Weighted Least Squares (WLS)

In this example (from Neter and Wasserman) n = 12 observations relate the preparation Cost (Y) of a bid to the Size (X) of the bid.

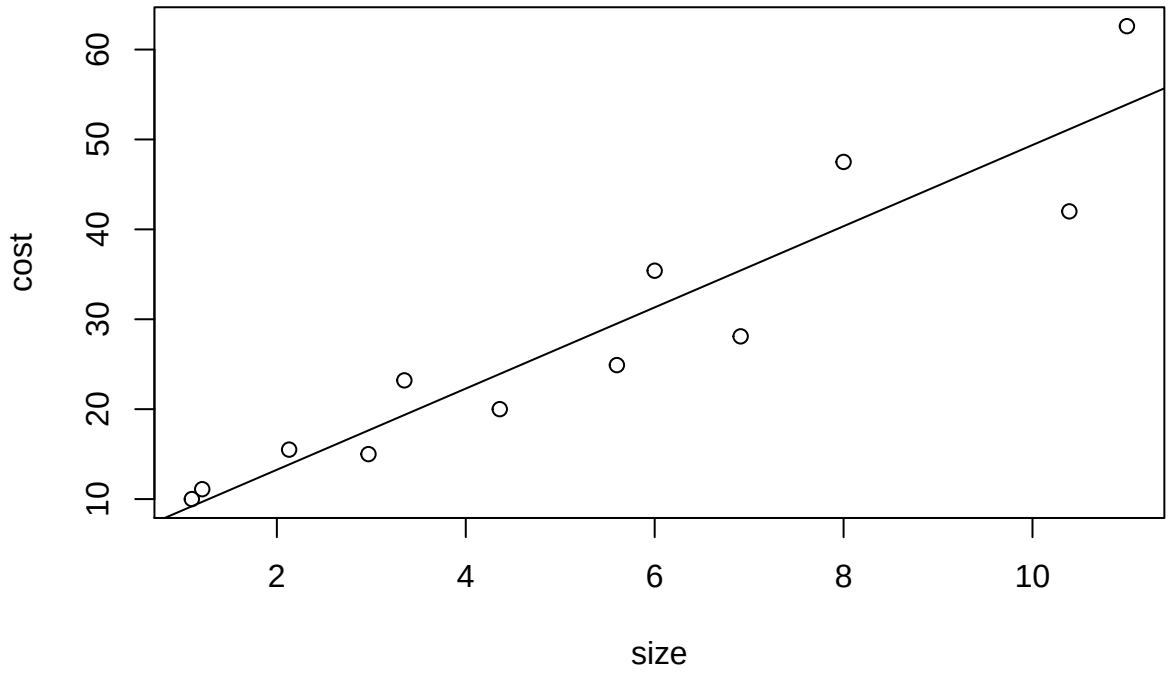
Based on the residual diagnostic plot, there is evidence of increasing variance with X. Instead of transforming, WLS is used.

Other approaches to consider include iteratively reweighted least squares (IRWLS) or robust regression.

```
InData <- read.csv("C:/hess/STAT512/RNotes/ExtraTopics/WLSdata.csv")
str(InData)
```

```
## 'data.frame':    12 obs. of  2 variables:
##  $ size: num  2.13 1.21 11 6 5.6 ...
##  $ cost: num  15.5 11.1 62.6 35.4 24.9 28.1 15 23.2 42 10 ...
```

```
plot(cost ~ size, data = InData)
abline(lm(cost ~ size, data = InData))
```

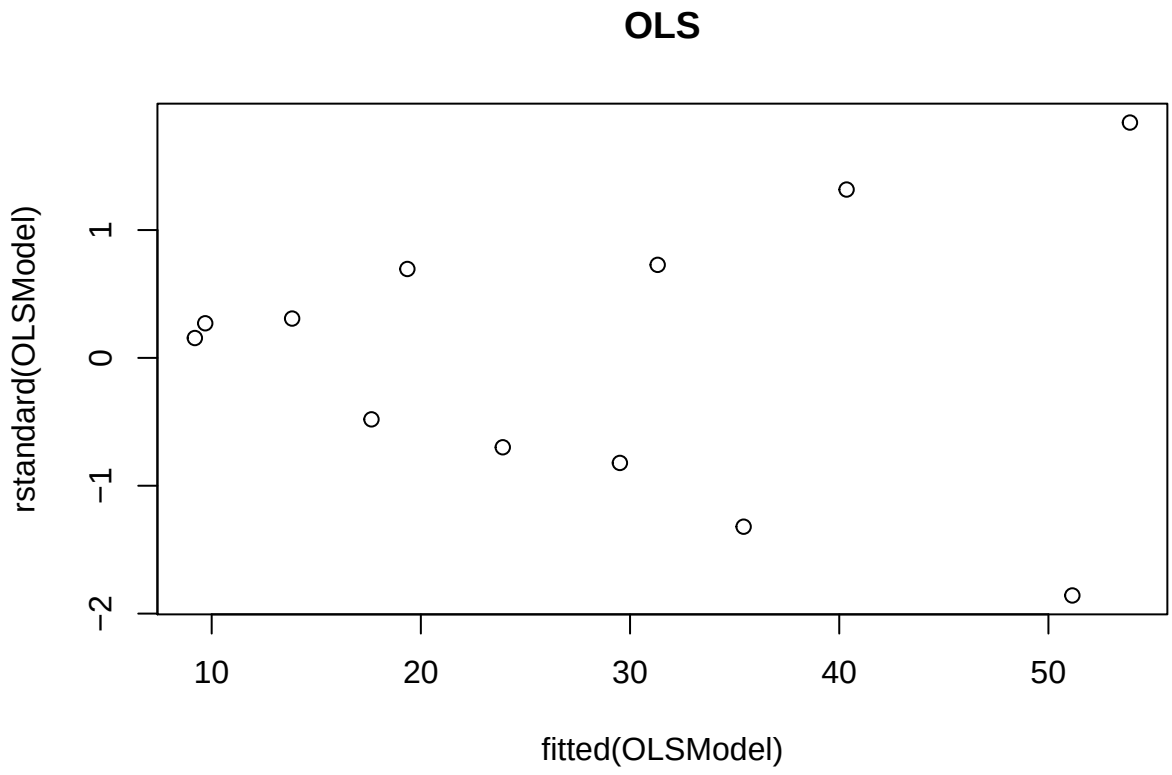


Ordinary Least Squares (OLS)

```
OLSModel <- lm(cost ~ size, data = InData)
summary(OLSModel)
```

```
##
## Call:
## lm(formula = cost ~ size, data = InData)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -9.143 -4.090  1.106  3.904  8.703
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   4.2289     3.2517   1.301    0.223
## size          4.5153     0.5285   8.544 6.59e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.87 on 10 degrees of freedom
## Multiple R-squared:  0.8795, Adjusted R-squared:  0.8675
## F-statistic:    73 on 1 and 10 DF,  p-value: 6.588e-06
```

```
plot(rstandard(OLSModel) ~ fitted(OLSModel), main = "OLS")
```

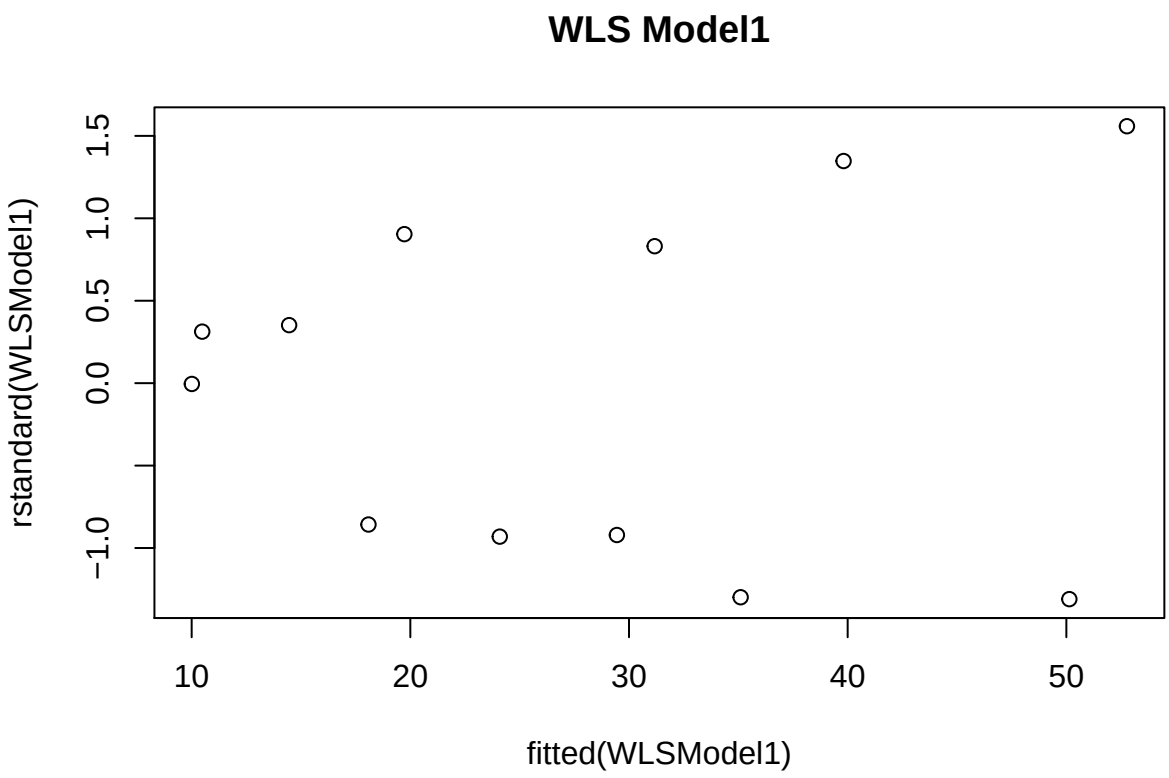


Weighted Least Squares

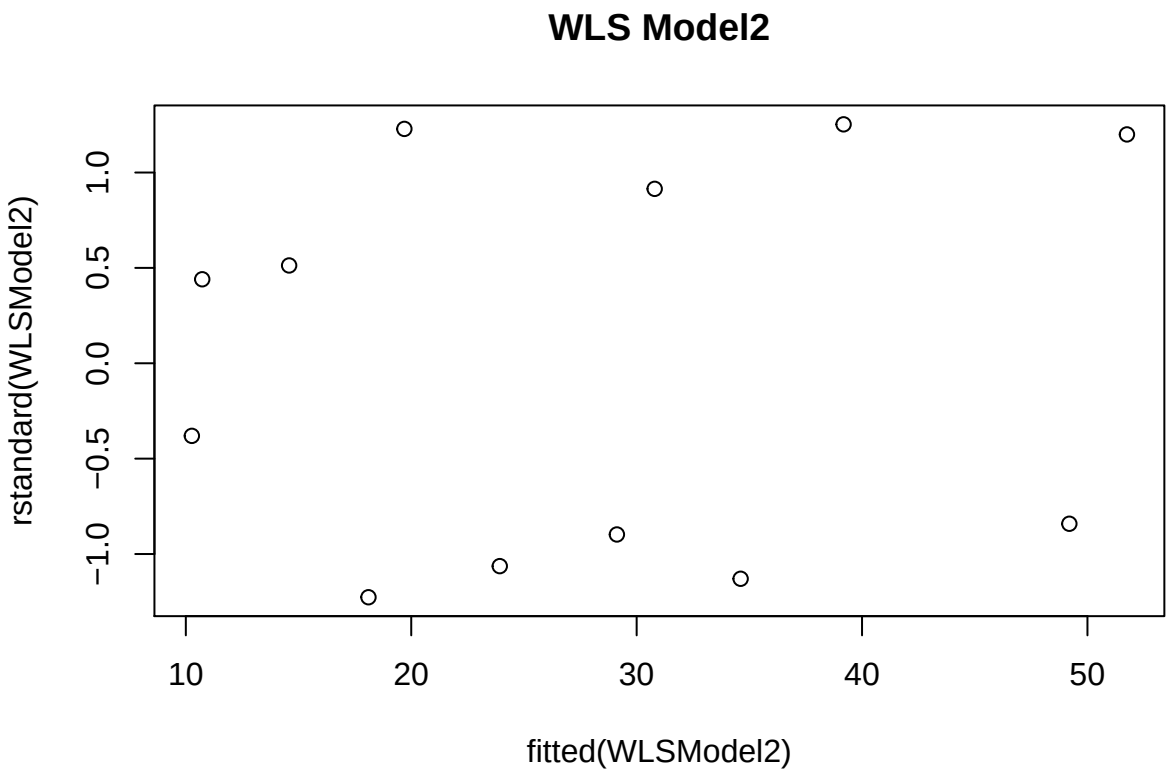
We try two different weights (1/cost and 1/cost^2). Note that we consider plots of the standardized residuals (not the raw residuals). The first weighted model still shows unequal variance. But the second weighted

model looks good!

```
InData$W1 <- 1/(InData$size)
InData$W2 <- 1/(InData$size*InData$size)
WLSModel1 <- lm(cost ~ size, weights = W1, data = InData)
plot(rstandard(WLSModel1) ~ fitted(WLSModel1), main = "WLS Model1")
```




```
WLSModel2 <- lm(cost ~ size, weights = W2, data = InData)
plot(rstandard(WLSModel2) ~ fitted(WLSModel2), main = "WLS Model2")
```




```
summary(WLSModel2)
```

```
##
## Call:
## lm(formula = cost ~ size, data = InData, weights = W2)
##
## Weighted Residuals:
##      Min       1Q   Median       3Q      Max
## -1.04471 -0.79092  0.03284  0.82150  1.04621
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   5.6569     0.9652   5.861 0.000159 ***
## size          4.1906     0.4037  10.381 1.13e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8897 on 10 degrees of freedom
## Multiple R-squared:  0.9151, Adjusted R-squared:  0.9066
## F-statistic: 107.8 on 1 and 10 DF,  p-value: 1.127e-06
```