	8
29	1	7	T7	2	7	8
29	1	2	T8	13	49	8
29	1	6	T9	10	35	8
29	1	4	T10	5	17	8
30	1	8	T1	11	53	8
30	1	1	T2	5	30	8
30	1	5	T3	12	49	8
30	1	3	T4	10	44	8
30	1	9	T5	16	73	8
30	1	7	T6	2	12	8
30	1	2	T7	5	21	8

30	1	6	T8	2	10	8
30	1	4	T9	3	17	8
30	1	10	T10	5	21	8
31	2	1	T1	10	370	9
31	2	5	T2	1	26	8
31	2	3	T3	9	36	9
31	2	9	T4	3	28	8
31	2	7	T5	2	20	8
31	2	2	T6	7	39	9
31	2	6	T7	1	25	8
31	2	4	T8	3	28	8
31	2	10	T9	4	45	8
31	2	8	T10	2	20	8
32	2	5	T1	3	56	8
32	2	3	T2	2	11	8
32	2	9	T3	4	23	8
32	2	7	T4	2	15	9
32	2	2	T5	4	20	8
32	2	6	T6	3	14	8
32	2	4	T7	1	10	8
32	2	10	T8	3	11	8
32	2	8	T9	1	10	8
32	2	1	T10	2	18	9
33	2	3	T1	5	180	8
33	2	9	T2	4	60	9
33	2	7	T3	6	64	8
33	2	2	T4	10	72	8
33	2	6	T5	3	20	8
33	2	4	T6	4	23	8
33	2	10	T7	3	20	8
33	2	8	T8	1	15	8
33	2	1	T9	4	30	8
33	2	5	T10	2	29	8
34	2	9	T1	2	76	9
34	2	7	T2	-	-	9
34	2	2	T3	3	47	8
34	2	6	T4	1	28	8
34	2	4	T5	3	24	8
34	2	10	T6	4	33	9
34	2	8	T7	2	20	8
34	2	1	T8	4	34	8
34	2	5	T9	1	18	8
34	2	3	T10	3	24	9
35	2	7	T1	5	27	8
35	2	2	T2	17	53	8
35	2	6	T3	3	16	8

35	2	4	T4	2	20	8
35	2	10	T5	3	26	8
35	2	8	T6	2	13	8
35	2	1	T7	5	35	8
35	2	5	T8	6	22	8
35	2	3	T9	3	18	8
35	2	9	T10	4	27	8
36	2	2	T1	6	55	9
36	2	6	T2	5	37	8
36	2	4	T3	6	43	8
36	2	10	T4	4	20	8
36	2	8	T5	1	17	8
36	2	1	T6	4	45	9
36	2	5	T7	1	24	8
36	2	3	T8	3	19	9
36	2	9	T9	3	25	8
36	2	7	T10	2	13	9
37	2	6	T1	6	83	8
37	2	4	T2	12	53	9
37	2	10	T3	5	27	8
37	2	8	T4	2	16	8
37	2	1	T5	5	38	9
37	2	5	T6	2	25	8
37	2	3	T7	3	12	8
37	2	9	T8	5	31	8
37	2	7	T9	3	16	8
37	2	2	T10	6	44	8
38	2	4	T1	9	55	9
38	2	10	T2	2	20	8
38	2	8	T3	2	17	8
38	2	1	T4	8	30	8
38	2	5	T5	5	24	8
38	2	3	T6	6	22	9
38	2	9	T7	1	26	8
38	2	7	T8	7	19	9
38	2	2	T9	8	33	9
38	2	6	T10	4	17	8
39	2	10	T1	5	27	8
39	2	8	T2	8	29	8
39	2	1	T3	3	29	8
39	2	5	T4	4	20	8
39	2	3	T5	4	19	8
39	2	9	T6	5	20	8
39	2	7	T7	2	21	9
39	2	2	T8	8	29	8
39	2	6	T9	3	18	8

39	2	4	T10	6	17	8
40	2	8	T1	4	24	8
40	2	1	T2	5	26	9
40	2	5	T3	6	30	8
40	2	3	T4	2	12	8
40	2	9	T5	4	19	8
40	2	7	T6	3	17	8
40	2	2	T7	8	60	9
40	2	6	T8	4	16	8
40	2	4	T9	2	13	8
40	2	10	T10	3	23	9
41	2	1	T1	9	60	9
41	2	5	T2	3	19	8
41	2	3	T3	4	18	8
41	2	9	T4	5	22	8
41	2	7	T5	2	13	9
41	2	2	T6	9	29	8
41	2	6	T7	5	16	8
41	2	4	T8	4	14	8
41	2	10	T9	2	20	8
41	2	8	T10	1	7	8
42	2	5	T1	10	77	9
42	2	3	T2	2	25	9
42	2	9	T3	6	30	9
42	2	7	T4	3	23	8
42	2	2	T5	5	33	9
42	2	6	T6	6	30	8
42	2	4	T7	7	24	8
42	2	10	T8	4	21	8
42	2	8	T9	1	9	8
42	2	1	T10	5	32	9
43	2	3	T1	2	28	8
43	2	9	T2	3	24	8
43	2	7	T3	3	22	8
43	2	2	T4	10	50	8
43	2	6	T5	1	9	8
43	2	4	T6	1	16	8
43	2	10	T7	3	10	8
43	2	8	T8	1	8	8
43	2	1	T9	8	39	9
43	2	5	T10	2	16	8
44	2	9	T1	4	21	8
44	2	7	T2	2	14	9
44	2	2	T3	4	20	8
44	2	6	T4	1	8	8
44	2	4	T5	3	14	8

44	2	10	T6	3	10	8
44	2	8	T7	1	8	8
44	2	1	T8	3	15	8
44	2	5	T9	1	9	8
44	2	3	T10	1	9	8
45	2	7	T1	2	23	9
45	2	2	T2	15	110	8
45	2	6	T3	2	14	8
45	2	4	T4	3	12	8
45	2	10	T5	3	11	8
45	2	8	T6	4	13	8
45	2	1	T7	5	19	9
45	2	5	T8	5	23	8
45	2	3	T9	6	20	8
45	2	9	T10	5	16	8
46	2	2	T1	10	210	8
46	2	6	T2	6	78	8
46	2	4	T3	4	56	9
46	2	10	T4	6	45	8
46	2	8	T5	4	30	8
46	2	1	T6	6	90	9
46	2	5	T7	4	40	8
46	2	3	T8	4	38	8
46	2	9	T9	5	42	8
46	2	7	T10	4	31	8
47	2	6	T1	8	66	8
47	2	4	T2	2	25	8
47	2	10	T3	2	14	8
47	2	8	T4	1	9	8
47	2	1	T5	5	50	8
47	2	5	T6	6	20	8
47	2	3	T7	4	15	8
47	2	9	T8	8	25	8
47	2	7	T9	3	16	8
47	2	2	T10	5	29	8
48	2	4	T1	4	58	8
48	2	10	T2	2	19	8
48	2	8	T3	4	31	8
48	2	1	T4	4	31	8
48	2	5	T5	1	16	8
48	2	3	T6	2	13	9
48	2	9	T7	3	22	9
48	2	7	T8	4	95	8
48	2	2	T9	3	19	8
48	2	6	T10	1	11	8
49	2	10	T1	2	17	8

49	2	8	T2	2	14	8
49	2	1	T3	5	35	9
49	2	5	T4	4	20	8
49	2	3	T5	4	17	8
49	2	9	T6	5	15	8
49	2	7	T7	2	20	8
49	2	2	T8	3	26	8
49	2	6	T9	1	11	8
49	2	4	T10	2	18	8
50	2	8	T1	2	33	8
50	2	1	T2	5	41	9
50	2	5	T3	1	25	8
50	2	3	T4	4	54	8
50	2	9	T5	4	39	8
50	2	7	T6	2	28	9
50	2	2	T7	5	50	8
50	2	6	T8	3	23	8
50	2	4	T9	5	23	8
50	2	10	T10	3	10	8
51	3	1	T1	2	15	8
51	3	5	T2	2	12	8
51	3	3	T3	5	35	8
51	3	9	T4	3	27	8
51	3	7	T5	4	28	8
51	3	2	T6	4	44	8
51	3	6	T7	1	5	8
51	3	4	T8	1	5	8
51	3	10	T9	6	69	8
51	3	8	T10	1	5	8
52	3	5	T1	5	8	8
52	3	3	T2	6	12	8
52	3	9	T3	3	9	8
52	3	7	T4	2	8	8
52	3	2	T5	8	19	8
52	3	6	T6	6	11	8
52	3	4	T7	5	13	8
52	3	10	T8	2	6	8
52	3	8	T9	1	4	8
52	3	1	T10	13	39	8
53	3	3	T1	1	7	8
53	3	9	T2	2	18	8
53	3	7	T3	1	6	8
53	3	2	T4	2	8	8
53	3	6	T5	1	4	8
53	3	4	T6	1	5	8
53	3	10	T7	5	21	8

53	3	8	T8	1	9	8
53	3	1	T9	7	40	8
53	3	5	T10	1	4	8
54	3	9	T1	1	4	8
54	3	7	T2	2	11	8
54	3	2	T3	11	40	8
54	3	6	T4	2	10	8
54	3	4	T5	2	10	8
54	3	10	T6	3	11	8
54	3	8	T7	1	4	8
54	3	1	T8	4	19	8
54	3	5	T9	1	4	8
54	3	3	T10	3	16	8
55	3	7	T1	4	11	8
55	3	2	T2	12	47	8
55	3	6	T3	1	3	8
55	3	4	T4	1	7	8
55	3	10	T5	8	47	8
55	3	8	T6	5	19	8
55	3	1	T7	4	24	8
55	3	5	T8	2	9	8
55	3	3	T9	3	14	8
55	3	9	T10	2	5	8
56	3	2	T1	4	20	8
56	3	6	T2	1	7	8
56	3	4	T3	5	25	8
56	3	10	T4	6	30	8
56	3	8	T5	8	35	8
56	3	1	T6	4	16	8
56	3	5	T7	1	6	8
56	3	3	T8	1	9	8
56	3	9	T9	1	6	8
56	3	7	T10	2	14	8
57	3	6	T1	1	8	8
57	3	4	T2	5	35	8
57	3	10	T3	4	20	8
57	3	8	T4	1	6	8
57	3	1	T5	2	15	8
57	3	5	T6	1	5	8
57	3	3	T7	2	8	8
57	3	9	T8	3	15	8
57	3	7	T9	14	66	8
57	3	2	T10	9	39	8
58	3	4	T1	3	18	8
58	3	10	T2	1	8	8
58	3	8	T3	1	4	8

58	3	1	T4	3	10	8
58	3	5	T5	1	4	8
58	3	3	T6	1	3	8
58	3	9	T7	3	8	8
58	3	7	T8	2	7	8
58	3	2	T9	2	6	8
58	3	6	T10	1	5	8
59	3	10	T1	1	5	8
59	3	8	T2	1	8	8
59	3	1	T3	12	57	8
59	3	5	T4	1	3	8
59	3	3	T5	1	7	8
59	3	9	T6	5	13	8
59	3	7	T7	1	5	8
59	3	2	T8	2	10	8
59	3	6	T9	1	5	8
59	3	4	T10	3	15	8
60	3	8	T1	3	6	8
60	3	1	T2	9	55	8
60	3	5	T3	1	3	8
60	3	3	T4	1	6	8
60	3	9	T5	2	10	8
60	3	7	T6	3	14	8
60	3	2	T7	4	21	8
60	3	6	T8	1	3	8
60	3	4	T9	2	12	8
60	3	10	T10	1	4	8
61	3	1	T1	11	31	8
61	3	5	T2	3	9	8
61	3	3	T3	5	17	8
61	3	9	T4	2	10	8
61	3	7	T5	6	22	8
61	3	2	T6	7	24	8
61	3	6	T7	1	5	8
61	3	4	T8	1	6	8
61	3	10	T9	1	5	8
61	3	8	T10	2	8	8
62	3	5	T1	3	11	8
62	3	3	T2	2	9	8
62	3	9	T3	4	18	8
62	3	7	T4	1	8	8
62	3	2	T5	4	19	8
62	3	6	T6	1	7	8
62	3	4	T7	1	7	8
62	3	10	T8	4	16	8
62	3	8	T9	3	13	8

62	3	1	T10	4	25	8
63	3	3	T1	6	17	8
63	3	9	T2	2	5	8
63	3	7	T3	1	4	8
63	3	2	T4	5	11	8
63	3	6	T5	1	4	8
63	3	4	T6	2	5	8
63	3	10	T7	2	5	8
63	3	8	T8	1	4	8
63	3	1	T9	2	5	8
63	3	5	T10	1	5	8
64	3	9	T1	3	10	8
64	3	7	T2	9	30	8
64	3	2	T3	4	11	8
64	3	6	T4	2	3	8
64	3	4	T5	4	13	8
64	3	10	T6	4	15	8
64	3	8	T7	1	4	8
64	3	1	T8	6	17	8
64	3	5	T9	2	6	8
64	3	3	T10	3	11	8
65	3	7	T1	1	3	8
65	3	2	T2	3	14	8
65	3	6	T3	1	4	8
65	3	4	T4	2	10	8
65	3	10	T5	1	4	8
65	3	8	T6	1	5	8
65	3	1	T7	6	24	8
65	3	5	T8	1	3	8
65	3	3	T9	1	3	8
65	3	9	T10	2	5	8
66	3	2	T1	2	12	8
66	3	6	T2	2	12	8
66	3	4	T3	2	14	8
66	3	10	T4	2	12	8
66	3	8	T5	1	5	8
66	3	1	T6	6	14	8
66	3	5	T7	5	8	8
66	3	3	T8	2	11	8
66	3	9	T9	1	6	8
66	3	7	T10	11	15	8
67	3	6	T1	2	9	8
67	3	4	T2	1	9	8
67	3	10	T3	2	14	8
67	3	8	T4	9	66	8
67	3	1	T5	3	21	8

67	3	5	T6	1	2	8
67	3	3	T7	2	8	8
67	3	9	T8	7	52	8
67	3	7	T9	1	3	8
67	3	2	T10	1	6	8
68	3	4	T1	1	14	8
68	3	10	T2	6	47	8
68	3	8	T3	2	17	8
68	3	1	T4	4	30	8
68	3	5	T5	1	5	8
68	3	3	T6	3	25	8
68	3	9	T7	2	9	8
68	3	7	T8	1	9	8
68	3	2	T9	5	36	8
68	3	6	T10	12	90	8
69	3	10	T1	3	5	8
69	3	8	T2	2	6	8
69	3	1	T3	1	5	8
69	3	5	T4	2	3	8
69	3	3	T5	1	5	8
69	3	9	T6	4	6	8
69	3	7	T7	3	8	8
69	3	2	T8	3	7	8
69	3	6	T9	2	4	8
69	3	4	T10	2	5	8
70	3	8	T1	1	5	8
70	3	1	T2	6	38	8
70	3	5	T3	1	6	8
70	3	3	T4	3	30	8
70	3	9	T5	2	13	8
70	3	7	T6	1	7	8
70	3	2	T7	3	18	8
70	3	6	T8	1	3	8
70	3	4	T9	1	5	8
70	3	10	T10	3	14	8

Tabelle 25: Primäre, untransformierte, korrekte Daten

Sekundäre Daten (Kontrolldaten)

Für die Bedeutung der Fragen und der entsprechenden Antworten siehe Tabelle 27.

Person	Geschlecht	Alter	Fragen											
			1.1	1.2					1.3	2.1	2.2	2.3	3.1	3.2
				-2	-1	0	1	2						
	6=Männlich 7=Weiblich		Für die Bedeutung der Fragen und der entsprechenden Antwortenskala [-2,-1,0,1,2] siehe Tabelle 27.											
11	6	22	0	-2	-1		1		-2	1	-2	1	1	-2
12	6	23	1		-1		1		-2	0	-2	1	0	-2
13	6	23	1	-2	-1	0	1		-2	1	-2	2	2	-2
14	6	23	1	-2	-1		1	2	-2	0	-2	1	-1	-2
15	6	22	0	-2	-1				-2	0	-2	2	0	-2
16	6	22	1		-1		1		-2	0	-2	2	0	-2
17	6	25	1		-1		1		-2	0	-2	1	1	-2
18	7	22	0		-1		1		-2	0	-2	2	1	-2
19	6	21	1	-2	-1	0	1		-2	0	-2	1	1	-2
20	6	24	1		-1		1	2	-2	0	-2	2	1	-2
21	6	24	1		-1		1	2	-2	1	-2	2	1	-2
22	6	21	1	-2	-1				-2	1	-2	2	0	-2
23	7	23	0		-1				-2	0	-2	2	1	-2
24	6	20	1	-2	-1	0			-2	1	-2	2	0	-2
25	6	22	1	-2	-1		1		-2	2	0	2	1	-2
26	6	22	1	-2	-1		1		-2	1	-2	2	1	-2
27	7	23	0		-1		1		-2	0	-2	2	1	-2
28	7	23	0		-1		1		-2	0	-2	1	-1	-2
29	7	25	0	-2					-2	0	-2	2	1	1
30	6	34	0	-2	-1		1		-2	1	-2	2	1	-2
31	6	20	0	-2	-1		1		-2	0	-2	-2	0	-2
32	6	21	0		-1	0			-2	1	-2	1	1	-2
33	6	22	1		-1		1		-2	0	-2	0	1	-2
34	6	23	0	-2	-1				-2	0	-1	0	1	-2
35	6	24	0	-2	-1				-2	1	-2	0	-1	-2
36	6	23	0	-2						0	-2	1	0	-2
37	6	24	1		-1				-2	1	-2	1	0	-2
38	6	23	1		-1		1	2	-2	1	-1	1	1	-2
39	6	22	1	-2	-1				-2	-1	-1	1	1	-2
40	6	21	1	-2	-1		1	2	-2	1	-1	0	1	-2
41	6	24	1	-2	-1		1		-2	1	-2	1	0	-2
42	6	27	0		-1		1		-2	0	-2	1	0	-2
43	6	24	1	-2	-1		1		-2	1	-2	0	0	-2
44	6	24	1		-1		1		-2	1	-1	1	1	-2
45	6	22	0	-2	-1		1		-2	1	-2	1	1	-2
46	6	22	0	-2	-1		1		-2	1	-2	1	1	-2

47	6	21	1		-1			2	-2	1	-2	0	1	-2
48	6	22	1	-2	-1		1		-2	0	-2	1	1	-2
49	6	23	1		-1		1		-2	1	-2	1	1	-2
50	6	21	1		-1		1		-2	1	-2	1	2	-2
51	6	22	1		-1	0			-2	0	-2	2	-2	-2
52	6	22	0		-1		1		-2	-1	-2	2	1	-2
53	6	23	1	-2	-1		1		-1	1	-2	2	0	-2
54	7	22	1		-1				-2	0	-2	2	0	-2
55	6	20	0	-2	-1		1		-2	0	-2	1	1	-2
56	6	22	0		-1		1	2	-2	1	-2	2	1	-1
57	6	24	0	-2	-1		1		-2	0	-2	2	0	-1
58	6	26	1		-1		1		-2	1	-2	2	2	-2
59	6	24	1	-2	-1	0	1		-2	0	-2	2	1	-2
60	6	23	0		-1		1		-2	1	-1	2	1	-2
61	6	25	1		-1				-2	0	-2	1	1	-2
62	6	23	0	-2	-1		1		1	0	-1	2	2	-1
63	6	25	1	-2	-1		1	2	-2	1	-2	2	-1	-2
64	6	26	1		-1		1		-2	1	-2	2	0	-2
65	6	25	1	-2	-1	0	1	2	-2	1	-2	2	2	2
66	6	26	1		-1	0			-2	1	-2	2	2	-2
67	6	22	1	-2	-1		1		-2	1	-2	2	1	-2
68	6	21	1	-2	-1	0	1	2	-2	1	-2	2	1	-1
69	6	23	1		-1		1	2	-2	1	-1	1	0	-2
70	6	25	1			0			-2	1	-1	1	-1	-2

Tabelle 26: Sekundäre Daten

Frage	Antwort				
	-2	-1	0	1	2
1.1 Wie oft benutzen Sie einen Rechner (PC, Workstation)?					
weniger als 1 Stunde Monatlich	minimum 1 Stunde Monatlich	minimum 1 Stunde Wöchentlich	minimum 1 Stunde täglich		
1.2 Wofür benutzen Sie einen Rechner? (Mehrfachnennungen möglich)					
Freizeit (Spiele usw.)	Kommunikation (e-mail, WWW...)	Planung und Konstruktion	Code implementieren, Textverarbeitung, Simulation	Andere Zwecke	
1.3 Wie gut sind Ihre Vorkenntnisse in BUILD-IT?					
Kenne BUILD-IT nicht	Habe von BUILD-IT gehört oder gelesen	Habe gesehen, wie BUILD-IT funktioniert	Habe BUILD-IT schon benutzt		
2.1 Wie empfanden Sie den Schwierigkeitsgrad der Aufgabenstellung?					
sehr schwer	schwer	neutral	leicht	sehr leicht	
2.2 Haben Sie früher ähnliche Aufgaben gelöst?					
noch nie	einige Male	ziemlich oft	sehr oft		
2.3 War die Aufgabenstellung verständlich formuliert?					
überhaupt nicht verständlich	nicht sehr verständlich	neutral	gut verständlich	sehr gut verständlich	
3.1 Wie schätzen Sie das Werkzeug zur Lösung dieser Aufgabe ein?					
überhaupt nicht hilfreich	nicht sehr hilfreich	angemessen	hilfreich	sehr hilfreich	
3.2 Wie oft haben Sie dieses Werkzeug eingesetzt?					
noch nie	einige Male	ziemlich oft	sehr oft	keine Antwort	

Tabelle 27: Fragebogen

A.5.2.2. Transformierte Daten

Person	Tool	Arithmetisches Mittel Zeit	Arithmetisches Mittel Anzahl Blöcke	sqrt Zeit	sqrt Anzahl Blöcke	In Zeit	In Anzahl Blöcke	2log Zeit	2log Anzahl Blöcke
	1=BUILD-IT 2=MODELER 3=PHYSICAL								
11	1	22.3	3.9	4.72	1.97	3.1	1.36	4.48	1.96
12	1	28.7	5.4	5.36	2.32	3.36	1.69	4.84	2.43
13	1	13.6	3.4	3.69	1.84	2.61	1.22	3.77	1.77
14	1	20.6	3.1	4.54	1.76	3.03	1.13	4.36	1.63
15	1	37.4	8.7	6.12	2.95	3.62	2.16	5.22	3.12
16	1	16.3	4.9	4.04	2.21	2.79	1.59	4.03	2.29
17	1	18	4.9	4.24	2.21	2.89	1.59	4.17	2.29
18	1	17.2	5.6	4.15	2.37	2.84	1.72	4.1	2.49
19	1	15.8	5.7	3.97	2.39	2.76	1.74	3.98	2.51
20	1	20.4	5.8	4.52	2.41	3.02	1.76	4.35	2.54
21	1	22.4	4.4	4.73	2.1	3.11	1.48	4.49	2.14
22	1	18.7	3.9	4.32	1.97	2.93	1.36	4.22	1.96
23	1	34.6	6.4	5.88	2.53	3.54	1.86	5.11	2.68
24	1	16.2	4.5	4.02	2.12	2.79	1.5	4.02	2.17
25	1	19	4.5	4.36	2.12	2.94	1.5	4.25	2.17
26	1	15.9	5.3	3.99	2.3	2.77	1.67	3.99	2.41
27	1	26.4	4.3	5.14	2.07	3.27	1.46	4.72	2.1
28	1	22.5	6.6	4.74	2.57	3.11	1.89	4.49	2.72
29	1	22.5	5.6	4.74	2.37	3.11	1.72	4.49	2.49
30	1	33	7.1	5.74	2.66	3.5	1.96	5.04	2.83
31	2	63.7	4.2	7.98	2.05	4.15	1.44	5.99	2.07
32	2	18.8	2.5	4.34	1.58	2.93	0.92	4.23	1.32
33	2	51.3	4.2	7.16	2.05	3.94	1.44	5.68	2.07
34	2	33.8	2.6	5.81	1.6	3.52	0.94	5.08	1.36
35	2	25.7	5	5.07	2.24	3.25	1.61	4.68	2.32
36	2	29.8	3.5	5.46	1.87	3.39	1.25	4.9	1.81
37	2	34.5	4.9	5.87	2.21	3.54	1.59	5.11	2.29
38	2	26.3	5.2	5.13	2.28	3.27	1.65	4.72	2.38
39	2	22.9	4.8	4.79	2.19	3.13	1.57	4.52	2.26
40	2	24	4.1	4.9	2.02	3.18	1.41	4.58	2.04
41	2	21.8	4.4	4.67	2.1	3.08	1.48	4.45	2.14
42	2	30.4	4.9	5.51	2.21	3.41	1.59	4.83	2.29
43	2	22.2	3.4	4.71	1.84	3.1	1.22	4.47	1.77
44	2	12.8	2.3	3.58	1.52	2.55	0.83	3.68	1.2

45	2	26.1	5	5.11	2.24	3.26	1.61	4.59	2.32
46	2	66	5.3	8.12	2.3	4.19	1.67	6.04	2.41
47	2	26.9	4.4	5.19	2.1	3.29	1.48	4.75	2.14
48	2	31.5	2.8	5.61	1.67	3.45	1.03	4.98	1.49
49	2	19.3	3	4.39	1.73	2.96	1.1	4.27	1.58
50	2	32.6	3.4	5.71	1.84	3.48	1.22	5.03	1.77
51	3	24.5	2.9	4.95	1.7	3.2	1.06	4.61	1.54
52	3	12.9	5.1	3.59	2.26	2.56	1.63	3.69	2.35
53	3	12.2	2.2	3.49	1.48	2.5	0.79	6.04	1.14
54	3	12.9	3	3.59	1.73	2.56	1.1	3.69	1.58
55	3	18.6	4.2	4.31	2.05	2.92	1.44	4.22	2.07
56	3	16.8	3.3	4.1	1.82	2.82	1.19	7.35	1.72
57	3	21.7	4.2	4.66	2.05	3.08	1.44	4.44	2.07
58	3	7.3	1.8	2.7	1.34	1.99	0.59	2.87	0.85
59	3	12.8	2.8	3.58	1.67	2.55	1.03	3.68	1.49
60	3	13.4	2.7	3.66	1.64	2.6	0.99	3.74	1.43
61	3	13.7	3.9	3.7	1.97	2.62	1.36	4.76	1.96
62	3	13.3	2.7	3.65	1.64	2.59	0.99	3.73	1.43
63	3	6.5	2.3	2.55	1.52	1.87	0.83	2.74	1.2
64	3	12	3.8	3.46	1.95	2.48	1.34	3.58	1.93
65	3	7.5	1.9	2.74	1.38	2.01	0.64	2.91	0.93
66	3	10.9	3.4	3.3	1.84	2.39	1.22	3.45	1.77
67	3	19	2.9	4.36	1.7	2.94	1.06	4.25	1.54
68	3	28.2	3.7	5.31	1.92	3.34	1.31	4.81	1.89
69	3	5.4	2.3	2.32	1.52	1.69	0.83	2.43	1.2
70	3	13.9	2.2	3.73	1.48	2.63	0.79	3.8	1.14

Tabelle 28: Transformierte Daten

A.6. SPSS Ausgabetafeln

A.6.1. Fehlerhafte Daten

A.6.1.1. Untransformierte Daten

		N	Mittelwert	Standardabweichung	Standardfehler	95%-Konfidenzintervall für den Mittelwert		Minimum	Maximum
						Untergrenze	Obergrenze		
Zeit (sec)	BUILD-IT	200	22.07	18.093	1.279	19.55	24.59	3	100
	MODELER	199	30.80	33.997	2.410	26.05	35.56	7	370
	PHYSICAL	200	24.84	79.715	5.637	13.72	35.95	2	864
	Gesamt	599	25.89	51.181	2.091	21.79	30.00	2	864
Anzahl Blöcke	BUILD-IT	200	5.20	4.088	.289	4.63	5.77	1	21
	MODELER	199	4.00	2.539	.180	3.65	4.35	1	17
	PHYSICAL	200	3.07	2.664	.188	2.69	3.44	1	14
	Gesamt	599	4.09	3.290	.134	3.82	4.35	1	21

Tabelle 29: SPSS Ausgabe; Standardabweichungen; fehlerhafte untransformierte Daten

A.6.1.2. Arithmetisches Mittel

		N	Mittelwert	Standardabweichung	Standardfehler	95%-Konfidenzintervall für den Mittelwert		Minimum	Maximum
						Untergrenze	Obergrenze		
Arithmetisches Mittel Zeit	BUILD-IT	20	22.070	6.7132	1.5011	18.928	25.212	13.6	37.4
	MODELER	20	30.819	14.0394	3.1393	24.248	37.390	12.8	66.0
	PHYSICAL	20	24.835	35.0378	7.8347	8.437	41.233	5.4	162.9
	Gesamt	60	25.908	22.0656	2.8487	20.208	31.608	5.4	162.9
Arithmetisches Mittel Anzahl Blöcke	BUILD-IT	20	5.200	1.3353	.2986	4.575	5.825	3.1	8.7
	MODELER	20	3.993	.9847	.2202	3.532	4.454	2.3	5.3
	PHYSICAL	20	3.065	.8750	.1956	2.656	3.474	1.8	5.1
	Gesamt	60	4.086	1.3820	.1784	3.729	4.443	1.8	8.7

Tabelle 30: SPSS Ausgabe; Standardabweichungen; fehlerhafte Daten, arithmetisches Mittel

A.6.1.3. Transformierte Daten – sqrt

		N	Mittelwert	Standardabweichung	Standardfehler	95%-Konfidenzintervall für den Mittelwert		Minimum	Maximum
						Untergrenze	Obergrenze		
sqrt Zeit	BUILD-IT	20	4.6506	.68230	.15257	4.3312	4.9699	3.69	6.12
	MODELER	20	5.4365	1.15347	.25792	4.8966	5.9763	3.58	8.12
	PHYSICAL	20	4.4288	2.34420	.52418	3.3317	5.5259	2.32	12.76
	Gesamt	60	4.8386	1.59315	.20567	4.4271	5.2502	2.32	12.76
sqrt Anzahl Blöcke	BUILD-IT	20	2.2632	.28682	.06413	2.1289	2.3974	1.76	2.95
	MODELER	20	1.9827	.25554	.05714	1.8631	2.1023	1.52	2.30
	PHYSICAL	20	1.7341	.24665	.05515	1.6187	1.8496	1.34	2.26
	Gesamt	60	1.9933	.33853	.04370	1.9059	2.0808	1.34	2.95

Tabelle 31: SPSS Ausgabe; Standardabweichungen; fehlerhafte transformierte (sqrt) Daten

A.6.2. Korrekte Daten

A.6.2.1. Untransformierte Daten

		N	Mittelwert	Standardabweichung	Standardfehler	95%-Konfidenzintervall für den Mittelwert		Minimum	Maximum
						Untergrenze	Obergrenze		
Zeit (sec)	BUILD-IT	200	22.08	18.098	1.280	19.55	24.60	3	100
	MODELER	199	31.01	34.169	2.422	26.23	35.78	7	370
	PHYSICAL	200	14.18	13.983	.989	12.23	16.12	2	90
	Gesamt	599	22.40	24.656	1.007	20.43	24.38	2	370
Anzahl Blöcke	BUILD-IT	200	5.20	4.088	.289	4.63	5.77	1	21
	MODELER	199	4.00	2.539	.180	3.65	4.35	1	17
	PHYSICAL	200	3.07	2.664	.188	2.69	3.44	1	14
	Gesamt	599	4.09	3.290	.134	3.82	4.35	1	21

Tabelle 32: SPSS Ausgabe; Standardabweichungen; korrekte untransformierte Daten

A.6.2.2. Arithmetisches Mittel

		N	Mittelwert	Standardabweichung	Standardfehler	95%-Konfidenzintervall für den Mittelwert		Minimum	Maximum
						Untergrenze	Obergrenze		
Arithmetisches Mittel Zeit	BUILD-IT	20	22.075	6.7184	1.5023	18.931	25.219	13.6	37.4
	MODELER	20	31.019	13.9842	3.1270	24.474	37.564	12.8	66.0
	PHYSICAL	20	14.175	5.9097	1.3215	11.409	16.941	5.4	28.2
	Gesamt	60	22.423	11.7008	1.5106	19.400	25.446	5.4	66.0
Arithmetisches Mittel Anzahl Blöcke	BUILD-IT	20	5.200	1.3353	.2986	4.575	5.825	3.1	8.7
	MODELER	20	3.993	.9847	.2202	3.532	4.454	2.3	5.3
	PHYSICAL	20	3.065	.8750	.1956	2.656	3.474	1.8	5.1
	Gesamt	60	4.086	1.3820	.1784	3.729	4.443	1.8	8.7

Tabelle 33: SPSS Ausgabe; Standardabweichungen; korrekte Daten, arithmetisches Mittel

A.6.2.3. Transformierte Daten

		N	Mittelwert	Standardabweichung	Standardfehler	95%-Konfidenzintervall für den Mittelwert		Minimum	Maximum
						Untergrenze	Obergrenze		
sqrt Zeit	BUILD-IT	20	4.6510	.68281	.15268	4.3315	4.9706	3.69	6.12
	MODELER	20	5.4557	1.14927	.25699	4.9178	5.9935	3.58	8.12
	PHYSICAL	20	3.6880	.77726	.17380	3.3242	4.0517	2.32	5.31
	Gesamt	60	4.5982	1.14065	.14726	4.3036	4.8929	2.32	8.12
sqrt Anzahl Blöcke	BUILD-IT	20	2.2632	.28682	.06413	2.1289	2.3974	1.76	2.95
	MODELER	20	1.9827	.25554	.05714	1.8631	2.1023	1.52	2.30
	PHYSICAL	20	1.7341	.24665	.05515	1.6187	1.8496	1.34	2.26
	Gesamt	60	1.9933	.33853	.04370	1.9059	2.0808	1.34	2.95
ln Zeit	BUILD-IT	20	3.0547	.28273	.06322	2.9224	3.1871	2.61	3.62
	MODELER	20	3.3545	.39752	.08889	3.1685	3.5406	2.55	4.19
	PHYSICAL	20	2.5668	.43165	.09652	2.3648	2.7688	1.69	3.34
	Gesamt	60	2.9920	.49375	.06374	2.8645	3.1196	1.69	4.19
ln Anzahl Blöcke	BUILD-IT	20	1.6184	.25164	.05627	1.5007	1.7362	1.13	2.16
	MODELER	20	1.3522	.26857	.06005	1.2265	1.4779	.83	1.67
	PHYSICAL	20	1.0819	.28361	.06342	.9492	1.2146	.59	1.63
	Gesamt	60	1.3509	.34397	.04441	1.2620	1.4397	.59	2.16
log2 Zeit	BUILD-IT	20	4.4068	.40761	.09114	4.2160	4.5976	3.77	5.22
	MODELER	20	4.8289	.57507	.12859	4.5597	5.0980	3.68	6.04
	PHYSICAL	20	4.0397	1.14061	.25505	3.5059	4.5735	2.43	7.35
	Gesamt	60	4.4251	.82746	.10682	4.2114	4.6389	2.43	7.35
log2 Anzahl Blöcke	BUILD-IT	20	2.3349	.36305	.08118	2.1650	2.5048	1.63	3.12
	MODELER	20	1.9508	.38747	.08664	1.7695	2.1322	1.20	2.41
	PHYSICAL	20	1.5608	.40916	.09149	1.3694	1.7523	.85	2.35
	Gesamt	60	1.9489	.49625	.06407	1.8207	2.0771	.85	3.12

Tabelle 34: SPSS Ausgabe; Standardabweichungen; korrekte transformierte Daten

	Levene-Statistik	df1	df2	Signifikanz
sqrt Anzahl Blöcke	.165	2	57	.849
sqrt Zeit	1.364	2	57	.264
ln Anzahl Blöcke	.303	2	57	.740
ln Zeit	.621	2	57	.541
log2 Anzahl Blöcke	.303	2	57	.740
log2 Zeit	5.120	2	57	.009

Tabelle 35: SPSS Ausgabe; Levene-Test

		sqrt Anzahl Blöcke	sqrt Zeit	In Anzahl Blöcke	In Zeit	log2 Anzahl Blöcke	log2 Zeit
N		20	20	20	20	20	20
Parameter der Normalverteilung(a,b)	Mittelwert	2.2632	4.6510	1.6184	3.0547	2.3349	4.4068
	Standardabweichung	.28682	.68281	.25164	.28273	.36305	.40761
Extremste Differenzen	Absolut	.106	.196	.090	.168	.090	.167
	Positiv	.106	.196	.090	.168	.090	.167
	Negativ	-.057	-.111	-.078	-.099	-.078	-.099
Kolmogorov-Smirnov-Z		.476	.877	.401	.750	.401	.748
Asymptotische Signifikanz (2-seitig)		.977	.425	.997	.628	.997	.630

Tabelle 36: SPSS Ausgabe; Kolmogorov-Smirnov-Test für BUILD-IT

- a Die zu testende Verteilung ist eine Normalverteilung.
b Aus den Daten berechnet.
c Tool = BUILD-IT

		sqrt Anzahl Blöcke	sqrt Zeit	In Anzahl Blöcke	In Zeit	log2 Anzahl Blöcke	log2 Zeit
N		20	20	20	20	20	20
Parameter der Normalverteilung(a,b)	Mittelwert	1.9827	5.4557	1.3522	3.3545	1.9508	4.8289
	Standardabweichung	.25554	1.14927	.26857	.39752	.38747	.57507
Extremste Differenzen	Absolut	.166	.208	.187	.170	.187	.163
	Positiv	.106	.208	.120	.170	.120	.163
	Negativ	-.166	-.115	-.187	-.096	-.187	-.103
Kolmogorov-Smirnov-Z		.740	.930	.834	.758	.834	.731
Asymptotische Signifikanz (2-seitig)		.643	.352	.489	.613	.489	.660

Tabelle 37: SPSS Ausgabe; Kolmogorov-Smirnov-Test für Modeler

- a Die zu testende Verteilung ist eine Normalverteilung.
b Aus den Daten berechnet.
c Tool = MODELER

		sqrt Anzahl Blöcke	sqrt Zeit	In Anzahl Blöcke	In Zeit	log2 Anzahl Blöcke	log2 Zeit
N		20	20	20	20	20	20
Parameter der Normalverteilung(a,b)	Mittelwert	1.7341	3.6880	1.0819	2.5668	1.5608	4.0397
	Standardabweichung	.24665	.77726	.28361	.43165	.40916	1.14061
Extremste Differenzen	Absolut	.111	.179	.110	.175	.110	.184
	Positiv	.111	.179	.110	.140	.110	.184
	Negativ	-.079	-.137	-.088	-.175	-.088	-.101
Kolmogorov-Smirnov-Z		.497	.802	.492	.781	.492	.824
Asymptotische Signifikanz (2-seitig)		.966	.541	.969	.575	.969	.506

Tabelle 38: SPSS Ausgabe; Kolmogorov-Smirnov-Test für Physical

a Die zu testende Verteilung ist eine Normalverteilung.

b Aus den Daten berechnet.

c Tool = PHYSICAL

A.6.3. Resultate

A.6.3.1. Zeit & Anzahl Blöcke

		Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
In Zeit	Zwischen den Gruppen	6.322	2	3.161	22.352	.000
	Innerhalb der Gruppen	8.061	57	.141		
	Gesamt	14.384	59			
sqrt Anzahl Blöcke	Zwischen den Gruppen	2.802	2	1.401	20.168	.000
	Innerhalb der Gruppen	3.960	57	.069		
	Gesamt	6.762	59			

Tabelle 39: SPSS Ausgabe; ANOVA

Abhängige Variable	(I) Tool	(J) Tool	Mittlere Differenz (I-J)	Standardfehler	Signifikanz	95%-Konfidenzintervall	
						Untergrenze	Obergrenze
ln Zeit	BUILD-IT	MODELER	-.29979(*)	.11892	.038	-.5860	-.0136
		PHYSICAL	.48790(*)	.11892	.000	.2017	.7741
	MODELER	BUILD-IT	.29979(*)	.11892	.038	.0136	.5860
		PHYSICAL	.78769(*)	.11892	.000	.5015	1.0739
	PHYSICAL	BUILD-IT	-.48790(*)	.11892	.000	-.7741	-.2017
		MODELER	-.78769(*)	.11892	.000	-1.0739	-.5015
sqrt Anzahl Blöcke	BUILD-IT	MODELER	.28048(*)	.08335	.004	.0799	.4810
		PHYSICAL	.52902(*)	.08335	.000	.3285	.7296
	MODELER	BUILD-IT	-.28048(*)	.08335	.004	-.4810	-.0799
		PHYSICAL	.24854(*)	.08335	.012	.0480	.4491
	PHYSICAL	BUILD-IT	-.52902(*)	.08335	.000	-.7296	-.3285
		MODELER	-.24854(*)	.08335	.012	-.4491	-.0480

Tabelle 40: SPSS Ausgabe; Tukey HSD Post-Hoc-Test zur ANOVA

* Die mittlere Differenz ist auf der Stufe .05 signifikant.

A.6.3.2. Benutzerzufriedenheit

	N	Mittelwert	Standardabweichung	Standardfehler	95%-Konfidenzintervall für den Mittelwert		Minimum	Maximum
					Untergrenze	Obergrenze		
BUILD-IT	20	.917	.4572	.1022	.703	1.131	.0	1.7
MODELER	20	.600	.4537	.1015	.388	.812	-.7	1.3
PHYSICAL	20	.983	.4648	.1039	.766	1.201	.0	1.7
Gesamt	60	.833	.4813	.0621	.709	.958	-.7	1.7

Tabelle 41: SPSS Ausgabe; Benutzerzufriedenheit; Arithmetisches Mittel

Levene-Statistik	df1	df2	Signifikanz
.149	2	57	.861

Tabelle 42: SPSS Ausgabe; Benutzerzufriedenheit; Levene-Test

		Benutzer- zufriedenheit
N		20
Parameter der Normalverteilung(a,b)	Mittelwert	.917
	Standardabweichung	.4572
Extremste Differenzen	Absolut	.222
	Positiv	.178
	Negativ	-.222
Kolmogorov-Smirnov-Z		.994
Asymptotische Signifikanz (2-seitig)		.276

Tabelle 43: SPSS Ausgabe; Benutzerzufriedenheit; Kolmogorov-Smirnov-Test für BUILD-IT

- a Die zu testende Verteilung ist eine Normalverteilung.
b Aus den Daten berechnet.
c Tool = BUILD-IT

		Benutzer- zufriedenheit
N		20
Parameter der Normalverteilung(a,b)	Mittelwert	.600
	Standardabweichung	.4537
Extremste Differenzen	Absolut	.178
	Positiv	.139
	Negativ	-.178
Kolmogorov-Smirnov-Z		.798
Asymptotische Signifikanz (2-seitig)		.548

Tabelle 44: SPSS Ausgabe; Benutzerzufriedenheit; Kolmogorov-Smirnov-Test für Modeler

- a Die zu testende Verteilung ist eine Normalverteilung.
b Aus den Daten berechnet.
c Tool = MODELER

		Benutzer- zufriedenheit
N		20
Parameter der Normalverteilung(a,b)	Mittelwert	.983
	Standardabweichung	.4648
Extremste Differenzen	Absolut	.202
	Positiv	.202
	Negativ	-.174
Kolmogorov-Smirnov-Z		.904
Asymptotische Signifikanz (2-seitig)		.387

Tabelle 45: SPSS Ausgabe; Benutzerzufriedenheit; Kolmogorov-Smirnov-Test für Physical

- a Die zu testende Verteilung ist eine Normalverteilung.
- b Aus den Daten berechnet.
- c Tool = PHYSICAL

	Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
Zwischen den Gruppen	1.678	2	.839	3.988	.024
Innerhalb der Gruppen	11.989	57	.210		
Gesamt	13.667	59			

Tabelle 46: SPSS Ausgabe; Benutzerzufriedenheit; ANOVA

(I) Tool	(J) Tool	Mittlere Differenz (I-J)	Standardfehler	Signifikanz	95%-Konfidenzintervall	
					Untergrenze	Obergrenze
BUILD-IT	MODELER	.3167	.1450	.083	-.032	.666
	PHYSICAL	-.0667	.1450	.890	-.416	.282
MODELER	BUILD-IT	-.3167	.1450	.083	-.666	.032
	PHYSICAL	-.3833(*)	.1450	.028	-.732	-.034
PHYSICAL	BUILD-IT	.0667	.1450	.890	-.282	.416
	MODELER	.3833(*)	.1450	.028	.034	.732

Tabelle 47: SPSS Ausgabe; Benutzerzufriedenheit; Tukey HSD Post-Hoc-Test für ANOVA

- * Die mittlere Differenz ist auf der Stufe .05 signifikant.

A.6.3.3. Versuchserfolg

	N	Mittelwert	Standardabweichung	Standardfehler	95%-Konfidenzintervall für den Mittelwert		Minimum	Maximum
					Untergrenze	Obergrenze		
BUILD-IT	200	8.04	.184	.013	8.01	8.06	8	9
MODELER	200	8.21	.408	.029	8.15	8.27	8	9
PHYSICAL	200	8.00	.000	.000	8.00	8.00	8	8
Gesamt	600	8.08	.274	.011	8.06	8.10	8	9

Tabelle 48: SPSS Ausgabe; Versuchserfolg; Standardabweichung

Levene-Statistik	df1	df2	Signifikanz
215.900	2	597	.000

Tabelle 49: SPSS Ausgabe; Versuchserfolg; Levene-Test

		Versuchserfolg
N		200
Parameter der Normalverteilung(a,b)	Mittelwert	8.04
	Standardabweichung	.184
Extremste Differenzen	Absolut	.540
	Positiv	.540
	Negativ	-.425
Kolmogorov-Smirnov-Z		7.641
Asymptotische Signifikanz (2-seitig)		.000

Tabelle 50: SPSS Ausgabe; Versuchserfolg; Kolmogorov-Smirnov-Test für BUILD-IT

- a Die zu testende Verteilung ist eine Normalverteilung.
 b Aus den Daten berechnet.
 c Tool = BUILD-IT

		Versuchs- erfolg
N		200
Parameter der Normalverteilung(a,b)	Mittelwert	8.21
	Standardabweichung	.408
Extremste Differenzen	Absolut	.486
	Positiv	.486
	Negativ	-.304
Kolmogorov-Smirnov-Z		6.880
Asymptotische Signifikanz (2-seitig)		.000

Tabelle 51: SPSS Ausgabe; Versuchserfolg; Kolmogorov-Smirnov-Test für Modeler

- a Die zu testende Verteilung ist eine Normalverteilung.
 b Aus den Daten berechnet.
 c Tool = MODELER

		Versuchs- erfolg
N		200
Parameter der Normalverteilung(a, b)	Mittelwert	8.00
	Standardabweichung	.000(c)

Tabelle 52: SPSS Ausgabe; Versuchserfolg; Kolmogorov-Smirnov-Test für Physical

- a Die zu testende Verteilung ist eine Normalverteilung.
 b Aus den Daten berechnet.
 c Die Verteilung weist keine Varianz für diese Variable auf. Der Kolmogorov-Smirnov-Anpassungstest kann nicht ausgeführt werden.
 d Tool = PHYSICAL

	Quadratsu mme	df	Mittel der Quadrate	F	Signifikanz
Zwischen den Gruppen	5.063	2	2.532	37.847	.000
Innerhalb der Gruppen	39.935	597	.067		
Gesamt	44.998	599			

Tabelle 53: SPSS Ausgabe; Versuchserfolg; ANOVA

(I) Tool	(J) Tool	Mittlere Differenz (I-J)	Standardfehler	Signifikanz	95%-Konfidenzintervall	
					Untergrenze	Obergrenze
BUILD-IT	MODELER	-.175(*)	.026	.000	-.24	-.11
	PHYSICAL	.035	.026	.366	-.03	.10
MODELER	BUILD-IT	.175(*)	.026	.000	.11	.24
	PHYSICAL	.210(*)	.026	.000	.15	.27
PHYSICAL	BUILD-IT	-.035	.026	.366	-.10	.03
	MODELER	-.210(*)	.026	.000	-.27	-.15

Tabelle 54: SPSS Ausgabe; Versuchserfolg; Tukey HSD Post-Hoc-Test für ANOVA

* Die mittlere Differenz ist auf der Stufe .05 signifikant.