STA 4320 CHAP 3.1

He Jiang

Aug 2024

Sec 3.1.1

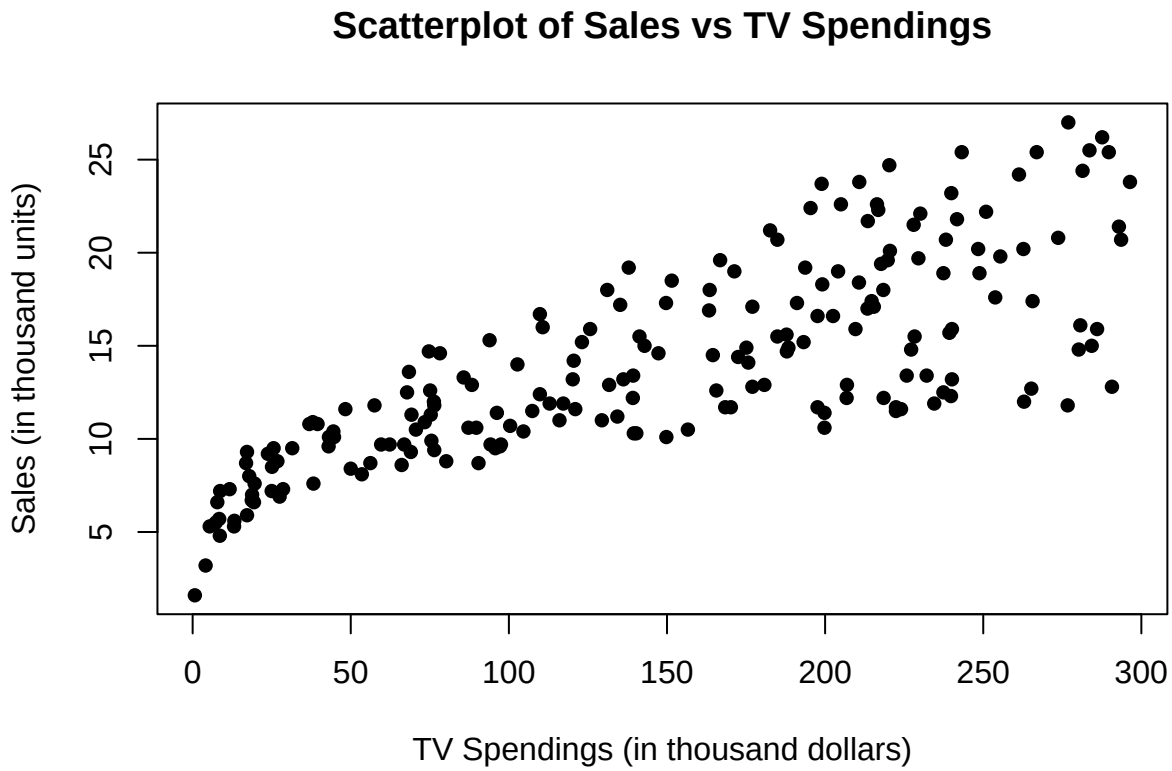
Advertising dataset

```
fpath = getwd()
Advertising = read.csv(paste0(fpath, "/Advertising.csv"))
```

Scatterplot

```
y = Advertising$sales
x = Advertising$TV

plot(x, y,
     main = "Scatterplot of Sales vs TV Spendings",
     pch = 16,
     xlab = "TV Spendings (in thousand dollars)",
     ylab = "Sales (in thousand units)")
```



To see some (x_i, y_i) values

```r
head( cbind(x, y) )
```

```
##          x    y
## [1,] 230.1 22.1
## [2,]  44.5 10.4
## [3,]  17.2  9.3
## [4,] 151.5 18.5
## [5,] 180.8 12.9
## [6,]   8.7  7.2
```

By hand

Sample values

```r
sigma_x = sum(x)
sigma_y = sum(y)

sigma_x2 = sum(x^2)
sigma_y2 = sum(y^2)

sigma_xy = sum(x*y)
```

By R function

R built-in function

```r
lm(y ~ x)
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Coefficients:
## (Intercept)            x
##     7.03259      0.04754
```

Summary of the regression result

```r
res = summary( lm(y ~ x) )
res
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.3860 -1.9545 -0.1913  2.0671  7.2124
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 7.032594   0.457843   15.36   <2e-16 ***
## x           0.047537   0.002691   17.67   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 3.259 on 198 degrees of freedom
## Multiple R-squared:  0.6119, Adjusted R-squared:  0.6099
## F-statistic: 312.1 on 1 and 198 DF,  p-value: < 2.2e-16
```

Scatterplot with regression line

```r
# scatterplot
plot(x, y,
     main = "Scatterplot of Sales vs TV Spendings",
     pch = 16,
     xlab = "TV Spendings (in thousand dollars)",
     ylab = "Sales (in thousand units)")

# overlay regression line
abline(lm(y ~ x),
       col = "blue",
       lwd = 2)

# add a legend
legend("bottomright",
       legend = c("regression line"),
       fill = "blue")
```