Assessed Activity 2

Omoegbemwen Aigbe 2301397

2023-10-18

# Assessed Activity 2

## Mathematics for the Biosciences

## MOD005667

## 2023/24 Tri 1

# Instructions

You have been provided with a data table that contains patient data from The Highbury and Hale End NHS Trust. The data table contains the patient weight, and the twice-daily amoxicillin dosage administered to patients with persistent otitis media in 2022. Your task is to audit the data to determine whether the hospital administered the correct dosage to each patient. To do so, you will have to analyse the data using your R and RStudio skills.

* You will need to complete this R markdown document by filling in your response to each of the questions.
* You will need to ‘knit’ your completed R markdown document to a Word document and submit the Word document to Canvas. **The final document must be submitted as a knitted Word document.**
* If the question requires you to write free text (i.e., the questions asking for a formula or the questions asking for your interpretation), write your response into the document in **bold** in the space provided below the question so that it is easy for me to see what you have done at a glance.
* If the question requires you to use R, insert an R block in the space below the question and complete your answer in the R block.
* Read the information provided closely. You will have to interpret this information to perform the audit correctly.
* **You may use chatGPT to help with any R code needed to complete this assessment.**
* When writing prompts for chatGPT, be as specific as possible. I have given clues in the wording of the questions to help you with this.
* It is a good idea to tell chatGPT the names of the variables in your workspace. For example, if your data frame is called “df”, tell chatGPT this so it can incorporate the correct data frame name in the code it provides you. Similarly, if the column in your data frame you want to work on is called “kg”, tell chatGPT this.
* It is also a good idea to tell chatGPT what you want to name functions, variables, and data columns.
* **You may not use chatGPT to answer any of the questions that do not require you to use R. These include the questions asking for formulas and interpretations.**

# Treatment of Otitis Media in Children with Amoxicillin

Otitis media is an infection of the middle ear common in infants and children. It causes inflammation and a build-up of fluid behind the eardrum. Persistent otitis media can be treated with the antibiotic amoxicillin. In 2022, The Highbury and Hale End NHS Trust treated 1,536 children with amoxicillin for persistent otitis media.

**Your task is to audit the Highbury and Hale End NHS Trust to determine if they correctly administered amoxicillin to their patients.**

The guidelines for treatment of otitis media in children is to administer 50 mg/Kg/day of amoxicillin. This dosage should be split over 2 doses per day by administering a certain volume of 50 mg/ml amoxicillin solution. Total daily dosages received should not exceed 1500 mg/day.

Miscalculations of the required dosage can lead to over or under-treatment, which in turn can lead to a range of side effects and consequences. You are performing this audit to identify any miscalculations in the treatments received by the children.

You have been given a data table in tab-delimited format. The first line of the file is a header that describes each data column, and each subsequent line consists of the data for one of the patients. There are three columns in the data called “patient”, “kg”, and “twice\_daily\_dose\_given\_ml”, which respectively give the patient ID, the patient weight in kilograms, and the twice daily dosage that each patient was administered. The twice-daily dose is given as the number of ml of 50 mg/ml solution administered.

# Exercise 1. Performing the Audit (60 pts)

Begin by editing the R block below to read in the data table and saving it as a data frame called ‘df’. You will not be marked on this task.

Loading the data into the table. The data is loaded and saved as ‘df’ and using a R code to produce the first 6 rows of the data to check if the data has been read into the markdown document

# read in the file You will have to edit the code below to give the correct path to the file on your computer  
  
file\_path <- "C:/Users/User/Downloads/assessment\_2\_patient\_data (1).txt"  
  
# Read the file into R as df  
df <- read.table(file\_path, header = TRUE)  
head(df)

## patient kg twice\_daily\_dose\_given\_ml  
## 1 patient\_1 16.1 8.05  
## 2 patient\_2 26.8 13.40  
## 3 patient\_3 13.5 6.75  
## 4 patient\_4 14.4 7.20  
## 5 patient\_5 15.2 7.60  
## 6 patient\_6 33.9 15.00

1. Write the mathematical formula for calculating the correct dosage (in mg) from the patient’s weight. For this, ignore the instruction that daily dosages cannot exceed 1500 mg/day

**Solution**

**From the question, the following constant was given:**

**The desired Dosage per kg = 50mg/kg/day**

**Dosage Frequency = 2 dosages per day**

**Volume of amoxicillin = 50mg/ml**

**Maximum total daily dosage = 1500**

**maximum weight for the maximum dosage to be correct = 1500/50 = 30kg**

**Patient-weight =?**

**To calculate the correct dosage from the patient's weight**

**correct dosage = Patient weight \* Desired dosage**

**From this formula, we can also derive that.**

**Patient weight = Total daily dosage / desired dosage**

1. Write an algorithm that combines your mathematical formula from 1a with the rule that daily dosages cannot exceed 1500 mg/day. This should take the form of:  
   If the patient weight is less than *x* kg then dosage is *y* mg*.*  
   Else if the patient weight is greater than *x* kg then dosage is *z* mg*.*

**Solution**

**An algorithm that gives a detailed description of how to calculate the required daily dosage using the mathematical formula and also satisfies that the dosage cannot exceed 1500mg.**

**The algorithm was written as a Pseudocode:**

**Mathematically,**

**Step 1: Start**

**Step 2: Input desired dosage (desired\_dosage)**

**Step 3: Input total daily dosage (total\_daily\_dosage)**

**Step 4: Let maximum\_weight = total\_daily\_dosage / desired\_dosage**

**Step 5: Create a conditional statement:   
- If patient weight (patient\_weight) is less than or equal to maximum\_weight, then - Set required dosage = patient\_weight \* desired\_dosage   
- Else, - Set required dosage = total\_daily\_dosage**

**Step 6: Print the required dosage.**

**Step 7: Let required twice daily dose = required dosage/2**

**Step 7: End**

calculate\_amoxicillin\_dosage <- function(patient\_weight) {  
 desired\_dosage <- 50  
 total\_daily\_dosage <- 1500  
  
 maximum\_weight <- total\_daily\_dosage / desired\_dosage  
  
 if (patient\_weight <= maximum\_weight) {  
 required\_dosage <- patient\_weight \* desired\_dosage  
 } else {   
 required\_dosage <- total\_daily\_dosage  
 }  
  
 cat("Required dosage:", required\_dosage, "mg\n")  
  
 twice\_daily\_required\_dosage <- required\_dosage / 2  
 cat("Twice daily required dosage:", twice\_daily\_required\_dosage, "mg\n")  
}

#testing the function.

#when the patient's weight is less than 30  
calculate\_amoxicillin\_dosage(23)

## Required dosage: 1150 mg  
## Twice daily required dosage: 575 mg

#when the patient's weight is greater than 30  
calculate\_amoxicillin\_dosage(37)

## Required dosage: 1500 mg  
## Twice daily required dosage: 750 mg

1. Calculate the correct twice-daily dose required for each patient in and save this to a new column in your data frame. To do so, write a function that takes weight in kg as an input and outputs the required twice daily dose in mg according to the algorithm you wrote in **1b**. Once you have your function, you can use the mapply() function to apply your function to the entire data set. Show your R code and use the head () command to print the first six lines of your data frame to your document.

calculate\_amoxicillin\_dosage <- function(patient\_weight) {  
 desired\_dosage <- 50  
 total\_daily\_dosage <- 1500  
  
 maximum\_weight <- total\_daily\_dosage / desired\_dosage  
  
 if (patient\_weight <= maximum\_weight) {  
 required\_dosage <- patient\_weight \* desired\_dosage  
 } else {  
 required\_dosage <- total\_daily\_dosage  
 }  
 return(required\_dosage / 2)  
}  
  
  
df$twice\_daily\_dose\_required\_mg <- mapply(calculate\_amoxicillin\_dosage, df$kg)  
  
  
head(df)

## patient kg twice\_daily\_dose\_given\_ml twice\_daily\_dose\_required\_mg  
## 1 patient\_1 16.1 8.05 402.5  
## 2 patient\_2 26.8 13.40 670.0  
## 3 patient\_3 13.5 6.75 337.5  
## 4 patient\_4 14.4 7.20 360.0  
## 5 patient\_5 15.2 7.60 380.0  
## 6 patient\_6 33.9 15.00 750.0

1. Write the formula for calculating the correct volume of 50 mg/ml solution that each patient should receive twice daily as a function of the twice-daily dosage.

**Solution**

**Volume of required dosage in ml = twice daily dose required mg/volume of amoxicillin,**

**Recall that the volume of amoxicillin = 50**

**therefore, the formula is.**

**The volume of required dosage in ml = twice daily dose required mg /50**

1. Use the formula you wrote in **1d** to calculate the correct volume of 50 mg/ml solution each patient should receive twice daily and save this to a new column in your data frame. This is a simple calculation, and you should be able to calculate by applying the maths directory to the data column you calculated in **1c**. Show your R code, and use the head() command to print the first six lines of your data frame to your document.

df$correct\_volume\_ml <- df$twice\_daily\_dose\_required\_mg / 50  
  
head(df)

## patient kg twice\_daily\_dose\_given\_ml twice\_daily\_dose\_required\_mg  
## 1 patient\_1 16.1 8.05 402.5  
## 2 patient\_2 26.8 13.40 670.0  
## 3 patient\_3 13.5 6.75 337.5  
## 4 patient\_4 14.4 7.20 360.0  
## 5 patient\_5 15.2 7.60 380.0  
## 6 patient\_6 33.9 15.00 750.0  
## correct\_volume\_ml  
## 1 8.05  
## 2 13.40  
## 3 6.75  
## 4 7.20  
## 5 7.60  
## 6 15.00

1. You should now be able to perform the audit by comparing the correct volume of 50 mg/ml solution to the actual dosage given in the data file. One way to compare the columns would be to subtract the correct dosage column you created from the dosage given column. If the correct dosage was administered, these values should be the same, and the result would be zero. If the patient received too much amoxicillin, the result would be a positive value, giving the number of ml of solution the patient was over-dosed, and if too little amoxicillin were given, the result would be a negative number indicating the number of ml of solution the patient was underdosed.  
   Create a new data column in your data frame that is the difference between the dosage given and the correct dosage. Again, show your R code and print the first six lines of the data frame to your document using the head () command.

**Solution**

df$difference\_in\_dosage\_ml <- df$correct\_volume\_ml - df$twice\_daily\_dose\_given\_ml   
   
head(df)

## patient kg twice\_daily\_dose\_given\_ml twice\_daily\_dose\_required\_mg  
## 1 patient\_1 16.1 8.05 402.5  
## 2 patient\_2 26.8 13.40 670.0  
## 3 patient\_3 13.5 6.75 337.5  
## 4 patient\_4 14.4 7.20 360.0  
## 5 patient\_5 15.2 7.60 380.0  
## 6 patient\_6 33.9 15.00 750.0  
## correct\_volume\_ml difference\_in\_dosage\_ml  
## 1 8.05 0  
## 2 13.40 0  
## 3 6.75 0  
## 4 7.20 0  
## 5 7.60 0  
## 6 15.00 0

1. In question **1f,** you calculated the numerical value of under-dosing and overdosing of each patient. You can also evaluate each patient categorically to determine the number of patients under-dosed, overdosed, or correctly dosed. Write an R function that takes as input the difference between the correct dose and the administered dose and returns “correct”, “low”, or “high” accordingly. Keep in mind when writing your function that very small differences between the administered dose and the correct dose (i.e differences very close to zero) should be considered correct. Consider a threshold of within +/- 0.5 ml of the correct volume to be correct. Make a new column in your data frame by applying your function to the column you made in **1f**. Show your R code and print the first six lines of your data frame using the head () command.

**Solution**

classify\_dose\_difference <- function(difference) {   
 threshold = 0.5   
   
 if (abs(difference) <= threshold) {   
 return("correct")   
 } else if (difference < -threshold) {   
 return("high")   
 } else {   
 return("low")   
 }   
}   
   
df$classify\_dose\_difference <- mapply(classify\_dose\_difference, df$difference\_in\_dosage\_ml)   
   
head(df)

## patient kg twice\_daily\_dose\_given\_ml twice\_daily\_dose\_required\_mg  
## 1 patient\_1 16.1 8.05 402.5  
## 2 patient\_2 26.8 13.40 670.0  
## 3 patient\_3 13.5 6.75 337.5  
## 4 patient\_4 14.4 7.20 360.0  
## 5 patient\_5 15.2 7.60 380.0  
## 6 patient\_6 33.9 15.00 750.0  
## correct\_volume\_ml difference\_in\_dosage\_ml classify\_dose\_difference  
## 1 8.05 0 correct  
## 2 13.40 0 correct  
## 3 6.75 0 correct  
## 4 7.20 0 correct  
## 5 7.60 0 correct  
## 6 15.00 0 correct

# Exercise 2. Evaluating the audit (40 pts)

1. How many patients received the correct, high, or low doses? Apply the table() function in R to the column in the data frame where you evaluate each patient for whether the dose was high, low, or correct, and print the table to your document.

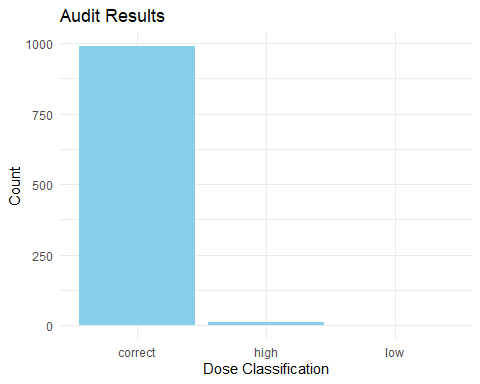
**Solution**

dose\_counts <- table(df$classify\_dose\_difference)   
print (dose\_counts)

##   
## correct high low   
## 990 9 1

**b)** Use ggplot2 in R to create a bar chart of the audit results. This will be a visual representation of the same data that went into your table in 2a.

library(ggplot2)   
   
   
classification\_df <- as.data.frame(dose\_counts)   
names(classification\_df) <- c("Classification", "Count")   
   
bar\_chart <- ggplot(data = classification\_df, aes(x = Classification, y = Count)) +   
 geom\_bar(stat = "identity", fill = "skyblue") +   
 labs(title = "Audit Results",   
 x = "Dose Classification",   
 y = "Count") +   
 theme\_minimal()   
   
print(bar\_chart)



1. Is under-dosing or overdosing of amoxicillin a bigger problem at this hospital? Why do you think this might be? Write a maximum of two sentences to give your interpretation of the results from **2a** and **2b**.

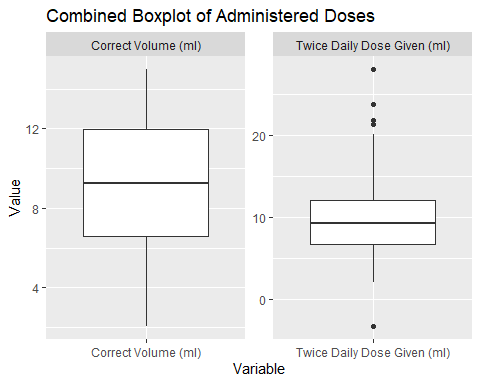
**Solution**

**From the data, we can infer that overdosing is a problem at this hospital. When required dosage - twice\_daily\_dose\_given\_ml, most of the incorrect results reported a negative value, meaning the given dose is higher than the required dose.**

**d) Challenge**. Can you think of a way to interrogate the data to test your interpretation from **2c**? Think of the pattern you are seeing in the data, and think of why this might be occurring given the dosing instructions for this treatment. Is there a way you can show this analytically or graphically using your R skills?

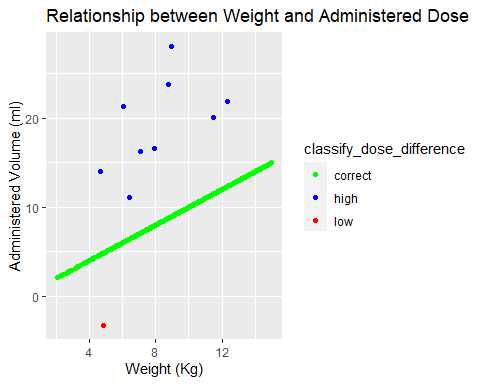
**Solution**

library(ggplot2)  
  
# Create a data frame for both columns  
combined\_df <- data.frame(  
 Variable = rep(c("Twice Daily Dose Given (ml)", "Correct Volume (ml)"), each = nrow(df)),  
 Value = c(df$twice\_daily\_dose\_given\_ml, df$correct\_volume\_ml)  
)  
  
# Create a boxplot with facets for each variable  
boxplot <- ggplot(data = combined\_df, aes(x = Variable, y = Value)) +  
 geom\_boxplot() +  
 labs(title = "Combined Boxplot of Administered Doses",  
 x = "Variable",  
 y = "Value") +  
 facet\_wrap(~Variable, scales = "free", ncol = 2)  
  
print(boxplot)

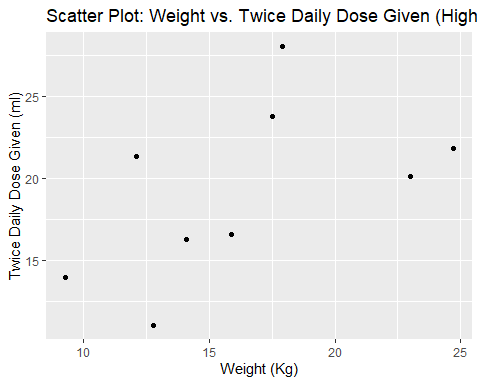


The Boxplot shows that there are outliers in the twice daily dose given. The box plot also shows a higher interquartile range in correct volume compared to the twice daily dose given.

library(ggplot2)  
  
scatter\_plot <- ggplot(data = df, aes(x = correct\_volume\_ml, y = twice\_daily\_dose\_given\_ml, color = classify\_dose\_difference)) +   
 geom\_point() +   
 labs(title = "Relationship between Weight and Administered Dose",   
 x = "Weight (Kg)",   
 y = "Administered Volume (ml)") +   
 scale\_color\_manual(values = c("correct" = "green", "low" = "red", "high" = "blue"))   
  
print(scatter\_plot)

 Printing the scatter **plot for the high dose alone.**

library(ggplot2)  
  
# Filter the data frame to include only rows with 'category' equal to 'high'  
filtered\_df <- df[df$classify\_dose\_difference == "high", ]  
  
scatter\_plot <- ggplot(data = filtered\_df, aes(x = kg, y = twice\_daily\_dose\_given\_ml)) +   
 geom\_point() +   
 labs(title = "Scatter Plot: Weight vs. Twice Daily Dose Given (High Category)",  
 x = "Weight (Kg)",  
 y = "Twice Daily Dose Given (ml)")  
  
print(scatter\_plot)



**finding out the correlation**

library(ggplot2)  
  
high\_category\_data <- subset(df, classify\_dose\_difference == "high")  
  
# Calculate the correlation coefficient  
correlation <- cor(high\_category\_data$kg, high\_category\_data$twice\_daily\_dose\_given\_ml)  
  
# Print the correlation coefficient (R value)  
cat("Correlation (R value):", correlation, "\n")

## Correlation (R value): 0.5265074

**The graph shows that a ‘High twice-daily dose’ given has a slightly positive correlation between dosage given and weight. When the data is whole, the high and low doses appear as outliers.**

**The reason for the outlier difference does not appear to be systematic. This is because the data could not show a consistent range of errors. The reason for overdosing can then be interpreted as human error, such as not using the correct weight when calculating or not taking note of the condition that dose cannot be higher than 1500mg.**