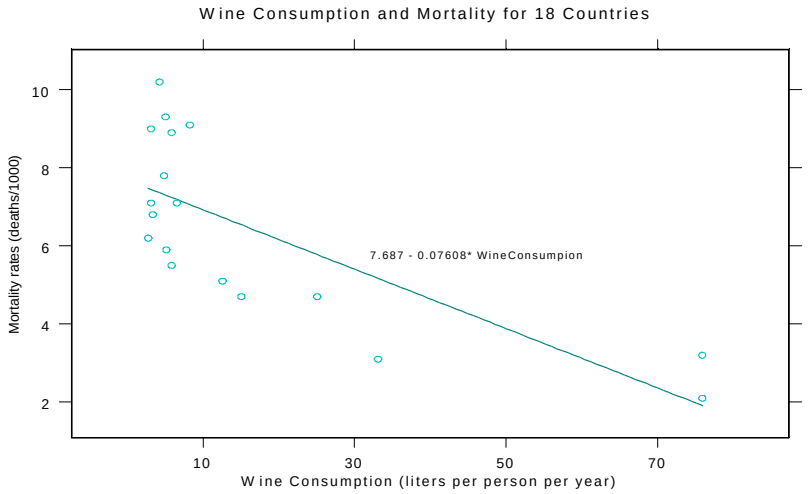
Analysis of Mortality and Wine Consumption



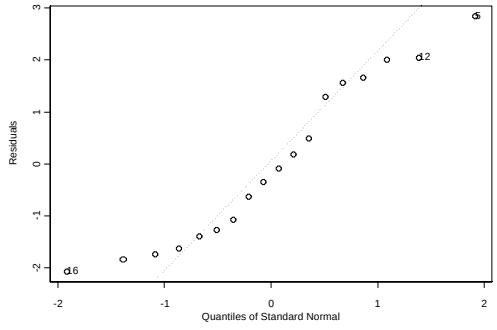

```
Coefficients:
                Value  Std. Error     t value    Pr(>|t|)
(Intercept)  7.6865549   0.4733221  16.2395861   0.0000000
       wine -0.0760809   0.0170013  -4.4750023   0.0003828

Multiple R-Squared: 0.5558719
```

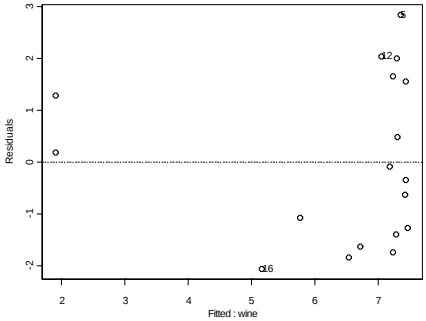
Depite the statistical significance of the slope (p-value = .00382), the fit of the linear regession model in not great.

Assumptions for linear regression:

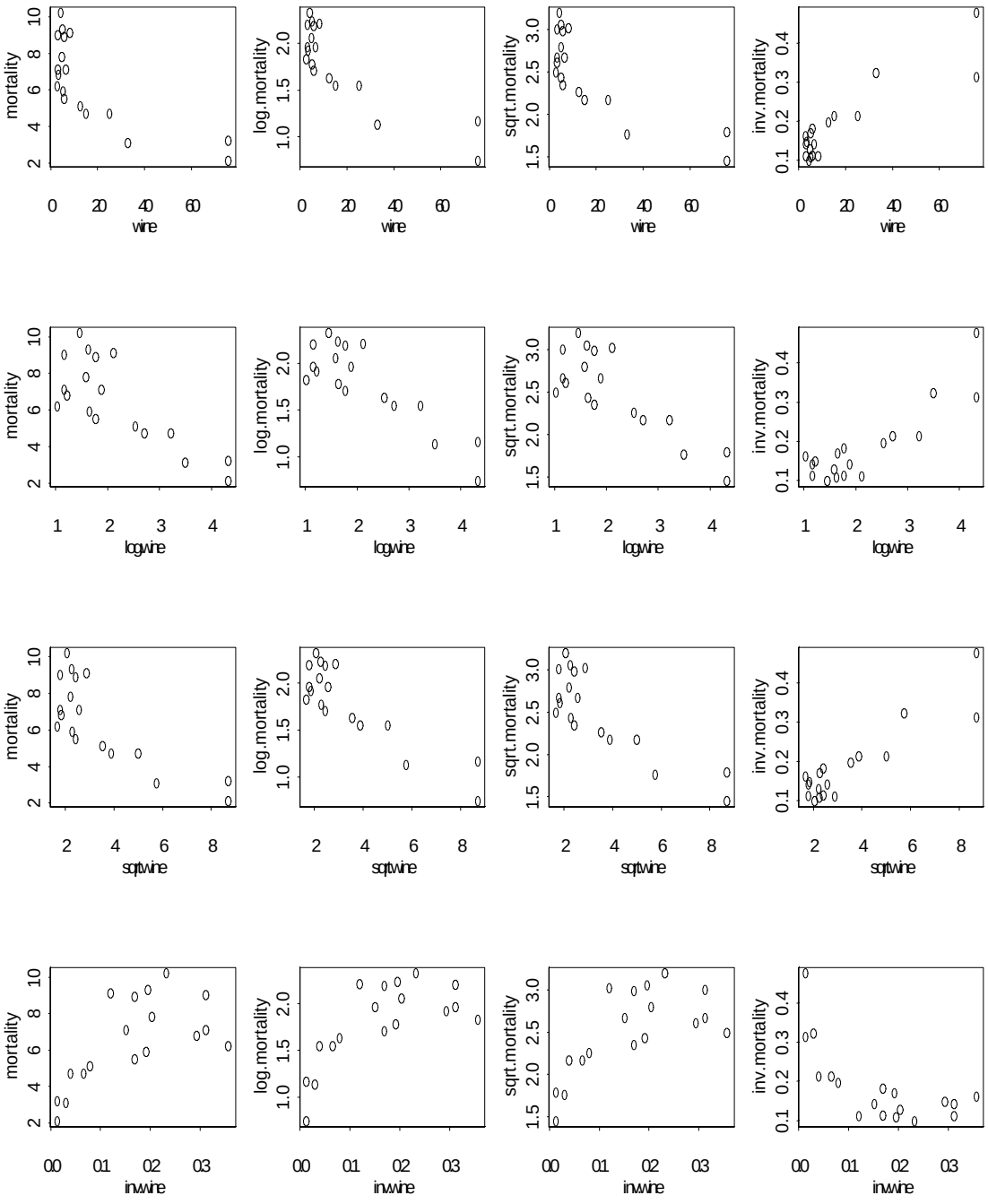
1. Mean is linear in Wine Consumption
2. Standard deviation of the errors is constant
3. Errors are independent (can't tell)
4. Errors are normally distributed

Scatter plot, residual plot and normal proability plot indicate assumptions 1, 2, and 4 are not .



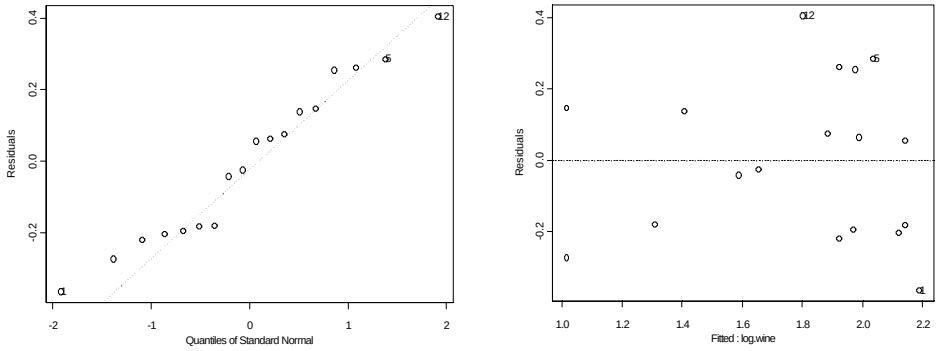


Goal: Find transformation of Wine Consumption and/or Mortality so that assumptions for simple linear model are valid.



Transformations of both Mortality and Wine Consumption

Log transformations of both look ok.... fit linear regression in log - log scale
Check residual plots... before proceeding with any tests, etc.

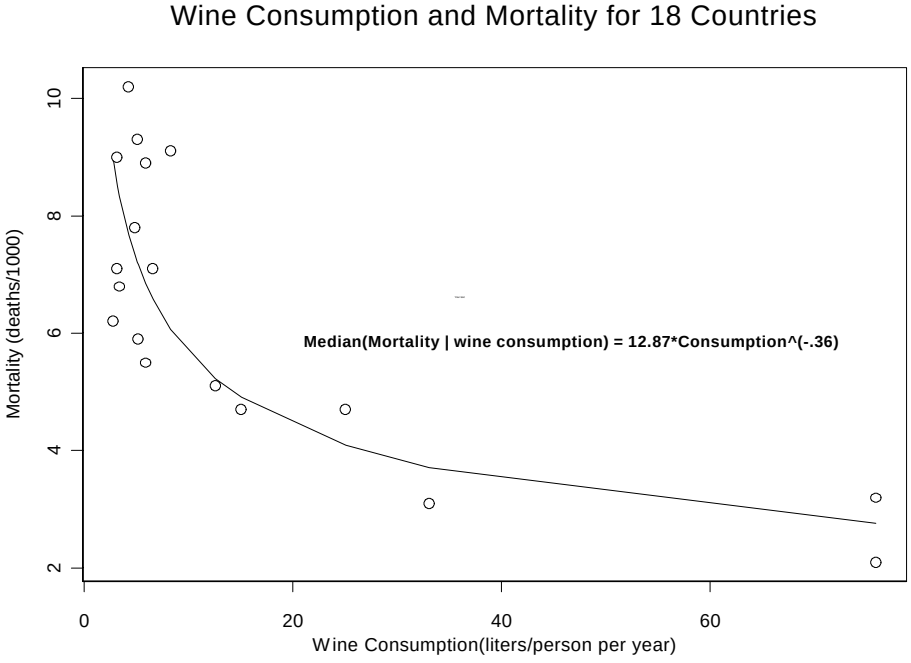


Not bad!
Because there are some countries that have the same level of wine consumption we can conduct a lack of fit test.

Ho: mean is linear in log(wine consumption)
Ha: mean is not linear in log(wine consumption)

F Value is 0.59813 with a p-value of 0.78.
Conclusion: no evidence of lack of fit of the linear model.

Now, we can get down to the business of interpreting the results!



*** Linear Model ***

```
Call: lm(formula = log.mortality ~ log.wine, data = Ex0822, na.action =
      na.omit)
Residuals:
        Min         1Q     Median        3Q       Max
 -0.3648742 -0.1912188 0.01497261 0.1448537 0.4052543

Coefficients:
                  Value  Std. Error     t value    Pr(>|t|)
(Intercept)   2.5555519   0.1268969  20.1388008   0.0000000
   log.wine  -0.3555959   0.0529094  -6.7208484   0.0000049

Residual standard error: 0.2285367 on 16 degrees of freedom
Multiple R-Squared: 0.738433
F-statistic: 45.1698 on 1 and 16 degrees of freedom, the p-value is 4.913645e-
006

Analysis of Variance Table

Response: log.mortality

Terms added sequentially (first to last)
          Df   Sum of Sq     Mean Sq     F Value            Pr(F)
 log.wine  1 2.359175573 2.359175573 45.16980284 4.913645278e-006
Residuals 16 0.835664687 0.052229043
```

*** Analysis of Variance Model ***

```
Call:
   aov(formula = log.mortality ~ as.factor(log.wine), data = Ex0822,
       na.action = na.omit)

Terms:
                as.factor(log.wine)   Residuals

 Sum of Squares         2.962188221 0.232652039
Deg. of Freedom                  14           3

Residual standard error: 0.2784792265
Estimated effects may be unbalanced

                     Df   Sum of Sq      Mean Sq     F Value
as.factor(log.wine) 14 2.962188221 0.2115848729 2.728343247
          Residuals  3 0.232652039 0.0775506796
```

Source	***DF***	***SS***	***MS***	***F-statistic***	***Pr(F)***
Residual (linear model)					
Lack of fit					0.78
Pure Error (Residual AOV)					