Homework 6

Sarah Adilijiang

Problem 1
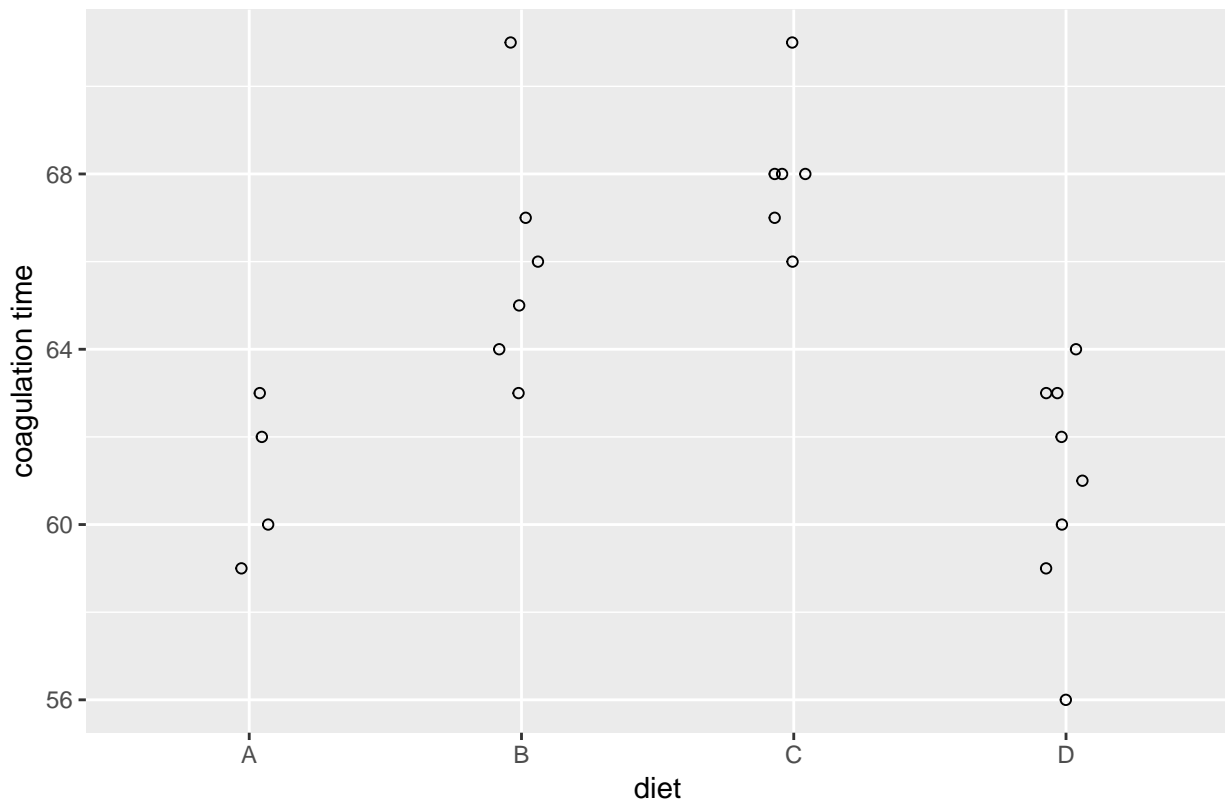
(a)

```r
library(faraway)
data(coagulation)

# plot the data
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.5.2
```

```r
ggplot(coagulation, aes(x=diet, y=coag)) +
    geom_point(position=position_jitter(width=0.1, height=0.0), shape=1) +
    labs(x="diet", y="coagulation time", caption="Note: Some jittering has been used to make coincident
    theme_grey()
```


Note: Some jittering has been used to make coincident points apparent.

Comment:

From the data plot, we can see that there is a significant difference in the blood coagulation time between different diets, so diet seems to be a significant effect for the response coagulation time.

(b)

The fixed effects model is:

$coag_{ij} = \mu + diet_i + \epsilon_{ij}$, where μ, $diet_i$ are fixed effects, and error term ϵ_{ij} i.i.d. $\sim N(0, \sigma^2)$.

```r
# fixed effect model
fmod = lm(coag~diet, coagulation)

# prediction & 95% prediction CI
predict(fmod, newdata=data.frame(diet="D"), interval="prediction", level=0.95)
```

```
##    fit      lwr      upr
## 1   61 55.76427 66.23573
```

```r
# using parametric bootstrap for 95% prediction CI
resid.sd = summary(fmod)$sigma

set.seed(123)
pred = numeric(1000)
for (i in 1:1000) {
    y = unlist(simulate(fmod))
    bmod = lm(y~diet, coagulation)
    pred[i] = predict(bmod, newdata=data.frame(diet="D")) + rnorm(n=1, sd=resid.sd)
}
quantile(pred, c(0.025, 0.975))
```

```
##     2.5%    97.5%
## 56.32447 65.92409
```

Answer:

The predicted value of the blood coagulation time of a new animal assigned to diet D is 61 seconds, and its 95% confidence interval is (55.76427, 66.23573) seconds.

While using parametric bootstrap, the 95% prediction CI is (56.32447, 65.92409) seconds.

(c)

The random effects model is:

$coag_{ij} = \mu + diet_i + \epsilon_{ij}$, where μ is fixed effects, and $diet_i$ is random effects.

Random effects $diet_i$ i.i.d. $\sim N(0, \sigma_b^2)$, error term ϵ_{ij} i.i.d. $\sim N(0, \sigma^2)$, and $diet_i$ and ϵ_{ij} are independent.

```r
# random effect model
library(lme4)
```

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

```r
rmod = lmer(coag~1+(1|diet), coagulation)

# two ways of prediction
fixef(rmod) + ranef(rmod)$diet["D",]
```

```
## (Intercept)
##    61.17017
```

```r
predict(rmod, newdata=data.frame(diet="D"))
```

```
##        1
```

2

```
## 61.17017
```

```
# using parametric bootstrap for 95% prediction CI
group.sd = as.data.frame(VarCorr(rmod))$sdcor[1]
resid.sd = as.data.frame(VarCorr(rmod))$sdcor[2]

set.seed(123)
pred = numeric(1000)
for (i in 1:1000) {
    y = unlist(simulate(rmod, use.u=TRUE))
    bmod = refit(rmod, y)
    pred[i] = predict(bmod, newdata=data.frame(diet="D")) + rnorm(n=1, sd=resid.sd)
}
quantile(pred, c(0.025, 0.975))
```

```
##     2.5%    97.5%
## 56.69907 66.22767
```

Answer:

The predicted value of the blood coagulation time of a new animal assigned to diet D is 61.17017 seconds, and its 95% confidence interval is (56.69907, 66.22767) seconds.

(d)

Here we are still using the random effect model.

```
# two ways of prediction
fixef(rmod)
```

```
## (Intercept)
##    64.01266
```

```
predict(rmod, re.form=~0)[1]
```

```
##        1
## 64.01266
```

```
# using parametric bootstrap for 95% prediction CI
set.seed(123)
pred = numeric(1000)
for (i in 1:1000) {
    y = unlist(simulate(rmod))
    bmod = refit(rmod, y)
    pred[i] = predict(bmod, re.form=~0)[1] +
        rnorm(n=1, sd=group.sd) + rnorm(n=1, sd=resid.sd)
}
quantile(pred, c(0.025, 0.975))
```

```
##     2.5%    97.5%
## 54.78558 72.62177
```

Answer:

The predicted value of the blood coagulation time of a new animal given a new diet is 64.01266 seconds, and its 95% confidence interval is (54.78558, 72.62177) seconds.

(e)

Here we are still using the random effect model.

```
# two ways of prediction
fixef(rmod)
```

```
## (Intercept)
##    64.01266
```

```
predict(rmod, re.form=~0)[1]
```

```
##        1
## 64.01266
```

```
# using parametric bootstrap for prediction & 95% prediction CI
set.seed(123)
pred = numeric(1000)
for (i in 1:1000) {
    y = unlist(simulate(rmod))
    bmod = refit(rmod, y)
    pred[i] = predict(bmod, re.form=~0)[1] +
        rnorm(n=1, sd=group.sd) + rnorm(n=1, sd=resid.sd)
}
quantile(pred, c(0.025, 0.975))
```

```
##     2.5%    97.5%
## 54.78558 72.62177
```

Answer:

The predicted value of the blood coagulation time of the first animal in the dataset when given a new diet is 64.01266 seconds, and its 95% confidence interval is (54.78558, 72.62177) seconds.

These values are exactly the same as in question (d), since which animal is given a new diet does not matter in this case, the only random effect in the model is the diet. Therefore, question (d) and (e) get the same answer.