STATS 500 HW2

Minxuan Chen

2023-09-17

Table of contents

Github repo: https://github.com/PKUniiiiice/STATS_500

1 Problem 1

1.1 (a)

```
library(faraway)
data(teengamb)
m1 <- lm(gamble ~., data=teengamb)
summary(m1)
```

```
Call:
lm(formula = gamble ~ ., data = teengamb)

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
```

Multiple R-squared gives the proportion of variation in the response variable, which can be explained by all the explanatory variables.

So, the percentage is 52.67%.

1.2 (b)

```
cat("Largest positive residual: Case Number", which.max(m1$residuals), "\n")
```

```
Largest positive residual: Case Number 24
```

```r
cat("Largest negative residual: Case Number", which.max(-m1$residuals), "\n")
```

```
Largest negative residual: Case Number 39
```

1.3 (c)

```r
cat("Mean of residuals: ", mean(m1$residuals), "\n")
```

```
Mean of residuals:  -1.556914e-16
```

```r
cat("Median of residuals: ", median(m1$residuals))
```

```
Median of residuals:  -1.451392
```

The mean value is slightly larger than the median, indicating a mild right skew in the residuals. This suggests a slight deviation from our assumption that the errors, $\epsilon_i \sim N(0, 1)$.

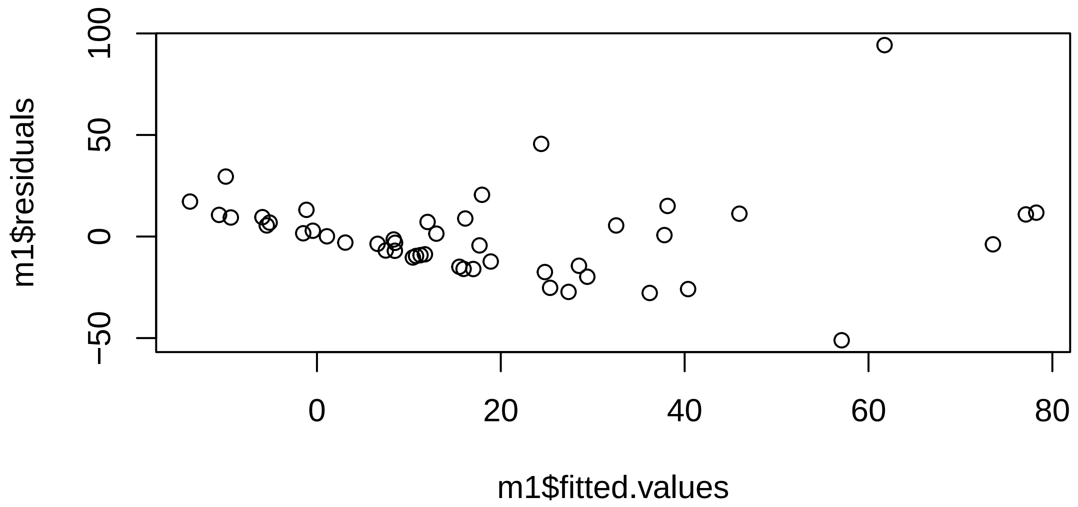
BTW, if we are confident that the assumption is correct, the small bias in the median may be attributed to the limited sample size. With a larger sample, we anticipate the median approaching zero.

1.4 (d)

```r
cat("Corr(res, fitted): ", cor(m1$residuals, m1$fitted.values))
```

```
Corr(res, fitted):  -6.215823e-17
```

```r
plot(x=m1$fitted.values, y=m1$residuals)
```



1.5 (e)

```r
for (i in c("income", "status", "verbal")){
  cat("Corr(res, ", i, "): ", cor(m1$residuals, teengamb[,i]), "\n")
}
```

```
Corr(res,  income ):  3.247058e-17
Corr(res,  status ):  1.542372e-17
Corr(res,  verbal ):  -2.559375e-16
```

We can observe that the correlations between the residuals and these predictors are very small, almost zero. This implies that the residuals and these predictors are uncorrelated.

It's worth to mention that the term "uncorrelated" typically applies to random variables. In our assumption within the classical linear regression model, X is not considered stochastic. Therefore, the result of $\mathrm{Corr}(X, \epsilon) = 0$ is a natural consequence.

However, when we say 'uncorrelated' in this context, we are referring to a property at the sample level. Specifically, it signifies that the columns of the design matrix X are orthogonal to the residual vector $\hat{\epsilon}$. This is a natural outcome due to the hat matrix H being a projection matrix.

A formal proof of this property can be provided.

$$\hat{\epsilon} = y - \hat{y} = (I - X(X^TX)^{-1}X^T)y, \rightarrow X^T\hat{\epsilon} = (X^T - X^T)y = 0$$

We consider the i-th row of X^T, which represents all values of the predictor X_i in the sample, then

$$\sum_{j=1}^{n} X_{ij}\hat{\epsilon}_j = 0$$

and the first row of X^T,

$$\sum_{j=1}^{n} 1 \cdot \hat{\epsilon}_j = 0$$

Therefore, the numerator of sample correlation between X_i and $\hat{\epsilon}$ is

$$n \sum_{j=1}^{n} X_{ij}\hat{\epsilon}_j - \sum_{j=1}^{n} X_{ij} \sum_{j=1}^{n} \hat{\epsilon}_j = 0$$

Hence, X_i and $\hat{\epsilon}$ are uncorrelated.

1.6 (f)

The difference is exactly equal to the coefficient of variable `sex`. Since 0 is male and 1 is female, the difference would be -22.11833 (female - male), i.e. female spends 22.11833 pounds less than male.

2 Problem 2

2.1 (a)

We check whether

$$[Cov(AU, BV)]_{ij} = \left[ACov(U, V)B^\top\right]_{ij}$$

$$\begin{aligned}
LHS &= Cov((AU)_i, (BV)_j) \\
&= Cov\left(\sum_{t=1}^{m} A_{it}U_t, \sum_{s=1}^{n} B_{js}V_s\right) \\
&= \sum_{t=1}^{m}\sum_{s=1}^{n} A_{it}B_{js}Cov(U_t, V_s)
\end{aligned}$$

$$\begin{aligned}
RHS &= \sum_{s=1}^{n} [ACov(U, V)]_{is} \, (B^\top)_{sj} \\
&= \sum_{s=1}^{n} \left[\sum_{t=1}^{m} A_{it}Cov(U, V)_{ts}\right] B_{js} \quad ((B^\top)_{sj} = B_{js}) \\
&= \sum_{s=1}^{n}\sum_{t=1}^{m} A_{it} \; B_{js}Cov(U_t, V_s) \\
&= \sum_{t=1}^{m}\sum_{s=1}^{n} A_{it}B_{js}Cov(U_t, V_s) = LHS
\end{aligned}$$

Therefore

$$Cov(AU, BV) = ACov(U, V)B^\top$$

2.2 (b)

It's known that

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

and

$$\hat{\epsilon} = (I - X(X^\top X)^{-1}X^\top)y = X\beta + (I - X(X^\top X)^{-1}X^\top)\epsilon$$

Note that in these two expressions, only ϵ is random. Thus

$$\begin{aligned}
&Cov(\hat{\beta}, \hat{\epsilon}) \\
=&Cov(\beta + (X^\top X)^{-1}X^\top\epsilon, X\beta + (I - X(X^\top X)^{-1}X^\top)\epsilon) \\
=&Cov((X^\top X)^{-1}X^\top\epsilon, (I - X(X^\top X)^{-1}X^\top)\epsilon) \\
=&(X^\top X)^{-1}X^\top Cov(\epsilon, \epsilon)\left[(I - X(X^\top X)^{-1}X^\top)\right]^\top \\
=&\sigma^2 I_n (X^\top X)^{-1}X^\top(I - X(X^\top X)^{-1}X^\top) \\
=&\sigma^2((X^\top X)^{-1}X^\top - (X^\top X)^{-1}X^\top)) \\
=&0
\end{aligned}$$

since $\hat{\beta}$ is of dim $(p+1) \times 1$, and $\hat{\epsilon}$ is of dim $n \times 1$, $Cov(\hat{\beta}, \hat{\epsilon})$ is a $(p+1) \times n$ matrix of zeros.