# Practical One

Sbonelo Gumede GMDSBO006

2025-02-06

## Air quality data set

Remove missing values

```
data <- na.omit(airquality)
```

Temperature statistics

```
temp_df <- data.frame(Statistics = c("Mean", "Standard Deviation", "Minimum", "Maximum"),
              Values = c(mean(data$Temp), sd(data$Temp), min(data$Temp), max(data$Temp)))
knitr::kable(temp_df)
```

| Statistics | Values |
|---|---:|
| Mean | 77.792793 |
| Standard Deviation | 9.529969 |
| Minimum | 57.000000 |
| Maximum | 97.000000 |

Ozone level statistics

```
ozone_df <- data.frame(Statistics = c("Mean", "Standard Deviation", "Minimum", "Maximum"),
               Values = c(mean(data$Ozone), sd(data$Ozone), min(data$Ozone), max(data$Ozone)))
knitr::kable(ozone_df)
```

| Statistics | Values |
|---|---:|
| Mean | 42.09910 |
| Standard Deviation | 33.27597 |
| Minimum | 1.00000 |
| Maximum | 168.00000 |

## Cars data set

Remove missing values

```
cars <- na.omit(cars)
```

First principle calculations function

```
f <- function(X, Y){
# Calculate beta estimates
B <- solve (t(X) %*% X) %*% t(X) %*% Y

n <- length(cars$speed)
k <- 2
df <- n-k

# Calculate standard errors
s_square <- (1/(df))* (t(Y - X%*%B) %*% (Y - X%*%B))
C <- solve(t(X) %*% X)
se <- sqrt(as.numeric(s_square) * diag(C))

# Calculate t-statistics
t_stats <- B/se

# Calculate p-values
p_values <- 2 * (1 - pt(abs(t_stats), df = df))

data_frame <- data.frame("Estimate"=B, "Std. Error"=se, "t value"=t_stats, "Pr(>|t|)"=p_values,
                         check.names=FALSE)
row.names(data_frame) <-  c("(Intercept)", "speed")
return(data_frame)
}
```

Using the `first principle calculations` function

```
X <- cbind(1, cars$speed)
Y <- cars$dist
knitr::kable(f(X, Y))
```

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | -17.579095 | 6.7584402 | -2.601058 | 0.0123188 |
| speed | 3.932409 | 0.4155128 | 9.463990 | 0.0000000 |

Using the `lm` function

```
# Fit a linear model
model <- lm (dist ~ speed, data = cars)

# Display the summary
summary_table <- as.data.frame(summary(model)$coefficients)
knitr::kable(summary_table)
```

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | -17.579095 | 6.7584402 | -2.601058 | 0.0123188 |
| speed | 3.932409 | 0.4155128 | 9.463990 | 0.0000000 |