PYTHON UYGULAMALI İSTATİKSEL VERİ BİLİMİ VE ANALİZİ

Yazar

Dr. Ahmet SEL

ISBN
978-625-7409-60-5

Kitap Adı
Python Uygulamalı İstatiksel Veri Bilimi ve Analizi

Yazarlar
Ahmet SEL
ORCID iD: 0000-0003-1914-5878

Yayın Koordinatörü
Yasin DİLMEN

Sayfa ve Kapak Tasarımı
Akademisyen Dizgi Ünitesi

Yayıncı Sertifika No
47518

Baskı ve Cilt
Vadi Matbaacılık

Bisac Code
SOC027000

DOI
xxx

GENEL DAĞITIM

Akademisyen Kitabevi A.Ş.

Halk Sokak 5 / A
Yenişehir / Ankara
Tel: 0312 431 16 33
siparis@akademisyen.com

www.akademisyen.com

ÖNSÖZ

Bilimsel gelişmelerin hızlı yaşandığı günümüzde giderek daha fazla veri toplanır ve depolanır hale gelmiştir. İlk başta veriler; sayı, metin, resim, video ve daha birçok kaynaktan elde edilen anlamsız bir yığın olarak görülebilir. Büyük veri, değişik kaynaklardan toparlanan ham verilerin anlamlı ve işlenebilir biçime dönüştürülmüş biçimine denir. Veri Bilimi; verilerin işlenmesi, analiz edilmesi ve yorumlanması adımlarını kullanarak faydalı bilgilerin elde edilmesini sağlar. Verilerin analiz edilmesinde istatistiksel yöntemlerin kullanılması verilerin yorumlanmasında hatalı sonuçların elde edilmesini önler. Çalışmada bu amaçla; veri temizleme, veri hazırlama, veri görselleştirme ve veri analizi dahil olmak üzere istatistiksel yöntemler kullanılmıştır.

Bu kitapta sunulan pratik kavramlar karar verme süreçlerinde bilgiyi geliştirmek ve yorumlamak için sosyoloji, pazarlama, işletme, kalite kontrol, eğitim, ekonomi, tıp ve mühendislik gibi veri analizi ve sentezinin gerekli olduğu birçok bilim dalında faydalı olacağı düşünülmektedir. Ayrıca kitapta yer alan örnek uygulama ve sorularda gerçek veri setleri ve güncel örnekler kullanılmıştır. Örneklerin tamamında uygulama adımları en baştan itibaren verilerek uygulamaların anlaşılabilirliği arttırılmıştır. Çalışmada Python programlama dili, öğrenmesi ve kullanması kolay olması yanında diğer programlama dillerinden daha az ayrıntılı ve okunaklı olduğu için tercih edilmiştir. Özellikle Python programında uygulamalar sonrası elde edilen çıktılar, analistler ve karar vericiler için tatmin edici olmaktadır.

İÇİNDEKİLER

KAYNAKLAR

Bartlett, M. S. (1937). Properties of sufficiency and statistical tests. *Proceedings of the Royal Society of London. Series A-Mathematical and Physical Sciences*, *160*(901), 268-282.

Bortz, J. (1993). Statistik für Sozialwissenschaftler. 4. Aufl., 419-420.

Conover, W. J., & Iman, R. L. (1979). *On multiple-comparisons procedures* (pp. pp1-14). Technical report, Los Alamos Scientific Laboratory.

Cressie, N., & Read, T. R. (1984). Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society: Series B (Methodological)*, *46*(3), 440-464.

D'Agostino, R. B. (1971). An omnibus test of normality for moderate and large size samples. *Biometrika*, *58*(2), 341-348.

Downey, A. (2014). *Think stats: exploratory data analysis*. " O'Reilly Media, Inc.".

Dunn, O. J. (1964). Multiple comparisons using rank sums. *Technometrics*, *6*(3), 241-252.

Glantz, S. A. (2002). Primer of biostatistics.

Haslwanter, T. (2016). An Introduction to Statistics with Python. *With applications in the life sciences. Switzerland: Springer International Publishing*.

https://www.kdnuggets.com/2020/07/complete-guide-survival-analysis-python-part1.html

https://www.kdnuggets.com/2020/07/guide-survival-analysis-python-part-2.html

https://www.kdnuggets.com/2020/07/guide-survival-analysis-python-part-3.html

https://www.kaggle.com/kashnitsky/topic-9-part-1-time-series-analysis-in-python

https://numpy.org/doc/stable/reference/routines.statistics.html

https://docs.scipy.org/doc/scipy/reference/stats.html

https://pandas.pydata.org/pandas-docs/stable/index.html

https://matplotlib.org/stable/users/index.html

https://pypi.org/

https://www.anaconda.com/products/individual#Downloads

Idris, I. (2014). *Python data analysis*. Packt Publishing Ltd.

Kaplan, D. (2009). *Statistical modeling: A fresh approach*. St Paul: Macalester College.

Kaplan, R. M., & Irvin, V. L. (2015). Likelihood of null effects of large NHLBI clinical trials has increased over time. *PloS one*, *10*(8), e0132382.

Kendall, M. G. (1948). Rank correlation methods.

Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association*, *47*(260), 583-621.

Levene, H. (1960). *In Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, I. Olkin et al. eds., Stanford University Press, pp. 278-292.

Loprinzi, C. L., Laurie, J. A., Wieand, H. S., Krook, J. E., Novotny, P. J., Kugler, J. W., ... & Klatt, N. E. (1994). Prospective evaluation of prognostic variables from patient-completed questionnaires. North Central Cancer Treatment Group. *Journal of Clinical Oncology*, *12*(3), 601-607.

Lothar S., (1997). *Angewandte Statistik.* Berlin: Springer. Pages: 395-397, 662-664.

Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The annals of mathematical statistics*, 50-60.

McKinney, W. (2012). *Python for data analysis: Data wrangling with Pandas, NumPy, and IPython*. " O'Reilly Media, Inc.".

Mukhiya S. K., Ahmed U. (2020). *Hands-On Exploratory Data Analysis with Python: Perform EDA techniques to understand, summarize, and investigate your data*. Packt Publishing Ltd.

Pratt, J. W. (1959). Remarks on zeros and ties in the Wilcoxon signed rank procedures. *Journal of the American Statistical Association*, *54*(287), 655-667.

Rajagopalan G. (2021). A Python Data Analyst's Toolkit: Learn Python and Python-based Libraries with Applications in Data Analysis and Statistics. *Switzerland: Springer International Publishing*.

Razali, N. M., & Wah, Y. B. (2011). Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. *Journal of statistical modeling and analytics*, *2*(1), 21-33.

Scheffé, H. (1953). A method for judging all contrasts in the analysis of variance. *Biometrika*, *40*(1-2), 87-110.

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, *52*(3/4), 591-611.

Stephens, M. A. (1976). Asymptotic results for goodness-of-fit statistics with unknown parameters. *The Annals of Statistics*, 357-369.

Tamhane, A. C. (1979). A comparison of procedures for multiple comparisons of means with unequal variances. *Journal of the American Statistical Association*, *74*(366a), 471-480.

Tukey, J. W. (1949). Comparing individual means in the analysis of variance. *Biometrics*, 99- 114.