Stat 526 **Assignment #10**

Linear Mixed Models

READING - Faraway Chapters 10, 11 and 13

0. (5 pts) Name

1. (8 pts) A school superintendent is concerned about the development of technology skills for research in high school. Since there are four high schools in his district, all of which go about this instruction differently, he decided to assess if there were any differences across schools. He first compiled a long list of "tech skills" and randomly selected four to be used in his study. He then randomly selected 24 students from each school and assigned each to one of the four tasks so that there were six students per task per school. Each student then performed the skill and was scored using a 0-100 numeric scale.

a) If a two-way ANOVA is to be used for the analysis, should it be treated as a fixed effects, random effects, or mixed effects model? State the model you would use and its assumptions.

I'd consider "tech skills" to be random because there is a population of skills and the ones in the study were randomly chosen. I'd consider school to be fixed since there are only these four schools. The model is

$$Y_{ijk} = \mu + \text{School}_i + \text{Task}_j + (\text{SchoolTask})_{ij} + \varepsilon_{ijk}$$

where

$$\text{Task}_i \sim N(0, \sigma_T), \quad (\text{SchoolTask})_{ij} \sim N(0, \sigma_{ST}), \quad \varepsilon_{ijk} \sim N(0, \sigma)$$

and the random terms are all independent of each other.

b) Complete the following ANOVA table and determine which effects are significant at the $\alpha = .05$ level. State your conclusions, making sure to estimate all variances and describing any additional mean comparisons you'd like to perform.

Source	DF	SS	MS	F
School	3	222.0	74.000	4.93
Task	3	96.0	32.000	2.13
School × Task	9	135.0	15.000	2.67
Error	80	450.0	5.625	

1. School effect :
   - Hypothesis : $H_0 : \text{School}_1 = \cdots = \text{School}_4 = 0$. Because $F^* > F_{(0.95,3,9)} = 3.86$, reject H_0 and conclude that there are school effects.
2. Task variance :
   - Hypothesis : $H_0 : \sigma_T^2 = 0$. Because $F^* < F_{(0.95,3,9)} = 3.86$, we do not reject H_0 and conclude that there is no variance among the tasks.
3. Interaction variance :
   - Hypothesis : $H_0 : \sigma_{ST}^2 = 0$. Because $F^* > F_{(0.95,9,80)} = 2.00$, we conclude that there is significant interaction variance.

Estimates of the non-zero variances are: $\hat{\sigma}^2 = MSE = 5.625$ and $\hat{\sigma}^2_{ST} = \frac{MSAB-MSE}{n} = \frac{15-5.625}{6} = 1.5625$. These will change slightly if we remove the random task effect from the model and refit. Currently, its estimate is $\frac{32-15}{24} = 0.708$. Because the School effects were found significantly different, I'd also want to compare the school means to see where the differences lie.

c) If the grand skill level of the high schools (average over the four schools) is of interest, describe how one would construct a 95% confidence interval.

Let μ be the overall mean. Its estimate is the grand sample mean, whose expected value can be expressed in terms of the model parameters

$$\hat{\mu} = \frac{1}{96}\sum_{i=1}^{3}\sum_{j=1}^{5}\sum_{k=1}^{4} y_{ijk} \rightarrow \mu + \frac{1}{4}\sum_{j=1}^{4}\text{Task}_j + \frac{1}{16}\sum_{i=1}^{4}\sum_{j=1}^{4}(\text{SchoolTask})_{ij} + \frac{1}{96}\sum_{i=1}^{4}\sum_{j=1}^{4}\sum_{k=1}^{6}\varepsilon_{ijk}$$

indicating that

$$\hat{\mu} \sim N(\mu, \frac{1}{4}\sigma^2_T + \frac{1}{16}\sigma^2_{ST} + \frac{1}{96}\sigma^2)$$

$s^2\{\hat{\mu}\} = \frac{1}{4}\cdot\frac{17}{24} + \frac{1}{16}\cdot\frac{9.375}{6} + \frac{1}{96}\cdot 5.625 = \frac{1}{3}$. This is also equal to $\sqrt{MSTask/96}$. Thus, we can use this standard error and 3 degrees of freedom to construct the 95% confidence interval for μ.

2. (15 pts) Nurse practitioners (NPs) are becoming the health partner choice for many Americans. A group of researchers study the performance of NPs in three specialties (neonatal care, women's health and oncology). They randomly selected four cities, and recorded competency scores of four nurses randomly selected within each specialty and city. The scores are on a continuous scale and shown below:

	City 1		City 2		City 3		City 4		Mean
Neonatal	71.5	58.9	68.5	64.8	59.1	67.1	77.2	75.2	
	72.9	67.9	71.2	74.2	62.2	62.5	84.7	67.3	69.075
Women's	83.8	76.9	70.5	65.6	71.0	75.7	72.4	81.6	
	73.2	79.3	74.0	78.2	63.5	65.0	79.6	81.2	74.469
Oncology	77.0	82.7	80.4	79.6	62.3	81.6	91.5	89.4	
	90.4	85.3	66.5	79.8	64.3	88.6	84.5	94.3	81.138
	76.650		72.775		68.575		81.575		74.894

a) State the linear mixed model that is appropriate for these data, as well as all the model assumptions. Also make sure to specify why each factor in your model is either random or fixed.

This study involves a 3×4 factorial structure with $n = 4$. The researchers are interested in three specialties (S) so I assume that specialty is a fixed factor. The cities (C) were randomly chosen so I'd consider that a factor random. The model is

$$Y_{ijk} = \mu + \text{S}_i + \text{C}_j + (\text{SC})_{ij} + \varepsilon_{ijk}$$

where

$$\text{C}_j \sim N(0, \sigma_C), (\text{SC})_{ij} \sim N(0, \sigma_{SC}) \text{ and } \varepsilon_{ijk} \sim N(0, \sigma)$$

The random terms are all independent of each other.

b) Provide the estimates of the fixed effects in your model. Make sure to specify the parametrization restriction you are using.

Estimates will depend on the choice of parameter restriction. Here, I assume the sum of specialty effects add to zero. Given that the study is balanced, the estimate $\hat{\mu} = 74.894$ is the grand sample mean and the specialty effect estimates are $\hat{S}_1 = 69.075 - 74.894 = -5.819$, $\hat{S}_2 = 74.469 - 74.894 = -0.425$, and $\hat{S}_3 = 81.138 - 74.894 = 6.244$.

c) Using the partial R output below, perform the appropriate F tests for the main effects and interaction. For each test, make sure to specify the null and alternative hypotheses.

```
> aov(y ~ city*spec, exer)
Call:
   aov(formula = y ~ city * spec, data = exer1)
Terms:
                     city      spec city:spec Residuals
Sum of Squares  1105.6706 1168.3663  156.8438 1429.8875
```

1. Specialty effect :
   - Hypothesis : $H_0 : \mathrm{S}_1 = \mathrm{S}_2 = \mathrm{S}_3 = 0$. $F^* = (1168.3663)/2)/(156.8438/6) = 22.3$. Because $F^* > F_{(0.95,2,6)} = 5.14$, reject H_0 and conclude that there are specialty effects.
2. City variance :
   - Hypothesis : $H_0 : \sigma_C^2 = 0$. $F^* = (1105.6706/3)/(156.8438/6) = 14.1$. Because $F^* > F_{(0.95,3,6)} = 4.76$, we reject H_0 and conclude that there is variance among the cities.
3. Interaction variance :
   - Hypothesis : $H_0 : \sigma_{SC}^2 = 0$. $F^* = (156.8438/6)/(1429.8875/36) = 0.66$. Because $F^* < F_{(0.95,6,36)} = 2.36$, we conclude that there is no significant interaction variance.

d) Estimate the error variance and any other variances found significantly different from zero in the previous part.

Estimates of the non-zero variances are: $\hat{\sigma}^2 = MSE = 1429.8875/36 = 39.72$ and $\hat{\sigma}_C^2 = \frac{MSC - MSSC}{bn} = \frac{1105.6706/3 - 156.8438/6}{12} = 28.53$. These would change slightly if we removed the random interaction effect from the model (set it equal to 0). Its current estimate is -3.3946 using the ANOVA table.

e) Using the table summary and previous calculations, test whether there is a difference between the average competency score for Neonatal care and for Women's health. Use the 0.05 significance level. There is no need for an multiple comparison adjustment.

In this balanced design, the SE for a difference between two specialty means is $\sqrt{(2MS_{SC}/(4*4))} = \sqrt{2(26.1406)/16} = 1.808$. Performing a t test, we get

$$t^* = \frac{74.469 - 69.075}{1.808} = 2.98$$

This is larger than $t_{0.05,6} = 2.447$ so we reject and conclude these two means are different.

f) Compute a 95% confidence interval for the average competency among the RN's.

In this balanced situation, the grand mean uses the MSCity in the calculation of the SE. Thus the confidence interval is

$$\begin{aligned} \hat{\mu} &\pm t_{0.05,3}\sqrt{MSCity/48} \\ 74.894 &\pm 3.182\sqrt{368.5569/48} \\ 74.894 &\pm 8.817 \end{aligned}$$

g) Use your variance estimates from part d) to estimate 1) the correlation between two observations from the same city and specification, 2) the correlation between two observations from the same city but different specifications, and 3) the correlation between two observations from the same specification but different cities.

Two observations from the same city and specification will have covariance $\sigma_C^2 + \sigma_{SC}^2$, two observations from the same city but different specifications will have covariance σ_C^2, and two observations from the same specification but different cities will have covariance 0. This means the correlations are estimated to be

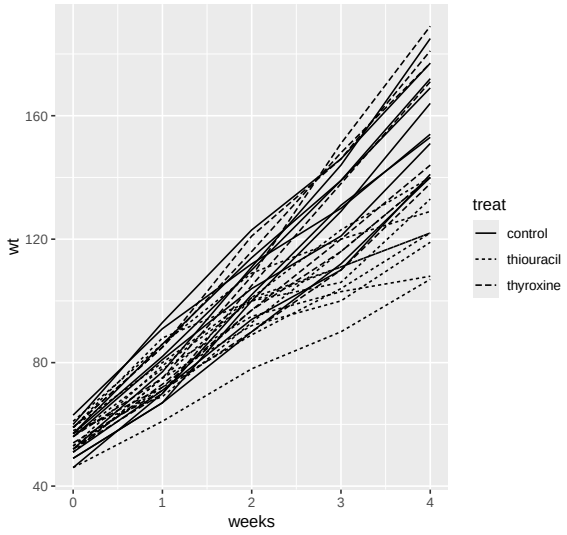
$$r_1 = \frac{28.53+0}{28.53+0+39.72} = 0.418 = r_2 \text{ and } r_3 = 0$$

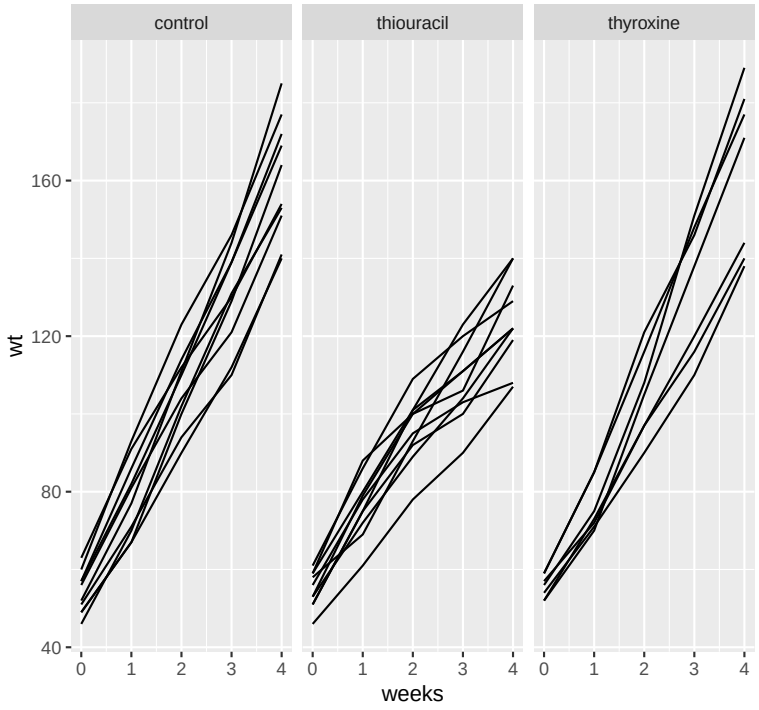
or

$$r_1 = \frac{28.53-3.39}{28.53-3.39+39.72} = 0.388, r_2 = \frac{28.53}{28.53-3.39+39.72} = 0.440, \text{ and } r_3 = 0$$

3. (12 pts) Faraway Chapter 11 Exercise 1

(a) From the plots it looks like we have a fairly linear trend over weeks but the rate of growth for mice in the thiouracil group seems slower.





(b) The following output summarizes the random coefficients model:

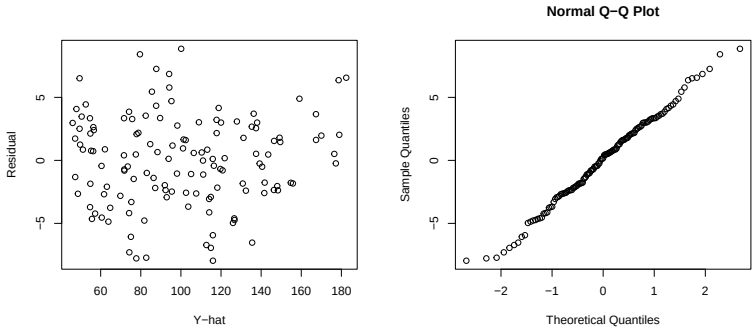
```
REML criterion at convergence: 878.7

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 subject  (Intercept) 32.50    5.700
          weeks       14.14    3.760    -0.13
 Residual             18.90    4.348
Number of obs: 135, groups:  subject, 27

Fixed effects:
                     Estimate Std. Error      df t value Pr(>|t|)
(Intercept)           52.8800     2.0938 23.9993  25.256  < 2e-16 ***
weeks                 26.4800     1.2661 23.9984  20.915  < 2e-16 ***
treatthiouracil        4.7800     2.9610 23.9993   1.614     0.12
treatthyroxine        -0.7943     3.2629 23.9993  -0.243     0.81
weeks:treatthiouracil -9.3700     1.7905 23.9984  -5.233 2.31e-05 ***
weeks:treatthyroxine   0.6629     1.9731 23.9984   0.336     0.74
```

Because of the restrictions, the intercept term represents the population intercept for the rats in the control group. The interaction term for thiouracil and weeks represents the difference in population slopes between the control mice and thiouracil mice. Finally the intercept random effect SD is describing the variability in intercepts among the mice in a particular treatment.

c) The following ANOVA results show there is a significant difference among the slopes of the three groups. Given that there are different slopes, there is no need to compare intercepts as the differences between lines are not constant.

```
Type III Analysis of Variance Table with Satterthwaite's method
             Sum Sq Mean Sq NumDF  DenDF  F value    Pr(>F)
weeks       17209.2 17209.2     1 23.998 910.3211 < 2.2e-16 ***
treat          71.9    36.0     2 23.999   1.9027     0.171
weeks:treat   692.6   346.3     2 23.998  18.3181 1.479e-05 ***
```

d) The residuals plots look quite reasonable suggesting the assumptions regarding the errors is reasonable.



e) Confidence intervals based on the bootstrap method are shown below. It appears the covariance between the mouse-specific slope and intercept could be zero.

```
                            2.5 %     97.5 %
.sig01                  3.6391893  7.9517699
.sig02                 -0.5739647  0.3933097
.sig03                  2.4647394  5.0482854
.sigma                  3.6055153  5.0939648
(Intercept)            48.8566035 57.0808232
weeks                  24.0628816 28.7067374
treatthiouracil        -0.8910029 10.3918832
treatthyroxine         -7.0401154  6.0875598
weeks:treatthiouracil -12.6413934 -5.5307281
weeks:treatthyroxine   -3.1733891  4.1064249
```

To assess if the thyroxine group is significantly different from the control, we create a new trt variable that takes the value 1 if thyroxine or control and 2 otherwise. We can then perform a likelihood ratio test or F test between the two models. The output below shows there is little evidence of a difference.

```
F-test with Kenward-Roger approximation; computing time: 0.16 sec.
large : wt ~ weeks + treat + (1 + weeks | subject) + weeks:treat
small : wt ~ weeks * trt + (1 + weeks | subject)
         stat     ndf     ddf F.scaling p.value
Ftest  0.0671  2.0000 23.0000   0.95833  0.9353
```

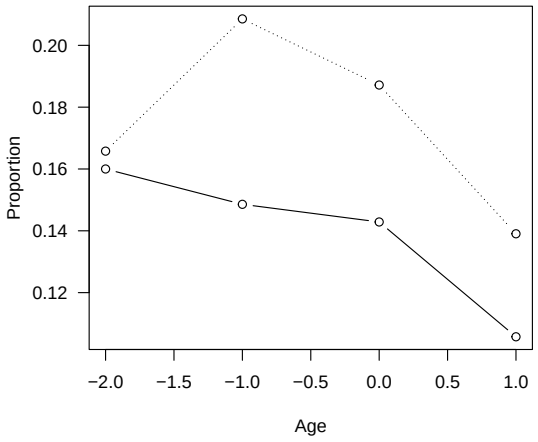
4. (10 pts) Faraway Chapter 13 Exercise 1

a) I believe the data do not provide this information. The variable `smoke` indicates the mother's status at the start of the study. If it instead represents the status at the beginning of each year, then the answer is no, none of the mothers change their status. There are 350 children who's mother did not smoke and 187 children with mothers who did.

b) The following table breaks down the distribution of "wheeze years" by smoking status. There are a slightly higher proportion of 0's and 1's in the non-smoking and slightly higher proportions of 2's, 3's, and 4's in the smoking group.

```
     times
smoke          0          1          2          3          4
    0 0.67714286 0.18571429 0.07142857 0.03428571 0.03142857
    1 0.63101604 0.17112299 0.10160428 0.05882353 0.03743316
```

c) Breaking things down this way, we find the proportion of children wheezing is always higher in the smoking group, especially when the children were 8 and 9 years of age.



d) Fitting a binomial to the 537 children using "wheeze years" as the response, we get the following output:

```
 Call: glm(cbind(V1, 4-V1) ~ smoke, family ="binomial", data=ohioc)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.82124    0.07719 -23.595   <2e-16 ***
smoke        0.27156    0.12334   2.202   0.0277 *

    Null deviance: 1045.3  on 536  degrees of freedom
Residual deviance: 1040.5  on 535  degrees of freedom
AIC: 1337.9
```

We can see that the deviance suggests poor fit. With $m = 4$ for each observation, asymptotics may be questionable. However, we know that these counts do not arise from Bernoulli trials as they 1) are all from the same subject (not independent) and 2) possibly varying p across years. Given the poor fit of the model, we cannot draw any definitive conclusions regarding maternal smoking. If, for example, we considered the quasibinomial distribution, the statistical significance of the smoking effect would disappear.

e) If we account for the possible correlation among observations and allow for the probability to vary over time and smoking status, we get the following output:

```
Linear mixed-effects model fit by maximum likelihood

Random effects:
 Formula: ~1 | id
        (Intercept)  Residual
StdDev:    2.058848 0.6352806

Fixed effects: resp ~ smoke * age
                 Value  Std.Error   DF    t-value p-value
(Intercept) -2.7946368 0.14450555 1609 -19.339305  0.0000
smoke        0.3932561 0.23805479  535   1.651956  0.0991
```

```
age         -0.2237722 0.05588650 1609  -4.004048  0.0001
smoke:age    0.1089708 0.08961638 1609   1.215970  0.2242
Number of Observations: 2148
Number of Groups: 537
```

The interaction is not significant so we can look at the main effects. There is a general decreases of roughly 20% ($\exp\{-0.224\}$) in the odds of wheezing each year ($P = 0.0001$) and the smoking group has greater odds of wheezing although that is not significant.

f) If we use Gaussian quadrature, the results are quite similar in terms of significance and parameter estimates.

```
Formula: resp ~ smoke * age + (1 | id)

     AIC      BIC   logLik deviance df.resid
  1604.7   1633.1   -797.4   1594.7     2143

Random effects:
 Groups Name        Variance Std.Dev.
 id     (Intercept) 4.694    2.167
Number of obs: 2148, groups:  id, 537

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -3.12840    0.22286 -14.038   <2e-16 ***
smoke        0.46204    0.28556   1.618   0.1057
age         -0.21637    0.08656  -2.500   0.0124 *
smoke:age    0.10533    0.13849   0.761   0.4469
```

i) The probabilities are relatively small for this data set so the distributions are skewed to the right. This means the GEE results should find generally larger probabilities than with GLMM and that is played out. Notice how much larger the intercept is for the GEE results.

```
 Call:
geeglm(formula = resp ~ smoke * age, family = binomial, data = ohio,
    id = id, corstr = "ar1")

 Coefficients:
            Estimate  Std.err    Wald Pr(>|W|)
(Intercept) -1.92477  0.12070 254.312   <2e-16 ***
smoke        0.28877  0.19138   2.277   0.1313
age         -0.14780  0.05984   6.101   0.0135 *
smoke:age    0.08355  0.09167   0.831   0.3621

Correlation structure = ar1
Estimated Scale Parameters:
            Estimate Std.err
(Intercept)    1.022  0.1254
  Link = identity

Estimated Correlation Parameters:
      Estimate Std.err
alpha   0.4914 0.06799
Number of clusters:   537  Maximum cluster size: 4
```

The general conclusions here are similar to the GLMMs. The age effect is not as strong here but reasonably close. We can see that the AR(1) correlation is roughly 0.5, suggesting someone already wheezing is likely to continue to wheeze.

j) Overall, we should trust the GLMM/GEE results more than the GLM results. The GLM model does not fit well and the results from GLMM explain why. There are changes in the probability of wheezing by age and the observations are correlated.

```
library(faraway)
library(ggplot2)
ggplot(ratdrink,aes(x=weeks, y=wt, linetype=treat, group=subject))+geom_line()
ggplot(ratdrink,aes(x=weeks, y=wt, group=subject))+geom_line() + facet_wrap(~treat)

library(lmerTest)
mod3 = lmer(wt~weeks*treat + (1+weeks|subject),ratdrink)
summary(mod3)
anova(mod3)

plot(fitted(mod3),residuals(mod3),xlab="Y-hat", ylab="Residual")
qqnorm(residuals(mod3))

confint(mod3,method="boot")

mod3ml = lmer(wt~weeks*treat + (1+weeks|subject),REML=F,ratdrink)
ratdrink$trt = factor(ifelse(ratdrink$treat=="thyroxine" | ratdrink$treat=="control",1,2))
mod4ml = lmer(wt~weeks*trt + (1+weeks|subject),REML=F,ratdrink)

library(pbkrtest)
KRmodcomp(mod3ml,mod4ml)


##### Problem #4
library(dplyr)
ohio1 = ohio %>% group_by(id,smoke) %>% summarize(times=sum(resp)) %>% xtabs(formula=~smoke+times)
prop.table(ohio1,1)

ohio %>% group_by(id,smoke) %>% summarize(times=sum(resp)) %>% xtabs(formula=~times+age+smoke)

###Create time plot - the crude way
ohio2 = xtabs(~age+resp+smoke,ohio)
x1=c(prop.table(ohio2[,,1],1)[,2])
x2=c(prop.table(ohio2[,,2],1)[,2])
plot(c(-2,-1,0,1),x1,type="b",las=1,ylim=c(min(x1,x2),max(x1,x2))
     ,xlab="Age",ylab="Proportion")
lines(c(-2,-1,0,1),x2,lty=3,type="b")

####Get data set in binomial format
library(plyr)
ohioc = ddply(ohio, .(id,smoke), function(x) sum(x[,1]))
mod1 = glm(cbind(V1,4-V1) ~ smoke,family="binomial",ohioc)
summary(mod1)

###Fitting a GLMM
library(MASS)
mod2a = glmmPQL(resp~smoke*age, random=~1|id, family=binomial,ohio)
summary(mod2a)

library(lme4)
mod2b = glmer(resp~ smoke*age+(1|id), nAGQ=25, family=binomial,ohio)
summary(mod2b)

library(geepack)
mod3 = geeglm(resp~smoke*age,id=id, family=binomial,corstr="ar1",ohio)
summary(mod3)
```