

참고문헌

- 개인정보보호위원회(2022), 『가명정보 처리 가이드라인』
- 국세청(2021), 『국세 데이터 활용 안내서』
- 금융위원회·금융감독원(2022), 『금융분야 가명의명처리 안내서』
- 김승현(2020), 「통계적 노출 제어 방법 및 활용사례 연구: 재현데이터 방법론을 중심으로」, 『주택금융리서치』, 9호, pp. 18~35
- 김정연·박민정(2019), 「다중대체와 재현자료 작성」, 『응용통계연구』, 32(1), pp. 83~97
- 박민정·김항준(2016), 「마이크로데이터 공표를 위한 통계적 노출제어 방법론 고찰」, 『응용통계연구』, 29(6), pp. 1041~1059
- 박민정(2016), 「통계적 노출제어 최신동향 검토: 차등정보보호를 중심으로」, 2016년 하반기 연구보고서, 제 I 권
- _____ (2020), 「통계데이터센터 DB를 활용한 재현자료 생성 방법 연구」, 통계청 통계개발원 데이터 정보보호연구
- 유성준·박나리(2020), 「CART 기법을 이용한 개인신용정보 재현자료 생성기법」, 『통계연구』, 25(1), pp. 1~30
- 통계청(2021a), 『통계적 노출제어 가이드라인』
- _____ (2021b), 『통계의 창』, 여름호
- 한국신용정보원(2020), 『진짜 같은 가짜! 재현데이터의 개념 및 활용 사례』, 이슈리포트, 2019-8
- 홍동숙(2022), 『부도 예측을 위한 인공지능 학습용 데이터 생성 및 검증 기법: GAN (Generative Adversarial Network) 기반 재현데이터 중심으로』, 신용정보원 CIS 보고서
- Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., and Zhang, L.(2016), “Deep learning with differential privacy”, In Proceedings of the 2016 ACM SIGSAC conference on computer and communications security,

pp. 308~318

- Calacci, D., Berke, A., and Larson, K.(2019). "The tradeoff between the utility and risk of location data and implications for public good", arXiv preprint arXiv:1905.09350
- Che, Z., Cheng, Y., Zhai, S., Sun, Z., and Liu, Y.(2017), "Boosting deep learning risk prediction with generative adversarial networks for electronic health records", In 2017 IEEE International Conference on Data Mining(ICDM), IEEE, pp. 787~792
- Chen, R. J., Lu, M. Y., Chen, T. Y., Williamson, D. F., and Mahmood, F.(2021), "Synthetic data in machine learning for medicine and healthcare", *Nature Biomedical Engineering*, 5(6), pp. 493~497
- Chib, S.(1996), "Calculating posterior distributions and modal estimates in Markov mixture models", *Journal of Econometrics*, 75(1), pp. 79~97
- Choi, E., Biswal, S., Malin, B., Duke, J., Stewart, W. F., and Sun, J.(2017), "Generating multi-label discrete patient records using generative adversarial networks", In Machine learning for healthcare conference, PMLR
- Dandekar, A., Zen, R. A., and Bressan, S.(2018), "A comparative study of synthetic dataset generation techniques", In Database and Expert Systems Applications: 29th International Conference, DEXA 2018, Regensburg, Germany, September 3–6, 2018, Proceedings, Part II 29, Springer International Publishing, pp. 387~395
- Dwork, C., McSherry, F., Nissim, K., and Smith, A.(2006), "Calibrating noise to sensitivity in private data analysis". In Theory of Cryptography: Third Theory of Cryptography Conference, TCC 2006, New York, NY, USA, March 4-7, 2006. Proceedings 3, Springer Berlin Heidelberg, pp. 265~284
- Geman, D., Geman, S., Hallonquist, N., and Younes, L.(2015). "Visual turing test for computer vision systems", Proceedings of the National Academy of Sciences, 112(12), pp. 3618~3623
- Goodfellow, I., Bengio, Y., and Courville, A.(2016), *Deep learning*, MIT press

- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y.(2020), “Generative adversarial networks”, Communications of the ACM, 63(11), pp. 139~144
- Gouweleeuw, J. M., Kooiman, P., and De Wolf, P. P.(1998), “Post randomisation for statistical disclosure control: Theory and implementation”, *Journal of official Statistics*
- Jeong, D., Kim, J. H., and Im, J.(2022), “A new global measure to simultaneously evaluate data utility and privacy risk”, IEEE Transactions on Information Forensics and Security, 18, pp. 715~729
- Ji, Z., Lipton, Z. C., and Elkan, C.(2014), “Differential privacy and machine learning: a survey and review”, arXiv preprint arXiv:1412.7584
- Li, N., Li, T., and Venkatasubramanian, S.(2007), “t-closeness: Privacy beyond k-anonymity and l-diversity”, In 2007 IEEE 23rd international conference on data engineering, IEEE, pp. 106~115
- Li, P., Stuart, E. A., and Allison, D. B.(2015), “Multiple imputation: a flexible tool for handling missing data”, Jama, 314(18), pp. 1966~1967
- Little, R. J.(1993), “Statistical analysis of masked data”, *Journal of official statistics-stockholm-*, 9, pp. 407~407
- Machanavajjhala, A., Kifer, D., Gehrke, J., and Venkitasubramaniam, M.(2007), “l-diversity: Privacy beyond k-anonymity”, ACM Transactions on Knowledge Discovery from Data(TKDD), 1(1), 3~es
- Nowok, B., Raab, G. M., and Dibben, C.(2016). “synthpop: Bespoke creation of synthetic data in R”, *Journal of statistical software*, 74, pp. 1~26
- Park, M. J., and Kim, H. J.(2016), “Statistical disclosure control for public microdata: present and future”, *The Korean Journal of Applied Statistics*, 29(6), pp. 1041~1059
- Rubin, D. B.(1993), “Statistical disclosure limitation”, *Journal of official Statistics*, 9(2),

pp. 461~468

- Samarati, P., and Sweeney, L.(1998), “Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression”
- Snoke, J., Raab, G. M., Nowok, B., Dibben, C., and Slavkovic, A.(2018), “General and specific utility measures for synthetic data”, *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 181(3), pp. 663~688
- Templ, M., Meindl, B., Kowarik, A., and Chen, S.(2013), “Introduction to statistical disclosure control(sdc)”, Project: Relative to the testing of SDC algorithms and provision of practical SDC, data analysis OG
- Wei, J., Nie, Y., and Xie, W.(2020), “The Study of the Theoretical Size and Node Probability of the Loop Cutset in Bayesian Networks”, Mathematics, 8(7), p. 1079
- Xu, Lei, Maria Skoularidou, Alfredo Cuesta-Infante, and Kalyan Veeramachaneni(2019), “Modeling tabular data using conditional gan” Advances in Neural Information Processing Systems, 32

개인정보보호법(<https://www.law.go.kr/>)

금융위원회(<https://www.fsc.go.kr/index>)

대한민국 정책브리핑(<https://www.korea.kr/special/policyCurationView.do?newsId=148867915>)

KCB 데이터스토어(<https://datastore.koreacb.com/support/utilizeCaseView8.do>)

Gartner(<https://www.gartner.com/en>)

Google developers(https://developers.google.com/machine-learning/gan/gan_structure?hl=ko)

SDC practical guide(<https://sdcpractice.readthedocs.io/en/latest/index.html>)

<https://gretel.ai>

<https://limoss.london/how-does-sdc-work>

<https://elise-deux.medium.com/new-list-of-synthetic-data-vendors-2022-f06dbe91784>

<https://www.kaggle.com/datasets/kondla/carinsurance>

<https://www.etnews.com/20220315000132>

<https://open-innovations.org/events/synae/>

<https://www.statice.ai/case-study/die-mobiliar>

<https://www.geeksforgeeks.org/cart-classification-and-regression-tree-in-machine-learning/>