homework03

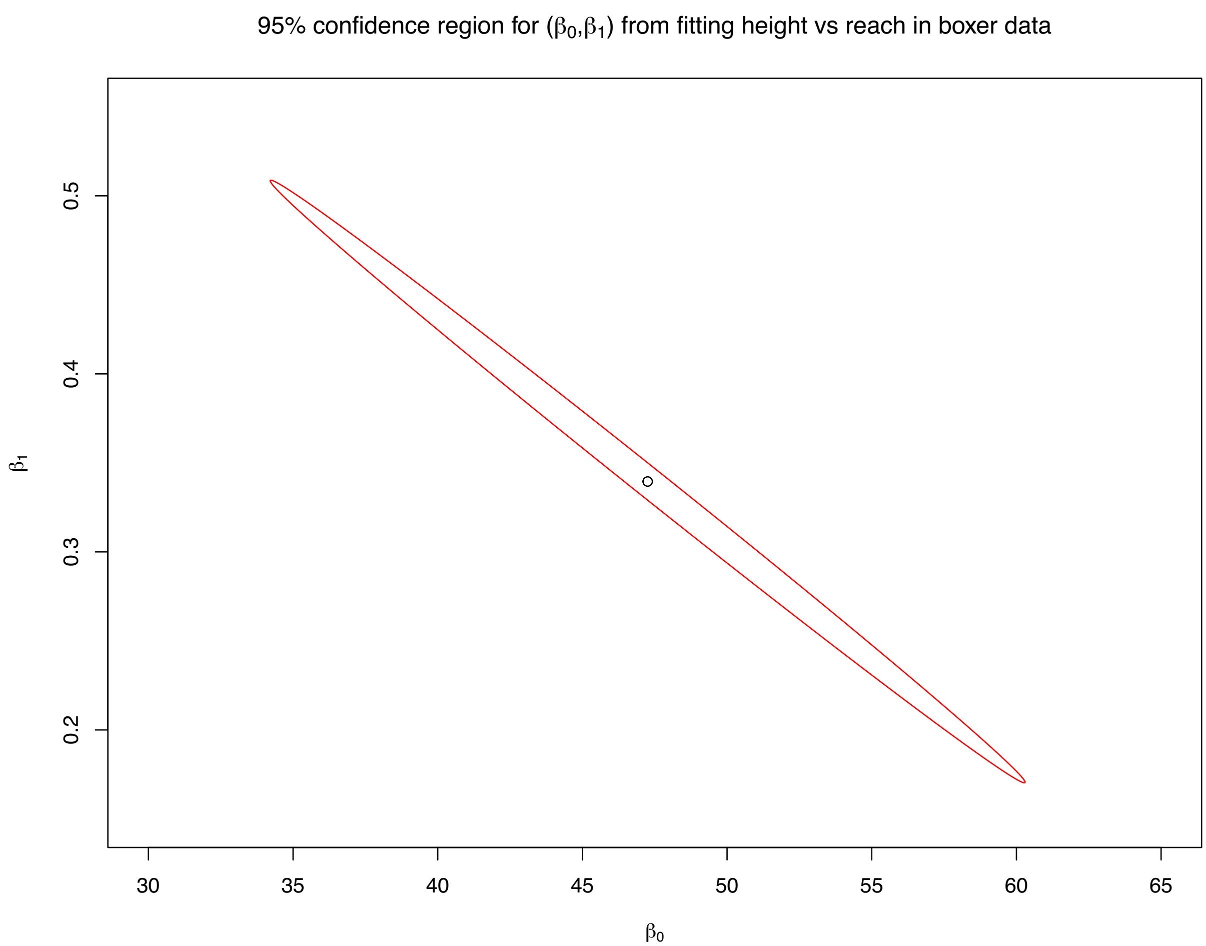
Jiapeng Wang

2023-10-12

Question1:

```
rm(list = ls())
data <- read.csv("boxers.csv", header = TRUE, sep = ",")
x = data$reach
y = data$height
model = lm(y ~ x)
row1 <- rep(1, 19)
X1 <- rbind(row1, data$reach)
X1 = t(X1)
L <- t(X1) %*% X1
y_hat = fitted(model)
e = y - y_hat
s = sqrt(sum(e^2) / (19 - 2))

beta_hat = c(model$coefficients[1], model$coefficients[2])
beta_0 <- seq(30, 65, length.out = 500)

F_value <- qf(0.95, df1=2, df2=17, lower.tail=TRUE)

s_y = 2 * s^2 * F_value
y_lower = rep(0,500)
y_upper = rep(0,500)

#Solve quadratic equations with one variable
getroot = function(a,b,c){
  x1 = (-b + (b^2 - 4 * a * c)^0.5) / 2 / a
  x2 = (-b - (b^2 - 4 * a * c)^0.5) / 2 / a
  ans=c(x1,x2)
  return(ans)
}

for (i in 1:500) {
  solutions <- getroot(L[2,2], (L[1,2] + L[2,1]) * (beta_hat[1] - beta_0[i]), L[1,1] * (beta_hat[1] - be
  y_lower[i] = beta_hat[2] - max(solutions)
  y_upper[i] = beta_hat[2] - min(solutions)
}

plot(beta_hat[1], beta_hat[2],
     main = expression(paste('95% confidence region for (', beta[0], ',', beta[1], ") from fitting heig
     xlab = expression(paste(beta[0])),
     ylab = expression(paste(beta[1])),
```

```
     xlim = c(30,65),
     ylim = c(0.15,0.55))
lines(beta_0, y_lower, col='red', lty = 1)
lines(beta_0, y_upper, col='red', lty = 1)
```



Q2:

2. Show that the Box-Cox λ can be computed as follows:

   (a) Let $\dot{y} = \sqrt[n]{y_1 y_2 \ldots y_n}$ denote the geometric mean of the y observations

   (b) For each λ, transform the y_i to

$$z_i(\lambda) = \begin{cases} \frac{y_i^\lambda - 1}{\lambda \dot{y}^{\lambda-1}}, & \lambda \neq 0 \\ \dot{y}\log y_i, & \lambda = 0 \end{cases}$$

   (c) Fit the desired linear regression model to $z_1(\lambda), z_2(\lambda), \ldots, z_n(\lambda)$ and compute its residual sum of squares $S(\lambda)$

   (d) The Box-Cox λ minimizes $S(\lambda)$

We already have maximum likelihood way to estimate λ, the likelihood function is

$$\frac{1}{(2\pi)^{n/2}\sigma^n}\exp\left\{-\frac{(y^{(\lambda)}-X\beta)'(y^{(\lambda)}-X\beta)}{2\sigma^2}\right\} J(\lambda, y) \quad \text{where } J=\prod_{i=1}^{n} y_i^{\lambda-1}$$

Using this we could find that $\hat{\beta}_{MLE} = (X'X)^{-1}X'y^{(\lambda)}$, $\hat{\sigma}^2_{MLE} = \frac{1}{n}y^{(\lambda)'}(I-X(X'X)^{-1}X')y^{(\lambda)}$

Then, $\log L_{max}(\lambda) = \log L(\hat{\beta}_{MLE}, \hat{\sigma}^2_{MLE}) = -\frac{n}{2}\log\left(y^{(\lambda)'}(I-X(X'X)^{-1}X')y^{(\lambda)}\right) + \log J$

$$= -\frac{n}{2}\log\left(\frac{y^{(\lambda)'}}{J^{\frac{1}{n}}}\cdot(I-X(X'X)^{-1}X')\frac{y^{(\lambda)}}{J^{\frac{1}{n}}}\right)$$

And we suppose:

$z_i(\lambda) = \frac{1}{J^{\frac{1}{n}}}y_i^{(\lambda)}$, $S(\lambda)$ is residual sum of squares for $z_i(\lambda)$

So $\log L_{max}(\lambda) = -\frac{n}{2}\log(S(\lambda))$

Thus we minimize $S(\lambda)$ $\Leftrightarrow$ we find the MLE of λ

So Box-Cox λ can be computed by minimizing $S(\lambda)$.

$\square$