The Power of Testing Parameter on the Specified Continuous Distributions

Budi Pratikno
Department of Mathematics
Faculty of Mathematics and Natural Science
Jenderal Soedirman University, Purwokerto, Indonesia,
Jl. Dr. Soeparno 61 Karangwangkal, Purwokerto, Banyumas, 53122, Central of Java, Indonesia
budi.pratikno@unsoed.ac.id

Alfa Annisa Pratama
Department of Mathematics
Faculty of Mathematics and Natural Science
Jenderal Soedirman University, Purwokerto, Indonesia,
Jl. Dr. Soeparno 61 Karangwangkal, Purwokerto, Banyumas, 53122, Central of Java, Indonesia
alfaannisaprat@gmail.com

Dian Permana
Department of Mathematics
Faculty of Mathematics and Natural Science
Jenderal Soedirman University, Purwokerto, Indonesia,
Jl. Dr. Soeparno 61 Karangwangkal, Purwokerto, Banyumas, 53122, Central of Java, Indonesia
dian.phw@gmail.com

Jajang
Department of Mathematics
Faculty of Mathematics and Natural Science
Jenderal Soedirman University, Purwokerto, Indonesia,
Jl. Dr. Soeparno 61 Karangwangkal, Purwokerto, Banyumas, 53122, Central of Java, Indonesia
jajang@unsoed.ac.id

Mashuri
Department of Mathematics
Faculty of Mathematics and Natural Science
Jenderal Soedirman University, Purwokerto, Indonesia,
Jl. Dr. Soeparno 61 Karangwangkal, Purwokerto, Banyumas, 53122, Central of Java, Indonesia
mashuri@unsoed.ac.id

Agustini Tripena Br.Sb.

Department of Mathematics
Faculty of Mathematics and Natural Science
Jenderal Soedirman University, Purwokerto, Indonesia,
Jl. Dr. Soeparno 61 Karangwangkal, Purwokerto, Banyumas, 53122, Central of Java, Indonesia
agustini.tripena@unsoed.ac.id

Abstract

The research studied the power on specified continuous distributions, namely Lognormal and Logistics distributions. Here, we derived the formula of the power of the Lognormal and Logistics distributions at least in two steps: (1) create sufficient statistic and rejection region for getting the bound of the area, and (2) derived the formula using the bound of the rejection area or using the basic concept of the definition of the power. To get the power and size, we here must define the hypothesis testing of the parameter shape in term of one-side hypothesis testing. This is due to the parameter shape are greater than zero. The Graphically analysed is given to make an easier interpretation, so a simulation is then given to figure their curve. R code is then used to compute and plot the curves. The result showed that Lognormal distribution depended on the degree of freedom n, bound of the rejection area and parameter shape (σ). Moreover, we also noted that the curves of the power are sigmoid and they increase (more) faster (going to be one) on the small parameter shape (σ) and large n. The size is constant and remain unchanged, and here the eligible size is 0.049 close to level of significance 0.05. So, we accepted this size as the minimum size (close to 0.05). In the context of the Logistics distribution, the result showed that the power of the Logistics distribution increase as the k increases, and the highestt curve occurs on large k ($k=10$), but not for the σ. Generally, the size is constant and it does not significantly change the curve on several k. Moreover, the size increase as the k increases. We noted here that the highest size occurs on $k=10$, and its value is around 1.0. This size is an impossible thing (not reasonable) to be used (far way from 0.05). However, we must choose the small size (less than 0.05), so we did not recommend this size. In this research, our target is to choose the maximum power and minimum size.

Keywords
Lognormal distribution, logistics distributions, the power function, hypothesis testing, R-code.

1. Introduction
In the statistics concept, there are three important concepts of the hypothesis testing in rejecting or accepting null hypothesis (H_0), namely (1) probability error type I (α), (2) a probability error type II (β) and (3) a power. One of them, the power, is a significant method to test the testing on the parameter shape or scale. Due to this, we then studied the power of the hypothesis testing on Lognorml distribution. Following Wackerly, et al. (2008), the power is defined as a probability to reject H_0 under H_1 in testing hypothesis $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$, for parameter θ. Similarly, the size is then defined as a probability to reject H_0 under H_0.

Many authors already discussed about the power in testing parameter shape on several distributions and or hypothesis testing, such as, Pratikno (2012) and Khan and Pratikno (2013), studied the power in testing intercept with non-sample prior information (NSPI). They used the probability integral of the cumulative distribution function (cdf) of the distributions to calculate the power. Moreover, Khan et al. (2016) used the power to compute the cdf of the bivariate noncentral F (BNCF) distribution in multivariate and multiple regression models. Here, we also noted that many authors, such as Khan (2003), Khan and Saleh [2008,1997, 2001], Khan and Hoque (2003), Saleh (2006), Yunus (2010), and Yunus and Khan [2007, 2011a, 2011b, 2011c], have contributed to the research of the power in the context of the estimator and hypothesis area. In the context of the hypothesis testing with NSPI on regression models, the BNCF distribution is used to compute the power using *R-code*. This is due to the computational of the probability integral of the probability distribution function (pdf) and cdf of the BNCF distribution are very complicated and hard. Here, we noted that the probability density function (pdf) and cumulative distribution function (cdf) really significant to use in computing the power and size function as well as on the Lognormal distribution.

From the above previous research, we then focused to do the similar research in testing parameter shape on the Lognormal and Logistics distributions. There are basically four steps to compute the power and size, that are: (1)

create the sufficiently statistics, (2) determine the rejection area using uniformly most powerful test (UMPT) or most powerful (MP) test, (3) derive the formula of the power, and (4) use a simulation using generate data from *R-code*. In this paper, the introduction is given in Section 1. An illustration in computing of the power is given in Section 2. The formula and graphically analysis of the power in testing parameter on the hypothesis are obtained in Section 3. The conclusion is provided in Section 4.

2. Method

There are basically four steps to compute the power and size, that are: (1) create the sufficiently statistics, (2) determine the rejection area using uniformly most powerful test (UMPT) or most powerful (MP) test, (3) derive the formula of the power, and (4) use a simulation using generate data from *R-code*. We have sometime the difficulties on the point (2), so we are then allowed to use the basic concept of the power, that is, a probability to reject H_0 under H_1.

3. Result

3.1. An Illustration to Compute the Power

Following previous research, we noted that the maximum power and minimum size of the tests are used to get the eligible testing. Here, the power is defined as a probability to reject H_0 under H_1 in testing parameter shape (or scale) on the hypothesis. Detail of the power and size on some continuous distributions and testing coefficient parameters on the regression model are found Pratikno's research. To illustrate the power function, we used case of the Binomial distribution, in testing $H_0 : p = p_0 = 0.3$ versus $H_1 : p > 0.3$, with rejection area $R = \{(x_1, x_2, ..., x_i) : Y \leq 4\}$ for several $m = 3, 4, 5$. The formula of the power function of this distribution is then given as $\pi(p) = P(\text{reject } H_0 | \text{Under } H_1 : p = p_1) = \sum_{y=0}^{4} \binom{m}{y} p_1^y (1-p_1)^{9-y}$, $m = 3, 4, 5$, and their simulation graphs are then given at Figure 1.
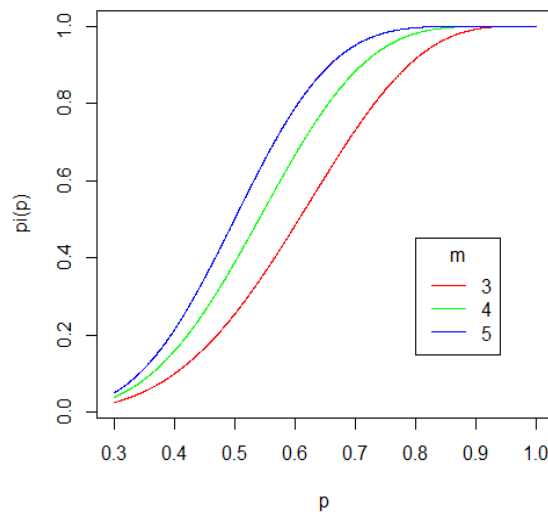

Figure 1. The power of Binomial distribution

From Figure 1, we see that the curve is sigmoid and tend to be one more faster for large *m*, we here then choose the blue (*m=5*) as a suitable curve then others. This is due to they are quickly to be one for small *p*. Therefore, we recommend them as the significant curves of the power function on the Binomial distribution.

3.2 The Graphs of the pdf and cdf of the Lognormal Distribution

Following Walpole, et al. (2012), Bain and Engelhardt(1992) and Lai and Balakrishnan (2009), the pdf of the random variable *X* of the Lognormal distribution is given as

$$f(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{1}{2\sigma^2}[\ln(x)-\mu]^2}, x \geq 0 \quad (1)$$ with parameter $-\infty < \mu < \infty$, and $0 < \sigma < \infty$. The X is

then notated as $X \sim LOGN(\mu, \sigma^2)$. By taking *ln* of the equation (1), we get $\ln(X) \sim N(\mu, \sigma^2)$. Furthermore,

the cdf of the lognormal distribution of $X \sim LOGN(\mu, \sigma^2)$ is written as

$$F_X(x) = P[X \leq x] = P[\ln X \leq \ln x] = \Phi\left(\frac{\ln x - \mu}{\sigma}\right) \quad (2)$$ where Φ is the cdf of the normal standard, X is *random*

variable of the lognormal distribution, and $\ln(X)$ is random variable of the normal distribution. Using equation (1), (2) and *R-code*, the graphs of the pdf and cdf are then produced in Figure 2.
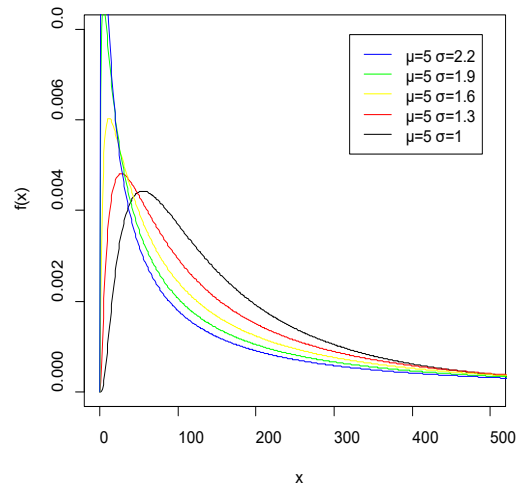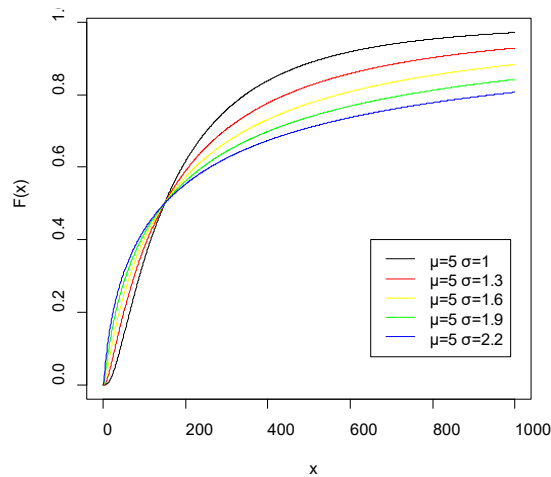

Figure 2.a.The pdf curve on several σ

Figure 2.b.The cdf curve on several σ

Figure 2. The pdf and cdf of Lognormal

From Figure 2., it is clear that the μ and σ are significant in affecting the skew-ness of the curve. The minimum curve is occurred for large μ and σ, it means the curve will be nearly normal. So, the skew-ness of the curves will be occurred for small μ and σ.

3.3 The Power and Size of the Lognormal Distribution

Following the process in deriving the formula of the power function the Section 2 and the steps of the research method on several case of the previous research, we then focused to conduct graphically analysis of the pdf, cdf and its power and size on lognormal distribution using the basic concept of the definition of the power. Here, we summarized the procedure in deriving the formula of the power as: (1) determine the sufficiently statistics, (2) create the rejection area using *uniformly most powerful test* (UMPT) or most powerful (MP) test, (3) derive the formula of the power of the lognormal distribution, and (4) finally, we simulate the graphs of the power using generate data on *R-code*.

Using the equation (1), the join distribution of $X_1, ..., X_n$ is then given as.

$$f\left(x_1,...,x_2|\mu\right) = \prod_{i=1}^{n} f\left(x_i|\mu\right) = \prod_{i=1}^{n} \left(\frac{1}{\sqrt{2\pi}\sigma x_i} e^{-\frac{1}{2\sigma^2}\left[\ln(x_i)-\mu\right]^2} \right)$$

$$= \frac{1}{\prod_{i=1}^{n} x_i \left(\sqrt{2\pi}\sigma\right)^n} e^{\frac{-\sum_{i=1}^{n}(\ln x_i)^2}{2\sigma^2}} e^{\frac{\sum_{i=1}^{n}\ln x_i \mu}{\sigma^2} - \frac{n\mu^2}{2\sigma^2}}. \tag{3}$$

Let, $s = \sum_{i=1}^{n} \ln X_i$, we then get $g\left(s|\mu\right) = e^{\frac{s\mu}{\sigma^2} - \frac{n\mu^2}{2\sigma^2}}$, where $h(x) = \frac{1}{\prod_{i=1}^{n} x_i \left(\sqrt{2\pi}\sigma\right)^n} e^{\frac{-\sum_{i=1}^{n}(\ln x_i)^2}{2\sigma^2}}$ and

$s = \sum_{i=1}^{n} \ln X_i \sim N\left(n\mu, n\sigma^2\right)$, is called sufficient statistics, with $\lambda(s)$ is monotone likelihood ratio (MLR).
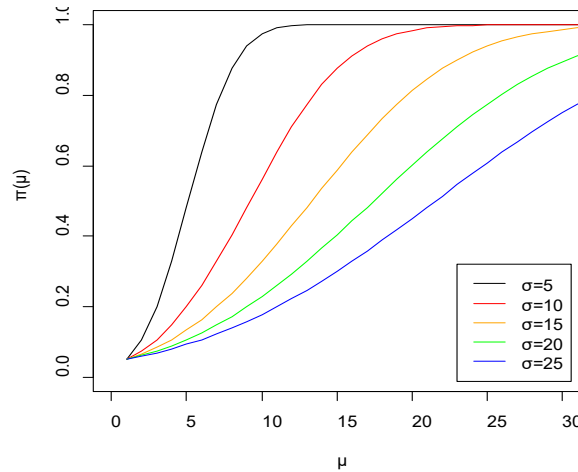
Using UMP test in testing $H_0 : \mu = \mu_0$ versus $H_1 : \mu > \mu_0$, we then reject H_0 when

$\sum_{i=1}^{n} \ln x_i \geq Z_\alpha \sigma \sqrt{n} + n\mu_0$. Based on this rejection area, we then derive the formula of the power as follows
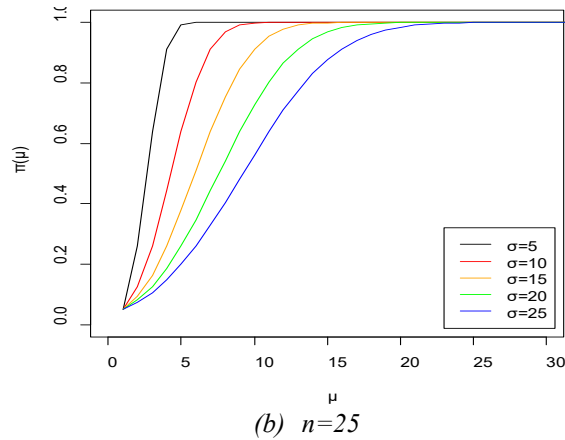
$$\pi(\mu) = P\left(\text{Re } ject \ H_0|\mu\right) = P\left(\sum_{i=1}^{n} \ln x_i \geq Z_\alpha \sigma \sqrt{n} + n\mu_0 \Big| \mu \right)$$

$$= 1 - P\left(\frac{\sum_{i=1}^{n} \ln x_i - n\mu}{\sigma\sqrt{n}} < Z_\alpha + \frac{\sqrt{n}\left(\mu_0 - \mu\right)}{\sigma} \right) \tag{4}$$

Using the equation (4), the graphs of the power function of the lognormal distribution at $\alpha = 0.05$ are presented at Figure 3.

(a) n=4

(b) n=25

Figure 3. The power curve of the lognormal

Figure 3. showed that the curves decrease as the σ increase, and the curves tend to be more faster to one when the n increases. Thus, we concluded that both n and σ significantly affect to the skew-ness of their curves. We here also compute the values of the size with different α^*, $\alpha = 0.01$ and $\alpha = 0.05$. Both size are constant, that are 0.10 and 0.049. Here, we must choose the minimum size (usually close, α *(alpha)*) .

From the simulation of the formula of the size, it is clear that the minimum size is around 0.049 and occurred on *alpha* is 0.05, but not on *alpha* 0.01 (the size is around 0.10). It means that we prefer to choose the minimum size when *alpha* is 0.05.

3.4 The Power and Size of the Logistics Distribution

The power function of the Logistics distribution in hypothesis testing $H_0 : \mu = \mu_0$ versus $H_1 : \mu > \mu_0$, under H_1, is given by

$$
\pi(\mu) = P(X \le k) = \int_{-\infty}^{k} f(x)dx
$$

$$
= \int_{-\infty}^{k} \frac{e^{-\left(\frac{\sum x_i - n\mu}{n\sigma}\right)}}{\prod_{i=1}^{n} \sigma\left(1 + e^{-\left(\frac{x_i - \mu}{\sigma}\right)}\right)^2} \, dx = \frac{1}{1 + e^{-\left(\frac{k-\mu}{\sigma}\right)}} \tag{5}
$$

Similarly, the size of the this distribution in testing $H_0 : \mu = \mu_0$ versus $H_1 : \mu > \mu_0$,, under H_0, is given as

$$
\alpha(\mu) = P(X \le k) = \int_{-\infty}^{k} f(x)dx
$$

$$
= \int_{-\infty}^{k} \frac{e^{-\left(\frac{\sum x_i - n\mu_0}{n\sigma}\right)}}{\prod_{i=1}^{n} \sigma\left(1 + e^{-\left(\frac{x_i - \mu_0}{\sigma}\right)}\right)^2} \, dx = \frac{1}{1 + e^{-\left(\frac{k-\mu_0}{\sigma}\right)}} \tag{6}
$$

Using the equation (5) and *R-code*, we then produced the plots of the power of the Logistics distribution at Figure 4 and Figure 5, respectively.
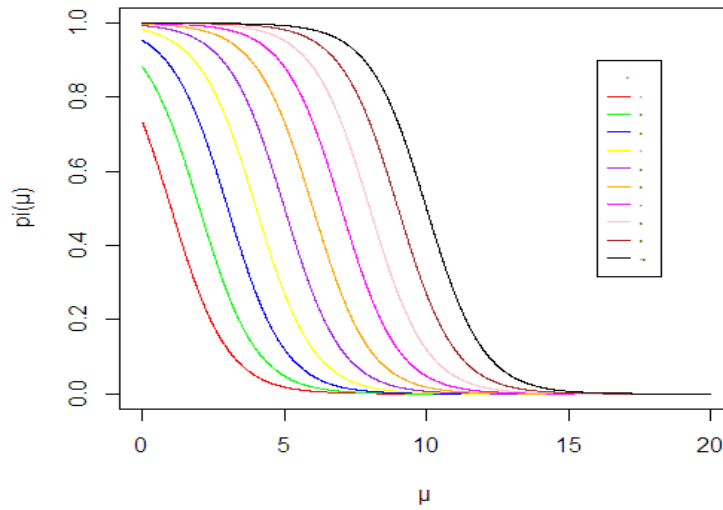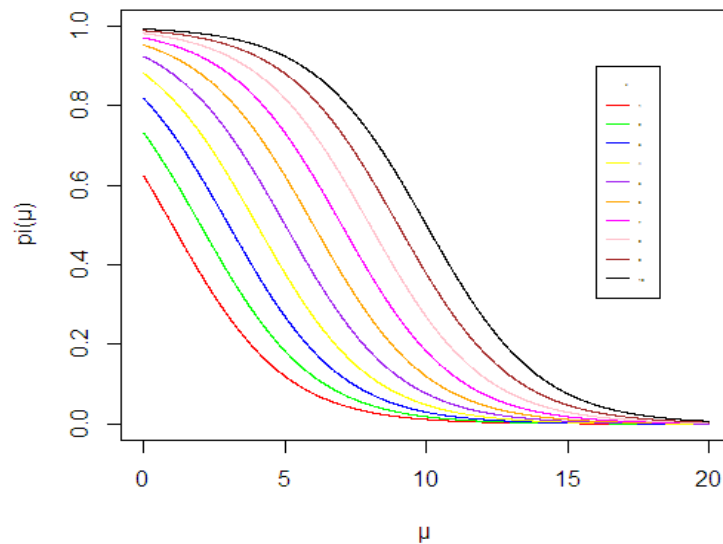
Figure 4. The *power* when $\sigma = 1$, *k =0,1,…,10*

Figure 5. The *power* when_$\sigma = 2$, *k =0,1,…,10*

From Figure 4-5, we see that the shape of the power depended on σ and *k*. Here, we noted that the curves increase as the *k* increases, and the highest curve occurs on large *k* (*k=10*). The power tend to be decrease as the σ increase. We then conclude that both *k* and σ really significant influenced to the shape of the power of the curves. Similarly, we then used equation (6) and *R-code* to produce the plots of the size of the Logistics distribution, respectively. From the simulation, we got that all the values of the size are constant. Here, we note that the maximum size is 0.10 and the minimum size is 0.05 for both $\sigma = 1$ and $\sigma = 2$. In this simulation, the **size** depended on σ and *k*, and it increases as the *k* increases (for small k, *k=1*, $\alpha = 0.05$ and the highest size occurs when *k=10*, $\alpha = 0.10$) , so it is an impossible (not reasonable) to be used (far way from 0.05). However, we must choose the small size (less than 0.05). Here, we have not got it yet, so we did not recommended this size.

Unlike previous research on some discrete or continuous distribution, here there is no specific distribution of the sufficiently statistics in finding the rejection area. We therefore derive the formula of the power using the Basic concept of the definition of the power, that is, the probability to reject H_o under H_1 .

4. Conclusion

The research studied the power and size of the hypothesis testing of the parameter shape on the Lognormal and Logistics distribution. We here then determined and graphically analysed the power and size on these distributions. There are four steps to derive the power and size functions as follow: (1) determine the sufficiently statistics, (2) compute the rejection region, (3) derive the formula of the power, and (4) create the graphs using generate data. The curve of the power is sigmoid and the curve increase faster (to be one) for small parameter shape (σ) and large n. In this research, the formula of the power is derived using at least the basic concept of the definition of the power. In this research, we used definition of the power on testing hypothesis of the parameter shape in term of one-side hypothesis. A simulation is then given to figure the power's and size's curve using R code. The result showed that Lognormal distribution depended on the degree of freedom n, bound of the rejection area, and parameter shape (σ). Moreover, the size is always constant and remain unchanged, and here the eligible size is 0.049 close to level of significance 0.05. So, we accepted this size as the minimum size (< 0.05). In the context of the Logistics distribution, the result showed that the power of the Logistics distribution increase as the k increases, and the highestt curve occurs on large k ($k=10$), but not for the σ . Generally, the size is constant and it does not significantly change the curve on several k. Moreover, the size increase as the k increases. We noted here that the highest size occurs on $k=10$, and its value is around 1.0. This size is an impossible thing (not reasonable) to be used (far way from 0.05). However, we must choose the small size (less than 0.05), so we did not recommend this size.

References

Bain, I.J., Mngelhardt, M. Introduction to Probability and Mathematical Statistics, 2th ed. USA: Duxbury Press., Belmont, CA, https://books.google.co.id/books/about/Introduction_to_Probability_and_Mathemat.html?hl=id&id=MkFRIAAACAAJ&redir_esc=y, 1992.

Khan, S., Pratikno, B. Testing Base Load with Non-Sample Prior Information on Process Load. Statistical Papers, 54 (3), 605-617, DOI: 10.1007/s00362-012-0451-40.81, 2013.

Khan, S., Pratikno, B., Ibrahim, A.I.N, Yunus, R.M. The correlated bivariate noncentral F distribution and Its application. Communications in Statistics—Simulation and Computation, 45 3491–3507, DOI: 10.1080/03610918.2014.999084, 2016.

Khan, S. Estimation of the Parameters of two Parallel Regression Lines Under Uncertain Prior Information. Biometrical Journal, **44**, 73-90, doi.org/ 10.1002/bimj.200290017, 2003.

Khan, S. Shrinkage estimators of intercept parameters of two simple regression models with suspected equal slopes. Communications in Statistics - Theory and Methods, **37**, 247-260, doi.org/10.1080/ 03610920701648961, 2008.

Khan, S., Saleh, A. K. Md. E. Shrinkage pre-test estimator of the intercept parameter for a regression model with multivariate Student-t errors. Biometrical Journal, **39**,1-17, doi.org/10.1001/bimj.4710380 202, 1997.

Khan, S., Saleh, A. K. Md. E. On the comparison of the pre-test and shrinkage estimators for the univariate normal mean. Statistical Papers, **42**(4), 451-473, doi.org/10.1007/s003620100073, 2001.

Khan, S., Hoque, Z. Preliminary test estimators for the multivariate normal mean based on the modified W, LR and LM tests. Journal of Statistical Research, Vol 37, 43-55, http://www.isrt.ac.bd/node/15, 2003.

Lai, C.D., Balakrishnan, N. Continous Bivariate Distributions, 2th ed., Springer, https://link.springer.com/book/10.1007/b101765, 2009.

Pratikno, B. Test of Hypothesis for Linear Models with Non-Sample Prior Information. Unpublished PhD Thesis, University of Southern Queensland, Australia, https://research.usq.edu.au/download/fd27eeffb76c7e0a66d0e8e885a132043def7de51f760e23e7fac7d511f4348a/904968/Pratikno_2012_whole.pdf., 2012.

Saleh, A. K. Md. E. Theory of preliminary test and Stein-type estimation with applications. John Wiley and Sons, Inc., New Jersey, DOI:10.1002/0471773751, 2006.

Yunus, R.M. Increasing power of M-test through pre-testing. Unpublished PhD Thesis, University of Southern Queensland, Australia, https://www.cambridge.org/core/journals/bulletin-of-the-australian-mathematical-society/article/increasing-power-of-mtest-through-pretesting/862C81029B3D370918414B10520B7CC1 , 2010.

Yunus, R.M., Khan, S. Test for intercept after pre-testing on slope a robust method. In: 9th Islamic Countries Conference on Statistical Sciences (ICCS-IX): Statistics in the Contemporary World - Theories, Methods and Applications,https://www.researchgate.net/publication/228779650_Test_for_intercept_after_pretesting_on_slope-a_robust_method, 2007.

Yunus, R.M., Khan, S. Increasing power of the test through pre-test a robust method. Communications in Statistics-Theory and Methods, **40**, 581-597, 2011a, https://doi.org/10.1080/03610920903350572, 2011a.

Yunus, R.M., Khan, S. M-tests for multivariate regression model. Journal of Nonparamatric Statistics, **23**, 201-218, https://doi.org/10.1080/ 10485252.2010.503896, 2011b.

Yunus, R.M., Khan, S. The bivariate noncentral chi-square distribution - Acompound distribution approach. Applied Mathematics and Computation, **217**, 6237-6247, doi.org/10.1016/j.amc.2010.12.112 , 2011c.

Wackerly, D.D., Mendenhall III, W, R. L.Scheaffer. Mathematical Statistics With Application, 7th Ed. Thomson Learning, Inc., Belmont, CA, USA, https://books.google.co.id/books/about/ Mathematical_Statistics_with_Application.html?id=lTgGAAAAQBAJ&redir_esc=y , 2008.

Walpole, R.E., Myers, R.H., Myers, S.L., Ye, K. Probability and Statistics in Engineers and Scientist. Prentice Hall, Perason., https:// spada.uns.ac.id/ pluginfile.php/221008/mod_resource/content/1/ ProbabilityStatistics_for_EngineersScientists%289th_Edition%29_Walpole.pdf , 2012.

Biographies

Budi Pratikno, currently works as a lecturer in Department of Mathematics, Universitas Jenderal Soedirman, author of several international journals related to testing intercepts with non-sample prior information (NSPI), which are widely published in international journals such as Statistical Papers, Far East Journal of Mathematics Science, Springer, JSTA, ISSOS, IJET, IJAST, IOP series and others. In the study of NSPI as the main research the author, we try to improve population inference using non-sample. The author completed his undergraduate education at the Mathematics Department (UGM), then continued his Masters in Statistical Science at La Trobe University, Melbourne, Australia, and his S3 in Statistical Science (Ph.D) at the University of Southern Quensland (USQ), Toowoomba, Brisbane, Australia.

Alfa Annisa Pratama is a student in Department of Mathematics, Universitas Jenderal Soedirman.

Dian Permana is a student in Department of Mathematics, Universitas Jenderal Soedirman.

Jajang, currently works as a lecturer in Department of Mathematics, Universitas Jenderal Soedirman. The author completed his undergraduate education in Statistics at the IPB University, then continued his Masters in Statistical Science at at the IPB University, and his Doctor at at the IPB University.

Mashuri currently works as a lecturer in Department of Mathematics, Universitas Jenderal Soedirman. The author completed his undergraduate education in Statistics at the UNPAD University, then continued his Masters at at the ITB University, and his Doctor at at the ITB University.

Agustini Tripena Br.Sb currently works as a lecturer in Department of Mathematics, Universitas Jenderal Soedirman. The author completed his undergraduate education in Statistics at the USUUniversity, then continued his Masters at at ITS University.