

Web-based tools for the teaching of statistics using R and MATLAB/Gnu Octave

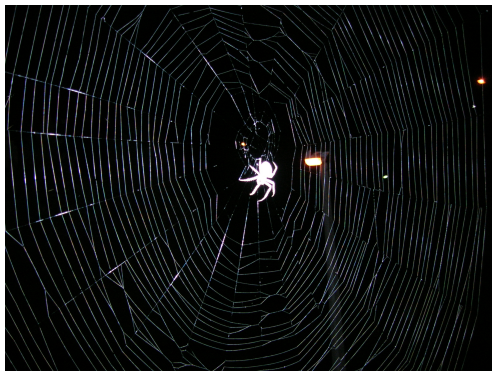
Sigbert Klinke

Humboldt-Universität zu Berlin
Ladislaus von Bortkiewicz Chair of Statistics



Applicable Semiparametrics - 12 October 2013

The web as hub



Source: Wikimedia Commons

- Teaching statistics
 - Lectures
 - Teaching material
 - Exercises
 - Exams

The teacher: spider or fly?

The university uses methods like a thousand years ago

Home | Video | Themen | Forum | English | DER SPIEGEL | SPIEGEL TV | Abo | Shop | Schlagzeilen | Wetter | TV-Programm | mehr ▼

SPIEGEL ONLINE UNISPIEGEL

Politik | Wirtschaft | Panorama | Sport | Kultur | Netze | Wissenschaft | Gesundheit | Einestages | Karriere | Uni | Schule | Reise | Auto

Nachrichten » UnISPIEGEL » Studium » Studium und Internet » Sebastian Thrun: Stanford University lehrt "wie vor 1000 Jahren"

Ex-Stanford-Professor Thrun: "Die Uni nutzt Methoden wie vor tausend Jahren"

Er war Professor an der US-Eliteschmiede Stanford - doch Sebastian Thrun, Experte für Künstliche Intelligenz, hat genug vom alten Uni-Geschäft. Im Interview erklärt er, warum er nur noch über eine Web-Plattform lehren will und was Hochschulen mit Ex-Freundinnen gemeinsam haben.


Montag, 19.03.2012 - 11:39 Uhr


Drucken | Versenden

Nutzungsrechte | Feedback



Empfehlen | 50 | +1

Zur Person

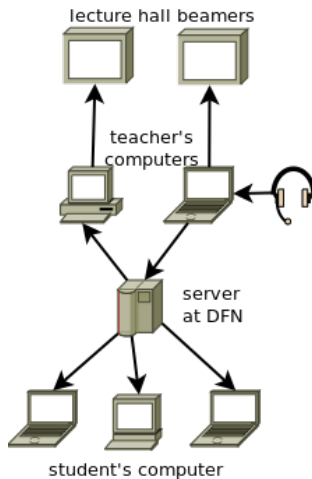
 **Sebastian Thrun**, 44, wurde in Solingen geboren. Er studierte Informatik in

 Getty Images

Source: Spiegel Online 19.03.12

- Thrun teaches since 2012 at [Udacity](#) , an online university
- Can we stream our lecture to internet? Yes, we can!
- Technical solutions
 - [WatchMyCam](#) 
 - [BigBlueButton](#) 
 - [Adobe Connect](#) 


Limitations and solution



■ Limitations

- Bandwidth for streaming
- Tuning image and voice
- Maintaining another server and software

■ Solutions

- [DFN](#)  provides a Adobe connect server for free
- Intended for web conferencing

Experiences

2 Deskriptive Statistik

1. Grundbegriffe
 - 1.1 Statistische Variablen
 - 1.2 Skalierung
 - 1.3 Klassierung
2. Univariate Statistik
 - 2.1 Verteilung
 - 2.2 Lageparameter
 - 2.3 Streuungsparameter
3. Bivariate Statistik
 - 3.1 Gemeinsame Verteilung
 - 3.2 Randverteilung
 - 3.3 Bedingte Verteilung
 - 3.4 Unabhängige Variablen
 - 3.5 Zusammenhang für zwei metrische Variablen
 - 3.6 Zusammenhang für zwei ordinale Variablen
 - 3.7 Zusammenhang für zwei nominale Variablen
4. Anhang

LeB (HU Berlin) Desktop-Statistik Folie 2 / 154

- Since summer 2012 streaming of lectures
 - ~ 50 lectures streamed
 - STAT 101 and 102 ~ 50-100 viewers
 - recording possible, not done yet
- Problems
 - anonymous student logins
 - connection

CD

Explained: A Deck of Cards

Assume you have shuffled a standard deck of 52 playing cards. You are interested in the *probability* of a randomly drawn card being a queen or a heart. We are thus interested in probability of the event $\{(Queen) \cup (Heart)\}$.

Following Laplace's notion of probability, we proceed as follows: There are 4 queens and 13 hearts in the deck. Hence,

- $P(\{(Queen)\}) = \frac{4}{52}$
- $P(\{(Heart)\}) = \frac{13}{52}$

But there is also one card which is both a queen and a heart. As this card is included in both events, we would overstate the probability of either queen or heart appearing if we simply added both probabilities. In fact, the addition rule of probability requires one to deduct the probability of this joint event

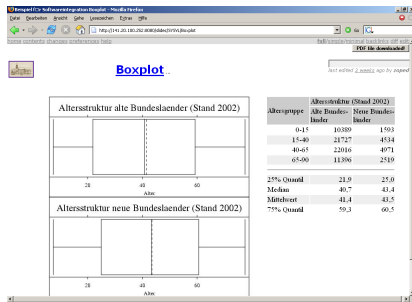
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Hence,

- $P(A \cap B) = P(\{(Queen) \cap (Heart)\}) = \frac{1}{52}$

- MM*Stat CD (1998) by Härdle et al.
- book structure
 - chapter and sections
 - hyperlinks
 - different kind of examples
 - XploRe server on CD
 - multiple choice exercises
- various languages
- problems
 - JScript used
 - only IE 5ff
 - *no maintenance!*

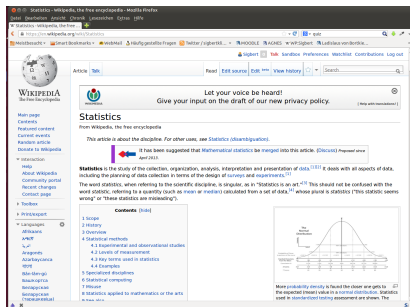
ZWiki









A wiki is usually a web application which allows people to add, modify, or delete content in a collaboration with others.

- wiki requirements
 - math (via \LaTeX)
 - exercises (quizzes)
 - integration of statistical software
- prequel wiki (2003)
 - with Zope, ZWiki and LatexWiki plugin
 - stopped in 2005 (*no maintenance!*)
 - integration of XploRe and R

MediaWiki



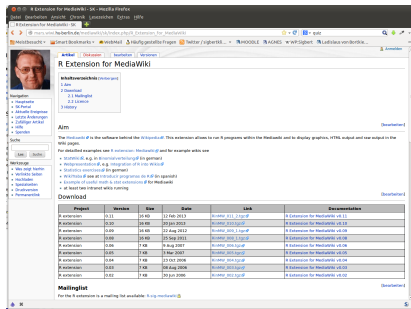
- Mediawiki  offers
 - quizzes
 - math rendering
 - a lot more extensions
- other wikis possible
 - DokuWiki 
 - TWiki 
 - for more see wikimatrix.org 
- our wikis (2005)
 - Teachwiki 
 - StatWiki 

Maintain content not software



- MM*Stat wiki (2006)
 - save the content of CD
 - use [Quiz](#) extension for exercises
 - statistical software?
- R extension for Mediawiki
 - several web-based interfaces for R
 - [R-php](#)
 - [RStudio](#) web server
- Integration of videos

R extension



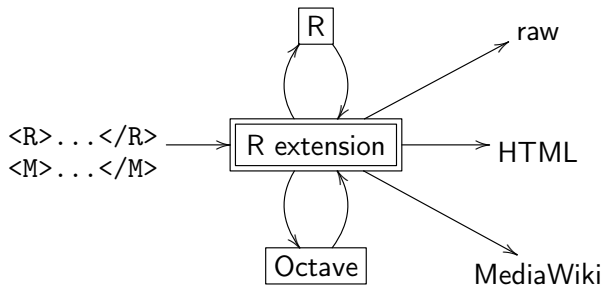
The screenshot shows the project page for 'R Extension for MediaWiki'. It includes a navigation menu, a list of versions, an 'AIM' section describing the software's purpose, a 'Download' section with a table of releases, and a 'Mailinglist' section.


Version	Release	Date	Link	Download
R.extension 0.11	10 KB	12 Feb 2013	Cvinko 011_1.jpg	R Extension for MediaWiki 01.11
R.extension 0.10	10 KB	29 Jan 2013	R.extension_010.jpg	R Extension for MediaWiki 01.10
R.extension 0.09	10 KB	27 Aug 2012	R.extension_009.jpg	R Extension for MediaWiki 01.09
R.extension 0.08	10 KB	25 Sep 2011	Cvinko_008_1.jpg	R Extension for MediaWiki 01.08
R.extension 0.06	1 KB	30 Aug 2007	R.extension_006.jpg	R Extension for MediaWiki 01.06
R.extension 0.05	1 KB	30 Aug 2007	R.extension_005.jpg	R Extension for MediaWiki 01.05
R.extension 0.04	1 KB	23 Oct 2006	Cvinko_004.jpg	R Extension for MediaWiki 01.04
R.extension 0.03	1 KB	28 Aug 2006	R.extension_003.jpg	R Extension for MediaWiki 01.03
R.extension 0.01	1 KB	29 Jan 2006	Cvinko_001.jpg	R Extension for MediaWiki 01.01

Klinke & Zlatkin-Troitschanskaia
(2007) *Embedding R in the Mediawiki*,
SFB 649 Discussion Papers.

- development start in 2006
- goal
 - generate tables
 - generate figures
- output
 - raw & graphics
 - HTML
 - wiki
- how many distribution function tables to display?
 - limited interactivity

Workflow



 running R or Octave code on a web server is a security risk!

Examples

Verteilungsfunktion

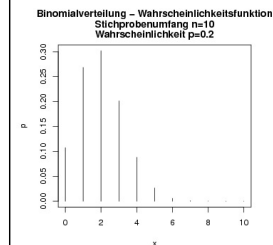
Für $n > 30$ sind geeignete Approximationen anzuwenden.

p:

p= 0.5 x/n	1	2	3	4	5
0	0.5000	0.2500	0.1250	0.0625	0.0312
1	1.0000	0.7500	0.5000	0.3125	0.1875
2	1.0000	1.0000	0.8750	0.6875	0.5000
3	1.0000	1.0000	1.0000	0.9375	0.8125
4	1.0000	1.0000	1.0000	1.0000	0.9688
5	1.0000	1.0000	1.0000	1.0000	1.0000
6	1.0000	1.0000	1.0000	1.0000	1.0000
7	1.0000	1.0000	1.0000	1.0000	1.0000
8	1.0000	1.0000	1.0000	1.0000	1.0000
9	1.0000	1.0000	1.0000	1.0000	1.0000
10	1.0000	1.0000	1.0000	1.0000	1.0000
11	1.0000	1.0000	1.0000	1.0000	1.0000
12	1.0000	1.0000	1.0000	1.0000	1.0000
13	1.0000	1.0000	1.0000	1.0000	1.0000
14	1.0000	1.0000	1.0000	1.0000	1.0000
15	1.0000	1.0000	1.0000	1.0000	1.0000
p= 0.5 x/n	16	17	18	19	20
0	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0003	0.0001	0.0001	0.0000	0.0000
2	0.0021	0.0012	0.0007	0.0004	0.0002
3	0.0106	0.0064	0.0038	0.0022	0.0013
4	0.0384	0.0245	0.0154	0.0096	0.0059

Graphische Darstellung der Wahrscheinlichkeitsfunktion

n: prob:



Security

- Highly dangerous
 - system calls
 - eval calls
- Further risks
 - file I/O → reading data?
 - user interaction
 - graphical user interface
 - system information
- Defence
 - scan input (forbidden routines)
 - scan output (PHP, JavaScript)
 - virtual machine
- *no 100% security*



Source: Wikimedia Commons

Homework & written exams



Source: Wikimedia Commons

- Homework
 - planned in 2014
 - lead students to statistics earlier
 - experience show: 60% of students are not teamworking
- Exams
 - number raises: +40% in last term
 - two exams after lecture
 - then nine months break
- Solution: put exercises online

MAUTS

Statistik I & II (SS 13 + WS 13/14)

Test: Test zum Thema "Univariate Deskriptive Statistik"
Dieser Test läuft nur bis Do, 10. Oktober 2013, 14:06
Versuche: 394

Test: Test zum Thema "Bivariate Deskriptive Statistik"
Dieser Test läuft nur bis Do, 10. Oktober 2013, 15:27
Versuche: 182

Test: Test zum Thema "Zufallsvariablen"
Dieser Test läuft nur bis Do, 10. Oktober 2013, 14:08
Versuche: 152

Test: Test zum Thema "Kombinatorik"
Dieser Test läuft nur bis So, 6. Oktober 2013, 23:59
Versuche: 226

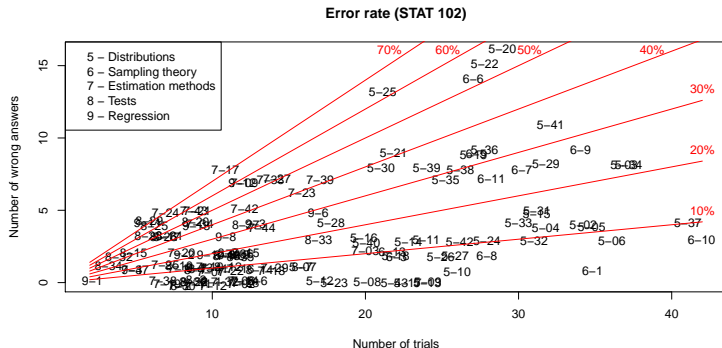
Test: Test zum Thema "Grundzüge der Wahrscheinlichkeitsrechnung"
Dieser Test läuft nur bis So, 6. Oktober 2013, 23:59
Versuche: 203

Test: Vorbereitung zur Klausur
Dieser Test läuft nur bis Do, 10. Oktober 2013, 12:13
Versuche: 310

From: 2 Oct 2013

- Moodle - the CMS of HU Berlin
 - offers a Test module
 - allows for random questions
 - answer format: numeric, free text, (multiple choice), ...
- MAUTS project
 - transfer >400 STAT 101/102 exercises to Moodle
 - generate tests for specific topics and an exam

Exercise analysis



Vorschau Frage Studentinnen 6-6

Frage 1

Bisher nicht beantwortet
Erreichbare Punkte: 1,00

Der Anteil der Studentinnen an allen Studierenden einer Hochschule beträgt 40%. Das Studentenwerk zieht für eine Erhebung eine einfache Zufallsstichprobe vom Umfang $n = 30$. Wie groß ist die Wahrscheinlichkeit dafür, dass in dieser Stichprobe weniger als 30% oder mehr als 50% Studentinnen sind?

Antwort:

Q&A

Questions and Answers 00.05

Description and overview:
 Questions and Answers is a tool to display (database) multiple choice questions to a user and to judge his/her answers. In case the user chooses the wrong answer then we know depending on the answer which mistake the user has done. Please try some exercises!

Currently we offer the following exercises:

- Questions from all statistics examinations (Lernaktiv)
- Examfragen Statistik I (L1, WS 05/06 (Lernaktiv))
- Examfragen Statistik I (L1, SS 06/07 (Lernaktiv))
- Examfragen Statistik I (L1, SS 07/08 (Lernaktiv))
- Examfragen (Lernaktiv)

Hint: Only exercises with a label are finished, others most probably will not work!

Copyright
 Copyright (C) 2003-2009 Ingrid Klinka [Impressum](#)

This program is free software; you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation; either version 2 of the License, or (at your option) any later version.

This program is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of

Statistical calculator

Verfahrenstechnik
[Statistik](#) [Statistikgrundvorlesung](#) [Zusätze](#) [Statistikgrundvorlesung](#) [Zusätze](#) [Statistikgrundvorlesung](#) [Zusätze](#) [Statistikgrundvorlesung](#) [Zusätze](#)

- **Questions & Answers** (2002)
 - multiple choice exercises
 - from written exams
 - multiple exercise generation
 - analyse and react on student answer

Verkehrsunfälle

Aufgabe
 Das statistische Bundesamt hat für eine Region in Deutschland die folgenden Zahlen für die im Straßenverkehr getöteten Personen im Alter von 15 bis 65 Jahren angegeben:

Alter	15 - 40 Jahre	40 - 45 Jahre	45 - 50 Jahre	50 - 55 Jahre	55 - 65 Jahre
Getötete	25	25	5	15	30

Berechnen Sie den Mittelwert des Alters der Getöteten.

Antwortmöglichkeiten

40.0 40.0 41.4 42.5 43.8 45.5 47.7 49.4 51.0 58.3

Verkehrsunfälle

Ihr Ergebnis ist falsch!

Lösung

1. Berechnen Sie die Häufigkeiten und Häufigkeitsdichten


x_j^*	$X \leq x_j^*$	$h(x_j)$	$f(x_j)$	$f(x_j)$
15 - 40	25	0.25	0.01	
40 - 45	25	0.25	0.05	
45 - 50	5	0.05	0.01	
50 - 55	15	0.15	0.03	
55 - 65	30	0.3	0.03	

2. Wählen Sie die Modellklasse
 3. Nehmen Sie die Punktberechnung vor

$$x_D = x_j^* + \frac{f(x_j) - f(x_{j-1})}{2f(x_j) - f(x_{j-1}) - f(x_{j+1})} \cdot (x_j^* - x_{j-1}^*)$$

Ihre falsche Lösung ist

exams package of R

- Developed by [Grün and Zeileis \(2009\)](#) 
- Random selection of exercises
- Uses \LaTeX & Sweave
- Generates
 - \LaTeX
 - PDF
 - Moodle
 - ...

⇒ Individual E-exam

```
myexam <- list(  
  list.files(path, "^stichprobe"),  
  c(list.files(path, "^maxlik"), ...),  
  list.files(path, "^missing"),  
  list.files(path, "^outlier"),  
  list.files(path, "^robust"),  
  list.files(path, "^univariate"),  
  list.files(path, "^transformation"),  
  list.files(path, "^verteilung"),  
  list.files(path, "^test_allg"),  
  ...  
  list.files(path, "^prak")  
)  
  
exams(myexam, n=30,...)
```

End of talk

