Data Analysis

**TSP instance d2103**

## 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 (0.6291 means that by knowing A,B,C,D we could explain 62.91% of the variance in F. A, B, C, D are all negative correlated to F. But P values show only C and D(10%, 30%, 40%) are statistically significant in the model.

##   
## Call:  
## lm(formula = F ~ A + B + C + D, data = myData1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -55.774 -15.298 -0.002 13.087 75.226   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.938e+02 1.170e+01 25.107 <2e-16 \*\*\*  
## A -2.474e-02 7.264e-02 -0.341 0.7343   
## B 5.398e-16 9.877e-02 0.000 1.0000   
## C -1.850e+00 1.512e-01 -12.234 <2e-16 \*\*\*  
## D10% -1.567e+01 7.809e+00 -2.006 0.0481 \*   
## D20% -1.361e+01 7.809e+00 -1.743 0.0851 .   
## D30% -1.561e+01 7.809e+00 -1.999 0.0489 \*   
## D40% -2.106e+01 7.809e+00 -2.696 0.0085 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 23.43 on 82 degrees of freedom  
## Multiple R-squared: 0.6582, Adjusted R-squared: 0.6291   
## F-statistic: 22.56 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 (0.8196means that by knowing A, B, C, D we could explain 81.96% of the variance in G). A is positive correlated to G. B, C, D are all negative correlated to G. But P values shows only C and D(20%, 40%) are statistically significant in the model.

##   
## Call:  
## lm(formula = G ~ A + B + C + D, data = myData1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -184.21 -45.73 11.26 51.88 93.57   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 672.31026 34.67348 19.390 <2e-16 \*\*\*  
## A 0.04564 0.21522 0.212 0.8326   
## B -0.03200 0.29264 -0.109 0.9132   
## C -8.98917 0.44802 -20.064 <2e-16 \*\*\*  
## D10% -45.44444 23.13543 -1.964 0.0529 .   
## D20% -63.88889 23.13543 -2.762 0.0071 \*\*   
## D30% -37.72222 23.13543 -1.630 0.1068   
## D40% -49.94444 23.13543 -2.159 0.0338 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 69.41 on 82 degrees of freedom  
## Multiple R-squared: 0.8338, Adjusted R-squared: 0.8196   
## F-statistic: 58.75 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 not that good (0.286 means that by knowing A, B, C, D we could only explain 28.6% of the variance in I). B is positive correlated to I. A, C, D are all negative correlated to I. P values show that C and D are statistically significant in the model.

##   
## Call:  
## lm(formula = I ~ A + B + C + D, data = myData1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -264.70 -69.36 -16.14 86.11 206.17   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 861.2392 55.0096 15.656 < 2e-16 \*\*\*  
## A -0.1388 0.3414 -0.407 0.685345   
## B 0.5376 0.4643 1.158 0.250256   
## C -3.1515 0.7108 -4.434 2.85e-05 \*\*\*  
## D10% -148.5056 36.7044 -4.046 0.000117 \*\*\*  
## D20% -127.3111 36.7044 -3.469 0.000836 \*\*\*  
## D30% -134.9889 36.7044 -3.678 0.000419 \*\*\*  
## D40% -117.6111 36.7044 -3.204 0.001929 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 110.1 on 82 degrees of freedom  
## Multiple R-squared: 0.3422, Adjusted R-squared: 0.286   
## F-statistic: 6.093 on 7 and 82 DF, p-value: 9.648e-06

## 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 not that good (0.1166 means that by knowing A, B, C, D we could only explain 11.66% of the variance in K). A and B are positive correlated to K. C and Dare all negative correlated to K. P values show that only C and D(20%) are statistically significant in the model.

##   
## Call:  
## lm(formula = K ~ A + B + C + D, data = myData1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -137.32 -39.53 10.51 41.65 129.66   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 216.68406 30.10821 7.197 2.66e-10 \*\*\*  
## A 0.08741 0.18688 0.468 0.64124   
## B 0.16221 0.25411 0.638 0.52502   
## C -1.15138 0.38903 -2.960 0.00403 \*\*  
## D10% -7.20444 20.08931 -0.359 0.72080   
## D20% -53.69667 20.08931 -2.673 0.00907 \*\*   
## D30% -23.26000 20.08931 -1.158 0.25029   
## D40% -5.88111 20.08931 -0.293 0.77045   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 60.27 on 82 degrees of freedom  
## Multiple R-squared: 0.1861, Adjusted R-squared: 0.1166   
## F-statistic: 2.678 on 7 and 82 DF, p-value: 0.01507

## 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 (0.06277 means that by knowing A,B,C,D we could only explain 6.277% of the variance in N). B, C, and D(10%, 20%, 30%) are positive correlated to N. A and D(40%)are negative correlated to N. P values show that only C and D(20%) are statistically significant in the model.

##   
## Call:  
## lm(formula = N ~ A + B + C + D, data = myData1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -11.2117 -4.0751 0.2902 3.8108 9.4409   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 43.117880 2.511515 17.168 <2e-16 \*\*\*  
## A -0.014302 0.015589 -0.917 0.3616   
## B 0.008298 0.021197 0.391 0.6965   
## C 0.074092 0.032451 2.283 0.0250 \*   
## D10% 0.810000 1.675776 0.483 0.6301   
## D20% 3.603889 1.675776 2.151 0.0345 \*   
## D30% 1.570556 1.675776 0.937 0.3514   
## D40% -0.208333 1.675776 -0.124 0.9014   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 5.027 on 82 degrees of freedom  
## Multiple R-squared: 0.1365, Adjusted R-squared: 0.06277   
## F-statistic: 1.851 on 7 and 82 DF, p-value: 0.08839

## 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 not that good (0.1274means that by knowing A,B,C,D we could only explain 12.74% of the variance in O). A, B, D(10%, 40%) are positive correlated to I. C and D(20%, 30%) are all negative correlated to O. P values show that C is statistically significant in the model.

##   
## Call:  
## lm(formula = O ~ A + B + C + D, data = myData1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -5.4545 -1.9111 0.3525 1.7647 5.8109   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 10.3621282 1.2835399 8.073 5.01e-12 \*\*\*  
## A 0.0009038 0.0079670 0.113 0.9100   
## B 0.0135244 0.0108330 1.248 0.2154   
## C -0.0375750 0.0165846 -2.266 0.0261 \*   
## D10% 1.0916667 0.8564256 1.275 0.2060   
## D20% -1.7027778 0.8564256 -1.988 0.0501 .   
## D30% -0.9333333 0.8564256 -1.090 0.2790   
## D40% 0.4111111 0.8564256 0.480 0.6325   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.569 on 82 degrees of freedom  
## Multiple R-squared: 0.196, Adjusted R-squared: 0.1274   
## F-statistic: 2.857 on 7 and 82 DF, p-value: 0.01019

## 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 good (0.8061 means that by knowing A, B, C, D we could explain 80.61% of the variance in E). C and D are positive correlated to E.A and Bare all negative correlated to E. P values show that C and D(10%, 20%, 30%) are statistically significant in the model.

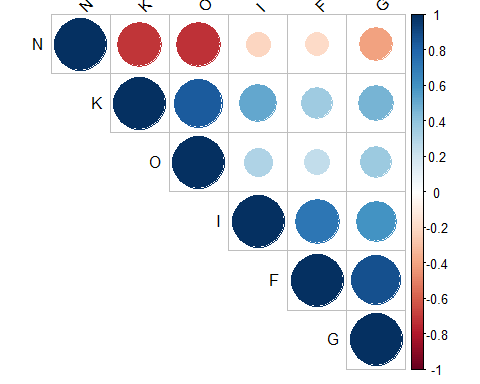
##   
## Call:  
## lm(formula = E ~ A + B + C + D, data = myData1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.31815 -0.11957 -0.01764 0.11520 0.54237   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.0103022 0.0963029 0.107 0.91507   
## A -0.0003021 0.0005978 -0.505 0.61466   
## B -0.0010120 0.0008128 -1.245 0.21665   
## C 0.0235334 0.0012443 18.913 < 2e-16 \*\*\*  
## D10% 0.1944056 0.0642569 3.025 0.00332 \*\*   
## D20% 0.1923444 0.0642569 2.993 0.00365 \*\*   
## D30% 0.1871389 0.0642569 2.912 0.00462 \*\*   
## D40% 0.0401833 0.0642569 0.625 0.53348   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.1928 on 82 degrees of freedom  
## Multiple R-squared: 0.8213, Adjusted R-squared: 0.8061   
## F-statistic: 53.85 on 7 and 82 DF, p-value: < 2.2e-16

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

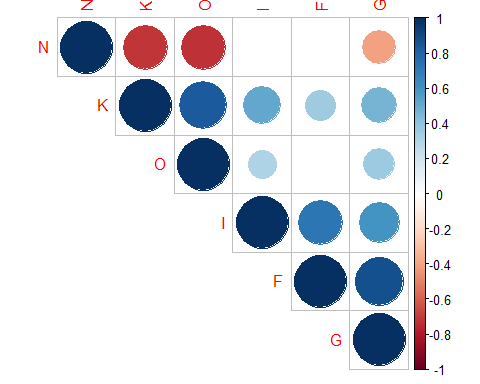
## Below are the correlations coefficients between the possible pairs of variables (F, G, I, K, N, O) and p values of correlations.

## F G I K N O  
## F 1.00 0.88 0.72 0.36 -0.20 0.24  
## G 0.88 1.00 0.60 0.47 -0.40 0.37  
## I 0.72 0.60 1.00 0.51 -0.21 0.30  
## K 0.36 0.47 0.51 1.00 -0.72 0.84  
## N -0.20 -0.40 -0.21 -0.72 1.00 -0.73  
## O 0.24 0.37 0.30 0.84 -0.73 1.00  
##   
## n= 90   
##   
##   
## P  
## F G I K N O   
## F 0.0000 0.0000 0.0006 0.0649 0.0225  
## G 0.0000 0.0000 0.0000 0.0000 0.0004  
## I 0.0000 0.0000 0.0000 0.0430 0.0040  
## K 0.0006 0.0000 0.0000 0.0000 0.0000  
## N 0.0649 0.0000 0.0430 0.0000 0.0000  
## O 0.0225 0.0004 0.0040 0.0000 0.0000

## 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.



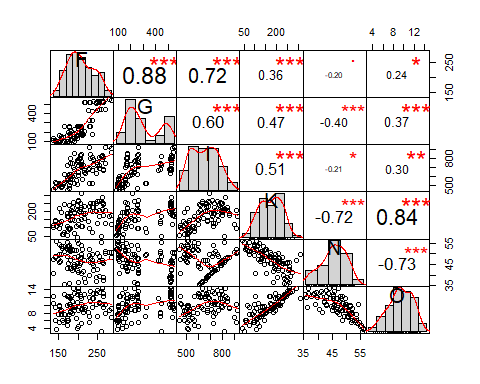
## In the plot below:

## 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 bad. There is no relation between H and A, B, C, D.

##   
## Call:  
## lm(formula = H ~ A + B + C + D, data = myData1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -149.58 -23.76 -6.34 14.61 491.21   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 8.058e+04 3.199e+01 2518.882 <2e-16 \*\*\*  
## A -2.363e-02 1.986e-01 -0.119 0.906   
## B 1.508e-01 2.700e-01 0.558 0.578   
## C 6.450e-02 4.134e-01 0.156 0.876   
## D10% -1.956e+01 2.135e+01 -0.916 0.362   
## D20% -2.938e+01 2.135e+01 -1.377 0.172   
## D30% -2.251e+01 2.135e+01 -1.054 0.295   
## D40% -1.838e+01 2.135e+01 -0.861 0.392   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 64.04 on 82 degrees of freedom  
## Multiple R-squared: 0.02891, Adjusted R-squared: -0.05398   
## F-statistic: 0.3488 on 7 and 82 DF, p-value: 0.9287

## 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   
## -152.01 -45.10 -12.40 18.55 1552.49   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 103.08562 89.64248 1.150 0.254  
## A -0.03137 0.55641 -0.056 0.955  
## B 0.78530 0.75658 1.038 0.302  
## C 0.03954 1.15827 0.034 0.973  
## D10% -86.44944 59.81280 -1.445 0.152  
## D20% -91.69833 59.81280 -1.533 0.129  
## D30% -98.08056 59.81280 -1.640 0.105  
## D40% -87.87222 59.81280 -1.469 0.146  
##   
## Residual standard error: 179.4 on 82 degrees of freedom  
## Multiple R-squared: 0.05565, Adjusted R-squared: -0.02496   
## F-statistic: 0.6904 on 7 and 82 DF, p-value: 0.6799

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

##   
## Call:  
## lm(formula = L ~ A + B + C + D, data = myData1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -36.320 -18.199 -2.799 15.243 76.674   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 8.048e+04 1.138e+01 7070.274 <2e-16 \*\*\*  
## A 9.359e-03 7.065e-02 0.132 0.895   
## B 4.444e-03 9.607e-02 0.046 0.963   
## C 1.917e-01 1.471e-01 1.303 0.196   
## D10% -1.167e+00 7.595e+00 -0.154 0.878   
## D20% 5.833e+00 7.595e+00 0.768 0.445   
## D30% -2.056e+00 7.595e+00 -0.271 0.787   
## D40% -8.333e-01 7.595e+00 -0.110 0.913   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 22.78 on 82 degrees of freedom  
## Multiple R-squared: 0.03636, Adjusted R-squared: -0.0459   
## F-statistic: 0.442 on 7 and 82 DF, p-value: 0.8729

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

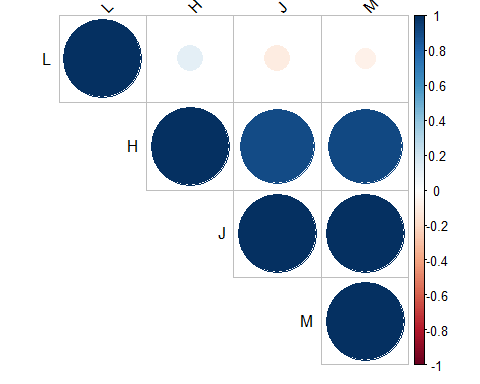
##   
## Call:  
## lm(formula = M ~ A + B + C + D, data = myData1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -476.8 -143.1 -41.4 62.0 4961.9   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 8.080e+04 2.875e+02 281.059 <2e-16 \*\*\*  
## A -3.756e-02 1.784e+00 -0.021 0.9833   
## B 2.589e+00 2.426e+00 1.067 0.2891   
## C 2.075e-01 3.715e+00 0.056 0.9556   
## D10% -2.653e+02 1.918e+02 -1.383 0.1703   
## D20% -2.971e+02 1.918e+02 -1.549 0.1253   
## D30% -3.225e+02 1.918e+02 -1.681 0.0965 .   
## D40% -2.870e+02 1.918e+02 -1.496 0.1384   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 575.4 on 82 degrees of freedom  
## Multiple R-squared: 0.05711, Adjusted R-squared: -0.02338   
## F-statistic: 0.7095 on 7 and 82 DF, p-value: 0.664

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

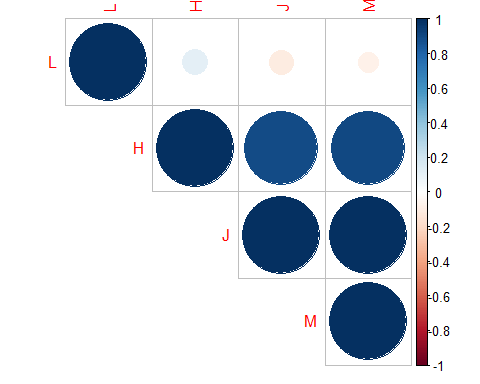
## Below are the correlation coefficients between the possible pairs of variables (H, J, L, M) and p values of correlations.

## H J L M  
## H 1.00 0.90 0.11 0.90  
## J 0.90 1.00 -0.11 1.00  
## L 0.11 -0.11 1.00 -0.07  
## M 0.90 1.00 -0.07 1.00  
##   
## n= 90   
##   
##   
## P  
## H J L M   
## H 0.0000 0.2944 0.0000  
## J 0.0000 0.3245 0.0000  
## L 0.2944 0.3245 0.4918  
## M 0.0000 0.0000 0.4918

## 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("*", ##"", "*", ".", "")

