STAT350 Assignment 2 Solution

Question 1

For simple linear regression, the variance of slope $\hat{\beta}_1$ is $Var(\hat{\beta}_1) = \frac{\sigma^2}{S_{xx}}$. Minimizing $Var(\hat{\beta}_1)$ is equivalent to maximizing $S_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2$, which implies each x_i should be as extreme as possible. One possible solution is to set half of x_i's to 0 and another half of x_i's to 1. For example, we can make $x_1 = x_2 = ... = x_5 = 0$ and $x_6 = x_7 = ... = x_{10} = 1$ to maximize $Var(\hat{\beta}_1)$.

Question 2

The joint likelihood is

$$L = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}(y - X\beta)^{'}(y - X\beta)\right).$$

Then the log-likelihood is

$$l = \log(L) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}(y - X\beta)^{'}(y - X\beta).$$

To get the maximum likelihood estimator for β, we set the derivative to 0 and solve

$$\begin{aligned}
\frac{\delta l}{\delta\beta} &= \frac{1}{2\sigma^2}\frac{\delta(y - X\beta)^{'}(y - X\beta)}{\delta\beta} \\
&= \frac{1}{2\sigma^2}\frac{\delta(y^{'}y - y^{'}X\beta - \beta^{'}X^{'}y - \beta^{'}X^{'}X\beta)}{\delta\beta} \\
&= \frac{1}{2\sigma^2}\frac{\delta(y^{'}y - 2\beta^{'}X^{'}y - \beta^{'}X^{'}X\beta)}{\delta\beta} \\
&= \frac{1}{2\sigma^2}(-2X^{'}y + 2X^{'}X\beta) \\
&= 0.
\end{aligned}$$

So $\hat{\beta} = (X^{'}X)^{-1}X^{'}y$.

Question 3

(a)

Given that $\hat{y} = X\hat{\beta}$ and $Var(\hat{\beta}) = \sigma^2(X^{'}X)^{-1}$, then

$$\begin{aligned}
Var(\hat{y}) &= Var(X\hat{\beta}) \\
&= XVar(\hat{\beta})X^{'} \\
&= \sigma^2X(X^{'}X)^{-1}X^{'} \\
&= \sigma^2H.
\end{aligned}$$

(b)

To prove that $\hat{y}_0$ is an unbiased estimator of $E(y|x_0)$, we have to show that $E(\hat{y}_0) = E(y|x_0)$:

$$\begin{aligned} E(y|x_0) &= E(x_0^{'}\beta + \epsilon) \\ &= E(x_0^{'}\beta) + E(\epsilon) \\ &= x_0^{'}\beta \text{ (because } E(\epsilon) = 0). \end{aligned}$$

Then

$$\begin{aligned} E(\hat{y}_0) &= E(x_0^{'}\hat{\beta}) \\ &= x_0^{'}E(\hat{\beta}) \\ &= x_0^{'}\beta \text{ (because } \hat{\beta} \text{ is an unbiased estimator of } \beta) \\ &= E(y|x_0). \end{aligned}$$

Question 4

(a)

```r
x1=rnorm(200,mean=0,sd=2)
x2=rnorm(200,mean=0,sd=2)
err=rnorm(200,mean=0,sd=1)
y=1+2*x1+5*x2+err
model=lm(y~x1+x2)
summary(model)
```

```
##
## Call:
## lm(formula = y ~ x1 + x2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.55273 -0.78248 -0.05568  0.76019  2.60427
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.91690    0.07594   12.07   <2e-16 ***
## x1           2.02447    0.03803   53.24   <2e-16 ***
## x2           5.00816    0.03648  137.29   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.071 on 197 degrees of freedom
## Multiple R-squared:  0.9905, Adjusted R-squared:  0.9904
## F-statistic: 1.026e+04 on 2 and 197 DF,  p-value: < 2.2e-16
```

```r
paste0("The estimated line is ",round(summary(model)$coef[1,1],3),"+",round(summary(model)$coef[2,1],3)
```

```
## [1] "The estimated line is 0.917+2.024x1+5.008x2"
```

###(b)

```r
round(vcov(model),10)
```

```
##                (Intercept)            x1            x2
## (Intercept)  0.0057670105 -0.0001798538 -0.0001019904
## x1          -0.0001798538  0.0014460838  0.0001262196
## x2          -0.0001019904  0.0001262196  0.0013307566
```

The theoretical values for the covariance matrix of the regression coefficients is $\sigma^2(X^{'}X)^{-1}$.

```
sigma=1
X=cbind(rep(1,200),x1,x2)
round(sigma^2*solve(t(X)%*%X),10)
```

```
##                            x1            x2
##     0.0050244043 -0.0001566944 -0.0000888573
## x1 -0.0001566944  0.0012598745  0.0001099666
## x2 -0.0000888573  0.0001099666  0.0011593977
```

(c)

```
summary(model)
```

```
##
## Call:
## lm(formula = y ~ x1 + x2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.55273 -0.78248 -0.05568  0.76019  2.60427
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.91690    0.07594   12.07   <2e-16 ***
## x1           2.02447    0.03803   53.24   <2e-16 ***
## x2           5.00816    0.03648  137.29   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.071 on 197 degrees of freedom
## Multiple R-squared:  0.9905, Adjusted R-squared:  0.9904
## F-statistic: 1.026e+04 on 2 and 197 DF,  p-value: < 2.2e-16
```

The t-statistic is 52.87 and the p-value is less than 0.05, so we reject the null hypothesis that $\beta_1 = 0$.

(d)

```
rej=NULL
for (i in 1:1000) {
  x1=rnorm(200,mean=0,sd=2)
  x2=rnorm(200,mean=0,sd=2)
  err=rnorm(200,mean=0,sd=1)
  y=1+2*x1+5*x2+err
  rej[i]=summary(model)$coef[2,4]<=0.05
}
sum(rej)/1000
```

```
## [1] 1
```

The null hypothesis is always rejected.