

Empirical Null Distribution of $-2\log(\lambda)$

Bushra Shamshad

Department of Statistics, University of Karachi
Main University Rd, Karachi, Karachi City, Sindh 75270, Pakistan
Phone: +92 21 99261300
bshamshad@uok.edu.pk

Junaid Saghir Siddiqi

Department of Statistics, University of Karachi
Main University Rd, Karachi, Karachi City, Sindh 75270, Pakistan
Phone: +92 21 99261300
jssdr123@yahoo.com

Abstract

Approximation of Non-central Chi-square distribution as an empirical distribution of log-likelihood ratio test statistics ($-2\log\lambda$; abbreviated as LRT) has been a concern in the field of structural equation modeling. Under extremely severe misspecification (Chun & Shapiro, 2009) reported that non-central Chi-square is not a good choice. In this paper, we have used a bootstrap sampling procedure to investigate the empirical null distribution of LRT specifically in the context of a latent class model (LCM) via frequentist framework (that is, EM algorithm). We used two types of data sets. The first type includes those sets of data on which LCM had been carried out (published results; named as “training data”). The other type is that of those data sets which are not published earlier (i.e. “real” collected data; named as “test data”). Non-central χ^2 distribution with degrees of freedom equals to the expected value of bootstrap LRT and non-centrality parameter equals to inverse of the variance of bootstrapped LRT is found to be very well fitted empirical null distribution of LRT in case of LCM. These results will help in obtaining the significance value of LRT for deciding on the number of classes present in a latent variable.

Keywords: Latent Class Model, Likelihood Ratio Test Statistics, Bootstrapping, Em-Algorithm, Non-Central χ^2 Distribution, Goodness of Fit.

1. Introduction

Likelihood ratio is defined as the ratio of the likelihood of one model (stated in the null hypothesis) over the other (the alternative hypothesis). It is a statistical test and used to compare the fit of one model over the other, where one of the two models is nested in the other. When all regularity conditions are satisfied, $-2\log\lambda$ follows Chi-square distribution and can be used for testing the significance of the fitted model. In the context of latent class analysis and mixture distributions, it has been known for years that the regularity condition of $-2\log\lambda$ does not hold, that is, the model parameters under the null hypothesis lie on the boundary of parameter space (Aitkin., Anderson & Hinde 1981). In other words, it is not possible to carry out the test of likelihood ratio (LR) as two models (under the null and alternative hypothesis) that have a different number of parameters, where one model is nested under the other. A subset of the parameters stated under the null hypothesis is set to zero in order to compare it with the other model stated in the alternative hypothesis. Therefore, it fails to follow the asymptotic chi-square distribution and the distribution is undefined for the likelihood ratio test.

(Wilks, 1935; 1938) showed that for large samples, the distribution of log-likelihood ratio test ($-2\log\lambda$) for nested models will be asymptotically χ^2 with degrees of freedom equal to the difference between the dimensions of the sets of parameters involve in the test statistics (where one (stated in the null hypothesis), among the two compare model, is the special case of the other (stated in an alternative model)). There are a number of studies done for finding whether chi-square can be used as an approximated distribution of $-2\log\lambda$ using simulation techniques. It was then suggested that the distribution of $-2\log\lambda$ can be approximated as chi-square with degrees of freedom equal to

twice the number of manifest items in the model (Wolfe, 1970). Further study (Hartigan, 1977) conducted also showed that the distribution can be approximated by chi-square with degrees of freedom p and $p+1$, where p is the number of manifest items.

The problem in using the LR test was discussed by (Aitkin & Wilson, 1980) reporting that in small samples the test might not follow the asymptotic distribution of likelihood ratio. Furthermore, (Aitkin et al., 1981) and (McLachlan, 1987) showed reservation regarding approximation adequacy for the null distribution of $-2\log\lambda$ to be chi-square since regularity condition does not hold. The use of prior distribution for the vector of mixing proportion was showed by (Aitkin & Rubin, 1985). A note written by (Quinn, McLachlan & Hjort, 1987) which showed that for the approach used by (Aitkin & Rubin, 1985) too the regularity conditions do not hold and therefore the standard asymptotic result cannot be applied.

Several studies conducted for assessing the null distribution of $-2\log\lambda$ by various researchers. As mentioned earlier that through the usual testing procedure, it is considered to be asymptotic chi-square with degrees of freedom equal to the difference between the number of parameters under the null and alternative model. Studies conducted in this respect show reservations, since the regularity condition do not hold for $-2\log\lambda$ with a mixture and latent class model and so the chances for fitting asymptotic χ^2 as a null distribution of $-2\log\lambda$ are small. (Aitkin, et al., 1981) mentioned the suggestions made by (Wolfe, 1970) and (Hartigan, 1977) that $-2\log\lambda$ could be approximated by χ^2_{2k} on the basis of small simulation and that it should be between χ^2_k and χ^2_{k+1} respectively. They assessed the null distribution of $-2\log\lambda$ by simulating its two sets (19 points each), from a single population, in which 38 items (of teaching style data) were independent and the response probabilities are estimated from the real data, to test the hypothesis of a homogenous population at 5% level of significance. Both the sets of values were rejected. The first 19 values were simulated under the null hypothesis of a no-class model against an alternative hypothesis of a two-class model and the other set of values generated under a two-class model against the alternative hypothesis of a 3-class model. Both the sets were also considered as a single sample but none of the results (individual and/or pooled) showed the likelihood of fitting of asymptotic distribution of χ^2 . Thus, we need alternative method(s) for testing the hypothesis on model parameters.

2. The Empirical Null Distribution of LRT ($-2\log\lambda$)

Simulation studies are done in the context of a mixture modeling to approximate the theoretical null distribution of LRT, which include studies/researches by (Titterington et al., 1985) (McLachlan, 1987) (McLachlan & Basford, 1988) and others (for details see McLachlan & Peel, 2000).

3. Bootstrap Likelihood Ratio Test (BLRT) statistic

BLRT was first introduced by (McLachlan, 1987) in his paper for the assessment of the p-value of the likelihood ratio test statistic for the number of component in a normal mixture taking the simplest situation for a specified value of g_1 , of a no-class model $H_0: g = g_1$ against a two-class model $H_1: g = g_2$. This method starts, under the null hypothesis, a bootstrap sample from the mixture model, where $\hat{\varphi}$ being the vector of the likelihood parameter of φ estimated through its MLE from the original data. The value of $-2\log\lambda$ is obtained for each bootstrap sample after fitting the mixture model for $g = g_1$ and g_2 in turn to it. The process is repeated independently M number of times, and the replicated values of $-2\log\lambda$ evaluated from successive bootstrap samples provide an assessment of the bootstrap, and hence of the true, null distribution of $-2\log\lambda$ (McLachlan and Peel, 2000). A test developed by (Lo, Mendell & Rubin, 2001) using the proposed theorem by (Voung, 1989) that under the null hypothesis that the random sample is drawn from g_1 – component normal mixture distribution versus the alternative that it is drawn from g_2 – component normal mixture, where $g_1 < g_2$. The likelihood ratio statistic based on Kullback-Leibler information criteria is asymptotically distributed as a weighted sum of independent χ^2 with 1 degree of freedom. Through simulation results, LMR test evaluates the improvement in the fitting of the two successive models

and provide significance value to see if there is any improvement in the fit for the higher component model.

The inconsistency in the mathematical proof of Lo-Mendell-Rubin test for normal outcome was pointed by (Jeffries, 2003). Despite such critic, many researchers are using LMR test empirically for determining the number of mixtures/ classes. “MPlus” software provides LMR p-value in fitting different models. Also a package named “MplusAutomation_0.5” (Hallquist, 2008)) in R software is available which provides programming for model fitting and calculations for significance values of related fit statistics and extractions of model parameters. The program is developed by Muthen and Muthen (www.statmodel.com).

(Nylund et al., 2007) uses the method proposed earlier of bootstrap likelihood ratio statistic for LCM and calculated the p-value obtained through BLRT estimates of the log likelihood difference distribution which indicates if a t-1-class model is rejected in favor of the t-class model. (Nylund et al., 2007), while comparing the performance of BLRT with LMR and a Naïve chi-square test she also examine the performance of information criteria’s (which include AIC, BIC, CAIC and ABIC) for the LCM, factor mixture model and growth mixture model. The method involves estimating t-1 and t class models for the log-likelihood ratio statistic, which considered as initial estimates. On the basis of the simulation results BIC was marked as the best indicator among other information criterions and the bootstrap likelihood ratio test (BLRT) came up to be a very consistent indicator for determining the number of classes in three types of mixture modeling, namely, LCM, factor mixture modeling and growth mixture modeling. They considered both discrete and continuous LCM and use normal distribution to calculate the significance value of LRT regarding the number of component/ classes in the model. Furthermore, they used Bayesian approach and MCMC simulation technique for the assessment of the significance value of LRT. (Chun & Shapiro, 2009) assess the remarks made by (Nylund et al., 2007) regarding normal approximation. They reported that non-central χ^2 can be approximated for the LRT statistics under reasonable misspecification. According to their research, the findings may vary for different models. The power computation of the test for LRT was done by (Gudicha, Schmittmann, & Vermunt, 2016), using two different methods, to estimate the non-centrality parameter of non-central χ^2 distribution. We have used the method in frequentist framework (i.e. method proposed by Bartholomew in 1987 to find the solution of unknown parameters for LCM through EM-algorithm). Thus, there is a room for investigation of the asymptotic distribution of $-2\log\lambda$ purely in the context of LCM for categorical data.

4. Our Approach for BLRT

To establish its empirical distribution we have used an approach proposed by (McLachlan, 1987) for a mixture of two multivariate normal. (Nylund et al., 2007) uses the same method proposed earlier for bootstrap likelihood ratio statistic. They used normal distribution to calculate the significance value. Although, they mentioned that it does not always fits.

The approach we are using, to find and establish the empirical distribution of likelihood ratio test statistic, is by using confirmatory LCM, which then further used for exploratory models. Initially, the hypothesis under consideration about log-likelihood ratio statistic ($LRT_{t-1,t}$) under the null hypothesis is of “t-1”- class model against the alternative hypothesis of “t”- class model. Bootstrap samples are based on fitting of the LCM (via EM algorithm) while using a set of parameters stated under the null hypothesis. For each bootstrap sample, LCM is estimated for “t-1” and “t” class model to compute the value of $-2\log\lambda$. The complete procedure for bootstrap sampling and storage of $LRT_{t-1,t}$ is presented below.

1. Estimate the model for t-1 and t-class model and calculate the initial estimates of the log-likelihood ratio statistic ($LRT_{t-1,t}$)_{initial} beforehand, which will be used for calculating the significance value of $LRT_{t-1,t}$.

2. Assume a hypothetical population based on the parameters of $(t - 1)$ -latent class model stated in the null hypothesis. From the assumed hypothetical population, generate bootstrap sample and estimate $(t-1)$ and (t) -latent class models to calculate $LRT_{t-1,t}$.
3. Repeat step 2, say “B” times independently to compute $LRT_{t-1,t}$ at each replication.
4. The resultant vector of $LRT_{t-1,t}$ obtained is then used to evaluate the empirical distribution through goodness of fit test.

The process is repeated independently “100” times. it should be noted that the estimation procedure used is same for both bootstrap samples and for the model fitted beforehand for obtaining parameter (McLachlan, 1987) (McLachlan & Krishnan, 2008).

In the following sub sections, we will be investigating the null distribution of LRT through the bootstrap sampling technique on training and test data sets. The hypothetical population considered here is from $(t-1)$ -class model. For the simulation of hypothetical population, LCM is fitted on the concerned data and the estimates thus obtained are treated as parameters. The $(t-1)$ and t -class models are fitted on the sample drawn from the hypothetical population. $LRT_{t-1,t}$ is then calculated using maximum likelihood estimates for $(t-1)$ and t -class models from each booteed sample. The process is repeated multiple number of times (say B-time) and $LRT_{t-1,t}$ of size B, is then assessed for the empirical distribution of $-2\log\lambda$. It is then investigated that the empirical distribution of $-2\log\lambda$ found out to be a non-central χ^2 distribution with degrees of freedom equal to the expected mean of $LRT_{t-1,t}$ and non-centrality parameter (ncp) equals to the inverse of the expected variance of $LRT_{t-1,t}$. The chi-square goodness of fit test shows that the non-central χ^2 distribution (with $df = E(LRT_{t-1,t})$ ncp = $(V(LRT_{t-1,t}))^{-1}$) is very well fitted to the data, for each situation and for each data considered.

In the next sections we present the procedure for each training data, i.e., Mastery, role conflict, and Karachi University Teachers Society (KUTS) panel data. Goodness of fit test and its graphical representation for a single case of size (B =) 100 is presented for bootstrap $LRT_{1,2}$, $LRT_{2,3}$ for both Mastery and Role Conflict data and $LRT_{1,2}$, $LRT_{2,3}$ & $LRT_{3,4}$ for KUTS panel data (since a 3-class model is best fitted on KUTS data).

5. Mastery Data

(Macready & Dayton, 1977) data known as “Mastery data” was also used by (Bartholomew, 1987) for applying latent class model. The data is about a test based on 4 dichotomous items constructed on solving problem of multiplications. The procedure for BLRT is applied to mastery data in two situations 1) 1 vs. 2-class model and 2) 2 vs. 3-class model. As, it is known that a mastery data decomposed into two classes of a single latent variable named as “Master” and “Non Master” class (Macready and Dayton, 1977) (Bartholomew, 1987). Although the procedure is repeated “100” times, we are presenting one simulation result of the goodness of fit test for situation 1 and 2.

It can be seen from the Table 1 (see also Figure 1), that in both situations the non-central χ^2 distribution (with $df = E(BLRT_{t-1,t})$, $ncp = (V(BLRT_{t-1,t}))^{-1}$) is a good fit to the null distribution of LRT, that is, p-value for χ^2 goodness of fit for $LRT_{1,2}$ is 0.46701 with 3 df (= no. of classes–2–1; as 2 parameters are estimated). Whereas, for $LRT_{2,3}$, the p-value equal to 0.36136 with 3 df, also indicate very well fitting of non-central χ^2 distribution.

Table 1. Mastery data; expected frequencies obtained through the fitting of non-central χ^2 distribution for size “B = 100”. Observed frequencies from the bootstrapped sample of; (a) $LRT_{1,2}$ (b) $LRT_{2,3}$

$LRT_{1,2}$	Observed frequency	Empirical frequency
(0.548,2.56]	18	16.99109
(2.56,4.56]	28	27.755298
(4.56,6.57]	27	23.037882
(6.57,8.58]	10	14.861048
(8.58,10.6]	9	8.40641
(10.6,12.6]	3	4.38686
(12.6,14.6]	2	2.167725

(14.6,36]	3	1.884808
Total	100	99.491121
χ^2 (goodness of fit)	2.546142	
Degrees of freedom (df)	3	
p-value	0.46701	

(a)

$LRT_{2,3}$	observed frequency	Empirical frequency
(0,1.67]	28	28.393741
(1.67,3.39]	32	30.546773
(3.39,5.11]	20	19.279113
(5.11,6.83]	8	10.702626
(6.83,8.55]	9	5.591767
(8.55,10.3]	1	2.819109
(10.3,12]	1	1.38809
(12,33.2]	1	1.278692
Total	100	99.999911
χ^2 (goodness of fit)	3.203071	
Degrees of freedom (df)	3	
p-value	0.361363	

(b)

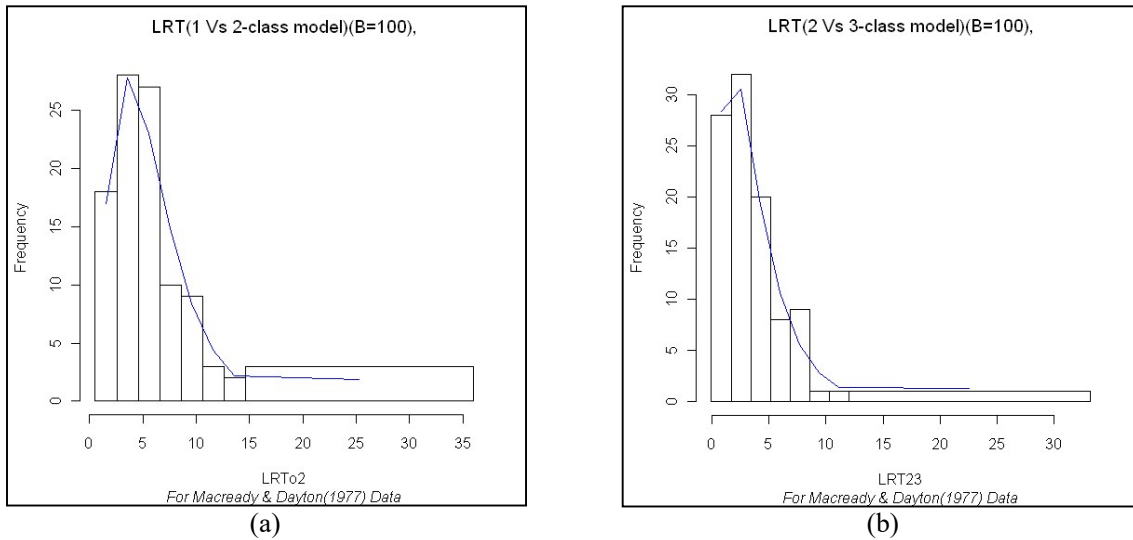


Figure 1. Mastery data (for $B = 100$); along with a superimposed non-central χ^2 distribution curve: (a) Histogram for $BLRT_{1,2}$; (b) Histogram for $BLRT_{2,3}$.

6. Role Conflict Data

Role conflict data is taken from (Coleman, 1964). (Goodman, 1974a; 1974b) used this data to explain a restricted LCM, which further discussed by (Bartholomew, 1987). The data is a panel data collected at two different points in time. Two questions were asked each time from individuals. Each question was responded as either ‘positive’ or ‘negative’. A restricted model with an assumption that there exist two latent variables, and that they altogether form four latent classes is the solution for the data. The empirical distribution of BLRT ($B = 100$) is presented in the Table 2 for (a) Situation 1 (H_0 : 1-class model Vs. H_1 : 2-class model) and (b) Situation 2 (H_0 : 2-class model Vs. H_1 : 3-class model), along with the expected frequencies and goodness of fit test.

In each case, Chi-square test indicates the goodness of fit for the non-central χ^2 (with $df = E(LRT_{t-1,t})$, $ncp = V(LRT_{t-1,t})^{-1}$) to the empirical null distribution of BLRT. That is, p-value for χ^2 goodness of fit in the first situation is 0.276083 with 2 df. Whereas, in the second situation the p-

value equal to 0.058646 with 3 df also indicate very well fitting of non-central χ^2 distribution. The goodness of fit can also be seen in Figure 2 for both $LRT_{1,2}$ and $LRT_{2,3}$.

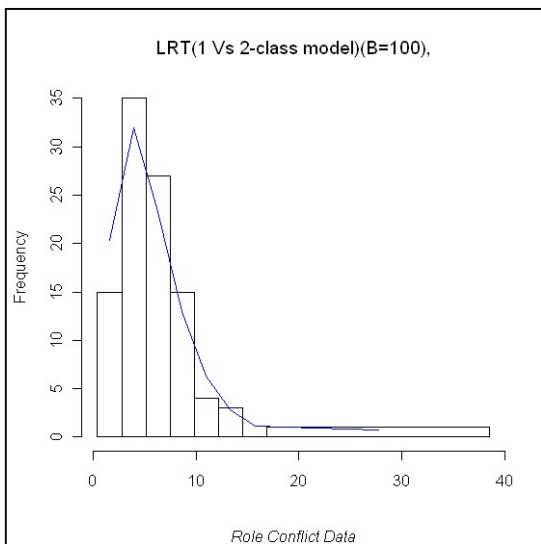
Table 2. Role conflict data; Expected frequencies obtained through the fitting of non-central χ^2 distribution for size “B = 100”. Observed frequencies from the bootstrapped sample of (a) $LRT_{1,2}$; (b) $LRT_{2,3}$.

$LRT_{1,2}$	Observed Frequency	Empirical Frequency
(0.45,2.8]	15	20.31083
(2.8,5.15]	35	31.96709
(5.15,7.49]	27	23.35075
(7.49,9.84]	15	12.99846
(9.84,12.2]	4	6.299644
(12.2,14.5]	3	2.806942
(14.5,16.9]	0	1.182026
(16.9,38.5]	1	0.779773
Total	100	99.6955
χ^2 (goodness of fit)		3.867928
Degrees of freedom (df)		2
p-value		0.276083

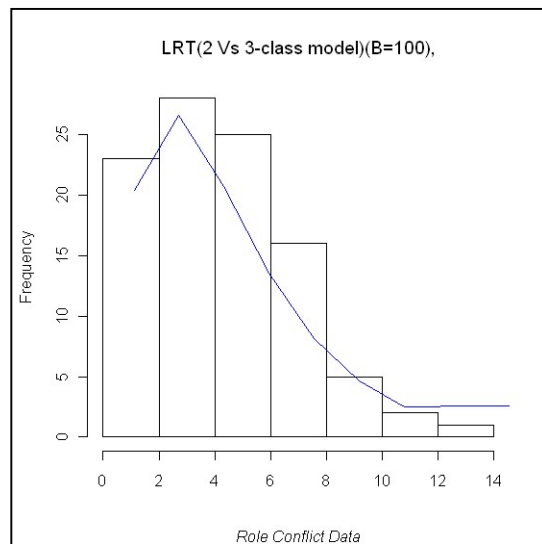
(a)

$LRT_{1,2}$	Observed Frequency	Empirical Frequency
(0.298,1.91]	23	20.37098
(1.91,3.53]	23	26.60688
(3.53,5.14]	19	20.69197
(5.14,6.75]	21	13.54058
(6.75,8.37]	9	8.11963
(8.37,9.98]	2	4.617904
(9.98,11.6]	2	2.534554
(11.6,32.7]	1	2.812054
Total	100	99.29455
χ^2 (goodness of fit)		7.458069
Degrees of freedom (df)		3
p-value		0.058646

(b)



(a)



(b)

Figure 2. Role Conflict data (for B = 100); Histogram along with superimposed non-central χ^2 curve for (a) $BLRT_{1,2}$; (b) $BLRT_{2,3}$.

7. KUTS Panel Data

KUTS panel data is the original results of the election held in 1993-94 of teachers of the University of Karachi. Two groups (panels) were contesting we named them as “Rightist” and “Mix” based on their manifesto. The data was first used by (Shamshad & Siddiqi, 2012) to fit LCM and found that a 3-class model is best among other class models. As described earlier the strategy for Mastery and Role conflict data we have assessed the empirical distribution of BLRT for KUTS panel data in three situations. (a) Situation 1; (H_0 : 1-class model Vs. H_1 : 2-class model) and (b) Situation 2; (H_0 : 2-class model Vs. H_1 : 3-class model) (c) Situation 3; (H_0 : 3-class model Vs. H_1 : 4-class model).

Non-central χ^2 distribution when fitted to BLRT in each situation gives test value (chi-square goodness of fit) very small (< 2), which is an indication of high p-value, that is, for test values 1.98, 0.1565 and 1.173 the p-values are 0.371, 0.924 and 0.882 for $BLRT_{1,2}$, $BLRT_{2,3}$ and $BLRT_{3,4}$ respectively (see Table 3; Figure 3).

Table 3. KUTS panel data; Observed and expected frequencies (obtained through fitting non-central χ^2 distribution) for size “B = 100”; Bootstrapped sample of (a) $LRT_{1,2}$ (b) $LRT_{2,3}$ (c) $LRT_{3,4}$

$LRT_{1,2}$	Observed Frequency	Expected frequency
(0,2.37]	21	21.5922
(2.37,4.8]	37	35.4063
(4.8,7.22]	19	23.1141
(7.22,9.65]	12	11.5248
(9.65,12.1]	8	5.05212
(12.1,14.5]	2	2.05339
(14.5,16.9]	0	0.79404
(16.9,38.6]	1	0.46302
Total	100	99.99994
χ^2 (goodness of fit)		1.980392
Degrees of freedom (df)		2
p-value		0.371504

(a)

$LRT_{2,3}$	Observed Frequency	Expected Frequency
(0,2.17]	23	22.83642
(2.17,4.4]	34	34.10453
(4.4,6.62]	23	22.26926
(6.62,8.84]	12	11.5264
(8.84,11.1]	5	5.340879
(11.1,13.3]	2	2.319014
(13.3,15.5]	0	0.964693
(15.5,37]	1	0.638765
Total	100	99.99995
χ^2 (goodness of fit)		0.156566
Degrees of freedom (df)		2
p-value		0.924703

(b)

$LRT_{3,4}$	Observed Frequency	Expected Frequency
(0,1.06]	21	21.84624
(1.06,2.16]	24	24.17677
(2.16,3.27]	18	18.18279
(3.27,4.38]	16	12.58786
(4.38,5.49]	7	8.380645
(5.49,6.59]	6	5.45261

(6.59,7.7]	3	3.493749
(7.7,18.5]	5	5.828512
Total	100	99.94918
χ^2 (goodness of fit)		1.173873
Degrees of freedom (df)		4
p-value		0.882381

(c)

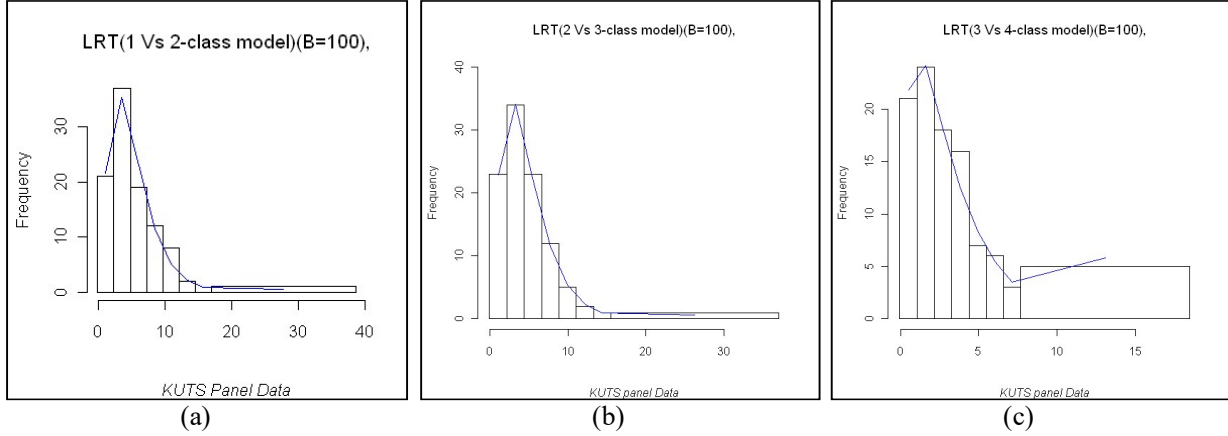


Figure 3. KUTS panel data (B=100); Histogram and a superimposed curve of non-central χ^2 distribution of bootstrapped (a) $LRT_{1,2}$; (b) $LRT_{2,3}$; (c) $LRT_{3,4}$.

Table 4. Percent Acceptance Rate of Non- Central χ^2 Distribution with $df = E(BLRT_{t-1,t})$, $n_{cp} = V(BLRT_{t-1,t})^{-1}$

		level of significance	100	200	500
Mastery data	LRT(1 vs. 2-class model)	5%	86.73%	87.88%	88.89%
		1%	95.92%	93.94%	91.92%
	LRT(2 vs. 3-class model) "	5%	80.81%	80.00%	55.00%
		1%	96.97%	90.00%	67.00%
role conflict data	LRT(1 vs. 2-class model) "	5%	85.86%	85.00%	85.98%
		1%	90.91%	94.00%	94.39%
	LRT(2 vs. 3-class model) "	5%	86.00%	81.08%	74.26%
		1%	95.00%	91.22%	87.13%
KUTS panel data	LRT(1 vs. 2-class model) "	5%	78.57%	81.82%	27.27%
		1%	88.78%	93.94%	43.43%
	LRT(2 vs. 3-class model) "	5%	82.83%	78.79%	60.61%
		1%	93.94%	92.93%	78.79%
	LRT(3 vs. 4-class model) "	5%	82.65%	62.63%	42.98%
		1%	96.94%	78.79%	55.26%

For sizes (i.e. B = 100, 200 and 500) approximately 100 repetitions are done to validate the fitting of non-central χ^2 distribution on the empirical null distribution of $-2\log\lambda$ for Mastery, Role Conflict and KUTS panel data. The summary presented in Table 4, show the percent acceptance rate of non-central χ^2 distribution at 5% and 1% level of significance. It can be concluded that the null distribution of BLRT is non-central χ^2 , when modeling is done using LCM. For B = 500, the percentage rate of acceptance is quite low, which can be improved by reevaluating the frequency distribution of BLRT, as we have used the same codes for the construction of frequency distribution of BLRT as used for size 100. Since, for larger samples the distribution of BLRT is highly skewed and need appropriate class intervals for the construction of frequency distribution of BLRT. Thus, once again the provision of accepting the decision, regarding the number of classes in the model, through calculating the significance value of BLRT, using the non-central χ^2 distribution (with $df = E(LRT_{t-1,t})$, $n_{cp} = V(LRT_{t-1,t})^{-1}$) is very strong.

8. Test Data Sets: Description of General Health Parameter (GHP) Data

GHP data set is collected through a survey of a long questionnaire (more than 76 questions) from more than 1500 students studied at different government and private sector Universities and Colleges in Karachi, Pakistan during years 2008-2010. The purpose is to check whether the respondent is aware about his/her health conditions. Asking health related questions not only give a chance to a respondent to overview his/her health condition, but also provide collective information about health of teenagers in the society. Even if the respondent would not answer correctly due to any reason, at least he/she would think and realize any sort of problem by themselves. Questionnaire was adopted from the survey of world health organization and constructed in both Urdu and English languages separately, keeping in mind that most of the targeted population might not be comfortable in the English version as they are in their learning stage and could have difficulty in understanding the language, which might result in misleading information. Each question has been asked about the difficulty the respondents have had in doing work, moving around, listening, seeing, understanding, recognizing and remembering thing etc., considering the last 30 days. In order to get true response, each question is scaled from 1 to 5, 1 being (at a minimum) “none of the time” or “no problem” and 5 being the “all of the time” or “extreme problem”.

From this survey we have focused our attention toward an important public health issue of problematic sleep which requires accurate diagnosis. Sleep problem is referred as both symptom and sign of specific disorder, known as “insomnia”. (Roth, 2007) defines “insomnia” in survey studies as a positive response to either question “Do you have difficulty falling or staying asleep?” or “Do you experience difficulty sleeping?” According to International Classification of Sleep Disorder, 2nd edition (ICSD-2) “insomnia” is defined as having complaints of difficulty in initiating and maintaining sleep or wake up too early or having a sleep that is of poor quality, such difficulties occur even with appropriate circumstances and adequate opportunities for sleep (Schutte-Rodin et al., 2008) (Buysse, 2008), which results in at least one of the following daytime dysfunctioning. That are, daytime sleepiness; fatigue or anxiety; concentration, remembering or focusing problems; difficulty in complex mental tasks; social or vocational dysfunction; poor school/job performance; irritable or bad mood. Having tension, headache or gastrointestinal symptoms in response to sleep loss; motivation, energy or initiative reduction; proneness for error/accidents at work or while driving and concern about sleep problem (ICSD-2). Here, we consider variables which are daytime impairment associated with insomnia as either symptom or diagnostic criteria. We named the set of these variables as “GHP-Insomnia data”. The selected questions are as follow:

R: How much difficulty did you have with concentrating and remembering things?
(**Concentration Problem**)

C: How much difficulty did you have with analyzing and solving problem in day to day life? (**Cognitive Issue**)

L: How much difficulty did you have with learning new tasks, for example, learning how to get a new place? (**Learning Issue**)

I: How much did you feel irritable or having a bad mood? (**Irritable**)

S: How much difficulty did you feel in falling asleep, waking up frequently during night or waking up too early in the morning? (**Problematic sleep**)

We have used a total of 1289 responses after discarding the responses having missing values for this analysis. For convenience, 5-level likert scale has been reduced to binary as “1-No difficulty” being marked as a negative response (“1”) and the rest of the level of having difficulty of any degree (That are; 2, 3, 4, 5) are marked as a positive response (“2”). Summary of latent class model fitting up to 5 classes are presented in Table 5, in which the AIC is minimum in 3-class model and BIC in 2-class model. The difference between the values of G^2 and χ^2 for 1-class and 2-class model is also an indication of the presence of a latent variable in the data.

Table 6 presents the estimated model parameters for 1 till 3-classes the probabilities presented are of a positive response to each item along with the respective estimated standard errors of estimates. In case of 2-class model the class proportions dividing the total sampled population in two groups are estimated as 67% and 33%. Class 1 (with the highest proportion of approximately 867 out of 1289 total respondents), show high probabilities of having difficulty in each and every statement. Approximately, 674 out of 867 individuals in this class faced difficulty in concentrating and remembering things (R), 765 and 781 (out of 867) had difficulty in the cognitive issue (C) and feel irritable (I), respectively, in the same duration. Whereas, approximately 75% and 70% of the respondents had had problematic sleep (S) and issues in learning new tasks (L), respectively. Class 2 on the other hand constitutes those respondents who felt irritable (I) (approximately, 311 out of 422) and had problematic sleep (S) (approximately; 218 out of 422). The probabilities of having positive response to questions related to the concentration issue (R), cognitive issue (C) and the learning issue (L) are very low in class 2. The probability of being irritable or having a bad mood is high among individuals in class 2 although the probability of having sleep issue is also high but not too high. Individual in class 1 seems to have at risk of having Insomnia. In a 3-class model (see Table 6) number of estimated parameters increased but with a clearer grouping of individuals with an additional class. There are only 8% individuals in the total sampled population who marked negative response to difficulty in any of the question asked with high probabilities. In Class 1 (with approximately 108 individual out of 1289) none of the respondent feel being irritable or in a bad mood as well as only 37, 28, 31 and 39 (out of 108 individual) show difficulty in the concentration issue, cognitive issue, learning issues and problematic sleep respectively, and these probabilities are not very high. We marked this group as “healthy fellows”.

Whereas, class 2 with maximum proportion of 835 out of 1289 individuals (64.7% class proportion) represents that group of people who are at great risk of insomnia since the probabilities are very high for each and every question asked, i.e. they have disturbed sleep (S) (624 out of 835 marked positive response) as well as they face difficulty in daytime activities such as in remembering and concentrating things (R), cognitive issues (C), learning new tasks (L) and have a bad mood (I) with around 78%, 88%, 74%, and 89% probabilities, respectively. We marked class 2 as “chronic-insomnia risk group”. Respondents in class 3 (class proportion is 26.89%) give 100% positive response to being irritable (I) and 204 out of 347 respondents complain problematic sleep (S), although this group of individuals have no difficulty, in analyzing and solving day to day problems in life (C) (with 62.8% proportion of 347), in concentrating issue (R) (with 51.38% probability) and 81.54% proportion of 347 in learning new tasks (L). For class 3, the probability of having disturbed sleep is slightly high; this might be due to reason of having a bad mood of all individuals of this group. Feeling irritable might be a sign of anxiety or depression due to social or psychological issues, which may result in the nighttime sleep problem. We marked class 3 as “chronic irritable group”.

Table 5. Results of Fitting Latent Class Models for GHP data

	1 class	2 class	3 class	4 class	5 class
AIC	7752.916	7545.709	7536.830	7540.031	7546.453
BIC	7778.724	7602.487	7624.577	7658.748	7696.140
G ² (Likelihood ratio /deviance statistic)	257.969	38.762	17.883	9.084	3.506
χ^2 (Chi-square goodness of fit)	341.947	37.597	17.706	8.896	3.406
Number of estimated parameters	5	11	17	23	29
maximum log-likelihood	-3871.45	-3761.85	-3751.41	-3747.01	-3744.22
$-2\log\lambda$	219.206	20.880	8.800	5.577	

Table 6. GHP survey data; Latent class model parameter estimates for 1 to 3-classes (probabilities of having difficulty)

	1-latent class	2-latent class model		3-latent class model		
	Model	Class 1	Class 2	Class 1	Class 2	Class 3
R	0.6633 [0.0136]	0.7772 [0.0207]	0.4290 [0.0391]	0.3459 [0.0567]	0.7778 [0.0222]	0.4862 [0.0457]
C	0.6920 [0.0137]	0.8823 [0.0255]	0.3005 [0.0543]	0.2632 [0.0627]	0.8804 [0.0302]	0.3716 [0.0695]
L	0.5516 [0.0145]	0.7062 [0.0254]	0.2335 [0.0434]	0.2970 [0.0558]	0.7368 [0.0345]	0.1846 [0.0703]
I	0.8472 [0.0102]	0.9006 [0.0130]	0.7372 [0.0285]	0.0000 [0.0000]	0.8930 [0.0143]	1.0000 [0.0000]
S	0.6734 [0.0133]	0.7493 [0.0192]	0.5172 [0.0346]	0.3677 [0.0565]	0.7481 [0.0199]	0.5884 [0.0378]
Class Proportion		0.6729 [0.0452]	0.3271 [0.0452]	0.0835 [0.0110]	0.6476 [0.0597]	0.2689 [0.0549]

The cross classification table (see Table 7b) of latent class membership against problematic sleep (S: insomnia) for the 3-class model present that 32.6% of the total individuals responded that they did not have problematic sleep at any time during the last month, whereas, 32.58% had problematic sleep, some of the time. The marginal totals for problematic sleep for responses “A good bit of the time”, “Most of the time” and “All of the time” are 17.22%, 10.7% and 6.8%, respectively. As, we have converted the responses dichotomously as “no difficulty (1)” and “having difficulty of any degree (2 → (2, 3, 4 & 5))”. Therefore, the cumulative for responses (2, 3, 4 & 5) of “having difficulty of any degree” gives 67.33% of the total respondents. The proportion of class membership obtained through the cross classification against problematic sleep (S) are approximately the same as obtained through the fitted model (see Table), that are, 8.9% (116 out of 1289), 66.4% (856 out of 1289) and 24.5% (317 out of 1289) for classes 1, 2 and 3, respectively (see Table:). The distribution of individuals in each of the 5-levels of responses clearly show that in class 1 (marked as “healthy fellows”), 66.37% of the individuals do not face any difficulty in sleeping, while only 2.58% and 6% of the total face difficulty “all of the time” and “most of the time”, respectively. Although, the cumulative percentage of responses “some of the time” and “a good bit of the time” is 25% (29 out of 116).

In class 2 (marked as “chronic-insomnia risk group”), 37.5% of 856 respondents marked no difficulty in having asleep during the nighttime. Even though, only 65 of 856 (6%) respondents had sleep problem all of the time, but the remaining cumulative of difficulty (low to marginally high degree; responses 2, 3 & 4) is 56.15% which is quite alarming. The percentage distribution in 5-level scale of having asleep at bedtime for the last group (class 3) of “chronic-irritable” show that, a total of 73.7% of 317 respondents faced difficulty (for responses; 2, 3, 4, & 5). The individual percentages of having difficulty are also very high, that are 35%, 19.8%, 11.2% and 7.59% for responses 2, 3, 4 & 5, in class 3, as compared to other classes. This might be due to irritability they were facing during that time, which could be a cause of any social, psychological or mental pressure. The cross classification table for a 2-class model is also presented (see Table 7a) for reader interest. We will now assess the empirical null distribution of $-2\log\lambda$ for GHP-Insomnia data. The estimated parameter of 1- till 3- latent class model provided in Table--- are used for booting hypothetical populations based on these estimates. For each $LRT_{1,2}$, $LRT_{2,3}$ and $LRT_{3,4}$, 100 replications of bootstrap samples are obtained on which non-central χ^2 distribution with $df = E(LRT_{l,t})$ and $n_{cp} = [V(LRT_{l,t})]^{-1}$ is assessed for fitting distribution.

Table 7. GHP-Insomnia data: Cross classification tables of latent class membership against problematic sleep (insomnia) for: (a) 2-class model; (b) 3-class model

Problematic sleep (insomnia)	Latent class membership		Total
	“Healthy Fellows group” Class 1	“Chronic-insomnia risk group” Class 2	
1-None of the time	189	232	421
2- Some of the time	92	328	420
3-A good bit of the time	44	178	222
4-Most of the time	32	106	138
5-All of the time	22	66	88
Total	379	910	1289

(a)

Problematic sleep (insomnia)	Latent class membership			Total
	“Healthy Fellows group” Class 1	“Chronic-insomnia risk group” Class 2	“Cornice-irritable group” Class 3	
1-None of the time	77	225	119	421
2- Some of the time	22	300	98	420
3-A good bit of the time	7	170	45	222
4-Most of the time	7	96	35	138
5-All of the time	3	65	20	88
Total	116	856	317	1289

(b)

The significance values obtained from $BLRT_{1,2}$, $BLRT_{2,3}$ and $BLRT_{3,4}$ are 0.265872, 0.664608 and 0.08312 against test values 219.206, 20.88 and 8.8 each with 3 degrees of freedom, respectively, (see Table 8; see also Figure 4), show good fit of non-central χ^2 distribution for each $BLRT_{t-1,t}$.

Table 8. GHP-Insomnia data (size B = 100); Observed frequencies; Expected frequencies (obtained through the fitting of non-central χ^2 distribution; (a) $LRT_{1,2}$; (b) $LRT_{2,3}$; (c) $LRT_{3,4}$

$LRT_{1,2}$	Observed Frequency	Expected Frequency
(0.73,3.57]	16	10.80113
(3.57,6.41]	24	29.29937
(6.41,9.24]	25	27.83557
(9.24,12.1]	18	17.3231
(12.1,14.9]	11	8.624018
(14.9,17.8]	4	3.742099
(17.8,20.6]	1	1.479546
(20.6,42.5]	1	0.835142
Total	100	99.93998
χ^2 (goodness of fit) Test Statistics		3.959493
Degrees of freedom (df)		3
P-value		0.265872

(a)

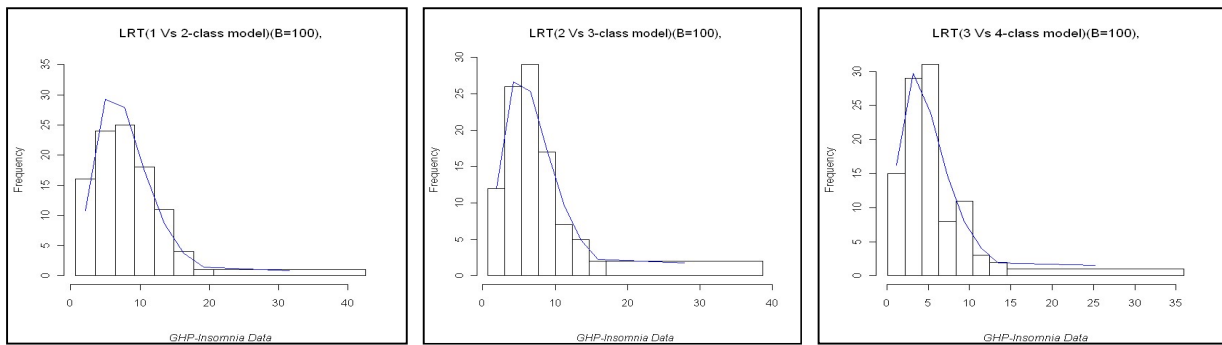
$LRT_{2,3}$	Observed Frequency	Expected Frequency
(0.75,3.08]	12	12.08323
(3.08,5.41]	26	26.62737
(5.41,7.74]	29	25.28383
(7.74,10.1]	17	17.12687
(10.1,12.4]	7	9.693352

(12.4,14.7]	5	4.908126
(14.7,17.1]	2	2.303954
(17.1,38.7]	2	1.758664
Total	100	99.7854
χ^2 (goodness of fit) Test Statistics		1.577037
Degrees of freedom (df)		3
P-value		0.664608

(b)

LRT _{3,4}	Observed Frequency	Expected Frequency
(0.155,2.21]	15	16.22409
(2.21,4.26]	29	29.66279
(4.26,6.31]	31	23.90222
(6.31,8.37]	8	14.75638
(8.37,10.4]	11	7.985343
(10.4,12.5]	3	3.994056
(12.5,14.5]	2	1.895745
(14.5,35.9]	1	1.542697
Total	100	99.96332
χ^2 (goodness of fit) Test Statistics		6.672024
Degrees of freedom (df)		3
P-value		0.08312

(c)



(a)

(b)

(c)

Figure 4. GHP-Insomnia data (B=100); Histogram along with a superimposed non-central χ^2 curve for bootstrap sample of; (a) LRT1,2 ; (b) LRT2,3; (c) LRT3,4

9. Conclusion

We have considered four different data sets: Mastery, Role Conflict, KUTS panel and GHP-Insomnia data. The task of establishing the empirical distribution of $-2\log\lambda$ when fitting latent class model reveals that non-central χ^2 distribution with $df = E(LRT_{t-1,t})$ and $ncp = [V(LRT_{t-1,t})]^{-1}$ is very well fitted (with high percentage) to each data set (either “training” or “test”) and that one can rely on calculating the significance value of $-2\log\lambda$ obtained through the empirical null distribution (i.e. Non-central χ^2 distribution).

References

Aitkin, M., & Rubin, D. B. (1985). Estimation and hypothesis testing in finite mixture models. *Journal of the Royal Statistical Society Series B (Methodological)*, 47 (1), 67-75.

- Aitkin, M., & Wilson, G. T. (1980). Mixture Models, Outliers, and the EM Algorithm. *Technometrics*, 22(3), 325-331.
- Aitkin, M., Anderson, D., & Hinde, J. (1981). Statistical Modeling of Data on Teaching Styles. *Journal of the Royal Statistical Society, Series A (General)*, 144(4), 419-461.
- American Academy of Sleep Medicine. Diagnostic and Coding Manual. 2nd ed. Westchester, Ill: American Academy of Sleep Medicine; 2005. The International Classification of Sleep Disorders.
- Bartholomew, D. J. (1987). *Latent Variable Models and Factor Analysis*. London: Charles Griffin & Co. Ltd.
- Buysse D. J. (2008). Chronic insomnia. *The American journal of psychiatry*, 165(6), 678–686.
- Chun, S. Y., & Shapiro, A. (2009). Normal versus Noncentral Chi-square asymptotics of misspecified models. *Multivariate Behavioral Research*, 44 (6), 803-27.
- Coleman, J.S. (1964). *Introduction to Mathematical Society*. New York: Free Press.
- Gudicha, D. W., Schmittmann, V. D., & Vermunt, J. K. (2016) Power Computation for Likelihood Ratio Tests for the Transition Parameters in Latent Markov Models. *Structural Equation Modeling: A Multidisciplinary Journal*, 23 (2), 234-245.
- Goodman, L. A. (1974a). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*, 61(2), 215-231.
- Goodman, L. A. (1974b). The analysis of systems of qualitative variables when some of the variables are unobservable. Part I-A modified latent structure approach. *American Journal of Sociology*, 79(5), 1179-1259.
- Hallquist (2008-2011). *MplusAutomation: Automating Mplus Model Estimation and Interpretation*. R package version 0.5. <http://CRAN.R-project.org/package=MplusAutomation>.
- Hartigan, J. A. (1977). Distribution Problems in Clustering. In J. V. Ryzin, ed., *Classification and Clustering*, New York: Academic Press.
- Jeffries, N. (2003). A note on “Testing the number of components in a normal mixture”. *Biometrika*, 90, 991–994.
- Lo, Y., Mendell, N., & Rubin, D. (2001). Testing the number of component in a normal mixture. *Biometrika* 88, 767-778.
- Macready, G. B., & Dayton, C. M. (1977). The use of probabilistic models in the assessment of mastery. *Journal of Educational Statistics*, 2, 99-120.
- McLachlan, G. J. (1987). On bootstrapping the likelihood ratio test statistic for the number of components in a normal mixture. *Applied Statistics*, 36, 318-324.
- McLachlan, G. J., & T. Krishnan. (2008). *The EM algorithm and Extensions*. Hoboken, N.J: Wiley-Interscience.
- McLachlan, G., & Peel, D. (2000). *Finite Mixture Models*. New York: Wiley.
- McLachlan, G. J., & Basford, K. E. (1988). *Mixture Models: Inference and Application to Clustering*. New York: Marcel Dekker.
- Nylund, K. L., Asparouhov, T., Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling*, 14(4), 535–569.
- Quinn, B. G., McLachlan, G. J., & Hjort, N. L. (1987). A note on the Aitkin-Rubin approach to hypothesis testing in mixture models. *Journal of the Royal Statistical Society B*, 49, 311-314.
- Roth, T. (2007). Insomnia; Definition, prevalence, etiology, and consequences. *Journal of Clinical Sleep Medicine*, 3(5 Suppl): S7-S10.
- Schutte-Rodin, S., Broch, L., Buysse, D., Dorsey, C., & Sateia, M. (2008). Clinical guideline for the evaluation and management of chronic insomnia in adults. *Journal of Clinical Sleep Medicine*, 4(5), 487–504.
- Shamshad, B., & Siddiqi, J. S. (2012) Exploration of groups through latent structural model. *Journal of Basic and Applied Science*, 8 (1), 145-150.
- Titterton, D. M., Smith, A. F. M., & Markov, U. E. (1985). *Statistical Analysis of Finite Mixture Distributions*. New York: Wiley.

- Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, 57, 307-33.
- Wilks, S. S. (1935). The likelihood test of independence in contingency tables. *The Annals of Mathematical Statistics*, 6(4), 190-196.
- Wilks, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics*, 9(1), 60-62.
- Wolfe, J.H. (1970). Pattern clustering by multivariate mixture analysis. *Multivariate Behavioral Research*, 5, 329–350.



Distribution Theory and Structure Equation Modeling.

Bushra Shamsad received her B.Sc (Honors), M.Sc and Ph.D degree in Statistics from Department of Statistics, University of Karachi, Pakistan, in 2002, 2003 and 2013, respectively. She is associated with the Department of Statistics, Karachi University since 2004, initially as a Co-operative Lecturer then became Full-time Lecturer in 2006. She is working as Assistant Professor since 2011. Her awards and honors include First-Class-First Position and 2 gold medals in M.Sc. and First-Class-Second Position in B.Sc. (Honors). Her research interests are Multivariate Analysis, Categorical Data Analysis,



Dr. Junaid Saghir Siddiqi born on 01-05-1952, He did his B.Sc. Hons. & M.Sc. in Statistics, in 1972 and 1973 respectively, from the Department of Statistics, University of Karachi. He did Ph.D. in Statistics from University of Exeter, England in 1992. He initially worked as Teaching / Research Assistant in Karachi University. He then joined Government of Sindh as Research Officer before joining the Department of Statistics, Karachi University as a Full-time Lecturer in October 1975. He has been retired as Professor of Statistics on 30th April 2012. Currently he is working as an adjunct Professor in the Department. His research interest includes Application of multivariate methods in several social and economic field. He has supervised 5 PhD students.