Applications (사례 연구)

Chanseok Park (박찬석)

Applied Statistics Laboratory
Department of Industrial Engineering
Pusan National University

August 5, 2020

Hosted by SEC

부산대학교
PUSAN NATIONAL UNIVERSITY

Overview

1. Robust statistics와 응용 사례
2. Missing · Incomplete Data와 응용 사례
3. Future work (missing with contamination)

Overview

1. Robust statistics와 응용 사례
2. Missing · Incomplete Data와 응용 사례
3. Future work (missing with contamination)

Overview

1. Robust statistics와 응용 사례
2. Missing · Incomplete Data와 응용 사례
3. Future work (missing with contamination)

Robust statistics와 응용 사례 (Basic applications)

Basic applications (estimating μ and σ)

Consider observations from $X_i \sim N(\mu, \sigma^2)$. We need to estimate μ and σ^2.

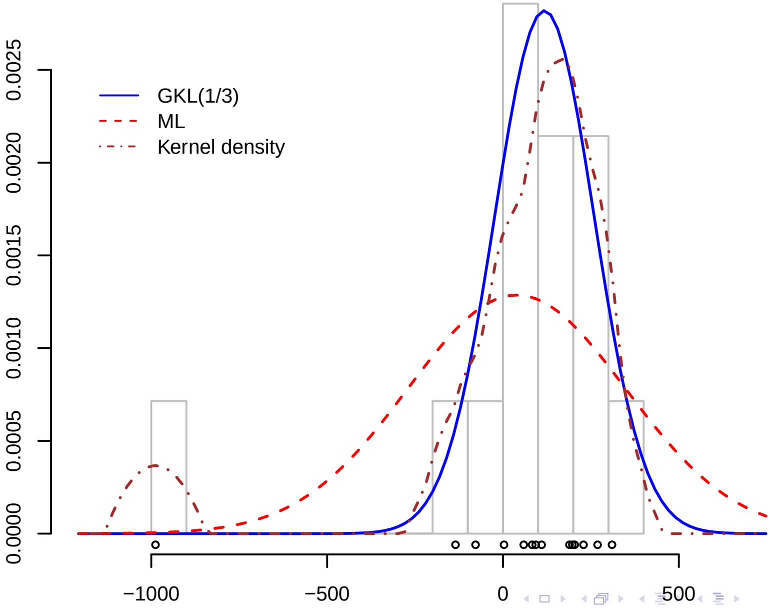
- MLE (maximum likelihood estimator)

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{and} \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \hat{\mu})^2$$

- BUE (best unbised estimator) or UMVUE

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{and} \quad S^2 = \frac{1}{\boldsymbol{n-1}}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

- MDE (minimum distance estimator): KL, GKL, etc.
  MLE is a special case of GKL (Basu et al., 2011) which can have robustness. MDE is asymptotically fully efficient, but its calculation is quite complex. Thus, HL and Shamos are recommended (Talk-2).


GKL(1/3)
ML
Kernel density
0.0000
0.0005
0.0010
0.0015
0.0020
0.0025
−1000
−500
0
500

Robust statistics와 응용 사례 (t-test)

We consider the ergonomic data set in Fernström and Ericson (1996). The data set was obtained from 16 workers. The amount of time, expressed as a proportion of total time, is observed during which arm elevation was below 30 degrees. The measurements were performed on two occasions separated by 18 months apart, that is, before and after a change of work organization. More more details, see Kim et al. (2018) and Park and Wang (2018b).

Subject	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Before	81	87	86	82	90	86	96	73	74	75	72	80	66	72	56	82
After	78	91	78	78	84	67	92	70	58	62	70	58	66	60	$\mathbf{65} + \delta$	73
Difference	3	-4	8	4	6	19	4	3	16	13	2	22	0	12	-9	9

We can test $H_0 : \mu_x = \mu_y \ \text{ and } \ H_1 : \mu_x \neq \mu_y$, equivalently, $H_0 : \mu_d = 0 \ \text{ and } \ H_1 : \mu_d \neq 0$, where $\mu_d = \mu_x - \mu_y$. This is a typical paired sample t-test with the test statistic given by

$$T = \frac{\bar{D} - 0}{S_D/\sqrt{n}},$$

where $\bar{D}$ is the sample mean of D_i and S_D is the sample standard deviation of D_i. Using D_i, this becomes a one-sample t-test.

Robust statistics와 응용 사례 (t-test)

Theorem 1 (Park, 2018)

Let $X_1, X_2, \ldots, X_n$ be a random sample from a normal distribution with mean μ and variance σ^2. Then we have

$$T_A = \sqrt{\frac{2n}{\pi}}\Phi^{-1}\left(\frac{3}{4}\right)\frac{\displaystyle\operatorname*{median}_{1\le i\le n} X_i - \mu}{\displaystyle\operatorname*{median}_{1\le i\le n}\left|X_i - \operatorname*{median}_{1\le i\le n} X_i\right|} \xrightarrow{d} N(0,1). \tag{1}$$

Theorem 2 (Jeong et al., 2018)

Let $X_1, X_2, \ldots, X_n$ be a random sample from a normal distribution with mean μ and variance σ^2. Then we have

$$T_B = \sqrt{\frac{3n}{2\pi}}\Phi^{-1}\left(\frac{3}{4}\right)\frac{\displaystyle\operatorname*{median}_{i\le j}\left(X_i + X_j\right) - 2\mu}{\displaystyle\operatorname*{median}_{i\le j}\left(\left|X_i - X_j\right|\right)} \xrightarrow{d} N(0,1). \tag{2}$$

Robust statistics와 응용 사례 (t-test)

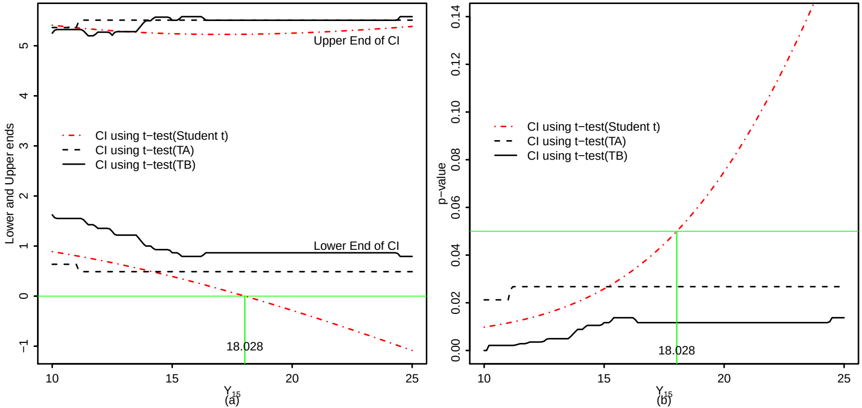
The above Theorems work well with a large sample size because these are based on asymptotic standard normal distribution.
Recently, Park and Wang (2018a) developed the `rt.test` R package used the empirical distributions instead of the asymptotic standard normal distribution. Using the `rt.test`, we can carry out robustified t-test easily.

This idea can be easily applied to control charting which is similar to a confidnce interval.

- Phase I: use robustifed control chart.
- Phase II: use conventional control chart.

Note: This work is in progress.

Robust statistics와 응용 사례 (Basic applications)



Missing data와 응용 사례

XXXX

- XXX
- XXX

Future work (missing with contamination)

XXXX

- XXX
- XXX

References

Basu, A., Shioya, H., and Park, C. (2011). Statistical Inference: The Minimum Distance Approach. Monographs on Statistics and Applied Probability. Chapman & Hall.

Fernström, E. A. C. and Ericson, M. O. (1996). Upper-arm elevation during office work. Ergonomics, 39(10):1221–1230.

Jeong, R., Son, S. B., Lee, H. J., and Kim, H. (2018). On the robustification of the z-test statistic. Presented at KIIE Conference, Gyeongju, Korea. April 6, 2018.

Kim, H., Park, C., and Wang, M. (2018). Paired t-test based on robustified statistics. Presented at KIIE Conference, Seoul, Korea. November 9.

Park, C. (2018). Note on the robustification of the Student t-test statistic using the median and the median absolute deviation. `https://arxiv.org/abs/1805.12256`. ArXiv e-prints.

References

Park, C. and Wang, M. (2018a). Empirical distributions of the robustified t-test statistics. `https://arxiv.org/abs/1807.02215`. ArXiv e-prints.

Park, C. and Wang, M. (2018b). `rt.test`: Robustified t-test. `https://CRAN.R-project.org/package=rt.test`. R package version 1.18.7.9.