Midterm Q3

Bhagyarathi Raman

3/9/2022

```
library(AER)
```

```
## Loading required package: car
```

```
## Loading required package: carData
```

```
## Loading required package: lmtest
```

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
## Loading required package: sandwich
```

```
## Loading required package: survival
```

```
data("Affairs")
names(Affairs)
```

```
## [1] "affairs"       "gender"        "age"           "yearsmarried"
## [5] "children"      "religiousness" "education"     "occupation"
## [9] "rating"
```

3. Model Fitting and Residual plots

a. Fit a linear model for affairs using all predictors. Call this model g. Show the coefficients and interpret the coefficent for childrenyes.

```
g=lm (affairs~gender+age+yearsmarried+children+education+religiousness+occupation+rating, data = Affairs)
summary(g)
```

```
##
## Call:
## lm(formula = affairs ~ gender + age + yearsmarried + children +
##     education + religiousness + occupation + rating, data = Affairs)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.0503 -1.7226 -0.7947  0.2101 12.7036
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    5.87201    1.13750   5.162 3.34e-07 ***
## gendermale     0.05409    0.30049   0.180   0.8572
## age           -0.05098    0.02262  -2.254   0.0246 *
## yearsmarried   0.16947    0.04122   4.111 4.50e-05 ***
## childrenyes   -0.14262    0.35020  -0.407   0.6840
## education     -0.01375    0.06414  -0.214   0.8303
## religiousness -0.47761    0.11173  -4.275 2.23e-05 ***
## occupation     0.10492    0.08888   1.180   0.2383
## rating        -0.71188    0.12001  -5.932 5.09e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.095 on 592 degrees of freedom
## Multiple R-squared:  0.1317, Adjusted R-squared:   0.12
## F-statistic: 11.23 on 8 and 592 DF,  p-value: 7.472e-15
```

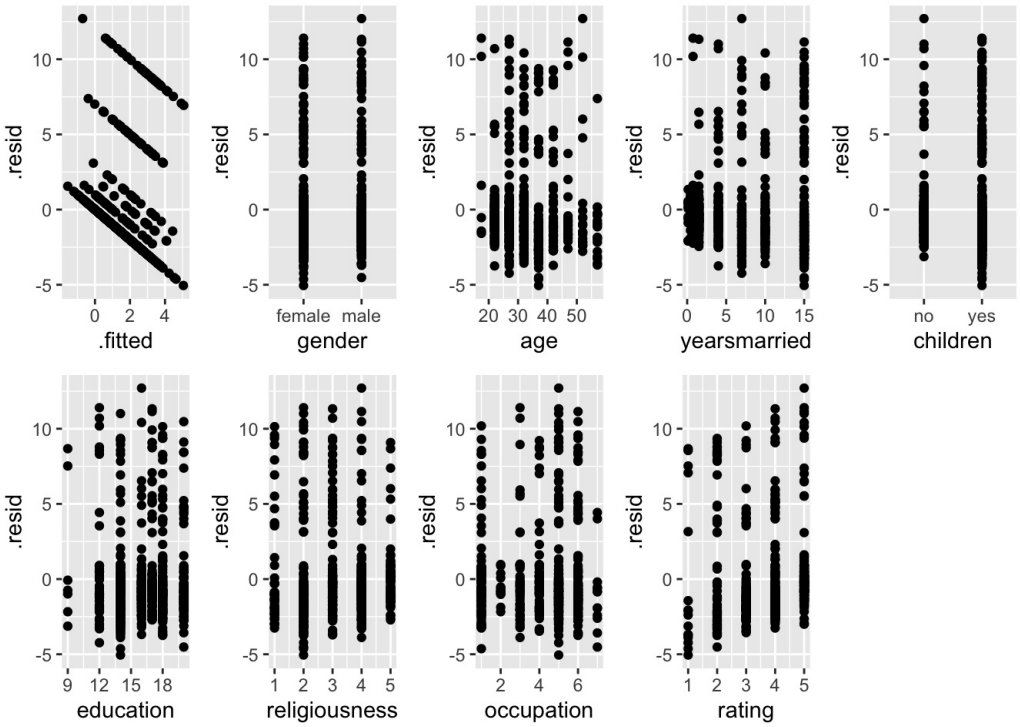
```
coef(g)
```

```
##   (Intercept)    gendermale           age  yearsmarried   childrenyes
##    5.87201014    0.05408587   -0.05097628    0.16947232   -0.14262446
##     education religiousness    occupation        rating
##   -0.01374903   -0.47761363    0.10491597   -0.71187692
```

```
# childrenyes   -0.14262
#The childrenyes coefficient is negative. the affairs tend to decrease when their are more children.
```

b. Obtain and show plots of the g-residuals against the fitted values. Does this plot reveal any abnormal patterns, if so why?

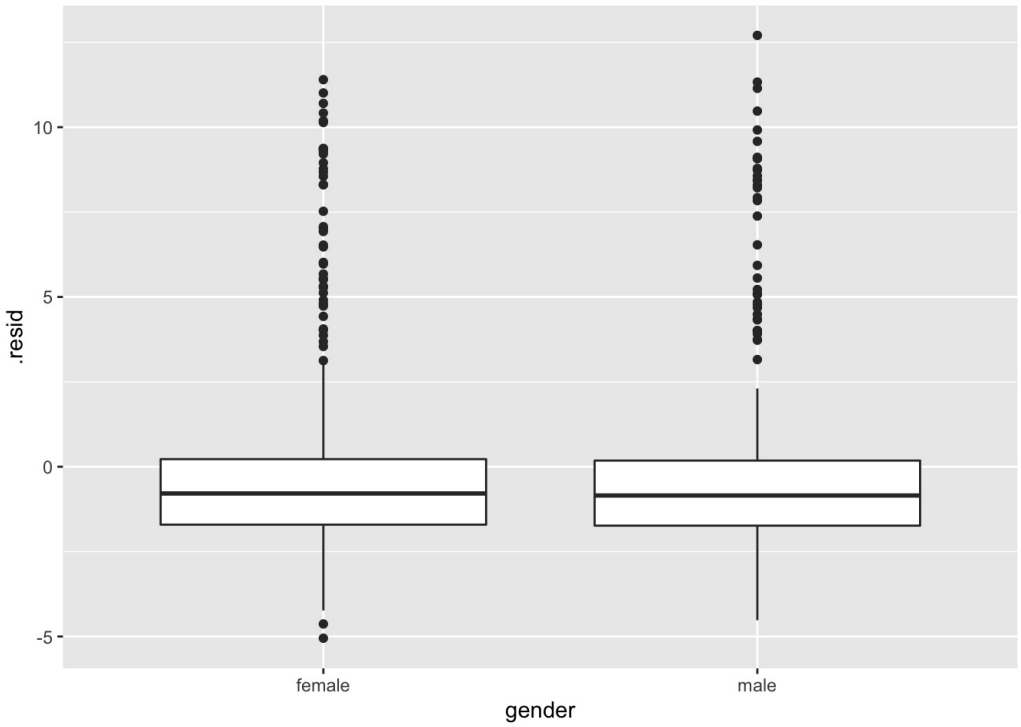
```
library(ggplot2)
mod.g <- fortify(g)
p1 <- ggplot(mod.g) +
  aes(x=.fitted, y=.resid) +
  geom_point()
p2 <- ggplot(mod.g) +
  aes(x=gender, y=.resid) +
  geom_point()
p3 <- ggplot(mod.g) +
  aes(x=age, y=.resid) +
  geom_point()
p4 <- ggplot(mod.g) +
  aes(x=yearsmarried, y=.resid) +
  geom_point()
p5<-ggplot(mod.g)+
  aes(x=children,y=.resid)+
  geom_point()
p6<-ggplot(mod.g)+
  aes(x=education,y=.resid)+
  geom_point()
p7<-ggplot(mod.g)+
  aes(x=religiousness,y=.resid)+
  geom_point()
p8<-ggplot(mod.g)+
  aes(x=occupation,y=.resid)+
  geom_point()
p9<-ggplot(mod.g)+
  aes(x=rating,y=.resid)+
  geom_point()

library(gridExtra)
grid.arrange(p1, p2, p3, p4,p5,p6,p7,p8,p9, nrow = 2)
```



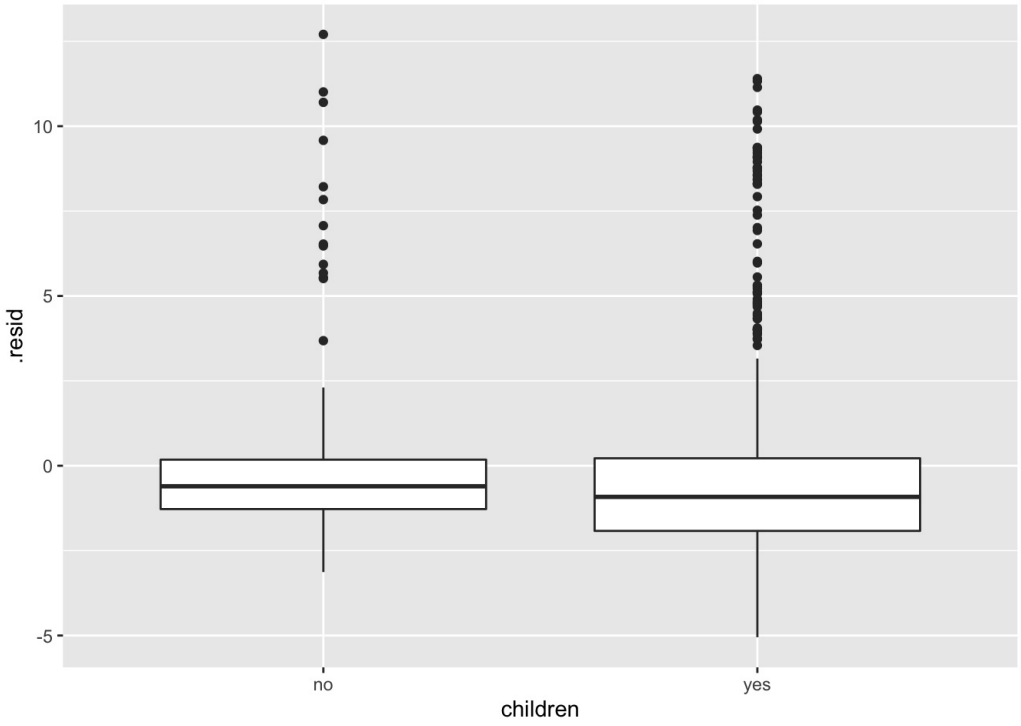

```
# The plot reveals that the error variance is not equal.
# All the plots are not linear and has various outliers.
```

c. Obtain and show the boxplots of the g-residuals versus gender and children. Does the plot reveal any patterns?

```
ggplot(mod.g)+
  aes(x=gender,.resid)+
  geom_boxplot()
```



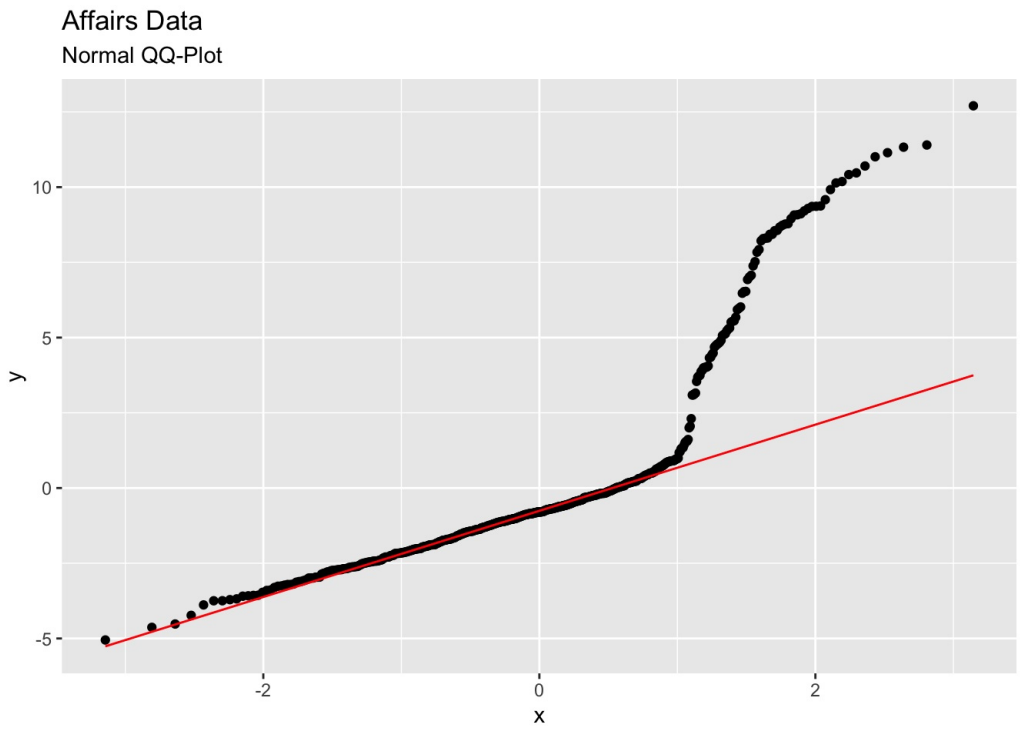

```
ggplot(mod.g)+
  aes(x=children,.resid)+
  geom_boxplot()
```




```
#The boxplots size (width) are different, indicating variance error between gender.
# The boxplot of the g-residuals with gender, the male shows outliers than female.
# The boxplot size of the g-residuals with childrenno shows outliers than childrenyes and the boxplots with child
renyes is bigger than the childrenno.
```

d. Using the g-residuals, obtain and show the Normal QQ-Plot. Does the plot indicate the residuals are normal or not normal? Explain your answer.

```
(p1 <- ggplot(mod.g) +
  aes(sample=.resid) +
  geom_qq() +
  geom_qq_line(color="red") +
  labs(title="Affairs Data",
       subtitle="Normal QQ-Plot"))
```




```
# The residuals are not normaly distributed, The qq plot is not straight and it is significantly deviating from t
he straight diagonal line. The p-value is also big.
```