HOW STUDENTS RELATE RESIDUALS BEHAVIOR WITH PARAMETER SIGNIFICANCE TESTS IN A LINEAR MODEL? A CASE OF STUDY IN A HIGHER EDUCATION MEXICAN INSTITUTION

José Luis Ángel Rodríguez Silva
Monterrey Institute of Technology, Campus Aguascalientes, México
jlars@itesm.mx

In the implementation of a linear regression model, it is crucial to carry out significance tests to assure the proper inclusion of parameters. In addition, in order to perform such a hypothesis test, students should verify the properties of the residuals and the most important is undoubtedly the normality of the residuals. However, it is clear that there is, in general, an important gap between the hypothesis testing and residuals verification stages. In this paper we present an empirical study about the most probable confusions and misunderstandings students have in relating these two mental statistical constructs in a course of Administrative Statistics in a Mexican Education Higher Institute. One of the main results is that the order in which these two concepts are taught is essential to guarantee there is not a disconnection between the two stages, and perhaps it would be better to first analyze the properties of the residuals and then to engage in the hypothesis testing phase.

BACKGROUND

Regression analysis is a critical assignment to many students in professional bachelors. Indeed, several of the most important problems in Marketing, Finance, Engineering, etc., can be solved only, in a complete sense, by using regression analysis techniques in order to establish a potential relationship between several variables and to predict the behavior of one (or several) of them in relation to the remaining ones. In this sense, for many university programs, the regression analysis course is fundamental and has an important place in college statistics. As a matter of fact, regression analysis is one of the most widely applied branches of statistics.

When performing regression analysis, one of the most important goals is to show the linear model itself and to determine if it fits the data in a reasonable sense. One of the key tasks we must carry out is to perform statistical hypothesis tests of the estimated parameters (based on the t-Student and F distributions for the individual and joint significance, respectively) to know if such parameters should be included in the linear model. After this, by the sequence of topics shown in many statistics books, it is usual to invest a great deal of time to study several aspects of the linear model, like fitness degree, the potential inclusion of some variables (to construct a multiple regression model), several model modifications and related aspects. Not until the last part of the regression analysis module, are presented some properties about the residual behavior, like non-normality, heteroscedasticity, autocorrelation, etc. These properties are crucial to assure the robustness of the hypotheses tests about the estimated model parameters. For example, if there is a severe lack about the normal behavior of residuals, it is very probable to conclude incorrectly about the potential inclusion of the estimated parameters. However, as we have mentioned, these residual properties are frequently not checked until the last part of the Regression Analysis module, leading to the following problems:

- This temporal gap could create potential confusion and a typical student is unlikely to understand the intimate and fundamental relationship between these topics.
- It is not infrequent to fail to study the residual properties due to lack of time and to pressure related to cover other topics of the course. Potentially the professor might not emphasize the intimate link between these two topics.

By the previous considerations, it seem clear that there is a need to have a deeper insight and more profound empirical results in order to perform a clearer framework in the regression analysis topic sequence and to give a coherent didactic proposal about it.

The importance of topic order has been studied by some researchers like Moore (1997), Hoerl, Hahn and Doganaksoy (1997), Chance and Rossman (2001) and Malone, Gabrosek, Curtiss and Race (2010), giving some approaches from different points of view about the “correct” order of topics in statistics courses. However, it is clear that in general terms, the literature is scarce and

In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9, July, 2014), Flagstaff, Arizona, USA.* Voorburg, The Netherlands: International Statistical Institute. iase-web.org

there is not enough analyses, especially empirical ones, to elucidate the complexity and possible solutions to this problem, which is critical to improve the statistics teaching-learning process.

Now, in the usual exposition of many books about regression analysis, although there is not a complete uniformity in the topic exposure, most of them have an order, something we could name "classic order". For example, in Salvatore (2002), uses the following sequence:

- Simple Regression Analysis.
- Multiple Regression Analysis.
- Further Techniques and Applications in Regression Analysis.
- Problems in Regression Analysis.
- Simultaneous Equation Models.

It is evident that there at least two ample academic units (Multiple Regression Analysis and Further Techniques and Applications in Regression Analysis) between the topics of testing hypotheses, checking model assumptions, and understanding the residual behavior (analyzed in Problems in Regression Analysis). For this reason we ask if this order is adequate, in a pedagogic sense, or if we should re-order such topic sequence to increase a better understanding.

With these reflections, we describe now the problem definition.

THE PROBLEM

To analyze, based on a study case in a Higher Education Mexican Institution, which one of the following didactic sequences seem to be more appropriate in the learning-teaching process of the hypothesis testing of the linear model parameters:

- *Didactic Scheme 1: Classical exposure of the Regression Analysis Models, described in textbooks like Salvatore (2002) or Walpole, Myers, Myers and Ye (2007), testing first the significance of the parameters of the model, spending some time in studying intermediate topics, and then (if there is enough time) analyzing the residual behavior.*
- *Didactic Scheme 2: Joint exposure of the significance parameter tests with residual analysis. That is, immediately after the study of the t-Student and F-Fisher tests for model parameter significance, analyze the residual behavior in order to test the suitability of such tests.*

METHOD

- In the semester August-December, 2012, a lecture on Regression Analysis was given to 33 economic-administrative students (finance, marketing and international business management) in a Higher Educative Mexican Institution, under the scheme 1 described in the problem definition section, i.e., under a "classic order".
- In the semester January-June, 2013, the same lecture was given to 35 students with identical profiles, i.e, students of finance, marketing and international business management, but this time under the didactic scheme 2, that is, with a joint exposure of significance test of parameters and immediately after with a complete study of residual behavior.
- All students had taken a previous assignment of *Administrative Statistics I*, with the following topics: univariate descriptive statistics (describing data sets with aid of histograms and pie graphs, and finding the usual statistics like mean, median, mode, variance, standard deviation, etc.), main random variables (discrete and continuous) and the principal properties of normal distribution (analyzing the Central Limit Theorem and the Law of Large Numbers).
- In fair terms, except by the didactic sequences, all other factors remained the same. That is both lectures used the same support textbook (Salvatore, 2002), the same informatics tool (Minitab®), the same schedule (7:00-8:00 am, three times per week) and almost the same independent academic burden (about the 5 hours of independent work per week). Moreover, the socio-demographic and economic students' conditions were very similar (almost the same average age, all of them full-time students and with a very similar mean income). By design, there were no confounding factors and the only significant variable we changed was the didactic sequences previously described.
- What was taught was the basic theory and practice of the linear regression model. Specifically, during the classroom sessions were exposed to the corresponding material of chapters 6 to 9 of Salvatore (2002).

- The way the lectures were given was by the Problem Based Learning (PBL) approach.
- In order to gather information, two items were employed: the first with a quantitative approach, and the second with a qualitative-perception flavor. These two items were:
  - *Suppose you have found a linear model of one dependent variable, Y, against two independent variables, X_1 and X_2. You are considering the linear model $Y=b_0+b_1X_1+b_2X_2$, and you have three t-Student empirical values: 3.45 for b_0, 3.87 for b_1 and 5.32 for b_2. The corresponding t-Student theoretical value is 2.83. Moreover, the empirical F-Fisher value is 5.21 and the corresponding theoretical value is 3.24. Finally, the Anderson-Darling p-value to test normality of residual is 0.001. a) Write the corresponding sets of hypothesis tests for the model parameters; b) What are your conclusions about this model, i.e., do we have a correct model? Explain.*
  - *Given the values of the t-Student and F-Fisher statistics, as well as the Anderson-Darling p-values testing normality of residuals, what is the relationship between these values?*
- In both items, we graded the students' answers in a 0 to 10 scale, 0 with no answer at all and 10 if a complete and satisfactory answer.
- Finally, a survey was carried with several students in both groups after the application of two items in order to capture a more intuitive appreciation. There were applied to questions: *a) Do you think it is important to check the regression model assumptions? b) Why? Please, be as specific as you can.*

RESULTS

- In table 1 are presented, in a compact way, the results of the two items. The confidence intervals were found with a 95% confidence level, in all instances.

Table 1: Results of the implementation of the two didactic schemes

Characteristic	ITEM 1		ITEM 2	
	AD 2012 Group	JJ 2013 Group	AD 2012 Group	JJ 2013 Group
Size group	33	35	33	35
Mean	5.38	7.50	5.15	7.57
Median	5	7.5	5	10
Standard Deviation	3.86	2.84	4.10	3.40
Mode	10	10	10	10
Modal frequency	9	15	11	19
Mean Interval Confidence	(4.01, 6.75)	(6.52, 8.48)	(3.70, 6.60)	(6.40, 8.74)
Mean Difference Confidence Interval (Group JJ Vs. Group AD)	(0.47, 3.77)		(0.59, 4.25)	
Grades ≥ 7.0	16	24	14	24
Grade proportion ≥ 7.0	0.48	0.69	0.42	0.69
Proportion Confidence Interval	(0.31, 0.66)	(0.53, 0.84)	(0.26, 0.59)	(0.53, 0.84)
Proportion Difference Confidence Interval (Group JJ Vs. Group AD)	(-0.03, 0.43)		(0.03, 0.49)	
Variance Interval	(9.63,26.05)	(5.29,13.88)	(10.85,29.34)	(7.57,19.87)
Quotient Variance Interval (Group JJ Vs. Group AD)	(0.27,1.08)		(0.34,1.38)	

Source: *Data analysis from source data*

- Besides, the intuitive survey results points that all the students of the JJ 2013 group, without any single exception, answered that they in fact consider very important to check model assumptions, concentrating their justification in the argument that without such validation, the model itself could be useless. Contrary, in the AD 2012 group, only 52% of students (17 of 33),

were able to answer affirmatively about the importance to check such model assumptions, and only the 33% of them (10 students) give a reasonable justification.

RESULT INTERPRETATIONS

- The academic performance in the two items of the JJ 2013 group seem to be superior, in a 95% confidence level, with respect to the AD 2015 group, as we can conclude by the fact that 0 does not belong to the mean difference confidence interval, although we can note that the grade mode was equal in both groups (10), i.e., in both didactic sequences the modal frequency was 10.
- Not only the grade itself is interesting but the proportions of students with a 7 approved note. We can observe that in item 1 it seems there is no significant difference (due to the fact that 0 belongs to the confidence interval), but, in contrast, in item 2 there seems to be a significant difference.
- Variances seem to be equal in both groups, because 1 is contained in both intervals.
- It is conclusive that the rate of students who recognize the importance of checking model assumptions if they have the chance to link immediately study of residual behavior with the construction of the model itself. If we do not carry out this connection, students could be lost and not realize the importance to check model assumptions, specially the normal residual behavior.
- Finally, by a direct interview with several students of both groups, it seems that the joint scheme is a more reasonable choice to emphasize the robustness of significance test with respect residuals behavior.

CONCLUSIONS

The didactic sequences in statistics courses are relevant, and for this reason, the choice to select a particular order can influence, in a significant way, the academic performance of students.

In this case of study research, it was proved that it is a good pedagogic choice to bind the significance tests of parameters of a linear model with the verification of the residual properties, mostly about normality. Equivalently, it does not seem to be a good strategy to allow a gap between these two topics if we wanted to maximize students understanding about this transcendental Regression Analysis construct.

Finally, there is a lot of work to do to find the best strategies in the topic sequences, not only en Regression Analysis modules, but virtually in all statistics courses. Both, theoretical and empirical work is crucial in order to propose the best educative alternatives to increase students understanding.

REFERENCES

Chance, B., & Rossman, A. (2001). Sequencing topics in introductory statistics: A debate on what to teach when. *The American Statistician, Volume 55*, 140-144.

Hoerl, R., Hahn, G., & Doganaksoy, N. (1997). Comment on “New Pedagogy and New Content: The Case of Statistics” by D. Moore. *International Statistical Review, 65*, 147–153.

Malone, C., Gabrosek, J., Curtiss, P., & Race, M. (2010). Resequencing topics in an introductory applied statistics course. *The American Statistician, 64*(1).

Moore, D. (1997). New pedagogy and new content: The case of statistics. *International Statistical Review, 65*, 123–127.

Salvatore, D., & Reagle, D. (2002). *Shaum’s outline of statistics and econometrics* (2nd ed.). New York: Schaum McGraw-Hill.

Walpole, R., Myers, R., Myers, S., & Ye, K. (2007). *Probability and statistics for engineers and scientists* (8th edition). New Jersey: Pearson, Prentice-Hall.