Correlation Analysis

ATIQ UR REHMAN

Date: 2/18/2021

INTRODUCTION

- Correlation analysis measures the relationship between the two variables.
- Correlation analysis shows us how to determine both the nature and strength of relationship between two variables.
- Three possible results in the correlational study: Positive, Negative and no correlation.
- Example positive correlation: Income and consumption
- Example of negative correlation: Cold weather and air conditioning cost
- Example of no correlation: price of clothes and car speed.

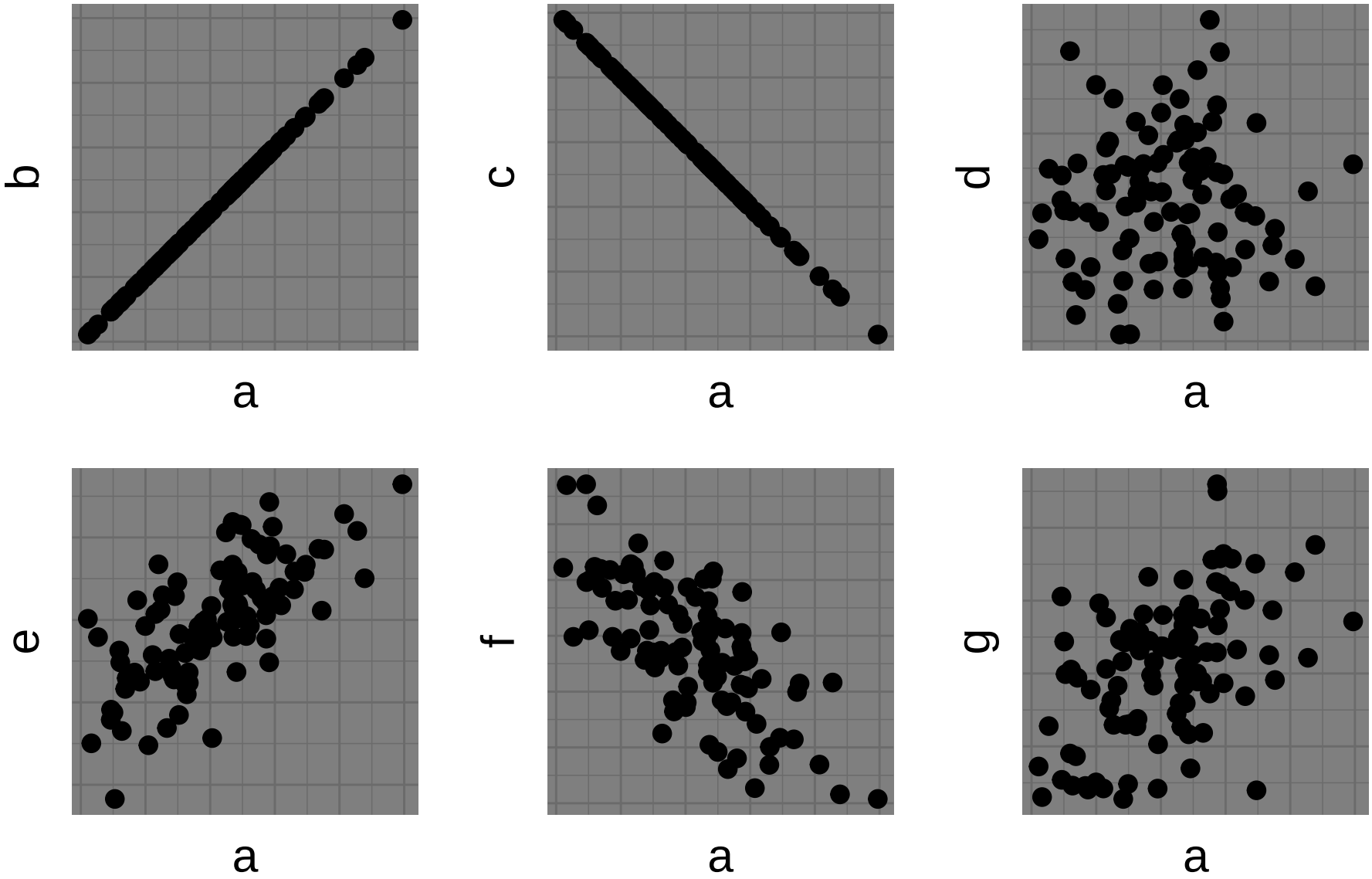
Types of Correlation

- Perfect positve correlation
- perfect negative correlation
- strong/weak positive correlation
- strong/weak negative correlation
- absolute no correlation

. . .

- A correlation of $+1$ indicates a perfect positive correlation.
- A correlation of -1 indicates a perfect negative correlation.
- Correlation value close to 1 shows strong and close to 0 indicates weak correlation.
- Also correlation value close to -1 shows strong and close to 0 indicates weak correlation.
- A zero correlation indicates that there is no relationship between the variables.

Some diagrams of correlatons



Method of studying correlation

- Scatter Diagram
- Karl Pearson' method of correlation

Scatter diagram

- Scatter diagram can only tell us about the nature of correlation.
- We can identify the direction of the relation visually.
- We cannot determine the magnitude of correlation in scatter diagram.

Karl Pearson method of correlation

- It is denoted by "r".
- The formula of r is

$$r = \frac{cov(x, y)}{\sqrt{Var(x)Var(y)}},$$

where $cov(x, y)$ is the shared variability of the two variables and $Var(x)$, $Var(y)$ are the individual variability each varibale.

- The range of r is $-1 \leq r \leq +1$

Real dataset

```
df_cars <- datasets::cars

str(df_cars)
```

```
## 'data.frame':    50 obs. of  2 variables:
##  $ speed: num  4 4 7 7 8 9 10 10 10 11 ...
##  $ dist : num  2 10 4 22 16 10 18 26 34 17 ...
```

- The correlation between the cars' speed and distance travelled by the cars is

```
cor(df_cars$speed, df_cars$dist)
```

```
## [1] 0.8068949
```

The scatter plot of cars' dataset

```
## `geom_smooth()` using formula 'y ~ x'
```