STAT30040 Homework 7

Sarah Adilijiang

Problem 1

(a)

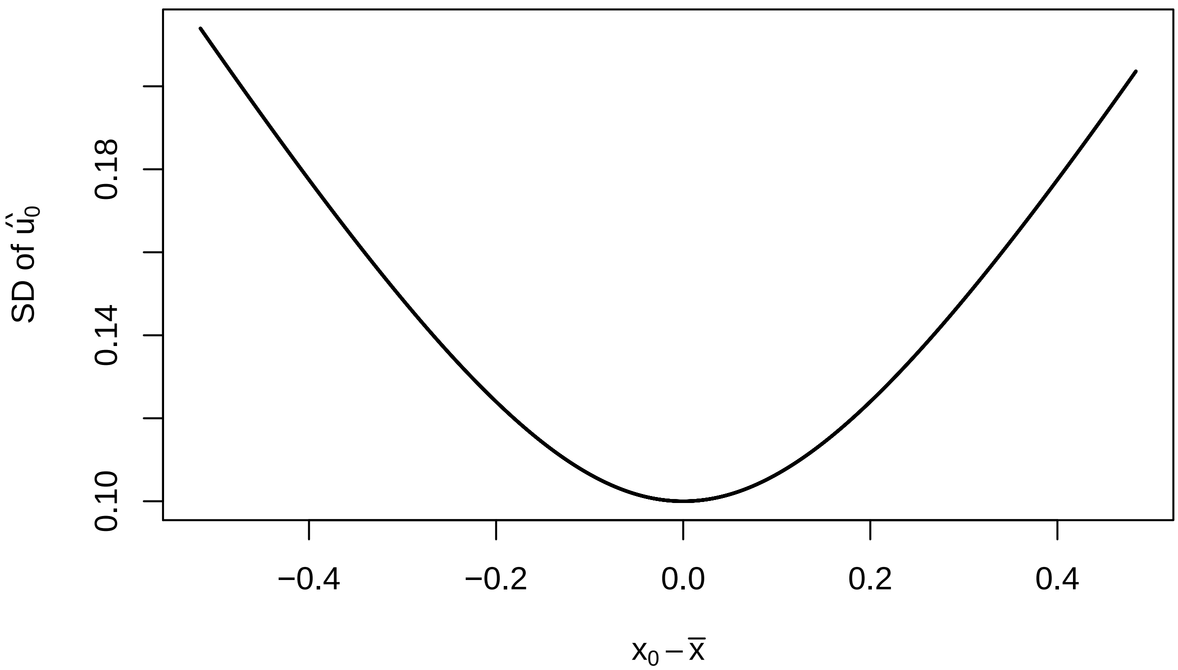
$$Var(\hat{\mu}_0) = Var(\hat{\beta}_0 + \hat{\beta}_1 x_0) = Var(\hat{\beta}_0) + x_0^2 Var(\hat{\beta}_1) + 2x_0 Cov(\hat{\beta}_0,\, \hat{\beta}_1) = \frac{\sigma^2}{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2} \left(\frac{\sum\limits_{i=1}^{n} x_i^2}{n} + x_0^2 - 2x_0\bar{x} \right)$$

$$= \frac{\sigma^2}{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2} \left(\frac{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2 + n\bar{x}^2}{n} + x_0^2 - 2x_0\bar{x} \right) = \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2 + x_0^2 - 2x_0\bar{x}}{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2} \right) = \sigma^2 \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2} \right)$$

(b)

$$SD(\hat{\mu}_0) = \sqrt{Var(\hat{\mu}_0)} = \sigma \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2}}$$

```r
sigma=1;   n=100
x = runif(100,0,1)
SXX = sum((x-mean(x))^2)
u = seq(0,1,by=0.0001) - mean(x)
sd = sigma * sqrt(1/n + u^2/SXX)
plot(sd~u, xlab=expression(x[0] - bar(x)), ylab=expression(paste("SD of ",hat(u[0]))), cex=0.1)
```



(c)

Since $E(\hat{\mu}_0) = E(\hat{\beta}_0 + \hat{\beta}_1 x_0) = \beta_0 + \beta_1 x_0 = \mu_0$ and $Var(\hat{\mu}_0) = \sigma^2 \left(\frac{1}{n} + \frac{(x_0-\bar{x})^2}{\sum\limits_{i=1}^{n}(x_i-\bar{x})^2} \right)$

Thus under the assumption of normality, we have the following distributions:

(1)$\hat{\mu}_0 \sim N\left(\mu_0,\, \sigma^2 \left(\frac{1}{n} + \frac{(x_0-\bar{x})^2}{\sum\limits_{i=1}^{n}(x_i-\bar{x})^2} \right)\right)$, hence $\frac{\hat{\mu}_0-\mu_0}{\sigma\sqrt{\frac{1}{n}+\frac{(x_0-\bar{x})^2}{\sum\limits_{i=1}^{n}(x_i-\bar{x})^2}}} \sim N(0,1)$

(2)$\frac{RSS}{\sigma^2} \sim \chi^2_{n-2}$

(3)These two distributions are independent

So from these three points, we can obtain:

$$\frac{\frac{\hat{\mu}_0-\mu_0}{\sigma\sqrt{\frac{1}{n}+\frac{(x_0-\bar{x})^2}{\sum\limits_{i=1}^{n}(x_i-\bar{x})^2}}}}{\sqrt{\frac{RSS}{\sigma^2}/(n-2)}} \sim t_{(n-2)}$$

Plug in $\hat{\sigma} = \frac{RSS}{n-2}$, we get:

$$\frac{\hat{\mu}_0 - \mu_0}{\hat{\sigma}\sqrt{\frac{1}{n} + \frac{(x_0-\bar{x})^2}{\sum\limits_{i=1}^{n}(x_i-\bar{x})^2}}} \sim t_{(n-2)}$$

Therefore, a $1-\alpha = 95\%$ confidence interval for μ_0 is:

$$\hat{\mu}_0 \;\pm\; t_{\frac{\alpha}{2},\,(n-2)}\;\hat{\sigma}\sqrt{\frac{1}{n} + \frac{(x_0-\bar{x})^2}{\sum\limits_{i=1}^{n}(x_i-\bar{x})^2}}$$

where $\hat{\sigma} = \frac{RSS}{n-2}$

Problem 2

(a)

Since the e_0 is independent of the original observations and has the variance σ^2, so we have:

$$Var(\hat{Y}_0 - Y_0) = Var(\hat{\beta}_0 + \hat{\beta}_1 x_0 - \beta_0 - \beta_1 x_0 - e_0) = Var(\hat{\beta}_0 + \hat{\beta}_1 x_0) + Var(e_0)$$

$$= \sigma^2\left(\frac{1}{n} + \frac{(x_0-\bar{x})^2}{\sum\limits_{i=1}^{n}(x_i-\bar{x})^2}\right) + \sigma^2 \;=\; \sigma^2\left(1 + \frac{1}{n} + \frac{(x_0-\bar{x})^2}{\sum\limits_{i=1}^{n}(x_i-\bar{x})^2}\right)$$

Comparing with $\hat{\mu}_0$ in problem 1, we have: $Var(\hat{Y}_0 - Y_0) - Var(\hat{\mu}_0) = \sigma^2$

(b)

Since $E(\hat{Y}_0 - Y_0) = E(\hat{\beta}_0 + \hat{\beta}_1 x_0 - \beta_0 - \beta_1 x_0 - e_0) = 0$ and $Var(\hat{Y}_0 - Y_0) = \sigma^2\left(1 + \frac{1}{n} + \frac{(x_0-\bar{x})^2}{\sum\limits_{i=1}^{n}(x_i-\bar{x})^2}\right)$

Thus under the assumption of normality and that e_0 is normally distributed, we have the distribution of $\hat{Y}_0 - Y_0$:

$$\hat{Y}_0 - Y_0 \sim N\left(0,\, \sigma^2\left(1 + \frac{1}{n} + \frac{(x_0-\bar{x})^2}{\sum\limits_{i=1}^{n}(x_i-\bar{x})^2}\right)\right)$$

So $\frac{\hat{Y}_0 - Y_0}{\sigma\sqrt{1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{\sum\limits_{i=1}^{n}(x_i-\bar{x})^2}}} \sim N(0,1)$, since $\frac{RSS}{\sigma^2} \sim \chi^2_{n-2}$ and these two distributions are independent, we can obtain that:

$$\frac{\frac{\hat{Y}_0 - Y_0}{\sigma\sqrt{1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{\sum\limits_{i=1}^{n}(x_i-\bar{x})^2}}}}{\sqrt{\frac{RSS}{\sigma^2}/(n-2)}} \sim t_{(n-2)}$$

Plug in $\hat{\sigma} = \frac{RSS}{n-2}$, we get:

$$\frac{\hat{Y}_0 - Y_0}{\hat{\sigma}\sqrt{1 + \frac{1}{n} + \frac{(x_0-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}}} \sim t_{(n-2)}$$

Therefore, a $100(1-\alpha)\%$ prediction interval for Y_0 is:

$$\hat{Y}_0 \ \pm \ t_{\frac{\alpha}{2},\ (n-2)}\ \hat{\sigma}\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$$

where $\hat{\sigma} = \frac{RSS}{n-2}$ and $\hat{Y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$

Problem 3

(a)

α_0 is the intercept term, which represents the average respiratory resistance of a child with asthma and zero height value. If setting α_0 to zero, we assume that the average respiratory resistance of a child with asthma and zero height value is zero.

α_1 is the coefficient of continuous predictor x_i (**height**), which represents the change in the average respiratory resistance of a child with asthma for an additional unit (cm) of height. If setting α_1 to zero, we assume that the average respiratory resistance of a child with asthma does not depend on the height of that child.

β_0 is the coefficient of indicator predictor z_i, which represents the amount that the average respiratory resistance of a child with cystic fibrosis is higher than that of a child with asthma when they both have zero height value. If setting β_0 to zero, we assume that the average respiratory resistance of a child with cystic fibrosis is the same as a child with asthma when they both have zero height value.

β_1 is the coefficient of the interaction term between x_i and z_i, which represents the amount that the change in the average respiratory resistance of a child with cystic fibrosis is higher than that of a child with asthma for an additional unit (cm) of height. If setting β_1 to zero, we assume that the change in the average respiratory resistance of a child with cystic fibrosis is the same as a child with asthma for an additional unit (cm) of height.

(b)

```r
mod = lm(resistance ~ height * z, mydata)
summary(mod)
```

```
##
## Call:
## lm(formula = resistance ~ height * z, data = mydata)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.9608 -2.8496 -0.5944  1.5739 12.6113
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 27.51570    6.15000   4.474 3.34e-05 ***
## height      -0.13606    0.04921  -2.765  0.00749 **
## z1          -3.70866   13.04125  -0.284  0.77707
## height:z1    0.01145    0.12270   0.093  0.92596
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.134 on 62 degrees of freedom
## Multiple R-squared:  0.1278, Adjusted R-squared:  0.08565
## F-statistic:  3.03 on 3 and 62 DF,  p-value: 0.03594
```

Answer:

In the model summary, it shows the results of the four t-tests for all the parameters, where the null hypothesis is that their parameter values are zero, respectively.

We can see that α_0 (intercept term) is significant at 0.1% significance level, which means the average respiratory resistance of a child with asthma and zero height value is not zero.

α_1 is significant at 1% significance level, which means the average respiratory resistance of a child with asthma does depend on the height of the child.

β_0 is not significant, which means there is no significant evidence that the average respiratory resistance of a child with cystic fibrosis is different from a child with asthma when they both have zero height value.

β_1 is also not significant, which means there is no significant evidence that the change in the average respiratory resistance of a child with cystic fibrosis is different from a child with asthma for an additional unit (cm) of height.

To sum up, the t-test results show that the average respiratory resistance of a child does depend on the height of the child, but there is no difference between two diseases when controlling for the height.

(c)

```
nmod = lm(resistance ~ height, mydata)
anova(nmod, mod)
```

```
## Analysis of Variance Table
##
## Model 1: resistance ~ height
## Model 2: resistance ~ height * z
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     64 1110.0
## 2     62 1059.7  2    50.358 1.4732 0.2371
```

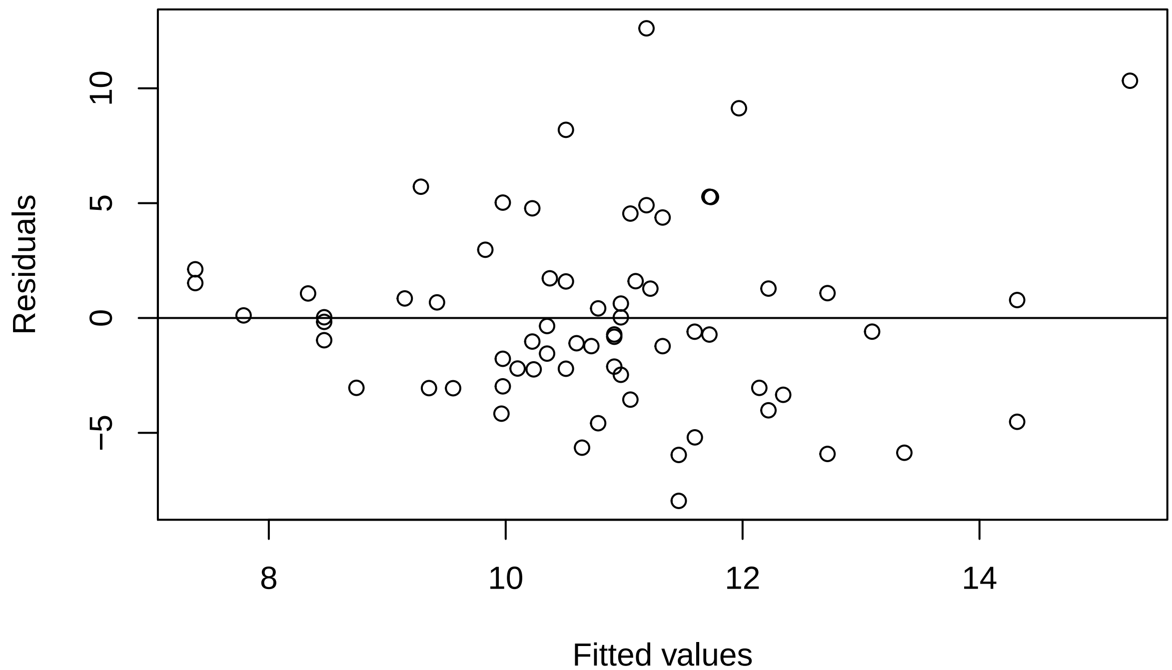
Answer:

$H_0 : \beta_0 = \beta_1 = 0$, this null hypothesis means that the average respiratory resistance of a child is the same for a child with asthma and cycstic fibrosis when controlling for the height of the child.

The p-value of F-test is $0.2371 > 0.05$, so we do not reject the null hypothesis. Therefore, there is no significant evidence that the average respiratory resistance of a child is different from a child with asthma and cycstic fibrosis when controlling for the height of the child.

(d)

```
plot(residuals(mod)~fitted(mod), xlab="Fitted values", ylab="Residuals")
abline(h=0)
```



Answer:

The variance of residuals seem to slightly increase when the fitted value increases, which shows a heteroskedasticity problem. Therefore, the model assumption of constant variance does not hold.

Problem 4

(a)

In the new model: $z_i = u_i\beta_0 + v_i\beta_1 + \delta_i$, where $z_i = \rho_i^{-1} y_i$, $u_i = \rho_i^{-1}$, $v_i = \rho_i^{-1} x_i$, $\delta_i = \rho_i^{-1} e_i$, since ρ_i are know constants, thus z_i is still a random variable conditioning on given data v_i, and the random error term now is δ_i.

And for the random error term, since $E(e_i) = 0$ and $Var(e_i) = \rho_i^2\sigma^2$, and e_i are independet, thus $E(\delta_i) = E(\rho_i^{-1} e_i) = 0$ and $Var(\delta_i) = Var(\rho_i^{-1} e_i) = \sigma^2$, and δ_i are independent. Therefore, δ_i have mean zero and constant variance and are independent.

As a result, the new model $z_i = u_i\beta_0 + v_i\beta_1 + \delta_i$ satisfies the assumptions of the standard statistical model.

(b)

In the least squares estimation, we want to minimize: $S(\beta_0, \beta_1) = \sum_{i=1}^{n} \delta_i^2 = \sum_{i=1}^{n} (z_i - u_i\beta_0 - v_i\beta_1)^2$

So we set the following derivatives to zero:

$$\frac{\partial S}{\partial \beta_0} = -2\sum_{i=1}^{n} u_i(z_i - u_i\beta_0 - v_i\beta_1) = 0$$

$$\frac{\partial S}{\partial \beta_1} = -2\sum_{i=1}^{n} v_i(z_i - u_i\beta_0 - v_i\beta_1) = 0$$

Thus we get that the minimizers $\hat{\beta}_0$ and $\hat{\beta}_1$ satisfy:

$$\sum_{i=1}^{n} u_i z_i = \hat{\beta}_0 \sum_{i=1}^{n} u_i^2 + \hat{\beta}_1 \sum_{i=1}^{n} u_i v_i$$

$$\sum_{i=1}^{n} v_i z_i = \hat{\beta}_0 \sum_{i=1}^{n} u_i v_i + \hat{\beta}_1 \sum_{i=1}^{n} v_i^2$$

Solving for $\hat{\beta}_0$ and $\hat{\beta}_1$, we obtain:

$$\hat{\beta}_0 = \frac{\sum\limits_{i=1}^{n} v_i^2 \sum\limits_{i=1}^{n} u_i z_i - \sum\limits_{i=1}^{n} u_i v_i \sum\limits_{i=1}^{n} v_i z_i}{\sum\limits_{i=1}^{n} u_i^2 \sum\limits_{i=1}^{n} v_i^2 - (\sum\limits_{i=1}^{n} u_i v_i)^2} = \frac{\sum\limits_{i=1}^{n} \rho_i^{-2} x_i^2 \sum\limits_{i=1}^{n} \rho_i^{-2} y_i - \sum\limits_{i=1}^{n} \rho_i^{-2} x_i \sum\limits_{i=1}^{n} \rho_i^{-2} x_i y_i}{\sum\limits_{i=1}^{n} \rho_i^{-2} \sum\limits_{i=1}^{n} \rho_i^{-2} x_i^2 - (\sum\limits_{i=1}^{n} \rho_i^{-2} x_i)^2}$$

$$\hat{\beta}_1 = \frac{\sum\limits_{i=1}^{n} u_i^2 \sum\limits_{i=1}^{n} v_i z_i - \sum\limits_{i=1}^{n} u_i v_i \sum\limits_{i=1}^{n} u_i z_i}{\sum\limits_{i=1}^{n} u_i^2 \sum\limits_{i=1}^{n} v_i^2 - (\sum\limits_{i=1}^{n} u_i v_i)^2} = \frac{\sum\limits_{i=1}^{n} \rho_i^{-2} \sum\limits_{i=1}^{n} \rho_i^{-2} x_i y_i - \sum\limits_{i=1}^{n} \rho_i^{-2} x_i \sum\limits_{i=1}^{n} \rho_i^{-2} y_i}{\sum\limits_{i=1}^{n} \rho_i^{-2} \sum\limits_{i=1}^{n} \rho_i^{-2} x_i^2 - (\sum\limits_{i=1}^{n} \rho_i^{-2} x_i)^2}$$

Since $\frac{\partial^2 S}{\partial \beta_0^2} = 2\sum\limits_{i=1}^{n} u_i^2 > 0$ and $\frac{\partial^2 S}{\partial \beta_1^2} = 2\sum\limits_{i=1}^{n} v_i^2 > 0$, so the $\hat{\beta}_0$ and $\hat{\beta}_1$ are the minimizer for function $S(\beta_0, \beta_1)$.

(c)

In question (b),

$$S(\beta_0, \beta_1) = \sum_{i=1}^{n}(z_i - u_i\beta_0 - v_i\beta_1)^2 = \sum_{i=1}^{n}(\rho_i^{-1} y_i - \rho_i^{-1}\beta_0 - \rho_i^{-1} x_i \beta_1)^2 = \sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2 \rho_i^{-2}$$

Therefore, minimizing $S(\beta_0, \beta_1)$ in the question (b) is equivalent to minimizing $\sum\limits_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2 \rho_i^{-2}$, which is the weighted least squares criterion.

(d)

Since $Var(z_i) = Var(u_i\beta_0 + v_i\beta_1 + \delta_i) = Var(\delta_i) = \sigma^2$, so we have:

$$Var(\hat{\beta}_0) = \frac{(\sum\limits_{i=1}^{n} v_i^2)^2 \sum\limits_{i=1}^{n} u_i^2 Var(z_i) + (\sum\limits_{i=1}^{n} u_i v_i)^2 \sum\limits_{i=1}^{n} v_i^2 Var(z_i) - 2(\sum\limits_{i=1}^{n} v_i^2)(\sum\limits_{i=1}^{n} u_i v_i) Cov(\sum\limits_{i=1}^{n} u_i z_i, \sum\limits_{i=1}^{n} v_i z_i)}{\left(\sum\limits_{i=1}^{n} u_i^2 \sum\limits_{i=1}^{n} v_i^2 - (\sum\limits_{i=1}^{n} u_i v_i)^2\right)^2}$$

$$= \frac{\sigma^2(\sum\limits_{i=1}^{n} v_i^2)^2 \sum\limits_{i=1}^{n} u_i^2 + \sigma^2(\sum\limits_{i=1}^{n} u_i v_i)^2 \sum\limits_{i=1}^{n} v_i^2 - 2(\sum\limits_{i=1}^{n} v_i^2)(\sum\limits_{i=1}^{n} u_i v_i) \sum\limits_{i=1}^{n} u_i v_i Var(z_i)}{\left(\sum\limits_{i=1}^{n} u_i^2 \sum\limits_{i=1}^{n} v_i^2 - (\sum\limits_{i=1}^{n} u_i v_i)^2\right)^2}$$

$$= \frac{\sigma^2 \sum_{i=1}^{n} v_i^2}{\sum_{i=1}^{n} u_i^2 \sum_{i=1}^{n} v_i^2 - (\sum_{i=1}^{n} u_i v_i)^2} = \frac{\sigma^2 \sum_{i=1}^{n} \rho_i^{-2} x_i^2}{\sum_{i=1}^{n} \rho_i^{-2} \sum_{i=1}^{n} \rho_i^{-2} x_i^2 - (\sum_{i=1}^{n} \rho_i^{-2} x_i)^2}$$

and:

$$Var(\hat{\beta}_1) = \frac{(\sum_{i=1}^{n} u_i^2)^2 \sum_{i=1}^{n} v_i^2 Var(z_i) + (\sum_{i=1}^{n} u_i v_i)^2 \sum_{i=1}^{n} u_i^2 Var(z_i) - 2(\sum_{i=1}^{n} u_i^2)(\sum_{i=1}^{n} u_i v_i) Cov(\sum_{i=1}^{n} v_i z_i, \sum_{i=1}^{n} u_i z_i)}{\left(\sum_{i=1}^{n} u_i^2 \sum_{i=1}^{n} v_i^2 - (\sum_{i=1}^{n} u_i v_i)^2\right)^2}$$

$$= \frac{\sigma^2(\sum_{i=1}^{n} u_i^2)^2 \sum_{i=1}^{n} v_i^2 + \sigma^2(\sum_{i=1}^{n} u_i v_i)^2 \sum_{i=1}^{n} u_i^2 - 2(\sum_{i=1}^{n} u_i^2)(\sum_{i=1}^{n} u_i v_i) \sum_{i=1}^{n} u_i v_i Var(z_i)}{\left(\sum_{i=1}^{n} u_i^2 \sum_{i=1}^{n} v_i^2 - (\sum_{i=1}^{n} u_i v_i)^2\right)^2}$$

$$= \frac{\sigma^2 \sum_{i=1}^{n} u_i^2}{\sum_{i=1}^{n} u_i^2 \sum_{i=1}^{n} v_i^2 - (\sum_{i=1}^{n} u_i v_i)^2} = \frac{\sigma^2 \sum_{i=1}^{n} \rho_i^{-2}}{\sum_{i=1}^{n} \rho_i^{-2} \sum_{i=1}^{n} \rho_i^{-2} x_i^2 - (\sum_{i=1}^{n} \rho_i^{-2} x_i)^2}$$

Problem 5

(a)

The original model is: $Y = X\beta + \varepsilon$, where $Y \in R^n, X \in R^{n\times p}, \beta \in R^p, \varepsilon \in R^n$, and ε_i's are independent with mean 0 and $Var(\varepsilon) = \Sigma = \sigma^2 \begin{bmatrix} \rho_1^2 & & \\ & \ddots & \\ & & \rho_n^2 \end{bmatrix} = \sigma^2 R^T R$, where $R = \begin{bmatrix} \rho_1 & & \\ & \ddots & \\ & & \rho_n \end{bmatrix}$

Now we transform the model into a one: $Z = V\beta + E$, where $Z = R^{-1}Y,\ V = R^{-1}X,\ E = R^{-1}\varepsilon$.

Since R is a constant matrix, thus Z is still a random variable matrix conditioning on given data design matrix V, and the random error matix now is E.

For the random error matix E, since $E(\varepsilon) = 0$ and $Var(\varepsilon) = \Sigma = \sigma^2 R^T R$, and ε_i's are independet, thus $E(E) = E(R^{-1}\varepsilon) = 0$ and $Var(E) = Var(R^{-1}\varepsilon) = R^{-1}\sigma^2 R^T R\ (R^{-1})^T = \sigma^2 I_n$, and e_i's (the components of E) are independent. Therefore, e_i's have mean zero and constant variance and are independent.

As a result, the new model $Z = V\beta + E$ satisfies the assumptions of the standard statistical model.

(b)

In the least squares estimation, we want to minimize: $S(\beta) = ||Z - V\beta||^2$

So we set the following derivative to zero:

$$\frac{\partial S}{\partial \beta} = \frac{\partial ||Z - V\beta||^2}{\partial \beta} = 0$$

Since:

$$||Z-V\beta||^2 = (Z-V\beta)^T(Z-V\beta) = (Z^T-\beta^T V^T)(Z-V\beta) = Z^T Z-\beta^T V^T Z-Z^T V\beta+\beta^T V^T V\beta = Z^T Z-2\beta^T V^T Z+\beta^T V^T V\beta$$

So:

$$\frac{\partial S}{\partial \beta} = \frac{\partial ||Z - V\beta||^2}{\partial \beta} = -2V^T Z + 2V^T V\beta = 0$$

Solving the equation, we obtain:

$$\hat{\beta} = (V^TV)^{-1}V^TZ$$

And since $\frac{\partial^2 S}{\partial \beta^2} = V^TV > 0$, so the $\hat{\beta} = (V^TV)^{-1}V^TZ$ is the minimizer for the score function $S(\beta)$.

Then plug in $V = R^{-1}X,\ Z = R^{-1}Y$, we get that:

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

(d)

Since $Var(Z) = Var(V\beta + E) = Var(E) = \sigma^2 I_n$, so we have:

$Var(\hat{\beta}) = Var((V^TV)^{-1}V^TZ) = (V^TV)^{-1}V^TVar(Z)((V^TV)^{-1}V^T)^T = \sigma^2(V^TV)^{-1}(V^TV)(V^TV)^{-1} = \sigma^2(V^TV)^{-1}$

Then plug in $V = R^{-1}X$, we get that: $Var(\hat{\beta}) = \sigma^2((R^{-1}X)^TR^{-1}X)^{-1} = \sigma^2(X^T(R^TR)^{-1}X)^{-1}$