Practical One

Sbonelo Gumede GMDSBO006

2025-02-07

Air quality data set

Remove missing values

```
# Rows with NA
knitr::kable(airquality[rowSums(is.na(airquality)) > 0, ])
```

	Ozone	Solar.R	Wind	Temp	Month	Day
5	NA	NA	14.3	56	5	5
6	28	NA	14.9	66	5	6
10	NA	194	8.6	69	5	10
11	7	NA	6.9	74	5	11
25	NA	66	16.6	57	5	25
26	NA	266	14.9	58	5	26
27	NA	NA	8.0	57	5	27
32	NA	286	8.6	78	6	1
33	NA	287	9.7	74	6	2
34	NA	242	16.1	67	6	3
35	NA	186	9.2	84	6	4
36	NA	220	8.6	85	6	5
37	NA	264	14.3	79	6	6
39	NA	273	6.9	87	6	8
42	NA	259	10.9	93	6	11
43	NA	250	9.2	92	6	12
45	NA	332	13.8	80	6	14
46	NA	322	11.5	79	6	15
52	NA	150	6.3	77	6	21
53	NA	59	1.7	76	6	22
54	NA	91	4.6	76	6	23
55	NA	250	6.3	76	6	24
56	NA	135	8.0	75	6	25
57	NA	127	8.0	78	6	26
58	NA	47	10.3	73	6	27
59	NA	98	11.5	80	6	28
60	NA	31	14.9	77	6	29
61	NA	138	8.0	83	6	30
65	NA	101	10.9	84	7	4
72	NA	139	8.6	82	7	11
75	NA	291	14.9	91	7	14
83	NA	258	9.7	81	7	22
84	NA	295	11.5	82	7	23
96	78	NA	6.9	86	8	4
97	35	NA	7.4	85	8	5
98	66	NA	4.6	87	8	6
102	NA	222	8.6	92	8	10
103	NA	137	11.5	86	8	11

	Ozone	Solar.R	Wind	Temp	Month	Day
107	NA	64	11.5	79	8	15
115	NA	255	12.6	75	8	23
119	NA	153	5.7	88	8	27
150	NA	145	13.2	77	9	27

```
# Remove rows with NA
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