# NTHU STAT 5410 - Linear Models Assignment 5 Report

105061110 周柏宇

1.

(a) Since , where and . The weighted least squares estimator is and its variance is . The reason why we should use WLS rather than OLS is that the variance of violets the assumption of equal variance from the Gauss-Markov theorem, which guarantees the variance of will be the smallest among all the linear unbiased estimators. However, we can fix the issue by fitting a transformed version of the original model , where . Then for the transformed model is exactly the for the original model.

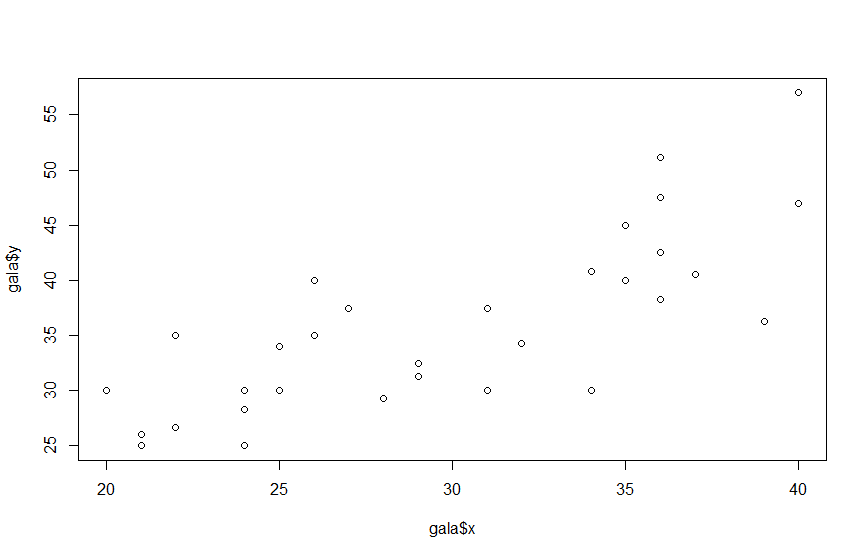
(b) In this case, , where and . The best linear unbiased estimator (BLUE) of β is and its variance is .

2.

(a)

> gala <- read.table("C:/Users/Thomas/Downloads/Linear\_models/hw5/E6.10.txt", header=T)

> plot(gala$x, gala$y)



Despite that the responses are averaged, we still can get a feeling of unequal variance among the observations due to their duplicates. If we assume the raw responses (i.e. ) are of constant variance . Then the covariance matrix for the sample mean will be

, where and . Therefore, .

> n <- gala[,2]

> x <- gala[,3]

> y <- gala[,4]

> WLS <- lm(y ~ x, weights=n)

> summary(WLS)

Call:

lm(formula = y ~ x, weights = n)

Weighted Residuals:

Min 1Q Median 3Q Max

-20.278 -7.661 -0.680 4.543 33.219

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 2.2932 4.5903 0.500 0.621

x 1.1319 0.1475 7.676 1.46e-08 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 10.01 on 30 degrees of freedom

Multiple R-squared: 0.6626, Adjusted R-squared: 0.6514

F-statistic: 58.93 on 1 and 30 DF, p-value: 1.458e-08

(b) OLS estimators

> OLS <- lm(y ~ x)

> summary(OLS)

Call:

lm(formula = y ~ x)

Residuals:

Min 1Q Median 3Q Max

-10.2223 -3.0279 -0.6581 3.7721 10.7442

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 6.0325 4.5137 1.336 0.191

x 1.0056 0.1475 6.819 1.45e-07 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.07 on 30 degrees of freedom

Multiple R-squared: 0.6079, Adjusted R-squared: 0.5948

F-statistic: 46.5 on 1 and 30 DF, p-value: 1.449e-07

Since the summary in WLS shows the weighted residuals as well as weighted RSE and weighted R-squared, we cannot compare the statistics without directly. We need to use

> WLS.RSS <- sum(WLS$residuals^2)

> WLS.RSS

[1] 790.0211

> OLS.RSS <- sum(OLS$residuals^2)

> OLS.RSS

[1] 771.0631

The RSS for OLS is indeed smaller since it minimize the rather than .

> plot(n, WLS$residuals, pch=1)

> legend("bottomleft", legend = c("WLS"), pch = c(1), lty = c(1))

> abline(a=0, b=0)

> plot(n, OLS$residuals, pch=4)

> legend("bottomleft", legend = c("WLS", "OLS"), pch = c(1, 4), lty = c(1, 1))

> abline(a=0, b=0)

(c)

> length(unique(x))

[1] 17

> length(x)

[1] 32

We can see that with 32 observations there are only 17 unique x. As the plot shows, we have several duplicates on some x.

3.