

To find out useful independent variables to explain “operational losses” (the dependent variable):

1. Create a correlation matrix for all variables in the Excel

- 1) Use Excel “Add-in” tool— “*Data Analysis*” to create a correlation matrix, like this (*sheet “matrix_1”*):

	A	B	C	D	E	F	G	H	I
1		Loss	Errors	ancels/Correct	Breaks	Fails	Trades	HeadCnt	Downtime
2	Loss	1							
3	Errors	-0.0199952	1						
4	Cancels/Corre	-0.0386812	-0.0268267	1					
5	Breaks	0.05571281	0.15592036	0.15050024	1				
6	Fails	-0.0596397	0.84031664	0.00187339	0.17816488	1			
7	Trades	0.01080568	-0.1588921	0.13851223	0.28907445	-0.0821499	1		
8	HeadCnt	-0.3860149	0.02879228	-0.0341622	-0.1696803	0.03357789	0.05884404	1	
9	Downtime	0.73283729	-0.1076204	-0.0902812	-0.0312905	-0.1460223	0.02934735	-0.539907	1

- 2) Seek out correlation coefficients whose abstract value is greater than 0.5, (i.e. $|corr| > 0.5$) which indicates that the corresponding two (independent) variables are highly correlated.

	A	B	C	D	E	F	G	H	I
1		Loss	Errors	ancels/Correct	Breaks	Fails	Trades	HeadCnt	Downtime
2	Loss	1							
3	Errors	-0.0199952	1						
4	Cancels/Corre	-0.0386812	-0.0268267	1					
5	Breaks	0.05571281	0.15592036	0.15050024	1				
6	Fails	-0.0596397	0.84031664	0.00187339	0.17816488	1			
7	Trades	0.01080568	-0.1588921	0.13851223	0.28907445	-0.0821499	1		
8	HeadCnt	-0.3860149	0.02879228	-0.0341622	-0.1696803	0.03357789	0.05884404	1	
9	Downtime	0.73283729	-0.1076204	-0.0902812	-0.0312905	-0.1460223	0.02934735	-0.539907	1

As shown above, $corr(Fails, Errors) \approx 0.84 > 0.5$, $corr(Downtime, Headcounts) \approx -0.54 < -0.5$, so we should select from variables “Fails” and “Errors”, and from “Downtime” and “Headcounts” to run regression on “Loss”.

3) Run regressions

Tips: To alleviate the collinearity (high correlation between independent variables), we use the method of dropping one of the two correlated variables, and run the regression with the remaining one. Here we have two groups of highly correlated variables, so there exit four types of combinations.

i. Keep “Fails” and “Downtime” (sheet 1.1)

	A	B			
1	SUMMARY OUTPUT			<i>Coefficients</i>	<i>Standard Error</i>
2				<i>t Stat</i>	<i>P-value</i>
3	<i>Regression Statistics</i>		Intercept	-235405.58	129365.122
4	Multiple R	0.73878204	Cancels/Corre	12.1260886	39.6554494
5	R Square	0.54579891	Breaks	1052.59175	952.958288
6	Adjusted R Sq	0.52308885	Fails	5.07988615	11.2008803
7	Standard Erro	194813.708	Trades	-11.673332	24.2154355
8	Observations	106	Downtime	58980.8189	5433.80395
9					

ii. Keep “Fails” and “Headcounts” (sheet 1.2)

1	SUMMARY OUTPUT			<i>Coefficients</i>	<i>Standard Error</i>
2				<i>t Stat</i>	<i>P-value</i>
3	<i>Regression Statistics</i>		Intercept	452520.484	192664.148
4	Multiple R	0.39408077	Cancels/Corre	-32.580286	53.851808
5	R Square	0.15529965	Breaks	-78.149856	1327.75593
6	Adjusted R Sq	0.11306463	Fails	-6.7491026	15.1765359
7	Standard Erro	265672.596	Trades	13.4943334	33.276425
8	Observations	106	HeadCnt	-65745.212	15888.4104
9					

iii. Keep “Errors” and “Downtime” (sheet 1.3)

	A	B			
1	SUMMARY OUTPUT			<i>Coefficients</i>	<i>Standard Error</i>
2				<i>t Stat</i>	<i>P-value</i>
3	<i>Regression Statistics</i>		Intercept	-239339.87	128664.259
4	Multiple R	0.73934421	Cancels/Corre	12.673897	39.6389391
5	R Square	0.54662986	Breaks	1016.03071	952.990006
6	Adjusted R Sq	0.52396135	Errors	116.705659	187.035489
7	Standard Erro	194635.422	Trades	-9.969303	24.5095355
8	Observations	106	Downtime	58975.8336	5402.81814
9					

iv. Keep “Errors” and “Headcounts” (sheet 1.4)

SUMMARY OUTPUT			Coefficients	Standard Error	t Stat	P-value
<i>Regression Statistics</i>						
Multiple R	0.39195567	Intercept	438839.726	192581.797	2.27871861	0.02481004
R Square	0.15362925	Cancels/Corre	-32.349081	53.9252373	-0.5998876	0.54993808
Adjusted R Sq	0.11131071	Breaks	-208.95278	1330.33811	-0.1570674	0.8755084
Standard Error	265935.152	Errors	-0.9182724	255.130941	-0.0035992	0.99713541
Observations	106	Trades	15.6722657	33.7666876	0.46413394	0.64356061
		HeadCnt	-66328.289	15904.3325	-4.1704541	6.4814E-05

2. We can **combine some variables** according to their economic meanings. For example, $\frac{\text{Trades}}{\text{Headcounts}} = \text{Workload}$. Thus, we can add in a new variable – “Workload” into the analysis.

- 1) To create the correlation matrix and find out highly correlated (independent) variables using the same method in #1 part (sheet “matrix_2”):

	A	B	C	D	E	F	G	H	I	J
1		Loss	cancels/Correct	Breaks	Errors	Fails	Trades	HeadCnt	Downtime	Workload
2	Loss	1								
3	Cancels/Corre	-0.0386812	1							
4	Breaks	0.05571281	0.15050024	1						
5	Errors	-0.0199952	-0.0268267	0.15592036	1					
6	Fails	-0.0596397	0.00187339	0.17816488	0.84031664	1				
7	Trades	0.01080568	0.13851223	0.28907445	-0.1588921	-0.0821499	1			
8	HeadCnt	-0.3860149	-0.0341622	-0.1696803	0.02879228	0.03357789	0.05884404	1		
9	Downtime	0.73283729	-0.0902812	-0.0312905	-0.1076204	-0.1460223	0.02934735	-0.539907	1	
10	Workload	0.55920468	0.05686906	0.18331256	-0.1168494	-0.1132616	0.3660611	-0.7663326	0.75583856	1

Besides the high correlation of “Fails” and “Errors”, we can observe that

$$\text{corr}(\text{Workload}, \text{Downtime}) \approx 0.756 > 0.5$$

$$\text{corr}(\text{Workload}, \text{Headcount}) \approx -0.766 < -0.5$$

$$\text{corr}(\text{Headcount}, \text{Downtime}) \approx -0.54 < -0.5$$

2) run regressions

Tips: Seemingly, we have **four** groups of highly correlated variables. But we can simplify it into **two** groups: one is to choose from “Fails” and “Errors” and the other is to choose from

“Workload”, “Downtime” and “Headcounts”. So, there are six cases to discuss.

i. Keep “Fails” and “Workload” (sheet 2.1)

SUMMARY OUTPUT			<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
		Intercept	-45408.753	150714.551	-0.3012898	0.76381914
<i>Regression Statistics</i>		Cancels/Corre	-26.538935	46.9311741	-0.5654863	0.57300967
Multiple R	0.5985186	Breaks	136.410889	1139.51904	0.11970918	0.9049537
R Square	0.35822452	Fails	-1.1304741	13.2647988	-0.0852236	0.93225401
Adjusted R Sq	0.32613574	Trades	-74.730694	30.3379618	-2.4632734	0.01547544
Standard Erro	231572.5	Workload	576.77352	78.3759405	7.35906347	5.2781E-11
Observations	106					

ii. Keep “Fails” and “Downtime” (sheet 2.2)

SUMMARY OUTPUT			<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
		Intercept	-235405.58	129365.122	-1.819699	0.0717967
<i>Regression Statistics</i>		Cancels/Corre	12.1260886	39.6554494	0.30578619	0.76040311
Multiple R	0.73878204	Breaks	1052.59175	952.958288	1.10455176	0.27200449
R Square	0.54579891	Fails	5.07988615	11.2008803	0.45352562	0.65115295
Adjusted R Sq	0.52308885	Trades	-11.673332	24.2154355	-0.4820616	0.63081559
Standard Erro	194813.708	Downtime	58980.8189	5433.80395	10.8544253	1.3366E-18
Observations	106					

iii. Keep “Fails” and “Headcounts” (sheet 2.3)

SUMMARY OUTPUT			<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
		Intercept	452520.484	192664.148	2.34875294	0.02080341
<i>Regression Statistics</i>		Cancels/Corre	-32.580286	53.851808	-0.6049989	0.54655016
Multiple R	0.39408077	Breaks	-78.149856	1327.75593	-0.0588586	0.95318219
R Square	0.15529965	Fails	-6.7491026	15.1765359	-0.4447064	0.65749297
Adjusted R Sq	0.11306463	Trades	13.4943334	33.276425	0.40552233	0.68595993
Standard Erro	265672.596	HeadCnt	-65745.212	15888.4104	-4.1379352	7.3163E-05
Observations	106					

iv. Keep “Errors” and “Workload” (sheet 2.4)

SUMMARY OUTPUT			<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
<i>Regression Statistics</i>		Intercept	-55931.248	150229.999	-0.3723041	0.71045428
Multiple R	0.59875167	Cancels/Corre	-26.161573	46.9403253	-0.5573368	0.57854271
R Square	0.35850356	Breaks	57.8259958	1139.6432	0.05074044	0.95963357
Adjusted R Sq	0.32642874	Errors	50.0515127	222.144355	0.22531076	0.82219734
Standard Error	231522.151	Trades	-73.27616	30.6584924	-2.3900771	0.01871782
Observations	106	Workload	579.009948	78.1559238	7.40839491	4.1526E-11

v. Keep “Errors” and “Downtime” (sheet 2.5)

SUMMARY OUTPUT			<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
<i>Regression Statistics</i>		Intercept	-239339.87	128664.2586	-1.8601892	0.06579762
Multiple R	0.73934421	Cancels/Corre	12.673897	39.63893908	0.31973351	0.74983739
R Square	0.54662986	Breaks	1016.03071	95299%	1.06615043	0.28892297
Adjusted R Sq	0.52396135	Errors	116.705659	187.0354886	0.62397601	0.53406419
Standard Error	194635.422	Trades	-9.969303	24.50953545	-0.406752	0.68505939
Observations	106	Downtime	58975.8336	5402.818141	10.915754	9.8227E-19

vi. Keep “Errors” and “Headcounts” (sheet 2.6)

SUMMARY OUTPUT			<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
<i>Regression Statistics</i>		Intercept	438839.726	192581.797	2.27871861	0.02481004
Multiple R	0.39195567	Cancels/Corre	-32.349081	53.9252373	-0.5998876	0.54993808
R Square	0.15362925	Breaks	-208.95278	1330.33811	-0.1570674	0.8755084
Adjusted R Sq	0.11131071	Errors	-0.9182724	255.130941	-0.0035992	0.99713541
Standard Error	265935.152	Trades	15.6722657	33.7666876	0.46413394	0.64356061
Observations	106	HeadCnt	-66328.289	15904.3325	-4.1704541	6.4814E-05