# exercise\_2 script

### dataset 1

datum <- read.csv("exercise\_2\_dataset1.csv")

# examine the data

head(datum)

# plot the data

plot(PredRate ~ PreyDens, data=datum,

xlab = expression("Prey density (prey/ha)"),

ylab = expression("Predation rate (prey killed per day)"))

# homoscedasticity is violated

# fit the regression

results <- lm(PredRate ~ PreyDens, data = datum)

abline(results)

# examine the results

summary(results)

confint(results)

# 95% ci

ci <- (confint(results)[2,2] - confint(results)[2,1])/2

p <- summary(results)$coefficients[2,4]

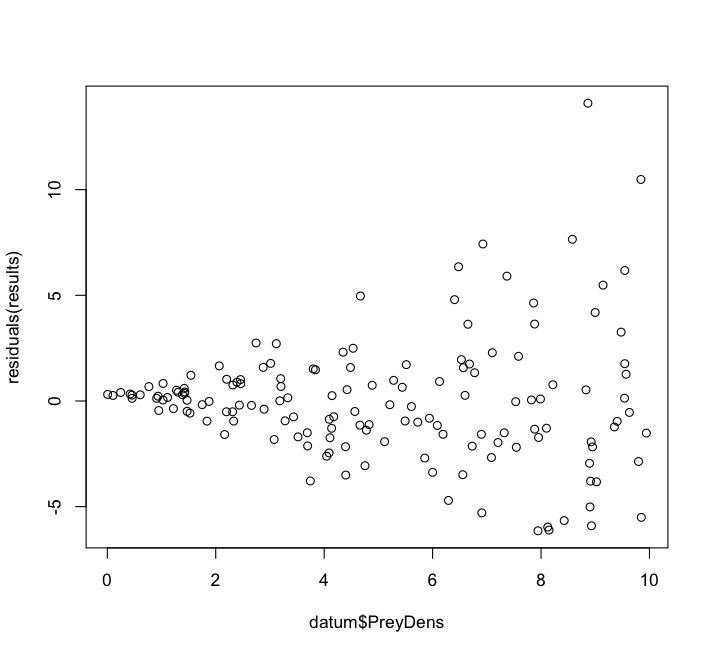
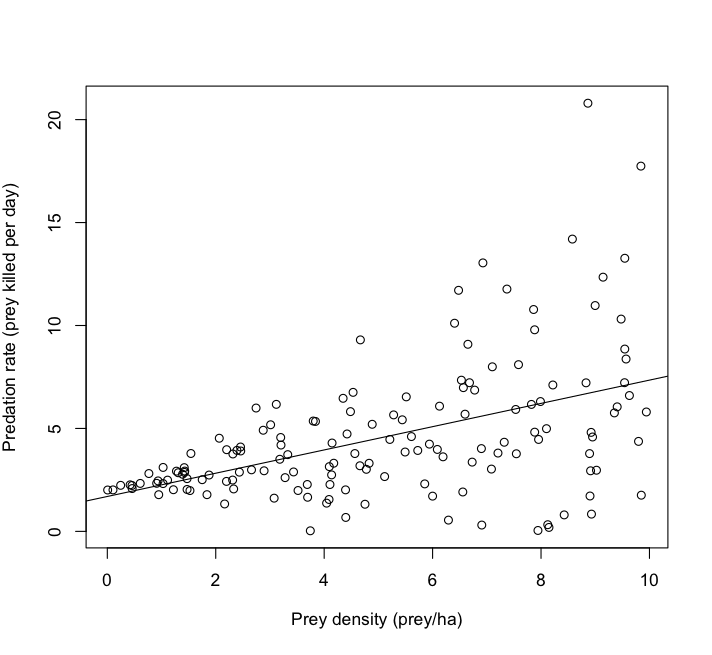
r2 <- summary(results)$r.squared

# for each 1 prey/ha increase in prey density, we observed a 0.56 prey killed

# per day (+/-0.17; +/-95% CI) increase in predation rate (p = 3.043404e-10; r2 = 0.24)

# homoscedasticity was violated, but this does not influence the parameter estimate.

# the parameter estimate and its 95% CI overlap Truth (0.5)



### dataset 2

datum <- read.csv("exercise\_2\_dataset2.csv")

# examine the data

head(datum)

# plot the data

plot(BasalArea ~ Age, data=datum, xlab = "Stand age (years)", ylab = "Tree basal area (m^2)")

# fit the regression

results <- lm(BasalArea ~ Age, data=datum)

abline(results)

# examine the results

summary(results)

confint(results)

# 95% ci

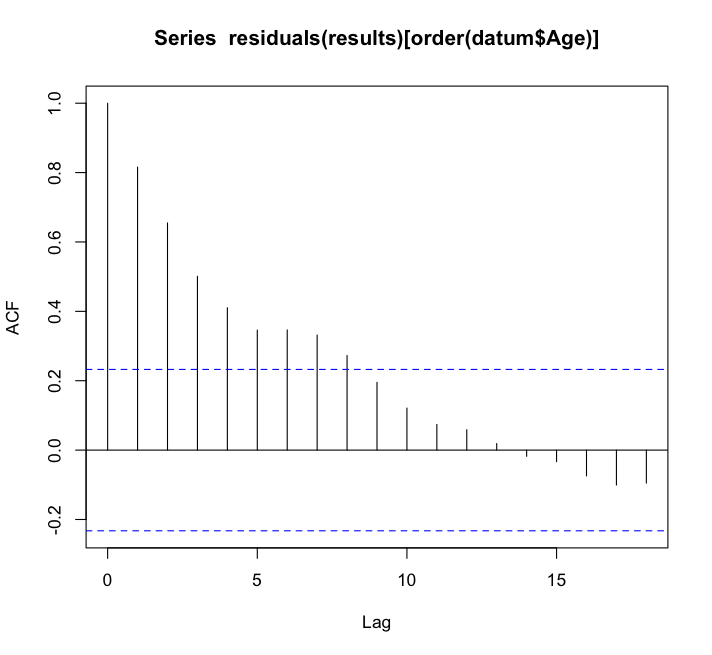
effect <- summary(results)$coefficients[2,1]

ci <- (confint(results)[2,2] - confint(results)[2,1])/2

ci <- (confint(results)[2,2] - results$coefficients[2])

p <- summary(results)$coefficients[2,4]

r2 <- summary(results)$r.squared



effect; ci; p; r2

# for each 1 year increase in stand age, we observed a 2.15 m^2 (+/-0.05; +/-95% CI)

# increase in tree basal area (p = < 2e-16; r2 = 0.99)

# autocorrelation was violated, but this does not influence the parameter estimate.

# the parameter estimate and its 95% CI overlap Truth (2)

acf(residuals(results)[order(datum$Age)])

### dataset 3

datum <- read.csv("exercise\_2\_dataset3.csv")

# examine the data

head(datum)

# plot the data

plot(LitterSize ~ PreyDensity, data=datum,

xlab = "Prey density (prey/ha)", ylab = "Bear litter size (pups)")

# fit the regression

results <- lm(LitterSize ~ PreyDensity, data=datum)

abline(results)

# examine the results

summary(results)

confint(results)

# 95% ci

effect <- summary(results)$coefficients[2,1]

ci <- (confint(results)[2,2] - confint(results)[2,1])/2

ci <- (confint(results)[2,2] - results$coefficients[2])

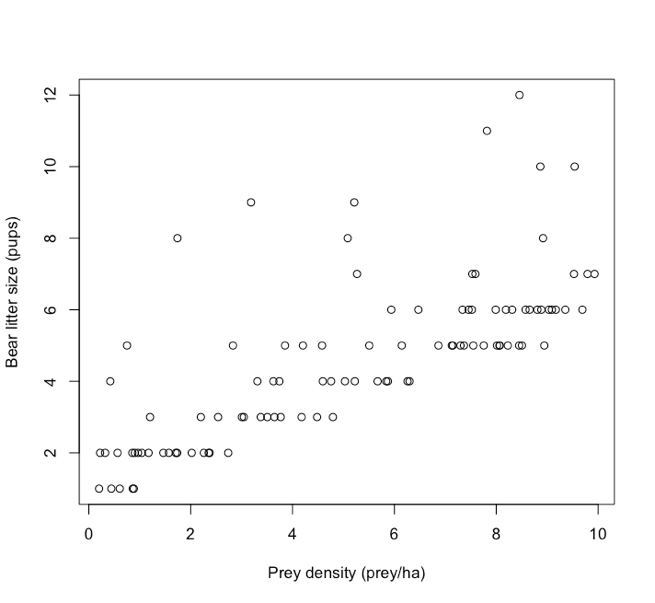
p <- summary(results)$coefficients[2,4]

r2 <- summary(results)$r.squared

effect; ci; p; r2

# for each 1 prey/ha increase in prey density, we observed a 0.54 cub (+/-0.10; +/-95% CI)

# increase in bear litter size (p = < 2e-16; r2 = 0.52)



# normality of error was violated, but this does not influence the parameter estimate.

# the parameter estimate and its 95% CI overlap Truth (0.5)

# three graphs can diagnose this:

# x-y plot

plot(LitterSize ~ PreyDensity, data=datum,

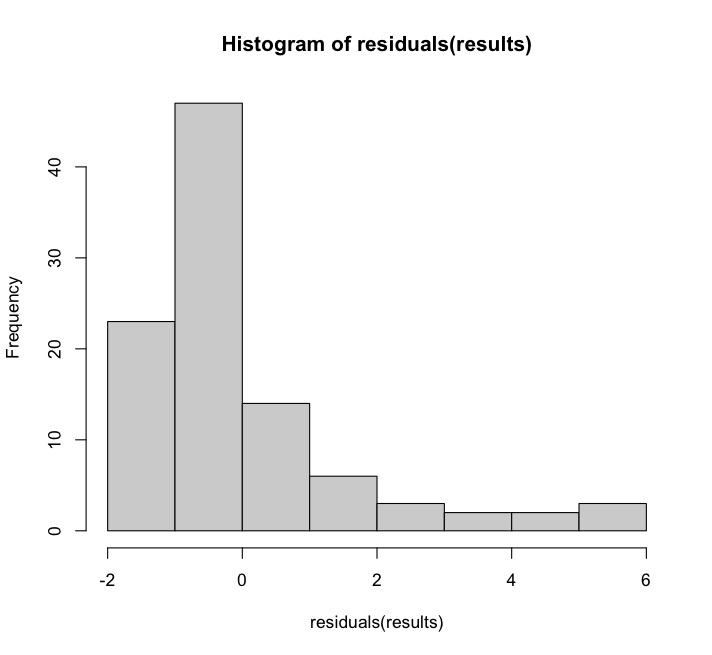
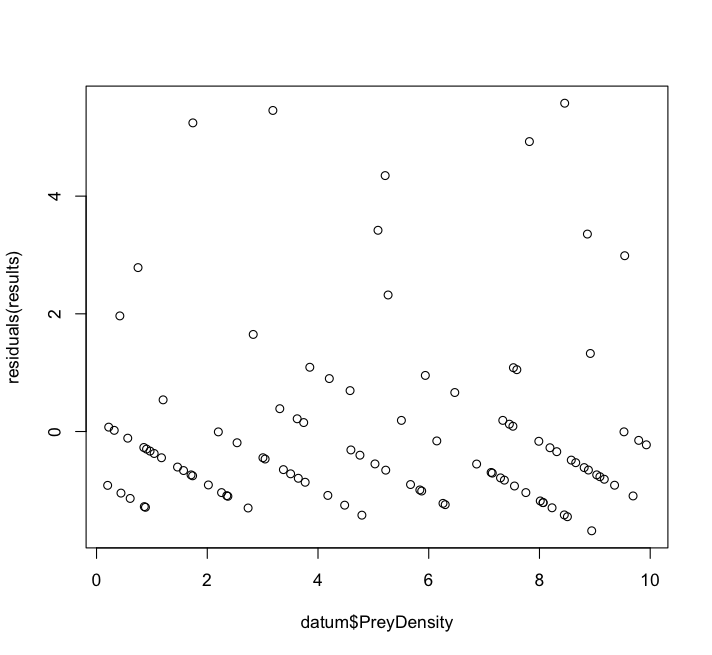
xlab = "Prey density (prey/ha)", ylab = "Bear litter size (pups)")

# residuals plot

plot(residuals(results) ~ datum$PreyDensity, data=datum)

# residuals hist

hist(residuals(results))



### dataset 4

datum <- read.csv("exercise\_2\_dataset4.csv")

# examine the data

head(datum)

# plot the data

plot(ProbUse ~ RoadDensity, data=datum,

xlab = "Road density (roads/km^2)",

ylab = "Probability of habitat use (%)")

# fit the regression

results <- lm(ProbUse ~ RoadDensity, data=datum)

abline(results)

# examine the results

summary(results)

confint(results)

# 95% ci

effect <- summary(results)$coefficients[2,1]

ci <- (confint(results)[2,2] - confint(results)[2,1])/2

ci <- (confint(results)[2,2] - results$coefficients[2])

p <- summary(results)$coefficients[2,4]

r2 <- summary(results)$r.squared

effect; ci; p; r2

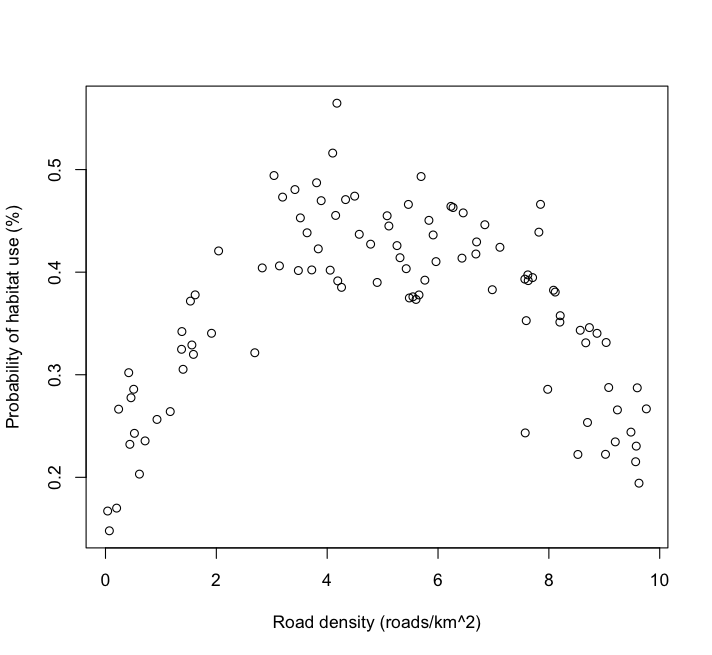
# for each 1 road/km2 increase in road density, we observed a 0.001 percent

# (+/-0.006; +/-95% CI) increase in probability of habitat use; however, this

# result was not statistically significant (p = < 2e-16; r2 = 0.002)

# linearity of error was violated, and this does influence the parameter estimate.

# the user should use a different model to estimate this relationship



# three graphs can diagnose this:

# x-y plot

plot(ProbUse ~ RoadDensity, data=datum,

xlab = "Road density (roads/km^2)",

ylab = "Probability of habitat use (%)")

# residuals plot

plot(residuals(results) ~ datum$RoadDensity, data=datum)

# autocorrelation function

acf(residuals(results)[order(datum$RoadDensity)])

### dataset 5

datum <- read.csv("exercise\_2\_dataset5.csv")

# examine the data

head(datum)

# plot the data

plot(CowsKilled ~ DeerDensity, data=datum,

xlab = "Native deer density (deer/ha)",

ylab = "Domestic livestock killed (livestock/km^2)")

# fit the regression

results <- lm(CowsKilled ~ DeerDensity, data=datum)

abline(results)

# examine the results

summary(results)

confint(results)

# 95% ci

effect <- summary(results)$coefficients[2,1]

ci <- (confint(results)[2,2] - confint(results)[2,1])/2

ci <- (confint(results)[2,2] - results$coefficients[2])

p <- summary(results)$coefficients[2,4]

r2 <- summary(results)$r.squared

effect; ci; p; r2

# for each 1 deer/ha increase in native deer density, we observed a 1.06 livestock/km2

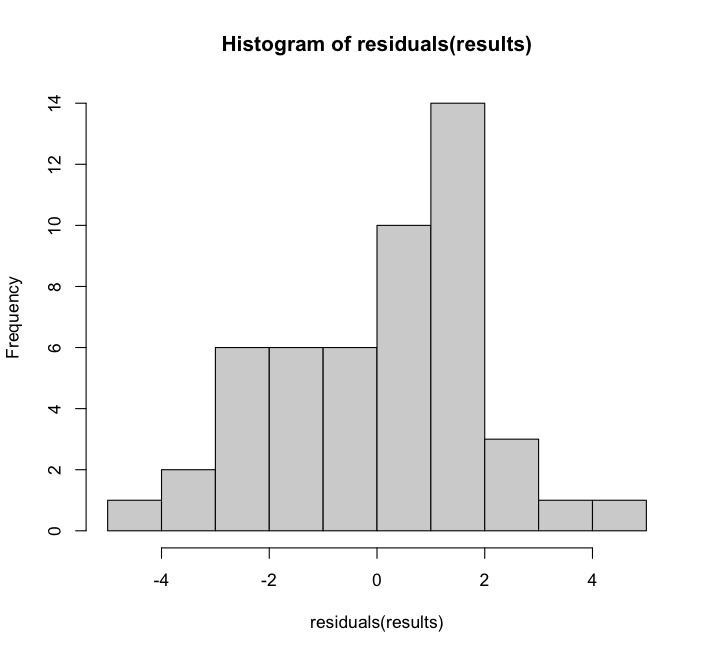
# (+/-0.35; +/-95% CI) decrease in domestic livestock livestock killed (p = < 2e-16; r2 = 0.43)

# normality of error was violated, and it was subtle. this is identified with a

# residuals histogram plot. however, this does not influence estimation

# histogram of residuals

hist(residuals(results))



#### number of livestock predicted to get killed when deer density is 6/ha

datumPredict <- data.frame(DeerDensity = 6)

predictions <- predict(results, datumPredict, interval = "prediction")

predictions

# 1.82 livestock/km2 (+/-4.22; +/-95% CI) would get killed at a native deer density of 6 deer/ha.

# make predictions across a range of densities

summary(datum)

# Minimum X was closer to 0, maximum was closer to 5

# Let's make predictions between 0 -- 5

# And let's try to make 20 predictions across this interval:

# 10 cm / 20 predictions = a prediction every 0.5 cm

xSequence <- seq(from = 0, to = 5, by = 0.25)

xSequence

# Turn this into a dataframe with a column called 'rainfall'

datumPredict <- data.frame(DeerDensity = xSequence)

datumPredict

# Make the predictions

predictions <- predict(results, datumPredict, interval = "prediction")

# Examine predictions

head(predictions)

# Make it a dataframe to make it easier to add data to it.

predictions <- as.data.frame(predictions)

# Add a new column called 'DeerDensity'

predictions$DeerDensity <- xSequence

# Examine

head(predictions)

# Very simple way to plot the prediction intervals

matplot(predictions$DeerDensity, predictions[,1:3], type = "l",

ylab = "Domestic livestock killed (livestock/km2)",

xlab = "Native deer density (deer/ha)")

# Plot the data, but this time specify what the x- and y-limits are.

# This is important to make sure everything will fit on here.

plot(CowsKilled ~ DeerDensity, data = datum, xlim = c(0, 5), ylim = c(0, 12),

ylab = "Domestic livestock killed (livestock/km2)",

xlab = "Native deer density (deer/ha)")

# But now we have to add other data onto here, without erasing the original data.

# Specify the parameter: new = TRUE

par(new = TRUE)

plot(fit ~ DeerDensity, data = predictions, xlim = c(0, 5), ylim = c(0, 12), type="l", ylab="", xlab="")

par(new = TRUE)

plot(lwr ~ DeerDensity, data = predictions, xlim = c(0, 5), ylim = c(0, 12), type="l", ylab="", xlab="")

par(new = TRUE)

plot(upr ~ DeerDensity, data = predictions, xlim = c(0, 5), ylim = c(0, 12), type="l", ylab="", xlab="")