1)By Producing Data

Packages

These are packages we need, first of all we have to install these packages.

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.2.3

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(MASS)
```

```
## Warning: package 'MASS' was built under R version 4.2.3

##
## Attaching package: 'MASS'

## The following object is masked from 'package:dplyr':
##
##     select
```

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.2.3
```

```
library(reshape2)
```

```
## Warning: package 'reshape2' was built under R version 4.2.3
```

This is the first step, I product a dataset These code use for making a random dataset consisting of 100 observations. x is normal disturbuited with mean 20 and standard deviation of 5, y=2*x+error normal disturbution with 0 mean and standart deviation of 2

```
set.seed(123)
```

Variables

```
x1 <- rnorm(30, mean = rep(c(1, 2, 3), each = 10), sd = 1)
x2 <- rnorm(30, mean = rep(c(5, 10, 15), each = 10), sd = 2)
y <- 3*x1 + 1.5*x2 + rnorm(30, mean = 0, sd = 1)
```

Making a data set

```
data <- data.frame( x1, x2, y)
```

I collect x and y observations togeher with data.frame()

```
rdata<- data.frame(x1=x1,x2=x2,y=y)
```

With this code I made a regresion model

```
model<-lm(y~x1+x2,data=rdata)
```

With this code, I can see the summary of my data

```
summary(model)
```

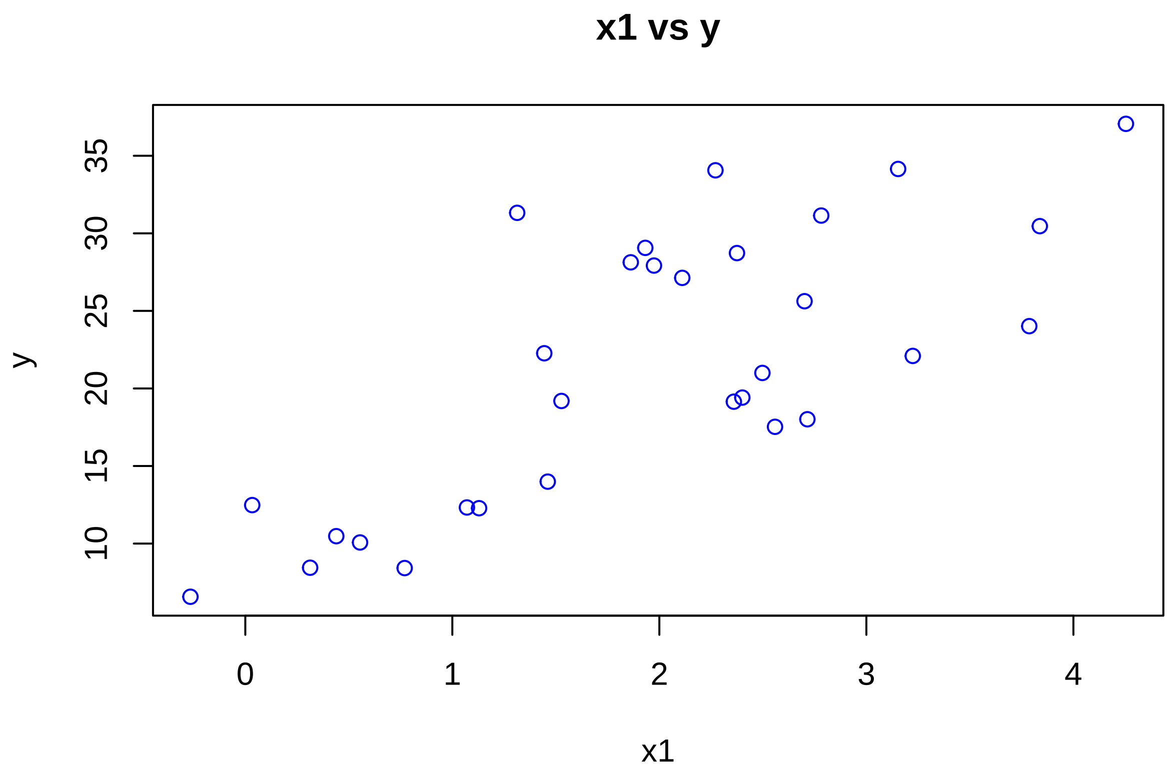
```
##
## Call:
## lm(formula = y ~ x1 + x2, data = rdata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.43626 -0.52913  0.02861  0.55710  2.14934
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.10088    0.43137  -0.234    0.817
## x1           3.20272    0.15913  20.126   <2e-16 ***
## x2           1.47387    0.04138  35.616   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8754 on 27 degrees of freedom
## Multiple R-squared:  0.9908, Adjusted R-squared:  0.9901
## F-statistic:  1452 on 2 and 27 DF,  p-value: < 2.2e-16
```

Comment:

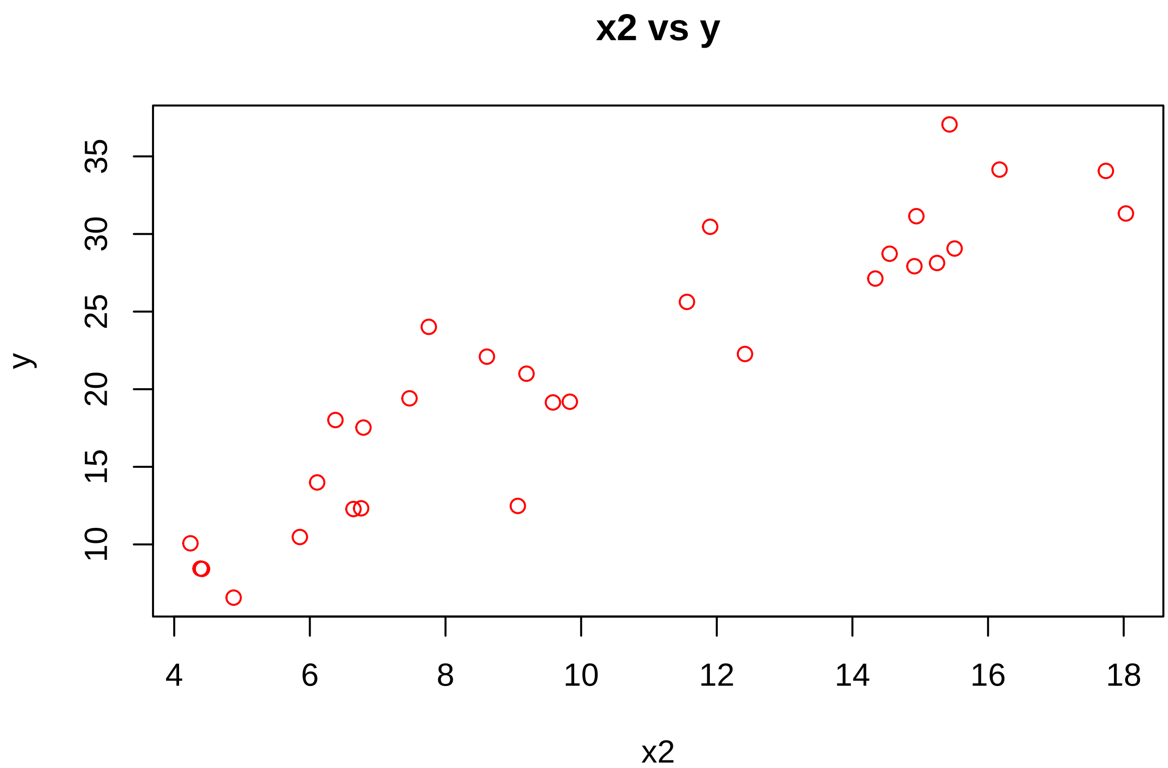
This model shows that "x1" and "x2" have positive associations with "y," with each unit increase in "x1" corresponding to approximately a 1.95456-unit increase in "y," and each unit increase in "x2" corresponding to approximately a 3.00953-unit increase in "y." The categorical variable "z" also has a significant effect on "y," with level "B" associated with a 5.67059-unit increase and level "C" associated with a 9.99279-unit increase in "y." These relationships are statistically significant with high t-values and low p-values. The model has a high R-squared value of 0.9927, indicating that it explains a substantial portion of the variance in "y," making it a strong predictor of the relationship between the variables.

Visualization

```
plot(rdata$x1, rdata$y, main = "x1 vs y", xlab = "x1", ylab = "y", col = "blue")
```




```
plot(rdata$x2, rdata$y, main = "x2 vs y", xlab = "x2", ylab = "y", col = "red")
```



As you can see in the graph x2 affects y positively and also there is a corellasion between x and y.

ANOVA

```
anova_c <- anova(model)
anova_c
```

```
## Analysis of Variance Table
##
## Response: y
##           Df  Sum Sq Mean Sq F value    Pr(>F)
## x1         1 1254.01 1254.01  1636.2 < 2.2e-16 ***
## x2         1  972.17  972.17  1268.5 < 2.2e-16 ***
## Residuals 27   20.69    0.77
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

When the F value is high and the p-value is low, it can be concluded that there is a statistically significant difference between groups

ANCOVA

```
ancova_C <- lm(y ~ x1 + x2, data = rdata)
summary(ancova_C)
```

```
##
## Call:
## lm(formula = y ~ x1 + x2, data = rdata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.43626 -0.52913  0.02861  0.55710  2.14934
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.10088    0.43137  -0.234    0.817
## x1           3.20272    0.15913  20.126   <2e-16 ***
## x2           1.47387    0.04138  35.616   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8754 on 27 degrees of freedom
## Multiple R-squared:  0.9908, Adjusted R-squared:  0.9901
## F-statistic:  1452 on 2 and 27 DF,  p-value: < 2.2e-16
```

Comment:

Based on the data in the table, we see that the model fits the data quite well. The coefficients of the independent variables are high and statistically significant, meaning that we see that these variables have a strong effect on the dependent variable. In short, this regression model seems to explain the relationship between the variables in the data set quite well.

2)Using Data from R-studio

As you can see we call the data from R-studio.

```
data(mtcars)
```

After that we make basic regression model

```
reg_model <- lm(mpg ~ wt, data = mtcars)
summary(reg_model)
```

```
##
## Call:
## lm(formula = mpg ~ wt, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.5432 -2.3647 -0.1252  1.4096  6.8727
##
```

```
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  37.2851     1.8776  19.858  < 2e-16 ***
## wt           -5.3445     0.5591  -9.559 1.29e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.046 on 30 degrees of freedom
## Multiple R-squared:  0.7528, Adjusted R-squared:  0.7446
## F-statistic: 91.38 on 1 and 30 DF,  p-value: 1.294e-10
```

Comment:

We found the p-value to be 1.294e-10. This is a very low value and very close to zero. We can say that this is highly statistically significant on (wt) / (mpg) in the model. That is, each unit increase in the weight of the vehicle leads to a significant decrease in fuel economy. Therefore, increasing weight is associated with decreasing miles per gallon. This finding tells us that heavier vehicles generally have lower fuel economy.

ANOVA

```r
anova_result <- anova(reg_model)
print(anova_result)
```

```
## Analysis of Variance Table
##
## Response: mpg
##           Df Sum Sq Mean Sq F value    Pr(>F)
## wt         1 847.73  847.73  91.375 1.294e-10 ***
## Residuals 30 278.32    9.28
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

ANCOVA

```r
ancova_model <- lm(mpg ~ wt + drat, data = mtcars)
ancova_result <- anova(ancova_model)
print(ancova_result)
```

```
## Analysis of Variance Table
##
## Response: mpg
##           Df Sum Sq Mean Sq F value    Pr(>F)
## wt         1 847.73  847.73 91.3086 1.832e-10 ***
## drat       1   9.08    9.08  0.9781    0.3309
## Residuals 29 269.24    9.28
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Comment:

In this output, the effects of both weight (wt) and rear drive ratio (drat) variables on the fuel economy of the vehicles were examined. First, we found the p-value for the weight (wt) variable to be 1.832e-10, which is an extremely low value. This is statistically significant over weight in miles per gallon. That is, an increase in vehicle weight is associated with a decrease in miles per gallon. However, the p-value for the rear drive ratio (drat) variable is 0.3309, which is not statistically significant. Therefore, there is not enough evidence to say that changes in rear axle ratio have a significant impact on miles per gallon. As a result, it can be said that the most important factor on the fuel economy of vehicles is weight and the rear drive ratio does not play an important role in explaining this effect.

Visualization

And Finally this is our heatmap with correlation. And Also we can see relations between with variables.

```
cor_matrix <- cor(mtcars)
melted_cor <- melt(cor_matrix)
ggplot(melted_cor, aes(Var1, Var2, fill = value)) +
  geom_tile() +
  scale_fill_gradient2(low = "blue", high = "red", mid = "white", midpoint = 0, limit = c(-1, 1), name =
  labs(title = "Correlation Heatmap of mtcars Variables",
       x = "Variables",
       y = "Variables") +
  theme_minimal()
```