Group Activity

Peyton Hall

2024-03-15

```
rm(list=ls())
```

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.4.4     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(tidyr)
library(readxl)
library(dplyr)
```

```
# Group Activity 01
# Question 1. Use the piggy data on d2l.
Piggy <- read_excel("~/Desktop/Data211/Week 10/Piggy.xlsx")
Piggy
```

```
## # A tibble: 16 x 6
##    pigID week0 week4 week8 week12 week16
##    <dbl> <dbl> <dbl> <dbl>  <dbl>  <dbl>
##  1     1   1     4.8  12.6   16     21
##  2     2   1     4.2   7     10     14
##  3     3   0.8   4     6      9     10
##  4     4   0.8   4     6      9     13
##  5     5   0.8   5     9.4   11     14
##  6     6   0.8   3.2   7     10     15
##  7     7   0.8   3.2   5.5    9.6   12
##  8     8   0.8   3.4   7      8.7   12.4
##  9     9   1     5.4  10     13     17.4
## 10    10   1.2   4.8  12.6   16     20
## 11    11   1     4.6  11     18     22
## 12    12   0.8   4.2   8     11     13
## 13    13   0.8   3.8   7      8.9   12
```

```
## 14    14   1     5.4  11     14     19
## 15    15   1     6     5.4   10     17
## 16    16   1     3.4   7.8   10     13
```

```
# a) Is that data a long format or a wide format?
#    - wide because you can see that the ids have single instances of each one,
#      and it is split apart into weeks.

# columns: "pigID", "week0", "week4", "week8", "week12", "week16"
Piggy_long <- pivot_longer(Piggy,
                           cols = -pigID,  # Columns to pivot (excluding pigID)
                           names_to = "Time",  # New column name for time
                           values_to = "Weight")  # New column name for values
Piggy_long
```

```
## # A tibble: 80 x 3
##    pigID Time   Weight
##    <dbl> <chr>   <dbl>
##  1     1 week0     1
##  2     1 week4     4.8
##  3     1 week8    12.6
##  4     1 week12   16
##  5     1 week16   21
##  6     2 week0     1
##  7     2 week4     4.2
##  8     2 week8     7
##  9     2 week12   10
## 10     2 week16   14
## # i 70 more rows
```

```
# Now Piggy_long will have a long format where each row represents one observation for each pig at a sp


# Question 2. Use the Diabete data on d2l.
# library(readxl)
Diabete <- read_excel("~/Desktop/Data211/Week 10/Diabete.xlsx")
# View(Diabete)
# a) Change the format with a variable SBP showing the systolic blood pressure at different
# times.
# Note: it is in long format.
# Columns: "ID", "Time", "SBP"
# Pivot from long to wide format
Diabete_wide <- pivot_wider(Diabete,
                            id_cols = ID,  # Column(s) to keep as identifier(s)
                            names_from = Time,  # Values from "Time" will become column names
                            values_from = SBP)  # Values from "SBP" will populate the wide format colum
Diabete_wide
```

```
## # A tibble: 18 x 4
##       ID Baseline followone followtwo
##    <dbl>    <dbl>     <dbl>     <dbl>
##  1     1      130       120       115
##  2     2      140       130       124
```

```
##  3     3      145       130       125
##  4     4      140       120       110
##  5     5      150       125       113
##  6     6      145       120       115
##  7     7      140       110       106
##  8     8      145       112       109
##  9     9      155       115       110
## 10    10      140       106       100
## 11    11      136       109       101
## 12    12      139       112       101
## 13    13      150       130       121
## 14    14      145       130       120
## 15    15      155       135       126
## 16    16      140       110       110
## 17    17      142       115       114
## 18    18      145       119       112
```

```r
# Group Activity 02:
# Use the Tree data from d2l, find the mean for each variable of group, branch
# density, branch thickness, and tree heights.
# (Hint: use for loop to identify the Group, and filter() to filter each group before you
# calculate mean of each group)

Tree_data <- read_excel("~/Desktop/Data211/Week 10/Tree data.xlsx")
# View(Tree_data)

means_list <- list() # Initialize a list to store means for each group

# Loop through each group
for (i in 1:3) {
  # Filter data for the current group using filter() function
  group_data <- filter(Tree_data, Group == i)
  # Calculate mean for each variable within the group
  means <- sapply(group_data[,2:4], mean)
  # Store means in the list
  means_list[[paste0("Group_", i)]] <- means
} # end for

# Combine means from the list into a data frame
means_df <- do.call(rbind, means_list)
# Rename rows and columns
row.names(means_df) <- paste0("Group_", 1:3)
# Rename columns
colnames(means_df) <- c("Meandensity", "Meanthickness", "Meanheight")
# Print the means for each group
means_df
```

```
##         Meandensity Meanthickness Meanheight
## Group_1       3.775          3.40       13.5
## Group_2       4.600          2.85       12.5
## Group_3       2.225          6.55       12.5
```