Homework 2

MacKenzie Ullman

2/15/2021

Calculate the residuals. Use the summary() function applied to your numeric vector of residuals to verify that this matches the Residuals: quantiles reported from the summary of your fitted model. Plot a histogram of your residuals. Do they appear reasonably Gaussian?

```
setwd("F:/Quant Eco")
getwd()
```

```
## [1] "F:/Quant Eco"
```

```
beach <- read.csv(file='beach.csv')
fit_i <- lm(OpenBeach ~ BeachID * Year, data = beach)
fit_i
```

```
##
## Call:
## lm(formula = OpenBeach ~ BeachID * Year, data = beach)
##
## Coefficients:
##   (Intercept)       BeachIDB       BeachIDC           Year  BeachIDB:Year
##      36945.73      -34211.98      -28776.11         -18.36          17.04
## BeachIDC:Year
##         14.37
```

```
res <- residuals.lm(fit_i)
res
```

```
##           1           2           3           4           5           6
##  74.98497092  79.14988575   7.01480058 -20.32028459   6.24463025 -50.82554009
##           7           8           9          10          11          12
## -10.86062526 -35.53079560 -20.76588076 -71.30096593  -5.27113627 -18.00622144
##          13          14          15          16          17          18
## -66.84130661 -29.17639177 -12.11147694   0.15343789  22.41835272  31.48326755
##          19          20          21          22          23          24
##  57.24818238  62.31309722  -8.23403540  23.18810936  15.91025413 -38.76760111
##          25          26          27          28          29          30
##  -9.94545634   4.39883319  -3.47902205  12.46526748   3.78741224 -29.69044299
##          31          32          33          34          35          36
##  -0.06829823   0.75384653  17.47599130  11.09813606  11.82028083  12.34242559
##          37          38          39          40          41          42
##   6.16457035  17.78671512 -21.69114012   1.63100465 -26.94685059  -8.57438940
##          43          44          45          46          47          48
##  70.81920301  32.51279541 -33.59361218  -2.80001978 -49.91283496  -7.31924256
##          49          50          51          52          53          54
## -32.73205775 -18.23846534 -40.74487294 -53.45128053 -11.95768812  76.13590428
##          55          56          57          58          59          60
```

```
##  89.52949669   6.12308909  84.21668150 -38.58972610  -3.19613369 -80.20254128
##           61           62
##  13.69105112   8.28464353
```

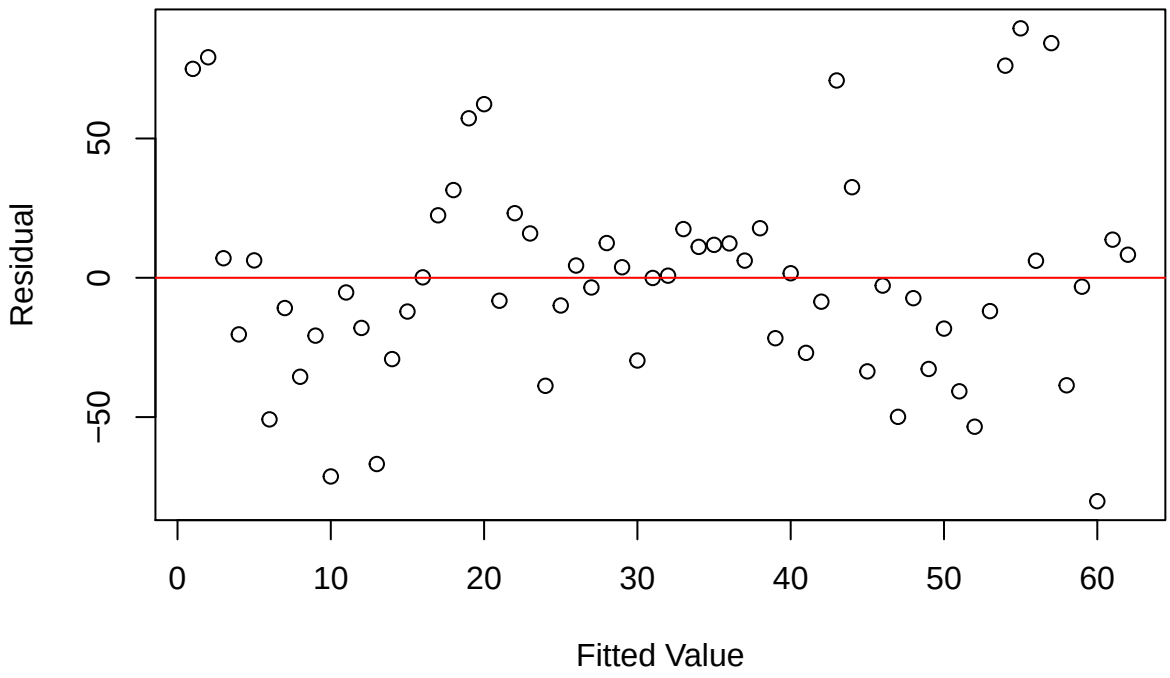
```
summary(res)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -80.203 -21.460  -1.434   0.000  13.385  89.529
```

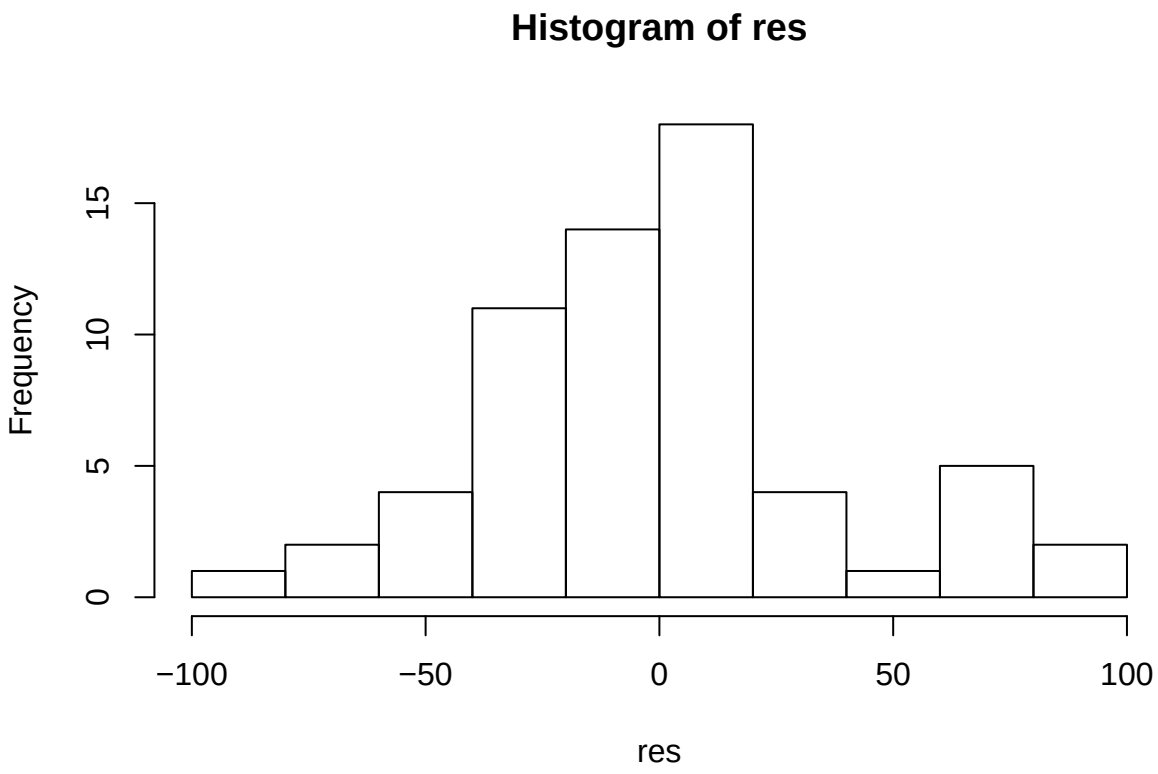
```
summary(fit_i)
```

```
##
## Call:
## lm(formula = OpenBeach ~ BeachID * Year, data = beach)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -80.203 -21.460  -1.434  13.385  89.529
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)    36945.725   2547.493  14.503  < 2e-16 ***
## BeachIDB      -34211.979   3602.404  -9.497 2.87e-13 ***
## BeachIDC      -28776.115   3602.404  -7.988 8.09e-11 ***
## Year             -18.365      1.272 -14.443  < 2e-16 ***
## BeachIDB:Year     17.043      1.798   9.478 3.07e-13 ***
## BeachIDC:Year     14.371      1.798   7.992 7.96e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 39.46 on 56 degrees of freedom
## Multiple R-squared:  0.8298, Adjusted R-squared:  0.8146
## F-statistic: 54.62 on 5 and 56 DF,  p-value: < 2.2e-16
```

```
plot(res, xlab= 'Fitted Value', ylab= 'Residual')
abline(h=0, col='red')
```




```
hist(res)
```



The numeric vector of residuals matche the Residuals: quantiles reported from the summary of the fitted model. The residuals appear reasonably Gaussian.

Calculate test statistics for your regression coefficients. Verify by comparing to test statistics reported from model output.

```
summary(fit_i)$coefficients
```

```
##                   Estimate  Std. Error    t value     Pr(>|t|)
## (Intercept)    36945.72537 2547.492995  14.502778 1.296196e-20
## BeachIDB      -34211.97897 3602.403682  -9.496986 2.868075e-13
## BeachIDC      -28776.11491 3602.403682  -7.988032 8.090204e-11
## Year             -18.36491    1.271577 -14.442627 1.558806e-20
## BeachIDB:Year     17.04277    1.798122   9.478096 3.074370e-13
## BeachIDC:Year     14.37132    1.798122   7.992408 7.957165e-11
```

```
b_1 <- coef(fit_i)[2]
s_b <- summary(fit_i)[['coefficients']]['BeachIDB', 'Std. Error']
B_1 <- 0
t1 <- (b_1 - B_1) / s_b
t1
```

```
##  BeachIDB
## -9.496986
```

```
b_2 <- coef(fit_i)[3]
s_b1 <- summary(fit_i)[['coefficients']]['BeachIDC', 'Std. Error']
B_2 <- 0
t2 <- (b_2 - B_2) / s_b1
```

```
t2
```

```
##  BeachIDC
## -7.988032
```

```
b_3 <- coef(fit_i)[4]
s_b2 <- summary(fit_i)[['coefficients']]['Year', 'Std. Error']
B_3 <- 0
t3 <- (b_3 - B_3) / s_b2
t3
```

```
##      Year
## -14.44263
```

```
b_4 <- coef(fit_i)[5]
s_b3 <- summary(fit_i)[['coefficients']]['BeachIDB:Year', 'Std. Error']
B_4 <- 0
t4 <- (b_4 - B_4) / s_b3
t4
```

```
## BeachIDB:Year
##      9.478096
```

```
b_5 <- coef(fit_i)[6]
s_b4 <- summary(fit_i)[['coefficients']]['BeachIDC:Year', 'Std. Error']
B_5 <- 0
t5 <- (b_5 - B_5) / s_b4
t5
```

```
## BeachIDC:Year
##      7.992408
```

Calculate p-values for your regression coefficients. Verify by comparing to p-values reported from modeloutput. What are the associated null hypotheses? Do you reject or fail to reject these null hypotheses?

```
p1<-pt(-1*abs(t1), df=56) + (1 - pt(abs(t1), df = 56))
p1
```

```
##     BeachIDB
## 2.868446e-13
```

```
p2<-pt(-1*abs(t2), df=56) + (1 - pt(abs(t2), df = 56))
p2
```

```
##     BeachIDC
## 8.090199e-11
```

```
p3<- 2*pt(-1*abs(t3), df=56) + (1 - pt(abs(t3), df = 56))
p3
```

```
##         Year
## 1.558806e-20
```

```
p4<-pt(-1*abs(t4), df=56) + (1 - pt(abs(t4), df = 56))
p4
```

```
## BeachIDB:Year
##  3.073734e-13
```

```
p5<-pt(-1*abs(t5), df=56) + (1 - pt(abs(t5), df = 56))
p5
```

```
## BeachIDC:Year
##  7.957167e-11
```

The null hypotheses associated with these p-values is that value of the slope coefficent(B_1,B_2,B_3,B_4,B_5) is equal to 0. Each p-value is less than 0.05 so we reject the null hypotheses.

Select a single regression coefficient (your choice) and devise a null hypothesis that is different from the default in lm(). Report the test statistics, your p-value, and whether you reject or fail to reject your null hypothesis.

H0:B_1=5 HA:B_1 > 5

```
ts_new <- (coef(fit_i)[2] - 5) /
summary(fit_i)[['coefficients']]['BeachIDB', 'Std. Error']
ts_new
```

```
##  BeachIDB
## -9.498374
```

```
pnew<-pt(-1*abs(ts_new), df=56) + (1 - pt(abs(ts_new), df = 56))
pnew
```

```
##     BeachIDB
## 2.853376e-13
```

The test statistic is -9.498374. The p-value is 2.853376e-13. We reject the null hypothesis.

Interpret output of your fitted model. Tell me how beach area does (or does not change) through time at each of the 3 beaches.

```
summary(fit_i)
```

```
##
## Call:
## lm(formula = OpenBeach ~ BeachID * Year, data = beach)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -80.203 -21.460  -1.434  13.385  89.529
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)    36945.725   2547.493  14.503  < 2e-16 ***
## BeachIDB      -34211.979   3602.404  -9.497 2.87e-13 ***
## BeachIDC      -28776.115   3602.404  -7.988 8.09e-11 ***
## Year             -18.365      1.272 -14.443  < 2e-16 ***
## BeachIDB:Year     17.043      1.798   9.478 3.07e-13 ***
## BeachIDC:Year     14.371      1.798   7.992 7.96e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 39.46 on 56 degrees of freedom
## Multiple R-squared:  0.8298, Adjusted R-squared:  0.8146
## F-statistic: 54.62 on 5 and 56 DF,  p-value: < 2.2e-16
```

The F-statistic p-value is $<$ 2.2e-16, which is highly significant.This means that, at least, one of the predictor variables is significantly related to the outcome variable.The predictor variables that are significant, according to the coefficients table, have t-statitic p-values less than 0.05 or have a ***. The beach area at beach A decreases by 18.36491 hectares per year. The beach area at beach B decreases by 1.32214 hectaeres per year. The beach area at beach C decreases by 3.99359 hectares per year.