The International Journal of Engineering and Science (IJES)
|| Volume || 6 || Issue || 5 || Pages || PP 07-15 || 2017 ||
ISSN (e): 2319 – 1813 ISSN (p): 2319 – 1805

Mixed Model Analysis for Overdispersion

B. Vedavathi [1], B. Muniswamy[2]
[1, 2] *Department of Statistics, Andhra University, Visakhapatnam*

----------ABSTRACT----------
In this paper we focus on mixed model analysis for regression model to take account of over dispersion in random effects. Moreover, we present the Data Exploration, Box plot, QQ plot, Analysis of variance, linear models, linear mixed –effects model for testing the over dispersion parameter in the mixed model. A mixed model is similar in many ways to a linear model. It estimates the effects of one or more explanatory variables on a response variable. In this article, the mixed model analysis was analyzed with the R-Language. The output of a mixed model will give you a list of explanatory values, estimates and confidence intervals of their effect sizes, *P*-values for each effect, and at least one measure of how well the model fits. The application of the model was tested using open-source dataset such as using numerical illustration and real datasets.
Keywords- Count data, Linear mixed model, Overdispersion, Random effects to study variations.

Date of Submission: 04 April 2017 Date of Accepted: 24 April 2017

I. INTRODUCTION

Over dispersion is the condition by which data appear more dispersed than is expected under a reference model. For count data, with repeated measurements on each subject over time or space, or to multiple related outcomes at one point in time we use mixed model analysis. This mixed model approach allows a wide variety of correlation patterns (**or variance –covariance structures**) to be explicitly modeled. The advantage of mixed models is that they naturally handle uneven spacing of repeated measurements whether intentional or un-intentional. Also important is the fact that mixed model analysis is often more interpretable than classical repeated measures. Finally mixed models can also be extended as generalized mixed models to non-normal outcomes. The term mixed model refers to the use of both fixed and random effects in the same analysis. Fixed effects have levels that are of primary interest and would be used again if the experiment were repeated. Random effects have levels that are not of primary interest, but rather are thought of as a random selection from a much larger set of levels. Subject effects are almost always random effects, while treatment levels are almost always fixed effects. Other examples of random effects include cities in a multi-site trial, batches in a chemical or industrial experiment, and classrooms in an educational setting. As an alternative to underlying normal variable models, previous authors have defined multivariate distributions for mixed outcomes by incorporating shared normally distributed random effects in generalized linear mixed models (**Moustaki, 1996; Sammel et al. 1997; Moustaki and Knott, 2000; Dunson, 2000, 2003**). Although models of this type are very flexible, the lack of simple expressions for the marginal mean and variance makes parameter interpretation difficult. In addition, model fitting tends to be highly computationally intensive, particularly when more than a few random effects are incorporated. Generalized Linear Model (**GLM**) context (**i.e., models without random effects**), and many software packages such as R (**R Core Team, 2014**) will calculate this value automatically for GLMs. For Generalized Linear Mixed Models (**GLMMs**), the situation becomes more complex due to uncertainty in how to calculate the residual degrees of freedom (d.f.) for a model that contains random effects. For mixed models, the dispersion parameter can be calculated as the ratio of the sum of the squared Pearson residuals to the residual degrees of freedom (**e.g., Zuur et al., 2009**)

The use of both fixed and random effects in the same model can be thought of hierarchically, and there is a very close relationship between mixed models and the class of models called hierarchical linear models. the fixed effects parameters tell how population means differ between any set of treatments, while the random effect parameters represent the general variability among subjects or other units.

II. LINEAR MIXED MODEL

The linear mixed model is defined as $Y = x_{ij}^{t}\beta + u_{ij}^{t}\gamma_i + \varepsilon_{ij}$

Where

Y_{ij} is the response of j^{th} member of cluster $i, i = 1, \ldots, m$ and $j = 1, \ldots, n_i$

m is the number of clusters.

n_i is size of cluster *i*

x_{ij} is the covariate vector of j^{th} member of cluster i for fixed effects, $\in R_P$

β is the fixed effects parameter $\in R_P$

u_{ij} is the covariate vector of j^{th} member.

III. FIXED AND RANDOM FACTORS/EFFECTS

How can we extend the linear model to allow for such dependent data structures?
Fixed factor = qualitative covariate (e.g. gender, age group)
Fixed effect = quantitative covariate (e.g. age)
Random factor = qualitative variable whose levels are randomly sampled from a population of levels being studied.
Random effect = quantitative variable whose levels are randomly sampled from a population of levels being studied.

IV. MIXED LINEAR MODEL (LMM) I

Assumptions:

$\gamma_i \sim Nq(0, D)$, and $D \in \square^{q \times q}$

$$\varepsilon_i = \begin{pmatrix} \varepsilon_{i1} \\ \vdots \\ \varepsilon_{in_i} \end{pmatrix} \square\ N_{n_i}(0,\ \Sigma_i)\ ,\quad \Sigma_i \in \square^{n_i \times n_i}$$

$\gamma_1, \ldots, \gamma_m, \varepsilon_{1,}\varepsilon_{1,} \ldots\ldots \varepsilon_{m,}$ are independent

D = covariance matrix of random effects γ_i

Σ_i = covariance matrix of error vector ε_i in cluster *i*

Mixed Linear Model (LMM) II

Matrix Notation:

$$X_i = \begin{pmatrix} x_{i1}^t \\ \vdots \\ x_{in_i}^t \end{pmatrix} \in \square^{n_i \times p}\ ,\ U_i = \begin{pmatrix} u_{i1}^t \\ \vdots \\ u_{in_i}^t \end{pmatrix} \in \square^{n_i \times q},\ Y_i = \begin{pmatrix} Y_{i1} \\ \vdots \\ Y_{in_i} \end{pmatrix} \in \square^{n_i}$$

This implies;

$$\begin{aligned} &Y_i = X_i\beta + U_i\gamma_i + \varepsilon_i \\ &\gamma_i \sim N_q(0, D) \\ &\gamma_1, \ldots, \gamma_m, \varepsilon_{1,}\varepsilon_{1,} \ldots\ldots \varepsilon_{m,} \textit{are independent.} \\ &\varepsilon_i \square\ N_{n_i}(0,\ \Sigma_i) \end{aligned}$$

V. LIKELIHOOD INFERENCE FOR LMM:

1) Estimation of β and γ for known *G* and *R*
Estimation of β: Using (5), we have as MLE or weighted LSE of β

$$\hat{\beta} = \left(X^t V^{-1} X\right)^{-1} X^t V^{-1} Y$$

This estimate is called the weighted LSE

Estimation of γ :

We know that $Y = N_n(X\beta,\ \mathrm{V}) \quad \gamma = N_{mq}(0, g)$

$$Cov(Y,\gamma) = Cov(X\beta + U\gamma,\ \gamma)$$
$$= \underbrace{Cov(X\beta,\ \gamma)}_{=0} + U\ \underbrace{\mathrm{var}(\gamma,\ \gamma)}_{g} + \underbrace{Cov(\varepsilon,\ \gamma)}_{=0}$$
$$= Ug$$
$$E(\gamma|Y) = 0 + gU^tV^{-10}(Y - X\beta)$$
$$= gU^tV^{-1}(Y - X\beta)$$

This is the best linear unbiased predictor of γ (BLUP)

Joint maximization of log likelihood of $\left(Y^t,\ \gamma^t\right)^t$ with respect to $(\beta,\ \gamma)$

$$f(y,\gamma) = f(y|\gamma)\, f(\gamma)$$
$$= \exp\left\{\frac{-1}{2}\gamma^t g^{-1}\gamma\right\}$$
$$\Rightarrow \ln f(y,\gamma) = \frac{-1}{2}(y - X\beta - U\gamma)^t R^{-1}(y - X\beta - U\gamma)$$
$$\frac{-1}{2}\ \underbrace{\gamma^t g^{-1}\gamma}_{\text{penality term of }\gamma} + \text{constants ind. of } (\beta,\gamma)$$

So, it is enough to minimize.

$$Q(\beta,\gamma) = (y - X\beta - U\gamma)^t R^{-1}(y - X\beta - U\gamma) - \gamma^t g^{-1}\gamma$$
$$= \gamma^t R^{-1}\gamma - 2\beta^t X^t R^{-1}y + 2\beta^t X^t R^{-1}U\gamma - 2\gamma^t U^t R^{-1}y$$
$$+ \beta^t X^t R^{-1}X\beta + \gamma^t U^t R^{-1}U\gamma + \gamma^t g^{-1}\gamma$$

Mixed Model Equation:

$$\frac{\partial Q(\beta,\gamma)}{\partial\beta} = -2X^t R^{-1}y + 2X^t U\gamma + 2X^t R^{-1}X\beta = 0$$
$$\frac{\partial Q(\beta,\gamma)}{\partial\beta} = -2U^t R^{-1}X\beta - 2U^t R^{-1}y + 2U^t R^{-1}U\gamma + 2G^{-1}\gamma = 0$$
$$\Leftrightarrow X^t R^{-1}X\beta + X^t R^{-1}U\gamma = X^t R^{-1}y$$
$$U^t R^{-1}X\beta + \left(U^t R^{-1}U + G^{-1}\right)\gamma = U^t R^{-1}y$$

ML Estimation in extended marginal model:

$Y = X\beta + \varepsilon^*, \varepsilon^* \square\ N_n(0, V(\upsilon))$ with $V(\upsilon) = UG(\upsilon)U^t + R(\upsilon)$

Log likelihood for (β,υ)

$$l(\beta,\upsilon) = -\frac{1}{2}\left\{\ln|V(\upsilon)| + (y - X\beta)^t V(\upsilon)^{-1}(y - X\beta)\right\} + cont.ind.of\ \beta,\upsilon$$

If we maximize (11) for fixed ϑ with regard to β, we get

$$\hat{\beta}(\upsilon) = \left(X^t V(\upsilon)^{-1} X\right)^{-1} X^t V(\upsilon)^{-1} y$$

Then the profile log likelihood is

$$l_p(\upsilon) = l\left(\beta(\upsilon), \upsilon\right)$$
$$= -\frac{1}{2}\left\{\ln\left|V(\upsilon)\right| + \left(y - X\hat{\beta}(\upsilon)\right)^t V(\upsilon)^{-1}\left(y - X\hat{\beta}(\upsilon)\right)\right\}$$

Maximizing $l_p(\upsilon)$ w.r.t to υ gives MLE $\hat{\upsilon}_{ML}$. $\hat{\upsilon}_{ML}$ is however biased; this is why one uses often restricted ML estimation (REML)

Summary: Estimation in LMM with unknown covariance.

For the linear mixed model $Y = X\beta + U\gamma + \varepsilon,$

$$\begin{pmatrix}\gamma\\ \varepsilon\end{pmatrix} \sim N_{mq+n}\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}G(\upsilon) & O_{mq\times n}\\ O_{n\times mq} & R(\upsilon)\end{pmatrix}\right)$$

With $V(\upsilon) = UG(\upsilon)U^t + R(\upsilon)$ the covariance parameter vector υ is estimated by either

$\hat{\upsilon}_{ML}$ which maximizes

$$l_p(\upsilon) = -\frac{1}{2}\left\{\ln\left|V(\upsilon)\right| + \left(y - X\hat{\beta}(\upsilon)\right)^t V(\upsilon)^{-1}\left(y - X\hat{\beta}(\upsilon)\right)\right\}$$

Where $\hat{\beta} = \left(X^t V(\upsilon)^{-1} X\right)^{-1} X^t V(\upsilon)^{-1} Y$

or by $\hat{\upsilon}_{REML}$ which maximizes $l_R(\upsilon) = l_p(\upsilon) - \frac{1}{2}\ln\left|X^t V(\upsilon)^{-1} X\right| + C$

The fixed effects β and random effects γ are estimated by

$$\hat{\beta} = \left(X^t V^{-1} X\right)^{-1} X^t V^{-1} Y$$
$$\hat{\gamma} = GU^t V^{-1}\left(Y - X\hat{\beta}\right)$$

Where $\hat{V} = V\left(\hat{\upsilon}_{ML}\right)$ or $V\left(\hat{\upsilon}_{REML}\right)$

Confidence interval and hypothesis tests:

Since $Y \sim N\left(X\beta, V(\vartheta)\right)$ hold, an approximation to the covariance of $\hat{\beta} = \left(X^t V^{-1}\left(\hat{\vartheta}\right) X\right)^{-1} X^{-t} V^{-1}\left(\hat{\vartheta}\right) Y$ is given by

$$\left(X^t V^{-1}\left(\hat{\vartheta}\right) X\right)^{-1}$$

Note: here one assumes that $V\left(\hat{\vartheta}\right)$ is fixed and does not depend on *Y*.

Therefore $\hat{\sigma}_j := \left(X^t V^{-1}\left(\hat{\vartheta}\right) X\right)^{-1}_{jj}$ are considered as estimates of $Var\left(\hat{\beta}_j\right)$.

Therefore

$$\hat{\beta}_j \pm z_{1-\alpha/2}\sqrt{\left(X^t V^{-1}\left(\hat{\vartheta}\right) X\right)^{-1}_{jj}}$$

Gives an approximate $100(1-\alpha)\%$ CI for β_j.

It is expected that $\left(X^t V^{-1}\left(\hat{\vartheta}\right) X\right)^{-1}_{jj}$ underestimated $Var\left(\hat{\beta}_j\right)$ since the variation in $\hat{\vartheta}$ is not taken into account.

A full Bayesian analysis using MCMC methods is preferable to these approximations.

Under the assumption that $\hat{\beta}$ is asymptotically normal with mean β and covariance matrix $A(\vartheta)$, then the usual hypothesis tests can be done; i.e., for

- $H_0 : \beta_j = 0$ versus $H_1 : \beta_j \neq 0$

  Reject $H_0 \Leftrightarrow |t_j| = \left|\frac{\hat{\beta}_j}{\hat{\sigma}_j}\right| > z_{1-\alpha/2}$

- $H_0 : C\beta = d$ versus $H_1 : C\beta \neq d$ $\text{rank}(C) = r$

  Reject $H_0 \Leftrightarrow W : (C\hat{\beta} - d)^t \left(C^t A(\hat{\vartheta}) C\right)^{-1} (C\hat{\beta} - d) > \chi^2_{1-\alpha,r}$ (Wald-Test)

  Or

  Reject $H_0 \Leftrightarrow -2\left[l(\hat{\beta}, \hat{\gamma}) - l(\hat{\beta}_R, \hat{\gamma}_R)\right] > \chi^2_{1-\alpha,r}$

  Where $\hat{\beta}, \hat{\gamma}$ estimates in unrestricted model

  $\hat{\beta}_R, \hat{\gamma}_R$ estimates in restricted model $(C\beta = d)$ (Likelihood Ratio Test)

Example:

Description Of Contents Of The Data:

The data set contains information on 910 persons about Diabetes .They weighted measured at baseline and again they returned to campl1 year later. Each time, a serum sample was taken from which a determination of hemoglobin $A1c$ (HgbA1C) was made.HgbA1C also called glycosylated hemoglobin. This is routinely monitored by insulin injections. Missing data are indicated by blanks. At the end of the variable name implies that the variable is being considered as a factor.

Field	Description
mon_a1c	Month A1c
day_a1c	Day A1c
yr_a1c	Yr A1c
age_yrs	Age in years
gly_a1c	Hemoglobin A1c
ht_cm	Height in cm missing=999.9
wt_kg	Weight in kg
sex	M-Male, F-female

Source Of The Data:

Fundamentals of Biostatistics; Seventh edition; by BERNARD ROSNER.

Display the names of variables in column order of the data frame, also explains the characteristics of the variable:

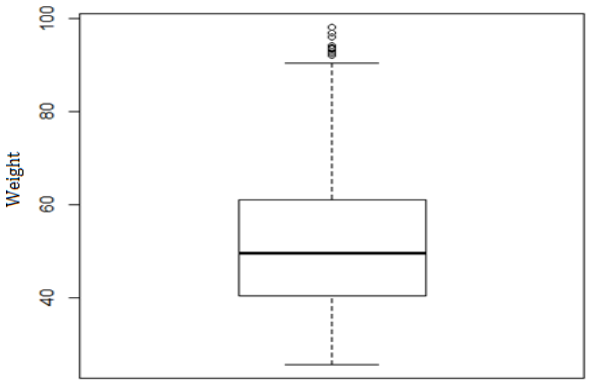
```
> data<-read.table("C://Users//vedavathi//Desktop//mydata.csv", header=TRUE, sep=",")
> names(data)
[1] "ID"      "mon_a1c" "day_a1c" "yr_a1c"  "age_yrs" "gly_a1c" "ht_cm"   "wt_kg"   "Sex"
> str(data)
'data.frame':   910 obs. of  9 variables:
 $ ID     : int  118130 118130 118130 120882 120882 120882 120882 120882 120882 124129 ...
 $ mon_a1c: int  4 10 3 5 5 11 7 7 7 4 ...
 $ day_a1c: int  26 4 21 11 15 1 23 1 21 20 ...
 $ yr_a1c : int  88 88 89 88 89 89 90 91 92 88 ...
 $ age_yrs: num  14.2 14.7 15.1 10.5 11.5 12 12.7 13.6 14.7 10.7 ...
 $ gly_a1c: num  9.1 8.77 9.26 9.93 8.36 ...
 $ ht_cm  : num  158 161 164 142 149 ...
 $ wt_kg  : num  54.9 57.1 61.5 40.5 45.5 47.5 50.1 59.3 67.3 43.2 ...
 $ Sex    : Factor w/ 2 levels "F","M": 2 2 2 1 1 2 1 2 1 1 ...
```

Data Exploration: The summary statistics for each variable defined in data is shown below:

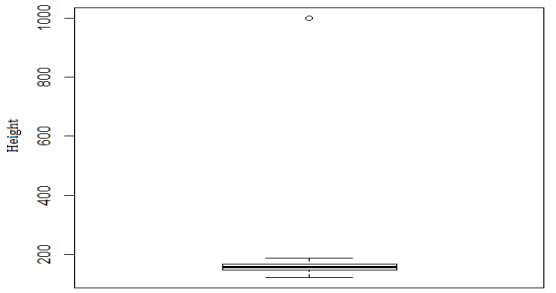
```
> summary(data)
       ID            mon_a1c          day_a1c          yr_a1c
 Min.   :118130   Min.   : 1.000   Min.   : 1.00   Min.   :85.00
 1st Qu.:149237   1st Qu.: 3.000   1st Qu.: 8.00   1st Qu.:90.00
 Median :158336   Median : 6.000   Median :17.00   Median :92.00
 Mean   :159564   Mean   : 6.343   Mean   :16.17   Mean   :92.51
 3rd Qu.:170820   3rd Qu.: 9.000   3rd Qu.:23.00   3rd Qu.:95.00
 Max.   :200889   Max.   :12.000   Max.   :31.00   Max.   :98.00
    age_yrs          gly_a1c           ht_cm           wt_kg         Sex
 Min.   : 9.00   Min.   : 5.298   Min.   :122.0   Min.   :25.70   F:465
 1st Qu.:11.40   1st Qu.: 7.777   1st Qu.:145.4   1st Qu.:40.50   M:445
 Median :13.00   Median : 8.603   Median :156.0   Median :49.50
 Mean   :12.75   Mean   : 8.808   Mean   :172.4   Mean   :51.29
 3rd Qu.:14.30   3rd Qu.: 9.595   3rd Qu.:167.2   3rd Qu.:60.98
 Max.   :15.50   Max.   :16.207   Max.   :999.9   Max.   :98.10
```

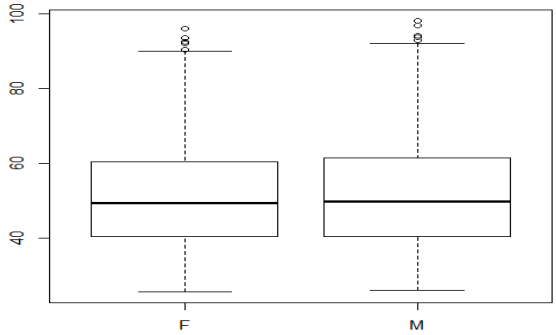
To check the normality of the data use Box plot:

```
> attach(data)
> boxplot(wt_kg)
> boxplot(ht_cm)
> boxplot(wt_kg~Sex)
> qqnorm(wt_kg,main="Normal Q-Q plot for Weight",ylab="weight")
> qqline(wt_kg)
```

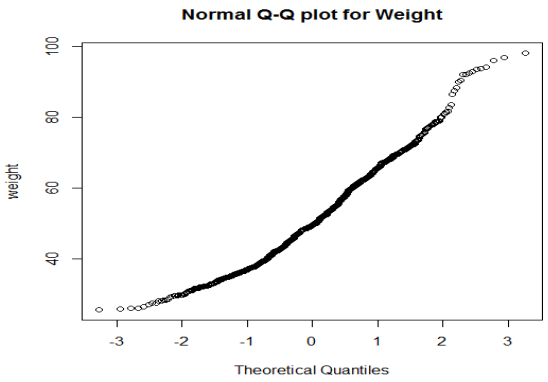


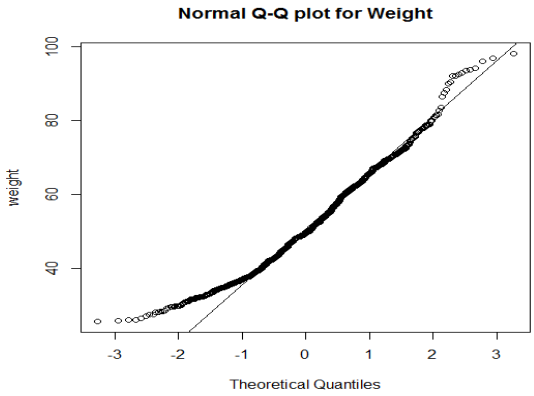
From the above plot, it is clear that the weight appears normally distributed. The central line indicates the median. Also the graph shows that there are some outliers.



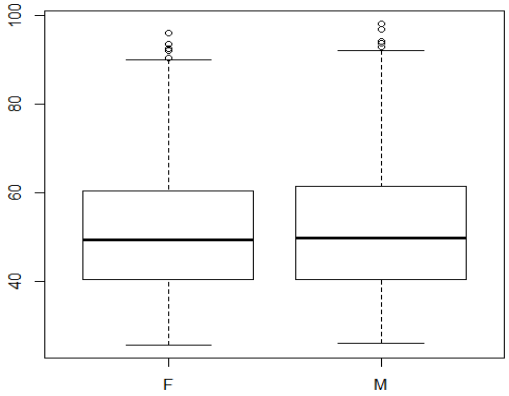


Q-Q Plot:





By observing the above plot, we can say that the weight follows normal distribution as the plotted point's forms approximately a straight line.



The above Boxplot reveals the weights of male and female's. The weights are normally distributed.

VI. DATA ANALYSIS

The following is the R-Code and output to run a generalized linear model to fit:

```
> fit1<-lm(wt_kg~ht_cm+gly_a1c+age_yrs+yr_a1c+day_a1c+mon_a1c+Sex,data)
> summary(fit1)

Call:
lm(formula = wt_kg ~ ht_cm + gly_a1c + age_yrs + yr_a1c + day_a1c +
    mon_a1c + Sex, data = data)

Residuals:
     Min       1Q   Median       3Q      Max
-18.2271  -5.9372  -0.7197   5.1082  31.5953

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -87.034002   9.593514  -9.072  <0.0000 ***
ht_cm         0.001145   0.002492   0.460   0.6460
gly_a1c      -0.351696   0.194614  -1.807   0.0711 .
age_yrs       5.783645   0.170632  33.895  <0.0000 ***
yr_a1c        0.721802   0.100891   7.154   0.0000 ***
day_a1c       0.034909   0.033000   1.058   0.2904
mon_a1c       0.056875   0.085070   0.669   0.5039
SexM         -0.448153   0.577837  -0.776   0.4382
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.671 on 902 degrees of freedom
Multiple R-squared: 0.5966,     Adjusted R-squared: 0.5934
F-statistic: 190.5 on 7 and 902 DF,  p-value: < 2.2e-16
```

```
> anova(fit1)
Analysis of Variance Table

Response: wt_kg
           Df Sum Sq Mean Sq   F value    Pr(>F)
ht_cm       1   1370    1370   18.2287 0.0000217 ***
gly_a1c     1    969     969   12.8941 0.0003474 ***
age_yrs     1  93993   93993 1250.2151 0.0000000 ***
yr_a1c      1   3780    3780   50.2768 0.0000000 ***
day_a1c     1     85      85    1.1356 0.2868598
mon_a1c     1     34      34    0.4514 0.5018594
Sex         1     45      45    0.6015 0.4382059
Residuals 902  67813      75
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

To introduce random effects models, the three models were compared.

1. Model 1: With yr_a1c and age_yrs as random effects.
2. Model 2: With yr_a1c as random effect.
3. Model 3: With only age_yrs as random effect.

VII. MODEL 1

Response variable: wt_kg
Fixed effects: Sex, mon_a1c, day_a1c, gly_a1c, ht_cm
Random effects: yr_a1c + age_yrs
The R-code and output:

```
> library(lme4)
> library(foreign)
> lmer(wt_kg~Sex+mon_a1c+day_a1c+gly_a1c+ht_cm+(1|yr_a1c)+(1|age_yrs),data=data)
Linear mixed model fit by REML ['lmerMod']
Formula: wt_kg ~ Sex + mon_a1c + day_a1c + gly_a1c + ht_cm + (1 | yr_a1c) + (1 | age_yrs)
   Data: data
REML criterion at convergence: 6713.362
Random effects:
 Groups   Name        Std.Dev.
 age_yrs  (Intercept) 10.604
 yr_a1c   (Intercept)  2.361
 Residual              8.535
Number of obs: 910, groups:  age_yrs, 66; yr_a1c, 13
Fixed Effects:
(Intercept)         SexM      mon_a1c      day_a1c      gly_a1c        ht_cm
 50.4814391   -0.5661842    0.0223023    0.0446197   -0.2717199    0.0007087
```

From the output, the residual is 8.535

VIII. MODEL 2

Response variable: wt_kg
Fixed effects: Sex, mon_a1c, day_a1c, gly_a1c, ht_cm, age_yrs
Random effects: yr_a1c

The R-code and output for Model 2:

```
> lmer(wt_kg~Sex+mon_a1c+day_a1c+gly_a1c+ht_cm+age_yrs+(1|yr_a1c),data=data)
Linear mixed model fit by REML ['lmerMod']
Formula: wt_kg ~ Sex + mon_a1c + day_a1c + gly_a1c + ht_cm + age_yrs + (1 | yr_a1c)
   Data: data
REML criterion at convergence: 6539.101
Random effects:
 Groups   Name        Std.Dev.
 yr_a1c   (Intercept) 2.464
 Residual             8.626
Number of obs: 910, groups:  yr_a1c, 13
Fixed Effects:
(Intercept)         SexM      mon_a1c      day_a1c      gly_a1c        ht_cm
 -2.038e+01   -5.361e-01    2.417e-02    3.558e-02   -3.516e-01    6.938e-04
     age_yrs
  5.805e+00
```

From the above model 2 output, the residual is 8.626

IX. MODEL 3

Response variable: wt_kg
Fixed effects: Sex, mon_a1c, day_a1c, gly_a1c, ht_cm, yr_a1c
Random effects: age_yrs

The R-code and output for Model 3:

```
> lmer(wt_kg~Sex+mon_a1c+day_a1c+gly_a1c+ht_cm+yr_a1c+(1|age_yrs),data=data)
Linear mixed model fit by REML ['lmerMod']
Formula: wt_kg ~ Sex + mon_a1c + day_a1c + gly_a1c + ht_cm + yr_a1c + (1 | age_yrs)
   Data: data
REML criterion at convergence: 6706.153
Random effects:
 Groups   Name        Std.Dev.
 age_yrs  (Intercept) 10.54
 Residual              8.59
Number of obs: 910, groups:  age_yrs, 66
Fixed Effects:
(Intercept)         SexM      mon_a1c      day_a1c   gly_a1c     ht_cm       yr_a1c
 -13.424430    -0.473324     0.055380     0.043073 -0.270892  0.001041     0.688729
```

From the above model 3 output, the residual is 8.59

X. RESULTS

Age in years contributes variation in the glycosylated hemoglobin we may choose Model 3 (With only age in years as random effects) on the basis of its REML value 6706.153.

Comparison of the yr_a1c and age in years variance components in model 1 indicates that the standard deviation component for age in years (10.604) yr_a1c (2.361) and residual (8.535). With the random terms (yr_a1c and age in years) included in the model, the variance from 112.4448 to 5.5743

With only yr_a1c as random effect in Model 2, the standard deviation component for yr_a1c is 2.464 and residual is 8.626 .

The mixed model with yr_a1c component alone included utilizes almost equivalent information as the mixed model with both yr_a1c and age in year's component included.

Yet, our fundamental target was to look at the fuse of arbitrary impacts to study variations among yr_a1c and age in years and their impact on person's weight. Subsequently to accomplish this objective we may pick Model 1 since it contains both random effects yr_a1c and age in years. The residuals among the three models, Model 1 has less residual with 8.535. Hence **Model 1 is suggestable**.

XI. REFERENCE

[1]. Fahrmeir, L., T. Kneib, and S. Lang (2007). Regression: Modelle, Methoden and Anwendungen. Berlin: Springer-Verlag.
[2]. West, B., K. E. Welch, and A. T. Galecki (2007). Linear Mixed Models: a practical guide using statistical software. Boca Raton: Chapman-Hall/CRC.
[3]. Mixed model analysis using R: Stephen Mbunzi & Sonal Nagada
[4]. Applied Multivariate Statisticsl Anaalysis 5E by Richard A. Johnson, Dean W. Wichern
[5]. Applied Multivariate Statisticsl Anaalysis 6E by Richard A. Johnson, Dean W. Wichern
[6]. MixedModels: A Julia Package for Fitting (Statistical) Mixed-Effects Models. Julia package version 0.3-22,
[7]. Mullen KM, Nash JC, Varadhan R (2014). minqa: Derivative-Free Optimization Algorithms by Quadratic Approximation. R package version 1.2.4
[8]. DebRoy S (2004). "Linear Mixed Models and Penalized Least Squares." Journal of Multivariate Analysis, 91(1),
[9]. Zero-Inflated Negative Binomial Mixed Regression Modeling of Over-Dispersed Count Data with Extra Zeros By Kelvin K.W.Yau, Kui wang, Andy H.Lee
[10]. Using observation-level random effects to model overdispersion in count data in ecology and evolution. Xavier A. Harrison Institute of Zoology, Zoological Society of London, London, UK