#Hypothesis Testing: Simple Linear Model

Kamesh Dubey

2024-05-25

1. Generate Data

```r
set.seed(0)

age = round(runif(20, min=20, max = 80))

heartrate = (220-age)
heartrate = round(heartrate + rnorm(20, mean=0, sd=15))

cat("Age:", age )
```

```
## Age: 74 36 42 54 74 32 74 77 60 58 24 32 31 61 43 66 50 63 80 43
```

```r
cat("\n\nHearbeat:", heartrate)
```

```
##
##
## Hearbeat: 157 172 161 162 142 182 150 130 167 143 193 194 191 171 176 162 186 147 121 178
```

Create a Dataframe

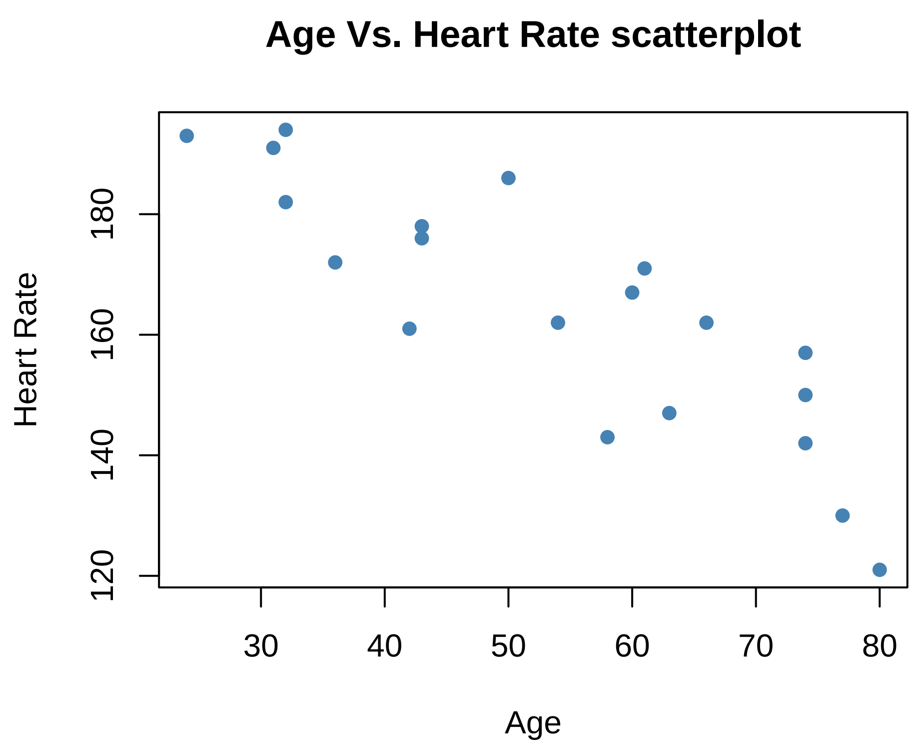
```r
data = data.frame(Age=age, Heartrate=heartrate)

print(data)
```

```
##    Age Heartrate
## 1   74       157
## 2   36       172
## 3   42       161
## 4   54       162
## 5   74       142
## 6   32       182
## 7   74       150
## 8   77       130
```

```
## 9   60        167
## 10  58        143
## 11  24        193
## 12  32        194
## 13  31        191
## 14  61        171
## 15  43        176
## 16  66        162
## 17  50        186
## 18  63        147
## 19  80        121
## 20  43        178
```

Plot Age and Heartbeart

```r
par(plt=c(0.2, 0.8, 0.2, 0.75))
plot(age, heartrate,
     xlab = "Age", ylab="Heart Rate",
     main = "Age Vs. Heart Rate scatterplot",
     pch=16, col="steelblue")
```



Simple Linear Model

```
model = lm(heartrate~age)
summary(model)
```

```
##
## Call:
## lm(formula = heartrate ~ age)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -16.8550  -8.3937  -0.3439   7.9752  17.9683
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 219.1365     7.9690  27.499 3.72e-16 ***
## age          -1.0221     0.1414  -7.228 1.01e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.82 on 18 degrees of freedom
## Multiple R-squared:  0.7438, Adjusted R-squared:  0.7295
## F-statistic: 52.25 on 1 and 18 DF,  p-value: 1.007e-06
```

Observations

- using p-values, we can see that both Intercept and age are significant.
- R-Squared: 0.743, Adjusted R-Squared:0.729. We can conclude that age explain the heart rate quite well
- The equation for heart rate from **Linear Model** can be expressed as: $\text{heart rate} = 219.1365 - 1.0221 \times \text{age}$

Residual Analysis

```
res = residuals(model)
pred = fitted(model)

cat("Fitted-Values: ", pred)
```

```
## Fitted-Values:  143.5015 182.3411 176.2085 163.9434 143.5015 186.4294 143.5015 140.4352 157.8108 159
```

```
cat("\n\nResidual: ", res)
```

```
##
##
## Residual:  13.49851 -10.34107 -15.2085 -1.943372 -1.501487 -4.429445 6.498513 -10.4352 9.189194 -16.
```

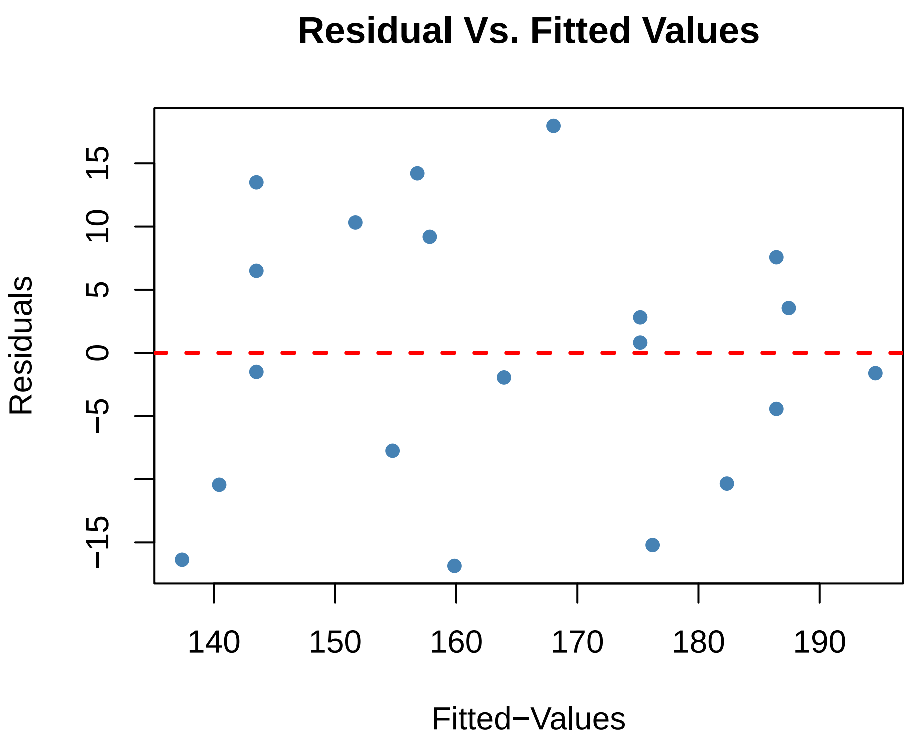
```r
cat("\n\nSummary of Residual\n")
```

```
##
##
## Summary of Residual
```

```r
print(summary(res))
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -16.8550  -8.3937  -0.3439   0.0000   7.9752  17.9683
```

- From Residual summary we have: $\sum_{i=1}^{20} e_i = 0$ or $\bar{e_i} = 0$



From the Residual vs Fitted Values plot, We can conclude that

- Residuals are Randomly Scattered around Zero mean.
- No Systematic pattern can be seen in the plot.
- Variance of Residual seems Constant
- No sign of outliers.

Testing of Hypothesis

```r
n = length(age)

beta_0_hat = coef(model)[["age"]]

RSS=sum(res^2)

sigma_hat = sqrt(RSS/n-2)

Sxx = sqrt(sum(age=mean(age)^2))
Std_err_beta_0 = sigma_hat/Sxx

cat("beta_0=", beta_0_hat)
```

```
## beta_0= -1.022094
```

```r
cat("\n\nSigma_hat=", sigma_hat)
```

```
##
##
## Sigma_hat= 10.16389
```

```r
cat("\n\nSE(beta_0)=", Std_err_beta_0)
```

```
##
##
## SE(beta_0)= 0.1892717
```

Test for betas

- **for** β_0

Consider

$$H_0 : \beta_0 = 0 \quad \text{vs} \quad H_1 : \beta_0 \neq 0$$

Then,

```r
t_0 = (beta_0_hat-(-0))/Std_err_beta_0

cat("       Under Ho,\n                 t-statistic value= ", t_0)
```

```
##         Under Ho,
##                   t-statistic value=  -5.400143
```

```r
p_value = pt(t_0, 20, lower.tail=FALSE)

cat("\n\n      The corresponding p-value for the test statistic will be:\n\n                     p-value
```

```
##
##
##       The corresponding p-value for the test statistic will be:
##
##                       p-value= 0.9999862
```

The p-value of 0.99 (»0.05) tell us that $\beta_0 = 0$ is not likely.

This time, Let's Consider

$$H_0: \beta_{0} = -1 \quad \text{vs} \quad H_1: \beta_{0} \neq -1$$

```r
Sxx = sqrt(sum(age=mean(age)^2))
Std_err_beta_0 = sigma_hat/Sxx

t = (beta_0_hat-(-1))/Std_err_beta_0
cat("Under Ho,\n            t-statistic value= ", t )
```

```
## Under Ho,
##             t-statistic value=  -0.116733
```

```r
p_value = pt(t, 20, lower.tail=FALSE)

cat("\n\n The corresponding p-value for the test statistic will be:\n\n              p-value=", p_value
```

```
##
##
##   The corresponding p-value for the test statistic will be:
##
##                 p-value= 0.5458821
```

The p-value of 0.99 (»0.05) tell us that $\beta_0 = -1$ is also not likely.