Project 2

Tony Nguyen

April 26, 2019

Introduction

Atomic Weapons exposure

During the 1940s and 1950s, the United States conducted 1,054 nuclear tests mostly in Nevada test site.

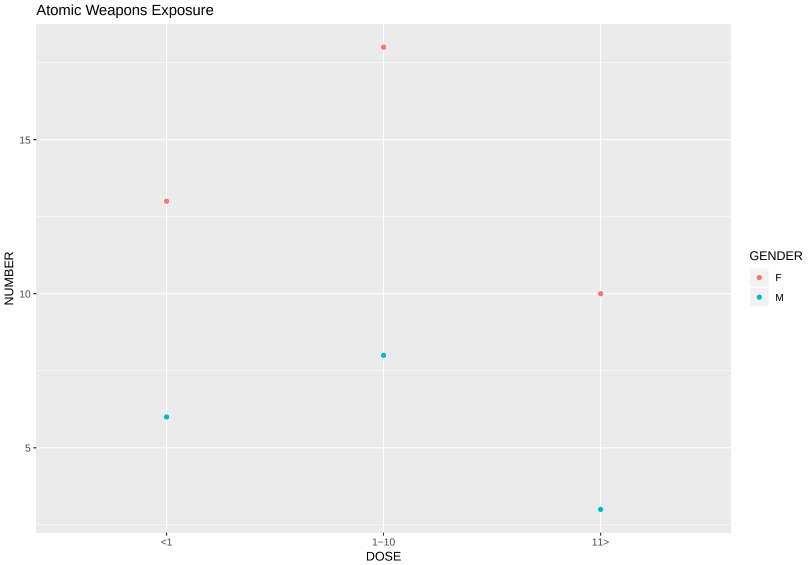
Below is Operation Desert Rock conducted between 1951 and 1957 (Image from the Atomic Heritage Foundation).

Problem

Because of the radiation, it has led researchers to look into the effects of these testings on workers exposed to these atomic weapons. It would be looking at doses and number of thyroid cancer cases.

```
##    DOSE GENDER NUMBER
## 1    <1      M      6
## 2  1-10      M      8
## 3   11>      M      3
## 4    <1      F     13
## 5  1-10      F     18
## 6   11>      F     10
```

Plot data and interpretation



Theory needed to carry out SLR:

$$Y = \beta_0 + \beta_1 X + \epsilon_i$$

Where β_0 and β_1 are unknown parameters. ϵ_i is the random error.

Assumptions

According to the book, we have to make these assumptions about ϵ:

1. The mean of the probability distribution of $\epsilon = 0$.
2. ϵ is normally distributed.
3. Errors from one observation are independent from others.
4. The variance of the probability distribution of ϵ is constant for all values of the independent variable - for a straight line model this means $V(\epsilon) =$ a constant for all values of x.

Method of Least Squares

```
##
## Call:
## lm(formula = NUMBER ~ DOSE, data = atomic_data)
##
## Residuals:
##    1    2    3    4    5    6
## -3.5 -5.0 -3.5  3.5  5.0  3.5
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    9.500      4.062   2.339    0.101
## DOSE1-10       3.500      5.745   0.609    0.585
## DOSE11>       -3.000      5.745  -0.522    0.638
##
## Residual standard error: 5.745 on 3 degrees of freedom
## Multiple R-squared:  0.2995, Adjusted R-squared:  -0.16
## F-statistic: 0.6414 on 2 and 3 DF,  p-value: 0.5863
```

Least Square Methods cont.

$$\hat{\beta}_0 = 9.500, \hat{\beta}_1 = 3.500, \hat{\beta}_2 = -3.000$$

Least Squares Estimate

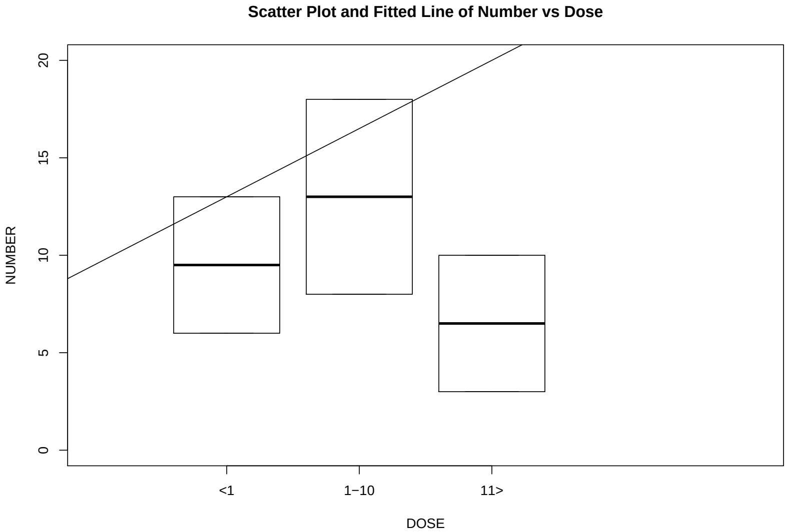
$$\hat{\beta}_0 + \hat{\beta}_1 X + \hat{\beta}_2 X^2 = 9.5 + 3.5X - 3.0X^2$$

Calculating Confidence Interval for Parameter Estimates

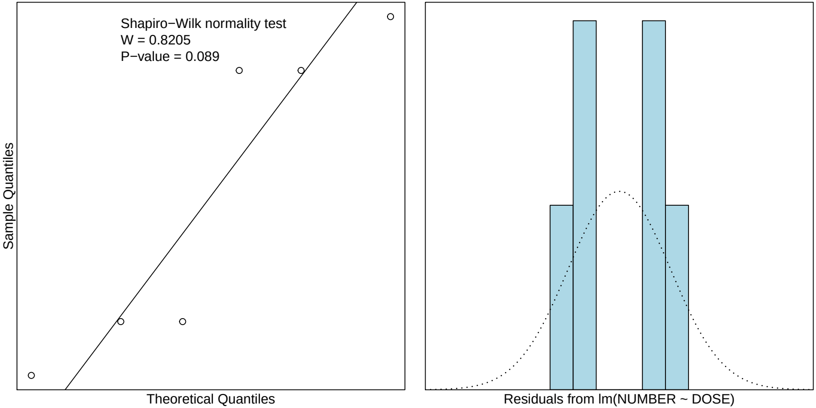
```
##              95 % C.I.lower    95 % C.I.upper
## (Intercept)        -3.42716          22.42716
## DOSE1-10          -14.78176          21.78176
## DOSE11>           -21.28176          15.28176
```

Validity with mathematical expressions

```
## Warning in abline(atomic.lm): only using the first two o
## coefficients
```



Shapiro-wilk



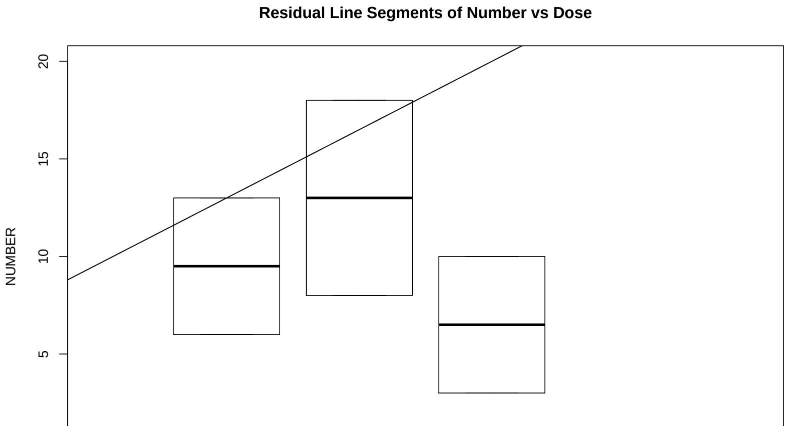
Shapiro-wilk cont.

After using shapiro-wilk normality test, the p-value for the atomic data is 0.089. Since the p-value of 0.089 is greater than 0.05, that means the data is distributed normally and we would not reject the null hypothesis.

Residual vs fitted values

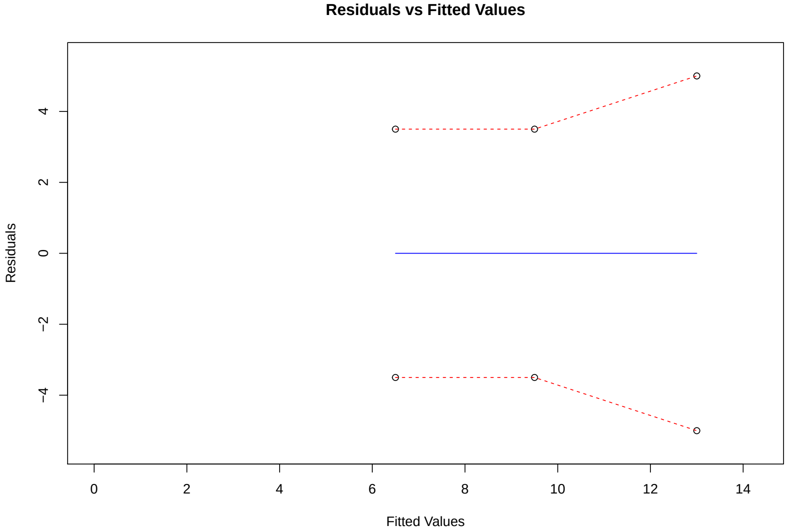
```
## Warning in abline(ht.lm): only using the first two of 3
## coefficients

## Warning in abline(ht.lm): only using the first two of 3
## coefficients
```



trendscatter on Residual Vs Fitted

There is some uniform to this plot so linear model would work.



Use adjusted R^2

$$R^2_{adj} = 1 - [\frac{(1 - R^2)(n - 1)}{n - k - 1}]$$

The closer R^2 is to 1 the better fit of the trend line. $R^2 = \frac{MSS}{TSS}$

```
## [1] 99
```

```
## [1] 42.33333
```

```
## [1] 141.3333
```

```
## [1] 0.2995283
```

R2 is about 0.29 and that means if not a very significant correct fit for the dataset.

Use of `predict()`

Lets try to predict the doses with the following number of cases of thyroid cancers 20, 25, 41, 55, 59, 68.

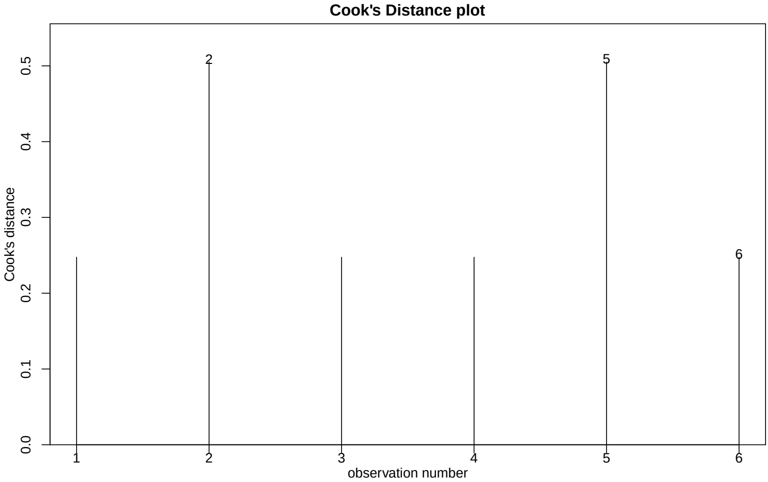
```
##    1    2    3    4    5    6
##  9.5 13.0  6.5  9.5 13.0  6.5
```

Use of ciReg()

```
##                 95 % C.I.lower    95 % C.I.upper
## (Intercept)           -3.42716          22.42716
## DOSE1-10             -14.78176          21.78176
## DOSE11>              -21.28176          15.28176
```

Check on outliers using cooks plots

Using cooks plots to check for outliers and how they may be distorting plot.



Conclusion

There isn't a significant relationship between number of thyroid cases and doses of exposure. Linear model isn't great for this type of problem.

References

U.S. Department of Energy Nevada Operations Office.2000."United States Nuclear Test July 1945 through September 1992."https://web.archive.org/web/20061012160826/http://www.nv.doe.gov/library/publications/historical/DOENV_209_REV15.pdf

Wikipedia. n.d.Comprehensive Nuclear-Test-Ban Treaty.https://en.wikipedia.org/wiki/Comprehensive_Nuclear-Test-Ban_Treaty