**Problem 1**: Load the vitcap dataset from the ISwR package. The vitcap data frame has 24 rows and 3 columns. It contains data on vital capacity for workers in the cadmium industry.

A screen shot of a computer code

AI-generated content may be incorrect.

A screenshot of a computer screen

AI-generated content may be incorrect.

Variables:

group: a numeric vector; group codes are 1: exposed > 10 years, 3: not exposed.

age: a numeric vector, age in years.

vital.capacity: a numeric vector, vital capacity (a measure of lung volume) in liters.

1. Perform some descriptive statistics on the vital.capacity variable (mean, standard deviation, histogram, etc.)

A screen shot of a computer code

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.A graph of a vital capacity

AI-generated content may be incorrect.

1. Does the data satisfy Normality?

A black background with white text

AI-generated content may be incorrect.

A screen shot of a computer code

AI-generated content may be incorrect.

Yes, because the p-value is greater than 0.05 we conclude that the vital capacity data is approximately normally distributed, therefore the data does satisfy Normality.

1. Conduct a hypothesis test to determine if the mean vital capacity for workers in the cadmium industry is significantly different than 4.7 liters.

A black and white text

AI-generated content may be incorrect.A screenshot of a computer program

AI-generated content may be incorrect.

The p-value (0.2767) is greater than 0.05, so we fail to reject the null hypothesis. The workers’ average vital capacity (4.47 L) is not significantly different from 4.7 L. Combined with the Shapiro–Wilk result (p = 0.098), the data are roughly normal, and the t-test assumptions are satisfied.

1. Graph the data to visualize the results of the hypothesis test.  
   A screen shot of a computer program

   AI-generated content may be incorrect.

A chart of a group of green squares

AI-generated content may be incorrect.

1. Write a few sentences to draw thorough conclusions about the analysis

The average vital capacity of the workers was 4.47 liters, which is close to the expected 4.7 liters. The t-test showed that this difference is not significant, so the workers’ lung capacity is about what we would expect. The data looks fairly normal, and the boxplot shows no big differences between exposure groups. Overall, exposure does not seem to have a clear effect on lung capacity.

**Problem 2**: A dataset named **marketing\_performance.csv** contains data on the performance of digital marketing campaigns run by a consulting firm. Each campaign used one of three **Marketing Strategies** and was implemented in one of two **Regions**.

| **Variable** | **Description** |
| --- | --- |
| SalesLift | Percent increase in sales after the campaign |
| Strategy | Marketing strategy used: Email, SocialMedia, or Influencer |
| Region | Region of the campaign: Urban or Rural |

a) Load the dataset marketing\_performance.csv into R and display the first few rows.

A screen shot of a computer

AI-generated content may be incorrect.

A screen shot of a computer

AI-generated content may be incorrect.

b) Create a **boxplot** showing SalesLift by Strategy, using color to distinguish Region.  
*(Hint: You can use either interaction(Strategy, Region) or ggplot2 with fill = Region.)*

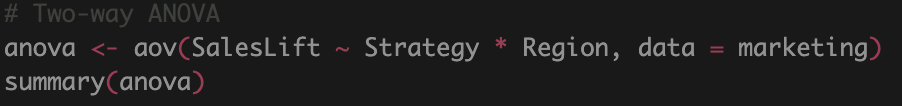
A close-up of a black background

AI-generated content may be incorrect.

A graph with different colored squares

AI-generated content may be incorrect.

c) Conduct a **Two-Way ANOVA** to determine whether **Strategy**, **Region**, or their **interaction** significantly affect sales performance.



A screenshot of a black screen

AI-generated content may be incorrect.

d) Check ANOVA assumptions using shapiro.test() on the residuals and bartlett.test() for equal variances.

A black background with white text

AI-generated content may be incorrect.A computer screen with white text

AI-generated content may be incorrect.

e) Write a short paragraph (3–4 sentences) interpreting the results. If there’s a significant interaction, explain what that means in the context of business decision-making.

Both marketing strategy and region affect sales lift, and the combination of the two matters too. This means that the best strategy depends on the region. Tests showed that the data meets the assumptions for ANOVA, so the results are reliable. Overall, marketing should be adjusted for different regions to get the best results.

**Problem 3**: Take a look at the 'iris' dataset that comes with R. The data can be loaded with the code:

data(iris)

iris

A description of the dataset can be found by running

?iris

In this dataset, write R code to find the mean of'Sepal.Length', specifically for the speciesvirginica**.** Take a screen-shot of your code and your numeric answer, paste it below.

A black screen with white text

AI-generated content may be incorrect.A screenshot of a computer

AI-generated content may be incorrect.

**Problem 4**: Continuing with the 'iris' dataset from the previous Question, use the ‘apply’ function to find a vector of means for the 4 variables 'Sepal.Length', 'Sepal.Width', 'Petal.Length', and 'Petal.Width'. Paste a screenshot of your code and your numeric answers, below.

**A number on a black background

AI-generated content may be incorrect.A black background with white text

AI-generated content may be incorrect.**