Assignment3 STAT291

2023-02-14

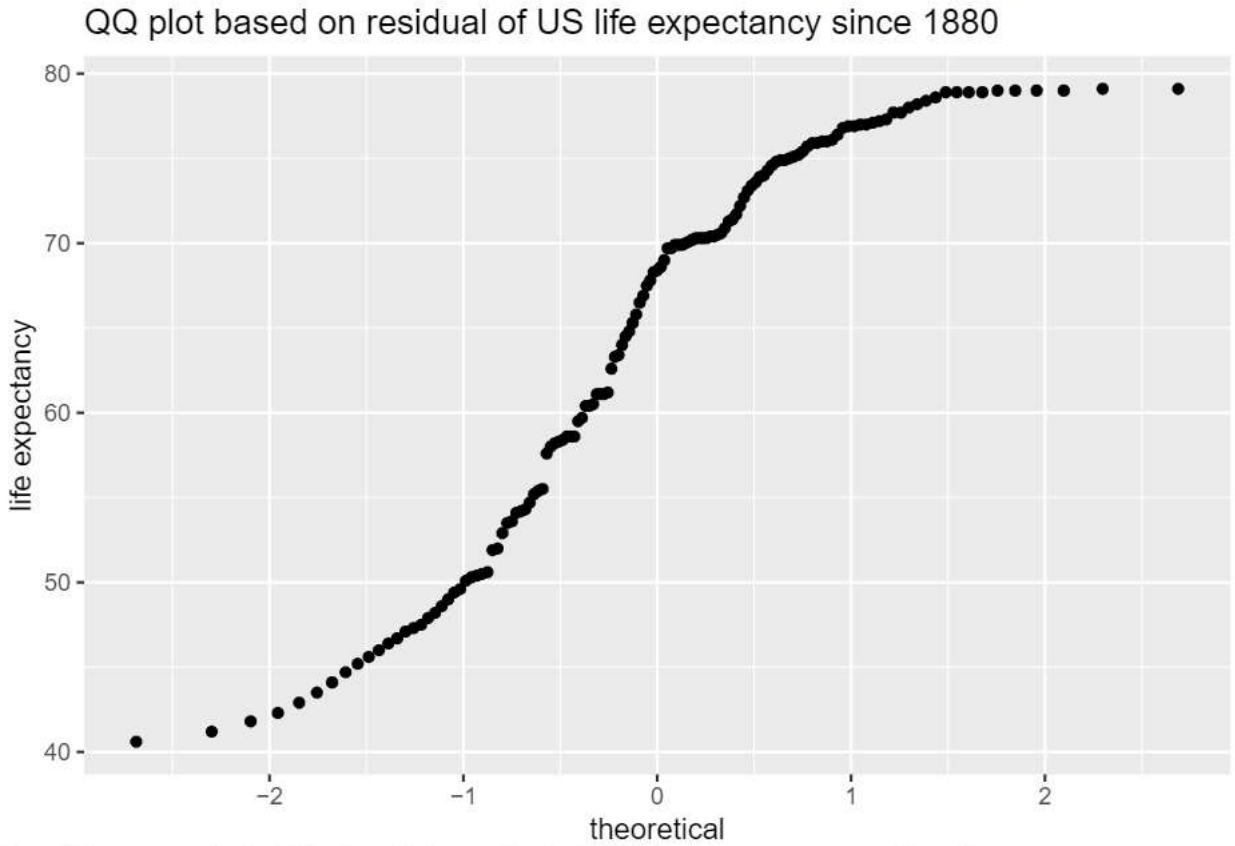
By, Lucas Weston

Section 1: Revisit the regression model of US life expectancy on number of years since 1880 in homework 2

Q1: Get the QQ plot of the residual

```
df <- read.csv("USlifehistory.csv")
rModel <- lm(life_expectancy~year, data = df)


ggplot(rModel, aes(sample=life_expectancy))+stat_qq()+
labs(title="QQ plot based on residual of US life expectancy since 1880", y = "life expectancy")
```

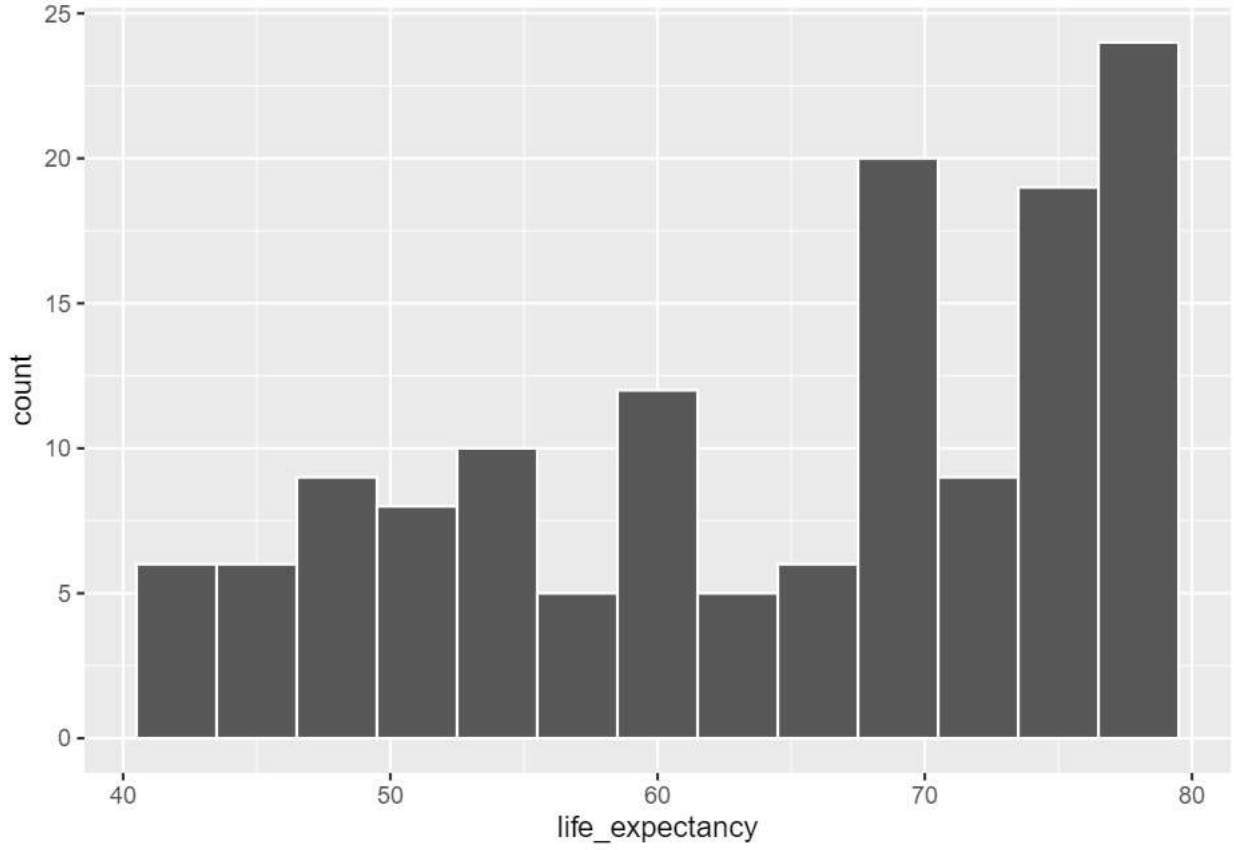


Here I have created the QQ plot of the residual of US life expectancy on number of years since 1880.

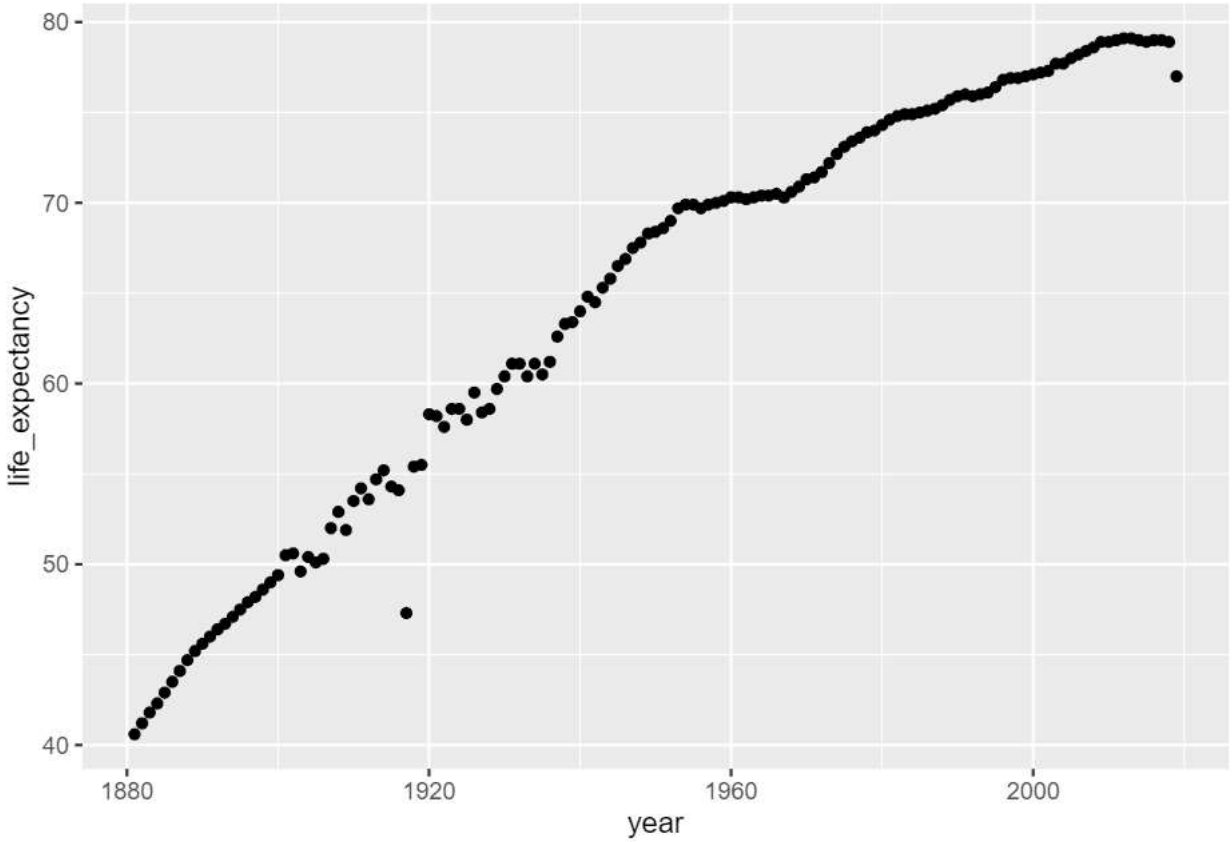
##Q2: Together with histogram of residual and scatter plot of residual vs. life expectancy, check the four assumptions in regression (LINE property).

```
df <- read.csv("USlifehistory.csv")
rModel <- lm(life_expectancy~year, data = df)

ggplot(data=rModel, aes(x=life_expectancy)) + geom_histogram(color="white", binwidth=3)
```




```
ggplot(rModel, aes(y=life_expectancy, x=year)) + geom_point()
```



Here I have created a histogram and scatterplot of the residual vs. the life expectancy.

Four assumptions for regression:

Linearity: Yes, it is linear.

Homoscedasticity: Yes, it is the same for any value.

Independence: Yes, observations are independent from each other.

Normality: Yes, they are normally distributed.

Section 2: Regression of child's height (gender adjusted) on mid-height of parent

Q1: Have a scatterplot of childs height vs. mid height of parent. On top of scatterplot, add the regression line and the diagonal line y=x, with different colors

```
gH <- read.csv("galton_height.csv")
```

Q2: What is the average children's height in the data? What is the average mid-height of the parent?

Q3: Among parents whose mid-height between 72 and 73 inches, what is the average height of their children?

Q4: Run regression, is the model significant?

Q5: If the parents' mid-height increases by 1 inch, what is the expected increase in child's height? Is the expected increase larger or smaller than 1 inch?

Q6: Estimate the child's height if the mid-height of parent is 64, 68, 70, 72, 76 respectively, and check their "closeness" to the mean height of all children

Section 3: Regression of mid-height of parent on child's height (gender adjusted)

Q1:Have a scatterplot of y vs. x. On top of scatterplot, add the regression line and the diagonal line y=x, with different colors

Q2: Among all children with height between 72 and 73 inches, what is the mean mid-height of their parents?

Q3: Run regression, is the model significant?

Q4: If the child's height increases by 1 inch, what is the expected increase in parent's mid-height? Is the expected increase larger or smaller than 1 inch?

Q5: Estimate the parent's mid-height if the child's height is 64, 68, 70, 72, 76 respectively,

and check their "closeness" to the mean mid-height of all parents

##Section 4: Use the above results to explain regression to the mean