

Learning Confidence Intervals with Mobile Devices

<http://dx.doi.org/10.3991/ijim.v7i4.2788>

F.J. Tapia-Moreno, H.A. Villa-Martinez
University of Sonora, Hermosillo, México

Abstract—Mobile learning (m-learning) enhances learning skills in some students. Mobile phones, Tablets, PDAs, Pocket PCs, and Internet can be used jointly in order to encourage and motivate learning wherever and whenever students want to learn. In this work, we show learning objects for teaching and learning inferential statistics using mobile devices. With these learning objects, students can calculate confidence intervals based in either a large or a small data sample obtained from a normal or a non-normal population. These objects have been designed for devices with Android operating system.

Index Terms—m-learning, inferential statistics, learning objects.

I. INTRODUCTION

In April of 2012 we presented a paper about the main results obtained in the project *Statistics-to-Go* [1] in which we developed learning objects for descriptive statistics by means of three Java-ME suites named *Module 1*, *Visual Summary Measures*, and *Skewness Measures*. We wrote the corresponding user manuals. These learning objects are currently used by the students who take the statistics courses offered in the University of Sonora. The success in the use of these learning objects has motivated us to build a module for inferential statistics.

In this paper, we present learning objects for calculate confidence intervals based in either a large or a small sample of data obtained from a normal or a non-normal population.

The remainder of this paper is composed as follows: Section II presents a short review about the state of art in mobile learning. Section III shows details about Android operating system and what can be created on mobile devices which use this operating system. Section IV presents basic confidence intervals theory. Section V shows the latest advances in the project *Statistics-to-Go*. Finally, Section VI presents our conclusions and future work.

II. MOBILE LEARNING

Mobile learning (also known as m-learning) is the acquisition of knowledge by means of some mobile device, for example a cellular phone or a tablet computer [2]. m-learning is extremely effective in classroom environments where modern technology may not be available to all students, or in schools where there are not sufficient computational resources for all students. In such situations, using a mobile device for learning, a device that can be shared among a group of students, is a great

way to ensure that pupils are still able to benefit from the diverse opportunities presented by educational technology.

Some of the reported uses of m-learning are: to enhance the user experience in visitors of an art museum, to take advantage of the learning context [3], to allow students to participate on an academic discussion forum [4], in training for personnel in disaster or emergency situations [5], to suggest didactic methods that are adequate to a specific teaching situation [6], and in the learning of science [7], languages [8], and medicine [9].

The differences between m-learning and other types of learning can be studied by considering both the technology involved and the educational experience. Regarding technology, m-learning differs in the use of portable equipment that allows students to access learning objects anytime and anywhere. Regarding the educational experience, Traxler [10] compares m-learning and traditional e-learning using keywords. This way, m-learning is “personal”, “spontaneous”, “opportunistic”, “informal”, “pervasive”, “private”, “context-aware”, “bite-sized”, and “portable”, whereas e-learning is “structured”, “media-rich”, “broadband”, “interactive”, “intelligent”, and “usable”. The same author notes that some of these distinctions can disappear as mobile technology advance, but properties as informality, mobility, and context will remain.

III. ANDROID OPERATING SYSTEM

Android is an operating system designed primarily for mobile devices such as cellular phones and tablet computers. Initially developed by Android, Inc., which Google bought in 2005, Android was unveiled in 2007 along with the founding of the Open Handset Alliance, a consortium of hardware, software, and telecommunication companies devoted to advancing open standards for mobile devices. The first Android-powered phone was sold in October 2008 [11].

Android-based cellular phones lead the market. According to Tech-Thoughts [12], Android has a share of almost 70% in the global smartphone market, while iOS, made by Apple, accounts to about 14%. In fact, Android has the largest share in every developed country, except in the United States and Japan where iOS is the dominant mobile operating system [13, 14].

In the tablet world the story is similar. According to BGR [15], Android worldwide market share grew from 39.8% in 2011 to 42.7% in 2012, while iOS share diminished from 56.3% to 53.8%.

IV. CONFIDENCE INTERVALS

In statistical inference, one wishes to estimate population parameters (for example average, variance, standard deviation, proportions, and differences between means and/or proportions) using observed sample data. A confidence interval gives an estimated range of values which is likely to include the unknown population parameter. This estimated range is calculated from a given set of sample data. Each confidence interval has associated a measure of the certainty that the interval confidence contains the population's parameter. This measure is called *confidence level* and is indicated by $1-\alpha$, where α , a number between 0 and 1, is called the significance grade. The most used confidence levels are 90%, 95%, and 99%. Each one of these numbers represents the probability that the confidence interval contains the real valor of the population parameter. Figure 1 shows a confidence interval for the population's average μ based in an observed sample data. $\mu_{\bar{x}}$ is the sample's average.

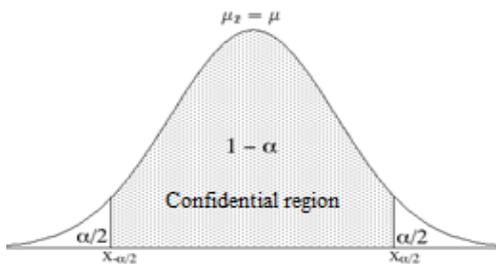


Figure 1.

According to Triola [16], from a population sample a confidence interval can be constructed for the population's mean using the normal distribution when: 1) the size of the sample is larger than 30 ($n \geq 30$) (this is due to the central limit theorem), or 2) when the sample size is less than 30 ($n < 30$), but the population has a normal distribution and its standard deviation is known. For this case, the confidence interval is given by Equation (1).

$$\bar{X} \pm z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \quad (1)$$

Note that if the population's standard deviation is unknown, then Equation (2) should be used for determining the confidence interval.

$$\bar{X} \pm z_{1-\alpha/2} \cdot \frac{S}{\sqrt{n}} \quad (2)$$

where \bar{X} is the sample's mean, S is the sample's standard deviation, n is the size of the sample, and $z_{1-\alpha/2}$ is the critical value of the normal standardized distribution [16].

If the size of the sample is less than 30 ($n < 30$), the population has normal distribution, and the standard deviation is unknown, then the Student's distribution is preferred for constructing the confidence interval for population's mean [2]. In this case, the confidence interval is given by Equation (3)

$$\bar{X} \pm t_{n-1,1-\alpha/2} \cdot \frac{S}{\sqrt{n}} \quad (3)$$

where \bar{X} is the sample's mean, S is the sample's standard deviation, n is the size of the sample, $n - 1$ are the degrees of freedom, and $t_{n-1,1-\alpha/2}$ is the corresponding critical value for the confidence grade $(1 - \alpha/2)$ of the Student's distribution.

When the size of the sample is less than 30 ($n < 30$) and the population has a non-normal distribution, then the Chebyshev's Theorem can be used to build a confidence interval for the population's mean [16]. Equations (4) and (5) show the confidence interval for this case when the population's standard deviation is known and unknown, respectively.

$$\bar{X} \pm k \cdot \frac{\sigma}{\sqrt{n}} \quad (4)$$

$$\bar{X} \pm k \cdot \frac{S}{\sqrt{n}} \quad (5)$$

where $k = \sqrt{1/\alpha}$.

In a similar way, confidence intervals for a population's proportion can be build using the normal distribution as an approximation to the binomial distribution if $n \geq 30$, and either $n \cdot p \geq 5$ or $n \cdot (1 - p) \geq 5$. Here, n is the sample size and p is the probability of success in the experiment [16]. The confidence interval for this case is presented in Equation (6).

$$\bar{X} \pm z_{1-\alpha/2} \cdot S_{\bar{p}} \quad (6)$$

where $S_{\bar{p}} = \sqrt{\frac{\bar{p} \cdot (1-\bar{p})}{n}}$ or $S_{\bar{p}} = \sqrt{\frac{\bar{p} \cdot (1-\bar{p})}{n} \cdot \frac{N-n}{N-1}}$ if $n < 0.05 \cdot N$. Here N is the population size, and \bar{p} is the probability of success in the experiment.

Similarly, confidence intervals can be constructed for the difference between the means of two populations and for the difference between the proportions of two populations. The confidence interval for each one of these cases is shown in Equations (7), (8), (9), (10), (11), and (12).

$$(\bar{X}_1 - \bar{X}_2) \pm z_{1-\alpha/2} \cdot \sigma_{\bar{X}_1 - \bar{X}_2} \quad (7)$$

$$(\bar{X}_1 - \bar{X}_2) \pm z_{1-\alpha/2} \cdot S_{\bar{X}_1 - \bar{X}_2} \quad (8)$$

$$(\bar{X}_1 - \bar{X}_2) \pm t_{(n_1+n_2-2),1-\alpha/2} \cdot S_{\bar{X}_1 - \bar{X}_2} \quad (9)$$

$$(\bar{X}_1 - \bar{X}_2) \pm k \cdot \sigma_{\bar{X}_1 - \bar{X}_2} \quad (10)$$

$$(\bar{X}_1 - \bar{X}_2) \pm k \cdot S_{\bar{X}_1 - \bar{X}_2} \quad (11)$$

$$(\bar{p}_1 - \bar{p}_2) \pm z_{1-\alpha/2} \cdot S_{\bar{p}_1 - \bar{p}_2} \quad (12)$$

where $\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$, $S_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$, $S_{\bar{p}_1 - \bar{p}_2} = \sqrt{\frac{\bar{p}_1 \cdot (1-\bar{p}_1)}{n_1} + \frac{\bar{p}_2 \cdot (1-\bar{p}_2)}{n_2}}$, and \bar{X}_1 , \bar{X}_2 , σ_1 , σ_2 , S_1 , S_2 , \bar{p}_1 , \bar{p}_2 , n_1 , and n_2 are respectively, the means, standard deviations of the populations, standard deviations

of the samples, proportions of the samples, and the samples sizes.

If a sample is taken from a normal distribution, confidence intervals can be built for both population's variance and standard deviation using the chi square distribution [17]. Both confidence intervals are given in Equations (13) and (14), respectively.

$$\frac{(n-1) \cdot S^2}{\chi_{n-1, \text{upper}}^2} \leq \sigma^2 \leq \frac{(n-1) \cdot S^2}{\chi_{n-1, \text{lower}}^2} \quad (13)$$

$$\sqrt{\frac{(n-1) \cdot S^2}{\chi_{n-1, \text{upper}}^2}} \leq \sigma \leq \sqrt{\frac{(n-1) \cdot S^2}{\chi_{n-1, \text{lower}}^2}} \quad (14)$$

where the upper and lower sub-indices identify the percentiles points of the chi-square distribution that was used for constructing the confidence interval.

Finally, we can build confidence interval for the ratio of variances of two normally distributed populations using the F distribution [16] employing Equation (15).

$$\frac{S_1^2/S_2^2}{F_{1-\alpha/2}} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2/S_2^2}{F_{\alpha/2}} \quad (15)$$

One last remark is that a person who wants to calculate confidence intervals needs statistics tables, like those in [17], in order to obtain the proportion of the area under the normal curve, the Student t , the chi-square, and the F factor. On the other hand, our system obviates this requirement.

IV. RESULTS

In this second stage of our project, we have developed learning and teaching objects for Inferential Statistics. Each one of the objects calculates a confidence interval for a distinct situation. The first object computes confidence intervals for the population's mean. The application uses Equations (1) or Equation (2) depending if the value of the population's standard deviation is known or not. Figure 2 shows the output of this object for a sample with size $n = 57$, a mean of 7.5, a standard deviation of 2.1, and a confidence level of 95%. That is, $\alpha = 0.05$.

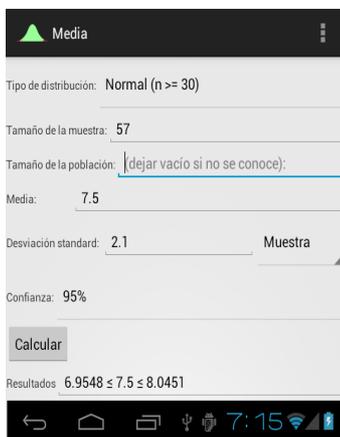


Figure 2.

The second object computes confidence intervals for the population's average when the size of the sample is less than 30 ($n < 30$), the population has a normal distribution, and the standard deviation is unknown. For this case, the Student's distribution is the preferred way for constructing the confidence interval for the population's mean. In this case, the system uses Equation (3) in order to calculate it. Figure 3 shows the output of this object for a sample with size $n = 21$, a mean of 7.5, a standard deviation of 2.1, and a confidence level of 95%.

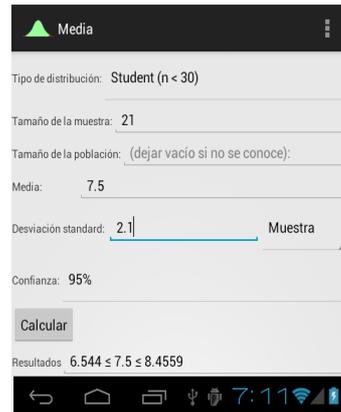


Figure 3.

The third object computes confidence intervals for the population's average when the population has a non-normal distribution. The Chebyshev's theorem is used to build this confidence interval. The application uses Equation (4) or Equation (5) depending if the value of the population's standard deviation is known or not. Figure 4 shows the output of this object for a sample with $n = 57$, a mean of 7.5, a standard deviation of 2.1, and a confidence level of 95%.



Figure 4.

The fourth object calculates confidence intervals for a population's proportion using Equation (6). This application can be used only when both conditions mentioned in section IV are satisfied. Figure 5 presents the output of this object for a population with size $N = 400$, and a sample of size $n = 57$ with a proportion of $\hat{p} = 0.58$ and a confidence level of 95%.



Figure 5.

The fifth object computes confidence intervals for the difference between the means of two populations using Equation (7) or Equation (8) depending if both population's standard deviations are known or not. Figure 6 shows the output of this object for two samples with sizes $n_1 = 71$ and $n_2 = 55$, sample's means of $\bar{X}_1 = 17.58$ and $\bar{X}_2 = 11.6$, and sample's standard deviations of $s_1 = 4.8$ and $s_2 = 1.256$, with confidence level of 95%.



Figure 6.

The sixth object computes confidence intervals for the difference between the means of two populations when both two samples are little ($n_1, n_2 < 30$), and both populations have approximately normal distributions. Equation (9) is utilized by the object to calculate these intervals.

The seventh object calculates confidence intervals for the difference between the means of two populations when these populations have non-normal distributions. The object uses Equation (10) or Equation (11) depending if both population's standard deviations are known or unknown. The input data, and the output is similar to the output of the fifth object, shown in Figure 6, with the difference that the seventh object uses Student's distribution in order to estimate the confidence interval.

The eighth object estimates confidence intervals for the difference between the proportions of two populations when the conditions mentioned in the section IV are satisfied. The object uses Equation (12) to calculate the confidence intervals. Figure 7 shows the output for two samples with sizes $n_1 = 45$, and $n_2 = 65$ with proportions of $\hat{p}_1 = 0.70$, and $\hat{p}_2 = 0.55$, with a confidence level of 95%.

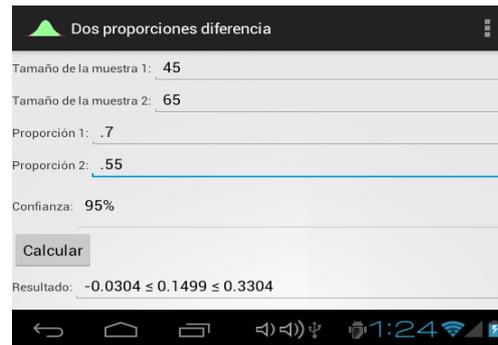


Figure 7.

The ninth object computes confidence intervals for the population's variance and standard deviation using the chi square distribution. The object uses Equation (13) or Equation (14) to estimate these confidence intervals. Figure 8 shows the result output for a sample with size $n = 70$, a standard deviation of 7.5, and a confidence level of 95%.



Figure 8.

The tenth object calculates confidence intervals for the ratio of the variances of two samples employing Equation (15). Figure 10 shows the output of this object for two samples with sizes $n_1 = 54$ and $n_2 = 84$, variances of $s_1 = 8.6$ and $s_2 = 5.25$, with a confidence level of 95%.

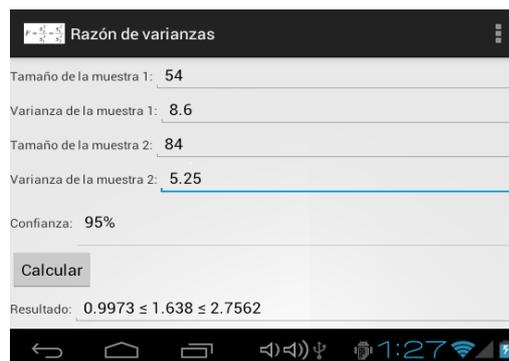


Figure 9.

VI. CONCLUSIONS

We have presented the main achievements obtained after completing the second stage of the project *Statistics-to-Go*. We have designed learning objects that enhance learning of statistical concepts exclusively using a cellular phone and taking advantage of the ubiquity of personal technology for educational purposes. We are exploiting the possibilities of mobile phones for teaching and learning statistics helping by the way those students who do not have possibilities for buying a computer or paying

for Internet service. These learning objects cover most of the syllabus of the introductory inferential statistics course offered in the first semester at the University of Sonora.

In the third stage of this project we will design learning and teaching objects to perform **statistical hypothesis tests, which is** a method of making decisions using data from a scientific study, and that the students learn in a second course of inferential statistics.

Also we will test the efficacy of these objects on the students' mobile devices and we will gather their opinion about the usability of our applications.

REFERENCES

- [1] Tapia et al. *Elaboration of Statistics Learning Objects for Mobile Devices*. iJIM, 6, 2, p. 51-54. April 2012
- [2] Molnár, Györgi, J.: *New Learning Spaces? M-Learning's, in Particular the iPad's Potentials in Education*. iJIM, 7, 1, p. 56-60. January 2013.
- [3] Bhaskar, N. Uday.; P. Govindarajulu *Advanced and Effective Learning in Context Aware and Adaptive Mobile Learning Scenarios*. iJIM, 4, 1, p. 9-13. January 2010.
- [4] Motiwalla, Luvai F. *Mobile learning: A framework and evaluation*. Computers & Education 49, p. 581-596. 2005. <http://dx.doi.org/10.1016/j.compedu.2005.10.011>
- [5] Colley, Jo; Claire Bradley; Geoff Stead; Jessica Wakelin. *Global MedAid – evolution of an mlearning app for international work-based learners*. mLearn 2012. Helsinki, Finland. 2012.
- [6] Möller, Andreas; Stefan Diewald; Luis Roalter; Barbara Beege; Matthias Kranz. *MobiDics: Cooperative Mobile E-Learning for Teachers*. mLearn 2012. Helsinki, Finland. 2012.
- [7] Sohaib, Ahmed; David Parsons. *Evaluating 'ThinkLearn': A Mobile Science Inquiry Based Learning Application in Practice*. mLearn 2012. Helsinki, Finland. 2012.
- [8] Mmaki, Jantjies; Mike Joy. *Multilingual mobile learning – A case study of four South African high schools*. mLearn 2012. Helsinki, Finland. 2012.
- [9] Hardy, Ernst; John Harrison; David Griffin. *Iterative design and delivery of high impact, multiple platform, scenario-based interactive mobile learning activities in the health sciences*. mLearn 2012. Helsinki, Finland. 2012.
- [10] Traxler, J.: *Defining, Discussing, and Evaluating Mobile Learning: The moving finger writes and having writ....* International Review of Research in Open and Distance Learning, 8, 2, 1-12. 2007.
- [11] Kerfs, Jeremy: *Beginning Android Tablet Games Programming*. p. 1-2. APress. 2011.
- [12] <http://www.tech-thoughts.net/2012/12/smartphone-market-share-trends-by-country.html#UQRSaL-mz2Y>. Last accessed Jan 26, 2013.
- [13] http://www.kantarworldpanel.com/dwl.php?sn=news_downloads&id=113. Last accessed Jan 26, 2013.
- [14] <http://www.kantarworldpanel.com/global/News/news-articles/Windows-sees-strong-European-growth>. Last accessed Jan 26, 2013.
- [15] <http://bgr.com/2012/12/05/tablet-market-share-2012/>. Last accessed Jan 26, 2013.
- [16] Triola, Mario F. *Elementary Statistics with Multimedia Study/Guide*, 10th Edition. Addison-Wesley Longman. 2007.
- [17] Neave, Henry. *Statistics Tables for Mathematicians, Engineers, Economists. and the Behavioural and Management Sciences*. College Audience. 2011.

AUTHORS

F. J. Tapia-Moreno is with the Department of Mathematics, University of Sonora, Hermosillo, Sonora 83000 Mexico (e-mail: ftapia@gauss.mat.uson.mx).

H. A. Villa-Martínez is with the Department of Mathematics, University of Sonora, Hermosillo, Sonora 83000 Mexico (e-mail: hvilla@gauss.mat.uson.mx).

This work was supported in part by a PIFI fund from the Mexican Federal Government. Submitted 18 May 2013. Published as resubmitted by the authors on 08 September 2013.