````
---
title: "DA2 Homework Assignment 01 - Linear Regression Analysis"
author: Gaire Ananta Prasad M24W0272
output:
  html_document:
    df_print: paged
editor_options:
  markdown:
    wrap: sentence
---

```{r load-data}
# Load the dataset
advertising_data <- read.csv("/Users/gaire/Documents/Assignement 1/Advertising.csv")
head(advertising_data)
```

M24W0272 Gaire Ananta Prasad Our goal is to investigate the effect that advertising costs through TV, Radio and Newspaper platforms have on product sales.
Through the statistical analysis method Linear Regression we identify and understand the connection between these variables.

```{r }
str(advertising_data)


```

```{r }
summary(advertising_data)


```

```{r tv advertising}
# Linear regression for TV vs. Sales
model_tv <- lm(Sales ~ TV, data = advertising_data)
summary(model_tv)


```

Gaire Ananta Prasad M24W0272 The basic linear regression shows that advertising on television creates both a robust and favorable impact on sales levels.
The value of 0.0475 found in the TV variable shows that sales grow about 47.5 units in response to each additional \$1,000 budget dedicated to television advertising.
The statistical significance of this result becomes evident because the p-value reaches below two inverse octillion (2e-16).
The analysis indicates that TV advertising accounts for 61% of the total factors affecting sales variation.
TV advertising establishes itself as a powerful tool that drives sales performance so it becomes a vital element for company success prediction.

```{r}
model <- lm(Sales ~ TV, data = advertising_data)
```

```{r}
plot(advertising_data$TV, advertising_data$Sales, xlab = "TV Advertising Expenses", ylab = "Sales", main = "Scatter Plot of TV vs. Sales")
abline(model, col = "red")

```
````

M24W0272 Gaire Ananta Prasad A plot depicting TV advertising efforts against sales demonstrates an evident positive linear relationship where greater TV advertisement investment leads to increased product sales.
The plot displays an outline regression line determined by our model which serves to fit the data best.
The data distribution pattern in the visual representation demonstrates how well the model has been fitted because most of the points cluster near the drawn line.
Analysis through the chart substantiates our discovery that product sales grow based on increased TV advertising investments.

```{r}
predicted_sales <- predict(model, newdata = data.frame(TV = advertising_data$TV))
residuals <- advertising_data$Sales - predicted_sales
residuals
```

Gaire Ananta Prasad M24W0272 The residuals resulted from subtracting actual sales data from predicted estimates using TV advertising expense data from our regression model.
The calculation of Residual spans from Actual Sales to Predicted Sales.
The model prediction deviations from actual data for each observation are measured by these residuals.
A beneficial residual indicates the model predicted values that were lower than actual sales figures whereas the model underestimated the figures.
The model overestimated when its prediction produced a negative residual value.
Our evaluation process depends on the examination of residuals which enable us to assess multiple aspects.
The quality of data fit throughout the complete dataset serves as an evaluation indicator.
The analysis checks for discoverable patterns and unusual points within the data.
The randomness of the errors indicates the model functions appropriately.

```{r}
mse <- mean(residuals^2)
mse
```

#Linear regression for Radio

```{r}
model_radio <- lm(Sales ~ Radio, data = advertising_data)
summary(model_radio)
```

Gaire Ananta Prasad M24W0272 This model evaluates the impact of Radio advertising on sales.
The coefficient for Radio is 0.2025, which suggests that for each additional \$1,000 spent on radio ads, sales are projected to increase by approximately 202.5 units.

With a p-value less than 2e-16, the relationship is highly statistically significant.
The R-squared value of 0.332 indicates that about 33% of the variation in sales can be explained by radio advertising alone.

While this is lower than the explanatory power of TV advertising, Radio still plays a significant role and could be a cost-effective channel for boosting sales.

```{r}
radio_model <- lm(Sales ~ Radio, data = advertising_data)
plot(advertising_data$Radio, advertising_data$Sales,
xlab = "Radio Advertising Expenses",
ylab = "Sales",
main = "Scatter Plot of Radio vs. Sales")
abline(radio_model, col = "blue")
```

M24W0272 Gaire Ananta Prasad The scatter plot reveals a moderate positive relationship between Radio advertising and Sales.

In general, as spending on radio ads increases, sales tend to rise as well.
The red line represents the best-fitting linear regression, highlighting this trend.
However, the noticeable spread of data points around the line suggests that other factors besides radio advertising also play a role in influencing sales.

```{r}
predicted_sales <- predict(model, newdata = data.frame(
TV = advertising_data$TV,
Radio = advertising_data$Radio,
Newspaper = advertising_data$Newspaper
))
residuals <- advertising_data$Sales - predicted_sales
residuals
```

M24W0272 Gaire Ananta Prasad The residuals from the Radio advertising model show the difference between the actual sales and what the model predicted.
A positive residual (e.g., +4.13) means the model underpredicted sales, while a negative residual (e.g., -2.73) means it overpredicted them.

When residuals are large, like +6.08 or -6.94, it suggests the model's predictions were significantly off for those observations.
This indicates that there may be additional factors influencing sales that the model doesn't account for.

By examining these residuals, we can pinpoint where the model may need improvement to better capture the true effect of Radio advertising on sales.

#Linear regression for Newspaper

```{r}
model_newspaper <- lm(Sales ~ Newspaper, data = advertising_data)
summary(model_newspaper)
```

M24W0272 Gaire Ananta Prasad This model analyzes the impact of Newspaper advertising on sales.
The coefficient of 0.0547 indicates a weak positive relationship — sales tend to increase slightly as newspaper ad spending goes up.

However, with a p-value of 0.06, the result is not statistically significant at the conventional 0.05 level, suggesting that the effect may not be reliable.

The R-squared value is 0.052, meaning that only 5.2% of the variation in sales can be explained by newspaper advertising, pointing to a relatively low predictive power for this variable.

```{r}
plot(advertising_data$Newspaper, advertising_data$Sales, xlab = "Newspaper Advertisin
g Expenses", ylab = "Sales", main = "Scatter Plot of Newspaper vs. Sales")
abline(model, col = "yellow")
```

M24W0272 Gaire Ananta Prasad The linear regression results suggest that Newspaper advertising has a minor yet statistically meaningful impact on sales.
For each additional unit spent on newspaper ads, sales are expected to rise by approximately 0.055 units.

Despite this, the model accounts for only around 5% of the variation in sales, implying that newspaper advertising on its own is a weak predictor of sales outcomes.

Overall, while there is some positive effect, Newspaper ads appear less effective when compared to more influential channels like TV or Radio.

```{r}
```

```
predicted_sales <- predict(model, newdata = data.frame(
TV = advertising_data$TV,
Radio = advertising_data$Radio,
Newspaper = advertising_data$Newspaper
))
residuals <- advertising_data$Sales - predicted_sales
residuals
```

M24W0272 Gaire Ananta Prasad The residuals from the Newspaper variable span a wide range, from -11.23 to 12.78, highlighting several cases where the model's predictions were significantly off.
While the majority of residuals are within a moderate range, the median value of -0.84 suggests a slight tendency to underestimate sales.

Despite this, the low R-squared value of 0.04733 indicates that Newspaper advertising has limited explanatory power — it accounts for only a small fraction of the variation in sales.
This reinforces the idea that newspaper ads alone are not a reliable predictor of sales performance.