USING THE TOULMIN MODEL OF ARGUMENTATION TO VALIDATE STUDENTS' INFERENTIAL REASONING

Ana Gómez-Blancarte[1] and María Guadalupe Tobías-Lara
[1]Instituto Politécnico Nacional, CICATA-Legaria
[2] Tecnológico de Monterrey
algomezb@ipn.mx

Assuming that formal ideas of statistical inference must be constructed progressively to gradually increase their level of formalization, exploring and promoting the development of informal inferential reasoning has been a focus of research on statistics education for over a decade. How to evaluate the development of students' inferential reasoning, and how they demonstrate such reasoning are two important, recent research issues. Given that inferential reasoning –whether formal or informal– may be linked to the activity of argumentation, we claim that students' inferential reasoning should be evaluated in terms of the strengths or weaknesses of their arguments. Drawing on the Toulmin model of argumentation, this paper proposes two levels of analysis to inform us about the space of reasons necessary to validate students' inferential reasoning.

INTRODUCTION

Interest in studying informal inferential reasoning emerged because this type of reasoning may improve our understanding of concepts in statistical inference (e.g., Garfield & Ben-Zvi, 2008; Zieffler, Garfield, delMas & Reading, 2008). Research in this area distinguishes between informal and formal inferential reasoning (IIR and FIR, respectively). The latter refers to a type of reasoning that utilizes formal methods of statistical inference. This distinction has raised questions about the cognitive development of informal ideas of statistical inference towards more formal ones (Reading 2007; Zieffler et al., 2008; Garfield & Zieffler, 2011; Makar & Rubin, 2014). However, research shows that both types of reasoning can be used by professional statisticians (Garfield, Le, Zieffler & Ben-Zvi, 2015), and that their use depends on the type of task or problem at hand (Pratt & Ainley, 2008). Also playing an important role is the context of the problem on which inferences are made, since the techniques considered informal in statistics may be seen as formal in certain fields of science (Bakker, Kent, Detty, Noss & Hoyles, 2008).

It is important to understand from the outset that inferential reasoning –formal or informal– is complex, because in addition to integrating various statistical concepts (Chance, delMas & Garfield, 2004), it requires information on the practical problem; for example, on the techniques, norms and proposals used in the field of science (see Bakker et al., 2008) to which the phenomenon under study pertains. Knowledge of context is also necessary because it provides a series of scientific reasons that may support the inference. It is in this sense that we consider that inferential reasoning is comparable to argumentation, as there is a need to construct persuasive arguments based on both data analysis (Ben-Zvi, 2006) and scientific reasons related to the problem. In this article, we propose using the Toulmin (2003) model of argumentation to analyze the validity of students' inferential reasoning. Our proposal suggests two levels of analysis that integrate statistical and scientific reasons.

THE LINK BETWEEN TOULMIN ELEMENTS OF ARGUMENTATION AND THE COMPONENTS OF INFERENTIAL REASONING

The activity known as argumentation has been proposed as a method for teaching statistical inference (Garfield & Ben-Zvi, 2008). Some researchers have used elements of the Toulmin (2003) model to understand the logic of hypothesis statistical testing (LeMire, 2010) and analyze students' justifications as the basis of their conclusions (Weber, Maher, Powell & Stohl, 2008; Osana, Leath & Thomson, 2004). The model, or scheme, of argumentation proposed by Toulmin to justify the process of defending a certain conclusion, consists of six elements –Claim, Data, Warrant, Backing, Qualifier and Rebuttal– that can be illustrated in any field of practical reasoning, though the argumentation procedures developed in different fields of knowledge or scientific disciplines may vary (Toulmin, Rieke & Janik, 1984). The connections among Claim (C), Data (D), Warrant (W) and Backing (B) give an argument its solidity, while the presence of Qualifiers (Q) and Rebuttals (R) demonstrates the strength

In M. A. Sorto, A. White, & L. Guyot (Eds.), *Looking back, looking forward. Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10, July, 2018), Kyoto, Japan.* Voorburg, The Netherlands: International Statistical Institute. iase-web.org

of the interconnections among these elements (Figure 1). The strength of an argument involves the reliability of its conclusions, inferences or claims, which are practically reliable but not formally irrefutable. Because inferential reasoning is a "probabilistic generalization from data" (Makar & Rubin, 2009, p. 85), a Q would reflect "recognition that judgments based on sample data are inherently uncertain" (Konold et al., 2011, p. 68).

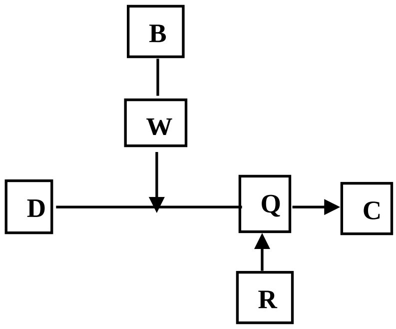


Figure 1. Toulmin model of argumentation

Based on the components of informal inferential reasoning proposed by Zieffler et al. (2008), and those of informal statistical inference put forth by Makar and Rubin (2009), we elaborate an analogy between these components and some of the elements of Toulmin's model of argumentation (see Table 1) to construct our frame of reference. Like Garfield et al. (2015), we consider that these components are not exclusive to IIR, because they are also present in FIR. Hence, we refer to them as characteristic of both formal and informal inferential reasoning. We further understand that IIR emphasizes a form of data analysis that uses non-formal methods (Zieffler et al., 2008) and identifies patterns in the data (Makar & Rubin, 2009); while in FIR data analysis includes formal methods.

Table 1. Analogy between the elements of the Toulmin model of argumentation and inferential reasoning

Elements	Relation to inferential reasoning
C	According to Toulmin et al. (1984), Claim refers to the element the merits of which one seeks to establish. It is the first element that must be identified in an argument. In inferential reasoning, Claim involves "predictions about populations based on samples" (Zieffler et al., 2008, p. 45), while for Makar and Rubin (2009, p. 85), it is a "generalization, including predictions, parameter estimates, and conclusions, that extend beyond describing the given data".
D	Data are explicit facts that illustrate how a conclusion was reached (Hitchcock, 2017). In inferential reasoning, the generalization or inference is based on evidence in the form of statistical data (Makar & Rubin, 2009) and "[d]raw[s] on, utilize[es], and integrat[es] prior knowledge (e.g., formal knowledge about foundational concepts, […]; informal knowledge about inference […]), to the extent that this knowledge is available" (Zieffler et al., 2008, p. 45).
W	Warrants are justifications of inferences; they may be rules or general principles that act as an authorizing bridge between conclusion and evidence (D), necessary to demonstrate that the transit from data to affirmation is appropriate and legitimate (Toulmin, 2003). They are general because, "one needs to be able to justify the warrant independently of the particular case to which it is applied, and that such an independent justification can only come if it makes no reference to the particular case; that is, is general" (Hitchcock, 2017, p. 83). Hitchcock emphasizes that the functional role of the Warrant is what distinguishes the evidence, not the implicit or explicit form in which this is expressed. In inferential reasoning, the determination of the Warrants depends largely on the context of the problem and what has been inferred about it. If the inference or generalization concerns the logic of an inferential concept (e.g., LeMire, 2010), then the Warrants may be statistical principles; if the inference involves a problem in a certain discipline, then the

	general principles may pertain to the field of that discipline, since the statistical results must be interpreted in terms of the context.
B	The term backing refers to a corpus of scientific theory or statute that gives solidity to the Warrants. Toulmin (2003) distinguishes between Warrants as hypothetical propositions and Backing as categorical propositions that need not be explicit. Toulmin clarifies that Warrants can be considered non-challenging, so their support is understood, adding that Backing is especially necessary when there are various possible ways to connect the step between evidence and conclusion. Thus, Backing supports conclusions that are conflictive in terms of the form of reasoning, but not with respect to the truth or falsity of the evidence. In inferential reasoning, Backing can be deemed a discourse that sustains the Warrant used. Like the Warrant, this discourse can come from either the field of statistics or the context of the problem.
Q/R	In statistical inference –which involves working with samples– generalizations cannot be expressed in absolute terms (Makar & Rubin, 2009), but require using the language of probability to express uncertain knowledge and, depending on the degree of formality required in the inference, as a numerical measure. In inferential reasoning, the use of probabilistic language as a modal qualifier (Q) emphasizes the importance of uncertainty when drawing inferences, and of the conditions (R) under which the statistical principles and context (W) may be inappropriate.

APPLICATION OF THE TOULMIN MODEL OF ARGUMENTATION TO VALIDATE INFERENTIAL REASONING

To exemplify our proposal, we analyzed the reasoning of a group of engineering students in an introductory statistics course who were asked to resolve the following situation:

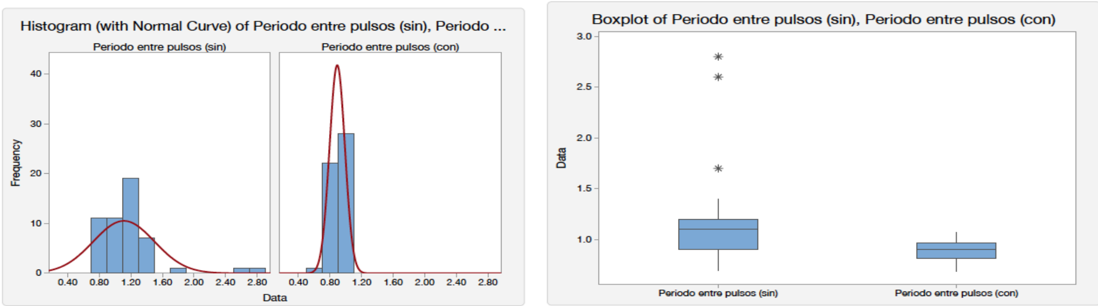
> A group of experimental doctors conducted tests of the utility of a new pacemaker to improve the functioning coronary patients' hearts. They are undecided as to whether they should begin to market the device on a broad scale, and so decide to conduct a statistical analysis to help them decide. They proceed to gather information on a sample of 102 heart patients, half of whom used the pacemaker, while the other half did not. The data from their sample was presented in *Excel*, including measures of the variable time between pulses in both groups of patients; that is, with, and without, the pacemaker. Based on an analysis of these data, is it possible to generalize that the pacemaker exerted an effect that improved the functioning of the hearts of every patient who used it?

A total of 13 teams worked on this situation, but this article examines only the response of team 1. In addition to resolving the problem, students were asked to investigate the broader context; for example, the function of a pacemaker, the variables used to measure heart function, and the normal physiological values for those variables. It is important to note that in a real context, pacemakers are programmed to produce a frequency within normal values; that is, 60-100 beats per minute. In the situation presented, the variable time between pulses measured the seconds per beat, expecting that this variable would fall within the range of 0.60-1 second per beat. This information is relevant to evaluating whether a pacemaker exerts an effect on heart functioning; that is, if it makes it beat at a pre-programmed frequency.

Like those in the other teams, the students in team 1 utilized Minitab software for data analysis. They analyzed the distribution of the variable using histograms and boxplots. Also, they obtained the measures of center and dispersion (standard deviation) for both experimental groups, and determined their conclusion based on these data (see Figure 2).

First level of analysis

In the first level of analysis, we evaluated the presence and connection of the three components of inferential reasoning (see Figure 3): 1) the inference (C); 2) should be based on a statistical analysis of the data (D); and 3) should include a language of uncertainty (Q), together with possible conditions under which the inference may not be appropriate (R).



(without pacemaker) (with pacemaker) (without pacemaker) (with pacemaker)

It is clear in these histograms and boxplots that the data on the period between pulses with the pacemaker are less disperse. In the data of patients with pacemakers, symmetry is visible in the distribution of results. In the data from the patients without pacemakers, the distribution of results is biased to the right, and there are three atypical or unusual values.

Descriptive Statistics: Periodo entre pulsos (sin), Periodo entre pulsos (con)

Statistics

Variable	N	N*	Mean	SE Mean	StDev	CoefVar	Minimum	Q1	Median	Q3	Maximum	Skewness
Periodo entre pulsos (sin)	51	0	1.11176	0.05421	0.38712	34.82	0.70000	0.90000	1.10000	1.20000	2.80000	2.80509
Periodo entre pulsos (con)	51	0	0.89118	0.01363	0.09736	10.92	0.69000	0.82000	0.90000	0.97000	1.07000	-0.21734

The mean (representative value) for the patients without pacemakers is 1.1, while for those with pacemakers it is 0.89118. The standard deviation of the data for the patients with, and without, pacemakers is 0.38712, and 0.09736, respectively.

Based on these results, it can be assumed that the patients with pacemakers have better control over the period between pulses, so having a functioning pacemaker does have a positive effect. It is important to note that these results could be affected by external factors such as age, gender and medical conditions.

Figure 2. Team 1's statistical analysis

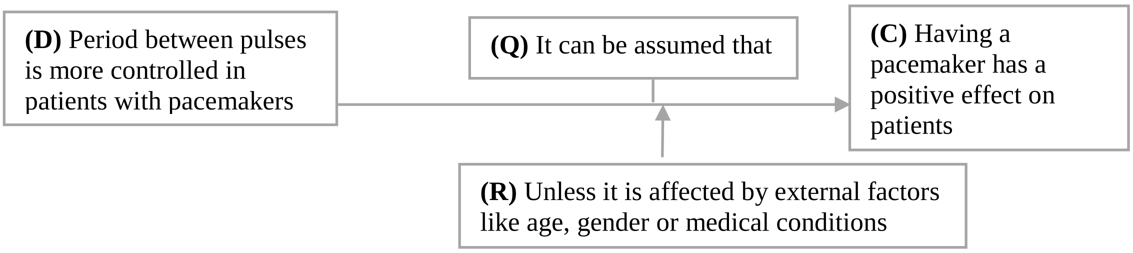


Figure 3. Example of first-level analysis (based on team 1's statistical analysis)

The presence of these elements and their interconnection allows us to determine whether the students used inferential reasoning. Of the 13 teams that resolved the situation, only team 1 employed a language of uncertainty (Q) and mentioned some factors that could invalidate the inference (R). At this first-level analysis, the validity of the statistical analysis, the language of uncertainty, possible rebuttals, and the inference itself, can all be questioned. In this case, the team's reasons are based on the idea of variability, which was perceived through an informal focus; namely, observing visual characteristics in the histograms and boxplots (e.g., center, form, atypical data), and on the numerical values of the measures for center and standard deviation. Those observations allowed the team to identify that "the patients with pacemakers had better control of the period between pulses". By assuming that this variable had less variability in the sample of patients with pacemakers, the team reached the conclusion, with some degree of uncertainty, that "having a pacemaker does provide some effectiveness". But, is the statistical evidence sufficient to assume that the pacemaker is effective? Is the lower variability in the sample of patients with pacemakers a sufficient, solid and reliable reason? Toulmin et al. (1984) point out that knowing the evidence on which the conclusion is based is "only the first step toward judging its solidity and reliability" (p. 26). The next point to be evaluated is whether the evidence provides legitimate support for the conclusion.

Second level of analysis

At the second level of analysis, we evaluated whether the evidence provided information relevant to the conclusion. To this end, it is necessary to inquire into the Warrants; that is, how did the students arrive at their conclusion based on a certain body of evidence?, and what general principle authorized the transit from evidence to conclusion? According to Toulmin et al. (1984), the type of Warrants "will depend on the kind of claim under discussion" (p. 26); that is, on the effectiveness of the pacemaker.

Medical norms establish that heart rate should be between 60 and 100 beats per minute (0.6-1 second per beat). The statistical data show that the period between pulses in the patients with pacemakers was around 0.89118 ± 0.09736 seconds per beat; while for those without pacemakers, it was around 1.11176 ± 0.38712 seconds per beat. Given that the period between the pulses of the patients with pacemakers is within the aforementioned physiological values (W), the effect of the device is justified. These values were determined in a community external to the statistical group; in this case, doctors who belong to cardiology associations (B). Figure 4 shows the connections among the six elements of Toulmin's model of argumentation.

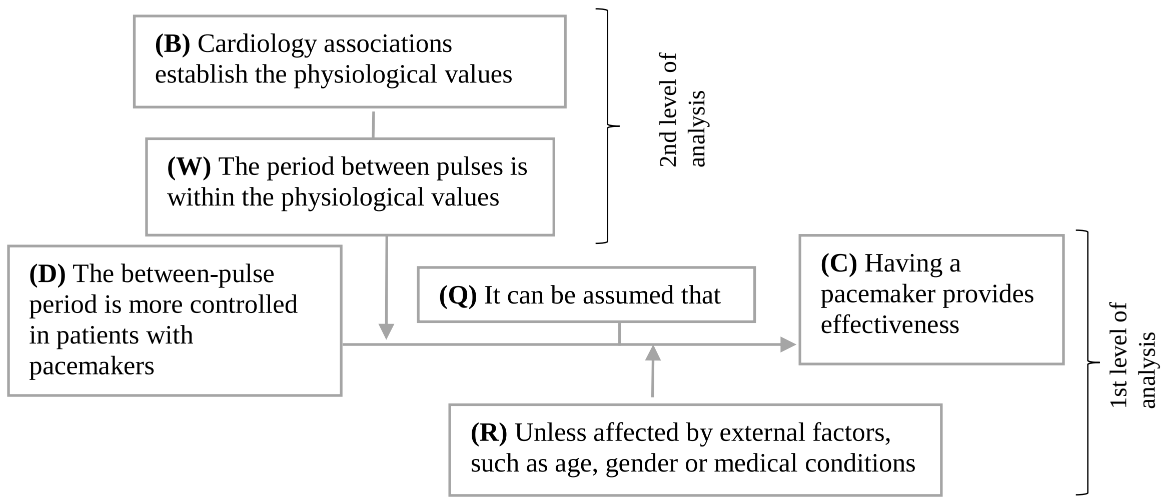


Figure 4. Example of the first and second levels of analysis

The Warrants and Backing were not made explicit by the students; rather, we show them here to exemplify our proposal. When students were asked about the Warrant, they only repeated the content of their statistical analysis in an attempt to justify how they had observed the variability.

CONCLUSION

Here, we have presented a proposal for analyzing the validity of students' inferential reasoning by applying Toulmin's model of argumentation. Based on the idea that the validity of an argument is field-dependent (Toulmin, 2003), we propose two levels of analysis. The first pertains more closely to the field of statistics, since it validates whether or not inferential reasoning is involved (i.e., if it includes all necessary components), and if the statistical analysis of the data is highly acceptable. For example, in the situation of the pacemakers, one could ask if the sample was random, about the characteristics of the patients selected that make it possible to compile a profile of the population to which the generalization is applied, or about the conditions of study design (e.g., placing the pacemakers and their effect on the patient). The second level of analysis corresponds more closely to the context or disciplinary field in which the problem is posited, since expert knowledge is required that must come from a context in order to judge whether the inference may be valid or invalid in light of the statistical results.

Exemplifying our proposal through the situation involving the pacemakers, limits us to one sole type of statistical inference. Nonetheless, we believe that it is necessary to consider that the validity of inferential reasoning resides not only in the students' statistical knowledge (whether formal or informal), but also in their ability to construct arguments based on reasons that emerge from the field to which the inferential situation pertains. In this sense, the proposal may provide a way to promote statistical knowledge and knowledge of context “as an organic part of a space of reasons” (Bakker et al., 2008, p. 142).

REFERENCES

Bakker, A., Kent, P., Derry, J., Noss, R., & Hoyles, C. (2008). Statistical inference at work: Statistical process control as an example. *Statistics Education Research Journal*, *7*(2), 130-145.

Ben-Zvi, D. (2006). Scaffolding students’ informal inference and argumentation. In A Rossman & B. Chance (Eds.), *Proceedings of the Seventh International Conference on Teaching Statistics.* Salvador, Brazil: International Association for Statistical Education.

Chance, B., delMas, R. & Garfield, J. (2004). Reasoning about sampling distributions. En D. Ben-Zvi, & J. Garfield (Eds.), *The Challenge of Developing Statistical Literacy, Reasoning and Thinking* (pp. 295-323). Minnesota, USA: Springer Science.

Garfield, J., Le, L., Zieffler, A., & Ben-Zvi, D. (2015). Developing students’ reasoning about samples and sampling variability as a path to expert statistical thinking. *Educational Studies in Mathematics*, *88* (3), 327-342. doi: 10.1007/s10649-014-9541-7

Garfield, J., & Zieffler, A. (2011). Discussion on the Paper by Wild, Pfannkuch, Regan & Horton: Towards more accessible conceptions of statistical inference. *Journal of the Royal Statistical Society*, *174*(2), 247-295.

Garfield, J., & Ben-Zvi, D. (2008). Learning to Reason about Statistical Inference. In J. Garfield & D. Ben-Zvi (Eds.), *Developing students’ statistical reasoning. Connecting Research and Teaching Practice* (pp. 351-389). Minnesota, USA: Springer Science.

Hitchcock, D. (2017). *On Reasoning and Argument: Essays in Informal Logic and on Critical Thinking* (1st ed.). Cham, Switzerland: Springer.

Konold, C., Madden, S., Pollatsek, A., Pfannkuch, M., Wild, C., Ziedins, I., Finzer, W., Horton, N., & Kazak, S. (2011). Conceptual challenges in coordination theoretical and data-centered estimates of probability. *Mathematical Thinking and Learning*, *13*(1-2), 68-86. doi: 10.1080/10986065.2011.538299

Lemire, S. D. (2010). An argument framework for the application of null hypothesis statistical testing in support of research. *Journal of Statistics Education*, *18*(2), 1-23.

Makar, K. & Rubin, A. (2014). Informal statistical inference revisited. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Proceedings of the 9th International Conference on Teaching Statistics*. Flagstaff, Arizona, USA: International Statistical Institute.

Makar, K., & Rubin, A. (2009). A framework for thinking about informal statistical inference. *Statistics Education Research Journal*, *8*(1), 82-105.

Osana, H. P., Leath, E. P., & Thompson, S. E. (2004). Improving evidential argumentation through statistical sampling: Evaluating the effects of a classroom intervention for at-risk 7th-graders. *Journal of Mathematical Behavior*, *23*(3), 351–370.

Pratt, D., & Ainley, J. (2008). Introducing the special issue on informal inferential reasoning. *Statistics Education Research Journal*, *7*(2), 3–4.

Reading, C. (2007). Cognitive development of reasoning about inference. Discussant reaction presented at the *Fifth International Research Forum on Statistical Reasoning, Thinking, and Literacy*. Warwick, England: University of Warwick.

Toulmin, S. (2003). *The Uses of Argument* (updated edition). Cambridge, UK: Cambridge University Press.

Toulmin, S., Rieke, R., & Janik, A. (1984). *An Introduction to Reasoning* (2nd Ed.). New York, USA: Macmillan Publishing Company.

Weber, K., Maher, C., Powell, A., & Lee, H. S. (2008). Learning opportunities from group discussions: warrants become the objects of debate. *Educational Studies in Mathematics*, *68*(3), 247-261.

Zieffler, A., Garfield, J., delMas, R., y Reading, C. (2008). A framework to support research on informal inferential reasoning. *Statistics Education Research Journal*, *7*(2), 40-58.