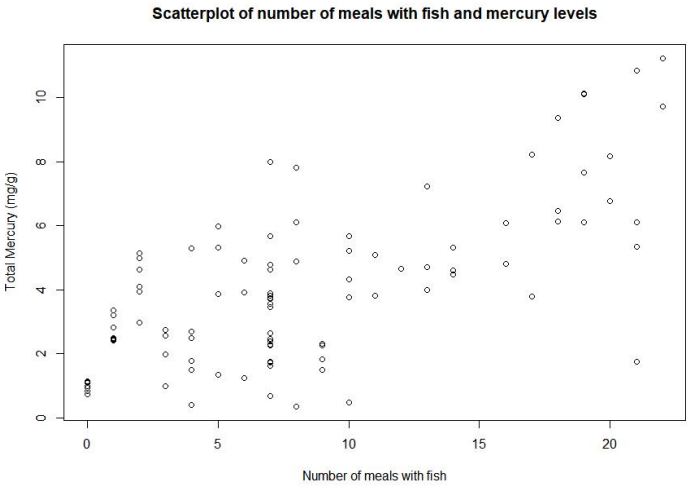
CS555

Weilin Lu

Assignment 3

1:

```
# read in data
data <- read.csv("/Users/12103/Desktop/CS555 HW/HW3/fish_mercury.csv")
#Q1
# create scatterplot with number of meals on the x-axis and mercury levels on the y-axis
plot(data$Number.of.meals.with.fish, data$Total.Mercury.in.mg.g,
     xlab = "Number of meals with fish", ylab = "Total Mercury (mg/g)",
     main = "Scatterplot of number of meals with fish and mercury levels")
```



2:

```
#Q2
cor(data$Number.of.meals.with.fish, data$Total.Mercury.in.mg.g)
```

```
> cor(data$Number.of.meals.with.fish, data$Total.Mercury.in.mg.g)
[1] 0.6991094
```

3:

```
#Q3
model <- lm(Total.Mercury.in.mg.g ~ Number.of.meals.with.fish, data=data)
summary(model)
plot(data$Number.of.meals.with.fish, data$Total.Mercury.in.mg.g,
     xlab="Number of meals with fish", ylab="Total Mercury in mg/g",
     main="Scatterplot of Number of meals with fish and Total Mercury levels")
abline(model, col="red")
```

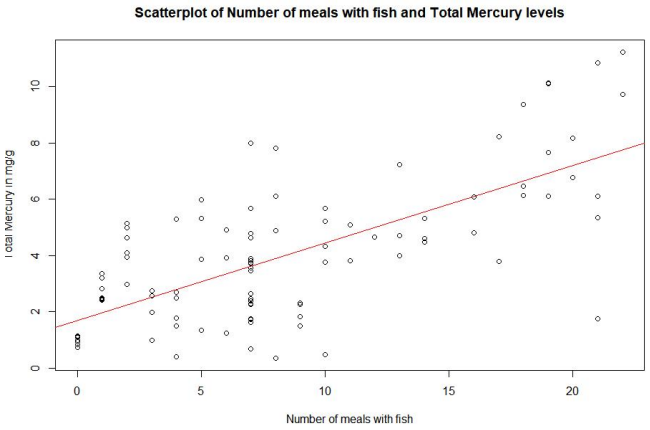
```
> summary(model)

Call:
lm(formula = Total.Mercury.in.mg.g ~ Number.of.meals.with.fish,
    data = data)

Residuals:
   Min     1Q Median     3Q    Max
-5.718 -1.143 -0.183  1.044  4.379

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)            1.68764    0.29833   5.657 1.53e-07 ***
Number.of.meals.with.fish  0.27595    0.02851   9.679 6.01e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.817 on 98 degrees of freedom
Multiple R-squared:  0.4888,    Adjusted R-squared:  0.4835
F-statistic: 93.69 on 1 and 98 DF,  p-value: 6.013e-16
```



4；

```
#Q4
# fitting the linear regression model
model <- lm(`Total Mercury in mg/g` ~ `Number of meals with fish`, data = data)

# extracting the regression coefficients
beta0 <- coef(model)[1]
beta1 <- coef(model)[2]

# interpreting the coefficients in the context of the data set
cat("The estimate for beta1 is", beta1, "which means that on average, for each additional meal with fish consumed per week, the total mercury in mg/g increases by", beta1, "mg/g.\n")
cat("The estimate for beta0 is", beta0, "which means that when the number of meals with fish consumed per week is 0, the total mercury in mg/g is", beta0, "mg/g.\n")
```

```
> cat("The estimate for beta1 is", beta1, "which means that on average, for each additional meal with fish consumed per week, the total mercury in mg/g increases by", beta1, "mg/g.\n")
The estimate for beta1 is 0.2759503 which means that on average, for each additional meal with fish consumed per week, the total mercury in mg/g increases by 0.2759503 mg/g.
> cat("The estimate for beta0 is", beta0, "which means that when the number of meals with fish consumed per week is 0, the total mercury in mg/g is", beta0, "mg/g.\n")
The estimate for beta0 is 1.687643 which means that when the number of meals with fish consumed per week is 0, the total mercury in mg/g is 1.687643 mg/g.
```

5：

```
#Q5
# ANOVA table
anova(model)
# standard error of beta1
summary(model)$coefficients[2, 2]
# F-test for beta1 = 0
summary(model)$fstatistic
# 5-step procedure for testing beta1 = 0 at alpha = 0.05
# Step 1: State the null and alternative hypotheses
# H0: beta1 = 0
# Ha: beta1 != 0
# Step 2: Determine the test statistic
# F-test, so the test statistic is the F-statistic from the ANOVA table
# Step 3: Determine the p-value
# From the ANOVA table, we see that the p-value is less than 0.05
# Step 4: Make a decision
# Since the p-value is less than 0.05, we reject the null hypothesis.
# Step 5: Interpret the results
# We have sufficient evidence to conclude that there is a significant linear relationship between the number of meals with fish consumed per week and the total mercury in mg/g.
# R-squared value
summary(model)$r.squared
# 90% confidence interval for beta1
confint(model, level = 0.90)
```

```
> anova(model)
Analysis of Variance Table

Response: Total.Mercury.in.mg.g
                       Df Sum Sq Mean Sq F value    Pr(>F)
Number.of.meals.with.fish  1 309.24 309.239  93.689 6.013e-16 ***
Residuals                 98 323.47   3.301
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> # standard error of beta1
> summary(model)$coefficients[2, 2]
[1] 0.02850937
> # F-test for beta1 = 0
> summary(model)$fstatistic
   value    numdf    dendf
93.68853  1.00000 98.00000
> # Ha: beta1 != 0
> # Step 2: Determine the test statistic
> # F-test, so the test statistic is the F-statistic from the ANOVA table
> # Step 3: Determine the p-value
> # From the ANOVA table, we see that the p-value is less than 0.05
> # Step 4: Make a decision
> # Since the p-value is less than 0.05, we reject the null hypothesis.
> # Step 5: Interpret the results
> # We have sufficient evidence to conclude that there is a significant linear relationship between the number of meals with fish consumed per week and the total mercury in mg/g.
> # R-squared value
> summary(model)$r.squared
[1] 0.488754
> # 90% confidence interval for beta1
> confint(model, level = 0.90)
                            5 %       95 %
(Intercept)            1.192253 2.1830324
Number.of.meals.with.fish 0.228609 0.3232916
```

EXTRA CREDIT

a) ：

NO. From the graph, if the seed weight is 1.6, the seed number is 12158. When the seed weight is 253, the seed number is 2475.So they are inversely proportional. An inverse relationship is not a linear model.

b) ：

Model a is a better choice. Model a is a very typical inverse graph. And the inverse graph is very

predictable

c) :
Weight should be:373
d) :
r^2=0.7016
r^2 is a statistical measure of the goodness of fit of the data. If r2 is closer to 1, the fit is better, otherwise it is worse. r2= 0.7016 shows that the logarithmic function can fit the relationship between the number of seeds and the weight of seeds relatively correctly.