

**Blanco-Fernández, Angela; Winker, Peter**

**Data generation processes and statistical management of interval data.** (English)

Zbl 1443.62017

AStA, Adv. Stat. Anal. 100, No. 4, 475-494 (2016).

Summary: Statistical methods for dealing with interval data have been developed for some time. Real intervals are the natural extension of real point values. They are commonly considered to generalize the nature of the experimental outcomes from the classical scenario to a more imprecise situation. Interval data have been mainly treated in the context of fuzzy models, as a particular case of increasing the level of imprecision of the data. However, specific methods to deal explicitly with interval data have also been developed. It is described which experimental settings might result in interval-valued data. Some of the major statistical procedures used to deal with interval data are presented. Given the quite different data generation processes resulting in interval data, it is discussed which method appears most appropriate for specific types of interval data. Some practical applications demonstrate the link between data generation processes, specific type of interval data, and statistical methods used for the analysis of these data.

**MSC:**

62A86 Fuzzy analysis in statistics

Cited in 4 Documents

**Keywords:**

experimental data generation process; interval data; epistemic view; ontic view; possibilistic approach; probabilistic approach

**Full Text:** DOI

**References:**

- [1] Abdallah, F; Gning, A; Bonnifait, P, Adapting particle filter on interval data for dynamic state estimation, IEEE Int. Conf. Acoust. Speech Signal Proc., ICASSP, 1153-1156, (2007)
- [2] Bertoluzza, C., Corral, N., Salas, A.: On a new class of distances between fuzzy numbers. *Mathware Soft Comput.* \textbf{2}(2), 71-84 (1995) · Zbl 0887.04003
- [3] Beunza, J; Toledo, E; Hu, F; Bes, M; Serrano, M; Sanchez, A; Martinez, JA; Martinez, MA, Adherence to the Mediterranean diet, long-term weight change, and incident overweight or obesity: the seguimiento universidad de navarra (SUN) cohort, *Am. J. Clin. Nutr.*, 92, 1484-1493, (2010) · doi:10.3945/ajcn.2010.29764
- [4] Billard, L; Diday, E, From the statistics of data to the statistics of knowledge: symbolic data analysis, *J. Am. Stat. Assoc.*, 98, 470-487, (2003) · doi:10.1198/016214503000242
- [5] Blanco-Fernández, A; Corral, N; González-Rodríguez, G, Estimation of a flexible simple linear model for interval data based on set arithmetic, *Comput. Stat. Data Anal.*, 55, 2568-2578, (2011) · Zbl 1464.62030 · doi:10.1016/j.csda.2011.03.005
- [6] Blanco-Fernández, A; Colubi, A; González-Rodríguez, G, Confidence sets in a linear regression model for interval data, *J. Stat. Plan Inference*, 142, 1320-1329, (2012) · Zbl 1242.62072 · doi:10.1016/j.jspi.2011.09.017
- [7] Blanco-Fernández, A; Casals, R; Colubi, A; Corral, N; García-Bárcana, M; Gil, MA; González-Rodríguez, G; López, T; Lubiano, A; Montenegro, M; Ramos-Guajardo, A; de la Rosa de Saa, S; Sinova, B, A distance-based statistical analysis of fuzzy number-valued data, *Int. J. Approx. Reason.*, 55, 1487-1501, (2014) · Zbl 1407.62094 · doi:10.1016/j.ijar.2013.09.020
- [8] Bock, H.H., Diday, E.: *Analysis of Symbolic Data. Exploratory Methods for Extracting Statistical Information from Complex Data.* Springer, Heidelberg (2000) · Zbl 1039.62501
- [9] Černý, M; Antochb, J; Hladík, M, On the possibilistic approach to linear regression models involving uncertain, indeterminate or interval data, *Inf. Sci.*, 244, 26-47, (2013) · Zbl 1357.62237 · doi:10.1016/j.ins.2013.04.035
- [10] Cerquera, D., Laisney, F., Ullrich, H.: Considerations on Partially Identified Regression Models. Working Papers of BETA No. 2012-07, ZEW. Centre for European Economic Research Discussion Paper No. 12-024 (2012). ftp://ftp.zew.de/pub/zew-docs/dp/dp12024
- [11] Chavent, M; Carvalho, FAT; Lechevallier, Y; Verde, R, New clustering methods for interval data, *Comput. Stat.*, 21, 211-229, (2006) · Zbl 1114.62069 · doi:10.1007/s00180-006-0260-0
- [12] Colubi, A; López-Díaz, M; Domínguez-Menchero, JS; Gil, MA, A generalized strong law of large numbers, *Probab. Theory Relat.*, 114, 401-417, (1999) · Zbl 0933.60023 · doi:10.1007/s004400050229
- [13] Colubi, A, Statistical inference about the means of fuzzy random variables: applications to the analysis of fuzzy- and real-

- valued data, *Fuzzy Set Syst.*, 160, 344-356, (2009) · [Zbl 1175.62021](#) · [doi:10.1016/j.fss.2007.12.019](#)
- [14] Corral, N; Gil, MA; Gil, P; Pardo, L (ed.); et al., Interval and fuzzy-valued approaches to the statistical management of imprecise data, 453-468, (2011), Heidelberg · [doi:10.1007/978-3-642-20853-9\\_31](#)
- [15] De la Rosa de Sáa, S; Gil, MA; González-Rodríguez, G; López, MT; Lubiano, MA, Fuzzy rating scale-based questionnaires and their statistical analysis, *IEEE T Fuzzy Syst.*, 23, 111-126, (2015) · [doi:10.1109/TFUZZ.2014.2307895](#)
- [16] Diamond, P, Least squares Fitting of compact set-valued data, *J. Math. Anal. Appl.*, 147, 531-544, (1990) · [Zbl 0704.65006](#) · [doi:10.1016/0022-247X\(90\)90353-H](#)
- [17] Domingues, MAO; Souza, R; Cysneiros, FJA, A robust method for linear regression of symbolic interval data, *Pattern Recogn. Lett.*, 31, 1991-1996, (2010) · [doi:10.1016/j.patrec.2010.06.008](#)
- [18] Duarte, AP; Brito, P, Linear discriminant analysis for interval data, *Comput. Stat.*, 21, 289-308, (2006) · [Zbl 1113.62080](#) · [doi:10.1007/s00180-006-0264-9](#)
- [19] Dubois, D; Couso, I, Statistical reasoning with set-valued information: ontic vs. epistemic views, *Int. J. Approx. Reason.*, 55, 1502-1518, (2014) · [Zbl 1407.62032](#) · [doi:10.1016/j.ijar.2013.07.002](#)
- [20] D'Urso, P; Giordani, P, A least squares approach to principal component analysis for interval valued data, *Chemom. Intell. Lab. Syst.*, 70, 179-192, (2004) · [doi:10.1016/j.chemolab.2003.11.005](#)
- [21] Fischer, H; García-Bárcana, M; Tillmann, P; Winker, P, Evaluating FOMC forecast ranges: an interval data approach, *Empir. Econ.*, 47, 365-388, (2013) · [doi:10.1007/s00181-013-0736-z](#)
- [22] Fischer, H., Blanco-Fernández, A., Winker, P.: Predicting stock return volatility: can we benefit from regression models for return intervals? *J. Forecast.* (2015) (forthcoming)
- [23] Gil, MA; López-García, MT; Lubiano, MA; Montenegro, M, Regression and correlation analyses of a linear relation between random intervals, *Test*, 10, 183-201, (2001) · [Zbl 0981.62062](#) · [doi:10.1007/BF02595831](#)
- [24] Gil, MA; González-Rodríguez, G; Colubi, A; Montenegro, M, Testing linear independence in linear models with interval-valued data, *Comput. Stat. Data Anal.*, 51, 3002-3015, (2007) · [Zbl 1161.62358](#) · [doi:10.1016/j.csda.2006.01.015](#)
- [25] Gil, MA; González-Rodríguez, G; Trillas, E (ed.); et al., Fuzzy vs likert scales in statistics, 407-420, (2012), Heidelberg
- [26] Giordani, P; Kiers, HAL, Three-way component analysis of interval valued data, *J. Chemometr.*, 18, 253-264, (2004) · [doi:10.1002/cem.868](#)
- [27] Gonzalez-Calvo, A., Hernandez-Leal, P.A., Arbelo, M.: Forest Fire Risk Dynamic Index. In: De la Riva, J. et al. (eds.) *Proceedings of 5th International Workshop on Remote Sensing and GIS Applications to Forest*, pp. 125-129 (2005)
- [28] González-Rodríguez, G; Blanco, A; Corral, N; Colubi, A, Least squares estimation of linear regression models for convex compact random sets, *Adv. Data Anal. Classif.*, 1, 67-81, (2007) · [Zbl 1131.62058](#) · [doi:10.1007/s11634-006-0003-7](#)
- [29] González-Rodríguez, G., Trutschnig, W., Colubi, A.: Confidence regions for the mean of a fuzzy random variable. In: *Abstracts of IFSA World Congress/EUSFLAT Conference (IFSA-EUSFLAT 2009, Lisbon, Portugal)* · [Zbl 1170.68045](#)
- [30] Hofer, E.P., Rauh, A.: Applications of Interval Algorithms in Engineering. In: Luther, W., Otten, W. (eds.) *International Symposium on Scientific Computing, Computer Arithmetic and Validated Numerics (12th GAMM—IMACS 2006, Germany)*. IEEE Computer Society Conference Publishing Services (2006)
- [31] Horowitz, JL; Manski, CF; Ponomareva, CF; Stoye, J, Computation of bounds on population parameters when the data are incomplete, *Reliab. Comput.*, 9, 419-440, (2003) · [Zbl 1140.62314](#) · [doi:10.1023/A:1025865520086](#)
- [32] Heitjan, DF; Rubin, DB, Ignorability and coarse data, *Ann. Stat.*, 19, 2244-2253, (1991) · [Zbl 0745.62004](#) · [doi:10.1214/aos/1176348396](#)
- [33] Hodge, AM; English, DR; Itsiopoulos, C; ODea, K; Giles, GG, Does a Mediterranean diet reduce the mortality risk associated with diabetes: evidence from the Melbourne collaborative cohort study, *Nutr. Metab. Cardiovasc. Dis.*, 21, 733-739, (2011) · [doi:10.1016/j.numecd.2010.10.014](#)
- [34] Hu, C; He, L, An application of interval methods to stock market forecasting, *Reliab. Comput.*, 13, 423-434, (2007) · [Zbl 1125.91348](#) · [doi:10.1007/s11155-007-9039-4](#)
- [35] Huang, J., Wellner, J.: In: Lin, D.Y., Fleming, T.R. (eds.) *Proceedings of First Seattle Symposium in Biostatistics, Lecture Notes in Statistics. Interval Censored Survival Data: A Review of Recent Progress*, pp. 123-169. Springer, New York (1997) · [Zbl 0916.62077](#)
- [36] Jahanshahloo, GR; Lotfi, FH; Malkhalifeh, MR; Namin, MA, A generalized model for data envelopment analysis with interval data, *Appl. Math. Model.*, 33, 3237-3244, (2008) · [Zbl 1205.90152](#) · [doi:10.1016/j.apm.2008.10.030](#)
- [37] Joslyn, C, Measurement of possibilistic histograms from interval data, *Int. J. Gen. Syst.*, 26, 9-33, (1997) · [Zbl 0897.94020](#) · [doi:10.1080/03081079708945167](#)
- [38] Kallithraka, S; Arvanitoyannis, IS; Kefalas, P; El-Zajouli, A; Soufleros, E; Psarra, E, Instrumental and sensory analysis of Greek wines; implementation of principal component analysis (PCA) for classification according to geographical origin, *Food Chem.*, 73, 501-514, (2001) · [doi:10.1016/S0308-8146\(00\)00327-7](#)
- [39] Körner, R., Näther, W.: *On the Variance of Random Fuzzy Variables. Statistical Modelling, Analysis and Management of Fuzzy Data*, pp. 22-39. Springer, Berlin (2002)
- [40] Kristiansen, L; Gronbaek, M; Becker, U; Tolstrup, J-S, Risk of pancreatitis according to alcohol drinking habits: a population-based cohort study, *Am. J. Epidemiol.*, 168, 932-937, (2008) · [doi:10.1093/aje/kwn222](#)
- [41] Lauro, N.C., Palumbo, F.: Principal component analysis for non-precise data. In: Vichi et al. (eds.) *New developments in classification and data analysis*, pp. 173-184. Springer (2005) · [Zbl 1341.62163](#)
- [42] Lima Neto, EA; de Carvalho, FAT, Constrained linear regression models for symbolic interval-valued variables, *Comput. Stat.*

- Data Anal., 54, 333-347, (2010) · [Zbl 1464.62055](#) · [doi:10.1016/j.csda.2009.08.010](#)
- [43] Liu, J; Liu, W; Wu, L; Yan, G, A flexible approach for multivariate mixed-effects models with non-ignorable missing values, *J. Stat. Comput. Simul.*, 85, 3727-3743, (2015) · [doi:10.1080/00949655.2015.1005014](#)
- [44] Manski, CF; Tamer, E, Inference on regressions with interval data on a regressor or outcome, *Econometrica*, 70, 519-546, (2002) · [Zbl 1121.62544](#) · [doi:10.1111/1468-0262.00294](#)
- [45] Manski, C.F.: *Partial Identification of Probability Distributions*. Springer, New York (2003) · [Zbl 1047.62001](#)
- [46] Matheron, G.: *Random Sets and Integral Geometry*. Wiley, New York (1975) · [Zbl 0321.60009](#)
- [47] Molchanov, I.: *Theory of Random Sets. Probability and its Applications*. Springer, London (2005)
- [48] Nakama, T; Colubi, A; Lubiano, MA; Borgelt, C (ed.); et al., Two-way analysis of variance for interval-valued data, 475-482, (2010), Heidelberg · [doi:10.1007/978-3-642-14746-3\\_59](#)
- [49] Pimentel, BA; Souza, MCR, Possibilistic clustering methods for interval-valued data, *Int. J. Uncertain. Fuzzy*, 22, 263-291, (2014) · [Zbl 1323.62055](#) · [doi:10.1142/S0218488514500135](#)
- [50] Puri, M; Ralescu, D, Fuzzy random variables, *J. Math. Anal. Appl.*, 114, 409-422, (1986) · [Zbl 0592.60004](#) · [doi:10.1016/0022-247X\(86\)90093-4](#)
- [51] Ramos-Guajardo, AB; González-Rodríguez, G; Borgelt, C (ed.); et al., Testing the variability of interval data: an application to tidal fluctuation, 65-74, (2013), Heidelberg · [Zbl 1348.62256](#) · [doi:10.1007/978-3-642-30278-7\\_6](#)
- [52] Ramos-Guajardo, AB; Colubi, A; González-Rodríguez, G, Inclusion degree tests for the Aumann expectation of a random interval, *Inf. Sci.*, 288, 412-422, (2014) · [Zbl 1357.62088](#) · [doi:10.1016/j.ins.2014.08.013](#)
- [53] Schneeweiss, H; Augustin, T, Some recent advances in measurement error models and methods, *Allgemeines Statistisches Archiv AStA*, 90, 183-197, (2006) · [Zbl 1103.62063](#) · [doi:10.1007/s10182-006-0229-x](#)
- [54] Schneeweiss, H; Komlos, J; Ahmad, AS, Symmetric and asymmetric rounding: a review and some new results, *Adv. Stat. Anal.*, 94, 247-271, (2010) · [Zbl 1443.62061](#) · [doi:10.1007/s10182-010-0125-2](#)
- [55] Schollmeyer, G; Augustin, T, Statistical modeling under partial identification: distinguishing three types of identification regions in regression analysis with interval data, *Int. J. Approx. Reason.*, 56, 224-248, (2015) · [Zbl 06372148](#) · [doi:10.1016/j.ijar.2014.07.003](#)
- [56] Shao, J., Tu, D.: *The Jackknife and Bootstrap*. Springer, New York (1995) · [Zbl 0947.62501](#) · [doi:10.1007/978-1-4612-0795-5](#)
- [57] Sinova, B; Casals, MR; Colubi, A; Gil, MA; Borgelt, C (ed.); et al., The Median of a random interval, 575-583, (2010), Heidelberg · [doi:10.1007/978-3-642-14746-3\\_71](#)
- [58] Spano, D., Georgiadis, T., Duce, P., Rossi, F., Delitala, A., Dessy, C., Bianco, G.: A fire index for mediterranean vegetation based on micrometeorological and ecophysiological measurements. *Am. Meteorol. Soc.* 3.1 (2003). <https://ams.confex.com/ams/pdfpapers/65497.pdf>
- [59] Stoye, J, Partial identification of spread parameters, *Quant. Econ.*, 1, 323-357, (2010) · [Zbl 1205.62015](#) · [doi:10.3982/QE24](#)
- [60] Tamer, E, Partial identification in econometrics, *Annu. Rev. Econ.*, 2, 167-195, (2010) · [doi:10.1146/annurev.economics.050708.143401](#)
- [61] Xu, S; Chen, X; Han, A; Huynh, VN (ed.); et al., Interval/probabilistic uncertainty and non-classical logics, 353-363, (2008), Heidelberg
- [62] Zadeh, LA, Fuzzy sets, *Inf. Control*, 8, 338-353, (1965) · [Zbl 0139.24606](#) · [doi:10.1016/S0019-9958\(65\)90241-X](#)
- [63] Zuccolotto, P, Principal component analysis with interval imputed missing values, *Adv Stat Anal*, 96, 123, (2012) · [Zbl 1443.62178](#) · [doi:10.1007/s10182-011-0164-3](#)

This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.