Meher Shrishti Nigam

20BRS1193

EDA Lab 3, Q2

options(prompt="MEHERSHRISHTI>", continue =" ")

options(prompt=">", continue =" ")

EDA-LAB-EXPERIMENT-3 (Date-13/1/2023)

The ends of chromosomes are called telomeres. These telomeres are shortened a bit

during each cell cycle as DNA is replicated. One of their purposes is to protect more

valuable DNA in the chromosome from degradation during replication. As people get older

and their cells have replicated more often, their telomeres shorten. There is evidence that

these shortened telomeres may play a role in aging. Telomeres can be lengthened in germ

cells and stem cells by an enzyme called telomerase, but this enzyme is not active in most

healthy somatic cells. (Cancer cells, on the other hand, usually express telomerase.)

Given that the length of telomeres is biologically important, it becomes interesting to know

whether telomere length varies between individuals and whether this variation is inherited. A

set of data was collected by Nordfjäll et al. (2005) on the telomere length of fathers and their

children; these data are in the file "telomere inheritance.csv".

df <- read.csv("telomere inheritance.csv")

df

```
MEHERSHRISHTI>df <- read.csv("telomere inheritance.csv")
MEHERSHRISHTI>df
   father_telomere_length offspring_telomere_length
1                   0.281                     0.410
2                   0.301                     0.311
3                   0.282                     0.582
4                   0.425                     0.574
5                   0.463                     0.592
6                   0.514                     0.614
7                   0.506                     0.657
8                   0.535                     0.704
9                   0.556                     0.661
10                  0.590                     0.692
11                  0.490                     0.756
```
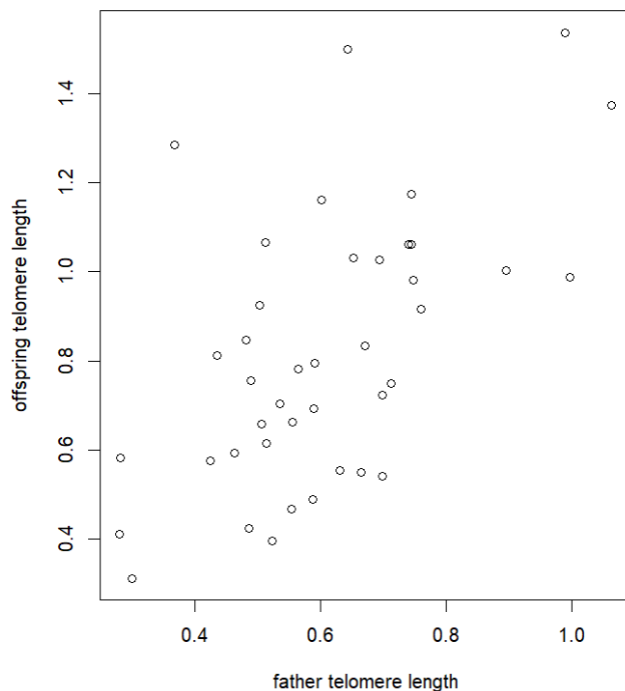
a. Create a scatter plot showing the relationship between father and offspring telomere length.

plot(x=df$father_telomere_length, y=df$offspring_telomere_length, xlab="father telomere length", ylab="offspring telomere length")

b. Do the data require any transformation before analysis using linear regression?

```
MEHERSHRISHTI>plot(x=df$father_telomere_length, y=df$offspring_telomere_lengt
h, xlab="father telomere length", ylab="offspring telomere length")
MEHERSHRISHTI># b. Do the data require any transformation before analysis usi
ng linear regression?
MEHERSHRISHTI>linear_model <- lm(df$offspring_telomere_length ~ df$father_tel
omere_length)
MEHERSHRISHTI>cor(df$offspring_telomere_length , df$father_telomere_length)
[1] 0.579086
MEHERSHRISHTI>summary(linear_model)

Call:
lm(formula = df$offspring_telomere_length ~ df$father_telomere_length)

Residuals:
     Min       1Q   Median       3Q      Max
-0.36768 -0.17438 -0.04824  0.12324  0.69947

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)                 0.2252     0.1396   1.613    0.115
df$father_telomere_length   0.9792     0.2208   4.436 7.29e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2498 on 39 degrees of freedom
Multiple R-squared:  0.3353,    Adjusted R-squared:  0.3183
F-statistic: 19.68 on 1 and 39 DF,  p-value: 7.287e-05

MEHERSHRISHTI>abline(linear_model)
MEHERSHRISHTI>|
```
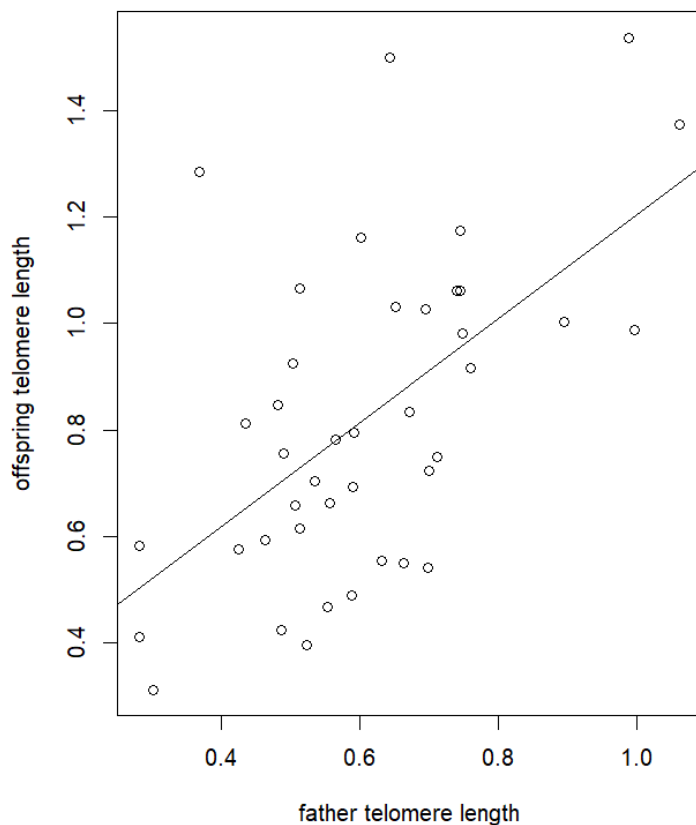
linear_model <- lm(df$offspring_telomere_length ~ df$father_telomere_length)

cor(df$offspring_telomere_length , df$father_telomere_length)

summary(linear_model)

abline(linear_model)

The linear model does not fit the data too well, and the attributes only have moderate positive correlation.

linear_model <- lm(df$offspring_telomere_length ~ sqrt(df$father_telomere_length))

plot(x=sqrt(df$father_telomere_length), y=df$offspring_telomere_length)

cor(df$offspring_telomere_length , sqrt(df$father_telomere_length))

summary(linear_model)

abline(linear_model)

```
MEHERSHRISHTI>linear_model <- lm(df$offspring_telomere_length ~ sqrt(df$fat
her_telomere_length))
MEHERSHRISHTI>plot(x=sqrt(df$father_telomere_length), y=df$offspring_telome
re_length)
MEHERSHRISHTI>cor(df$offspring_telomere_length , sqrt(df$father_telomere_le
ngth))
[1] 0.5668155
MEHERSHRISHTI>summary(linear_model)

Call:
lm(formula = df$offspring_telomere_length ~ sqrt(df$father_telomere_lengt
h))

Residuals:
     Min      1Q  Median      3Q     Max
-0.37498 -0.17118 -0.05309  0.14043  0.71002

Coefficients:
                                 Estimate Std. Error t value Pr(>|t|)
(Intercept)                       -0.3290     0.2703  -1.217 0.230819
sqrt(df$father_telomere_length)    1.4901     0.3468   4.297 0.000112

(Intercept)
sqrt(df$father_telomere_length) ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2524 on 39 degrees of freedom
Multiple R-squared:  0.3213,     Adjusted R-squared:  0.3039
F-statistic: 18.46 on 1 and 39 DF,  p-value: 0.0001117

MEHERSHRISHTI>abline(linear_model)
```
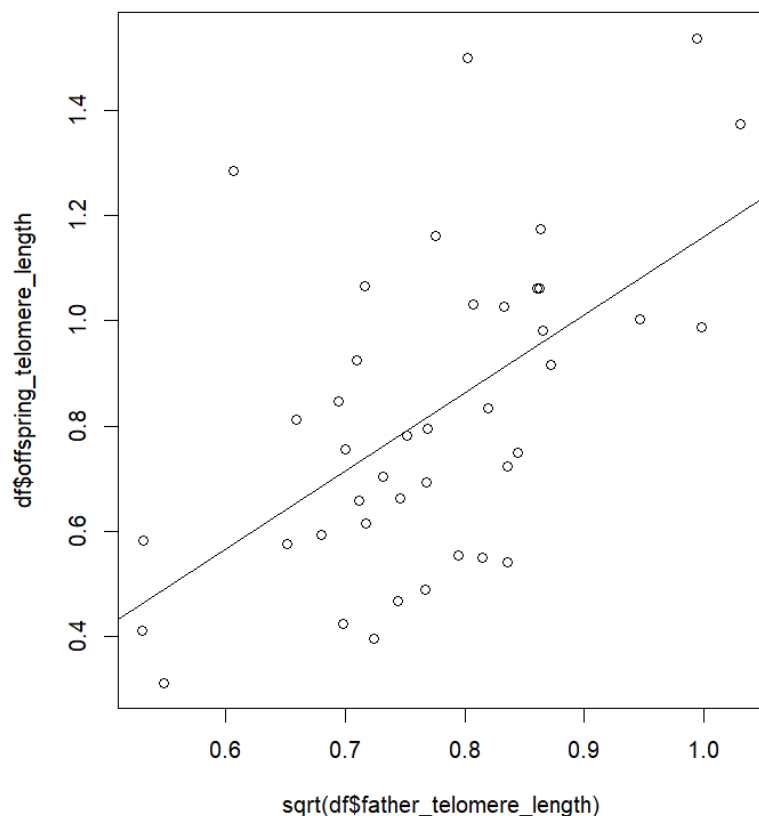
This transformation does not improve the correlation or linear model, it is similar to the original.

install.packages("pracma")

library(pracma)

linear_model <- lm(df$offspring_telomere_length ~ nthroot(df$father_telomere_length, 3))

plot(x=nthroot(df$father_telomere_length, 3), y=df$offspring_telomere_length)

cor(df$offspring_telomere_length , nthroot(df$father_telomere_length, 3))

summary(linear_model)

abline(linear_model)

```
MEHERSHRISHTI>library(pracma)
MEHERSHRISHTI>linear_model <- lm(df$offspring_telomere_length ~ nthroot(df$father_telomere_length, 3))
MEHERSHRISHTI>plot(x=nthroot(df$father_telomere_length, 3), y=df$offspring_telomere_length)
MEHERSHRISHTI>cor(df$offspring_telomere_length , nthroot(df$father_telomere_length, 3))
[1] 0.5613966
MEHERSHRISHTI>summary(linear_model)

Call:
lm(formula = df$offspring_telomere_length ~ nthroot(df$father_telomere_length,
    3))

Residuals:
    Min      1Q   Median      3Q      Max
-0.37647 -0.15590 -0.05202  0.13656  0.71180

Coefficients:
                                       Estimate Std. Error t value Pr(>|t|)
(Intercept)                             -0.8742     0.4018  -2.176 0.035708 *
nthroot(df$father_telomere_length, 3)    2.0198     0.4768   4.237 0.000134 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2535 on 39 degrees of freedom
Multiple R-squared:  0.3152,    Adjusted R-squared:  0.2976
F-statistic: 17.95 on 1 and 39 DF,  p-value: 0.0001342

MEHERSHRISHTI>abline(linear_model)
MEHERSHRISHTI>
```
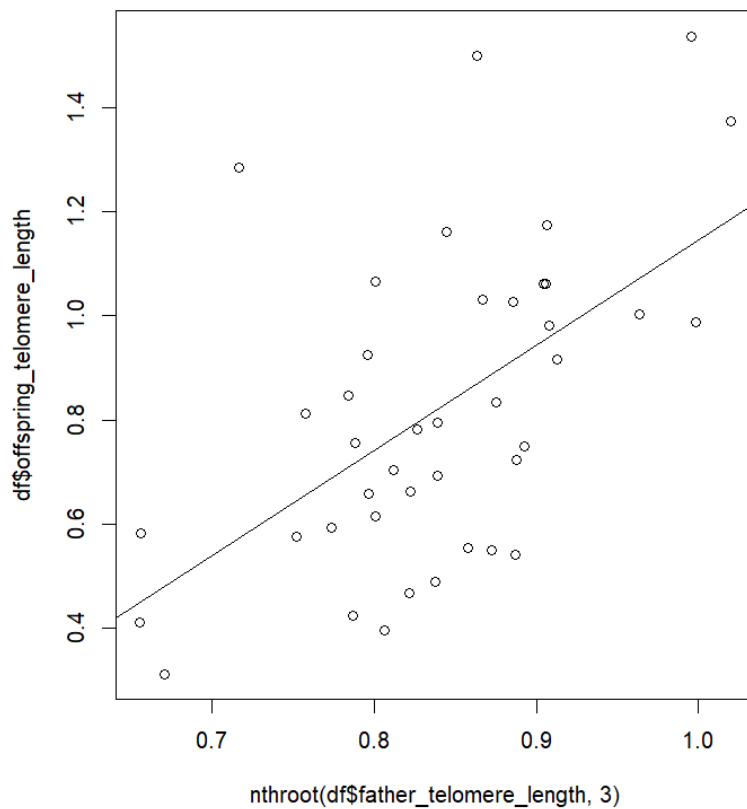
This as well does not improve the linear model significantly.

 # c. Estimate an equation that predicts the offspring telomere length from its father's. Is

 # there evidence that the father's telomere length predicts his offspring's value?


```
linear_model <- lm(df$offspring_telomere_length ~ sqrt(df$father_telomere_length))

plot(x=sqrt(df$father_telomere_length), y=df$offspring_telomere_length)

cor(df$offspring_telomere_length , sqrt(df$father_telomere_length))

summary(linear_model)

abline(linear_model)


sqrt_ftl <- sqrt(df$father_telomere_length)

linear_model <- lm(df$offspring_telomere_length ~ sqrt_ftl)

telomere_test <- data.frame(sqrt_ftl = c(0.8))

predict(linear_model, newdata = telomere_test)
```

```
MEHERSHRISHTI>linear_model <- lm(df$offspring_telomere_length ~ sqrt(df$father_telomere_length))
MEHERSHRISHTI>plot(x=sqrt(df$father_telomere_length), y=df$offspring_telomere_length)
MEHERSHRISHTI>cor(df$offspring_telomere_length , sqrt(df$father_telomere_length))
[1] 0.5668155
MEHERSHRISHTI>summary(linear_model)

Call:
lm(formula = df$offspring_telomere_length ~ sqrt(df$father_telomere_length))

Residuals:
     Min       1Q   Median       3Q      Max
-0.37498 -0.17118 -0.05309  0.14043  0.71002

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                      -0.3290     0.2703  -1.217 0.230819
sqrt(df$father_telomere_length)   1.4901     0.3468   4.297 0.000112 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2524 on 39 degrees of freedom
Multiple R-squared:  0.3213,    Adjusted R-squared:  0.3039
F-statistic: 18.46 on 1 and 39 DF,  p-value: 0.0001117

MEHERSHRISHTI>abline(linear_model)
MEHERSHRISHTI>sqrt_ftl <- sqrt(df$father_telomere_length)
MEHERSHRISHTI>linear_model <- lm(df$offspring_telomere_length ~ sqrt_ftl)
MEHERSHRISHTI>telomere_test <- data.frame(sqrt_ftl = c(0.8))
MEHERSHRISHTI>predict(linear_model, newdata = telomere_test)
       1
0.8631302
```
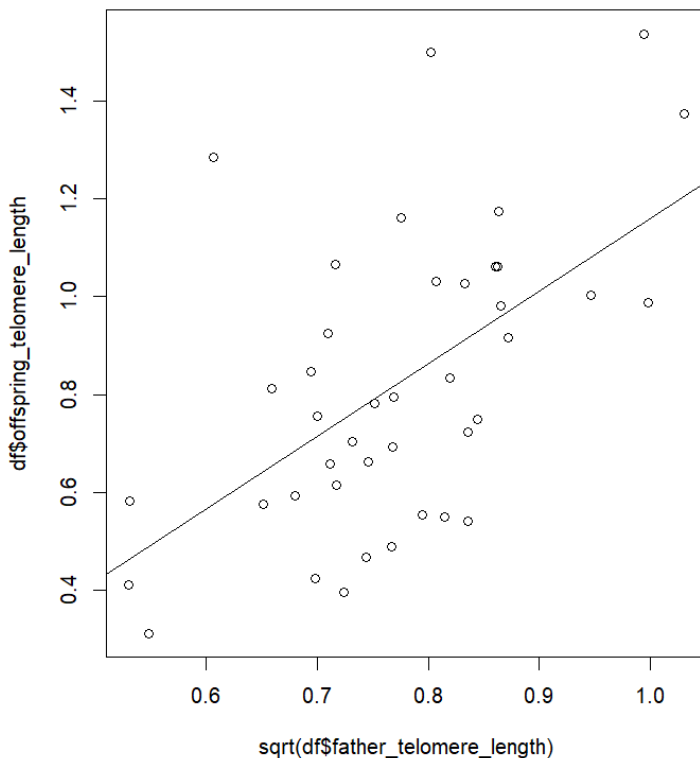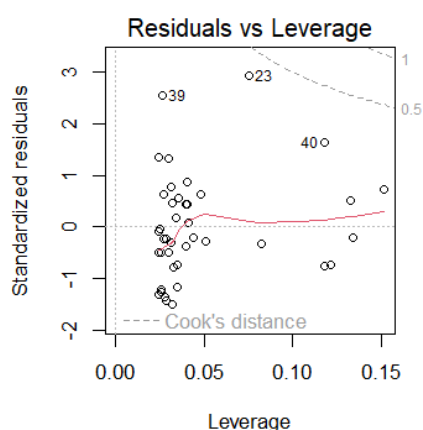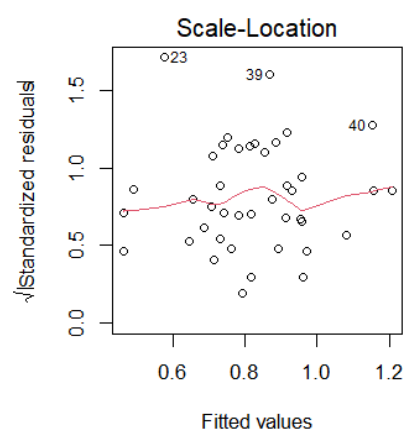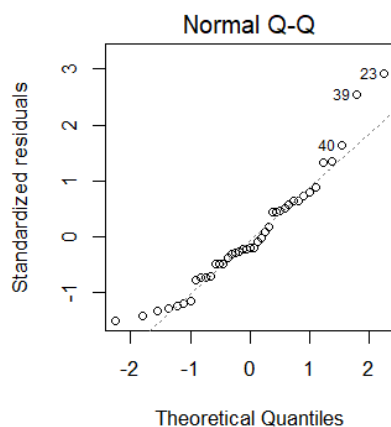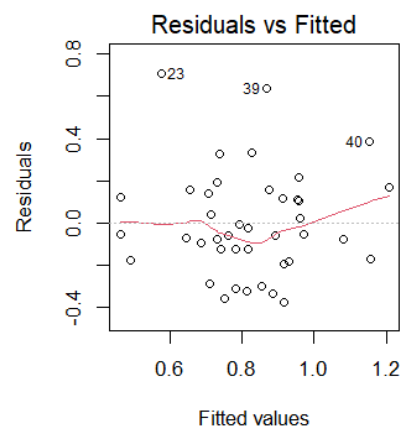


```
par(mfrow = c(2, 2))

plot(linear_model)
```

```
MEHERSHRISHTI>par(mfrow = c(2, 2))
MEHERSHRISHTI>plot(linear_model)
MEHERSHRISHTI>
```

There is little evidence that says father's telomere length predicts offspring's value.
The residuals graph shows that the data moderately fits in the equation Y − sqrt(X) .

Although some correlation is deduced, the relationship is not linear, sqrt or cuberoot.