Diagnostics Assignment Marvin

Marvin Williams

12/5/2020

```r
model1 <- lm(cpoverty ~ unempl.y + ben.ink, data =  data)
```

Model

The model I've constructed looks at the percentage of the population aged 15-29 who are neither unemployed nor in any educational/training program, as well as public expenditure on services and in-kind benefits for families as a % of GDP, as predictors of the child poverty rate. I chose these 2 predictor variables as I figured unemployment rates for the youngest working population could give a sense of where things are financially for the country, and in turn, child poverty rates. Also with public expenditure, government spending on services and in-kind benefits for families would be assumed to have a substantial impact on the likelihood of a child living in poverty.

```r
summary(model1)
```

```
##
## Call:
## lm(formula = cpoverty ~ unempl.y + ben.ink, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.9975 -1.8289  0.1968  2.1072  4.2318
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   7.8092     2.3432   3.333  0.00422 **
## unempl.y      0.4575     0.1332   3.435  0.00340 **
## ben.ink      -2.9230     1.2328  -2.371  0.03063 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.847 on 16 degrees of freedom
##   (15 observations deleted due to missingness)
## Multiple R-squared:  0.5447, Adjusted R-squared:  0.4878
## F-statistic: 9.572 on 2 and 16 DF,  p-value: 0.001846
```

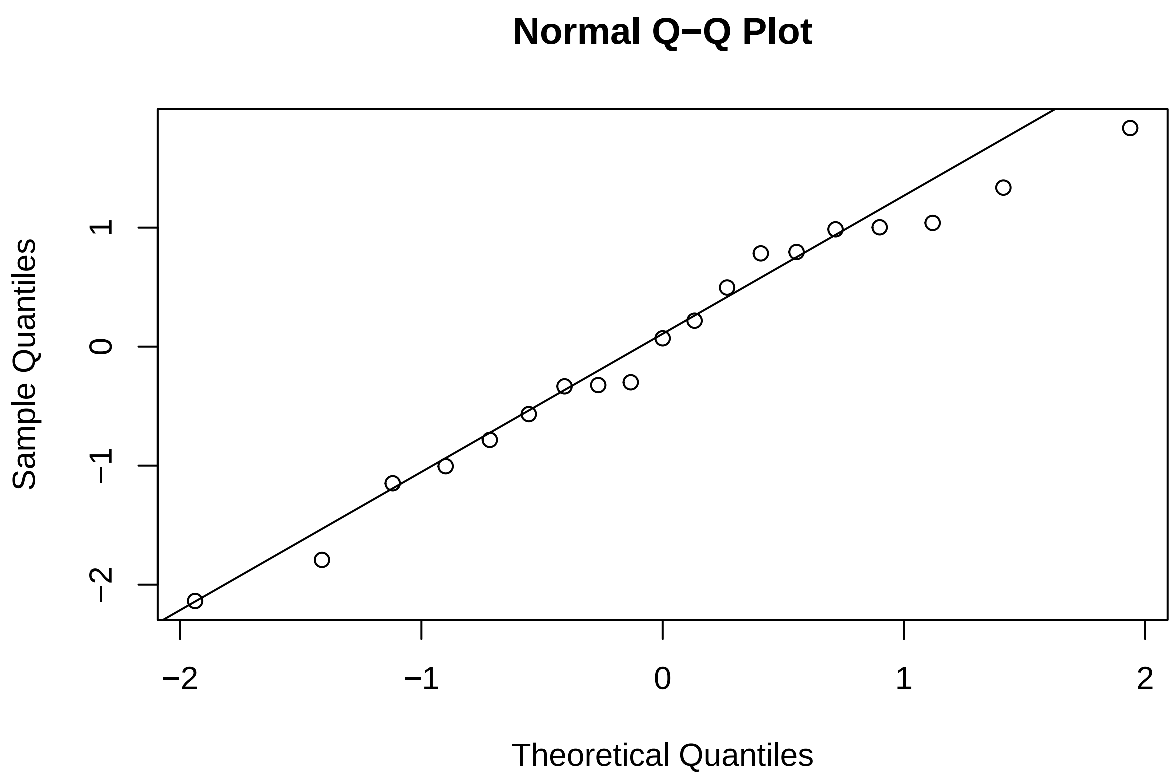
The result of the model shows both variables to be significant predictors of child poverty rates at the .05 alpha level. With a coefficient of 0.4575, this tells us that unemployment is positively correlated with the child poverty rate. With every unit increase in unemployment for ages 15-29-year-olds, we can expect child an increase in the child poverty rate of 0.4575. For public expenditure on services and in-kind benefits, with a coefficient of -2.9230, we can observe a negative correlation with the child poverty rate. For every unit increase in public expenditure, we see a decrease in child poverty by -2.9230. With an adjusted R-Squared of 0.4878, we can say this model accounts for approximately 48* of variation in child poverty rates.

Normally Distributed Errors

```r
resid1 <- studres(model1)
summary(resid1)
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## -2.137247 -0.675482  0.069880  0.009147  0.890434  1.836427
```

```r
qqnorm(resid1)
qqline(resid1)
```

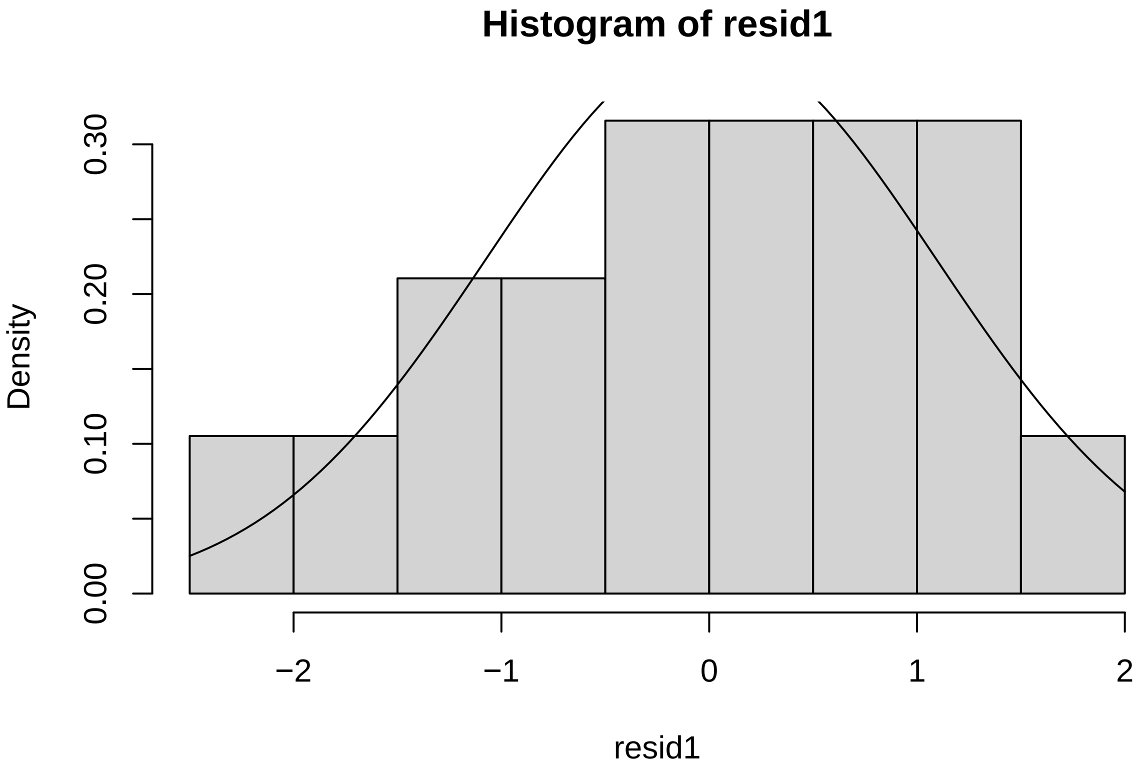



```r
shapiro.test(resid1)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  resid1
## W = 0.97172, p-value = 0.8103
```

```r
mean.r1 <- mean(resid1)
sd.r1 <- sd(resid1)
hist(resid1, freq = FALSE)
```

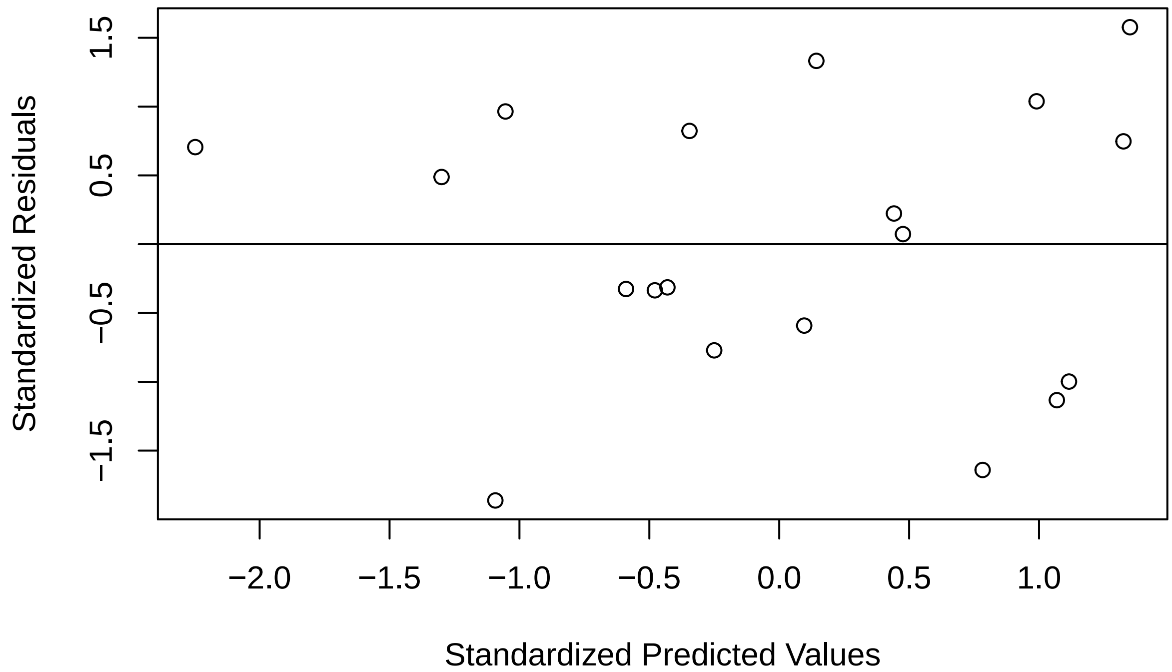
```r
curve(dnorm(x, mean=mean.r1, sd=sd.r1), add=TRUE)
```



There were multiple tests used to determine the normality distribution of the model. The generated studentized residuals show there to be a mean of 0. The generated Q-Q plot shows that the residuals lie relatively tightly among the normality line. Plotting the residuals on a histogram, overlapped by a normal curve line, we can see that the residuals are relatively normally distributed with the bulk of the distribution in the center and fewer as you move away from the center. Finally, the Shapiro-Wilk normality test has generated a p-value of 0.8103 for the residuals, in which we fail to reject the null that the residuals are normally distributed.

Heteroskedacity

```r
p.1 <- predict(model1)
std.p.1 <- (p.1 - mean(p.1))/sd(p.1)
r.1 <- resid(model1)
std.r.1 <- (r.1 - mean(r.1))/sd(r.1)
plot(std.p.1, std.r.1, xlab = "Standardized Predicted Values", ylab = "Standardized Residuals")
abline(0,0)
```




```r
bptest(model1)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  model1
## BP = 4.2542, df = 2, p-value = 0.1192
```

From a visual test of heteroskedasticity using the residual plots, there is a semi-pattern, although nothing clear pattern. The results of the Breusch-Pagan test give us a p-value of 0.1192, well above .05 which shows a lack of significance. From this, we fail to reject the null which concludes that homoskedasticity is present, thus our OLS assumptions aren't violated.

Collinearity

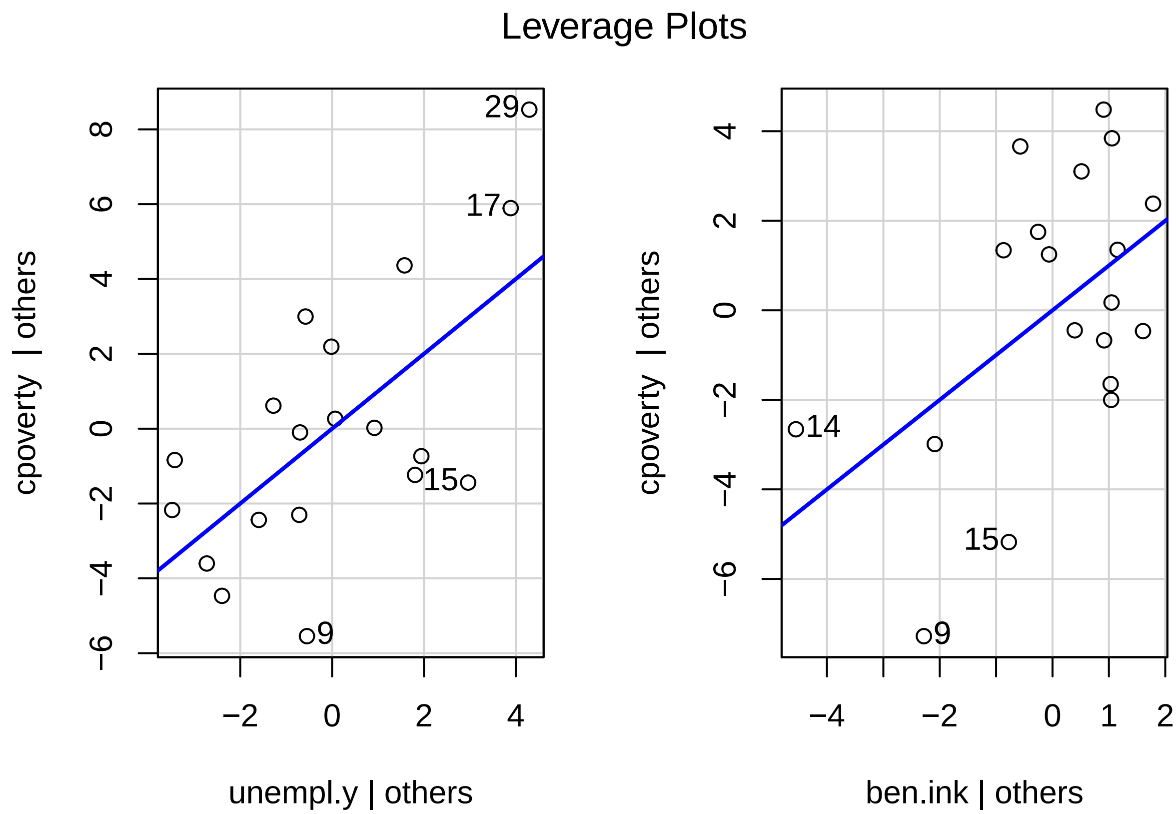
```r
pred1 <- model1$mode[-1]
vifs1 <- vif(pred1)
vifs1
```

```
##   Variables      VIF
## 1  unempl.y 1.009105
## 2   ben.ink 1.009105
```

With a variable inflation factor much less than 4, there is enough evidence to assume collinearity is not an issue with our model.

Outliers

```r
leveragePlots(model1)
```



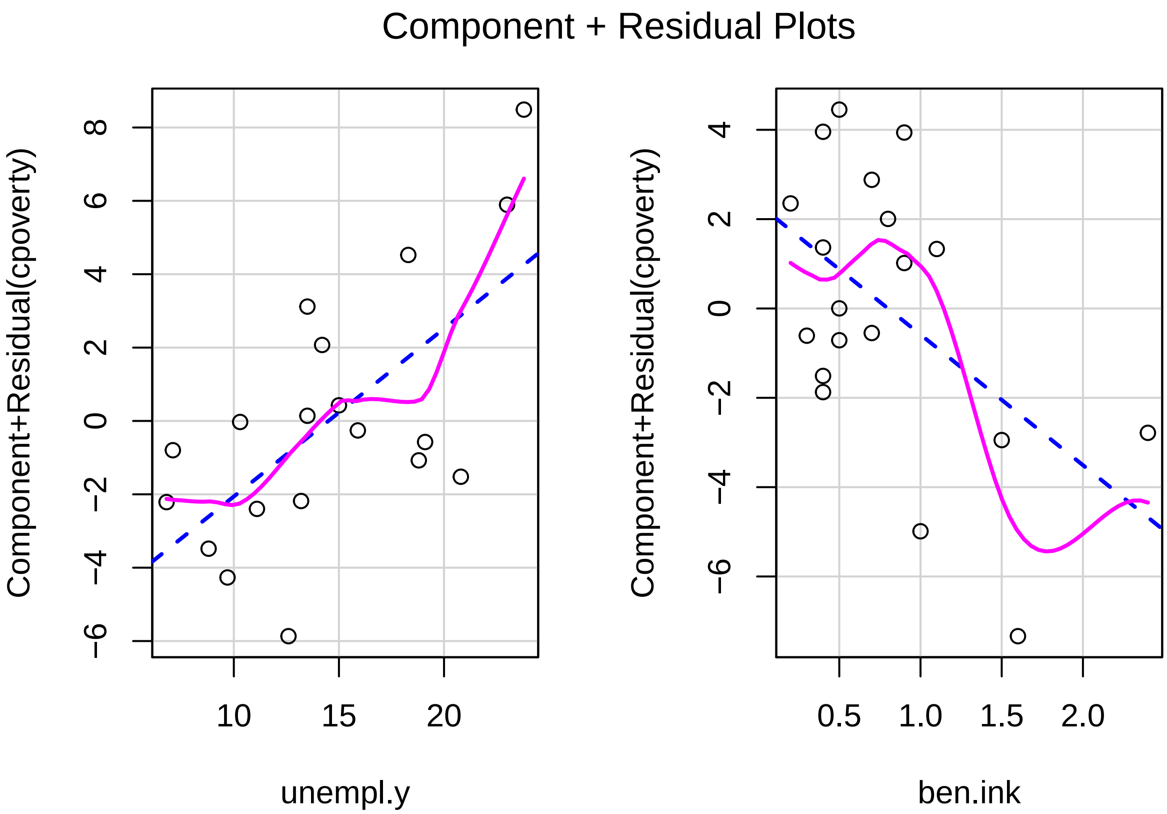

```r
outlierTest(model1)
```

```
## No Studentized residuals with Bonferroni p < 0.05
## Largest |rstudent|:
##    rstudent unadjusted p-value Bonferroni p
## 9 -2.137247           0.049455      0.93964
```

The results from the Bonferroni Outlier Test show a p of 0.93, which we fail to reject the null hypothesis that states no significant outliers.

Linearity

```r
crPlots(model1)
```



###

The results from the Component-Residual Plot show a linear relationship between both unemployment and public expenditure on child poverty rates.

After running diagnostics on the model, there shows no violating of any OLS assumptions including it being Normally-Distributed, Mean Zero Errors, Homoskedasticity, Collinearity, Lack of Outliers, and existence of Linearity. Due to these assumptions being met, there doesn't seem to be a need to adjust the model any further.