W25_HW2

Cameron Bentz

February 20, 2025

Question 1

Find the distance between objects 1 and 3 by using the formula provided on the slides. Notice that we have mixed type of attributes. (You can scan and submit your handwritten calculation)

Object Identifier	test-1 (nominal)	test-2 (ordinal)	test-3 (numeric)
1	A	excellent	45
2	B	fair	22
3	C	good	64
4	A	excellent	28

Manual calculation of mixed-type distance

- Nominal Distance: A vs C = 1 (different)
- Ordinal Distance: excellent (3) vs good (2) = 1
- Numeric Distance: |45-64| = 19

Distance Formula:

$$Distance = \sqrt{w_1(d_{nominal})^2 + w_2(d_{ordinal})^2 + w_3(d_{numeric})^2}$$

Assuming equal weights:

$$Distance = \sqrt{1^2 + 1^2 + 19^2}$$

```
distance <- sqrt(1^2 + 1^2 + 19^2)
cat("Distance between Object 1 and Object 3:", distance, "\n")
```

```
## Distance between Object 1 and Object 3: 19.05256
```

Question 2

Write a program in any language which can compute Manhattan and Euclidean distances between any two given vectors with any length. You can pass the length to your function, but please don't limit the dimension to 2. You can test your function on vectors you fill in your code without asking user input.

Manhattan Distance Function

$$D_{Manhattan} = \sum_{i=1}^{n} |a_i - b_i|$$

```
manhattan_distance <- function(a, b) {
  return(sum(abs(a-b)))
}
```

Euclidean Distance Function

$$D_{Euclidean} = \sqrt{\sum_{i=1}^{n} (a_i - b_i)^2}$$

```
euclidean_distance <- function(a, b) {
  return(sqrt(sum((a-b)^2)))
}
```

```
vector1 <- c(1, 2, 3)
vector2 <- c(4, 5, 6)

cat("Manhattan Distance:", manhattan_distance(vector1, vector2), "\n")
```

Test the functions

```
## Manhattan Distance: 9
```

```
cat("Euclidean Distance:", euclidean_distance(vector1, vector2), "\n")
```

```
## Euclidean Distance: 5.196152
```

Question 3

In the table below, determine whether passing a class has a dependency on attendance by using Chi-square test.Please refer to the formula in the slides.
(For the expected value for each cell, multiply the total counts in the rows and columns of the cell and divide by total count.
For example: Expected value for Attended-Pass=33*31/54 = 18.94. You can scan and submit your hand-written calculation)

	Passed	Failed	**Total**
Attended	25	6	**31**
Skipped	8	15	**23**
Total	**33**	**21**	**54**

Expected values

$$E = \frac{(\text{Row Total} \times \text{Column Total})}{\text{Grand Total}}$$

- Attended-Pass = (31 x 33)/54 = 18.94
- Attended-Fail = (31 x 21)/54 = 12.06
- Skipped-Pass = (23 x 33)/54 = 14.06
- Skipped-Fail = (23 x 21)/54 = 8.94

```r
# Value frequencies
values <- matrix(c(25, 6, 8, 15), nrow = 2, byrow = TRUE)
colnames(values) <- c("Passed", "Failed")
rownames(values) <- c("Attended", "Skipped")

# Chi-Square Test
chisq_result <- chisq.test(values, correct = FALSE)
chisq_result
```

Chi-Square Test of Independence

```
##
##  Pearson's Chi-squared test
##
## data:  values
## X-squared = 11.686, df = 1, p-value = 0.0006297
```

Question 4

In R, there is a built-in data frame called mtcars. Please calculate the correlation between mpg and wt attributes of mtcars by using cor() function. Then generate scatter plot based on these two attributes. Your scatter plot should be like the one below. You don't need to submit the image, but R script should be submitted.

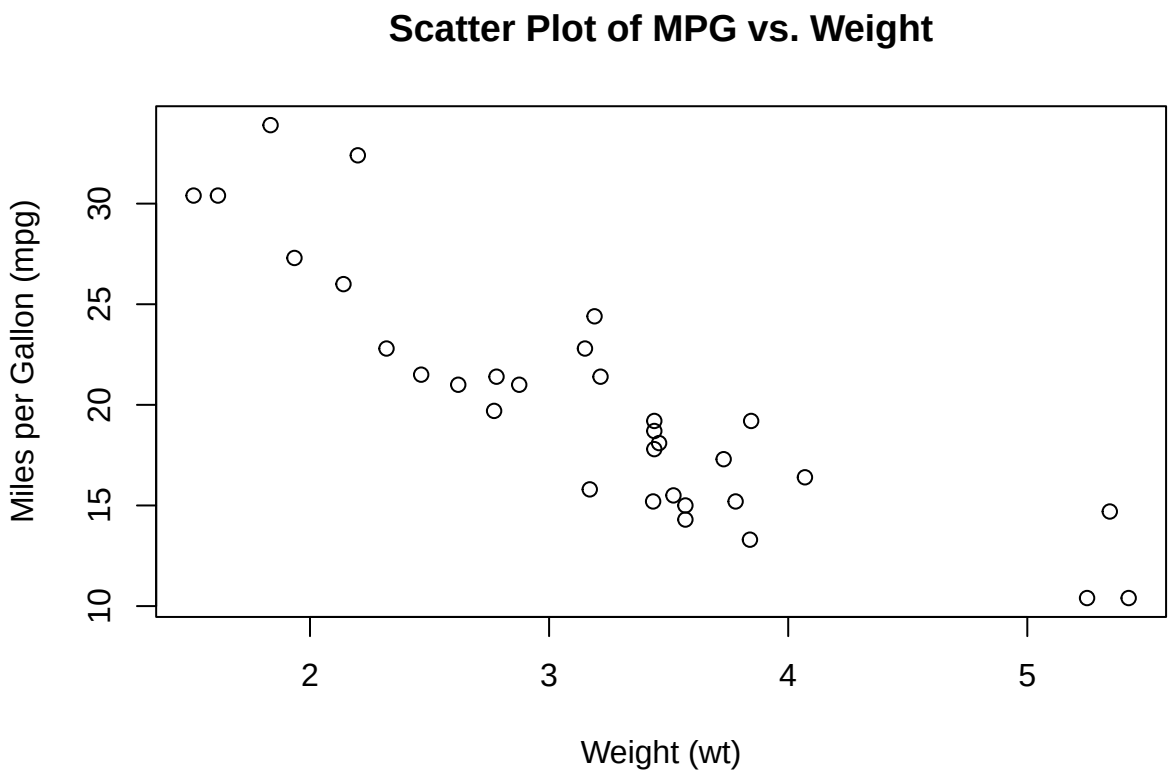
```r
# Load Dataset
data(mtcars)

# Calculate Correlation
correlation <- cor(mtcars$mpg, mtcars$wt)
cat("Correlation between Miles Per Gallon and weight:", correlation, "\n")
```

Calculate correlation of mpg and wt

```
## Correlation between Miles Per Gallon and weight: -0.8676594
```

```r
plot(mtcars$wt, mtcars$mpg,
     main = "Scatter Plot of MPG vs. Weight",
     xlab = "Weight (wt)",
     ylab = "Miles per Gallon (mpg)",
     col = "black",
     pch = 21)
```



Scatter plot