

Entwicklung eines Statistik-Szenarios aus der Chemietechnik innerhalb der Lehr- und Lernsoftware e-stat

Prof. Dr. C. Weihs, Dipl. Stat. M. Kappler*

Uni Dortmund, Fachbereich Statistik, Lehrstuhl Computergestützte Statistik

1 Einführung

EMILeA-stat (kurz e-stat) ist die Abkürzung für "Eine multimediale internetbasierte und interaktive Lehr- und Lernumgebung in der angewandten Statistik".

Damit sind die Hauptschwerpunkte und -zielsetzungen von EMILeA-stat schon vorgegeben. In dem Projekt soll eine internetbasierte Lehr- und Lernsoftware erstellt werden, die sich zu einem breiten Informationsforum in der Statistik für zahlreiche Gruppen entwickeln soll und als effizientes Werkzeug in Aus- und Weiterbildung eingesetzt werden soll.

Der Name e-stat wird in diesem Bericht synonym zu EMILeA-stat verwendet, da e-stat der eigentlich angedachte Name für die Software war, jedoch nicht eingetragen werden kann. Sowohl intern als auch extern werden beide Begriffe nebeneinander verwendet.

Der Bericht ist in zwei Teile geteilt. Im ersten wird die Lehr- und Lernsoftware EMILeA-stat sowie deren Ziele, Funktionalitäten und Kernmerkmale geschildert. Im zweiten Teil wird das von den Autoren entwickelte Statistik-Szenario aus der Chemietechnik, das in EMILeA-stat eingebunden wird, erklärt und dessen Aufbau näher dargestellt.

2 EMILeA-stat

Das Projekt zur Entwicklung von EMILeA-stat wird vom Bundesministerium für Bildung und Forschung (bmb+f) im Rahmen des Programms "Neue Medien in der Bildung (Förderbereich Hochschulen)" mit einem Volumen von insgesamt 1,5 Mio. € gefördert und hat eine Laufzeit von April 2001 bis März 2004. Das Projekt wird geleitet von Prof. Dr. Uwe Kamps, Uni Oldenburg.

Um möglichst die gesamte Bandbreite der Statistik abzudecken und den Anforderungen der zukünftigen Nutzer gerecht zu werden, sind mehrere Universitäten mit verschiedenen Fachbereichen und Fachrichtungen an dem Projekt beteiligt. Antragsteller sind:

- Prof. Dr. Joachim Buhmann, Uni Bonn, Inst. für Informatik
- Dr. Erhard Cramer, Uni Oldenburg, FB Mathematik
- Prof. Dr. Ursula Gather, Uni Dortmund, FB Statistik
- Prof. Dr. Heinz Lothar Grob, Uni Münster, Inst. für Wirtschaftsinformatik
- Prof. Dr. Wolfgang Härdle, Humboldt-Uni zu Berlin, wirtschaftswiss. Fakultät
- Prof. Dr. Udo Kamps (Projektleiter), Uni Oldenburg, FB Mathematik
- Prof. Dr. Claus Möbus, Uni Oldenburg, FB Informatik
- Prof. Dr. Dietmar Pfeifer, Uni Oldenburg, FB Mathematik
- Prof. Dr. Kristina Reiss, Uni Oldenburg, FB Mathematik
- Prof. Dr. Bernd Rönz, Humboldt-Uni zu Berlin, wirtschaftswiss. Fakultät
- Prof. Dr. Hans Gerhard Strohe, Uni Potsdam, wirtschafts- und sozialwiss. Fakultät
- Prof. Dr. Karl-Heinz Waldmann, Uni Karlsruhe, Inst. für Wirtschaftstheorie und Operations Research
- Prof. Dr. Claus Weihs, Uni Dortmund, FB Statistik

Weitere Partner sind in Beratung und Entwicklung der zu bietenden Funktionalitäten sowie für die Evaluation der entstehenden Software im Projekt beteiligt. Schon in der Entwicklungsphase soll e-stat in Kursen und als Vorlesungen eingesetzt werden, in denen die Lernenden am Ende ein Feedback zu Bedienfreundlichkeit, Design etc. geben können.

Zu den folgenden Abschnitten siehe auch <http://www.emilea.de/>.

2.1 Zielgruppen und Zielsetzung

Als Zielgruppen für e-stat sind alle Gruppen zu nennen, die in irgendeiner Weise Statistik lernen, lehren oder statistische Begriffe verstehen und in ihren Kontext einordnen wollen. Dabei sind an erster Stelle Schulen zu nennen, an denen häufig eine Grundausbildung in Statistik stattfindet. Hier soll e-stat eingesetzt werden, um ein tieferes Verständnis für die Statistik zu wecken und einen fundierten Einstieg in die Statistik zu geben.

Ebenso ist Statistik ein wichtiger Bestandteil in den unterschiedlichsten Ausbildungen und Studiengängen. Als erstes ist hier das Hauptfach Statistik zu nennen. Aber vor allem in den Studiengängen, die im Studienverlaufsplan einen Teil Statistik vorgesehen haben, dem von den Studierenden oft starke Abneigung entgegengebracht wird, ist e-stat sinnvoll einzusetzen. Als Beispiel seien hier nur wirtschaftswissenschaftliche Studiengänge oder Medizin genannt.

Auch in vielen berufsbegleitenden Fortbildungsmaßnahmen findet eine statistische Ausbildung statt. In all diesen Bereichen bietet e-stat eine Möglichkeit, das Lernen und Lehren zu ergänzen und durch Selbststudium zu erweitern.

Weiterhin stellt e-stat auch eine Art Lexikon dar, in der sowohl Vorgebildete als auch interessierte Laien schnell und effizient einzelne aufgeschnappte statistische Begriffe nachschlagen können. Als Beispiel sei hier ein Journalist genannt, der sich über den Begriff Signifikanzniveau informieren will.

Auf diese Weise können mit einem einzigen System unterschiedlichste Gruppen erreicht werden. e-stat soll sich zu einem breiten und alltagstauglichen Informationsforum innerhalb der Statistikaus- und -weiterbildung entwickeln.

2.2 Funktionalität

e-stat ist als multimediale und internetbasierte Lernsoftware konzipiert und bietet dadurch vielfältige Möglichkeiten, Lernen eigenverantwortlich zu gestalten.

Lernen ist hier unabhängig von Zeit und Ort möglich. Von jedem internetfähigen Rechner ist der Zugriff auf die Software möglich, und der Lernende kann dort ansetzen, wo er seine Lektion unterbrochen hat. Ebenso ist eine individuelle Zeitplanung des Lernstoffes möglich. Tempo und Umfang des Lernens können an die individuellen Bedürfnisse angepasst werden.

Zusätzlich zu diesen in der Anlegung von e-stat als internetbasierte Lernsoftware liegenden Eigenschaften sind einige weitere Punkte geplant, die sich am Lernenden orientieren und die es erlauben, auf seine individuellen Bedürfnisse einzugehen. Drei dieser Konzepte sollen kurz genannt werden.

2.2.1 Kurse

Ein Nachteil der Möglichkeit, das Lerntempo und Lernumfang selbst zu bestimmen, liegt darin, dass dies in hohem Maße selbstgesteuertes und eigenverantwortliches Lernen erfordert, was bei vielen Lernenden, insbesondere wenn die Statistik nicht der Hauptzweck des Studiums ist, oft nicht vorausgesetzt werden kann. Dem kann insofern teilweise entgegen gewirkt werden, indem Kurse mit klar abgegrenztem Lernziel zur Verfügung gestellt werden, die der Lernende durchlaufen kann und nach denen er sein Wissen anhand von Tests überprüfen kann. Des weiteren sind zahlreiche interaktive Java-Applets vorgesehen, mit deren Hilfe statistische Sachverhalte "spielerisch" erfahren werden können, z.B. die Auswirkung von einzelnen Ausreißern etwa auf den Mittelwert oder auf Regressionsgeraden.

Diese Kurse innerhalb von e-stat können auch als Begleitung von Vorlesungen oder Unterricht fungieren. Lehrende erhalten die Möglichkeit, Kurse neu zusammenzustellen, um so

den Schülern Begleitmaterial für den Unterricht zu liefern, anhand dessen der Unterricht vor- und nachbereitet sowie auch erweitert werden kann.

2.2.2 Szenarien

Szenarien in EMILeA-stat sind eigenständige Umgebungen, in denen der Lernende sein erlerntes Wissen "spielerisch" in realistischen Situationen anwenden kann und sich neue Verfahren aneignen soll, um das Ziel des Szenarios zu erreichen. EMILeA-stat dient hierbei lediglich als Hilfesystem, um Probleme, die im Szenario auftauchen, zu lösen, und um die statistischen Verfahren nachschlagen zu können, die Verwendung finden sollen.

2.2.3 Level

Um den Erfahrungshintergrund und das statistische Vorwissen der Benutzer zu berücksichtigen, wurden verschiedene Abstraktionsebenen geschaffen.

Die unterste Ebene, Level A, orientiert sich an Benutzern, die über kein oder nur sehr wenig mathematisches und statistisches Hintergrundwissen verfügen. Hier werden die Sachverhalte überwiegend sehr ausführlich und mit viel Text und, wo dies möglich ist, mit Verzicht auf mathematische Formeln erklärt. Diese Ebene eignet sich z.B. für Schulgruppen, die einen ersten Einstieg in Statistik erhalten wollen, oder den o.g. Journalisten.

Die mittlere Ebene, Level B, wendet sich an Benutzer, die ein gewisses mathematisches Verständnis mitbringen und auch teilweise mit mathematischer Symbolik vertraut sind. Zielgruppe für diese Ebene sind im Wesentlichen Anwender aus Fachwissenschaften, in denen Statistik angewandt wird.

Die dritte Ebene, Level C, ist die mathematisch statistisch formale Ebene, auf der die Methoden mathematisch korrekt eingeführt und erklärt werden. Hier findet sich am ehesten eine Struktur der Art "Definition-Satz-Beweis" wieder. Als Zielgruppe auf dieser Ebene sind folglich Mathematiker und Statistiker gedacht.

Ziel von e-stat ist es, mit den gebotenen Funktionalitäten und den sich daraus ergebenden Vorteilen einen innovativen Beitrag zur Qualitätsverbesserung in der praxisnahen Aus- und Weiterbildung zu leisten.

2.3 Kernmerkmale

e-stat ist als benutzerorientiertes Produkt konzipiert. Dazu soll die Software schon während der Entwicklungsphase eingesetzt und mit den Nutzern Rücksprache bzgl. Design, Bedienfreundlichkeit etc. gehalten werden. Auch durch die verschiedenen Konzepte wie die oben beschriebenen Kurse und Abstraktionsebenen steht der Nutzer im Mittelpunkt der Entwicklung.

Nach Ende der Entwicklungszeit ist EMILeA-stat nicht als fertiges Produkt anzusehen. Durch eine offene Architektur ist von Anfang an vorgesehen, dass interessierte Anwender die Software ausbauen und eigene statistische Inhalte und Szenarien einbringen können. Ebenso besteht die Möglichkeit die bereits vorhandenen statistischen Inhalte durch eine neue Verknüpfung zu neuen Kursen zusammen zu fassen. Hier soll nach dem Baukastenprinzip aus allen Bereichen gewählt werden können (Inhalte, Beispiele, Java-Applets etc.). Diese Erweiterungsmöglichkeiten werden dabei nicht nur lediglich bereitgestellt, es ist sogar ausdrücklich erwünscht sie einzusetzen, damit sich e-stat zu einem lebendigen und breit angelegten Forum entwickeln kann.

Zentrale Ziele sind weiterhin die effiziente Einsetzbarkeit der Software in allen geplanten Bereichen sowohl als reines Nachschlagewerk als auch zur Unterrichtsbegleitung sowie Vor- und Nachbereitung.

Ein weiterer wichtiger Punkt von EMILeA-stat ist seine Open-Source-Konzeption. Ausbildung und Schulung soll auch nach Ende der Entwicklungsphase dauerhaft kostenlos und frei zugänglich sein.

2.4 Themenfelder

Da bei der breit angelegten Konzeption von EMILeA-stat hier nicht alle Themenfelder genannt werden können, seien hier lediglich die Hauptschwerpunkte aufgeführt.

Zum einen sind dies die klassischen Bereiche der Statistik wie z.B. deskriptive und induktive Statistik, grafische Methoden, Wahrscheinlichkeitsrechnung, Versuchsplanung etc. Zum anderen werden speziellere Verfahren der Statistik vorgestellt, die Anwendung in den verschiedenen Fachgebieten finden, wie z.B. die Finanz- und Versicherungsmathematik oder Statistik in den Ingenieurwissenschaften.

Ein weiteres wichtiges Themenfeld stellen die Szenarien dar; Umgebungen, in denen der Lernende in einem realistischen Umfeld sein erworbenes Wissen erproben kann. Beispiele für solche Umgebungen sind ein Börsenspiel, eine virtuelle Unternehmung und das im folgenden vorgestellte Produktionsszenario aus der Chemietechnik.

3 Statistik-Szenario aus der Chemietechnik

3.1 Ziele

Das Lernen durch eigene Aktionen stellt für Moore (1997) das neue Modell der Pädagogik dar. Insbesondere die Kombination von Inhalt, Lernen in realen Datensituationen und Einsatz von Multimedia mit zahlreichen Interaktionsmöglichkeiten ermöglicht nach Moore ein effektives und modernes Lernen. Auch die Möglichkeit das Tempo und die Ausführlichkeit der einzelnen Schritte selbst zu wählen, ermöglichen es, den Lernprozess individuell auf den Benutzer abzustimmen. All diese Konzepte werden in dem hier vorgestellten Szenario der Lernsoftware EMILeA-stat verwirklicht.

In diesem Szenario wird dem Benutzer erlaubt, die im Rahmen eines Kurses in EMILeA-stat erlernten Fähigkeiten in einer realistischen Umgebung anzuwenden. Ebenso soll es möglich sein, ohne große Vorkenntnisse in das Szenario einzusteigen und sich anhand der Fragestellungen im Szenario neue Kenntnisse anzueignen. Hierzu wird der Benutzer in eine Chemietechnik-Umgebung versetzt und soll die Produktion von Styrol auswerten und optimieren. Der Schwerpunkt wird hierbei auf die explorative Datenanalyse und Versuchsplanung gelegt.

Der Benutzer soll anhand eines intuitiven Aufbaus des Szenarios Lernschritte selbst bewältigen, Strukturen in den Daten erkennen, über diese zu den Hauptfragestellungen gelangen und eine Versuchsplanung durchführen, die die Ausbeute der Zielgröße optimiert.

Anhand des Styrol-Prozesses soll dem Lernenden in EMILeA-stat die Möglichkeit gegeben werden, einerseits sich mit der Datenanalyse mit realen Daten vertraut zu machen und andererseits in einem realistischen Umfeld an Sinn/Zweck und Problematik insbesondere der Versuchsplanung herangeführt zu werden.

Zu den Abschnitten 3.2 bis 3.6 vergleiche Weihs und Jessenberger (1999).

3.2 Styrol-Prozess

Styrol ist eines der wichtigsten Zwischenprodukte der chemischen Industrie. Es dient u.a. als Grundlage zur Herstellung von Styropor[®], Plastiktüten und vielen technischen Kunststoffen. Styrol wird dabei in einer katalytisch unterstützten Reaktion in einem adiabatischen Rohrreaktor (siehe Abb. 1) aus Ethylbenzol und Wasserdampf gewonnen.

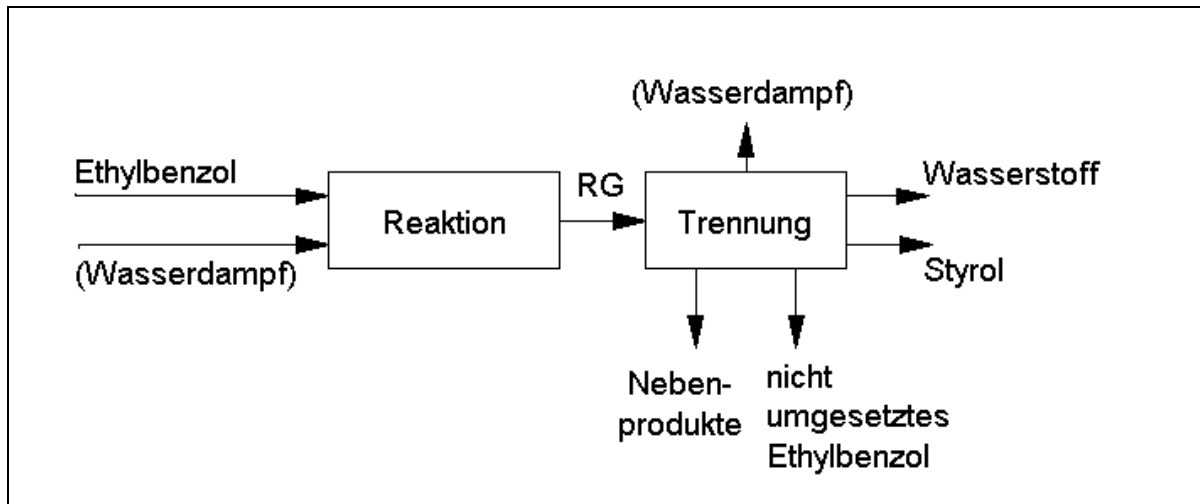


Abbildung 1: Schematische Darstellung des Reaktors

3.3 Variablen

In dem skizzierten Ablauf existieren zahlreiche Einflussgrößen auf die Styrol-Ausbeute. Dies sind zum einen Variablen, auf die direkt Einfluss genommen werden kann, so genannte Einstellgrößen, und Variablen, die zwar nicht konstant sind, die aber lediglich beobachtet werden können, so genannte Messgrößen.

3.3.1 Einflussgrößen

- Versuchsnummer
- Katalysator-Lieferant
- Arbeiterschicht
- Durchmesser Katalysatorschüttung
- Dichte Katalysatorschüttung
- Umgebungstemperatur
- Reaktordruck
- Reaktortemperatur
- Temperatur Ethylbenzol ($t=0$)
- Temperatur Wasserdampf ($t=0$)
- Massenstrom Ethylbenzol ($t=0$)
- Massenstrom Wasserdampf ($t=0$)

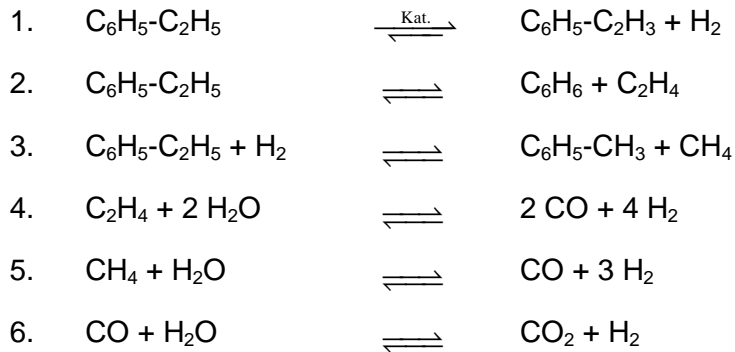
3.3.2 Messgrößen

- Temperatur Ethylbenzol
- Temperatur Wasserdampf
- Massenstrom Ethylbenzol
- Massenstrom Wasserdampf
- Styrol-Ausbeute
- 10 Nebenprodukte

All diese Variablen können einen Einfluss auf die Styrol-Ausbeute besitzen, müssen dies aber nicht. Im Verlauf des Szenarios soll nun der Anwender diejenigen Variablen identifizieren, die wirklich bedeutsam sind.

3.4 Reaktionen

Hier sind die im Reaktor ablaufenden Reaktionen im einzelnen dargestellt. Die erste Reaktionsgleichung stellt die Hauptreaktion dar. $C_6H_5-C_2H_5$ ist die Summenformel für Ethylbenzol und $C_6H_5-C_2H_3$ ist die Summenformel für Styrol. Diese Reaktion wird durch einen Katalysator unterstützt, wodurch sich die Gleichgewichtsreaktion auf die rechte Seite verschieben soll. Die restlichen Reaktionsgleichungen sind Nebenreaktionen, die zwar unerwünscht, aber im Reaktor nicht zu unterdrücken sind.



3.5 Daten

Aus den o.g. Reaktionsgleichungen kann über ein Modell zur Massenbilanzierung und Lösung eines Differentialgleichungssystems bei Kenntnis der Einstellgrößen die Styrol-Ausbeute theoretisch bestimmt werden. Damit ist die Styrol-Produktion für das Szenario simulierbar und es ist möglich, dass der Anwender seine eigenen Einstellungen vornimmt.

3.5.1 Qualitätsschwachstellenanalyse

Im ersten Schritt des Szenarios soll der Benutzer mit einem festen Datensatz von 100 Beobachtungen arbeiten, der vergangene Prozessläufe darstellen soll. Anhand dieser Daten soll er herausfinden, welche Variablen einen Haupteinfluss auf die Styrol-Ausbeute nehmen und in weiteren Untersuchungen optimal eingestellt werden sollen.

Da dieser Datensatz fest ist und von allen Anwendern bearbeitet werden soll, besteht die Möglichkeit, bestimmte Lernziele in den Daten zu verstecken, wie etwa die Behandlung von unvollständigen Daten, Eingabefehlern, Ausreißer etc.

3.5.2 Qualitätsoptimierung

Im Anschluss an die Qualitätsschwachstellenanalyse, in der die wichtigsten Einflussvariablen auf die Styrol-Ausbeute herausgefiltert wurden, soll anhand von Versuchsplanung die optimale Einstellung dieser Variablen bestimmt werden.

Versuchsplanung ist eine Methode, mit möglichst wenig Versuchen ein Set von Einflussvariablen so einzustellen, dass die Zielvariable optimal wird.

Mit den so gefundenen Einstellungen kann dann aus der Lösung des Differentialgleichungssystems ein Datensatz erzeugt werden, an Hand dessen überprüft werden kann, ob der Prozess verbessert wurde, d.h. eine höhere Styrol-Ausbeute ergibt als im ersten Datensatz.

3.6 Statistische Methoden

Die statistischen Methoden, die den Lernenden zur Verfügung stehen, sind durch die Lernziele in den beiden o.g. Teilen des Szenarios bestimmt. Im Teil der Qualitätsanalyse kommen Methoden der explorativen Datenanalyse zum Einsatz. Die Lernenden sollen die Variablen kennenlernen und mit der Datensituation vertraut werden. Einzelne Variablen können mit Mittelwert, Median, Varianz, Spannweite, Histogramm, Boxplot etc. charakterisiert werden.

In einem zweiten Schritt kann über Zusammenhangsanalysen die Abhängigkeitsstruktur im Datensatz betrachtet werden. Hierbei können Korrelationen, Streudiagrammmatrizen, Kontingenztafelanalyse, lineare Regression, Zeitreihenanalyse etc. Anwendung finden. Insbesondere der Einfluss einzelner Variablen auf die Styrol-Ausbeute ist hierbei für die Anwender von Interesse.

Nachdem die wichtigsten Einflussvariablen bestimmt sind, findet im dritten Schritt die statistische Versuchsplanung statt, in der die Einstellung der Variablen für eine optimale Styrol-Ausbeute bestimmt wird.

Weitere Schritte im Szenario, die allerdings aus Zeit- und Personalmangel derzeit nicht geplant sind, könnten in den Bereichen Qualitätsstabilisierung, -kontrolle und -charakterisierung erfolgen.

3.7 Aufbau und Ablauf

Beim Aufruf des Szenarios gelangt der Anwender auf eine Startseite (siehe Abb. 2), auf der er aufgefordert wird, sich anzumelden. Die Anmeldung ist notwendig, weil der Nutzer im weiteren Verlauf die Möglichkeit haben soll, seine bisherigen Aktivitäten und Notizen abzuspeichern. Des Weiteren kann auf diese Weise auf den individuellen Kenntnisstand des Benutzers mit entsprechenden Hilfestellungen oder Aufgaben eingegangen werden. Ein Gastzugang wird ebenfalls möglich sein.

Nach der Anmeldung gelangt der Anwender auf die Szenario-Hauptseite. Von hier aus kann er sich zunächst über den Aufbau des Szenarios, seine Ziele und die Bedienung in einer Führung informieren. Hierbei sind Hintergrundinformationen zum Prozess, Reaktionsgleichungen und Einflussgrößen erhältlich.

Ebenfalls von der Hauptseite kann der Anwender auf die in früheren Sitzungen abgespeicherten Auswertungen, Grafiken und Notizen zugreifen. Des Weiteren besteht jederzeit die Möglichkeit, zu e-stat zurückzukehren, um dort statistische Methoden zu lernen oder nachzuschlagen.

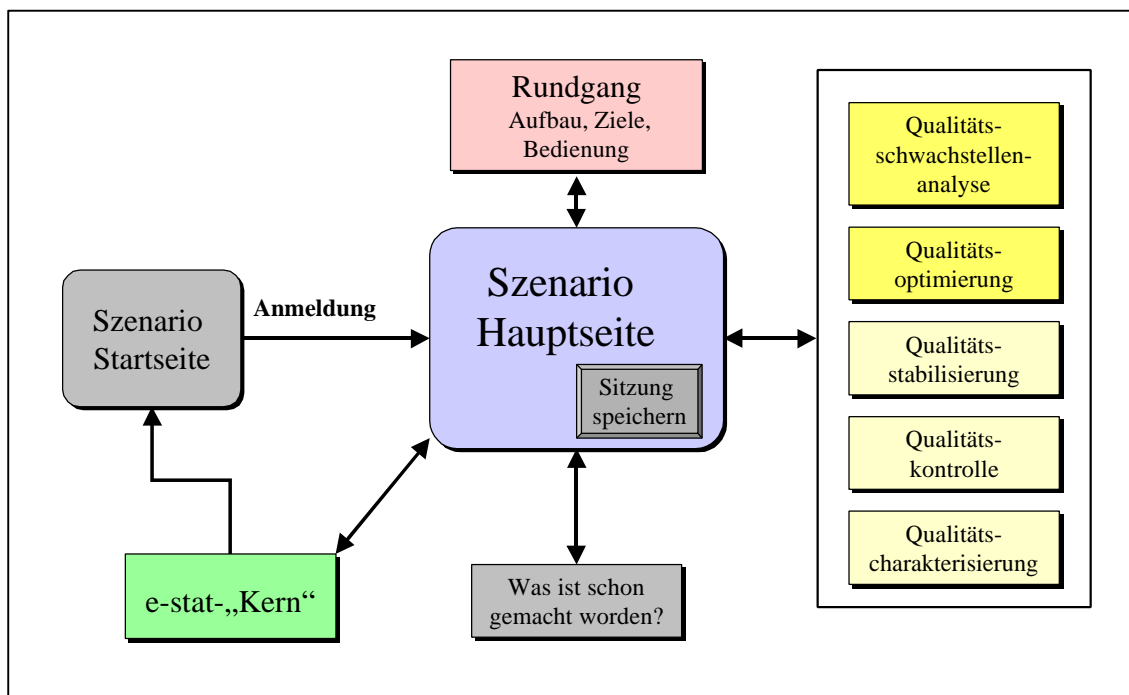


Abbildung 2: Aufbau des Szenarios, Teil 1

Die zu durchlaufenden statistischen Kapitel sind ebenfalls von der Hauptseite des Szenarios zugänglich. Dort erhält der Anwender Zugriff auf die Variablen sowie die statistischen Methoden. Er kann Auswertungen auswählen, die entsprechenden Auswertungsprogramme werden gestartet und liefern die Ergebnisse zurück, die betrachtet werden können (siehe Abb. 3).

Zur Seite gestellt wird dem Anwender dabei ein kontextsensitives Hilfesystem, das zum tieferen Verständnis der statistischen Methoden direkt mit den entsprechenden statistischen Inhalten in e-stat verknüpft ist. Von hier können direkt wieder neue Auswertungen ausgewählt und gestartet werden.

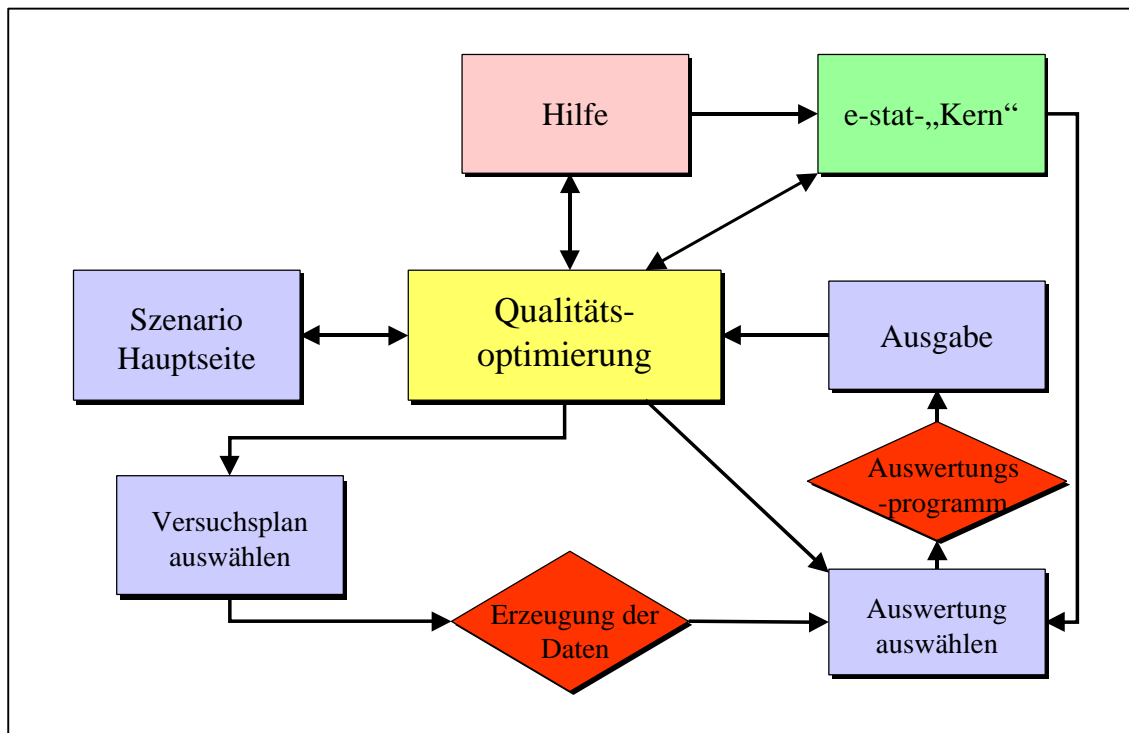


Abbildung 3: Aufbau des Szenarios, Teil 2

4 Kontakt

Fragen, Anregungen, Kritik oder andere Bemerkungen zu EMILeA-stat sind willkommen. Diese können an die Projektleitung in Oldenburg gerichtet werden.

Projektkoordinatorin
Frau Dipl.-Math. Katharina Cramer
Universität Oldenburg
Fachbereich Mathematik
26111 Oldenburg
Tel.: 0441/798-3216
e-stat@uni-oldenburg.de

Projektleiter
Prof. Dr. Udo Kamps
Universität Oldenburg
Fachbereich Mathematik
26111 Oldenburg
Tel.: 0441/798-3243
kamps@mathematik.uni-oldenburg.de

Bei Fragen zum Chemietechnik-Szenario wenden Sie sich bitte an:

Prof. Dr. Claus Weihs
Universität Dortmund
Fachbereich Statistik
44221 Dortmund
Tel.: 0231/755-4363
weihs@statistik.uni-dortmund.de

Dipl.Stat. Martin Kappler
Universität Dortmund
Fachbereich Statistik
44221 Dortmund
Tel.: 0231/755-5409
kappler@statistik.uni-dortmund.de

5 Literatur

D. S. Moore (1997): *New Pedagogy and New Content: The Case of Statistics*. Int. Stat. Rev., **65**, 2, 123-165

C. Weihs, J. Jessenberger (1999): *Statistische Methoden zur Qualitätssicherung und -optimierung*. Wiley-VCH, Weinheim

<http://www.emilea.de/>, Homepage von EMILeA-stat