

# Test1

2023-04-04

```
library(dplyr)

##
## 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(ggplot2)
text1 <- read.delim("~/Desktop/text1.txt")
data(text1)

## Warning in data(text1): data set 'text1' not found

names(text1)

## [1] "Age"          "Expenditures"

data <- data.frame(Age = text1$Age, Expen. = text1$Expenditures)
Regression <- lm(Age~Expen., data = data)
summary(Regression)$coefficients

##               Estimate Std. Error   t value    Pr(>|t|)
## (Intercept) 141.695823 29.3112307   4.834182 6.331258e-05
## Expen.      -1.012207  0.2893719  -3.497944 1.851756e-03
```

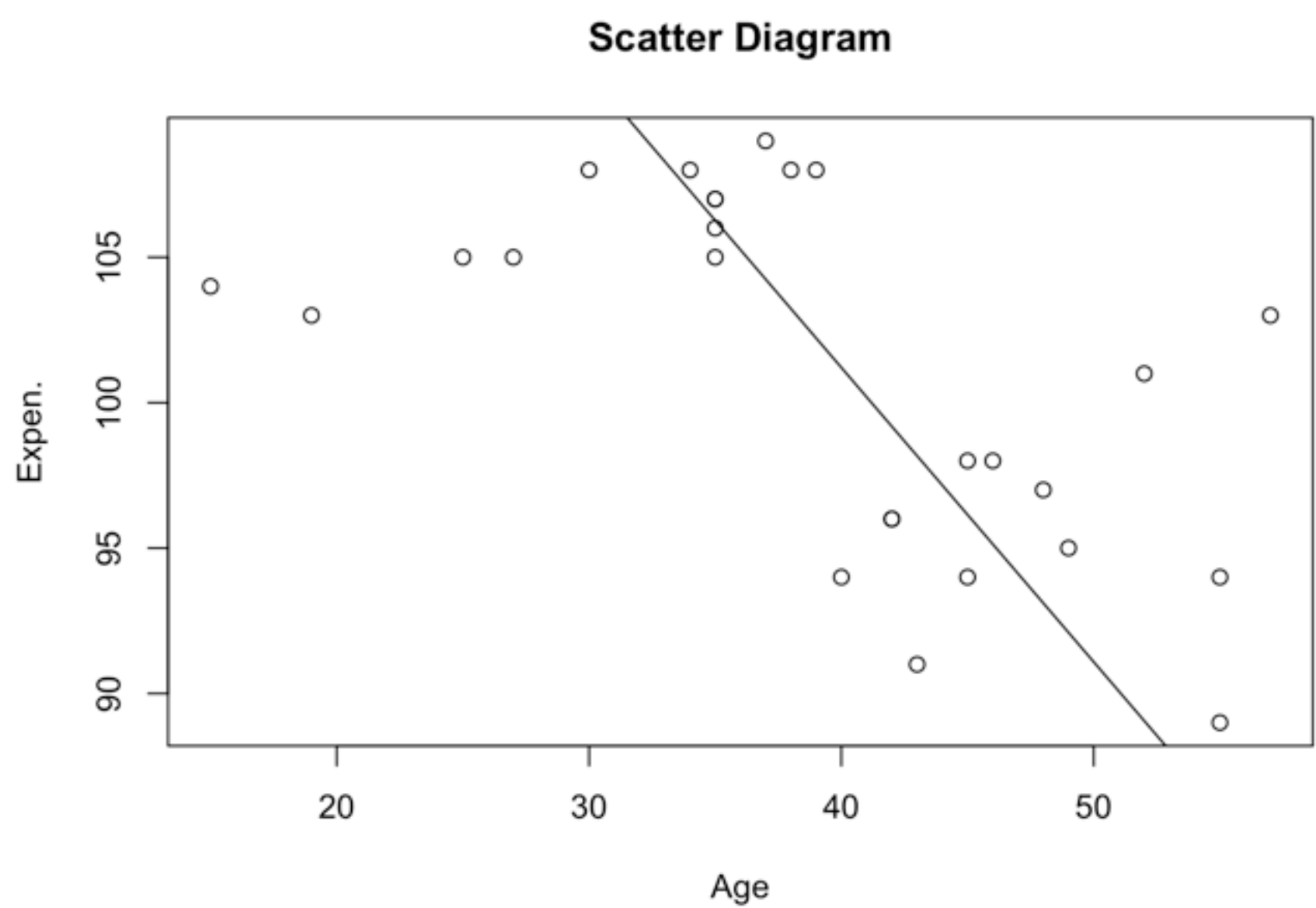
## 1.

the coefficients a = 141.695823

the coefficients b = -1.012207

T-value of b = -3.497944

```
plot(data, main = "Scatter Diagram")
abline(Regression)
```



## 2.

This diagram appears to be divided into two groups based on age 40.

Overall, the regression line  $y = a + bx$  shows the form of a straight line going downward.

This means that expenditures gradually decline with age.

```
data <- data %>%
  mutate(more_40 = ifelse(Age >= 40, "more_40", "less_40"))

more_40_age <- data %>%
  filter(more_40 == "more_40")

Regression <- lm(Age~Expen., data = more_40_age)
summary(Regression)$coefficients

##               Estimate Std. Error   t value    Pr(>|t|)
## (Intercept) 16.419764 42.0691450   0.3903042 0.7037638
## Expen.       0.325476  0.4386122   0.7420587 0.4735991

less_40_age <- data %>%
  filter(more_40 == "less_40")

Regression <- lm(Age~Expen., data = less_40_age)
summary(Regression)$coefficients

##               Estimate Std. Error   t value    Pr(>|t|)
## (Intercept) -314.971910 77.5923676  -4.059316 0.0018854001
## Expen.        3.252809  0.7292552   4.460453 0.0009619365
```

## 3.

### (1) : Age >= 40,

the coefficients a = 16.419764

the coefficients b = 0.325476

T-value of b = 0.7420587

### (2) : Age < 40,

the coefficients a = -314.971910

the coefficients b = 3.252809

T-value of b = 4.460453

## 4.

The difference in the results in (a) and (c) is thought to be caused by the very different distribution of data between the divided groups.

As a result, this statistic only looks effective in people over 40, but if the values vary so greatly depending on the grouping, I think the statistic should have been constructed for a sample of people over 40 in the first place.

Therefore, we learned that in statistics, along with consideration of which sample to choose, it is essential to also verify partial errors, including issues such as grouping.