FORENSIC SCIENCE JOURNAL SINCE 2002

Forensic Science Journal
2023;22(1):1-20
fsjournal.cpu.edu.tw
DOI:10.6593/FSJ.202312_22(1).0001

# Improvement of Principal Component Analysis (PCA) by Using Log-transformed Fermentation Congeners for the Authentication of Scotch Whiskies

**Hsiao-Wen Huang [1], M.S.; Wei-Tun Chang [1*], Ph.D**

*[1] Department of Forensic Science, Central Police University, No.56, Shuren Rd., Guishan Dist., Taoyuan City 333322, Taiwan (R.O.C.)*

Received: October 8, 2022; Accepted: January 30, 2023

## Abstract

In order to improve the authentication effectiveness by principal component analysis (PCA), 52 single malt Scotch whiskies, 31 blended Scotch whiskies, and 3 adulterated Scotch whiskies were analyzed to select appropriate fermentation congeners with gas chromatography-mass spectrometry (GC-MS). After the evaluation process, six fermentation congeners, including acetaldehyde, ethyl acetate, ethyl octanoate, 2-methyl propanol, 2-methyl butanol, and 3-methyl butanol, were selected based on the total explained variances results of PCA. Subsequently, the distribution fitting method demonstrated that the data distribution of these six fermentation congeners more suitably corresponded to the lognormal distribution for both authentic single malt Scotch whiskies and authentic blended Scotch whiskies. Along with comparing original data and log-transformed data of the integrated value of selected fermentation congeners, PCA charts with log-transformed data were demonstrated to express better exclusory authentication results for single malt Scotch whiskies. For blended Scotch whiskies, both PCA and LDA, with log-transformed data, can enhance the discrimination of adulterated samples. As a result, it was verified that PCA with log-transformed data could effectively improve the discriminating power in determining the authenticity of adulterated Scotch whiskies.

***Keywords:** forensic science, authentication, adulterated Scotch whisky, fermentation congeners, normality test, lognormal distribution, principal component analysis (PCA), gas chromatography-mass spectrometry (GC-MS)*

## Introduction

Based on the dietetic culture, the Taiwanese's growing preference for Whisky has been driving the imported market to supply more Scotch whiskies. Scotch whisky is one of the most popular distilled alcoholic beverages in the world, which must only be made in Scotland and manufactured following the rigorous regulation by the statutory instrument of Scotland [1]. Taiwan was ranked as the 3rd of the top 10 Scotch whisky export markets by value in 2021, up from 4th in 2019, according to the Scotch whisky Association report [2]. In light of its high market value, Scotch whiskies have become subject to illegal adulterated and counterfeited. Therefore, consumer protection agencies and producers eagerly require a more valuable and efficient method to support authenticity analyses.

In general principle, there are six processes to make Scotch whiskies in the manufacture protocol, including malting, mashing, fermentation, distilling, and

* Corresponding author: Wei-Tun Chang, Department of Forensic Science, Central Police University, No.56, Shuren Rd., Guishan Dist., Taoyuan City 333322, Taiwan (R.O.C.)
Fax: 886-3-3284118.
E-mail: una030@mail.cpu.edu.tw

maturation, sequentially. As the process of whisky wort fermentation by distiller's yeast, it produces not only ethanol (EtOH) as a major product but also a variety of essential congeners, such as aldehydes, higher alcohols, organic acids, and then esters, as flavor compounds. The fermentation process of whisky is similar to beer fermentation, but the wort used in the process is not boiled. Therefore, the enzymes in the wort can continue to react with the carbohydrate molecules/oligosaccharides to increase the yields of congeners [3]. Many researchers, mainly in genetic and proteomic fields [4-7], have deduced the pathway for showing the syntheses of higher alcohols and their esters. These studies discussed the knowledge of the pathways involving the synthesis of higher alcohols and esters by yeasts and fermentation parameters during biosynthesis. Also, some instrumental analysis has investigated the fermentation congeners to differentiate the counterfeit whiskies by gas chromatography (GC), gas chromatography-mass spectrometry (GC-MS), and liquid chromatography-quadrupole-orbitrap mass spectrometry (LC-Q-Orbitrap) [8-10], but rare of them studied the forensic application based on statistical characteristics via data distribution principles for the authentication purpose.

According to the literature survey related to principal component analysis (PCA) applications, most studies have merely described their analytical results via statistical methods [8, 11-16]. However, they rarely discussed how to improve PCA results deriving from data distribution fitness. A normal distribution is a widely used statistical model to describe observations for the natural and social sciences. Over the past decades, however, more and more studies have revealed that the observed data for most natural and technological processes are lognormal while showing a skewed distribution pattern [17]. For many studies, especially in earth and environmental science, lognormal distributions are observed to describe the random variation or mixing factors that occur in the data [18-20]. For example, ascribing a series of complex chemical reactions in organic substrates to fermentation, the concentration of fermentation congeners in Scotch whiskies might not distribute as normal distribution.

According to ILAC-G19:2002, Guidelines for Forensic Science Laboratories, reference collections of data should be uniquely identified and controlled for identification, comparison, or interpretation purposes [21]. However, it is not practical to collect all reference samples for authentication of adulterated or counterfeited Scotch whiskies. It is necessary to figure out a proper statistical distribution model to describe the data population of authentic Scotch whiskies. One such log-normal distribution of methanol (MeOH) has already been identified that the concentration of MeOH could serve as an exclusion marker to distinguish authentic Scotch whiskies without any other reference samples of authentic whisky in our previous work [22]. Moreover, the analysis of $\delta^{13}$C-ethanol distribution also found the differential between the authentic Scotch whiskies and rectified spirits to authenticate the seized Scotch whiskies via establishing criteria from the log-normal distribution of authentic samples [23].

In this study, we tried to investigate the log-normal distribution of fermentation congeners and develop a method to improve the discriminating power of PCA for the authentication of Scotch whiskies. Thus, a set of model fitting procedures for fermentation congeners was conducted on both authentic single malt Scotch whiskies and authentic blended Scotch whiskies to receive the appropriate statistical models. Finally, we compared the PCA charts with log-transformed and original data to evaluate authentication effectiveness in both single malt Scotch whiskies and blended Scotch whiskies.

## Materials and methods

### *Preparation of Whisky Samples*

The analytical samples consisted of 83 authentic Scotch whiskies and 3 adulterated whisky samples were measured by GC-MS in this study. The authentic samples, including 52 single malt Scotch whiskies and 31 blended Scotch whiskies imported from Scotland to Taiwan, were purchased from legal tobacco and liquor stores in Taiwan. Three adulterated samples, including one adulterated single malt Scotch whisky and two adulterated blended Scotch whiskies, were seized by the police agency in Taiwan.

Before measurement, all the samples were added into the same amount of internal standard (4-methyl-2-pentanol), which is prepared as 0.2%(v/v) in 40% Ethanol/DIW solution, to acquire the ratios of fermentation congeners/internal standard for the sequent principal component analysis.

### *GC-MS Analysis*

All Scotch whisky samples with the internal standard were analyzed by Agilent gas chromatography 6890 coupled to a quadrupole mass spectrometer 5973N (GC-MS) with full scan mode (m/z:35 to 300). The GC was equipped with an Agilent J&W CP-Wax 57 CB fused capillary column (50m×0.25mmID×0.2μm) and the carrier gas was helium at the flow rate of 1.0mL/min. The analytical parameters of the GC/MS system were set as follows: injection port and detector: 230°C and split injection mode (split ratio 1/10); oven temperature: initially 35°C for 2min, elevated to 80°C at 3°C min-1 and held 20min, and then 200°C at 25°C min-1 and maintained for 3 min. An example of the TIC chromatogram of GC-MS for authentic Scotch whiskies, which labels selected targets, is shown in Fig.1.

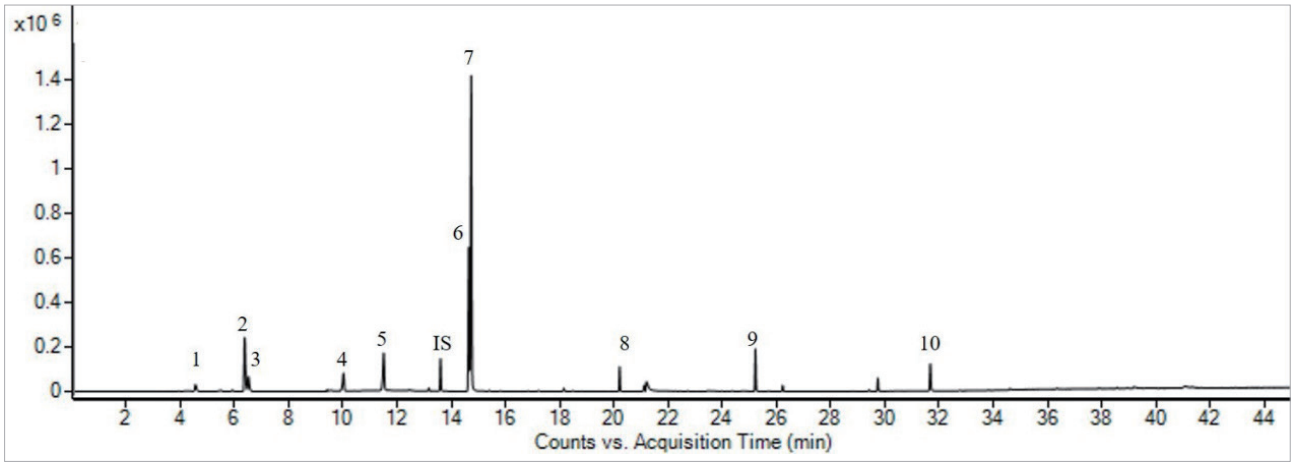


**Fig. 1** The TIC chromatogram of authentic Scotch whiskies in GC-MS. Peaks selected in this work: 1: acetaldehyde; 2: ethyl acetate; 3: acetal; 4: 1-propanol; 5: 2-methyl propanol; 6: 2-methyl butanol; 7: 3-methyl butanol; 8: ethyl octanoate; 9: ethyl decanoate; 10: 2-phenyl ethanol; IS: internal standard (4-methyl-2-pentanol).

### *Statistical Analysis of PCA and Model Fitting Process*

The evaluation process for the statistical analysis of PCA in this study is shown in the following descriptions. The integrated ratios of selected congeners/IS measured by GC-MS for 52 bottles of authentic single malt Scotch whiskies and 31 bottles of authentic blended Scotch whiskies were first plotted as a histogram. Then, to determine suitable distribution, the integrated area ratio of 10 components/IS, whose signal-to-noise (S/N) is higher than 10, from the TIC chromatogram of GC-MS was considered based on the definition of LOQ [24] and adopted to conduct the PCA process. After this evaluation step, it was found that the cumulative total explained variances would be higher than 80% for both single malt Scotch whiskies and blended Scotch whiskies, while the total explained variances were established by the six fermentation congeners which were acetaldehyde, ethyl acetate, 2-methyl propanol, 2-methyl butanol, 3-methyl butanol, and ethyl octanoate.

The total explained variances and their cumulative values of PCA extraction is described in Table 1. For single malt Scotch whiskies with original data, the ratio of the variance of PC1 and PC2 are 44.303 and 39.244. With log-transformed data, the variance ratio of PC1 and PC2 are 42.266 and 41.202, respectively. Therefore, the cumulative explained variance for raw and log-transformed data of single malt Scotch whiskies explains more than 83% of variances. However, for the blended Scotch whiskies, the ratio of the cumulative explained variance to the log-transformed data is 90.771%, which is higher than the ratio of the original data. To clarify the statistical characteristics of these selected fermentation congeners, these six components were further analyzed by the model fitting procedures and the normality test. Finally, PCAs with original data and transformed data were both discussed to compare the discriminating power for authentication of Scotch whiskies.

**Table 1** The total explained variance and their cumulative explained variance of PCAs in Scotch whiskies.

| | | with original data | | with log-transformed data | |
|---|---|---|---|---|---|
| | | PC1 | PC2 | PC1 | PC2 |
| single malt Scotch whisky | % of variance | 44.303 | 39.244 | 42.266 | 41.202 |
| | cumulative explained variance (%) | 44.303 | 83.547 | 42.266 | 83.457 |
| blended Scotch whisky | % of variance | 72.209 | 16.698 | 73.434 | 17.337 |
| | cumulative explained variance (%) | 72.209 | 88.907 | 73.434 | 90.771 |

The statistical analysis was executed by two software, SPSS(IBM) and SPC (BPI Consulting, LLC). First, considering the rigor evaluation to deduce the suitable distribution model, the software SPC checked all data for their distribution fitting results. Afterward, the Anderson-Darling test was performed to confirm the normality of fitting distribution models with a p-value of >0.05, which was considered to confirm normality. The normality test on the log scale is shown in Table 2. Also, for comparison, the normality test results of fermentation congeners with original data of authentic Scotch whiskies were listed in Table 3.

As shown in Table 3, for the data on the original scale of single malt Scotch whisky, the AD test demonstrated that those of ethyl acetate and 2-methyl propanol did not pass the normality examination, whose p-values <0.05. As a result, the AD test certified that the data of these two components on the original scale of single male Scotch whisky did not belong to normal distribution. Also, while discussing the data on the original scale of blended Scotch whisky, the p values of acetaldehyde, 2-methyl butanol, 3-methyl butanol, and ethyl octanoate were all <0.05. Therefore, for blended Scotch whisky, it is certified that the data distributions of these four components did not belong to the normal one.

On the other hand, the p-values for the selected six fermentation congeners with log-transformed data for both single malt Scotch whisky and blended Scotch whisky were all above 0.05. These results imply that the data of these components on the log scale belonged to a normal distribution, as shown in Table 2.

**Table 2** The test results of normality for fermentation congeners with log-transformed data in authentic Scotch whiskies.

| Anderson-Darling Test | | | |
|---|---|---|---|
| with log-transformed data | | single malt Scotch whisky | blended Scotch whisky |
| acetaldehyde | Statistic | 0.537 | 0.643 |
| | df | 52 | 31 |
| | Sig. | 0.161 | 0.086 |
| ethyl acetate | Statistic | 0.536 | 0.322 |
| | df | 52 | 31 |
| | Sig. | 0.162 | 0.514 |
| 2-methyl propanol | Statistic | 0.307 | 0.233 |
| | df | 52 | 31 |
| | Sig. | 0.292 | 0.789 |
| 2-methyl butanol | Statistic | 0.543 | 0.467 |
| | df | 52 | 31 |
| | Sig. | 0.155 | 0.234 |
| 3-methyl butanol | Statistic | 0.209 | 0.578 |
| | df | 52 | 31 |
| | Sig. | 0.856 | 0.122 |
| ethyl octanoate | Statistic | 0.232 | 0.425 |
| | df | 52 | 31 |
| | Sig. | 0.792 | 0.298 |

**Table 3** The test results of normality for fermentation congeners with original data in authentic Scotch whiskies.

| Anderson-Darling Test | | | |
|---|---|---|---|
| with original data | | single malt Scotch whisky | blended Scotch whisky |
| acetaldehyde | Statistic | 0.467 | 0.955 |
| | df | 52 | 31 |
| | Sig. | 0.241 | 0.014 |
| ethyl acetate | Statistic | 1.018 | 0.540 |
| | df | 52 | 31 |
| | Sig. | 0.010 | 0.155 |
| 2-methyl propanol | Statistic | 1.014 | 0.557 |
| | df | 52 | 31 |
| | Sig. | 0.011 | 0.140 |
| 2-methyl butanol | Statistic | 0.680 | 1.059 |
| | df | 52 | 31 |
| | Sig. | 0.072 | 0.008 |
| 3-methyl butanol | Statistic | 0.216 | 0.925 |
| | df | 52 | 31 |
| | Sig. | 0.837 | 0.017 |
| ethyl octanoate | Statistic | 0.210 | 2.083 |
| | df | 52 | 31 |
| | Sig. | 0.853 | 0.000 |

## Results and Discussion

### *Model Fitting Results of the Fermentation Congeners in Authentic Scotch whiskies*

A series of experiments are performed to measure the fermentation congeners of authentic Scotch whiskies. Due to the chromatographic characteristics, the concentration of fermentation congeners is proportional to the integrated area of signals. Therefore, integrated ratio analysis is adopted in this work in order to rapidly screen the fermentation congeners suitable for evaluating the distribution.

As regards the suitable statistical modeling to describe the fermentation congeners, the six fermentation congeners selected from the PCA evaluation method, including acetaldehyde, ethyl acetate, 2-methyl propanol, 2-methyl butanol, 3-methyl butanol, and ethyl octanoate, were discussed in the chemical route and operated a series of the statistical model fitting process.

*1. Aldehyde in Authentic Scotch whiskies*

In the whisky fermentation process, acetaldehyde formed as the side product of amino acid synthesis [25-26]. In other words, it is produced as a result of the decarboxyl of pyruvate, as shown in Fig. 2 This suggests that this enzymic mechanism might push the concentration of acetaldehyde towards other distributions due to kinetic effects.

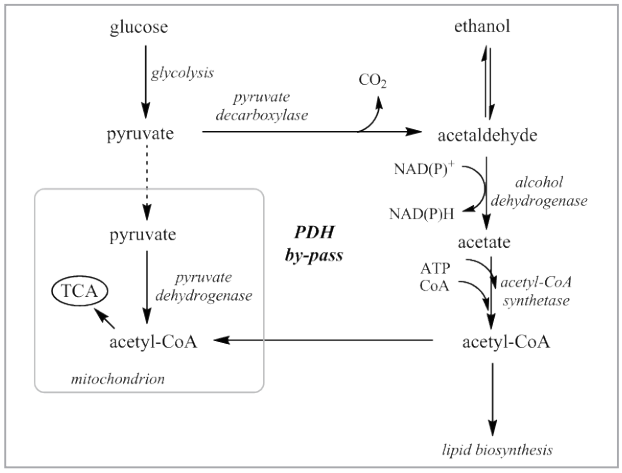


**Fig. 2** The production route of acetaldehyde during fermentation [26].

Fig. 3(a) and Fig. 4(a) are the raw data histograms of the integrated ratio of acetaldehyde/IS in authentic single malt Scotch whiskies and authentic blended Scotch whiskies, respectively. In Fig. 3(a), the data distribution displays a slightly positive skewness with a value of 0.2347, implying that the integrated value distribution of acetaldehyde/IS may not only fit the normal distribution. Therefore, the measurement value is evaluated under a series of data distribution processes and replot the distribution with both original data and log-transformed data to verify the distribution correctness, as shown in Fig. 3(b) and Fig. 3(c). As a result, it is demonstrated that the integrated ratio of acetaldehyde/IS in authentic single malt Scotch whiskies conformed not only to a normal distribution but also to a lognormal distribution. Moreover, in the case of authentic blended Scotch whiskies, as in Fig. 4(a), the skewness value is higher than that of Fig. 3(a), suggesting that the integrated value of acetaldehyde/IS may be distributed rather other distribution than a normal one. After the same evaluation process as the previous description, Fig. 4(b) and Fig. 4(c) demonstrate that the integrated value distribution of acetaldehyde/IS in authentic blended Scotch whiskies belongs to the lognormal distribution.

In order to confirm the correctness of the data fitting procedure, the P-P plot method is employed to present the visual comparison of two cumulative distribution functions against each other. As shown in Fig. 3(d) and Fig. 4(d), a straight diagonal line verified that the fitting distribution result of acetaldehyde in authentic single malt Scotch whiskies and authentic blended Scotch whiskies correspond to a lognormal distribution.

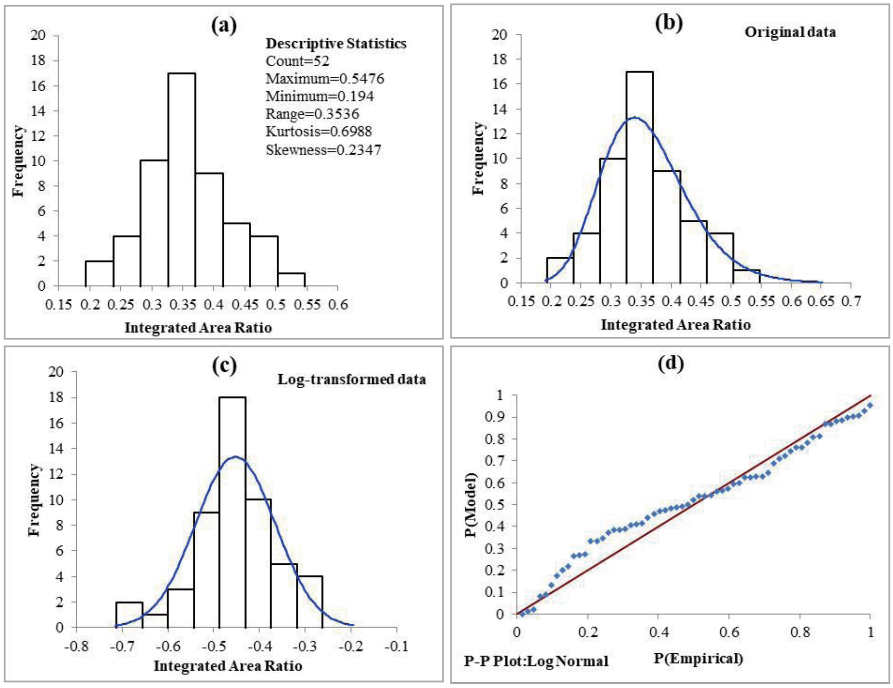


**Fig. 3** Statistical analysis of acetaldehyde/IS of authentic single malt Scotch whiskies. (a) The histogram of the raw data. (b) The distribution fitting results showed that the original data conformed with a lognormal distribution. (c) The distribution fitting results showed that the log-transformed data conformed with a normal distribution. (d) The P-P plot of distribution fitting result.

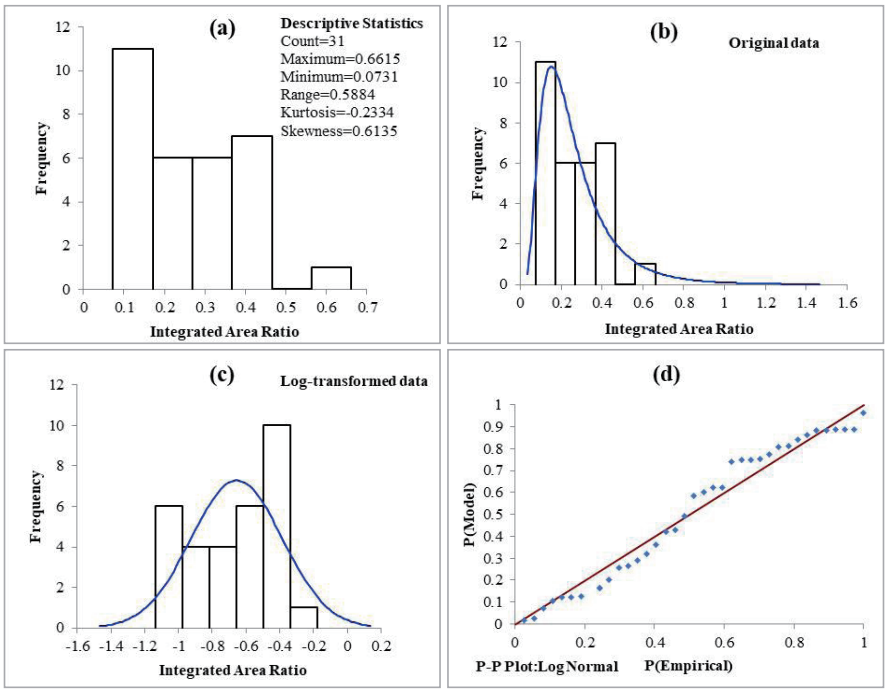


**Fig. 4** Statistical analysis of acetaldehyde/IS of blended Scotch whiskies. (a) The histogram of the raw data. (b) The distribution fitting results showed that the original data conformed with a lognormal distribution. (c) The distribution fitting results showed that the log-transformed data conformed with a normal distribution. (d) The P-P plot of distribution fitting result.

*2. Ester in Authentic Scotch whiskies*

Esters are compounds formed during yeast fermentation by the enzymatic condensation of organic acids with alcohols and distribute a flowery or fruit-like aroma to spirits. Ethyl acetate is the principal ester formed by yeast, principally by enzymic reactions during fermentation [5, 7]. In clarity, chemical reactions form ethyl acetate between ethanol and acetic acid. The chemical pathway of esters is described in Fig. 5. In this work, two esters, which are ethyl acetate and ethyl octanoate, were found suitable to establish the distribution models.

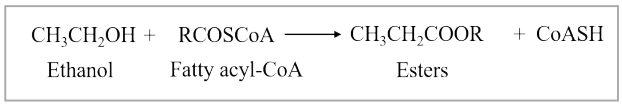


**Fig. 5** The production route of esters during fermentation [25].

The raw data histograms, plotted in Fig. 6(a) and Fig. 7(a), are the integrated ratio of ethyl acetate/IS in authentic single malt Scotch whiskies and authentic blended Scotch whiskies, respectively. As shown in Fig. 6(a), the obvious skewness plot with a value of 0.7405 shows that the integrated value distribution of ethyl acetate/IS is highly possible to correspond to other distributions rather than to normal distribution. After operating the evaluation process established in this study, the corresponding distributions with both the original data and log-transformed data are displayed in Fig. 6(b) and 6(c). These two figures show strong evidence that the distribution of the integrated ratio of ethyl acetate/IS in authentic single malt Scotch whiskies is lognormal due to the long tail of figures representing the higher frequencies of greater abundances.

Similarly, as in Fig. 7(b). and Fig. 7(c), the integrated value of ethyl acetate/IS in authentic blended Scotch whiskies is also distributed as a lognormal distribution. The P-P plots of ethyl acetate, either for authentic single malt Scotch whiskies or authentic blended Scotch whiskies, confirm the fitting distributions as lognormal distributions.

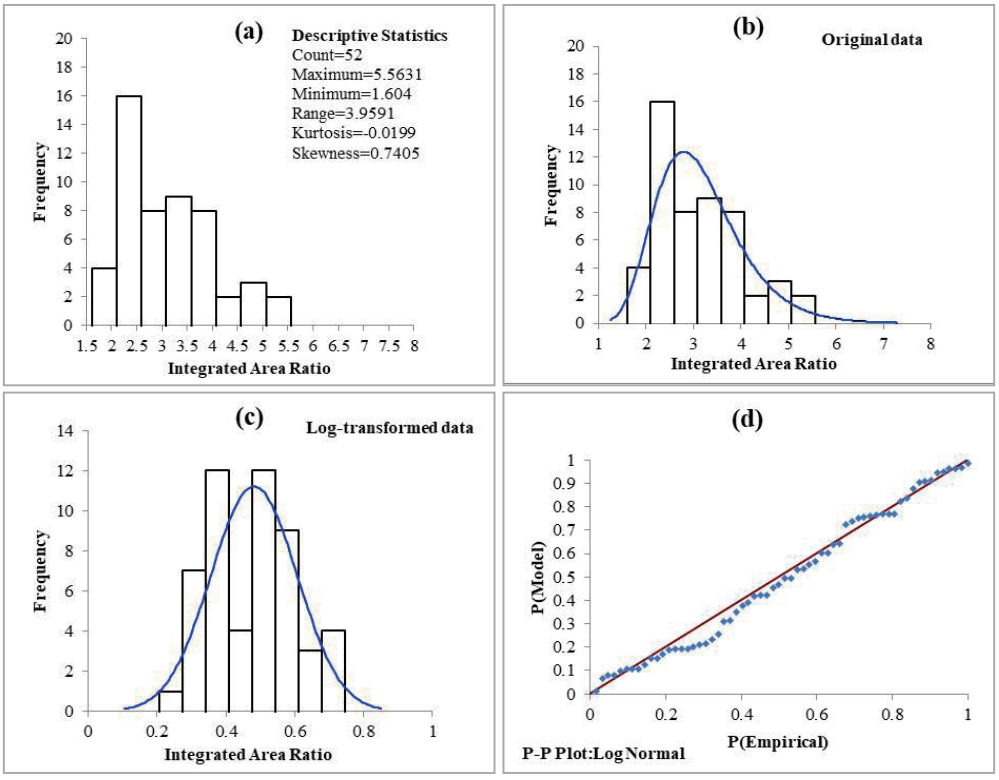


**Fig. 6** Statistical analysis of ethyl acetate/IS of authentic single malt Scotch whiskies. (a) The histogram of the raw data. (b) The distribution fitting results showed that the original data conformed with a lognormal distribution. (c) The distribution fitting results showed that the log-transformed data conformed with a normal distribution. (d) The P-P plot of distribution fitting result.

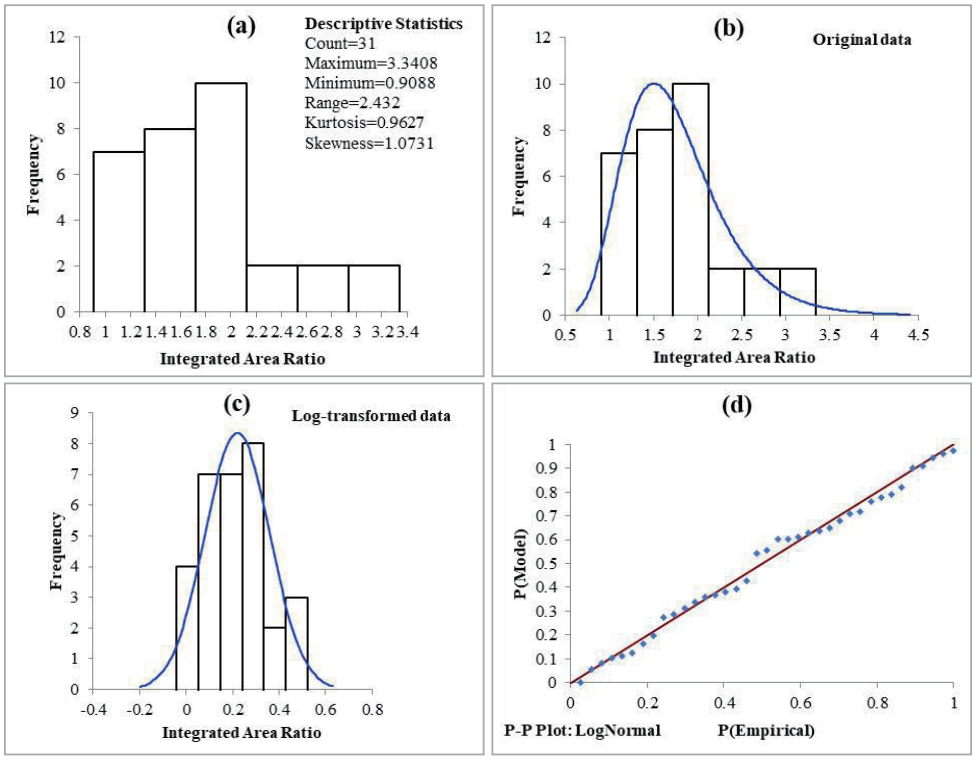


**Fig. 7** Statistical analysis of ethyl acetate/IS of authentic blended Scotch whiskies. (a) The histogram of the raw data. (b) The distribution fitting results showed that the original data conformed with a lognormal distribution. (c) The distribution fitting results showed that the log-transformed data conformed with a normal distribution. (d) The P-P plot of distribution fitting result.

For ethyl octanoate, the statistical evaluation methods used to analyze ethyl acetate are also employed to study its integrated ratio. The results shown in Fig. 8 and Fig. 9 indicated that the integrated ratio of ethyl octanoate/IS in authentic single Scotch whiskies and authentic blended Scotch whiskies are distributed as a lognormal distribution which reflects the multifactorial processes during synthesis, as expected.

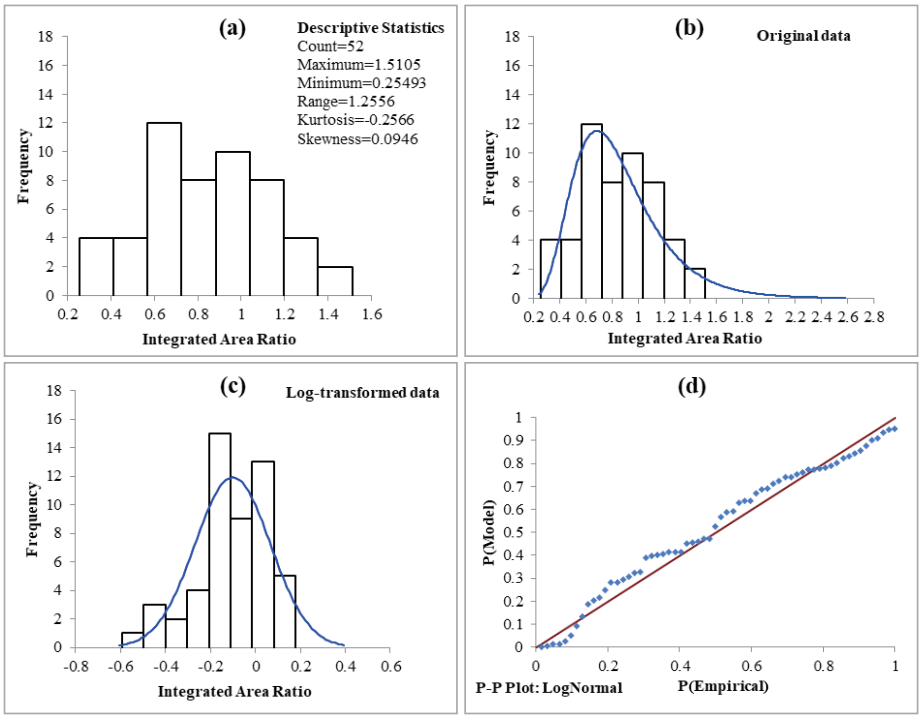


**Fig. 8** Statistical analysis of ethyl octanoate/IS of authentic single malt Scotch whiskies. (a) The histogram of the raw data. (b) The distribution fitting results showed that the original data conformed with a lognormal distribution. (c) The distribution fitting results showed that the log-transformed data conformed with a normal distribution. (d) The P-P plot of distribution fitting result.

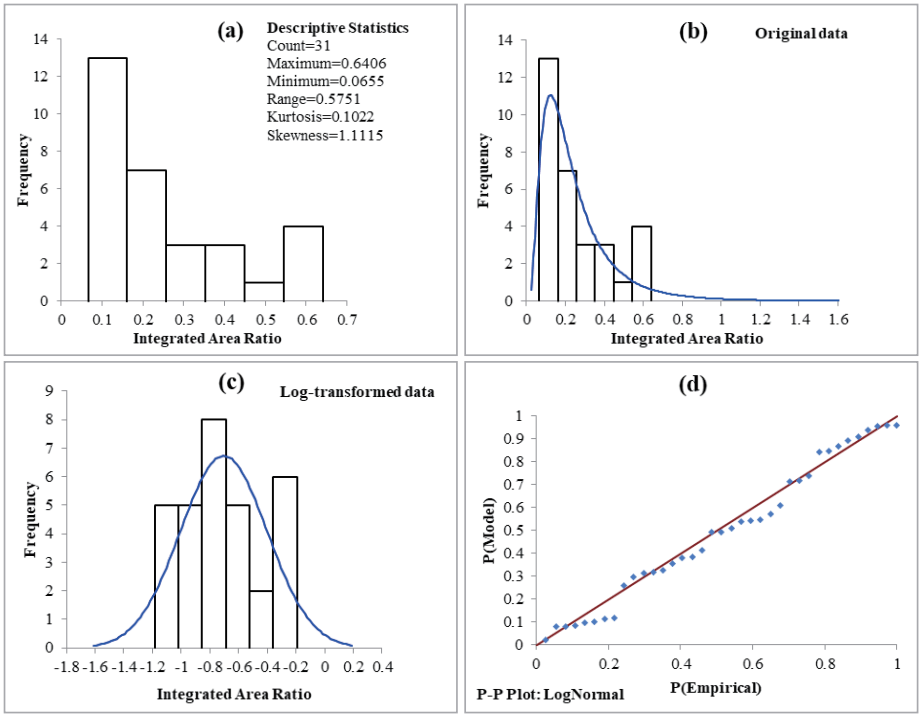


**Fig. 9** Statistical analysis of ethyl octanoate/IS of authentic blended Scotch whiskies. (a) The histogram of the raw data. (b) The distribution fitting results showed that the original data conformed with a lognormal distribution. (c) The distribution fitting results showed that the log-transformed data conformed with a normal distribution. (d) The P-P plot of distribution fitting result.

*3. Higher Alcohols in Authentic Scotch whiskies*

Higher alcohols in spirits are formed mainly by anabolism or catabolism reaction called the Ehrlich pathway of amino acids. The higher alcohols are produced by yeast fermentation through biochemical reactions with amino acids and carbohydrates. Therefore, the yield of higher alcohols mainly results from yeast growth [3, 6-7, 27]. The Ehrlich pathway of amino acids, which produce the higher alcohol, is summarized in Fig. 10. Among the higher-alcohol congeners found in Scotch whiskies in this study, 2-methyl propanol, 2-methyl-butanol and 3-methyl-1-butanol are filtered to investigate the possibility of serving as the chemical target to confirm the authenticity of Scotch whiskies.

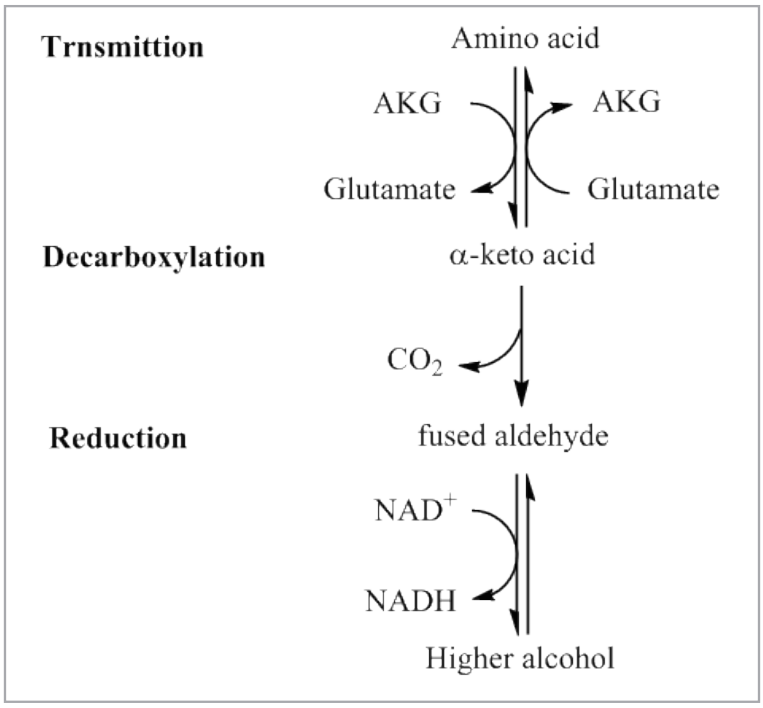


**Fig. 10** The production route of higher alcohols during fermentation [27].

Fig. 11, Fig. 12, and Fig. 13 show the statistical description of the integrated ratio of 2-methyl propanol/IS, 2-methyl butanol/IS and 3-methyl butanol/IS in authentic single malt Scotch whiskies, respectively. As an apparent skewness pattern with a tail on the right side, it is noted that all of the integrated value distributions of these three higher-alcohols are possibly identified as lognormal distributions. As shown in Fig. 11 to Fig. 13, the corresponding distribution tests demonstrated that the inference about lognormal distribution to describe 2-methyl propanol, 2-methyl butanol and 3-methyl butanol in authentic single malt Scotch whiskies is correct.

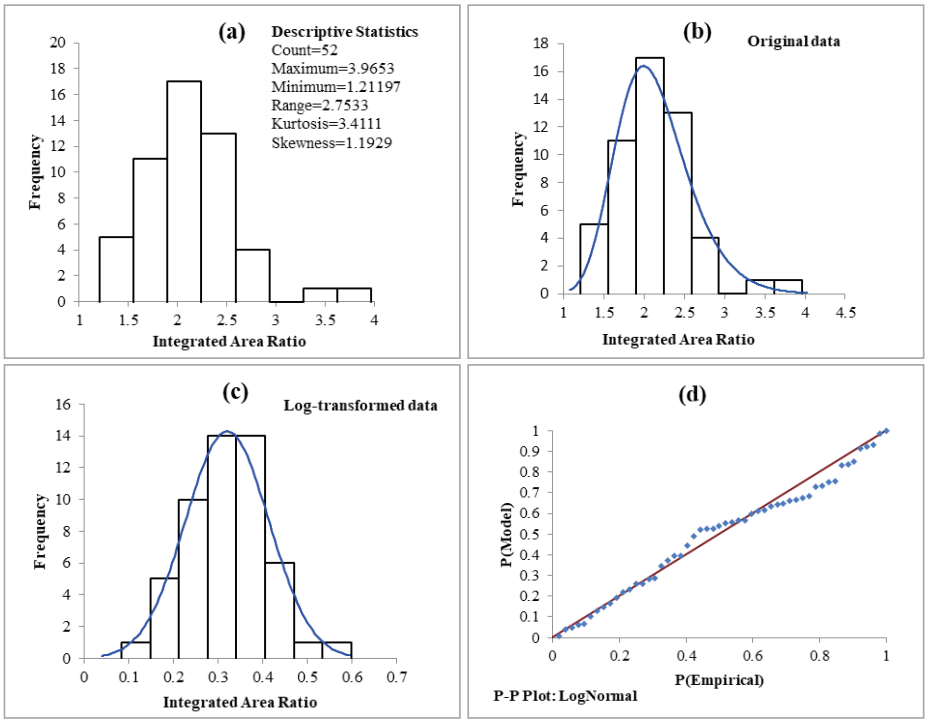


**Fig. 11** Statistical analysis of 2-methyl propanol/IS of authentic single malt Scotch whiskies. (a) The histogram of the raw data. (b) The distribution fitting results showed that the original data conformed with a lognormal distribution. (c) The distribution fitting results showed that the log-transformed data conformed with a normal distribution. (d) The P-P plot of distribution fitting result.

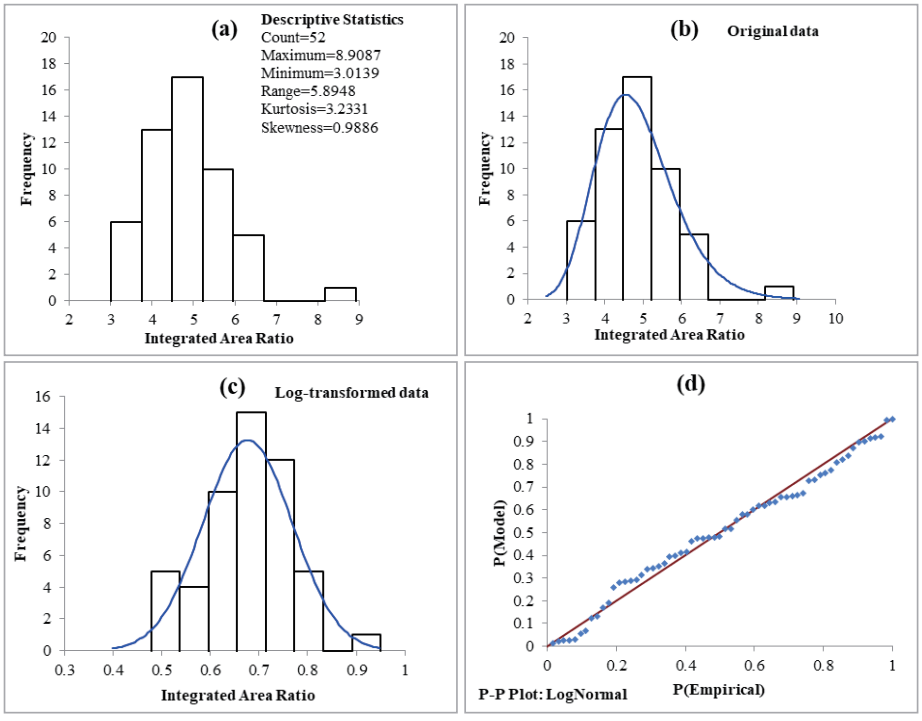


**Fig. 12** Statistical analysis of 2-methyl butanol/IS of authentic single malt Scotch whiskies. (a) The histogram of the raw data. (b) The distribution fitting results showed that the original data conformed with a lognormal distribution. (c) The distribution fitting results showed that the log-transformed data conformed with a normal distribution. (d) The P-P plot of distribution fitting result.

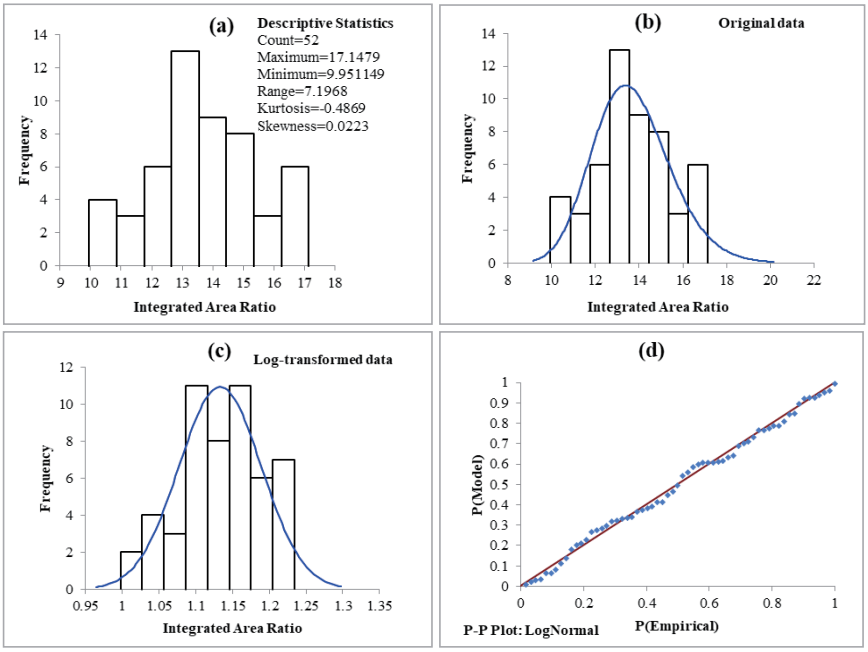


**Fig. 13** Statistical analysis of 3-methyl butanol/IS of authentic single malt Scotch whiskies. (a) The histogram of the raw data. (b) The distribution fitting results showed that the original data conformed with a lognormal distribution. (c) The distribution fitting results showed that the log-transformed data conformed with a normal distribution. (d) The P-P plot of distribution fitting result.

On the other hand, the integrated values of these three higher alcohols in authentic blended Scotch whiskies are also distributed to a lognormal distribution, as shown in Fig. 14, Fig. 15, and Fig. 16. Therefore, considering all these tests, the integrated ratio of 2-methyl propanol/IS, 2-methyl butanol/IS and 3-methyl butanol/IS, either in authentic single malt Scotch whiskies or authentic blended Scotch whiskies, are demonstrated to be lognormal distributions.

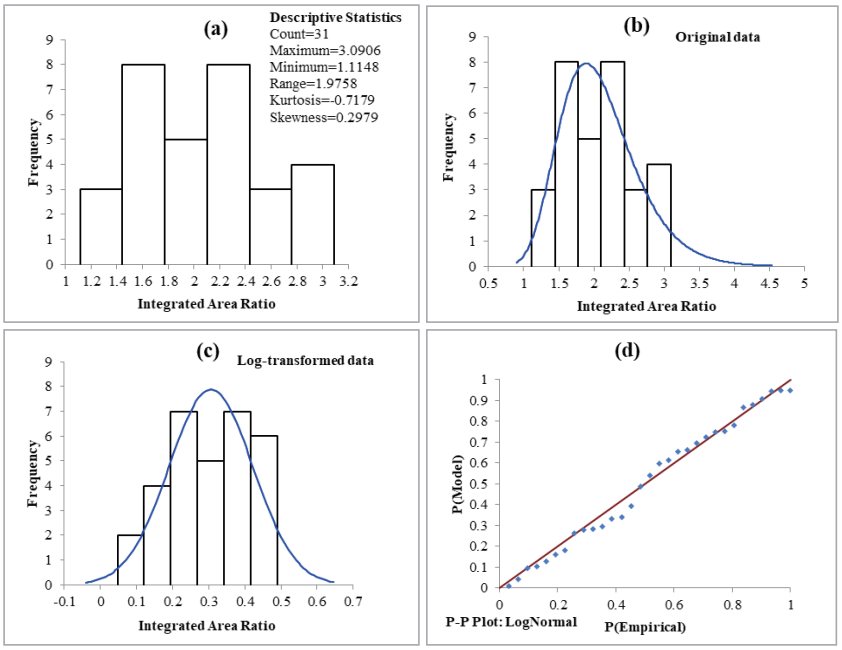


**Fig. 14** Statistical analysis of 2-methyl propanol/IS of authentic blended Scotch whiskies. (a) The histogram of the raw data. (b) The distribution fitting results showed that the original data conformed with a lognormal distribution. (c) The distribution fitting results showed that the log-transformed data conformed with a normal distribution. (d) The P-P plot of distribution fitting result.

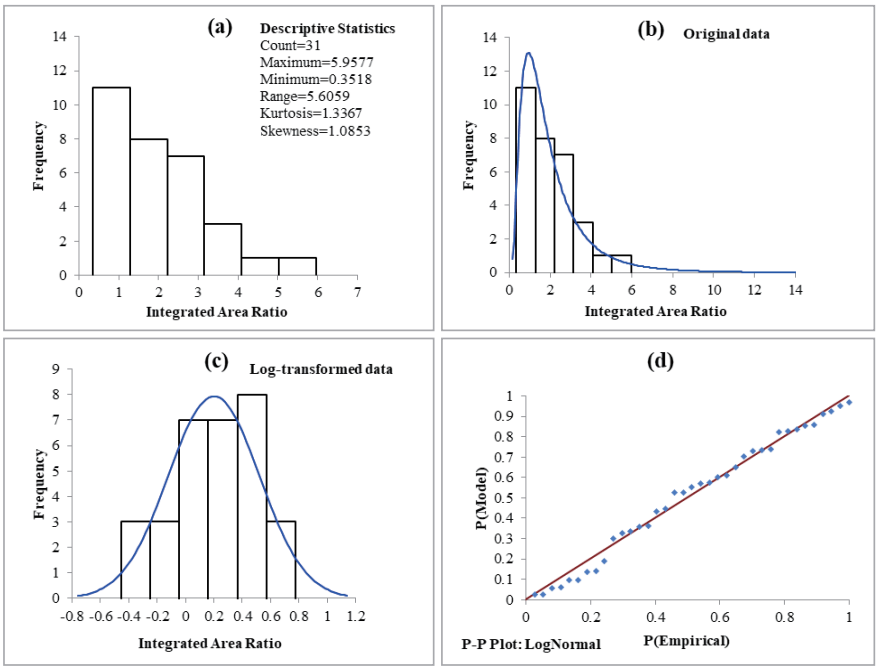


**Fig. 15** Statistical analysis of 2-methyl butanol/IS of authentic blended Scotch whiskies. (a) The histogram of the raw data. (b) The distribution fitting results showed that the original data conformed with a lognormal distribution. (c) The distribution fitting results showed that the log-transformed data conformed with a normal distribution. (d) The P-P plot of distribution fitting result.

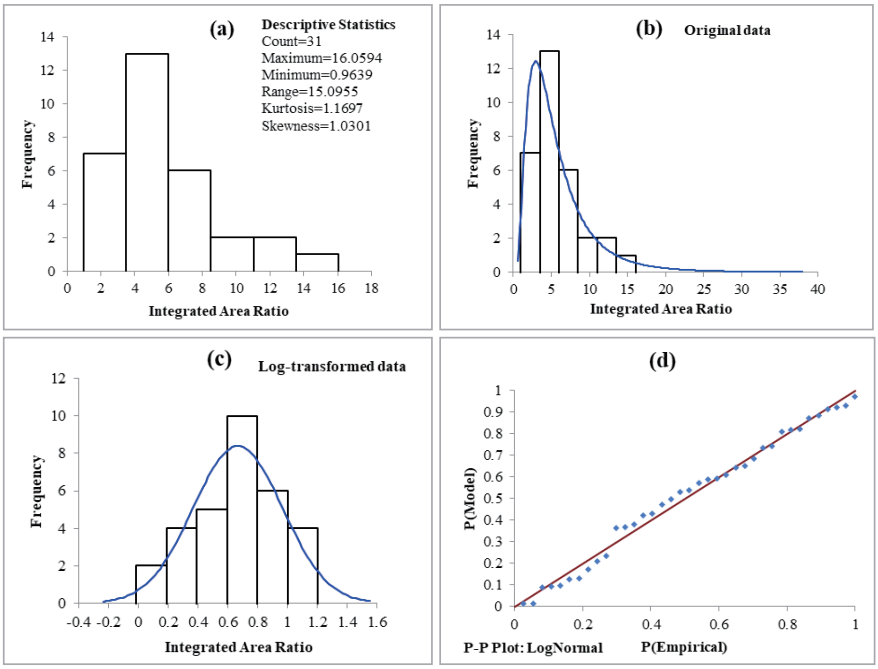


**Fig. 16** Statistical analysis of 3-methyl butanol/IS of authentic blended Scotch whiskies. (a) The histogram of the raw data. (b) The distribution fitting results showed that the original data conformed with a lognormal distribution. (c) The distribution fitting results showed that the log-transformed data conformed with a normal distribution. (d) The P-P plot of distribution fitting result.

### *Implications of Lognormal Distribution for Integrated Ratio of Congeners/IS*

Scotch whisky is one of the best-selling spirits in the world. According to the diversity of raw materials and production process, malt type of Scotch whisky and grain type of Scotch whisky are two significant types of Scotch whiskies [1,3]. Furthermore, blended Scotch whisky is manufactured by mixing these two types of Scotch whiskies with a ratio determined by distilleries. Therefore, some apparent factors affect the distribution of the integrated ratio of congeners/IS, including raw materials, fermentation process by yeast, distillation process, and blended process. The stochastic variability of all factors leads to a lognormal distribution of the integrated ratio of congeners/IS, while constant variance brings about normal distribution [28, 29].

According to the metabolic routes of yeast, the primary metabolites that occur during the fermentation or wort process are ethanol, carbon dioxide, and glycerol. The secondary metabolic products of yeast, such as aldehyde, esters, higher alcohols, and sulfur compounds, result in spirits' flavor characteristics [25]. As shown in this study, 6 secondary metabolic products of yeast, including acetaldehyde, ethyl acetate, ethyl octanoate, 2-methyl propanol, 2-methyl butanol, and 3-methyl butanol, are demonstrated that the integrated ratio of congeners/IS is distributed as lognormal distribution in both authentic single malt Scotch whiskies and authentic blended Scotch whiskies. In view of a large number of authentic samples and a series of statistical tests, the distribution assessments in this study are representative.

### *Exclusive Authentication of Scotch whiskies by Using Principal Component Analysis (PCA) with Log-Transformed Data.*

To visualize the discriminating power of the fermentation congeners, PCA is selected to describe the grouping cluster. Based on the lognormal distribution and statistical characteristics, the PCA plots with log-transformed data might display a better-discriminating feature than those with original data. The data matrix of authentic and adulterated Scotch whiskies was subjected to two principal components, PC1 and PC2, to investigate the eliminative result for authentication.

With original data of the fermentation congeners in Scotch whiskies, the PCA charts were plotted in Fig.17. For a better description of authentication, the sample distribution in the PC1 and PC2 axis of Fig. 17 were depicted in Fig. 19(a), Fig. 19(c), Fig. 20(a), and Fig. 20(c), respectively. With the original data of the fermentation congeners in single malt Scotch whiskies, Fig. 17(a) shows that PC1 expresses a slightly positive skewness with a right tail, and PC2 displays a symmetrical pattern. After distribution evaluation as aforementioned, it is certified that the data of PC1 in Fig. 17(a) belongs to the lognormal distribution, and those of PC2 belong to normal distribution. To explain the discriminating power, the confidence interval (CI) is estimated based on the statistical empirical rule. For a normal distribution, 68.3% of the observed data will fall within the first standard deviation ($\bar{x} \pm s$), 95.5% within the first two standard deviations ($\bar{x} \pm 2s$), and 99.7% within the first three standard deviations ($\bar{x} \pm 3s$). For a lognormal distribution, the corresponding confidence intervals become $\bar{x}^* \times/ s^*$ for 68.3% probability, $\bar{x}^* \times/ (s^*)^2$ for 95.5% probability, and $\bar{x}^* \times/ (s^*)^3$ for 99.7% probability in the original scale [29]. The estimated CI boundaries are labeled in Fig. 19 and Fig. 20. For classification results for the PC axis in Fig. 19(a) and Fig. 19(c), the adulterated sample, S1, is located within the CI range of 95.5% and 99.7% in PC2, meaning there is less discrimination for PCA charts with original data for authentication. In the case of blended Scotch whiskies, Fig. 17(b), similar to Fig. 17(a), presents that the data distribution of PCA with original data does not belong to the normal distribution. Moreover, the seized blended Scotch whiskies, S2 and S3, were found within the authentic blended Scotch whiskies distribution, as shown in Fig. 20(a). As a result, the PCA analysis with original data expresses less eliminative ability for Scotch whiskies authentication.

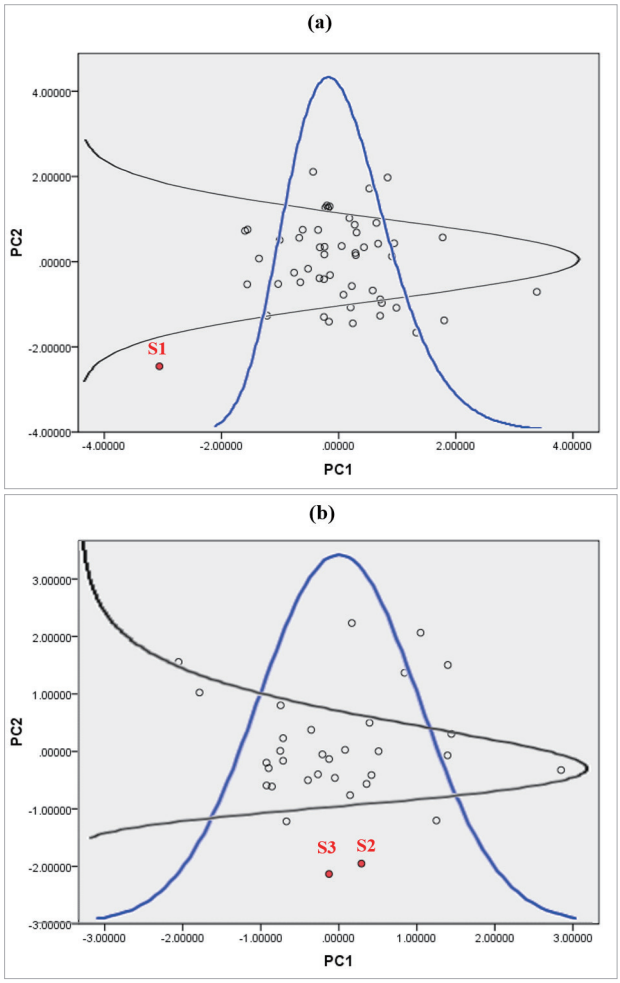


**Fig. 17** PCA charts with original data for authentication. (a) for single malt Scotch whiskies. (b) for blended Scotch whiskies.

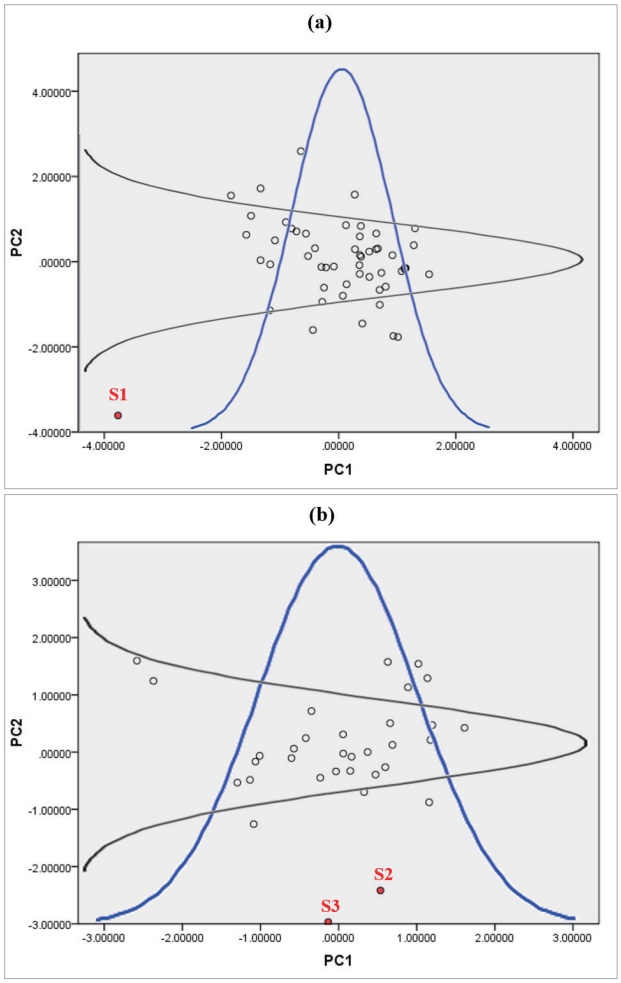


**Fig. 18** PCA charts with log-transformed data for authentication. (a) for single malt Scotch whiskies. (b) for blended Scotch whiskies.

On the other hand, the log-transformed data in Fig. 18(a) displays two groups, indicating that authentic single malt Scotch whiskies could be grouped homogeneously, and adulterated sample S1 behaves as another group. Specifically speaking, adulterated sample S1 is located out of the CI boundaries of 99.7% for both PC1 and PC2, as shown in Fig. 19(b) and Fig. 19(d). Therefore, compared to the classification result of original data in Fig.17(a), it received a better discrimination result. In the case of the log-transformed data of blended Scotch whiskies, however, the PCA chart appears to be a more scattered feature than that of single malt Scotch whiskies. As shown in Fig. 18(b), two adulterated blended Scotch whiskies, S2 and S3, are located within the sample distribution of the authentic group in PC1. Nevertheless, S2 and S3 in Fig. 18(b) express more discrimination power in PC2 than in Fig. 17(b).

Fig. 19 and Fig. 20 show the distribution of the PCA axis, including those with original data and log-transformed data. Obviously, for all the distribution with log-transformed data in Fig. 19 and Fig. 20, the adulterated samples have located a distance with a higher degree of discrimination from the authentic groups for both single malt Scotch whiskies and blended Scotch whiskies, compared to the distribution with original data. Due to the complex factors during the manufacturing process, it is reasonable that the discriminating power of PCA for blended Scotch whiskies is less than that for single malt Scotch whiskies.

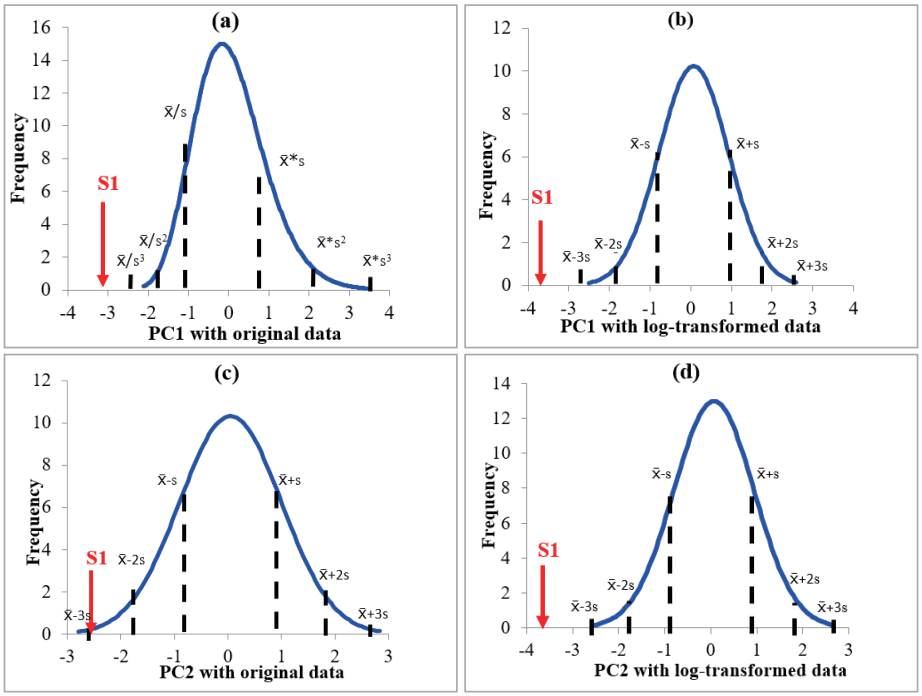


**Fig. 19** Statistical distribution PCA axis of authentication for single malt Scotch whiskies. (a) on PC1 with original data. (b) on PC1 with log-transformed data. (c) on PC2 with original data. (d) on PC2 with log-transformed data.

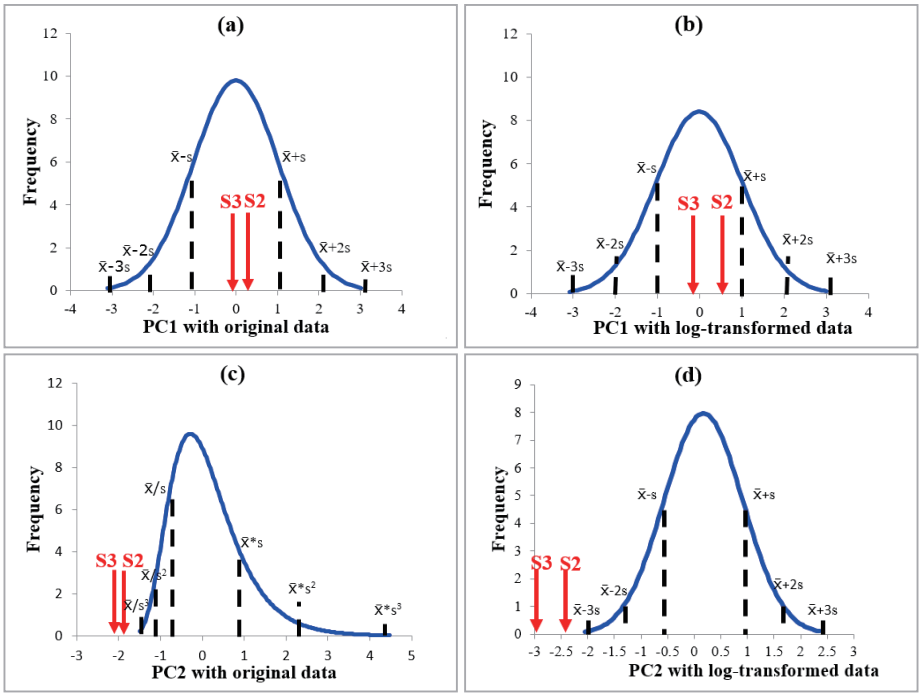


**Fig. 20** Statistical distribution PCA axis of authentication for blended Scotch whiskies. (a) on PC1 with original data. (b) on PC1 with log-transformed data. (c) on PC2 with original data. (d) on PC2 with log-transformed data.

In order to enhance the discrimination power of blended Scotch whiskies, linear discriminant analysis (LDA) is employed to classify the adulterated and authentic samples, a statistical analysis method used to describe these differences between groups. Table 4 shows the casewise statistical results of LDA analysis. As shown in Table 4, with original data, there are two misclassified cases within the authentic blended Scotch whisky group. However, all the cases with log-transformed data were accurately classified. The aforementioned classification results were plotted in Fig. 21. As shown in Fig. 21(a), the two adulterated samples, S2 and S3, were classified as different groups from the authentic samples. However, two authentic samples, with the green triangle label and green diamond label in Fig. 21(a) are predicted to be in the adulterated groups, implying the wrong classification. Therefore, it is proved that LDA with original data would lead the error discrimination.

On the other hand, with log-transformed data of blended Scotch whiskies, LDA can successfully classify the adulterated samples, S2 and S3, from authentic samples with 100% correctness. Especially, as shown in Fig. 21(b), S2 and S3 can be quickly excluded from authentic blended Scotch whiskies on the LDA1 axis. As a result, LDA with log-transformed data can enhance the discrimination power from authentic blended Scotch whiskies.

In conclusion, a PCA study with log-transformed data of the integrated ratio of fermentation congeners/IS in Scotch whiskies can be used as a primary eliminative method for the authentication of seized Scotch whiskies. In addition, with log-transformed data, both PCA and LDA can provide better classification results for blended Scotch whiskies.

**Table 4** The casewise statistics of authentic blended Scotch whiskies and adulterated whisky samples by LDA analysis.

| Casewise Statistics | | | | | |
|---|---|---|---|---|---|
| Samples | | with original data | | with log-transformed data | |
| | | actual group | predicted group | actual group | predicted group |
| authentic blended Scotch whisky | 1 | 1 | 1 | 1 | 1 |
| | 2 | 1 | 1 | 1 | 1 |
| | 3 | 1 | 1 | 1 | 1 |
| | 4 | 1 | 1 | 1 | 1 |
| | 5 | 1 | 1 | 1 | 1 |
| | 6 | 1 | 1 | 1 | 1 |
| | 7 | 1 | 1 | 1 | 1 |
| | 8 | 1 | 1 | 1 | 1 |
| | 9 | 1 | 1 | 1 | 1 |
| | 10 | 1 | 1 | 1 | 1 |
| | 11 | 1 | 1 | 1 | 1 |
| | 12 | 1 | 1 | 1 | 1 |
| | 13 | 1 | **3*** | 1 | 1 |
| | 14 | 1 | 1 | 1 | 1 |
| | 15 | 1 | 1 | 1 | 1 |
| | 16 | 1 | 1 | 1 | 1 |
| | 17 | 1 | 1 | 1 | 1 |

Table 4 (continued)

| Casewise Statistics | | | | | |
|---|---|---|---|---|---|
| Samples | | with original data | | with log-transformed data | |
| | | actual group | predicted group | actual group | predicted group |
| authentic blended Scotch whisky | 18 | 1 | 1 | 1 | 1 |
| | 19 | 1 | 1 | 1 | 1 |
| | 20 | 1 | 1 | 1 | 1 |
| | 21 | 1 | 1 | 1 | 1 |
| | 22 | 1 | 1 | 1 | 1 |
| | 23 | 1 | 1 | 1 | 1 |
| | 24 | 1 | 1 | 1 | 1 |
| | 25 | 1 | 1 | 1 | 1 |
| | 26 | 1 | 1 | 1 | 1 |
| | 27 | 1 | 1 | 1 | 1 |
| | 28 | 1 | **2*** | 1 | 1 |
| | 29 | 1 | 1 | 1 | 1 |
| | 30 | 1 | 1 | 1 | 1 |
| | 31 | 1 | 1 | 1 | 1 |
| adulterated whisky samples | S2 | 2 | 2 | 2 | 2 |
| | S3 | 3 | 3 | 3 | 3 |

* misclassified case

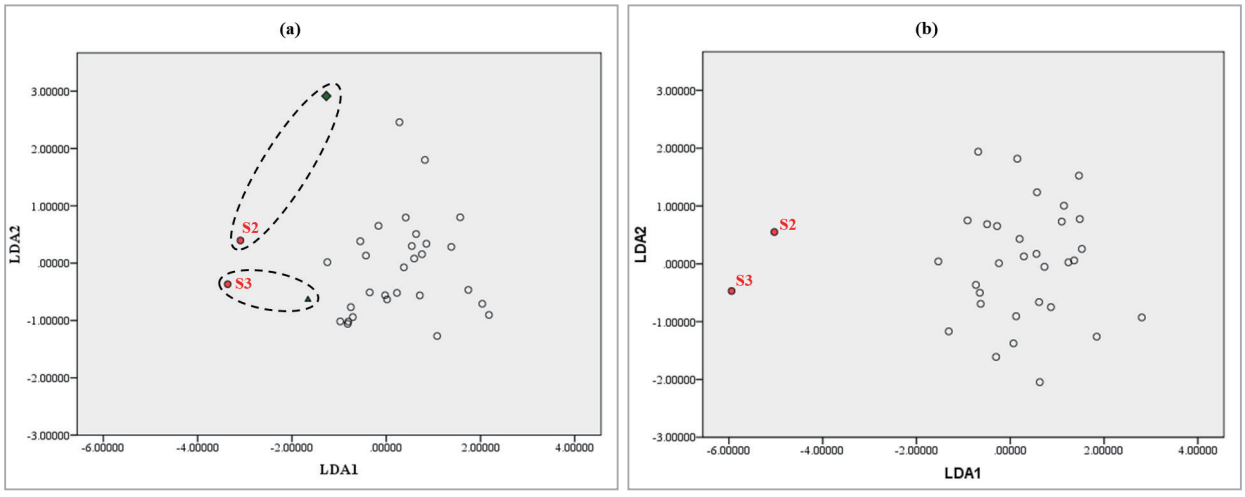


**Fig. 21** LDA results for blended Scotch whiskies. (a) with original data. (b) with log-transformed data.

## Conclusions

Instrumental measurement combined with statistical evaluation is an efficient method nowadays. While using the statistical model to describe the observed data, it should be noted that the statistical model is suitable and correct. In this work, we developed a method to improve the discriminating result of PCA. A series of statistical processes, including PCA evaluation and data fitting method, to select the appropriate fermentation congeners were operated to examine the data of the integrated ratio of fermentation congeners/IS in Scotch whiskies. As a result, we demonstrated that the data of selected six fermentation congeners in Scotch whiskies are proposed as lognormal distribution, which is reasonable based on nature distribution and could be served as helpful components to improve the discriminating power of the PCA method for Scotch whiskies. For those data with less discrimination power from the PCA method, LDA with log-transformed data could also effectively provide the classification result. On the basis of the observed data characteristics, it was certified that PCA with log-transformed data was critical in Scotch whiskies in determining the authenticity of suspected bottled whiskies.

## Acknowledgments

The authors would like to thank the Ministry of the Interior, Republic of China (Taiwan) for the financial support (project no. 110-0805-10-17-01) and Dr. Ya-Ting Chang for the statistical suggestions.

## References

1. The Scotch whisky Regulations 2009 No. 2890. https://www.legislation.gov.uk/uksi/2009/2890/pdfs/uksi_20092890_en.pdf.
2. The Scotch whisky Association. 2021 exports show industry on road to recovery. https://www.scotch-whisky.org.uk/newsroom/2021-exports-show-industry-on-road-to- recovery/.
3. Stewart G, Russell I. Whisky: Technology, Production and Marketing. 2nd edition, ISBN: 978-0124017351, Wiley-Blackwell, 2014.
4. Suomalainen H. Yeast and its effect on the flavour of alcoholic beverages. J. Inst. Brew. 1971; 77: 164–177. https://doi.org/10.1002/j.2050-0416.1971.tb03370.x.
5. Suomalainen H, Lehtonen, M. The production of aroma compounds by yeast. J. Inst. Brew. 1979; 85: 149–156. https://doi.org/10.1002/j.2050-0416.1979.tb06846.x.
6. Chen EH. Relative contribution of Ehrlich and biosynthetic pathways to the formation of fusel alcohols. J. Am. Soc. Brew. Chem. 1978; 36: 39–43. https://doi.org/10.1094/ASBCJ-36-0039.
7. Pires EJ, Teixeira JA. Brányik, T. Vicente, A. A. Yeast: the soul of beer's aroma—a review of flavour-active esters and higher alcohols produced by the brewing yeast. Appl. Microbio. Biotechnol. 2014; 98: 1937–1949. doi: 10.1007/s00253-013-5470-0.
8. Herranz A, Serna P, Barro C, Martin PJ, Cabezudo MD. Application of the statistical multivariate analysis to the differentiation of whiskies of different brands. Food Chem. 1989; 31: 73–81. https://doi.org/10.1016/0308-8146(89)90152-0.
9. Demyttenaere JCR, Sánchez Martínez JI, Verhé R, Sandra P, De Kimpe N. Analysis of volatiles of malt whisky by solid-phase microextraction and stir bar sorptive extraction. J. Chromatogr. A. 2003; 985: 221–32. doi: 10.1016/s0021-9673(02)01471-1.
10. Barnaba C, Dellacassa E, Nicolini G, Nardin T, Malacarne M, Larcher R. Identification and quantification of 56 targeted phenols in wines, spirits, and vinegars by online solid-phase extraction-ultrahigh-performance liquid chromatography-quadrupole-orbitrap mass spectrometry. J. Chromatogr A. 2015; 1423: 124–135. doi: 10.1016/j.chroma.2015.10.085.
11. Wiśniewska P, Boqué R, Borràs E, Busto O, Wardencki W, Namieśnik J, Dymerski T. Authenti-cation of whisky due to its botanical origin and way of production by instrumental analysis and multivariate classification methods, Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2017; 173: 849–853, https://doi.org/10.1016/j.saa.2016.10.042.
12. Picque D, Lieben P, Corrieu G, Cantagrel R, Lablanquie O, Snakkers G. Discrimination of Cognacs and Other Distilled Drinks by Mid-infrared Spectroscopy. J. Agric. Food Chem. 2006; 54: 5220–5226. DOI: 10.1021/jf060465u.
13. Contreras U, Barbosa-García O, Pichardo-Molina

JL, Ramos-Ortíz G, Maldonado JL, Meneses-Nava MA, Ornelas-Soto NE, López-de-Alba PL. Screening method for identification of adulterate and fake tequilas by using UV–VIS spectroscopy and chemometrics, Food Res. Int. 2010; 43: 2356–2362, https://doi.org/10.1016/j.foodres.2010.09.001.

14. González-Arjona D, González-Gallero V, Pablos F, Gustavo González A. Authentication and differentiation of irish whiskeys by higher-alcohol congener analysis. Anal. Chim. Acta. 1999; 381: 257–264, https://doi.org/10.1016/S0003-2670(98)00764-8.
15. Tamaki T, Takamiya Y, Miyagi T, Nishiya T. Changes in ester compounds and higher alcohols of awamori during aging, J. Ferment. Technol. 1986; 64: 17–24, https://doi.org/10.1016/0385-6380(86)90052-X.
16. Nicolli KP, Biasoto ACT, Souza-Silva EA, Guerra CC, dos Santos HP, Welke JE, Zini CA. Sensory, olfactometry and comprehensive two-dimensional gas chromatography analyses as appropriate tools to characterize the effects of vine management on wine aroma, Food Chem. 2018; 243: 103–117, https://doi.org/10.1016/j.foodchem.2017.09.078.
17. Annila A, Grönholm T. Natural Distribution. Math. Biosci. 2007; 210: 659–667. https://www.mv.helsinki.fi/home/aannila/arto/naturaldistribution.pdf.
18. Siano DB. The log-normal distribution function, J. Chem. Educ. 1972; 49: 755–757. https://doi.org/10.1021/ed049p755.
19. Graddum JH. Lognormal distribution. Nature. 1945; 156: 463–466.
20. Koch AL. The logarithm in biology. I. Mechanisms generating the log-normal distribution exactly. J. Theor. Biol. 1966; 23: 276–290. https://doi.org/10.1016/0022-5193(66)90119-6.
21. ILAC-G19:2002, Guidelines for Forensic Science Laboratories, 2002. http://www.sadcmet.org/SADCWaterLab/Archived_Reports/2006%20Reports%20and%20Docs/Ilac-g19.pdf.
22. Huang HW, Chang WT. Methanol Concentration as a Preceding Eliminative Marker for the Authentication of Scotch whiskies in Taiwan. Forensic Sci. Int. 2022; 339: 111413. https://doi.org/10.1016/j.forsciint.2022.111413.
23. Huang HW, Chang WT. $\delta^{13}$C-Ethanol as a Potential Exclusionary Criterium for the Authentication of Scotch whiskies in Taiwan: Normal vs. 3-Parameter Lognormal Distributions of $\delta^{13}$C-Ethanol Found in Single malt and Blended Scotch whiskies. Beverages, 2022, Submitted.
24. Lister A. Chapter 7-Validation of HPLC Methods in Pharmaceutical Analysis, Separation Science and Technology, Academic Press, 2005; 6: 191–217, https://doi.org/10.1016/S0149-6395(05)80051-0.
25. Stewart GG. The Production of Secondary Metabolites with Flavors Potential during Brewing and Distilling Wort Fermentations. Fermentation 2017 ; 3: 63. https://doi.org/10.3390/fermentation3040063.
26. Fabienne R, Emilie A, Sylvie D. Engineering of the Pyruvate Dehydrogenase Bypass inSaccharomyces cerevisiae: Role of the Cytosolic Mg2+ and Mitochondrial K+ Acetaldehyde Dehydrogenases Ald6p and Ald4p in Acetate Formation during Alcoholic Fermentation. ASM Journals: Applied and Environmental Microbiology, 2000; 66: 3151–3159. https://doi.org/10.1128/AEM.66.8.3151-3159.2000.
27. Hazelwood LA, Daran JM, van Maris AJ, Pronk JT, Dickinson JR. The Ehrlich pathway for fusel alcohol production: a century of research on Saccharomyces cerevisiae metabolism. Appl. Environ. Microbiol. 2008; 74: 2259–2266. doi: 10.1128/AEM.02625-07.
28. Andersson A. Mechanisms for log normal concentration distributions in the environment. Sci. Rep. 2021; 11: 16418. https://doi.org/10.1038/s41598-021-96010-6.
29. Limpert E, Stahel WA, Abbt M. Log-normal distributions across the sciences: keys and clues. BioScience 2001; 51: 341–352. https://doi.org/10.1641/0006-3568(2001)051[0341:LNDATS]2.0.CO;2.