# exercise 4 key

# dataset 1

datum <- read.csv("exercise\_4\_dataset1.csv")

head(datum)

# 1) y = beta0 + beta1 \* elevation + beta2 \* latitude + beta3 \* Nevada + beta4 \* Utah + error ~ N(0, sigma)

# 2) beta0 - the average body size when elevation and latitude are 0 for the reference group (Colorado)

# beta1 - the slope of the elevation/body size relationship; change in body size for each unit increase in eleevation

# beta2 - the slope of the latitude/body size relationship; change in body size for each unit increase in latitude

# beta3 - the difference in body size between animals in Nevada and the Colorado

# beta4 - the difference in body size betwen animals in Utah and Colorado

results <- lm(Size ~ Latitude + Elevation + Nevada + Utah, data = datum)

summary(results)

# for each one degree increase in latitude, we observed a 0.89 (+/-0.15; +/-95% CI) decrease in body size (P = 2e-16).

# for each 100 m increase in elevation, we observed a 0.16 (+/-0.08; +/-95% CI) decrease in body size (P = 0.000239).

# we found that animals in Nevada were 1.95 kg (+/-0.53; +/-95% CI) heavier than animals in Colorado (1.31e-10).

# we found that animals in Utah were 1.27 kg (+/-0.53; +/-95% CI) heavier than animals in Colorado (6.10e-06).

TukeyHSD(aov(results))

# or

results\_aov <- aov(Size ~ Latitude + Elevation + State, data = datum)

TukeyHSD(results\_aov, "State")

# the effects changed, which is wrong

# dataset2

#1

datum <- read.csv("exercise\_4\_dataset2.csv")

#2

plot(Cover ~ Food, data = datum)

#3

summary(lm(Cover ~ Food, data = datum))

summary(lm(Food ~ Cover, data = datum)) # same results for r^2 for each

#4

plot(MarmotDensity ~ Cover, data = datum)

#5

plot(MarmotDensity ~ Food, data = datum)

#6

results1 <- lm(MarmotDensity ~ Food, data = datum)

summary(results1)

#7

results2 <- lm(MarmotDensity ~ Cover, data = datum)

summary(results2)

#8

results3 <- lm(MarmotDensity ~ Food + Cover, data = datum)

summary(results3)

#9

car::vif(results3)

#10

# - multiple variable model best estimates truth, unbiased

# - final model would be the multivariable (multiple regression) model, and I

# dealt with collinearity simply by accounting for all confounding variables

# - i learned that i accounted for collinearity simply by accounting for all

# confounding variables

# dataset3

#1

datum <- read.csv("exercise\_4\_dataset3.csv")

head(datum)

#2

plot(Sediment ~ Organic, data = datum)

#3

summary(lm(Sediment ~ Organic, data = datum))

#4

plot(Clarity ~ Organic, data = datum)

#5

summary(lm(Clarity ~ Organic, data = datum))

#6

results1 <- lm(Clarity ~ Sediment, data = datum)

summary(results1)

#7

results2 <- lm(Clarity ~ Organic, data = datum)

summary(results2)

#8

results3 <- lm(Clarity ~ Sediment + Organic, data = datum)

summary(results3)

# report results here

#9

car::vif(results3)

#10

# - univariate models both recovered significant effects, but the effect of organic

# decreased to zero in the full model. this suggests it is a redundant variable.

# - my final model would be the simple univariate model of 'results1'. the multi-variable

# model suggests that 'Organic' is a redundant variable that does not explain

# significant extra variation. we can thus drop it from the model to reduce variance

# inflation, so i would prefer the results from results1

# - we can drop redundant variables from full models when they don't contribute much.