

Lab8

For this lab we will explore ANCOVA models using simulated data. Consider the motivation presented in class where the experimental units are piglets of different weights.

Q1

Consider the 60 “fake” piglets and write code to add a column titled `treatment` to the `fake_pigs` tibble. Use a CRD to allocate 20 EUs to treatments 1, 2, and 3.

```
set.seed(03212022)
total_pigs <- 60
fake_pigs <- tibble(piglet = 1:total_pigs, piglet_weight = runif(total_pigs, min = 10, max = 30))
```

Q2

Consider the following statistical model:

$$Y_{ij} = \tau_i + x_{ij}\beta + E_{ij}$$

where Y_{ij} is the weight after a study for the j^{th} piglet in the i^{th} treatment, τ_i is the treatment effect associated with treatment i (when holding $\mu = 0$ for identifiability), x_{ij} is the starting weight for the ij^{th} piglet, β is effect associated with the starting weight, and E_{ij} is the error term in the model, where $E_{ij} \sim N(0, \sigma^2)$.

Data has been simulated for each of the treatments using values below for $\tau_1, \tau_2, \tau_3, \beta$, and σ simulate values for Y . Create a figure that shows Y and includes indicators (color / shapes) for the different treatments. (Note: you'll need to make sure you add a column titled `treatment` in your `fake_pigs` tibble and remove the `eval = F` tag on the R code chunk below.)

```
tau1 <- 10
tau2 <- 15
tau3 <- 20
beta <- 1
sigma <- 2

X_matrix <- model.matrix(~factor(treatment) - 1 + piglet_weight, data = fake_pigs)
param_vec <- c(tau1, tau2, tau3, beta)
Y <- rnorm(total_pigs, mean = X_matrix %*% param_vec, sd = sigma)
fake_pigs <- fake_pigs %>% mutate(Y = Y)
```

Q3

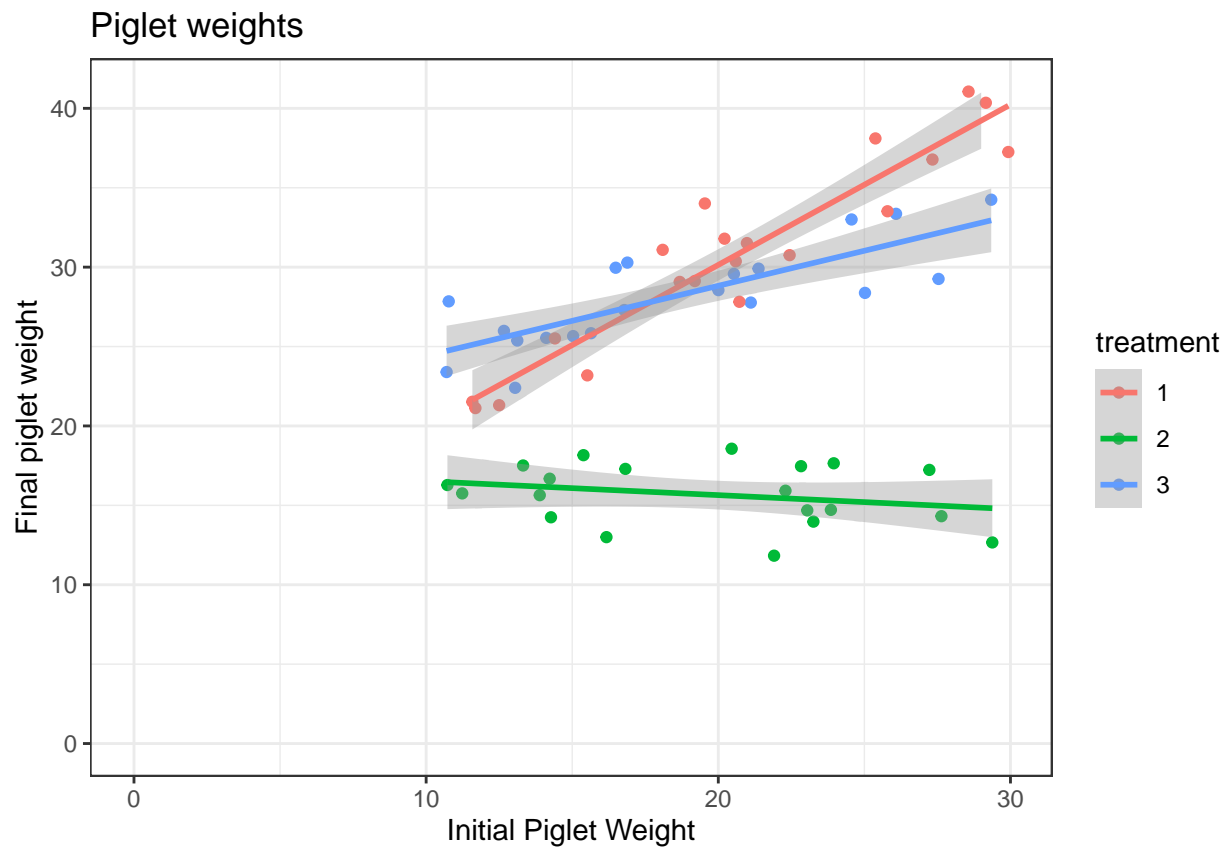
Use `lm` to fit the model spelled in in Q2 to the sythetic data. Print the model output here. Use the cell means specification.

Q4

Now consider the model in Q2, does the treatment effect differ depending on the starting weight? Why or why not?

Q5

Consider the figure below and explain how the initial piglet weights and the treatment factors influence the final weights.



Q6 (541 only)

Fit a model to the data set created for Q5 that includes an interaction term. Interpret the parameters in this model - for this you don't need to talk about the actual values, just the meaning of the coefficients.