Linear Models: Homework 1

Emanuel Gloria

2024-2025

Answers to the questions

Question 1a

Based on the regression analysis result, the effect of ADDF17 on the mean IgG1 blood level concentration is positive and statistically significant. Specifically, for every 1 unit increase of ADDF17 concentration, the mean IgG1 concentration increases by approximately 1.071 mg/mL on average. This suggests that higher concentration of ADDF17 are associated with increased IgG1 levels in the blood.

Question 1b

The standard error of the coefficients are 0.17 and 0.69 for the intercept and ADDF17, respectively. While the mean square of the residual standard error is 0.9127^2 or 0.833 with 58 degrees of freedom. This indicates a moderate level of precision in the estimate of the effect.

The 95% confidence interval for the effect size is (0.441, 1.702), meaning we are 95% confident that the true increase in IgG1 per unit increase in ADDF17 concentration lies between 0.437 mg/mL and 1.705 mg/mL. The relatively wide interval suggests some imprecision in the estimate, but the positive effect remains significant.

Question 1c

We are testing whether the concentration of ADDF17 has a statistical significant effect on IgG1 levels. A non-zero slope Beta in the regression would indicate that ADDF17 concentration influences IgG1 levels.

Thus, our hypotheses are:

Ho: Beta = 0, where Beta is the slope (the effect of concentration)

Ha: Beta != 0, meaning, there is a significant effect of ADDF17 on the mean of IgG1 levels

Based on the results at 5% level of significance, there is a significant effect of ADDF17 on the mean of IgG1 blood levels. Therefore, changes in ADDF17 concentration are statistically associated with the changes in IgG1.

Question 1d

After excluding the data for the mice that received a dose of 2 units, the effects of ADDF17 concentration on IgG1 levels is no longer statistically significant at 5% level. Removing those data reduced the range of concentrations, limiting the observed effect and increasing uncertainty. The high dose of 2 units likely had a stronger effect on IgG1 levels, and without it, the overall relationship between ADDF17 and IgG1 weakened. Hence, less statistically significant.

Question 2

This simulation study uses repeated sampling (1000 times) to interpret the effect of ADDF17 concentration on IgG1 levels, applying the principle of sampling from a known data-generating model with epsilon represented by $N(0,0.9)$. In this context, the model assumes we know the "true" relationship between concentration and IgG1 levels, using specified parameters to simulate data.

The averaging done on p2 and p1 in the simulation study measures how often the p-value from each regression model falls below the significance threshold of 0.05, indicating a statistically significant result.

- With the full dataset ($\text{mean}(p2 < 0.05) = 0.707$): On average, 70% of the time, the p-value is below 0.05, meaning that we detect a significant effect of ADDF17 concentration in about three-quarters of the simulations. This shows strong evidence of a relationship when the full data is used, including the higher concentrations.

- Without the 2-unit data ($\text{mean}(p1 < 0.05) = 0.21$): Only 21% of the time, the p-value is below 0.05, meaning that the concentration effect is detected far less frequently. Excluding the higher concentration leads to a much lower chance of finding statistical significance.

The comparison highlights that including the 2-unit dose increases the power to detect the effect of ADDF17 on IgG1 levels. Averaging the p-values from repeated simulations demonstrates how crucial the 2-unit data is in consistently revealing the true relationship between concentration and IgG1 levels. Without the higher dose, the model has much less power to detect the effect.

Appendix with R code

Question 1

```r
# Add the data
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
load("/Users/eman/Documents/3560_LIMO/Assignments/mice.RData")
mice2 <- mice %>% filter(concentration != 2) # Remove outlier

# Plot the data to visualize the relationship
par(mfrow = c(1,2))

plot(mice$concentration, mice$IgG1,
     main="ADDF17 vs IgG1 Levels (Full)",
     xlab="ADDF17 (µg/mL)",
     ylab="IgG1 (mg/mL)",
     pch=19, col="blue")

plot(mice2$concentration, mice2$IgG1,
     main="ADDF17 vs IgG1 Levels (~Full)",
     xlab="ADDF17 (µg/mL)",
     ylab="IgG1 (mg/mL)",
     pch=19, col="red")
```
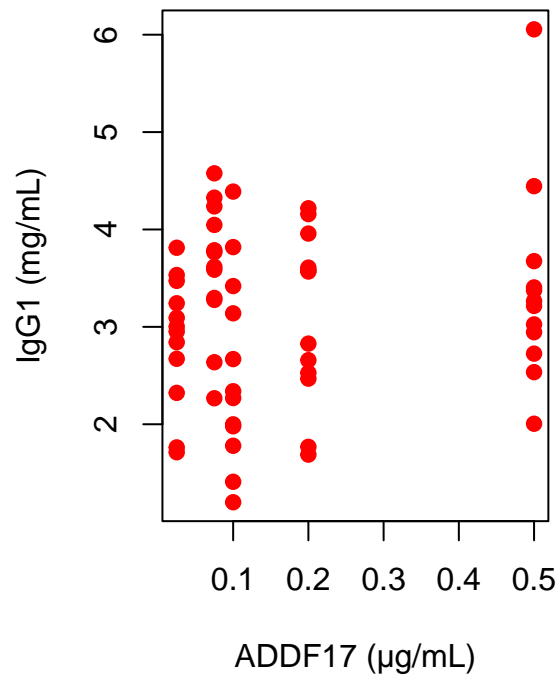
ADDF17 vs IgG1 Levels (Full)
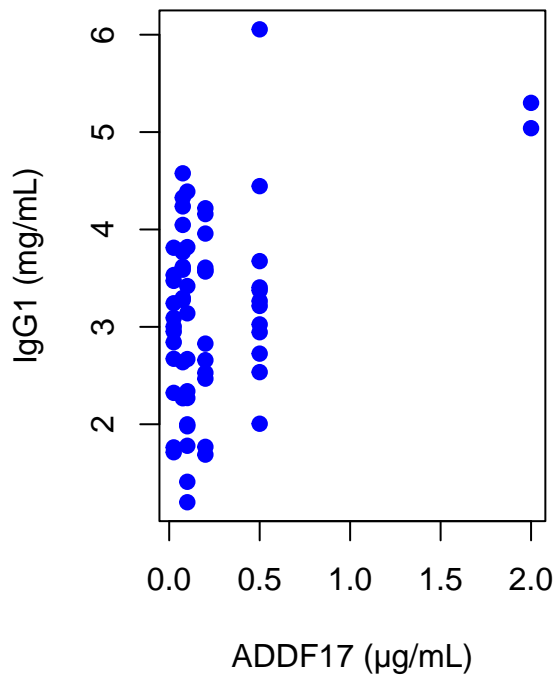
ADDF17 vs IgG1 Levels (~Full)


```r
m = lm(IgG1 ~ concentration, data = mice)
m2 = lm(IgG1 ~ concentration, data = mice2)

summary(m)
```

```
##
## Call:
## lm(formula = IgG1 ~ concentration, data = mice)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.81822 -0.64593  0.00123  0.59608  2.60921
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)     2.9101     0.1367  21.287   <2e-16 ***
## concentration   1.0714     0.3150   3.401   0.0012 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8989 on 60 degrees of freedom
## Multiple R-squared:  0.1616, Adjusted R-squared:  0.1476
## F-statistic: 11.57 on 1 and 60 DF,  p-value: 0.0012
```

```
summary(m2)
```

```
##
## Call:
## lm(formula = IgG1 ~ concentration, data = mice2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.83402 -0.64787  0.02151  0.58186  2.69190
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)     2.9505     0.1717   17.18   <2e-16 ***
## concentration   0.8252     0.6937    1.19    0.239
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9127 on 58 degrees of freedom
## Multiple R-squared:  0.02382,    Adjusted R-squared:  0.006985
## F-statistic: 1.415 on 1 and 58 DF,  p-value: 0.2391
```

```
confint(m)
```

```
##                    2.5 %   97.5 %
## (Intercept)   2.6366273 3.183535
## concentration 0.4412574 1.701578
```

```
confint(m2)
```

```
##                     2.5 %   97.5 %
## (Intercept)    2.6068235 3.294167
## concentration -0.5634331 2.213855
```

Question 2

```
set.seed(2678)
```

```
p1<-p2<-vector(length=1000)
for(i in 1:1000) {
  conc2<-mice$concentration
  mice2<-data.frame(
    concentration=mice$concentration,
    IgG1=3+0.83*conc2+rnorm(62,sd=0.9)
  )
  conc1<-mice$concentration[mice$concentration<2]
  mice1<-data.frame(
    concentration=conc1,
    IgG1=3+0.83*conc1+rnorm(60,sd=0.9)
  )
```

```
  m2<-lm(IgG1~concentration,data=mice2)
  p2[i]<-summary(m2)$coef[2,4]

  m1<-lm(IgG1~concentration,data=mice1)
  p1[i]<-summary(m1)$coef[2,4]
}

mean(p2<0.05)
```

```
## [1] 0.707
```

```
mean(p1<0.05)
```

```
## [1] 0.21
```