Homework 2

Igor Kuivjogi Fernandes

2023-02-02

1. An ANOVA output is shown below. Fill in the missing information.
   One-way ANOVA

Source	DF	SS	MS	F	P
Factor	3	36.15	?	?	?
Error	?	?	?		
Total	19	196.04			

Completing the cells we got:

Source	DF	SS	MS	F	P
Factor	3	36.15	12.05	1.213418	0.3369274
Error	16	158.89	9.930625		
Total	19	196.04			

The p-value is the $P(F > 1.123418) = 0.3369274$, with $\text{df}_{\text{factor}} = 3$ and $\text{df}_{\text{error}} = 16$.

2. I belong to a golf club in my neighborhood. I divide the year into three golf seasons: summer (June–September), winter (November–March), and shoulder (October, April, and May). I believe that I play my best golf during the summer (because I have more time and the course isn't crowded) and shoulder (because the course isn't crowded) seasons and my worst golf is during the winter (because when all of the part-year residents show up, the course is crowded, play is slow, and I get frustrated). Data from the last year are shown in the following table.

We can write a hypothesis test as follows:

$$
\begin{aligned}
H_0 &: \text{the golf performance in the seasons are equal} \\
H_1 &: \text{at least one golf performance differs}
\end{aligned}
$$

```r
df <- data.frame(
  season = c(rep('summer', 10), rep('shoulder', 7), rep('winter', 8)),
  y = c(83, 85, 85, 87, 90, 88, 88, 84, 91, 90, 91, 87, 84,
        87, 85, 86, 83, 94, 91, 87, 85, 87, 91, 92, 86)
)

# frequency table
cat('summer:', df[df$season == 'summer', 'y'], '\n',
    'shoulder:', df[df$season == 'shoulder', 'y'], '\n',
    'winter:', df[df$season == 'winter', 'y'], '\n\n')
```

```
## summer: 83 85 85 87 90 88 88 84 91 90
##  shoulder: 91 87 84 87 85 86 83
##  winter: 94 91 87 85 87 91 92 86
```

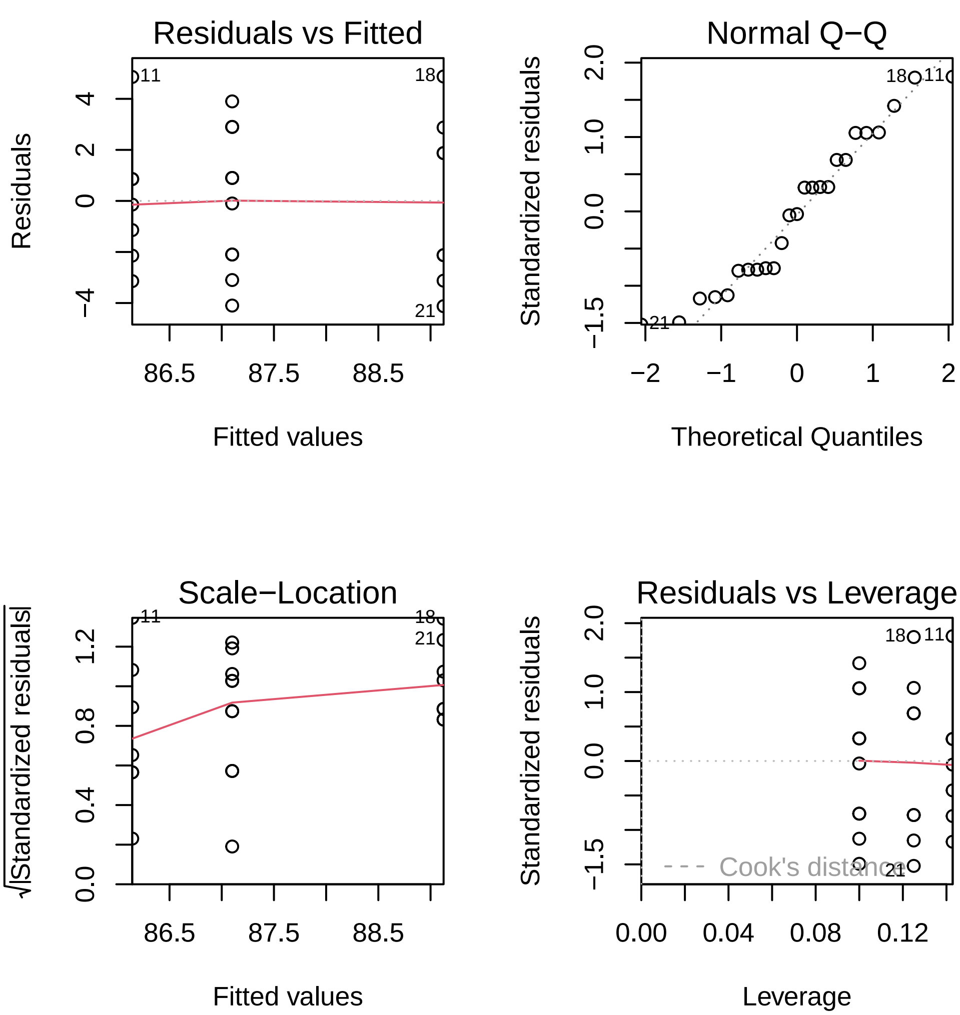
```r
# one-way ANOVA
fit <- aov(y ~ season, data = df)
summary(fit)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## season       2  35.61  17.804   2.121  0.144
## Residuals   22 184.63   8.392
```

a) Do the data indicate that my opinion is correct? Use alpha 0.05.
   No, we don't reject the null hypothesis that the golf performance is equal, because p-value $= 0.144 > \alpha$, which means that the golf performance was the same for all the seasons.

b) Analyze the residuals from this experiment and comment on model adequacy.

```r
par(mfrow = c(2, 2))
plot(fit, xaxs='i', yaxs='i')
```



The Residual vs Fitted plot shows homoscedasticity, i.e., the variance is constant along the x-axis, and are equally distributed around zero.
The Normal Q-Q plot shows that the residuals are normally distributed because the points are very aligned with the Q-Q line.
The Scale-Location plot presents more variance in the lower portion of the fitted values, whereas the larger portion shows a lower variance. The variance is decreasing as the fitted values increase.
The Residuals vs Leverage does not show any points with big Cook's distance, therefore there are no influential points in the model.
In general, although the Scale-Location shows a slightly non horizontal pattern, the residual plots represent a good model adequacy.

3. An article in Environment International (Vol. 18, No. 4, 1992) describes an experiment in which the amount of radon released in showers was investigated. Radon-enriched water was used in the experiment, and six different orifice diameters were tested in shower heads. The data from the experiment are shown in the following table: