Assignment_7

Oluwanifemi

2024-11-03

```r
library(faraway)
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.4.2
```

```r
library(alr4)
```

```
## Loading required package: car

## Loading required package: carData

##
## Attaching package: 'car'

## The following objects are masked from 'package:faraway':
##
##     logit, vif

## Loading required package: effects

## lattice theme set by effectsTheme()
## See ?effectsTheme for details.

##
## Attaching package: 'alr4'

## The following objects are masked from 'package:faraway':
##
##     cathedral, pipeline, twins
```

```r
library(ggpubr)
```

```
## Warning: package 'ggpubr' was built under R version 4.4.2
```

```r
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.4.2
```

```
## Warning: package 'forcats' was built under R version 4.4.2

## Warning: package 'lubridate' was built under R version 4.4.2

## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v lubridate 1.9.3     v tibble    3.2.1
## v purrr     1.0.2     v tidyr     1.3.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## x dplyr::recode() masks car::recode()
## x purrr::some()   masks car::some()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become error
```

```
library(broom)
```

```
## Warning: package 'broom' was built under R version 4.4.2
```

```
library(AICcmodavg)
```

```
## Warning: package 'AICcmodavg' was built under R version 4.4.2
```

```
View(fuel2001)
```

Question 2(c)

```
fuel2001$Fuel <- fuel2001$FuelC/fuel2001$Pop
fuel2001$Dlic <- fuel2001$Drivers/fuel2001$Pop
ll<-lm(formula = Fuel ~ Tax + Dlic + Income + I(log(Miles)), data = fuel2001)
summary(ll)
```

```
##
## Call:
## lm(formula = Fuel ~ Tax + Dlic + Income + I(log(Miles)), data = fuel2001)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.163145 -0.033039  0.005895  0.031989  0.183499
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.542e-01  1.949e-01   0.791 0.432938
## Tax           -4.228e-03  2.030e-03  -2.083 0.042873 *
## Dlic           4.719e-01  1.285e-01   3.672 0.000626 ***
## Income        -6.135e-06  2.194e-06  -2.797 0.007508 **
## I(log(Miles))  2.676e-02  9.337e-03   2.865 0.006259 **
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.06489 on 46 degrees of freedom
## Multiple R-squared:  0.5105, Adjusted R-squared:  0.4679
## F-statistic: 11.99 on 4 and 46 DF,  p-value: 9.331e-07
```

```r
anova_table <- anova(ll)
print(anova_table)
```

```
## Analysis of Variance Table
##
## Response: Fuel
##                 Df   Sum Sq  Mean Sq F value    Pr(>F)
## Tax              1 0.026635 0.026635  6.3254 0.0154602 *
## Dlic             1 0.079378 0.079378 18.8506 7.692e-05 ***
## Income           1 0.061408 0.061408 14.5833 0.0003997 ***
## I(log(Miles))    1 0.034573 0.034573  8.2104 0.0062592 **
## Residuals       46 0.193700 0.004211
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
anova_model <- aov(Fuel ~ Tax + Dlic + Income + I(log(Miles)), data = fuel2001)
print(anova_model)
```

```
## Call:
##    aov(formula = Fuel ~ Tax + Dlic + Income + I(log(Miles)), data = fuel2001)
##
## Terms:
##                        Tax       Dlic     Income I(log(Miles))  Residuals
## Sum of Squares  0.02663528 0.07937753 0.06140818    0.03457307 0.19370002
## Deg. of Freedom          1          1          1             1         46
##
## Residual standard error: 0.06489122
## Estimated effects may be unbalanced
```

```r
summary(anova_model)
```

```
##                 Df  Sum Sq Mean Sq F value   Pr(>F)
## Tax              1 0.02664 0.02664   6.325  0.01546 *
## Dlic             1 0.07938 0.07938  18.851 7.69e-05 ***
## Income           1 0.06141 0.06141  14.583  0.00040 ***
## I(log(Miles))    1 0.03457 0.03457   8.210  0.00626 **
## Residuals       46 0.19370 0.00421
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Explanation of ANOVA Table Components

- **RSS** represents unexplained variance, showing how much variability in *Fuel* is not explained by the predictors.

- **Res.Df** tells us the remaining degrees of freedom after fitting the predictors.
- The **F-statistic of 11.99** (from the full model) suggests the overall model explains the response variable variance significantly better than a model with no predictors. The RSS in the aov table is 0.193700, which is about the same as my calculated RSS which was 0.1936784. The residual DF is 46, whichis the same as what I has calculated which was n-p (50-4) which is 46. The F-statistic is 11.9, the same as my calculation

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

QUESTION 3

Given Information

- We know that $y \sim N_n(X\beta, \sigma^2 I_n)$, where:
  - y is an n-dimensional vector of observed values,
  - X is an $n \times (p+1)$ matrix of predictors,
  - β is a $(p+1)$-dimensional vector of coefficients,
  - $\sigma^2 I_n$ is the covariance matrix of y.
- The projection (or "hat") matrix $H = X(X'X)^{-1}X'$, is symmetric and idempotent:
  - **Symmetric**: $H = H'$,
  - **Idempotent**: $H^2 = H$.
- The residual vector $\hat{e} = y - \hat{y}$, where $\hat{y} = Hy$ represents the fitted values:

$$\hat{e} = y - \hat{y} = y - Hy = (I - H)y$$

Part (a): Show that $\hat{e} \sim N_n(0_n, \sigma^2(I - H))$

The residual vector $\hat{e}$ is given by:

$$\hat{e} = (I - H)y$$

Since $y \sim N_n(X\beta, \sigma^2 I_n)$, we can write $y = X\beta + \epsilon$, where $\epsilon \sim N_n(0_n, \sigma^2 I_n)$. Therefore:

$$\hat{e} = (I - H)(X\beta + \epsilon) = (I - H)X\beta + (I - H)\epsilon$$

Taking the expectation of $\hat{e}$:

$$\mathbb{E}(\hat{e}) = (I - H)X\beta + (I - H)\mathbb{E}(\epsilon)$$

Since ϵ has mean zero, $\mathbb{E}(\epsilon) = 0_n$, this reduces to:

$$\mathbb{E}(\hat{e}) = (I - H)X\beta$$

Note: H is a projection matrix onto the column space of X, so $HX\beta = X\beta$. This implies:

$$(I - H)X\beta = X\beta - HX\beta = X\beta - X\beta = 0_n$$

Thus, $\mathbb{E}(\hat{e}) = 0_n$.

To determine the covariance matrix of $\hat{e}$, we calculate $\mathrm{Var}(\hat{e})$:

$$\mathrm{Var}(\hat{e}) = \mathrm{Var}((I - H)y) = (I - H)\mathrm{Var}(y)(I - H)'$$

Since $\mathrm{Var}(y) = \sigma^2 I_n$, we get:

$$\mathrm{Var}(\hat{e}) = (I - H)(\sigma^2 I_n)(I - H)' = \sigma^2 (I - H)(I - H)'$$

Because $I - H$ is symmetric and idempotent, $(I - H)' = I - H$ and $(I - H)^2 = I - H$, so:

$$\mathrm{Var}(\hat{e}) = \sigma^2 (I - H)$$

Conclusion for Part (a)

$\hat{e}$ has a mean of 0_n and covariance matrix $\sigma^2(I - H)$, so:

$$\hat{e} \sim N_n(0_n, \sigma^2(I - H))$$

Part (b): Show that $\hat{\beta} \sim N_{p+1}(\beta, \sigma^2(X'X)^{-1})$

The estimator $\hat{\beta}$ for β is given by:

$$\hat{\beta} = (X'X)^{-1}X'y$$

Substitute $y = X\beta + \epsilon$ into the expression for $\hat{\beta}$:

$$\hat{\beta} = (X'X)^{-1}X'(X\beta + \epsilon)$$

Distribute X':

$$\hat{\beta} = (X'X)^{-1}X'X\beta + (X'X)^{-1}X'\epsilon$$

Since $(X'X)^{-1}X'X = I$, we have:

$$\hat{\beta} = \beta + (X'X)^{-1}X'\epsilon$$

Since ϵ is mean-zero, the expectation of $\hat{\beta}$ is:

$$\mathbb{E}(\hat{\beta}) = \mathbb{E}(\beta + (X'X)^{-1}X'\epsilon) = \beta + (X'X)^{-1}X'\mathbb{E}(\epsilon) = \beta$$

Thus, $\hat{\beta}$ is an unbiased estimator of β.

To determine the covariance matrix of $\hat{\beta}$, we calculate $\mathrm{Var}(\hat{\beta})$:

$$\mathrm{Var}(\hat{\beta}) = \mathrm{Var}(\beta + (X'X)^{-1}X'\epsilon)$$

Since β is a constant and does not contribute to the variance, this simplifies to:

$$\mathrm{Var}(\hat{\beta}) = (X'X)^{-1}X' \,\mathrm{Var}(\epsilon)\, X(X'X)^{-1}$$

Since $\mathrm{Var}(\epsilon) = \sigma^2 I_n$, we have:

$$\mathrm{Var}(\hat{\beta}) = (X'X)^{-1}X'(\sigma^2 I_n)X(X'X)^{-1} = \sigma^2(X'X)^{-1}$$

Conclusion for Part (b)

$\hat{\beta}$ has a mean of β and covariance matrix $\sigma^2(X'X)^{-1}$. Thus:

$$\hat{\beta} \sim N_{p+1}(\beta, \sigma^2(X'X)^{-1})$$