

STATS 500 HW2

Minxuan Chen

2023-09-17

Table of contents

1	Problem 1	1
1.1	(a)	1
1.2	(b)	1
1.3	(c)	2
1.4	(d)	2
1.5	(e)	3
1.6	(f)	4
2	Problem 2	4
2.1	(a)	4
2.2	(b)	5

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

1 Problem 1

1.1 (a)

```
1 library(faraway)
2 data(teengamb)
3 m1 <- lm(gamble ~., data=teengamb)
4 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)

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

Largest positive residual: Case Number 24

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

Largest negative residual: Case Number 39

1.3 (c)

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

Mean of residuals: -1.556914e-16

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

Median of residuals: -1.451392

Note that mean value is larger than median but not very severe, so we conclude that the residuals is slightly skewed to the right. In other words, our assumption of the error, $\sigma_i \sim N(0, 1)$ is slightly violated.

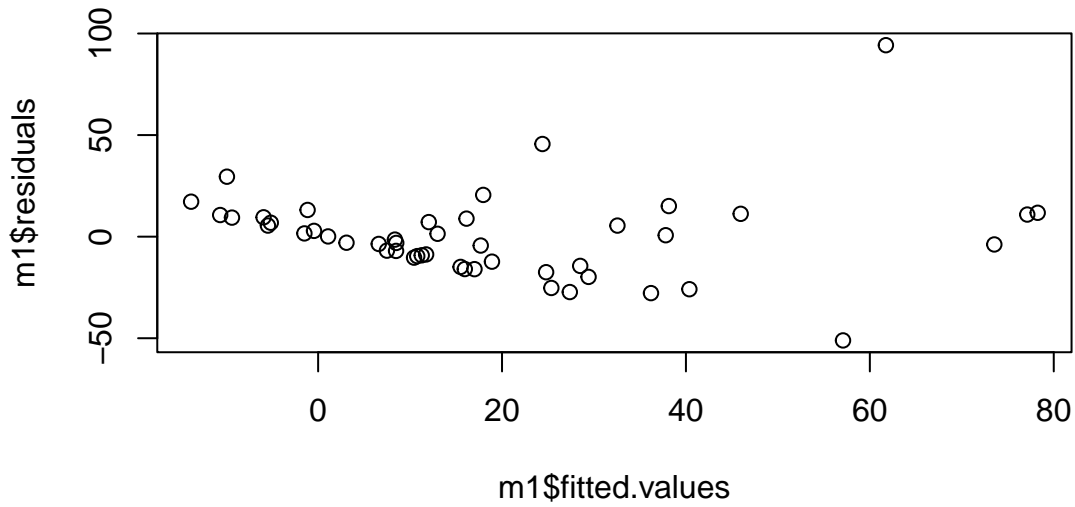
BTW, if we're sure that the assumption is right, then the small bias of median is due to limited sample size. When we get more samples, we expect to see a zero median as well.

1.4 (d)

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

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

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



1.5 (e)

```

1 for (i in c("income", "status", "verbal")){
2   cat("Corr(res, ", i, "): ", cor(m1$residuals, teengamb[,i]), "\n")
3 }

```

```

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

```

We observe that the correlations between the residuals and these predictors are so small that we can regard them as 0. In other words, residuals and these predictors are uncorrelated.

To correspond this result to theory, it's worthy to mention that “uncorrelated” is a term of random variable. In our assumption of classical linear regression model, X is not stochastic, so $\text{Corr}(X, \epsilon) = 0$ is a natural result. However, here we say “uncorrelated” points at sample level. It more means columns of design matrix X is orthogonal to residual vector $\hat{\epsilon}$. This is the natural result of hat matrix H is a projection.

We can give a proof.

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

We consider the i -th row of X^T , which denotes all values of predictor X_i , 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 of 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 the meaning of 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

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

Therefore

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

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

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

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.