Multivariate Analysis of Waiting and Treatment Time in Emergency Departments

Abera Wouhib, Myron J. Katzoff, PhD, and Meena Khare, National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention, 3311 Toledo Road, Room 3113, Hyattsville, MD 20782

Abstract

Checking on normality is desirable in multivariate data analyses, if statistical inferences are made under the assumption of normal distribution. Normal probability plots or Q-Q plots provide a good visual check and are considered to be adequate for the purpose of assessing normality by many researchers.[1] Transforming the data to make it 'normal' is an appropriate procedure if departure from normality is suspected in the plots. Once the problem of normality is resolved by means of transformation, the desired statistical inference can proceed in a fairly standard fashion. To illustrate data transformations and to apply multivariate statistical inferences, waiting and treatment times for emergency department visits along with some socio-demographical characteristics were used from the National Hospital Ambulatory Medical Care Survey (NHAMC).

[*] **Key Words**: **Multivariate Analysis: Waiting time: Treatment time**

1. Introduction:

Many statistical analyses are based on the assumption that the data were generated from a normal distribution. When that assumption holds, the problems of parameter estimation are mathematically tractable and 'suitable' results can be obtained. However, real data are never exactly normally distributed and the normal density is often only a useful approximation to the true population distribution. Moreover, for large sample sizes the normal distribution serves as a very reasonable population model in some instances. In particular, the sampling distributions of estimators of many survey population parameters are approximately normal, regardless of the form of the parent population, due to a so-called 'central limit effect'. Furthermore, since many real-world problems fall naturally within the framework of normal theory, the normal distribution plays an essential role in many statistical inference applications. The inference problem needs serious attention and a different approach has to be crafted when the above assumptions do not hold. The main purpose of this paper is to explore an alternative when sample survey data do not meet the required assumptions for appropriate statistical inferences of interest, particularly with respect to multivariate normal statistical analysis, and to fix the problems when the assumptions are not met. When normality is not a viable assumption, an alternative is to make non-normal data more 'normal appearing' by transforming it. This paper deals with the latter when performing multivariate analyses. Once the problem of normality is resolved satisfactorily, the desired statistical inference can proceed in a fairly standard fashion: considering a random sample of size n_1 from one population and a random sample of size n_2 from a second population , p-variate observations are used in testing the equalities of the mean vectors of p-variables under a null hypothesis of no difference between the group mean vectors against the alternative hypothesis of a difference in the group mean vectors. When a statistically significant difference is identified by the multivariate testing procedure, an accepted statistical procedure is to examine hypotheses involving parameters of the original mean vectors for the purpose of identifying as might be expected, where differences occur. This is accompanied by conducting tests with subvectors of the data on waiting and treatment time means since p=2. We demonstrate the full range of statistical testing procedures with an example using data from the National Hospital Ambulatory Medical Care Survey (NHAMCS).[**]

[*] "The findings and conclusions in this paper are those of the authors and do not necessarily represent the views of the National Center for Health Statistics, Centers for Disease Control and Prevention."

[**] "The results in this paper are solely for illustrative purposes of the statistical methods and no conclusion should be made based on the outcome."

The National Hospital Ambulatory Medical Care Survey, a national probability sample survey of nonfederal general and short-stay hospitals sponsored by the National Center for Health Statistics (NCHS) of the Centers for Disease Control and Prevention has been collecting data on sample patients from emergency departments (EDs) since 1992 [2] To illustrate data transformation steps before analyzing the data and to apply multivariate statistical inferences accordingly, we used waiting and treatment time data along with some relevant socio-demographical characteristics of the ED patients. An appropriate data transformation method was applied to both waiting and treatment times, and then the multivariate statistical analyses were carried out to test the difference between the mean vectors of population groups for some important socio-demographic categories.

2. Methods

a) Data Transformation

Transformations are re-expressions of the data in different units. Suitable transformations are usually suggested by the nature of the data themselves. If the data are multidimensional and cannot be approximated by a normal distribution, the data may suggest the transformation needed to get around the problem of non-normality. Power transformations in the form of $y = x^{\lambda}$ are one useful family of transformations for the purpose of producing approximate normality and let only the appearance of the data themselves influences the choice of appropriate λ. Power transformations are also defined only for positive variables.

Box and Cox defined a slightly modified family of power transformations to achieve multivariate normality with dimension p >1 as follow: [3]

$$y_k^{\lambda_k} = \begin{cases} \dfrac{x_k^{\lambda_k} - 1}{\lambda_k} & \lambda_k \neq 0 \\ \ln \quad x_k & \lambda_k = 0 \end{cases} \qquad for \quad k = 1, 2, \ldots p$$

This transformation is continuous in λ_k for $x_k > 0$; here, x_k is the kth coordinate of the data vector. The $\lambda_1, \lambda_2, \ldots, \lambda_p$ are parameters in the power transformations for the p measured characteristics. The Box-Cox solution for the choice of an appropriate power λ_k is selected by maximizing

$\ell_k(\lambda) = -\dfrac{n}{2} \ln\left[\dfrac{1}{n}\sum_{i=1}^{n}\left(y_{ik}^{\lambda_k} - \overline{y_k^{\lambda_k}}\right)^2\right] + (\lambda_k - 1)\sum_{i=1}^{n} \ln \quad x_{ik}$, where $x_{1k}, x_{2k}, \ldots, x_{nk}$ are the n observation of the k^{th} variable, k = 1,2,…,p and $\overline{y_k^{\lambda_k}} = \dfrac{1}{n}\sum_{i=1}^{n} y_{ik}^{\lambda_k} = \dfrac{1}{n}\sum_{i=1}^{n}\left(\dfrac{x_{ik}^{\lambda_k} - 1}{\lambda_k}\right)$ is the arithmetic average of the transformed observations.

The transformation procedure is equivalent to making each marginal distribution approximately normal. In theory although normal marginal distributions are not sufficient to ensure that the joint distribution is normal, in practical applications maximizing $\ell_k(\lambda)$ at each marginal should be sufficient and the results are not usually different from maximizing p equations simultaneously. Generally, maximizing $\ell_k(\lambda)$ for p-equations simultaneously is mathematically cumbersome: however, when p is small, it could be manageable. The optimal choices of λ will not guarantee that the transformed set of values adequately conform to a multivariate normal distribution. But, the outcome produced by the transformation should always be examined for possible violation of other underlying assumptions.

b) Inferences

Once the transformed multivariate data are checked for approximate normality, the statistical inferences of interest would proceed. In multivariate statistical analysis, comparing the mean vectors of populations from sample survey data can be developed by analogy with the univariate procedure.

We compare the mean vectors by testing under a null hypothesis, $H_0: \mu_1 = \mu_2 \Rightarrow \mu_1 - \mu_2 = 0$ versus an alternative, $H_1: \mu_1 \neq \mu_2 \Rightarrow \mu_1 - \mu_2 \neq 0$ where μ_1 is the mean vector of population 1 and μ_2 is the mean vector of population 2. Consider a random sample of size n_1 from population 1 and a sample of size n_2 from population 2 of p-variables multivariate data. For large sample size survey data, we can run the test by making a few additional assumptions.

- The p-variate samples of size n_1 from population 1 are distributed independently and normally with mean vector μ_1 and p x p variance-covariance matrix Σ_1.

- The p-variate samples of size n_2 from population 2 are distributed independently and normally with mean vector μ_2 and a p x p variance-covariance matrix Σ_2.

These assumptions impact the problem of making an inference about the p x 1 vector $\mu_1 - \mu_2$ under one of the following two cases.

Case 1: $\Sigma_1 \neq \Sigma_2$ (the variance-covariance matrices of the two population groups are unequal)

When $\Sigma_1 \neq \Sigma_2$, we are unable to find a measure whose distribution does not depend on the unknowns Σ_1 and Σ_2. However, when n_1 and n_2 are large, we can reduce the complexities due to unequal variance-covariance matrices. If $\overline{Y_1}$ and $\overline{Y_2}$ are the p-variate sample mean vectors for populations 1 and 2, then for large n_1 and n_2, $\overline{Y_1} - \overline{Y_2}$ is distributed nearly normal, $N_p[\mu_1 - \mu_2,\ n_1^{-1}\Sigma_1 + n_2^{-1}\Sigma_2]$. If Σ_1 and Σ_2 were known, the square of the statistical distance from $\overline{Y_1} - \overline{Y_2}$ to $\mu_1 - \mu_2$ is: $T^{2*} = \left[\left(\overline{Y_1} - \overline{Y_2}\right) - \left(\mu_1 - \mu_2\right)\right]^T \left(n_1^{-1}\Sigma_1 + n_2^{-1}\Sigma_2\right)^{-1} \left[\left(\overline{Y_1} - \overline{Y_2}\right) - \left(\mu_1 - \mu_2\right)\right]$, where $[\ \ ]^T$ is the transpose of the matrix. Then, under H_0, T^{2*} is approximately distributed as χ^2_p. In practice, the variance-covariance matrices Σ_1 and Σ_2 are unknown and they are estimated by corresponding p x p sample variance-covariance matrices (Wishart matrices), S_1 and S_2. The p x p sample variance-covariance matrix from population 1 is

$$S_1 = \begin{pmatrix} s_{11} & \dots & s_{1k} & \dots & s_{1p} \\ s_{i1} & \dots & s_{ik} & \dots & s_{ip} \\ s_{p1} & \dots & s_{pk} & \dots & s_{pp} \end{pmatrix},\ \text{where} \left(n_1 - 1\right) s_{ik} = \sum_{j=1}^{n_1} \left(y_{ji} - \overline{y_i}\right)\left(y_{jk} - \overline{y_k}\right) \text{for } \bar{Y}_1 = \left(\bar{y}_1, \bar{y}_2, \dots \bar{y}_p\right).$$

The S_2 can be computed analogously using sample from population 2. The decision rule for the multivariate test procedure is to reject H_0, if $T^{2*} > \chi^2_p$.

When variance-covariance matrices of two population groups may not be equal and the assumption of equal variance-covariance matrix is unjustifiable, then it is called a Behrens-Fisher multivariate problem. Under this circumstance, statistical inferences based on the square distance, T^{2*} work better when the sample sizes from the two population groups are large. Nevertheless, T^{2*} is distributed approximately χ^2_p only if we assume that the sample variance-covariance matrices are equal to their corresponding population values. Therefore, care must be taken when inferences based on the T^{2*} are made. More practical experience is needed with this test before recommending it unconditionally. [3]

Case 2: [3] $\Sigma_1 = \Sigma_2 = \Sigma$ (the variance-covariance matrices of the two population groups are equal)

This assumption can be generalized to compare mean vectors of g populations. Under this assumption, we can compare mean vectors by testing the null hypothesis: $H_0 : \mu_1 = \mu_2 = ... = \mu_g$.

The model $Y_{ij} = \mu_i + e_{ij}$ or equivalently as $Y_{ij} = \mu + \tau_i + e_{ij}$, where Y_{ij}, μ, τ_i and e_{ij} are all p-variate vectors. Equivalently, we can re-write the above model for all population groups g in terms of a regression model as: $Y = XB + \varepsilon$, with the null hypothesis: $H_0 : \beta_1 = \beta_2 = ... = \beta_g$.
All Y, X and ε are in multidimensional p-variable matrices and X is a matrix of dummy variables that depends on the population group in the model. B is the matrix of the population means, and ε is the matrix of the model errors. Note that, $E\left(\varepsilon^T \varepsilon\right) = \Sigma$ where E is the expectation operator.

For $\bar{Y}$ the grand sample mean and $\bar{Y}_1, \bar{Y}_2, ..., \bar{Y}_g$ the corresponding sample mean vector of population 1, population 2, up to population g, then,

$$\sum_{i=1}^{g} \sum_{j=1}^{n_i} \left(Y_{ij} - \bar{Y}\right)\left(Y_{ij} - \bar{Y}\right)^T = \sum_{i=1}^{g} n_i \left(\bar{Y}_i - \bar{Y}\right)\left(\bar{Y}_i - \bar{Y}\right)^T + \sum_{i=1}^{g} \sum_{j=}^{n_i} \left(Y_{ij} - \bar{Y}_i\right)\left(Y_{ij} - \bar{Y}_i\right)^T$$

(Corrected total Sum of squares) = (between sum of squares) + (within sum of squares),

$$B = \sum_{i=1}^{g} n_i \left(\bar{Y}_i - \bar{Y}\right)\left(\bar{Y}_i - \bar{Y}\right)^T \quad and \quad W = \sum_{i=1}^{g} \sum_{j=}^{n_i} \left(Y_{ij} - \bar{Y}_i\right)\left(Y_{ij} - \bar{Y}_i\right)^T .$$

One test of $H_0 : \mu_1 = \mu_2 = ... = \mu_g$ involves B and W variances. We reject H_0 if the ratio $\Lambda^* = \frac{|W|}{|B+W|}$ $> F_{2(g-1),2(n-g-1)}$, where Λ^* is Wilk's Lambda statistics, $n = \sum_{i=1}^{g} n_i$, if p =2 (the degree of freedom for F distribution varies with p and g), and | * | is determinant of the matrix.

The Λ^* is equivalent to the likelihood ratio test. Besides the Wilk's Lambda statistics, the other multivariate tests, the Lawley-Hotelling trace $= tr\left(BW^{-1}\right)$, Pillai trace= $tr[B\left(B+W\right)^{-1}]$, and Roy's largest root = maximum eigenvalue of $W\left(B+W\right)^{-1}$ are nearly equivalent for extremely large samples. Also these tests are available in most statistical packages and it is convenient to use them once the assumption of $\Sigma_1 = \Sigma_2 = ... = \Sigma_g = \Sigma$ for g population groups has been established.

3. Data and Application:

The National Hospital Ambulatory Medical Care Survey (NHAMCS) has been collecting data on waiting time and treatment time in emergency departments (EDs) of nonfederal general and short-stay hospitals selected in a probability sample survey conducted annually. Data from the ED sample visits in 2004 and 2005 on waiting and treatment time along with some socio-demographic characteristics were used in the analysis of this paper. Here, waiting time is defined as time from arrival at an ED up to seeing health professionals and treatment time is time under the care of health professionals until discharged from the ED. The NHAMCS ED sample has a total of 33,605 records from 2005 and 36,589 from 2004. Records with complete information on waiting and length of visit were considered along with the categorical variables of interest in the analysis. Treatment time was computed by subtracting waiting time from length of visit. A total of 52,142 complete records of 25,668 from 2005 and 26,474 from 2004 NHAMCS ED were used in these analyses. [2]

a) Transformation:

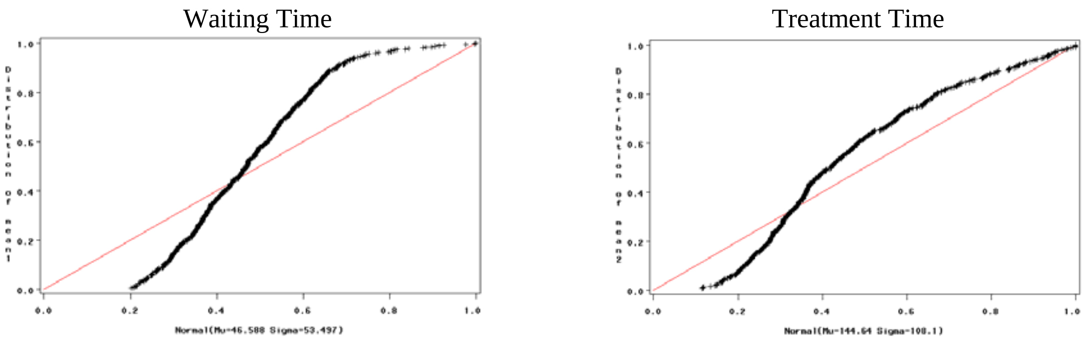
The time that patients spend in emergency departments (EDs) is divided between the time spent in a queue waiting for service (waiting time) and that receiving service from medical professionals (treatment time). The waiting and treatment times have some level of interdependency (correlation coefficient, $\hat{\rho}$ =0.11) and at the same time have distinctive features as they are used to measure completely different aspects of time in emergency departments, as shown in figure 1. As variables measured in time units, the two variables display non-normality even in large sample sizes. Under this scenario applying data transformation methods on waiting and treatment times and transforming the data is customary before considering any statistical inferences based on the assumed normal distributions. As part of the multivariate statistical analysis, first we explore for normality and as indicated in Figures 1 and 2, the data for both waiting and treatment times were not distributed normally. The Box-Cox power transformation was applied and possible values of λ that maximize $\ell_k(\lambda)$ are shown in Table 1 for waiting time and treatment time, respectively.

Table 1: The Box-Cox Power Transformation of Waiting and Treatment Time for Some Values of λ

λ	$\ell_k(\lambda)$ of waiting time (k=1)	$\ell_k(\lambda)$ of treatment time (k=2)
-1.6	-729373.03	-396865.98
-1.4	-693502.89	-361220.16
-1.2	-657624.87	-328835.76
-1.0	-621735.42	-300844.18
-0.8	-585828.78	-278136.73
-0.6	-549895.25	-260968.10
-0.4	-513917.89	-248991.91
-0.2	-477865.77	-241663.97
0.2	-405242.36	-239474.97 **
0.4	-368310.01	-244301.42
0.6	-330387.55	-252961.55
0.8	-290811.27	-265281.07
1.0	-253950.48	-280950.80
1.2	-250325.52 **	-299554.61
1.4	-273581.93	-320637.61
1.6	-299719.99	-343769.04
1.8	-326955.06	-368576.77
2.0	-355455.41	-394755.66
2.2	-385061.18	-422061.68
2.4	-415540.30	-450301.14
2.6	-446701.42	-479320.10

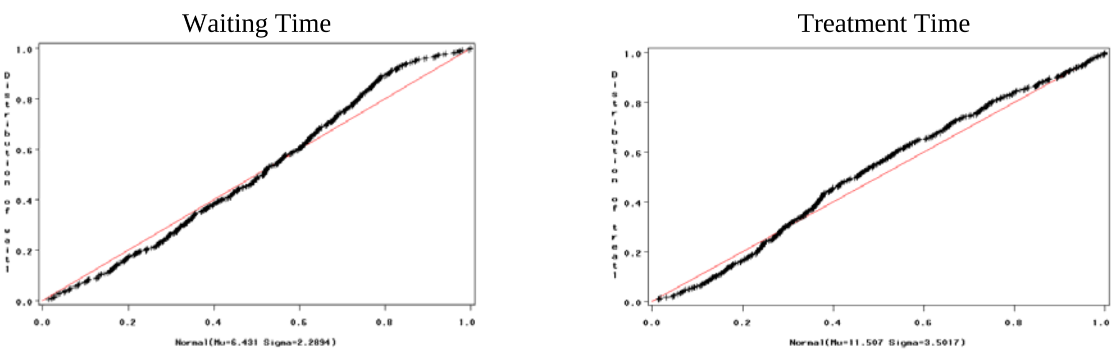
As displayed in Table 1, $\ell_k(\lambda)$ is maximum for $k = 1$ and $k = 2$ at (**), which is at around λ_1 = 1.2 and λ_2 = 0.2 for waiting time and treatment time, respectively. The powers are determined for marginal distributions of waiting time and treatment time. We can also determine simultaneously the pair of powers (λ_1, λ_2) that makes the joint distribution approximately normal, however, the 'best' transformations for the bivariate case did not differ substantially from those obtained by considering each marginal. For simplicity, we chose $\lambda = 0.5$, a value between λ_1 and λ_2, maximized $\ell_1(\lambda)$ and $\ell_2(\lambda)$ for both waiting and treatment time, and approximated the Box-Cox transformation by the square root transformation. It is clear from Figures 1 and 2 that the transformed NHAMCS are much closer to normal in the central portions of the data than the original, which are important for comparing EDs waiting and treatment time mean vectors.

Figure 1: Normal Probability Plots for Original NHAMCS ED Visits



The probability plots for both waiting and treatment times are off from the diagonal line. Under perfect normal distribution scenario, the two lines coincide. Similarly, the Q&Q plots (not shown) also exhibit the original data being far from normal distribution.

Figure 2: Normal Probability Plots of NHAMCS ED Visits after Square-Root Transformation



After applying the square-root transformation, both waiting and treatment time appear to have near normal distribution.

b) Inferences:

In NHAMCS ED visits, the mutually exclusive subcategories of socio-demographic characteristics and geographic locations were taken into account as population groups of the survey. Under the assumption of $\Sigma_1 = \Sigma_2 = ... = \Sigma_g = \Sigma$ (2x2) for g mutually exclusive subcategories of the NHAMCS ED, the model: $Y = XB + \varepsilon \Leftrightarrow Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_g X_g + \varepsilon$, where the Y matrix of transformed waiting and treatment time by population groups $X_1, X_2, ..., X_g$ are dummy variables of population group 1, population group 2, and so on. The β_0 is the two-dimensional mean vector of transformed waiting and treatment time for base population group, and $\beta_1, \ldots, \beta_g$ are the two dimensional vector parameters of the corresponding population group mean vector minus the base population group mean vector.

The NHAMCS Population groups in the model are the following categories and subcategories.

- **Insurance Status**: these include the three broad and mutually exclusive subcategories of insurance status of the ED patient such as government, if payment is covered by Medicare, Medicaid or State Children's Health Insurance (SCHIP), Private insurance, if private insurances are the source of the payment during the visit, and No insurance, which includes self-pay, charity, no charge or no means of payment. [4]
- **Race Ethnicity** (Non-Hispanic Black, Hispanic, and All others including non Hispanic White)
- **Age Group** in Years ($\leq$ 2 (Infant), 2 - 17 (Child), 18 – 64 (Adult), and $\geq$65 (elderly))

- **Geographic** Area (Metropolitan Statistical Area (MSA), and Non-MSA)
- **Gender** (Male, Female)
- **Mode of Arrival** (Ambulance, Non-ambulance)
- **Nature of emergency** (Injury/poison, Other reasons)

The model is estimated by: $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_{GI}\ X_{GI} + \hat{\beta}_{NI}\ X_{NI} + ... + \hat{\beta}_{Ip}\ X_{IP}$.

The $\hat{\beta}_0$ an estimated mean vector of transformed waiting and treatment times of the reference (base) population group, where the group is defined as estimated waiting and treatment time for population group with private insurance, race/ethnicity of all others (excluding Hispanics and Non-Hispanic Black), adult male, from MSA and arrived in ED by ambulance because of injury or poisoning. The $\hat{\beta}_{GI}$ is an estimated difference from the reference (base) population group due to government insurance, $\hat{\beta}_{NI}$ is an estimated difference from the reference (base) group due to no insurance, and $\hat{\beta}_{Ip}$ is an estimated difference from the reference (base) base group due to injury or poison.

The test statistic shows that the regression model above is significant, once all the seven categories are fitted. It has P-values less than 0.0001 for all tests including Wilks' Lambda, Pillai's Trace, Hotelling-Lawley Trace, and Roy's Greatest Root. But, the partial tests are not significantly different for all population groups (categories) at 0.05% level. Here the test results reflect the two-dimension of both waiting and treatment times simultaneously in multivariate context. In terms of mean waiting and treatment times the test for the difference between private and government insurance can be re-written as: $H_0 : \beta_{GI} = \mu_{PI} - \mu_{GI} = 0 \Leftrightarrow H_0 : \begin{pmatrix} \mu_{PI(waiting)} - \mu_{GI(waiting)} \\ \mu_{PI(treatment)} - \mu_{GI(treatment)} \end{pmatrix} = \begin{pmatrix} \beta_{GI(waiting)} \\ \beta_{GI(treatment)} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$.

Regarding the population groups for the insurance status variable, there are three mutually exclusive groups, where g =3. Then under the assumption of an equal variance-covariance matrix, to test $H_0 : \mu_{PI} = \mu_{GI} = \mu_{NI}$, which is similar to $H_0 : \begin{matrix} \beta_{GI} = \mu_{PI} - \mu_{GI} = 0 \\ \beta_{NI} = \mu_{PI} - \mu_{NI} = 0 \end{matrix}$, where μ_{PI} is the mean vector of waiting and treatment times for the population group with private insurance, μ_{gI} is the mean vector of waiting and treatment times for the population group with government insurance, μ_{nI} is the mean vector of waiting and treatment time for population group with no insurance, β_{pg} is the mean difference of μ_{PI} and μ_{gI} , and β_{pn} is the mean difference of μ_{PI} and μ_{nI} .

Using the SAS[5] statistical package for the computation, the p-value for the three multivariate statistics, Wilks' Lambda (Λ^*), Pillai's Trace and Hotelling-Lawley Trace is 0.43. For the Roy's Greatest Root, the p-value is 0.15.

At the 5% significance level, we failed to reject the null hypothesis. We found no significant difference between mean vectors of patients with the three subcategories of insurance, private, government and no insurance from NHAMCS waiting time and treatment time in EDs. Similarly, we did not find significant difference among subpopulations by race/ethnicity and gender (data not shown).

However, for the age groups consisting of Infants (<2), Children (2 – 17 years), and elderly (65 +), we found significant difference in ED waiting and treatment times compared to adults ($18 \leq$ age < 65), which is the reference (base) population group in the model. The test statistics, Wilks' Lambda, Pillai's Trace, Hotelling-lawley Trace, and Roy's Greatest Root all have p-value <0.0001. The results also show significant differences in the mean vectors for population groups by geographic areas, mode of arrival and nature of emergency (data not shown).

The model including all population groups was statistically significant using the original (pre transformed) data with all partial tests also significant at 5% (data not shown).

4. Conclusion:

NHAMCS is a complex survey with multiple stages of sample selection such that each visit in the covered population has a known non-zero probability of selection. The probabilities of selection, along with the adjustments for nonresponse and post-stratification, are reflected in the survey weights that are provided in the data files. [3] These weights were used in the analysis, particularly in estimating the mean waiting and treatment times of each subcategory. The variance-covariance matrix associated with each subcategory (population group) was assumed to be equal (does not vary among population groups) as stated under case 2 which may not be always true. The unequal variance approach is difficult to apply, unless the variance-covariance matrices of the sample are assumed to be equal with the corresponding population variance-covariance, which is a very strong assumption to consider in survey data.

In multivariate analysis, conclusions based on hypothesis testing reflect simultaneously the features of all variables involved in the analysis. The data transformation made a difference regarding statistical inferences based on the partial test results of insurance status, race/ethnicity and gender. The test results could be different from conclusions based on univariate analysis or analyses. In this paper, if the null hypothesis is rejected in the NHAMCS ED case, the conclusion is based on both waiting and treatment times simultaneously; thus, it may differ from the conclusions if waiting time and treatment time were analyzed separately or analyzed as a total time in EDs by combining both waiting and treatment times. Therefore, the results are solely for illustrative purposes of the statistical methods and no conclusion should be made based on the outcome of the analyses.

Generally, when the data do not tend to be normal even for large sample sizes, using the normal distribution to serve as a very reasonable assumption is not always correct. Population models in some instances need transformation before statistical inferences. The methods mentioned in this paper can be applied in survey data where variables are measured in time units, such as queuing in line, or time taken for some activities, waiting time, particularly in multivariate model analyses, where the underlying assumption of normality is usually ignored and taken for granted.

REFERENCES

1. Rencher, A.C., Methods of Multivariate Analysis, 2nd Edition, John Wiley & Sons, 2002

2. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics. National Hospital Ambulatory Medical Care Survey Public Use File Micro Data 2004 and 2005.

3. Johnson R.A., Wichern, D.W. Applied Multivariate Statistical Analysis, 5th Edition, Prentice Hall press, 2002

4. Nawar EW, Niska RW, Xu J. National Hospital Ambulatory Medical Care Survey: 2005 Emergency Department Summary. Advance data from vital and health statistics; no. 386. Hyattsville, MD: National Center for Health Statistics. 2007.

5. SAS software, version 9.1, SAS inc.

6. D.B. Suits, Use of Dummy Variables in Regression Equations, Journal of Americal Statistical Association, Vol. 52, No. 280, December 1957, 548 - 551

7. R. Khattree, D.N. Naik, Applied Multivariate Statistics with SAS software, 2nd Edition, SAS Institute, Cary, NC, 1999