國立中山大學 應用數學 學系(研究所)

碩士論文

手術時間分佈之混合模型估計

Mixture models for estimating operation time distributions

研究生：陳怡綾 撰

指導教授：羅夢娜 教授

中華民國 九十四 年 七月

Mixture models for estimating operation time distributions

by
Yi-Ling Chen

Advisor
Mong-Na Lo Huang

Department of Applied Mathematics
National Sun Yat-sen University
Kaohsiung, Taiwan, 804, R.O.C.

July, 2005

Abstract

Surgeon operation time is a useful and important information for hospital management, which involves operation time estimation for patients under different diagnoses, operation room scheduling, operating room utilization improvements and so on. In this work, we will focus on studying the operation time distributions of thirteen operations performed in the gynecology (GYN) department of one major teaching hospital in southern Taiwan. We firstly investigate what types of distributions are suitable in describing these operation times empirically, where log-normal and mixture log-normal distribution are identified to be acceptable statistically in describing these operation times. Then we compare and characterize the operations into different categories based on the operation time distribution estimates. Later we try to illustrate the possible reason why distributions for some operations with large data set turn out to be mixture of certain log-normal distributions. Finally we end with discussions on possible future work.

Keywords: classification, EM algorithm, gynecology, likelihood ratio test, MLE, mixture of log-normal distributions.

Contents

List of Tables **1**

List of Figures **3**

1. Introduction **4**

2. Mixture distributions and method of estimation of parameters ... **9**
- **2.1 Expectation-maximization (EM) algorithm** 9
- **2.2 The standard errors of MLE** 10
- **2.3 Kolmogorov-Smirnov test** 11
- **2.4 Tamhane T2 test** 12

3. Effect testing **12**
- **3.1 Likelihood ratio test** 13

4. Statistical analysis results **15**
- **4.1 Classification of operations based on operation time** 15
- **4.2 Mixture log-normal distribution** 16
- **4.3 Effect of doctor for three operations** 20
- **4.4 Effect of stage for V4 operation time** 23
- **4.5 Other variables for V4 operation** 25
- **4.6 Poisson regression model for V4 operation** 27

5. Conclusion and discussion **29**

References **30**

Appendix **32**

List of Tables

Table 1: Thirteen operations in the GYN department. 6

Table 2: Descriptive statistics of operation times. 7

Table 3: Descriptive statistics of the logarithm of the operation times. 7

Table 4: Results of K-S test. 15

Table 5: One-way ANOVA. 16

Table 6: Results of fitting a three-component mixture of log-normal distributions to V4 operation time. 18

Table 7: Results of fitting a three-component mixture of log-normal distributions to V11 operation time. 18

Table 8: Results of fitting a three-component mixture of log-normal distributions to V12 operation time. 19

Table 9: Descriptive statistics of doctors at V4 operation time. 21

Table 10: Results of fitting a three-component mixture of log-normal distributions to data sets of doctors at V4 operation time. 21

Table 11: Descriptive statistics of doctors at V11 operation time. 21

Table 12: Results of fitting a three-component mixture of log-normal distributions to data sets of doctors at V11 operation time. 22

Table 13: Descriptive statistics of doctors at V12 operation time. 22

Table 14: Results of fitting a three-component mixture of log-normal distributions to data sets of doctors at V12 operation time. 23

Table 15: Likelihood ratio test under the hypothesis-H_0 : the effect of doctor is not significant. 23

Table 16: Summary statistic of stages at V4 operation time 24

Table 17: Results of fitting a three-component mixture of log-normal distributions to data sets of stages at $V4_{(148)}$ operation time. 24

Table 18: Likelihood ratio test for the effect of stage. 25

Table 19: Correlation matrix 26

Table 20: The fitted results.(Comp. component) 26

Table 21: Results of fitting poisson regression model 28

Table A1: Results of fitting poisson regression model using S-PLUS software. 35

List of Figures

Figure 1: Boxplots for operation times. 8

Figure 2: Boxplots for the logarithm of the operation times. 8

Figure 3: Results for Tamhane T2 test. 16

Figure 4: Plots of mixture of log-normal distributions for V4, V11, and V12 operation times. 19

Figure 5: Boxplots of stages for operation times. 25

Figure 6: Scatter plot with confidence ellipse for V4 operation time in stage 2. 27

Figure 7: Scatter plot of Blood L and Blood T for V4 operation in stage2 27

Figure 8: Scatter plot of Blood T and Age for V4 operation in stage 227

Figure 9: Response variable versus the fitted values 28

Figures A1: Histograms of the thirteen operation times. 36

Figures A2: Histograms of the logarithm of the thirteen operation times. 38

1. Introduction

In recent years, the managers of hospitals have been confronted with the severe competition for operation; besides, the policies of our Health insurance program also make them pay more attention to the cost control. Not only the high cost of human resource but the capital-intensive attribute of surgeries compel the managers to think more about appropriate administration and one of the efficient uses of all resource. Therefore, the surgery management has became important issues to hospital manager. Surgeon operation time is a useful and important information for hospital management, which involves operation time estimation for patients under different diagnoses, operation room scheduling, operating room utilization improvements and so on.

In our study, the data of the operation time was collected from the gynecology (GYN) department at one major teaching hospital in southern Taiwan from January 2000 to June 2003. There are nine doctors in the gynecology department. Because there are many kinds of operations performed in the GYN department, we analyze thirteen operations where the data counts exceed thirty. Corresponding variable name for these operations are listed in Table 1. The descriptive statistics and boxplots of the operation time for each of the operation (Vi), i= 1,. . . ,13, are presented in Table 2 and Figure 1. After taking the logarithm of the original time data of each operation, the descriptive statistics and boxplots (LVi), i= 1,. . . ,13, are separately listed in Table 3 and Figure 2. Later each operation time distribution has been fitted by a log-normal distribution verified by Kolmogorov-Smirnov goodness of fit test (K-S test) respectively. We investigate the characteristics of these distributions of the logarithm of operation time by statistics.

For those data sets which can be fitted by log-normal distributions, we use standard one-way analysis of variance to compare the differences between means. If it is considered to have significant differences between them, we then use Tamhane T2 test to compare how they are differed in means and characterize the operations into different categories with "short", "medium", "long", or "ultra long" operation times. For those data sets which are rejected as fitted by the K-S test, they are further fitted by a more general mixture log-normal distribution instead.

In this work, a two-factor factorial design with factor doctor and stage which means the disease severity of the patients is considered although the response of the time for each operation which are rejected by the K-S test is assumed to be a random variable with mixture log-normal distribution, where the parameters of the distribution such as the mixture proportions, parameters for each distribution in

the mixtures, depend on the factor. In the following, we will consider the general mixture model and describe the method we are going to use to do the estimation and perform the appropriate tests. In the practical problem here it is of interest to know whether the factor of doctor and stage have effect on the operation time which be fitted by mixture log-normal distribution. In order to do that, likelihood ratio tests are used here to accomplish our investigations.

In fact, we should take all the variables which affect operation time distributions into account when we examine the factor effect on operation time and try to find other factors which affect operation time distributions to explain why distribution for some operations turn out to be mixture log-normal distributions. The other variables which affect operation time distributions are like patient's age, residents practicing years, blood transfusion and blood loss et al. For operation time which are fitted by log-normal distribution, we can use the method of multiple regression to test the relationship between independent variables (doctor, stage, patient's age, residents practicing-years, blood transfusion and blood loss et al) and dependent variable (operating time) and generalized linear regression to develop a operating time forecasting model. In multiple regression, we maybe use different models to find what kind of variables effect operating time most. Hence we not only provide these relationship results to help the operation room scheduling more efficient, but also build a surgeon-based model to forecast most operating time by different variables.

Operating room managers who seek to maximize utilization in their operating room suite may attempt to build an efficient operating room scheduling. Accurate estimation of operating times is a prerequisite for the efficient scheduling of the operating suite. Operation time distributions of GYN department generally can be consider two types; log-normal and mixture log-normal distributions. How to calculate the operation time finishing probability and decide the order of operation be performed for the operation room scheduling are primary problems. For example, if three operations are performed today, the operation time of X_1 is fitted by log-normal distribution, the operation time of X_2 and X_3 are fitted by mixture log-normal distributions, the finishing probability of operation time, $P(X_1 + X_2 + X_3 \geq T)$ where T is critical value of operation time, is hard to calculate since it doesn't have close form. Therefore the method of estimating the finishing probability of operation time for mixture log-normal distribution is worthy to discuss in the future.

In section 2 we introduce the definition of a finite mixture distribution and use the method of the maximum likelihood to estimate the parameters. The computa-

tion algorithm for finding the MLE of parameter vectors of mixture distributions under different considerations to accommodate the practical situation are also introduced, namely, the EM algorithm proposed first by Dempster. at al(1977), the EM algorithm for grouped and truncated data proposed by McLachlan and Jones (1988). Later the mixture of log-normal distribution will be used to fit our data. The standard errors estimates for MLE (θ) can be computed by taking the square root of the corresponding diagonal element of $I^{-1}(\theta)$, where $I(\theta)$ is corresponding Fisher information matrix. The Kolmogorov-Smirnov goodness of fit test for ascertaining whether an assumed probability distribution is consistent with a given set of data is also stated. In Section 3 likelihood ratio tests for testing the effects of main factors are formulated. In Section 4 we applied all these methods to the time data of the operation, and in Section 5 we conclude with a conclusion and discussion.

Table 1 : Thirteen operations in the GYN department.

Code	Operation
V1	Anterior-posterior colporrhaphy
V2	BSO + omentec. + ATH + retrope.Lm.R.D
V3	Enucleation of ovarian cyst
V4	Hysterectomy rad. cervical cance
V5	Hysterectomy, total extended
V6	Lapa.oophorectomy, partial/total
V7	Laparoscopy operative
V8	Laparotomy abdomen for 2nd look
V9	Myomectomy
V10	Salpingo-oophorectomy
V11	Total Hysterectomy (ATH/VTH)
V12	Total hysterectomy (LAVTH)
V13	Total vaginectomy resection

Table 2 : Descriptive statistics of operation times.

Operation	Total	Min	Max	Mean	Std.Dev.
V1	49	15	204	89.04	31.93
V2	131	80	959	247.15	109.56
V3	243	60	330	126.13	44.22
V4	403	84	540	256.89	59.34
V5	50	154	360	219.62	47.09
V6	96	60	520	137.96	68.94
V7	135	60	475	156.33	79.89
V8	60	90	435	185.07	65.34
V9	260	60	370	131.26	46.37
V10	113	65	520	138.68	64.07
V11	785	60	580	139.66	51.89
V12	947	65	575	149.41	54.19
V13	40	34	129	75.30	24.33

Table 3 : Descriptive statistics of the logarithm of the operation times.

Operation	Total	Min	Max	Mean	Std.Dev.
LV1	49	2.71	5.32	4.42	.40
LV2	131	4.28	6.87	5.43	.40
LV3	243	4.09	5.80	4.78	.32
LV4	403	4.43	6.29	5.52	.22
LV5	50	5.04	5.89	5.37	.20
LV6	96	4.09	6.25	4.84	.40
LV7	135	4.09	6.16	4.95	.43
LV8	60	4.50	6.05	5.17	.33
LV9	260	4.09	5.91	4.82	.32
LV10	113	4.17	6.25	4.85	.39
LV11	785	4.09	6.36	4.89	.30
LV12	947	4.17	6.35	4.95	.32
LV13	40	3.53	4.86	4.27	.34

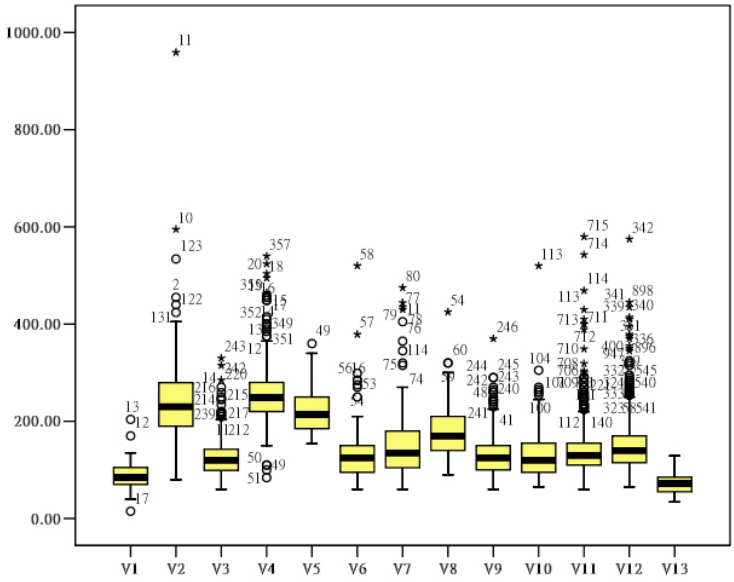


Figure 1 : Boxplots for operation times.

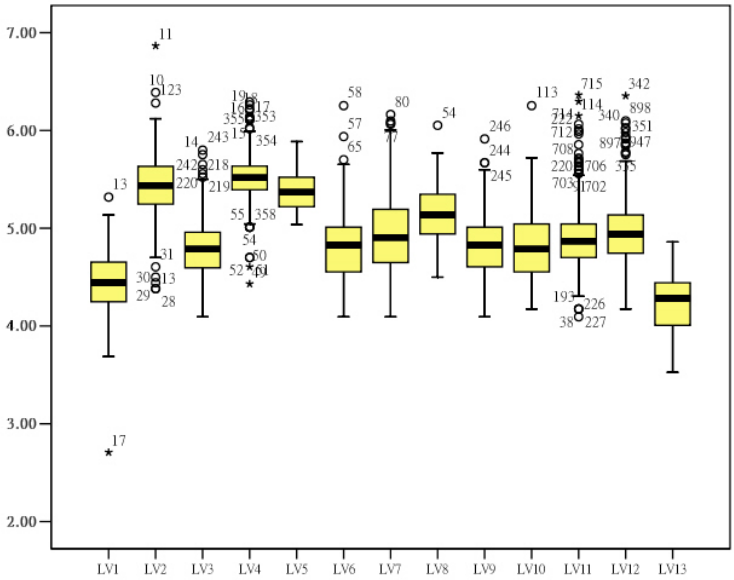


Figure 2 : Boxplots for the logarithm of the operation times.

2. Mixture distributions and method of estimation of parameters

Let the random variable X have probability density function(p.d.f) $f(x)$ of a mixture distribution with finite components, the mixture distributions can be represented in this form

$$f(x;\Phi) = \sum_{i=1}^{g} p_i f_i(x;\theta), \tag{1}$$

where $f_i(x;\theta)$ is the p.d.f corresponding to the ith component of the mixture and θ denotes the vector of all unknown parameters in the parametric forms adopted for these g component densities. $\Phi = (p', \theta')'$ be the vector of all unknown parameters, where $p = (p_1, ..., p_g)'$ is the vector of mixing proportions satisfying $\sum_{i=1}^{g} p_i = 1$.

Assume n independent observations, x_1,...,x_n, were obtained from a mixture distributions, then the vector of all unknown parameters of (1), $\Phi = (p', \theta')'$, will be estimated. For estimating the unknown parameters, we apply the standard maximum likelihood estimation (MLE) method. Not only is it appealing on intuitive grounds, but it also possesses desirable statistical properties such as, under very general conditions, the estimates obtained by the method are consistent.

In the case of mixture distributions, one of the well-know numerical method for finding the MLE is described in the following subsection.

2.1 Expectation-maximization (EM) algorithm

The expectation-maximization (EM) algorithm (Dempster et al., 1977) is an iterative procedure of maximum likelihood estimation for data containing missing values. Such a missing value problem includes as a particular case the estimation of the parameters of a mixture distributions from an observed sample. In order to train the mixture distributions, the EM algorithm is often applied to optimize model parameters because of it is easily programmable and satisfies a monotonic convergence property.

EM algorithm for computing the maximum likelihood estimation (MLE)

The EM algorithm is a standard technique and useful tool for obtaining maximum likelihood estimates for finite mixture models. Maximum likelihood estimation (MLE) can be undertaken via the EM algorithm of Dempster, Laird, and Rubin(1977); see also the monographs on mixture distributions by Everitt and Hand

(1981), Titterington, Smith, and Makov (1985), and more recently, McLachlan and Basford (1988). The M-steps and E-steps are repeated iteratively until some convergence criterion is satisfied.

EM algorithm for grouped and truncated data

Frequently in practice, data collected on the phenomenon of interest are available only in grouped form and may also be truncated. We consider here the fitting of finite mixture distributions to such data. Dempster et al. (1977) showed how EM algorithm can be used to carry out MLE for grouped and truncated data, although they did not consider specifically mixture distributions in this context. More recently, McLaren, Brittenham, and Hasselblad (1986) used the EM algorithm to fit a doubly truncated log-normal distribution in the modeling of the distribution of red blood cell volumes in healthy individuals and patients with anemia.

In our study, the operation time of the GYN department was recorded a unit of one minute. Hence the data sets we obtained can be regarded as in grouped form, and we selected the data sets which the operation time exceed sixty minutes since operation time is usually at least one hour. That is to say we truncated our observations before sixty minutes. Our observations are considered to be in grouped and truncated form.

Mclachlan and Jones (1988) considered the fitting of finite normal mixture models via the EM algorithm for data which are available only in grouped form and which may also be truncated. The detail of EM algorithm is given in Appendix (A). In our study here, we adopt the method of EM algorithm which be proposed by Mclachlan and Jones (1988) to estimate parameters.

2.2 The standard errors of MLE

Suppose that random variable X has a density function $f(x|\theta)$, where θ denotes unknown parameters, we define the Fisher information matrix $I(\theta)$ by

$$I(\theta) = E[\frac{\partial}{\partial\theta}\log f(X|\theta)]^2.$$

Under appropriate smoothness conditions on $f(x|\theta)$, $I(\theta)$ may also be expressed as

$$I(\theta) = -E[\frac{\partial^2}{\partial\theta^2}\log f(X|\theta)]. \tag{2}$$

The large sample distribution of a maximum likelihood estimate $(\hat{\theta})$ is approximately

normal distribution with mean θ_0 and variance $1/nI(\theta_0)$ where θ_0 is the true value of θ.

A corresponding result can be proved from the multidimensional case. The vector of maximum likelihood estimates is asymptotically normally distributed. The mean of the asymptotic distribution is the vector of true parameters, θ_0. The covariance of the estimates $\hat{\theta_i}$ and $\hat{\theta_j}$ is given by the ij entry of the matrix $n^{-1}I^{-1}(\theta_0)$, where $I(\theta)$ is the matrix with ij component

$$E[\frac{\partial}{\partial\theta_i}\log f(X|\theta)\frac{\partial}{\partial\theta_j}\log f(X|\theta)] = -E[\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log f(X|\theta)]. \tag{3}$$

The standard errors estimates for $\hat{\theta}$ can be computed by taking the square root of the corresponding diagonal elements of $n^{-2}I^{-1}(\hat{\theta})$.

2.3 Kolmogorov-Smirnov test

As soon as we have obtained the estimation of parameters, we need to test whether the particular estimated p.d.f. is consistent with those observed data. The Kolmogorov-Smirnov goodness of fit test will be used.

The data consist of a random sample $X_1, X_2, \ldots, X_n$ of size n associated with some unknown distribution function by $F(x)$. Let $S(x)$ be the empirical distribution function based on the random sample $X_1, X_2, \ldots, X_n$, and $F^*(x)$ be a completely specified hypothesized distribution function. If we had wished to test the null hypothesis

$$\begin{aligned} H_0 &: F(x) = F^*(x) \qquad \text{for all } x, \\ H_1 &: F(x) \neq F^*(x) \qquad \text{for at least one value of } x. \end{aligned}$$

Test statistic T is defined

$$T = \sup_x |F^*(x) - S(x)|.$$

Reject H_0 at the level of significance α if T exceeds the $1-\alpha$ quantile, or by using p-value given by

$$p = t\sum_{j=0}^{[n(1-t)]}\binom{n}{j}(1-t-\frac{j}{n})^{n-j}(t+\frac{j}{n})^{j-1} \tag{4}$$

where t is the observed value of the test statistic.

The Kolmogorov-Smirnov test may be preferred over the Chi-squared test for goodness of fit if the sample size is small; the Kolmogorov-Smirnov test is exact even for small samples, while the Chi-squared test assumes that number of observations is large enough so that the Chi-squared distribution provides a good approximation as the distribution of the test statistic. There is controversy over which test is the more powerful, but the general feeling seems to be that the Kolmogorov-Smirnov test is probably more powerful than the Chi-squared test in most situations involving ordinal data. For further comparisons please see Slakter(1965).

2.4 Tamhane T2 test

We hope to classify operation times base on log-normal distributions. To investigate pairs of significantly different operation time means, we need to use multiple comparison. Tamhane T2 procedure based on *Student's t* distribution was proposed by Tamhane (1977, 1979). Tamhane T2 test is based on assumptions that variances are unequal and sample sizes are unequal. SPSS software provides Tamhane T2 test for pairwise comparisons.

Consider an experiment with data y_{ij} satisfying the one-way, fixed-effects analysis of variance model

$$y_{ij} = \mu_i + e_{ij} \ ,$$

where the e_{ij} are independent with $e_{ij} \sim N(0, \sigma_i^2)$ and the μ_i and σ_i^2 are unknown, for i= 1, 2,. . . , k and j= 1, 2,. . . , n_i. Denote by $\overline{y}_i$ the ith sample mean and by s_i^2 an unbiased estimate of σ_i^2 based on ν_i degrees of freedom for level i, $n_i - 1$.

Two means are significantly different if

$$|\overline{y}_i - \overline{y}_j| \geq t_{\gamma,\ \hat{\nu}_{ij}}(\frac{s_i^2}{n_i} + \frac{s_j^2}{n_j})^{1/2},$$

where $t_{\gamma,\ \hat{\nu}_{ij}}$ is the two-sided γ point of *Student's t* distribution with $\hat{\nu}_{ij}$ df, where $\gamma = 1-(1-\alpha)^{1/k^*}$, α is experimentwise error rate under the complete null hypothesis, $k^* = k(k-1)/2$, and $\hat{\nu}_{ij}$ denotes as

$$\hat{\nu}_{ij} = \frac{(s_i^2/n_i + s_j^2/n_j)^2}{s_i^4/n_i^2\nu_i + s_j^4/n_j^2\nu_j}.$$

3. Effect testing

We now discuss that testing of factor effects to from finite mixture log-normal distributions. Yong. et al (2004) investigated the asymptotic properties of the likelihood ratio statistic for testing homogeneity in normal mixture models in the presence of a structural parameter. They showed that the ordinary likelihood ratio test has the simple χ^2-type null limiting distribution under some assumptions. More recently, Zhang. et al (2004) used the likelihood ratio test to test the existence of any QTL affecting the expression of an embryo or endosperm trait for mapping QTL is based on a mixture model. He thought the log-likelihood ratio test statistics is asymptotically χ^2 distributed. Hence we also use likelihood ratio tests to test whether factors of interest have significant effect to the mixture log-normal distribution model.

3.1 Likelihood ratio test

If $X_1, \ldots, X_n$ is a random sample from a population with p.d.f $f(x|\theta)$ where θ denotes the parameter vector of distribution. The likelihood function is defined to be

$$L(\theta) = \prod_{i=1}^{n} f(x_i|\theta).$$

To test $\theta \in \Omega_0$ versus $\theta \in \Omega_1$, the well known likelihood ratio test statistic is defined through

$$\lambda = \frac{\sup_{\theta\in\Omega_0} L(\theta)}{\sup_{\theta\in\Omega} L(\theta)},$$

where $\Omega = \Omega_0 \cup \Omega_1$. Under some additional regularity conditions, the asymptotic distribution of the statistic $-2\log\lambda$ is χ^2_k, where $k = \dim\Omega - \dim\Omega_0$. One rejects $\theta \in \Omega_0$ whenever $-2\log\lambda > C$, where C is determined by the desired level of the test.

In our study here, we are interested in knowing whether the factor of doctor has effect on the operation time. Let $\hat{\Phi}_{\mathrm{V}i}$, i= 4, 11, and 12, denote the parameter vectors of the distribution of V4, V11, and V12 operation time, respectively. If we want to test the effect of doctor at V4 operation, the test statistic used here is through

$$\lambda_{\mathrm{V4Doc}} = \frac{L(\hat{\Phi}_{\mathrm{V4}})}{L(\hat{\Phi}_{\mathrm{V4Doc4}})L(\hat{\Phi}_{\mathrm{V4Doc7}})L(\hat{\Phi}_{\mathrm{V4Doc9}})}, \tag{5}$$

where $L(\hat{\Phi}_{\mathrm{V4}}) = \prod_i^n f(x_i|\hat{\Phi}_{\mathrm{V4}})$ is the maximized likelihood for V4 operation time. $\hat{\Phi}_{\mathrm{V4Doc}i}$ is the estimation of the parameter vector of the time distribution of the ith doctor at V4 operation, i= 4, 7, and 9. $L(\hat{\Phi}_{\mathrm{V4Doc}i})$ is the maximized likelihood of

the ith time data set of doctor at V4 operation, i= 4, 7, and 9. Log-likelihood ratio $(-2\log\lambda_{\text{V4Doc}})$ is asymptotically χ^2 distributed with the degrees of freedom equals to the difference between the sum of the number of parameters of $\hat{\Phi}_{\text{V4Doc4}}$, $\hat{\Phi}_{\text{V4Doc7}}$, and $\hat{\Phi}_{\text{V4Doc9}}$ and the number of parameters of $\hat{\Phi}_{\text{V4}}$.

To test the effect of doctor at V11 operation, the test statistic λ_{V11Doc} is

$$\lambda_{\text{V11Doc}} = \frac{L(\hat{\Phi}_{\text{V11}})}{L(\hat{\Phi}_{\text{V11Doc19}})L(\hat{\Phi}_{\text{V11Doc267}})L(\hat{\Phi}_{\text{V11Doc3458}})}, \tag{6}$$

where $L(\hat{\Phi}_{\text{V11}})$ is the maximized likelihood for V11 operation. $\hat{\Phi}_{\text{V11Doc19}}$ denote the parameter vector of the combined time data sets of the 1st and 9th doctor at V11 operation. $\hat{\Phi}_{\text{V11Doc267}}$ is the parameter vector of the combined time data sets of the 2nd, 6th, and 7th doctor at V11 operation. Similarly, $\hat{\Phi}_{\text{V11Doc3458}}$ is the parameter vectors of the combined time data sets of the 3rd, 4th, 5th and 8th doctor at V11 operation. Log-likelihood ratio $(-2\log\lambda_{\text{V11Doc}})$ is asymptotically χ^2 distributed with the degrees of freedom equal to the difference in dimension between the sum of the number of parameters of $\hat{\Phi}_{\text{V11Doc19}}$, $\hat{\Phi}_{\text{V11Doc267}}$, and $\hat{\Phi}_{\text{V11Doc3458}}$ and the sum of the number of parameters of $\hat{\Phi}_{\text{V11}}$.

To test the effect of doctor at V12 operation, the test statistic is adopted in the same manner, where

$$\lambda_{\text{V12Doc}} = \frac{L(\hat{\Phi}_{\text{V12}})}{L(\hat{\Phi}_{\text{V12Doc1379}})L(\hat{\Phi}_{\text{V12Doc2}})L(\hat{\Phi}_{\text{V12Doc4568}})}, \tag{7}$$

where $L(\hat{\Phi}_{\text{V12}})$ is the maximized likelihood for V12 operation. $\hat{\Phi}_{\text{V12Doc1379}}$ denote the parameter vector of the combined time data sets of the 1st, 3rd, 7th, and 9th doctor at V12 operation. $\hat{\Phi}_{\text{V12Doc2}}$ is the parameter vector of the time data set of the 2nddoctor at V12 operation. Similarly, $\hat{\Phi}_{\text{V11Doc4568}}$ is the parameter vectors of the combined time data sets of the 4th, 5th, 6th and 8th doctor at V12 operation. Log-likelihood ratio $(-2\log\lambda_{\text{V12Doc}})$ is asymptotically χ^2 distributed with the degrees of freedom equal to the difference in dimension between the sum of the number of parameters of $\hat{\Phi}_{\text{V12Doc1379}}$, $\hat{\Phi}_{\text{V12Doc2}}$, and $\hat{\Phi}_{\text{V12Doc4568}}$ and the sum of the number of parameters of $\hat{\Phi}_{\text{V12}}$.

Regardless of doctors, the effect of stage of the patients' disease at V4 operation can be tested by

$$\lambda_{\text{V4Stage}} = \frac{L(\hat{\Phi}_{\text{V4}})}{L(\hat{\Phi}_{\text{V4Stage1}})L(\hat{\Phi}_{\text{V4Stage2}})L(\hat{\Phi}_{\text{V4Stage3}})}. \tag{8}$$

$\hat{\Phi}_{\text{V4Stage}i}$ is the estimation of the parameter vector of the time distribution of the ith stage at V4 operation, i= 1, 2, and 3. $L(\hat{\Phi}_{\text{V4Stage}i})$ is the maximized likelihood

of the ith time data set of stage at V4 operation, i= 1, 2, and 3. Log-likelihood ratio ($-2\log\lambda_{\text{V4Doc}}$) is asymptotically χ^2 distributed with the degrees of freedom equals to the difference between the sum of the number of parameters of $\hat{\Phi}_{\text{V4Stage1}}$, $\hat{\Phi}_{\text{V4Stage2}}$, and $\hat{\Phi}_{\text{V4Stage3}}$ and the number of parameters of $\hat{\Phi}_{\text{V4}}$.

4. Statistical analysis results

In our work, we focus on studying the operation time distributions of thirteen operations of the gynecology (GYN) department. The sets of observed frequency counts in histogram form are given in Appendix (C). We fit log-normal distributions for those data and test results are given in Table 4. For those data sets which can be fitted by log-normal distributions, we use standard one way analysis of variance to compare the differences between means and characterize the operations into different categories with "short", "medium", or "long" operation times. For those data sets which are rejected as fitted by the K-S test, they are further fitted by a three-component mixture of log-normal distribution instead, and later we examine the effect of doctor and stage of the patients' disease on the operation time which those fitting are rejected.

4.1 Classification of operations based on operation time

Each operation time distribution is fitted by log-normal distribution and verified by Kolmogorov-Smirov gooodness of fit test respectively. From Table 4, we know the results of fitting log-normal distribution.

Table 4 : Results of K-S test.

Operation	LV1	LV2	LV3	LV4	LV5	LV6	LV7
K-S Test	.354	.180	.349	.004	.804	.249	.465
Operation	LV8	LV9	LV10	LV11	LV12	LV13	
K-S Test	.768	.244	.216	.000	.017	.158	

For those data sets of operation time (V1, V2, V3, V5, V6, V7, V8, V9, V10, and V13) which can be fitted by log-normal distributions, we use standard one way analysis of variance to compare the differences between means. From Table 5, because Sig. = .000, we conclude that the means of those data sets of the operation time have significant differences.

Because it is significant differences between them, we then use Tamhane T2 test to compare how they are differed in means. Tamhane T2 test is used to compare all pair of means with unequal group variances. A graph underlying those means of the operation time that are not significantly different is shown in Figure 3. We can characterize the operations into different categories with "short", "medium", "long", or "ultra-long" operation times from results of Tamhane T2 test. From the analysis we see that there are significant differences between all pairs of means except V1 and V13, V3, V6, V7, V9, and V10, V2 and V5. Therefore we categorize V1 and V13 with short operation times; V3, V6, V7, V9, and V10 with medium operation times; V8 with long operation times; and V2 and V5 with ultra-long operation times, respectively.

Table 5 : One way ANOVA.

Source	Sum of squares	df	Mean square	F_0	Sig.
Operations	82.806	9	9.201	72.353	.000
Error	146.748	1154	.127		
Total	229.555	1163			

V1	V13	V3	V9	V6	V10	V7	V8	V2	V5
4.25	4.42	4.78	4.82	4.83	4.85	4.95	5.16	5.37	5.43

Figure 3 : Results of Tamhane T2 test

4.2 Mixture log-normal distribution

From Table 4, we know that the time data sets of V4, V11, and V12 operation are rejected as fitted by the K-S test. The sets of observed frequency counts in histogram form presented in Appendix (D) suggests that the operation time distributions are mixture models, which also seem to be able to explain the real situation reasonably. Hence we will fit a more general mixture log-normal distribution for those data sets of V4, V11, and V12 operation time.

A random variable Y is said to be log-normally distributed if $X{=}\log Y$ is normally distributed. Only positive values are possible for the variable Y, and the distribution is skewed to the left. The p.d.f. of log-normal distribution is

$$f(y|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}\frac{e^{-(\log y-\mu)^2/2\sigma^2}}{y}, 0 \le y < \infty, -\infty < \mu < \infty. \tag{9}$$

For those operation (V4, V11, and V12) times which are rejected as fitted by the K-S test, they are fitted by a more general mixture log-normal distribution instead. Since the operation time was recorded a unit of one minute, the time data sets we obtained can be regarded as in grouped form, and because the operation time is usually at least one hour, we selected the data with operation time exceeding sixty minutes. Hence our observations are considered to be in grouped and truncated form. We can think of the data which operation time is less than sixty minutes as special case and discuss them further. In order to obtain the MLE of mixture of log-normal distributions for our data, the formulas for performing EM algorithm and EM algorithm for grouped and truncated data are used in our estimates. We first fit three-component mixture of log-normal distributions to these data set of operation time (V4, V11, and V12) via formulas deduced by McLachlan and Jones (1988).

To estimate the MLE of parameter vector $\Phi = (p_1, p_2, \mu_1, \mu_2, \mu_3, \sigma_1, \sigma_2, \sigma_3)$ of a mixture of three log-normal distributions, we only take the logarithm of the original data and estimate the MLE of parameter vector of a mixture of three normal distributions. Because the minute is a unit of operation time, we try to group the operation time into five types, that is, each one of the interval of operation time is one, two, three, four and five minutes respectively. For example, if we consider six observations of operation time are 60, 61, 62, 63, 64, 65 minutes, then the grouping is over six intervals of equal width of one minute, that is, one frequency for each interval. For the interval of equal width of two minutes, we have [60,62), [62,64), and [64,66) three intervals, and each interval has two frequencies. According to the EM algorithm for mixture of log-normal distributions, we obtain the estimates of parameter vectors $\hat{\Phi}_i$, i= 4, 11, and 12, listed at Table 6, 7, and 8 respectively.

To test whether the three-component mixture of log-normal distributions is consistent with the observed data, the Kolmogorov-Smirnov goodness-of-fit statistic is considered. The test results are also given in Table 6, 7, and 8 for each time data of V4, V11, and V12 operation along with the associated p-value. According to the results of Table 6, 7, and 8 we find that the interval of different width influence the results of fitting mixture distributions to operation time. It seems interested in the results. Here we adopt the results which each one of the interval of operation time is one minute. Hence in the following the discussion will be restricted to the estimates obtained by each one of the interval of operation time is one minute. The plots of the estimated density function for the time data sets of V4, V11, and V12 operations are presented in Figure 4 and it can be seen that three-component log-normal distribution fits the data sets of V4, V11, and V12 operation time.

Table 6 : Results of fitting a three-component mixture of log-normal distributions to V4 operation time. (Standard errors are in parentheses)

	Minute(s)/Group				
V4 $\hat{\Phi}_4$	1	2	3	4	5
$\hat{p}_1$	.010(.005)	.010(.005)	.010(.005)	.009(.005)	.009(.005)
$\hat{p}_2$	.745(.047)	.729(.047)	.729(.047)	.727(.047)	.728(.047)
$\hat{\mu}_1$	4.613(.062)	4.619(.063)	4.633(.066)	4.634(.067)	4.634(.059)
$\hat{\mu}_2$	5.494(.010)	5.496(.010)	5.497(.010)	5.504(.010)	5.502(.010)
$\hat{\mu}_3$	5.718(.034)	5.668(.035)	5.666(.033)	5.667(.033)	5.660(.034)
$\hat{\sigma}_1$	.110(.047)	.113(.047)	.119(.050)	.118(.051)	.105(.045)
$\hat{\sigma}_2$	.144(.008)	.144(.008)	.144(.008)	.142(.008)	.140(.008)
$\hat{\sigma}_3$	.285(.024)	.288(.024)	.273(.023)	.277(.023)	.284(.024)
K-S p-value	.630	.395	.247	.087	.132

Table 7 : Results of fitting a three-component mixture of log-normal distributions to V11 operation time. (Standard errors are in parentheses)

	Minute(s)/Group				
V11 $\hat{\Phi}_{11}$	1	2	3	4	5
$\hat{p}_1$	.269(.031)	.219(.031)	.202(.031)	.198(.031)	.193(.030)
$\hat{p}_2$	.374(.032)	.343(.031)	.340(.031)	.337(.031)	.331(.031)
$\hat{\mu}_1$	4.720(.024)	4.734(.030)	4.737(.032)	4.747(.034)	4.721(.033)
$\hat{\mu}_2$	4.824(.017)	4.824(.016)	4.826(.018)	4.830(.019)	4.822(.019)
$\hat{\mu}_3$	5.100(.027)	5.037(.024)	5.307(.023)	5.042(.025)	5.039(.023)
$\hat{\sigma}_1$	.214(.018)	.205(.025)	.199(.026)	.202(.028)	.195(.026)
$\hat{\sigma}_2$	.166(.014)	.149(.014)	.162(.015)	.167(.015)	.165(.015)
$\hat{\sigma}_3$	.358(.018)	.391(.016)	.388(.016)	.422(.017)	.378(.015)
K-S p-value	.263	.205	.056	.023	.050

Table 8 : Results of fitting a three-component mixture of log-normal distributions to V12 operation time. (Standard errors are in parentheses)

	Minute(s)/Group				
V12 $\hat{\Phi}_{12}$	1	2	3	4	5
$\hat{p}_1$	.188(.023)	.141(.022)	.121(.021)	.126(.020)	.121(.019)
$\hat{p}_2$	.306(.028)	.278(.027)	.243(.026)	.255(.026)	.250(.026)
$\hat{\mu}_1$	4.643(.037)	4.648(.044)	4.645(.050)	4.622(.048)	4.568(.052)
$\hat{\mu}_2$	4.854(.022)	4.803(.021)	4.807(.024)	4.817(.022)	4.620(.023)
$\hat{\mu}_3$	5.100(.019)	5.095(.017)	5.079(.016)	5.094(.016)	5.077(.016)
$\hat{\sigma}_1$	.344(.027)	.205(.031)	.199(.035)	.202(.034)	.195(.037)
$\hat{\sigma}_2$	.197(.019)	.149(.018)	.162(.020)	.167(.018)	.165(.019)
$\hat{\sigma}_3$	.324(.014)	.391(.012)	.388(.011)	.422(.011)	.378(.012)
K-S p-value	.303	.159	.023	.009	.059

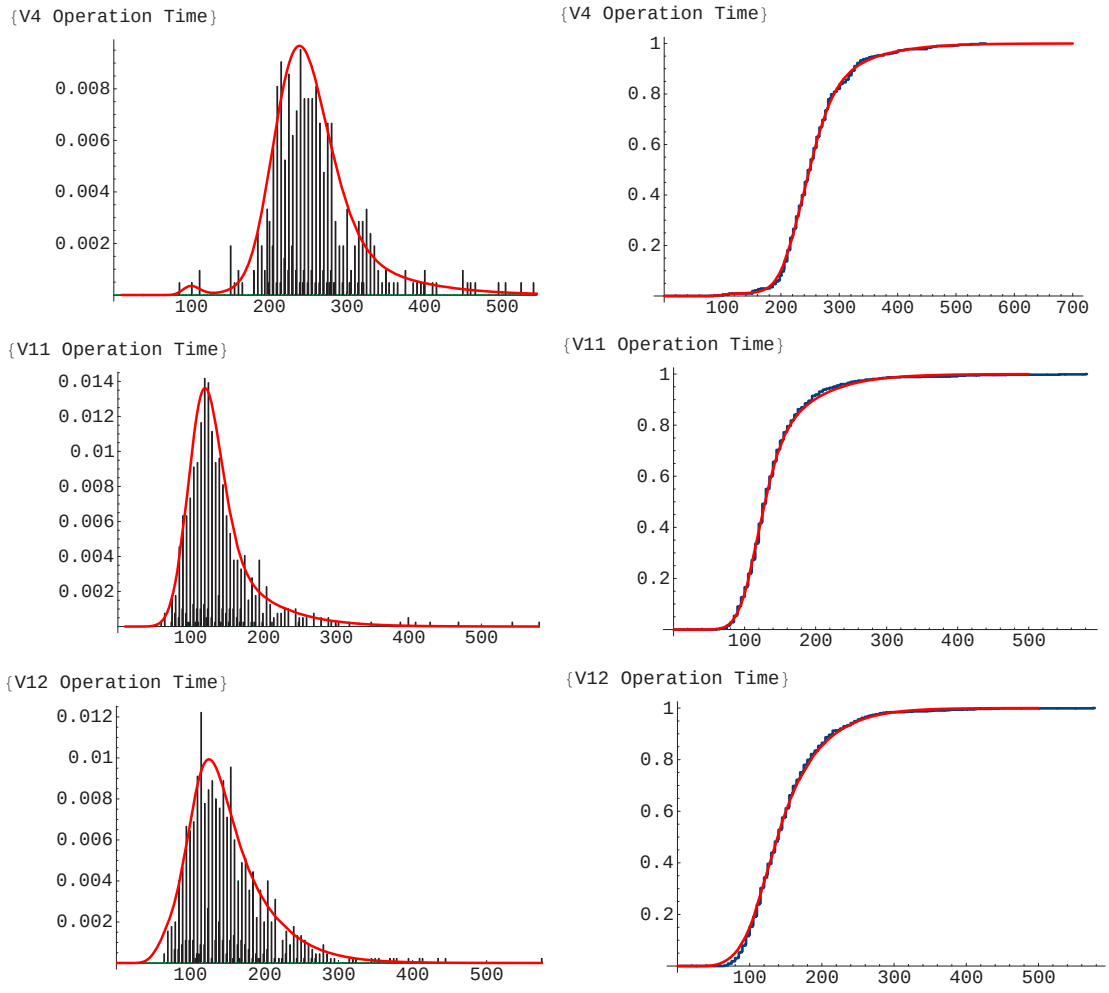


Figure 4 : Plots of mixture of log-normal distributions for V4, V11, and V12 operation times.

4.3 Effect of doctor for three operations

We now examine the effect of doctor on the data sets of operation time which are accepted as fitted three-component mixture of log-normal distributions by the K-S test. Regarding the data sets of V4, V11 and V12 operation time, we may proceed to consider the likelihood ratio test to test the effect of doctor.

In order to test the effect of doctor, we combine the data sets for different doctors combinations and fit mixture of log-normal distributions to each combined data set. The descriptive statistics of the time data sets of doctors at V4 operation are given in Table 9. Because Doctor 4 , Doctor 7, and Doctor 9 performed the V4 operation, we draw out the time data sets of Doctor 4, Doctor 7, and Doctor 9 at V4 operation, and fit mixture of log-normal distributions, respectively. The MLE for the time data sets of ith doctor, i= 4, 7, and 9, at V4 operation are denoted by $\hat{\Phi}_{\text{V4Doc}i}$, which are displayed in Table 10. The maximized likelihood $L(\hat{\Phi})$ of each data sets of doctors at V4 operation time are also listed in Table 10.

The descriptive statistics of the time data sets of doctors at V11 operation are given in Table 11. We combine the time data sets of 1st and 9th; 2nd, 6th, and 7th; 3rd, 4th, 5th, and 8th, doctor at V11 operation and fit mixture of log-normal distributions, respectively. The analysis of the MLE for combined time data sets and the maximized likelihood of each data sets of doctors at V11 operation time are are shown in Table 12.

Similarly, for V12 operation the descriptive statistics of the time data sets of doctors are given in Table 13. We also combine the time data sets of 2nd; 1st, 3rd, 7th, and 9th; 4th, 5th, 6th, and 8th, doctor at V12 operation and fit mixture of log-normal distributions, respectively. The results of the MLE for combined data sets and the maximized likelihood of each data sets of doctors at V12 operation time are listed in Table 14.

In the practical problem here it is of interest to know whether the factor of doctor has effect on the operation time. Hence we test the effect of doctor, i.e., the null hypothesis is H_0: the effect of doctor is not significant. We consider the likelihood ratio test statistics λ in and (5), (6), and (7). The results listed in Table 15 show that at level α=.05 the effect of doctor at V4, V11, and V12 operation time which are indicated by λ_{V4Doc}, λ_{V11Doc}, and λ_{V12Doc} separately is significant, since p-value are almost zero. Therefore we conclude that there is obviously indication that there is significant effect of doctor at V4, V11, and V12 operation time.

Table 9 : Descriptive statistic of doctors at V4 operation time.

Doctor	N	%	Mean	Std. Dev.
4	23	.057	5.805	.297
7	24	.060	5.467	.162
9	355	.883	5.510	.208

Table 10: Results of fitting mixture of log-normal distributions to data sets of doctors at V4 operation time. (Standard errors are in parentheses.)

	V4 Total $\hat{\Phi}_{\text{V4}}$	Doctor 4 $\hat{\Phi}_{\text{V4Doc4}}$	Doctor 7 $\hat{\Phi}_{\text{V4Doc7}}$	Doctor 9 $\hat{\Phi}_{\text{V4Doc9}}$
$\hat{p}_1$	.010(.005)	-	-	.027(.010)
$\hat{p}_2$	.745(.047)	-	-	.832(.052)
$\hat{\mu}_1$	4.613(.062)	5.805(.062)	5.467(.033)	4.644(.238)
$\hat{\mu}_2$	5.494(.010)	-	-	5.513(.057)
$\hat{\mu}_3$	5.718(.034)	-	-	5.559(.010)
$\hat{\sigma}_1$	.110(.047)	.297(.044)	.162(.023)	.533(.162)
$\hat{\sigma}_2$	.144(.008)	-	-	.153(.040)
$\hat{\sigma}_3$	.285(.024)	-	-	.289(.008)
K-S p-value	.630	.942	.786	.556
$\log L(\hat{\Phi})$	-2155.189	-137.749	-121.68	-1874.383

Table 11 : Descriptive statistic of doctors at V11 operation time.

Doctor	N	%	Mean	Std. Dev.
1	37	.047	4.867	.202
2	29	.037	4.607	.267
3	33	.042	5.037	.267
4	17	.023	5.124	.440
5	33	.042	5.029	.204
6	2	.003	5.074	.978
7	48	.061	4.761	.225
8	26	.033	4.990	.447
9	560	.712	4.887	.286

Table 12 : Results of fitting mixture of log-normal distributions to data sets of doctors at V11 operation time. (Standard errors are in parentheses.)

	V11 Total $\hat{\Phi}_{\text{V11}}$	Doctor 1.9 $\hat{\Phi}_{\text{V11Doc19}}$	Doctor 2.6.7 $\hat{\Phi}_{\text{V11Doc267}}$	Doctor 3.4.5.8 $\hat{\Phi}_{\text{V11Doc3458}}$
$\hat{p}_1$	.269(.031)	.276(.035)	-	-
$\hat{p}_2$	.374(.032)	.370(.038)	-	-
$\hat{\mu}_1$	4.720(.024)	4.736(.023)	4.712(.031)	5.037(.032)
$\hat{\mu}_2$	4.824(.017)	4.864(.020)	-	-
$\hat{\mu}_3$	5.100(.027)	5.020(.028)	-	-
$\hat{\sigma}_1$	.214(.018)	.169(.019)	.278(.022)	.332(.022)
$\hat{\sigma}_2$	.166(.014)	.168(.017)	-	-
$\hat{\sigma}_3$	.358(.018)	.343(.019)	-	-
K-S p-value	.263	.110	.380	.450
$\log L(\hat{\Phi})$	-3963.138	-2969.740	-382.755	-582.98

Table 13 : Descriptive statistic of doctors at V12 operation time.

Doctor	N	%	Mean	Std. Dev.
1	58	.061	4.920	.293
2	70	.074	4.607	.261
3	62	.065	4.931	.246
4	161	.170	5.157	.353
5	35	.037	5.156	.201
6	14	.015	5.046	.465
7	64	.068	4.828	.269
8	92	.097	5.065	.285
9	391	.413	4.908	.273

Table 14 : Results of fitting mixture of log-normal distributions to data sets of doctors at V12 operation time. (Standard errors are in parentheses.)

	V12 Total $\hat{\Phi}_{\text{V12}}$	Doctor 2 $\hat{\Phi}_{\text{V12Doc2}}$	Doctor 1.3.7.9 $\hat{\Phi}_{\text{V12Doc1379}}$	Doctor 4.5.6.8 $\hat{\Phi}_{\text{V12Doc4568}}$
$\hat{p}_1$	.188(.023)	-	.207(.037)	-
$\hat{p}_2$	.306(.028)	-	.399(.042)	-
$\hat{\mu}_1$	4.643(.037)	4.607(.031)	4.781(.042)	5.124(.019)
$\hat{\mu}_2$	4.854(.022)	-	4.894(.028)	-
$\hat{\mu}_3$	5.100(.019)	-	4.982(.028)	-
$\hat{\sigma}_1$	.344(.027)	.261(.022)	.209(.033)	.327(.013)
$\hat{\sigma}_2$	.197(.019)	-	.261(.021)	-
$\hat{\sigma}_3$	.324(.014)	-	.292(.019)	-
K-S p-value	.303	.251	.312	.647
$\log L(\hat{\Phi})$	-4962.055	-327.212	-2887.211	-1730.165

Table 15 : Likelihood ratio test under the hypothesis- H_0 : the effect of doctor is not significant.

	V4	V11	V12	Total Effect
-2log λ	42.75	55.33	34.94	133.02
$d.f.$	4	4	4	12
p-value	10^{-8}	10^{-11}	10^{-7}	0

4.4 Effect of stage for V4 operation time

Generally speaking, intuitively we feel that that the disease severity of the patients concerns the operation time for longer operation time. We conjecture that the disease severity of the patients influence the length of the operation time. Here we gathered the data sets where doctors made diagnoses of the disease severity of the patients for V4 operation time. We use "stage" to express the disease severity of the patients. Stage 1 means that the disease severity of the patients is incipient, stage 2 means that the disease severity of the patients is middle, and stage 3 means that the disease severity of the patients is terminal. Since the stage data sets are not to collect, we have only collected 148 data sets on stage and some other variables for V4 operation in number and denoted by $\text{V4}_{(148)}$. Now we want to examine the effect of stage of the patients' disease on the time data sets of $\text{V4}_{(148)}$ operation and try to illustrate the reason why the overall operation time has the special pattern

of mixture log-normal distribution.

Since we have only partial data sets, we combine the data sets for three different stages of the patients' disease. In order to test the effect of stage of the patients' disease, we fit mixture of log-normal distributions to each combined data set. The summary statistics of the time data sets of stages of the patients' disease at V4 operation are given in Table 16. We draw out the time data sets of stage 1, stage 2, and stage 3 at $V4_{(148)}$ operation, and fit mixture of log-normal distributions, respectively. The MLE for the time data sets of ith stage, i= 1, 2, and 3, at $V4_{(148)}$ operation are denoted by $\hat{\Phi}_{\text{V4Stage}i}$, which are displayed in Table 17. The maximized likelihood $L(\hat{\Phi})$ of each data sets of stage of the patients' disease at $V4_{(148)}$ operation time are also listed in Table 17.

We test the effect of stage at V4 operation time, i.e., the null hypothesis is H_0: the effect of stage at $V4_{(148)}$ operation time is not significant. We consider the likelihood ratio test statistics λ in (8) and the results listed in Table 18 show that at level α= .05 the effect of stage at V4 operation time which is indicated by λ_{V4Stage} is significant, since p-value is .001 smaller than .05 . For that reason we conclude that there is obviously indication that there is significant effect of stage at $V4_{(148)}$ operation time.

Table 16 : Summary statistic of stages at $V4_{(148)}$ operation time.

Stage	N	%	Mean	Std. Dev.
1	11	.074	5.427	.262
2	109	.736	5.410	.141
3	28	.190	5.461	.336

Table 17: Results of fitting mixture of log-normal distributions to data sets of stages at $V4_{(148)}$ operation time. (Standard errors are in parentheses.)

	$V4_{(148)}$ $\hat{\Phi}_{\text{V4}}$	Stage 1 $\hat{\Phi}_{\text{V4Stage1}}$	Stage 2 $\hat{\Phi}_{\text{V4Stage2}}$	Stage 3 $\hat{\Phi}_{\text{V4Stage3}}$
$\hat{p}_1$	.165(.040)	-	.170(.040)	-
$\hat{\mu}_1$	5.237(.100)	5.427(.023)	5.225(.087)	5.461(.030)
$\hat{\mu}_2$	5.419(.008)	-	5.414(.007)	-
$\hat{\sigma}_1$	.478(.072)	.262(.017)	.420(.062)	.336(.021)
$\hat{\sigma}_2$	.077(.006)	-	.067(.005)	-
K-S p-value	.713	.877	.958	.575
$\log L(\hat{\Phi})$	-740.616	-60.087	-509.431	-161.571

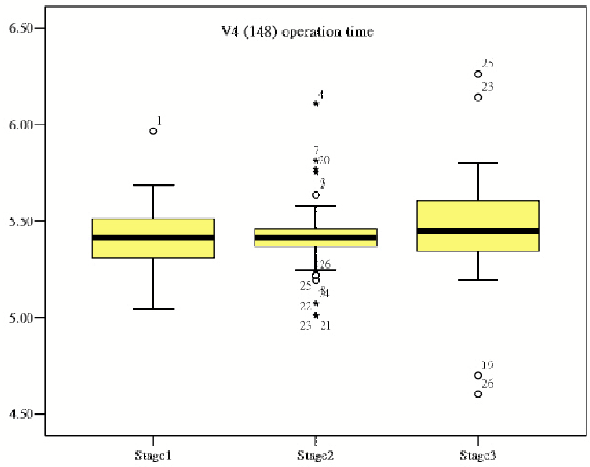


Figure 5 : Boxplots of stage for $V4_{(148)}$ operation time.

Table 18 : Likelihood ratio test for the effect of stage.

	$V4_{(148)}$
-2log λ	19.052
$d.f.$	4
p-value	.001

4.5 Other variables for V4 operation

Besides doctor and stage, we also collected data on four factor variables which may affect V4 operation time distribution: residents practicing years (R. Year), blood transfusion (Blood T), blood loss (Blood L), and patient's age (Age). We only focus on analyzing variables for V4 operation time in stage 2 as there are more data in that category. The correlation matrix of those variables are shown in Table 19. Comparing the correlation matrix, we find the correlations between Time and Blood L, Blood L and Blood T and Blood T and Age are more significant than others.

In multiple regression, we may use different models to find what kind of variables affect the operating time most. But under the assumption that the error follows normal distribution, the usual method in multiple regression dose not work for the operation time while fitted by mixture log-normal distribution. How to find a method to develop a operation time forecasting model should be investigated.

We take the logarithm of the data set of operation time in stage 2 and fit two-component mixture of bivariate normal distributions to those data sets of operation time and blood loss in stage 2 via the EM algorithm. Table 20 presents the estimated results, i.e. mean vector, covariance matrices and the weight. Then we try to make 95% prediction confidence ellipse for operation time based on two-component mixture of bivariate normal distributions. We do the scatter plot with 95% prediction confidence ellipses from component 1 and component 2 with mean vector ($\hat{\mu}$) and covariance matrix ($\hat{\Sigma}$) to explore relationships among pairs of variables respectively. Figure 6 displays a scatter plot with 95% prediction confidence ellipses.

The broken curve is 95% confidence ellipse for the distribution corresponding to the component 1 of the mixture and the solid curve is 95% confidence ellipse for the distribution corresponding to the component 2 of the mixture. We think that the probability which is on the ellipse of component 1 is $p_1(.356) \times 0.95 = 0.338$ and the probability which is on the ellipse of component 2 is $p_2(.644) \times 0.95 = 0.612$. The probability of joint confidence region is 0.272. Scatter plots of Blood L and Blood T and Blood T and Age are presented in Figure 7 and Figure 8 respectively. The ellipse shows graphically a positive correlation between variables Time and Blood L. If operation time falls outside the ellipse, we say it is not normal case and need further discussion.

Table 19 : Correlation Matrix

Correlation	Time(Stage2)	R. Year	Blood L	Blood T	Age
Time(Stage2)	1	-.032	.371	.187	-.050
R. Year	-.032	1	-.043	.047	-.098
Blood L	.371	-.043	1	.688	.131
Blood T	.187	.047	.688	1	.326
Age	-.050	-.098	.131	.326	1

Table 20 : The fitted results.(Comp. component)

	Comp. 1	Comp. 2
Mean($\hat{\mu}^T$)	5.338 431.018	5.439 881.123
Covariance	$\hat{\Sigma}_1$	$\hat{\Sigma}_2$
Weight($\hat{p}$)	.356	.644

$$\left(\hat{\Sigma}_1 = \begin{bmatrix} .017 & 4.563 \\ 4.563 & 12821.580 \end{bmatrix}, \quad \hat{\Sigma}_2 = \begin{bmatrix} .011 & 18.548 \\ 18.548 & 145349.079 \end{bmatrix} \right)$$

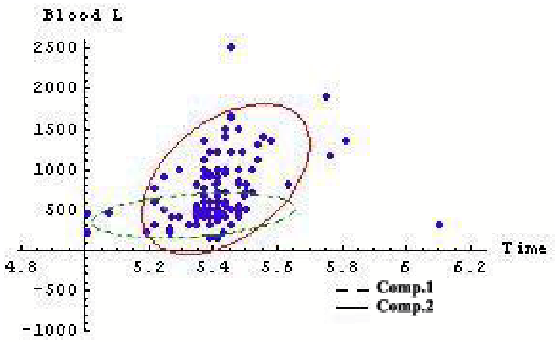


Figure 6 : Scatter plot with confidence ellipse for V4 operation time in stage 2.

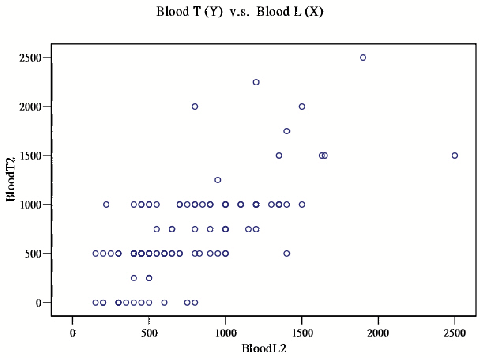


Figure 7 : Scatter plot of Blood L and Blood T for V4 operation in stage 2

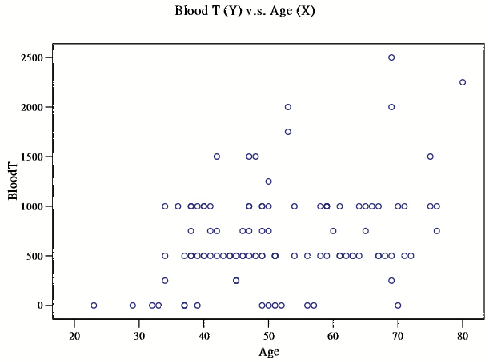


Figure 8 : Scatter plot of Blood T and Age for V4 operation in stage 2

4.6 Poisson regression model for V4 operation

The best known models for counts data assume a poisson distribution for the response. Suppose Y has poisson distribution with mean λ, then using a generalized linear model with log link, we get a poisson regression model

$$\log(\lambda) = \alpha + \beta X.$$

Goal of the study here is to find a model for relations between blood transfusion (Blood T), blood loss (Blood L), and patient's age (Age). The main effect of "Blood L" and "Age" are denoted by X_{bloodl} and X_{age}. We only analyze variables for V4 operation time in stage 2 and use the statistical software S-PLUS to fit those data but only consider the main effect of "Blood L" and "Age". All of fitted results using S-PLUS software are listed in Appendix (B) and partial results are shown in Table 21.

According Table 21, we know assuming that the blood transfusion (Blood T) has a poisson distribution, we get a model of

$$\log(\lambda) = \alpha + \beta_1 X_{\text{bloodl}} + \beta_2 X_{\text{age}}, \tag{10}$$

where $\hat{\alpha} = -0.443$, $\hat{\beta}_1 = 0.001$, and $\hat{\beta}_2 = 0.012$. All of main factor terms with large absolute t value (i.e. > 2) are the candidate of the significant terms. The poisson regression model is good because the deviance is 87.28839, which is much smaller than the critical value of chi-square distribution with 107 degrees of freedom at level 0.05 ($\chi^2_{105,.05} = 129.918$). Hence we accept H_0: the poisson model provides a good fit. In other words, the model in this fitting is a good description of those data. We display a plot of the response variable versus the fitted values in Figure 9. The broken line is a diagonal line and the solid line is a poisson regression line.

Specifically, blood transfusion can be forecasted by using poisson regression model (10). For example, if a patient who age and blood loss are 50 years old and 1000 c.c, average blood transfusion of the patient is

$$\exp\left(-0.44285 + 0.00101 \times 1000 + 0.01218 \times 50\right) = 3.242.$$

We can infer that the patient is transfused about three bags of blood.

Table 21 : Results of fitting poisson regression model

	Value	Std. Error	t value
Intercept	-0.44285	0.26379	-1.67882
BloodL	0.00101	0.00014	7.37748
Age	0.01218	0.00465	2.61800
Null Deviance: 154.7614 on 107 degrees of freedom			
Residual Deviance: 87.28839 on 105 degrees of freedom			

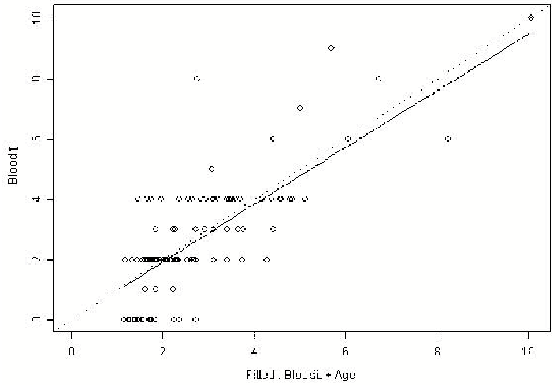


Figure 9 : Response variable versus the fitted values.
The solid line is a poisson regression line.

5. Conclusion and discussion

In this work, we have studied the operation time distributions of thirteen operations. Log-normal and mixture log-normal distribution are identified to be acceptable statistically in describing these operation times respectively. In order to obtain the MLE of mixture log-normal distributions, the method of computing EM algorithm for grouped and truncated data is used in our estimates. In order to classify those operations with operation times fitted by log-normal distributions into different categories such as "short", "medium", "long", "ultra-long" operation times, we use standard one-way analysis of variance and Tamhane T2 test to compare the differences between means of operation time. Finally we investigate the effect of doctor and stage on the data sets of operation time for those identified as three-component mixture of log-normal distributions by the K-S test.

According to the results from Table 6, 7, and 8, we find that bandwidth interval selection in fitting the distribution would make a difference in the fitting of the mixture distributions on operation time. For V11 operation, test result with one minute as the basic interval unit situation is different from test result with that of four minutes. It is worth noting that through rigorous statistical analysis presented here, it helps to provide an objective estimation on the distribution of the operation time.

To investigate effects of other factor variables which may affect V4 operation time distribution, we only have some partial information. Those factor variables reveals only part of the affecting operation time distribution, not express all of situations. If we can collect and discuss all of factor variables of V4, V11, and V12 operation time, it could enable us to understand further about operation time distributions.

On the whole, it seems one main factor affecting the operation time is the disease severity of the patients. The doctors usually diagnose the disease severity of the patients into three groups, that is, incipient, middle, and terminal. In general, the stage of patients' disease is in proportion to operation time. We conjecture that the patients with disease severity as terminal have longer operation time, and the patients with disease severity as incipient have shorter operation time. For operations with large data set, operation time appears to be affected by the stage of patients' disease more clearly. It is of interest to see if three proportions of the disease severity of the patients (incipient, middle, and terminal) are significantly related to the estimates of the vector of mixing proportions.

Furthermore, we have also established a poisson regression model for relationship

for blood transfusion with respect to blood loss and age of the patient. Later blood transfusion can be predicted by using poisson regression model. We relationship between operation time and blood loss are also interested in finding the which seem to be distributed as mixture of bivariate normal distributed. We can judge further whether operation time is normal or not . Hence it is worth discussing that how to make a 95% prediction confidence region based on mixture of bivariate normal distribution.

Understanding the operation time distributions and the factor effect to operation time is helpful in making an efficient operating room scheduling in the future. It is of interest to develop a method to model the operating time and make estimation of the probability to finish the operation on time with mixture log-normal distribution, and predict the extra cost due to over time operation.

References

1. Conover W. J. (1999) Parctical nonparametric statistics. New York: Wiley.

2. Dempster, A. P., Laird, N. M., and Rubin, D. R. (1977). Maximum likelihood from incomplete data (with Discussion). *Journal of the Royal Statistical Society, Series B* **39**, 1–38.

3. Dunnett C. W. (1980). Pairwise multiple comparisons in the unequal variance case. *Journal of the American Statistical Association* **75**, 796–800.

4. Everitt, B. S. and Hand, D.J. (1981). *Finite mixture distribution.* London: Chapman and Hill.

5. Hasselblad, V. (1966). Estimation of parameters for a mixture of normal distributions. *Technometrics* **8**, 431–446.

6. Huang, M.-N. L., Lin C. S. and Soong K. (2004). Factor effects testing for mixture distributions with application to the study of emergence of Pontomyia Oceana. *Journal of Data Science* **2**, 213–230.

7. McLachlan, G. J. and Jones, P. N. (1988). Fitting mixture models to grouped and truncated data via the EM algorithm. *Biometrics* **44**, 571–578.

8. McLachlan, G. J. and Krishnan, T.(1997). *The EM algorithm and extensions.* New York: Wiley.

9. McLachlan, G. J. and Peel, D.(2000). *Finite mixture models.* New York: Wiley.

10. McLaren, C. E., Brittenham, G. M., and Hasselblad, V. (1986). Analysis of the volume of red blood cells: Application of the expectation-maximization algorithm to grouped data from the doubly-truncated log-normal distribution. *Biometrics* **42**, 143–158.

11. Montgomery D. C. (1997) *Design and analysis of experiments.* New York: Wiley.

12. Rice, J. A. (1995). *Mathematical statistics and data analysis.* New York: Wiley.

13. Slakter, M. J. (1965). A comparison of the Pearson chi-square and Kolmogorov goodness-of-fit tests with respect to validity. *Journal of the American Statistical Association* **60**, 854–858.

14. Tamhane, A. C. (1977). Multiple comparison in model I one-way anova with unequal variances. *Communications in Statistics* **A6**, 15–32.

15. Tamhane, A. C. (1979). A comparison of procedure for multiple comparisons of means with unequal variances. *Journal of the American Statistical Association* **74**, 471–480.

16. Titterington, D. M., Smith, A. F. M., and Makov, U. E. (1985). *Statistical analysis of finite mixture distributions.* New York: Wiley.

17. Yong S. Q. and Bruce S. (2004). Likelihood ratio test for homogeneity in normal mixtures in the presence of a structural parameter. *Statistica Sinica* **14**, 1165–1177.

18. Zhang L. , Yang M. C. K. , Wang X. , Larkings B. A. , Gallo-Meagher M. , and Wu R. (2004). A model for estimating joint maternal-offspring effects on seed development in autogamous plants. *Physiol Genomics* **19**, 262–269.

Appendix

(A) EM algorithm for grouped and truncated data.

We consider fitting of finite normal mixture models via the EM algorithm with grouped and truncated data. Suppose the sample space of X, Ω, is partitioned into $\nu = r+t$ mutually exclusive subsets Ω_j $(j = 1, \ldots, \nu)$. Independent observations are made on X, but only the number n_j falling in Ω_j $(j = 1, \ldots, r)$ is recorded. That is, in addition to the individual observations, the number of observations n_j falling in Ω_j $(j = r+1, \ldots, \nu)$ are not available for the subsequent estimation of Φ.

For given $n = \sum_{j=1}^{r} n_j$, it is assumed that observed data $y = (n_1, \ldots, n_r)$ has a multinomial distribution consisting of n draws on r categories with probabilities $P_j(\Phi)/P(\Phi)$,$j = 1, \ldots, r$, where

$$P_j(\Phi) = \int_{\Omega_j} f(x; \Phi) dx$$

and

$$P(\Phi) = \sum_{j=1}^{r} P_j(\Phi).$$

This gives the likelihood function

$$L(\Phi; \mathbf{y}) = \left(\frac{n!}{\prod_{j=1}^{r} n_j!} \right) \prod_{j=1}^{r} \left[\frac{P_j(\Phi)}{P(\Phi)} \right]^{n_j}. \tag{A.1}$$

We can undertake solving the likelihood equation $\partial L(\Phi;\ \mathbf{y})/\partial\Phi$=0 within the EM framework by introducing the random variables $\mathbf{u} = (n_{r+1}, \ldots, n_\nu)'$ and

$$\mathbf{x}_j = (x_{j1}, \ldots, x_{jn_j})' \quad (j = 1, \ldots, \nu),$$

where, conditional on $\mathbf{y}$, $\mathbf{u}$ has probability function

$$\left[\frac{(\sum_{j=r+1}^{\nu} n_j + n = 1)!}{(n-1)!} \right] [P(\Phi)]^n \prod_{j=r+1}^{\nu} \frac{[P_j(\Phi)]^{n_j}}{n_j!}.$$

Conditional on $\mathbf{y}$ and $\mathbf{u}$, the x_{jk} $(k = 1, \ldots, n_j)$ denote n_j independent observations with density $f(x, \Phi)/P_j(\Phi)$ for $j = 1, \ldots, \nu$. The EM machinery is then invoked by declaring

$$\mathbf{w} = (\mathbf{y}', \mathbf{x}_1', \ldots, \mathbf{x}_\nu', \mathbf{u}')' \tag{A.2}$$

as the complete data vector. The log-likelihood for this complete data specification is equal to

$$\sum_{j=1}^{\nu} \sum_{k=1}^{n_j} \log f(x_{jk}; \Phi), \tag{A.3}$$

and it implies the likelihood (A.1) for the incomplete data $\mathbf{y}$.

If we work with the complete data vector $\mathbf{w}$ specified by (A.2) and $f(x;\Phi)$ is a mixture density, then the M-step of the EM algorithm will require itself an iterative procedure. Consequently, for mixture densities we propose a further extension of the complete data vector w to include the zero-one indicator variables in $\mathbf{z}_{ik} = (z_{1jk},\ldots,z_{gjk})'$ $(j=1,\ldots,\nu; k=1,\ldots,n_j)$, where $\sum_{i=1}^{g} z_{ijk} = 1$ and where, given the x_{jk}, the z_{jk} are conditionally independent with

$$\Pr\{z_{ijk}=1|x_{jk}\} = \frac{\pi_i f_i(x_{jk};\theta)}{f(x_{jk};\Phi)} = \tau_i(x_{jk};\Phi), \text{ say,}$$

for $i=1,\ldots,g$. Note that $\tau_i(x_{jk};\Phi)$ is the probability that x_{jk} belongs to the ith component given its value.

With the inclusion of these indicator variables z_{ijk} in the complete data specification, the log-likelihood (A.3) becomes

$$\sum_{i=1}^{g}\sum_{j=1}^{\nu}\sum_{k=1}^{n_j} z_{ijk}[\log f_i(x_{jk};\theta+\log\pi_i)]. \tag{A.4}$$

E-step

The E-step of the EM algorithm at $(p+1)$th stage requires taking its expectation conditional on the observed data $\mathbf{y}$, using the current fit $\Phi^{(p)}$ for Φ.

$$\begin{aligned}
& Q(\Phi;\Phi^{(p)}) \\
= \;& E(\sum_{i=1}^{g}\sum_{j=1}^{\nu}\sum_{k=1}^{n_j} z_{ijk}[\log f_i(x_{jk};\theta)+\log\pi_i]|\mathbf{y}=(n_1,\ldots,n_r)') \\
= \;& \sum_{i=1}^{g}\sum_{j=1}^{\nu} E(\sum_{k=1}^{n_j} z_{ijk}[\log f_i(x_{jk};\theta)+\log\pi_i]|\mathbf{y}=(n_1,\ldots,n_r)') \\
= \;& \sum_{i=1}^{g}\sum_{j=1}^{\nu} E(n_j|\mathbf{y}=(n_1,\ldots,n_r)')E(z_{ijk}[\log f_i(x_{jk};\theta)+\log\pi_i]|\mathbf{y}=(n_1,\ldots,n_r)') \\
= \;& \sum_{i=1}^{g}\sum_{j=1}^{\nu} m_j(\Phi^{(p)})E_j^{(p)}\{\tau_i(X;\Phi^{(p)})[\log f_i(X;\theta)+\log\pi_i]\},
\end{aligned}$$

where $E_j^{(p)}$ refers to expectation with respect to the density $f(x,\Phi^{(p)})/P_j(\Phi^{(p)})$ and

$$m_j(\Phi^{(p)}) = \begin{cases} n_j & ,j=1,\ldots,r; \\ n\dfrac{P_j\Phi^{(p)}}{P(\Phi^{(p)})} & ,j=r+1,\ldots,\nu. \end{cases}$$

M-step for Normal mixtures

At the $(p+1)$th stage of the M-step, the intent is to maximize $Q(\Phi;\Phi^{(p)})$ with respect to Φ, to produce a new estimate of Φ, $\Phi^{(p+1)}$. We consider this now, first for mixtures with normal component densities having mean μ_i and variance σ_i^2 $(i=1,\ldots,g)$.

The estimates of π_i, μ_i, and $\sigma_i^2 (i=1,\ldots,g)$ so obtained at the $(p+1)$th stage of the M-step satisfy

$$\begin{aligned}\frac{\partial Q(\Phi;\Phi^{(p)})}{\partial \mu_i} &= \sum_{j=1}^{\nu} m_j(\Phi^{(p)}) E_j^{(p)}\{\tau_i(X;\Phi^{(p)})[\frac{\partial}{\partial \mu_i}\log f_i(X;\theta)+\log\pi_i)]\\ &= \sum_{j=1}^{\nu} m_j(\Phi^{(p)}) E_j^{(p)}\{\tau_i(X;\Phi^{(p)})(X-\mu_i)\}\end{aligned}$$

$$\begin{aligned}\frac{\partial Q(\Phi;\Phi^{(p)})}{\partial \sigma_i^2} &= \sum_{j=1}^{\nu} m_j(\Phi^{(p)}) E_j^{(p)}\{\tau_i(X;\Phi^{(p)})[\frac{\partial}{\partial \sigma_i^2}\log f_i(X;\theta)+\log\pi_i)]\\ &= \sum_{j=1}^{\nu} m_j(\Phi^{(p)}) E_j^{(p)}\{\tau_i(X;\Phi^{(p)})[\sigma_i^{-2}(X-\mu_i)^2]\}\end{aligned}$$

$$\begin{aligned}\frac{\partial Q(\Phi;\Phi^{(p)})}{\partial \pi_i} &= \frac{\partial}{\partial \pi_i}\sum_{j=1}^{\nu} m_j(\Phi^{(p)}) E_j^{(p)}\{\tau_i(X;\Phi^{(p)})[\log f_i(X;\theta)+\log\pi_i)]\\ &= \frac{\sum_{j=1}^{\nu} m_j(\Phi^{(p)}) E_j^{(p)}\{\tau_i(X;\Phi^{(p)})\}}{\pi_i} - \sum_{j=1}^{\nu} m_j(\Phi^{(p)})\end{aligned}$$

Therefore

$$\begin{aligned}\hat{\mu}_i^{(p+1)} &= \frac{\sum_{j=1}^{\nu} m_j(\Phi^{(p)}) E_j^{(p)}\{\tau_i(X;\Phi^{(p)})X\}}{c_i(\Phi^{(p)})},\\ [\hat{\sigma}_i^{(p+1)}]^2 &= \frac{\sum_{j=1}^{\nu} m_j(\Phi^{(p)}) E_j^{(p)}\{\tau_i(X;\Phi^{(p)})(X-\mu_i^{(p+1)})^2\}}{c_i(\Phi^{(p)})},\\ \hat{\pi}_i^{(p+1)} &= \frac{\sum_{j=1}^{\nu} m_j(\Phi^{(p)}) E_j^{(p)}\{\tau_i(X;\Phi^{(p)})\}}{\sum_{j=1}^{\nu} m_j(\Phi^{(p)})},\end{aligned}$$

where

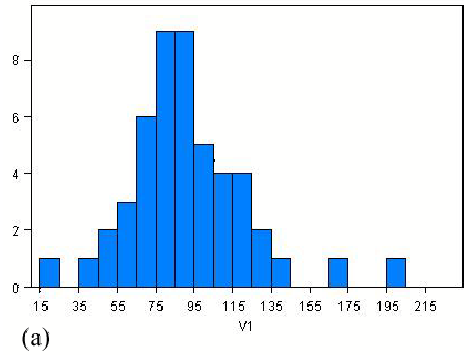
$$c_i(\Phi^{(p)}) = \sum_{j=1}^{\nu} m_j(\Phi^{(p)}) E_j^{(p)}\{\tau_i(X;\Phi^{(p)})\}.$$

(B)

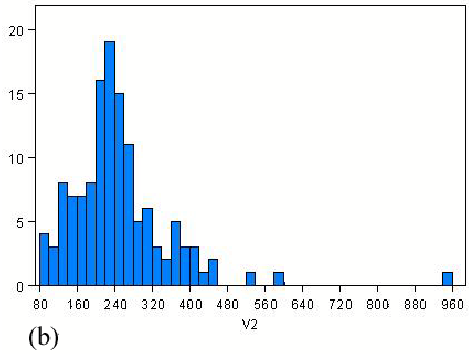
Table A1 : Results of fitting poisson regression model using S-PLUS software.

* * * Generalized Linear Model * * *				
Call : glm (formula = Bloodt ∼ BloodL + Age, family= poisson (link = log), data = SDF100, na.action = na.exclude, control= list (epsilon = 0.0001, maxit=50, trace=T))				
Deviance Residuals:				
Min	1Q	Median	3Q	Max
-2.338	-0.433	-0.020	0.318	2.576
Coefficients:				
	Value	Std. Error	t value	
(Intercept)	-0.442854279	0.2637894141	-1.678817	
BloodL	0.001005496	0.0001362925	7.377483	
Age	0.012179204	0.0046521065	2.617998	
(Dispersion Parameter for Poisson family taken to be 1)				
Null Deviance: 154.7614 on 107 degrees of freedom				
Residual Deviance: 87.28839 on 105 degrees of freedom				
Number of Fisher Scoring Iterations: 4				

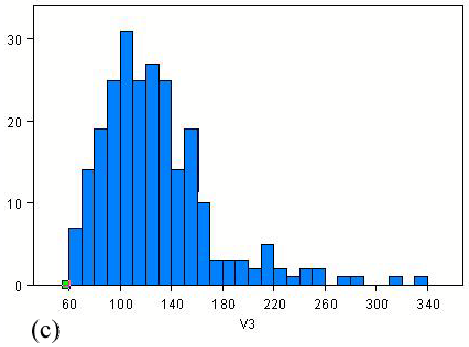
(C) Histograms for the thirteen operation times



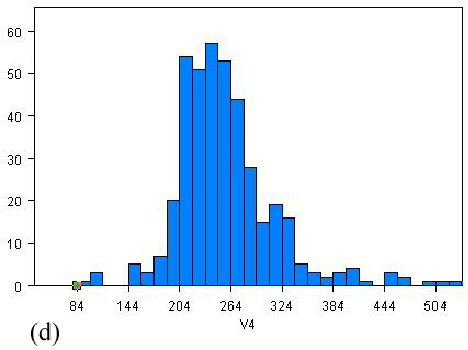
(a)



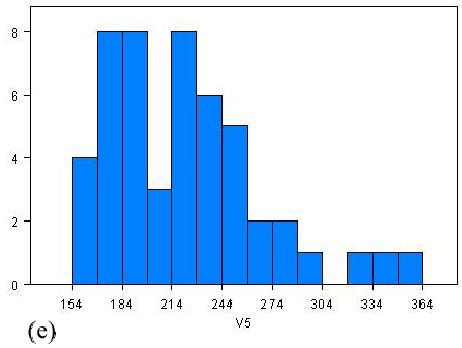
(b)



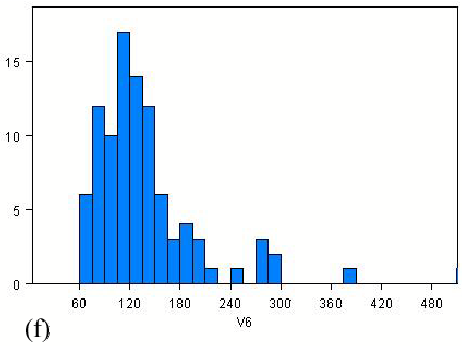
(c)



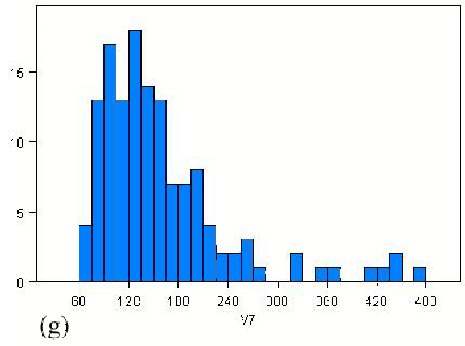
(d)



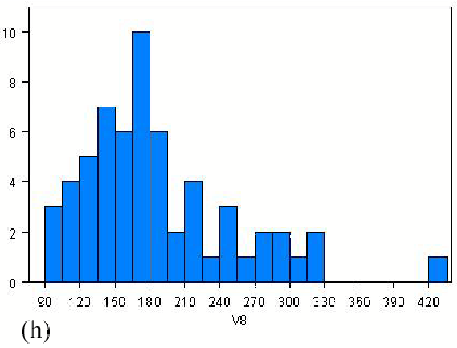
(e)



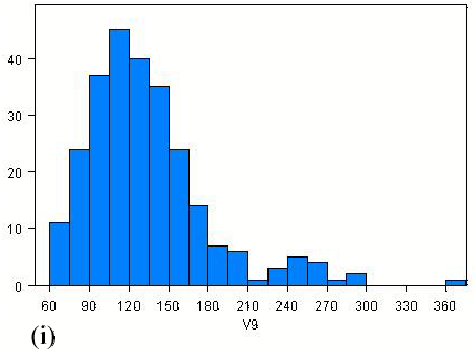
(f)



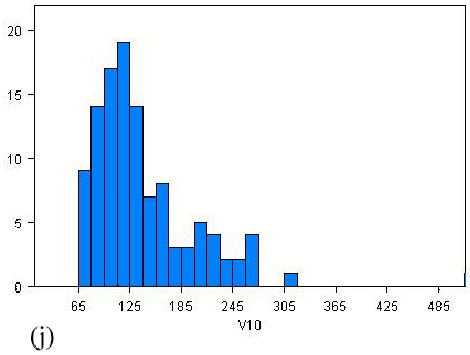
(g)



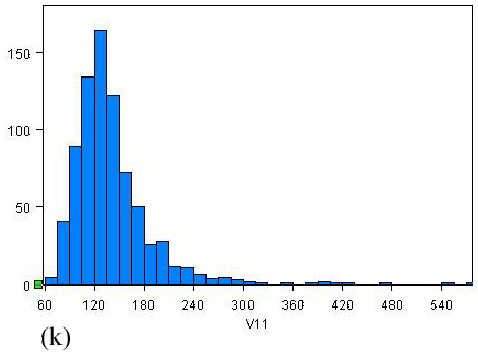
(h)



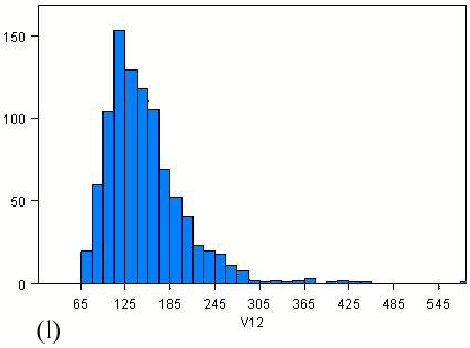
(i)



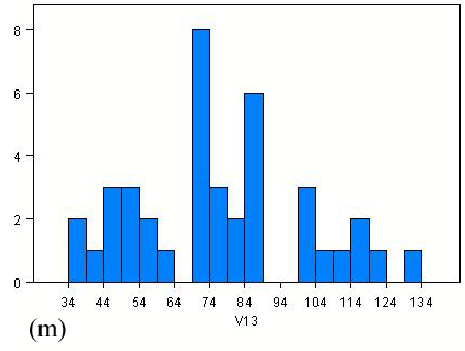
(j)



(k)



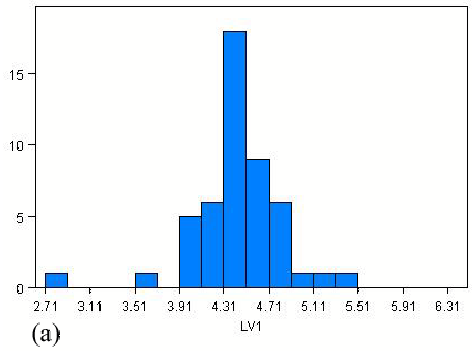
(l)



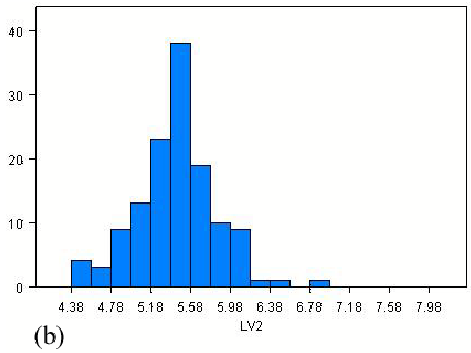
(m)

Figures A1: (a)-(m) are histograms for the thirteen operation time in GYN department.

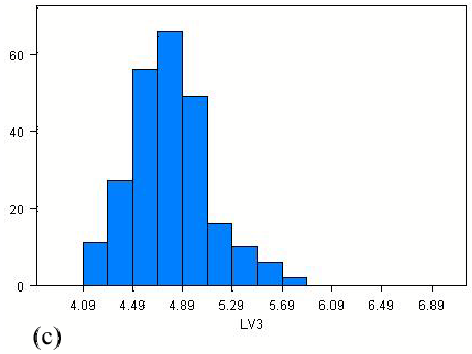
(D) Histograms for the logarithm of the thirteen operation times



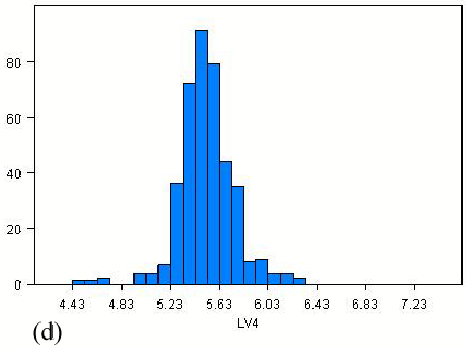
(a)



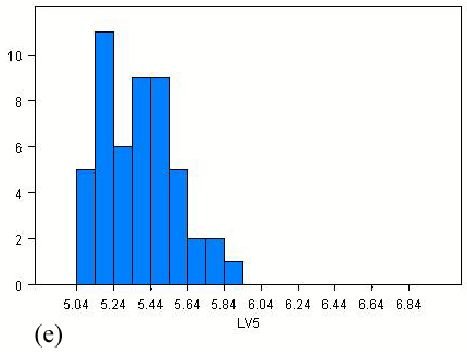
(b)



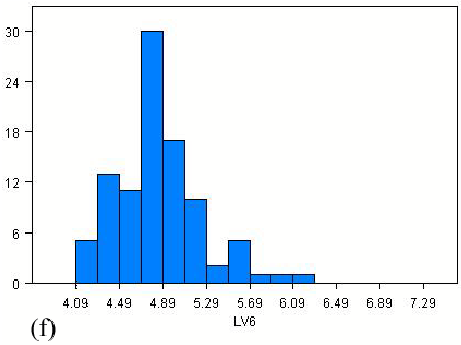
(c)



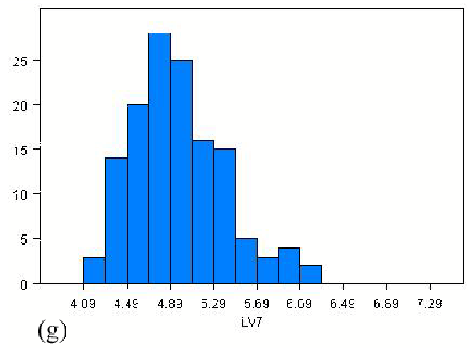
(d)



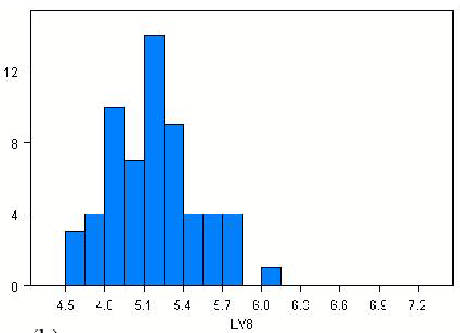
(e)



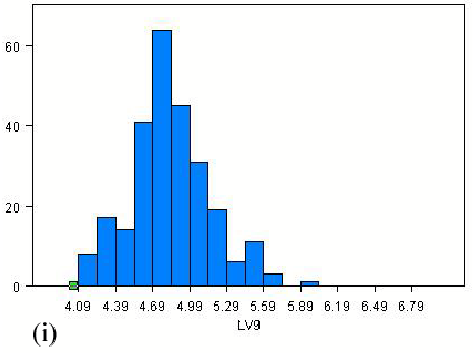
(f)



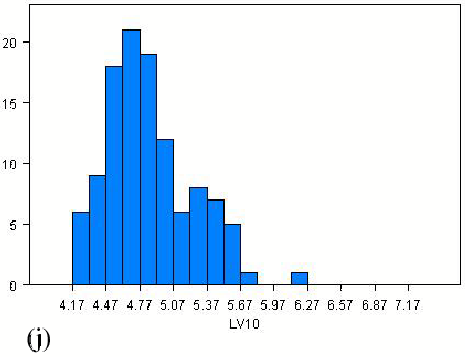
(g)



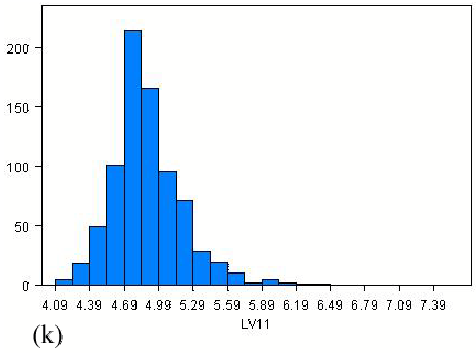
(h)



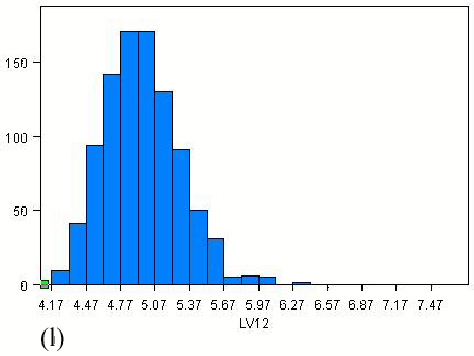
(i)



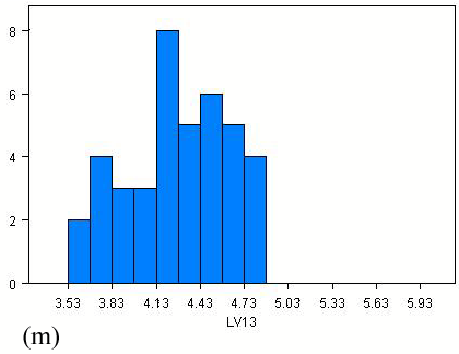
(j)



(k)



(l)



(m)

Figures A2: (a)-(m) are histograms for the logarithm of the thirteen operation time in GYN department.

(E) EM algorithm for grouped and truncated data using Mathematica 5 software

[1] *“a” is list of number from shortest operation time to longest opsration time.*

```
<<Statistics`ContinuousDistributions`
a=Table[i,{i,60,540}];
b=Append[a, ];
x=Prepend[b,0];
```

[2] *Three component mixture of log-normal distribution.*

```
f[x_, u_, s_] := PDF[NormalDistribution[u, s], x]
ff[x_, pi1_, pi2_, u1_, s1_, u2_, s2_, u3_, s3_] :=
  pi1 * PDF[NormalDistribution[u1, s1], x] +
   pi2 * PDF[NormalDistribution[u2, s2], x] +
   (1 - pi1 - pi2) * PDF[NormalDistribution[u3, s3], x];
P[j_, pi1_, pi2_, u1_, s1_, u2_, s2_, u3_, s3_] :=
  ∫_Log[x[[j]]]^Log[x[[j+1]]] ff[y, pi1, pi2, u1, s1, u2, s2, u3, s3] ⅆy;
```

[3] *Initial values*

```
pi1=0.2625078816031413`;
pi2=0.009303529391563604`;
u1=5.654585814499704`;
u2=4.610202770196282`;
u3=5.491969502758988`;
s1=0.27985722968656435`;
s2=0.11113668260933902`;
s3=0.1480334356538293`;pi3=1-pi1-pi2;
```

[4] *data : each one of the interval of operation time is one minute.*

```
Do[m={403*P[1,pi1,pi2,u1,s1,u2,s2,u3,s3]/(1-P[1,pi1,pi2,u1,s1,u2,s2,u3,
s3]-P[482,pi1,pi2,u1,s1,u2,s2,u3,s3]),data,
403*P[482,pi1,pi2,u1,s1,u2,s2,u3,s3]/(1-P[1,pi1,pi2,u1,s1,u2,s2,u3,s3]-
P[482,pi1,pi2,u1,s1,u2,s2,u3,s3])};
```

```
aa = Table[{P[j, pi1, pi2, u1, s1, u2, s2, u3, s3],
    {x[[j]], x[[j + 1]]}, m[[j]]}, {j, 482}];
pp = {}; xx = {}; mm = {};
Do[If[aa[[h, 1]] ≠ 0, AppendTo[pp, aa[[h, 1]]];
   AppendTo[mm, aa[[h, 3]]]; AppendTo[xx, aa[[h, 2]]]],
  {h, Length[aa]}];
EX1[j_] := 1/pp[[j]] ∫_Log[xx[[j,1]]]^Log[xx[[j,2]]] pi1 * f[y, u1, s1] dy;
EX2[j_] := 1/pp[[j]] ∫_Log[xx[[j,1]]]^Log[xx[[j,2]]] pi2 * f[y, u2, s2] dy;
EX3[j_] :=
   1/pp[[j]] ∫_Log[xx[[j,1]]]^Log[xx[[j,2]]] (1 - pi1 - pi2) * f[y, u3, s3] dy;

c1 = Sum_k=1^Length[pp] mm[[k]] * EX1[k];
c2 = Sum_k=1^Length[pp] mm[[k]] * EX2[k];
c3 = Sum_k=1^Length[pp] mm[[k]] * EX3[k];

pi11 = (Sum_k=1^Length[pp] mm[[k]] * EX1[k]) / (Sum_k=1^Length[pp] mm[[k]]);
pi22 = (Sum_k=1^Length[pp] mm[[k]] * EX2[k]) / (Sum_k=1^Length[pp] mm[[k]]);
pi33 = (Sum_k=1^Length[pp] mm[[k]] * EX3[k]) / (Sum_k=1^Length[pp] mm[[k]]);

u11 =
   1/c1
    (Sum_k=1^Length[pp] mm[[k]] * 1/pp[[k]]
       ∫_Log[xx[[k,1]]]^Log[xx[[k,2]]] y * pi1 * f[y, u1, s1] dy);

u22 =
   1/c2
    (Sum_k=1^Length[pp] mm[[k]] * 1/pp[[k]]
       ∫_Log[xx[[k,1]]]^Log[xx[[k,2]]] y * pi2 * f[y, u2, s2] dy);
```

$$u33 = \frac{1}{c3}\left(\sum_{k=1}^{Length[pp]} mm[[k]] * \frac{1}{pp[[k]]} \int_{Log[xx[[k,1]]]}^{Log[xx[[k,2]]]} y * (1 - pi1 - pi2) * f[y, u3, s3]\, dy\right);$$

$$s11 = \frac{1}{c1}\left(\sum_{k=1}^{Length[pp]} mm[[k]] * \frac{1}{pp[[k]]} \int_{Log[xx[[k,1]]]}^{Log[xx[[k,2]]]} (y - u1)^2 * pi1 * f[y, u1, s1]\, dy\right);$$

$$s22 = \frac{1}{c2}\left(\sum_{k=1}^{Length[pp]} mm[[k]] * \frac{1}{pp[[k]]} \int_{Log[xx[[k,1]]]}^{Log[xx[[k,2]]]} (y - u2)^2 * pi2 * f[y, u2, s2]\, dy\right);$$

$$s33 = \frac{1}{c3}\left(\sum_{k=1}^{Length[pp]} mm[[k]] * \frac{1}{pp[[k]]} \int_{Log[xx[[k,1]]]}^{Log[xx[[k,2]]]} (y - u3)^2 * pi3 * f[y, u3, s3]\, dy\right);$$

```
pi1 = pi11;
 pi2 = pi22;
 pi3 = pi33;
 u1 = u11;
 u2 = u22;
 u3 = u33;
 s1 = √s11;
 s2 = √s22; s3 = √s33;, {z, 1}]
```

$$pi1 = pi11$$
$$pi2 = pi22$$
$$pi3 = pi33$$
$$u1 = u11$$
$$u2 = u22$$
$$u3 = u33$$
$$s1 = \sqrt{s11}$$
$$s2 = \sqrt{s22}$$
$$s3 = \sqrt{s33}$$