Отчет по домашнему заданию

Фахртдинов Т. А.

24 февраля 2020 г.

Восьмая задача. Исследование зависимости между количественными признаками.
Вариант 3.

```
h <- c(1.1, 4.6, 0.7, 4.0, 0.7, 27.6, 20.9, 55.6, 69.0, 23.0, 19.5, 8.9, 50.4,
       39.9, 20.7, 26.6, 13.9, 23.6, 16.2, 29.9, 13.9, 65.2, 31.4, 26.0, 25.0)
t <- c(73.2, 71.9, 76.4, 72.1, 71.9, 83.5, 85.9, 85, 82.3, 84.5, 85.3, 85.3,
       84.1, 74.3, 71, 69.5, 59.1, 73.1, 75.7, 81.9, 62.1, 72.6, 72.6, 77.9, 77.7)
```

Получаем оценки параметров линейной регрессии.

```
n <- length(h)
beta <- (sum(h * t) - n * mean(h) * mean(t)) /
        (sum(h * h) - n * mean(h)^2)
alpha <- mean(t) - beta * mean(h)
```

Коэффициенты α, β соответственно равны:

```
## [1] 73.3662933  0.1208841
```

Проверим значимость оценок:

```
Y <- alpha + beta * h
Qe <- sum((t - Y)^2)
S2 <- Qe / (n - 2)
S2.alpha <- S2 * sum(h * h) / (n * sum((h - mean(h))^2))
S2.beta <- S2 / sum((h - mean(h))^2)
T_a <- alpha / sqrt(S2.alpha)
T_b <- beta / sqrt(S2.beta)
```

Значения критерия и p-value, для α и β соответственно:

```
## [1] 31.44714  0.00000
## [1] 1.6049144 0.1221572
```

Найдем значение параметров с помощью встроенной функции:

```
summary(lm(t~h))

##
## Call:
## lm(formula = t ~ h)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -15.9466  -3.8896   0.3754   4.9125  10.8578
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 73.36629    2.33300  31.447   <2e-16 ***
## h            0.12088    0.07532   1.605    0.122
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.023 on 23 degrees of freedom
## Multiple R-squared:  0.1007,Adjusted R-squared:  0.06161
## F-statistic: 2.576 on 1 and 23 DF,  p-value: 0.1222
```

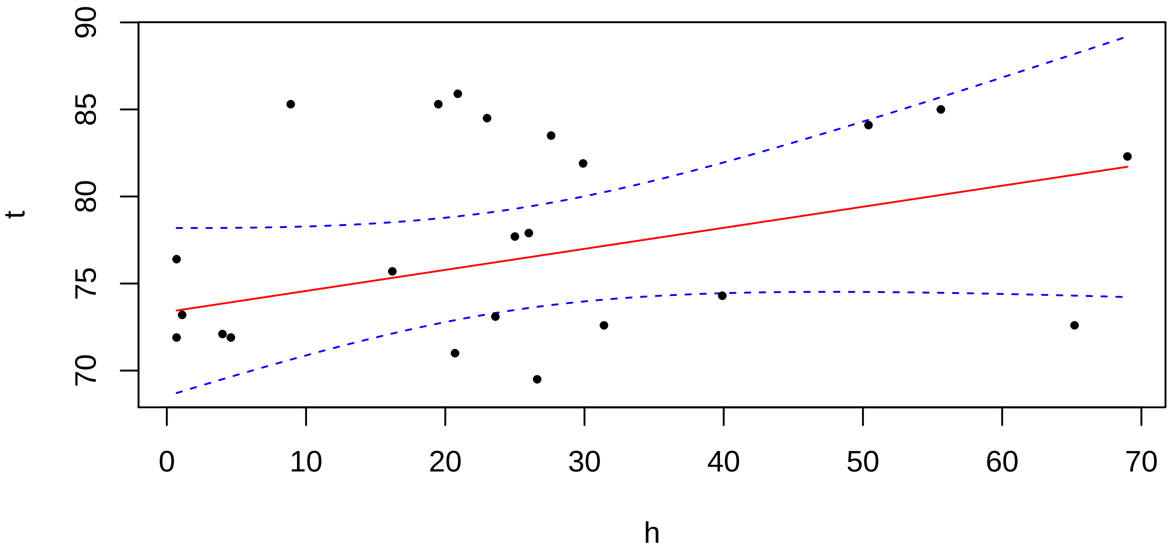
Полученные значения параметров совпадают со значениями, полученными нами. Так же, полученные значения критериев и p-value совпадают с полученными нами.

$p - value_{\alpha} < 0.05$, отклоняем гипотезу о равенстве коэффициента нулю.

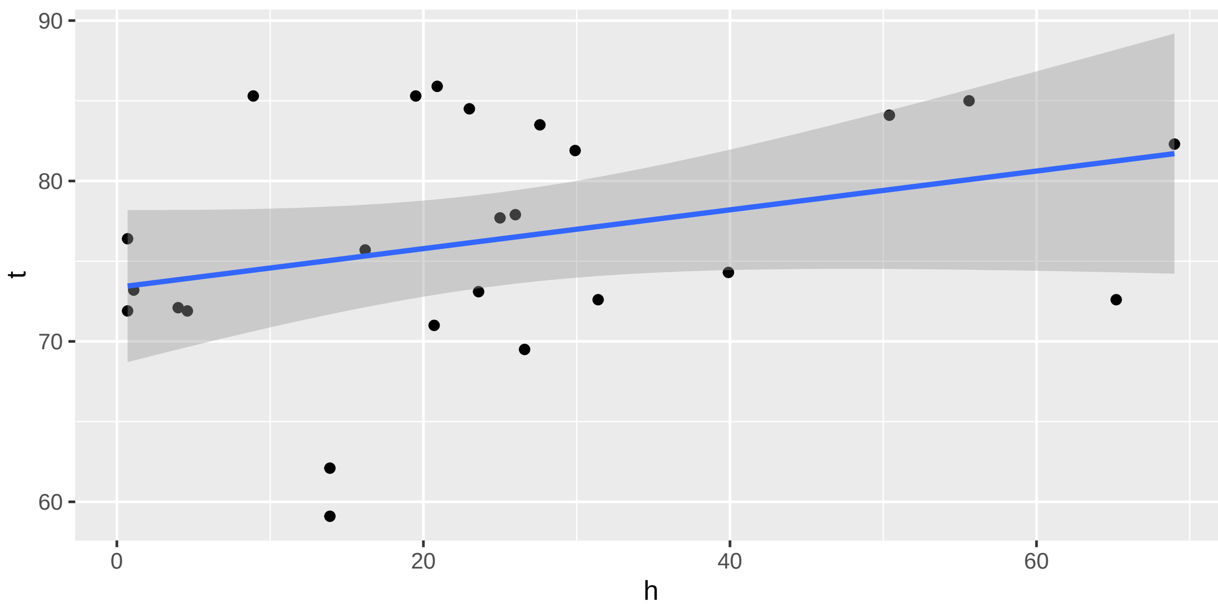
$p - value_{\beta} > 0.05$, нет оснований отклонить гипотезу о равенстве коэффициента нулю.

На одном графике строим двумерную диаграмму признаков и проводим линию регрессии.

```
step <- seq(min(h), max(h), by = 0.1)
Y <- alpha + beta*step
S <- sqrt(S2*( 1/n + (step - mean(h))^2/sum((h - mean(h))^2)))
confInt1 <- Y - qt(1-0.05/2, n - 2) * S
confInt2 <- Y + qt(1-0.05/2, n - 2) * S
confInt <- cbind(confInt1, confInt2, Y)
matplot(step,
confInt, col=c("blue","blue","red"),lty=c(2,2,1),
type= "l", ylab = "t", xlab = "h")
points(h, t, pch = 19, cex = 0.5)
```



Теперь воспользуемся функцией пакета ggplot, для построения:



Вычислим коэффициенты корреляции и детерминации.

Коэффициент детерминации:

```
Q <- sum((t - mean(t))^2)
R_2 <- 1 - Qe / Q
R_2

## [1] 0.1007106
```

Коэффициент корреляции:

```
rho <- sum((h - mean(h)) * (t - mean(t)))
rho <- rho / (sqrt(sum((h - mean(h))^2)) * sqrt(sum((t - mean(t))^2)))
rho

## [1] 0.3173494
```

Значимость коэффициента корреляции:

Проверяется гипотеза H_0: $\rho = 0$

```
T_rho<- rho * (n - 2) / sqrt(1 - rho^2)
pval_rho <- 2 * min(1 - pt(T_rho, n - 2), pt(T_rho, n - 2))
```

Значение критерия и p-value:

```
## [1] 7.696899e+00 8.266966e-08
```

8.266966e-08 $<$ 0.05, отклоняем гипотезу о равенстве коэффициента корреляции нулю.

Значимость коэффициента детерминации:

Проверяется гипотеза H_0: $R^2 = 0$

```
T_r2<- sqrt(R_2) * (n - 2) / sqrt(1 - R_2)
pval_r2 <- 2 * min(1 - pt(T_r2, n - 2), pt(T_r2, n - 2))
```

Значение критерия и p-value:

```
## [1] 7.696899e+00 8.266966e-08
```

8.266966e-08 $<$ 0.05, отклоняем гипотезу о равенстве коэффициента детерминации нулю. Что ожидаемо, так как проверка гипотезы о равенстве коэффициента детерминации нулю эквивалентна проверке гипотезы о равенстве нулю коэффициента корреляции.