Case Study 1: The Ice Cream Sales Conundrum

Julian Gradziel

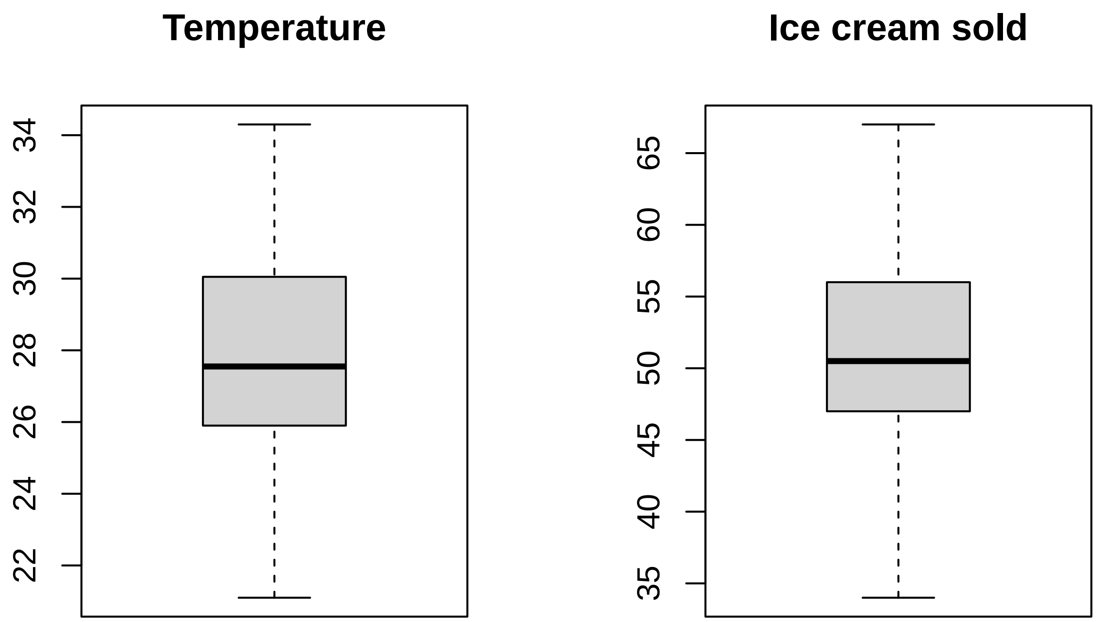
2023-10

##Task 1

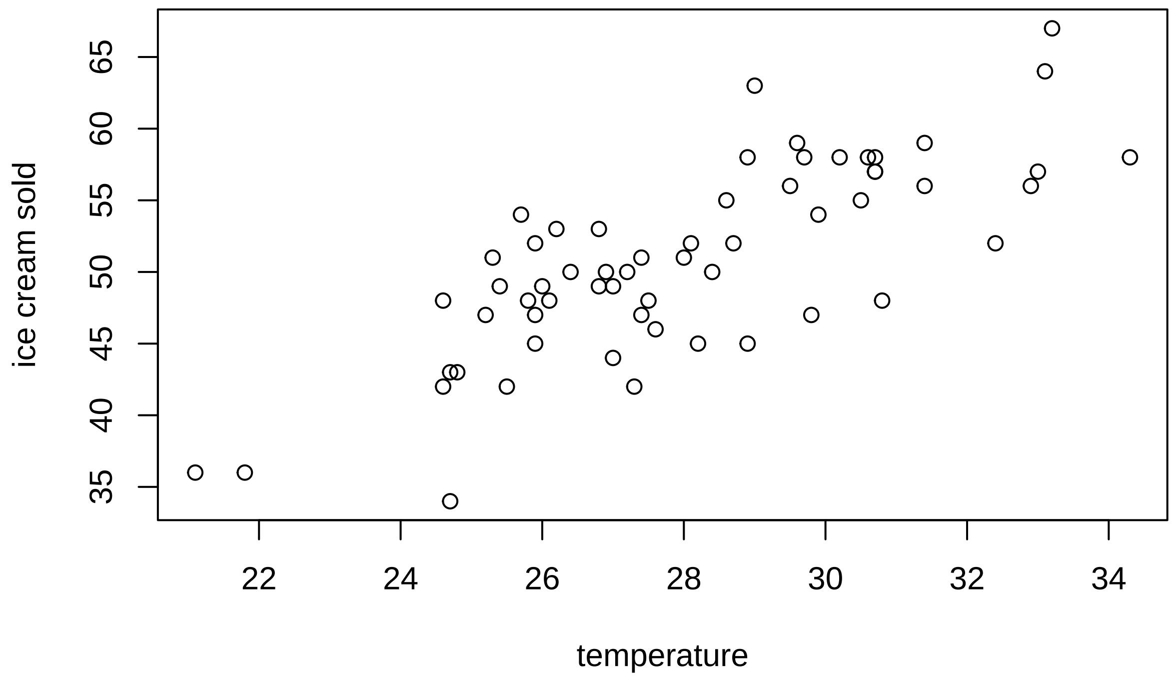
```
setwd("C:/Users/julia/Desktop/na pendrive/data for R")
kwality <- read.csv("kwality_sales.csv")

summary(kwality)
```

```
##        X          temperature    ice_creams_sold
##  Min.   : 0.00   Min.   :21.10   Min.   :34.00
##  1st Qu.:14.75   1st Qu.:25.90   1st Qu.:47.00
##  Median :29.50   Median :27.55   Median :50.50
##  Mean   :29.50   Mean   :28.03   Mean   :50.85
##  3rd Qu.:44.25   3rd Qu.:29.98   3rd Qu.:56.00
##  Max.   :59.00   Max.   :34.30   Max.   :67.00
```

```
par(mfrow=c(1,2))
boxplot(kwality$temperature, main = "Temperature")
boxplot(kwality$ice_creams_sold, main = "Ice cream sold")
```




```r
plot(kwality$temperature, kwality$ice_creams_sold, xlab="temperature", ylab="ice cream sold")
```



##Task 2

```r
cor(kwality$temperature,kwality$ice_creams_sold)
```

```
## [1] 0.7765052
```

The correlation coefficient indicates supports the information obtained in task 1 that there is a moderately strong positive linear relation between the two variables.

##Task 3

```r
linear_regression <- lm(kwality$ice_creams_sold~kwality$temperature)
summary(linear_regression)
```

```
##
## Call:
## lm(formula = kwality$ice_creams_sold ~ kwality$temperature)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.6698  -2.5339   0.5594   2.5706  10.3458
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)           -1.1940     5.5726  -0.214    0.831
```

```
## kwality$temperature   1.8568     0.1978   9.385 3.09e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.264 on 58 degrees of freedom
## Multiple R-squared:  0.603,  Adjusted R-squared:  0.5961
## F-statistic: 88.08 on 1 and 58 DF,  p-value: 3.089e-13
```

the standard error is 4.264 error is really small which is reflected in the p-value of 3.089e-13

```r
linear_regression$coefficients
```

```
##         (Intercept) kwality$temperature
##           -1.194008            1.856836
```

For an increase in temperature of 1°C, ice cream sales increase by 1.86, i.e. roughly 2. Since ice cream sales cannot be negative, the intercept of -1.194 itself does not have a significant meaning in this context. However, it indicates that ice cream sales start only at a positive temperature (first ice cream sold at roughly 2°C).

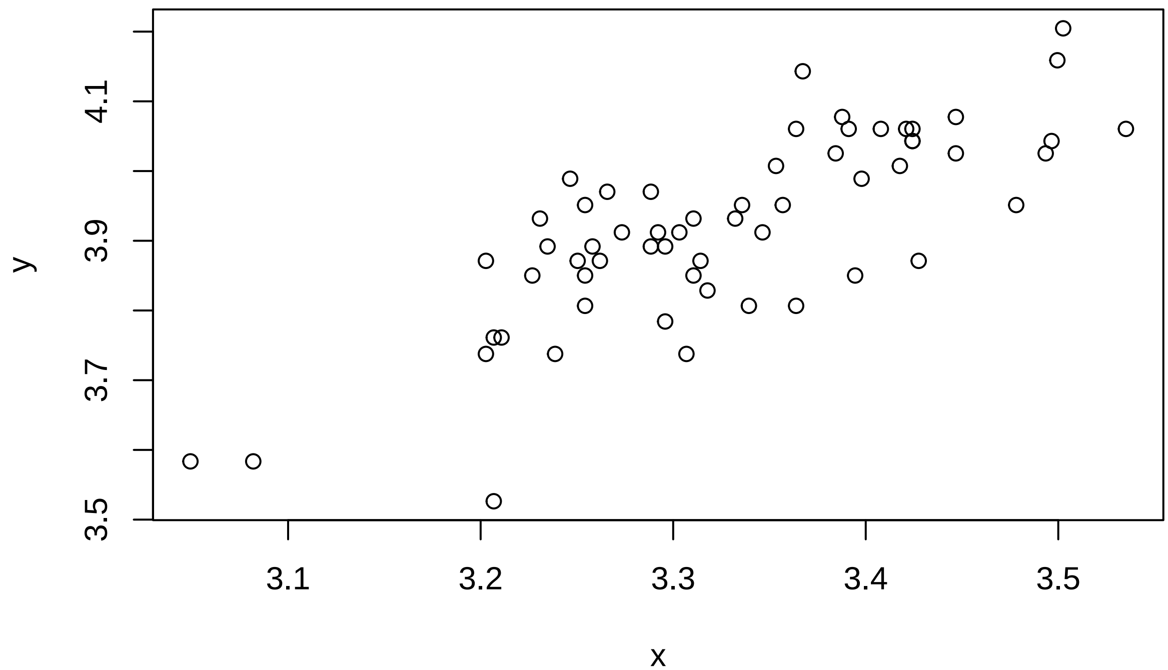
##Task 4

```r
sales_at_34_degrees <- linear_regression$coefficients[1]+linear_regression$coefficients[2]*34
print(sales_at_34_degrees)
```

```
## (Intercept)
##     61.9384
```

##Task 5

```r
y <- log(kwality$ice_creams_sold)
x <- log(kwality$temperature)
plot(x,y)
```




```
linear_regression_2 <- lm(y~x)
```

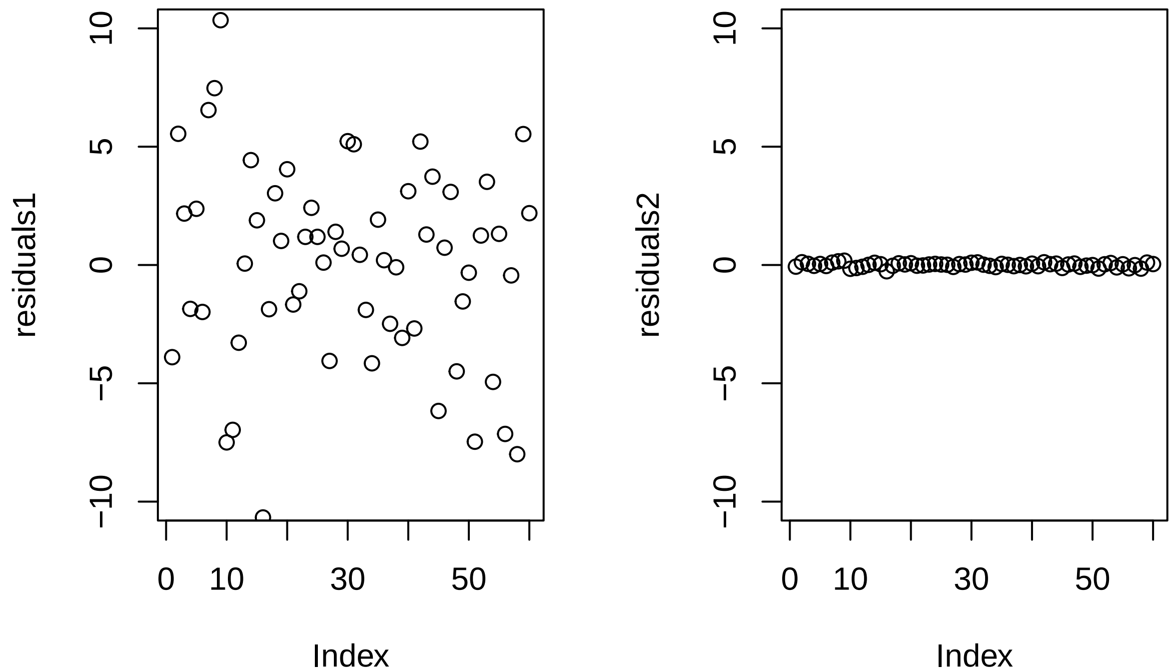
```
exp(linear_regression_2$coefficients[1]+linear_regression_2$coefficients[2]*log(34))
```

```
## (Intercept)
##    62.14772
```

The obtained number deviates a bit from the one we got using the standard linear regression model. This can be explained by the fact that …

Task 6

```
residuals1 <- residuals(linear_regression)
residuals2 <- residuals(linear_regression_2)
par(mfrow=c(1,2))
plot(residuals1, ylim=c(-10,10))
plot(residuals2, ylim=c(-10,10))
```



Residuals are randomly scattered. As shown in the two graphs, residuals for the log-log model are much smaller indicating that the log-log model provides a much better fit