Code Appendix

Zirui Zhang

```r
library(tidyverse)
library(ggplot2)
library(patchwork)
library(nlme)
library(lme4)
library("car")
library(geepack)
```

```r
load("~/Documents/2023Fall/P8157/P8157/WtLoss.RData")
data = wtloss
p0 = ggplot(data, aes(x = time, y = weight, group = id, color = as.factor(diet))) +
  geom_line() +
  facet_grid(~diet) +
  theme_classic()
d0 = data |>
  mutate(diet = as.factor(diet))
```

```r
fit1 = lme(fixed = weight ~ diet*time, random=reStruct(~ 1 | id), data=data, method="ML")
fit2 = lme(fixed = weight ~ diet*time, random=reStruct(~ time | id), data=data, method="ML")
# fixed effect
fixed = data.frame(
  coef.fit1 = c(summary(fit1)$coefficients$fixed),
  sd.fit1 = c(sqrt(diag(summary(fit1)$varFix))),
  coef.fit2 = c(summary(fit2)$coefficients$fixed),
  sd.fit2 = c(sqrt(diag(summary(fit2)$varFix)))
)
rownames(fixed) = c("Intercept, b0", "Main effect for diet, b1",
                    "Main effect for time, b2", "Interaction, b3")
colnames(fixed) = c("Est.fit1", "SE.fit1","Est.fit2", "SE.fit2")
# random effect
random = data.frame(
  ran.fit1 = c(as.numeric(VarCorr(summary(fit1))[1,2]), NA, summary(fit1)$sigma),
  ran.fit2 = c(as.numeric(VarCorr(summary(fit2))[1,2]), as.numeric(VarCorr(summary(fit2))[2,2]), summary
)
rownames(random) = c("SD of random intercepts", "SD of random slope",
                     "SD of errors")
colnames(random) = c("fit1", "fit2")
```

```r
sim = simulate(fit1, nsim = 1000, seed = 1504, fit2, method = "ML")
lrt = data.frame(stat = -2*(-sim$alt$ML[, "logLik"]+sim$null$ML[, "logLik"]))
data1 = data.frame(x = rchisq(1000, df = 1))
data2 = data.frame(x = rchisq(1000, df = 2))
p1 = ggplot() +
```

```r
  geom_histogram(data = data1, aes(x = x, y = ..density..),
                 bins = 100, fill = "blue", alpha = 0.5) +
  geom_histogram(data = data2, aes(x = x, y = ..density..),
                 bins = 100, fill = "green", alpha = 0.5) +
  geom_histogram(data = lrt, aes(x = stat, y = ..density..),
                 bins = 100, fill = "black", alpha = 0.8) +
  labs(title = "Chi-Square Distributions",
       x = "Value",
       y = "Density") +
  theme_minimal()
```

```r
# residuals - stage 1 and random intercept
epsHat = data.frame(eps = resid(fit1, type="normalized"))
gammaHat = data.frame(gam = ranef(fit1)[,1])
epsHat$diet = as.factor(data$diet)
epsHat$time = as.factor(floor(data$time/2))
# box plot mean model - diet - stage 1
p2 = ggplot(epsHat, aes(x = diet, y = eps)) + geom_boxplot()
# box plot mean model - time - stage 1
p3 = ggplot(epsHat, aes(x = time, y = eps)) + geom_boxplot()
# scatterplot for dependence model - stage 1
p4 = epsHat |> mutate(time = as.numeric(time)) |> filter(time > 3) |>
  ggplot(aes(x = lag(eps), y = eps)) + geom_point() +geom_smooth()
```

```r
fit.I = geeglm(scale(weight) ~ diet*time, id=id, data, family=gaussian, scale.fix=TRUE, corstr="independ
fit.E = geeglm(weight ~ diet*time, id=id, data, family=gaussian, scale.fix=TRUE, corstr="exchangeable")
fit.AR = geeglm(weight ~ diet*time, id=id, data, family=gaussian, scale.fix=TRUE, corstr="ar1")
est = data.frame(
  est = c(summary(fit.I)$coefficients[1][[1]][2:4],
          summary(fit.E)$coefficients[1][[1]][2:4],
          summary(fit.E)$geese$correlation[1,1],
          summary(fit.AR)$coefficients[1][[1]][2:4],
          summary(fit.AR)$geese$correlation[1,1]),
  se = c(summary(fit.I)$coefficients[2][[1]][2:4],
         summary(fit.E)$coefficients[2][[1]][2:4],
         summary(fit.E)$geese$correlation[1,2],
         summary(fit.AR)$coefficients[2][[1]][2:4],
         summary(fit.AR)$geese$correlation[1,2])
)
rownames(est) = c("GEE-I: diet", "GEE-I: time", "GEE-I: time*diet",
                  "GEE-E: diet", "GEE-E: time", "GEE-E: time*diet", "GEE-E: rho",
                  "GEE-AR1: diet", "GEE-AR1: time", "GEE-AR1: time*diet", "GEE-AR1: ")
colnames(est) = c("Est", "SE")
```

0.1 Question (a)

p0
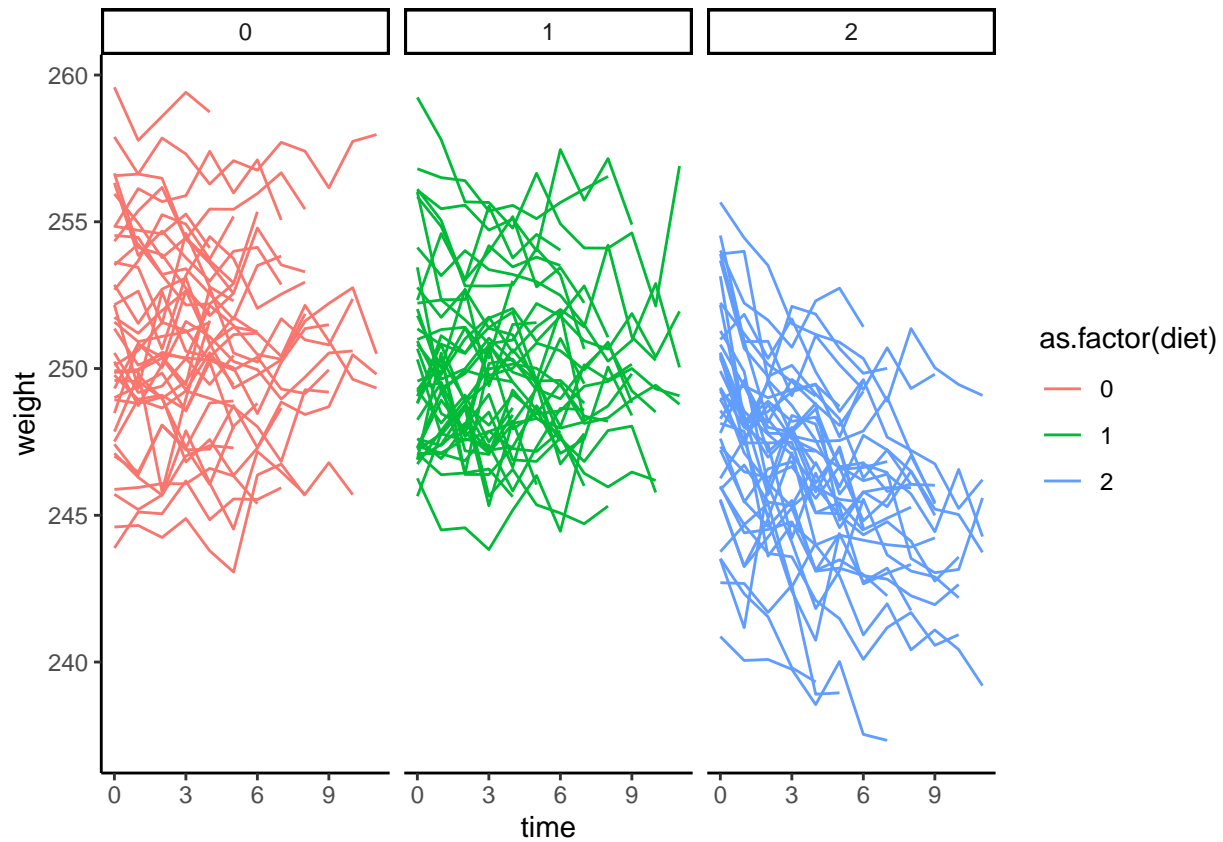

```
knitr::kable(fixed, format = "markdown")
```

	Est.fit1	SE.fit1	Est.fit2	SE.fit2
Intercept, b0	251.1588355	0.4464424	251.1989578	0.4659681
Main effect for diet, b1	-1.3191012	0.3456541	-1.3351304	0.3607745
Main effect for time, b2	0.0894331	0.0221724	0.0713879	0.0342238
Interaction, b3	-0.2556355	0.0164954	-0.2503117	0.0259901

```
knitr::kable(random, format = "markdown")
```

|	fit1	fit2
SD of random intercepts	3.018501	3.1652455
SD of random slope	NA	0.1702491
SD of errors	1.111389	1.0008228

0.2 Question (b)

$$H_0 : G(a) = \begin{bmatrix} \Sigma_{\gamma,00} & 0 \\ 0 & 0 \end{bmatrix}$$

$$H_1 : G(a) = \begin{bmatrix} \Sigma_{\gamma,00} & \Sigma_{\gamma,01} \\ \Sigma_{\gamma,10} & \Sigma_{\gamma,11} \end{bmatrix}$$
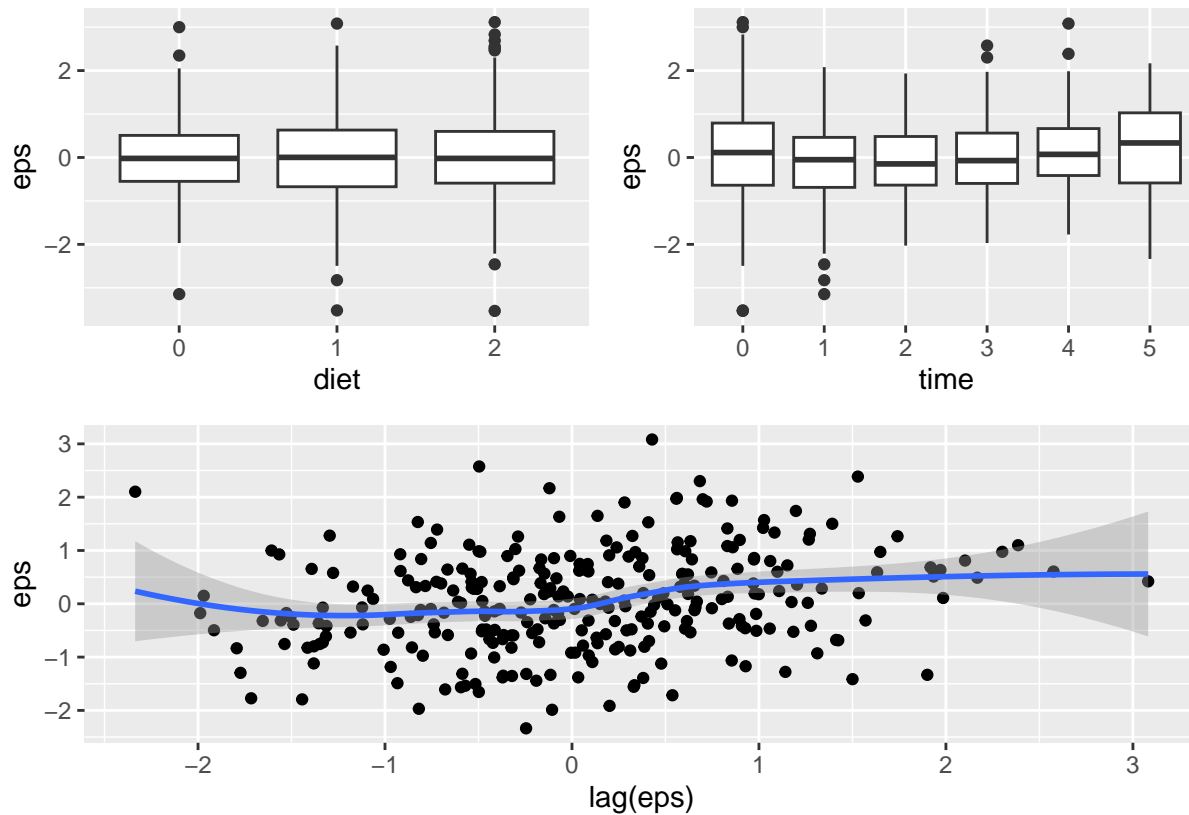
p1

```
## Warning: The dot-dot notation ('..density..') was deprecated in ggplot2 3.4.0.
## i Please use 'after_stat(density)' instead.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```
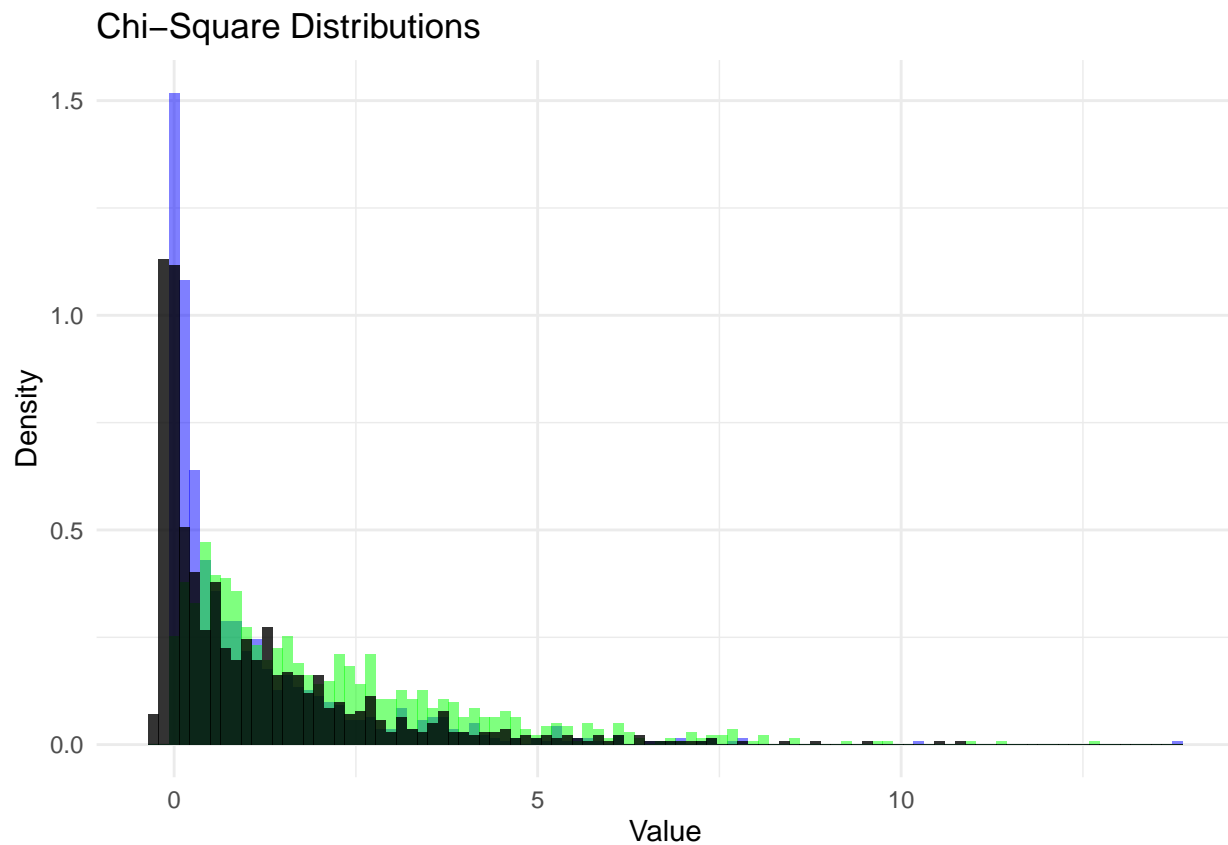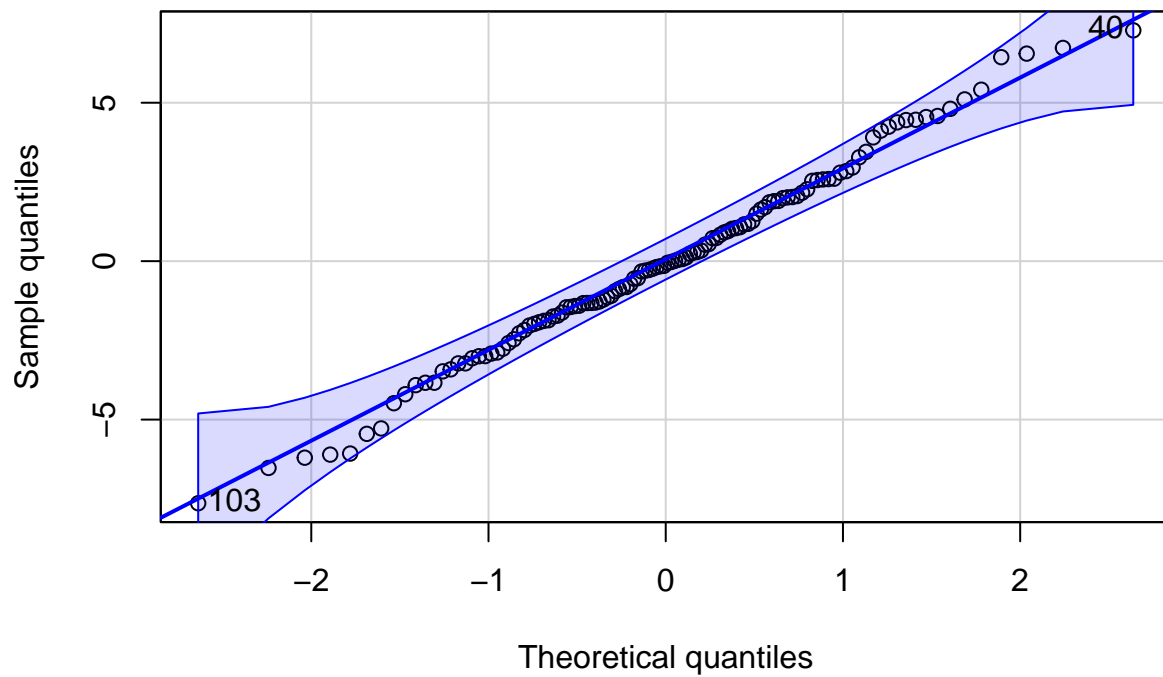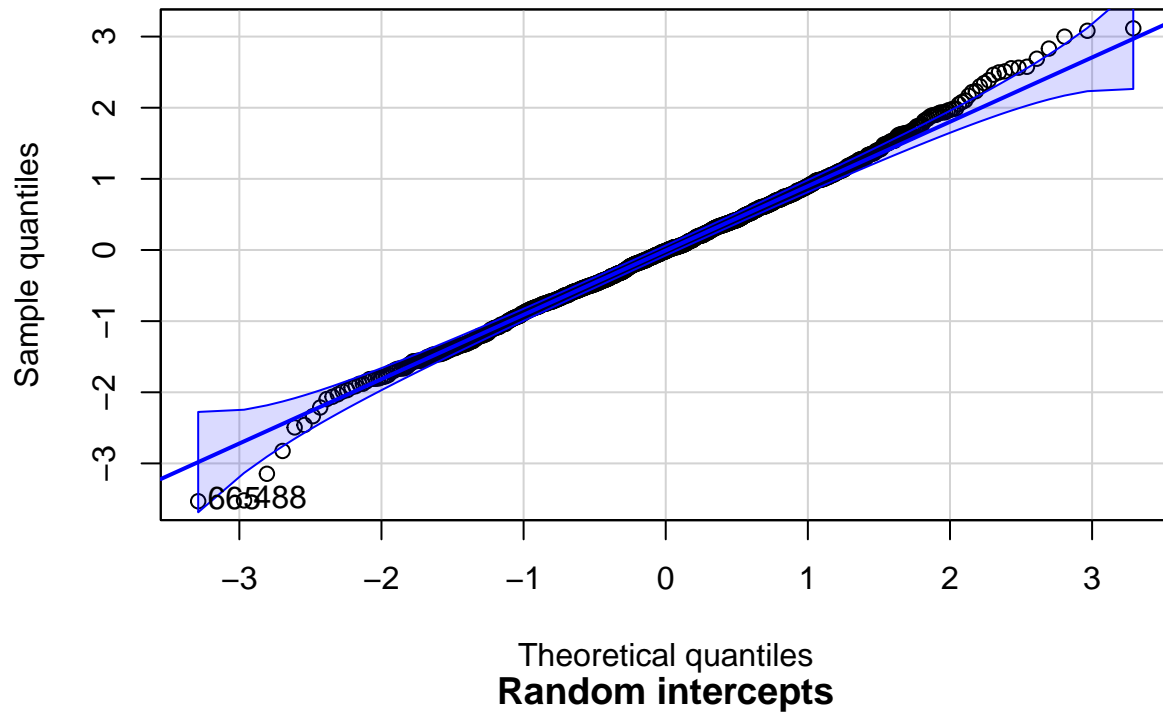
0.3 Question (c)

```
# box plot mean model - diet - stage 1
# box plot mean model - time - stage 1
# scatterplot for dependence model - stage 1
(p2+p3)/p4
```



```
# qqplot for normality - both
qqPlot(epsHat$eps, xlab = "Theoretical quantiles", ylab = "Sample quantiles", main = "Stage 1 residuals
qqPlot(gammaHat$gam, xlab = "Theoretical quantiles", ylab = "Sample quantiles", main = "Random intercept
```

Stage 1 residuals

Random intercepts


```
## [1] 768 528
```

0.4 Question (d)

```
knitr::kable(est, format = "markdown")
```

	Est	SE
GEE-I: diet	-0.3648114	0.1045868
GEE-I: time	0.0205598	0.0191428
GEE-I: time*diet	-0.0579918	0.0139632
GEE-E: diet	-1.3191729	0.3746549
GEE-E: time	0.0894030	0.0338822
GEE-E: time*diet	-0.2555805	0.0254618
GEE-E: rho	0.8746175	0.1266974
GEE-AR1: diet	-1.1194725	0.4007161
GEE-AR1: time	0.0621281	0.0392733
GEE-AR1: time*diet	-0.2647457	0.0280059
GEE-AR1:	0.9685963	0.0333514

0.5 Question (e)

0.6 Question (f)