Intermediate Statistics Final Report

Section 1:

Table 1 summarizes the describing statistics of each variable in the data set. It displays that each variable’s mean is larger than the median, which means each variable is skewed to the right. Table 2 displays the correlations between each variable. Each pair of variables has a positive correlation, which means that, in a single organism, if there is more than one PCB present, there will be a larger overall number of PCB’s in that organism. Table 3 is the regression analysis of 4 different PCBs. This table, along with a graphical representation, shows that there are outliers present in the data, which are removed for a better, less varied display of the data.

Table 4 displays the data when testing the significance of PCBs. The data, after the removal of PCB180, holds that 97% of the variance from the response variable is from explanatory variables. That data has displayed where the majority of our variance is coming from, however, has not helped us reduce the overall standard error. Table 5 displays the results of successfully reducing the standard error. Using a logarithmic function applied to each observation of each variable, and running the regression on that data greatly reduces the overall standard error (from ~6.4 to ~0.13) making it a much more reliable test.

Table 1: Statistics summary

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Count | Min | Q1 | Median | Mean | Q3 | Max | Std.dev. |
| PCB | 69 | 6.0996 | 30.1830 | 47.9596 | 68.4674 | 91.6305 | 318.7461 | 3.019 |
| PCB52 | 69 | 0.020 | 0.228 | 0.477 | 0.958 | 0.892 | 9.060 | 1.598 |
| PCB118 | 69 | 0.236 | 1.490 | 2.420 | 3.256 | 3.890 | 18.900 | 3.019 |
| PCB138 | 69 | 0.64 | 3.18 | 4.92 | 6.83 | 8.65 | 32.30 | 5.8627 |
| PCB180 | 69 | 0.395 | 1.240 | 2.690 | 4.158 | 4.490 | 31.500 | 4.986 |

Table 2: Variable Correlations

|  |  |  |
| --- | --- | --- |
| Variable 1 | Variable 2 | Correlation |
| PCB | PCB52 | 0.5963572 |
| PCB118 | 0.843298 |
| PCB138 | 0.9288353 |
| PCB180 | 0.8008549 |
| PCB52 | PCB118 | 0.6849073 |
| PCB138 | 0.3008983 |
| PCB180 | 0.08692971 |
| PCB118 | PCB138 | 0.7293792 |
| PCB180 | 0.4374443 |
| PCB138 | PCB180 | 0.8823022 |

Table 3: Regression Analysis before removal of PCB180

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Standard Error | t-value | P(T>|t|) |
| Intercept | 0.9369 | 1.2293 | 0.762 | 0.449 |
| PCB52 | 11.8727 | 0.7290 | 16.287 | <2e-16 |
| PCB118 | 3.7611 | 0.6424 | 5.855 | 1.79e-07 |
| PCB138 | 3.8842 | 0.4978 | 7.803 | 7.19e-11 |
| PCB180 | 4.1823 | 0.4318 | 9.687 | 3.64e-14 |

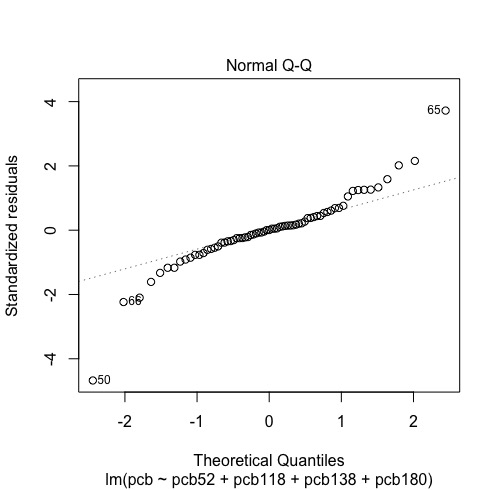


Table 4: Regression Analysis after removal of PCB180

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Standard Error | t-value | P(T>|t|) |
| Intercept | -1.02 | 1.8895 | -0.539 | 0.592 |
| PCB52 | 12.6442 | 1.1291 | 11.198 | <2e-16 |
| PCB118 | 0.3131 | 0.8333 | 0.376 | 0.708 |
| PCB138 | 8.2546 | 0.3279 | 25.177 | <2e-16 |

Table 5: Effect of Logs

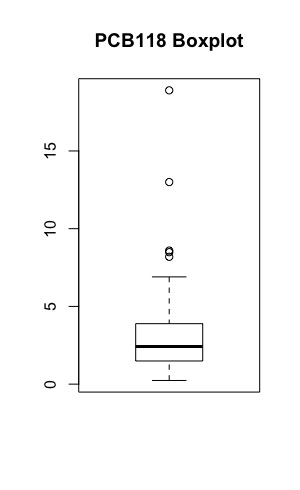
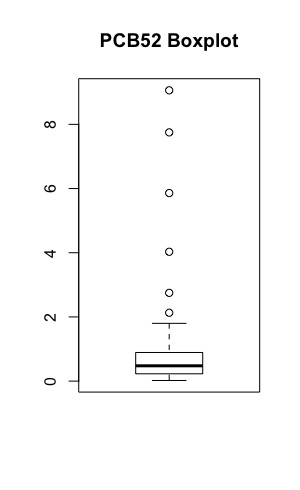
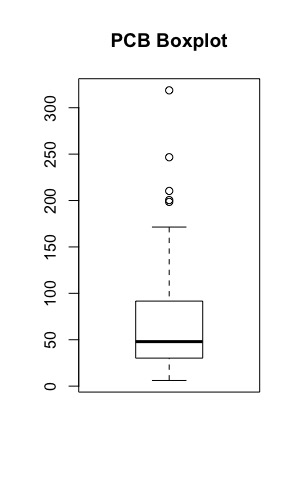
|  |  |  |  |
| --- | --- | --- | --- |
| S | R-sq | R-sq (adj) | R-sq (pred) |
| 0.134974 | 97.46% | 97.17% | 96.27% |

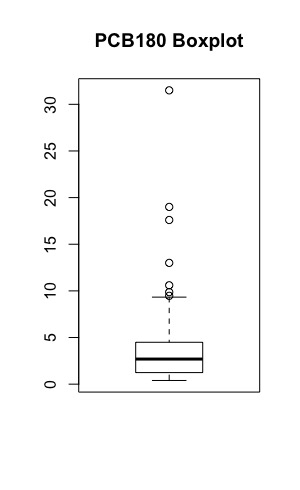
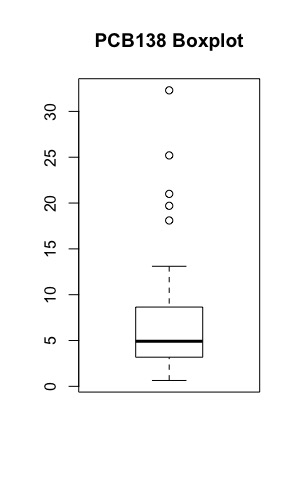
Section 2:

11.42: Relationships among PCB congeners.

a)

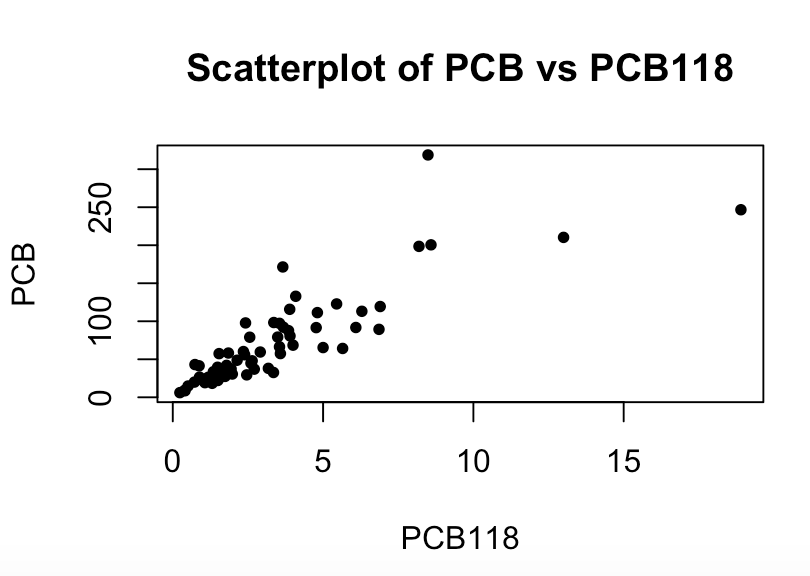
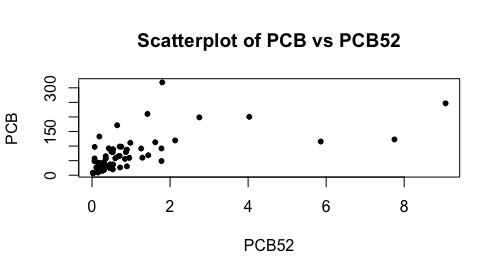
|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Count | Mean | Std.dev. | Min | Q1 | Median | Q3 | Max |
| PCB | 69 | 68.4674 | 3.019 | 6.0996 | 30.1830 | 47.9596 | 91.6305 | 318.7461 |
| PCB52 | 69 | 0.958 | 1.598 | 0.020 | 0.228 | 0.477 | 0.892 | 9.060 |
| PCB118 | 69 | 3.256 | 3.019 | 0.236 | 1.490 | 2.420 | 3.890 | 18.900 |
| PCB138 | 69 | 6.83 | 5.8627 | 0.64 | 3.18 | 4.92 | 8.65 | 32.30 |
| PCB180 | 69 | 4.158 | 4.986 | 0.395 | 1.240 | 2.690 | 4.490 | 31.500 |

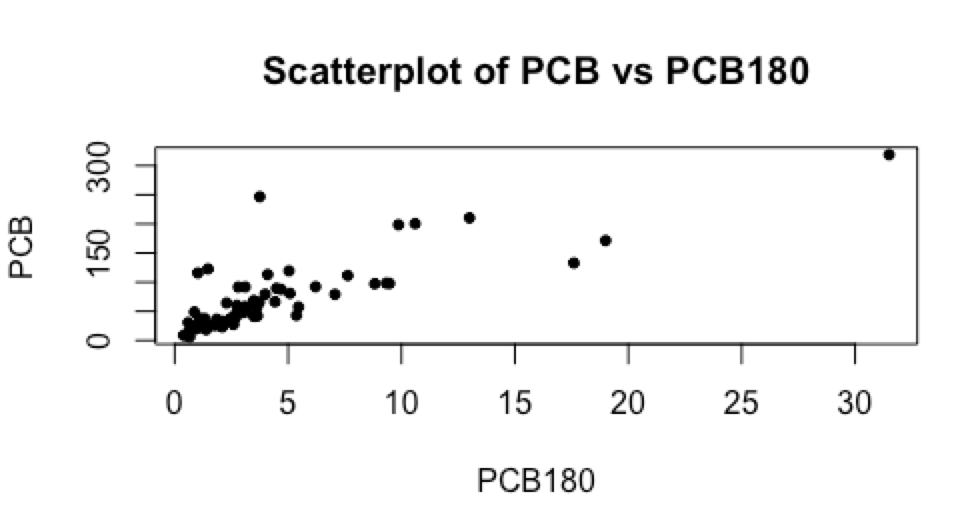
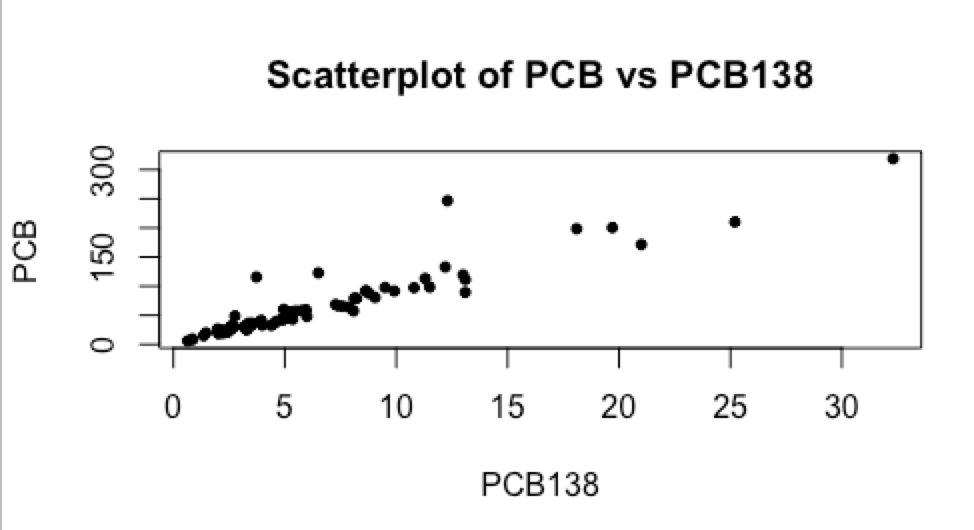


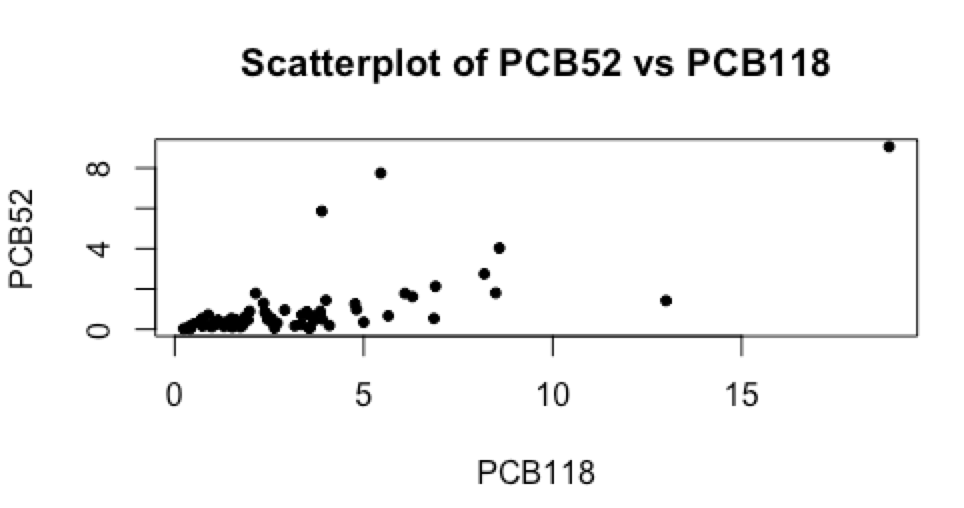


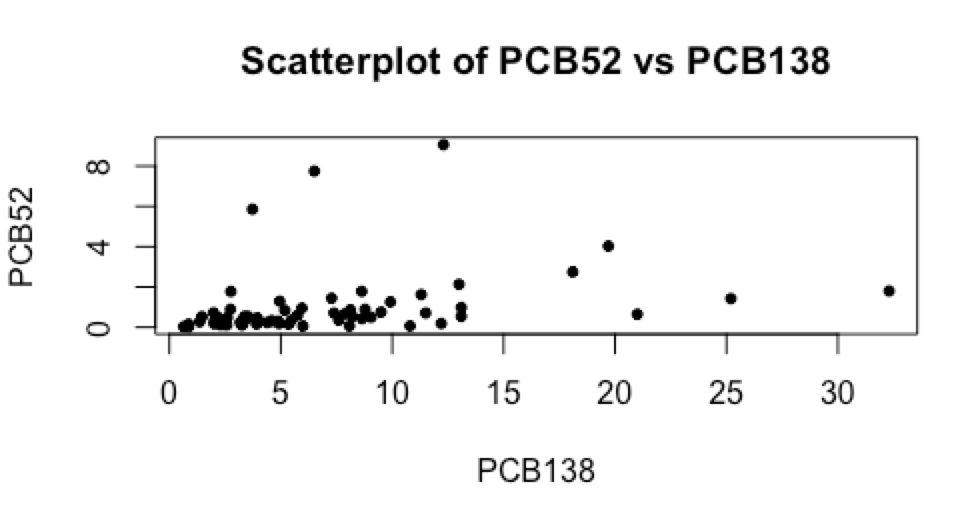
b)

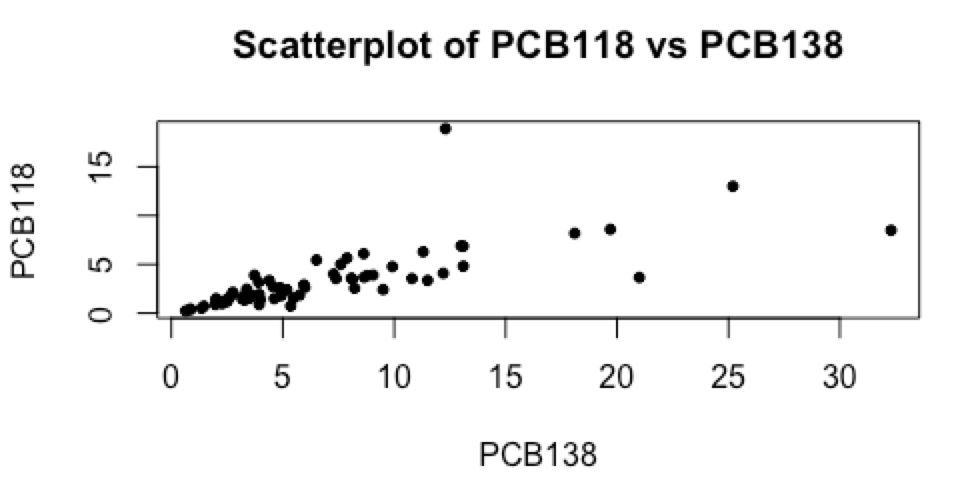
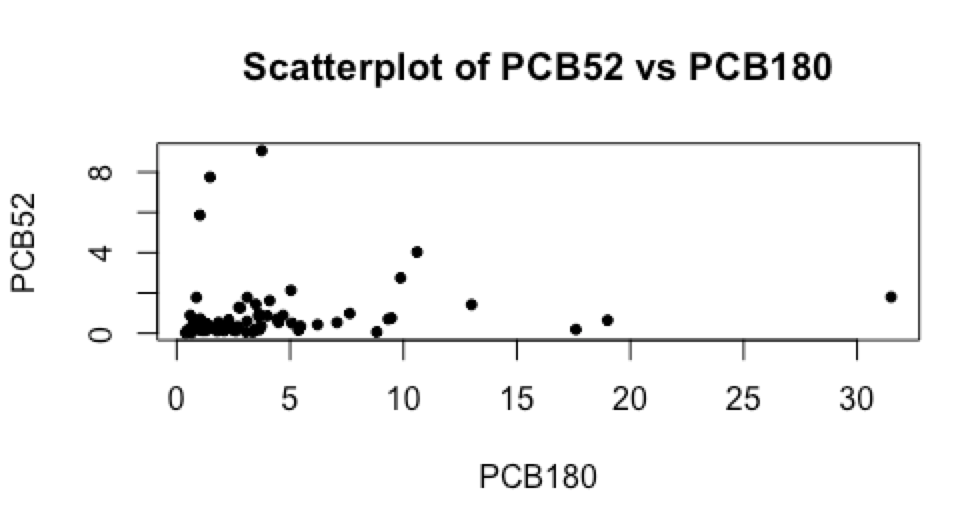
|  |  |  |
| --- | --- | --- |
| Variable 1 | Variable 2 | Correlation |
| PCB | PCB52 | 0.5963572 |
| PCB118 | 0.843298 |
| PCB138 | 0.9288353 |
| PCB180 | 0.8008549 |
| PCB52 | PCB118 | 0.6849073 |
| PCB138 | 0.3008983 |
| PCB180 | 0.08692971 |
| PCB118 | PCB138 | 0.7293792 |
| PCB180 | 0.4374443 |
| PCB138 | PCB180 | 0.8823022 |

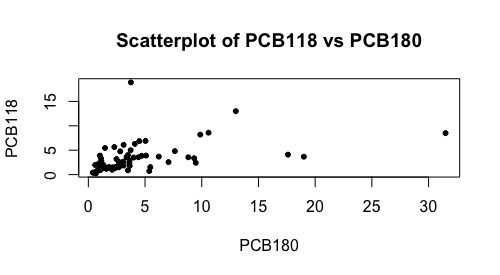


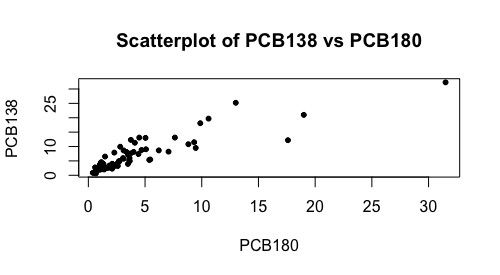












11.43: Predicting the total amount of PCB.

a) Statistical Model:

p = 4, n = 69

Explanatory variables: x­1, x2, x3, x4

yi = β0 + β1xi1 + β2xi2 + β3xi3 + β4xi4 + εi i = 1, 2, …, 69

Assumptions:

The deviations εi about the mean are independent and normally distributed N(0, σ). The standard deviation does not depend on the values of the explanatory variables.

Parameters: β0 , β1 , β2 , β3 , β4

b) Regression:

Response variable: PCB

Explanatory variables: PCB52, PCB118, PCB138, PCB180

Residuals:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Min | Q1 | Median | Q3 | Max |
| -22.0864 | -2.4554 | 0.0278 | 2.7726 | 22.5487 |

Coefficients:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Standard Error | t-value | P(T>|t|) |
| Intercept | 0.9369 | 1.2293 | 0.762 | 0.449 |
| PCB52 | 11.8727 | 0.7290 | 16.287 | <2e-16 |
| PCB118 | 3.7611 | 0.6424 | 5.855 | 1.79e-07 |
| PCB138 | 3.8842 | 0.4978 | 7.803 | 7.19e-11 |
| PCB180 | 4.1823 | 0.4318 | 9.687 | 3.64e-14 |

Residual standard error: 6.382 on 64 degrees of freedom

Multiple R-squared: 0.9891, Adjusted R-squared: 0.9885

F-statistic: 1456 on 4 and 64 DF, p-value: < 2.2e-16

Multiple Regression equation:

Interpretation:

Intercept term: 0.94 (fixed amount of PCB when all four congeners are 0)

PCB52: total amount of PCB is expected to increase by 11.87 percent when PCB52 is increased by 1.

PCB118: total amount of PCB is expected to increase by 3.76 percent when PCB118 is increased by 1.

PCB138: total amount of PCB is expected to increase by 3.88 percent when PCB138 is increased by 1.

PCB180: total amount of PCB is expected to increase by 4.18 percent when PCB180 is increased by 1.

Test: α = 0.05

Null Hypothesis: H0: β0 = β1 = β2 = β3 = β4  = 0

Alternate Hypothesis: Ha: at least one βi does not equal 0.

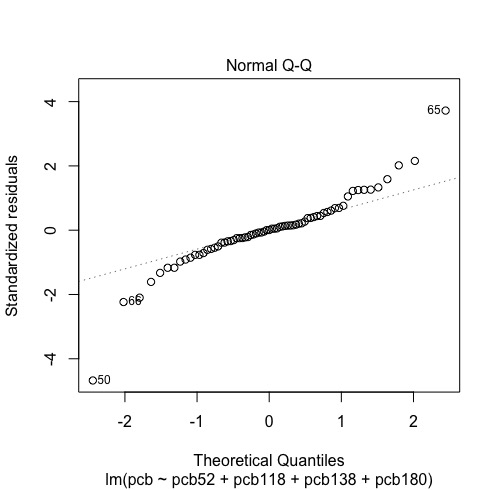
ANOVA:

Response: pcb

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | df | Sum of Squares | Mean of Squares | F-value | P(F>f) |
| PCB52 | 1 | 85302 | 85302 | 2094.273 | < 2.2e-16 |
| PCB118 | 1 | 85429 | 85429 | 2097.405 | < 2.2e-16 |
| PCB138 | 1 | 62693 | 62693 | 1539.202 | < 2.2e-16 |
| PCB180 | 1 | 3822 | 3822 | 93.834 | 3.64e-14 |
| Residual | 64 | 2607 | 41 |  |  |

Since the P-value is less than the level of significance, null hypothesis is rejected. All of the regression coefficients are significantly different from 0, except for the constant. R2 = 0.989 which means 98.9% of the total PCB can be explained by the explanatory variables it the regression equation. The rest is unknown. The residual standard error is 6.382

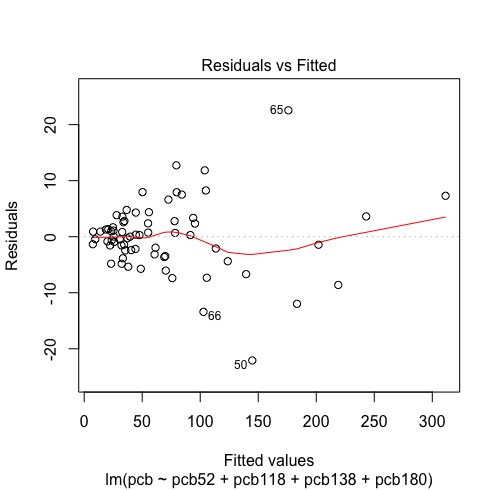
c) Results:



The residuals look normal, with 2 outliers. There are no specific patterns when the residuals are plotted against each explanatory variable; PCB52,PCB118, PCB138, PCB180.

11.44: Adjusting the analysis for potential outliers.

a)



There are no patterns here, but there are 2 outliers. The low residual at line 50, and the high residual at line 65. The overestimate would be the one at line 65.

b) The new multiple linear regression equation is:

Residuals:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Min | Q1 | Median | Q3 | Max |
| -12.2421 | -2.1762 | -0.1378 | 1.7036 | 14.2051 |

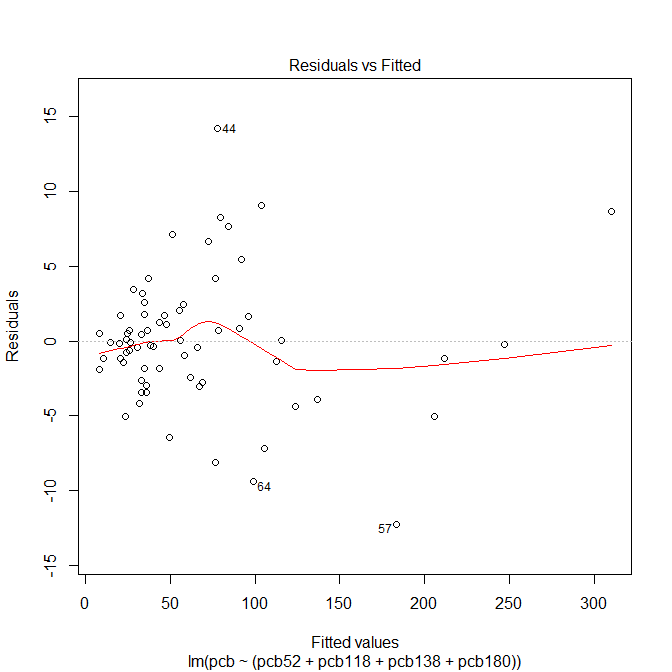
Coefficients:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Standard error | t-value | P(T>|t|) |
| Intercept | 1.6277 | 0.8858 | 1.838 | 0.0709 |
| PCB52 | 14.4420 | 0.6960 | 20.751 | < 2e-16 |
| PCB118 | 2.5996 | 0.5164 | 5.034 | 4.40e-06 |
| PCB138 | 4.0541 | 0.3752 | 10.805 | 6.89e-16 |
| PCB180 | 4.1086 | 0.3175 | 12.942 | < 2e-16 |

Residual standard error: 4.555 on 62 degrees of freedom

Multiple R-squared: 0.9941, Adjusted R-squared: 0.9938

F-statistic: 2629 on 4 and 62 DF, p-value: < 2.2e-16



When the outliers are removed, the residual standard error is reduced from 6.382 to 4.555. Within the table, the errors for the explanatory variables have also decreased.

11.45: More on predicting the total amount of PCB.

a) The multiple linear regression equation for:

Response Variable: PCB

Explanatory Variables: PCB52, PCB118, PCB138

Residuals:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Min | Q1 | Median | Q3 | Max |
| -29.6219 | -3.3502 | 0.8791 | 3.3785 | 29.5217 |

Coefficients:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Standard Error | t-value | P(T>|t|) |
| Intercept | -1.02 | 1.8895 | -0.539 | 0.592 |
| PCB52 | 12.6442 | 1.1291 | 11.198 | <2e-16 |
| PCB118 | 0.3131 | 0.8333 | 0.376 | 0.708 |
| PCB138 | 8.2546 | 0.3279 | 25.177 | <2e-16 |

Residual standard error: 9.945 on 65 degrees of freedom

Multiple R-squared: 0.9732, Adjusted R-squared: 0.972

F-statistic: 786.7 on 3 and 65 DF, p-value: < 2.2e-16

Interpretation:

Intercept term: -1.02 (fixed amount of PCB when all four congeners are 0)

PCB52: total amount of PCB is expected to increase by 12.64 percent when PCB52 is increased by 1.

PCB118: total amount of PCB is expected to increase by 0.31 percent when PCB118 is increased by 1.

PCB138: total amount of PCB is expected to increase by 8.25 percent when PCB138 is increased by 1.

b) Significance level: α = 0.05

Coefficient for PCB118 = 0.313

p-val for PCB118 = 0.708

Significance Test: reject if p>α

P = 0.708 > α. Therefore, reject.

c) In 11.43:

Coefficient for PCB118 = 3.761

p-val for PCB118 = 1.79e-07

Significance test: reject if p>α

P = 1.79e-07 < α.

d) Excluding a variable in a multiple regression analysis can drastically change the results. The p-value of PCB118 in part A is much larger than the original analysis, and therefore is not significant, while the original analysis has a much smaller p-value, making it more significant.

11.46: Multiple regression model for total TEQ:

a)

β0 = 0, β1 = β2 = β3 = 1

b) σ = 7.95e-06

c) Residuals:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Min | Q1 | Median | Q3 | Max |
| -5.638e-06 | -2.844e-06 | -1.680e-06 | -1.130e-06 | 3.714e-05 |

Coefficients:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Standard Error | t-value | P(T>|t|) |
| Intercept | 3.426e-07 | 1.917e-06 | 1.790e-01 | 0.859 |
| TEQPCB | 1 | 8.239e-07 | 1.214e+06 | <2e-16 |
| TEQDIOXIN | 1 | 1.761e-06 | 5.677e+05 | <2e-16 |
| TEQFURAN | 1 | 5.664e-06 | 1.766e+05 | <2e-16 |

Residual standard error: 7.95e-06 on 65 degrees of freedom

Multiple R-squared: 1, Adjusted R-squared: 1

F-statistic: 9.581e+11 on 3 and 65 DF, p-value: < 2.2e-16

Interpretation:

Intercept term: 0 (fixed amount of TEQ when all four congeners are 0)

TEQPCB: total amount of TEQ is expected to increase by 1 percent when TEQPCB is increased by 1.

TEQDIOXIN: total amount of TEQ is expected to increase by 1 percent when TEQDIIOXIN is increased by 1.

TEQFURAN: total amount of TEQ is expected to increase by 1 percent when TEQFURAN is increased by 1.

Test: α = 0.05

Null Hypothesis: H0: β0 = β1 = β2 = β3 = β4  = 0

Alternate Hypothesis: Ha: at least one βj does not equal 0.

ANOVA:

Response: teq

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | df | Sum of Squares | Mean of Squares | F-value | P(F>f) |
| TEQPCB | 1 | 152.801 | 152.801 | 2.4174e+12 | < 2.2e-16 |
| TEQDIOXIN | 1 | 26.903 | 26.903 | 4.2562e+11 | < 2.2e-16 |
| TEQFURAN | 1 | 1.970 | 1.970 | 3.1174e+10 | < 2.2e-16 |
| Residual | 65 | 0 | 0 |  |  |

Since the P-value is less than the level of significance, null hypothesis is rejected.

All of the regression coefficients are significantly different from 0, except for the constant.

R2 = 1 which means 100% of the total TEQ can be explained by the explanatory variables in the regression equation. The residual standard error is 7.95e-06.

11.47: Multiple regression model for total TEQ continued.

a) Residuals:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Min | Q1 | Median | Q3 | Max |
| -1.6655 | -0.6000 | -0.1814 | 0.5162 | 2.7025 |

Coefficients:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Standard error | t-value | P(T>|t|) |
| Intercept | 1.059965 | 0.184450 | 5.747 | 2.73e-07 |
| PCB52 | -0.097277 | 0. 109383 | -0.889 | 0.37716 |
| PCB118 | 0.306184 | 0.096388 | 3.177 | 0.00229 |
| PCB138 | 0.105786 | 0.074697 | 1.416 | 0.16156 |
| PCB180 | -0.003905 | 0.064784 | -0.060 | 0.95212 |

Residual standard error: 0.9576 on 64 degrees of freedom

Multiple R-squared: 0.6769, Adjusted R-squared: 0.6568

F-statistic: 33.53 on 4 and 64 DF, p-value: 4.489e-15

Interpretation:

Intercept term: 1.06 (fixed amount of PCB when all four congeners are 0)

PCB52: total amount of PCB is expected to decrease by 0.01 percent when PCB52 is increased by 1.

PCB118: total amount of PCB is expected to increase by 0.31 percent when PCB118 is increased by 1.

PCB138: total amount of PCB is expected to increase by 0.106 percent when PCB138 is increased by 1.

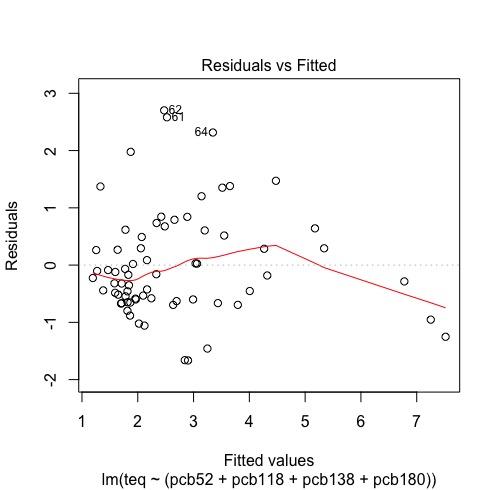
PCB180: total amount of PCB is expected to decrease by 0.004 percent when PCB180 is increased by 1.

ANOVA:

Response: teq

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | df | Sum of Squares | Mean of Squares | F-value | P(F>f) |
| PCB52 | 1 | 29.85 | 29.85 | 32.553 | 3.21e-07 |
| PCB118 | 1 | 83.61 | 83.61 | 91.174 | 6.30e-14 |
| PCB138 | 1 | 9.52 | 9.52 | 10.378 | 0.00201 |
| PCB180 | 1 | 0 | 0 | 0.004 | 0.95212 |
| Residual | 64 | 58.69 | 0.92 |  |  |

R2 = .6769, which means that 67.69% of TEQ is explained by the explanatory variables. The rest is unknown. The residual standard error is 0.9576.



The residuals are skewed to the right. When plotted against the explanatory variables, there are no clear patterns visible.

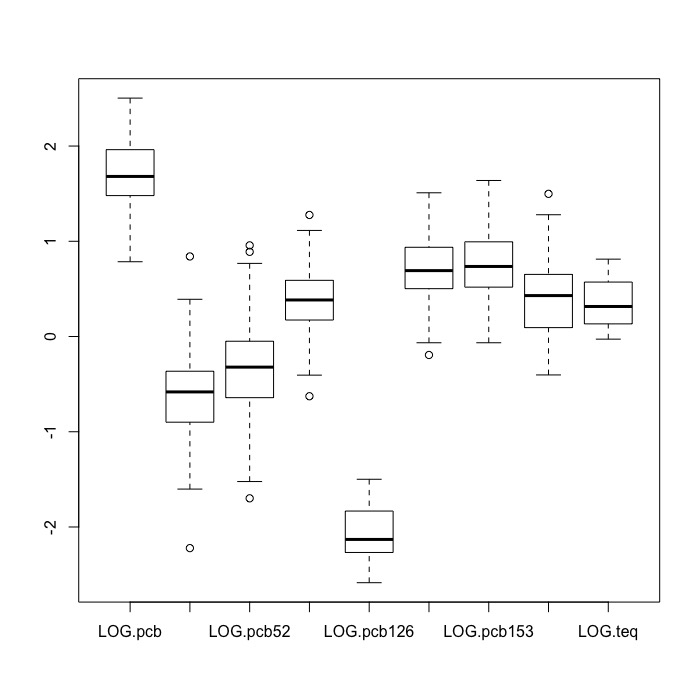
11.48: Predicting total amount of PCB using transformed variables.

a) If you do not do anything about the 16 zero values of PCB126, the software will change it to “#NUM!”

b) If you attempt to run a regression to predict the log of PCB using the log of PCB126 and PCB52 without accounting for the zero values, the software will not run the regression on “#NUM!” because it is not a numerical value. The software will throw an error rather than ignore them. It is not a good way of handling the situation, it is inconvenient to refuse those values when there is a large amount of 0 values.

c) Numerical and graphical representation of the data.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Min | Q1 | Median | Mean | Q3 | Max |
| LOG.pcb | 0.7853 | 1.4798 | 1.6809 | 1.7011 | 1.9620 | 2.5034 |
| LOG.pcb28 | -2.2218 | -0.8996 | -0.5817 | -0.5793 | -0.3645 | 0.8407 |
| LOG.pcb52 | -1.69897 | -0.64207 | -0.32148 | -0.33537 | -0.04964 | 0.95713 |
| LOG.pcb118 | -0.6271 | 0.1732 | 0.3838 | 0.3717 | 0.5899 | 1.2765 |
| LOG.pcb126 | -2.585 | -2.268 | -2.131 | -2.104 | -1.833 | -1.499 |
| LOG.pcb138 | -0.1938 | 0.5024 | 0.6920 | 0.7009 | 0.9370 | 1.5092 |
| LOG.pcb153 | -0.0655 | 0.5185 | 0.7356 | 0.7397 | 0.9943 | 1.6385 |
| LOG.pcb180 | -0.40340 | 0.09342 | 0.42975 | 0.42354 | 0.65225 | 1.49831 |
| LOG.teq | -0.02761 | 0.13274 | 0.31534 | 0.34950 | 0.57174 | 0.81245 |

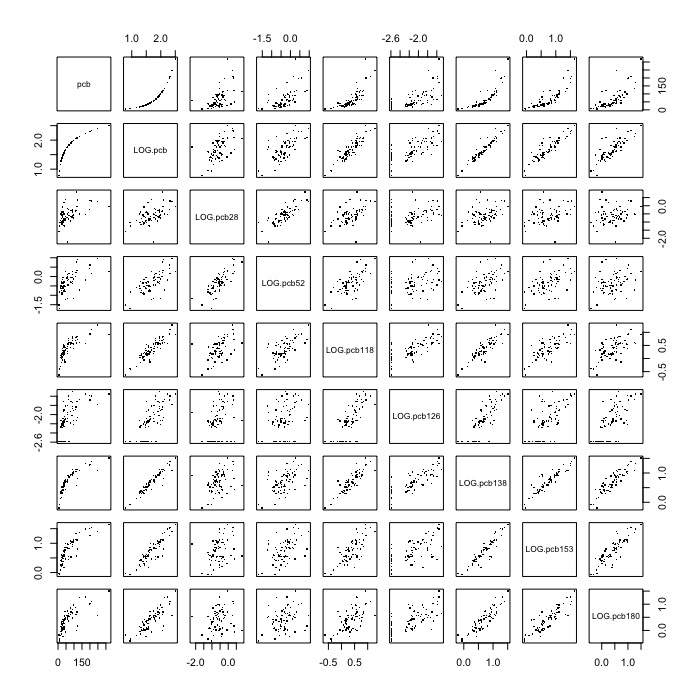


11.49: Predicting total amount of PCB using transformed variables, continued.

a) Each pair of values has a positive correlation. The only variable that has any outliers is LOG.pcb28, however, each scatterplot related to the transformed variables is linearly correlated.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | pcb | LOG.pcb | LOG.pcb28 | LOG.pcb52 | LOG.pcb118 |
| pcb | 1.0000000 | 0.8861681 | 0.5406772 | 0.6314956 | 0.7857074 |
| LOG.pcb | 0.8861681 | 1.0000000 | 0.5699256 | 0.7005905 | 0.9064775 |
| LOG.pcb28 | 0.5406772 | 0.5699256 | 1.0000000 | 0.7950316 | 0.5336685 |
| LOG.pcb52 | 0.6314956 | 0.7005905 | 0.7950316 | 1.0000000 | 0.6709082 |
| LOG.pcb118 | 0.7857074 | 0.9064775 | 0.5336685 | 0.6709082 | 1.0000000 |
| LOG.pcb126 | 0.6518593 | 0.7292267 | 0.2721924 | 0.3308594 | 0.7394002 |
| LOG.pcb138 | 0.8278464 | 0.9560549 | 0.3876895 | 0.5404601 | 0.8897442 |
| LOG.pcb153 | 0.8286488 | 0.9049176 | 0.3260234 | 0.5192283 | 0.7798756 |
| LOG.pcb180 | 0.7447056 | 0.8288974 | 0.2272701 | 0.3015365 | 0.6538711 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | LOG.pcb126 | LOG.pcb138 | LOG.pcb153 | LOG.pcb180 |
| pcb | 0.6518593 | 0.8278464 | 0.8286488 | 0.7447056 |
| LOG.pcb | 0.7292267 | 0.9560549 | 0.9049176 | 0.8288974 |
| LOG.pcb28 | 0.2721924 | 0.3876895 | 0.3260234 | 0.2272701 |
| LOG.pcb52 | 0.3308594 | 0.5404601 | 0.5192283 | 0.3015365 |
| LOG.pcb118 | 0.7394002 | 0.8897442 | 0.7798756 | 0.6538711 |
| LOG.pcb126 | 1.0000000 | 0.7923915 | 0.6465768 | 0.6954466 |
| LOG.pcb138 | 0.7923915 | 1.0000000 | 0.9219441 | 0.8963662 |
| LOG.pcb153 | 0.6465768 | 0.9219441 | 1.0000000 | 0.8668080 |
| LOG.pcb180 | 0.6954466 | 0.8963662 | 0.8668080 | 1.0000000 |

d

Compared to the correlations from 11.42, each pair in each question has a positive correlation. However, the correlations are, in some cases, much higher than the original corrlations.

11.50: Even more on predicting the total amount of PCB using transformed variables.

a) Regression Equation:

Residuals:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Min | Q1 | Median | Q3 | Max |
| -2.3366 | -0.7636 | -0.1818 | 0.5010 | 2.2728 |

Coefficients:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Standard Error | t-value | P(T>|t|) |
| Intercept | 1.51847 | 2.87722 | 0.528 | 0.5996 |
| LN.pcb | -0.04296 | 0.92888 | -0.046 | 0.9633 |
| LN.pcb52 | 0.03077 | 0.22590 | 0.136 | 0.8921 |
| LN.pcb118 | 1.20625 | 0.54254 | 2.223 | 0.0298 |
| LN.pcb138 | 0.70015 | 0.98289 | 0.712 | 0.4789 |
| LN.pcb153 | -0.71591 | 0.40666 | -1.760 | 0.0833 |
| LN.pcb180 | 0.41045 | 0.50083 | 0.820 | 0.4156 |

Residual standard error: 1.069 on 62 degrees of freedom

Multiple R-squared: 0.6103, Adjusted R-squared: 0.5726

F-statistic: 16.18 on 6 and 62 DF, p-value: 4.207e-11

Interpretation:

Intercept term: 1.52 (fixed amount of PCB when all other congeners are 0)

LN.PCB: total amount of PCB is expected to decrease by 0.043 percent when PCB52 is increased by 1.

LN.PCB52: total amount of PCB is expected to increase by 0.030 percent when PCB52 is increased by 1.

LN.PCB118: total amount of PCB is expected to increase by 1.21 percent when PCB118 is increased by 1.

LN.PCB138: total amount of PCB is expected to increase by 0.700 percent when PCB138 is increased by 1.

LN.PCB153: total amount of PCB is expected to decrease by 0.716 percent when PCB153 is increased by 1.

LN.PCB180: total amount of PCB is expected to increase by 0.410 percent when PCB180 is increased by 1.

ANOVA:

Response: teq

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | df | Sum of Squares | Mean of Squares | F-value | P(F>f) |
| LN.pcb | 1 | 92.444 | 92.444 | 80.9532 | 7.508e-13 |
| LN.pcb52 | 1 | 0.344 | 0.344 | 0.3008 | 0.585345 |
| LN.pcb118 | 1 | 13.205 | 13.205 | 11.5640 | 0.001181 |
| LN.pcb138 | 1 | 0.757 | 0.757 | 0.6625 | 0.418782 |
| LN.pcb153 | 1 | 3.358 | 3.358 | 2.9403 | 0.091391 |
| LN.pcb180 | 1 | 0.767 | 0.767 | 0.6717 | 0.415613 |
| Residuals | 62 | 70.800 | 1.142 |  |  |

R2 = .6103, which means that 61.03% of TEQ is explained by the explanatory variables. The rest is unknown. The residual standard error is 1.069.

11.51: Predicting total TEQ using transformed variables.

a) Regression Equation:

Residuals:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Min | Q1 | Median | Q3 | Max |
| -0.23924 | -0.07739 | 0.01298 | 0.06054 | 0.41781 |

Coefficients:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Standard Error | t-value | P(T>|t|) |
| Intercept | 1.72132 | 0.09879 | 17.424 | < 2e-16 |
| LOG.pcb126 | 0.62222 | 0.04801 | 12.960 | < 2e-16 |
| LOG.pcb28 | 0.10770 | 0.03246 | 3.318 | 0.00148 |

Residual standard error: 0.1267 on 66 degrees of freedom

Multiple R-squared: 0.7681, Adjusted R-squared: 0.7611

F-statistic: 109.3 on 2 and 66 DF, p-value: < 2.2e-16

Interpretation:

Intercept term: 1.72 (fixed amount of PCB when all other congeners are 0)

LN.PCB28: total amount of PCB is expected to increase by 0.622 percent when PCB52 is increased by 1.

LN.PCB126: total amount of PCB is expected to increase by 0.108 percent when PCB180 is increased by 1.

ANOVA:

Response: LOG.teq

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | df | Sum of Squares | Mean of Squares | F-value | P(F>f) |
| LOG.pcb126 | 1 | 3.330 | 3.330 | 207.58 | < 2e-16 |
| LOG.pcb28 | 1 | 0.177 | 0.177 | 11.01 | 0.00148 |
| Residuals | 66 | 1.059 | 0.016 |  |  |

R2 = .7681, which means that 76.81% of TEQ is explained by the variables PCB126 and PCB28. The rest is unknown. The residual standard error is 0.1267.

11.52: Interpretation of coefficients in log PCB regressions.

a) Regression Equation:

Residuals:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Min | Q1 | Median | Q3 | Max |
| -0.23310 | -0.07925 | 0.00317 | 0.06473 | 0.43508 |

Coefficients:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Standard error | t-value | P(T>|t|) |
| Intercept | 1.60616 | 0.23982 | 6.697 | 7.65e-09 |
| LOG.pcb28 | 0.04508 | 0.05626 | 0.801 | 0.426 |
| LOG.pcb52 | 0.04209 | 0.06483 | 0.649 | 0.519 |
| LOG.pcb118 | 0.19173 | 0.14548 | 1.318 | 0.192 |
| LOG.pcb126 | 0.56299 | 0.08430 | 6.678 | 8.25e-09 |
| LOG.pcb138 | -0.08939 | 0.27753 | -0.322 | 0.748 |
| LOG.pcb153 | -0.09030 | 0.11660 | -0.774 | 0.442 |
| LOG.pcb180 | 0.06266 | 0.13489 | 0.465 | 0.644 |

Residual standard error: 0.1277 on 61 degrees of freedom

Multiple R-squared: 0.7822, Adjusted R-squared: 0.7572

F-statistic: 31.29 on 7 and 61 DF, p-value: < 2.2e-16

Interpretation:

Intercept term: 1.61 (fixed amount of PCB when all other congeners are 0)

LN.PCB28: total amount of PCB is expected to decrease by 0.045 percent when PCB52 is increased by 1.

LN.PCB52: total amount of PCB is expected to increase by 0.042 percent when PCB52 is increased by 1.

LN.PCB118: total amount of PCB is expected to increase by 0.192 percent when PCB118 is increased by 1.

LN.PCB126: total amount of PCB is expected to increase by 0.563 percent when PCB138 is increased by 1.

LN.PCB138: total amount of PCB is expected to decrease by 0.089 percent when PCB153 is increased by 1.

LN.PCB180: total amount of PCB is expected to decrease by 0.090 percent when PCB180 is increased by 1.

LN.PCB180: total amount of PCB is expected to increase by 0.063 percent when PCB180 is increased by 1.

Test:

Null Hypothesis: H0 : βj = 0 where j = 1, 2, 3, 4, 5, 6

Alternate Hypothesis: Ha : βj ≠ 0

All of the regression coefficients are significantly different from 0, except for the constant, so we reject H0.

ANOVA:

Response: LOG.teq

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | df | Sum of Squares | Mean of Squares | F-value | P(F>f) |
| LOG.pcb28 | 1 | 0.81201 | 0.81201 | 49.8069 | 1.846e-09 |
| LOG.pcb52 | 1 | 0.20151 | 0.20151 | 12.3601 | 0.0008338 |
| LOG.pcb118 | 1 | 1.61853 | 1.61853 | 99.2774 | 2.048e-14 |
| LOG.pcb126 | 1 | 0.91372 | 0.91372 | 56.0458 | 3.372e-10 |
| LOG.pcb138 | 1 | 0.01388 | 0.01388 | 0.8514 | 0.3597981 |
| LOG.pcb153 | 1 | 0.00776 | 0.00776 | 0.4757 | 0.4929960 |
| LOG.pcb180 | 1 | 0.00352 | 0.00352 | 0.2158 | 0.6439272 |
| Residuals | 61 | 0.99449 | 0.01630 |  |  |

R2 = .7822, which means that 78.22% of TEQ is explained by the explanatory variables. The rest is unknown. The residual standard error is 0.1277.

Our original calculations stated that the standard error was 6.382. This means that taking the logarithm of PCB when all log congener PCB variables are considered because it led to an extreme reduction in error.