

# Assessment of Reproducible Computing as an E-Learning Tool in Statistics Education

E-Learn 2008, Las Vegas, Nevada, USA

Patrick Wessa

# Acknowledgments

- Funding:

- K.U.Leuven Association, OOF 2007/13
- Donations from private companies

- Contributors:

Bart Baesens, Eric Bloemen, Eddy Borghers, Christophe Croux, Claude Doom, Dirk Janssens, Christine Lourdon, Koen Milis, Stephan Poelmans, Riko van Dijk, Guido Van Rompuy, Ed van Stee, Larry Weldon, Patrick Wessa  
([www.freestatistics.org](http://www.freestatistics.org))

# Claerbout's principle\*

- An article about computational science in a scientific publication is not the scholarship itself, it is merely **advertising** of the scholarship. The actual scholarship is the complete software development environment and that complete set of instructions that generated the figures.

\*Source: Jan de Leeuw

# My question

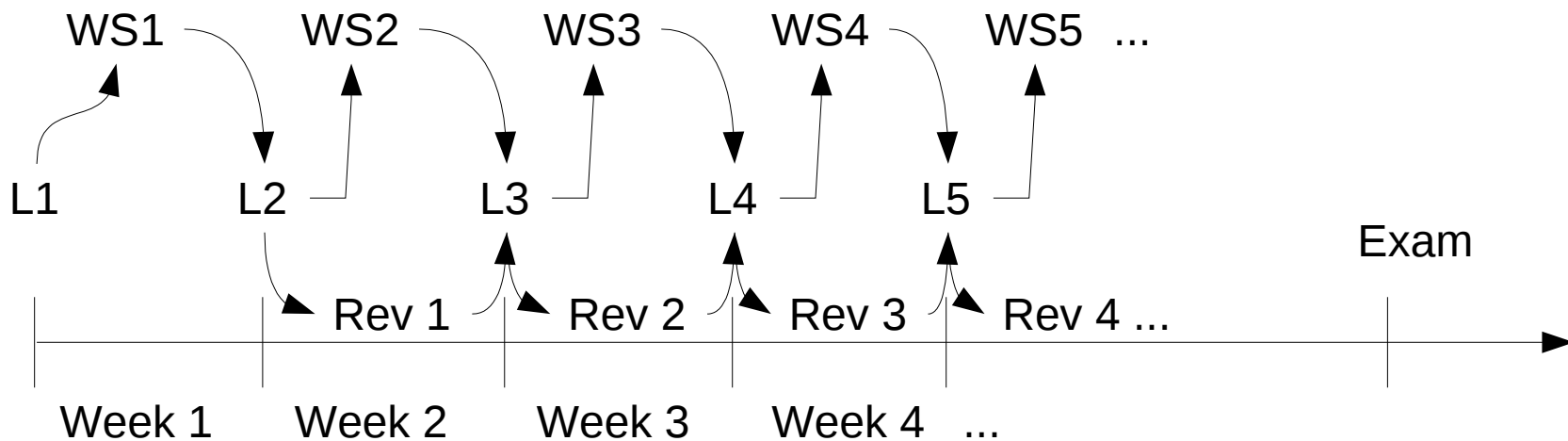
- If academic statisticians find it hard (if not impossible) to verify or review the results in empirical papers, how could we possibly expect students to learn from statistical results without the proper tools to easily review, verify, or challenge them?

# My proposed solution

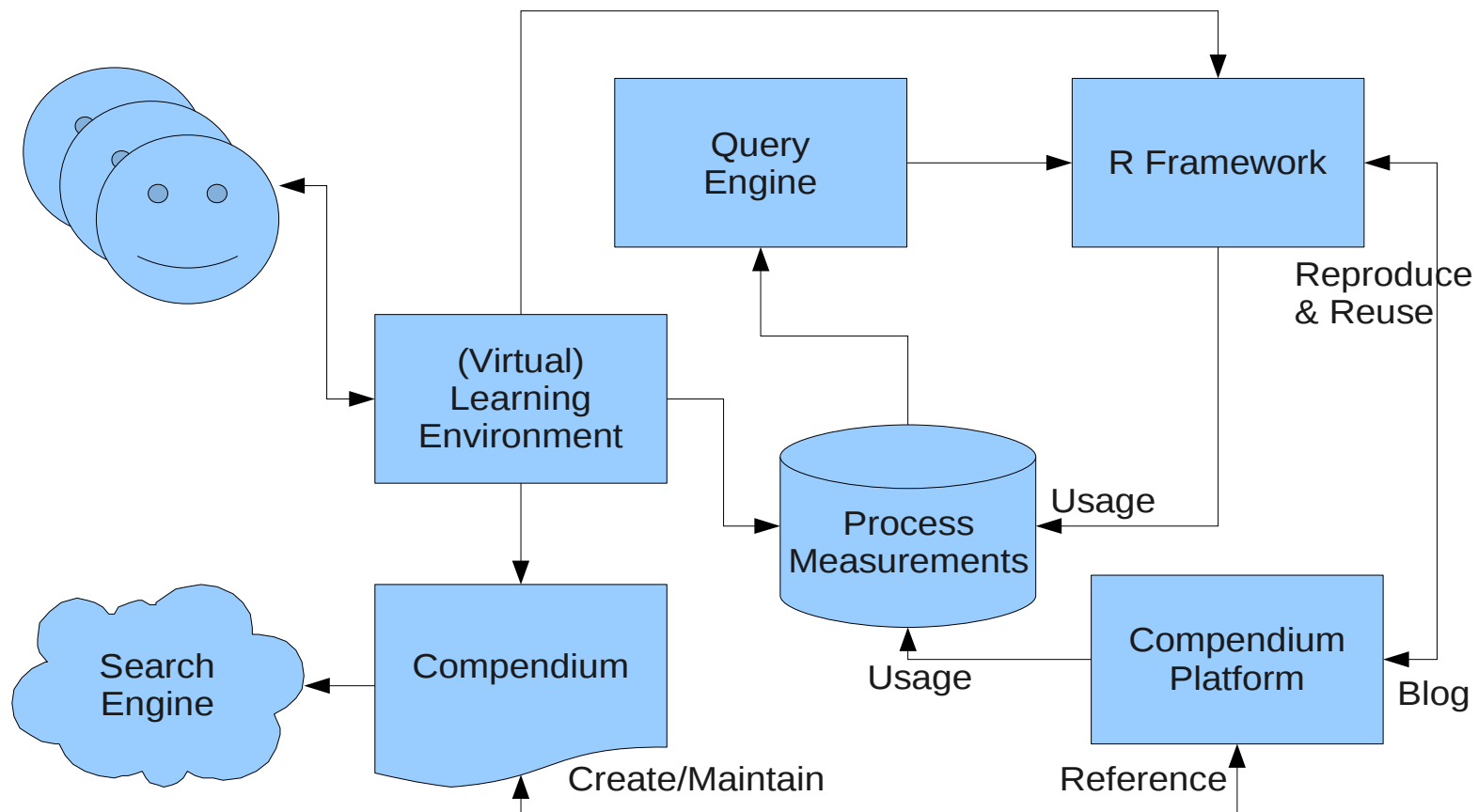
- Poelmans, S., Wessa, P., Milis, K., Bloemen, E., and Doom, C.: Usability and Acceptance of E-Learning in Statistics Education, based on the Compendium Platform, Proceedings of the International Conference of Education, Research and Innovation, International Association of Technology, Education and Development, 2008
- Wessa, P.: A framework for statistical software development, maintenance, and publishing within an open-access business model, Computational Statistics, 2008
- Wessa, P.: Let us free statistics of irreproducible research, Statistics Seminar at Simon Fraser University, Vancouver, Canada, 2008
- Wessa, P.: Learning Statistics based on the Compendium and Reproducible Computing, Proceedings of the World Congress on Engineering and Computer Science (International Conference on Education and Information Technology), UC Berkeley, San Francisco, USA, 2008
- Wessa, P.: Assessment of Reproducible Computing as an E-Learning Tool in Statistics Education, Proceedings of the World Conference on E-Learning in Corporate, Government, Healthcare, & Higher Education, 2008
- Wessa, P.: Measurement and Control of Statistics Learning Processes based on Constructivist Feedback and Reproducible Computing, Proceedings of the 3rd International Conference on Virtual Learning, 2008
- Wessa, P.: How Reproducible Research Leads to Non-Rote Learning Within a Socially Constructivist E-Learning Environment, Proceedings of the 7th European Conference on e-Learning, 2008
- Wessa, P.: A Compendium Platform for Reproducible, R-based Research with a focus on Statistics Education, UseR Conference, Dortmund, Germany, 2008
- ...

# Lectures

- 13 weeks (semester)
- Week 1: Introduction (explanation) + workshop assignment
- Week 2-12: Workshops + Peer Assessments
- Week 13: Final Exam (multiple choice)
- Grades received from Peers do NOT count => there is no penalty for making mistakes!!
- The quality of feedback messages is graded by the educator



# Learning System or Educational Laboratory?



Multiple Regression - Free Statistics and Forecasting Software (Calculators) v.1.1.23-r1 - Mozilla

File Edit View History Bookmarks Tools Help

http://www.wessa.net/rwasp\_multipleregression.wasp

Server St... Gmail - In... Blog & Sh... Course: A... Blog & Sh... Multiple ...

**Fixed Seasonal Effects**  
 Include Monthly Dummies

**Type of Equation**  
 Linear Trend

**Chart options**  
 Width: 600  
 Height: 400

**R Code**

```

library(lattice)
par1 <- as.numeric(par1)
x <- t(y)
k <- length(x[,1])
n <- length(x[,1])
x1 <- cbind(x[,par1], x[,1:kt=par1])
mycolnames <- c(colnames(x)[par1], colnames(x)[1:kt=par1])
colnames(x1) <- mycolnames #colnames(x)[par1]
x <- x1
if (par3 == 'First Differences'){
  x2 <- array(0, dim=c(n-1,k), dimnames=list(1:(n-1), paste('(1-
B)'.colnames(x), sep='')))
  for (i in 1:n-1) {
    for (j in 1:k) {
      x2[i,j] <- x[i+1,j] - x[i,j]
    }
  }
  x <- x2
}
if (par2 == 'Include Monthly Dummies'){
  x2 <- array(0, dim=c(n,11), dimnames=list(1:n, paste('M', seq(1:11), sep='')))
  for (i in 1:11){
    x2[seq(i,n,12),i] <- 1
  }
  x <- cbind(x, x2)
}

```

Compute

Summary of computational transaction

Raw Input	view raw input (R code)
Raw Output	view raw output of R engine
Computing time	6 seconds
R Server	'Sir Ronald Aylmer Fisher' @ 193.190.124.24

Multiple Linear Regression - Estimated Regression Equation

```

ongevallen[t] = + 2324.06337310277 - 226.389033602698[t] - 451.374973296311M[t] - 626.461053323709M2[t] - 983.133097991392M3[t] - 694.266342698015M4[t]
- 526.4789873226630M5[t] - 629.464131594261M6[t] - 532.674270661884M7[t] - 515.434421329607M8[t] - 460.88706997131M9[t] - 319.717210064794M10[t]
- 118.309626332377M11[t] - 1.764266332376026t + e[t]

```

Multiple Linear Regression - Ordinary Least Squares

Variable	Parameter	S.D.	T-STAT	2-tail p-value	1-tail p-value
Intercept	2324.06337310277	44.025939	52.7837	0	0
x	-226.389033602698	41.037226	-5.5166	0	0

Done

A framework for statistical software development, maintenance, and publishing within an open-access business model, 2008, Computational Statistics, Springer



# Computations are “blogged”

Blog & Share - Free Statistics and Forecasting Software (Calculators) v.1.1.23-r1 - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.wessa.net/blogshare.wasp?outtype=&id=8&command=blog&check=Sun, 03 Aug 2008 04:33

Server Status ... Gmail - Inbox (... Blog & Share S... Course: Applie... Blog & Share S... International C... Blog & Share - ...

http://www.freestatics.org/) where it is permanently archived for reference purposes. In addition visitors of the Blog can Discuss, Reproduce, and Reuse all Statistical Computations in the archive.

**Submit your Statistical Computation to the FreeStatistics.org Archive**

Field	Value
<b>Title</b> (optional, meaningful title)	this is my title
<b>Keywords</b> (optional, comma-delimited list)	statistics, assignment 5, hypothesis testing, any other keyword
<b>Your Comments</b> (optional, any meaningful text)	I computed this hypothesis test to answer question 3 in assignment 5.
<b>E-mail</b> (optional, private - this is required if you want to edit/delete the post at a later time)	patrick@wessa.net
<b>Type of Access</b> (optional, do you want to grant everyone access to your archived computation?)	Public (anybody can access my computation)
<b>Moratorium date</b> (enter the moratorium date - only needed if 'Moratorium' is selected in 'Type of Access')	YYYY-MM-DD
<b>Captcha</b>	

Done

Multiple Regression (old)  
Descriptive Statistics  
Statistical Distributions  
Hypothesis Testing  
Statistics Education

Academic citations  
Computations Archive  
Search Computations  
R Project  
FAQ  
About Wessa.net  
Powered by Linux

Server status page  
History list

Google

Web wessa.net

Search

workshop - OpenOffice.org Writer

File Edit View Insert Format Table Tools Window Help

Equation Expand Config

Default Garamond 12

**Question 2: Investigate the prediction errors of the model that you used in question 1. Are the underlying regression assumptions satisfied?**

De Adjusted R-squared is gelijk aan 0,6412. Dit wil zeggen dat we 64% van de wijzigingen van het aantal verkeersslachtoffers kunnen verklaren. Het resterende gedeelte (36%) kunnen we niet verklaren aan de hand van ons model maar zijn bijvoorbeeld te wijten aan uitzonderlijke weersomstandigheden.

Het volgende waar we heen zullen kijken is de Interpolation Plot. De stippellijn op deze grafiek geeft het werkelijke aantal slachtoffers weer. De volle lijn geeft het aantal verkeersslachtoffers ~~we~~er die voorspeld zijn door het model.

<http://www.freeststatistics.org/blog/index.php?v=date/2007/Nov/14/t1195074007ni07puuvacjlu0w.htm>

**Learning Statistics based on the Compendium and Reproducible Computing, Proceedings of the World Congress on Engineering and Computer Science 2008, ISBN: 978-988-98671-0-2, UC Berkeley, San Francisco, USA**

# Snapshot of “Blogged” Computation

The screenshot shows a web browser window with the following elements:

- Browser Title Bar:** Blog & Share Statistical Computations at FreeStatistics.org
- Address Bar:** <http://www.freeststatistics.org/blog/index.php?v=date/2007/Nov/14/t1195074007ni07puuvacjlu0w.htm>
- Navigation:** Home » date » 2007 » Nov
- Buttons:** Print, PDF, TeX, Statistics, Search, Edit, Post Comment, **Reproduce**, Reuse
- Section Header:** WS 8 - Q1 (3)
- Metadata:**
  - R Software Module: `rwasp_multipleregression.wasp` (opens new window with default values)
  - Title produced by software: Multiple Regression
  - Date of computation: Wed, 14 Nov 2007 14:04:40 -0700
- Citation:**

Cite this page as follows:  
**Statistical Computations at FreeStatistics.org**, Office for Research Development and Education, URL <http://www.freeststatistics.org/blog/date/2007/Nov/14/t1195074007ni07puuvacjlu0w.htm>, Retrieved Sun, 03 Aug 2008 09:58:02 +0000
- Keywords:** IsPrivate?, User-defined keywords:
- Data:**

Dataseries X:  
» [Textbox](#) « » [Textfile](#) « » [CSV](#) «

```
1687 0 1508 0 1507 0 1385 0 1632 0 1511 0 1559 0 1630 0 1579 0 1653 0 2152 0 2148 0 1752 0 1765 0 1717 0 1558 0 1575 0 1520 0 1805 0 1800 0 1719 0 2008 0 2242 0 2478 0 2030 0
1655 0 1693 0 1623 0 1805 0 1746 0 1795 0 1926 0 1619 0 1992 0 2233 0 2192 0 2080 0 1768 0 1835 0 1569 0 1976 0 1853 0 1965 0 1689 0 1778 0 1976 0 2397 0 2654 0 2097 0 1963 0
1677 0 1941 0 2003 0 1813 0 2012 0 1912 0 2084 0 2080 0 2118 0 2150 0 1608 0 1503 0 1548 0 1382 0 1731 0 1798 0 1779 0 1887 0 2004 0 2077 0 2092 0 2051 0 1577 0 1356 0 1652
0 1382 0 1519 0 1421 0 1442 0 1543 0 1656 0 1561 0 1905 0 2199 0 1473 0 1655 0 1407 0 1395 0 1530 0 1309 0 1526 0 1327 0 1627 0 1748 0 1958 0 2274 0 1648 0 1401 0 1411 0
1403 0 1394 0 1520 0 1528 0 1643 0 1515 0 1685 0 2000 0 2215 0 1956 0 1462 0 1563 0 1459 0 1446 0 1622 0 1657 0 1638 0 1643 0 1683 0 2050 0 2262 0 1813 0 1445 0 1762 0 1461
0 1556 0 1431 0 1427 0 1554 0 1645 0 1653 0 2016 0 2207 0 1665 0 1361 0 1506 0 1360 0 1453 0 1522 0 1460 0 1552 0 1548 0 1827 0 1737 0 1941 0 1474 0 1458 0 1542 0 1404 0
1522 0 1385 0 1641 0 1510 0 1681 0 1938 0 1868 0 1726 0 1456 0 1445 0 1456 0 1365 0 1487 0 1558 0 1488 0 1684 0 1594 0 1850 0 1998 0 2079 0 1494 0 1057 1 1218 1 1168 1 1236
1 1076 1 1174 1 1139 1 1427 1 1487 1 1483 1 1513 1 1357 1 1165 1 1282 1 1110 1 1297 1 1185 1 1222 1 1284 1 1444 1 1575 1 1737 1 1763 1
```
- Footer:** Text written by user: Done

Two callout boxes are present:

- A box pointing to the **Reproduce** button with the text: "Reproduce or Reuse at wessa.net"
- A box pointing to the citation text with the text: "Cite the computation as follows"

# Feedback (Peer Review)

The screenshot shows a Moodle workshop assessment page titled "ABS/SHW: Case: the Seatbelt Law". The page displays two evaluation elements. Each element has a "Grade" section with radio buttons for "Excellent", "Very Poor", and an unselected middle option. The "Feedback" section contains text from the reviewer. The first element's feedback is partially cut off. The second element's feedback discusses the lack of discussion on r-squared values and prediction errors. The page footer shows "Done" and "Geen titel".

Element 1: Evaluate Q1.	Weight: 1.00
Grade: Excellent <input type="radio"/> <input type="radio"/> <input checked="" type="radio"/> <input type="radio"/> <input type="radio"/> Very Poor	
Feedback: Soms wat onduidelijke uitleg over de p-waarde. Ik mis ook wat concrete uitleg over de verschillende onderdelen van het model zelf.	
Element 2: Evaluate Q2.	
Grade: Excellent <input type="radio"/> <input type="radio"/> <input type="radio"/> <input checked="" type="radio"/> <input type="radio"/> Very Poor	
Feedback: Hier staat alleen een link naar een berekening. Je had de r-squared waarden moeten bespreken. Ook had je kunnen nagaan of de voorspellingsfouten zich aan bepaalde assumpties voldaan hadden. Zo zijn er twee voorwaarden voor een goed model: het gemiddelde van de voorspellingsfouten moet 0 zijn en het gemiddelde moet constant zijn. Als we de voorspellingsfouten invullen in de central tendency calculator kunnen we zien dat aan deze assumpties niet voldaan is. Dus er is nog verbetering mogelijk aan ons model.	

Done

"Geen titel"

Submitting Peer Review (feedback) is a good learning activity – not a good grading procedure

How Reproducible Research Leads to Non-Rote Learning Within a Socially Constructivist E-Learning Environment, Proceedings of the 7th European Conference on e-Learning (ECEL'08), Cyprus

# Assessment Methodology

The analysis of the survey responses is performed in such a way that anyone is able to interpret the results. Each question was based on a 5-point Likert scale (5 is excellent, 3 is neutral, and 1 is poor). By subtracting a fixed constant (= 3) we obtained scores that are contained in the interval  $[-2, 2]$  where the neutral score is zero valued. This score  $S_{i,j}$  represents the transformed reply (for all questions  $i = 1, \dots, Q$  and for all students  $j = 1, \dots, N$ ) for which the following definitions can be formulated:

- $D_{i,j}^+ = 1$  if  $S_{i,j} > 0$ ,  $D_{i,j}^+ = 0$  and  $S_{i,j} \leq 0$
- $D_{i,j}^- = 1$  if  $S_{i,j} < 0$ ,  $D_{i,j}^- = 0$  and  $S_{i,j} \geq 0$
- $P_i^s$  is the sum of all positive scores:  $P_i^s = \sum_{j=1}^N D_{i,j}^+ S_{i,j}$  for  $i = 1, \dots, Q$
- $N_i^s$  is the sum of all absolute values of negative scores:  $N_i^s = \sum_{j=1}^N D_{i,j}^- |S_{i,j}|$  for  $i = 1, \dots, Q$
- $P_i^c$  is the number of positive scores  $P_i^c = \sum_{j=1}^N D_{i,j}^+$  for  $i = 1, \dots, Q$
- $N_i^c$  is the number of negative scores  $N_i^c = \sum_{j=1}^N D_{i,j}^-$  for  $i = 1, \dots, Q$

It is now possible to define three aggregated measures (AM) for each question:

1. the arithmetic mean:  $\frac{1}{N} \sum_{j=1}^N S_{i,j}$  for  $i = 1, \dots, Q$
2. the difference between positive and (absolute) negative scores, divided by the absolute sum of all scores  $\frac{(P_i^s - N_i^s)}{(P_i^s + N_i^s)}$  for  $i = 1, \dots, Q$
3. the difference between the number of positive and negative scores, divided by the sum of all absolute scores  $\frac{(P_i^c - N_i^c)}{(P_i^c + N_i^c)}$  voor alle  $i = 1, \dots, Q$

The first two measures can be used if a quasi-interval scale can be assumed. The third measure does not make the assumption of a quasi-interval scale because the scores are substituted by frequencies (counts). The drawback of the third measure is that it does not differentiate between extreme answers ( $\pm 2$ ) and moderate answers ( $\pm 1$ ). In other words, the third measure has the advantages that are associated with ordinal (rank-based) measures but at a cost of loss of information. The first measure is contained in the interval  $[-2, 2]$  and last two measures lie in the interval  $[-1, 1]$ .

# ATTLES (table 1)

Question	mean	$(Ps-Ns)/(Ps+Ns)$	$(Pc-Nc)/(Pc+Nc)$
1	0.45	0.55	0.51
2	0.68	0.74	0.72
3	0.58	0.63	0.6
4	0.47	0.57	0.55
5	0.76	0.79	0.77
6	0.74	0.73	0.72
7	0.7	0.73	0.68
8	0.98	0.86	0.84
9	-0.11	-0.16	-0.11
10	0.83	0.82	0.78
11	-0.37	-0.38	-0.35
12	0.91	0.86	0.84
13	1.15	0.79	0.73
14	0.5	0.59	0.54
15	0.22	0.36	0.36
16	-0.59	-0.63	-0.59
17	0.41	0.54	0.5
18	0.53	0.61	0.61
19	0.52	0.66	0.62
20	-0.17	-0.2	-0.13

**Table 1. ATTLES survey scores [Wessa, 2008e]**

# Negative Scores

- The most important part of my education has been learning to understand people who are very different to me.
- I like playing devil's advocate - arguing the opposite of what someone is saying.
- I often find myself arguing with the authors of books that I read, trying to logically figure out why they're wrong.
- I spend time figuring out what's 'wrong' with things. For example, I'll look for something in a literary interpretation that isn't argued well enough.

# Critical Thinking and Attitudes

- Negative AM scores: students lack the most fundamental attitude of good scientists which allows them to be critical and question any assumption that underlies our thinking or analysis.
- Hence, the introduction of new learning technologies that allow students to reproduce (c.q. challenge) computations from peers is expected to be difficult and lead to negative learning experiences.
- If students dislike to challenge the analysis of others, they are not likely to appreciate assignments that are related to Reproducible Computing and Peer Assessment.



# COLLES (table 2)

Q	$(Pc-Nc)/(Pc+Nc)$	Q	$(Pc-Nc)/(Pc+Nc)$	Q	$(Pc-Nc)/(Pc+Nc)$	Q	$(Pc-Nc)/(Pc+Nc)$
1	0.65	13	0.81	25	0.91	37	0.61
2	0.9	14	0.85	26	0.95	38	0.81
3	0.69	15	0.61	27	0.84	39	0.1
4	0.93	16	0.81	28	0.93	40	0.33
5	0.54	17	0.51	29	0.84	41	0.69
6	0.92	18	0.69	30	0.96	42	0.88
7	0.54	19	0.57	31	0.86	43	0.6
8	0.91	20	0.81	32	0.92	44	0.86
9	0.76	21	0.24	33	0.5	45	0.87
10	0.91	22	0.51	34	0.76	46	0.97
11	0.84	23	0.43	35	0.4	47	0.86
12	0.91	24	0.81	36	0.7	48	0.94

**Table 2. COLLES survey scores (count-based aggregated measures) [Wessa, 2008f]**

# Learning Experiences

- Overwhelming evidence that students had a positive learning experience (at the end of the semester).
- This is surprising because:
  - heavy workload
  - Reproducible Computing involves critical thinking which is disliked by students
- All questions have a positive AM - some are even close to the maximum score.

# CSUQ+ (table 3)

Q	$(Pc-Nc)/(Pc+Nc)$	Q	$(Pc-Nc)/(Pc+Nc)$	Q	$(Pc-Nc)/(Pc+Nc)$
1	0.89	12	0.57	23	0.45
2	0.82	13	0.15	24	0.87
3	0.84	14	0.7	25	0.49
4	0.62	15	0.77	26	0.93
5	0.67	16	0.51	27	0.54
6	0.73	17	0.52	28	0.94
7	0.62	18	0.82	29	0.71
8	0.67	19	0.88	30	0.9
9	-0.26	20	0.61	31	0.72
10	0.53	21	0.61	32	0.95
11	0.69	22	0.8	33	0.81

**Table 3. CSUQ survey scores (count-based aggregated measures) [Wessa, 2008g]**

# Perceived Usability

- The web-based software was highly rated by students.
- The only exception is question 9: “The website gives error messages that clearly tell me how to fix problems.”
  - Reproducible Computing offers quicker and better problem-solving
  - Communication between Developer and User

# Comparative Advantage

- AM ratings  $> 0.5$ 
  - Q20: Overall, the website was helpful in learning statistics
  - Q21: Learning Statistics with this website is more effective than with a traditional handbook
  - Q22: I intend to use this website when I need to apply statistics in the future
  - Q27: To learn statistics, this website is better than the statistical courses I have had so far

# Some advertising

- Slides:  
<http://www.wessa.net/download/elearn2008.pdf>
- Websites:
  - <http://www.wessa.net>
  - <http://www.freestatistics.org> >> Publications
- E-mail: [patrick@wessa.net](mailto:patrick@wessa.net)