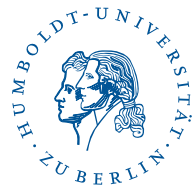
e-Learning Statistics – A Selective Review

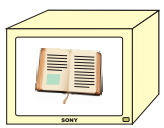
Wolfgang Härdle
Sigbert Klinke
Uwe Ziegenhagen

Institute for Statistics and Econometrics
Humboldt-Universität zu Berlin
http://ise.wiwi.hu-berlin.de



e-Learning

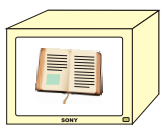
- web-based teaching materials and hypermedia
- authoring systems
- simulations and animations
- e-mail, blogs and wikis
- learning management/Learning content management systems
- videoconferencing and teleteaching



e-Learning

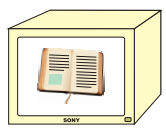
- *e*-learning is a big theme in modern education
- large-scale projects: NYU: 20'000'000 \$, eCornell: 40'000'000 \$, UK e-Universities: 62'000'000 £
- Institute for Statistics and Econometrics (ISE) involved in different smaller projects

"E-learning has not really revolutionized learning and teaching to date. Far-reaching, novel ways of teaching and learning, facilitated by ICT, remain nascent or still to be invented."
OECD (2005)



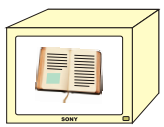
Outline

1. Motivation ✓
2. A typical education in statistics
3. *e*-Learning materials
   - MM*Stat
   - Electronic books
   - Q & A
   - e-stat
   - Moodle
4. Conclusion



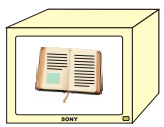
Course Structure – Bachelor (B.Sc.)

Group	Class	*e*-Media
Introduction	Statistics I & II	MM*Stat
		Q & A
Multivariate Statistics	Multivariate Statistics I	e-stat
	XploRe Introductory Course	XploRe
Applied Statistics	Computerbased Statistics I & II	e-Books
	Data Mining/Statistical Learning	Excel
	Numerical Introductory Course	R
Privatissimum	Privatissimum	
	Bachelor Thesis	



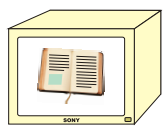
Course Structure – Master (M.Sc.)

Group	Class	*e*-Media
Multivariate Statistics	Multivariate Statistics I	e-stat
	XploRe Introductory Course	XploRe
Statistics of Fin. Markets	Statistics of Fin. Markets I & II	R
Advanced Statistics	Multivariate Statistics II	Matlab
	Non- and Semiparametrics I & II	
	Applied Quantitative Methods	
Privatissimum	Privatissimum	
	Master Thesis	



Course Structure - PhD

Group	Class	*e*-Media
Financial Statistics	Quantitative Finance Seminar	XploRe
	Adv. Stat. Methods in Finance	eBooks
	Mathematical Statistics Seminar	Matlab
	Statistical Tools in Finance and Insurance	



MM*Stat

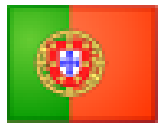
- HTML/JavaScript/CSS
- generated with LaTeX
- filing-card structure
- contents:
  - lectures, explanations
  - examples, multiple choice questions
  - interactive examples

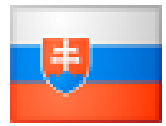


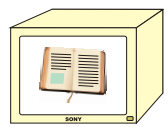










1.5 Measurement Scales

The values random variables take can differ distinctively:

Symbol	Variable	Sample space
X	Age (rounded to years)	$\{0, 1, 2, \ldots\}$
S	Sex	{female, male}
T	Marital status	{single,married, divorced}
Y	Monthly income	$[0, \infty)$

They can be classified into quantitative, i.e. numerically valued (age and income) and qualitative, i.e. categorical, (sex, marital status) variables. As numerical values are usually assigned to observations of qualitative variables, they may appear quantitative. Yet such synthetic assignments aren't of the same quality as numerical measurements that naturally arise in observing a phenomenon. The crucial distinction between quantitative and qualitative variables lies in the properties of the actual **scale of measurement**, which in turn is crucial to the applicability of statistical methods. In developing new tools statisticians make assumptions about permissible measurement scales.

A measurement is a numerical assignment to an observation. Some measurements appear more natural than others. By measuring the height of persons, for example, we apply a yardstick that ensures comparability between observations up to almost any desired precision--regardless of the units (such as inches or centimeters). School grades, on the other hand, represent a relatively rough classification indicating a certain ranking, yet putting many pupils into the same category. The values assigned to qualitative statements like `very good', `average' etc. are an arbitrary yet practical short cut in assessing people's achievements. As there is no conceptual reasoning behind a school grade scale, one should not try to interpret the 'distances' between grades.

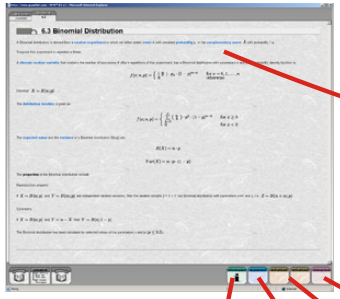
Clearly, height measurements convey more information than school marks, as distances between measurements can consistently be compared. Statements such as `Tom is twice as tall as his son' or `Manuela is 35 centimeters smaller than her partner' are permissible.

As statistical methods are developed in mathematical terms, the applicable scales are also defined in terms of mathematical concepts. These are the transformations that can be imposed on them without loss of information. The wider the range of permissible transformations, the less information the scale can convey. The following table lists common measurement scales in increasing order of information content. Scales carrying more information can always be transformed into less informative scales.

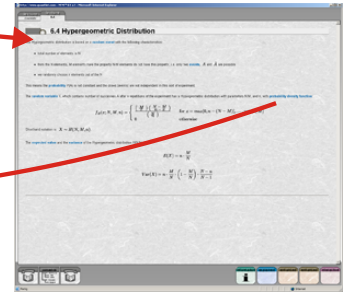
Variable	Measurement Scale	Statements	Permissible Transformations
Qualitative	**Nominal Scale**	equivalence	any equivalence preserving mapping

MM*Stat Linking of Topics

Lecture

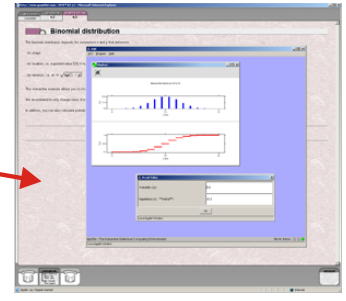


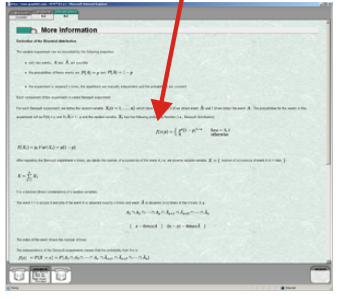
Lecture



Glossary

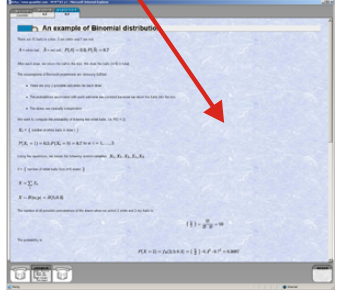
Interactive



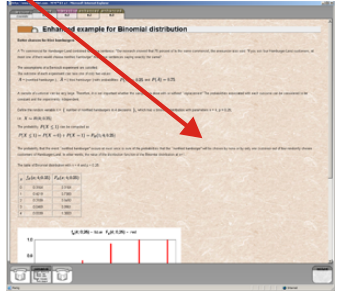


Derivations

Example

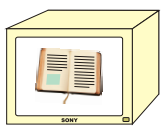


Enh. Example



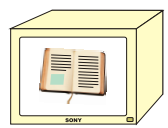
MM*Stat – Review

+ positive feedback from students, partly replaced traditional media such as books
+ simple access via webbrowser
- no significant improvement of grades
- complicated technical structure (browser war)
- focus on economics
- cultural differences and accessibility issues not considered
- sustainability



Electronic Books (MD*Book Framework)

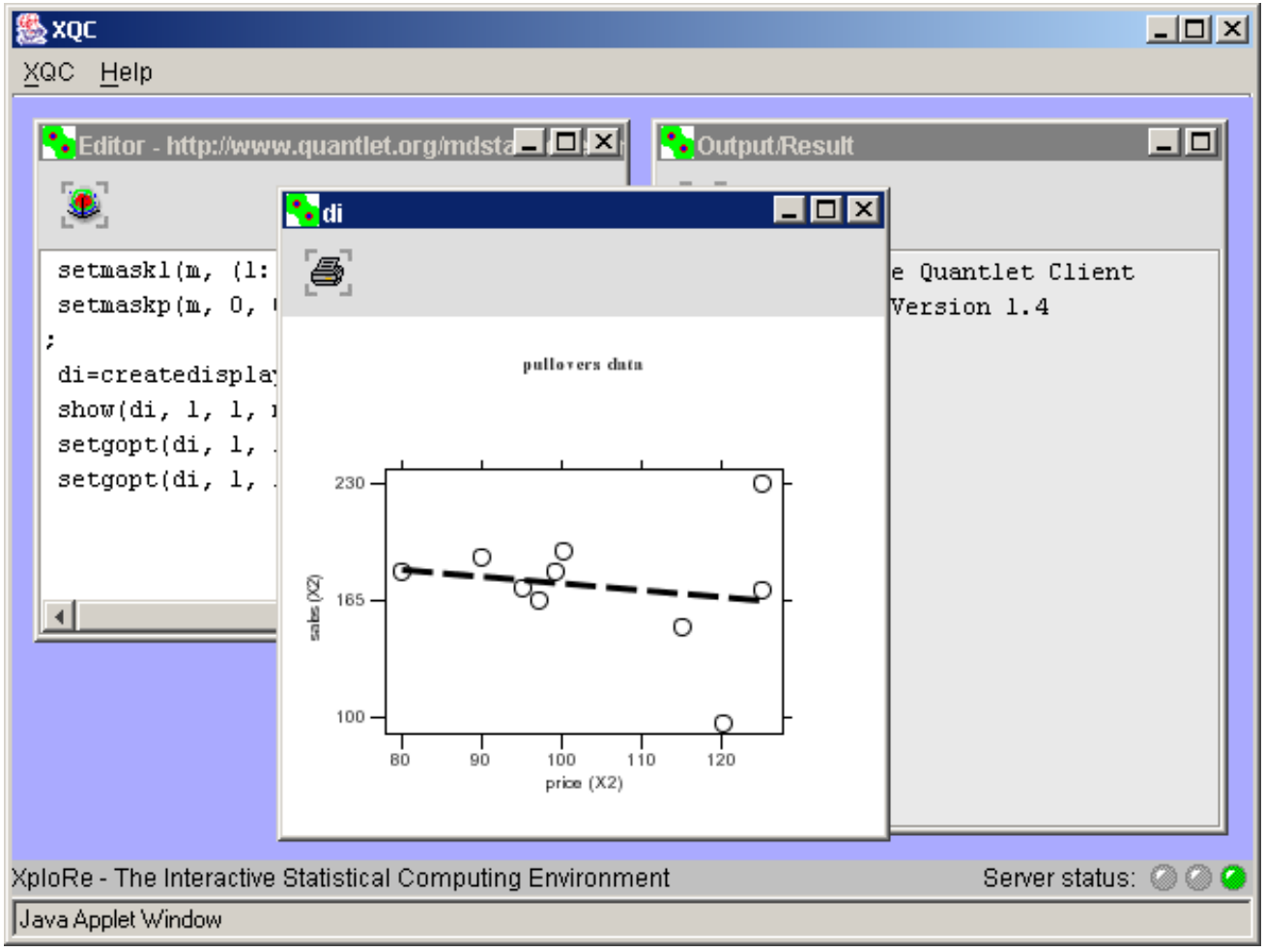
- 'golden solution', used for various books
  - embed special links in the LaTeX source
  - generate respective HTML-pages
  - pages for 'edit', 'execute' and download of XploRe files
- HTML pages use Java applets with XploRe Quantlet Server
- additionally: easily generate PDF and PostScript versions

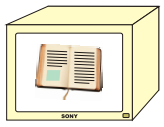



Applied Multivariate Statistical Analysis - Microsoft Internet Explorer
Datei Bearbeiten Ansicht Favoriten Extras ?
Adresse http://www.quantlet.org/mdstat/codes/mva/MVAregpull.html
Wechseln zu
QUANTLETS
home
execute
edit
help
MVAregpull
Description: MVAregpull computes a linear regression of sales (X1) on price (X2) from the pullovers data set ("pullover.dat")
Download: MVAregpull.xpl
Code:
library("xplore")
x=read("pullover") ; reads the pullovers data
y=x[,1] ; prices (X2)
x=matrix(rows(x))~x[,2] ; constant & sales (X1)
beta=gls(x, y) ; computes beta (lin. regression)
d=x[,2]~y ; data points
line=min(x[,2])|max(x[,2])
m=line~(((1|1)~line)*beta) ; regression line
setmaskl(m, (1: rows(m))', 0, 2, 2)
setmaskp(m, 0, 0, 0)
di=createdisplay(1, 1)
show(di, 1, 1, m, d) ; shows data and regression line
setgopt(di, 1, 1, "title", "pullovers data")
setgopt(di, 1, 1, "xlabel","price (X2)","ylabel","sales (X2)")
MD*Tech Method and Data Technologies
Internet

⇒ Link

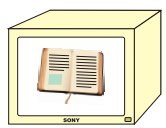
Electronic Books – Example Applet





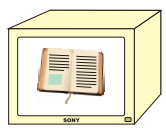
Electronic Books – Review

+ based mostly on standard software (LaTeX)
+ different output formats, searchable index (Google)
+ allows real-world data to be used
- complex structure, requires profound knowledge
- not independent from underlying computing engine
- accessibility not considered



e-stat

- developed by a group of German universities
- focus: provide materials for students and people with interest in statistics
- content delivered as XML modules following a DTD
- XML code arranged in views and scenarios with three levels of abstraction: introductory (A), applied (B), advanced (C)
- different ways of learning, based on methods, problems or views
- http://www.emilea.de



Inhaltsverzeichnis

Support Vektor Maschinen

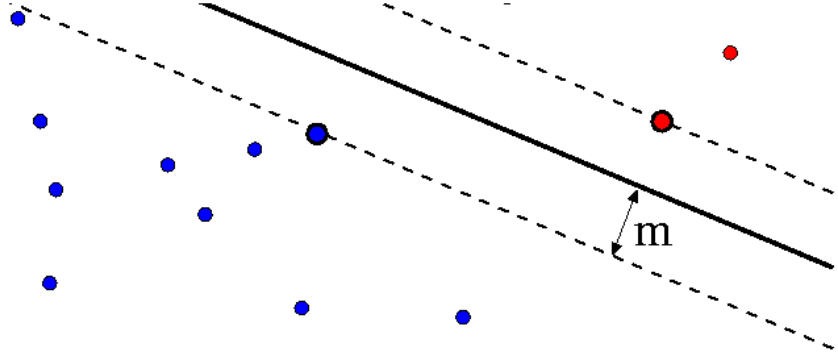


Abbildung Das Prinzip einer Support Vektor Maschine: Es soll die Hyperebene gefunden werden, die den Rand m maximiert.

Abbildung ■ veranschaulicht das Prinzip der Support Vektor Maschine. Der Rand m einer Hyperebene ist der Abstand zum nächst Trainingsvektor. Die Länge des Gewichtsvektors $\mathbf{a}$ wird normiert, ansonsten könnte man den Rand durch Skalierung des Gewichtsvektors über jede Schranke heben.

Die Trainingsvektoren $\mathbf{x}_i$ die auf dem Rand liegen, d.h. für die gilt

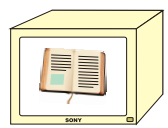
$$\frac{y_i \mathbf{a}^t \mathbf{x}_i}{||\mathbf{a}||} = m$$

Formel : (3.131)

heißen **Supportvektoren** (Stützvektoren). Die Supportvektoren bestimmen die separierende Hyperebene. Läßt man die übrigen Vektoren weg, so ändert sich die optimale Entscheidungsgrenze im Sinne der Supportvektormaschine nicht.

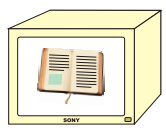
e-stat – Review

+ covers different courses of studies
+ different levels of abstraction and ways of accessing topics
+ sustainability
- complex framework required to add new content
- for only a subset of topics all levels have been implemented
- accessibility not considered

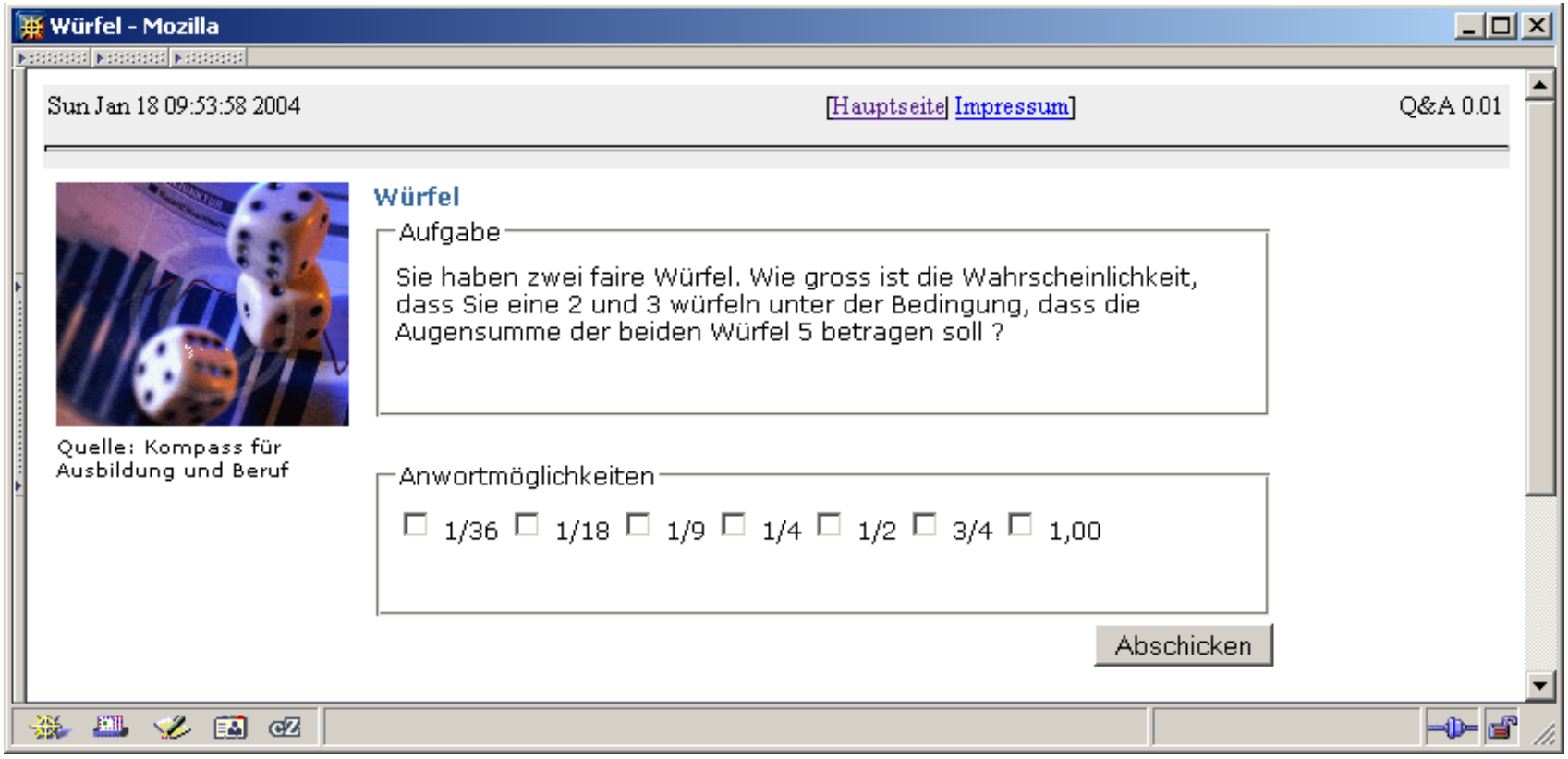


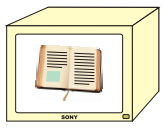
Q & A (Questions and Answers)

- idea: interactive, web-based environment for practicing
- typically 300-400 students per year in undergraduate statistics
- evaluation of exams showed 'typical' errors
- students can choose via checkbox correct solution
- if wrong answer is given, hint on the error is presented



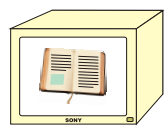
Q & A (Questions and Answers)





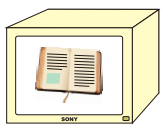
Q & A – Review

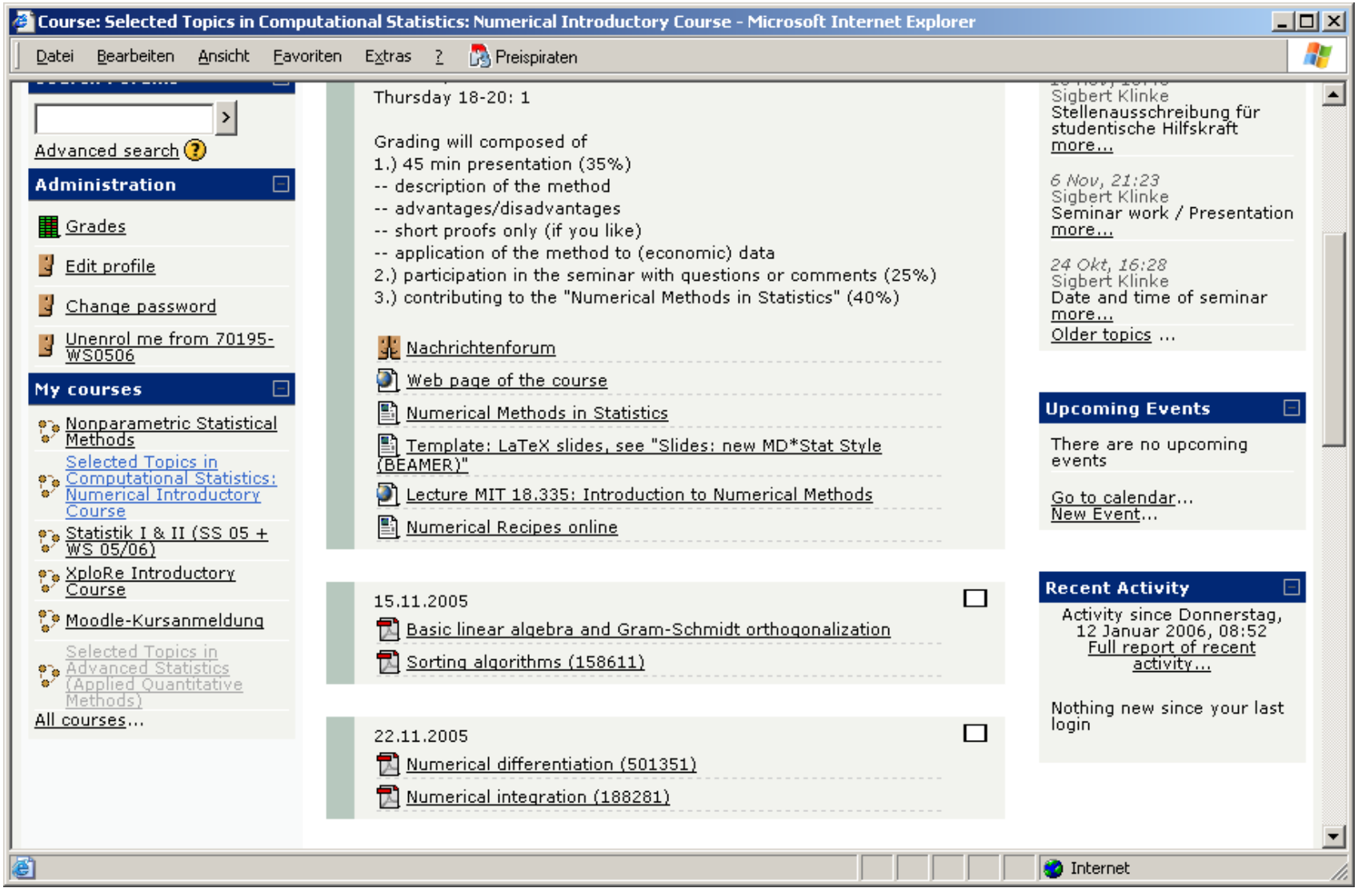
+ covers typical error sources
+ positive feedback and hints from students
+ follows structure of the examination
- requires huge amount of time from the lecturer
- students tend to trial-and-error to find correct solution
- accessibility not considered



Moodle

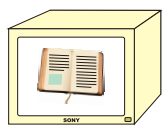
- learning management software
- open source software based on PHP and MySQL
- helps lecturers create effective online learning communities
- Moodle allows to
  1. structure lectures according to a set of topics or a time table,
  2. upload and manage documents (slides, exercises) for lecturers and students
  3. set up and maintain forums for student/student and student/teacher communication
  4. generate online exercises with direct evaluation.




Course: Selected Topics in Computational Statistics: Numerical Introductory Course - Microsoft Internet Explorer
Datei Bearbeiten Ansicht Favoriten Extras ? Preispiraten
Advanced search
Administration
Grades
Edit profile
Change password
Unenrol me from 70195-WS0506
My courses
Nonparametric Statistical Methods
Selected Topics in Computational Statistics: Numerical Introductory Course
Statistik I & II (SS 05 + WS 05/06)
XploRe Introductory Course
Moodle-Kursanmeldung
Selected Topics in Advanced Statistics (Applied Quantitative Methods)
All courses...
Thursday 18-20: 1
Grading will composed of
1.) 45 min presentation (35%)
-- description of the method
-- advantages/disadvantages
-- short proofs only (if you like)
-- application of the method to (economic) data
2.) participation in the seminar with questions or comments (25%)
3.) contributing to the "Numerical Methods in Statistics" (40%)
Nachrichtenforum
Web page of the course
Numerical Methods in Statistics
Template: LaTeX slides, see "Slides: new MD*Stat Style (BEAMER)"
Lecture MIT 18.335: Introduction to Numerical Methods
Numerical Recipes online
15.11.2005
Basic linear algebra and Gram-Schmidt orthogonalization
Sorting algorithms (158611)
22.11.2005
Numerical differentiation (501351)
Numerical integration (188281)
Sigbert Klinke
Stellenausschreibung für studentische Hilfskraft
more...
6 Nov, 21:23
Sigbert Klinke
Seminar work / Presentation
more...
24 Okt, 16:28
Sigbert Klinke
Date and time of seminar
more...
Older topics ...
Upcoming Events
There are no upcoming events
Go to calendar...
New Event...
Recent Activity
Activity since Donnerstag, 12 Januar 2006, 08:52
Full report of recent activity...
Nothing new since your last login
Internet

Moodle – Review

+ gives good overview about the number of students
+ allows communication with all registered students
+ simplified file management (no HTML editing necessary)
+ accessibility for physically handicapped
- requires a certain initial amount of time
- advantage for students

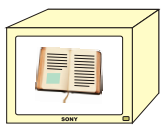


Other Packages

Microsoft Excel available everywhere, well-known interface

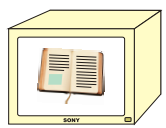
DoLStat@d by Okayama University, Japan: web based learning system using real world data together with their analysis stories

'Neue Statistik' by Freie Universität Berlin: problem-oriented and practical approach using animations, diagrams and video sequences



Conclusion

- *e*-Learning may provide valuable help for students & lecturers
- often huge time investment on the teachers' side without noticable improvement in students' understanding
- consequent evaluation of is rarely done, accesslogs are unsufficient
- require willingness to adjust the own behavior from students and teachers
- students only use them if they feel a real advantage
- best practice: start with simple solution, have students evaluate, evaluate, evaluate...

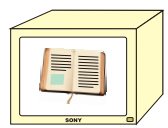


References

Grune, C. (2002): Neue Statistik, Talk at Learntec 2002, Karlsruhe

Klinke, S. (2004): Q&A - Variable multiple choice exercises with commented answers. In: Antoch, J. (ed.) COMPSTAT. Proceedings in Computational Statistics - 16th Symposium held in Prague, Physika Verlag, Heidelberg, p. 1299-1304

Kristöfl, R. (2005): Evaluation von Lernplattformen: Verfahren, Ergebnisse und Empfehlungen, Report of the Austrian ministry for Education, Science and Culture, `http://moodle.de/mod/resource/view.php?id=706`



References


Klinke, S., Witzel, R. (2002): MD*Book online - A tool for creating interactive documents. In: COMPSTAT. Proceedings in Computational Statistics - 15th Symposium held in Berlin, Physika Verlag, Heidelberg, p. 98-99


Müller, M., Rönz, B., Ziegenhagen, U. (2000): The multimedia project MM*STAT for teaching statistics. In: Bethlehem, J., van der Heijden, P. (ed.) COMPSTAT. Proceedings in Computational Statistics - 14th Symposium held in Utrecht, Physika Verlag, Heidelberg, p. 409-415


Mori, Y., Yamamoto, Y. and Yadohisa, H. (2003): Data-oriented Learning System of Statistics based on Analysis Scenario/Story (DoLStat). Bulletin of the International Statistical Institute (ISI).


OECD (2005): *e*-Learning in Tertiary Education. Where do we stand?