

Class Activity 2 – SNOWGEESE

Leonardo L Sanchez

2026-01-21

Consider question 3.46 form the textbook and the data set: SNOWGEESE.

Botanists at the University of Toronto conducted a series of experiments to investigate the feeding habits of baby snow geese (Journal of Applied Ecology, Vol. 32, 1995). Goslings were deprived of food until their guts were empty, then were allowed to feed for 6 hours on a diet of plants or Purina Duck Chow. For each of 42 feeding trials, the change in the weight of the gosling after 2.5 hours was recorded as a percentage of initial weight. Two other variables recorded were digestion efficiency (measured as a percentage) and amount of acid-detergent fiber in the digestive tract (also measured as a percentage). The data for selected feeding trials are listed in the table below.

```
setwd("P:/langara/term 3/dana 4810")
SNOWGESE <- read.table("SNOWGESE.txt", header = TRUE)

#View(SNOWGESE)
str(SNOWGESE)

## 'data.frame':    42 obs. of  5 variables:
## $ Trial   : int  1 2 3 4 5 6 7 8 9 10 ...
## $ Diet     : chr  "Plants" "Plants" "Plants" ...
## $ WtChange: num  -6 -5 -4.5 0 2 3.5 -2 -2.5 -3.5 -2.5 ...
## $ DigEff   : num  0 2.5 5 0 0 1 2.5 10 20 12.5 ...
## $ ADFiber  : num  28.5 27.5 27.5 32.5 32 30 34 36.5 28.5 29 ...

sapply(SNOWGESE, class)

##      Trial      Diet     WtChange     DigEff     ADFiber
##   "integer" "character"  "numeric"  "numeric"  "numeric"
```

- 1) [1] List the variable(s) of interest in this study along with their types, list of categories, unit of measurements when appropriate. (Note: To identify how many categories and what categories the categorical variable(s) have, use an R code.)

```
# 1) Variables of interest: type, categories, units

# variable types
sapply(SNOWGESE, class)

##      Trial      Diet     WtChange     DigEff     ADFiber
##   "integer" "character"  "numeric"  "numeric"  "numeric"

# categorical variables + categories
categorical_vars <- names(SNOWGESE)[sapply(SNOWGESE, function(x) is.factor(x) || is.character(x))]
categorical_vars
```

```

## [1] "Diet"
lapply(SNOWGEESE[categorical_vars], unique)

## $Diet
## [1] "Plants" "Chow"

# quantitative variables
quantitative_vars <- names(SNOWGEESE) [sapply(SNOWGEESE, is.numeric)]
quantitative_vars

## [1] "Trial"      "WtChange"   "DigEff"     "ADFiber"

# units
units <- data.frame(
  Variable = c("WtChange", "DigEff", "ADFiber"),
  Unit = c(
    "Percent change in initial weight (%)",
    "Digestion efficiency (%)",
    "Acid-detergent fiber (%)"
  )
)
units

##   Variable           Unit
## 1 WtChange Percent change in initial weight (%)
## 2   DigEff   Digestion efficiency (%)
## 3   ADFiber Acid-detergent fiber (%)

```

The categorical variable has two categories: Chow and Plants

2) [1] Create a scatterplot and describe a possible relationship between weight change and digestion efficiency.

```

# closes any broken plot state
# if it errors, ignore and keep going

SNOWGEESE$Diet <- factor(SNOWGEESE$Diet)

model <- lm(WtChange ~ DigEff, data = SNOWGEESE)

plot(
  WtChange ~ DigEff,
  data = SNOWGEESE,
  pch = 19,
  col = as.integer(Diet),
  xlab = "Digestion Efficiency (%)",
  ylab = "Weight Change (%)",
  main = "Weight Change vs Digestion Efficiency"
)

legend(
  "topleft",
  legend = levels(SNOWGEESE$Diet),
  col = seq_along(levels(SNOWGEESE$Diet)),

```

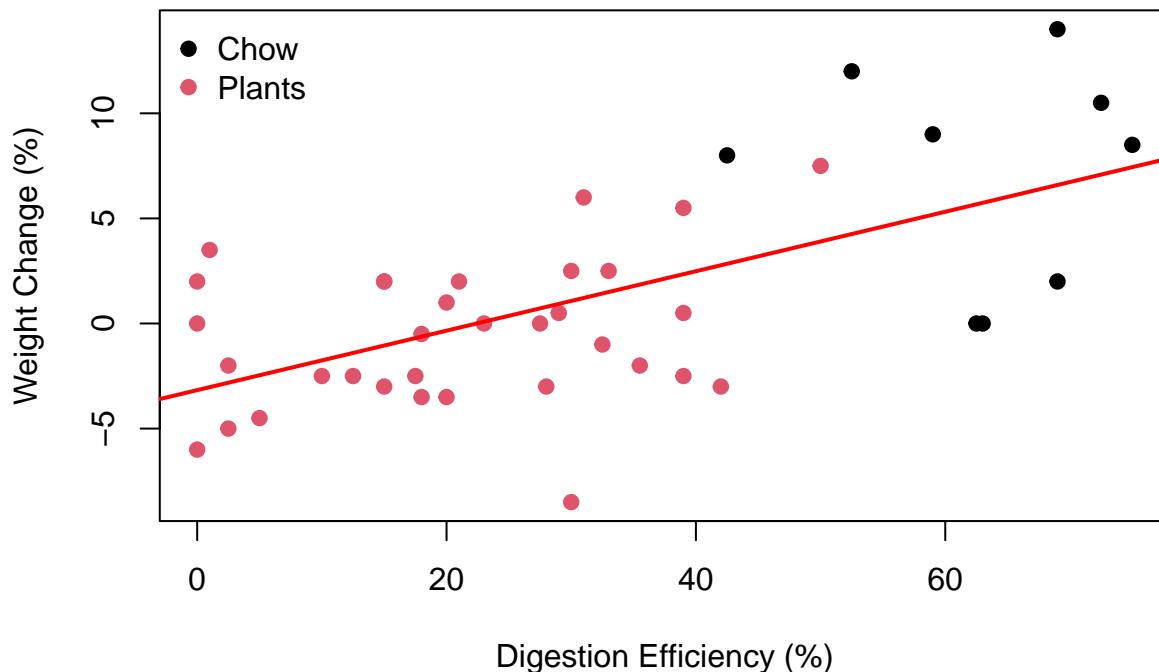
```

    pch = 19,
    bty = "n"
)

abline(model, col = "red", lwd = 2)

```

Weight Change vs Digestion Efficiency



It seems that Weight Change and Digestion Efficiency have a positive relationship. As Digestion Efficiency increases, Weight Change also tends to increase.

```

corWtChangevsDigEff <- cor(SNOWGEESE$WtChange, SNOWGEESE$DigEff)
corWtChangevsDigEff

```

3) [1] Calculate the coefficient of correlation relating weight change to digestion efficiency. Interpret this value.

```

## [1] 0.6122317

# 4) Hypothesis test for correlation (alpha = 0.01)
cor_test <- cor.test(
  SNOWGEESE$WtChange,
  SNOWGEESE$DigEff,
  alternative = "two.sided",
  conf.level = 0.99
)
cor_test

##
## Pearson's product-moment correlation
##
## data: SNOWGEESE$WtChange and SNOWGEESE$DigEff

```

```

## t = 4.8972, df = 40, p-value = 1.642e-05
## alternative hypothesis: true correlation is not equal to 0
## 99 percent confidence interval:
## 0.2913314 0.8092825
## sample estimates:
## cor
## 0.6122317

```

The correlation coefficient ($R = 0.6123$) suggests a moderate- strong positive correlation. This means that as Digestion Efficiency increases, Weight Change also tends to increase. Since the correlation is positive, it indicates that goslings with higher digestion efficiency generally gain more weight.

4) [7] Conduct a test to determine whether weight change is correlated with a digestion efficiency. Use alpha = :01.

```

cor_test <- cor.test(SNOWGEESE$WtChange, SNOWGEESE$DigEff, conf.level = 0.99)
cor_test

```

```

##
## Pearson's product-moment correlation
##
## data: SNOWGEESE$WtChange and SNOWGEESE$DigEff
## t = 4.8972, df = 40, p-value = 1.642e-05
## alternative hypothesis: true correlation is not equal to 0
## 99 percent confidence interval:
## 0.2913314 0.8092825
## sample estimates:
## cor
## 0.6122317

```

The p-value is much smaller than the significance level (p-value 0.00001642 vs $\alpha = 0.01$).

This means we reject the null hypothesis that there is no correlation between weight change and digestion efficiency.

There is strong statistical evidence that a significant relationship exists between these variables.

The positive correlation means goslings with higher digestion efficiency tend to gain more weight.

```

model <- lm(WtChange ~ DigEff, data = SNOWGEESE)
summary(model)

```

5) [4] Assuming the relationship between the variables mentioned in part (2) is best described by a straight line, use the method of least squares to estimate the y-intercept and slope of the line. (Also, provide the Least Squares Regression Line in the context of the question.)

```

##
## Call:
## lm(formula = WtChange ~ DigEff, data = SNOWGEESE)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -9.5733 -2.7288 -0.2575  3.1273  7.7436
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.7288    0.2575  10.60   <2e-16 ***
## DigEff      0.2575    0.0643   3.95   0.0001642 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```

## (Intercept) -3.17067    1.06784  -2.969  0.00503 **
## DigEff        0.14147    0.02889   4.897  1.64e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.003 on 40 degrees of freedom
## Multiple R-squared:  0.3748, Adjusted R-squared:  0.3592
## F-statistic: 23.98 on 1 and 40 DF,  p-value: 1.642e-05
anova(model)

## Analysis of Variance Table
##
## Response: WtChange
##             Df Sum Sq Mean Sq F value    Pr(>F)
## DigEff      1 384.24 384.24  23.982 1.642e-05 ***
## Residuals  40 640.88   16.02
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
intercept <- coef(model)[1]
slope <- coef(model)[2]

```

The Least Squares Regression Line is:

$$\text{WtChange} = -3.17 + 0.1414 * \text{DigEff}$$

- 6) [2] Provide an interpretation of the slope estimate and intercept of the Least Square Regression Line in the context of the question. The slope (0.1414) indicates that for each 1% increase in digestion efficiency, the goslings' weight is expected to increase by 0.1414%.

```
confint(model, level = 0.90)
```

- 7) [4] Find a 90% confidence interval for the slope and provide an interpretation for the calculated interval.

```

##               5 %         95 %
## (Intercept) -4.96874788 -1.3725857
## DigEff       0.09282487  0.1901092
intercept <- coef(model)[1]
slope <- coef(model)[2]

intercept

## (Intercept)
## -3.170667
slope

## DigEff
## 0.141467

```

We are 90% confident that the true slope of the relationship between Weight Change and Digestion Efficiency lies between 0.0928 and 0.1901. This means that for every 1% increase in Digestion Efficiency, the goslings' Weight Change is expected to increase between 0.0928% and 0.1901% of their initial weight. Since the confidence interval does not include 0, it confirms that there is a statistically significant positive relationship between the two variables.

```

new_data <- data.frame(DigEff = 17)
conf_pred <- predict(model, new_data, interval = "confidence", level = 0.90)
print(conf_pred)

```

8) [4] Find a 90% confidence interval for the mean weight change of all goslings with digestion efficiency of 17%. Interpret the result.

```

##          fit      lwr      upr
## 1 -0.7657271 -1.986815 0.4553605

```

We are 90% confident that the mean value of weight change for goslings with digestion efficiency of 17% is between -1.9868% and 0.4553%

```

pred_interval <- predict(model, new_data, interval = "prediction", level = 0.90)

print(pred_interval)

```

9) [4] Find a 90% prediction interval for the weight change of a gosling with digestion efficiency of 17%. Interpret the result.

```

##          fit      lwr      upr
## 1 -0.7657271 -7.615459 6.084004

```

With 90% confidence, we predict that the weight change for a single gosling with Digestion Efficiency of 17% falls between -7.6154% and 6.084%.

```

plot(
  SNOWGEESE$DigEff,
  SNOWGEESE$WtChange,
  main = "Regression Line with Confidence and Prediction Intervals",
  xlab = "Digestive Efficiency (%)",
  ylab = "Weight Change (%)",
  pch = 19,
  col = "blue"
)

abline(model, col = "red", lwd = 2)

dig_eff_seq <- seq(min(SNOWGEESE$DigEff), max(SNOWGEESE$DigEff), length.out = 100)
pred_values <- data.frame(DigEff = dig_eff_seq)

conf_band <- predict(model, newdata = pred_values, interval = "confidence", level = 0.90)
pred_band <- predict(model, newdata = pred_values, interval = "prediction", level = 0.90)

lines(dig_eff_seq, conf_band[,2], col = "green", lty = 2)
lines(dig_eff_seq, conf_band[,3], col = "green", lty = 2)

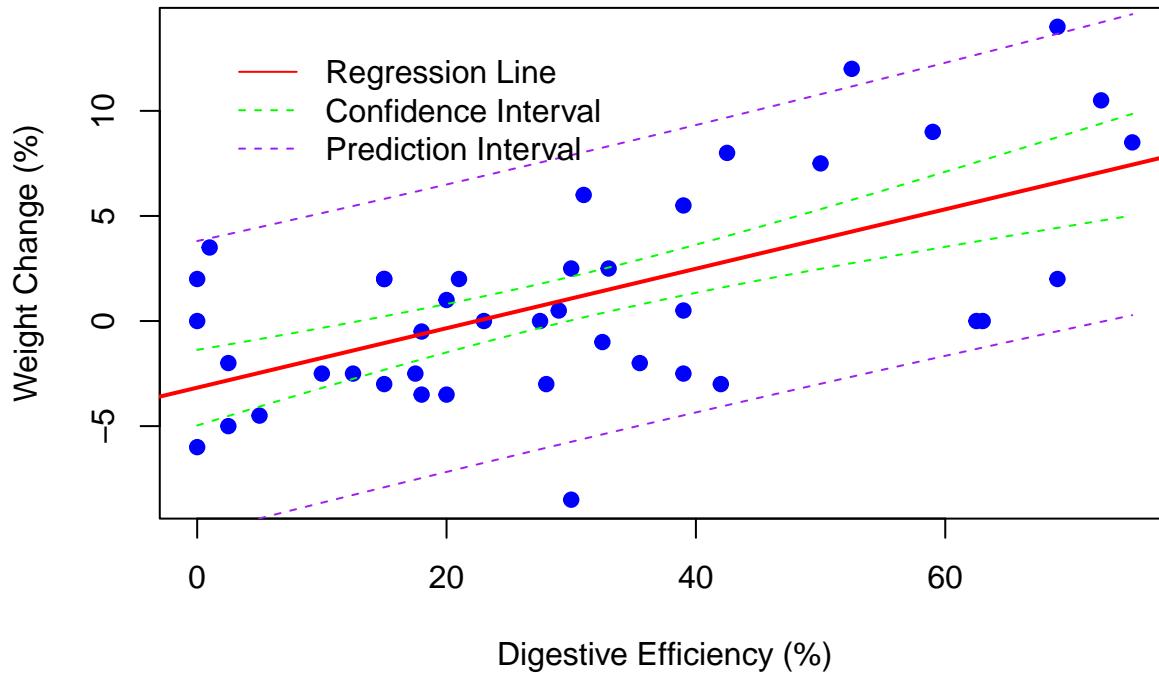
lines(dig_eff_seq, pred_band[,2], col = "purple", lty = 2)
lines(dig_eff_seq, pred_band[,3], col = "purple", lty = 2)

legend("topleft", legend = c("Regression Line", "Confidence Interval", "Prediction Interval"),
       col = c("red", "green", "purple"), lty = c(1,2,2), bty = "n", inset = 0.05)

```

10) [5] Plot the least squares line on your scatterplot and depict 90% confidence intervals for mean values and prediction intervals for new values over the entire range of the regression line.

Regression Line with Confidence and Prediction Intervals



```
r_squared <- summary(model)$r.squared  
r_squared
```

11) [1] Find the coefficient of determination from the RStudio output and interpret the value.

```
## [1] 0.3748277
```