Data Analysis

**TSP instance fl3795**

## A. The way the dependent variables F, G, I, K, N, O depend on A, B, C, D and how are these inter-correlated (F, G, I, K, N, O are different measures that show us how fast the altered applications work). E is a ratio of two such measures (F and G). Study on how E evolves depending on A,B,C,D.

## Below is the way the dependent variable F depends on A, B, C, D. Adjusted R-Squared index shows that the linear regression model is good (0.8245 means that by knowing A, B, C, D we could explain 82.45% of the variance in F). B, C, D are all negative correlated to F. A is negative correlated to F. P values show only C and D(20%, 40%) are statistically significant in the model.

##   
## Call:  
## lm(formula = F ~ A + B + C + D, data = myData1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -15.444 -5.311 0.516 3.983 19.270   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 75.25128 3.49348 21.540 < 2e-16 \*\*\*  
## A 0.02404 0.02168 1.109 0.27086   
## B -0.02400 0.02948 -0.814 0.41802   
## C -0.91417 0.04514 -20.252 < 2e-16 \*\*\*  
## D10% -3.38889 2.33098 -1.454 0.14981   
## D20% -7.16667 2.33098 -3.075 0.00286 \*\*   
## D30% -4.44444 2.33098 -1.907 0.06006 .   
## D40% -7.16667 2.33098 -3.075 0.00286 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 6.993 on 82 degrees of freedom  
## (1 observation deleted due to missingness)  
## Multiple R-squared: 0.8383, Adjusted R-squared: 0.8245   
## F-statistic: 60.74 on 7 and 82 DF, p-value: < 2.2e-16

## Below is the way the dependent variable G depends on A, B, C, D. Adjusted R-Squared index shows that the linear regression model is good (0.8498 means that by knowing A, B, C, D we could explain 84.98% of the variance in G). A and D(10%) are positive correlated to G. Others are all negative correlated to G. P values shows C and D(20%, 30%, 40%) are statistically significant in the model.

##   
## Call:  
## lm(formula = G ~ A + B + C + D, data = myData1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -34.463 -14.401 3.968 10.768 28.421   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 156.82821 7.34589 21.349 < 2e-16 \*\*\*  
## A 0.05301 0.04560 1.163 0.248341   
## B -0.01067 0.06200 -0.172 0.863825   
## C -2.05750 0.09492 -21.677 < 2e-16 \*\*\*  
## D10% 0.83333 4.90145 0.170 0.865415   
## D20% -18.33333 4.90145 -3.740 0.000339 \*\*\*  
## D30% -23.00000 4.90145 -4.692 1.07e-05 \*\*\*  
## D40% -14.83333 4.90145 -3.026 0.003307 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 14.7 on 82 degrees of freedom  
## (1 observation deleted due to missingness)  
## Multiple R-squared: 0.8616, Adjusted R-squared: 0.8498   
## F-statistic: 72.96 on 7 and 82 DF, p-value: < 2.2e-16

## Below is the way the dependent variable I depends on A,B,C,D. Adjusted R-Squared index shows that the linear regression model is bad. Dependent variable does not have any linear relation with independents. P values show that none of the independents is statistically significant in the model.

##   
## Call:  
## lm(formula = I ~ A + B + C + D, data = myData1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2324.9 -662.0 3.7 189.1 27497.6   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 1699.7103 1579.3631 1.076 0.285  
## A -9.4942 9.8032 -0.968 0.336  
## B -13.3576 13.3298 -1.002 0.319  
## C -0.3596 20.4069 -0.018 0.986  
## D10% 3.1333 1053.8098 0.003 0.998  
## D20% 1603.0056 1053.8098 1.521 0.132  
## D30% -56.4500 1053.8098 -0.054 0.957  
## D40% -68.0056 1053.8098 -0.065 0.949  
##   
## Residual standard error: 3161 on 82 degrees of freedom  
## (1 observation deleted due to missingness)  
## Multiple R-squared: 0.06599, Adjusted R-squared: -0.01374   
## F-statistic: 0.8277 on 7 and 82 DF, p-value: 0.5673

## Below is the way the dependent variable K depends on A, B, C, D. Adjusted R-Squared index shows that the linear regression model is fine (0.5712 means that by knowing A,B,C,D we could only explain 57.12% of the variance in K). D(10%) is positive correlated to K. Other independents are all negative correlated to K. P values show that only C and D(20%, 30%, 40%) are statistically significant in the model.

##   
## Call:  
## lm(formula = K ~ A + B + C + D, data = myData1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -21.536 -7.852 -0.871 7.179 35.418   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 86.777410 6.119481 14.181 < 2e-16 \*\*\*  
## A -0.014499 0.037984 -0.382 0.704   
## B -0.007742 0.051648 -0.150 0.881   
## C -0.073558 0.079070 -0.930 0.355   
## D10% 0.747222 4.083145 0.183 0.855   
## D20% -24.150556 4.083145 -5.915 7.37e-08 \*\*\*  
## D30% -32.229444 4.083145 -7.893 1.14e-11 \*\*\*  
## D40% -29.287222 4.083145 -7.173 2.96e-10 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 12.25 on 82 degrees of freedom  
## (1 observation deleted due to missingness)  
## Multiple R-squared: 0.6049, Adjusted R-squared: 0.5712   
## F-statistic: 17.94 on 7 and 82 DF, p-value: 3.096e-14

## Below is the way the dependent variable N depends on A,B,C,D. Adjusted R-Squared index shows that the linear regression model is not that good (0.1359 means that by knowing A,B,C,D we could only explain 13.59% of the variance in N). D(10%) is positive correlated to N. Other independentsare all negative correlated to N. P values show that only D(20% and 40%) are statistically significant in the model.

##   
## Call:  
## lm(formula = N ~ A + B + C + D, data = myData1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -8.6715 -3.5844 -0.3329 3.8977 10.6701   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 43.468205 2.441201 17.806 < 2e-16 \*\*\*  
## A -0.009071 0.015153 -0.599 0.55108   
## B -0.015871 0.020604 -0.770 0.44333   
## C -0.038783 0.031543 -1.230 0.22238   
## D10% 0.467222 1.628860 0.287 0.77496   
## D20% -3.243333 1.628860 -1.991 0.04979 \*   
## D30% -1.202778 1.628860 -0.738 0.46237   
## D40% -5.517778 1.628860 -3.388 0.00109 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 4.887 on 82 degrees of freedom  
## (1 observation deleted due to missingness)  
## Multiple R-squared: 0.2039, Adjusted R-squared: 0.1359   
## F-statistic: 3 on 7 and 82 DF, p-value: 0.007432

## Below is the way the dependent variable O depends on A, B, C, D. Adjusted R-Squared index shows that the linear regression model is just fine (0.23 means that by knowing A, B, C, D we could only explain 23% of the variance in O). B, C, D(10%, 40%) are positive correlated to I. Other independents are all negative correlated to O. P values show that D(20%, 30%, 40%) are statistically significant in the model.

##   
## Call:  
## lm(formula = O ~ A + B + C + D, data = myData1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -6.506 -1.653 0.183 1.546 5.084   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 16.652684 1.208125 13.784 < 2e-16 \*\*\*  
## A -0.007154 0.007499 -0.954 0.34285   
## B 0.003458 0.010197 0.339 0.73539   
## C 0.016558 0.015610 1.061 0.29192   
## D10% 0.142778 0.806106 0.177 0.85985   
## D20% -2.132222 0.806106 -2.645 0.00979 \*\*   
## D30% -3.447222 0.806106 -4.276 5.1e-05 \*\*\*  
## D40% -2.559444 0.806106 -3.175 0.00211 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.418 on 82 degrees of freedom  
## (1 observation deleted due to missingness)  
## Multiple R-squared: 0.2906, Adjusted R-squared: 0.23   
## F-statistic: 4.798 on 7 and 82 DF, p-value: 0.0001462

## E is a ratio of the measures (F and G). How does it evolve depending on A,B,C,D?

## Below is the way the dependent variable E depends on A, B, C, D. Adjusted R-Squared index shows that the linear regression model is bad (0.01584 means that by knowing A, B, C, D we could explain 1.584% of the variance in E). So there is no linear relation between independents and dependents.

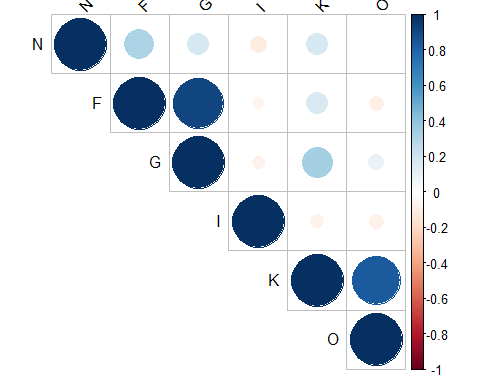
##   
## Call:  
## lm(formula = E ~ A + B + C + D, data = myData1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -10.991 -3.594 -0.893 1.503 112.522   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) -2.52477 6.55924 -0.385 0.701  
## A 0.05625 0.04071 1.382 0.171  
## B 0.05549 0.05536 1.002 0.319  
## C -0.10271 0.08475 -1.212 0.229  
## D10% -0.04463 4.37657 -0.010 0.992  
## D20% 0.01811 4.37657 0.004 0.997  
## D30% 0.10873 4.37657 0.025 0.980  
## D40% 6.98307 4.37657 1.596 0.114  
##   
## Residual standard error: 13.13 on 82 degrees of freedom  
## (1 observation deleted due to missingness)  
## Multiple R-squared: 0.09325, Adjusted R-squared: 0.01584   
## F-statistic: 1.205 on 7 and 82 DF, p-value: 0.3096

## How F, G, I, K, N, O inter-correlated. See below matrix and plots.

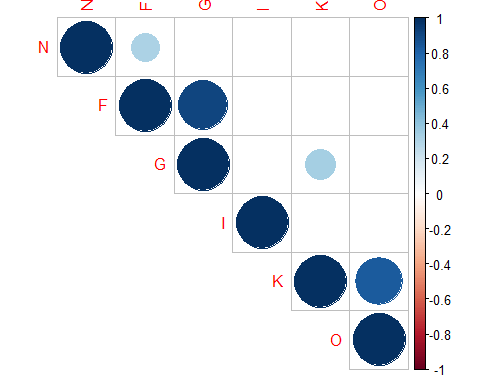
## Below are the correlations coefficients between the possible pairs of variables (F, G, I, K, N, O) and p values of correlations.

## F G I K N O  
## F 1.00 0.91 -0.05 0.16 0.31 -0.08  
## G 0.91 1.00 -0.06 0.34 0.17 0.09  
## I -0.05 -0.06 1.00 -0.07 -0.10 -0.08  
## K 0.16 0.34 -0.07 1.00 0.17 0.83  
## N 0.31 0.17 -0.10 0.17 1.00 -0.01  
## O -0.08 0.09 -0.08 0.83 -0.01 1.00  
##   
## n= 90   
##   
##   
## P  
## F G I K N O   
## F 0.0000 0.6338 0.1211 0.0027 0.4341  
## G 0.0000 0.5509 0.0012 0.1081 0.3868  
## I 0.6338 0.5509 0.5185 0.3479 0.4584  
## K 0.1211 0.0012 0.5185 0.1001 0.0000  
## N 0.0027 0.1081 0.3479 0.1001 0.9356  
## O 0.4341 0.3868 0.4584 0.0000 0.9356

## Below is the correlogram. Positive correlations are displayed in blue and negative correlations in red color. Color intensity and the size of the circle are proportional to the correlation coefficients. In the right side of the correlogram, the legend color shows the correlation coefficients and the corresponding colors.



## In the plot below, correlations with p-value > 0.01 are considered as insignificant. In this case the correlation coefficient values are leaved blank or crosses are added.



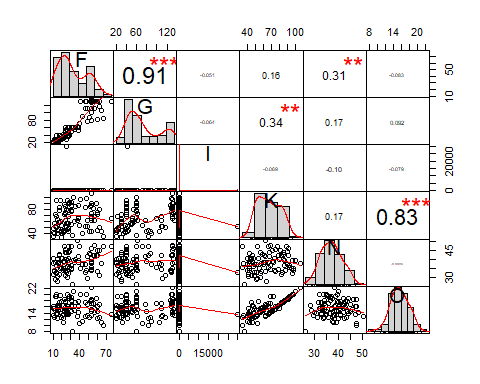
## In the plot below:

## The distribution of each variable is shown on the diagonal.

## On the bottom of the diagonal: the bivariate scatter plots with a fitted line are displayed.

## On the top of the diagonal : the value of the correlation plus the significance level as stars.

## Each significance level is associated to a symbol: p-values(0, 0.001, 0.01, 0.05, 0.1, 1) <=> symbols("", ##"", "", ".", "")



## B. The way the dependent variables H, J, L, M depend on A, B, C, D and how are these inter-correlated (H, J, L, M are: H - the objective function and J, L, M - three measures that show us how good the altered applications still perform).

## Below is the way the dependent variable H depends on A, B, C, D. Adjusted R-Squared index shows that the linear regression model is fine (0.5084 means that by knowing A,B,C,D we could only explain 50.84% of the variance in H). D(10%) is negative correlated to H. Other independents are all positive correlated to H. P values show that D(20%, 30%, 40%) are statistically significant in the model.

##   
## Call:  
## lm(formula = H ~ A + B + C + D, data = myData1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -105.583 -34.775 -1.943 27.548 185.639   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.976e+04 2.666e+01 1116.149 < 2e-16 \*\*\*  
## A 2.951e-02 1.655e-01 0.178 0.859   
## B 1.722e-01 2.250e-01 0.765 0.446   
## C 5.094e-01 3.445e-01 1.479 0.143   
## D10% -1.551e+01 1.779e+01 -0.872 0.386   
## D20% 1.003e+02 1.779e+01 5.640 2.35e-07 \*\*\*  
## D30% 1.143e+02 1.779e+01 6.426 8.13e-09 \*\*\*  
## D40% 9.683e+01 1.779e+01 5.442 5.34e-07 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 53.38 on 82 degrees of freedom  
## (1 observation deleted due to missingness)  
## Multiple R-squared: 0.5471, Adjusted R-squared: 0.5084   
## F-statistic: 14.15 on 7 and 82 DF, p-value: 6.583e-12

## Below is the way the dependent variable J depends on A,B,C,D. Adjusted R-Squared index shows that the linear regression model is bad. There is no relation between J and A, B, C, D.

##   
## Call:  
## lm(formula = J ~ A + B + C + D, data = myData1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -98.32 -22.29 4.39 16.71 108.83   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 172.87266 19.77078 8.744 2.33e-13 \*\*\*  
## A -0.06652 0.12272 -0.542 0.5893   
## B -0.24848 0.16686 -1.489 0.1403   
## C 0.13537 0.25546 0.530 0.5976   
## D10% 3.27444 13.19180 0.248 0.8046   
## D20% 7.19111 13.19180 0.545 0.5872   
## D30% 25.28556 13.19180 1.917 0.0588 .   
## D40% 11.58278 13.19180 0.878 0.3825   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 39.58 on 82 degrees of freedom  
## (1 observation deleted due to missingness)  
## Multiple R-squared: 0.08123, Adjusted R-squared: 0.002797   
## F-statistic: 1.036 on 7 and 82 DF, p-value: 0.4129

## Below is the way the dependent variable L depends on A, B, C, D. Adjusted R-Squared index shows that the linear regression model is not that good (0.1467 means that by knowing A, B, C, D we could only explain 14.67% of the variance in L). D(10%) is negative correlated to L. Other independents are all positive correlated to L. P values show that B and D(20%) are statistically significant in the model.

##   
## Call:  
## lm(formula = L ~ A + B + C + D, data = myData1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -276.520 -78.474 -0.649 80.211 247.384   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.945e+04 5.393e+01 546.047 <2e-16 \*\*\*  
## A 2.922e-01 3.347e-01 0.873 0.3852   
## B 1.094e+00 4.552e-01 2.404 0.0185 \*   
## C 1.083e-02 6.968e-01 0.016 0.9876   
## D10% -3.156e+01 3.598e+01 -0.877 0.3831   
## D20% 8.850e+01 3.598e+01 2.459 0.0160 \*   
## D30% 6.283e+01 3.598e+01 1.746 0.0845 .   
## D40% 6.550e+01 3.598e+01 1.820 0.0724 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 108 on 82 degrees of freedom  
## (1 observation deleted due to missingness)  
## Multiple R-squared: 0.2138, Adjusted R-squared: 0.1467   
## F-statistic: 3.185 on 7 and 82 DF, p-value: 0.004942

## Below is the way the dependent variables M depend on A, B, C, D. Adjusted R-Squared index shows that the linear regression model is not that good (0.1779 means that by knowing A, B, C, D we could only explain 17.79% of the variance in M). A and D(10%) are negative correlated to M. Other independents are all positive correlated to M. P values show that D(20%, 30%, 40%) are statistically significant in the model.

##   
## Call:  
## lm(formula = M ~ A + B + C + D, data = myData1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -532.62 -59.35 -8.31 45.58 351.80   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 29992.1709 57.4539 522.022 < 2e-16 \*\*\*  
## A -0.3082 0.3566 -0.864 0.38998   
## B 0.5756 0.4849 1.187 0.23868   
## C 0.4233 0.7424 0.570 0.57007   
## D10% -16.6111 38.3354 -0.433 0.66593   
## D20% 118.3333 38.3354 3.087 0.00276 \*\*   
## D30% 115.5556 38.3354 3.014 0.00343 \*\*   
## D40% 99.6111 38.3354 2.598 0.01110 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 115 on 82 degrees of freedom  
## (1 observation deleted due to missingness)  
## Multiple R-squared: 0.2425, Adjusted R-squared: 0.1779   
## F-statistic: 3.751 on 7 and 82 DF, p-value: 0.001422

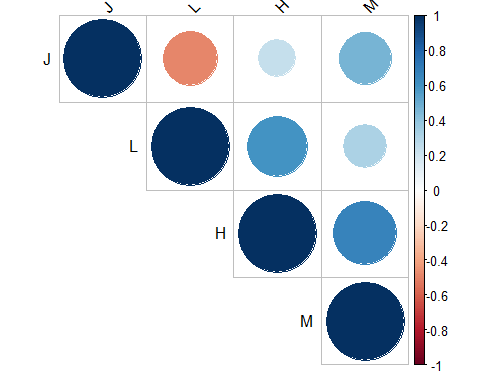
## How are H, J, L, M inter-correlated. See below matrix and plots.

## Below are the correlation coefficients between the possible pairs of variables (H, J, L, M) and p values of correlations.

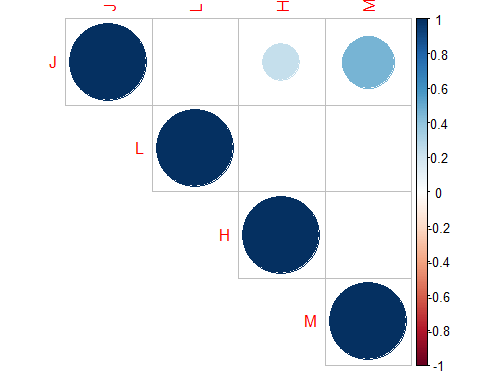
## H J L M  
## H 1.00 0.24 0.60 0.67  
## J 0.24 1.00 -0.49 0.46  
## L 0.60 -0.49 1.00 0.31  
## M 0.67 0.46 0.31 1.00  
##   
## n= 90   
##   
##   
## P  
## H J L M   
## H 0.0242 0.0000 0.0000  
## J 0.0242 0.0000 0.0000  
## L 0.0000 0.0000 0.0026  
## M 0.0000 0.0000 0.0026

## H J L M  
## H 1.00 0.24 0.60 0.67  
## J 0.24 1.00 -0.49 0.46  
## L 0.60 -0.49 1.00 0.31  
## M 0.67 0.46 0.31 1.00

## Below is the correlogram. Positive correlations are displayed in blue and negative correlations in red color. Color intensity and the size of the circle are proportional to the correlation coefficients. In the right side of the correlogram, the legend color shows the correlation coefficients and the corresponding colors.



## In the plot below, correlations with p-value > 0.01 are considered as insignificant. In this case the correlation coefficient values are leaved blank or crosses are added.



## In the plot below:

## The distribution of each variable is shown on the diagonal.

## On the bottom of the diagonal: the bivariate scatter plots with a fitted line are displayed.

## On the top of the diagonal: the value of the correlation plus the significance level as stars.

## Each significance level is associated to a symbol: p-values(0, 0.001, 0.01, 0.05, 0.1, 1) <=> symbols("*", ##"", "*", ".", "")

