MATH 2310 Confidence Intervals

Jakob Balkovec

2024-05-17

For this lab, we will be working with the same snow geese data that we used in our first lab assignment. The columns are not labeled in this data set. The first variable is just an index of the trial number. The second variable is a description of the type of diet fed to the goose. The third variable is their percent weight change after being allowed to feed for 2.5 hours. The fourth variable is their digestion efficiency. The fifth variable is the amount of acid-detergent fiber in their digestive tract.

Goals for this Assignment

- Refresh how to calculate summary statistics in R
- Practice calculating confidence intervals
- Interpret results of confidence intervals

Activity One

```r
suppressPackageStartupMessages({
  library(dplyr)
  library(readxl)
  library(knitr)
})

library(readxl)

path <- "Snowgeese.xls"
snowgeese <- suppressMessages(read_excel(path, col_names = FALSE))

colnames(snowgeese) <- c("Index", "Plants", "PercentWeightChange", "DigestionEfficiency", "AcidDetergentFiber")

# Subsets
chow_geese <- snowgeese %>% filter(Plants == "Chow")
plants_geese <- snowgeese %>% filter(Plants == "Plants")
```

We will first analyze the weight change for the snow geese.

Skill Objective:

Use R to calculate the means and standard deviations for weight change for geese fed a plant diet, and for geese fed a chow diet.

```r
# Data: snowgeese (type: "tbl_df" || "tbl" || "data.frame")

# Chow fed geese - diet type
chow_pwc <- chow_geese$PercentWeightChange # Subset

chow_mean <- mean(chow_pwc, na.rm = FALSE)
chow_sd <- sd(chow_pwc, na.rm = FALSE)


# Plant fed geese - diet type
plants_pwc <- plants_geese$PercentWeightChange # Subset

plants_mean <- mean(plants_pwc, na.rm = FALSE)
plants_sd <- sd(plants_pwc, na.rm = FALSE)

# Combine results into a data frame
results_diet <- data.frame(
  Group = c("Chow", "Plants"),
  Mean = c(round(chow_mean, digits = 2), round(plants_mean, digits = 2)),
  SD = c(round(chow_sd, digits = 2), round(plants_sd, digits = 2))
)

kable(results_diet, col.names = c("Group", "Mean", "Standard Deviation"), caption = "Summary Statistics for Perce
nt Weight Change")
```

Summary Statistics for Percent Weight Change

Group	Mean	Standard Deviation
Chow	7.11	5.20
Plants	-0.55	3.51

Then, using these results, calculate a 99% confidence interval for the mean weight change for geese fed a plant diet, and a 99% confidence interval for the mean weight change for geese fed a chow diet.

```r
# Document
get_critical_value <- function(alpha = 0.01) {

  return (qnorm(1 - alpha/2))
}

# Document
confidence_interval <- function(mean, sd, n, alpha = 0.01) {

  sem <- sd / sqrt(n)
  margin_of_error <- get_critical_value(alpha) * sem
  ci_lower <- mean - margin_of_error
  ci_upper <- mean + margin_of_error

  return(list(ci_lower = ci_lower, ci_upper = ci_upper))
}

# Chow confidence interval
chow_ci <- confidence_interval(chow_mean, chow_sd, nrow(chow_geese))

# Plants confidence interval
plants_ci <- confidence_interval(plants_mean, plants_sd, nrow(plants_geese))

# Combine results into a data frame
confidence_intervals <- data.frame(
  Group = c("Chow", "Plants"),
  CI_Lower = c(round(chow_ci$ci_lower, digits = 2), round(plants_ci$ci_lower, digits = 2)),
  CI_Upper = c(round(chow_ci$ci_upper, digits = 2), round(plants_ci$ci_upper, digits = 2))
)

# Lower: 99% confidence interval lower bound
# Upper: 99% confidence interval upper bound
kable(confidence_intervals, col.names = c("Group", "Lower", "Upper"), caption = "99% Confidence Intervals for Per
cent Weight Change")
```

99% Confidence Intervals for Percent Weight Change

Group	Lower	Upper
Chow	2.65	11.58
Plants	-2.12	1.03

```
##  Chow Fed Geese -> CI :  ( 2.65 ,  11.58 )
##  Plant Fed Geese -> CI : ( -2.12 ,  1.03 )
```

Analysis Objective:

Is there evidence that geese fed a plant diet will, on average, either gain or lose weight? What about for geese fed a chow diet?

Analysis:

Chow Fed Geese:
Confidence Interval: (2.65, 11.58)

Interpretation:

With 99% confidence, we can estimate that the true mean weight change for **chow-fed geese** lies somewhere between 2.65 and 11.58 percent. Since both bounds are positive, it suggests that, on average, chow-fed geese are likely to gain weight.

Plant Fed Geese:
Confidence Interval: (-2.12, 1.03)

Interpretation:

With 99% confidence, we can estimate that the true mean weight change for **plant-fed geese** lies between -2.12 and 1.03 percent. Since this interval includes zero, it suggests that there is no strong evidence to conclude that plant-fed geese will either gain or lose weight on average, we can only speculate. Since the “scale” tips towards the lower bound we can assume that **plant-fed geese** are more likely to lose weight.

Conclusion:

Chow Fed Geese: There is evidence to suggest that chow-fed geese will, on average, gain weight.
Plant Fed Geese: There is really no strong evidence to conclude that plant-fed geese will either gain or lose weight on average.

Activity Two

Now we will examine digestion efficiency.

Skill Objective:

Use R to calculate the means and standard deviations for digestion efficiency for geese fed a plant diet, and for geese fed a chow diet.

```r
# Chow fed geese - digestion efficiency (de)
chow_de <- chow_geese$DigestionEfficiency

chow_de_mean <- mean(chow_de, na.rm = FALSE)
chow_de_sd <- sd(chow_de, na.rm = FALSE)

# Plant fed geese - digestion efficiency (de)
plants_de <- plants_geese$DigestionEfficiency

plants_de_mean <- mean(plants_de, na.rm = FALSE)
plants_de_sd <- sd(plants_de, na.rm = FALSE)

# Combine results into a data frame
results_de <- data.frame(
  Group = c("Chow", "Plants"),
  Mean = c(round(chow_de_mean, digits = 2), round(plants_de_mean, digits = 2)),
  SD = c(round(chow_de_sd, digits = 2), round(plants_de_sd, digits = 2))
)

kable(results_de, col.names = c("Group", "Mean", "Standard Deviation"), caption = "Summary Statistics for Digesti
on Efficiency")
```

Summary Statistics for Digestion Efficiency

Group	Mean	Standard Deviation
Chow	62.78	10.32
Plants	21.26	13.87

Then, using these results, calculate a single 99% confidence interval for the difference in mean digestion efficiencies between these two groups.

```r
std_dev <- function(n_chow, n_plants) {
  return (pooled_sd <- sqrt(((n_chow - 1) * chow_de_sd^2 + (n_plants - 1) * plants_de_sd^2) / (n_chow + n_plants
- 2)))
}

std_err <- function(pooled_sd) {
  return(pooled_sd * sqrt(1/n_chow + 1 / n_plants))
}

n_chow <- length(chow_de)
n_plants <- length(plants_de)

pooled_sd <- std_dev(n_chow, n_plants)
se_diff <- std_err(pooled_sd)

# 99% Confidence interval
alpha <- 0.01
crit_value <- qnorm(1 - alpha / 2)

margin_of_error <- crit_value * se_diff
diff_in_means <- chow_de_mean - plants_de_mean

confidence_interval <- data.frame(
  Difference = round(diff_in_means, digits = 2),
  CI_Lower = round((diff_in_means - margin_of_error), digits = 2 ),
  CI_Upper = round((diff_in_means + margin_of_error), digits = 2)
)

kable(confidence_interval, col.names = c("Difference in Means", "Lower", "Upper"), caption = "99% Confidence Inte
rval for Digestion Efficiency")
```

99% Confidence Interval for Digestion Efficiency

Difference in Means	Lower	Upper
41.52	28.7	54.34

```
## Plant & Chow Fed Geese -> CI : ( 28.7 ,  54.34 )
```

Analysis Objective:

Based on your confidence interval, what can you say about how the average digestion efficiency compares between the two diets?

Analysis:

Single 99% confidence interval for the difference in mean digestion efficiencies between Chow and Plants fed geese:
Confidence Interval: (28.7 , 54.34)

Interpretation:

The confidence interval suggests that we can be 99% confident that the true difference in mean digestion efficiencies between geese fed a plant diet and those fed a chow diet lies between 28.7 and 54.34 percent.

Conclusion:

Since the confidence interval does not include zero, we can conclude that there is a statistically significant difference in the average digestion efficiency between the two diets. Specifically, geese fed a chow diet have a significantly higher average digestion efficiency compared to those fed a plant diet.