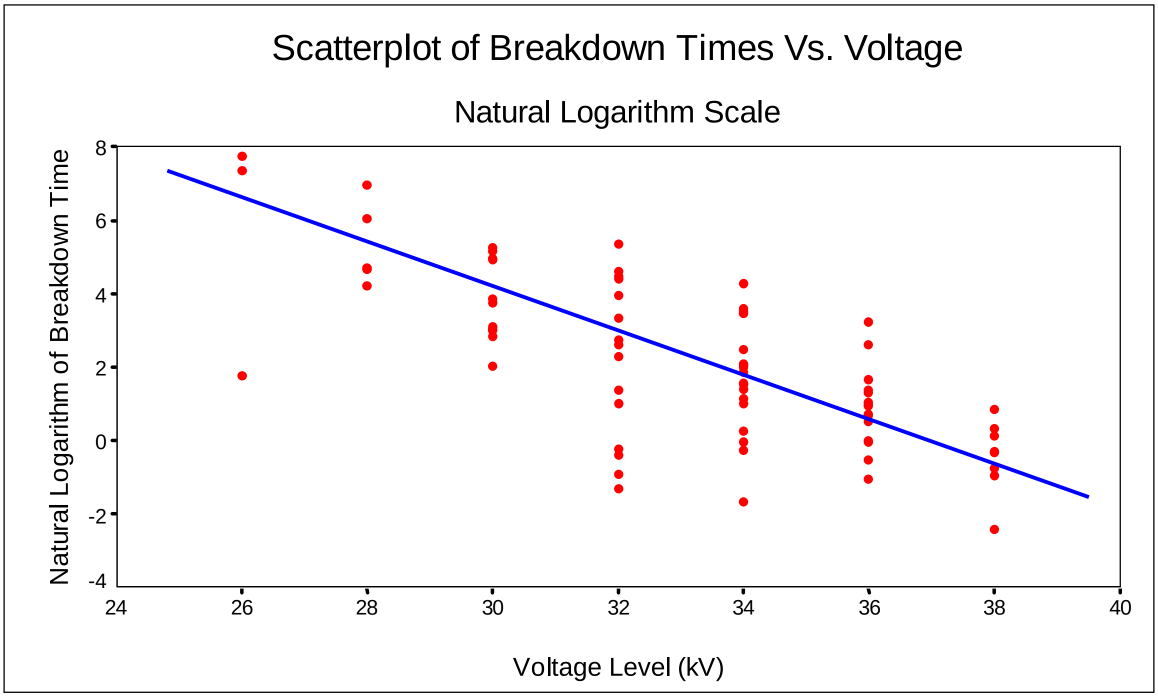
BREAKDOWN TIMES

6. Simple Linear Regression Model

As you have seen before, the scatterplot of the log-transformed breakdown times versus voltage obtained in the previous section displays a linear pattern.



Thus the following simple regression model is suitable:

$$Ln(Time\,|\,Voltage) = \beta_0 + \beta_1 * Voltage + ERROR.$$

Here *Time* denotes the time until breakdown and *Voltage* is the voltage level with possible values at 26, 28, …,38. The random variable *ERROR* is assumed to follow a normal distribution with the mean of zero and an unknown standard deviation σ. The standard deviation is constant at all levels of *Voltage*. The variable *ERROR* follows a normal distribution at each voltage level.

The simple linear regression model can be stated equivalently as follows:

$$\mu\{Ln(Time\,|\,Voltage)\} = \beta_0 + \beta_1 * Voltage.$$

The above model with *Voltage* as a predictor is useful only if the slope β_1 is different from zero. The hypothesis that $\beta_1 = 0$ (the model is useful) can be tested using either t or F tests. The F-statistic is the square of the t-statistic and the corresponding p-values of the two tests are identical.

A quick glance at the scatterplot of the log-transformed breakdown times versus voltage shows that the data provide strong evidence of the utility of the linear regression model.

The SPSS simple linear regression model output for the problem has the following form:

LINEAR REGRESSION

Multiple R	.71667
R Square	.51361
Adjusted R Square	.50704
Standard Error	1.55995

Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	1	190.15149	190.15149
Residual	74	180.07484	2.43344

F = 78.14090 Signif F = .0000

According to the output, the value of the correlation coefficient between breakdown time and voltage is -0.71667. The value of R^2 (0.5136) says that 51.36% of the variation in the log-breakdown times was explained by the linear regression on voltage. The remaining variation was due to some other variables.

We analyze the ANOVA table associated with the simple regression. The sum of squares due to the regression model is reported as 190.15149, and the sum of squares due to error (residual sum of squares) is 180.07484. The residual mean square is an estimate of the variance σ^2 and is equal to 2.4334.

The value of the F statistic is equal to 78.1409 with the corresponding p-value of 0 provides very strong evidence of the utility of the model. This is what we expected by examining the scatterplot of the log-transformed responses.

Now we analyze the part of the output providing the estimates of the regression parameters.

---------------------- **Variables in the Equation** ------------------------------------

Variable	**B**	**SE B**	**95% Confidence Interval B**		**Beta**
VOLTAGE	-.507365	.057396	-.621729	-.393001	-.716665
(Constant)	18.955459	1.910019	15.149663	22.761254	

Variable	**T**	**Sig T**
VOLTAGE	-8.840	.0000
(Constant)	9.924	.0000

According to the output, the estimated regression line of the breakdown time of insulating fluid on voltage is

$$\mu\{Ln(Time\,|\,Voltage)\} = 18.955 - 0.507 * Voltage.$$

The negative sign of the slope is logical because the relationship between the log-breakdown time and voltage is negative. Any predictions of the natural logarithm of the breakdown time based on the above estimated regression line are valid only in the range 26 to 38 kV of voltage.

The estimated regression line was obtained for the log-transformed data. We remember that if the log-transformed responses have a symmetric distribution, then taking the antilogarithm of the slope of the estimated regression line for the log-transformed data, shows a multiplicative change in the median response as the explanatory variable increases by 1 unit.

The slope of -0.507 shows the average change in the natural logarithm of the breakdown time as voltage increases by 1 kV. Thus a one kV increase in voltage is associated with a multiplicative change in median breakdown time of exp(-0.507) = 0.60. So, the median breakdown time at 28 kV is 60% of what it is at 27 kV; the median breakdown time at 29 kV is 60% of what it is at 28 kV.

Since a 95% confidence interval for β_1 is -0.622 to -0.393, a 95% confidence interval for $\exp(\beta_1)$ is exp(-0.622) to exp(-0.393), or 0.54 to 0.68.

If any plots are requested in SPSS, the following summary statistics are displayed for predicted values and residuals at the given voltage levels (*PRED and *RESID) and standardized predicted values and standardized residuals (*ZPRED and *ZRESID). The statistics for our example are given below.

Residuals Statistics

	Min	**Max**	**Mean**	**Std Dev**	**N**
*PRED	-.3244	5.7640	2.1457	1.5923	76
*RESID	-4.0291	2.6513	.0000	1.5495	76
*ZPRED	-1.5513	2.2724	.0000	1.0000	76
*ZRESID	-2.5828	1.6996	.0000	.9933	76

Total Cases = 76

The predicted means, residuals, and standard errors of the predicted means can also be obtained as separate variables in SPSS. In particular, the predicted values (mean estimates) and their standard errors can be obtained in SPSS and added to the data file as the variables pre_1 and sep_1, respectively.

Estimates of Group Means

Voltage (kV)	n	Predicted Values	Standard Error
26	3	5.76397	0.44673
28	5	4.74924	0.34463
30	11	3.73451	0.25362
32	15	2.71978	0.19036
34	19	1.70505	0.18575
36	15	0.69032	0.24315
38	8	-0.32441	0.33181

The magnitude of the standard errors indicates the accuracy of the estimated means. You will see later that the standard errors are smaller than the standard errors obtained from the separate-means model (one-way ANOVA).