Giang Vu - 4445745
STAT 2131
Oct 8, 2020

Homework 7

1) a) • $(v - Hv)^T (Hv) = 0$

We have $(v - Hv)^T (Hv)$

$$= v^T H v - v^T H^T H v$$

$$= v^T H v - v^T H H v \quad (\text{as } H^T = H)$$

$$= v^T H v - v^T H v \quad (\text{as } HH = H^2 = H)$$

$$= 0$$

• Prove $Hv = \underset{u \in Im(x)}{\operatorname{argmin}} \| v - u \|_2^2$

Scanned with CamScanner

ii) $$\|v-u\|_2^2 = \|v - Hv + Hv - u\|_2^2$$
$$= \|(v-Hv) + (Hv-u)\|_2^2$$
$$= [(v-Hv)+(Hv-u)]^T[(v-Hv)+(Hv-u)]$$
$$= \underbrace{(v-Hv)^T(v-Hv)}_{(1)} + \underbrace{(v-Hv)^T(Hv-u)}_{(2)} + \underbrace{(Hv-u)^T(v-Hv)}_{(3)} + \underbrace{(Hv-u)^T(Hv-u)}_{(4)}$$

Let's look at $(1) = (v-Hv)^T(v-Hv) = \|v-Hv\|_2^2$

$(4) = (Hv-u)^T(Hv-u) = \|Hv-u\|_2^2$

$$(2) = (v-Hv)^T(Hv-u)$$
$$= v^THv - v^Tu - v^TH^THv + v^TH^Tu$$
$$= v^THv - v^Tu - v^THv + v^Tu$$
(because $H^TH = HH = H$
and $H^Tu = Hu = u$)

$\Rightarrow (2) = (v^THv - v^THv) + (v^Tu - v^Tu) = 0$

$$(3) = (Hv-u)^T(v-Hv)$$
$$= v^TH^Tv - v^THHv - u^Tv + u^THv$$
$$= v^THv - v^THv - u^Tv + u^Tv$$
(because $H = H^T = HH$
and $u^TH = (Hu)^T = u^T$)
$$= 0$$

$\Rightarrow \|v-u\|_2^2 = (1) + (2) + (3) + (4)$
$$= \|v-Hv\|_2^2 + 0 + 0 + \|Hv-u\|_2^2$$
$$= \|v-Hv\|_2^2 + \|Hv-u\|_2^2$$
which is $\geq \|v-Hv\|_2^2$

Therefore, $u = Hv = \underset{u\in Im(X)}{\arg\min} \|v-u\|_2^2$

b) We have H & P are both orthogonal projection matrices onto $X_{n \times p}$

Suppose we have $u \in Im(X)$

$\Rightarrow Hu = u$ and $Pu = u$

$\Rightarrow Hu = Pu \Rightarrow H = P$

$\Rightarrow H$ is unique

c) 4 properties : (1) $XX^+X = X$ (3) $(XX^+)^T = XX^+$

(2) $X^+XX^+ = X^+$ (4) $(X^+X)^T = X^+X$

Scanned with CamScanner

We have $H = XX^+$

- Symmetry

$$H^T = (XX^+)^T = XX^+ = H \quad \text{(using property (3))}$$

- Idempotence

$$HH = XX^+XX^+ = (XX^+X)X^+ = XX^+ = H$$

(using property (1))

$$\Rightarrow H^2 = H$$

- $im(H) = im(X)$

Let $a \in Im(H)$, $a \in \mathbb{R}^n$

$$\Rightarrow a = \underset{n\times n}{H}\underset{n\times 1}{v_1} = XX^+ v_1 = X\underbrace{(\underset{p\times n}{X^+}\underset{n\times 1}{v_1})}_{\text{a } p\times 1 \text{ vecter}}$$

$$\Rightarrow a \in Im(X)$$

$$\Rightarrow Im(H) \subset Im(X)$$

Let $b \in Im(X)$ $\Rightarrow$ $\underset{n\times 1}{b} = \underset{n\times p}{X}\underset{p\times 1}{v_2} = (\underset{n\times p}{X}\underset{p\times n}{X^+}\underset{n\times p}{X})\underset{p\times 1}{v_2}$

$b \in \mathbb{R}^n$

$$= XX^+(\underset{n\times p}{X}\underset{p\times 1}{v_2})$$

$$= H(\underbrace{Xv_2}_{\text{a } n\times 1 \text{ vecter}})$$

$$\Rightarrow b \in Im(H)$$

$$\Rightarrow Im(X) \subset Im(H)$$

$$\Rightarrow Im(H) = Im(X)$$

Scanned with CamScanner

2a) $\hat{\beta}_{OLS} = (X^T X)^{-1} X^T y$

$$Var(\hat{\beta}_{OLS}) = Var((X^T X)^{-1} X^T y)$$

$$= (X^T X)^{-1} X^T \, Var(y) \, X (X^T X)^{-1}$$

$$= (X^T X)^{-1} X^T \sigma^2 I \, X (X^T X)^{-1}$$

$$= \sigma^2 (X^T X)^{-1} X^T X (X^T X)^{-1}$$

$$= \sigma^2 (X^T X)^{-1}$$

$H = X(X^T X)^{-1} X^T$

Scanned with CamScanner

b) • Prove $A^T X = I_p$

Because $\tilde{\beta}$ is a linear unbiased estimator for $\beta \Rightarrow E(\tilde{\beta}) = \beta$

$\Rightarrow E(A^T y) = \beta \Rightarrow E(A^T X \beta) = \beta \Rightarrow A^T X \beta = \beta$

$\Rightarrow A^T X = I_p$

Scanned with CamScanner

- $Var(\tilde{\beta}) = Var(A^T y)$

$= A^T \, Var(y) \, A$

$= A^T \sigma^2 I A = \sigma^2 A^T A$

Scanned with CamScanner

c) $Var(\tilde{\beta}) = \sigma^2 A^T A$ (from part b)

$= \sigma^2 A^T I A = \sigma^2 A^T (H + I - H) A$

$= \sigma^2 A^T H A + \sigma^2 A^T (I - H) A$

$= \sigma^2 A^T X (X^T X)^{-1} X^T A + \sigma^2 A^T (I - H) A$

$= \sigma^2 \cdot I \cdot (X^T X)^{-1} \cdot I + \sigma^2 A^T (I - H) A$

$= \sigma^2 (X^T X)^{-1} + \sigma^2 A^T (I - H) A$

$= Var(\hat{\beta}_{OLS}) + \sigma^2 A^T (I - H) A$

We can see that $A^T (I - H) A$ is a quadratic form, and $I - H = Q$ is an ~~pro~~ orthogonal projection, hence it is symmetric

From HW5, we showed that all the eigenvalues of Q (or $I - H$) are either 0 or 1, so the eigenvalues of $(I - H)$ are all ≥ 0

$\Rightarrow \sigma^2 A^T (I - H) A$ is a positive semidefinite matrix, we call this M

$\Rightarrow Var(\tilde{\beta}) = Var(\hat{\beta}_{OLS}) + M$

d) $Var(q^T \tilde{\beta}) = q^T Var(\tilde{\beta}) (q^T)^T$

$= q^T (Var \hat{\beta}_{OLS} + M) q$

$= q^T Var(\hat{\beta}_{OLS}) q + q^T M q$

$= Var(q^T \hat{\beta}_{OLS}) + \underbrace{q^T M q}_{\geq 0}$ (from part c)

$\Rightarrow Var(q^T \tilde{\beta}) \geq Var(q^T \hat{\beta}_{OLS})$

Scanned with CamScanner

because

e) i) $E(\tilde{Y}) = E(\tilde{X}\beta + \tilde{\epsilon})$

We have $\tilde{Y} = \tilde{X}\beta + \tilde{\epsilon}$

$\Rightarrow R^{-1}Y = R^{-1}X\beta + \tilde{\epsilon} = R^{-1}(X\beta + \frac{1}{R^{-1}}\tilde{\epsilon})$

$\Rightarrow Y = X\beta + \frac{1}{R^{-1}}\tilde{\epsilon}$

But $Y = X\beta + \epsilon \Rightarrow \tilde{\epsilon} = R^{-1}\epsilon$

$\Rightarrow E(\tilde{Y}) = E(R^{-1}X\beta + R^{-1}\epsilon)$

$= R^{-1}E(X\beta) + R^{-1}\underbrace{E(\epsilon)}_{0}$

$= R^{-1}X\beta$

$Var(\tilde{Y}) = Var(R^{-1}Y)$

$= R^{-1}\,Var(Y)\,(R^{-1})^T$

$= R^{-1}\sigma^2\Sigma(R^T)^{-1}$

$= \sigma^2(R^{-1}R)(R^T(R^T)^{-1})$ (as $\Sigma = RR^T$)

$= \sigma^2$

ii) $\Sigma = RR^T \Rightarrow \Sigma^{-1} = (R^T)^{-1}R^{-1}$

$\Sigma^T = (R^T)^T R^T = RR^T \Rightarrow (\Sigma^T)^{-1} = (\Sigma^{-1})^T = (R^T)^{-1}R^{-1}$

$\hat{\beta}_{GLS}$ mimizes the function $f(\beta) = (\tilde{Y} - \tilde{X}\beta)^T(\tilde{Y} - \tilde{X}\beta)$

Set $\nabla_\beta f(\beta) = 0$ and solve for $\hat{\beta}_{GLS}$

We have $f(\beta) = (R^{-1}(Y - X\beta))^T(R^{-1}(Y - X\beta))$

$= (Y - X\beta)^T(R^T)^{-1}R^{-1}(Y - X\beta)$

$= (Y - X\beta)^T\Sigma^{-1}(Y - X\beta)$

$\Rightarrow f(\beta) = Y^T\Sigma^{-1}Y - Y^T\Sigma^{-1}X\beta - \beta^TX^T\Sigma^{-1}Y + \beta^TX^T\Sigma^{-1}X\beta$

($Y^T\Sigma^{-1}X\beta = (Y^T\Sigma^{-1}X\beta)^T = \beta^TX^T\Sigma^{-1}Y$ is a scalar)

$\Rightarrow f(\beta) = Y^T\Sigma^{-1}Y - 2\beta^TX^T\Sigma^{-1}Y + \beta^TX^T\Sigma^{-1}X\beta$

$\Rightarrow \nabla_\beta f(\beta) = 0 - 2X^T\Sigma^{-1}Y + 2X^T\Sigma^{-1}X\beta$

Set $\nabla_\beta f(\beta) = 0$

$\Rightarrow -2X^T\Sigma^{-1}Y + 2X^T\Sigma^{-1}X\hat{\beta}_{GLS} = 0$

$\Rightarrow X^T\Sigma^{-1}X\hat{\beta}_{GLS} = X^T\Sigma^{-1}Y$

$\Rightarrow \hat{\beta}_{GLS} = (X^T\Sigma^{-1}X)^{-1}X^T\Sigma^{-1}Y$

Scanned with CamScanner

HW7

Giang Vu

10/6/2020

Problem 3 (6.10 & 7.4)

6.10 - a

```
#read data
hw7 <- read.csv("/Users/giangvu/Desktop/STAT 2131 - Applied Stat Methods 1/HW/hw7/hw7.3data.csv",sep =
mod73 <- lm(Y ~ X1 + X2 + X3, data = hw7)
sum73 <- summary(mod73)
sum73
```

```
##
## Call:
## lm(formula = Y ~ X1 + X2 + X3, data = hw7)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -264.05 -110.73  -22.52   79.29  295.75
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.150e+03  1.956e+02  21.220  < 2e-16 ***
## X1           7.871e-04  3.646e-04   2.159   0.0359 *
## X2          -1.317e+01  2.309e+01  -0.570   0.5712
## X3           6.236e+02  6.264e+01   9.954 2.94e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 143.3 on 48 degrees of freedom
## Multiple R-squared:  0.6883, Adjusted R-squared:  0.6689
## F-statistic: 35.34 on 3 and 48 DF,  p-value: 3.316e-12
```

The estimated regression function is

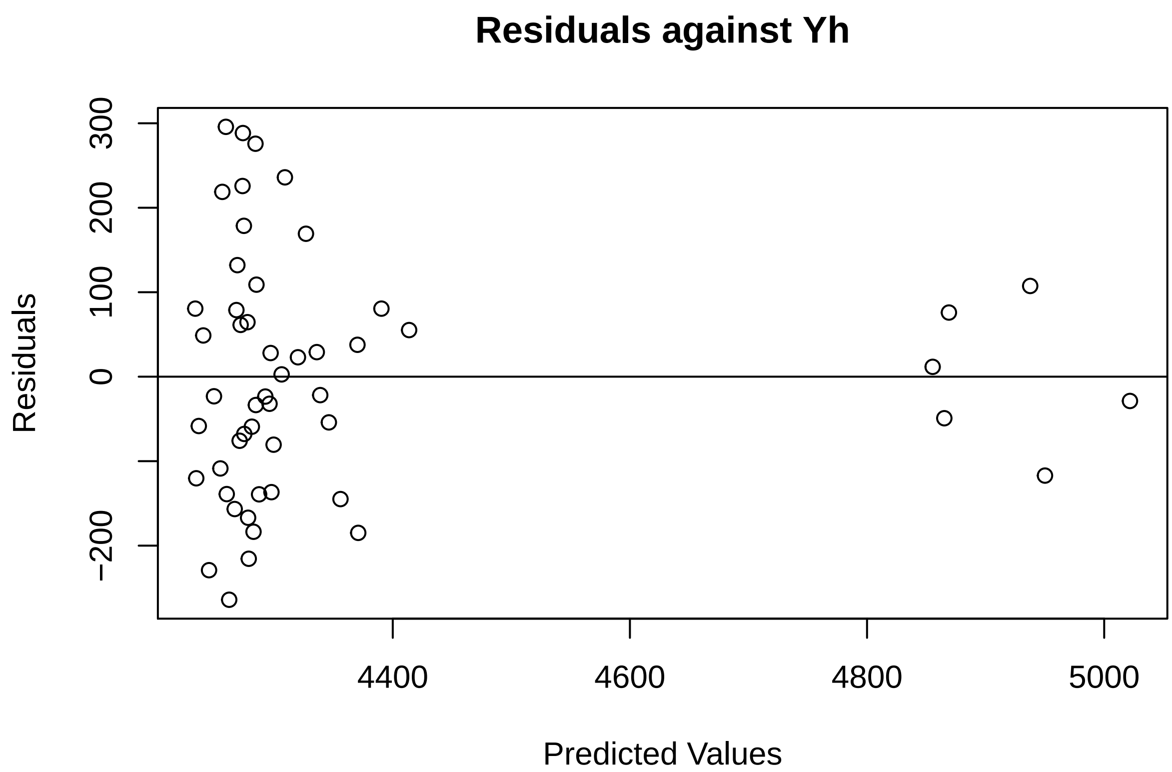
$$\hat{Y} = 4150 + 0.0007871X_1 - 13.17X_2 + 623.6X_3$$

Interpretation of coefficients:
When number of cases shipped increases by 1 case, the total labor hours is expected to increase by 0.0007871 hour.
When indirect costs of total labor hours increases by 1 percentage point, the total labor hours is expected to decrease by 13.17 hours.

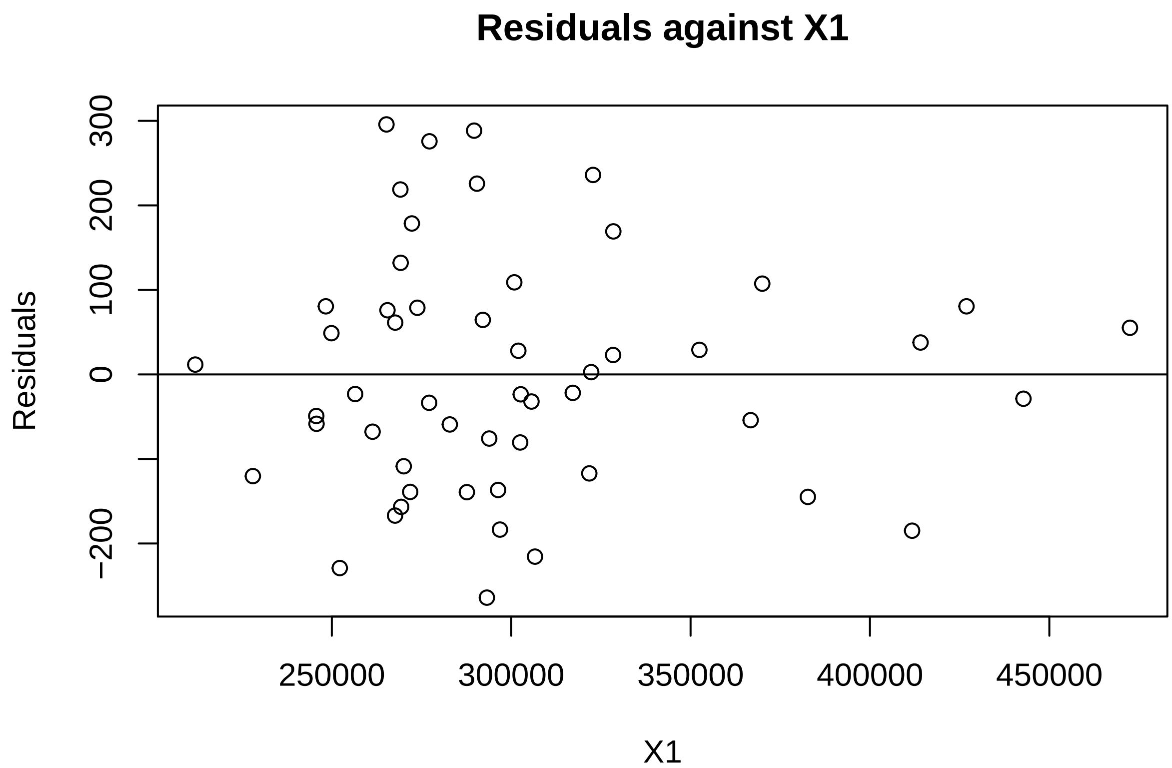
When the week has a holiday, the total labor hours is expected to increase by 623.6 hours, and there's no change expected with total labor hours when the week has no holiday.

6.10 - c

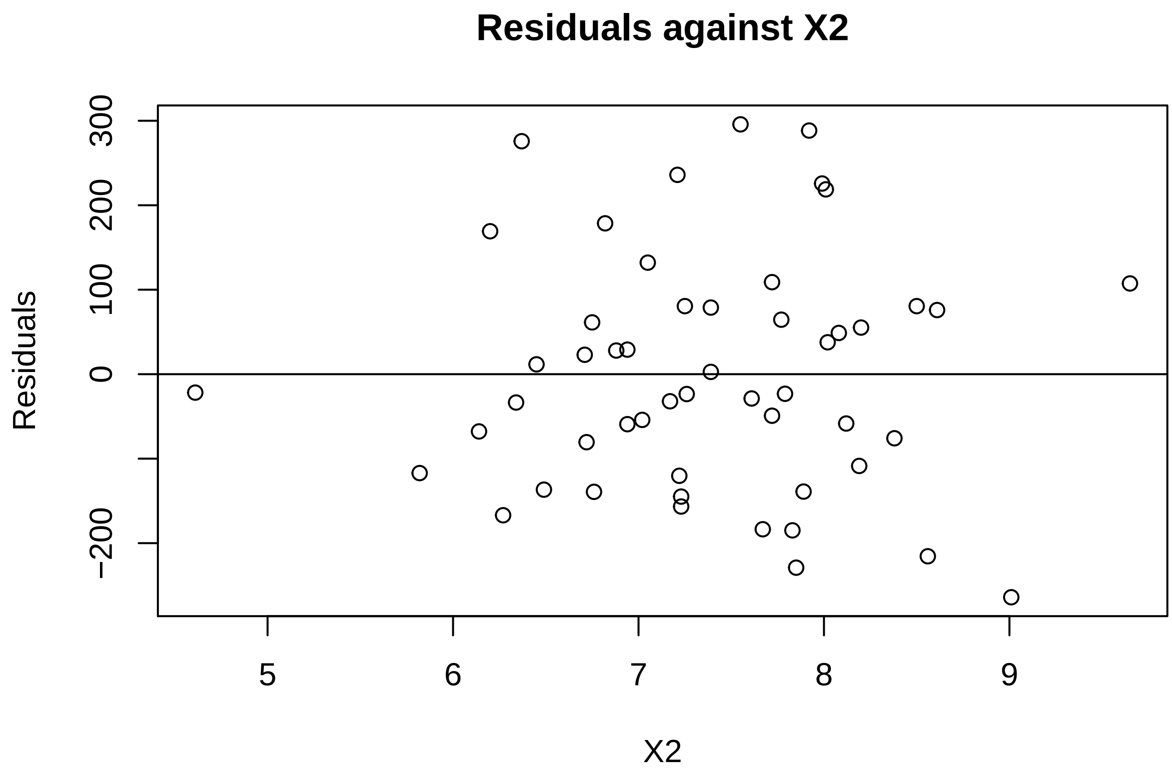
```
#residuals against Yh
plot(predict(mod73), resid(mod73),
     ylab  = "Residuals", xlab = "Predicted Values", main ="Residuals against Yh")
abline(0,0)
```



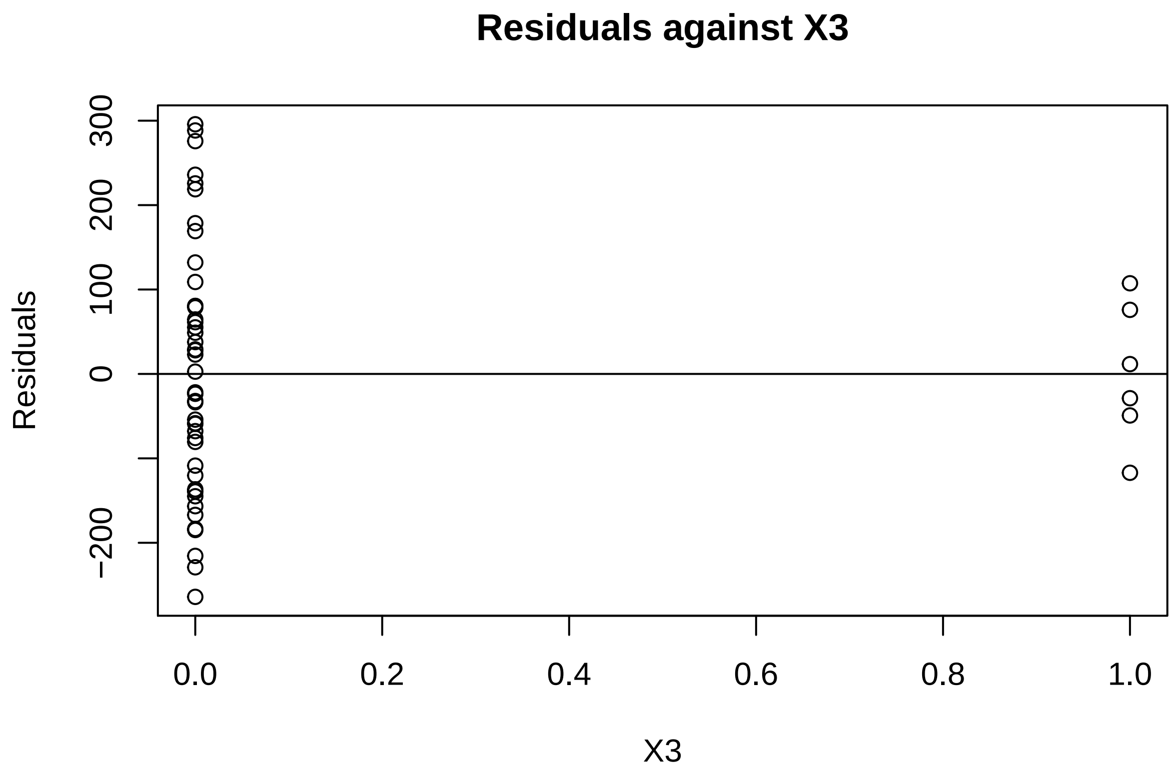

```
#residuals against X1
plot(hw7$X1, resid(mod73),
     ylab  = "Residuals", xlab = "X1", main ="Residuals against X1")
abline(0,0)
```



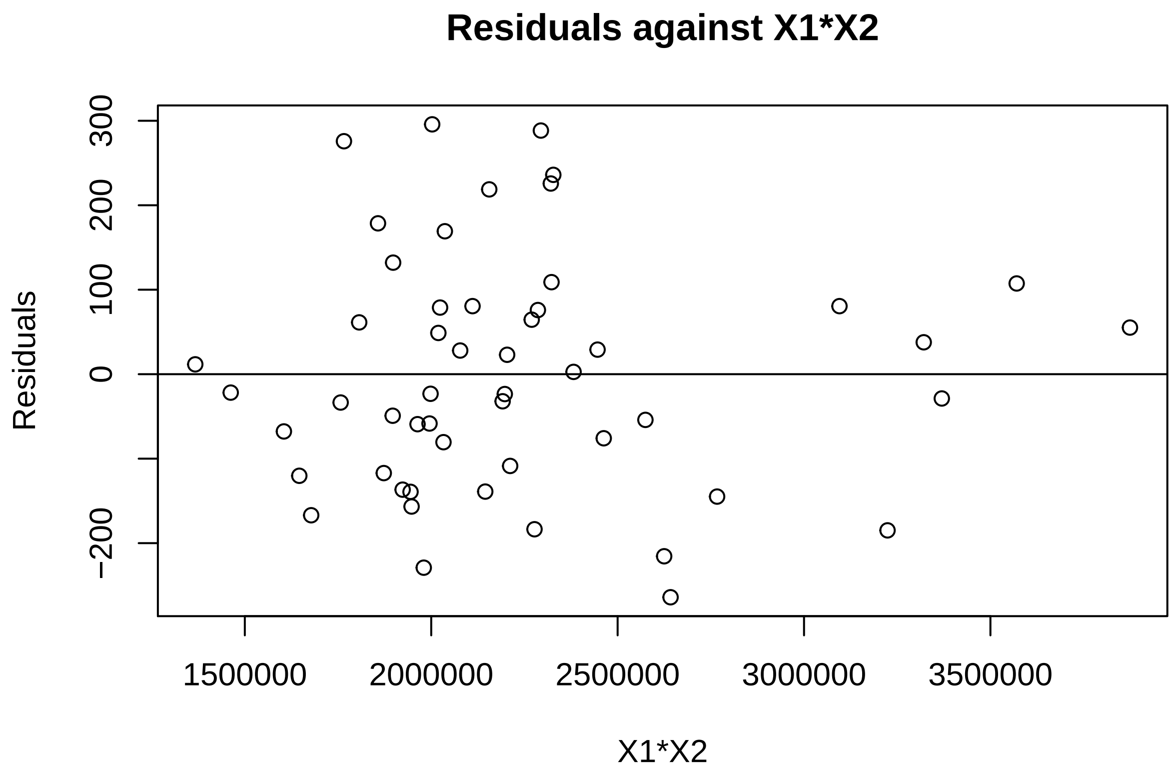

```r
#residuals against X2
plot(hw7$X2, resid(mod73),
     ylab  = "Residuals", xlab = "X2", main ="Residuals against X2")
abline(0,0)
```



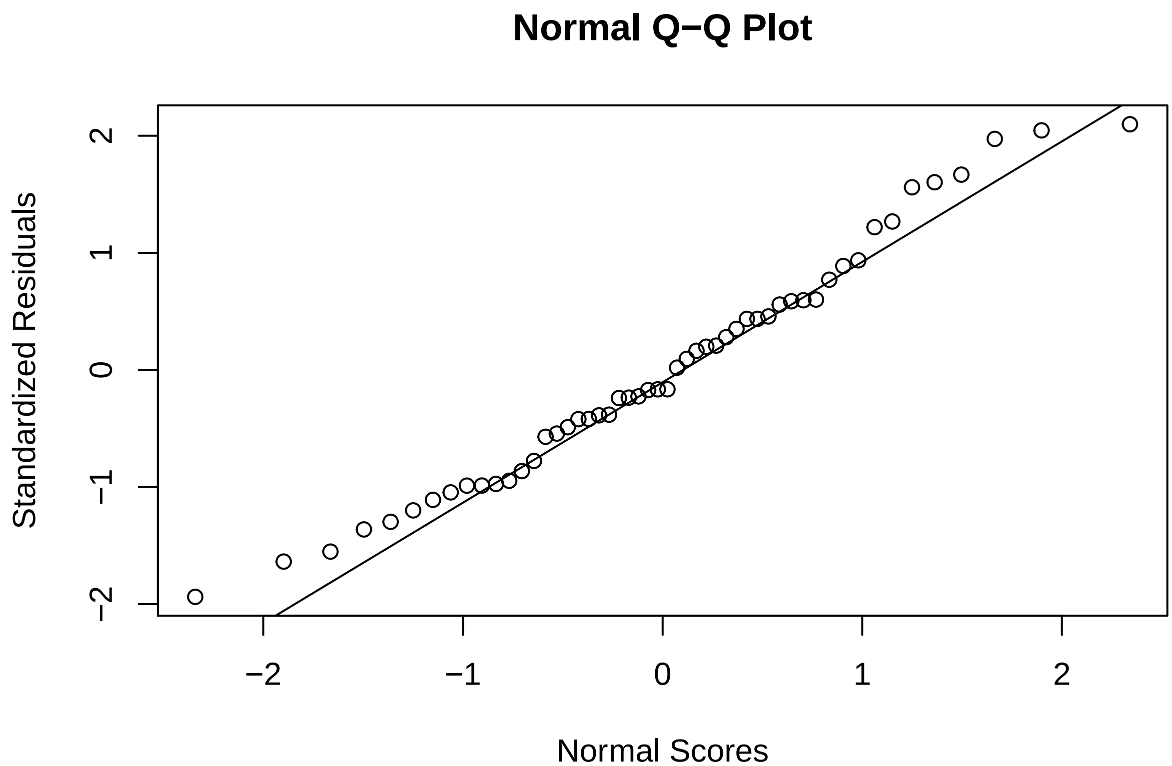

```
#Residuals against X3
plot(hw7$X3, resid(mod73),
     ylab  = "Residuals", xlab = "X3", main ="Residuals against X3")
abline(0,0)
```




```r
#Residuals against X1*X2
plot(hw7$X1*hw7$X2, resid(mod73),
     ylab  = "Residuals", xlab = "X1*X2", main ="Residuals against X1*X2")
abline(0,0)
```




```r
#normal prob plot
qqnorm(rstandard(mod73),
       ylab="Standardized Residuals",
       xlab="Normal Scores")
qqline(rstandard(mod73))
```



From the residual plots, we can see that there are no relationships between the residuals and X1, X2, and X1*X2. There's some pattern between residuals and predicted values of Y, and between residuals and X3 as well. From the normal probability plot, we can see that the residuals have a relatively normal distribution.

7.4 - b

```
linearHypothesis(mod73,c("X2=0"))
```

```
## Linear hypothesis test
##
## Hypothesis:
## X2 = 0
##
## Model 1: restricted model
## Model 2: Y ~ X1 + X2 + X3
##
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     49 992204
## 2     48 985530  1    6674.6 0.3251 0.5712
```

We test the following hypotheses

$$H_0 : \beta_2 = 0$$

$$H_a : \beta_2 \neq 0$$

$$SSR(X_2|X_1, X_3) = 6674.6$$

$$SSE(X_1, X_2, X_3) = 985530$$

$$F^* = \frac{6674.6/(2-1)}{985530/(52-4)} = 0.325 < F(0.95, 1, 48) = 4.04$$

If F* is smaller than or equal to F(0.95,1,48), we fail to reject H0, otherwise we reject H0.
In this situation, we fail to reject H0 and conclude that we can drop X2 from our regression.
From the test, we get p-value = 0.5712, which is greater than alpha = 0.05.

Problem 4 (6.16)

#6.16 - a

```
#read data
hw7.4 <- read.csv("/Users/giangvu/Desktop/STAT 2131 - Applied Stat Methods 1/HW/hw7/hw7.4data.csv",sep=
#fit model
mod74 <- lm(Y ~ X1 + X2 + X3, data = hw7.4)
sum74 <- summary(mod74)
sum74
```

```
##
## Call:
## lm(formula = Y ~ X1 + X2 + X3, data = hw7.4)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -18.3524  -6.4230   0.5196   8.3715  17.1601
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 158.4913    18.1259   8.744 5.26e-11 ***
## X1           -1.1416     0.2148  -5.315 3.81e-06 ***
## X2           -0.4420     0.4920  -0.898   0.3741
## X3          -13.4702     7.0997  -1.897   0.0647 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.06 on 42 degrees of freedom
## Multiple R-squared:  0.6822, Adjusted R-squared:  0.6595
## F-statistic: 30.05 on 3 and 42 DF,  p-value: 1.542e-10
```

```
#hypothesis testing
linearHypothesis(mod74,c("X1=0","X2=0","X3=0"))
```

```
## Linear hypothesis test
```

```
##
## Hypothesis:
## X1 = 0
## X2 = 0
## X3 = 0
##
## Model 1: restricted model
## Model 2: Y ~ X1 + X2 + X3
##
##   Res.Df     RSS Df Sum of Sq      F    Pr(>F)
## 1     45 13369.3
## 2     42  4248.8  3    9120.5 30.052 1.542e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We test the following hypotheses

$$H_0 : \beta_1 = \beta_2 = \beta_3 = 0$$

$$H_a : not\ all\ beta's\ = 0$$

$$SSR = 9120.5$$

$$SSE = 4248.8$$

$$F^* = \frac{9120.5/(4-1)}{4248.8/(46-4)} = 30.05 > F(0.9, 3, 42) = 2.23$$

If F* is smaller than or equal to F(0.9,3,42), we fail to reject H0, otherwise we reject H0.
In this situation, we reject H0 and conclude that there is a regression relation between Y and the 3 X's.
From the test, we get p-value = 1.542e-10, which is smaller than alpha = 0.1.

#6.16 - b
The interval estimates of B1, B2, B3 using a 90% family confidence coefficient is calculated below

```
confint(mod74,level=1-0.10/(2*3))
```

```
##                 0.833 %    99.167 %
## (Intercept) 113.291314 203.6911891
## X1           -1.677249  -0.6059751
## X2           -1.668803   0.7847947
## X3          -31.174356   4.2340294
```

From the result, we could see that only the interval for variable X1 (patient's age) doesn't contain 0, while X2 (illness severity) and X3 (anxiety level) do. Therefore, we could say that X1 might be the only significant predictor here.

#6.16 - c

$$SSR = 9120.5$$

$$SSTO = 13369.3$$

The coefficient of multiple determination is

$$R^2 = \frac{9120.5}{13369.3} = 0.68$$

This coefficient of multiple determination tells us that about 68% of the variability in patient satisfaction can be explained by patient's age, severity of illness and anxiety level together.

5) a) X has n rows and there are r rows that are identical to the ith row

The ith diagonal entry of H is h_i or H_{ii}

Scanned with CamScanner

We have $H_{ii} = (H^2)_{ii} = \sum_{j=1}^{n} H_{ij} H_{ji}$ $\left(\overset{\text{because}}{H^2 = H}\right)$

Because H is also symmetric $\Rightarrow H_{ij} = H_{ji}$

$\Rightarrow h_i = H_{ii} = \sum_{j=1}^{n} (H_{ij})^2 \geq \sum_{\text{only identical rows}} (H_{ij})^2$

But we only r identical rows

$\Rightarrow h_i \geq r H_{ij}^2 = r h_{ij}^2$

$\Rightarrow h_{ii} \leq \frac{1}{r}$

Now define a matrix $A = H - \frac{1}{n} I_n I_n^T$

A is an orthogonal projection matrix, A is symmetric, idempotent, and positive semidefinite

We have $H_{ii} = \underset{(\geq 0)}{A_{ii}} + \underset{(= \frac{1}{n})}{(\frac{1}{n} I_n I_n^T)_{ii}}$

$\Rightarrow H_{ii} \geq \frac{1}{n}$

$\Rightarrow h_i \geq \frac{1}{n}$

So we have $\frac{1}{n} \leq h_i \leq \frac{1}{r}$

b) The equalities exist when the number of identical rows in X are exactly n

$\Rightarrow$ All rows of X are identical $\Rightarrow n = r$

$\Rightarrow \frac{1}{n} = \frac{1}{r} = h_i$

Scanned with CamScanner