

Q3-SLR-water consumption

Pablo E. Gutiérrez-Fonseca

2024-12-08

Question 3

Exam Question:

Understanding water consumption patterns is essential for managing resources effectively, particularly in growing urban areas. One common approach is to model water consumption as a function of population size, which allows researchers to estimate water demand based on demographic data.

In this question, you will work with data from Tang and Ang (2007), which examines water consumption (WC) as a response variable and population size (POP) as a predictor.

- Perform a linear regression to determine whether population size is a significant predictor of water consumption.
- Evaluate the model's assumptions, including normality of residuals and homoscedasticity.
- Summarize your analysis in a concise paragraph, including your regression results, assumption checks, and any changes made to refine the model.

```
wc_data <- read.csv('Q3-SLR-water consumption.csv')
head(wc_data)
```

```
##   CITY   POP  WC
## 1    1 50000 100
## 2    2 100000 110
## 3    3 200000 110
## 4    4 250000 113
## 5    5 300000 125
## 6    6 400000 130
```

Set up regression

```
mod3 <- lm(WC ~ POP, wc_data)
summary(mod3)
```

```
##
## Call:
## lm(formula = WC ~ POP, data = wc_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -6.0967  -3.6530  -0.0098   3.7402   6.0488
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 98.932363014  2.845932507   34.76 0.000000000513 ***
```

```
## POP          0.000071455  0.000006203   11.52 0.000002925192 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.74 on 8 degrees of freedom
## Multiple R-squared:  0.9431, Adjusted R-squared:  0.936
## F-statistic: 132.7 on 1 and 8 DF,  p-value: 0.000002925
```

Get standardized residuals

```
mod3.res <- rstandard(mod3)
shapiro.test(mod3.res)

##
## Shapiro-Wilk normality test
##
## data:  mod3.res
## W = 0.91235, p-value = 0.2975
```

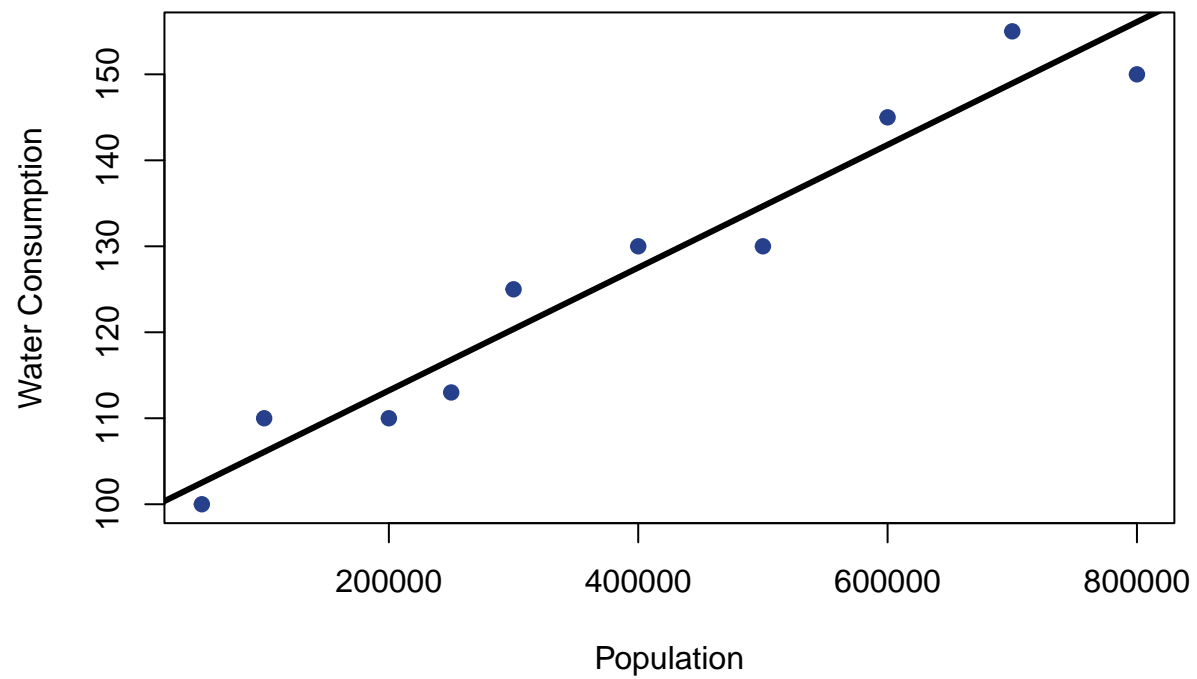
The Breusch-Pagan test to assess homoscedasticity

```
ncvTest(mod3)

## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 0.9224111, Df = 1, p = 0.33684
```

Make plot of Water Consumption and Population

```
plot(WC ~ POP, wc_data,
     pch = 19,
     col = "royalblue4",
     ylab = "Water Consumption",
     xlab = "Population")
#add regression line
#make line width thicker
abline(mod3, lwd=3)
```



References:

- <http://avinashu.com/tutorial/indexLR.html>