

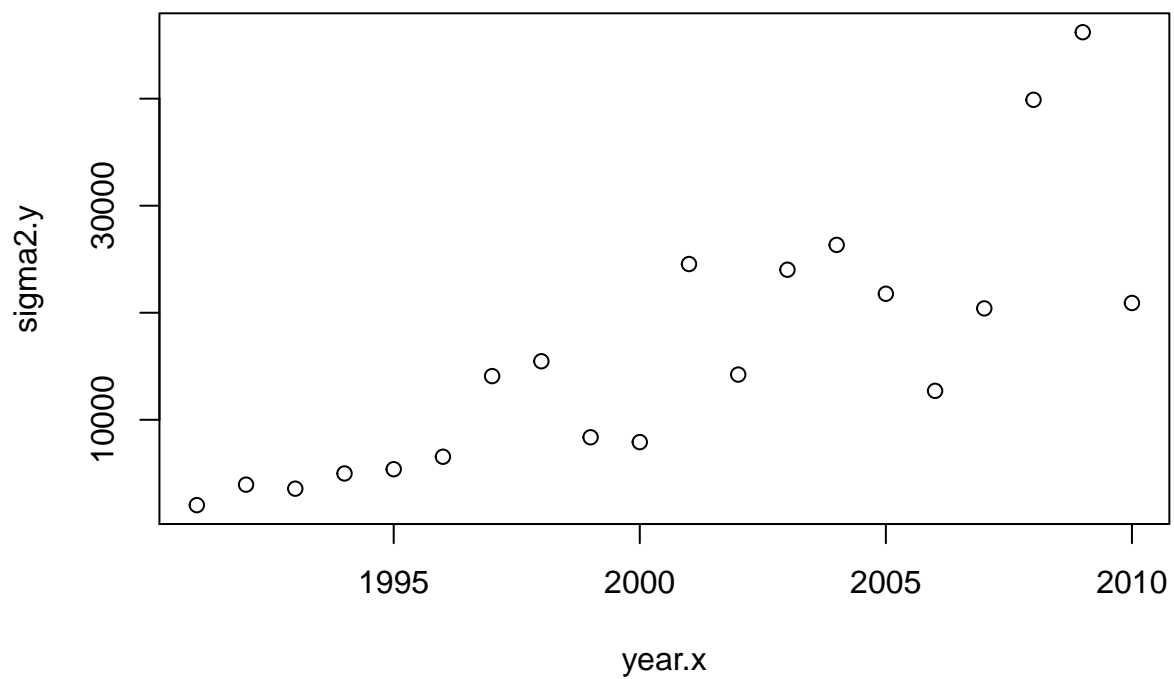
a1q3

Mushi Wang

02/06/2020

q3 (a)

```
sales = read.table("../JaxSales.txt", header = TRUE)
year.x = seq(1991, 2010)
sigma2.y = vector(length = 20)
for (i in year.x) {
  sigma2.y[i + 1 - 1991] = var(sales$Sales[which(sales$Year == i)])
}
plot(year.x, sigma2.y)
```



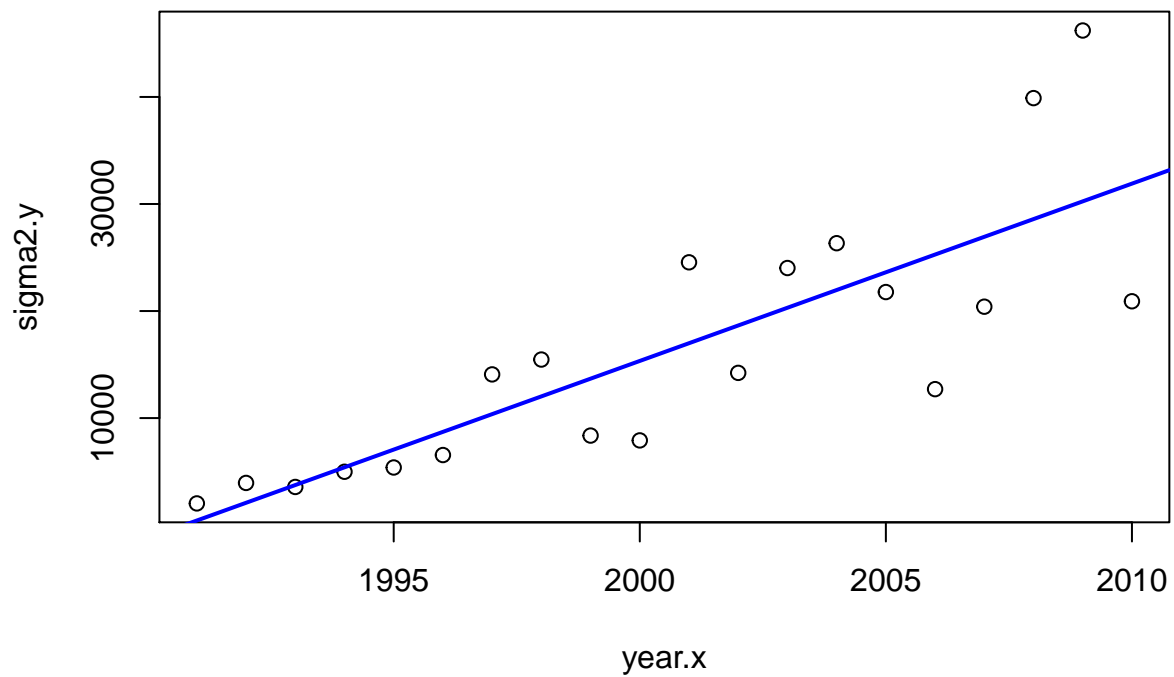
(b)

```
varModel = lm(sigma2.y ~ year.x)
varModel$coefficients
```

```
## (Intercept)      year.x
## -3297581.598    1656.459
```

$\hat{\alpha}_0 = -3297581.598, \hat{\alpha}_1 = 1656.459$

```
plot(year.x, sigma2.y)
abline(lm(sigma2.y ~ year.x), col="blue", lwd = 2)
```

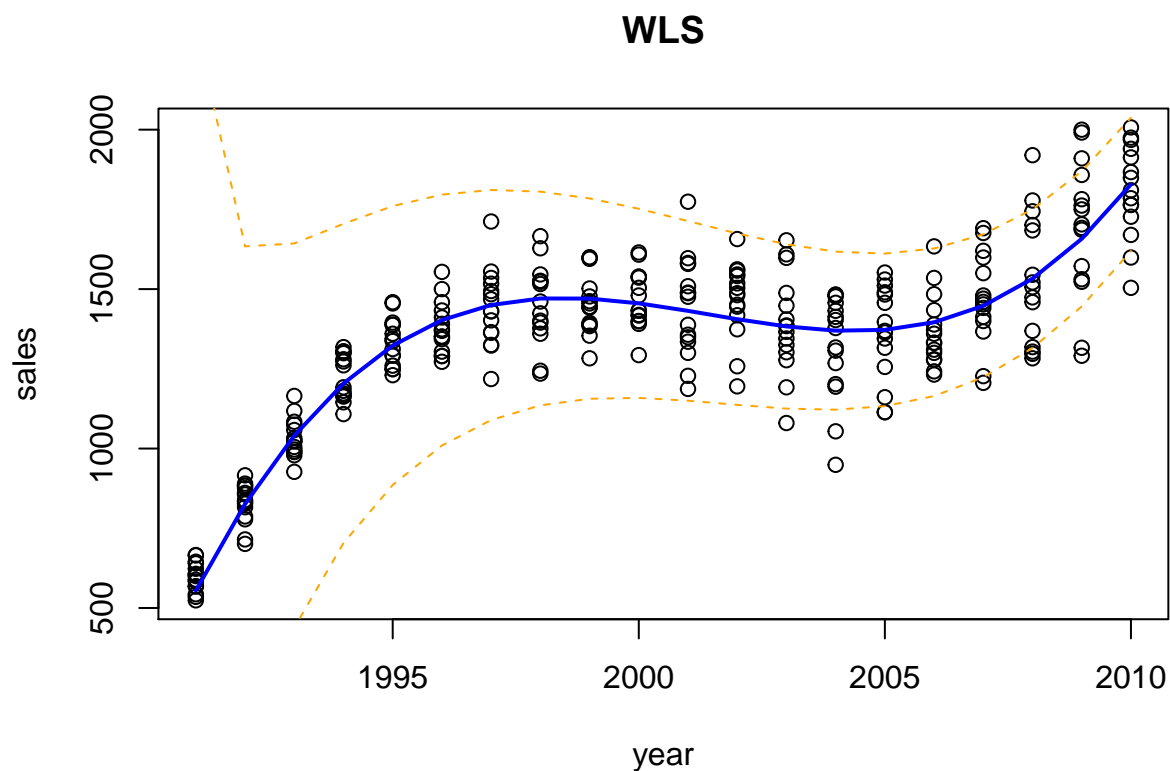


(c)(1)

```
w1 = vector(length = length(sales$Year))
w1 = varModel$coefficients[1] + varModel$coefficients[2] * sales$Year
wls1 = lm(sales$Sales~poly(sales$Year, 3), weight = w1)

plot(sales$Year, sales$Sales, xlab = "year", ylab = "sales", main = "WLS")
lines(sales$Year, predict(wls1), type="l", col="blue", lwd=2)

pred_interval1 <- predict(wls1, newdata = data.frame(year=sales$Year),
                          interval="prediction", level = 0.95, weight = w1)
lines(sales$Year, pred_interval1[,2], col="orange", lty=2)
lines(sales$Year, pred_interval1[,3], col="orange", lty=2)
```



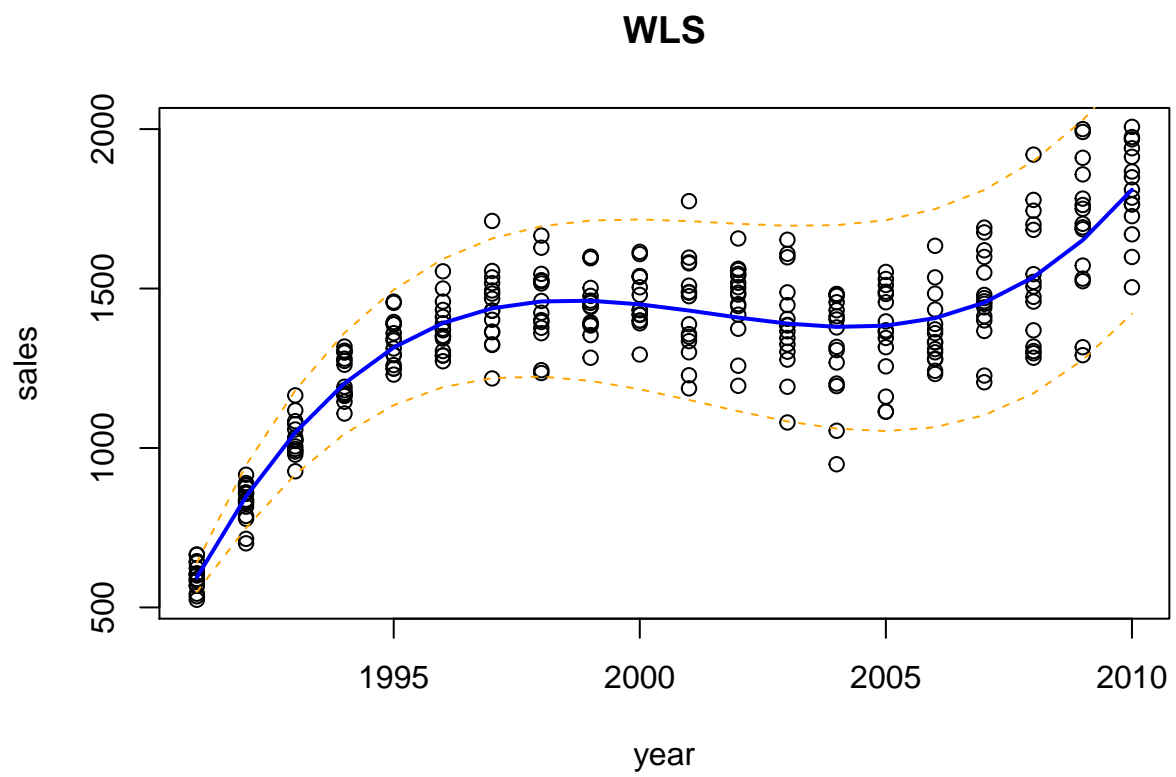
(2)

```
w2 = 1 / w1
wls2 = lm(sales$Sales~poly(sales$Year, 3), weight = w2)

plot(sales$Year, sales$Sales, xlab = "year", ylab = "sales", main = "WLS")
lines(sales$Year, predict(wls2), type="l", col="blue", lwd=2)

pred_interval2 <- predict(wls2, newdata = data.frame(year=sales$Year),
                          interval="prediction", level = 0.95, weight = w2)
lines(sales$Year, pred_interval2[,2], col="orange", lty=2)
```

```
lines(sales$Year, pred_interval2[,3], col="orange", lty=2)
```



(d)

In the first model, since $\hat{\alpha}_1$ is positive, the weight increases as year increases. So the greastest influence points are when $x = 2010$. So the least influence points are when $x = 1991$.

In the second model, since $\hat{\alpha}_1$ is positive and $w = \frac{1}{\sigma^2(x)}$. The weight decreases as year increases. So the greastest influence points are when $x = 1991$. So the least influence points are when $x = 2010$.

We want to give greater weight to the points that has low variance($x = 1991$) and less weight to the points that has high variance $x = 2010$. So we prefer model 2.