

Eksik Veri Tiplerinde (MAR, MCAR, MNAR) Tamamlama Algoritmalarının Parametre Tahmin Analizi

Editör

Yıldırım DEMİR

Yazarlar

Dr. Öğr. Üyesi Şakir İŞLEYEN

Serdar İNAN



AKADEMİSYEN
KİTABEVİ

© Copyright 2019

Bu kitabın, basım, yayın ve satış hakları Akademisyen Kitabevi A.Ş.'ne aittir. Anılan kuruluşun izni alınmadan kitabın tümü ya da bölümleri mekanik, elektronik, fotokopi, manyetik kağıt ve/veya başka yöntemlerle çoğaltılamaz, basılamaz, dağıtılamaz. Tablo, şekil ve grafikler izin alınmadan, ticari amaçlı kullanılamaz. Bu kitap T.C. Kültür Bakanlığı bandrolü ile satılmaktadır.

Bu Çalışma Serdar İnan'ın Tezinden Üretilmiştir.

ISBN

978-605-258-705-8

Kitap Adı

Eksik Veri Tiplerinde (MAR, MCAR, MNAR) Tamamlama
Algoritmalarının Parametre Tahmin Analizi

Editör

Yıldırım DEMİR

Yazarlar

Dr. Öğr. Üyesi Şakir İŞLEYEN
Serdar İNAN

Yayın Koordinatörü

Yasin DİLMEN

Sayfa Tasarımı

Akademisyen Dizgi Ünitesi

Yayıncı Sertifika No

25465

Baskı ve Cilt

Uzun Dijital

Bisac Code

BUS000000

GENEL DAĞITIM

Akademisyen Kitabevi A.Ş.

Halk Sokak 5 / A

Yenişehir / Ankara

Tel: 0312 431 16 33

siparis@akademisyen.com

www.akademisyen.com

İÇİNDEKİLER

1. GİRİŞ	1
2. GENEL BİLGİLER	5
2.1. Veri Matrisi (Data Matrix)	5
2.2. Eksik Veri Tanımı	6
2.3. Eksik Verinin Tarihsel Gelişimi	9
2.4. Eksik Veri Mekanizmaları (Türleri)	10
2.4.1. Tamamen rastgele eksik veri (missing completely at random- mcar)	12
2.4.2. Rastgele eksik veri (missing at random -mar)	14
2.4.3. Rastgele olmayan eksik veri (missing not at random - mmar)	16
3. EKSİK VERİLERİ YÖNETME YÖNTEMLERİ	25
3.1. Geleneksel Yöntemler	28
3.1.1. Liste boyunca silme (listwise deletion -dizin silme)	28
3.1.2. Çift yönlü silme (pairwise deletion-ikili silme)	32
3.1.3. Ortalama atama (mean imputation, ortalama değerlendirme)	36
3.1.4. Regresyon atama (regression imputation)	38
3.1.5. Tahmini ortalama eşleştirme (sıcak deste-hot deck imputation)	42
3.1.6. Soğuk deste ataması (cold deck imputation)	44
3.1.7. En yakın komşu yöntemi (nearest neighbor)	45
3.1.8. Son gözlemi ileri taşıma (last observation carried forward)	45
3.2. Modern Atama Yöntemleri	46
3.2.1. Çoklu atama (multiple imputation)	46
3.2.2. Beklenti maksimizasyonu (em: expectation maximization)	48
4. YÖNTEM	51
4.1. Veri Üretimi	51
4.2. Veri Eksiltilmesi	52
4.3. Veri Tamamlanması	52
4.4. Parametrelerin Hesaplanması	52

5.1. Normal Dağılım 55

5.2. Üstel Dağılım..... 73

5.3. Gamma Dağılımı 91

5.4. Multinomial Dağılım 109

6. TARTIŞMA VE SONUÇ..... 115

Kaynaklar 119

KAYNAKLAR

1. Akbaş, U. Tavşancıl, E. 2015. Farklı Örneklem Büyüklüklerinde ve Kayıp Veri Örüntülerinde Ölçeklerin Psikometrik Özelliklerinin Kayıp Veri Baş Etme Teknikleri ile İncelenmesi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 6(1): 38-57.
2. Allison, P. D., 2000. Multiple İmputation for Missing Data: A Cautionary Tale. *Sociological Methods and Research*, 28: 301-309.
3. Allison, P. D., 2003 . Missing Data Techniques for Structural Equation Modeling. *Journal of Abnormal Psychology*, 4(1): 545-557.
4. Allison, Paul D. 2009. *Missing Data, Handbook of Quantitative Methods in Psychology* (Editör: Roger E. Millsap ve Alberto Maydeu-Olivares). Sage Publications.72-89.
5. Allison, Paul D. 2012. *Handling Missing Data by Maximum Likelihood*. SAS Global Forum.32.
6. Alpar, R., 2017. **Uygulamalı Çok Değişkenli İstatistiksel Yöntemler**. Detay Yayıncılık. Ankara. 840.
7. Bal, C., 2003. Çok Gruplu Veri Setlerinde Eksik Gözlem Sorununun Çözümlemesi ve **Sağlık Alanında Bir Uygulama**. (Doktora Tezi), Osmangazi Üniversitesi, Sağlık Bilimleri Enstitüsü, Eskişehir.
8. Baraldi A.N., Enders. C.,2010. An İntroduction to Modern Missing Data Analyses, *Journal of School Psychology* 48 : 5-37
9. Baygül, A., 2007, *Kayıp Veri Analizinde Sıklıkla Kullanılan Etkin Yöntemlerin Değerlendirilmesi*, Yüksek Lisans Tezi, İ. Ü. Sağlık Bilimleri Enstitüsü, İstanbul.
10. Bergmeir, C., Benitez, J.M., 2012. On The Use Of Cross-Validation for Time Series Predictor Evaluation, *Inform Sciences*, 191:192-213
11. Bernaards, C.A., Sijtsma, K., 2000. Influence of İmputation and EM Methods on Factor Analysis When İtem Nonresponse in



- Questionnaire Data is Nonignorable. *Multivariate Behavioral Research*, 35(3): 321-364.
12. Buuren, S. V., Oudshoorn K.G., 2011. Mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software* 45 (3): 1-67.
 13. Buuren, S.V., 2012. *Flexible Imputation of Missing Data*. Boca Raton, FL: Chapman & Hall/CRC.416.
 14. Carpita, M., Manisera, M., 2011. On The Imputation of Missing Data in Surveys With Likert-Type Scales. *Journal of Classical*, 28: 93-112
 15. Cheng, T. S., Victoria-Feser, M. P., 2002, High-Breakdown Estimation of Multivariate Mean And Covariance With Missing Observations. *British J. Math. Statist. Psych.*, 55: 317-335.
 16. Cook, R. J., Zeng, L., Yi, G. Y., 2004. Marginal Analysis Of Incomplete Longitudinal Binary Data: A Cautionary Note On LOCF Imputation. *Biometrics*, 60:820-828.
 17. Cumming, P., 2013. Missing Data and Multiple Imputation. *Clinical Review & Education*, 167 (7): 656-661.
 18. Demir, E., 2013. Kayıp Verilerin Varlığında Çoktan Seçmeli Testlerde Madde Ve Test Parametrelerinin Kestirilmesi: SBS Örneği. *Eğitim Bilimleri Araştırmaları Dergisi*, 3(2): 48-68.
 19. Demir, E., Parlak, B., 2012. Türkiye’de Eğitim Araştırmalarında Kayıp Veri Sorunu. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 3 (1): 20-241.
 20. Demir, E., 2012. Araştırmalarda Kayıp Veri Sorunu, Çağdaş Eğitim Dergisi Akademik, 1(3): 78-87
 21. Dempster, A. P., Laird N. M., Rubin D. B., 1977. Maximum Likelihood Estimation from Incomplete Data via the EM Algorithm (with Discussion). *Journal of the Royal Statistical Society*, 39 (1): 1-38.
 22. Donders, A.R.T., Heijden, G.J.M.G., Stijnen, T., Moons, K.G.M., 2006. A Gentle Introduction To Imputation Of Missing Values. *Journal of Clinical Epidemiology*, 59: 1087-1091



23. Downey, R.G., King, C.V. 1998. Missing Data in Likert Ratings: A Comparison of Replacement Methods. *The Journal of General Psychology*, 125 (2): 175-191
24. Duncan, T.E., Duncan, S.C. ve Li, F. 1998. A Comparison Of Model- And Multiple İmputation-Based Approaches To Longitudinal Analyses With Partial Missingness. *Structural Equation Modeling: A Multidisciplinary Journal*, 5 (1): 1-21
25. Enders, C. K., 2010. *Applied Missing Data Analysis*. The Guilford Press, New York,401.
26. Enders, C.K., 2004. The İmpact Of Missing Data On Sample Reliability Estimates: Implications For Reliability Reporting Practices. *Educational and Psychological Measurement*, 64(3): 419-436.
27. Genc, S., Boran, F.E., Akay, D., Xu, Z.S., 2010. Interval Multiplicative Transitivity for Consistency, Missing Values and Priority Weights of İnterval Fuzzy Preference Relations, *Inform Sciences*, 180: 4877-4891.
28. Graham, J.W., 2009. Missing Data Analysis: Making İt Work İn The Real World. *Annual Review of Psychology*, 60: 549-576.
29. King, G., J. Honaker, A. Joseph,K. Scheve., 2001. Analyzing Incomplete Political Science Data: An Alternative Algorithm for Multiple İmputation. *American Political Science Review*, 95 (1): 49-69.
30. Kürşad,M.Ş., Nartgün. Z , (2015). Eksik Veri Sorununun Çözümünde Kullanılan Farklı Yöntemlerin Ölçeklerin Geçerlik ve Güvenirliği Bağlamında Karşılaştırılması. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*,6(2): 254-267
31. Lee, H., Rancourt, E., Srndal, C. E., 1994. Experiments with Variance Estimation from Survey Data with İmputed Values. *Journal of Official Statistics* 10: 231-238.
32. Leite, W. ve Beretvas S.N. 2010. The Performance Of Multiple İmputation For Likert-Type İtems With Missing Data. *Journal of Modern Applied Statistical Methods*,9(1): 64-74.



33. Little, R. J. A., Rubin, D. B., 1987. *Statistical Analysis with Missing Data*, John Wiley & Sons, New York. 371.
34. Little, R. J. A., Rubin, D. B., 2002. *Statistical Analysis with Missing Data (2nd ed.)*, John Wiley & Sons, New York. 381.
35. Little, R.J.A., 1988. Missing Data Adjustments in Large Surveys. *Journal of Business & Economic Statistics*, 6 (3): 287-296.
36. Liu, G., Gould, A. L., 2002. Comparison Of Alternative Strategies For Analysis Of Longitudinal Trials. *Journal of Biopharmaceutical Statistics*, 12: 207-226.
37. Longford, N.T., 2006. *Missing Data and Small-area Estimation*, Springer, 376 .
38. Mallinckrodt, C. H., Clark, W. S., David, S. R., 2001. Accounting For Dropout Bias Using Mixed Effects Models, *Journal of Biopharmaceutical Statistics*, 11: 9-21.
39. Marsh, H. W., 1998. Pairwise Deletion for Missing Data in Structural Equation Models: Nonpositive Definite Matrices, Parameter Estimates, Goodness of Fit, and Adjusted Sample Sizes. *Structural Equation Modeling* 5 (1): 22-36.
40. McKnight, P. M., 2007. *Missing Data : A Gentle Introduction* The Guilford Press. . New York.251.
41. Meng, Z.Q., Shi, Z.Z., 2012. Extended Rough Set-Based Attribute Reduction In Inconsistent Incomplete Decision Systems. *Inform Sciences*, 204: 44-69.
42. Musil,C.M., Warner, C.B, Yobas, P.K., Jones, S.L., 2002. A Comparison Of Imputation Techniques For Handling Missing Data. *Western Journal of Nursing Research*, 24(7):815-829.
43. Nartgün, Z.,2015. Kayıp Veri Sorununun Çözümünde Kullanılan Farklı Yöntemlerin Farklı Kayıp Veri Koşulları Altında Ölçeklerin Psikometrik Nitelikleri Ve Ölçme Sonuçları Bağlamında Karşılaştırılması. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, 7(4): 252-265.
44. Nelwamondo, F.V., Marwala, T.,2008. Techniques For Handling Missing Data: Applications to Online Condition Monitoring. *International Journal of Innovative Computing*, 4:13-41.



45. Orchard, T., Woodbury, M.A., 1972. A Missing Information Principle: Theory and Applications, *Proc. Sixth Berkeley Symp. on Math. Statist. and Prob.*, 1: 697-715.
46. Osborne, J. W., 2013. *Best Practices In Data Cleaning*. Sage Publication, California. 226.
47. Özdamar, K., 2018. *Paket Programlar ile İstatistiksel Veri Analizi-1*. Nisan Kitabevi, Eskişehir. Türkiye. 584.
48. Özdamar, K., 2018. *Paket Programlar ile İstatistiksel Veri Analizi-2*. Nisan Kitabevi, Eskişehir. Türkiye. 474.
49. Pelckmans, K., De Brabanter, J., Suykens, J.A.K., De Moor, B., 2005. Handling Missing Values In Support Vector Machine Classifiers, *Neural Networks*, 18: 684-692.
50. Peng, C.-Y. J., Harwell, M., Liou, S.-M., Ehman L. H. , 2006. *Advances In Missing Data Methods And Implications For Educational Research*. (Editör: S. Sawilowsky) , Real Data Analysis. Greenwich, CT: Information Age Publishing Inc. 31-78.
51. Pigott, T. D., 2001). A Review Of Methods For Missing Data. *Educational Resarch and Evaluation*, 7(1): 353-383.
52. Praag V., Dijkstra T. K., Velzen J. V., 1985. Least-Squares Theory Based on General Distributional Assumptions with an Application to the Incomplete Observations Problem. *Psychometrika* 50 (1): 25-36.
53. Rubin, D. B., 1976. *Inference and Missing Data*, Biometrika, 63: 581-592.
54. Rubin, D.B. , 1987. *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons Inc., New York.
55. Satıcı, E., 2009. *Kayıp Gözlem Olması Durumunda Kitle Ortalaması Tahmini*, Doktora Tezi, H.Ü. Fen Bilimleri Enstitüsü, Ankara.
56. Satıcı, E., Kadılar, C., 2009. Kayıp Gözlem Olduğunda Kitle Ortalamasının Tahmini. *Anadolu Üniversitesi Bilim ve Teknoloji Dergisi*, 10(2): 549-556.
57. Schafer, J. L., J. W. Graham. 2002. Missing Data: Our View of the State of the Art. *Psychological Methods* 7 (2): 147-177.



58. Schafer, J. L., N. Schenker., 2000. Inference with Imputed Conditional Means. **Journal of the American Statistical Association**, **95** (449): 144-154.
59. Schafer, J. L., 1999. Multiple Imputation: A Primer. *Statistical Methods On Medical Research*, **8**(1): 3-15.
60. Schaffer, J.L., (1997). *Analysis of Incomplete Multivariate Data*, Chapman&Hall, London. 444.
61. Sehgal, M.S., Gondal, L., Dooley, L.S., 2005. Collateral Missing Value Imputation: A New Robust Missing Value Estimation Algorithm For Microarray Data, *Bioinformatics*, **21**: 2417-2423,
62. Ser G., Bati C.T., 2015., Bati Eksik Veri Analizinde Çoklu Atama Yönteminin Değerlendirilmesi: Hayvancılıkta Tekrarlı Ölçüm Verisi Üzerine Bir Uygulama. *Türk Tarım - Gıda Bilim ve Teknoloji Dergisi*, **3**(12): 926-932.
63. Silva, L., Zarate, L., 2014. A Brief Review of The Main Approaches for Treatment of Missing Data. *Intelligent Data Analysis*, **1177-1198**.
64. Streiner, D.V., 2002. The Case Of The Missing Data: Methods Of Dealing With Dropouts And Other Research Vagaries. *Research methods in Psychiatry*, **47**: 68-75.
65. Şeker Ş. E., Eşmekaya E., 2017. Eksik Verilerin Tamamlanması (Imputation) *YBS Ansiklopedi*, **4**(3):10-17
66. Toka.O., 2012. **Kayıp Veri Durumunda Sağlam Kestirim(-Yüksek Lisans Tezi)** Hacettepe Üniversitesi Fen Bilimleri Enstitüsü. Ankara.
67. Twisk J, De Boer M, De Vente W, Heymans M. 2013. Multiple Imputation Of Missing Values Was Not Necessary Before Performing A Longitudinal Mixed-Model Analysis. *J. Clin. Epidemiol.*, **66**: 1022-1028.
68. Van der Ark, L. A., Vermunt, J. K., 2010. New Developments in Missing Data Analysis. *Methodology*, **6**(1), 1-2



69. Wilkinson, L., 1999. Statistical Methods In Psychology Journals: Guidelines And Explanations. *American Psychologist, Task Force on Statistical Inference*, 54: 594-604.
70. Young, W., Weckman, G., Holland, W. 2011. A Survey Of Methodologies For The Treatment Of Missing Values Within Datasets: Limitations And Benefits, *Theoretical Issues in Ergonomics Science*, 12 (1): 15-43.
71. Zhu X. 2014. Comparison Of Four Methods For Handling Missing Data In Longitudinal Data Analysis Through A Simulation Study. *Open Journal of Statistics*, 4: 933-944.