Lab8

key

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))
```

```
fake_pigs <- fake_pigs %>% mutate(treatment = factor(sample(rep(1:3, total_pigs/3), total_pigs)))
```

Q2

Consider the following statistical model:

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

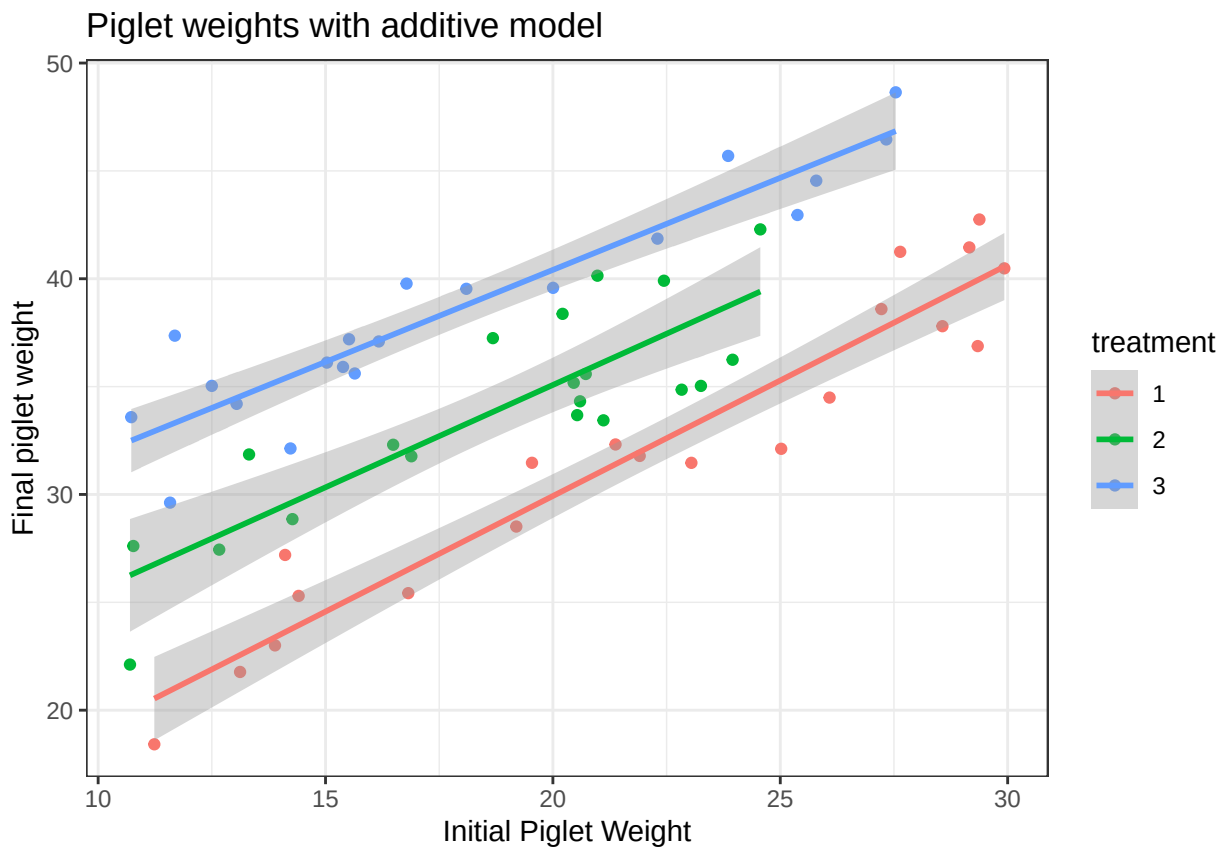
where Y_{ij} is the weight after a study for the j^{th} piglet in the i^{th} treatment, τ_i is the treatment effect associated with treatment i (when holding $\mu = 0$ for identifiability), x_{ij} is the starting weight for the ij_{th} piglet, β is effect associated with the starting weight, and E_{ij} 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 τ_1, τ_2, τ_3, β, 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 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)
```

```
fake_pigs %>%
  ggplot(aes(y = Y, x = piglet_weight, color = treatment)) +
  geom_point() +
  geom_smooth(method = 'lm', formula = 'y~x') +
  theme_bw() + xlab('Initial Piglet Weight') +
  ylab('Final piglet weight') +
  ggtitle('Piglet weights with additive model')
```



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.

```
fake_pigs %>% lm(Y ~ treatment + piglet_weight -1, data = .) %>% summary()
```

```
##
## Call:
## lm(formula = Y ~ treatment + piglet_weight - 1, data = .)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.9553 -1.6192 -0.2403  1.5182  4.7824
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## treatment1    10.69996    1.27943   8.363 1.96e-11 ***
## treatment2    15.67270    1.11990  13.995  < 2e-16 ***
## treatment3    21.22416    1.07968  19.658  < 2e-16 ***
## piglet_weight  0.97162    0.05349  18.165  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.218 on 56 degrees of freedom
## Multiple R-squared:  0.9963, Adjusted R-squared:  0.9961
## F-statistic:  3819 on 4 and 56 DF,  p-value: < 2.2e-16
```

Q4

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

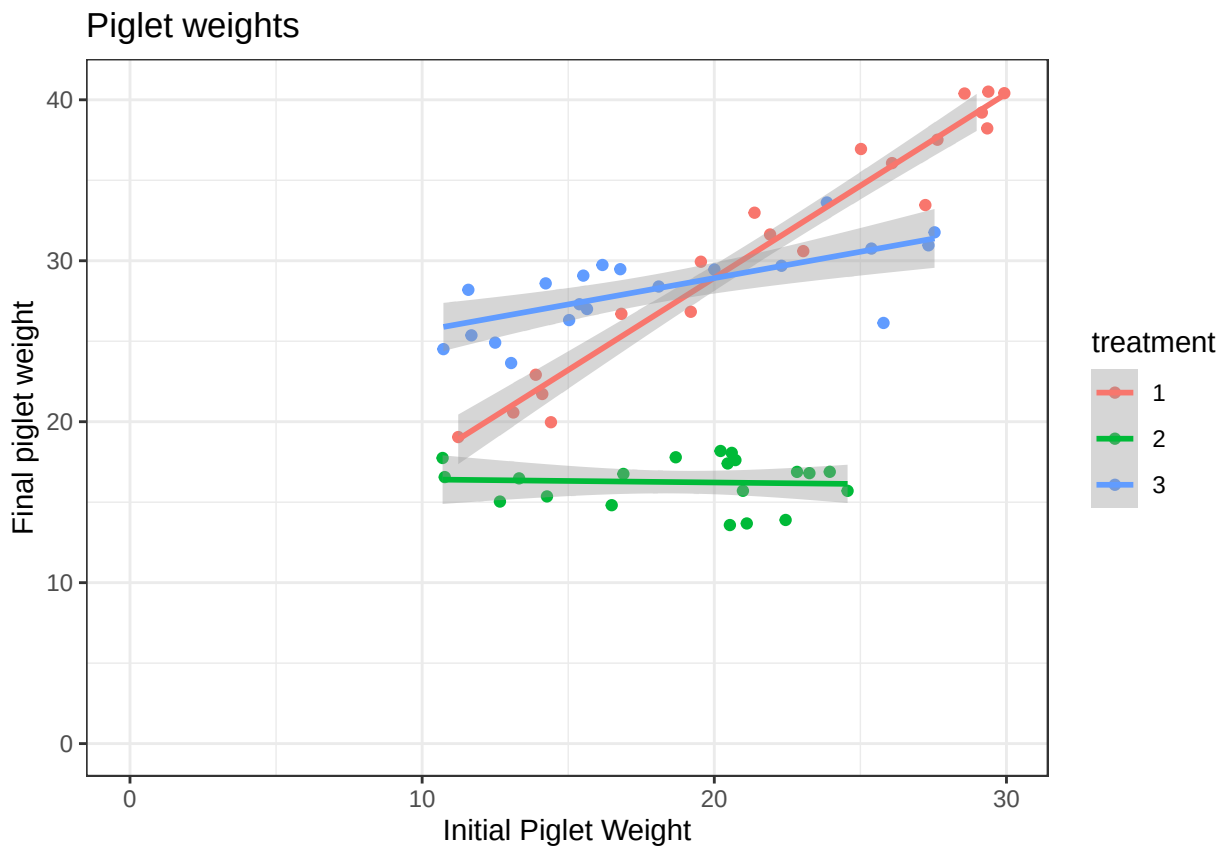
In this particular case, the final weight is just the starting weight plus the treatment weight (with error), so weight gained does not depend on starting weight - although final weight certainly does. Even if $\beta \neq 1$ the treatment effects do not depend on starting weights (although the final weight would).

Q5

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

```
X_matrix <- model.matrix(~factor(treatment) * piglet_weight -1, data = fake_pigs)
param_vec <- c(tau1, tau2, tau3, beta, -1, -.6)
new_Y <- rnorm(total_pigs, mean = X_matrix %*% param_vec, sd = sigma)
fake_pigs <- fake_pigs %>% mutate(Y2 = new_Y)

fake_pigs %>%
  ggplot(aes(y = Y2, x = piglet_weight, color = treatment)) +
  geom_point() +
  geom_smooth(method = 'lm', formula = 'y~x') +
  theme_bw() + xlab('Initial Piglet Weight') +
  ylab('Final piglet weight') +
  ggtitle('Piglet weights') +
  ylim(0, max(new_Y)) + xlim(0, max(fake_pigs$piglet_weight))
```



Piglets receiving treatment 2 all end up at about the same weight, regardless of starting weight. Many of these lose considerable weight. For treatments 1 and 3, the final weight seems to depend on a combination of the initial weight and the treatment level. For piglets starting at closer to 10 units, the treatment 3 leads to higher weights; however, for initial weights closer to 30 treatment 1 leads to higher weights. This is all to say the the treatment effect depends on an interaction with the intitial weight. Another way to state this is if you are hoping to have the largest pig, the best treatment would depend on the starting weight.

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.

```
fake_pigs %>% lm(Y2 ~ treatment * piglet_weight -1, data = .) %>% summary()
```

```
##
## Call:
## lm(formula = Y2 ~ treatment * piglet_weight - 1, data = .)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.6814 -0.9767  0.1202  1.3596  3.4357
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## treatment1                  6.04327    1.39193   4.342 6.26e-05 ***
## treatment2                 16.60369    1.71090   9.705 1.95e-13 ***
## treatment3                 22.37841    1.30792  17.110  < 2e-16 ***
## piglet_weight               1.14461    0.06076  18.838  < 2e-16 ***
## treatment2:piglet_weight   -1.16361    0.10767 -10.807 4.08e-15 ***
## treatment3:piglet_weight   -0.81745    0.09257  -8.831 4.61e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.689 on 54 degrees of freedom
## Multiple R-squared:  0.9963, Adjusted R-squared:  0.9959
## F-statistic:  2446 on 6 and 54 DF,  p-value: < 2.2e-16
```

- `treatment1`: is the intercept or simple effect for treatment 1 when initial piglet weight = 0 (not scientifically meaningful)
- `treatment2`: is the intercept or simple effect for treatment 2 when initial piglet weight = 0 (not scientifically meaningful)
- `treatment3` is the intercept or simple effect for treatment 3 when initial piglet weight = 0 (not scientifically meaningful)
- `piglet_weight` is the slope or expected increase for each unit of initial weight for a piglet receiving treatment one
- `treatment2:piglet_weight` is the difference in slope between treatments one and two, if this is not zero the weight gain depends on treatment *AND* starting weight
- `treatment3:piglet_weight` is the difference in slope between treatments one and three, if this is not zero the weight gain depends on treatment *AND* starting weight