HW5

Xinrui Hu

2023-03-10

```
library(faraway)
data = teengamb
```

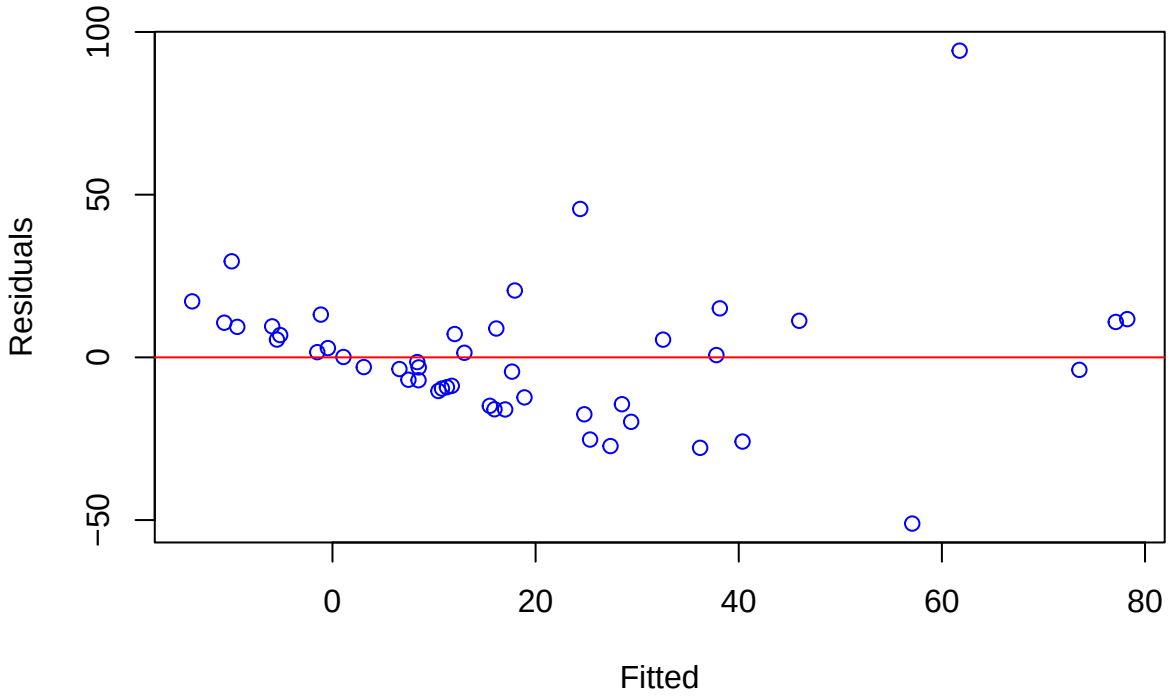
Question 1: Create a regression with gamble as the outcome variable and sex, status, income, and verbal as predictors.

```
model1 <- lm(gamble ~ sex + status + income + verbal, data)
summary(model1)
```

```
##
## Call:
## lm(formula = gamble ~ sex + status + income + verbal, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -51.082 -11.320  -1.451   9.452  94.252
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  22.55565   17.19680   1.312   0.1968
## sex         -22.11833    8.21111  -2.694   0.0101 *
## status        0.05223    0.28111   0.186   0.8535
## income        4.96198    1.02539   4.839 1.79e-05 ***
## verbal       -2.95949    2.17215  -1.362   0.1803
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 22.69 on 42 degrees of freedom
## Multiple R-squared:  0.5267, Adjusted R-squared:  0.4816
## F-statistic: 11.69 on 4 and 42 DF,  p-value: 1.815e-06
```

Question 2 Create a residuals vs. fitted plot. Do you think the variance is constant?

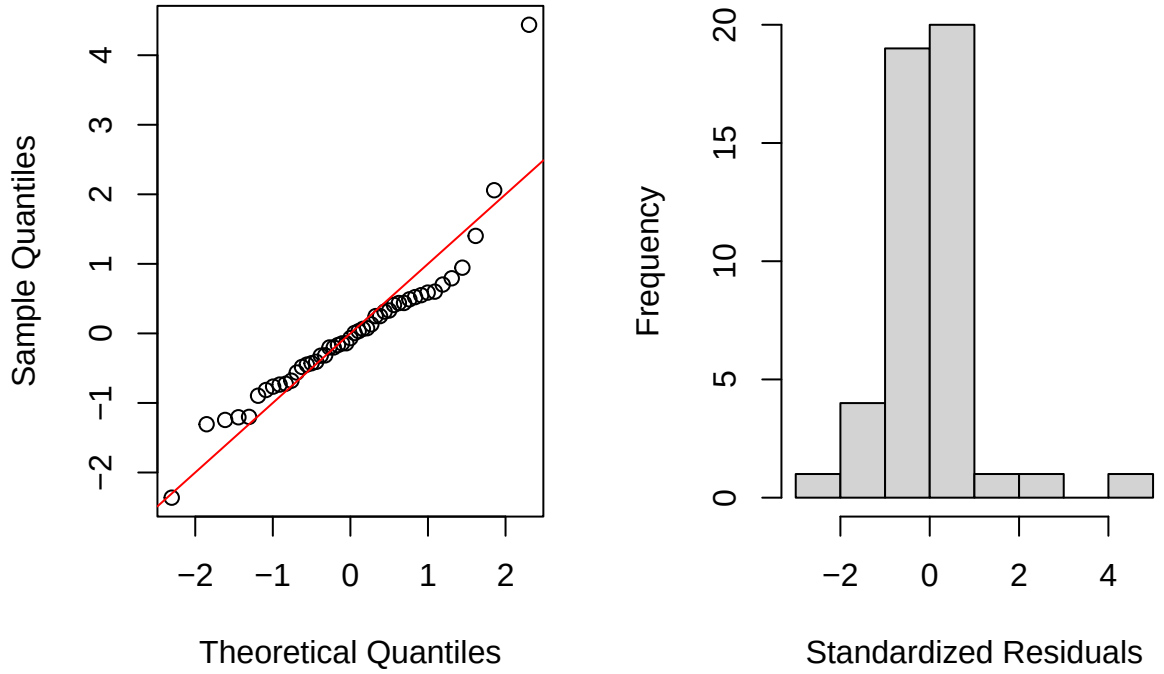
```
plot(fitted(model1), residuals(model1), xlab = "Fitted",ylab = "Residuals", col = "blue")
abline(h=0, col = "red")
```




```
# No the variance is not constant
```

Question 3 Create a Quantile-Quantile plot and a histogram based on the standardized residuals. Does the distribution of residuals look normal?

```
par (mfrow = c (1,2))
qqnorm(rstandard(model1), main = "")
abline(0,1, col = "red")
hist (rstandard(model1), main = "", xlab = "Standardized Residuals")
```



```
par (mfrow = c (1,1))
# The distribution of the residual is not normal
```

Question 4 Print all the standardized residuals. Are there any observations with standardized residuals greater than 3 or smaller than -3? If so, which ones? What are their standardized residuals?

```
rstandard(model1)
```

```
##            1            2            3            4            5            6
##  0.489347271  0.437414812  0.248790150 -0.814014679  1.401788808 -0.143846620
##            7            8            9           10           11           12
## -0.320827472 -0.562723398  0.313115783 -0.482362840  0.074066172 -0.141483007
##           13           14           15           16           17           18
##  0.005464581  0.436520105  0.129377367  0.792401904 -1.200579053 -1.306012119
##           19           20           21           22           23           24
##  0.600408769 -0.736030247 -0.723692228 -0.448798461 -1.244245565  4.437619557
##           25           26           27           28           29           30
##  0.032537476 -0.428799340 -1.205941923 -0.408580166 -0.315672285 -0.896386840
##           31           32           33           34           35           36
##  0.549812067  0.702266263  0.586657467 -0.167587874 -0.765095281  2.058415021
##           37           38           39           40           41           42
##  0.945193672  0.521605507 -2.362002273  0.409064868 -0.068051161 -0.202304946
##           43           44           45           46           47
## -0.202185610 -0.680286218  0.248743895  0.065005443  0.329071926
```

```
# There are one standardized residual greater than 3 but no standardized residual smaller than -3. The
```

Question 5 Identify points with leverages that are at least two times the average leverage. Did you find any points? If so, which points did you find?

```
hatv <- hatvalues(model1)
mean(hatv)
```

```
## [1] 0.106383
```

```
hatv [hatv > 2 * mean (hatv)]
```

```
##        31        33        35        42
## 0.2395031 0.2213439 0.3118029 0.3016088
```

```
# I found 4 points, the 31st, 33rd, 35th, and 42nd points.
```

Question 6 Detect outliers using studentized residuals. Use the Bonferroni correction. Which observations are detected, and what is their studentized residual?

```
stud <- rstudent(model1)
stud[which.max(abs(stud))]
```

```
##       24
## 6.016116
```

```
p <- 47 * 0.05
n <- 47
qt(1-.05 /(n*2),n-p-1)
```

```
## [1] 3.506683
```

```
which(abs(stud) > qt(1-.05/(n*2),n-p-1))
```

```
## 24
```

```
## 24
```

```
#Question 7
#Show the values of any outlier observation that you found in question 6
#The 24th observation detected and  the studentized residual is 6.016
```

Question 8 Use Cook's distances to search for influential points. Does any point have a Cook's distance above 0.5?

```
cook <- cooks.distance(model1)
cook[which(cook > 0.5)]
```

```
##        24
## 0.5565011
```

```
# The 24th point has the cook's distance larger than 0.5
```

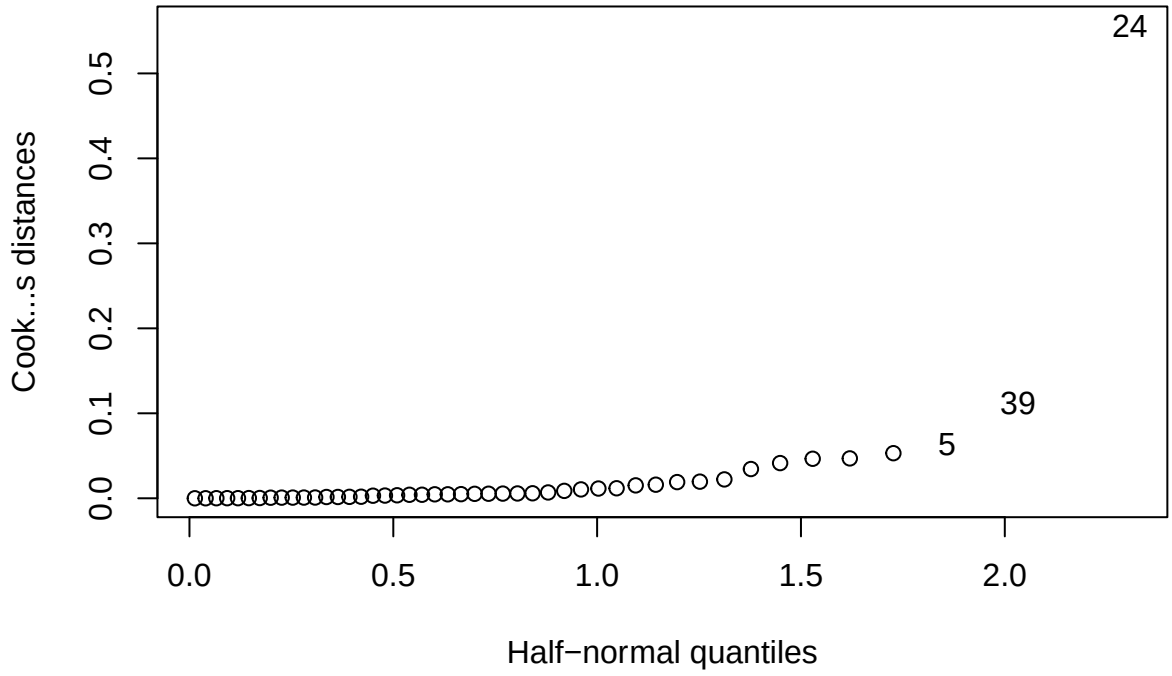
Question 9 Create a half normal plot. Which observation has the highest Cooks distance?

```
countries <- row.names(data)
halfnorm(cook,3, ylab = "Cook's distances")
```

```
## Warning in title(...): conversion failure on 'Cook's distances' in
## 'mbcsToSbcs': dot substituted for <e2>
```

```
## Warning in title(...): conversion failure on 'Cook's distances' in
## 'mbcsToSbcs': dot substituted for <80>
```

```
## Warning in title(...): conversion failure on 'Cook's distances' in
## 'mbcsToSbcs': dot substituted for <99>
```




```
# The 24th point has the highest distance
```