

hw2q28

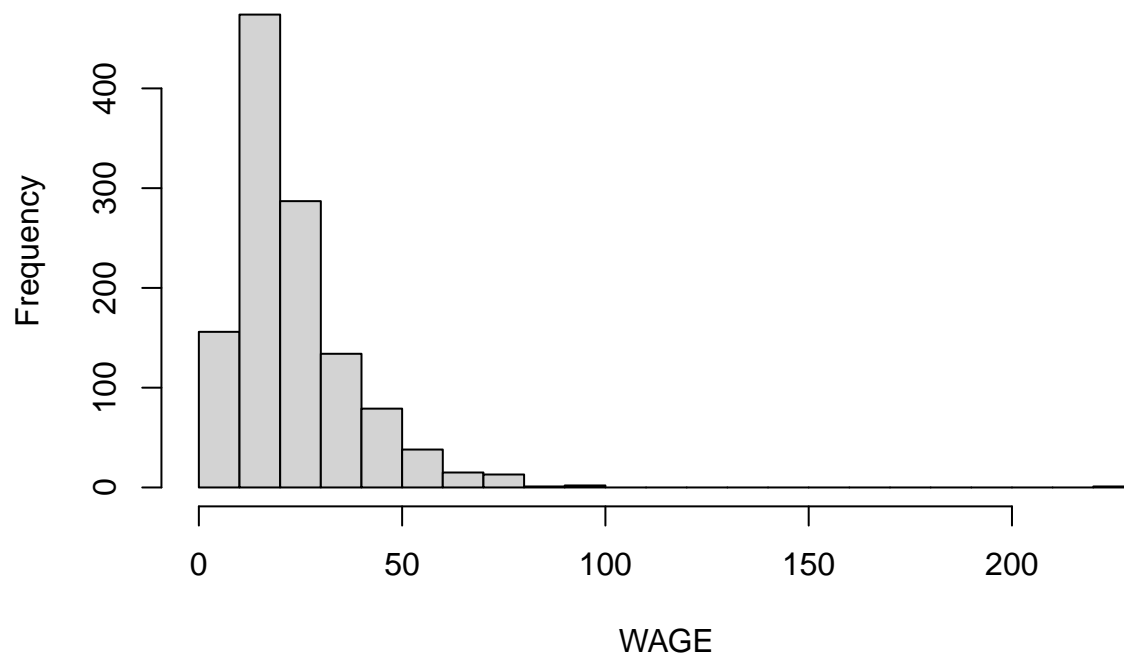
313707025 jebuhdah

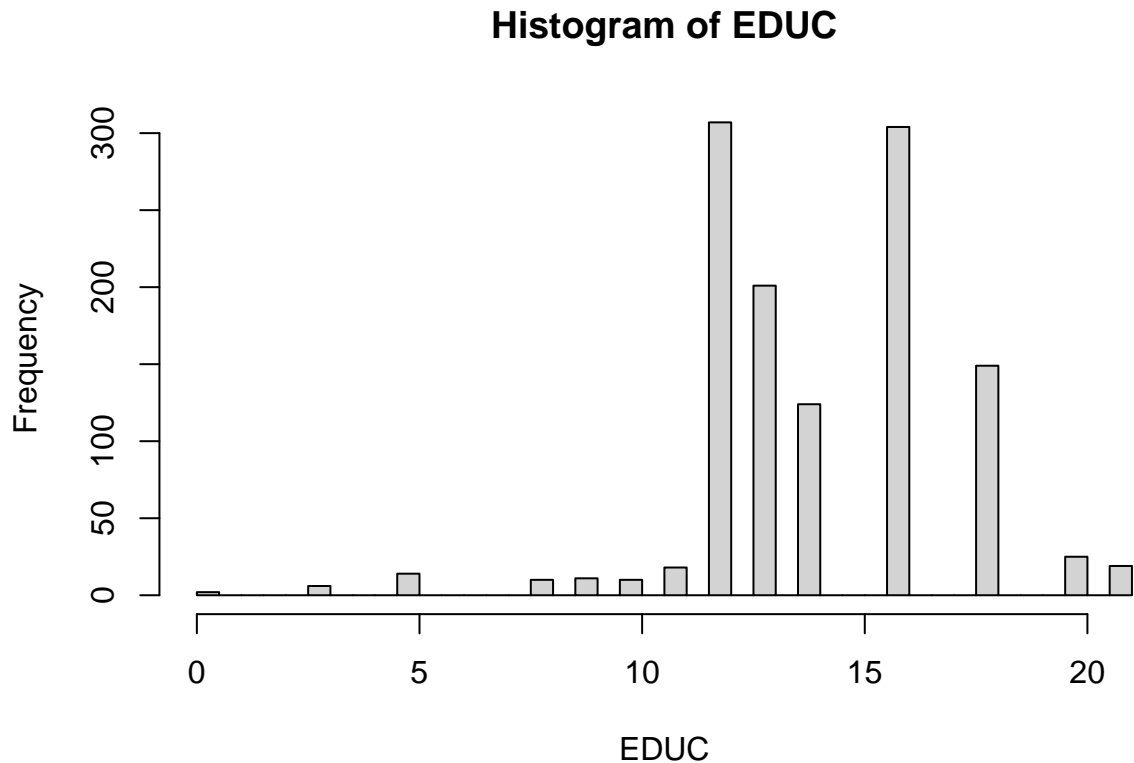
2025-03-10

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	3.94	13.00	19.30	23.64	29.80	221.10

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	0.0	12.0	14.0	14.2	16.0	21.0

Histogram of WAGE

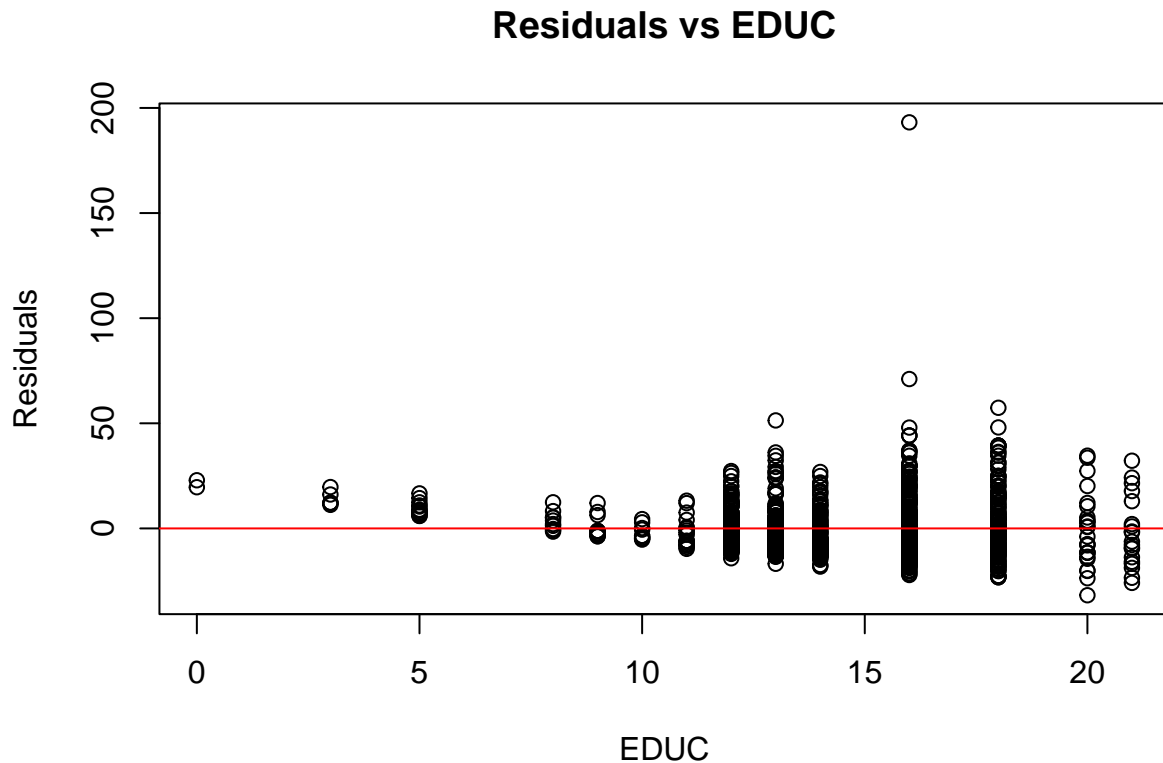




a: there is a left skew for wage and right skew for educ. there is a decline before 15 years and after 15 years, assuming that's after high school and uni which individuals seek jobs?

```
##
## Call:
## lm(formula = wage ~ educ, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -31.785  -8.381  -3.166   5.708  193.152
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -10.4000     1.9624   -5.3 1.38e-07 ***
## educ         2.3968     0.1354   17.7 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.55 on 1198 degrees of freedom
## Multiple R-squared:  0.2073, Adjusted R-squared:  0.2067
## F-statistic: 313.3 on 1 and 1198 DF, p-value: < 2.2e-16
```

b: the beta is 2.40, so every increase in education year increase income by 2.40.



c: it seems like there is a pattern and it doesn't seem random, thus violating homoskedasticity.

```
##
## Call:
## lm(formula = wage ~ educ, data = male_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -27.643  -9.279  -2.957   5.663  191.329
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -8.2849     2.6738  -3.099  0.00203 **
## educ           2.3785     0.1881  12.648 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 14.71 on 670 degrees of freedom
## Multiple R-squared:  0.1927, Adjusted R-squared:  0.1915
## F-statistic: 160 on 1 and 670 DF, p-value: < 2.2e-16

##
## Call:
## lm(formula = wage ~ educ, data = female_data)
##
## Residuals:
```

```

##      Min      1Q  Median      3Q      Max
## -30.837  -6.971  -2.811   5.102  49.502
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -16.6028      2.7837  -5.964 4.51e-09 ***
## educ         2.6595       0.1876  14.174 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.5 on 526 degrees of freedom
## Multiple R-squared:  0.2764, Adjusted R-squared:  0.275
## F-statistic: 200.9 on 1 and 526 DF,  p-value: < 2.2e-16

##
## Call:
## lm(formula = wage ~ educ, data = black_data)
##
## Residuals:
##      Min      1Q  Median      3Q      Max
## -15.673  -6.719  -2.673   4.321  40.381
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -6.2541      5.5539  -1.126   0.263
## educ          1.9233      0.3983   4.829 4.79e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.51 on 103 degrees of freedom
## Multiple R-squared:  0.1846, Adjusted R-squared:  0.1767
## F-statistic: 23.32 on 1 and 103 DF,  p-value: 4.788e-06

##
## Call:
## lm(formula = wage ~ educ, data = white_data)
##
## Residuals:
##      Min      1Q  Median      3Q      Max
## -32.131  -8.539  -3.119   5.960 192.890
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -10.475      2.081  -5.034 5.6e-07 ***
## educ         2.418       0.143  16.902 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.79 on 1093 degrees of freedom
## Multiple R-squared:  0.2072, Adjusted R-squared:  0.2065
## F-statistic: 285.7 on 1 and 1093 DF,  p-value: < 2.2e-16

```

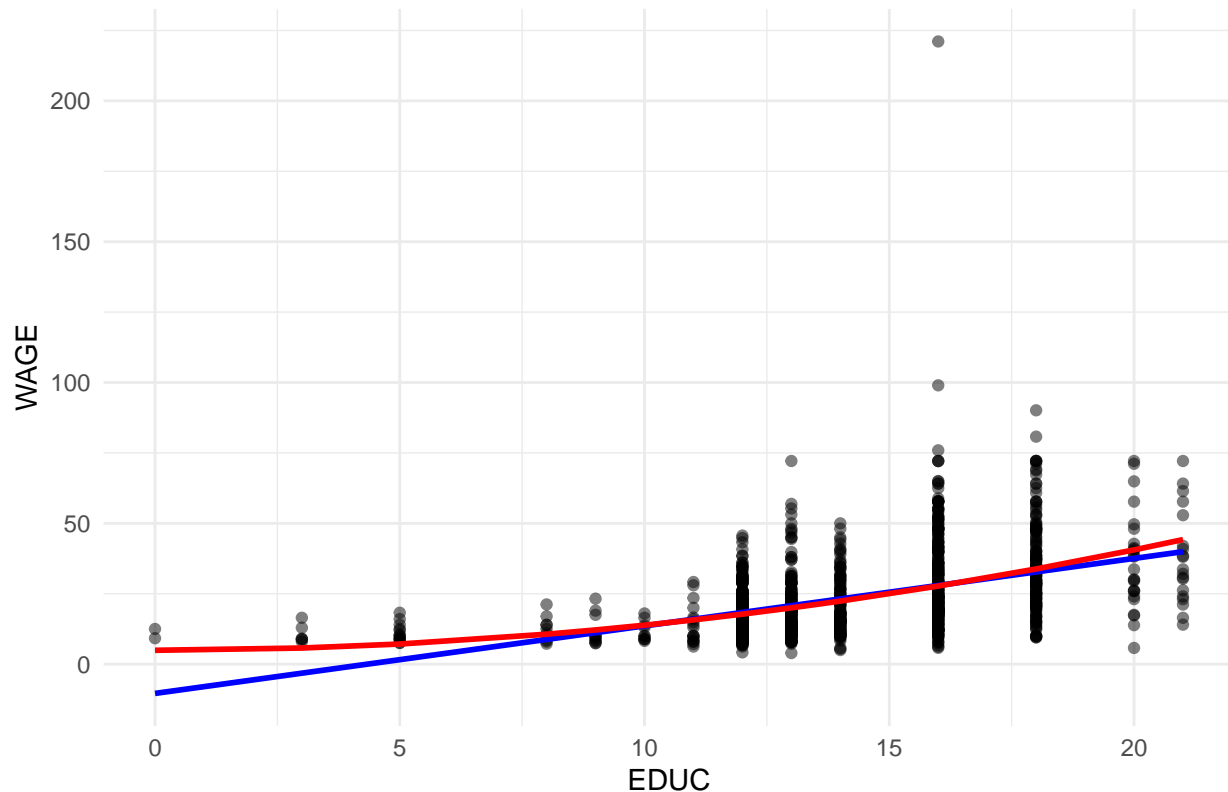
d: analyzing on beta educ, it seems like the male female comparasion only differents by minor margin while showing that females has higher growth rates with levels of educaution, while female have lower starting

wage than males. but in terms of white black comparasion, whites have much higher increase in wage when education level increases.

```
##
## Call:
## lm(formula = wage ~ educ_squared, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -34.820  -8.117  -2.752   5.248  193.365
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.916477   1.091864   4.503 7.36e-06 ***
## educ_squared 0.089134   0.004858  18.347 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.45 on 1198 degrees of freedom
## Multiple R-squared:  0.2194, Adjusted R-squared:  0.2187
## F-statistic: 336.6 on 1 and 1198 DF,  p-value: < 2.2e-16
```

e: marginal effect = $\text{wage} / \text{educ} = 2(a_2)\text{educ} = 2(0.09)\text{educ}$ which we have 12yr=2.16, and 16yr=2.88
marginal effect of the linear model would be $\text{wage} / \text{educ} = a_2 = 2.40$. so in comparasion, it tells us that
quac model have smaller me 12 years of educ, and larger me for 16 years of educ than linear model.

Fitted Linear and Quadratic Models



f: it seems like quadratic model would fit better for the data provided.