

# On the Difficulty to Design Arabic E-learning System in Statistics

Taleb Ahmad

Wolfgang Härdle

CASE-Center for Applied Statistics and  
Economics

Institut für Statistik and Ökonometrie  
Humboldt-Universität zu Berlin



## Motivation

- E-learning platform, e-books, MD\*booklets are useful for Statistics education.
- The system MM\*STAT was developed at the School for Business and Economics of Humboldt-Universität zu Berlin.
- MM\*STAT exists in more 10 languages.



Figure 1: MM\*STAT with the multiple languages



هذا الكتاب الإلكتروني يتطلب [Internet Explorer 5](#) أو أكثر. لمزيد من المعلومات شاهد المساعدة التقنية.  
مقدمة تطبيقية في عالم الإحصاء

رونز ليرند من ترخيص فيسديو و هيردلي فولفغابغ من ترخيص فيسديو

Syrien	Ägypten	Kuweit	Lebanon	Tunesien	Saudi-Arabia
					

Figure 2: MM\*STAT with the Arabic language



- The implementation of MM\*STAT needed for Arab educational institutions is however difficult.
- MD\*Book created the interactive e-learning documents .
- This tool uses the LATEX format file to enable compilation into PDF, Java, HTML.
- We aim to use the LATEX file format for creating an Arabic text.

- As we will show this remains a problem, basically caused by the incompatibility of the standard Latex2html with the phenomenons of the Arabic typography.
- We need therefore to adapt the MM\*STAT architecture in a hand made fashion.

## Outline

1. Motivation ✓
2. What is e-learning ?
3. What is e-book ?
4. Arabic MM\*STAT
5. Difficulties to design, an Arabic MM\*STAT
6. Arabic MM\*STAT Realisation
7. Conclusion

## What is e-learning ?

- E-learning most often means an approach to facilitate and enhance learning through the use of devices based on computer and communications technology.
- Such devices include personal computers, CDROMs, Digital Television, Mobile Phones.
- Communications technology enables the use of the Internet, email, discussion forums, and collaborative software.

- E-Learning is the modern model of education, transferring the classical text books to an electronic medium.
- It has become an important stream of education in present and has a promising future.
- E-Learning uses net work technology to manage, design, deliver, select, and extend classical learning.

## What is an e-book ?

- Definitions of e-books differ. Hawkins (2000) states that " an e-book is the contents of a book made available in an electronic form."
- Morgan (1999) limits the definition of e-books, as opposed to e-texts, to being a hardware/software combination used to read electronic data on a specially designed portable device.
- An e-Book is an electronic version of a traditional print book that can be read by using a personal computer.

- Most books can be read as PDF files HTML, or any other encrypted format.
- Most e-book pages take the normal book page form. It may be in full colored, include graphics and photos.

## Arabic MM\*STAT

- Arabic MM\*STAT is a Web-based learning platform that provides a fully integrated student environment to learn statistics.
- MM\*STAT allows an interactive introduction into the world of statistics.



The goal is to:

- Contribute to the evolution of Statistics e-learning by the development of an efficient system.
- To create an Arabic MM\*STAT, because there were not enough Arabic statistic e-learning platforms. The purpose of this study is to develop a platform that will be used completely in Arabic, together with other languages such as English and German.

We know that the Arabic language is different from English and other languages in a number of respects:

- Arabic language is written from right to left.
- It is possible to form hundreds of words from one root word (al- Fedaghi and al-Sadoun, 1990).
- In Arabic the definite article and prepositions, are not separated from their following word by a space, on another hand some words have different meanings but have the same written form.

- The peculiar morphology of Arabic might render methods used for English text retrieval inappropriate.
- For example, the English phrase "and she wrote it" comprising of four words would be written in Arabic as one word "wakatabathu" (wa=and, kataba=wrote, t=she,hu=it). Moukdad, H., Andrew, L., (2001).
- Most Arabic letters can be connected to other letters on both the left and right sides.
- In order to look up an Arabic word in a dictionary or index, it must be reduced to its root; unlike English,
- Figure 3 presents the example of Arabic Symbols

Aussprache	Arabic	English
Sabah alhair	صباح الخير	Good morning
Naharak Said	نهارك سعيد	Good afternoon
Masa alhair	مساء الخير	Good evening
Marhaba	مرحبا	Hello
Has said	حظ سعيد	Good luck
wadaan	وداعا	Goodbye
Kaif ahwalak	كيف احوالك	How are you
Arak samatan	اراك ثانية	See you again
schukran	شكرا	thanks
tamam	تمام	fine
ana	انا	I
anta	انت	you
Schukran laka	شكرا لك	Thank you
li	لي	For me
laka	لك	For you
bisurur	بسرور	With pleasure
masa	ماذا	what
Masa turid	ماذا تريد	What do you want
Masa tafail	ماذا تفعل	What are you doing
Taal	تعال	Come here
saidn	ساعدني	Help me
Sahaba	ذهب	go
Inisami	انتظري	Wait for me
hasaman	حسنا	All right
Mata	متى	when
Kam	كم	How many
Schukran limesaadatik	شكرا لمتساعدتك	Thanks for your help
Hasi hadia laka	هذه هدية لك	This is a present for you
Kaif	كيف	how
Hasi	هذه	this
Alan	الآن	Now
gadan	غدا	tomorrow
Ams	امس	yesterday
Aljaum	اليوم	today
Huna	هنا	here
Hunaka	هناك	there
Naam	نعم	yes
La	لا	No
Mumkin	ممكنا	possible
mafhum	مفهوم	understood
Ana uraid	انا اريد	I want
uhab	احب	I like
Atamana	اتمنى	I wish
Arif	اعرف	I know
Aschtari	اشترى	I buy
La arif	لا اعرف	I do not know
Amal	اصل	I work

Figure 3: Example:Arabic Symbols

- For example, there are many terms in Statistics that simply do not have corresponding Arabic terms.
- A classic example in Statistics is that Arabic makes no distinction between "administration" and "management" - both are "idara" in Arabic. This can create unacceptable ambiguities in translation.
- Other examples of terms that have ambiguous meanings in translated Arabic are (Calculate, Compute ) and (Calculator, Computer).
- Figure 4 presents these examples:

Arabic	English
ادارة	Administration
ادارة	Management
حساب	Calculate
حساب	Compute
حاسوب	Calculator
حاسوب	Computer

Figure 4: Example:Arabic Symbols

- Generally, the difficulty with someone who has translated a software product into another language to get a realistic picture of the challenges.
- Not only are there vocabularies the translator might not know, but they also must adapt sentence structure, significance, and colloquialisms to convey equivalent meaning.

## Difficulties to design, an Arabic MM\*STAT

We now discuss some practical and technical problems.

- The development of MM\*Stat (Müller, M., Rönz, B., Ziegenhagen, U., 2000) has been the first step in the transition from traditional textbooks to integrated learning environments.
- This project is with multiple languages and our goal is to apply this project in Arabic language.
- To do this, one needs Arabic Windows XP version, as a tool to write the Arabic code.



## Technical Procedure

- Translate MM\*STAT documents in an Arabic Microsoft Word, and save the files with unicode (file.unicode).
- Direction: Writing in Arabic is from right to left but the direction in HTML appears from left to right, which technical problem for working in Arabic.
- Solution : To solve this problem we use the code *dir = " rtl"* in HTML file at every paragraph and link - Figure - Table and the statistics forms.

- Then the direction of the writing changes from right to left.
- MD\*Book software environment requires LATEX . Another problem is the LATEX Format.
- Arab text does not work with LATEX . We aim to finding a solution to this problem in our project.
- There was a project for Arabtex in LATEX presented by Prof. Klaus Lagally - Universität Stuttgart.

- This project was a good step but it is not enough, because the letter in Arab text in this project was written with English, not with Arabic
- This project did not have the ability to reduce Statistics on an Arabic Website.
- This project uses only PDF, not HTML.
- To summarize, the goal of my work is to create an electronic book in Arabic using LATEX source code and the MD\*Book tool.

## Arabic MM\*STAT Realisation

- Figure 5 presents Arabic MM\*STAT : Cover Page
- Figure 6 presents Arabic MM\*STAT : Content Page
- The user can enter the courses topics via a list of contents and can go to any desired topic or course chapter. By hovering the mouse pointer over a selected course file for a few seconds will appear to identify the course topic.

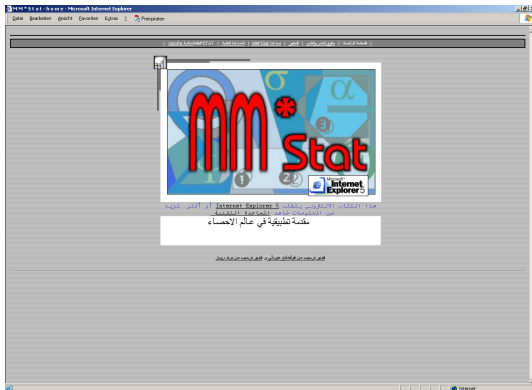


Figure 5: Arabic MM\*Stat: Cover Page

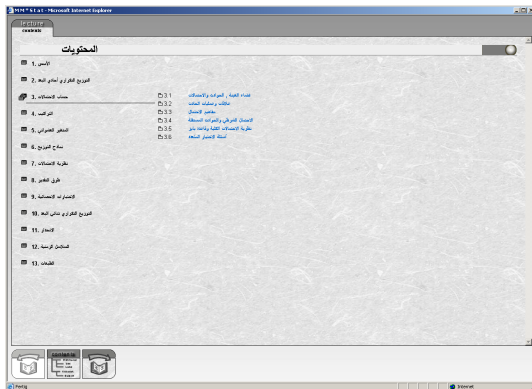


Figure 6: Arabic MM\*Stat: Content Page

- Figure 7 shows a graphical user interface (GUI) of a statistic topic in Arabic MM\*STAT and Figure 8 shows the corresponding English version. ( Both the figures explain the lecture 1.8: Grouping Continuous Data ).
- The Arabic MM\*STAT Course has five components structures; the lecture, information, explained , enhanced and interactive.

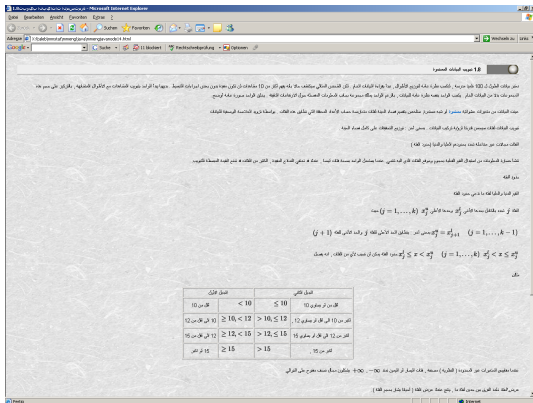


Figure 7: Arabic MM\*Stat: Layer Architecture and Screenshot



**1.8 Grouping Continuous Data**

Consider height data on 100 school boys. In order to gain an overview of the distribution of heights, you start 'reading' the raw data. But the typical person will soon discover that making sense of more than, say, 10 observations without some process of simplification is not useful. Instead, one starts to group individuals with similar heights. By focusing on the size of these groupings rather than on the raw data itself one gains an overview of the data. Even though one has not seen detailed information about exact heights, one has created a clearer overall picture.

Data sampled from continuous random variables can be condensed by partitioning the sample space into mutually exclusive classes. Counting the number of realizations falling into each of these classes is a means of pooling a descriptive summary of the data.

Grouping data into classes can greatly enhance our ability to 'see' the structure of the data, i.e. the distribution of the realizations over the sample space.

Classes are non-overlapping intervals specified by their upper and lower limits (class boundaries).

Loss of information arises from replacing the actual values by the sizes and location of the classes into which they fall. If one uses too few classes, then useful patterns may be concealed. Too many classes may inhibit the positional value of grouping.

**Class boundaries**

The upper and lower values of a class are called class boundaries.

A class  $j$  is fully specified by its lower boundary  $a_j^L$  and upper boundary  $a_j^U$  ( $j = 1, \dots, k$ ), where

$$a_j^L = a_{j-1}^U \quad (j = 1, \dots, k-1), \text{ i.e. upper boundary of the } j\text{th class and lower boundary of the } (j+1)\text{th class coincide}$$

$$a_j^L < a \leq a_j^U \text{ or } a_j^L \leq a < a_j^U \quad (j = 1, \dots, k), \text{ i.e. the class boundary can be attributed to either of the classes it separates.}$$

**Example**

	1st alternative	2nd alternative
less than 10	$< 10$	$\leq 10$ less than or equal to 10
10 to less than 12	$\geq 10, < 12$	$> 10, \leq 12$ greater than 10 to less than or equal to 12
12 to less than 15	$\geq 12, < 15$	$> 12, \leq 15$ greater than 12 to less than or equal to 15
15 or greater	$\geq 15$	$> 15$ greater than 15

Figure 8: English MM\*Stat: Layer Architecture and Screenshot

- Figure 9 presents the different elements of a statistics lecture. This structure called MD\*Booklet.
- Each lecture gives the basic concepts of the general statistical theory, definitions, formulae, mathematical proofs .
- Arabic MM\*Stat provides the students to compute distribution functions, graphics and derive results for statistical tests. These examples have been written in XploRe.

The screenshot displays the MM\*Stat software interface, specifically the 'Lecture 6.5 Poisson Distribution' section. The window title is 'MM\*Stat - Microsoft Internet Explorer'. The content includes:

**6.5 Poisson Distribution**

The Poisson distribution can describe an experiment in which an outcome can be obtained a number of times (for example, accidental deaths).

The random variable  $X$  denotes the outcomes and is **discrete** in nature. This random variable will be described by a **probability density function** referred to as a Poisson distribution with parameter  $\lambda$ .

$$f_{PO}(x; \lambda) = \begin{cases} \frac{\lambda^x e^{-\lambda}}{x!} & \text{for } x = 0, 1, 2, \dots; \lambda > 0 \\ 0 & \text{otherwise} \end{cases}$$

The **distribution function** is:

$$F_{PO}(x; \lambda) = \begin{cases} \sum_{k=0}^x \frac{\lambda^k e^{-\lambda}}{k!} & \text{for } k \geq 0; \lambda > 0 \\ 0 & \text{for } k \leq 0 \end{cases}$$

The **expected value** and **variance** of the Poisson distribution are:

$$E(X) = \lambda \quad \text{Var}(X) = \lambda$$

The table presents the Poisson distribution for different values of  $\lambda$ .

**Properties** of the poisson distribution - **Reproductivity**: Consider two independent variables  $X \sim PO(\lambda_1)$  &  $Y \sim PO(\lambda_2)$ , then the random variable  $Z = X+Y$  is a Poisson distributed with parameter  $\lambda_1 + \lambda_2$ .  $Z \sim PO(\lambda_1 + \lambda_2)$

-Poisson distribution for an arbitrary interval length:

If the number of outcomes in a single interval is poisson distributed, then the number of outcomes in an interval of length  $t$  will also be Poisson distributed with parameter  $\lambda t$ .

$$f_{PO}(t; \lambda \cdot t) = \frac{(\lambda t)^t}{t!} e^{-\lambda t}$$

The interface includes navigation icons (Home, Back, Forward, Stop, Refresh, Print) and a status bar at the bottom showing 'http://www.mstat.com'.

Figure 9: MM\*Stat: with course Poisson Distribution.

## Conclusion

- There is the possibility of creating an e-learning system with Arabic MM\*STAT
- Problems are encountered, both in the context of the Arabic language itself and the technical problems, which is the nonexistence of a possibility to use LATEX code in Arabic.
- Our next goal is to explore possible solutions to the problem of Arabic LATEX, actually the goal has to do with XML!

## References

- AbdulJaleel, N., Larkey, L., *English to Arabic Transliteration for Information Retrieval: A Statistical Approach*, Center for Intelligent Information Retrieval Computer Science, University of Massachusetts.
- Al-Fedaghi, S and H. Al-Sadoun. Morphological compression of Arabic text. *Information Processing Management* 26(2) , 1990.
- Ahmad, T., Mungo, J., *On the Difficulty to Design Arabic E-learning System in Statistics* Humboldt-Universität zu Berlin, 2006.
- Bates, T., *Forum on National strategies for E-learning*, IIEP Virtual Institute, 21 October to 15 November 2002.

- Jemni, M., Bahattab, A., *A collaborative approach for development of arabic courses for E-learning, a case study of Tunisian-Saudi Arabian experience.*
- Hammad, S., AL-ayyoub,A.,Sarie,T., *combining existing E-learning components towards an ivle*, Arab Open University, Amman.
- Härdle,W., Klinke,S., Ziegenhagen,U., *eLearning - A Critical Review*, Humboldt-Universität zu Berlin, January 5, 2006.
- Lagally, K., *ArabTeX : a System for Typesetting Arabic : User Manual Version 3.00*, Universität Stuttgart, 1993.

- Mungo, J., *e-learning/e-teaching An Implementation and Evaluation of a Finance Introductory Course*, Humboldt-Universität zu Berlin, 2004.
- Moukdad, H., Andrew, L., *Information Retrieval from Full-Text Arabic Databases, Can Search Engines Designed for English Do the Job?*, Canada, 2001.
- Müller, M., Rönz, B., Ziegenhagen, U., *The Multimedia Project MM\*Stat for Teaching Statistics, In COMPSTAT. Proceedings in Computational Statistics*, Physica Verlag, (2000).
- Maegaard, B., Choukri, K., Mokbel, C., Yaseen, M., *Language Technology for Arabic*, NEMLAR, Center for Sprogteknologi. University of Copenhagen, July 2005.