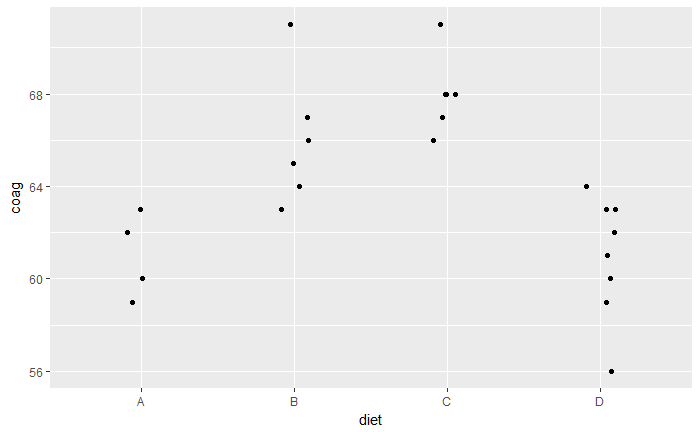
**STAT34700-HW4**

**Problem 2**

**(a)**



**Comments:**

From the plot of coagulation time vs diet type, we can see that for different diet types, the coagulation time are different. For example, for diet type A and D, the coagulation times are smaller than the coagulation times for diet type B and C. In addition, there is variance for coagulation times of each diet type. The variance of coagulation time for different diet types are not the same.

**(b)**

**The fixed effects model:**

Fit the fixed effects model:

> fit\_fix = lm(coag ~ diet, data = coagulation)

**Output:**

> summary(fit\_fix)

Call:

lm(formula = coag ~ diet, data = coagulation)

Residuals:

Min 1Q Median 3Q Max

-5.00 -1.25 0.00 1.25 5.00

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 6.100e+01 1.183e+00 51.554 < 2e-16 \*\*\*

dietB 5.000e+00 1.528e+00 3.273 0.003803 \*\*

dietC 7.000e+00 1.528e+00 4.583 0.000181 \*\*\*

dietD 2.991e-15 1.449e+00 0.000 1.000000

---

Signif. codes: 0 ?\*\*?0.001 ?\*?0.01 ??0.05 ??0.1 ??1

Residual standard error: 2.366 on 20 degrees of freedom

Multiple R-squared: 0.6706, Adjusted R-squared: 0.6212

F-statistic: 13.57 on 3 and 20 DF, p-value: 4.658e-05

Then we calculate the prediction and the 95% prediction interval for the response of a new animal assigned to diet D:

**Output:**

> predict(fit\_fix, new=data.frame(diet="D"), interval="prediction")

fit lwr upr

1 61 55.76427 66.23573

The prediction value is 61. The 95%prediction interval is [55.76, 66.24]

**(c)**

**The random effects model:**

Fit the random effects model:

> fit\_ran=lmer(coag~1+(1|diet),data=coagulation)

**Output:**

> summary(fit\_ran)

Linear mixed model fit by REML ['lmerMod']

Formula: coag ~ 1 + (1 | diet)

Data: coagulation

REML criterion at convergence: 115.8

Scaled residuals:

Min 1Q Median 3Q Max

-2.18491 -0.59921 0.09332 0.54078 2.17508

Random effects:

Groups Name Variance Std.Dev.

diet (Intercept) 11.692 3.419

Residual 5.599 2.366

Number of obs: 24, groups: diet, 4

Fixed effects:

Estimate Std. Error t value

(Intercept) 64.01 1.78 35.96

Then for the response of a new animal assigned to diet D, we calculate the prediction value by R. For the 95% prediction interval for blood coagulation time of a new animal assigned to diet D, use the bootstrap method. First, we fit the model 1000 times and get 1000 predictions. Second, we use the 0.025 and 0.975 quantiles of these predictions as the lower bound and higher bound for the 95% prediction interval:

**Output:**

> predict(fit\_ran, newdata=data.frame(diet="D"))

1

61.17017

> quantile(pv, c(0.025, 0.975))

2.5% 97.5%

56.69907 66.22767

The prediction value is 61.17. The 95%prediction interval is [56.70, 66.23].

**(d)**

Then for the response of a new diet is given to a new animal, we calculate the prediction value

by R. For the 95% prediction interval for blood coagulation time, we use the bootstrap method.

First, we fit the model 1000 times and get 1000 predictions for the response. Second, we use the 0.025 and 0.975 quantiles of these predictions as the lower bound and higher bound for the 95% prediction interval:

**Output:**

> predict(fit\_ran, re.form=~0)[1]

1

64.01266

> quantile(pv2, c(0.025, 0.975))

2.5% 97.5%

55.65531 72.31752

The prediction value is 64.01. The 95% prediction interval is [55.66, 72.32].

**(e)**

Because the effects of the initial diet for this animal have washed out, this animal is the same as a new animal. If a new diet is given to this animal in the dataset, it is equivalent to a new diet is given to a new animal. Hence, the result should be the same as the result we get in part (d).

The prediction value is 64.01. The 95% prediction interval is [55.66, 72.32].

**R Code**

library(faraway)

library(lme4)

library(ggplot2)

#a

data(coagulation)

ggplot(coagulation, aes(x=diet, y=coag))+geom\_point(position = position\_jitter(width=0.1, height=0.0))

#b

fit\_fix = lm (coag ~ diet, data = coagulation)

summary(fit\_fix)

predict(fit\_fix, new=data.frame(diet="D"), interval="prediction")

#c

fit\_ran=lmer(coag~1+(1|diet),data=coagulation)

summary(fit\_ran)

#prediction

predict(fit\_ran, newdata=data.frame(diet="D"))

#bootstrap to get CI of response

resid.sd <- as.data.frame(VarCorr(fit\_ran))$sdcor[2]

pv = numeric(1000)

set.seed(123)

for(i in 1:1000){

y <- unlist(simulate(fit\_ran, use.u=TRUE))

fit\_ran1 <- refit(fit\_ran, y)

pv[i] <- predict(fit\_ran1, newdata=data.frame(diet="D")) + rnorm(n=1,sd=resid.sd)

}

quantile(pv, c(0.025, 0.975))

#d

#prediction

predict(fit\_ran, re.form=~0)[1]

#bootstrap to get CI of response

group.sd <- as.data.frame(VarCorr(fit\_ran))$sdcor[1]

pv2 <- numeric(1000)

for(i in 1:1000){

y2 <- unlist(simulate(fit\_ran))

fit\_ran2 <- refit(fit\_ran, y2)

pv2[i] <- predict(fit\_ran2, re.form=~0)[1] + rnorm(n=1,sd=group.sd) + rnorm(n=1,sd=resid.sd)

}

quantile(pv2, c(0.025, 0.975))