# Ball Bearing Analysis

Anjani Chaudhary and Kyle Garza
Math 5345, D. Sun
Texas State University

# Ball Bearing Analysis: In two parts

1. Recreation of analysis from a theoretical model from an article by Chrys Caroni (2002) of data originally collected and analyzed by Lieblein and Zelen (1956).

2. An exploratory analysis of the Lieblein-Zelen data where we build a model from the ground up.

# Checking the Reliability of Ball Bearings

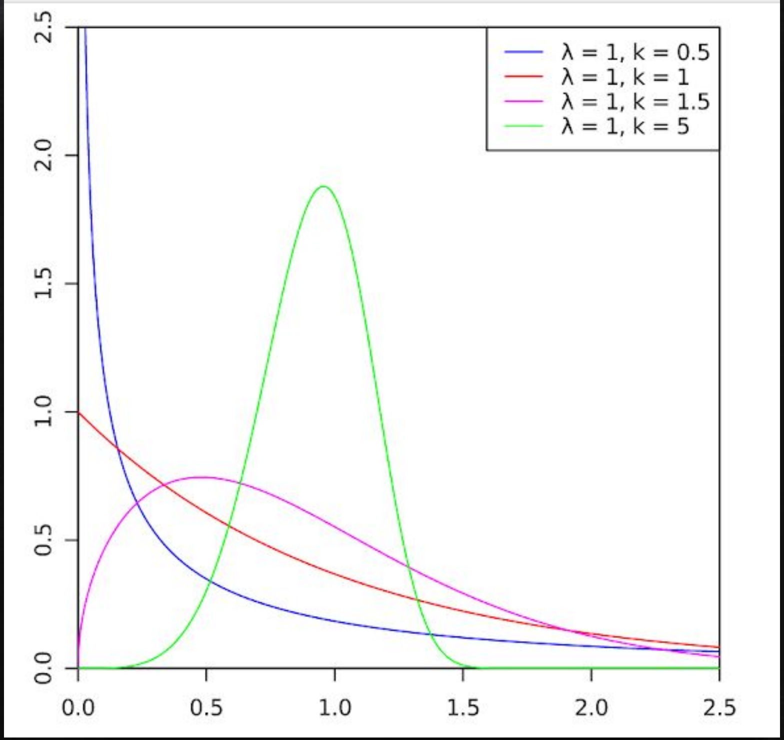
- Load ($P$), Diameter of Balls ($D$), Number of Balls ($Z$)
- Bearing Life ($L_{10}$) – Number of revolutions in Millions that 10% of bearings are expected to fail
- Standard model used in the industry:

$$L=\left(\frac{fZ^a C^b}{P}\right)^p$$

# How the Data was Collected

- 210 Observations completed by three companies
- Each observation consisted of 8 to 94 bearings tested until they failed
- Measured the number of revolutions in millions at time of failure
- $L_{10}$ was estimated using a Weibull distribution

## Weibull Distribution



# Tested Hypotheses

(a) All the parameters of the equation are the same for each one of the three companies
(b) The parameter $\beta_3$ (hence, $-\rho$) is the same for each company
(c) All the parameters of the equation are the same for each one of three types of bearing produced by Company B
(d) The parameter $\beta_3$ is the same for each type of bearing produced by Company B.

# Linearization of the Model

Lieblein and Zelen's main concern was estimating $\rho$, believed to be equal to 3. This can be estimated with the parameter $-\beta_3$.

$$L = \left( \frac{f Z^a D^b}{P} \right)^p$$

$$\ln(L) = \ln\left( \frac{f Z^a D^b}{P} \right)^p$$

$$\ln(L) = p \ln f + ap \ln Z + bp \ln D - p \ln P$$

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

# To test our hypotheses we need two additional models and some indicator variables

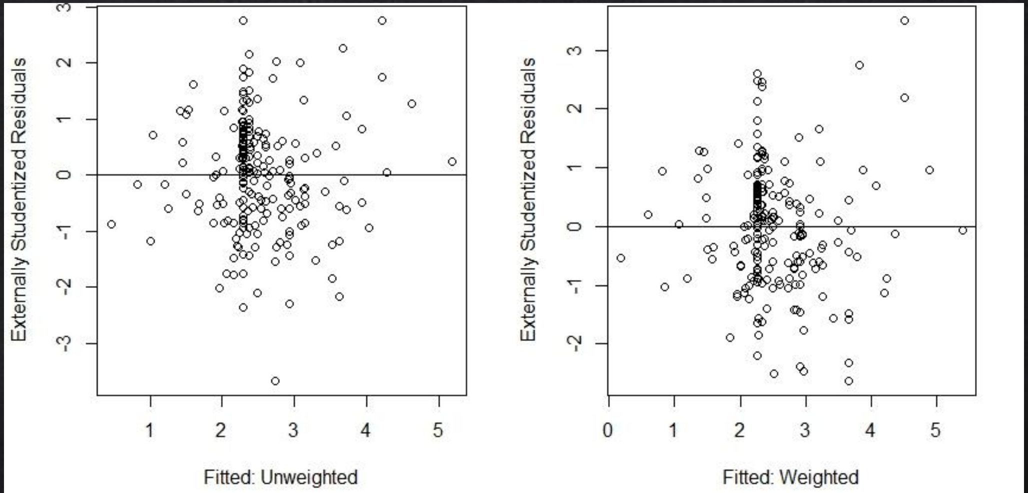
| | |
|---|---|
| Model 1 | $ln(L_{10}) \sim ln(Z) + ln(D) + ln(P) + B + C + Bln(Z) + Bln(D) + Bln(P) + Cln(Z) + Cln(D) + Cln(P)$ |
| Model 2 | $ln((L_{10}) \sim ln(Z) + ln(D) + ln(P)$ |
| Model 3 | $ln(L_{10}) \sim ln(Z) + ln(D) + ln(P) + B + C + Bln(Z) + Bln(D) + Cln(Z) + Cln(D)$ |

| | Company | B | C |
|---|---|---|---|
| Company A | 1 | 0 | 0 |
| Company B | 2 | 1 | 0 |
| Company C | 3 | 0 | 1 |

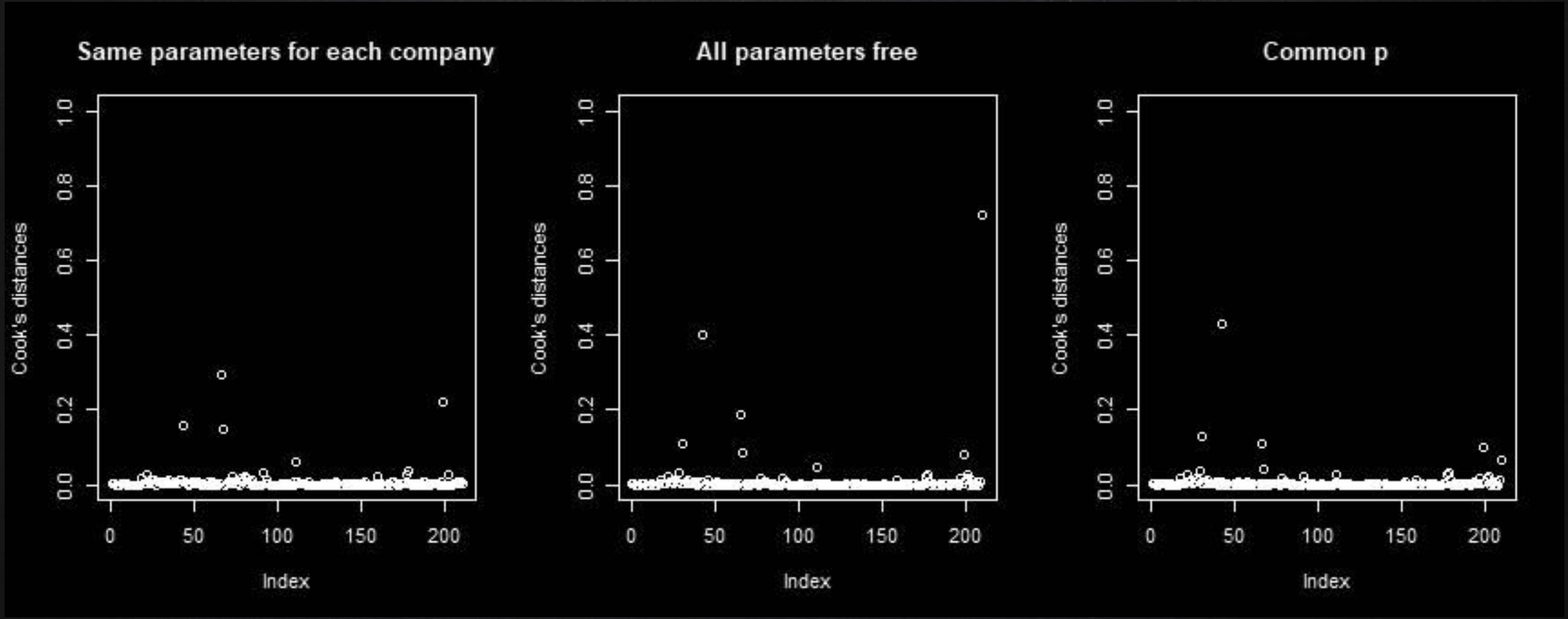
$B$ = ifelse($Company$ == 2, 1, 0)

$C$ = ifelse($Company$ == 3, 1, 0)

# Weighted Least-Squares

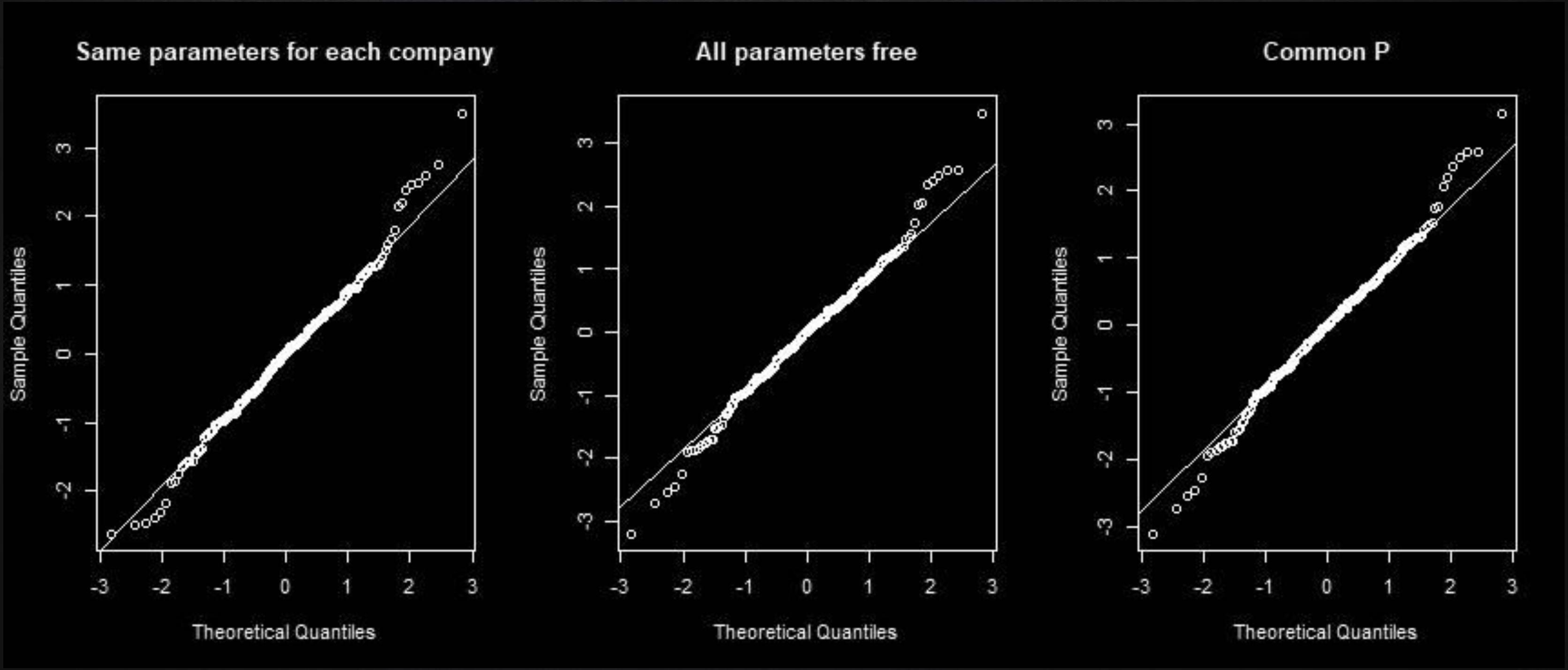



Fitted versus Residuals for the unweighted and weighted model: $\ln L_{10} \sim \ln Z + \ln D + \ln P$




No outliers found using cook's distance

# Normality? Yes!



# Results of the Analysis of Variance

| Analysis of Variance Table | | | | | | |
|---|---|---|---|---|---|---|
| Model 1 | $ln(L_{10}) \sim ln(Z) + ln(D) + ln(P) + B + C + Bln(Z) + Bln(D) + Bln(P) + Cln(Z) + Cln(D) + Cln(P)$ | | | | | |
| Model 2 | $ln((L_{10}) \sim ln(Z) + ln(D) + ln(P)$ | | | | | |
| Model 3 | $ln(L_{10}) \sim ln(Z) + ln(D) + ln(P) + B + C + Bln(Z) + Bln(D) + Cln(Z) + Cln(D)$ | | | | | |
| | Res. DF | RSS | DF | SS | F | P(>F) |
| 1 | 198 | 1944.5 | | | | |
| 2 (2-1) | 206 | 2215.4 | 8 | 270.88 | 3.4478 | 0.0009674 |
| 3 (3-1) | 200 | 1950.1 | 2 | 5.5869 | 0.2844 | 0.7527 |

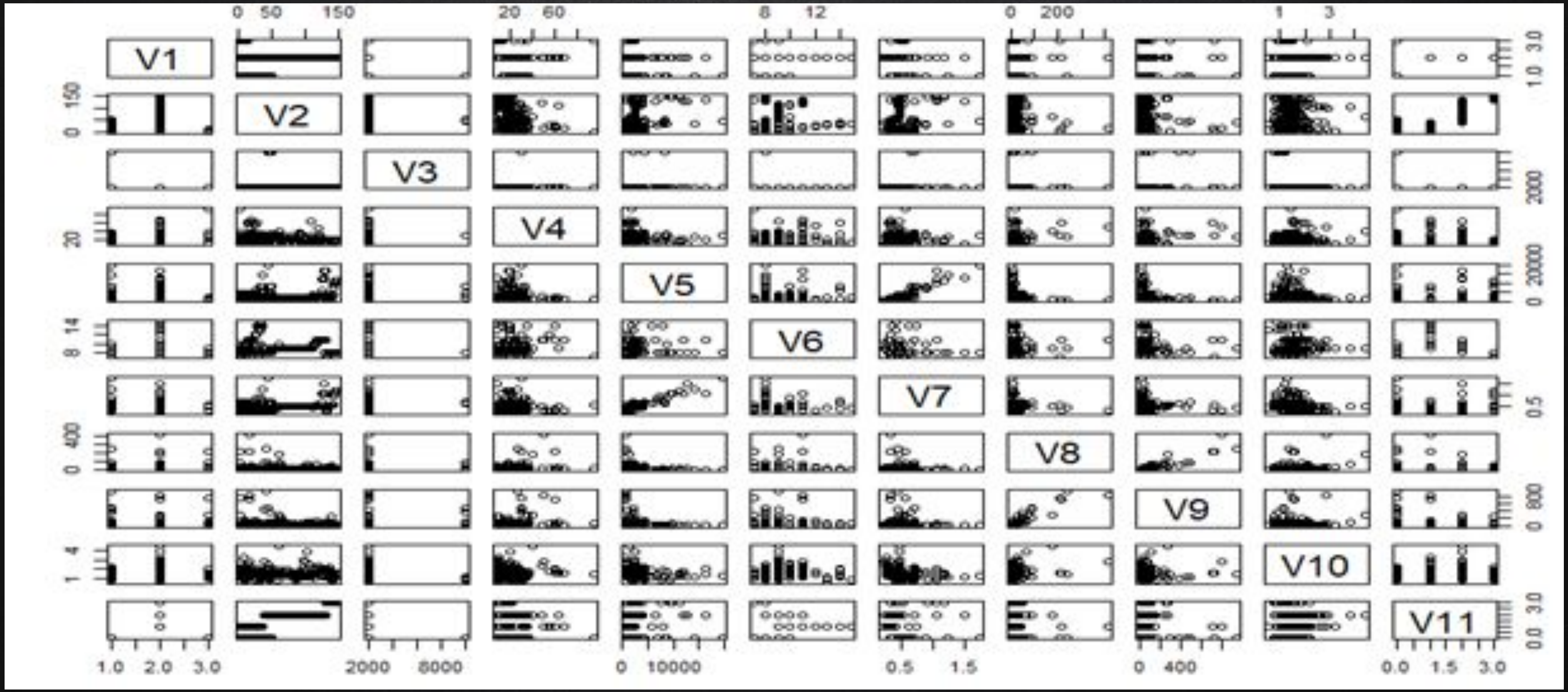
# Results of the Analysis of Variance (R-output)

```
> #Analysis for Model 3 - Model 1
> anova(bb.lm.w, bb.lmfull)
Analysis of Variance Table

Model 1: lnL10 ~ lnZ + lnD + lnP
Model 2: lnL10 ~ lnZ + lnD + lnP + B + C + BlnZ + BlnD + BlnP + ClnZ +
    ClnD + ClnP
  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1    206 2215.4
2    198 1944.5  8    270.88 3.4478 0.0009674 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> #Analysis for Model 2 - Model 1
> anova(bb.lmCP, bb.lmfull)
Analysis of Variance Table

Model 1: lnL10 ~ lnZ + lnD + lnP + B + C + BlnZ + BlnD + ClnZ + ClnD
Model 2: lnL10 ~ lnZ + lnD + lnP + B + C + BlnZ + BlnD + BlnP + ClnZ +
    ClnD + ClnP
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1    200 1950.1
2    198 1944.5  2    5.5869 0.2844 0.7527
```

# Ball Bearing

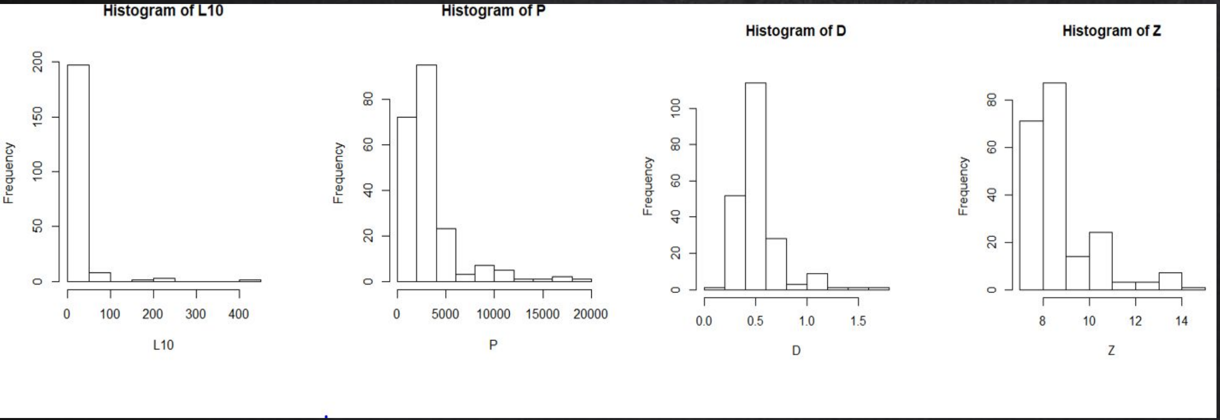
Our Exploratory Analysis



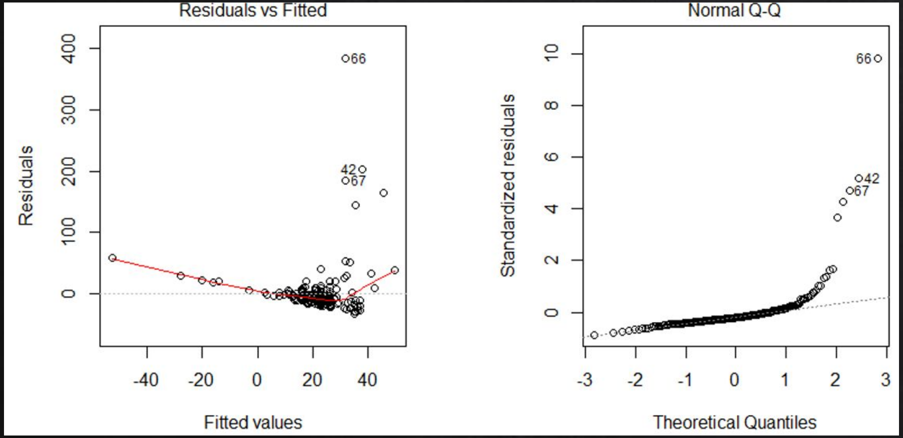

Scatter plot of all variables ( L10 = β0 + β1Z + β2D + β3P)

# Original Equation (without transformation)

## L10 = β0 + β1Z + β2D + β3P




- Mean square error: points not distributed in equal no, below and above ab line
- Normality: light tailed in on the right
- Constant variance :Non-constant
- Linearity: $R^2$ value very small and some kind of irregular shape




| $R^2$ | Adjusted $R^2$ |
|---|---|
| 8.589% | 7.258% |

# Possible transformations

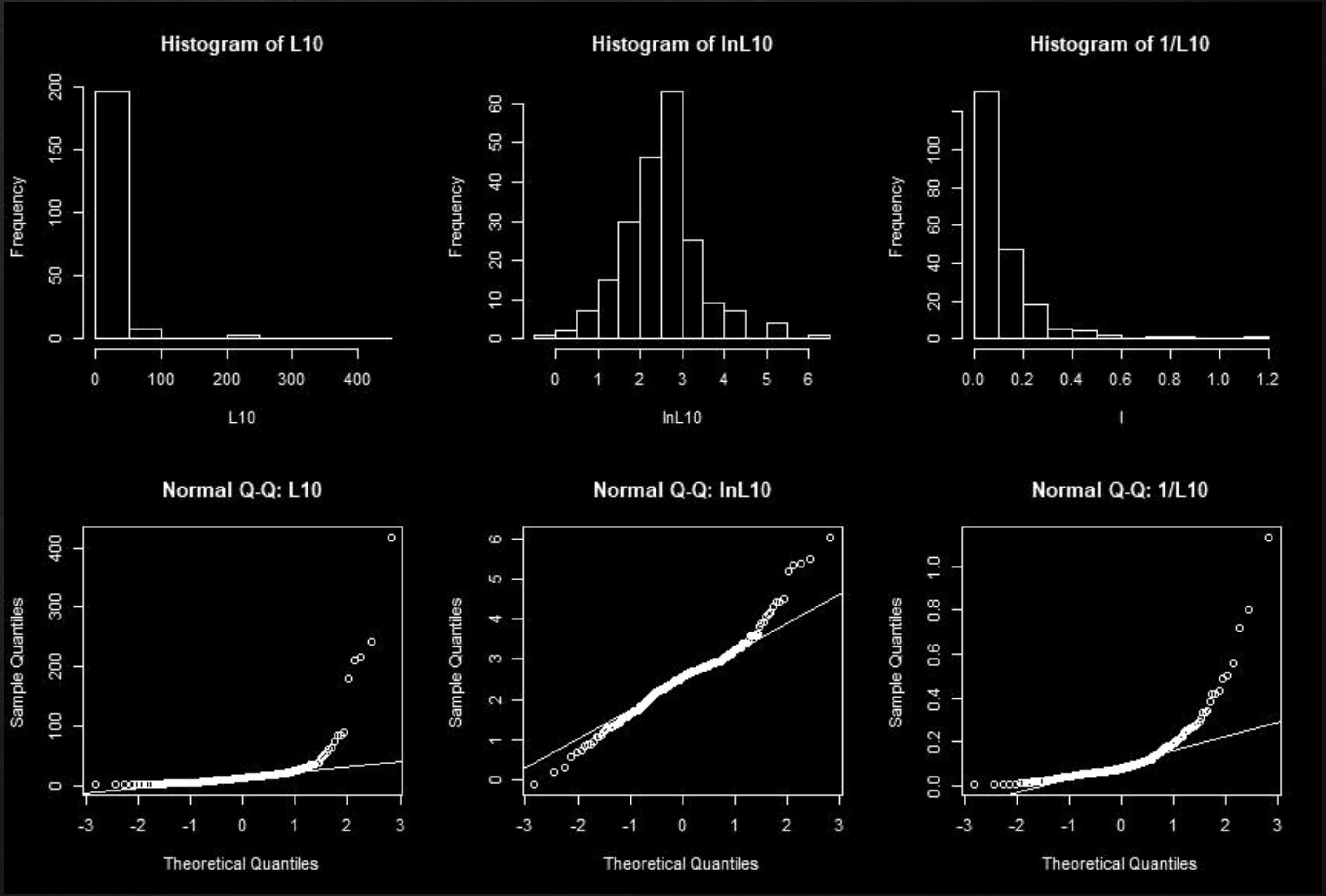
$L_{10}$

vs.

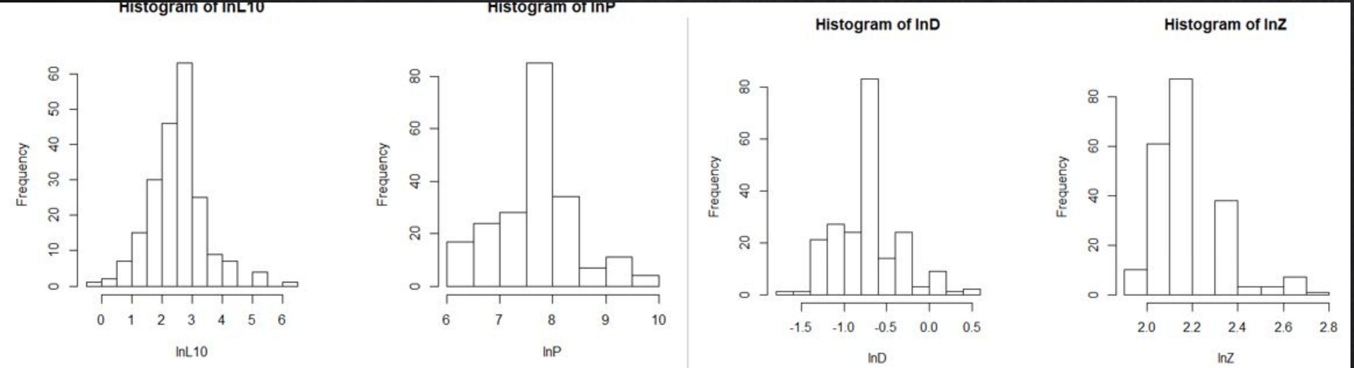
$\ln(L_{10})$

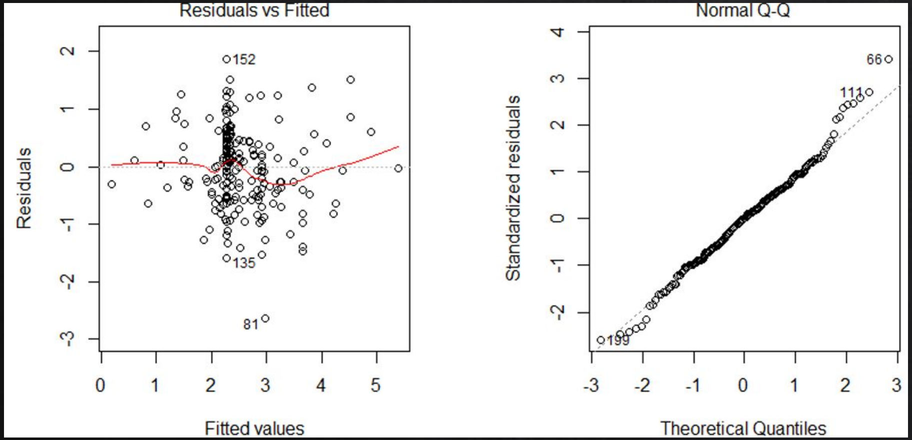
vs.

$1/L_{10}$



# After transformation




- Mean square error: points distributed in equal no, below and above ab line
- Normality: slighly light tailed in on the right but better than before
- Constant variance :

  No specific pattern
- Linearity: $R^2$ value greater and no specific shape

| $R^2$ | Adjusted $R^2$ |
| --- | --- |
| 55.08% | 54.42% |

# Considering Differences Between Companies

$$\ln(L_{10}) = 26.8505 + 1.1577\ln(Z) + 5.1646\ln(D) - 2.9908\ln(P) - 3.1904B + 1.7921C + 0.6804B\ln(Z) - 0.1973B\ln(D) + 0.2331B\ln(P) - 1.1759C\ln(Z) - 0.8365C\ln(D) - 0.1430C\ln(P)$$

```
> reg2 = lm(lnL10~lnP+lnZ+lnD+B+C+BlnZ+BlnD+BlnP+ClnP+ClnZ+ClnD, weights = N)
> summary(reg2)

Call:
lm(formula = lnL10 ~ lnP + lnZ + lnD + B + C + BlnZ + BlnD +
    BlnP + ClnP + ClnZ + ClnD, weights = N)

Weighted Residuals:
    Min      1Q  Median      3Q     Max
-8.2686 -2.0230  0.0308  1.7236  9.6884

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  26.8505     3.3231   8.080 6.27e-14 ***
lnP          -2.9908     0.2940 -10.174  < 2e-16 ***
lnZ           1.1577     1.2181   0.950    0.343
lnD           5.1646     0.6381   8.093 5.77e-14 ***
B            -3.1904     3.8987  -0.818    0.414
C             1.7921     5.9996   0.299    0.765
BlnZ          0.6804     1.2834   0.530    0.597
BlnD         -0.1973     0.8252  -0.239    0.811
BlnP          0.2331     0.3811   0.612    0.541
ClnP         -0.1430     0.6773  -0.211    0.833
ClnZ         -1.1759     2.6884  -0.437    0.662
ClnD         -0.8365     1.9812  -0.422    0.673
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.134 on 198 degrees of freedom
Multiple R-squared:  0.6057,    Adjusted R-squared:  0.5838
F-statistic: 27.65 on 11 and 198 DF,  p-value: < 2.2e-16
```

| $R^2$ | Adjusted $R^2$ |
|---|---|
| 60.57% | 58.38% |

# Variable Selection Algorithms

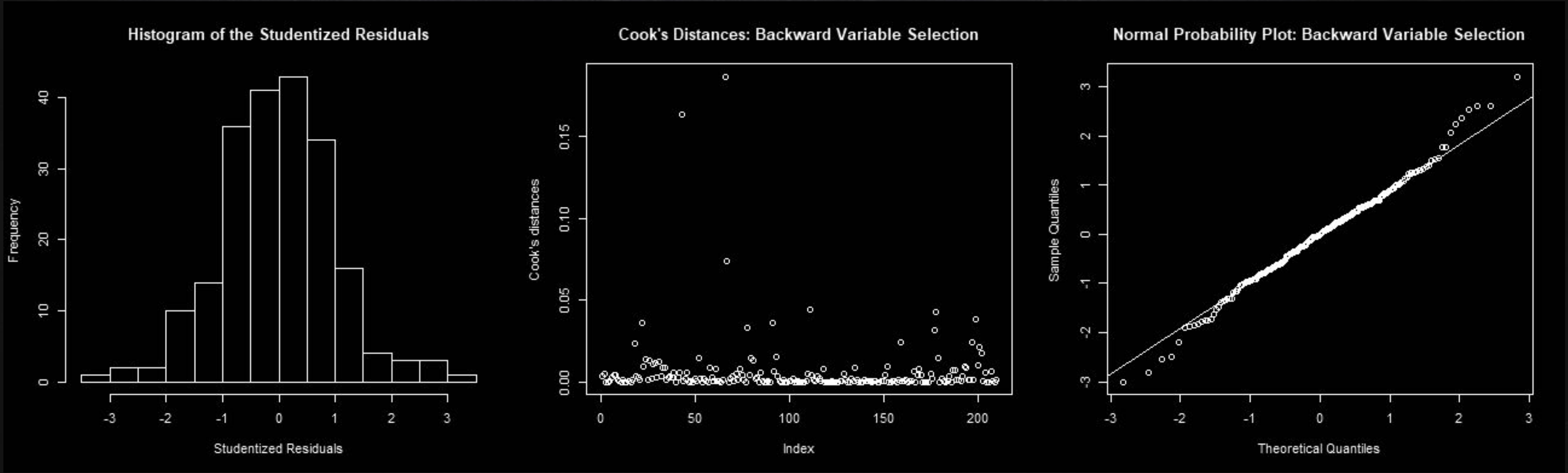
```
> summary(forw)

Call:
lm(formula = lnL10 ~ lnP + lnD + ClnP + lnZ + BlnP)

Residuals:
     Min       1Q   Median       3Q      Max 
-2.36024 -0.41123  0.02573  0.41273  1.78835 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 22.76945    1.60693  14.170  < 2e-16 ***
lnP         -2.53692    0.19206 -13.209  < 2e-16 ***
lnD          4.41488    0.40262  10.965  < 2e-16 ***
ClnP        -0.10893    0.03066  -3.553 0.000473 ***
lnZ          1.17859    0.37766   3.121 0.002065 ** 
BlnP         0.02228    0.01555   1.432 0.153606    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6645 on 204 degrees of freedom
Multiple R-squared:  0.499,	Adjusted R-squared:  0.4867 
F-statistic: 40.64 on 5 and 204 DF,  p-value: < 2.2e-16
```

```
> summary(back)

Call:
lm(formula = lnL10 ~ lnZ + lnD + lnP + BlnP + ClnZ, weights = N)

Weighted Residuals:
    Min      1Q  Median      3Q     Max 
-8.6622 -2.0509 -0.0315  1.7694  9.2115 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 24.69388    1.50123  16.449  < 2e-16 ***
lnZ          1.79801    0.36536   4.921 1.77e-06 ***
lnD          5.14191    0.37549  13.694  < 2e-16 ***
lnP         -2.89425    0.17493 -16.545  < 2e-16 ***
BlnP         0.03236    0.01552   2.085   0.0383 *  
ClnZ        -0.46229    0.10926  -4.231 3.51e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.101 on 204 degrees of freedom
Multiple R-squared:  0.6021,	Adjusted R-squared:  0.5923 
F-statistic: 61.74 on 5 and 204 DF,  p-value: < 2.2e-16
```

```
> summary(stepw)

Call:
lm(formula = lnL10 ~ lnP + lnD + ClnP + lnZ + BlnP)

Residuals:
     Min       1Q   Median       3Q      Max 
-2.36024 -0.41123  0.02573  0.41273  1.78835 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 22.76945    1.60693  14.170  < 2e-16 ***
lnP         -2.53692    0.19206 -13.209  < 2e-16 ***
lnD          4.41488    0.40262  10.965  < 2e-16 ***
ClnP        -0.10893    0.03066  -3.553 0.000473 ***
lnZ          1.17859    0.37766   3.121 0.002065 ** 
BlnP         0.02228    0.01555   1.432 0.153606    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6645 on 204 degrees of freedom
Multiple R-squared:  0.499,	Adjusted R-squared:  0.4867 
F-statistic: 40.64 on 5 and 204 DF,  p-value: < 2.2e-16
```

# Variable Selection

$$ln(L_{10}) = 24.69388 + 1.79801 ln(Z) + 5.14191 ln(D) - 2.89425 ln(P) + 0.03236 B ln(P) - 0.46229 C ln(Z)$$




| $R^2$ | Adjusted $R^2$ |
|---|---|
| 60.21% | 59.23% |

| VIF: $\ln Z$ | $\ln D$ | $\ln P$ | $B\ln P$ | $C\ln Z$ |
|---|---|---|---|---|
| 1.672802 | 9.880358 | 8.966071 | 1.542267 | 1.379378 |

# Conclusion

- For our study, we did not consider the type of bearing used by company B and the year in which all the tests were done (for additional research)
- From our analysis also we found that all parameters are not same for all companies
- Theoretically, the Caroni article says, parameter ‘p’ is same for all companies i.e value is 3. But in are analysis, we get ‘p’ slightly less than 3.

# References


[1] Caroni, Chrys. “Modeling the Reliability of Ball Bearings.” *Journal of Statistics Education*, Volume 10, Number 3, 2002.

[2] Lieblein, J. and Zelen, M. “Statistical Investigation of the fatigue Life of Deep-Groove Ball Bearings.” *Journal of Research of the National Bureau of Standards*, Volume 57, Number 5, November 1956, Research Paper 2719, 273-316.

[3] Montgomery, Douglas C. et al. *Introduction to Linear Regression Analysis*. 5th ed., Wiley, 2012.