**Forward Selection**

Import data set

|  |
| --- |
| fdata <- read\_excel("F:/Christ/lab-LRM-5CMS and EMS/forward selection-data.xlsx")  > View(fdata)  Run the full model  > fullmodel=lm(y~., data=fdata)  > formula(fullmodel)  y ~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9  > summary(fullmodel)  Call:  lm(formula = y ~ ., data = fdata)  Residuals:  Min 1Q Median 3Q Max  -3.8504 -1.4017 0.0929 1.7541 3.7206  Coefficients:  Estimate Std. Error t value Pr(>|t|)  (Intercept) 17.11351 5.88549 2.908 0.0131 \*  x1 2.39009 1.05740 2.260 0.0432 \*  x2 5.74422 4.35113 1.320 0.2114  x3 0.12998 0.52530 0.247 0.8087  x4 2.63623 4.34493 0.607 0.5553  x5 2.32382 1.46160 1.590 0.1378  x6 -1.62471 2.40137 -0.677 0.5115  x7 -0.09723 3.38794 -0.029 0.9776  x8 -0.04445 0.06212 -0.716 0.4879  x9 2.03656 1.97372 1.032 0.3225  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  Residual standard error: 2.841 on 12 degrees of freedom  Multiple R-squared: 0.8774, Adjusted R-squared: 0.7854  F-statistic: 9.539 on 9 and 12 DF, p-value: 0.0003125  Forward selection (Start with no regressors)  > f1=lm(y~1, data=fdata)  > summary(f1)  Call:  lm(formula = y ~ 1, data = fdata)  Residuals:  Min 1Q Median 3Q Max  -9.095 -5.071 1.405 3.655 10.805  Coefficients:  Estimate Std. Error t value Pr(>|t|)  (Intercept) 34.995 1.308 26.77 <2e-16 \*\*\*  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  Residual standard error: 6.133 on 21 degrees of freedom  > step(f1, direction="forward", scope=formula(fullmodel))  Start: AIC=80.78  y ~ 1  Df Sum of Sq RSS AIC  + x1 1 616.67 173.16 49.390  + x2 1 386.12 403.71 68.012  + x4 1 385.20 404.63 68.062  + x3 1 339.51 450.32 70.416  + x6 1 208.00 581.83 76.053  + x5 1 199.58 590.25 76.369  + x8 1 132.88 656.95 78.725  <none> 789.83 80.777  + x7 1 57.89 731.94 81.103  + x9 1 43.18 746.65 81.540  Step: AIC=49.39  y ~ x1  Df Sum of Sq RSS AIC  + x2 1 22.9619 150.20 48.260  + x9 1 15.9875 157.17 49.259  <none> 173.16 49.390  + x4 1 7.8167 165.34 50.374  + x5 1 5.6693 167.49 50.657  + x6 1 3.5583 169.60 50.933  + x3 1 3.2496 169.91 50.973  + x7 1 2.4360 170.72 51.078  + x8 1 1.7536 171.41 51.166  Step: AIC=48.26  y ~ x1 + x2  Df Sum of Sq RSS AIC  <none> 150.20 48.260  + x9 1 11.3028 138.90 48.539  + x7 1 8.3644 141.83 49.000  + x5 1 7.6678 142.53 49.107  + x6 1 6.5813 143.62 49.274  + x8 1 6.2771 143.92 49.321  + x3 1 4.2757 145.92 49.625  + x4 1 0.3311 149.87 50.212  Call:  lm(formula = y ~ x1 + x2, data = fdata)  Coefficients:  (Intercept) x1 x2  9.321 2.923 5.550  > cm=lm(y~x1+x2, data=fdata)  > summary(cm)  Call:  lm(formula = y ~ x1 + x2, data = fdata)  Residuals:  Min 1Q Median 3Q Max  -4.8510 -2.0998 0.0266 1.3604 5.1215  Coefficients:  Estimate Std. Error t value Pr(>|t|)  (Intercept) 9.3207 3.1373 2.971 0.00785 \*\*  x1 2.9232 0.5162 5.663 1.85e-05 \*\*\*  x2 5.5497 3.2563 1.704 0.10462  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  Residual standard error: 2.812 on 19 degrees of freedom  Multiple R-squared: 0.8098, Adjusted R-squared: 0.7898  F-statistic: 40.46 on 2 and 19 DF, p-value: 1.418e-07 |
|  |
| |  | | --- | |  | |

**Backward Elimination**

|  |
| --- |
| step(fullmodel, direction="backward")  Start: AIC=52.61  y ~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9  Df Sum of Sq RSS AIC  - x7 1 0.007 96.868 50.611  - x3 1 0.494 97.355 50.721  - x4 1 2.971 99.832 51.274  - x6 1 3.695 100.556 51.433  - x8 1 4.133 100.994 51.528  - x9 1 8.594 105.455 52.479  <none> 96.861 52.609  - x2 1 14.068 110.929 53.593  - x5 1 20.404 117.265 54.815  - x1 1 41.240 138.101 58.413  Step: AIC=50.61  y ~ x1 + x2 + x3 + x4 + x5 + x6 + x8 + x9  Df Sum of Sq RSS AIC  - x3 1 0.488 97.356 48.721  - x4 1 2.983 99.850 49.278  - x8 1 4.613 101.480 49.634  <none> 96.868 50.611  - x9 1 12.927 109.795 51.367  - x6 1 14.410 111.277 51.662  - x2 1 15.494 112.362 51.875  - x5 1 21.721 118.588 53.062  - x1 1 55.163 152.031 58.527  Step: AIC=48.72  y ~ x1 + x2 + x4 + x5 + x6 + x8 + x9  Df Sum of Sq RSS AIC  - x4 1 4.954 102.310 47.813  - x8 1 5.101 102.457 47.845  <none> 97.356 48.721  - x6 1 14.955 112.311 49.865  - x2 1 15.330 112.686 49.938  - x9 1 19.445 116.800 50.727  - x5 1 21.698 119.054 51.148  - x1 1 77.178 174.534 59.564  Step: AIC=47.81  y ~ x1 + x2 + x5 + x6 + x8 + x9  Df Sum of Sq RSS AIC  - x8 1 5.452 107.76 46.955  <none> 102.31 47.813  - x6 1 10.963 113.27 48.053  - x9 1 17.202 119.51 49.232  - x5 1 21.654 123.96 50.037  - x2 1 31.807 134.12 51.769  - x1 1 88.167 190.48 59.487  Step: AIC=46.96  y ~ x1 + x2 + x5 + x6 + x9  Df Sum of Sq RSS AIC  <none> 107.76 46.955  - x9 1 13.719 121.48 47.592  - x5 1 21.365 129.13 48.935  - x6 1 25.110 132.87 49.564  - x2 1 28.366 136.13 50.096  - x1 1 202.251 310.01 68.203  Call:  lm(formula = y ~ x1 + x2 + x5 + x6 + x9, data = fdata)  Coefficients:  (Intercept) x1 x2 x5 x6 x9  16.182 3.055 6.364 2.195 -1.837 1.823 |
|  |
| |  |  |  |  |  | | --- | --- | --- | --- | --- | | >   |  | | --- | | be=lm(y~x1+x2+x5+x6+x9, data=fdata)  > summary(be)  Call:  lm(formula = y ~ x1 + x2 + x5 + x6 + x9, data = fdata)  Residuals:  Min 1Q Median 3Q Max  -3.8842 -1.5551 0.1727 1.5600 3.5507  Coefficients:  Estimate Std. Error t value Pr(>|t|)  (Intercept) 16.1817 4.5945 3.522 0.00283 \*\*  x1 3.0551 0.5575 5.480 5.04e-05 \*\*\*  x2 6.3639 3.1010 2.052 0.05688 .  x5 2.1949 1.2324 1.781 0.09389 .  x6 -1.8373 0.9516 -1.931 0.07141 .  x9 1.8231 1.2774 1.427 0.17274  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  Residual standard error: 2.595 on 16 degrees of freedom  Multiple R-squared: 0.8636, Adjusted R-squared: 0.8209  F-statistic: 20.25 on 5 and 16 DF, p-value: 2.091e-06 | |  | | |  | | --- | |  | | | |

**Stepwise procedure**

|  |
| --- |
| step(fullmodel, direction="both")  Start: AIC=52.61  y ~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9  Df Sum of Sq RSS AIC  - x7 1 0.007 96.868 50.611  - x3 1 0.494 97.355 50.721  - x4 1 2.971 99.832 51.274  - x6 1 3.695 100.556 51.433  - x8 1 4.133 100.994 51.528  - x9 1 8.594 105.455 52.479  <none> 96.861 52.609  - x2 1 14.068 110.929 53.593  - x5 1 20.404 117.265 54.815  - x1 1 41.240 138.101 58.413  Step: AIC=50.61  y ~ x1 + x2 + x3 + x4 + x5 + x6 + x8 + x9  Df Sum of Sq RSS AIC  - x3 1 0.488 97.356 48.721  - x4 1 2.983 99.850 49.278  - x8 1 4.613 101.480 49.634  <none> 96.868 50.611  - x9 1 12.927 109.795 51.367  - x6 1 14.410 111.277 51.662  - x2 1 15.494 112.362 51.875  + x7 1 0.007 96.861 52.609  - x5 1 21.721 118.588 53.062  - x1 1 55.163 152.031 58.527  Step: AIC=48.72  y ~ x1 + x2 + x4 + x5 + x6 + x8 + x9  Df Sum of Sq RSS AIC  - x4 1 4.954 102.310 47.813  - x8 1 5.101 102.457 47.845  <none> 97.356 48.721  - x6 1 14.955 112.311 49.865  - x2 1 15.330 112.686 49.938  + x3 1 0.488 96.868 50.611  + x7 1 0.001 97.355 50.721  - x9 1 19.445 116.800 50.727  - x5 1 21.698 119.054 51.148  - x1 1 77.178 174.534 59.564  Step: AIC=47.81  y ~ x1 + x2 + x5 + x6 + x8 + x9  Df Sum of Sq RSS AIC  - x8 1 5.452 107.762 46.955  <none> 102.310 47.813  - x6 1 10.963 113.273 48.053  + x4 1 4.954 97.356 48.721  - x9 1 17.202 119.513 49.232  + x3 1 2.460 99.850 49.278  + x7 1 0.133 102.177 49.785  - x5 1 21.654 123.964 50.037  - x2 1 31.807 134.117 51.769  - x1 1 88.167 190.478 59.487  Step: AIC=46.96  y ~ x1 + x2 + x5 + x6 + x9  Df Sum of Sq RSS AIC  <none> 107.76 46.955  - x9 1 13.719 121.48 47.592  + x8 1 5.452 102.31 47.813  + x4 1 5.305 102.46 47.845  + x3 1 3.477 104.28 48.234  - x5 1 21.365 129.13 48.935  + x7 1 0.079 107.68 48.939  - x6 1 25.110 132.87 49.564  - x2 1 28.366 136.13 50.096  - x1 1 202.251 310.01 68.203  Call:  lm(formula = y ~ x1 + x2 + x5 + x6 + x9, data = fdata)  Coefficients:  (Intercept) x1 x2 x5 x6 x9  16.182 3.055 6.364 2.195 -1.837 1.823 |
|  |
| |  | | --- | | > | |