Data Analysis

**TSP instance rl1889**

## Warning: package 'corrplot' was built under R version 3.4.4

## corrplot 0.84 loaded

## Warning: package 'Hmisc' was built under R version 3.4.4

## Loading required package: lattice

## Loading required package: survival

## Loading required package: Formula

## Warning: package 'Formula' was built under R version 3.4.4

## Loading required package: ggplot2

## Warning: package 'ggplot2' was built under R version 3.4.2

##   
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:base':  
##   
## format.pval, units

## Warning: package 'PerformanceAnalytics' was built under R version 3.4.4

## Loading required package: xts

## Warning: package 'xts' was built under R version 3.4.4

## Loading required package: zoo

## Warning: package 'zoo' was built under R version 3.4.2

##   
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':  
##   
## as.Date, as.Date.numeric

##   
## Attaching package: 'PerformanceAnalytics'

## The following object is masked from 'package:graphics':  
##   
## legend

## A B C D E F G H I J K L  
## 1 100 100 20 10% 0.9206 290 315 318079.8 547.6 417.28 125.12 317350  
## 2 100 100 20 20% 0.8745 202 231 318087.0 424.0 710.66 93.07 318048  
## 3 100 100 20 30% 0.8284 182 212 318094.2 400.2 581.12 30.89 318746  
## 4 100 100 20 40% 0.9728 179 184 318101.4 349.0 1284.76 26.47 318756  
## 5 100 100 20 0 0.9250 185 200 318108.6 369.3 1452.55 135.31 318766  
## 6 40 100 20 10% 0.7172 246 343 318115.8 563.3 509.48 90.04 318776  
## M N O X X.1 X.2  
## 1 318927 53.50 4.99 NA NA NA  
## 2 320337 56.03 3.15 NA NA NA  
## 3 321747 58.56 3.31 NA NA NA  
## 4 321816 56.85 4.32 NA NA NA  
## 5 321885 53.02 11.23 NA NA NA  
## 6 321954 49.95 7.58 NA NA NA

## 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.

## The way the dependent variable F depends on A, B, C, D.

##   
## Call:  
## lm(formula = F ~ A + B + C + D, data = myData1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -163.55 -30.94 11.85 36.70 89.96   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 228.9205 26.6421 8.592 4.66e-13 \*\*\*  
## A 0.2971 0.1654 1.797 0.0761 .   
## B 0.2653 0.2249 1.180 0.2414   
## C -1.4958 0.3442 -4.345 3.96e-05 \*\*\*  
## D10% 28.1667 17.7766 1.584 0.1169   
## D20% 9.9444 17.7766 0.559 0.5774   
## D30% 16.8333 17.7766 0.947 0.3465   
## D40% 11.7222 17.7766 0.659 0.5115   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 53.33 on 82 degrees of freedom  
## Multiple R-squared: 0.242, Adjusted R-squared: 0.1773   
## F-statistic: 3.741 on 7 and 82 DF, p-value: 0.001454

## The way the dependent variable G depends on A, B, C, D.

##   
## Call:  
## lm(formula = G ~ A + B + C + D, data = myData1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -129.660 -33.017 0.821 33.011 199.414   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 342.90085 27.10933 12.649 <2e-16 \*\*\*  
## A -0.05647 0.16827 -0.336 0.738   
## B 0.09111 0.22880 0.398 0.692   
## C -4.74167 0.35028 -13.537 <2e-16 \*\*\*  
## D10% 28.66667 18.08835 1.585 0.117   
## D20% 19.33333 18.08835 1.069 0.288   
## D30% 16.88889 18.08835 0.934 0.353   
## D40% 18.22222 18.08835 1.007 0.317   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 54.27 on 82 degrees of freedom  
## Multiple R-squared: 0.6942, Adjusted R-squared: 0.6681   
## F-statistic: 26.59 on 7 and 82 DF, p-value: < 2.2e-16

## The way the dependent variable I depends on A, B, C, D.

##   
## Call:  
## lm(formula = I ~ A + B + C + D, data = myData1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -407.79 -101.53 38.90 84.14 336.21   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 363.3008 72.0959 5.039 2.74e-06 \*\*\*  
## A 0.5673 0.4475 1.268 0.2085   
## B 1.3111 0.6085 2.155 0.0341 \*   
## C 0.5004 0.9316 0.537 0.5926   
## D10% 29.4667 48.1051 0.613 0.5419   
## D20% -1.3056 48.1051 -0.027 0.9784   
## D30% 30.3700 48.1051 0.631 0.5296   
## D40% 29.8339 48.1051 0.620 0.5369   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 144.3 on 82 degrees of freedom  
## Multiple R-squared: 0.08387, Adjusted R-squared: 0.005664   
## F-statistic: 1.072 on 7 and 82 DF, p-value: 0.3886

## The way the dependent variable K depends on A, B, C, D.

##   
## Call:  
## lm(formula = K ~ A + B + C + D, data = myData1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -73.005 -20.906 -4.859 20.008 118.928   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 55.4936 18.3387 3.026 0.00331 \*\*  
## A 0.1295 0.1138 1.138 0.25847   
## B 0.1376 0.1548 0.889 0.37646   
## C -0.1352 0.2370 -0.571 0.56985   
## D10% 15.4239 12.2363 1.261 0.21106   
## D20% -2.9617 12.2363 -0.242 0.80935   
## D30% -2.6222 12.2363 -0.214 0.83085   
## D40% 19.9683 12.2363 1.632 0.10654   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 36.71 on 82 degrees of freedom  
## Multiple R-squared: 0.09642, Adjusted R-squared: 0.01929   
## F-statistic: 1.25 on 7 and 82 DF, p-value: 0.2855

## The way the dependent variable N depends on A, B, C, D.

##   
## Call:  
## lm(formula = N ~ A + B + C + D, data = myData1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -11.708 -1.666 0.220 1.616 13.450   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 54.019043 1.571372 34.377 <2e-16 \*\*\*  
## A 0.006851 0.009754 0.702 0.4844   
## B 0.007249 0.013262 0.547 0.5862   
## C 0.014367 0.020304 0.708 0.4812   
## D10% -1.996111 1.048478 -1.904 0.0604 .   
## D20% -1.222778 1.048478 -1.166 0.2469   
## D30% 0.073333 1.048478 0.070 0.9444   
## D40% -0.845556 1.048478 -0.806 0.4223   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3.145 on 82 degrees of freedom  
## Multiple R-squared: 0.07633, Adjusted R-squared: -0.002516   
## F-statistic: 0.9681 on 7 and 82 DF, p-value: 0.4601

## The way the dependent variable O depends on A, B, C, D.

##   
## Call:  
## lm(formula = O ~ A + B + C + D, data = myData1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -4.3082 -1.4771 -0.4027 1.5335 8.5021   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 5.294701 1.248177 4.242 5.79e-05 \*\*\*  
## A 0.016658 0.007747 2.150 0.0345 \*   
## B -0.002769 0.010535 -0.263 0.7933   
## C -0.016442 0.016128 -1.019 0.3110   
## D10% 0.721111 0.832830 0.866 0.3891   
## D20% 0.135000 0.832830 0.162 0.8716   
## D30% -0.857778 0.832830 -1.030 0.3061   
## D40% 0.714444 0.832830 0.858 0.3935   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.498 on 82 degrees of freedom  
## Multiple R-squared: 0.1143, Adjusted R-squared: 0.03871   
## F-statistic: 1.512 on 7 and 82 DF, p-value: 0.1747

## E is a ratio of F and G. Study on how E evolves depending on A, B, C, D.

##   
## Call:  
## lm(formula = E ~ A + B + C + D, data = myData1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.59491 -0.12230 -0.00215 0.12432 0.56298   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.2704582 0.1120153 2.414 0.018 \*   
## A 0.0010162 0.0006953 1.462 0.148   
## B 0.0007799 0.0009454 0.825 0.412   
## C 0.0275353 0.0014473 19.025 <2e-16 \*\*\*  
## D10% -0.0380889 0.0747408 -0.510 0.612   
## D20% -0.0471611 0.0747408 -0.631 0.530   
## D30% -0.0201778 0.0747408 -0.270 0.788   
## D40% -0.0285889 0.0747408 -0.383 0.703   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.2242 on 82 degrees of freedom  
## Multiple R-squared: 0.8166, Adjusted R-squared: 0.801   
## F-statistic: 52.17 on 7 and 82 DF, p-value: < 2.2e-16

## Inter-correlations of the variables F, G, I, K, N, O.

## F G I K N O  
## 1 290 315 547.6 125.12 53.50 4.99  
## 2 202 231 424.0 93.07 56.03 3.15  
## 3 182 212 400.2 30.89 58.56 3.31  
## 4 179 184 349.0 26.47 56.85 4.32  
## 5 185 200 369.3 135.31 53.02 11.23  
## 6 246 343 563.3 90.04 49.95 7.58

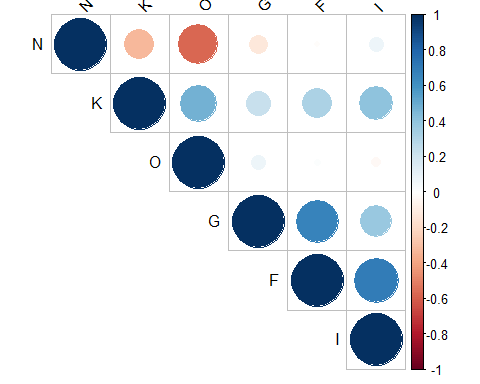
## F G I K N O  
## F 1.00 0.67 0.70 0.31 -0.01 0.02  
## G 0.67 1.00 0.37 0.22 -0.12 0.08  
## I 0.70 0.37 1.00 0.40 0.08 -0.03  
## K 0.31 0.22 0.40 1.00 -0.33 0.48  
## N -0.01 -0.12 0.08 -0.33 1.00 -0.58  
## O 0.02 0.08 -0.03 0.48 -0.58 1.00  
##   
## n= 90   
##   
##   
## P  
## F G I K N O   
## F 0.0000 0.0000 0.0026 0.9197 0.8851  
## G 0.0000 0.0003 0.0348 0.2440 0.4575  
## I 0.0000 0.0003 0.0000 0.4624 0.7507  
## K 0.0026 0.0348 0.0000 0.0017 0.0000  
## N 0.9197 0.2440 0.4624 0.0017 0.0000  
## O 0.8851 0.4575 0.7507 0.0000 0.0000

## In the table above correlation coefficients between the possible pairs of variables (F, G, I, K, N, O) are shown.

## Draw a correlogram, ignore the NA values.

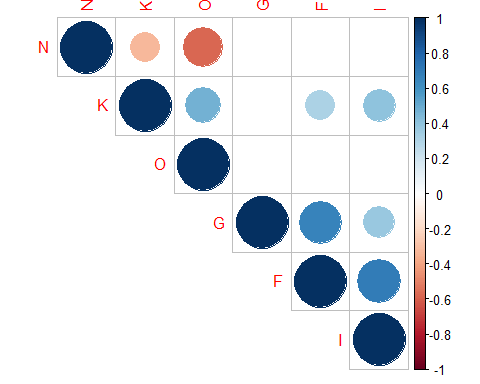
## F G I K N O  
## F 1.00000000 0.66744241 0.69967941 0.3135882 -0.01078315 0.01545118  
## G 0.66744241 1.00000000 0.37253019 0.2227946 -0.12406778 0.07930729  
## I 0.69967941 0.37253019 1.00000000 0.4030971 0.07843945 -0.03395521  
## K 0.31358818 0.22279462 0.40309714 1.0000000 -0.32605375 0.47526804  
## N -0.01078315 -0.12406778 0.07843945 -0.3260538 1.00000000 -0.57885906  
## O 0.01545118 0.07930729 -0.03395521 0.4752680 -0.57885906 1.00000000

## F G I K N O  
## F 1.00 0.67 0.70 0.31 -0.01 0.02  
## G 0.67 1.00 0.37 0.22 -0.12 0.08  
## I 0.70 0.37 1.00 0.40 0.08 -0.03  
## K 0.31 0.22 0.40 1.00 -0.33 0.48  
## N -0.01 -0.12 0.08 -0.33 1.00 -0.58  
## O 0.02 0.08 -0.03 0.48 -0.58 1.00



## 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.



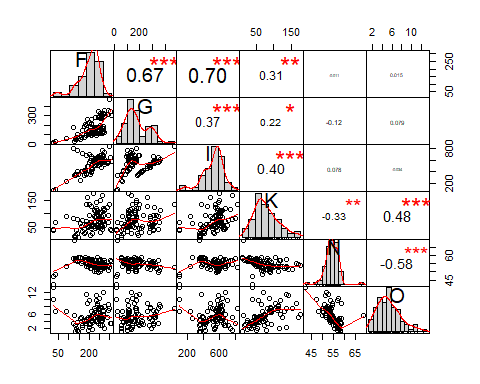
## In the plot below:

## 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).

## The way the dependent variable H depends on A, B, C, D:

##   
## Call:  
## lm(formula = H ~ A + B + C + D, data = myData1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -231.23 -156.69 -8.88 113.31 359.31   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.183e+05 9.001e+01 3536.748 <2e-16 \*\*\*  
## A 1.731e-01 5.587e-01 0.310 0.7575   
## B -1.200e+00 7.597e-01 -1.580 0.1180   
## C 4.050e+00 1.163e+00 3.482 0.0008 \*\*\*  
## D10% -2.880e+01 6.006e+01 -0.480 0.6328   
## D20% -2.160e+01 6.006e+01 -0.360 0.7200   
## D30% -1.440e+01 6.006e+01 -0.240 0.8111   
## D40% -7.200e+00 6.006e+01 -0.120 0.9049   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 180.2 on 82 degrees of freedom  
## Multiple R-squared: 0.1547, Adjusted R-squared: 0.08253   
## F-statistic: 2.144 on 7 and 82 DF, p-value: 0.04789

## The way the dependent variable J depends on A, B, C, D.

##   
## Call:  
## lm(formula = J ~ A + B + C + D, data = myData1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -491.21 -145.85 -81.19 126.44 1143.94   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 792.1615 129.4168 6.121 3.05e-08 \*\*\*  
## A -0.3864 0.8033 -0.481 0.6318   
## B 0.6413 1.0923 0.587 0.5587   
## C -1.6069 1.6722 -0.961 0.3394   
## D10% -200.2178 86.3517 -2.319 0.0229 \*   
## D20% -183.8794 86.3517 -2.129 0.0362 \*   
## D30% -110.7867 86.3517 -1.283 0.2031   
## D40% 57.1589 86.3517 0.662 0.5099   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 259.1 on 82 degrees of freedom  
## Multiple R-squared: 0.1565, Adjusted R-squared: 0.08446   
## F-statistic: 2.173 on 7 and 82 DF, p-value: 0.045

## The way the dependent variable L depends on A, B, C, D.

##   
## Call:  
## lm(formula = L ~ A + B + C + D, data = myData1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1481.0 -180.5 15.8 204.8 574.6   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.191e+05 1.592e+02 2004.650 < 2e-16 \*\*\*  
## A -6.858e-01 9.881e-01 -0.694 0.489636   
## B -2.584e+00 1.344e+00 -1.923 0.057925 .   
## C 7.345e+00 2.057e+00 3.571 0.000599 \*\*\*  
## D10% -1.164e+02 1.062e+02 -1.096 0.276175   
## D20% -6.822e+01 1.062e+02 -0.642 0.522487   
## D30% -2.000e+01 1.062e+02 -0.188 0.851116   
## D40% -1.000e+01 1.062e+02 -0.094 0.925224   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 318.7 on 82 degrees of freedom  
## Multiple R-squared: 0.185, Adjusted R-squared: 0.1154   
## F-statistic: 2.659 on 7 and 82 DF, p-value: 0.01571

## The way the dependent variable M depends on A, B, C, D.

##   
## Call:  
## lm(formula = M ~ A + B + C + D, data = myData1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -4283.2 -1408.3 0.3 1112.9 3590.6   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.241e+05 9.028e+02 359.015 < 2e-16 \*\*\*  
## A -1.465e-01 5.604e+00 -0.026 0.979202   
## B -1.329e+01 7.620e+00 -1.744 0.084936 .   
## C 4.216e+01 1.167e+01 3.614 0.000518 \*\*\*  
## D10% -4.250e+02 6.024e+02 -0.705 0.482501   
## D20% -2.815e+02 6.024e+02 -0.467 0.641534   
## D30% -1.380e+02 6.024e+02 -0.229 0.819378   
## D40% -6.900e+01 6.024e+02 -0.115 0.909090   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1807 on 82 degrees of freedom  
## Multiple R-squared: 0.1696, Adjusted R-squared: 0.09873   
## F-statistic: 2.393 on 7 and 82 DF, p-value: 0.02806

## How are H, J, L, M inter-correlated.

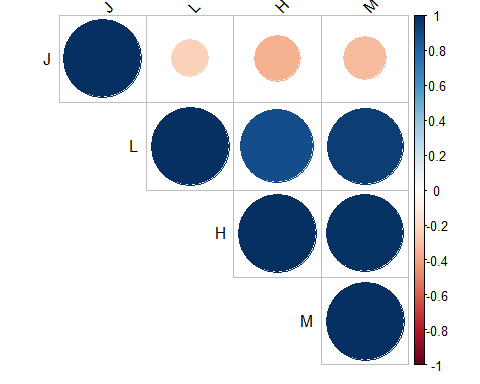
## H J L M  
## 1 318079.8 417.28 317350 318927  
## 2 318087.0 710.66 318048 320337  
## 3 318094.2 581.12 318746 321747  
## 4 318101.4 1284.76 318756 321816  
## 5 318108.6 1452.55 318766 321885  
## 6 318115.8 509.48 318776 321954

## H J L M  
## H 1.00 -0.35 0.89 0.99  
## J -0.35 1.00 -0.23 -0.32  
## L 0.89 -0.23 1.00 0.95  
## M 0.99 -0.32 0.95 1.00  
##   
## n= 90   
##   
##   
## P  
## H J L M   
## H 0.0007 0.0000 0.0000  
## J 0.0007 0.0280 0.0022  
## L 0.0000 0.0280 0.0000  
## M 0.0000 0.0022 0.0000

## H J L M  
## H 1.00 -0.35 0.89 0.99  
## J -0.35 1.00 -0.23 -0.32  
## L 0.89 -0.23 1.00 0.95  
## M 0.99 -0.32 0.95 1.00

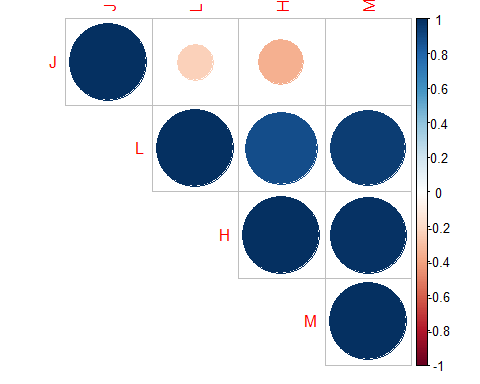
## In the table above correlation coefficients between the possible pairs of variables (F, G, I, K, N, O) are shown.

## Draw a 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 below plot, 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 below plot:

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

