Stats112 Final

Viraj Vijaywargiya

2023-06-10

```
library( nnet )
library(lattice)
library(nlme)
library(lme4)
```

```
## Loading required package: Matrix
```

```
##
## Attaching package: 'lme4'
```

```
## The following object is masked from 'package:nlme':
##
##     lmList
```

```
library(geepack)
library(survival)
library(emdbook)
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
```

```
## v ggplot2 3.3.6      v purrr   0.3.5
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.3      v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::collapse() masks nlme::collapse()
## x tidyr::expand()   masks Matrix::expand()
## x dplyr::filter()   masks stats::filter()
## x dplyr::lag()      masks stats::lag()
## x tidyr::pack()     masks Matrix::pack()
## x tidyr::unpack()   masks Matrix::unpack()
```

```
BoneWide = read.csv("/Users/virajvijaywargiya/Downloads/BoneWide.csv")
BoneLong = read.csv("/Users/virajvijaywargiya/Downloads/BoneLong.csv")
BoneWide$Trt = as.factor(BoneWide$Trt)
BoneLong$Trt = as.factor(BoneLong$Trt)
BoneWide$Trt = relevel(BoneWide$Trt, ref="P")
BoneLong$Trt = relevel(BoneLong$Trt, ref="P")
```

```
ID.comp = BoneWide$ID[complete.cases(BoneWide)]
BoneLong.comp = BoneLong[BoneLong$ID %in% ID.comp,]
BoneWide.comp = BoneWide[BoneWide$ID %in% ID.comp,]
```

1. ```
   na.count = function(x){ return( sum(is.na(x)) ) }
   NumberOfObservations = 5-apply(BoneWide, 1, na.count)/2
   table(NumberOfObservations)
   ```

   ```
   ## NumberOfObservations
   ##  1  2  3  4  5
   ##  7  6  5  3 91
   ```

   Above is the completed table, where the first row represents "Number of repeated measurements" and the second row represents "Number of subjects". The table shows the distribution of the number of observations across subjects. The "Number of subjects" shows the count of subjects corresponding to each number of repeated measurements. For example, there are 7 subjects who have only one measurement, 6 subjects with two measurements, 5 subjects with three measurements, 3 subjects with four measurements, and 91 subjects with the complete set of five measurements.
   This table provides an overview of the distribution of observations and highlights the impact of missing data on the number of measurements available for each subject. It helps identify the subjects who dropped out and provides insights into the issue of missing data in this study. The table shows that there are 21 subjects who have fewer than five measurements. These subjects are likely the ones who dropped out of the study or were lost to follow-up before all measurements could be obtained.
   In terms of the 21 subjects that dropped out of the study and were lost to follow-up, we can infer that they have fewer than five measurements, suggesting missing data due to dropout or loss to follow-up. The missing data could be due to various reasons such as non-compliance, withdrawal from the study, or other factors. Among the 21 subjects who dropped out, some have only one measurement, indicating that they left the study early and did not participate in subsequent measurements. The largest number of observations is seen in subjects who have completed the study, with 91 subjects having the full set of five measurements.

2. For a complete case analysis to be valid with these data, we need to assume that the missing data mechanism is Missing Completely At Random (MCAR). Under the MCAR assumption, the probability of missingness is unrelated to both observed and unobserved data. In this scenario, the missing subjects are considered to be a random subset of the overall sample, and their absence does not introduce bias into the analysis.

   However, if the missing data mechanism is not MCAR, it can cause issues for our analysis. Specifically, if the missing data mechanism is Missing Not At Random (MNAR) or Missing At Random (MAR), the missingness may be related to the unobserved data or variables of interest. This can lead to biased estimates and potentially affect the validity of inference.

The potential causes of this type of missing data in the study could include various factors such as non-compliance with the treatment protocol, withdrawal from the study due to adverse effects, lack of follow-up appointments, or personal reasons.

3. ```
   table(BoneWide$Trt)
   ```

   ```
   ##
   ##  P  C
   ## 57 55
   ```

Number of subjects in the control group (Placebo): 57
Number of subjects in the reatment group (Calcium): 55

```
table(BoneWide$BMICat)
```

```
##
## Normal   Over  Under
##     65     31     16
```

Normal: The "Normal" category has a count of 65, indicating that there are 65 individuals in the study whose body mass index falls within the "Normal" category.

Over: The "Over" category has a count of 31, suggesting that there are 31 individuals whose body mass index is classified as "Over," indicating they have a higher body mass index compared to the normal range.

Under: The "Under" category has a count of 16, indicating that there are 16 individuals classified as having a body mass index categorized as "Under," implying they have a lower body mass index compared to the normal range.

4. Mean and standard deviation (respectively) for age, for each visit index by treatment:

```
tapply(BoneLong$Age, list(BoneLong$Trt,BoneLong$Visit), mean, na.rm=TRUE)
```

```
##          1        2        3        4        5
## P 11.05614 11.57170 12.02157 12.55417 13.03617
## C 11.05455 11.56346 12.02500 12.55000 13.03864
```

```
tapply(BoneLong$Age, list(BoneLong$Trt,BoneLong$Visit), sd, na.rm=TRUE)
```

```
##           1          2          3         4         5
## P 0.1018008 0.10261459 0.09233358 0.1110076 0.1030524
## C 0.1015038 0.09707253 0.10619132 0.1295291 0.1125103
```

Mean and standard deviation (respectively) for BMD, for each visit index by treatment:

```
tapply(BoneLong$BMD, list(BoneLong$Trt,BoneLong$Visit), mean, na.rm=TRUE)
```

```
##           1         2         3         4         5
## P 0.8700702 0.8896792 0.9141569 0.9416458 0.9582340
## C 0.8804545 0.9032692 0.9382500 0.9645000 0.9881591
```

```
tapply(BoneLong$BMD, list(BoneLong$Trt,BoneLong$Visit), sd, na.rm=TRUE)
```

```
##            1          2          3          4          5
## P 0.06577012 0.07509936 0.07802830 0.07530519 0.07363931
## C 0.05968895 0.05932731 0.05878214 0.06510120 0.06287194
```

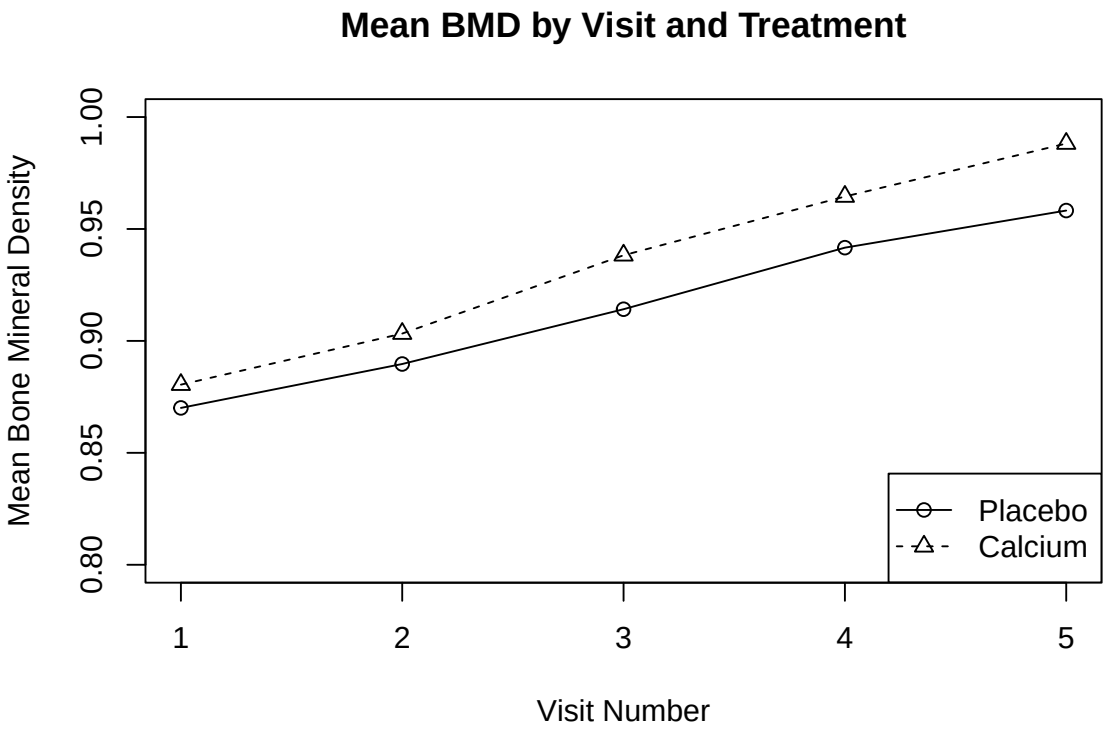
Number of missing observations for each visit index by treatment:

```
tapply(BoneLong$BMD, list(BoneLong$Trt,BoneLong$Visit), na.count)
```

```
##   1 2 3 4  5
## P 0 4 6 9 10
## C 0 3 7 9 11
```

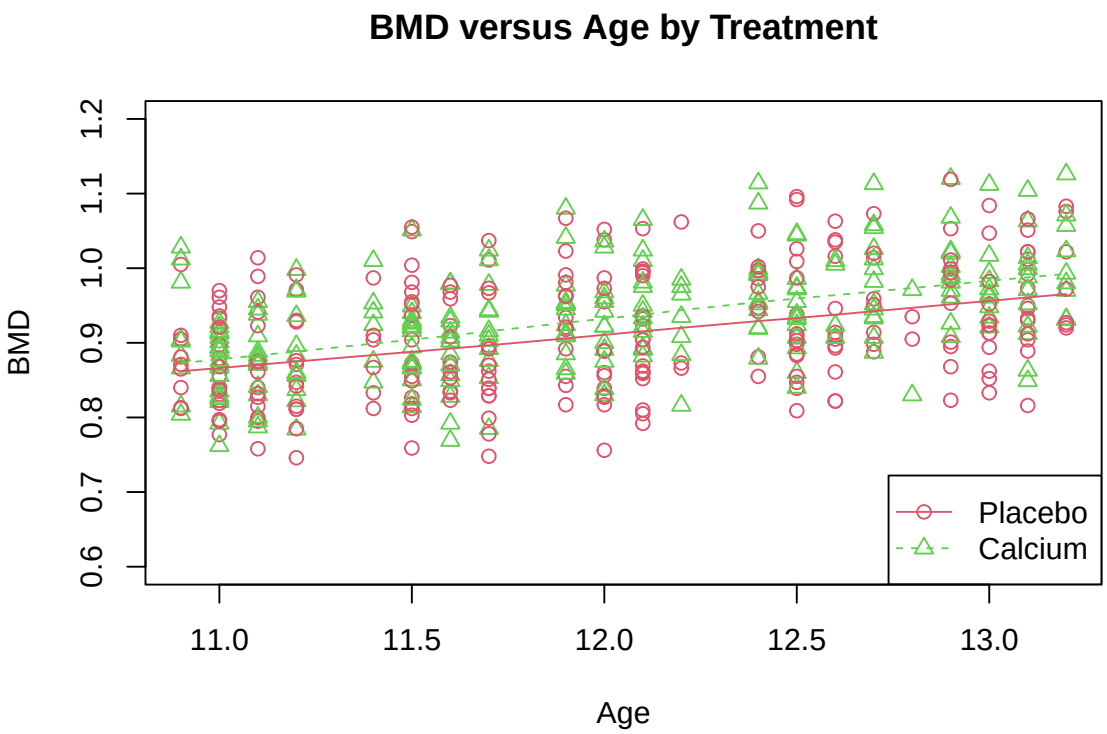
From the above table, we can see that it seems like the missingness pattern along the visits is similar for the Placebo and Calcium groups. Number of missing observations for both treatment groups increases by visit, with similar values.

5.
```
means = tapply(BoneLong$BMD, list(BoneLong$Trt,BoneLong$Visit), mean, na.rm=TRUE)
plot(c(1:5),means[1,],type="o",xlab="Visit Number", ylab="Mean Bone Mineral Density", ylim=c(0.8,1)
lines(c(1:5),means[2,],type="o",xlab="Visit Number", ylab="Mean Bone Mineral Density", lty=2,pch=2)
legend("bottomright",c("Placebo","Calcium"),lty=c(1,2),pch=c(1,2))
```



The plot "Mean BMD by Visit and Treatment" shows the mean Bone Mineral Density (BMD) for the Placebo and Calcium groups across five visits. Both groups show an increasing trend in BMD over the visits. However, the Calcium group consistently exhibits a higher mean BMD than the Placebo group at each visit, suggesting a potential increasing difference in BMD between the two groups.

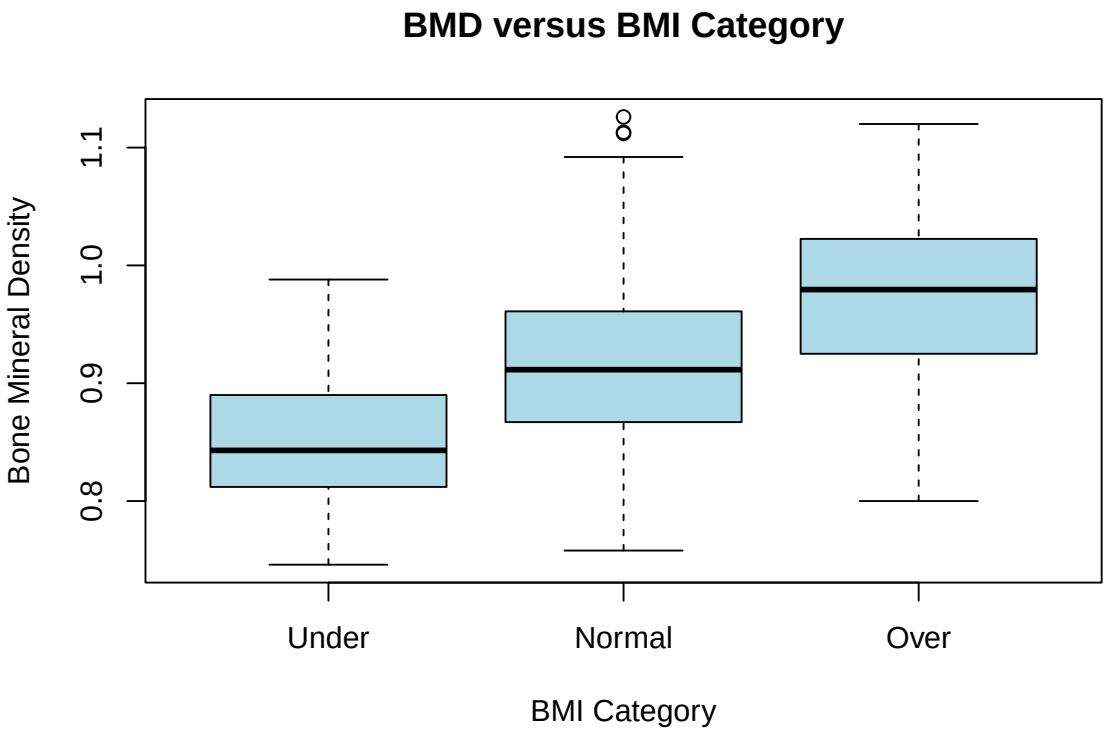
6.
```
plot(BMD ~ Age, data=BoneLong, pch=as.numeric(BoneLong$Trt), col=as.numeric(BoneLong$Trt)+1, main='
lines(lowess(BoneLong.comp$Age[BoneLong.comp$Trt=="P"], BoneLong.comp$BMD[BoneLong.comp$Trt=="P"]),
lines(lowess(BoneLong.comp$Age[BoneLong.comp$Trt=="C"], BoneLong.comp$BMD[BoneLong.comp$Trt=="C"]),
legend("bottomright", c("Placebo","Calcium"), lty=c(1,2), col=c(2,3), pch=c(1,2))
```



7.
- **Trends in Bone Mineral Density (BMD) with Age**: The plot "BMD versus Age by Treatment" shows an increasing trend in BMD with age for both the Placebo and Calcium groups. This suggests that as age increases, so does BMD. BMD for the calcium group remains higher than that of the placebo group with age.
- **Differences between Treatments**: The "Mean BMD by Visit and Treatment" plot and the summary statistics show that the Calcium group consistently has a higher mean BMD than the Placebo group across all visits. This suggests that the Calcium treatment may be more effective in increasing BMD than the Placebo.
- **Variability among Individuals**: The standard deviation values for BMD for each visit index by treatment indicate variability in BMD among individuals within each treatment group. The variability seems to be slightly higher in the Placebo group compared to the Calcium group. This could be due to individual differences in response to treatment, lifestyle factors, or other unmeasured variables.

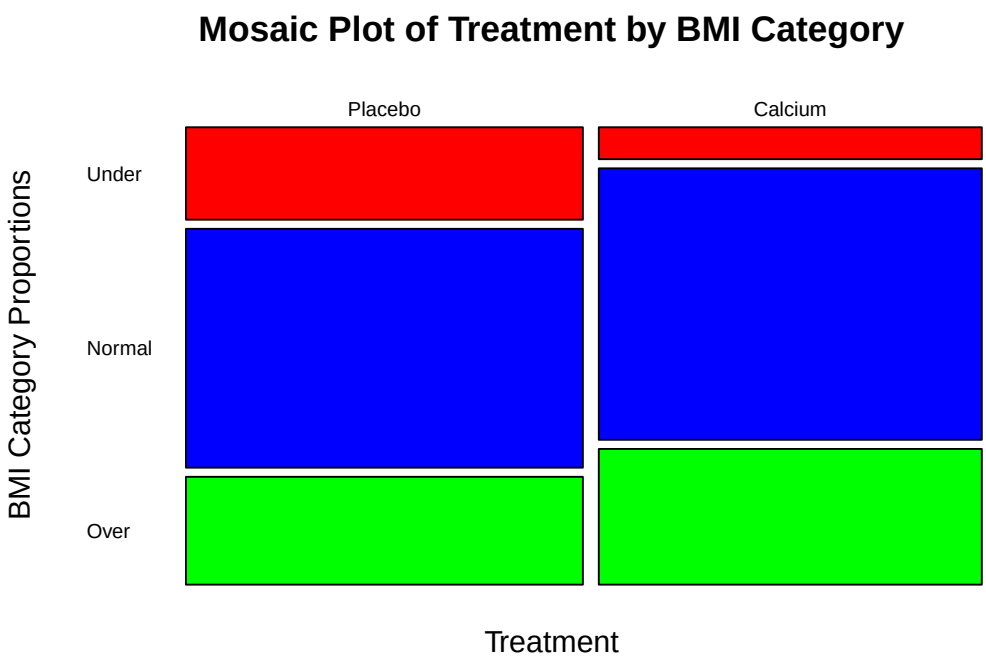
  The number of missing observations for each visit index by treatment suggests that data completeness may be an issue, especially in later visits. This could potentially affect the reliability of the observed trends and differences.

8.
```
plot(BMD ~ relevel(as.factor(BMICat),ref="Under"), data=BoneLong, main="BMD versus BMI Category",co
```



BMD versus BMI Category: This plot shows the distribution of BMD across different BMI categories - Under, Normal, and Over. It provides a visual representation of how BMD varies with BMI category. However, without specific details from the plot, it's difficult to provide a precise interpretation. Generally, such a plot could help identify if certain BMI categories are associated with higher or lower BMD.

```
mosaicplot( ~ factor(Trt, levels=c("P","C"), labels=c("Placebo","Calcium")) + relevel(as.factor(BMI
```



Mosaic Plot of Treatment by BMI Category: This plot shows the distribution of the two treatments (Placebo and Calcium) across different BMI categories. Each segment's size in the mosaic plot corresponds to the proportion of individuals in that category. This plot can help understand if there's a relationship between the treatment received and the BMI category of the individuals. For example, if larger segments are observed for the Calcium treatment in the "Over" BMI category, it could suggest that a higher proportion of individuals with "Over" BMI received the Calcium treatment.

9. Using only the complete cases, fit the generalized least squares model,
BMD ~ Trt+I(Visit-1)+Trt*I(Visit-1)
assuming a compound symmetry covariance structure with constant variance.
Note: We do I(Visit-1) to set our first visit (Visit=1) to be a baseline Visit 0. Also note: I(.) in R is not the indicator function, it is the identity function. What we are doing is taking Visits and subtracting 1 from them (so Visit=1 is the baseline Visit 0 now).

9a) When using the compound symmetry covariance structure in the generalized least squares model, we are assuming that the correlation of bone mineral density (BMD) across visits is constant and equal for all pairs of visits. This means that the correlation between any two BMD measurements is the same regardless of the time interval between them. In other words, the correlation between BMD at Visit 1 and Visit 2 is the same as the correlation between BMD at Visit 1 and Visit 3, Visit 1 and Visit 4, and so on. The compound symmetry structure imposes a homogeneous correlation structure on the BMD measurements, assuming that the temporal correlation remains constant throughout the study period. It simplifies the modeling of the correlation and allows for a single correlation parameter to be estimated for all pairs of visits.

On the other hand, if we were using an AR1 (auto regressive order 1) covariance structure, we would assume that the correlation between BMD measurements decreases as the time interval between visits increases. In this case, the correlation between BMD at Visit i and Visit j would be higher when the time interval |i - j| is smaller. The AR1 structure accounts for the temporal dependence between consecutive visits and allows for a more flexible modeling of the correlation structure that considers the ordering of the visits.

9b)

```
mod1 = gls(BMD ~ Trt+I(Visit-1)+Trt*I(Visit-1), correlation=corCompSymm(, form=~Visit | ID), metho
summary(mod1)
```

```
## Generalized least squares fit by REML
##   Model: BMD ~ Trt + I(Visit - 1) + Trt * I(Visit - 1)
##   Data: BoneLong.comp
##       AIC       BIC  logLik
##   -2046.5 -2021.831 1029.25
##
## Correlation Structure: Compound symmetry
##  Formula: ~Visit | ID
##  Parameter estimate(s):
##       Rho
## 0.9480044
##
## Coefficients:
##                        Value   Std.Error  t-value p-value
## (Intercept)       0.8699447 0.009751713 89.20942  0.0000
## TrtC              0.0123462 0.014024109  0.88036  0.3791
## I(Visit - 1)      0.0226894 0.000710605 31.92965  0.0000
## TrtC:I(Visit - 1) 0.0043243 0.001021933  4.23146  0.0000
##
##  Correlation:
##                   (Intr) TrtC   I(V-1)
## TrtC              -0.695
## I(Visit - 1)      -0.146  0.101
## TrtC:I(Visit - 1)  0.101 -0.146 -0.695
##
## Standardized residuals:
##        Min         Q1        Med         Q3        Max
## -2.1418111 -0.7471617 -0.1227180  0.6908222  2.4032626
##
## Residual standard error: 0.06756064
## Degrees of freedom: 455 total; 451 residual
```

Equation for the fitted mean bone mineral density:
BMD = B0 + B1 TrtC + B2 I(Visit-1) + B3 TrtC*I(Visit-1) + e
BMD = 0.87 + 0.012 TrtC + 0.023 I(Visit-1) + 0.0043 TrtCI(Visit-1)

In the equation,
BMD is the mean Bone Mineral Density. *TrtC* is the Treatment group (1 if Calcium, 0 if Placebo). *I(Visit-1)* is the Visit number minus 1, to set the first visit as the baseline (Visit 0). *TrtC*I(Visit-1)* is the interaction term that allows for the effect of the treatment on BMD to vary with the visit number. e is the error term.

The coefficients B1 and B3 represent the additional effect on BMD of the Calcium treatment at the baseline visit and at subsequent visits, respectively.

9c) Interpreting each of the estimated B coefficients, in context of the problem:

B0(intercept) is the estimated mean Bone Mineral Density (BMD) for the Placebo group (TrtC = 0) at the baseline visit (Visit 0).

B1(TrtC) is the estimated additional effect on the mean BMD for the Calcium group compared to the Placebo group at the baseline visit. In other words, at the baseline visit, the mean BMD for the

Calcium group is estimated to be 0.012 units higher than that for the Placebo group, all else being equal.

B2(I(Visit-1)) represents the estimated change in mean BMD for each unit increase in the visit number (after subtracting 1 to set the first visit as the baseline) for the Placebo group. In other words, for the Placebo group, the mean BMD is estimated to increase by 0.023 units for each subsequent visit.

*B3(TrtC*I(Visit-1))* is the estimated additional change in mean BMD for each unit increase in the visit number for the Calcium group compared to the Placebo group. In other words, for each subsequent visit, the mean BMD for the Calcium group is estimated to increase by an additional 0.0043 units compared to the Placebo group, all else being equal.

9d) BMD(Trt = Calcium, Visit = 1): $[0.87 + 0.012(1) + 0.023(0) + 0.0043(1)(0)] = 0.882$
BMD(Trt = Placebo, Visit = 1): $[0.87 + 0.012(0) + 0.023(0) + 0.0043(0)(0)] = 0.87$
Therefore, estimated difference in mean bone mineral density between the two treatments on the first visit: BMD(Trt = Calcium, Visit = 1) - BMD(Trt = Placebo, Visit = 1) = 0.012

BMD(Trt = Calcium, Visit = 5): $[0.87 + 0.012(1) + 0.023(4) + 0.0043(1)(4)] = 0.9912$
BMD(Trt = Placebo, Visit = 5): $[0.87 + 0.012(0) + 0.023(4) + 0.0043(0)(4)] = 0.962$

Therefore, estimated difference in mean bone mineral density between the two treatments on the fifth visit: BMD(Trt = Calcium, Visit = 5) - BMD(Trt = Placebo, Visit = 5) = 0.0292

9e) Test for if change in mean bone mineral density across visits differs between the two treatments.

Comparing compound symmetry models, with and without interaction term (Trt*I(Visit-1)),

```
mod1.ML = gls(BMD ~ Trt+I(Visit-1)+Trt*I(Visit-1), correlation=corCompSymm(, form=~Visit | ID), met
mod0.ML = gls(BMD ~ Trt+I(Visit-1), correlation=corCompSymm(, form=~Visit | ID), method="ML", data=

anova(mod0.ML, mod1.ML)
```

```
##         Model df       AIC       BIC   logLik   Test  L.Ratio p-value
## mod0.ML     1  5 -2071.033 -2050.432 1040.516
## mod1.ML     2  6 -2086.606 -2061.884 1049.303 1 vs 2 17.57312  <.0001
```

mod1.ML (model with interaction term) has lower AIC value and p-value of <.0001.

Comparing AR1 models, with and without interaction term (Trt*I(Visit-1)),

```
mod1.1.ML = gls(BMD ~ Trt+I(Visit-1)+Trt*I(Visit-1), correlation=corAR1(, form=~Visit | ID), methoc
mod0.1.ML = gls(BMD ~ Trt+I(Visit-1), correlation= corAR1(, form=~Visit | ID), method="ML", data=Bo

anova(mod0.1.ML, mod1.1.ML)
```

```
##           Model df       AIC       BIC   logLik   Test  L.Ratio p-value
## mod0.1.ML     1  5 -2200.650 -2180.048 1105.325
## mod1.1.ML     2  6 -2206.868 -2182.146 1109.434 1 vs 2 8.218095  0.0041
```

mod1.1.ML (model with interaction term) has lower AIC value and p-value of 0.0041.

Based on the tests and R outputs above, there is evidence that the change in mean bone mineral density (BMD) across visits differs between the two treatments. This is indicated by the p-value of <.0001 when comparing compound symmetry models with and without the interaction term (Trt*I(Visit-1)). This p-value is significantly less than the common significance level of 0.05, suggesting that the interaction term is statistically significant.

The interaction term represents the difference in the rate of change in BMD across visits between the two treatments. The positive coefficient for the interaction term in the model (B3 = 0.0043243)

suggests that the mean BMD increases at a faster rate for the Calcium group compared to the Placebo group across visits.

Therefore, the evidence suggests that the Calcium treatment is associated with a greater increase in BMD over time compared to the Placebo treatment.

9f) Estimated marginal variance-covariance matrix of the responses,

```
getVarCov(mod1)
```

```
## Marginal variance covariance matrix
##           [,1]      [,2]      [,3]      [,4]      [,5]
## [1,] 0.0045644 0.0043271 0.0043271 0.0043271 0.0043271
## [2,] 0.0043271 0.0045644 0.0043271 0.0043271 0.0043271
## [3,] 0.0043271 0.0043271 0.0045644 0.0043271 0.0043271
## [4,] 0.0043271 0.0043271 0.0043271 0.0045644 0.0043271
## [5,] 0.0043271 0.0043271 0.0043271 0.0043271 0.0045644
##   Standard Deviations: 0.067561 0.067561 0.067561 0.067561 0.067561
```

This matrix represents the estimated covariances between the BMD measurements at different visits. The diagonal elements (e.g., 0.0045644) represent the variances of the BMD measurements at each visit, while the off-diagonal elements (e.g., 0.0043271) represent the covariances between the BMD measurements at different visits.

The constant off-diagonal elements reflect the compound symmetry assumption, which assumes that the correlation between BMD measurements is the same for all pairs of visits. The standard deviations for all visits are 0.067561.

9g) Model comparison test (using anova function) to determine if the variance in bone mineral density differs across visits or we can use constant variance across visits.

```
mod2 = gls(BMD ~ Trt+I(Visit-1)+Trt*I(Visit-1), correlation=corCompSymm(, form=~Visit | ID), metho

anova(mod1, mod2)
```

```
##      Model df       AIC       BIC   logLik   Test  L.Ratio p-value
## mod1     1  6 -2046.500 -2021.831 1029.250
## mod2     2 10 -2057.456 -2016.341 1038.728 1 vs 2 18.95576   8e-04
```

Null hypothesis: The variance in bone mineral density (BMD) is constant across visits.
Alternate hypothesis: The variance in bone mineral density (BMD) differs across visits.

From the test and output above, we can see that mod2 (model with a weight function to account for heteroscedasticity, unequal variances, across visits) has a lower AIC value and p-value of 8e-04, indicating a better fit to the data compared to the model assuming constant variance across visits. Therefore, we reject the null hypothesis and conclude that there is evidence to suggest that the variance in BMD differs across visits.

10. Using all available data, that is to say all available observations that are not missing (use argument na.action = na.omit in the lme function), we now will fit a linear mixed effects model (LME) where:
    - the marginal mean of BMD varies with age,
    - each treatment can have a different marginal mean BMD trajectory both in intercept and in slope (with respect to age), so that is to say treatment has a fixed effect and an interaction effect with age,
    - the model is not adjusted for BMI category (that is to say BMI is not in the model in any way), and
    - subject-specific mean response trajectories can differ in intercept but not in slope (random intercept only).

10a) Theoretical notational form of the true conditional model:
Yij = B0 + B1 Trt(i) + B2 a(ij) + B3 Trt(i)*a(ij) + b_0i + eij

Yij is the bone mineral density (BMD) of the i-th individual on the jth visit. *Trt(i)* is the treatment group for subject i (1 if Calcium, 0 if Placebo). *a(ij)* is the age of the i-th individual on the j-th visit. *Trt(i)*a(ij)* is the interaction term between treatment and age for i-th individual on the j-th visit. *b_0i* is the random intercept for subject i. *eij* is the error term.

The fixed effects B0, B1, B2, and B3 are assumed to be constant across all subjects. B0 is the intercept, B1 is the effect of the Calcium treatment, B2 is the effect of age, and B3 is the interaction effect between the Calcium treatment and age.
The random effect b_0i is assumed to vary across subjects according to a normal distribution with mean 0 and variance sigmabˆ2.
The error term eij is assumed to be normally distributed with mean 0 and constant variance sigmaˆ2.
The random effects b_0i and the error terms eij are assumed to be independent.

10b) Fitted LME model,

```
mod3 = lme(BMD ~ Trt+Age+Trt*Age, random = ~ 1|ID, data=BoneLong, na.action=na.omit, method="ML")
summary(mod3)
```

```
## Linear mixed-effects model fit by maximum likelihood
##   Data: BoneLong
##         AIC       BIC   logLik
##   -2278.122 -2252.822 1145.061
##
## Random effects:
##  Formula: ~1 | ID
##         (Intercept)   Residual
## StdDev:  0.06653545 0.01501359
##
## Fixed effects:  BMD ~ Trt + Age + Trt * Age
##                  Value  Std.Error  DF  t-value p-value
## (Intercept)  0.3631525 0.01874861 387 19.36956  0.0000
## TrtC        -0.0863222 0.02683117 110 -3.21724  0.0017
## Age          0.0457723 0.00138325 387 33.09041  0.0000
## TrtC:Age     0.0088754 0.00198223 387  4.47751  0.0000
##  Correlation:
##          (Intr) TrtC   Age
## TrtC     -0.699
## Age      -0.880  0.615
## TrtC:Age  0.614 -0.881 -0.698
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -3.36572939 -0.59950246  0.02745521  0.62033724  2.68297920
##
## Number of Observations: 501
## Number of Groups: 112
```

BMD(ij) = 0.36 - 0.086 Trt(i) + 0.046 a(ij) + 0.0089 Trt(i)*a(ij) + b_0i

10c) From the R output in part a, the estimated standard deviation of the random intercepts is 0.06653545.
In the context of this problem, this value represents the variability in the baseline bone mineral density (BMD) across different subjects, after accounting for the effects of treatment and age. A larger standard

deviation would indicate greater variability in baseline BMD across subjects, while a smaller standard deviation would indicate less variability.
This suggests that while the treatment and age are important factors in determining BMD, there is still substantial individual variability in baseline BMD that is not explained by these factors. This could be due to other unmeasured factors or inherent individual differences.

10d) Likelihood ratio test to determine if a random slope on Age should be added to the model,

```
mod3.REML = lme(BMD ~ Trt+Age+Trt*Age, random = ~ 1|ID, data=BoneLong, na.action=na.omit, method="R

mod3s.REML = lme(BMD ~ Trt+Age+Trt*Age, random = ~ 1+Age|ID, data=BoneLong, na.action=na.omit, meth

anova(mod3.REML, mod3s.REML)
```

```
##             Model df       AIC       BIC   logLik   Test  L.Ratio p-value
## mod3.REML       1  6 -2240.336 -2215.085 1126.168
## mod3s.REML      2  8 -2335.238 -2301.569 1175.619 1 vs 2 98.90186  <.0001
```

Null hypothesis: the simpler model (without the random slope on Age) is sufficient.
Alternate hypothesis: the more complex model (with the random slope on Age) is better.

From the output above, mod3s.REML (model with the random slope on Age) has a lower AIC value and a p-value of <.0001. Therefore, we reject the null hypothesis and conclude that the more complex model with a random slope on Age provides a significantly better fit to the data than the simpler model without the random slope on Age.

10e) Likelihood ratio test to determine if we should adjust for BMI category in our analysis (fixed effect),

```
mod4.0 = lme(BMD ~ Trt+Age+Trt*Age + BMICat, random = ~ 1 + Age|ID, data=BoneLong, na.action=na.omi

mod4.1 = lme(BMD ~ Trt+Age+Trt*Age, random = ~ 1 + Age|ID, data=BoneLong, na.action=na.omit, method

anova(mod4.0, mod4.1)
```

```
##         Model df       AIC       BIC   logLik   Test L.Ratio p-value
## mod4.0      1 10 -2408.607 -2366.441 1214.304
## mod4.1      2  8 -2371.392 -2337.659 1193.696 1 vs 2 41.2159  <.0001
```

Null hypothesis: the simpler model (without adjusting for BMI category) is sufficient.
Alternate hypothesis: the more complex model (adjusting for BMI category) is better.

From the output above, mod4.0 (model with BMICat) has a lower AIC value, and the test has a p-value of <.0001. Therefore, we reject the null hypothesis and conclude that the more complex model adjusting for BMI category provides a significantly better fit to the data than the simpler model without adjusting for BMI category.

10f) When we're working with a Linear Mixed Effects (LME) model, the residuals (differences between our observed and predicted values) aren't independent from each other. They have a certain pattern or structure because of the random effects we've included in our model. This pattern can make it difficult to check if our model meets certain assumptions, like having a constant variance (homoscedasticity) and following a normal distribution.

However, we can transform, or change, the residuals. This process helps in getting rid of the influence of the random effects and gives us "normalized" or "standardized" residuals. These transformed residuals have a constant variance, which makes it easier to check if our model meets its assumptions. This step is really important because it helps in making sure that the conclusions we draw from our model are valid.

11. The study has a few limitations. The study was conducted on a volunteer sample of 112 adolescent girls recruited from local school districts in primarily white middle-class neighborhoods. This limits the generalizability of the results to other demographic groups. The sample may not be representative of the broader population, including different ethnicities, socioeconomic statuses, or males. Secondly, 21 participants dropped out before the study ended, which could skew the results if they left because of the treatment or the outcome.

    The study design could also be improved by having a randomized block design or a stratified random design. In a randomized block design, participants would be grouped based on certain characteristics before being randomly assigned to a treatment. This could make the results more precise. A stratified random design would ensure the sample represents key demographic variables. For example, the researchers could stratify by age, ethnicity, and socioeconomic status, and then randomly sample within each stratum.

    Lastly, to address issue of loss to follow-up, the researchers could use strategies like regular check-ins, flexible schedules, and incentives to keep participants in the study. They could also use statistical methods designed to handle missing data, such as multiple imputation or full information maximum likelihood.

12. The best LME model from part 10 (model adjusting for BMI category, along with a random intercept and random slope on Age) takes into account the treatment, age, the interaction between treatment and age, and BMI category. It also considers the random effects of age and the individual subjects (ID).

    In context of the study, this model shows that calcium supplementation (Trt) has a significant positive effect on bone mineral density (BMD) among adolescent girls. The interaction term (Trt*Age) suggests that the effect of calcium supplementation on BMD changes with age. This indicates that the positive effect of calcium supplementation on BMD becomes more pronounced, as the girls grow older.

    The model also adjusts for BMI category, suggesting that the effect of calcium supplementation on BMD is independent of the girls' BMI. This is important as it shows that the benefits of calcium supplementation apply to girls across different BMI categories.

    Additionally, the random effects in the model account for individual differences among the girls and the potential influence of age on BMD. This ensures that the model captures the variability in BMD not only due to the treatment but also due to individual differences and age.

    Therefore, the LME model and study provides evidence supporting the use of calcium supplementation for improving bone health in adolescent girls.