Simple Linear Regression Analysis: LDL and Systolic Blood Pressure

Statisticians World

October 15, 2024

Package Setup

```
library(tidyverse)
library(gtsummary)
library(ggplot2)
library(ggpubr)
library(GGally)
library(rsq)
library(broom)
library(broom.helpers)
library(labelled)
```

Importing Data

```
setwd("C:/Users/O&1/OneDrive/Documents/R-Youtube")
data <- read.csv("CHDdata.csv")
```

Exploratory Data Analysis

```
glimpse(data)
```

```
## Rows: 462
## Columns: 10
## $ sbp       <int> 160, 144, 118, 170, 134, 132, 142, 114, 114, 132, 206, 134, ~
## $ tobacco   <dbl> 12.00, 0.01, 0.08, 7.50, 13.60, 6.20, 4.05, 4.08, 0.00, 0.00~
## $ ldl       <dbl> 5.73, 4.41, 3.48, 6.41, 3.50, 6.47, 3.38, 4.59, 3.83, 5.80, ~
## $ adiposity <dbl> 23.11, 28.61, 32.28, 38.03, 27.78, 36.21, 16.20, 14.60, 19.4~
## $ famhist   <chr> "Present", "Absent", "Present", "Present", "Present", "Prese~
## $ typea     <int> 49, 55, 52, 51, 60, 62, 59, 62, 49, 69, 72, 65, 59, 49, 54, ~
## $ obesity   <dbl> 25.30, 28.87, 29.14, 31.99, 25.99, 30.77, 20.81, 23.11, 24.8~
## $ alcohol   <dbl> 97.20, 2.06, 3.81, 24.26, 57.34, 14.14, 2.62, 6.72, 2.49, 0.~
## $ age       <int> 52, 63, 46, 58, 49, 45, 38, 58, 29, 53, 60, 40, 17, 15, 53, ~
## $ chd       <int> 1, 1, 0, 1, 1, 0, 0, 1, 0, 1, 1, 1, 0, 0, 0, 0, 0, 1, 1, 1, ~
```

```
data %>%
  select(ldl, sbp) %>%
  tbl_summary()
```

Characteristic	N = 462[1]
ldl	4.34 (3.28, 5.80)
sbp	134 (124, 148)

[1] Median (Q1, Q3)

Visualizing Distributions & Outliers

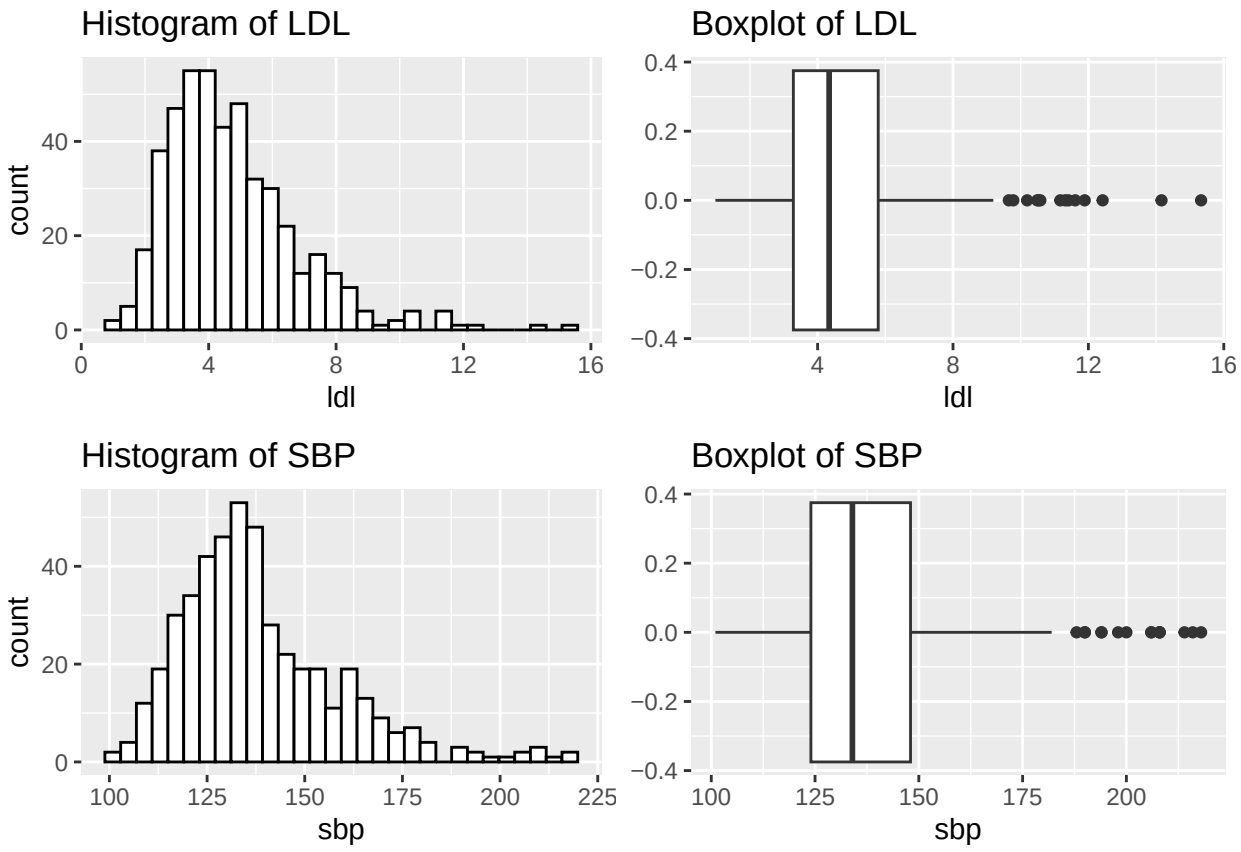
```
hist_ldl <- ggplot(data, aes(ldl)) +
  geom_histogram(color = "black", fill = "white") +
  labs(title = "Histogram of LDL")

hist_sbp <- ggplot(data, aes(sbp)) +
  geom_histogram(color = "black", fill = "white") +
  labs(title = "Histogram of SBP")

bplot_ldl <- ggplot(data, aes(ldl)) +
  geom_boxplot() +
  labs(title = "Boxplot of LDL")

bplot_sbp <- ggplot(data, aes(sbp)) +
  geom_boxplot() +
  labs(title = "Boxplot of SBP")

ggarrange(hist_ldl, bplot_ldl, hist_sbp, bplot_sbp,
          ncol = 2, nrow = 2)
```



Simple Linear Regression Model

Model: ldl ~ sbp

```
slr_ldl <- lm(ldl ~ sbp, data = data)
summary(slr_ldl)
```

```
##
## Call:
## lm(formula = ldl ~ sbp, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.6591 -1.3770 -0.4412  1.0589 10.8828
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 2.527930   0.650441   3.886 0.000117 ***
## sbp         0.015994   0.004652   3.438 0.000638 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.047 on 460 degrees of freedom
## Multiple R-squared:  0.02506,    Adjusted R-squared:  0.02294
## F-statistic: 11.82 on 1 and 460 DF,  p-value: 0.0006383
```

```r
tidy(slr_ldl, conf.int = TRUE)
```

```
## # A tibble: 2 x 7
##   term        estimate std.error statistic  p.value conf.low conf.high
##   <chr>          <dbl>     <dbl>     <dbl>    <dbl>    <dbl>     <dbl>
## 1 (Intercept)   2.53     0.650        3.89 0.000117  1.25       3.81
## 2 sbp           0.0160   0.00465      3.44 0.000638  0.00685    0.0251
```

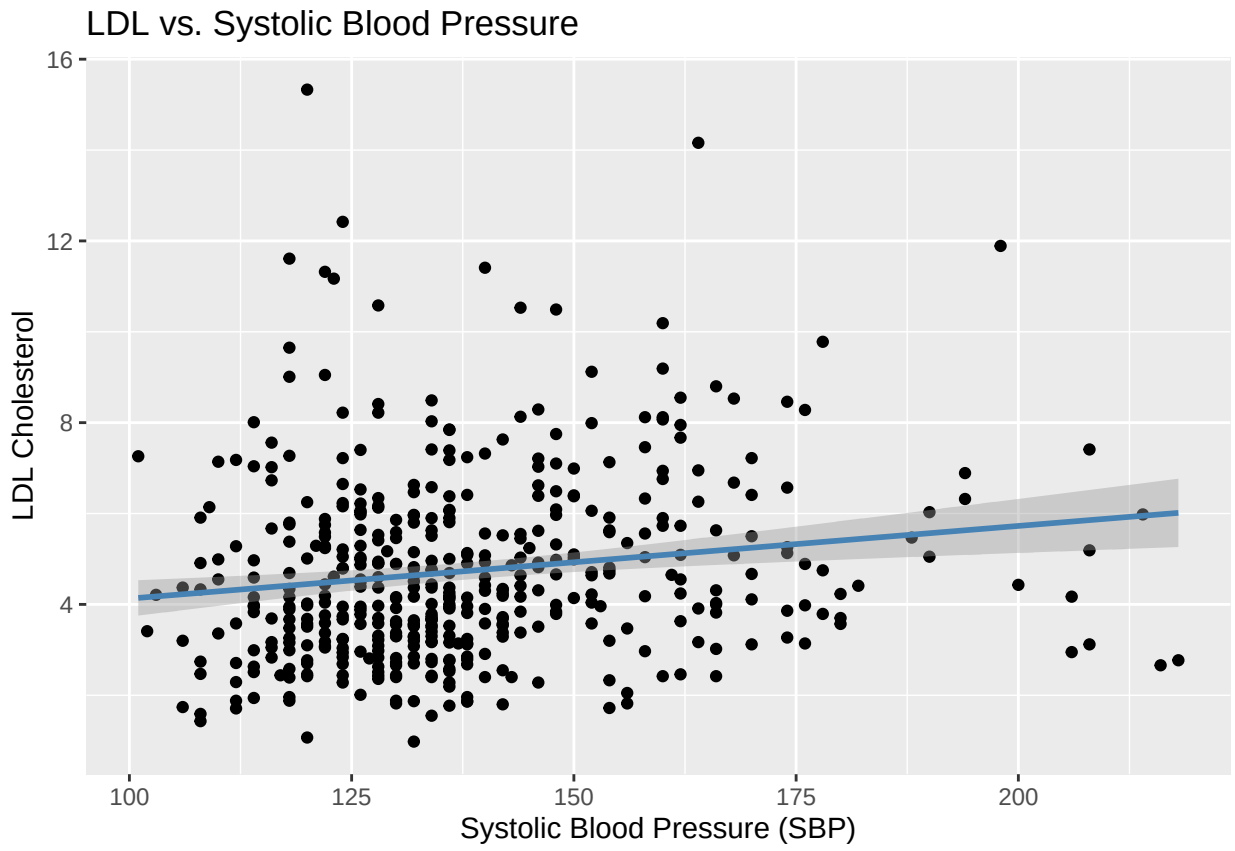
Model Fit

```r
rsq(slr_ldl)
```

```
## [1] 0.02505773
```

Regression Line Visualization

```r
ggplot(data, aes(x = sbp, y = ldl)) +
  geom_point() +
  geom_smooth(method = "lm", se = TRUE, color = "steelblue") +
  labs(title = "LDL vs. Systolic Blood Pressure",
       x = "Systolic Blood Pressure (SBP)",
       y = "LDL Cholesterol")
```



Characteristic	Beta	95% CI[1]	p-value
sbp	0.02	0.01, 0.03	<0.001

[1] CI = Confidence Interval

Regression Output Table

```
tbl_regression(slr_ldl)
```

Conclusion

This analysis explored the relationship between LDL cholesterol and systolic blood pressure using simple linear regression. Preliminary results suggest a statistically significant linear relationship between the two variables. Always check model assumptions and consider clinical significance alongside statistical results.