

Linear Models: Homework 1

Juan Andres VANEGAS JADAN

2024-2025

Answers to the questions

Question 1a

Each ug increase of ADDF17 has the effect of increasing 1.0714mg/L on the mean of IgG1 levels.

Question 1b

The 95% confidence interval lies between 0.44 and 1.7 mg/L, indicating that the effect is likely possible, as 0 is not included. This interval suggest that for each ug of ADDF17, the mean of IgG1 would increase between 0.44 and 1.7 mg/L with 95% of probability. With a standard error of 0.31, the estimation of the effect of ADDF17 over IgG1 has a moderate variability.

Question 1c

The effect of ADDF17 over IgG1 levels is statistically significant, with a p-value of 0.001, This result suggests that ADDF17 concentration has a meaningful impact on IgG1 levels.

The p-value was calculated with a two-sided test $H1 : \neq 0$, as there was no prior information on wheather ADDF17 would increase or decrease IgG1 levels.

Question 1d

If we do not include the doses of 2ug, we obtain a lower value for the estimation, equals to 0.8252. Furthermore with a p-value of 0.239, this estimation would not be significant.

The non significance can also be reiterated looking at the confidence interval. This would include values from -0.56 to 2.21, giving inconclusive information if the ADDF17 concentration would increase or decrease the IgG1 levels.

Question 2

The simulations reveal that the 2ug doses samples play a key role for obtaining significant estimations. When we include them, 71% of the simulations return a significant estimations. In contrast, when we exclude them, only 20% of the simulations return significant estimations. One reason could be that the higher doses samples have a larger and more consistent effect.

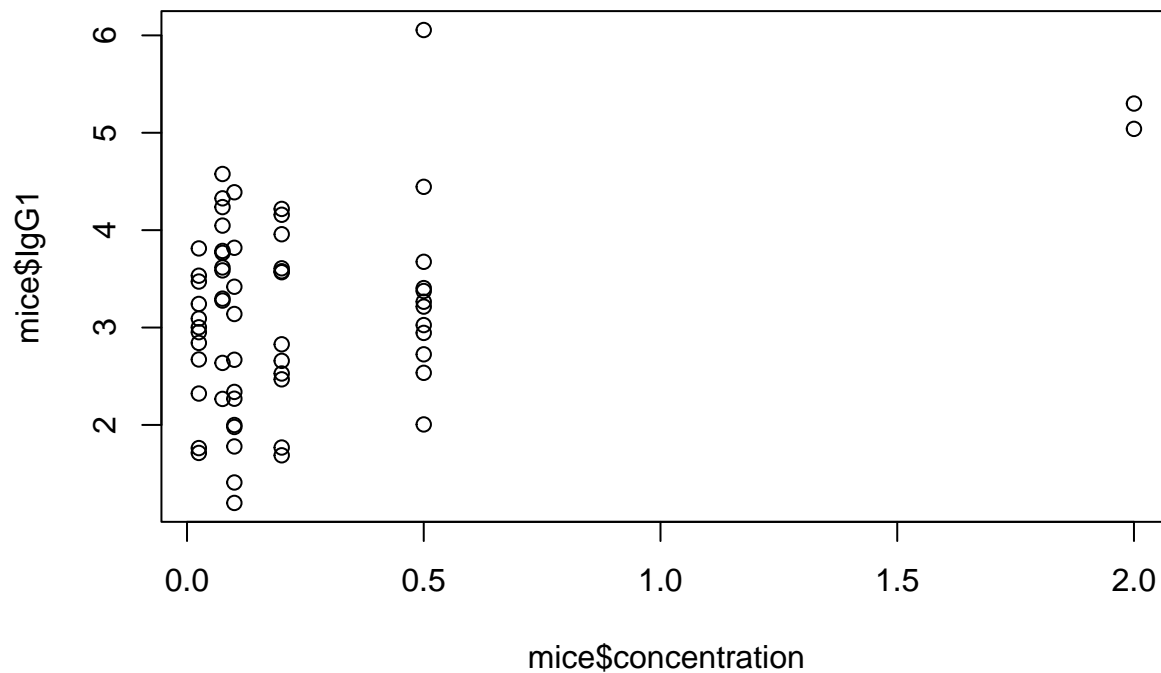
Appendix with R code

Question 1

```
home_path <- Sys.getenv("HOME")
setwd(paste(home_path, "/academics/hasselt/linear-models", sep = ""))
load(file = "Data/mice.RData")
str(mice)
```

```
## 'data.frame': 62 obs. of 2 variables:
## $ concentration: num 0.025 0.025 0.025 0.025 0.025 0.025 0.025 0.025 0.025 0.025 ...
## $ IgG1 : num 3 3.47 3.81 1.76 2.84 ...
```

```
plot(mice$concentration, mice$IgG1)
```



Definition of the linear model.

```
m <- lm(mice$IgG1 ~ mice$concentration)
summary(m)
```

```
##
## Call:
## lm(formula = mice$IgG1 ~ mice$concentration)
##
## 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 ***
## mice$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
```

Estimation and variance of the estimation give us:

```
coef_est <- coef(summary(m))
beta1_hat <- coef_est["mice$concentration", "Estimate"]
sigma2_beta1 <- coef_est["mice$concentration", "Std. Error"]
print(paste("The parameter beta1 is estimated to: ", beta1_hat))
```

```
## [1] "The parameter beta1 is estimated to:  1.07141776113785"
```

```
print(paste("and it has a std deviation of ", sigma2_beta1))
```

```
## [1] "and it has a std deviation of  0.315033268359725"
```

Computation of the confidence interval give us:

```
lowerci <- beta1_hat - qt(0.975, df = nrow(mice) - 2) * sigma2_beta1
upperci <- beta1_hat + qt(0.975, df = nrow(mice) - 2) * sigma2_beta1
print(paste("The confidence intervale is beetween ", lowerci, " and ", upperci))
```

```
## [1] "The confidence intervale is beetween  0.441257400575863  and  1.70157812169984"
```

Computation of the p-value at two sides give us:

```
t_observed <- beta1_hat / sigma2_beta1
p.value.twosided <- 2 * (1 - pt(abs(t_observed), df = nrow(mice) - 2))
print(paste("The p-value of the parameter estimation is of ", p.value.twosided))
```

```
## [1] "The p-value of the parameter estimation is of  0.00119987396850418"
```

If we evaluate the same data but excluding the dose of 2ug

```
mice_filtered <- mice %>% filter(concentration != 2)
m_filtered <- lm(IgG1 ~ concentration, data = mice_filtered)
summary(m_filtered)
```

```
##
## Call:
## lm(formula = IgG1 ~ concentration, data = mice_filtered)
##
## 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_filtered)
```

```
##              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
```