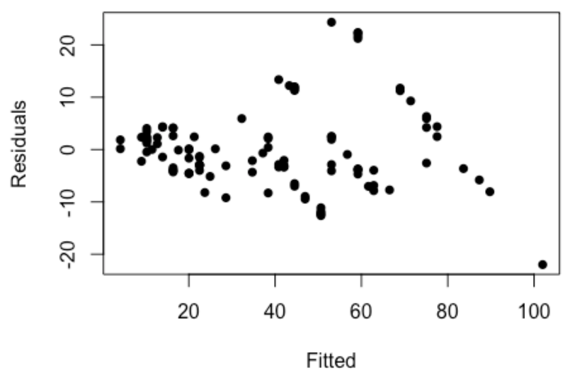
HW#5, Nan Deng

(a)

```r
library(faraway)
data(pipeline)
pipeline_fit <- lm(Lab ~ Field, data=pipeline)
plot(fitted(pipeline_fit),resid(pipeline_fit), xlab="Fitted", ylab="Res
iduals",pch=16)
```



According to the residual distribution against y, it shows Heteroscedasticity (non-constant variance). Although the variance keeps concentrated at the head, it starts to become seperate when x gets greater.

(b)

$$\log(varlab) = \log\hat{a}_0 + \hat{a}_1 \log(meanfield)$$

$$\log(varlab) = \log(\hat{a}_0 \times meanfield^{\hat{a}_1})$$

$$varlab = \hat{a}_0 \times meanfield^{\hat{a}_1}$$

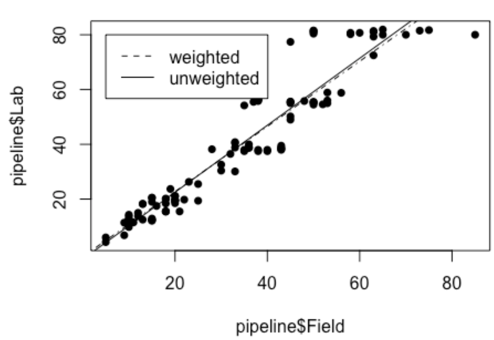
$$w = field^{-\hat{a}_1}/\hat{a}_0 = x^{-\hat{a}_1}/\hat{a}_0$$

```r
i <- order(pipeline$Field)
npipe <- pipeline[i,]
ff <- gl(12,9)[-108]
meanfield <- unlist(lapply(split(npipe$Field,ff),mean))
varlab <- unlist(lapply(split(npipe$Lab,ff),var))
log_fit <- lm(I(log(varlab)) ~ I(log(meanfield)), data=data.frame(varla
b,meanfield)[-length(varlab),])
summary(log_fit)

##
## Call:
## lm(formula = I(log(varlab)) ~ I(log(meanfield)), data = data.frame(v
arlab,
```

```
##     meanfield)[-length(varlab), ])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.00477 -0.42268  0.05989  0.37854  0.93815
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)         -1.9352     1.0929  -1.771 0.110403
## I(log(meanfield))    1.6707     0.3296   5.070 0.000672 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.657 on 9 degrees of freedom
## Multiple R-squared:  0.7406, Adjusted R-squared:  0.7118
## F-statistic:  25.7 on 1 and 9 DF,  p-value: 0.0006723

a0 <- exp(log_fit$coefficients[1])
a1 <- log_fit$coefficients[2]
paste("a0=",a0)

## [1] "a0= 0.144400092461675"

paste("a1=",a1)

## [1] "a1= 1.67072344424889"
```

(c)

```
pipeline$w <- pipeline$Field^(-a1)/a0
pipeline_fit_1 <- lm(Lab ~ Field,weights=pipeline$w,data=pipeline)
plot(pipeline$Field, pipeline$Lab, pch=16)
abline(pipeline_fit)
abline(pipeline_fit_1,lty=4)
legend(5,80,legend=c('weighted','unweighted'),lty=c(2,1))
```



(d)

```
sum_ei <- sum(resid(pipeline_fit_1))
sum_wiei <- sum(resid(pipeline_fit_1)*pipeline$wi)
sum_ei

## [1] 22.18607

sum_wiei

## [1] 0
```

The residuals do not sum to zero, while the sum of wiei does. Considering the model pipeline_fit_1 is impacted by the weight, the true ei of this model should also incorporate wi, which is wiei.

(e)

```
pipeline_fit_2 = lm(Lab ~ 1, weight = w, data=pipeline)
anova(pipeline_fit_2,pipeline_fit_1)

## Analysis of Variance Table
##
## Model 1: Lab ~ 1
## Model 2: Lab ~ Field
##   Res.Df     RSS Df Sum of Sq      F    Pr(>F)
## 1    106 1288.19
## 2    105  101.79  1    1186.4 1223.8 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

y_bar <- coef(pipeline_fit_2)
y_bar

## (Intercept)
##    18.32086

SSR <- sum(pipeline$w*(fitted(pipeline_fit_1)-y_bar)^2)
SSR

## [1] 1186.407

SST <- sum(pipeline$w*(pipeline$Lab-y_bar)^2)
SST

## [1] 1288.195
```

Since 1288.19(SST)=1186.407(SSR)+101.79(SSR), SST is equal to the sumation of SSE and SSR.