# #1

① By definition we have $R^2 = 1 - \dfrac{\sum_{n=1}^{N}(y_n - \hat{y}_n)^2}{\sum_{n=1}^{N}(y_n - \bar{y})^2}$.

And we know $\sum_{n=1}^{N}(y_n - \hat{y}_n)^2$ is the sum of residual of square, when the modules values exactly match the observed value which $(y_n - \hat{y}_n)$ is equal to 0. the sum is 0, then the $R^2$ is 1. Since the residuals is producing a model that minimises residuals, the residuals of the model must be smaller than in the $y=\bar{y}$ case, so $\sum_{n=1}^{N}(y_n - \hat{y}_n)^2$ must smaller than $\sum_{n=1}^{N}(y_n - \bar{y})^2$ which smaller than 1. one minus a number smaller than 1 imply bigger than 0. which $0 \le R^2 \le 1$.

② No, the $R_a^2$ can be negative and it always smaller or equal to $R^2$. Since we known $0 \le R \le 1$, by the definition $R_a^2 = 1 - (1-R^2)\left[\dfrac{N-1}{N-M-1}\right]$,

a) $R^2 \to 1$ : $R_a^2 = 1 - (1-1)\left[\dfrac{N-1}{N-M-1}\right] = 1$

b) $R^2 \to 0$ : $R_a^2 = 1 - (1-0)\left[\dfrac{N-1}{N-M-1}\right] = 1 - \dfrac{N-1}{N-M-1} = \dfrac{\cancel{N-M-1-N+1}}{N-M-1}$

$= \dfrac{-M}{N-M-1}$

$M+1 < N. \Rightarrow M < N-1$

then $N-1-M$ must be positive, then $\dfrac{-M}{N-1-M}$ will be negative. so when $R^2$ approach to 0 and the number of sample is bigger than variable., the $R_a^2$ could be negative.

#2

Since noise / error are normally - distributed, then $Y_n$ must be normally - distributed as well. Since we have

$$e_n = \epsilon_n - \sum_{i=1}^{N} h_{ni} \epsilon_i$$

$$= (1 - h_{nn}) \epsilon_n - \sum_{i=1}^{n-1} h_{ni} \epsilon_i - \sum_{i=n+1}^{N} h_{ni} \epsilon_i$$

Since a linear combination of independent normally - distributed ramdon variable is also normally - distributed, this mean that the residual $e_n$ are also normally - distributed.

#3

From lecture 8, we get

1) The estimator $\hat{\beta}_m$ for the regression parameters $\beta_m$ is unbiased. $E(\hat{\beta}_0) = \beta_0$.

2) The $\hat{\beta}_m$ are of minimum variance among all unbiased. linear estimators for $\beta_m$. Among all unbiased, linear estimators $\alpha_m$, the error $E[(\alpha_m - \beta_m)] = E[(\alpha_m - E[\alpha_m])^2]$ is minimized when $\alpha_m = \hat{\beta}_m$.

# #4

```
Call:
lm(formula = Temp ~ Ozone + Wind + Month, data = airquality)

Residuals:
     Min       1Q   Median       3Q      Max 
-21.4801  -4.2884   0.4907   4.6106  12.4071 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 58.97480    4.10572  14.364  < 2e-16 ***
Ozone        0.17060    0.02183   7.816 3.22e-12 ***
Wind        -0.24236    0.20302  -1.194    0.235    
Month        1.95866    0.39830   4.918 3.03e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.159 on 112 degrees of freedom
  (37 observations deleted due to missingness)
Multiple R-squared:  0.5893,	Adjusted R-squared:  0.5783 
F-statistic: 53.58 on 3 and 112 DF,  p-value: < 2.2e-16
```

① Yes, this model is significant, since the p-value is $< 2.2e-16$.

② The Ozone and Month are important for the model, since they value are $3.22e-12$ and $3.03e-06$. By comparing with those two values the wind has higher p-value.

③ We know the residual should be normally distribution, and base on the median of residual is close to 0 and 1Q & 3Q nearly symmetric, then we can imply this is a vaild model.

#5

```
> n = nrow(mtcars)
> p = length(coef(mtcars))
> X = cbind(rep(1, n), mtcars$disp, mtcars$hp)
> y = mtcars$mpg
>
> (beta_hat = solve(t(X) %*% X) %*% t(X) %*% y)
            [,1]
[1,] 30.73590425
[2,] -0.03034628
[3,] -0.02484008
> x <- lm(mpg ~ disp + hp, data = mtcars)
> x

Call:
lm(formula = mpg ~ disp + hp, data = mtcars)

Coefficients:
(Intercept)         disp           hp
   30.73590     -0.03035     -0.02484
```

Yes, they are same.

#6

①

```
Call:
lm(formula = dheight ~ mheight, data = Heights)

Residuals:
   Min     1Q Median     3Q    Max 
-7.397 -1.529  0.036  1.492  9.053 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 29.91744    1.62247   18.44   <2e-16 ***
mheight      0.54175    0.02596   20.87   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.266 on 1373 degrees of freedom
Multiple R-squared:  0.2408,	Adjusted R-squared:  0.2402 
F-statistic: 435.5 on 1 and 1373 DF,  p-value: < 2.2e-16
```

The estimate of daughter is 29.91744 and mother is 0.54175. The SD of daughter is 1.62247 and 0.02596, the value of coefficient is 0.2408, the estimate of variance is $(2.26)^2 \approx 5.13$. And since P close to 0, we can know $\beta_1 \neq 0$. Base on $R^2 = 0.241$ which is small that we can't sure this modle is valid.

②

```
> confint(H, level = 0.99)
                 0.5 %     99.5 %
(Intercept) 25.7324151 34.1024585
mheight      0.4747836  0.6087104
>
```

③

```
> predict(H, data.frame(mheight=64), interval="prediction", level=.99)
       fit      lwr      upr
1 64.58925 58.74045 70.43805
>
```

# #7

①

```
> B <- lm(Length ~ Age, data = wblake)
> predict(B, data.frame(Age=c(2, 4, 6)), interval="prediction", level = 0.95)
       fit       lwr      upr
1 126.1749  69.73151 182.6184
2 186.8227 130.45720 243.1882
3 247.4705 191.05332 303.8877
>
```

②

```
> predict(B, data.frame(Age= 9), interval="prediction", level = 0.95)
       fit      lwr      upr
1 338.4422 281.7056 395.1788
>
```

Since the dataset don't have any fishs older than 8, we can't know if the function apply for the Age 9 fishs.

#8

$$E(y|x) = \beta_1 x$$

① $\hat{\beta}_1$ is unbiased :

$$RSS = \Sigma(y_i - \beta_1 x_i)^2$$

$$\frac{\partial \beta_1 RSS}{\partial \beta_1} = -2\Sigma(y_i - \beta_1 x_i)\cdot x_i = 0$$

$$\Sigma(y_i x_i - \beta_1 x_i^2) = 0$$

$$\Sigma y_i x_i - \Sigma \beta_1 x_i^2 = 0$$

$$\Sigma y_i x_i - \beta_1 \Sigma x_i^2 = 0$$

$$\frac{\Sigma y_i x_i}{\Sigma x_i^2} = \beta_1$$

$$E(\hat{\beta}_1|x) = \frac{E(\Sigma x_i y_i)}{E(\Sigma x_i^2)} = \frac{\Sigma x_i \cdot E(\Sigma y_i)}{\Sigma x_i^2} = \frac{\Sigma x_i \cdot (\beta_1 x_i)}{\Sigma x_i^2}$$

$$= \frac{\cancel{\Sigma x_i^2}\cdot \beta_1}{\cancel{\Sigma x_i^2}}$$

$$= \beta_1$$

Show $Var(\hat{\beta}_1|x) = \dfrac{\sigma^2}{\Sigma x_i^2}$.

$$Var(\hat{\beta}_1|x) = Var\left(\frac{\Sigma x_i y_i}{\Sigma x_i^2}\right) = \frac{Var(y_i)\cdot \Sigma x_i^2}{(\Sigma x_i^2)^2}$$

$$= \frac{\sigma^2 \cdot}{\Sigma x_i^2}$$

Expression of $\sigma^2$:

$$
\begin{aligned}
\sigma^2 = RSS &= \Sigma(y_i - \hat{\beta}_1 x_i)^2 \\
&= \Sigma(y_i^2 - 2y_i\hat{\beta}_1 x_i + (\hat{\beta}_1 x_i)^2) \\
&= \Sigma y_i^2 - 2\hat{\beta}_1 \Sigma y_i x_i + \hat{\beta}_1^2 \Sigma x_i^2 \\
&= \Sigma y_i^2 - 2\left(\frac{\Sigma x_i y_i}{\Sigma x_i^2}\right)\cdot \Sigma y_i x_i + \left(\frac{\Sigma x_i y_i}{\Sigma x_i^2}\right)^2 \cdot \Sigma x_i^2 \\
&= \Sigma y_i^2 - 2\frac{\Sigma(x_i y_i)^2}{\Sigma x_i^2} + \frac{\Sigma(x_i y_i)^2}{\Sigma x_i^2} \\
&= \Sigma y_i^2 - \frac{\Sigma(x_i y_i)^2}{\Sigma x_i^2}
\end{aligned}
$$

$df = n-1$

②

```
Call:
lm(formula = Y ~ X - 1, data = snake)

Coefficients:
     X
0.5204
```

$\hat{\beta}_1 = 0.5204$