statmodata1

Ezo YEDİGÖL

2024-04-18

#DATASET-1 #The data set consists of 100 observations with 4 variables: Sales, Advertising, Price, and Season. #Sales, Advertising, and Price are numerical variables, while Season is a categorical variable.

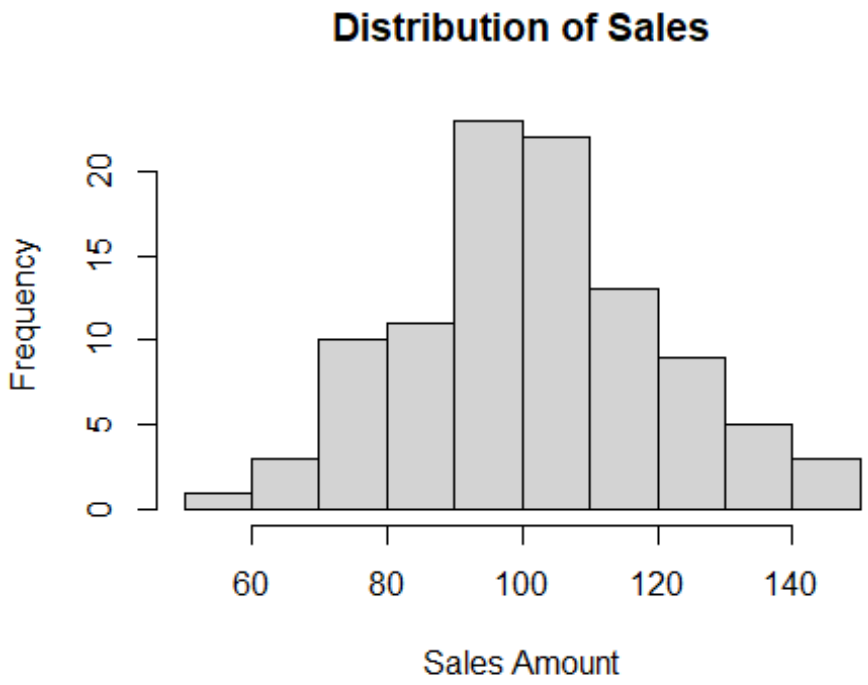
```
# Example Data Set-1
set.seed(123)  # Setting seed for reproducibility

# Sample data set size
n_samples <- 100

# Dependent variable: Sales amount
sales <- rnorm(n_samples, mean = 100, sd = 20)

# Independent variables: Advertising spend, product price, season
advertising_spend <- rnorm(n_samples, mean = 500, sd = 100)
product_price <- rnorm(n_samples, mean = 50, sd = 10)
season <- sample(c('Spring', 'Summer', 'Fall', 'Winter'), n_samples, replace
= TRUE)

# Creating the data frame
data <- data.frame(Sales = sales, Advertising = advertising_spend, Price = pr
oduct_price, Season = season)

# Displaying the head of the generated data set
head(data)
```

```
##       Sales Advertising    Price Season
## 1  88.79049    428.9593 71.98810 Winter
## 2  95.39645    525.6884 63.12413 Summer
## 3 131.17417    475.3308 47.34855 Winter
## 4 101.41017    465.2457 55.43194 Spring
## 5 102.58575    404.8381 45.85660   Fall
## 6 134.30130    495.4972 45.23753   Fall
```

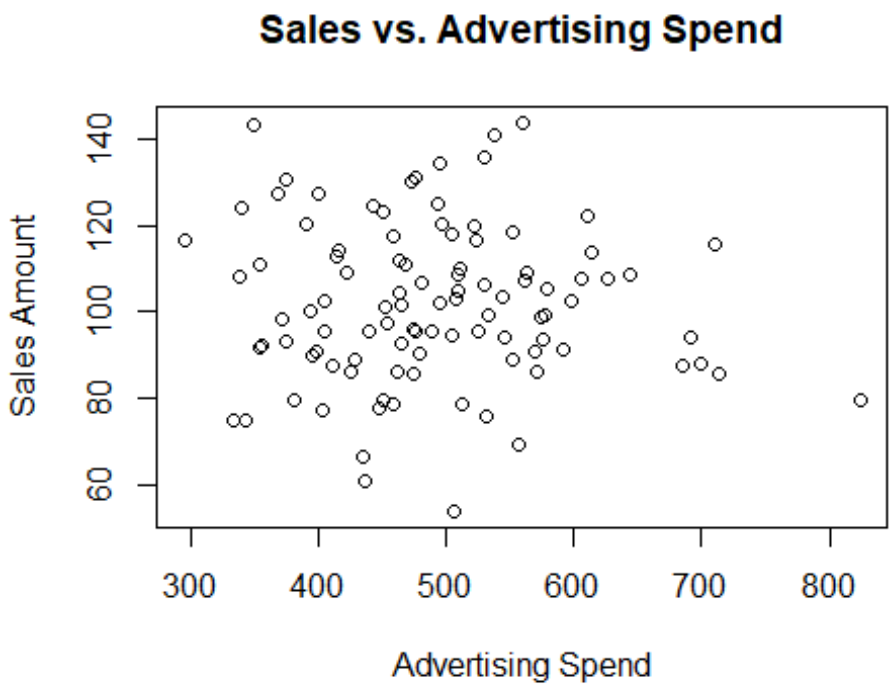
```
# Structural Characteristics of the Data Set
# Displaying the structure of the data set
str(data)
```

```
## 'data.frame':    100 obs. of  4 variables:
##  $ Sales      : num  88.8 95.4 131.2 101.4 102.6 ...
##  $ Advertising: num  429 526 475 465 405 ...
```

```
##  $ Price      : num  72 63.1 47.3 55.4 45.9 ...
##  $ Season     : chr  "Winter" "Summer" "Winter" "Spring" ...

# Summary Statistics for Numerical Variables
# Displaying summary statistics for numerical variables
summary(data)

##      Sales          Advertising        Price          Season
##  Min.   : 53.82   Min.   :294.7   Min.   :32.43   Length:100
##  1st Qu.: 90.12   1st Qu.:419.9   1st Qu.:44.69   Class :character
##  Median :101.24   Median :477.4   Median :50.36   Mode  :character
##  Mean   :101.81   Mean   :489.2   Mean   :51.20
##  3rd Qu.:113.84   3rd Qu.:546.8   3rd Qu.:57.64
##  Max.   :143.75   Max.   :824.1   Max.   :72.93

# Graphical Exploratory Data Analysis
# Histogram of Sales
hist(data$Sales, main = "Distribution of Sales", xlab = "Sales Amount", ylab
= "Frequency")
```

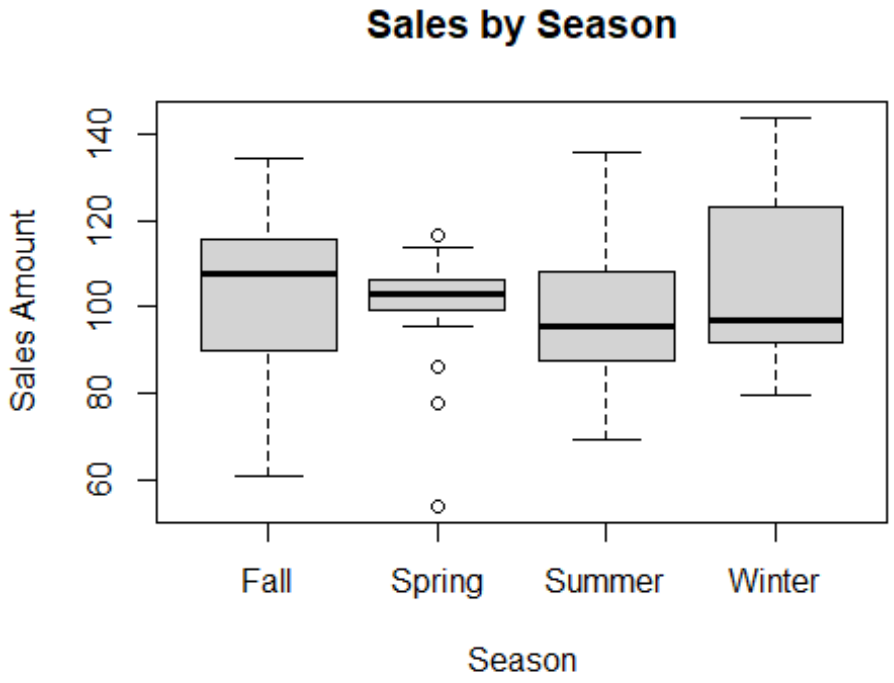



```
# Scatterplot of Sales vs. Advertising Spend
plot(data$Advertising, data$Sales, main = "Sales vs. Advertising Spend", xlab
= "Advertising Spend", ylab = "Sales Amount")
```




```r
# Scatterplot of Sales vs. Product Price
plot(data$Price, data$Sales, main = "Sales vs. Product Price", xlab = "Produc
t Price", ylab = "Sales Amount")
```

```r
# Boxplot of Sales by Season
boxplot(Sales ~ Season, data = data, main = "Sales by Season", xlab = "Season
", ylab = "Sales Amount")
```




```r
# Correlation Matrix
# Computing the correlation matrix for numerical variables
cor(data[, c("Sales", "Advertising", "Price")])
```

```
##                   Sales Advertising       Price
## Sales        1.00000000 -0.04953215 -0.12917601
## Advertising -0.04953215  1.00000000  0.03057903
## Price       -0.12917601  0.03057903  1.00000000
```

```r
# Distribution of Categorical Variable
# Displaying the distribution of observations across different seasons
table(data$Season)
```

```
##
##   Fall Spring Summer Winter
##     30     17     29     24
```

```r
# Regression Analysis
# Modeling sales amount based on advertising spend, product price, and season

# Creating the model
reg_model <- lm(Sales ~ Advertising + Price + Season, data = data)
```

```r
# Displaying the model summary
summary(reg_model)
```

```
##
## Call:
## lm(formula = Sales ~ Advertising + Price + Season, data = data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -44.23 -13.26   0.69  12.24  42.61
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)  120.069657  13.865573   8.660  1.3e-13 ***
## Advertising   -0.008525   0.018964  -0.450    0.654
## Price         -0.267752   0.192339  -1.392    0.167
## SeasonSpring  -2.428627   5.507749  -0.441    0.660
## SeasonSummer  -4.550053   4.741543  -0.960    0.340
## SeasonWinter   5.633448   4.978207   1.132    0.261
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 18.14 on 94 degrees of freedom
## Multiple R-squared:  0.06251,    Adjusted R-squared:  0.01265
## F-statistic: 1.254 on 5 and 94 DF,  p-value: 0.2906
```

```r
# ANOVA Analysis
# Checking the effect of the season factor on sales amount

# Creating the ANOVA model
anova_model <- lm(Sales ~ Season, data = data)

# Displaying ANOVA results
anova_result <- anova(anova_model)
print(anova_result)
```

```
## Analysis of Variance Table
##
## Response: Sales
##           Df Sum Sq Mean Sq F value Pr(>F)
## Season     3   1346  448.61  1.3607 0.2595
## Residuals 96  31650  329.69
```

```r
# ANCOVA Analysis
# Checking the effect of season factor on sales amount controlling for advert
ising spend and product price

# Creating the ANCOVA model
ancova_model <- lm(Sales ~ Advertising + Price + Season, data = data)
```

```r
# Displaying ANCOVA results
summary(ancova_model)
```

```
##
## Call:
## lm(formula = Sales ~ Advertising + Price + Season, data = data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -44.23 -13.26   0.69  12.24  42.61
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)  120.069657  13.865573   8.660  1.3e-13 ***
## Advertising   -0.008525   0.018964  -0.450    0.654
## Price         -0.267752   0.192339  -1.392    0.167
## SeasonSpring  -2.428627   5.507749  -0.441    0.660
## SeasonSummer  -4.550053   4.741543  -0.960    0.340
## SeasonWinter   5.633448   4.978207   1.132    0.261
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 18.14 on 94 degrees of freedom
## Multiple R-squared:  0.06251,    Adjusted R-squared:  0.01265
## F-statistic: 1.254 on 5 and 94 DF,  p-value: 0.2906
```

```r
#conclusions
#Regression Analysis Result:
#None of the variables (advertising spend, product price, and season) had a s
ignificant effect on sales amount.
#The R-squared value is low, indicating that the variables explain only a sma
ll portion of the variance in sales amount.

#ANOVA Analysis Result:
#The season factor does not have a significant effect on sales amount.

#ANCOVA Analysis Result:
#The results from ANCOVA analysis confirm that there is no significant effect
of advertising spend and product price on sales amount, controlling for the s
eason factor.
```

```
```