CS664 ASSIGNMENT 1

Methods for Statistical Consulting

PRITHWISH GANGULY
PG422

Question 1

```
##
## Call:
## lm(formula = log(C) ~ D + log(T2) + log(S) + PR + NE + CT + log(N) +
##     PT)
##
## Residuals:
##     Min      1Q   Median      3Q      Max
## -0.29131 -0.09935  0.02178  0.09351  0.24800
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -15.22561    3.37328  -4.514 0.000156 ***
## D             0.22722    0.04394   5.171 3.06e-05 ***
## log(T2)       0.30186    0.22833   1.322 0.199155
## log(S)        0.68246    0.12805   5.330 2.07e-05 ***
## PR           -0.09336    0.07022  -1.330 0.196709
## NE            0.25895    0.07379   3.509 0.001886 **
## CT            0.11462    0.06227   1.841 0.078631 .
## log(N)       -0.07873    0.04249  -1.853 0.076751 .
## PT           -0.21572    0.11451  -1.884 0.072280 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1578 on 23 degrees of freedom
## Multiple R-squared:  0.8706, Adjusted R-squared:  0.8256
## F-statistic: 19.34 on 8 and 23 DF,  p-value: 1.709e-08
```

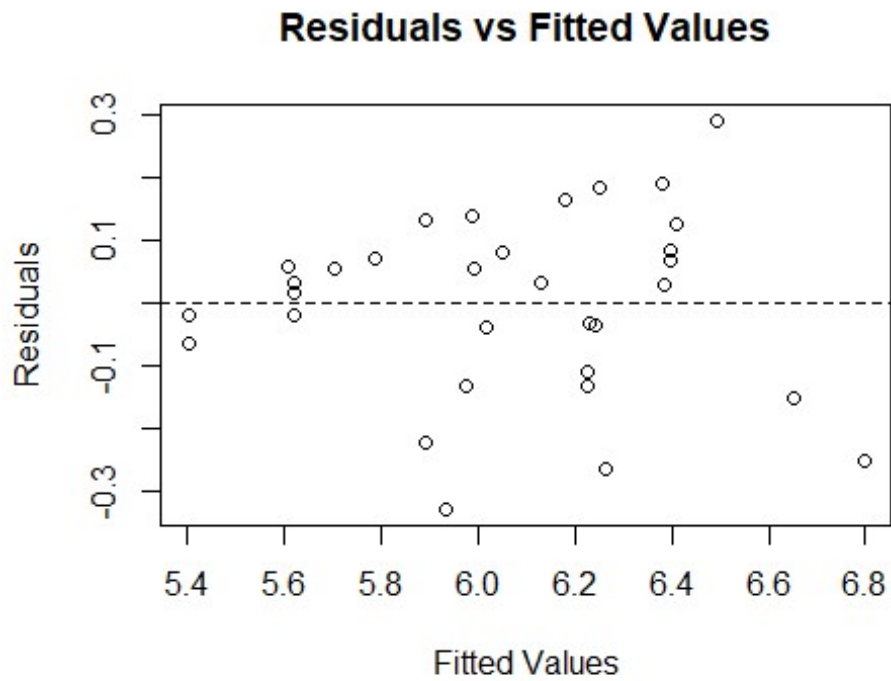**Estimate for PT**: -0.21451 Standard error for PT: 0.12229

Nuclear plants with partial turnkey guarantee cost 0.21451 less than those that do not have the guarantee.
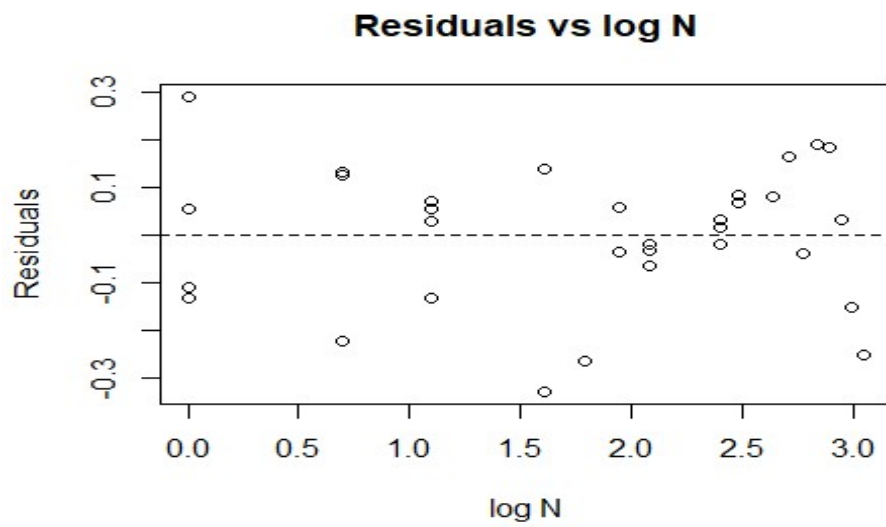
Question 2

```
##
## Call:
## lm(formula = log(C) ~ PT + CT + log(N) + log(S) + D + NE)
##
## Residuals:
##     Min      1Q   Median      3Q      Max
## -0.32721 -0.07620  0.02920  0.08115  0.28946
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -13.26031    3.13950  -4.224 0.000278 ***
## PT           -0.22610    0.11355  -1.991 0.057490 .
## CT            0.14039    0.06042   2.323 0.028582 *
## log(N)       -0.08758    0.04147  -2.112 0.044891 *
## log(S)        0.72341    0.11882   6.088 2.31e-06 ***
## D             0.21241    0.04326   4.910 4.70e-05 ***
## NE            0.24902    0.07414   3.359 0.002510 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1592 on 25 degrees of freedom
## Multiple R-squared:  0.8569, Adjusted R-squared:  0.8225
## F-statistic: 24.95 on 6 and 25 DF,  p-value: 2.058e-09
```
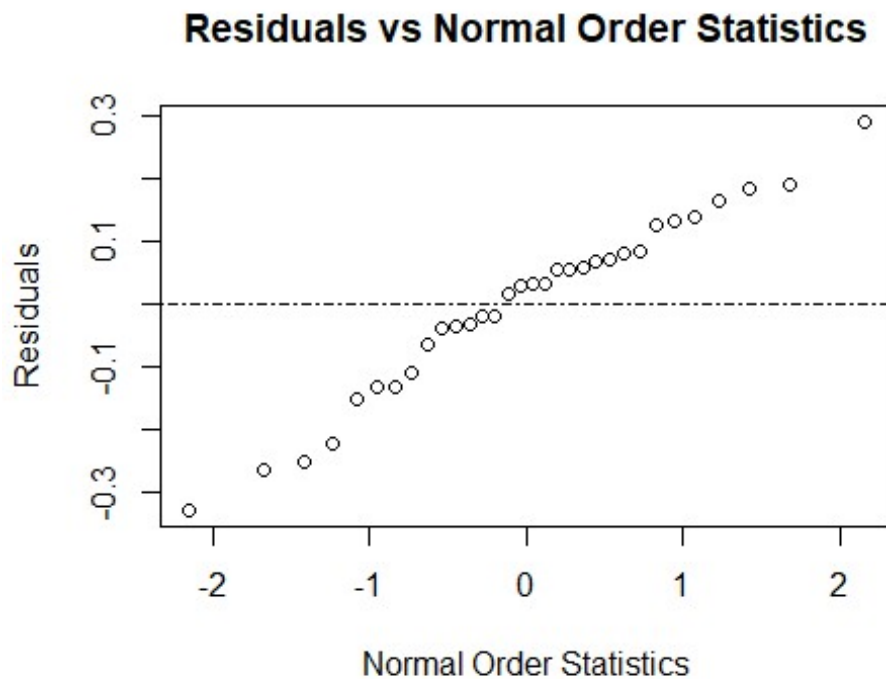
Question 3

Part a

Residuals vs Fitted Values

Part b

Residuals vs log N

Part c

Residuals vs Normal Order Statistics

Question 4

Residuals vs Fitted values - This checks for the assumption that the relationship is linear and also appears if the data appears homoscedastic. When this assumption is guaranteed, the residuals are grouped around the 0 line. If the data is homoscedastic, the points are scattered randomly around the x-axis. If they are not, e.g. if they form a curve, bowtie etc., then data doesn't meet the assumption.

Residuals vs log N - This checks for outliers generated by the log N variable in particular. The assumption being made here is that the relationship between residuals and log N in particular is also linear and this can be observed by the fact that points are grouped randomly around the 0 line as observed here.

Residuals vs Normal Order Statistics - This checks if the residuals are normally distributed. When this assumption is guaranteed, the data points closely follows a straight line at a 45 degree angle upwards.

The above plots also help us look for outliers which are points that are further away from other points.

Question 5

```
##
## Call:
## lm(formula = log(C) ~ PT + CT + log(N) + log(S) + D + NE + I(PT *
##     Z))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.32866 -0.05714  0.02067  0.07979  0.29282
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -13.08645    3.23858  -4.041 0.000475 ***
## PT           -2.18759    5.85357  -0.374 0.711895
## CT            0.13998    0.06154   2.275 0.032156 *
## log(N)       -0.08683    0.04229  -2.053 0.051102 .
## log(S)        0.71761    0.12222   5.872 4.68e-06 ***
## D             0.21044    0.04444   4.735 8.14e-05 ***
## NE            0.24841    0.07551   3.290 0.003088 **
## I(PT * Z)     0.29159    0.87002   0.335 0.740418
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1621 on 24 degrees of freedom
## Multiple R-squared:  0.8575, Adjusted R-squared:  0.816
## F-statistic: 20.64 on 7 and 24 DF,  p-value: 1.033e-08
```

References

Applied Statistics, Principles and Examples by D. R. Cox & E. J. Snell